

Data-driven modeling of intermittent turbulence in the stably stratified atmospheric boundary layer

Dissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
Dr. rer. nat.

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von
Master of Science
Vyacheslav Boyko
geb. in Kharkiv

Berlin 2022

Erstgutachterin: Prof. Dr. Nikki Vercauteren
Zweitgutachter: Prof. Dr. Gabriel G. Katul
Tag der Disputation: 16. June 2022

Declaration of authorship

Name: Boyko

First name: Vyacheslav

I declare to the Freie Universität Berlin that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. The present thesis is free of plagiarism. I have marked as such all statements that are taken literally or in content from other writings. This dissertation has not been submitted in the same or similar form in any previous doctoral procedure.

I agree to have my thesis examined by a plagiarism examination software.

Date:

Signature:

Zusammenfassung

Modellierung der kleinskaligen Turbulenz spielt in den numerischen Wettervorhersagesystemen eine entscheidende Rolle, denn die Turbulenz trägt zum Transport und Durchmischung von skalaren Größen, z.B. Temperatur und Momentum, signifikant bei. Prinzipiell wird die unaufgelöste Durchmischung in einem Modell durch die aufgelösten - und somit berechneten - Fluidbewegungen mathematisch modelliert. Dabei werden Turbulenzmodelle basierend auf statistischen Eigenschaften der Turbulenz formuliert und durch Experimente kalibriert. In Abhängigkeit vom Zustand der atmosphärischen turbulenten Grenzschicht ergibt sich ein Spektrum an Möglichkeiten ein Turbulenzmodell zu kalibrieren. Eine Herausforderung an die Modellierung stellt somit eine sogenannte stabil geschichtete Grenzschicht dar. Markant für diesen Zustand ist die gekühlte Erdoberfläche, die eine sogenannte intermittierende Turbulenz, welche gegenwärtig unzureichend erforscht ist, verursacht. Ein Forschungsansatz beruht darauf, dass die intermittierende Turbulenz mit einer Reihe von unaufgelösten Fluidbewegungen, welche sub-mesoskalige Prozesse genannt werden, in Verbindung steht. Es wird dabei angenommen, dass für den stabilen Zustand die Sub-Mesoskalen unabhängig von den berechneten großen Skalen sind und somit sporadisch im Simulationsgebiet entstehen. Dabei ist das Rechengitter des Wettervorhersagesystems zu grob, um die nötigen Prozesse aufzulösen. Die sich mit dieser Problematik befassenden Studien zeigen sogar, dass selbst bei höheren Modellauflösungen die Beobachtungsdaten nur in den statistischen Größen repräsentiert werden und zurzeit nicht exakt abzubilden sind. Motivierend daraus wird in dieser Arbeit ein stochastisches Turbulenzmodell für die Atmosphärische Stabile Grenzschicht (ASG) entwickelt, um sporadischen Ereignissen, die eine intermittierende Durchmischung verursachen, Rechnung zu tragen. Somit befasst sich die vorliegende Arbeit mit der Signifikanz der Sub-Mesoskalen bei der Erzeugung der Turbulenz in der ASG. Es wird eine datengetriebene Gleichung, welche modular genug und somit leicht mit bestehenden Reynolds-gemittelten Navier-Stokes Modellen kombinierbar ist, entwickelt. Dafür umfasst die Arbeit die Analyse von vorhandenen Feldmessungen, die über mehrere Monate in einem flachen Gelände gesammelt wurden. Die Messungen werden mit einem statistischen modellbasierten Clustering-Verfahren bearbeitet. Diese Methodik erlaubt es, die instationären Zustände in der ASG, basierend auf unterschiedlichen Skaleninteraktionen, gleichzeitig zu modellieren und zu klassifizieren. Das Clustering-Verfahren wird dann zu einem Multiskalen-Parametrisierungsverfahren weiterentwickelt, das zur Identifizierung der gesuchten Gleichung aus den Daten verwendet wird. Daraus resultierend, besteht der wichtigste Beitrag dieser Arbeit in der entwickelten stochastischen Stabilitätsgleichung, die den Prozess der transienten Stabilitätskorrektur vorhersagt. Dieser Prozess beschreibt die Auswirkung der statischen Stabilität der Strömung auf die gemittelten Strömungsvariablen und berücksichtigt dabei zufällige Störungen, die hypothetisch durch die Untergitterprozesse verursacht werden. Das vorgestellte Turbulenzmodell bietet einen allgemeinen Ansatz, um den stochastisch erscheinenden Durchmischungseffekt von nicht aufgelösten und von der mittleren Skala unabhängigen Prozessen zu berücksichtigen, ohne diese explizit zu modellieren.

Abstract

Modeling small-scale turbulence is essential in numerical weather prediction systems because turbulence contributes significantly to the transport and mixing of scalar quantities, e.g., temperature and momentum. In principle, the unresolved mixing in a model is mathematically modeled based on the resolved - and thus calculated - fluid motions. Turbulence models are formulated based on statistical flow properties and calibrated with experiments. Depending on the state of the atmospheric turbulent boundary layer, there is a range of possible conditions to calibrate a turbulence model. The boundary layer is particularly challenging to model in conditions of stable stratification. The cooled surface and low wind velocity are characteristic of this condition, which induces an insufficiently explored, so-called intermittent, turbulence. One research approach is based on the observation that intermittent turbulence is associated with unresolved fluid motions, called sub-mesoscale motions. The sub-mesoscales are assumed to be independent of the large resolvable scales and thus occur sporadically in the simulation domain. The computational grid of the weather forecasting system is too coarse to resolve these necessary processes. The studies dealing with this problem show that with higher model resolutions, the observational data are only represented in the statistical quantities and cannot be predicted precisely. Motivated by this, this work develops a stochastic turbulence model for the atmospheric Stable Boundary Layer (SBL) to account for sporadic events that cause intermittent mixing. Thus, the present work addresses the importance of sub-mesoscales in generating turbulence in the SBL. A data-driven equation is developed, which is modular and easily combinable with existing Reynolds-averaged Navier-Stokes models. For this, the work analyzes existing field measurements collected over several months in even terrain. The measurements are processed using a statistical model-based clustering technique. This methodology simultaneously models and classifies the nonstationary SBL based on different scale interactions. The clustering procedure is then developed into a stochastic parameterization method used to identify an equation from the data. Resulting from this, the main contribution of this work is the developed stochastic stability equation that predicts the process of transient stability correction. This process describes the effect of the static stability of the flow on the averaged flow variables, taking into account random perturbations hypothetically caused by the sub-grid processes. The presented turbulence model provides a general approach to account for the stochastic mixing effects of unresolved and mean-scale independent processes without explicitly modeling them.

Dedicated to my father
Vitaliy Viktorovič Boyko

Acknowledgements

My special appreciation and gratitude go to my supervisor, Nikki Vercauteren. I am grateful for her guidance, encouragement, constant support, and patience; for giving me the scientific freedom to explore the frontiers of knowledge. I could not have imagined a better supervisor for my doctoral studies. I would also like to thank Gabriel G. Katul for his interest and time in serving as my committee member. I also thank my collaborator Sebastian Krumscheid for sharing his knowledge and experience in inverse stochastic modeling; his help was greatly appreciated. I would also like to thank the following people: Sergey Lesnik, Metodija Shapkalijeovski, Abhishek Harikrishnan, Cedrick Ansorge, Ivan Bašták Ďurán, Felix Oertel for stimulating discussions and everything else. I am grateful for the financial support provided by the German Research Council (DFG) through Grand No. VE 933/2-1 and VE 933/2-2, which made this research possible. I am grateful to Anna Becker for her love, support, listening, and understanding throughout the time of this work.

Table of Contents

Zusammenfassung	v
Abstract	vii
List of Figures	xvii
List of Tables	xxi
Abbreviations	xxiii
Symbols	xxv
1 Introduction	1
1.1 Numerical Weather Prediction	4
1.1.1 Stable Boundary Layer Over Flat Terrain	6
1.1.2 Sub-mesoscale Motions and Turbulence	8
1.1.3 Global Intermittency	10
1.1.4 Modeling Errors and Uncertainty	12
1.2 Turbulence Parameterizations	13
1.2.1 Deterministic Parameterization	13
1.2.2 The Gray Zone of Turbulence	14
1.2.3 Stochastic Parameterization at Sub-Grid Scale	15
1.2.4 Stochastic Parameterization of Intermittent Turbulence	16
1.3 Data-Driven Modeling of the Stable Boundary Layer	17
1.4 Summary and Research Proposition	18
1.5 Outline	20
1.6 Publications	21
2 Model-Based Nonstationary Time Series Modeling	23
2.1 Stationary Models	24
2.1.1 Linear Discrete-Time Models	24
2.1.2 Nonlinear Stochastic Differential Equations (SDE)	26
2.2 Nonparametric Clustering Methods for Nonstationary Models	28
2.2.1 Smooth Variational Clustering of SDEs	28
2.2.2 Bounded Variation of Linear Discrete-Time Models	37
2.3 Parametrization of SDEs with an External Process	38

TABLE OF CONTENTS

2.4	Estimation of Hyperparameters for the Clustering Framework	41
2.4.1	Optimal Regularization Parameter	42
2.4.2	Optimal Number of Clusters	45
2.5	Implementation	46
2.6	Numerical Studies	47
2.6.1	Nonstationary ORNSTEIN–UHLENBECK Process	48
2.6.2	SDE with Two Independent Auxiliary Processes	52
2.6.3	SDE with Nonlinear Drift and Nonlinear Diffusion	54
2.7	Discussion and Conclusion	57
3	Nonstationary Analysis of the Stable Boundary Layer	61
3.1	Fluxes Over Snow Surfaces II Dataset	62
3.2	Classifying Vertical Turbulent Velocity Fluctuations	63
3.2.1	Data Preprocessing and Scales	63
3.2.2	Model Formulation for Clustering	64
3.2.3	Estimation of the Persistence Parameter	65
3.2.4	Clustering and Modeling Results	70
3.2.5	Regime-Dependent Multiresolution Decomposition	72
3.3	Classifying Turbulence Kinetic Energy	75
3.3.1	Data Preprocessing Strategy	76
3.3.2	Definition of Scales	77
3.3.3	Filtering Methods	79
3.3.4	Scalewise Visualization of the Wind Profiles	82
3.3.5	Model Formulation for Clustering	86
3.3.6	Clustering and Modeling Results	87
3.3.7	Regime-Dependent Turbulence-Submeso-Intensity Diagram	91
3.3.8	Regime-Dependent Flux–Profile Relationship	96
3.4	Discussion and Conclusion	98
4	Stochastic Stability Equation	101
4.1	Definition of Variables and Temporal Scales	103
4.2	Modeling Strategy	105
4.2.1	Two Distinct Scale-Driven Regimes	105
4.2.2	Construction of the Model Structure	107
4.2.3	Parameter Perturbation Approach	109
4.2.4	Variation of the Model Complexity	111
4.3	Data-Driven Rejection of Model Alternatives	112
4.4	Identification of Local Parameter Scaling	114
4.4.1	Piecewise Parameterization Based on Entire Dataset	116
4.4.2	Continuous Parametrization Based on Reduced Dataset	118
4.5	Vertical Covariance Structure of the Modeled Variable	120
4.6	Model Validation	122
4.6.1	Stationary Probability Density Function	123
4.6.2	Out-Of-Sample Forecast	126

4.6.3	Temporal Autocorrelation and Self-Similarity	132
4.7	Conclusion and Discussion	134
5	Stochastically Unsteady Reynolds-Averaged Navier–Stokes Equations	137
5.1	Model Formulation	138
5.1.1	Extending a Reynolds-Averaged Navier–Stokes Model	138
5.1.2	Stochastic Hybrid Formulation of the Turbulent Mixing Length	141
5.1.3	Initial and Boundary Conditions	143
5.1.4	Inertial Oscillations	145
5.2	Numerics	146
5.2.1	Discretization with the Finite Element Method	146
5.2.2	Numerical Grid	149
5.2.3	Model Calibration	150
5.3	Numerical Analysis of the Stochastic Turbulence Model	151
5.3.1	Neutral Boundary Layer	152
5.3.2	Stably Stratified Boundary Layer	154
5.3.3	Variable Geostrophic Wind and Net Radiation	157
5.3.4	Diffusion-Induced Transitions of Temperature Inversion	164
5.4	Conclusions	168
6	Conclusion	171
6.1	Summary	171
6.2	Outlook	174
	References	177
	Appendix A	193
A.1	Coefficients of the HERMITE Expansion	194

List of Figures

1.1	Sketch illustrating different boundary layer types.	3
1.2	Measured time series of TKE in a SBL.	8
1.3	Observation of a sub-mesoscale flow structure.	9
1.4	DNS of an EKMAN flow.	12
2.1	Modeling nonstationary data using stationary SDEs.	29
2.2	An approximation of the time-dependent parameter function with piecewise constant subdivisions.	31
2.3	Construction of a self-contained prediction model by applying stochastic modeling.	36
2.4	Construction of a self-contained predictive model with the FEM- H^1 -SDE method.	42
2.5	Effect of the regularization value on the estimate of $\mathbf{\Gamma}^*(t)$	43
2.6	Estimation of the optimal regularization parameter ϵ^2 by filtering the affiliation function.	44
2.7	Data for the example with nonstationary OU Process.	48
2.8	Estimating the number of clusters for the nonstationary OU example.	50
2.9	Result of clustering the nonstationary OU process.	51
2.10	Recovered parameter scaling function of the nonstationary OU process.	51
2.11	Data for the example with nonstationary and nonlinear SDE where two parameters are modulated by two independent processes with non-overlapping time scales.	52
2.12	Estimating the number of clusters for the clustering example with two auxiliary processes.	53
2.13	Result of clustering the nonstationary and nonlinear process with two auxiliary processes.	54
2.14	Recovered parameter scaling functions from the nonstationary and nonlinear sample path with two auxiliary processes.	54
2.15	Data for the example with a nonlinear stochastic system where the bifurcation parameter is time-dependent.	55
2.16	Estimation of the number of clusters for the example with a nonlinear stochastic system where the bifurcation parameter is time-dependent.	55
2.17	Clustering results for the example with a nonlinear stochastic system where the bifurcation parameter is time-dependent.	56

LIST OF FIGURES

2.18	Recovered parameter scaling functions for the example with a nonlinear stochastic system where the bifurcation parameter is time-dependent.	57
3.1	Pictures of the measuring site from the FLOSS2 experiment.	63
3.2	Residual analysis by modeling certain variables with the FEM-BV-VARX method.	66
3.3	Analysis of the consistency of the affiliation function.	66
3.4	A sketch illustrating an approach to find an optimum using the FEM-BV-VARX method by varying the C_p value.	68
3.5	Residual value and consistency of the estimated affiliation vector plotted against the persistence parameter C_p	68
3.6	A sketch illustrating the modified algorithm for obtaining a better solution starting from the solutions at lower C_p values.	69
3.7	Analysis of the pre-run consistency.	69
3.8	Histogram of bulk Richardson numbers in the clustered flow regimes.	71
3.9	Classification and modeling of the 68 nights with the model defined in section 3.2.2, where the sub-mesoscales force the vertical fluctuations of the turbulent scales σ_w	72
3.10	Multiresolution cospectra of the heat flux.	73
3.11	Multiresolution cospectra of the vertical momentum flux.	74
3.12	Data analysis flowchart.	77
3.13	Average heat flux of the FLOSS2 dataset based on 24-hour records.	78
3.14	Illustration of the defined scales for the analysis of the FLOSS2 dataset and the filtering methods used.	78
3.15	An example of an intermittent mixing events and scale activity.	81
3.16	Example demonstrating the filtering results of the \hat{u} component with the DWT.	82
3.17	Scale activity and turbulence response for a night with moderate wind.	83
3.18	Scale activity and turbulence response for a night with weak wind.	84
3.19	Scale activity and turbulence response for a night with almost no mean wind.	85
3.20	Coefficient of determination over regime persistence parameter C_p	88
3.21	In-sample prediction of TI by the estimated model	89
3.22	Stack plots of the parameter values for the forcing terms.	90
3.23	TSI diagram for three heights.	92
3.24	Classified TSI diagram by limiting ζ value.	93
3.25	Classified TSI diagram with FEM-BV-VARX method.	94
3.26	Comparison of the ζ classifier with the FEM-BV-VARX classifier for the strongly stable regime only.	95
3.27	A sketch of a TSI diagram illustrating the intermittent turbulence states.	95
3.28	Classified flux–profile relationship with FEM-BV-VARX method.	97
4.1	A sketch of the scales and associated variables defined for the FLOSS2 dataset to identify the SSE.	103
4.2	Comparison of several stability correction functions ϕ_f with the steady state solutions of the proposed ODE.	106

4.3	Bifurcation diagrams for the considered ODEs displaying the steady-state solutions.	108
4.4	Association between the variability of the Ri number and the non-dimensional sub-mesoscale variability for the scales between 3 m - 1 h.	110
4.5	Selection of the best fitting model from a considered pool of models based on the model fitness function.	114
4.6	Estimated optimal number of clusters for each of the considered model types.	115
4.7	Local scaling of the time-dependent model parameters $\lambda(t), \nu(t), \sigma_p(t)$ for the SSE and their respective piecewise parametrization functions identified utilizing entire FLOSS2 dataset for training.	117
4.8	Local scaling of the time-dependent model parameters $\lambda(t), \nu(t), \sigma_p(t)$ for the SSE and their respective continuous parametrization functions.	119
4.9	Estimation of the regime-dependent vertical correlation function $\rho(z, l_z)$ for the process ϕ	122
4.10	Stationary PDF for the variable ϕ calculated from SSE parameterized with the piecewise functions.	124
4.11	Stationary PDF for the variable ϕ calculated from SSE parameterized with continuous functions.	126
4.12	Comparison of in-sample prediction using the piecewise parameterized SSE with MATÉRN covariance for the vertical dimension.	129
4.13	Comparing the profile evolution of the modeled ϕ against the validation dataset.	131
4.14	Quantile-wise comparison of height-averaged PDFs of the model within selected periods in Figure 4.13.	132
4.15	Analysis of the temporal autocorrelation function.	134
5.1	Statistical analysis of the stochastic turbulent mixing length.	142
5.2	Coupling strength of the deterministic and stochastic equations of the SURANS model solved with a segregated approach.	148
5.3	The computation of the hybrid stochastic mixed length correction variable φ using two different grids.	150
5.4	Comparison of the predicted TKE by the SURANS and RANS models in the condition of neutral stratification ($Ri = 0$) for three heights ($z = 0.5, 70, 150$ m).	154
5.5	Comparison of the TKE predicted by the SURANS and RANS models in the condition of strongly stable stratification (bulk $Ri \approx 0.6$) for the height $z = 20$ m.	156
5.6	Temporal evolution of the profiles of TKE for a realization of the SURANS and RANS models.	157
5.7	Temporal evolution of the temperature profiles for a realization of the SURANS and RANS models.	157
5.8	Temporal evolution of the wind profiles (u component) for one realization of the SURANS and RANS models.	158
5.9	Temporal evolution of the Ri number profile for a realization of the models SURANS and RANS.	158
5.10	Solution of the SURANS model with variable forcing parameters R_n (net radiation) and the geostrophic wind u_g	159

LIST OF FIGURES

5.11	The relative difference in profiles between the SURANS (the central tendency of 100 realizations) and the RANS model according to the numerical study in Fig. 5.10.	162
5.12	The relative difference between noise-free limit SURANS ($\sigma_s \approx -1$) and the noise present SURANS models (central tendency of 100 realizations) simulation.	163
5.13	Modeled relationship between 50-m temperature inversion and wind speed under the influence of assumed sub-mesoscale turbulent diffusion.	166
A.1	Estimation of the number of clusters K for different heights.	193

List of Tables

3.1	Definition of the discrete-wavelet-filter-bank cascade.	81
3.2	R^2 values for each identified regime for each height.	88
3.3	Statistics of stability parameter ζ for each identified regime and height.	92
4.1	Optimal set of hyperparameters $(K_{\text{opt}}, \epsilon_{\text{opt}}^2)$ for the SSE.	118
4.2	Parametrization of the SSE's parameters with continuous scaling functions. . .	119
4.3	Regime dependent estimates of a Gaussian model for the vertical ACF.	122
5.1	Summary of the settings of the SURANS solver.	152
5.2	Relevant solver settings for the numerical study of the neutral stratified boundary layer.	153
5.3	Relevant solver settings for the numerical study of the stably stratified boundary layer.	155
5.4	Relevant solver settings for the numerical study with unsteady forcing variables.	160
5.5	Relevant solver settings for numerical study of the temperature inversion. . . .	165
A.1	Model library tested on data.	195

Abbreviations

- ABL** Atmospheric Boundary Layer 1, 171
- ACF** Autocorrelation Function 121
- ARMAX** AutoRegressive Moving Average model with eXogenous variables 24
- BV** Bounded Variation 37
- c.t.d.f.** conditional transition density function 27
- CBL** Convective Boundary Layer 5
- DNS** Direct Numerical Simulation 4
- DWT** Discrete Wavelet Transform 44
- ECMWF** European Centre for Medium-Range Weather Forecasts 14
- FEM** Finite Element Method 29
- FEM- H^1** Finite-Element, H^1 -Regularization 29
- FEM- H^1 -SDE** Finite-Element, H^1 -Regularization, Stochastic Differential Equation 37, 172
- FEM-BV-VARX** Finite-Element, Bounded-Variation, Vector AutoRegressive (X) Factor 37, 172
- FLOSS2** Fluxes Over Snow Surfaces Phase II 30, 171
- GPU** Graphics Processing Unit 34
- i.i.d.** independent and identically distributed 25
- IFS** Integrated Forecasting System 14
- KDE** Kernel Density Estimation 125
- LES** Large Eddy Simulation 4
- MetOffice** Met Office institution in the United Kingdom 106
- MLE** Maximum Likelihood Estimation 27, 172
- MOST** MONIN–OBUKHOV Similarity Theory 6, 173
- MRD** Multiresolution Decomposition 72, 175
- MSC** Meteorological Service of Canada 106
- NWP** Numerical Weather Prediction 4, 171
- ODE** Ordinary Differential Equation 105, 193
- ODT** One-Dimensional Turbulence model 16
- OU** ORNSTEIN–UHLENBECK 47

Abbreviations

- PDE** Partial Differential Equation 146
- PDF** Probability Density Function 56, 173
- PSD** Power Spectral Density 134
- PSP** Physically based Stochastic Perturbation 15
- QP** Quadratic Programming 34
- RANS** REYNOLDS-averaged NAVIER-STOKES 4, 171
- SBL** Stable Boundary Layer 5, 171
- SDE** Stochastic Differential Equation 16, 172
- SI** Submeso Intensity 82
- SPPT** Stochastically Perturbed Parameterization Tendency 15
- SSE** Stochastic Stability Equation 105, 173
- SURANS** Stochastically Unsteady REYNOLDS-averaged NAVIER-STOKES Equations 137
- TI** Turbulence Intensity 82
- TKE** Turbulence Kinetic Energy 7, 171
- TPE** Turbulence Potential Energy 13
- TSI** Turbulence-Submeso-Intensity 76, 172

Symbols

Symbol	Description	Unit
$\mathbf{B}(\tau)$	Vector of modeling errors	–
C_g	Thermal capacity of the soil	$\text{J K}^{-1} \text{m}^{-3}$
C_p	Hyperparameter restricting the maximum number of jumps in the affiliation vector of the FEM-BV-VARX framework	–
D	Diagonal matrix with eigenvalues	–
E_γ	Signal energy of the affiliation vector	–
E_q	Function describing a stable equilibrium solution of an ODE	–
F	Fourier matrix	–
$\mathbf{\Gamma}$	Affiliation vector	–
H_0	Turbulent sensible heat flux at the surface	$\text{J m}^{-2} \text{s}^{-1}$
H_c	Starting height of the atmospheric lapse rate	m
H_j	HERMITE polynomials	–
\mathbf{H}_k	Stiffness matrix of a cluster	–
H	Height of the computational domain	m
$I_{A_j^{(k)}}$	Clustered indicator functions	–
I_{A_j}	Indicator functions	–
$\Im \mathbf{m}$	Imaginary part of a complex number	–
I	Measure of global intermittency for the SBL	–
J	Log-likelihood value for a given model	–
K_h	Eddy heat conductivity coefficient	$\text{m}^2 \text{s}^{-1}$
K_m	Kinematic eddy-viscosity coefficient	$\text{m}^2 \text{s}^{-1}$
K_{opt}	Optimal number of cluster	–
K	Number of cluster	–
$L_{\widehat{N}}^\epsilon$	Regularized averaged clustering functional on a coarser grid	–
L_N^ϵ	Regularized averaged clustering functional	–
L_N	Averaged clustering functional	–
Λ	Scaling function for the suppression rate in the SSE	–
L	MONIN–OBUKHOV length	m
N_u	Relaxation term towards the geostrophic wind component in x direction	m s^{-2}

Symbols

Symbol	Description	Unit
N_v	Relaxation term towards the geostrophic wind component in y direction	m s^{-2}
N	Number of samples in a time series	–
Pr_{t,z_0}	Turbulent PRANDTL number at the surface	–
Pr_t	Turbulent PRANDTL number	–
Q	Normalization constant	–
R^2	Coefficient of determination	–
R_n	Net radiation	W m^{-2}
\Re	Real part of a complex number	–
Re	REYNOLDS number	–
P	Circulant covariance matrix	–
\widetilde{Ri}_{1h}	Fluctuation intensity of the log-transformed Ri number confined between 1 h and 3 min	–
\widetilde{Ri}	Fluctuation intensity of the Ri number confined between 1 h and 3 min	–
\overline{Ri}	Gradient RICHARDSON number at a resolved mean scale of 1 h	–
\widehat{Ri}	Gradient RICHARDSON number at a scale of 3 min	–
Ri_b	Bulk RICHARDSON number	–
Ri_c	Critical RICHARDSON number (0.25)	–
Ri_k	RICHARDSON number conditioned on the classified regime k with the SSE	–
Ri	RICHARDSON number	–
\mathcal{H}	Shape factor	–
Σ	Scaling function for the perturbation intensity in the SSE	–
T_0	Reference potential temperature	K
T_g	Temperature directly at the surface	K
T_s	Soil temperature below the surface at some finite depth	K
\mathcal{T}_j	Time intervals of a reduce grid	s
Θ	Time-varying parameter vector	–
$\overline{\Theta}^*$	Estimate of the cluster-averaged parameter vector	–
$\overline{\Theta}$	Cluster-averaged parameter vector	–
U_s	Submeso horizontal motions	m s^{-1}
Γ	Atmospheric lapse rate	K m^{-1}
\mathcal{V}	Scaling function for the production rate in the SSE	–
\overline{U}	Mean horizontal wind speed	m s^{-1}
W_K^*	Diversity measure of clustering for the zero-reference dataset	–
W_K	Diversity measure of clustering	–
X, Y, Z	Variables for stochastic processes	–
a_i, b_i, c_i, d_i	Model parameters to be estimated	–
α_N	Grid size reduction factor	–

Symbol	Description	Unit
α_ε	Modeling constant of the turbulent dissipation rate	–
α	Modeling constant of the turbulent kinematic eddy-viscosity coefficient	–
\mathbf{A}	Total stiffness matrix of a variational problem	–
b	Discretized model fitness function	–
β	Exponent of a nonlinear term	–
b	Parameter of a nonlinear term in the SSE	–
c_P	Specific heat of air at constant pressure	$\text{J kg}^{-1} \text{K}^{-1}$
$\Delta\tau$	Time step on a reduced grid	s
δ_1	Displacement thickness	m
δ_2	Momentum thickness	m
δ	Denotes a test function	–
d_{ij}	Weighted diversity matrix	–
Δt	Time step	s
ε_t	Random process that mirrors the effect of random turbulent mixing	–
η_j	Coefficients of the HERMITE expansion	–
\mathbb{E}	Expected value	–
e	Turbulence kinetic energy	$\text{m}^2 \text{s}^{-2}$
f	Temporarily used function	–
f_b	Band limit of a signal	s^{-1}
f_c	Coriolis parameter	rads s^{-1}
f_s	Sampling frequency	s^{-1}
\hat{u}_*	Friction velocity computed with 3 min averaging window	m s^{-1}
f	Drift term of a SDE	–
g	Acceleration of gravity	–
γ_k^*	Estimate of the cluster/model affiliation function	–
$\mathbf{\Gamma}^*$	Estimate of the cluster affiliation vector	–
γ	Cluster/model affiliation function	–
g	Diffusion term of a SDE	–
h_b	Boundary layer height	m
i, j, n, k	Indices	–
k_s	Steepness of the parameter of the sigmoid function	–
κ_m	Soil heat transfer coefficient	$\text{W m}^{-2} \text{K}^{-1}$
κ	Von Kármán's constant	–
l_N	Fitness function	–
ℓ	Time-dependent fitness function	–
l_m	Turbulent mixing length for momentum	m
l_z	Correlation length parameter of the variable ϕ in height	–
λ_b	Blackadar's extension modeling parameter	–
λ_i	Parameters of the scaling function Λ	–

Symbols

Symbol	Description	Unit
λ	Suppression rate of the turbulent mixing length	s^{-1}
\mathcal{L}	LAMPERTI transform	–
\mathbb{M}	Most probable value (mode)	–
μ	Drift term of the process Y	
n	Discrete-time coordinate index	s
ν_j	Weights calculated from the affiliation vector	–
ν	Kinematic viscosity of a fluid	$\text{m}^2 \text{s}^{-1}$
ω	Angular velocity of the Earth	rads s^{-1}
p_X	Conditional transition density function of the process X	–
p_Z	Conditional transition density function of the process Z	–
p_ϕ	Transition probability density function of a particular model for the process ϕ	–
p_s	Stationary probability density function for a SDE	–
Ω_θ	Acceptable parameter space	–
ϕ_f	Momentum stability correction function following the MOST	–
ϕ	Stochastic process representing momentum stability correction at the integral scale	–
p_k^*	Stationary distribution function of a SDE sub-model	–
p	Memory depth of the autoregressive term	–
q	Memory depth of the exogenous term	–
ϵ_{opt}^2	Optimal regularization parameter	–
ϵ^2	Regularization parameter	–
$R_{\phi\phi}$	Temporal autocorrelation function	–
ρ_a	Air density	–
ρ	Autocorrelation function for the height	–
r	Memory depth of the noise term	–
σ_p	Intensity of a random process in a parameter perturbation ansatz	$\sqrt{\text{h}}$
$\tilde{\sigma}$	Variability of the submesoscale of motions confined between 1 h and 3 min	–
σ_i	Parameters of the scaling function Σ	–
σ_s	Stratification independent noise intensity in the SSE	–
sig	Logistic sigmoid function	–
s	Coordinate of a grid where the SSE is solved	–
t_e	End value of a time interval	s
t_*	$t_* = t/t_{\text{night}}$	–
τ_h	Constant transforming the time units in the SSE	–
τ_{macro}	ODE's macroscopic time scale of evolution	s
τ_{perturb}	Perturbation time scale of sub-mesoscale motions	s
τ_r	Relaxation time scale in a RANS model	s
τ	Time coordinate of a reduced grid	s

Symbol	Description	Unit
\hat{V}	Test space	–
$\bar{\theta}_k^*$	Estimate of a cluster-averaged parameter vector	–
θ_i	Time-varying parameter	
$\bar{\theta}_k$	Cluster-averaged locally valid parameter set	
ϑ^*	Estimate of a stationary parameter vector	
ϑ	Stationary parameter vector	
\tilde{u}	Streamwise velocity component of the sub-mesoscale motions in a stream-lined coordinate system	m s^{-1}
\tilde{v}	Cross-streamwise velocity component of the sub-mesoscale motions in a stream-lined coordinate system	m s^{-1}
\top	Transpose	–
\tilde{V}	Test space	–
t	Time coordinate	s
u_g	Geostrophic wind component in x direction	$\text{m}^2 \text{s}^{-1}$
u_∞	Free stream velocity	m s^{-1}
$\tilde{\mathbf{u}}$	Dataset specific jet scales ($\approx 1 \text{ h} - 3 \text{ h}$)	s
\bar{u}_k	Cluster respective averaged value of the auxiliary process $u(t)$	
$u_{*,z0}$	Boundary condition for the friction velocity at the surface	m s^{-1}
$\tilde{\mathbf{u}}$	Dataset specific sub-mesoscales ($\approx 3 \text{ min} - 60 \text{ min}$)	s
$\hat{\mathbf{u}}$	Estimate of a turbulent scale for the nocturnal FLOSS2 dataset ($\approx 1 - 5 \text{ min}$)	s
$u(t)$	Auxiliary process	
α	Limiting value for a PDF	–
γ	Semivariogram	–
λ	Eigenvalue	–
μ	A measure	
ν	Eigenvector	–
ϕ	Weight function	–
v_i	Parameters of the scaling function \mathcal{V}	–
v	Production rate of the turbulent mixing length	s^{-1}
φ	Latitude	grad
u	Streamwise velocity component in the stream-lined coordinate system	m s^{-1}
v_g	Geostrophic wind component in y direction	$\text{m}^2 \text{s}^{-1}$
σ_{MLE}^2	Variance of the MLE	–
φ	Stochastic process as a convex combination of the process ϕ and the function ϕ_f	–
\mathbb{V}	Variance	
ϑ_i	Stationary parameter of some SDE	
v	Cross-streamwise velocity component in the stream-lined coordinate system	m s^{-1}

Symbols

Symbol	Description	Unit
\widehat{N}	Number of grid points on a reduced grid	–
$W(t)$	WIENER process	–
w	Vertical velocity component in the stream-lined coordinate system	m s^{-1}
ξ	i.i.d. process	–
z_{0h}	Roughness length for heat	m
z_0	Grid point directly at the surface	–
z_p	Limiting height of the stochastic fade-out layer	m
z_*	$z_* = z/z_{\text{tower}}$	–
z_s	Limiting height for the SSE	m
ζ	Stability parameter	–
z	Height coordinate	m

1

Introduction

The Atmospheric Boundary Layer (ABL) is a fundamental element in agricultural and aviation meteorology, hydrology, and is particularly important for climate and weather forecasting. Accurate mathematical modeling of the ABL is critical for controlling and managing air quality and predicting current and future climate threats. This thesis introduces a novel turbulence model that improves the representation of the surface ABL.

The nature and complexity of the ABL can be intuitively explained through one of canonical examples from fluid dynamics, which is denoted as plate turbulence. Imagine air flowing over a horizontal and smooth plate (see Fig. 1.1a). At a certain distance from the surface of the plate, the air flows uniformly and at a certain velocity u_∞ . Nevertheless, the velocity of the flow directly at the surface is zero. The boundary layer is the velocity vector profile \mathbf{u} that develops from the surface upward to a certain finite height h_b .

The boundary layer becomes turbulent at high flow velocities u_∞ due to the significant vertical gradient of the velocity profile. The turbulent flow condition is characterized by a large REYNOLDS number

$$Re = \frac{h_b u_\infty}{\nu}, \quad (1.1)$$

where ν is the kinematic viscosity of a fluid. Turbulence can be severely idealized as a chaotic multitude of vortices. These eddies populate the entire boundary layer and shrink the closer they are to the surface. As the air continues to flow across the plate the entire flow field reaches a state of equilibrium. In such a quasi-stationary state the mean velocity of the flow is constant, but instantaneous fluid motions remain their chaotic nature on smaller scales. To study the flow, the velocity vector is typically decomposed by distinguishing between the mean velocity $\bar{\mathbf{u}}$ and the velocity of the turbulent fluctuations \mathbf{u}' , the latter being described statistically. Such an approach is referred to as a REYNOLDS decomposition (Stull, 1988, Sec. 2.3) and is applied spatially or temporally:

$$\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}', \quad (1.2)$$

where $\mathbf{u} = (u, v, w)$ is the velocity vector field. Thereby, one of the assumptions is the *scale separation*, which allows one to describe the influence of the turbulent eddies on the mean velocity profile. The assumption of scale separation implies that there is a spectral gap between the forcing scale u_∞ and the fluctuations \mathbf{u}' . The description of the effects of turbulent motions on the mean velocity of the flow is the goal of turbulence modeling. Similarly, the boundary layer turbulence is an essential feature of any surface-bounded flow and is necessary to capture for the representation of the ABL (Stewart, 1979).

Compared to the plate turbulence (see Fig. 1.1a), the ABL is an overwhelmingly complex and highly turbulent system whose Re number is on the order of $10^7 - 10^8$ (see Fig. 1.1b,c). Half of the kinetic energy loss of the atmosphere occurs in the ABL (Palmén & Newton, 1969). Heat, moisture, momentum, and pollutants are transported at a high rate in the vertical direction due to turbulence. The height of the boundary layer is primarily related to the size of the turbulent eddies. Depending on the conditions, the turbulent eddies vary significantly in size, from a few meters to kilometers (Holton, 2004; Honnert et al., 2020). For example, the large Re number indicates that the turbulent eddies occupy a significant spectral range, hence there is room for other processes which can interfere with turbulent eddies. The interference of different physical processes with turbulence violates the assumption of scale separation. The scale separation is necessary to use REYNOLDS decomposition for modeling turbulence and is only valid under certain conditions (Kang & Won, 2016; Larsén et al., 2016; Muschinski et al., 2004). The boundary layer turbulence is not the only characteristic feature of the ABL.

One of the prominent simplified representation of the ABL is the EKMAN layer model. This model balances the geostrophic winds, Coriolis force and the parameterized turbulent transport. As a result the EKMAN layer model predicts a spiral structure for the boundary layer. The effect of the EKMAN spiral is typically negligible when turbulence is strong (especially at the surface) as it is the case for the daytime ABL. Nevertheless, at low winds and in the nighttime the EKMAN effect shows an influence on the boundary layer (Holton, 2004, Sec. 5.3.4).

In addition to the EKMAN effect, the ABL is characterized by numerous local flow conditions created by irregular ground surface or highly variable tree canopy and urban terrain (see Fig. 1.1b,c). The resulting flow field of the ABL is strongly non-uniform, making experimental field measurements unique and challenging to reproduce in the laboratory. Experimental studies of the ABL are as well hampered by several environmental conditions. For instance, the ABL is driven by the complex nonstationary flow in the upper atmosphere and thermodynamic processes within. The solar radiation and the resulting diurnal cycle are crucial for determining the ABL and its turbulent features. The cloud cover and the shadows cast by it, soil properties, water content with its phase transitions significantly influence the shape of the boundary layer profile and affect the anthropogenic and natural environment. However, the focus of this work lies on the turbulence in the ABL.

Turbulence is one of the most complex problems and has been studied since the middle of the 19th century when NAVIER and STOKES formulated the equations of hydrodynamics. It is important to consider the simplified fluid-dynamic experiment of plate turbulence presented in Fig. 1.1a, as it is an excellent illustrative example for the study of nonstationary ABL. From such experiments we know, for example, that turbulence is inherently a three-dimensional phenomenon and has a characteristic statistical properties under idealized conditions.

The fundamental statistical properties of turbulence were developed by KOLMOGOROV and are still a cornerstone of turbulence modeling today. One prominent illustration is that the turbulent vortices form a self-similar cascade by continuously breaking apart. The stretching of the vortices and the self-intensification of the strain field causes the breakdown of the large vortices into smaller ones (Bos, 2021; Carbone & Bragg, 2020). This cascade has its well-known slope of $-5/3$ in the power spectrum diagram. The scale range spanned in the spectrum by this cascade is called the inertial subrange. Therefore, vortices associated with the inertial subrange are considered turbulent and referred to as KOLMOGOROV-like turbulence. The entire flow field must be in the so-called statistical equilibrium to observe KOLMOGOROV turbulence. However, the ABL is seldom in this state of balance since the mechanical and thermodynamic forcing conditions are constantly evolving. Therefore, the turbulence is continuously trying to establish a state of equilibrium with the mean flow which results in continuously evolving boundary layer state.

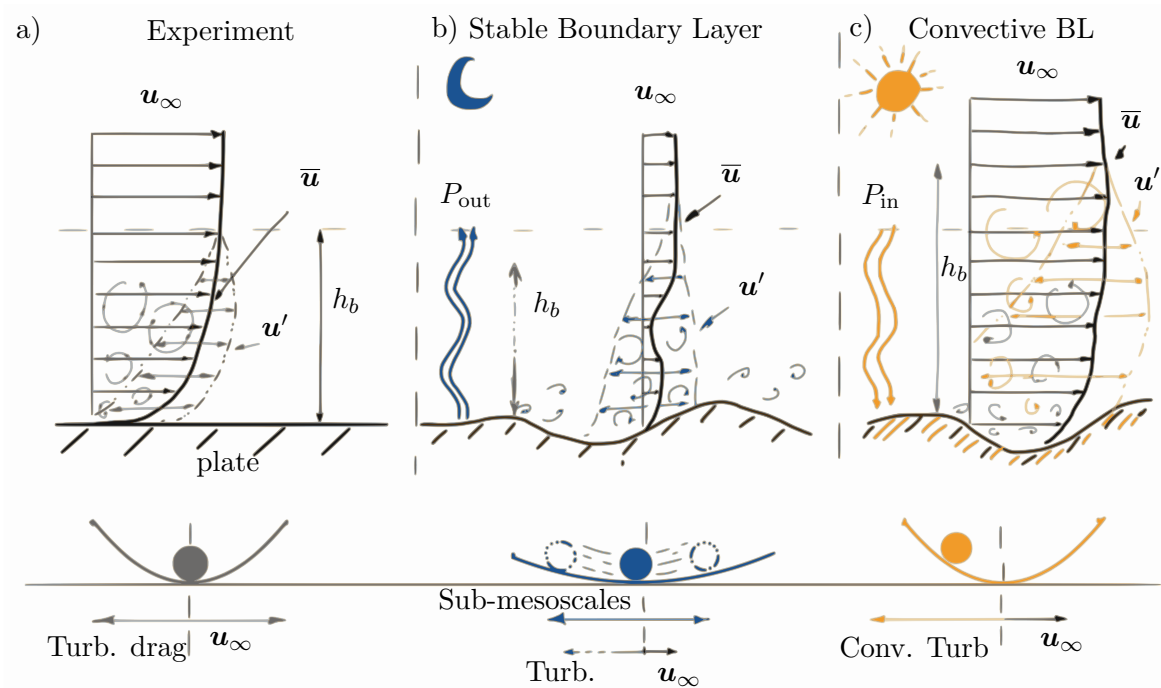


Figure 1.1: Simplified sketch illustrates different types of boundary layers. a) An idealized experimental setup where the fluid flows over a smooth plate inducing mechanical turbulence. An equilibrium is established between the forcing flow u_∞ and the induced turbulent drag. b) Stable ABLs occur, for example, at night when the longwave radiation P_{out} cools the surface suppressing thereby the turbulence. In the low-wind condition, the sub-mesoscales (see Sec. 1.1.2) dominate the flow. The generation of turbulence under such conditions is currently under research. c) A convective atmospheric boundary layer is formed during the day when the surface is heated by solar radiation P_{in} . Convective turbulence induced by buoyant instabilities causes the boundary layer to grow. This work aims to model the conditions depicted in panel b).

Turbulent eddies associated with the inertial subrange have the property of isotropization. Therefore, different vortex structures are found at the beginning (large wave number) of the cascade than at the end of the cascade. The large vortices are more disorganized and anisotropic, and the small vortices are typically isotropic. Stiperski et al. (2021) report on a preferable and distinct mechanism of isotropization in the atmospheric turbulence and proposes a mathematical description of the trajectories in a two-dimensional barycentric mapping of

the six-dimensional REYNOLDS stress tensor. Such dimensionally-reduced descriptions are especially helpful for construction of turbulence models. In the inertial subrange, statistically, the turbulence organizes large, stretched flow structures into small and solitary vortices, which are then converted into heat. The statistical description of turbulence is a simplification, and still a well-established modeling approach, but there are possible structure-based approaches in research (Marusic & Monty, 2019). Looking at the interior of the boundary layer locally in space and time, one finds a variety of different vortex configurations (Bermejo-Moreno & Pullin, 2008; Harikrishnan et al., 2021; Robinson, 1991). These turbulent structures are challenging to describe mathematically (Lindheim et al., 2021), especially in the ABL, where observational data are sparse and often limited to point measurements.

Turbulence is one of the critical phenomena that constrains our ability to predict weather and climate. To anticipate the average state variables like wind velocity, temperature, humidity accurately, one requires to resolve the entire inertial subrange. As a practical matter, it is computationally infeasible at this point to resolve the entire turbulent cascade in a realistic ABL setting (Stoll et al., 2020; Voller & Porté-Agel, 2002). Instead, efficient turbulence models are used, which, for example, exploit the statistical properties of turbulence (Meneveau, 2012; Pope, 2000).

Depending on available computational resources and how detailed the simulation of turbulent eddies needs to be, one can choose different strategies. Direct Numerical Simulation (DNS) resolves all possible vortices, making it the most expensive and accurate method. This method is mainly used in the study of turbulent flows. Then there are methods that resolve only the dominant turbulent eddies and model the small eddies. Such approaches are called Large Eddy Simulation (LES) and based on filtering the flow equations. The LES is used in both research and application. The most inexpensive models are based on the idea of modeling all possible turbulent eddies. They are based on the REYNOLDS decomposition and are therefore known as REYNOLDS-averaged NAVIER–STOKES (RANS) models (also called closure models). An atmospheric RANS model is utilized in this work and a detailed description is given in Ch. 5. Typically, such mathematical models parameterize the mixing properties of the turbulence within a simulation cell. The computing power limits the smallest possible size of the cell. A comprehensive simulation domain of an ABL may have millions of these cells, and nevertheless, there will be unresolved turbulent motions that require subgrid-scale modeling.

1.1 Numerical Weather Prediction

The ABL is a critical component of the Numerical Weather Prediction (NWP) models. The NWP system calculates the state of atmosphere at various temporal and spatial scales by solving mathematical models with supercomputers. A real-time weather forecast combines several complex numerical methods and delivers one final product. First, satellites and weather stations collect meteorological observations. Second, in the so-called assimilation (Lorenc et al., 2015) step, the data is processed and integrated into the solver as initial conditions. Third, the fluid-dynamic solver simulates the state of atmospheric variables for the future time on a numerical grid with a particular resolution. For the unresolved processes, simplified

mathematical models are developed (also called parameterizations), calibrated and executed for the range of a prediction. In the final step, the simulation results are further processed by applying a statistical downscaling procedure (Tang et al., 2016). This allows the prediction of surface temperature or precipitation on a much finer resolution than the simulation.

The NAVIER–STOKES equations and the mass continuity equations are the fluid-dynamic core of the NWP models. Additional equations describing the thermodynamic processes are tightly coupled to the dynamics of air. All equations are discretized and numerically solved on a computational grid to predict the wind velocity, etc. The grid resolution is limited and can only be increased by reducing the size of the entire simulation area or by increasing the number of cores in a highly parallel algorithm. For example, Ullrich & Jablonowski (2012) provide a detailed description of the equations and numerical methods used to solve them, together with mesh descriptions and simulation examples.

The NWP models are utilized for short- and long-term forecasts, the latter of which is considered in the assessment of climate. The short-range NWP models are currently run operationally on grid cells ranging in size from 3 to 9 km, depending on the simulation area (Beljaars et al., 2018; Lee et al., 2019b). For comparison, in the years 1990–2004, the grid cell size of the Met Office models was 60 km for the global forecast system and 12 km for the British model (Golding et al., 2004). The performance of NWP models has steadily improved over the past decades. The advances in predictive capabilities result primarily from the scientific knowledge and technological advances in high-performance computing (Bauer et al., 2015).

Advances in high-performance computing technology have also enabled the development of scalable machine learning approaches. Schultz et al. (2021) report on recent ideas in combining operational NWP workflows with machine learning approaches. Schultz et al. (2021) suggest to replace the entire chain of assimilation, prediction, and statistical downscaling by a direct mapping from observation to prediction using a deep neural network.

Nevertheless, the parameterization of unresolved processes at multiple scales and understanding the related physical mechanisms in the ABL is an active area of research. With the growing computing resources, it becomes possible to use more accurate numerical simulation tools. The parameterization of turbulence is usually performed with RANS models. On the other hand, the LES approach offers the possibility of accurately computing the large turbulent eddies and modeling only the small vortices, thereby increasing prediction accuracy. The LES models have the potential to be used in the NWP at regional scales in the future (Schalkwijk et al., 2015; Stevens et al., 2020), although this is still a far reach goal.

The ABL and its spectrum of physical processes are difficult to simulate and predict due to turbulence. The two opposing turbulence conditions in the ABL are encountered in the diurnal cycle. During the day, the surface is being heated up, and the air begins to rise (see Fig. 1.1c). The convectively driven air masses contribute to turbulent mixing and generate unstable or Convective Boundary Layer (CBL). In contrast, over ice and on cloudless nights, the surface is cooled by the radiative effect, which leads to the formation of the Stable Boundary Layer (SBL) (Fernando & Weil, 2010) (see Fig. 1.1b). Under such conditions, the turbulent eddies have to counteract buoyant forces due to the stratification and thus transport the colder and heavier air from below to the higher layers. This additional work against the rotation of the turbulent eddies weakens mixing and thus also the averaged transport properties of the turbulence. This

typically results in a boundary layer that is much shallower than that during the day. The representation of SBL in NWP applications is an active research area and remains a challenge (Holtslag et al., 2013). This work focuses on improving the parametrization of the strongly SBL over flat terrain.

1.1.1 Stable Boundary Layer Over Flat Terrain

The preferred method in modeling SBL is the RANS model. Derived from the NAVIER–STOKES equations under multiple assumptions (see Sec. 5.1.1 for more details) there exist a multitude of formulations with different complexity and number of equations involved (Stull, 1988). The MONIN–OBUKHOV Similarity Theory (MOST) is the established approach that is still used today as a component of the RANS models to account for stratification effects (Monin & Obukhov, 1954) (for a historical review see (Foken, 2006)). Derived with the help of dimensional analysis and under certain assumptions (Garratt, 1994), the theory states that the dimensionless mean wind shear $\partial\bar{u}/\partial z$ and gradient of the mean potential temperature $\partial\bar{T}/\partial z$ in the surface layer is a function of only the stability parameter ζ .

$$\frac{\kappa z}{u_*} \frac{\partial\bar{u}}{\partial z} = \phi_m(\zeta), \quad \frac{\kappa z}{T_*} \frac{\partial\bar{T}}{\partial z} = \phi_h(\zeta), \quad (1.3)$$

where z is the height over the ground, κ the VON KÁRMÁN constant, u_* is the friction velocity, $T_* = -(\overline{w'T'})_0/u_*$, and $\zeta = z/L$ is a non-dimensional variable in a range $\zeta > 0$ for stable conditions. The height above the surface z is divided by the MONIN–OBUKHOV length L , which combines the turbulent fluxes for momentum and heat:

$$L = -\frac{u_*/\kappa}{(g/T_0)(\overline{w'T'})_0}, \quad (1.4)$$

where $(\overline{w'T'})_0$ is the turbulent heat flux at the surface (denoted with 0) and T_0 a reference potential temperature. Stability functions $\phi_m(\zeta)$, $\phi_h(\zeta)$ are applied to take the effect of stratification on the vertical profiles into account. The forms of the stability functions are determined empirically from observations, and their simplest definition is, for example (Högström, 1996):

$$\phi_m = 1 + 5.3\zeta, \quad \phi_h = 1 + 8\zeta. \quad (1.5)$$

Several authors propose various similar functional forms for geographically different locations. These functions play an important role in the design of turbulence models for SBL. The topic of their effectiveness is extensively studied in the literature (for an overview of the available functions, see Gryanik et al. (2020)). The MOST is a theory of the surface layer and describes the boundary layer as a single entity. Assuming that the turbulent eddies are uniformly distributed in space, the time-averaged quantities at the surface describe the entire boundary layer. Nieuwstadt (1984) shows that by defining a local (surface-independent) Obukhov length, the MOST, is also valid within the boundary layer and causes a local scaling of the turbulent mixing.

The SBL is also characterized by the dimensionless gradient RICHARDSON number, which is another widely used stability parameter:

$$Ri = \frac{\frac{g}{T_0} \frac{\partial \bar{T}}{\partial z}}{\left(\frac{\partial \bar{u}}{\partial z}\right)^2 + \left(\frac{\partial \bar{v}}{\partial z}\right)^2}, \quad (1.6)$$

where g is the gravitational acceleration. Compared to the ζ parameter, the Ri number is determined from the averaged variables resolved by the model. In contrast, the parameter ζ is a function of the turbulent sub-grid fluxes. Nieuwstadt (1984) shows in his Fig. 3 that the two stability parameters are related as follows:

$$Ri = \zeta \frac{\phi_h}{\phi_m^2}, \quad (1.7)$$

and there is a functional mapping $\zeta \mapsto Ri$ with the asymptotic value $Ri \approx 0.25$. Therefore, the stability functions ϕ_m, ϕ_h can be parameterized with the Ri number and are also used in turbulence modeling (Cuxart et al., 2006). Regardless of the functional form or stability parameter that is used to model the SBL, in this work, the terms "classical" or "traditional scaling" refer to the concepts of MOST.

For the canonical example of turbulent flow over a flat plate introduced at the beginning (see Fig. 1.1a), critical $Ri_c = 0.25$ means an extreme stratification, which leads to a significant suppression and damping of the turbulent vortices (Howard, 1961; Miles, 1961). However, this flow condition is typically transient, and turbulent mixing can be rapidly reestablished when a disturbance is added to the flow (Donda et al., 2015). Moreover, in atmospheric SBL, the existence of a critical Ri_c number is controversial, so that complete turbulence cutoff is practically rarely observed (Galperin et al., 2007). For $Ri > 0.25$ a weak and intermittent mixing is observed (Mahrt & Vickers, 2006).

The resulting characteristic feature of SBL is its unsteadiness. This can be observed from the time evolution of the Turbulence Kinetic Energy (TKE), which is calculated from the point measurements as follows:

$$e = 0.5(\overline{u'u'} + \overline{v'v'} + \overline{w'w'}). \quad (1.8)$$

An observational sample is given in Fig. 1.2. By performing a spectral analysis of the TKE signal, one can recognize the nonstationarity due to the increased energy of the high frequencies. The nonstationarity is to be understood as a relative property. The TKE of the neutral boundary layer (see Fig. 1.1a) is considered as the reference condition. In this state, the turbulent vortices are continuously distributed in space (see Fig. 1.4a). Kinetic energy is thought to be injected on one scale and transported through the vortex cascade, creating turbulence. Therefore, the TKE value near the surface can be estimated from the velocity of the free stream (e.g., the velocity at the top of the boundary layer). In the fundamental modeling of turbulence, it is assumed that the TKE follows any change in the velocity of the free stream with a specific dynamic. This assumption means that any frequency in the free stream velocity should not induce significant harmonics in the TKE signal. According to the

1. Introduction

measurements of the SBL, the TKE response can be intermittent (see the spikes in Fig. 1.2), which means that there are new frequencies that are not related to the mean flow velocity.

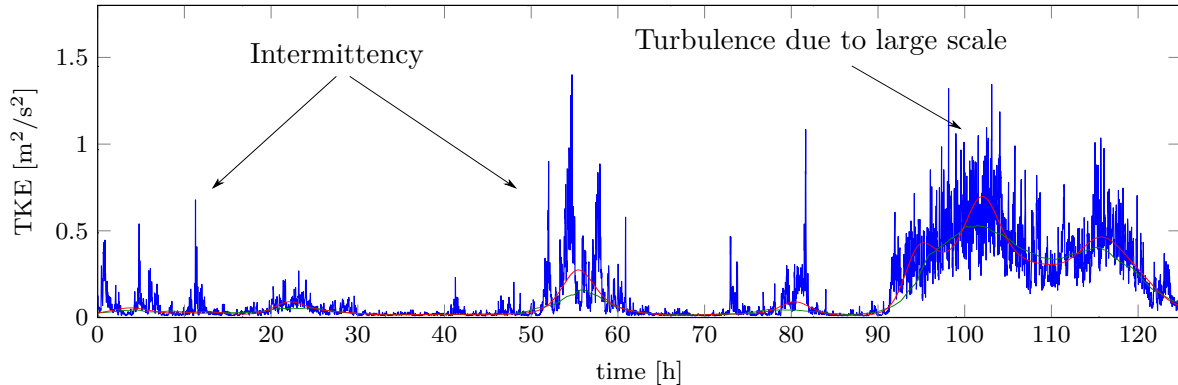


Figure 1.2: Measured time series of TKE. A detailed description of the dataset is given in Sec. 3.1. The figure shows two different regimes. There is a period with strong winds and continuous turbulence, where the turbulence is driven by the large scale and reasonably modeled with the MOST. And there are periods of weak winds and intermittent mixing. The intermittent TKE bursts are not represented in the RANS turbulence model. A model accounting for such bursts is developed in this thesis.

The reasons for irregular mixing under stable conditions are under research and challenging to unify (Mahrt & Bou-Zeid, 2020). The two prerequisites are the stratification of the atmospheric flow and weak geostrophic winds. Both circumstances lead to boundary layers that are decoupled from the ground in terms of turbulent transport. In the ABL at sufficiently high wind velocities, the turbulence is generated primarily at the surface, leading to the formation of attached boundary layers, as in the example of plate turbulence (see Fig. 1.1a). At low forcing wind velocities, other processes with comparable kinetic energy as the background mean flow cause additional local turbulence generation, which no longer attaches to the surface and leads to the formation of the detached boundary layers (Acevedo et al., 2016; Lan et al., 2018). In such a case, turbulence production is no longer associated with the surface and leads to an underrepresentation of the turbulent transport in the RANS models.

1.1.2 Sub-mesoscale Motions and Turbulence

The unresolved processes that occur within the numerical grid of a real-time NWP model combine diverse physical principles. For example, drainage flows (Mahrt et al., 2001), sharp microfronts (Mahrt, 2010a, 2019), internal gravity waves (Sun et al., 2004; Zaitseva et al., 2018), density currents (Sun et al., 2002), ducted waves propagating over several kilometers (Viana et al., 2009), thermal sub-mesoscale fronts (Pfister et al., 2021), meandering motions (Cava et al., 2019), ramp-like structures (Shi & Hu, 2020), top-down turbulent events (Sun et al., 2012) that are associated with KELVIN–HELMHOLTZ instability and wave-turbulence interaction (Sun et al., 2015). All these processes are described in the recent literature with the term sub-mesoscale motions (see Fig. 1.3 for an example). Mahrt (2014) defines that these non-turbulent motions are of size less than 2 km.

Research on the systematization and study of the sub-mesoscale motions reveals the great diversity of complex fluid-dynamic motions (Kang et al., 2015; Vercauteren et al., 2019b). One

can specify phenomenological mechanisms behind the patterns, but a clear and comprehensive definition is lacking. In studying the SBL, it is difficult to quantify KOLMOGOROV-like turbulence and to identify sub-mesoscale motions, because the scales of motion in the spectral domain have an overlap (Vercauteren et al., 2016). The alternative wavelet-based analysis, sometimes used to filter out flow motions, is limited by a specific form of the basis function and provides a criterion for fluid motions that is too restrictive. Assume that we can determine the position of the largest turbulent eddies in the SBL at any time (e.g., the inertial sub-range). Assume also that we can separate turbulence from the sub-mesoscale motions. Then "the energy-containing motions (e.g., sub-mesoscale motions) do not have a universal form brought about by a statistical equilibrium" (Pope, 2000, Sec. 6.6). This makes the generalization of sub-mesoscales challenging.



Figure 1.3: Observation of a sub-mesoscale flow structure at the surface through the fog. The fog injection occurred before sunrise, when stratification was at its highest. Such flow structures are considered non-turbulent, and their induced turbulence hypothetically leads to intermittent mixing in the boundary layer. Such mixing events are not considered by conventional models and are investigated in this work. Modified picture from <http://submeso.org>, used with permission of Dr. Larry Mahrt.

Hypothetically, since the sub-mesoscale motions involve multiple physical processes, the turbulence produced by each process can be assumed to be independent of the others (although this is rather an idealization). It seems that the multitude of energetic motions produces turbulence on several scales in the SBL. Laboratory experiments showed that the turbulence generated with a space-filling fractal grid with different wire diameters (i.e., several multiple scales) exhibits nontraditional equilibrium properties (Hurst & Vassilicos, 2007; Stresing et al., 2010). For idealized canonical turbulence like the plate turbulence, the outer scale (e.g., the largest turbulent eddy) and the dissipation scale (the smallest turbulent eddy) is proportional to the free stream velocity (Pope, 2000, p. 186). In the fractal-generated turbulence, there is only one relevant length scale, which can simplify the modeling approaches in certain cases. It is not known to what extent these findings are related to the turbulence and sub-mesoscale perturbations in the SBL.

The critical Ri number marks a theoretical value above which turbulent mixing is largely suppressed, and the definition of this number depends on the flow configuration. The principle of turbulence cutoff is also taken into account in the modeling. As long as some large energetic motions are present (and these are resolved by the model), the turbulence parametrization in the SBL remains relatively accurate. This regime is also known as the weakly SBL. The turbulence models relying on the Ri number refer to the most significant resolved scale and are bound to it. If the large scale loses kinetic energy (for example, due to pressure drop in the higher atmosphere), the SBL enters a regime determined by weak winds. Consequently, the shear of the flow at the surface decreases, the critical Ri number is exceeded, and the turbulence model predicts the collapse of the turbulence. At this point, the critical Ri number becomes a poor characteristic of the boundary layer state because it is constructed from shear whose kinetic energy is of the same order of magnitude as the kinetic energy of the sub-mesoscale motions.

As mentioned earlier, the sub-mesoscale motions typically fall within the sub-grid part of the model, and there is a lack of information on velocity gradients at those scales. The uncertainty in the Ri numbers results from the uncertainty in the sub-mesoscale motions. The more the kinetic energy of the sub-mesoscale motions corresponds to that of the resolved flow, the more undifferentiated the Ri number becomes. In addition, the velocity gradients are less associated with the surface and more associated with localized sub-mesoscale motions away from the surface. Consequently, the SBL is considered to be nonstationary.

To improve the representation of the SBL, this work focuses on the analysis and systematization of the sub-mesoscale motions. Thus, a parameterization will be suggested and developed that considers the effect of the sub-mesoscale motions in a RANS model.

1.1.3 Global Intermittency

Global intermittency is an alternate term for the nonstationary TKE dynamics introduced earlier. Mahrt (1999) defines the global intermittency as one of the properties of the very SBL and connects this property directly with the turbulence and not with the mean state of the flow. Global intermittency explains a state of turbulence within the SBL in which KOLMOGOROV-like vortices are absent at all scales for some regions, and an inertial subrange cannot be sustained for long periods. Since all turbulent flow exhibits some sort of intermittent behavior, Mahrt (1999) emphasizes that intermittency within a given vortex is a different term and is referred to as fine-scale intermittency. In this work, "intermittency" means global intermittency.

In other words, the intermittency would be caused due to the external perturbation of the flow field, where one can hypothetically associate for each perturbation process an individual localized turbulence cascade. The average statistical properties of such flows are particularly of significance for the modeling. In this context, external means that there is already a dominant flow-inducing scale. A large number of such external sub-mesoscale events overlap strongly within a specific spectral range. The turbulence and the associated intermittent mixing cannot be considered equilibrated in such a case. Chapter 5 explains in detail how such concepts of non-equilibrated turbulence can be expressed in a stochastic way.

Surprising is the recent discovery that global intermittency is not exclusively a consequence of sub-mesoscale motions. Ansonge & Mellado (2014) performed a DNS study and analysis of the SBL at impressively large Re and Ri numbers (see Fig. 1.4). New insights are gained by resolving all scales of fluid motions. The corresponding simulation of the EKMAN flow (which better reflects the ABL than a channel flow or a flow over a plate) shows that beyond the critical Ri number, there exists intermittent turbulence (see Fig. 1.4b). The flow in the numerical experiment is thereby forced by stationary boundary conditions and corresponds to the forcing with one large scale. Ansonge & Mellado (2014) concluded that global intermittency occurs despite the absence of external mechanisms classified as the sub-mesoscale motions. Ansonge & Mellado (2014) are convinced and suggest:

“ Global intermittency effects could be incorporated into turbulence closures for Reynolds-averaged NWP models once its dependency on Ri and Re is quantitatively understood – without depending strongly on the local details of the flow such as surface heterogeneities. ”

Global intermittency directly manifests in the stability function parameterization and causes a larger scatter in the observed data points (see Fig. 4.2). This dispersion is more pronounced with increasing stratification. However, in a DNS study (Ansonge, 2017, Sec. 9.3), it is shown that this dispersion can be explained. By conditioning the estimation of the relevant variables of the MOST on the locally turbulent regions (see Fig. 1.4b blue color) within the EKMAN flow, the data points coincide for all positive values of the Ri number, supporting the theory beyond $Ri_c = 0.25$. However, this finding has limited applicability to modeling because the localization of turbulent regions is a posteriori possible in a DNS and practically unknown in a NWP situation. Ansonge reports that the formation of the turbulent local areas depends on the internal properties of the flow configuration and is thus only indirectly linked to the external influences.

From the modeling perspective, the sub-mesoscale motions should be described stochastically, as suggested by Mahrt (2014). The intermittent TKE could be modeled directly by adding a stochastic term in one of the RANS equations. The modeling approach is motivated by the following studies. Belušić & Güttler (2010) performed a high-resolution numerical study (grid size 1/3 km) with a mesoscale model capable of resolving meandering motions in mountainous terrain. For the model to represent the intended scales, the authors had to eliminate turbulent diffusion in the horizontal direction. The variability and mean of the velocity components appear to agree reasonably well with observations in their simulations. However, the agreement is only in statistics, and locally predicted values of the wind differ significantly from observations. This difference can be caused by small changes and deviations in the model initialization. Uncertainties in the assimilation of the model affect the prediction accuracy proportionally to this uncertainty. Therefore, above a particular model resolution, further reduction of the grid size cannot achieve a refined representation of the observations. Thus, Belušić & Güttler (2010) states that the incorporation of the missing sub-mesoscale motions seems to be feasible only by reproducing meandering statistics, i.e., parameterizations.

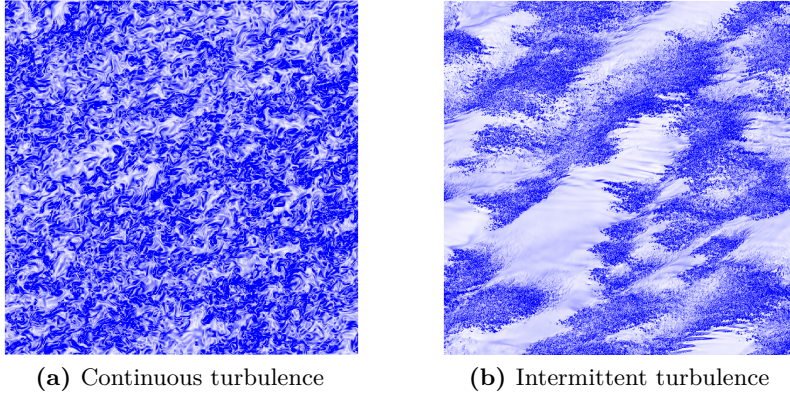


Figure 1.4: The magnitude of vorticity (blue color indicates high value and the white color indicates low value). A horizontal section of an DNS of EKMAN flow under two different stratification conditions is shown, revealing the phenomenological difference of turbulent motions. (a) Neutral stratified flow; homogeneous space-filling turbulent motions (b) Turbulence in a SBL at $Ri \approx Ri_c$; patchy and intermittent turbulent regions with locally laminar flow conditions. A constant geostrophic wind drives the flow. Used with permission of Dr. Cedrick Ansoerge. Figure courtesy of Abhishek Harikrishnan.

In a similar fine-resolution numerical study, Seaman et al. (2012) also report that a numerical model can predict realistic flow profiles, intermittent low-level jet turbulence, and submeso flow variability. However, a tendency was absent to improve the accurate individual realizations in observational data on finer scale by refining the resolution of the computational grid, such that only the statistics were adequately represented.

1.1.4 Modeling Errors and Uncertainty

Errors and uncertainties exist in NWP models during the simulation of turbulent exchange in the SBL and are one of the significant challenges in the parameterization task (Holtslag et al., 2013). Graham et al. (2019) analyze six atmospheric reanalysis datasets over arctic sea ice and continues to report on the known surface temperature bias associated with the SBL. The authors emphasize that the simulation of turbulent heat fluxes is inadequately represented and often has the wrong magnitude and/or direction. The authors explain this by the weakly expressed temperature inversions at the surface during stable weather conditions. The temperature profile directly responds to the turbulent mixing process, and the warm bias results from the turbulent diffusion in the models being intentionally tuned.

Turbulent mixing is tuned in the NWP to optimally fit the model results to the data, which is no surprise in the turbulence modeling community. Sandu et al. (2013) performed a series of sensitivity forecasting experiments varying turbulent diffusion and shows that this affects not only the boundary layer winds but also synoptic cyclones, anticyclones, and the planetary-scale standing waves. The tuning of turbulent diffusion leads to two opposite effects. According to Sandu et al. (2013) reducing diffusivity leads to improved boundary layer winds but biases the large-scale flow and 2m temperature. Therefore, the turbulent diffusion is set to higher values to represent the large scales adequately. The work presented in this thesis could provide an alternative approach since the new model (see Ch. 4) maintains the averaged diffusion level of the traditional model but increases the effective transport due to the non-Gaussian statistics in the TKE (see Fig. 5.5).

The correct representation of the large scales has priority over the representation of the winds at the surface. If turbulent diffusion is intentionally set higher, the averaged temperature transport towards the ground increases and leads to a warmer surface temperature. In contrast, the use of appropriate values for turbulent diffusion according to MOST, which does not take into account the additional mixing due to unresolved sub-mesoscales, leads to the collapse of the turbulence and thus to the breakdown of the transport from the upper atmosphere to the surface. It is important to note that heat transport in such a case is excluded from the surface energy balance equation due to decoupling. This equation is responsible for predicting the surface temperature. In the surface balance equation, the main counterforce to the turbulent heat flux is radiative cooling. Since turbulence is suppressed, radiative cooling dominates and the surface temperature decreases, leading to so-called runaway cooling issue (Jiménez & Cuxart, 2005). Thus, the compromise is to make the boundary layer more turbulent in order to maintain the correct representation of the large scales.

The cornerstone for the parameterization of the SBL remains the MOST with its restriction for strongly stable condition. The breakdown of MOST is related to the divergence of the stability function at higher stratification (see, for instance Fig. 4 in (Baklanov & Grisogono, 2007)). Optis et al. (2016) explains this problem and provides evidence that the divergence is related to the sub-mesoscale motions, as well as a possible way to model it. Nevertheless, Optis et al. (2016) also points out that additional reasons are the rotation of the wind vector and surface decoupling.

1.2 Turbulence Parameterizations

The recent advances in ABL parameterization approaches cover a considerable range. They can be broadly divided into deterministic and stochastic approaches.

1.2.1 Deterministic Parameterization

A more comprehensive understanding of stationary SBL is achieved by introducing a Turbulence Potential Energy (TPE) equation and provides therefore a much more complete theory than the MOST. Following Zilitinkevich et al. (2007) the TPE (E_p) is proportional to the potential temperature fluctuations:

$$E_p = \frac{1}{2} \left(\frac{g/T_0}{N_b} \right)^2 \overline{T'T'}, \quad (1.9)$$

where $g = 9.81\text{ms}^{-1}$, T_0 is a reference temperature and N_b the BRUNT-VÄISÄLÄ frequency. By multiplying the prognostic equation for $\overline{T'T'}$ by $(\cdot)^2$, an additional energy balance equation is obtained. Zilitinkevich et al. (2009, 2007, 2013, 2008) introduce the concept of total turbulent energy as the sum of the TKE and the TPE and emphasizes that the MOST do not account for the critical energy exchange between these two equations through the buoyancy flux term. Zilitinkevich et al. (2013) suggest that the relationship between the dissipation time scales for TKE and TPE is an asymptotic constant $C_\rho = 0.86$ as $Ri \rightarrow \infty$, implying that at high stratification, the dominant energy sink is in the temperature fluctuation equation and not in the TKE as might be assumed. Therefore, to represent an appropriate energy balance at the sub-grid level, Zilitinkevich proposes to use the so-called energy- and flux-budget closure

models. In particular, for the NWP, a three-equation version is recommended, consisting of the two energy balance equations and a diagnostic formulation for the vertical TKE. Kurbatskii & Kurbatskaya (2012) studied such an approach using numerical and observational data and confirmed the functional dependence of the ratio of potential energy to total energy as a function of the gradient Ri number.

Building on the ideas of Zilitinkevich et al. (2013), Āuran et al. (2014, 2018) developed a turbulence scheme with two prognostic turbulence energies that are valid for all atmospheric stratification conditions, taking into account moisture and phase changes. Such a scheme is attractive because the stability parameter is independent of the local gradients due to the two prognostic turbulence energies and becomes a predictive variable, thus circumventing the turbulence cutoff problem. Their model can represent a more continuous evolution of turbulent fluxes and has a deeper mixing property but tends to overestimate the cloud coverage. However, the authors report that their model outperforms the standard Integrated Forecasting System (IFS) model developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) in anticipating the 2-m temperature in the arctic region.

Other deterministic approaches aim to improve turbulent closures by keeping the number of equations minimal. Such equations are formulated based on the turbulent mixing length. This length is used in RANS models to model the turbulence diffusion coefficient, and in the case of SBL, the turbulent mixing length is controlled by the stability function. Designing the mixing length provides flexibility for specific modeling objectives, but the general principle is that with increasing stratification, the turbulent mixing length is decreased (Lenderink & Holtslag, 2004).

In a series of works, Huang & Bou-Zeid (2013); Huang et al. (2013) conducted an LES study of SBL and investigated the effects of stratification on turbulent structures. The authors proposed a novel mixing length formulation and found the improved performance of a RANS model for the unsteady simulation cases. Baas et al. (2018) validated a similar modeling principle. Baas et al. (2018) presents a special scale formulation that is additionally used in SBL and is deployed under very stable conditions. Such a model shows an improved simulation results for the Cabauw dataset. The present work builds on a similar idea and introduces a stochastic version of the turbulent mixing length.

1.2.2 The Gray Zone of Turbulence

The term "gray zone" refers primarily to the physics of CBL and is a term associated with the mesoscale modeling community. Wyngaard (2004) and Honnert et al. (2016, 2011) define the "gray zone" as a turbulence modeling issue. It occurs when the effective model resolution is less than or equal to the size of the prevailing KOLMOGOROV-like turbulence vortices. In particular, during the day when the surface is heated, the turbulent convective eddies dominate the mean flow and may become larger than the model grid size. Therefore, the growing vortices must appear as resolved in the simulated domain, even though their physical origin is within the sub-grid where they are parameterized. If we think of turbulence in terms of KOLMOGOROV theory, the large eddies are strongly anisotropic. This property makes it difficult to represent them as an independent entity.

The term "gray zone" has its justification for existence in the SBL community. During stable stratification, the large turbulent eddies are reduced to a smaller size. However, the absence of turbulence at high stratification allows sub-mesoscale motions to generate and maintain weak turbulent mixing. The relationship between such turbulence production scales and large-scale flow is unknown, and it is assumed that the sub-mesoscales are only transiently present at the subgrid-scale. These non-turbulent motions are unresolved and therefore require parameterizations. In general, as the Ri number increases, a model cannot adjust the grid resolution to match the decreasing size of the turbulence eddies. The presence of sub-mesoscale motions becomes progressively important for the generation of turbulence as stratification increases. This circumstance can therefore also be considered as a gray zone issue. Unlike the models that do not represent the turbulent convective eddies in the CBL, the models in the SBL do not represent the sub-mesoscale motions.

To address the issue of turbulence modeling in the "gray zone" of the CBL Zhang et al. (2018) propose a deterministic concept based on the work of Wyngaard (2004). The novelty of their approach is a model that combines two different ideas of turbulence parameterization. The first scheme is the LES closure regime, where the energetic motions are resolved, and the small, turbulent ones are parameterized according to the SMAGORINSKY–LILLY model. The second scheme is a RANS approach, which is utilized to simulate mesoscale dynamics. Zhang et al. (2018) combined both modeling strategies in a single code capable of transitioning between an LES and an RANS mode of operation, thus adapting to convective processes.

In his review, Honnert et al. (2020) shows several additional ways to model CBL in the gray zone. Of particular interest to this work are the mentioned stochastic approaches that improve the prediction of daily small-scale convective rain events.

1.2.3 Stochastic Parameterization at Sub-Grid Scale

Existing stochastic parameterizations are focused on modeling turbulence above the sub-grid scale and are mainly related to the CBL with a strong focus on diabatic processes and deep convection. The general idea is to add random perturbations of temperature and velocity at the smallest resolved scale level and reflect the representation of model uncertainties. Leutbecher et al. (2017) provides an overview of the Stochastically Perturbed Parameterization Tendency (SPPT) that is increasingly being implemented in NWP models. The SPPT method explicitly assumes that the nature of the perturbation is multiplicative, meaning that the added perturbations are multiplied by the physical tendencies. The multiplicative approach to perturbations is debatable and is currently being investigated. Christensen (2020) provides justifications for the approach and emphasizes that the stochastic perturbations should be positively skewed to be physically realistic. Finding ways to consistently and physically incorporate the element of randomness into NWP models is important.

Kober & Craig (2016) introduce the so-called Physically based Stochastic Perturbation (PSP) method. The amplitude of the perturbations is linked to the turbulence parameterizations in the sub-grid, resulting in a missing coupling mechanism between resolved and parameterized scales. Moreover, the perturbation's spatial and temporal correlation metrics closely mimic the turbulent statistical properties. Hirt et al. (2019) extended the idea of the PSP method to include the important non-divergent condition for the 3D stochastic velocity field.

Sakradzija & Klocke (2018) presented another example where stochastic approaches are superior to deterministic ones in predicting shallow convection. Their stochastic modeling idea differs significantly from the principles explained above. Model-relevant quantities such as cloud number, mass flux, and cloud lifetime are taken from assumed probability density functions, which are made dependent on the model’s resolution. The developed sampling procedure is uniquely adapted to the problem by incorporating two physical constraints. The resulting numerical study by Sakradzija & Klocke (2018) of their new approach shows an improved mean vertical structure of the cloud cover and variable profiles of the boundary layer.

The stochastic parametrization approaches in the CBL provide exciting ideas which could be adapted to the SBL. A lot of effort goes into making the stochastic strategies physically consistent. One component plays an essential role in this – the parameter dependence of stochastic models on the resolved scale dynamics, which is addressed in this work.

1.2.4 Stochastic Parameterization of Intermittent Turbulence

This work focuses on stochastic modeling of nonstationary turbulence in SBL. It is, therefore, appropriate to highlight an existing and closely related approach. This approach is not explicitly applicable to SBL, but it has a lot of overlap with the subject and is an introduction to stochastic turbulence modeling. In particular, the model calibration technique developed in Ch. 2 can be used to tailor in the following mentioned stochastic model to the SBL problem.

Pope (1991) proposed a probability density function model for inhomogeneous turbulent flows and demonstrated, using examples such as the momentumless wake and the plane mixing layer, that the method reproduces intermittency. In particular, (Pope, 1991, Eq. 11) has established a Stochastic Differential Equation (SDE) for the relaxation rate $\omega(\mathbf{x}, t)$, defining it as a mixed variable:

$$\omega(\mathbf{x}, t) = \epsilon(\mathbf{x}, t) / \tilde{k}(\mathbf{x}, t), \quad (1.10)$$

where $\epsilon(\mathbf{x}, t)$ is the local instantaneous dissipation rate and $\tilde{k}(\mathbf{x}, t)$ is the deterministic TKE normalized by the averaged dissipation. The complete turbulence model is thus a stochastic version of the k - ω two-equations model family. However, it is debatable if the model can be applied to SBL to model global intermittency because the stochastic dissipation is defined locally. Pope (1991) argues that the parameters of his SDE lacks an explicit dependence on the external large-scale flow variables such as Re number and that such a dependence would further improve the model. Chapters 4 and 5 offer an approach and an application example how the described issue can be addressed for a certain inhomogeneous turbulent flow.

Calibration of Eq. 11 in Pope (1991) for the SBL flow using the method in Sec. 2.2.1 is hindered by the estimation of the instantaneous dissipation rate. This variable is difficult to estimate correctly from the time series of field measurements because of the uncertainties associated with the required data processing steps. These steps include the TAYLOR hypothesis, which is invalid for non-neutral conditions in the ABL (Cheng et al., 2017). Possible assumption of KOLMOGOROV theory and associated assumption of steady-state flow conditions are violated in SBL. Nevertheless, one possible way to incorporate Pope’s model into the SBL scenario could be to analyze DNS data since the dissipation rate can be determined directly from the nonstationary simulation results.

Stochastic approaches to model turbulence can also be effective tools if certain abstractions and simplifications are agreed upon. One such example is the One-Dimensional Turbulence model (ODT) developed by Kerstein (1999); Kerstein et al. (2001). The main idea is to replace the complex nonlinear term in the NAVIER–STOKES equation with a stochastic mapping event confined to a one-dimensional line in the simulation domain, representing thereby the nonlinearity stochastically in the equations. The random mapping event in the ODT represents a mathematically modeled mixing occurrence of turbulent eddies. The sampling procedure for such events is in accordance with the principles of conservation of energy. As a result of the approach, the turbulent statistical and averaged properties of the flow variables are recovered by the ODT without any closure assumptions. Klein & Schmidt (2019) provides a numerical investigation of the canonical RAYLEIGH–BÉNARD convection problem using the ODT model and reports a valid numerical representation of the physical phenomenon. Freire & Chamecki (2018) investigated the ABL with surface canopy using the ODT model. The canopy effect is introduced through the energetics of stochastic vortices. The authors report that the model’s overall performance is quantitatively accurate and emphasize that some improvements are still possible. The method is attractive for research projects since it can simulate adequate turbulence statistics at significantly large Re numbers with low costs.

Due to the unreproducible field measurements, the construction of stochastic, nonstationary turbulence models for SBL requires data-driven stochastic approaches (Vercauteren & Klein, 2015; von Larcher et al., 2015). Monahan et al. (2015) applies a hidden MARKOV model to partition the SBL into two dominant and highly overlapping regimes. The approach offers new insights since their classification does not depend on a stability parameter. Instead, it considers the temporal and cross-dimensional dependence of a three-dimensional vector time series, which sufficiently describes the SBL. After a comprehensive analysis of nine geographically distinct measurement sites, Abraham et al. (2019) proposed a stochastic model prototype based on the MARKOV chain approach to simulate the observed stochastic regime transitions in SBL. Further research will investigate the properties of the regime transition, including the spatial dimension. Abraham & Monahan (2020) reports that SBLs are influenced by local process activity and are related to topography.

1.3 Data-Driven Modeling of the Stable Boundary Layer

Adequate representation of the strongly SBL in NWP remains challenging. The turbulence responsible for the transfer of momentum, temperature, and moisture becomes intermittent at high stratification and rarely disappears completely. Intermittent mixing events fall into the unresolved sub-grid scale of the models and are therefore underrepresented. Crucially, some energy-containing motions - the sub-mesoscale motions - also can fall within the subgrid-scale of the model. These turbulence-generating processes and the turbulent mixing are difficult to distinguish from each other. It is, therefore, unclear to what extent the sub-mesoscale motions generate intermittent turbulence.

Stochastic modeling of ABL has become increasingly relevant in the last decade. In this regard, the approaches in CBL differ from those in SBL. There is a clear tendency for stochastic modeling to be more advanced in CBL than in SBL. Although the physical principles vary

significantly between the two types of boundary layers, there is a realistic prospect that the transfer of the stochastic approaches to SBL is feasible.

Besides the noticeable difference in the number of research articles, there is also a difference in the general approach, which is physically motivated. The ABL in the convective regime has a different turbulence character than in the strongly stable regime. This encourages applying the stochastic approaches in a way that is based on some modeling principles. In comparison, the processes and turbulence in the strong SBL are more uncertain. The application of first-principles modeling in SBL is complicated by the numerous overlaps and interactions between physical concepts. For this reason, the research direction in SBL relies heavily on the use of data-driven exploratory analysis methods that aim to classify and systematize.

The relationship between the non-turbulent motions and the induced intermittent mixing is generally unknown and not sufficiently studied. To improve the models, it is important to address how the sub-mesoscale motions generates turbulence in strongly SBL to understand better how it could be represented stochastically. To construct a stochastic closure for the SBL that incorporates the effects of sub-mesoscale motions, it is important to identify a possible set of stochastic surrogate models with time-dependent parameters. The main objective of the research is to use appropriate inverse modeling methods to determine scaling functions from the SBL data that map large-scale flow variables to the time-dependent parameters. This will allow us to calibrate the stochastic mathematical model and combine it with the existing turbulence models. Therefore, the objective is to use a suitable data-driven parameterization approach for nonstationary time series modeling.

1.4 Summary and Research Proposition

The idea for initiating the research is to separate the SBL in terms of the regime-averaged transport property following Vercauteren & Klein (2015). The authors propose to separate the regimes in terms of scale interactions, which makes more physical sense in contrast to the Ri number method. A model-based statistical clustering method is applied to identify the regimes. Thereby, the model structure is formed by a linear combination of different time-lagged regression variables. These variables are composed on multiple scales from the wind speed measurements. After classification, a multiscale spectral analysis is performed to identify the scales that produce average transport and are identifiable as turbulence (Vercauteren et al., 2016). This analysis brings insight into the scale-wise transport properties of intermittent turbulence within the SBL and is practiced in this work on data from a different measurement site.

The first objective of this work is therefore:

- Analysis and classification of intermittent turbulence and sub-mesoscale motions within regimes based on interactions at different scales following Vercauteren & Klein (2015).

To build a turbulence model that accounts for the turbulent mixing from sub-mesoscales, it is important to understand the generating mechanisms of intermittent turbulence in the SBL. Realistically, this goal is not achievable within the scope of this work due to the enormous amount of processes to be analyzed and the complexity of the SBL. However, with the clustering method, it is possible to test a more general hypothesis statistically. One of the

known precursors of turbulence is the gradient of the stream velocity. Since we are dealing with a model-based method, the idea is to model the measured TKE by a linear combination of the velocity gradients. The peculiarity of the approach is that the gradients fed into the model are calculated from the wavelet-filtered velocity at multiple scales. This should lead to a more specific classification and modeling of the TKE. It is hypothesized that in a regime of strongly stable stratification, the gradients of wind speed corresponding to the scale range of sub-mesoscale motions predict a predominant fraction of the intermittent TKE variability. Conversely, gradients of large and resolvable scale should control the regime of weak stratification.

The second objective is therefore:

- Explore potentially suitable model structures for statistical clustering and construct a surrogate model for the temporal evolution of the TKE at a fixed point in space for the SBL.

The sub-mesoscale motions include a variety of different processes with unique physical mechanisms. In particular, the relationship between these unresolved processes and large-scale flow is unknown. Therefore, it is unclear how the sub-mesoscale motions within a sub-grid cell of a NWP model can be predicted using the resolved scales. This means that even if a suitable model existed that predicts the TKE from the sub-mesoscale motions, the lack of connection to the large scale prevents us from simulating the sub-mesoscale motions in the first place. Therefore, a model that predicts TKE solely from the sub-mesoscale motions is more suitable for research purposes.

The two principal insights from the previous objectives are as follows. First, the sub-mesoscale motions should be represented in a rudimentary way as a random process. Structural-spatial properties of such a stochastic process are irrelevant in the first place, as they would only add complexity at this stage of understanding. The only important variable is the intensity of the noise process, which is used to represent the sub-mesoscale activity. Hypothetically, as stratification increases and mean wind speed decreases, the importance of sub-mesoscale motions and thus noise intensity should increase. The intensity of the turbulence diffusion noise process is indeed related to sub-mesoscales, as reported in this work (see Fig. 4.4).

Second, the model-based statistical clustering approach applied in objective one can be used as a parameterization tool for nonstationary stochastic equation models. The mathematical formulation of the initially used tool has to be readjusted for this goal. The main drawbacks are the discrete-time formulation and the linear model structure. Discrete-time models are trained for a fixed step size given by the training data. This makes it inconvenient to integrate such a data-driven model into a turbulence closure in the most general way. The third objective is therefore primarily the further development of the clustering method:

- Modification of the statistical clustering algorithm for modeling time series with nonlinear and nonstationary stochastic differential equations of continuous type.

Once a suitable turbulence parameterization tool has been developed, the next important step is to select an appropriate variable to be modeled and to be considered for the inclusion of random perturbations. The strategy is to look for a non-dimensional variable that is already part of an existing RANS turbulence model structure. In conjunction with this variable, it is

hypothesized that there exists a low-dimensional model with nonstationary, time-dependent parameters. It is assumed that the time dependence of the parameters reflects the regime occupation and the regime transitions within the SBL. Here, the noise process of the SDE is a proxy that mimics the activity of the sub-mesoscale motions.

The fourth objective is therefore:

- Identify a continuous-type model structure with a small number of parameters for a modeling variable under consideration.

The development of the stochastic closure for the SBL is divided into identifying a model structure and the corresponding calibration with measured data. The time-varying model parameters are re-expressed using the variables on coarser temporal scales by a so-called scaling function. Therefore, to make the task feasible, it is essential to keep the number of parameters small since each additional parameter requires another scaling function that is valid together with the previous one. The fifth objective is therefore:

- Discovery of the functional forms of the scaling functions linking the time-dependent SDE model parameters to the resolved variables of the large scale.

After a characteristic variable is modeled, parameterized, and validated, the new equation is evaluated in a RANS model. The sixth objective is therefore:

- Incorporation of the stochastically modeled characteristic variables into a turbulence closure and analyzing the ability to predict the strongly SBL.

The validation of the novel stochastic turbulence model is based on the bottom-up principle. The predictive power of the discovered SDE, which is the central element of this work, is validated utilizing the large scales calculated from a dataset. Thus, at the lowest level of model validation, the representation of the characteristic stochastic variables is verified against data independently of additional modeling assumptions. At the next level of complexity, the stochastic equation is integrated into an autonomous turbulence closure. This stage examines whether the desired characteristic intermittent turbulence is represented and whether unexpected model behavior occurs. The final test would incorporate the new stochastic equation into a NWP system, but that is left for future research.

1.5 Outline

This thesis is organized as follows. Chapter 2 discusses methods for analyzing and modeling turbulence data, explaining the statistical discrete-time clustering method and presenting an advanced continuous-time version of this method. Chapter 3 focuses on the analysis of SBL using the discrete-time classification technique. Furthermore, chapter 3 emphasizes that the clustering approach can also be used as a parameterization and modeling tool. Consequently, in chapter 4, the version of the continuous-time clustering method developed in Ch. 2 is used to formulate, parameterize, and validate a nonstationary stochastic stability equation that models a characteristic turbulence variable for the SBL. Chapter 5 shows how the novel equation is integrated into a RANS model. Several well-controlled numerical case studies are performed to validate the novel model suitable for representing SBLs.

1.6 Publications

Boyko, V., Vercauteren, N. Multiscale Shear Forcing of Turbulence in the Nocturnal Boundary Layer: A Statistical Analysis. *Boundary-Layer Meteorol* 179, 43-72 (2021). <https://doi.org/10.1007/s10546-020-00583-0>

V. Boyko conducted the research. N. Vercauteren performed a supporting role. Parts of this paper are included in Ch. 3.

Boyko, V., Krumscheid, S., Vercauteren, N. Statistical learning of nonlinear stochastic differential equations from nonstationary time series using variational clustering. Accepted for publication in *Journal of Multiscale Modeling and Simulation* on 27 April 2022. ArXiv: 2102.12395.

V. Boyko conducted the research. N. Vercauteren performed a supporting role. S. Krumscheid performed a supporting role and provided helpful instructions. Furthermore, S. Krumscheid provided a framework element that computes the transition probability density function. Parts of this paper are included in Ch. 2.

Vercauteren, N, Boyko, V, Feranda, D, Stiperski, I. Scale interactions and anisotropy in stable boundary layers. *Q J R Meteorol Soc.* 2019; 145: 1799-1813. <https://doi.org/10.1002/qj.3524>

N. Vercauteren conducted the research. D. Faranda and I. Stiperski had a supporting role and provided helpful instructions along with the corresponding computational scripts. V. Boyko performed the calculations and produced the figures.

Vercauteren, N., Boyko, V., Kaiser, A. Belusic, D. Statistical Investigation of Flow Structures in Different Regimes of the Stable Boundary Layer. *Boundary-Layer Meteorol* 173, 143-164 (2019). <https://doi.org/10.1007/s10546-019-00464-1>

N. Vercauteren performed the research. A. Kaiser performed calculations. D. Belusic provided helpful instructions. V. Boyko performed calculations and produced the figures. Figures from this work are included into Ch. 3.

2

Model-Based Nonstationary Time Series Modeling

This work is utilizing time series measurements recorded continuously over several months. These are multi-point measurements taken on a tower. Therefore, the spatial resolution in the vertical direction away from the surface is sparse. The vertical transport is essential in connecting the upper atmosphere with the surface, and is the primary research interest.

The so-called eddy covariance methods involve point measurements that record a time series of a 3D wind velocity vector and temperature. The instruments used are sonic anemometers with a typical sampling frequency of up to 20 or 60 Hz (Lee et al., 2004). Such temporal resolution is sufficient to resolve the turbulent eddies contributing to transport but is insufficient to measure the relevant turbulent dissipation rate (this variable is not in focus for this work). As a side note, it is necessary to point out that the dissipation rate is crucial for improving the turbulence models in the SBL; therefore, different measurement techniques must be applied to capture this quantity. Kit et al. (2017) established a measurement setup where the sonic anemometers were utilized to calibrate and use the hot-wire/film anemometry technique. Kit et al. (2017) were able to resolve air motions at a temporal sampling frequency of 2 kHz and examine fine-scale bursts of turbulence, providing insights into the entire inertial subrange of the ABL.

The general idea of a data-driven stochastic modeling strategy is parameterization of the unresolved degrees of freedom in a complex system. Consider a multiscale system composed of a model representing the macroscale (RANS), while a low-dimensional model represents the fast degrees of freedom (in this case it is the equation for the TKE). Theoretically, the fast degrees of freedom are driven by the macroscale variables in some concealed way which one seeks to identify. Suppose to have access to at least one measurement set for this modeling task, including all scales forming the training data set. It is assumed that the resolved scales control the dynamics by modulating the parameters of yet unknown low-dimensional model. One tries to describe the fast degree of freedom with a low-dimensional model, recover its temporal parameter modulation and re-express them by some appropriated resolved characteristic

variable of a dynamical system. Thereby it is assumed that there is a hidden functional relationship, which one aims to infer, between the model parameters and some resolved deterministic variable.

This work focuses on inverse modeling of general time series data with discrete and continuous type models and the inverse modeling will be applied to turbulence measurements later in this work. This chapter provides a brief overview of the nonstationary linear discrete-time statistical clustering method (Horenko, 2010b) applied in Ch. 3 to study and classify the SBL regimes. Thus, the used model structures entails physically meaningful hypotheses that contribute to the main objective of the clustering. Nevertheless, the central part of this chapter is concerned with developing a statistical parameterization framework for continuous-type nonlinear models based on the ideas of the discrete-time framework (Horenko, 2010b). The objective in designing such a method is explicitly to model a nonstationary temporal evolution of some relevant turbulence variable (see Ch. 4). The imposed model structure should be modular and directly integrated into a RANS turbulence model. This integration is the subject of Ch. 5.

2.1 Stationary Models

This section summarizes information and methods for modeling stationary time series. The models presented are then combined with a clustering procedure to model nonstationary data.

2.1.1 Linear Discrete-Time Models

This section briefly reviews the theoretical background for discrete-time models, focusing on one particular type. This specific model type is thus the model structure for the variational clustering approach applied in Ch. 3. The clustering method will only be briefly described to provide the context for further methodological development.

Consider a time series $Y[n]$, where $n = 1, 2, \dots, N$ is the experimental observations at time points $0 = t_0 < t_1 < \dots < t_N = t_e$ with a constant time step Δt and a total number of N observations. The square brackets indicate that this is a discrete-time variable. In particular, this notation implies, for example, that some discretization operation of an underlying system has already been performed and applied to the continuous counterpart variable or system. The time series $Y[n]$ is the target process to be modeled. Moreover, process $X[n]$ is said to force or control process $Y[n]$ under a hypothetical assumption. This process acts as an exogenous variable on the system. In such a case, a stationary linear AutoRegressive Moving Average model with eXogenous variables (ARMAX) has the general form, characterized by the model terms containing discrete time-lagged information, as:

$$Y^*[n] = a_0 + \sum_{i=1}^p a_i Y^*[n - i\Delta t] + \sum_{j=1}^q b_j X[n - j\Delta t] + \sum_{k=1}^r c_k \xi[n - k\Delta t] + d_0 \xi[n]. \quad (2.1)$$

The idea is to match a prediction process $Y^*[n]$ to the observed process $Y[n]$ by using a model structure based on p past observables of itself, q past observation values of an external process X together with an stochastic process ξ which is independent of Y^* and also provided with information about the past r values. The process ξ has a special meaning and will be explained

in the following. Equation (2.1) is defined as a stationary model by assuming the parameter values a_0, a_i, b_j, c_k, d_0 to be constant within the N observations. When we attempt to fit the model to the data, we must ensure that the process we are trying to model is stationary and does not contradict this assumption.

Another model detail is the discrete time step Δt , shown in Eq. (2.1) for demonstration purposes. The time step is implicitly included in the model by processing the variables Y and X and is not explicitly incorporated into the procedure for estimating the model parameters. For this reason, one must always additionally specify the time step when evaluating such models since the estimated parameters are dependent on Δt . In later applications, Y will be a part of a turbulence closure model to be embedded in RANS. The turbulence model may use a different or variable time step, leading to discrepancies with the fitted ARMAX. Integrating ARMAX into other systems requires additional care, e.g., converting them into continuous-time models. With additional nonlinear terms, the complexity of such a transformation increases significantly (for an approach see (Billings, 2013)).

To understand the role of the random process ξ in inverse modeling, it is helpful to introduce the concepts of deterministic and stochastic processes. Deterministic processes are processes that are known to have a representative system of equations. Stochastic processes are those processes that we observe but cannot establish an exact equation. A process that we do not understand and for which there is no descriptive equation seems entirely random to us.

During an identification process of the system from data, continuous invention and testing of the model structure contributes to quantifying the variability, and thus the variance, of the time series $Y[n]$ under study (see Eq. (2.1)). Such a modeling principle is based on the WOLD representation theorem (Hannan, 1970, p. 137).

The following elementary components create Eq. (2.1). The simplest imaginable model would consist of the variance d_0 and the mean a_0 , which are essentially the first two statistical moments of the time series Y . This operation is usually the first step to comprehension when we know nothing about the data. Next, one could test a pure autoregressive (AR) model by minimizing the L_2 norm between the data and the model concerning its parameters

$$\mathbf{a}^* \in \operatorname{argmin}_{\mathbf{a}} \left\{ \left\| Y[n] - a_0 - \sum_{i=1}^p a_i Y[n - i\Delta t] \right\|_2^2 \right\}, \quad (2.2)$$

where $\mathbf{a}^* = (a_0, a_1, \dots, a_p)$ is the vector of optimal model parameters. The inverse problem is solved by forming a well-posed system known as the YULE-WALKER equations. The residual after minimization should then be the independent and identically distributed (i.i.d.) process ξ ; it must be orthogonal to the time series generated by the assumed model structure. Furthermore, it is also preferable that the residual variance is small such that $d_0 \rightarrow 0$ in Eq. (2.1).

If the residual is not i.i.d. and has some autocorrelation, one further refines the modeling hypothesis and adds a deterministic variable X . The estimation of the autoregressive model with exogenous variable (ARX) is done in a similar way by minimizing:

$$\mathbf{a}^*, \mathbf{b}^* \in \operatorname{argmin}_{\mathbf{a}, \mathbf{b}} \left\{ \left\| Y[n] - a_0 - \sum_{i=1}^p a_i Y[n - i\Delta t] - \sum_{j=1}^q b_j X[n - j\Delta t] \right\|_2^2 \right\}, \quad (2.3)$$

The residual of the minimization should then be the i.i.d. process ξ . If the residual ξ still significantly correlates with the system variable one models the residual ξ with a moving average process in the final step. This stage is called error modeling and is a separate procedure after which the ARMAX is entirely identified.

When using discrete-time models, always ensure that the residuals resemble white noise. Otherwise, the model is biased, i.e., it is only valid for the uniquely observed data and cannot be used for predictions with other input signals. This is an inherent property of discrete-time models since the general model validity can be directly quantified by analyzing the residuals.

The general approach to identifying discrete-time linear models was briefly addressed. However, a more comprehensive and far-reaching explanation can be found in (Billings, 2013). Billings (2013) covers all relevant and advanced topics such as model structure estimation, error modeling, estimation of nonlinear discrete-time models, the transformation of discrete-time linear and nonlinear models to continuous-time models. The ideas of discrete-time linear models are applied in Ch. 3 for modeling nonstationary processes; in the next section, we will look at continuous and stationary models.

2.1.2 Nonlinear Stochastic Differential Equations (SDE)

The advantage of using SDE in modeling time series data is that they are of the continuous type. With the specific formulation used here, they also model the nonlinear noise terms. Error modeling is included in the estimation procedure and does not need to be performed separately as in discrete-time models. In addition, the time step Δt is included in the estimation procedure so that after the identification of the SDE, direct coupling to the RANS turbulence model can be implemented. These models are combined next with a variational clustering approach in Sec. 2.2.1 to identify nonstationary SDEs. The whole procedure of nonstationary model identification is then applied in Ch. 4 to develop a turbulence closure for the SBL.

The use of SDEs in the given formulation below has a disadvantage compared to the linear discrete-time models presented in Sec. 2.1.1. The used SDEs have one time derivative; therefore, the smoothness or regularity of the predicted time series is limited. In contrast, the discrete-time model contains several time-delayed terms that can model a time series with smooth wave patterns. Considering the variable to be modeled for the RANS closure (Eqs. (5.1) - (5.3)), the regularity of the proposed SDE is sufficient and is analyzed in Sec. 4.6.3 in more details. The methodology is discussed next.

Assume a stochastic process denoted by a one-dimensional, time-dependent variable $X(t)$. Let $X(t_i)$ be the (experimental) measurements at the observation times $0 = t_0 < t_1 < t_i < t_N = t_e$. The signal is recorded at a constant sampling frequency such that $t_{i+1} - t_i = \Delta t > 0$ for $i = 0, \dots, N$. The dataset consists of N observations. Let us consider as underlying data generating process a general one-dimensional ITÔ SDE of the form

$$dX = f(X; \boldsymbol{\vartheta}) dt + g(X; \boldsymbol{\vartheta}) dW(t), \quad X(t_0) = X_0, \quad (2.4)$$

over a bounded time interval $[0, t_e]$, $t_e > 0$. The functional form for the drift $f(\cdot; \cdot)$ and diffusion $g(\cdot; \cdot)$ is nonlinear in X and linear in $\boldsymbol{\vartheta}$, depending on an unknown vector-valued parameter $\boldsymbol{\vartheta} = [\vartheta_1, \vartheta_2, \dots, \vartheta_n] \in \Omega_{\boldsymbol{\vartheta}} \subset \mathbb{R}^n$, where $\Omega_{\boldsymbol{\vartheta}}$ denotes the acceptable parameter space. The vector $\boldsymbol{\vartheta}$ includes the parameters of the drift f and diffusion term g . Furthermore, suppose $W(t)$ is

an admissible one-dimensional WIENER process. It is assumed that Eq. (2.4) has a unique, strong solution for each $\boldsymbol{\vartheta}$. The goal is to estimate the parameter $\boldsymbol{\vartheta}$ from given N observations using the maximum likelihood approach. The discrete-time negative log-likelihood function is defined as (Fuchs, 2013)

$$l_N(\boldsymbol{\vartheta}) := - \sum_{i=0}^{N-1} \ln [(p_X(\Delta t, X(t_{i+1})|X(t_i); \boldsymbol{\vartheta}))], \quad (2.5)$$

where p_X indicates the conditional transition density function (c.t.d.f.) of the stationary process X . From now on, we refer to l_N as the *fitness function* to match the definition of the variational framework of Horenko (2010a), but in general, l_N is a cost function. The optimum of the fitness function with respect to $\boldsymbol{\vartheta}$ results in the parameter estimate

$$\boldsymbol{\vartheta}^* \in \arg \min_{\boldsymbol{\vartheta} \in \Omega_{\boldsymbol{\vartheta}}} l_N(\boldsymbol{\vartheta}), \quad (2.6)$$

which is the Maximum Likelihood Estimation (MLE). It is known that the MLE converges to the true parameter $\boldsymbol{\vartheta}_0$ in probability as $N \rightarrow \infty$ for a fixed Δt . Moreover, it is consistent and asymptotically normal and independent of the discretization (Dacunha-Castelle & Florens-Zmirou, 1986),

$$\sqrt{N}(\boldsymbol{\vartheta}^* - \boldsymbol{\vartheta}_0) \rightarrow^d \mathcal{N}(0, \boldsymbol{\sigma}_{MLE}^2), \quad (2.7)$$

where $\boldsymbol{\sigma}_{MLE}^2$ is the variance of the MLE $\boldsymbol{\vartheta}^*$. This estimator converges in distribution \rightarrow^d to the unknown parameter at a rate $1/\sqrt{N}$.

For a general nonlinear SDE of the type (2.4) the conditional transition density function is unknown but can be approximated using the closed-form expansion following Ait-Sahalia (1999, 2002). Process X is transformed into process Z with two steps. The transition density of process Z , namely p_Z , is evaluated via HERMITE polynomials, and the required density p_X is obtained by inverting the transformations. The first transformation step, $Y := \mathcal{L}(X)$, standardizes the process X such that Y has unit diffusion:

$$dY = \mu(Y; \boldsymbol{\vartheta}) dt + dW(t), \quad Y(t_0) = Y_0, \quad (2.8)$$

where the drift μ is

$$\mu(Y; \boldsymbol{\vartheta}) = \frac{f(\mathcal{L}^{-1}(Y); \boldsymbol{\vartheta})}{g(\mathcal{L}^{-1}(Y); \boldsymbol{\vartheta})} - \frac{1}{2} \frac{\partial g(\mathcal{L}^{-1}(Y); \boldsymbol{\vartheta})}{\partial X}. \quad (2.9)$$

The LAMPERTI transform is defined

$$\mathcal{L}(X) := \int g(X, \boldsymbol{\vartheta})^{-1} dX, \quad (2.10)$$

such that the inverse $\mathcal{L}^{-1}(Y)$ exists. The second transformation is introduced to normalize process Y with respect to the step size:

$$Z := (Y - y_0)/\sqrt{\Delta t}. \quad (2.11)$$

The factor $1/\sqrt{\Delta t}$ prevents that the transition density turns into a Dirac delta function for small Δt and the shift operation resets the initial value for the process Z to $Z_0 = 0$ (Fuchs, 2013)[ch. 6.3]. The truncated HERMITE expansion of the transition density function p_Z is defined as

$$p_Z^{(J)}(\Delta t, z|y_0; \boldsymbol{\vartheta}) := \phi(z) \sum_{j=0}^J \eta_j(y_0, \Delta t; \boldsymbol{\vartheta}) H_j(z), \quad (2.12)$$

where $\phi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ is the weight function,

$$\eta_j(y_0, \Delta t; \boldsymbol{\vartheta}) := (1/j!) \int_{-\infty}^{+\infty} H_j(z) p_Z(\Delta t, z|y_0; \boldsymbol{\vartheta}) dz \quad (2.13)$$

are the coefficients and $H_j(z)$ are the HERMITE polynomials. By reverting the transformation from Z to X , one obtains the approximated density $p_X^{(J)}$:

$$p_X^{(J)}(\Delta t, X(t_{i+1})|X(t_i); \boldsymbol{\vartheta}) = \frac{1}{g(X(t_{i+1}); \boldsymbol{\vartheta})\sqrt{\Delta t}} p_Z^{(J)}(\Delta t, z|y_0; \boldsymbol{\vartheta}). \quad (2.14)$$

Theorem 1 of Aït-Sahalia (2002) established the necessary conditions for the convergence of

$$p_X^{(J)}(\Delta t, X(t_{i+1})|X(t_i); \boldsymbol{\vartheta}) \rightarrow p_X(\Delta t, X(t_{i+1})|X(t_i); \boldsymbol{\vartheta}) \quad \text{as } J \rightarrow \infty, \quad (2.15)$$

both uniformly in $X(t_i)$ and in $\boldsymbol{\vartheta}$; this yields the actual transition density function p_X . To evaluate the integral in Eq. (2.13), Aït-Sahalia (2002) proposes a TAYLOR series expansion in Δt for the coefficients η_j up to the order $M = 3$ (see Appendix A.1). The convergence rate of the truncation error is dependent on the choice of J :

$$\eta_j^{(M)}(y_0|\Delta t; \boldsymbol{\vartheta}) = \eta_j(y_0, \Delta t; \boldsymbol{\vartheta}) + \mathcal{O}(\Delta t^{M+1}), \quad j = 1, \dots, J; \quad M = J/2. \quad (2.16)$$

In this work, $J = 6$ is chosen as it is also used by Aït-Sahalia (2002); this is a sufficiently accurate approximation and agrees with the proposed clustering approach presented below.

In summary, consider an abstract sketch (see Fig. 2.1) that reflects the scope of autonomous SDE modeling. As one would expect, when an SDE with stationary parameters is used to estimate a strongly nonstationary time series, the predictive ability could be poor outside of the observed samples (see Fig. 2.1). In the following section, the modeling is refined with a more detailed approximation.

2.2 Nonparametric Clustering Methods for Nonstationary Models

This section addresses a nonparametric clustering method used to model and analyze turbulence time series data with a pre-specified model structures.

2.2.1 Smooth Variational Clustering of SDEs

The nonparametric variational clustering method relaxes the assumptions of constant parameters in a generally defined SDE (2.4) (see Fig. 2.1). It thus contributes to the analysis

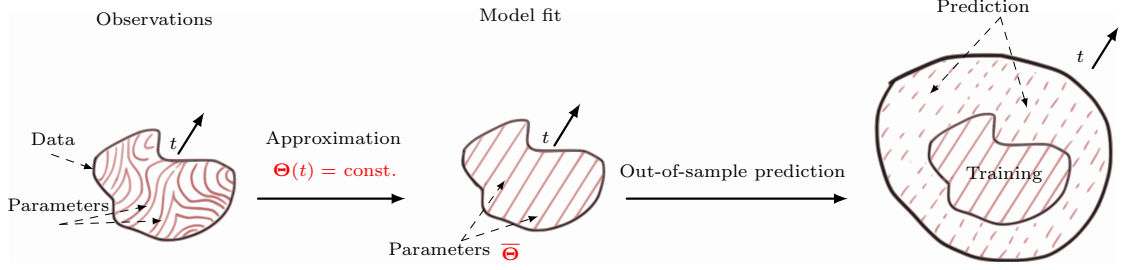


Figure 2.1: An illustration of a data-driven modeling approach with stationary SDE. In the left diagram, the curved red lines illustrate the parameter space of a possible model for some set of observations. Fitting an SDE with constant parameters assumes the variability of the parameters to be time-independent, as shown in the middle cartoon. Out-of-sample prediction is inaccurate because the parameters’ evolution is unknown outside training data, as shown in the right cartoon.

of complex and nonstationary time series. Data recorded during an experiment in the ABL is challenging to reproduce due to uncontrollable environmental conditions. The duration of a field measurement campaign can span several months, resulting in a single time series with a wide range of dynamically different regimes (Mahrt, 2014; Sun et al., 2012; Vercauteren et al., 2019b). Therefore, it is necessary to cluster the data to isolate and understand the physical mechanisms in each of the reasonable regimes.

Clustering and modeling of transient wind measurements based on SDE with time-dependent parameters is an excellent tool. The task of classification improves when the model structure is adapted to the process dynamics. An improved classification allows a more accurate understanding of the SBL regimes and thus leads to more sophisticated suggestions and improvements of the model structure. The application of model structure design and data classification cycle makes it possible to find scalable surrogate models.

This section addresses the technical aspects of the nonparametric variational clustering method and the necessary steps to use it in combination with generally defined nonlinear SDE. Thus, the goal is to estimate the following type of models:

$$dX = f(X; \Theta(t))dt + g(X; \Theta(t))dW(t), \quad X(t_0) = X_0, \quad (2.17)$$

$$\Theta : [0, t_e] \rightarrow \mathbb{R}^n, \quad \Theta(t) = [\theta_1(t), \theta_2(t), \dots, \theta_n(t)] \in \Omega_\Theta, \quad (2.18)$$

where the possibly nonlinear drift f and the nonlinear diffusion term g are linearly dependent on an unknown time-dependent vector function $\Theta(t)$, which in this case, are also the unknown parameters of the SDE. The vector $\Theta(t)$ summarizes the parameters of the SDE function f and g , where the index n in Eq. (2.18) indicates the total number of parameters.

The nonparametric clustering approach with smooth regularization, introduced by Horenko (2010a) (called Finite-Element, H^1 -Regularization (FEM- H^1)) and further developed by Pospisil et al. (2018), is adopted next to estimate the time-varying parameters. In general, it is an optimization framework with a specific regularization technique. Here the abbreviation Finite Element Method (FEM) indicates that the finite-dimensional projection space used for regularization is constructed by following the basic principles of the FEM. Moreover, " H^1 " symbolizes that a specific smoothness condition on the functions (affiliation function) are imposed.

2. Model-Based Nonstationary Time Series Modeling

Here, identifying $\Theta(t)$ using FEM- H^1 involves solving two different inverse problems coupled into one estimation procedure. The first problem is determining the parameter values for which the relevant procedure was discussed in Sec. 2.1.2. The second problem is to identify their locations in time, which is the focus of this section. The FEM approach reduces the inverse problem's complexity and the computational power required to determine the parameters $\Theta(t)$. Since the term FEM is usually associated with partial differential equations, it is emphasized that this is not the case here.

The formulation of the fitness function (2.5) for Eq. (2.17) becomes a problem. The problem of parameter estimation (see Eq. (2.6)) is ill-posed in such a case because one searches for a parameter vector size equal to the length of the time series times the number of model parameters. For the considered Fluxes Over Snow Surfaces Phase II (FLOSS2) data, this means about 10^6 parameter values and is intractable. To overcome this difficulty, Horenko proposes regularizing the fitness function by assuming that the process $X(t)$ varies much faster than the parameter function $\Theta(t)$. Formally, this means that the fitness function is expressed by a convex combination of K locally-optimal fitness functions,

$$l_N(\Theta(t)) := \sum_{i=0}^{N-1} \sum_{k=1}^K \gamma_k(t_i) \ell(t_i; \bar{\theta}_k), \quad K \geq 2, \quad (2.19)$$

where $\ell(t_i; \bar{\theta}_k)$ is the time-dependent fitness function formed by the time-averaged locally valid parameter $\bar{\theta}_k$. The overbar indicates the averaged value, which is assumed to be constant on a temporally local period. The value of K indicates the number of clusters or sub-models. The functions $\gamma_k(t)$ are the affiliation functions with specific properties that will be discussed below. They are unknown and are estimated differently than the model parameters. The function $\ell(\cdot; \cdot)$ is close related to Eq. (2.5) and is defined as

$$\ell(t_i; \bar{\theta}_k) := -\ln \left[p_X(\Delta t, X(t_{i+1}) | X(t_i); \bar{\theta}_k) \right] \quad i = 1, \dots, N-1, \quad (2.20)$$

where the index i iterates over the samples. To simplify the implementation process, the sample size of $\ell(\cdot; \cdot)$ is set equal to the sample size of $X(t)$ by setting the value at the last time step N equal to the first $\ell(t_N; \bar{\theta}_k) = \ell(t_1; \bar{\theta}_k)$.

In each cluster K the function value of ℓ varies depending on a different locally-valid parameter set $\bar{\theta}_k$, and therefore, when considering the particular periods in isolation, several local minima can occur within the entire fitness function. The intention of the clustering method is to discover local minima and thus estimate the corresponding parameter values.

The affiliation functions $\gamma_k(t)$ encode the information about the location of the minima for each of the k -th $f(\cdot; \cdot)$ and are stored in the so-called affiliation vector $\Gamma(t) = [\gamma_1(t), \gamma_2(t), \dots, \gamma_K(t)] \in \mathbb{R}^{K \times N}$. The functions $\gamma_k(t)$ satisfy the convexity constraints and thus form the partition of unity,

$$\sum_{k=1}^K \gamma_k(t) = 1, \quad \forall t, \quad (2.21)$$

$$1 \geq \gamma_k(t) \geq 0, \quad \forall t, k. \quad (2.22)$$

The vector $\mathbf{\Gamma}(t)$ has several possible interpretations. It can be interpreted as a vector of time-dependent weights. Following Pospisil et al. (2018), the vector $\mathbf{\Gamma}(t)$ represents the activity of a corresponding k -th SDE at given time t_i . It can also be understood as a K -dimensional time-dependent probability vector expressing the probability of observing the k -th stationary model at time t_i . For demonstrative reasons, assuming that we know in advance the affiliation vector, the temporal evolution of a parameter function is approximated as follows

$$\theta(t) \approx \sum_{k=1}^K \gamma_k(t) \bar{\theta}_k + \varepsilon(t) \quad (2.23)$$

where $\theta(t)$ is approximated by $K \in \mathbb{N}$ piecewise constant $\bar{\theta}_k$ (see Fig. 2.2) and where $\varepsilon(t)$ is an approximation error.

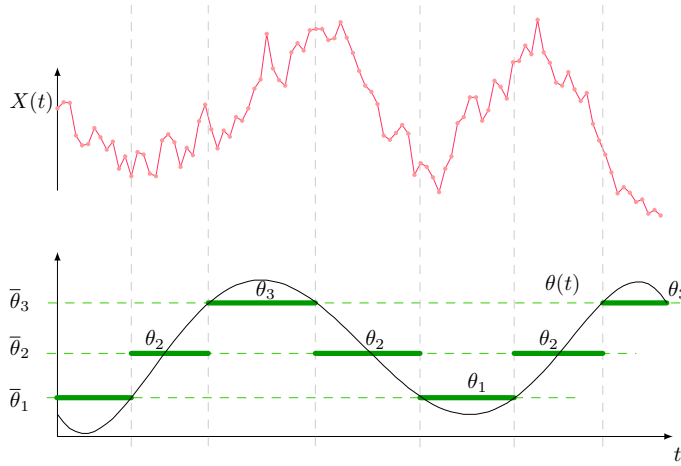


Figure 2.2: Sketch illustrating an approximation of a time-dependent parameter function with piecewise constant subdivisions. The upper panel shows a non-stationary time series $X(t)$ that is observed. The lower panel shows the time-dependent parameter of some model with a black line for $X(t)$ (which is not observed). The green horizontal lines indicate a possible approximation to the black line as an example with three constant values.

Figure 2.2 illustrates as an example the approximation of a nonstationary parameter with $K = 3$. The figure shows a variable position and period of the individual subdivisions, depending on the evolution of $\theta(t)$ and the number of clusters used. This information is contained in the $\mathbf{\Gamma}(t)$ function. For $K \rightarrow N$, the approximation of $\theta(t)$ becomes progressively refined. With each additional cluster added, another set of complementary parameters must be estimated. In the limiting case, when the number of clusters is equal to the number of points in the time series ($K = N$), each point should potentially be described by an equation with its unique parameter values. For example, a SDE model with 3 parameters and a time series of length N yields $3N$ parameters to be estimated. Such an approximation is interpreted as overfitting. Another consequence is that the uncertainty for each estimate of $\bar{\theta}_k$ increases (see Eq. (2.7)) because the set of available data points for each cluster shrinks (i.e., for each SDE model).

Following the definition of Horenko (2010a), the averaged clustering functional L_N together with the assumptions (2.19), (2.21), (2.22), and the fitness function Eq. (2.20) for the SDE is

2. Model-Based Nonstationary Time Series Modeling

defined as follows,

$$L_N(\mathbf{\Gamma}(t), \bar{\boldsymbol{\theta}}_1, \dots, \bar{\boldsymbol{\theta}}_K) = - \sum_{i=0}^{N-1} \sum_{k=1}^K \gamma_k(t_i) \ln \left[p_X(\Delta t, X(t_{i+1}) | X(t_i); \bar{\boldsymbol{\theta}}_k) \right]. \quad (2.24)$$

Some important properties are summarized below from the corresponding literature. Metzner et al. (2012) proved that the average cluster functional is the lower bound of the total negative log-likelihood by using JENSEN'S inequality. The functional $L_N(\mathbf{\Gamma}(t), \bar{\boldsymbol{\theta}}_1, \dots, \bar{\boldsymbol{\theta}}_K)$ is privileged over the functional $l_N(\boldsymbol{\theta}(t))$ for the following rationale. Instead of estimating $|\boldsymbol{\theta}(t)| = N \times n$ parameters (where n is the number of parameters in the basic model), one needs to estimate a reduced set $\bar{\boldsymbol{\Theta}} = (\bar{\boldsymbol{\theta}}_1, \dots, \bar{\boldsymbol{\theta}}_K)$ with a cardinality $|\bar{\boldsymbol{\Theta}}| = K \times n$ and affiliation vector $\mathbf{\Gamma}(t)$. One still needs to specify a hyperparameter K , but the number of parameters is reduced by several orders of magnitude, and the optimization problem is simplified. The entire variation problem changes to

$$(\mathbf{\Gamma}^*(t), \bar{\boldsymbol{\Theta}}^*) \in \arg \min_{\substack{\boldsymbol{\Theta} \in \Omega_{\boldsymbol{\theta}} \\ \mathbf{\Gamma} \in \Omega_{\mathbf{\Gamma}}}} L_N(\mathbf{\Gamma}(t), \bar{\boldsymbol{\theta}}_1, \dots, \bar{\boldsymbol{\theta}}_K), \quad (2.25)$$

where $\bar{\boldsymbol{\Theta}}^* = [\bar{\boldsymbol{\theta}}_1^*, \bar{\boldsymbol{\theta}}_2^*, \dots, \bar{\boldsymbol{\theta}}_K^*]$ denotes the estimate for the reduced set of averaged parameters, and $\mathbf{\Gamma}^*(t)$ denotes the estimate for the affiliation vector under the constraints (2.21) and (2.22). The vector $\mathbf{\Gamma}^*(t)$ still yields irregular solutions. To achieve permanent separation of K clusters, the problem (2.25) is further regularized by Horenko (2010a)'s TIKHONOV approach,

$$L_N^\epsilon(\mathbf{\Gamma}(t), \bar{\boldsymbol{\Theta}}, \epsilon^2) = L_N(\mathbf{\Gamma}(t), \bar{\boldsymbol{\Theta}}) + \epsilon^2 \sum_{k=1}^K \int_0^{t_e} \frac{d\gamma_k(t)}{dt} \frac{d\gamma_k(t)}{dt} dt, \quad (2.26)$$

where ϵ^2 is the regularization value and, together with K , represents a hyperparameter of the whole methodology. To find a reasonable solution vector $\mathbf{\Gamma}^*(t)$, one must specify the function space. For example, the original authors also used bounded variation functions to model jump transitions with $\gamma_k(t)$; this form of bounded variation is discussed in Sec. 2.2.2.

To parameterize the SDE, the $\gamma_k(t) \in H^1$ with the appropriate smoothing measure $\|\frac{d\gamma_k(t)}{dt}\|_{L^2}$ is convenient. One of the reasons for this is that the smooth change of $\gamma_k(t)$ offers us insights into the transition zones between the different regimes. It gives us, for example, the possibility to determine the time of the transitions and understand the sensitivity of the parameter excitation of the system. Smooth regime transitions are more typical in SBL.

After the construction of the functional (2.26) and the selection of the function space for $\gamma_k(t)$, it is relevant to discuss the discretization of $L_N^\epsilon(\cdot, \cdot, \cdot)$ by finite elements since the introduced fitness function (2.20) requires modifications to the numerical solution of the problem. For the initial discretization of the problem, see (Horenko, 2010a). The main focus here is on the interpolation and reduction steps, which are new compared to the original algorithm.

The reason for the discretization is to simplify the solution of the following quadratic programming problem. Recall that the data is sampled with a time step Δt . The discretization reduces the size of the time series by a specific factor $\alpha_N \in (0, 1)$. After the reduction, we work

with a reduced time series of length $\widehat{N} < N$ and a new time step $\Delta\tau > \Delta t$. The discretized version of $L_N^\epsilon(\cdot, \cdot, \cdot)$ is

$$L_{\widehat{N}}^\epsilon(\mathbf{\Gamma}(\tau), \overline{\boldsymbol{\Theta}}, \epsilon^2) = \sum_{k=1}^K \left[b(\tau, \overline{\boldsymbol{\theta}}_k)^\top \gamma_k(\tau) + \epsilon^2 \gamma_k(\tau)^\top \mathbf{A} \gamma_k(\tau) \right] \rightarrow \min_{\mathbf{\Gamma}(\tau), \overline{\boldsymbol{\Theta}}} . \quad (2.27)$$

where the vector $\mathbf{\Gamma}(\tau) = [\gamma_1(\tau), \gamma_2(\tau), \dots, \gamma_K(\tau)] \in \mathbb{R}^{K \times \widehat{N}}$ is the coarsened version of $\mathbf{\Gamma}(t)$. Here τ denotes the reduced t -grid by a factor $\alpha_N \in (0, 1)$ and distinguishes the notation of the vector $\mathbf{\Gamma}(t)$ from $\mathbf{\Gamma}(\tau)$. For more details on the FEM approximation of $\gamma_k(\tau)$ and the construction of \mathbf{A} , see (Horenko, 2010a) and (Metzner et al., 2012)).

The coarser and uniform mesh with its size $\Delta\tau$ is a set of points $(\tau_j)_{0 \leq j \leq \widehat{N}+1}$ with respective intervals $\mathcal{T}_j = [\tau_j, \tau_{j+1}]$ with $0 = \tau_0 < \tau_1 < \dots < \tau_{\widehat{N}+1} = t_e$ such that the resulting number of discretization points $\widehat{N} < N$. The LAGRANGE \mathbb{P}_1 elements associated with the space of globally continuous, locally linear functions for every interval $\Delta\tau$:

$$V_{\Delta\tau} = \{ \delta v_{\Delta\tau} \in C^0([0, t_e]) : \delta v_{\Delta\tau}|_{\mathcal{T}_j} \in \mathbb{P}_1, 0 \leq j \leq \widehat{N} \}, \quad (2.28)$$

with the subspace:

$$V_{0, \Delta\tau} = \{ \delta v_{\Delta\tau} \in V_{\Delta\tau} : \delta v_{\Delta\tau}(0) = \delta v_{\Delta\tau}(t_e) = 0 \}. \quad (2.29)$$

The first term, $b(\cdot, \cdot)^\top \gamma_k(\cdot)$ in Eq. (2.27) is the projected version of $L_N(\cdot, \cdot)$ (see Eq. (2.24)) and is formulated as a dot product. The symbol \top denotes the transpose of the corresponding vector. The vector $b(\tau, \overline{\boldsymbol{\theta}}_k)$ denotes the discretized model fitness function,

$$b(\tau, \overline{\boldsymbol{\theta}}_k) = \left(\int_{\tau_1}^{\tau_2} \ell(t, \overline{\boldsymbol{\theta}}_k) \delta v_1(t) d\tau, \dots, \int_{\tau_{\widehat{N}-1}}^{\tau_{\widehat{N}}} \ell(t, \overline{\boldsymbol{\theta}}_k) \delta v_{\widehat{N}}(t) d\tau \right), \quad (2.30)$$

where $\delta v(t)$ is the hat function and $\ell(\cdot, \cdot)$ the function in Eq. (2.20). The operation in Eq. (2.30) is termed a reduction and projects a vector into the space of the finite element. Consequently, the temporal evolution of the function value $\ell(\cdot, \cdot)$ changes its sampling rate from Δt to a larger mesh size $\Delta\tau$. The second term $\epsilon^2 \gamma_k(\tau)^\top \mathbf{A} \gamma_k(\tau)$ penalizes the functional by controlling the regularity of the vector $\mathbf{\Gamma}(\tau)$. By selecting the \mathbb{P}_1 elements, the respective stiffness matrix of a cluster \mathbf{H}_k is composed as follows

$$\mathbf{H}_k = \begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{pmatrix} \in \mathbb{R}^{\widehat{N} \times \widehat{N}}, \quad k = 1, \dots, K.$$

and the total stiffness matrix of the variational problem is

$$\mathbf{A} = \begin{pmatrix} \mathbf{H}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{H}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{H}_K \end{pmatrix} \in \mathbb{R}^{K\hat{N} \times K\hat{N}}.$$

In the literature, the regularized variational problem (2.27) is solved iteratively in two steps:

1. Fix $\mathbf{\Gamma}(t)$ and find $\bar{\mathbf{\Theta}}^*$ by solving $L_N(\mathbf{\Gamma}(t), \bar{\mathbf{\Theta}}) \rightarrow \min_{\bar{\mathbf{\Theta}}} \quad (2.24)$
2. Fix $\bar{\mathbf{\Theta}}$ and find $\mathbf{\Gamma}^*(\tau)$ by solving $L_{\hat{N}}^\epsilon(\mathbf{\Gamma}(\tau), \bar{\mathbf{\Theta}}, \epsilon^2) \rightarrow \min_{\mathbf{\Gamma}(\tau)} \quad (2.27)$

The second step is a quadratic cost function problem involving linear equality and inequality constraints. Equation (2.27) is reformulated by defining column vectors from K clusters. The so-called vector of modeling errors $\mathbf{B}(\tau)$ (Pospisil et al., 2018) is constructed using the reduced local fitness function see Eq. (2.30),

$$\mathbf{B}(\tau) := [b(\tau, \bar{\boldsymbol{\theta}}_1), b(\tau, \bar{\boldsymbol{\theta}}_2), \dots, b(\tau, \bar{\boldsymbol{\theta}}_K)] \in \mathbb{R}^{K\hat{N}}, \quad (2.31)$$

From Eq. (2.27) the block-structured Quadratic Programming (QP) problem is:

$$\mathbf{\Gamma}^*(\tau) := \arg \min_{\mathbf{\Gamma} \in \Omega_\Gamma} L_{\hat{N}}^\epsilon(\mathbf{\Gamma}(\tau), \bar{\mathbf{\Theta}}, \epsilon^2), \quad (2.32)$$

$$L_{\hat{N}}^\epsilon := \frac{1}{\hat{N}} \mathbf{B}(\tau)^\top \mathbf{\Gamma}(\tau) + \frac{1}{\hat{N}} \epsilon^2 \mathbf{\Gamma}(\tau)^\top \mathbf{A} \mathbf{\Gamma}(\tau), \quad (2.33)$$

$$\Omega_\Gamma := \{\mathbf{\Gamma}(\tau) \in \mathbb{R}^{K\hat{N}} : \mathbf{\Gamma}(\tau) \geq 0 \wedge \sum_{k=1}^K \gamma_k(\tau) = \mathbf{1} \forall t \in [0, T]\}, \quad (2.34)$$

Where $1/\hat{N}$ is an introduced scaling coefficient to avoid large numbers. The fitted spectral projected gradient method developed by Pospisil et al. (2018) is applied to solve the QP. Their approach is characterized by a high granularity of the parallelization and is suitable for execution on Graphics Processing Unit (GPU) clusters. It is also efficient and outperforms traditional denoising algorithms in terms of the signal-to-noise ratio.

Reducing the samples size to $\hat{N} < N$ decreases the size of the QP problem and its computational complexity. Alternatively, higher-order elements also require additional collocation points but provide smoother approximations. However, the accuracy in determining $\bar{\mathbf{\Theta}}^*$ depends on the number of available observations since it scales with $1/\sqrt{\hat{N}}$ (see Eq. (2.7)) and therefore must be calculated with the smallest available time step Δt . If we now reduce the size of the QP problem (see Eq. (2.30)), the accuracy of the estimate $\bar{\boldsymbol{\theta}}^*$ will decrease, and this is undesirable. Consequently, we cannot solve both problems with the same time step $\Delta \tau$ and must interpolate the solution of QP at each iteration step between $\Delta \tau$ and Δt . This results in the following modification of the original procedure:

1. Fix $\mathbf{\Gamma}(t)$ and find $\bar{\mathbf{\Theta}}^*$ by solving $L_N(\mathbf{\Gamma}(t), \bar{\mathbf{\Theta}}) \rightarrow \min_{\bar{\mathbf{\Theta}}}$
2. reduce the fitness function to the τ -grid

3. Fix $\bar{\Theta}$ and find $\Gamma^*(\tau)$ by solving $L_N^\epsilon(\Gamma(\tau), \bar{\Theta}, \epsilon^2) \rightarrow \min_{\Gamma(\tau)}$
4. interpolate $\Gamma^*(\tau)$ to the t -grid step size Δt

Algorithm 1: The adapted subspace algorithm for SDE models. The main change to Horenko (2010a)'s algorithm is in line 9, where the functional L_N is formulated for a general SDE. Furthermore, lines 14 and 18 are added to maintain the accuracy of the estimated $\bar{\Theta}^*$. The reduced variables are indicated by the time variable τ and the non-reduced with t , respectively.

```

Input : Time series  $X$ , number of clusters  $K$ , regularization value  $\epsilon^2$ , and the
          reduction value  $\alpha_N$ 
Output:  $\Gamma^*(t)$  and  $\bar{\Theta}^*$ 
1 Generate random initial  $\Gamma^0(t)$  satisfying (2.21) and (2.22)
2 /* We need the full  $\Gamma^0(t)$  to find  $\bar{\Theta}$  and the reduced  $\Gamma^0(\tau)$  to solve the QP
   problem */
3 Reduce the initial vector  $\Gamma^0(t)$  to  $\Gamma^0(\tau)$  on the coarser grid  $\Delta\tau$ 
4 Set iteration counter  $j = 0$  for the main optimization loop
5 /* main optimization loop */
6 while  $|L_N^\epsilon(\Gamma^{j+1}(\tau), \bar{\Theta}^{j+1}, \epsilon^2) - L_N^\epsilon(\Gamma^j(\tau), \bar{\Theta}^j, \epsilon^2)| > tol$  do
7     /* Estimate  $\bar{\Theta}^j$  from the full (This will ensure maximum accuracy for
       the estimate of  $\bar{\Theta}^j$ )  $\Gamma^j(t)$  applying  $K$  times unconstrained
       minimization */
8     for  $k \leftarrow 1$  to  $K$  do
9          $\bar{\theta}_k^{j+1} = \arg \min_{\theta} L_N(\gamma_k^j(t), \theta)$  // See Eq. (2.24).
10    end for
11    /* From the found  $\bar{\theta}_k^{j+1}$  compute the fitness functions  $f^{j+1}(t, \bar{\theta}_k^{j+1})$  or
       save it from the step in line 9 */
12    Compute  $f^{j+1}(t, \bar{\theta}_k^{j+1})$ , if not saved from line 9 //
13    /* Reduce  $f^{j+1}(t, \bar{\theta}_k^{j+1})$  because  $\Gamma(\tau)$  will be estimated on the coarser
       grid  $\Delta\tau$  */
14    Compute  $b^{j+1}(\tau, \bar{\theta}_k^{j+1}) = reduction(f^{j+1}(t, \bar{\theta}_k^{j+1}), \Delta\tau, \alpha_N)$  // See Eq. (2.30)
15    /* Apply quadratic programming to solve the constrained minimization
       */
16     $\Gamma^{j+1}(\tau) = \arg \min_{\Gamma} L_N^\epsilon(\Gamma(\tau), b^{j+1}(\tau, \bar{\theta}_k^{j+1}), \epsilon^2)$  satisfying (2.21) and (2.22)
17    /* Interpolate  $\Gamma^{j+1}(\tau)$  onto the original grid with a scale  $\Delta t$  */
18     $\Gamma^{j+1}(t) = interpolate(\Gamma^{j+1}(\tau), \Delta t)$ 
19    /* Incrementing the counter  $j$  of the main loop */
20     $j = j + 1$ 
21 end while
22 return  $\Gamma^{j+1}(t), \bar{\Theta}^{j+1}$ 

```

In this way, the discretization step of the theta solver (Δt) is decoupled from the discretization step of the gamma solver ($\Delta\tau$), ensuring an accurate approximation of $\bar{\Theta}$ while preserving the advantages of FEM. Algorithm 1 to solve the variational problem (2.27) is called the subspace clustering algorithm (Horenko, 2010a) and is reproduced with the proposed modification.

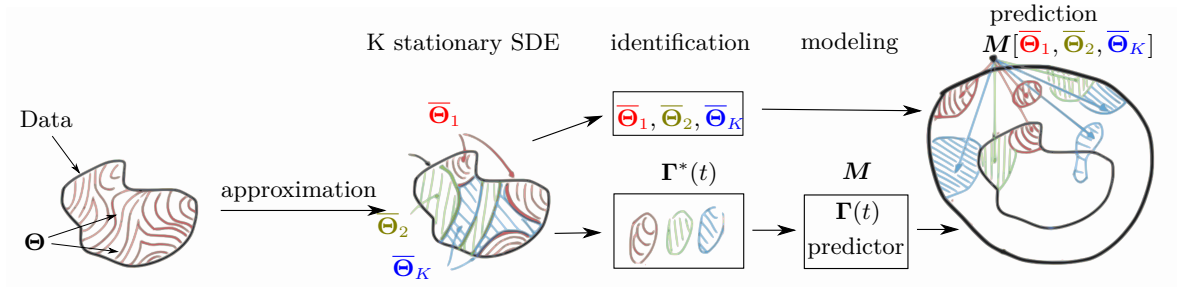


Figure 2.3: The infographic shows the workflow to construct a self-contained prediction model by applying additional stochastic modeling of the affiliation function with a MARKOV chain approach. The stochastic predictor of the affiliation function \mathcal{M} provides a prediction of the regime occupation on average. For more details, see (Metzner et al., 2012).

After the successful application of the clustering method, there are several options for further proceeding. For example, the vector $\Gamma^*(t)$ provides a classifier that can be used to analyze the quantities of interest with respect to the clusters. Instead, the idea is to use the vector $\Gamma^*(t)$ to construct stochastic closures. One difficulty here is the lack of information about the future realization of this vector. Vector $\Gamma^*(t)$ is an estimate based on the data used in the identification. Although we assume that we have a sufficient amount of data to capture the full spectrum of regime dynamics, we cannot make predictions of the future state of the vector $\Gamma(t)$ without additional modeling work. One can associate the affiliation function with a MARKOV process under the assumption that it is homogeneous and stationary (Metzner et al., 2012). Such assumptions are hard to imagine in SBL because the regime transitions might depend on the larger scales and the thermodynamical processes. Alternatively, Kaiser et al. (2017) used a neural network in the context of clustering to predict the velocity components of internal gravity waves outside the training data set.

Here, an approach is proposed in which the clustered parameter values are traced back to an auxiliary process $u(t)$. This approach is attractive because the deterministic process $u(t)$ should live on a larger scale than the process of interest $X(t)$. The only limitation, then, is that it is not always possible to find such a process. In the case of existence, however, no further assumptions are required.

The section is concluded by presenting a summarizing abstract sketch that shows the scope of the nonparametric clustering methodology (see Fig. 2.3). The figure shows a possible further modeling process to create a fully predictive model.

2.2.2 Bounded Variation of Linear Discrete-Time Models

This section provides a brief overview of the clustering framework based on the Finite-Element, Bounded-Variation, Vector AutoRegressive (X) Factor (FEM-BV-VARX) method. The FEM-BV-VARX method is applied in this work to analyze the regimes of the SBL (see Ch. 3), whereas the new version of this methodology, presented in Sec. 2.2.1 and referred to as Finite-Element, H^1 -Regularization, Stochastic Differential Equation (FEM- H^1 -SDE), is primarily used to develop the stochastic turbulence closure in Ch. 4.

The FEM-BV-VARX method (Horenko, 2010b) is widely used to study multidimensional transient atmospheric data. For example, Franzke et al. (2015) examined the response of Southern Hemisphere circulation trends to several external forcings and found that anthropogenic CO₂ is a more relevant driver than ozone depletion. Risbey et al. (2015) examined mid-tropospheric flows and showed that teleconnection patterns in the Northern Hemisphere exhibit a characteristic alternation of metastable states. O’Kane et al. (2017) investigated the influence of memory and dimensionality related to the FEM-BV-VARX method in defining the quasi-stationary states of the troposphere.

The mathematical formulation of the FEM-BV-VARX tool is based on the discrete-time linear models presented in Sec. 2.1.1 and uses the basic variational clustering principle given in Sec. 2.2.1, but with some modifications. Igdalov et al. (2019) is one of the authors who developed the version of the FEM-BV-VARX code applied in this work.

Based on the theory presented in Sec. 2.2.1, the components characterizing the FEM-BV-VARX method are as follows. The FEM approximation remains in operation to reduce the size of the problem, projecting the vector of modeling errors onto the finite element basis. The regularization is performed with a class of functions with Bounded Variation (BV), which leads to the so-called persistence condition:

$$\|\gamma\|_{\text{BV}} = \sum_{i=0}^{N-1} |\gamma_k[t_{i+1}] - \gamma_k[t_i]| \leq C_p \quad k = 1, \dots, K \quad (2.35)$$

where the hyperparameter $C_p \geq 0$ restricts the maximum number of jumps allowed in the affiliation vector $\mathbf{\Gamma}(t)$. By implementing such a condition in the regularization term in Eq. (2.26), the parameter C_p replaces the functionality of the parameter ϵ^2 . For more details and numerical implementation, see (Metzner et al., 2012, p. 189). The bounded variation offers two benefits compared to the smooth regularization approach. First, the regularization is based on the control of the hyperparameter that sets the maximum allowed regime transitions rather than on the smoothness of the affiliation vector. Second, the affiliation function is of the discrete type, i.e., the transitions between regimes occur as discrete switches requiring only one discrete time step. Such an affiliation function includes smooth transitions (Metzner et al., 2012) but does not provide detailed information about how the transition unfolds.

Despite these exciting features of the regularization with BV, the smooth regularization was used for clustering the SDE for the following reason. The main drawback of BV regularization is that it provides a discrete classifier. As shown in Fig. 2.13a, BV regularization obscures vital information about the probability of observing the regimes. For example, in regime states with probability 60% might be too uncertain, and one might exclude them from further analysis. Similarly, if one observes a short regime state and wants to know how much of that

time was spent fading in and out. All this information is not directly available with the BV regularization approach.

Finally, the fitness function in the FEM-BV-VARX method is in vector form. All parameters are matrices, and therefore the procedure can be applied to multi-dimensional data. In this case, the off-diagonal entries in the parameter matrices indicate the cross-dimensional time-delayed interactions. The fitness function takes the form

$$\ell(\mathbf{Y}; \boldsymbol{\theta}) := \left\| \mathbf{Y}[n] - \mathbf{A}_0[n] - \sum_{i=1}^p \mathbf{A}_i[n] \mathbf{Y}[n - i\Delta t] - \sum_{j=1}^q \mathbf{B}_j[n] \mathbf{X}[n - j\Delta t] \right\|_{\mathbf{P}(t)}, \quad (2.36)$$

with $n = 0, \dots, N - 1$, and where $\boldsymbol{\Theta} = (\mathbf{A}_0, \mathbf{A}_i, \mathbf{B}_j)$ ($i = 1, \dots, p$ and $j = 1, \dots, q$) is the vector of the model parameter matrix. The bold notation indicates that the modeled time series \mathbf{Y} is multidimensional. Assuming the weighted norm $\|\cdot\|_{\mathbf{P}(t)} = \sqrt{\|\cdot \mathbf{P}(t), \mathbf{P}^\top(t) \cdot\|_2}$, the norm for the least-squares problem is preserved in the essential noise dimensions. The matrix $\mathbf{P}(t)$ is orthogonal and multiplied by the diagonal matrix $\boldsymbol{\Lambda}(t)$ to get the noise matrix

$$\mathbf{C}(t) = \mathbf{P}(t)\boldsymbol{\Lambda}(t). \quad (2.37)$$

The matrix $\mathbf{C}(t)$ is the residual between the estimated model and the observed time series. For more details, see (Horenko, 2010b). The FEM-BV-VARX method is obtained by inserting Eq. (2.36) into Eq. (2.19) and using the functions with bounded variations in Eq. (2.26) to express the affiliation vector.

The individual model structures of the FEM-BV-VARX are presented in the corresponding Ch. 3, where the clustering approach is used for data analysis.

2.3 Parametrization of SDEs with an External Process

In this work, the formulation of turbulence closure is based on the determination of a functional dependence from data between the parameters of an SDE and some arbitrary auxiliary process $u(t)$, which is predictable and operates on a time scale larger than the scale of the process $X(t)$. It is assumed that the process $u(t)$ is known and related to the dynamics of $X(t)$ in some way. Otherwise, one must sequentially probe each candidate for the role of $u(t)$. This section shows that the classifier $\boldsymbol{\Gamma}^*(t)$ is a natural component for identifying functional relationships between variables of different scales.

The classifier estimate $\boldsymbol{\Gamma}^*(t)$ found by the algorithm 1 is utilized to formulate a LEBESGUE integral. Then, it is shown that by conserving these integral values, one can account for a possible mismatch of units, which would require a search for a functional relation between some model parameters $\bar{\theta}_k^*$ and $u(t)$. Furthermore, it is shown that with an increasing number of clusters, the resolution of the functional link between $\bar{\theta}_k^*$ and $u(t)$ grows.

Let us denote the temporal space by \mathcal{T} , the σ -algebra by \mathcal{F} , which has the subsets A_j of this space, and a measure μ which is defined on \mathcal{F} . For the sake of simplicity and illustration, consider only one function of the parameter vector $\boldsymbol{\Theta}$ and denote it by $\theta : \mathcal{T} \rightarrow \mathbb{R}$. Moreover, the auxiliary function $u : \mathcal{T} \rightarrow \mathbb{R}$. To define the LEBESGUE integral, consider the sets $\{A_j : 1 \leq j \leq m\}$ partitioning the space \mathcal{T} such that. $A_j \cap A_i = \emptyset$ for $j \neq i$. Then the

approximation reads:

$$\mathcal{I}(\theta) \approx \sum_{j=1}^m \tilde{\theta}_j \mu(A_j) \approx \sum_{j=1}^m \tilde{\theta}_j \mu(\theta^{-1}(J_j)), \quad (2.38)$$

where $\mathcal{I}(\cdot)$ is the value of the integral calculated over an interval, $\tilde{\theta}_j$ is the function value of θ within the subpartition J_j , and the set A_j is the inverse image of the subpartition J_j . The measure μ assigns a value to the quantity A_j . Increasing the number of partitions J_j improves the approximation of the integral value. To calculate the values $\mu(A_j)$, it is necessary to define the so-called indicator functions $I_{A_j} : \mathcal{T} \rightarrow \mathbb{R}$,

$$I_{A_j}(t) = \begin{cases} 1 & \text{if } t \in A_j \\ 0 & \text{otherwise} \end{cases} \quad (2.39)$$

which describe the position of the set A_j in time.

Next, consider the estimated $\mathbf{\Gamma}^*(t)$, which by construction is a partition of unity and at the same time a partition of space \mathcal{T} . It is used to define the LEBESGUE integral on the interval $[0, t_e]$. In this context, the valuable property of $\mathbf{\Gamma}^*(t)$ is the grouping of sets A_j into K clusters.

$$\{A_j^{(k)} : 1 \leq j \leq m_k ; 1 \leq k \leq K\}, \quad (2.40)$$

where index k denotes the group of sets identified as a cluster, index j denotes a set of the cluster k , and m_k is the maximum number of sets in cluster k . Using the estimated $\mathbf{\Gamma}^*(t)$, the ordered partition of the space \mathcal{T} is

$$\{A_1^{(1)}, \dots, A_{m_1}^{(1)}, A_1^{(2)}, \dots, A_{m_2}^{(2)}, \dots, \dots, A_1^{(k)}, \dots, A_{m_k}^{(k)}\} \quad k = 1, \dots, K, \quad (2.41)$$

such that $A_j^{(k)} \subseteq \mathcal{T}$, $A_j^{(k)} \in \mathcal{F}$ and $A_j^{(k)} \cap A_i^{(p)} = \emptyset$ for $j \neq i$ and $k \neq p$. To define the nonoverlapping sets $A_j^{(k)}$ the $\mathbf{\Gamma}^*(t)$ is limited with a value $c_{\text{th}} \in (0, 1)$, leading to the definition of the clustered indicator functions

$$I_{A_j^{(k)}}(t) := \begin{cases} 1 & \text{if } t \in A_j^{(k)} \Leftrightarrow \gamma_k(t) > c_{\text{th}}, \\ 0 & \text{otherwise.} \end{cases} \quad (2.42)$$

The value of a LEBESGUE integral is then expressed approximately as

$$\int \theta(t) \mu(dt) \approx \sum_{k=1}^K \sum_{j=1}^{m_k} \tilde{\theta}_j \mu(A_j^{(k)}). \quad (2.43)$$

It is important to note that the approximation accuracy is independent of the number of clusters K since the corresponding sum only specifies an order in which the integral is composed. However, the approximation quality depends on the estimate $\mathbf{\Gamma}^*(t)$ and thus on the division of the space \mathcal{T} . The uncertainty in $\gamma(t)$, where it is < 1 , indicates the deviations from the averaged values of the clusters. For example, in transition regions between regimes or when several clusters occupy the same period (violating the rule of non-overlapping sets in the

definition of the LEBESGUE integral). The threshold c_{th} provides a way to select precisely clustered sets A_j and reduce uncertainties.

The affiliation function enables the division of the LEBESGUE integral into the individual parts:

$$\int_0^{t_e} \theta \mu(dt) \approx \int_{A^{(1)}} \theta \mu(dt) + \int_{A^{(2)}} \theta \mu(dt), \dots, \int_{A^{(k)}} \theta \mu(dt) \quad k = 1, \dots, K. \quad (2.44)$$

Recall that the MLE (2.5) is an approximation to a time-series averaged $\bar{\theta}^*$ and can also be expressed as an integral. Algorithm 1 yields the decomposition of the integral value into individual subsets with the help of the estimated affiliation function

$$\int_0^{t_e} \theta \mu(dt) = \frac{\sum_k \int_{A^{(k)}} \theta \mu(dt)}{\sum_k \int_{A^{(k)}} \mu(dt)} = \sum_k \bar{\theta}_k^* \quad k = 1, \dots, K, \quad (2.45)$$

where $\bar{\theta}_k^*$ is the estimate obtained in solving the problem (2.25). To parameterize the time dependence of θ by a deterministic process u it is demanded:

$$\int_0^{t_e} \theta \mu(dt) \stackrel{!}{=} \int_0^{t_e} u \mu(dt) \quad (2.46)$$

If the variables have units, at this point, the units on the left must correspond to the units on the right. One should include problem-specific characteristic variables such as characteristic lengths, times, or masses to resolve any unit specific inconsistencies. The trivial case is a one-to-one relationship between the functions θ and u . In general, however, we consider splitting the integrals and scaling the respective partial integrals by a factor s_k to achieve partial equality. The left-hand side in Eq. (2.46) is known by the estimated parameters (see Eq. (2.25)) and given directly by algorithm 1. Accordingly, the right side is evaluated using the vector $\mathbf{\Gamma}^*(t)$, which is used to construct the sets $A^{(k)}$:

$$\bar{\theta}_1^* + \bar{\theta}_2^* + \dots + \bar{\theta}_k^* = s_1 \frac{\int_{A^{(1)}} u \mu(dt)}{\int_{A^{(1)}} \mu(dt)} + s_2 \frac{\int_{A^{(2)}} u \mu(dt)}{\int_{A^{(2)}} \mu(dt)} + \dots + s_k \frac{\int_{A^{(k)}} u \mu(dt)}{\int_{A^{(k)}} \mu(dt)} \quad k = 1, \dots, K. \quad (2.47)$$

The quantities $A^{(k)}$ allow us to calculate the scaling factors s_k since the corresponding values on the left and right sides in Eq. (2.47) are known. Taking the limit over the number of clusters k , one redefines the factors s_k to a mapping $S : u \rightarrow \theta$, where $u = 0$ is excluded. The rate at which the number of partitions A_j within a cluster k increases is larger than that of the number of clusters. Or in other words, the convergence of the definition of LEBESGUE integral Eq. (2.38) must be significant to guarantee its convergence in the first place.

To construct a stochastic data-driven closure, consider performing the following procedures. For simplicity, assume that we observe a slowly evolving, known process $u(t)$ and a fast, nonstationary, stochastic process $X(t)$ described by some nonlinear SDE. Assume further that we have expert knowledge, intuition, or a first principle approach that suggests that $u(t)$ controls the nonstationary process $X(t)$ by modulating its model parameters. The approach to

parameterization is as follows. First, one estimates the time-varying parameters of the SDE:

$$dX = f(X; \Theta(t))dt + g(X; \Theta(t))dW_t, \quad X(t_0) = X_0, \quad (2.48)$$

$$\Theta : [0, t_e] \rightarrow \mathbb{R}^n, \quad \Theta(t) = [\theta_1(t), \theta_2(t), \dots, \theta_n(t)] \in \Omega_\Theta, \quad (2.49)$$

using the clustering approach presented in Sec. 2.2.1, i.e., by solving the variation problem:

$$(\mathbf{\Gamma}^*, \bar{\Theta}^*) \in \arg \min_{\substack{\Theta \in \Omega_\Theta \\ \mathbf{\Gamma} \in \Omega_\mathbf{\Gamma}}} L_{\hat{N}}^\epsilon(\mathbf{\Gamma}, \bar{\Theta}, \epsilon^2), \quad (2.50)$$

where the number of clusters $K \geq 2$. The functional $L_{\hat{N}}^\epsilon$ is defined in Eq. (2.27).

Second, using $\mathbf{\Gamma}^*(t)$, we estimate the cluster-related values of the process $u(t)$ by calculating the weighted average:

$$\bar{u}_k = \frac{\sum_{i=1}^N u(t_i) \gamma_k(t_i)}{\sum_{i=1}^N \gamma_k(t_i)}, \quad k = 1, \dots, K. \quad (2.51)$$

where \bar{u}_k denotes the averaged value of the auxiliary process $u(t)$ in a cluster k . Recall that $\gamma_k(t_i)$ is the affiliation function for cluster k (see Fig. 2.9a the color-coded regions). In the last step, one performs regression analysis to parameterize the parameter variability with some appropriate functions S_n that map the values of $u(t)$ to $\theta_n(t)$:

$$S_n : \bar{u}_k \rightarrow \bar{\theta}_{k,n}, \quad \bar{\theta}_k = [\bar{\theta}_{k,1}, \bar{\theta}_{k,2}, \dots, \bar{\theta}_{k,n}], \quad k = 1, \dots, K \quad n = 1, \dots, m. \quad (2.52)$$

The differentiability of the functions S_n is unconstrained. Parameterization is performed separately for each element of the parameter vector Θ . The vector of scaling functions $\mathbf{S} = [S_1, S_2, \dots, S_n] \in \mathbb{R}^n$ defines the modulation of model parameters through a process $u(t)$. The complexity of the relationship depends on the application and should be investigated individually for given problems. If all scaling functions could be retrieved, we obtain a closed-form system to predict the process

$$dX = f(X; \mathbf{S}(u(t)))dt + g(X; \mathbf{S}(u(t)))dW_t, \quad X(t_0) = X_0, \quad (2.53)$$

where $u(t)$ is known for every t . The approach allows modeling of non-equilibrium processes characterized by nonlinearity in f and g for those cases where the nonstationarity of the process is challenging to understand. Figure 2.4 summarizes the idea of this concept. In practice, it is assumed that the functional relationship between the process $u(t)$ and the SDE's modulated parameters $\Theta(t)$ holds for the estimated range.

2.4 Estimation of Hyperparameters for the Clustering Framework

Suitable model configuration includes the determination of the parameter ϵ^2 that regulates the smoothness of the affiliation vector, the determination of the number of clusters K , and the selection of the functional forms $f(\cdot; \cdot)$ and $g(\cdot; \cdot)$. A variety of strategies with examples are described in Horenko (2010a,b); Metzner et al. (2012), and are based on the Information

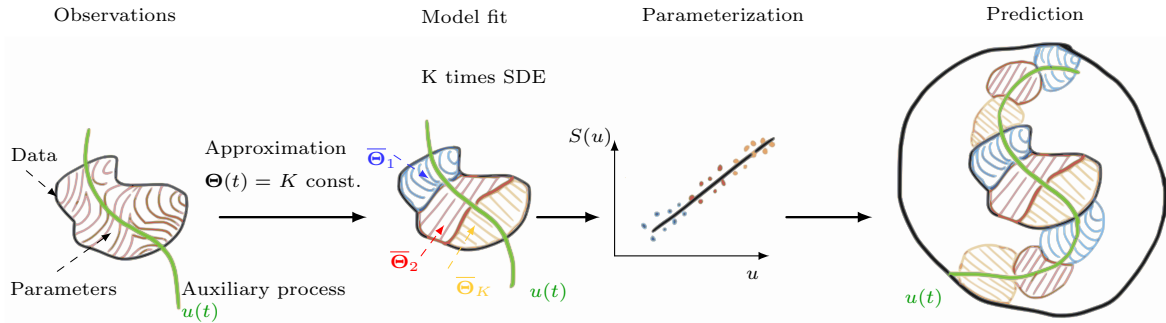


Figure 2.4: The infographic shows a workflow for building a self-contained predictive model with the FEM- H^1 -SDE method by regressing time-varying parameters on a deterministic process $u(t)$. The functional relationship $S(u)$ is hidden and must be discovered a posteriori.

Criteria approaches. New methods that involve the selection of ϵ^2 and K independently are developed and tested next. The parameters are estimated independently from each other and provide a robust estimation. The method of selecting the parameter ϵ^2 takes into account the frequency information of the auxiliary process and thus allows to establish a correspondence between the parameters and the descriptive process $u(t)$.

2.4.1 Optimal Regularization Parameter

Horenko (2010a) analyzed the influence of the parameter ϵ^2 on the regularity of the cluster affiliation function $\mathbf{\Gamma}^*(t)$. In summary, the optimal parameter ϵ_{opt}^2 is characterized by a sharp disjunction between the individual functions $\gamma_k(t)$ so that the regime transitions are as short as possible. Conversely, poor segmentation exists when two or more functions $\gamma_k(t)$ occupy the same period. Recall, the functions $\gamma_k(t)$ obey the convexity condition (see Eqs. (2.21); (2.22)) and if they have the same time frame, it means that none of the model parameters fit the data sufficiently well. Such results indicate that the number of clusters is overstated or that the selected model structure is inappropriate for the data under consideration. Thus, a relevant consideration in picking an appropriate value for ϵ^2 is tracking the regularity of the affiliation function $\mathbf{\Gamma}^*(t)$. Figure 2.5 shows a part of $\mathbf{\Gamma}^*(t)$ with $K = 3$ for a well-posed toy example.

To motivate the approach in searching for ϵ_{opt}^2 , we consider $\mathbf{\Gamma}^*(t)$ for different parameter values (see Figs. 2.5a, b, c). Consider the white regions between function values 0 and 1 which are uncolored. The cumulative size of these regions is dissimilar in each of the panels and is approximately the smallest at the value of ϵ_{opt}^2 (panel b). If we consider the function $\mathbf{\Gamma}^*(t)$ to be an oscillatory signal, then, to minimize the white area, we seek a $\mathbf{\Gamma}^*(t)$ with the largest energy. One can easily calculate and plot the signal energy of the affiliation vector over ϵ^2 and find a maximum as follows.

$$E_\gamma(\epsilon^2) = \frac{1}{K} \sum_{k=1}^K \int_0^{t_e} \gamma_k(t, \epsilon^2)^2 dt, \quad (2.54)$$

where $\gamma_k^2(t, \epsilon^2)$ denotes the solution of Eq. (2.27) for a given ϵ^2 value. The maximum of the curve $E_\gamma(\epsilon^2)$ characterizes the optimal regularization value (see Fig. 2.6 color code value 0). The current affiliation function used to generate the training data consists of clear transitions and without any irregularities. In addition, the same SDE model is used for parameter

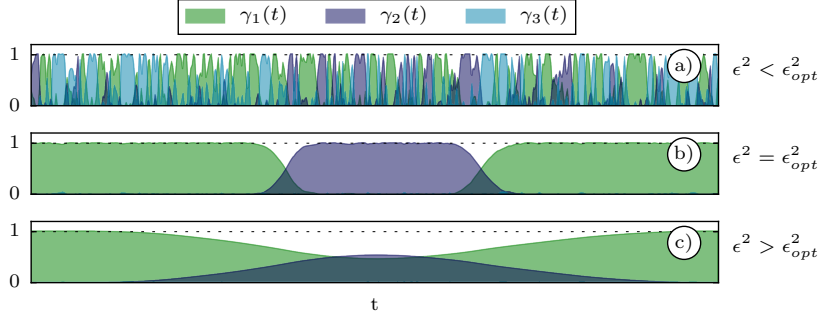


Figure 2.5: Effect of the regularization value on the estimate of $\Gamma^*(t)$. Each panel shows the same time window but for different values of ϵ^2 . Panel a) shows weak regularization: the regime periods are irregular and noisy. Their time scale is close to dt . Panel b) shows the optimal regularization values: the $\gamma_1(t)$ and $\gamma_2(t)$ functions show a relatively sharp and clear separation. Panel c) shows too strong regularization: the function $\gamma_2(t)$ is blurred and suppressed.

identification and training data generation. In other words, it's a well-posed example. However, when working with actual, unexplored data, we may, unfortunately, be unaware of the true functional form of the SDE, or the data may exhibit measurement errors. This potentially leads to a scenario where the functions $\gamma_k(t)$ become indistinct and cause the curve $E_\gamma(\epsilon^2)$ to become less smooth and lose a pronounced maximum. To make the procedure more robust, we investigate the frequency content of the vector $\Gamma^*(t)$.

One way to emphasize the maximum of the curve $E_\gamma(\epsilon^2)$ is to gradually eliminate the high-frequency oscillations in $\Gamma^*(t)$. The frequencies to be eliminated imply rapid state transitions. Since the variability of the parameters would have to be tied to a slow scale, rapid state transitions should be avoided.

Since the affiliation function takes step-like shapes in optimality, a multiresolution analysis with wavelets is performed (Kumar & Fofoula-Georgiou, 1997; Mallat, 1999). The affiliation vector is decomposed using the discrete wavelet transform to obtain C levels with $C+1$ frequency bands. The function $\gamma_k(t)$ is then expressed as the sum of the low-frequency component $\gamma_{k,C}^A(t)$ after C stages of transformations and the sum of the high-frequency components $\gamma_{k,c}^D(t)$ over the previous decomposition stages

$$\gamma_k(t) = \gamma_{k,C}^A(t) + \sum_{c=1}^C \gamma_{k,c}^D(t). \quad (2.55)$$

The HAAR wavelet is chosen as the basis function for filtering. This wavelet accurately represents the affiliation vector's sharp jumps, especially when approaching the optimal parameter value ϵ_{opt}^2 . Using the orthogonality property of HAAR wavelets, one substitutes Eq. (2.55) into (2.54) and obtains

$$E_\gamma(\epsilon^2) = \frac{1}{K} \sum_{k=1}^K \int_0^{t_e} [\gamma_{k,C}^A(t)]^2 dt + \frac{1}{K} \sum_{k=1}^K \int_0^{t_e} \left[\sum_{c=1}^C \gamma_{k,c}^D(t) \right]^2 dt. \quad (2.56)$$

The signal $\gamma_k(t)$ is decomposed with the maximum possible number of levels C , depending on the number of samples. The first term in Eq. (2.56) is the mean value of the entire time

2. Model-Based Nonstationary Time Series Modeling

series and is irrelevant for the purpose. The property of local support of wavelets allows for extracting out the sum sign from the integral,

$$E_\gamma(\epsilon^2) = \frac{1}{K} \sum_{k=1}^K \sum_{c=1}^C \int_{-\infty}^{\infty} [\gamma_{k,c}^D(t)]^2 dt. \quad (2.57)$$

To perform the filtering, the sample size of the vector $\gamma_k(t)$ is expanded to the nearest value 2^i using the reflect padding method, where $i \in \mathbb{N}^+$. The continuous integration is expressed in a discrete form as a sum

$$E_\gamma(\epsilon^2) = \frac{1}{K} \sum_{k=1}^K \sum_{c=1}^C \sum_{n=1}^{C_n} [\gamma_{k,c}^D(n)]^2 \Delta t, \quad (2.58)$$

where C_n denotes the number of coefficients at the decomposition level c . Due to the efficient implementation of the multiresolution decomposition, the number of coefficients at each level is reduced sequentially. With successive decomposition levels, the cutoff frequency gets halved, and the energy of the resulting frequency bands can be canceled by setting the corresponding filter coefficients to zero.

To magnify the optimum in the choice of the regularization parameter, the optimal cutoff frequency can be estimated in several ways. A rational choice would be to place it at the scale of the spectral gap between the process $X(t)$ and $u(t)$. A spectral gap exists when the smallest scale of $u(t)$ is larger than the largest scale of $X(t)$. In this way, the scale of the process $u(t)$ limits the smallest possible scale of regime jumps in $\Theta(t)$.

The curve $E_\gamma(\epsilon^2)$ naturally falls off at higher values of ϵ^2 due to the over-smoothing of the vector $\mathbf{\Gamma}^*(t)$. The shape of the smoother transitions differs from the sharper shape of the HAAR wavelet. Consequently, the wavelets stop registering energy at their respective scales. As a result, the energy curves are plotted by successively removing the high frequencies and looking for a maximum magnification without changing its position (see Fig. 2.6). The multiresolution analysis is performed with the help of the Python library PyWavelets (Lee et al., 2019a).

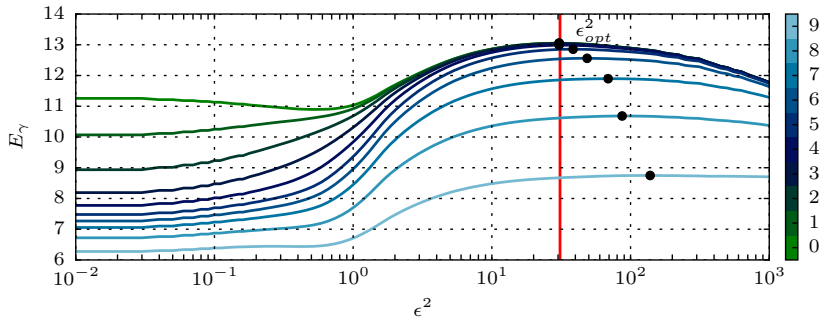


Figure 2.6: Estimation of the optimal regularization parameter ϵ^2 by filtering the affiliation function for a toy example with three predefined clusters. The y-axis is the cluster-averaged signal energy of $\mathbf{\Gamma}^*(t)$ (see Eq. (2.58)). Subsequent removal of the high-frequency components from the estimated affiliation vector highlights the maximum. This maximum indicates the optimal parameter value. The color bar marks the number of detail coefficient levels removed with a Discrete Wavelet Transform (DWT) (9 from the maximum 16). The zero (green line) marks the unfiltered $\mathbf{\Gamma}^*(t)$.

2.4.2 Optimal Number of Clusters

The approach to determine the optimal number of clusters K_{opt} follows the elbow method applied in the K-means clustering approach (Yuan & Yang, 2019). The idea is to observe an error measure for the clustering procedure over the number of clusters and look for a characteristic "kink" in the graph. The choice of error measure is therefore significant and specific to the method used.

For example, a measure of compactness in K-Means clustering is defined as the sum of Euclidean intra-cluster distances between points in a given cluster. In contrast, a different error measure is used here because the SDE clustering method differentiates the temporal dynamics of a time series within clusters rather than a distance in space. Two clusters can be highly overlapping but different in terms of their induced stationary probability density functions. Euclidean distance would be a poor choice in this case. Instead, the KULLBACK–LEIBLER (KL) divergence is used, which is closely related to the stationary probability density of the SDE (this is an assumption within each of the clusters):

$$KL(P||Q) := \int P(x) \log \left(\frac{P(x)}{Q(x)} \right) dx. \quad (2.59)$$

The KL divergence describes the information gain when switching from the distribution with density P to the distribution with density Q . Here the P and Q denote the stationary distributions of two different clusters classified, for example, by $\gamma_1(t)$ and $\gamma_2(t)$. The goal is to quantify the difference between P and Q , or generally, the *diversity* between K stationary distributions. The more diverse all found distributions are, the more appropriate is the number of clusters. Because of the asymmetry of the KL divergence, the values of the opposite divergences $KL(Q||P)$ are also taken into account.

Denote by p_k^* the stationary distribution of an SDE with the estimated parameter set $\bar{\theta}_k^*$ and the corresponding $\gamma_k(t)$ for clusters $k = 1, \dots, K$. Thus, for each cluster, we have a different SDE model with a stationary distribution. As a measure, the weighted diversity matrix $d_{ij} \in \mathbb{R}^{K \times K}$ is defined as

$$d_{ij} := KL(p_i^*||p_j^*) = \nu_j \int p_i^* \log \left(\frac{p_i^*}{p_j^*} \right) dx \geq 0 \quad i, j = 1, \dots, K, \quad (2.60)$$

where ν_j are the weights calculated from the affiliation vector that quantify the relative prevalence of cluster j :

$$\nu_j = \int_0^{t_e} \gamma_j(t) dt, \quad j = 1, \dots, K. \quad (2.61)$$

These weights are introduced for the following reason. By adding more clusters, the less likely $\gamma_j(t)$ is suppressed (see, e.g., Fig. 2.5c) and hardly reaches values equal to 1 or a cluster has a short lifetime relative to the entire horizon t_e . These suppressed, less likely, inactive regimes should therefore contain less diversity and compensate the measure. To find the optimal number of clusters within a reasonable range $2 < K < K_{\text{max}}$, the diversity measure of clustering with K clusters is defined:

$$W_K = \sum_{i=1}^K \sum_{j=1}^K d_{ij}, \quad K = 2, 3, \dots, K_{\text{max}}, \quad (2.62)$$

where K_{\max} denotes the maximum of the considered clusters, assuming that $2 < K_{\text{opt}} < K_{\max}$. Reiterate that p_k^* in Eq. (2.60) indicates the stationary distribution of the SDE model, which is characterized by the parameter set $\bar{\theta}_k^*$. Specifically, for the SDE model

$$dX = f(X; \bar{\theta}_k^*)dt + g(X; \bar{\theta}_k^*)dW(t), \quad k = 1, \dots, K, \quad (2.63)$$

the invariant density p_k^* is

$$p_k^*(X) = \frac{N_c}{g^2(X; \bar{\theta}_k^*)} \exp\left(\int \frac{2f(X; \bar{\theta}_k^*)}{g^2(X; \bar{\theta}_k^*)} dx\right), \quad k = 1, \dots, K, \quad (2.64)$$

where the normalization constant N_c is as follows

$$N_c^{-1} = \int \frac{1}{g^2(X; \bar{\theta}_k^*)} \exp\left(\int \frac{2f(X; \bar{\theta}_k^*)}{g^2(X; \bar{\theta}_k^*)} dx\right), \quad k = 1, \dots, K, \quad (2.65)$$

provided all terms are well-defined (see Horsthemke (1984)[Ch. 6.1] for details).

In real-world applications, the increase in diversity W_K may be prohibitively steady, making the detection of K_{opt} difficult. A further extension of the elbow method to estimate K_{opt} is applied.

Tibshirani et al. (2001) proposed to measure the increase in W_K with respect to a so-called zero reference distribution: a dataset with no apparent clustering pattern. In the work of Tibshirani et al. (2001), the random distribution of points in two-dimensional space serves as a null reference dataset suitable for clustering using the K-means method. Considering the properties of the SDE clustering method used, the WIENER process with reflective boundaries is chosen as the zero-reference time series. The optimal number of clusters is estimated as the point at which the smallest distance between the measure of clustering diversity of the analyzed data and the reference data is reached:

$$\text{Gap}(K) = \mathbb{E}[\log(W_K^*)] - \log(W_K), \quad (2.66)$$

where W_K^* denotes the diversity measure of the zero-reference dataset, and the expected value $\mathbb{E}[\log(W_k^*)]$ is estimated by clustering $B \gg 1$ different independent reference time series. The boundaries for the reference WIENER process are limited to the domain of the analyzed process. Sec. 2.6 demonstrates the application of the presented approach on controlled numerical examples.

2.5 Implementation

Algorithm 1 is implemented in the C++ programming language. The core components of the FEM framework, namely: the spectral projected gradient method for quadratic programming, the FEM reduction and interpolation procedures, and the initial gamma generation procedure, are reimplemented from Matlab code created by Pospisil et al. (2018).

For the solution of the theta problem (see algorithm 1 line 9), the nonlinear optimization package NLOpt (G. Johnson, 2021) is used. This interface is convenient and contains many global and local optimization algorithms that can be probed by changing a single parameter.

By testing different methods, a combination of Controlled Random Search with local mutation (Price, 1977, 1983) and the gradient-free local optimization "PRAXIS" via the "Principal-Axis-Method" (Richard P, 2013) proved to be the most resilient.

The code is accelerated on the GPU with the CUDA library cuSPARSE, Thrust, and requires CUDA version 10.2. The transition density is reimplemented from a Maple code written by Sebastian Krumscheid into the Matlab symbolic toolbox to automate export into a C++ executable function. The Thrust template library allowed the machine-generated code of the transition density p_X (see Eq. (2.14)) to be incorporated into the main program flow, allowing testing of various model structures.

2.6 Numerical Studies

The numerical study is composed of three examples with different complexity. The following auxiliary process $u(t)$ is stochastic with a smooth sampling path and with adequate scale separation to $X(t)$. However, when working with data obtained from real applications, the process $u(t)$ may instead represent a characteristic quantity of a larger dynamical system.

The numerical examples involve the stepwise generation of a synthetic time series $X(t)$ and identifying the underlying functional relationship between the parameters of a predefined SDE and a process $u(t)$. It is assumed that the process $u(t)$ is available at each time point and serves as a predictor for the parameter vector $\Theta(t)$.

To create the synthetic dataset $X(t)$ for testing, the process $u(t)$ is used together with the functions $S_n(u)$ to generate the evolution of the parameter vector $\Theta(t)$. Then the sample path $X(t)$ is modeled with the predefined SDE and $\Theta(t)$. The time series $X(t)$ forms the dataset of test observations from which the functions $S_n(u)$ are recovered and compared to the function applied in the generation step.

A stochastic dynamic system is used to define an auxiliary process that maintains the required scale separation between $X(t)$ and $\Theta(t)$. Following the same principle, a realization of the auxiliary process $u(t)$ is constructed for all examples. A four-dimensional ORNSTEIN–UHLENBECK (OU) process is used by augmenting its parameters with the normalized BUTTERWORTH polynomial coefficients

$$d\mathbf{U} = -\frac{1}{T_c} \mathbf{A}_D \mathbf{U} dt + \mathcal{B} d\mathbf{W}(t), \quad \mathbf{U} = [u(t), u_1(t), u_2(t), u_3(t)], \quad (2.67)$$

where $T_c > 0$ controls the cutoff frequency, $\mathcal{B} > 0$ is the diffusion coefficient. The auxiliary process $u(t)$ is then the first component of the vector $\mathbf{U}(t)$. The diffusion matrix \mathcal{B} is almost everywhere zero, and together with the drift matrix \mathbf{A}_D takes the following form:

$$\mathbf{A}_D = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_0 & a_1 & a_2 & a_3 \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_0 \end{bmatrix}. \quad (2.68)$$

The coefficients in the matrix \mathbf{A}_D are: $[a_0, a_1, a_2, a_3] \approx [1, 2.61, 3.41, 2.61]$; $b_0 = 1$. The effect of the coefficients in \mathbf{A}_D is the smoothing of the first component of the vector \mathbf{U} , which

corresponds to filtering. The matrix \mathbf{A}_D is constructed based on the filter design theory. The drift term in Eq. (2.67) acts like a filter on the diffusion process, which acts as an input signal.

Consider that the process $u(t)$ can also be generated differently. One solves a one-dimensional OU process and then processes the result with a linear filter by fixing the cutoff frequency T_c . The filtering step gives us a sufficiently smooth time series. Alternatively, one could specify a power spectrum density in the frequency domain and then transform it into physical space using the inverse FOURIER transform. The above approach is favorite and straightforward because we simultaneously solve $u(t)$ and $X(t)$ and control the time step. However, an analytical solution, which also exists, would be the most elegant approach.

2.6.1 Nonstationary Ornstein–Uhlenbeck Process

We begin the numerical analysis with a trivial case where the sample path $X(t)$ is modeled by an SDE with a linear drift term and additive noise. Let us consider the OU process of the following form:

$$dX = (\theta_1(t) - \theta_2(t)X)dt + \theta_3(t)dW(t), \quad X(t_0) = 0, \quad (2.69)$$

in this case, the analytical expression for the transition density $p_X(\cdot, \cdot)$ is known. Nevertheless, the approximation with the HERMITE expansion is applied to test the implementation of the clustering framework. Figure 2.7a shows a realization of the sample path for process

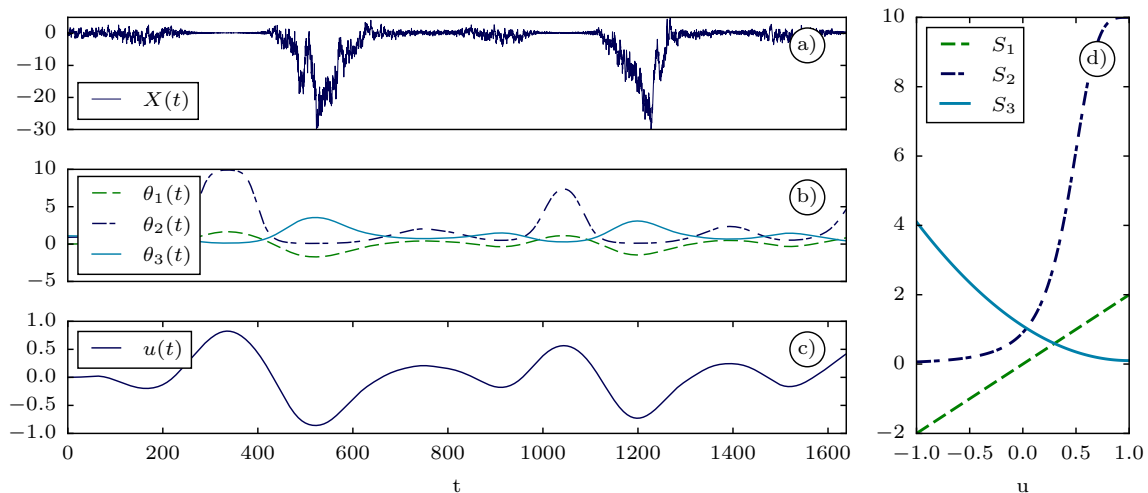


Figure 2.7: The summary of the functions to generate the nonstationary training data $X(t)$ according to Eq. 2.69. a) The sampling path of the process $X(t)$. b) The temporal evolution of the model parameters. c) The parameter auxiliary process $u(t)$. d) The scaling functions $\theta_n(t) = S_n(u(t))$.

$X(t)$. The SDE 2.69 is simultaneously implemented with the Eq. 2.67 and solved using the EULER–MARUYAMA method. The discretization time step $\Delta t = 0.1$. The resulting time series has 16384 grid points and is a small size dataset. During simulation, the time step between each 16384 samples is further reduced to a value of 10^{-6} to increase accuracy. The scaling

functions are defined as follows (see Fig. 2.7d)

$$\theta_1(t) := S_1(u) = 2u(t), \quad (2.70)$$

$$\theta_2(t) := S_2(u) = [(u(t) - 1)^4 + 0.1]^{-1}, \quad (2.71)$$

$$\theta_3(t) := S_3(u) = (u(t) - 1)^2 + 0.1. \quad (2.72)$$

The evolution of the vector $\Theta(t)$ (see Fig. 2.7b) and its functional relationship with the process $u(t)$ is hidden and unknown in real applications. The following demonstrates how to apply the FEM- H^1 -SDE method to find the scaling functions S_n when the discrete values of $X(t)$ and $u(t)$ are known.

One starts by estimating the hyperparameter values K and ϵ^2 by going through the optimization problem for each parameter combination. In this example, the following intervals for $K = (2, 10)$ and $\epsilon^2 = (10^{-1}, 10^2)$ are selected, the latter having 100 subdivisions on the logarithmic scale. Theoretically, for any parameter combination (K, ϵ^2) , the optimization procedure can be launched with a random initial guess $\bar{\Theta}_0$ and $\Gamma_0(t)$. In practice, however, it is efficient to use the optimal values from the previous run in the following way.

1. Choose a value for K
2. Start the first optimization run with the smallest or the largest value of ϵ_0^2 together with a random estimate for $\bar{\Theta}_0$ and $\Gamma_0(t)$
3. Algorithm 1 converges to a solution $\bar{\Theta}_1^*$ and Γ_1^*
4. Change the parameter value ϵ^2 to the next closest value ϵ_1^2
5. Start the optimization procedure with ϵ_1^2 and the solution from the previous run: $\bar{\Theta}_1^*$ and Γ_1^*
6. Algorithm 1 converges to the solution $\bar{\Theta}_2^*$ and Γ_2^*
7. Solve the variational problem for all considered ϵ^2 by changing its value progressively
8. Change the value of K and repeat the procedure

The approach where the initial estimation is random requires more computational time overall than with reusable solutions. By plotting the signal energy of the affiliation vector against the parameter ϵ^2 , we estimate the optimal value ϵ_{opt}^2 for each value of K separately (see for $K = 6$ Fig. 2.8c). According to Fig. 2.8c, the maximum energy is found at a value $\epsilon_{\text{opt}}^2 = 20$, indicating the clearest separation of the regimes. As shown in Sec. 2.4.1, the maximum in the curve E_γ can be further emphasized by suppressing the high frequencies in the vector $\Gamma^*(t)$. This is highlighted in Fig. 2.8c, where the index 0 represents the unfiltered result. Each additional integer denotes the continuous reduction of the frequency content by a factor of two. We note that the frequency suppression shows the desired effect but is unnecessary in this example. In addition, the affiliation function was examined for each value of ϵ^2 (not shown), and it was found to have relatively small differences in its value for a wide range of the ϵ^2 value. This indicates a robust estimate and may be attributed to the simplicity of the example.

2. Model-Based Nonstationary Time Series Modeling

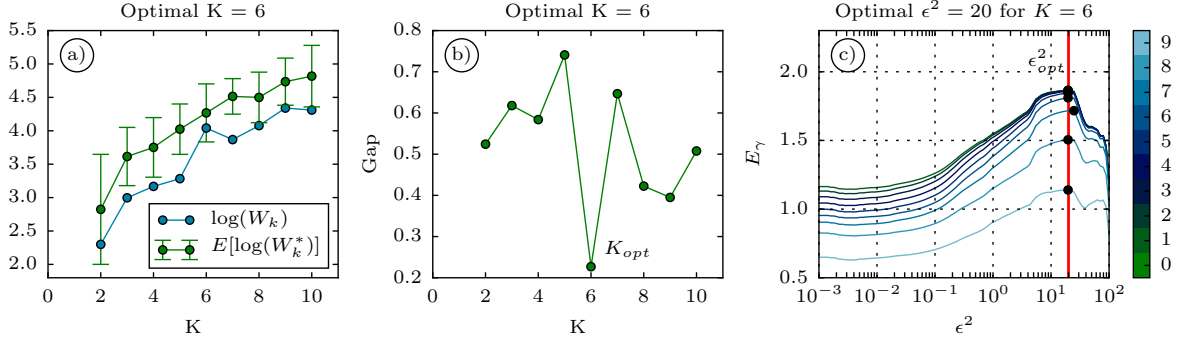


Figure 2.8: Estimating number of clusters for the nonstationary OU example. Panel a) shows that for $K > 6$, the diversity of the training data increases at approximately the same rate as in the reference data sets. Panel b) shows the distance between the two curves in panel a). This suggests an optimal K value of 6. (c) The estimate of the regularization parameter ϵ^2 for clustering with $K = 6$. The color bar encodes the successive suppression of the high-frequency bands to highlight the maximum. The effect is minimal, and filtering is unnecessary in this example.

The procedure for estimating the optimal K_{opt} requires more effort than ϵ_{opt}^2 . Section 2.4.2 introduced the diversity measure to quantify the incremental information gain with each additional cluster. Figure 2.8a shows the graph of this measure (blue) for the training data $X(t)$ and the reference data (green). According to the original approach, the reference dataset must be constructed to have no evident clustering. It is recalled that for this purpose, we chose the constrained WIENER process. The boundaries are set to the boundaries of the analyzed time series (see Fig. 2.7a) so that the WIENER process is reflected at these boundaries. This reference process is generated B times with the same time step used in the identification step. Clustering is repeated for each of these data sets with the value $\epsilon^2 = 20$ for the entire range of K values. The purpose is to leave all the parameters of the framework unchanged and replace only the data set. The whole procedure is computationally intensive, as we must repeat the clustering B times to determine the expected value. The resulting graph of the information gain for clustering the reference time series is shown in Figure 2.8a (green). The important consideration is the rate at which the information gain occurs when clustering the analyzed data compared to the rate at which the reference time series is clustered.

Figure 2.8b shows the distance between the two cluster curves. A reasonable choice is $K_{opt} = 6$ since the curve of the analyzed data does not approach the reference curve after that, which means that the model diversity does not increase significantly with respect to the bounded WIENER process. The expected value is formed based on 10 reference datasets. This number is increased in the following examples to obtain a better estimate.

The following paragraph explains the details of the internal parameters of the clustering framework. The minimization algorithm for the gamma solver has two termination criteria, which are as follows. The maximum number of iterations is set to 100, and the minimum difference between successive cost function values is 10^{-8} . The reduction value for the gamma solver accelerates the QP problem and is set to 1/3. In the theta solver, the global optimizer is a random search algorithm complemented by a local search algorithm. For both, the maximum number of calls is set to 300, and the termination tolerance is 10^{-10} . The population size of the global optimizer is set to 3000. The final point to mention is the parameter bounds

for the theta solver. In this example (and for clustering the reference dataset), they are $-20 < \theta_1 < 20$; $0 < \theta_2 < 20$; $0 < \theta_3 < 20$. Finally, the results are discussed.

Figure 2.9a shows $\Gamma^*(t)$ obtained after clustering the time series $X(t)$ with $K = 6$ (see Fig. 2.7a). The result is considered valid since a high value of $\gamma(t)$ is observed almost at every time instant. In Fig. 2.9b, the affiliation function is used to label the training data. Although the sample path of $X(t)$ at $t \approx 500$ appears to be similar to the sample path at $t \approx 1200$, the method distinguishes a change in the parameters of the SDE. It divides these times into two different clusters correctly.

In Figure 2.9c, the affiliation function is used to classify the auxiliary process $u(t)$. The duration of the regime occupation changes between clusters. It can be seen to be related to the rate of change of the auxiliary process. If this process changes rapidly, the occupation duration is short, and the process is nearly constant; the regime occupation is long.

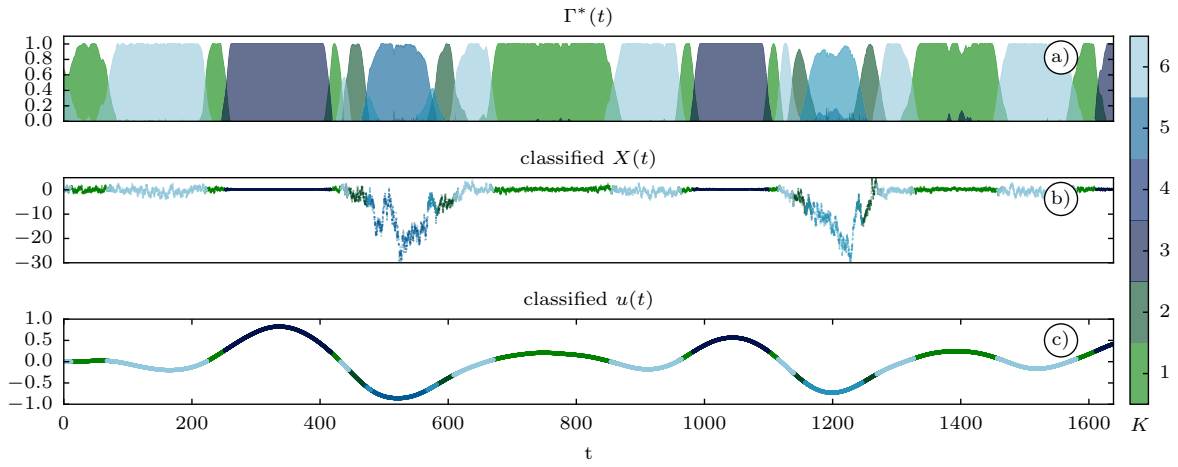


Figure 2.9: Result of clustering the nonstationary OU example in Fig. 2.7. a) The affiliation function found when solving the variational problem (2.27) with $K = 6$ and $\epsilon^2 = 57$. b) The classified time series $X(t)$. c) The classified auxiliary process $u(t)$. The color bar indicates the six different regimes.

One averages the classified auxiliary process (see Fig. 2.9c) to estimate the scaling functions and plots the values versus the estimated parameters for each cluster, respectively, using the weighted-average approach. Figure 2.10 shows the estimated parameter values obtained with

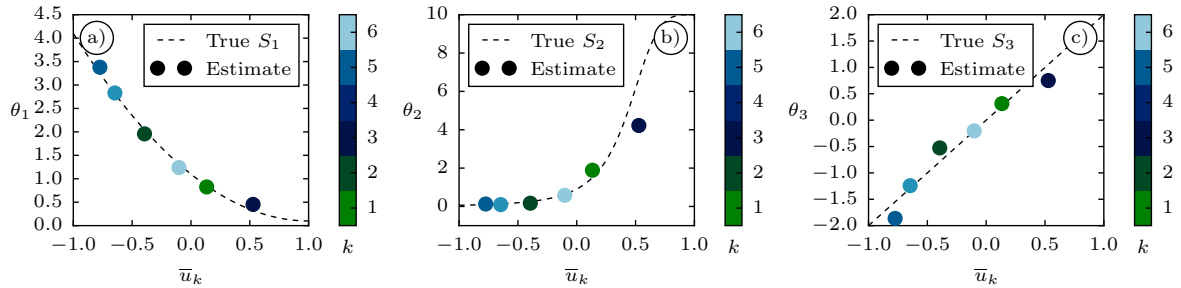


Figure 2.10: The estimated parameters are plotted over the cluster-averaged auxiliary process. The model identification is performed with an optimal number of clusters, corresponding to six. The color bar shows the different clusters as in Fig. 2.9. The dots represent the estimated parameter values. The dashed line shows the actual scaling functions used in generating the training data.

six clusters. The corresponding six points are sufficient to estimate the hidden scaling functions.

2. Model-Based Nonstationary Time Series Modeling

When working with unexplored data, the scaling functions would need to be parameterized using the scatter plots. This step is left out here because the estimated parameters agree well with the true functions. The largest error in this example is found for the parameter θ_2 in the cluster $k = 3$. This error can be explained because this cluster has the largest variability of $u(t)$. Please note the classified periods of $u(t)$ for $k = 3$ in Fig. 2.9c. One way to improve the results would be to consider more sample points and repeat the clustering with more clusters.

2.6.2 SDE with Two Independent Auxiliary Processes

In this example, an SDE with nonlinear drift and a multiplicative noise term is considered. In addition, this example aims to show that the clustering method can handle the modulation by two independent parameters with different time scales. The following functional form of the SDE is investigated:

$$dX = (2 - \theta_1(t)X - \log(X^2))dt + \theta_2(t)XdW(t), \quad X(t_0) = 1, \quad (2.73)$$

The relationship between the parameters and the auxiliary processes $u(t)$ and $v(t)$ is

$$\theta_1(t) := S_1(u) = (u(t) + 1)^4 + 0.1, \quad (2.74)$$

$$\theta_2(t) := S_2(v) = -2v(t) + 2, \quad (2.75)$$

The only difference between $u(t)$ and $v(t)$ is that their cutoff scale is $T_{c,v} = 4T_{c,u}$ (see Eq. (2.67)). The task is to recover the scaling functions given only one realization of $X(t)$, $u(t)$ and $v(t)$. The sample size of the time series $X(t)$ is 131072 points, which is larger than in the example

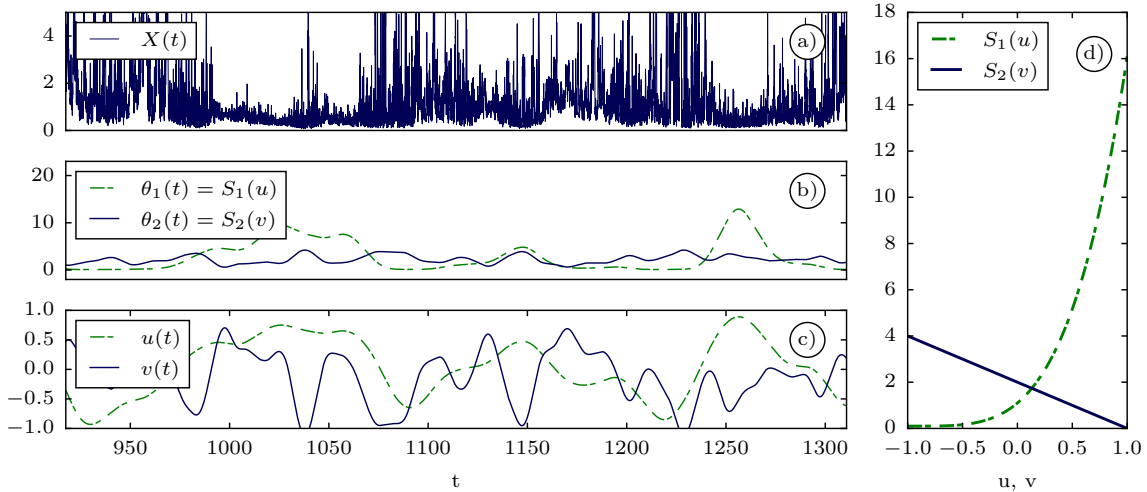


Figure 2.11: Training data for nonstationary and nonlinear SDE where two parameters are modulated by two independent processes with non-overlapping time scales. a) The sampling path of the process $X(t)$. b) The temporal evolution of the model parameters. c) The parameter auxiliary processes $u(t)$ and $v(t)$. d) The parameter scaling functions $\theta_n(t) = S_n$.

with the OU process (see Sec. 2.6.1). This helps the methodology to distinguish the unique parameter pairs. (The processes $u(t)$ and $v(t)$ are uncorrelated, and the re-occurrence of the unique parameter pairs in the sample path $X(t)$ is not guaranteed. This makes identification of the scaling functions difficult. To mitigate this problem, one must increase the number of

clusters to improve the resolution of the scaling functions. Increasing the number of clusters reduces the number of samples available to estimate the parameter values within a cluster.

Consequently, the sample size of the entire time series $X(t)$ must also be increased to maintain the accuracy of the parameter estimates. It is essential to consider the dimensions of the parameter vector and the vector of auxiliary processes since there may be a limited number of data points to resolve the dependencies. One can deliberately fix one of the model parameters to compensate for overparameterization.

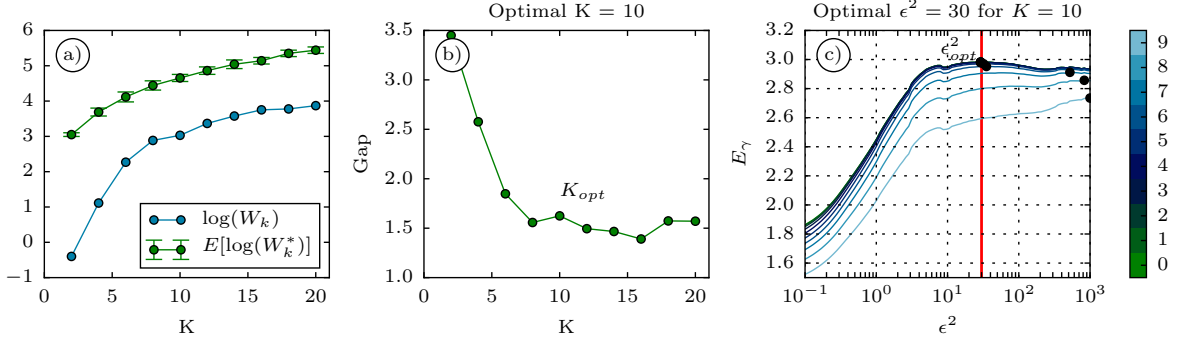


Figure 2.12: Estimation of the number of clusters for the sample with two auxiliary processes (see Fig. 2.11 and Eq. 2.73) using the gap statistics approach in a) and b). In panel a), the diversity of the training data increases at the same rate as in the reference datasets after $K = 10$. This suggests that the optimal K value is around 10. (c) The estimation of the regularization parameter ϵ^2 for clustering with $K = 10$. The color-coding shows the subsequent suppression of the high-frequency bands to accentuate the maximum. The filtering is not necessary in this example.

Figure 2.11a shows the example path of the SDE (2.73) along with the auxiliary processes $u(t)$ and $v(t)$ (see Fig. 2.11c) used to generate $\Theta(t) = [\theta_1(t), \theta_2(t)]$ (see Fig. 2.11b). The scaling functions are shown in Fig. 2.11d.

The selection of the hyperparameters K_{opt} and ϵ^2_{opt} is performed in the same way described in the OU process example (see Sec. 2.6.1). The corresponding graphs are shown in Fig. 2.12.

The choice of K_{opt} requires some clarification. The diversity measure of clustering for process $X(t)$ and the reference dataset is shown for comparison in Fig. 2.12a. The peak in the evolution of diversity in the clustering of the dataset $X(t)$ is reached approximately at $K_{opt} = 10$. Plotting the distance between the two curves, we find that for $K > 10$, clustering does not add significantly more new information than clustering the bounded WIENER process. Selecting ϵ^2_{opt} for this example is characterized by a flat maximum (see Fig. 2.12c).

The optimal classifier $\Gamma^*(t)$ and the partitioning of the auxiliary processes are shown in Fig. 2.13, where the clustering quality can be examined. Figure 2.14 shows the estimated parameters and the true scaling functions. Visually inspecting these results, we find a consistent spread of estimates for the θ_2 parameter, but the estimates for the θ_1 parameter are disproportionately distributed. For the estimates of θ_2 , it is noticeable that three parameters appear to have been estimated twice and reside close to each other. However, the respective cluster values for the parameter θ_1 scatter apart, so that the scaling function is sampled at significantly different points. This means that the methodology in this example distinguishes pairs of parameters that have a single common element. The methodology classifies SDE with at least two uncorrelated auxiliary processes.

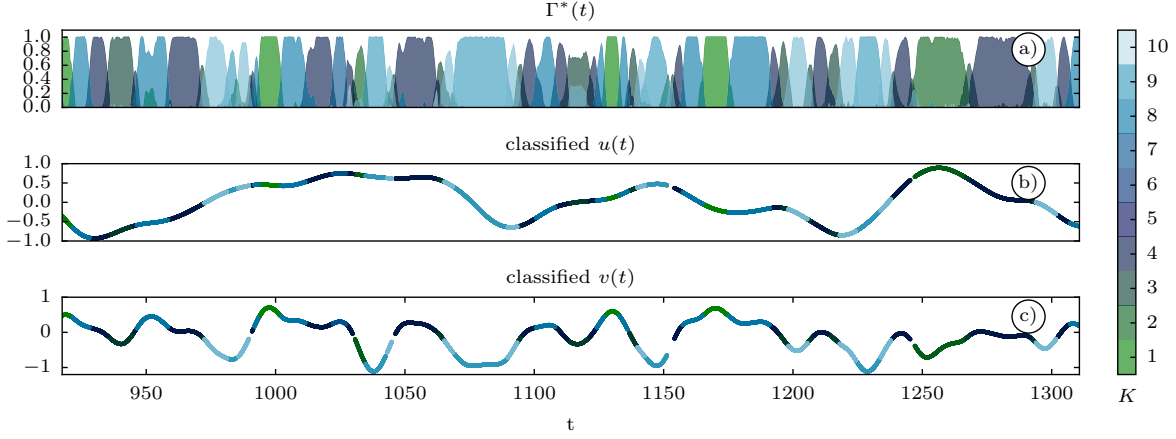


Figure 2.13: The result of clustering the example time series shown in Fig.2.11. a) The affiliation function found when solving the variational problem (2.27) with $K = 10$ and $\epsilon^2 = 30$. b) The classified auxiliary process $u(t)$. c) The classified auxiliary process $v(t)$. The figure shows 1\3 of the training dataset.

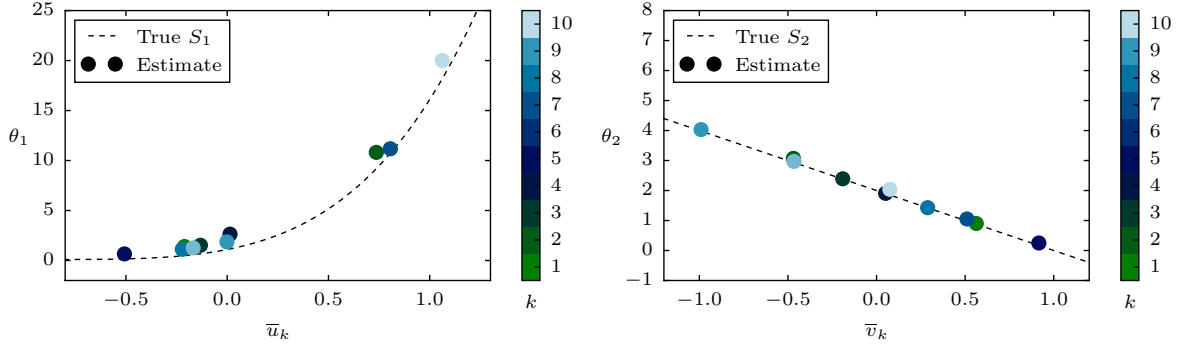


Figure 2.14: The estimated parameters are plotted against the cluster-averaged auxiliary processes values. Model identification is performed with an optimal number of clusters, which is ten. The color bar indicates the different clusters, as in Fig. 2.9. The dots represent the estimated parameter values, and the dashed line shows the true scaling functions used in generating the training data.

2.6.3 SDE with Nonlinear Drift and Nonlinear Diffusion

In this example, we study an SDE with nonlinear drift and nonlinear diffusion. For simplicity, we consider only one auxiliary process $u(t)$ and two parameters: one in the drift term and one in the diffusion term. The functional form of the SDE is

$$dX = (\theta_1(t)X - X^3)dt + \theta_2(t)\sqrt{1 + X^2}dW(t), \quad X(t_0) = 0. \quad (2.76)$$

It is possible to include other parameters in the model structure, for example, before the quadratic or the cubic term. The system's time-dependent bifurcation parameter $\theta_1(t)$ causes a continuous variation of the metastability in time. The goal is to show that the clustering method can capture transitions in the underlying dipole potential coming from the cubic nonlinearity of an SDE. The relationship between the parameters and the auxiliary processes is

$$\theta_1(t) \equiv S_1(U_1) := -0.4(U_1(t) - 1)^2 + 2.5, \quad (2.77)$$

$$\theta_2(t) \equiv S_2(U_1) := -4U_1(t) + 5. \quad (2.78)$$

As in the previous examples, the task is to determine the scaling functions when knowing only one realization of the process $X(t)$ and $u(t)$. In this section, the details of the simulation

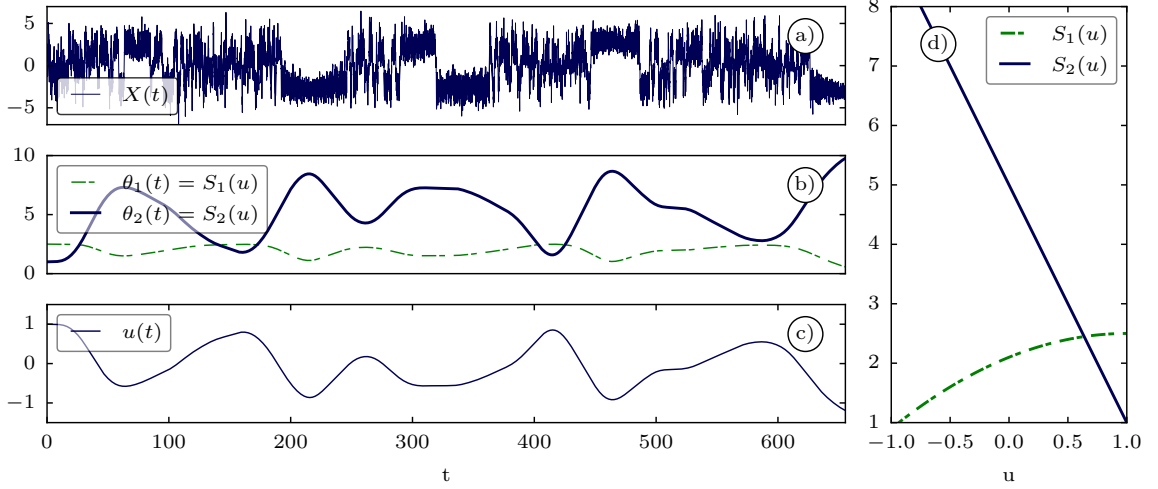


Figure 2.15: The example of a non-stationary process associated with Eq. (2.76). a) The sample path of $X(t)$. b) The time evolution of the model parameters. c) The parameter auxiliary process $u(t)$. d) The scaling functions $S_n(u(t)) = \theta_n(t)$.

of the sample path $X(t)$ are described (see Fig. 2.15), which forms the training data for the clustering test. The time series $X(t)$ has 65536 samples and a constant time step $\Delta t = 0.01$. The sample path is obtained by solving Eq. (2.76) with the MILSTEIN method, and the sample path of $u(t)$ is obtained by solving Eq. (2.67) with the EULER–MARUYAMA method. During the simulation, the time step between each 65536 samples is reduced to a value of 10^{-4} to obtain a more accurate representation of the sample path. The initial value for Eq. (2.76) is set to $X(t_0) = 0$, and for Eq. (2.67) - to $U(t_0) = [1, 0, 0, 0]$. The parameters in Eq. (2.67) have the following values: $T_c = 15$; $b_0 = 0.2$.

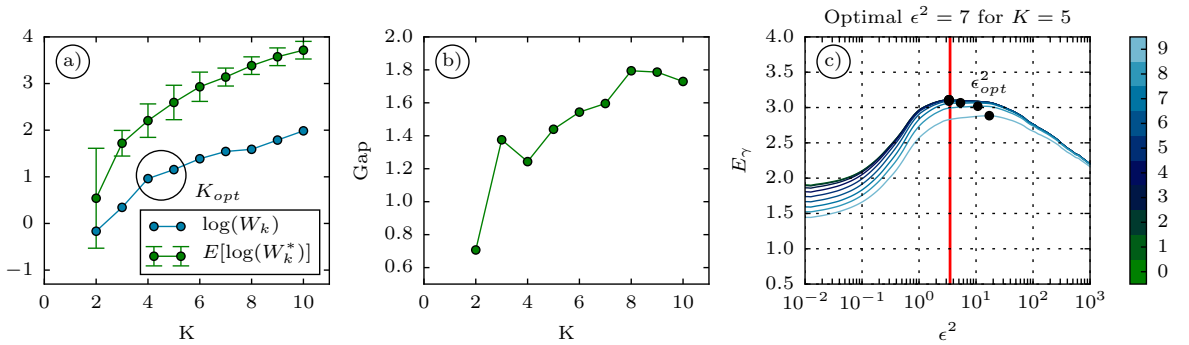


Figure 2.16: Estimating the number of clusters for the example associated with Eq. 2.76 (see Fig. 2.15) using the gap statistics approach a) (see Eq. (2.62)) and b) (see Eq. (2.66)). In panel a), it can be seen that after $K = 4$, the diversity of the zero-reference dataset increases at a constant rate. b) The gap statistic shows $K_{opt} = 4$. c) The estimate of the regularization parameter ϵ^2 for clustering with $K = 5$. The color bar encodes the successive suppression of the high-frequency bands to highlight the maximum.

In this example, the estimation of K_{opt} does not provide a clear result as in the previous examples. However, it is still beneficial to analyze the tendencies of the curves in Figure 2.16a, b.

2. Model-Based Nonstationary Time Series Modeling

The goal is to recover the scaling functions, and in the case of a nonlinear scaling, one would like to have more than two points and thus more clusters. According to the second minimum in Fig. 2.16b, the second-best choice is $K_{\text{opt}} = 4$. Examining the curve $\log(W_k)$ in Fig. 2.16a, we find a small characteristic change in slope at $K_{\text{opt}} = 4$. To resolve the scaling functions, $K_{\text{opt}} = 5$ is selected for further analysis.

The following paragraph provides details on the parameters of the clustering framework. In this example, the maximum number of iterations is set to 100, and the minimum difference in successive evaluations of the cost function is set to 10^{-8} . The reduction value for the gamma solver is set to 0.1. A global optimizer is a random search algorithm with an additional gradient-free local search algorithm for the theta solver. For both, the maximum number of calls is set to 500, and the termination tolerance is set to 10^{-10} . The population size of the global optimizer is set to 1000. The bounds for the parameter limits of the theta solver (and for the clustering of the reference dataset) are $-10 < \theta_1 < 10$; $0 < \theta_2 < 10$.

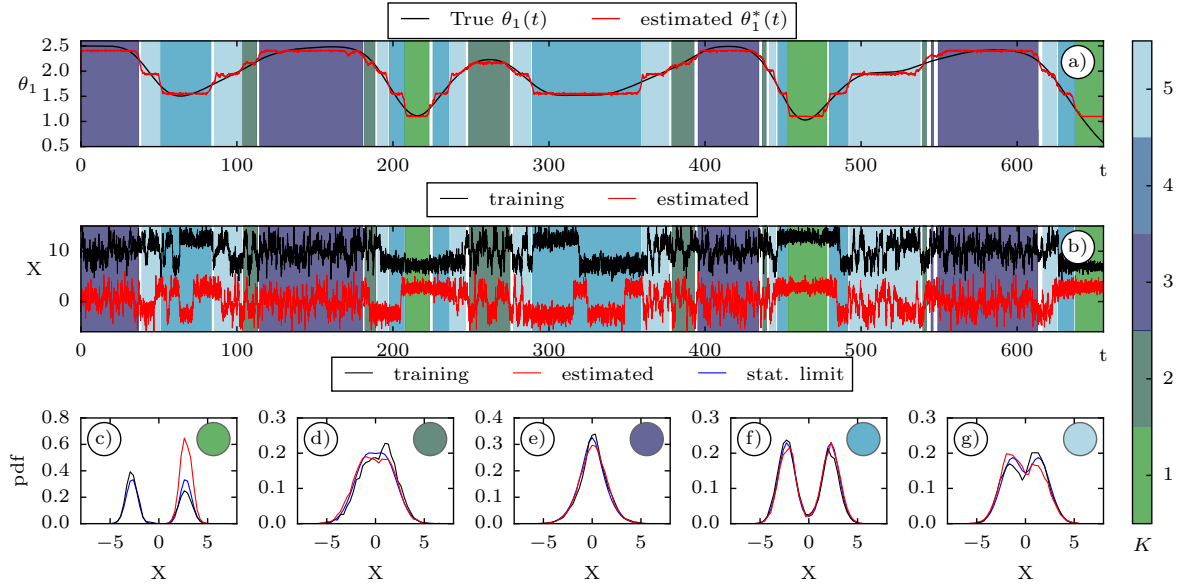


Figure 2.17: The clustered time series is generated using the Eq. 2.76. In a) and b), the colored background is the estimated affiliation vector for $\Gamma(t) > 0.8$. The color bar on the right side of the panels encodes the corresponding 5 clusters. a) The temporal evolution (black) of the diffusion parameter and its recovered version (red). The second parameter looks similar and is not shown. b) The simulation of the example path with the estimated parameter values compared to the training data. Panels c) to g) show the cluster-related Probability Density Functions (PDFs): the histograms for the training (black) and the simulation (red). The blue lines mark the steady-state PDFs calculated from the parameter estimates. The colored circles mark the corresponding clusters.

Discussion of a solution for $K_{\text{opt}} = 5$ will follow next (see Fig. 2.17). Knowing the affiliation function and the averaged values of the parameters, the time evolution of the model parameters (see Fig. 2.17a) is approximated by

$$\theta^*(t) \approx \sum_{k=1}^{K_{\text{opt}}} \bar{\theta}_k^* \gamma_k(t), \quad (2.79)$$

where $\bar{\theta}_k^*$ is a constant vector of length N . Figure 2.18a shows that the methodology reproduces the hidden evolution of the parameters quite well. The approximation gets better as more

clusters are used (not shown). The drawback is that the uncertainty in $\bar{\theta}^*$ increases with the number of clusters.

The approximated path of the model parameters (see Fig. 2.17a in red) is used to simulate the in-sample prediction (see Fig. 2.17b in red). In this figure, the time series of the training dataset is shifted up by 10 for better comparison. The SDE under consideration exhibits a temporal modulation of the bifurcation. It is expected that the times with two metastable states have different sampling paths since the local state of the system depends on the particular realization of the WIENER process. Two distinct WIENER processes were used to generate the training and simulation solution.

Figure 2.17b confirms the correctly captured signal dynamics. The corresponding cluster characteristics agree well with the features of the training data and with the respective stationary distributions computed from the parameter estimates. The cluster with index 1 shows the largest discrepancy (see Fig. 2.17c). This cluster has a life span that is too brief to capture the regime jump in the current realization. However, the stationary distribution calculated from the estimated parameter values shows the correct two-mode distribution (see Fig. 2.17c).

Figure 2.17 demonstrates the performance of the model within the training data. To construct a self-contained, nonstationary predictive model, it is shown that the scaling functions between the auxiliary process $u(t)$ and the parameters are also recovered (see Fig. 2.18). The

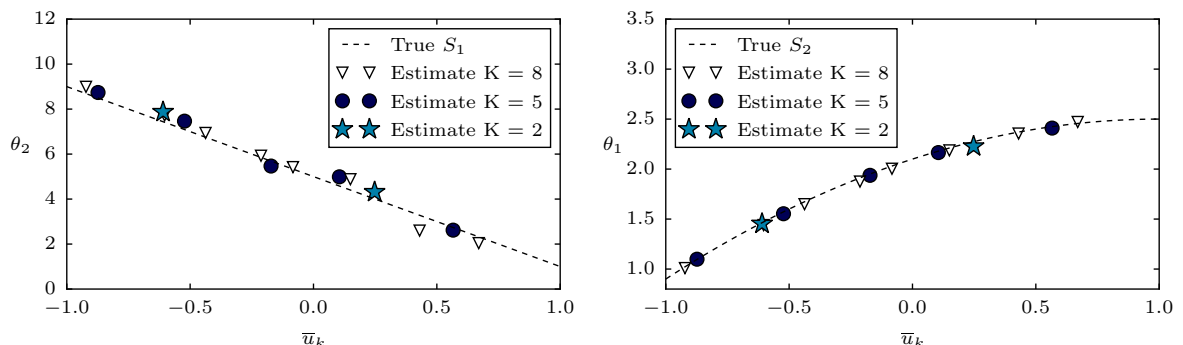


Figure 2.18: The estimated parameters are plotted against the cluster-averaged auxiliary process for various clusters. The dots represent the estimated parameter values, and the dashed line shows the actual scaling functions used in generating the training data. The number of clusters is irrelevant since all solutions are valid.

accuracy of the clustering method is sufficient to obtain the correct parameter values for a wide range of clusters. It should be noted that the optimal number of clusters is of secondary importance. The number of clusters should be as low as possible and high enough to identify the scaling function. Once the figures have been obtained as Fig. 2.18, the final step is to perform a regression analysis to parameterize the scaling behavior. This step is omitted here as it will be specific to each application.

2.7 Discussion and Conclusion

Data-driven modeling techniques can advance the understanding of complex dynamical systems by revealing underlying patterns or hidden rules. A particular challenge is the identification and parameterization of nonlinear and nonstationary SDEs. System identification aims to

determine a model structure and estimate the appropriate model parameters from the data. After its completion, one can analyze the newly estimated model to better understand the process and derive hidden relationships. Pure classification may not explicitly reflect these relationships, so uncovering them will require additional research. For example, in Vercauteren et al. (2016), the influence of non-turbulent motions on turbulence was studied after applying a model-based data clustering method (Vercauteren & Klein, 2015). In a study of the North Atlantic Oscillation in the climate system, Quinn et al. (2021) used clustering results to study dynamical processes associated with atmospheric regime transitions.

The approach presented in Sec. 2.2.1 subdivides the time series into segments and models each with an individual, locally stationary SDE. Such a decomposition of a complex dynamic behavior may not be optimal since it is not always clear how to distinguish between a nonlinear and nonstationary time series without knowing the underlying structure of the data generating process. Defining a suitable model structure is, therefore, an important step. Billings (2013) has developed an estimation algorithm for discrete-time systems that can sequentially evaluate and collect important model terms. Following this idea, Wei et al. (2004) found an effective model for a highly nonlinear dynamical system in the terrestrial magnetosphere. In their framework, the time-varying parameters can be estimated using a multi-wavelet (Wei et al., 2010) or a sliding window approach (Li et al., 2016). In comparison, variational clustering optimally estimates all positions and time window lengths to find the locally optimal parameter values. As shown in Fig. 2.17, the variability of the time window is essential and is associated with the rate of change of the auxiliary process.

Section 2.2.1 illustrates the theory of the variational clustering framework, which allows simultaneous analysis and modeling of a nonstationary time series. A framework variant (Horenko, 2010b) is based on discretized mathematical models is also briefly discussed in Sec. 2.2.2. To integrate an identified discrete-time model into an existing continuous-time system such as a RANS turbulence closure, one must consider the discretization operation. Otherwise, the solution is limited to the sampling frequency imposed by the data. One way to overcome this problem is to preserve the frequency response of the identified model during the transformation from a discrete system to a continuous analog system (Kuznetsov & Clarke, 1994). This transformation is widely applied in linear systems theory and can also be generalized to nonlinear systems (Billings, 2013)[p. 342]. The approach developed in this chapter avoids the difficulty because the MLE used is derived from a reasonable likelihood (fitness) function for the discrete-time observations of a continuous-type SDE.

In general, the main challenge in applying the MLE is the lack of a transition density function, which requires, for example, a solution of the FOKKER–PLANCK Equation for a considered structure of the SDE. In particular, for nonlinear SDE, this function is not known analytically and requires estimation (Fuchs, 2013). In this work, dividing the available time series into K clusters reduces the number of available samples for estimation, leading to uncertain estimates in each cluster. The closed-form approximation based on the HERMITE expansion provides an accurate approximation for the transition density, with $O(\Delta t^4)$ (Aït-Sahalia, 2002). This accuracy makes it possible to increase the number of clusters and improve the resolution of the hidden scaling functions, allowing accurate parameterization (see Fig. 2.18).

This work combines the closed-form likelihood function approach based on appropriate HERMITE expansion with the nonparametric clustering framework. In particular, the necessary modifications to the original subspace clustering algorithm are developed to obtain accurate estimates of the model parameters independent of the discretization of the QP problem. In addition, an extensive numerical study validates with three controlled examples of varying complexity both the parameterization methodology and the novel methods for estimating the framework’s hyperparameters.

In particular, the approach to estimate the optimal number of clusters is based on measuring the degree of model diversity and comparing it to the clustering of the reflected reference WIENER process. The numerical experiments confirm that this approach provides sufficiently reliable estimates. However, the associated graphs turn out to depend on the model structure under consideration so that the decision requires a case-specific interpretation. The proposed method is still very computationally intensive, as the clustering of the reference dataset must be repeated several times to obtain a statistical measures.

The method for determining the optimal regularization value of the affiliation function is shown to be robust for the test cases studied. The method is extended by filtering out the small-scale variability from the affiliation vector to articulate the optimum. This method has been successfully tested on a real-world data application and proven effective (see Ch. 4).

In summary, the data-driven approach of variational clustering enables modeling nonlinear and nonstationary processes to develop models for dynamical systems with multiple scales. The model structure of the SDE is unknown and must be specified or derived in a complementary way. The method is accurate and provides good results on small datasets consisting of 2×10^4 points. It is also scalable on the GPU, as implemented in this work. A dataset with 3×10^6 points were also tested, where the computation time on a GeForce 1080 was about 24 hours. The numerical examples demonstrate the correct identification of the required scaling functions, which allows the parameterization of a considered SDE by relating additional auxiliary processes to the model parameters.

This work focuses on one-dimensional time series; however, it is possible to extend the approach to a multidimensional case. One could alternatively use the closed-form likelihood expansion for multivariate diffusion (Aït-Sahalia, 2008). A modification of the variational clustering method is unnecessary for this purpose. In practice, however, overparameterization problems can arise because the cross-dimensional interaction terms and their temporal variability contribute to poorly constrained systems. One way to mitigate this is to enforce diagonally sparse diffusion and drift matrices, as demonstrated by O’Kane et al. (2017) in a similar application.

Another research direction is the study of multidimensional or even spatial problems. Then it is convenient to replace the MLE with an ensemble KALMAN inversion procedure (Albers et al., 2019). The variation clustering framework is also generalized to classify spatial parameter variability (Kaiser, 2015, p. 81). Pospisil et al. (2018) implemented this approach in the openly available code linked in their work. The proposed methods appear promising; however, as the dimension increases, it becomes increasingly important to estimate the correct model terms rather than the parameter values.

2. Model-Based Nonstationary Time Series Modeling

Another direction for future research is to extend the nonstationary data-driven learning approach developed in this work to SDE models with (at least) two widely separated time scales, for which the MLE is known to be asymptotically biased. The proposed methodology must be combined with robust multiscale inference techniques (Kalliadasis et al., 2015; Krumscheid, 2018; Krumscheid et al., 2013). The approach presented here provides an ideal framework for deriving subgrid-scale models modulated by resolved variables. The application of the methodology is presented in Ch. 4

3

Nonstationary Analysis of the Stable Boundary Layer

Any attempt to divide the stable boundary layer into a few classes or states, as in this study, is an oversimplification.

—Larry Mahrt, *Stratified Atmospheric Boundary Layers (1999)*

In the previous chapter, the model-based variational clustering methods were introduced and explained. One of them is the discrete-time model-based clustering method FEM-BV-VARX, which is applied in the following to study the nonstationary SBL in terms of a few dynamical states characterized by distinct interactions between non-turbulent and turbulent scales. The modeling assumption is that the nonstationarity of the SBL is encoded by the time variation of some specific model parameters (see Ch. 2). Two model structures for statistical clustering framework are proposed, guided by some physical considerations. The model structure is guided by the hypothesis that turbulence is partially modulated by wave-like fluid motions leading to localized shear generation. Moreover, the analysis focuses on the post-processing of the classified regimes of the SBL. In addition, the predictive capabilities of the models are also covered. The models developed in this chapter are not suitable for constructing RANS closures because they are of discrete type. However, the next chapter describes how a data-driven and modular stochastic extension for the RANS turbulence closure can be constructed from the eddy-correlation data using a continuous-type clustering approach (see Ch. 4). The exploratory data analysis and modeling experience in this chapter serves as the groundwork for this.

The FLOSS2 dataset is examined by improving the discrete-time model structure gradually. Since two different model structures are tested, the chapter is divided into two corresponding sections, reporting on data processing, evaluation, and post-processing for each model. In the first part (Sec. 3.2), a model structure following Vercauteren & Klein (2015) is used, with the new contribution being that the dataset is different. In addition, the methodological approach is slightly adapted from the study of Vercauteren & Klein (2015). After revising the model used in Sec. 3.2, an improved model structure is proposed in Sec. 3.3, which aims to investigate

the relevance of velocity gradients at the sub-mesoscale in the generation of intermittent turbulence. The subsequently introduced diagram and cluster analysis in Sec. 3.3.7 confirm and explain the contribution of the sub-mesoscale motions within an SBL to the intermittent state of turbulence.

3.1 Fluxes Over Snow Surfaces II Dataset

The dataset consists of high-resolution eddy correlation data collected during the (FLOSS2) field program. This experiment was conducted from 20 November 2002 to 4 April 2003. The SBL eventually evolves during the night, with the most stable condition occurring before sunrise. It is, therefore, preferable to measure in winter to increase the number of observations of the stable regimes. The dataset is officially available and can be accessed through the Earth Observing Laboratory (EOL data, 1999).

The experimental site is privileged for the study of SBL. The site is locally flat with variable bush height from 0.2 m to 0.5 m (Mahrt & Vickers, 2005) (see Fig. 3.1a) and is located in the North Park region of Colorado, near Walden. During the field campaign, a thin layer of snow covered the ground for about 20 days. This resulted in measurements where the entire day was in stable condition, and thus such records are of particular interest for this work. A flat surface is important for studying scale interactions because it eliminates the flow complexity induced by the terrain.

Seven sonic anemometers (Campbell CSAT3) recorded three wind velocity components and temperature at heights of 1, 2, 5, 10, 15, 20, and 30 m. The instruments were attached to a rigid truss tower that allows air masses to pass through (see Fig. 3.1b). The permeable yet rigid structure of the tower is important because otherwise, an impermeable or flexible body would induce large flow structures that would contaminate the signal of atmospheric turbulence. The natural turbulence structures are of particular interest. The dataset was quality controlled following Vickers & Mahrt (1997) and carefully prepared for clustering by the author.

When working with temperature data, one must be aware of uncorrected bias. The bias manifests itself in the deviation from the averaged temperature variables due to the influence of humidity on the sonic anemometers. In a personal communication, Danijel Belušić suggested an approximate correction:

$$T_{\text{bias}} = [0, -0.2991, 0.0897, -0.1781, -0.0155, -0.0951, -0.1026][\text{T}], \quad (3.1)$$

where the first value in the series corresponds to the lowest measurement height. The correction is determined by the method of deviations from the expected constant potential temperature profile at near-neutral conditions. Alternatively, if humidity measurements are available, one can correct for the sensible heat flux following (Foken, 2017, Section 4.2.3.5). This work focuses mainly on modeling turbulent momentum transport, so the turbulent heat flux is of secondary importance. However, for a SBL, the mentioned errors are of minor importance (Mahrt, 2011). Despite the temperature bias, the excellent quality and continuity of the dataset are reported, which is essential for the data analysis methods applied.



Figure 3.1: (a) The site of the FLOSS2 experiment. (b) The 30 m high measurement tower with seven mounted sonic anemometers. ©2021 UCAR.

3.2 Classifying Vertical Turbulent Velocity Fluctuations

The section covers the following topics: Data preprocessing, preparing the modeling variables, performing the model-based clustering, evaluating the model’s prediction on the training data, and post-processing the identified SBL regimes.

The model structure used below follows the design of Vercauteren & Klein (2015), where the variable of interest is the vertical turbulent velocity fluctuations, and the exogenous variable is the sub-mesoscale wind speed. The aim is to investigate whether a data-driven analysis of a nonstationary time series can identify periods in which non-turbulent motions and small-scale turbulence interact differently. Vercauteren & Klein (2015) performed such an analysis using the Snow-Horizontal Array Turbulence Study dataset collected over Plaine Morte Glacier. The goal of this study is to reproduce the modeling results using the FLOSS2 dataset. Vercauteren et al. (2019b) report on the comparison of the clustering results of these two datasets by analyzing and comparing the statistical properties of the non-turbulent structures within the identified regime.

3.2.1 Data Preprocessing and Scales

The SBL must be selected from the continuously recorded measurements. For this, the data are restricted to the night time between 18:00 and 07:00 local time. To maximize the continuity of a time series, several data-reduction steps are performed. Nights with missing periods (probably due to iced instruments) in the time series longer than 80 minutes in a single night (which equals 12 nights in total) are removed. Such long data gaps are exceptions, so the resulting nights have gaps with smaller periods (≈ 3 min); these were linearly interpolated. Fifty-one nights with periods where the mean wind vector approaches the instruments through the tower structure for longer than a total of 5 min were removed to minimize the tower’s influence. This results in 68 nights for the further analysis.

Next, two relevant variables are prepared for the clustering procedure. For this study, only the time series at the 2 m level is considered. Following Mahrt (2010b), the submeso velocity scale for one velocity component is

$$X^* := [X]^{1\text{min}} - [X]^{30\text{min}}, \quad (3.2)$$

where the square brackets denote the time-averaging operator, and the superscript number indicates the time window. The 1-min scale is assumed to be the turbulent scale for the strongly SBL of the FLOSS2 dataset. This scale is dependent on the Ri number. However, to simplify the analysis, the scale is fixed to be valid in the strongly stable regime. Reestimation of this scale value after clustering in Sec. 3.2.5 confirms this assumption (see Fig. 3.10).

Consider the modeling variables next. The first variable (being the forcing for the turbulence to be modeled) is the strength of the submeso horizontal motions defined as:

$$U_s := \sqrt{(u^*)^2 + (v^*)^2}. \quad (3.3)$$

As conjectured by Mahrt (2010b), the value U_s vanishes in the neutral steady-state flow limit. This aspect is investigated in a separate clustering study using an extended model structure (see Sec. 3.3). Moreover, Sec. 3.3.7 quantifies how the turbulence energy changes when approaching such a limit of low submeso activity (see Fig. 3.25).

The second variable in the analysis, which is classified based on its interaction with U_s , is the standard deviation of the vertical turbulent fluctuations σ_w and is estimated with REYNOLDS averaging using a scale of 1 min.

$$w' = w - [w]^{1\text{min}}, \quad (3.4)$$

$$\sigma_w = \sqrt{[w'w']^{1\text{min}}}. \quad (3.5)$$

The vertical component (σ_w) of the TKE is accounted for by giving priority to vertical transport in the boundary layer. However, the entire TKE is modeled in Sec. 3.3.5. Next, the model structure is assembled by assuming that the variable U_s drives the variable σ_w in strongly SBL.

3.2.2 Model Formulation for Clustering

Given the one-dimensional time series of variables σ_w and U_s with a sampling rate of 1 cycle per hour for the considered 68 nights, the goal is to estimate the temporal parameter variation of the following model. The fitness function (l_{loc}) includes the proposed model structure (see Sec. 2.1.1 to comprehend the structure of the model):

$$l_{\text{loc}}(\sigma_w; \boldsymbol{\theta}) := \left\| \sigma_w[n] - a_0[n] - \sum_{j=0}^q b_j[n] U_s[n - j\Delta t] \right\|_2 \quad n = 0, \dots, N - 1, \quad (3.6)$$

where N is the length of the time series, q is the memory depth, and at the same time a hyperparameter of the FEM-BV-VARX method. The parentheses $[\cdot]$ emphasize the discrete-time formulation of the model. The brackets $\|\cdot\|_2$ denote the Euclidean distance and $\boldsymbol{\Theta} = (a_0[n], b_j[n])$ denotes the vector of unknown time-dependent parameters. These parameters are obtained by minimizing the functional

$$L(\boldsymbol{\Theta}, \boldsymbol{\Gamma}[n]) = \sum_{k=1}^K \sum_{n=0}^{N-1} \gamma_k[n] l_{\text{loc}}(\sigma_w; \boldsymbol{\Theta}) \rightarrow \min_{\boldsymbol{\Theta}, \boldsymbol{\Gamma}[n]}, \quad (3.7)$$

with the methodology presented in Sec. (2.2.2) under the constraints (2.21), (2.22) for the affiliation vector $\Gamma[n]$ together with the bounded variation approach (see Sec. 2.2.2). The hyperparameter C_p from the bounded variation defines the maximum allowed number of regime transitions and prevents the algorithm from overfitting the temporal variability of the parameter Θ . The hyperparameter K defines the number of sub-models (clusters/regimes) to resolve the temporal variation of Θ .

Excluding the autoregressive terms in the model (3.6), the time series is explicitly tested for the GRANGER causal relationship between σ_w and U_s as a way to assess if σ_w evolves according to the dynamics of U_s . Such a selective model structure uses the smallest set of model parameters and, in particular, tests the hypothesis of whether sub-mesoscale variability is critical for predicting the part of the TKE responsible for vertical transport.

Nevertheless, the autoregressive terms could improve the ability of the model to predict the variance of the variable σ_w . This was tested but is not reported. The results show that the residual of the clustering remains significantly non-Gaussian, suggesting that the representation of the dynamics in the variable σ_w with linear models is incomplete and requires further improvement. However, the goal of the following analysis is classification, and improvement of the model structure is addressed in the follow up study.

3.2.3 Estimation of the Persistence Parameter

The search for the optimal persistence parameter C_p of the FEM-BV-VARX method (see Sec. 2.2.2) is related to the search for a global minimum of the clustering functional. The minimization of (3.7) is difficult in this case because the variables σ_w and U_s are subject to different underlying distribution types. In the limiting case of strong stratification, the variable σ_w resembles an exponential distribution, and U_s resembles a Gaussian distribution. The assumption that the relationship between σ_w and U_s holds primarily in the strong SBL is insufficient to estimate reproducible regime segmentation and requires the development of an adapted global search algorithm.

The assumption that for the model (3.6) the residuals are Gaussian after parameter estimation is not met, as the study shows, and it is better to transform the variable before clustering (as it is done in Sec. 3.3.5). Moreover, the forcing variable oscillates around a mean value, while σ_w , in contrast, follows an extreme value distribution with exponential peaks and decays. Therefore, it is questionable whether the oscillatory patterns in U_s are present in the σ_w variable. The listed discrepancies also lead to difficulties in estimating the optimal number of clusters K . However, in this section, it is shown how to improve the solution procedure of the FEM-BV-VARX framework when one encounters difficulties finding reproducible solutions for the affiliation function $\Gamma(t)$.

The following study evaluates the reproducibility of the solution and suggests an optimal C_p value for a fixed value of K . One starts by measuring the goodness of the fit with a norm of the residual:

$$\|r\|^2 = \|x_s - x_m\|_2^2, \quad (3.8)$$

where x_s is the data and x_m is the signal modeled by FEM-BV-VARX. The residual decreases as C_p increases. The following experiments are performed with the variables σ_w and U_s , setting the memory depth of the model (3.6) to $q = 3$ and $K = 4$ for the rest of the study.

3. Nonstationary Analysis of the Stable Boundary Layer

In practice, it is common to run a minimization procedure several times to explore the search space and to identify the global minimum. Based on this idea, a small experiment is performed below to evaluate whether a similar solution can be observed repeatedly and how different it might be. In the experiment, the entire set of FEM-BV-VARX calls is divided into several blocks. Each block contains 500 individual realizations. Each run in a block is performed with a randomly generated initial value $\Gamma_0(t)$. The best result of a block, corresponding to the smallest residual, is shown in Fig. 3.2 for demonstration purposes.

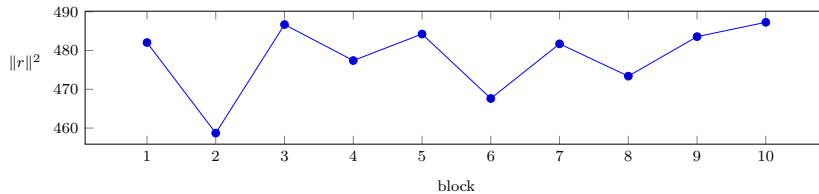


Figure 3.2: The residual value of the best solution is found in 500 consecutive runs and grouped into a block. The method FEM-BV-VARX is called 5000 times. The variability of the residual indicates uncertainty in the particular affiliation function identified.

The residuals of the best results over 10 blocks are in the range $[455, 490]$. The significance of this scatter is challenging to assess but suggests that each solution preferably represents a local minimum. As shown in Fig. 3.2, it is not straightforward to find a solution that has similar residual values every time. The vector $\Gamma(t)$ differs slightly for each of these solutions. So the question is: how much dissimilarity in the optimal solution $\Gamma(t)$ should one tolerate? This question is answered after analyzing the discrepancy between solutions. For this purpose, Fig. 3.3 shows the difference between the two best solutions found after 500.

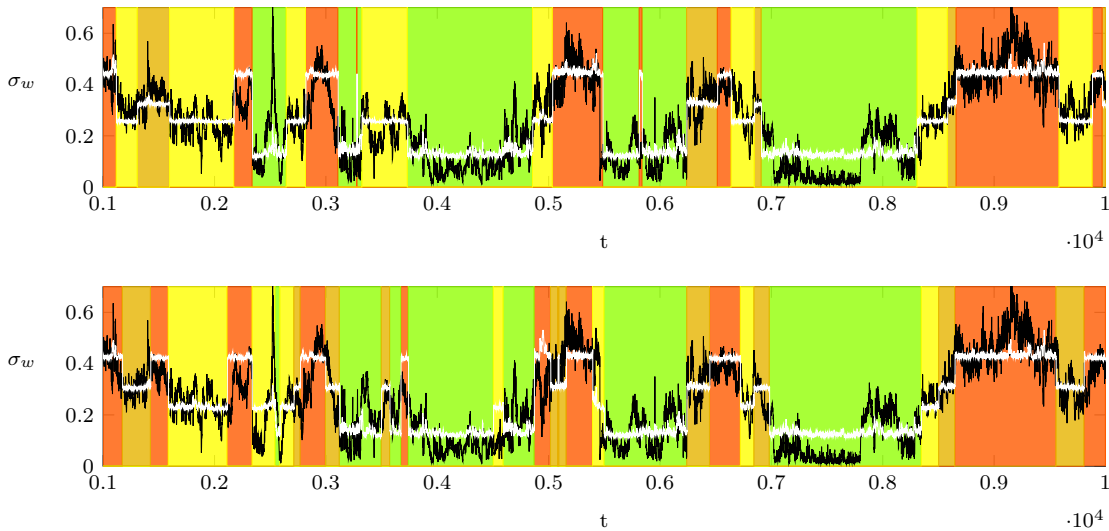


Figure 3.3: A plot of the estimated function $\Gamma(t)$ for the first (top panel) and second (bottom panel) best solutions from a pool of 500 solutions selected based on the residual value. Black is the signal of interest (σ_w); white is the modeled signal. The colors indicate the cluster states. Only a portion of the entire time series is shown. The entire time series is 54000 min long.

Comparing the results in Fig. 3.3, one can see some agreement between the $\gamma_i(t)$ functions. Indeed, the yellow clusters at $t = 0.2$, the green ones at $t = 0.45$; $t = 0.6$; $t = 0.77$, and the red ones at $t = 0.9$ are sufficiently similar. However, the transition sequences between these long clusters show dissimilarities. For example, at $t = 0.65$, the mean of the cluster states is

the same in both solutions, but the duration is different. This discrepancy in the duration of the cluster states can be tolerated because the resulting change in sample size within the cluster states is negligible for further statistical analysis. More importantly, the sequence of cluster states in the function $\gamma_i(t)$ remains the same between the different clustering results (see Fig. 3.3 at $t = 0.65$). In contrast, the shape of the cluster states at the following time points is clearly dissimilar: $t = 0.11$; $t = 0.25$; $t = 0.5$, which is undesirable and a sign of inconsistency.

Next, the correlation of the regime paths between two solutions is investigated. This allows the examination and estimation of uniqueness for a set of solutions. The format of the vector $\Gamma(t) \in \mathbb{Z}^{K \times N}$ is a binary vector. Multiplying it by the vector $[1, 2, \dots, K]$ from the left-hand side, one obtains a labeled path vector. The correspondence between the path vectors of different solutions is directly evaluated by counting the matching elements and calculating the percentage in relation to the total sample size. For example, comparing the first six best affiliation functions of a solution block, the following path correlation matrix is obtained:

$$\begin{array}{c}
 \begin{array}{cccccc}
 & 1 & 2 & 3 & 4 & 5 & \mathbf{6} \\
 \mathbf{1} & \left(\begin{array}{cccccc}
 100 & 69 & 73 & 72 & 59 & 72 \\
 \mathbf{2} & - & 100 & 65 & 68 & 65 & \mathbf{71} \\
 3 & - & - & 100 & 73 & 56 & 76 \\
 4 & - & - & - & 100 & 57 & 72 \\
 5 & - & - & - & - & 100 & 59 \\
 6 & - & - & - & - & - & 100
 \end{array} \right)
 \end{array}
 \end{array}$$

The two optimal solutions with the lowest residual value are marked red and match to 71%. This example shows that selecting solutions based on residual value alone does not guarantee that one will get unique solutions. For example, the most similarity is found between solution 3 and 6; their overlap is 76%. When working with the FEM-BV-VARX framework provided by the original authors, one can observe the following. When searching for the parameter range of C_p , the solver starts with the lowest C_p , i.e., long and steady states are identified. After finding the best solution within a block, the function $\Gamma(t)$ is used as the initial condition for the next block. This new block then contains the solutions found with the subsequent increase in the C_p value. By continuously increasing the C_p parameter, one will find that the lowest residual in most cases contains the solution of the previous best model (see Fig. 3.4). The idea of such an approach is to follow the path of the best solution found at the low value C_p by letting the solution incrementally explore the neighborhood for better solutions at a less restrictive regularization value.

However, in a series of experiments, it is shown that this strategy does not work optimally for the considered variables, since the subsequent increase of the parameter C_p alone does not cause the algorithm to find a better optimum on the path of C_p , and thus leads to a waste of computation time. This means that the best solution found at the lowest value of C_p remains the best, regardless of the change in C_p . This behavior is obvious because it is much more difficult to estimate a solution from a random guess at $C_p = 100$ than to propagate the best solution from $C_p = 5$, which is equivalent to refining the solution. This is illustrated in Fig. 3.5.

The black dots on the right side of the graph show the residual of a solution found by running the algorithm with a random initial guess for $C_p = 100$ (a total of 7300 minimizations

3. Nonstationary Analysis of the Stable Boundary Layer

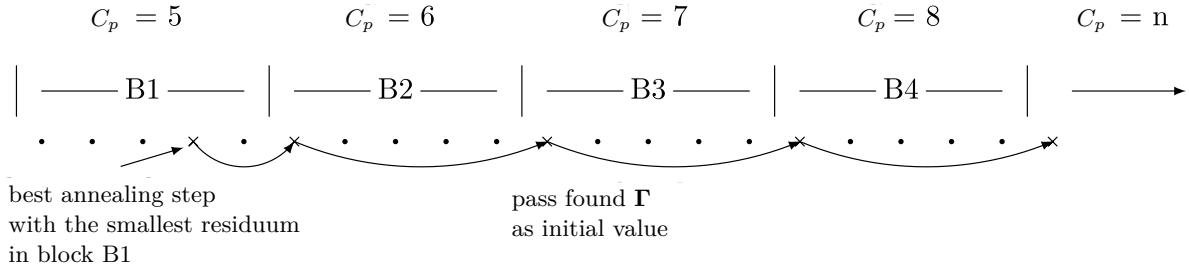


Figure 3.4: A sketch illustrating an approach to find an optimum using the FEM-BV-VARX method by varying the C_p value. In this example, each block has five annealing steps. The only frame parameter that changes between blocks is the regularization parameter C_p that controls the maximum number of jumps. The cross indicates the best solution with the lowest residual value. The first pass in a successive block is started with the optimal solution of the previous block.

were performed to produce the best four solutions in the figure). Consider the descending colored lines for comparison. These lines represent the path by which the best solution propagates from $C_p = 3$ to $C_p = 100$, producing a much better result. The procedure for obtaining such results is described below.

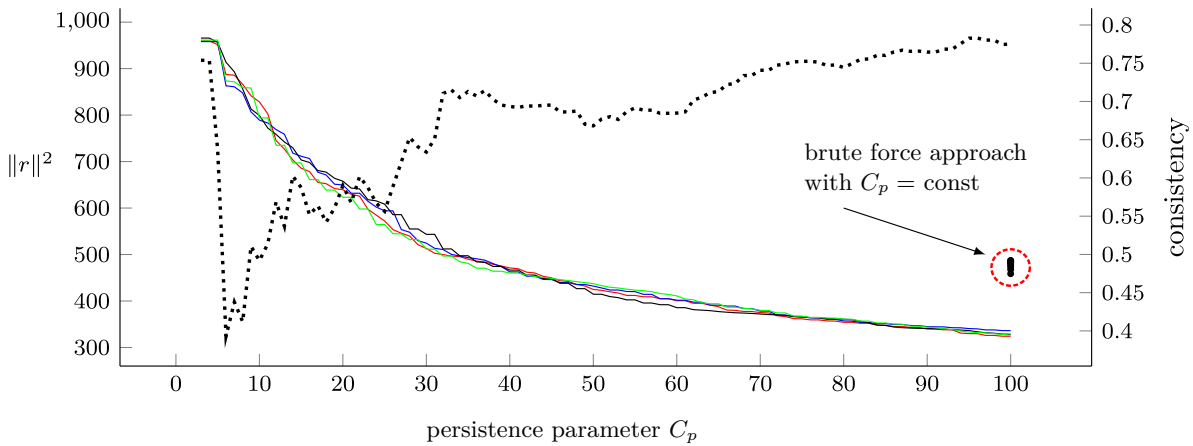


Figure 3.5: Residual value and consistency of the estimated affiliation vector plotted against the persistence parameter C_p . The colored solid lines refer to the left axis and denote the residual value obtained by the method shown in Fig. 3.6. The right axis shows the consistency of the estimated affiliation vector (dotted black line), which is measured as the percentage overlap between four solutions (marked with colored solid lines). The brute force approach denoted with the red circle cannot obtain lower residual values.

The procedure for finding a better solution is optimized by strengthening the search for the best solution at the smallest values of C_p . The idea is to find the solution with the smallest possible residual for small C_p values and propagate the $\Gamma(t)$ to the higher C_p values by transferring the optimal solution as an initial estimate only once. The procedure is outlined in Fig. 3.6. The phase in which the algorithm searches for the best solution with a fixed low value of C_p is called the search state, in which the best solution is found. In the subsequent confirmation state, the algorithm cannot find an improved solution for several blocks. Therefore, for the confirmation state, one specifies a limited number of blocks, after which the parameter C_p is incremented. By applying the above scheme, one obtains the solution shown in Fig. 3.5 with the left axis and the colored lines.

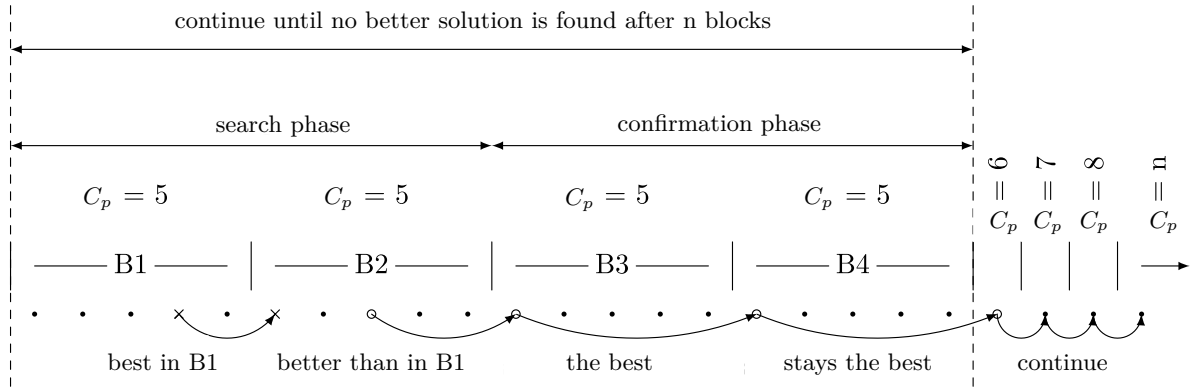


Figure 3.6: A sketch illustrating the modified algorithm for obtaining a better solution starting from the solutions at lower C_p values. The sketch demonstrates how to find the best possible solution with a lower C_p value and gradually increase the C_p parameter.

The pre-run routine provides a better solution in terms of residuals and is also faster. Most of the time is spent calculating the pre-run. In this example, a block contains 20 solutions, and to terminate the pre-run, the confirmation phase is set to be 50 blocks long.

Evaluation of several experiments showed that a better solution at a low C_p value did not lead to a better-propagated solution at higher values of C_p (see the colored lines in Fig. 3.5, which show four similar experiments).

The pre-run approach produces more consistent results. To investigate how the uniqueness of the solution changes with respect to C_p , we consider the consistency of clustering in all four pre-run solutions, labeled with different colors in Fig. 3.5. The respective consistency graph for the four minimizations is shown with a second y -axis in Fig. 3.5. The consistency of the solution improves by propagating the best solution. Compared to the local search approach, the per-run approach increases the consistency of a solution $\Gamma(t)$ from 34% to 78% at the value $C_p = 100$.

Figure 3.7 shows a similar experiment where the residual of the pre-run is included and the maximum C_p is set to 425.

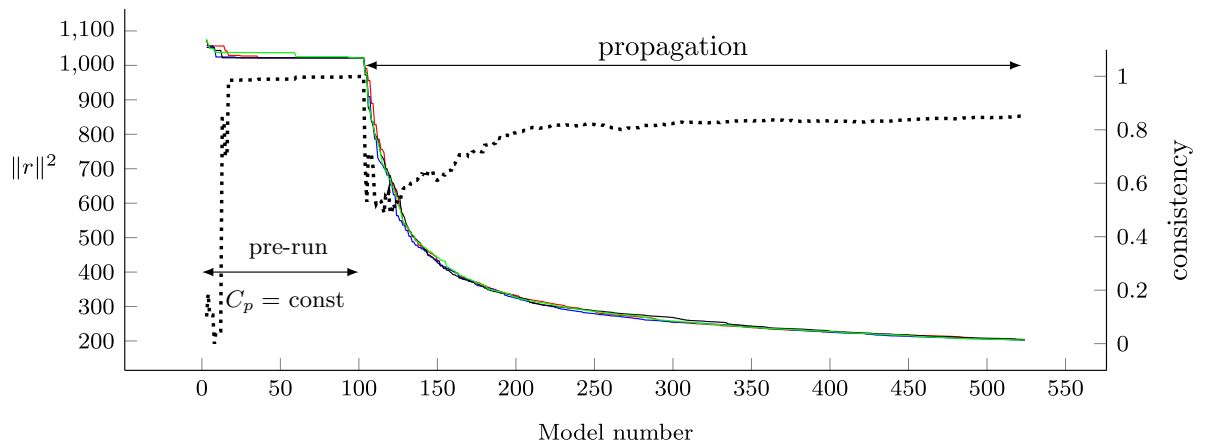


Figure 3.7: Analysis of the pre-run consistency. The colors indicate four repetitive calculations. The dotted line corresponds to the right axis. The right axis shows the consistency of the estimated membership vector (dotted black line), measured as the percentage overlap between four solutions (marked with colored solid lines).

The different color indicates four evaluations of the global search algorithm. In this case, the x-axis indicates the model number characterized by a C_p value. The drop in consistency at the beginning of the minimization procedure in Fig. 3.5 and 3.7 indicates that a good solution in the pre-run does not provide higher consistency for further propagation of the $\Gamma(t)$ result. This means that it is not necessary to stay in the pre-run state for a long time. Note that the solution has converged in the pre-run since the consistency is almost equal to one.

In summary, the pre-run approach provides a way to find the solution of Eq. (3.7) faster and with higher consistency in the regime transition path compared to the brute force approach. Moreover, the asymptotic convergence of the consistency measure provides the necessary criteria for selecting the parameter C_p when the residual does not provide relevant changes for that. The optimal C_p is chosen at the point where the consistency has reached a steady state. The method presented above is applied in the following section.

3.2.4 Clustering and Modeling Results

The results are based on a carefully conducted and extensive study by selecting the hyperparameters of the FEM-BV-VARX framework using the model (3.6). Namely, the number of clusters K , the regularization parameter C_p , and the memory depth q . Before using the framework, the memory depth is estimated by evaluating the partial cross-correlation function (Brockwell & Davis, 2002), averaged over 68 nights, between the variables σ_w and U_s . The analysis of the pacf shows that the memory size $q = 6$ (corresponding to 6 min for this analysis) is adequate. The parameter $C_p = 150$ and the optimal clustering solution is selected based on the procedure developed in Sec. 3.2.3 by examining the range $C_p = [2, 302]$.

Selecting the optimal parameter K for this particular analysis is challenging. This may be due to the weak relationship between σ_w and U_s during dominantly present, weakly stable regimes. Consequently, the estimated regime affiliation function is ambiguous and subject to variations in the duration of each regime state. The AKAIKE and BAYESIAN information criteria generally used to determine the optimal K (Metzner et al., 2012) are ineffective in this example. A range of $K = 2, 3, 4, 5, 6, 7$ values was tested, and the information criteria did not indicate a minimum with respect to K , and thus did not indicate an optimal K . Instead, the information criteria plot shows asymptotic behavior (not shown), suggesting that the optimum could only be found beyond $K = 10$ and is inappropriately complex for post-analysis.

A solution with a significantly large number of clusters is, however, plausible if the following property of the variables σ_m and U_s is considered. As reported by Vercauteren & Klein (2015), the relationship between σ_m and U_s is significant in a smaller study with mostly very stable conditions. In contrast, the amount of very SBL regimes in the selected 68 FLOSS2 nights was found to be low. Therefore, the periods with a correlation between σ_m and U_s are sparse relative to the entire time series.

Finding an appropriate number of clusters in such a case is problematic. There is a lack of significant correlation between σ_w and U_s due to weakly stable regimes. Therefore, the FEM-BV-VARX distinguishes the clusters based on their mean value. However, one of these clusters is the strongly stable regime (the regime of interest), in which the relationship between σ_w and U_s is relevant. In this regime, the respective parameter values of the model are

non-zero, which gives more weight to the functional compared to the regimes with weakly stable conditions, where only the parameter of the mean dominates.

Given the above data configuration, the FEM-BV-VARX method attempts to resolve precisely those regimes for which the given model structure provides the most significant responses to the clustering functional. However, to resolve the SBL in multiple meaningful regimes, the method must first split the existing weakly stable regimes into multiple clusters. As a result, the information criterion exhibits asymptotic behavior as the functional value falls faster than the penalty term increases as a function of K .

This problem is circumvented by the recursive application of the FEM-BV-VARX method. In the first clustering step, the time series is clustered with $K = 3$, resulting in clusters labeled C1, C2, C_s. Then, the strongly stable regime (C_s) is selected and clustered again with $K = 2$, leading to the clusters labeled with C3, C4. The second clustering is not biased by the weakly stable regimes and is therefore driven by the relationship between σ_w and U_s rather than the mean cluster value.

In this study, the bulk RICHARDSON number is used as an indicator for the stability of the flow

$$Ri_b = \frac{\Delta z(\bar{T}(z_2) - \bar{T}(z_1))(g/T_0)}{(\bar{U}(z_2) - \bar{U}(z_1))^2}, \quad (3.9)$$

where $\bar{U} = \sqrt{([u]^{1h})^2 + ([v]^{1h})^2}$ and $\bar{T} = [T]^{1h}$ are averaged on a time scale of 1 h. The resulting classified distribution of the Ri_b is reported in Fig. 3.8. The histogram confirms

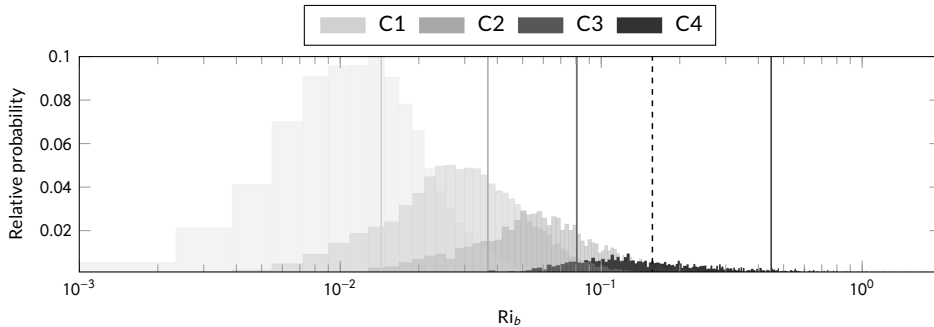


Figure 3.8: Histogram of bulk Richardson numbers in the clustered flow regimes. The vertical lines represent the median of the underlying distribution. From left to right, the values are $Ri_{b(C1)}=0.03$, $Ri_{b(C2)}=0.07$, $Ri_{b(C3)}=0.14$, $Ri_{b(crit)}=0.25$, $Ri_{b(C4)}=0.59$. The x-axis shows a maximum $Ri_{b(C1)}=2$ to zoom in on the distributions because the long tail of the cluster C4 extends to $Ri_{b(C1)}=6$. This also explains the abnormal median value.

that periods with $Ri_b > 0.25$ are rare in the selected 68 nights. The distribution of the most stable regime C4 shows a long tail that persists for values $Ri_b > 0.25$. Because of this tail, the median of the cluster is 0.59. An abnormal median value in C4 indicates that the strongly SBL should probably be free of turbulent mixing on average. Nevertheless, the distribution of the Ri_b number shows that the maximum value of the histogram is below $Ri_c = 0.25$ when considering the most probable Ri_b number, symbolizing probable weak mixing.

The performance of the model is presented in Fig. 3.9. The middle panel shows an example from the 68 nights, with the σ_w variable denoted in black and the prediction of this variable by the FEM-BV-VARX model colored in red. The middle panel also shows the clustered regime states by the background color. During the first clustering step, the time series is divided into

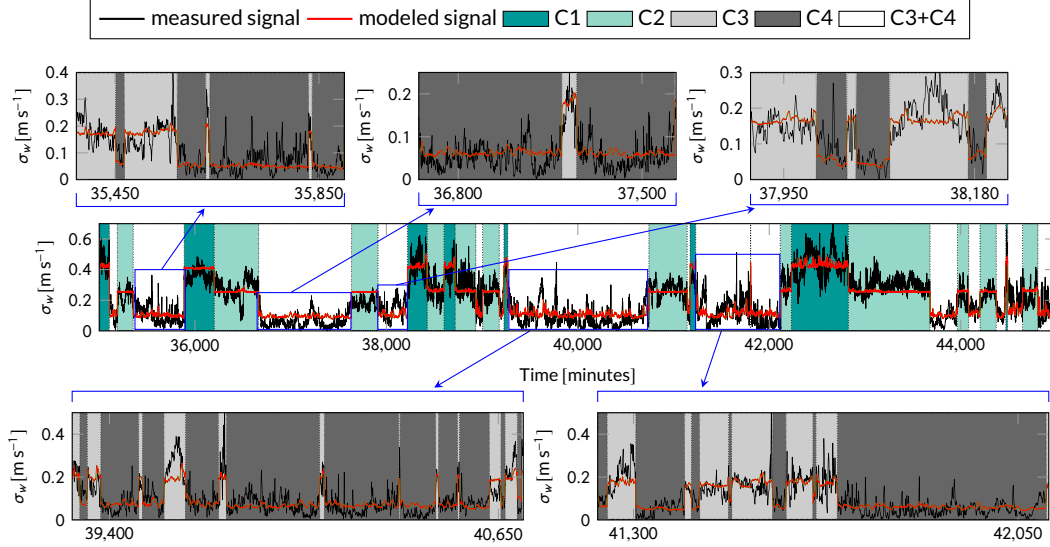


Figure 3.9: Classification and modeling of the 68 nights with the model defined in section 3.2.2, where the sub-mesoscales force the vertical fluctuations of the turbulent scales σ_w . The blue insets in the middle panel describe the recursive clustering result of the strongly stable regime discovered in the first clustering step. Only a fraction of the data is displayed.

three clusters. Then the periods highlighted in white are clustered again with $K = 2$. The result of the second clustering is shown in the top and bottom panels.

Looking closely at the insets in Fig. 3.9, one can see that the model can capture an oscillating pattern in the variable σ_w induced by the forcing variable U_s . This is particularly evident in the lower-left panel. Comparing regime C4 with the weakly stable regimes C1, C2, we find that these regimes have less variance and are mainly explained by the mean. Next, we analyze the scale iterations of the FLOSS2 dataset as a function of the four estimated regimes.

3.2.5 Regime-Dependent Multiresolution Decomposition

Multiresolution Decomposition (MRD) is a valuable technique for studying turbulent fluxes. In particular, when analyzing SBLs, turbulent momentum and heat fluxes decomposed into multiple dyadic scales with MRD provide insight into scale-averaged transport statistics. Examining turbulent transport with MRD allows us to identify the averaged turbulent time scale and thus evaluate the choice of scale in REYNOLDS averaging. In the previous section, the nonstationarity of turbulent mixing was divided into four regimes based on scale interactions, resulting in an ordered classification of the SBL. By applying MRD analysis in each regime, the nonstationary scale interactions are revealed, and their dependence on the flow stability is clarified.

In the following, the MRD is estimated according to Vickers & Mahrt (2003) using a wavelet basis set. The statistics of the MR spectra are obtained by evaluating MR spectra across multiple time windows. The MR spectrum estimated with wavelets is more accurate than one calculated with the FOURIER basis (Howell & Mahrt, 1997). The reason for this is the presence of sub-mesoscale events in the SBL. These events are localized in time and are therefore better captured with a wavelet approach.

The MRD calculation is performed within the clusters indicated by the respective Ri_b number: C1($Ri_b = 0.03$), C2($Ri_b = 0.07$), C3($Ri_b = 0.14$), C4($Ri_b = 0.59$) (see Fig. 3.8).

Regimes C1 and C2 are close to the neutral condition and serve as references rather than being the focus of the analysis. Therefore, the most interesting regime is C4, which exhibits supercritical stratification.

The calculation of the MR spectra in combination with the estimated regime affiliation vector $\Gamma(t)$ requires a detailed explanation. Consider the lower-left panel in Fig. 3.9, where the regime states of clusters C3 and C4 are shown with gray colors. The duration of some regime states for cluster C3 is shorter than the size of the processing window of the MRD, in this case, 30 min. On the other hand, the duration of the regime states in cluster C4 may span a few integer values of 30 min windows, leaving periods < 30 min unconnected.

Therefore, the length of the cluster states is taken into account when calculating the MR spectra. Cluster states shorter than 30 min are excluded. Parts corresponding to multiples of the 30 min window are taken from the more persistent states. Such a procedure avoids time series jumps that would artificially occur if the regime states were collected sequentially in a single array from a cluster.

Figure 3.10 shows the MRD of the heat flux for three heights (horizontal fields corresponding to 2, 15, and 30 m) and four clusters from C1 to C4 (vertical panels from weakly to strongly stable conditions). Cospectras are calculated over a window length of about 30 min. Boxplots are used for each scale to show the respective cluster statistics over several windows.

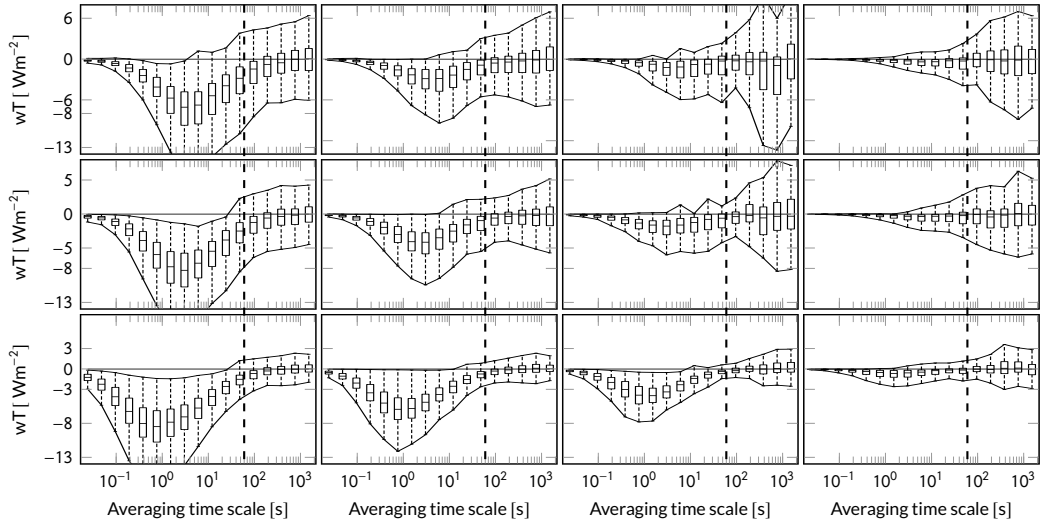


Figure 3.10: Multiresolution cospectra of the heat flux. Viewed from top to bottom, each row of panels corresponds to heights of 30 m, 15 m, and 2 m, respectively. Viewed from left to right, each column of panels corresponds to regimes C1-C4. Here C1 is a weakly stable regime and C4 is a strongly stable regime. The regime-corresponding Ri number is given in Fig. 3.8. The large dashed vertical lines in each panel mark the 1-min scale. Each box contains 17 boxplots, each representing the distribution of heat flux on a corresponding scale. The boxes represent the 25th and 75th percentiles, with whiskers connected by a solid line across the scales. The horizontal line in each box indicates the median.

The vertical dashed line in each of the panels in Fig. 3.10 represents the selected turbulent length scale equal to 1 min. The REYNOLDS averaging at this scale captures the relevant part of the average turbulent transport in all four regimes and satisfies the MOST assumptions only for the weakly stable regime. The TKE in weakly stable regimes is slightly underrepresented. Above the 1-min scale, the average turbulent heat transport flattens out in the clusters. This means that the average effect of this scale (on a scale of several days) is zero. Nevertheless,

3. Nonstationary Analysis of the Stable Boundary Layer

the variance of the distributions at scales larger than 1-min suggests that these sub-mesoscale motions are active and can contribute locally (on a scale of hours) to a positive or negative heat flux.

The importance of sub-mesoscale activity can be assessed by relating the variance of scales > 1 min to the median of scales < 1 min in Fig. 3.10. The reason for this is that the primary mechanism of sustained heat transport is KOLMOGOROV-like turbulence. These are the scales that cause sustained transport (scales < 1 min). For example, consider the lowest panel in the left column (C1 regime at the height of 2 m). In this case, the turbulence generated by the shear produced at the surface is responsible for the main transport. The sub-mesoscale variance is relatively small and may not contribute significantly to the transport; however, with increasing altitude (consider the C1 regime at 30 m heights in the upper left column), the sub-mesoscale variability increases so that sporadic events may contribute more strongly to the KOLMOGOROV-like turbulent mixing (on scale of hours).

Consider that the flow stability affects the KOLMOGOROV-like turbulent vortices so that they are suppressed. This becomes clear in the cospectra as we move from the weakly stable regime C1 to the strongly stable regime C4. Here, the averaged turbulent transport decreases on scales < 1 min, but not the variance on scales > 1 min. Since the effect of stability is not proportional along the scale spectrum, the sporadic sub-mesoscale variability dominates the heat transport at higher stabilities. This leads to intermittency as the sub-mesoscales are random with respect to the mean flow, and the KOLMOGOROV-like turbulent eddies are absent. Similar cospectra are shown by the MRD of the momentum flux in Fig. 3.11. The figure shows

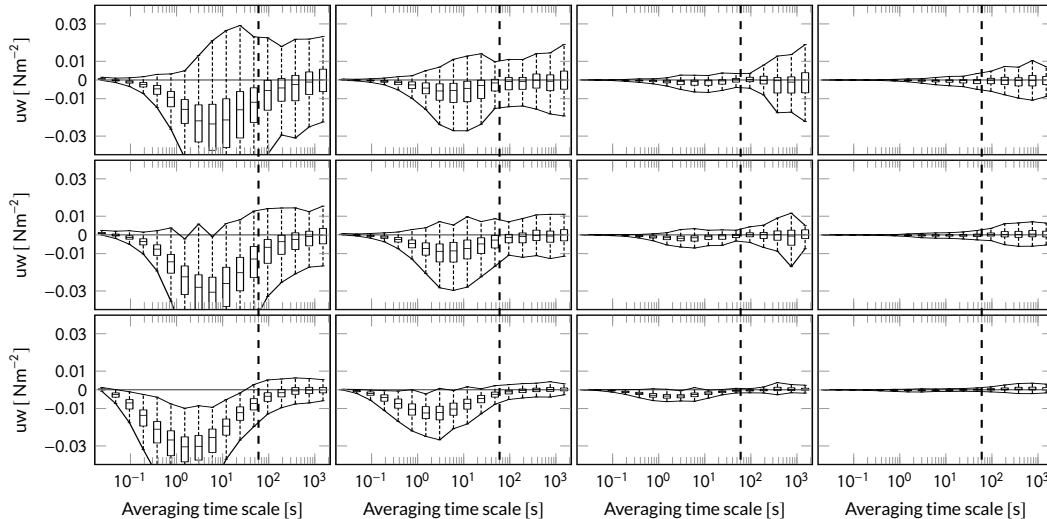


Figure 3.11: Multiresolution cospectra of the vertical momentum flux. Viewed from top to bottom, each row of panels corresponds to heights of 30 m, 15 m, and 2 m, respectively. Viewed from left to right, each column of panels corresponds to regimes C1-C4. Here C1 is a weakly stable regime and C4 is a strongly stable regime. The regime corresponding Ri number is given in Fig. 3.8. The large dashed vertical lines in each panel mark the 1-min scale. Each box contains 17 boxplots, each representing the distribution of heat flux on a corresponding scale. The boxes represent the 25th and 75th percentiles, with whiskers connected by a solid line across the scales. The horizontal line in each box indicates the median.

that in C2, C3, and C4, choosing a 1-min averaging scale is a good approximation. However, the TKE calculated in cluster C1 is slightly underrepresented in higher altitudes.

3.3 Classifying Turbulence Kinetic Energy

The clustering analysis in the previous section shows a systematic classification of regimes based on a statistical regression model that allows granular segmentation of the turbulent scale versus the flow stability (see Fig. 3.8). The core assumption of such a model is the interaction between the sub-mesoscale wind speed and the integrated vertical turbulent mixing.

Analysis of the clustered data provides valuable insight into the turbulence within the detected regimes. The model in the previous section confirms a valid relationship between σ_w and U_s , but the variance explained by the model should be improved. Therefore, the next step towards turbulence closure is to improve the predictive capability within the regimes by constructing an advanced model structure to use in the FEM-BV-VARX framework. Accordingly, building on the ideas of the previous section, the TKE will replace the σ_w variable, and the forcing variable is replaced by the vertical derivative of the wind speed filtered into different scales. In addition, the TKE is log-transformed to reflect the assumptions of the clustering framework better.

This section discusses both the modeling and analysis of the SBL resulting from an improved model structure, and it is organized as follows. The first step is to separate the velocity vector into the mean, jet, and submeso scales using wavelet filtering. Then, the most characteristic transient nights are analyzed to answer how the mean, jet, and sub-mesoscales resemble turbulence intensity and local stability.

The selected nights illustrate a typical case of complex multiscale interactions in nocturnal boundary layer flows. The three most characteristic nights were selected as examples from the entire dataset and ordered by decreasing mean wind to visualize the dynamics of the SBL. The rest of the nights (not shown) are an irregular mixture of the dynamics of the selected nights (Figs. 3.17, 3.18, 3.19). The height coordinate is normalized by the tower height (30 m) $z_* = z/z_{\text{tower}}$ and the time coordinate is normalized to the length of the night (15.38 h) $t_* = t/t_{\text{night}}$.

The model structure is introduced next. In contrast to the previous section, multiple heights are modeled but classified independently. Moreover, the estimated parameters in the boundary layer show and quantify the relationship between the vertical gradient of the considered scales and the turbulence response. The prediction performance of the FEM-BV-VARX model with the observed TKE confirms a well-explained variance.

A Turbulence-Submeso-Intensity (TSI) diagram is introduced to unify the analysis of scale interactions. The TSI diagram plots the normalized turbulent scales as a function of the normalized sub-mesoscales. The normalization factor here is the mean wind. The appropriate subdivision of the scales using wavelet filtering is unique to the dataset and is motivated by the classification results from the previous study (see Sec. 3.2).

To develop a better understanding of the TSI diagram, it is examined in terms of several variables. For example, the regions characterized by the values of ζ and the shape of the mean wind profile provide information about the boundary layer structure and classify the phase space of the TSI diagram into physically meaningful regions. Moreover, the point cloud of the TSI diagram is classified by relating it to the stability value ζ and the statistical classifier obtained with the applied FEM-BV-VARX model. Table 3.3 relates the stability value ζ to

the statistical classifier by summarizing the regime-dependent statistics of stability for each measurement height.

The study of the statistical clustering of the flux-profile relationship (see Sec. 3.3.8) led to ideas for developing the turbulence model presented in Ch. 5.

3.3.1 Data Preprocessing Strategy

The data preprocessing steps are described in the order in which they are performed. This order is important to minimize the adverse effects of filtering in the boundary regions between the selected nights. These boundary regions occur when stable conditions are extracted from continuous records.

The data processing workflow for this study is summarized in Fig. 3.12. The total dataset contains 132 days of continuously recorded time series. In individual cases, the instruments were covered with ice and therefore did not register any records, resulting in long data gaps (on the order of several hours). The days with these long periods and those with unrealistically large velocity gradients were removed, leaving 102 days for analysis. The remaining shorter minute-scale data gaps (quality control results) were linearly interpolated in time to increase continuity for the FEM-BV-VARX clustering method.

The periods during which the flow passes through the tower have been preserved compared to the previous study (see Sec. 3.2.1). This is justified by the fact that the vortices induced by the tower are of the same order of magnitude as the truss rods (about 10 cm). It is assumed that the fluctuation energy induced by the truss rods is small compared to the 1-min turbulent scale of the boundary layer. Removing the periods when the flow crosses the tower for comparison, as was done in Sec. 3.2.1, results in a smaller dataset of 68 days, which unfortunately excludes relevant parts with strongly stable periods. Therefore, 102 nights are considered in this analysis. Here, the velocity vector is rotated in the mean wind direction twice using a sliding window approach. The natural mean wind direction used here is calculated from the 30 m sonic anemometer using a moving average on a scale of one hour.

The velocity vector is separated into several scales using two filtering methods. First, a running time average is applied to estimate the REYNOLDS stress and turbulent heat flux. Then, the larger scales are filtered into different scale ranges using the DWT (see Sec. 3.3.3). The nighttime is then selected based on the average negative heat flux, where the average is based on 24-hour records (see Fig. 3.13). Nighttime is selected after wavelet filtering to mitigate the boundary effects of the wavelet transform, known as the cone of influence of a mother wavelet (Torrence & Compo, 1998). Vertical gradients are computed with finite differences: central differences within the boundary layer and one-sided at the lowest and highest measurement points. The unequal mesh size is taken into account. In the next step, the applied clustering procedure uses the estimated gradients as exogenous variables in the model structure. After performing the clustering of the TKE, the resulting affiliation function is used to classify the TSI diagram and the flux-profile relationship.

The following section defines the scales for the analysis in question. Details on the filter techniques follow in the next section.

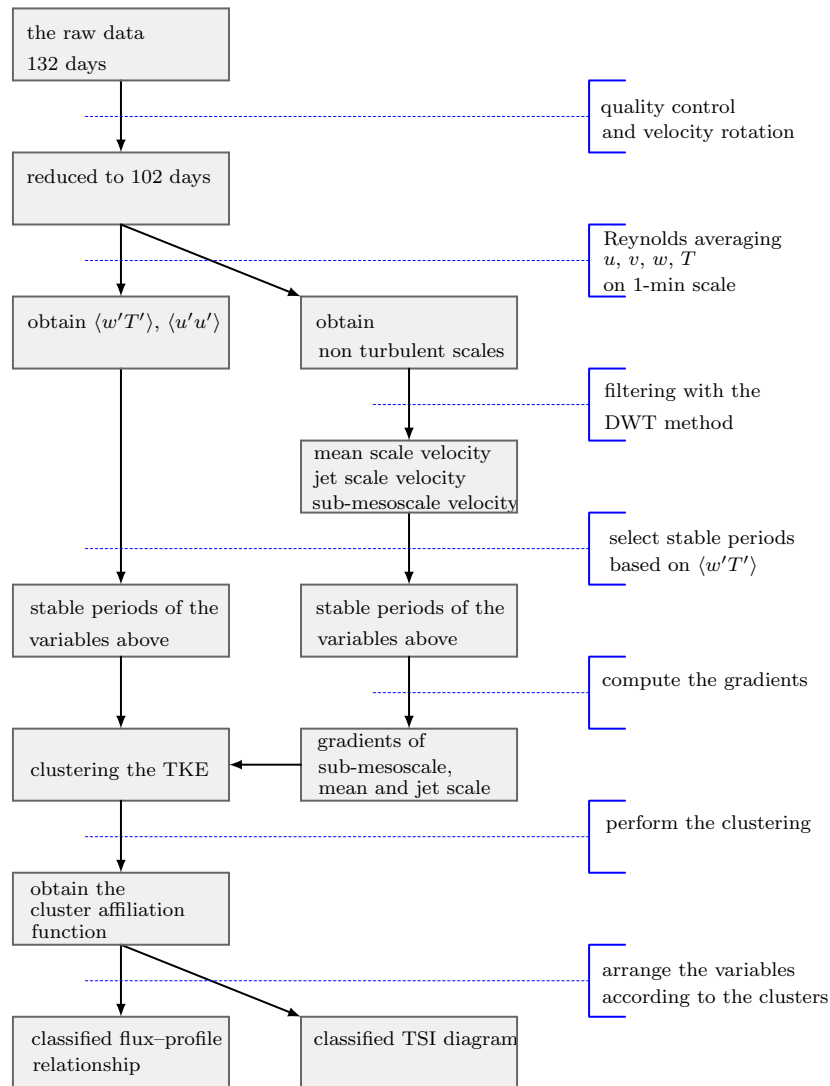


Figure 3.12: Data analysis flowchart. The gray rectangles show the results of the individual processing steps. The blue horizontal section lines mark the processing steps. The flux-profile relationship and the TSI diagram are presented in Sec. 3.3.7 and Sec. 3.3.8, respectively. The non-turbulent scales at the beginning of the middle diagram column are assumed to be non-turbulent. The chosen REYNOLDS scale is adjusted for the stable regime. To assess the error for the neutral state, see Fig. 3.11.

3.3.2 Definition of Scales

Figure 3.14 proposes the separation of scales for the analysis. The idea is first to apply REYNOLDS averaging to separate the KOLMOGOROV-like turbulent scales. The resulting mean scale after REYNOLDS averaging is further decomposed by wavelet filtering. The reason for the mixed filtering is that the turbulent eddies are continuously present, and therefore one does not require information about their temporal location. In contrast, the (assumed) non-turbulent motions are transient and not present every night, so wavelet filtering is better suited to filter them out accurately.

In the previous section, a turbulence scale was suggested based on the calculated momentum and heat flux cospectra (see Fig. 3.10). The analysis indicates that scales below one minute correspond mainly to turbulence in the nighttime FLOSS2 dataset; accordingly, this threshold is used here in the first step of the REYNOLDS averaging. In the following, the resulting

3. Nonstationary Analysis of the Stable Boundary Layer

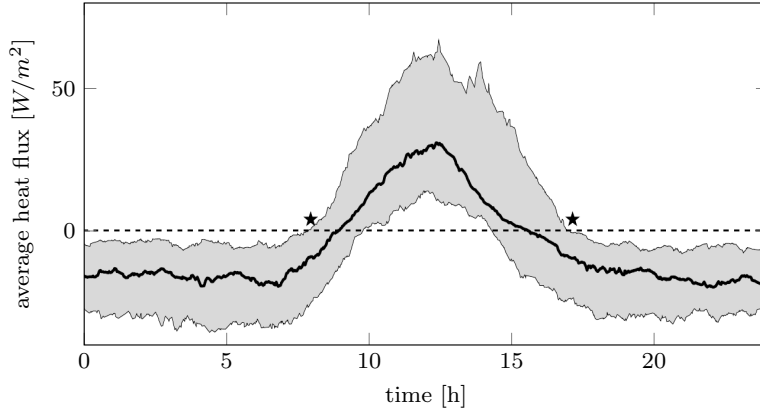


Figure 3.13: Average heat flux of the FLOSS2 dataset based on 24-hour records. The solid thick black line is the median. The gray area marks the interquartile range. This plot is used to identify the time series with stable conditions for clustering analysis. The stars mark the beginning and end of the unstable conditions, which are removed.

REYNOLDS-averaged mean signal is divided into three frequency bands, which are then used for the clustering procedure (see Sec. 3.3.5). The turbulent fluctuations are then $\mathbf{u}' = \mathbf{u} - \widehat{\mathbf{u}}$, where $\widehat{(\cdot)}$ denotes a dataset specific estimated turbulent scale. Next, $\widehat{\mathbf{u}}$ is filtered into different scales and then used to estimate the vertical gradients. Linear filters constructed in the FOURIER domain may not filter $\widehat{\mathbf{u}}$ correctly due to the nonstationarity of the non-turbulent scales. The DWT, implemented as a filter bank in the Python package of Lee et al. (2019a), is used to perform the filtering.

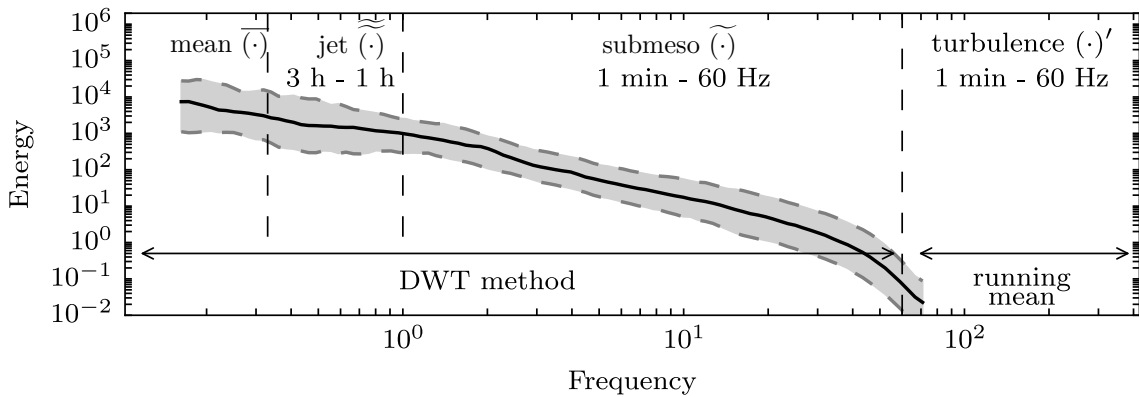


Figure 3.14: Illustration of the defined scales for the analysis of the FLOSS2 dataset and the filtering methods used. The vertical dashed lines from left to right indicate the frequencies of 3 h, 1 h, and 1 min. The 1 min scale is estimated in Fig. 3.10.

Filtering with DWT further divides the velocity vector into three scale ranges $\widehat{\mathbf{u}} = \bar{\mathbf{u}} + \tilde{\mathbf{u}} + \tilde{\tilde{\mathbf{u}}}$ (see Sec. 3.3.3 and for an example see Fig. 3.16). The sub-mesoscale motions $\tilde{\mathbf{u}} = (\tilde{u}, \tilde{v}, \tilde{w})$ occupies the period from one minute to one hour, the jet scale $\tilde{\tilde{\mathbf{u}}} = (\tilde{\tilde{u}}, \tilde{\tilde{v}}, \tilde{\tilde{w}})$ occupies the range from one to three hours, and the mean scale $\bar{\mathbf{u}} = (\bar{u}, \bar{v}, \bar{w})$ is three hours and above (see Fig. 3.14).

A three-hour mean scale is selected because it is large enough to provide a logarithmic wind profile from the surface to 30 m. The jet scale was introduced for several reasons. This scale has enough energy to create turbulence close to the ground but not enough to create a

logarithmic profile up to 30 m. Also, the jet scale adds inflection points to the mean wind profile, changing the height at which a maximum wind speed is observed. DWT filtering aims to find a mean scale where the highest measurement point (30 m) corresponds to a wind maximum, and thus the MOST is applicable. Any fluctuations faster than three hours are considered as unknown fluctuations. Therefore, the scaling velocity for the following analysis is

$$\bar{u}_\infty(t) := \sqrt{\bar{u}(z = 30, t)^2 + \bar{v}(z = 30, t)^2}, \quad (3.10)$$

with a band limit of three hours (see Fig. 3.14). By normalizing with $\bar{u}_\infty(t)$, unknown fluctuations faster than 3 h are examined in terms of the well-known similarity theory. On the scale of 3 h, the boundary layer structure for the dataset with the respective measurement height of 30 m shows canonical boundary layer behavior.

The presented scale separation is valid only for the considered dataset because, as mentioned before, the mean scale depends on the measurement height. Therefore, the exact definition of the mean scale may be different in each study. However, the objective judgment is to choose the mean scale for a controlled height, such that the boundary layer follows the MOST. The following section describes the filtering methods.

3.3.3 Filtering Methods

The decomposition of the scales is done in two filtering steps. In the first step, the turbulent fluctuations are separated using the running time average:

$$\widehat{X}(t) = \frac{1}{2t_e} \int_{t-t_e}^{t+t_e} X(t^*)G(t^*)dt^*, \quad (3.11)$$

where $2t_e$ equals 60 seconds, so $\widehat{\cdot}$ denotes the time average on a scale of one minute. The filtered velocity component is denoted with $X(t)$, and G is the convolution kernel equal to 1 (i.e., this corresponds to the block averages). The turbulent fluctuations are then $\mathbf{u}' = \mathbf{u} - \widehat{\mathbf{u}}$, where $\mathbf{u} = (u, v, w)$ is the measured wind vector. Suppose the boundary layer is in the weakly stable region with typically larger turbulent eddies. In that case, the 1-min average will result in an underestimated TKE because the filter scale is not matched to the turbulence scale. A moving variance is used to calculate the TKE:

$$\widehat{X'X'} = \frac{1}{2t_e} \int_{t-t_e}^{t+t_e} X'(t^*)X'(t^*)dt^*, \quad (3.12)$$

where $2t_e$ is equal to 60 seconds and the mean has been removed before (see Eq. (3.11)). Consequently, the turbulent kinetic energy $e := 0.5(\widehat{u'u'} + \widehat{v'v'} + \widehat{w'w'})$. The running-window statistics allow the selection of the sampling frequency for the subsequent analysis. After filtering with Eq. (3.11) and (3.12), the sampling frequency of $e, \widehat{\mathbf{u}}$ is unchanged ($f_s = 60$ Hz), but the frequency content has changed. The highest frequency in $\widehat{\mathbf{u}}$ is 60 cycles/hour and is the *band limit* f_b of a signal. According to Billings (2013, p.476), SHANNON's theorem states that to recover all information in a signal band limited to f_b Hz, the signal should be sampled at least with $2f_b$ Hz. For parameter estimates, a sampling rate of about $5f_b/2$ is often sufficient. Consequently, the variables: $e, \widehat{\mathbf{u}}, \widehat{w'T'}$ are down sampled from $f_s = 60$ Hz to 180 cycles/hour, resolving the 1-min fluctuation with three points. This sampling rate is important for extracting

3. Nonstationary Analysis of the Stable Boundary Layer

statistical relationships between scales when using the FEM-BV-VARX clustering approach (see Sec. 3.3.5).

In the second step, wavelet filtering is performed. A DAUBECHIES wavelet of order 20 is chosen as the mother wavelet (ψ). It forms an orthonormal basis:

$$\int \psi_{i,j}(t)\psi_{n,k}(t)dt = \delta_{in}\delta_{jk} \quad \psi_{i,j}(t) = \frac{1}{\sqrt{2^i}}\psi\left(\frac{t-j2^i}{2^i}\right), \quad (3.13)$$

where δ_{ij} is the Kronecker delta function, i is the dilation, and j is the translation index of the dyadic scale. Thus, functions can be approximated by a linear combination with arbitrary accuracy if the basis is chosen appropriately. The coefficients of a wavelet transform are:

$$D_{i,j} = \langle f, \psi_{i,j} \rangle \equiv \int_{-\infty}^{\infty} f(t)\psi_{i,j}(t)dt, \quad (3.14)$$

where $f(t)$ stands for the analyzed function. The dyadic scale discretization with wavelets allows a sparse representation of a signal in 1D and provides a time scale representation of a process (Kumar & Foufoula-Georgiou, 1997):

$$f(t) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} D_{i,j}\psi_{i,j}(t). \quad (3.15)$$

The reason for using a dyadic filter bank is efficiency. The basic idea is to construct a cascade of high and low-pass filters (Mallat, 1999). After a decomposition stage, the output consists of approximation coefficients (low-frequency band) and detail coefficients (high-frequency band). Then the approximation coefficients are reduced by a factor of two and the discrete convolutions with low-pass and high-pass filters are repeated. The filter bank can have multiple decomposition levels, and by back-transforming the cascade, the signal is fully reconstructed. By setting the wavelet coefficients at a decomposition level to zero, the user removes unwanted scales and thus can extract those of interest. According to the NYQUIST sampling theorem, the highest frequency of the signal is $180/2$ c/h = 90 c/h. After the first stage of the decomposition, the detail coefficients contain frequencies between $45 - 90$ c/h. The complete filter bank cascade consists of nine stages and is summarized in Tab. 3.1.

The choice of a wavelet basis depends on the time series being analyzed. A suitable basis must represent most of the signal's energy content with the least number of wavelet coefficients (Kumar & Foufoula-Georgiou, 1997). The SHANNON measure of entropy can be used to find a basis from the considered library of wavelets, as shown by Katul & Vidakovic (1996). Here, different reasoning is used.

The scales ($\bar{\mathbf{u}}, \tilde{\mathbf{u}}, \tilde{\mathbf{u}}$) need to be separated in physical space since the study aims to investigate from which scales the TKE is triggered. To separate scales, a basis with strong localization properties in the frequency domain is needed. Weak localization properties are characterized in particular by a smooth frequency response function. The function extends into the adjacent bands beyond the predefined cutoff in the spectral domain. This results in energy loss across scales. This weakness of the DWT is referred to as spectral leakage (Qiu et al., 1995). The sensitivity study of Peng et al. (2009) shows that this problem can be mitigated by selecting a higher order mother wavelet, resulting in a sharper spectral cutoff. Higher-order wavelets

require a higher temporal resolution of the signal. The selected sampling frequency of 180 c/h in this analysis supports these requirements.

Table 3.1: Definition of the discrete-wavelet-filter-bank cascade.

Level	Frequency [cph]	Scale label	Time scale
0	<60	turbulence	
1	60 – 45	sub-mesoscale	1 min
2	45 – 22.5	sub-mesoscale	
3	22.5 – 11.2	sub-mesoscale	
4	11.2 – 5.6	sub-mesoscale	
5	5.6 – 2.8	sub-mesoscale	
6	2.8 – 1.4	sub-mesoscale	1 h
7	1.4 – 0.7	jet	
8	0.7 – 0.35	jet	3 h
9	>0.35	mean	

To analyze the scale interactions with a TSI diagram (see Sec. 3.3.7) the definition of Turbulence Intensity (TI) (Pope, 2000) is

$$\text{TI} := \frac{1}{\bar{u}_\infty} \sqrt{\frac{1}{3}(\widehat{u'u'} + \widehat{v'v'} + \widehat{w'w'})}, \quad (3.16)$$

and the Submeso Intensity (SI) is defined analogically as

$$\text{SI} := \frac{1}{\bar{u}_\infty} \sqrt{\frac{1}{3}(\tilde{u}^2 + \tilde{v}^2 + \tilde{w}^2)}, \quad (3.17)$$

denoting the energy content of the sub-mesoscales. The jet intensity is not analyzed because it is expected to generate turbulence mainly due to ground shear, similar to the mean scale. The horizontal wind speed for the mean, jet, and sub-mesoscale bands is defined respectively as: $\bar{U} = \sqrt{\bar{u}^2 + \bar{v}^2}$, $\tilde{U} = \sqrt{\tilde{u}^2 + \tilde{v}^2}$, $\tilde{U} = \sqrt{\tilde{u}^2 + \tilde{v}^2}$.

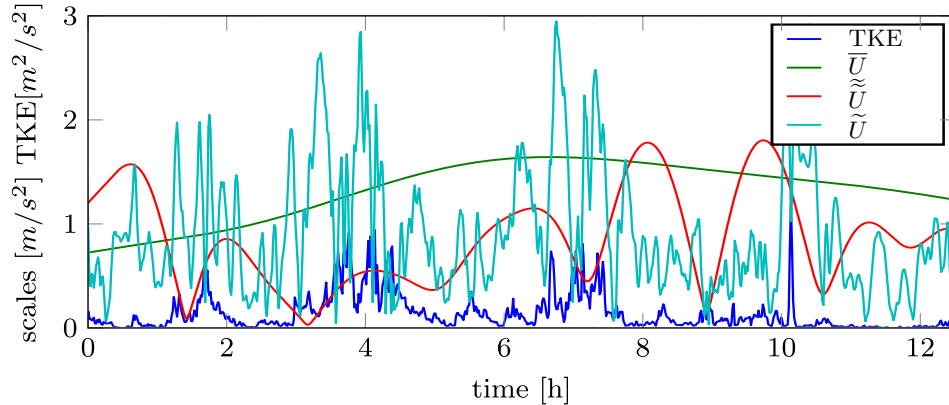


Figure 3.15: An example of an intermittent mixing events and scale activity. The TKE is calculated on the scale of 1 min. The scales are defined in Fig. 3.14. The value of the mean wind is too low to sustain turbulence. However, it is interesting to note that the sub-mesoscale activity is consistent with the intermittent bursts in the TKE signal, but not with the jet scale.

The choice of filtering is based on the following consideration. A possible alternative would be to use the DWT to filter the entire spectrum. But in this case, the choice of wavelet basis is

not obvious. As noted by Kumar & Foufoula-Georgiou (1997), the HAAR wavelet is a good and simple choice for applications where the process has sharp fluctuations. In the inertial subrange, for example, this would be a reasonable basis. A high-order DAUBECHIES is more appropriate outside the turbulence cascade, where sub-mesoscale motions are reported as smoother signals. As the scale changes from turbulent to submeso to jet to mean scale, the wavelet basis should continuously change from HAAR wavelet to DAUBECHIES. As mentioned earlier, the localization property of wavelets is only important for the sub-mesoscales. Therefore, it is easier to filter the turbulent scales with REYNOLDS averaging.

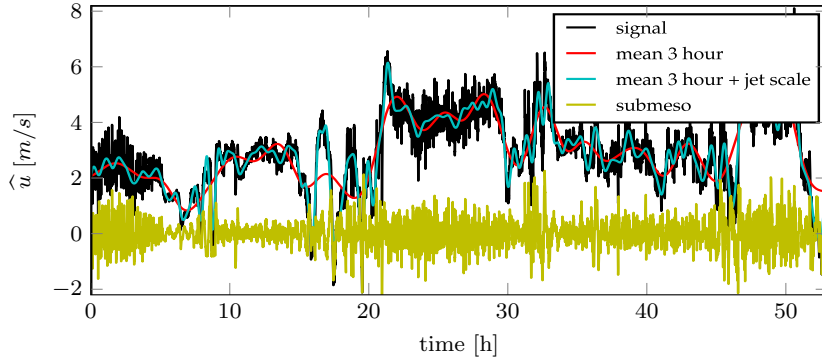


Figure 3.16: Example demonstrating the filtering results of the \hat{u} component with the DWT following the scales definition in Fig. 3.14.

3.3.4 Scalewise Visualization of the Wind Profiles

In this section, the results of wavelet filtering are presented by plotting wind profiles at different scales for three specific nights. The objective here is to visualize the variety of scale interaction patterns.

The observational study begins by evaluating the scale activity for a night (see Fig. 3.17) with a moderate averaged wind speed of $\mathbb{E}(\bar{u}_\infty) = 5.7 \text{ m s}^{-1}$. Within FLOSS2, nights with $\bar{u}_\infty > 6 \text{ m s}^{-1}$ unmistakably show a well-developed logarithmic mean wind profile along with a canonical TKE profile. The profiles are related to the neutral or weakly stable regime and do not show any interesting dynamics relevant to this study and therefore serving as a reference. Similarly, a critical wind speed of $5\text{--}7 \text{ m s}^{-1}$ was found by Van de Wiel et al. (2012) above which turbulence is active in a classical manner, for clear sky conditions.

Intermittency is weakly present in the weakly stable conditions, where the local stability exhibits weak spatio-temporal (vertical space) variations (see Fig. 3.17e, where $t_* \in (0.3, 0.5)$ and $t_* \in (0.8, 1.0)$). The mean velocity profile during this night is well pronounced and appears to be relatively stable (see Fig. 3.17a). The TI is on average 0.1 and shows an extreme only at $t_* = 0.4$ (see Fig. 3.17d). This peak coincides with increased sub-mesoscale and jet-scale activity (see Figs. 3.17b,c). Looking more closely at the scale fluctuations of the submeso motions (see Fig. 3.17c), we detect an increase in relative magnitude at $t_* \in (0.1, 0.2)$ and $t_* \in (0.7, 0.9)$. For the time $t_* \in (0.7, 0.9)$, the fluctuation of the sub-mesoscale band visually matches the streaky pattern in the TI, while the jet-scale band is less pronounced. In contrast, the jet scale during $t_* \in (0.1, 0.2)$ slows down $t_* \in (0.1, 0.15)$ and then accelerates at $t_* \in (0.15, 0.2)$. At this point, the TI is responding to the temporal evolution of the jet rather

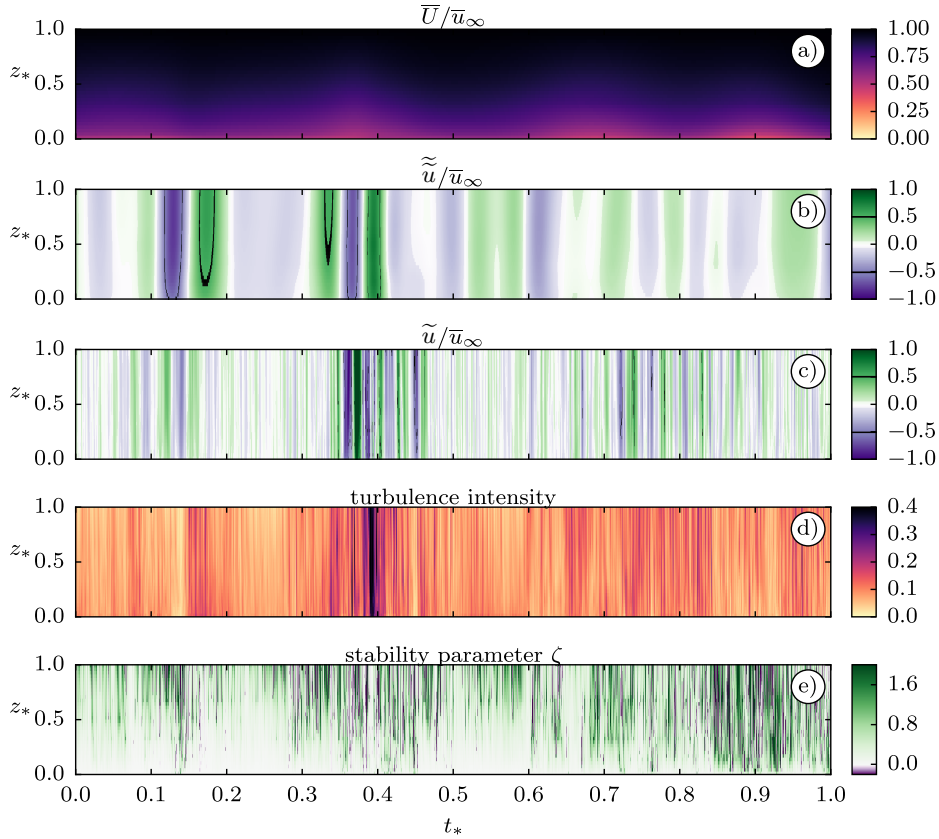


Figure 3.17: Scale activity and turbulence response for a night with moderate wind. The wind speed averaged over this night is $\mathbb{E}(\bar{u}_\infty) = 5.7 \text{ m s}^{-1}$. The panels show: a) the mean wind profile, b) the jet scale, c) the sub-mesoscale, d) the TI, and e) ζ . For the jet and sub-mesoscales, only the component in the direction of flow is shown to identify the times of deceleration. The component across the flow behaves similarly.

than the sub-mesoscale (see Fig. 3.17d with $t_* \in (0.1, 0.2)$). The spatio-temporal structure of the TI exhibits weak vertical variation. No significant variability in height is observed in either the jet or sub-mesoscales. It can be concluded that, on average, this night is in a weakly stable condition with insignificant departures as confirmed by the stability parameter (see Fig. 3.17e).

Figure 3.18 shows a night with an averaged velocity magnitude of $\mathbb{E}(\bar{u}_\infty) = 2.7 \text{ m s}^{-1}$ and significant deviations in the profile of the mean wind (see Fig. 3.18a). The mean velocity profile during this night is unsteady and shows at the beginning $t_* \in (0.1, 0.3)$ and at the end of the night $t_* \in (0.8, 1.0)$ rather a laminar profile. Irregularities in the profile exist in the middle of the night $t_* \in (0.4, 0.7)$.

The ζ value confirms the complex evolution of the boundary layer (see Fig. 3.18e). The stability parameter indicates a shallow boundary layer at $t_* \in (0.05, 0.1)$, followed by a sequence of fully to partially neutral states $t_* \in (0, 1, 0, 3)$. This is followed by an irregular variation from neutral to strongly stable states at $t_* \in (0.3, 0.5)$ and the boundary layer decouples from the ground. This dynamics of ζ is then accompanied by high-frequency noise in ζ throughout the range of stability values between unstable and stable states (see $t_* \in (0.5, 0.9)$ Fig. 3.18e). In the first half of the night ($t_* \in (0.3, 0.5)$), the ζ response occurs more on a larger scale; in the second part of the night, ζ resembles more small-scale dynamics. The average relative magnitude (estimate from the figure) of the jet and sub-mesoscale (see Fig. 3.18b,c) is about

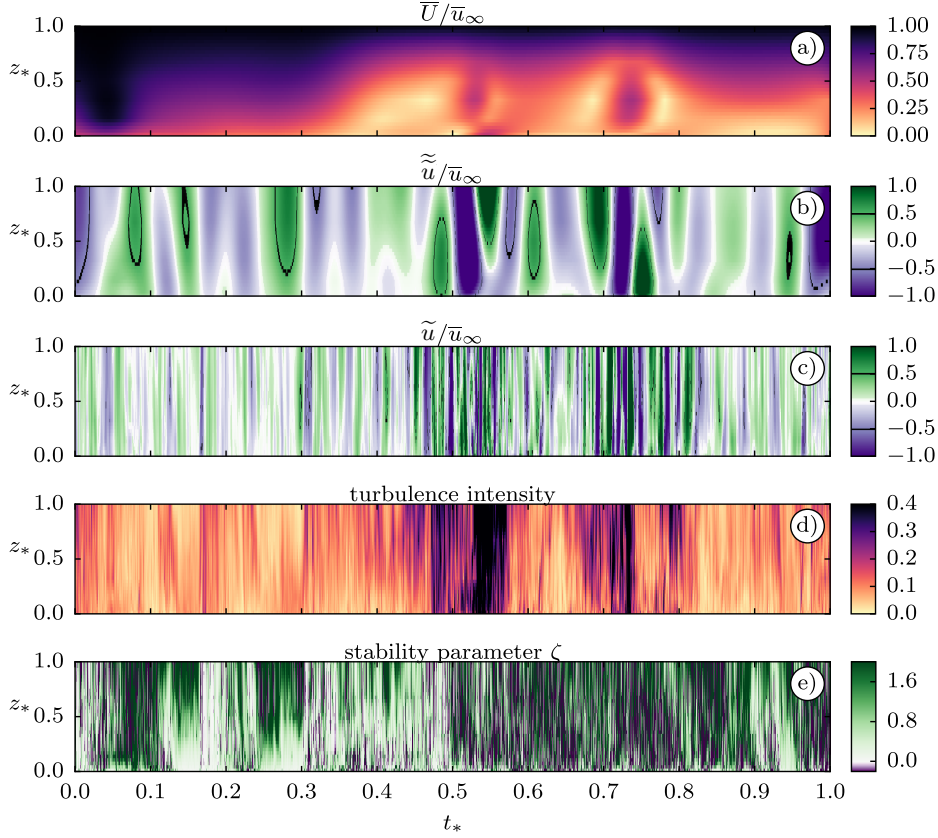


Figure 3.18: Scale activity and turbulence response for a night with weak wind. The wind velocity averaged over this night is $\mathbb{E}(\bar{u}_\infty) = 2.7 \text{ m s}^{-1}$. The panels show: a) the mean wind profile, b) the jet scale, c) the sub-mesoscales, d) the TI, and e) ζ . For jet and sub-mesoscales, only the streamwise component is displayed to identify times of deceleration. Cross-stream component behaves similarly.

0.5 times that of \bar{u}_∞ . In the second half of the night, the jet and submeso scales are more pronounced than in the first half of the night because of the lower mean-scale energy. Sharp peaks in turbulence intensity at $t_* \in (0.5, 0.6)$ and $t_* \in (0.7, 0.8)$ correspond to increased absolute values of the jet and submeso scales along with a relatively comparable shape of the mean wind profile $t_* \in (0.4, 0.8)$.

The variability of the jet and sub-mesoscales with respect to height becomes more evident at night with the lower mean wind (compare above in Fig. 3.18b,c with fig. 3.19b,c). The significance of the vertical gradients increases and therefore is used later in the construction of a model structure for the clustering analysis. At both scales, the sporadic accelerations occupy significant portions of the boundary layer. In general, the TI is intermittent in time and shows evidence of a pronounced dependence on height (see Fig. 3.19d; $t_* \in (0.3, 0.4)$). For wind speeds below $\mathbb{E}(\bar{u}_\infty) = 2.7 \text{ m s}^{-1}$ the energetically largest scale should be the jet scale.

Figure 3.19 shows a night with an averaged velocity magnitude of $\mathbb{E}(\bar{u}_\infty) = 1.3 \text{ m s}^{-1}$. The variability of the sub-mesoscales is of the order of 1 m s^{-1} . The profile of the mean scale is irregular (see Fig. 3.19a). The TI responds more to the sub-mesoscale band than to the jet scale. This response can be inferred from the patterns (see Figs. 3.19c,d). Because of their chaotic behavior, it is difficult to derive structural information from these profiles. There is a noticeable artifact at the end of the night ($t_* = (0.9, 1.0)$) that is likely due to filtering.

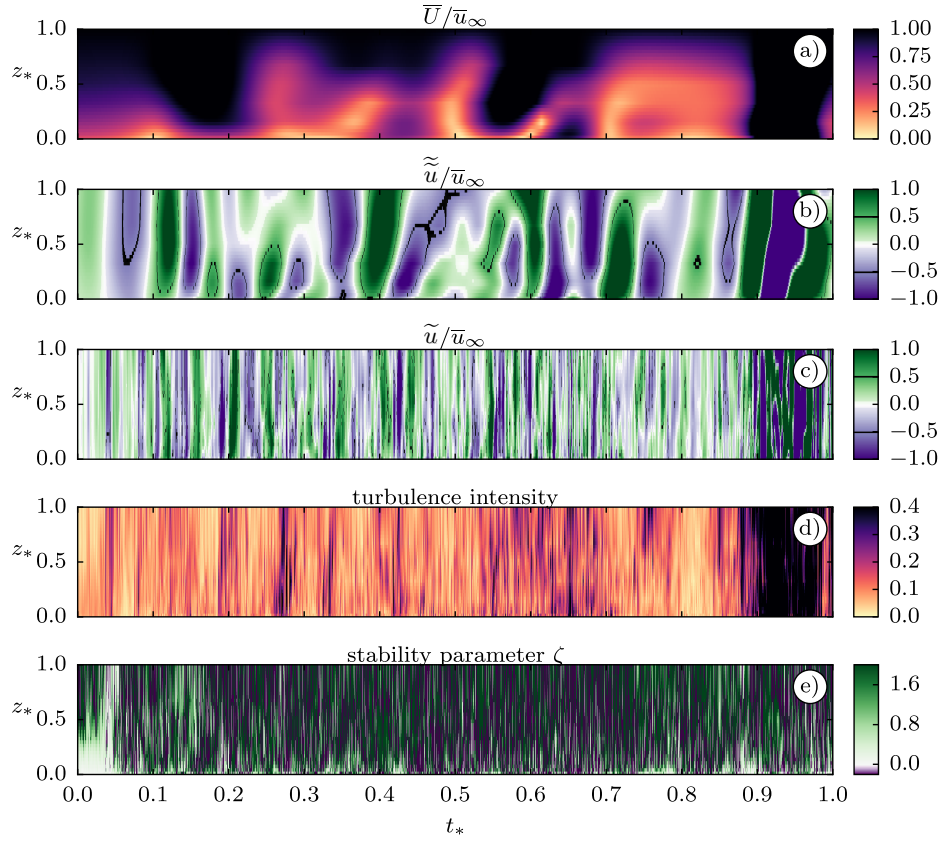


Figure 3.19: Scale activity and turbulence response for a night with almost no mean wind. The wind velocity averaged over this night is $\mathbb{E}(\bar{u}_\infty) = 1.3 \text{ m s}^{-1}$. The panels show: a) the mean wind profile, b) the jet scale, c) the sub-mesoscales, d) the TI, and e) ζ . For jet and sub-mesoscales, only the streamwise component is displayed to identify times of deceleration. Cross-stream component behaves similarly.

We studied three nights with a wide range of scale interactions. The vertical gradients of the jet and sub-mesoscale wind profiles become more pronounced as the mean wind decreases. For the following model construction, it is assumed that these gradients contribute to the generation of the TKE. Therefore, the nonstationary modeling is performed with this assumption next.

3.3.5 Model Formulation for Clustering

Recall, that the clustering method (see Sec. 2.2.2) aims to describes the nonstationary time series as a combination of $K > 1$ stationary sub-models that alternate in time. The difference between the individual models is in their parameter values. Each set of parameters of the K th model is constant for an a priori unknown period. A statistical learning algorithm determines this period and the model parameters. This section establishes a one-dimensional autoregressive factor model with exogenous variables to model the TKE.

More precisely, the goal is to test a hypothesis that consists in obtaining a GRANGER causal relationship (Granger, 1988) between the measured TKE (e_M) and the gradients of the prefiltered scales (exogenous variable): $\bar{F}_M = \partial\bar{u}/\partial z$; $\tilde{\tilde{F}}_M = \partial\tilde{\tilde{u}}/\partial z$; $\tilde{F}_M = \partial\tilde{u}/\partial z$ (for more details on the scale definition see Sec. 3.3.2). When the causal relationship changes in time, the clustering method identifies a regime switch. The variables are temporarily marked with a subscript (M) to indicate that they will be rescaled.

Rescaling of variables is considered to improve the model identification. The type of rescaling should be carefully selected depending on the research task. On the one hand, statistical rescaling is preferred when the clustering aspect of the methodology is of interest. On the other hand, physically meaningful scaling should be applied when the modeling aspect is important (the trained model should be used for out-of-sample prediction). However, nondimensionalization of an unknown system is not always possible in exploratory data analysis because little knowledge about the model structure is available. Since the focus is on clustering the statistical scaling is preferred (see Eq. (3.18) below). In the case of strong causality, the variance of the data explained by the model should be high and, therefore, should also confirm the choice of the model structure.

Removing the detectable nonlinearity a priori in the data benefits clustering. Since the statistical model will be linear, the algorithm does not need to learn the nonlinearity. Otherwise, the model would require more clusters. According to the previous study in Sec. 3.2, the TKE has a heavy tail distribution in the strongly stable condition. Consequently, taking the logarithm of the TKE: $e = \ln(e_M)$ and making the distribution more Gaussian efficiently reduces the number of clusters needed to describe the intermittent regime. Otherwise, clustering was found to separate cases based on mean σ_w values rather than dynamic interactions with the imposed exogenous variables.

In the following, the variables are standardized with:

$$Z = \frac{Z_M - \mathbb{E}[Z_M]}{\sqrt{\mathbb{V}[Z_M]}}, \quad (3.18)$$

where Z_M is $\ln(e_M)$, \bar{F}_M , $\tilde{\tilde{F}}_M$, \tilde{F}_M and Z in this section stands for e , \bar{F} , $\tilde{\tilde{F}}$, \tilde{F} . The clustering framework is generalized to multiple dimensions and can be used with all seven measurement heights of the FLOSS2 dataset. Unfortunately, this leads to a large number of parameters. Therefore, the one-dimensional model is estimated separately for each height to keep the parameter space small. The structure of the linear time-delayed model is:

$$e_t = \mu(t) + \sum_{n=1}^p a_n(t)e_{t-n} + \sum_{\bar{m}=0}^q \bar{b}_n(t)\bar{F}_{t-\bar{m}} + \sum_{\tilde{\tilde{m}}=0}^q \tilde{\tilde{b}}_m(t)\tilde{\tilde{F}}_{t-\tilde{\tilde{m}}} + \sum_{\tilde{m}=0}^q \tilde{b}_m(t)\tilde{F}_{t-\tilde{m}} + \varepsilon(t), \quad (3.19)$$

where p and q are memory lags of the statistical process, the index t denotes the discrete-time step, and $\Theta(t) = \{\mu(t), \mathbf{A}(t), \mathbf{B}(t), \varepsilon(t)\}$ denotes the parameter set. The first parameter of the set $\Theta(t)$ is an offset $\mu(t)$, which becomes a constant mean of a cluster under the assumptions of the clustering method. The vector $\mathbf{A}(t) = [a_1(t), \dots, a_p(t)]$ consists of the coefficients of the autoregression and $\mathbf{B}(t) = [\bar{b}_1(t), \dots, \bar{b}_p(t), \tilde{b}_1(t), \dots, \tilde{b}_p(t), \tilde{b}_1(t), \dots, \tilde{b}_p(t)]$ consist of the coefficients of the exogenous variables. The last parameter of the set $\Theta(t)$ is the residual error $\varepsilon(t)$ between TKE (e_t) and the linear model.

The model fitness function is defined by the least square residual norm:

$$\ell(e_t, \Theta) = \|e_t - \sum_{n=1}^p a_n(t)e_{t-n} - \sum_{\bar{m}=0}^q \bar{b}_n(t)\bar{F}_{t-\bar{m}} - \sum_{\tilde{m}=0}^q \tilde{b}_n(t)\tilde{F}_{t-\tilde{m}} - \sum_{\tilde{m}=0}^q \tilde{b}_n(t)\tilde{F}_{t-\tilde{m}}\|, \quad (3.20)$$

The average clustering functional takes the form:

$$L[\Theta, \Gamma(t)] = \sum_{k=1}^K \sum_{t=0}^T \gamma_k(t) \ell(e_t, \Theta_k) \rightarrow \min_{\Theta, \Gamma}, \quad (3.21)$$

The solution of such a functional is described in Sec. 2.1 and 2.2. Next, consider the model performance.

3.3.6 Clustering and Modeling Results

In this section, the TKE (or the corresponding non-dimensional variable TI see Eq. (3.16)) is modeled using the FEM-BV-VARX method, with the model structure given in Sec. 3.3.5. The clustering method is associated with several hyperparameters that are estimated a posteriori. The number of clusters is set to three to resolve weakly, moderately, and strongly stable regimes (see Tab. 3.3). Test runs with two, and four clusters were performed. Two clusters did not separate the right hand side branch of the TSI diagram (see Fig. 3.23), and four clusters resolved the intermediate regime into two. Moreover, more clusters added complexity to the data analysis, and no additional relevant insights were obtained. To select an optimal memory size, several minimizations studies were performed for a fixed value $K = 3$ over a range of $C_p = (4, 200)$. After several tests, the parameters $p > 3$ and $q > 10$ provide no significant improvement in model performance.

The optimal parameter C_p for the maximum number of regime transitions is selected based on the adjusted coefficient of determination R^2 (Lewis-Beck & Lewis-Beck, 2015): as:

$$R^2 := 1 - \frac{\text{unexplained variance}}{\text{total variance}} = 1 - \frac{\sum_{k=1}^K \sum_{i=1}^{T_k} \varepsilon_{k,i}^2}{\sum_{k=1}^K \sum_{i=1}^{T_k} (e_{k,i} - \mathbb{E}[e_{k,i}])^2}, \quad (3.22)$$

where $\mathbb{E}[\cdot]$ is the expected value and T_k is the number of samples in the cluster k .

The coefficient of determination R^2 represents the relative explained variance. This coefficient serves as a performance indicator for the model. If one plots its value above the persistence threshold, an optimal value C_{opt} (see Fig. 3.20) is indicated by a maximum. The left panel of this figure shows that the largest variance explained by the model is at the surface. Rescaling the graphs for each altitude shows that the optimal value of C_p is slightly dependent on altitude. The averaged value of C_p over all heights is 103. The parameter C_{opt}

3. Nonstationary Analysis of the Stable Boundary Layer

experiences scatter when the minimization of the functional is repeated. This is because the global optimum is approached from different, random initial conditions. The variability of C_{opt} is of order ten (\pm regime jumps for 102 nights), and therefore the locations of regime jumps in the affiliation function differ between solution marginal. The balance between the number of minimization trials and solution accuracy is considered acceptable for this study.

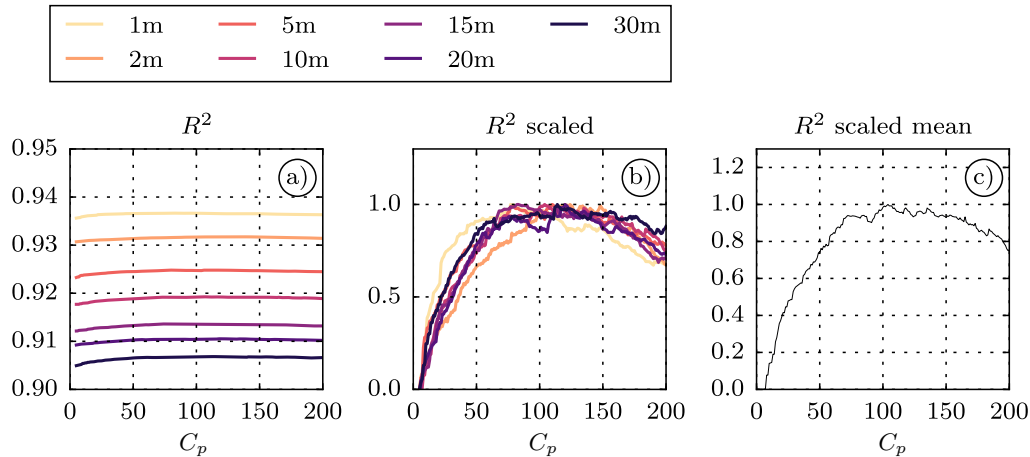


Figure 3.20: Coefficient of determination over regime persistence parameter C_p . a) R^2 at each height, b) R^2 normalized to emphasize the maximum, and c) the mean R^2 for all seven heights. The peak of a curve indicates the optimal parameter C_p .

The R^2 value is summarized in Tab. 3.2 for each height and regime. Overall, the variance explained by the model ranges from 0.8 to 0.93 and is considered satisfactory. The model performs consistently in the weak and intermediate clusters with respect to the modeled height. A steady decrease in R^2 with increasing height is observed (see Tab. 3.2 row: weakly and intermediate). This is because in the weakly stable regime the turbulence is well described by the mechanical shear generation at the ground. In contrast, the strongly stable regime has a lower value.

Table 3.2: R^2 values for each identified regime for each height.

Stability regime	1 m	2 m	5 m	10 m	15 m	20 m	30 m
Weak	0.93	0.90	0.90	0.88	0.89	0.85	0.80
Intermediate	0.92	0.93	0.92	0.91	0.88	0.82	0.80
Strong	0.88	0.81	0.81	0.84	0.86	0.88	0.85

Figure 3.21 shows the modeled and measured TI for the intermittent night shown in Fig. 3.18. The large deviation of the model is found at $z_* = 0.4$. The small-scale details at the top of the boundary layer are captured, but do not extend deep enough towards ground, mainly because the model does not account for spatial correlation. The temporal evolution of the TI is plotted for $z_* = 1.9$ (see Fig. 3.21c). The model reproduces the main trend, but with lower variance.

Figure 3.22 illustrates the statistical causality between the TKE and the vertical gradients of the jet and sub-mesoscales in terms of regression coefficients. The absolute values of the parameters cannot be compared between regimes because the averaged value of the TKE is

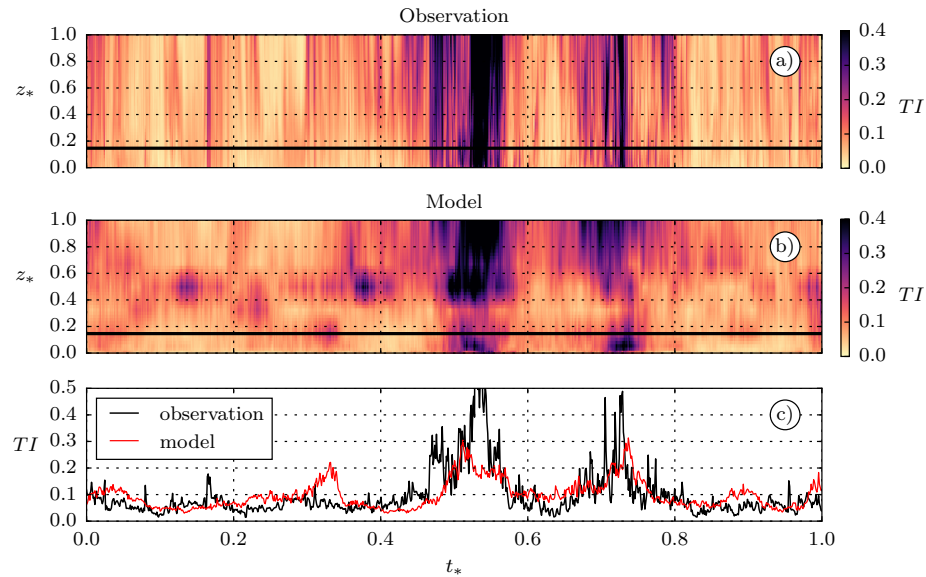


Figure 3.21: In-sample prediction of TI by the estimated model for the night shown in Fig. 3.18. Panel a) shows the measured TI and panel b) the predicted one. The solid black line in a) and b) marks a sample height of interest. Panel c) shows the time series of TI at that height. TI is linearly interpolated in height.

different in these regimes. However, a comparison between the coefficients of the exogenous variables is justified within a cluster at one altitude. In Fig. 3.22, the right column of panels shows the strongly stable regime (intermittent). At the height of one meter (see Fig. 3.22i), the memory effect (more time delays are important) is more significant than at 15 m (see Fig. 3.22f) and 30 m (see Fig. 3.22c). A comparison of the color-coded regions shows that the contribution of each scale to the TKE is approximately equal at ground level (see Fig. 3.22i). The influence of the gradients of the jet and sub-mesoscales on the TKE becomes smaller with increasing altitude (see Figs. 3.22f,c). The expected behavior is seen in the detected weakly stable regime (see Figs. 3.22a,d,g). Near the ground (see Fig. 3.22g), the gradient of the mean scale is the dominant factor. At greater distance from the ground (see Fig. 3.22a), the jet scale and the mean scale are dominant. The intermediate regime (see Figs. 3.22b,e,h) indicates a contribution of the jet scale at all three altitudes.

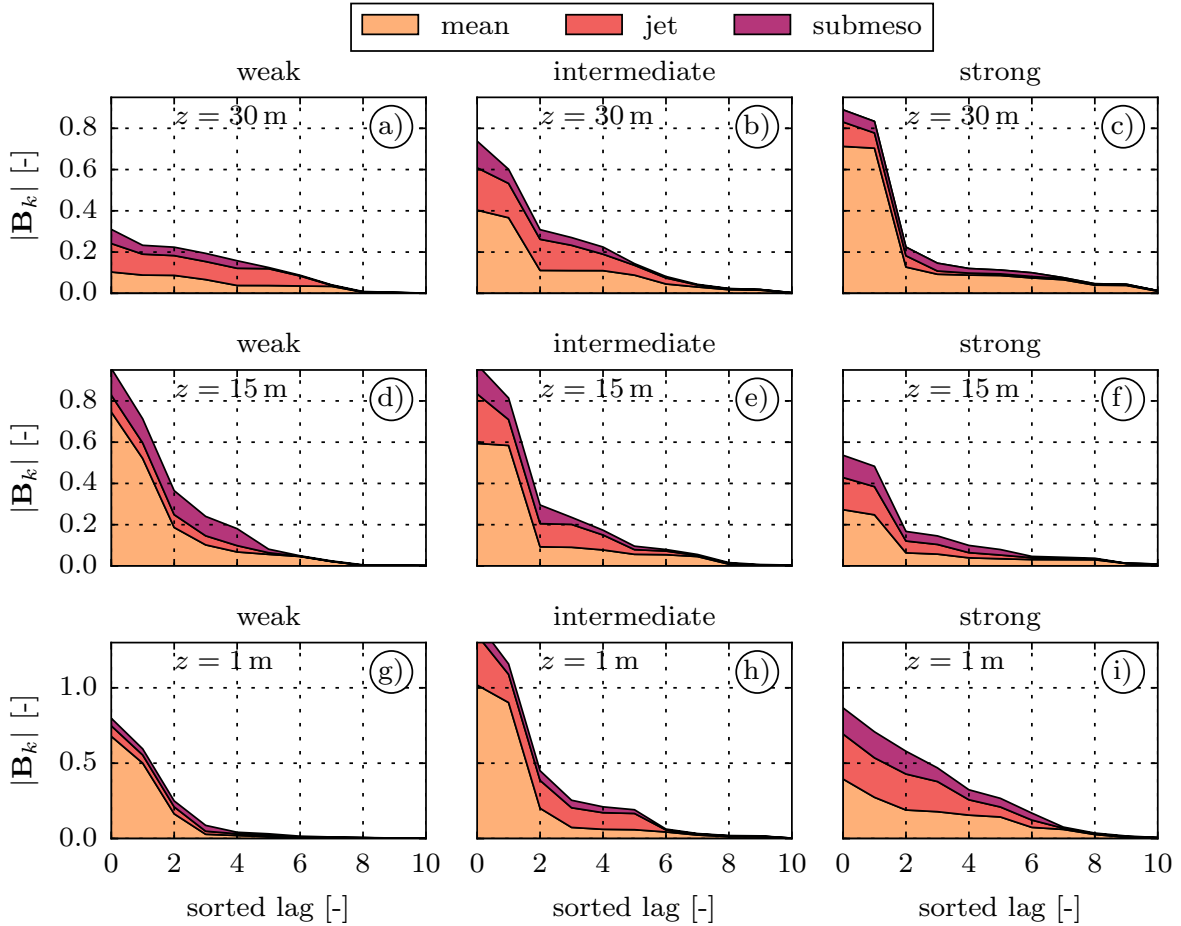


Figure 3.22: Stack plots of the parameter values for the forcing terms only, where $|\mathbf{B}_k|$ is the sorted absolute parameter vector of the forcing variables, where $k = 3$; $\mathbf{B}_1 = \text{sort}(|\bar{b}_1|, \dots, |\bar{b}_{11}|)$ and holds parameters for the lagged $\partial\bar{U}/\partial z$. $\mathbf{B}_2 = \text{sort}(|\tilde{b}_1|, \dots, |\tilde{b}_{11}|)$ and holds parameters for the lagged $\partial\tilde{U}/\partial z$. $\mathbf{B}_3 = \text{sort}(|\tilde{\tilde{b}}_1|, \dots, |\tilde{\tilde{b}}_{11}|)$ and holds parameters for the lagged $\partial\tilde{\tilde{U}}/\partial z$. From left to right, the columns are the: weak (a, d, g), intermediate (b, e, h), and strong (c, f, i) stability regimes. Each row is representing the heights: 30 m (a, b, c), 15 m (d, e, f) and 1 m (g, h, i). Each panel shows 11 parameter values for each forcing variable, with the absolute values of the parameters sorted in decreasing order. The occupied volume represents the explanatory power of each forcing variable relative to the other variables within each panel.

3.3.7 Regime-Dependent Turbulence-Submeso-Intensity Diagram

The Turbulence-Submeso-Intensity (TSI) diagram illustrates the relationship between the energy of the sub-mesoscales and the energy of the turbulent scales. It is expressed relative to the mean flow so that it has a common denominator (\bar{u}_∞). Looking only at the scatter plot, one can conclude that the value of TI at some region is correlated with SI (see Fig. 3.23a, where $SI > 10^{-1}$). This dependence is of particular concern since the slope is almost unity. One could not trust the simple scatter plot because of a possible self-correlation (Kim, 1999) for $SI > 10^{-1}$. On the other hand, there is a visible change at $SI \approx 10^{-1}$ from no correlation to almost 1 : 1, which contradicts the self-correlation problem.

The goal is to analyze the energy distribution between sub-mesoscale and turbulence scales in the phase space spanned by TI and SI (see Fig. 3.25). It is helpful to know whether some regions can be associated with specific stability values or the mean wind profile shape to understand the TSI diagram better.

The stability parameter and the shape of the wind profile are inherently connected. For example, damping of turbulence changes the shape of the mean wind profile from logarithmic towards linear. Such properties of the boundary layer can be quantified by the shape factor $\mathcal{H} = \delta_1/\delta_2$, where δ_1 [m] is the displacement thickness and δ_2 [m] is the momentum thickness (Schlichting & Gersten, 2017) and are defined as follows:

$$\delta_1 = \int_0^{30} \left(1 - \frac{\bar{U}}{\bar{u}_\infty}\right) dz, \quad \delta_2 = \int_0^{30} \frac{\bar{U}}{\bar{u}_\infty} \left(1 - \frac{\bar{U}}{\bar{u}_\infty}\right) dz. \quad (3.23)$$

The characteristic values are: $\mathcal{H} = 2.3$ for a laminar profile and $\mathcal{H} = 1.3 - 1.4$ for a turbulent profile. The value \mathcal{H} quantifies the boundary layer types in different regimes based on their shape.

To investigate the relationship between the values of TI and SI, it is reliable to use several factors such as the stability threshold ζ , the clustering method, and the shape factor \mathcal{H} of the boundary layer, each of which partitions the scatter plot in different ways based on the stability, the GRANGER causality between TKE and the gradients of the filtered wind speed (see Sec. 3.3.5), and the type of boundary layer profile. With all these indicators, the TSI diagram can be estimated more reliably.

Figure 3.23 shows the TSI diagram for three heights: 1, 15, and 30 m (column-wise from left to right), where each point is color-coded with the stability (see Figs. 3.23a, b, c) and shape factor (see Figs. 3.23d, e, f). First, it can be observed that the data points change their basic outer shape, indicating a height dependence. Second, the average turbulence level ($TI = 0.11$; see Fig. 3.23a, left corner of the point cloud) within the weak stability condition ($\zeta < 0.5$) for the height of 1 m is evident. The shape factor (see Fig. 3.23d) has low values in this region and is an additional confirmation that the boundary layer profile is turbulent.

Periods of strongly stable stratification ($\zeta > 1$) are observed more frequently (darker area) near the ground (see Fig. 3.23c) than at the surface (see Fig. 3.23a). With increasing altitude, the left corner of the point cloud (an indicator of average TI) decreases in value and merges with the center. Detection of the average TI is complicated by the increasing scatter at the left edge of the point cloud (see Fig. 3.23c). The shape factor indicates that the left corner of the cloud is mostly in a state of developed turbulence (see Fig. 3.23f).

3. Nonstationary Analysis of the Stable Boundary Layer

Table 3.3: Statistics of stability parameter ζ for each identified regime and height. Q1 denotes the 0.25 quantile, Q2 denotes the median and Q3 denotes the 0.75 quantile.

Height	Stability regime								
	Weak			Intermediate			Strong		
	Q1	Q2	Q3	Q1	Q2	Q3	Q1	Q2	Q3
1 m	0.01	0.02	0.07	0.02	0.11	0.48	0.02	0.12	0.53
2 m	0.02	0.04	0.09	0.05	0.16	0.52	0.04	0.20	0.78
5 m	0.03	0.07	0.17	0.10	0.23	0.58	0.01	0.41	1.42
10 m	0.05	0.14	0.34	0.18	0.57	1.43	-0.04	0.50	1.83
15 m	0.08	0.21	0.52	0.13	0.64	1.84	0.00	0.77	2.51
20 m	0.09	0.25	0.59	0.17	0.70	1.83	0.00	0.83	2.80
30 m	0.12	0.35	0.86	0.17	0.81	2.02	-0.22	1.05	3.85

An additional benefit of the shape factor is that the upper right corner of the point cloud can be partially distinguished as a more laminar profile in the TSI diagram (see Fig. 3.23d,e,f). Near the ground (see Fig. 3.23f), \mathcal{H} shows the region with high SI and TI as a more laminar profile ($\mathcal{H} \approx 2$) and low SI and TI as a turbulent profile ($\mathcal{H} \approx 1.65$). A contrasting separation of the TSI diagram is unobserved utilizing ζ (see Fig. 3.23c). A possible explanation for this is that the average value of ζ (see Fig. 3.24f) better fits the separation by the shape factor (see Fig. 3.23f). It is likely that one observes high values of ζ at low values of TI.

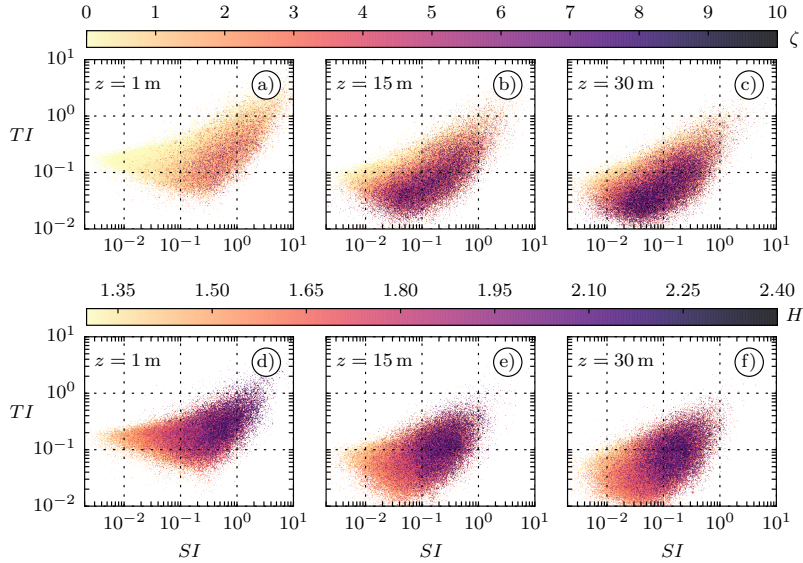


Figure 3.23: TSI diagram for three heights: 1 m, 15 m, 30 m. In (a,b,c) the points are coded with value of ζ and in (d,e,f) with the shape factor \mathcal{H} . The scatter plots contain a set of points where: $0 < \zeta < 10$ and $1.3 < \mathcal{H} < 2.3$. The points with a higher value of ζ and \mathcal{H} are plotted on top of the points with a lower value (they are sorted). $\mathcal{H} = 2.3$ is for laminar profile and $\mathcal{H} = 1.3 - 1.4$ for turbulent.

Figure 3.24 shows the classification of the TSI diagram according to the threshold ζ . The subcritical cluster $0 < \zeta < 1$ is shown in the top row and the supercritical cluster $1 < \zeta < 10$ is shown in the bottom row for three heights of interest. In addition, the subfigures show the PDF for the limited sets of points.

For subcritical stability, the densest region shows a decrease in TI with height, as expected (see Fig. 3.24a,b,c). A barely noticeable decrease in the sub-mesoscale energy across the boundary layer height is present (note the shift of the densest region to the left). At supercritical stability, the elliptical shape of the densest region changes its location and orientation in the TSI diagram (see Fig. 3.24d,e,f). In the supercritical region, the SI decreases stronger with height than in the subcritical region. At ground level, the high-density regions for the subcritical (see Fig. 3.24a) and supercritical (see Fig. 3.24d) regimes are separate, but near ground level they overlap (see Fig. 3.24c,f).

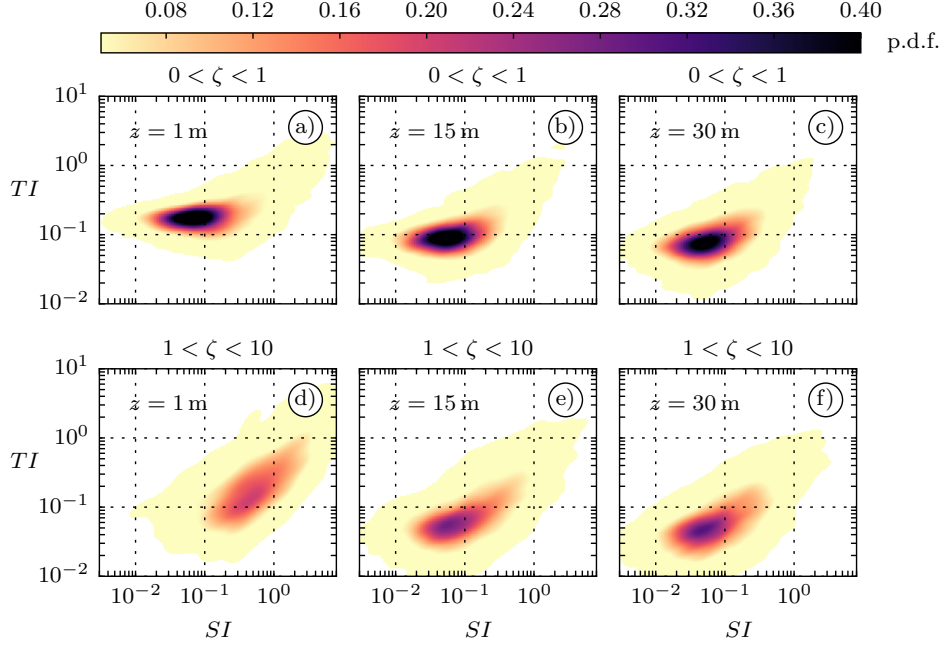


Figure 3.24: Classified TSI diagram by limiting with ζ value for three heights: 1 m, 15 m, 30 m. Subcritical stability $0 < \zeta < 1$ (a,b,c) and supercritical stability $1 < \zeta < 10$ (d,e,f). For each set of points, an PDF is estimated and plotted. Each column of subfigures represents a height.

Figure 3.25 shows the classification of the TSI diagram by the FEM-BV-VARX model (see Sec. 3.3.5) into weakly, intermediate, and strongly stable regimes for three heights: 1, 15, and 30 m, respectively. Both classifiers: ζ and FEM-BV-VARX show almost identical PDFs. in the TSI diagram for the weakly stable regime (compare 3.24a,b,c with 3.25g,d,a). The evaluated ζ value within the regimes supports this observation within the FEM-BV-VARX regimes (see Tab. 3.3). The strongly stable regime in Fig. 3.25i,f,c shows more stretched density functions. With increasing height, the shape changes from an angular feather-like structure to an ellipse. At 30 m height, there is a high probability of observing variability in TI values in the range $(1.1 \times 10^{-2}, 1.1 \times 10^{-1})$ (see Fig. 3.25c). This variability is smaller in the weakly stable regime (see Figs. 3.25a,d,g).

Using the TSI diagram, it is shown that FEM-BV-VARX characterizes the SBL differently compared to the classification with ζ (see Fig. 3.26). In Fig. 3.26, the high-density regions are estimated using the method of Hyndman (1996).

Hyndman defines the regions as follows. For a density function $f(\mathbf{x})$ of a random variable \mathbf{X} , the $100(1 - \alpha) \%$ region of high density is the subset $R(f_\alpha)$ of the sample space of \mathbf{X} such that $R(f_\alpha) = \{\mathbf{x} : f(\mathbf{x}) \geq f_\alpha\}$, where f_α is the largest constant, so that $P(\mathbf{X} \in R(f_\alpha)) \geq 1 - \alpha$.

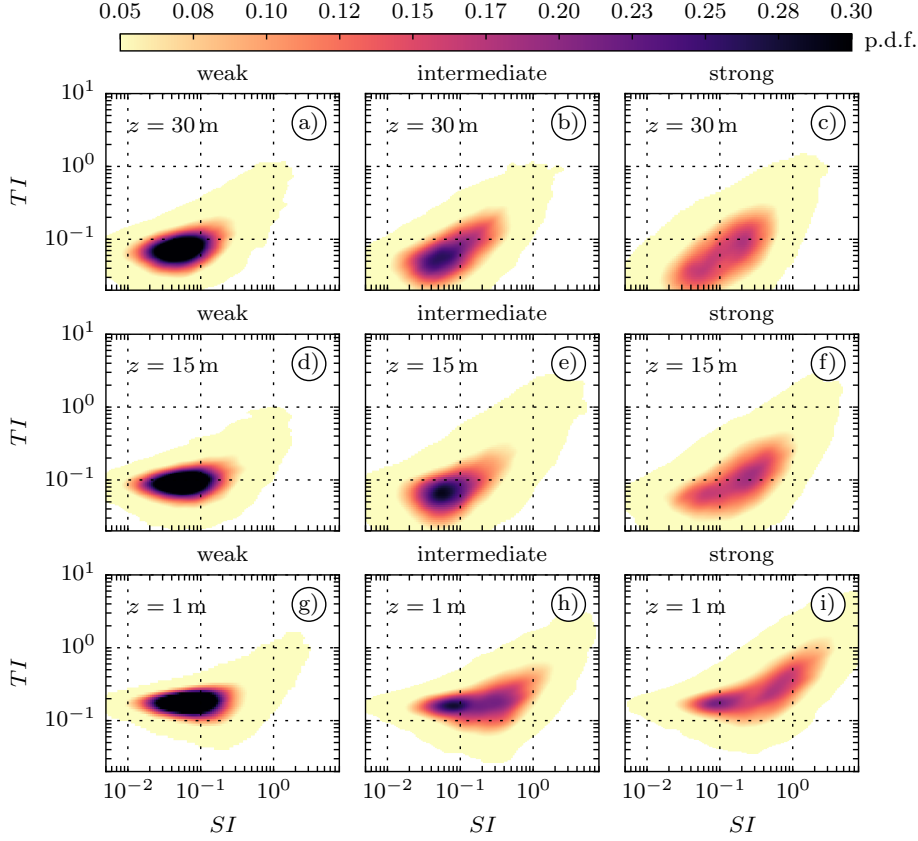


Figure 3.25: Classified TSI diagram with FEM-BV-VARX method. Weakly stable (a,d,g), medium (b,e,h) and strongly stable (c,f,i). The stability value of these regimes is given in Tab. 3.3. Three heights are presented: 30 m (a,b,c), 15 m (d,e,f), and 1 m (g,h,i). For each set of points, an PDF is estimated and plotted.

So f_α is the α quantile of $f(\mathbf{X})$. Selecting regions in this way allows us to see the bimodality of the distribution even in higher dimensional space.

Compare the strongly stable regime estimated with FEM-BV-VARX to the regime selected with $1 < \zeta < 10$ for a value $\alpha = 0.85$ (see Fig. 3.26). This α threshold means that all PDFs in the subfigures have a 15 % volume and are bounded from below by a plane.

The statistical clustering method identifies a larger region in the TSI diagram than the ζ classifier (see Fig. 3.26a, b, c with Fig. 3.26d, e, f). The intervening cluster (see Fig. 3.25h, e, b) is also part of the set $1 < \zeta < 10$. Nevertheless, clustering based on the relationship between turbulence and velocity gradients shows a more intense relationship between turbulence and sub-mesoscale motion in the TSI diagram. Interestingly, the local ζ value can indicate a range of TI at high stability (see Fig. 3.23c), but only highlights a reduced low value on average (see Fig. 3.26c). Surprisingly, the clustering method finds an additional state of TI (compare Fig. 3.26f with Fig. 3.26c), thereby forming a dipole. The metastable states indicate that the value jumps between a state with low TKE and low sub-mesoscale activity and a state with high TKE and high sub-mesoscale activity. Such a dipole structure is comparable to the intermittent nature of the SBL. At each altitude, the statistical classifier finds a more elongated relationship. The ζ classifier selects more concentrated patches and failing to detect the increased TI values at higher stabilities, suggesting that the temporal dependence of the

sub-mesoscale motions plays an essential role in describing the intermittent turbulence within the SBL.

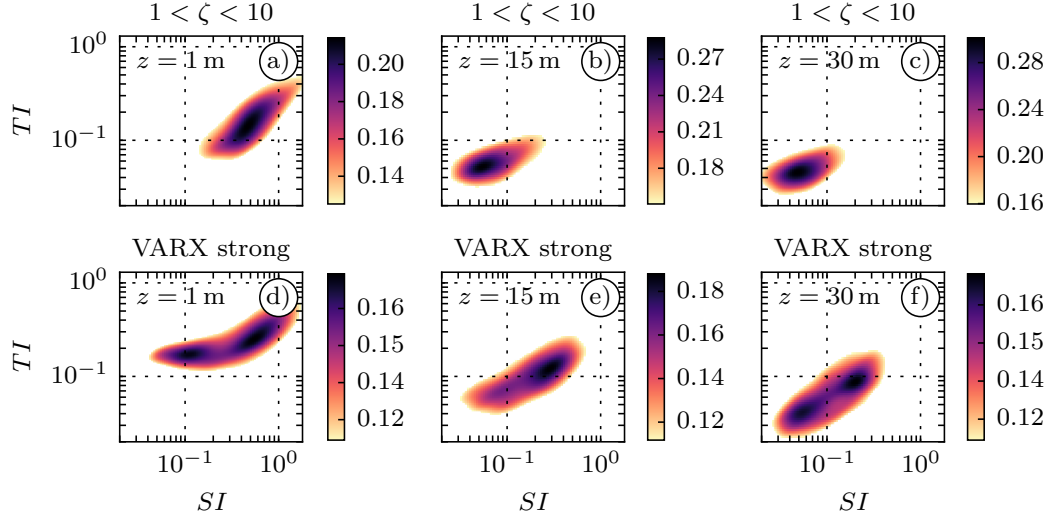


Figure 3.26: Comparison of the ζ -classifier (a,b,c) with the FEM-BV-VARX classifier (d,e,f) for the strongly stable regime only. The heights represent columnwise 1, 15, and 30 m. A PDF is estimated in each subgraph. The PDFs are limited to indicate the regions with the highest density. The threshold is chosen so that the projected region of maximum density corresponds to 15% of the total volume of the PDF.

The findings on the interaction between turbulence and sub-mesoscale motions are summarized in an abstract sketch from the analysis of the TSI diagrams (see Fig. 3.27) based on figures 3.26, 3.25, and 3.24. Ground shear is not the only source of turbulence, as other scales generate additional turbulence above the mean level (see Fig. 3.27a, right red ellipse). The situation changes with increasing altitude (see Fig. 3.27b). The mean wind speed is low; consequently, the gradient of the mean wind speed over the ground is close to zero, resulting in no mechanical generation of turbulence. In addition to the low wind speed, stratification suppresses turbulence. Therefore, the overlap of the intermittent (red) region with the non-intermittent region is less distinct (see Fig. 3.27b). As a result, the turbulence is driven more by the sub-mesoscale than by the gradient of the mean wind speed.

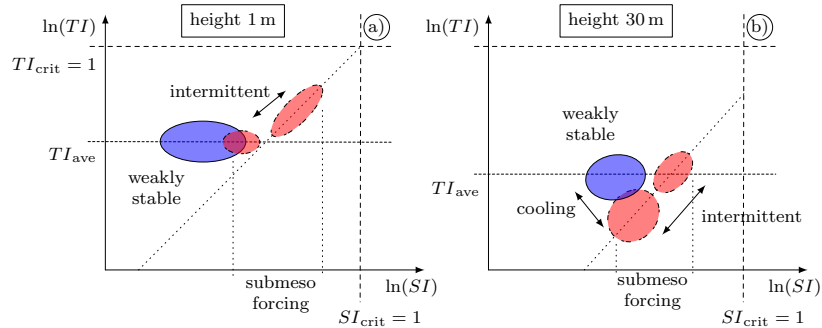


Figure 3.27: A sketch of a TSI diagram illustrating the intermittent turbulence states in the FLOSS2 dataset. The weak and strong stability regimes are marked with blue and red, respectively, TI_{ave} is the averaged value of TI; $TI_{crit} = 1$ if the energy of the mean wind speed is equivalent to the TKE and $SI_{crit} = 1$ if the energy of the mean wind speed is equal the energy of the sub-mesoscale motions.

3.3.8 Regime-Dependent Flux–Profile Relationship

The MOST is commonly used in single-column models (He et al., 2019; Rodrigo & Anderson, 2013), but fails as stratification increases due to greater dispersion in the data points. The estimated regime classifier $\mathbf{\Gamma}(t)$ is used (see Eq. (3.21)) to examine the flux–profile relationship for momentum in three different clusters. Table 3.3 shows the relationship between the detected clusters and the stability parameter ζ . To avoid self-correlation in the flux–profile relationship, a different normalization factor is chosen

$$\phi_m \phi_w^{-1} = \frac{\kappa z}{\sigma_w} \frac{\partial \bar{U}}{\partial z} \quad (3.24)$$

following Grachev et al. (2018), where the friction velocity is replaced by the standard deviation of the vertical velocity component σ_w . The quantity $\phi_m \phi_w^{-1}$ is expressed as a function of stability $\zeta = z/L$.

Figure 3.28 displays the flux–profile relationship for each cluster. Recall that the cluster stabilities have different origins than in MOST and are grouped based on the model affiliation function $\mathbf{\Gamma}(t)$ that is determined by solving the minimization problem (3.21). Equation 3.19 prescribes the rule that is used to identify a cluster. For each regime, a PDF is estimated, and the high-density region is constrained with $\alpha = 0.32$. The three flux–profile relationships show substantial overlap with increasing scatter. The strong stability regime (see Fig. 3.28c) reveals an outbreak towards the horizontal axis, which means more sustained turbulence. The maximum error (the distance between the black curve and the edge of the color coded PDF) relative to the MOST can be found in the intermittent cluster approximately at $\zeta \approx 2$. The error decreases for higher ζ . This observation is not sensitive to the value of the PDF threshold α . Besides calculating the flux–profile relationship with the gradient of the mean scale, the flux–profile relationship constructed with the gradients of the jet and sub-mesoscale, revealed substantial scatter regardless of the classified regimes and, therefore, are omitted.

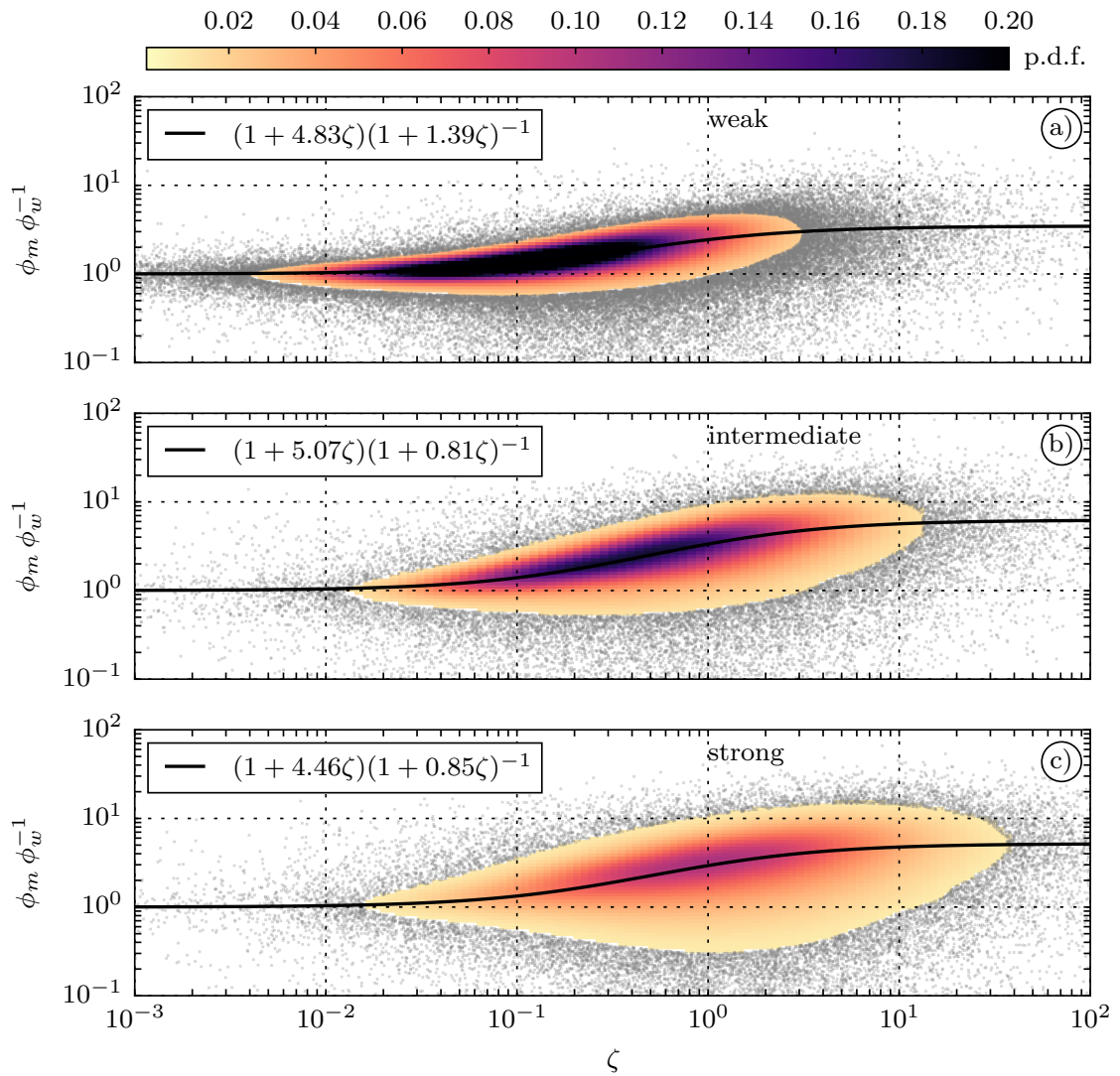


Figure 3.28: Classified flux–profile relationship with FEM-BV-VARX method. a) Weakly, b) intermediate, and c) strongly stable regimes. A PDF is estimated and limited to 68 % of the total volume. The gray points relate to the remaining 32% volume of the PDF. A modified stability function is fitted on the points within the limited probability density.

3.4 Discussion and Conclusion

In this chapter, a detailed study of the SBL is conducted by modeling the nonstationary time series using a model-based statistical clustering approach and analyzing the scale interactions within the detected clusters.

In the first part of the chapter (see Sec. 3.2), an adapted clustering strategy of Vercauteren & Klein (2015) is applied to analyze the FLOSS2 dataset. Difficulties were encountered in applying the FEM-BV-VARX method in terms of reproducibility and selection of the regime persistence parameter. This is attributed to the proposed model structure, which is slightly different from the one used by Vercauteren & Klein (2015). Namely, it only includes the interactions between the vertical turbulent fluctuation energy and the sub-mesoscale motions, thereby excluding the autoregressive terms. In addition, the FLOSS2 dataset was subjected to rigorous preprocessing, which resulted in a significant reduction in the number of strongly stable conditions. These two factors made the application of the FEM-BV-VARX challenging. However, an improved algorithm for determining the persistence parameter of the framework (see Sec. 3.2.3) and the recursive clustering procedure (see Sec. 3.2.4) enabled a reasonable classification. The resulting multiresolution analysis of the classified wind data (see Fig. 3.2.5) provided the necessary estimate of the stability-dependent time scale of the turbulent motions for the FLOSS2 dataset.

In the second part of the chapter (see Sec. 3.3), an improved model for clustering the TKE in the FLOSS2 dataset is presented. Robust classification results are obtained due to the logarithmic transformation of the TKE. Moreover, the variance of the modeled TKE is $\approx 80\%$. Such good predictive skill is attributed to the gradients of the sub-mean scales used to force the TKE.

Section 3.3.4 demonstrates scale decomposition utilizing the DWT. The activity of the sub-mean scales and their spatio-temporal interaction with the TKE and stability becomes nonstationary in the SBL. There is a time-averaging scale long enough for the considered maximum observation height of 30 m to obtain the logarithmic profile for the mean wind speed, corresponding to 3 hours. The variation in the mean wind profile that does not maintain the logarithmic shape is considered a deviation from the MOST. In the Sec. 3.3.2, the choice of sub-mean scales is subjective and is used to visualize the scales acting on the turbulence under stable conditions. The profiles studied show that intermittent turbulence is associated with decoupled boundary layers (Lan et al., 2018) and that top-down intermittency (Sun et al., 2012) is related to the activity of sub-mesoscale.

The improved statistical model 3.3.5 reproduces the value of TI (see Fig. 3.21) within the training data. While accounting for altitude correlation in the model would further improve performance, using this model to predict TKE under stable conditions is inappropriate for several reasons. The model requires the gradients of the sub-mean scales and the affiliation function, but these quantities are unknown in NWP models. In particular, the regime affiliation function needs to be parameterized. In addition, the number of parameters must be reduced (for the model 3.3.5 one has 42 parameters). In the following chapter, the sub-mean scales are modeled as a stochastic process, and the regime affiliation is parameterized with the Ri number.

Section 3.3.7 presents the TSI diagram showing how the sub-mesoscales are related to turbulence intensity using the ζ classifier and statistical clustering. The largest differences between these classifiers are found in the very stable region ($\zeta > 1$). The clustering method detects states in the TSI diagram with the interscale relationship because the clustering is based on GRANGER causality between the vertical gradients of the wind at multiple scales and the TKE. In contrast, the stability parameter ζ does not find these intermittent states.

The probability density function of the TSI diagram is analyzed for the regime of strong stability for seven available heights. The gradient-based classification yields a bimodal distribution (see Figs. d, e, f). Still, the bimodality is weak, and to separate the modes, one must select the high-density region containing 1% of the total PDF volume. The results show how the two modes change position as a function of altitude. In a strong SBL, a mode corresponding to high TI and SI values are found to be associated with a state transition to a higher TKE level due to sub-mesoscale motions. The TSI diagram shows the relative importance of sub-mesoscale motions in generating turbulence in terms of stability and height (see Fig. 3.25). Indeed, it is not known whether the footprint in the TSI diagram would look different at a different location. The footprint also cannot be thoroughly investigated with the stability parameter ζ because it does not take into account the important temporal dependence between sub-mesoscale of motions and the TKE, as shown in Fig. 3.26, leading to a significant underrepresentation of the SBL regimes.

This chapter presents an application of the FEM-BV-VARX method, where classification and modeling are performed simultaneously with discrete-time linear models. Using a different model structure compared to the previous approach of Vercauteren & Klein (2015), as shown in Sec. 3.2 and Sec. 3.3, improves both aspects: Clustering and prediction, leading to a better understanding of SBLs. The analysis of the classified flux-profile relationship in Sec. 3.3.8 suggests an idea to construct a stochastic RANS model. Since the main modeling variable in this chapter was the TKE, a better variable would be the nondimensional flux-correction variable ϕ . The following chapter explains the modeling procedure by utilizing specially for this aim developed FEM- H^1 -SDE method (see Sec. 2.2.1).

4

Stochastic Stability Equation

In the last chapter, a detailed analysis of the nocturnal boundary layer is performed to investigate the scale interactions using different approaches. The study is conducted within the distinct clusters identified by the FEM-BV-VARX method (see Sec. 2.2.2), with less attention paid to the predictive capabilities of the framework. Two main ideas for further research on modeling intermittent turbulence from the previous chapter are summarized below.

The first idea is to understand the nonstationarity of the boundary layer in terms of assumed hypothetical statistical functional relationships between variables at different scales. A detailed analysis of the statistics within the clusters proves the validity of the assumed correlations. The second important idea is that the proposed extension of the discrete model structure leads to an improved representation of the modeled TKE (see Fig. 3.21).

The previous modeling results naturally suggest using the model-based cluster framework for the construction of a stochastic turbulence closure. For this reason, the methodology developed in Sec. 2.2.1 is, in general, an essential step towards embedding data-driven non-stationary stochastic models into deterministic multi-scale dynamical systems. Several modeling opportunities arise when the framework is seen as an application-specific parameterization tool rather than a general data analysis method. The new FEM- H^1 -SDE framework is suitable for developing stochastic surrogate models of a continuous type and allows for time-varying parameterization of the noise term, which is required to generalize the effect of sub-mesoscales. In contrast, noise modeling is not addressed by the FEM-BV-VARX method.

The traditional scaling relationships used in the parameterization of the SBL (see Sec. 1.1.1) are limited by the assumptions of stationarity and homogeneity imposed on turbulence. In particular, the MOST assumes a persistent geostrophic wind that drives the boundary layer with significant kinetic energy. For such boundary layers, the MOST provides the necessary scaling of turbulent diffusion as a function of the static stability of the flow. In this work, the MOST concept is extended from a functional relation to a formulation that uses a SDE whose parameters also follow a scaling relationship. Correcting for turbulent diffusion using an SDE allows us to relax the stationarity assumption and introduce perturbations that represent the net effect of sub-mesoscales that are unresolved by a RANS closure. This concept enables us

4. Stochastic Stability Equation

to express the mean impact of static flow stability on the wind profile and include intermittent deviations and nonstationary mixing due to turbulence generated by sub-mean scales.

Finding a stochastic model for the TKE, as treated in Ch. 3, is suboptimal because the intermittency is indistinguishable from the mean turbulence level, which is already well predicted by RANS models in weakly stable regimes. As tested (but not reported) in this work, it is possible to estimate a valid nonstationary and nonlinear stochastic surrogate model for the evolution of the TKE using STUART-LANDAU type models. Unfortunately, such modeling ideas induce more research questions than answers. In particular, it is not clear how to define the reference level of TKE based on which the stochastic extension should contribute missing TKE from sub-mesoscale activity. This means that by using different RANS formulations with pre-tuned parameters will yield different reference levels of TKE and therefore result in other parameter estimates of the stochastic model part from the same data. Therefore, the data-driven identification of a stochastic extension would depend highly on the particular parameterization of the RANS model and thus would lack modularity. Another critical aspect is that due to the stochastic modeling of the TKE, the vertical correlation of the turbulence intermittency is unclear. This results in the need for further research to develop the FEM- H^1 -SDE method to classify spatial parameter modulations.

The above problems are avoided by selecting an appropriate variable for modeling. A promising direction is to find a nondimensional variable, which in this case is the stability correction variable. By stochastically modeling this variable, the turbulent diffusion coefficient becomes a stochastic process. Moreover, the KOLMOGOROV assumptions in modeling the dissipation term in a RANS closure are conserved because the stochastic modification of the turbulent mixing length influence the production and dissipation terms respectively without changing their proportion. Moreover, the stability correction variable is sufficiently isolated from the basic structure of the RANS closure and can be added in a highly modular fashion. This chapter shows how to motivate, construct, parameterize, and validate an SDE that extends the canonical stability function. The resulting SDE can be integrated into existing atmospheric RANS closures up to order 1.5. Surprisingly, the proposed SDE was found to be locally valid and thus independent of z up to a height of 30 m, suggesting a description of the vertical correlation structure of the stochastic perturbations. The proposed formulation captures stochastic spatial perturbations in height - practical suggestions for adjusting and tuning the stochastic model are provided as well.

Modeling of the stochastic stability equation is initiated by defining the averaging scales and variables. Then, the modeling steps to deduce the stochastic equation's functional form are given and supported with a corresponding, short data analysis. A sensitivity study is provided to ensure the right choice of the model structure by comparing the performance of a less or more complex model structure in a data-driven fashion. After establishing the model structure, its parameterization is performed, which provides the necessary functions to connect the stochastic model to a RANS closure. The first parametrization attempt is intentionally used on the complete dataset to check if the overall concept of stochastic parametrization is working and the relevant scaling functions are, in fact, present. The second attempt is a more optimized parametrization study on a reduced training dataset. The carefully selected validation data is used to test the model in predicting the stability correction variable. Investigation of the spatial

covariance is conducted to enable a valid spatial-temporal prediction (vertical). The classified variability of the spatial covariance is also parametrized based on stability and included in the formulation of the stochastic model. In the final part of the chapter, the validation is carried out by comparing the stochastic stability equation's prediction against the data in the previously selected validation dataset. An approach to tune the stochastic stability equation is shown. The final sections cover the analysis of the stochastic equation.

4.1 Definition of Variables and Temporal Scales

Typically, a single-column RANS turbulence model of order 1.5 (Stull, 1988) is formulated for the SBL by prescribing, among other things, the kinematic eddy-viscosity coefficient which models the momentum diffusion due to turbulent eddies:

$$K_m = \alpha l_m \sqrt{e}, \quad (4.1)$$

where α is a modeling constant, l_m is the turbulent mixing length for momentum, and e is the TKE. Similarly, the eddy heat conductivity coefficient K_h responsible for the heat diffusion due to turbulence should be specified to achieve a complete closure. Here, the stochastic modeling focuses only on the coefficient K_m by assuming a constant value for the turbulent PRANDTL number $Pr_t = K_m/K_h$.

Before modeling the variable l_m , it is convenient to examine several time scales, where the REYNOLDS-averaging operator is applied in the following. Figure 4.1 shows a sketch introducing the variables with their associated scale.

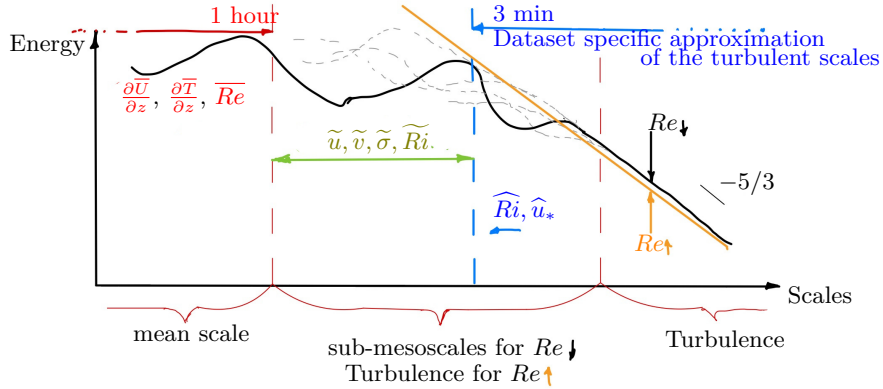


Figure 4.1: A sketch of the scales and associated variables defined for the stable conditions of the FLOSS2 dataset to identify the SSE. For strong wind conditions, the inertial subrange extends into the defined scale band 1 h - 3 min. For stable conditions and weak winds, this scale band is filled with sub-mesoscales. The turbulence averaging scale of 3 min is larger than the inertial subrange because it is estimated based on an entire dataset with averaged MRD and therefore the turbulence scale may differ locally.

For this study of the FLOSS2 dataset, the *long-time average* $\overline{(\cdot)}$ is set to 1 h. The long-time average should be considered as the time scale at which the MOST is valid. This means that the mean wind energy is sufficient to maintain a substantial amount of continuous KOLMOGOROV-like turbulence connecting the boundary layer to the ground throughout. More rigorous considerations would also require logarithmic scaling of the mean wind profile. However, to

4. Stochastic Stability Equation

avoid further complications in the modeling of jet dynamics (see Fig. 3.18), it is assumed that the 1-h mean is sufficient and that this scale is adequately represented with an appropriate RANS formulation.

The definition of the short-time mean scale is related to the averaged integral scale of turbulence in the stably stratified condition. Based on the previous multiresolution analysis (see Sec. 3.2.5), the *short-time mean* $\widehat{(\cdot)}$ is set to 3 min. The integral scale of turbulence scales with the Re number (Pope, 2000). In contrast, the uncertainty of the integral scale is assumed to increase with increasing Ri number because, in the strongly stable regime, the sub-mesoscales begin to dominate over the mean scale in the generation of turbulence (see Fig. 3.3.7). The 3 min period is dataset-specific and has an averaged value that suits nighttime. From the definition of the long- and short-time averages, it follows that a scale band corresponding to sub-mesoscale motions exists between 1 h and 3 min. The horizontal velocity components in this band are referred to as \tilde{u} and \tilde{v} .

Referencing the introduced scales map in Fig. 4.1, the variable l_m (Mahrt & Vickers, 2003) is defined as:

$$l_m := \frac{\widehat{u}_*}{\partial\overline{U}/\partial z} = \frac{\kappa z}{\phi_f}, \quad (4.2)$$

where $\widehat{u}_* = (\widehat{u'w'^2} + \widehat{v'w'^2})^{1/4}$ is the friction velocity estimated on the short-time scale, $\kappa = 0.4$ is the VON KÁRMÁN constant, z is the height coordinate, ϕ_f is the momentum stability correction function following the MOST, and $\overline{U} = \sqrt{\overline{u^2} + \overline{v^2}}$ is the mean horizontal wind speed, where the overline denotes the 1-h moving average. The function ϕ_f is determined from the flux-profile analysis and reduces turbulent mixing in a turbulence closure scheme of order 1.5. In the following only the mixing length for momentum is modeled stochastically. The stochastic process representing momentum stability correction is:

$$\phi(t) := \frac{\kappa z |\partial\overline{U}/\partial z|}{\widehat{u}_*}, \quad (4.3)$$

and will be modeled next. The interpretation of the definition is as follows. The scale inequality between the definition of \widehat{u}_* and $|\partial\overline{U}/\partial z|$ excludes in the long-time averaging operator $\overline{(\cdot)}$ the sub-mesoscale variations. Importantly, the scale inequality between the definition of \widehat{u}_* and $|\partial\overline{U}/\partial z|$ includes in the short-term averaging operator $\widehat{(\cdot)}$ the turbulent mixing generated by the sub-mesoscales. Therefore, the scale inequality introduces uncertainty in the ϕ variable related to the scale gap involving the sub-mesoscales. It is assumed that the sub-mesoscale motions induce transient changes in the mixture length, and to represent those, the uncertainty in ϕ is modeled with an SDE.

An essential assumption in the above formulation is that the sub-mesoscale is independent of the mean scale for any flow stability. This assumption does not hold in the weakly stable regime, especially when the geostrophic wind value is immense because the sub-mesoscale band exhibits an inertial subrange which is dependent on the mean wind (KOLMOGOROV turbulence). As the subsequent analysis of the model shows (see Fig. 5.1 and Fig. 5.10), violation of this assumption leads to an artificial variance at high winds in weakly stable conditions and shows that the stochastic term in the SDE must be penalized by the Re number.

4.2 Modeling Strategy

The objective is to establish a SDE for the temporal evolution of the variable ϕ which is denoted as the Stochastic Stability Equation (SSE). The parameters of this new stochastic equation depend on the Ri number as discovered in the study. The functional form of the equation and the functions that map the Ri number to the parameter values are specified in the modeling process. The former is derived from MOST, and the latter is subject to parameter estimation using the methodology developed in (see Sec. 2.2.1). In the following steps, the strategy is formulated for determining the functional form of SDE.

1. Substituting the function ϕ_f with an Ordinary Differential Equation (ODE) under the condition that the steady state of the ODE obeys the function ϕ_f .
2. Conversion of the ODE into a SSE by perturbing one of its parameters.
3. Fitting the proposed SSE with nonstationary parameters to the FLOSS2 data and identifying the scaling of the parameters with the Ri number.

To begin identifying the model structure, we consider the weakly stable regime and the strongly stable regime. Both characterize the SBL and highlight specific scale activity. In the weakly stable regime, the mean scale drives turbulent diffusion due to mechanical shear at the surface. In the strongly stable regime, the stochastic sub-mesoscale motions govern the intermittent turbulent mixing (see Sec. 3.3.4). The data used for modeling is pre-processed following Sec. 3.3.1.

4.2.1 Two Distinct Scale-Driven Regimes

It is known that turbulence is preserved under idealized conditions up to a certain critical value $Ri_c = 0.25$ (Galperin et al., 2007). Up to this value, the integral scale of turbulence may be associated with a large scale of turbulence production. However, in the presence of other non-turbulent scales (as it is the case in the SBL), the separation and quantification of the integral scale are ambiguous since, at high values of the Ri number, there exist multiple overlapping scales in physical space. One of the reasons for this is that the velocity gradients used in forming the Ri number may be associated with different processes (see Sec. 1.1.2). The energetic levels of these processes are close to each other, so they are challenging to differentiate. As a result the turbulence is sustained way beyond the Ri_c constructed with the mean scale (Mahrt, 2010b; Mahrt & Vickers, 2006).

Instead of considering some limiting Ri number as a transition value, the SBL is roughly divided into two limiting regions: the weakly stable region ($Ri \approx 0.05$) and the strongly stable region ($Ri \approx 1$) (the values given are indicative). The continuous unsteady dynamics of the SBL across Ri_c becomes visible after the parameterization of the stochastic stability equation (see Fig. 4.10). By the weakly stable regime, it is understood that the solution of the SSE is in agreement with the MOST. In particular, this condition is characterized by equilibrated turbulence when the mean wind energy is the dominant origin of energy injection. In terms of Ri number, the weakly stable regime is associated with the value $Ri < Ri_c$. In contrast, the strongly stable regime characterizes turbulence driven by the sub-mesoscale motions. This

4. Stochastic Stability Equation

regime is associated with a value of $Ri > Ri_c$. The transition between the two regimes is controlled by the stratification, the mean wind scale, and the sub-mesoscales.

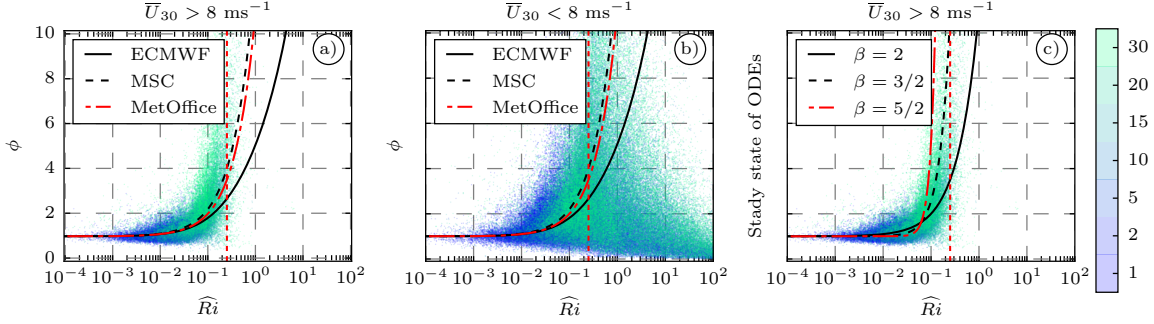


Figure 4.2: Comparison of different stability correction functions with the steady state of the ODE proposed in Sec. 4.2.2. The FLOSS2 dataset is partitioned according to the wind threshold 8 ms^{-1} . Panel a) shows the weakly stable regime in which the classical scaling ϕ_f (Cuxart et al., 2006) is present up to the Ri_c number marked by the vertical red dashed line. c) The nonlinearity parameter ($\beta = 2, 3/2, 5/2$) of the three ODEs 4.7 is set so that the steady state scales accordingly with \widehat{Ri} (3 min scale). The color encodes the height of the measurements. Panel b) shows the data part that does not fully follow the scaling and is modeled below.

We examine some existing stability functions that parameterize the SBL in some operational models. The most straightforward stability function $\phi_f(Ri)$ is linear since it provides a simple tuning procedure. The function is estimated to significantly suppress turbulent mixing near Ri_c . However, due to the sub-mean scales, some models apply a mild cut so that some mixing prevails beyond the critical value (see Fig. 4.2a). Following the comparative study of Cuxart et al. (2006), 4.2a and 4.2b show several functions: the ECMWF model, the model of the Met Office institution in the United Kingdom (MetOffice), and the Meteorological Service of Canada (MSC) model. Sandu et al. (2013) provides a historical overview of the functional forms considered in the ECMWF model. They all obey a similar structure with some modifications considering turbulence cut-off characteristics.

In Fig. 4.2, the data of the FLOSS2 experiment are separated exemplarily using the mean wind threshold 8 m s^{-1} to illustrate two regimes and to motivate the stochastic modeling. The two regimes are the weakly stable regime (see Fig. 4.2a), in which classical scaling theory applies, and the strongly stable regime (see Fig. 4.2b) with intermittent and unsteady turbulence. The functions shown in Figs. 4.2a, b have been plotted as they can be found in the literature and are not adapted to the data under consideration. Furthermore, no specific partitioning of the data into the individual regimes is performed. Good extraction of periods with constant mean winds reduces the scatter in the plots. Such pre-processing of nonstationarity is unnecessary in the proposed stochastic parameterization since the procedure already takes it into account.

Figure 4.2 shows the trend of the FLOSS2 datasets and that this trend is captured by the parameterization of the considered models for the weakly stable regime. However, at low mean wind speeds, the scaling deviates from the observations in the strongly stable condition (see Fig. 4.2b). To capture that variability (see Fig. 4.2b) and the transition into this regime, the intention is to model the nonstationary and nonlinear time series of the variable ϕ with a SDE.

4.2.2 Construction of the Model Structure

The goal is to create an initial model structure that will form the basis for further development. Thus, the modeling procedure begins by replacing the stability function with an ODE to open the possibility of capturing the temporal dynamics of the process ϕ . To my knowledge, there is no functional form of such an equation with respect to the Ri number. As mentioned earlier, the first step requires that the steady-state of the ODE reflects the scaling trend of the data in the weakly stable regime.

Starting with a simple candidate and adapting it to the design goal, consider:

$$\dot{\phi} = \widehat{Ri} \phi - \phi^2, \quad (4.4)$$

where \widehat{Ri} is the gradient RICHARDSON number at a scale of 3 min (the scale where the inertial range is approximated for the FLOSS2 dataset, see Sec. 3.2.5). This parameter changes the steady-state solution and ϕ is the stability correction variable defined in Eq. (4.3). The dot notation denotes the time derivative. Equilibrium occurs at $E_q(\phi, \widehat{Ri}) := \widehat{Ri} \phi - \phi^2 = 0$ and is equal to $(0, \widehat{Ri})$. Linear scaling is present, indicating that the structure is suitable.

Next, a rationale for considering nonlinear terms is given. The slow increase of the flow stability causes a linear increase of ϕ and thus a suppression of turbulence at the mean scale of the flow. At higher stabilities, the sub-mesoscales induce turbulent mixing faster than the mean scale because of their rapid impact on the boundary layer. To represent such dynamics, the reduction term should therefore be larger and is consequently quadratic. The quadratic term acts against the linear suppression term and generates turbulence. Depending on the stability and sub-mesoscale activity, the relationship between these terms must change. This is estimated later from data with the method presented in Sec. 2.2.1.

To ensure that the found states are attractive and not repelling, one can perform a bifurcation analysis. The stability of the equilibrium points is obtained by analyzing $dE_q/d\phi$ at these points. The bifurcation diagram of Eq. (4.4) is shown in Fig. 4.3, left panel. The equilibrium branches exchange stability at $\widehat{Ri} = 0$. The goal is to parameterize the SBL; the parameter range $\widehat{Ri} > 0$ is of interest since the equilibrium points start to scale linearly with \widehat{Ri} , as in the MOST.

Currently, only the model structure is investigated, but further analysis includes non-stationary parameter estimation. Therefore, one would like to reduce the model complexity by possibly exploring appropriate nonlinearity terms. To do this, one may consider that the scaling of the steady-state does not strictly follow a linear trend in the data. To change the scaling rate, one can modify the exponent in the second term:

$$\dot{\phi} = \widehat{Ri} \phi - \phi^\beta, \quad (4.5)$$

where two parameter values are of interest: $\beta = 3/2$ and $\beta = 5/4$. The equilibrium points are $(0, \widehat{Ri}^2)$ and $(0, \widehat{Ri}^4)$, respectively. The bifurcation diagrams are shown in the middle panel of Figure 4.3. The equilibrium branches exchange their stability at $\widehat{Ri} = 0$ and now the stable branch scales faster for $\widehat{Ri} > 0$ than in the case $\beta = 2$. These model structures should be considered as the data can support them in the further identification process. At this stage, it is unclear how nonstationarity affects the steady-state, and it is, therefore, useful to consider a

4. Stochastic Stability Equation

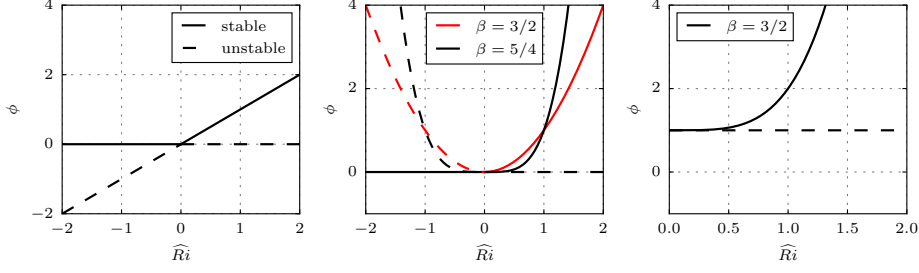


Figure 4.3: Bifurcation diagrams for the considered ODEs, showing the stationary solutions. The left panel shows the steady states of the Eq. (4.5) with $\beta = 2$. Decreasing the exponent to the value one increases the scaling rate of the steady state, as in the middle panel. The right panel shows the shifted steady state of Eq. (4.6) converging to the value 1 when $\widehat{Ri} \rightarrow 0$.

range of feasible model structures and later exclude those that are not supported by the data (see Sec. 4.3).

One further tailors the model structure by including properties of the RANS model characteristic to the neutral state. It is known that $l_m = \kappa z$ in the neutral condition and thus requires $\phi \rightarrow 1$ as $\widehat{Ri} \rightarrow 0$ from $+\infty$. For the differential equation, this means that the equilibrium branch must be shifted up by a value equal to one:

$$\dot{\phi} = 1 + \widehat{Ri} \phi - \phi^\beta. \quad (4.6)$$

To allow greater flexibility in fitting the ODE equilibrium to the data so that the critical Ri_c is adequately approached, one introduces a parameter before the second term.

$$\dot{\phi} = 1 + \widehat{Ri} \phi - b\phi^\beta, \quad b > 0. \quad (4.7)$$

This parameter is understood as a shift of the stable brunch along the x-axis in the log-linear plot (see Fig. 4.3). Moreover, the value of b must be positive for stable solutions of ODE. The equilibrium points for $\beta = 2$ are $(\widehat{Ri} \pm \sqrt{\widehat{Ri}^2 + 4b})/2b$. For $\beta = 3/2, 5/4$ there is only one real stable brunch, which for $\widehat{Ri} > 0$ can be approximated by $\widehat{Ri}^2/b^2 + 1$ and $\widehat{Ri}^4/b^4 + 1$, respectively. In the following, it is shown that the considered values of β cover a plausible range for ϕ in the regime where the classical stability correction of turbulence works.

The steady-state is estimated for the three nonlinear ODEs such that the solution branch follows the trend in the data (see Fig. 4.2c). The three hypothetically suitable equations are:

$$\dot{\phi} = 1 + \widehat{Ri} \phi - b \phi^\beta; \quad \beta = 2, \frac{3}{2}, \frac{5}{4}; \quad b = 0.1, 0.08, 0.07. \quad (4.8)$$

The stable branch of these equations does not follow an asymptote at Ri_c , but becomes asymptotically better in the approximation when $\beta \rightarrow 1$. These three equations form the basis for parameter estimation and provide a reasonable collection of models to test with the clustering method (see Sec. 4.3). The variations of the parameter b are introduced here to show that the ODE obeys the traditional scaling theory in the stationary, unperturbed limiting case.

In the next step, the effects of sub-mesoscale motions are included. It is known that their importance for turbulence generation becomes relevant at low wind speeds and large values of the Ri number (see Fig. 3.26).

4.2.3 Parameter Perturbation Approach

The stochastic parameter perturbation approach and the resulting structural change in Eq. (4.7) require a foundation. Therefore, the following hypothesis is investigated. The sub-mesoscale motions induce changes in the stratification and should be reflected in the local Ri gradient number. In particular, the perturbations of the velocity profile in the scale band bounded by the largest turbulent eddy and the mean scale should correlate with the fluctuation of the Ri number in the same scale band. In what follows, we examine whether the variance of the Ri number correlates with the sub-mesoscale motion's variance. This relationship provides us with a crudely simplified motivation of the parameter perturbation approach. The temporal averaging scales are discussed next.

The temporal evolution of the gradient RICHARDSON number at a scale of 3 min is denoted with \widehat{Ri} . It refers to the scale of the largest turbulent eddy (KOLMOGOROV-like). This value is subject to some variability because the integral scale estimate is based on the entire dataset average. It is, therefore, possible that the integral scale varies locally.

This uncertainty of \widehat{Ri} is measured as a variance by calculating the running variance at the scale of 1 h. This is a convenient operation since the variable \widehat{Ri} has no fluctuations shorter than the period of 3 min, and providing an approximation to the fluctuation intensity $\widetilde{Ri} := \overline{\widehat{Ri}\widehat{Ri}}$ of the Ri number related to the frequency band between 3 min and 1 h.

Next, we consider the sub-mesoscale motions, denoted by \tilde{u} and \tilde{v} . These velocity components are separated with a bandpass filter using the DWT (see Sec. 3.3.3), selecting the wind oscillations in the scale range between 3 min and 1 h. The non-dimensional variable rendering the variability of the sub-mesoscale motions is defined as:

$$\tilde{\sigma}(t) := \frac{1}{2\overline{U}^2}(\overline{\tilde{u}\tilde{u}} + \overline{\tilde{v}\tilde{v}}), \quad (4.9)$$

where \overline{U} is used to nondimensionalize the running 1 h variance of the sub-mesoscales. The factor 2 averages the two components.

The correlation between $\tilde{\sigma}$ and \widetilde{Ri} is not present in a scatter plot. The PEARSON's correlation coefficient (Rodgers & Nicewander, 1988; Scott, 2015) quantifies this judgment for each observation level (see Fig. 4.4, right panel). An important assumption is that the PEARSON coefficient applies to normally distributed variables and tests for linear correlation.

However, the SPEARMAN's correlation coefficient (Scott, 2015) analyzes the strength of the monotonic nonlinear scaling (see Fig. 4.4 right panel) between two random variables. Indeed, it is found that the considered variables $\tilde{\sigma}$ and \widetilde{Ri} are nonlinearly related. A value of ≈ 0.8 indicates a nonlinear relationship over the entire measurement height.

A linear relationship can be achieved if the random variables are logarithmically transformed. In this case, the variability of the natural logarithm of the \widehat{Ri} number $\widetilde{Ri}_{\ln} := \ln(\widehat{Ri})\ln(\widehat{Ri})$ and $\ln(\tilde{\sigma})$ is considered. A scatter plot of these variables for the height of 10 m is shown as an example in the left panel of Fig. 4.4. After logarithmic transformation, PEARSON's and

4. Stochastic Stability Equation

SPEARMAN's correlation coefficients show almost the same values for the total measured height (see Fig. 4.4 right panel). The interpretation of Fig. 4.4 left panel is improved by conditioning the points on the value of the mean wind. The results show that, especially for the strongly stable regime ($\bar{U} < 3\text{ms}^{-1}$), the value of the two variables $\tilde{\sigma}$ and \tilde{Ri} increases.

The above analysis reasons that in the spectral sub-mesoscale band, an increase in the variance of the Ri number can be observed with an increased relative variance of the sub-mesoscale motions. However, no conclusions about the causality can be drawn from the analysis. It could be bidirectional: strong stratification produces the sub-mesoscale motions or the opposite. However, the necessary conclusion is that the variability of the \tilde{Ri} parameter for the strongly stable regime (see Fig. 4.4 left panel) arises together with the variability of the sub-mesoscale motions $\tilde{\sigma}$. This observation supports the perturbation approach used to motivate the SSE below.

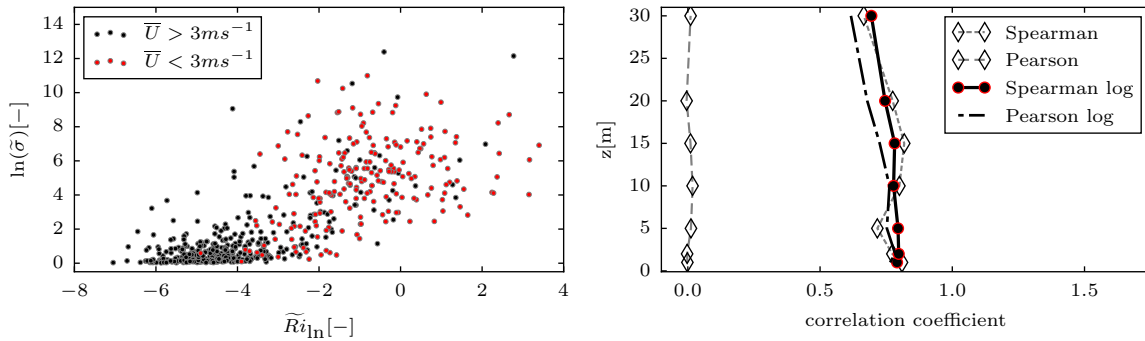


Figure 4.4: Association between the variability of the \tilde{Ri} number and the non-dimensional sub-mesoscale variability for the scales 3 m - 1 h (see Fig. 4.1). The left panel shows the scatter in a log-log plot for 10 m measurement by a wind value 3ms^{-1} . For mean wind values below 3ms^{-1} , the variability in both variables increases, suggesting the perturbation approach Eq. (4.10). The right panel compares the PEARSON and SPEARMAN correlation coefficients between the two variables for each measurement height, with the data being untransformed and log-transformed.

Consider a fixed value of parameters b and β in the proposed ODE (4.8) and focus on parameter \widehat{Ri} . The macroscopic time scale of evolution τ_{macro} for this equation is inversely proportional to the \widehat{Ri} value ($\tau_{\text{macro}} = \widehat{Ri}^{-1}$) (Horsthemke, 1984, Sec. 6.4). Let us assume that the boundary layer undergoes a typical radiative cooling process, as it usually occurs during the night. Accordingly, the mean value of the parameter \widehat{Ri} oscillates slowly, say with a scale of 1 h. With a continuously increasing absolute value of the \widehat{Ri} number, the time scale τ_{macro} of the dynamical system (4.8) becomes small.

Next, we assume that the sub-mesoscale motions act permanently and randomly in the boundary layer (see Fig. 4.4 left panel). With their time scale τ_{perturb} , these motions represent the perturbation of the dynamical system. The important question is how the dynamical system reacts when the two scales (τ_{macro} and τ_{perturb}) are close or farther apart?

At small values of \widehat{Ri} ($\tau_{\text{macro}} \gg \tau_{\text{perturb}}$), near the neutral state, the temporal fluctuation scale of the system is large. Meaning that the large scale of the boundary layer drives the turbulence; the wind profile is fully developed and is logarithmic. By consistently increasing the stratification, the macroscopic time scale of the evolution of Eq. (4.8) now decreases and becomes sensitive to the fast-scale perturbations ($\tau_{\text{macro}} \approx \tau_{\text{perturb}}$). The sporadic processes in the sub-mesoscale region can now affect the \widehat{Ri} number (see left panel in Fig. 4.4) and change

the evolution of the system (see Eq. (4.8)). Therefore, it is assumed that the perturbation of the system occurs through the \widehat{Ri} parameter. Additionally, decorrelation time scales of the following Eq. (4.11) as a function of the \widehat{Ri} number is examined in Sec. 4.6.3 Fig. 4.15 and supports the perturbation approach. Therefore, the parameter is perturbed as:

$$\widehat{Ri} = \overline{Ri} + \sigma_p \varepsilon_t, \quad (4.10)$$

where the value \widehat{Ri} is perturbed from its mean \overline{Ri} (the overline denotes the 1-h mean) with a i.i.d. random process ε_t that mirrors the effect of random turbulent mixing assumed to be related to the sporadic processes in the sub-mesoscale band (see Fig. 4.1). The intensity of perturbation is denoted with σ_p and is modeling the variable $\tilde{\sigma}$ (see Fig. 4.4).

The disturbance type in this study is limited to pure white noise (i.i.d.). However, it is more accurate to expect a process whose energy spectrum is non-flat (colored noise), band-limited, and has a particular regularity value. The decomposition approach Eq. (4.10) is the simplest possible and can be replaced by a more reasonable modeling hypothesis.

The selected parameter perturbation is included in the Eq. 4.8. Thereby, the process ε_t is defined to be the time derivative of the WIENER process $dW(t)/dt$. The SSE takes the following structure.

$$d\phi = \left(1 + \overline{Ri} \phi - b \phi^\beta\right) dt + \sigma_p \phi dW(t), \quad \phi(0) = \phi_0, \quad (4.11)$$

where ϕ is the stochastic correction of the turbulent mixing length for the momentum under the influence of stratification and the random external perturbations of intensity σ_p . The stochastic forcing term $\sigma_p \phi dW(t)$ is a multiplicative type obtained by inserting Eq. (4.10) into Eq. (4.8).

Equation (4.11) is different from its unperturbed form (4.8), and therefore it would be wrong to interpret the parameters in the same way. After parameter perturbation, the interpretation of how the \widehat{Ri} number changes the steady-state of the solution is altered by the noise. To investigate this, one can perform a stochastic bifurcation analysis (Stuart & Ord, 1994). Such study is of secondary importance because the parameters are estimated from data, and the steady-state properties can change significantly due to nonstationarity. It is the task of the model-based clustering method (see Sec. 2.2.1) to resolve the nonstationarity and identify the hidden relationships with the resolved scales. Accordingly, the model to be estimated is:

$$d\phi = \left[1 + a_1(t)\phi - a_2(t)\phi^\beta\right] dt + a_3(t)\phi dW(t), \quad \phi(0) = \phi_0, \quad (4.12)$$

where the total time-varying parameter vector $\Theta(t) = [a_1(t), a_2(t), a_3(t)]$ is subject to parameterization via data. The parameter β is constant; it is investigated in the following analysis. Thus, the time-varying parameters are assumed to be characterized by the \overline{Ri} number to construct a closed-form model for a RANS closure.

4.2.4 Variation of the Model Complexity

The parameter estimation procedure (see Sec. 2.2.1) is applied to various model structures to investigate if a simpler or more appropriate model than the one given in the previous

4. Stochastic Stability Equation

section (see Eq. (4.11) with $\beta = 2, 5/4, 3/2$) can be found. Additional models are constructed by changing the degree of nonlinearity and the complexity of the noise. Testing of different models is considered as another validation experiment that turns out to support the choice of Eq. (4.11) as it is determined next.

The following list summarizes the suggested model variants:

1. $d\phi = (a_1 - a_2\phi)dt + a_3dW(t)$ (simplification)
2. $d\phi = (a_1 - a_2\phi)dt + a_3\phi dW(t)$ (simplification)
3. $d\phi = (1 + a_1\phi - a_2\phi^{5/4})dt + a_3\phi dW(t)$ (base)
4. $d\phi = (1 + a_1\phi - a_2\phi^{3/2})dt + a_3\phi dW(t)$ (base)
5. $d\phi = (1 + a_1\phi - a_2\phi^2)dt + a_3\phi dW(t)$ (base)
6. $d\phi = (1 + a_1\phi - a_2\phi^3)dt + a_3\phi dW(t)$ (complication)

The model parameters are considered to be time-dependent. Even more model structures have been tested and are listed in the appendix A along with a brief description of their performance.

Consider the models labeled with index 3 - 5. The constant 1 is justified in Sec. 4.2.2. However, it was also tested whether the clustering method can identify this constant value 1 by the parameter estimation procedure. The results show that the value cannot be identified together with other parameters. A plausible reason could be an overparameterization of the model. This suggests investigating a model where the parameter of nonlinearity is constant. Namely, to estimate:

$$d\phi = \left[a_1(t) + a_2(t)\phi - \phi^\beta \right] dt + a_3(t)\phi dW(t), \quad \phi(0) = \phi_0, \quad (4.13)$$

Under the condition that the parameter $a_1(t)$ converges to 1 and $a_3(t)$ is close to zero in the neutral state ($\bar{R}i = 0$). This model structure could be tested in further studies. In the next section, the selected models are estimated and compared in terms of their likelihood.

4.3 Data-Driven Rejection of Model Alternatives

Although the models induced by the perturbation approach seem reasonable, it is common practice to review different models. For this purpose, a collection of models to work with was presented in Sec. 4.2.4. The models proposed in Sec. 4.2.2 are estimated using the clustering procedure to identify an appropriate model structure and possible scaling of the nonstationary parameters. In this section, the model performance is examined in terms of the likelihood value:

$$J = \sum_{k=1}^K \sum_{i=0}^N \gamma_k(t_i) \ln \left[p_\phi(\Delta t, \phi(t_{i+1}) | \phi(t_i); \bar{\theta}_k^*) \right], \quad (4.14)$$

where p_ϕ is the transition probability density function of a particular model (for its computation see Sec. 2.1.2), ϕ the analyzed stability correction process, $\gamma_k(t_i)$ and $\bar{\theta}_k^*$ are the identified regime affiliation and regime model parameters with the FEM- H^1 -SDE method (see Sec. 2.2.1).

Note, that the cost is unpenalized, because the considered models have an equal number of parameters. Typically, the information criteria used in model selection have a penalty to compensate for the models' complexity (Burnham et al., 2002). The number of clusters K changes the complexity of the models as well. Therefore, the models' cost is compared over a reasonable fixed range of hyperparameter K .

Multiple definitions of the BAYESIAN information criteria for the variational framework (Gerber & Horenko, 2014; Horenko, 2010b; Metzner et al., 2012) were tested on controlled numerical examples (see Sec. 2.6). They were found to be inconsistent in estimating the optimal number of clusters for different models structures (probably due to the nonlinearity of the models). The parameter ϵ_{opt}^2 of the clustering framework is estimated using the procedure introduced in Sec. 2.4.1. Then, all fluctuations of the function γ_k faster than 1 h are removed with a discrete wavelet filtering approach using HAAR wavelet bases. This ensures that the regimes do not switch faster than 1 h, which is the averaging scale of the RANS model.

This section uses data initially preprocessed according to Sec. 3.3.1. The vertically averaged value of ϕ is examined to identify the optimal model structure:

$$\langle \phi \rangle = \frac{1}{z_{30} - z_1} \int_{z_1}^{z_{30}} \phi(t, z) dz, \quad (4.15)$$

where $z_{30} = 30$ m is the height of the highest measurement and $z_1 = 1$ m is the height of the lowest. Once the model structure is established in the following analysis, the clustering is repeated for different heights.

Figure 4.5 shows the change in model likelihood according to Eq. (4.14) over the number of clusters used up to 10. This is more than sufficient to understand which model structure is the most appropriate. The y-axis shows the scaled model cost according to Eq. (4.14). The scaling between 0 and 1 is subject to the absolute minimum and maximum values and is done here for a visual purpose. The lower the value, the better the model approximates the data.

To estimate an appropriate value for the number of clusters, the cost tradeoff is analyzed and combined with the information gain measure introduced in Sec. 2.4.2. Since the cost and information gain are not normalized, the tradeoff and the resulting number of clusters must be well-considered.

In Figure 4.6, the cost (red) and the information gain (blue) are plotted on the dual axis so that they overlap graphically. This allows us to assess whether the two curves are approaching an asymptotic value, and to find a range for the number of clusters where the improvement of the fit in terms of information gain is justified. The green areas in the figure 4.6 mark a plausible range for the choice of the hyperparameter K for each of the models considered.

From Figure 4.6, it can be seen that the proposed range for the number of clusters K is between 4 and 7 and that the range is insensitive to the model structure. The results in Fig. 4.5 show that within the considered range, the best model is type 5, which has multiplicative noise and quadratic nonlinearity. The worst model is labeled with index 0 and has linear structure and additive noise, where the absolute fitness value is so insufficient that it is not even considered in the figure due to the significant cost values (compare the absolute cost value in Fig. 4.6). Moreover, according to Fig. 4.5, the inclusion of cubic nonlinearity (model type 6) leads to an increase in the cost value. Reducing to linear drift while retaining multiplicative noise (model type 1) also does not improve performance.

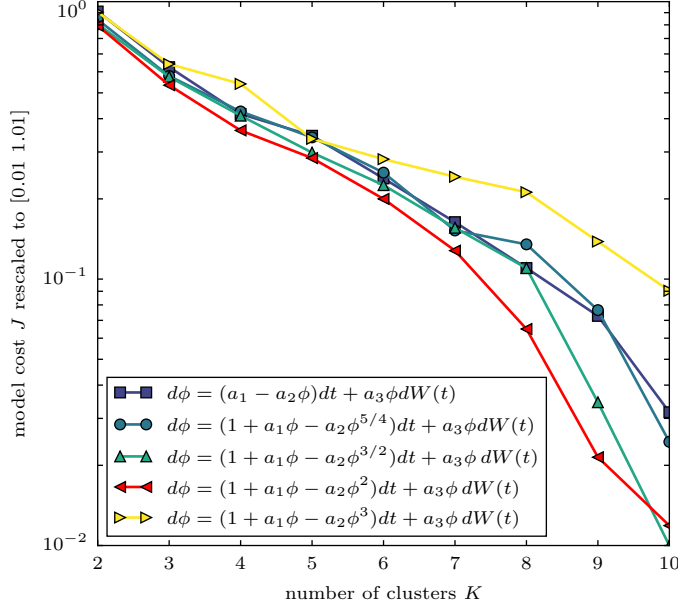


Figure 4.5: Selection of the best fitting model from a considered pool of models based on the model fitness function (see Eq. (4.14)). The likelihood value is plotted against the number of clusters used to resolve the time-varying parameters. The y-axis is rescaled for better illustration. The model structure with quadratic nonlinearity performs best over almost the entire range of clusters used. The simplest model structure: $d\phi = (a_1 - a_2\phi)dt + a_3dW(t)$ is not shown because its fitness value is significantly higher than the value of the displayed models (see Fig. 4.6).

The formed sensitivity study is a plausibility check for the derivation of the model structure Eq. (4.13). The data-driven testing of different model structures confirms that the quadratic nonlinearity and multiplicative noise (Eq. (4.12)) is the correct candidate for the simplest possible model structure within the considered dataset.

Finally, it should be noted that an attempt was made to simplify the quadratic model so that the parameter $a_1(t)$ scales linearly with \overline{Ri} . In addition, when identifying the parameter scaling, an attempt was made to express one of the parameters by another to obtain a model with effectively two parameters. None of the above approaches lead to a reasonable stochastic closure.

4.4 Identification of Local Parameter Scaling

The SSE is defined as follows and one strives to obtain a closed-form as a function of Ri :

$$d\phi = \left[1 + \lambda(t)\phi - v(t)\phi^2\right] dt + \sigma_p(t)\phi dW(t), \quad \phi(0) = \phi_0, \quad (4.16)$$

where $\lambda(t)$ is the suppression rate of the turbulent mixing length $l_m \sim 1/\phi$ due to processes above the mean scale, $v(t)$ is the production rate of the turbulent mixing, which becomes relevant at larger values of the \overline{Ri} number, and $\sigma_p(t)$ is assumingly the perturbation intensity of the turbulent mixing as a result of random sub-grid processes not resolved by the RANS model (e.g., sub-mesoscales). The unknown time-dependent parameters $\lambda(t)$, $v(t)$, $\sigma_p(t)$ are assumed to relate to the \overline{Ri} number formed from the local gradients of wind and temperature at the mean scale (marked with an overline and equal to 1 h in this study). Therefore, the

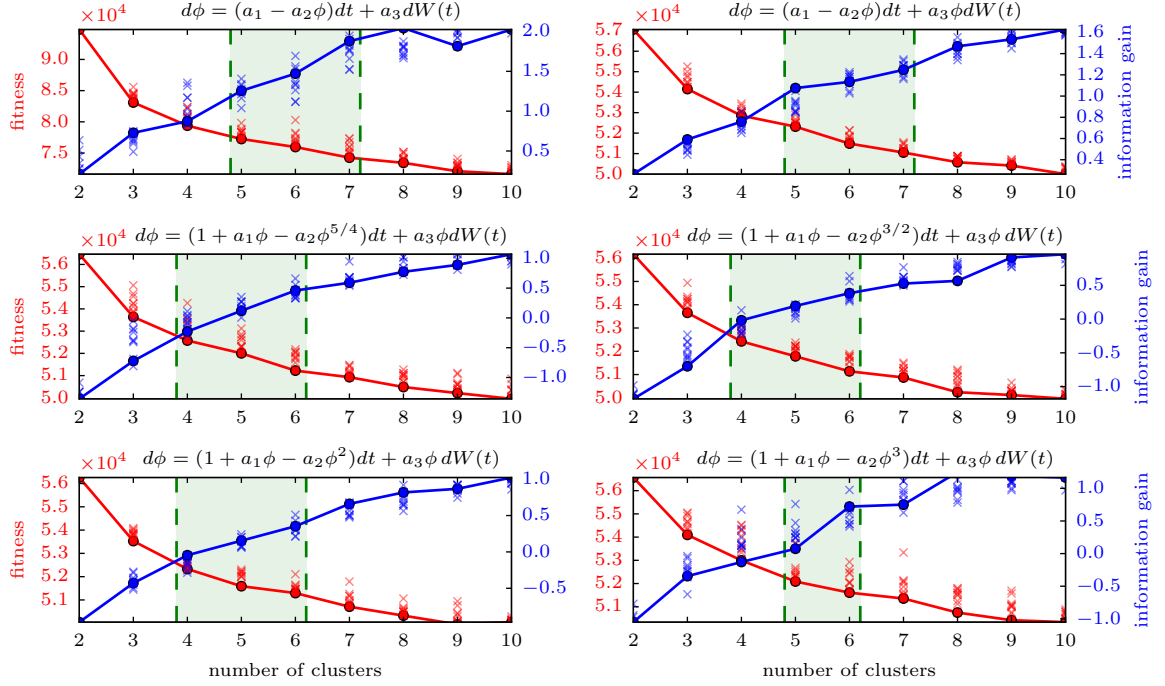


Figure 4.6: Estimated optimal number of clusters for each of the considered model types. As a trade-off between the optimal (low) fitness value and the information gain. The green color indicates a proposed trade-off region. Each box represents one model type. The information gain is estimated with Eq. (2.62), and logarithmically transformed.

objective is to determine the following scaling functions:

$$\lambda(t) = \Lambda(\overline{Ri}), \quad (4.17)$$

$$v(t) = \mathcal{V}(\overline{Ri}), \quad (4.18)$$

$$\sigma_p(t) = \Sigma(\overline{Ri}). \quad (4.19)$$

The assumption that such function exists is sensible to the choice of the model structure. For that reason, close attention is given in Sec. 4.3 to determine a pool of suitable model structures. Furthermore, in general, the number of free parameters (here 3) demands to be small. Otherwise, it becomes unlikely to identify all scaling functions due to over-parameterization.

Furthermore, the parameters $\lambda(t)$, $v(t)$, and $\sigma_p(t)$ are assumed to be independent of height, and as the results show, this is indeed the case (see Fig. 4.8). Since the variable ϕ is estimated independently at each height, and the stochastic model is independent of z by construction, information about the vertical correlation of the process $W(t)$ is missing and assessed later.

The vertical correlation is studied a posteriori. Physically, the sub-mesoscale motions are non-uniformly distributed throughout the boundary layer and depend on height. In the parameter identification, the sub-mesoscale motions are assumed to occur vertically with equal probability in the observed boundary layer. In other words, they are modeled by a stochastic process that is homogeneous in height. This is a strict assumption that must be made because of the one-dimensional form of the parameterization method (see Sec. 2.2.1).

The study aims to investigate the scaling of the parameters of the model and is performed twice. In the first experiment, the entire dataset is used for training. Since the results proved

promising, the second estimation is performed on a reduced data set with higher estimation accuracy and cross-validated with an excluded period. The very stable period excluded from the training is carefully selected to form a validation dataset. Therefore, both experiments differ in the amount of training data, estimation accuracy, and parameter parametrization.

In the first experiment, where the complete dataset is modeled, piecewise scaling functions are proposed. Such a definition provides fine control over the shape of the stationary distribution that the model predicts within the specific stability regions. The second parameterization is practiced based on a reduced dataset. The scaling functions are designed to parameterize the entire stability range continuously. It is worthwhile to consider both studies to investigate the sensitivity of the parameterization.

After applying the model-based clustering method, the scaling of the model parameters with \overline{Ri} is examined using the estimated regime classifier (see Fig. 4.8). The cluster-averaged value of the \overline{Ri} number is calculated using the following equation for each height and cluster individually:

$$Ri_k = \frac{\sum_{i=0}^N \log_{10}(\overline{Ri}(t_i)) \gamma_k^*(t_i)}{\sum_{i=0}^N \gamma_k^*(t_i)}, \quad k = 1, \dots, K, \quad (4.20)$$

where γ_k^* is the estimated model affiliation function from the vector $\mathbf{\Gamma}^*$ obtained with the FEM- H^1 -SDE method (see Sec. 2.2.1). The above equation can be recognized as a weighted arithmetic mean, where the weights are the model affiliation function due to the imposed constraints (2.21) and (2.22).

The scaling of the model parameters was also investigated with the mean local wind and temperature gradients in different combinations. This investigation did not result in an appropriate parametrization and hence is unreported.

4.4.1 Piecewise Parameterization Based on Entire Dataset

This identification study is performed by resolving the nonstationarity with 4 clusters for each height by determining the regularization parameter of the cluster method a posteriori using the procedure presented in (see Sec. 2.4.1). The 200 evaluations of the clustering method ensure the adequate exploration of the cost function landscape in searching for the global optimum. The regularization parameter varies between minimization attempts. The optimal solutions of the classifier are used as initial estimates for optimization with the next closest regularization parameter in some realistic range to estimate a reasonable parameter.

The results are summarized below, with Fig. 4.7 showing the scatter plots of the estimated parameters over the Ri_k number along with the proposed scaling functions that quantify the scaling relationships.

The parameter of the linear term is parameterized with (red line in Fig. 4.7a):

$$\Lambda(\overline{Ri}) = \lambda_1 \tanh[\lambda_2 \log_{10}(\overline{Ri})] + \lambda_3, \quad (4.21)$$

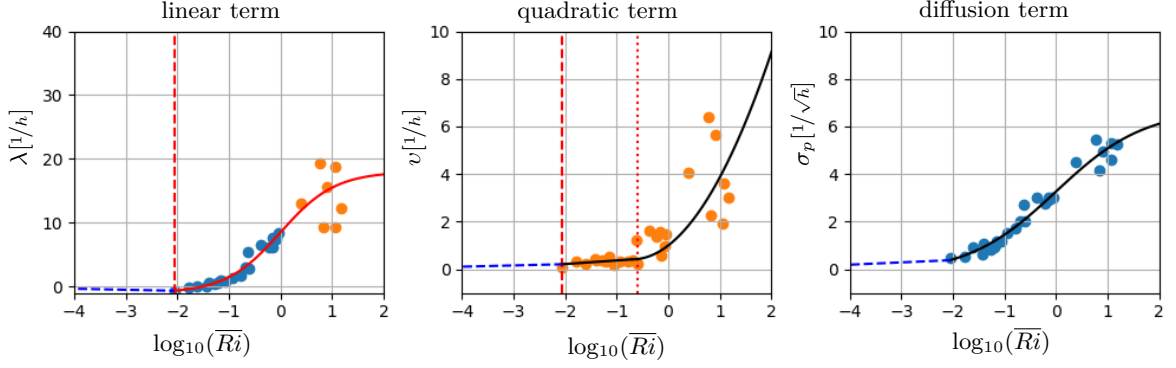


Figure 4.7: Local scaling of the time-dependent model parameters $\lambda(t)$, $v(t)$, $\sigma_p(t)$ in Eq. (4.16) and their respective piecewise parametrization functions identified utilizing entire stably stratified FLOSS2 dataset for training. The estimation of the parameter values is performed from a time series with sampling rate units cycles/hour. Resulting units of the linear and quadratic terms are in hours. The units of the diffusion parameter is balancing the units of $dW(t)/dt$ [$1/\sqrt{h}$] to result in [$1/h$]. The scatter points represent the seven clustered time series from the respective heights. In the left panel the orange points are removed from the calibration procedure of the scaling function. The blue dashed line is a linear function extrapolating towards the limiting value such that $\phi = 1$.

with $\lambda = (9.5266, 0.9401, 8.4955)$. The parameter of the nonlinear term is parameterized with (see Fig. 4.7b):

$$\mathcal{V}(\overline{Ri}) = \begin{cases} v_1 \log_{10}(\overline{Ri}) + v_2 & \text{if } \overline{Ri} < 0.0088, \\ v_3 \log_{10}(\overline{Ri}) + v_4 & \text{if } \overline{Ri} \in [0.0088, 0.2], \\ v_5 [\log_{10}(\overline{Ri}) + v_6]^2 + v_7 & \text{if } \overline{Ri} > 0.2, \end{cases} \quad (4.22)$$

where $\mathbf{v} = (0.0532, 0.3193, 0.1505, 0.5192, 1.1930, 0.6990, 0.4140)$. The diffusion parameter is parameterized by (see Fig. 4.7c):

$$\Sigma(\overline{Ri}) = \begin{cases} \sigma_1 \log_{10}(\overline{Ri}) + \sigma_2 & \text{if } \overline{Ri} < 0.0088, \\ \sigma_3 \tanh[\sigma_4 \log_{10}(\overline{Ri})] + \sigma_5 & \text{if } \overline{Ri} > 0.0088, \end{cases} \quad (4.23)$$

where $\boldsymbol{\sigma} = (0.0959, 0.5757, 3.4704, 0.5795, 3.2621)$.

This parameterization approach describes several regions with different functions. For all three parameter values of SDE, the values are bounded within $\overline{Ri} \in [0, 01, 10]$. The dispersion of the parameter values takes for $\overline{Ri} > 1$. Extrapolation beyond $\overline{Ri} = 10$ should be done with caution. In contrast, extrapolation to the limit $\overline{Ri} = 0$ is reliable and is based on the following physical consideration.

The expected value of the SSE should converge to 1 when $\overline{Ri} \rightarrow 0$, which represents the neutral boundary layer. The most sensible parameters for this condition are those of the drift term. The limiting values of the parameters at $\overline{Ri} = 0.0001$ are determined here by trial and error after tuning some fitted values. An extended analytical investigation can improve the parameter estimation values by incorporating the limiting states as a condition in an estimation procedure. Such an investigation is omitted because the following parameterization approach (see Sec. 4.4.2) finds the limiting value resulting from a regression analysis without tuning.

4. Stochastic Stability Equation

All three panels of the Fig. 4.7 lack point estimates for the limits of the \overline{Ri} number range. The scaling function is set so that the regime-dependent stationary PDF of Eq. (4.16) (see Fig. 4.10) obeys the Gaussian distribution in the neutral condition. The range $\overline{Ri} \in [0.0088, 0.2]$ in the parameter v is described with an individual linear function. For the range $\overline{Ri} > 0.2$ the parameter v follows a quadratic scaling.

The proposed piecewise parameterization provides a physically reasonable separation of the x-axis at the Ri_c number. The piecewise description of the pre- and post-critical regions controls the quality of the resulting stationary probability density of the SDE (see Fig. 4.10). However, the drawback is the increased complexity. It is challenging to obtain the correct smooth transient PDF and the correct limit state at $\overline{Ri} = 0$. Moreover, the total number of parameters in this approach is 17 and is considered too large. The following identification study aims to reduce this number. A scaling function for an SDE parameter and the appropriate limit state at $\overline{Ri} = 0$ are identified next without any post-tuning.

4.4.2 Continuous Parametrization Based on Reduced Dataset

Four clusters were used to resolve the nonstationarity in the previous section to test the overall stochastic parameterization procedure. Now, the clustering is repeated in more detail on a reduced dataset to obtain the best possible estimates. Table 4.1 summarizes the estimated hyperparameters for this study. The estimate of the affiliation function (not shown) is insensitive to the choice of the ϵ^2 parameter. For each height, the shape of the affiliation function shows marginal differences concerning the slight change in ϵ^2 . The regime occupation time was investigated to see if it matches across the boundary layer height and found to depend on the height z . This is mainly because shallow boundary layers can be intermittent at the top and neutral at the surface.

The optimal number of clusters is selected in this study based on the idea of the best tradeoff between the functional cost value and the information gain. A summary figure can be found in the appendix A.1. According to this figure, the resolution of nonstationarity with 5 clusters is sufficient. Subsequent analysis suggests that this number is rather too high, as the affiliation function shows more suppressed transition regimes than clustering with $K = 3, 4$ (not shown). Nevertheless, the following results are based on $K = 5$.

Table 4.1: The estimated hyperparameters ($K_{\text{opt}}, \epsilon_{\text{opt}}^2$) for each height of the model structure Eq. (4.16).

	1 m	2 m	5 m	10 m	15 m	20 m	30 m
K_{opt}	6	5	5	5	5	5	5
ϵ_{opt}^2	15	10	21.5	14.6	68.1	46.4	21.5

The parameterization of Eq. (4.16) is shown in Fig. 4.8. The results show a sufficiently reliable scaling where the entire parameter range is described by a continuous function (see Tab. 4.2). It is plausible to logarithmically transform both axes in the case of the nonlinear terms (see Fig. 4.8b) and the diffusion parameter (see Fig. 4.8c), since these parameter values must always be positive. The parameter values of the nonlinear term seem to have a larger scatter. A sensitivity study has shown that a slight change in the slope of the parameterized function leads to only a marginal change in the curves for the moments of the PDF (see

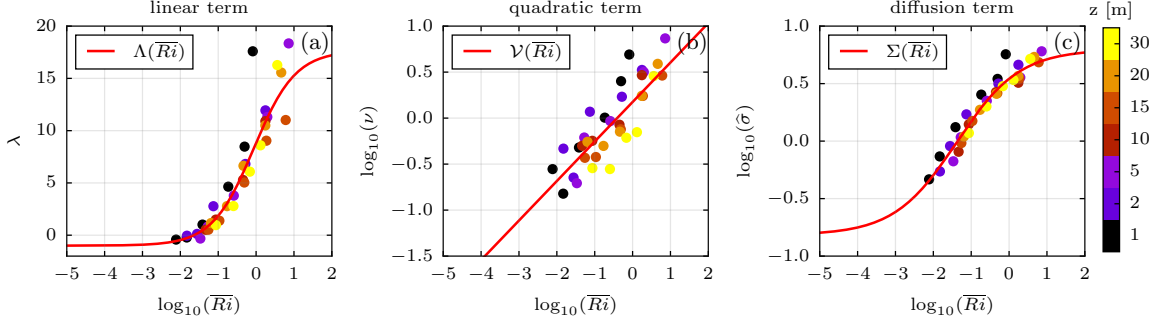


Figure 4.8: Local scaling of the time-dependent model parameters $\lambda(t), \nu(t), \sigma_p(t)$ in Eq. (4.16) and their respective continuous parametrization functions. The color bar is valid for each of the panels and labels the seven heights. The estimation of the parameter values is performed from a time series with sampling rate units cycles/hour. Resulting units of the linear and quadratic terms are $[1/h]$. The units of the diffusion parameter $[1/\sqrt{h}]$ is balancing the units of $dW(t)/dt [1/\sqrt{h}]$ to result in $[1/h]$.

Fig. 4.11). However, the greatest sensitivity is observed in the parameterization of the linear term (see Fig. 4.8a). Changes in this parameter have a large impact on the scaling properties of the expected value of the PDF.

Studying the function in Fig. 4.8a one may wonder why the tangens hyperbolicus was chosen, considering that the function for $\overline{Ri} > 1$ does not obey the trend of the data. Several variants of different polynomials were also tested to find a more straightforward function. It turns out that they cannot accurately capture the scaling in the essential range $\overline{Ri} \in [0, 01, 1]$. The inappropriate scaling problems in the presence of extreme turbulence suppression are illustrated by the PDF in Fig. 4.34), especially for the considered range of \overline{Ri} numbers (this effect is not shown graphically). The tail of the scaling function for $\overline{Ri} > 1$ is an inappropriate suggestion for extrapolation. Furthermore, it is important to note that the scatter increases for $\overline{Ri} > 1.0$. Up to this value, there is a relatively moderate, persistent uncertainty in the parameters, including the transition across the Ri_c . The lack of increased uncertainty in the parameter values at Ri_c indicates a robust representation of the statistical properties of the variable ϕ .

Table 4.2: Continuous scaling functions for the SSE parameters as a function of the local gradient \overline{Ri} number. The parameter values (indexed by 1, 2, 3, 4) of the descriptive function are estimated with regression analysis. The parameter σ_s controls the noise intensity independently of the stratification and is therefore useful for tuning. Value $\sigma_s = 0$ corresponds to the FLOSS2 dataset.

Param.	Descriptive function	1	2	3	4
$\lambda(t, z)$	$\Lambda(\overline{Ri}) = \lambda_1 \tanh[\lambda_2 \log_{10}(\overline{Ri}) - \lambda_3] + \lambda_4$	9.3212	0.9088	0.0738	8.3220
$\nu(t, z)$	$\mathcal{V}(\overline{Ri}) = 10^{(v_1 \log_{10}(\overline{Ri}) + v_2)}$	0.4294	0.1749	–	–
$\sigma_p(t, z)$	$\Sigma(\overline{Ri}) = 10^{(\sigma_1 \tanh[\sigma_2 \log_{10}(\overline{Ri}) - \sigma_3] + \sigma_s)}$	0.8069	0.6044	0.8368	$\sigma_s = 0$

It is also worth mentioning that, unlike the previous parameterization approach (see Sec. 4.4.1), the parameter values of the scaling function (see Sec. 4.2) are determined by least-squares estimation. As mentioned earlier, the important target is to obtain a PDF with an expected value equal to one when the number Ri approaches zero. The regression analysis

naturally yields the correct values, so this condition is satisfied (see Fig. 4.8) without any additional modification of the parameter estimates.

The diffusion term parameter (see Fig. 4.8c) cannot be set to zero in the neutral state and must approach a small value. Otherwise, the stationary probability distribution function is considered non-computable. This does not mean that the strong solution of SDE does not exist. The fact that the stationary distribution is not computable for some finite, small values of σ_p can be attributed to the numerical computation. The most important observation related to the diffusion parameter is the ability to control the scaling of the noise and adjust it as needed. Since the intensity of the noise reflects the activity of the sub-mesoscale motions, the ability to prescribe it is beneficial when incorporating the SSE into a turbulence model.

It is found that the expected value of the stationary PDF (see Fig. 4.8) does not change significantly when the function $\Sigma(\overline{Ri})$ is replaced by a suitable linear function in the log-log representation. A log-log linear scaling is less favorable for the FLOSS2 dataset than the proposed one but still valid. The $\Sigma(\overline{Ri})$ function can be adjusted along the y-axis with the parameter σ_s to enhance or suppress the occurrence of the sub-mesoscale motions regardless the stratification.

The identification study presented achieves the goal of finding a simpler parameterization of the SSE. Compared to the approach in Sec. 4.4.1, a reduced dataset is used. The splitting of the dataset into training and validation cannot be generalized by a percentage rule (Xu, 2018). The main reason for this is that the set of highly intermittent regimes (which are of interest) is present at the beginning of the dataset. Since the research interest is in modeling these regimes, one must choose the validation dataset wisely. For the identification study here, $\approx 5\%$ of the dataset is used as validation. The data selected consists mainly of intermittent regimens, as the goal is to validate these dynamics. This removes the most important data from the training. A rough estimate is that $\approx 20\%$ of the highly intermittent regimes is removed from the training data. It could already be seen that there are some differences in the parameterization of the scaling functions (see Fig. 4.7 and Fig. 4.8). In the following section on validation, the differences are shown more clearly. However, they are relatively acceptable and should give a sense of how sensible the parameterization is.

4.5 Vertical Covariance Structure of the Modeled Variable

The previous section shows that the parameter values of the SSE scale with the local value of the \overline{Ri} number. These results suggest that the proposed equation is suitable for prediction within the first 30 m of the SBL. The spatial correlation of ϕ is expected to be a consequence for spatially coherent sub-mesoscale motions affecting the local mixing length over an area.

It is assumed that the scaling of the parameters occurs as a logical consequence of the nondimensionalization with the height variable z in Eq. 4.3 occurs. This nondimensionalization is known for the neutral boundary layer from the logarithmic law. The more surprising results imply that this nondimensionalization also holds for highly stratified boundary layers with intermittent mixing. The important detail assisting to explain the surprising scaling properties is the study of the vertical correlation of the process ϕ .

This section examines the significance of the vertical covariance as a function of different \overline{Ri} numbers. The results of this section, as well as the validation of the spatio-temporal simulation of the SSE (see Sec. 4.6.2), suggest that the scaling of the parameters of the stochastic models is likely due to the minimum possible correlation length of the variable ϕ , which is, on average, slightly smaller than the measurement height during highly stratified regimes. The vertical covariance of the noise term would be characteristic of the measurement area. It should therefore be accounted for by incorporating the SDE model into the existing RANS model (see Ch. 5). In particular, attention in combining RANS and the SSE is mandatory to avoid excessive mixing due to sub-mesoscale motions in the upper boundary layer ($z > 100$ m)

Understanding the vertical covariance of the ϕ process is essential in facing the task of turbulence modeling. To incorporate the prediction of ϕ via the SSE into a turbulence model, one must specify the vertical covariance in the increments of the WIENER process. There are two primary purposes for the previous. The first has to do with the numerical coupling between the solution of the RANS model and the SDE. The turbulence model has derivatives in height and consequently imposes those constraints on the solution of SDE. Accordingly, the solution of SDE must have at least one derivative in height so the RANS model will stay stable. For further details, the reader is referred to Ch. 5. The second reason is purely physical. Since the sub-mesoscale motions are structures in space, the turbulent mixing generated should also be spatially correlated to some degree.

The FLOSS2 dataset has seven measurement points, and this set of points is not sufficient to obtain a valid estimate for the Autocorrelation Function (ACF) by inspecting the lagged distance of the variables ϕ . Instead, spatial correlation is examined using a semivariogram (Journal & Huijbregts, 1978; Matheron, 1963; Olea, 1999) given by an unbiased estimator:

$$\gamma(\Delta z) = \frac{1}{2N} \sum_{i=1}^N (\phi(z_i + \Delta z) - \phi(z_i))^2, \quad (4.24)$$

where N is the length of the time series and $\Delta z = [1, 4, 9, 14, 19, 29]$ m. The semivariogram is widely used in geostatistics and helps to deduce the autocorrelation structure for the process ϕ .

For the analysis of the nonstationary ACF, the spatially integrated variable $\langle \phi \rangle$ is clustered with the model (see Eq. 4.16) using the variational approach from Sec. 2.2.1 and the model 4.16 by resolving nonstationarity with four clusters. Within each cluster, a semivariogram is estimated. It is found that in all clusters, the $\gamma(\Delta z)$ is adequately represented by a Gaussian model:

$$\gamma(\Delta z) \sim a_1(1 - \rho(z)), \quad (4.25)$$

$$\rho(z) = \exp(-(z/l_z)^2), \quad (4.26)$$

where $\rho(z)$ is the Gaussian ACF, l_z is the correlation length parameter of the variable ϕ in height. The parameter a_1 is introduced to rescale the semivariogram to the range $[0, 1]$. The value 1 in Eq. (4.25) is the transformation between the functional form of the semivariogram and the ACF. The parameter estimates a_1^* and l_z^* are obtained using the least squares estimator and are reported in Tab. 4.3 along the identified regimes. The resulting ACF and the scaling of the variable l_z with \overline{Ri} are shown in Fig. 4.9. Figure 4.9a shows a clear separation into the

4. Stochastic Stability Equation

Table 4.3: Regime-dependent estimates of a Gaussian model for the vertical ACF. The indexed columns denote the clusters and the rows the parameters.

	1	2	3	4
l_z^*	21	34	2736	3323
a_1^*	14	34	33905	35775

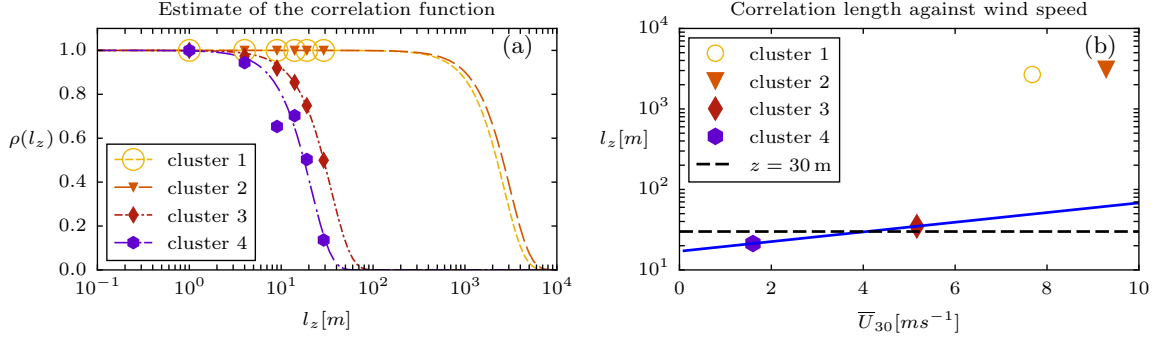


Figure 4.9: Estimation of the regime-dependent vertical correlation function $\rho(z, l_z)$ for the process ϕ . Each autocorrelation function $\rho(z) = \exp(-(z/l_z)^2)$ in panel (a) is estimated using the variogram data (not shown). The data are then rescaled to the form shown in (a). Panel (b) shows the regime-dependent estimates of length l_z as a function of wind speed at the highest measurement point of $z = 30$ m. Clusters 3 and 4 are the strongly stable regimes.

two regimes. Clusters one and two show that the variable ϕ is strongly correlated within the first 30m. The estimated correlation length for these regimes is a result of extrapolation and cannot be considered true. A reasonable interpretation is that a fully spatially connected ϕ is typical for the neutral boundary layer at large Re numbers and low \overline{Ri} numbers.

Clusters three and four in Fig. 4.9 show that the correlation length is within the measurement height. Examining the change in l_z over the mean wind speed at 30m height, we find evidence of scaling (see Fig. 4.9b). Given the maximum available height, it is reasonable to conclude that there is a log-linear relationship in FLOSS2 between the variables examined in Fig. 4.9b.

4.6 Model Validation

The identified SDE (4.16) is validated in three different ways. Comparison of the conditional probability density function of ϕ between data and model shows us how adequately the latter captures the long-term statistics. Simulating the sample evolution of the SDE and comparing it to the evolution of the measured ϕ provides insight into how adequately it reflects nonstationarity and thus interference in a localized space-time domain. Finally, examination of the temporal ACF provides insight into the dynamic memory effects present in the data and how well the model captures them. The given SDE validation tests aim to examine the performance of the identified model to recognize its weaknesses. The tests are also intended to provide a basis to support the embedding of the SDE model in a RANS closure.

4.6.1 Stationary Probability Density Function

For a fixed parameter \overline{Ri} , the SSE (4.16) will reach a stationary behavior after a some time. This behavior is described by a time-independent probability density function $p_s(\phi, \overline{Ri})$ of the stationary stochastic process ϕ . Although $p_s(\phi, \overline{Ri} = \text{const})$ is fixed, individual realizations of this process continue to fluctuate, so that individual observations $\phi(t_0, \overline{Ri})$ and $\phi(t_0 + \Delta t, \overline{Ri})$ belong to the density $p_s(\phi, \overline{Ri})$.

Stationary statistical properties such as the expected value $\mathbb{E}[\phi]$ and the most probable value $\mathbb{M}[\phi]$ can be calculated directly from p_s . The difference

$$I(\overline{Ri}) := \mathbb{E}[\phi] - \mathbb{M}[\phi] \quad (4.27)$$

is a function of \overline{Ri} and is defined as the measure of global intermittency for the SBL. The most significant parameter affecting the objectivity of the measure I is the two averaging scales in Eq. (4.3). However, this choice is determined to some extent by the multiresolution analysis (see Fig. 3.10).

The measure I is interpreted as follows. In the neutral state, $I \approx 0$ and $p_s(\phi, \overline{Ri})$ represents a Gaussian distribution. As \overline{Ri} , increases the expected value diverge from the mode, the measure I increases, and the shape of $p_s(\phi, \overline{Ri})$ changes toward an extreme value distribution.

The function $p_s(\phi, \overline{Ri})$ is computed from an ITÔ one-dimensional SDE assuming natural boundary conditions (Horsthemke, 1984):

$$p_s(\phi, \overline{Ri}) = \frac{Q}{g^2(\phi, \overline{Ri})} \exp \left(\int \frac{2f(\phi, \overline{Ri})}{g^2(\phi, \overline{Ri})} d\phi \right), \quad (4.28)$$

where the normalization constant Q is:

$$Q^{-1} = \int_0^\infty \frac{1}{g^2(\phi, \overline{Ri})} \exp \left(\int \frac{2f(\phi, \overline{Ri})}{g^2(\phi, \overline{Ri})} d\phi \right) d\phi. \quad (4.29)$$

The drift term is $f(\phi, \overline{Ri}) = 1 + \Lambda(\overline{Ri})\phi - \mathcal{V}(\overline{Ri})\phi^2$ and the diffusion term is $g(\phi, \overline{Ri}) = \Sigma(\overline{Ri})\phi$. To my knowledge, there is no closed-form solution for the normalization constant, and therefore it is evaluated numerically. However, given the constant Q a solution for the stationary density is found, which is:

$$p_s(\phi, \overline{Ri}, Q) = \frac{Q}{\Sigma^2(\overline{Ri})} \exp \left(-2 \frac{\phi^2 \mathcal{V}(\overline{Ri}) + 1}{\phi \Sigma^2(\overline{Ri})} \right) \phi^{2[\Lambda(\overline{Ri})/\Sigma^2(\overline{Ri})-1]}. \quad (4.30)$$

Next, the stationary distributions of the different parameterization strategies are examined. Recall that two parameterizations are performed. The piecewise parameterization (see Sec. 4.4.1) is denoted with P-parameterization, and the continuous parameterization (see Sec. 4.4.2) is denoted with C-parameterization. It is also important to remember that C parameterization is obtained from a reduced dataset and has nine parameters, and the P-parameterization has 17 parameters. The differences between the two approaches are discussed below.

Figure 4.10 shows the scatter plot for the measured quantities ϕ and Ri from all seven heights. The resulting p_s of the model is shown with solid contour lines. For comparison, the

4. Stochastic Stability Equation

stability function of the MSC atmospheric model is plotted with a dashed black line. The value of the Ri_c number is shown with a dashed red line. The expected value and distribution's mode are plotted above the \overline{Ri} number to perceive the distribution type better. The color of the data points indicates the height. The model captures the structure of the distribution (see

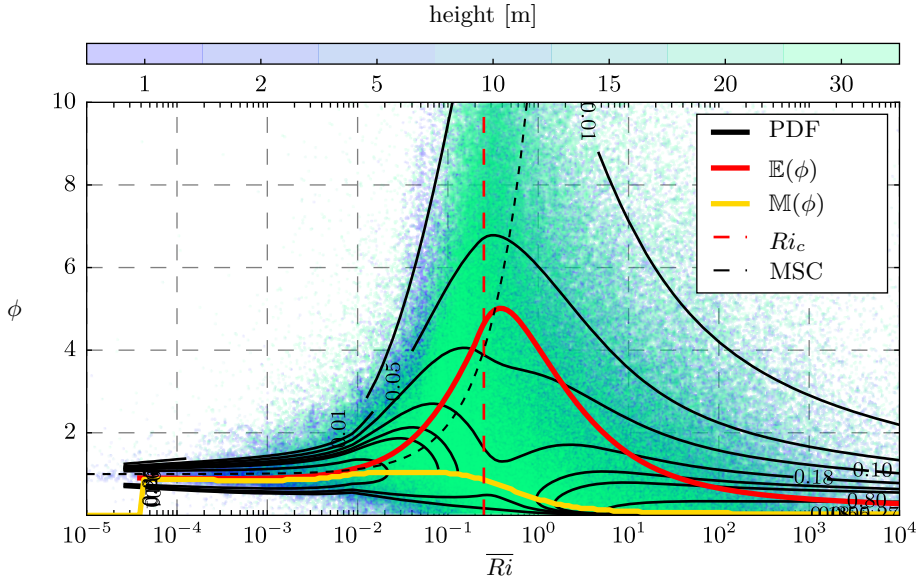


Figure 4.10: Stationary PDF for the variable ϕ calculated from SSE parameterized with the piecewise functions (see Sec. 4.4.1). The measured values are presented as a scatter plot, with the height color coded. The color code has a transparency factor of 0.2. The solid black lines show the contour lines of the $p_s(\phi, \overline{Ri})$ calculated from the model. The solid red line shows the expected value of the stationary PDF and the solid yellow line - the mode.

Fig. 4.10) as the data points fall into the distributional shape. The characteristic limit state $\overline{Ri} \rightarrow 0$ obeys a Gaussian distribution with expected value 1 and decreasing variance, which is also perceived by the reduced dispersion. The distribution for $\overline{Ri} = 10^{-5}$ apparently cannot be computed because the intensity of diffusion exceeds a possible minimum value. The expected value in the limit of small noise intensity is meaningful only for the parameter values of the drift term. In the limit, $\overline{Ri} \rightarrow 0$ the diffusion parameter does not change the expected value. This observation is obtained here empirically but should be studied analytically in the future. The analytical study will be beneficial if the proposed SDE is confirmed to some extent on more than one dataset.

It can be seen in the Fig. 4.10 that as the Ri_c number is approached from the left, the expected value begins to diverge from the mode of the distribution. Due to this divergence, the time evolution in a prediction of stability correction destabilizes over time and loses its equilibrium so that the coefficient of turbulent mixing exhibits intermittency.

The expected value starts to decrease after the Ri_c number (see Fig. 4.10 and Fig. 4.11 is more quantitative), which implies a relative increase in turbulent mixing. This behavior is not expected and represents the largest difference from traditional scaling. Such a large contrast is created by the nonstationary events in the mean wind profile. These events are considered in the identification of SDE. As noted in the multiscale analysis, the 1-hour mean wind profile loses its logarithmic shape (see Sec. 3.3.4). The velocity profile is prone to jet-like accelerations and decelerations (see Fig. 3.18). This energy of jet-like dynamics dominates the

flow. The activity of jet-like motions may be the reason for the increased averaged turbulent mixing captured by stochastic parametrization.

These results indicate that the effects of the sub-mesoscale motions are modeled on a scale from 3 min to 1 h along with the jet-like dynamics > 1 h. Quantification of the contribution of the sub-mesoscale motions is not possible with this stochastic modeling approach. On the other hand, it should be noted that the noise intensity of the stochastic model increases with the \overline{Ri} number (see Fig. 4.8c). As shown in the multiscale analysis, the sub-mesoscale motion band has a specific energy limit (see Ch. 3). Therefore, the noise intensity associated with the sub-mesoscale motions should also be limited. Consequently, it is reasonable to assume that in the context of the stochastic model (see Eq. (4.16)), the intensity of the noise exceeds a suitable limit representative of sub-mesoscale motions. After exceeding this threshold, the noise intensity should be physically related to the jet scale representation. Further research is needed to investigate this hypothesis.

The result that the model describes the more or less correct distribution of ϕ as the combined effect of the jet and sub-mesoscale motion is an indication that a more accurate period selection without the jet scales may yield a different limiting distribution for supercritical regimes $10 > \overline{Ri} > 1$. Further research on the classification of the motion scales is needed to gain a better understanding of the physics involved for the regime quantified by the SDE for $10 > \overline{Ri} > 1$. The isolation of the jet-like motions should eventually lead to the curve of the expected value in Fig. 4.10 following the traditional scaling. For practical reasons, one should consider a particular RANS model in parallel when parameterizing the SSE. If a RANS model predicts the jet dynamics, the stochastic model can be mixed with the traditional scaling for supercritical stabilities.

Some conclusions for the P-parameterization in Fig. 4.10. One has fine control over the stationary distribution for each stability regime of interest, with the drawback of having too many parameters for the scaling functions. In addition, adjusting particular predictions is difficult due to the large number of parameters. A more general parameterization is presented below to overcome the mentioned complications.

For a proper evaluation of the C-parameterization, where the scaling function is given by (see Tab. 4.2), the conditional density of the data is computed as a reference. Figure 4.11 shows in the upper panel the conditional probability density estimated with a Kernel Density Estimation (KDE) method using the training data. All seven heights are used for this purpose. The expected value and mode of the distribution are estimated from the density. A characteristic divergence of the mode and expected value is present after the critical value of \overline{Ri} . The distribution of the model is shown in Fig. 4.11 in the bottom panel and is largely similar to the distribution found in the data. The main difference is how the expected value deviates from the model. In the data, both statistical measures stay close to each other over a longer range of the \overline{Ri} number, and then, at the critical value, they diverge abruptly. According to the SSE model, the divergence is more uniform across the \overline{Ri} number. It is important to note that the conditional PDF of the data was also examined separately for each measurement height. It was found (not shown) that some individual heights exhibit the same type of moment divergence as given by the SSE model. Therefore, the discrepancy can be attributed to averaging effect over height. It is also likely that the logarithmic spacing of the

4. Stochastic Stability Equation

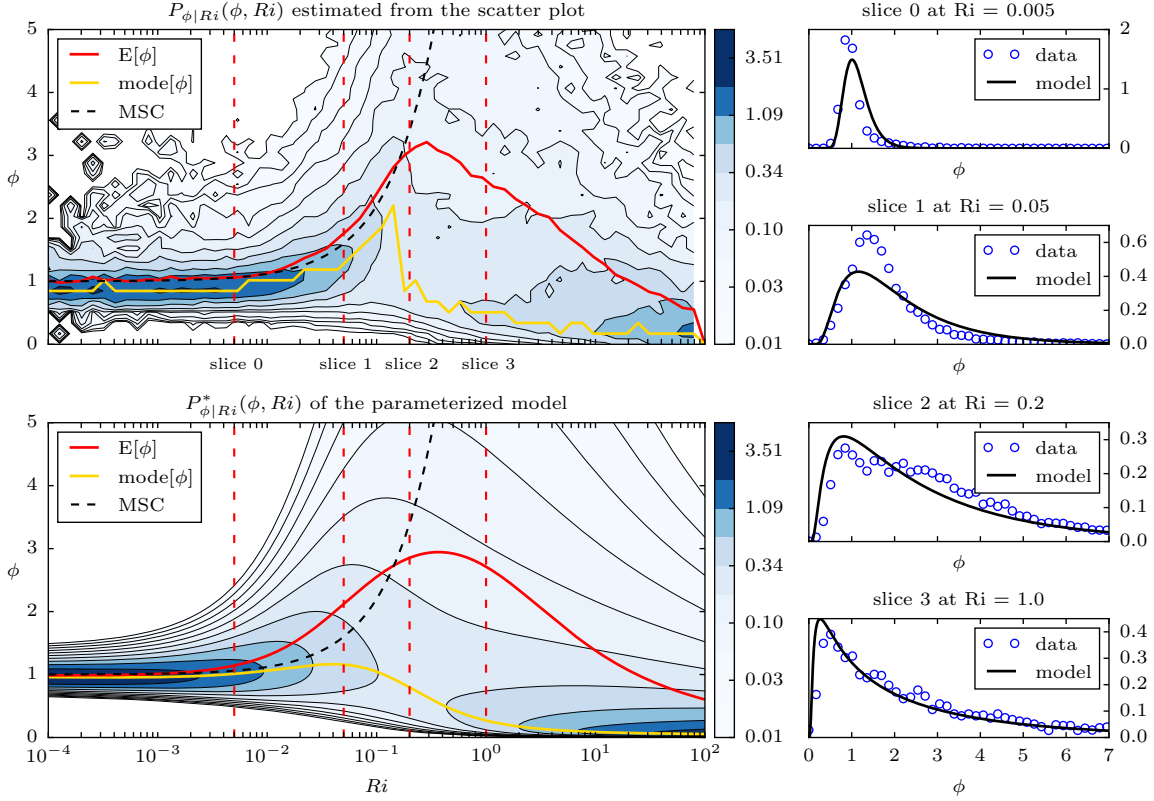


Figure 4.11: Stationary PDF for the variable ϕ calculated from SSE parameterized with continuous functions. The parametrization is performed on a reduced training dataset which is used in the upper panel. The stationary PDF conditional on Ri is estimated from the training data (top panel). The distribution predicted by the model is shown in the lower panel. The expected value (red) and mode (yellow) are plotted as a function of Ri for better comparison. The four panels on the right show the distributions for fixed values of Ri , indicated by dashed vertical lines in the left panels. The parameterization function of the MSC model is given in the left panels for reference.

measurements introduces some bias, as the region with higher wind gradients is sampled more frequently than the upper boundary layer.

The right panel in Fig. 4.11 compares the distributions of observed and modeled probability density of ϕ for four fixed \overline{Ri} values, represented by the vertical dashed lines in the panels on the left. The neutral, pre-critical, and post-critical regimes are adequately captured. The distributions for $0.05 > \overline{Ri} > 0.1$ of the model have a more pronounced right tail of the distribution. The left tail of the distribution is captured quite well. Attempts were made to better represent the model's distribution; for that, the scaling functions were slightly modified but without significant improvement. Therefore, the solution found by estimating the scaling functions using the least-squares estimation method is optimal.

As a concluding remark, both parameterization approaches produce reasonable results. The C-parameterization is closer to the data statistics and consists of a simple set of functions and is therefore preferable for use in a RANS closure.

4.6.2 Out-Of-Sample Forecast

The simulated time evolution of ϕ is validated on data excluded from training. The setup of the numerical experiment is to compute sample paths of the SSE by providing the evolution of

the \overline{Ri} number estimated from data. To do this, one needs the gradients of mean wind and temperature. To obtain these, Reynolds averaging is applied on a time scale of 1 h, separating the mean from the fluctuations $\overline{u}_i = u_i - u'_i$ and $\overline{T} = T - T'$. Then, at each time step, a least-squares fit of the log-linear equation following (Nieuwstadt, 1984) is performed:

$$f(z) = a_1 z + a_2 \ln(z) + a_3, \quad (4.31)$$

where $f(z)$ denotes $\overline{U} = \sqrt{\overline{u}^2 + \overline{v}^2}$ or \overline{T} . Local gradients are easily obtained via:

$$\frac{\partial f}{\partial z} = a_1 + \frac{a_2}{z}. \quad (4.32)$$

Then the predictor variable is estimated from data:

$$\overline{Ri} = \frac{\frac{g}{T_0} \frac{\partial \overline{T}}{\partial z}}{\left(\frac{\partial \overline{u}}{\partial z}\right)^2 + \left(\frac{\partial \overline{v}}{\partial z}\right)^2}, \quad (4.33)$$

where the overbar highlights the averaging scale of 1 h. The \overline{Ri} is used to solve the SSE:

$$d\phi = \left[1 + \Lambda(\overline{Ri})\phi - \mathcal{V}(\overline{Ri})\phi^2\right] dt + \Sigma(\overline{Ri})\phi dW(t), \quad \phi(0) = \phi_0, \quad (4.34)$$

where the parameter functions ($\Lambda, \mathcal{V}, \Sigma$) are given in Tab. 4.2.

It remains to specify the sampling procedure of the WIENER process. Recall that Eq. (4.34) is locally valid since the parameters are expressed by the local value of the \overline{Ri} number. Thus, for a one-dimensional SDE in a numerical discretization procedure, the discrete-time increment of the WIENER process $\Delta W = W(t + \Delta t) - W(t)$ is i.i.d. random variable with expected value zero and variance Δt : $\mathcal{N}(0, \Delta t)$ (Horsthemke, 1984). To combine the variable ϕ with a RANS model, it is required to prescribe the correlation structure $\rho(z)$ of ΔW in height such that the vertical gradients of ϕ exist. Moreover, the reference data show some correlation patterns in the intermittent regime that need to be modeled appropriately (see Fig. 4.13b). The spatial covariance is excluded from the classification analysis, so the obtained parameter scaling depends on the preprocessed nondimensionalization with height z (see Eq. (4.3)). Introducing a spatial covariance into ΔW requires a strong assumption that the spatial structure of ΔW is independent of the parameter-scaling function Σ .

The sampling method for the variable ΔW that is independent in time but correlated in the vertical direction is discussed next. For the numerical solution of Eq. (4.34), a stationary Gaussian process with zero mean and a spatial covariance function ρ must be sampled such that the samples are independent. This is achieved by using the circulant embedding method of Lord et al. (2014), which is an exact sampling method and is discussed below.

Let us denote the vector of samples, which represents the spatially correlated perturbations in height z by \mathbf{X} ,

$$\mathbf{X} := [X(z_1), X(z_2), \dots, X(z_{N-1})]^\top, \quad (4.35)$$

with uniformly spaced sample points $z_n = n\Delta z$ for $n = 0, \dots, N - 1$ on a space interval $z = [1, 30]$ m.

4. Stochastic Stability Equation

The vector \mathbf{X} must be sampled such that the autocovariance $Cov(X, X) = \mathbb{E}(X - \mathbb{E}(X))^2$ obeys a parametric covariance function $\rho(z)$. For a given covariance function $\rho(z)$, there is an associated real-valued, symmetric, nonnegative definite and circulant covariance matrix $\mathbf{P} \in \mathbb{R}^{N \times N}$. The function $\rho(z)$ evaluated on a discrete grid is the first column $\boldsymbol{\rho}_1 = [\rho_0, \rho_1, \dots, \rho_{N-1}]$ of \mathbf{P} . Given the specific matrix \mathbf{P} (for more details see Lord et al. (2014)), two time-independent samples $\mathbf{X}(t)$, $\mathbf{X}(t + \Delta t)$ can be sampled from $\mathcal{N}(0, \mathbf{P})$ applying the discrete KARHUNEN-LOEVE expansion:

$$\mathbf{Z} = \mathbf{X}(t) + \sqrt{-1} \mathbf{X}(t + \Delta t) = F D^{1/2} \boldsymbol{\xi} = \sum_{i=1}^N \mathbf{v}_i \sqrt{\lambda_i} \xi_i, \quad (4.36)$$

where \mathbf{Z} is a complex-valued Gaussian vector and F is a FOURIER matrix typically used in the discrete FOURIER transform. Together with the spectral decomposition of:

$$\mathbf{P} = F D^{1/2} D^{1/2} F^*, \quad (4.37)$$

the symbol D denotes the diagonal matrix with eigenvalues λ_i . The columns of F are orthonormal eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$ and $\boldsymbol{\xi} = [\xi_1, \xi_2, \dots, \xi_N]^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{1}_N)$. The matrix F^* is the complex conjugate of F . The multiplication of F by $D^{1/2} \boldsymbol{\xi}$ in Eq. (4.36) is formally the operation of the discrete FOURIER transform. To use the Eq. (4.36) in practice, the circulant matrix \mathbf{P} is constructed following the FOURIER representation of circulant matrices (see Proposition 6.45 in Lord et al. (2014)[p.243]), the diagonal entries of D are given by the inverse discrete FOURIER transform of the first column of the matrix \mathbf{P} :

$$\mathbf{v} := \sqrt{N} F^* \boldsymbol{\rho}_1. \quad (4.38)$$

Finally, two time-independent samples $\mathbf{X}(t, z)$, $\mathbf{X}(t + \Delta t, z)$ are obtained using Algorithm 6.7 of Lord et al. (2014) with the following steps:

1. Evaluate $\rho(z)$ on a discrete grid to obtain the vector $\boldsymbol{\rho}_1$
2. Compute $\mathbf{v} = N F^* \boldsymbol{\rho}_1$
3. Compute $\mathbf{y} = \mathbf{v}^{1/2} \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \mathcal{CN}(\mathbf{0}, 2 \cdot \mathbf{1}_N)$ is complex-valued
4. Compute $\mathbf{Z} = (F \mathbf{y}) / \sqrt{N}$
5. Get the real and imaginary parts: $\mathbf{X}(t, z) = \Re(\mathbf{Z})$, $\mathbf{X}(t + \Delta t, z) = \Im(\mathbf{Z})$

Now that we know how to sample the spatially correlated perturbation, we are ready to solve Eq. (4.34). The first simulation is examined using the P -parametrization, with the corresponding conditional probability distribution in Fig. 4.10 and the parameter functions $(\Lambda, \mathcal{V}, \Sigma)$ are given by Eqs. 4.21, 4.22, 4.23. The covariance function of the perturbations is a MATÉRN covariance (Lord et al., 2014) with a constant correlation length of 15 m as an example. The actual correlation function of the dataset is examined in Sec. 4.5.

Figure 4.12 shows the measured and simulated profiles of the variable ϕ throughout 100 hours within the training data. The sampling frequency is 1 min for both variables. The right panel shows the statistics averaged over the period (shown with the boxplots) and

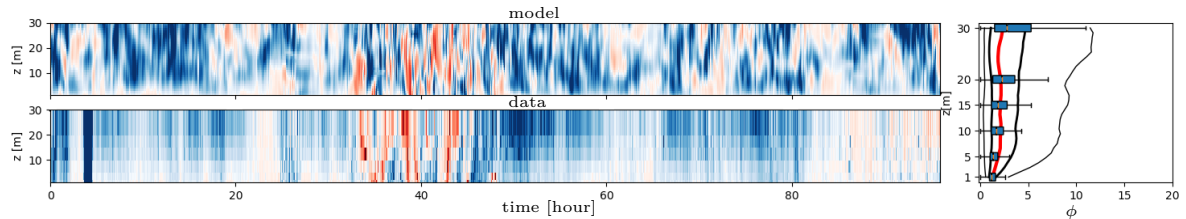


Figure 4.12: In-sample prediction comparison using the piecewise parameterized SSE with MATÉRN covariance function for the vertical dimension. The top panel shows the prediction of the model within the training data shown in the bottom panel. The right panel compares the total statistics profile found in the data (boxplots) and those predicted by the model (solid lines).

the model (marked with solid lines) for the height considered. The vertical red line is the median and closely follows the median of the boxplots. The black solid lines represent the percentiles 5% 25% 75% and 95% of the distributions. It is noticeable that the percentiles 5% 25% are in excellent agreement with the respective quantiles of the box plots. The 75% and 95% percentiles are noticeably larger than the data. The overall evolution of the simulated ϕ adequately captures the highly intermittent period around time frame 40 h. The model resolves the fine structure in the upper height, which is difficult to see in the measured variable due to larger spacing intervals. On average, the real data show a more regular, uniform development pattern within neutral conditions. In addition, the model predicts some intermittent local patches of increased mixing over the ground, which the data do not confirm. A possible explanation would be that the disturbance changes its correlation length as a function of stability conditions. The P-parameterization approach reproduces the ϕ trend. It is also important to recall that the period considered is part of the training data set and is expected to be valid in terms of record-averaged statistics.

Next, we compare the evolution of the ϕ profiles using the C-parameterization (see Sec. 4.4.2) within the validation data not used in the clustering procedure. The input variable is the \overline{Ri} number estimated on a scale of one hour. Equation (4.16) is solved using the parameterization functions in Tab. 4.2. The vertical correlation structure of the WIENER process is given by the estimates in Fig. 4.9. The correlation length l_z is given by the scaling found in the data and is calculated from the value of the mean wind at the highest point \overline{U}_{30} as

$$\log_{10}(l_z) = 0.06 \overline{U}_{30} + 1.23. \quad (4.39)$$

The parameters of the previous function are estimated using only the values of the weak wind regimes (see Fig. 4.9). The correlation length for the strong wind regime is considered unreliable because it exceeds the measurement height.

Equation 4.16 is solved numerically using the MILSTEIN method with a time step $\Delta t = 1.6 \cdot 10^{-4}$ and then downsampled to $\Delta t = 1.6 \cdot 10^{-2}$. The spatial height coordinate is resolved using 32 points. However, the spatial resolution does not affect the solution since there are no gradients in the current formulation. The spatial resolution is set to resolve the smallest possible correlation length about 10 m.

The predictor variable \overline{Ri} varies on a time scale slower than 1 h. Since the traditional scaling law for predicting ϕ is a static function, its time evolution is also slower than 1 hour

($\bar{\phi} = 1 + 12\bar{Ri}$). Solving Eq. (4.16) yields a ϕ with stochastic fluctuations much faster than 1 hour. Therefore, for comparison, the modeled variable ϕ is temporarily averaged with a moving window of length 1 hour ($\bar{\phi}$).

The circular embedding method generates the spatially correlated perturbation patterns with a Gaussian correlation function. Since the method is based on periodic functions, the size of the circulant matrix here is set ten times larger than the number of spatial grid points. This avoids the undesirable effect of a pattern entering the area from another side due to the periodic boundary conditions.

Figure 4.13a shows the evolution of the \bar{Ri} number and Fig. 4.13b shows the corresponding evolution of $\bar{\phi}$ for several nights of the validation dataset. The maximum value of \bar{Ri} in the forecast is limited to 10. The color bar in Fig. 4.13a is set so that the white areas indicate the critical value $Ri_c = 0.25$. Accordingly, the red color indicates the neutral conditions and the blue color indicates the highly stratified conditions. The color bar in Figure 4.13b is set so that the white areas denote $\bar{\phi} = 1$, which corresponds to neutral conditions. Accordingly, the blue color denotes the suppressed turbulence level $\bar{\phi} \gg 1$ and the red regions represent the enhanced mixing due to turbulence $\bar{\phi} < 1$. Different flow conditions are marked with a red frame as an example. It can be seen that for the intermittent conditions with high values of \bar{Ri} (see Fig. 4.13a), the variable $\bar{\phi}$ (see Fig. 4.13b) shows extreme mixing events. Recall that $\bar{\phi}$ represents the relative deviation of turbulent mixing as a function of the mean wind gradient. For the intermittent conditions in Fig. 4.13, the gradient of the mean wind is lower compared to the other two marked regimes.

The performance of the stochastic model is analyzed on the selected data by simulating $\bar{\phi}$ using the evolution of \bar{Ri} calculated from the data (see Figs. 4.13c,d,e). The solution in the three panels differs by a model parameter σ_s , representing the sub-mesoscale motion's intensity. This parameter is defined in Tab. 4.2 in the function that parameterizes the scaling of the noise process in the SSE. The identification study (see Sec. 4.4.2) showed that $\sigma_s = 0$ is optimal based on the training dataset. However, the parameter $\sigma_s = 0$ is not optimal for the validation dataset since the observations show a slightly weaker sub-mesoscale activity. Based on the found optimum $\sigma_s = 0$, it is easy to adjust the simulation of the model to the considered realization of the measured $\bar{\phi}$.

The activity of the sub-mesoscale motions is controlled by changing the value of the parameter σ_s that shifts the scaling function (see Fig. 4.8c) along the y-axis. Consequently, it is easy to suppress the intensity of the sub-mesoscale motions by decreasing the value of σ_s . This is done in Fig. 4.13d,e. Figure 4.13a shows the solution of the stochastic equation with the identified parameter $\sigma_s = 0$. Comparing this solution to the data within the selected regimes (see Figure 4.13b), we find that the intensity of the extreme mixing events is too high. Figure 4.13d shows a slightly adjusted value of $\sigma_s = -0.07$, which lowers the intensity of the extreme events to a reasonable level for all three regimes considered. More quantitative analysis for fitting the σ_s parameter is presented later in this section.

A further decrease in the value of the parameter σ_s provides an even stronger elimination of fast perturbations (see Fig. 4.13e). This limit is interesting because one can wonder how the model behaves in the regimes of precritical stability. To analyze this question, one can examine the evolution of $\bar{\phi}$ given by traditional scaling (compare Fig. 4.13e with 4.13f). It is assumed

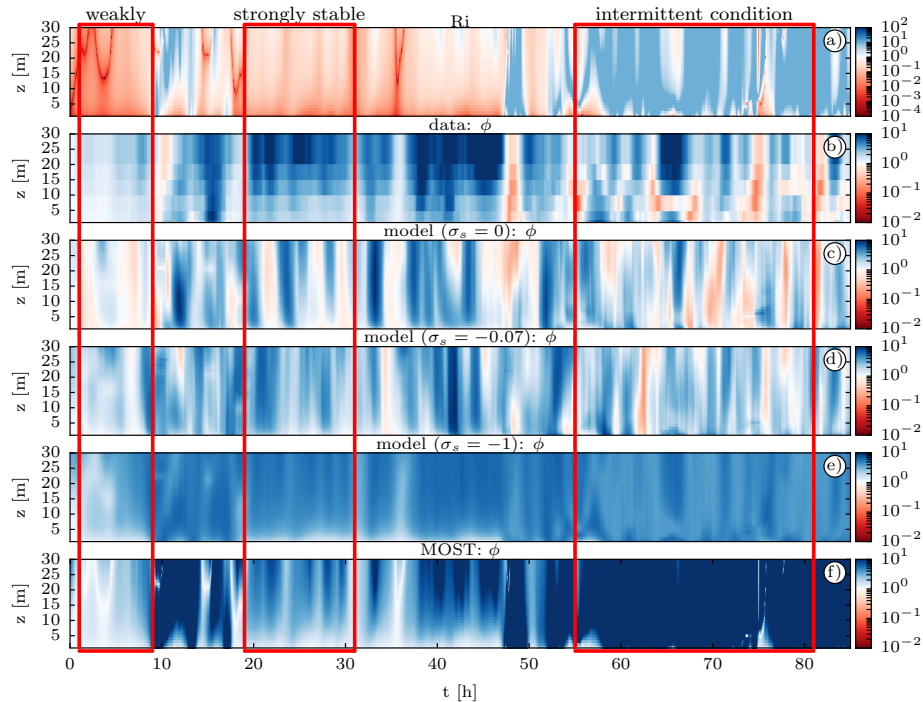


Figure 4.13: Comparison of the profile evolution of the modeled ϕ on the validation dataset. a) Evolution of \overline{Ri} number is calculated on a scale of one hour $(\overline{\cdot})$ and is used to predict ϕ evolving on a scale of three minutes. b) Observed variable $\overline{\phi}$ in the validation dataset. c) Prediction of $\overline{\phi}$ with the C-parameterized model (see Eq. 4.34). d) Prediction of $\overline{\phi}$ with empirically fitted noise level ($\sigma_s = -0.07$). e) Prediction of $\overline{\phi}$ with the suppressed noise level. f) Prediction of $\overline{\phi}$ with traditional scaling. The red frames mark dynamically different regimes.

that the stochastic model in the limiting case without small-scale perturbations should converge against the classical scaling for $\overline{Ri} < Ri_c$. Comparing the weakly and strongly stable conditions (Figs. 4.13e with 4.13f), we find that $\overline{\phi}$ of the stochastic model resembles the classical scaling prediction. It also captures the main tendency by evolving from the weakly stable to the strongly stable regime. However, it appears that classical scaling performs favorably with respect to observations in the weakly and strongly stable regimes (compare Fig. 4.13f with 4.13b). The comparison suggests that the stochastic model is not sensitive enough to the vertical gradient of the \overline{Ri} number in the limiting case of low small-scale perturbation. This is because the expected value of the invariant measure starts to fall after criticality (see Fig. 4.11). However, if we compare the intermittent regime (see Fig. 4.13e with 4.13f), the stochastic model with parameter $\sigma_s = -1$ yields a finite turbulence suppression level, while the classical theory predicts the absence of turbulence.

In summary, the stochastic model captures the full range of regimes in the nighttime boundary layer exceptionally well. The model reproduces the spatio-temporal patterns in the data. In addition, the intensity of the sub-mesoscale motions will be adjusted by changing a parameter value. How the adjustment of this parameter can be controlled and quantified is explained below.

Consider the selected regimes in Fig. 4.13 (red frames). One way to understand the effects of parameter σ_s is to compare the distributions of the vertically averaged variables $\overline{\phi}$. The distributions are compared quantile-wise in a so-called quantile-quantile plot (see Fig. 4.14).

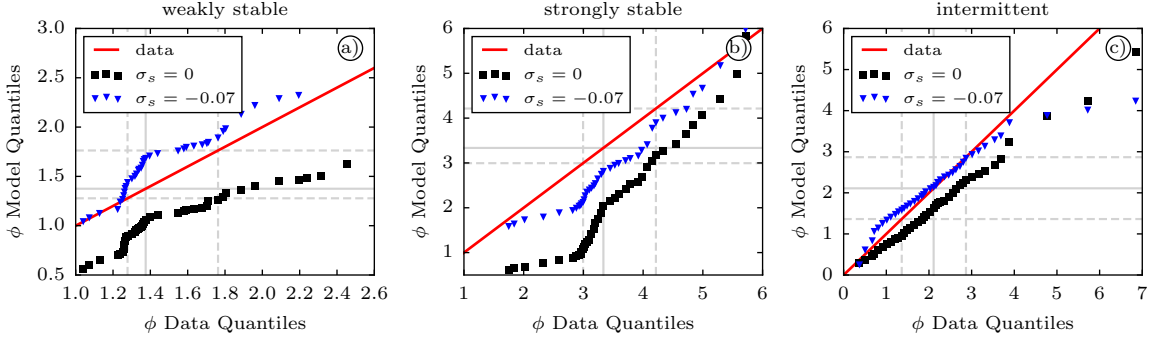


Figure 4.14: Quantile comparison of height-averaged PDFs of the model within selected periods in Figure 4.13. a) Selected weakly stable state. b) Selected strongly stable state. c) Selected intermittent state. The solid red line in each box represents the quantiles of the data distribution for reference. The distribution is recovered when the points lie on this line. The solid gray line is the median of the data. The dashed gray lines mark the respective upper and lower quartiles that make up the interquartile range on each axis. Points above the red line indicate weaker mixing, those below indicate stronger mixing than found in the data per quantile. Regions outside the interquartile range can be interpreted as rare mixing events.

The red reference curve in all three panels is created by plotting the quantiles of the data distribution against itself. Any other distribution that is to be compared to the reference is plotted in terms of its quantiles over the quantiles of the reference distribution. If the points on the reference line coincide, then the model captures the data distribution well. Another useful feature of such a plot is the derivation of the effects of turbulent mixing. Points below the red line indicate intense mixing, and points above indicate too weak mixing. Depending on how far the points are from the median (in all three plots, the median is represented by the solid gray line on both axes), one can judge how rare the extreme events are and whether they cause stronger or weaker mixing than the observations.

In the following, we investigate the effect of changing the intensity of the sub-mesoscale motions of the model. The originally parameterized model with parameter $\sigma_s = 0$ produces too much mixing for the considered validation dataset (the quantile marked in black is below the red reference line (see Fig. 4.14)). For all three regimes, reducing the parameter value from $\sigma_s = 0$ to $\sigma_s = -0.07$ is beneficial because the quantiles are closer to the reference line (see Figs. 4.14a, b, c), with the best estimate found in the intermittent regime. The median of the weakly and strongly stable periods is overestimated and underestimated, respectively. The right tail of the modeled distribution $\sigma_s = -0.07$ shows that extreme events cause slightly less mixing. In contrast, the model $\sigma_s = -0.07$ in the strongly stable period tends to overestimate mixing in general but is better than the model with $\sigma_s = 0$. Overall, fitting the parameter leads to improved performance; however, it is important to remember that the plots represent one realization of the solution. Indeed, the quantiles are different in another realization but still obey the trends described above with respect to the change in σ_s .

4.6.3 Temporal Autocorrelation and Self-Similarity

The temporal ACF is estimated exemplary only for the height of 15 m, although more heights are available. The study is based on the results of clustering the time series of ϕ with Eq. (4.16) using five clusters. The research interest aims to understand the time evolution dynamics in

the identified clusters and the relationship with the Ri number. It is also of interest whether the proposed SDE reproduces the temporal ACF of the data. The temporal ACF is defined as:

$$R_{\phi\phi} := \mathbb{E}[\phi(t)\phi(t + \tau)], \quad (4.40)$$

assuming that the process ϕ is stationary. The assumption of stationarity is justified within each identified cluster by the stationary parameter values of the underlying SDE. However, constant parameter values of a SDE do not directly imply a stationary behavior of the process. Since the considered SDE is nonlinear, the sample path and data still exhibit some degree of nonstationarity within each cluster, especially for higher values of Ri number.

Moreover, the regime affiliation function shows that the transitions between regimes are sometimes quite long. Accurate quantification of regime states is questionable during transient states. Therefore, the time series evolution and the corresponding ACFs are expected to experience some variability within the identified regimes.

To account for this variation, the ACF within each regime are consequently estimated several times and then averaged. Because the total occupation time of the regimes varies across regimes, the number of ACFs available to estimate the average also varies. However, the estimated confidence intervals reflect a sufficient number of observations within each regime. The left column of panels in Fig. 4.15 shows the regime-based ACF for the data and model, labeled with the median Ri number for each. Panels are sorted from top to bottom by increasing Ri number.

For reference, an exponential ACF is estimated from the data ACF (blue) and normalized by its variance. The variance within each regime is known indirectly through the identified function (see Fig. 4.8c). The ACF is computed with a maximum lag of 3 hours and normalized by 1 hour. This is a reasonable choice since the mean wind profile is on this scale. The window for estimating ACF is set to 6 hours to account for the large lag values. The light blue areas in Fig. 4.15 show the confidence intervals.

According to Fig. 4.15, the autocorrelation function for ϕ follows an exponential decay only for the first few lags in the data. Moreover, the decorrelation time shrinks as the Ri number increases (see the panels from top to bottom). Above a specific value, the autocorrelation deviates from the exponential curve and begins to decay more slowly. This behavior indicates the presence of some long-lasting memory effects. The observed delay in the decay rate of ACF is present in all regimes. A comparison of the model results shows that the found kink in ACF is not captured in any of the regimes. However, the model reproduces the decorrelation time (where the autocorrelation meets the x-axis) relatively well with respect to the Ri number. However, the decorrelation time in each regime is shorter than in the data. The confidence intervals at the bottom of each panel indicate the standard deviation and variability of the ACF number. The largest variability is found at $\overline{Ri} = 0.56$, shown in the middle panel.

According to the WIENER–CHINTSCHIN theorem (Leibovich & Barkai, 2015), the FOURIER transform of the ACF gives the power spectral density of the process ϕ and is shown in the right column in Fig. 4.15. The frequency is constrained on the x-axis between 0.3 and 30 cycles per hour. The spectral energy decay follows a power law with an exponent value of -0.75 and is reproduced by the model in each range. These results suggest that the process is ϕ self-similar, assuming the proposed model structure. The model structure may bias the

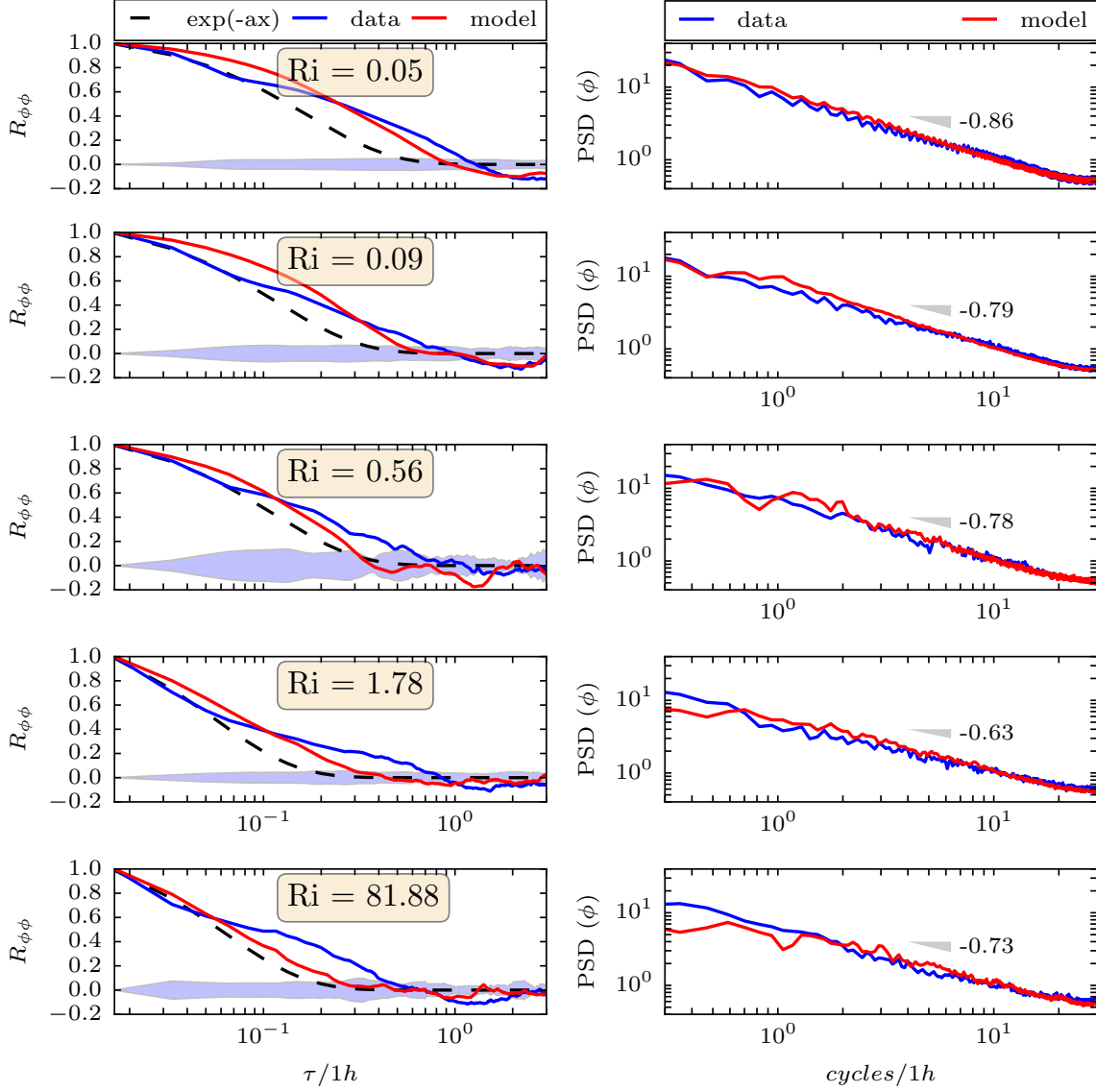


Figure 4.15: The left column shows the regime-averaged temporal autocorrelation function of the variable ϕ for a time series modeled at height 15 m. Blue labels the observations, and red color labels the ACF of the model. The right column shows the regime-dependent Power Spectral Density (PSD) of the data and the model. The slope is marked in each panel with a gray triangle.

classification. For example, a linear model may model the time series better than the proposed one at low Ri number. Therefore, it is possible that the optimal model structure depends on the Ri number and consequently leads to different spectral densities.

4.7 Conclusion and Discussion

In this chapter, a stochastic model is introduced and validated to parameterize the intermittent turbulence in SBL. The approach accounts for turbulent mixing arising from the sub-mesoscale of motions. This is achieved by formulating a SDE based on the existing scaling relation in combination with a parameter perturbation approach. The validity of the adopted functional form of the SDE is confirmed by comparing several closely related functional forms. The complexity of the proposed functional form is shown to be simplest way to achieve a valid

representation of the dynamics for the considered variable ϕ . However, further simplifications of the functional form are possible. In this work, the SDE has three nonstationary parameters and a quadratic nonlinearity.

The proposed SDE parameterization is achieved using the developed model-based clustering method (see Sec. 2.2.1). The methodology discovers a suitable scaling of the stochastic model parameters with the local Ri number (see Fig. 4.8), which allows application in a RANS turbulence model. These results support that the SDE is suitable to model the entire regime range. The parameter estimation procedure shows that nonstationarity is captured; additional isolated regime studies should be performed concerning the SSE. In particular, numerical or wind tunnel experiments can provide controlled case studies, for example, by perturbing the stable SBL at a constant Ri number with small momentum injections. Under controlled conditions, the SSE can be fitted in time assuming constant parameters. Such parameterization studies are more accessible to perform than applying the model-based clustering method to field measurements.

The SSE model adequately captures the invariant density and the resulting dispersion of the stability correction variables. The observation of non-Gaussian statistics can explain the intermittent behavior as follows. Global intermittency is explained by the increased divergence of the expected value from the distribution mode. For sufficiently large observation periods, such as months and years, the expected value corresponds to the degree of turbulence suppression (see red line in Fig. 4.11). However, if we start observing in a short time frame such as hours or nights, the most likely value is the mode of the distribution (see yellow line in Fig. 4.11), and thus turbulent mixing is present.

In this study, only one data set was analyzed. A broader validity of the stochastic equation is questionable and should be explored in other field experiments. The important question to investigate is: How general is the functional form of the stochastic model? As geostrophic conditions change from place to place, flow patterns also change, and therefore the parameterization functions of the SDE can be expected to change. If a particular parameterization design is fixed, the identification of the parameters can be made directly by adjusting the PDF moments of the process ϕ , thereby avoiding the clustering method.

It is shown that the out-of-sample prediction of the variable ϕ is a good approximation of the spatio-temporal evolution of the data (see Fig. 4.13). The presented stochastic model can be tuned to optimize the intensity of intermittent turbulent mixing by changing one parameter. The stochastic activity is also present in the weakly stable regimes and may be undesirable when modeled with the RANS model. However, the induced variability in the mixing is Gaussian and does not change the long-term behavior. To suppress the variability in the weakly stable regime, one must change the presented parameterization of the diffusion term in the desired Ri range. A discontinuity in the scaling is possible and might be exploited. However, when changing parameters rapidly, one wants to account for the delay required for equilibration. Consequently, discontinuity of the parameters of the SSE should be treated carefully.

A detailed examination of the temporal ACF revealed a characteristic kink that causes a longer decaying correlation compared to an exponential decaying ACF. The stochastic model cannot represent the kink observed in data but captures the decorrelation time scale correctly.

4. Stochastic Stability Equation

Some additional observations arising from unreported observations of the model performance analysis are summarized next. The C-parameterization (see Tab. 4.2) appears to be too over diffusive compared to the P-parameterization (see Sec. 4.4.1). This is mainly because the divergence of the moments is too smooth. Another reason could be that the data from the training was excluded for validation – also, the integral scale changes with the Ri number. The integral scale is a constant averaged value within the strongly stable regime. Fixing the integral scale leads to increased variance of ϕ for the weakly stable regime present in the C-parameterization. It is possible to improve this problem by integrating the temporal variability of the turbulent scale into the pre-analysis step for estimating ϕ . To recover the temporal variability of the integral scale, one must perform its estimation within each identified regime and reconstruct its temporal evolution using the affiliation function (see Fig. 2.17). Once the time-dependent integral scale is known, REYNOLDS-averaging is repeated with a time-dependent scale. Such time-dependent averaging results in a more accurate estimation of the ϕ signal for the weak and neutral conditions.

5

Stochastically Unsteady Reynolds-Averaged Navier–Stokes Equations

This work aims to model intermittent turbulence by identifying and developing a data-driven surrogate model. Previous chapter defines and validates the SSE (4.33) that models the momentum stability correction in the SBL as a stochastic process. This process represents a non-dimensional quantity by which turbulent diffusion is suppressed due to stratification. The SSE accounts for the random turbulent mixing phenomena likely to be generated by non-turbulent sub-mesoscale motions (see Ch. 3).

In chapter 4, the SSE is validated in a restricted setting where only the modeled variable is examined. Although the results anticipate promising applications, one would like to understand how to use the stochastic equation in predicting SBLs. To complete the data-driven development cycle of an intermittent turbulence closure, the SSE is coupled with a 1.5 order RANS turbulence model.

An innovative turbulent mixing length is introduced to combine a RANS model with the SSE model. Details of the numerical discretization are presented, using a blending procedure to combine the deterministic and stochastic system structures. The resulting turbulence closure is termed Stochastically Unsteady REYNOLDS-averaged NAVIER–STOKES Equations (SURANS) because the introduced sub-grid perturbations prevent the model from reaching an equilibrium state.

The overall modeling concept is critically evaluated against a turbulence model without stochasticity. Several numerical experiments are performed to assess the solution properties of the SURANS closure in simulating neutral and strongly stable regimes. A simulation with varying forcings is also performed to evaluate the transient dynamics of the model. In addition, the effect of non-turbulent sub-mesoscale motions on the temperature inversion within the SBL is investigated. This chapter provides insights into the numerical implementation and numerical stability, performance, and validation of the SURANS model. The results provide a

solid foundation for further research questions and may even motivate testing the SSE in a global circulation model.

5.1 Model Formulation

5.1.1 Extending a Reynolds-Averaged Navier–Stokes Model

The SURANS model is a set of prognostic equations for simulating intermittent SBL. The main difference to the RANS model is an added SSE representing the evolution of a stability correction variable. The SSE is limited at this stage of research to the surface boundary layer, and hence the effect of intermittent mixing is modeled to a certain maximum height above the surface because the SSE is calibrated based on measurements up to 30 m.

Consider the following assumptions for a classical RANS to simulate a SBL. It is assumed that the BOUSSINESQ approximation and the geostrophic balance holds. The coordinate system is aligned with the direction of the geostrophic wind. The surface is considered to be flat. The turbulence is assumed to be homogeneous in the horizontal plane, and the dissipation is in balance with the production. The vertical turbulent fluxes are more relevant than the horizontal ones. The flow is incompressible, the humidity and the air viscosity are neglected. A more detailed derivation of the RANS model can be found in (Stull, 1988). Hence, the set of equations for the SURANS model is:

$$\frac{\partial u}{\partial t} = (v - v_g)f_c + \frac{\partial}{\partial z} \left(K_m(\phi) \frac{\partial u}{\partial z} \right) - N_u \quad (5.1)$$

$$\frac{\partial v}{\partial t} = -(u - u_g)f_c + \frac{\partial}{\partial z} \left(K_m(\phi) \frac{\partial v}{\partial z} \right) - N_v \quad (5.2)$$

$$\frac{\partial e}{\partial t} = \frac{\partial}{\partial z} \left(K_m(\phi) \frac{\partial e}{\partial z} \right) + K_m(\phi) \left[\left(\frac{\partial u}{\partial z} \right)^2 + \left(\frac{\partial v}{\partial z} \right)^2 \right] - \frac{g}{T_0} K_h(\phi) \frac{\partial T}{\partial z} - \frac{(\alpha_\varepsilon e)^{3/2}}{l_m(\phi)} \quad (5.3)$$

$$\partial \phi = \tau_h^{-1} (1 + \Lambda(Ri)\phi - \mathcal{V}(Ri)\phi^2) \partial t + \tau_h^{-1/2} \Sigma(Ri)\phi \partial W(t) \quad (5.4)$$

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial z} \left(K_h(\phi) \frac{\partial T}{\partial z} \right) \quad (5.5)$$

$$\frac{dT_g}{dt} = \frac{1}{C_g} (R_n - H_0) - \kappa_m (T_g - T_s) \quad (5.6)$$

Equation (5.5) provides the evolution of the temperature without radiative effects. The surface temperature T_g is the boundary condition of Eq. (5.5) and is modeled utilizing the force-restore method (Garratt, 1994; Stull, 1988, Ch. 7) in formulation given by Acevedo et al. (2021). The thermal capacity of the soil is denoted with C_g . The resulting soil heat transfer coefficient $\kappa_m = 1.18\omega$ is parametrized relating the Earth’s angular frequency. The term H_0 denotes the turbulent sensible heat flux and is in more detail explained in Sec. 5.1.3; R_n is the net radiation and incorporates, among others, the effect of cloud cover. More detailed parametrization of this term is used by (Maroneze et al., 2019). The temperature below the surface T_s at some finite depth is nearly constant and fluctuates on a seasonal scale. It is, therefore, deemed fixed for the simulation of individual nights.

Equation (5.4) is the novel contribution to the classical RANS model and is detailed next. Equation (5.1) and (5.2) are the REYNOLDS-averaged momentum equations in geostrophic

balance forced by the mean geostrophic wind components (u_g, v_g) . The profile of these components is uniform in height. The wind vector components are denoted with (u, v) , t is time in seconds, z is height in meters, and f_c is the CORIOLIS parameter. The kinematic turbulent fluxes for momentum and heat are modeled according the eddy-viscosity theory ($\overline{w'u'} = -K_m \frac{\partial u}{\partial z}$, $\overline{w'v'} = -K_m \frac{\partial v}{\partial z}$, $\overline{w'T'} = -K_h \frac{\partial T}{\partial z}$). The turbulent PRANDTL number $Pr_t = \frac{K_m}{K_h}$ is for simplicity set to one. (Although stochastic modeling of K_h is possible with additional data analysis, and even as a coupled 2D system for K_h and K_m , this numerical study focuses solely on modeling K_m). Relaxation term $N_u = (u - u_g)/\tau_r$ and $N_v = (v - v_g)/\tau_r$, where τ_r is the relaxation time, forces the solution over the entire column towards the geostrophic wind (Vignon et al., 2018). The relaxation terms are a trade-off to suppress the inertial oscillations (detailed explanation is provided in Sec. 5.1.4). The value of τ_r [h] is set in a range of 3 – 6 hours, such that the solution is dictated by (5.1) and (5.6) rather than by the geostrophic forcing (u_g, v_g) . A study with changing geostrophic wind is conducted in Sec 5.3.3.

The prognostic Eq. 5.3 describes the evolution of the TKE denoted with e , where $g = 9,81$ [m s^{-2}] is the gravitational acceleration, $T_0 = 300$ [K] is the reference potential temperature, $\alpha_\varepsilon = 0.1$ is the modeling constant of the dissipation, and l_m is the turbulent mixing length. The first production term includes the vertical gradient of the wind components. The third buoyancy term is in this work always positive (stable stratification) and hence functions as a sink. The last term in Eq. (5.3) is the parametrized turbulence dissipation rate. Furthermore, it is assumed that the diffusion coefficient of the LAPLACE operator in Eq. (5.3) is equal to K_m . The reader is referred to Rodrigo & Anderson (2013) for more details on the parametrization of the TKE equation. Next, the gradient *Ri* number is used:

$$Ri = \frac{\frac{g}{T_0} \frac{\partial T}{\partial z}}{\left(\frac{\partial u}{\partial z}\right)^2 + \left(\frac{\partial v}{\partial z}\right)^2}, \quad (5.7)$$

where its value is confined in the range $[10^{-5}, 10]$. The essential closure elements of the model are explained next. The first one being the kinematic eddy-viscosity coefficient, which is modeled as:

$$K_m = \alpha l_m \sqrt{e}, \quad (5.8)$$

where $\alpha = 0.46$ is a modeling constant for neutral boundary layer (Rodrigo & Anderson, 2013). The turbulent mixing length is:

$$l_m = \frac{\kappa z}{\varphi(t, Ri) + \frac{\kappa z}{\lambda_b}}, \quad (5.9)$$

where $\varphi(t, Ri)$ is a nondimensional stochastic process as a convex combination of the process ϕ and function ϕ_f that corrects the turbulent mixing length of stratified flow depending on the *Ri* number. Variable φ is defined later in text. Also, this process accounts for the variation of the mixing length induced by the sum of different non-turbulent sub-mesoscale mixing events. The term $\frac{\kappa z}{\lambda_b}$ is the BLACKADAR's extension (Blackadar, 1962). Following Rodrigo & Anderson (2013), the value λ_b , which restrains the size of the largest turbulent eddies in

neutral stratification is parametrized as:

$$\lambda_b = 2.7 \times 10^{-4} \frac{u_g}{|f_c|}, \quad (5.10)$$

where $f_c = 2\omega \sin(\varphi)$, with $\varphi = 40^\circ\text{N}$ and $\omega = 7.27 \times 10^{-5} \text{s}^{-1}$. The stochastic process ϕ is expressed by the prognostic Eq. (5.4) with the scaling functions obtained in Ch. 4:

$$\Lambda(Ri) = 9.3 \tanh[0.9 \log_{10}(Ri) - 0.1] + 8.3, \quad (5.11)$$

$$\mathcal{V}(Ri) = 10^{(0.4 \log_{10}(Ri) + 0.2)}, \quad (5.12)$$

$$\Sigma(Ri) = 10^{(0.8 \tanh[0.6 \log_{10}(Ri) - 0.8] + \sigma_s)}, \quad (5.13)$$

where σ_s (see Eq. (5.13)) regulates the intensity of the stochastic turbulent diffusion assumed to be related to sub-mesoscale motions. The parameter σ_s can be adjusted in the range $[-1, 0]$. The value -1 equals to the considerably low intensity of the noise, such that the solution of Eq. (5.4) becomes nearly deterministic. The value 0 corresponds to the level of the FLOSS2 dataset and models relatively intense perturbations. Parameter values obtained for FLOSS2 are given in Tab. 4.2. The identification of the scaling function is performed from tower data (see Ch. 4).

The following critical assumption is implicitly asserted. Consider first that the scales of motions smaller than the geostrophic wind influence the model's local mixing length stochastically. The noise term in Eq. (5.4) imitates this circumstance. Then it is assumed that for each sub-mesoscale type of motion, the KOLMOGOROV theory of inertial subrange holds locally in space, and the parametrization of the dissipation is identical to the neutral case. Essentially, each non-turbulent sub-mesoscale motion cascades to the dissipative eddies raising KOLMOGOROV-like turbulence. However, this assumption is rather doubtful and demands further research. For instance, based on experimental and numerical studies, Zilitinkevich et al. (2019) find that the dissipation rate depends on the stratification and propose a stability-dependent parametrization. Another example of nontraditional scaling of the dissipation rate is observed in turbulence produced at multiple scales by a space-filling fractal grid. Laizet & Vassilicos (2011) provide evidence that this kind of homogeneous isotropic turbulence decays characterized by a single length scale. It is unclear how to model the dissipation rate in the stochastic formulation of the mixing length for the SBL, especially at critical Ri numbers, where the distribution of the value ϕ is non-Gaussian.

Another important detail is the necessity to adjust the units of time between the Eq. (5.4) and the rest of the model. Recall that the Eq. 5.4 is estimated from data based on a sampling frequency with units cycles/hour. Hence the unit of the parameters is hours. The constant $\tau_h = 3600$ in Eq. (5.4) transforms the units of the equation into seconds. Consider that due to $\mathbb{E}(dW(t))^2 = dt$, the process $dW(t)$ has the units of $\sqrt{\text{time}}$ (Horsthemke, 1984), and hence the transformation of units for the noise (stochastic) term is different than in the drift (deterministic) term. The detailed formulation of the stochastic process ϕ is discussed next.

5.1.2 Stochastic Hybrid Formulation of the Turbulent Mixing Length

The turbulent mixing length for momentum is constructed using a mixture of deterministic and stochastic formalism. Equation (5.4) predicts the correction value ϕ from the surface up to some chosen height z_s . After which $z > z_s$ the traditional scaling function $\phi_f(Ri)$ is in operation taken from (Cuxart et al., 2006):

$$\phi_f(Ri) = 1 + 12Ri \quad \text{for } z > z_s. \quad (5.14)$$

The parametrization of Eq. (5.4) is based on the height captured by the measurement tower, which is equal to 30 m. Therefore, with a slightly larger cut-off height, $z_s = 50$ m is set to consider the following blending zone between the deterministic (ϕ_f) and stochastic correction value (ϕ). Finally, both descriptions are joined utilizing the logistic sigmoid function:

$$\text{sig}(z) = \frac{1}{1 + \exp(-k_s(z - z_s))}, \quad (5.15)$$

where z_s is the sigmoid's midpoint, and $k_s = 0.1$ is the steepness of the curve, which regulates the sharpness of transition from ϕ to ϕ_f at the height z_s . Then the linear-convex composite is defined:

$$\varphi(t, Ri) = \phi_f(Ri)\text{sig}(z) + \phi(t, Ri)(1 - \text{sig}(z)), \quad (5.16)$$

and is inserted into (5.9). Due to the stochasticity of φ the mixing length l_m and hence the entire turbulence closure becomes stochastic. It is necessary to analyze how the expected value $\mathbb{E}(l_m)$ and the variance $\mathbb{V}(l_m)$ varies as a function of the Ri number and the geostrophic wind u_g . By estimating these measures, one attains a general idea of how the mixed formulation operates concerning the two main forcing variables.

Following the work of Elandt-Johnson et al. (1980); Lee & Forthofer (2006); Stuart & Ord (1994) the expected value (\mathbb{E}) of a quotient of two random variables (X and Y) is determined with a first-order TAYLOR series approximation

$$\mathbb{E}(X/Y) \approx \mathbb{E}(X)/\mathbb{E}(Y). \quad (5.17)$$

The variance of a quotient of two random variables as a first-order approximation is:

$$\mathbb{V}(X/Y) \approx \frac{(\mathbb{E}(X))^2}{(\mathbb{E}(Y))^2} \left[\frac{\mathbb{V}(X)}{(\mathbb{E}(X))^2} - 2\frac{\text{Cov}(X, Y)}{\mathbb{E}(X)\mathbb{E}(Y)} + \frac{\mathbb{V}(Y)}{(\mathbb{E}(Y))^2} \right], \quad (5.18)$$

One obtains the following estimates for the mixing length by applying the formulas (5.17), (5.18) to (5.9):

$$\mathbb{E}(l_m) \approx \frac{\kappa z}{\mathbb{E}(\varphi) + \frac{\kappa z}{\lambda_b}}, \quad (5.19)$$

$$\mathbb{V}(l_m) \approx \frac{(\kappa z)^2}{\left[\mathbb{E}(\varphi) + \frac{\kappa z}{\lambda_b}\right]^4} \mathbb{V}(\varphi). \quad (5.20)$$

Figure 5.1 shows the values of $\mathbb{E}(l_m)$ (left column), and $\mathbb{V}(l_m)$ (right column) computed exemplary for three heights. To comprehend the physically possible values of $\mathbb{E}(l_m)$ and $\mathbb{V}(l_m)$,

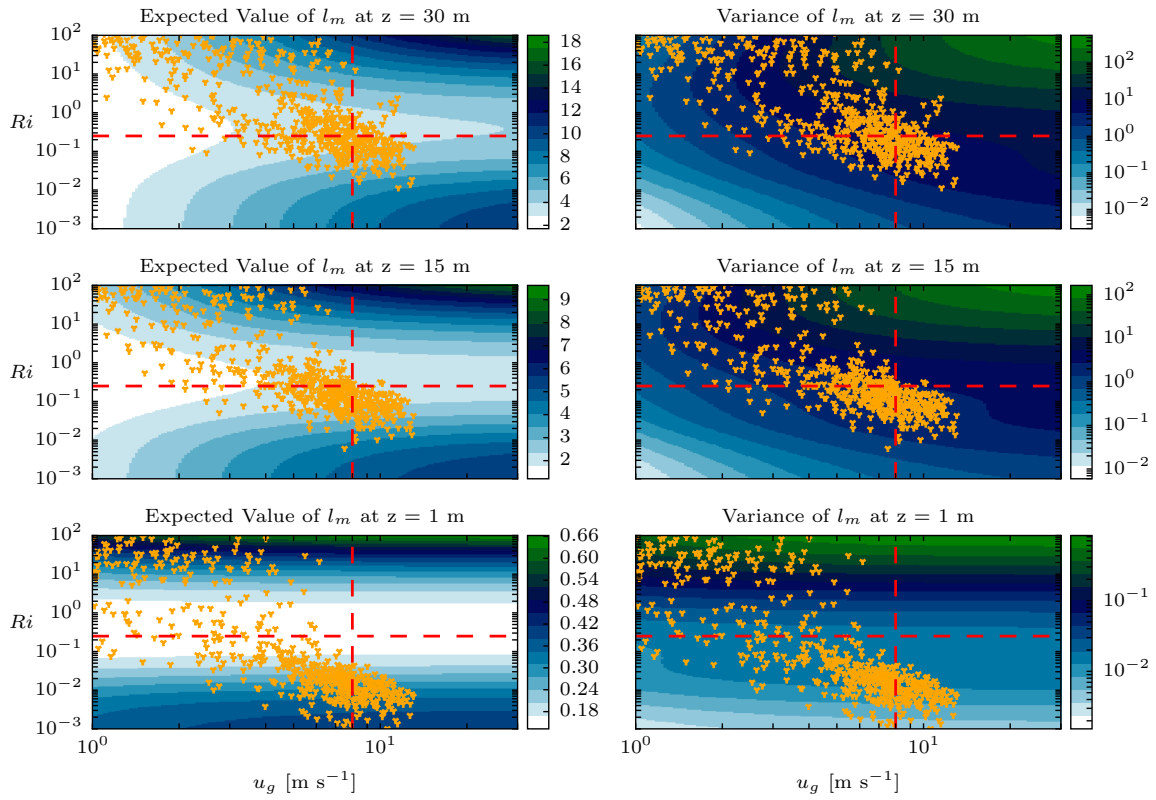


Figure 5.1: The expected value (left panels, calculated with Eq. (5.19)) and variance (right panels, calculated with Eq. (5.20)) of the stochastic mixed length for the three measured heights 30, 15, 1 m as a function of Ri number and u_g . The orange dots show observed l_m values for (Ri, u_g) to capture physically meaningful ranges. The proposed SURANS model should not operate outside the point cloud. The points visualize physically plausible regions for the variable l_m . The red dashed horizontal line is $Ri_c = 0.25$. The red dashed vertical line is a proposed critical $u_g = 9$ m s $^{-1}$. Above this critical wind, the SSE should operate in the lower right quadrant.

the orange coded points represent observed values at (Ri, u_g) in the FLOSS2 dataset. The value u_g for the data points is selected at the tower height 30 m.

Figure 5.1 also suggests according to the model that at low wind speeds and overcritical values of the Ri number, turbulent mixing is persevered. The overcritical values of the Ri number represent a state of the boundary layer where the energy of the mean wind becomes approximately equal to the energy of the sub-mesoscale motions (see Sec. 3.3.7). The separation between the sub-mesoscale motions and the mean wind is non-trivial. Hence the computation of the Ri is sensible to the choice of the temporal averaging scale.

The horizontal dashed line represents the Ri_c number, and the vertical dashed line represents some suggested $u_g = 9 \text{ ms}^{-1}$ which separates the wind limit after which the SBL is weakly stratified. The dashed lines separate the phase-space into four quadrants. The primary remark is focused on the upper right quadrant, which corresponds to high wind and high stability. The orange data points suggest that there are no observations with such characteristics. This is not surprising as at high values of the geostrophic wind, the heat transport is enhanced and the temperature profile becomes mixed relatively rapidly.

The present stochastic formulation shows an undesired performance in the upper right quadrant region. The value of $\mathbb{V}(l_m)$ (see Fig. 5.1 right column) is boundless in that quadrant and continues to grow with increasing Ri and u_g . This is somehow contradicting the physical understanding. It is expected that the $\mathbb{V}(l_m)$ takes some limiting value because l_m becomes fully linked to the scale of the geostrophic wind. The undesired increase in variance is discovered after identifying Eq. (5.4) in Ch. 4. This analysis suggests that the noise term in (5.4) should additionally be dependent on the value of the geostrophic wind. Consider that the scales of motions are always related in the physical domain. Especially in the neutral condition, the assumption that the sub-mesoscale motions are independent of the mean scale does not hold. This performance is qualitatively unveiled in the numerical study (see Sec. 5.3.3). Therefore, the noise term in the SSE requires an additional constrain to limit the noise intensity by the Re number or some critical mean wind speed value. This aspect is not analyzed in this work and future work should incorporate the value of the geostrophic wind into the parametrization of the noise intensity.

5.1.3 Initial and Boundary Conditions

Appropriated initial conditions for Eqs. (5.1) – (5.6) are required to initialize the boundary-layer state at the start of the simulation. Depending on the contour of the initial profiles, the solver's time to reach the quasi-stationary solution varies. To decrease this time, the profiles of neutral conditions are imposed. Under the assumption that the coordinate system is aligned with the geostrophic wind vector and applying the law of the wall, the initial turbulent profiles for one velocity component is:

$$u(z, t = 0) = \frac{u_{*,\text{init}}}{\kappa} \ln(z/z_0), \quad (5.21)$$

$$v(z, t = 0) = 0, \quad (5.22)$$

where $u_{*,\text{init}} = (0.5 C_f u_g^2)^{1/2}$. Here $C_f \approx 4 \times 10^{-4}$ is a tuning parameter and is adjusted such that at the domain top, the value u_g is justified. An appropriated profile of the v -component

is challenging to define. Hence this profile is set to zero since its magnitude is smaller than the one of the u -component. The initial profile for e is estimated following Parente et al. (2011):

$$e(z, t = 0) = a_1 \ln(z) + a_2 \quad (5.23)$$

The coefficients a_1 and a_2 are estimated using the following boundary values:

$$e(z = z_0, t = 0) = u_{*,\text{init}}^2 (0.087)^{-1/2}, \quad (5.24)$$

$$e(z = H, t = 0) = 0, \quad (5.25)$$

where H is the domain height. The initial profile of the potential temperature is constant $T_0 = 300$ K up to a certain height $H_c = 200$ m and then increases according to the atmospheric lapse rate $\Gamma = 0.01$ K m⁻¹ as used by Sorbjan (2012):

$$T(z, t = 0) = \begin{cases} T_0, & \text{for } z \leq H_c, \\ T_0 + \Gamma z, & \text{for } z \geq H_c. \end{cases} \quad (5.26)$$

The boundary conditions are discussed next. For the wind components no-slip conditions (DIRICHLET condition) are set at the surface. At the top boundary, the vertical gradients are set to zero (NEUMANN condition):

$$u(z = z_0) = v(z = z_0) = 0, \quad (5.27)$$

$$\frac{\partial u}{\partial z}(z = H) = \frac{\partial v}{\partial z}(z = H) = 0. \quad (5.28)$$

For the TKE and temperature, the boundary conditions are set following Želi et al. (2019):

$$e(z = z_0) = C_{fm}^{-1/2} u_{*,z0}^2 \quad (5.29)$$

$$e(z = H) = 0, \quad (5.30)$$

where H is defined large enough, such that the boundary layer height fits within the domain and $C_{fm} = 0.087$ is chosen following Želi et al. (2019). The value $u_{*,z0}$ is estimated according to the logarithmic law:

$$u_{*,z0} = \frac{\kappa}{\ln(z_1/z_0)} \sqrt{u(z_1)^2 + v(z_1)^2}, \quad (5.31)$$

where $z_1 = z_0 + \Delta t$ is the next grid point. The upper boundary condition of the potential temperature is dictated by the lapse rate:

$$\frac{\partial T}{\partial z}(z = H) = \Gamma. \quad (5.32)$$

The potential temperature at the surface $T_g = T(z = z_0)$ results from the surface energy balance and is modeled with (5.6). The term responsible for the coupling of the surface temperature and the turbulent diffusion of heat is the surface sensible-heat flux H_0 . It is

defined following Želi et al. (2019):

$$H_0 = \rho_a c_P \overline{w'T'}_{z_0} \frac{Pr_{t,z_0}}{\kappa} \ln \frac{z_0}{z_{0h}}, \quad (5.33)$$

where ρ_a is the air density, c_P the specific heat of air at constant pressure. The turbulent PRANDTL number at the surface is set constant $Pr_{t,z_0} = 0.75$ following Želi et al. (2019). The roughness length for heat is a sensible variable $z_{0h} = 0.1z_0$ and is set following the work of Sanz Rodrigo et al. (2017). This roughness length is known to vary with the wind speed and to depend on vegetation (Brutsaert, 1982; Malhi, 1996). The heat flux at the surface is

$$\overline{w'T'}_{z_0} = -K_h(z_0) \left(\frac{\partial T}{\partial z} \right)_{z_0}. \quad (5.34)$$

The initial conditions presented are used to determine valid steady-state solutions for further simulations. The solver simulates neutral stratification over ten hours to obtain quasi-stationary profiles with the initial conditions presented above. This solution is then used as the initial condition for subsequent studies to analyze different stratification and sub-mesoscale intensities.

5.1.4 Inertial Oscillations

The nudging term in Eq. (5.1) and Eq. (5.2) has an important purpose in the considered boundary layer model. Absence of the nudging terms leads to unwanted inertial oscillations. These undamped oscillations occur in the velocity components around the equilibrium wind vector with a period of $2\pi/f_c$. The amplitude of the oscillations equals the difference between the magnitudes of the initial wind profile vector and the imposed forcing by the geostrophic wind vector (u_g, v_g) (Van de Wiel et al., 2010). Since the initial profiles substantially differ from the wanted wind profiles in stable stratification, the oscillations in the present implementation show amplitudes that cannot be ignored. Furthermore, their undamped nature makes it difficult to obtain a steady, equilibrated wind profile even after > 50 hours simulation. The oscillations develop as soon the geostrophic vector is changed during the simulation time, even though one may start with an adequate initial profile. The absence of turbulent diffusion in the outer layer appears to be rarely observed in nature (Drüe & Heinemann, 2007). The absence of damping in the model hampers the numerical study. Why do those oscillations occur, and how physically correct is the phenomenon?

The inertial oscillations emerge from the model formulation, but it is more due to the over-simplistic representation of the true atmosphere rather than the modeling assumptions. The full system of equations (see Eqs. (5.1) - (5.6)) can be understood to be centered on the turbulent diffusion. Under strongly stable stratification, the boundary layer height shrinks and becomes smaller than the resolved domain height. In such a scenario, the solution above the boundary layer shows diminishing vertical gradients of the wind and a turbulent diffusion coefficient close to zero. Consequently, the solution above the boundary layer is only governed by the CORIOLIS terms (see Eq. (5.1) and Eq. (5.2)), resulting in an undamped oscillatory system. Limiting the turbulent diffusion by the value of the molecular diffusion did not resolve the issue.

Another important aspect worth considering is that the nighttime (12 hour) is often too short to encounter an entire oscillation period (Baas et al., 2012). The TKE can be produced through the variation of the geostrophic wind speed or the inertial oscillations. The interpretation of the results is then difficult with the newly developed stochastic extension of the model. To diminish misinterpretations, nudging terms are applied. Their relaxation time is set to five hours, such that the solver has enough time to develop its own dynamics and still be forcible with a varying geostrophic wind vector. The nudging strategy offers a good compromise, such that the desired study can be performed with a relatively simple model setup.

Nevertheless, inertial oscillations are physically likely. Baas et al. (2012) finds these oscillations in the Cabauw measurement site and demonstrates that the idealized RANS model correctly represents those, at least almost one cycle with an, unfortunately, large case-to-case variability. According to the study above, the conditions needed to observe such oscillations are rather infrequent. Baas et al. (2012) found only 28 nights within a time period of ten years of continuous observations.

The oscillation events are also related to the low-level jet problem (Wang et al., 2007). Although the characteristic nocturnal low-level jet is known to be associated with the inertial oscillations (Davis, 2000) one should also keep in mind that low-level jets induced by other phenomena (Banta, 2008; Lundquist, 2003; Ostdiek & Blumen, 1997) are also observed in the ABL. These several causes make it therefore difficult to differentiate the associated shear-induced turbulence within isolated numerical studies. For instance internal oscillations were also found by Ansorge (2017) in a DNS of the stably stratified boundary layer at high Re and Ri numbers.

5.2 Numerics

This section treats the implementation and calibration of the stochastic single-column model.

5.2.1 Discretization with the Finite Element Method

Equations 5.1, 5.2, 5.3, and 5.5 are discretized and solved utilizing the FEM library FEniCS (Alnæs et al., 2015; Logg et al., 2012). Dunbar et al. (2008) also applied the FEM to simulate the SBL and showed that an adaptive grid refinement approach significantly increases the accuracy of the solution. Nevertheless, the adaptive grid technique is not used here. Instead, a finer grid resolution is set and found to be affordable for the single-column study done here.

The spatial discretization is performed with Lagrangian P1 elements where the error in the H^1 norm scales as $O(\Delta t)$ and in the L^2 norm as $O((\Delta t)^2)$ (Quarteroni, 2017, p. 98). The temporal derivative is discretized with the implicit EULER method. Solver stability and short computation time are prioritized over accuracy because the goal is to investigate the performance of the stochastic extension and analyze ensemble predictions. Nevertheless, P2 elements were tested and the solution did not show any qualitative difference in the shape of the stationary profiles.

The advantage of working with FEniCS is that the library performs the discretization of the nonlinear system and assembles the resulting matrices. For that, the library requires

writing the equations in a weak form, which is explained next. Consider the following short notation for the system of Partial Differential Equations (PDEs) :

$$\frac{\partial \mathbf{v}}{\partial t} = \frac{\partial}{\partial z} \left(K_m \frac{\partial \mathbf{v}}{\partial z} \right) + \mathbf{F}. \quad (5.35)$$

The cumulative variable vector $\mathbf{v} = (u, v, T, e)$ holds the four unknowns. The forcing vector is

$$\mathbf{F} = \left((v - v_g) f_c - N_u(u), -(u - u_g) f_c - N_v(v), P(u, v, T, K_m) - \mathcal{E}(\phi, e), 0 \right), \quad (5.36)$$

The turbulent diffusion coefficient $K_m(\phi, e) = K_h$ depends on the stochastic variable ϕ and the deterministic variable e . The TKE production term consists of

$$P(u, v, T, K_m) = K_m \left[\left(\frac{\partial u}{\partial z} \right)^2 + \left(\frac{\partial v}{\partial z} \right)^2 \right] - \frac{g}{T_0} K_m \frac{\partial T}{\partial z}, \quad (5.37)$$

and is unbounded from below causing numerical instabilities in stable conditions. Therefore the value e is reset to 10^{-4} if negative values appear. The dissipation is

$$\mathcal{E}(\phi, e) = \frac{(\alpha_\varepsilon e)^{3/2}}{l_m(\phi)}. \quad (5.38)$$

Let superscript n denote a quantity at a time t_n . The time-discrete version of the PDE system applying the implicit EULER method is:

$$\frac{\mathbf{v}^{n+1} - \mathbf{v}^n}{\Delta t} = \frac{\partial}{\partial z} \left(K_m(\phi^n, e^{n+1}) \frac{\partial \mathbf{v}^{n+1}}{\partial z} \right) + \mathbf{F}^{n+1} \quad (5.39)$$

where stochastic variable ϕ^n of the turbulent diffusion coefficient and turbulent dissipation is dependent on the current time step t_n and the TKE on the future time step t_{n+1} . In such a formulation, the solution of Eq. (5.39) is segregated from the solution of the stochastic Eq. (5.4). The benefit of a segregated solution is a simple approach to introduce the sampling method (see Sec. 4.6.2) for the spatially correlated noise into the FEniCS framework. Furthermore, the variable ϕ is resolved on a different grid than the other four variables, and the solution is interpolated in between. The stochastic Eq. (5.4) is solved with a fast sampling method, where the grid is required to be equidistant to exploit the efficiency of the fast FOURIER transform (see Sec. 4.6.2). In contrast, the grid for the other four prognostic Eq. (5.39) is non-equidistant and is much finer at the surface than at the top. The objective of this grid is to resolve the gradients. Due to the discrepancies in the numerical grid, it is not possible to construct an entirely coupled system because the stochastic model dictates the grid spacing. Hence, the segregated solution proposed above requires to repeat in each time step the evaluation of the right-hand side of the Eq. (5.4) and Eq. (5.39) iteratively until convergence of the the total system is reached. Such a procedure ensures the coupling between the segregated solutions and demands additional computational overhead. The following Fig. 5.2 investigates the relative error within one time step with respect to the number of coupled solution calls.

For example, consider the evaluation of two sub-iterations for one time step $t_n \rightarrow t_{n+1}$.

1. use $\phi^n, u^n, v^n, T^n, e^n$ as initial values

5. Stochastically Unsteady Reynolds-Averaged Navier–Stokes Equations

2. solve Eq. (5.39) with $\phi^n, u^n, v^n, T^n, e^n$ to obtain: u^1, v^1, T^1, e^1
3. solve Eq. (5.4) using u^1, v^1, T^1 to obtain ϕ^1
4. solve Eq. (5.39) again, but now with $\phi^1, u^n, v^n, T^n, e^n$ to obtain u^2, v^2, T^2, e^2
5. solve Eq. (5.4) using u^2, v^2, T^2 to obtain ϕ^2
6. for more iterations repeat steps 4 and 5
7. exit the sub iterations, where the last iteration provides $\phi^{n+1}, u^{n+1}, v^{n+1}, T^{n+1}, e^{n+1}$

Note that steps 1 to 3 are the natural procedure for obtaining the system update for the next step. It is the following steps 4 and 5 which improve the coupling strength between the deterministic and stochastic equations. However, Fig 5.2 suggests and validates that the system is already sufficiently coupled and additional sub-iterations are unnecessary because the relative error is negligible.

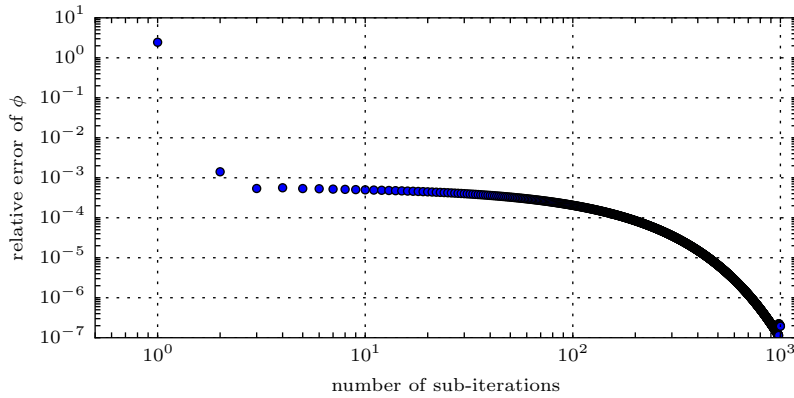


Figure 5.2: Coupling strength of the deterministic and stochastic equations of the SURANS model solved with a segregated approach. Convergence of the relative error for the variable ϕ as a function of the recursive evaluations of the deterministic and stochastic parts of the overall SURANS system. The solution of the SSE is segregated from the solution of the RANS model and may not adapt quickly enough to the update $t + \Delta t$. Low error values indicate that the solutions of the RANS model and the SSE model are in equilibrium with each other within a discrete time step. The first point on the left shows an error relative to the solution of the previous time step. It should be large since it is the time update from Δt . The second point from the left shows the error when the RANS model and the SSE model are updated twice in a sequence recursive, starting with t . The following points represent the relative error, defined as $\|(\phi^i - \phi^{i+1})/\phi^i\|_2$, where i denotes subsequent sub-iterations for a time step Δt .

In figure 5.2, we observe a relative difference of 10^{-3} if we use one additional sub-iteration within a time step. Approximately for the next, 100 sub iteration steps, the relative error stays almost constant and then starts decreasing rapidly to the defined precision of the NEWTON solver, which solves the Eq. (5.39).

Next, one converts the Eq. (5.39) into weak form by multiplying them with the test function vector $\mathbf{w} = (\delta u, \delta v, \delta T, \delta k) \in \hat{V}$ and integrate second derivatives by parts over the domain $\Omega \times [z_0, H]$, where H is the height of the domain is given in meters. The test functions belong to the so-called *test space* \hat{V} and the vector of interest \mathbf{v} belongs to the so-called *trial space* \tilde{V} .

The trial and test spaces are defined as:

$$\tilde{V} = \{\mathbf{v} \in H^1(\Omega)\}, \quad (5.40)$$

$$\hat{V} = \{\mathbf{w} \in H^1(\Omega) : \mathbf{w} = 0 \text{ on } \partial\Omega\}, \quad (5.41)$$

where $H^1(\Omega)$ is the SOBOLEV space and $\partial\Omega$ the boundary of the domain. The weak form of the system reads:

$$\int_{\Omega} \frac{\mathbf{v}^{n+1} - \mathbf{v}^n}{\Delta t} \mathbf{w} \, dz = - \int_{\Omega} K_m(\phi^n, e^{n+1}) \frac{\partial \mathbf{v}^{n+1}}{\partial z} \frac{\partial \mathbf{w}}{\partial z} \, dz + \int_{\Omega} \mathbf{F}^{n+1} \mathbf{w} \, dz + \int_{\partial\Omega} \mathbf{B} \, ds, \quad (5.42)$$

where $\mathbf{B} = (0, 0, 0, K_m \frac{\partial T}{\partial z} \delta T)$ is the NEUMANN boundary condition vector at $z = H$ introducing the atmospheric lapse rate for the temperature equation ($K_m \frac{\partial T}{\partial z} = \Gamma$). Equation 5.42 is solved utilizing the NEWTON method implemented within the FEniCS framework.

Equation 5.6, which solves for the boundary condition of the temperature equation, is discretized with the explicit EULER method in time. The stochastic Eq. (5.4) is discretized with the MILSTEIN method in time (Lord et al., 2014). Next, the numerical grid is addressed.

5.2.2 Numerical Grid

The system of Eqs. (5.1) – (5.6) is solved with two different numerical grids. For the variables u, v, T , and e , a power-three transform on the z -axis is imposed to improve the resolution of the gradients in the vicinity of the surface. Such a non-equidistant grid cannot be used to solve the stochastic Eq. (5.4) due to the sampling algorithm (see Sec. 4.6.2), which utilizes the FOURIER transform. Furthermore, since the stochastic perturbations are included in the lower portion of the boundary layer ($z < 50$ m) the stochastic grid is confined to some height. This limiting of the grid saves computational resources and improves the vertical resolution of the stochastic perturbations.

Figure 5.3 shows the description of the numerical grids along with the computation steps to obtain the hybrid stochastic mixing length correction φ defined by (5.16).

The total domain is organized into three sub-layers, as indicated on the left in Fig. 5.3. The stochastic layer reaches up to the height $z_s = 50$ m. In this sub-domain, the dynamic of the stability correction variable is entirely determined by the SDE (5.4). From the height of z_s up to the height $1.2z_s < z_p < 2.0z_s$, the stochastic fade-out layer is defined. The layer is responsible for the smooth transition from the stochastic to the deterministic value. The transition layer is also responsible for providing sufficient buffer length needed by the sampling algorithm to obtain random structures which do not re-enter the domain at the surface s_0 . (If we do not use a buffer layer, the stochastic structures will re-enter at the surface due to periodic assumptions of the Fourier transform.)

The s grid is per construction bounded to the size of the spatial step of the z grid at the surface (indicated by the Δz_1 in Fig. 5.3). To ensure proper interpolation of the Ri value from the z to s grid the grid step $\Delta s_1 < \Delta z_1$. The value of the stochastic variable ϕ at the time step t_{n+1} is computed on the s grid and then interpolated back to the z grid within the stochastic fade-out layer ($z < z_p$; see Fig. 5.3). The residual values, which correspond to

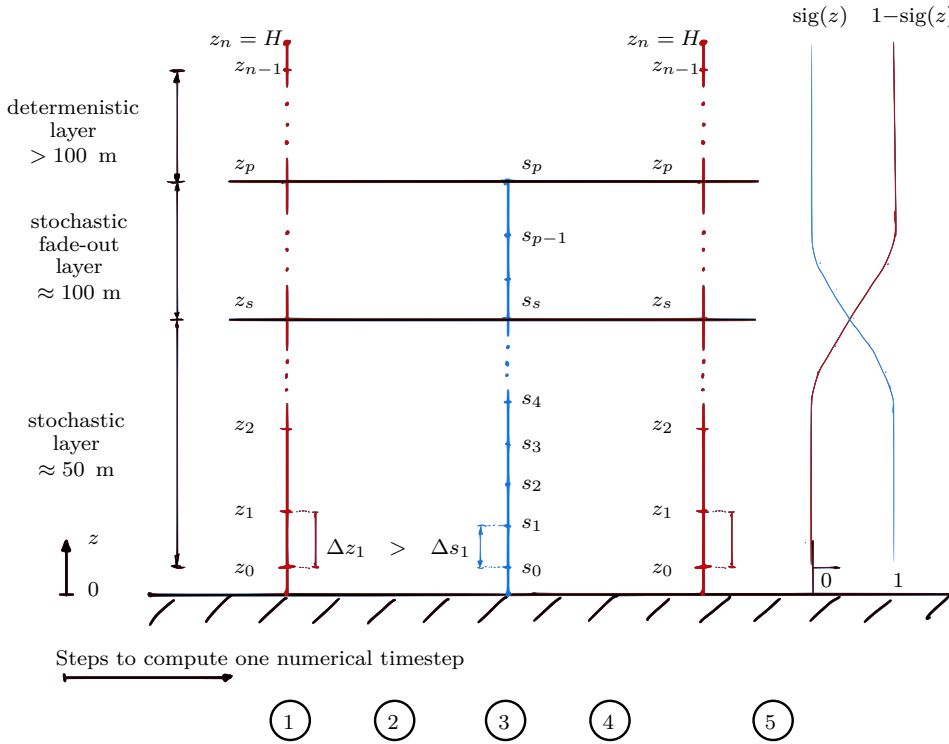


Figure 5.3: The computation of the hybrid stochastic mixed length correction ϕ using two different grids. The z -grid in red is non-equidistant and is used to solve the variables u, v, T, e . The s -grid in blue is equidistant and is used to solve the stochastic variables ϕ . The circled numbers below mark the five steps to calculate the value of ϕ . 1) Calculate the Ri number on the grid z . 2) Interpolate the Ri number on the equidistant s -grid. 3) Evolve the stochastic variable ϕ to the next time step by solving the SSE. 4) Interpolate ϕ to the non-equidistant grid within the height z_p . 5) Compute the linear-convex combination between the deterministic ϕ_f and stochastic ϕ variables on the z grid using the sigmoid function $\text{sig}(z)$ (see Eq. (5.15)).

$z > z_p$, are filled with the value 1. This artificial value is a placeholder and never enters the solution because of subsequent application of the blending function $\text{sig}(z)$.

In the final step 5 a linear-convex combination is performed between the stochastic ϕ and the deterministic ϕ_f variables (see Eq. (5.16); also marked in Fig. 5.3 with the step 5). The height $z_s = 50$ m characterizes the smooth blending between stochastic ϕ and the deterministic ϕ_f variables. Its value is set slightly larger than the measurement tower, which was used to calibrate the stochastic part of the model. Hence, only the lowest 50 m of the simulations have a randomized stability correction in the application of MOST.

5.2.3 Model Calibration

The model calibration is separated into three categories: the parameters of the TKE equation, the parameters of the stochastic perturbation equation, and the one of the soil properties. Almost every parameter value related to the turbulence closure is set according to Rodrigo & Anderson (2013) where different SBL regimes were simulated with a single-column RANS model. Their model was calibrated on the data from the Clean Air Sector of the British Antarctic Survey’s Halley Research Station.

The parameters of the stochastic Eq. (5.4) are entirely based on the estimation from the FLOSS2 dataset described in Ch. 3.1. The calibration of the Eq. (5.4) is described in Ch. 4.

Although the equation has numerous parameters, practically, one parameter is responsible for the perturbation intensity.

The study conducted by Acevedo et al. (2021) is utilized to calibrate the soil parameters, based partially on the FLOSS2 dataset. Acevedo et al. (2021) studied the external controls of the temperature inversion within the SBL and found a linear relationship between the net radiation and the regime transition wind speed. The soil parameters such as the thermal capacity C_g , net radiation R_n , and the soil heat transfer coefficient are tested and slightly adapted for the case studies. Similar parameter values may also be found in (Stull, 1988).

Furthermore, the given value of the dissipation constant by Rodrigo & Anderson (2013) is slightly adapted such that the unperturbed TKE value at the surface is close to the value found in the study of Acevedo et al. (2021). The values of parameters for the entire SURANS model are summarized in Tab. 5.1.

5.3 Numerical Analysis of the Stochastic Turbulence Model

The following study analyzes the implemented SURANS model and studies its differences to the classical 1.5 order closure. Comparison of the model with real observations is not performed for several reasons. Due to the lack of geostrophic wind measurements, reanalysis data needs to be incorporated. Although reanalysis data is accessible, the model is missing cloud parameterizations, for instance. Clouds drastically influence the net radiation and, therefore, the entire boundary layer state. Missing components in the physics parametrization are expected to cause deviations from the observations and require specific period selection and model tuning.

It is important to have a reduced setup at the current model development state to validate the numerical stability first. Furthermore, the outcome of the stochastic perturbations is adequately comprehended with an isolated, low-complexity model.

Several numerical experiments are performed to study the consequence of the perturbations applied to represent the impact of sub-mesoscale motions in comparison to the unperturbed model. The neutral stratification is studied first. This study is a validation case because, per construction, the ensemble mean of the SURANS model must match the RANS model in neutral stratification. Next, the strongly SBL with intermittent mixing is analyzed. The SURANS model reproduces intermittent TKE state. By analyzing this state the ensemble mean is not a representative measure due to non-Gaussian statistics. A more appropriated measure is the central tendency (the most probable value), and its evolution is used to evaluate the performance of the models.

The two studies introduced above are performed for a quasi-stationary case, where the geostrophic forcing and the soil properties are constant in time. However, the stochastic perturbations may also alter the solution under conditions with variable forcing. Therefore, this aspect is analyzed in a separate study. Finally, the consequence of the sub-mesoscale turbulent diffusion on the temperature inversion is analyzed. The results show the response in the inversion shape as a function of the perturbations.

Table 5.1 summarizes the parameter settings of the SURANS model. Empty parameter values indicate that they will be adjusted for each case studies individually.

5. Stochastically Unsteady Reynolds-Averaged Navier–Stokes Equations

Table 5.1: Summary of the settings of the SURANS solver. The parameters marked with ‘-’ are given individually in the following case studies. The symbols in this table are those used in the program code.

Description	Symbol	Value	Source
Total simulation time [h]	T_end_h	–	set
Timestep [s]	dt	–	tuned
Grid resolution (z grid)	Nz	100	set
Roughness length [m]	z0	0.044	Acevedo et al. (2021)
Roughness length for heat [m]	z0h	$z0 \times 0.1$	Sanz Rodrigo et al. (2017)
Domain height [m]	H	300	set
Restoring temperature [K]	theta_m	290	set
Reference pot. Temperature [K]	T_ref	300	set
Air density [kg/m ³]	rho	1.225	set
Air spec. heat capacity [J/kg/K]	C_p	–	set
Soil heat capacity [J/m ² /K]	C_g	as in	Acevedo et al. (2021)
Net radiation	R_n	-60 – 0	Acevedo et al. (2021)
Geostrophic wind [m/s]	U_top	–	set
Geostrophic wind [m/s]	V_top	0	set
Latitude [°N]	latitude	40	FLOSS2 dataset
CORIOLIS parameter [rads/s]	f_c	9.34e-05	FLOSS2 dataset
Atmospheric lapse rate [K/m]	gamma	0.01	Rodrigo & Anderson (2013)
Relaxation time scale [s]	tau	3600×5	tuned
Minimum TKE level [m ² /s ²]	min_tke	10^{-4}	tuned
Turbulent PRANDTL number for BC	Pr_t	0.85	Želi et al. (2019)
Eddy viscosity constant [–]	alpha	0.46	Rodrigo & Anderson (2013)
Dissipation constant [–]	alpha_e	0.1	tuned
Sub-mesoscale intensity [–]	d	-0.07	FLOSS2 dataset (ref)
Stochastic model height [m]	z_l	50	FLOSS2 dataset
Covariance length [m]	lz	20	FLOSS2 dataset
Von Karman’s constant [–]	kappa	0.41	Rodrigo & Anderson (2013)

5.3.1 Neutral Boundary Layer

The neutral boundary layer is the first experiment which validates that the central tendency of the SURANS model is equivalent to the RANS solution. The initial conditions are given in Sec. 5.1.3 and the simulation period is set to 15 hours. The solver specific settings for this

experiment are given in Tab. 5.2 and the rest in Tab. 5.1. The forcing parameters are set to be constant.

Table 5.2: Relevant solver settings for the numerical study of the neutral layer.

Description	Symbol	Value
Total simulation time [h]	T_end_h	15
Time step [s]	dt	10
Grid resolution (z grid)	Nz	100
Domain height [m]	H	300
Restoring temperature [K]	theta_m	300
Reference pot. Temperature [K]	T_ref	300
Net radiation	R_n	0
Geostrophic wind [m/s]	U_top	5

The stratification is controlled with two parameters of the surface balance in Eq. (5.6). The first one is net radiation. One of the effective physical mechanisms behind the term is the cloud cover. To stimulate neutral conditions the net radiation is turned off. The second parameter is the restoring temperature in Eq. (5.6). This is the temperature of the soil below the surface in some finite depth. Depending on the conditions, the restoring temperature is varying slower than the diurnal cycle. Therefore, it is expected that some difference between the soil and air temperature is present after the sunset provoking at least weakly stable boundary layers. Hence, a less realistic condition is imposed by setting the difference between the soil and air temperature to zero, mainly for validation purposes. This condition ensures strictly neutral stratification.

Figure 5.4 shows the comparison of the TKE at three different heights among the model runs with and without the stochastic mixing. A quasi steady-state solution is reached approximately after six hours with the RANS model. The central tendency of the SURANS model, which is estimated from 100 realizations, is nearly identical to the solution of the RANS model. The regularity of the sample paths (indicated with the thin colored lines) varies across the height. More rapid changes are found closer to the surface (sample paths in gray), and smooth oscillations with smaller variances occur at $z = 150$ m (sample path in green). The stochastic mixing length equation is only active up to the height $z = 100$ m. As indicated by the sample paths in green ($z = 170$ m), the variability induced at the surface is propagating into the upper boundary layer. The MOST operates in the the upper boundary layer.

We find that the distributions of the TKE from the 100 SURANS simulations are close to being Gaussian, but more importantly, symmetrical. This symmetry indicates that the modeled type of turbulence is in statistical equilibrium, and the perturbations maintain their path around the central tendency, which is the mean value. The physical interpretation of the stochastic oscillations is that the integral scale of turbulence is draining energy not only from the mean scale (heir one hour and more) but also obtain stochastically added effect from unresolved random mixing events. Since the scale of the perturbation is smaller than the mean

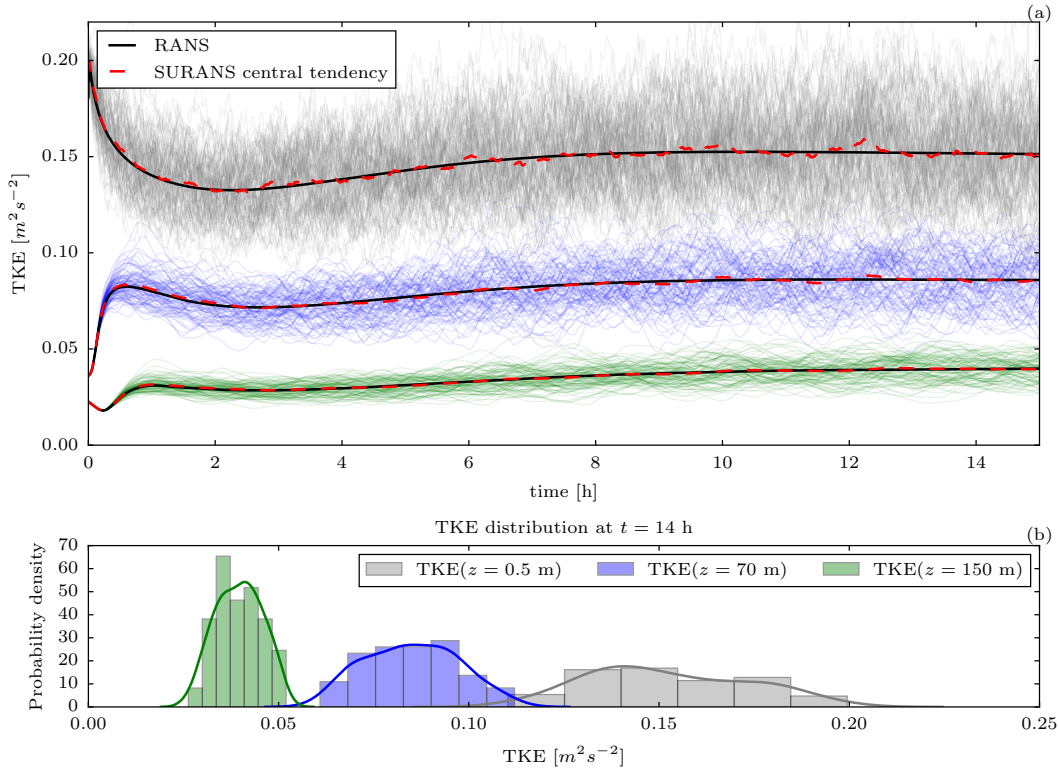


Figure 5.4: Comparison of the predicted TKE by the SURANS and RANS models in the condition of neutral stratification ($Ri = 0$) for three heights ($z = 0.5, 70, 150$ m). The evolution of TKE is shown in (a) and the corresponding color legend is given in (b). Panel (a) shows the RANS solution with a solid black line. The many lines in different colors indicate the 100 realizations of the SURANS model for their heights. The central tendency of the SURANS model is indicated by a dashed red line. The respective probability distribution of the TKE ensemble at $t = 14$ h is given in panel (b).

scale the central tendency coincides with the expected value. The equation of the TKE is in stable statistical equilibrium. As shown later, the equilibrium becomes weaker and more sensible to the perturbations at a larger Ri number leading to intermittency.

5.3.2 Stably Stratified Boundary Layer

By far the most interesting results are associated with the stably stratified boundary layer in the presents of random mixing events. As in the previous section, the initial conditions are given in Sec. 5.1.3, and the simulation period is set to 15 hours. The solver-specific settings for this experiment are given in Tab. 5.3 and the rest in Tab. 5.1. The forcing parameters are set to be constant.

The stratification is imposed with two mechanisms and is likely to be found in real observations. The first mechanism of stratification is the difference between the restoring (soil) temperature of 290 K and the potential temperature of the air 300 K. Such condition arises during the transition from day to night time. Imagine that the air is heated up in the daytime before the night, but the deep soil temperature remains unchanged from the previous cold and snowy week. In addition, the second mechanism of radiative cooling also enhances the stratification. The net radiation of -30 W m^{-2} is selected from (Acevedo et al., 2021) and

Table 5.3: Relevant solver settings for the numerical study of the stably stratified boundary layer.

Description	Symbol	Value
Total simulation time [h]	T_end_h	15
Time step [s]	dt	5
Grid resolution (z grid)	Nz	100
Domain height [m]	H	300
Restoring temperature [K]	theta_m	290
Reference pot. Temperature [K]	T_ref	300
Net radiation	R_n	-30
Geostrophic wind [m/s]	U_top	5

considered as the FLOSS2 dataset average value. This setup may describe a possible cloud-free night in springtime.

Figure 5.5 illustrates the solution of the SURANS model. The TKE at the height $z = 20$ m is compared against the solution of the RANS model using several statistical metrics. The Fig. 5.5a shows in blue a characteristic signature of intermittent TKE simulated with the SURANS model. Thin gray lines display other realizations. Note the two different types of spikes found at $t = 6$ h and $t = 10$ h. Their magnitude is significantly larger than the ensemble mean (solid yellow) and the central tendency (solid red). The duration of these events is approximately one hour and falls within the characteristic range of sub-mesoscale motions.

Consider Fig. 5.5b, which shows the TKE's distribution of 100 model runs at $t = 14$ h. The solid black line represents the prediction of the RANS model for comparison. The solid yellow line is the ensemble mean of the SURANS model, and the solid red line is the central tendency. The central tendency is estimated from the PDF, which is fitted to the histogram by applying the KDE method (Scott, 2015). The estimation is poor because it violates the boundary condition on the left side. Nevertheless, the KDE is a time-efficient method to approximate the most probable value. The histogram indicates a smaller value of the central tendency than the estimated one. A better estimation can be achieved if a specific distribution type is assumed. However, one should keep in mind that the distribution type should depend on the stratification. This dependence makes the fitting task less trivial. Therefore, the study is refrained from using more complex estimation approaches for studying the distributions of the TKE.

The first statistic in Fig. 5.5a for comparison is the ensemble mean, and it is slightly above the RANS prediction. However, the central tendency is significantly smaller and indicates that it is likely to observe an absence of turbulent mixing. The heavy tail in the ensembles distributions is significant and induces sporadic rare events. Some realizations of the model (not shown) predict a low TKE level throughout the entire simulation period. The wide variety of TKE signatures favors the representative capabilities of the model.

Figure 5.6 shows a clearly intermittent profile that has several characteristics. The largest intensity of each burst is found at the surface. Recall that the stochastic correction of the

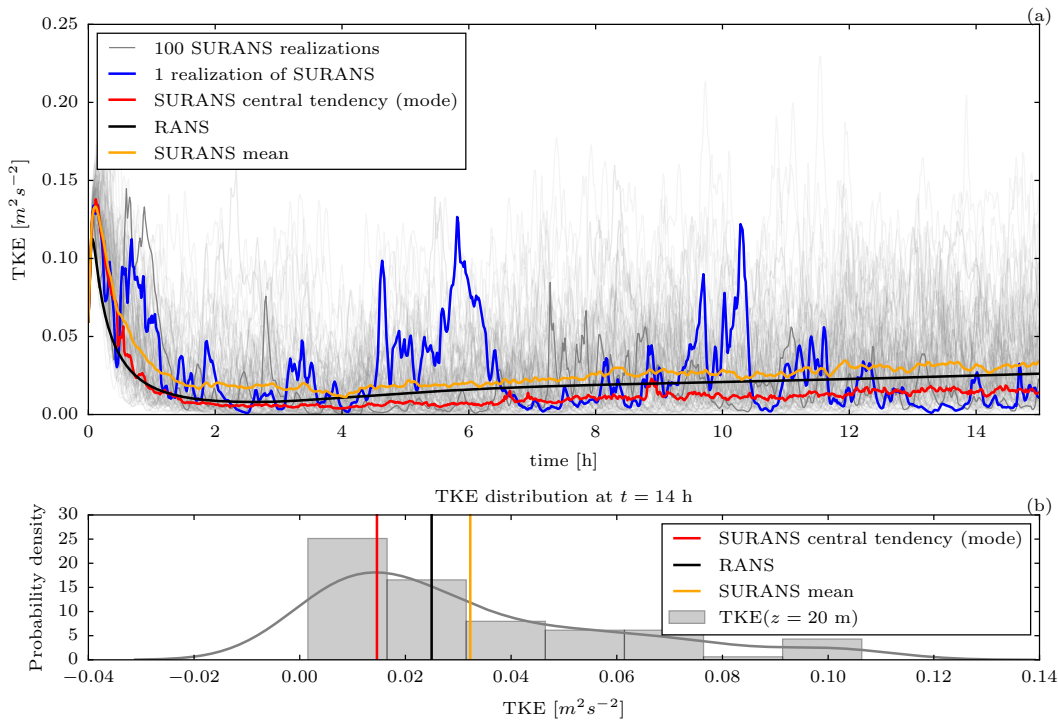


Figure 5.5: Comparison of SURANS and RANS models predicted TKE under the condition of strongly stable stratification (mass $Ri \approx 0.6$) for height $z = 20$ m. For the visualization of the Ri number profiles, see Fig. 5.9. The evolution of the TKE is shown in (a). The ensemble distribution of 100 sample paths of the SURANS model at $t = 14$ h is shown in panel (b) along with the fitted probability density function (solid gray line) using a KDE method. The thin gray lines show the 100 realizations of the SURANS model.

turbulent diffusion is detached from the ground (see Fig. 4.13). Still, we cannot find any turbulent patches that are clearly aloft, which is typical for observations (see Fig. 3.18). The bursts are absent aloft because the turbulent diffusion is multiplied with the gradient of the mean wind, and hence the spatial distribution of the TKE is intrinsically constrained by the wind gradient.

A slight inclination (as somebody brushed it from left to right) in the bursts is also present. Some events show that some turbulence is still maintained aloft from the surface (see Fig. 5.6a $t = 3.5$ h and $t = 5$ h). The model is representing the decoupled TKE from the surface after a burst, but only as a consequence and not as a source. Important to add to this observation is that the flow is forced with a steady mean wind. Changing the forcing changes the gradient and hence provide room for the stochastic perturbations to appear. It is not clear if such scenario represents observational data.

Next, we investigate the capability of the models to transport the temperature (see Fig. 5.7). It is evident that in the case of stochastic perturbations, the mixing is performed faster. The mixing rate is higher, and the temperature inversion is also shifted up and is less abrupt. The temperature profile changes its shape in an unsteady way (compare Fig. 5.7a to 5.7b), which can be related to the activity of the intermittent burst periods (see Fig. 5.6a). The stochastic model shows a qualitatively different solution of the temperature inversion. Considering the humidity (which is excluded from the model), the cloud and fog formation prediction will

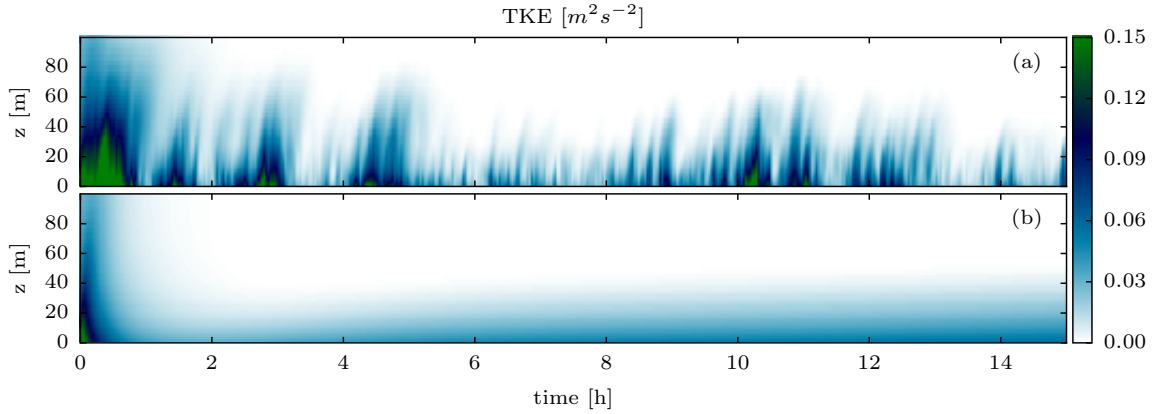


Figure 5.6: Temporal evolution of the profiles of TKE for a realization of the SURANS (a) and RANS (b) models. The color bar applies to both panels.

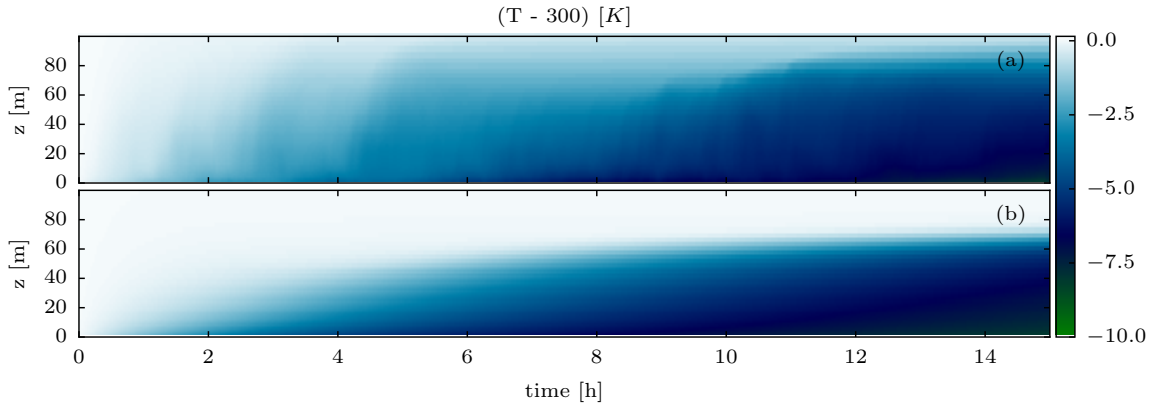


Figure 5.7: Temporal evolution of the temperature profiles for one realization of the SURANS (a) and RANS (b) models. The color bar is valid for both panels.

experience a significant difference from the unperturbed model. A dedicated case study should investigate the performance against observational data.

Finally, we analyze the profiles of the dominant wind velocity component u to comprehend the boundary layer height (see Fig. 5.8). In Fig. 5.8a we can recognize the repeating pattern of the TKE bursts (compare Fig. 5.6a). The dominant stochastic turbulent diffusion dictates the boundary layer shape as a consequence of random mixing events.

Figure 5.9 shows the evolution of the profiles for the Ri number. The SURANS model predicts the strongly unsteady Ri number, but the bulk Ri number is computed for the layer between the z_0 level and $z = 80$ m. Deviations are found during random mixing events when the temperature profile is mixed sporadically, reducing the local bulk Ri number (compare with Fig. 5.7).

5.3.3 Variable Geostrophic Wind and Net Radiation

In previous examples, a stationary forcing of the model is considered. However, predicting real weather conditions implies continuously changing forces. Hence the intention is to investigate if the stochastic perturbations change the solutions during the transient states. Such states are characterized by some rate of change in the forcing variables.

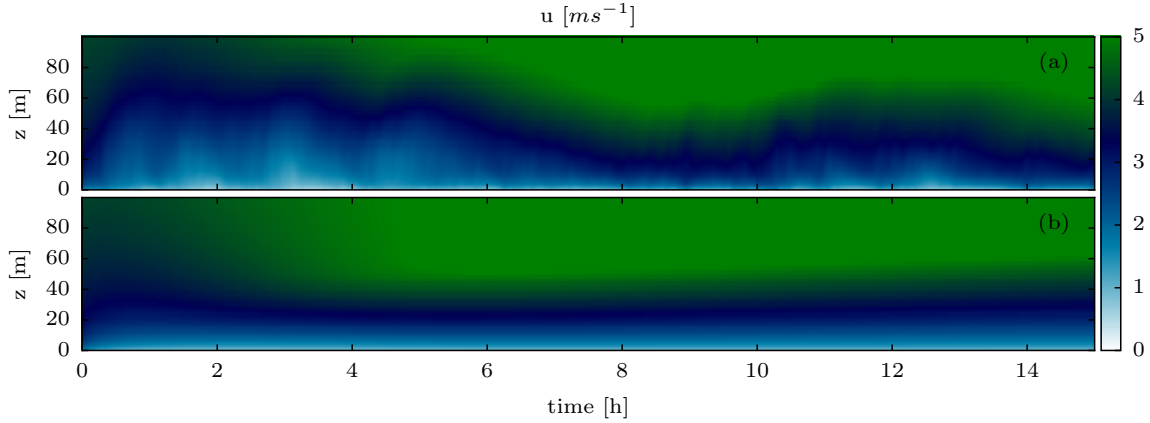


Figure 5.8: Temporal evolution of the wind profiles (u component) for one realization of the SURANS and RANS models. The color bar is valid for both panels.

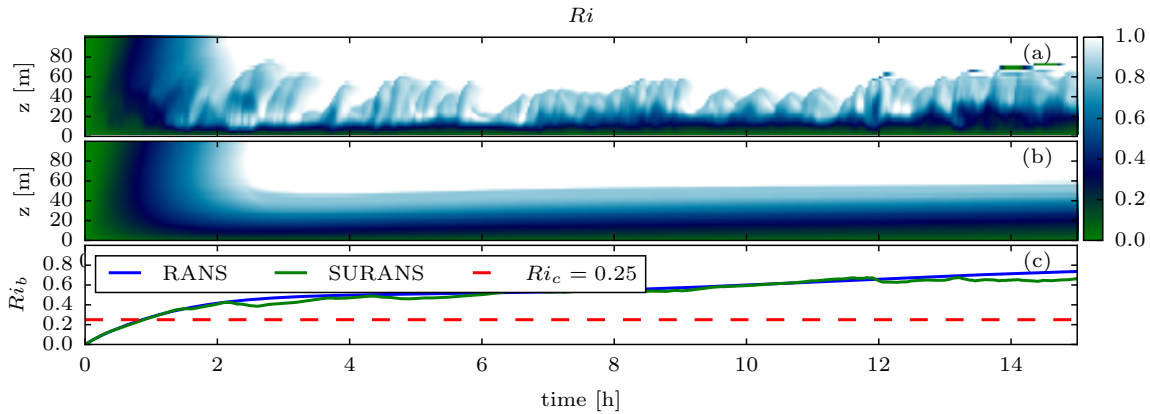


Figure 5.9: Temporal evolution of the Ri number profile for a realization of the SURANS (a) and RANS (b). The color bar applies to panels (a) and (b). Panel (c) shows the bulk Ri number calculated from the surface to $z = 80$ m.

The initial conditions are given in Sec. 5.1.3, and the simulation period is set to 30 hours, which is longer than the average nighttime. In this experiment, the focus lies on computing the transitions between weakly and strongly SBL, as in (Acevedo et al., 2021; Maroneze et al., 2019). The novelty of this study provide predictions of the regime transitions with random sub-grid mixing events.

The setting of the nonstationary forcing is discussed next. Since the nudging time scale is set to five hours, the rate of change in the geostrophic wind is slow. In combination with the change of the radiative cooling, we will cover within one simulation four possible forcing combinations. The solver-specific settings for this experiment are given in Tab. 5.4 and the rest in Tab. 5.1.

To answer some questions, consider Fig. 5.10 first. The forcing of the model is given in Fig. 5.10a, the temporal evolution of the TKE at the height of $z = 9$ m and the profiles for the variables $e, T, U = \sqrt{u^2 + v^2}$ at three different times (note the arrows in Fig. 5.10b). The quantities to compare in Fig. 5.10 are:

- The 100 realizations of the SURANS model (gray, thin lines).

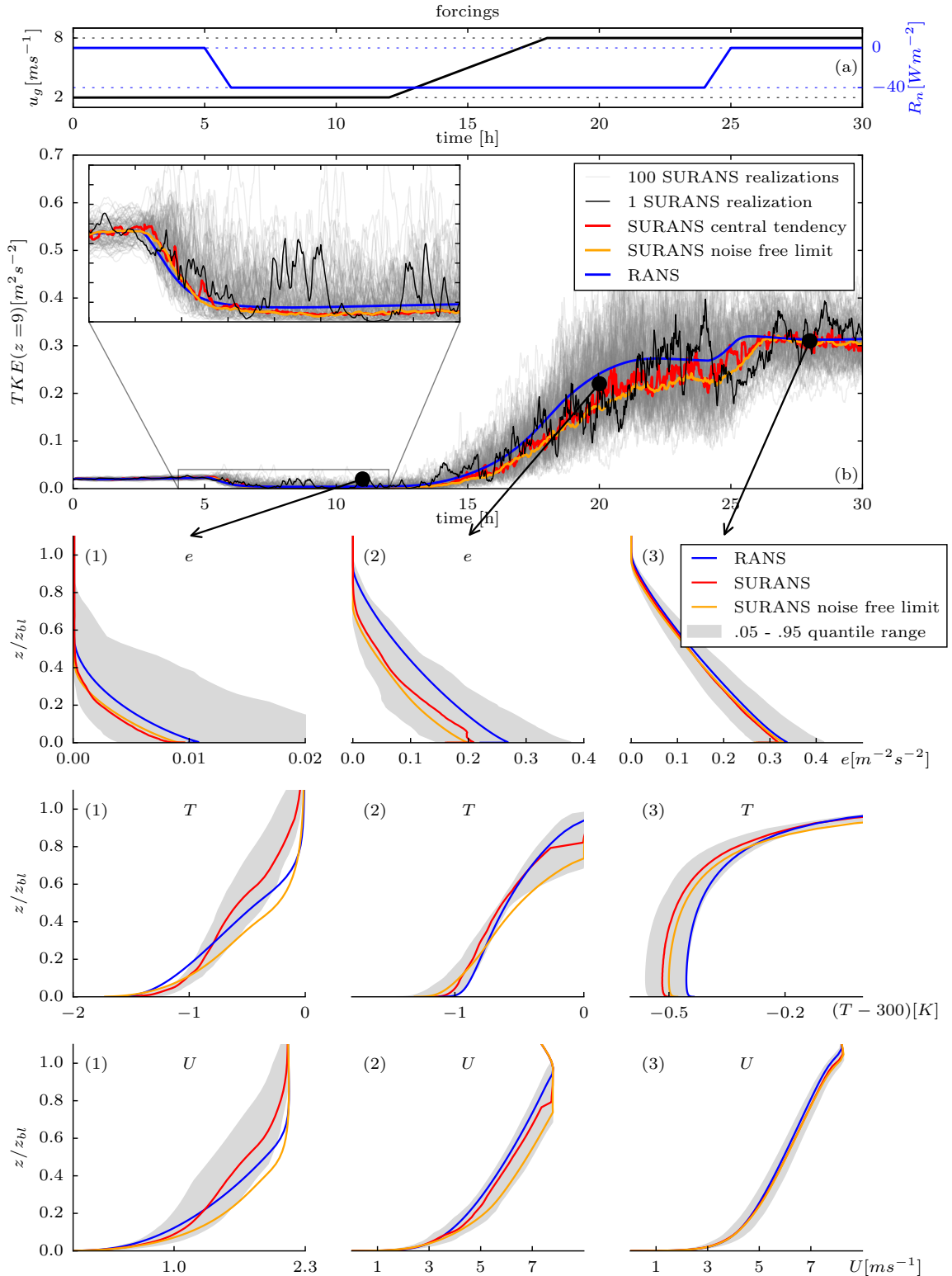


Figure 5.10: Solution of the SURANS model with variable forcing parameters R_n (net radiation) and the geostrophic wind u_g . The total simulation period is 30 hours. The nudging time scale is set to 5 hours. Panel (b) shows the evolution of TKE at 9 m for 100 realizations, marked with gray lines. The zoom area highlights the transition to stable stratification in weak winds by increasing net radiation. The evolution of the forcing is shown in a). The sub-images in b) show profiles of the variables at 3 different times marked with black dots in b). The SURANS profiles represent the central tendency, with the gray area showing the quantile range. The boundary layer height z_{bl} used for normalization is 50, 200, 240 m.

Table 5.4: Relevant solver settings for the numerical study with unsteady forcing variables.

Description	Symbol	Value
Total simulation time [h]	T_end_h	30
Time step [s]	dt	2
Grid resolution (z grid)	Nz	100
Domain height [m]	H	300
Restoring temperature [K]	theta_m	300
Reference pot. Temperature [K]	T_ref	300
Net radiation	R_n	variable (see Fig. 5.10a)
Geostrophic wind [m/s]	U_top	variable (see Fig. 5.10a)

- The central tendency (solid red), estimated as the most probable value from the fitted distribution (see Fig. 5.5).
- The noise-free limit of the SURANS model. In this case, the stochastic equation is solved once with a sufficiently low value of the noise, such that the dynamical evolution can be considered deterministic (solid yellow line). The noise-free limit is introduced to eliminate the effect of the difference between the MOST stability function and the deterministic steady-state of the prognostic Eq. (5.4) (the expected value). One realization of the SURANS model is provided to comprehend the rare events during the stable low-wind conditions (solid black line).
- The prediction of the RANS model (solid blue line).

In the following, we cover some questions to learn more about the stochastic model. First of all, what is the resulting difference in the central tendencies due to perturbations? For that, we compare a solution of SURANS in the noise-free limit with the central tendency estimated from the 100 noise-present realizations (see yellow and red lines in Fig. 5.10). There are no significant differences in terms of the considered quantities for the TKE (see 5.10b and the corresponding profiles). However, there is a substantially different temperature and velocity profile. The temperature is transported more effectively within the stable condition and surprisingly above the average boundary layer height (see Fig. 5.10 panel (1)). The central tendency of the velocity profile experiences a deceleration. The unsteady profile shape is, in fact, the consequence of the solution and is not a side effect of grid interpolation. The blending height between the stochastic and deterministic quantity is 50 m above the analyzed profiles in Fig. 5.10 panels (1). This means that for higher geostrophic winds, the perturbation of the turbulent diffusion is propagated to the top of the boundary layer (200 m), although the actual perturbations are limited at 50 m.

What is the difference between the central tendencies of the SURANS and the RANS models? Asking this question, we are confronting the differences under the circumstance of the parametrization effort. This issue is related to the approach one decides to follow. The stochastic approach is more complex than fitting a function to the data (MOST) but is

accurate and predicts the uncertainties. By comparing the RANS solution (blue line) to the central tendency (red line) predicted by the SURANS model, we find that RANS is predicting higher levels of TKE throughout the entire simulation. In comparison, the SURANS model is predicting lower levels of TKE but controversially more enhanced transport of temperature and velocity (see Fig. 5.10 panels (1)). This nontrivial effect may result from non-equilibrium statistics in the stochastic formulation of the turbulent mixing length. Consider the next question.

Is there something unique in the uncertainty of the prognostic variable's profile? In Figure 5.10, the lower panels show the profiles, and the gray area represents the 0.05 – 0.95 quantile range constraining the variability of the 100 different model runs. Thereby the central tendency is given by the solid red line. For stable stratification (see Fig. 5.10 panels (1)), the quantile range for the TKE is asymmetrical, showing the largest spread at the surface. In neutral conditions, the quantile range is symmetrical (see Fig. 5.10 panels (3)). The shape of the gray area (the uncertainty) and its change with the height in the TKE are significantly different than in the temperature and velocity profiles. The largest uncertainty is found in the middle of the boundary layer, and its value decreases at the surface and the boundary layer top.

What is the noise impact on the transient states? The influence of the noise is manifested in terms of the possible solution paths which the solver predicts (see Fig. 5.10b thin gray lines). Following the inset in Fig. 5.10b, we observe a transition from neutral to stable stratification induced by the increase in net radiation. The central tendency and the noise-free limit of the SURANS model overlap. However, multiple individual realizations (thin gray lines) show a pronounced tendency to lag the transition rather than induce early transition. In contrast, by transitioning from low wind to high wind (see Fig. 5.10b from $t = 15$ to $t = 20$), the solution path can show both fast and delayed transitions.

Is the variance of the TKE represented adequately? The variance is bounded, responds to the stratification, and to the geostrophic wind speed change. Figure 5.10b demonstrates one skeptical point, as u_g increases, the variance in the TKE increases as well. This relation is plausible since the largest turbulent eddies develop with the value of geostrophic wind. To what extent the variance should grow is not bounded by the stratification only. The proposed model does not have any mechanisms that link the stochastic perturbations' variance with the large scale. Physically there is a connection because of the sub-mesoscale dependence on a large scale. Figure 5.10b demonstrates that the variance reduces to some lower value as soon the stratification is applied (see $t = 25$ h). The results indicate that under high stratification and high wind, the model experience unusual variance. The reason for this is the parametrization of the noise term in the stochastic equation. Besides the Ri number, a dependence on the geostrophic wind speed should be considered.

What is the relative difference between the SURANS and RANS solutions? This question is interesting because one would prefer to identify the areas where a significant change is expected. In this question, it is important to differentiate the following two cases.

Given the data of the correction variable ϕ_f , what is the difference between the parametrization strategies in terms of the profiles (see Fig. 5.11)? We would like to compare the usual parameterization approach (MOST), which leads to the RANS solution against the stochastic approach, which fits the nonstationary SDE instead and leads to the SURANS model.

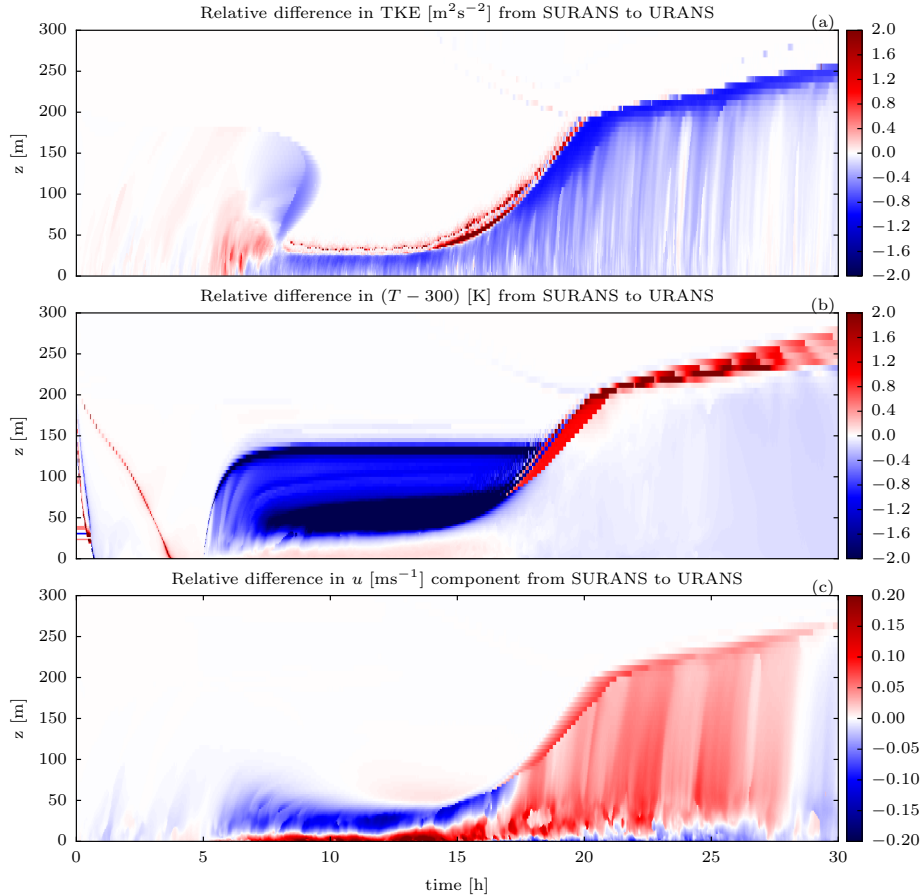


Figure 5.11: The relative difference in profiles between the SURANS (the central tendency of 100 realizations) and the RANS model according to the numerical study in Fig. 5.10. The red color denotes the area where the variables of the SURANS model have larger magnitude than those of the RANS model. The white color denotes no differences. Panel (a) shows the TKE, (b) the temperature ($-2 [K] < (T - 300)[K] < 0 [K]$), and (c) the dominant u component of the wind. The forcing variables change with time and are shown in Fig. 5.10a. Condition of stable stratification for $t \in (5, 15)$ h and condition of high wind pressure for $t \in (5, 15)$ h.

The SURANS model provides an adequate representation of the intermittent turbulence than the RANS model, because it also takes into account the effect of the unresolved small scale mixing processes that are independent of the mean wind at high Ri number.

Figure 5.11a shows the relative difference in the TKE. The relative abrupt transition from blue to white color indicates approximately the boundary layer height. The boundary layer grows after $t = 15$ as the geostrophic wind is increased. For the time $t > 25$ h the net radiation is turned off, and the central tendency of the SURANS model becomes equal to the RANS solution. For the time $t > 6$ h the value of the TKE predicted by the SURANS model is 50% smaller than predicted by RANS on average, indicating a shallower boundary layer (see Fig. 5.10 the TKE profiles for additional comprehension)

Figure 5.11b shows the relative difference in the temperature. Within the boundary layer (this is the region between the x-axis and the bright blue and red spots) the differences in the models are insignificant. Contrary, the SURANS model shows a deviation from the RANS solution at the boundary layer top. For the stable conditions ($t \in (5, 15)$ in Fig. 5.11b), the central tendency of the SURANS solution predicts almost a 200% lower value of the

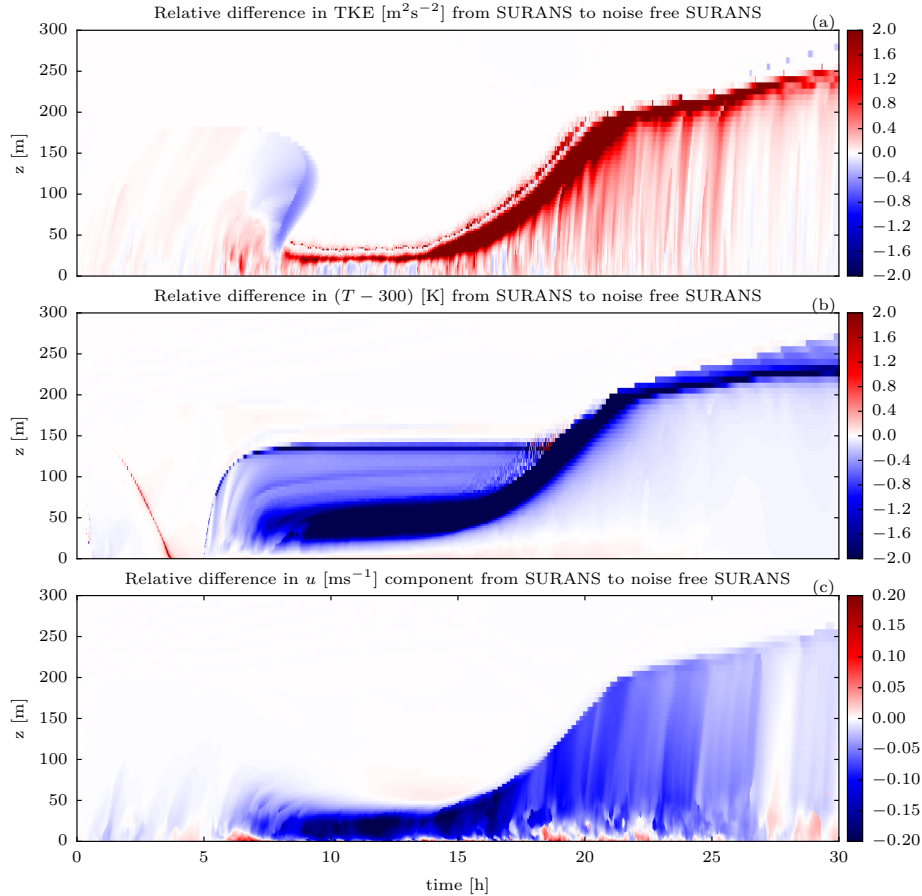


Figure 5.12: The relative difference between noise-free limit SURANS ($\sigma_s \approx -1$) and the noise present SURANS models (central tendency of 100 realizations) simulation. These differences represent the pure effect of the noise on the central tendencies and are invariable to the choice of the stability function. In this example $-2 [K] < (T - 300)[K] < 0 [K]$. Panel (a) shows the TKE, (b) the temperature and (c) the dominant u component of the wind. The forcing variables are changing in time and can be found in Fig. 5.10a.

temperature than the RANS model for a large area above the boundary layer (see the blue area in Fig. 5.11b). At the same time, the differences at the surface are relatively small. This can be explained by the enhanced transport due to intermittent turbulence. Recall, per construction, the stochastic perturbations start to decay for $z > 50$ m. At the same time, the boundary layer height is approximately 25 m. Such that the stochastic mixing determines the boundary layer height. This observation can also be found in Fig. 5.10(1) in the profiles of TKE. In that figure, the uncertainty (gray color) is above the boundary layer height.

The red area at the top of the boundary layer in Fig. 5.11b for $t > 17$ h (the high-wind regime) means that the central tendency of the SURANS model is predicting an increased value of the relative temperature. Hence, the errors produced in the stable regime ($5 < t < 15$ h) are propagated into the high-wind regime ($t > 20$ h) at the boundary layer top. This findings suggest that the different transport of temperature and moisture (not in this model) may impact the creation of clouds in the early morning with large geostrophic winds.

One would also like to know how the noise influences the solution and under which circumstances the differences emerge? To answer this question, we cannot compare the SURANS and RANS models because steady-state solution of the SDE does not exactly match

the static, linear parametrization function $\phi_f(Ri)$. Mainly because the steady-state of the SDE is implicitly determined from the time series dynamics and the function $\phi_f(Ri)$ from a scatter plot.

For that reason, Fig. 5.12 compares the noise-free limit of the SURANS model with the central tendency of the noise-present cases. The results indicate that the largest differences in the TKE and temperature appears above the boundary layer. The internal profile of the temperature (see Fig. 5.12b) is mostly unchanged, where the internal profile of the TKE (see Fig. 5.12a) shows deviations on a finer scale. The velocity profile (see Fig. 5.12c) shows that the added noise reduces the magnitude almost in the entire boundary layer. Patchy, intermittent accelerations can be detected near the surface.

In summary, the noise perturbations induce large differences at the boundary layer top. Parametrization of the intermittent diffusion by using a SDE is costly in term of number of parameters but shows significantly different profile characteristics. One of the model’s known biases is the warm temperature at the surface (see Sec. 1.1.4). Particularly for climate models, the first boundary layer node, representing the surface, may be at the height of $z = 50$ due to coarse resolution. In such a case, the shallow boundary layer lies within and is unresolved. As the comparison shows (see Fig. 5.12b and Fig. 5.11b), the SURANS model predicts a relative temperature which is by 200% smaller than the traditional RANS model. These results demonstrate that the new model can reduce the warm bias in an isolated setting.

In the next section, we investigate the stochastic impact of the sub-mesoscale turbulent diffusion on the temperature inversion.

5.3.4 Diffusion-Induced Transitions of Temperature Inversion

The temperature inversion at the surface is studied in terms of a wind-dependent temperature difference (Van de Wiel et al., 2017). This difference is typically constructed by using temperature values at two different location close to the surface and plotting it against the geostrophic wind (van der Linden et al., 2017) or the internal boundary layer velocity (Baas et al., 2019). Although the variables used to study the a temperature inversion in each study undergoes slight variations a prominent characteristic is common to all of them.

Based on the inversion strength, there exist two regimes. A strong temperature inversion regime occurs for low winds and substantial radiative cooling. High values of geostrophic wind speeds characterize the second regime. The increased turbulent transport resists the radiative cooling, and a weak inversion is formed. The sharpness of the transition between the mentioned regimes depends on multiple factors such as thermal properties of the soil, turbulent mixing, and the sub-mesoscale intensity, as will be discussed next.

Baas et al. (2019) identified the dominating processes that drive the transitions and the inversion strength. The authors find that the major formation of the inversion is induced mainly by the change in geostrophic wind speed and by the large-scale subsidence in a specific field study. Furthermore, from the analysis of the momentum and temperature budgets, the authors find that the heat flux divergence plays an important role in the transition from the regime of large to small temperature inversion. This transition is called erosion. As highlighted above, large-scale processes control the formation or erosion of the temperature inversion. It is, however, not known, if the random mixing events from non-turbulent processes

can induce regime transitions. Abraham et al. (2019) showed that while the onset of the vary SBL can be stochastically predicted, the erosions during a night happens apparently randomly, suggesting relevance of local processes This aspect is analyzed next with the help of the developed SURANS model.

In the following experiment, we analyze the inversion in four ways, where the turbulent properties differ. By changing the turbulent properties of the heat transfer towards (we analyze only stable conditions), the surface will differ and hence resist the radiative cooling depending on the intensity of the turbulence. The solver-specific settings for this experiment are given in Tab. 5.5 and the rest in Tab. 5.1. The initial conditions for this experiment are constructed by running the model for ten hours in neutral conditions. The profiles at the final time step form the initial conditions.

Table 5.5: Relevant solver settings for numerical study of the temperature inversion.

Description	Symbol	Value
Total simulation time [h]	T_end_h	5
Time step [s]	dt	1
Grid resolution (z grid)	Nz	200
Domain height [m]	H	800
Restoring temperature [K]	theta_m	290
Reference pot. Temperature [K]	T_ref	300
Net radiation	R_n	-55
Geostrophic wind [m/s]	U_top	variable (see Fig. 5.13)

The idea of the experiment is to run 100 realizations starting from neutral conditions by imposing strong stratification. Each realization experience various stochastic perturbations of the turbulent diffusion due to random sub-mesoscale activity. For each fixed geostrophic wind, 100 realizations of the SURANS models are constructed to investigate temperature inversion. After five hours of simulation time, the temperature inversion will be different in each realization due to the stochastic nature of the sub-mesoscale motions. By constructing a scatter plot from all realizations, we discover the possible transition diagrams as a function of sub-mesoscale intensity and geostrophic wind speed (see Fig. 5.13).

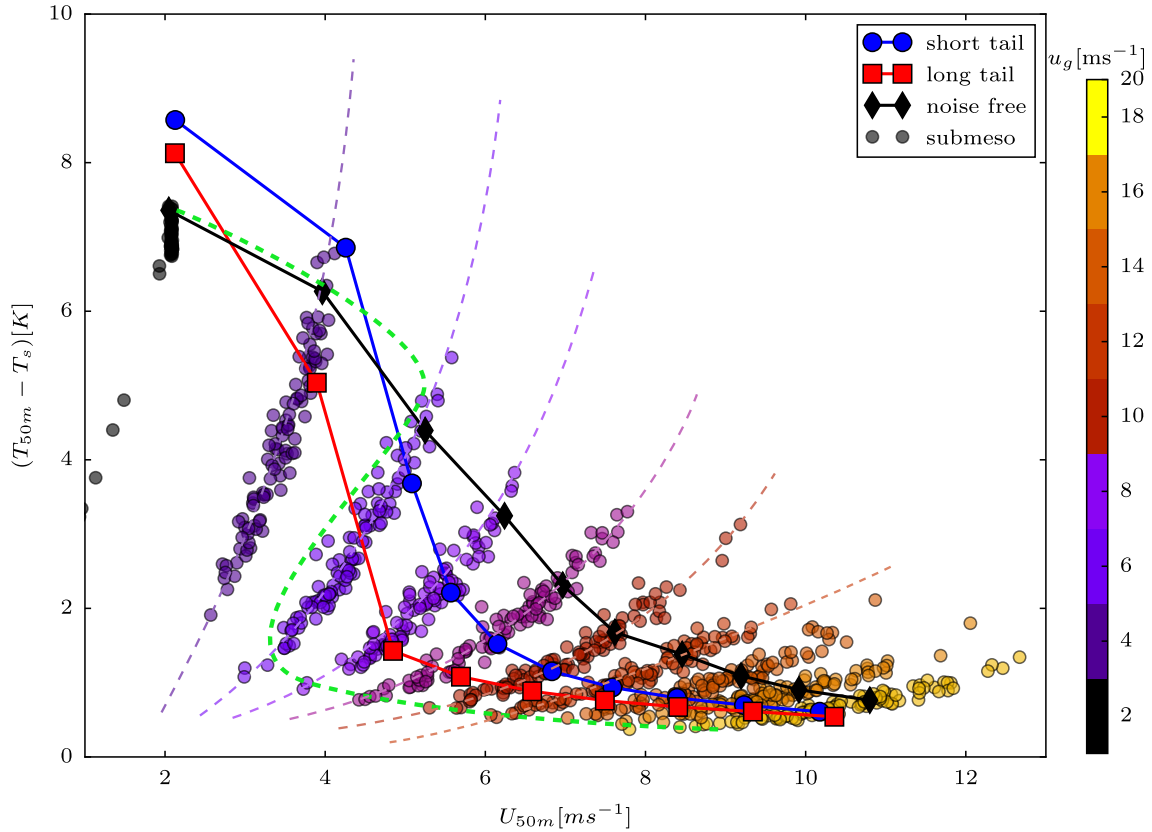


Figure 5.13: Modeled relationship between 50-m temperature inversion and wind speed under the influence of assumed sub-mesoscale turbulent diffusion and various stability functions. The temperature difference on the y axis is constructed using the surface temperature T_s measured at the height z_0 . The legend entries "short tail" ($\phi = 1 + 12Ri$) and "long tail" ($\phi = 1 + 4Ri$) denote the temperature inversion generated by the RANS model. For comparison, the "noise-free" mark denotes the solution of the SURANS model, where the intensity of the noise is set to a negligible value to be considered deterministic ($\sigma_s = -1$). The scatter points denote the SURANS solutions corresponding to the noise intensity found in the FLOSS2 dataset ($\sigma_s = 0$). The color of the points encodes the geostrophic wind value. The model SURANS is evaluated 100 times to produce the scatter plot. Each point is another realization of the noise. The colored dashed lines mark solutions with constant u_g but a different realization value of the Wiener process. For high u_g values, the inversion is independent of the noise because the colored stripes are horizontally aligned. For low values of u_g , the strength of the inversion is determined by the local realization of the noise process. A characteristic S-curve (Van de Wiel et al., 2017) in the current simulation setting can only be imagined if certain scattering points, corresponding to different noise realizations, are selectively plotted (as indicated by the dashed green line).

Figure 5.13 shows the temperature inversion produced with several configurations of the turbulence model. The modification of the turbulence is realized by changing the stability correction function in the RANS model and varying the noise intensity in the SURANS model. The red squares and the red line mark the long-tail model, meaning more turbulence is maintained at a higher Ri number. The blue dots and the blue line mark the short-tail model, where the turbulence is suppressed rapidly with increasing Ri number. The black diamonds and the black line show the noise-free solution of the SURANS model, meaning that the level of the noise is significantly low, such that the SURANS solution is considered almost deterministic. The scattered data points represent the multitude of noise-induced solutions of the SURANS model. The color bar labels the value of the geostrophic wind forcing.

The change in the curves (see Fig. 5.13) from long-tail to short-tail model is the increased temperature inversion at the critical wind speed of 5 ms^{-1} . Such an increase is expected because reducing the turbulence level reduces the resistance of the radiative cooling (see Eq. 5.6). If we compare the rate at which the temperature transition is happening at the critical wind speed of 5 ms^{-1} , we find that it is almost unchanged in the long-tail and short-tail models. The difference between the long-tail and short-tail models is the slope of the stability function. Hence, this slope is not responsible for the rate $\frac{d\Delta T}{dU_{50m}}(U_{50m} = 5)$. Regarding this aspect, the noise-free SURANS model provides a different solution. The temperature inversion is weaker and smoother. The critical wind speed is less pronounced because the rate $\frac{d\Delta T}{dU_{50m}}$ at which the black line (see Fig. 5.13) changes is smaller in comparison to the short-tail model. The differences between the noise-free SURANS and the RANS model are due to the implicitly determined nonlinear scaling function of the stability correction variable (see the red solid line in Fig. 4.11). This nonlinear function is implicit because it results from the estimation of the SDE from the time series. In contrast, the short-tail is directly fitted to the scatter plot. The estimation of the SDE makes a more reliable representation because the approach captures the dynamical properties of the variable ϕ .

In the analysis of the temperature inversion with the SURANS model, the starting point is the noise-free solutions (see black line in Fig. 5.13). In addition, the colored scattered points represent the various noise-induced solutions. We find that the solutions manifest a specific organizational pattern in the form of strips. The start of the stripes is marked with black diamonds, which symbolize the noise-free limit. Starting from the noise-free limit, the increase of the noise intensity is associated with the direction of the stripes towards the lower values of ΔT . The path from the noise-free limit towards reducing the temperature inversion is associated with the increased level of sub-mesoscale activity and goes through the respective values of the RANS solution (relate the black squares and blue circles in Fig. 5.13). The sub-mesoscale activity assumingly induces turbulent mixing and hence resists radiative cooling. This observation coincides with the findings that a long-tail model is reducing the temperature inversion.

A more interesting finding, which supports the conceptual validity of the SURANS model is the inclination of the scatter point stripes. As the geostrophic wind increases, the strips take a more horizontal orientation. This means that along each stripe, the change in the temperature inversion with respect to U_{50m} reduces. This means that the temperature inversion is insensible to the sub-mesoscale perturbations at high geostrophic wind speeds. This makes sense since large geostrophic winds dictate the formation of the boundary layer. In contrary, at low geostrophic winds, the strips take almost vertical positions, indicating that the temperature inversion is highly responsive to the perturbation of the turbulent mixing.

In the presented setting above (see Fig. 5.13), a critical wind speed may be defined as the value at which the largest variability $\mathbb{V}(\Delta T)$ and $\mathbb{V}(U_{50m})$ in both variables arises (\mathbb{V} denotes variance). Due to the perturbations, the simultaneous increase in both variables' variability indicates a sensible response of the entire turbulence model. The two contrasting regimes are then defined as $\mathbb{V}(\Delta T) > \mathbb{V}(U_{50m})$ for the regime of strong temperature inversion and $\mathbb{V}(\Delta T) < \mathbb{V}(U_{50m})$ for the regime of weak temperature inversion. For the considered example, the critical geostrophic wind speed, estimated from the scatter plot, is 6 ms^{-1} .

5.4 Conclusions

This chapter provides the formulation and numerical analysis of a SURANS model for a stably stratified atmospheric boundary layer. To improve the representation of intermittent turbulence in regimes of low geostrophic wind speeds and increased sub-mesoscale activity. The new finding is summarized in the following points.

- The SURANS model adequately represents the intermittent TKE patterns within the regime of strong stratification (see Fig. 5.6).
- The SURANS model shows enhanced mixing properties in comparison to a RANS with a linear stability correction function (see Fig. 5.7). The temperature profile is mixed faster and reaches over larger heights. In comparison to the RANS solution, the stochastic model, predicts lower temperature value just above the shallow, stably stratified boundary layer (see Fig 5.12).
- The SURANS model shows that in the regime of stable stratification the boundary layer height is unsteady and is dictated by the stochastic events (see Fig. 5.8).
- The stochastic perturbation in the SURANS model is parameterized based on the local gradient Ri number. Therefore, the regime-dependent intermittent statistical properties of the TKE are captured continuously as a function of flow stability (see Fig. 5.10).
- The analysis results suggest that in the condition of weak stratification and large geostrophic wind speeds the SURANS model appears to show disproportionately enhanced variance (see Fig. 5.10a). Therefore, besides parameterizing the noise term in Eq. (5.4) with the Ri number one needs to investigate and constrain the variance of the noise with the value of geostrophic wind speed.
- The stochastic equation is confined to the surface portion of the boundary layer and needs to be blended with the deterministic part of the system. It is unknown if the proposed closure is locally valid in the outer boundary layer.
- In comparison with the RANS model, the SURANS model show significant differences in the solution, which is located at the boundary layer top (see Fig. 5.11).
- The SURANS model allowed to investigate the solution states of the temperature inversion diagram induced by stochastic turbulent diffusion Sec. 5.3.4).
- Although the stochastic correction variable show detached mixing events from the surface, the intermittency in the turbulence remains bounded to the surface for the considered case studies. This is because the origin of the intermittent bursts in the TKE are always constrained by the gradient of the mean wind due to multiplication. Stochastic diffusion will act in places where a wind gradient is present. Hence, the SURANS is able to simulate detached TKE if the gradient of the mean wind is detached from the surface as well.

- The effect of stochastic diffusion reaches beyond the boundary layer height as well as beyond the limiting height of the perturbations (see Fig. 5.7). This results in qualitatively different profiles compared to the RANS solutions in the outer boundary layer. Furthermore, the boundary layer height becomes highly variable in strongly SBL and is determined by the random turbulent mixing events.

The presented SURANS model shows the potential to be used as an exploratory or even predictive tool. To investigate the use of the SSE for less idealized setups, future studies should validate the performance of the SURANS against observational data.

6

Conclusion

This work presents a REYNOLDS-averaged NAVIER–STOKES (RANS) type stochastic turbulence closure developed based on the Fluxes Over Snow Surfaces Phase II (FLOSS2) dataset for simulating the surface atmospheric Stable Boundary Layer (SBL) up to a height of 50 m. The turbulence closure covers a stability range $0 < Ri < 10$ and represents intermittent turbulent mixing events. The intermittency of turbulence beyond the critical $Ri_c = 0.25$ number is assumed to be caused by physical processes occurring at scales unresolved by the RANS model. These unresolved, so-called sub-mesoscale motions determine turbulent mixing in conditions of high stratification and low mean wind speed. The non-turbulent sub-mesoscale motions initiate at the model’s subgrid-scale and are independent of the geostrophic wind forcing for $Ri > Ri_c$. Therefore, small-scale turbulence resulting from sub-mesoscales is assumed to occur randomly in a RANS model. The following sections provide a detailed summary and future research directions.

6.1 Summary

Parameterization of the SBL is essential for the Numerical Weather Prediction (NWP) systems. The NWP models depend on the processes at the surface responsible for the transport of scalar quantities such as temperature and moisture. Turbulence, which is the most important transport phenomenon in the surface Atmospheric Boundary Layer (ABL), links the large-scale flow of the upper atmosphere with properties of the Earth surface. Turbulence modeling is challenging especially in the SBL, partly because of the influence of unresolved sub-mesoscale motions.

The sub-mesoscale motions summarize significant processes in the nonstationary SBL. Analysis of the nonstationarity and scale interactions in the FLOSS2 dataset shows that the intermittency of the turbulent scales is related to the increase in sub-mesoscale energy. The relationship between these two scales of motion is investigated utilizing GRANGER causality. GRANGER causality evaluates whether the sub-mesoscale can effectively predict the Turbulence Kinetic Energy (TKE) in the strongly SBL by exploiting the statistical memory effects in the time-series dynamics. The study shows that vertical gradients of sub-mesoscale wind speed

are a better predictor of TKE than sub-mesoscale wind speed alone. The modeling results suggest that there is remaining residual variability in the TKE that is unpredicted by the vertical gradient of the sub-mesoscale wind. Therefore, the modeling results suggest that other turbulence-generating mechanisms such as gravity wave breaking or rotational shear are important.

In this work, the SBL is classified in different ways to analyze the nonstationary scale interactions. Usually, the SBL is classified by limiting a stability parameter (e.g., Ri , ζ). This approach is compared with the performance of classifying SBL with statistical models. Clustering with statistical models allows to include hypothetical relationships for the analyzed variables. Therefore, such classification provides a tool for discovering novel approaches to analyze the nonstationary SBL. This study shows that model-based statistical classification identifies an intermittent regime. This regime is characterized by the Turbulence-Submeso-Intensity (TSI) diagram, which indicates the relative contributions of sub-mesoscale and turbulence intensity. The TSI diagram shows a weakly bimodal distribution representative of an intermittent boundary layer, while the stability parameters do not reveal a similar pattern (see Fig. 3.26). The reason for this difference is that the stability parameters lack the scale interactions between the sub-mesoscale motions and the TKE. The observed differences in the performance of the regime classifiers suggest that the sub-mesoscales are essential in the strongly SBL. The stability parameter should account for the effect of the sub-mesoscale processes to improve the RANS models.

The classified cospectra of the scale activities, are investigated as a function of stability to conclude the relevance of the sub-mesoscales in the SBL. The classified flux cospectra (see Fig. 3.10) highlights dissimilarities between the transport properties of the sub-mesoscales and the KOLMOGOROV-like turbulence. The small-scale turbulence provides sustained averaged transport, while the sub-mesoscale exhibits zero average transport over time. With increasing stability of the flow, the scale variability in the flux cospectra for the sub-mesoscale range increases. In combination with low-wind conditions, the production of small-scale turbulence becomes more related to the sub-mesoscale than to the mean wind. Chapter 3 details the above mentioned findings, where the main tool to obtain them is the Finite-Element, Bounded-Variation, Vector AutoRegressive (X) Factor (FEM-BV-VARX) clustering method.

The FEM-BV-VARX method is used in the first step of this work to analyze the SBL. In the next step, the FEM-BV-VARX method is further developed to the Finite-Element, H^1 -Regularization, Stochastic Differential Equation (FEM- H^1 -SDE) variant. The novelty here is the combination of an existing FEM- H^1 -variational clustering method with an existing Maximum Likelihood Estimation (MLE) approach for nonlinear Stochastic Differential Equations (SDEs). The FEM- H^1 -SDE method models the nonstationary turbulence measurements with nonlinear one-dimensional SDEs. The method provides an option to work with generally defined SDEs with appropriate nonlinear drift and nonlinear diffusion terms. Including estimation of the time-dependent parameters of nonlinear diffusion makes the FEM- H^1 -SDE method more suitable for parameterization approaches in turbulence than the FEM-BV-VARX method, since the FEM- H^1 -SDE method explicitly models the noise of a system. In addition, the FEM- H^1 -SDE method uses a continuous-type mathematical model, which makes it possible to transfer the identified models from data into other continuous

systems, being a RANS model one of them. In contrast, the noise of a discrete-time system must be modeled a posteriori in the FEM-BV-VARX method. The disadvantage of the FEM- H^1 -SDE method is that it is limited to one dimension and does not take into account the memory effect. The FEM-BV-VARX method models memory effects and works with multiple dimensions, so this tool is better suited for exploratory data analysis.

The main objective of this work is fulfilled by utilizing the FEM- H^1 -SDE method in modeling turbulence measurements of the FLOSS2 dataset. Data-driven parameterization of intermittent turbulence in atmospheric SBL is achieved by modeling the nondimensional stability correction variable ϕ . In a RANS closure, the variable ϕ modifies the turbulent mixing length of the momentum to suppress turbulent mixing with increasing flow stability according to the MONIN–OBUKHOV Similarity Theory (MOST). In this work, the variable ϕ is considered as a time-dependent stochastic process that reflects the variety of mixing events, including those, assumingly, induced by the sub-mesoscale motions. The model structure for the stochastic process ϕ is postulated following the scaling of the MOST (see Sec. 4.2). The model structure for ϕ is further developed into a SDE using a parameter perturbation approach. Data analysis is performed to justify the parameter perturbation, which suggests that the sporadic turbulent mixing events perturb the ϕ system through the Ri number. The resulting SDE for modeling the evolution of ϕ is referred to as Stochastic Stability Equation (SSE). The proposed SSE calibration is investigated on the FLOSS2 dataset using the FEM- H^1 -SDE method along with other model alternatives of varying complexity. Examination of the multiple models showed that the proposed SSE performs better than the alternatives and, most importantly, provides the ability to parameterize the nonstationary parameter values with the vertically local value of the gradient Ri number. The discovered scaling functions (see Fig. 4.8) provide the closed-form expression for the stochastic process ϕ , so that the SSE can be directly integrated into an atmospheric RANS model.

The advantage of using SSE to model intermittent SBL is that the effect of random mixing events is controlled by one parameter: noise intensity. Since the noise intensity scales with the flow stability, the respective scaling function can be shifted to enhance or eliminate the random perturbations. By completely suppressing the noise intensity, the steady-state of the SSE converges to the MOST scaling for the sub-critical range $0 < Ri < 0.25$. Such a property originates purely from the identified scaling function using the FEM- H^1 -SDE method and justifies the initial choice of model structure, yet there may be better alternatives. For the overcritical regime $0.25 < Ri < 10$, the SSE in the limit of zero perturbations shows on average low turbulent mixing, which is a consequence of the parameter calibration procedure with the FEM- H^1 -SDE method. Simulating the variable ϕ in the $0.25 < Ri < 10$ regime with perturbations of finite intensity shows intermittent dynamic. The parameterized SSE provides a measure to quantify the intermittency of the flow. The proposed SSE induces a Probability Density Function (PDF) for ϕ as a function of flow stability. With increasing flow stability, the type of the PDF transforms from a Gaussian distribution to an extreme value distribution. Together with the transformation of the PDF's shape the moments become a function of stability. With increasing stability the expected value of the PDF diverges from the most probable value of the PDF. The divergence of the expected value from the most likely value is a measure of global intermittency for the stratified flow (see Fig. 4.11).

The SSE is combined with a general atmospheric RANS model to investigate and validate the utility of the SSE. The FLOSS2 dataset is not used to validate the stochastic RANS model directly. The numerical studies yielded a reasonable qualitative representation of the intermittent dynamics of the TKE in the strongly stable regime (see Fig. 5.5). Moreover, the study demonstrated that the parametrization of the noise term in the SSE requires additional parameter to properly model the weakly stable regimes. By construction of the SSE, the perturbations are independent of the mean wind forcing scale. This modeling assumption is true for the strongly stable condition and invalid for the weakly stable condition. In the weakly stable condition, the inertial subrange becomes larger than the operational time scale of the SSE. This means that the inertial subrange overlaps with the sub-mesoscales so that the KOLMOGOROV turbulence dominates. The resulting drawback of the SSE parameterization is the unconstrained variance in condition of a large geostrophic wind paired with weakly stable stratification. This shortcoming is highlighted in Fig. 5.10. It can be remedied by constraining the noise term in SSE with a limiting geostrophic wind value. Such a value can be approximated from the flux cospectras, where the scale activity is tracked. Additionally, the limiting geostrophic wind value is directly readable from Fig. 5.1.

6.2 Outlook

Future research directions may include several topics including: further development of the FEM- H^1 -SDE clustering method, investigation of the scaling functions for the SSE, further extension of the stochastic parameterization to include the turbulent mixing length for heat, and inclusion of the regime-dependent integral scale of turbulence in the SSE.

The FEM- H^1 -SDE is expandable to classify spatially varying parameters of a model. Details of the corresponding literature is provided in Sec. 2.7. A practical first goal would be to identify vertically varying parameters of the SSE. This becomes relevant when considering boundary layers in complex terrain because the scaling coordinate of the variable l_m is no longer linear and may depend on local topography. Therefore, the local topography changes the turbulent patterns and can lead to different spatial scaling laws. These laws can be discovered by learning the spatial parameter variability from data. In the present work, this aspect is eliminated because, for a flat surface, the vertical scaling is known ($l_m = \kappa z$) and is directly incorporated into the preprocessing step of the variable ϕ so that FEM- H^1 -SDE method does not need to discover it.

The parameter scaling functions of SSE in this work are proposed based on the FLOSS2 dataset. More data needs to be examined to validate the discovered scaling. It is also of interest to investigate the general validity of the parameter scaling. Since the parameter scaling depends on the model structure, it is beneficial to look for a model structure with two time-dependent parameters by simplifying the SSE structure further. In addition, any work dealing with the derivation of the model structure using first principle approaches would be constructive.

Another interesting research direction is the development of a 2D system of SSEs that combines the effects of the kinematic eddy-viscosity (K_m) and eddy heat conductivity (K_h) coefficients under the effect of sub-mesoscales. By using an additional SSE for the K_h coefficient

in a coupled system with the K_m coefficient, a stochastic model could be obtained for the Pr_t as a function of random mixing events in the subgrid-scale for the conditions $Ri > Ri_c$.

The regularity of SSE solution and thus memory is not investigated in this study. The identification of memory effects can be integrated into the FEM- H^1 -SDE framework. This aspect is useful in identifying continuous systems with memory for the variable ϕ . The variable ϕ will require memory if a regime-dependent turbulence averaging scale is utilized to model the stochastic process ϕ . The regime-dependent integral scale of turbulence can be calculated directly from the results of this study. The estimated model affiliation function can be used in combination with Multiresolution Decomposition (MRD) to estimate the time-varying turbulent averaging scale.

Finally, it would be interesting to study the consequences of the SSE in a larger NWP system. The numerical study (see Sec. 5.3.3) has shown that the random perturbations within the boundary layer induce the most important changes at the top of the boundary layer. Moreover, the induced differences at the top of the boundary layer are propagated into the weakly stable regime in case of increasing wind forcing (see Fig. 5.12). Hypothetically, this could influence cloud formation in the morning hours. The effects of SSE in a NWP system are inconclusive without further research.

References

- Abraham, C., Holdsworth, A. M., & Monahan, A. H. (2019). A prototype stochastic parameterization of regime behaviour in the stably stratified atmospheric boundary layer. *Nonlinear Processes in Geophysics*, *26*(4), 401–427.
- Abraham, C., & Monahan, A. H. (2020). Spatial Dependence of Stably Stratified Nocturnal Boundary-Layer Regimes in Complex Terrain. *Boundary-Layer Meteorology*, *177*(1), 19–47.
- Acevedo, O. C., Costa, F. D., Maroneze, R., Carvalho, A. D., Puhales, F. S., & Oliveira, P. E. S. (2021). External controls on the transition between stable boundary-layer turbulence regimes. *Quarterly Journal of the Royal Meteorological Society*, (p. qj.4027).
- Acevedo, O. C., Mahrt, L., Puhales, F. S., Costa, F. D., Medeiros, L. E., & Degrazia, G. A. (2016). Contrasting structures between the decoupled and coupled states of the stable boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 693–702.
- Aït-Sahalia, Y. (1999). Transition Densities for Interest Rate and Other Nonlinear Diffusions. *The Journal of Finance*, (p. 35).
- Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach. *Econometrica*, *70*(1), 223–262.
- Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, *36*(2).
- Albers, D. J., Blancquart, P.-A., Levine, M. E., Esmailzadeh Seylabi, E., & Stuart, A. (2019). Ensemble Kalman methods with constraints. *Inverse Problems*, *35*(9), 095007.
- Alnæs, M., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., & Wells, G. N. (2015). The FEniCS Project Version 1.5. *Archive of Numerical Software*, Vol 3. Publisher: University Library Heidelberg.
- Ansorge, C. (2017). *Analyses of Turbulence in the Neutrally and Stably Stratified Planetary Boundary Layer*. Springer Theses. Cham: Springer International Publishing.
- Ansorge, C., & Mellado, J. P. (2014). Global Intermittency and Collapsing Turbulence in the Stratified Planetary Boundary Layer. *Boundary-Layer Meteorology*, *153*(1), 89–116.
- Baas, P., van de Wiel, B., Van der Linden, S., & Bosveld, F. (2018). From near-neutral to strongly stratified: Adequately modelling the clear-sky nocturnal boundary layer at Cabauw. *Boundary-Layer Meteorology*, *166*(2), 217–238. Publisher: Springer.

REFERENCES

- Baas, P., van de Wiel, B. J., Van Meijgaard, E., Vignon, E., Genthon, C., van der Linden, S. J., & de Roode, S. R. (2019). Transitions in the wintertime near-surface temperature inversion at Dome C, Antarctica. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 930–946. Publisher: Wiley Online Library.
- Baas, P., van de Wiel, B. J. H., van den Brink, L., & Holtslag, A. (2012). Composite hodographs and inertial oscillations in the nocturnal boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *138*(663), 528–535.
- Baklanov, A., & Grisogono, B. (Eds.) (2007). *Atmospheric boundary layers: nature, theory, and application to environmental modelling and security*. New York ; London: Springer Science+Business Media. Meeting Name: NATO Advanced Research Workshop "Atmospheric Boundary Layers: Modelling and Applications to Environmental Security" OCLC: ocn173498958.
- Banta, R. M. (2008). Stable boundary layer regimes from the perspective of the low-level jet. *Acta Geophysica*, *56*(1), 58–87.
- Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55.
- Beljaars, A., Balsamo, G., Bechtold, P., Bozzo, A., Forbes, R., Hogan, R. J., Köhler, M., Morcrette, J.-J., Tompkins, A. M., Viterbo, P., & Wedi, N. (2018). The Numerics of Physical Parametrization in the ECMWF Model. *Frontiers in Earth Science*, *6*, 137.
- Belušić, D., & Güttler, I. (2010). Can mesoscale models reproduce meandering motions? *Quarterly Journal of the Royal Meteorological Society*, (pp. n/a–n/a).
- Bermejo-Moreno, I., & Pullin, D. I. (2008). On the non-local geometry of turbulence. *Journal of Fluid Mechanics*, *603*, 101–135.
- Billings, S. A. (2013). *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. Chichester, West Sussex, United Kingdom: John Wiley & Sons, Inc.
- Blackadar, A. K. (1962). The vertical distribution of wind and turbulent exchange in a neutral atmosphere. *Journal of Geophysical Research*, *67*(8), 3095–3102.
- Bos, W. J. (2021). Three-dimensional turbulence without vortex stretching. *Journal of Fluid Mechanics*, *915*, A121.
- Boyko, V., Krumscheid, S., & Vercauteren, N. (2021). Statistical learning of nonlinear stochastic differential equations from nonstationary time series using variational clustering. *arXiv:2102.12395 [math]*. ArXiv: 2102.12395. Submitted to Multiscale Modeling and Simulation.
- Boyko, V., & Vercauteren, N. (2020). Multiscale Shear Forcing of Turbulence in the Nocturnal Boundary Layer: A Statistical Analysis. *Boundary-Layer Meteorology*.

- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer texts in statistics. New York: Springer, 2nd ed ed.
- Brutsaert, W. (1982). *Evaporation into the Atmosphere*. Dordrecht: Springer Netherlands.
- Burnham, K. P., Anderson, D. R., & Burnham, K. P. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer, 2nd ed ed. OCLC: ocm48557578.
- Carbone, M., & Bragg, A. D. (2020). Is vortex stretching the main cause of the turbulent energy cascade? *Journal of Fluid Mechanics*, 883, R2.
- Cava, D., Mortarini, L., Anfossi, D., & Giostra, U. (2019). Interaction of Submeso Motions in the Antarctic Stable Boundary Layer. *Boundary-Layer Meteorology*, 171(2), 151–173.
- Cheng, Y., Sayde, C., Li, Q., Basara, J., Selker, J., Tanner, E., & Gentine, P. (2017). Failure of Taylor’s hypothesis in the atmospheric surface layer and its correction for eddy-covariance measurements: Correction of Taylor’s Hypothesis. *Geophysical Research Letters*, 44(9), 4287–4295.
- Christensen, H. M. (2020). Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society*, 146(727), 938–962.
- Cuxart, J., Holtslag, A. A. M., Beare, R. J., Bazile, E., Beljaars, A., Cheng, A., Conangla, L., Ek, M., Freedman, F., Hamdi, R., Kerstein, A., Kitagawa, H., Lenderink, G., Lewellen, D., Maillhot, J., Mauritsen, T., Perov, V., Schayes, G., Steeneveld, G.-J., Svensson, G., Taylor, P., Weng, W., Wunsch, S., & Xu, K.-M. (2006). Single-Column Model Intercomparison for a Stably Stratified Atmospheric Boundary Layer. *Boundary-Layer Meteorology*, 118(2), 273–303.
- Dacunha-Castelle, D., & Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19(4), 263–284.
- Davis, P. A. (2000). Development and mechanisms of the nocturnal jet. *Meteorological Applications*, 7(3), 239–246.
- Donda, J. M. M., van Hooijdonk, I. G. S., Moene, A. F., Jonker, H. J. J., van Heijst, G. J. F., Clercx, H. J. H., & van de Wiel, B. J. H. (2015). Collapse of turbulence in stably stratified channel flow: a transient phenomenon: Turbulence Collapse in Stably Stratified Channel Flow. *Quarterly Journal of the Royal Meteorological Society*, 141(691), 2137–2147.
- Drüe, C., & Heinemann, G. (2007). Characteristics of intermittent turbulence in the upper stable boundary layer over Greenland. *Boundary-Layer Meteorology*, 124(3), 361–381.
- Dunbar, T. M., Hanert, E., & Hogan, R. J. (2008). A One-Dimensional Finite-Element Boundary-Layer Model with a Vertical Adaptive Grid. *Boundary-Layer Meteorology*, 128(3), 459–472.

REFERENCES

- Ďurán, I. B., Geleyn, J.-F., & Váňa, F. (2014). A Compact Model for the Stability Dependency of TKE Production–Destruction–Conversion Terms Valid for the Whole Range of Richardson Numbers. *Journal of the Atmospheric Sciences*, *71*(8), 3004–3026.
- Ďurán, I. B., Geleyn, J.-F., Váňa, F., Schmidli, J., & Brožková, R. (2018). A Turbulence Scheme with Two Prognostic Turbulence Energies. *Journal of the Atmospheric Sciences*, *75*(10), 3381–3402.
- Elandt-Johnson, R. C., Johnson, N. L., & Statistiker, M. (1980). *Survival models and data analysis*. 312.0151 E37. Wiley Online Library.
- EOL data (1999). Earth Observing Laboratory.
- Fernando, H. J. S., & Weil, J. C. (2010). Whither the Stable Boundary Layer?: A Shift in the Research Agenda. *Bulletin of the American Meteorological Society*, *91*(11), 1475–1484.
- Foken, T. (2006). 50 Years of the Monin–Obukhov Similarity Theory. *Boundary-Layer Meteorology*, *119*(3), 431–447.
- Foken, T. (2017). *Micrometeorology*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Franzke, C. L. E., O’Kane, T. J., Monselesan, D. P., Risbey, J. S., & Horenko, I. (2015). Systematic attribution of observed Southern Hemisphere circulation trends to external forcing and internal variability. *Nonlinear Processes in Geophysics*, *22*(5), 513–525.
- Freire, L. S., & Chamecki, M. (2018). A one dimensional stochastic model of turbulence within and above plant canopies. *Agricultural and Forest Meteorology*, *250-251*, 9–23.
- Fuchs, C. (2013). *Inference for Diffusion Processes*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- G. Johnson, S. (2021). The NLOpt nonlinear-optimization package.
- Galperin, B., Sukoriansky, S., & Anderson, P. S. (2007). On the critical Richardson number in stably stratified turbulence. *Atmospheric Science Letters*, *8*(3), 65–69.
- Garratt, J. R. (1994). The atmospheric boundary layer. *Earth-Science Reviews*, *37*(1-2), 89–134. Publisher: Elsevier.
- Gerber, S., & Horenko, I. (2014). On inference of causality for discrete state models in a multiscale context. *Proceedings of the National Academy of Sciences*, *111*(41), 14651–14656.
- Golding, B., Mylne, K., & Clark, P. (2004). The history and future of numerical weather prediction in the Met Office. *Weather*, *59*(11), 299–306.
- Grachev, A. A., Leo, L. S., Fernando, H. J., Fairall, C. W., Creegan, E., Blomquist, B. W., Christman, A. J., & Hocut, C. M. (2018). Air–sea/land interaction in the coastal zone. *Boundary-Layer Meteorol*, *167*(2), 181–210. Publisher: Springer.
- Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., & Hudson, S. R. (2019). Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer. *Journal of Climate*, *32*(14), 4121–4143.

- Granger, C. W. J. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, 39(1), 199–211.
- Gryanik, V. M., Lüpkes, C., Grachev, A., & Sidorenko, D. (2020). New Modified and Extended Stability Functions for the Stable Boundary Layer based on SHEBA and Parametrizations of Bulk Transfer Coefficients for Climate Models. *Journal of the Atmospheric Sciences*, 77(8), 2687–2716.
- Hannan, E. J. (1970). *Multiple time series*. Wiley series in probability and mathematical statistics. New York: Wiley.
- Harikrishnan, A., Ansorge, C., Klein, R., & Vercauteren, N. (2021). Geometry and organization of coherent structures in stably stratified atmospheric boundary layers. [eprint: 2110.02253](#).
- He, Y., McFarlane, N., & Monahan, A. (2019). A New TKE-Based Parameterization of Atmospheric Turbulence in the Canadian Global and Regional Climate Models. *J Adv Model Earth Syst*. Publisher: Wiley Online Library.
- Hirt, M., Rasp, S., Blahak, U., & Craig, G. C. (2019). Stochastic Parameterization of Processes Leading to Convective Initiation in Kilometer-Scale Models. *Monthly Weather Review*, 147, 18.
- Högström, U. (1996). Review of some basic characteristics of the atmospheric surface layer. *Boundary-Layer Meteorology*, 78(3), 215–246. Publisher: Springer.
- Holton, J. R. (2004). *An introduction to dynamic meteorology*. No. v. 88 in International geophysics series. Burlington, MA: Elsevier Academic Press, 4th ed ed. OCLC: 54400282.
- Holtslag, A. A. M., Svensson, G., Baas, P., Basu, S., Beare, B., Beljaars, A. C. M., Bosveld, F. C., Cuxart, J., Lindvall, J., Steeneveld, G. J., Tjernström, M., & Van De Wiel, B. J. H. (2013). Stable Atmospheric Boundary Layers and Diurnal Cycles: Challenges for Weather and Climate Models. *Bulletin of the American Meteorological Society*, 94(11), 1691–1706.
- Honnert, R., Couvreux, F., Masson, V., & Lancz, D. (2016). Sampling the Structure of Convective Turbulence and Implications for Grey-Zone Parametrizations. *Boundary-Layer Meteorology*, 160(1), 133–156.
- Honnert, R., Efstathiou, G. A., Beare, R. J., Ito, J., Lock, A., Neggers, R., Plant, R. S., Shin, H. H., Tomassini, L., & Zhou, B. (2020). The Atmospheric Boundary Layer and the “Gray Zone” of Turbulence: A Critical Review. *Journal of Geophysical Research: Atmospheres*, 125(13).
- Honnert, R., Masson, V., & Couvreux, F. (2011). A Diagnostic for Evaluating the Representation of Turbulence in Atmospheric Models at the Kilometric Scale. *Journal of the Atmospheric Sciences*, 68(12), 3112–3131.
- Horenko, I. (2010a). Finite Element Approach to Clustering of Multidimensional Time Series. *SIAM Journal on Scientific Computing*, 32(1), 62–83.

REFERENCES

- Horenko, I. (2010b). On the Identification of Nonstationary Factor Models and Their Application to Atmospheric Data Analysis. *Journal of the Atmospheric Sciences*, 67(5), 1559–1574.
- Horsthemke, W. (1984). Noise Induced Transitions. *Non-Equilibrium Dynamics in Chemical Systems*, (pp. 150–160). Publisher: Springer, Berlin, Heidelberg.
- Howard, L. N. (1961). Note on a paper of John W. Miles. *Journal of Fluid Mechanics*, 10(4), 509–512. Publisher: Cambridge University Press.
- Howell, J. F., & Mahrt, L. (1997). Multiresolution Flux Decomposition. *Boundary-Layer Meteorology*, 83(1), 117–137.
- Huang, J., & Bou-Zeid, E. (2013). Turbulence and Vertical Fluxes in the Stable Atmospheric Boundary Layer. Part I: A Large-Eddy Simulation Study. *Journal of the Atmospheric Sciences*, 70(6), 1513–1527.
- Huang, J., Bou-Zeid, E., & Golaz, J.-C. (2013). Turbulence and Vertical Fluxes in the Stable Atmospheric Boundary Layer. Part II: A Novel Mixing-Length Model. *Journal of the Atmospheric Sciences*, 70(6), 1528–1542.
- Hurst, D., & Vassilicos, J. C. (2007). Scalings and decay of fractal-generated turbulence. *Physics of Fluids*, 19(3), 035103.
- Hyndman, R. J. (1996). Computing and graphing highest density regions. *The American Statistician*, 50(2), 120–126. Publisher: Taylor & Francis Group.
- Igdalov, D., Kaiser, O., & Horenko, I. (2019). On a simultaneous parameter inference and missing data imputation for nonstationary autoregressive models. *arXiv:1912.12894 [stat]*. ArXiv: 1912.12894.
- Jiménez, M. A., & Cuxart, J. (2005). Large-Eddy Simulations of the Stable Boundary Layer Using the Standard Kolmogorov Theory: Range of Applicability. *Boundary-Layer Meteorology*, 115(2), 241–261.
- Journel, A., & Huijbregts, C. J. (1978). Mining geostatistics. Academic Press, New York. *Mining geostatistics. Academic Press, New York.*
- Kaiser, O. (2015). *Data-based analysis of extreme events: inference, numerics and applications*. Ph.D. thesis, Università della Svizzera italiana.
- Kaiser, O., Hien, S., Achatz, U., & Horenko, I. (2017). Stochastic Parametrization for the Impact of Neglected Variability Patterns. In *EGU General Assembly Conference Abstracts*, (p. 10397).
- Kalliadasis, S., Krumscheid, S., & Pavliotis, G. A. (2015). A new framework for extracting coarse-grained models from time series with multiscale structure. *J. Comput. Phys.*, 296, 314–328.

- Kang, S.-L., & Won, H. (2016). Spectral structure of 5 year time series of horizontal wind speed at the Boulder Atmospheric Observatory: Spectra of 5 Year Time Series. *Journal of Geophysical Research: Atmospheres*, *121*(20), 11,946–11,967.
- Kang, Y., Belušić, D., & Smith-Miles, K. (2015). Classes of structures in the stable atmospheric boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *141*(691), 2057–2069.
- Katul, G., & Vidakovic, B. (1996). The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. *Boundary-Layer Meteorology*, *77*(2), 153–172. Publisher: Springer.
- Kerstein, A. R. (1999). One-dimensional turbulence: model formulation and application to homogeneous turbulence, shear flows, and buoyant stratified flows. *Journal of Fluid Mechanics*, *392*, 277–334.
- Kerstein, A. R., Ashurst, W. T., Wunsch, S., & Nilsen, V. (2001). One-dimensional turbulence: vector formulation and application to free shear flows. *Journal of Fluid Mechanics*, *447*, 85–109.
- Kim, J.-H. (1999). Spurious correlation between ratios with a common divisor. *Statistics & probability letters*, *44*(4), 383–386. Publisher: Elsevier.
- Kit, E., Hocut, C. M., Liberzon, D., & Fernando, H. J. S. (2017). Fine scale turbulent bursts in stable atmospheric boundary layer in complex terrain. *Journal of Fluid Mechanics*, *833*, 745–772.
- Klein, M., & Schmidt, H. (2019). Investigating Rayleigh Benard convection at low Prandtl numbers using one-dimensional turbulence modeling. In *11th International Symposium on Turbulence and Shear Flow Phenomena, Southampton, UK, July*, vol. 30, (pp. 1–6).
- Kober, K., & Craig, G. C. (2016). Physically Based Stochastic Perturbations (PSP) in the Boundary Layer to Represent Uncertainty in Convective Initiation. *Journal of the Atmospheric Sciences*, *73*(7), 2893–2911.
- Krumscheid, S. (2018). Perturbation-based inference for diffusion processes: Obtaining effective models from multiscale data. *Math. Models Methods Appl. Sci.*, *28*(8), 1565–1597.
- Krumscheid, S., Pavliotis, G. A., & Kalliadasis, S. (2013). Semiparametric Drift and Diffusion Estimation for Multiscale Diffusions. *Multiscale Model. Simul.*, *11*(2), 442–473.
- Kumar, P., & Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of geophysics*, *35*(4), 385–412. Publisher: Wiley Online Library.
- Kurbatskii, A. F., & Kurbatskaya, L. I. (2012). On the eddy mixing and energetics of turbulence in a stable atmospheric boundary layer. *Izvestiya, Atmospheric and Oceanic Physics*, *48*(6), 595–602.
- Kuznetsov, A., & Clarke, D. (1994). Simple Numerical Algorithms for Continuous-To-Discrete and Discrete-To-Continuous Conversion of the Systems with Time Delay. *IFAC Proceedings Volumes*, *27*(8), 1549–1554.

REFERENCES

- Laizet, S., & Vassilicos, J. C. (2011). DNS of Fractal-Generated Turbulence. *Flow, Turbulence and Combustion*, 87(4), 673–705.
- Lan, C., Liu, H., Li, D., Katul, G. G., & Finn, D. (2018). Distinct Turbulence Structures in Stably Stratified Boundary Layers With Weak and Strong Surface Shear. *Journal of Geophysical Research: Atmospheres*, 123(15), 7839–7854.
- Larsén, X. G., Larsen, S. E., & Petersen, E. L. (2016). Full-Scale Spectrum of Boundary-Layer Winds. *Boundary-Layer Meteorology*, 159(2), 349–371.
- Lee, E. S., & Forthofer, R. N. (2006). *Analyzing complex survey data*. No. no. 07-071 in Quantitative applications in the social sciences. Thousand Oaks, Calif: Sage Publications, 2nd ed ed. OCLC: ocm59002567.
- Lee, G. R., Gommers, R., Wasilewski, F., Wohlfahrt, K., & O’Leary, A. (2019a). Py-Wavelets/pywt: PyWavelets v1.0.3.
- Lee, T. R., Buban, M., Turner, D. D., Meyers, T. P., & Baker, C. B. (2019b). Evaluation of the High-Resolution Rapid Refresh (HRRR) Model Using Near-Surface Meteorological and Flux Observations from Northern Alabama. *Weather and Forecasting*, 34(3), 635–663.
- Lee, X., Massman, W. J., & Law, B. E. (2004). *Handbook of Micrometeorology*. Springer Science & Business Media.
- Leibovich, N., & Barkai, E. (2015). Aging Wiener-Khinchin Theorem. *Physical Review Letters*, 115(8), 080602.
- Lenderink, G., & Holtslag, A. (2004). An updated length-scale formulation for turbulent mixing in clear and cloudy boundary layers. *Quarterly Journal of the Royal Meteorological Society*, 130(604), 3405–3427.
- Leutbecher, M., Lock, S.-J., Ollinaho, P., Lang, S. T. K., Balsamo, G., Bechtold, P., Bonavita, M., Christensen, H. M., Diamantakis, M., Dutra, E., English, S., Fisher, M., Forbes, R. M., Goddard, J., Haiden, T., Hogan, R. J., Juricke, S., Lawrence, H., MacLeod, D., Magnusson, L., Malardel, S., Massart, S., Sandu, I., Smolarkiewicz, P. K., Subramanian, A., Vitart, F., Wedi, N., & Weisheimer, A. (2017). Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society*, 143(707), 2315–2339.
- Lewis-Beck, C., & Lewis-Beck, M. (2015). *Applied regression: An introduction*, vol. 22. Sage publications.
- Li, Y., Wei, H.-L., Billings, S. A., & Sarrigiannis, P. (2016). Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG. *International Journal of Systems Science*, 47(11), 2671–2681.
- Lindheim, J. v., Harikrishnan, A., Dörffel, T., Klein, R., Koltai, P., Mikula, N., Müller, A., Névir, P., Pacey, G., Polzin, R., & Vercauteren, N. (2021). Definition, detection, and tracking of persistent structures in atmospheric flows. [_eprint: 2111.13645](#).

- Logg, A., Mardal, K.-A., & Wells, G. (Eds.) (2012). *Automated Solution of Differential Equations by the Finite Element Method*. Springer Berlin Heidelberg.
- Lord, G. J., Powell, C. E., & Shardlow, T. (2014). *An introduction to computational stochastic PDEs*. No. 50 in Cambridge texts in applied mathematics. New York, NY, USA: Cambridge University Press.
- Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., & Fairbairn, D. (2015). Comparison of Hybrid-4DVar and Hybrid-4DVar Data Assimilation Methods for Global NWP. *Monthly Weather Review*, *143*(1), 212–229.
- Lundquist, J. K. (2003). Intermittent and Elliptical Inertial Oscillations in the Atmospheric Boundary Layer. *Journal of the Atmospheric Sciences*, *60*, 13.
- Mahrt, L. (1999). Stratified Atmospheric Boundary Layers. *Boundary-Layer Meteorology*, *90*(3), 375–396.
- Mahrt, L. (2010a). Common microfronts and other solitary events in the nocturnal boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *136*(652), 1712–1722.
- Mahrt, L. (2010b). Variability and Maintenance of Turbulence in the Very Stable Boundary Layer. *Boundary-Layer Meteorology*, *135*(1), 1–18.
- Mahrt, L. (2011). The Near-Calm Stable Boundary Layer. *Boundary-Layer Meteorology*, *140*(3), 343–360.
- Mahrt, L. (2014). Stably Stratified Atmospheric Boundary Layers. *Annual Review of Fluid Mechanics*, *46*(1), 23–45.
- Mahrt, L. (2019). Microfronts in the nocturnal boundary layer. *Quarterly Journal of the Royal Meteorological Society*, *145*(719), 546–562.
- Mahrt, L., & Bou-Zeid, E. (2020). Non-stationary Boundary Layers. *Boundary-Layer Meteorology*.
- Mahrt, L., & Vickers, D. (2003). Formulation of Turbulent Fluxes in the Stable Boundary Layer. *Journal of the Atmospheric Sciences*, *60*, 11.
- Mahrt, L., & Vickers, D. (2005). Boundary-Layer Adjustment Over Small-Scale Changes of Surface Heat Flux. *Boundary-Layer Meteorology*, *116*(2), 313–330.
- Mahrt, L., & Vickers, D. (2006). Extremely Weak Mixing in Stable Conditions. *Boundary-Layer Meteorology*, *119*(1), 19–39.
- Mahrt, L., Vickers, D., Nakamura, R., Soler, M. R., Sun, J., Burns, S., & Lenschow, D. (2001). Shallow Drainage Flows. *Boundary-Layer Meteorology*, *101*(2), 243–260.
- Malhi, Y. (1996). The behaviour of the roughness length for temperature over heterogeneous surfaces. *Quarterly Journal of the Royal Meteorological Society*, *122*(533), 1095–1125.
- Mallat, S. (1999). *A wavelet tour of signal processing*. Elsevier.

REFERENCES

- Maroneze, R., Acevedo, O. C., Costa, F. D., & Sun, J. (2019). Simulating the regime transition of the stable boundary layer using different simplified models. *Boundary-Layer Meteorology*, *170*(2), 305–321. Publisher: Springer.
- Marusic, I., & Monty, J. P. (2019). Attached Eddy Model of Wall Turbulence. *Annual Review of Fluid Mechanics*, *51*(1), 49–74.
- Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, *58*(8), 1246–1266.
- Meneveau, C. (2012). Germano identity-based subgrid-scale modeling: A brief survey of variations on a fertile theme. *Physics of Fluids*, *24*(12), 121301.
- Metzner, P., Putzig, L., & Horenko, I. (2012). Analysis of persistent nonstationary time series and applications. *Communications in Applied Mathematics and Computational Science*, *7*(2), 175–229.
- Miles, J. W. (1961). On the stability of heterogeneous shear flows. *Journal of Fluid Mechanics*, *10*(4), 496–508. Publisher: Cambridge University Press.
- Monahan, A. H., Rees, T., He, Y., & McFarlane, N. (2015). Multiple Regimes of Wind, Stratification, and Turbulence in the Stable Boundary Layer. *Journal of the Atmospheric Sciences*, *72*(8), 3178–3198.
- Monin, A. S., & Obukhov, A. M. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, *151*(163), e187.
- Muschinski, A., Frehlich, R. G., & Balsley, B. B. (2004). Small-scale and large-scale intermittency in the nocturnal boundary layer and the residual layer. *Journal of Fluid Mechanics*, *515*, 319–351.
- Nieuwstadt, F. T. M. (1984). The Turbulent Structure of the Stable, Nocturnal Boundary Layer. *Journal of the Atmospheric Sciences*, *41*(14), 2202–2216. Publisher: American Meteorological Society.
- O’Kane, T. J., Monselesan, D. P., Risbey, J. S., Horenko, I., & Franzke, C. L. E. (2017). Research Article. On memory, dimension, and atmospheric teleconnections. *Mathematics of Climate and Weather Forecasting*, *3*(1), 1–27.
- Olea, R. A. (1999). *Geostatistics for Engineers and Earth Scientists*. Boston, MA: Springer US.
- Optis, M., Monahan, A., & Bosveld, F. C. (2016). Limitations and breakdown of Monin-Obukhov similarity theory for wind profile extrapolation under stable stratification: Limitations and breakdown of Monin-Obukhov similarity theory for wind profile extrapolation under stable stratification. *Wind Energy*, *19*(6), 1053–1072.
- Ostdiek, V., & Blumen, W. (1997). A Dynamic Trio: Inertial Oscillation, Deformation Frontogenesis, and the Ekman–Taylor Boundary Layer. *Journal of the Atmospheric Sciences*, *54*, 13.

- Palmén, E. H., & Newton, C. W. (1969). *Atmospheric circulation systems: their structure and physical interpretation*, vol. 13. Academic press.
- Parente, A., Górlé, C., van Beeck, J., & Benocci, C. (2011). A Comprehensive Modelling Approach for the Neutral Atmospheric Boundary Layer: Consistent Inflow Conditions, Wall Function and Turbulence Model. *Boundary-Layer Meteorology*, *140*(3), 411–428.
- Peng, Z., Jackson, M. R., Rongong, J., Chu, F., & Parkin, R. M. (2009). On the energy leakage of discrete wavelet transform. *Mechanical Systems and Signal Processing*, *23*(2), 330–343. Publisher: Elsevier.
- Pfister, L., Lapo, K., Mahrt, L., & Thomas, C. K. (2021). Thermal Submeso Motions in the Nocturnal Stable Boundary Layer. Part 2: Generating Mechanisms and Implications. *Boundary-Layer Meteorology*, *180*(2), 203–224.
- Pope, S. B. (1991). Application of the velocity-dissipation probability density function model to inhomogeneous turbulent flows. *Physics of Fluids A: Fluid Dynamics*, *3*(8), 1947–1957.
- Pope, S. B. (2000). *Turbulent flows*. Cambridge university press.
- Pospisil, L., Gagliardini, P., Sawyer, W., & Horenko, I. (2018). On a scalable nonparametric denoising of time series signals. *Communications in Applied Mathematics and Computational Science*, *13*(1), 107–138.
- Price, W. L. (1977). A controlled random search procedure for global optimisation. *The Computer Journal*, *20*(4), 367–370.
- Price, W. L. (1983). Global optimization by controlled random search. *Journal of Optimization Theory and Applications*, *40*(3), 333–348.
- Qiu, J., Paw U, K. T., & Shaw, R. H. (1995). The leakage problem of orthonormal wavelet transforms when applied to atmospheric turbulence. *Journal of Geophysical Research: Atmospheres*, *100*(D12), 25769–25779. Publisher: Wiley Online Library.
- Quarteroni, A. (2017). *Numerical Models for Differential Problems*, vol. 16 of *MS&A*. Cham: Springer International Publishing.
- Quinn, C., Harries, D., & O’Kane, T. J. (2021). Dynamical analysis of a reduced model for the North Atlantic Oscillation. *Journal of the Atmospheric Sciences*. Publisher: American Meteorological Society.
- Richard P, B. (2013). *Algorithms for minimization without derivatives*. Courier Corporation.
- Risbey, J. S., O’Kane, T. J., Monselesan, D. P., Franzke, C., & Horenko, I. (2015). Metastability of Northern Hemisphere Teleconnection Modes. *Journal of the Atmospheric Sciences*, *72*(1), 35–54.
- Robinson, S. K. (1991). *The kinematics of turbulent boundary layer structure*. PhD Thesis, Stanford University.

REFERENCES

- Rodgers, J. L., & Nicewander, W. A. (1988). Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician*, 42(1), 59. Publisher: JSTOR.
- Rodrigo, J. S., & Anderson, P. S. (2013). Investigation of the Stable Atmospheric Boundary Layer at Halley Antarctica. *Boundary-Layer Meteorology*, 148(3), 517–539.
- Sakradzija, M., & Klocke, D. (2018). Physically Constrained Stochastic Shallow Convection in Realistic Kilometer-Scale Simulations. *Journal of Advances in Modeling Earth Systems*, 10(11), 2755–2776.
- Sandu, I., Beljaars, A., Bechtold, P., Mauritsen, T., & Balsamo, G. (2013). Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *Journal of Advances in Modeling Earth Systems*, 5(2), 117–133.
- Sanz Rodrigo, J., Churchfield, M., & Kosovic, B. (2017). A methodology for the design and testing of atmospheric boundary layer models for wind energy applications. *Wind Energy Science*, 2(1), 35–54.
- Schalkwijk, J., Jonker, H. J. J., Siebesma, A. P., & Van Meijgaard, E. (2015). Weather Forecasting Using GPU-Based Large-Eddy Simulations. *Bulletin of the American Meteorological Society*, 96(5), 715–723.
- Schlichting, H., & Gersten, K. (2017). *Boundary-Layer Theory*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Schultz, M. G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L. H., Mozaffari, A., & Stadtler, S. (2021). Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), 20200097.
- Scott, D. W. (2015). *Multivariate Density Estimation*. Wiley.
- Seaman, N. L., Gaudet, B. J., Stauffer, D. R., Mahrt, L., Richardson, S. J., Zielonka, J. R., & Wyngaard, J. C. (2012). Numerical Prediction of Submesoscale Flow in the Nocturnal Stable Boundary Layer over Complex Terrain. *Monthly Weather Review*, 140(3), 956–977.
- Shi, Y., & Hu, F. (2020). Ramp-Like PM Accumulation Process and Z-Less Similarity in the Stable Boundary Layer. *Geophysical Research Letters*, 47(3).
- Sorbjan, Z. (2012). A study of the stable boundary layer based on a single-column K-theory model. *Boundary-layer meteorology*, 142(1), 33–53. Publisher: Springer.
- Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klingner, C., Klocke, D., Rybka, H., Schubotz, W., Windmiller, J., Adamidis, P., Arka, I., Barlakas, V., Biercamp, J., Brueck, M., Brune, S., Buehler, S. A., Burkhardt, U., Cioni, G., Costa-Surós, M., Crewell, S., Crüger, T., Deneke, H., Friederichs, P., Henken, C. C., Hohenegger, C., Jacob, M., Jakub, F., Kalthoff, N., Köhler, M., van LAAR, T. W., Li, P., Löhnert, U., Macke, A., Madenach, N., Mayer, B., Nam, C., Naumann, A. K., Peters, K., Poll, S., Quaas, J., Röber, N., Rochetin, N., Scheck, L., Schemann, V., Schnitt, S., Seifert, A., Senf, F., Shapkalijevski, M., Simmer, C., Singh,

- S., Sourdeval, O., Spickermann, D., Strandgren, J., Tessiot, O., Vercauteren, N., Vial, J., Voigt, A., & Zängl, G. (2020). The Added Value of Large-eddy and Storm-resolving Models for Simulating Clouds and Precipitation. *Journal of the Meteorological Society of Japan. Ser. II*, 98(2), 395–435.
- Stewart, R. W. (1979). *The atmospheric boundary layer*. WMO ; no. 523. Geneva: World Meteorological Organization.
- Stiperski, I., Katul, G. G., & Calaf, M. (2021). Universal Return to Isotropy of Inhomogeneous Atmospheric Boundary Layer Turbulence. *Physical Review Letters*, 126(19), 194501.
- Stoll, R., Gibbs, J. A., Salesky, S. T., Anderson, W., & Calaf, M. (2020). Large-Eddy Simulation of the Atmospheric Boundary Layer. *Boundary-Layer Meteorology*, 177, 541–581.
- Stresing, R., Peinke, J., Seoud, R. E., & Vassilicos, J. C. (2010). Defining a New Class of Turbulent Flows. *Physical Review Letters*, 104(19), 194501.
- Stuart, A., & Ord, K. (1994). *Kendalls Advanced Theory of Statistics Arnold*. London.
- Stull, R. B. (1988). *An introduction to boundary layer meteorology*, vol. 13. Springer Science & Business Media.
- Sun, J., Burns, S. P., Lenschow, D. H., Banta, R., Newsom, R., Coulter, R., Frasier, S., Ince, T., Nappo, C., Cuxart, J., Blumen, W., Lee, X., & Hu, X.-Z. (2002). Intermittent Turbulence Associated with a Density Current Passage in the Stable Boundary Layer. *Boundary-Layer Meteorology*, 105(2), 199–219.
- Sun, J., Lenschow, D. H., Burns, S. P., Banta, R. M., Newsom, R. K., Coulter, R., Frasier, S., Ince, T., Nappo, C., Balsley, B. B., Jensen, M., Mahrt, L., Miller, D., & Skelly, B. (2004). Atmospheric Disturbances that Generate Intermittent Turbulence in Nocturnal Boundary Layers. *Boundary-Layer Meteorology*, 110(2), 255–279.
- Sun, J., Mahrt, L., Banta, R. M., & Pichugina, Y. L. (2012). Turbulence Regimes and Turbulence Intermittency in the Stable Boundary Layer during CASES-99. *Journal of the Atmospheric Sciences*, 69(1), 338–351. Publisher: American Meteorological Society.
- Sun, J., Nappo, C. J., Mahrt, L., Belušić, D., Grisogono, B., Stauffer, D. R., Pulido, M., Staquet, C., Jiang, Q., Pouquet, A., Yagüe, C., Galperin, B., Smith, R. B., Finnigan, J. J., Mayor, S. D., Svensson, G., Grachev, A. A., & Neff, W. D. (2015). Review of wave-turbulence interactions in the stable atmospheric boundary layer. *Reviews of Geophysics*, 53(3), 956–993.
- Tang, J., Niu, X., Wang, S., Gao, H., Wang, X., & Wu, J. (2016). Statistical downscaling and dynamical downscaling of regional climate in China: Present climate evaluations and future climate projections. *Journal of Geophysical Research: Atmospheres*, 121(5), 2110–2129.
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411–423.

REFERENCES

- Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), 61–78. Publisher: American Meteorological Society.
- Ullrich, P. A., & Jablonowski, C. (2012). MCore: A non-hydrostatic atmospheric dynamical core utilizing high-order finite-volume methods. *Journal of Computational Physics*, 231(15), 5078–5108.
- Van de Wiel, B. J. H., Moene, A. F., Jonker, H. J. J., Baas, P., Basu, S., Donda, J. M. M., Sun, J., & Holtslag, A. A. M. (2012). The Minimum Wind Speed for Sustainable Turbulence in the Nocturnal Boundary Layer. *Journal of the Atmospheric Sciences*, 69(11), 3116–3127.
- Van de Wiel, B. J. H., Moene, A. F., Steeneveld, G. J., Baas, P., Bosveld, F. C., & Holtslag, A. A. M. (2010). A Conceptual View on Inertial Oscillations and Nocturnal Low-Level Jets. *Journal of the Atmospheric Sciences*, 67(8), 2679–2689.
- Van de Wiel, B. J. H., Vignon, E., Baas, P., van Hooijdonk, I. G. S., van der Linden, S. J. A., Antoon van Hooft, J., Bosveld, F. C., de Roode, S. R., Moene, A. F., & Genthon, C. (2017). Regime Transitions in Near-Surface Temperature Inversions: A Conceptual Model. *Journal of the Atmospheric Sciences*, 74(4), 1057–1073.
- van der Linden, S. J. A., Baas, P., Antoon van Hooft, J., van Hooijdonk, I. G. S., Bosveld, F. C., & van de Wiel, B. J. H. (2017). Local Characteristics of the Nocturnal Boundary Layer in Response to External Pressure Forcing. *Journal of Applied Meteorology and Climatology*, 56(11), 3035–3047.
- Vercauteren, N., Boyko, V., Faranda, D., & Stiperski, I. (2019a). Scale interactions and anisotropy in stable boundary layers. *Quarterly Journal of the Royal Meteorological Society*, 145(722), 1799–1813.
- Vercauteren, N., Boyko, V., Kaiser, A., & Belusic, D. (2019b). Statistical Investigation of Flow Structures in Different Regimes of the Stable Boundary Layer. *Boundary-Layer Meteorology*, 173(2), 143–164.
- Vercauteren, N., & Klein, R. (2015). A Clustering Method to Characterize Intermittent Bursts of Turbulence and Interaction with Submesoscale Motions in the Stable Boundary Layer. *Journal of the Atmospheric Sciences*, 72(4), 1504–1517.
- Vercauteren, N., Mahrt, L., & Klein, R. (2016). Investigation of interactions between scales of motion in the stable boundary layer: Interactions between Scales of Motion in the Stable Boundary Layer. *Quarterly Journal of the Royal Meteorological Society*, 142(699), 2424–2433.
- Viana, S., Yagüe, C., & Maqueda, G. (2009). Propagation and effects of a mesoscale gravity wave over a weakly-stratified nocturnal boundary layer during the SABLES2006 field campaign. *Boundary-layer meteorology*, 133(2), 165–188. Publisher: Springer.
- Vickers, D., & Mahrt, L. (1997). Quality Control and Flux Sampling Problems for Tower and Aircraft Data. *Journal of Atmospheric and Oceanic Technology*, 14, 15.

- Vickers, D., & Mahrt, L. (2003). The Cospectral Gap and Turbulent Flux Calculations. *Journal of Atmospheric and Oceanic Technology*, *20*, 13.
- Vignon, E., Hourdin, F., Genthon, C., Van de Wiel, B. J. H., Gallée, H., Madeleine, J.-B., & Beaumet, J. (2018). Modeling the Dynamics of the Atmospheric Boundary Layer Over the Antarctic Plateau With a General Circulation Model. *Journal of Advances in Modeling Earth Systems*, *10*(1), 98–125.
- Voller, V. R., & Porté-Agel, F. (2002). More’s law and numerical modeling. *Journal of Computational Physics*, *179*, 698–703.
- von Larcher, T., Beck, A., Klein, R., Horenko, I., Metzner, P., Waidmann, M., Igdalov, D., Gassner, G., & Munz, C.-D. (2015). Towards a Framework for the Stochastic Modelling of Subgrid Scale Fluxes for Large Eddy Simulation. *Meteorologische Zeitschrift*, *24*(3), 313–342.
- Wang, Y., Klipp, C. L., Garvey, D. M., Ligon, D. A., Williamson, C. C., Chang, S. S., Newsom, R. K., & Calhoun, R. (2007). Nocturnal Low-Level-Jet-Dominated Atmospheric Boundary Layer Observed by a Doppler Lidar over Oklahoma City during JU2003. *Journal of Applied Meteorology and Climatology*, *46*(12), 2098–2109.
- Wei, H. L., Billings, S. A., & Liu, J. (2004). Term and variable selection for non-linear system identification. *International Journal of Control*, *77*(1), 86–110.
- Wei, H. L., Billings, S. A., & Liu, J. J. (2010). Time-varying parametric modelling and time-dependent spectral characterisation with applications to EEG signals using multiwavelets. *International Journal of Modelling, Identification and Control*, *9*(3), 215.
- Wyngaard, J. C. (2004). Toward Numerical Modeling in the “Terra Incognita”. *Journal of the Atmospheric Sciences*, *61*, 11.
- Xu, Y. (2018). On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, (p. 14).
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2), 226–235.
- Zaitseva, D. V., Kallistratova, M. A., Lyulyukin, V. S., Kouznetsov, R. D., & Kuznetsov, D. D. (2018). The Effect of Internal Gravity Waves on Fluctuations in Meteorological Parameters of the Atmospheric Boundary Layer. *Izvestiya, Atmospheric and Oceanic Physics*, *54*(2), 173–181.
- Želi, V., Brethouwer, G., Wallin, S., & Johansson, A. V. (2019). Consistent boundary-condition treatment for computation of the atmospheric boundary layer using the explicit algebraic reynolds-stress model. *Boundary-Layer Meteorology*, *171*(1), 53–77. Publisher: Springer.
- Zhang, X., Bao, J.-W., Chen, B., & Grell, E. D. (2018). A Three-Dimensional Scale-Adaptive Turbulent Kinetic Energy Scheme in the WRF-ARW Model. *Monthly Weather Review*, *146*(7), 2023–2045.

REFERENCES

- Zilitinkevich, S., Druzhinin, O., Glazunov, A., Kadantsev, E., Mortikov, E., Repina, I., & Troitskaya, Y. (2019). Dissipation rate of turbulent kinetic energy in stably stratified sheared flows. *Atmospheric Chemistry and Physics*, *19*(4), 2489–2496.
- Zilitinkevich, S. S., Elperin, T., Kleorin, N., L'vov, V., & Rogachevskii, I. (2009). Energy- and Flux-Budget Turbulence Closure Model for Stably Stratified Flows. Part II: The Role of Internal Gravity Waves. *Boundary-Layer Meteorology*, *133*(2), 139–164.
- Zilitinkevich, S. S., Elperin, T., Kleorin, N., & Rogachevskii, I. (2007). Energy- and flux-budget (EFB) turbulence closure model for stably stratified flows. Part I: steady-state, homogeneous regimes. *Boundary-Layer Meteorology*, *125*(2), 167–191.
- Zilitinkevich, S. S., Elperin, T., Kleorin, N., Rogachevskii, I., & Esau, I. (2013). A Hierarchy of Energy- and Flux-Budget (EFB) Turbulence Closure Models for Stably-Stratified Geophysical Flows. *Boundary-Layer Meteorology*, *146*(3), 341–373.
- Zilitinkevich, S. S., Elperin, T., Kleorin, N., Rogachevskii, I., Esau, I., Mauritsen, T., & Miles, M. W. (2008). Turbulence energetics in stably stratified geophysical flows: Strong and weak mixing regimes. *Quarterly Journal of the Royal Meteorological Society*, *134*(633), 793–799.



In Section 4.2.2 the Ordinary Differential Equation (ODE) is introduced to model the stability correction variable ϕ . In the follow up section this ODE is perturbed through the one of its parameters.

Estimation of the number of cluster to determine the scaling function in section 4.4.2 is presented in the following figure A.1

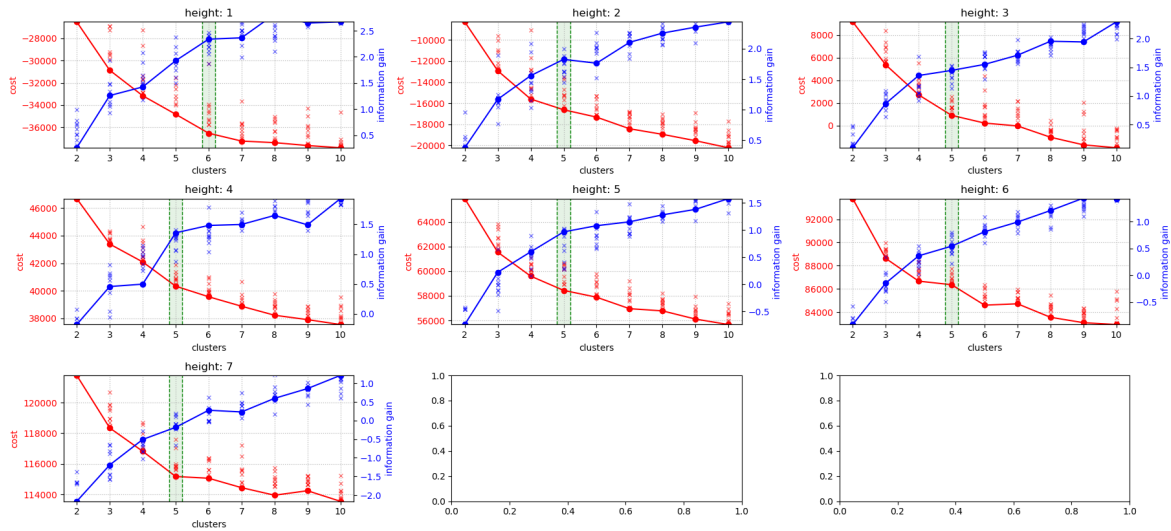


Figure A.1: Estimation of the number of clusters K for the identification study in section 4.4.2. Each box shows the tradeoff between minimization cost and gain in cluster diversity. The 1 height index denotes the 1 m height above the surface, and the 7 height index represents the 30 m height. The red curve shows a decreasing fitness function value with the number of clusters. The blue curve shows the diversity added with each new cluster. A good choice (highlighted with a green stripe) for K is when the change rate of both measures reduces simultaneously. The information gain is the logarithmically transformed diversity measure.

A.1 Coefficients of the Hermite Expansion

The coefficients η_j in Eq. (2.12) are approximated with a truncated TAYLOR series of length $M = J/2$ as:

$$\begin{aligned}
\eta_0^{(3)} &= 1, \\
\eta_1^{(3)} &= -\mu h^{1/2} - \frac{2\mu\mu' + \mu''}{4} h^{3/2} - \frac{4\mu\mu'^2 + 4\mu^2\mu'' + 6\mu'\mu'' + 4\mu\mu^{(3)} + \mu^{(4)}}{24} h^{5/2}, \\
\eta_2^{(3)} &= \frac{\mu^2 + \mu'}{2} h + \frac{6\mu^2\mu' + 4(\mu')^2 + 7\mu\mu'' + 2\mu^{(3)}}{12} h^2 + \frac{1}{96} \left(28\mu^2(\mu')^2 + 28\mu^2\mu^{(3)} \right. \\
&\quad \left. + 16(\mu')^3 + 16\mu^3\mu'' + 88\mu\mu'\mu'' + 21(\mu'')^2 + 32\mu'\mu^{(3)} + 16\mu\mu^{(4)} + 3\mu^{(5)} \right) h^3, \\
\eta_3^{(3)} &= -\frac{\mu^3 + 3\mu\mu' + \mu''}{6} h^{3/2} - \frac{1}{48} \left(12\mu^3\mu' + 28\mu(\mu')^2 + 22\mu^2\mu'' + 24\mu'\mu'' \right. \\
&\quad \left. + 14\mu\mu^{(3)} + 3\mu^{(4)} \right) h^{5/2}, \\
\eta_4^{(3)} &= \frac{\mu^4 + 6\mu^2\mu' + 3(\mu')^2 + 4\mu\mu'' + \mu^{(3)}}{24} h^2 + \frac{1}{240} \left(20\mu^4\mu' + 50\mu^3\mu'' + 100\mu^2(\mu')^2 \right. \\
&\quad \left. + 50\mu^2\mu^{(3)} + 23\mu\mu^{(4)} + 180\mu\mu'\mu'' + 40(\mu')^3 + 34(\mu'')^2 + 52\mu'\mu^{(3)} + 4\mu^{(5)} \right) h^3, \\
\eta_5^{(3)} &= -\frac{\mu^5 + 10\mu^3\mu' + 15\mu(\mu')^2 + 10\mu^2\mu'' + 10\mu'\mu'' + 5\mu\mu^{(3)} + \mu^{(4)}}{120} h^{5/2}, \\
\eta_6^{(3)} &= \frac{1}{720} \left(\mu^6 + 15\mu^4\mu' + 15(\mu')^3 + 20\mu^3\mu'' + 15\mu'\mu^{(3)} + 45\mu^2(\mu')^2 + 10(\mu'')^2 \right. \\
&\quad \left. + 15\mu^2\mu^{(3)} + 60\mu\mu'\mu'' + 6\mu\mu^{(4)} + \mu^{(5)} \right) h^3,
\end{aligned}$$

where μ is the drift of the Y process evaluated at y_0 (see Eq. (2.11)), "''" denotes a derivative and $h = \Delta t$. The HERMITE polynomials for the expansion are: $H_0(z) = 1, H_1(z) = -z, H_2(z) = -1 + z^2, H_3(z) = 3z - z^3, H_4(z) = 3 - 6z^2 + z^4, H_5(z) = -15z + 10z^3 - z^5, H_6(z) = -15 + 45z^2 - 15z^4 + z^6$.

Table A.1: The SDE model library is tested on the data to validate the model structure. The starting point is the favored model structures given in section 4.2.2. The degree of nonlinearity (parameter β) of the noise term is systematically varied. It is assumed that all parameters are time-varying. The model structure is presented in a short notation by denoting the noise term with dW_t .

	Perturbation type	Explanation
	additive noise	
$\beta = 1$	$a_1 - a_2\phi + s_1d$	Invalid because $\phi < 0$
$\beta = 5/4$	$1 + a_2\phi - a_2\phi^{5/4} + s_1dW_t$	Unstable after identification
$\beta = 3/2$	$1 + a_1\phi - a_2\phi^{3/2} + s_1dW_t$	Unstable after identification
$\beta = 2$	$1 + a_1\phi - a_2\phi^2 + s_1dW_t$	Unstable after identification
$\beta = 3$	$1 + a_1\phi - a_2\phi^3 + s_1dW_t$	Unstable after identification
	Linear term	
$\beta = 1$	$a_1 - a_2\phi + s_1\phi dW_t$	Poor performance
$\beta = 5/4$	$1 + a_1\phi - a_2\phi^{5/4} + s_1\phi dW_t$	Favorite
$\beta = 3/2$	$1 + a_1\phi - a_2\phi^{3/2} + s_1\phi dW_t$	Favorite
$\beta = 2$	$1 + a_1\phi - a_2\phi^2 + s_1\phi dW_t$	Favorite
$\beta = 3$	$1 + a_1\phi - a_2\phi^3 + s_1\phi dW_t$	Poor param. scaling
	Nonlinear term	
$\beta = 1$	–	
$\beta = 5/4$	$1 + a_1\phi - a_2\phi^{5/4} + s_1\phi^{5/4} dW_t$	Parameters unidentifiable
$\beta = 3/2$	$1 + a_1\phi - a_2\phi^{3/2} + s_1\phi^{3/2} dW_t$	Parameters unidentifiable
$\beta = 2$	$1 + a_1\phi - a_2\phi^2 + s_1\phi^2 dW_t$	Parameters unidentifiable
$\beta = 3$	$1 + a_1\phi - a_2\phi^3 + s_1\phi^3 dW_t$	Parameters unidentifiable
	Other diffusion types	
$\beta = 3$	$1 + a_1\phi - a_2\phi^3 + \sqrt{s_1 + s_2\phi^2} dW_t$	No scaling function for a_2
	Amplitude equation of Stuart-Landau type	–
	$b + a_0\phi + \phi \sum_{k=1}^2 a_k \phi ^{2k} + \sqrt{s_1 + s_2\phi^2} dW_t$	Unstable after identification