Freie Universität Berlin

Enhancing the Upper-Air Observational Temperature Record to Improve Satellite Validation and Weather Forecasts

Jordis S. Tradowsky

February 2018

Doctoral thesis

Institute of Meteorology
Department of Earth Sciences
Freie Universität Berlin

Bodeker Scientific

NIWA ATMOSPHERIC RESEARCH STATION
LAUDER
Since 1961
NDACC-BSRN-TCCON-GRUAN
NIWA
Taihoro Nukurangi
www.niwa.co.nz
45.038°S, 169.684°E, 370 m

Supervisor (Gutachter) at Freie Universität Berlin, Germany

Prof. Dr Peter Builtjes (Erstgutachter)

Prof. Dr Jürgen Fischer (Zweitgutachter)

Supervisor at Bodeker Scientific, New Zealand

Dr Greg Bodeker

Supervisor at National Institute of Water and Atmospheric Research, New Zealand

Dr Richard Querel

Acknowledgement

During the course of this thesis, I got supported by various people - scientist who taught me the required scientific skill; technicians who trained me in conducting reference-quality measurements; and colleagues and friends that made me feel welcome in a new environment. Furthermore, I got the full support of my family, who always encourage me to make my dreams come true, even if this means that I am living on the other side of the world. I really appreciate this and would like to thank you very warmly.

I have worked at several institutions and would like to name colleagues from each of them, but there are far more people whose support I appreciated.

I would like to thank my supervisor and mentor Greg Bodeker from Bodeker Scientific in New Zealand, for the weekly discussions, the support and the trust he gave me throughout my PhD. It enabled me to work with many of the outstanding scientists named below. Furthermore I would like to thank the whole team at Bodeker Scientific; Stefanie Kremser, Jared Lewis, Jono Conway, Chris Cameron, Nicky Brookes, and Laura Revell for the discussions, the support and the fun we had.

I would like to thank the whole team of the New Zealand National Institute of Water and Atmospheric Research at Lauder, and specifically my supervisor Richard Querel and my colleague Wills Dobson, for their support and the skills they taught me. Performing the weekly balloon launches with you has always been a pleasure.

I would like to thank my professors at Freie Universität Berlin, Germany, Peter Builtjes and Jürgen Fischer for their supervision through the course of my thesis.

I would like to thank the Satellite Applications team of the UK Met Office. Specifically, I would like to thank Chris Burrows, for the excellent supervision and his expertise, Sean Healy (ECMWF) for his expertise and the useful discussions and John Eyre, who mentored the project and supported me with his vast expertise. Furthermore I would like to thank Bruce Ingleby (ECMWF), Bill Bell, Ian Culverwell, and Axel von Engeln (ECMWF) for the fruitful discussions during this project. This project was organised as a Radio Occultation Meteorology Satellite Application Facility (ROM SAF) Visiting Scientist Project and I would like to thank the ROM SAF for their support.

I would also like to thank Karin Kreher who proofread parts of the thesis.

The funding for the doctoral thesis was granted by the German Academic Exchange Service (DAAD), Bodeker Scientific, and the New Zealand National Institute of Water and Atmospheric Research.

Summary

This thesis describes the development of novel methods to enhance the upper-air record of temperature with the goal of improving weather forecast and serving the satellite validation community. The thesis begins by demonstrating a new method to enhance the value of global radiosonde (RS) and radio occultation (RO) observations in numerical weather prediction (NWP) models. The space-based RO[1] technique provides highly accurate, globally distributed profile measurements in all weather conditions. These homogeneous RO data are used to bias correct stratospheric RS temperatures from the operational observing network, which uses various RS types, prior to their data assimilation into an NWP model. Such a bias corrections improves the consistency between the observation types which is especially important for RO and RS as both observation types anchor the temperature in NWP models.

This method is then applied to study, in detail, the differences between RO measurements and the reference-quality RS data products from the GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN) at 6 upper-air sites. In contrast to RS data from the operational network, GRUAN processed RS data are corrected for all known biases and provide an uncertainty estimate with every value.

In a follow-on to this piece of work, the upper-air temperature time series at the GRUAN site in Lauder, New Zealand, is improved by combining measurements made at two locations into a Site Atmospheric State Best Estimate (SASBE) of the temperature. The construction of the data set is described in detail, including the estimation of uncertainties on every value of the SASBE.

An imperative for creating homogeneous time series of upper-air measurements is careful management of transitions from one instrument type to another. Therefore, a final study reported on in this thesis tests if the bias between two different RS types can be estimated from interlaced measurements, i.e. a time series of profiles obtained by alternating between two RS types from day to day (one instrument measuring on even dates and the other one on odd dates).

The following scientific questions are addressed in this thesis and in the associated publications which are part of this thesis:

1. How can the spatial and temporal homogeneity of one measurement technique or one instrument type be used to homogenise measurements from other sensors?
2. How can the *a priori* information required as input to RO retrievals be reduced to diminish the dependence of derived products on *a priori* information?
3. How can the structural uncertainty in the new tangent linear RO retrieval presented here be assessed?
4. How do RO measurements compare to the GRUAN RS data product in the stratosphere?
5. How can measurements from distributed, i.e. not perfectly collocated, measurement sites be combined?

[1] All abbreviations can be found in the Glossary, Appendix B

6. How can the difference in biases between two RS types be assessed?

These questions are addressed in this thesis and in the peer-reviewed publications listed below which are also part of this thesis.

- Tradowsky, J.S. (2015), Characterisation of radiosonde temperature biases and errors using radio occultation measurements. *Radio Occultation Meteorology Satellite Application Facility Visiting Scientist Report 26*. Available at http://www.romsaf.org/Publications/reports/romsaf_vs26_rep_v12.pdf, (Tradowsky 2015).
- Tradowsky, J.S. (2016), Radiosonde Temperature Bias Corrections using Radio Occultation Bending Angles as Reference. *Radio Occultation Meteorology Satellite Application Facility Visiting Scientist Report 31*. Available at http://www.romsaf.org/Publications/reports/romsaf_vs31_rep_v10.pdf, (Tradowsky 2016).
- Tradowsky, J.S., G.E. Bodeker, P.W. Thorne, F. Carminati, and W. Bell (2016). GRUAN in the service of GSICS: Using reference ground-based profile measurements to provide traceable radiance calibration for space-based radiometers. *Global Space-based Inter-Calibration System Quarterly Newsletter* 10.2. doi:10.7289/V5GT5K7S, (Tradowsky et al. 2016).
- Tradowsky, J.S., C.P. Burrows, S.B. Healy, and J.R. Eyre (2017). A New Method to Correct Radiosonde Temperature Biases Using Radio Occultation Data. *Journal of Applied Meteorology and Climatology* doi: 10.1175/JAMC-D-16-0136.1, (Tradowsky et al. 2017).

While the content of these publications is part of this doctoral thesis, not all publications are reprinted here in full in the interest of brevity. The interested reader is referred to the listed publications for further details.

Additionally two articles, listed below, are currently submitted or under review in *Earth System Science Data Discussions* and *Atmospheric Measurement Techniques Discussions*, respectively, and will be reprinted in slightly adapted form here.

- Tradowsky, J.S., G.E. Bodeker, R.R. Querel, P.J.H Builtjes, and J. Fischer (2018). Combining Data from the Distributed GRUAN Site Lauder-Invercargill, New Zealand, to Provide a Site Atmospheric State Best Estimate of Temperature. Paper submitted to *Earth System Science Data Discussions* (February 2018, Tradowsky et al. 2018, reference added prior to the official publication of thesis).
- Kremser, S., J.S. Tradowsky, H.W. Rust, and G.E. Bodeker (2018). Is it feasible to estimate radiosonde biases from interlaced measurements? Paper in review at *Atmospheric Measurement Techniques Discussions*, (Kremser et al. 2018a, under review). Final version published as Kremser et al. (2018b) (reference added prior to the official publication of thesis).

This thesis is motivated by the international research and NWP communities need to make use of the limited upper-air measurements available. While there are more than 800 upper-air sites launching RSs worldwide, improvements in accuracy, homogeneity, and temporal resolution of the data are needed to make them best-suited for, among other things, NWP and validation purposes. To assimilate data into an NWP model, it is desirable to eliminate biases between different observations. RS temperature biases can result in erroneous temperature fields which lead to spurious features in the wind field. Furthermore, for the validation of space-based instruments, high-quality data, including

traceable uncertainty estimates, are required. To increase the number of collocations with space-based measurements, high temporal resolution of the validation data set is desirable.

This doctoral thesis, which supports the activities of GRUAN, focusses on treating uncertainty estimates as an imperative rather than as a supplement. Measurement uncertainties, as well as structural uncertainties, and representativeness uncertainties are therefore discussed here.

To address the first scientific question posed, a novel method to correct RS temperature biases using RO measurements as an unbiased reference has been developed and is presented here. Such a bias correction improves the use of globally distributed RS temperature profiles within NWP as they are spatially homogenized using homogeneous RO data. An NWP model is used as a transfer standard to minimise effects of imperfect collocation between RS and RO measurements. This new double-differencing method to correct the RS record was developed in cooperation with colleagues from the United Kingdom Met Office and the European Centre for Medium-Range Weather Forecasts (ECMWF). The author of this thesis spent 4.5 months working at the Met Office as part of two EUMETSAT[2] Radio Occultation Meteorology Satellite Application Facility visiting scientist projects. A first trial of the method in the Met Office NWP model indicates promising results for temperatures and heights, but mixed results for other parameters. Prior to implementation of any method into an NWP model extensive testing of suggested changes is required to ensure reliability of the forecasting system. Therefore, a detailed forecast impact study is planned at the Met Office to carefully evaluate effects of implementing this novel bias correction approach. Once such a forecast impact study has been performed, the method may be implemented in NWP model to improve global weather forecasts.

As part of the development of the method described above to correct RS biases in NWP models, also a new tangent linear RO retrieval has been developed, addressing the second research question. While the conventional RO retrieval requires smoothing with a climatological bending angle profile above about 40 km altitude, the tangent linear retrieval does not include such *a priori* knowledge. Thus, a method to significantly reduce *a priori* knowledge required in RO retrievals is presented as part of this thesis. This retrieval may also prove valuable to study changes in climate based on RO bending angle statistics as it can be applied to propagate changes detected in stratospheric bending angles into temperatures.

An important part of any retrieval should be the assessment of structural uncertainties, which are caused by inadequacies in retrievals. To address this third scientific question, a method to estimate structural uncertainties in the new tangent linear RO retrievals is demonstrated. The method allows to estimate regional or global structural uncertainties. The upper limit of structural uncertainties in the tangent linear RO retrieval for global measurements is estimates as approximately 0.4 K at 10 hPa, which decreases to below 0.2 K at pressures above 30 hPa. These estimates compare well with structural uncertainties estimated by comparison of conventional RO retrievals from different retrieval centres. The presented method to estimate uncertainties in the tangent linear RO retrieval, however, enables estimating structural uncertainties depending on the desired application, and

[2]European Organisation for the Exploitation of Meteorological Satellites

the structural uncertainty can be significantly smaller than the upper limit of uncertainties given above for specific applications. For example, the highest structural uncertainties per region is between 0.078 K and 0.192 K for comparison between GRUAN and RO data.

Addressing the forth scientific question posed above, this thesis demonstrated how the double-differencing method developed in question 1 is applied to compare RO data with a reference-quality GRUAN RS data product, taking GRUAN measurement uncertainties, as well as RO sampling and structural uncertainties into account. At most pressure levels GRUAN and RO data are consistent or in agreement with each other, but inconsistencies at some levels suggest further research.

In response to the fifth scientific question, a novel method to combine upper-air measurements from two sites is demonstrated. The RS temperature measurements performed at the distributed GRUAN site Lauder-Invercargill in New Zealand are combined into a temperature SASBE. A regression model is used to transfer knowledge from RSs launched at Invercargill to the vertical column above Lauder. In the absence of any measurements at Lauder or Invercargill, the SASBE is based on a diurnal temperature cycle which is obtained by fitting Fourier series to 7 years of hourly ERA5[3] temperature reanalysis data which has been interpolated to Lauder. The SASBE includes an estimate of the uncertainty on every data point, which is calculated by (i) propagating uncertainties on the input variables through all calculations, (ii) accounting for the uncertainty introduced by the transfer algorithm, and (iii) taking the representativeness uncertainty in the ERA5 reanalysis into account. The temperature profiles are available at hourly resolution for the years 1997-2012 at 16 vertical levels from the surface to 10 hPa. The uncertainty on the temperature is smallest in the direct vicinity of a measurement made at Lauder and increases with time distance from that measurement. It is demonstrated that including Invercargill data in the SASBE improves representativeness of the SASBE with respect to the temperatures measured at Lauder. However, further research into the uncertainty resulting from the transfer of data is suggested to reflect this improvement in the overall uncertainty on the best-estimate temperature. Consequently, the value of distributed upper-air sites, which exist at several GRUAN locations is demonstrated. The temperature SASBE has been constructed for the GRUAN site at Lauder, New Zealand, but the methodology can be implemented across GRUAN. This SASBE enhances the upper-air record of temperature above Lauder by (i) combining the data from several instruments, (ii) increasing the temporal resolution, and (ii) providing a robust estimate of the uncertainty. This SASBE can be used to validate models and space-based sensors, as well as to provide *a priori* knowledge for the retrieval of other parameters observed at Lauder.

Within GRUAN any change of instrumentation needs careful transition management to ensure the long-term homogeneity of the time series. Any change in instrumentation could introduce inhomogeneities into the time series, which may prevent the detection of changes in climate. Typically, the bias between instruments is assessed by performing dual measurements with both instruments for an overlap period. These parallel measurements can then be used to estimate biases between instruments. However, addressing scientific question 6, the feasibility of an alternative approach, where two instrument types are used at alternating days, is studied using a synthetic data set. It is shown that, for transition management of RS within GRUAN, dual launches are indispensable to confidently esti-

[3]ERA5 is the 5th generation of reanalysis by the European Centre for Medium-Range Weather Forecast

mate the bias between both instruments. However, the studied approach may be used for transition between other instruments or for inter-calibration of space-based sensors.

Zusammenfassung

Die vorliegende Doktorarbeit beschreibt neu entwickelte Methoden die zu Verbesserungen atmosphärischer Temperaturdatensätze dienen, welche wiederum die Wettervorhersage verbessern und zur Satellitenvalidierung genutzt werden können.

Die Doktorarbeit stellt zunächst eine Methode zur verbesserten Nutzung von Radiosonden- und Radiookkultationsmessungen in der numerischen Wettervorhersage vor. Die satellitengestützte Radiookkultationstechnik liefert globale und qualitativ hochwertige Profilmessungen in allen Wetterbedingungen. Dieser homogene Radiookkultationsdatensatz dient als Referenz für die Korrektur von stratosphärischen Temperaturfehlern in Radiosondenprofilen des globalen Messnetzwerkes, welches verschiedenste Radiosonden mit unterschiedlichen Fehlercharakteristiken verwendet. Die Fehlerkorrektur ist zur Anwendung vor der Assimilation der Radiosondenprofile in die numerische Wettervorhersage gedacht, welche unter Anderem auch die Übereinstimmung von Radiosonden- und Radiookkultationsmessungen verbessert.

Diese entwickelte Methode wird anschließend genutzt, um die Unterschiede zwischen Radiookkultationsmessungen und den Radiosondenmessungen eines globalen Referenzmessnetzwerkes. In der vorliegenden Arbeit werden Radiosondenmessungen von sechs Messstationen des GCOS (Global Climate Observing System) Reference Upper-Air Networks (GRUANs) Messnetzwerkes mit Radiookkultationsmessungen verglichen.

Des Weiteren wird eine neue Methode zur Kombinierung verschiedener Temperaturmessungen beschrieben, welche die Messungen von zwei Messtationen auf der Südinsel Neuseelands vereint. Diese bestmögliche Abschätzung des Temperaturprofils und der Unsicherheit in der Temperatur über Lauder, Neuseeland, führt zu einer deutlichen Verbesserung der Temperaturzeitreihe über Lauder.

Jeglicher Austausch von Messinstrumenten bedarf einer sorgfältiger Übergangsplanung um eine homogene Zeitreihe von atmosphärischen Variablen sicherzustellen. Typischerweise werden Parallelmessungen mit dem alten und dem neuen Instrument durchgeführt, um die Unterschiede in den Messungen abzuschätzen. In dieser Doktorarbeit wird untersucht, ob es möglich wäre den Unterschied in den Radiosondenmessungen mit zwei unterschiedlichen Instrumenten aus einer Datenreihe von abwechselnden Messungen abzuschätzen.

Die nachfolgenden wissenschaftlichen Fragestellungen werden in dieser Doktorarbeit und in den dazugehörigen wissenschaftlichen Veröffentlichungen behandelt.

1. Wie kann die räumliche und zeitliche Homogenität einer Messtechnik oder eines Instrumentes genutzt werden um die Datenreihe anderer Messgeräte zu homogenisieren?
2. Wie kann die *a priori* Information welche im Radiookkultationsretrieval benötigt wird, reduziert werden um die Abhängigkeit der abgeleiteten atmosphärischen Variablen (z.B. Temperatur) von der *a priori* Information zu vermindern?
3. Wie kann die strukturelle Unsicherheit in dem tangentenlinearen Radiookkultationsretrieval welches in dieser Doktorarbeit entwickelt wurde, berechnet werden?
4. Wie gut stimmen die Messungen des Referenzmessnetzwerkes GRUAN mit Ra-

diookkultationsmessungen in der Stratosphäre überein?

5. Wie können die Messungen von zwei Standorten kombiniert werden um eine bestmögliche Abschätzung des Temperaturprofils zu erhalten?
6. Wie können die Unterschiede in den Messungen zweier Radiosondentypen abgeschätzt werden?

Die folgenden wissenschaftlichen Veröffentlichungen bilden einen Teil der hier vorliegenden Doktorarbeit, werden hier jedoch nicht vollständig wiedergegeben, um die Doktorarbeit übersichtlich zu gestalten:

- Tradowsky, J.S. (2015), Characterisation of radiosonde temperature biases and errors using radio occultation measurements. *Radio Occultation Meteorology Satellite Application Facility Visiting Scientist Report 26.* Available at `http://www.romsaf.org/Publications/reports/romsaf_vs26_rep_v12.pdf`, (Tradowsky 2015).
- Tradowsky, J.S. (2016), Radiosonde Temperature Bias Corrections using Radio Occultation Bending Angles as Reference. *Radio Occultation Meteorology Satellite Application Facility Visiting Scientist Report 31.* Available at `http://www.romsaf.org/Publications/reports/romsaf_vs31_rep_v10.pdf`, (Tradowsky 2016).
- Tradowsky, J.S., G.E. Bodeker, P.W. Thorne, F. Carminati, and W. Bell (2016). GRUAN in the service of GSICS: Using reference ground-based profile measurements to provide traceable radiance calibration for space-based radiometers. *Global Space-based Inter-Calibration System Quarterly Newsletter* 10.2. doi:10.7289/V5GT5K7S, (Tradowsky et al. 2016).
- Tradowsky, J.S., C.P. Burrows, S.B. Healy, and J.R. Eyre (2017). A New Method to Correct Radiosonde Temperature Biases Using Radio Occultation Data. *Journal of Applied Meteorology and Climatology* doi: 10.1175/JAMC-D-16-0136.1, (Tradowsky et al. 2017).

Zusätzliche sind zwei Artikel zur Veröffentlichung eingereicht, beziehungsweise unter Review. Diese zwei Artikel werden in dieser Doktorarbeit in leicht abgeänderter Form wiedergegeben.

- Tradowsky, J.S., G.E. Bodeker, R.R. Querel, P.J.H Builtjes, and J. Fischer (2018). Combining Data from the Distributed GRUAN Site Lauder-Invercargill, New Zealand, to Provide a Site Atmospheric State Best Estimate of Temperature. Paper submitted to *Earth System Science Data Discussions* (February 2018).
- Kremser, S., J.S. Tradowsky, H.W. Rust, and G.E. Bodeker (2018). Is it feasible to estimate radiosonde biases from interlaced measurements? Paper in review at *Atmospheric Measurement Techniques Discussions*, (Kremser et al. 2018a, under review).

Da die Anzahl an atmosphärischen Messungen gering ist, ist es umso wichtiger für die Klimaforschung und die Wettervorhersage den maximalen Nutzen aus den vorhandenen Messungen zu extrahieren. Global existieren mehr als 800 Messstationen welche Radiosondenmessungen durchführen. Verbesserungen in der Homogenität der Messzeitreihe, der zeitlichen Auflösung und der Genauigkeit der Messungen, werden jedoch benötigt um den Nutzen dieser Messungen für die Wettervorhersage und Satellitenvalidierung zu erhöhen. Vor die Assimilation von Messdaten in Wettervorhersagemodelle ist es wünschenswert

entgegengesetzte Fehler in den Messdaten zu beseitigen. Messfehler in Radiosondenmessungen können zu Fehler im modellierten Luftfeuchtigkeits- und Temperaturfeld führen, und dies kann wiederum zu einem fehlerhaften modellierten Windfeld führen. Für die Validierung von satellitengestützten Messungen werden qualitativ hochwertige Messdatensätze benötigt, welche eine Abschätzung der Unsicherheiten beinhalten. Um die Anzahl der vergleichbaren Profilen des Referenzdatensatzes mit dem Satellitendatensatz zu erhöhen, ist eine höhere zeitliche Auflösung des Referenzdatensatzes wünschenswert.

Eine sorgfältige Abschätzung der Unsicherheiten ist ein wichtiger Teil dieser Arbeit, welche damit die Ziele GRUANs unterstützt. Messunsicherheiten, strukturelle Unsicherheiten und Repräsentativitätsunsicherheiten werden diskutiert und in die Berechnungen einbezogen.

Zur Beantwortung der ersten wissenschaftlichen Fragestellung wird eine neuartige Methode zum Korrigieren von Radiosondenfehlern dargestellt. Homogenen Radiookkultationsmessungen werden als Referenz für die entwickelte Korrektur verwendet. Solch eine Fehlerkorrektur verbessert den Nutzen der globalen Radiosondenmessungen in der Wettervorhersage. Ein numerisches Wettervorhersagemodell wird als Transferstandard genutzt um negative Effekte, die durch die räumliche und zeitliche Distanz zwischen den Messungen Zustandekommen, zu verringern. Diese neue Doppeldifferenztechnik zur Korrektur von Radiosondenfehlern wurde in Zusammenarbeit mit dem Met Office des Vereinigten Königreiches und mit dem Europäischen Zentrum für Mittelfristige Wettervorhersage (ECMWF) während eines mehrmonatigen Arbeitsaufenthalts[4] entwickelt. Erste Ergebnisse dieser neuen Methode deuten auf verbesserte Temperaturen und geopotentielle Höhen in Wettervorhersagemodellen hin. Jedoch sind die Auswirkungen auf die anderen Modellvariablen durchwachsen. Bevor eine verbesserte Methode in ein Wettervorhersagemodell implementiert werden kann, muss eine umfassende Untersuchung zu den Auswirkungen auf das Modell stattfinden, um die Stabilität des Vorhersagemodells sicherzustellen. Diesbezüglich ist eine weitere detailliertere Studie der Vorhersagegüte basierend auf der Einführung der hier beschriebenen Methode geplant. Basierend auf den Ergebnissen dieser Studie, könnte die Methode in Vorhersagemodelle implementiert werden um die globale Wettervorhersage zu verbessern.

Ein neues tangentenlineares Radiookkultationsretrieval wurde entwickelt, welches zur Beantwortung der zweiten wissenschaftlichen Frage dient. Dieses entwickelte Retrieval benötigt deutlich weniger *a priori* Informationen, als das konventionelle Radiookkultationsretrieval. Während das gemessene Biegewinkelprofil im konventionellen Radiookkultationsretrieval mit einem klimatologischen Profil vereinigt wird, ist dies im hier beschriebenen Retrieval nicht nötig. Daher kann die dargestellte Methode die Abhängigkeit der abgeleiteten Radiookkultationsvariablen von *a priori* Information deutlich verringern. Das tangentenlineare Radiookkultationsretrieval kann möglicherweise auch zur Untersuchung des Klimawandels genutzt werden, jedoch benötigt dies einer detaillierten Studie.

Um die dritte wissenschaftliche Frage zu beantworten, wurde eine Methode zur Berechnung der strukturellen Unsicherheit im tangentenlinearen Radiookkultationsretrieval entwickelt. Es können sowohl globale Unsicherheiten, als auch die Unsicherheit im Retrieval in einer bestimmten Region berechnet werden. Das Höchstmaß an struktureller Unsicherheit für globale Radiookkultationen wird als 0,4 K auf dem 10 hPa Druckniveau

[4] Organisiert durch die EUMETSAT Radio Occultation Meteorology Satellite Application Facility

abgeschätzt. Die Unsicherheit nimmt mit zunehmenden Druck ab. Diese Abschätzungen der Unsicherheit sind vergleichbar mit der strukturellen Unsicherheit im konventionellen Radiookkultationsretrieval. Die hier dargestellte Methode zur Abschätzung der strukturellen Unsicherheit ermöglicht es allerdings eine passgenaue Abschätzung basierend auf der geplanten Anwendung zu berechnen. Die Unsicherheit kann je nach Anwendung deutlich kleiner sein, als das berechnete Höchstmaß. Für den Vergleich von GRUAN- und Radiookkultataionsmessungen ergeben sich zum Beispiel strukturelle Unsicherheiten zwischen 0.078 K und 0.192 K im Radiookkultationsretrieval.

Anschließend wird, zur Beantwortung der vierten wissenschaftlichen Fragestellung, erläutert wie die Doppeldifferenztechnik genutzt werden kann um GRUAN Radiosondenprofile mit Radiookkultationsmessungen zu vergleichen. Die Messunsicherheiten, Stichprobenunsicherheit und die strukturelle Unsicherheit werden hierbei berücksichtigt. Übereinstimmung der Messungen ist auf den meisten Druckniveaus gegeben, jedoch geben Unstimmigkeiten auf bestimmten Druckniveaus Anlass zur weiteren Forschung.

Eine neue Methode zur Vereinigung der atmosphärischen Messungen an zwei verschiedenen Messstationen wurde entwickelt welche die fünfte wissenschaftliche Fragestellung behandelt. Die Radiosondentemperaturprofile der Messstationen in Lauder und Invercargill (Neuseeland) werden genutzt, um eine bestmögliche Abschätzung des Temperaturprofils über Lauder zu erstellen. Ein Regressionsmodell wird verwendet, um die Temperaturinformation von Invercargill nach Lauder zu übertragen. Je weiter der zeitliche Abstand zu einer Radiosondenmessung in Lauder oder Invercargill ist, umso mehr beruht die abgeschätzte Temperatur auf einem Temperaturtagesgang, welcher aus dem ERA5 Datensatz abgeschätzt wird. Dieses sogenannte "Site Atmospheric State Best Estimate (SASBE)" beinhaltet eine Abschätzung der Unsicherheit jedes Temperaturwertes. Diese Unsicherheiten berücksichtigen (i) die Messunsicherheiten, (ii) die Unsicherheit welche durch die Übertragung der Daten von Invercargill nach Lauder hervorgerufen wurde und (iii) die Repräsentativitätsunsicherheit im Temperaturtagesgang. Das SASBE hat eine stündliche zeitliche Auflösung und die Temperaturprofile reichen vom Erdboden bis zum Druckniveau von 10 hPa. Das SASBE wurde für die Jahre 1996-2012 erstellt und ist öffentlich erhältlich. Die Ungenauigkeit des SASBEs ist am geringsten zu Zeiten an denen eine Radiosondenmessung in Lauder vorhanden ist. Es wird gezeigt, dass die Verwendung der Radiosondenmessungen aus Invercargill die Repräsentativität des SASBEs verbessert, jedoch ist eine weiterführende Analyse der Unsicherheiten nötig, um diese Verbesserung auch in der Gesamtunsicherheit des SASBEs widerzuspiegeln. Das Temperatur-SASBE wurde für die dezentralisierte GRUAN Messstation Lauder-Invercargill erstellt, allerdings eignet sich die Methode zur Implementierung an anderen GRUAN Messstationen. Das SASBE verbessert den Temperaturdatensatz über Lauder indem es (i) die Messungen verschiedener Standorte miteinander vereint, (ii) die zeitliche Auflösung deutlich verbessert, und (iii) eine robuste Abschätzung der Unsicherheit miteinbezieht. Das SASBE kann sowohl zur Satelliten- und Modellvalidierung, als auch als *a priori* für das Retrieval anderer Datensätze für Lauder genutzt werden.

Für GRUAN als Referenzmessnetzwerk ist es unabdingbar jegliche Umstellung der Messgeräte zu koordinieren, um den Unterschied in den Messungen der beiden Geräte bestimmen zu können. Typischerweise wird der Unterschied in den Messungen aus einer parallelen Messzeitreihe bestimmt. In dieser Doktorarbeit wird jedoch untersucht, ob es möglich wäre die Unterschiede aus abwechselnden Messungen zweier Instrumente zu bes-

timmen. In Antwort auf die sechste wissenschaftliche Fragestellung kann schlussgefolgert werden, dass parallele Radiosondenmessungen für die Umstellung des Radiosondentyps an GRUAN Stationen unabdingbar sind. Die Methode zur Abschätzung von Unterschieden in den abwechselnden Messungen kann allerdings potentiell für andere Instrumente oder für die Inter-Kalibrierung von satellitengestützten Messgeräten verwendet werden.

Contents

List of Figures

List of Tables

1. Introduction

Upper-air measurements of atmospheric climate variables such as temperature and water vapour are used for weather prediction, for atmospheric process studies and to detect changes in the climate system. The upper-air comprises those areas of the atmosphere which can not be described based on surface observations alone, i.e. upper-air extends from just above the surface (about 10 m) to the Kármán line at 100 km above sea level, which builds the boundary between the atmosphere and outer space (Norbert 2013). The troposphere is the lowers layer of the atmosphere which extends from the surface to about 16 km at the equator and to about 10 km at higher latitudes. Most weather phenomena on our planet, e.g. high and low pressure systems, clouds, and rain, occur in the troposphere and temperatures decrease with altitude (see Fig. 1.1). The tropopause, at which temperatures reach their minimum, is the boundary that separates troposphere from stratosphere. In the stratosphere the temperature increases due to the absorption of ultraviolet sunlight by ozone. Maximum concentrations of ozone are found in the lower stratosphere forming the "ozone layer" which protects life on Earth from harmful radiation. Temperatures peak at the stratopause before decreasing again through the mesosphere. Above the mesosphere is the thermosphere, in which particles are ionised by the radiations from the sun. This layer of the atmosphere is referred to as the ionosphere and it is important for the transmission of radio signals as it acts like a mirror that bounces the signal back towards the Earth and thereby enables long-distance radio communications. The layer below the ionosphere is also known as the neutral atmosphere. The different layers of the atmosphere and their temperature structure, as well as the altitude, pressure, density, and the molecules per square centimetre are shown in Fig. 1.1.

The upper-air is monitored with either *in situ* measurements or remote sensing measurements. An *in situ* measurement, Latin for "in position", is made at the location of the instrument. For example, to measure an atmospheric temperature profile *in situ*, the instrument needs to be carried upwards through the atmosphere by e.g. a balloon, to measure the temperature along its path. In contrast, remote sensing techniques, measure the state of the atmosphere from a distant position and atmospheric climate variables are inferred from the measurement.

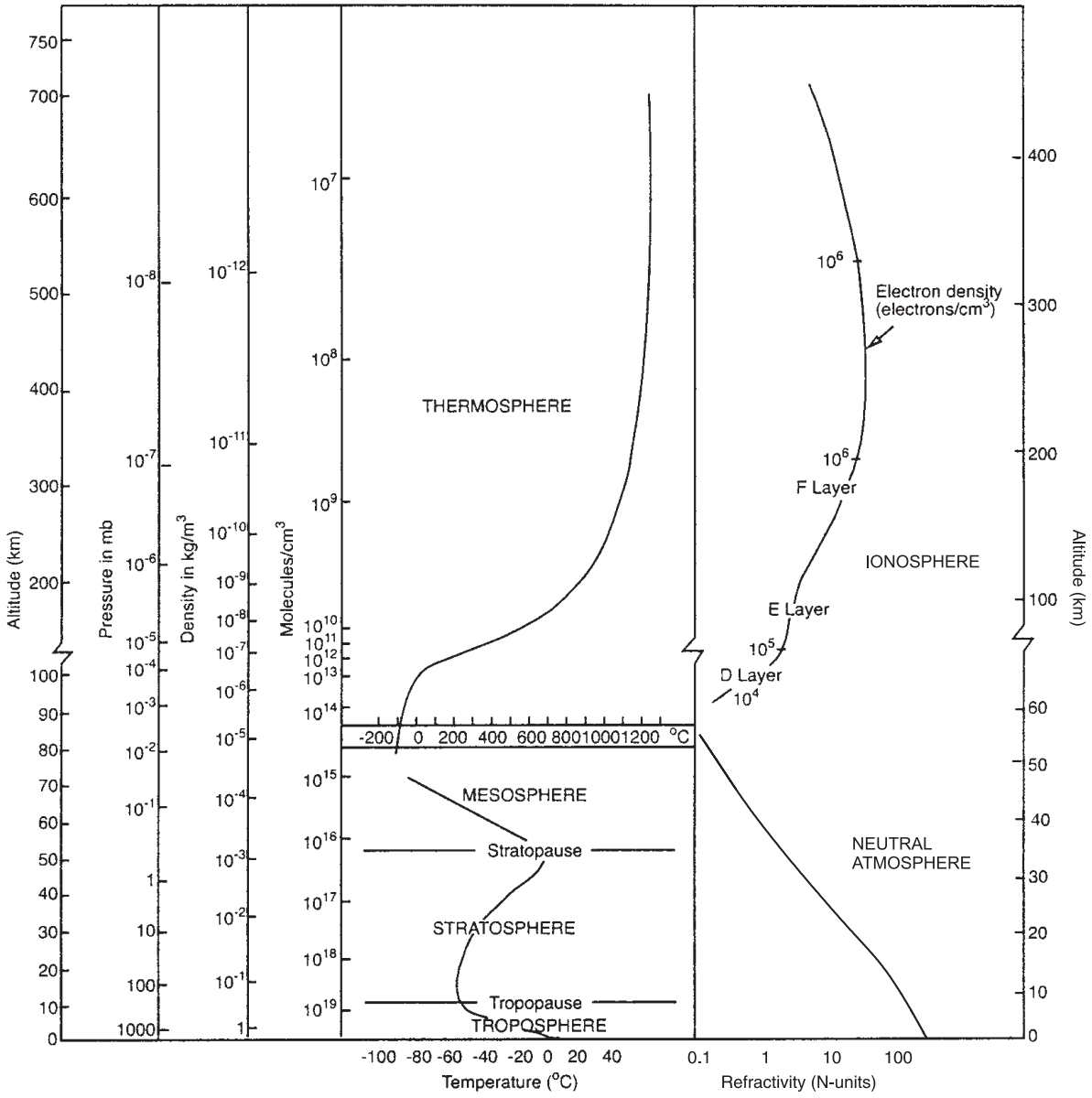


Figure 1.1.: The terminology of atmospheric layers and levels together with the temperature structure through the atmosphere. Also given are altitude, pressure, density, the molecules per square centimetre, the electron density, and the refractivity. Figure adapted from Syndergaard (1999).

1.1. *In situ* observations of the upper-air

In situ upper-air observations are made with balloon-borne sondes, long-duration high altitude balloons, aircraft, drones, or sounding rockets. The release of an atmospheric measurement balloon is referred to as a "launch" (Fig. 1.2). The balloon is filled with a lifting gas, typically helium or hydrogen. Below the ballon is a parachute which ensures that the payload is decelerated after the balloon burst. Between the parachute and the balloon is a string which unwinds to about 50 m following the launch of the payload. This distance between the instrument and the parachute/balloon minimises the effect of disturbing the atmosphere with the balloon prior to the measurement.

Figure 1.2.: Balloon launch at the atmospheric research facility in Lauder, New Zealand. ©Otago Daily Times.

The most commonly used *in situ* instrument for atmospheric measurements is a radiosonde (RS), which measures temperature, humidity, and depending on the instrument, pressure. RS are are launched daily at more than 800 upper-air sites (Ingleby 2017), mainly in support of numerical weather predictions (NWPs), and can reach altitudes of up to 40 km, depending on the balloon size. Many modern RSs use the Global Navigation Satellite System (GNSS) to determine their position, wind speed, wind direction, and pressure.

Figure 1.3 shows the commercially available RS[5] type *Vaisala RS92*, which consists of a temperature, pressure, and two humidity sensors. Furthermore, it has two antennae, one GNSS antenna for the positioning and one antenna to transfer the data to the ground station, which is located in the area where the balloon was launched.

In addition to RS, atmospheric research facilities are launching specific instruments to

[5]All abbreviations can be found in the Glossary, Appendix B

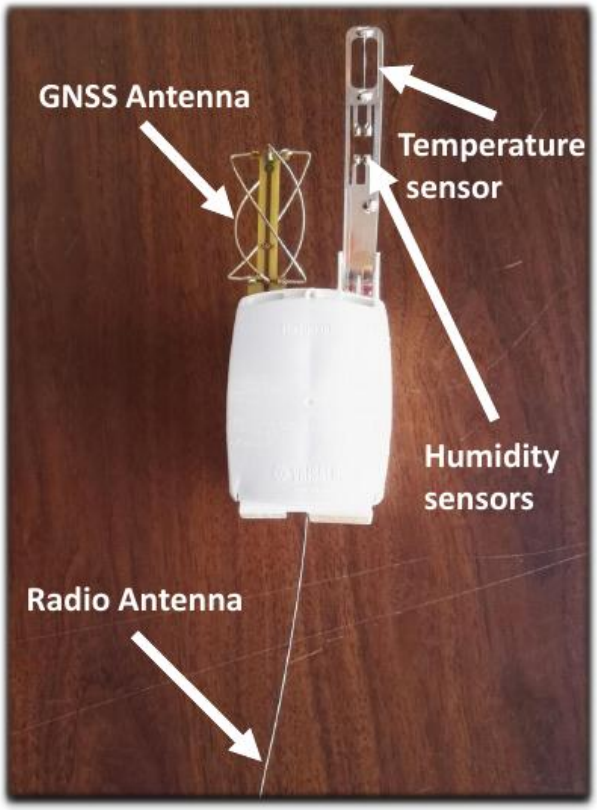


Figure 1.3.: Vaisala RS92 radiosonde instrument.©Wills Dobson.

measure ozone concentrations (ozonesonde; Bodeker et al. 1998), stratospheric water vapour (frost point hygrometer; Vömel et al. 2007; Hurst et al. 2014) or aerosol content of the atmosphere (backscatter sonde, Beyerle et al. 2001). The word "sonde" is commonly used for balloon-borne instrumentation, either for individual instruments or for a payload consisting of several instruments. Figure 1.4 shows a balloon payload launched at Lauder, New Zealand, including three different RSs, a frost point hygrometer, and an ozonesonde.

Prior to the development of modern RS that transmit information through radio signal to a ground station, manned balloons and tethered meteorological kites were used to observe selected climate variables in the upper-air. The record altitude reached by a tethered kite was 9750 m (Seidel et al. 2009).

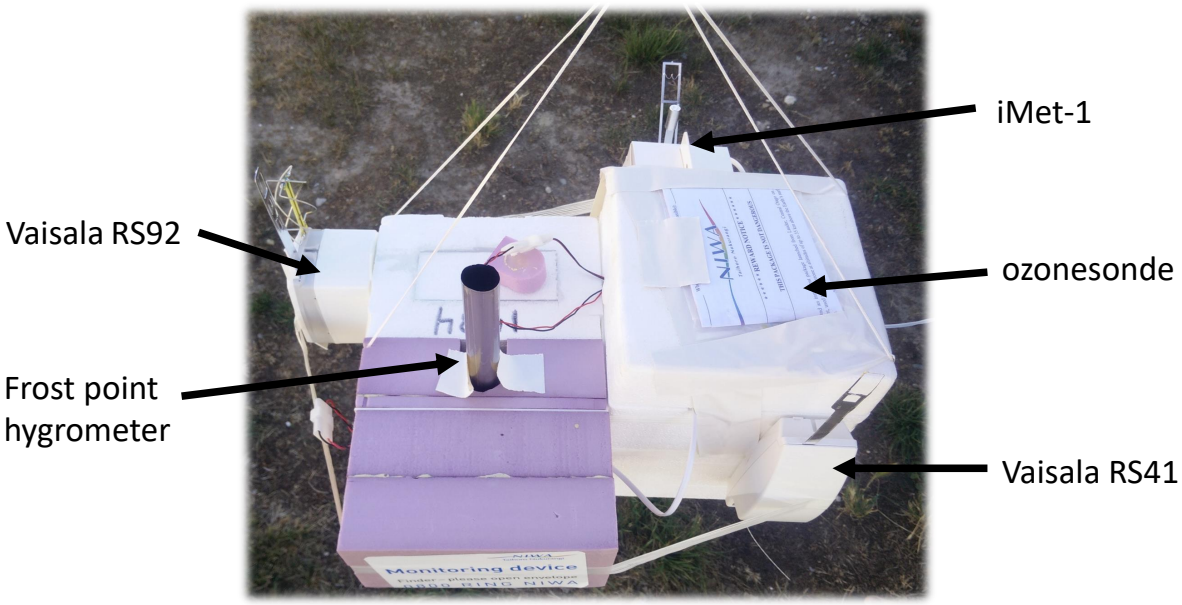


Figure 1.4.: Balloon payload with a frost point hygrometer, an ozonesonde, and three radiosondes, i.e. Vaisala RS41, Vaisala RS92, and iMet-1.

1.2. Remote sensing of the upper-air

Another source of upper-air observations are instruments that remotely sense the atmosphere, i.e. information about the upper-air is collected without making physical contact with the probed area of the atmosphere. Remote sensing techniques do not measure atmospheric variables directly, but it is possible to infer information about the state of the atmosphere from the raw measurement, such as radiance (radiometers; e.g. Aumann et al. 2003) time delay of a signal (radio occultation; e.g. Kursinski et al. 1997), or absorption spectra (Dobson spectrophotometer Dobson 1968). A retrieval algorithm (Rodgers 2000) is required to derive the vertical distribution of atmospheric climate variables such as temperature from the raw measurement. The retrieval algorithm is used to process measurements, their uncertainty informations, the relationship between the measurements and the unknown state of of the target variable, as well as initial information (*a priori*) to estimate the variable of interest.

There are ground-based remote sensors looking upwards, and space-based remote sensors looking downwards. Numerous ground-based instruments are in use globally to monitor, e.g. humidity and temperature (Hewisen 2007), atmospheric constituents such as ozone (Parrish et al. 1992) and aerosols (Uchino et al. 1995), as well as total column ozone measurements (Dobson 1968).

To derive atmospheric climate variables from ground-based or satellite-based measurements a retrieval of some sort is requires. For example, radiometers such as the Infrared Atmospheric Sounding Interferometer (IASI Aumann et al. 2003) and the Atmospheric Infrared Sounder (AIRS Siméoni et al. 1997), measure the top-of-the-atmosphere (TOA) radiance from which upper-air temperature or water vapour can be obtained by using a

radiative transfer model (see e.g. Aires et al. 2002). The GNSS radio occultation (RO) method (Kursinski et al. 1997; Kursinski et al. 2000) retrieves *inter alia* bending angle, temperature and water vapour profiles from the measured time delay of a radio signal. The RO technique provides measurements with global coverage and in all weather conditions (Anthes 2011). An occultation event occurs when the radio signal transmitted by a GNSS satellite passes through the atmosphere, where the signal is bent before being received by a low Earth orbit (LEO) satellite (see Fig. 1.5). The raw measurement is the time delay of the radio signal, which has been caused by the bending in the atmosphere. From the measured time delay, bending angle, refractivity, temperature, and humidity can be derived. However, every step in the retrieval chain, from the time delay to the humidity, requires additional *a priori* information. Thus, while the bending angle profile mainly consists of measured information, an increasing amount of *a priori* information constrains the refractivity, temperature and humidity retrieval. This is further discussed in Chapter 3.

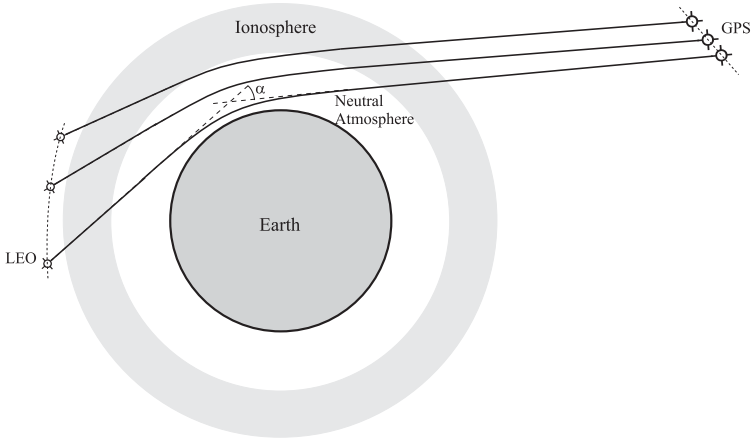


Figure 1.5.: The radio occultation technique. Figure from Syndergaard (1999).

The need for *a priori* information in the retrieval implies that structural uncertainty is inherent, i.e. depending on the choice of *a priori*, the retrieval will produce a different profile of, e.g. temperature for the same measured time delay. Within the RO community efforts are ongoing to characterise this structural uncertainty, see e.g Ho et al. (2009), Ho et al. (2012), Steiner et al. (2011), and Steiner et al. (2013). Within this thesis an alternative method to estimate the structural uncertainty is presented, which was developed during a visiting scientist project within the Radio Occultation Meteorology Satellite Application Facility. This new method is described in detail in Chapter 3 as well as in Tradowsky (2016) and Tradowsky et al. (2017).

1.3. Use of upper-air observations in numerical weather predictions

Numerical weather prediction (NWP) models forecast the weather based on the mathematical description of atmospheric circulation and weather phenomena. Measurements of the current state of the atmosphere are ingested into the model to provide boundary

conditions on which the prediction is based. The process of ingesting data into an NWP model is referred to as data assimilation. *In situ* and remote sensed observations of the upper-air are extensively used in NWP. Most atmospheric observations require correction of systematic errors, i.e. biases. Such bias corrections can be part of the assimilation system, a process called variational bias corrections (see e.g. Derber and Wu 1998; Dee 2004). However, to apply variational bias correction in a forecasting model, sufficient anchor measurements are required (Eyre 2016). Anchor measurements provide a reference for the assimilation system in which the other measurements are bias corrected. Consistency between the observations that anchor a data assimilation system is important for the stability of the assimilation system.

RS and RO measurements anchor temperatures in NWP models. Therefore, the consistency between these two observation types is essential. However, while the RO data record is spatially and temporally homogeneous, temperature biases in RS measurements depend, *inter alia*, on instrument type and the daytime at which the temperature was measured. Such biases between different RS can lead to spurious features in the modelled temperature field which, in turn, has the potential of causing spurious features in the modelled wind field. Therefore, the homogenization of anchor measurements prior to their assimilation into NWP models has the potential to improve NWPs. Within this thesis, a new method to correct stratospheric RS temperature biases using homogeneous RO measurements as a reference has been developed and is presented in Chapter 3 as well as in Tradowsky (2015), Tradowsky (2016), and Tradowsky et al. (2017).

1.4. Use of upper-air observations in climate research

Historically upper-air observation were primarily used for NWP and process studies. For NWP the amount of data available is of primary importance, while the total accuracy and the long-term homogeneity is secondary. In recent decades, upper-air observations are additionally used to investigate changes in climate. Unfortunately, most upper-air records are not best-suited to detect climate change (Immler et al. 2010). This is caused by, in part, a lack of instrument calibration, undocumented changes in instrumentation, and relocation of measurement sites, all of which can introduce offsets and shifts in the long-term measurement time series. Extensive effort was invested in the homogenisation of upper-air records of RSs (Thorne et al. 2005; Haimberger 2007; Dai et al. 2011; Haimberger et al. 2012). However, it was recognised that there is a need for long-term, high-quality data records of upper-air climate variables (Trenberth et al. 2002). These measurements should be traceable to the international system of unites (SI, Système International d'Unités; NIST 2008). Metrological traceability refers to the "property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty" (JCGM/WG 1 2012).

To provide guidance for those who need to concentrate their efforts on a limited set of critical variables, the Global Climate Observing System (GCOS) developed the concept of "essential climate variable" (ECVs), which are physical, chemical, or biological variables that are essential to the understanding of Earth's climate system (Bojinski et al. 2014). The current list of ECVs is available in GCOS-200 (2016), however, further variables can

be added if they comply with the criteria of relevance, feasibility and cost effectiveness. To be classified as an ECV, a variable must be critical to the understanding of changes in climate; it must be feasible to observe the variable on a global scale; and making measurements must be affordable (Bojinski et al. 2014). ECVs are defined for the atmosphere, the ocean and terrestrial areas. In this thesis, atmospheric ECVs, which are further divided into (i) surface ECV such as air temperature, water vapour, pressure, (ii) upper-air ECVs such as temperature, water vapour, wind speed and direction, radiation budget, and (iii) ECVs concerning the composition of the atmosphere such as carbon dioxide, methane, ozone, will be used.

The GCOS (Global Climate Observing System) Reference Upper-Air Network (GRUAN; Seidel et al. 2009; Immler et al. 2010; Bodeker et al. 2016) was established to provide long-term reference quality data sets of ECVs suitable for detecting changes in climate. Within GRUAN, reference-quality means that the traceability of the calibration chain is maintained, the measurement uncertainty is estimated for every measured value, and GRUAN data products are corrected for all known biases (Immler et al. 2010). As the understanding of the measurement and its bias is changing over time, an essential aspect of GRUAN is the long-term storage of the raw measurements together with a wide range of metadata, providing the opportunity to reprocess.

At the beginning of 2018, 28 upper-air sites have joined GRUAN (Fig. 1.6). Nine of those sites have already passed the stringent certification procedure required to be part of the network, and others are in the process of getting certified. A first GRUAN data product, which accounts for all known biases and includes an estimate of the uncertainty on every datum, has been developed for the Vaisala RS92 instrument. GRUAN data products for other RS types, ozonesondes, frost point hygrometers, and GNSS precipitable water vapour measurements are currently under development. Atmospheric upper-air measurements from GRUAN have been used in numerous studies such as Bodeker and Kremser (2015), who describe how GRUAN data can be used to analyse temperature trends, Ladstädter et al. (2015) who compare GRUAN temperature profiles with Vaisala RS92 temperature profiles, and RO retrievals of temperature, and Noh et al. (2016) who compare temperature and water vapour profiles from reanalysis, and the UK Met Office NWP Model with GRUAN RS92 profiles. Furthermore a software has been developed, known as the GRUAN processor (Carminati et al. 2016), that uses GRUAN profiles to simulate TOA radiances and brightness temperatures, including uncertainty estimates. The GRUAN processor will be used to characterise uncertainties in NWP models of the UK Met Office and ECMWF.

Building a reference-quality data record of upper-air observations suitable to detect changes in climate is an ambitious task and GRUAN faces many challenges. Five task teams have been established within GRUAN. One task team is, among other things, evaluating how measurements from different sources can be combined into best-estimate data products. Within GRUAN a cluster of instruments can build a distributed GRUAN site (GCOS-170 2013). However, when combining measurements taken at those decentralized sites, measurement uncertainties and the uncertainties resulting from transferring data to a different location need to be considered.

A novel method has been developed to combine temperature measurements made at a distributed GRUAN site into a Site Atmospheric State Best Estimate (SASBE; Tobin

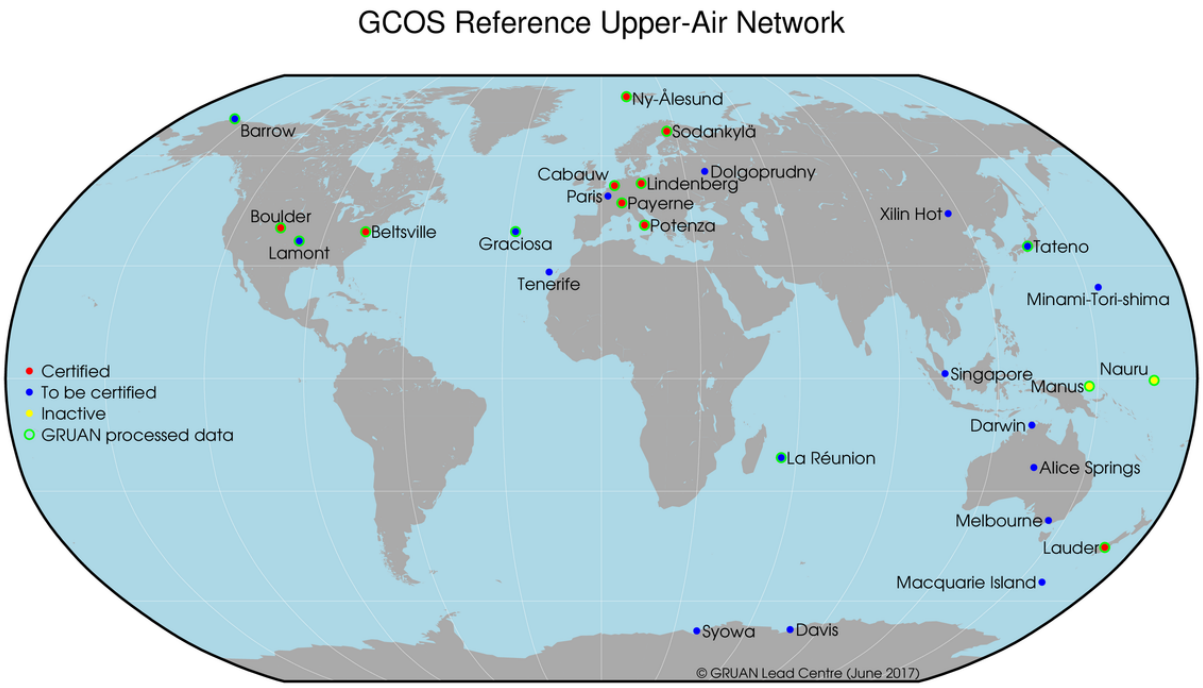


Figure 1.6.: Upper-air sites which are part of GRUAN as of June 2017. From https://www.gruan.org/network/sites/, accessed February 2018.

et al. 2006). The temperature SASBE for Lauder, New Zealand combines weekly RS measurements above Lauder, with 12-hourly RS measurements made at Invercargill, New Zealand, and a diurnal temperature cycle that has been obtained from hourly ERA5[6] reanalysis data (Hersbach and Dee 2016).

Another challenge for GRUAN, that is tackled by the radiosonde task team, is how to best manage the transition to a new radiosonde type. Changes from old to new instrumentation need to be well-documented and all biases between the old an new instrumentation need to be assessed to prevent step changes in the measurement time series (see e.g. Jensen et al. 2016). Over the recent years, GRUAN sites have changed from the Vaisala RS92 to the new Vaisala RS41 instrument. Dual RS launches were performed at sites that introduce instrument changes, and the measurement were provided to a GRUAN-wide archive of dual launches. Currently there are some studies investigating the biases between the new and old instrument types (e.g. Jensen et al. 2016). To perform dual launches a second Vaisala ground station is required and twice as many RSs are needed, increasing the costs for the measurement site. Therefore, theoretical methods to estimate the biases between two instrument types may be useful. Within this thesis an alternative approach to estimate biases between instrument types is discussed.

To maintain a reference-quality record of upper-air ECVs the measurements need to be performed to the highest possible standard. GRUAN is setting new standards for the operation at reference-quality upper-air sites, which can also provide guidance to other upper-air network, such as the GCOS Upper-Air Network (GUAN) and the wider operational upper-air network which includes all measurement sites operated by national

[6]ERA5 is the 5th generation of reanalysis by the European Centre for Medium-Range Weather Forecast, see also Section 5.1.3 and Hersbach and Dee (2016).

weather services.

1.5. Scientific questions addressed in this thesis

New methods to improve the homogeneity, temporal resolution, and uncertainty characterization of upper-air measurements on a global and local scale have been developed to address the following six science questions:

1. How can the spatial and temporal homogeneity of one measurement technique or one instrument type be used to homogenise measurements from other sensors?

   A novel method to correct RS temperature biases using RO measurements as a reference has been developed to improve global upper-air data used within NWP models. Making use of the spatial homogeneity and stability of RO measurements, stratospheric RS temperature measurements are homogenized prior to their assimilation into NWP models. This has the potential to improve NWPs, as opposing biases in assimilated temperatures can lead to spurious features in the modelled fields. This new method is described in Chapter 3 and in several publications which build part of this thesis (Tradowsky 2015; Tradowsky 2016; Tradowsky et al. 2017).
2. How can the *a priori* information required as input to RO retrievals be reduced, to diminish the dependence of derived products on *a priori* information?

   A new method to derive information about the temperature from RO measurements using a tangent linear retrieval method is described in Chapter 3, and in the publications Tradowsky (2016) and Tradowsky et al. (2017). This method is less dependent on *a priori* information compared to the conventional RO retrieval.
3. How can the structural uncertainty in the tangent linear RO retrieval be assessed? How does the estimated structural uncertainty compare with structural uncertainty estimates from conventional RO retrievals?

   A novel approach to estimate the structural uncertainty in the tangent linear RO retrieval has been developed and is presented in Chapter 3, Tradowsky (2016) and Tradowsky et al. (2017).
4. How do RO measurements compare to the GRUAN RS data product in the stratosphere? Due to their high quality, RO and GRUAN data are commonly used for comparison with measurements from other sources. But how well do the measurements from these two observation types agree with each other? A comparison of these entirely independent observations can reveal measurement biases and inadequacies in the RO retrieval or in the GRUAN processing.

   The method to estimate and correct biases in the RS temperature measurement using RO measurements as reference (Question 1), can also be used to compare stratospheric GRUAN and RO profiles. Such a comparison is presented for six GRUAN sites in Chapter 4 and in Tradowsky (2016).
5. How can measurements from distributed, i.e. not perfectly collocated, measurement sites be combined?

   A methodology to combine measurements from a distributed GRUAN site in a temperature SASBE for Lauder, New Zealand, has been developed and is presented

in Chapter 5 and Tradowsky et al. (2018, submitted).

The temperature SASBE is available at 16 levels from the surface to 10 hPa for the years 1997 to 2012. An estimate of the uncertainty is included on every data point. The following uncertainty estimate includes (i) measurement uncertainties, (ii) uncertainties resulting from the transfer of information to another location, and (iii) the representativeness uncertainty in the ERA5 diurnal cycle. The methodology is intended to be implemented across GRUAN. A selection of possible applications for SASBEs is presented in Chapter 6.

6. How can the difference in biases between two RS types be assessed? One possibility to assess the bias between two RS types is through dual balloon flights.

   This thesis presents a study that investigates if the difference in temperature biases between two RSs can also be estimated from the time series of interlaced RS measurements, see Chapter 7 and Kremser et al. (2018a, under review). Interlaced measurements of RSs are performed by alternating the instruments from one day to another, i.e. instrument A is used on every even day of the year and instrument B on every uneven day of the year. Synthetic data sets are used to explore the applicability of this statistical approach to RS change management.

An overall conclusion and discussion of the results are presented in Chapter 8.

2. Uncertainty terminology

Within this thesis the measurement uncertainty terminology as used within the GRUAN[7] community is applied. This terminology is based on JCGM/WG 1 (2008). The definitions below are taken from the GRUAN Uncertainty Glossary (Bodeker n.d.), a GRUAN technical white paper whose development was lead by Greg Bodeker.

Measurand: The quantity to be measured. As its true value is unknown, it is characterised by a probability density function that describes the likelihood of any particular value occurring when an infinite number of measurements are made under identical conditions (see upper panel of Fig. 2.1). In the case of measurements in the atmosphere, this is seldom possible.

Uncertainty (U): An expression of the degree of 'doubt' in a measurement resulting from contributions by systematic errors and random errors. Where possible the uncertainty should be evaluated from experiments but can also be estimated based on other information.

Systematic error: The difference between the mean of a large number of measurements of the measurand and the true value of the measurand. This is the bias arising from systematic effects. Because the true value of the measurand is not known, the systematic error can only be estimated, e.g. through intercomparisons. Systematic errors lead to a bias in the mean of a large set of measurements such that, unlike random errors, averaging does not reduce the systematic error in the mean.

Accuracy: Closeness of agreement between the mean of a large number of measurements of a measurand and the true value of the measurand. The preference is to refer to this as the systematic error. In some situations the median of the large number of measurements may be a better estimate of the most representative value than the mean.

Standard uncertainty (u): The 1σ width (standard deviation) of the probability of a large number of measurements of the measurand once all systematic errors have been accounted for and all covariance in errors have been removed (this cannot usually be achieved in practice).

Random error: The result of stochastic variation in quantities that influence the measurement that can never be completely avoided. Averaging over a large set of measurements reduces the random error in the mean.

Once corrections have been made for systematic errors and correlated measurement errors, it is assumed that the probability density function of the measurements becomes Gaussian in shape and the mean value of that Gaussian is assumed to be the true value of the measurand (lower panel of Fig. 2.1). This is not always the case e.g. ocean surface winds retrieved from scatterometer data have a probability density function which is inherently non-Guassian. If, after systematic and correlated measurement errors have been accounted for, the probability density function becomes Gaussian in shape, the probability density function then provides a valid description of the measurement uncertainty i.e. the uncertainty on the measurement can be expressed as the standard uncertainty (u)

[7] All abbreviations can be found in the Glossary, Appendix B

which is calculated as the standard deviation of U. Provided that corrections have been made for all systematic errors, the expected value of the uncertainty (U) is zero.

Type A evaluation of uncertainty: Evaluation of a component of the measurement uncertainty via a statistical analysis of measurements of the measurand obtained under defined measurement conditions.

Type B evaluation of uncertainty: Evaluation of a component of the measurement uncertainty determined by means other than a Type A evaluation.

Since it is seldom possible to make multiple measurements of the measurand in the atmosphere at the same location and time, Type B evaluation plays a major role for determining the uncertainty of upper-air measurements made within GRUAN. In a Type B evaluation, the variance, or the standard uncertainty (u), are evaluated using scientific judgement based on all of the available information on what the structure of the probability density function of multiple simultaneous measurements would look like. GRUAN provides estimates of standard uncertainty and measurement uncertainty (U) after correction for known systematic biases, together with uncertainties resulting from signal smoothing and instrumental time-lags.

Bias: The contribution to the uncertainty arising from systematic effects.

Coverage probability: The probability that the true value of a measurand is contained within a specified coverage interval.

Metrological traceability: A property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty.

Representativeness error: Random and systematic contributions to this source of error result from sampling and collocation mismatches, horizontally, vertically and temporally. Representativeness error can only be defined relative to some 'true' scale e.g. it is necessary to decide whether the true value is a point value or a spatial mean.

Traceability: A property of the result of a measurement, or the value of a standard, whereby it can be related to stated references, usually national or international standards, through an unbroken chain of comparisons all having stated uncertainties.

Variability: The standard deviation of a set of measurements of a variable in a given temporal or spatial range. Not to be confused with the measurement uncertainty.

This uncertainty terminology will be used throughout the thesis and uncertainties are always given at the 1σ level. The 1σ uncertainty range (i.e. the mean value $\pm$ 1σ) contains 68.3% of the data, which means that 31.7% of data are expected to be outside of the 1σ uncertainty interval. 95.4% of the data are expected to be in the 2σ uncertainty range.

While an uncertainty estimate of 1σ is used here, this uncertainty can easily be expanded by multiplication with the coverage factor k, where $k = 1$ for 1σ and $k = 2$ for 2σ.

As it is not possible to repeat any of the measurements used in this thesis under defined measurement conditions, type B evaluations of the uncertainty will be used throughout the thesis.

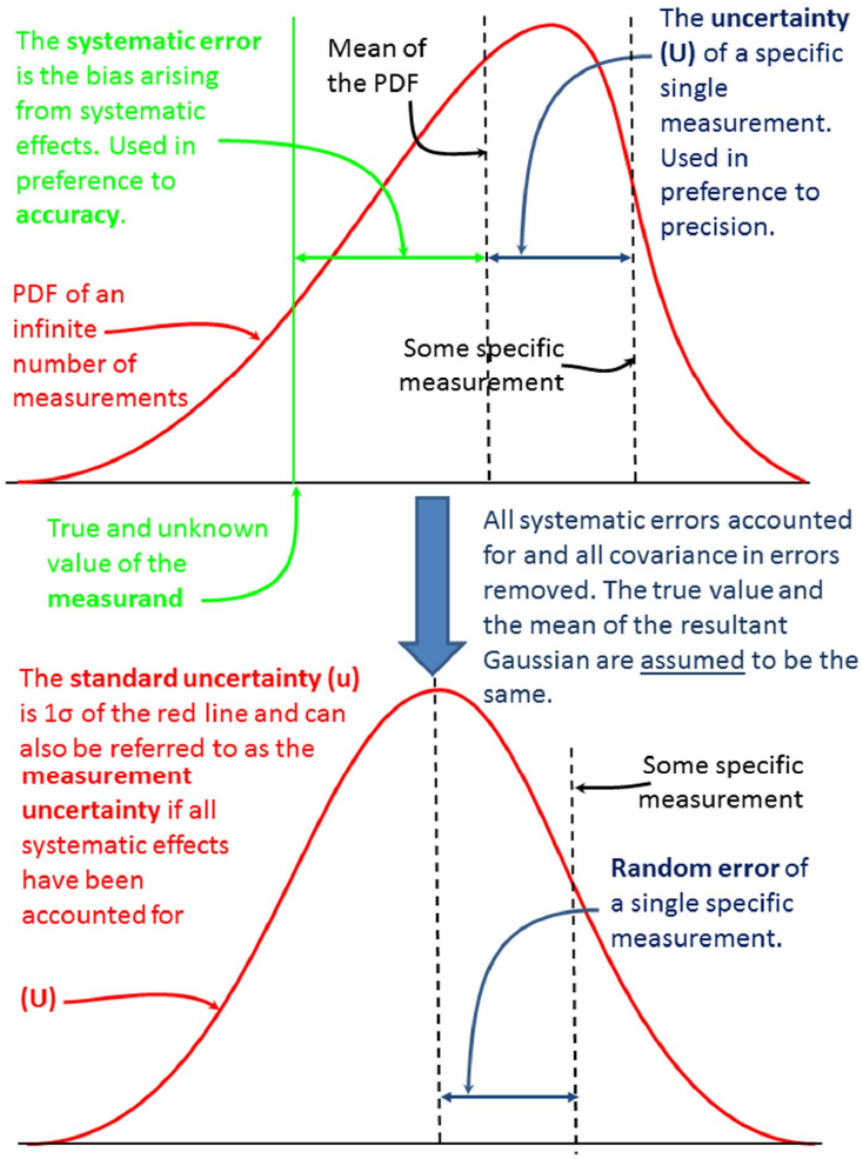


Figure 2.1.: A graphical representation of the GRUAN uncertainty lexicon. The word 'error' is reserved for those quantities pertaining directly to the (unknown) value of the measurand, while the word 'uncertainty' is reserved for those quantities that are directly observable or derivable. The large vertical arrow in the centre represents an ideal situation. In practice it is usually not possible to account for and remove all systematic errors and all covariance in errors.

3. Enhancing the value of radiosonde and radio occultation measurements in the numerical weather prediction

Upper-air measurements from RSs[8] are used extensively in NWP models and are an essential anchor measurement for the model temperature. However, RS data exhibit biases which require correction prior to their assimilation into an NWP model. This chapter describes a new method to estimate and correct RS temperature biases on a station-by-station basis using RO BAs as a reference. This part of the thesis was conducted as two visiting scientist projects within the Radio Occultation Meteorology Satellite Application Facility (ROM SAF), which is a decentralised facility under EUMETSAT. The ROM SAF Visiting Scientist Report 26 (Tradowsky 2015) describes the new method to correct RS temperature biases within NWP models and outlines the results of a sensitivity study. ROM SAF Visiting Scientist Report 31 (Tradowsky 2016) describes the preparation of a forecast impact study, develops a method to estimate the structural uncertainty in the applied RO retrieval and compares GRUAN and RO data.

While both ROM SAF reports underwent an internal peer-review process, the main results from the projects where published in the *Journal of Applied Meteorology and Climatology* of the American Meteorological Society to ensure a full anonymous peer-review process. This publication is reprinted in the following section with minor changes to the formatting.

The following section is a slightly adapted reprint of **Tradowsky, J.S., Burrows, C.B., Healy, S.B., and Eyre, J.R. (2017). "A new Method to Correct Radiosonde Temperature Biases Using Radio Occultation Data", *Journal of Applied Meteorology and Climatology*, 65, 6, 1643-1661, doi:10.1175/JAMC-D-16-0136.1**

3.1. Paper I: A new method to correct radiosonde temperature biases using radio occultation data

3.1.1. Abstract

A new method to estimate RS temperature biases using RO measurements as a reference has been developed. The bias is estimated as the difference between mean RO and mean RS departures from co-located profiles extracted from the Met Office global NWP system. Using NWP background profiles reduces the impact of spatial and temporal collocation errors. The use of NWP output also permits determination of the lowest level at which

[8] All abbreviations can be found in the Glossary, Appendix B

the atmosphere is sufficiently dry to analyse RO dry temperature (Tdry) retrievals. We demonstrate the advantages of using a new tangent linear version of the Tdry retrieval algorithm to propagate bending angle (BA) departures to Tdry departures. This reduces the influence of *a priori* assumptions, compared to a non-linear retrieval. RS temperature biases which depend on altitude and the solar elevation angle (SEA) are presented for five carefully chosen upper-air sites and show strong inter-site differences, with biases exceeding 2 K at one of the sites. If implemented in NWP models to correct RS temperature biases prior to assimilation, this method could aid the need for consistent anchor measurements in the assimilation system. The method presented here is therefore relevant to NWP centres and the results will be of interest to the RS community by providing site-specific temperature bias profiles. The new tangent linear version of the linear Abel transform and the hydrostatic integration are described in the interests of the RO community.

3.1.2. Introduction

For decades, RS profiles have been assimilated into NWP systems and, since 2006, RO data have been assimilated, demonstrating a positive impact on weather forecasts (see Healy 2008b; Poli et al. 2010; Rennie 2010b). However, the impact of the high quality RO data may be limited by opposing biases between the observation types. Furthermore, local and regional variations in RS temperature biases have the potential to cause false horizontal temperature gradients in NWP analyses, which would lead to spurious features in the wind field. Satellite radiance measurements require bias corrections when assimilated into NWP models, and these corrections are computed relative to the model background or analysis, either statically (Eyre 1992; Harris and Kelly 2001), or via variational methods (Derber and Wu 1998; Dee 2004; Auligné et al. 2007). This can only be done consistently if sufficient “anchor" measurements are present in the assimilation system. Amongst these, RO and RS are key contributors and therefore ensuring consistency between the bias characteristics of these observation types is important for the stability of the assimilation systems.

This study describes a method to calculate the RS temperature bias correction on a station-by-station basis using high quality RO profiles (see e.g. Anthes 2011) as a reference. A forecast impact study using the method described here to correct the RS temperature biases prior to assimilation into an NWP system is planned and the results will be published separately.

The RO variable used here is the BA as a function of the impact parameter (native RO coordinate), but some assumptions (see Section 3.1.4) allow the retrieval of a Tdry profile, which is a valid approximation of the physical temperature when water vapour effects are negligible. In contrast to other studies (see e.g. He et al. 2009; Sun et al. 2010; Sun et al. 2013; Ladstädter et al. 2015), which analyse the difference between RS and RO profiles based on spatial and temporal collocations, here short-range global forecast (i.e. “background”) fields from the Met Office global NWP system (Davies et al. 2005; Rawlins et al. 2007) are used as a transfer medium; we calculate the background departures (observation minus model background, O-B) for RO and RS respectively and compare the mean departures of the two observation types. The BA background fields are computed from the model fields using the forward model described in Healy and Thépaut

(2006) and Burrows et al. (2014). The method applied here has three advantages: (i) compared to direct observation-to-observation collocations, the influence of differences in time and space is minimised, since every measurement (both RO and RS) has a co-located model background profile, i.e. the model is interpolated to the position and time of each measurement, (ii) the lowest level at which the model humidity is negligible can be determined for each RO profile, which enables the use of the retrieved Tdry departures as low in the atmosphere as is reasonable, and (iii) a tangent linear (TL) retrieval can be used. The knowledge of atmospheric humidity facilitates the investigation of the RS bias using Tdry from 10 hPa to a location-dependent pressure level of about 100 hPa in the tropics and 300-400 hPa in the high latitudes (see also Ladstädter et al. 2015). The bias calculated at the lowest levels might though not be representative for all atmospheric conditions, as it is calculated from a subset of RO profiles which are sampled in especially dry conditions.

To avoid including *a priori* knowledge about the humidity in the RO retrieval chain, we calculate the bias corrections only at altitudes for which water vapour effects are negligible. For operational use in NWP systems a gradual transition of the bias corrections below this altitude must be applied (see Tradowsky 2016) to avoid discontinuities in the assimilated temperature profiles. As the RS bias tends to be largest in the stratosphere, we expect this approach to be reasonable. In contrast, e.g. Sun et al. (2013) use Tdry down to 150 hPa globally and wet temperature retrievals which include *a priori* knowledge from the National Centers for Environmental Prediction (NCEP) 12 hour forecasts further down in the atmosphere.

This paper is based on the results of the ROM SAF Visiting Scientist projects Number 26 (see Tradowsky 2015, for more details on the method), and Number 31 (see Tradowsky 2016, for details on the preparation of the forecast impact study, and on the structural uncertainty).

3.1.3. Observational data and the NWP system

Radiosonde data

More than 800 upper-air sites launch RS on weather balloons to measure vertical profiles of temperature, humidity and, depending on the RS type, pressure. Many manufacturers correct, among other things, the radiation bias in the RS temperature and humidity profiles before the data are released. Even after this correction, the temperature profiles can have substantial biases that vary with the SEA, but also depend on the post-flight processing applied at the ground station. The RS temperature bias is commonly calculated per sonde type (see e.g. He et al. 2009; Sun et al. 2010; Sun et al. 2013), though Milan and Haimberger (2015) found a variation of the temperature bias for stations launching the same sonde type, supporting our approach to calculate the bias separately for each station. A thorough evaluation of sources of biases in RS measurements can be found in Dirksen et al. (2014). Here, we provide a method to correct the remaining temperature bias prior to assimilation into NWP models. A bias correction is calculated for most (for technical reasons not all) land-based RS launch sites that regularly disseminated operational data in 2014. The results for five example sites (see Table 3.1) comprising different climate regimes are analysed here.

Table 3.1.: Country, latitude, longitude, World Meteorological Organization (WMO) station identifier (ID) and RS/radar type for the example upper-air sites analysed in this study. The RS type "unknown, not specified" actually includes three Russian RS types, namely I-2012, MRZ-3MK, AK2m (Bruce Ingleby, personal communication).

Country	Lat	Lon	ID	RS/radar type
Germany	52.22	14.12	10393	Vaisala RS92
Russia (West)	69.35	88.27	23078	AVK-AK2-02, MARL-A/Vektor-M AK2-02, MARL-A/Vector-M-BAR
Russia (East)	59.55	150.78	25913	MARL-A/Vektor-M AK2-02, AVK-BAR, MARL-A/Vector-M-BAR, unknown
Indonesia	-1.18	136.12	97560	Meisei
Antarctica	-69.0	39.58	89532	Meisei

The temperature bias is calculated at those standard pressure levels (SPLs) of RS profiles submitted in the alphanumeric TEMP format (i.e. 1000 hPa, 925 hPa, 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 250 hPa, 200 hPa, 150 hPa, 100 hPa, 70 hPa, 50 hPa, 30 hPa, 20 hPa and 10 hPa) that have negligible humidity. In the Met Office NWP system TEMP profiles have a cold bias of -0.05 K that is caused by the conversion from degree Celsius to Kelvin (see Ingleby and Edwards 2015). Additionally, the encoding/decoding of the RS profiles in TEMP format can cause an offset, e.g. -0.095 K for the RS92 with DigiCORAIII processing (see Ingleby and Edwards 2015), leading to a combined bias of almost -0.15 K in the RS92 data within the Met Office NWP system.
In this study only those RS profiles that passed the operational quality control of the Met Office Observation Processing System (OPS) in 2014 are analysed. Outliers in the RS dataset are rejected based on the Median Absolute Deviation (MAD). To calculate the MAD, first, the RS departures for all profiles at one upper-air site are calculated. Then the absolute deviation is calculated by subtracting the median of all RS O-Bs from each RS departure profile. The median value of the absolute deviation at one SPL is the MAD for the given level. Based on Leys et al. (2013) a moderately conservative threshold of 2.5 is chosen to reject outliers, such that the RS temperature at a certain level is rejected if a RS O-B is more than 2.5 × MAD away from the median of all RS departures at that level. While mean and standard deviation (SD) are especially sensitive to outliers, and are therefore problematic for their detection, the MAD presents a robust measure to reveal outliers.

Radio occultation data

The signal transmitted by the GNSS at about 20,000 km altitude is received by a Low Earth Orbit (LEO) satellite. The measurements are made during the RO event, i.e. when the GNSS satellite rises above or sets behind the horizon. In this configuration the signal passes through the limb of the atmosphere, where it is bent and delayed before it is received by the LEO satellite. The measured phase shift of the received signal allows the retrieval

of atmospheric variables, mainly temperature and pressure, and given additional *a priori* knowledge about the atmospheric state, water vapour. Since the temperature and pressure values are retrieved assuming that water vapour is negligible, these quantities are often referred to as Tdry and dry pressure, respectively, by the RO community. A description of the RO technique can be found in Kursinski et al. (1997) and Kursinski et al. (2000). Since the RO measurement is based on precise timing available from atomic clocks, it offers the possibility to be traceable to the SI standard of time (Leroy et al. 2006). This ensures the long-term stability of RO data and makes them valuable for climate studies. In contrast to other remote sensing techniques, RO measurements are nearly independent of rain and cloud and the profiles are retrieved from the higher atmosphere down into the boundary layer, with a vertical resolution of 0.1-1 km and a horizontal resolution of 100-300 km around the tangent point (see Fig. 3 in Anthes 2011).

As the noise in the RO profile increases at high altitudes, the measured BA profile is usually merged with a climatological BA profile using statistical optimisation in order to retrieve refractivity. This is applied above a certain altitude, e.g. above 40 km in Healy (2001). Thus, the influence of the RO measurement in the retrieval of refractivity and temperature decreases with altitude while the influence of the climatology increases.

This study uses the near-real-time BA data of the US-Taiwanese Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) - Formosa Satellite Mission 3 (FORMOSAT-3) (hereafter COSMIC, see e.g. Anthes et al. 2008). The profiles from both rising and setting occultations are used, as the influence of separating them becomes negligible given an appropriate sample size (see Tradowsky 2015). In a preprocessing step prior to assimilation into the Met Office global data assimilation system, the BA profiles are assessed for their quality. Central to this procedure is a 1D-Var algorithm that uses co-located model background information to obtain an optimal solution for each observation profile. Complete profiles are flagged for rejection if (i) the 1D-Var assimilation fails to converge within 20 iterations, (ii) the initial cost function over all levels is greater than 2.5 or (iii) the final cost function over all levels, i.e. at convergence, is greater than 2.0. The values in the (conservative) rejection criteria and the number of iterations required were chosen based on experience. Approximately 10% of the RO profiles are rejected. Furthermore, BAs are rejected on a level-by-level basis if the absolute value of the observation minus the 1D-Var solution is greater than the assumed observation error multiplied by 5.

The radio occultation temperature null space It should be emphasised that RO is not a direct measurement of temperature, and some *a priori* information is required to make a temperature retrieval well-posed. The need for *a priori* information implies that there is a RO measurement null space, meaning there are some atmospheric profile perturbations that do not affect the measured values. Conversely, information about these perturbations cannot be retrieved directly from the measurement alone (see Rodgers 2000, section 2.2.1).

RO measurements in the dry atmosphere are sensitive to density ρ as a function of altitude h. The density, however, is proportional to the quotient of pressure and temperature ($\rho \propto P/T$). Therefore *a priori* information is needed to split into the influence caused by $T(h)$ and $P(h)$. As a consequence, temperature perturbations ΔT for which the

amplitude grows exponentially with the density scale height $H = \frac{RT}{g}$ as:

$$\Delta T(h) = k \cdot \exp^{(h/H)} \tag{3.1}$$

where k is a constant small perturbation, are difficult to detect with RO. For example, assume two altitude levels h_1 (lower) and h_2 (upper) separated by Δh, with a temperature variation $T(h)$ between the two levels. Perturbing the temperature based on Eq.(3.1), leads to increased temperature at h_2 and therefore the thickness of the layer increases which leads to an increase of the pressure at level h_2. Thus, while $P(h_2)$ and $T(h_2)$ are changing, the ratio $P(h_2)/T(h_2)$ stays approximately constant and thus the density as a function of altitude is not significantly affected.

This temperature null space is a fundamental limitation of the RO technique and is distinct from the exponentially decreasing influence of the top level pressure (see e.g. Kursinski et al. 1997) which is calculated from an *a priori* temperature and the refractivity using Eq.(3.9). Temperature perturbations which grow exponentially based on Eq.(3.1) quickly produce unphysical temperature values on altitude levels. However, more subtle perturbation patterns – partly composed of this exponential growth – remain potentially problematic. For example, the temperature bias highlighted by Steiner et al. (1999, see Figure 8b), produced by perturbing the *a priori* information used in their geophysical retrieval, is in the RO measurement null space.

Due to this RO null space the technique presented here will only be able to estimate the contribution to the RS biases that the BA observations can determine uniquely. However, the assimilation of RO in NWP systems (Healy 2008a) and reanalysis (Poli et al. 2010; Simmons et al. 2014; Kobayashi et al. 2015) is seen to anchor the temperatures at around 100 hPa, which is an indication that RO is able to provide useful bias information at these levels.

NWP system

In this study the Met Office global NWP system provides the model backgrounds for the RO and RS profiles. Using model fields as transfer medium reduces differences that would be caused by imperfect collocations.
During the investigated year, 2014, the model was due for an update of the dynamical core from the version described in Davies et al. (2005) to a version described in Walters et al. (2014). As a result, the resolution changed on the 15 July from N512L70 to N768L70. This corresponds to a decrease in the grid length from about 25 km in mid-latitudes to about 17 km. Also, the time step decreased from 10 minutes to 7.5 minutes.

3.1.4. Method

The RS temperature bias corrections are calculated on a station-by-station basis, giving a vertical bias correction profile for each of the 762 studied sites ("site" and "station" are used synonymously). In this paper a carefully chosen subset of 5 sites comprising different climate regimes and different sonde types is presented. The results for all sites are available as supplementary material to Tradowsky (2015). The bias correction profiles extend from 10 hPa to at least 100 hPa, and for most of the non-tropical stations, considerably

lower into the atmosphere. The lower boundary is determined by the humidity in the background profile as described below. Since the bias corrections at the lowest levels are calculated from a subset of RO profiles sampled in especially dry air masses, these bias corrections might not be representative for all atmospheric conditions.

For each upper-air site, the RO BA profiles within 500 km of the site are selected, and the following are calculated: (i) the mean RO BA profile, (ii) the mean NWP background BA profile and (iii) the mean BA departure (observation minus background) profile. Thus, instead of using only those RS and RO profiles that are closely collocated in time and space, all occultations within 500 km of the site, and all RSs launched at the site, independent of the sonde type, are analysed. Increasing the vicinity radius has only a minor influence on the departure statistics (radii between 300 km and 2000 km have been studied in Tradowsky (2015); see also Sun et al. (2010)). There is only a slight tendency to increase the SD as the radius is increased. A 500 km radius is therefore used to obtain a sufficient sample size while keeping the SD low.

The RS temperature biases depend on the solar elevation, as can be seen in e.g. Philipona et al. (2013), Sun et al. (2013), and Dirksen et al. (2014), and mean RO Tdry departures also show a slight dependence on the SEA as can be seen in Tradowsky (2015). The small influence of solar elevation on RO BAs, which is caused by the ionosphere, propagates through to refractivity and temperature in the retrieval chain, as described by Healy and Culverwell (2015). Therefore both data sets are divided into four SEA ranges, i.e. high (SEA $> 22.5°$), low ($7.5° <$ SEA $< 22.5°$), dusk ($-7.5° <$ SEA $< 7.5°$) and night (SEA $< -7.5°$). RSs are typically launched at 0 UTC and 12 UTC (at some stations additionally 6 UTC and 18 UTC) and the SEA at these times depends on geographic location and time of the year. Thus, for a given site, the separation by SEA may include preferential sampling for particular seasons.
To estimate biases in RS temperature profiles, the mean BA departure profile is propagated to a Tdry departure profile. First, the linear Abel transform is used to obtain inferred refractivity departures, and then the hydrostatic integration of the refractivities gives Tdry departures as described below.

The RS temperature bias correction is calculated as the difference between the mean RO Tdry departures and the mean RS temperature departures, i.e.

$$\overline{O_{RO} - O_{RS}} \simeq \overline{O_{RO} - B_{RO}} - \overline{O_{RS} - B_{RS}} \tag{3.2}$$

where O is the observation and B is the background, forward-modelled into observation space. The $\simeq$ denotes that the assumption has been made that B_{RO} and B_{RS} are equally representative of the true values at the RO and RS locations respectively, i.e the central assumption is that the NWP forecast bias does not vary between the RO and RS locations. This is a more robust assumption compared to direct collocations between measurements, which are made assuming that the atmosphere does not vary over the separation distance. A similar double differencing approach is used by Haimberger et al. (2012) to homogenise RS temperature records.

In this investigation a new tangent linear Tdry retrieval, which calculates the Tdry departures from BA departures, is used. This retrieval is described here for the first time in the peer-reviewed literature, though it is based on the linear calculations developed in Syndergaard (1999), who originally proposed them to assess error propagation in the retrieval

chain. The tangent linear retrieval allows to estimate the Tdry departures implied by any subset of BA departures. Here, the BA departures up to 35 km impact height are used as this is the altitude range of RSs. The BA departures above 35 km impact height are set to zero. In practice, this enables us to reduce the influence of *a priori* information on our results. More detail about the retrieval can be found in Tradowsky (2015). Up to now all Tdry retrievals use a combination of RO BA and a climatological/model BA at altitudes above 30-40 km (Ho et al. 2012, Table 1) to minimise the effect of increasing noise in high altitude RO BAs. Usually this is done with statistical optimisation, which blends RO and climatological BAs, decreasing the weight of the RO BAs with altitude. Depending on the processing centre, different high level initialisations are applied, as it is described in Ho et al. (2012, Appendix A). Thus, no consensus about how best to initialise upper level BAs exists in the community, but the spread between the different retrievals can be used to estimate the structural uncertainty (e.g. Ho et al. 2012; Steiner et al. 2013). The influence of the statistically optimised upper level BAs propagates down into lower levels when refractivity and Tdry are retrieved. Thus, the whole retrieval depends on the climatological BA profile chosen by the processing centre (see estimate of differences in e.g. Ho et al. 2012; Steiner et al. 2013), although the influence of climatology decreases with decreasing altitude. The method proposed here eliminates the use of a climatological BA profile, by using a cut-off of the BA departures above 35 km impact height, which can be seen as an extreme case of the statistical optimisation. As noted above, this altitude range is also a better match to the altitudes sampled by RSs.

(Tangent) linear Abel transform

BA profiles are retrieved from the measured phase shift of the GNSS signal, assuming local spherical symmetry of the atmospheric refractive index (see Kursinski et al. 2000). Under this assumption the relation between the BA, α, and the refractive index, n as a function of tangent radius r is described by the Abel transform (Fjeldbo et al. 1971).

$$n(r) = \exp\left(\frac{1}{\pi}\int_{x}^{\infty}\frac{\alpha(a)}{\sqrt{a^2-x^2}}da\right) \tag{3.3}$$

where $x = nr$ equals the impact parameter a for the ray with tangent radius r. The impact parameter for a given ray represents the distance of closest approach of the undeflected part of the ray from the local centre of curvature of the Earth at the position of the ray tangent point (see e.g. Fig.1 in Kursinski et al. 2000, for occultation geometry). For convenience, the refractivity N is defined as $N = 10^6(n-1)$. Assuming linear variation of the BA between successive observation levels, Syndergaard (1999) linearised Eq.(3.3) giving the refractivity as a function of the tangent radius, which is implicitly related to the refractive index as $r = a/n(r)$.

$$N(r) \approx \frac{10^6}{\pi}\int_{x}^{\infty}\frac{\alpha(a)}{\sqrt{a^2-x^2}}da \tag{3.4}$$

Discretising Eq.(3.4) (see Appendix A in Syndergaard 1999) gives:

$$\mathbf{N} = \mathbf{A}\boldsymbol{\alpha} \tag{3.5}$$

with $\mathbf{A}$ (see Eq.(3.33) in Syndergaard 1999) being an $m \times m$ triangular matrix, where m is the dimension of the profile.
A priori information is required to solve the Abel transform, as the upper limit of the integral is infinity, but the BA are only available up to $\approx 60\,\mathrm{km}$. Within this project, for each profile the BA value at the highest level and the mean NWP background temperature at this level are used to initialise a (fairly arbitrary) extrapolation allowing the computation of the Abel integral up to infinity (for details see Burrows and Healy 2016). The refractivity at the highest level is then used (again, with the background temperature at this level) to initialise the hydrostatic integral to obtain Tdry at all levels below. As a consequence of these two integrals, each retrieved value of Tdry contains information from every value of refractivity above, and each of these refractivity values contains information from the BA at the respective level and above, including the uppermost BA value and those extrapolated values defined by the uppermost BA. There is, therefore, considerable sensitivity of Tdry to the highest level of the BA profile. To minimise the influence of *a priori*, TL calculations as described below are used here.

As long as the departure from the linearisation state is small, the linear approximation of a non-linear function provides a good approximation of the function for small perturbations. TL models are used extensively in data assimilation systems (Hoffman et al. 1992), e.g. to estimate the development of the atmospheric state at short time scales, as well as in observation operators. To build the TL retrieval code, the partial derivative of each equation is calculated one after another.
In this study observed and NWP background BAs are closely collocated as the model is interpolated to the measurement position and time and therefore are sufficiently similar to consider the matrix $\mathbf{A}$ to be equal for RO and model BA. Hence the matrix $\mathbf{A}$ can also be used to propagate BA departures, since
$\mathbf{N_2} - \mathbf{N_1} \simeq \mathbf{A}\boldsymbol{\alpha_2} - \mathbf{A}\boldsymbol{\alpha_1} = \mathbf{A}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)$. This TL version of the Abel transform, where the BA departures $\boldsymbol{\delta\alpha}$ are used to calculate the refractivity departures $\boldsymbol{\delta}\mathbf{N}$, can be expressed in the same notation as Eq.(3.5):

$$\delta\mathbf{N} = \mathbf{A}\delta\boldsymbol{\alpha} \tag{3.6}$$

The TL Abel transform, together with the TL version of the Tdry calculation, is used throughout this paper unless stated otherwise.

Dry temperature calculation

The relation between refractivity, N, temperature T, pressure P and water vapour pressure P_w is described in the Smith-Weintraub equation (Smith and Weintraub 1953)

$$N = c_1 \frac{P}{T} + c_2 \frac{P_w}{T^2} \tag{3.7}$$

where $c_1 = 77.6\,K\ hPa^{-1}$ and $c_2 = 3.73 \times 10^5\,K^2\ hPa^{-1}$ are empirical constants.
To compute the pressure from the refractivity assuming a dry atmosphere, the hydrostatic equation is integrated using a gravitational acceleration, g, that varies with latitude and

altitude. With the ideal gas law and assuming the refractivity varies exponentially with the geopotential height z, the contribution to the pressure from each discrete layer can be calculated as (for details see Tradowsky 2015):

$$\Delta P = \frac{g}{Rc_1} \frac{N_i - N_{i+1}}{ln\left(\frac{N_i}{N_{i+1}}\right)} (z_{i+1} - z_i) \tag{3.8}$$

where R is the specific gas constant. The calculation of the pressure at each impact height level (impact height IH is the native coordinate for BAs and is related to the geometric height (altitude) h via $IH \approx h + 6.3 \cdot N(h)$, see Healy et al. 2007) is initiated at the highest level, which requires *a priori* knowledge about the temperature at this level. The *a priori* temperature used here is a mean value of the NWP background temperature at 59.06 km geopotential height which is closest to the highest impact height level. The average includes background temperatures at 59.06 km for all those profiles that are used to calculate the RO O-B at a given site. With this temperature, the top level pressure is calculated as:

$$P(z_{top}) = \frac{N(z_{top})T(z_{top})}{c_1} \tag{3.9}$$

At altitudes where the water vapor contribution is negligible, Tdry is calculated from the pressure and Abel-derived refractivity with the dry term of Eq.(3.7) as:

$$T_{dry}(z) = c_1 \frac{P(z)}{N(z)} \tag{3.10}$$

A TL Tdry calculation, comprising the Abel transform, the hydrostatic integration, and the Smith-Weintraub equation is used here to calculate the Tdry departures (δT_{dry}) and can be expressed with the linear Tdry operator $\boldsymbol{K}$ and the BA departure as:

$$\delta \boldsymbol{T} = \frac{\partial \boldsymbol{T}}{\partial \boldsymbol{\alpha}} \delta \boldsymbol{\alpha} = \boldsymbol{K} \delta \boldsymbol{\alpha} \tag{3.11}$$

where $\boldsymbol{K}$ is the ma6trix of partial derivatives of Tdry with respect to the BA, assuming all *a priori* information required in the retrieval is fixed.
A calculation with the original non-linear equations is done to provide linearisation states, for the purpose of comparison, and to study the performance of the Tdry calculation (see Section 3.1.8).

Determining the lowest dry level

Since output from an NWP model is used as a transfer medium, it is possible to determine the lowest level of negligible humidity in the background fields separately for each RO profile, enabling us to use the Tdry profile as low in the atmosphere as is reasonable. The metric used to determine the deviation of the Tdry from the actual temperature is described in Scherllin-Pirscher et al. (2011) as: $T_{dry} - T \approx -4/5 \cdot c_{q2T} \cdot q$, where q is the model specific humidity and $c_{q2T} = (c_2/c_1)/a_w = 7728$. a_w is the ratio of the dry air to water vapour gas constants, and c_1 and c_2 are the constants from Eq.(3.7). The level with lowest altitude above which the difference in physical temperature and Tdry

remains below 0.09 K is determined for each profile. This value is chosen based on the suggestions in (Scherllin-Pirscher et al. 2011) and also Fig. 3.8 (and analogous figures for different profiles, not shown here) supports this decision, since the model temperature and the model Tdry (computed from forward-modelled BAs) agree well in the lower levels that satisfy the criterion.

The BA departure statistics are calculated using BA values at all dry levels. If fewer than 10 dry profiles contributed to these statistics, then no subsequent analysis is performed as the averages may be unreliable.

Upper cut-off criterion for bending angle departures

The GNSS-RO Tdry retrieval is an example of an "exact retrieval" (Rodgers 2000), because it does not consider the impact of measurement noise and tries to fit the measured values exactly, including the noise they contain. As the noise in the RO BAs increases with altitude in the upper stratosphere and above, conventional retrievals of refractivity, and hence Tdry, apply statistical optimisation which blends the RO BA profile with a climatological BA profile, and the resulting profile is smoother than the original observed profile (see Appendix B for a brief description of statistical optimisation). Therefore, when statistical optimisation is used in the retrieval algorithm, the measured BAs contribute less information to the retrieval at higher altitudes, where the climatology dominates. As stated above, the Abel transform and the hydrostatic integral both serve to propagate BA information down through the whole profile. Therefore, if statistical optimisation is used, the influence of the climatology will have some impact on the resulting Tdry retrieval at all levels below which the climatology has a finite contribution. This is not desirable when estimating biases of RS, so here we intend to use the "exact" Tdry retrieval described above in a way which minimises the impact of noisy or biased data. Therefore a linear approach is used, that allows to restrict the vertical range of BAs that contribute to the Tdry departures and furthermore avoids the use of prior information. By setting the BAs to zero above 35 km, the method bears similarity to an extreme case of statistical optimisation where, in the BA departures, the observed BA is replaced with the background values above 35 km (or, equivalently, both observation and background are replaced by an independent climatology above this impact height). Here Tdry departures that have been calculated from BA departures below 35 km are analysed, which eliminates the influence of *a priori* assumptions and model bias above this impact height. To calculate RS bias corrections, we use a double differencing approach, which is valid under the assumption that the biases in the model background profiles that are collocated with the RS observations are the same as those collocated with the RO observations. Therefore it is important to use model information in the same altitude range for both observing systems. To reiterate, using the BA departures up to an altitude far above the RS balloon burst altitude would invalidate this central assumption, because the impact of any model bias at these uppermost levels would contribute to the Tdry departures below (see Fig. 3.1 and section 3.1.9), while these high-level biases will *not* contribute to RS temperature departures.

Burrows and Healy (2016)(Fig. 8) show how the global Tdry departures for different cut-off impact heights differ when using either the Met Office or the ECMWF BA background fields, highlighting the importance of minimizing any influence from the model which will

not be removed by the double differencing. When no cut-off is applied the difference between the Tdry departures is >1 K at 10 hPa. The BA departure statistics of Met Office and ECMWF are very similar up to an impact height of about 35 km. This implies that the differences in the Tdry departures are caused by model BA biases of opposite sign at higher altitudes (see Fig. 7 in Burrows and Healy 2016) that are propagated down in the retrieval chain.

In Fig. 3.1 the TL Tdry departures for different cut-off impact heights (including no cut-off) are compared with Tdry departures calculated using BA profiles that are blended with climatology (similar to statistical optimisation). In addition, Fig. 3.1 shows the non-linear Tdry departures for the example site in western Russia. Tradowsky (2015) shows similar figures (excluding the statistical optimisation) for all five example sites, showing that the departures for most stations tend to become more positive with increasing cut-off impact height (Burrows and Healy 2016, show that this is caused by the model background fields). Note that the similarity of the green line (no cut-off) and the dashed black line (non-linear) indicates the accuracy of the linearisation. A major impact on the Tdry departures results from setting solely the BA departure value at the highest level to zero prior to the TL Tdry calculation (compare green and grey line). This large sensitivity is an artefact of using the RO BA value at the highest level to extrapolate the observation vector to infinity. In Section 3.1.9 this sensitivity is evaluated in detail by analysing covariance and Jacobian matrices.

The dot-dashed curves show Tdry departures calculated using model/RO BAs that are blended with a smooth climatological BA (in this case the background profiles are used), with 50% weight on *a priori* information at 47 km (red) and 32 km (violet), respectively. Conventional RO retrievals typically use statistical optimisation to smooth the profiles and we aim to illustrate what effect different implementations of statistical optimisation can have on the retrieval (see Section 3.1.9 for a description of the applied smoothing). When using 50% *a priori* at 32 km, the Tdry departures closely follow the line for 35 km cut-off (blue), with the biggest difference of about 0.1 K at the top level. Similarly, using 50% *a priori* at 47 km closely resembles a cut-off at 50 km. Thus, setting the departures above a certain altitude to zero gives results similar to blending the RO and model BA profile with climatology, but does not require *a priori* information.
Like every assumption that is made in a retrieval, reasonable variations in the choice of the upper cut-off impact height contribute to the structural uncertainty. This is analogous to the structural uncertainty related to choosing a climatology/weighting to be used in the calculation of statistically-optimised BAs (see difference in dash-dotted lines in Fig. 3.1). Ho et al. (2012) estimates the structural uncertainty resulting from different processing schemes as the differences and SDs between individual centres and the inter-centre mean. For global RO Tdry calculated from different processing centres, Ho et al. (2012) finds the mean differences to be between -0.27 K and 0.15 K in an altitude range of 8-30 km.

To give an estimate of the structural uncertainty caused by the choice of the cut-off impact height, the range (largest minus smallest value) of the departures calculated for cut-off impact heights between 35 km and 55 km is shown in Fig. 3.2. The black line is calculated from ≈ 77,000 globally distributed COSMIC-6 BA profiles measured during 2014, while the other lines are calculated from all COSMIC BA profiles within 500 km around the respective upper-air site (the station ID is given in the legend). In general, the structural uncertainty increases with decreasing pressure. The structural uncertainty of

the global statistics is calculated to estimate the typical behaviour. It stays below 0.2 K at pressures above 50 hPa and increases to about 0.47 K at 10hPa. Our estimation of the global structural uncertainty resulting from choosing an upper cut-off is thus similar to the structural uncertainty resulting from different processing as presented in Ho et al. (2012). The structural uncertainty is appreciably lower at the Russian upper-air site with the station ID 25913 and appreciably larger at the Indonesian site (ID 97560). Strong vertical gradients, caused by gravity waves above this tropical site, could explain the strong dependence on the cut-off impact height, but also the comparably small sample size ($\leq$ 242 ROs, compared to $\leq$ 1093 for site 25913) could affect the results. Estimating the structural uncertainty is of major importance for climate studies. For the purpose of estimating RS biases, the spread of the departures between 35 km and 55 km might though be an overestimation of the uncertainty, as a cut-off impact height of 55 km would invalidate the central assumption for the double-differencing technique, i.e. the model bias is constant over the separation distance (see also Burrows and Healy 2016).

In summary, the application of the 35 km cut-off is justified for two reasons. First, this reduces the effect of noisy, and arbitrarily extrapolated RO data from the upper stratosphere and mesosphere that would otherwise affect all calculated bias corrections. Second, it avoids the influence of model biases above the RS altitude range which, if the cut-off was not applied, would result in biased RO departures whose influence would persist even after the double differencing.

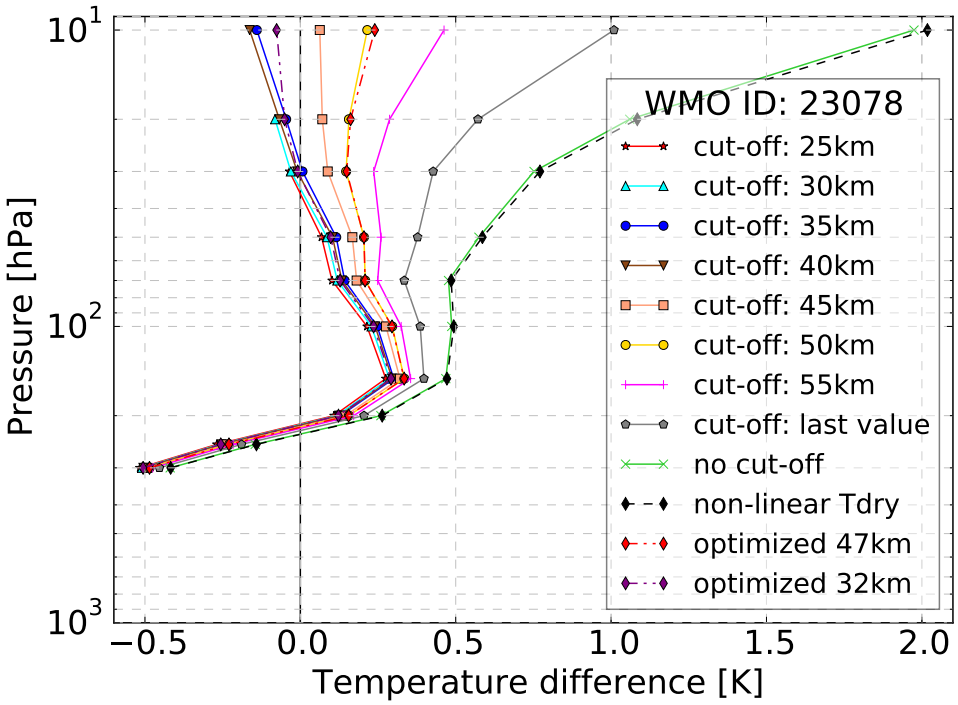


Figure 3.1.: Sensitivity of the mean Tdry departure to different upper cut-off impact heights. The mean Tdry departure is calculated from up to 842 RO profiles within a 500 km radius around the example site 23078 in western Russia. Also shown is the Tdry departure calculated with the non-linear retrieval (black dashed) and the Tdry departures calculated using an approach similar to statistical optimisation with 50% of the background BA used at 32 km (purple dash-dotted) and at 47 km (red dash-dotted), respectively.

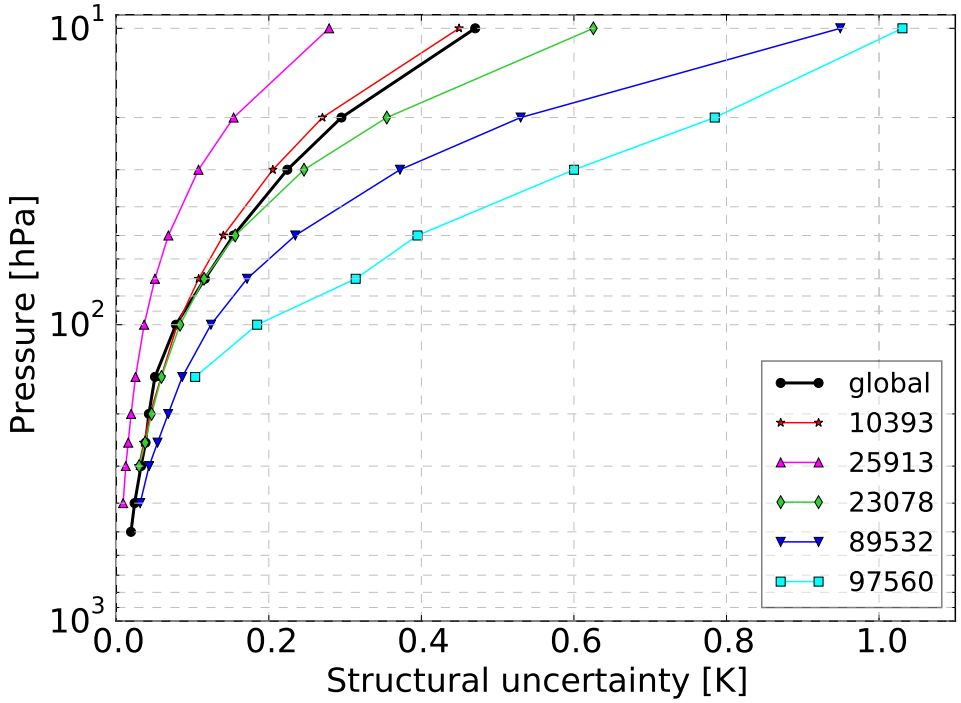


Figure 3.2.: Estimate of the structural uncertainty in Tdry departures for five example stations and for the mean global Tdry departures calculated from ≈77,000 COSMIC-6 profiles. The structural uncertainty is calculated as the range (largest departure minus smallest departure) of Tdry departures inferred from BA departures with different upper impact height cut-offs between 35 km and 55 km.

3.1.5. Comparison of departure statistics for radiosonde and radio occultation

The RO Tdry departures interpolated to SPLs (blue), the RS temperature departures (pink) and the bias corrections (green) are presented in Fig. 3.3-Fig. 3.7 for five example upper-air sites (see Table 3.1). Since the RS and RO departure statistics depend on the SEA, the bias and hence the bias correction profile (the bias correction is shown in the plots, identical to bias but opposite sign) is calculated separately for high, low, dusk and night launches as defined in section 3.1.4. The departure statistics and the bias correction are displayed with error bars representing the associated sampling uncertainty (standard error, SE) of the mean, see Eq.(3.13) and Eq.(3.14)), taking into account the sample size (dashed blue and pink lines) and the SD of the means. An estimate of the structural uncertainty, which is not included in the error bars, is given in Fig. 3.2. The horizontal cyan coloured line in Fig. 3.3-Fig. 3.7 marks the highest SPL (lowest altitude) where at least 95% of the RO profiles are used, i.e. not more than 5% of the profiles are excluded due to atmospheric humidity exceeding the threshold defined in Section 3.1.4 or because they failed the quality check. Below this level the bias corrections are not representative for all atmospheric conditions at the respective site, as the selected RO profiles are sampled in especially dry conditions. This could lead to particular bias characteristics in the RO departures, probably originating from the model, which would invalidate our central assumption as no masking is applied for RS. Thus, while the RO departures below the horizontal cyan line in Fig. 3.3-Fig. 3.7 are valid for certain atmospheric conditions, the resulting bias correction cannot be used to correct all RS profiles for the site.

From the five example sites shown here, the RS temperature bias correction is smallest at the German site (Fig. 3.3, bias < 0.5 K), which uses a Vaisala RS92 sonde, and at the Antarctic station (Fig. 3.7, bias < 1 K) using Meisei sondes. At the German example site we find a small warm bias at the 10 hPa level for "high" SEAs. Also Ladstädter et al. (2015) found this warm bias however with a stronger extend. Between $\approx$ 200-30 hPa, our analysis shows a small negative RS bias, which is not obvious in Ladstädter et al. (2015), who analysed the differences in RS and RO for the years 2002-2013. The differences in the results between this study and Ladstädter et al. (2015) might be caused by the choice of analysed years, as the vendor correction in the RS92 profiles might have changed. Furthermore, as described in Section 3.1.3, RS92 TEMP profiles are almost 0.15 K too cold in the Met Office global NWP system which partly explains the negative bias found in our analysis. Although the Indonesian site (Fig. 3.6) launches Meisei sondes like the Antarctic example site, the bias corrections are larger, reaching about 1.3 K. This indicates that the RS bias does not depend solely on the RS type (see also Milan and Haimberger 2015), but also on the station, possibly related to varying ground station software, different climate regimes (Sun et al. 2013, analyses the RS bias for different latitude bands), and the season in which the profiles are sampled. This finding supports our approach to calculate the bias correction on a station-by-station basis rather than based on the RS type.

While the RS bias correction at the site in western Russia (Fig. 3.4) stays below 0.7 K for dusk and night and is only larger at the lowest pressure level for high and low SEAs, a larger bias correction is needed at the site in eastern Russia (Fig. 3.5). Here, the bias correction is largest for dusk and night, reaching values up to about 2.5 K. Interestingly the RS temperatures at many Russian sites show cold biases, leading to a positive bias

correction being required, for all SEAs and at most levels. This is in contrast with the theoretically expected radiation bias, i.e. a warm bias during daytime and a cold bias during nighttime. Our results, however, agree with the findings by Rennie (2010a) for RSs tracked with the Russian AVK radar. The cold bias that prevails for all SEAs could be caused by the correction of biases in the ground system software. In general, the raw RS profiles are not disseminated, but rather vendor-corrected profiles, for which the applied corrections are not always traceable.

The temperature bias correction is calculated here on a station-by-station basis, disregarding which type of sonde/radar is used. This is a compromise made to achieve a sufficient sample size for statistical significance. Though it may lead to problems in the operational use for individual cases (e.g. if a Vaisala sonde would be launched at a site with a bias correction of +2.5 K), we expect that applying the bias correction operationally will improve the RS temperature profiles on average. The RS bias corrections are calculated from the highest dry SPL (lowest altitude) up to the lowest SPL (highest altitude) where enough profiles are available. Thus, bias correction profiles are calculated to a minimum pressure level of 10 hPa. For RSs that reach lower pressures, Tradowsky (2016) describes how the bias correction is extended upwards for the use in a forecast impact study.

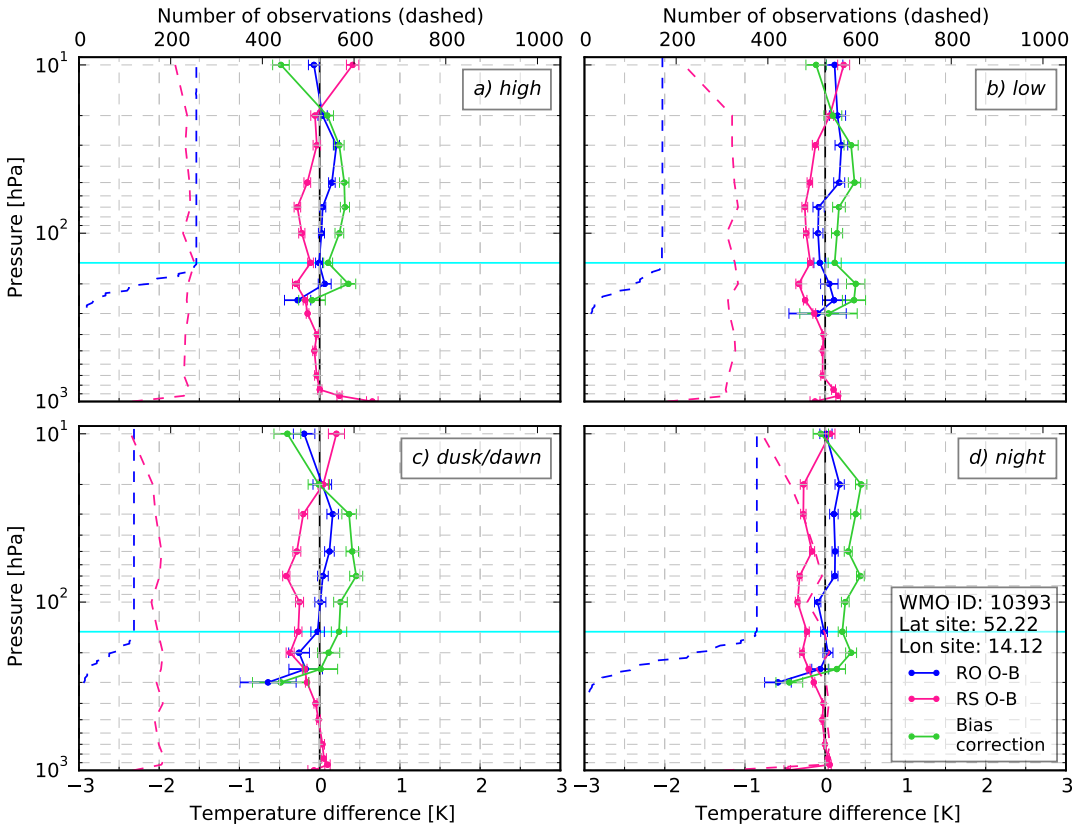


Figure 3.3.: Mean RO Tdry departures (blue), mean RS temperature departures (pink) and bias correction (RO O-B minus RS O-B, green) at the example site in Germany for different SEA ranges (a)-(d). The horizontal cyan line indicates the highest SPL (lowest altitude) where at least 95% of the RO profiles are included. The error bars represent the standard errors (sampling uncertainty) as calculated in Eq.(3.13) and Eq.(3.14). An estimate of the structural uncertainty is provided in Fig. 3.2.

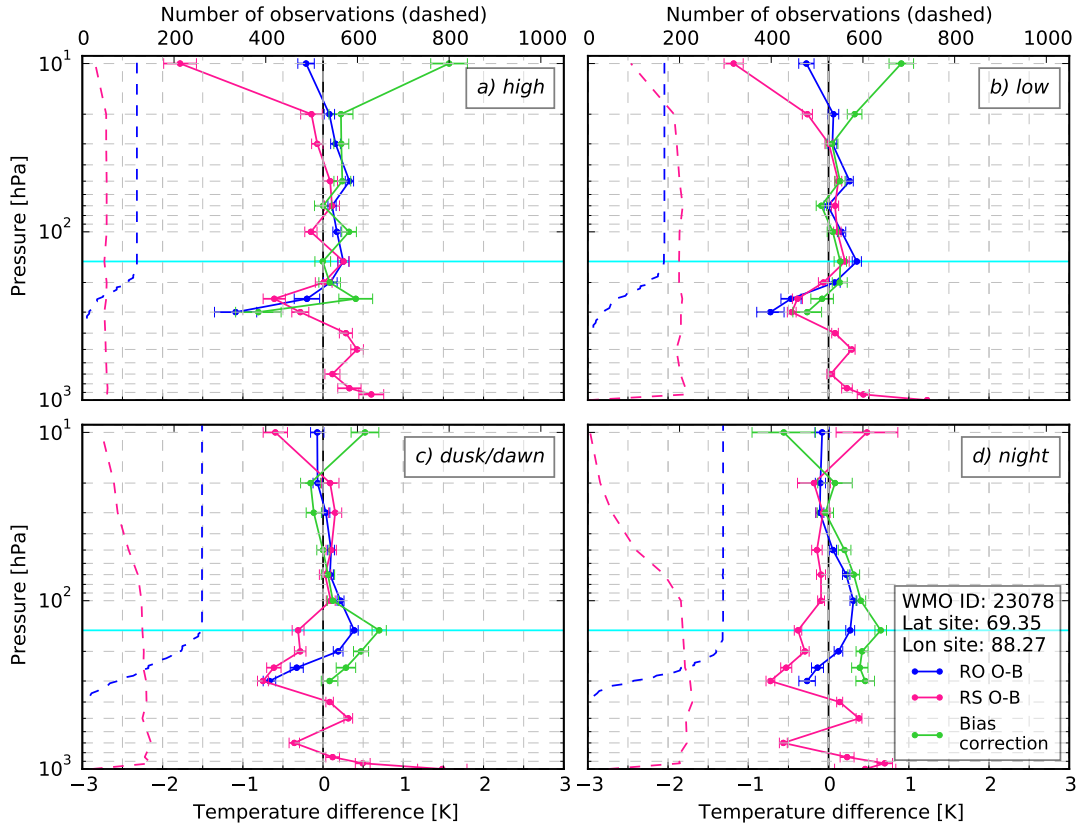


Figure 3.4.: Like Fig. 3.3, but for the example site in western Russia.

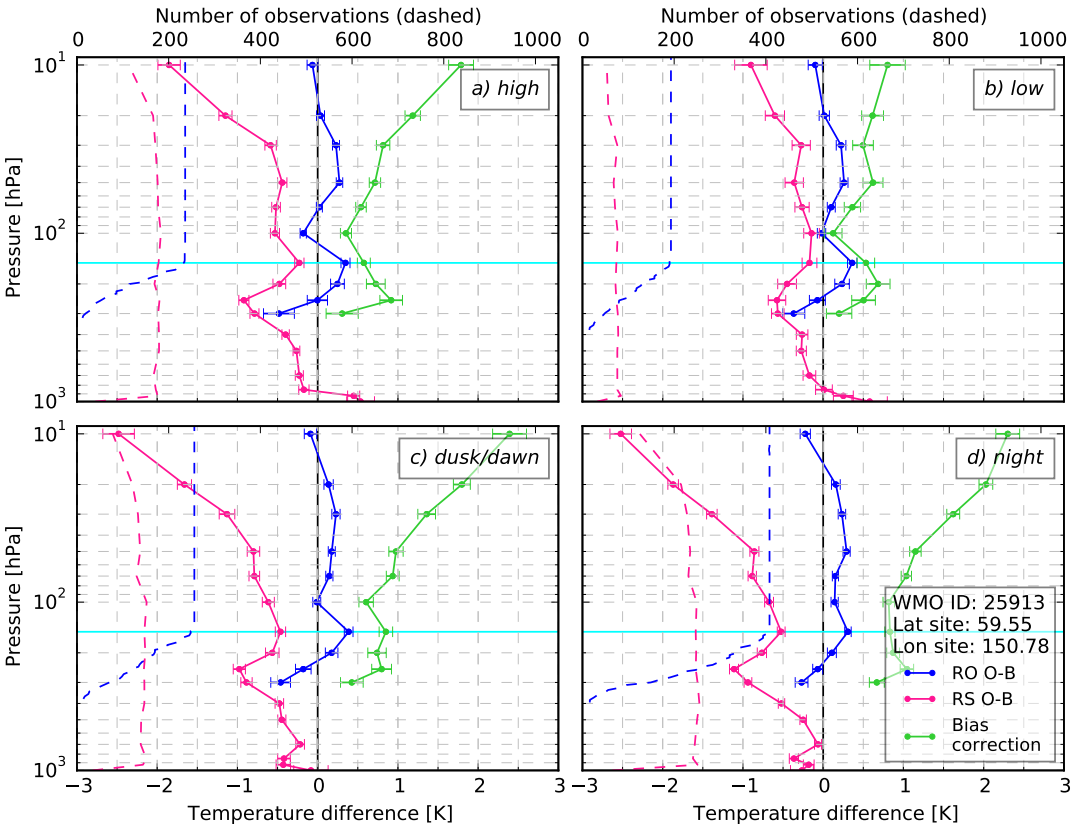


Figure 3.5.: Like Fig. 3.3, but for the example site in eastern Russia.

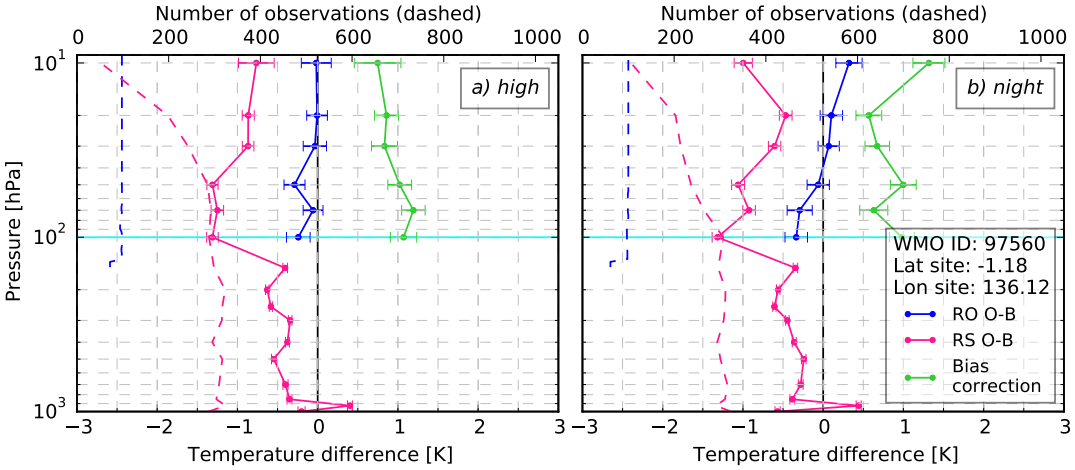


Figure 3.6.: Like Fig. 3.3, but for the example site in Indonesia.

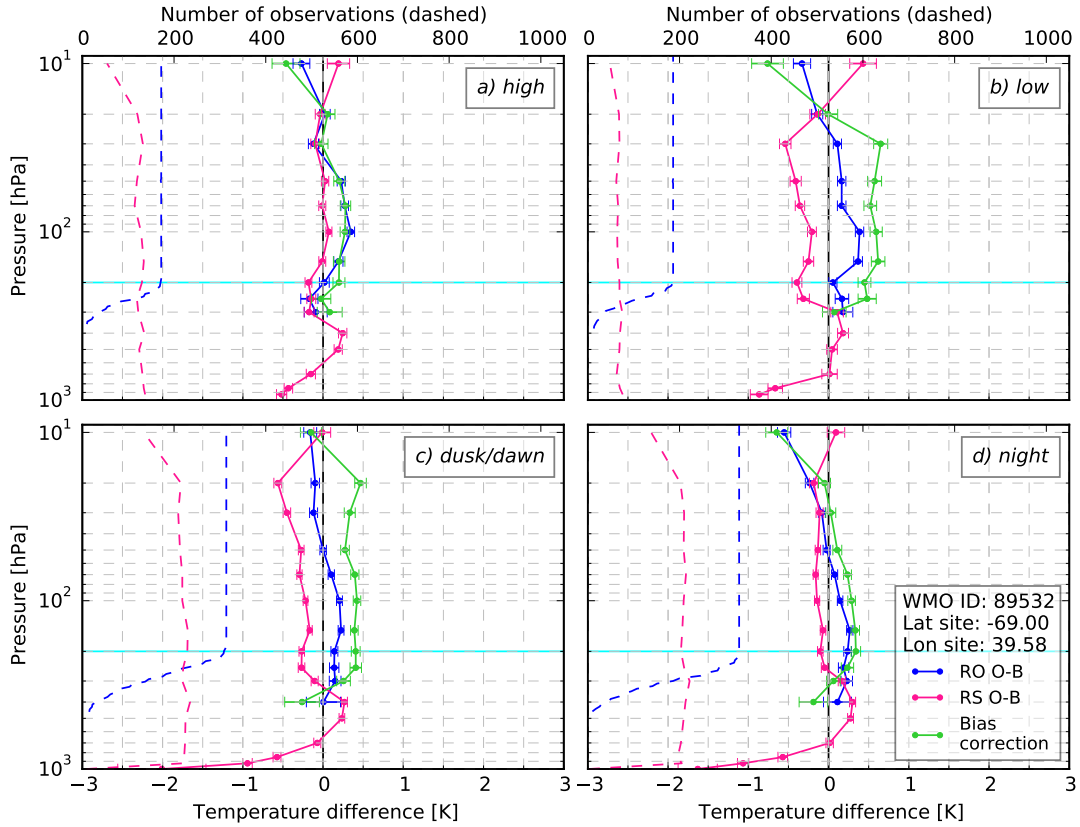


Figure 3.7.: Like Fig. 3.3, but for the example site in Antarctica.

3.1.6. Summary and Conclusions

A method to estimate the RS temperature bias based on a double differencing approach of RS and RO background departure statistics is presented and the bias corrections are shown for five carefully chosen example sites.

First, RS departure statistics from a given site are calculated as the mean difference between the RS temperature and the NWP system background temperature. Similarly, the RO Tdry departure statistics, including all profiles within 500 km of the launch site are calculated from the BA departures with a linear retrieval algorithm. Then, the difference between the RO and RS departure statistics (RO O-B minus RS O-B) estimates the temperature bias correction to be applied to RS temperature profiles, thereby using the RO measurement as an unbiased reference. In this method, the NWP fields serve as a transfer medium and this reduces the errors caused by imperfect collocation. The approach is based on the assumption that the bias in the NWP system does not vary within the vicinity radius of 500 km. Compared to the assumption of a non-varying atmosphere, which is implicitly used for direct observation-to-observation collocations, this assumption leads to relatively small SDs. The double differencing technique has two further advantages: (i) model humidity information allows the determination of the lowest level where RO Tdry can be used, and (ii) it is possible to use a TL retrieval of Tdry departures from BA departures which reduces the sensitivity to *a priori* information.

In contrast to the non-linear calculation of Tdry profiles from BA profiles, the TL version calculates Tdry departures from BA departures. As the aim is to use model background information from the same vertical range for both observation types, BAs are used from the lowest sufficiently dry level in the atmosphere to an impact height of 35 km. The BA departures above 35 km are set to zero, which limits the influence of *a priori* information and model biases at higher levels. The applied cut-off is comparable to an extreme case of statistical optimisation, which uses 100% measured/forward-modelled BA values below 35 km and 100% climatological BA above 35 km in both the measured and the forward-modelled profiles. As the mean Tdry departures and their covariances depend on the choice of upper impact height cut-off (which is similar to how the conventional RO retrieval depends on the weighting in the statistical optimisation), a detailed investigation of the cut-off is shown in Section 3.1.9.

In theory a positive RS temperature bias is expected during daytime from solar radiation, while the emission of radiation by the RS can cause a negative bias during nighttime. Most RS manufacturers account for these biases in the ground station software and disseminate a corrected RS profile. Such corrected profiles are analysed here, which might explain why the sign of the bias does not always agree with the expectation based on theory. This is especially clear for the analysed Russian upper-air sites where the temperature bias tends to be negative for all SEA ranges and can exceed -2 K at the 10 hPa level.
Depending on the station, the RS type and the SEA, the magnitude of the bias varies. For some stations the estimated temperature bias stays below ± 0.5 K throughout the whole profile, while it exceeds ± 2 K at others stations. An increase of the bias with increasing altitude is found for many stations, especially those that show large biases. Sites launching the Vaisala RS92 sonde, which is often used as reference (see e.g. Agustí-Panareda et al. 2009), tend to show a small negative bias in the lower levels and a positive bias in the highest level during daytime and a slightly negative bias at most levels during

nighttime. The bias is calculated for each upper-air site separately and differences in the bias characteristic occur even if the same sonde type is launched, as was also found by Milan and Haimberger (2015). This can be caused, for example, by different versions of the vendor software. As described in Ingleby and Edwards (2015) and in Section 3.1.3, the bias correction is not fully independent of the NWP system, as the TEMP encoding/decoding and the temperature conversion introduces a cold bias for RS in the Met Office NWP system.

The technique presented here provides the basis for developing bias corrections for all operational RS stations (see Tradowsky 2016), which will then be applied in a forecast impact study using the Met Office global NWP system. Within the forecast impact study, for each observed RS profile, the bias correction for the corresponding site and SEA would be interpolated onto the observation levels. The interpolated bias correction can then be added to the observed temperatures prior to assimilation. The bias corrections described here are only available in the stratosphere down to an altitude where humidity starts to become significant. Tradowsky (2016) describes how the bias correction below this altitude is reduced smoothly to a Met Office specific default value in the preparation of a forecast impact study.

The method presented here can also be used to investigate how the reference-quality RS profiles from GRUAN (GCOS-112 2007) compare to RO measurements (see Tradowsky 2016). In operational weather forecasts, the method presented here could be used to calculate the bias corrections on a regular basis. It could replace the Hawson correction (Hawson and Caton 1961) that is currently applied for some RS stations in the Met Office NWP system. The correction of RS temperatures prior to assimilation should serve the need for consistent measurements to anchor NWP models. Anchor measurements are not bias corrected as part of the data assimilation cycle and Eyre (2016) shows the importance of having a sufficient number of observations to anchor the system. There are two aspects of anchoring measurements within the scope of NWP, i.e. they provide the mean anchor to which the mean analysis is pulled, and they provide local anchors. Using the proposed method, the mean anchor is provided by the mean RO observations to which the mean RS temperatures, averaged over all stations, are corrected. The local anchor, however, is provided by both the RO and RS observations. Enhancing the consistency between the observation types that anchor the temperature in a data assimilation system has the potential to improve the NWP skill and further investigation will evaluate the performance of the proposed method.

3.1.7. Acknowledgements

This study was performed as ROM SAF Visiting Scientist project, with J.S. Tradowsky as Visiting Scientist. C.B. Burrows and S.B. Healy are members of the ROM SAF which is a decentralised facility of EUMETSAT. The authors are grateful that the ROM SAF enabled this study. We are thankful to Bruce Ingleby who provided knowledge about RS, Axel von Engeln who reviewed the ROM SAF Visiting Scientist report, Greg Bodeker who provided helpful comments about the draft paper and Ian Culverwell who supported the study with his knowledge about statistics. The first author expresses her gratitude to Peter Builtjes, Jürgen Fischer, Greg Bodeker, and Richard Querel, who supervise her work, and to the German Academic Exchange Service for supporting her with a doctoral

research grant. Furthermore the authors are thankful to three anonymous reviewers who's comments improved this paper.

3.1.8. Appendix A: Performance of the dry temperature calculation

Since the model BA is calculated from the model variables as described in Burrows et al. (2014), it is possible to test the performance of the non-linear Tdry calculation by comparing one model background temperature profile with the Tdry profile retrieved from the associated forward-modelled BA profile. If the assumptions in the forward (temperature to BA) and inverse (BA to temperature) calculations were identical, the differences of model temperature and model Tdry would be negligible in the dry atmosphere. But, due to differing assumptions, differences that vary with altitude are expected.

Figure 3.8 shows one typical profile of the background temperature (blue) and the Tdry (pink) retrieved from the associated background BA profile in the vicinity of the German example site (see Table 3.1). The profiles comprise pressure levels with negligible humidity, as determined following Section 3.1.4. Temperature differences of about 5 K around 1 hPa, occur with decreasing magnitude towards lower altitudes (<2 K at 10 hPa). The differences are caused by different assumptions in the forward and inverse calculations, especially about the variation of the quantities between levels (compare Burrows et al. (2014) with Section 3.1.4) and above the highest level. In this project the BA is assumed to fall exponentially with altitude above the uppermost observation in the Tdry retrieval (which is equivalent with the assumption of an isothermal atmosphere above 60 km impact height).

The highest level Tdry is plotted at a slightly too high dry pressure compared to the model pressure, which is caused by different assumptions in the calculation of pressure in the model and Tdry calculations. Otherwise, the top level Tdry agrees well with the background temperature, emphasizing the value of initializing the hydrostatic integration at the highest level with the model temperature as described in Section 3.1.4. Although some differences in the model temperature and Tdry are present, resulting from differences in the assumptions, it can be concluded that the non-linear Tdry calculation as described in Section 3.1.4 and further discussed in Burrows and Healy (2016) performs well.

3.1.9. Appendix B: Analysis of sensitivity to the upper cut-off height

Dry temperature departure covariance matrices for different upper cut-off impact heights

The method presented here uses a double differencing approach that relies on the assumption of a constant model bias within the separation distance of the RO and RS measurements. Therefore, the profile of BA departures in the approximate altitude range of the RS departures must be propagated into Tdry departures, thus using model background information within the same altitude range for both observation types. The BA departures above the approximate burst altitude of RSs are set to zero to achieve this.

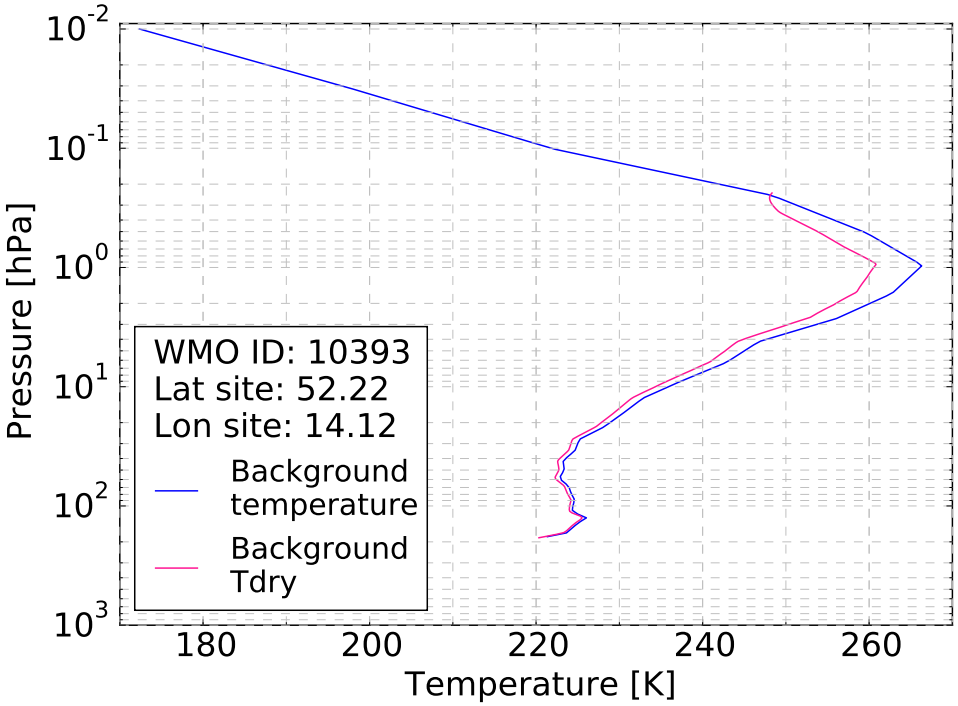


Figure 3.8.: Background temperature (blue) and Tdry (pink) for one individual profile from around the German example site. Tdry is calculated from the background BA using the non-linear Tdry calculation.

To ensure an optimal choice of this upper cut-off impact height and to understand the effect of setting high-level departure values to zero, the respective covariance matrices are analysed here. An estimate of the population covariance using a finite sample can be calculated with Eq.(3.12), where x and y are variables, $\bar{x}$ and $\bar{y}$ are the sample means and n is the sample size.

$$Cov(x,y) = \frac{\sum_{i=1}^{n}(x-\bar{x})(y-\bar{y})}{n-1} \tag{3.12}$$

The departure covariance is first calculated in BA space and is then propagated through the TL Tdry calculation to derive the Tdry departure covariances $\mathbf{C}_{Tdry}$ as $\mathbf{C}_{Tdry} = \mathbf{K}\mathbf{C}_{\alpha}\mathbf{K}^{T}$, where $\mathbf{K}$ is the linear Tdry operator in matrix form, $\mathbf{C}_{\alpha}$ is the covariance matrix of BA departures and $\mathbf{C}_{Tdry}$ is the inferred covariance matrix of Tdry departures. The diagonal elements of the covariance matrix are the variances (squared SDs), while the non-diagonal elements indicate the co-variation of the Tdry departures at one level with those at other levels.

Figure 3.9 shows the Tdry departure covariance matrices calculated from different subsamples of the BA departure profile (i.e. different upper cut-off impact heights). The colour scale suits a cut-off at 35 km, which leads to saturation of the colour for the panels with higher impact height cut-offs. The variances and covariances increase as the cut-off impact height is raised (from panel a) to f)) and a large influence is caused by setting the highest BA departure value to zero (compare panel e) to panel f)). The sensitivity to the BA at the highest observation level is caused by arbitrary assumptions about the behaviour of the BAs above the measured profile. The sensitivity to other BAs close to the

top is mainly caused by noisy BA departures at high altitudes, which potentially contain significant biases from the NWP backgrounds. The Tdry departures at each level depend on the refractivity departures at *all* levels above and each of these refractivity departures depend on *all* BA departures above. This causes the sensitivity to the (potentially large) BA departures at the top of the original profile, see also Section 3.1.4. Decreasing the upper cut-off impact height diminishes this dependence on the uppermost departures and therewith reduces the variance and covariance.

Using the TL Tdry calculation enables us to calculate the Tdry departures from a subset of BA departures, while the non-linear version would not allow this flexibility. Since RS only reach pressure levels of about 10 hPa, the influence of the NWP system and *a priori* information from altitudes above 10 hPa should be minimised. While a cut-off at 25 or 30 km is too low to allow comparisons to be made at 10 hPa, the impact height of 35 km appears to be a good choice as variance and covariance are low and, importantly, these statistics are not significantly sensitive to small changes in the cut-off impact height at around 35 km (see also Burrows and Healy 2016). Therefore the subset of the BA departures below the impact height of 35 km is used in this study and the BA departures at higher levels are set to zero. The influence caused by choosing an upper cut-off is comparable to the impact of the choice for a high-level initialisation (i.e. choice of climatology and smoothing) in the conventional retrieval chain based on statistical optimisation (see Ho et al. 2009; Ho et al. 2012; Steiner et al. 2013).

The non-linear Tdry calculation would, at all but the lowest levels, be highly dependent on the uppermost BAs, and lead to large SDs, which reach up to ≈ 7 K at the upper-air sites shown in Table 3.1 (figures not shown here). Setting the highest BA departure to zero at the least halves the SD at the 10 hPa level and further reducing the cut-off impact height decreases the SD as the Jacobians are unable to propagate information from noisy high level BA departure downwards. Given a 35 km cut-off, the SD remains under 1 K for most example stations and at most levels. Reducing the SD (σ) positively effects the standard error SE calculated as:

$$SE = \sigma / \sqrt{(n-1)} \tag{3.13}$$

The standard error indicates the quality of the estimation of the mean. Knowing the standard error of the bias correction is important to decide whether a bias correction is needed. The standard error of the bias correction is calculated following Burns and Dobson (1981), as:

$$SE_{bias\_correction} = \sqrt{\frac{\sigma_{RO}^2}{n_{RO}-1} + \frac{\sigma_{RS}^2}{n_{RS}-1}} = \sqrt{SE_{RO}^2 + SE_{RS}^2} \tag{3.14}$$

Dependence on the cut-off impact height

To understand the strong dependence of the Tdry departures to the cut-off impact height, Fig. 3.10 shows the different steps in the TL Tdry calculation, a) BA departures, b) inferred refractivity departures, c) inferred dry pressure departures, and d) inferred Tdry departures. The variables are plotted for a cut-off impact height of 55 km (pink) and no

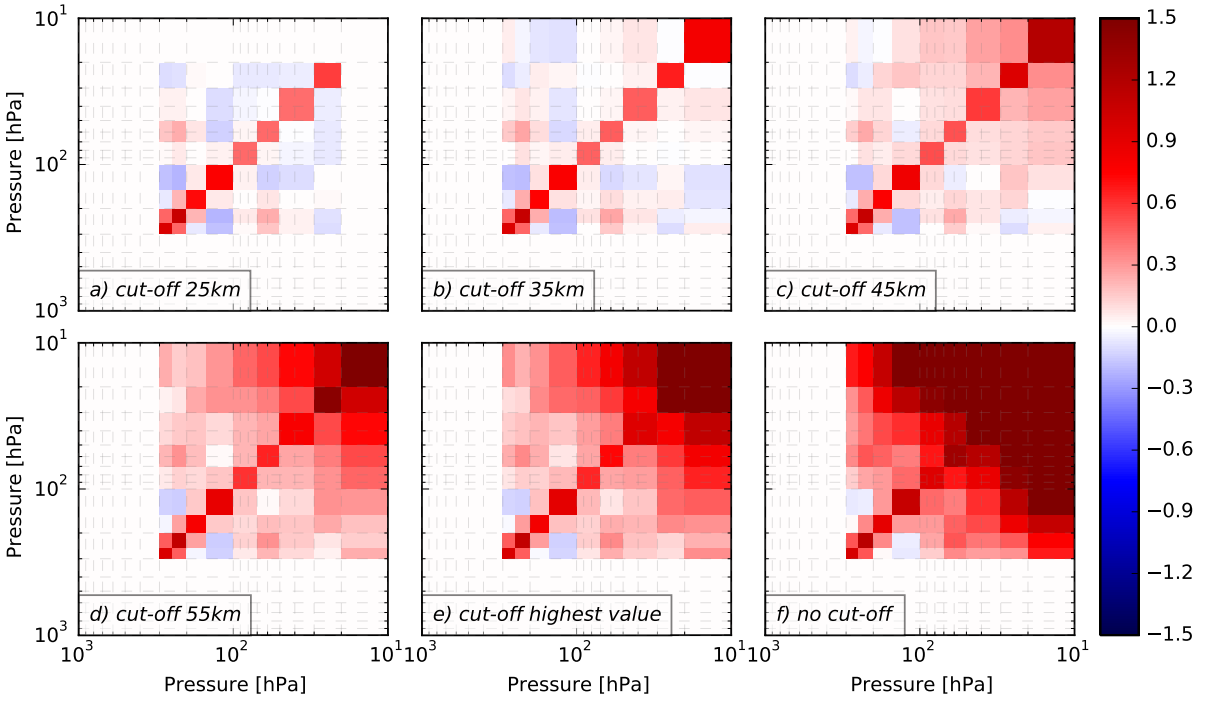


Figure 3.9.: Tdry departure covariance matrices for different upper cut-off impact heights; example site in western Russia. Colour scale for temperature squared $[K^2]$ is adjusted to a cut-off at 35 km impact height, which means that the colours are saturated for the higher cut-offs (they reach values of $\approx 36\,K^2$).

cut-off (blue).
Above 55 km the BA departures without cut-off are higher than those with cut-off at 55 km (Fig. 3.10a)), which leads to a systematic difference in the TL refractivity departures $\delta N_{no\ cut\text{-}off} > \delta N_{55km}$ (Fig. 3.10b)) calculated as $\delta\mathbf{N} = \mathbf{A}\delta\boldsymbol{\alpha}$. The pressure departure profile in Fig. 3.10c)), which is calculated as the sum of the top level pressure and the integral of the refractivity as described in Section 3.1.4, has a systematic difference of $\delta P_{no\ cut\text{-}off} > \delta P_{55km}$. Notably the differences in refractivity and pressure are present further down into the atmosphere than the differences in BA departures. Thus, a change in the upper cut-off impact height results in a different Tdry departure at all levels as can be seen in Fig. 3.10d).

How differences in the highest level BA departures propagate further down into the atmosphere can be illustrated with the Jacobian matrix of Tdry with respect to the BA (Burrows and Healy 2016, additionally show the Jacobian matrices of refractivity and pressure with respect to the BA). Figure 3.11 shows a subset of 3 typical Tdry Jacobians. Each Jacobian shows a sharp spike at the pressure level corresponding to the impact height level of the Tdry retrievals (blue 19 km impact height, pink 29 km impact height, green 39 km impact height), i.e. Tdry at a certain pressure level is highly sensitive to the BA at the same pressure level, and those just above. The sharpness of this spike is a feature of the Abel integral transform. However, the Tdry is also influenced by BAs higher up in the atmosphere and especially by the highest BA, as the long tail in the Jacobians (a feature of the hydrostatic integration in the Tdry computation) and the spike at the

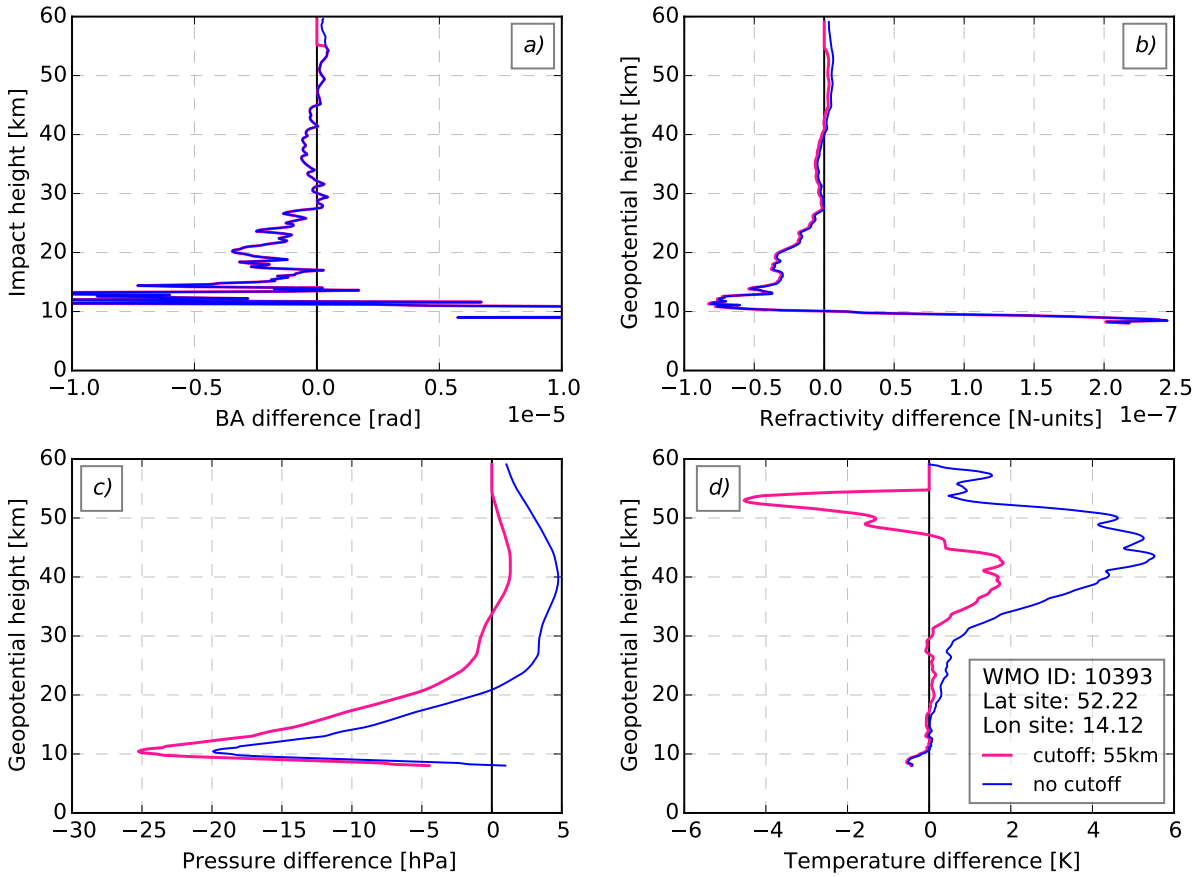


Figure 3.10.: The different steps in TL Tdry calculation for no cut-off (blue) and 55 km cut-off (pink), German example site. a) shows the BA departures, b) the inferred (TL) refractivity departures, c) the inferred (TL) dry pressure departures and d) the inferred (TL) Tdry departures.

highest level demonstrate.

3.1.10. Optimisation of high level bending angles

In the conventional RO retrieval the RO BAs are blended with a smooth *a priori* profile using optimisation techniques. In general, the statistical optimisation processing step can be written in a matrix/vector form as

$$\alpha_{smooth} = \alpha_{clim} + \mathbf{K}(\alpha_{RO} - \alpha_{clim}) \tag{3.15}$$

where α_{smooth}, α_{clim} and α_{RO} are the vectors of optimised, climatological (*a priori*) and observed BA profiles, respectively, and $\mathbf{K}$ is the gain matrix. The gain matrix can be written as $\mathbf{K} = \mathbf{B}(\mathbf{B}+\mathbf{R})^{-1}$, where $\mathbf{R}$ and $\mathbf{B}$ are the assumed covariance matrices for the observed and climatological BA profiles, respectively. These matrices are often assumed to be diagonal and typically the observation error statistics are ~ 2 microradians and the climatological errors are written as a scalar (~ 0.1) times the simulated BA value. This formulation ensures a smooth transition from observation to the climatological model in

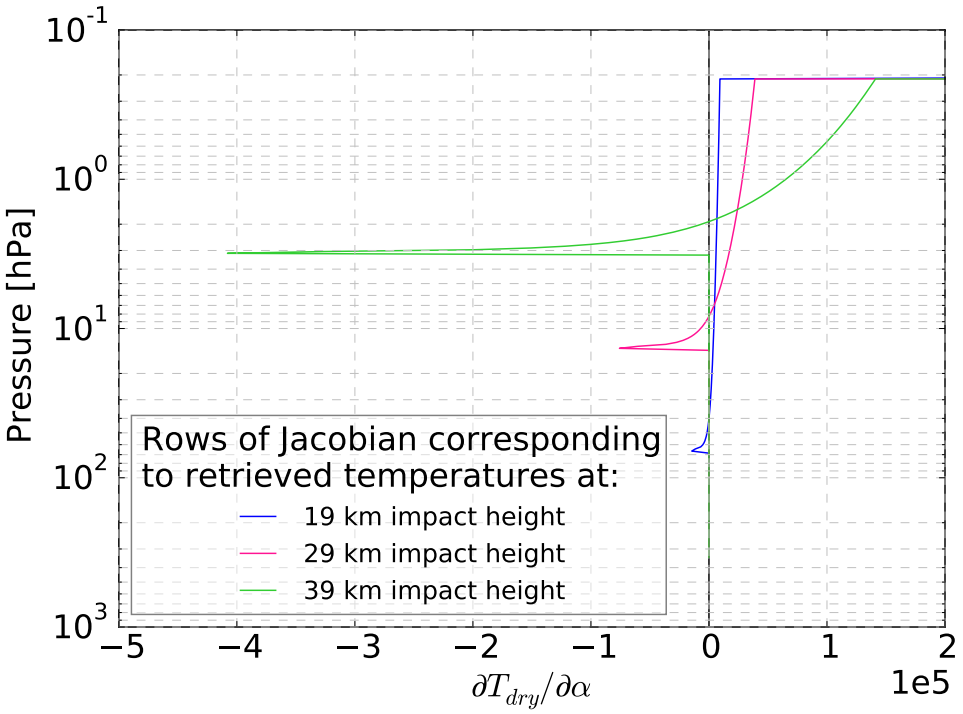


Figure 3.11.: Subset of a Jacobian matrix (partial derivative) for Tdry with respect to the BAs. The spike at the highest level is off scale, with the Jacobians reaching values of: blue: 3.8×10^5, pink: 1.65×10^6, green: 6.45×10^6.

the upper stratosphere.

The smoothed RO and model BAs which are plotted in Fig. 3.1 are calculated using a simplified smoothing as: $\alpha_{RO_{smooth}} = wt \cdot \alpha_{RO} + (1 - wt) \cdot \alpha_{clim}$, where the weight wt is calculated as $wt = \frac{\sigma^2_{clim}}{\sigma^2_{clim}+\sigma^2_{RO}}$. The assumed observation error is set to $\sigma_{RO} =$ 2 microradians and the error in the climatology is assumed to be (for two example cases) $\sigma_{clim} = 0.01 * \alpha_{clim}$ (violet dot-dashed line in Fig. 3.1) and $\sigma_{clim} = 0.1 * \alpha_{clim}$ (red dot-dashed line in Fig. 3.1). With typically-used values of climatological BA at 55 km impact height ($\approx$ 10 microradians) and an assumed 10% error on the climatology, the weight would be 0.2, which means that, at 55 km, the RO observation provides 20% of the information and the climatology 80%. Therefore the method presented here does not neglect much more observational information than other, more conventional approaches.

Bibliography

Agustí-Panareda, A, D. Vasiljevic, A. Beljaars, O. Bock, F. Guichard, M. Nuret, A. Garcia-Mendez, E. Andersson, P. Bechtold, A. Fink, H. Hersbach, J.-P. Lafore, J.-B. Ngamini, D.J. Parker, J.-L. Redelsperger, and A.M. Tompkins (2009). "Radiosonde humidity bias correction over the West African region for the special AMMA reanalysis at ECMWF". In: *Quart. J. Roy. Meteor. Soc.* 135, pp. 595–617. DOI: 10.1002/qj.396.

Anthes, R.A. (2011). "Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather". In: *Atmos. Meas. Tech.* 4, pp. 1077–1103. DOI: 10.5194/amt-4-1077-2011.

Anthes, R.A., D. Ector, D.C. Hunt, Y.-H. Kuo, C. Rocken, W.S. Schreiner, S.V. Sokolovskiy, S. Syndergaard, T.-K. Wee, Z. Zeng, P.A. Bernhardt, K.F. Dymond, Y. Chen, H. Liu, K. Manning, W.J. Randel, K.E. Trenberth, L. Cucurull, S.B. Healy, S.-P. Ho, C. McCormick, T.K. Meehan, D.C. Thompson, and N.L. Yen (2008). "The COSMIC/FORMOSAT-3 Mission: Early Results". In: *Bull. Amer. Meteor. Soc.* 313–333. DOI: 10.1175/BAMS-89-3-313.

Auligné, T., A.P. McNally, and D.P. Dee (2007). "Adaptive bias correction for satellite data in a numerical weather prediction system". In: *Quart. J. Roy. Meteor. Soc.* 133, pp. 631–642. DOI: 10.1002/qj.56.

Burns, R.B. and C.B. Dobson (1981). "Standard error of the difference between means". In: *Experimental Psychology Research Methods and Statistics*. Springer Verlag, pp. 151–157.

Burrows, C.P. and S.B. Healy (2016). *Sensitivity of radio occultation-based dry temperature retrievals to upper-level information and its relevance to radiosonde bias corrections*. Forecasting Research Technical Report 615. Met Office. URL: http://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/library/frtr_615_2016_2p.pdf.

Burrows, C.P., S.B. Healy, and I.D. Culverwell (2014). "Improving the bias characteristics of the ROPP refractivity and bending angle operators". In: *Atmos. Meas. Tech.* 7, pp. 3445–3458. DOI: 10.5194/amt-7-3445-2014.

Davies, T., M.J.P. Cullen, A.J. Malcolm, M.H. Mawson, A. Staniforth, A.A. White, and N. Wood (2005). "A new dynamical core for the Met Office's global and regional modelling of the atmosphere". In: *Quart. J. Roy. Meteor. Soc.* 131, pp. 1759–178. DOI: 10.1256/qj.04.101.

Dee, D.P. (2004). *Variational bias correction of radiance data in the ECMWF system*. ECMWF conference paper. European Centre for Medium-Range Weather Forecasts. URL: http://www.ecmwf.int/sites/default/files/elibrary/2004/8930-variational-bias-correction-radiance-data-ecmwf-system.pdf.

Derber, J.C. and W.-S. Wu (1998). "The Use of TOVS Cloud-Cleared Radiances in the NCEP SSI Analysis System". In: *Mon. Wea. Rev.* 126, pp. 2287–2299.

Dirksen, R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, and H. Vömel (2014). "Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde". In: *Atmos. Meas. Tech.* 7, pp. 4463–4490. DOI: 10.5194/amt-7-4463-2014.

Eyre, J.R. (1992). *A bias correction scheme for simulated TOVS brightness temperatures*. ECMWF Technical Memorandum 186. European Centre for Medium-Range Weather Forecasts. URL: http://www.ecmwf.int/sites/default/files/elibrary/1992/9330-bias-correction-scheme-simulated-tovs-brightness-temperatures.pdf.

– (2016). "Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties ". In: *Quart. J. Roy. Meteor. Soc.* 142.699, pp. 2284–2291.

Fjeldbo, G., A.J. Kliore, and V.R. Eshleman (1971). "The neutral atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments". In: *The Astronomical Journal* 76.2, pp. 123–140.

GCOS-112 (2007). *GCOS REFERENCE UPPER-AIR NETWORK (GRUAN): Justification, requirements, siting and instrumentation options*. Tech. rep. GCOS-112 (WMO/TD No. 1379). World Meteorological Organization. URL: https://www.wmo.int/pages/prog/gcos/Publications/gcos-112.pdf.

Haimberger, L., C. Tavolato, and S. Sperka (2012). "Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations". In: *J. Climate* 25, pp. 8108–3131. DOI: :10.1175/JCLI-D-11-00668.1.

Harris, B.A. and G. Kelly (2001). "A satellite radiance-bias correction scheme for data assimilation". In: *Quart. J. Roy. Meteor. Soc.* 127, pp. 1453–1468. DOI: 10.1002/qj.49712757418.

Hawson, C.L. and P.G.F. Caton (1961). "A synoptic method for the international comparison of geopotential observations". In: *The Meteorological Magazine* 90.1073, pp. 333–364.

He, W., S.-P. Ho, H. Chen, X. Zhou, D. Hunt, and Y.-H. Kuo (2009). "Assessment of radiosonde temperature measurements in the upper troposphere and lower stratosphere using COSMIC radio occultation data". In: *Geophys. Res. Lett.* 36. DOI: 10.1029/2009GL038712.

Healy, S.B. (2001). "Smoothing radio occultation bending angles above 40 km". In: *Ann. Geophysicae* 19, pp. 459–468.

– (2008a). "Assimilation of GPS Radio Occultation Measurements at ECMWF". In: *GRAS SAF Workshop on Applications of GPSRO Measurements, 16-18 June 2008*. Available at http://www.romsaf.org/Workshops/agrom_prog/Healy.pdf. European Centre for Medium-Range Weather Forecasts. ECMWF, Shinfield Park, Reading, UK, pp. 99–109.

– (2008b). "Forecast impact experiment with a constellation of GPS radio occultation receivers". In: *Atmos. Sci. Let.* 9, pp. 111–118. DOI: 10.1002/asl.169.

Healy, S.B. and I.D. Culverwell (2015). "A modification to the standard ionospheric correction method used in GPS radio occultation". In: *Atmos. Meas. Tech.* 8, pp. 3385–3393. DOI: 10.5194/amt-8-3385-2015.

Healy, S.B. and J.-N. Thépaut (2006). "Assimilation experiments with CHAMP GPS radio occultation measurements". In: *Q. J. R. Meteorol. Soc.* 132, pp. 605–623. DOI: 10.1256/qj.04.182.

Healy, S.B., J. Wickert, G. Michalak, T. Schmidt, and G. Beyerle (2007). "Combined forecast impact of GRACE-A and CHAMP GPS radio occultation bending angle profiles". In: *Atmos. Sci. Let.* 8, pp. 43–50. DOI: 10.1002/asl.149.

Ho, S.-P., D. Hunt, A.K. Steiner, A.J. Mannucci, G. Kirchengast, H. Gleisner, S. Heise, A. von Engeln, C. Marquardt, S. Sokolovskiy, W. Schreiner, B. Scherrlin-Pirscher, C. Ao, J. Wickert, S. Syndergaard, K.B. Lauritsen, S.S. Leroy, E.R. Kursinski, Y.-H. Kuo, U. Foelsche, T. Schmidt, and M. Gorbunov (2012). "Reproducibility of GPS radio occultation data for climate monitoring: Profile-to-profile inter-comparison of CHAMP climate records 2002 to 2008 from six data centers". In: *J. Geophys. Res.* 117. DOI: 10.1029/2012JD017665.

Ho, S.-P., G. Kirchengast, S. Leroy, J. Wickert, A.J. Mannucci, A.K. Steiner, D. Hunt, W. Schreiner, S. Sokolovskiy, C. Ao, M. Borsche, A. von Engeln, U. Foelsche, S. Heise, B. Iijima, Y.-H. Kuo, R. Kursinski, B. Pirscher, M. Ringer, C. Rocken, and T. Schmidt (2009). "Estimating the uncertainty of using GPS radio occultation data for climate monitoring: Intercomparison of CHAMP refractivity climate records from 2002 to 2006 from different data centers". In: *J. Geophys. Res.* 114. DOI: 10.1029/2009JD011969.

Hoffman, R.N., J.-F. Louis, and T. Nehrkorn (1992). *A method for implementing adjoint calculations in the discrete case*. ECMWF Technical Memorandum 184. European Centre for Medium-Range Weather Forecasts.

Ingleby, B. and D. Edwards (2015). "Changes to radiosonde reports and their processing for numerical weather prediction". In: *Atmos. Sci. Let.* 16, pp. 44–49. DOI: 10.1002/asl2.518.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Honoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi (2015). "The JRA-55 Reanalysis: General Specifications and Basic Characteristics". In: *J. Meteorol. Soc. Jpn.* 93, pp. 5–48. DOI: 10.2151/jmsj.2015-001.

Kursinski, E.R., G.A. Hajj, S.S. Leroy, and B. Herman (2000). "The GPS Radio Occultation Technique". In: *Terrestrial, Atmospheric and Oceanic Sciences* 11.1, pp. 53–114.

Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, and K.R. Hardy (1997). "Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System". In: *J. Geophys. Res.* 102, pp. 23429–23465.

Ladstädter, F., A.K. Steiner, M. Schwärz, and G. Kirchengast (2015). "Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002-2013". In: *Atmos. Meas. Tech.* 8.3, pp. 1819–1834. DOI: 10.5194/amt-8-1819-2015.

Leroy, S.S., J.A. Dykema, and J.G. Anderson (2006). "Climate Benchmarking Using GNSS Occultation". In: *Atmosphere and Climate Studies by Occultation Methods*. Ed. by U. Foelsche, G. Kirchengast, and A. Steiner. Springer-Verlag, pp. 287–301.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata (2013). "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the media". In: *J. Exp. Soc. Psychol.* 49, pp. 764–766. DOI: 10.1016/j.jesp.2013.03.013.

Milan, M. and L. Haimberger (2015). "Predictors and grouping for bias correction of radiosonde temperature observations". In: *J. Geophys. Res. Atmos.* 120, pp. 10736–10766. DOI: 10.1002/2015JD023635.

Philipona, R., A. Kräuchi, G. Romanens, G. Levrat, P. Ruppert, E. Brocard, P. Jeannet, D. Ruffieux, and B. Calpini (2013). "Solar and Thermal Radiation Errors on Upper-Air Radiosonde Temperature Measurements". In: *J. Atmos. Oceanic Technol.* 30, pp. 2382–2392. DOI: 10.1175/JTECH-D-13-00047.1.

Poli, P., S.B. Healy, and D.P. Dee (2010). "Assimilation of Global Positioning System radio occultation data in the ECMWF ERA-Interim reanalysis". In: *Q. J. R. Meteorol. Soc.* 136, pp. 1972–1990. DOI: 10.1002/qj.722.

Rawlins, F., S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc, and T.J. Payne (2007). "The Met Office global four-dimensional variational data assimilation scheme". In: *Quart. J. Roy. Meteor. Soc.* 133, pp. 347–362. DOI: 10.1002/qj.32.

Rennie, M.P. (2010a). *Investigation into the consistency of radiosonde and GPSRO observations and their application in NWP at the Met Office*. Met Office Forecasting R&D Technical Report 541.

– (2010b). "The impact of GPS radio occultation assimilation at the Met Office". In: *Quart. J. Roy. Meteor. Soc.* 136, pp. 116–131. DOI: 10.1002/qj.521.

Rodgers, C.D. (2000). *Inverse Methods for Atmospheric Sounding, Theory and Practice*. World Scientific.

Scherllin-Pirscher, B., G. Kirchengast, A.K. Steiner, Y.-H. Kuo, and U. Foelsche (2011). "Quantifying uncertainty in climatological fields from GPS radio occultation: an empirical-analytical error model". In: *Atmos. Meas. Tech.* 4, pp. 2019–2034. DOI: 10.5194/amt-4-2019-2011.

Simmons, A.J., P. Poli, D.P. Dee, P. Berrisford, H. Hersbach, S. Kobayashi, and C. Peubey (2014). "Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim". In: *Quart. J. Roy. Meteor. Soc.* 140, pp. 329–353. DOI: 10.1002/qj.2317.

Smith, E.K. and S. Weintraub (1953). "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies". In: *J. Res. Natl. Bur. Stand. (U. S.)* 50.1, pp. 39–41.

Steiner, A.K., D. Hunt, S.-P. Ho, G. Kirchengast, A.J. Mannucci, B. Scherllin-Pirscher, H. Gleisner, A. von Engeln, T. Schmidt, C. Ao, S.S. Leroy, E.R. Kursinski, U. Foelsche, M. Gorbunov, S. Heise, Y.-H. Kuo, K.B. Lauritsen, C. Marquardt, C. Rocken, W. Schreiner, S. Sokolovskiy, S. Syndergaard, and J. Wickert (2013). "Quantification of structural uncertainty in climate data records from GPS radio occultation". In: *Atmos. Chem. Phys.* 13.3, pp. 1469–1484. DOI: 10.5194/acp-13-1469-2013.

Steiner, A.K., G. Kirchengast, and H.P. Ladreiter (1999). "Inversion, error analysis, and validation of GPS/MET occultation data". In: *Ann. Geophysicae* 17, pp. 122–138.

Sun, B., A. Reale, S. Schroeder, D.J. Seidel, and B. Ballish (2013). "Toward improved corrections for radiation-induced biases in radiosonde temperature observations". In: *J. Geophys. Res.* 118, pp. 4231–4243. DOI: 10.1002/jgrd.50369.

Sun, B., A. Reale, D.J. Seidel, and D.C. Hunt (2010). "Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics". In: *J. Geophys. Res.* 115. DOI: 10.1029/2010JD014457.

Syndergaard, S. (1999). *Retrieval Analysis and Methodologies in Atmospheric Limb Sounding Using the GNSS Radio Occultation Technique*. Danish Meteorological Institute Scientific Report 99-6. Danish Meteorological Institute.

Tradowsky, J.S. (2015). *Characterisation of radiosonde temperature biases and errors using radio occultation measurements*. ROM SAF Visiting Scientist report 26. Radio Occultation Meteorology Satellite Application Facility. URL: http://www.romsaf.org/Publications/reports/romsaf_vs26_rep_v12.pdf.

– (2016). *Radiosonde Temperature Bias Corrections using Radio Occultation Bending Angles as Reference*. ROM SAF Visiting Scientist report 31. Radio Occultation Meteorology Satellite Application Facility. URL: http://www.romsaf.org/Publications/reports/romsaf_vs31_rep_v10.pdf.

Walters, D., N. Wood, S. Vosper, and S. Milton (2014). *ENDGame: A new dynamical core for seamless atmospheric prediction*. Met Office Report. Available at http://www.metoffice.gov.uk/media/pdf/s/h/ENDGameGOVSci_v2.0.pdf. UK Met Office.

3.2. Summary of the ROM SAF visiting scientist reports

In the interest of brevity, the two ROM SAF visiting scientist reports (Tradowsky 2015; Tradowsky 2016), on which the paper reprinted above is based, are not fully reprinted here although they give far more detailed and additional information. To give an overview of the content, a slightly adapted version of their executive summaries is given below. Te interested reader is referred to the publicly available reports for further details.

Executive summary of Tradowsky (2015) Measurements of atmospheric temperature have been made for decades and nowadays RS and RO measurements are used to anchor the temperature in NWP models. However, the impact of the high quality RO data might be limited due to opposing biases between the observation types. Therefore, for better exploitation of highly accurate RO measurements an improved bias correction in RS temperatures is needed. Here we use a new method to calculate the RS temperature bias on a station-by-station basis for different solar elevation angle (SEA) ranges. The temperature bias corrections are calculated using fields from the Met Office Unified Model as a transfer medium. Thus, the bias correction is calculated based on the background departure statistics, i.e. the difference between the measurement and the model background (short-range forecast). This method has two advantages: (i) compared to direct observation-to-observation collocations, the influence of differences in time and space is minimised since every measurement (both RO and RS) has a collocated model background profile, and (ii) the lowest level where the calculated model dry temperature (Tdry) is acceptably close to the model temperature, i.e. where the atmospheric humidity is negligible, is determined for each profile, which allows the RO Tdry profiles to be used as low in the atmosphere as is reasonable.

For this investigation the RO departure statistics are first calculated in BA space and are then propagated to Tdry departures using tangent linear versions of the linear Abel transform to obtain refractivity departures, and the hydrostatic integration of the refractivity to obtain Tdry departures. This method was suggested by Sean Healy (personal communication) and is, for the first time described here in detail. Using the tangent linear version of the Tdry calculation enables us to select the maximum impact height at which the RO BA departures are used, while the full non-linear version would not allow this flexibility. Choosing this upper limit above which the BA departures are set to zero is a crucial decision, since it influences the results. Since RSs only reach pressure levels of about 10 hPa, the aim is to use the minimum amount of data above 10 hPa in order to eliminate the effect of model biases at higher levels.

The results of this project show that the RS temperature bias varies from station-to-station. The temperature bias depends *inter alia* on the RS type, the correction software of the vendor and the position of the sun. While the calculated bias correction stays below 0.5 - 1 K for sites launching e.g. the Finish Vaisala RS92, the biases at other stations can be considerably higher. For most stations the bias increases with the altitude and reaches its maximum at 10 hPa, which is the lowest analysed pressure level. For some Russian sites biases of 2-3 K occur at the higher level in the atmosphere. Interestingly, the Russian sondes tend to have a cold bias at higher levels for all SEAs. In contrast the Vaisala RS92,

which is known as one of the most accurate RSs, tend to have a slight cold bias at the lower levels and a slight warm bias at the highest levels for high SEAs. RS92s launched at night tend to have a slight cold bias at most altitudes.

The large temperature biases found at some stations show the importance of applying an up-to-date bias correction to the RS temperature profiles before they are assimilated into NWP models. Thereby opposing biases in RO and RS measurements can be minimised. The results are intended to be used in a forecast impact study and could afterwards be implemented into operational weather forecast systems, where they would need to be updated on a regular basis.

Executive summary of Tradowsky (2016) For the purpose of NWP, both RO and RS measurements are used to anchor the temperature. Thus, it is important to try to eliminate biases between these two observation types prior to the assimilation into the numerical weather forecasts. Following the development of a new method to estimate corrections for stratospheric RS temperature biases using RO BAs as an unbiased reference as presented in ROM SAF VS26 (Tradowsky 2015), this study focuses on three work packages.

1. Station-by-station temperature bias corrections to be applied in a forecast impact study using the Met Office global NWP system are calculated for 852 upper-air stations. While the stratospheric correction is calculated using a double differencing approach of RS and RO background departures, smoothing towards the default correction is done in the troposphere, where RS biases are generally smaller.
2. Rather than using statistical optimisation of high level BAs as is done in the conventional RO retrieval, the method applied here (see Tradowsky 2015) applies a cut-off impact height above which all BA departures are set to zero before they are propagated into Tdry departures using a linear operator. The cut-off is applied for two purposes, (i) to eliminate the effect of noisy high level BAs and (ii) to enable the use of double differencing, which relies on the assumption of a constant model bias over the separation distance of the RO and RS measurements — hence, including model background information up to a significantly higher level in the RO departures than in the RS departures would invalidate this central assumption. While the cut-off is therefore needed in this project, it also adds structural uncertainty to the derived Tdry departures, similar to the way in which choosing an implementation of statistical optimisation would add structural uncertainty.
3. The GRUAN reference-quality RS data product is compared with RO in the stratosphere using the double differencing method developed for the correction of RS biases. As the RO technique and the GRUAN data product both offer high quality data, good agreement is expected.

The bias correction profiles calculated here are prepared for a forecast impact study to be performed at the Met Office and include corrections for four different SEA ranges. If the sample size within a SEA range is too small to calculate a correction, SEA ranges are combined and it is shown that this is a suitable estimate of the bias correction for the majority of sites. This report also discusses how the tangent linear retrieval which is used to calculate the Tdry departures from BA departures compares to other RO retrievals. While optimisation of high level BAs with a smooth *a priori* profile is used in the conven-

tional retrieval, here we explicitly set the high level BA departures to zero and calculate the Tdry departures from a subset of BA departures. This is essential in order to use an NWP model as the transfer medium in the comparison of two measurement techniques. It is shown that cutting off high level BA departures indeed gives similar results to applying a simplified algorithm that approximates the results that would be obtained using statistical optimisation. The comparison of GRUAN temperature departures with RO Tdry departures illustrates another usage of the technique. For most GRUAN sites, good agreement between RO and GRUAN is found at most levels. The detailed analysis will be presented in Chapter 4.

4. Comparison of GRUAN radiosonde temperature profiles with dry temperatures derived from radio occultation bending angles

The method developed in Chapter 3 to bias correct RSs[9] prior to their assimilation into NWP can also be used to compare GRUAN data products and RO observations. As both GRUAN data products and RO data are commonly used as a reference in the comparison with other data sets, the conformity of those observation types will be evaluated within this chapter. Agreement or consistency (as defined in Immler et al. 2010) is expected between data sets that account for all biases and provide valid uncertainty estimates. Therefore, the comparison of these two independent measurement techniques has the potential to reveal unaccounted measurement biases and retrieval problems. RO profiles from the COSMIC mission will be compared with version 2 of the GRUAN data product for the Vaisala RS92 instrument (Dirksen et al. 2014). The UK Met Office model will serve as a transfer standard as required for the double differencing technique. The tangent linear RO retrieval described in Tradowsky et al. (2017) and Chapter 3 is used to calculate Tdry departures from BA departures. This analysis was performed as part of the ROM SAF visiting scientist project 31 and the following text has been adapted from Chapter 6 of Tradowsky (2016) to fit the context of this thesis.

The double differencing technique described in Tradowsky (2015) and Tradowsky et al. (2017) is used here to compare GRUAN and RO observation minus background (O-B) statistics on a station-by-station basis. First, the departures of the GRUAN temperatures from the model background temperature are calculated for every GRUAN profile and then the mean departure profile at a given station is calculated on SPLs.

As the GRUAN data processing does not change over years, it is highly unlikely that there would be any discontinuities in the time series at any one GRUAN location resulting from data processing. The same could not be the case for vendor provided RS data which could be subject to changes in the processing. It is therefore justifiable to use the GRUAN-processed RS92 measurements made during 2014 and 2015 for this analysis. As described in Section 3.1.4 COSMIC BA data within 500 km of a GRUAN site are used to calculate the mean BA departure profile which is then propagated into a mean Tdry departure profile using the tangent linear retrieval chain described in Tradowsky (2015). The sensitivity to the vicinity radius was studies in Tradowsky (2015) and it was found that that increasing the radius is not altering the mean results, but is increasing the SD slightly. The double differencing method applied here is using the model background as a transfer standard to diminish the effect of imperfect collocation which leads to significantly

[9] All abbreviations can be found in the Glossary, Appendix B

smaller SDs compared to the use of direct collocations as they have been analysed in e.g. Sun et al. (2010), Sun et al. (2013), and Ladstädter et al. (2015). The double differencing method is based on the assumption of a constant NWP forecast bias (but a varying atmosphere) over the separation distance between the measurements (chosen to be ≤ 500 km). This is a more robust assumption compared to assuming a non-varying atmosphere over the separation distance as it is required for direct collocations. The RO Tdry departures calculated from all COSMIC profiles measured within 500 km of the GRUAN site during 2014 and 2015 are then compared to the GRUAN temperature departures in the sufficiently dry part of the atmosphere (above the tropopause).

As both RO and GRUAN provide high quality data which are used as reference data products, good agreement between the profiles is expected. However, while GRUAN has the lowest random uncertainty in the troposphere, the "core-region" of RO is in the in the upper troposphere and lower stratosphere. Within this altitude region RO profiles exhibits the highest accuracy (Kursinski et al. 2000) and the effect of *a priori* information used for the upper-level initialisation and to resolve the temperature/water vapour ambiguity in the troposphere is small (Steiner et al. 2013; Tradowsky et al. 2017). To investigate if both data sets are consistent, it is essential to include an estimate of the uncertainty on both data sets in any comparison. The GRUAN data product includes an estimate of the uncertainty on every datum, which is propagated into the uncertainty on the mean GRUAN O-B departures. The following sections describe (i) how an uncertainty weighted average of the GRUAN departures is calculated and (ii) how the uncertainty on the mean GRUAN temperature departure is calculated from the uncertainties on the individual profiles. Furthermore, the structural uncertainty in the RO Tdry departures is estimated and finally the RO and GRUAN departures are compared and discussed.

4.1. Calculation of the mean GRUAN departures and the associated uncertainties

As the GRUAN profiles come with an estimate of the uncertainty on each datum, these uncertainties can be propagated into uncertainties on the mean GRUAN departures from model background values. For this purpose, the variable *u_temp* which is the standard uncertainty (k=1) of the air temperature provided in the GRUAN data files, is used. This is the total uncertainty of temperature, which is composed of correlated and uncorrelated components (see Sommer et al. 2014; Dirksen et al. 2014). This uncertainty is propagated through the calculation of mean departures and is displayed in the uncertainty bars of GRUAN departures. The different steps in the calculation of the GRUAN departures are, (i) interpolation of temperatures to SPLs, (ii) calculation of background departures, and (iii) calculation of mean background departures for every available GRUAN site.

4.1.1. Interpolation of GRUAN temperatures to standard pressure levels

The linear interpolation to the SPLs is done in logarithmic pressure using the following equation

$$T_{GRUAN}(P) = (1-\beta)\cdot T_0 + \beta \cdot T_1 \tag{4.1}$$

where $\beta = \frac{ln(P)-ln(P_0)}{ln(P_1)-ln(P_0)}$, T_0 is the temperature at the next highest pressure P_0 (lower altitude), T_1 is the temperature at the next lowest pressure P_1 (higher altitude) and P is the pressure level to which the temperature is interpolated. The uncertainty on the interpolated temperature can be calculated using the error propagation equation (see e.g. Eq. (3.13) in Bevington and Robinson 2003)), taking into account the uncertainty on the temperature profile, but not the uncertainty on the pressure, as:

$$\sigma_{T(P)} = \sqrt{\sigma_{T_0}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)^2 + \sigma_{T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)^2 + 2\cdot \sigma_{T_0T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)\cdot\left(\frac{\partial T(P)}{\partial T_1}\right)} \tag{4.2}$$

The GRUAN RS92 version 2 data products does not include the covariance matrix which contains $\sigma_{T_0T_1}$. Therefore, the limits of the uncertainty are estimated by analysing the three limits of the correlation coefficient R, which can be used to rewrite Eq. (4.2) as:

$$\begin{aligned}\sigma_{T(P)} &= \\ &= \sqrt{\sigma_{T_0}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)^2 + \sigma_{T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)^2 + 2\cdot \sigma_{T_0}\cdot\sigma_{T_1} \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)\cdot\left(\frac{\partial T(P)}{\partial T_1}\right)\cdot R}\end{aligned} \tag{4.3}$$

where $-1 \leq R \leq 1$ is a coefficient describing the interdependence of σ_{T_0} and σ_{T_1}.

In the following, uncertainty estimates for three values of R are discussed to inform the calculation of the uncertainty on the interpolated GRUAN temperature.

1. Uncorrelated uncertainties, $R = 0$:
   If the uncertainties are uncorrelated, the covariance term is zero and the uncertainty on the interpolated temperature can be calculated as:

$$\sigma_{T(P)} = \sqrt{\sigma_{T_0}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)^2 + \sigma_{T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)^2} = \sqrt{\sigma_{T_0}^2\cdot(1-\beta)^2 + \sigma_{T_1}^2\cdot\beta^2} \tag{4.4}$$

2. Completely negatively correlated uncertainties ($R = -1$):
   If σ_{T_0} and σ_{T_1} are anti-correlated, $\sigma_{T(P)}$ is lower than in case 1., i.e.

$$\begin{aligned}\sigma_{T(P)} &= \\ &= \sqrt{\sigma_{T_0}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)^2 + \sigma_{T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)^2 - 2\cdot \sigma_{T0}\cdot\sigma_{T1} \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)\cdot\left(\frac{\partial T(P)}{\partial T_1}\right)} \\ &= |(1-\beta)\cdot\sigma_{T_0} - \beta\cdot\sigma_{T_1}|\end{aligned} \tag{4.5}$$

The uncertainty in $T_{GRUAN}(P)$ can become very small if the uncertainties on T_0 and T_1 are anti-correlated. In particular, $\sigma_{T(P)} = 0$ for $\beta = \sigma_{T_0}/(\sigma_{T_1} - \sigma_{T_0})$.

3. Completely positively correlated uncertainties ($R = 1$):
   If σ_{T_0} and σ_{T_1} are completely positively correlated, $\sigma_{T_{GRUAN}(P)}$ is calculated as a linear interpolation of the temperature uncertainties in logarithmic pressure.

$$\begin{aligned}\sigma_{T(P)} &= \\ &= \sqrt{\sigma_{T_0}^2 \cdot \left(\frac{\partial T(P)}{\partial T_0}\right)^2 + \sigma_{T_1}^2 \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)^2 + 2 \cdot \sigma_{T0} \cdot \sigma_{T1} \cdot \left(\frac{\partial T(P)}{\partial T_0}\right) \cdot \left(\frac{\partial T(P)}{\partial T_1}\right)} \\ &= \sigma_{T_0} \cdot (1 - \beta) + \sigma_{T_1} \cdot \beta\end{aligned} \tag{4.6}$$

For this study we assume the uncertainties on T_1 and T_2 to be fully positive correlated. If the error covariances are made available for the GRUAN data product at some time in the future, it would be possible to improve the estimate of the uncertainty.

The estimation of uncertainties on interpolated values, as applied here, does not take into account that the temperature at a level between two measurement levels may be outside the range of temperatures observed at the level above and below. It is not straight forward to account for this uncertainty and, due to the high vertical resolution of GRUAN profiles, the effect is expected to be small. It could be examined if kriging would provide a better means to interpolate and estimate the uncertainty on interpolated values.

4.1.2. Calculation of background departures (O-Bs)

The departures of the Met Office Unified Model temperatures from the GRUAN temperatures are calculated as:

$$OB = T_{GRUAN} - T_{background} \tag{4.7}$$

The model uncertainty is assumed to be zero as the model is only used as a transfer standard, i.e. the central assumption is that the model bias is constant over the distance between the GRUAN and RO profiles. Thus, the uncertainty on the departures equals the uncertainty on T_{GRUAN}.

$$\sigma_{OB} = \sigma_T \tag{4.8}$$

Propagating the GRUAN uncertainty through the calculations, results in a profile of uncertainties associated with each departure profile.

4.1.3. Calculation of mean GRUAN temperature departures

Following Bevington and Robinson (2003), the mean GRUAN temperature departure is calculated as a weighted average, where the weights are determined by the uncertainties

on the individual temperature departures σ_{OB} as:

$$\overline{OB_{GRUAN}} = \frac{\sum_{i=1}^{N} \frac{1}{\sigma^2_{OB_i}} \cdot OB_i}{\sum_{i=1}^{N} \frac{1}{\sigma^2_{OB_i}}} \tag{4.9}$$

The uncertainty of the weighted mean is calculated as (see Bevington and Robinson 2003):

$$\sigma_{\overline{OB_{GRUAN}}} = \sqrt{\frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma^2_{OB_i}}} \cdot \frac{1}{N-1} \sum_{i=1}^{N} \frac{(OB_i - \overline{OB})^2}{\sigma^2_{OB_i}}} \tag{4.10}$$

taking into account the SD of the departures, the sample size and the weights, which are determined by the GRUAN uncertainties.

4.2. Analysing structural uncertainty in the RO retrieval at GRUAN sites

In this section, the RO Tdry departures calculated from the COSMIC BA departures around the GRUAN sites are shown for different upper cut-off impact heights. For an explanation of the upper cut-off height, see Tradowsky et al. (2017) and Chapter 3. Within this project, an impact height cut-off at 35km is used to analyse the difference between RO and GRUAN O-B departures, but reasonable variations in the upper cut-off impact height would be conceivable. The Figs 4.1-4.6 show the Tdry departures for cut-off impact heights of 25 km, 27.6 km, 30 km, 32.6 km, 35 km, 37.6 km, 40 km and 42.6 km. While the spread between departures stays below ~ 0.1 K at some stations (i.e. Barrow, Ny Ålesund) it is larger at others, reaching ~ 0.25 K at 10 hPa.

A cut-off close to the altitude of the balloon burst is needed to use the double differencing method which relies on the assumption of a constant model bias over the separation distance between the measurements. However, choosing a cut-off impact height adds structural uncertainty as the O-B statistics depend on the choice. While using a cut-off at 35 km seems to be a reasonable choice in this project, cutting off at 30 km or 40 km could also be adequate. Therefore the range (maximum minus minimum value) of the Tdry departures with cut-offs between 30 km and 40 km impact height is used to estimate the structural uncertainty of the Tdry departure. The structural uncertainty reaches the following values depending on the upper air site: Lindenberg: 0.176 K, Cabauw: 0.183 K, Barrow: 0.078 K, Southern Great Plains: 0.192 K, Sodankylä: 0.145 K, and Ny Ålesund: 0.091 K. A general discussion of the structural uncertainty induced by the cut-off and a comparison to the structural uncertainty in the conventional RO retrieval, is given in Chapter 3, Tradowsky (2016) and Tradowsky et al. (2017).

Based on the burst altitudes of balloons launched at GRUAN sites, it seems to be reason-

able to choose cut-off impact heights between 30 km and 40 km to estimate the structural uncertainty. Recall that it is important to use a cut-off impact height approximately coincident with burst point altitude of the RSs. If the cut-off impact height is chosen too low, it precludes the calculation of RO Tdry departures at the highest level where GRUAN departures are available. On the other hand, choosing a cut-off impact height above the altitude that is reached by the RSs, results in the downward propagation of model biases in the RO retrieval (see Burrows and Healy 2016), invalidating the central assumption needed for double differencing.

In general, it would be possible to choose a station-dependent upper impact height cut-off for GRUAN stations. This would allow the average burst altitude of RSs to be taken into account. For example, the majority of the RSs launched in Lindenberg reach 10 hPa, while this is exceptional for the Southern Great Plains site. The structural uncertainty could then be estimated by variation of this station dependent cut-off impact height. This improvement could be facilitated in a future extension of this study.

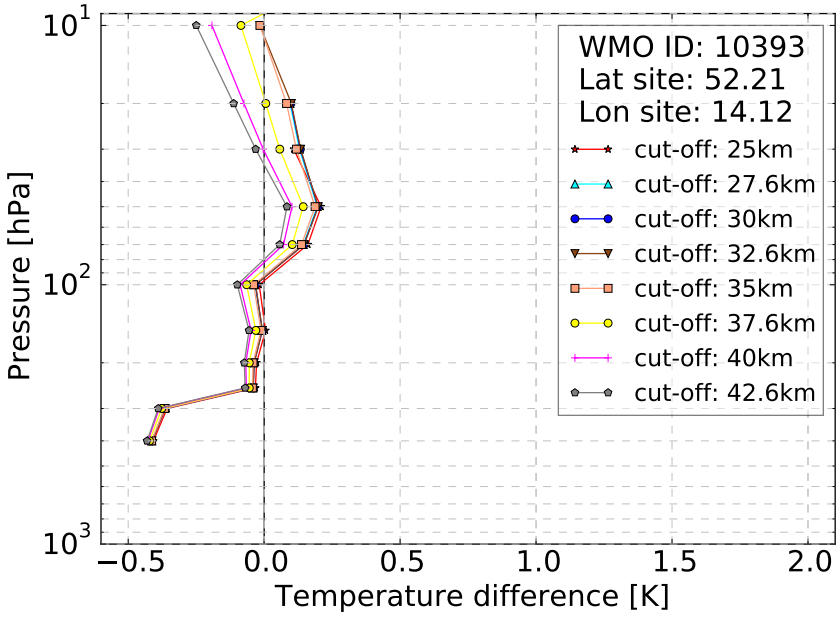


Figure 4.1.: Sensitivity to the upper cut-off impact height at the GRUAN station in Lindenberg.

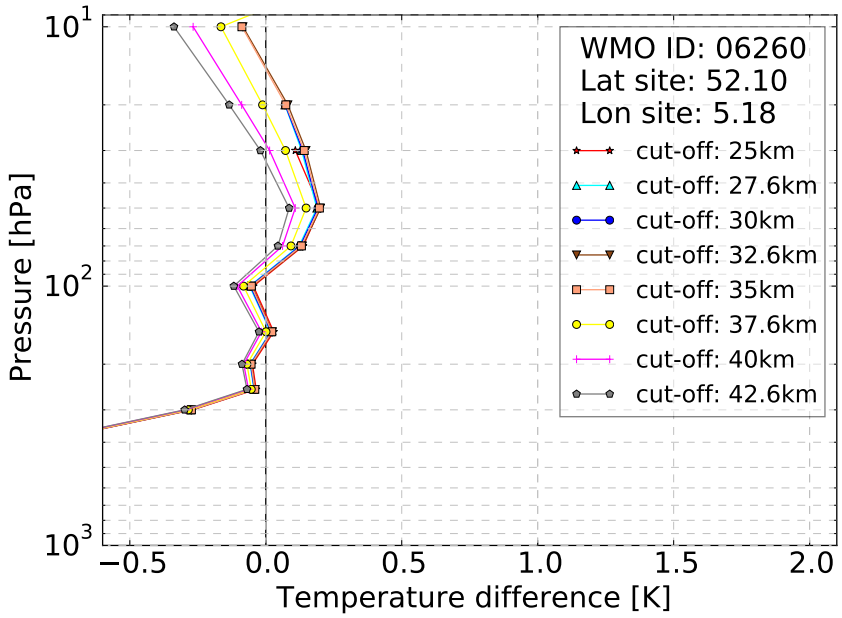


Figure 4.2.: Sensitivity to the upper cut-off impact height at the GRUAN station in Cabauw.

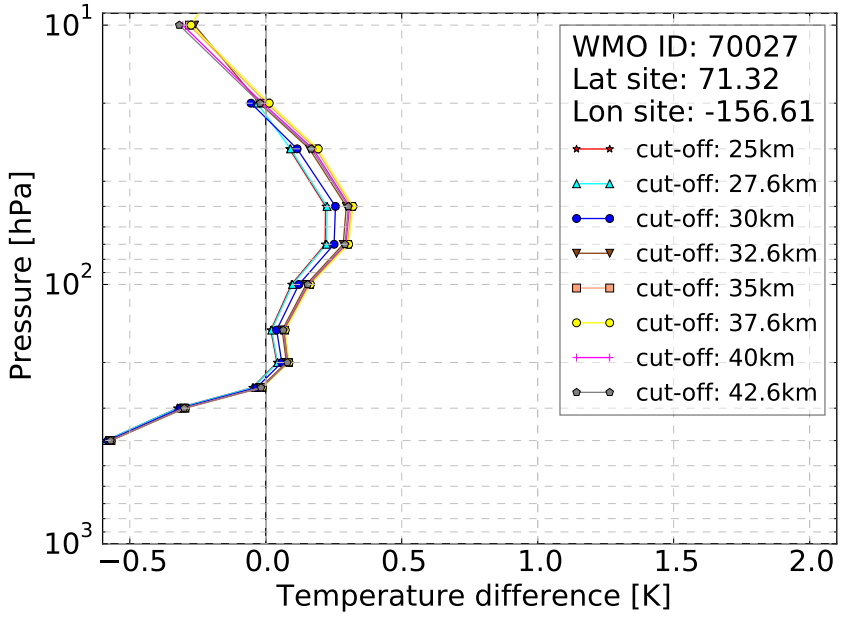


Figure 4.3.: Sensitivity to the upper cut-off impact height at the GRUAN station in Barrow.

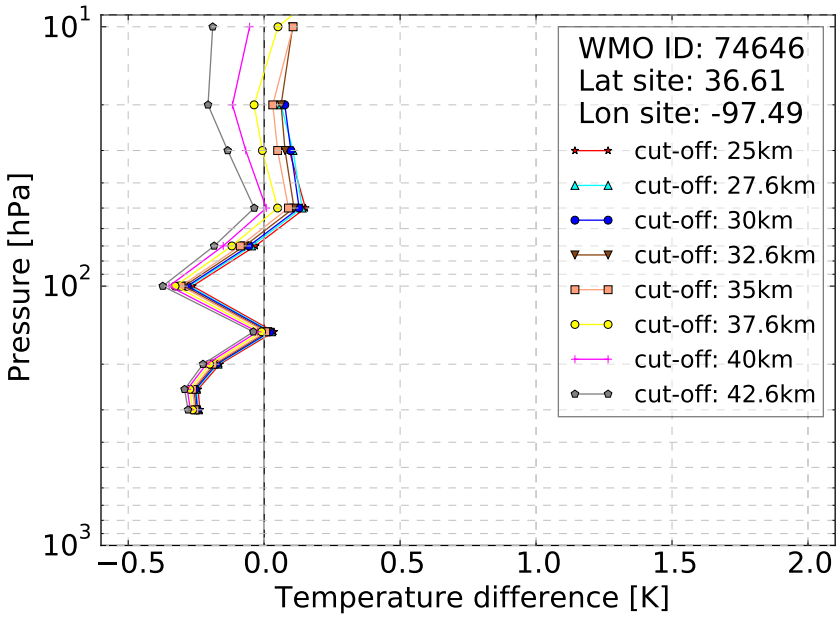


Figure 4.4.: Sensitivity to the upper cut-off impact height at the GRUAN station Southern Great Plains.

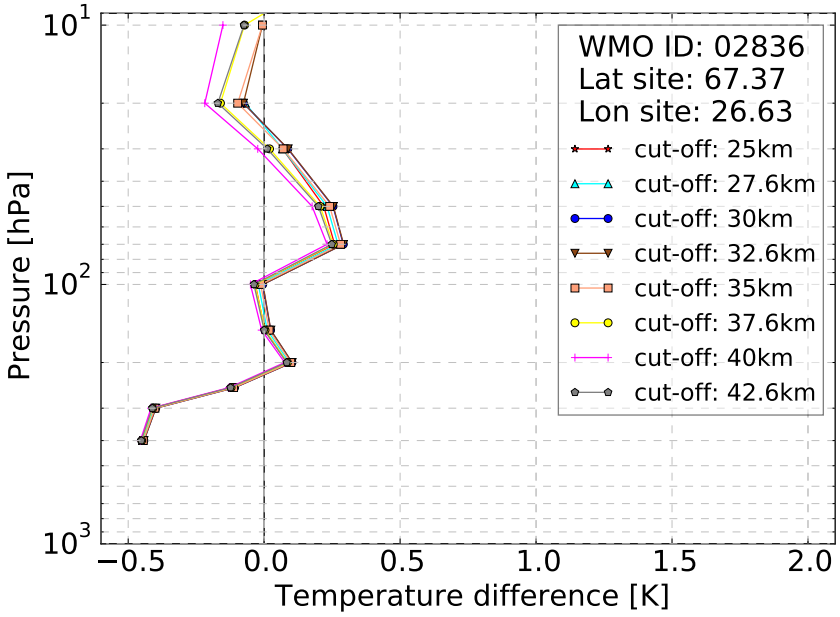


Figure 4.5.: Sensitivity to the upper cut-off impact height at the GRUAN station in Sodankylä.

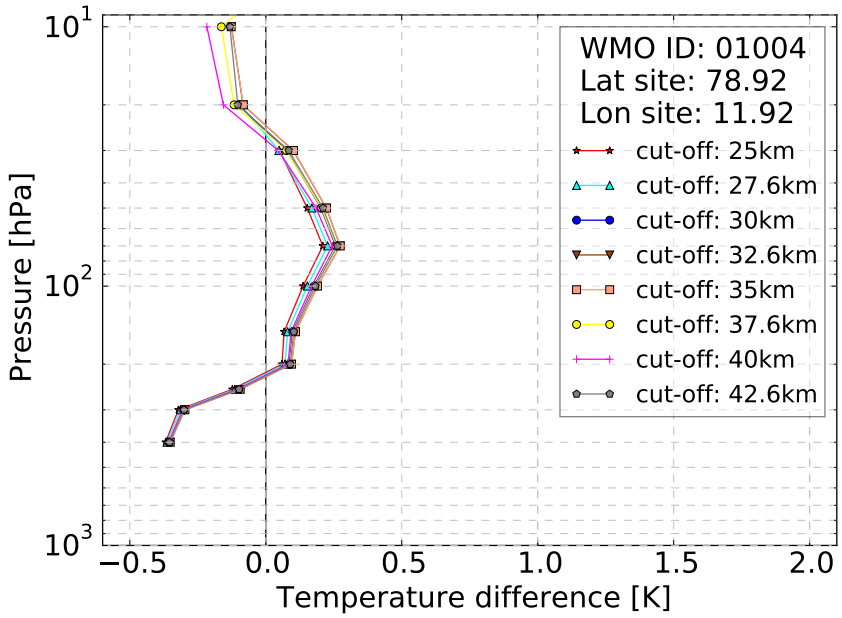


Figure 4.6.: Sensitivity to the upper cut-off impact height at the GRUAN station in Ny Ålesund.

4.3. Comparison of GRUAN and RO departure statistics

In this section, the GRUAN departures are compared to RO departures and the differences are analysed. As GRUAN is offering data products accounting for all known systematic biases, and RO measurements are known to have a high accuracy in the upper troposphere and lower stratosphere, the expectation is that the data are mostly either consistent or in agreement. Using temporal and spatial collocations (criteria of 3 hours and 300 km), Ladstädter et al. (2015) found a warm bias in GRUAN data product for the Vaisala RS92 at high altitudes, with good agreement otherwise. Using the method applied here, a comparison of RO BAs propagated into Tdry space and GRUAN temperatures is possible at altitudes above ∼150 hPa. At lower altitudes, Tdry is not a valid estimate of the temperature due to the effects of humidity. As for the RS statistics, the highest pressure level at which the comparison can be made is shown with a cyan coloured horizontal line in Figs. 4.7-4.20. At altitudes below this line, the number of RO profiles decreases, and those RO profiles that are included are sampled in especially dry airmasses.

In the context of GRUAN, consistency between two independent measurements is achieved if

$$|m_1 - m_2| < k \cdot \sqrt{u_1^2 + u_2^2} \tag{4.11}$$

is true for k=1 (Immler et al. 2010). k is the coverage factor "which determines an interval about the mean value as a multiple of standard uncertainty" (Immler et al. 2010). If Eq. (4.11) is true for $k = 2$ the data are in statistical agreement. The 1σ uncertainty is plotted as the error bars on the difference (RO O-B minus GRUAN O-B). Thus, if the uncertainty range in Figs. 4.7, 4.9, 4.11, 4.13, 4.15, 4.17 encompasses the vertical zero

difference line, the GRUAN and RO temperatures are consistent at a given pressure level. For every analysed GRUAN station, it is tested if the GRUAN O-Bs and RO O-Bs are consist or in agreement. If they are not in agreement (i.e. if $|m_1 - m_2| > 2 \cdot \sqrt{u_1^2 + u_2^2}$) potential reasons are discussed.

In the following subsections, the O-B statistics of GRUAN and RO are discussed for those GRUAN sites that have a sufficiently large sample size of profiles to calculate valid statistics. For every station, two plots are shown, as explained below:

1. Figures 4.7, 4.9, 4.11, 4.13, 4.15, 4.17 show the GRUAN departures in pink, the RO Tdry departures in blue and the difference between the RO O-B and GRUAN O-B in green. The dashed lines give the RO (blue) and GRUAN (pink) sample size. The horizontal cyan line indicates the highest SPL (lowest altitude) where at least 95% of the RO profiles are included. The RO uncertainty bars include the sampling error plus the structural uncertainty, where sampling uncertainty is calculated as $SE_{RO} = \sigma_{RO}/\sqrt{(n-1)}$ and the structural uncertainty as described in Section 4.2, i.e. as half of the range of RO departures for cut-off impact heights between 30 km and 40 km. Thus, the RO uncertainty is calculated as $uncertainty_{RO} = SE_{RO} + range/2$. The GRUAN uncertainty bars are calculated as explained in Section 4.1.3. The uncertainty bars on RO O-B minus GRUAN O-B are $SE_{difference} = \sqrt{uncertainty_{RO}^2 + \sigma_{OB_{GRUAN}}^2}$.
2. Figures 4.8, 4.10, 4.12, 4.14, 4.16, 4.18 show the uncertainty on the differences between the RO and GRUAN departures for $k = 1$ (consistency, red) and $k = 2$ (agreement, red dashed). If the absolute value of the RO O-B minus GRUAN O-B (green) is lower than the threshold for consistency (agreement), the GRUAN departures are consistent (in agreement) with the RO departures.

4.3.1. Lindenberg, Germany

The GRUAN station in Lindenberg, Germany, is operated by *Deutscher Wetterdienst*, the German weather service and is situated at the GRUAN Lead-Centre. RSs are launched four times daily and the Vaisala RS product is submitted to the Global Telecommunication System (GTS).

In Fig. 4.7, the RO and GRUAN departures are consistent at most levels for high and low SEA (Fig. 4.7(a),(b)) from 150 hPa to 20 hPa. At the highest altitude (lowest pressure level) a negative difference between RO O-B and GRUAN O-B exists, implying that the GRUAN temperature is nearly 0.5 K warmer at this level. This agrees with the results from Ladstädter et al. (2015), who found that GRUAN temperatures are slightly warmer than RO Tdry retrievals from the Wegener Center for Climate and Global Change. For dusk/dawn (Fig. 4.7(c)) and night (Fig. 4.7(d)) the RO O-B minus GRUAN O-B profile is mostly slightly positive, with consistency at some levels. As a comparison can only be performed if a sufficient number of RO profiles sampled in dry conditions is available, all data below the cyan coloured line are not discussed here.

For high SEA, consistency or agreement is found at all but the very highest altitude (see Fig. 4.8). For low SEA all levels show consistency or agreement. For dusk/dawn and night SEAs the GRUAN and RO departures are inconsistent at the 70 hPa or 70 hPa and 50 hPa level respectively. The reason for the inconsistencies is currently not known.

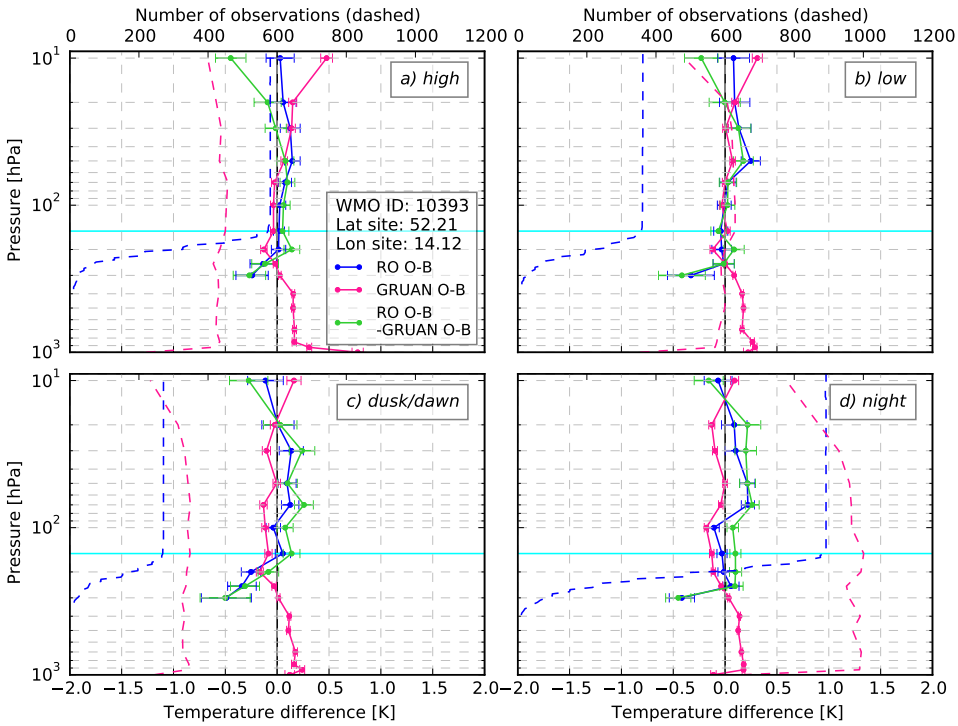


Figure 4.7.: Mean RO Tdry departures (blue), mean GRUAN temperature departures (pink) and RO O-B minus GRUAN O-B (green) at the GRUAN site in Germany for different SEA ranges (a)-(d). The horizontal cyan line indicates the highest SPL (lowest altitude) where at least 95% of the RO profiles are included. The RO error bars represent the structural uncertainty in addition to the sampling uncertainty, i.e. $uncertainty_{RO} = SE_{RO} + range/2$. The dashed lines show the GRUAN (pink) and RO (blue) sample sizes.

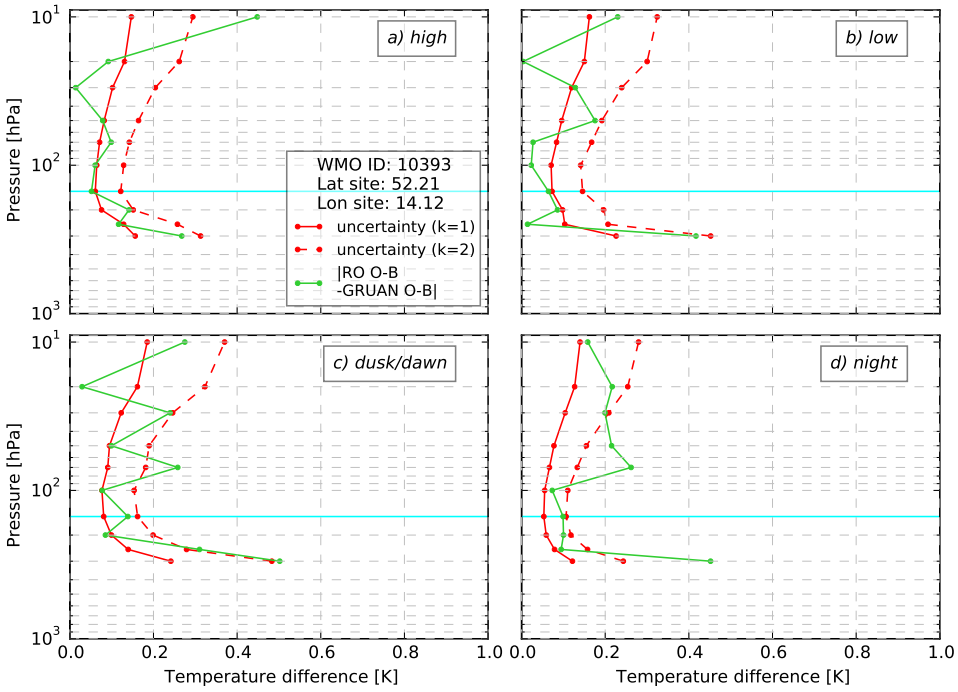


Figure 4.8.: Uncertainties of RO O-B minus GRUAN O-B for $k = 1$ (red) and $k = 2$ (red dashed) and the absolute value of RO O-B minus GRUAN O-B (green). Statistics are presented for the GRUAN site in Germany for different SEA ranges (a)-(d). The horizontal cyan line indicates the highest SPL (lowest altitude) where at least 95% of the RO profiles are included.

4.3.2. Cabauw, Netherlands

The distributed measurement site Cabauw/De Bilt in the Netherlands is operated by the Dutch Weather Service *Koninklijk Nederlands Meteorologisch Instituut* (KNMI). During 2016 Cabauw was the 8^{th} upper-air site that received formal certification for its GRUAN RS operations.

494 GRUAN profiles are available for the upper-air site in De Bilt during 2014 and 2015, with most launches performed during night-time. For high and low SEAs the GRUAN and RO departures presented in Fig. 4.9a),b) are consistent for most pressure levels between 150 hPa and 20 hPa, see Fig. 4.10a),b). At 10 hPa they are not consistent and the difference between the departures is negative. As the GRUAN sample size for high and low SEAs is small, the results might not be robust.

The RO and GRUAN departures are neither consistent nor in agreement at most pressure levels for night-time launches (Fig. 4.10d)). Agreement is found at 150 hPa and at 30 hPa only. It is not clear what causes the lack of agreement at night-time. However, at night-time the GRUAN sample size decrease steadily at pressures below 50 hPa and thus a sampling bias caused by the dependence of the burst altitude on atmospheric parameters such as humidity and temperature, might be present. Thus, the GRUAN profiles at high altitudes might preferentially be sampled in specific conditions. As the RO profiles are sampled in all atmospheric conditions, the double differencing method might not offer a valid estimation of the differences between the RO and GRUAN departures if there is a GRUAN sampling bias at high altitudes.

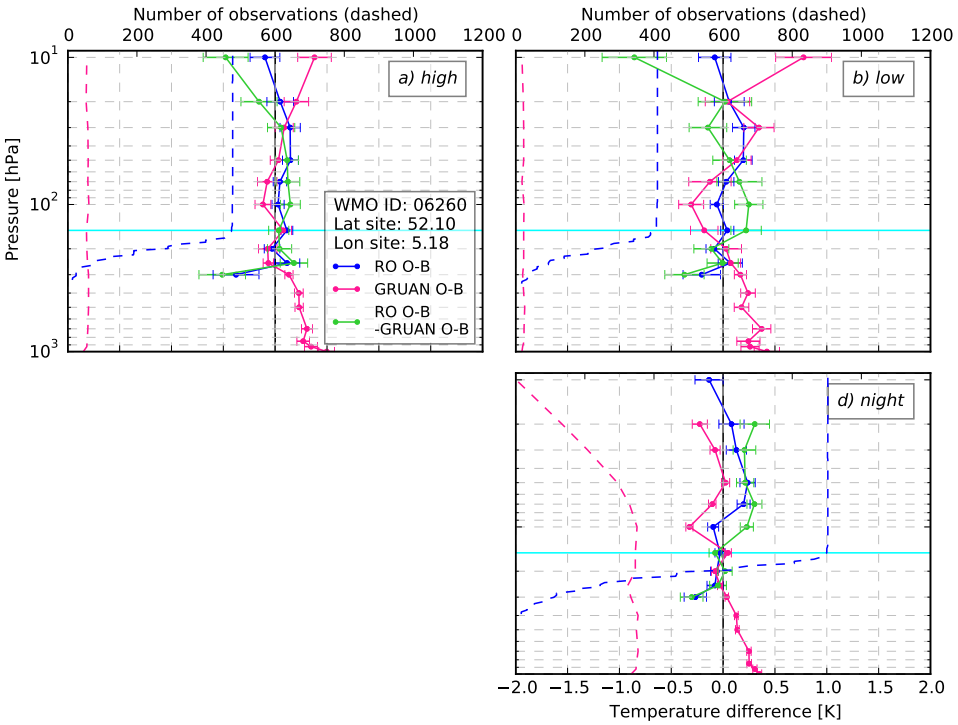


Figure 4.9.: As in Fig. 4.7, but for the GRUAN site Cabauw.

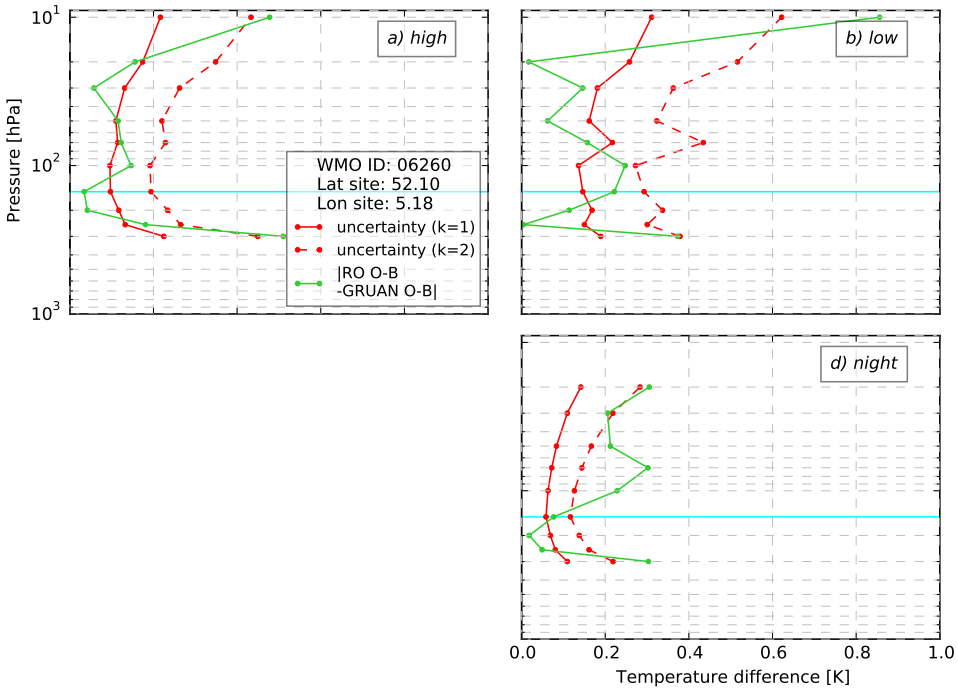


Figure 4.10.: As in Fig. 4.8, but for the GRUAN site Cabauw.

4.3.3. Barrow, Alaska, United States

The GRUAN site in Barrow, Alaska, is operated by the US *Department of Energy* as part of the *Atmospheric Radiation Measurement* program. More than 100 GRUAN profiles are available per SEA range, although most of them do not reach 10 hPa. For high and low SEA ranges (see Fig. 4.11a),b)), the RO and GRUAN departures are consistent at all pressure levels between 150 hPa and 30 hPa. At the 20 hPa level, the GRUAN sample size is very small and thus the results at this level are unreliable.

At dusk/dawn, the GRUAN and RO departures are consistent at all but the 100 hPa level, where they are statistically significantly different (see Fig. 4.11c) and Fig. 4.12c)). At night-time, the GRUAN and RO departures are in agreement at some levels, but show statistically significantly differences at others (see Fig. 4.12(d)).

At the majority of pressure levels, RO and GRUAN departures from model background are more similar to each other than the magnitude of their value, i.e. RO and RS are closer to each other than either is to the model background. This indicates an error in the model. Similar behaviour is also be found at some other stations, but to a varying extent.

4.3.4. Southern Great Plains, Kansas, United States

The GRUAN site Southern Great Plains in Lamont, Kansas is a super-site operated by the US *Department of Energy* as part of the *Atmospheric Radiation Measurement* program.

Compared to the other sites, larger differences between the RO and GRUAN O-Bs are found at the Southern Great Plains site. The pressure range over which sufficient RO and

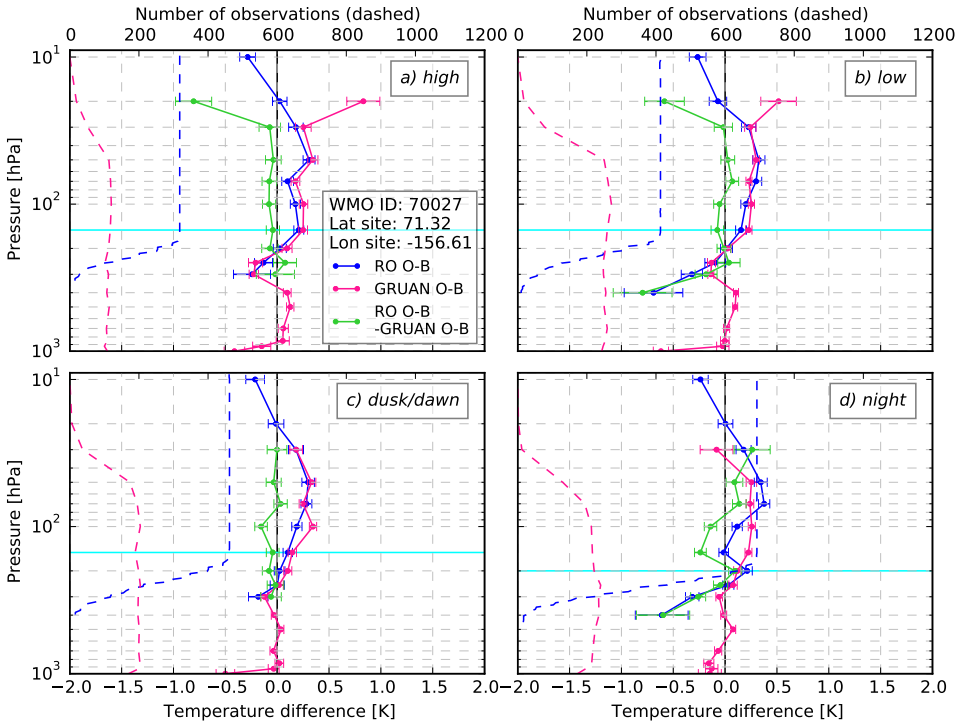


Figure 4.11.: As in Fig. 4.7, but for the GRUAN site Barrow.

GRUAN profiles are available is comparably small, as the number of GRUAN profiles decreases at pressures below approximately 50 hPa and a sufficient RO sample size is only available starting from 100 hPa.

For high, low and dusk/dawn launches (Fig. 4.14a)-c)) agreement is found at some levels, while the RO and GRUAN departures are statistically significantly different at other levels. For night-time launches (Fig. 4.14d)), statistically significant differences are found at all levels. The reason for the lack of consistency/agreement between the GRUAN and RO departures at this specific site is not clear. As the RS launched at this site do commonly not reach the 10 hPa level, the choice of a lower cut-off impact height would be conceivable. However, the RO O-B profile for a cut-off impact height of 25 km at the Southern Great Plains site is similar to the profile calculated with a cut-off at 35km as can be seen in Fig. 4.4. Thus, lowering the cut-off impact height does not result in major changes.

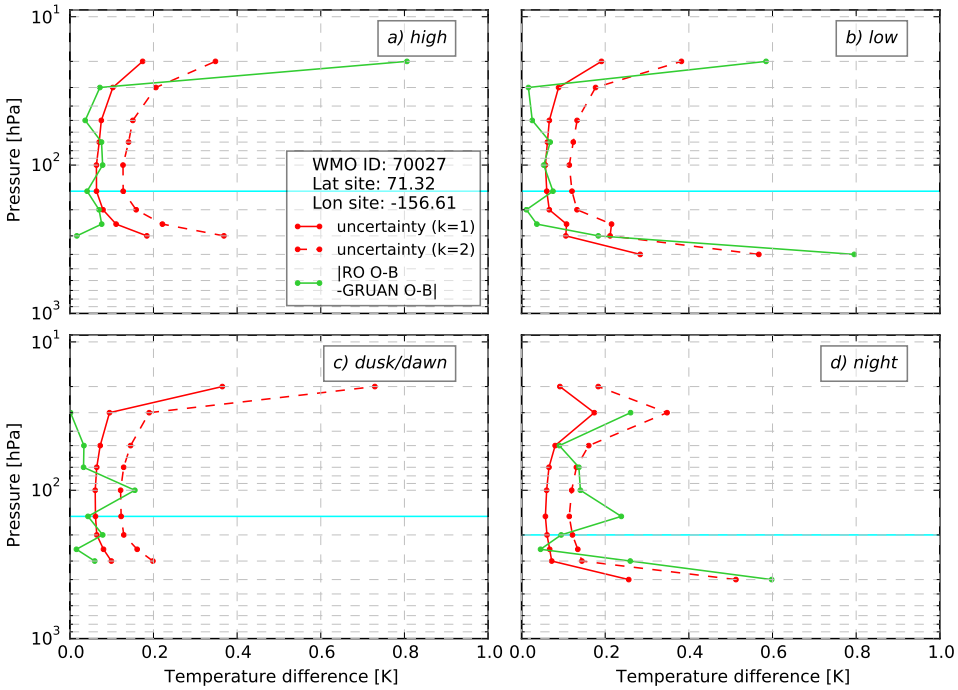


Figure 4.12.: As in Fig. 4.8, but for the GRUAN site Barrow.

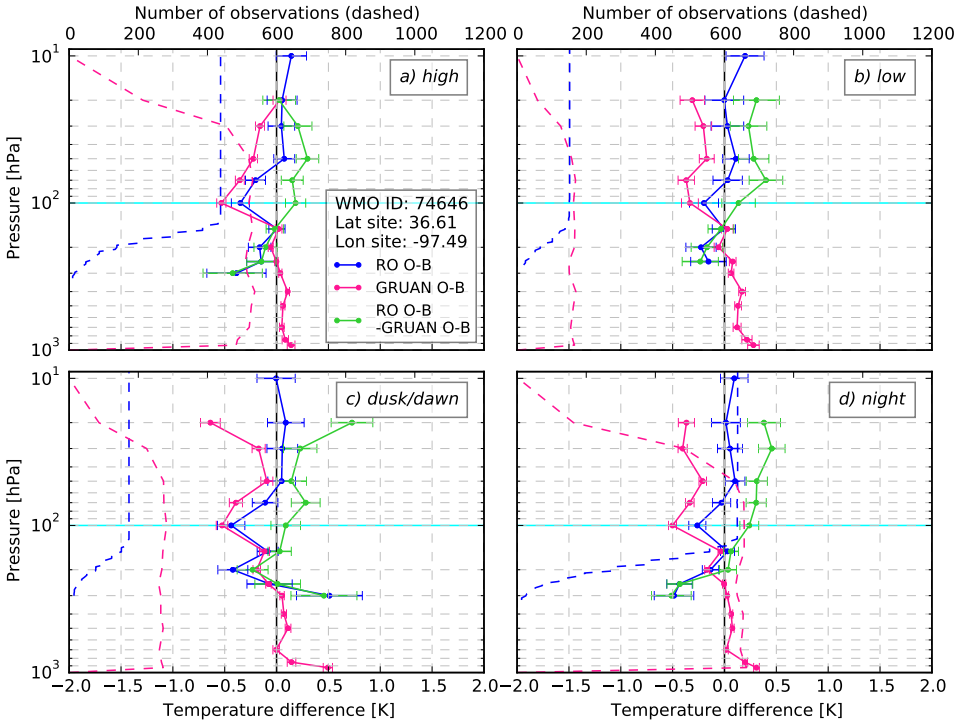


Figure 4.13.: As in Fig. 4.7, but for the GRUAN Southern Great Plains site.

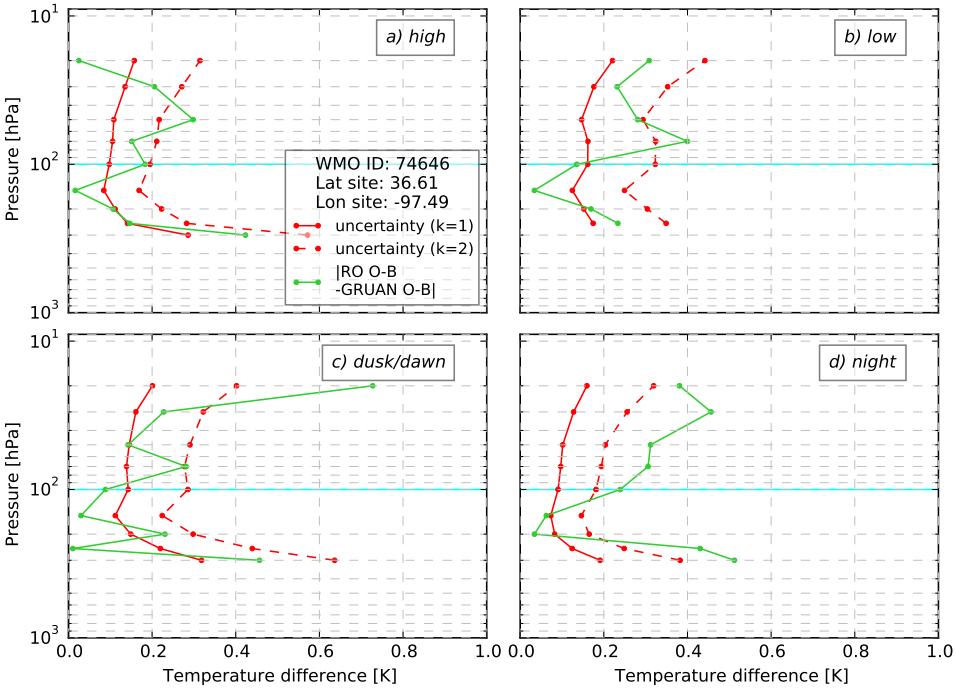


Figure 4.14.: As in Fig. 4.8, but for the GRUAN Southern Great Plains site.

4.3.5. Sodankylä, Finland

The GRUAN site in Sodankylä, Finland, is operated by the *Finnish Meteorological Institute*. Most RS launches are performed with an auto launcher, but some manual launches are also included in the GRUAN data set.

For all SEA ranges, consistency or agreement is found at all but the 10 hPa level, where the RO O-B minus the GRUAN O-B is consistently negative for all SEA ranges. As the sample size is diminished at the 10 hPa level, a sampling bias for the GRUAN sondes, as described in Section 4.3.2, might cause the statistically significant differences between the RO and GRUAN departures.

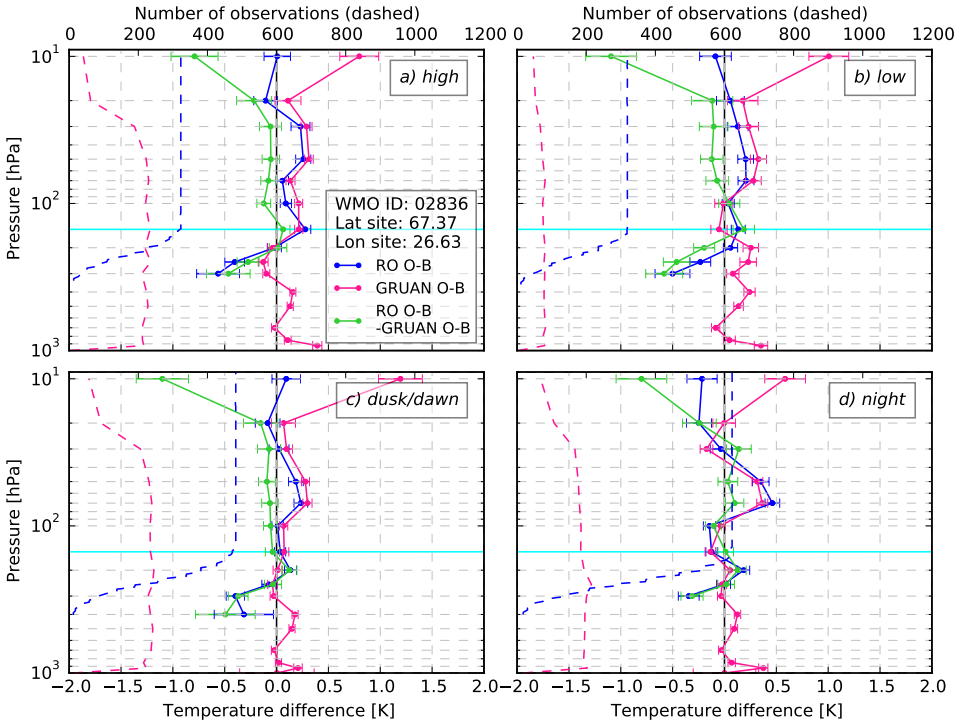


Figure 4.15.: As in Fig. 4.7, but for the GRUAN site at Sodankylä.

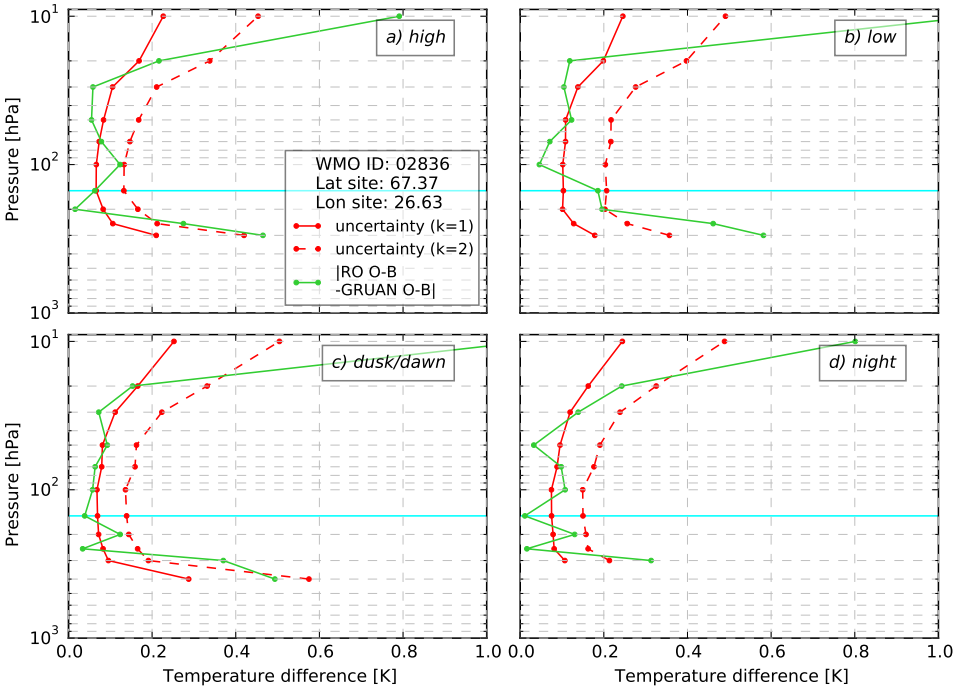


Figure 4.16.: As in Fig. 4.8, but for the GRUAN site at Sodankylä.

4.3.6. Ny Ålesund, Norway

The upper-air site in Ny Ålesund, Svalbard, Norway, at a latitude of 78.92° is the northernmost station within GRUAN. The station is operated by *Alfred-Wegener-Institut* based in Potsdam, Germany.

For all SEAs and at most levels with pressure larger than 30 hPa, the GRUAN and RO departures are consistent or in agreement, as can be seen in Fig. 4.18. But, at the top level(s) statistically significant differences are found for all SEA ranges. Here the RO O-B minus GRUAN O-B is without exception negative, which means that the GRUAN temperatures are warmer than the RO temperatures. In contrast to situation at Sodankylä, the sample size stays approximately constant up to 10 hPa and thus the differences cannot be caused by preferential sampling at certain atmospheric conditions. There are at least two other possible explanations for the significant differences between RO and GRUAN departure statistics at the highest level for all SEA ranges, i.e.

- The proximity to the polar jet could violate the central assumption of a constant model bias over the separation distance between the RO and RS location. Given that the model bias could vary between areas affected by the polar jet and the surrounding area, caution needs to be taken when analysing the results. Some RO and GRUAN profiles will be sampled inside and others outside of the polar vortex. A quantitative analysis of the implications is not part of this project, but could be a useful extension of this work.
- Preferential sampling of GRUAN profiles in a specific season. RSs are launched at specific times of the day (typically a subset of 00 UTC, 6 UTC, 12 UTC, and 18 UTC). These fixed times can lead to preferential sampling during one season, especially at high latitudes, i.e. the dusk/dawn launches could all be made within

two months of the year when the standard launch time coincides with dusk or dawn. As the RO departures are calculated from the RO profiles measured in all seasons, a potentially existing seasonal model bias could influence the results. To investigate the seasonal influence, the departure statistics are calculated separately for the summer and winter half year in Section 4.3.7.

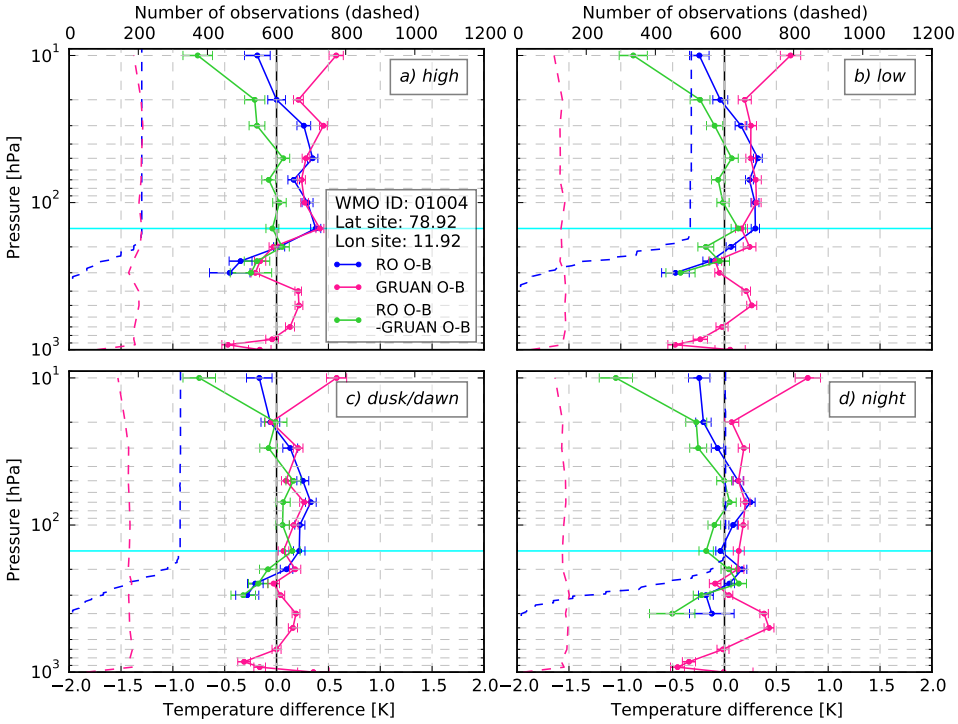


Figure 4.17.: As in Fig. 4.7, but for the GRUAN site at Ny Ålesund.

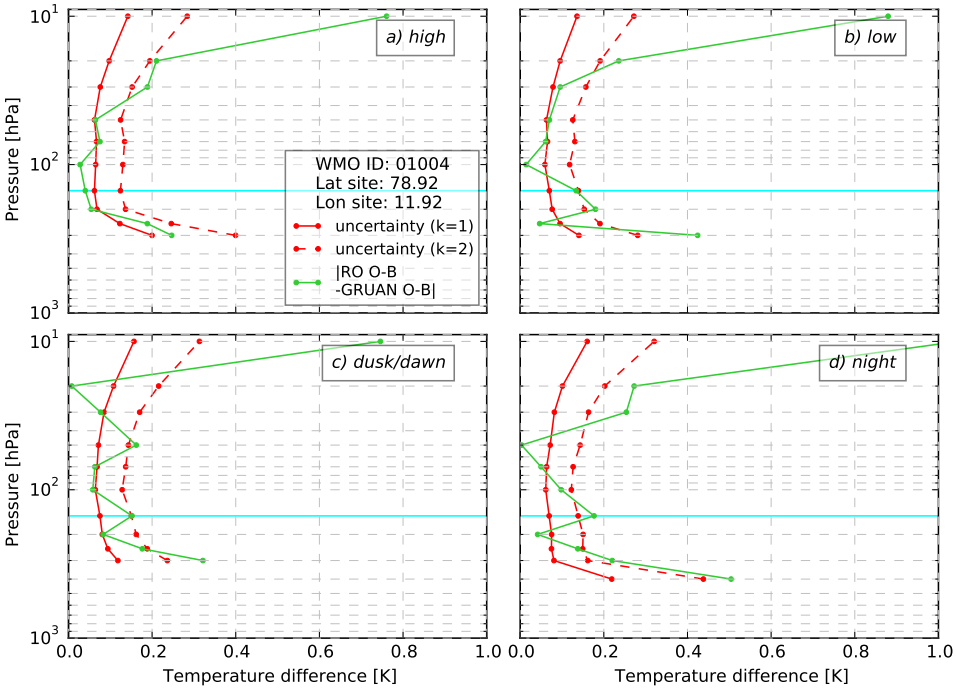


Figure 4.18.: As in Fig. 4.8, but for the GRUAN site at Ny Ålesund.

4.3.7. Analysis of seasonal dependence for the GRUAN sites at Sodankylä and Ny Ålesund

Especially for those upper-air sites located at high latitudes, it is possible that a seasonal bias in the model propagates into a seasonal bias in the departure statistics. This is caused by the fact that the RSs are launched around a subset of standard launch times of 0 UTC, 6 UTC, 12 UTC and 18 UTC. At high latitudes, it is conceivable that all RS measurements within a given SEA range are made during a certain season, while the RO departures include measurements from all/most seasons. This could invalidate the central assumption on which the double differencing technique is based, i.e. the model background bias is constant over the separation distance. This assumption could be invalidated if the model bias has a seasonal dependence.

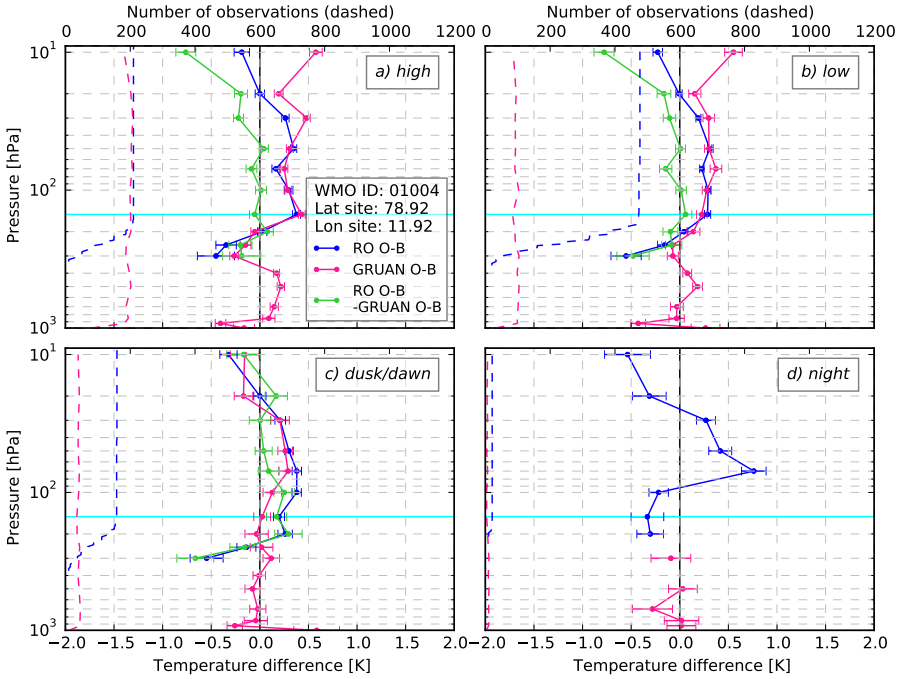
To test if the difference between GRUAN and RO departures at the lowest pressure level (highest altitude), which is found at all SEA ranges, for the sites at Sodankylä and Ny Ålesund is caused by a seasonally varying model bias, the departures are calculated separately for the summer and winter half year, i.e. for the months April-September (summer half-year) and October-March (winter half-year).

Figure 4.19 shows the departure statistics at Ny Ålesund for the summer half-year (a) and the winter half-year (b) separately. For some SEA ranges the number of profiles available is too small to calculate the departure statistics for a half-year. The statistics for high SEA ranges are calculated only for the summer months and valid night-time departures are only available in winter. The very low GRUAN sample size at low SEA during the winter months precludes a robust statistical analysis. For the dusk/dawn SEA range, better agreement is found during the summer months than for the whole year. Here GRUAN and RO departures are similar at most pressure levels. During the winter

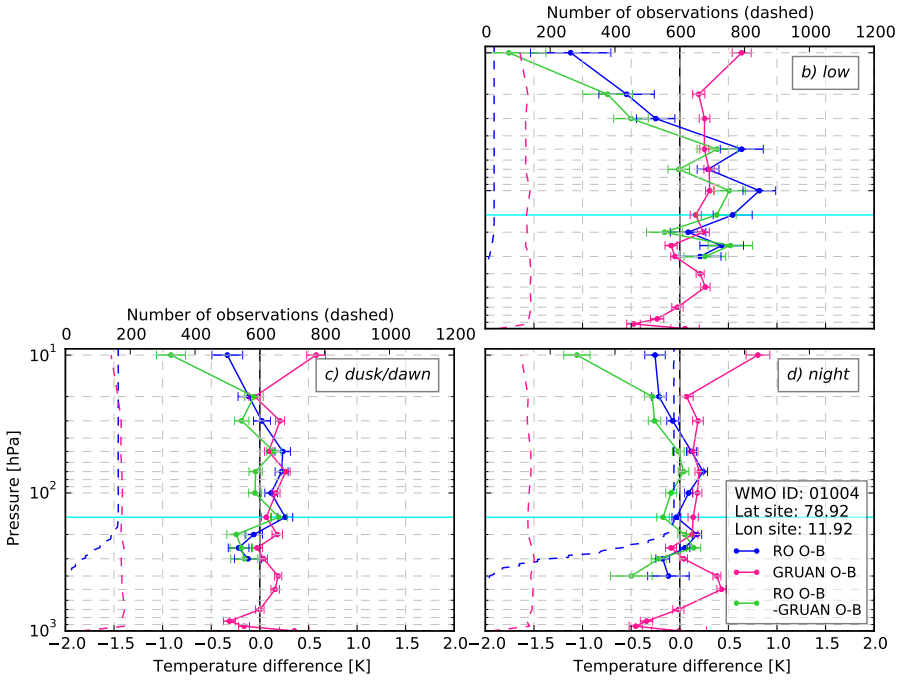
months at dusk/dawn, the RO O-B minus GRUAN O-B is ∼-0.8 K. Similar values are also found for the departures calculated from night-time launches performed in winter.

The departure statistics calculated for summer and winter months separately for Sodankylä are shown in Fig. 4.20. Unfortunately the number of measurements at the highest level is too low for most SEA ranges to analyse the results. It is clear, however, that there are large positive O-Bs for GRUAN in the winter months at the highest levels for all launch times. As discussed in Section 4.3.5, the decreasing GRUAN sample at the highest levels, could lead to a sampling bias which could cause the large difference between the GRUAN and RO O-B statistics.

Figure 4.20 furthermore indicates that a seasonal model bias between 70 hPa and 50 hPa at the Sodankylä site might exist. During the winter months, both the GRUAN and the RO departures are positive at 70-50 hPa and consistent/in agreement during night-time. This means that GRUAN and RO temperatures agree with each other, but are both different to the model. The model bias is also apparent for dusk/dawn in the winter months, but to a lesser degree.

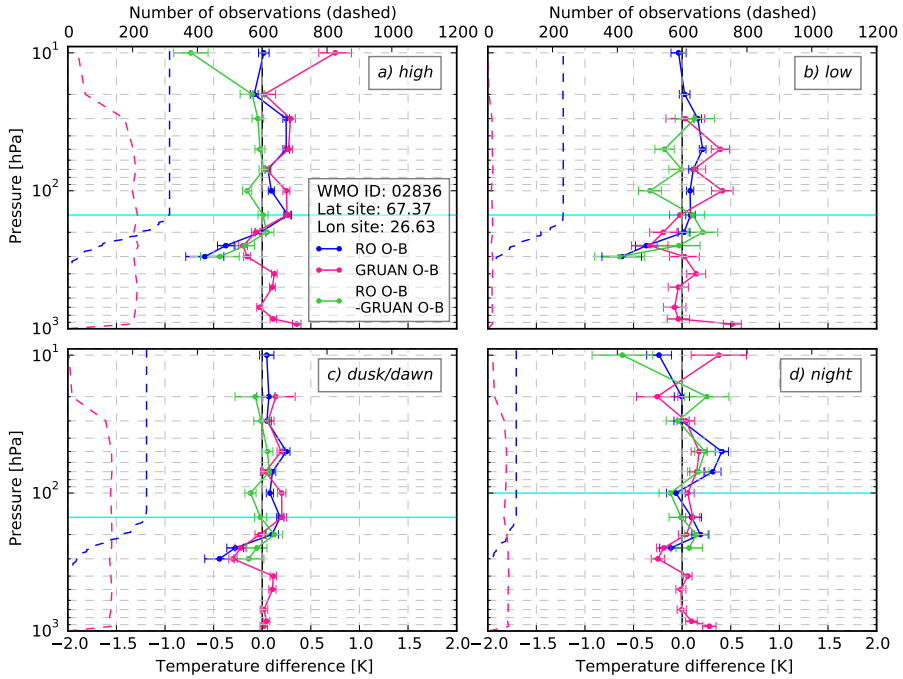


(a) Summer half-year

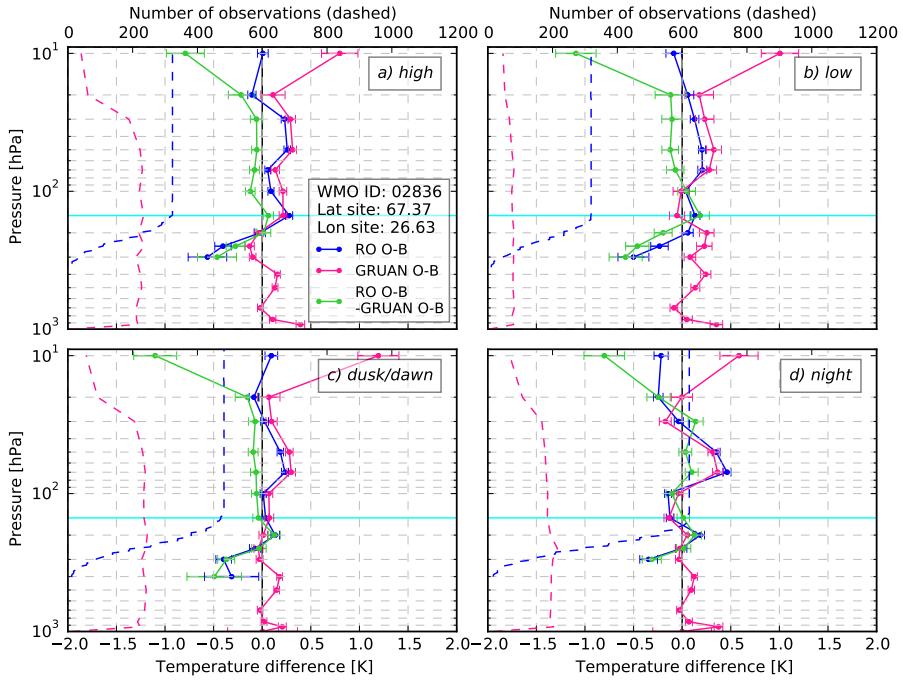


(b) Winter half-year

Figure 4.19.: Mean RO Tdry departures (blue), mean GRUAN temperature departures (pink) and RO O-B minus GRUAN O-B (green) at the GRUAN site Ny Ålesund for different SEA ranges (a)-(d). The horizontal cyan line indicates the highest SPL (lowest altitude) where at least 95% of the RO profiles are included. The RO error bars represent the standard errors (sampling uncertainty) calculated as $SE = \sigma/\sqrt{(n-1)}$. (a) summer half-year, (b) winter half-year.



(a) Summer half-year



(b) Winter half-year

Figure 4.20.: As in Fig. 4.19, but for the GRUAN site at Sodankylä.

4.4. Discussion of GRUAN and radio occultation comparison

As both RO and GRUAN offer high-quality data sets, similarity between the measurements is expected when taking the uncertainties in account. The GRUAN uncertainty estimate is propagated into the mean GRUAN departure and is shown in the associated uncertainty bars. However, as discussed in Section 4.1, the estimate of the uncertainty in the temperatures which are interpolated to the standard pressure levels could be improved if the covariances of the temperature profile would be known. Furthermore, the interpolation itself should add uncertainty, as the variable to be interpolated could peak between the two measured levels. While this effect is expected to be small in the analysis here, as GRUAN profiles have a very high vertical resolution, it is a general problem when estimating the uncertainty on an interpolated value.

The RO uncertainty includes sampling and structural uncertainty. Consistency and agreement between the RO O-Bs and the GRUAN O-Bs is discussed based on Immler et al. (2010) for six GRUAN upper-sir sites, i.e. Lindenberg, Cabauw, Southern Great Plains, Barrow, Ny Ålesund, and Sodankylä. For most GRUAN sites consistency or agreement is found at high, low, and dusk/dawn SEAs at most levels, excluding the 10 hPa level. At the 10 hPa level, the RO O-B minus the GRUAN O-B are negative for high SEA at all sites. This could be caused by a warm temperature bias in the GRUAN data product for the Vaisala RS92 version 2 which is analysed here. A similar results was observed by Ladstädter et al. (2015).

At Ny Ålesund and Sodankylä negative double differences are also found for dusk and night SEA at the highest level. Further investigation shows that this warm bias at night-time/dusk mainly occurs in the winter months. Possible reasons for this difference between the GRUAN and RO departure are: (i) preferential sampling of the GRUAN data in certain seasons, (ii) proximity to the polar jet; the central assumption of the double differencing technique could be invalidated if the model bias in proximity to the jet is not constant within 500 km around the GRUAN site, and (iii) a sampling bias caused by the dependence of the burst altitude on atmospheric conditions could explain differences in Sodankylä.

During night-time statistically significant positive double differences (RO O-B minus GRUAN O-B) are found at some stations in the lower stratosphere. The largest differences between the GRUAN and the RO departures are found at the Southern Great Plains site, where the departures are statistically significant different for all levels at night-time. It is currently not clear what is causing these differences and further investigation is required.

5. Enhancing the value of temperature measurements made at the lower South Island of New Zealand

After the improvement of the global RS[10] record and the comparison of measurements from a subset of GRUAN sites with RO measurements, the thesis will now describe a method to improve the upper-air record of temperature at one specific GRUAN site located in Lauder, New Zealand.

Ground-based and *in situ* observations of ECVs are especially sparse in the Southern Hemisphere, due to the remote locations and the vast areas covered by ocean. At Lauder, located centrally on the lower South Island of New Zealand, a state-of-the-art atmospheric measurement facility is operating since 1961. Lauder operates a wide range of instruments; microwave radiometers, lidars, UV/Vis spectrometers, automatic weather stations, flask sampling stations, and a balloon-borne observation program. IN 2015 the research facility in Lauder received GRUAN certification and GRUAN processed RS data from Lauder are available. To increase the sample size of GRUAN-processed RS profiles from New Zealand, an agreement was found between The New Zealand National Institute of Water and Atmospheric Research (NIWA) and the New Zealand MetService to operate the upper-air facilities in Lauder and Invercargill, 180 km southwest of Lauder, as a decentralized GRUAN site.

This chapter described a methodology to combine measurements from Lauder and Invercargill in a temperature SASBE. While this study is not using GRUAN processed profiles, as the time frame processed here was prior to the availability of GRUAN profiles, the validity of the method is in no way affected by this. Based on the scarcity of measurements in the Southern Hemisphere, the distributed GRUAN site at Lauder and Invercargill is well-suited to demonstrate a concept to combine data from distributed sites.

The term SASBE was introduced by Tobin et al. (2006) who combine measurements made with RSs and a microwave radiometers. Generally, SASBEs combine measurements from multiple instruments to create a best possible vertically resolved high temporal resolution time series of the parameter of interest above one site. The SASBE should aim to contain all suitable knowledge about the state of the target variable and include an estimate of the uncertainty on every value. As a results SASBEs are well suited for the validation of space-based instruments.

The following section is a slightly adapted reprint of **Tradowsky, J.S., G.E. Bodeker, R.R. Querel, P.J.H. Builtjes, and J. Fischer (2018). "Combining Data from the Distributed GRUAN Site Lauder-Invercargill, New Zealand, to Provide a Site Atmospheric State Best Estimate of Temperature"**, submitted to *Earth System Science Data Discussions*, (Tradowsky et al. 2018, submitted).

[10] All abbreviations can be found in the Glossary, Appendix B

5.1. Paper II: Combining data from the distributed GRUAN site Lauder-Invercargill, New Zealand, to provide a site atmospheric state best estimate of temperature

5.1.1. Abstract

A SASBE of the temperature profile above the GRUAN site at Lauder, New Zealand, has been developed. Data from multiple sources are combined within the SASBE to generate a high temporal resolution data set that includes an estimate of the uncertainty on every value.

The SASBE has been developed to enhance the value of measurements made at the distributed GRUAN site at Lauder and Invercargill (about 180 km apart), and to demonstrate a methodology which can be adapted to other distributed sites. Within GRUAN, a distributed site consists of a cluster of instruments at different locations.

The temperature SASBE combines measurements from RSs and automatic weather stations at Lauder and Invercargill, and ERA5 reanalysis, which is used to calculate a diurnal temperature cycle to which the SASBE converges in the absence of any measurements.

The SASBE provides hourly temperature profiles at 16 pressure levels between the surface and 10 hPa for the years 1997 to 2012. Every temperature value has an associated uncertainty which is calculated by propagating the measurement uncertainties, the ERA5 ensemble SDs, and the ERA5 representativeness uncertainty through the retrieval chain. The SASBE is available free of charge.

The study demonstrates a method to combine data collected at distributed sites. The resulting best-estimate temperature data product for Lauder is expected to be valuable for satellite and model validation as measurements of atmospheric ECVs are sparse in the Southern Hemisphere. The SASBE could, for example, be used to constrain a radiative transfer model to provide TOA radiances with traceable uncertainty estimates.

5.1.2. Introduction

Measurements of the upper-air are essential for atmospheric research and weather forecasts. While high vertical resolution temperature profiles can be retrieved from space-based instruments, these retrievals require validation, which is typically done by comparison with ground-based or *in situ* measurements. Ground-based techniques to observe upper-air temperatures include lidar and microwave radiometer, while *in situ* measurements are typically made using balloon-borne RSs. While RSs provide vertically highly-resolved profiles of temperature, pressure and humidity, they are only used at about 800 upper-air sites worldwide (Ingleby 2017), which typically launch two sondes per day. Given the limited spatio-temporal sampling of the RS measurements, their use for satellite validation can be challenging as the number of collocations within a given time interval and distance is small (see e.g. Calbet 2016). However, if measurements from different instruments, or from collocated sites, can be combined in a best-estimate data product, the value of those measurements can be enhanced.

Site Atmospheric State Best Estimates (SASBEs; Tobin et al. 2006) combine measurements from multiple instruments to create a vertically-resolved, high temporal resolution time series of an atmospheric ECV (ECV; GCOS-200 2016; Bojinski et al. 2014) above a site. SASBEs aim to encompass all suitable knowledge of the state of the target variable at the specific site and include an estimate of the uncertainty on each value therewith satisfying the requirements of GCOS-170 (2013). Tobin et al. (2006) developed a SASBE for the Atmospheric Radiation Measurement (ARM) site Southern Great Plains[11], which was used to validate retrievals of temperature and water vapour from the Atmospheric Infrared Sounder (AIRS; Aumann et al. 2003). Maillard Barras et al. (2015) present a methodology to combine *in situ* ozonesonde and space-based microwave radiometer measurements from SOMORA into an ozone SASBE. The authors found improved agreement between the ozone SASBE and the Microwave Limb Sounder (AURA/MLS) in comparison to the agreement with the operational SOMORA retrieval which does not include ozonesonde profiles.

The combination of measurements from different sensors with ancillary information in a SASBE, enhances the information content of available observations. Furthermore, it can simplify the use of observations if the SASBE is provided in a well-documented open-access database.

As upper-air measurements in the Southern Hemisphere are especially sparse, it is essential to exploit all available observations. A well-equipped atmospheric measurements site, operated by NIWA, is based at Lauder on the South Island of New Zealand. This site is part of GRUAN (GCOS-112 2007), which was established to fill the need for reference-quality measurements of upper-air ECVs. Bodeker et al. (2016) document the development of GRUAN over the first ten years after its establishment; a governance structure is in place, 24 sites have joined the network as of 2016, and a first reference-quality data product, accounting for all known biases, and including a traceable uncertainty estimate, is publicly available (see Dirksen et al. 2014).

Within GRUAN, a cluster of instruments at different locations can be operated as a distributed, i.e. decentralized site. If the measurements made at one location are used to estimate measurements made elsewhere, the uncertainty estimates on the estimated measurements must include the original measurement uncertainty plus the additional uncertainty induced by the transfer algorithm (GCOS-170 2013). A method to enhance the value of a distributed GRUAN site by combining measurements from distributed instruments is presented here. Measurements made at Invercargill (46.413° S, 168.3173° E, 1 m) are used as predictors of the temperatures above Lauder (45.0383° S, 169.6843° E, 370 m) and, as required, the uncertainty added by the transfer algorithm is calculated and included in the best-estimate of the temperature. GRUAN is also setting new standards regarding the careful documentation of data sets, i.e. every GRUAN data product has an accompanying peer-reviewed publication describing the bias correction, retrieval steps and uncertainty components. Following GRUAN's tenets, this paper documents a method to create a temperature SASBE for Lauder, New Zealand. This SASBE will be published receive a digital object identifier.

The temperature SASBE presented here combines measurements performed at the distributed GRUAN site Lauder and Invercargill both based in the lower South Island of

[11] the site has since joined GRUAN

New Zealand. The SASBE comprises hourly temperature profiles above Lauder and associated uncertainty estimates at 16 vertical levels, i.e. surface, 925 hPa, 850 hPa, 700 hPa, 500 hPa, 400 hPa, 300 hPa, 250 hPa, 200 hPa, 150 hPa, 100 hPa, 70 hPa, 50 hPa, 30 hPa, 20 hPa, and 10 hPa. The value of the upper-air observations is enhanced by (i) considerable improvement of the temporal resolution, (ii) the availability of an uncertainty estimate with every temperature value, and (iii) better documentation of the data set.

This paper presents the observational data sets used herein, followed by a detailed description of the SASBE methodology used. The SASBE is presented in Section 5.1.5 and the paper concludes with a discussion outlining possible applications for the temperature SASBE at Lauder.

5.1.3. Observational data and reanalysis

The temperature SASBE combines RS and automatic weather station data collected at Lauder and Invercargill. These measurements are supplemented with ERA5 reanalysis (Hersbach and Dee 2016) which are used to provide an estimate of the diurnal temperature cycle above the sites. The atmospheric research facility at Lauder is operated by NIWA and the operational measurement site at Invercargill is run by the New Zealand MetService. Lauder is a certified GRUAN site which also includes RS profiles from Invercargill in its data stream, based on the concept of a distributed site. The RS data used in the SASBE described here were not GRUAN processed as the time frame used here was prior to the availability of GRUAN profiles for Lauder. Nonetheless, the purpose if this paper is to demonstrate the construction of a SASBE for a GRUAN site for implementation across GRUAN. Lack of GRUAN data for this demonstration in no way compromises the validity of the method. Within this study, basic estimates of the 1σ uncertainty for RS temperatures are used. If available, traceable uncertainty estimates should be used instead, such as those provided with the GRUAN RS data product.

The ERA5 reanalysis is the fifth generation of atmospheric reanalysis produced by ECMWF. The ERA5 data are used in this study to provide diurnal temperature cycles above Lauder and Invercargill from which anomalies can be calculated. ERA5 reanalysis data for the years 2010 to 2017 are used here as this was the period of data available at the beginning of 2018. Eventually, ERA5 will be available from 1950 onwards and monthly updates with a maximum latency of three months are planned. ERA5 has an hourly temporal resolution, a 31 km horizontal grid worldwide with 137 vertical levels up to 0.01 hPa. It is the first reanalysis which includes uncertainties calculated by running a 10-member ensemble of data assimilations at 62 km resolution with 3-hourly output. The temperature and its uncertainty are interpolated to the location of the Lauder and Invercargill upper-air sites. Additional temporal interpolation is required to obtain uncertainty estimates on the hourly ERA5 temperatures.

To clarify the sources of the data and the various identification numbers for the stations, Table 5.1 summarises the data sources, while Table 5.2 provides the station details.

Dataset	Source	Available at
Lauder RS	Network for the Detection of Atmospheric Composition Change	ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/lauder/
Lauder AWS	New Zealand's National Climate Database	https://cliflo.niwa.co.nz/
Invercargill RS	MetService; New Zealand's National Climate Database	https://cliflo.niwa.co.nz/
Invercargill AWS	New Zealand's National Climate Database	https://cliflo.niwa.co.nz/
ERA5 reanalysis	ECMWF Copernicus Climate Change Service	https://climate.copernicus.eu/products/climate-reanalysis

Table 5.1.: The data sets used. For those data sets available in New Zealand's National Climate Database, Table 5.2 gives the Agent number and Network number which are required to identify the site. The Invercargill RS measurements are obtained from two sources, (i) the New Zealand's Climate Database (low resolution), and (ii) as high resolution profiles in the original Vaisala output format (MetService). The high resolution Vaisala files, which also include a range of metadata, have been processed into netCDF files (Bruno Kinoshita, personal communication) and are used for the years they are available. AWS abbreviates automatic weather station.

Station	WMO ID	Cliflo Identifiers	OSCAR Entry
Lauder	93817		https://oscar.wmo.int/surface/index.html#/search/station/stationReportDetails/12986
Lauder AWS		Agent 5535, Network I59065	https://oscar.wmo.int/surface/index.html#/search/station/stationReportDetails/12986
Invercargill	93844	Agent: 5814, Network: I68433	https://oscar.wmo.int/surface/index.html#/search/station/stationReportDetails/12991
Invercargill AWS	93845	Agent: 12444, Network: I68437	https://oscar.wmo.int/surface/index.html#/search/station/stationReportDetails/12992

Table 5.2.: Station details for Lauder and Invercargill, including the WMO station identifier (ID). The Agent and Network number are required to identify the station in New Zealand's National Climate Database. The Observing Systems Capability Analysis and Review (OSCAR) Tool is WMO's official repository of metadata on surface-based observations.

5.1.4. Methodology

The availability of upper-air measurements is limited, especially in the Southern Hemisphere. Therefore, combining those observations available with ancillary information in a SASBE will enhance the value of the upper-air record. The temperature SASBE presented here is aiming to combine all available knowledge about the temperature profile at the lower South Island of New Zealand in a temperature SASBE for Lauder, located centrally within the lower South Island. One source of knowledge about the temperatures above Lauder is the ERA5 reanalysis which is used to calculate a diurnal temperature cycle. This diurnal cycle builds the foundation for the SASBE presented here, i.e. the SASBE converges towards the ERA5 diurnal cycle in the absence of any other sources of data. However, the upper-air site at Lauder, New Zealand, performs weekly RSs launches. The temperature anomaly of the RS measurement with respect to the ERA5 diurnal cycle is calculated and, around the launch time of the RS, the temperature SASBE diverts from the diurnal cycle towards the observed temperature. At the exact time of the measurement, the SASBE adapts the measured value, and with further time difference to the measurement relax back to the diurnal cycle. The persistence (autocorrelation) of temperature anomalies can be used to determine how fast the SASBE converges towards the diurnal cycle. At the exact time of the RS measurement, the autocorrelation function is 1, i.e. the measurement is perfectly correlated with itself. With further time difference the autocorrelation decreases and the SASBE converges towards the diurnal cycle.

While the upper-air research site in Lauder has a wide range of instruments, the temperature profile is only measured once weekly with RSs. Fortunately, about 180 km southwest of Lauder, the operational upper-air site at Invercargill measures the temperature profile twice daily. While the absolute temperatures above Lauder and Invercargill might differ, the temperature anomalies are well-correlated (see Section 5.1.4). Therefore, the Invercargill temperature anomalies with respect to the ERA5 diurnal cycle may be used to infer temperature anomalies above Lauder. Because these inferred anomalies provide knowledge about the temperature above Lauder they can be integrated into the SASBE to induce excursions from the diurnal cycle. As for the Lauder RS anomalies, their influence should decrease with further time distance from the measurement time.

Within the SASBE, the inferred temperature anomalies should have less weight than the measurements made at Lauder, i.e. they should be punished for originating from Invercargill, and should furthermore not be granted any weight at the time of a Lauder measurement. This is achieved by including the weight ϕ which is again based on the autocorrelation of temperature anomalies at Lauder. At any given time t, ϕ is the value of the autocorrelation function for the time difference to the closest Lauder launch time (see Fig. 5.1).

The concept of the SASBE is schematically shown in Fig. 5.1. The details of the preparation of the individual components and the combination of them in the temperature SASBE is described in the following sections. In the following all calculations are presented for one pressure level only.

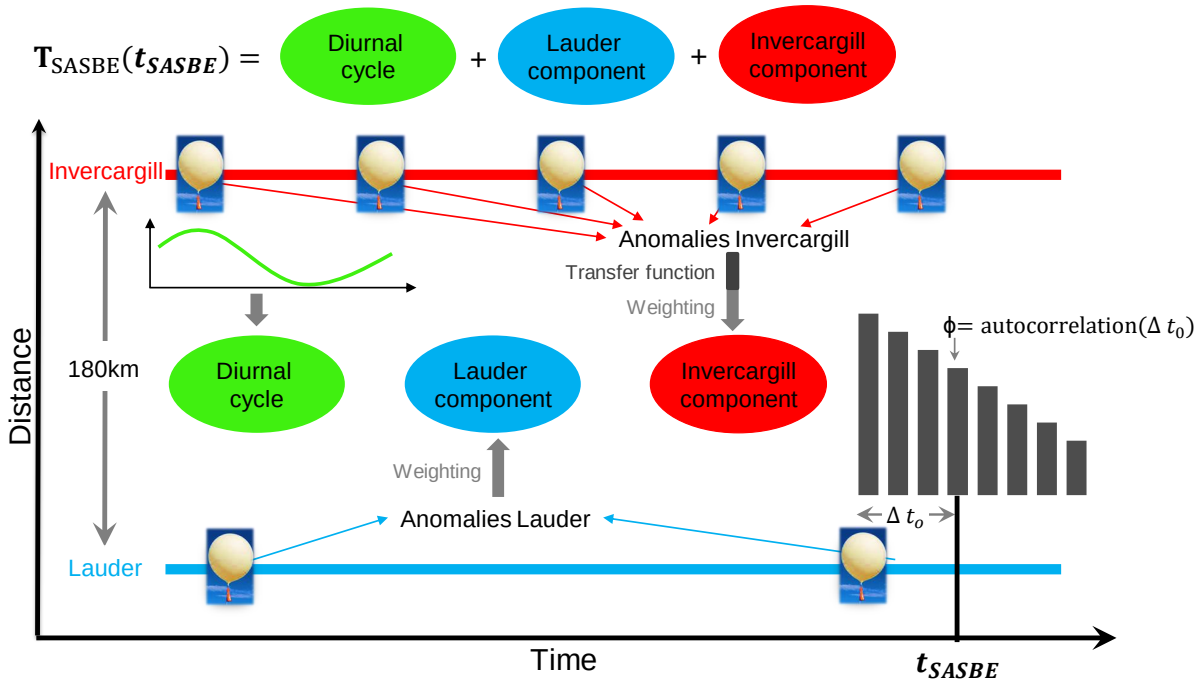


Figure 5.1.: Schematic explanation of the temperature SASBE and its components. The diurnal cycle is determined from ERA5. To calculate the Invercargill component, a transfer function and some weighting is applied to the temperature anomalies at Invercargill. To calculate the Lauder component, the temperature anomalies are weighted. Also shown is the weight ϕ, which is determined by the value of the autocorrelation function for the time difference between t and the closest Lauder RS launch.

Calculating the upper-air diurnal temperature cycle above Lauder and Invercargill

The diurnal temperature cycle above Lauder and Invercargill is calculated from the hourly ERA5 reanalysis by fitting four Fourier pairs for the hour of the day (h), and four Fourier pairs for the day of the year (d). Bilinear interpolation in space is used to retrieve ERA5 reanalysis temperature at the exact location of the Lauder and Invercargill upper-air sites, respectively.

$$T_{Diur} = \zeta_0 + \sum_{i=1}^{4} \zeta_{i\cdot 2-1} \cdot \sin\left(\frac{i \cdot 2\pi h}{24}\right) + \zeta_{i\cdot 2} \cdot \cos\left(\frac{i \cdot 2\pi h}{24}\right) \tag{5.1}$$

where all ζs are expanded in a Fourier series with 4 pairs to account for the annual cycle, i.e.:

$$\zeta_i = \zeta_{i_0} + \sum_{j=1}^{4} \zeta_{i_{j\cdot 2-1}} \cdot \sin\left(\frac{j \cdot 2\pi d}{365.25}\right) + \zeta_{i_{j\cdot 2}} \cdot \cos\left(\frac{j \cdot 2\pi d}{365.25}\right) \tag{5.2}$$

This leads to $9 \cdot 9 = 81$ fit coefficient which are estimated for Lauder and for Invercargill and can be used to calculate a climatological mean diurnal temperature cycle for every day of the year, with a smooth transition from day to day.

The 1σ uncertainty on the regression coefficients is calculated based on the method

described in Bodeker and Kremser (2015), using the 3-hourly 10-member ensemble SDs provided within ERA5, interpolated to an hourly resolution. The fitting uncertainty on the diurnal cycle is calculated as:

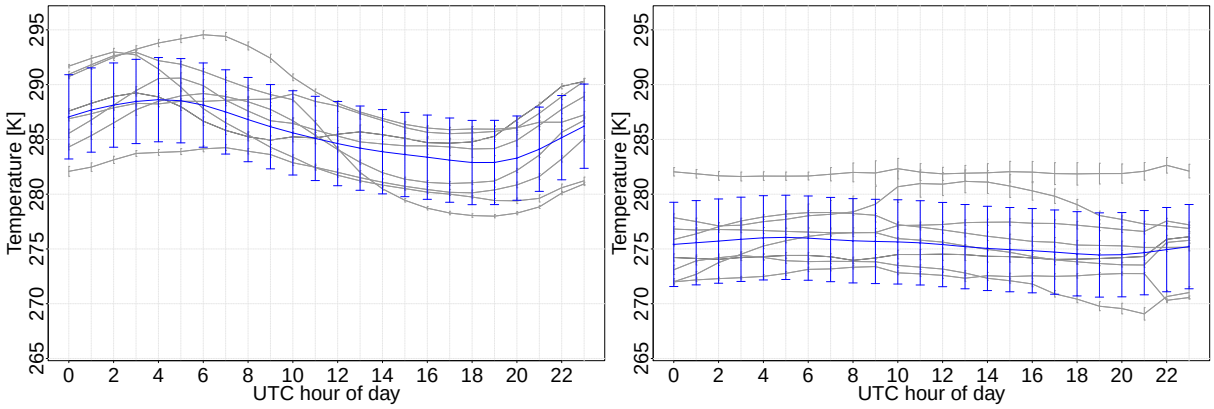
$$\sigma_{fit} = \sqrt{\sum_{i=1}^{81} \sigma_{\zeta_i}^2 \left(\frac{\partial T_{Diur}}{\partial \zeta_i} \right)^2} \tag{5.3}$$

This uncertainty indicates how good the regression model fits the 8 years of hourly reanalysis data. However, this uncertainty does not indicate how representative the ERA5 diurnal cycle is for the temperatures measured above Lauder or Invercargill. Therefore, another uncertainty component, the representativeness uncertainty ($\sigma_{representativeness}$), is included and the uncertainty on the diurnal cycle is estimated as:

$$\sigma_{T_{Diur}} = \sqrt{\sigma_{fit}^2 + \sigma_{representativeness}^2} \tag{5.4}$$

The representativeness uncertainty is estimated as the standard deviation of the differences between the RS temperature and T_{Diur} for every available RS launch at Lauder/Invercargill, respectively.

The fitted diurnal cycle with its associated uncertainty and the ERA5 temperatures at the respective day of the year are shown in Fig. 5.2 at 925 hPa for two selected days of the year. Depending on the site and the pressure level, the representativeness uncertainty reaches values from approximately 2.8 K to 5.6 K and dominates the uncertainty on the diurnal cycle.



(a) 1st day of year (b) 183rd day of year

Figure 5.2.: Fitted diurnal cycle (blue) and ERA5 temperatures for two example days of the year from 2010 to 2016 (grey), at 925 hPa above Lauder. The ERA5 ensemble SD is shown as vertical bars around the grey lines and the uncertainty on the fitted diurnal cycle is shown in the blue error bars.

Inferring temperature anomalies above Lauder from temperature measurements above Invercargill

While RSs are launched weekly at the research facility in Lauder, twice daily RS observations are performed at the operational upper-air site in Invercargill. Hence, using RS profiles from Invercargill to estimate the temperature above Lauder greatly increases the availability of data to be included in the SASBE. Using the Invercargill measurements as a proxy for the atmospheric state above Lauder demonstrates how data collected at distributed GRUAN sites can be combined.

While the absolute temperature values at Lauder (inland) and Invercargill (coastal) are likely quite different, the temperature anomalies with respect to their respective diurnal cycles are correlated. For a maximum time difference between the Lauder and Invercargill RS launch of 1.5 hours, the correlation of the anomalies is between $R = 0.69$ and $R = 0.92$, with a mean correlation of $R = 0.85$. This correlation of anomalies is exploited to estimate temperature anomalies above Lauder based on temperature profiles from RSs launched at Invercargill.

The temperature anomaly (T') with respect to the diurnal cycle is calculated from the RS measurement and the diurnal cycle calculated at the respective site as $T'(t) = T_{RS}(t) - T_{Diur}(t)$. The 1σ uncertainty on the temperature anomalies is calculated as $\sigma_{T'(t)} = \sqrt{\sigma^2_{T_{RS}(t)} + \sigma^2_{T_{Diur}(t)}}$. When the separation in launch time between Lauder and Invercargill is less than 1.5 hours, the data are used to train a regression model (see Eq. (5.5)) that is applied to estimate temperature anomalies above Lauder from the temperature anomaly of a RS launched from Invercargill.

Four basis functions are used in the regression, i.e. the offset term that accounts for the constant bias between Lauder and Invercargill (Term 1), the temperature anomaly above Invercargill (Term 2), the difference (always Invercargill minus Lauder) in the surface pressure (ΔSP, Term 3), and the difference in the surface temperature anomaly ($\Delta ST'$, Term 4).

$$\widehat{T}'_{Lau} = \underbrace{\gamma}_{\text{Term 1}} + \underbrace{\beta \cdot T'_{Inv}}_{\text{Term 2}} + \underbrace{\eta \cdot \Delta SP}_{\text{Term 3}} + \underbrace{\kappa \cdot \Delta ST'}_{\text{Term 4}} + \underbrace{\epsilon}_{\text{Term 5}} \tag{5.5}$$

γ, β, η, and κ are the regression coefficients and ϵ (Term 5) is the residual which cannot be explained with the regression model. Term 2 is expanded with a Fourier series, where θ is the wind direction at the respective pressure level above Invercargill.

$$\beta = \beta_0 + \beta_1 sin(2\pi\theta) + \beta_2 cos(2\pi\theta) + \beta_3 sin(4\pi\theta) + \beta_4 cos(4\pi\theta) \tag{5.6}$$

Including the Fourier pairs and normalising all basis functions (except γ) by subtracting

their mean value leads to:

$$\begin{aligned}\widehat{T}'_{Lau} = \gamma + \beta_0 \cdot (T'_{Inv} - \overline{T'_{Inv}}) + \beta_1 \cdot \left(T'_{Inv}\sin(2\pi\theta) - \overline{T'_{Inv}\sin(2\pi\theta)}\right) \\ +\beta_2 \cdot \left(T'_{Inv}\cos(2\pi\theta) - \overline{T'_{Inv}\cos(2\pi\theta)}\right) + \beta_3 \cdot \left(T'_{Inv}\sin(4\pi\theta) - \overline{T'_{Inv}\sin(4\pi\theta)}\right) \\ +\beta_4 \cdot \left(T'_{Inv}\cos(4\pi\theta) - \overline{T'_{Inv}\cos(4\pi\theta)}\right) + \eta \cdot (\Delta SP - \overline{\Delta SP}) + \kappa \cdot (\Delta ST' - \overline{\Delta ST'}) + \epsilon\end{aligned} \tag{5.7}$$

The overbar denotes the mean of the basis function. At 925 hPa the wind direction is not recorded in the input dataset and therefore a simplified regression model (see Eq. (5.8)), excluding the Fourier expansion is applied.

$$\widehat{T}'_{Lau} = \gamma + \beta \cdot (T'_{Inv} - \overline{T'_{Inv}}) + \eta \cdot (\Delta SP - \overline{\Delta SP}) + \kappa \cdot (\Delta ST' - \overline{\Delta ST'}) + \epsilon \tag{5.8}$$

The residuals (ϵ) are plotted in Fig. 5.3 for varying maximum time differences $\Delta t_{regression}$ between the Lauder and Invercargill launch times. Decreasing the time difference, in general, decreases the residuals at most levels slightly. However, at 30 hPa the regression model trained on collocations from within 6 hours leads to the smallest residuals. Further analysis shows that the correlation between Lauder and Invercargill temperature anomalies is larger for a maximum time difference of 3 hours ($R = 0.78$) and 6 hours ($R = 0.76$) than for 1.5 hours ($R = 0.69$). The sample size for collocations within 1.5 hours is small and contains some large differences between the Lauder and Invercargill anomalies. The increased persistence in the stratosphere, combined with the small sample size, leads to a larger correlation when expanding the time difference. Since the maximum time difference of 1.5 hours minimises the residuals at most pressure levels, it is used as $\Delta t_{regression}$. This choice is ultimately an expert decision and further options could be tested to ensure an optimal choice is made, e.g. different maximum time differences could be applied at each pressure level. For the purpose of demonstrating this methodology, a constant value has been selected with the expectation that the results would not change significantly for reasonable variation of $\Delta t_{regression}$. At the surface, plotted at 965 hPa in Fig. 5.3 for convenience, the residuals are zero based on the availability of hourly surface data. The calculation of the residuals at the surface level provides a test of the regression model. Within the SASBE, the measurements of the automatic weather station are used at the surface level.

The uncertainties on the Invercargill and Lauder measurements and anomalies, and on the regression coefficients, are propagated through Eq. (5.7) and Eq. (5.8), following standard error propagation rules (e.g. Bevington and Robinson 2003). Uncertainties of different variables are assumed to be uncorrelated since the error covariance matrices are unknown. The uncertainty on the estimated Lauder temperature anomaly is then calculated including the partial derivatives with respect to all variables. The uncertainty for the full regression model is shown in Eq. (5.9) and a simplified version is applied at the 925 hPa level.

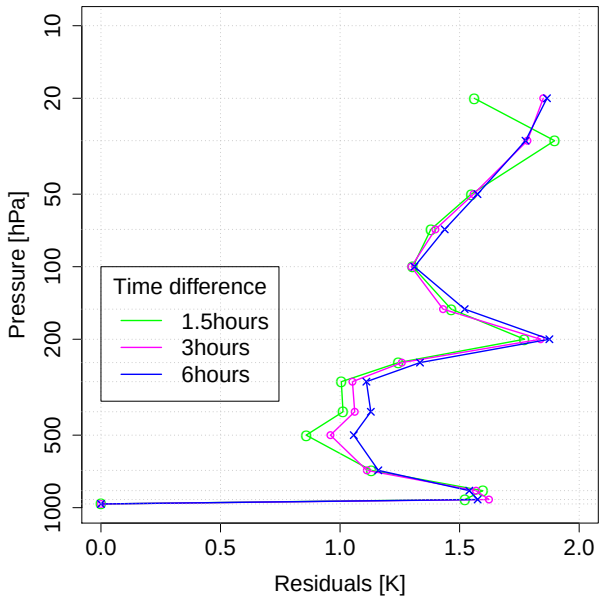


Figure 5.3.: Temperature residuals for the regression model described in Eq. (5.5). The regression model is trained with three different datasets containing Lauder and Invercargill RS measurements with launch times having a maximum time difference of 1.5 hours (green, 324 collocations), 3 hours (pink, 646 collocations), and 6 hours (blue, 879 collocations).

$$\sigma_{\widehat{T}'_{Lau}} = \sqrt{\sigma_{\gamma}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \gamma}\right)^2 + \sigma_{\beta_0}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \beta_0}\right)^2 + \sigma_{\beta_1}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \beta_1}\right)^2 + \sigma_{\beta_2}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \beta_2}\right)^2 + \sigma_{\beta_3}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \beta_3}\right)^2 + \sigma_{\beta_4}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \beta_4}\right)^2 + \sigma_{T'_{Inv}}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial T'_{Inv}}\right)^2 + \sigma_{\theta}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \theta}\right)^2 + \sigma_{\eta}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \eta}\right)^2 + \sigma_{\Delta SP}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \Delta SP}\right)^2 + \sigma_{\kappa}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \kappa}\right)^2 + \sigma_{\Delta ST'}^2\left(\frac{\partial \widehat{T}'_{Lau}}{\partial \Delta ST'}\right)^2} \tag{5.9}$$

As a result, every estimated Lauder temperature anomaly has an individual uncertainty estimate which includes the uncertainty on the measurements and Invercargill anomaly, as well as the uncertainty introduced by the regression model.

Unfortunately, there appears to be considerable uncertainty about uncertainties associated with RS measurements. With the exception of the GRUAN RS data product, which is corrected for all known biases, and provides an uncertainty best-estimate on every value, the different terminology used by different authors of publications using RS measurements leads to confusion. Furthermore, the uncertainty estimates are not traceable and are commonly provided as a general estimate, rather than as an individual value for each measurement.

For the purpose of this temperature SASBE, rather than unravelling the different esti-

Instrument and variable	Uncertainty
Automatic weather station temperature	0.2 K
Automatic weather station pressure sensor	0.5 hPa
Vaisala RS92 temperature	0.25 K
Vaisala RS80 temperature	0.5 K
Vaisala RS80/RS92 wind direction	3°

Table 5.3.: Estimates of uncertainties associated with the measurements used in the SASBE.

mates of uncertainties for the same instruments, we set the 1σ uncertainty to the values given in Table 5.3. The exact values of the uncertainties might differ from these estimates, but this is irrelevant to the mechanics of the methodology and to the temperature estimate itself, but could inflate or deflate the uncertainty bars. The 1σ temperature uncertainty for the Vaisala RS92 is set to 0.25 K based on Vaisala (2013), which is in agreement with Steinbrecht et al. (2008). This uncertainty is doubled for the Vaisala RS80, which is the older RS model manufactured by Vaisala. The uncertainty on the wind direction measured with the RS92 and RS80 instruments is estimated as 3° at the 1σ level. The 1σ uncertainties for the automatic weather stations at Lauder and Invercargill are set to 0.2 K and 0.5 hPa for temperature and pressure respectively.

Combining the individual components of the site atmospheric state best estimate

The temperature SASBE for Lauder is calculated from the different components presented above. The diurnal temperature T_{Diur} builds the foundation for the SASBE. Estimates of the temperature anomaly above Lauder ($\widehat{T'}_{Lau}$) inferred from RS launches at Invercargill are available approximately 12-hourly (the $\widehat{\ }$ symbol is used for inferred quantities). The Lauder temperature anomalies (T'_{Lau}) are available about once weekly.

At any time t the SASBE temperature $T_{SASBE}(t)$ may be influenced by several RS measurements made at Lauder and Invercargill (the maximum time difference of measurements influencing the temperature SASBE is set to 30 days). Weighting determines how much weight is given to a certain temperature anomaly. The weight ϕ determines how much influence the Lauder and Invercargill measurements have. At the time of a RS measurement at Lauder $\phi = 1$ which implies that the entire weight is given to the Lauder temperature anomalies. With further time difference from the closest launch at Lauder (Δt_0), ϕ decreases based on the autocorrelation function (acf). Autocorrelation is the correlation of a time series with a lagged (delayed) version of itself (Wilks 2011). Here ϕ is calculated from ERA5 temperature anomalies with respect to the ERA5 diurnal cycle as:

$$\phi = acf(\Delta t_0), \quad \text{if} \quad \phi < 0.5 \quad => \quad \phi = 0.5 \tag{5.10}$$

The lowest value of ϕ is restricted to 0.5 to minimise the maximum weight given to the Invercargill component. The choice of this threshold is ultimately an expert decision. A quantitative comparison of different values for ϕ could be made by analysing the effects of different lower thresholds for ϕ during the period of an extensive measurement campaign.

Furthermore, another weight is required to determine how much influence any individual temperature anomaly from Lauder or Invercargill has. These weights are taking into

account the time difference (Δt_i,Δt_j) of an individual launch to the time for which the SASBE is calculated. For Lauder temperature anomalies these weights, w_i, are calculated as:

$$w_i = \frac{1}{\Delta t_i^2} \cdot \frac{1}{\sum_{i=1}^{N} \Delta t_i^2} \tag{5.11}$$

from the N Lauder launch times t_i. For the estimated anomalies inferred from Invercargill measurement these weights, w_j are calculated from the M Invercargill launch times t_j as:

$$w_j = \frac{1}{\Delta t_j^2} \cdot \frac{1}{\sum_{i=1}^{N} \frac{1}{\Delta t_i^2} + \sum_{i=1}^{M} \frac{1}{\Delta t_j^2}} \tag{5.12}$$

The weights for Lauder anomalies are normalised with the sum over all Lauder weights and the anomalies implied from measurements in Invercargill are normalised with the sum of their own weight plus the sum of the Lauder weights. This insures that the inferred temperatures have limited influence.

If a positive temperature anomaly was determined from a RS measurement at a given time, there is some confidence that the anomaly would still be positive in the hours before and afterwards. However, as a best-estimate, the temperature anomaly a few hours after the measurement may conservatively be assumed lower than at the time of the measurement. This effect is included into the SASBE by attenuating the temperature anomaly with the autocorrelation function. This attenuated anomaly T^* is calculated as:

$$T^*_{i_{Lau}}(t) = T'_{Lau}(t_i) \cdot acf(\Delta t_i) \tag{5.13}$$

with $\Delta t_i = |t - t_i|$ for Lauder. The attenuated estimate of anomalies above Lauder inferred from Invercargill RS measurements is calculated as:

$$\widehat{T}^*_{j_{Lau}}(t) = \widehat{T}'_{Lau}(t_j) \cdot acf(\Delta t_j) \tag{5.14}$$

where $\Delta t_j = |t - t_j|$.

Taking the different weights and the decaying temperature anomalies into account, the SASBE temperature is calculated as:

$$T_{SASBE}(t) = \underbrace{T_{Diur}(t)}_{\text{Diurnal cycle}} + \underbrace{\sum_{i=1}^{N} \phi \cdot w_i \cdot T^*_{i_{Lau}}(t)}_{\text{Lauder component}} + \underbrace{\sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot \widehat{T}^*_{j_{Lau}}(t)}_{\text{Invercargill component}} \tag{5.15}$$

As the SASBE requires an uncertainty estimate on every temperature value, the uncertainties are propagated through Eq. (5.15). In a first step, Eq. (5.15) is rewritten, by expanding the terms for the weighted temperature anomaly, as:

$$T_{SASBE}(t) = T_{Diur}(t) + \sum_{i=1}^{N} \phi \cdot w_i \cdot acf(\Delta t_i) \cdot [T_{RS_{Lau}}(t_i) - T_{Diur}(t_i)] + \sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot acf(\Delta t_j) \cdot \left[\widehat{T}_{RS_{Lau}}(t_j) - T_{Diur}(t_j)\right] \tag{5.16}$$

where $T_{RS_{Lau}}$ is the temperature measured with a RS above Lauder and $\widehat{T}_{RS_{Lau}}$ is a regression model estimated temperature above Lauder based on the Invercargill RS flight and the diurnal temperature cycle above Lauder. The uncertainty on the temperature SASBE is calculated from the following components:

$$\sigma_{T_{SASBE}(t)} = \sqrt{\sigma^2_{T_{Diur}} \left(\frac{\partial T_{SASBE}}{\partial T_{Diur}}\right)^2 + \sigma^2_{T_{RS_{Lau}}} \left(\frac{\partial T_{SASBE}}{\partial T_{RS_{Lau}}}\right)^2 + \sigma^2_{\widehat{T}_{RS_{Lau}}} \left(\frac{\partial T_{SASBE}}{\partial \widehat{T}_{RS_{Lau}}}\right)^2} \tag{5.17}$$

with the partial derivatives:

$$\frac{\partial T_{SASBE}}{\partial T_{Diur}} = 1 - \sum_{i=1}^{N} \phi \cdot w_i \cdot acf(\Delta t_i) - \sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot acf(\Delta t_j) \tag{5.18}$$

$$\frac{\partial T_{SASBE}}{\partial T_{RS_{Lau}}} = \sum_{i=1}^{N} \phi \cdot w_i \cdot acf(\Delta t_i) \tag{5.19}$$

and

$$\frac{\partial T_{SASBE}}{\partial \widehat{T}_{RS_{Lau}}} = \sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot acf(\Delta t_j) \tag{5.20}$$

respectively, and therefore:

$$\sigma_{T_{SASBE}(t)} = \sqrt{\sigma^2_{T_{Diur}} \left(1 - \sum_{i=1}^{N} \phi \cdot w_i \cdot acf(\Delta t_i) - \sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot acf(\Delta t_j)\right)^2 + \sigma^2_{T_{RS_{Lau}}} \left(\sum_{i=1}^{N} \phi \cdot w_{i_{Lau}} \cdot acf(\Delta t_i)\right)^2 + \sigma^2_{\widehat{T}_{RS_{Lau}}} \left(\sum_{j=1}^{M} (1-\phi) \cdot w_j \cdot acf(\Delta t_j)\right)^2} \tag{5.21}$$

This uncertainty, at the 1σ level, was calculated for every temperature value in the SASBE and is shown as black uncertainty bars in the top panels of Figs. 5.4-5.7

5.1.5. Results

The temperature SASBE for Lauder is available in hourly resolution for 1997 to 2012 at 16 vertical levels. An estimate of the uncertainty is available with every temperature value. At the surface, the temperature SASBE consists of the measurements from the automatic weather station in Lauder. If no temperature or pressure measurement is available at a given time, the values are linearly interpolated if measurements are available within ±1 hour (temperature) or ±12 hours (pressure). In this study, the uncertainty added by the interpolation is not taken into account.

As an example, Figs. 5.4-5.7 show the hourly temperature SASBE at 925 hPa, and its individual components, for 6th to 9th December 2010, respectively. The panel (a) in each figure shows the SASBE temperature (black line) and its associated 1σ uncertainty bars. Also shown is the diurnal cycle and its uncertainty (green), as well as the results of a denial study (red), which calculates a best-estimate of the temperature assuming that no RS measurements are available at Lauder. At times when no RS measurement above Lauder influences the temperature SASBE, the temperature of the denial study equals the temperature SASBE. Blue stars show the temperature measured with a RS above Lauder, if available, and red stars show the temperature that is inferred from the Invercargill RSs by transferring the anomalies to Lauder using a regression model as described in Section 5.1.4. At times when a measurement is available above Lauder, the SASBE temperature is identical to the RS measurement, as it receives the full weight, i.e. $\phi = 1$ (see panel (d)).

The panel (b) shows the uncertainty added by the individual components and the total uncertainty on the SASBE, i.e. the terms of squared Eq. (5.17). At the exact time of a Lauder RS measurement, the uncertainty on the SASBE equals the RS uncertainty, and increases with time away from a measurement at Lauder, as the other terms influence the SASBE and therefore increase the uncertainty.

Panel (c) shows the weighted contribution from Lauder RS anomalies $\left(\sum_{i=1}^{N} \phi \cdot w \cdot T^*_{i_{Lau}}(t)\right)$ in blue and the weighted contribution from estimated Lauder temperature anomalies $\left(\sum_{j=1}^{M} \phi \cdot w_j \cdot \widehat{T}^*_{j_{Lau}}(t)\right)$ in red.

The bottom panel shows the weights ϕ and $(1 - \phi)$ which are given to the Lauder anomalies and estimated Lauder anomalies, respectively. Typically two RSs are launched daily from Invercargill at about 10 UTC and 22 UTC (as shown in Figs. 5.4 to 5.7), while typically one RS is launched from Lauder each week, here visible in Fig. 5.5. Close to the time of the Lauder RS launch, the uncertainty on the SASBE decreases as it is increasingly dominated by the RS uncertainty. The uncertainty reaches its smallest value at the exact time of a RS measurement at Lauder. It can further be seen in Fig. 5.5 that using the Invercargill measurements as a proxy for the temperature above Lauder is an improvement compared to using only *a priori*. This becomes obvious when comparing the differences between the SASBE and (i) denial study, and (ii) the diurnal cycle at the time of a Lauder RS measurement. As the temperature of the denial study is closer to the SASBE than the diurnal cycle, including information from Invercargill would be beneficial at this instance. However, it can be questioned whether using estimates of temperature anomalies above Lauder which are based on measurements from Invercargill is generally superior to using the ERA5 diurnal cycle alone. Analysis of the normalises

squared residuals between the Lauder RS measurements and (i) the diurnal cycle and (ii) the diurnal cycle plus temperature anomalies estimated using the regression model and the Invercargill RSs indicate that integrating measurements from Invercargill into the temperature SASBE for Lauder improves the SASBE (figures not shown here). However, it can be seen in Fig. 5.4 and Fig. 5.7 that the uncertainty is larger close to the time of an Invercargill launch than with further distance. Further investigation is required to ensure that the improvement in the temperature estimate based on the Invercargill measurements is reflected in uncertainty estimate of the temperature SASBE.

The results of this study indicate that RS measurements from Invercargill, together with a transfer algorithm, add value to the temperature SASBE for Lauder. To transfer the information of RS temperatures, the correlation of temperature anomalies between Lauder and Invercargill is exploited. While the SASBE presented here will not at all times give an accurate estimate of the true temperature, it is a best estimate of this unknown true value, which includes an estimate of the uncertainty to indicate the confidence that users of the SASBE should have in the temperature estimate at a given time.

5.1.6. Conclusions

This study demonstrates a method to combine the temperature measurements made at a distributed upper-air site into a SASBE. Distributed sites within the context of GRUAN are clusters of instruments at different locations which collectively submit data to the GRUAN archive. Examples of such GRUAN sites are Cabauw/De Bilt, The Netherlands, Beltsville/Sterling, US, and Lauder/Invercargill, New Zealand. While the instrumentation might differ at other upper-air sites, the methodology described here can be adapted for other distributed sites.

Within the development of this data product, some pragmatic choices have been made. While other researchers might choose some parameters differently, the methodology itself remains unaffected. Parameters to be selected include the maximum separation in time used to choose data to train the regression model, and the maximum weight given to the estimated anomalies. Furthermore, different options for the calculation of the weights might be possible.

The new data product presented here, provides an hourly resolved best-estimate of the temperature above Lauder, New Zealand, at 16 vertical levels from the surface to 10 hPa. This SASBE combines RS and automatic weather station measurements from Lauder and Invercargill with a diurnal cycle calculated from the new ERA5 reanalysis by ECMWF. The information from RS measurements made at Invercargill is transferred to Lauder using a regression model. While it would be preferable to have high temporal resolution (e.g daily) measurements available at Lauder, using the RS measurements from Invercargill improves the estimate of the temperature compared to merely using the ERA5 diurnal cycle. Using the temperature anomalies of RSs launched from Invercargill with respect to the diurnal cycle, introduces uncertainty, which is calculated by error propagation through the regression model. This meets the requirements outlined by GCOS-170 (2013) regarding the handling of uncertainties when transferring information from the location of a measurement to another location. While using Invercargill data in the SASBE improves representativeness of the SASBE with respect to the temperatures measured at Lauder,

further research into the uncertainty resulting from the transfer of data is suggested to reflect this improvement in the overall uncertainty on the best estimate-temperature.

As ERA5 has an hourly resolution and provides uncertainties calculated by a 10-member ensemble, it is well-suited to provide a default diurnal cycle which defines the SASBE in the absence of any measurements. As measurements become available, the temperature SASBE diverges from the diurnal cycle towards the measurements. At the time a RS measurement is made above Lauder, the temperature SASBE is defined entirely by the measurement and the uncertainty equals the measurement uncertainty. While the SASBE will not always represent the true temperature, it is a best-estimate of the true temperature constructed by combining different sources of information about the temperature above Lauder. The uncertainty on the data product indicates the confidence that a user should have in the temperature estimate at a given time. The uncertainty is an essential part of the SASBE, and users of the data product are urged to be cognizant of the uncertainties when using the temperature SASBE.

Due to its high temporal resolution, and the included estimate of the uncertainty, the SASBE is well-qualified for the validation of space-based sensors. Most overpasses of a satellite during the years 1997-2012 will lead to a close collocation in time, given the hourly resolution of the SASBE. Using individual RS measurements rather than a SASBE can minimise the number of collocations significantly, as was the case in Calbet et al. (see e.g. 2017), who analysed differences between the Infrared Atmospheric Sounding Interferometer (IASI, Siméoni et al. 1997) and GRUAN RS profiles, but was limited to using data from one GRUAN site only. If the satellite data product, i.e. the retrieved temperature profile, is provided with an uncertainty estimate, a quantitative comparison between the SASBE and the temperature retrieval is possible. The comparisons can either be made using direct collocations or by applying an NWP model as a transfer standard (Tradowsky et al. 2017) to eliminate effects caused by imperfect spatial and, to a smaller degree, temporal collocation. However, most space-based sensors do not measure ECVs directly. For example, radiometers such as AIRS and IASI measure the TOA radiance and a radiative transfer model is required to retrieve atmospheric ECVs. As the retrieval of atmospheric temperatures from radiance is an ill-posed problem, it can be preferable to validate space-based radiometers in their native measurement space (Calbet et al. 2017). With a radiative transfer model and additional water vapour profiles, the temperature SASBE can be used to provide such a radiance space validation as it has been described in Tradowsky et al. (2016). The uncertainties provided in the SASBEs can be propagated into uncertainties in the TOA radiances, by using a Monte Carlo approach. For this purpose, the TOA radiances are calculated many times (e.g. 100s of model runs) with slightly changed input variables (i.e. temperature, ozone and water vapour). Every input variable is varied between its upper and lower uncertainty bound and different combinations of the variables are used to calculate the TOA radiances. The uncertainty on the modelled radiances is then obtained from the spread of the calculated TOA radiances for different variations of the input variables. The modelled radiances can then be compared with radiances obtained from satellite-based radiometers above the specific location. This method has the advantage that it validates the space-based radiometers in radiance space without requiring a retrieval and the *a priori* information required within the retrieval.

The temperature SASBE may also be used as *a priori* for the retrieval of other ECVs

at Lauder, i.e. a temperature profile is required as *a priori* for the retrieval of an ozone profile from the stratospheric ozone lidar based at Lauder (Brinksma et al. 1997). As the RS measurements made at Lauder have not been assimilated into NWP models prior to 2016, the SASBE can be used as an approximately independent validation data set for model calculations (not entirely independent as the Invercargill measurements have been assimilated). Its hourly resolution and the availability of uncertainties make the SASBE especially valuable. The ERA5 reanalysis is used to calculate the diurnal cycle and the anomalies with regards to this diurnal cycle above Lauder and Invercargill. However at times close to a Lauder RS measurements, the SASBE is primarily determined by the Lauder RS temperature. Thus, the SASBE may be valuable to assess the accuracy of the ERA5 reanalysis when applying close collocation criteria for the time (i.e. around the RS measurement time). While this is not different to using the Lauder RSs directly, the SASBE is available in an easier accessible and consistent format.

This publication describes the development of a temperature SASBE for the GRUAN site at Lauder, New Zealand. A method to combine measurements from distributed sites is presented which accounts for the additional uncertainty that is induced by using the measurement of an ECV made at one location as a proxy for the ECV at another location nearby. This is an essential first step to enhance the value of distributed GRUAN sites.

Acknowledgement The authors would like to thank the New Zealand MetService for the RS data from Invercargill and Bruno Kinoshita (NIWA) for converting a subset of RSs profiles from PCCORA binary into netCDF format. The measurements and activities at Lauder are supported by the Strategic Science Investment Fund (SSIF) through the New Zealand Ministry of Business, Innovation and Employment (MBIE). The SASBE contains modified Copernicus Climate Change Service information [2017].

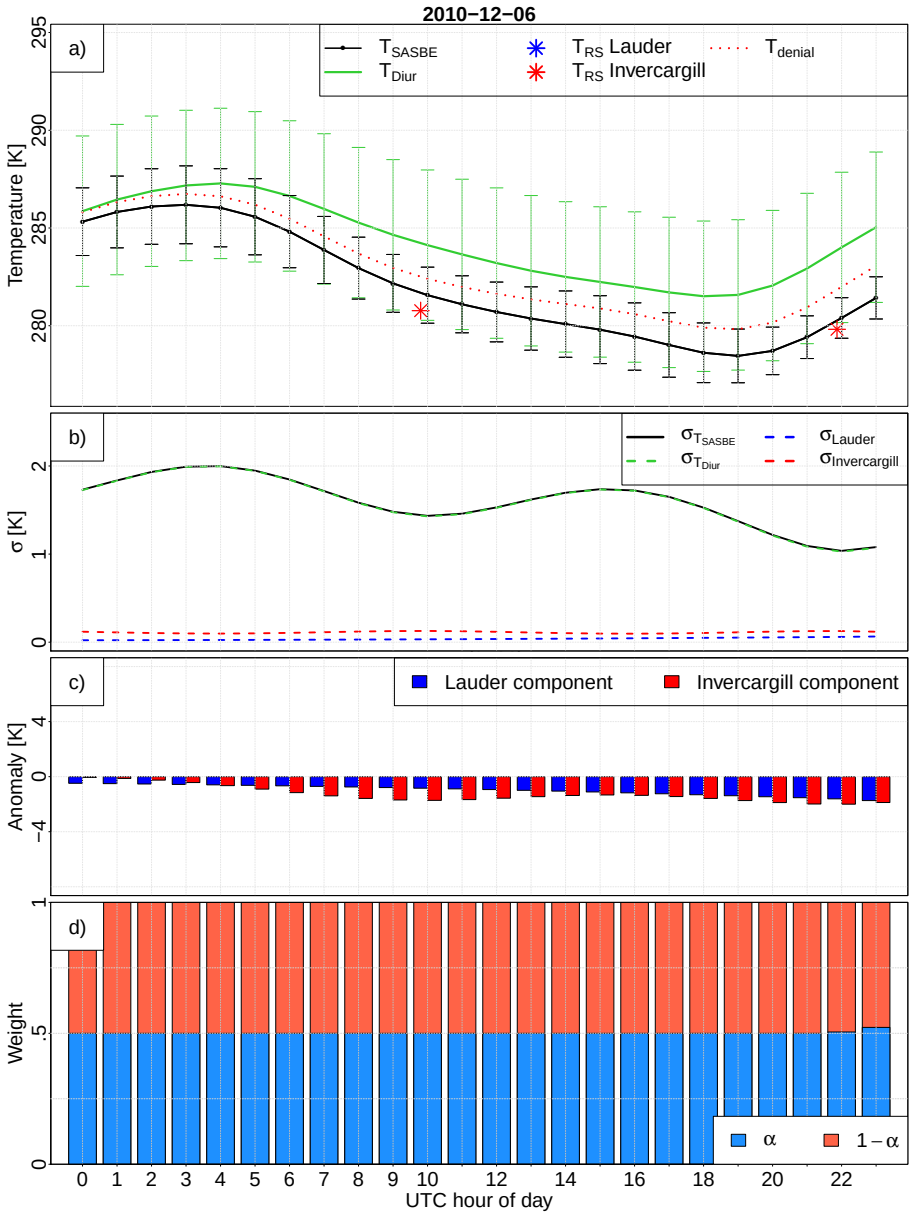


Figure 5.4.: Temperature SASBE at 925 hPa and its individual components for the 6th December 2010. Top panel: Temperature SASBE (black line), diurnal cycle (green line), denial study (i.e. pretending no measurements exist at Lauder, dotted red line), and RS temperature measurement if available (Invercargill: red star, Lauder: blue star; only shown on days when an observation is available). Second panel: squared uncertainty components from Eq. (5.17), i.e. uncertainty added by: the diurnal cycle (red), the Lauder RS anomaly (blue), and the estimated Lauder anomaly (red). The squared combined uncertainty is shown as black line. Third panel: Weighted Lauder RS temperature anomaly ($\sum_{i=1}^{N} \phi \cdot w \cdot T^*_{i_{Lau}}(t)$, blue bars) and weighted estimated Lauder temperature anomaly ($\sum_{j=1}^{M} \phi \cdot w_j \cdot \widehat{T}^*_{j_{Lau}}(t)$, red bars). Bottom panel: weights ϕ and $(1-\phi)$ which determine how much weight is given to the Lauder RS temperature anomalies and estimated Lauder anomalies based on RS measurements taken above Invercargill.

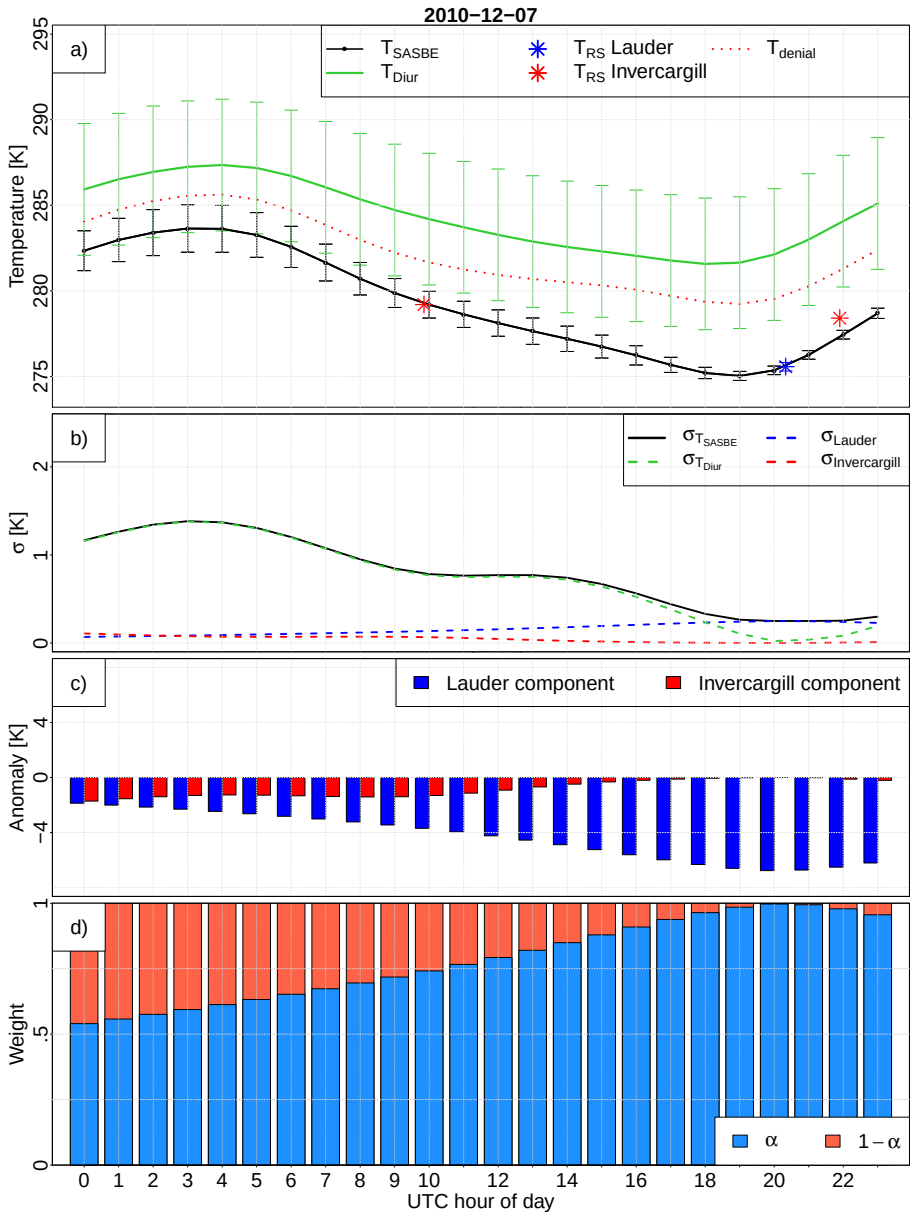


Figure 5.5.: As Fig 5.4, but for the 7th December 2010.

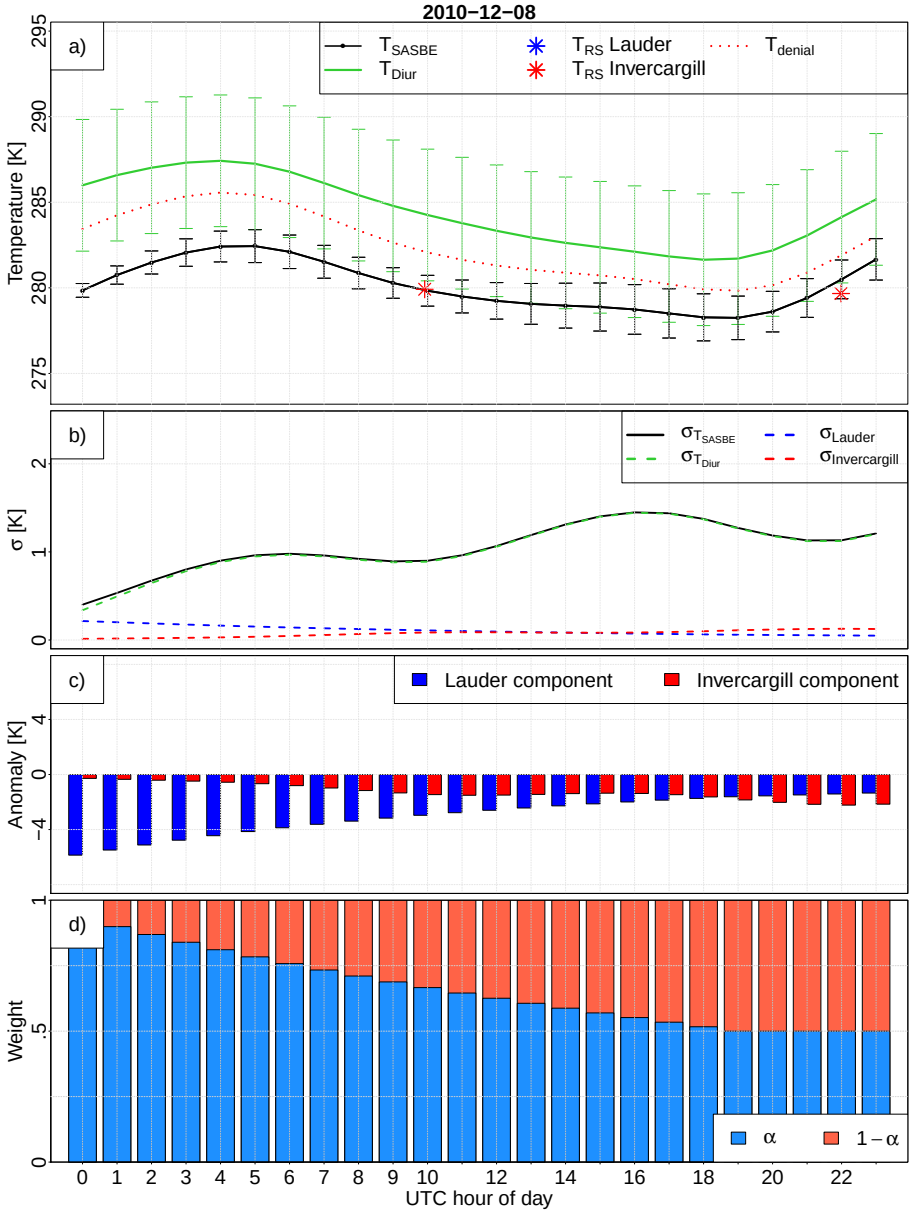


Figure 5.6.: As Fig 5.4, but for the 8th December 2010.

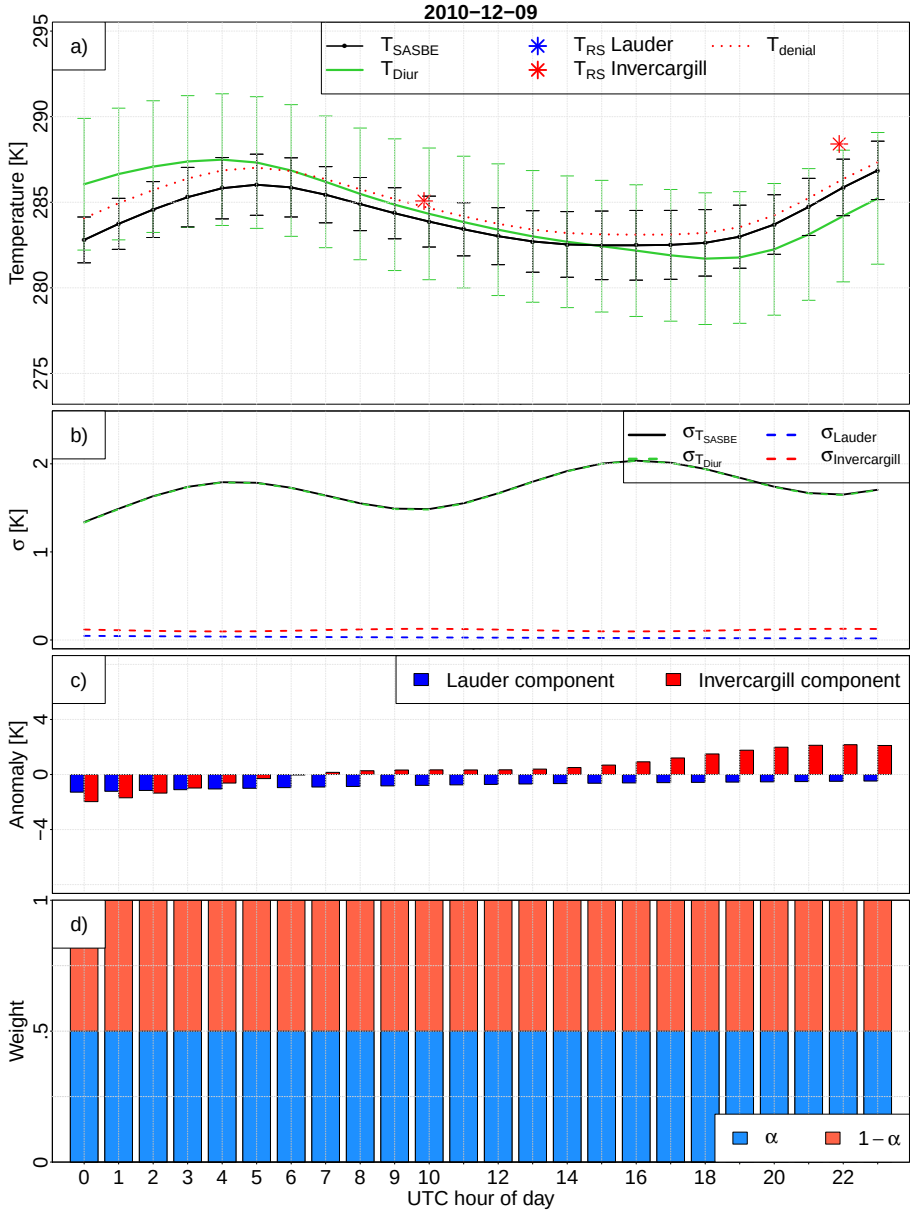


Figure 5.7.: As Fig 5.4, but for the 9th December 2010.

Bibliography

Aumann, H.H., M.T. Chahine, C. Gautier, M.D Goldberg, E. Kalnay, L.M. McMillin, H. Revercomb, P.W. Rosenkranz, W.L. Smith, D.H. Staelin, and et al. (2003). "AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2, pp. 253–264.

Bevington, P.R. and D.K. Robinson (2003). *Data Reduction and Error Analysis for the Physical Sciences*. 3rd ed. Mc Graw Hill.

Bodeker, G.E., S. Bojinski, D. Cimini, R.J. Dirksen, M. Haeffelin, J.W. Hannigan, D.F. Hurst, T. Leblanc, F. Madonna, M. Maturilli, A.C. Mikalsen, R. Philipona, T. Reale, D.J. Seidel, D.G.H. Tan, P.W. Thorne, H. Vömel, and J. Wang (2016). "Reference Upper-Air Observations for Climate: From Concept to Reality". In: *Bull. Amer. Meteor. Soc.* 97, pp. 123–135. DOI: 10.1175/bams-d-14-00072.1.

Bodeker, G.E. and S. Kremser (2015). "Techniques for analyses of trends in GRUAN data". In: *Atmos. Meas. Tech.* 8, 1673–1684. DOI: 10.5194/amt-8-1673-2015.

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zemp (2014). "The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy". In: *Bull. Amer. Meteor. Soc.* 95.9, pp. 1431–1443. DOI: 10.1175/BAMS-D-13-00047.1.

Brinksma, E.J., D.P.J. Swart, J.B. Bergwerff, Y.J. Meijer, and F.T Ormel (1997). "Advances in Atmospheric Remote Sensing with Lidar". In: ed. by A. Ansmann, R. Neuber, P. Rairoux, and U. Wandinger. Springer, Berlin, Heidelberg. Chap. Stratospheric and Mesospheric Profiling, RIVM Stratospheric Ozone Lidar at NDSC Station Lauder: Routine Measurements and Validation During the OPAL Campaign, pp. 529–532. DOI: 10.1007/978-3-642-60612-0_128.

Calbet, X. (2016). "Assessment of adequate quality and collocation of reference measurements with space-borne hyperspectral infrared instruments to validate retrievals of temperature and water vapour". In: *Atmos. Meas. Tech.* 9, pp. 1–8. DOI: 10.5194/amt-9-1-2016.

Calbet, X., N. Peinado-Galan, P Rípodas, T. Trent, R. Dirksen, and M. Sommer (2017). "Consistency between GRUAN sondes, LBLRTM and IASI". In: *Atmos. Meas. Tech.* 10, pp. 2323–2335. DOI: 10.5194/amt-10-2323-2017.

Dirksen, R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, and H. Vömel (2014). "Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde". In: *Atmos. Meas. Tech.* 7, pp. 4463–4490. DOI: 10.5194/amt-7-4463-2014.

GCOS-112 (2007). *GCOS REFERENCE UPPER-AIR NETWORK (GRUAN): Justification, requirements, siting and instrumentation options*. Tech. rep. GCOS-112 (WMO/TD No. 1379). World Meteorological Organization. URL: https://www.wmo.int/pages/prog/gcos/Publications/gcos-112.pdf.

GCOS-170 (2013). *The GCOS Reference Upper-Air Network (GRUAN)*. WIGOS Technical Report 2013-2. World Meteorological Organization.

GCOS-200 (2016). *The Global Observing System for Climate: Implementation Needs*. Tech. rep. World Meteorological Organization.

Hersbach, H. and D. Dee (2016). *ERA5 reanalysis is in production*. ECMWF Newsletter No.147. URL: https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production.

Ingleby, B. (2017). *An assessment of different radiosonde types 2015/2016*. Technical Memorandum 807. European Centre for Medium-Range Weather Forecasts.

Maillard Barras, E., A. Haefele, R. Stübi, and D. Ruffieux (2015). "A method to derive the Site Atmospheric State Best Estimate (SASBE) of ozone profiles from radiosonde and passive microwave data". In: *Atmos. Meas. Tech. Discuss.* 8.3, pp. 3399–3422. DOI: 10.5194/amtd-8-3399-2015.

Siméoni, D., C. Singer, and G. Chalon (1997). "Infrared atmospheric sounding interferometer". In: *Acta Astronautica* 40.2-8, pp. 113–118.

Steinbrecht, W., H. Claude, F. Schönenborn, U. Leiterer, H. Dier, and E. Lanzinger (2008). "Pressure and Temperature Differences between Vaisala RS80 and RS92 Radiosonde Systems". In: *J. Atmos. Oceanic Technol.* DOI: 10.1175/2007JTECHA999.1.

Tobin, D.C., H.E. Revercomb, R.O. Knuteson, B.M. Lesht, L.L. Strow, S.E. Hannon, W.F. Feltz, L.A. Moy, E.J. Fetzer, and T.S. Cress (2006). "Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapour retrieval". In: *J. Geophys. Res.* 111. DOI: 10.1029/2005JD006103.

Tradowsky, J.S., G.E. Bodeker, P.W. Thorne, F. Carminati, and W. Bell (2016). "GRUAN in the service of GSICS: Using reference ground-based profile measurements to provide traceable radiance calibration for space-based radiometers". In: *GSICS Quarterly Newsletter* 10.2. DOI: 10.7289/V5GT5K7S.

Tradowsky, J.S., C.P. Burrows, S.B. Healy, and J.R. Eyre (2017). "A New Method to Correct Radiosonde Temperature Biases Using Radio Occultation Data". In: *Journal of Applied Meteorology and Climatology* 56.6, pp. 1643–1661. DOI: 10.1175/JAMC-D-16-0136.1.

Vaisala (2013). *Vaisala Radiosonde RS92 Performance in the WMO Intercomparison of High Quality Radiosonde Systems / Yangjiang, China 2010*. Vaisala White Paper.

Wilks, D.S. (2011). *Statistical methods in the atmospheric sciences*. 3rd. San Diego, CA: Academic Press.

6. Applications for site atmospheric state best estimates

SASBEs[12] can be used for various purposes, including the validation of satellite measurements and model output. This chapter will give a brief overview of possible applications for the temperature SASBE in Lauder and then discusses in more detail how SASBEs over the Antarctic continent could be used to robustly analyse the sensitivity of radiative transfer calculations to the spectrally resolved surface emissivity.

SASBEs combine the measurements from different sensors with additional information about the state of the target variable and therefore have benefits compared to observations with one specific instrument. An outstanding feature of SASBEs is the uncertainty estimate on every value, which is calculated by propagating the uncertainty of the input variables through all calculations, and taking representativeness uncertainties in account.

Using the temperature SASBE for Lauder, New Zealand, as *a priori* for the retrieval of atmospheric variables from ground-based remote sensing instruments Retrieving atmospheric variables using remote sensing techniques typically requires *a priori* information. Due to its high temporal resolution the temperature SASBE can provide a valuable source of *a priori* knowledge to be used in the retrieval of atmospheric variables from ground-based remote sensing instruments at Lauder. As an example, the use of the temperature SASBE in the ozone retrieval from lidar measurements is presented below, but other retrievals also require *a priori* knowledge.

The stratospheric ozone lidar was installed by the Dutch National Institute for Public Health and the Environment (RIVM) at the NIWA site in Lauder in November 1994 (Brinksma et al. 1997). For the retrieval of ozone profiles *a priori* information about the air density is required to calculate the Rayleigh correction and a temperature profile is required to determine the temperature dependent ozone absorption cross Section (Brinksma et al. 2000). In Brinksma et al. (2000) the temperature and density profile from the closest collocated RS launch at Lauder is used for the lidar ozone retrieval. Due to the weekly launch schedule at Lauder the time difference between the RS and lidar measurements can be several days. As the atmosphere is not constant over this time frame, the *a priori* used in the lidar retrieval may be an inadequate estimate of the true state of the atmosphere at the time of the lidar measurement. Using the hourly resolved temperature SASBE instead of the closest RS profile, therefore, has the potential to improve the lidar ozone retrieval which is regularly submitted to the Network for the Detection of Atmospheric Composition Change (NDACC) archive (`http://www.ndsc.ncep.noaa.gov/data/`).

Using the temperature SASBE for Lauder to validate satellite data products and top-of-the-atmosphere radiances *In situ* profile measurements of the atmosphere are especially sparse in the Southern Hemisphere. Therefore, the SASBE for Lauder, New

[12] All abbreviations can be found in the Glossary, Appendix B

Zealand, which provides hourly temperature profiles including an estimate of the uncertainty on every datum, can be very valuable to validate space-based measurement. This validation can be done (i) on product basis, i.e. the temperature profile retrieved from a satellite measurement is validated against the SASBE, or (ii) in radiance space, i.e. the temperature SASBE together with water vapour and ozone profiles and some information about the temperature above the highest level of the SASBE, are used as input to a radiative transfer model to calculate top-of-the-atmosphere (TOA) radiances which are then compared to radiances measured with a space-based radiometer.
To validate satellites measurements on product basis, the raw measurements, e.g. the spectrally resolved TOA radiances are used to retrieve profiles of e.g. temperature. A retrieval algorithm and typically some *a priori* information is required to calculate the target variable, here temperature, from the radiances. The retrieval of temperature profiles enable the direct comparison with the temperature SASBE. Comparisons can either be made using direct collocations, or by applying an NWP model as transfer standard (Tradowsky et al. 2017). For the direct collocation method SASBE profiles and retrieved profiles which are closely collocated in time and space are used. Calbet et al. (2017) uses collocation criteria of 25 km and 30 min to compare individual profiles with each other and Sun et al. (2010) and Sun et al. (2013) uses 250 km and 6 hours collocation criteria to calculate statistics of the difference between a large number of collocated RS and RO profiles. The direct collocation method relies on very strict collocation criteria, or on a large sample size to minimise the effect of variations in the atmospheric state over the separation distance between the compared measurements. The standard deviation of the differences between the compared measurements increases with relaxation of the collocation criteria (Sun et al. 2010). Different instruments typically sample a different airmass, i.e. RSs provide a profiles of *in situ* measurements while radiometers instantaneously measures the radiances in a larger region. Therefore, it is important to take into account uncertainties added by imperfect collocation and representativeness uncertainties.

To minimise the standard deviation in the comparison, a double differencing method, using an NWP model as transfer standard, can be used instead of direct collocations. Rather than comparing the imperfectly collocated data of two instruments directly with each other, the observation minus model background departures are calculated for both observation types, and their differences are analysed. Under the assumption that the model bias is constant over the separation distance between the observations, the model background becomes extraneous and the double difference provides a valid comparison of the two observation types. The idea to use double differencing is described in Haimberger et al. (2012) and an application of the method to correct RS temperature biases using RO data has been demonstrated in this thesis, and in Tradowsky (2015), Tradowsky (2016), and Tradowsky et al. (2017).

While the validation of satellite retrievals in product space, e.g. in temperature space is comparably straight forward, it can only provide limited information. A more valuable validation of the measurements of space-based radiometers is the validation in their native measurement space, i.e the validation of the TOA radiances. SASBEs and a radiative transfer model can be used to provide such a radiance space validation.The uncertainties provided in the SASBEs can be propagated into uncertainties of the TOA radiances, by using a Monte Carlo approach. For this purpose the TOA radiances are calculated many times (e.g. 100s of model runs) with slightly changed input variables (i.e. temperature,

ozone and water vapour). Every input variable is varied between its upper and lower uncertainty bound and different combinations of the variables are used to calculate the TOA radiances. The uncertainty of the modelled radiances is then obtained from the spread of the calculated TOA radiances for different variations of the input variables. The modelled radiances can then be compared with radiances obtained from satellite-based radiometers above the specific location. This method has the advantage that it validates the space-based radiometers in measurement space without requiring a retrieval and *a priori* information.

The idea to use SASBEs to provide a radiance space validation of space-based radiometers was outlines in Tradowsky et al. (2016), which is reprinted with permission below as it builds a part of this thesis.

Tradowsky et al. (2016): GRUAN in the service of GSICS: Using reference ground-based profile measurements to provide traceable radiance calibration for space-based radiometers by Jordis Tradowsky (Bodeker Scientific), Greg Bodeker (Bodeker Scientific), Peter Thorne (Maynooth University), Fabien Carminati (Met Office), William Bell (Met Office). Published in the Global Space-based Inter-Calibration System Quarterly Newsletter Vol. 10 Nr. 2, 2016,
doi:10.7289/V5GT5K7S (Tradowsky et al. 2016). Reprinted with permission.

GRUAN comprises 24 sites that measure vertical profiles of ECVs such as temperature, pressure, and water vapour. These measurements are reference-quality measurements in that all systematic biases having been accounted for and measurement uncertainties are traceable to internationally recognised measurement standards (Immler et al. 2010). The resultant long-term homogeneity of the measurement series, as well as their network-wide uniformity and coherence, makes them ideally suited for providing a reference standard for space-based radiometric measurements. Reference measurements of the atmospheric state variables influencing radiative transfer through the column, together with, for example, surface emissivity and surface temperature, can be used as input to a state-of-the-art radiative transfer model to simulate TOA radiances. Propagating the SI traceable uncertainties in the measured vertical profiles of the state variables through to uncertainties in the TOA radiances provides a degree of SI traceability for the simulated radiances. Bootstrapping methods, that also account for vertical autocorrelation in the profiles, can be used to propagate uncertainties from measured variables to the modelled radiances. The modelled radiances, with their robustly determined uncertainty estimates, are also suitable for comparison with space-based radiometric measurements. If satellite measurements are provided with their associated uncertainty estimates, a quantitative comparison between those measurements and the modelled radiances can be performed. Such a radiance-space validation supports the creation of seamless, stable, and long-term measurement series from many satellite-based instruments that are then suitable for detecting trends and variability in a wide range of atmospheric variables describing the state of Earth's atmosphere and surface.

At present, a GRUAN data product for Vaisala RS92 RS is available. Data products for other RS types, GNSS precipitable water vapour, ozonesondes, lidars, and microwave radiometers are currently under development. GRUAN is working towards providing a set of ECV profile measurements, suitable for describing the clear-sky radiative transfer

in the column, above many of the GRUAN sites. Reale et al. (2016) gives a summary of activities within the GRUAN community.

Currently under development at the Met Office is a software package, referred to as the GRUAN processor (Carminati et al. 2016), which is designed to simulate TOA spectra (L1B radiance or brightness temperatures), including uncertainties propagated to observation (radiance) space, from GRUAN measurements. The processor will also simulate TOA radiances from NWP model fields interpolated to the GRUAN locations. The GRUAN processor will therefore provide statistics on the differences between NWP fields and GRUAN observations in terms of temperature and water vapour, but also in terms of simulated TOA radiances/brightness temperatures. The collocation algorithm of the GRUAN processor uses a trilinear interpolation of latitude, longitude, and time that accounts for the RS drift and ascent time. Collocated GRUAN and model profiles are converted into radiances (or brightness temperatures) using the RTTOV (Radiative Transfer for TOVS, http://nwpsaf.eu/site/software/rttov/) fast radiative transfer model. Measurements by various past, present and future satellite instruments, which are supported by RTTOV, can be simulated. The GRUAN processor will be used to characterise the uncertainties in Met Office and ECMWF NWP models, and simulations based on them. Its development is part of the Gap Analysis for Integrated Atmospheric ECV CLIMate Monitoring (GAIA-CLIM, http://www.gaia-clim.eu/) project, which is funded by the European Union's Horizon 2020 research and innovation programme. The GRUAN processor will be integrated into the GAIA-CLIM Virtual Observatory, a freely accessible portal that will enable users to access and work with the data.

To promote the use of GRUAN measurements for the calibration and validation of space-based measurements, further activities are underway within the GRUAN community, including: The construction of Site Atmospheric State Best Estimates (SASBEs): Redundant measurements of ECVs from different instruments (e.g. lidar, sonde, microwave radiometer) can be combined into a SASBE to provide a best estimate of the spatial and temporal variability of that ECV above the site of interest, together with traceable uncertainty estimates. GRUAN sites seek to make redundant measurements of various ECVs and therefore are well equipped to deliver observations suitable for the construction of SASBEs. The establishment of GRUAN operating protocols to encourage sites to time measurements with satellite overpass times. The extent to which satellite overpass times for GRUAN sites can be calculated in advance and provided to GRUAN sites will determine, in part, the ability of those sites to provide targeted measurements for the calibration of space-based instruments. If these targeted measurements are subsequently used in a SASBE, the uncertainty in the SASBE can be reduced at the overpass time of the satellite. Currently SASBEs for temperature, water vapour and ozone concentration profiles above the GRUAN site in Lauder, New Zealand (45.038°S, 169.684°E) are under development. When finalised, these SASBEs can be used as input for radiative transfer calculations. An interactive version of the temperature SASBE above Lauder is available at http://sasbe.bodekerscientific.com/ where further information about the data product can also be found. Viewers can subscribe to be informed when the SASBEs become available to the scientific community.

After the publication of the the newsletter article reprinted above, the work on the temperature SASBE has proceeded and the second version of the SASBE is finalized and a manuscript describing the construction of the SASBE has been submitted to *Earth*

System Science Data Discussions (Tradowsky et al. 2018, submitted).

Using SASBE for the West Antarctic Ice Sheet to study the sensitivity of radiative transfer calculations to far-infrared surface emissivities Within this thesis, a temperature SASBE has been developed for Lauder, New Zealand. A similar methodology can be applied at other upper-air stations and the resulting SASBEs can be used for various purposes. Below one desired extension of the work will be described. A proposal has been submitted to secure funding for the described project.

Outgoing longwave radiation (OLR) emitted by our planet plays an important role in the global energy balance and is the main way in which Earth cools itself. The OLR depends on the temperature profile, profiles of absorbing atmospheric trace gases (e.g. water vapour), the surface emissivity, and the surface temperature. Of these, it is the longwave surface emissivity that is most poorly known and which is hardest to measure. It is essential, however, to accurately represent the Antarctic longwave surface emissivity in global climate models as roughly 50% of the terrestrial radiation from Antarctica is emitted in the far-infrared band (Chen et al. 2014; Feldman et al. 2014) i.e. at wavelengths between 15 μm and 100 μm. For much of the globe, due to water vapour in the atmosphere, the atmosphere is opaque to the surface at these wavelengths. However, the atmosphere becomes partially transparent in this far-infrared band under very dry conditions (where total column water vapour is less than 10 mm, see Fig. 1A of Feldman et al. 2014). The effect of far-infrared emissivity on OLR becomes much more pronounced if the total column water is below 1 mm. The total column water vapour is the height of liquid water at standard surface pressure and temperature, resulting from condensing the entire water vapour from the atmospheric column (Mims III et al. 2011).

Such dry conditions with total column water vapour of less than 1 mm exist above Antarctica, the Himalayas and Greenland (see e.g. Fig 2A of Feldman et al. 2014). Thus, to accurately simulate radiative balance in the atmosphere, i.e. the difference between incoming solar radiation and outgoing infrared radiation, climate models require spectrally resolved emissivities as surface boundary conditions. While space-based instruments provide global measurements of spectrally resolved outgoing longwave radiation (OLR) between 5 μm and 15.4 μm, few measurements are available beyond 15.4 μm (Feldman et al. 2014). In fact, despite its key role in the radiative energy budget over Antarctica, there are no spatially-resolved measurements of far-infrared surface emissivity over Antarctica. Currently, researchers use estimates of far-infrared emissivities in their models, but the quality of these estimates in unknown. Uncertainties in the effects of mineralogy and snow grain size on far-infrared surface emissivity can drive changes as large as 1.8-2 W/m^2 in OLR with consequences for modelled patterns of surface temperature and frozen surface extend (Feldman et al. 2014). To highlight its significance, Feldman et al. (2014) stated that this is equivalent to the effect that a surface temperature change of 2 K would have. Thus, above Antarctica, an accurate representation of the spectrally resolved far-infrared surface emissivity is essential for robust modelling of the response of the continent to climate change.

The far-infrared radiation budget is affected by the surface emissivity, by clouds and by water vapour in the atmosphere. Surface emissivity quantifies the effectiveness of a surface in emitting energy as thermal radiation. While some studies have focussed on the

influence of water vapour and clouds on the far-infrared radiation budget (e.g. Turner and Mlawer 2010; Palchetti et al. 2015), effects of far-infrared surface emissivity are currently underexplored. While space-based measurements of emissivities at shorter wavelengths ($<$15.4 μm) are available from e.g. the Atmospheric Infrared Sounder (AIRS, Aumann et al. 2003) and the Infrared Atmospheric Sounding Interferometer (IASI, Siméoni et al. 1997), very little is known about the spectrally resolved surface emissivity at wavelength beyond 15.4 μm. The surface emissivity in this band depends on, *inter alia*, the surface type, mineralogy, and the snow grain size, and is temporally and spatially heterogeneous. Surface emissivities beyond 15.4 μm are inferred from other wavelength bands, typically the mid-infrared, and can potentially be wrong. Estimates for surface, cloud and water vapour emission in the far-infrared can be checked for consistency using OLR observations collected with the Clouds and the Earth's Radiant Energy System (CERES) instrument, but compensating errors complicate the effort (Feldman et al. 2014). The quality of the estimates for surface emissivity beyond 15.4 μm is thus highly uncertain. Currently there are no global measurements to constrain the surface emissivity at wavelengths beyond 15.4 μm. Due to the heterogeneity of the surface emissivity, a downward looking instrument from a high-altitude platform is required to make the required spectrally and spatially resolved measurements of the surface emissivity. The proposed space-based CLimate Absolute Radiance and Refractivity Observatory (CLARREO) mission would have provided such measurements globally, but it is currently unclear when or if this mission will launch due to funding restrictions. Measurements of far-infrared emissivities above Antarctica made e.g. from a balloon-based platform would provide essential data over a continent where such measurements are desperately needed to reliable model local climate, which also effects the global climate.

A study to provide a robust analysis of the uncertainties in the Antarctic radiation budget caused by our knowledge gap in spectrally resolved far-infrared emissivity above Antarctica is envisioned as a first step towards performing a measurement campaign. The planned project, if funding can be ensured, will focus on the West Antarctic Ice Sheet because (i) the atmosphere there is sufficiently dry to be affected by uncertainties in far-infrared surface emissivity, and (ii) a substantial data set was recently collected there by the US Atmospheric Radiation Measurement (ARM) West Antarctic Radiation Experiment (AWARE, Lubin et al. 2017) field campaign. These data can be combined into Site Atmospheric State Best Estimates (SASBEs), which constrain a radiative transfer model (RTM) which can then be used to study the OLR sensitivity to changes in the surface emissivity. The temperature and water vapour SASBEs will be used to prescribe the atmospheric state in the RTM. The uncertainties provided as part of the SASBE can be propagated, using a Monte Carlo approach, into uncertainties in the TOA radiances (Tradowsky et al. 2016). For this purpose, the TOA radiances are calculated many times (e.g. 100s of model runs) with temperature and water vapour values that are perturbed within the distributions defined by their uncertainties i.e. every input variable is varied within the Gaussian profile of its error bounds and provided as an ensemble of inputs to the RTM used to calculate the TOA radiances. The uncertainties of the modelled radiances are then derived from the spread of the calculated TOA radiances for the ensemble of input variables.

The RTM and the SASBEs are then applied to study the sensitivity to changes in the spectrally resolved surface emissivity at wavelengths longer than 15 μm. An ensemble

of RTM simulations with different far-infrared emissivities will be generated. Based on the very limited knowledge of far-infrared emissivities above Antarctica, it will be challenging to determine the range of emissivities to be studied. Therefore, in a first step, currently available knowledge will be critically assessed to arrive at a 'best estimate' of the far-infrared surface emissivity. The assessment of current knowledge will include the range of surface emissivities used by different climate models, as well as measurements performed in laboratories. For every member of the ensemble, a Monte Carlo approach as described above will be used to estimate the uncertainty on the TOA radiance caused by the uncertainty in the atmospheric conditions (i.e. the uncertainty in the SASBEs). This allows to quantify if the effects of changes in the surface emissivities on the TOA radiances are statistically significant. The spread of the ensemble will quantify the uncertainties in the Antarctic infrared radiation budget caused by lack of knowledge of far-infrared emissivities.

Such a robust analysis of the problem is an essential first step before attempting to mitigate the knowledge gap through a measurement campaign. Ultimately, to close the current gap in our knowledge, measurements of the emissivity in the far-infrared need to be made; no amount of modelling is going to solve the problem. The described project could provide the robust basis to secure funding for such a measurement campaign following the publication of the results.

Using the temperature SASBE for Lauder, New Zealand to validate model output The high temporal resolution of the SASBE and Lauders locational in the sparsely observed Southern Hemisphere makes the SASBE valuable for the validation of models. The avialability of an uncertainty estimate on every datum allows the quantitative comparison between model output and the SASBE. The RS measurements made at Lauder have not been assimilated into NWP models before 2016, however the measurements from Invercargill have been assimilated since many years. Therefore, the SASBE can only be seen as independent of the model in the close vicinity of a Lauder RS measurement.

A subset of potential applications for SASBEs is presented here, but various other applications could be identified by specific user groups. SASBEs of specific target variables could be specifically developed for a certain task, as it was done in Tobin et al. (2006). This might include targeted measurements or the use of existing data sets. It should be encouraged to publish the methodology as an essential part of the SASBE and make the final data product freely available to the international research community. GRUAN has extended its Scheduling Task Team to include research into the combination of measurements. The new Task Team on Measurement Scheduling and Combination (https://www.gruan.org/network/task-teams/tt-scheduling/) is currently chaired by Tom Gardiner and Fabio Madonna and has 12 members, including the author of this thesis.

7. Estimating the difference in biases between two radiosonde types

Ensuring the long-term stability of Essential Climate Variables (ECVs) is of major importance to study the effects of anthropogenic climate change. However, improvements in technology lead to constant changes in the instrumentation used to perform measurements in the atmosphere. This can lead to jumps in the time series of a variable, which should be prevented if possible. To prevent temporal inhomogeneities in a data record, the time series of two instruments should overlap. The bias between the two instruments can then be estimated from the dual measurements. Performing dual RS[13] launches to produce an overlapping time series causes extra costs and is therefore uncommon within the operational upper-air network. However, within GRUAN, careful transition management is required whenever an instrument is changed. The transition from the older Vaisala RS92 instrument to the improved Vaisala RS41 is currently under way at a number of GRUAN sites which have therefore performed dual launches. In this chapter an alternative approach to dual RS launches is evaluated which could provide a cost effective option for change management. Therefore, this approach, if feasible, could also be employed at the wider operational network of upper-air sites. The results of the study indicate, however, that for the typical size of a bias between two high-quality RSs, the evaluated method is unlikely to provide reliable estimates of the difference between the instruments. The author of this thesis is the second author of the presented paper. She performed the simulation study described the paper and provided valuable input to the discussions around the methodology and to the writing of the paper.

The following section is a slightly adapted version of **Kremser, S., J.S. Tradowsky, H.W. Rust, and G.E. Bodeker (2018). "Is it feasible to estimate radiosonde biases from interlaced measurements?", *Atmospheric Measurement Techniques Discussions*, doi: 10.5194/amt-2018-6** (Kremser et al. 2018a, under review). A slightly revised version of the paper has been published in Kremser et al. (2018b) (reference added prior to the official publication of thesis).

7.1. Paper III: Is it feasible to estimate radiosonde biases from interlaced measurements?

7.1.1. Abstract

Upper-air measurements of ECVs, such as temperature, are crucial for climate monitoring and climate change detection. Because of the internal variability of the climate system, many decades of measurements are typically required to robustly detect any trend in the climate data record. It is imperative for the records to be temporally homogeneous over many decades to confidently estimate any trend. Historically, records of upper-air

[13] All abbreviations can be found in the Glossary, Appendix B

measurements were primarily made for short-term weather forecasts and, as such, are seldom suitable for studying long-term climate change as they lack the required continuity and homogeneity. Recognizing this, GRUAN has been established to provide reference-quality measurements of climate variables, such as temperature, pressure and humidity, together with well characterised and traceable estimates of the measurement uncertainty. To ensure that GRUAN data products are suitable to detect climate change, a scientifically robust instrument replacement strategy must always be adopted whenever there is a change in instrumentation. By fully characterizing any systematic differences between the old and new measurement system a temporally homogeneous data series can be created. One strategy is to operate both the old and new instruments in tandem for some overlap period to characterise any inter-instrument biases. However, this strategy can be prohibitively expensive for measurement sites operated by national weather services or research institutes. An alternative strategy that has been proposed is to alternate between the old and new instruments, so-called interlacing, and then statistically derive the systematic biases between the two instruments. Here we investigate the feasibility of such an approach specifically for RSs, i.e. flying the old and new instruments on alternating days. Synthetic data sets are used to explore the applicability of this statistical approach to RS change management.

7.1.2. Introduction

RSs are indispensable for monitoring the upper-air as they provide high vertical resolution *in situ* observations of temperature, pressure and water vapour between the surface and the upper troposphere/lower stratosphere. Determining long-term temperature trends from RS measurements is challenging because changes in instrumentation can, among other things, introduce discontinuities in the measurement time series (see Fig. 7.1). Since RS measurements are primarily made to provide the data needed to constrain weather forecasts and not to detect long-term changes in climate, little attention has been paid to ensuring the long-term homogeneity of the measurement record when changing from one instrument to another. As a result, RS data records typically fall short of the standard required to reliably detect changes in climate. Another cause of inhomogeneities in the record is undocumented changes in data processing (Thorne et al. 2011). While much effort has been spent attempting to remove discontinuities in RS data records (e.g. Sherwood et al. 2005; Randel and Wu 2006; Haimberger et al. 2012), lack of confidence in the long-term homogeneity erodes confidence in derived trends. Seidel and Free (2006) used upper-air temperatures from the NCEP-NCAR[14] reanalysis (Saha et al. 2010) to investigate the effects of sampling frequency, changes in observation schedule, and the introduction of inhomogeneities, to the RS climate data record. Their results indicate that introducing inhomogeneities into a temperature time series provides the most significant source of uncertainty on trend estimates. Maintaining the temperature measurement stability to within 0.1 K for periods of 20 to 50 years, avoids uncertainties in trend estimates in at least 99% of cases (Seidel and Free 2006). With a weaker stability requirement of 0.25 K, the uncertainty on a 50 year trend estimate increases by about 5% for twice-daily sampling. Rust et al. (2008) showed that inhomogeneities in temperature measurements

[14]National Centers for Environmental Prediction-National Center for Atmospheric Research

can cause spurious memory, leading to larger uncertainty for statistics derived from these series. The results of these studies demonstrate the need to account for any inhomogeneities in the measurement time series prior to any trend analysis.

GRUAN was established to provide reference-quality measurements of atmospheric ECVs, suitable for reliably detecting changes in global and regional climate on decadal scales. To avoid compromising the integrity of the long-term climate record, it is essential that any change, e.g. in the instrumentation or data processing, is adequately assessed before the change is implemented. For example, when transitioning from one RS type to another, inter-comparison between both RS types is required to assess a potential systematic difference between the RSs and to correct for it, ensuring a continuous homogeneous data set without any introduced discontinuities. Typically, intercomparisons of measurements from dual or quadruple (two of each instrument-type) RS flights are used to robustly detect systematic differences between the instruments (e.g. Luers and Eskridge 1998; Steinbrecht et al. 2008; Jensen et al. 2016). Results presented in Steinbrecht et al. (2008) indicated that temperature biases often increase significantly with increasing altitude, particularly in the lower stratosphere. Instrument biases are also influenced by clouds as shown in Jensen et al. (2016) who found systematic differences in temperature measurements greater than 2 K between the Vaisala RS92 and RS41 RS when exiting cloud layers. This large difference in temperature measurements between the two RSs was attributed to the wet-bulb effect, where the temperature sensor gets wet while passing through a cloud layer and is subject to evaporative cooling after entering dryer parts of the atmosphere. Below 28 km altitude, Jensen et al. (2016) found a mean systematic difference between the temperature measurements of the two RSs of 0.13 K.

For RS measurements performed at GRUAN sites, it is suggested that sites conduct dual sonde launches for at least 6 months when changing from one instrument type to another (GCOS-171 2013). However, analysis of data from dual sonde launches conducted at the GRUAN Lead Centre suggests that at least 200 dual flights over a period of one year are required to accurately assess the systematic difference between the two sonde-types (GCOS-171 2013). The number of dual sonde flights required may be site dependent and therefore, site specific analysis is likely required to determine the required number of dual flights at any site. Furthermore, it is possible that instrument biases at one site may not be the same in different atmospheric conditions at other sites, though this has not been extensively evaluated. Therefore, it would be ideal if all GRUAN sites could complete thorough RS intercomparisons by performing dual RS launches for at least 6 months prior to any instrument change. However, the costs of such a measurement campaign can be significant, preventing some stations from performing extensive dual launches.

In this study, we investigate the feasibility of quantifying the difference in biases of two instrument types by alternating between the two different instruments and then applying a statistical model to infer any systematic biases between the two instruments. For this study, we conduct the investigation by applying the developed statistical model to synthetic data sets, where persistence of weather conditions is a controllable parameter, that represent such interlaced RS flights. Specifically, we investigate (i) if a combination of interlaced measurements together with an appropriate statistical model can be used to estimate the differences in biases of two instrument types, and (ii) if so, how effective the approach is. This method, if feasible, could reduce the financial burden for sites seeking to manage such a transition, since an interlacing approach would not require additional

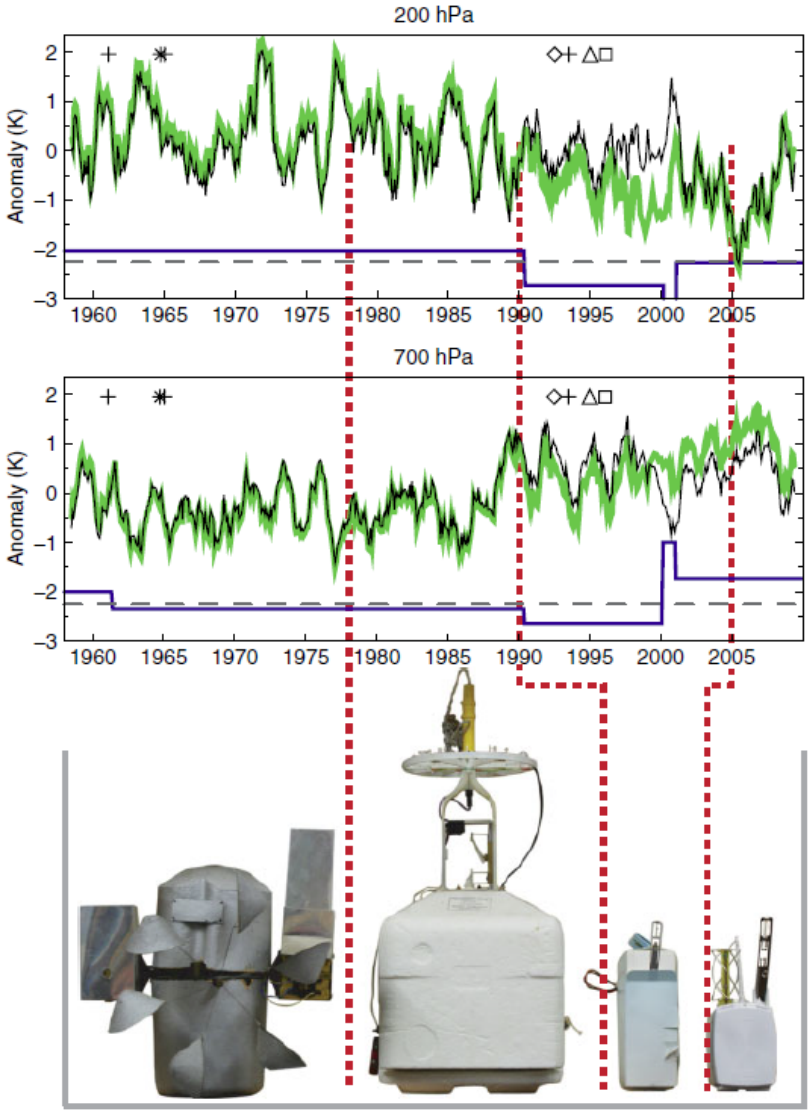


Figure 7.1.: Top two panels: Monthly temperature anomalies (smoothed with a 13-point running mean) during 1958-2009 from RS observations at Camborne, Cornwall, UK, at 200 hPa (near tropopause) and 700 hPa (lower-troposphere). Included are raw (black) and adjusted (green) RS temperature data from the Hadley Centre (HadAT). The smoothed difference series between the two shows the adjustments (offset by 2.25 K). Bottom panel: the four RS types used over this period (from left to right, with typical periods of operation): Phillips Mark IIb (1950-1970); Phillips MK3 (mid 1970s to early 1990s); Vaisala RS-80 (early 1990s to 2005/2006); and Vaisala RS-92 (since 2005/2006). Dates of RS changes are indicated by red dotted lines. Five other potential sources of inconsistencies in the data sets include: Change in the radiation correction procedure (cross); Change in the data cut-off (star); Change of pressure sensor (diamond); Change of wind equipment (triangle); Change of relative humidity sensor (square). Figure adapted from Thorne et al. (2011).

measurements above what is needed for normal daily operation.

7.1.3. Methodology

Background

Any modification of instrumentation might introduce a systematic change to the measurement time series. This change is typically assumed to be a constant difference (Δ) as a first order approximation, resulting from differences in the individual instrument biases, i.e. their systematic deviations from the true value. As the true value of the quantity being measured is unknown in practice, it is not possible to estimate each instrument's individual bias. It is possible, however, to estimate the difference $\Delta = \mathrm{Bias}_A - \mathrm{Bias}_B$ in biases Bias_A and Bias_B of instruments A and B. If temporally and spatially coincident measurements are made using instrument A and B (i.e. dual flights), this difference can be easily obtained: Consider some quantity of interest, e.g. air temperature (T), measured with instrument A and instrument B at the same location and time t. The bias of each instrument is the difference between the expectation value of the instrument's measurement and the unknown true value T_t:

$$\mathrm{Bias}(T_{t,A}) = E[T_{t,A}] - T_t \quad \text{and} \quad \mathrm{Bias}(T_{t,B}) = E[T_{t,B}] - T_t. \tag{7.1}$$

where $T_{t,A}$ and $T_{t,B}$ is the temperature at time t measured with instrument A and B, respectively. The difference in the instrumental bias is therefore:

$$\Delta_t = \mathrm{Bias}(T_{t,A}) - \mathrm{Bias}(T_{t,B}) = E[T_{t,A}] - E[T_{t,B}], \tag{7.2}$$

Consider now that $T_{t,B}$ differs from $T_{t,A}$ only by a constant offset Δ, i.e.:

$$T_{t,A} = T_{t,B} + \Delta \tag{7.3}$$

which is independent of the true value and thus the measurement time t. Under this assumption, an estimate for the stationary difference in biases can be obtained from N dual measurements according to:

$$\hat{\Delta} = \frac{1}{N}\sum_{t=1}^{N}(T_{t,A} - T_{t,B}) = \frac{1}{N}\sum_{t=1}^{N}\left((T_{t,A} - T_t) - (T_{t,B} - T_t)\right), \tag{7.4}$$

with $\hat{\Delta}$ denoting an estimate of the constant offset Δ. This equation applies even if the true value T_t is changing with time as it depends only on anomalies $T_{t,A/B} - T_t$. Under suitable conditions, the uncertainty (expressed in terms of standard deviation (SD)) of this estimate decreases with $\sqrt{N}$ and depends on the persistence (i.e. autocorrelation) of the time series (Wilks 2011).

A statistical model for interlaced measurements

As dual measurements, using both instrument types, require additional resources, and therefore inherent additional costs, estimating a systematic difference between the instru-

ments using interlaced measurements, i.e. using instrument A at odd days $t \in \{1, 3, 5, \ldots\}$ and instrument B at even days $t \in \{2, 4, 6, \ldots\}$ is explored in this study. Using this approach, at every time t only *one* measurement from *one* instrument is available, hence Eq. (7.4) is not applicable.

The underlying assumption for the approach outlined here to work is that the quantity of interest fluctuates around a smooth climatological signal (i.e. a seasonal cycle) and the fluctuations show a certain degree of persistence at the weather time scale, e.g. the fluctuations show a day to day dependence. For a typical difference in the biases between RSs this persistence (i.e. autocorrelation) is key to the idea of estimating a bias from interlaced measurements. The difference in the biases tested here is smaller than the day-to-day fluctuations themselves as it carries information from the measurement A to the measurement B.

In the following, a simplified model for air temperatures time series complying with the above mentioned assumptions is constructed. The true (unobserved) time series is represented by a smooth seasonal cycle with an auto-regressive process of first order (AR[1], e.g. Box and Jenkins 1976; Wilks 2011) added to the time series, i.e.:

$$T_t = \mu_0 + \mu_1 \sin\left(2\,\pi\,\frac{d_t}{365} - \frac{\pi}{2}\right) + \mu_2 \sin\left(2\,\pi\,\frac{2\,d_t}{365} - \frac{\pi}{2}\right) + \epsilon_t \tag{7.5}$$

$$\epsilon_t = a\,\epsilon_{t-1} + \eta_t\,, \tag{7.6}$$

with $d_t \in [1, \ldots, 365]$ giving the day in the year for date t, a is the autocorrelation coefficient which describes the degree of persistence in the time series, and $\eta_t \sim \mathcal{N}(0, \sigma^2)$ being the driving noise of the AR[1] process, selected randomly from a Gaussian distribution. The latter is taken to be Gaussian white noise with zero mean and variance σ^2. This is a well established model for the persistence of, e.g. daily air temperatures (e.g. Wilks 2011).

Pseudo-observations are now obtained from a realisation of T_t (Eq. (7.5)) with an instrument bias and random measurement noise added. Here, we aim for interlaced temperature measurements $T_{t,A}$ and $T_{t,B}$ from instruments A and B and thus add the instrument biases c_A and c_B, respectively, and independent Gaussian measurement uncertainties $\epsilon_{t,A} \sim \mathcal{N}(0, \sigma_A^2)$ and $\epsilon_{t,B} \sim \mathcal{N}(0, \sigma_B^2)$:

$$T_{t,A} = T_t + c_A + \epsilon_{t,A} \quad t \in t_A = \{1, 3, 5\ldots\} \quad \text{and} \tag{7.7}$$

$$T_{t,B} = T_t + c_B + \epsilon_{t,B} \quad t \in t_B = \{2, 4, 6\ldots\}\,. \tag{7.8}$$

For simplicity, we assume equal variances $\sigma_A^2 = \sigma_B^2$ for the measurement uncertainties. The continuous series of combined interlaced measurements $T_{t,AB}$ for $t \in \{1, 2, 3, \ldots\}$ is therefore:

$$T_{t,AB} = T_t + c_A\,\chi(t \in t_A) + c_B\,\chi(t \in t_b) + \epsilon_t, \tag{7.9}$$

with indicator function χ being 1 if t is a member of the set t_A or t_B, respectively, and 0 otherwise. Figure 7.2 shows an example of such a synthetic time series of interlaced measurements. This example is based on a simulated temperature time series using a realisation of an AR[1] process using an autocorrelation coefficient of $a = 0.5$ in Eq. (7.6), similar to the autocorrelation coefficient of RS measurements at 300 hPa above Lindenberg, Germany (cf. Section 7.1.3).

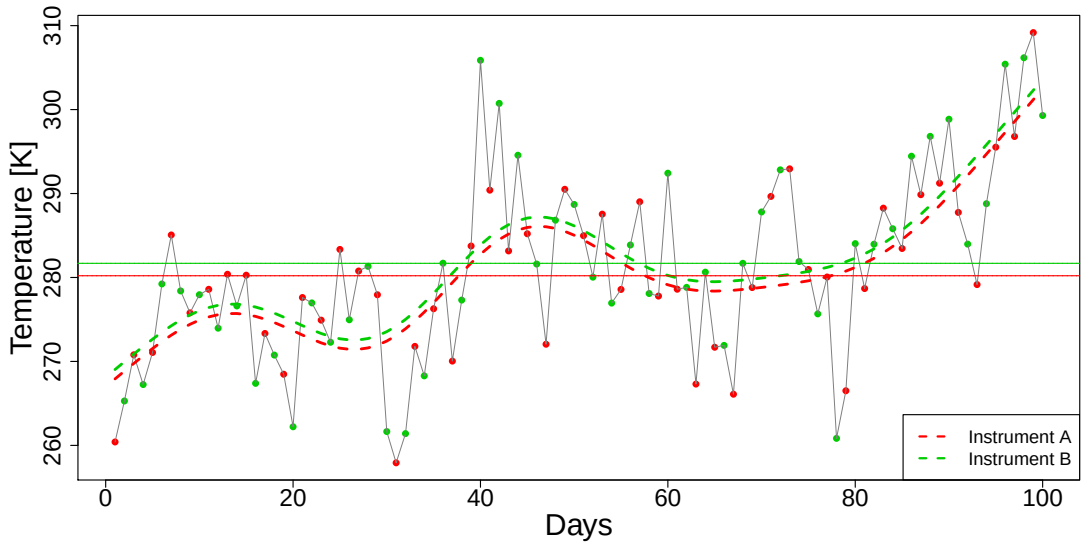


Figure 7.2.: Example time series for interlaced measurements of instrument A (red dots) and instrument B (green dots). Horizontal lines are the means of the measurements using instrument A (red) and instrument B (green). Smooth dashed lines (red for instrument A, green for instrument B) are spline estimates with the differences being an estimate for the differences in the instrument biases.

Estimating the difference in instrument biases

A direct approach to estimate the difference in instrument biases $\Delta = c_A - c_B$ is an estimation using the differences in means $\bar{T}_A$ and $\bar{T}_B$ of instrument A and B, respectively, over a common time period t_1 to t_2, i.e.

$$\widehat{\Delta}_{\text{mean}} = \bar{T}_A - \bar{T}_B \tag{7.10}$$

with

$$\bar{T}_A = \frac{1}{N_A} \sum_{t \geq t_1}^{t \leq t_2} T_{t,A} \quad \text{for} \quad t \in t_A \quad \text{and} \quad \bar{T}_B = \frac{1}{N_B} \sum_{t \geq t_1}^{t \leq t_2} T_{t,B} \quad \text{for} \quad t \in t_B \tag{7.11}$$

being the arithmetic means for the individual instruments; N_A and N_B are the number of measurements made by instrument A and B, respectively, in the given time period. The uncertainty on this estimate of the difference in instrument biases decreases with increasing N_A and N_B but depends also on the persistence of the underlying time series: larger persistence leads to larger uncertainties when calculating arithmetic means (e.g. von Storch and Zwiers 1999).

Here, we exploit the persistence and suggest an approach based on the estimation of a slowly varying signal common to both instruments. Imagine, for example, a smooth temperature time series in the absence of weather-induced noise. Measurements are then made of that signal using instrument A and this measurement series is represented by $s(t)$ and an additional measurement noise ϵ_t. Analogously, measurements of the same slowly varying signal are made using instrument B and can be represented by the same

$s(t)$ but with the difference in instrumental biases Δ and again measurement noise ϵ_t; i.e. $s(t) + \Delta + \epsilon_t$. A model for these interlaced measurements $T_{t,AB}$ is constructed using the indicator function χ:

$$\hat{T}_{t,AB} = s(t) + \Delta\,\chi(t \in t_B) + \epsilon_t\,. \tag{7.12}$$

For $t \in t_B$, the indicator function $\chi(t \in t_B)$ returns 1 and we obtain a measurement with instrument B, i.e. $\hat{T}_{t,B} = s(t) + \Delta + \epsilon_t$. For other time steps $t \in t_A$ the indicator function returns 0 and we obtain a measurement of instrument A, i.e. $\hat{T}_{t,A} = s(t) + \epsilon_t$, excluding the difference in instrumental bias Δ. The statistical model described in Eq. (7.12) belongs to the class of generalised additive models (GAMs, e.g. Chambers and Hastie 1992) and the smooth term s can be estimated using a smooth spline fit with degrees of freedom determined by generalised cross validation (Wood 2006). This functionality is implemented in the `R`-package `mgcv` (Wood 2006).

Simulation set-up

To investigate whether interlaced measurements, diagnosed using the methodology described above, can be used to estimate potential biases between instruments, we design a simulation study wherein an ensemble of synthetic upper-air temperature time series is generated using a stochastic process. For each member of the ensemble, interlaced measurements for two instruments are obtained by adding a systematic measurement uncertainty (i.e. bias) for each instrument plus some random measurement noise. As the instrument biases are known, their difference Δ is also known. The questions to be answered in this study are:

1. Can a combination of interlaced measurements, together with an adequate statistical model, be used to estimate the difference in instrument biases?
2. If so, how effective is this estimation compared to an approach requiring dual measurements?

An analysis of the 300 hPa temperatures measured by RSs at Lindenberg, Germany, forms the basis for this simulation study. After subtracting the seasonal cycle, the temperature anomalies show a variance of about $\sigma^2_{\text{anomalies}} = 10\,K^2$ and can be adequately described with an AR[1] process as in Eq. (7.6) with $a \sim 0.5$. To provide a realistic synthetic time series for analysis, we use driving Gaussian white noise $\eta \sim \mathcal{N}(0, \sigma_a^2)$ with variance $\sigma_a^2 = (1 - a^2)\,\sigma^2_{\text{anomalies}}$. This choice of σ_a^2 ensures that the anomaly variance is fixed at $\sigma^2_{\text{anomalies}} = 10\,K^2$ independent of the value of a. This is necessary as we vary the persistence parameter (i.e. the autocorrelation coefficient) $a \in (0, 1)$ to study time series with different persistence but identical anomaly variance.

The synthetic temperature series is generated using Eq (7.9) that includes a seasonal cycle and a realisation of an AR[1] process. The instrument biases in Eq. (7.9), are prescribed at $c_A = -0.1\,\text{K}$ and $c_B = 0.2\,\text{K}$ and are added to the time series together with a measurement uncertainty being specified as Gaussian white noise $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The resulting two time series for instruments A and B are combined to a) a synthetic time series of dual measurements, and b) an interlaced observational counterpart. The difference in instrument biases between both time series is prescribed as $\Delta = c_A - c_B = -0.1 - 0.2 = -0.3\,K$. To investigate the influence of (i) persistence in the temperature series, (ii) measurement noise, and (iii) number of measurements on our ability to estimate

the difference in biases between two instruments, the following parameters are prescribed and controlled in our study:

persistence of the time series $a \in \{0.5, 0.7, 0.8, 0.9, 0.95, 0.99\}$

number of measurements $N \in \{50, 100, 250, 500, 1000, 2000, 3000\}$,

leading to $6 \times 7 = 42$ combinations, i.e. 42 synthetic time series to be analysed. The instrument noise is fixed at $\sigma^2 \in 0.1$. To generate a synthetic time series for a given a, N and σ, the following steps were taken:

1. Generate a time series of length N consisting of an annual cycle and a realisation of an AR[1] process as described above.
2. Add an offset of -0.1 K (instrument bias of instrument A) and Gaussian noise with variance $\sigma^2 = 0.1$ to produce a synthetic time series for instrument A.
3. Add an offset of 0.2 K (instrument bias of instrument B) and Gaussian noise with variance $\sigma^2 = 0.1$ to produce a synthetic time series for instrument B.
4. Select measurements from A for odd days and from B for even days to generate an interlaced time series.
5. Repeat steps 1 to 4 many times (e.g. $M = 1000$) to generate 1000 synthetic time series to derive statistically robust estimates of $\hat{\Delta}$.

The difference in instrument biases is then estimated based on

1. the calculated mean values of N dual measurements (Eq. (7.10)), i.e. N measurements for A and N measurements for B made simultaneously, and
2. results from the statistical model (Eq. (7.12) using the time series of N interlaced measurement, i.e. $N/2$ measurements for A and $N/2$ measurements for B.

7.1.4. Results

The box plots in Fig. 7.3 summarise the distribution of $M = 1000$ bias estimates $\widehat{\Delta}$ for a varying numbers of interlaced flights N. The upper panel of Fig. 7.3 is based on the simulated temperature time series with an AR[1] coefficient $a = 0.5$, being similar to the autocorrelation coefficient found for temperature measurements at 300 hPa above Lindenberg. The middle and bottom rows are examples for stronger persistence, i.e. $a = 0.8$ and $a = 0.9$, respectively. All panels show that the spread in the estimated difference in bias between instruments A and B ($\widehat{\Delta}$) converges towards the true value ($\Delta = -0.3$) for increasing N in all cases. The rate at which this converges with increasing N depends on the persistence (i.e. autocorrelation) in the underlying time series. Weak persistence (small a) leads to slower convergence (Fig. 7.3, top row), while strong persistence (a approaching 1) shows faster convergence.

The SD of $\widehat{\Delta}$ (see Fig. 7.4), representing the uncertainty with which the difference in the bias between instruments A and B can be estimated, depends on the number of interlaced flights and on the AR[1] coefficient a (coloured lines in Fig. 7.4). The SD can be used to construct asymptotic confidence intervals for the estimates using the standard normal assumption (e.g. Wilks 2011, Chapt. 5), i.e. for a 95% confidence interval, the estimated bias needs to be within 1.96 times the SD. For all a, the SD decreases with increasing N; however, the SD is generally larger for weak persistence (small $a \in (0, 1)$) and smaller for

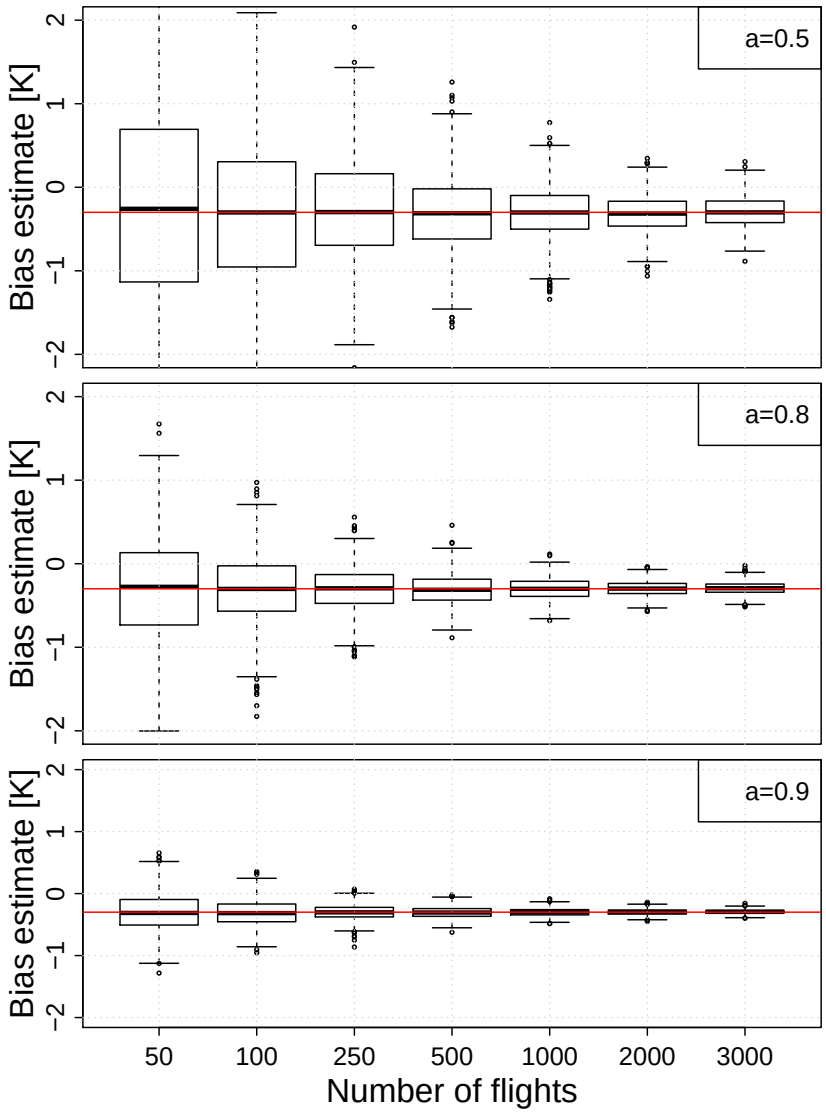


Figure 7.3.: Box and whisker plots of bias estimates ($\widehat{\Delta}$) against number of interlaced flights N (50 flights means 25 flights of instrument A and 25 flights of instrument B) as derived from $M = 1000$ simulations using an autocorrelation coefficient of $a = 0.5$ (top), $a = 0.8$ (middle) and $a = 0.9$ (bottom) and a measurement noise of $\sigma^2 = 0.1$. The boxes show the inter-quartile range. The upper and lower whiskers represent the maximum (excluding outliers) and minimum (excluding outliers). Suspected outliers are shown as dots and are located outside the fences (whiskers) of the boxplot (e.g: outside 1.5 times the interquartile range above the upper quartile and below the lower quartile). The true difference in biases $\Delta = -0.3\,K$ is marked with a red line.

strong persistent (large $a \in (0, 1)$).

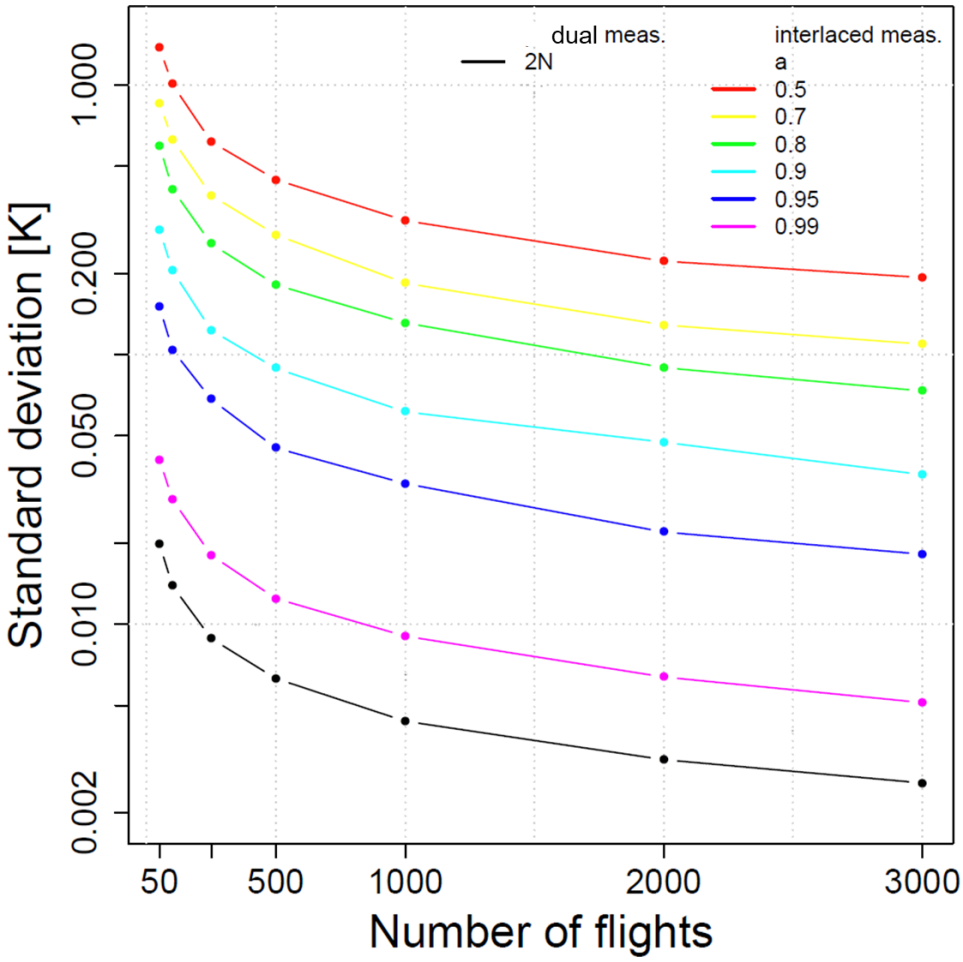


Figure 7.4.: SD of $\widehat{\Delta}$ against number of flights N for different AR[1] coefficients a. The black solid line represents the reference experiment with dual flights of instruments A and B, i.e. 2 N measurements. To compare the results from the dual flights (black solid line) with the results obtained from interlaced flights, the number of dual flights have to be doubled. Note the logarithmic vertical scale.

The synthetic time series of dual flights performed with instrument A and B simultaneously at N times (i.e. $2\,N$ measurements, solid black line in Fig. 7.4) provides the most reliable estimate of the biases between the instruments, i.e. the SD is smallest for any N. To provide a robust comparison of the results from the dual flights to the results from N interlaced measurements, the results from the dual flights need to be compared to the results of doubled N interlaced flights. For a time series with an autocorrelation coefficient of $a = 0.5$, at least 2000 days of consecutive interlaced daily measurements would be required to estimate the difference in instrument's biases with a SD of 0.22 K. Consider the following example, a station operator seeks to detect the difference in bias between two RSs in a temperature time series showing an autocorrelation coefficient of 0.95. The station operator requires a SD of $\hat{\Delta} \leq 0.05\,\mathrm{K}$ which leads to a 95% confidence interval of about 0.1 K ($\approx 0.05 * 1.96$), then, from Fig. 7.4 it can be inferred that 500

interlaced measurements are required to achieve this. Furthermore we conclude that, if an operator has a given amount of two types of RSs available from which the difference in instrument biases needs to be estimated, it is clear from Fig. 7.4 that dual flights result in better estimates (i.e. smaller SD in Fig. 7.4) than to interlace the instrument types from one day to the next. The results presented here (from dual and interlaced flights) also depend on the variance of the signal; for a higher measurement noise, the number of required days will increase and vice versa (not shown).

The results indicate that for typical difference in biases between RS types, the presented method on interlaced measurements is unlikely to provide a robust estimate of the difference in biases for a reasonable length of the measurement period (reasonable is considered as 2 years here). That said, there might be cases of larger instrument biases and/or larger persistence where the interlaced method could provide an alternative method to dual measurements, requiring fewer resources. This, however, is outside the scope of this study which focuses on RS temperature measurements.

7.1.5. Conclusions

We have used synthetic time series representing temperature measurements to investigate the possibility of using interlaced measurements performed with two different instruments types together with Generalised Additive Models to obtain an estimate of the difference in the bias between the two instrument types. Performing dual RS flights with both instrument types is costly, and therefore we investigated the feasibility of using interlaced flights to obtain an estimate of the difference in the bias. This would be more sustainable and less costly. Information about typically small differences in instrument biases can be obtained from non-simultaneous measurements using a persistence assumption, i.e. some information of some day's measurement is carried over to the next day. As atmospheric temperatures tend to be autocorrelated in time (e.g. Wilks 2011; Maraun et al. 2004), the persistence assumption is justifiable. However, the strength of the autocorrelation depends, in part, on the geographical location of the measurement site and on altitude. Here, we investigated how a statistical approach to estimate the difference between two instrument biases is affected by the persistence of a time series.

The results presented here indicate that while it is in principle possible to estimate the difference between two instrument's biases from interlaced measurements, the number of interlaced flights required to obtain a satisfying accuracy is very large for reasonable values of the autocorrelation coefficient. Strongly autocorrelated signals require fewer data for an accurate estimate of the difference in biases and therefore fewer interlaced flights, than time series with low autocorrelation. The results show that for very strong persistence (e.g. an AR[1] coefficient of 0.99) about twice the number of measurements is needed compared to parallel measurements to obtain a comparable uncertainty in estimates for interlaced measurements. Hence, the described approach may be used for measurements with very strong persistence or where the costs for sufficient parallel measurements exceeds the costs for sufficient interlaced measurements to confidently infer the difference in the instrument bias. However, if, for example, it would be possible to derive a robust estimate of the difference in instrument biases from interlaced measurements in some reasonable time period (e.g. 2 years) and even if this period was more than 2 or 3 times longer than would be required from a dual measurement strategy to achieve the same level of

confidence, the interlacing approach would present a cost saving over an approach that would start with dual flights and then continue with flights using only the new instrument.

Acknowledgements We would like to thank the NOAA[15] GCOS office, through the Meteorological Service of New Zealand Limited, for supporting this research. HR acknowledges support from the Freie Universität Berlin within the Excellence Initiative of the German Research Foundation. We would also like to thank Fabio Madonna and Alessandro Fasso for helpful discussion around the alternative approach of interlaced measurements. We thank the GRUAN for providing the data used in this publication. The data were sourced from
ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/. The authors confirm that these data have been used in a manner consistent with the GRUAN data use policy, as articulated in the GRUAN Guide, and have not been used for commercial gain.

[15]National Oceanic and Atmospheric Administration

Bibliography

Box, G.E.P. and G.M. Jenkins (1976). *Time Series Analysis: forecasting and control*. New Jersey: Prentice Hall.

Chambers, J.M. and T.H. Hastie, eds. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole. Chap. 7. Generalized additive models.

GCOS-171 (2013). *The GCOS Reference Upper-Air Network (GRUAN) Guide*. WIGOS TECHNICAL REPORT. World Meteorological Organization.

Haimberger, L., C. Tavolato, and S. Sperka (2012). "Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations". In: *J. Climate* 25, pp. 8108–3131. DOI: : 10.1175/JCLI-D-11-00668.1.

Jensen, M.P., D. Holdridge, P. Survo, R. Lehtinen, S. Baxter, T. Toto, and K.L. Johnson (2016). "Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains Site". In: *Atmos. Meas. Tech.* 9, pp. 3115–3129. DOI: 10.5194/amt-9-3115-2016.

Luers, J.K. and R.E. Eskridge (1998). "Use of radiosonde temperature data in climate studies". In: *Journal of Climate* 11, pp. 1002–1019.

Maraun, D., H.W. Rust, and J. Timmer (2004). "Tempting long-memory - on the interpretation of DFA results". In: *Nonlin. Proc. Geophys.* 11, pp. 495–503.

Randel, W.J. and F. Wu (2006). "Biases in Stratospheric and Tropospheric Temperature Trends Derived from Historical Radiosonde Data". In: *Journal of Climate* 19, pp. 2094–2104.

Rust, H.W., O. Mestre, and V.K.C. Venema (2008). "Fewer jumps, less memory: Homogenized temperature records and long memory". In: *J. Geophys. Res.* 113.D19110.

Saha, S., S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, and et al. (2010). "The NCEP Climate Forecast System Reanalysis". In: *Bull. Amer. Meteor. Soc.* 91.8, pp. 1015–1057. ISSN: 1520-0477. DOI: 10.1175/2010bams3001.1. URL: http://dx.doi.org/10.1175/2010BAMS3001.1.

Seidel, D.J. and M. Free (2006). "Measurement Requirements for Climate Monitoring of Upper-Air Temperature Derived from Reanalysis Data". In: *J. Climate* 19, pp. 854–871.

Sherwood, S.C., J.R. Lanzante, and C.L. Meyer (2005). "Radiosonde Daytime Biases and Late 20th Century Warming". In: *Science* 309, pp. 1556–1559.

Steinbrecht, W., H. Claude, F. Schönenborn, U. Leiterer, H. Dier, and E. Lanzinger (2008). "Pressure and Temperature Differences between Vaisala RS80 and RS92 Radiosonde Systems". In: *J. Atmos. Oceanic Technol.* DOI: 10.1175/2007JTECHA999.1.

Thorne, P.W., J.R. Lanzante, T.C. Peterson, D.J Seidel, and K.P. Shine (2011). "Tropospheric temperature trends: history of an ongoing controversy". In: *WIREs Climate Change* 2.

von Storch, H. and F.W. Zwiers (1999). *Statistical analysis in Climate Research*. Cambridge University Press.

Wilks, D.S. (2011). *Statistical methods in the atmospheric sciences*. 3rd. San Diego, CA: Academic Press.

Wood, S.N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

8. Conclusion and outlook

This thesis presents approaches to improve the upper-air observational record at both global and local scales. These methods have been published in peer-reviewed reports, and peer-reviewed journal articles (Tradowsky 2015; Tradowsky et al. 2016; Tradowsky 2016; Tradowsky et al. 2017), as discussion papers (Kremser et al. 2018a, under review), or have been submitted for publication (Tradowsky et al. 2018, submitted). These publications form the foundation for this thesis and are, in parts, reprinted in the previous chapters. Since each chapter includes a detailed discussion of the results, only a brief overview is given here before a general discussion of the thesis outcomes.

Ground-based and space-based measurements of the upper-air are extensively used in numerical weather prediction (NWP). Radiosonde (RS) and radio occultation (RO) profiles anchor the temperature in data assimilation systems of NWP[16] models. Consistency between these two observation types is essential as inconsistencies between anchor observations can cause spurious features in the NWP model. A challenge, defined in the first scientific question in Chapter 1, is ensuring that the homogeneity of one observation type can be transferred to measurements made with sensors that provide less homogeneous data. This challenge is addressed by developing a novel method to correct heterogeneous temperature biases in RS measurements using spatially homogeneous and long-term stable RO data record as a reference. Usually RS temperature bias corrections, which are based on RO, are calculated using direct collocations between RO and RS profiles. A direct collocation exists if profiles from each sensor are measured within a given time interval and distance of each other, e.g. 3 hours and 300 km. As the atmosphere is not homogeneous over such a separation in space and time, applying direct collocation leads to large uncertainties in the temperature bias corrections. To compensate for this, a large sample size is required and, for this reason, the temperature biases are usually calculated on the basis of RS types, averaging all profiles globally that use the same RS types.

Here, a method is presented that uses a new approach to calculate RS temperature biases which does not require direct collocations. Rather, the UK Met Office NWP model is used as a transfer standard and averaged observation minus background departures of all RO profiles measured within 500 km of an upper-air site are compared with averaged RS observation minus background profiles for the selected site. Assuming that the model background error is constant over the separation distance of up to 500 km, the model background becomes extraneous and the double difference estimates the difference between RO and RS profile. The observation minus background for RO measurements is calculated in bending angle space and a tangent linear retrieval has been developed to propagate the bending angle departures into dry temperature departures (Tradowsky et al. 2017). This approach significantly reduces the *a priori* required in the retrieval since no smooth climatological bending angle is required unlike in the conventional RO retrieval. Dry temperatures are approximately equal to physical temperatures in those parts of the atmosphere where water vapour is negligible - primarily above the tropopause. The lowest dry level in the atmosphere has been determined for every RO profile indi-

[16] All abbreviations can be found in the Glossary, Appendix B

vidually. The use of a transfer standard significantly reduces the standard deviation and therefore allows the calculation of RS temperature biases on a station-by-station basis. This also ensures that bias profiles for all upper-air sites can be calculated, even if no RS type is supplied.

Remote sensing methods, in general, rely on a retrieval that calculates a physical variable from the raw measurement. Typically some *a priori* is required for the retrieval. For RO, the raw measurement is the time delay (phase shift) of a radio signal which passes through the atmosphere. During a setting occultation event the radio signal penetrates deeper and deeper in the atmosphere and from the measurement of the time delay at different impact heights above ground a profile of time delays is obtained. In a first retrieval step, the bending angle profile is calculated from the profile of time delay measurements. As the bending angle profile is noisy at the higher altitudes, it is blended with a smooth climatological bending angle profile in the conventional RO retrieval. The use of *a priori* usually implies structural uncertainty, which is caused by the lack of knowledge about the ideal *a priori*, whose choice effects the results of the retrieval. The spread of retrieved dry temperature profiles from different retrieval centres, which use different climatological bending angle profiles and different optimisation algorithms, gives an estimate of the structural uncertainty.

The work presented in this thesis has considered the possibility of reducing the influence of the *a priori* in the RO retrieval to address the second scientific question posed in Chapter 1. The tangent linear RO retrieval can be used to propagate bending angle differences into dry temperature differences. It does not require smoothing with a climatological bending angle profile. Instead, this retrieval allows the altitude range of interest to be chosen depending on the purpose of the retrieval. The bending angle departures above this chosen threshold are set to zero, implying that the observation and background are identical above the threshold. This method requires significantly less *a priori* information than the conventional RO retrieval.

The tangent linear RO retrieval, as any retrieval, has structural uncertainty, which can be estimated by varying the upper threshold above which the departures are assumed to be zero. This is addressing the third scientific question which build the basis of this thesis. An upper limit of the structural uncertainty estimates is calculated in Chapter 3 as 0.4 K at 10 hPa, which decreases to values below 0.2 K at pressures above 30 hPa and agrees well with the structural uncertainty estimated for conventional RO retrievals (Ho et al. 2009; Ho et al. 2012; Steiner et al. 2013). The presented method to estimate uncertainties in the tangent linear RO retrieval, however, enables estimating structural uncertainties depending on the desired application, and the structural uncertainty can be significantly smaller than the upper limit of uncertain- ties given above for specific applications. For example, the highest structural uncertainties per region is between 0.078 K and 0.192 K for comparison between GRUAN and RO data (see Chapter 4).

An ongoing cooperation with the UK Met Office, ECMWF and the EUMETSAT Radio Occultation Meteorology Satellite Application Facility is planned to further investigate RS temperature bias corrections, the tangent linear RO retrieval and the structural uncertainty in this retrieval.

As the tangent linear RO retrieval is able to propagate bending angle differences into dry temperature differences, this retrieval could help to detect climate changes at lower

stratospheric levels. To date, the time series of RO measurements may be too short for a statistically robust study of trends in climate, i.e. the first pilot RO satellite was launched in 1995, and operational data are available only since 2006. However, RO satellites are intended to be maintained into the foreseeable future and thus, the ability to propagate changes detected in bending angle space into dry temperature space might prove valuable and should be further investigated. Detailed investigation of this approach is however required first.

While the double differencing method described above was developed to correct RS temperature biases prior to the assimilation of the RS temperatures into the NWP model, it was also applied to investigate how the GRUAN RS data product compares with RO data on a station-by-station basis to address the fourth scientific question from Chapter 1. GRUAN RS profiles and RO data rely on entirely different measurement techniques and they are both used as reference in comparisons with other sensors. Taking the GRUAN uncertainty estimate into account, as well as the radio occultation structural and sampling uncertainty, the temperature departures, which are again calculated with respect to UK Met Office model background fields, are compared to assess whether the measurement types are consistent or in agreement (Immler et al. 2010).

At those pressure levels where a representative number of profiles is available for comparison, the RO and GRUAN profiles are mostly consistent or in agreement. A common exception is found in the lower stratosphere during the night. The reason for these inconsistent values is currently unknown and suggests the need for further research, which could be undertaken as a cooperation between the GRUAN and RO communities. An updated GRUAN RS92 data product (version 3) is currently in development and it would be worthwhile to repeat the analysis presented here with the updated GRUAN data product. Furthermore, a large number of RS92 profiles submitted to GRUAN have not yet been processed due to a change in the ground station software file format. As soon as these profiles become available, the sample size of GRUAN profiles to be used in the comparison will significantly increase and enable a more detailed analysis of differences at high altitudes, where the sample size is currently insufficient at several sites.

Following the development of methods to improve the upper-air observational record on a global scale, and for a subset of GRUAN sites, a method to combine temperature measurements made at Lauder and Invercargill in the lower South Island of New Zealand into a temperature Site Atmospheric State Best Estimate (SASBE) for Lauder was demonstrated addressing the fifth research question of this thesis, see Chapter 1. The measurement sites at Lauder (inland) and Invercargill (coastal) are about 180 km apart and constitute a distributed GRUAN site. The SASBE combines RS measurements from Lauder and Invercargill and a diurnal temperature cycle from ERA5 reanalysis towards which the SASBE relaxes with increasing time separation to a measurement. A method to use the temperature measurements at Invercargill as predictors of temperatures at Lauder is presented for the purpose of implementation within GRUAN. The uncertainty added by the transfer algorithm is taken into account as required by GCOS-170 (2013). A publication describing the methodology of the construction of the temperature SASBE has been submitted to *Earth System Science Data Discussions* (Tradowsky et al. 2018, submitted). The temperature SASBE is available at 16 pressure levels at hourly resolution for the years 1997 to 2012. Every temperature value in the SASBE has an associated uncertainty which has been propagated through the calculations. It is demonstrated that

including Invercargill data in the SASBE improves the representativeness of the SASBE with respect to the temperatures measured at Lauder. However, further research into the uncertainty resulting from the transfer of data is suggested to reflect this improvement in the overall uncertainty on the best estimate-temperature.

While the GRUAN data product for the Vaisala RS92 RS has not been used in the SASBE presented here, the method for the construction of such a temperature SASBE at a distributed GRUAN site has been demonstrated for implementation across GRUAN. The absence of GRUAN data in this demonstration in no way compromises the validity of the method. However, in a future project, the method could be applied to several distributed GRUAN sites using temperature profiles and their associated uncertainties that have been GRUAN processed.

As measurements of upper-air essential climate variables (ECVs) are extremely sparse in the Southern Hemisphere, this SASBE may be especially valuable to the scientific community. Possible applications include the use of the SASBE as *a priori* information in the retrieval of other variables in Lauder, e.g. for the ozone lidar retrieval. The SASBE may also be used to validate models, which may be especially informative as the Lauder RSs were not assimilated into NWP models prior to 2016. However, the Invercargill RSs have been assimilated into NWP models and researchers will need to be cognizant of this before using the SASBE for any model validation.

The upper-air facility at Lauder is regularly used for validating space-based sensors given the scarcity of reference-quality measurements in the Southern Hemisphere. Thus, it is expected that the SASBE will be used for the validation of space-based instruments. This can be done either on the basis of validating derived products i.e. the retrieval of temperature from a satellite instrument is compared with the temperature SASBE, or in native measurement space, e.g. in radiance measurement space. Performing the comparison in radiance space has the advantage that no retrieval is required to calculate the ECV from a raw measurement, which usually is an ill-posed problem, i.e. several equally likely solutions exist. With a radiative transfer model, some additional information about the temperature above the highest altitude of the SASBE, and given knowledge about the profiles of some other ECVs, the temperature SASBE can be used to calculate top-of-the-atmosphere (TOA) radiances with uncertainty estimates that are based on the temperature uncertainties as estimated in the SASBE. Ideally, a water vapour SASBE may be used together with the temperature SASBE in radiative transfer calculations. Thus, a radiance-space validation of space-based radiometers using SASBEs would be conceivable.

A study has been planned and an associated funding proposal has been submitted to evaluate the extend to which the energy balance is influenced by changes in the highly uncertain surface emissivity at wavelengths above 15.4 μm. To this end SASBEs of temperature and water vapour above Antarctica will be required. These SASBEs can be used to constrain a radiative transfer model to calculate the top-of-the-atmosphere (TOA) radiances. The surface emissivities in the radiative transfer model can then be varied between reasonable thresholds to analyse to what extend reasonable changes in the far-infrared surface emissivity effect the TOA radiances. Once the extent of our current lack in knowledge about far-infrared surface emissivities on the Antarctic energy budget is clarified, further steps to close the knowledge gap can be planned. This project will be pursued by

the author of this thesis once funding is secured.

Within GRUAN, every change in instrumentation has to be carefully managed to ensure the long-term stability of the time series. Currently the change from the Vaisala RS92 to the Vaisala RS41 radiosonde is underway and dual launches are being conducted at GRUAN sites to quantify the possible biases between these two instruments. Careful assessment of the measurement biases between the two instruments is required to maintain the long-term stability of the GRUAN RS data record. Even small step changes in the time series of e.g. temperature, may prevent the small signal of changes in upper-air climate to be detected. However, as dual launches are costly, a study has been performed to analyse if it is feasible to estimate the difference between biases of two instruments from alternating measurements with the two instrument types, i.e. alternating the instrument type from day to day (sixth scientific question) While this interlacing approach is promising for highly autocorrelated variables, it proves less valuable for estimating of biases between RSs at GRUAN sites. However, it may be feasible to apply the method to other instrumentation, e.g. space-based instruments. On a satellite, where the energy supply is limited, it might prove valuable to alternate between two instruments for the purpose of comparison or inter-calibration. Depending on the parameter, the persistence of the measured quantity may be high enough to estimate the bias between the instruments from these alternating measurements.

Several scientific questions have been raised and were addressed in this thesis. Three methodologies to improve the upper-air observational record from the global to the local scale have been presented and discussed. On the global scale, improvements to aid the NWP community and the RO community were pursued. The results have been presented at conferences and were discussed in the scientific community before they were published in the peer-reviewed literature. A close collaboration between the author of this thesis, the Satellite Application group at the UK Met Office, ECMWF and the EUMETSAT Radio Occultation Meteorology Satellite Application Facility is ongoing and a future project is planned to answer questions that emerged based on the outcomes of the cooperation. Such a project would include: (i) a detailed study of effects of applying the new method to correct RS biases in the Met Office NWP model, including a study of regional effects and stratospheric effects, (ii) analysis of the structural uncertainty in tangent linear RO retrievals, e.g. dependence on latitude and area, and (iii) further comparison of GRUAN and RO using the double differencing technique, e.g. a study of the RS41 data product and inclusion of more stations.

A temperature SASBE for Lauder, New Zealand, was developed in the interests of the GRUAN community and to serve the satellite validation and modelling communities. The methodology is described in detail and the high temporal resolution data product is publicly available. It has been demonstrated how the measurements of temperature from Invercargill can be used to estimate the temperature above Lauder. Such a transfer of information may be part of a best-estimate data product or may be conducted to obtain *a priori* information for retrievals by other instruments which are spatially separated. However, the method may not be used to transfer information to another location prior to the assimilation into a forecast model, as this does not lead to benefits. Instead, the forecast model should process the measurements together with their location information in the assimilation system.

In addition to the theoretical work, the author has actively participated in reference-quality measurements at the GRUAN site at Lauder and received a certificate of appreciation for her support of the Southern Hemisphere ADditional OZonesondes (SHADOZ) program. She also organised and conducted an ozonesonde campaign at Palau in the Western Pacific Warm Pool. Gaining a detailed understanding of the processes involved in taking measurements provides a deeper understanding of difficulties in the measurements and the sources of uncertainties in measured data. Within atmospheric sciences, ongoing efforts should be invested for improved estimation, propagation and communication of uncertainties. Without a robust assessment of uncertainties, it is impossible to determine the level of confidence in a result. The close collaboration between the GRUAN and RO community has benefited both scientific fields and such a close exchange between different disciplines is encouraged in the interest of scientific advancement.

A. Contributions to the scientific community

During this thesis, I have supported the international scientific community and have share my knowledge with peers. Examples of these contributions are providing reviews for scientific journals or presenting GRUAN at international conferences. In this chapter an overview of the services provided to the scientific community is given.

A.1. Supporting the effort to receive WMO station identifiers for all GRUAN sites

Many GRUAN sites have traditionally been used for research purposes only, rather than making the measurements available to the NWP community. Though, the measurements made at GRUAN sites can be valuable for NWP centres in several ways, i.e.:

1. Direct effect via assimilation: Given a GRUAN site is providing a high quantity of near real-time data, or data from a remote location, assimilation of the measurements could positively affect NWPs. The near real-time data submitted by GRUAN sites to the Global Telecommunication System (GTS), which distributes meteorological data, are not GRUAN data products, but rather the manufacturer provided output from a measurement system, which is consistent with the data submitted from operational sites (Ingleby and Edwards 2015).
2. Indirect effect via validation: An NWP centre might deliberately decide not to assimilate the data from GRUAN sites, but rather use the GRUAN data products to validate the model. As GRUAN data products are corrected for all known biases and include an estimate of the uncertainty on every value, the value they offer to validate NWP models is far superior to that of manufacturer provided data products.

A prerequisite for any use of measurements performed at a GRUAN site by an NWP centre is knowledge about the existence of the station. NWP centres retrieve the measurements to be assimilated into the model from the GTS and, given a station is not submitting to the GTS, the centres might not be aware of its existence. Furthermore every measurement facility requires a WMO station identifier, which is a unique 5-digit number. During 2016, several GRUAN stations[17] were lacking assignment of a WMO station identifier, preventing the use of measurements made at GRUAN sites in the NWP. While working as a visiting scientist at the Met Office during 2015, I became aware of this issue and took on the challenge of mitigating the problem. While I am not in the position to assign station identifiers myself, I was (i) helping to clarify the problem, (ii) providing support to the GRUAN station representatives to find the responsible people at the National Weather Service, (iii) persistently discussing the issue with the involved parties. By mid-2017 progress was made at most GRUAN sites.

[17] Barrow, Beltsville, Boulder, Manus, Nauru, Potenza, La Reunion, Southern Great Plains

Several steps, as described below, are required before a GRUAN site is able to submit data to the GTS and can be used in NWP:

1. Assignment of a WMO station identifier. The assignment is typically done by a member of the national weather service of the country hosting the GRUAN site, e.g. the WMO focal point.
2. The GRUAN site needs to be added to the OSCAR Tool (https://oscar.wmo.int/surface/index.html#/), where information about the station including station identifier and location are summarised. There seems to be confusion about who is responsible to do this, possibly the WMO focal point of the respective country.
3. The GRUAN site needs to start submitting data to the GTS. This might involve changes in the ground station software to set up automatic submission of the data. For example, a BUFR report needs to be produced automatically after the balloon burst of a RS and the data need to be submitted to the GTS.
4. Finally an NWP centre should check that the data are arriving and can be used. This is important as otherwise the data might be automatically rejected by the data assimilation system and the effort of taking steps 1.-3. was wasted.

The following table gives an overview of the GRUAN stations including WMO station identifier. If problems regarding the station identifier or the submission of RS data to the GTS are known to myself, the issue is described as a comment. The data are up to date as of the 4th July 2017.

Table A.1.: List of GRUAN upper-air sites. The official WMO station identifier (ID) is given where available, even if the station ID is currently still missing in the WMO OSCAR Surface system (https://oscar.wmo.int/surface/index.html. If there are any issues with the receiving of a WMO ID, or with the submission of the data to the GTS, a comment explains the problem. Latitude and longitude are taken from the GRUAN website (https://www.gruan.org/network/sites/) and do not always agree with the exact station position in the GRUAN data product for the Vaisala RS92. Table up to date as of the 4th July 2017.

Station Name, Country	WMO ID	Latitude	Longitude	Altitude	Comment
Alice Springs (ALC), Australia	94326	-23.70°	133.87°	546 m	
Barrow (BAR), USA	70027	71.32°	-156.61°	8 m	Donna Holdridge organised a station ID and started to submit BUFR data to the GTS.
Beltsville (BEL), USA	-	39.05°	-76.88°	53 m	Belay Demoz is trying to get a station ID or use the station ID of the co-located Sterling facility
Boulder (BOU), USA	72471	39.95°	-105.20°	1743 m	Dale Hurst got the ID organised but it is not in OSCAR yet. Problems with BUFR submission.
Cabauw/De Bilt (CAB), The Netherlands	06260	51.97°	4.92°	1 m	
Darwin (DAR), Australia	94120	-12.43°	130.89°	30 m	
Davis (DVS), Antarctica (operated by Australia)	89571	-68.58°	77.97°	18 m	
Dolgoprudny (DLG), Russia	27612	55.75°	37.62°	185 m	
Lauder (LAU), New Zealand	93817	-45.05°	169.68°	370 m	
Lindenberg (LIN), Germany	10393	52.21°	14.12°	98 m	
Macquarie Island (MAQ), Australia	94998	-54.62°	158.85°	6 m	

Continued on next page

Table A.1 – continued from previous page

Station Name, Country	WMO ID	Latitude	Longitude	Altitude	Comment
Manus (MAN), Papua New Guinea (operated by USA)	-	-2.06°	147.42°	6 m	inactive site without an ID
Melbourne (MEL), Australia	94866	-37.81°	144.96°	113 m	
Nauru (NAU), Nauru (operated by USA)	-	-0.52°	166.92°	7 m	inactive site without an ID
Ny Ålesund (NYA), Norway (operated by Germany)	01004	78.92°	11.92°	5 m	
Payerne (PAY), Switzerland	06610	46.81°	6.95°	491 m	
Potenza (POT), Italy	-	40.60°	15.72°	720 m	Fabio Madonna is trying to get a station ID, seems to be challenging.
La Reunion (REU), France	-	55.38°	-21.08°	2200 m	Stefanie Evan is evaluating the option to use the station ID from the co-located operational site.
Southern Great Plains (SGP), USA	74646	36.60°	-97.49°	320 m	Donna Holdridge invested much effort to get the ID, but still problems with BUFR transmission
Paris (SIR), France	07151	48.7°	2.2°	156 m	
Singapore (SNG), Singapore	48698	1.30°	103.80°	21 m	
Sodankylä (SOD) Finland	02836	67.37°	26.63°	179 m	
Tateno (TAT), Japan	47646	36.06°	140.13°	25 m	
Tenerife (TEN), Spain (operated by USA)	60018	28.32°	-16.38°	115 m	
Xilin Hot (XIL), China	54102	43.95°	116.12°	1013 m	

A.2. Peer-reviews

The anonymous peer-review system builds the backbone of the scientific literature. The anonymous review of papers that are submitted to a scientific journal by peers is essential to ensure the high-quality of scientific literature. Providing peer-reviews is a voluntary work that can take a considerable amount of time, while not leading to any king of recognition. Though, it is essential and it is a service provided by scientists to ensure high quality scientific publications.

During her PhD project the author of this thesis provided three reviews to a paper in the *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. Furthermore, I provided one anonymous peer-review to *Atmospheric Measurement Techniques*.

A.3. Conference contributions

During the course of this PhD project, I joined various international conferences to present the methodologies I developed and results of my research. These conference contributions are summarised in Table A.2.

Table A.2.: My conference presentations during the course of my thesis.

Conference	Date, Place	Contribution	Comment
American Geophysical Society Fall Meeting	December 2015, San Francisco, CA, USA	Poster "A Site Atmospheric State Best Estimate of Temperature for Lauder, New Zealand"	Won an Outstanding Student Paper Award
8th GRUAN Implementation and Coordination Meeting	April 2016, Boulder, CO, USA	Invited talk "A new method to correct radiosonde temperature biases using radio occultation data"	
8th GRUAN Implementation and Coordination Meeting	April 2016, Boulder, CO, USA	Talk "A Site Atmospheric State Best Estimate for the temperature above Lauder, New Zealand"	
5th International Radio Occultation Working Group Meeting	September 2016, Leibnitz, Austria	Invited talk "The GCOS Reference Upper-Air Network (GRUAN) and its Relevance to the Radio Occultation Community "	Representing the GRUAN community
9th GRUAN Implementation and Coordination Meeting	June 2017, Helsinki, Finland	Talk "Comparison of GRUAN profiles with radio occultation bending angles propagated into temperature space"	
9th GRUAN Implementation and Coordination Meeting	June 2017, Helsinki, Finland	Poster "A Site Atmospheric State Best Estimate of Temperature for Lauder, New Zealand"	
9th GRUAN Implementation and Coordination Meeting	June 2017, Helsinki, Finland	Talk "An Update from the International Radio Occultation Working Group Meeting 2016"	
6th International Radio Occultation Working Group meeting	September 2017, Estes Park, CO, USA	Talk “Comparison of GRUAN profiles with radio occultation bending angles propagated into temperature space"	Representing the GRUAN community
2017 Conference of the New Zealand Meteorological Society	November 2017, Dunedin, New Zealand	Talk “A Site Atmospheric State Best Estimate of Temperature for Lauder, New Zealand"	

B. Glossary

AIRS	Atmospheric Infrared Sounder
BA	Bending Angle
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
doi	digital object identifier
ECMWF	The European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GPS	Global Positioning System
GTS	Global Telecommunication System
GCOS	Global Climate Observing System
GNSS	Global Navigation Satellite System
GUAN	GCOS Upper-Air Network
GRUAN	GCOS Reference Upper-Air Network
GSICS	Global Space-based Inter-Calibration System
IASI	Infrared Atmospheric Sounding Interferometer
LEO	Low Earth orbit
MetService	Meteorological Service of New Zealand Ltd.
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NDACC	Network for the Detection of Atmospheric Composition Change
NIWA	National Institute of Water and Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration (USA)
NWP	Numerical Weather Prediction
O-B	Observation minus Background
OSCAR	Observing System Capability Analysis and Review tool
PDF	Probability Density Function
RO	Radio Occultation
ROM SAF	Radio Occultation Meteorology Satellite Application Facility
RS	Radiosonde
SEA	Solar Elevation Angle
SE	Standard Error
SI	Système international d'unités (International System of Units)
SPL	Standard Pressure Level (of radiosondes profiles in TEMP format)
SD	Standard Deviation
SHADOZ	Southern Hemisphere ADditional OZonesondes
Tdry	Dry temperature
TL	Tangent linear
TOA	Top-of-the-atmosphere
UK	United Kingdon
WMO	World Meteorological Organization

Bibliography

Aires, F., W.B. Rossow, N.A. Scott, and A. Chédin (2002). "Remote sensing from the infrared atmospheric sounding interferometer instrument 2. Simultaneous retrieval of temperature, water vapor, and ozone atmospheric profiles". In: *J. Geophys. Res. Atmos.* 107.D22. DOI: 10.1029/2001JD001591.

Anthes, R.A. (2011). "Exploring Earth's atmosphere with radio occultation: contributions to weather, climate and space weather". In: *Atmos. Meas. Tech.* 4, pp. 1077–1103. DOI: 10.5194/amt-4-1077-2011.

Aumann, H.H., M.T. Chahine, C. Gautier, M.D Goldberg, E. Kalnay, L.M. McMillin, H. Revercomb, P.W. Rosenkranz, W.L. Smith, D.H. Staelin, and et al. (2003). "AIRS/AMSU/HSB on the Aqua mission: Design, science objectives, data products, and processing systems". In: *IEEE Transactions on Geoscience and Remote Sensing* 41.2, pp. 253–264.

Bevington, P.R. and D.K. Robinson (2003). *Data Reduction and Error Analysis for the Physical Sciences*. 3rd ed. Mc Graw Hill.

Beyerle, G., M.R. Gross, D.A. Haner, N.T. Kjome, I.S. McDermid, T.J. McGee, J.M. Rosen, H.-J. Schäfer, and O. Schrems (2001). "A Lidar and Backscatter Sonde Measurement Campaign at Table Mountain during February-March 1997: Observations of Cirrus Clouds". In: *J. Atmos. Sci.* 58.10, pp. 1275–1287.

Bodeker, G.E. "GRUAN Uncertainty Glossary".

Bodeker, G.E., S. Bojinski, D. Cimini, R.J. Dirksen, M. Haeffelin, J.W. Hannigan, D.F. Hurst, T. Leblanc, F. Madonna, M. Maturilli, A.C. Mikalsen, R. Philipona, T. Reale, D.J. Seidel, D.G.H. Tan, P.W. Thorne, H. Vömel, and J. Wang (2016). "Reference Upper-Air Observations for Climate: From Concept to Reality". In: *Bull. Amer. Meteor. Soc.* 97, pp. 123–135. DOI: 10.1175/bams-d-14-00072.1.

Bodeker, G.E., I.S. Boyd, and W.A. Matthews (1998). "Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand' 1986-1996". In: *J. Geophys. Res.* 103, 28661–28681. DOI: 10.1029/98JD02581.

Bodeker, G.E. and S. Kremser (2015). "Techniques for analyses of trends in GRUAN data". In: *Atmos. Meas. Tech.* 8, 1673–1684. DOI: 10.5194/amt-8-1673-2015.

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zemp (2014). "The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy". In: *Bull. Amer. Meteor. Soc.* 95.9, pp. 1431–1443. DOI: 10.1175/BAMS-D-13-00047.1.

Brinksma, E.J., J. Ajtic, J.B. Bergwerff, G.E. Bodeker, I.S. Boyd, J.F. de Haan, W. Hogervorst, J.W. Hovenier, and D.P.J. Swart (2000). "Validation of 3 years of ozone measurements over Network for the Detection of Stratospheric Change station Lauder, New Zealand". In: *J. Geophys. Res.* 105.D13, pp. 17291–17306. DOI: 10.1029/2000JD900143.

Brinksma, E.J., D.P.J. Swart, J.B. Bergwerff, Y.J. Meijer, and F.T Ormel (1997). "Advances in Atmospheric Remote Sensing with Lidar". In: ed. by A. Ansmann, R. Neuber, P. Rairoux, and U. Wandinger. Springer, Berlin, Heidelberg. Chap. Stratospheric and Mesospheric Profiling, RIVM Stratospheric Ozone Lidar at NDSC Station Lauder: Routine Measurements and Validation During the OPAL Campaign, pp. 529–532. DOI: 10.1007/978-3-642-60612-0_128.

Burrows, C.P. and S.B. Healy (2016). *Sensitivity of radio occultation-based dry temperature retrievals to upper-level information and its relevance to radiosonde bias corrections*. Forecasting Research Technical Report 615. Met Office. URL: http://www.metoffice.gov.uk/binaries/content/assets/mohippo/pdf/library/frtr_615_2016_2p.pdf.

Calbet, X., N. Peinado-Galan, P Rípodas, T. Trent, R. Dirksen, and M. Sommer (2017). "Consistency between GRUAN sondes, LBLRTM and IASI". In: *Atmos. Meas. Tech.* 10, pp. 2323–2335. DOI: 10.5194/amt-10-2323-2017.

Carminati, F., W. Bell, S. Migliorini, S. Newman, and A. Smith (2016). *An introduction to the GRUAN processor*. Satellite Applications Technical Memo 46. UK Met Office. URL: http://www.gaia-clim.eu/system/files/publications/Carminati_2016.pdf.

Chen, X., X. Huang, and M.G. Flanner (2014). "Sensitivity of modeled far-IR radiation budgets in polar continents to treatments of snow surface and ice cloud radiative properties". In: *Geophys. Res. Lett.* 41. DOI: 10.1002/2014GL061216.

Dai, A., J. Wang, P.W. Thorne, D.E. Parker, L. Haimberger, and X.L. Wang (2011). "A New Approach to Homogenize Daily Radiosonde Humidity Data". In: *J. Climate* 24.4, pp. 965–991. DOI: 10.1175/2010JCLI3816.1.

Dee, D.P. (2004). *Variational bias correction of radiance data in the ECMWF system*. ECMWF conference paper. European Centre for Medium-Range Weather Forecasts. URL: http://www.ecmwf.int/sites/default/files/elibrary/2004/8930-variational-bias-correction-radiance-data-ecmwf-system.pdf.

Derber, J.C. and W.-S. Wu (1998). "The Use of TOVS Cloud-Cleared Radiances in the NCEP SSI Analysis System". In: *Mon. Wea. Rev.* 126, pp. 2287–2299.

Dirksen, R.J., M. Sommer, F.J. Immler, D.F. Hurst, R. Kivi, and H. Vömel (2014). "Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde". In: *Atmos. Meas. Tech.* 7, pp. 4463–4490. DOI: 10.5194/amt-7-4463-2014.

Dobson, G.M.B. (1968). "Forty Years' Research on Atmospheric Ozone at Oxford: a History". In: *Appl. Opt.* 7.3, pp. 387–405. DOI: 10.1364/AO.7.000387.

Eyre, J.R. (2016). "Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties ". In: *Quart. J. Roy. Meteor. Soc.* 142.699, pp. 2284–2291.

Feldman, D.R., W.D. Collins, R. Pincus, X. Huang, and X. Chen (2014). "Far-infrared surface emissivity and climate". In: *Proc. Natl. Acad. Sci. U. S. A.* 111. DOI: 10.1073/pnas.1413640111,.

GCOS-170 (2013). *The GCOS Reference Upper-Air Network (GRUAN)*. WIGOS Technical Report 2013-2. World Meteorological Organization.

GCOS-200 (2016). *The Global Observing System for Climate: Implementation Needs*. Tech. rep. World Meteorological Organization.

Haimberger, L. (2007). "Homogenization of Radiosonde Temperature Time Series Using Innovation Statistics". In: *J. Climate* 20.7, pp. 1377–1403. DOI: 10.1175/JCLI4050.1. eprint: https://doi.org/10.1175/JCLI4050.1. URL: https://doi.org/10.1175/JCLI4050.1.

Haimberger, L., C. Tavolato, and S. Sperka (2012). "Homogenization of the Global Radiosonde Temperature Dataset through Combined Comparison with Reanalysis Background Series and Neighboring Stations". In: *J. Climate* 25, pp. 8108–3131. DOI: :10.1175/JCLI-D-11-00668.1.

Hersbach, H. and D. Dee (2016). *ERA5 reanalysis is in production*. ECMWF Newsletter No.147. URL: https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production.

Hewisen, T.J. (2007). "1D-VAR Retrieval of Temperature and Humidity Profiles From a Ground-Based Microwave Radiometer". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.7, pp. 2163–2168. DOI: 10.1109/TGRS.2007.898091.

Ho, S.-P., D. Hunt, A.K. Steiner, A.J. Mannucci, G. Kirchengast, H. Gleisner, S. Heise, A. von Engeln, C. Marquardt, S. Sokolovskiy, W. Schreiner, B. Scherrlin-Pirscher, C. Ao, J. Wickert, S. Syndergaard, K.B. Lauritsen, S.S. Leroy, E.R. Kursinski, Y.-H. Kuo, U. Foelsche, T. Schmidt, and M. Gorbunov (2012). "Reproducibility of GPS radio occultation data for climate monitoring: Profile-to-profile inter-comparison of CHAMP climate records 2002 to 2008 from six data centers". In: *J. Geophys. Res.* 117. DOI: 10.1029/2012JD017665.

Ho, S.-P., G. Kirchengast, S. Leroy, J. Wickert, A.J. Mannucci, A.K. Steiner, D. Hunt, W. Schreiner, S. Sokolovskiy, C. Ao, M. Borsche, A. von Engeln, U. Foelsche, S. Heise, B. Iijima, Y.-H. Kuo, R. Kursinski, B. Pirscher, M. Ringer, C. Rocken, and T. Schmidt (2009). "Estimating the uncertainty of using GPS radio occultation data for climate monitoring: Intercomparison of CHAMP refractivity climate records from 2002 to 2006 from different data centers". In: *J. Geophys. Res.* 114. DOI: 10.1029/2009JD011969.

Hurst, D.F., A. Lambert, W.G. Read, S.M. Davis, K.H. Rosenlof, E.G. Hall, A.F. Jordan, and S.J. Oltmans (2014). "Validation of Aura Microwave Limb Sounder stratospheric water vapor measurements by the NOAA frost point hygrometer". In: *J. Geophys. Res. Atmos.* 119.3, pp. 1612–1625. DOI: 10.1002/2013JD020757.

Immler, F.J., J. Dykema, T. Gardiner, D.N. Whiteman, P.W. Thorne, and H. Vömel (2010). "Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products". In: *Atmos. Meas. Tech.* 3.5, pp. 1217–1231. DOI: 10.5194/amt-3-1217-2010.

Ingleby, B. (2017). *An assessment of different radiosonde types 2015/2016*. Technical Memorandum 807. European Centre for Medium-Range Weather Forecasts.

Ingleby, B. and D. Edwards (2015). "Changes to radiosonde reports and their processing for numerical weather prediction". In: *Atmos. Sci. Let.* 16, pp. 44–49. DOI: 10.1002/asl2.518.

JCGM/WG 1 (2008). *Evaluation of measurement data - Guide to the expression of uncertainty in measurement (GUM)*. Tech. rep. JCGM 100:2008. International Bureau of Weights and Measures / Bureau International des Poids et Mesures.

– (2012). *International vocabulary of metrology - Basic and general concepts and associated terms (VIM)*. Tech. rep. JCGM 200:2012. International Bureau of Weights and Measures / Bureau International des Poids et Mesures.

Jensen, M.P., D. Holdridge, P. Survo, R. Lehtinen, S. Baxter, T. Toto, and K.L. Johnson (2016). "Comparison of Vaisala radiosondes RS41 and RS92 at the ARM Southern Great Plains Site". In: *Atmos. Meas. Tech.* 9, pp. 3115–3129. DOI: 10.5194/amt-9-3115-2016.

Kremser, S., J.S. Tradowsky, H.W. Rust, and G.E. Bodeker (2018a). "Is it feasible to estimate radiosonde biases from interlaced measurements?" In: *Atmos. Meas. Tech. Discuss.* Under review. DOI: 10.5194/amt-2018-6.

Kremser, S., J.S. Tradowsky, H.W. Rust, and G.E. Bodeker (2018b). "Is it feasible to estimate radiosonde biases from interlaced measurements?" In: *Atmos. Meas. Tech.* 11, pp. 3021–3029. DOI: 10.5194/amt-11-3021-2018.

Kursinski, E.R., G.A. Hajj, S.S. Leroy, and B. Herman (2000). "The GPS Radio Occultation Technique". In: *Terrestrial, Atmospheric and Oceanic Sciences* 11.1, pp. 53–114.

Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, and K.R. Hardy (1997). "Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System". In: *J. Geophys. Res.* 102, pp. 23429–23465.

Ladstädter, F., A.K. Steiner, M. Schwärz, and G. Kirchengast (2015). "Climate intercomparison of GPS radio occultation, RS90/92 radiosondes and GRUAN from 2002-2013". In: *Atmos. Meas. Tech.* 8.3, pp. 1819–1834. DOI: 10.5194/amt-8-1819-2015.

Lubin, D., J. Verlinde, D.H. Bromwich, A.M. Vogelmann, and L.M. Russell (2017). *ARM West Antarctic Radiation Experiment (AWARE) Field Campaign Report*. Tech. rep. DOE/SC-ARM-17-028. U.S. Department of Energy, Office of Sciences. URL: https://www.arm.gov/publications/programdocs/doe-sc-arm-17-028.pdf.

Mims III, F.M., L. Hartung Chambers, and D.R. Brooks (2011). "Measuring Total Column Water Vapor by Pointing an Infrared Thermometer at the Sky". In: *Bull. Amer. Meteor. Society*. DOI: 10.1175/2011BAMS3215.1.

NIST (2008). *The International System of Units (SI)*. Ed. by B.N. Taylor and A. Thompson. 2008 Edition. US National Institute of Standards and Technology (NIST).

Noh, Y.-C., B.J. Sohn, Y. Kim, S. Joo, and W. Bell (2016). "Evaluation of Temperature and Humidity Profiles of Unified Model and ECMWF Analyses Using GRUAN Radiosonde Observations". In: *Atmosphere* 7. DOI: 10.3390/atmos7070094.

Norbert, C. (2013). "Human Spaceflight and Exploration". In: ed. by C. Norbert. Springer, Berlin, Heidelberg. Chap. Introduction, pp. 1–9. DOI: 10.1007/978-3-642-23725-6_1.

Palchetti, L., G. Bianchini, G. Di Natale, and M. Del Guasta (2015). "Far-infared radiative properties of water vapor and clouds in Antarctica". In: *Bull. Amer. Meteor. Society*. DOI: 10.1175/BAMS-D-13-00286.1.

Parrish, A., B.J. Connor, J.J. Tsou, I.S. McDermid, and W.P. Chu (1992). "Ground-based microwave monitoring of stratospheric ozone". In: *J. Geophys. Res. Atmos.* 97.D2, pp. 2541–2546. ISSN: 2156-2202. DOI: 10.1029/91JD02914. URL: http://dx.doi.org/10.1029/91JD02914.

Reale, T., P.W. Thorne, and B. Sun (2016). "8th GRUAN Implementation and Coordination Meeting (ICM-8)," in: *GSICS Quarterly Newsletter*. DOI: 10.7289/V5222RSK.

Rodgers, C.D. (2000). *Inverse Methods for Atmospheric Sounding, Theory and Practice*. World Scientific.

Seidel, D.J, F.H. Berger, F. Immler, H.J. Diamond, J. Dykema, D. Goodrich, W. Murray, T. Peterson, D. Sisterson, M. Sommer, P.W. Thorne, H. Vömel, and J. Wang (2009). "Reference upper-air observations for climate: Rationale, progress, and plans". In: *Bull. Amer. Meteor. Soc.* 90.3, pp. 361–369.

Siméoni, D., C. Singer, and G. Chalon (1997). "Infrared atmospheric sounding interferometer". In: *Acta Astronautica* 40.2-8, pp. 113–118.

Sommer, M., R. Dirksen, and C. von Rohden (2014). *Brief Description of the RS92 GRUAN Data Product (RS92-GDP)*. GRUAN Technical Document 4. GCOS Reference Upper-Air Network. URL: http://www.dwd.de/EN/research/international_programme/gruan/download/gruan_td-4.pdf?__blob=publicationFile&v=2.

Steiner, A.K., D. Hunt, S.-P. Ho, G. Kirchengast, A.J. Mannucci, B. Scherllin-Pirscher, H. Gleisner, A. von Engeln, T. Schmidt, C. Ao, S.S. Leroy, E.R. Kursinski, U. Foelsche, M. Gorbunov, S. Heise, Y.-H. Kuo, K.B. Lauritsen, C. Marquardt, C. Rocken, W. Schreiner, S. Sokolovskiy, S. Syndergaard, and J. Wickert (2013). "Quantification of structural uncertainty in climate data records from GPS radio occultation". In: *Atmos. Chem. Phys.* 13.3, pp. 1469–1484. DOI: 10.5194/acp-13-1469-2013.

Steiner, A.K., B.C. Lackner, F. Ladstädter, B. Scherllin–Pirscher, U. Foelsche, and G. Kirchengast (2011). "GPS radio occultation for climate monitoring and change detection". In: *Radio Science* 46.RSOD24. DOI: 10.1029/2010RS004614.

Sun, B., A. Reale, S. Schroeder, D.J. Seidel, and B. Ballish (2013). "Toward improved corrections for radiation-induced biases in radiosonde temperature observations". In: *J. Geophys. Res.* 118, pp. 4231–4243. DOI: 10.1002/jgrd.50369.

Sun, B., A. Reale, D.J. Seidel, and D.C. Hunt (2010). "Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics". In: *J. Geophys. Res.* 115. DOI: 10.1029/2010JD014457.

Syndergaard, S. (1999). *Retrieval Analysis and Methodologies in Atmospheric Limb Sounding Using the GNSS Radio Occultation Technique*. Danish Meteorological Institute Scientific Report 99-6. Danish Meteorological Institute.

Thorne, P.W., D.E. Parker, S.F.B. Tett, P.D. Jones, M. McCarthy, H. Coleman, and P. Brohan (2005). "Revisiting radiosonde upper air temperatures from 1958 to 2002". In: *J. Geophys. Res. Atmos.* 110.D18. ISSN: 2156-2202. DOI: 10.1029/2004JD005753.

Tobin, D.C., H.E. Revercomb, R.O. Knuteson, B.M. Lesht, L.L. Strow, S.E. Hannon, W.F. Feltz, L.A. Moy, E.J. Fetzer, and T.S. Cress (2006). "Atmospheric Radiation Measurement site atmospheric state best estimates for Atmospheric Infrared Sounder temperature and water vapour retrieval". In: *J. Geophys. Res.* 111. DOI: 10.1029/2005JD006103.

Tradowsky, J.S. (2015). *Characterisation of radiosonde temperature biases and errors using radio occultation measurements*. ROM SAF Visiting Scientist report 26. Radio Occultation Meteorology Satellite Application Facility. URL: http://www.romsaf.org/Publications/reports/romsaf_vs26_rep_v12.pdf.

– (2016). *Radiosonde Temperature Bias Corrections using Radio Occultation Bending Angles as Reference*. ROM SAF Visiting Scientist report 31. Radio Occultation Meteorology Satellite Application Facility. URL: http://www.romsaf.org/Publications/reports/romsaf_vs31_rep_v10.pdf.

Tradowsky, J.S., G.E. Bodeker, R.R. Querel, P.J.H. Builtjes, and J. Fischer (2018). "Combining Data from the Distributed GRUAN Site Lauder-Invercargill, New Zealand, to Provide a Site Atmospheric State Best Estimate of Temperature". In: *Earth System Science Data Discussions*. Publication submitted. DOI: 10.5194/essd-2018-20.

Tradowsky, J.S., G.E. Bodeker, P.W. Thorne, F. Carminati, and W. Bell (2016). "GRUAN in the service of GSICS: Using reference ground-based profile measurements to provide traceable radiance calibration for space-based radiometers". In: *GSICS Quarterly Newsletter* 10.2. DOI: 10.7289/V5GT5K7S.

Tradowsky, J.S., C.P. Burrows, S.B. Healy, and J.R. Eyre (2017). "A New Method to Correct Radiosonde Temperature Biases Using Radio Occultation Data". In: *Journal of*

Applied Meteorology and Climatology 56.6, pp. 1643–1661. DOI: 10.1175/JAMC-D-16-0136.1.

Trenberth, K.E., T.R. Karl, and T.W. Spence (2002). "The Need for a Systems Approach to Climate Observations". In: *Bull. Amer. Meteor. Soc.* 83.11, pp. 1593–1602. DOI: 10.1175/BAMS-83-11-1593.

Turner, D.D. and E.J. Mlawer (2010). "The radiative heating in underexplored bands campaigns". In: *Bull. Amer. Meteor. Society*. DOI: 10.1175/2010BAMS2904.1.

Uchino, O., T. Nagai, T. Fujimoto, W.A. Matthews, and J. Orange (1995). "Extensive lidar observations of the Pinatubo aerosol layers at Tsukuba (36.1N), Naha (26.2N), Japan and Lauder (45.0S), New Zealand". In: *Geophys. Res. Lett.* 22.1, pp. 57–60. ISSN: 1944-8007. DOI: 10.1029/94GL02735.

Vömel, H., D.E. David, and K. Smith (2007). "Accuracy of tropospheric and stratospheric water vapor measurements by the cryogenic frost point hygrometer: Instrumental details and observations". In: *J. Geophys. Res. Atmos.* 112.D8. DOI: 10.1029/2006JD007224.