



Performance of one-dimensional hydrodynamic lake models during short-term extreme weather events

J.P. Mesman^{a,b,*}, A.I. Ayala^{a,b}, R. Adrian^{c,d}, E. De Eyto^e, M.A. Frassl^f, S. Goyette^{a,g}, J. Kasparian^{a,g}, M. Perroud^a, J.A.A. Stelzer^{a,c,d}, D.C. Pierson^b, B.W. Ibelings^a

^a University of Geneva, Institute for Environmental Sciences, 1211, Geneva 4, Switzerland

^b Uppsala University, Department of Ecology and Genetics, 75236, Uppsala, Sweden

^c Leibniz Institute of Freshwater Ecology and Inland Fisheries, Department of Ecosystem Research, Berlin, Germany

^d Freie Universität Berlin, Department of Biology, Chemistry, and Pharmacy, 14195, Berlin, Germany

^e Marine Institute, F28PF65, County Mayo, Ireland

^f Griffith University, Australian Rivers Institute, Brisbane, Australia

^g University of Geneva, Group of Applied Physics, 1211, Geneva 4, Switzerland

ARTICLE INFO

Keywords:

Storm
Heatwave
Model validation
Simstrat
GOTM
General lake model

ABSTRACT

Numerical lake models are useful tools to study hydrodynamics in lakes, and are increasingly applied to extreme weather events. However, little is known about the accuracy of such models during these short-term events. We used high-frequency data from three lakes to test the performance of three one-dimensional (1D) hydrodynamic models (Simstrat, GOTM, GLM) during storms and heatwaves. Models reproduced the overall direction and magnitude of changes during the extreme events, with accurate timing and little bias. Changes in volume-averaged and surface temperatures and Schmidt stability were simulated more accurately than changes in bottom temperature, maximum buoyancy frequency, or mixed layer depth. However, in most cases the model error was higher (30–100%) during extreme events compared to reference periods. As a consequence, while 1D lake models can be used to study effects of extreme weather events, the increased uncertainty in the simulations should be taken into account when interpreting results.

1. Introduction

Over the last few years, limnologists have devoted increased attention to extreme weather events (e.g. Bertani et al., 2016; Kasprzak et al., 2017; Andersen et al., 2020; Chen et al., 2020). These are predicted to become more frequent and intense with climate change (IPCC, 2014; Bailey and Pol, 2016), and can have profound effects on lake ecosystems. Extreme weather events, such as storms and heatwaves, have a direct effect on lake physics. Wind storms can induce mixing events, cooling of the surface layer, resuspension of sediments, and deepening of the thermocline (Jennings et al., 2012; Andersen et al., 2020). Heatwaves have effects that are largely opposite to storms, as these cause heating of the water column and strengthen thermal stratification (Jankowski et al., 2006; Huber et al., 2012). The disturbance in the thermal profile and inflow conditions caused by these events often mediates further changes in nutrients, oxygen, and phytoplankton community (Huber et al., 2010; Klug et al., 2012; Kasprzak et al., 2017). Physical

disturbances in the water column due to an extreme event are often short-lived (Wilhelm and Adrian, 2008; Jennings et al., 2012; Kuha et al., 2016; Stockwell et al., 2020), although they can also have a longer effect (Huber et al., 2012; Andersen et al., 2020), depending on the time of year when they occur (Mi et al., 2018), or whether water transparency is affected as part of the event (e.g. dissolved organic carbon loading, or suspended particles) (Klug et al., 2012; De Eyto et al., 2016; Perga et al., 2018). Moreover, a short physical disturbance does not automatically imply a short-lived effect on biogeochemistry and ecology. For example, short-term mixing events can be a major factor affecting the transport of nutrients, stimulating phytoplankton growth (Soranno et al., 1997; Crockford et al., 2015).

Numerical lake models are useful tools for understanding aquatic processes, disentangling causal factors, and for estimating future trajectories of the system (forecasting, climate scenarios). Recently, several studies have applied one-dimensional (1D) lake models to study the effects of extreme weather conditions on lake thermal structure or lake

* Corresponding author. University of Geneva, Institute for Environmental Sciences, 1211, Geneva 4, Switzerland.

E-mail address: Jorrit.Mesman@unige.ch (J.P. Mesman).

ecology (Bueche et al., 2017; Mi et al., 2018; Perga et al., 2018; Soares et al., 2019; Chen et al., 2020). However, there is still a lack of understanding on how accurately lake models actually simulate observed conditions during these short-term events. It is common practice to assess models on the basis of their goodness-of-fit (e.g. root mean square error or mean absolute error) over the whole calibration and/or validation period, but these long timescales obscure any potential errors during disturbance by and recovery from short-term events. Evaluating event-specific errors will help to understand and minimise the uncertainty in model studies concerning ecosystem effects of extreme weather events. Showing that a model is capable to accurately simulate system changes caused by extreme weather events, will increase our confidence in their capability to provide reliable estimates for future effects of climate change. This advanced model testing is an important step in a multi-level model assessment (Hipsey et al., 2020).

In the present validation study, we used more than 10 years of hourly and sub-hourly in-situ measurements of meteorological variables and water temperature from three lakes of varying depths and mixing dynamics, to assess model performance during short-term extreme wind and temperature events. The analyses were done with three 1D hydrodynamic models - Simstrat (Goudsmit et al., 2002; Gaudard et al., 2019), GOTM (General Ocean Turbulence Model, Umlauf et al., 2005), and GLM (General Lake Model, Hipsey et al., 2019). These three models differ in turbulence schemes, calibration procedures, forcing variables, and parameterisations. Additionally, we report on observed changes in lake thermal metrics during storms and heatwaves in lakes with different morphology and mixing regimes. We focus our analysis on lake temperature (full profile, volume-averaged, surface, and bottom temperatures) and stratification metrics (Schmidt stability, maximum buoyancy frequency, and mixed layer depth), based on high-frequency temperature profile observations and simulations. Changes in these thermal metrics can translate into further changes in water transparency, and distribution and transport of oxygen and nutrients, with repercussions on biological processes.

We assessed whether the models could reproduce the direction, magnitude, and timing of change during an event, what the accuracy of the models was during extreme events compared to standard conditions, and if there was a consistent tendency of the models to over- or underestimate changes during an event. Following this, we draw conclusions on the implications of our findings for applying 1D hydrodynamic models to short-term extreme wind and temperature events in different types of studies.

2. Methods

2.1. Observational data

Meteorological and water temperature profile data from Lough Feeagh (Ireland), Lake Erken (Sweden), and Müggelsee (Germany) were used for this study. Long-term records of sub-hourly water temperature profile and surface meteorological data were available for the period 2004–2017 at Müggelsee and between 2005 and 2017 for Lough Feeagh and Lake Erken. Measurements of air temperature, wind speed and wind direction, relative humidity, incoming solar radiation, and air pressure were collected at each lake. Cloud cover was available at hourly intervals in the database generated by Moras et al. (2019) for Lake Erken and from the airport Berlin-Schönefeld for Müggelsee (10 km distance from the lake), and daily observations on-site were made for Lough Feeagh.

Lough Feeagh (53°56'21" N, 9°34'33" W; mean depth 14.5 m; maximum depth 46 m) is a monomictic lake, located on the west coast of Ireland. It experiences high rainfall and wind speeds, has no winter ice cover, and is rich in dissolved organic carbon (DOC) as a result of drainage from surrounding peatlands (De Eyto et al., 2016). Meteorological records and lake temperature profile data are available with measurement frequencies of 1 and 2 min, respectively. Meteorological

data were collected on the shore of the lake (Met Éireann, 2018). The water temperature profiles were measured by an automated monitoring buoy above the deepest point of the lake, with temperature sensors at 0.9, 2.5, 5, 8, 11, 14, 16, 18, 20, 22, 27, 32, and 42 m below the surface (De Eyto et al., 2020).

Lake Erken (59°50'37" N, 18°35'38" E; mean depth 9 m; maximum depth 21 m) is a dimictic lake in the eastern part of Sweden. It has a surface area of 24 km² and experiences ice cover in winter and stratification in summer (Persson and Jones, 2008). High-frequency (30-min) lake temperature data from 2005 onwards were used for this study. Meteorological forcing data with a 5-min frequency are available from July 2008 onwards, and hourly forcing data were used before this point in time. Meteorological data were collected from a station on a small island 500 m off shore, while water temperature was measured at a monitoring buoy that was located a further 500 m from the island, at 15 m depth. Water temperature measurements were made at 0.5 m depth intervals prior to 2016, and at 0.25 m intervals after 2016.

Müggelsee (52°26'24" N, 13°38'58" E; mean depth 4.9 m; maximum depth 8 m), located in Berlin, Germany, is the shallowest lake in this study. It has a surface area of 7.3 km², is wind-exposed, and classifies as a polymictic lake (Wilhelm and Adrian, 2008). It experiences ice cover in most winters and stratifies during summer at high air temperatures and moderate wind conditions. Lake temperature data were collected by a floating monitoring station anchored 300 m from the northern lake shore at a depth of 5.5 m. Water temperature at 2 m depth was measured every 5 min, and a profile with 0.5 m depth intervals from the surface up to 5 m depth was measured every hour. Meteorological data were available every 5 min, measured at the monitoring station (for more details on used sensors and methodology, see Wilhelm et al., 2006).

Precipitation and in- and outflows were not included in this study, and the water level was kept constant in the simulations. The reason for this is that these data were available at lower frequencies than the forcing and model time steps, and potentially at lower frequencies than the effects of the investigated events. Additionally, annual water level fluctuations are generally less than 1 m in Lough Feeagh and Lake Erken (Moras et al., 2019; Kelly et al., 2020), and only around 0.25 m in Müggelsee (Driescher et al., 1993), so water level was assumed to be of minor importance for thermal stratification patterns. Water transparency was also kept constant in all three models. The light attenuation coefficient was calculated from the average Secchi depth (S_d) observed over the simulated period with equations specific for the conditions in each lake; attenuation coefficient = $2.7/S_d$ (Feeagh; Koenings and Edmundson, 1991), $2.4/S_d$ (Lake Erken; based on observed Secchi depths and light profiles, unpublished data), and $1.3611 * S_d^{-0.7105}$ (Müggelsee; Hilt et al., 2010).

For information on gap-filling procedures, see Suppl. Mat. A.

2.2. Lake thermal metrics

Model fit was assessed for the following thermal metrics: lake temperature (full profile), volume-averaged temperature, surface temperature (≤ 1 m depth), bottom temperature (deepest observation), Schmidt stability (Schmidt, 1928; Idso, 1973), maximum buoyancy frequency squared (N^2 , hereafter referred to as "maximum buoyancy frequency"), and mixed layer depth. The R package "rLakeAnalyzer" (Winslow et al., 2019) was used to calculate volume-averaged temperature, Schmidt stability, and maximum buoyancy frequency, see Read et al. (2011) for formulas. The mixed layer depth was defined using an absolute density difference from the surface (following De Boyer Montégut et al., 2004; Wilson et al, submitted). A threshold of 0.15 kg/m³ was chosen, which gave robust estimates of the depth of stratification for all three lakes. If the density of the deepest measured temperature was within this density threshold, the water column was assumed to be completely mixed and mixed layer depth was set to the deepest measurement. The relation between water temperature and density by Martin and McCutcheon (1999) was used.

2.3. Lake models

Three 1D hydrodynamic lake models were used in this study: Simstrat, GOTM, and GLM. The models take into account lake morphology in turbulence equations, but otherwise assume horizontal homogeneity. These models all simulate the vertical thermal lake structure and are forced by the same meteorological input, but are different in their code structure, processes included (such as seiche-induced mixing, or different wavelenghts of light), and parameterizations for surface fluxes and turbulence, so that each model could result in potentially different outcomes. A full description of the governing equations used by each of these open source models can be found in Goudsmit et al. (2002) for Simstrat, Umlauf et al. (2005) for GOTM, and Hipsey et al. (2019) for GLM, in addition to manuals and support on the respective websites (Simstrat: <https://github.com/Eawag-AppliedSystemAnalysis/Simstrat>, GOTM: <https://gotm.net/>, GLM: <http://aed.see.uwa.edu.au/research/models/GLM/>. Last access: 2020-08-20). Specific settings for each model used in this study are provided in the Suppl. Mat. B. The main differences between the models are mentioned below.

Simstrat and GOTM have a fixed layer structure, resolving turbulent kinetic energy production and diffusion between layers of fixed thickness. Layers in GLM can vary in thickness or merge depending on the degree of turbulent kinetic energy. Simstrat was forced with wind direction as an additional input variable; this is used to resolve mixing caused by seiches. Ice cover modules are present in Simstrat and GLM, while there is no ice module in the version of GOTM used in this study. Air pressure is a constant value in Simstrat and GLM, and the average value over the simulated period was used in this study, while measured air pressure was used as input in GOTM. Additionally, the used version of GLM could not be run with sub-hourly forcing due to the inherent structure of the code, while a forcing frequency of 10 min was used for Simstrat and GOTM. To account for this difference, additional runs with hourly forcing were performed for Simstrat and GOTM. Whenever these hourly forcing runs were used instead of the ones with 10-min forcing, this is specifically mentioned.

2.4. Calibration

A period of one year was used for model spin-up and calibration. Automatic calibration procedures were applied to minimise the error in water temperature at all depths. The standard calibration procedures available for each model were different. Simstrat applied the PEST (model-independent Parameter ESTimation and uncertainty analysis) software to minimise the sum of squares of the error (Doherty, 2015). The ParSAC python package was used for GOTM. It maximises the log-likelihood using a differential evolution method. GLM was calibrated with the “nloptr” R package (Johnson, 2014), using the Nelder-Mead simplex algorithm (Nelder and Mead, 1965) to minimise the root mean square error. Model parameters and calibration ranges can be found in Suppl. Mat. C. The remainder of the data series was used as validation period and to identify extreme events.

2.5. Storm and heatwave events

Model performance during extreme weather events was assessed on a selection of ten storms and ten heatwaves per lake.

The storm events were defined using 10-min wind speed observations. For the purpose of identifying storms missing data were not filled (Suppl. Mat. A) so that only actual measured data were used. The period April–October was used due to the frequent absence of winter profile data in Lake Erken and Müggelsee, due to ice cover. We chose to base the events on the turbulent wind energy flux at 10 m above the surface (P_{10} , $W m^{-2}$) instead of wind speed, because it is a more direct measure of the amount of energy transfer to the lake, and thus a more direct measure of the atmospheric impact on thermal stratification. P_{10} was calculated as

$P_{10} = \rho_{air} C_D U_{10}^3$, using a fixed drag coefficient (C_D) of $0.9 \cdot 10^{-3}$ (Wüest et al., 2000), where ρ_{air} is air density ($kg m^{-3}$) and U_{10} is wind speed at 10 m above the surface ($m s^{-1}$). The top 5% of daily sums of P_{10} were selected, and days within this selection were considered as a single event if they occurred within two days from each other. Events with less than 10 h of measured water temperature data, or no prior thermal stratification (Lough Feeagh and Lake Erken only), were excluded. The exact timing of the start and end of an event were defined when the 8-h moving average of wind speed passed the 75th percentile of all observed wind speed data. Lastly, P_{10} was recalculated for the whole duration of an event, but the 75th percentile of all P_{10} data was subtracted, to attach value only to the periods with extremely high wind speeds. The events were then ordered by the summed P_{10} and the top 10 events were selected.

The heatwave events were defined using air temperature data. To select warm spells relative to the time of the year, that is, also outside of the middle of summer, the two warmest three-day degree-day periods for each month in the period April–August were taken, always in two separate years. If the temperature on the days before and after this three-day period was above the 95th percentile of that month, these days were also included in the event. Events that had insufficient water temperature data, or that were within one week of another heatwave event, were excluded. In that case, the next warmest period was chosen, until an event with enough lake data was found. For Lake Erken, only one event of the four warmest degree-day periods in April had enough data. Instead of picking a colder period in April, an extra event in August was selected.

In order to compare the response of the models during extreme events with average weather conditions, ten “reference” wind and temperature periods were defined. The selection methods and time periods were identical to the methods and periods used for the extreme events, but instead of selecting events with the highest daily sums of P_{10} or highest three-day summed temperature, periods with values closest to the median were chosen. Reference events could not be within one week of an extreme event and the duration was fixed to 24 h for wind periods, and three days for temperature periods (Fig. 1). For Lake Erken, reference temperature periods were shifted one month (May–September), due to frequently missing data in mid-April because of ice cover.

Simulations were initialised one week before each event. This initialisation was done to minimise model error and differences between models at the onset of an event, but at the same time to allow spin-up time of the simulation. Restricting the simulation period before the extreme event allowed for direct quantification of model performance during extreme weather conditions and isolation of the effects of the event, avoiding the effects of accumulated model error during pre-event normal weather conditions.

2.6. Assessment of model performance

Model performance was evaluated by comparing measured and simulated temperature profiles and the lake thermal metrics calculated from them, using Mean Absolute Error (MAE) as a measure for goodness of model fit. MAE was first calculated for the calibration and validation periods. Then, the MAE of the water temperature profile was compared between extreme and reference events with a *t*-test or a Wilcoxon rank-sum test (in case of non-normality or outliers) for each lake and storms and heatwaves separately. To see if different lakes and event types had a different effect on model fit, a two-way ANOVA on the MAE during extreme events only was performed. A post hoc Tukey test was done to compare lakes with each other. A one-way ANOVA on the MAE was done to compare the performance of the different models during extreme events, followed by a post hoc Tukey test.

In addition, the difference in thermal metrics between the two pre-event days and the two post-event days was defined as the change in a metric during an event. This change for each metric in observations was

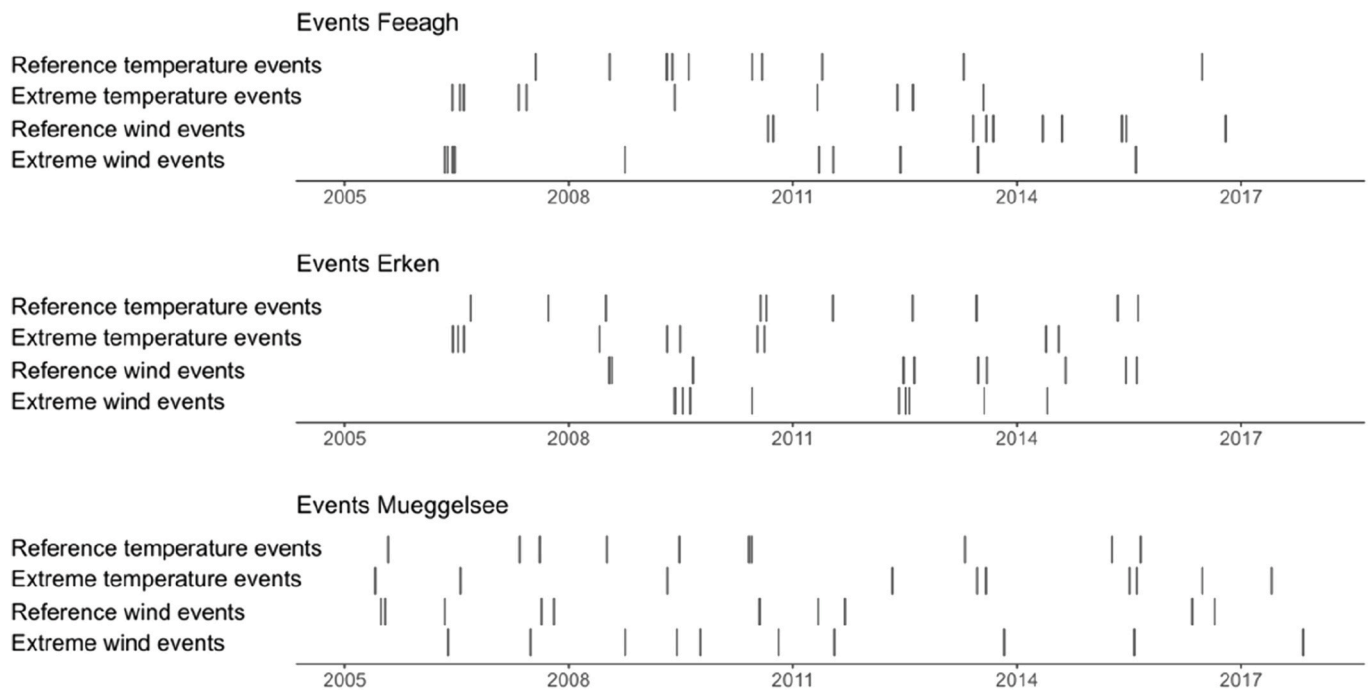


Fig. 1. Timing of the extreme and reference events used in the analyses. The start and end dates of all events can be found in Suppl. Mat. D.

tested for significance with a t -test, or with a Wilcoxon sign test in case of non-normal data (assessed by QQ-plots and Shapiro-Wilk tests) or outliers. The performance of the models in simulating the change in a metric during events was assessed by inspecting plots and by calculating the Concordance Correlation Coefficient (CCC; Lin, 1989) between the simulated and observed change in metric. The CCC is similar to Pearson's correlation coefficient, but penalises for a deviation from the 1:1 line and was therefore deemed a more accurate statistic for model comparison. To test for consistent bias in model simulations during events, a one-way ANOVA was performed on the change during an extreme event, for each lake, event type (storm/heatwave), and metric. A post hoc Tukey test was used to compare models with observations. In case the data was non-normally distributed (assessed by Shapiro-Wilk tests), a Kruskal-Wallis test and post hoc Dunn test were performed instead, using the "dunn.test" R package (Dinno, 2017).

To evaluate model accuracy in simulating the timing of events, a temporal cross-correlation analysis was performed on the simulated and observed datasets for each event and each metric. The cross-correlation analysis temporally shifted the two datasets relative to each other, and the time lag with the highest cross-correlation coefficient was taken as the time lag in the simulation. Data gaps up to 2 h were linearly interpolated. Larger gaps were considered exclusion criteria for the cross-correlation analysis. Also, if the maximum cross-correlation coefficient between simulation and observations was below 0.3, the simulation was deemed too inaccurate to determine a time lag.

All analyses were done with the software R (version 3.6.2, R Core Team, 2019). In those cases where the p -value of a statistical test was used to distinguish between significant and non-significant, an alpha of 0.05 was used.

3. Results

3.1. Model performance for the whole simulation period

The models successfully reproduced the seasonal cycles of temperature and stratification (Suppl. Mat. E). All models performed reasonably well, although GLM showed a poorer performance compared to the other two models, based on MAE during the calibration and validation

periods (Fig. 2).

3.2. Observations during events

The observed data confirmed the opposite effects of storms and heatwaves on surface temperature, volume-averaged temperature, Schmidt stability and maximum buoyancy frequency (Fig. 3, Suppl. Mat. F). Differences between lakes could be observed. In the two deeper lakes of this study, Lough Feeagh and Lake Erken, Schmidt stability decreased and the mixed layer deepened during extreme wind events. Volume-averaged temperature was not strongly affected, but surface temperature decreased and bottom temperatures increased, indicating mixing between top and bottom waters. In Müggelsee, complete mixing occurred during all studied storm events, and the water column was often well-mixed already before the start of the actual event due to the lake's shallow depth (data not shown). For four out of the ten storms in Müggelsee, stratification formed again within a few days after the end of an event, which caused no change in Schmidt stability or mixed layer depth compared to before the event. Cooling of all water layers occurred during all ten storm events in Müggelsee.

During high temperature events, Schmidt stability tended to increase in Lough Feeagh and Lake Erken (Fig. 3). There was no change in the mixed layer depth during these events. Water temperatures at all depths increased, but the increase was stronger near the surface than near the bottom. After heatwave events in Müggelsee, temperature in all water layers had increased to a similar extent, because effects of the heatwaves on stratification dissipated soon after the end of the events. Stratification occurred during nine of the ten events, but increases in Schmidt stability and mixed layer depth did not remain significant after the events.

In all lakes and during both storm and heatwave events, changes in maximum buoyancy frequency tended to follow the same trend as Schmidt stability, but were not significantly different from zero.

3.3. Model performance during events

Generally, models performed better during the reference events compared to the extreme events. Only the simulations for Lough Feeagh had significantly higher MAE during storm events compared to reference

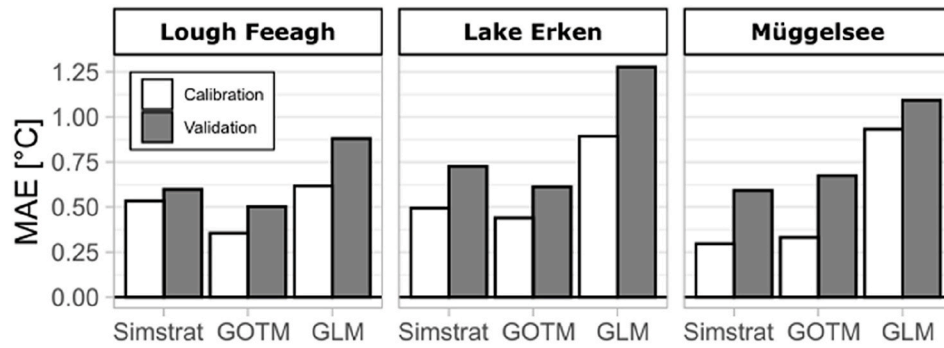


Fig. 2. Bar graph of Mean Absolute Error (MAE) of lake temperature in °C for the calibration (first year) and validation period of each model and lake. For the exact values and other measures of goodness-of-fit, see Suppl. Mat. E.

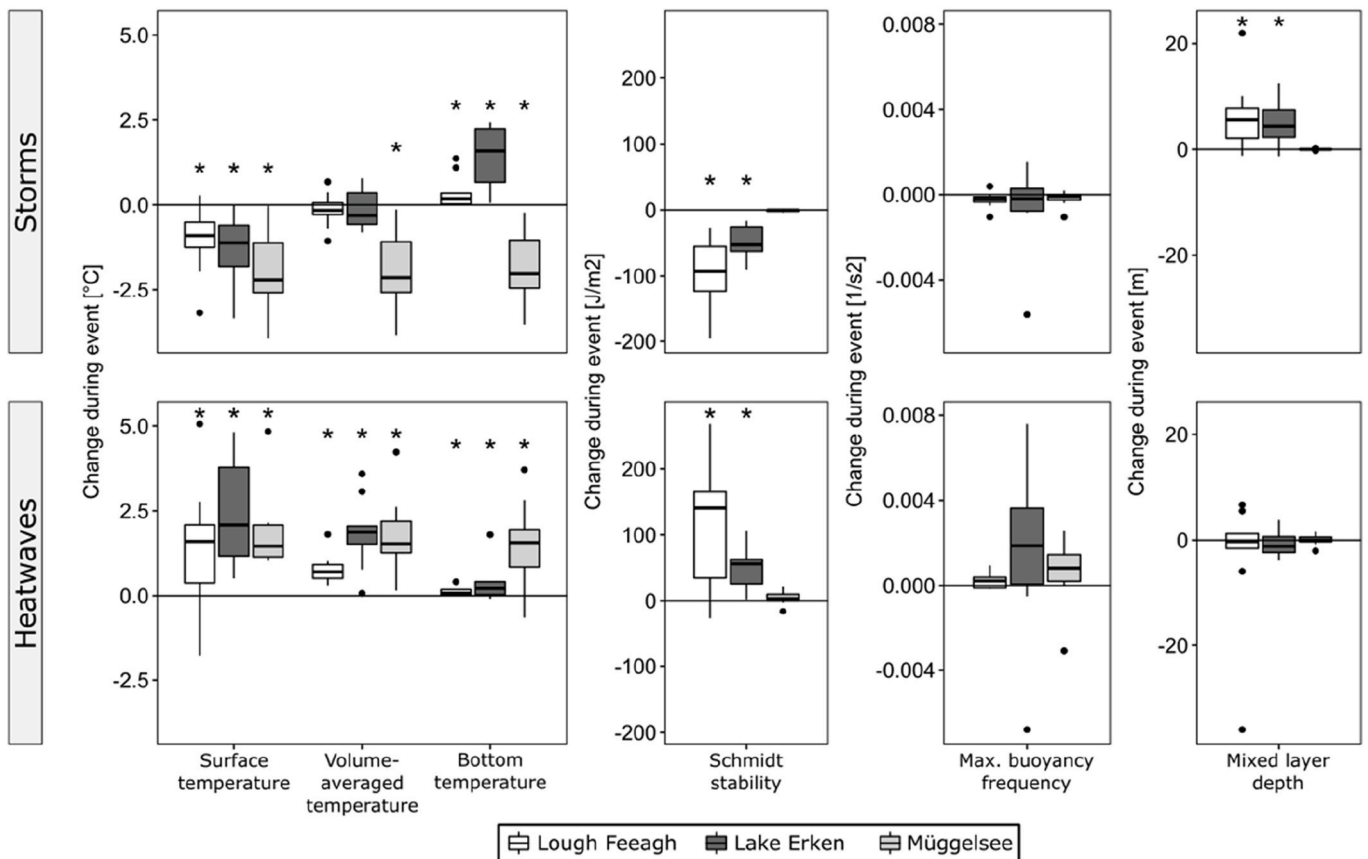


Fig. 3. Boxplots showing the observed changes in temperature metrics, Schmidt stability, maximum buoyancy frequency, and mixed layer depth during the identified storms and heatwaves. Change during the event is calculated as the average over the two days after the event minus the average over the two days before the event. The boxplots show the median and first and third quartile. Whiskers extend to the smallest and largest value within 1.5 times the inter-quartile range from the nearest quartile (see geom_boxplot function in ggplot2 R package, Wickham, 2016). Values outside this range are defined as outliers (•). * indicates a significant difference from zero change (See Suppl. Mat. F).

wind events, while MAE during storm events in Müggelsee was significantly lower than the reference (Fig. 4). In all lakes, the MAE of the water temperature profile was higher during the heatwave events compared to the MAE during the reference temperature events. A two-way ANOVA on the MAE during extreme events showed that different lakes ($F = 18.58, p < 0.001$), different event types (storm/heatwave, $F = 6.54, p = 0.01$) and the interaction between the two ($F = 6.91, p = 0.001$) had significant effects on MAE. Lough Feeagh had the lowest MAE's during storm and heatwave events, compared to the other lakes (0.3–0.4 °C lower, Tukey test, $p < 0.001$), and MAE's were slightly higher during the heatwave events compared to the storm events

(0.16 °C higher, Tukey test, $p = 0.01$). During the extreme events, Simstrat and GOTM had a similar MAE (mean of 0.62 °C), while the MAE for GLM was 0.24 °C (39%) higher (One-way ANOVA, $F = 5.32, p = 0.005$, Tukey test, $p < 0.05$, Suppl. Mat. G).

Despite these increases in model error, the direction and magnitude of change during extreme events was often reproduced by the models (Figs. 5, 6); during storms, the changes in surface- and volume-averaged temperature were accurately reproduced by all models (Concordance Correlation Coefficient, $CCC > 0.7$, Fig. 6), while the bottom temperature was reproduced with less accuracy. Simstrat and GOTM reproduced changes in Schmidt stability and buoyancy frequency during storms

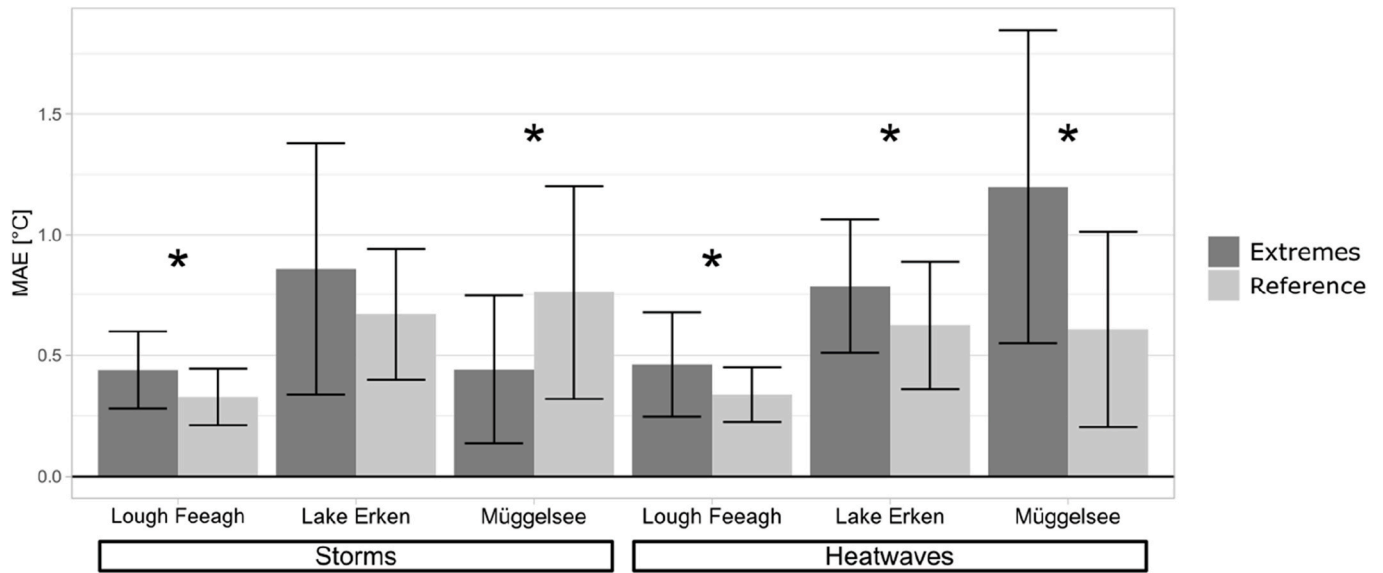


Fig. 4. Mean Absolute Error (MAE) of the water temperature profile, comparison between extreme and reference events. The bars denote the average MAE during extreme or reference events, averaged over all models, and the error bars represent one standard deviation. Statistical differences in MAE were tested with t-tests, or Wilcoxon rank sum tests for Lough Feeagh storms and Müggelsee heatwaves due to non-normal distributions. Statistically significant results ($p < 0.05$) are indicated with a * (Suppl. Mat G).

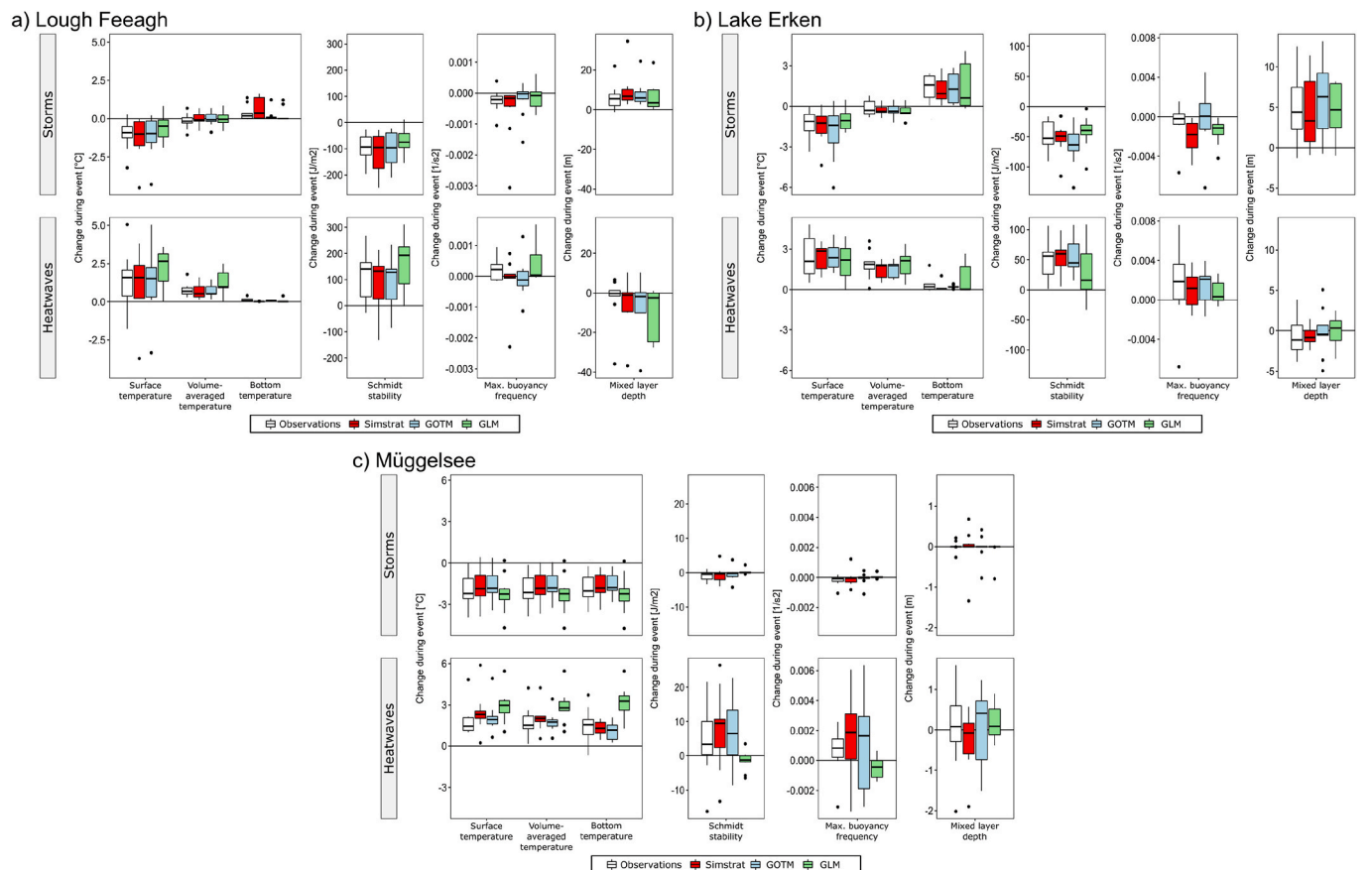


Fig. 5. Boxplots showing the change in temperature metrics, Schmidt stability, maximum buoyancy frequency, and mixed layer depth during the identified storms and heatwaves. The change during the event is calculated as the average over the two days before the event minus the average over the two days after the event. The boxplots show the changes in the observations (white), Simstrat (red), GOTM (blue), and GLM (green) for a) Lough Feeagh, b) Lake Erken, and c) Müggelsee. The boxplots show the median and first and third quartile. Whiskers extend to the smallest and largest value within 1.5 times the inter-quartile range from the nearest quartile (see geom_boxplot function in the ggplot2 R package, Wickham, 2016). Values outside this range are defined as outliers (•).

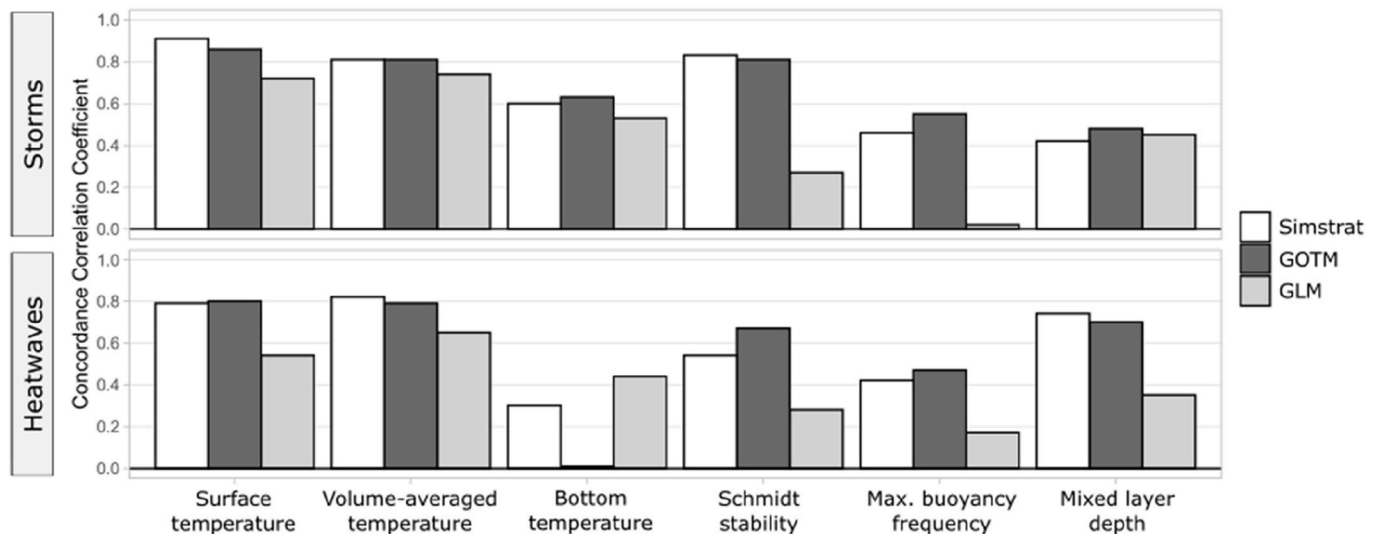


Fig. 6. Average concordance correlation coefficient (CCC) between simulated and observed changes in temperature and stratification metrics, for each model and metric during storms and heatwaves. For the exact values and a comparison between CCC and Pearson's correlation coefficient, see Suppl. Mat. H.

better than GLM (Fig. 6). The change in mixed layer depth during storms was reproduced with an average CCC of 0.5 for all models. The simulated changes during heatwaves had slightly lower performance for surface- and volume-averaged temperature than during storms (Fig. 6). During the heatwaves, Simstrat and GOTM performed better than GLM for all metrics, except for bottom temperature, where GOTM and Simstrat performed poorly. Simstrat and GOTM simulated the change in mixed layer depth better during heatwaves compared to storm events. On average, changes in surface temperature, volume-average temperature, and Schmidt stability were simulated more accurately than bottom temperature, maximum buoyancy frequency, and mixed layer depth.

Simstrat and GLM underestimated the increases in bottom temperature during heatwaves in Lough Feeagh (Kruskal-Wallis test, $Chi-sq = 21.1$, $p < 0.001$). GLM overestimated the increases in both surface (Kruskal-Wallis test, $Chi-sq = 8.2$, $p = 0.04$) and bottom temperatures (one-way ANOVA, $F = 9.1$, $p < 0.001$) during heatwaves in Müggelsee. None of the other extreme events showed a statistically significant difference in the mean change during the event between models and observations, for any metric. However, GLM underestimated the change in bottom temperature during reference wind events as well (Kruskal-Wallis test, $Chi-sq = 15.5$, $p = 0.001$). This was likely due to GLM showing very little heating of deep-water layers during the reference wind events in Lough Feeagh, resulting in low variance and a significant difference with observations. Some heating of bottom layers is expected even under non-extreme conditions, for example as a result of vertical turbulent diffusion (Livingstone, 1997).

Temporal cross-correlation could be performed for more than 80% of the events for Schmidt stability and volume-averaged and surface temperature, to calculate the simulation lag. Where the lags were calculated for these metrics, they were less than 1 h in more than 80% of the events (Suppl. Mat. I). For bottom temperature, maximum buoyancy frequency, and mixed layer depth, lags could only be calculated for about half of the events due to inaccurate simulations (see Material & Methods). About 70% (maximum buoyancy frequency) and 80% (bottom temperature and mixed layer depth) of the calculated lags were below 1 h. GLM was slightly worse in reproducing the timing of the simulations compared to the other models, but still had more than 50% (Schmidt stability, maximum buoyancy frequency, mixed layer depth) or more than 80% (temperature metrics) of the lags at or below 1 h. Differences between lakes varied per metric, but the timing of the simulations was not consistently better in any of the lakes.

In Suppl. Mat. J we show the temperature profiles and the corresponding detailed model simulations for three example events.

4. Discussion

In the present study we tested the model performance of three 1D models to capture responses to two kinds of extreme weather events in three different lakes. Firstly, we assessed the model performance during the generic validation period. The model fit for the full validation period was comparable to other studies (RMSE ranging from 0.5 to 2.0 °C, e.g. Fang et al., 2012; Stepanenko et al., 2013; Bruce et al., 2018; Moras et al., 2019; Schwefel et al., 2019). All models performed within the margins commonly found in literature, although GOTM and Simstrat performed better than GLM. A potential reason for this could have been a consequence of forcing GLM with hourly data, as compared to 10-min data for GOTM and Simstrat. However, this was found not to be the reason for the lower performance of GLM, because when GOTM and Simstrat were calibrated and run with hourly data, model errors were still about 40% lower than for GLM (Suppl. Mat. E). Model validation studies like this are valuable to better understand in which systems the models perform well and where they may have limitations. It could be that the different layer structure is beneficial for GOTM and Simstrat in this case of short-term extreme events, whereas GLM with adaptive layers may perform better in water bodies with fluctuating water levels. The different calibration routines between the models might also have influenced the model fit. More studies of this type are required to understand structural uncertainty in lake models (Frassl et al., 2019).

In agreement with previous studies (Jennings et al., 2012; Kasprzak et al., 2017; Andersen et al., 2020), wind events caused reduced Schmidt stability, deepened mixed layers, and cooled surface waters while the bottom water warmed in the two deep lakes (Lough Feeagh and Lake Erken). The shallow Müggelsee was always completely mixed during the storm events. Heatwaves are associated with increased surface water temperatures and stronger stratification (Jankowski et al., 2006; Jöhnk et al., 2008). In this study, temperatures in all water layers increased during the high temperature events. In Lough Feeagh and Lake Erken, the surface temperature increase was stronger than near the bottom and stratification strengthened. In Müggelsee, stratification occurred during most of the heatwave events, in line with the findings of Wilhelm and Adrian (2008). However, within two days after the heatwave events, stratification had reached levels similar to before the event. This caused

the temperature increase between two days before and two days after the events to be more or less uniform with depth.

In general, all models were able to reproduce the overall trends during either heating or wind events. Changes in surface and volume-averaged temperature and Schmidt stability were simulated most accurately, while changes in bottom temperatures especially during heatwaves were simulated less well. Also, the simulations of changes in maximum buoyancy frequency during storms and heatwaves, and of changes in mixed layer depth during heatwaves, were less accurate. The present study is amongst the first to look at model performance during short-term events. In the scenario study by [Mi et al. \(2018\)](#), GLM also simulated credible changes in hypolimnetic temperature, mixed layer depth, and Schmidt stability after a wind perturbation, although a comparison with observations during wind events was not performed.

In addition to reproducing the general trends, only in a few cases did models consistently over- or underestimate a change during events. Increases in bottom temperatures were underestimated during heatwaves in Lough Feeagh by Simstrat and GLM, which suggests that these models fail to adequately simulate increases in bottom temperatures in deep lakes, at least over the short time intervals evaluated here. However, the increases in Lough Feeagh bottom temperatures during heatwaves were only around 0.2 °C. GLM overestimated temperature increase in the whole water column during heatwaves in Müggelsee, often by more than 1 °C, while not showing such a bias over the full validation period. We have not explored further why only GLM showed this overestimation during heatwaves. It may be related to the combination of GLM's flexible grid structure and the depth of the lake, with Müggelsee being a shallow lake. The positive bias to warmer temperatures during a heatwave was not observed in the GLM simulations of Lough Feeagh and Lake Erken. This aligns with a GLM simulation of Lake Ammersee (mean depth 38.6 m), where surface temperature was also not overestimated during a heatwave year ([Bueche et al., 2017](#)).

As with the overall model performance in this study, GLM displayed higher model errors than Simstrat and GOTM during extreme events. Like the performance during the calibration and validation periods, we found that even when Simstrat and GOTM were forced with hourly inputs, these models still showed lower errors than GLM (Supp. Mat. G). The example results show that the surface heat fluxes had different values for each model (Suppl. Mat. J). This is partially the result of different calibration outcomes. The heat fluxes in the different models followed the same pattern, except for the longwave heat flux, which was notably different in GLM than in the other two models. This was likely due to a different parameterisation of the incoming longwave radiation. The behaviour of Simstrat and GOTM under extreme weather conditions was more similar to each other than to GLM (e.g. Suppl. Mat. J). This similarity is likely the result of a similar model structure, as both are k-epsilon turbulence models ([Rodi, 1980](#)), while GLM calculates mixing based on energy and density gradients (see [Hipsey et al., 2019](#)). The reason for using multiple models in this study was to ascertain if certain models performed significantly better than others, but also to provide results that are representative of 1D models in general, rather than any one particular model. Because all three models, despite their differences, tended to simulate the same general trends, but showed a higher MAE during extreme weather events, we can assume that strengths and weakness in event simulations found here are likely to occur to a similar extent in other 1D hydrodynamic lake models as well.

Most simulations captured the observed timing of the extreme events, that is, most of the effects were simulated within 1 h of the observations, and more than 90% of the modelled events had lags of less than 4 h, for all metrics. It should be noted, however, that we could only determine the lag if a reasonable model fit after the cross-correlation analysis was obtained (cross-correlation coefficient of 0.3 or higher). So, there is a bias towards events that were simulated well. For Schmidt stability, volume-averaged, and surface temperature, lags could be determined in 80–90% of the cases, but for the other metrics only in 40–60% of the cases. To our knowledge, accuracy of timing of short-

term events in hydrodynamic lake models has rarely been tested, yet it is a crucial aspect of model performance, especially for forecasting purposes. In studies aimed at forecasting phytoplankton blooms, timing is sometimes included in model assessment ([Gurkan et al., 2006](#); [Page et al., 2018](#)), and changes in hydrodynamics can be an important driver in phytoplankton dynamics ([Wilhelm and Adrian, 2008](#); [Kasprzak et al., 2017](#)).

Despite the reproduction of the overall trends, the low degree of bias, and the accurate timing of simulations, model error increased during extreme events compared to the reference periods by roughly 30% during storm events in Lough Feeagh, and during heatwaves by 30% (Lough Feeagh, Lake Erken) to 100% (Müggelsee). This lower performance shows that predictions made by hydrodynamic models during extreme weather events should be treated with additional caution. Notable exceptions were the storm events in Müggelsee, where the model error was 40% lower than during the reference periods. This likely has to do with the shallow depth of Müggelsee and might be systematic for shallow lakes in general; the selected storm events were some of the most extreme in a 14-year period and as a result this shallow lake mixed completely. This was correctly simulated by the models, and errors estimating these isothermal conditions tended to be lower than the errors than during the reference periods, when stratification sometimes occurred.

The larger errors during the storms in the deep lakes and during heatwaves can have multiple causes. Firstly, many of the models' parameterizations are nonlinear, and thus the magnitude of energy and turbulence fluxes might increase faster than linearly under more extreme conditions. By using high-frequency driving data, averaging errors relating to removing high frequency variation in meteorological forcing data were reduced. However, it is still possible that the values assigned to model coefficients during long-term calibration may not be appropriate for the extreme conditions of specific events and this would then automatically cause a larger error. Secondly, the assumption of one-dimensionality in the models holds less well during extreme events. During storms, the leeside of a lake and bays experience notably less wind forcing, internal waves can form, and wave breaking creates turbulence on underwater slopes ([Wüest et al., 2000](#); [MacIntyre and Jellison, 2001](#)). Shallow areas tend to stratify earlier and warm faster than deep areas ([Woolway and Merchant, 2018](#)), potentially creating more horizontal heterogeneity during heatwaves. These three-dimensional processes are not included in 1D models, and these sources of error may be accentuated during extreme events. Lastly, extreme events could also increase the importance of processes that were not included or kept constant in this study, such as precipitation, inflow, or turbidity.

We found that extreme weather generally resulted in momentarily less accurate simulation of lake conditions, even with high-frequency forcing data collected on-site, and with all three models. But to what extent is this a problem? Numerical process-based lake models are still amongst the best tools we have to simulate thermal dynamics in lakes during extreme weather events and the fact that uncertainty increases during these conditions does not invalidate their usefulness. In flood and hurricane forecasting, it is acknowledged that numerical models have large uncertainty during extreme weather conditions ([Todini, 2004](#); [Heming et al., 2019](#)). The uncertainty connected to these forecasts is an important aspect of the output that is included when informing decision makers and the public. In the case of extreme events in lakes, uncertainty can be taken into account partially by simply being aware of it. For example, since the timing of event impacts was simulated accurately, for some purposes of modelling it might be sufficient to take the timing of the event as information and knowing that the magnitude of the impact could differ from the simulations. However, to quantify the uncertainty during extreme events, a potential pathway would be ensemble modelling with forcing scenarios of varying intensity. Because we found little consistent bias, model runs with higher and lower wind speeds or temperatures could provide an uncertainty band during extreme weather events. More research would be needed to determine

what methods would be best suited to quantify uncertainty during extreme events.

The models in this study captured the overall trends, and the range of error during the extreme events (MAE 0.4–1.2 °C) is similar to the level of uncertainty found in other lake modelling studies during regular conditions (e.g. Soullignac et al., 2018; Moras et al., 2019). Larger model uncertainty during extreme events is, to a certain extent, expected because of greater spatial variations in lake thermal structure, larger energy fluxes, and more rapid changes in thermal gradients in the water column at small temporal scale, compared to non-extreme circumstances. It depends on the objective of the modeller if this reduced accuracy poses a problem. Larger error during extreme events might not pose a problem for long-term climate forecasting, as model fit during these short periods is generally not of interest for this type of studies. An exception to this statement would be if there are long-term consequences of extreme events, as in the case of tipping points (Scheffer et al., 2001). For short-term forecasting, however, extreme events are amongst the most important events to capture. This study shows that 1D lake models can be used to simulate these events, but the short-term predictions may be less precise than would occur under more normal conditions. This should be kept in mind when interpreting the forecasts. The results in the present study suggest that forecasts for temperature data and Schmidt stability will be more precise than for maximum buoyancy frequency and mixed layer depth. For scenario studies (as in Mi et al., 2018), the increased uncertainty during events is likely not a major issue. The absolute magnitude of the effect of an event might differ from observations, but the overall response is simulated. Coupling of physical models and biogeochemical models involves a risk of error propagation; a wrong estimation of water temperature could lead to wrong growth rates, or a too shallow mixing event results in less nutrient upwelling than in reality. Because of this, it is likely that uncertainty during extreme events also increases for biogeochemical models.

5. Conclusion

Extreme weather events are projected to increase in magnitude and frequency and can have large and diverse effects on lake ecosystems. One-dimensional hydrodynamic lake models could help in elucidating their impacts on lakes, but so far no studies have investigated how well these models perform during such events. In this study, Simstrat, GOTM, and GLM were run during multiple selected storms and heatwaves in three lakes in order to assess model performance. The overall effects of extreme weather on lake temperature and stratification metrics were captured by the models with correct timing and little bias, but the precision of the model output was reduced compared to non-extreme conditions. As with the model fit during calibration and validation, Simstrat and GOTM performed better during extreme events than GLM.

The implications of these findings ultimately depend on a modeller's objectives, but we are convinced that the findings in this paper can help to elucidate the uncertainty of model predictions during extreme weather events. This would lead to a more responsible use of 1D lake models, as uncertainty is an important part of model simulations. We propose that 1D lake models can be adequate tools to evaluate changes in hydrodynamics during extreme weather events, provided that the increased uncertainty during these events is kept in mind when interpreting the results.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their thanks to the Marine Institute, the Swedish

Infrastructure for Ecosystem Science (SITES) and Uppsala University, and the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB) for collecting and sharing the water temperature data of Lough Feeagh, Lake Erken, and Müggelsee.

J.P.M., A.I.A., and J.A.A.S. were funded by the MANTEL ITN (Management of climatic extreme events in lakes and reservoirs for the protection of ecosystem services) through the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 722518. R.A. acknowledges support by the "LimnoSCenES" Project (AD 91/22–1) within the BiodivERsA and the Belmont Forum Programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envsoft.2020.104852>.

Software availability

Instructions to download and install the models used in this study can be found at:

- Simstrat (version 2.1): <https://github.com/Eawag-AppliedSystemAnalysis/Simstrat/tree/v2.1>
- GOTM, (version 5.1): <https://gotm.net/>
- GLM, (version 3.0.4): <https://aed.see.uwa.edu.au/research/models/GLM/index.html>

Last access: 2020-08-20.

References

- Andersen, M.R., de Eyto, E., Dillane, M., Poole, R., Jennings, E., 2020. 13 Years of storms: An analysis of the effects of storms on lake physics on the atlantic fringe of europe. *Water* 12 (2), 318. <https://doi.org/10.3390/w12020318>.
- Bailey, L.D., Pol, M., 2016. Tackling extremes: Challenges for ecological and evolutionary research on extreme climatic events. *J. Anim. Ecol.* 85 (1), 85–96. <https://doi.org/10.1111/1365-2656.12451>.
- Bertani, I., Primicerio, R., Rossetti, G., 2016. Extreme climatic event triggers a lake regime shift that propagates across multiple trophic levels. *Ecosystems* 19 (1), 16–31. <https://doi.org/10.1007/s10021-015-9914-5>.
- Bruce, L.C., Frassl, M.A., Arhonditsis, G.B., Gal, G., Hamilton, D.P., Hanson, P.C., Hetherington, A.L., Melack, J.M., Read, J.S., Rinke, K., Rigosi, A., Trolle, D., Winslow, L., Adrian, R., Ayala, A.I., Bocaniov, S.A., Boehrer, B., Boon, C., Brookes, J. D., Bueche, T., Busch, B.D., Copetti, D., Cortés, A., de Eyto, E., Elliott, J.A., Gallina, N., Gilboa, Y., Guyennon, N., Huang, L., Kerimoglu, O., Lenters, J.D., MacIntyre, S., Makler-Pick, V., McBride, C.G., Moreira, S., Ozkundakci, D., Piloti, M., Rueda, F.J., Rusak, J.A., Samal, N.R., Schmid, M., Shatwell, T., Snorheim, C., Soullignac, F., Valerio, G., van der Linden, L., Vetter, M., Vinçon-Leite, B., Wang, J., Weber, M., Wickramaratne, C., Woolway, R.I., Yao, H., Hipsey, M.R., 2018. A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network. *Environ. Model. Software* 102, 274–291. <https://doi.org/10.1016/j.envsoft.2017.11.016>.
- Bueche, T., Hamilton, D.P., Vetter, M., 2017. Using the General Lake Model (GLM) to simulate water temperatures and ice cover of a medium-sized lake: a case study of Lake Ammersee, Germany. *Environmental Earth Sciences* 76 (13). <https://doi.org/10.1007/s12665-017-6790-7>.
- Chen, W., Nielsen, A., Andersen, T.K., Hu, F., Chou, Q., Søndergaard, M., Jeppesen, E., Trolle, D., 2020. Modeling the ecological response of a temporarily summer-stratified lake to extreme heatwaves. *Water* 12 (1), 94. <https://doi.org/10.3390/w12010094>.
- Crockford, L., Jordan, P., Melland, A., Taylor, D., 2015. Storm-triggered, increased supply of sediment-derived phosphorus to the epilimnion in a small freshwater lake. *Inland Waters* 5 (1), 15–26. <https://doi.org/10.5268/iw-5.1.738>.
- De Boyer Montégut, C., Madec, G., Fisher, A.S., Lazar, A., Iudicone, D., 2004. Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. *J. Geophys. Res.* 109 (C12), C12003. <https://doi.org/10.1029/2004jc002378>.
- De Eyto, E., Dillane, M., Moore, T., Wilson, H., Cooney, J., Hughes, P., Murphy, M., Nixon, P., Sweeney, D., Poole, R., 2020. Lough Feeagh water temperature profiles marine Institute. Ireland. URL. <http://data.marine.ie/geonetwork/srv/eng/catalog.search#/metadata/ie.marine.data:dataset.2817>.
- De Eyto, E., Jennings, E., Ryder, E., Sparber, K., Dillane, M., Dalton, C., Poole, R., 2016. Response of a humic lake ecosystem to an extreme precipitation event: physical, chemical, and biological implications. *Inland Waters* 6 (4), 483–498. <https://doi.org/10.1080/IW-6.4.875>.

- Dinno, A., 2017. *dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums*, 5 ed, Version 1.3.
- Doherty, J., 2015. Calibration and Uncertainty Analysis for Complex Environmental Models. *Watermark Numerical Computing*.
- Driescher, E., Behrendt, H., Schellenberger, G., Stellmacher, R., 1993. Lake Müggelsee and its environment—natural conditions and anthropogenic impacts. *Internationale Revue der gesamten Hydrobiologie und Hydrographie* 78 (3), 327–343.
- Fang, X., Alam, S.R., Stefan, H.G., Jiang, L., Jacobson, P.C., Pereira, D.L., 2012. Simulations of water quality and oxythermal cisco habitat in Minnesota lakes under past and future climate scenarios. *Water Qual. Res. J. Can.* 47, 3–4. <https://doi.org/10.2166/wqrj.2012.031>.
- Frassl, M.A., Abell, J.M., Botelho, D.A., Cinque, K., Gibbes, B.R., Jöhnk, K.D., Muraoka, K., Robson, B.J., Wolski, M., Xiao, M., Hamilton, D.P., 2019. A short review of contemporary developments in aquatic ecosystem modelling of lakes and reservoirs. *Environ. Model. Software* 117, 181–187. <https://doi.org/10.1016/j.envsoft.2019.03.024>.
- Gaudard, A., Råman Vinnå, L., Bärenbold, F., Schmid, M., Bouffard, D., 2019. Toward an open access to high-frequency lake modeling and statistics data for scientists and practitioners—the case of Swiss lakes using Simstrat v2.1. *Geosci. Model Dev. (GMD)* 12 (9), 3955–3974. <https://doi.org/10.5194/gmd-12-3955-2019>.
- Goudsmit, G.H., Burchard, H., Peeters, F., Wüest, A., 2002. Application of $k-\epsilon$ turbulence models to enclosed basins: the role of internal seiches. *J. Geophys. Res.: Oceans* 107 (C12). <https://doi.org/10.1029/2001JC000954>.
- Gurkan, Z., Zhang, J., Jørgensen, S.E., 2006. Development of a structurally dynamic model for forecasting the effects of restoration of Lake Fure, Denmark. *Ecol. Model.* 197 (1–2), 89–102. <https://doi.org/10.1016/j.ecolmodel.2006.03.006>.
- Heming, J.T., Prates, F., Bender, M.A., Bowyer, R., Cangialosi, J., Caroff, P., Coleman, T., Doyle, J.D., Dube, A., Faure, G., Fraser, J., Howell, B.C., Igarashi, Y., McTaggart-Cowan, R., Mohapatra, M., Moskaitis, J.R., Murtha, J., Rivett, R., Sharma, M., Short, C.J., Singh, A.A., Tallapragada, V., Titley, H.A., Xiao, Y., 2019. Review of recent progress in tropical cyclone track forecasting and expression of uncertainties. *Tropical Cyclone Research and Review* 8 (4), 181–218. <https://doi.org/10.1016/j.tcr.2020.01.001>.
- Hilt, S., Henschke, I., Rucker, J., Nixdorf, B., 2010. Can submerged macrophytes influence turbidity and trophic state in deep lakes? Suggestions from a case study. *J. Environ. Qual.* 39 (2), 725–733. <https://doi.org/10.2134/jeq2009.0122>.
- Hipsey, M.R., Bruce, L.C., Boon, C., Busch, B., Carey, C.C., Hamilton, D.P., Hanson, P.C., Read, J.S., de Sousa, E., Weber, M., Winslow, L.A., 2019. A General Lake model (GLM 3.0) for linking with high-frequency sensor data from the global lake ecological observatory network (GLEON). *Geosci. Model Dev. (GMD)* 12 (1), 473–523. <https://doi.org/10.5194/gmd-12-473-2019>.
- Hipsey, M.R., Gal, G., Arhonditsis, G.B., Carey, C.C., Elliott, J.A., Frassl, M.A., Janse, J.H., de Mora, L., Robson, B.J., 2020. A system of metrics for the assessment and improvement of aquatic ecosystem models. *Environ. Model. Software* 128. <https://doi.org/10.1016/j.envsoft.2020.104697>.
- Huber, V., Adrian, R., Gerten, D., 2010. A matter of timing: heat wave impact on crustacean zooplankton. *Freshw. Biol.* <https://doi.org/10.1111/j.1365-2427.2010.02411.x>.
- Huber, V., Wagner, C., Gerten, D., Adrian, R., 2012. To bloom or not to bloom: contrasting responses of cyanobacteria to recent heat waves explained by critical thresholds of abiotic drivers. *Oecologia* 169 (1), 245–256. <https://doi.org/10.1007/s00442-011-2186-7>.
- Idso, S.B., 1973. On the concept of lake stability. *Limnol. Oceanogr.* 18 (4), 681–683. <https://doi.org/10.4319/lo.1973.18.4.0681>.
- IPCC, 2014. *Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change*. In: Pachauri, R.K., Meyer, L.A. (Eds.), *Core Writing Team. IPCC, Geneva, Switzerland*, p. 151.
- Jankowski, T., Livingstone, D.M., Bührer, H., Forster, R., Niederhauser, P., 2006. Consequences of the 2003 European heat wave for lake temperature profiles, thermal stability, and hypolimnetic oxygen depletion: implications for a warmer world. *Limnol. Oceanogr.* 51 (2), 815–819.
- Jennings, E., Jones, S., Arvola, L., Staehr, P.A., Gaiser, E., Jones, I.D., Weathers, K.C., Weyhenmeyer, G.A., Chiu, C.Y., De Eyto, E., 2012. Effects of weather-related episodic events in lakes: An analysis based on high-frequency data. *Freshw. Biol.* 57 (3), 589–601. <https://doi.org/10.1111/j.1365-2427.2011.02729.x>.
- Jöhnk, K.D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biol.* 14 (3), 495–512. <https://doi.org/10.1111/j.1365-2486.2007.01510.x>.
- Johnson, S.G., 2014. *The NLOpt Nonlinear-Optimization Package*.
- Kasprzak, P., Shatwell, T., Gessner, M.O., Gonsiorczyk, T., Kirillin, G., Selmeçy, G., Padišák, J., Engelhardt, C., 2017. Extreme weather event triggers cascade towards extreme turbidity in a clear-water lake. *Ecosystems* 20 (8), 1407–1420. <https://doi.org/10.1007/s10021-017-0121-4>.
- Kelly, S., Doyle, B., de Eyto, E., Dillane, M., McGinnity, P., Poole, R., White, M., Jennings, E., 2020. Impacts of a record-breaking storm on physical and biogeochemical regimes along a catchment-to-coast continuum. *PLoS One* 15 (7), e0235963. <https://doi.org/10.1371/journal.pone.0235963>.
- Klug, J.L., Richardson, D.C., Ewing, H.A., Hargreaves, B.R., Samal, N.R., Vachon, D., Pierson, D.C., Lindsey, A.M., O'Donnell, D.M., Effler, S.W., Weathers, K.C., 2012. Ecosystem effects of a tropical cyclone on a network of lakes in northeastern North America. *Environmental science & technology* 46 (21), 11693–11701. <https://doi.org/10.1021/es302063v>.
- Koenings, J.P., Edmundson, J.A., 1991. Secchi disk and photometer estimates of light regimes in Alaskan lakes: Effects of yellow color and turbidity. *Limnol. Oceanogr.* 36 (1), 91–105.
- Kuha, J., Arvola, L., Hanson, P.C., Huotari, J., Huttula, T., Juntunen, J., Järvinen, M., Kallio, K., Ketola, M., Kuoppamäki, K., Lepistö, A., Lohila, A., Paavola, R., Vuorenmaa, J., Winslow, L., Karjalainen, J., 2016. Response of boreal lakes to episodic weather-induced events. *Inland Waters* 6 (4), 523–534. <https://doi.org/10.1080/1W-6.4.886>.
- Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 255–268.
- Livingstone, D.M., 1997. An example of the simultaneous occurrence of climate-driven “sawtooth” deep-water warming/cooling episodes in several Swiss lakes. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verh. Proc. Trav. SIL* 26 (2), 822–828. <https://doi.org/10.1080/03680770.1995.11900832>.
- MacIntyre, S., Jellison, R., 2001. *Nutrient Fluxes from Upwelling and Enhanced Turbulence at the Top of the Pycnocline in Mono Lake, California, Saline Lakes*. Springer, pp. 13–29.
- Martin, J.L., McCutcheon, S.C., 1999. *Hydrodynamics and Transport for Water Quality Modeling*. CRC press.
- Met Éireann, 2018. *Met Éireann Data Licensed under a Creative Commons Attribution-ShareAlike 4.0 International Licence*.
- Mi, C., Frassl, M.A., Boehrer, B., Rinke, K., 2018. Episodic wind events induce persistent shifts in the thermal stratification of a reservoir (Rappbode Reservoir, Germany). *Int. Rev. Hydrobiol.* 103 (3–4), 71–82. <https://doi.org/10.1002/iroh.201701916>.
- Moras, S., Ayala, A.I., Pierson, D.C., 2019. Historical modelling of changes in Lake Erken thermal conditions. *Hydrol. Earth Syst. Sci.* 23 (12), 5001–5016. <https://doi.org/10.5194/hess-23-5001-2019>.
- Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. *Comput. J.* 7 (4), 308–313.
- Page, T., Smith, P.J., Beven, K.J., Jones, I.D., Elliott, J.A., Maberly, S.C., Mackay, E.B., De Ville, M., Feuchtmayr, H., 2018. Adaptive forecasting of phytoplankton communities. *Water Res.* 134, 74–85. <https://doi.org/10.1016/j.watres.2018.01.046>.
- Perga, M.E., Bruel, R., Rodriguez, L., Guénand, Y., Bouffard, D., 2018. Storm impacts on alpine lakes: Antecedent weather conditions matter more than the event intensity. *Global Change Biol.* <https://doi.org/10.1111/gcb.14384>.
- Persson, I., Jones, I.D., 2008. The effect of water colour on lake hydrodynamics: A modelling study. *Freshw. Biol.* 53 (12), 2345–2355. <https://doi.org/10.1111/j.1365-2427.2008.02049.x>.
- R Core Team, 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Read, J.S., Hamilton, D.P., Jones, I.D., Muraoka, K., Winslow, L.A., Kroiss, R., Wu, C.H., Gaiser, E., 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environ. Model. Software* 26 (11), 1325–1336. <https://doi.org/10.1016/j.envsoft.2011.05.006>.
- Rodi, W., 1980. *Turbulence Models and Their Application in Hydraulics - A State-Of-The-Art Review*. CRC Press.
- Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. *Nature* 413 (6856), 591. <https://doi.org/10.1038/35098000>.
- Schmidt, W., 1928. Über die Temperatur- und Stabilitätsverhältnisse Von Seen. *Geogr. Ann.* 10 (1–2), 145–177. <https://doi.org/10.1080/20014422.1928.11880475>.
- Schwefel, R., Müller, B., Boisgontier, H., Wüest, A., 2019. Global warming affects nutrient upwelling in deep lakes. *Aquat. Sci.* 81 (3), 50. <https://doi.org/10.1007/s00027-019-0637-0>.
- Soares, L.M.V., Silva, T.F.d.G., Vinçon-Leite, B., Eleutério, J.C., Lima, L.C.d., Nascimento, N.d.O., 2019. Modelling drought impacts on the hydrodynamics of a tropical water supply reservoir. *Inland Waters* 1–16. <https://doi.org/10.1080/20442041.2019.1596015>.
- Soranno, P.A., Carpenter, S.R., Lathrop, R.C., 1997. Internal phosphorus loading in Lake Mendota: Response to external loads and weather. *Can. J. Fish. Aquat. Sci.* 54 (8), 1883–1893. <https://doi.org/10.1139/f97-095>.
- Soullignac, F., Danis, P.-A., Bouffard, D., Chanut, V., Dambrine, E., Guénand, Y., Harmel, T., Ibelings, B.W., Trevisan, D., Uittenboogaard, R., Anneville, O., 2018. Using 3D modeling and remote sensing capabilities for a better understanding of spatio-temporal heterogeneities of phytoplankton abundance in large lakes. *J. Great Lake Res.* 44 (4), 756–764. <https://doi.org/10.1016/j.jglr.2018.05.008>.
- Stepanenko, V.M., Martynov, A., Jöhnk, K.D., Subin, Z.M., Perroud, M., Fang, X., Beyrich, F., Mironov, D., Goyette, S., 2013. A one-dimensional model intercomparison study of thermal regime of a shallow, turbid midlatitude lake. *Geosci. Model Dev. (GMD)* 6 (4), 1337–1352. <https://doi.org/10.5194/gmd-6-1337-2013>.
- Stockwell, J.D., Doubek, J.P., Adrian, R., Anneville, O., Carey, C.C., Carvalho, L., De Senerpont Domis, L.N., Dur, G., Frassl, M.A., Grossart, H.P., Ibelings, B.W., Lajeunesse, M.J., Lewandowska, A.M., Llamas, M.E., Matsuzaki, S.I.S., Nodine, E.R., Nøges, P., Patil, V.P., Pomati, F., Rinke, K., Rudstam, L.G., Rusak, J.A., Salmaso, N., Seltmann, C.T., Straile, D., Thackeray, S.J., Thiery, W., Urrutia-Cordero, P., Venail, P., Verburg, P., Woolway, R.I., Zohary, T., Andersen, M.R., Bhattacharya, R., Hejzlar, J., Janatian, N., Kpodonu, A.T.N.K., Williamson, T.J., Wilson, H.L., 2020. Storm impacts on phytoplankton community dynamics in lakes. *Global Change Biol.* 26 (5), 2756–2784. <https://doi.org/10.1111/gcb.15033>.
- Todini, E., 2004. Role and treatment of uncertainty in real-time flood forecasting. *Hydrol. Process.* 18 (14), 2743–2746. <https://doi.org/10.1002/hyp.5687>.
- Umlauf, L., Burchard, H., Bolding, K., 2005. *GOTM: Sourcecode and Test Case Documentation*.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Wilhelm, S., Adrian, R., 2008. Impact of summer warming on the thermal characteristics of a polymictic lake and consequences for oxygen, nutrients and phytoplankton. *Freshw. Biol.* 53 (2), 226–237. <https://doi.org/10.1111/j.1365-2427.2007.01887.x>.

- Wilhelm, S., Hintze, T., Livingstone, D.M., Adrian, R., 2006. Long-term response of daily epilimnetic temperature extrema to climate forcing. *Can. J. Fish. Aquat. Sci.* 63 (11), 2467–2477. <https://doi.org/10.1139/F06-140>.
- Wilson, H.L., Ayala, A.I., Jones, I.D., De Eyto, E., Grossart, H.-P., Perga, M.-E., Pierson, D., Rolston, A., Woolway, R.I., Jennings, E., Submitted. Variability in epilimnion depth estimations in lakes. *Hydrol. Earth Syst. Sci.* doi:10.5194/hess-2020-222.
- Winslow, L.A., Read, J.S., Woolway, R.I., Brenttrup, J.A., Leach, T., Zwart, J., Albers, S., Collinge, D., 2019. rLakeAnalyzer: Lake Physics Tools. R Package Version 1.11.4.1.
- Woolway, R.I., Merchant, C.J., 2018. Intralake heterogeneity of thermal responses to climate change: a study of large northern hemisphere lakes. *J. Geophys. Res.: Atmosphere* 123 (6), 3087–3098. <https://doi.org/10.1002/2017jd027661>.
- Wüest, A., Piepke, G., Van Senden, D.C., 2000. Turbulent kinetic energy balance as a tool for estimating vertical diffusivity in wind-forced stratified waters. *Limnol. Oceanogr.* 45 (6), 1388–1400. <https://doi.org/10.4319/lo.2000.45.6.1388>.