Water Research X 30 (2026) 100457

Contents lists available at ScienceDirect

Water Research X

journal homepage: www.sciencedirect.com/journal/water-research-x

How climate change erodes short-term lake-temperature predictability: Informing climate resilient lake forecasting

D. Atton Beckmann [a,*], M. Werther [b], T. Shatwell [c], E. Spyrakos [a], P. Hunter [a,d], I.D. Jones [a]

[a] Biological and Environmental Sciences, School of Natural Sciences, University of Stirling, Stirling, United Kingdom
[b] Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters - Research and Management, Dübendorf, Switzerland
[c] Helmholtz Centre for Environmental Research (UFZ), Department of Lake Research, Brückstrasse 3a, 39114 Magdeburg, Germany
[d] Scotland's International Environment Centre, School of Natural Sciences, University of Stirling, Stirling, United Kingdom

ARTICLE INFO

Keywords:
Forecasting
Machine learning
Water quality
Climate change

ABSTRACT

Climate warming threatens short-term environmental forecast skill, yet its effect on water quality predictability is largely unquantified. Here, we demonstrate a new approach for assessing climate change effects on lake forecasts. Random forest (RF) and gated recurrent unit network models were trained on data from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) Local Lakes Sector (five central-European lakes, four hydrodynamic models) and then used to forecast daily lake surface temperature 14 days ahead for 2060 - 2100 under four climate scenarios. We then varied (i) sensor-sampling interval (3, 7, 14 days) and (ii) training-set length (1 - 30 years). Under the strongest forcing (SSP585), the summer mean absolute error (MAE) of worst-affected lake, Esthwaite, rose by 0.14 °C (from 1.75 to 1.89 °C), driven by higher day-to-day temperature volatility ($R^2 = 0.78$). For this lake, extending the training set from 5 to 20 years or shortening sampling from 14 to 3 days reduced summer MAE by 0.11 and 0.17 °C, effectively offsetting the volatility caused by climate change. In winter, forecast error declined for four lakes because warmer, more stratified conditions simplified surface-layer dynamics. Thus, modest investments in monitoring cadence or historical record length can preserve forecast skill, even under extreme climate change. More broadly, this highlights a largely unexplored potential use for climate scenario projections: informing the design of climate resilient lake monitoring systems.

1. Introduction

Climate change is altering freshwater ecosystems globally (Capon et al., 2021). Rising lake surface temperatures affect metabolic rates, species survival, and biogeography (Dokulil et al., 2021), while shifts in stratification and mixing influence light, carbon, and oxygen dynamics (Woolway et al., 2021). Increased frequency and severity of droughts and extreme precipitation events are also projected to increase (Dai, 2013; Tabari, 2020), heightening the risk of harmful algal blooms in many lakes (Paerl et al., 2020). Simultaneously, water quality monitoring is advancing, with automated sensors complementing manual sampling and generating high-frequency, high-volume datasets (Marcé et al., 2016). Machine learning (ML) is increasingly applied to manage this complexity - both for real-time monitoring and short-term forecasting of variables such as water temperature (Di Nunno et al., 2023; Yousefi and Toffolon, 2022), dissolved oxygen (Ziyad Sami et al., 2022), and algal abundance (Schaeffer et al., 2024). By generating daily to weekly forecasts, ML can support real-time management, allowing for timely execution of interventions such as water supply switching, aeration system activation, or issue of public health advisories (Stroom and Kardinaal, 2016).

As water quality forecasting capabilities increase, it is likely that some water managers will become increasingly reliant on such approaches to ensure consistent and safe provision of many of the ecosystem services provided by freshwater bodies, such as water supply, recreation, food, and carbon sequestration (Carey et al., 2025; Inácio et al., 2022). However, in the context of a rapidly changing climate, it is possible that short-term water quality forecasting will become more inaccurate for several reasons. Firstly, under climate change, water bodies should be considered non-stationary systems, and will therefore become inherently less predictable than they would be given a more stable climate (Milly et al., 2008; Rollinson et al., 2021). Secondly, data-driven forecasting approaches such as ML rely on representative training datasets, typically historical monitoring data, to make future

* Corresponding author.
E-mail address: daniel.atton.beckmann@stir.ac.uk (D. Atton Beckmann).

https://doi.org/10.1016/j.wroa.2025.100457
Received 16 September 2025; Received in revised form 18 November 2025; Accepted 22 November 2025
Available online 23 November 2025

forecasts. With a sufficiently large training dataset, performance improvements tend to show diminishing returns as the length of the training data is increased (Atton Beckmann et al., 2025; Last, 2007). Importantly, this effect may be stronger for non-stationary problems as older training data may be less representative of the present system, and therefore less useful for informing predictions, particularly if the system is changing rapidly (Valavi et al., 2022). Lastly, the forecast uncertainty associated with certain meteorological variables such as precipitation is expected to increase and this might in turn affect water quality forecast ability (Xu et al., 2020).

Lake surface water temperature is a fundamental variable in aquatic systems because it strongly influences numerous physical, chemical, and biological processes. These include dissolved oxygen dynamics (Hutchings et al., 2024), algal bloom development (Glibert and Burkholder, 2018), fish habitat suitability (Fang and Stefan, 2012), and the spread of pathogens (van der Wielen et al., 2023). Consequently, accurate forecasting of lake surface water temperature is central to effective lake and water resource management, particularly under changing climatic conditions (Piccolroaz et al., 2024).

In recent years, there has been growing interest in applying ML techniques to model and forecast lake surface water temperature across a range of spatial and temporal scales. Many studies have focused on short-term prediction or *nowcasting*, often comparing the performance of different ML algorithms (Heddam et al., 2020; Liu and Chen, 2012; Quan et al., 2022; Yousefi and Toffolon, 2022). Others have developed hybrid approaches that integrate ML with process-based models to improve temperature simulations (Jia et al., 2021; Read et al., 2019). Additional work has explored ML-based approaches for both short- and long-term forecasting (Saber et al., 2020; Zhu et al., 2020). However, the vast majority of these studies have assessed model performance only under present-day or historical conditions. A few have considered limited cross-condition testing - for instance, Read et al. (2019) examined models trained on relatively cold years and tested on warmer ones - but systematic evaluation of ML model performance under projected future climate scenarios remains largely unexplored.

This study addresses this gap by leveraging lake-specific surface water temperature projections available through the Lakes Sector of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP; Golub et al., 2022) to investigate how climate change may affect ML short-term forecasting skill. Understanding this relationship is crucial, as some variables and lakes are likely to become increasingly difficult to forecast under future climatic conditions (Milly et al., 2008). Beyond quantifying this effect, it is also essential to assess whether modifications to the monitoring or forecasting approach can compensate for potential performance degradation. Because ML approaches tend to benefit from longer training datasets (Atton Beckmann et al., 2025; Mahmood et al., 2022), it is important to understand monitoring requirements for climate-resilient forecasting as soon as possible to give sufficient time for data collection.

Therefore, this study presents an approach for informing the design of water quality monitoring systems that can provide sufficient data for training data-driven short-term forecast models, even in extreme climate change scenarios. To demonstrate this, we investigate how the performance of short-term ML forecasts of surface water temperature is affected under different climate change scenarios for five moderately-sized northern temperate lakes. While the lakes in this study are geographically specific, this serves as a proof of concept for the approach and offers valuable insights relevant to many northern temperate lakes. Additionally, by comparing performance across different sampling frequencies and training dataset durations, we also test how climate-driven performance degradation can be mitigated by adjusting the monitoring strategy. Therefore, the key research questions of this study are:

1. Does the performance of ML forecasts of surface water temperature reduce in more extreme climate change scenarios?
2. Are the benefits of increased training dataset duration lessened in more extreme climate change scenarios?
3. Can management strategies such as increasing sampling frequency compensate for the effects of more extreme climate change scenarios on forecast performance?
4. Which variables can be considered drivers of any link between forecast performance and climate scenario?

Ultimately, by addressing these questions, we aim to demonstrate the value of using climate projections and modelled lake data to inform the design of water quality monitoring and forecasting systems in the context of a rapidly changing climate.

2. Materials and methods

2.1. Dataset and study sites

The ISIMIP framework provides both historical and projected climate and socioeconomic forcing data over a range of different policy relevant climate scenarios. The ISIMIP Lake Sector has used these data to force physical lake models both for 'general' lakes across the globe, and 'local' lakes which have been calibrated to a smaller number of specific lakes (Golub et al., 2022). This study uses ISIMIP Local Lakes data from five lakes located in England, Ireland, and Germany (Table S1, Figure S1, Figure S2). These lakes were chosen as they are all located in the northern temperate region, and are moderately sized (~1–15 km^2) and therefore can be considered representative of a large number of lakes in this part of the world given the high prevalence of lakes this size (Cael and Seekell, 2016). Daily surface water temperature projections from the ISIMIP3b simulation round for each of these lakes from 2015–2100 were downloaded under four different scenarios, including a pre-industrial control (PIC) and three of the shared socioeconomic pathways (SSP) described by O'Neill et al. (2016) (Table 1).

For each scenario, projections were obtained from four different physics-driven, one-dimensional lake models: Flake (Mironov et al., 2010), General Lake Model (GLM) (Hipsey et al., 2019), General Ocean Turbulence Model (GOTM) (Umlauf et al., 2005), and Simstrat (Gaudard et al., 2019); and five different ensemble climate models: GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1) (Dunne et al., 2020), IPSL-CM6A-LR (Boucher et al., 2020), Max Planck Institute Earth System Model (MPI-ESM1.2) (Gutjahr et al., 2019), Meteorological Research Institute Earth System Model Version 2.0 (MRI-ESM2.0) (Yukimoto et al., 2019), and the UK Earth System Model (UKESM) (Sellar et al., 2019).

Alongside surface water temperature data, daily meteorological data were also downloaded from the ISIMIP repository for the same five climate models. Specifically, these variables were: near-surface relative humidity, near-surface specific humidity, precipitation, snowfall flux, surface air pressure, surface downwelling longwave radiation, surface downwelling shortwave radiation, near-surface wind speed, near-surface air temperature, daily-maximum near-surface air temperature

Table 1
Climate change scenarios used for experiments.

Scenario	Description	Anthropogenic Radiative Forcing in 2100 (Wm^{-2})
PIC	Model driven by invariant solar, greenhouse gases, ozone, tropospheric aerosol and land-use forcings from the year 1850.	0
SSP126	'Sustainable and green' pathway where substantial climate change mitigation actions are taken.	2.6
SSP370	'Regional rivalry' pathway where climate change action is not prioritised.	7.0
SSP585	'Fossil-fuelled development'. Most extreme climate change scenario.	8.5

and daily minimum near-surface air temperature.

2.2. Forecasting model and approach

This study used entirely modelled data from ISIMIP3b lake simulations both for training and evaluation. This enabled systematic analysis of the effects of climate change on forecast skill under well-controlled conditions, independent of observational gaps or other inconsistencies. The effect of climate change was explored through a comparison of forecast performance in more extreme scenarios with the PIC scenario used as a baseline. Therefore, the 2100 anthropogenic radiative forcing associated with each scenario (Table 1) is used as the independent variable associated with climate change for the experimental procedure.

To ensure that the day-to-day variation of the ISIMIP-driven lake model outputs was sufficiently realistic to be a good test of forecast skill, surface water temperature measurements provided by the United Kingdom Centre for Ecology and Hydrology from Windermere (hourly data from 2016–18) and Esthwaite (fortnightly data from 2015–18), were compared with simulation outputs (Feuchtmayr et al., 2021b, 2021a). First-order time series differencing was used to assess the daily variability for Windermere and fortnightly variability for Esthwaite. Differencing involves calculating the change in temperature between successive time steps (e.g., day-to-day or fortnight-to-fortnight), providing a measure of short-term variability. The mean of the absolute differenced values and the standard deviation of the raw differenced time series were calculated for the in situ data and the corresponding simulated data from all four lake models, providing a comparison of the magnitude and variability of short-term temperature changes. These variability metrics for the observed and simulated time series matched closely for both lakes, indicating that the simulated data were sufficiently realistic for testing forecasting skill (Figure S3, Table S2). Additionally, the mean absolute error (MAE) between the in situ and simulated time series was sufficiently low (average = 1.75 °C) to indicate that this data captured well the seasonal lake variation. However, it is important to note that a very close match between the simulated and in situ records is not expected, as the simulated time series represents a climate-scenario projection rather than a hindcast of observed conditions.

Two ML forecasting models were used for the core experiments of the study. Firstly, as an example of a widely used *classical* ML approach, a Random Forest (RF) model implemented using Scikit Learn was used (Breiman, 2001; Pedregosa et al., 2011). Specifically, the RF was chosen because it has been shown to be effective for similar short-term forecasting tasks even with a training dataset of relatively short duration (Atton Beckmann et al., 2025; Bhogal et al., 2023; Giamalaki et al., 2022). In addition to the RF, a Gated Recurrent Unit (GRU) network implemented in *Tensorflow* was used (Martín Abadi et al., 2015). GRUs are a type of recurrent neural network which are generally very effective for time series forecasting tasks as they explicitly account for temporal dependencies in sequential data. Compared to Long Short-Term Memory (LSTM) networks, which are widely used for similar forecasting tasks, GRUs provide similar predictive capability whilst generally requiring smaller training datasets due to their streamlined architecture (Chung et al., 2014; Waqas and Humphries, 2024).

Prior to running forecasting experiments, input data were downsampled by keeping every T^{th} observation and discarding the others for three different sampling periods, $T = 3, 7, 14$ days. All forecasts were made 2 weeks ahead, using data from the most recent 28 days as input. Therefore, for 14-day data, the model would have access to data from the two most recent observations to make a prediction, whereas for 3-day data, the model would have access to the nine most recent observations. Each observation consisted of measurements of all eleven meteorological variables, as well as the surface water temperature estimate from that day.

The RF and GRU hyperparameters were tuned using modelled data from all lakes, lake models climate models, and scenarios for the period from 2015–2060 using 100 iterations of a randomised search with five-fold cross validation. The resulting tuned parameters are detailed in Table S3.

2.3. Statistical analysis

Forecast runs were carried out for the test years 2061–2100 at the three different sampling frequencies for all four climate scenarios and each of the lake/climate model combinations. For each test year, the model was trained using the N years directly prior to the test year for a range of different training dataset durations, $N = 1, 5, 10, 15, 20, 30$ years. As surface water temperatures were expected to be, on average, higher in the more extreme climate change scenarios, for statistical analysis it was necessary to use a relative error metric which does not scale with the magnitude of the predicted variable. Therefore, mean absolute percentage error (MAPE) was used to evaluate performance as this is a widely used relative error metric (Koutsandreas et al., 2022). Alongside MAPE, MAE was also calculated as this is a widely used non-normalised average error metric (Willmott and Matsuura, 2005).

A linear mixed model (LMM) was used to test the following:

1) is forecast performance reduced in more extreme climate change scenarios?
2) do the benefits of training dataset duration lessen in more extreme climate change scenarios?
3) can increasing sampling frequency or training dataset duration compensate for performance reductions in more extreme climate change scenarios?

LMM statistical significance was calculated using Satterthwaite's method to estimate degrees of freedom in order to obtain p-values (Kuznetsova et al., 2017). MAPE was used as the dependent variable in the LMM, with a natural logarithm transform used to approximate normality. Climate scenario was included as a fixed effect to understand how this affects forecast performance. In addition, training dataset duration and its interaction with climate scenario were included as fixed effects to understand if the benefits of long training datasets diminish in more extreme climate scenarios. Prior to analysis, the training dataset duration variable was inverse transformed to obtain an approximately linear relationship with the error. Further, sampling frequency and its interaction with training dataset duration were included as fixed effects to understand if increasing sampling frequency could compensate for any negative effects on performance due to more extreme climate change scenarios. Additionally, season and its interaction with climate scenario were included as fixed effects as this was flagged in preliminary data exploration as an important interaction. Test year was also included as a fixed effect to understand if forecast performance could be expected to continually degrade further into the future. Finally, climate model, lake model and lake were all included as random effects to understand their contribution to the variance of forecast performance. This resulted in the following model specification:

$$y = \beta_0 + \beta_1 N + \beta_2 S + \beta_3 R + \beta_4 T + \beta_5 Y + \gamma_1 NS + \gamma_2 SR + \gamma_3 TN + u_C + u_{LM} + u_L + \varepsilon \quad (1)$$

Where y is the dependent variable, log*(MAPE)*; N the inverse of the number of training years; S the climate scenario; R the season; T the sampling period; Y the test year; β_{0-5} the fixed effect coefficients; γ_{1-3} the interaction effect coefficients; u_C, u_{LM}, u_L the random effects associated with climate model, lake model and lake respectively, and ε the residual error. To make meaningful comparisons between the effect sizes of each of the fixed effects, all variables used in the LMM were first standardised to have a mean of zero and standard deviation of one. Following standardisation, two LMMs were fitted – one for the outputs from the GRU model, and one for the RF, allowing cross-model comparison of effect sizes.

As an additional demonstration of the practical effect of climate scenario on performance, the maximum summer forecast error was compared between scenarios for each model. Maximum error was defined as the largest absolute difference between simulated and observed surface water temperature within each summer season. This metric was selected because in summer months extreme warm events can, for example, exceed thermal limits for freshwater fish or trigger other heat-related impacts and are therefore of significant management concern (Till et al., 2019). Maximum summer error therefore provides a conservative measure of model reliability under conditions critical for decision-making (Duan et al., 2024; Woolway et al., 2022).

Finally, additional linear regression analyses were used to explore the relationships between the suspected drivers of forecast error and climate change scenario. Firstly, to understand which variables to focus on, a feature importance analysis was carried out using a permutation importance approach using MAE as the error metric (Breiman, 2001; Ernst, 2004). The highest ranking features accounting for the vast majority (> 90 %) of the total increase in MAE found in the permutation importance analysis were selected for further analysis. These variables were then aggregated in two ways: 1) using monthly averages, and 2) using monthly averages of the absolute values of the differenced time series of each variable. This second aggregation approach was used as it is effectively a measure of volatility, which for some variables, such as temperature, may increase under climate change (Guo et al., 2021). Therefore, this enabled testing if increased predictor volatility associated with more extreme climate change scenarios was also associated with increased forecast error. To do this, firstly the aggregated variables were regressed against climate change scenario. Following this, MAPE was regressed against each of the aggregations. This was investigated for a medium/high data availability scenario with a training dataset duration of 20 years, and a sampling period of seven days.

3. Results

3.1. Effect of climate scenario on forecast performance

In the more extreme climate change scenarios, forecast error was on average higher, particularly in the summer months (Fig. 1a). However, for all lakes except Müggelsee, the reverse effect was observed in the winter months: forecast error was on average lowest in the most severe climate change scenarios (Fig. 1c). These effects were statistically significant for both models ($p < 0.001$, Table 2).

Table 2
Summary of fixed effects from standardised LMM regression with log(MAPE) as the dependent variable. Statistical significance is indicated using asterisk notation: *** - $p < 0.001$, ** - $p < 0.01$, * - $p < 0.05$, n.s. – not significant ($p \geq 0.05$).

	Estimate (± Standard Error)		Significance	
Fixed Effect	RF	GRU	RF	GRU
β_0 (Intercept)	−0.073 ± 0.163	0.031 ± 0.161	n.s.	n.s.
β_1 (Inverse no of train years)	0.131 ± 0.000	0.252 ± 0.000	***	***
β_2 (Scenario)	0.023 ± 0.000	0.026 ± 0.000	***	***
β_3 (Season – spring)	0.047 ± 0.001	0.048 ± 0.001	***	***
β_3 (Season – summer)	0.614 ± 0.001	0.562 ± 0.001	***	***
β_3 (Season – winter)	−0.601 ± 0.001	−0.501 ± 0.001	***	***
β_4 (Sample period)	0.021 ± 0.000	0.013 ± 0.000	***	***
β_5 (Year)	0.003 ± 0.000	0.004 ± 0.000	***	***
γ_1 (Inverse no of train years: Scenario)	0.003 ± 0.000	−0.002 ± 0.000	***	***
γ_2 (Scenario: Season – spring)	−0.001 ± 0.001	−0.014 ± 0.001	n.s.	***
γ_2 (Scenario: Season – summer)	0.028 ± 0.001	0.018 ± 0.001	***	***
γ_2 (Scenario: Season – winter)	−0.072 ± 0.001	−0.076 ± 0.001	***	***
γ_3 (Sample period: No. of train years)	−0.030 ± 0.000	0.002 ± 0.000	***	***

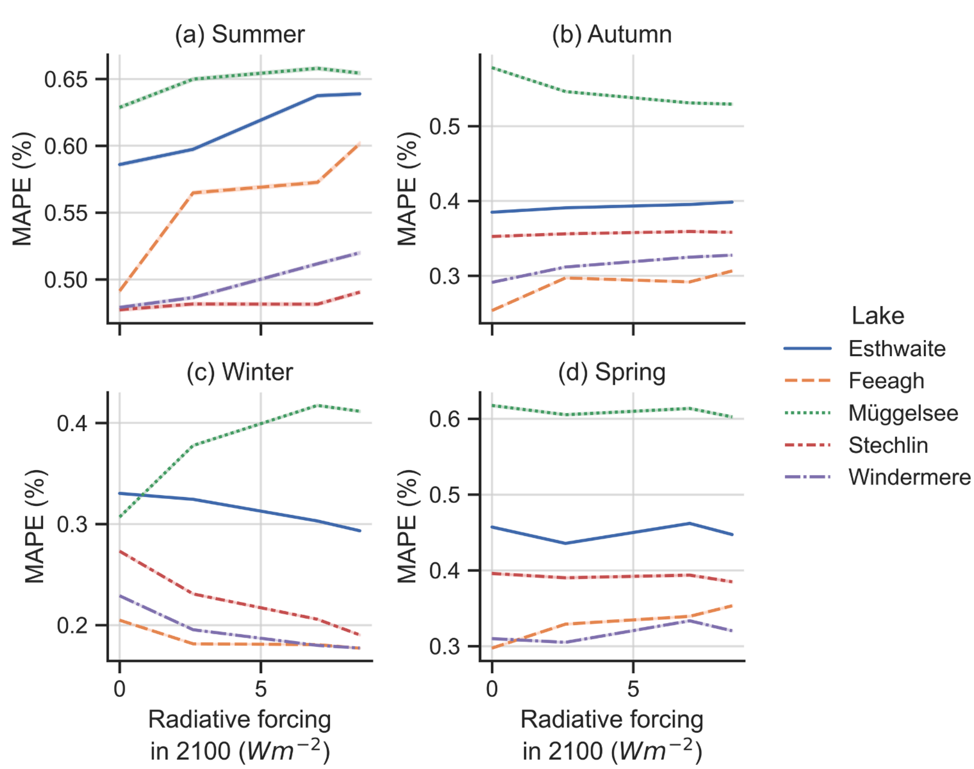

Fig. 1. MAPE against 2100 anthropogenic radiative forcing separated by season and lake (Esthwaite – blue, Feeagh – orange, Müggelsee – green, Stechlin – red, Windermere - purple).

The standardised regression coefficient for the year variable was positive and statistically significant ($p < 0.001$) for both forecast models, but much smaller in magnitude than most other predictor coefficients. This indicates that forecast error increased further into the future, but that this effect was small in comparison to the effects of training dataset duration, sampling period, and climate change scenario.

The variance associated with lake (RF = 0.12, GRU = 0.12) and the unexplained residual variance (RF = 0.68, GRU = 0.70) were of a similar order of magnitude to the largest of the fixed effect coefficients: season and training dataset duration (Table 2, Table S4). The random effects associated with climate model and lake model were much smaller suggesting approximate agreement between models.

With 30 years of weekly training data, averaged across all model runs and years, the GRU model's maximum summer forecast error was 0.16 °C higher under SSP585 compared to SSP126, while the RF model's maximum summer error increased by 0.30 °C (Table S5).

3.2. Interaction between training dataset duration and climate scenario

Across all climate change scenarios, training dataset duration scaled with error according to an approximate inverse power law (Fig. 2). This has the consequence that the majority of the performance benefit was obtained with the first ten years of training data, with diminishing returns beyond this point. The effect of training dataset duration was stronger for the GRU (= 0.252) than the RF (= 0.131) model. As such, in low data-availability scenarios, the RF consistently out-performed the GRU, but with the maximum training dataset size and sampling frequency, the performance of the two models was approximately equal (Fig. 2). The interaction between the number of training years and climate change scenario was statistically significant ($p < 0.001$) for both forecast models, but with standardised regression coefficients over 40 times smaller than that of training dataset duration this effect is effectively negligible (Table 2).

3.3. Analysis of relative effect sizes

Asides from the seasonal effects, training dataset duration had the largest effect on performance. However, this was most prevalent for low numbers of training years. Whilst the difference in performance between low and high training dataset duration scenarios was smaller at higher sampling frequencies, the overall effect of increasing sampling frequency was that this improved performance. Standardised regression coefficients for sampling frequency (RF = 0.021, GRU = 0.013) and climate change scenario (RF = 0.023, GRU = 0.026) were of similar magnitudes indicating that increasing sampling frequency has a similar effect size on performance as more extreme climate scenarios (Table 2, Fig. 3). For the RF model, reducing the sampling period by 1 day would have comparable impact on average forecast performance as a 1.0 Wm^{-2} reduction in radiative forcing by 2100.

For all lakes, various interventions such as increasing sampling frequency, or training dataset duration were sufficient to compensate for the forecast performance difference between SSP126 and SSP585, even in the summer months where this effect was strongest (Fig. 4). Notably, the German lakes required much smaller interventions than the English and Irish lakes. For example, if using the RF model to make forecasts for Stechlin, increasing the sampling frequency from fortnightly to weekly sampling was sufficient to compensate for the effect of climate change scenario, whereas with Windermere, the equivalent compensation required a sampling frequency increase from fortnightly to 3-day sampling (Fig. 4). Additionally, GRU forecasts generally required smaller interventions to compensate for the effects of climate change than the RF. For example, for Windermere, increasing the sampling from fortnightly to weekly was sufficient compensation for the GRU, whereas for the RF this required an increasing from fortnightly to 3-day sampling (Fig. 4).

3.4. Drivers of performance differences

On average across the two forecast models, the three most important variables were shortwave downwelling radiation, water surface

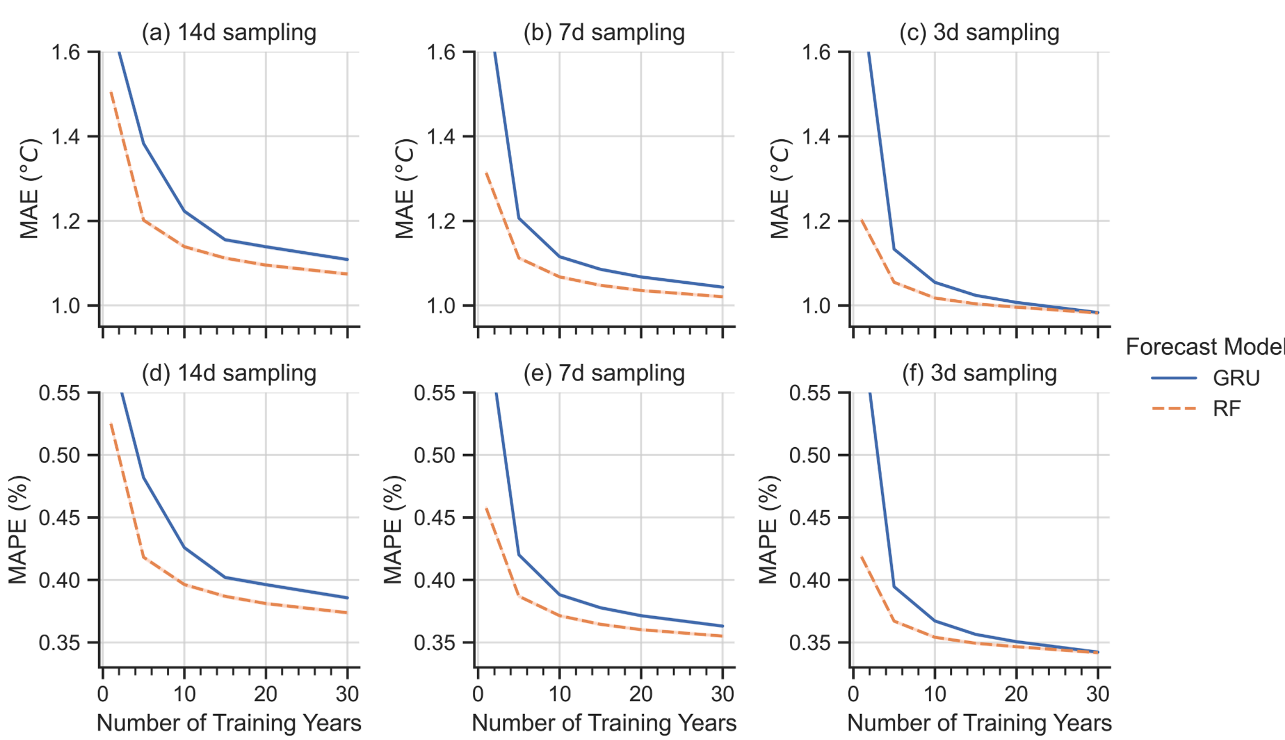

Fig. 2. Forecast model MAE (top row) and MAPE (bottom row) plotted against number of training years for different sampling periods (14 days – left, 7 days – centre, 3 days – right). Blue solid lines show errors for the GRU model, and dashed orange lines the RF.

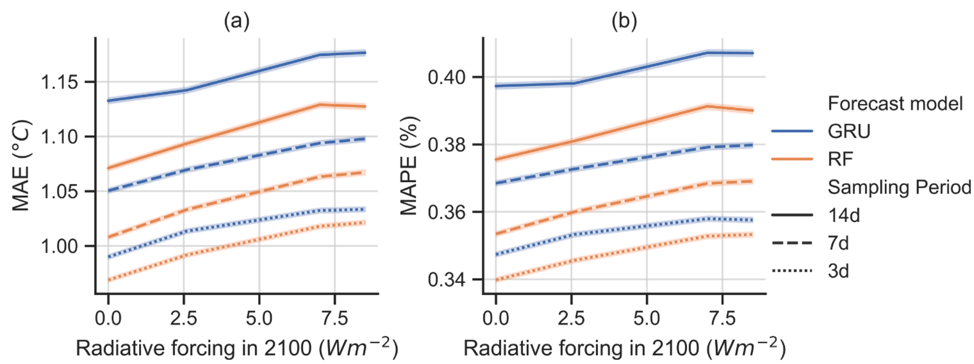

Fig. 3. Forecast model MAE (a) and MAPE (b), plotted against the 2100 anthropogenic radiative forcing associated with each of the four climate scenarios for different sampling periods for higher data availability scenarios (number of training years ≥ 10). Sampling period is indicated by line style (14 days – solid, 7 days – dashed, 3 days – dotted), with blue lines showing errors for the GRU model and orange lines the RF.

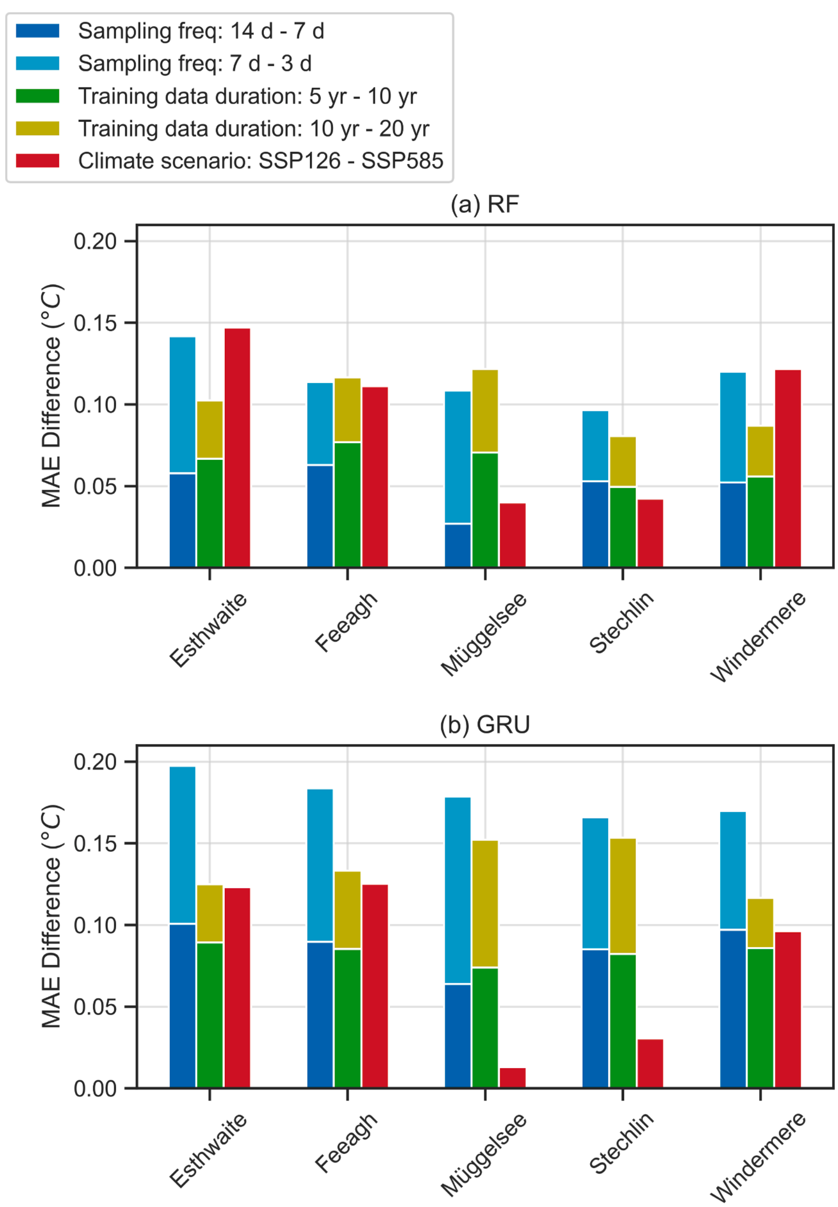

Fig. 4. Absolute difference in summer MAE for (a) RF and (b) GRU resulting from increasing sampling frequency from 14 to 7 to 3 days (blue and light blue bars) and increasing training dataset duration from 5 to 10 to 20 years (green and olive bars) under climate scenario SSP585. Red bars show the absolute difference in summer MAE between SSP126 and SSP585.

temperature, and specific humidity. Together, these accounted for 90.2 % of the total increase in MAE found in the permutation importance analysis (Figure S4), and thus were selected for further investigation in relation to climate scenario and forecast error.

In summer months, it was found that climate change scenario was significantly ($p < 0.05$) correlated with both surface temperature and specific humidity, as well as the absolute values of these following differencing (Fig. 5a, b, c, d). Conversely, shortwave downwelling radiation was not found to be associated with climate change scenario (Fig. 5e, f).

Similarly, both mean and differenced versions of surface temperature and specific humidity were associated with MAPE (Fig. 6a, b, c, d), whereas shortwave downwelling radiation was not found to be significantly correlated with forecast error (Fig. 6e, f). The variable with the strongest relationship with forecast error was absolute differenced surface temperature ($R^2 = 0.78$).

Similar, but mostly weaker patterns were observed in the other seasons. For example, absolute differenced surface temperature was significantly correlated with MAPE in all seasons (Table S7), but only significantly correlated with climate change scenario in autumn and summer (Table S6). Generally, the associations between aggregated variables and MAPE were strongest in the spring and summer months.

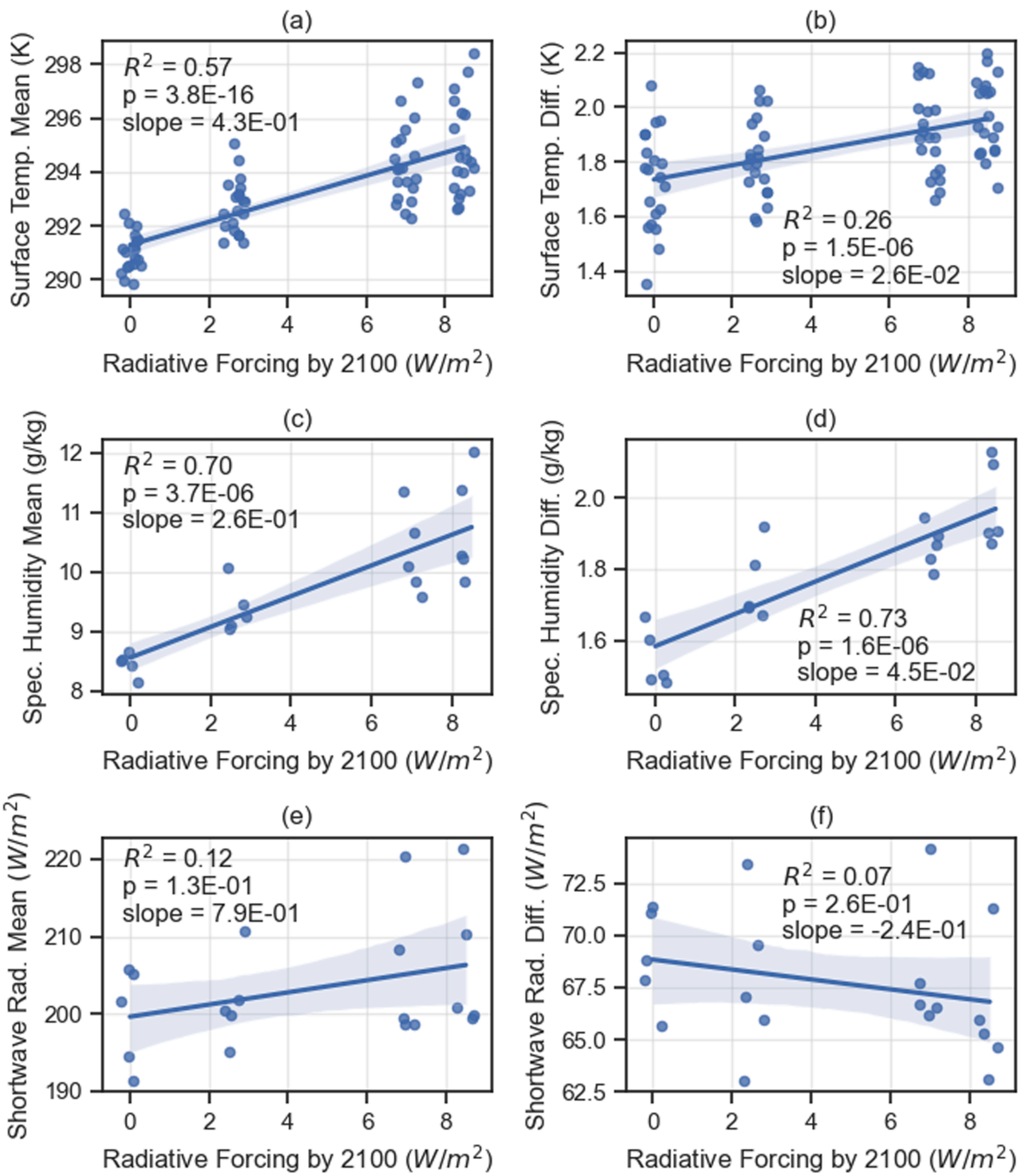

Fig. 5. Linear regression analyses from summer months between climate change scenarios and key feature means (left column) and absolute values of the differenced time series of key features (right column). Each scatter point represents the average from a single climate/lake model combination. Note that specific humidity (c, d) and shortwave radiation (e, f) do not vary with lake model, and so have fewer points than surface temperature (a, b). 'Jitter' in the x axis has been added to improve readability.

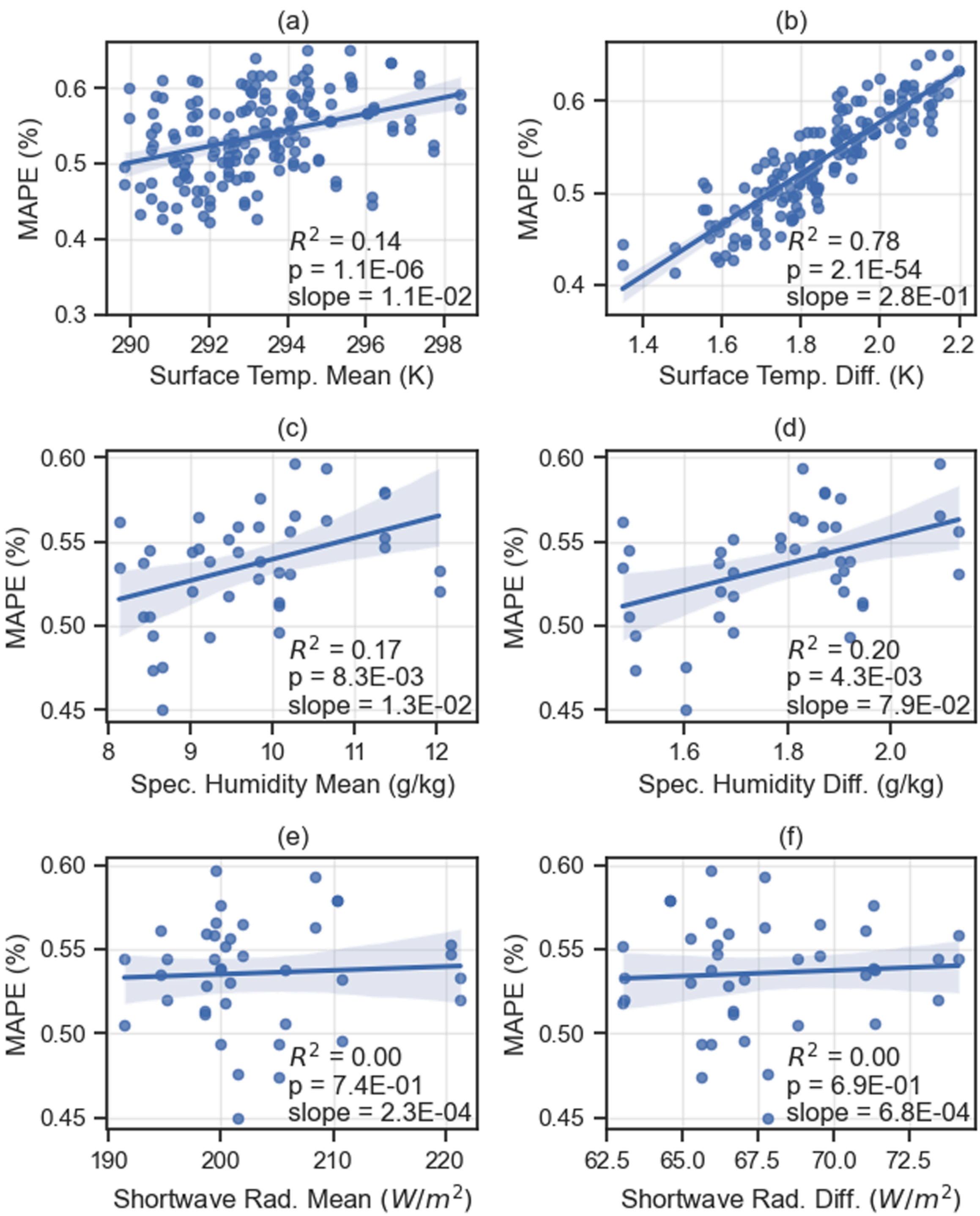

Fig. 6. Linear regression analyses from summer months between key feature means (left column), absolute values of the differenced time series of key features (right column) and MAPE. Each scatter point represents the average from a single climate/lake model/forecast model combination. Note that specific humidity (c, d) and shortwave radiation (e, f) do not vary with lake model, and so have fewer points than surface temperature (a, b).

4. Discussion

4.1. The impact of climate change on water temperature forecasting

It was found that surface water temperature forecasts did suffer from reduced performance in more extreme climate change scenarios, although this varied between seasons, lakes and forecast models. Whilst the size of this effect may appear small, it was found to be comparable to the performance benefit gained from quite substantial monitoring improvements.

In summer, MAE in the most severely affected lake (Esthwaite) was on average 0.14 °C higher in SSP585 compared with SSP126. Additionally, across all lakes, using a 30-year weekly training dataset, the maximum summer error for the best performing model (RF) was on average 0.30 °C higher in SSP585 than SSP126. Consequently, if these forecasts were used as an early warning system to trigger specific management interventions, then these error increases could be conceivably large enough to have negative management effects, particularly if the common approach of threshold-based decision triggers is used (Herman and Giuliani, 2018). Specifically, this could potentially determine whether a forecast system correctly triggers a response such as reservoir withdrawal for thermal regulation (Feldbauer et al., 2020; Rheinheimer et al., 2015; Weber et al., 2017).

With the RF forecast model, one way to compensate for the increased error between SSP126 and SSP585 would be to increase the sampling frequency approximately fourfold. Although automated in situ sampling technology now provides a means for consistent high frequency sampling of water quality indicators, this is still costly, particularly when considering the maintenance of such systems over long time periods (van der Wielen et al., 2023). This highlights a critical design consideration for climate-resilient monitoring systems - the trade-off between the cost of monitoring infrastructure and forecast performance degradation under climate change which, as demonstrated, can be quantified using the approach outlined in this study.

4.2. Variation across lakes and seasons

All lakes behaved in a similar way during the summer, that is forecast performance generally reduced in more extreme climate change scenarios (Fig. 1a). However, all lakes except Müggelsee showed the opposite pattern in the winter – forecast performance improved in more extreme climate change scenarios (Fig. 1c). This contrast can be explained by shifts in lake mixing regimes: climate change is expected to push some lakes from dimictic toward monomictic behaviour, reducing winter mixing events and increasing thermal stability (Shatwell et al., 2019). Mixed conditions tend to lower surface temperature variability due to a deeper, more thermally inert surface layer, making winter conditions more predictable. Müggelsee, by contrast, is relatively much more shallow and polymictic (Table S1), with a more stable mixing regime that does not shift under warming (Shatwell and Köhler, 2019). It likely represents the default pattern for lakes that remain polymictic, lacking the "reset" effect that makes deeper lakes more predictable in winter. As a result, Müggelsee shows a winter forecast pattern more similar to its summer behaviour. This underlines the need for site-specific monitoring design that is informed by climate projections, as no single rule applies uniformly across lakes. This also illustrates the importance of seasonality, suggesting that forecasts for lakes located in different climate zones, with alternative seasonal patterns, may respond differently to climate change. For example, climate change is likely increasing over-lake atmospheric instability, which varies by lake size and latitude, and may lead to a decoupling of air and surface water temperatures, potentially reducing forecast skill more strongly in some lakes than others (Woolway et al., 2017). Additionally, shifts in lake thermal structure due to climate change have also been shown to be altitude dependent, which indicates that lake forecast performance degradation due to climate change may also vary across elevation gradients (Råman Vinnå et al., 2021). Accordingly, it is evident that the results of this study are not globally generalisable but are still relevant for the large number of northern temperate lakes with similar latitudes, morphologies, and elevations to the five studied lakes.

4.3. Relevance to other water quality metrics

Under the assumption that increased difficulty in forecasting water temperature would also be associated with increased difficulty in forecasting other variables which are driven or affected by it, the results of this study may generalise to other variables. This assumption is likely most valid for variables which are largely physics-driven, such as stratification onset, or hypolimnetic anoxia (Boehrer and Schultze, 2008; Snortheim et al., 2017). For variables strongly influenced by both the physics and ecology of the study site, the picture is more complex. For example, in the case of cyanobacterial bloom forecasting, increased uncertainty in surface temperature suggests higher uncertainty in growth rates (Davis et al., 2009). However for waterbodies where cyanobacterial blooms transition from being rare events to more commonplace phenomena due to climate change, they may inherently become more predictable (Feng et al., 2023; Paerl et al., 2020). Given that it is possible to simulate processes such as algal growth under various climate scenarios (Zwart et al., 2019), it is likely that soon a more diverse and complex set of water quality variables will become available through projects like ISIMIP (Golub et al., 2022). This therefore opens the possibility for informing the design of climate-resilient monitoring and forecasting systems for more complex water quality phenomena such as algal blooms.

Additionally, whilst it was found that there was a significant interaction ($p < 0.001$) between training dataset duration and climate scenario, the relative size of this effect was small in comparison to other factors, such as climate scenario (Table 2). This means that older data were approximately equally useful for training models, regardless of climate change intensity. A likely explanation for this is that surface water temperature is well-understood to be primarily governed by air temperature and surface heat exchange, which, from a process perspective, can be considered largely stationary. In other words, the underlying heat transfer equations which govern it generally do not change in structure, except during periods of ice cover (Hipsey et al., 2019). In contrast, other variables such as algal biomass are influenced by a far more complex set of interacting drivers, including nutrients, light availability, and grazing, which can exhibit strong non-stationarity, particularly under ecological regime shifts (Zingone and Oksfeldt Enevoldsen, 2000). These complexities may render older data less representative of current or future conditions, underscoring that future work applying the approach presented here to additional water quality indicators would be valuable for designing climate-resilient monitoring and forecasting systems.

4.4. Forecast model choice implications

Whilst the structure of recurrent neural network models such as GRU is inherently suited to time series forecasting tasks, they often do not outperform other forecasting approaches, particularly in data-limited scenarios (Atton Beckmann et al., 2025; Hewamalage et al., 2021; Yousefi and Toffolon, 2022). Accordingly, in this study, in all but the highest data availability scenario, the RF outperformed the GRU (Fig. 2). This is likely due to the higher complexity of the GRU relative to the RF, which therefore necessitates a larger volume of training data to achieve competitive performance (Atton Beckmann et al., 2025). However, with sufficient training data (30 years of 3-day samples), the GRU's performance did marginally exceed the RF. Furthermore, given that the GRU's training dataset size / performance gradient (β_1) was almost double that of the RF (Table 2), it can be inferred that in even higher data-availability scenarios, the GRU would be expected to outperform the RF by an increasing margin.

Whilst both models' performance was affected by climate change scenario, this effect varied according to season, and in some cases was notably different between the two forecast models (Table 2). In autumn, both models suffered similar performance degradations in more extreme climate scenarios. Likewise, in winter, the forecast performance improvements due to intensified climate warming were comparable between forecast models. However, in summer, climate scenario decreased the RF's performance ($\gamma_2 = 0.028$) more than the GRU's performance ($\gamma_2 = 0.018$). Similarly, in spring, the effect of climate scenario on the RF's performance was approximately equivalent to that in autumn ($\gamma_2 = n.s.$), whereas the GRU's performance was less negatively affected ($\gamma_2 = -0.014$). Therefore, on average, whilst the RF performed better with smaller training datasets, it was also more negatively affected by climate scenario than the GRU. This suggests that where large volumes of high frequency measurement data are available, using recurrent neural network models such as GRU may be beneficial for minimising the effects of climate change on forecast performance.

4.5. Evaluating forecast skill with synthetic data

All the data used in this study were synthetic, and therefore there is some uncertainty concerning how the findings would translate to actual surface lake temperature forecasting with field-measured data. To some extent this was quantified by using a suite of climate and lake models, and it was found that the variances associated with the random effects of lake model ($\sigma^2 = 0.0087$) and climate model ($\sigma^2 = 0.0018$) were both small compared with the variance associated with lake ($\sigma^2 = 0.12$). This suggests good agreement between the models used in this study. Furthermore, the comparison of simulated data with in situ observations demonstrated that the magnitude and variation in daily and fortnightly changes were closely matched (Table S2). It was found that GLM tended to have slightly higher short-term variations than the corresponding in situ observations, whereas Flake tended to vary less. This demonstrates that it is certainly possible to systemically under- or over-estimate forecast skill when testing this with simulated data. However, by using a suite of models, any model-specific biases will likely have been reduced. Ultimately, despite the uncertainties inherent with modelled data, using these allows exploration of forecast performance across multiple climate scenarios and decades, which would be impossible with field measurements alone.

4.6. Designing climate-resilient water quality forecasting systems

Unlike weather forecasting, which benefits from decades of methodological refinement (Lynch, 2008), water quality forecasting is still developing, particularly in the context of ecological complexity and climate change (Dietze et al., 2018). Interest in forecasting complex ecological phenomena, such as algal blooms, continues to grow (Zahir et al., 2024), yet these events often have site-specific drivers that differ markedly between lakes (Ding and Qin, 2024). Climate change further complicates this challenge by simultaneously increasing the need for reliable forecasts (Bartlett and Dedekorkut-Howes, 2022) and, as this study demonstrates, potentially making the forecasting task inherently more difficult. This illustrates the importance of designing water quality monitoring infrastructure with both climate resilience and site-specific variability in mind. While ongoing evaluation and adaptation of monitoring systems is critical, we argue that a purely iterative design approach is not efficient - particularly when it may take five or more years to determine whether a given sampling approach adequately supports the forecasting objective (Atton Beckmann et al., 2025).

Instead, we propose that the design of monitoring and forecasting systems should also be informed through simulation, by testing forecasting models on data generated by physical and biological models driven by climate projections. This approach enables systematic testing of key design decisions, such as which variables to monitor, how frequently to sample, which forecast models to apply, and how much training data is needed before reliable forecasts can be generated. As demonstrated in this study, climate projections can also be used to evaluate how system performance may change under different future scenarios, supporting the development of forecasting systems capable of maintaining sufficient accuracy even under extreme climate change. Whilst this simulation-based approach is constrained by the underlying input data - here, physical, one-dimensional lake models and ISIMIP climate projections - these models are continually improving (Frieler et al., 2024; Golub et al., 2022). Moreover, two-way coupling of climate projections with three-dimensional lake models has already been successfully demonstrated for some regions, which is particularly important for resolving hydro-dynamics in larger and more complex lake systems (Xue et al., 2022). Therefore, as coupling capabilities expand and lake and climate models continue to advance, they will offer increasing potential to support the development of effective monitoring and forecasting strategies for managing freshwater resources in a changing climate.

5. Conclusions

This study highlights how climate change can affect the performance of short-term lake water temperature forecasts, and how simulation-based approaches can be used to design climate-resilient monitoring and forecasting systems. The key conclusions are:

- Forecast skill is sensitive to climate change, especially in summer, due to increased surface temperature variability. However, this performance degradation can be offset through targeted improvements in monitoring.
- Monitoring system design should balance cost and performance, as increasing sampling frequency or extending historical records can compensate for climate-driven forecast degradation. These trade-offs can be quantified using synthetic data.
- Responses vary by lake and season, with deeper lakes often becoming more predictable in winter under warming scenarios. This emphasizes the need for site-specific, simulation-informed monitoring strategies.
- Findings likely extend beyond temperature forecasts to other water quality indicators influenced by surface physics, such as stratification or hypoxia. For more complex ecological phenomena, responses may differ due to nonlinear regime shifts.
- Simulation-based system design offers a proactive approach to developing resilient water quality monitoring infrastructure, complementing traditional iterative methods that require years of real-world testing.
- As climate and lake models improve, this approach will become increasingly powerful in guiding the development of sustainable, future-proof forecasting systems for freshwater management.

These findings support a shift towards forward-looking, data-informed monitoring strategies that account for the challenges posed by a changing climate.

Funding

This work was carried out as part of D. Atton Beckmann's PhD, which was funded by the Scottish Government's Hydro Nation Scholars programme. Mortimer Werther was supported by the Swiss Science Foundation grant Lake3P, no 204783.

CRediT authorship contribution statement

D. Atton Beckmann: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **M. Werther:** Writing – review & editing. **T. Shatwell:** Writing – review & editing. **E. Spyrakos:** Writing – review & editing, Supervision. **P. Hunter:** Writing – review & editing, Supervision. **I.D. Jones:** Writing – review & editing, Supervision, Conceptualization.

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

We acknowledge Johannes Feldbauer (TU Dresden) and Jorrit Mesman (Uppsala University), who were responsible for the ISIMIP3b local lake simulations using the FLake, GLM, GOTM, and Simstrat models. We also acknowledge the ISIMIP lake sector coordinators, Ana Ayala-Zamora (Uppsala University) and Robert Ladwig (Aarhus University).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.wroa.2025.100457.

Data availability

All data used for this study are presently publicly available from the ISIMIP data repositories (https://data.isimip.org/) (Frieler et al., 2024; Golub et al., 2022).

Outputs from forecast models, statistical test results etc. can be made available on request by contacting the corresponding author.

References

Atton Beckmann, D., Werther, M., Mackay, E.B., Spyrakos, E., Hunter, P., Jones, I.D., 2025. Are more data always better? – Machine learning forecasting of algae based on long-term observations. J. Environ. Manage. 373, 123478. https://doi.org/10.1016/j.jenvman.2024.123478.

Bartlett, J.A., Dedekorkut-Howes, A., 2022. Adaptation strategies for climate change impacts on water quality: a systematic review of the literature. J. Water Clim. Change 14, 651–675. https://doi.org/10.2166/wcc.2022.279.

Bhogal, D., Zhao, Y., Kapoor, R., Toscas, P., 2023. A comparative study of water temperature forecasting methods. In: 2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Presented at the 2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 325–329. https://doi.org/10.1109/WI-IAT59888.2023.00052.

Boehrer, B., Schultze, M., 2008. Stratification of lakes. Rev. Geophys. 46. https://doi.org/10.1029/2006RG000210.

Boucher, O., Servonnat, J., Albright, A.L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., Lionel, Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A.K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., Vuichard, N., 2020. Presentation and evaluation of the IPSL-CM6A-LR climate model. J. Adv. Model. Earth Syst. 12, e2019MS002010. https://doi.org/10.1029/2019MS002010.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Cael, B.B., Seekell, D.A., 2016. The size-distribution of Earth's lakes. Sci. Rep. 6, 29633. https://doi.org/10.1038/srep29633.

Capon, S.J., Stewart-Koster, B., Bunn, S.E., 2021. Future of freshwater ecosystems in a 1.5 °C warmer world. Front. Environ. Sci. 9. https://doi.org/10.3389/fenvs.2021.784642.

Carey, C.C., Calder, R.S.D., Figueiredo, R.J., Gramacy, R.B., Lofton, M.E., Schreiber, M. E., Thomas, R.Q., 2025. A framework for developing a real-time lake phytoplankton forecasting system to support water quality management in the face of global change. Ambio 54, 475–487. https://doi.org/10.1007/s13280-024-02076-7.

Chung, J., Gulcehre, C., Cho, K., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. https://doi.org/10.48550/arXiv.1412.3555.

Dai, A., 2013. Increasing drought under global warming in observations and models. Nat. Clim. Change 3, 52–58. https://doi.org/10.1038/nclimate1633.

Davis, T.W., Berry, D.L., Boyer, G.L., Gobler, C.J., 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of *microcystis* during cyanobacteria blooms. Harmful algae. This issue contains the special section on "Strains" 8, 715–725. https://doi.org/10.1016/j.hal.2009.02.004.

Di Nunno, F., Zhu, S., Ptak, M., Sojka, M., Granata, F., 2023. A stacked machine learning model for multi-step ahead prediction of lake surface water temperature. Sci. Total Environ. 890, 164323. https://doi.org/10.1016/j.scitotenv.2023.164323.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G., Loescher, H.W., Lunch, C.K., Pijanowski, B.C., Randerson, J.T., Read, E.K., Tredennick, A.T., Vargas, R., Weathers, K.C., White, E.P., 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proc. Natl. Acad. Sci. 115, 1424–1432. https://doi.org/10.1073/pnas.1710231115.

Ding, K., Qin, B., 2024. Global drivers of algal bloom occurrence: the role of climate, hydrology, and morphometry. Water 16, 3624. https://doi.org/10.3390/w16243624.

Dokulil, M.T., de Eyto, E., Maberly, S.C., May, L., Weyhenmeyer, G.A., Woolway, R.I., 2021. Increasing maximum lake surface temperature under climate change. Clim. Change 165, 56. https://doi.org/10.1007/s10584-021-03085-1.

Duan, Z., Gao, W., Cheng, G., Zhang, Y., Chang, X., 2024. Warming surface and Lake heatwaves as key drivers to harmful algal blooms: a case study of Lake Dianchi, China. J. Hydrol. 632, 130971. https://doi.org/10.1016/j.jhydrol.2024.130971.

Dunne, J.P., Horowitz, L.W., Adcroft, A.J., Ginoux, P., Held, I.M., John, J.G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C.A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K.A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P.G., Griffies, S.M., Guo, H., Hallberg, R.W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P.C.D., Nikonov, S., Paynter, D.J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B.G., Robinson, T., Schwarzkopf, D.M., Sentman, L.T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A.T., Wyman, B., Zeng, Y., Zhao, M., 2020. The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): overall coupled model description and simulation characteristics. J. Adv. Model. Earth Syst. 12. https://doi.org/10.1029/2019MS002015 e2019MS002015.

Ernst, M.D., 2004. Permutation methods: a basis for exact inference. Stat. Sci. 19, 676–685.

Fang, X., Stefan, H.G., 2012. Impacts of climatic changes on water quality and fish habitat in aquatic systems. In: Chen, W.-Y., Seiner, J., Suzuki, T., Lackner, M. (Eds.), Handbook of Climate Change Mitigation. Springer US, New York, NY, pp. 531–569. https://doi.org/10.1007/978-1-4419-7991-9_16.

Feldbauer, J., Kneis, D., Hegewald, T., Berendonk, T.U., Petzoldt, T., 2020. Managing climate change in drinking water reservoirs: potentials and limitations of dynamic withdrawal strategies. Environ. Sci. Eur. 32, 48. https://doi.org/10.1186/s12302-020-00324-7.

Feng, C., Li, L., Xu, C., 2023. Advancements in predicting and modeling rare event outcomes for enhanced decision-making. BMC Med. Res. Methodol. 23, 243. https://doi.org/10.1186/s12874-023-02060-x.

Feuchtmayr, H., Clarke, M.A., De Ville, M.M., Dodd, B.A., Fletcher, J., James, J.B., Mackay, E.B., Pereira, M.G., Rhodes, G., Thackeray, S.J., Maberly, S.C., 2021a. Surface temperature, surface oxygen, water clarity, water chemistry and phytoplankton chlorophyll a data from Esthwaite Water, 2014 to 2018. https://doi.org/10.5285/89ff964d-4aed-4b66-8669-71d42c5add52.

Feuchtmayr, H., Jones, I.D., Clarke, M., Maberly, S.C., 2021b Hourly data from automatic water monitoring buoy from Windermere South Basin, 2016 to 2018. https://doi.org/10.5285/e1b6e387-1bd6-444a-8de9-035df0b8baec.

Frieler, K., Volkholz, J., Lange, S., Schewe, J., Mengel, M., del Rocío Rivas López, M., Otto, C., Reyer, C.P.O., Karger, D.N., Malle, J.T., Treu, S., Menz, C., Blanchard, J.L., Harrison, C.S., Petrik, C.M., Eddy, T.D., Ortega-Cisneros, K., Novaglio, C., Rousseau, Y., Watson, R.A., Stock, C., Liu, X., Heneghan, R., Tittensor, D., Maury, O., Büchner, M., Vogt, T., Wang, T., Sun, F., Sauer, I.J., Koch, J., Vanderkelen, I., Jägermeyr, J., Müller, C., Rabin, S., Klar, J., Vega del Valle, I.D., Lasslop, G., Chadburn, S., Burke, E., Gallego-Sala, A., Smith, N., Chang, J., Hantson, S., Burton, C., Gädeke, A., Li, F., Gosling, S.N., Müller Schmied, H., Hattermann, F., Wang, J., Yao, F., Hickler, T., Marcé, R., Pierson, D., Thiery, W., Mercado-Bettín, D., Ladwig, R., Ayala-Zamora, A.I., Forrest, M., Bechtold, M., 2024. Scenario setup and forcing data for impact model evaluation and impact attribution within the third round of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP3a). Geosci. Model Dev. 17, 1–51. https://doi.org/10.5194/gmd-17-1-2024.

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. 12, 3955–3974. https://doi.org/10.5194/gmd-12-3955-2019.

Giamalaki, K., Beaulieu, C., Prochaska, J.X., 2022. Assessing predictability of marine heatwaves with random forests. Geophys. Res. Lett. 49. https://doi.org/10.1029/2022GL099069 e2022GL099069.

Glibert, P.M., Burkholder, J.M., 2018. Causes of harmful algal blooms. Harmful Algal Blooms. John Wiley & Sons, Ltd, pp. 1–38. https://doi.org/10.1002/9781118994672.ch1.

Golub, M., Thiery, W., Marcé, R., Pierson, D., Vanderkelen, I., Mercado-Bettin, D., Woolway, R.I., Grant, L., Jennings, E., Kraemer, B.M., Schewe, J., Zhao, F., Frieler, K., Mengel, M., Bogomolov, V.Y., Bouffard, D., Côté, M., Couture, R.-M., Debolskiy, A.V., Droppers, B., Gal, G., Guo, M., Janssen, A.B.G., Kirillin, G., Ladwig, R., Magee, M., Moore, T., Perroud, M., Piccolroaz, S., Raaman Vinnaa, L., Schmid, M., Shatwell, T., Stepanenko, V.M., Tan, Z., Woodward, B., Yao, H., Adrian, R., Allan, M., Anneville, O., Arvola, L., Atkins, K., Boegman, L., Carey, C., Christianson, K., de Eyto, E., DeGasperi, C., Grechushnikova, M., Hejzlar, J., Joehnk, K., Jones, I.D., Laas, A., Mackay, E.B., Mammarella, I., Markensten, H., McBride, C., Özkundakci, D., Potes, M., Rinke, K., Robertson, D., Rusak, J.A., Salgado, R., van der Linden, L., Verburg, P., Wain, D., Ward, N.K., Wollrab, S., Zdorovennova, G., 2022. A framework for ensemble modelling of climate change impacts on lakes worldwide: the ISIMIP Lake Sector. Geosci. Model Dev. 15, 4597–4623. https://doi.org/10.5194/gmd-15-4597-2022.

Guo, F., Do, V., Cooper, R., Huang, Y., Zhang, P., Ran, J., Zhang, Q., Tian, L., Fu, Z., 2021. Trends of temperature variability: which variability and what health implications? Sci. Total Environ. 768, 144487. https://doi.org/10.1016/j.scitotenv.2020.144487.

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J.H., von Storch, J.-S., Brüggemann, N., Haak, H., Stössel, A., 2019. Max Planck Institute Earth System Model (MPI-ESM1.2) for the high-resolution model Intercomparison Project (HighResMIP). Geosci. Model Dev. 12, 3241–3281. https://doi.org/10.5194/gmd-12-3241-2019.

Heddam, S., Ptak, M., Zhu, S., 2020. Modelling of daily lake surface water temperature from air temperature: extremely randomized trees (ERT) versus Air2Water, MARS, M5Tree, RF and MLPNN. J. Hydrol. 588, 125130. https://doi.org/10.1016/j.jhydrol.2020.125130.

Herman, J.D., Giuliani, M., 2018. Policy tree optimization for threshold-based water resources management over multiple timescales. Environ. Model. Softw. 99, 39–51. https://doi.org/10.1016/j.envsoft.2017.09.016.
Hewamalage, H., Bergmeir, C., Bandara, K., 2021. Recurrent Neural Networks for Time Series Forecasting: current status and future directions. Int. J. Forecast. 37, 388–427. https://doi.org/10.1016/j.ijforecast.2020.06.008.
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. 12, 473–523. https://doi.org/10.5194/gmd-12-473-2019.
Hutchings, A.M., de Vries, C.S., Hayes, N.R., Orr, H.G., 2024. Temperature and dissolved oxygen trends in english estuaries over the past 30 years. Estuar. Coast. Shelf. Sci. 306, 108892. https://doi.org/10.1016/j.ecss.2024.108892.
Inácio, M., Barceló, D., Zhao, W., Pereira, P., 2022. Mapping lake ecosystem services: a systematic review. Sci. Total Environ. 847, 157561. https://doi.org/10.1016/j.scitotenv.2022.157561.
Jia, X., Willard, J., Karpatne, A., Read, J.S., Zwart, J.A., Steinbach, M., Kumar, V., 2021. Physics-guided machine learning for scientific discovery: an application in simulating lake temperature profiles. ACM/IMS Trans. Data Sci. 2 (20), 1–20. https://doi.org/10.1145/3447814, 26.
Koutsandreas, D., Spiliotis, E., Petropoulos, F., Assimakopoulos, V., 2022. On the selection of forecasting accuracy measures. J. Oper. Res. Soc. 73, 937–954. https://doi.org/10.1080/01605682.2021.1892464.
Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2017. lmerTest Package: tests in linear mixed effects models. J. Stat. Softw. 82, 1–26. https://doi.org/10.18637/jss.v082.i13.
Last, M., 2007. Predicting and optimizing classifier utility with the power law. In: Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007). Presented at the Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), pp. 219–224. https://doi.org/10.1109/ICDMW.2007.31.
Liu, W.-C., Chen, W.-B., 2012. Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models. Comput. Geosci. 45, 13–25. https://doi.org/10.1016/j.cageo.2012.03.010.
Lynch, P., 2008. The origins of computer weather prediction and climate modeling. J. Comput. Phys. Predict. Weather Clim. Extreme Events 227, 3431–3444. https://doi.org/10.1016/j.jcp.2007.02.034.
Mahmood, R., Lucas, J., Acuna, D., Li, D., Philion, J., Alvarez, J.M., Yu, Z., Fidler, S., Law, M.T., 2022. How much more data do I need? Estim. Requir. Downstr. Tasks. https://doi.org/10.48550/arXiv.2207.01725.
Marcé, R., George, G., Buscarinu, P., Deidda, M., Dunalska, J., de Eyto, E., Flaim, G., Grossart, H.-P., Istvanovics, V., Lenhardt, M., Moreno-Ostos, E., Obrador, B., Ostrovsky, I., Pierson, D.C., Potužák, J., Poikane, S., Rinke, K., Rodríguez-Mozaz, S., Staehr, P.A., Šumberová, K., Waajen, G., Weyhenmeyer, G.A., Weathers, K.C., Zion, M., Ibelings, B.W., Jennings, E., 2016. Automatic high frequency monitoring for improved lake and reservoir management. Environ. Sci. Technol. 50, 10780–10794. https://doi.org/10.1021/acs.est.6b01604.
Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian, Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jia, Y., Jozefowicz, Rafal, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, Mané, Dandelion, Monga, Rajat, Moore, Sherry, Murray, Derek, Olah, Chris, Schuster, Mike, Shlens, Jonathon, Steiner, Benoit, Sutskever, Ilya, Talwar, Kunal, Tucker, Paul, Vanhoucke, Vincent, Vasudevan, Vijay, Viégas, Fernanda, Vinyals, Oriol, Warden, Pete, Wattenberg, Martin, Wicke, Martin, Yu, Yuan, Zheng, Xiaoqiang, 2015. In: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.
Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., Stouffer, R.J., 2008. Stationarity is dead: whither water management? Science 319, 573–574. https://doi.org/10.1126/science.1151915.
Mironov, D., Heise, E., Kourzeneva, E., Ritter, B., Schneider, N., Terzhevik, A., 2010. Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO. Boreal Environ. Res. 15.
O'Neill, B.C., Tebaldi, C., van Vuuren, D.P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G.A., Moss, R., Riahi, K., Sanderson, B.M., 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geosci. Model Dev. 9, 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016.
Paerl, H., Havens, K., Hall, N., Otten, T., Zhu, M., Xu, H., Zhu, G., Qin, B., 2020. Mitigating a global expansion of toxic cyanobacterial blooms: confounding effects and challenges posed by climate change. Mar. Freshw. Res. 71, 579–592. https://doi.org/10.1071/MF18392.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Piccolroaz, S., Zhu, S., Ladwig, R., Carrea, L., Oliver, S., Piotrowski, A.P., Ptak, M., Shinohara, R., Sojka, M., Woolway, R.I., Zhu, D.Z., 2024. Lake water temperature modeling in an era of climate change: data sources, models, and future prospects. Rev. Geophys. 62. https://doi.org/10.1029/2023RG000816 e2023RG000816.
Quan, Q., Hao, Z., Xifeng, H., Jingchun, L., 2022. Research on water temperature prediction based on improved support vector regression. Neural Comput. Appl. 34, 8501–8510. https://doi.org/10.1007/s00521-020-04836-4.
Råman Vinnå, L., Medhaug, I., Schmid, M., Bouffard, D., 2021. The vulnerability of lakes to climate change along an altitudinal gradient. Commun. Earth Env. 2, 35. https://doi.org/10.1038/s43247-021-00106-w.
Read, J.S., Jia, X., Willard, J., Appling, A.P., Zwart, J.A., Oliver, S.K., Karpatne, A., Hansen, G.J.A., Hanson, P.C., Watkins, W., Steinbach, M., Kumar, V., 2019. Process-guided deep learning predictions of lake water temperature. Water Resour. Res. 55, 9173–9190. https://doi.org/10.1029/2019WR024922.
Rheinheimer, D.E., Null, S.E., Lund, J.R., 2015. Optimizing selective withdrawal from reservoirs to manage downstream temperatures with climate warming. J. Water Resour. Plan. Manag. 141, 04014063. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000447.
Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee, S., Dixon Hamil, K.-A., Koenig, L.E., Locke, D.H., DeMarche, M.L., Tingley, M.W., Wheeler, K., Youngflesh, C., Zipkin, E.F., 2021. Working across space and time: nonstationarity in ecological research and application. Front. Ecol. Environ. 19, 66–72. https://doi.org/10.1002/fee.2298.
Saber, A., James, D.E., Hayes, D.F., 2020. Long-term forecast of water temperature and dissolved oxygen profiles in deep lakes using artificial neural networks conjugated with wavelet transform. Limnol. Ocean. 65, 1297–1317. https://doi.org/10.1002/lno.11390.
Schaeffer, B.A., Reynolds, N., Ferriby, H., Salls, W., Smith, D., Johnston, J.M., Myer, M., 2024. Forecasting freshwater cyanobacterial harmful algal blooms for Sentinel-3 satellite resolved U.S. lakes and reservoirs. J. Environ. Manage. 349, 119518. https://doi.org/10.1016/j.jenvman.2023.119518.
Sellar, A.A., Jones, C.G., Mulcahy, J.P., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F.M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., de Mora, L., Kuhlbrodt, T., Rumbold, S.T., Kelley, D.I., Ellis, R., Johnson, C.E., Walton, J., Abraham, N.L., Andrews, M.B., Andrews, T., Archibald, A.T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G.A., Gedney, N., Griffiths, P.T., Harper, A.B., Hendry, M.A., Hewitt, A.J., Johnson, B., Jones, A., Jones, C.D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R.J., Predoi, V., Robertson, E., Siahaan, A., Smith, R.S., Swaminathan, R., Woodhouse, M.T., Zeng, G., Zerroukat, M., 2019. UKESM1: description and evaluation of the U.K. Earth System model. J. Adv. Model. Earth Syst. 11, 4513–4558. https://doi.org/10.1029/2019MS001739.
Shatwell, T., Köhler, J., 2019. Decreased nitrogen loading controls summer cyanobacterial blooms without promoting nitrogen-fixing taxa: long-term response of a shallow lake. Limnol. Ocean. 64, S166–S178. https://doi.org/10.1002/lno.11002.
Shatwell, T., Thiery, W., Kirillin, G., 2019. Future projections of temperature and mixing regime of European temperate lakes. Hydrol. Earth Syst. Sci. 23, 1533–1551. https://doi.org/10.5194/hess-23-1533-2019.
Snortheim, C.A., Hanson, P.C., McMahon, K.D., Read, J.S., Carey, C.C., Dugan, H.A., 2017. Meteorological drivers of hypolimnetic anoxia in a eutrophic, north temperate lake. Ecol. Modell. 343, 39–53. https://doi.org/10.1016/j.ecolmodel.2016.10.014.
Stroom, J.M., Kardinaal, W.E.A., 2016. How to combat cyanobacterial blooms: strategy toward preventive lake restoration and reactive control measures. Aquat. Ecol. 50, 541–576. https://doi.org/10.1007/s10452-016-9593-0.
Tabari, H., 2020. Climate change impact on flood and extreme precipitation increases with water availability. Sci. Rep. 10, 13768. https://doi.org/10.1038/s41598-020-70816-2.
Till, A., Rypel, A.L., Bray, A., Fey, S.B., 2019. Fish die-offs are concurrent with thermal extremes in north temperate lakes. Nat. Clim. Chang. 9, 637–641. https://doi.org/10.1038/s41558-019-0520-y.
Umlauf, L., Burchard, H., Bolding, K., 2005. GOTM Sourcecode and Test Case Documentation. Devel version-pre, p. 4.
Valavi, E., Hestness, J., Ardalani, N., Iansiti, M., 2022. Time and the value of data. https://doi.org/10.48550/arXiv.2203.09118.
van der Wielen, P.W.J.J., Dignum, M., Donocik, A., Prest, E.I., 2023. Influence of temperature on growth of four different opportunistic pathogens in drinking water biofilms. Microorganisms 11, 1574. https://doi.org/10.3390/microorganisms11061574.
Waqas, M., Humphries, U.W., 2024. A critical review of RNN and LSTM variants in hydrological time series predictions. MethodsX 13, 102946. https://doi.org/10.1016/j.mex.2024.102946.
Weber, M., Rinke, K., Hipsey, M.R., Boehrer, B., 2017. Optimizing withdrawal from drinking water reservoirs to reduce downstream temperature pollution and reservoir hypoxia. J. Environ. Manage. 197, 96–105. https://doi.org/10.1016/j.jenvman.2017.03.020.
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82. https://doi.org/10.3354/cr030079.
Woolway, R.I., Sharma, S., Smol, J.P., 2022. Lakes in hot water: the impacts of a changing climate on aquatic ecosystems. Bioscience 72, 1050–1061. https://doi.org/10.1093/biosci/biac052.
Woolway, R.I., Sharma, S., Weyhenmeyer, G.A., Debolskiy, A., Golub, M., Mercado-Bettín, D., Perroud, M., Stepanenko, V., Tan, Z., Grant, L., Ladwig, R., Mesman, J., Moore, T.N., Shatwell, T., Vanderkelen, I., Austin, J.A., DeGasperi, C.L., Dokulil, M., La Fuente, S., Mackay, E.B., Schladow, S.G., Watanabe, S., Marcé, R., Pierson, D.C., Thiery, W., Jennings, E., 2021. Phenological shifts in lake stratification under climate change. Nat. Commun. 12, 2318. https://doi.org/10.1038/s41467-021-22657-4.
Woolway, R.I., Verburg, P., Merchant, C.J., Lenters, J.D., Hamilton, D.P., Brookes, J., Kelly, S., Hook, S., Laas, A., Pierson, D., Rimmer, A., Rusak, J.A., Jones, I.D., 2017. Latitude and lake size are important predictors of over-lake atmospheric stability. Geophys. Res. Lett. 44, 8875–8883. https://doi.org/10.1002/2017GL073941.
Xu, L., Zhang, C., Chen, N., Moradkhani, H., Chu, P.-S., Zhang, X., 2020. Potential precipitation predictability decreases under future warming. Geophys. Res. Lett. 47. https://doi.org/10.1029/2020GL090798 e2020GL090798.

Xue, P., Ye, X., Pal, J.S., Chu, P.Y., Kayastha, M.B., Huang, C., 2022. Climate projections over the Great Lakes Region: using two-way coupling of a regional climate model with a 3-D lake model. Geosci. Model Dev. 15, 4425–4446. https://doi.org/10.5194/gmd-15-4425-2022.

Yousefi, A., Toffolon, M., 2022. Critical factors for the use of machine learning to predict lake surface water temperature. J. Hydrol. 606, 127418. https://doi.org/10.1016/j.jhydrol.2021.127418.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., Ishii, M., 2019. The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: description and basic evaluation of the physical component. J. Meteorol. Soc. Jpn. II 97, 931–965. https://doi.org/10.2151/jmsj.2019-051.

Zahir, M., Su, Y., Shahzad, M.I., Ayub, G., Rehman, S.U., Ijaz, J., 2024. A review on monitoring, forecasting, and early warning of harmful algal bloom. Aquaculture 741351. https://doi.org/10.1016/j.aquaculture.2024.741351.

Zhu, S., Ptak, M., Yaseen, Z.M., Dai, J., Sivakumar, B., 2020. Forecasting surface water temperature in lakes: a comparison of approaches. J. Hydrol. 585, 124809. https://doi.org/10.1016/j.jhydrol.2020.124809.

Zingone, A., Oksfeldt Enevoldsen, H., 2000. The diversity of harmful algal blooms: a challenge for science and management. Ocean. Coast. Manage 43, 725–748. https://doi.org/10.1016/S0964-5691(00)00056-9.

Ziyad Sami, B.F., Latif, S.D., Ahmed, A.N., Chow, M.F., Murti, M.A., Suhendi, A., Ziyad Sami, B.H., Wong, J.K., Birima, A.H., El-Shafie, A., 2022. Machine learning algorithm as a sustainable tool for dissolved oxygen prediction: a case study of Feitsui Reservoir. Taiwan. Sci. Rep. 12, 3649. https://doi.org/10.1038/s41598-022-06969-z.

Zwart, J.A., Hanson, Z.J., Read, J.S., Fienen, M.N., Hamlet, A.F., Bolster, D., Jones, S.E., 2019. Cross-scale interactions dictate regional lake carbon flux and productivity response to future climate. Geophys. Res. Lett. 46, 8840–8851. https://doi.org/10.1029/2019GL083478.