Uncertainty in projections of future lake thermal dynamics is differentially driven by lake and global climate models

Jacob H. Wynne[1,2], Whitney M. Woelmer[1], Tadhg N. Moore[1,3], R. Quinn Thomas[1,4], Kathleen C. Weathers[5], Cayelan C. Carey[1]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA

[2]Current address: Department of Microbiology, Oregon State University, Corvallis, Oregon, USA

[3]Current address: School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

[4]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, USA

[5]Cary Institute of Ecosystem Studies, Millbrook, New York, USA

Corresponding Author:

Whitney Woelmer[1]

926 W. Campus Drive, Blacksburg, VA, 24060 USA

Email: wwoelmer@vt.edu

Abstract

Freshwater ecosystems provide vital services, yet are facing increasing risks from global change. In particular, lake thermal dynamics have been altered around the world as a result of climate change, necessitating a predictive understanding of how climate will continue to alter lakes in the future as well as the associated uncertainty in these predictions. Numerous sources of uncertainty affect projections of future lake conditions but few are quantified, limiting the use of lake modeling projections as management tools. To quantify and evaluate the effects of two potentially important sources of uncertainty, lake model selection uncertainty and climate model selection uncertainty, we developed ensemble projections of lake thermal dynamics for a dimictic lake in New Hampshire, USA (Lake Sunapee). Our ensemble projections used four different climate models as inputs to five vertical one-dimensional (1-D) hydrodynamic lake models under three different climate change scenarios to simulate thermal metrics from 2006 to 2099. We found that almost all the lake thermal metrics modeled (surface water temperature, bottom water temperature, Schmidt stability, stratification duration, and ice cover, but not thermocline depth) are projected to change over the next century. Importantly, we found that the dominant source of uncertainty varied among the thermal metrics, as thermal metrics associated with the surface waters (surface water temperature, total ice duration) were driven primarily by climate model selection uncertainty, while metrics associated with deeper depths (bottom water temperature, stratification duration) were dominated by lake model selection uncertainty. Consequently, our results indicate

that researchers generating projections of lake bottom water metrics should prioritize including multiple lake models for best capturing projection uncertainty, while those focusing on lake surface metrics should prioritize including multiple climate models. Overall, our ensemble modeling study reveals important information on how climate change will affect lake thermal properties, and also provides some of the first analyses on how climate model selection uncertainty and lake model selection uncertainty interact to affect projections of future lake dynamics.

Introduction

Freshwater ecosystems provide essential ecosystem services, including water for drinking, irrigation, and fisheries, and substantial cultural and economic value (Janssen et al., 2021). However, freshwater ecosystems have been severely affected by human activities (IPCC, 2021), with abrupt and severe water quality degradation occurring in response to climate change (Woolway et al., 2019; Ho & Michalak, 2020), which is predicted to accelerate in the future (Weyhenmeyer, Westöö & Willén, 2008; Sharma et al., 2019; Woolway & Merchant, 2019). Thus, there is an increasing need for model projections that represent future lake ecosystem conditions to help decision-makers anticipate, prepare for, and potentially mitigate changes in lake ecosystem services (Brookes et al., 2014; Khan et al., 2015).

Among lake water quality metrics, lake thermal structure (which encompasses water column temperatures, stratification, and duration of ice coverage) plays a key role in lake ecosystem functioning and is extremely sensitive to altered climate (O'Reilly et al., 2015; Woolway & Merchant, 2018; Sharma et al., 2021b; Woolway, Sharma & Smol, 2022). For example, thermal stratification (i.e., the presence of a strong temperature gradient from the surface to the bottom of the lake) directly influences mixing regimes (the yearly pattern of thermal stratification; Lewis Jr., 1983; Wetzel, 2001), which are expected to shift in lakes under most climate change scenarios (Woolway & Merchant, 2019). For example, in many temperate, dimictic lakes, Woolway & Merchant (2019) projected a shift from two mixing events annually to a single mixing event as lakes lose ice cover. Mixing regimes have major implications for lake ecological processes such as primary productivity, availability of fish habitat, nutrient availability, and atmospheric gas exchange (Wetzel, 2001; Kirillin, 2010; Richardson et al., 2017). In addition, the duration of ice cover, which directly influences many lake mixing regimes, is expected to decrease on average by 29 ± 8 days by 2080-2100 in seasonally ice-covered lakes globally under possible future climate change scenarios (Woolway & Merchant, 2019). Changes in ice cover can fundamentally alter lake ecosystems, including lake hydrodynamics, oxygen availability, and organismal habitat (Salonen et al., 2009; Hampton et al., 2017; Flaim et al., 2020). Overall, given that different metrics of lake thermal structure will likely have varied future responses to climate change, developing projections for multiple thermal metrics - as well as quantifying their associated uncertainty - is critical when considering the ecological impacts of climate change on lakes.

To better interpret and appropriately use projections of future lake thermal structure, quantifying the different sources of uncertainty associated with model projections is critical for bounding predictions of future lake ecosystem changes. While many sources of uncertainty exist, the sources most commonly quantified in environmental forecasts and projections are model driver data uncertainty (i.e., uncertainty in estimates of model inputs, such as meteorological driver data), initial condition uncertainty (i.e., uncertainty in the model's initial states), observational uncertainty (i.e., uncertainty in the actual measurement of the variables being modeled), parameter uncertainty (i.e., uncertainty in the values of a model's parameters), and model process uncertainty (i.e., uncertainty in the modeled representation of complex ecosystem processes within a model; Dietze, 2017; Her et al., 2019; Heilman et al., 2022; Golub et al., 2022). Process model selection uncertainty (i.e., the uncertainty in having multiple process models simulate the same target variable) and driver model selection uncertainty (i.e., the uncertainty in having multiple driver models simulate the same variable that becomes input to the process model) are largely overlooked, and, to the best of our knowledge, have never been compared with one another in lake thermal projection studies (Moore et al., 2021; Feldbauer et al., 2022). While interactions across some of these uncertainty types have been examined in projection studies in other ecosystems (Wada et al., 2013; Hoan, Khoi & Nhi, 2020; Heilman et al., 2022), the relative importance of different sources of uncertainty remains largely unexplored in lake projections.

One commonly-used method to estimate uncertainty in projections is ensemble modeling (Parker, 2011). For longer-term lake projections specifically, this approach entails using one or more climate scenarios, fed into one or more climate models that generate weather data, which are then used as inputs (i.e., drivers) to one or more lake models to produce an ensemble projection of different lake thermal metrics. Comparing the output of ensemble members provides a more realistic representation of the diverse spread of model outcomes, as well as an opportunity to examine the effects of interactions between climate models and lake models on lake projections. Importantly, by predicting lake thermal properties using multiple lake models and climate models, we can better quantify the uncertainty in future lake responses to climate change.

To date, long-term (decade to century) lake thermal projection studies have generally only quantified one or two possible sources of uncertainty in ensemble modeling. For these long-term studies, lake model selection uncertainty (i.e., uncertainty derived from the decision of choosing a single lake process model from a suite of possible models), driver uncertainty, and parameter uncertainty are occasionally examined individually, with comparisons across uncertainty sources largely neglected (Kobler & Schmid, 2019; Golub et al., 2022). Feldbauer et al. (2022) found that lake model selection uncertainty was the dominant contributor in projections of water temperature for a German reservoir. However, they used a single climate model to produce driver inputs and therefore did not account for the role of climate model selection uncertainty, which limits their ability to quantify the overall and proportional contribution of both climate model and lake model selection uncertainty. Climate model selection uncertainty, a type of driver data uncertainty derived from the decision of choosing a single climate driver model from a suite of possible climate driver models, has previously been examined in projections for other ecosystems but not lakes specifically. Thus, an opportunity to expand on Feldbauer et al. (2022) would include analyzing both lake model selection uncertainty and climate model selection uncertainty with an ensemble approach using multiple climate scenarios, as we are unaware of any studies that analyze uncertainty across all three components of multiple climate scenarios, climate models, and lake process models.

In this study, we generated ensemble projections of future lake thermal dynamics for a dimictic lake in New Hampshire, USA (Lake Sunapee). We quantified the role of lake model selection uncertainty and climate model selection uncertainty in lake thermal projections for this seasonally ice-covered lake by using four General Circulation Models (GCMs) as inputs to five lake models. Our ensemble projections also used three climate scenarios as inputs to the GCMs, which represent a range of possible greenhouse gas radiative forcing pathways. We made projections for a set of lake thermal metrics over the next century relative to a historical baseline. Our objectives were to: 1) project future thermal structure for Lake Sunapee with quantified uncertainty up to the year 2099, 2) examine uncertainty dynamics among lake models, climate models, and their interactions between mid-century and end of century, and 3) partition the relative contributions of climate model selection uncertainty (from the four GCMs) and lake model selection uncertainty (from the five lake models) over time for each thermal metric. Altogether, our study aimed to improve our understanding of lake thermal responses to climate change and quantify the contribution of overlooked sources of uncertainty for projections of future lake dynamics.

Materials & Methods

Overview. To create ensemble projections of lake thermal dynamics, we used four GCMs to drive five vertical one-dimensional (1-D) hydrodynamic lake models. To explore the effects of different potential radiative forcing scenarios on future climate, we used GCM output from three representative concentration pathways (RCPs). Each lake model was calibrated with ten years of historical water temperature data (2005-2015) using standard methods (described below in *Materials & Methods: Lake Model Calibration and Validation*) to a minimum Root Mean Square Error (RMSE) for six lake thermal metrics (described below in *Materials & Methods: Thermal Metrics*). The relative performance of each model following calibration was evaluated using five years of validation data (2015-2020; Fig. 1.1). After calibration and validation, projections were run from 1938 to 2099, including a spin-up period (1938-1974) to minimize the impact of initial conditions on the simulations, a historical period used to calculate a historical baseline (1975-2005), and a future climate projection period (2006-2099; Fig. 1.2). Anomalies were calculated for all future projections based on the difference between the historical and projection periods to determine the change in each thermal metric. We grouped projection output by lake model and GCM to determine model interactions over mid- and end-century. Lastly, we partitioned the relative contributions of lake model selection uncertainty and GCM climate model selection uncertainty over time for all thermal metrics across the paired RCP, GCM, and lake model combinations (Fig. 1.3). We note that our study calculated lake model and climate model selection uncertainty using an "ensemble of opportunity" (following Tebaldi & Knutti, 2007), in that we quantified the uncertainty in our selection of lake models and climate models across a set of available models, which was not exhaustive of all possible ways the system could be modeled (Parker, 2011).

Study site. We generated lake thermal projections for Lake Sunapee, an oligotrophic lake located between Merrimack and Sullivan Counties in New Hampshire, USA (43.37, -72.05; Fig. 2). The lake is deep (Z_{maximum} = 33 m) and dimictic, with ice cover ranging from December or January until March or April (Bruesewitz et al., 2015; LSPA & Town of Sunapee, 2022). From 2016 to 2020, the mean observed ice duration was 102 days, with a range between 55 and 128 days. Summer stratification typically occurs from mid-June to late September, with a summer thermocline depth of 6-8 m (Carey et al., 2014). From 1979 to present, the Lake Sunapee region has experienced an increase in observed air temperature at a rate of 0.42 °C per decade and substantial land use change in the surrounding catchment (Cobourn et al., 2018; Ward et al., 2020).

Representative Concentration Pathways. To encompass a range of potential future climate scenarios, we used three representative concentration pathways (RCPs) in our study: RCP 2.6, RCP 6.0, and RCP 8.5 (Table S1). RCP scenarios were developed by the Intergovernmental Panel on Climate Change (IPCC) based on several socioeconomic factors, including land use and cover data, as well as greenhouse gas emissions (van Vuuren et al., 2011; Frieler et al., 2017). The three RCP scenarios range from low to high climate forcing impact, with RCP 2.6 having the lowest radiative forcing level of 2.6 W/m^2 by the end of the century, RCP 6.0 representing a medium forcing level of 6.0 W/m^2 by the end of the century, and RCP 8.5 representing a high forcing level of 8.5 W/m^2 by the end of the century (Table S1). To date, RCP 8.5 is the best match with current trends to at least mid-century under current and stated policies (Schwalm, Glendon & Duffy, 2020).

General Circulation Models. To represent the atmospheric conditions of future climate, we coupled each RCP scenario with four general circulation models (GCMs): MIROC5, IPSL-CM5A-LR, GFDL-ESM2M, and HADGEM2-ES (Table 1). The models were developed in four different climate research laboratories across the world, using different approaches to represent global climate processes. This model-grouping protocol follows the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), which is an international effort to better understand climate projections and their uncertainties using ensemble modeling (Frieler et al., 2017; Golub et al., 2022). All GCM outputs were downloaded from the ISIMIP database, where they were bias-corrected to a 0.5°x0.5° grid using the CDF-t (cumulative distribution function - transform) method (Lange,

2017) with the EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI) data (Table 1), and archived with all model code for this analysis for reproducibility (Wynne et al., 2022b). GCM output ranged from 1861-2005 based on historically reconstructed climatic conditions representing the industrialization period, as well as from 2006 to 2099, based on each individual RCP.

Lake models. To represent future thermal conditions within Lake Sunapee, we coupled RCP-GCM output with an ensemble of five one-dimensional (1-D) lake models which simulate lake thermal properties (Table 1). These models, which use different deterministic modeling approaches to simulate lake hydrodynamics with model-specific parameters and calculations, include the: 1) Freshwater Lake model (FLake) v1.0, which simulates lake systems using a two-layer parametric representation focusing on heat budget (Mironov, 2010); 2) General Lake Model (GLM) v3.1.0, which applies a Lagrangian structure to replicate mixing dynamics (Hipsey et al., 2019); 3) General Ocean Turbulence Model (GOTM) v3.2, which is a vertical 1-D hydrodynamic k-epsilon turbulence model (Li et al., 2021); 4) Multi-year Lake simulation model (MyLake) v1.2, which simulates daily vertical profiles of water temperature, seasonal ice and snow cover as well as other variables (Saloranta & Andersen, 2007); and 5) Simstrat v2.4.1, which is a vertical 1-D hydrodynamic model combining a buoyancy-extended k-epsilon model with seiche parameterization (Table 1; (Moore et al., 2021). Each model uses daily meteorological and lake hypsography input data to simulate water column temperature and ice cover on a daily timescale. We used LakeEnsemblR (LER), an R package for predicting lake thermal dynamics using a suite of lake models (Moore et al., 2021), for the ensemble lake modeling in our study. We used the LakeEnsemblR model output over our calibration, validation, and projection time periods to calculate the lake thermal metrics described below. These five models have been iteratively updated over time, however, we used the same model versions as Moore et al.'s (2021) LakeEnsemblR v1.0 release to maintain consistency throughout our study and for comparison with that earlier work.

Thermal metrics. We chose six ecologically important thermal metrics for comparison across models: summer mean surface and summer mean bottom temperatures (hereafter, surface temperature and bottom temperature, respectively), Schimdt stability, thermocline depth, summer stratification duration, and total ice cover duration. Schmidt stability (the stability of a water body's thermal stratification and its resistance to mixing, in J/m^2) and thermocline depth (the depth of greatest density change in the water column due to differences in water temperature) were calculated using the package rLakeAnalyzer (Read et al., 2014). The summer (June-August) mean surface and bottom temperatures, summer stratification duration (length of time stratified over the whole year not including inverse stratification during ice cover), as well as total length of ice duration were calculated using the LakeEnsemblR package (Moore et al., 2021).

Input data for lake model calibration and evaluation. Observations of water temperature at Lake Sunapee from multiple data sources were used to calibrate and validate the five hydrodynamic models in LER. First, water temperature observations were collected approximately every meter from the surface to 30.0 meters using manual thermal profile measurements collected approximately monthly in the summer from 1986-2021 (Steele, Weathers & Association, 2021). Second, a monitoring buoy was deployed by the Lake Sunapee Protective Association (LSPA) in 2007, providing high-frequency (10-minute collection interval) temperature profiles every meter from the surface to 10.0 m from 2007 to present (LSPA, Steele & Weathers, 2022a). Observations of the date of ice-off have been collected yearly since 1869 and ice-on since 2016 (LSPA & Sunapee, 2022) were used during both calibration and validation to estimate the ability of the models to simulate ice-off dynamics.

Lake hypsography and meteorological data were used as lake model inputs during calibration and validation. The EWEMBI data product from ISIMIP (1979-2016; Lange, 2019) was used as meteorological forcing data for calibrating all LER models. EWEMBI data were used in place of locally collected meteorological data to calibrate the lake models to maintain consistency with GCM projections, which were bias-corrected by ISIMIP using EWEMBI data. Because the EWEMBI data product was revised in 2016, we used the next-generation EWEMBI data product ECMWF Reanalysis v5 (ERA5) to drive the lake models during validation. A comparison between EWEMBI and ERA5 during an overlapping time period of 1975-2016 showed very similar trends between the two data products indicating negligible influence on our lake temperature simulations (Fig. S1).

Lake model calibration and validation. All five of the lake models in LER (Table 1) were calibrated over a 10-year period (27 June 2005 to 1 January 2015) and validated over a five-year period (11 June 2015-1 January 2020; Fig. 1). The calibration years were chosen because they covered a wide range in annual temperature and precipitation (LSPA, Steele & Weathers, 2022b) and included the only continuous winter of high-frequency thermal profiles under ice (2007-2008; Bruesewitz et al., 2015; Brentrup et al., 2021). To avoid errors associated with initial conditions, a model spin-up period of 180 days was removed from the beginning of the calibration period as well as the validation period.

Due to limited availability of observational data during the winter, calibration and validation of the lake thermal metrics (except for ice duration) focused on summer conditions. Specifically, during calibration and validation periods, we compared the daily average of surface and bottom temperatures to observations during only summer months (June-August). Similarly, due to the low data availability of deep-water temperature profiles, we calculated observed stratification duration only using observed and modeled output from the surface to 10.0 m, which encompassed typical summer thermocline depths but missed deeper temperatures. We removed 2007 from the calibration period due to limited data availability, as it had zero days of observations during the summer stratified period. Because the data record for ice-on began only recently relative to the record for ice-off (Bruesewitz et al., 2015; LSPA & Sunapee, 2022), we compared observations of the date of ice-off to model output during calibration and validation, while during the projection time period we calculated the total length of ice duration (between ice onset and ice-off).

Calibration was carried out using Latin Hypercube (LHC) sampling of parameters within LER, which uses upper and lower bounds for selected parameters and samples evenly within the bounded parameter space (Mckay, Beckman & Conover, 2000). The scaling factors for wind speed and shortwave radiation were calibrated for all models. However, due to structural differences among models, different model-specific parameters were calibrated for each model, which resulted in slight variation in performance among models (Moore et al., 2021). Because the LHC calibrated all models using the same method but targeted different key parameters within each model, there was a small range of model skill across all five models, but all were within reasonable skill levels for the whole water column RMSE (~2°C; Moore et al., 2021). We calibrated each of the models to match either the whole water column temperature (GLM, GOTM, Simstrat, MyLake) or the temperature to the mean water column depth (FLake). We chose to use mean water column depth for all Flake simulations as the model was shown to have better performance using this method in comparison with full water column depth (Woolway & Merchant, 2019).

Following calibration of the whole water column temperature, we assessed each of our six target thermal metrics (see *Materials & Methods: Thermal Metrics*) using RMSE and bias (mean error), as well as a Taylor diagram. Taylor diagrams evaluate multiple aspects of complex models by quantifying and visualizing the correlation (shown via straight dotted lines), centered root-mean square difference (shown via arced lines), and the magnitude of variability represented by standard deviation of the observations (x-axis) and the model output (y-axis; Taylor, 2001). We compared observations to modeled values of daily mean surface and bottom temperature and annual mean Schmidt stability, thermocline depth, total stratification duration, and total ice duration. After calibration, we used the final parameters from the calibration period (2005-2015) to validate our models from 2015-2020, for which we used the same goodness-of-fit metrics (RMSE, bias). While maintaining constant parameter values prevented us from quantifying the contribution of parameter uncertainty in our lake thermal projections, other lake and hydrological modeling studies have found parameter uncertainty to contribute a small fraction of total uncertainty (Her et al., 2019; Thomas et al., 2020), and our overarching goal was to isolate the effects of climate model and lake model selection uncertainty.

Climate projections. Following calibration, lake models were run for the entire simulation period from 1938-2099. This included a spin-up period (1938-1974), historical mean calculation period (1975-2005), and a climate projection period (2006-2099; Fig. 1). From 1938-2005, the GCMs were run using a historical reconstruction, while from 2006-2099, GCMs were run using three

RCP climate scenarios.

To represent deviation from historical trends, we calculated an annual anomaly from the historical baseline (1975-2005) for each thermal metric over our projection time period (2006-2099). First, we calculated the mean annual value for each thermal metric over the entire historical mean calculation period, except for surface and bottom water temperatures, for which we calculated a summer seasonal average across the summer period for each year. Then, anomalies from this historical period were calculated on a daily time step within the projection period for each thermal metric and averaged to an annual value for each metric.

Ensemble means across RCP scenarios. From our 60 unique model projections (3 RCP scenarios $\times$ 4 GCMs $\times$ 5 lake models), we calculated ensemble means aggregated for each RCP scenario to summarize the impact of each climate scenario on six chosen thermal metrics (*Methods and Materials: Thermal Metrics)*. These ensemble means included all GCMs and lake model projections, resulting in three ensemble means of 20 projections (4 GCMs $\times$ 5 lake models) for each of the six thermal metrics. Ensemble means were calculated annually over the entire projection time period (2006-2099). To represent uncertainty across projections, we also calculated the standard deviation across the 20 projections for each RCP scenario.

Ensemble model interactions. Ensemble modeling incorporates multiple models which represent the same processes in different ways, resulting in distinct dynamics for each model output (e.g., differences between GLM and Simstrat). Additionally, different combinations across model types (e.g., a GCM and a lake model) can potentially produce unique interactions which may further influence the magnitude and direction of model output. To examine the effects of individual models as well as the interactions between model types, we calculated model-type specific ensemble means which captured variability across all GCMs for a single lake model and vice versa. We did this separately for each RCP scenario.

To calculate model-type ensemble means, within a single RCP scenario, projection output was grouped by uncertainty type (e.g., a GCM or lake model), and the mean was calculated across an individual model within each type. For example, to compare the impact of an individual GCM (e.g., GFDL-ESM2M) on all lake models, the ensemble mean was calculated across all five lake model outputs annually (and then grouped by 30-year periods) driven by GFDL-ESM2M climate data (Fig. S2). Similarly, to calculate the impact of a single lake model (e.g., GOTM), the mean of all five GCM $\times$ GOTM outputs was taken (Fig. S2).

From the annual model-type means, we further grouped each model-type mean by 30-year intervals and examined distributions at mid-century (2020-2050) and end-century (2069-2099). Examining projection output over these 30-year intervals reduces climatic noise due to inter-annual variation in climate projections (Fischer et al., 2012). Within the model-type ensemble means (e.g., distribution of individual outputs that make up a model-type; Fig. S2), we specifically looked for multimodality in the distributions, which would signify a disagreement between models, and unimodality, which would signify a high level of agreement between models. We also examined the level of agreement across model-type means (e.g., lake models vs. GCMs) by examining their respective distributions with one another. Distributions with similar means and ranges of projected values among model-types were considered more robust than those with very different distributions. Within a distribution, we identified individual lake and GCM models which resulted in outliers by examining the full time series of each model-type mean (Fig. S3-S8).

Uncertainty partitioning and quantification. To determine the relative influence of climate model selection uncertainty and lake model selection uncertainty on total projection uncertainty, we partitioned the relative contribution of each uncertainty type for each lake thermal metric across all projections (n = 60 total). For each thermal metric, we calculated lake model selection uncertainty by calculating the variance across the five lake models for all GCM and RCP combinations (Moore et al., 2021). Further, we calculated climate model selection uncertainty by calculating the variance across the four GCMs for all lake model and RCP combinations. Total projection uncertainty was calculated as the total standard deviation across all projection outputs, which included climate model selection uncertainty and lake model selection uncertainty. To calculate proportional variance for the two uncertainty types, we divided the variance of each respective uncertainty type by their combined total uncertainty. In the results below, we primarily report our findings of RCP 8.5 for the uncertainty analyses because it has been shown to be the most accurate RCP relative to recent CO_2 concentrations and global temperatures (Schwalm, Glendon & Duffy, 2020); all other RCP scenario results presented in the SI (Fig. S3-S6).

All analyses were conducted in R v.4.0.2 (R Core Team, 2020). The water temperature observations and ice data are available in the Environmental Data Initiative repository (Steele, Weathers & Association, 2021; LSPA, Steele & Weathers, 2022a; LSPA & Town of Sunapee, 2022). All code to run the analyses, including the downloaded GCM x RCP output and model initialization files are available on Zenodo (Wynne et al., 2022a, Wynne et al. 2022b).

Results

Lake models reproduce observed thermal dynamics. Throughout the calibration period, all lake models reproduced observed Lake Sunapee thermal structure and stratification patterns (Fig. 3). Four out of five of the lake models reproduced modeled whole water column temperature with a root mean square error (RMSE) of <2 °C for whole water column temperatures (Table S2). FLake was the one exception among the five lake models, and exhibited a 2.23 °C RMSE (Table S2). The ensemble mean across all lake models performed as well or better for multiple metrics during calibration and validation compared to the best performing individual model, with a whole water temperature RMSE of 1.29 °C, and bias of -0.15 °C (Tables S2, S3).

Within the water column, model performance for water temperature varied with depth. All models reproduced surface temperature observations well, with high correlation between modeled output and observations (r=0.99, Fig. 4A), and RMSE <1.51 °C (Table S2). In contrast, the models reproduced observed bottom temperature with less skill, with RMSE ranging from 2.52-4.69 °C (Table S2) and r<0.4 (Fig. 4B). Because FLake only simulates the surface layer of lakes (see *Materials & Methods: Calibration*), output from this model was removed from the bottom water temperature calibration and validation calculations. Similarly, MyLake only simulates from the surface to 1.0 m above the bottom (here, 32.0m) and thus its output was analyzed accordingly across these depths.

All models reproduced Schmidt stability well but were less skillful at capturing summer thermocline depth and summer stratification duration. Modeled Schmidt stability from FLake exhibited a high correlation to the observed data (r>0.95, Fig. 4), but a low standard deviation compared to the observations, indicating that the variability in modeled output was smaller than observed variation (Fig. 4C). Overall, Simstrat reproduced Schmidt stability better than the other models (Fig. 4C; RMSE = 46.36 J/m^2, Table S2). Summer thermocline depth was reproduced poorly by all lake models, with r<0.40 (Fig. 4D) and a lower standard deviation for the modeled output than the observations. The worst performing model for thermocline depth was FLake, with r<0.10 (Fig. 4D). Summer stratification duration was also poorly reproduced, similar to thermocline depth, with an RMSE among models of 34-75 days and overall correlation of r<0.40 (Table S2, Fig. 4E). However, these poor goodness-of-fit values were likely due to the low temporal resolution of the observed data, which limited our ability to compare thermocline and stratification dynamics.

The day of ice-off was captured well by Simstrat, MyLake and FLake, with high goodness-of-fit metrics across these models (r>0.95, RMSE ~5 days) and similar standard deviation in comparison to the observed data (Table S2, Fig. 4F). In contrast, GOTM did not perform well in predicting the day of ice-off, with a higher RMSE (60 days; Table S2) and bias (59 days; Table S3) than all other models. The version of GLM (v3.1.0) in LakeEnsemblR (v1.0) did not simulate ice and thus was not included in the ice goodness-of-fit calculations or projections.

Warmer, more summer-stratified projections of Lake Sunapee's future. Our projections show that all six thermal metrics of Lake Sunapee will change substantially in response to climate change over the next century (Fig. 5). Surface temperature is projected to increase by 1-5 °C above historical conditions by the end of this century (Fig. 5A). Similarly, bottom temperature is also projected to increase, but to a lesser extent (1-2 °C; Fig. 5B). Metrics of stratification indicate a longer and stronger summer stratification period within the lake annually, with the duration of summer stratification increasing by 10-40 days (Fig. 5E). In addition, the strength of thermal stratification, Schmidt stability, is projected to increase by 20-100 J/m^2 (Fig. 5C). In contrast, total ice duration is projected to decrease by 20-75 days (Fig. 5F). Interestingly, thermocline depth is projected to stay the same over the course of the century, with increased variability by the end of the century (Fig. 5D).

As a result of longer summer stratification and less ice cover, Lake Sunapee's mixing dynamics will be altered, with up to 50 additional days spent stratified in the summer and up to 75 fewer days spent inversely stratified due to decreasing ice cover in the winter. The magnitude of anomalies for each metric were largely driven by RCP scenario, with the smallest ranges projected by RCP 2.6, followed by RCP 6.0, and the highest ranges projected by RCP 8.5 (Fig. 5F), which largely followed expected patterns corresponding to the magnitude of climate change associated with each RCP scenario. For example, under RCP 2.6, which includes reduced carbon emissions by mid-century (van Vuuren et al., 2011), anomalies decreased from mid- to end-century, with lower predicted temperatures, less change in stratification duration, and lower Schmidt stability values at end-century compared to mid-century, in line with the socioeconomic trajectory of this scenario.

Interactions between models from mid-century and end-century projections show model disagreement for lake bottom temperature and ice cover. From mid- to end-century, distributions of model-type means (ensemble means grouped by climate or lake models) varied with thermal metric (Fig. 6) as well as RCP scenario (Figs. S9-S10). Surface temperature had a wider climate model distribution for mid- and end-century than lake model distribution, indicating a wider range of possible values due to differences across climate models (Fig. 6A). However, within a model-type both climate and lake model distributions were unimodal, indicating a high degree of model agreement for surface temperature projections across models. In contrast, mid- and end-century bottom temperature projections showed a bimodal response due to differences across lake models but not across climate models (Fig. 6B). The bimodality was primarily due to GLM performance, in which GLM projected minimal changes in bottom temperatures under all climate models (Fig. S8B), as opposed to other lake models, which projected an increase of 2-4ºC.

In contrast to bottom temperature, Schmidt stability had a high degree of agreement between lake models and climate models, as shown by similar distributions between lake and climate models in mid-century and end-century (Fig. 6C). While distributions of thermocline depth anomalies were centered around zero, indicating minimal change at both mid-century and end-century, there was a large spread in the distribution of lake models as compared to GCMs by end-century, showing a wide range of disagreement in future thermocline depth due to differences in lake models (Fig. 6D). Similar to bottom temperature, this was primarily driven by GLM output, which projected a larger increase in thermocline depth (i.e., negative anomaly) than other lake models (Fig. S8). FLake also contributed to the disagreement in projections of thermocline depth, as the only model which projected a decrease in thermocline depth (i.e., positive anomaly). However, this result should be interpreted with caution due to FLake's poor fit in simulating thermocline depth during the calibration and validation period

(Fig. 4, Table S2) as well as that it only simulated water temperature to the mean water column depth and not the whole water column (see *Materials & Methods: Calibration*).

The model-type distributions showed model disagreement within and across model-types. Summer stratification duration had a bimodal distribution across lake models, particularly by end-century (Fig. 6E), driven by a lower projected anomaly from FLake (Fig. S7E). Thermocline depth showed a similar pattern (Fig. 6D). Lastly, total ice duration was generally unimodal for both climate and lake models, indicating agreement within model-type (Fig. 6F). However, for the end-century projections, the lake model distribution was centered around -40 (i.e., 40 fewer days of ice) while the climate model distribution was centered around -75 (i.e., 75 fewer days of ice). The difference in these distributions suggest a greater degree of ice loss due to variation in climate models than lake models.

The dominant source of uncertainty varied over time and thermal metric. The relative proportion of uncertainty due to climate and lake models varied among thermal metrics and over time for RCP 8.5 scenarios (Fig. 7). Uncertainty in surface temperature was consistently dominated by climate model selection uncertainty (>80%) throughout the entire projection period (Fig. 7A). In contrast, bottom temperature was dominated by climate model selection uncertainty up until mid-century, after which ~75% of uncertainty was due to lake model selection uncertainty (Fig. 7B). Uncertainty in Schmidt stability was dominated by climate model selection uncertainty until mid-century (~75%), after which lake model and climate model selection uncertainty contributed equally (Fig. 7C). Uncertainty in thermocline depth was evenly split by lake model and climate model selection uncertainty from the beginning of the projection period, but lake model selection uncertainty increased over time to an overall proportion of >75% by the end of the century (Fig. 7D). Total stratification duration was initially dominated by climate model selection uncertainty until mid-century (~75%), when lake model selection uncertainty became the primary source (~75%). Lastly, uncertainty in total ice duration was dominated by climate model selection uncertainty over the entire projection period (60-75%; Fig. 7E).

In all metrics but surface temperature, the proportional contribution of lake model selection uncertainty increased over time (Fig. 7). In contrast, no metrics exhibited a shift from lake model dominance to climate model dominance in proportional uncertainty over time. Specifically, for bottom temperature and total stratification duration, the dominant source of uncertainty switched from climate model to lake model mid-century (Fig. 7B, 7E). For surface temperature and total ice duration, climate model selection remained the dominant source of uncertainty throughout the projection time period (Fig. 7A, 7F). Interestingly, lake model selection uncertainty in surface temperature, Schmidt stability, thermocline depth, and total stratification duration did not increase at a constant rate and increased more quickly in the beginning of the projection period up to mid-century (Fig. 7).

The dominant source of uncertainty varied among RCP scenarios for some thermal metrics (Figs. S6, S7). Under RCP 2.6, climate model selection uncertainty was the dominant source of uncertainty for all thermal metrics but thermocline depth (Fig. S11). For bottom temperature and Schmidt stability, the relative contribution of lake model selection uncertainty increased up to mid-century and then began decreasing towards end-century. Under RCP 6.0, uncertainty dynamics were similar to RCP 8.5, in which climate model selection uncertainty was the dominant source for surface temperature and total ice duration, while lake model selection uncertainty was the dominant source for bottom temperature, thermocline depth, and total stratification duration by end-century (Fig. 7D). Schmidt stability had similar contributions between GLM and lake model selection uncertainty for both RCP 6.0 and RCP 8.5 (Fig. 7C). For the metrics dominated by lake model selection uncertainty under RCP 6.0, uncertainty continued to increase to 2099 as opposed to leveling off (Fig. S12B, S12C, S12E), similar to RCP 8.5 (Fig. 7B, 7C, 7E), which exhibited increased uncertainty earlier in the projection period.

Discussion

Overview. If decision-makers are to prepare for and mitigate the effects of climate change on lakes, they must have access to robust model projections with quantified uncertainties. Consequently, identifying the dominant sources of uncertainty (e.g., lake models, climate models) and their interactions are critical for both improving the accuracy of future projections and their interpretation, as well as building more robust coupled model frameworks. This study approaches these challenges by using ensemble modeling across multiple climate models, lake models, and socioeconomically-driven climate scenarios. Coupling ensembles of multiple model-types increases the robustness of our projections, as previous lake modeling studies generally have only used just one climate model, one lake model, and/or one climate scenario (e.g., Her et al., 2019; Golub et al., 2022; Feldbauer et al., 2022).

Across our ensemble projections, the majority of the thermal changes projected for Lake Sunapee were in line with previous studies, which predict warmer, longer summer stratification periods and less winter ice cover. While most models in our ensemble projections agreed on the projected changes in thermal metrics, some models exhibited entirely different trajectories from the ensemble. For example, GLM projected no change in bottom temperature, compared to the overall 1-2ºC increase in this metric projected by FLake, MyLake, Simstrat, and GOTM (Fig. 6, Fig. S3). Lastly, we found that the dominant source of uncertainty for each thermal metric was sensitive to depth (surface-level metrics were more sensitive to variation in the climate model), time (the dominant source of uncertainty changed mid-century for most metrics), and climate scenario (variation among lake models increased more quickly into the future under more severe RCPs). Below, we explore further our uncertainty findings, as well as the projected changes to thermal metrics in Lake Sunapee.

Dominant uncertainties vary for each thermal stratification metric: depth matters. Thermal metrics for the lake surface were affected more by variation across the climate models used as inputs, while metrics within the water column were affected more by variation across lake models (Fig. 7). Specifically, surface temperatures and ice duration were dominated by climate model selection uncertainty. In contrast, the dominant source of uncertainty for bottom temperature, thermocline depth, and stratification duration was lake model selection. Schmidt stability, which incorporates full water column dynamics, was equally dominated by climate model and lake model selection uncertainty by end-century.

Our findings are likely due to the direct influence of atmospheric conditions on surface temperature and ice cover, as opposed to hypolimnetic or whole-water column metrics (e.g., bottom temperatures, stratification metrics), which may be more sensitive to within-lake thermal stratification and mixing processes (Kraemer et al., 2015). While few studies exist which separately partition the relative contributions of different sources of uncertainty in predicted surface vs. bottom lake hydrodynamics, it has been previously shown that surface water thermal dynamics, including ice cover, are primarily driven by atmospheric forcing (Livingstone & Padisák, 2007; Sharma, Walker & Jackson, 2008; Piccolroaz, Toffolon & Majone, 2013) and that bottom waters are more weakly correlated to changing air temperatures (Butcher et al., 2015). Metrics of thermal stratification (e.g., Schmidt stability, thermocline depth) are driven by a suite of processes that were simulated differently in all of the five hydrodynamic lake models, such as sediment-water interactions, inflow dynamics, light transparency, solutes, and other processes (Saloranta & Andersen, 2007; Perroud et al., 2009; Mironov, 2010; Hipsey et al., 2019; Li et al., 2021), which may explain why these metrics were more sensitive to lake model selection uncertainty than climate model selection uncertainty.

Dominant sources of uncertainty in the lake metrics also changed over time into the future. Differences among climate models were the greatest contributor of total uncertainty for all metrics except thermocline depth in the beginning of the century, but switched for most metrics in mid-century. By end-century, bottom temperature, Schmidt stability, thermocline depth, and stratification duration were dominated by lake model selection uncertainty. In contrast, uncertainty in surface temperature as well as ice duration continued to be dominated by climate model selection. Changing uncertainty dynamics over time may indicate how different sources of uncertainty (e.g., climate model selection or lake model selection) propagate through time, having important implications for the projection horizon at which each of these sources dominate.

Comparisons of uncertainty dynamics across time scales: different contributions of uncertainty between short- and long-term predictions. Our uncertainty findings add context to the few studies that quantify similar sources of uncertainty in lake thermal projections and forecasts. In their long-term thermal projection study using the LakeEnsemblR framework for a

German reservoir, Feldbauer et al. (2022) found that lake model selection uncertainty was as high or higher than all other quantified uncertainties (which included a single type of meteorology driver data uncertainty, parameter uncertainty, and their interactions) for surface (3.0 m) and bottom (25.0 m) water temperatures, summer stratification duration, and total ice duration. By quantifying driver selection uncertainty across multiple climate models, as opposed to one, our study builds upon this work to identify the importance of driver model selection uncertainty relative to other sources. Indeed, we found that climate model selection uncertainty was the dominant source of uncertainty in surface-level metrics such as water temperature and total ice duration (Figs. 7A, 7F), highlighting the importance of quantifying uncertainty across multiple climate models for future lake studies. We note that Feldbauer et al. (2022) and our study used different approaches to quantify uncertainty, emphasizing the need for consistent approaches for uncertainty quantification to produce robust comparisons across lakes and analyses.

Comparing uncertainty dynamics across long-term lake projections and short-term forecasts can provide valuable insight into predictability of lake thermal dynamics across time scales. In their near-term forecasting study, Thomas et al. (2020) found that 16-day surface temperature forecasts were dominated by meteorological driver and downscaling uncertainty, while process uncertainty dominated bottom water forecasts almost entirely during the summer stratified period, similar to our findings. However, during the fall mixed period, process and driver uncertainty were nearly equal across surface and bottom water forecasts, suggesting that mixing dynamics are a key factor in the dominant source of uncertainty in lake thermal predictions (Thomas et al., 2020). Due to limited data availability for Lake Sunapee, we did not quantify changes during the mixed period, motivating the need for future work to examine uncertainty dynamics of long-term lake projections during mixed periods.

Overall, short-term lake thermal forecasts may have different uncertainty dynamics than long-term climate change projections, especially under RCP 8.5. For example, Thomas et al. (2020) found that driver uncertainty increased more than process uncertainty with time into the future across their 16-day forecast horizon. Across full century projections, we found the opposite effect, with the contribution of process uncertainty increasing to dominate or nearly dominate total uncertainty by the end of the century for bottom temperature, Schmidt stability, thermocline depth, and stratification duration under RCP 8.5. Altogether, this result emphasizes the importance of partitioning total uncertainty into individual contributions for comparing uncertainty dynamics across forecasts and projection horizons.

Uncertainty findings inform future lake projections: surface- and bottom-level thermal metrics require different uncertainty quantification approaches. Our findings support the importance of using multiple lake models to inform uncertainty dynamics and improve overall projection performance. However, running multiple lake models and climate models poses computational and logistical challenges, necessitating guidelines for how to prioritize conducting multi-model ensembles. For studies focused on estimating primarily surface water variables (e.g., analyses that use satellite data for lake thermal modeling), we suggest focusing on quantifying meteorological driver uncertainty, which can be done by including numerous GCMs (Her et al., 2019) or other relevant meteorological driver models. In contrast, in studies focusing on simulating whole water column stratification metrics, we suggest focusing on quantifying and reducing overall lake model uncertainty. The guidelines we suggest here were developed from our work in dimictic Lake Sunapee, and will be strengthened by additional studies in other lakes, which will test the robustness of these uncertainty contributions across lake and reservoir ecosystems (e.g., shallow lakes, tropical systems, reservoirs).

Importance of ensemble means for lake modeling. Over the calibration and validation period, we found that the ensemble mean across lake models was frequently the highest performing simulation relative to observations (Table S2). This finding is not novel to lake modeling alone, and has been well-documented across other ensemble modeling studies, e.g., in meteorological forecasts (Scher & Messori, 2021), epidemic forecasts (Sharma et al., 2021a), ecological forecasts (Lynch et al., 2012), and political forecasts (Beger, Dorff & Ward, 2014). Within aquatic ecosystems, the ensemble mean of phytoplankton projections from three models was the best predictor of phytoplankton in a temperate lake relative to any individual model (Trolle et al., 2014). While modeling two lakes using the LakeEnsemblR framework, Moore et al. (2021) found that the ensemble mean of water temperature frequently outperformed all other models. The findings by Moore et al. (2021) across two lakes coupled with our findings on Lake Sunapee provides support for the importance of using an ensemble of models for predicting lake thermal structure.

While ensemble modeling often has the benefit of improving model performance, ensembles with models that perform very poorly relative to others may have adverse effects on the overall ensemble mean. For example, we observed individual models that negatively skewed the ensemble mean in comparison to other lake models (e.g., poor performance of GOTM relative to other models in predicting ice-off, Table S2). As a result, the ensemble mean for ice-off during calibration performed worse than individual model performance of FLake, Simstrat, and MyLake. We also found differential model behavior within our projections which affected the overall projection mean. For example, GLM's bottom temperature projections showed little to no change under all RCP scenarios, while all other lake models showed an increase (Fig 6, S8B, S10B, S12B). As a result, GLM skewed the total ensemble mean to be colder for all RCP scenarios (Fig 5). Overall, in most cases the ensemble mean was the best performer, but it is important to assess all of the individual models' performance when interpreting ensemble mean projections, and potentially weight ensemble means based on historical performance (e.g., Raftery et al. 2005).

RCP scenario intensity impacts uncertainty dynamics. Our results suggest that changes in proportional uncertainty over time may be directly related to the magnitude of the RCP scenario, especially for bottom temperature, Schmidt stability, and stratification duration. Specifically, we found that for RCP 2.6, lake model selection uncertainty increased until mid-century, but then decreased towards end-century for bottom temperature, Schmidt stability, and total stratification duration (Fig. S6). In contrast, lake model selection uncertainty in RCP 6.0 exhibited a continuous increase throughout the projection period for the same metrics (Fig. S7). Further, RCP 8.5 exhibited a faster, non-linear increase in lake model selection uncertainty compared to RCP 6.0, which had a higher proportion of overall uncertainty due to lake model selection than RCP 2.6 or RCP 6.0. A possible explanation is that as the magnitude and variability of meteorological drivers changes from current and historical conditions (e.g., increased air temperatures or more variable rainfall) used to calibrate each model, uncertainty around their projected values increases. Altogether, lower atmospheric forcing values were associated with lower lake model uncertainty and higher climate model uncertainty. These findings remain largely unexplored and require further research across multiple RCP uncertainty analyses.

Future projections of Lake Sunapee's thermal structure. Our projections of Lake Sunapee generally align with (or show slightly greater warming than) global lake projection studies, which also examine thermal metrics across the next century. Stratification projections are similar, with our study predicting 10-40 more days of summer stratification duration by end-century (2099), compared to 10-35 days of more stratified conditions on average projected for lakes globally (Woolway et al., 2021; Woolway, Sharma & Smol, 2022). Similarly, our summer mean surface water temperature change projected for Lake Sunapee was higher (range of 1-5ºC) compared to the global surface temperature projection of 1-4ºC by end-century (Woolway, Sharma & Smol, 2022). Changes in lake bottom temperatures show less consistency across studies, with some lake bottom waters found to be warming and others cooling (Pilla et al., 2020). Based on our projections, Lake Sunapee's bottom waters are likely to warm by 1-2ºC during summer months. Lake Sunapee is projected to lose more ice than the global average (10-40 day loss), with a 20-75 day loss of winter ice cover by end-century (Woolway, Sharma & Smol, 2022). However, due to Lake Sunapee's historical pattern of ice duration, which lasts from December or January until March or April (LSPA and Town of Sunapee, 2022), Lake Sunapee is unlikely to be one of ~5,700 lakes at risk of completely losing ice cover this century (Sharma et al., 2021b).

Interestingly, our projections found no definitive change in thermocline depth under any RCP scenario. Across other lake thermal projection studies, thermocline depth also showed a variable response to climate change scenarios, with some studies reporting up to a 0.49 m shallower thermocline under RCP 6.0 (Ayala, Moras & Pierson, 2020), while others report a slight deepening of thermocline depth in response to multiple RCP scenarios (Prats et al., 2018; Barbosa et al., 2021), indicating uncertainty in the directionality of this metric. These inconsistent thermocline depth responses to climate change motivate the need for future studies which examine changes in this important thermal metric.

The changes in the thermal dynamics of Lake Sunapee, as documented by our projections, could lead to fundamental changes in other physical, chemical and biological processes in the lake. For example, stronger lake stratification can lead to increased hypolimnetic anoxia as a result of reduced mixing (Jankowski et al., 2006; Piccolroaz, Toffolon & Majone, 2015), leading to increases in greenhouse gas emissions and fluxes of nutrients and carbon from the sediments into the water column (Hounshell et al., 2021; Bartosiewicz et al., 2021; Carey et al., 2022a). Thermal habitat availability for a range of organisms (e.g., coldwater fishes) is likely to decrease under a warmer thermal regime (Hansen et al., 2017; Stetler et al., 2020; Kraemer et al., 2021), especially in the northern hemisphere (Comte & Olden, 2017). Warmer temperatures and changes in ice cover may promote the dominance of cyanobacteria in the phytoplankton community (Wagner & Adrian, 2009; Markensten, Moore & Persson, 2010; Elliott, 2012; Janse et al., 2015), leading to deleterious consequences for drinking water and fisheries (Paerl & Paul, 2012; Rigosi et al., 2014).

Study challenges and opportunities for future work. Our study provides several opportunities to build upon in future research. First, while this study focuses on model selection uncertainty through the use of ensemble modeling, numerous other sources of uncertainty may be important to lake thermal projections. In particular, uncertainty within a lake or climate model, potentially due to parameters, initial conditions, or specific processes, was not explored in this study (Pike et al., 2013; Huang et al., 2013; Kim et al., 2014; Page et al., 2018; Thomas et al., 2020; Carey et al., 2022b), and may be important to lake thermal projections. For example, because lake projections involve coupled process-based models (here, GCMs and lake models), each with many parameters, it is possible that parameter uncertainty may be an important contributor to total uncertainty (Ficklin & Barnhart, 2014), but this remains unexamined in lake projections, to our knowledge. In addition, focusing on summer temperatures in our study could have magnified the projected changes, as the highest increase in temperatures in some regions has been found to occur from May to August (Czernecki & Ptak, 2018). This summer focus also precluded us from examining uncertainty dynamics during the mixed season (e.g., October to May), which have yielded different results from the summer stratified period (Thomas et al. 2020). Altogether, important next steps building upon this study include examining uncertainty dynamics for additional water quality variables, seasons, and other sources of uncertainty not included in our study.

Conclusions. Overall, our study demonstrates that uncertainty in lake thermal projections varies across depth, time, and RCP scenario for a north temperate, dimictic lake. We found that the dominant source of uncertainty varied among thermal metrics, with metrics that are more sensitive to atmospheric influence (e.g., surface temperatures) dominated by differences among climate models, whereas within-lake metrics (e.g., stratification duration) dominated by differences among lake models. However, the dominant source of uncertainty also varied over time within the projection period, with lake model selection uncertainty increasing more than climate model selection uncertainty. Additionally, uncertainty contributions appear to be different between short-term and long-term projections, calling for an improved understanding of uncertainty propagation over longer time scales and across metrics. Altogether, there was agreement in the overarching changes Lake Sunapee is likely to experience by the end of the century: Lake Sunapee's surface water will likely warm by 1-5°C and will lead to 10-40 more days of summer stratification, as well as 20-75 fewer days of ice coverage annually, substantially changing the annual thermal structure of the lake. Our findings regarding Lake Sunapee's potential future highlight the importance of using uncertainty partitioning to improve our confidence in future climate projections of lake thermal dynamics.

Acknowledgments. We gratefully acknowledge the Lake Sunapee Protective Association staff and members, especially June Fichter, Geoffrey Lizotte, John Merriman, Robert Wood, and Teriko MacConnell for sensor installation, data access, and for their long-term contributions to monitoring on Lake Sunapee, laying the framework to make this study possible. We thank Bethel Steele and Nicole Ward for data collation and modeling support on Lake Sunapee. We thank the Carey Lab for their helpful input on analyses and results during our study, as well as the Advanced Research Computing at Virginia Tech for providing computational resources.

Funding support. Funding was provided by the National Science Foundation (DBI-1933016, DBI-1933102), the Virginia Tech College of Science Luther and Alice Hamlett Scholarship, as well as the Lake Sunapee Protective Association (LSPA) and the LSPA-VT Calhoun Fellowship program.

Statement of contribution. All authors conceptualized the study. Data were curated by JHW, WMW, CCC, and KCW. JHW and TNM developed the codebase, with support from RQT in developing the uncertainty analysis framework. All projection and data analyses were carried out by JHW, with substantial support from TNM and WMW. Manuscript writing was led by JHW, with significant help and guidance from WMW and CCC. KCW enabled access to long-term Lake Sunapee data and broader framing of the study. Project management was led by WMW, CCC, and JHW. All authors edited and approved the manuscript.

References

Ayala AI, Moras S, Pierson DC. 2020. Simulations of future changes in thermal structure of Lake Erken: proof of concept for ISIMIP2b lake sector local simulation strategy. *Hydrology and Earth System Sciences* 24:3311–3330. DOI: 10.5194/hess-24-3311-2020.

Barbosa CC, Calijuri M do C, dos Santos ACA, Ladwig R, de Oliveira LFA, Buarque ACS. 2021. Future projections of water level and thermal regime changes of a multipurpose subtropical reservoir (Sao Paulo, Brazil). *Science of The Total Environment* 770:144741. DOI: 10.1016/j.scitotenv.2020.144741.

Bartosiewicz M, Maranger R, Przytulska A, Laurion I. 2021. Effects of phytoplankton blooms on fluxes and emissions of greenhouse gases in a eutrophic lake. *Water Research* 196. DOI: 10.1016/j.watres.2021.116985.

Beger A, Dorff CL, Ward MD. 2014. Ensemble forecasting of irregular leadership change. *Research & Politics* 1:2053168014557511. DOI: 10.1177/2053168014557511.

Brentrup JA, Richardson DC, Carey CC, Ward NK, Bruesewitz DA, Weathers KC. 2021. Under-ice respiration rates shift the annual carbon cycle in the mixed layer of an oligotrophic lake from autotrophy to heterotrophy. *Inland Waters* 11:114–123. DOI: 10.1080/20442041.2020.1805261.

Brookes JD, Carey CC, Hamilton DP, Ho L, van der Linden L, Renner R, Rigosi A. 2014. Emerging Challenges for the Drinking Water Industry. *Environmental Science & Technology* 48:2099–2101. DOI: 10.1021/es405606t.

Bruesewitz DA, Carey CC, Richardson DC, Weathers KC. 2015. Under-ice thermal stratification dynamics of a large, deep lake revealed by high-frequency data. *Limnology and Oceanography* 60:347–359. DOI: 10.1002/lno.10014.

Butcher JB, Nover D, Johnson TE, Clark CM. 2015. Sensitivity of lake thermal and mixing dynamics to climate change. *Climatic Change* 129:295–305. DOI: 10.1007/s10584-015-1326-1.

Carey CC, Cottingham KL, Weathers KC, Brentrup JA, Ruppertsberger NM, Ewing H, Hairston NG. 2014. Experimental blooms of the cyanobacterium Gloeotrichia echinulata increase phytoplankton biomass, richness and diversity in an oligotrophic lake. *Journal of Plankton Research* 36:364–377. DOI: 10.1093/plankt/fbt105.

Carey CC, Hanson PC, Thomas RQ, Gerling AB, Hounshell AG, Lewis ASL, Lofton ME, McClure RP, Wander HL, Woelmer WM, Niederlehner BR, Schreiber ME. 2022a. Anoxia decreases the magnitude of the carbon, nitrogen, and phosphorus sink in freshwaters. *Global Change Biology* 28:4861–4881. DOI: 10.1111/gcb.16228.

Carey CC, Woelmer WM, Lofton ME, Figueiredo RJ, Bookout BJ, Corrigan RS, Daneshmand V, Hounshell AG, Howard DW, Lewis ASL, McClure RP, Wander HL, Ward NK, Thomas RQ. 2022b. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters* 12:107–120. DOI: 10.1080/20442041.2020.1816421.

Cobourn KM, Carey CC, Boyle KJ, Duffy C, Dugan HA, Farrell KJ, Fitchett L, Hanson PC, Hart JA, Henson VR, Hetherington AL, Kemanian AR, Rudstam LG, Shu L, Soranno PA, Sorice MG, Stachelek J, Ward NK, Weathers KC, Weng W, Zhang Y. 2018. From concept to practice to policy: modeling coupled natural and human systems in lake catchments. *Ecosphere* 9. DOI: 10.1002/ecs2.2209.

Collins WJ, Bellouin N, Doutriaux-Boucher M, Gedney N, Halloran P, Hinton T, Hughes J, Jones CD, Joshi M, Liddicoat S, Martin G, O'Connor F, Rae J, Senior C, Sitch S, Totterdell I, Wiltshire A, Woodward S. 2011. Development and evaluation of an Earth-System model – HadGEM2. *Geoscientific Model Development* 4:1051–1075. DOI: 10.5194/gmd-4-1051-2011.

Comte L, Olden JD. 2017. Climatic vulnerability of the world's freshwater and marine fishes. *Nature Climate Change* 7:718–722. DOI: 10.1038/nclimate3382.

Czernecki B, Ptak M. 2018. The impact of global warming on lake surface water temperature in Poland - the application of empirical-statistical downscaling, 1971-2100. *Journal of Limnology* DOI: 10.4081/jlimnol.2018.1707.

Dietze M. 2017. *Ecological Forecasting.* Princeton University Press.

Dufresne J-L, Foujols M-A, Denvil S, Caubel A, Marti O, Aumont O, Balkanski Y, Bekki S, Bellenger H, Benshila R, Bony S, Bopp L, Braconnot P, Brockmann P, Cadule P, Cheruy F, Codron F, Cozic A, Cugnet D, Noblet N de, Duvel J-P, Ethé C, Fairhead L, Fichefet T, Flavoni S, Friedlingstein P, Grandpeix J-Y, Guez L, Guilyardi E, Hauglustaine D, Hourdin F, Idelkadi A, Ghattas J, Joussaume S, Kageyama M, Krinner G, Labetoulle S, Lahellec A, Lefebvre M-P, Lefevre F, Levy C, Li ZX, Lloyd J, Lott F, Madec G, Mancip M, Marchand M, Masson S, Meurdesoif Y, Mignot J, Musat I, Parouty S, Polcher J, Rio C, Schulz M, Swingedouw D, Szopa S, Talandier C, Terray P, Viovy N, Vuichard N. 2013. Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. *Climate Dynamics* 40:2123–2165. DOI: 10.1007/s00382-012-1636-1.

Dunne JP, John JG, Adcroft AJ, Griffies SM, Hallberg RW, Shevliakova E, Stouffer RJ, Cooke W, Dunne KA, Harrison MJ, Krasting JP, Malyshev SL, Milly PCD, Phillipps PJ, Sentman LT, Samuels BL, Spelman MJ, Winton M, Wittenberg AT, Zadeh N. 2012. GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics. *Journal of Climate* 25:6646–6665. DOI: 10.1175/JCLI-D-11-00560.1.

Elliott JA. 2012. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. *Water Research* 46:1364–1371. DOI: 10.1016/j.watres.2011.12.018.

Feldbauer J, Ladwig R, Mesman JP, Moore TN, Zündorf H, Berendonk TU, Petzoldt T. 2022. Ensemble of models shows coherent response of a reservoir's stratification and ice cover to climate warming. *Aquatic Sciences* 84:50. DOI: 10.1007/s00027-022-00883-2.

Ficklin DL, Barnhart BL. 2014. SWAT hydrologic model parameter uncertainty and its implications for hydroclimatic projections in snowmelt-dependent watersheds. *Journal of Hydrology* 519:2081–2090. DOI: 10.1016/j.jhydrol.2014.09.082.

Fischer AM, Weigel AP, Buser CM, Knutti R, Künsch HR, Liniger MA, Schär C, Appenzeller C. 2012. Climate change projections for Switzerland based on a Bayesian multi-model approach: CLIMATE CHANGE PROJECTIONS FOR SWITZERLAND. *International Journal of Climatology* 32:2348–2371. DOI: 10.1002/joc.3396.

Flaim G, Andreis D, Piccolroaz S, Obertegger U. 2020. Ice Cover and Extreme Events Determine Dissolved Oxygen in a Placid Mountain Lake. *Water Resources Research* 56:e2020WR027321. DOI: 10.1029/2020WR027321.

Frieler K, Lange S, Piontek F, Reyer CPO, Schewe J, Warszawski L, Zhao F, Chini L, Denvil S, Emanuel K, Geiger T, Halladay K, Hurtt G, Mengel M, Murakami D, Ostberg S, Popp A, Riva R, Stevanovic M, SuzGBRi T, Volkholz J, Burke E, Ciais P, Ebi K, Eddy TD, Elliott J, Galbraith E, Gosling SN, Hattermann F, Hickler T, Hinkel J, Hof C, Huber V, Jägermeyr J, Krysanova V, Marcé R, Müller Schmied H, Mouratiadou I, Pierson D, Tittensor DP, Vautard R, Van Vliet M, Biber MF, Betts RA, Leon Bodirsky B, Deryng D, Frolking S, Jones CD, Lotze HK, Lotze-Campen H, Sahajpal R, Thonicke K, Tian H, Yamagata Y. 2017. Assessing the impacts of 1.5ĝ€°C global warming - Simulation protocol of the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP2b). *Geoscientific Model Development* 10:4321–4345. DOI: 10.5194/gmd-10-4321-2017.

Fujino J, Nair R, Kainuma M, Masui T, Matsuoka Y. 2006. Multi-gas Mitigation Analysis on Stabilization Scenarios Using Aim Global Model. *The Energy Journal* SI2006. DOI: 10.5547/ISSN0195-6574-EJ-VolSI2006-NoSI3-17.

Golub M, Thiery W, Marcé R, Pierson D, Vanderkelen I, Mercado D, Woolway RI, Grant L, Jennings E, Schewe J, Zhao F, Frieler K, Mengel M, Bogomolov VY, Bouffard D, Couture R-M, Debolskiy AV, Droppers B, Gal G, Guo M, Janssen ABG, Kirillin G, Ladwig R, Magee M, Moore T, Perroud M, Piccolroaz S, Raaman Vinnaa L, Schmid M, Shatwell T, Stepanenko VM, Tan Z, 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 ID, Laas A, Mackay EB, Mammarella I, Markensten H, McBride C, Özkundakci D, Potes M, Rinke K, Robertson D, Rusak J, Salgado R, van den Linden L, Verburg P, Wain D, Ward NK, Wollrab S, Zdorovennova G. 2022. A framework for ensemble modelling of climate change impacts on lakes worldwide: the ISIMIP Lake Sector. *Geoscientific Model Development Discussions*:1–57. DOI: 10.5194/gmd-2021-433.

Hampton SE, Galloway AWE, Powers SM, Ozersky T, Woo KH, Batt RD, Labou SG, O'Reilly CM, Sharma S, Lottig NR, Stanley EH, North RL, Stockwell JD, Adrian R, Weyhenmeyer GA, Arvola L, Baulch HM, Bertani I, Bowman LL, Carey CC, Catalan J, Colom-Montero W, Domine LM, Felip M, Granados I, Gries C, Grossart H-P, Haberman J, Haldna M, Hayden B, Higgins SN, Jolley JC, Kahilainen KK, Kaup E, Kehoe MJ, MacIntyre S, Mackay AW, Mariash HL, McKay RM, Nixdorf B, Nõges P, Nõges T, Palmer M, Pierson DC, Post DM, Pruett MJ, Rautio M, Read JS, Roberts SL, Rücker J, Sadro S, Silow EA, Smith DE, Sterner RW, Swann GEA, Timofeyev MA, Toro M, Twiss MR, Vogt RJ, Watson SB, Whiteford EJ, Xenopoulos MA. 2017. Ecology under lake ice. *Ecology Letters* 20:98–111. DOI: 10.1111/ele.12699.

Hansen GJA, Read JS, Hansen JF, Winslow LA. 2017. Projected shifts in fish species dominance in Wisconsin lakes under climate change. *Global Change Biology* 23:1463–1476. DOI: 10.1111/gcb.13462.

Heilman KA, Dietze MC, Arizpe AA, Aragon J, Gray A, Shaw JD, Finley AO, Klesse S, DeRose RJ, Evans MEK. 2022. Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Global Change Biology* 28:2442–2460. DOI: 10.1111/gcb.16038.

Her Y, Yoo S-H, Cho J, Hwang S, Jeong J, Seong C. 2019. Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports* 9:4974. DOI: 10.1038/s41598-019-41334-7.

Hipsey MR, Bruce LC, Boon C, Busch B, Carey CC, Hamilton DP, Hanson PC, Read JS, De Sousa E, Weber M, Winslow LA. 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geoscientific Model Development* 12:473–523. DOI: 10.5194/gmd-12-473-2019.

Ho JC, Michalak AM. 2020. Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. *Limnology and Oceanography* 65:992–1009. DOI: 10.1002/lno.11365.

Hoan NX, Khoi DN, Nhi PTT. 2020. Uncertainty assessment of streamflow projection under the impact of climate change in the Lower Mekong Basin: a case study of the Srepok River Basin, Vietnam. *Water and Environment Journal* 34:131–142. DOI: 10.1111/wej.12447.

Hounshell AG, McClure RP, Lofton ME, Carey CC. 2021. Whole-ecosystem oxygenation experiments reveal substantially greater hypolimnetic methane concentrations in reservoirs during anoxia. *Limnology and Oceanography Letters* 6:33–42. DOI: 10.1002/lol2.10173.

Huang J, Gao J, Liu J, Zhang Y. 2013. State and parameter update of a hydrodynamic-phytoplankton model using ensemble Kalman filter. *Ecological Modelling* 263:81–91. DOI: 10.1016/j.ecolmodel.2013.04.022.

Jankowski T, Livingstone DM, 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. *Limnology and Oceanography* 51:815–819. DOI: 10.4319/lo.2006.51.2.0815.

Janse JH, Kuiper JJ, Weijters MJ, Westerbeek EP, Jeuken MHJL, Bakkenes M, Alkemade R, Mooij WM, Verhoeven JTA. 2015. GLOBIO-Aquatic, a global model of human impact on the biodiversity of inland aquatic ecosystems. *Environmental Science & Policy* 48:99–114. DOI: 10.1016/j.envsci.2014.12.007.

Janssen ABG, Hilt S, Kosten S, de Klein JJM, Paerl HW, Van de Waal DB. 2021. Shifting states, shifting services: Linking regime shifts to changes in ecosystem services of shallow lakes. *Freshwater Biology* 66:1–12. DOI: 10.1111/fwb.13582.

Khan SJ, Deere D, Leusch FDL, Humpage A, Jenkins M, Cunliffe D. 2015. Extreme weather events: Should drinking water quality management systems adapt to changing risk profiles? *Water Research* 85:124–136. DOI: 10.1016/j.watres.2015.08.018.

Kim K, Park M, Min J-H, Ryu I, Kang M-R, Park LJ. 2014. Simulation of algal bloom dynamics in a river with the ensemble Kalman filter. *Journal of Hydrology* 519:2810–2821. DOI: 10.1016/j.jhydrol.2014.09.073.

Kirillin G. 2010. Modeling the impact of global warming on water temperature and seasonal mixing regimes in small temperate lakes. *Boreal Environment Resesarch* 15:279–293.

Kobler UG, Schmid M. 2019. Ensemble modelling of ice cover for a reservoir affected by pumped-storage operation and climate change. *Hydrological Processes* 33:2676–2690. DOI: 10.1002/hyp.13519.

Kraemer BM, Anneville O, Chandra S, Dix M, Kuusisto E, Livingstone DM, Rimmer A, Schladow SG, Silow E, Sitoki LM, Tamatamah R, Vadeboncoeur Y, McIntyre PB. 2015. Morphometry and average temperature affect lake stratification responses to climate change. *Geophysical Research Letters* 42:4981–4988. DOI: 10.1002/2015GL064097.

Kraemer BM, Pilla RM, Woolway RI, Anneville O, Ban S, Colom-Montero W, Devlin SP, Dokulil MT, Gaiser EE, Hambright KD, Hessen DO, Higgins SN, Jöhnk KD, Keller W, Knoll LB, Leavitt PR, Lepori F, Luger MS, Maberly SC, Müller-Navarra DC, Paterson AM, Pierson DC, Richardson DC, Rogora M, Rusak JA, Sadro S, Salmaso N, Schmid M, Silow EA, Sommaruga R, Stelzer JAA, Straile D, Thiery W, Timofeyev MA, Verburg P, Weyhenmeyer GA, Adrian R. 2021. Climate change drives widespread shifts in lake thermal habitat. *Nature Climate Change* 11:521–529. DOI: 10.1038/s41558-021-01060-3.

Lange S. 2017. Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. *Earth System Dynamics* 9:1–30. DOI: 10.5194/esd-2017-81.

Lange S. 2019. EartH2Observe, WFDEI and ERA-Interim data Merged and Bias-corrected for ISIMIP (EWEMBI). DOI: 10.5880/PIK.2019.004.

Lewis Jr. WM. 1983. A Revised Classification of Lakes Based on Mixing. *Canadian Journal of Fisheries and Aquatic Sciences* 40:1779–1787. DOI: 10.1139/f83-207.

Li Q, Bruggeman J, Burchard H, Klingbeil K, Umlauf L, Bolding K. 2021. Integrating CVMix into GOTM (v6.0): a consistent framework for testing, comparing, and applying ocean mixing schemes. *Geoscientific Model Development* 14:4261–4282. DOI: 10.5194/gmd-14-4261-2021.

Livingstone DM, Padisák J. 2007. Large-scale coherence in the response of lake surface-water temperatures to synoptic-scale climate forcing during summer. *Limnology and Oceanography* 52:896–902. DOI: 10.4319/lo.2007.52.2.0896.

LSPA, Steele BG, Weathers KC. 2022a. Lake Sunapee Instrumented Buoy: High Frequency Water Quality Data - 2007-2021. DOI: 10.6073/PASTA/8FC8E53B0AD784437B87E2DD5D9B961A.

LSPA, Steele BG, Weathers KC. 2022b. Lake Sunapee Instrumented Buoy: High-Frequency Weather Data, 2007-2021. DOI: 10.6073/PASTA/DD60713BAA41FE48193A25DD10805BE9.

LSPA and Town of Sunapee. 2022. Ice-off dates for Lake Sunapee, NH, USA, 1869-2022. DOI: 10.6073/PASTA/3C60C873C18C2D4811A084831B3F375A.

Lynch HJ, Naveen R, Trathan PN, Fagan WF. 2012. Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. *Ecology* 93:1367–1377. DOI: 10.1890/11-1588.1.

Markensten H, Moore K, Persson I. 2010. Simulated lake phytoplankton composition shifts toward cyanobacteria dominance in a future warmer climate. *Ecological Applications* 20:752–767. DOI: 10.1890/08-2109.1.

Masson-Delmotte V, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelecki, R. Yu, B. Zhou. IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA* 3–32. DOI: https://dx.doi.org/10.1017/9781009157896.001.

Mckay MD, Beckman RJ, Conover WJ. 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42:55–61. DOI: 10.1080/00401706.2000.10485979.

Mironov D. 2010. Implementation of the lake parameterisation scheme FLake into the numerical weather prediction model COSMO. *Boreal Environment Research* 15:218–230.

Moore TN, Mesman JP, Ladwig R, Feldbauer J, Olsson F, Pilla RM, Shatwell T, Venkiteswaran JJ, Delany AD, Dugan H, Rose KC, Read JS. 2021. LakeEnsemblR: An R package that facilitates ensemble modelling of lakes. *Environmental Modelling and Software* 143:105101. DOI: 10.1016/j.envsoft.2021.105101

O'Reilly CM, Sharma S, Gray DK, Hampton SE, Read JS, Rowley RJ, Schneider P, Lenters JD, McIntyre PB, Kraemer BM, Weyhenmeyer GA, Straile D, Dong B, Adrian R, Allan MG, Anneville O, Arvola L, Austin J, Bailey JL, Baron JS, Brookes JD, De Eyto E, Dokulil MT, Hamilton DP, Havens K, Hetherington AL, Higgins SN, Hook S, Izmest'Eva LR, Joehnk KD, Kangur K, Kasprzak P, Kumagai M, Kuusisto E, Leshkevich G, Livingstone DM, MacIntyre S, May L, Melack JM, Mueller-Navarra DC, Naumenko M, Noges P, Noges T, North RP, Plisnier PD, Rigosi A, Rimmer A, Rogora M, Rudstam LG, Rusak JA, Salmaso

N, Samal NR, Schindler DE, Schladow SG, Schmid M, Schmidt SR, Silow E, Soylu ME, Teubner K, Verburg P, Voutilainen A, Watkinson A, Williamson CE, Zhang G. 2015. Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters* 42:10773–10781. DOI: 10.1002/2015GL066235.

Paerl HW, Paul VJ. 2012. Climate change: Links to global expansion of harmful cyanobacteria. *Water Research* 46:1349–1363. DOI: 10.1016/j.watres.2011.08.002.

Page T, Smith PJ, Beven KJ, Jones ID, Elliott JA, Maberly SC, Mackay EB, De Ville M, Feuchtmayr H. 2018. Adaptive forecasting of phytoplankton communities. *Water Research* 134:74–85. DOI: 10.1016/j.watres.2018.01.046.

Parker WS. 2011. When Climate Models Agree: The Significance of Robust Model Predictions. *Philosophy of Science* 78:579–600. DOI: 10.1086/661566.

Perroud M, Goyette S, Martynov A, Beniston M, Annevillec O. 2009. Simulation of multiannual thermal profiles in deep Lake Geneva: A comparison of one-dimensional lake models. *Limnology and Oceanography* 54:1574–1594. DOI: 10.4319/lo.2009.54.5.1574.

Piccolroaz S, Toffolon M, Majone B. 2013. A simple lumped model to convert air temperature into surface water temperature in lakes. *Hydrology and Earth System Sciences* 17:3323–3338. DOI: 10.5194/hess-17-3323-2013.

Piccolroaz S, Toffolon M, Majone B. 2015. The role of stratification on lakes' thermal response: The case of Lake Superior. *Water Resources Research* 51:7878–7894. DOI: 10.1002/2014WR016555.

Pike A, Danner E, Boughton D, Melton F, Nemani R, Rajagopalan B, Lindley S. 2013. Forecasting river temperatures in real time using a stochastic dynamics approach. *Water Resources Research* 49:5168–5182. DOI: 10.1002/wrcr.20389.

Pilla RM, Williamson CE, Adamovich BV, Adrian R, Anneville O, Chandra S, Colom-Montero W, Devlin SP, Dix MA, Dokulil MT, Gaiser EE, Girdner SF, Hambright KD, Hamilton DP, Havens K, Hessen DO, Higgins SN, Huttula TH, Huuskonen H, Isles PDF, Joehnk KD, Jones ID, Keller WB, Knoll LB, Korhonen J, Kraemer BM, Leavitt PR, Lepori F, Luger MS, Maberly SC, Melack JM, Melles SJ, Müller-Navarra DC, Pierson DC, Pislegina HV, Plisnier P-D, Richardson DC, Rimmer A, Rogora M, Rusak JA, Sadro S, Salmaso N, Saros JE, Saulnier-Talbot É, Schindler DE, Schmid M, Shimaraeva SV, Silow EA, Sitoki LM, Sommaruga R, Straile D, Strock KE, Thiery W, Timofeyev MA, Verburg P, Vinebrooke RD, Weyhenmeyer GA, Zadereev E. 2020. Deeper waters are changing less consistently than surface waters in a global analysis of 102 lakes. *Scientific Reports* 10:20514. DOI: 10.1038/s41598-020-76873-x.

Prats J, Salençon M-J, Gant M, Danis P-A. 2018. Simulation of the hydrodynamic behaviour of a Mediterranean reservoir under different climate change and management scenarios. *Journal of Limnology* 77. DOI: 10.4081/jlimnol.2017.1567.

Raftery, AE, Gneitting T, Balabdaoui F, Polakowski M. 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* 133:1155-1174. DOI: 10.1175/MWR2906.1.

Read JS, Winslow LA, Hansen GJA, Van Den Hoek J, Hanson PC, Bruce LC, Markfort CD. 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecological Modelling* 291:142–150. DOI: 10.1016/j.ecolmodel.2014.07.029.

Riahi K, Rao S, Krey V, Cho C, Chirkov V, Fischer G, Kindermann G, Nakicenovic N, Rafaj P. 2011. RCP 8.5-A scenario of comparatively high greenhouse gas emissions. *Climatic Change* 109:33–57. DOI: 10.1007/s10584-011-0149-y.

Richardson DC, Carey CC, Bruesewitz DA, Weathers KC. 2017. Intra- and inter-annual variability in metabolism in an oligotrophic lake. *Aquatic Sciences* 79:319–333. DOI: 10.1007/s00027-016-0499-7.

Rigosi A, Carey CC, Ibelings BW, Brookes JD. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. *Limnology and Oceanography* 59:99–114. DOI: 10.4319/lo.2014.59.1.0099.

Salonen K, Leppäranta M, Viljanen M, Gulati RD. 2009. Perspectives in winter limnology: Closing the annual cycle of freezing lakes. *Aquatic Ecology* 43:609–616. DOI: 10.1007/s10452-009-9278-z.

Saloranta TM, Andersen T. 2007. MyLake—A multi-year lake simulation model code suitable for uncertainty and sensitivity analysis simulations. *Ecological Modelling* 207:45–60. DOI: 10.1016/j.ecolmodel.2007.03.018.

Scher S, Messori G. 2021. Ensemble Methods for Neural Network-Based Weather Forecasts. *Journal of Advances in Modeling Earth Systems* 13:2020MS002331. DOI: 10.1029/2020MS002331.

Schwalm CR, Glendon S, Duffy PB. 2020. RCP8.5 tracks cumulative CO2 emissions. *Proceedings of the National Academy of Sciences of the United States of America* 117:19656–19657. DOI: 10.1073/PNAS.2007117117.

Sharma S, Blagrave K, Magnuson JJ, O'Reilly CM, Oliver S, Batt RD, Magee MR, Straile D, Weyhenmeyer GA, Winslow L, Woolway RI. 2019. Widespread loss of lake ice around the Northern Hemisphere in a warming world. *Nature Climate Change* 9:227–231. DOI: 10.1038/s41558-018-0393-5.

Sharma N, Dev J, Mangla M, Wadhwa VM, Mohanty SN, Kakkar D. 2021a. A Heterogeneous Ensemble Forecasting Model for Disease Prediction. *New Generation Computing* 39:701–715. DOI: 10.1007/s00354-020-00119-7.

Sharma S, Richardson DC, Woolway RI, Imrit MA, Bouffard D, Blagrave K, Daly J, Filazzola A, Granin N, Korhonen J, Magnuson J, Marszelewski W, Matsuzaki S-IS, Perry W, Robertson DM, Rudstam LG, Weyhenmeyer GA, Yao H. 2021b. Loss of Ice Cover, Shifting Phenology, and More Extreme Events in Northern Hemisphere Lakes. *Journal of Geophysical Research: Biogeosciences* 126:e2021JG006348. DOI: 10.1029/2021JG006348.

Sharma S, Walker SC, Jackson DA. 2008. Empirical modelling of lake water-temperature relationships: a comparison of approaches. *Freshwater Biology* 53:897–911. DOI: 10.1111/j.1365-2427.2008.01943.x.

Steele BG, Weathers KC, Association LSP. 2021. Quality controlled in situ data from multiple locations in Lake Sunapee, NH, USA from the Lake Sunapee Protective Association's Long-term Monitoring Program, 1986-2020. DOI: 10.5281/zenodo.4652076.

Stetler JT, Girdner S, Mack J, Winslow LA, Leach TH, Rose KC. 2020. Atmospheric stilling and warming air temperatures drive long-term changes in lake stratification in a large oligotrophic lake. *Limnology and Oceanography* 9999:1–11. DOI: 10.1002/lno.11654.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* 106:7183–7192. DOI: 10.1029/2000JD900719.

Tebaldi C, Knutti R. 2007. The Use of the Multi-Model Ensemble in Probabilistic Climate Projections. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 365:2053–2075.

Thomas RQ, Figueiredo RJ, Daneshmand V, Bookout BJ, Puckett LK, Carey CC. 2020. A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time. *Water Resources Research* 56. DOI: 10.1029/2019WR026138.

Trolle D, Elliott JA, Mooij WM, Janse JH, Bolding K, Hamilton DP, Jeppesen E. 2014. Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software* 61:371–379. DOI: 10.1016/j.envsoft.2014.01.032.

van Vuuren DP, den Elzen MGJ, Lucas PL, Eickhout B, Strengers BJ, van Ruijven B, Wonink S, van Houdt R. 2007. Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Climatic Change* 81:119–159. DOI: 10.1007/s10584-006-9172-9.

van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK. 2011. The representative concentration pathways: an overview. *Climatic Change* 109:5. DOI: 10.1007/s10584-011-0148-z.

Wada Y, Wisser D, Eisner S, Flörke M, Gerten D, Haddeland I, Hanasaki N, Masaki Y, Portmann FT, Stacke T, Tessler Z, Schewe J. 2013. Multimodel projections and uncertainties of irrigation water demand under climate change. *Geophysical Research Letters* 40:4626–4632. DOI: 10.1002/grl.50686.

Wagner C, Adrian R. 2009. Cyanobacteria dominance: Quantifying the effects of climate change. *Limnology and Oceanography* 54:2460–2468. DOI: 10.4319/lo.2009.54.6_part_2.2460.

Ward NK, Steele BG, Weathers KC, Cottingham KL, Ewing HA, Hanson PC, Carey CC. 2020. Differential Responses of Maximum Versus Median Chlorophyll- *a* to Air Temperature and Nutrient Loads in an Oligotrophic Lake Over 31 Years. *Water Resources Research* 56:e2020WR027296. DOI: 10.1029/2020WR027296.

Watanabe M, Suzuki T, O'ishi R, Komuro Y, Watanabe S, Emori S, Takemura T, Chikira M, Ogura T, Sekiguchi M, Takata K, Yamazaki D, Yokohata T, Nozawa T, Hasumi H, Tatebe H, Kimoto M. 2010. Improved Climate Simulation by MIROC5: Mean States, Variability, and Climate Sensitivity. *Journal of Climate* 23:6312–6335. DOI: 10.1175/2010JCLI3679.1.

Wetzel R. 2001. *Limnology: lake and river ecosystems*. Gulf Professional Publishing.

Weyhenmeyer GA, Westöö AK, Willén E. 2008. Increasingly ice-free winters and their effects on water quality in Sweden's largest lakes. In: *Hydrobiologia*. 111–118. DOI: 10.1007/s10750-007-9188-9.

Woolway RI, Merchant CJ. 2018. Intralake Heterogeneity of Thermal Responses to Climate Change: A Study of Large Northern Hemisphere Lakes. *Journal of Geophysical Research: Atmospheres* 123:3087–3098. DOI: 10.1002/2017JD027661.

Woolway RI, Merchant CJ. 2019. Worldwide alteration of lake mixing regimes in response to climate change. *Nature Geoscience* 12:271–276. DOI: 10.1038/s41561-019-0322-x.

Woolway RI, Sharma S, Smol JP. 2022. Lakes in Hot Water: The Impacts of a Changing Climate on Aquatic Ecosystems. *BioScience* :biac052. DOI: 10.1093/biosci/biac052.

Woolway RI, Sharma S, Weyhenmeyer GA, Debolskiy A, Golub M, Mercado-Bettín D, Perroud M, Stepanenko V, Tan Z, Grant L, Ladwig R, Mesman J, Moore TN, Shatwell T, Vanderkelen I, Austin JA, DeGasperi CL, Dokulil M, La Fuente S, Mackay EB, Schladow SG, Watanabe S, Marcé R, Pierson DC, Thiery W, Jennings E. 2021. Phenological shifts in lake stratification under climate change. *Nature Communications* 12:2318. DOI: 10.1038/s41467-021-22657-4.

Woolway RI, Weyhenmeyer GA, Schmid M, Dokulil MT, de Eyto E, Maberly SC, May L, Merchant CJ. 2019. Substantial increase in minimum lake surface temperatures under climate change. *Climatic Change* 155:81–94. DOI: 10.1007/s10584-019-02465-y.

Wynne JH, Woelmer WM, Moore TN, Thomas RQ, Weathers KC, Carey CC. 2022a. Ensembled projection outputs of Lake Sunapee using multiple General

Circulation models and Lake models. DOI: 10.5281/zenodo.7232735.

Wynne JH, Woelmer WM, Moore T, Thomas RQ, Weathers KC, Carey CC. 2022b. jacob8776/sunapee_LER_projections: Repository for November 2022 submission to PeerJ. DOI: 10.5281/zenodo.7262287.

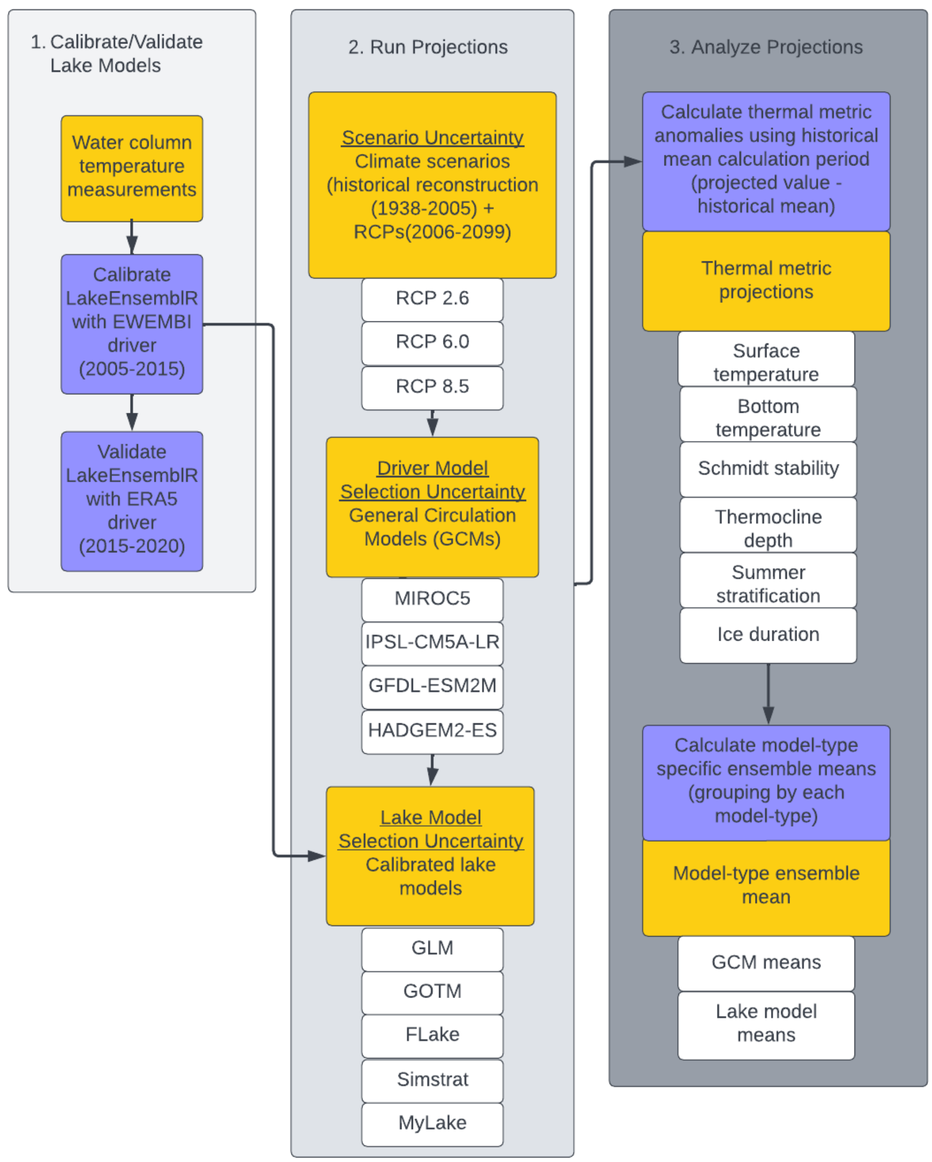


Figure 1: Conceptual workflow representing the methodology used

to produce lake thermal projections. Methodological workflow explaining the full projection process starting from Step 1, using observational data to calibrate and validate lake models; Step 2, using a scaffolded approach to create an ensemble of n = 60 model projections which incorporates climate model selection uncertainty (via 4 GCMs), and lake model selection uncertainty (via 5 lake models); and Step 3, analyze the projection output by first calculating anomalies for each thermal metric using the historical mean calculation period, and then calculating model-type ensemble means (see *Methods and Materials: Ensemble means across RCP scenarios*). Yellow boxes represent outputs or inputs in the workflow, while blue boxes represent actions.

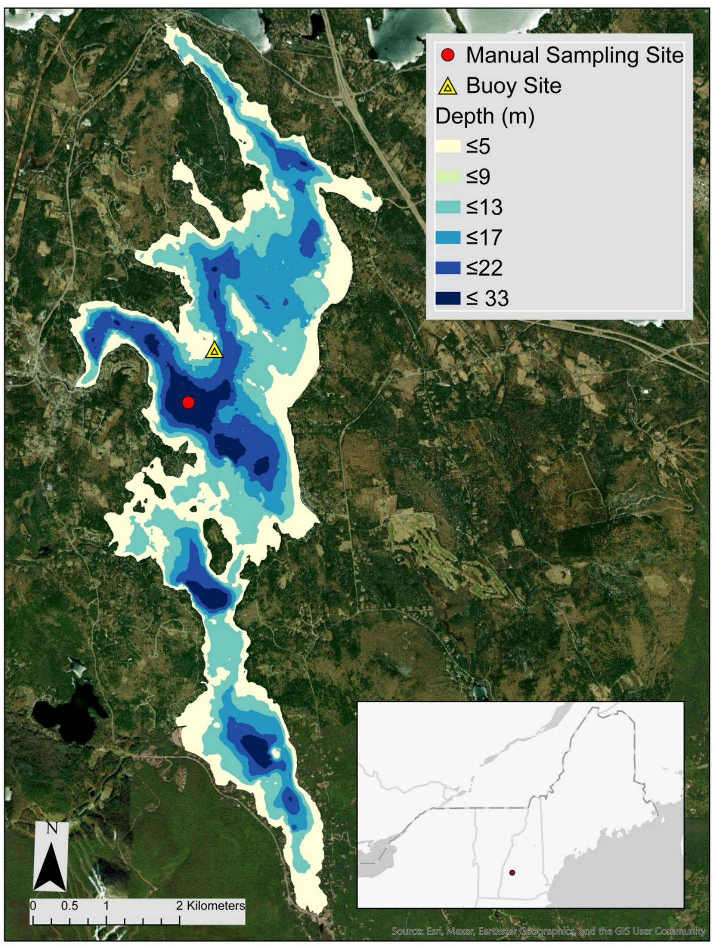


Figure 2. Map of Lake Sunapee, NH. Location and bathymetry of Lake Sunapee, New Hampshire, USA (43.39745° N, 72.05065° W), showing the site of the Lake Sunapee Protective Association GLEON buoy where high-frequency

temperature measurements are collected on m intervals to 10 m depth. The red dot denotes a manual sampling location, where monthly water temperature profiles were measured on m intervals to 30 m depth.

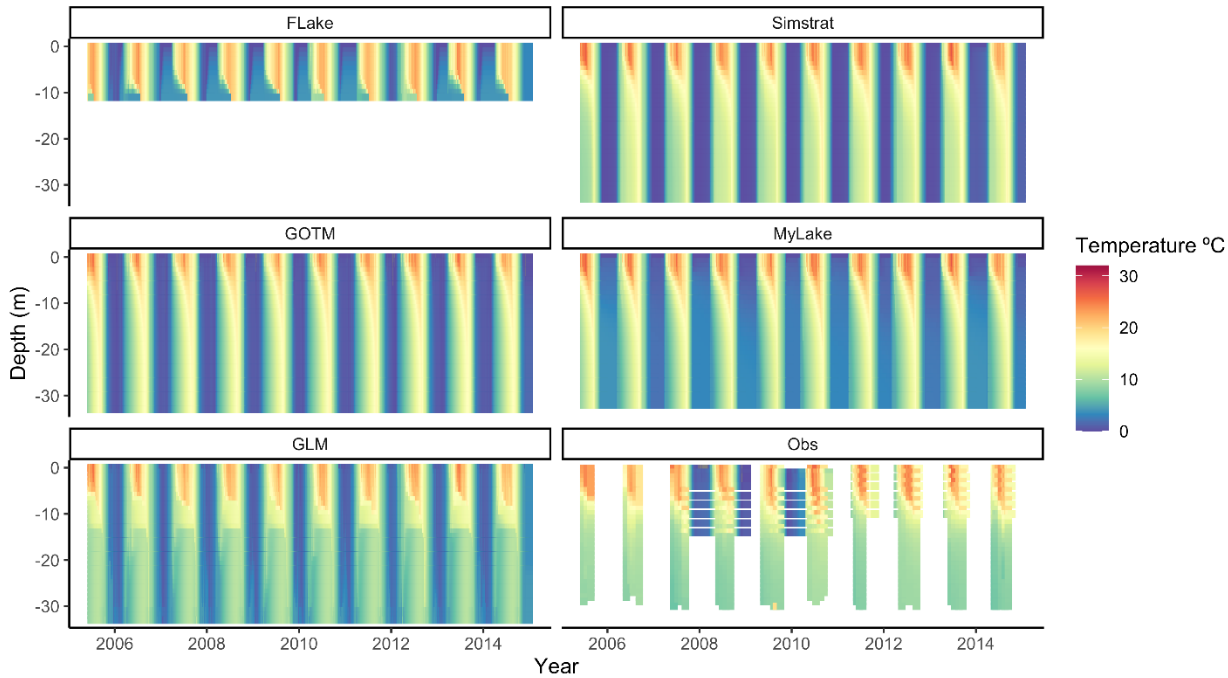


Figure 3. Modeled and observed water column temperatures over the calibration period. Contour plot of whole water column temperature profiles from the surface to the bottom of the lake (with the exception of FLake (surface to 11.0 m) and MyLake (surface to 32.0 m); see Methods: Calibration) during the calibration period (2005-2015) for each lake model (FLake, GLM, GOTM, Simstrat and MyLake). Observed water temperature data (Obs) for Lake Sunapee, NH are shown in the bottom right panel.

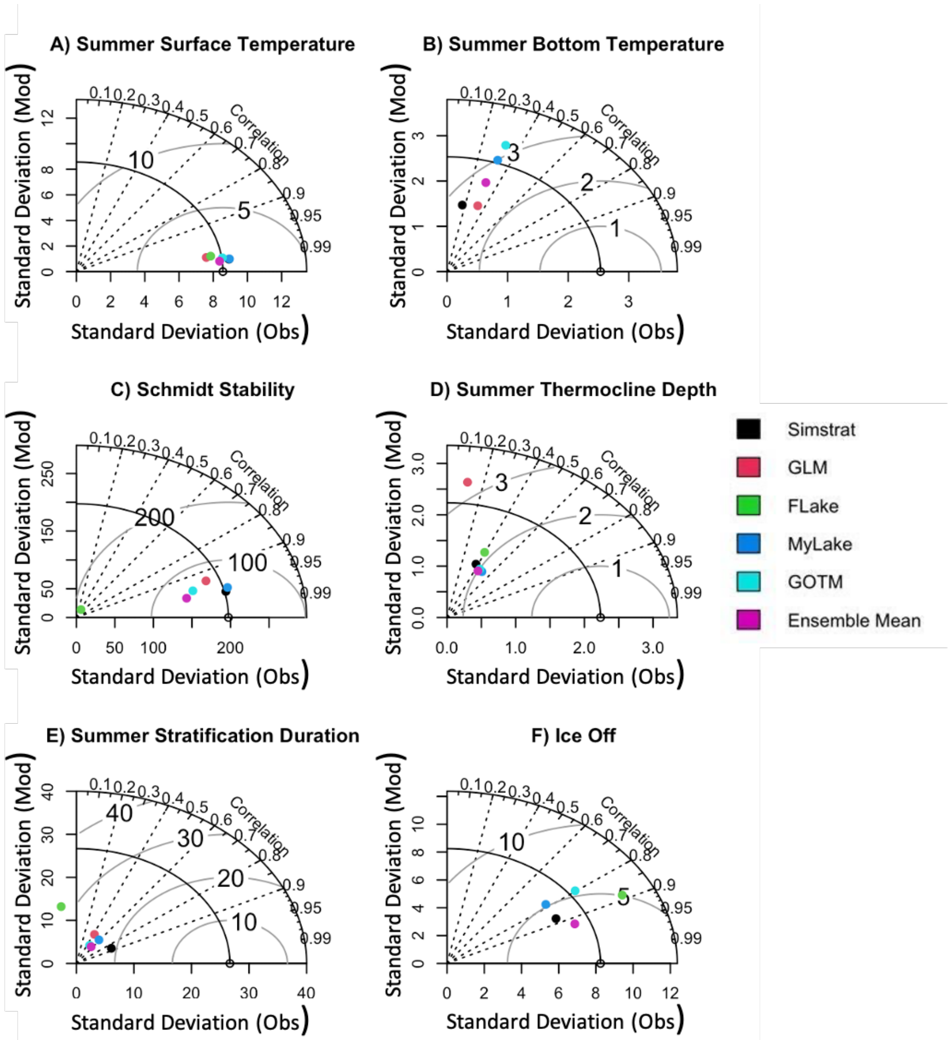


Figure 4. Taylor diagrams showing model performance. Taylor diagrams showing the standard deviation of observations (x-axis) and modeled thermal metrics (y-axis) for Lake Sunapee, NH, USA during 2005-2015. The correlation between observations and modeled output is shown as dotted lines, and root mean squared error (RMSE) between observed and modeled standard deviation is shown as arcs. Correlation (r) represents the linear trend between observed and modeled data. Each colored point represents a specific lake model compared with observed data. Thermal metrics include A) summer surface temperature mean, B) summer bottom temperature mean, C) Schmidt stability, D) summer thermocline depth, E) stratification duration, and F) Date of Ice-off. Summer is defined as June-August. Values to the left of the y-axis (e.g., panel E) represent negative correlation.

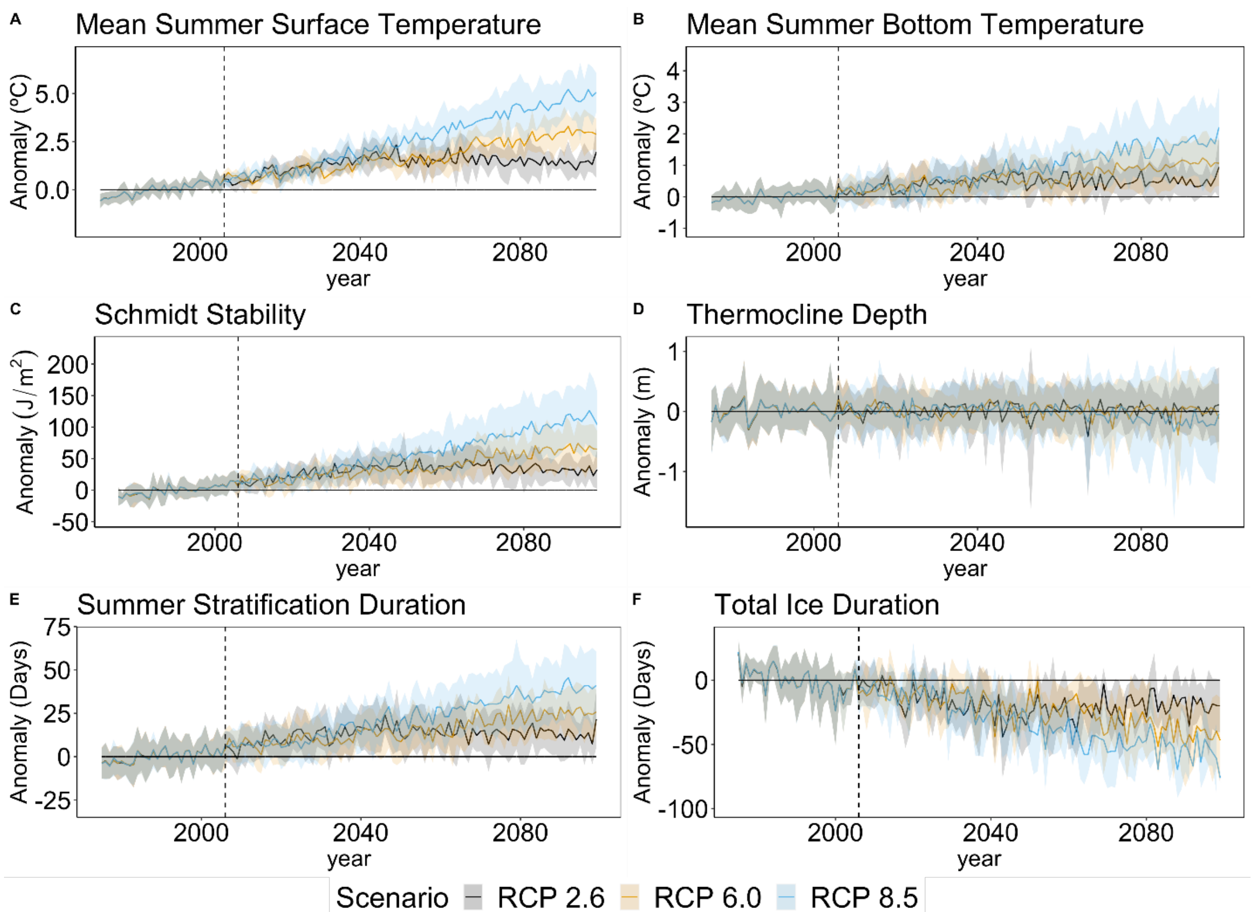
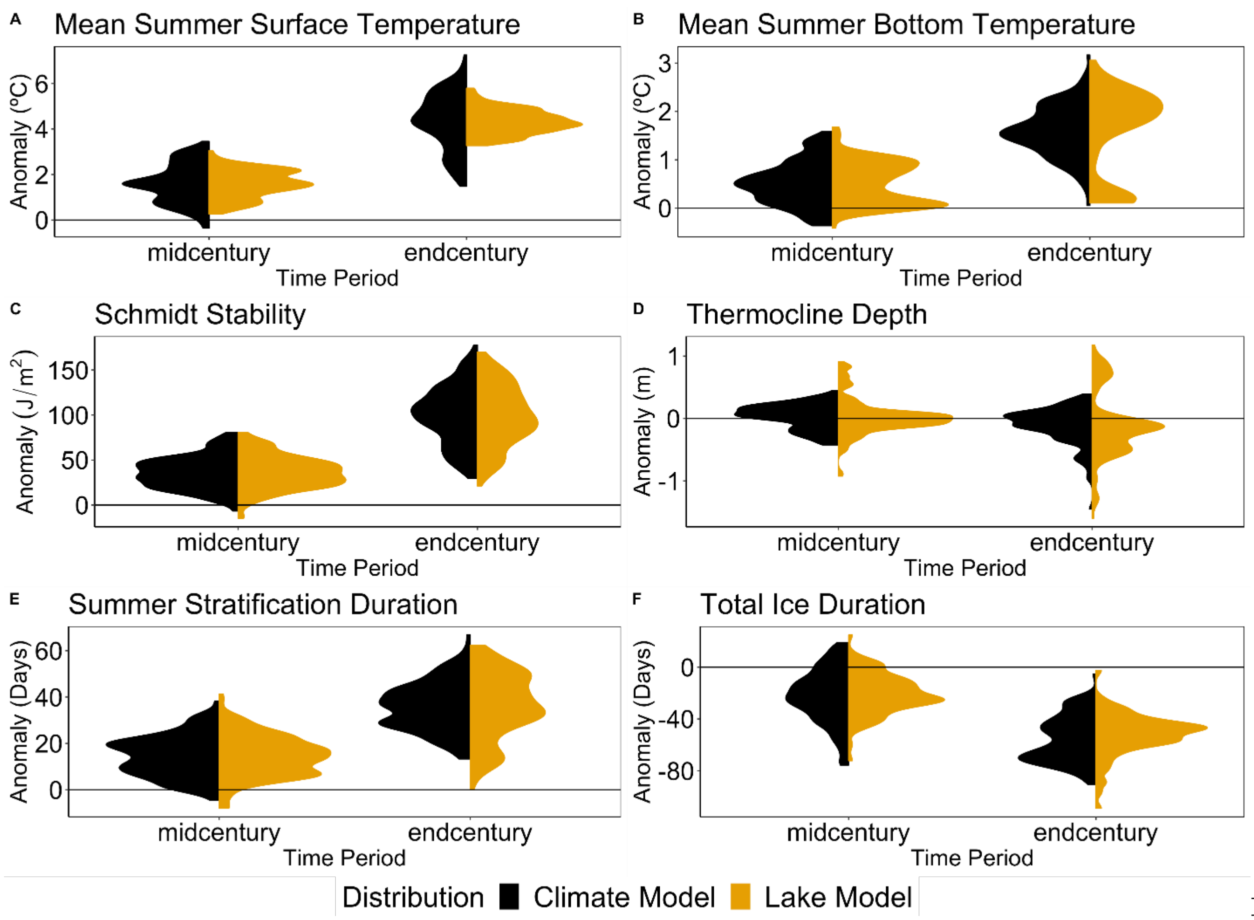
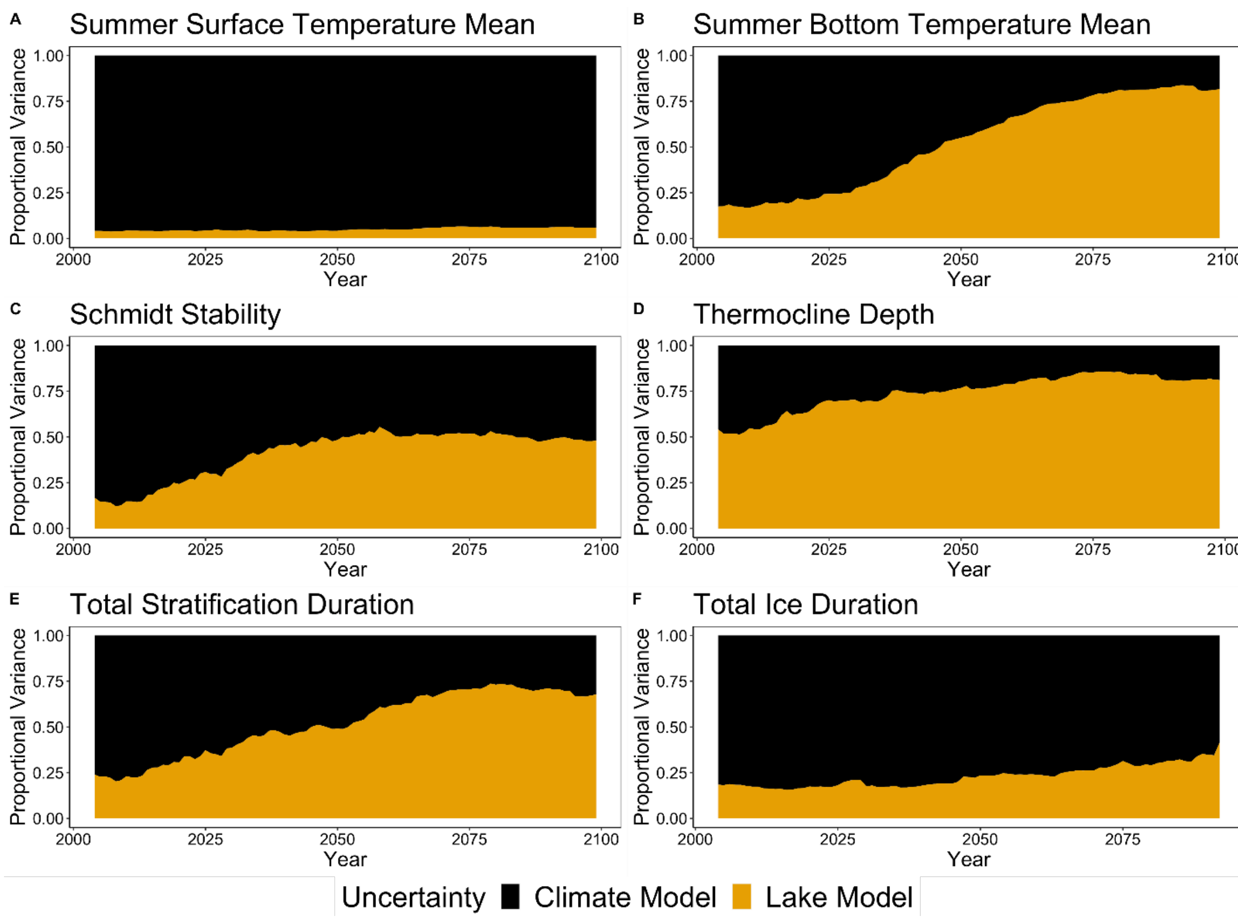


Figure 5. Lake thermal projections through end-century. Projected anomalies for A) mean summer surface temperature (surface temperature), B) mean summer bottom temperature (bottom temperature), C) Schmidt stability, D) thermocline depth, E) summer stratification duration, and F) total ice duration from 2006-2099. The vertical dashed line represents the beginning of the projection time period, with the left of the dashed line representing the historical mean calculation period on which anomalies were based (1975-2005). Each solid line represents the ensemble mean of the lake models under RCP 2.6, 6.0, or 8.5 and each shaded area around the solid lines represents total projection uncertainty under RCP 2.6, 6.0, or 8.5.



Fi

6. Model-type distributions for mid- and end-century across lake model and climate model uncertainty. Distributions of the model-type mean anomalies of the climate models and lake models under RCP 8.5. These distributions were calculated for A) mean summer surface temperature (surface temperature), B) mean summer bottom temperature (bottom temperature), C) Schmidt stability, D) thermocline depth, E) summer stratification duration, and F) total ice duration during mid-century (2020-2050) and end-century (2069-2099).



Fi

7. Proportional variance of the contribution of climate model selection uncertainty (black) and lake model selection uncertainty (orange) to total uncertainty. A) surface temperature mean (surface temperature), B) bottom temperature mean (bottom temperature), C) Schmidt stability, D) thermocline depth, E) total stratification duration, and F) total ice duration from 2006-2099. Proportional variance was calculated for all models using the RCP 8.5 climate scenario.