Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.

R. Quinn Thomas[1], Ryan McClure[1], Tadhg Moore[1], Whitney Woelmer[1], Carl Boettiger[2], Renato Figueiredo[3], Robert Hensley[4], and Cayelan Carey[1]

[1]Virginia Tech
[2]University of California Berkeley
[3]University of Florida
[4]National Ecological Observatory Network

November 30, 2022

Abstract

The National Ecological Observatory Network (NEON)'s standardized monitoring program provides an unprecedented opportunity for comparing the predictability of ecosystems. To harness the power of NEON data for examining environmental predictability, we scaled a near-term, iterative water temperature forecasting system to all six conterminous NEON lakes. We generated 1 to 35-day ahead forecasts using a process-based hydrodynamic model that was updated with observations as they became available. Forecasts were more accurate than a null model up to 35-days ahead among lakes, with an aggregated 1-day ahead RMSE (root-mean square error) of 0.60 and 35-days ahead RMSE of 2.17. Water temperature forecast accuracy was positively associated with lake depth and water clarity, and negatively associated with catchment size and fetch. Our results suggest that lake characteristics interact with weather to control the predictability of thermal structure. Our work provides some of the first probabilistic forecasts of NEON sites and a framework for examining continental-scale predictability.

Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.

R. Quinn Thomas[1,2]*, Ryan P. McClure[2], Tadhg N. Moore[1,2], Whitney M. Woelmer[1], Carl Boettiger[3], Renato J. Figueiredo[4], Robert T. Hensley[5], Cayelan C. Carey[2]

[1]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, USA 24061

[2]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA 24061

[3]Department of Environmental Science, Policy, and Management, University of California-Berkeley, Berkeley, California, USA 94720

[4]Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA 32611

[5]Battelle - National Ecological Observatory Network, Boulder, Colorado, USA 80301

*Corresponding Author: rqthomas@vt.edu

Submitted as a Research Communications to *Frontiers in Ecology and the Environment*

Open Research

All data analyzed in this manuscript are published and publicly available at Thomas et al. 2022a This submission uses novel code, which is provided in Thomas *et al*. (2022b) and Thomas *et al*. (2022c). The analysis is executable as a Binder at https://mybinder.org/v2/zenodo/10.5281/zenodo.6267616/?urlpath=rstudio with Binder instructions available in the Readme file and Web Panel 1.

Thomas RQ, McClure RP, Moore TM, Woelmer WM, Boettiger C, Figueiredo RJ, Hensley RT, and Carey CC. 2022a. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.: code (v1.1). Zenodo repository. https://doi.org/10.5281/zenodo.6674487

Thomas RQ, McClure RP, Moore TM, Woelmer WM, Boettiger C, Figueiredo RJ, Hensley RT, and Carey CC. 2022b. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.: data, forecasts, and scores. Zenodo repository. https://doi.org/10.5281/zenodo.6643596

Thomas RQ, Moore TN, and Daneshmand V. 2022c. Forecasting Lakes and Reservoir Ecosystems R-package (FLAREr): Version 2.2.1 (v2.2.1). Zenodo repository. https://doi.org/10.5281/zenodo.6098517

Abstract

The National Ecological Observatory Network (NEON)'s standardized monitoring program provides an unprecedented opportunity for comparing the predictability of ecosystems. To harness the power of NEON data for examining environmental predictability, we scaled a near-term, iterative water temperature forecasting system to all six conterminous NEON lakes. We generated 1 to 35-day ahead forecasts using a process-based hydrodynamic model that was updated with observations as they became available. Forecasts were more accurate than a null model up to 35-days ahead among lakes, with an aggregated 1-day ahead RMSE (root-mean square error) of 0.60°C and 35-days ahead RMSE of 2.17°C. Water temperature forecast accuracy was positively associated with lake depth and water clarity, and negatively associated with catchment size and fetch. Our results suggest that lake characteristics interact with weather to control the predictability of thermal structure. Our work provides some of the first probabilistic forecasts of NEON sites and a framework for examining continental-scale predictability.

Introduction

A primary goal of the U.S. National Ecological Observatory Network (NEON) is to "understand and forecast continental-scale environmental change" (National Research Council, 2004). With standardized data available across multiple sites, NEON is uniquely positioned to advance the emerging discipline of near-term, iterative environmental forecasting – i.e., the prediction of future environmental conditions and their uncertainty that are updated when observations are available (Dietze *et al.* 2018). However, NEON data have yet to be broadly used for forecasting, a major gap in realizing the potential of the network.

In particular, forecasting the same environmental variables across sites has the potential to reveal gradients of predictability at multiple temporal and spatial scales, a fundamental ecological challenge (Petchey *et al.* 2015; Houlahan *et al.* 2017). While it has been established that forecast accuracy (i.e., realized predictability) declines with horizon (i.e., time into the future), it remains unknown how far into the future different ecological variables can be predicted, and how predictability varies among different sites (Adler *et al.* 2020; Lewis *et al.* 2021). It is likely that both site-level characteristics (e.g., lake depth) and regional-scale characteristics (e.g., weather) affect forecast accuracy at different horizons (Heffernan *et al.* 2014), but the drivers and gradients of predictability remain unknown and may differ among environmental variables.

Lake water temperature is a promising first forecast variable for fulfilling NEON's mission of forecasting environmental change. NEON currently has high-frequency water temperature sensors deployed in six lake sites in the conterminous U.S., providing a range of water temperature dynamics to forecast. Water temperature is a fundamental property of lakes that governs water chemistry, habitat for biota, and other ecological interactions, yet varies substantially throughout a year as a function of lake morphometry, hydrology, ecology, and weather (Wetzel 2001), making it an ideal forecasting case study. Moreover, lake water temperature forecasts have practical benefits, as they could help managers choose which depths to extract water for treatment or preemptively apply interventions to mitigate water quality impairment (Carey *et al.* 2022).

Here, we developed the first known standardized, network-wide forecasts of NEON sites across the U.S. We applied an open-source forecasting system that uses forecasted weather data and a process-based hydrodynamic model to generate future predictions of lake water temperature for 1-35 days ahead. These iterative forecasts were updated with NEON data when they became available. We analyzed the forecasts to address two research questions: 1) How accurately can we predict lake water temperature 1-35 days into the future? and 2) How does forecast accuracy vary among lakes with different site-level characteristics and regional-scale weather?

Methods

Forecasting framework

We developed water temperature forecasts for all six conterminous U.S. NEON lake sites, paired within three NEON-defined ecoclimatic domains (Figure 1). Forecasts were developed using standardized configurations of FLARE (Forecasting Lake And Reservoir Ecosystems), an open-source forecasting system (Thomas *et al.* 2020; Daneshmand *et al.* 2021). The lakes vary in multiple characteristics, including morphometry (depth, volume, surface area, fetch); hydrology (residence time, catchment size); ecology (water clarity); and weather (air temperature, precipitation; Figure 1, see WebTable 1 for lake metadata). FLARE has previously been deployed on a reservoir in Virginia, USA with similar sensor infrastructure to a NEON site but heretofore had not been deployed on other lakes (Thomas *et al.* 2020). FLARE forecasts water temperature at multiple depths in the water column using the General Lake Model (GLM), an open-source hydrodynamic model (Hipsey *et al.* 2019).

FLARE's iterative forecasting cycle is summarized as: 1) each day, the output from the previous day's ensemble forecast (i.e., a set of equally likely simulations of potential future conditions) is used to initialize an ensemble forecast of the current day’s water temperature; 2) FLARE updates the current day’s ensemble forecast and key model parameters to be consistent with the current day's observations using data assimilation; and 3) after updating the forecast, a 1 to 35-day-ahead ensemble forecast of the future is generated, for which no observations are yet available for assimilation. We forecasted water temperature at every 0.25–0.5 m depth interval in each lake, which encompassed all depths with sensors as well as depths without sensors. The forecasts into the future are driven by 35-day-ahead meteorological forecasts from NOAA's Global Ensemble Forecasting System (Li *et al.* 2019). We used NEON's water temperature data (NEON 2022b, c; Hensley 2022) for data assimilation and forecast evaluation (WebPanel 1).

We used the ensemble Kalman filter (EnKF) for data assimilation (Evensen 2009). The EnKF updates model states and parameters based on differences between the ensemble forecast and observations from lake temperature sensors (following Thomas *et al.* 2020). We used this data assimilation approach, rather than directly initiating the forecast with observations, for multiple reasons. First, data assimilation provided initial conditions for forecasting water temperatures at depths without sensor observations. Second, data assimilation provided initial conditions on days when observations were not available. Third, data assimilation generated initial conditions that combined model predictions and observations based on the relative magnitudes of sensor observation and model error. Finally, data assimilation allowed us to dynamically calibrate the model by updating key model parameters.

Altogether, the ensemble forecasts from FLARE represented uncertainty in initial water temperatures when the forecast was initiated (whereby each ensemble member had a different starting temperature profile set by data assimilation), future meteorology (by associating each ensemble member with a different future weather trajectory from NOAA GEFS), a select set of GLM parameters (whereby each ensemble member was associated with different parameter values set by data assimilation), and GLM model equations (whereby normally-distributed error representing model process uncertainty was added to each ensemble member at each time step; Thomas *et al.* 2020).

Our application of FLARE for each lake was initiated on 18 April 2021, the first date when all six lakes had consistent data availability after ice-off. Water temperature data were assimilated but no forecasts were generated from 18 April–18 May 2021, a spin-up period for initial parameter tuning. Other than this one-month spin-up period, we performed no model calibration, with all lakes sharing the same initial parameters at the beginning of the spin-up period. Beginning on 18 May 2021, 35 day-ahead forecasts were produced every day for each lake through 22 October 2021, when data availability ended at the Northern Plains lakes for the year. During May-October, data were assimilated and the forecast initial conditions and parameters were updated each day with observations. Data assimilation resulted in a temporally dynamic calibration of the GLM model for each lake. This iterative forecasting cycle resulted in 159 unique 35-day forecasts , each with 200 ensemble members, for each of the six lakes. Our results below focus on the top 1 m (hereafter, surface).

Evaluation of forecasts

We evaluated forecast performance for each day in the 1–35 day horizon using root-mean square error (RMSE) of the forecasted mean water temperature across ensemble members at each depth and for each horizon (i.e., the 5 day-ahead RMSE included the 5th day of all forecasts at 1 m depth). Furthermore, we quantified: 1) forecast accuracy, defined as RMSE for the first day of the forecast, and 2) accuracy degradation, defined as the difference in maximum and minimum RMSE across the 35-day forecast horizon. We used Spearman rank correlations to quantify the relationships between lake characteristics (morphometry, hydrology, ecology, and weather) and mean forecast accuracy and accuracy degradation for each lake. We used Spearman rank correlations because the sample size was low (n=6 lakes) and many of the variables were non-normally distributed. To ease interpretation of the correlation coefficient, we negated RMSE so positive correlations were associated with higher accuracy. Our RMSE calculations only included dates for a given lake when forecasts were available at all 1–35 day horizons.

Additionally, we compared the forecasts generated using FLARE to null model forecasts that assumed the forecasted mean water temperature for a date and depth was equal to the mean water temperature observed historically on that day of year (DOY). The null model evaluated whether FLARE had higher forecast accuracy than a simple historical mean. The DOY null model was based on all historical NEON data available for a lake (WebTable 1).

Results

Overall, aggregated across the forecasting period, the forecasts were able to accurately predict surface water temperature within 2.60°C RMSE (root-mean square error) 1 to 35 days-ahead for all six lakes (Figure 2a; see WebFigure 1 for two example forecasts). The forecasts performed better than a DOY null model at least 35 days-ahead for the Northern Plains domain lakes; at least 30 days-ahead for the Great Lakes domain lakes; and at least 5 days-ahead for the Southeast lakes (Figure 2b). The forecasts for surface water temperature in each lake had similar accuracy when aggregating forecasts across all depths with observations (WebFigure 2).

Forecast accuracy decreased as the forecast horizon increased among all lakes (Figure 2a). At 1 day-ahead, the mean RMSE of all lakes’ forecasts was 0.61°C (range across lakes: 0.41-0.90°C); at 7 days-ahead, the mean RMSE of all lakes' forecasts was 1.21°C (range: 0.68-1.55°C); at 21 days-ahead, the RMSE of all lakes' forecasts was 2.03°C (range: 1.20-2.45°C); and at 35 days-ahead, the RMSE of all lakes' forecasts was 2.17°C (range: 1.14-2.60°C). The decrease in forecast accuracy as the forecast horizon increased was much lower for BARC than the other lakes (Figure 2a). The Southeast and Northern Plains domain lakes exhibited near-linear decreases in forecast accuracy until ~15-20 days-ahead, when the declines in accuracy saturated (Figure 2a). In comparison, the Great Lakes domain lakes exhibited a more constant decrease in accuracy throughout the 35-day horizon.

Differences in water temperature forecast accuracy and accuracy degradation among lakes were associated with multiple lake morphometric, hydrological, ecological, and weather characteristics. Although our inference space is extremely limited with n=6 lakes, we observed that forecast accuracy was positively correlated to maximum depth and water clarity, and negatively correlated to fetch and catchment size (Figure 3, WebTable 2, WebFigure 3). In contrast, accuracy degradation was positively correlated to volume and water clarity, and negatively correlated to mean annual air temperature (Figure 3, WebTable 2, WebFigure 4).

Conclusions

Here, we present the first continental-scale forecasts of lakes uniquely enabled by NEON. We applied the same forecasting framework to six NEON lakes (i.e., the hydrodynamic model was configured identically among lakes, all lakes had the same initial model parameters, each lake received similar amounts of data for assimilation), thus creating a standardized analysis that can shed light on differences in realized predictability (i.e., forecast accuracy) among sites. Overall, our forecasts had high accuracy among lakes, with consistent patterns in degradation of forecast accuracy with horizon. Below, we explore gradients in accuracy observed among lakes, as well as how our study provides a framework for future NEON forecasting efforts.

Among lakes, water temperature forecast accuracy was high overall, with a mean 1-day-ahead RMSE of 0.62°C and 35-day-ahead RMSE of 2.21°C. Data assimilation resulted in high accuracy at shorter horizons, with decreased forecast accuracy at longer horizons likely due to degradation in weather forecast accuracy. Regardless of horizon, we observed an overall high level of accuracy despite using forecasted, not observed, meteorological data as model inputs. Our forecast accuracy compares favorably to other multi-lake modeling studies that used observed meteorology as inputs: for example, Kreakie *et al.* (2021) predicted upper water column temperatures with an RMSE of 1.48°C for lakes across the U.S with a random forest model. Similarly, Read *et al.* (2014) predicted upper water column temperatures with an RMSE of 1.74°C for Wisconsin, USA lakes with a prior version of the GLM model. By comparing our forecasts to these studies and a DOY null, FLARE's use of automated sensors, data assimilation, and iterative forecasting adds substantial predictive power, especially for the northern lakes where the forecasts all beat the null model >27 days ahead.

Environmental drivers of predictability

The correlation analysis suggests potential relationships between forecast accuracy and environmental drivers that informs future research expanding beyond these six NEON lakes (Figure 3). Lake maximum depth, catchment size, fetch, and water clarity exhibited relationships with forecast accuracy. Deeper lakes have stronger thermal stratification and more resistance to wind-driven mixing (Gorham and Boyce 1989), thereby stabilizing their temperatures and increasing their predictability. In contrast, lakes with larger catchments experience greater inflow volumes (Messager *et al.* 2016) and lakes with greater fetch have greater wind-driven mixing (Rueda and Schladow 2009), both potentially resulting in more variable water temperatures and lower predictability. We observed a positive relationship between forecast accuracy and water clarity, as highlighted in the contrast between the two Southeast lakes: BARC had approximately ~10× higher water clarity than SUGG, and much higher forecast accuracy (Figure 2a, WebTable 1). Deeper penetration of solar radiation results in more uniform heating of the surface waters, thereby increasing deep water temperatures and decreasing vertical temperature gradients (Kirillin and Shatwell 2016). Altogether, the higher predictability of water temperature in BARC than SUGG may be due to the interacting drivers of greater depth, smaller fetch, and greater clarity, as well as other factors.

Forecast accuracy degradation was negatively related to mean annual temperature and positively related to water clarity and volume. The colder northern lakes (Northern Plains and Great Lakes domains) exhibited much greater degradation than one of the warmer Southeast lakes (BARC; Fig. 2a), potentially driving the relationship between air temperature and forecast degradation. While the two lakes with the highest water clarity (CRAM and LIRO in the Great Lakes domain) had a greater decline in forecast accuracy over the 35-day horizon than the three lakes with the lowest water clarity (PRLA, PRLO, and SUGG), thus driving the correlation, BARC was an important outlier because it had the highest water clarity yet the lowest decline in forecast accuracy (WebPanel 4). The patterns between degradation and water clarity/volume may have been an artifact of the lakes in the analysis, as the Great Lakes domain lakes had the greatest water clarity and volume and were the only lakes for which forecast accuracy did not saturate with horizon (Figure 2a, WebTable 1). We did not observe strong correlations between forecast accuracy/degradation and the other lake characteristics (Figure 3), though as noted above, our inference space with six lakes was limited. However, this initial analysis helps develop hypotheses on the drivers of lake water temperature predictability that can be tested in future work.

Using FLARE to forecast NEON lakes

Our application of FLARE to the NEON lakes both extends its current application from one reservoir in Virginia (Thomas *et al.* 2020) to six lakes across the USA, as well as increases its maximum forecast horizon from 16 days in the prior application to 35 days. FLARE forecasts of water temperature in the Virginia reservoir have similar accuracy as observed for the lakes in this study (RMSE of 0.52℃ at 1 day-ahead and 1.62℃ at 16 days-ahead at 1-m depth), and similar degradation of water temperature forecast accuracy with horizon (Thomas *et al.* 2020). This study also provides more evidence that FLARE can generate accurate forecasts rapidly, with only 1 month of spin-up following spring sensor deployment at the NEON lakes and initiating the spin-up with default model parameters. Interestingly, this study reveals that water temperature forecast degradation may saturate at longer horizons for some lakes (Figure 2a), which was only made possible by the recently extended duration of the NOAA meteorological forecasts as FLARE inputs.

We note caveats of this work. First, forecast accuracy/degradation is related to the ability of the GLM to simulate water temperature, so using a different model may influence the relationships we observed between the lake characteristics and accuracy/degradation (Figure 3). Second, our DOY null was limited to <4 years of data, depending on site (WebTable 1). As additional data become available, this null will potentially become more accurate, and may outcompete the forecasts at more horizons. Third, we only forecasted one year of water temperature due to the recent deployment of NEON infrastructure in the study lakes. Our findings may change as we forecast water temperature in future years due to interannual variability. As NEON continues monitoring these lakes into the future (National Research Council 2004), we can test the hypotheses generated in this initial analysis. Fourth, the correlation analyses were constrained by low sample size, low variability in characteristics within an ecoclimatic domain (e.g., the Northern Plains lakes are similar along many axes of potential variation), and collinear variation across domains (e.g., the deep lakes and dimictic lakes are only in the Great Lakes domain; WebTable 1), an inherent limitation of the NEON sampling design. Supplementing future NEON cross-lake forecast comparisons with other lakes (e.g., those in the Global Lake Ecological Observatory Network; Weathers *et al.* 2013) would extend key environmental gradients as well as evaluate whether our observed patterns are supported by a larger sample of forecasts. This extension is important as the six conterminous NEON lakes are not representative of the full range of lakes across the U.S, and the addition of larger and deeper lakes with surface inflows would greatly benefit our analysis.

Power and limitations of NEON for cross-lake forecasting

Similar to weather forecasting, which exhibited a large increase in the number of forecasts and prediction accuracy after an increase in data availability from sensors and satellites, improved models, and advanced data assimilation techniques (Bauer *et al.* 2015), we envision that NEON could catalyze a leap in continental-scale environmental forecasting. NEON's standardized measurements, well-documented metadata, and rigorous data QA/QC provide a critical foundation for forecasting. However, we note that data latency currently limits the ability to generate real-time forecasts. An automated near-term, iterative forecasting system benefits from near-real time data availability. Given the 2-week–1.5-month lag in data availability in

NEON's current pipeline, our analysis here was based on hindcasts – i.e., generating forecasts using forecasted drivers to the perspective of the model but for a past date (Jolliffe and Stephenson 2012). Unless NEON's data latency decreases, forecast analyses such as ours are limited to predicting the past.

Our study provides a framework that can be adapted for additional lakes - as well as terrestrial NEON sites - for forecasting a range of environmental variables and exploring the drivers of predictability. Next steps for this work include forecasting water temperature in future years for the NEON lakes, as well as adding in forecasts for additional water quality variables that NEON monitors, such as dissolved oxygen and chlorophyll-*a*. Forecasting additional water quality variables would greatly expand the utility of the FLARE workflow for informing management, as well as using the NEON lakes as a multi-region test-bed for developing forecasting methods that can be applied to other waterbodies. Following Dietze and Lynch (2019), the future is bright for forecasting in ecology, in large part due to observatory networks like NEON.

Acknowledgements

We thank Vahid Daneshmand, Bethel Steele, Kathleen Weathers, and the FLARE CIBR team for helpful insights and research support. Virginia Tech's Advanced Research Computing and Ben Sandbrook provided computational resources and support. This work was supported by NSF grants DEB-1926388, CNS-1737424, DBI-1933016, DBI-1933102, DBI-1942280, and DEB-1926050.

Authorship contribution statement

RQT, CCC, and RJF co-developed the FLARE forecasting framework and co-lead the FLARE project. RPM led the development of NEON data processing and FLARE forecasting workflows with assistance from RQT. RPM calibrated lake models with assistance from CCC. TNM assisted with GLM model setup and FLARE configuration. WMW co-developed the code for generating historical weather forecasts with RQT. CB led the development of the *neonstore* package for downloading NEON data and co-developed the code for forecast scoring with RQT. RTH provided lake metadata and assisted with NEON data interpretation. CCC and RQT drafted the manuscript with feedback from all co-authors. No author has a conflict of interest.

References

Adler PB, White EP, and Cortez MH. 2020. Matching the forecast horizon with the relevant spatial and temporal processes and data sources. *Ecography* **43**: 1729–39.

Bauer P, Thorpe A, and Brunet G. 2015. The quiet revolution of numerical weather prediction. *Nature* **525**: 47–55.

Carey CC, Woelmer WM, Lofton ME, *et al.* 2022. Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters* **12**: 107-120.

Daneshmand V, Breef-Pilz A, Carey CC, *et al.* 2021. Edge-to-cloud virtualized cyberinfrastructure for near real-time water quality forecasting in lakes and reservoirs. In: 2021 IEEE 17th International Conference on eScience (eScience). Innsbruck, Austria: IEEE.

Dietze MC, Fox A, Beck-Johnson LM, *et al.* 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc Natl Acad Sci U S A* **115**: 1424–32.

Dietze M and Lynch H. 2019. Forecasting a bright future for ecology. *Front. Ecol. Environ* **17**: 3.

Evensen G. 2009. Data Assimilation. Berlin, Heidelberg: Springer Berlin Heidelberg.

Gorham E and Boyce FM. 1989. Influence of lake surface area and depth upon thermal stratification and the depth of the summer thermocline. *J Great Lakes Res* **15**: 233–45.

Hensley, RT. 2022. NEON lakes Level 0 multisonde temperature data - 2021 ver 1. Environmental Data Initiative repository. https://doi.org/10.6073/pasta/fbbd2d5f59a8d92c6865d57e7abae379 (Accessed 25 January 2022).

Hipsey MR, Bruce LC, Boon C, *et al.* 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geosci Model Dev* **12**: 473–523.

Houlahan JE, McKinney ST, Anderson TM, and McGill BJ. 2017. The priority of prediction in ecological understanding. *Oikos* **126**: 1–7.

Jolliffe IT and Stephenson DB (Eds). 2012. Forecast verification: a practitioner's guide in atmospheric science. Oxford: Wiley-Blackwell.

Kirillin G and Shatwell T. 2016. Generalized scaling of seasonal thermal stratification in lakes. *Earth Sci Rev* **161**: 179-190.

Kreakie BJ, Shivers SD, Hollister JW, and Milstead WB. 2021. Predictive model of lake photic zone temperature across the conterminous United States. *Front Environ Sci* **9**: 707874.

Lewis ASL, Woelmer WM, Wander HL, *et al.* 2022. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecol Appl*. **32**: e02500

Li W, Guan H, Zhu Y, *et al.* 2019. Prediction skill of the MJO, NAO and PNA in the NCEP FV3-GEFS 35-day experiments. In: Science and Technology Infusion Climate Bulletin. Durham, NC: NOAA's National Weather Service.

Messager ML, Lehner B, Grill G, *et al.* 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nat Commun* **7**: 13603.

NEON. 2022a. Depth profile at specific depths (DP1.20254.001), RELEASE-2022. https://doi.org/10.48443/d2pr-h658. Dataset available at https://data.neonscience.org on (accessed 25 January 2022)

NEON. 2022b. Temperature at specific depth in surface water (DP1.20264.001). Dataset available at https://data.neonscience.org (accessed 25 January 2022)

NEON. 2022c. Temperature at specific depth in surface water, RELEASE-2022 (DP1.20264.001). https://doi.org/10.48443/g7bs-7j57. Dataset available at https://data.neonscience.org (accessed 25 January 2022)

Petchey OL, Pontarp M, Massie TM, *et al.* 2015. The ecological forecast horizon, and examples of its uses and determinants. *Ecol Lett* **18**: 597–611.

R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Read JS, Winslow LA, Hansen GJA, *et al.* 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecol Model* **291**: 142–50.

Rueda F and Schladow G. 2009. Mixing and stratification in lakes of varying horizontal length scales: Scaling arguments and energy partitioning. *Limnol Oceanogr* **54**: 2003–17.

Thomas RQ, Figueiredo RJ, Daneshmand V, *et al.* 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resour Res* **56**: e2019WR026138.

Weathers KC, Hanson PC, Arzberger P, *et al.* 2013. The Global Lake Ecological Observatory Network (GLEON): The evolution of grassroots network science. *Limnol Oceanogr Bull* **22**: 71–3.

Wetzel RG. 2001. Limnology: lake and river ecosystems. San Diego: Academic Press.

Figure captions

Figure 1. Map showing the locations of the six NEON (National Ecological Observatory Network) lakes forecasted in this study. The inset figures show a year of water temperature depth profiles, as measured by automated sensors deployed from a buoy (NEON 2022bc; Hensley 2022) and monthly handheld probe data collection at each lake (NEON 2022a). The automated sensor data were used in the data assimilation and forecast analysis at depths provided in WebTable 1; the handheld probe data were only used in this figure to better characterize the full water temperature profile. The inset table provides each lake's NEON Site ID, lake name, and NEON ecoclimatic domain. Summary statistics of each lake's morphometry, hydrology, ecology, and weather characteristics are available in WebTable 1.

Figure 2. (a) Surface water temperature (top 1 m) forecast accuracy, defined by RMSE (root-mean square error in ^{o}C), for 1 to 35-day ahead (horizon) forecasts at the six NEON lakes. (b) A skill score of the RMSE (in ^{o}C) of the null day-of-year model vs. forecasts generated by the FLARE (Forecasting Lake And Reservoir Ecosystems) system for each lake. Positive values indicate that FLARE forecasts outperformed the null at a given horizon, zero indicates that the forecasts and null performed similarly, and negative values indicate that the null outperformed the forecasts.

Figure 3. Spearman correlations between two metrics defining predictability at the six lakes: forecast accuracy (red points), defined as RMSE at 1-day ahead, and forecast accuracy degradation (blue points), defined as the difference in maximum and minimum RMSE across the 35-day forecast horizon. To ease interpretation of the correlation coefficient, we negated RMSE

so positive correlations are associated with higher accuracy. Given the extremely limited sample size of lakes (n=6), which is too small for reliable p-values for rho, we focused our interpretation on Spearman rho correlations |≥| 0.5 (above the dashed line). WebFigures 3 and 4 show the relationships as scatterplots.

Figures

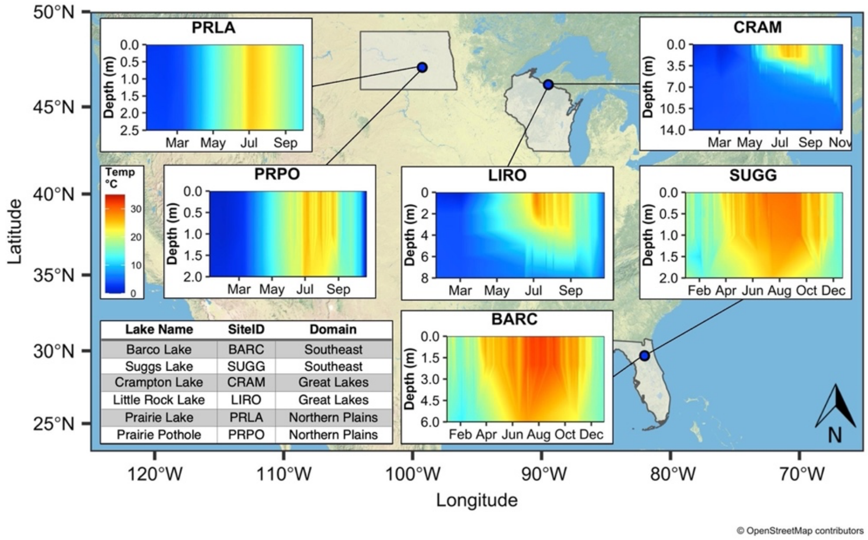


Figure 1. Map showing the locations of the six NEON (National Ecological Observatory Network) lakes forecasted in this study. The inset figures show a year of water temperature depth profiles, as measured by automated sensors deployed from a buoy (NEON 2022bc; Hensley 2022) and monthly handheld probe data collection at each lake (NEON 2022a). The automated sensor data were used in the data assimilation and forecast analysis at depths provided in WebTable 1; the handheld probe data were only used in this figure to better characterize the full water temperature profile. The inset table provides each lake's NEON Site ID, lake name, and NEON ecoclimatic domain. Summary statistics of each lake's morphometry, hydrology, ecology, and weather characteristics are available in WebTable 1.

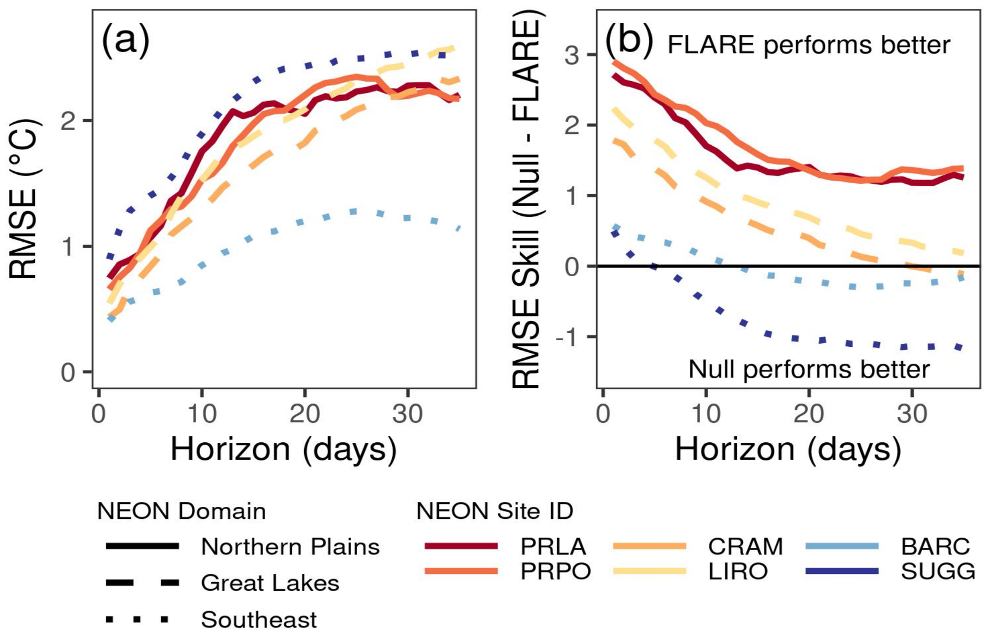


Figure 2. (a) Surface water temperature (top 1 m) forecast accuracy, defined by RMSE (root-mean square error in ºC), for 1 to 35-day ahead (horizon) forecasts at the six NEON lakes. (b) A skill score of the RMSE (in ºC) of the null day-of-year model vs. forecasts generated by the FLARE (Forecasting Lake And Reservoir Ecosystems) system for each lake. Positive values indicate that FLARE forecasts outperformed the null at a given horizon, zero indicates that the forecasts and null performed similarly, and negative values indicate that the null outperformed the forecasts.

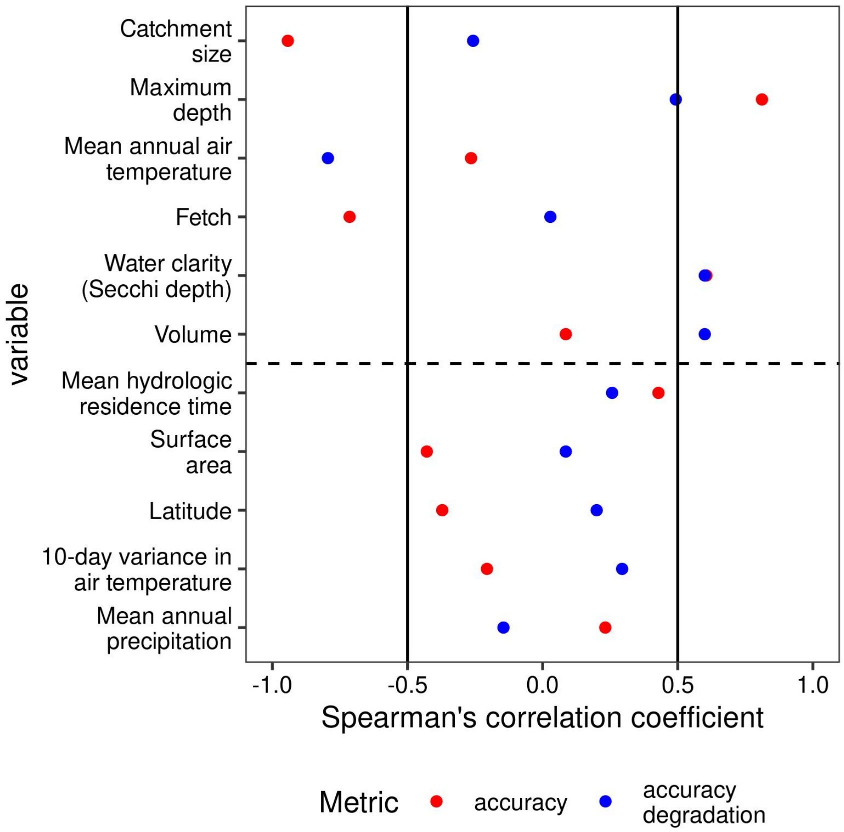


Figure 3. Spearman correlations between two metrics defining predictability at the six lakes: forecast accuracy (red points), defined as RMSE at 1-day ahead, and forecast accuracy degradation (blue points), defined as the difference in maximum and minimum RMSE across the 35-day forecast horizon. To ease interpretation of the correlation coefficient, we negated RMSE so positive correlations are associated with higher accuracy. Given the extremely limited sample size of lakes (n=6), which is too small for reliable p-values for rho, we focused our interpretation on Spearman rho correlations |≥| 0.5 (above the dashed line). WebFigures 3 and 4 show the relationships as scatterplots.

Supplemental Information for “Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.”

R. Quinn Thomas*, Ryan P. McClure, Tadhg N. Moore, Whitney M. Woelmer, Carl Boettiger, Renato J. Figueiredo, Robert T. Hensley, Cayelan C. Carey

*Corresponding author, rqthomas@vt.edu

This supplementary information includes:
WebPanel: 1
WebTables: 2
WebFigures: 4

WebPanel 1. Description of the forecasted NEON lakes, overview of the FLARE configuration for each lake, meteorological driver data, and mean day-of-year null model

Lake and descriptions

We generated forecasts for the six NEON lakes in the conterminous USA (WebTable 1). The six forecast sites were two paired lakes in the Great Lakes NEON ecoclimatic domain (Crampton Lake, NEON site ID – CRAM; Little Rock Lake, NEON site ID - LIRO), two paired lakes in the Northern Plains domain (Prairie Lake, NEON siteID – PRLA; Prairie Pothole, NEON siteID - PRPO), and two paired lakes in the Southeastern domain (Barco Lake, NEON siteID – BARC; Suggs Lake, NEON siteID - SUGG). We excluded the seventh NEON lake site (Toolik Lake) since it was not part of a paired NEON set and it has major surface inflows, unlike the other lakes.

Each lake had 5-10 water temperature sensors (Precision Measurement Engineering Inc. T-Chain RS 232/485 thermistors) deployed at various depths in the water column. The first sensor is deployed 0.05 m below the surface, with remaining depths dependent on the total depth of the lake. Generally, sensors are deployed at more frequent intervals within the upper 1.05 m than at deeper depths. These discrete depth water temperature data are available from NEON (NEON 2022a, b), and were accessed using the *neonstore* R package, which creates a "store" of NEON data on a local computer and eases the iterative downloading of additional NEON data without re-downloading data already within the store (Boettiger *et al.* 2021).

All data were filtered using the quality assurance codes provided by NEON. The 30-minute data product was aggregated to the hour and only the 00:00-01:00 UTC hour was used each day for assimilation and evaluation. The NEON (NEON 2022a, b) data were exported using the *neon_export* function in the *neonstore* R package and archived at Thomas and Boettiger (2022). Gaps in NEON's discrete depth water temperature dataset were filled using water temperature data collected by a YSI EXO2 multiparameter sonde as part of NEON's water quality data product (Hensley 2022).

FLARE and GLM configuration

Adapting FLARE to NEON lakes required configuring six unique GLM models with each lake's bathymetry and physical specifications and developing functions to download and process NEON water temperature data. Across all six lakes, we used the same initial default GLM hydrodynamic parameters (Hipsey *et al.* 2019) and tuned the same set of three parameters governing lake water temperature during data assimilation (lw_factor, kw, and sed_mean_temp). Since none of the six NEON lakes have major surface inflows or outflows and prior applications at a reservoir in Virginia showed limited sensitivity of forecast uncertainty to inflows (Thomas *et al.* 2020), we parameterized each lake without inflows or outflows.

We parameterized the process uncertainty in water temperature to be the same across sites and throughout the water column (standard deviation = 0.75°C). This value was based on the findings of Thomas *et al.* (2020), in which FLARE's process uncertainty was estimated across water column depths at a reservoir in Virginia. The process uncertainty was added to each ensemble member and modeled depth at each daily timestep. Since we expect this uncertainty to be correlated with depth (e.g., if the modeled temperature at a certain depth was 1°C warmer than observed, nearby depths should also likely be too warm as well), we included a correlation length that represents an exponential decay of correlations across depths (following Appendix A in Lenartz *et al.* 2007). The decay in correlation results in stronger correlations in water temperature at closer depths than further away depths. This decorrelation length parameter was set to 2 m.

Similarly, observation uncertainty in water temperature data was set to be the same across lakes and depths (standard deviation = 0.1°C), based on the FLARE application in Thomas *et al.* (2020). Since observation uncertainty represents sensor and sampling uncertainty, we did not expect observation uncertainty to be correlated with depth, and therefore the decorrelation length for this uncertainty source was set to 0 m.

Parameter estimation using the ensemble Kalman filter (EnKF) uses the estimated correlation between parameter values and the size of the errors between the predicted and observed states across ensemble members (Evensen 2009). Ensemble members that require large adjustments in the states to be consistent with observations will also adjust parameters that are correlated with that error. One challenge with estimating parameters using the EnKF is that the variation in parameter values across ensemble members collapses over time. The small variance among ensemble members prevents the parameters from further adjusting to reduce new biases in the model predictions (i.e., the calibration does not change through time).

As a result, parameter estimation methods using the EnKF need to use a technique to prevent a collapse in variance. Here, we use a method called variance inflation, in which the variance in parameter values among the ensemble members is increased at each time-step when data assimilation occurs. The variance inflation increases the spread in the parameters among ensemble members while maintaining the rank order of ensemble members. We used the same variance inflation factor across all parameters and lakes (0.04).

The FLAREr R package that contains FLARE functions can be found in the Zenodo repository (Thomas *et al.* 2022b), as well as the scripts for running FLARE at the six NEON lakes (Thomas *et al.* 2022a). All analyses were conducted in R software version 4.1.1 (R Core Team 2021).

Meteorological inputs

The forecasts were driven by numerical meteorological forecasts produced by NOAA's Global Ensemble Forecasting System (GEFS) version 12 (Li *et al.* 2019). We automated the downloading of ensemble members (n=31 total) from the NOAA GEFS output for each 0.5°×0.5° grid cell that included a NEON lake. NOAA GEFS generates weather forecasts at multiple times per day (00:00, 06:00, 12:00, and 18:00 UTC), which vary in their forecast horizon length (i.e., days into the future). We focused on the GEFS weather forecast that started at 00:00 UTC each day, as 30 of its 31 ensemble members extended 35 days into the future on a 6-hour time step and included all meteorological variables required by the GLM as model driver data. The 6-hour output resolution of each of the 30 ensemble members was temporally downscaled to 1-hour resolution for use in the GLM following Thomas *et al.* (2020).

We used a "stacked" GEFS product during the 1-month spin-up period. One challenge when using data assimilation to set initial conditions and tune parameters is a potential mismatch between the meteorological data used in the spin-up and data used for generating future forecasts. Since observed and forecasted meteorology are rarely a 1:1 match, a smooth transition from data assimilation to forecasting requires either the forecasted meteorology to be corrected for the site or past meteorological forecasts to be used in place of observed meteorology for data assimilation. Here, we used the latter option because NEON meteorological data has a 1.5-month latency and often has gaps for some of the required meteorological variables. To develop a "stacked" GEFS product, we downloaded the first time step of the forecasts that were initiated at

06:00, 12:00, and 18:00 UTC. We then combined the meteorological forecast at the first time step of the 00:00, 06:00, 12:00, and 18:00 UTC forecasts together to generate a 6-hr data product starting on 18 April 2021. The first time step is used because it directly follows data assimilation in the GEFS, and therefore is most closely aligned with observed meteorology. The "stacked" data product is generated each time new GEFS forecasts are available, and thus is near-real time.

To estimate the 10-day variance in air temperature that was used in the predictability correlation analysis, we calculated the running standard deviation over a rolling 10-day window between 18 May 2021 and 31 October 2021 from the "stacked" GEFS product. We used the mean of the 10-day running standard deviation to represent air temperature variance for each lake during the period that forecasts were generated.

All NOAA GEFS 1-hour forecasts and "stacked" products for the six NEON lakes are archived at Thomas and Woelmer (2022).

Mean Day-of-Year Null Forecast

We note that while the 1 to 3.5 years of data at the NEON lakes available for this day-of-year (DOY) null model (see WebTable 1) is lower than the ~30 years of data typically used in weather forecasting null climatology models, it still included all NEON data available for each lake. Moreover, the DOY null model for the lake with just one year of data (PRLA) performed similarly to the DOY null model for its paired lake (PRPO), which had three years of data (Figure 2b).

Analysis

Thomas and Boettiger (2022) and Thomas and Woelmer (2022). This submission uses novel code, which is provided in Thomas *et al.* (2022a) and Thomas *et al.* (2022b).

WebReferences

Boettiger C, Thomas RQ, Laney C, and Lunch C. 2021. neonstore: NEON Data Store. R package. CRAN repository. https://cran.r-project.org/web/packages/neonstore/index.html

Evensen G. 2009. Data Assimilation. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hensley, R.T. 2022. NEON lakes Level 0 multisonde temperature data - 2021 ver 1. Environmental Data Initiative repository. https://doi.org/10.6073/pasta/fbbd2d5f59a8d92c6865d57e7abae379 (Accessed 2022-01-25).

Hipsey MR, Bruce LC, Boon C, *et al.* 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geosci Model Dev* **12**: 473–523.

Lenartz F, Raick C, Soetaert K, and Grégoire M. 2007. Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea. *J Mar Syst* **68**: 327–48.

Li W, Guan H, Zhu Y, *et al.* 2019. Prediction Skill of the MJO, NAO and PNA in the NCEP FV3-GEFS 35-day Experiments. In: Science and Technology Infusion Climate Bulletin. Durham, NC: NOAA's National Weather Service.

NEON. 2022a. Temperature at specific depth in surface water (DP1.20264.001). Dataset available at https://data.neonscience.org (accessed 25 January 2022)

NEON. 2022b. Temperature at specific depth in surface water, RELEASE-2022 (DP1.20264.001). https://doi.org/10.48443/g7bs-7j57. Dataset available at https://data.neonscience.org (accessed 25 January 2022)

R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Thomas RQ and Boettiger C. 2022. RELEASE-2022 and provisional data for NEON DP1.20264.001 at BARC, SUGG, CRAM, LIRO, PRLA, and PRPO. Zenodo repository. https://doi.org/10.5281/zenodo.5918679

Thomas RQ, Figueiredo RJ, Daneshmand V, *et al.* 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resour Res* **56**: e2019WR026138.

Thomas RQ, McClure RP, and Moore TN. 2022a. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.: code (v1.0). Zenodo repository. https://doi.org/10.5281/zenodo.6267617

Thomas RQ, Moore TN, and Daneshmand V. 2022b. Forecasting Lakes and Reservoir Ecosystems R-package (FLAREr): Version 2.2.1 (v2.2.1). Zenodo repository. https://doi.org/10.5281/zenodo.6098517

Thomas RQ and Woelmer WM. 2022. Daily NOAA Global Ensemble Forecasting System forecasts for six National Ecological Observatory Network lakes (2021-05-18 to 2021-10-24). Zenodo repository. https://doi.org/10.5281/zenodo.5918357

WebTable 1. Metadata of the six conterminous U.S. lake sites in the National Ecological Observatory Network. Variables that were included in the predictability correlation analysis included: latitude, maximum lake depth, fetch, volume, surface area, mean Secchi depth, mean annual temperature, mean annual precipitation, variance in air temperature, mean hydrological residence time, and catchment size.

siteID	Lake name	NEON Ecoclimatic domain	Latitude (°N)	Longitude (°E)	Elevation (m)	Maximum lake depth (m)	Fetch (m)	Volume (m^3)	Surface area (km^2)
BARC	Barco Lake	Southeast	29.675982	-82.008414	27	6	425	256888	0.12
SUGG	Suggs Lake	Southeast	29.68778	-82.017745	32	3	867	415356	0.31
CRAM	Crampton Lake	Great Lakes	46.209675	-89.473688	509	19	782	889734	0.26
LIRO	Little Rock Lake	Great Lakes	45.998269	-89.704767	501	10	623	466757	0.19
PRLA	Prairie Lake	Northern Plains	47.15909	-99.11388	565	4	1010	389429	0.23
PRPO	Prairie Pothole	Northern Plains	47.129839	-99.253147	579	4	511	158520	0.11

WebTable 1. Continued

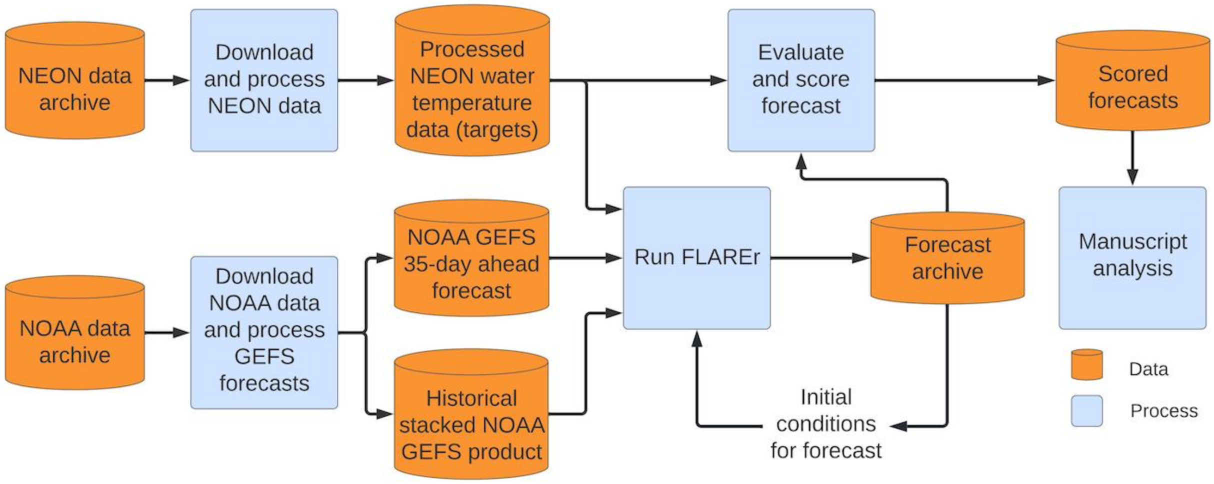
siteID	Mean Secchi depth (m)	Mixing regime	Mean annual temperature (°C)	Mean annual precipitation (mm)	Variance in air temperature (10-day standard deviation, °C)	Mean hydrological residence time (yrs)	Catchment size (km^2)	Number of years in time series for day-of-year null model
BARC	4.08	Polymictic	20.9	1308	1.09	3.3	0.8	2.4
SUGG	0.43	Polymictic	20.9	1308	1.09	1.6	36.9	3.4
CRAM	4.16	Dimictic	4.3	794	2.86	4.9	0.6	2.3
LIRO	4.37	Dimictic	4.4	796	2.86	3.4	0.9	3.1
PRLA	0.33	Polymictic	4.9	490	3.34	3.8	4.5	1.0
PRPO	0.40	Polymictic	4.9	494	3.39	3.2	1.4	2.0

WebTable 1. Continued

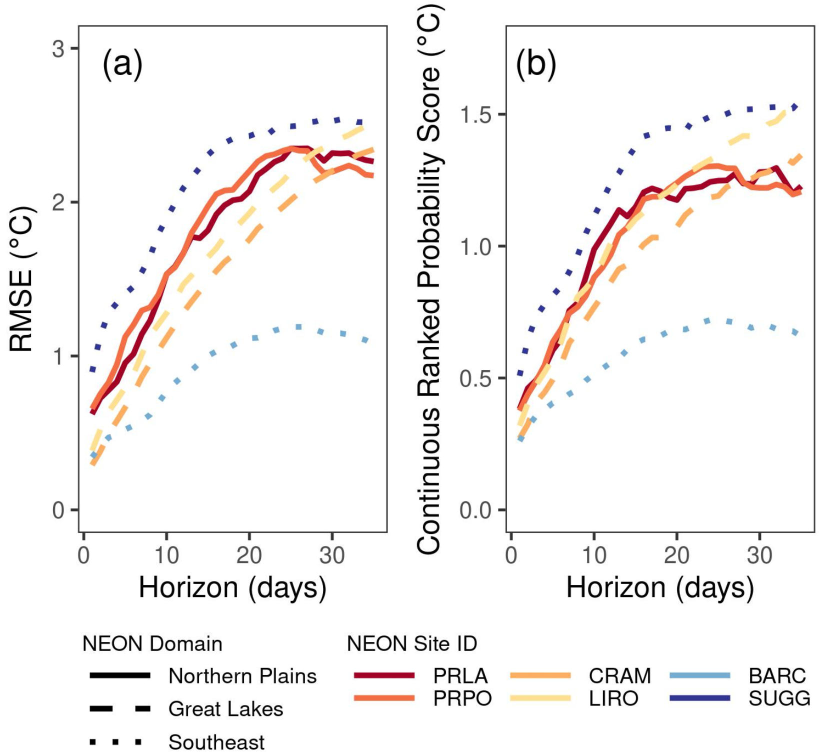
siteID	Catchment land cover	NEON Website
BARC	shrub/scrub	https://www.neonscience.org/field-sites/barc
SUGG	evergreen/forest; woody wetlands	https://www.neonscience.org/field-sites/sugg
CRAM	woody wetlands	https://www.neonscience.org/field-sites/cram
LIRO	deciduous forest; mixed forest	https://www.neonscience.org/field-sites/liro
PRLA	grassland/herbaceous	https://www.neonscience.org/field-sites/prla
PRPO	grassland/herbaceous	https://www.neonscience.org/field-sites/prpo

WebTable 2. Forecast accuracy, defined as root-mean square error (RMSE) at 1-day ahead, and forecast accuracy degradation, defined as the difference in maximum and minimum RMSE across the 35-day forecast horizon. We used Spearman rank correlations to quantify the relationships between morphometric, hydrological, ecological, and meteorological characteristics and mean forecast accuracy and accuracy degradation for each lake. To ease interpretation of the correlation coefficient, we negated RMSE so positive correlations are associated with higher accuracy. Given the extremely limited sample size of lakes (n=6), which is too small for reliable p-values for rho, we focused our interpretation on Spearman rho correlations |≥| 0.5 (included here).

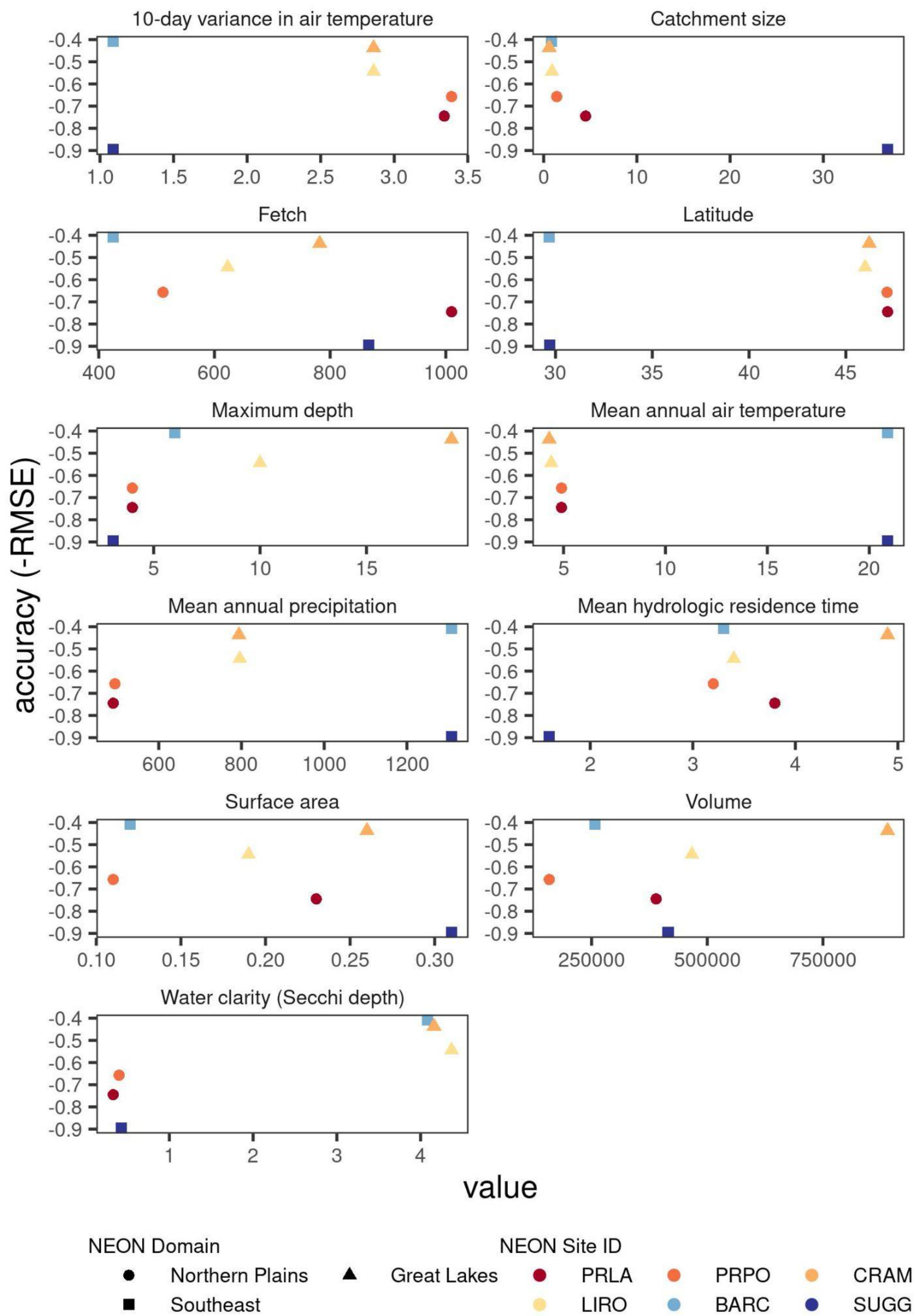
variable	metric	rho
Catchment size	accuracy	-0.94
Fetch	accuracy	-0.71
Maximum depth	accuracy	0.81
Water clarity (Secchi depth)	accuracy	0.60
Mean annual air temperature	degradation	-0.79
Water clarity (Secchi depth)	degradation	0.60
Volume	degradation	0.60



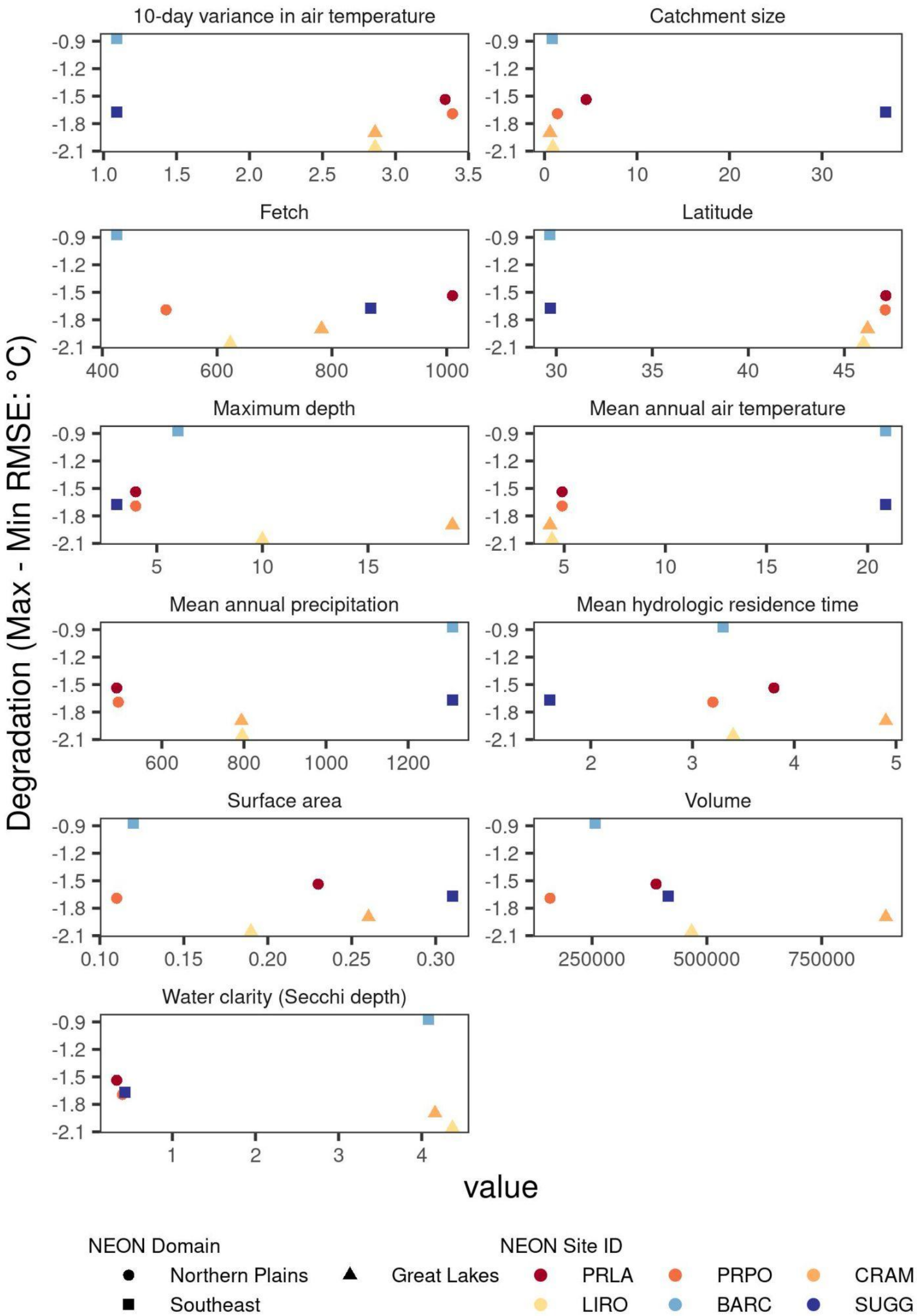
WebFigure 1. A diagram of the workflow used to generate the daily iterative forecasts using NOAA Global Ensemble Forecasting System (GEFS) meteorology forecasts, National Ecological Observatory Network (NEON) water temperature data, and the Forecasting Lake and Reservoir Ecosystems R package (FLAREr).



WebFigure 2. (a) Forecast accuracy for water temperature at all depths in each lake aggregated together. Accuracy is defined by RMSE (root-mean square error in ºC), calculated separately for each 1 to 35-days ahead (horizon) at the six NEON lakes. (b) Surface water temperature forecast accuracy, defined by the Continuous Ranked Probability Score (CRPS, in ºC), a metric that uses the entire ensemble to evaluate the forecast, which is analogous to mean absolute error.



WebFigure 3. Relationships between forecast accuracy (y-axis) and the morphometric, hydrological, ecological, and weather characteristics included in Figure 3 (x-axis). We negated RMSE (root-mean square error in ^{o}C), so positive correlations are associated with higher accuracy. WebTable 1 includes the units for each variable.



WebFigure 4. Relationships between forecast accuracy degradation (y-axis) and the morphometric, hydrological, ecological, and weather characteristics included in Figure 3 (x-axis). Degradation is defined as the difference in RMSE (root-mean square error in ºC) between the maximum and minimum RMSE over the 35-day forecast horizon. WebTable 1 includes the units for each variable.

WebPanel 1. Description of the forecasted NEON lakes, overview of the FLARE configuration for each lake, meteorological driver data, mean day-of-year null model, and guide to reproducibility.

NEON Lake temperature data

We generated forecasts for the six NEON lakes in the conterminous USA (WebTable 1). The six forecast sites were two paired lakes in the Great Lakes NEON ecoclimatic domain (Crampton Lake, NEON site ID – CRAM; Little Rock Lake, NEON site ID - LIRO), two paired lakes in the Northern Plains domain (Prairie Lake, NEON siteID – PRLA; Prairie Pothole, NEON siteID - PRPO), and two paired lakes in the Southeastern domain (Barco Lake, NEON siteID – BARC; Suggs Lake, NEON siteID - SUGG). We excluded the seventh NEON lake site (Toolik Lake) since it was not part of a paired NEON set and it has major surface inflows, unlike the other lakes.

Each lake had 5-10 water temperature sensors (Precision Measurement Engineering Inc. T-Chain RS 232/485 thermistors) deployed at various depths in the water column. The first sensor was deployed 0.05 m below the surface, with remaining depths dependent on the total depth of the lake. Generally, sensors were deployed at more frequent intervals within the upper 1.05 m than at deeper depths. These discrete depth water temperature data are available from NEON (NEON 2022a, b), and were accessed using the *neonstore* R package, which creates a "store" of NEON data on a local computer and eases the iterative downloading of additional NEON data without re-downloading data already within the store (Boettiger *et al.* 2021).

All data were filtered using the quality assurance codes provided by NEON. The 30-minute data product was aggregated to the hour and only the 00:00-01:00 UTC hour was used each day for assimilation and evaluation. The NEON (NEON 2022a, b) data were exported using the *neon_export* function in the *neonstore* R package and archived at Thomas and Boettiger (2022). Gaps in NEON's discrete depth water temperature dataset were filled using water temperature data collected by a YSI EXO2 multiparameter sonde as part of NEON's water quality data product (Hensley 2022).

FLARE and GLM configuration

Adapting FLARE to NEON lakes required configuring six unique GLM models with each lake's bathymetry and physical specifications and developing functions to download and process NEON water temperature data. Across all six lakes, we used the same initial default GLM hydrodynamic parameters (Hipsey *et al.* 2019) and tuned the same set of three parameters governing lake water temperature during data assimilation (lw_factor, kw, and sed_mean_temp). Since none of the six NEON lakes have major surface inflows or outflows and prior applications at a reservoir in Virginia showed limited sensitivity of forecast uncertainty to inflows (Thomas *et al.* 2020), we parameterized each lake without inflows or outflows.

We parameterized the process uncertainty in water temperature to be the same across sites and throughout the water column (standard deviation = 0.75°C). This value was based on the findings of Thomas *et al.* (2020), in which FLARE's process uncertainty was estimated across water column depths at a reservoir in Virginia. The process uncertainty was added to each ensemble member and modeled depth at each daily timestep. Since we expect this uncertainty to be correlated with depth (e.g., if the modeled temperature at a certain depth was 1°C warmer than observed, nearby depths should also likely be too warm as well), we included a correlation length that represents an exponential decay of correlations across depths (following Appendix A in Lenartz *et al.* 2007). The decay in correlation results in stronger correlations in water temperature at closer depths than further away depths. This decorrelation length parameter was set to 2 m.

Similarly, observation uncertainty in water temperature data was set to be the same across lakes and depths (standard deviation = 0.1°C), based on the FLARE application in Thomas *et al.* (2020). Since observation uncertainty represents sensor and sampling uncertainty, we did not expect observation uncertainty to be correlated with depth, and therefore the decorrelation length for this uncertainty source was set to 0 m.

Parameter estimation using the ensemble Kalman filter (EnKF) uses the estimated correlation between parameter values and the size of the errors between the predicted and observed states across ensemble members (Evensen 2009). Ensemble members that require large adjustments in the states to be consistent with observations will also adjust parameters that are correlated with that error. One challenge with estimating parameters using the EnKF is that the variation in parameter values across ensemble members collapses over time. The small variance among ensemble members prevents the parameters from further adjusting to reduce new biases in the model predictions (i.e., the calibration does not change through time).

As a result, parameter estimation methods using the EnKF need to use a technique to prevent a collapse in variance. Here, we use a method called variance inflation, in which the variance in parameter values among the ensemble members is increased at each time-step when data assimilation occurs. The variance inflation increases the spread in the parameters among ensemble members while maintaining the rank order of ensemble members. We used the same variance inflation factor across all parameters and lakes (0.04).

The FLAREr R package that contains FLARE functions can be found in the Zenodo repository (Thomas *et al.* 2022b), as well as the scripts for running FLARE at the six NEON lakes (Thomas *et al.* 2022a). All analyses were conducted in R software version 4.1.1 (R Core Team 2021).

Meteorological inputs

The forecasts were driven by numerical meteorological forecasts produced by NOAA's Global Ensemble Forecasting System (GEFS) version 12 (Li *et al.* 2019). We automated the downloading of ensemble members (n=31 total) from the NOAA GEFS output for each 0.5°×0.5° grid cell that included a NEON lake. NOAA GEFS generates weather forecasts at multiple times per day (00:00, 06:00, 12:00, and 18:00 UTC), which vary in their forecast horizon length (i.e., days into the future). We focused on the GEFS weather forecast that started at 00:00 UTC each day, as 30 of its 31 ensemble members extended 35 days into the future on a 6-hour time step and included all meteorological variables required by the GLM as model driver data. The 6-hour output resolution of each of the 30 ensemble members was temporally disaggregated to 1-hour resolution for use in the GLM following Thomas *et al.* (2020).

We used a "stacked" GEFS product during the 1-month spin-up period. One challenge when using data assimilation to set initial conditions and tune parameters is a potential mismatch between the meteorological data used in the spin-up and data used for generating future forecasts. Since observed and forecasted meteorology are rarely a 1:1 match, a smooth transition from data assimilation to forecasting requires either the forecasted meteorology to be corrected for the site or past meteorological forecasts to be used in place of observed meteorology for data assimilation. Here, we used the latter option because NEON meteorological data has a 1.5-month latency and often has gaps for some of the required meteorological variables. To develop a "stacked" GEFS product, we also downloaded the 0-hour and 6-hour horizon of the forecasts that were initiated every six hours at 06:00, 12:00, and 18:00 UTC each day (the 0-hour and 6-hour for the 00:00 UTC forecast were already downloaded as part of the full 35-day horizon). We then combined the temperature, relative humidity, and wind speed from the 0-hour horizon for all NOAA GEFS forecasts. The flux variables (precipitation, longwave radiation, and shortwave radiation) required using the 6-hour horizon because they integrate the 0^{th} to 6^{th} hour. The 0 and 6-hour horizons were used because they directly follow data assimilation in the GEFS, and therefore are most closely aligned with observed meteorology. The resulting "stacked" product was a 6-hr time-step meteorology product because the time step between the initiation of new forecasts was six hours. The stacked data product was updated each time new GEFS forecasts are available, and thus was near-real time.

To estimate the 10-day variance in air temperature that was used in the predictability correlation analysis, we calculated the running standard deviation over a rolling 10-day window between 18 May 2021 and 31 October 2021 from the "stacked" GEFS product. We used the mean of the 10-day running standard deviation to represent air temperature variance for each lake during the period that forecasts were generated.

All NOAA GEFS 1-hour forecasts and "stacked" products for the six NEON lakes are archived at Thomas et al (2022b).

Mean Day-of-Year Null Forecast

We note that while the 1 to 3.5 years of data at the NEON lakes available for this day-of-year (DOY) null model (see WebTable 1) is a shorter duration than the ~30 years of data typically used in weather forecasting null climatology models, it still included all NEON data available for each lake. Moreover, the DOY null model for the lake with just one year of data (PRLA) performed similarly to the DOY null model for its paired lake (PRPO), which had three years of data (Figure 2b).

Guide to Reproducibility

We have provided all code used to generate forecasts, analyze forecasts, and recreate figures in this manuscript as a GitHub repository that has been archived on Zenodo (Thomas et al. 2022a). There are three steps to the analysis that are documented as separate R scripts within the repository. First, the "01_combined_paper_workflow.R" in the "workflows/neon_lakes_ms/" directory of the repository obtains the NEON data and NOAA GEFS weather forecasts and then runs FLARE on the six sites. Since this script runs 159 separate 35-day horizon forecasts for the six lakes, the time required to generate all forecasts depends on the number and speed of computer processors available and can be a multi-day execution. This first step produces a set of output files for the GLM-based and day-of-year null forecasts in a "forecasts" directory.

Second, each ensemble forecast from the first step is aggregated to a mean with predictive intervals and scored (by matching to the corresponding observation, if available), with the summary statistics and observations saved as a set of scored files (one per output file) in a "scores" directory in the repository. The scoring is generated by the "02_score_forecasts.R" script located in the "workflows/neon_lakes_ms/" directory of the repository. While the scores can be generated using output files from the first step, we also provide the output files as an additional Zenodo repository (Thomas et al. 2022b) that can be downloaded and scored using the script without needing to re-run the forecasts.

Third, the scored files are analyzed using an Rmarkdown script located in the main directory of repository ("analysis_notebook.Rmd") to produce the figures and data reported in the text. The Rmarkdown script can use the scored files produced by the second step or the scores files available in the additional Zenodo repository (Thomas et al. 2022b).

Our analysis can be reproduced by downloading the Zenodo GitHub repository and running the three scripts associated with the steps described above. Re-running the full analysis requires downloading R, Rstudio, and all the required packages, and as noted above, can take multiple days of execution, depending on the computation available. We provide a script that downloads the required packages ("install.R" in the main directory of the repository). However, there is no guarantee that other versions of R and packages will produce the same results as presented here.

To enable greater reproducibility, we adapted the GitHub repository (Thomas et al. 2022a) to generate a Binder that is produced by mybinder.org (Jupyter et al 2018). Mybinder.org provides a web-based version of Rstudio for re-running our GitHub repository code that uses the same version of R and R packages that we used in this analysis (https://mybinder.org/v2/zenodo/10.5281/zenodo.6267616/?urlpath=rstudio). As a result, there is more confidence that the analysis can be reproduced by harnessing the Binder infrastructure, which directly re-runs the analysis on a remote server and provides an Rstudio interface via a web browser for running the scripts described above for each of the three analysis steps.

There are important caveats to using the Binder. First, at the time of this analysis, mybinder.org is free to use, and therefore its computational resources have limits and processing times can be slow. Consequently, we do not recommend running the full generation of the 35-day forecasts in the Binder. The Binder is ideally suited for exploring the scored forecasts and reproducing the figures and values presented in the text (i.e., the "analysis_notebook.Rmd" script described in the third step above). Second, at the time of this analysis, the Binder does not always consistently launch when accessing the Binder link and occasionally the connection times out. It may require accessing the Binder link again to get a successful launch of the R studio interface.

WebReferences

Boettiger C, Thomas RQ, Laney C, and Lunch C. 2021. neonstore: NEON Data Store. R package. CRAN repository. https://cran.r-project.org/web/packages/neonstore/index.html

Evensen G. 2009. Data Assimilation. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hensley, R.T. 2022. NEON lakes Level 0 multisonde temperature data - 2021 ver 1. Environmental Data Initiative repository. https://doi.org/10.6073/pasta/fbbd2d5f59a8d92c6865d57e7abae379 (Accessed 2022-01-25).

Hipsey MR, Bruce LC, Boon C, *et al.* 2019. A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geosci Model Dev* **12**: 473–523.

Jupyter et al. 2018. Binder 2.0 - reproducible, interactive, sharable environments for science at scale." Proceedings of the 17th Python in Science Conference (SCIPY 2018). Austin, Texas: SciPy. https://doi.org/10.25080/Majora-4af1f417-011.

Lenartz F, Raick C, Soetaert K, and Grégoire M. 2007. Application of an Ensemble Kalman filter to a 1-D coupled hydrodynamic-ecosystem model of the Ligurian Sea. *J Mar Syst* **68**: 327–48.

Li W, Guan H, Zhu Y, *et al.* 2019. Prediction Skill of the MJO, NAO and PNA in the NCEP FV3-GEFS 35-day Experiments. In: Science and Technology Infusion Climate Bulletin. Durham, NC: NOAA's National Weather Service.

NEON. 2022a. Temperature at specific depth in surface water (DP1.20264.001). Dataset available at https://data.neonscience.org (accessed 25 January 2022)

NEON. 2022b. Temperature at specific depth in surface water, RELEASE-2022 (DP1.20264.001). https://doi.org/10.48443/g7bs-7j57. Dataset available at https://data.neonscience.org (accessed 25 January 2022)

R Core Team. 2021. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Thomas RQ and Boettiger C. 2022. RELEASE-2022 and provisional data for NEON DP1.20264.001 at BARC, SUGG, CRAM, LIRO, PRLA, and PRPO. Zenodo repository. https://doi.org/10.5281/zenodo.5918679

Thomas RQ, Figueiredo RJ, Daneshmand V, *et al.* 2020. A near-term iterative forecasting system successfully predicts reservoir hydrodynamics and partitions uncertainty in real time. *Water Resour Res* **56**: e2019WR026138.

Thomas RQ, McClure RP, Moore TM, Woelmer WM, Boettiger C, Figueiredo RJ, Hensley RT, and Carey CC. 2022a. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.: code (v1.1). Zenodo repository. https://doi.org/10.5281/zenodo.6674487

Thomas RQ, McClure RP, Moore TM, *et al*. 2022b. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S.: data, forecasts, and scores. Zenodo repository. https://doi.org/10.5281/zenodo.6643596

WebTable 1. Metadata of the six conterminous U.S. lake sites in the National Ecological Observatory Network. Variables that were included in the predictability correlation analysis included: latitude, maximum lake depth, fetch, volume, surface area, mean Secchi depth, mean annual temperature, mean annual precipitation, variance in air temperature, mean hydrological residence time, and catchment size.

siteID	Lake name	NEON Ecoclimatic domain	Latitude (°N)	Longitude (°E)	Elevation (m)	Maximum lake depth (m)	Fetch (m)	Volume (m^3)	Surface area (km^2)
BARC	Barco Lake	Southeast	29.675982	-82.008414	27	6	425	256888	0.12
SUGG	Suggs Lake	Southeast	29.68778	-82.017745	32	3	867	415356	0.31
CRAM	Crampton Lake	Great Lakes	46.209675	-89.473688	509	19	782	889734	0.26
LIRO	Little Rock Lake	Great Lakes	45.998269	-89.704767	501	10	623	466757	0.19
PRLA	Prairie Lake	Northern Plains	47.15909	-99.11388	565	4	1010	389429	0.23
PRPO	Prairie Pothole	Northern Plains	47.129839	-99.253147	579	4	511	158520	0.11

WebTable 1. Continued

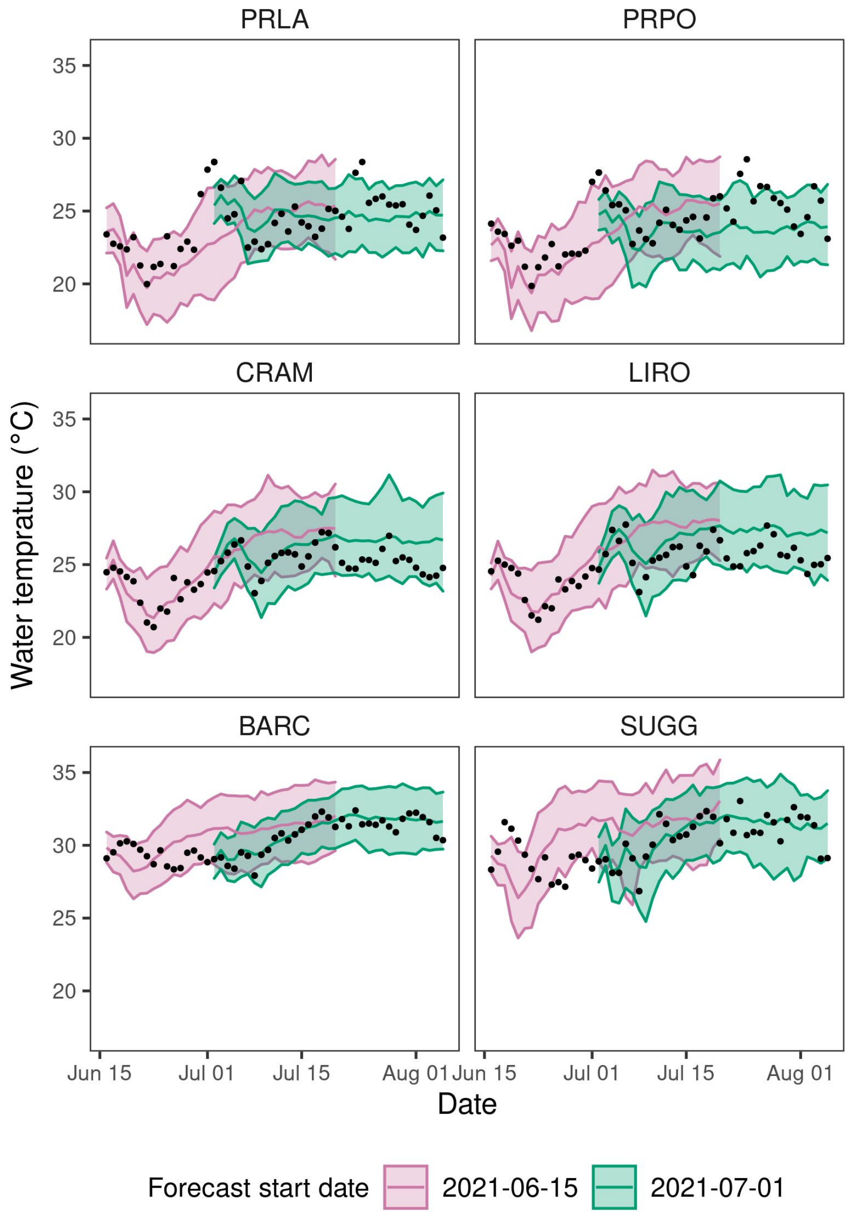
siteID	Mean Secchi depth (m)	Mixing regime	Mean annual temperature (°C)	Mean annual precipitation (mm)	Variance in air temperature (10-day standard deviation, °C)	Mean hydrological residence time (yrs)	Catchment size (km^2)	Number of years in time series for day-of-year null model
BARC	4.08	Polymictic	20.9	1308	1.09	3.3	0.8	2.4
SUGG	0.43	Polymictic	20.9	1308	1.09	1.6	36.9	3.4
CRAM	4.16	Dimictic	4.3	794	2.86	4.9	0.6	2.3
LIRO	4.37	Dimictic	4.4	796	2.86	3.4	0.9	3.1
PRLA	0.33	Polymictic	4.9	490	3.34	3.8	4.5	1.0
PRPO	0.40	Polymictic	4.9	494	3.39	3.2	1.4	2.0

WebTable 1. Continued

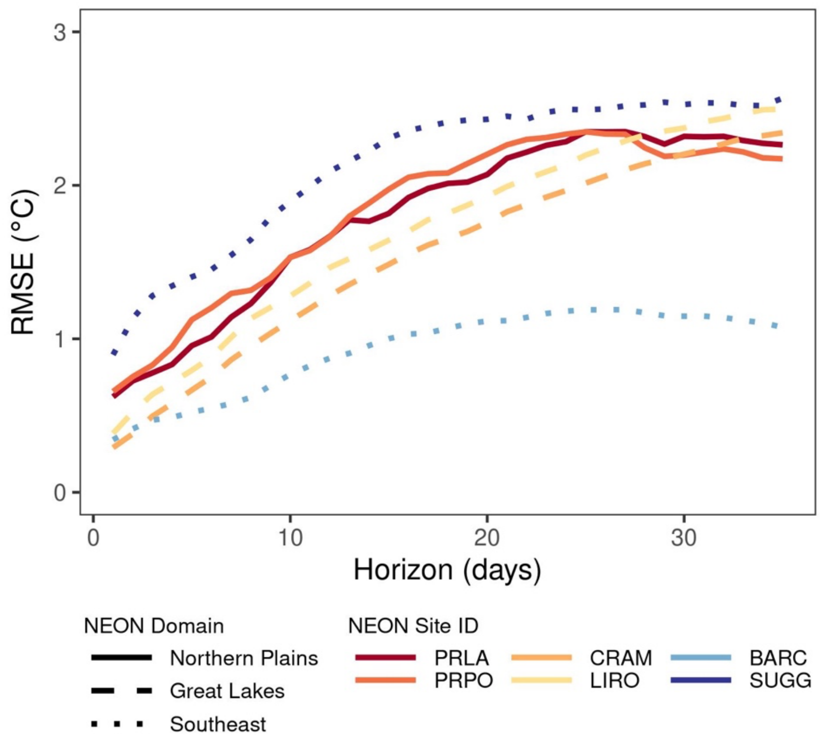
siteID	Catchment land cover	Depths with sensor observations (value is top of 0.25 m thick bin)	NEON Website
BARC	shrub/scrub	0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00 2.50, 3.00	https://www.neonscience.org/field-sites/barc
SUGG	evergreen/forest; woody wetlands	0.00, 0.25, 0.50, 0.75, 1.00	https://www.neonscience.org/field-sites/sugg
CRAM	woody wetlands	0.00, 0.25, 0.50, 0.75, 1.00, 1.75, 2.00, 2.50, 3.25, 3.50, 4.25, 4.75, 5.00, 6.25, 6.50, 6.75, 7.75, 8.00, 8.50, 9.25, 9.50, 10.25, 10.75, 11.00 12.00, 12.50, 13.50, 14.00, 15.50	https://www.neonscience.org/field-sites/cram
LIRO	deciduous forest; mixed forest	0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 4.00, 4.25, 4.50, 4.75, 5.00, 5.75, 6.00, 6.75	https://www.neonscience.org/field-sites/liro
PRLA	grassland/herbaceous	0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00	https://www.neonscience.org/field-sites/prla
PRPO	grassland/herbaceous	0.00, 0.25, 0.50, 0.75, 1.00	https://www.neonscience.org/field-sites/prpo

WebTable 2. Forecast accuracy, defined as root-mean square error (RMSE) at 1-day ahead, and forecast accuracy degradation, defined as the difference in maximum and minimum RMSE across the 35-day forecast horizon. We used Spearman rank correlations to quantify the relationships between morphometric, hydrological, ecological, and meteorological characteristics and mean forecast accuracy and accuracy degradation for each lake. To ease interpretation of the correlation coefficient, we negated RMSE so positive correlations are associated with higher accuracy. Given the extremely limited sample size of lakes (n=6), which is too small for reliable p-values for rho, we focused our interpretation on Spearman rho correlations |≥| 0.5 (included here).

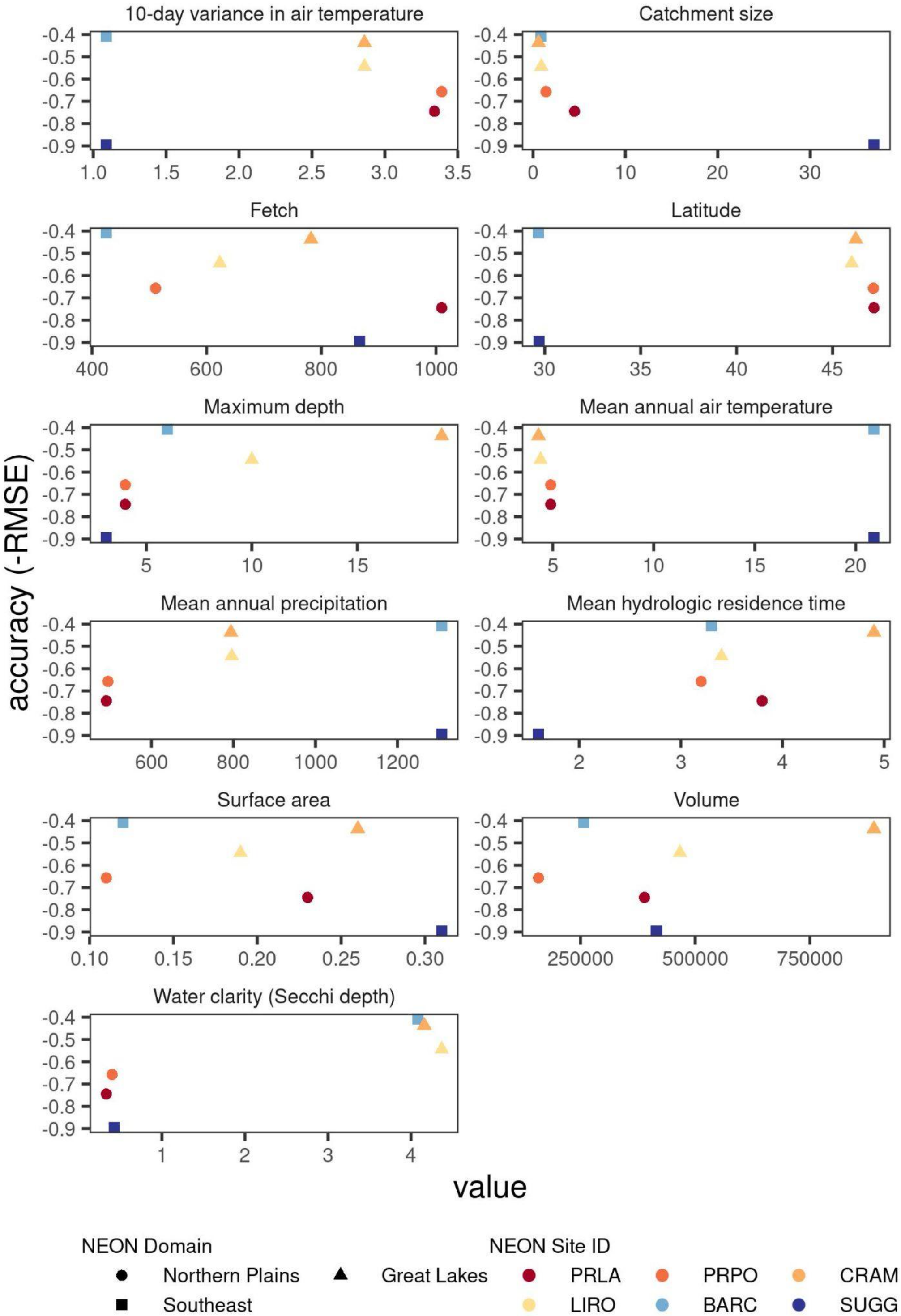
variable	metric	rho
Catchment size	accuracy	-0.94
Fetch	accuracy	-0.71
Maximum depth	accuracy	0.81
Water clarity (Secchi depth)	accuracy	0.60
Mean annual air temperature	degradation	-0.79
Water clarity (Secchi depth)	degradation	0.60
Volume	degradation	0.60



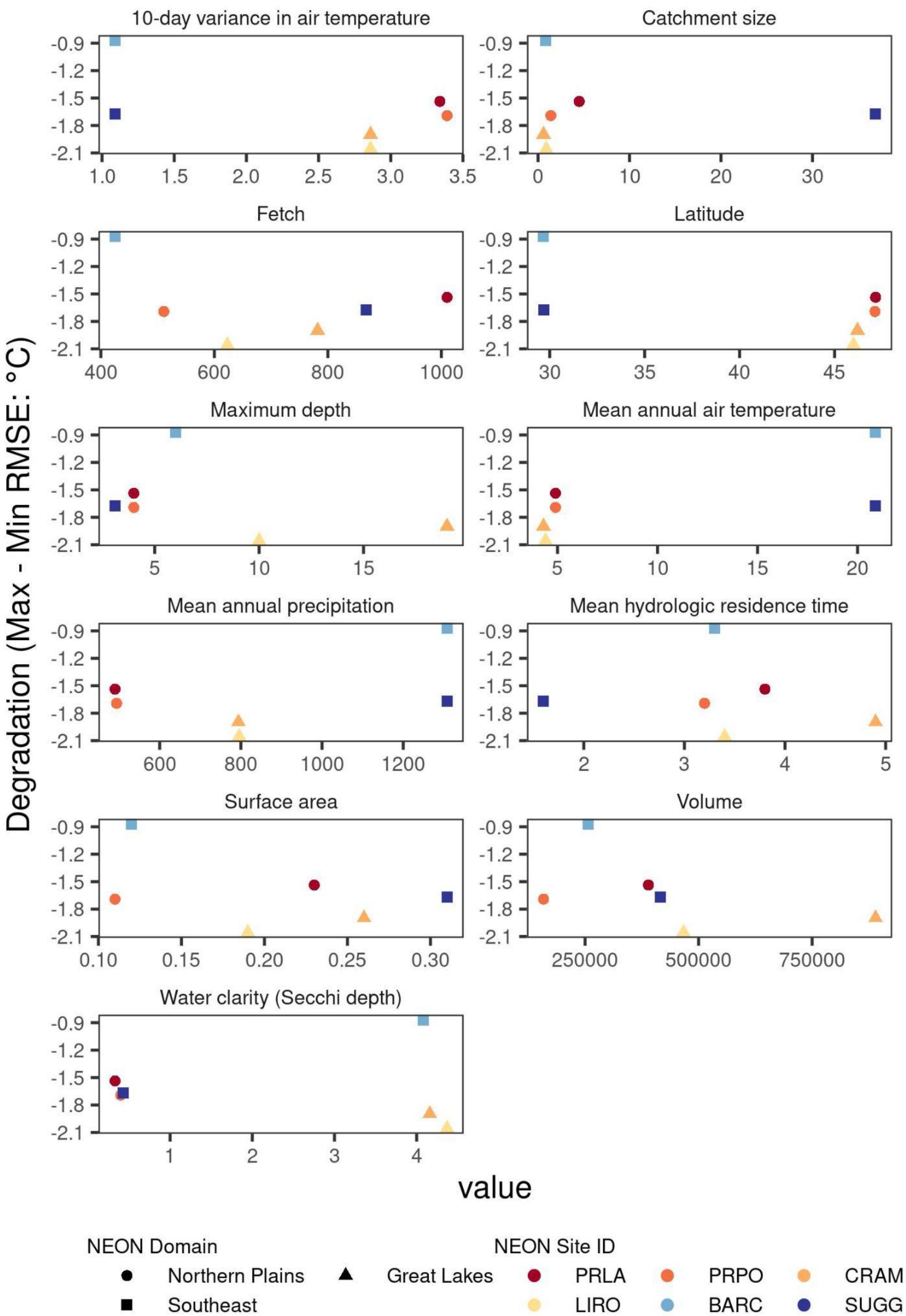
WebFigure 1. Example 35-day forecasts of surface water temperature that were initiated on 2021-06-15 and 2021-07-01. The shaded region represents the 10% and 90% quantiles. The observations (black dots) are provided for reference.



WebFigure 2. Forecast accuracy for water temperature at all depths in each lake aggregated together. Accuracy is defined by RMSE (root-mean square error in ^{o}C), calculated separately for each 1 to 35-days ahead (horizon) at the six NEON lakes.



WebFigure 3. Relationships between forecast accuracy (y-axis) and the morphometric, hydrological, ecological, and weather characteristics included in Figure 3 (x-axis). We negated RMSE (root-mean square error in ºC), so positive correlations are associated with higher accuracy. WebTable 1 includes the units for each variable.



WebFigure 4. Relationships between forecast accuracy degradation (y-axis) and the morphometric, hydrological, ecological, and weather characteristics included in Figure 3 (x-axis). Degradation is defined as the difference in RMSE (root-mean square error in ºC) between the maximum and minimum RMSE over the 35-day forecast horizon. WebTable 1 includes the units for each variable.