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

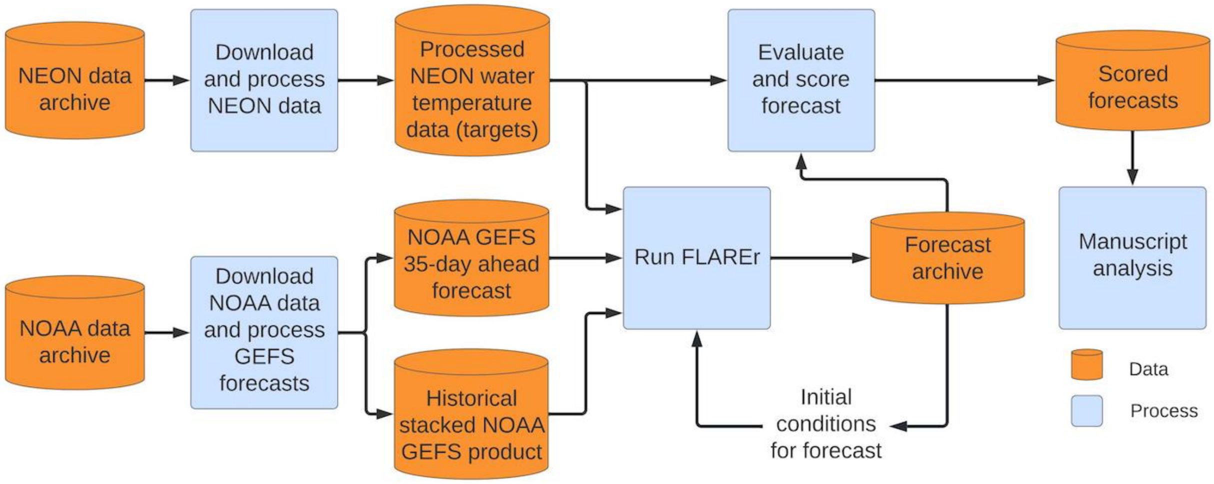
siteID	Mean Secchi depth (m)	Mixing regime	Mean annual temperature (°C)	Mean annual precipitation (mm)	Variance in air temperature (10-day standard deviation, °C)	Mean hydrological residence time (yrs)	Catchment size (km^2)	Number of years in time series for day-of-year null model
BARC	4.08	Polymictic	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

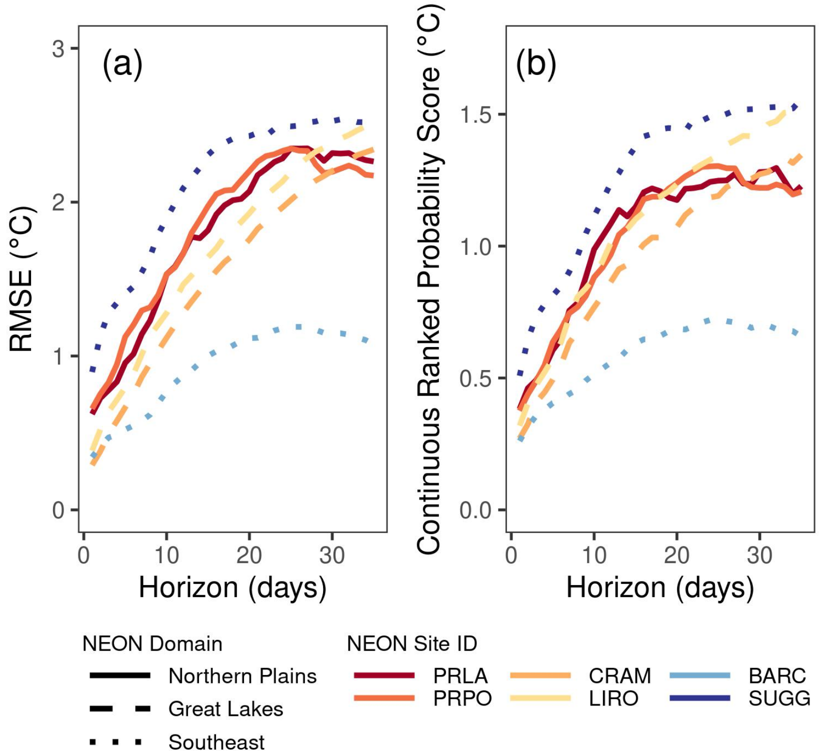
siteID	Catchment land cover	NEON Website
BARC	shrub/scrub	https://www.neonscience.org/field-sites/barc
SUGG	evergreen/forest; woody wetlands	https://www.neonscience.org/field-sites/sugg
CRAM	woody wetlands	https://www.neonscience.org/field-sites/cram
LIRO	deciduous forest; mixed forest	https://www.neonscience.org/field-sites/liro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

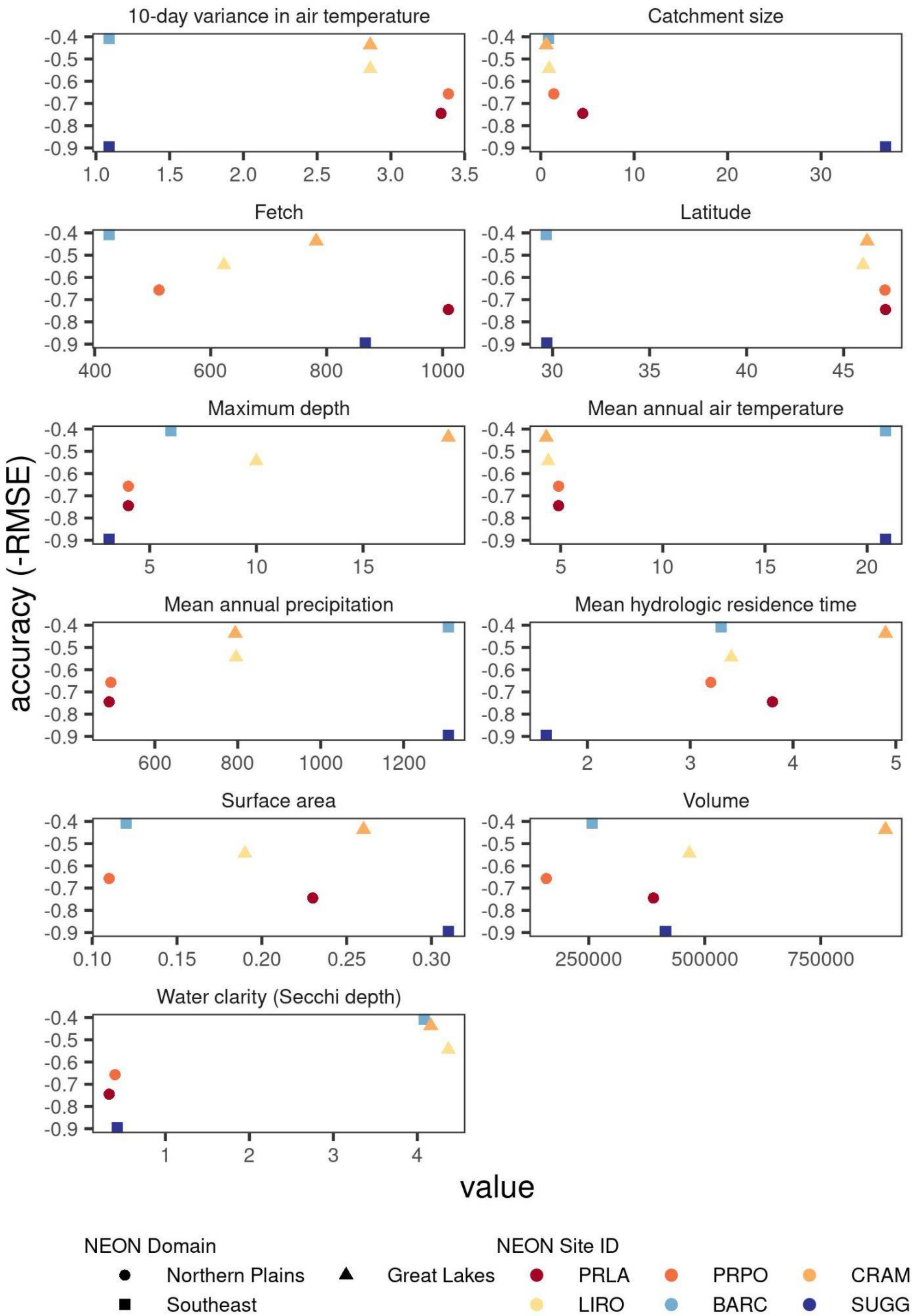
variable	metric	rho
Catchment size	accuracy	-0.94
Fetch	accuracy	-0.71
Maximum depth	accuracy	0.81
Water clarity (Secchi depth)	accuracy	0.60
Mean annual air temperature	degradation	-0.79
Water clarity (Secchi depth)	degradation	0.60
Volume	degradation	0.60



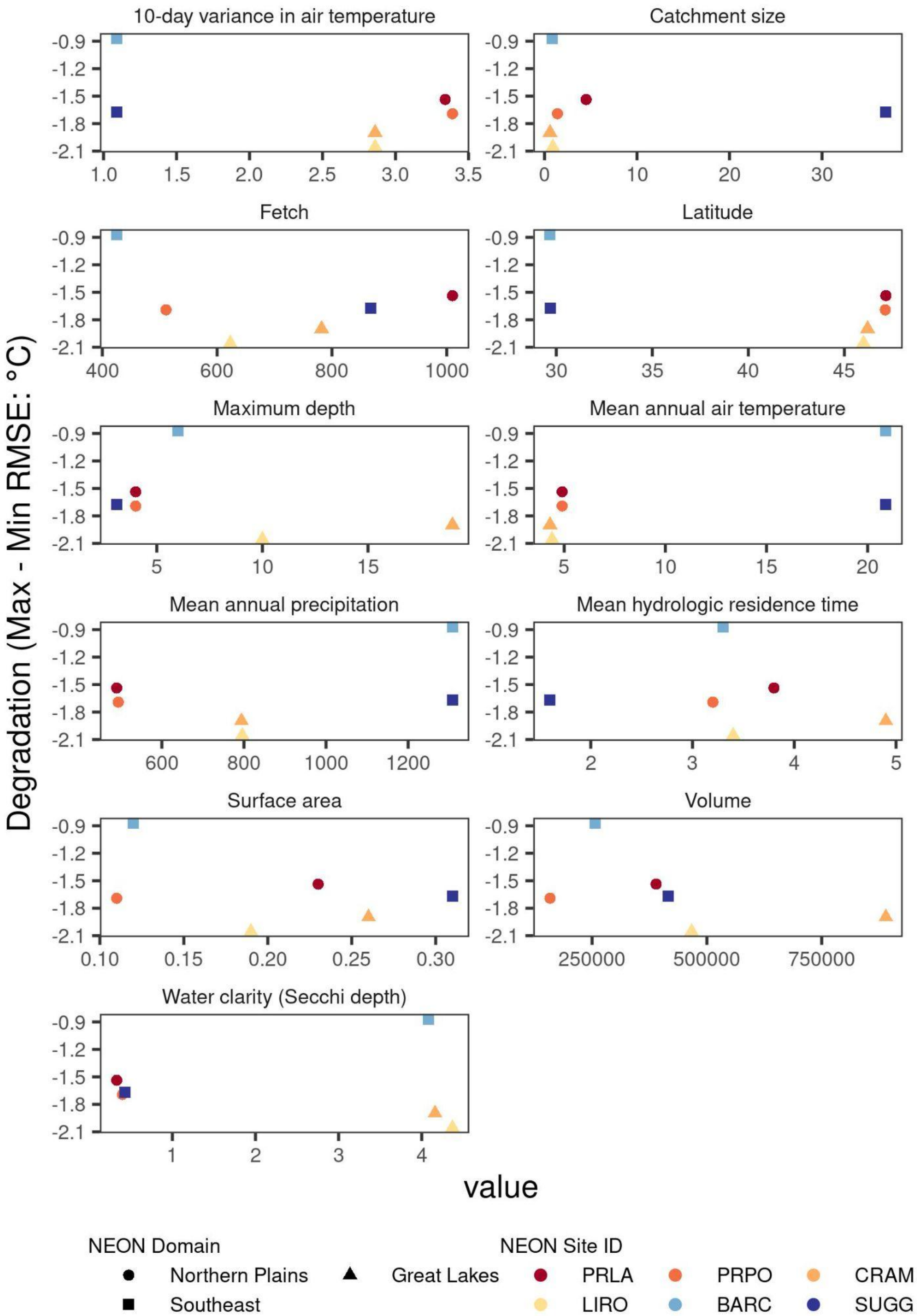
WebFigure 1. A diagram of the workflow used to generate the daily iterative forecasts using NOAA Global Ensemble Forecasting System (GEFS) meteorology forecasts, National Ecological Observatory Network (NEON) water temperature data, and the Forecasting Lake and Reservoir Ecosystems R package (FLAREr).



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 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.