








OPEN

DATA DESCRIPTOR

High-resolution freshwater dissolved calcium and pH data layers for Canada and the United States

Andrew J. Guerin¹  , Andréa M. Weise¹  , Jackson W. F. Chu² , Mark A. Wilcox³, Erin Sowerby Greene³ & Thomas W. Therriault³

Freshwater ecosystems are biologically important habitats that provide many ecosystem services. Calcium concentration and pH are two key variables that are linked to multiple chemical processes in these environments, influence the biology of organisms from diverse taxa, and can be important factors affecting the distribution of native and non-native species. However, it can be challenging to obtain high-resolution data for these variables at regional and national scales. To address this data gap, water quality data for lakes and rivers in Canada and the continental USA were compiled and used to generate high-resolution (10 × 10 km) interpolated raster layers, after comparing multiple spatial interpolation approaches. This is the first time that such data have been made available at this scale and resolution, providing a valuable resource for research, including projects evaluating risks from environmental change, pollution, and invasive species. This will aid the development of conservation and management strategies for these vital habitats.

Background and Summary

Calcium concentration and pH are key determinants of many environmental and biological processes in freshwater ecosystems. Both variables regulate metabolic physiology in aquatic organisms, influencing reproduction, growth, and predator-prey interactions across a wide range of taxa including bacteria^{1,2}, aquatic algae and diatoms^{3,4}, molluscs⁵, crustacea⁶, and fish^{4,7,8}. Since differences in these parameters can lead to detectable biological effects on individuals, populations, and communities^{9,10}, pH and calcium concentration can both be important predictors of species distributions^{11,12} and are often used to evaluate the risk of establishment for invasive species, such as dreissenid mussels^{13–15}. pH and dissolved calcium content of lakes influence their susceptibility to acidification^{16,17}. They affect nutrient availability^{18,19}, and play an important role in determining the environmental risks posed by metals and other contaminants by influencing their dissolution, mobilization, bioavailability, and toxicity^{20–22}, as well as mediating their adsorption and desorption by microplastics^{23,24}.

For large-scale studies at regional, national and continental levels, a common challenge facing freshwater researchers and resource managers is the availability of water quality data¹⁷, including calcium and pH. Such data are not readily available for all areas of North America, and given the large number of lakes, rivers, and other water bodies in Canada and the USA, measuring these variables at all sites would be prohibitively expensive and impractical. One way of improving water quality data coverage is to use existing measurements to predict values for unsampled locations *via* spatial interpolation^{14,25,26}. This approach has several advantages: large amounts of data from multiple sources can be combined, no complex mechanistic modelling is required, and a range of established interpolation methods are available.

The goal of this work was therefore to use spatial interpolation to generate calcium and pH raster layers for the entirety of Canada and the continental USA at higher spatial resolution and coverage than previously

¹Maurice Lamontagne Institute, Fisheries and Oceans Canada, 850 route de la mer, PO Box 1000, Mont Joli, Quebec, G5H 3Z4, Canada. ²Pacific Science Enterprise Centre, Fisheries and Oceans Canada, 4160 Marine Drive, West Vancouver, British Columbia, V7V 1N6, Canada. ³Pacific Biological Station, Fisheries and Oceans Canada, 3190 Hammond Bay Road, Nanaimo, British Columbia, V9T 6N7, Canada. ✉e-mail: andrewjguerin@gmail.com; Andrea.Weise@dfo-mpo.gc.ca

available^{13,15}. An expansive dataset covering Canada and the USA (1,347,887 calcium measurements from 97,648 locations, and 8,789,005 pH measurements from 208,784 locations) was compiled from multiple governmental, non-governmental, and academic sources, and used to generate spatially interpolated maps of these variables at a 10 × 10 km resolution. These layers will be of value for projects requiring calcium and pH data at regional to continental scales, including understanding past and present sensitivity of lakes and rivers to acidification^{16,17}, assessment of regional variation in the risks posed by contaminants²⁰, ecological niche modelling²⁷, and invasive species risk assessment^{13,15}.

Methods

Data sources. Since Canada lacks a centralised repository for water quality data, georeferenced Canadian water quality records were obtained from multiple sources: publicly-accessible federal^{28–31}, provincial and territorial agency databases^{32–40}; non-governmental open access data repositories - the Atlantic Datastream (<https://atlanticdatastream.ca/>)^{41–93} and the Mackenzie Datastream (<https://mackenziedatastream.ca/>)^{52,94–115}; published reports and primary literature^{116–124}; a previous invasive species risk assessment¹⁵; and directly from contacts in relevant agencies in each of the provinces and territories (Table 1). Records for the United States (including Alaska, but excluding Hawaii) were obtained from the Water Quality Portal¹²⁵, which combines data from federal, state, tribal and local agencies; the *dataRetrieval* package¹²⁶ was used to directly download data for sites with calcium and pH data collected between 2000 and 2021 (Water Quality Portal accessed 15th February 2021). To ensure that records were as contemporary as possible while retaining high spatial coverage, records from before 2000 were excluded for most sources. However, older records were retained for some areas of Canada (particularly the Territories) where fewer data were generally available. All data handling, processing and interpolation was conducted in R v4.1.0¹²⁷.

Data processing and preparation for interpolation. Records from appropriate site types (lakes, rivers, ponds, and streams) were selected where possible, although most data sources did not provide this information. Records from marine waters or in proximity to mines, industrial facilities, wastewater treatment infrastructure or other potentially-contaminated sites were excluded if this information was provided. For USA Water Quality Portal data, for example, this was done by excluding records with certain keywords (e.g. “WASTEWATER”) in the site name or site description fields. Records with various map datums (NAD27, NAD83, WGS84) were included without correction; differences among these three major datums are generally less than a few hundred meters, which is an acceptable degree of positional error given the intended final resolution of the interpolated data layers. In any case, most records did not include map datum information, although records which specified unusual or unrecognised map datums were excluded. Data were inspected for clearly incorrect positions (e.g., points plotting outside of the relevant state, province, territory, or points plotting in the ocean); these were corrected where possible. Records that lacked critical metadata (i.e., coordinates, date, etc.), had obvious position or date errors that could not be easily rectified, were flagged at the source with quality control concerns, or had impossible (e.g., negative) measurements, were excluded.

‘Total’ and ‘Dissolved’ calcium were the most commonly recorded fractions, but data for other fractions were sometimes provided. Analysis of data from samples where more than one fraction was measured demonstrated strong positive correlations with slope close to 1 among the most commonly measured fractions (Table 2). Consequently, where data for multiple fractions were provided, measurements of ‘Dissolved’ calcium were preferred, but most other fractions were treated as equivalent and used where ‘Dissolved’ data were not provided. Other fractions were occasionally provided, including ‘Filterable’ and ‘Fixed’ calcium; insufficient data were available to compare these with ‘Dissolved’ calcium, and since they were extremely rarely encountered, they were excluded. In any case, large numbers of records did not provide information on the fraction analysed; their removal would have had a highly detrimental impact on the extent of the available data, so they were retained and assumed to be equivalent to ‘Dissolved’. Calcium concentrations were converted to consistent units (mg L^{-1}) and records without units (<0.01% of records) were excluded.

Some records had extremely high calcium concentrations, including some well over 1000 mg L^{-1} ; these values were generally considered unfeasible, as freshwater calcium concentrations rarely exceed 450 mg L^{-1} and are typically much lower¹²⁸. Anomalously high calcium concentrations may result from inclusion of inappropriate sample types (e.g., contaminated water, industrial effluents, marine samples), equipment malfunction, and data entry errors. A cut-off of 500 mg L^{-1} was therefore set and all records with higher calcium concentrations were excluded; this represented <0.2% of all records. The only exceptions to this rule were samples from the Pecos and Wichita River systems in Texas; calcium concentrations above 500 mg L^{-1} are not unusual in this area^{129,130}, and removing all such records left a notable gap in spatial coverage in an area with already sparse data coverage. Instead, all records above 500 mg L^{-1} for this area were set to 500 mg L^{-1} to maintain consistency with the rest of the data, while avoiding the loss of spatial coverage. For records with calcium concentrations below 0.05 mg L^{-1} (a common detection limit), one of two approaches was taken. Where records were flagged as being ‘below detection limit’, or where an explicit detection limit was given for values below 0.05 mg L^{-1} , records were set to 0.05 mg L^{-1} for consistency across the dataset (<0.005% of all records). Other records with calcium concentrations less than 0.05 mg L^{-1} were excluded (<0.05% of all records). For pH data, records with values lower than 2.5 or above 12.5 were excluded, although for most sources all records fell within this range.

Duplicate data (duplicate records present in individual data sources, presence of the same data in multiple sources) and pseudo-duplicate data (lab and field replicates, samples collected simultaneously from different depths at a location) were handled by calculating an average (median) using all records for each site on each date. For each variable, these site-date medians were then used to calculate the following summary statistics for each site across all dates: mean, standard deviation, 25th percentile, 50th percentile (median), 75th percentile, minimum, maximum (all of these summary statistics are included in the shared databases, see Data Records

Type	Source	Spatial coverage	Temporal coverage	n Sites	
				Ca	pH
FAD	Environment and Climate Change Canada ²⁸	Canada, AK	2000–2019	282	279
FAD	Environment and Climate Change Canada ²⁹	ON	2000–2018	4172	4357
FAD	Environment and Climate Change Canada ³⁰	Canada	2002–2018	—	267
FAD	USGS ³¹	AK, YT	2009–2014	85	92
PDR	Atlantic DataStream ^{41–93}	NB, NL, NS, PE	2000–2020	555	2214
PDR	Mackenzie DataStream ^{52,94–115}	AB, BC, NT, YT	2000–2018	480	707
PDR	Water Quality Portal ¹²⁵	USA	2000–2021	76013	188951
PTA‡	Government of Alberta ³²	AB	2000–2020	254	302
PTA	Government of British Columbia ³³	BC	2000–2020	2903	3033
PTA†	Government of Manitoba	MB	2011–2018	262	302
PTA	Government of New Brunswick ³⁴	NB	2000–2020	635	663
PTA	Government of Newfoundland and Labrador ³⁵	NL	2019–2020	—	29
PTA	Government of Nova Scotia ³⁶	NS	2002–2017	—	5
PTA†	Government of the Northwest Territories	NT	1982–2021	97	97
PTA†	Natural Resources and Forestry Ontario	ON	2008–2017	1327	1346
PTA	Environment, Conservation and Parks Ontario ³⁷	ON	2000–2019	541	594
PTA	Government of Prince Edward Island ³⁸	PE	2001–2020	72	219
PTA‡	Gouvernement du Québec ³⁹	QC	2005–2020	1065	652
PTA‡	Water Security Agency, Saskatchewan ⁴⁰	SK	2010–2020	682	476
PTA†	Government of Yukon, Environment Department	YT	2010–2021	325	343
PRP	Antoniades <i>et al.</i> ^{116,117}	NT, NU	1996–2000	86	86
PRP	Filazzola <i>et al.</i> ¹¹⁸	North of 50°N	1998–2016	—	148
PRP	Joynt and Wolfe ¹¹⁹	NU	1995	—	56
PRP	Michelutti <i>et al.</i> ^{120,121}	NT, NU	1997–1998	37	37
PRP	Pienitz <i>et al.</i> ¹²³	NT	1991	24	24
PRP	Ruhland <i>et al.</i> ¹²⁴	NT, NU	1996, 1998	53	53
OT	2012 Risk Assessment ¹⁵ , 2004 Acid Deposition Assessment ¹²²	AB, BC, MB, NB, NL, NS, ON, QC, SK	1983–2011	8703	4088
OT†	Conseil de gouvernance de l'eau des bassins versant de la rivière Saint-François (and partners)	QC	2018–2020	116	418
OT†	Kivalliq Inuit Association, Crown-Indigenous Relations and Northern Affairs Canada	NU	2004–2020	56	54

Table 1. Summary of data sources used. FAD = Federal Agency Database (Canada/US); PDR = Public Data Repository; PTA = Provincial or Territorial Agency (Canada); PRP = Peer-Reviewed Publication; OT = Other Organisation or Data Source. † data provided directly by agency contacts, not publicly accessible; ‡ includes data provided by agency contacts and publicly accessible agency databases; all other data were obtained from publicly accessible sources.

section). For spatial interpolation, the median value for each site was used, since this measure is comparatively robust to outliers. Medians for each site were converted to spatial data and reprojected into the North America Albers Equal Area Conic projection, using the *sf* package¹³¹.

Spatial interpolation methods. To select the approach used to generate the interpolated data layers, three interpolation methods were compared (Table 3): nearest neighbour (NN), inverse distance weighting (IDW), and ordinary kriging (OK). NN is the simplest method, providing a baseline against which the more advanced methods can be compared; each point for which an interpolated value was required was assigned the value from the closest available data point. IDW uses a combination of values from multiple data points, weighted by distance. For IDW, arbitrary or 'standard' values for *nmax* (the maximum number of points to be considered when predicting a value for a specific grid cell) and *idp* (the inverse distance power parameter, which controls how the weighting of data points varies with distance) are often used¹³². In this case, however, the *optim* function was used to find values of *idp* and *nmax* for the calcium and pH data which minimised two different error metrics, root-mean-square error (RMSE) and mean absolute error (MAE), during preliminary 5-fold cross-validation (Table 4). OK is a geostatistical technique, which uses a fitted model of the spatial autocorrelation among data points (a 'semi-variogram' or 'variogram') to derive the weights used for the interpolation of values to each grid cell. OK often generates superior results to IDW¹³³, but this is not always the case^{132,134}. An additional advantage of OK is that it generates a measure of statistical uncertainty (Kriging variance) for each interpolated value; this is not typically provided by other methods. Kriging variance is influenced by the distances from the interpolated points to locations with data, and by the spatial covariance relationship determined by the fitted variogram; greater variance indicates greater distance from measured values and thus greater uncertainty in the interpolated values. Variograms for each variable were fitted using the *automap* package¹³⁵, which automatically selects

Comparison	n	ρ	p-value	Slope
“Dissolved” vs “Total”	139453	0.977	<0.001	1.019
“Dissolved” vs “Total Recoverable”	17791	0.982	<0.001	1.019
“Dissolved” vs “Recoverable”	9513	0.967	<0.001	1.023
“Dissolved” vs “Extractable”	490	0.947	<0.001	0.969
“Dissolved” vs “Soluble”	267	0.977	<0.001	0.998

Table 2. Comparison of “Dissolved” calcium versus different calcium fractions measured for samples where measurements for more than one fraction were supplied in the source data. ρ = Pearson’s correlation coefficient.

Interpolation Method	Calcium	pH
Nearest Neighbour (NN)	NN interpolation uses only data from nearest point ($nmax = 1$)	
Inverse Distance Weighting, Optimised for RMSE (IDW-OR)	$nmax = 15, idp = 1.2$	$nmax = 19, idp = 1$
Inverse Distance Weighting, Optimised for MAE (IDW-OM)	$nmax = 14, idp = 2.2$	$nmax = 16, idp = 1.1$
Ordinary Kriging (OK)	Model: Matern/Stein	Model: Matern/Stein
	Nugget: 700	Nugget: 0.4
	Sill: 3413	Sill: 0.86
Ordinary Kriging with zero nugget (OK-ZN)	Model: Matern/Stein	Model: Matern/Stein
	Sill: 2076	Sill: 0.92

Table 3. List of Interpolation methods used, including parameter values (where applicable).

Metric	Formula	Interpretation
Correlation coefficient, r	Correlation between Z_i' and Z_i	High correlation between observed and interpolated values suggests accurate interpolation
Mean Absolute Error (MAE)	$\frac{\sum_{i=1}^n Z_i' - Z_i }{n}$	A measure of the average magnitude of prediction errors
Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_{i=1}^n (Z_i' - Z_i)^2}{n}}$	Another measure of the average prediction error. Compared to MAE, RMSE is more sensitive to large outliers.
Mean Bias Error (MBE)	$\frac{\sum_{i=1}^n Z_i' - Z_i}{n}$	By including the sign of individual residuals, MBE shows whether (on average) the interpolation tends to over- or under-predict
Median Symmetric Accuracy (MSA)	$\exp(\text{Median}(\log_e(Z_i/Z_i'))) - 1$	Measure of ‘typical’ proportional error, symmetric and robust to outliers

Table 4. Error metrics considered for comparison of interpolation methods. Z_i is the observed value at a point, Z_i' is the value interpolated for that point during LOOCV.

relevant models and parameter values that best fit the empirical variogram (Table 3), although constraints can be applied to the process. In this case, variograms were fitted with and without a fixed ‘nugget’ of zero, since manual setting of this parameter can sometimes be advantageous¹³⁶. In all cases, OK was restricted to a $nmax$ of 100 and $nmin$ (the minimum number of data points to consider) of 15; changes to these numbers had little to no impact on the error metrics obtained during preliminary 5-fold cross-validation. Spatial models for all interpolation methods were fit using the *gstat* package¹³⁷.

Leave-one-out cross-validation (LOOCV) was performed for each method to compare their predictive accuracies. This technique drops an individual point from the dataset and then uses the remaining data to interpolate a value for the location of the dropped data; this is repeated for all available data points. The interpolated values for each point were compared to the real measured values and used to calculate multiple performance metrics (Table 4): the correlation coefficient r ; three absolute error measures, RMSE, MAE, and the mean bias error (MBE); and a measure of relative error, the median symmetric accuracy¹³⁸ (MSA). The interpolation methods were compared by considering their scores in these metrics. Initially, the intention was to use an ‘objective’ function¹³⁶ to integrate these metrics into a single performance score. However, this was not necessary, since for both calcium concentration and pH one method had the best scores in all key metrics (see Technical Validation). Since predictive accuracy of any interpolation method can vary spatially^{132,136}, error metrics were also calculated using data from each individual province, territory and state.

On the basis of these comparisons a final interpolation method was selected for each variable. Calcium and pH values were then interpolated onto a grid with a cell size of $10 \times 10 \text{ km}^2$ using the *gstat::predict* function and the resulting grids were converted to raster format¹³⁹. Interpolated rasters were masked using outlines of Canada and the USA from the *rnaturalearth* package¹⁴⁰. Rasters of Kriging variance for each variable were also generated at the same resolution.

	File name	Description
Rasters: Ca	calcium-KR-97648-median_10km_ZN_masked.tif	interpolation, masked
	calcium-KR-97648-median_10km_ZN_variance_masked.tif	kriging variance, masked
	calcium-KR-97648-median-10km-ZN.tif	interpolation, unmasked
	calcium-KR-97648-median-10km-ZN_variance.tif	kriging variance, unmasked
Rasters: pH	ph-KR-208784-median_10km_ZN_masked.tif	interpolation, masked
	ph-KR-208784-median_10km_ZN_variance_masked.tif	kriging variance, masked
	ph-KR-208784-median_10km_ZN.tif	interpolation, unmasked
	ph-KR-208784-median_10km_ZN_variance.tif	kriging variance, unmasked
Data: Ca	sites_calcium_85747.csv	site summary data
	metadata_sites_calcium_85747.csv	
	site-date_calcium_985955.csv	summary data for all dates with records at all sites
	metadata_site-date_calcium_985955.csv	
Data: pH	sites_ph_201183.csv	site summary data
	metadata_sites_ph_201183.csv	
	site-date_ph_3525217.csv	summary data for all dates with records at all sites
	metadata_site-date_ph_3525217.csv	
Other	source_ids.csv	list of data source IDs

Table 5. List of data provided in the associated data repository¹⁴¹. Masked rasters have been masked using country outlines such that values are only given for land area, and are therefore ‘ready-to-use’. Note that databases (.csv files) only include data from shareable (open source) providers.

Data Records

Project data are available at Data Dryad¹⁴¹. The data provided include the final interpolated rasters (and kriging variance rasters) for calcium and pH, the point data used for the interpolations (summary statistics for each site) and the underlying data for each site on each date (Table 5).

Rasters were generated by Ordinary Kriging with a pre-defined zero nugget: for both variables this was the best-performing method (see Technical Validation). All rasters use the North America Albers Equal Area projection (ESRI:102008), have a resolution of 10×10 km, and have been provided in geotiff format. For each interpolation, the associated kriging variance rasters have been provided; these can be used to identify areas of higher uncertainty resulting from low availability of water quality data. Rasters are provided both ‘masked’ (using country outlines for the USA and Canada from the *rnaturalearth* package¹⁴⁰ such that values are only provided for land area) and ‘unmasked’. The latter allow users to use their own territorial outlines for masking, to resample the rasters at different resolutions, or to reproject the rasters (for example into a latitude-longitude projection). It is advisable to perform these latter operations prior to masking the rasters with territory outlines. Some example R scripts to facilitate masking and reprojection can be found in the associated GitHub repository¹⁴² (see Code Availability, below).

The ‘sites’ files contain the site data used for the interpolations (Table 5). This includes the following summary statistics for each site: median (used for the interpolations), number of dates with data, total number of records included, mean, standard deviation, minimum, maximum, 25th percentile, 75th percentile. Information on data sources and years with data is also included. The ‘site-date’ files contain summary data for each site on each date with available data: the median, mean, standard deviation, minimum, maximum, 25th percentile, 75th percentile, and number of records. Data sharing agreements with some organisations do not permit open sharing of their data and thus these records are not included in the databases (11,901 sites for calcium, 7601 sites for pH). For a small number of sites for which both public and proprietary data were available (calcium: 383 sites, pH: 61 sites), summary statistics have been recalculated using only public data; consequently, the values provided may not exactly match those used for the interpolations. The associated metadata files include full information on the contents of each of the data files. Finally, the *source_ids.csv* file includes identifying information for the data sources included in the databases.

There are obvious spatial patterns in freshwater calcium concentrations across the continent (Fig. 1), reflecting the relationship with the chemical composition of the underlying bedrock. These include large areas of comparatively low calcium on the east and west coasts of Canada and the USA, and a large area corresponding to the Canadian Shield geological region. Calcium concentrations are comparatively high (30 mg L⁻¹ or greater) across a continuous broad area running from the southern United States up to Yukon and Alaska. Areas of high and low calcium tend to correspond with areas of high and low pH (Fig. 2).

Technical Validation

Calcium. The final calcium database used for the interpolations included records for 97,648 sites; the publicly shareable dataset includes 85,747 sites. Median calcium concentrations for individual sites ranged from 0.06 to 500 mg L⁻¹, but 95% of sites had median calcium concentrations of 115 mg L⁻¹ or lower (Fig. 3). The highest concentrations of sites were mostly in the eastern United States and parts of southern Ontario, Quebec, and New Brunswick, while coverage was lowest in Alaska, the Canadian territories (Yukon, Northwest Territories, and Nunavut), and parts of northern Quebec (Fig. 4a). The majority of sites (56%) were sampled multiple times;

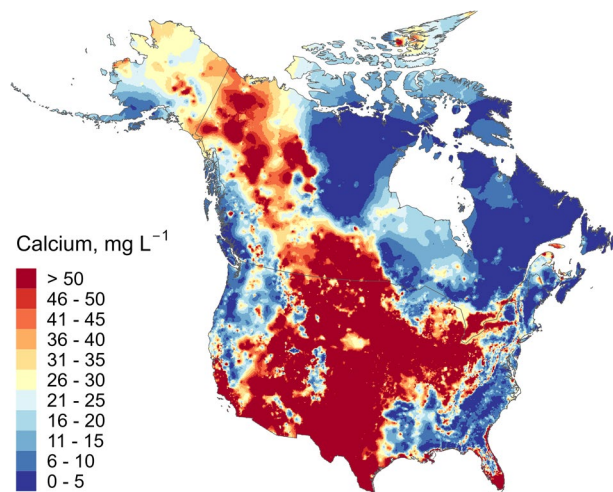


Fig. 1 Interpolated freshwater calcium concentration map for Canada and the USA, generated using zero-nugget Kriging interpolation.

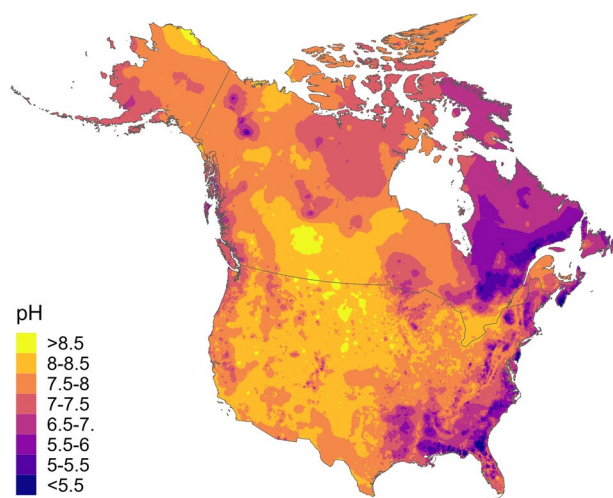


Fig. 2 Interpolated freshwater pH map for Canada and the USA, generated using zero-nugget Kriging interpolation.

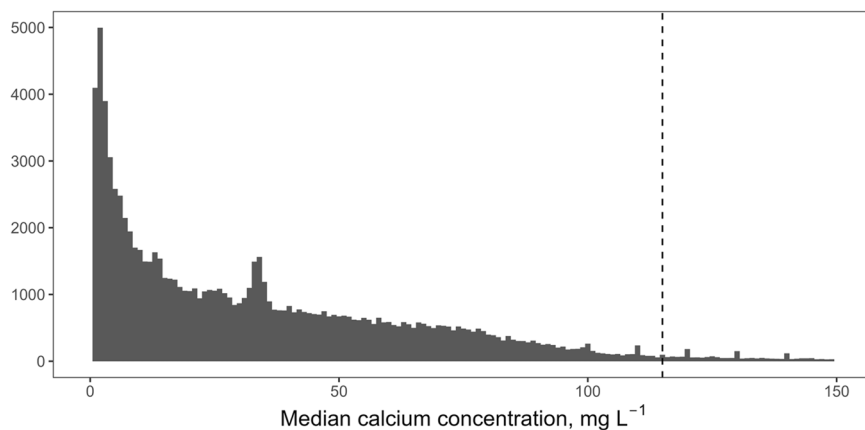


Fig. 3 Frequency distribution of site median calcium concentrations. Dashed line marks the 95th percentile. X-axis is truncated at 150 mg L⁻¹; a small proportion of sites (~3%) had higher values.

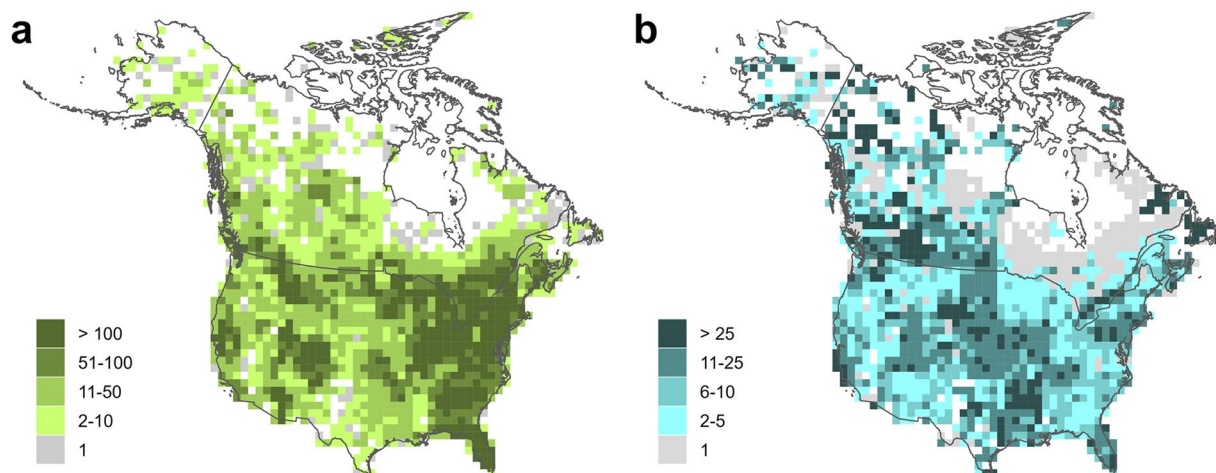


Fig. 4 Calcium concentration data coverage. **(a)** Sites per 10,000 km² **(b)** Sampling intensity (mean dates sampled per site per 10,000 km²).

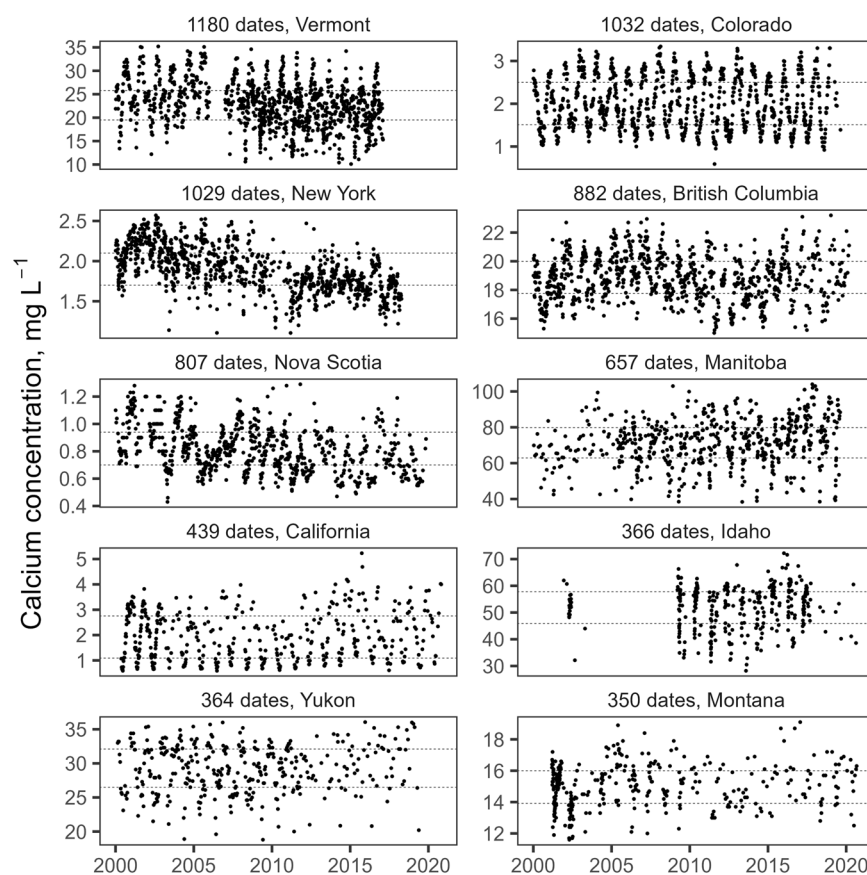


Fig. 5 Temporal variation in dissolved calcium concentration for ten sites, selected (from among the 100 sites with data for the most dates) to have the longest temporal coverage and to come from 10 different administrative regions; source region given for each plot, along with number of dates with data. Seasonal fluctuations are evident for most of the sites, but are small compared to spatial variation across Canada and the USA. Dotted lines mark the interquartile range for each site. Outliers have been removed for presentation, but were not excluded from calculation of site statistics. Note the differences in vertical scales for individual plots.

individual sites were sampled from 1 to over 1000 times (median dates sampled = 2, mean dates sampled = 10.6). In most areas sites were, on average, sampled at least twice; however, there were areas of northern Ontario and Quebec where only a single data point was provided for most sites (Fig. 4b). Some of the provided data for this

Method	<i>r</i>	RMSE	MAE	MBE	MSA
NN	0.78	32.77	13.25	0.09	0.21
IDW-OR	0.82	28.03	12.25	0.16	0.23
IDW-OM	0.82	28.63	12.09	0.13	0.21
OK	0.82	28.61	13.39	-0.01	0.30
OK-ZN	0.84	27.06	11.79	0.01	0.22

Table 6. Error Metrics for calcium interpolation methods, generated via leave-one-out cross-validation. The best interpolation methods for each metric are highlighted in bold. Metrics: *r* = correlation coefficient, RMSE = Root Mean Square Error, MAE = Mean Absolute Error, MBE = Mean Bias Error, MSA = Median Symmetric Accuracy. Methods: NN = nearest neighbour; IDW-OR = inverse distance weighting, RMSE-optimised; IDW-OM = inverse distance weighting, MAE-optimised; OK = ordinary kriging, unconstrained nugget; OK-ZN = ordinary kriging, zero nugget.

Province / Territory	Calcium		pH	
	n	Best	n	Best
Alberta	841	IDW-OR (OK-ZN)	475	IDW-OM
British Columbia	3384	OK-ZN	3154	OK-ZN
Manitoba	899	OK-ZN	814	OK-ZN (IDW-OM)
New Brunswick	936	OK-ZN	1221	OK-ZN
Newfoundland and Labrador	117	OK-ZN	330	OK-ZN
Nova Scotia	270	OK-ZN	657	IDW-OM
Northwest Territories	428	OK-ZN	504	IDW-OR/IDW-OM
Nunavut	164	OK	313	OK-ZN
Ontario	7819	OK-ZN	6537	OK-ZN
Prince Edward Island	75	OK-ZN	478	OK-ZN
Quebec	4619	OK-ZN	4259	OK-ZN
Saskatchewan	1657	OK-ZN	630	OK-ZN
Yukon Territory	365	OK-ZN	398	IDW-OM

Table 7. Best-performing calcium and pH interpolation methods for each Province / Territory (Canada), according to LOOCV error score and correlation between observed and predicted values. n = number of sites. Methods: NN = Nearest Neighbour; IDW-OR = Inverse distance weighting, RMSE-optimised; IDW-OM = Inverse distance weighting, MAE-optimised; OK = Kriging; OK-ZN = Kriging, zero nugget. Where two methods are presented separated by '/' this indicates that the two methods performed equally well. Methods presented in parentheses were those with the highest correlation coefficient, *r*, in cases where this was not the same method with the best scores in the other error metrics.

area were already temporally-averaged values, so this does not mean that all data points for these areas were based on single measurements. Temporal variation in measurements from individual sites is to be expected as a result of measurement error and temporal change, such as seasonal fluctuations in calcium concentrations (Fig. 5). However, the scale of temporal variation at individual sites was generally smaller than the spatial variation among sites. The interquartile range for temporally-averaged calcium concentrations across all sites was 48.5 mg L⁻¹, while the median interquartile range for calcium measurements at individual sites was 5 mg L⁻¹, and 75% of sites had an interquartile range of 12.5 or less.

For the interpolation of calcium concentrations, the zero-nugget Kriging method (OK-ZN) had the highest *r* value and the lowest error metrics (excluding the proportional error, MSA, which was very similar to the lowest value); in particular, the bias error (MBE) was lower than the other methods (Table 6). The IDW interpolations, however, were not substantially worse. At the province, territory, and state level, the outcome was mostly similar: OK-ZN was the best or joint-best method in 53 out of 62 cases (Tables 7, 8); OK was slightly superior for four areas, the IDW methods (IDW-OR and IDW-OM) were superior for 4 areas, and in one US state NN was the best method. There appeared to be no tendency for the best approach to vary with number of data points in each area; OK-ZN was generally superior for states, provinces and territories with low (e.g., Mississippi, n = 75) and high (e.g., Florida, n = 13,591) numbers of data points. Consequently, the zero-nugget kriging interpolation (OK-ZN) was selected as the preferred interpolation method.

pH. The final pH database used for the interpolations included records for 208,784 sites; the publicly shareable dataset includes 201,183 sites. The median pH across all sampled sites was 7.9, and 95% of sites had a median pH between 5.4 and 8.74 (Fig. 6). Density of sites was high across much of the USA, with a considerably higher number of sites than for calcium (Fig. 7a). Coverage tended to be sparser for Canada, with some areas, such as northern Saskatchewan and northern Manitoba, having fewer sites with available data compared to calcium. Compared to the calcium data, a greater proportion (67%) of sites had data from more than one date, and sites tended to have data from a greater number of sampling dates (median dates sampled = 4, mean dates sampled = 17.2). However,

State	Calcium		pH	
	n	Best	n	Best
Alaska	396	OK-ZN	1020	OK-ZN
Alabama	1274	OK-ZN	2635	IDW-OM/OK-ZN
Arkansas	848	OK-ZN	1258	IDW-OM
Arizona	2102	OK-ZN	3416	IDW-OM
California	2904	OK-ZN	6451	OK-ZN
Colorado	4355	OK-ZN	6296	OK-ZN
Connecticut	78	NN	168	IDW-OM (NN)
Delaware	75	OK-ZN	559	OK-ZN
Florida	13591	OK-ZN	29471	OK-ZN
Georgia	1954	IDW-OM	3554	OK-ZN (IDW-OM)
Iowa	266	OK	5001	OK-ZN
Idaho	591	OK-ZN	1322	OK-ZN
Illinois	2129	OK-ZN	3406	OK-ZN
Indiana	2260	OK-ZN	8845	IDW-OM (OK-ZN)
Kansas	1198	OK-ZN	1045	OK-ZN
Kentucky	751	OK-ZN	1394	IDW-OM (OK-ZN)
Louisiana	357	OK	935	IDW-OM (OK-ZN)
Massachusetts	344	OK-ZN	1442	OK-ZN
Maryland	579	OK	2942	OK-ZN
Maine	262	OK-ZN	1506	OK-ZN
Michigan	2367	OK-ZN	3046	IDW-OM
Minnesota	1994	OK-ZN	14282	OK-ZN
Missouri	847	OK-ZN	2975	OK-ZN
Mississippi	73	OK-ZN	1408	OK-ZN
Montana	4081	OK-ZN	5860	OK-ZN
North Carolina	938	OK/OK-ZN	2148	OK-ZN
North Dakota	1321	OK-ZN	1856	IDW-OM/OK-ZN
Nebraska	265	OK-ZN	1368	OK-ZN (IDW-OM)
New Hampshire	324	OK-ZN	2792	OK-ZN
New Jersey	1090	OK-ZN	4108	IDW-OM
New Mexico	1100	OK-ZN	1491	OK-ZN
Nevada	836	OK-ZN	1816	OK-ZN
New York	1819	OK-ZN	2611	OK-ZN
Ohio	5836	OK-ZN	6159	OK-ZN
Oklahoma	402	OK-ZN	4261	OK-ZN (IDW-OR)
Oregon	855	OK-ZN	3750	OK-ZN
Pennsylvania	1831	OK-ZN	2927	IDW-OM (OK-ZN)
Rhode Island	26	IDW-OM	59	IDW-OM
South Carolina	1509	IDW-OR	1958	OK-ZN
South Dakota	497	OK-ZN	3200	OK-ZN
Tennessee	795	OK-ZN	5811	OK-ZN
Texas	528	OK-ZN	5077	OK-ZN
Utah	2554	OK-ZN	3641	OK-ZN
Virginia	1665	OK-ZN	8271	OK-ZN
Vermont	1893	OK-ZN	2169	IDW-OM (OK-ZN)
Washington	335	OK-ZN	1876	IDW-OM/OK-ZN
Wisconsin	1952	OK-ZN	7308	OK-ZN
West Virginia	1339	OK-ZN	2553	OK-ZN
Wyoming	684	OK-ZN	1567	OK-ZN

Table 8. Best-performing calcium and pH interpolations for each State (USA). See Table 7 for abbreviations.

there were again areas of Quebec and Ontario where the data tended to be based on single values for each site (Fig. 7b). Temporal fluctuation at individual sites was also evident for pH (Fig. 8). However, the scale of temporal variation for individual sites was again smaller than the spatial variation among sites. The median interquartile range for pH measurements at individual sites was 0.3, with 75% of sites having an interquartile range of less than 0.46; the interquartile range across all sites (spatial variability) was 0.97.

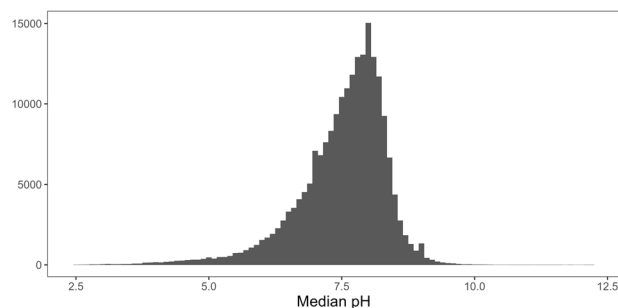


Fig. 6 Frequency distribution of site median pH values.

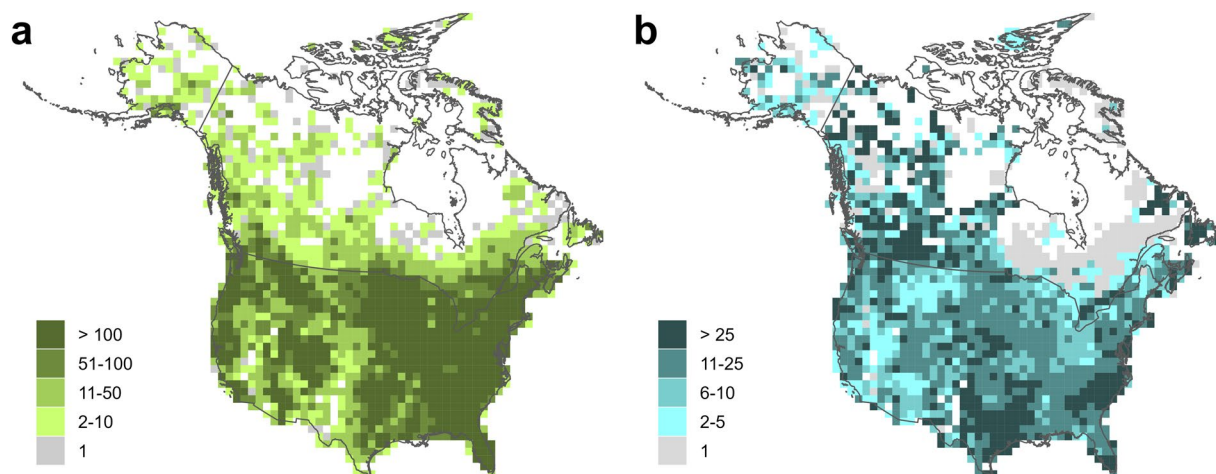


Fig. 7 pH data coverage. (a) Sites per 10,000 km² (b) Sampling intensity (mean dates sampled per site per 10,000 km²).

Error metrics for the pH interpolations were generally very low, including RMSE and MAE; this is to be expected, since the restricted range of feasible pH values makes extremely large errors impossible. While it is not valid to directly compare most metrics between interpolations with different scales and based on different data, it is worth noting that the proportional errors (MSA) were considerably lower for the pH interpolations compared to the calcium interpolations. It is important to be aware, however, that since pH is measured on a logarithmic scale, apparently small differences may have comparatively large physical and chemical implications. There was little variation in the accuracy of the different interpolation methods, with most error metrics being similar for most of the methods (Table 9). However, OK-ZN had the best (or equal-best) scores in every metric excluding the proportional error, which was very close to the lowest value. For individual provinces, territories and states, the situation was similar (Tables 7, 8); OK-ZN was the best or equal-best method in 46 cases, with IDW-OM / IDW-OR being slightly better for the others. Consequently, the zero-nugget kriging interpolation (OK-ZN) was selected as the preferred interpolation method.

Kriging variance maps generated by the selected interpolation methods can be used to identify areas of higher uncertainty in the interpolated values, and maps for the two variables show broadly similar patterns (Fig. 9). Across much of the USA and some of the Canadian provinces, there were high densities of sites (Figs. 4a, 7a); kriging variance was lower in these areas, indicating comparatively lower uncertainty in the interpolated values. Variance, and therefore uncertainty, was highest in the northern areas of the continent, where there were fewer sites with data. Interpolated values in such areas should be treated with some caution, since they are more distant from locations with measured values. These areas could be prioritised for future sampling if more certainty is required for estimates of pH and calcium concentrations.

Usage Notes

Example application – invasive species risk assessment for dreissenid mussels. To illustrate the advantages of high-resolution calcium and pH data, Wilcox *et al.*¹⁴³ performed a continental-scale risk assessment for two species of invasive, freshwater dreissenid mussels using the new data layers. The two species, the zebra mussel *Dreissena polymorpha* and the quagga mussel *D. rostriformis bugensis*, have significant ecological and economic impacts on freshwater ecosystems in North America¹⁴⁴, but can only survive, grow, and reproduce in waters with sufficiently high concentrations of dissolved calcium and within a particular pH range⁵.

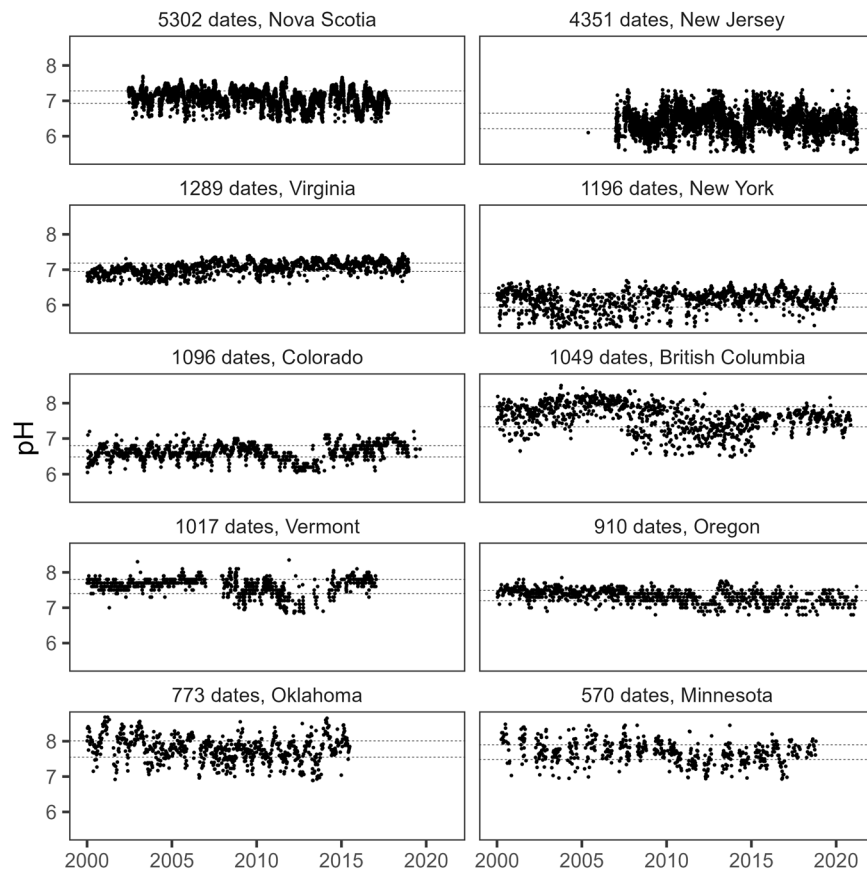


Fig. 8 Temporal trends in pH for ten sites, selected (from among the 100 sites with data for the most dates) to have the longest temporal coverage and to come from 10 different administrative regions; source region given for each plot, along with number of dates with data. Dotted lines mark the interquartile range for each site. Outliers have been removed for presentation, but were not excluded from calculation of site statistics.

Method	r	RMSE	MAE	MBE	MSA
NN	0.78	0.55	0.35	0.007	0.029
IDW-OR	0.84	0.46	0.30	0.007	0.027
IDW-OM	0.84	0.46	0.30	0.008	0.027
OK	0.81	0.49	0.33	0.001	0.031
OK-ZN	0.84	0.45	0.30	0.001	0.028

Table 9. Error Metrics for pH interpolation methods, generated via leave-one-out cross-validation. The best interpolation methods for each metric are highlighted in bold. Metrics: RMSE = Root Mean Square Error, MAE = Mean Absolute Error, MBE = Mean Bias Error, MSA = Median Symmetric Accuracy. Methods: NN = nearest neighbour; IDW-OR = inverse distance weighting, RMSE-optimised; IDW-OM = inverse distance weighting, MAE-optimised; OK = ordinary kriging, unconstrained nugget; OK-ZN = ordinary kriging, zero nugget.

Due to the lack of high-resolution calcium and pH data, previous risk assessments^{13,15} for these species have been limited to ecoregion- or sub-drainage-level resolution, and have not included Alaska and large areas of Canada (the Maritime provinces, Newfoundland and Labrador, and the Arctic). By combining the new calcium and pH data layers with additional high-resolution bioclimatic variables (e.g. temperature) from WorldClim¹⁴⁵, Wilcox *et al.*¹⁴³ were able to model habitat suitability for both dreissenid species for the entire extent of Canada and the continental USA, assess the importance of calcium and pH relative to additional bioclimatic drivers of mussel distributions, and calculate the relative risk of invasion for every Canadian province and territory at a 10 km² resolution.

Limitations. ‘Big data’ approaches can be a useful tool for water quality projects, but are not without limitations¹⁴⁶, and the aggregation of many different data sources, with highly variable quality control standards, necessitates some care in their use. Given the extremely large number of data points involved, inspection of individual data records was not possible. Therefore, some of the data filtering and cleaning approaches may have resulted in the exclusion of some valid records. On the other hand, it is likely that some low-quality data remain in the final database used

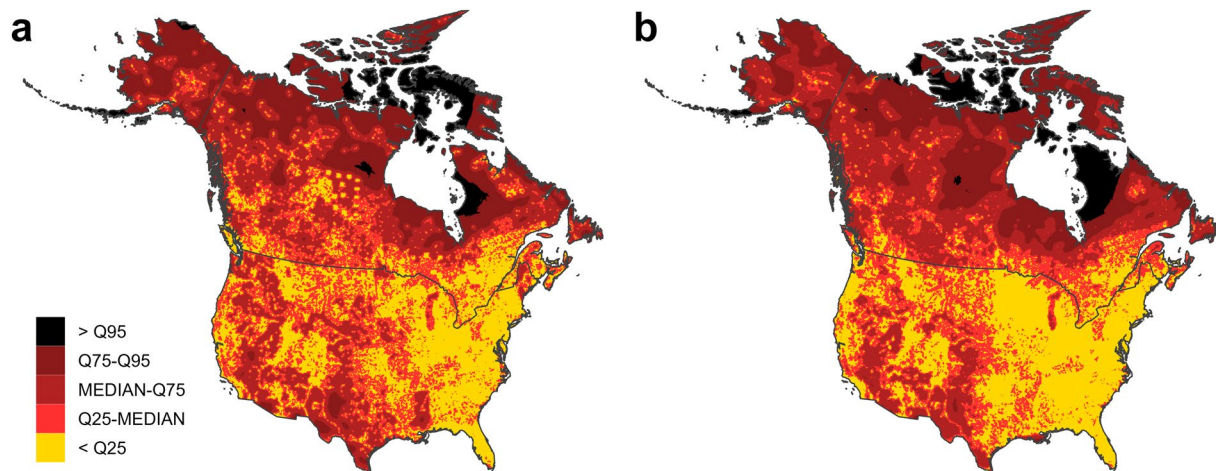


Fig. 9 Kriging variance maps for (a) calcium and (b) pH interpolations. Q25, Q75, and Q95 are 25th, 75th, and 95th percentile values.

for the interpolations. For example, while certain sources provided enough information to be able to quickly screen out inappropriate sample types, most sources did not. Points with large or obvious errors in their values or positions were easy to identify, and therefore to remove or correct. Records with incorrect or inaccurate – but plausible – values or positions were effectively impossible to identify and remove. The large number of records used for the final calcium and pH databases should, however, minimise the impact of these types of error on the final interpolations.

There are also a few limitations to using spatial interpolation to create such large-scale maps of water quality variables. Calcium concentrations and pH are primarily driven by the underlying geology¹⁴⁷; transitions between underlying rock types can be relatively well delineated, and interpolation across such boundaries may produce results that do not reflect reality. This problem is likely to be minimised in areas with a high density of data points but may be important in data-poor regions. For example, there are relatively large swathes of northern Quebec and Arctic Canada for which no calcium or pH data were available. This may not be problematic in some contexts; for example, in the case of invasive species risk assessment, there are other factors (low temperature, remote location) that may make these areas low risk for many non-native organisms. Geological proxies could be used to predict calcium concentration in locations with no water quality data¹⁴⁸, but this requires detailed geological information and validated mechanistic models and does not account for effects of plant cover and land use, which influence water chemistry^{4,149}. Despite these limitations, geological data could be used to help improve predictions in areas with lower data coverage, or higher uncertainty, via co-kriging, which allows relationships with additional variables to be used during the interpolation process^{150,151}. Alternatively, a range of machine learning approaches are available which are also able to use additional information, such as geological data and other environmental covariates; these methods can perform better than traditional geostatistical methods for generating spatial interpolations^{152,153}, particularly when the density of data points for the primary variable of interest is low¹⁵⁴. However, they do not always generate more accurate interpolations; a combined approach, which uses an ensemble of outputs from different interpolation methods with spatially-varying weightings dependant on density of available data, may result in better overall accuracy¹³⁶. Such exercises are good candidates for future improvement to these data layers.

Code availability

The associated GitHub repository¹⁴² (https://github.com/andrew-guerin/water_quality_interpolations) contains code used to perform the final interpolations, scripts for reprojection and resampling of rasters, copies of the interpolated data layers, copies of the calcium and pH databases used for the interpolations (excluding proprietary data from third parties, which cannot be publicly shared under existing data agreements), and Shiny app scripts for interactive maps which show the distribution of data points, with summary data (where these can be shared). Please note that, as a result of the large number of data points included, the Shiny maps may take a few moments to load.

Received: 21 July 2023; Accepted: 20 March 2024;

Published online: 11 April 2024

References

- Niño-García, J. P., Ruiz-González, C. & del Giorgio, P. A. Interactions between hydrology and water chemistry shape bacterioplankton biogeography across boreal freshwater networks. *ISME J* **10**, 1755–1766, <https://doi.org/10.1038/ismej.2015.226> (2016).
- Ortiz-Álvarez, R., Cáliz, J., Camarero, L. & Casamayor, E. O. Regional community assembly drivers and microbial environmental sources shaping bacterioplankton in an alpine lacustrine district (Pyrenees, Spain). *Environmental Microbiology* **22**, 297–309, <https://doi.org/10.1111/1462-2920.14848> (2020).
- Juggins, S., Kelly, M., Allott, T., Kelly-Quinn, M. & Monteith, D. A Water Framework Directive-compatible metric for assessing acidification in UK and Irish rivers using diatoms. *Sci. Total. Environ.* **568**, 671–678, <https://doi.org/10.1016/j.scitotenv.2016.02.163> (2016).

4. Jüttner, I. *et al.* Assessing the impact of land use and liming on stream quality, diatom assemblages and juvenile salmon in Wales, United Kingdom. *Ecol. Indic.* **121**, 107057, <https://doi.org/10.1016/j.ecolind.2020.107057> (2021).
5. Hincks, S. S. & Mackie, G. L. Effects of pH, calcium, alkalinity, hardness, and chlorophyll on the survival, growth, and reproductive success of zebra mussel (*Dreissena polymorpha*) in Ontario lakes. *Can. J. Fish. Aquat. Sci.* **54**, 2049–2057, <https://doi.org/10.1139/f97-114> (1997).
6. Ramaekers, L., Vanschoenwinkel, B., Brendonck, L. & Pincheel, T. Elevated dissolved carbon dioxide and associated acidification delays maturation and decreases calcification and survival in the freshwater crustacean *Daphnia magna*. *Limnol. Oceanogr.*, <https://doi.org/10.1002/lno.12372> (2023).
7. Messina, S., Costantini, D. & Eens, M. Impacts of rising temperatures and water acidification on the oxidative status and immune system of aquatic ectothermic vertebrates: A meta-analysis. *Sci. Total. Environ.* **868**, 161580, <https://doi.org/10.1016/j.scitotenv.2023.161580> (2023).
8. Sayer, M. D. J., Reader, J. P. & Dalziel, T. R. K. Freshwater acidification: effects on the early life stages of fish. *Rev. Fish. Biol. Fisheries* **3**, 95–132, <https://doi.org/10.1007/BF00045228> (1993).
9. Townsend, C. R., Hildrew, A. G. & Francis, J. Community structure in some southern English streams: the influence of physicochemical factors. *Freshwater Biol.* **13**, 521–544, <https://doi.org/10.1111/j.1365-2427.1983.tb00011.x> (1983).
10. Hasler, C. T. *et al.* Biological consequences of weak acidification caused by elevated carbon dioxide in freshwater ecosystems. *Hydrobiologia* **806**, 1–12, <https://doi.org/10.1007/s10750-017-3332-y> (2018).
11. Jia, C. *et al.* Effect of complex hydraulic variables and physicochemical factors on freshwater mussel density in the largest floodplain lake, China. *Ecol. Process.* **12**, 15, <https://doi.org/10.1186/s13717-023-00427-y> (2023).
12. Roy, S., Ray, S. & Saikia, S. K. Indicator environmental variables in regulating the distribution patterns of small freshwater fish *Amblypharyngodon mola* in India and Bangladesh. *Ecol. Indic.* **120**, 106906, <https://doi.org/10.1016/j.ecolind.2020.106906> (2021).
13. Whittier, T. R., Ringold, P. L., Herlihy, A. T. & Pierson, S. M. A calcium-based invasion risk assessment for zebra and quagga mussels (*Dreissena* spp.). *Front. Ecol. Environ.* **6**, 180–184, <https://doi.org/10.1890/070073> (2008).
14. Wells, S. W., Counihan, T. D., Puls, A., Sytsma, M. & Adair, B. Prioritizing zebra and quagga mussel monitoring in the Columbia River Basin. *Center for Lakes and Reservoirs Publications and Presentations* **10**, (2011).
15. Therriault, T. W., Weise, A. M., Higgins, S. N., Guo, Y. & Duhaime, J. Risk assessment for three dreissenid mussels (*Dreissena polymorpha*, *Dreissena rostriformis bugensis*, and *Mytilopsis leucophaea*) in Canadian freshwater ecosystems. *DFO Can. Sci. Advis. Sec. Res. Doc.* **2012/174**, 88 pp. (2013).
16. Clair, T. A., Witteman, J. P. & Whitlow, S. H. Acid precipitation sensitivity of Canada's Atlantic provinces. *Environment Canada Technical Bulletin* **124**, v + 12 pp. (1982).
17. Krzyzanowski, J. & Innes, J. L. Back to the basics – Estimating the sensitivity of freshwater to acidification using traditional approaches. *J. Environ. Manage.* **91**, 1227–1236, <https://doi.org/10.1016/j.jenvman.2010.01.013> (2010).
18. Schneider, S. C., Kahlert, M. & Kelly, M. G. Interactions between pH and nutrients on benthic algae in streams and consequences for ecological status assessment and species richness patterns. *Science of The Total Environment* **444**, 73–84, <https://doi.org/10.1016/j.scitotenv.2012.11.034> (2013).
19. Finlay, K. & Bogard, M. J. pH of Inland Waters. in *Encyclopedia of Inland Waters (Second Edition)* (eds. Mehner, T. & Tockner, K.) 112–122, <https://doi.org/10.1016/B978-0-12-819166-8.00045-1> (2022).
20. De Schampelaere, K. A. C., Lofts, S. & Janssen, C. R. Bioavailability models for predicting acute and chronic toxicity of zinc to algae, daphnids, and fish in natural surface waters. *Environ. Toxicol. Chem.* **24**, 1190, <https://doi.org/10.1897/04-229R.1> (2005).
21. Bethke, K., Kropidłowska, K., Stepnowski, P. & Caban, M. Review of warming and acidification effects to the ecotoxicity of pharmaceuticals on aquatic organisms in the era of climate change. *Sci. Total. Environ.* **877**, 162829, <https://doi.org/10.1016/j.scitotenv.2023.162829> (2023).
22. Hart, K. A., Kennedy, G. W. & Sterling, S. M. Distribution, drivers, and threats of aluminum in groundwater in Nova Scotia, Canada. *Water* **13**, 1578, <https://doi.org/10.3390/w13111578> (2021).
23. Khalid, N., Aqeel, M., Noman, A., Khan, S. M. & Akhter, N. Interactions and effects of microplastics with heavy metals in aquatic and terrestrial environments. *Environ. Pollut.* **290**, 118104, <https://doi.org/10.1016/j.envpol.2021.118104> (2021).
24. Jian, M. *et al.* How do microplastics adsorb metals? A preliminary study under simulated wetland conditions. *Chemosphere* **309**, 136547, <https://doi.org/10.1016/j.chemosphere.2022.136547> (2022).
25. Murphy, R. R., Curriero, F. C. & Ball, W. P. Comparison of spatial interpolation methods for water quality evaluation in the Chesapeake Bay. *J. Environ. Eng.* **136**, 160–171, [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000121](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000121) (2010).
26. Harris, P. & Juggins, S. Estimating freshwater acidification critical load exceedance data for Great Britain using space-varying relationship models. *Math. Geosci.* **43**, 265–292, <https://doi.org/10.1007/s11004-011-9331-z> (2011).
27. Dobson, B. *et al.* Predicting catchment suitability for biodiversity at national scales. *Water Res.* **221**, 118764, <https://doi.org/10.1016/j.watres.2022.118764> (2022).
28. Environment and Climate Change Canada. National long-term water quality monitoring data, <https://open.canada.ca/data/en/dataset/67b44816-9764-4609-ace1-68dc1764e9ea>.
29. Environment and Climate Change Canada. Great Lakes water quality monitoring and surveillance data, <https://open.canada.ca/data/en/dataset/cfdafa0c-a644-47cc-ad54-460304facf2e>.
30. Environment and Climate Change Canada. Water Quality in Canadian Rivers, <https://www.canada.ca/en/environment-climate-change/services/environmental-indicators/water-quality-canadian-rivers.html>.
31. Herman-Mercer, N. M. Water-Quality Data from the Yukon River Basin in Alaska and Canada, <https://doi.org/10.5066/F77D2S7B> (2016).
32. Government of Alberta, Ministry of Environment and Protected Areas. Water Quality Data Portal, <https://environment.extranet.gov.ab.ca/apps/WaterQuality/dataportal/>.
33. Government of British Columbia. Environmental Monitoring System, <https://catalogue.data.gov.bc.ca/dataset/bc-environmental-monitoring-system-results>.
34. Government of New Brunswick. Water Quality Data Portals, <https://www2.gnb.ca/content/gnb/en/departments/elg/environment/content/water/content/water-quality-data-portals.html>.
35. Government of Newfoundland and Labrador. Real time water quality monitoring stations, <https://www.gov.nl.ca/ecc/waterres/watermonitoring/rtwq/stations/>.
36. Government of Nova Scotia. Surface water quality monitoring network data, <https://novascotia.ca/nse/surface.water/automatedqualitymonitoringdata.asp>.
37. Ministry of Environment, Conservation and Parks Ontario. Provincial stream water quality monitoring network, <https://data.ontario.ca/dataset/provincial-stream-water-quality-monitoring-network>.
38. Government of Prince Edward Island, Environment, Water and Climate Change. Surface water quality, <https://www.princeedwardisland.ca/en/service/view-surface-water-quality>.
39. Ministère de l'Environnement, de la lutte contre les changements climatiques, de la Faune et des Parcs, Gouvernement du Québec. Duretés médianes des eaux de surface, <https://www.donneesquebec.ca/recherche/fr/dataset/duretes-medianes-des-eaux-de-surface>.
40. Government of Saskatchewan, Water Security Agency. Primary station water quality, <https://waterquality.saskatchewan.ca/PrimaryStation>.

41. Atlantic Coastal Action Program (ACAP) Saint John. ACAP Saint John: Community-Based Water Monitoring Program, DataStream <https://doi.org/10.25976/4MF6-K783> (2023).
42. Banook Area Residents Association (BARA). BARA - Dartmouth, NS - Sawmill River Watershed - Baseline YSI Study, DataStream <https://doi.org/10.25976/8SX1-NP85> (2022).
43. Bedeque Bay Environmental Management Association. Bedeque Bay Environmental Management Association Water Monitoring Program, DataStream <https://doi.org/10.25976/ZRRZ-R515> (2023).
44. Belleisle Watershed Coalition. Belleisle Watershed Coalition Water Quality Monitoring, DataStream <https://doi.org/10.25976/BEQ6-R303> (2021).
45. EOS Eco-Energy. Cape Tormentine Peninsula Watershed Water Quality Monitoring, DataStream <https://doi.org/10.25976/AF39-MV83> (2023).
46. Clean Annapolis River Project. River Guardians Water Quality Monitoring Program, DataStream <https://doi.org/10.25976/YE7N-BD78> (2022).
47. Clean Foundation. Clean Foundation Watershed Restoration Monitoring Data, DataStream <https://doi.org/10.25976/SW2M-0492> (2022).
48. Mi'kmaw Conservation Group. Cornwallis River Watershed Water Quality by Mikmaq Communities and Mikmaq Conservation Group, DataStream <https://doi.org/10.25976/N02Z-MM23> (2022).
49. Eastern Charlotte Waterways Inc. 2018 Freshwater Quality Data from the Outer Bay of Fundy Watershed Complex, DataStream <https://doi.org/10.25976/JBPT-BF54> (2022).
50. Eastern Charlotte Waterways Inc. Freshwater quality data from the Outer Bay of Fundy watershed complex, DataStream <https://doi.org/10.25976/KC1X-BC80> (2022).
51. Northeast Avalon ACAP. Enhancement of an Urban Wetland, Lundrigan's Marsh, DataStream <https://doi.org/10.25976/C4D2-E296> (2022).
52. Water Rangers. Equipping communities in data-deficient areas, DataStream <https://doi.org/10.25976/AQPR-E356> (2023).
53. Coastal Action. Fox Point Lake Water Quality Monitoring Dataset, DataStream <https://doi.org/10.25976/P1WT-1520> (2022).
54. ACAP Humber Arm. Freshwater Quality Monitoring: Bay of Islands and Humber Valley, Newfoundland and Labrador, DataStream <https://doi.org/10.25976/SM6C-FY36> (2023).
55. Hammond River Angling Association. Hammond River Angling Association Water Quality Monitoring Program, DataStream <https://doi.org/10.25976/9XBE-0V09> (2023).
56. Saint Mary's University, Dynamic Environment & Ecosystem Health Research Lab. Historic Gold Mine Tailings Wetland Sites, DataStream <https://doi.org/10.25976/X2E0-7126> (2022).
57. Indian Bay Ecosystem Corporation. Indian Bay Watershed Monitoring Project, DataStream <https://doi.org/10.25976/VAHX-DQ27> (2022).
58. Kelligrews Ecological Enhancement Program. Kelligrews Ecological Enhancement Program (KEEP) Water Quality Monitoring, DataStream <https://doi.org/10.25976/96AH-ZG44> (2022).
59. Coastal Action. LaHave River Water Quality Monitoring, DataStream <https://doi.org/10.25976/R245-Z904> (2022).
60. Manuels River. Manuels River Water Quality Monitoring, DataStream <https://doi.org/10.25976/A5TD-E253> (2022).
61. Environment and Climate Change Canada (ECCC); Parks Canada. Maritime Coastal Basin Long-term Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/6761-TK42> (2023).
62. Meduxnekeag River Association Inc. DataStream Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/Z9P3-B253> (2022).
63. Eastern Charlotte Waterways Inc. *et al.* Monitoring results from select estuaries in the four Atlantic provinces, DataStream <https://doi.org/10.25976/99FY-8J65> (2022).
64. Nashwaak Watershed Association. Nashwaak Watershed Water Quality Data, DataStream <https://doi.org/10.25976/1JAM-Q207> (2023).
65. Environment and Climate Change Canada, Government of Newfoundland Municipal Affairs and Environment, & Parks Canada. Newfoundland and Labrador Long-term Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/6YQ8-QM85> (2023).
66. Environment and Climate Change Canada. North Shore - Gaspé Basin Long-term Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/5XJ3-S893> (2023).
67. Government of Nova Scotia, Department of Fisheries and Aquaculture. NS Government Lake Survey Data, DataStream <https://doi.org/10.25976/R2B8-7966> (2022).
68. Oathill Lake Conservation Society. Oathill Lake Conservation Society Water Monitoring, DataStream <https://doi.org/10.25976/WTT6-6K82> (2023).
69. Kennebecasis Watershed Restoration Committee. Kennebecasis Watershed Restoration Committee Water Quality Monitoring Program, DataStream <https://doi.org/10.25976/H7BC-3Y92> (2022).
70. Passamaquoddy Recognition Group Inc. Peskotomuhkati Nation Coastal Restoration, DataStream <https://doi.org/10.25976/4Y34-RN27> (2023).
71. Petitcodiac Watershed Alliance (PWA) Inc. Petitcodiac Watershed Water Quality Monitoring, DataStream <https://doi.org/10.25976/1D9Y-0A25> (2023).
72. Coastal Action. Petite Riviere Lakes and Headwaters Dataset, DataStream <https://doi.org/10.25976/MB11-3303> (2022).
73. Coastal Action. Petite Riviere Watershed Water Quality Dataset, DataStream <https://doi.org/10.25976/0S2N-0D98> (2022).
74. Pictou County Rivers Association. Pictou County Water Quality data, DataStream <https://doi.org/10.25976/0MG9-ET50> (2023).
75. Government of Prince Edward Island, Environment, Water and Climate Change. Province of Prince Edward Island - Surface Water Quality Monitoring, DataStream <https://doi.org/10.25976/G5S5-YJ38> (2023).
76. Southeastern Anglers Association. Southeastern Anglers Association Water Quality, DataStream <https://doi.org/10.25976/DS9A-KC30> (2023).
77. Sackville Rivers Association. Sackville River Watershed water quality monitoring, DataStream <https://doi.org/10.25976/0YG6-GR73> (2022).
78. Environment and Climate Change Canada. Saint John River and St. Croix River Basin Long-term Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/7JVG-3N77> (2023).
79. Stratford Area Watershed Improvement Group (SAWIG). SAWIG Water Quality Dataset, DataStream <https://doi.org/10.25976/TY8W-FV06> (2023).
80. St. Croix International Waterway Commission. St. Croix International Waterway Commission, DataStream <https://doi.org/10.25976/CXJD-EZ84> (2023).
81. Shediac Bay Watershed Association (SBWA). Shediac Bay Watershed monitoring of the Shediac and Scoudouc rivers, DataStream <https://doi.org/10.25976/TSBM-8561> (2023).
82. Coastal Action. Sherbrooke Lake Water Quality Dataset, DataStream <https://doi.org/10.25976/4294-7804> (2022).
83. Shubenacadie Watershed Environmental Protection Society - Water Resources subcommittee. Shubenacadie Watershed Environmental Protection Society - Soldier and Miller Lakes Monitoring Program (SWEPS), DataStream <https://doi.org/10.25976/3TT9-GH44> (2022).

84. South Shore Watershed Association. South Shore Watershed Association Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/D6TS-A051> (2022).
85. Nova Scotia Environment. Surface Water Quality Monitoring Network Grab Sample Water Quality Data, DataStream <https://doi.org/10.25976/SFMJ-KF56> (2023).
86. EOS Eco-Energy. Tantramar River Watershed Water Quality Monitoring, DataStream <https://doi.org/10.25976/E0VQ-CV76> (2022).
87. Tobique Watershed Association. Tobique River Project, DataStream <https://doi.org/10.25976/FTZA-4B38> (2020).
88. Tusket River Environmental Protection Association. Water Quality data from the Tusket Catchment, DataStream <https://doi.org/10.25976/T5RX-QR11> (2022).
89. Vision H2O. Vision H2O Water Quality Dataset, DataStream <https://doi.org/10.25976/QGWY-A975> (2022).
90. Northeast Avalon ACAP. Water Quality Monitoring of Regional Rivers (Northeast Avalon), DataStream <https://doi.org/10.25976/BV8Z-1Z40> (2022).
91. Woodstock First Nation Climate Monitoring Program. Woodstock First Nation Climate Monitoring Program, DataStream <https://doi.org/10.25976/5FF8-0E09> (2022).
92. Willowbrook Watershed Services. WQ Response to Geologic Variation, DataStream <https://doi.org/10.25976/FBXM-3B32> (2022).
93. Winter River-Tracadie Bay Watershed Association. WRTBWA Quality and Quantity, DataStream <https://doi.org/10.25976/K85M-PV46> (2023).
94. Aboriginal Affairs and Northern Development Canada Division. CIMP 140: Changing hydrology in the Taiga Shield—Geochemical and resource management implications, DataStream <https://doi.org/10.25976/1ED9-GR10> (2022).
95. Alberta Environment and Parks, Regional Aquatic Monitoring Program (RAMP). Athabasca Basin: Tailing Ponds and Impacts on Aquifers, DataStream <https://doi.org/10.25976/388F-H804> (2022).
96. Alberta Lake Management Society. LakeWatch Water Quality Data, DataStream <https://doi.org/10.25976/VMET-CT64> (2022).
97. Athabasca Chipewyan First Nation. Athabasca Chipewyan First Nation Community Based Monitoring Program, DataStream <https://doi.org/10.25976/50S6-8W42> (2023).
98. Communities of the Northwest Territories, NWT-wide Community Based Water Quality Monitoring Program, & Government of the Northwest Territories, Environment and Climate Change. NWT-wide Community-based Monitoring Program, DataStream <https://doi.org/10.25976/4DER-GD31> (2023).
99. Environment and Climate Change Canada National Wildlife Research Centre. CIMP 161: Cumulative Impacts Monitoring of Aquatic Ecosystem Health of Yellowknife Bay, Great Slave Lake, NWT. DataStream <https://doi.org/10.25976/48P4-ST77> (2022).
100. Environment and Climate Change Canada & Carleton University. CIMP 177: The influence of forest fires on metal deposition to lakes and peatlands in the North Slave Region, NWT, DataStream <https://doi.org/10.25976/AW55-BX08> (2022).
101. Environment and Climate Change Canada & Parks Canada. Lower Mackenzie River Basin long-term water quality monitoring data, DataStream <https://doi.org/10.25976/1021-MV86> (2023).
102. Environment and Climate Change Canada (ECCC) / Environnement et Changement climatique Canada (ECCC); Partners. Peace-Athabasca River Basin Long-term Water Quality Monitoring Data, DataStream <https://doi.org/10.25976/9Q5K-PJ02> (2023).
103. Fort Nelson First Nation, Kerr Wood Leidal Associates Ltd., Peace Country Technical Services Ltd., & Chevron Resources. Fort Nelson First Nation Water Quality Monitoring, DataStream <https://doi.org/10.25976/2548-6V29> (2023).
104. GW Solutions, Municipality of Hudson's Hope, & Saulteau First Nation. Peace River Regional District Water Quality Baseline - Municipality of Hudson's Hope Lynx and Brenot Creek, DataStream <https://doi.org/10.25976/1J42-K176> (2022).
105. GW Solutions. Peace River Regional District Water Quality Baseline - BC Environmental Monitoring System (EMS), DataStream <https://doi.org/10.25976/MD0Q-9X60> (2022).
106. K'ágeé Tú First Nation (KTFN). K'ágeé Tu First Nation (KTFN) Community Based Monitoring of Kakisa River watershed, DataStream <https://doi.org/10.25976/EQ10-ZN81> (2022).
107. Lac La Biche County. Lac La Biche County Lake Water Quality Monitoring Program, DataStream <https://doi.org/10.25976/3PFH-3689> (2022).
108. Lesser Slave Watershed Council. LSWC tributary monitoring program, DataStream <https://doi.org/10.25976/T1JE-T809> (2023).
109. Mikisew Cree First Nation. Mikisew Cree First Nation - Community Based Monitoring Program, DataStream <https://doi.org/10.25976/AFJ8-CC25> (2023).
110. Pisaric, M. CIMP 174: The Impacts of Recent Wildfires on Northern Stream Ecosystems, DataStream <https://doi.org/10.25976/1X4V-8J34> (2022).
111. University of Alberta. CIMP 180: The impact of wildfire on diverse aquatic ecosystems of the NWT, DataStream <https://doi.org/10.25976/487W-F297> (2022).
112. University of Alberta. CIMP 199: Water quality of peatland ponds and streams on a latitudinal transect. DataStream <https://doi.org/10.25976/RZKG-7N02> (2022).
113. University of Waterloo. CIMP167: Changes in dissolved organic carbon quality and quantity: Implications for aquatic ecosystems and drinking water quality for northern communities. DataStream <https://doi.org/10.25976/HX2S-M606> (2022).
114. Upper Athabasca Community Based Monitoring (UATHCBM). Upper Athabasca Community Based Monitoring (UATHCBM). DataStream <https://doi.org/10.25976/YKHJ-0Z76> (2022).
115. Wilfrid Laurier University. CIMP 197. DataStream <https://doi.org/10.25976/SWMN-P451> (2020).
116. Antoniadis, D., Douglas, M. S. V. & Smol, J. P. Comparative physical and chemical limnology of two Canadian High Arctic regions: Alert (Ellesmere Island, NU) and Mould Bay (Prince Patrick Island, NWT). *Arch. Hydrobiol.* **158**, 485–516, <https://doi.org/10.1127/0003-9136/2003/0158-0485> (2003).
117. Antoniadis, D., Douglas, M. S. V. & Smol, J. P. The physical and chemical limnology of 24 ponds and one lake from Isachsen, Ellef Ringnes Island, Canadian High Arctic. *Internat. Rev. Hydrobiol.* **88**, 519–538, <https://doi.org/10.1002/iroh.200310665> (2003).
118. Filazzola, A. *et al.* A database of chlorophyll and water chemistry in freshwater lakes. *Sci. Data* **7**, 310, <https://doi.org/10.1038/s41597-020-00648-2> (2020).
119. Joynt, E. H. III & Wolfe, A. P. Paleoenvironmental inference models from sediment diatom assemblages in Baffin Island lakes (Nunavut, Canada) and reconstruction of summer water temperature. *Can. J. Fish. Aquat. Sci.* **58**, 1222–1243, <https://doi.org/10.1139/f01-071> (2001).
120. Michelutti, N., Douglas, M. S. V., Muir, D. C. G., Wang, X. & Smol, J. P. Limnological characteristics of 38 lakes and ponds on Axel Heiberg Island, High Arctic Canada. *Internat. Rev. Hydrobiol.* **87**, 385, [https://doi.org/10.1002/1522-2632\(200207\)87:4<385::AID-IROH385>3.0.CO;2-3](https://doi.org/10.1002/1522-2632(200207)87:4<385::AID-IROH385>3.0.CO;2-3) (2002).
121. Michelutti, N., Douglas, M. S. V., Lean, D. R. S. & Smol, J. P. Physical and chemical limnology of 34 ultra-oligotrophic lakes and ponds near Wynnai Bay, Victoria Island, Arctic Canada. *Hydrobiologia* **482**, 1–13, <https://doi.org/10.1023/A:1021201704844> (2002).
122. Morrison, H. A. & Carou, S. *Canadian Acid Deposition Science Assessment*. (Environment Canada, 2005).
123. Lienitz, R., Smol, J. P. & Lean, D. R. S. Physical and chemical limnology of 24 lakes located between Yellowknife and Contwoyto Lake, Northwest Territories (Canada). *Can. J. Fish. Aquat. Sci.* **54**, 12, <https://doi.org/10.1139/f96-275> (1997).
124. Rühland, K. M., Smol, J. P., Wang, X. & Muir, D. C. G. Limnological characteristics of 56 lakes in the Central Canadian Arctic Treeline Region. *J. Limnol.* **62**, 9, <https://doi.org/10.4081/jlimnol.2003.9> (2003).
125. Environmental Protection Agency & United States Geological Survey. Water Quality Portal. *USGS DOI Tool Production Environment* <https://doi.org/10.5066/P9QRKUVJ> (2021).

126. De Cicco, L. A., Hirsch, R. M., Lorenz, D. & Watkins, D. dataRetrieval, <https://doi.org/10.5066/P9X4L3GE> (2018).
127. R Development Core Team. R, a language and environment for statistical computing. (2021).
128. Weyhenmeyer, G. A. *et al.* Widespread diminishing anthropogenic effects on calcium in freshwaters. *Sci. Rep.* **9**, 10450, <https://doi.org/10.1038/s41598-019-46838-w> (2019).
129. Yuan, F. & Miyamoto, S. Dominant processes controlling water chemistry of the Pecos River in American southwest. *Geophys. Res. Lett.* **32**, L17406, <https://doi.org/10.1029/2005GL023359> (2005).
130. Collins, W. D. & Riffenburg, H. B. Quality of water of Pecos River in Texas. *Contributions to Hydrology of the United States* 67–88, (1927).
131. Pebesma, E. Simple features for R: standardized support for spatial vector data. *R J.* **10**, 439, <https://doi.org/10.32614/RJ-2018-009> (2018).
132. Yang, X., Xie, X., Liu, D. L., Ji, F. & Wang, L. Spatial interpolation of daily rainfall data for local climate impact assessment over Greater Sydney Region. *Adv. Meteorol.* **2015**, 1–12, <https://doi.org/10.1155/2015/563629> (2015).
133. Zimmerman, D., Pavlik, C., Ruggles, A. & Armstrong, M. P. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Math. Geol.* **31**, 375–390, <https://doi.org/10.1023/A:1007586507433> (1999).
134. Keskin, M. *et al.* Comparing spatial interpolation methods for mapping meteorological data in Turkey. in *Energy Systems and Management* (eds Bilge, A. N., Toy, A. Ö. & Günay, M. E.) 33–42, (Springer International Publishing, 2015).
135. Hiemstra, P. H., Pebesma, E. J., Twenhöfel, C. J. W. & Heuvelink, G. B. M. Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Comput. Geosci.* **35**, 1711–1721, <https://doi.org/10.1016/j.cageo.2008.10.011> (2009).
136. Granville, K., Woolford, D. G., Dean, C. B., Boychuk, D. & McFayden, C. B. On the selection of an interpolation method with an application to the Fire Weather Index in Ontario, Canada. *Environmetrics* **e2758**, <https://doi.org/10.1002/env.2758> (2022).
137. Pebesma, E. J. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* **30**, 683–691, <https://doi.org/10.1016/j.cageo.2004.03.012> (2004).
138. Morley, S. K., Brito, T. V. & Welling, D. T. Measures of Model Performance Based On the Log Accuracy Ratio. *Space Weather* **16**, 69–88, <https://doi.org/10.1002/2017SW001669> (2018).
139. Hijmans, R. J. Raster: geographic data analysis and modelling. R package version 3.4-13. (2021).
140. South, A. rnaturalearth: world map data from Natural Earth. R package version 0.1.0. <https://CRAN.R-project.org/package=rnaturalearth> (2017).
141. Guerin, A. J., Weise, A. M. & Therriault, T. W. Dissolved calcium and pH raster layers for freshwater environments in Canada and the USA. *Dryad* <https://doi.org/10.5061/dryad.dv41ns24c> (2024).
142. Guerin, A. J. Scripts and data for freshwater calcium concentration and pH interpolations covering Canada and the USA. *Zenodo* <https://doi.org/10.5281/ZENODO.10573970> (2024).
143. Wilcox, M. A., Weise, A.M., Guerin, A. J., Chu, J. W. F. & Therriault, T. W. National aquatic invasive species (AIS) risk assessment for Zebra (*Dreissena polymorpha*) and Quagga (*Dreissena rostriformis bugensis*) mussels, April 2022. *DFO Can. Sci. Advis. Sec. Res. Doc.* **2024/008**, ix + 91, <https://waves-vagues.dfo-mpo.gc.ca/library-bibliotheque/41229988.pdf> (2024).
144. Mackie, G. L. & Claudi, R. *Monitoring and Control of Macrofouling Mollusks in Fresh Water Systems*. (CRC Press, 2009).
145. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315, <https://doi.org/10.1002/joc.5086> (2017).
146. Sepulveda, A. J., Gage, J. A., Counihan, T. D. & Prisciandaro, A. F. Can big data inform invasive dreissenid mussel risk assessments of habitat suitability? *Hydrobiologia*, <https://doi.org/10.1007/s10750-023-05156-z> (2023).
147. Le, T. D. H. *et al.* Predicting current and future background ion concentrations in German surface water under climate change. *Phil. Trans. R. Soc. B* **374**, 20180004, <https://doi.org/10.1098/rstb.2018.0004> (2019).
148. Olson, J. R. & Hawkins, C. P. Predicting natural base-flow stream water chemistry in the western United States. *Water Resour. Res.* **48**, W02504, <https://doi.org/10.1029/2011WR011088> (2012).
149. Peterson, E. E., Merton, A. A., Theobald, D. M. & Urquhart, N. S. Patterns of spatial autocorrelation in stream water chemistry. *Environ. Monit. Assess.* **121**, 571–596, <https://doi.org/10.1007/s10661-005-9156-7> (2006).
150. Dowd, P. A. & Pardo-Igúzquiza, E. The Many Forms of Co-kriging: A Diversity of Multivariate Spatial Estimators. *Math Geosci* <https://doi.org/10.1007/s11004-023-10104-7> (2023).
151. Asante, D., Appiah-Adjei, E. K. & Asare, A. Delineation of groundwater potential zones using cokriging and weighted overlay techniques in the Assin Municipalities of Ghana. *Sustain. Water Resour. Manag.* **8**, 55, <https://doi.org/10.1007/s40899-022-00639-8> (2022).
152. Guo, X., Zhao, C., Li, G., Peng, M. & Zhang, Q. A Multifactor-Based Random Forest Regression Model to Reconstruct a Continuous Deformation Map in Xi'an, China. *Remote Sensing* **15**, 4795, <https://doi.org/10.3390/rs15194795> (2023).
153. Zhang, Y. *et al.* Prediction of Spatial Distribution of Soil Organic Carbon in Helan Farmland Based on Different Prediction Models. *Land* **12**, 1984, <https://doi.org/10.3390/land12111984> (2023).
154. Qu, L. *et al.* Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas. *Catena* **234**, 107572, <https://doi.org/10.1016/j.catena.2023.107572> (2024).

Acknowledgements

Funding for this project was provided by the Fisheries and Oceans Canada (DFO) Aquatic Invasive Species program. The authors also wish to acknowledge the following individuals and organisations who provided data or assisted with data acquisition: Sarah Forté (Crown-Indigenous Relations and Northern Affairs Canada); Kivalliq Inuit Association; Victoria Millette (Water Security Agency of Saskatchewan); Marie Ducharme, Cameron Sinclair and Nicole Novodvrosky (Department of Environment, Government of Yukon); Daniel Rheault (Department of Agriculture and Resource Development, Government of Manitoba); Jason LeBlanc (Department of Fisheries and Aquaculture, Government of Nova Scotia); Cindy Chu (Ontario Ministry of Natural Resources & Forestry); Annick Drouin, Sylvie Normand, Mario Bérubé, and Manon Ouellet (Ministère de l'Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs, Gouvernement du Québec); Julie Grenier (Conseil de gouvernance de l'eau des bassins versants de la rivière Saint-François); Robin Staples (Environment and Natural Resources, Government of the Northwest Territories); Thomas Clair, Suzanne Couture, Tanya Johnston, Mary Raven, Nelda Craig, Pat Shaw, Justin Shead, Dean Jeffries, Leif-Matthias Herborg, and others who provided data for DFO's 2012 risk assessment. Chris McKindsey and David Drolet provided constructive comments which led to improvements to the paper.

Author contributions

A.W. and T.T. secured funding and coordinated the work; T.T., A.W., A.G., M.W. and J.C. contributed to discussions on interpolation methods; AW provided existing data; T.T., A.W., E.S.G. and A.G. liaised with multiple organisations to obtain data; E.S.G. and A.G. downloaded data from public repositories and wrote scripts to perform data compilation and cleaning; A.G. assembled the final databases and carried out all spatial interpolations; A.G. and A.W. drafted the manuscript; all authors assisted with editing and revision.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.J.G. or A.M.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2024