Environmental Science & Technology
pubs.acs.org/est
Article

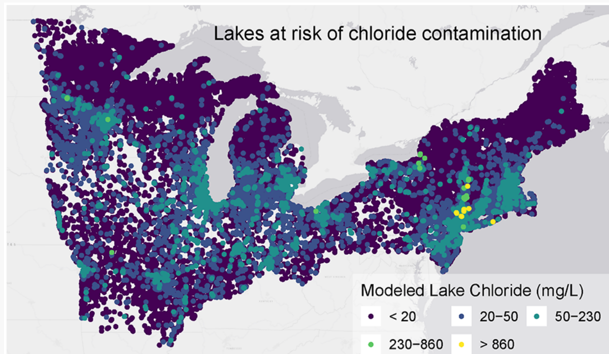
Lakes at Risk of Chloride Contamination

Hilary A. Dugan,* Nicholas K. Skaff, Jonathan P. Doubek, Sarah L. Bartlett, Samantha M. Burke, Flora E. Krivak-Tetley, Jamie C. Summers, Paul C. Hanson, and Kathleen C. Weathers

Cite This: https://dx.doi.org/10.1021/acs.est.9b07718 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Lakes in the Midwest and Northeast United States are at risk of anthropogenic chloride contamination, but there is little knowledge of the prevalence and spatial distribution of freshwater salinization. Here, we use a quantile regression forest (QRF) to leverage information from 2773 lakes to predict the chloride concentration of all 49 432 lakes greater than 4 ha in a 17-state area. The QRF incorporated 22 predictor variables, which included lake morphometry characteristics, watershed land use, and distance to the nearest road and interstate. Model predictions had an r^2 of 0.94 for all chloride observations, and an r^2 of 0.86 for predictions of the median chloride concentration observed at each lake. The four predictors with the largest influence on lake chloride concentrations were low and medium intensity development in the watershed, crop density in the watershed, and distance to the nearest interstate. Almost 2000 lakes are predicted to have chloride concentrations above 50 mg L^{-1} and should be monitored. We encourage management and governing agencies to use lake-specific model predictions to assess salt contamination risk as well as to augment their monitoring strategies to more comprehensively protect freshwater ecosystems from salinization.



■ INTRODUCTION

Sixty years of unbridled road salt application across a vast area of North America has led to the salinization of freshwater ecosystems.[1,2] Studies have documented a long-term rise in chloride and salinity in groundwater, streams, and lakes; including the Laurentian Great Lakes.[2−6] The freshwater-rich U.S. Midwest and Northeast states, which contain 49 000 large lakes (>4 ha), as well as hundreds of thousands of wetlands, streams, and ponds, are particularly reliant on road salt application during winter weather. Given the ubiquity and broad spatial distribution of road salt application, it is highly probable that many of these waterbodies are undergoing salinization. However, with water quality data available from only ~5% of the lakes in this region, the prevalence and spatial distribution of the problem is poorly understood.

When added to the landscape, deicing salts readily dissolve into their associated ions, often sodium (sometimes calcium and potassium) and chloride. Because chloride is the common anion in all salts, is mostly unreactive, and occurs naturally at extremely low concentrations in freshwater lakes,[7,8] its abundance in water is considered a robust tracer of anthropogenic inputs. The concern over elevated chloride concentrations is 2-fold: First, high concentrations are toxic to a variety of freshwater aquatic organisms.[9−11] Government agencies have set chloride thresholds to protect aquatic ecosystem function, but evidence is mounting that these criterion lack ecological realism and should not be standardized across all aquatic systems.[12−14] In the United States, the Environmental Protection Agency's (EPA) acute and chronic chloride concentration toxicity standards are 860 mg L^{-1} and 230 mg L^{-1}, respectively.[15] Second, chloride and associated sodium concentrations can compromise drinking water sources as concentrations above 250 mg L^{-1} taste salty, and high sodium concentrations contribute to hypertension in humans.[16] Salts are also known to mobilize toxic heavy metals from soils to groundwater, which have the potential to bioaccumulate in aquatic food webs and compromise human health.[17−19] Removing dissolved ions from water sources is an added expense to drinking water treatment, and identifying waterbodies at risk of long-term salinization will enable water managers to consider reduction of source inputs as an alternative to removal through costly filtration.

While road salt application is the most pervasive and common salt source, water softeners, synthetic fertilizers (namely KCl), and livestock excretion contribute sizable loads of salt to the environment.[20] Lakes are particularly susceptible to long-term salinization because of their longer water residence time and their large watersheds, which combine many exogenous salt

Received: December 18, 2019
Revised: April 30, 2020
Accepted: April 30, 2020
Published: April 30, 2020

Table 1. Predictor Variables (n = 22) Used in the Training Dataset

predictor	predictor.ID	data source	data processing
month of sampling	month	WQP	month of *ActivityStartDate*
lake area (ha)	LakeArea	LAGOS-NE: derived from National Hydrography data set	Subset to lakes of interest using NHD ID
watershed area (ha)	WS.Area	LAGOS-NE: custom geoprocessing using National Elevation Data set DEMs	Subset to lakes of interest using NHD ID
land use (watershed %):			
ice and snow	WS.IceSnow,	LAGOS-NE: derived from the National Land Cover Data set (NLCD). Refer to NLCD metadata for definitions of land cover classes.	subset to lakes of interest using NHD ID
open development	WS.Dev.Open,		
low development	WS.Dev.Low,		
medium development	WS.Dev.Med,		
high development	WS.Dev.High,		
barren land	WS.Barren,		
deciduous forest	WS.DeciduousForest,		
evergreen forest	WS.EvergreenForest,		
mixed forest	WS.MixedForest,		
shrub	WS.Shrub,		
grassland	WS.Grassland,		
pasture/hay	WS.PastureHay,		
crops	WS.Crops,		
woody wetlands	WS.WoodyWetlands,		
emergent wetlands	WS.EmergentWetlands		
road density in the watershed (meters ha^{-1})	WS.RoadDensity		
index of winter severity (unitless, range: 7.6−168.1, Figure S1)	WinterSeverity	ClearRoads (national research consortium, clearroads.org). Calculated from 2000 to 2010 as 0.50 × (average annual snowfall in inches) + 0.05 × (annual duration of snowfall in hours) + 0.05 × (annual duration of blowing snow in hours) + 0.10 × (annual duration of freezing rain in hours).	spatial join of lake locations with gridded winter severity product. Grid resolution of 0.25° lat/long.
distance to the nearest interstate (km)	InterstateDistance	U.S. Census Bureau	Euclidean distances to the nearest feature in R using the st_distance function in the sf package[66]
distance to the nearest road (km)	RoadDistance	U.S. Census Bureau (Openstreets data, where *fclass* is equal to motorway_link, motorway, primary, trunk_link, primary_link, or trunk)	

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

sources.[21] Unlike streams and rivers, which may experience extreme chloride loading during spring melt periods from surface inflows[22] or summer groundwater discharge,[23] lake chloride concentrations are more likely to display low interannual variability if their water residence time is greater than one year, even if these waterbodies are also receiving runoff events. The longer the residence time, the longer it will take for a waterbody to reach dynamic equilibrium with the inflows, and the more representative a single water sample from a lake may be of overall chloride load in its watershed. Despite the capacity for low-resolution water quality sampling to assess lake health, chloride is not routinely monitored in many lakes and states.

To identify at-risk lakes, we use machine learning and regional databases of lake characteristics and water quality to determine the predictors of chloride contamination. This analysis was motivated by previous regional and local studies of long-term chloride trends indicating that developed land, including impervious surface cover and the density of the road network surrounding lakes and rivers, are strongly correlated with elevated chloride concentrations.[5,24−26] These studies suggest that landscape watershed characteristics, along with lake characteristics, may provide the predictive power to identify sites at risk of chloride contamination, even in unmonitored areas.

Our analysis covers a 17-state area of the United States, including Connecticut, Illinois, Indiana, Iowa, Maine, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Vermont, and Wisconsin. Road salt application across this area is widespread, intensive, and expensive. Total road salt application in these states averages 6.9 million tons of salt and 68 million gallons of liquid brine per year, at an annual cost of $367 million for materials and $1.4 billion for application (years 2014−2017, Clear Roads 2018). This area is also lake-rich, containing 23% of the lakes in the United States.[28] The convergence of abundant freshwater and rampant salt use underscores the need to prioritize monitoring and predictive modeling in this region, with the aim of ultimately reducing anthropogenic chloride pollution.

■ METHODS

We leveraged publicly available land use, lake catchment and morphometry, and climate data across a 17-state area of the Midwest and Northeast United States, to predict chloride concentrations in 49 432 lakes. Our general methodology included: (Step i) Acquiring and geoprocessing lake water quality data and site characteristics. (Step ii) Harmonizing training data sets. (Step iii) Harmonizing hold-out/testing data sets. (Step iv) Building a machine learning model and selecting optimum hyperparameters for chloride prediction. (Step v) Calculating model fit on training and hold-out data sets. (Step vi) Evaluating variable importance of predictors. (Step vii) Building a prediction data set for 49 432 lakes.

Step i: Data Acquisition. Observational chloride measurements from lakes, reservoirs, and impoundments were downloaded from the U.S. water quality portal (WQP, waterqualitydata.us).[29] All results were converted to mg L^{-1}, and only data with *ResultStatusIdentifier* as "Accepted" or "Final" noted in the data set were retained. The initial search of 115 389 observations was then filtered to data collected after 1990, chloride concentrations <10 000 mg L^{-1}, and water samples less than 10 m deep or with depth not listed (where the assumption was an epilimnion measurement). These quality control steps were taken to limit inclusion of historical data that may not represent current conditions, remove naturally saline waterbodies ($n = 5$, adjacent/connected to the Atlantic Ocean), and remove potentially meromictic lakes ($n = 0$). Multiple observations collected on the same day were averaged. Lakes with missing watershed information were removed, resulting in 29 010 unique daily observations from 2773 lakes, which represent 5% of the lakes in the region. Three states (Illinois, Iowa, and Rhode Island) had no chloride data, and three states (Pennsylvania, Connecticut, and New Hampshire) had chloride data from only one lake.

The WQP, from which our observational data were drawn, combines data from state, local, and tribal monitoring agencies and acts as a standardized data warehouse for the procurement of regional to national-scale data. This simplified avenue for data access enables modeling efforts at a multistate level, but this secondary use of data ("the use of data beyond the original intent determined by the organization that collected the data") comes with the challenge of quality control.[30] A study by the United States Geological Survey (USGS) on the quality of multisource data found that, of the 25 million records examined, more than 14 million had missing or ambiguous metadata.[30] Ambiguity can arise from parameter names, filtration methods, chemical form, and units. In our experience, chloride observations are less cryptic than other water quality parameters, in that the parameter name is more standardized, the filtration and storage method is less prone to error, there is effectively only one chemical form, and the reported units are logical (mg L^{-1}, ppm, eq L^{-1}, μg L^{-1}, and μmol). The occurrence of outliers and data quality assurance are discussed in the Supporting Information (SI).

Step ii: Training Data Set. WQP site identification numbers (IDs) from the data set were linked to the high-resolution National Hydrography Data set (NHD) using a Python script that accessed bounding box information on each NHD shapefile and ran a spatial join (see SI). The resulting relational table linked each chloride observation to an individual lake through an NHD ID. For every NHD lake ID, geospatial lake data were obtained from the LAGOS-NE database,[31] which provides watershed ecological context for all lakes greater than 4 ha in the 17-state area (Table 1). Additional site characteristics were extracted from GIS line features of U.S. interstates and U.S. primary roads, and from gridded winter severity data (Table 1, Figure S1). Across all predictor variables in the training data set (Table 1), minimum values (excluding zero data) were ≥0.01. To log-transform data, we converted zero values to 0.001; thereby allowing us to distinguish zero values from positive values. Predictors related to climate, salt application rates at the state level, and coastal distance were investigated during initial model runs but found to be uninformative on our training data set, and not used in our final model (see SI).

Step iii: Holdout Data Sets. Four hold-out testing data sets (data not available in the WQP) were used to provide an independent evaluation of the model fit:

a. Rhode Island lakes (93 lakes). The University of Rhode Island Watershed Watch Program provided data for lakes in Rhode Island with chloride observations from 1990 to 2010.[32] Rhode Island lakes comprise some of the most urban lakes in the region.

b. Minnesota lakes (137 lakes). The Minnesota Pollution Control Agency provided chloride data from 1990 to

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

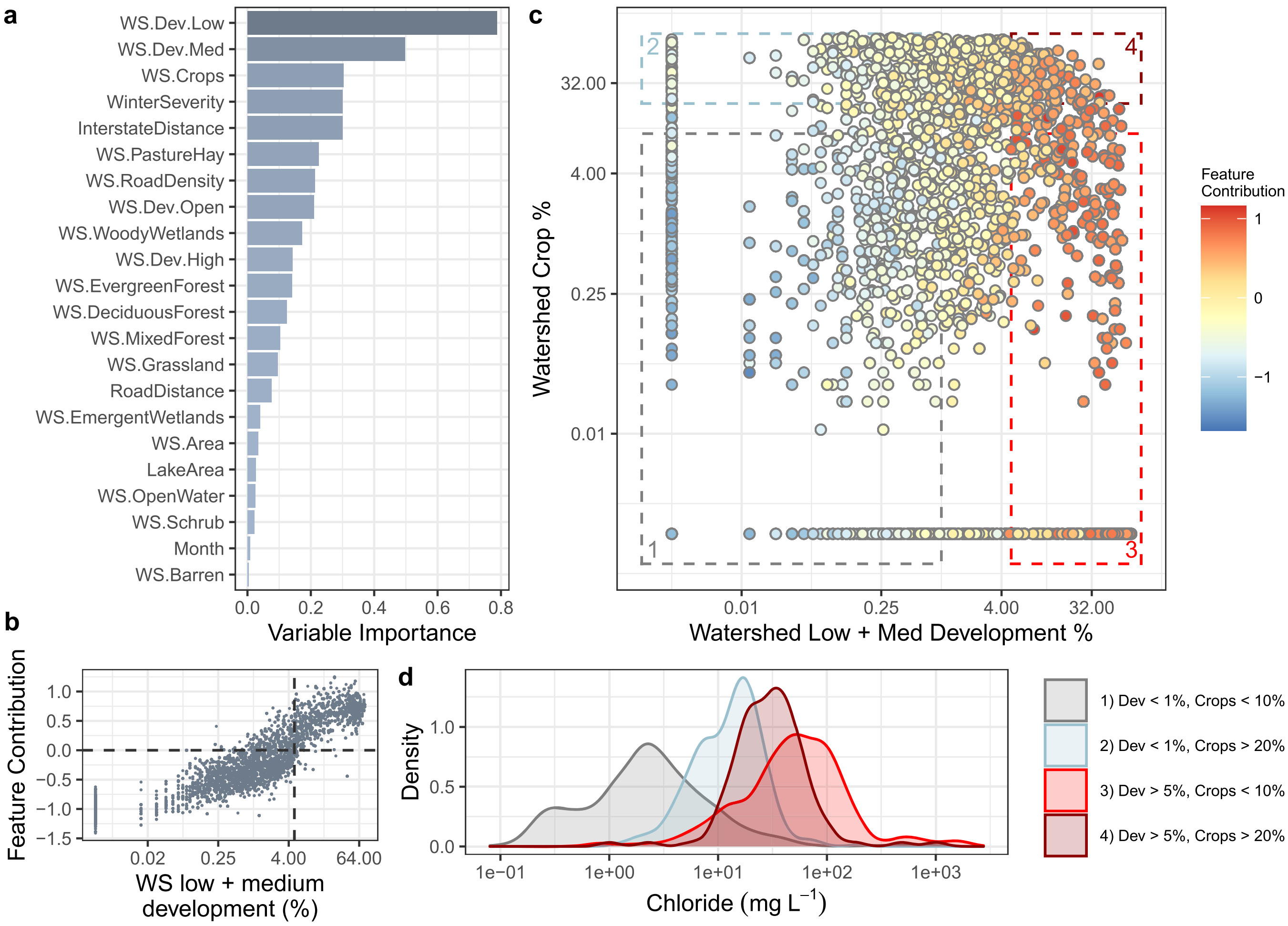


Figure 1. (a) Predictor variable importance calculated by permutation, with higher values on the *x*-axis corresponding with greater importance. (b) Feature contribution plot of the top two predictors—percent of low and medium development in the watershed. Feature contributions >0 indicate that the combination of predictor values resulted in a predicted chloride concentration greater than the mean predicted value. Feature contributions <0 indicate that the combination of predictors resulted in a predicted chloride concentration less than the mean predicted value. Dashed lines highlight FC = 0 and 5% development. (c) Combined feature contribution of watershed low and medium development (*x*-axis) and percentage of cropland in the watershed (*y*-axis) to lake chloride concentration. Each point represents an individual observation in the training data set and the associated feature contribution value represents the influence of the predictors on the total predicted chloride concentration. Dashed boxes 1−4 are referenced in (d). (d) Chloride concentration density distributions for four groupings of watershed characteristics. Data are the median observed chloride concentrations from the 2773 training lakes.

2018 for lakes that were not available in the WQP. The observed chloride was ≤1 mg L^{-1} for 104 of 137 lakes.

c. Wisconsin lakes (134 lakes). Chloride concentrations from Wisconsin lakes were compiled from four independent surveys available from the North Temperate Lakes Long-Term Ecological Research Site (NTL-LTER). Because these surveys only represented lakes in Northern Wisconsin, we sampled eight lakes in Southern Wisconsin in 2018 to better represent the diversity of lakes in the state.

d. The U.S. EPA National Lakes Assessment (NLA) lakes (488 lakes). The NLA is a snapshot survey of summer lake water quality across the U.S. that was conducted in 2007 and 2012.[33] Data from both sampling years were retained within the 17-state area used in this study, and any duplicate sampling was averaged. Sixty-one lakes were too small (<4 ha) to be included in the LAGOS data set. Surprisingly, none of the remaining 488 lakes were represented in the WQP training data.

We elected not to incorporate test data into our training data set to examine the model's predictive performance on regions not covered by the training data. By examining test-set error we could examine whether the model was robust to a variety of regional characteristics. Ultimately, our goal was to develop a model that is useful outside of well-studied locations, and the approach of hold-out testing data sets was the best method to evaluate our success in this aim.

Step iv: Machine Learning Model. A quantile regression forest (QRF) was used to model the relationship between observed chloride concentrations and lake and watershed characteristics. This model was chosen to accommodate a large number of correlated predictor variables (Figure S2), the presence of nonlinear responses, and the potential importance of interactions among predictor variables. The QRF consisted of 1000 trees and was implemented with the "ranger" package in R.[34]

In a typical QRF routine, a random fraction of observations would be sampled. However, since our training data set is unevenly weighted with respect to number of observations per

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

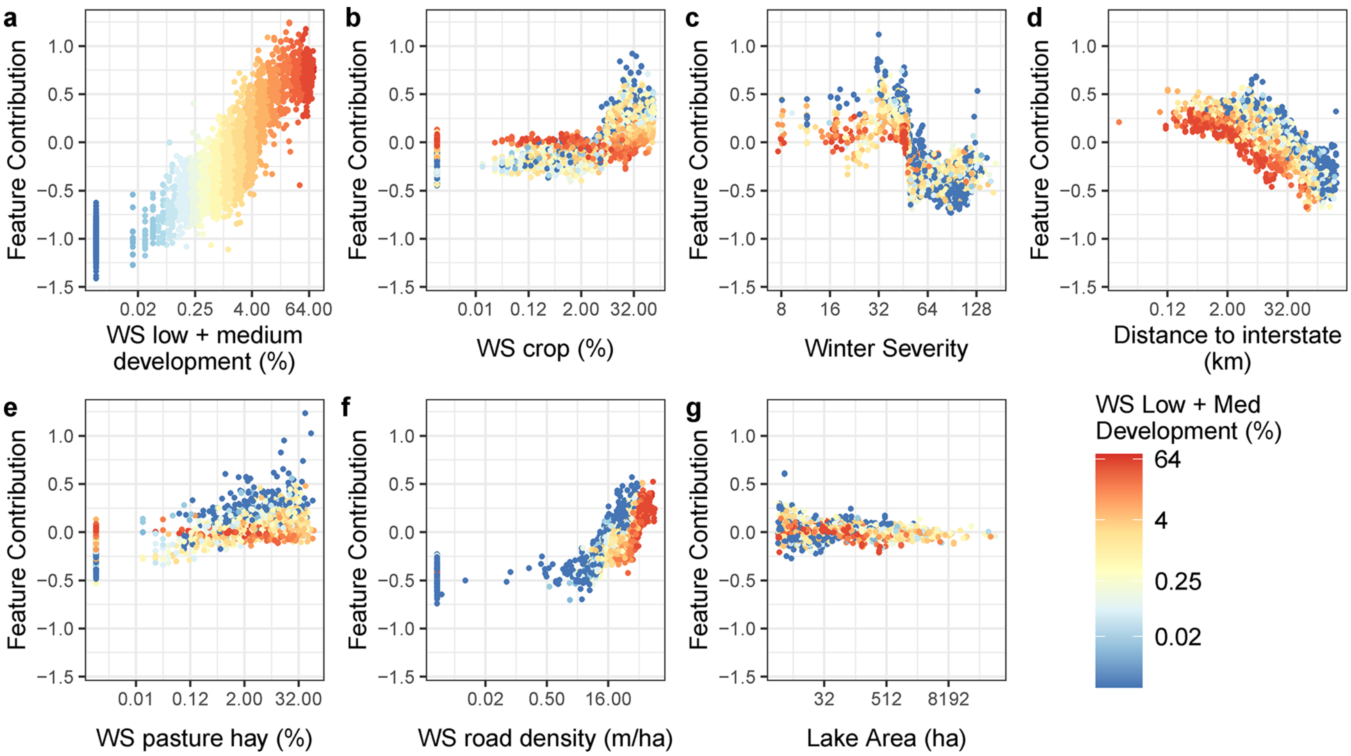


Figure 2. (a−f) Feature contribution plots for the top six predictor variables and (g) one noninformative predictor. *Y*-axis feature contribution values represent the additive contribution of that predictor to the predicted chloride concentrations. Feature contributions >0 indicate that the combination of predictor values resulted in a predicted chloride concentration greater than the mean predicted value. *X*-axis values are the predictor value for individual lakes, and each point represents an individual chloride observation in the training data set. Points are colored by the sum of low and medium development in the watershed to highlight the interaction of multiple drivers.

lake, this approach would overfit the QRF to lakes with a greater number of chloride observations. Therefore, we developed a customized sampling routine that constructed individual trees using the observations from a random subset of the study lakes (95% subset: the "in-bag samples"). Each resulting tree was used to make out-of-bag predictions on the remaining observations from the 5% of excluded lakes. For each test lake, there were 1000 terminal node values that created a distribution of predicted chloride concentrations. All predictions are reported as the back-transformed median of the distribution. Median terminal node values were chosen over mean values because they had superior predictive performance on out-of-bag observations, and to avoid the errors associated with back-transforming mean values. A 90% prediction interval can be calculated from the 0.05 and 0.95 quantiles of the estimated conditional distribution of the response variable.[35] The optimum number of candidate predictor variables considered at each tree split ($mtry = 4$) was selected based on out-of-bag error in the training model.

Step v: Model Fit. Model fits were evaluated using the coefficient of determination (r^2) on log-transformed chloride concentrations, and root-mean-square log error (RMSLE).

Step vi: Variable Importance. Variable importance was calculated via permutation, which measures the change in model prediction error after randomly shuffling each predictor variable. The order and absolute value of the importance of the predictor variables is more of a guidepost than a strict hierarchy of influence, because estimates are affected by the predictors included, the hyperparameters selected and the degree of correlation between predictors.[36] Hyperparameters in machine learning are parameters with a fixed value that are not changed by model training.

Step vii: Prediction Data Set. A prediction data set was constructed for the full LAGOS-NE data set, which contained 51 102 lakes and reservoirs greater than 4 ha in the 17-state area.[31] After removing lakes with no available land-use data because the watersheds crossed the US/Canada border, 49 432 lakes remained, of which 2773 were used for training the model. The prediction data set was identical in structure to the training data set but contained no observational chloride data. The month of prediction was set to July, as this had the greatest overlap with the training data set. In general, we would expect natural, ambient background chloride concentrations across the entire region to be <10 mg L^{-1}.[7,8] The distribution of lakes in the training and prediction data sets were unequal, with training lakes biased toward larger lakes in more urban environments (Figures S3 and S4).

RESULTS

The two most important predictor variables among QRF models constructed with a range of different hyperparameters were the percentage of low (WS.Dev.Low) and medium development (WS.Dev.Med) (Figure 1a). Considering that WS.Dev.Low and WS.Dev.Med were highly correlated, the top four predictors across all model tests were the percentage of low and medium development in a lake's watershed, followed by crop abundance (WS.Crops), a climate index of winter severity (Winter-Severity), and distance to the nearest interstate (Interstate Distance). Surprisingly, morphometric parameters including lake and watershed area, as well as month of sampling, were relatively poor predictors (Figure 1a).

Feature contribution (FC) plots are a useful method to illustrate the influence and additive contributions of predictor

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

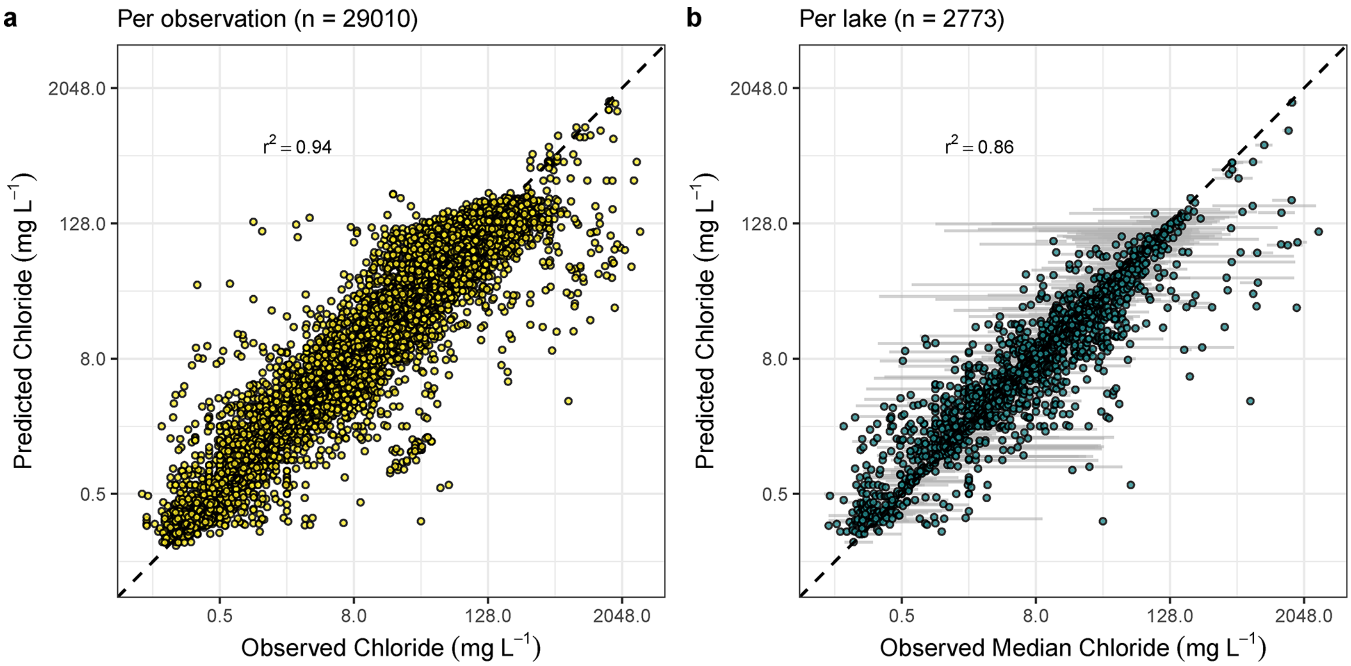


Figure 3. (a) Observed chloride vs out-of-bag predictions for all observations in the training data set. (b) Observed data in (a) grouped by individual lake and plotted as the median. Where multiple observations are available, ranges are shown for min−max observed data.

variables to chloride concentration (Figure 1b). FC values represent the change in predicted probability a sample receives in the QRF when split by the given variable, and can give insight into how the QRF decisions are made. Every test lake is associated with 22 FC values, one for each predictor. The sum of these 22 values added to the overall mean predicted chloride concentration equals the predicted chloride concentration for the test lake. Therefore, an FC equal to 0 would indicate that the predictor variable had no influence on the predicted outcome. FC values >0 indicate the predictor variable had a positive influence (predicted chloride concentration will be greater than the overall mean), and FC values <0 indicate the predictor variable had a negative influence (predicted chloride concentration will be less than the overall mean). The higher the absolute value of an FC, the more influence it has on the final prediction. For instance, the FC plot of the top predictor (low and medium development in a watershed) shows that overall low development results in low chloride concentrations and high development results in high chloride concentrations. Almost all sites with >5% low and medium development in the watershed had FC values >0 (Figure 1b).

Feature contributions are additive, such that a positive contribution from one predictor may cancel out a negative contribution from another predictor. In examining the additive contribution of low and medium development and crop abundance, it is clear that watershed development has a much stronger positive contribution to chloride concentrations than crop percentage (Figure 1c). We can isolate quadrants of the FC plot to show the relationship between modeled feature contributions and observed chloride concentrations. Data in group 1 (development <1%, crops <10%) had the overall lowest feature contributions (median −0.99), and a median observed chloride concentration of 2.2 mg L^{-1}. Data in group 2 (development <1%, crops >20%) still had relatively low feature contributions (median −0.39), and a median observed chloride concentration of 13.8 mg L^{-1}. Data in group 3 (development >5%, crops <10%) had the highest feature contributions (median 0.67), and a median observed chloride concentration of 51.2 mg L^{-1}. Data in group 4 (development >5%, crops >20%) had slightly lower feature contributions than group 3 (median 0.43), and a median observed chloride concentration of 30.7 mg L^{-1} (Figure 1d). Lakes in group 3 had higher overall development than lakes in group 4.

Individual FC plots revealed that several different forms of anthropogenic development in a lake's watershed contribute to higher chloride concentrations (Figure 2). For instance, a higher percentage of low-intensity development, close proximity to interstate highways, and high road density were associated with higher chloride concentration (Figure 2a,d,f). These predictors exhibit threshold effects in some cases. For instance, road density begins to have a positive contribution to chloride concentrations only when there is at least 16 m ha^{-1} of roads in the watershed, and the contribution of interstate highways to chloride concentrations rapidly declines when the distance from lakeshore is greater than 2 km.

FC relationships are modified by co-occurring variables. Here we highlight the interaction of predictor variables with low and medium development. For instance, high agricultural land use, both crop cultivation (Figure 2b) and pasture/hay (Figure 2e), contributed to higher chloride concentrations, but only under low development conditions. Likewise, distance to the nearest interstate was more likely to positively contribute to chloride concentrations when there was low development (Figure 2d). Climate was also associated with differences in lake chloride concentrations; lakes in a winter severity zone greater than 50 generally had lower chloride concentrations than lakes in warmer locations with a severity less than 50 (Figures 2c and S1 for maps). Finally, feature contribution plots of predictors with low variable importance, such as lake area, showed little influence on chloride concentrations across their range (Figure 2g).

QRF Model Predictions. The model predictions for out-of-bag observations within the training data set had an r^2 of 0.94 and an RMSLE of 0.41 for all observations, and an r^2 of 0.86 and

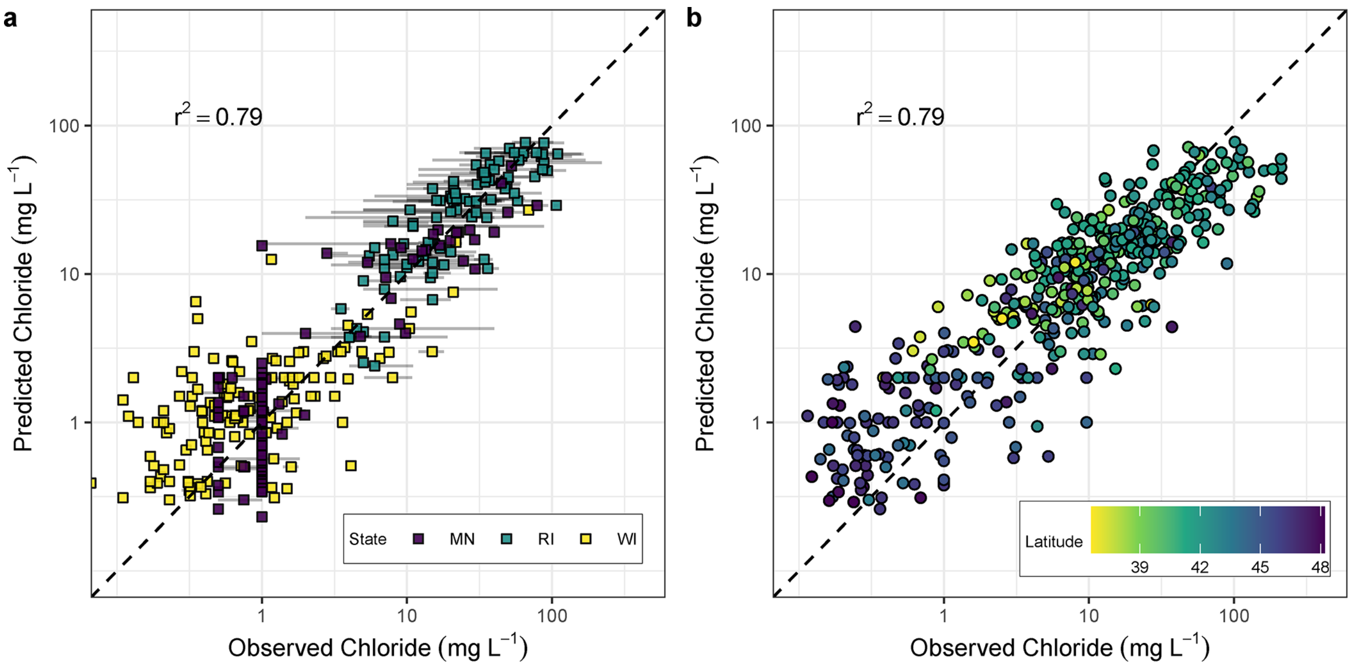


Figure 4. (a) Predicted chloride versus observed chloride for hold-out data sets from Minnesota, Rhode Island, and Wisconsin. Where multiple observations are available, ranges are shown for min−max observed data. (b) Predicted chloride versus observed chloride for lakes in the National Lakes Assessment. Individual lakes are colored by latitude.

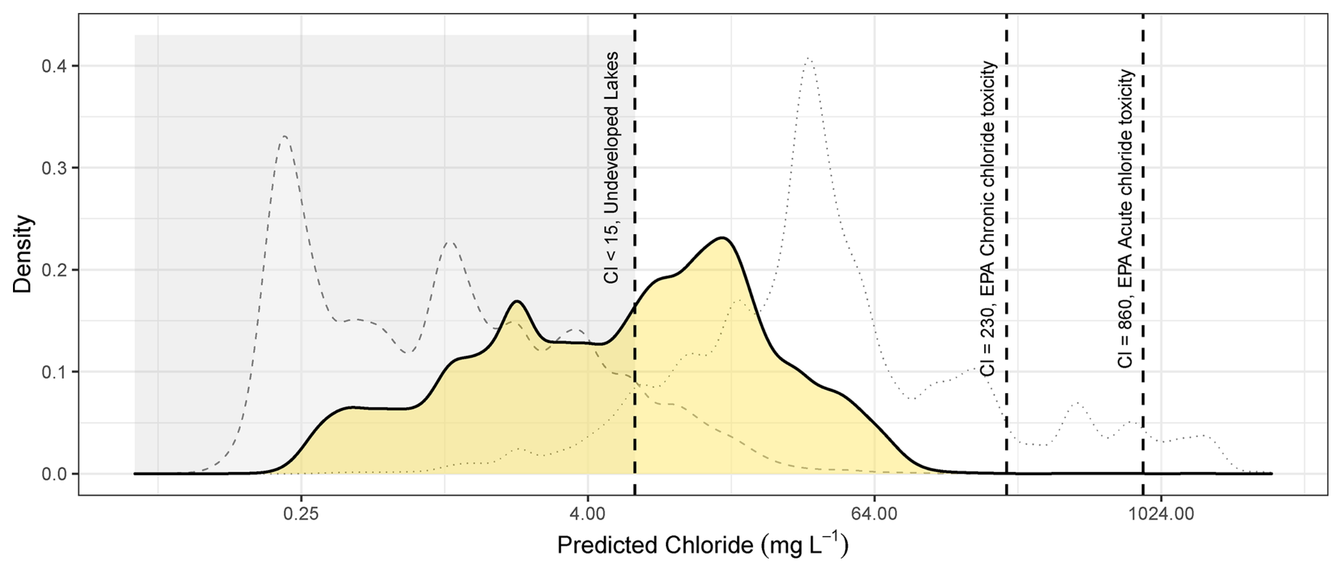


Figure 5. Density distribution of the median predicted chloride concentration for all 49,432 LAGOS-NE lakes (filled yellow). Dashed and dotted lines represent the 5% and 95% prediction quantiles. The gray rectangle denotes the range of predicted values (max 15 mg L^{-1}) for undeveloped lakes. Vertical lines represent the chronic and acute chloride toxicity thresholds as set by the U.S. EPA.

an RMSLE of 0.60 when data were summarized by taking the observed median chloride concentration for each individual lake (Figure 3). Model predictions were not biased toward over or under-prediction, with the exception of ~20 lakes where the chloride concentrations were consistently under-predicted, in some cases by 1000 to 2000 mg L^{-1}. Of the 20 lakes with observed median chloride concentrations >500 mg L^{-1} and predicted concentrations <200 mg L^{-1}, 18 are located in New York state, and all observation data were collected during a single year (2003, see SI). There were no cases where high-concentration lakes were overpredicted by more than 100 mg L^{-1} (Figure 3b).

Holdout Data Sets. The holdout data from Rhode Island, Minnesota, and Wisconsin compare well to the predicted values except at low concentrations (r^2 = 0.79, RMSLE = 0.86). The log−log scale of the prediction plots overemphasizes differences at low concentrations and, therefore, the Wisconsin and Minnesota lakes fall off the 1:1 line at concentrations <4 mg L^{-1}. Many of the Rhode Island lakes had multiple observations, and the model prediction fell within the range of observed values (Figure 4a). In these three states, there is a pattern of decreasing chloride concentration with latitude. This relationship is due to lower population densities in the northern regions of this 17-state area. The observed NLA data likewise compared well with the predictions, with an r^2 of 0.79 and RMSLE of 0.84 (Figure

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

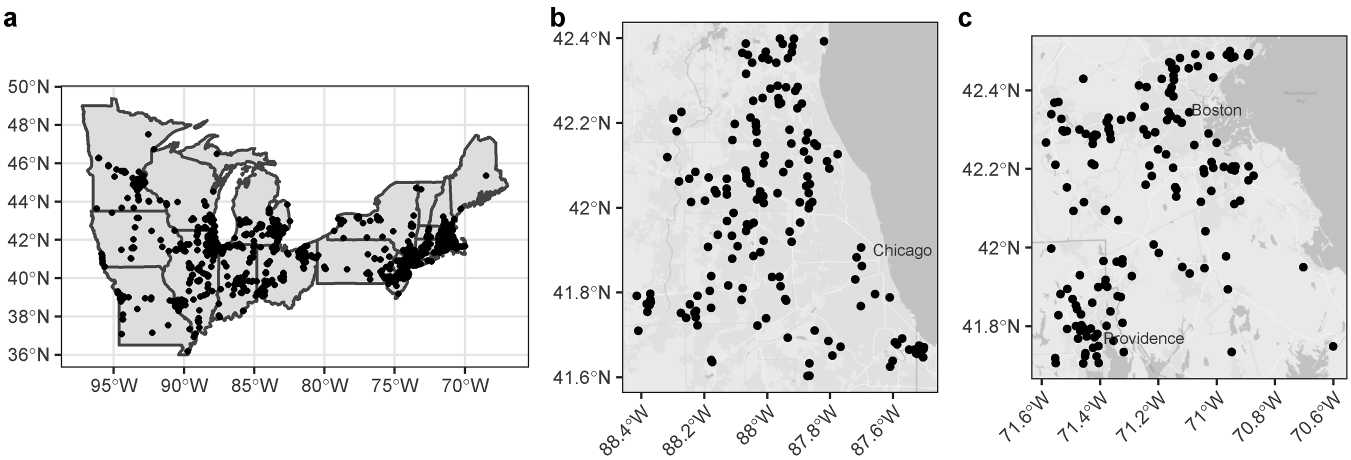


Figure 6. (a) Location of lakes predicted to have chloride concentrations above 50 mg L^{-1}. (b) Chicagoland, and (c) Boston, MA and Providence, RI are highlighted to show the density of at-risk lakes in regions of high development.

4b). While the model under-predicted chloride concentrations for the lakes with the highest observed values, similar to the training data set, the three highest observed lakes (>200 mg L^{-1}) are all located in Illinois, which is a state for which there was no training data.

Predicting Chloride Concentrations Across 17-States. The QRF model was used to predict chloride concentrations in the 49 432 lakes greater than 4 ha in a 17-state area in the U.S. Northeast and Midwest (Figure 5); 2773 lakes were in the training data set, and 46 659 lakes were test lakes. From historical empirical studies, the ambient background chloride concentrations across the entire region is assumed to be <10 mg L^{-1}.[7,8] Using our model, if we include only test lakes with no development or agriculture in their watersheds that are >100 km from an interstate (n = 2100), then the median predicted chloride concentration was 0.42 mg L^{-1}, and the maximum predicted chloride concentration was 15.0 mg L^{-1}.

Overall, the median predicted chloride in the 49 432 lakes was 6.8 mg L^{-1}, compared with a median of 7.9 mg L^{-1} for the training lakes. We identified 1972 lakes with predicted median chloride concentrations above 50 mg L^{-1}, including 245 lakes that were in the training data set. These lakes are typically located along U.S. interstate highways and near major metropolitan areas, including densely populated areas of New York, Massachusetts, and Illinois (Figure 6). Only 114 lakes, including 95 in the training set, are predicted to have median chloride concentrations exceeding 100 mg L^{-1}.

The prediction interval for many of the lakes in this region was high, especially in the Northeast U.S. (Figure S5). The prediction interval is kept as a log value as the difference between 5% and 95% quantiles of the estimated distribution of predicted chloride. When back transformed, the prediction interval scales with chloride concentration, and prevents evaluation of the drivers. The median and 95% quantile of the prediction interval are 3.73 and 5.89. In the 5% of lakes with a prediction interval >5.89, there was a small bias toward smaller lake and watershed area, with more development. The spatial distribution of the high prediction interval lakes reveals many are located in New Jersey, southern New York, and Connecticut (see SI Figure S5).

■ DISCUSSION

Of the 49 432 lakes greater than 4 ha in the 17-state area, about 4% were predicted to have chloride concentrations above 50 mg L^{-1}, and an additional 14% were predicted to have chloride concentrations between 20 and 50 mg L^{-1}. Fortunately, this result indicates that the vast majority of lakes (82%) are likely at concentrations <20 mg L^{-1}. However, with continued land-use change, even pristine lakes should be monitored for early warnings of chloride pollution.

We demonstrate that chloride is predictable at the regional scale, and that we can leverage information from 5% of the region's waterbodies to assess the status of all 49 432 lakes in the region. Overall our training data set was representative of the region in terms of watershed properties (Figure S3), but was skewed toward large lakes with high watershed development (Figure S4). This bias is to be expected, as lakes are rarely sampled to represent the regional population, unless through a stratified sampling regime designed to specifically address this challenge.[37] In reality, lakes are sampled for a variety of reasons, which skews the representativeness.[38] Sampling is often conducted in contaminated lakes or in lakes that are easy to access, which results in observational data that disproportionally represent large lakes in urban areas with water quality issues.[38]

Importance of Predictors. Top predictors of lake chloride concentration were the percentage of low and medium development, the percentage of cropland in a lake's watershed, and the distance to the nearest interstate highway. These predictors have rarely been studied at the same time, but all have been identified as major sources of anthropogenic chloride to freshwater lakes. Studies in North America have found positive correlations between lake and stream chloride concentrations and road density,[24,39] road proximity,[24] and percent impervious surface;[5,40] which are all proxies for road salt application. In addition, studies in Europe and North America have identified agricultural land cover as a major source of chloride to lakes and rivers,[41−43] due to the use of potassium chloride as a synthetic fertilizer. In the United States, potassium chloride accounts for over 90% of potash fertilizers, with over 5 M tons applied per year since 1990.[44]

While we do not quantify salt loads for individual watersheds, we illustrate the relative significance of these predictors as related to salt loading to a watershed. In our 17-state region, a common road salt application rate is 84 kg km^{-1} (300 lb per mile).[45] In a given winter, multiple snowfall events cumulate in an average annual application rate of 15−30 tons per mile (8.45−16.9 kg m^{-1}).[27] Multiplying this rate by the median road density of the 49 432 lakes (19.3 m ha^{-1}), results in an

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

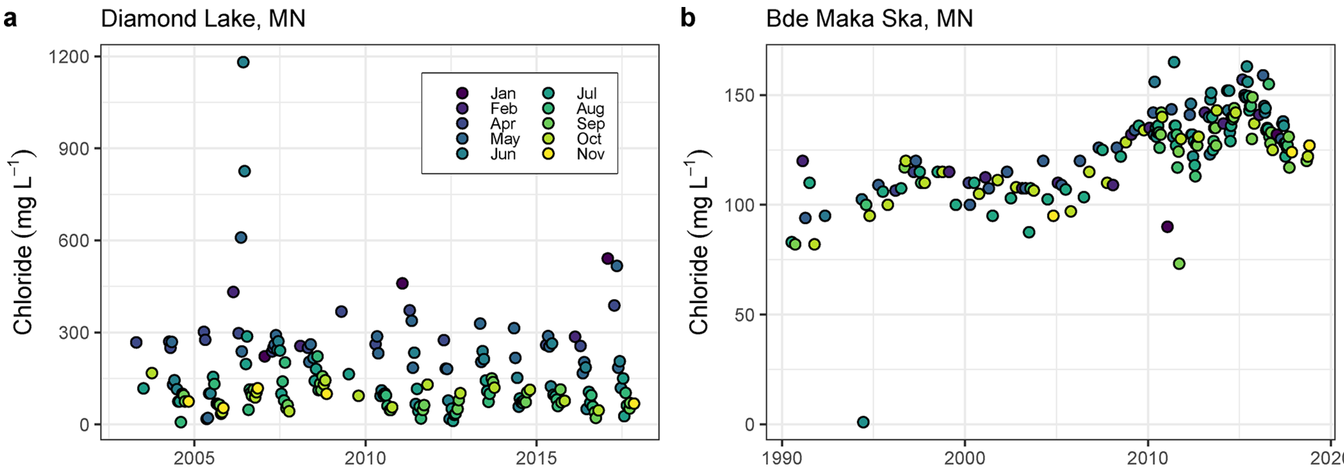


Figure 7. Observed chloride concentrations from (a) Diamond Lake (20 ha), Minnesota, and (b) Bde Maka Ska (170 ha), Minnesota. Observations are colored by the month of sampling. Data obtained from the water quality portal.

application rate of 163−326 kg ha^{-1}. The 95th percentile for road density was 85.1 m ha^{-1}, which could result in salt application rates of 719−1439 kg ha^{-1}. In comparison, the mean application rate of potash to corn and soybean crops in the United States in 2018 was 97 kg ha^{-1}, and 34 kg ha^{-1} to wheat in 2017.[44] While these rates are a liberal estimate—as not all crops are fertilized, and not all roads are salted—they are an indication that chloride from road deicers can easily be an order of magnitude greater than chloride from fertilizer in a watershed, and they support the relative importance of predictors determined in our model.

Winter severity had an unexpected relationship with chloride concentrations where high winter severity (>50) is found to be related to low chloride concentrations (Figure 2c). At first glance, this is counterintuitive, as more severe winters should require more deicing application. However, there are two explanations for this pattern. First, salt as a deicing agent loses efficacy with temperature. Many regions with severe winters are too cold for salt to be effective, and therefore opt for plowing and sanding for winter road maintenance. Second, the population density, and therefore road density and urban land use, is higher in regions with less severe winters. Still, the feature contribution of winter severity is an intriguing pattern and requires further investigation; especially in light of climate change and projected changes in winter severity.[46]

Poor predictors of chloride concentration include the month of sampling, lake area, and watershed area. We had hypothesized that month of sampling would influence chloride concentrations, as there is evidence that some waterbodies experience annual spring peaks in concentration following winter runoff (Figure 7a). However, it is likely that the effects of these factors were small at a regional scale for two reasons. First, depending on climate, hydrological inflows, and residence time, lakes may experience seasonal peaks anywhere from late winter to early summer, and many lakes likely experience no seasonal peaks due to watershed storage that delays inflows (Figure 7b). Second, 83% of observational data were collected between May and September and may not have included enough samples from the winter and spring to resolve the importance of sampling date (Figure S6).

Lake area and watershed area were also relatively poor predictors of chloride concentration. Likewise, maximum lake depth was not an important predictor for the 89% of the observed lakes for which lake depth was available. Had lake depth been an important predictor, the model could only have been applied to the 20% of the LAGOS-NE data set for which lake depth is documented.[47] Residence time is often considered a "master factor" for water quality in lakes,[48,49] as hydraulic residence time is proportional to the concentration of nonconservative solutes that are processed in a lake.[50] While we did not explicitly test residence time in our model, as the requisite data were not available, there was little indication from our morphometric parameters (lake depth, watershed area, and lake area) that residence time was a strong predictor of chloride concentration. This is consistent with models that predict the concentration of a conservative solute (nonreactive, not sedimented) in a steady-state lake will be proportional to flow-weighted concentration in the inflow.[51] Residence time will only impact the time it takes to reach dynamic equilibrium.

Unquantified Sources. As with any broad-scale environmental model, we were unable to account for all potential sources of chloride to lakes. On the basis of our model evaluation, we are confident that our model predictors correlate with road salt use, but we did not explicitly account for spatial variation in application rates, as these data were not available at high resolution at the time of analysis. A recently released a 1 km^2 resolution data set of estimated annual salt application used similar predictors to our model, including road density and developed land use, snowfall, and salt sources by state.[52]

We were unable to account for point sources of chloride including wastewater treatment plants, septic systems,[53] and industrial and mining effluents.[54,55] Wastewater treatment plants are point-sources of chloride due to discharges from water softeners and industrial sources (human waste is a small contributor).[20] In Minnesota, where water softeners are common in both residential and industrial settings, wastewater treatment discharges account for 22% of chloride discharges.[20] However, a large regional study of watersheds along the eastern U.S. found that septic and wastewater were small contributors to overall chloride loads.[56] On a regional scale, it is unknown how many wastewater treatment plants discharge effluent into lakes, and headwater lakes may be largely protected from point-source pollution. On individual lakes, septic systems, industrial effluents, or dust control measures may be significant sources of chloride, but enumeration is beyond the scope of this study.

Prediction at High Concentrations. Freshwater lakes in the U.S. Midwest and Northeast that are not impacted by

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

anthropogenic chloride often have low temporal variability in their chloride concentrations. Therefore, median chloride concentrations of these lakes are largely representative of the overall range in concentrations. However, small lakes that do experience severe chloride contamination may have high seasonal variability in chloride concentrations. Take the extreme example of Diamond Lake, which is a very shallow 20-ha lake in Minneapolis, Minnesota that receives direct runoff from nearby U.S. Interstate-35. The median chloride concentration of the lake is 115 mg L^{-1}, but the range is from 7.5 to 1181 mg L^{-1} (Figure 7a). Arguably, the median concentration of the lake should be viewed as a conservative indicator of ecosystem state, because extreme water quality values likely have disproportionately dire consequences for ecosystem function. However, the median concentration (115 mg L^{-1}) and the median prediction (94 mg L^{-1}) provide more than enough information to indicate that the lake is subject to anthropogenic chloride inputs.

The relationships between mean/median chloride concentration, maximum chloride concentration, and the threshold of chloride toxicity are central to our motivation in undertaking a predictive chloride model. Although we do not model or report on maximum chloride concentrations, maximums can greatly exceed median chloride concentrations, as shown for Diamond Lake, MN (Figure 7a). Many governing agencies have water quality criteria that set thresholds for toxicity. For instance, the U.S. EPA threshold for chronic chloride toxicity is 230 mg L^{-1}, while the acute toxicity threshold is 860 mg L^{-1}.[15] The Canadian water quality guidelines for freshwater set the long-term chloride exposure limit at 120 mg L^{-1}, and the short-term limit at 640 mg L^{-1}.[57] These thresholds are based on laboratory tests of select species to a range of chloride concentrations and represent a conservative percentile at which most species survive. In reality, salt sensitivity of freshwater organisms varies by taxa, with invertebrates typically being more sensitive than fish.[58] Some species, such as the glochidia of certain freshwater mussels[59] are impacted at chloride concentrations below the EPA water quality thresholds, while other species, such as invasive Asian Clams (*Corbicula fluminea*) and the Common Reed (*Phragmites australis*) can thrive above these thresholds.[60,61] The fact that some invasive taxa are salt tolerant is particularly concerning as they may be able to outcompete more sensitive, native taxa. In addition, other water quality characteristics (e.g., water hardness) can influence the vulnerability of species to chloride toxicity. Studies have found that the presence of major cations in freshwater (Na^+ and Ca^{2+}) can mitigate ion toxicity to invertebrates and fish.[58,62−64] Given the temporal variability of chloride at high concentrations, we suggest lakes predicted to have median chloride concentrations above 50 mg L^{-1} are a good starting point for field verification of chloride contamination.

The rampant use of salt persists in part due to the scarcity of incentives or policies aimed at limiting salt use, and it comes with large economic, environmental, and public health costs.[65] Our results indicated that 82% of 49,432 lakes >4 ha across a 17-state area from Maine to Minnesota are likely near, or at, natural chloride concentrations. The remaining 1972 lakes may be at risk due to salinization and can serve as a starting point for locating and assessing vulnerable ecosystems. Our model fit identified two major predictors of chloride concentrations: developed areas/roads and agricultural crops. We recognize that the model does not take into account other potential point sources, like wastewater discharge, which includes water softener effluent, or industrial sources. Nonetheless, the robust model fit indicates that we are capturing the major sources of chloride to most lakes. With more observational data, we could improve the model fit and prediction confidence (Figure S7).

Managing lakes to minimize chloride toxicity is not as simple as maintaining concentrations below a given threshold. There is significant seasonality and spatial heterogeneity in chloride concentrations, as well as the presence of at-risk resident biota. Lake managers and governing agencies are encouraged to use model predictions to augment their monitoring strategies to identify and protect freshwater ecosystems threatened by salinization.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.9b07718.

Supporting methods, discussion, and figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Hilary A. Dugan − *Center for Limnology, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States;* orcid.org/0000-0003-4674-1149; Email: hdugan@wisc.edu

Authors

Nicholas K. Skaff − *Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824, United States*

Jonathan P. Doubek − *School of Natural Resources & Environment and Center for Freshwater Research and Education, Lake Superior State University, Sault Sainte Marie, Michigan 49783, United States*

Sarah L. Bartlett − *NEW Water, Green Bay, Wisconsin 54302, United States*

Samantha M. Burke − *University of Guelph, School of Environmental Sciences, Guelph, Ontario N1G 2W1, Canada; Aquatic Contaminants Research Division, Environment & Climate Change Canada, Burlington, Ontario L7S 1A1, Canada;* orcid.org/0000-0002-7941-2096

Flora E. Krivak-Tetley − *Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire 03768, United States*

Jamie C. Summers − *WSP Canada Incorporated, Toronto, Ontario M4P 1E4, Canada*

Paul C. Hanson − *Center for Limnology, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States*

Kathleen C. Weathers − *Cary Institute of Ecosystem Studies, Millbrook, New York 12545, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.9b07718

Author Contributions

H.A.D., N.K.S., J.P.D., S.L.B., S.M.B., F.E.K., J.C.S., and K.C.W. initiated the ideas that led to this manuscript. H.A.D. contributed data and collected and harmonized data sets. H.A.D. and N.K.S. wrote the model code and performed all analyses. All authors contributed to writing and/or editing of the manuscript.

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

Notes

Observational data, model outputs, and QRF R code are archived and publicly available at doi: 10.6073/pasta/d581077c23dcfe33f48baf835339fa2f.

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research, as a continuation of GLEON Fellowship Program working groups (National Science Foundation [NSF] EF-1137327 and EF-1702991), would not have been possible without data contributions from a range of agencies, and the foresight and hard work of the LAGOS group (NSF EF-1638679). Chloride data were acquired from the Minnesota Pollution Control Agency, the University of Rhode Island Watershed Watch, and the North Temperate Lakes LTER (NSF DEB-1440297). The University of Wisconsin Data Science Initiative provided additional support for processing WQP data. Thank you to the Cary Institute of Ecosystem Studies, and Shannon LaDeau, for early brainstorming sessions, and Thomas Shannon and Alex Mathers for Wisconsin lake sampling.

■ REFERENCES

(1) Jackson, R. B.; Jobbágy, E. G. From Icy Roads to Salt Streams. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (38), 14487−14488.

(2) Corsi, S. R.; De Cicco, L. A.; Lutz, M. A.; Hirsch, R. M. River Chloride Trends in Snow-Affected Urban Watersheds: Increasing Concentrations Outpace Urban Growth Rate and Are Common among All Seasons. *Sci. Total Environ.* **2015**, *508*, 488−497.

(3) Kaushal, S. S.; Groffman, P. M.; Likens, G. E.; Belt, K. T.; Stack, W. P.; Kelly, V. R.; Band, L. E.; Fisher, G. T. Increased Salinization of Fresh Water in the Northeastern United States. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (38), 13517−13520.

(4) Chapra, S. C.; Dove, A.; Warren, G. J. Long-Term Trends of Great Lakes Major Ion Chemistry. *J. Great Lakes Res.* **2012**, *38* (3), 550−560.

(5) Dugan, H. A.; Bartlett, S. L.; Burke, S. M.; Doubek, J.; Krivak-Tetley, F.; Skaff, N. K.; Summers, J. C.; Farrell, K.; McCullough, I.; Morales-Williams, A.; Roberts, D.; Ouyang, Z.; Scordo, F.; Hanson, P. C.; Weathers, K. C. Salting Our Freshwater Lakes. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (17), 4453−4458.

(6) Kaushal, S. S.; Likens, G. E.; Pace, M. L.; Utz, R. M.; Haq, S.; Gorman, J.; Grese, M. Freshwater Salinization Syndrome on a Continental Scale. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (4), E574−E583.

(7) Brakke, D. F.; Landers, D. H.; Eilers, J. M. Chemical and Physical Characteristics of Lakes in the Northeastern United States. *Environ. Sci. Technol.* **1988**, *22*, 155−163.

(8) Eilers, J. M.; Brakke, D. F.; Landers, D. H. Chemical and Physical Characteristics of Lakes in the Upper Midwest, United States. *Environ. Sci. Technol.* **1988**, *22*, 164−172.

(9) Elphick, J. R. F.; Bergh, K. D.; Bailey, H. C. Chronic Toxicity of Chloride to Freshwater Species: Effects of Hardness and Implications for Water Quality Guidelines. *Environ. Toxicol. Chem.* **2011**, *30* (1), 239−246.

(10) Van Meter, R. J.; Swan, C. M. Road Salts as Environmental Constraints in Urban Pond Food Webs. *PLoS One* **2014**, *9* (2), e90168.

(11) Herbert, E.; Boon, P.; Burgin, A. J.; Neubauer, S. C.; Franklin, R. B.; Ardon, M.; Hopfensperger, K. N.; Lamers, L.; Gell, P. A Global Perspective on Wetland Salinization: Ecological Consequences of a Growing Threat to Freshwater Wetlands. *Ecosphere* **2015**, *6* (10), art206.

(12) Buchwalter, D. B.; Clements, W. H.; Luoma, S. N. Modernizing Water Quality Criteria in the United States: A Need to Expand the Definition of Acceptable Data. *Environ. Toxicol. Chem.* **2017**, *36* (2), 285−291.

(13) Castillo, A. M.; Sharpe, D. M. T.; Ghalambor, C. K.; De León, L. F. Exploring the Effects of Salinization on Trophic Diversity in Freshwater Ecosystems: A Quantitative Review. *Hydrobiologia*; Springer International Publishing, February 1, 2018; pp 1−17 DOI: 10.1007/s10750-017-3403-0.

(14) Cormier, S. M.; Zheng, L.; Flaherty, C. M. A Field-Based Model of the Relationship between Extirpation of Salt-Intolerant Benthic Invertebrates and Background Conductivity. *Sci. Total Environ.* **2018**, *633*, 1629−1636.

(15) EPA. *Ambient Water Quality Criteria for Chloride*; US Environmental Protection Agency: Washington DC, 1988.

(16) Vineis, P.; Chan, Q.; Khan, A. Climate Change Impacts on Water Salinity and Health. *J. Epidemiol. Glob. Health* **2011**, *1* (1), 5−10.

(17) Norrström, A. C.; Jacks, G. Concentration and Fractionation of Heavy Metals in Roadside Soils Receiving De-Icing Salts. *Sci. Total Environ.* **1998**, *218* (2−3), 161−174.

(18) Bäckström, M.; Karlsson, S.; Bäckman, L.; Folkeson, L.; Lind, B. Mobilisation of Heavy Metals by Deicing Salts in a Roadside Environment. *Water Res.* **2004**, *38* (3), 720−732.

(19) Schuler, M. S.; Relyea, R. A. A Review of the Combined Threats of Road Salts and Heavy Metals to Freshwater Systems. *BioScience* **2018**, *68* (5), 327−335.

(20) Overbo, A.; Heger, S.; Kyser, S.; Asleson, B.; Gulliver, J. *Chloride Contributions from Water Softeners and Other Domestic, Commercial, Industrial, and Agricultural Sources to Minnesota Waters*; Minneapolis, MN, 2019.

(21) Hotchkiss, E. R.; Sadro, S.; Hanson, P. C. Toward a More Integrative Perspective on Carbon Metabolism across Lentic and Lotic Inland Waters. *Limnol. Oceanogr. Lett.* **2018**, *3* (3), 57−63.

(22) Morgan, R. P.; Kline, K. M.; Kline, M. J.; Cushman, S. F.; Sell, M. T.; Weitzell, R. E.; Churchill, J. B. Stream Conductivity: Relationships to Land Use, Chloride, and Fishes in Maryland Streams. *North Am. J. Fish. Manag.* **2012**, *32* (5), 941−952.

(23) Kelly, V. R.; Findlay, S. E.; Hamilton, S. K.; Lovett, G. M.; Weathers, K. C. Seasonal and Long-Term Dynamics in Stream Water Sodium Chloride Concentrations and the Effectiveness of Road Salt Best Management Practices. *Water, Air, Soil Pollut.* **2019**, *230* (1), 13.

(24) Kelting, D. L.; Laxson, C. L.; Yerger, E. C. Regional Analysis of the Effect of Paved Roads on Sodium and Chloride in Lakes. *Water Res.* **2012**, *46* (8), 2749−2758.

(25) Scott, R.; Goulden, T.; Letman, M.; Hayward, J.; Jamieson, R. Long-Term Evaluation of the Impact of Urbanization on Chloride Levels in Lakes in a Temperate Region. *J. Environ. Manage.* **2019**, *244*, 285−293.

(26) Baker, M. E.; Schley, M. L.; Sexton, J. O. Impacts of Expanding Impervious Surface on Specific Conductance in Urbanizing Streams. *Water Resour. Res.* **2019**, *55* (8), 6482−6498.

(27) Clear Roads. *Annual Survey of State Winter Maintenance Data*; 2018.

(28) Winslow, L. A.; Read, J. S.; Hanson, P. C.; Stanley, E. H. Lake Shoreline in the Contiguous United States: Quantity, Distribution and Sensitivity to Observation Resolution. *Freshwater Biol.* **2014**, *59* (2), 213−223.

(29) Read, E. K.; Carr, L.; De Cicco, L.; Dugan, H. A.; Hanson, P. C.; Hart, J. A.; Kreft, J.; Read, J. S.; Winslow, L. A. Water Quality Data for National-Scale Aquatic Research: The Water Quality Portal. *Water Resour. Res.* **2017**, *53* (2), 1735−1745.

(30) Sprague, L. A.; Oelsner, G. P.; Argue, D. M. Challenges with Secondary Use of Multi-Source Water-Quality Data in the United States. *Water Res.* **2017**, *110*, 252−261.

(31) Soranno, P. A.; Bacon, L. C.; Beauchene, M.; Bednar, K. E.; Bissell, E. G.; Boudreau, C. K.; Boyer, M. G.; Bremigan, M. T.; Carpenter, S. R.; Carr, J. W.; Cheruvelil, K. S.; Christel, S. T.; Claucherty, M.; Collins, S. M.; Conroy, J. D.; Downing, J. A.; Dukett, J.; Fergus, C. E.; Filstrup, C. T.; Funk, C.; Gonzalez, M. J.; Green, L. T.; Gries, C.; Halfman, J. D.; Hamilton, S. K.; Hanson, P. C.; Henry, E. N.; Herron, E. M.; Hockings, C.; Jackson, J. R.; Jacobson-Hedin, K.; Janus, L. L.; Jones, W. W.; Jones, J. R.; Keson, C. M.; King, K. B. S.; Kishbaugh, S. A.; Lapierre, J.-F.; Lathrop, B.; Latimore, J. A.; Lee, Y.; Lottig, N. R.; Lynch, J. A.; Matthews, L. J.; McDowell, W. H.; Moore, K. E. B.; Neff, B. P.; Nelson, S. J.; Oliver, S. K.; Pace, M. L.; Pierson, D. C.; Poisson, A. C.;

Pollard, A. I.; Post, D. M.; Reyes, P. O.; Rosenberry, D. O.; Roy, K. M.; Rudstam, L. G.; Sarnelle, O.; Schuldt, N. J.; Scott, C. E.; Skaff, N. K.; Smith, N. J.; Spinelli, N. R.; Stachelek, J. J.; Stanley, E. H.; Stoddard, J. L.; Stopyak, S. B.; Stow, C. A.; Tallant, J. M.; Tan, P.-N.; Thorpe, A. P.; Vanni, M. J.; Wagner, T.; Watkins, G.; Weathers, K. C.; Webster, K. E.; White, J. D.; Wilmes, M. K.; Yuan, S. LAGOS-NE: A Multi-Scaled Geospatial and Temporal Database of Lake Ecological Context and Water Quality for Thousands of US Lakes. *GigaScience* **2017**, *6* (12), 1−22.

(32) Dugan, H. A.; Summers, J. C.; Skaff, N. K.; Krivak-Tetley, F. E.; Doubek, J. P.; Burke, S. M.; Bartlett, S. L.; Arvola, L.; Jarjanazi, H.; Korponai, J.; Kleeberg, A.; Monet, G.; Monteith, D.; Moore, K.; Rogora, M.; Hanson, P. C.; Weathers, K. C. Long-Term Chloride Concentrations in North American and European Freshwater Lakes. *Sci. Data* **2017**, *4*, 170101.

(33) U.S. EPA. National Lakes Assessment: A Collaborative Survey of the Nation's Lakes. 2018.

(34) Wright, M. N.; Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1−17.

(35) Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983−999.

(36) Nicodemus, K. K.; Malley, J. D.; Strobl, C.; Ziegler, A. The Behaviour of Random Forest Permutation-Based Variable Importance Measures under Predictor Correlation. *BMC Bioinf.* **2010**, *11* (1), 110.

(37) Peck, D. V.; Olsen, A. R.; Weber, M. H.; Paulsen, S. G.; Peterson, C.; Holdsworth, S. M. Survey Design and Extent Estimates for the National Lakes Assessment. *Freshw. Sci.* **2013**, *32* (4), 1231−1245.

(38) Stanley, E. H.; Collins, S. M.; Lottig, N. R.; Oliver, S. K.; Webster, K. E.; Cheruvelil, K. S.; Soranno, P. A. Biases in Lake Water Quality Sampling and Implications for Macroscale Research. *Limnol. Oceanogr.* **2019**, *64* (4), 1572−1585.

(39) Todd, A. K.; Kaltenecker, M. G. Warm Season Chloride Concentrations in Stream Habitats of Freshwater Mussel Species at Risk. *Environ. Pollut.* **2012**, *171*, 199−206.

(40) Daley, M. L.; Potter, J. D.; McDowell, W. H. Salinization of Urbanizing New Hampshire Streams and Groundwater: Effects of Road Salt and Hydrologic Variability. *J. North Am. Benthol. Soc.* **2009**, *28* (4), 929−940.

(41) Müller, B.; Gächter, R. Increasing Chloride Concentrations in Lake Constance: Characterization of Sources and Estimation of Loads. *Aquat. Sci.* **2012**, *74* (1), 101−112.

(42) Estévez, E.; Rodríguez-Castillo, T.; González-Ferreras, A. M.; Cañedo-Argüelles, M.; Barquín, J. Drivers of Spatio-Temporal Patterns of Salinity in Spanish Rivers: A Nationwide Assessment. *Philos. Trans. R. Soc., B* **2019**, *374* (1764), 20180022.

(43) Cherkauer, D. S.; McKereghan, P. F.; Schalch, L. H. Delivery of Chloride and Nitrate by Ground Water to the Great Lakes: Case Study for the Door Peninsula, Wisconsin. *Groundwater* **1992**, *30* (6), 885−894.

(44) USDA. Fertilizer Use and Price. *US Dep. Agric. Econ. Res. Serv.* 2019.

(45) LRRB. *Minnesota Snowa and Ice Control: Field Handbook for Snowplow Operators*; Manual Num.; Minnesota Local Road Research Board. MnDOT Office of Maintenance: St. Paul, MN, 2012.

(46) Notaro, M.; Lorenz, D.; Hoving, C.; Schummer, M. Twenty-First-Century Projections of Snowfall and Winter Severity across Central-Eastern North America. *J. Clim.* **2014**, *27* (17), 6526−6550.

(47) Oliver, S. K.; Soranno, P. A.; Fergus, C. E.; Wagner, T.; Winslow, L. A.; Scott, C. E.; Webster, K. E.; Downing, J. A.; Stanley, E. H. Prediction of Lake Depth across a 17-State Region in the United States. *Inland Waters* **2016**, *6* (3), 314−324.

(48) Evans, C. D.; Futter, M. N.; Moldan, F.; Valinia, S.; Frogbrook, Z.; Kothawala, D. N. Variability in Organic Carbon Reactivity across Lake Residence Time and Trophic Gradients. *Nat. Geosci.* **2017**, *10* (11), 832−835.

(49) Zwart, J. A.; Sebestyen, S. D.; Solomon, C. T.; Jones, S. E. The Influence of Hydrologic Residence Time on Lake Carbon Cycling Dynamics Following Extreme Precipitation Events. *Ecosystems* **2017**, *20* (5), 1000−1014.

(50) Vollenweider, R. Advances in Defining Critical Loading Levels for Phosphorus in Lake Eutrophication. *Mem. dell'Istituto Ital. di Idrobiol.* **1976**, *33*, 53−83.

(51) Bowser, C. J. Groundwater Pathways for Chloride Pollution of Lakes. *In Chemical Deicers and the Environment*; D'Itri, F. M., Ed.; Lewis Publishers Inc.: Chelsea, MI, 1992; pp 283−301.

(52) Bock, A. R.; Falcone, J. A.; Oelsner, G. P. Estimates of Road Salt Application across the Conterminous United States, 1992−2015—ScienceBase-Catalog. *U.S. Geol. Surv. Data Release* **2018**, DOI: 10.5066/P96IX385.

(53) Kelly, V. R.; Lovett, G. M.; Weathers, K. C.; Findlay, S. E. G.; Strayer, D. L.; Burns, D. J.; Likens, G. E. Long-Term Sodium Chloride Retention in a Rural Watershed: Legacy Effects of Road Salt on Streamwater Concentration. *Environ. Sci. Technol.* **2008**, *42*, 410−415.

(54) Ficker, H.; Luger, M.; Pamminger-Lahnsteiner, B.; Achleitner, D.; Jagsch, A.; Gassner, H. Diluting a Salty Soup: Impact of Long-Lasting Salt Pollution on a Deep Alpine Lake (Traunsee, Austria) and the Downside of Recent Recovery from Salinization. *Aquat. Sci.* **2019**, *81* (1), 7.

(55) Schulz, C.-J.; Cañedo-Argüelles, M. Lost in Translation: The German Literature on Freshwater Salinization. *Philos. Trans. R. Soc., B* **2019**, *374* (1764), 20180007.

(56) Moore, J.; Fanelli, R. M.; Sekellick, A. J. High-Frequency Data Reveal Deicing Salts Drive Elevated Specific Conductance and Chloride along with Pervasive and Frequent Exceedances of the U.S. Environmental Protection Agency Aquatic Life Criteria for Chloride in Urban Streams. *Environ. Sci. Technol.* **2020**, *54* (2), 778−789.

(57) Canadian Council of Ministers of the Environment. *Canadian Water Quality Guidelines for the Protection of Aquatic Life: Chloride*; Winnipeg, MB, 2011.

(58) Soucek, D. J.; Linton, T. K.; Tarr, C. D.; Dickinson, A.; Wickramanayake, N.; Delos, C. G.; Cruz, L. A. Influence of Water Hardness and Sulfate on the Acute Toxicity of Chloride to Sensitive Freshwater Invertebrates. *Environ. Toxicol. Chem.* **2011**, *30* (4), 930−938.

(59) Gillis, P. L. Assessing the Toxicity of Sodium Chloride to the Glochidia of Freshwater Mussels: Implications for Salinization of Surface Waters. *Environ. Pollut.* **2011**, *159* (6), 1702−1708.

(60) Coldsnow, K. D.; Relyea, R. A. Toxicity of Various Road-Deicing Salts to Asian Clams (*Corbicula Fluminea*). *Environ. Toxicol. Chem.* **2018**, *37* (7), 1839−1845.

(61) Vasquez, E. A.; Glenn, E. P.; Guntenspergen, G. R.; Brown, J. J.; Nelson, S. G. Salt Tolerance and Osmotic Adjustment of Spartina Alterniflora (Poaceae) and the Invasive M Haplotype of Phragmites Australis (Poaceae) along a Salinity Gradient. *Am. J. Bot.* **2006**, *93* (12), 1784−1790.

(62) Dwyer, F. J.; Burch, S. A.; Ingersoll, C. G.; Hunn, J. B. Toxicity of Trace Element and Salinity Mixtures to Striped Bass (*Morone Saxatilis*) and *Daphnia Magna*. *Environ. Toxicol. Chem.* **1992**, *11* (4), 513−520.

(63) Mount, D. R.; Gulley, D. D.; Hockett, J. R.; Garrison, T. D.; Evans, J. M. Statistical Models to Predict the Toxicity of Major Ions to *Ceriodaphnia Dubia, Daphnia Magna* and *Pimephales Promelas* (Fathead Minnows). *Environ. Toxicol. Chem.* **1997**, *16* (10), 2009−2019.

(64) Mount, D. R.; Erickson, R. J.; Highland, T. L.; Hockett, J. R.; Hoff, D. J.; Jenson, C. T.; Norberg-King, T. J.; Peterson, K. N.; Polaske, Z. M.; Wisniewski, S. The Acute Toxicity of Major Ion Salts to *Ceriodaphnia Dubia* : I. Influence of Background Water Chemistry. *Environ. Toxicol. Chem.* **2016**, *35* (12), 3039−3057.

(65) Strifling, D. A. Reducing Chloride Discharges to Surface Water and Groundwater: A Menu of Options for Policymakers. *Environmental Law*; Lewis & Clark Law School, 2018; pp 167−210 DOI: 10.2307/44633535.

(66) Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10* (1), 439−446.

https://dx.doi.org/10.1021/acs.est.9b07718
Environ. Sci. Technol. XXXX, XXX, XXX−XXX