Predicting atrazine concentrations in waterbodies across the contiguous United States: The importance of land use, hydrology, and water physicochemistry

Marieke Beaulieu,1* Hubert Cabana,1 Zofia Taranu,2 Yannick Huot3
1Department of Civil Engineering, Université de Sherbrooke, Sherbrooke, Canada
2Environment and Climate Change Canada, Aquatic Contaminants Research Division, Montreal, Canada
3Department of Geomatics, Université de Sherbrooke, Sherbrooke, Canada

Abstract

Atrazine contamination is ubiquitous in North American surface waters, but the dependency of the herbicide’s distribution on landscape and within-lake processes is currently poorly known. We sought to identify novel predictors of atrazine and to build a coherent framework to model its concentration in waterbodies through the development of binomial-gamma hurdle models and LASSO regression models. We constructed models for over 900 waterbodies in the contiguous United States using data from the 2012 U.S. EPA National Lake Assessment, the 2012 U.S. Department of Agriculture CropScape and the Global HydroLAB HydroLAKES databases. Atrazine was detected in 32% of U.S. waterbodies, with a mean concentration of 0.17 μg L⁻¹ when detected. The two-part hurdle model explained as much as 75% of the variance in atrazine across a spatially and temporally heterogeneous landscape. Three predictors explained 31% of the variability in atrazine detection in U.S. waterbodies, where the proportion of corn + soy cultures in the watershed was the most important variable. Once atrazine was detected, our models explained an additional 29% of the variability in atrazine concentrations, where the estimated areal weight of atrazine application (kg atrazine km²) in the watershed was the most important predictor. Spatially, water quality variables associated with eutrophication were linked to increased levels of atrazine contamination while cooler water temperatures and natural lakes and landscapes were associated with decreased levels of contamination. Our results suggest that changes in land-use practices may be the most effective way to mitigate atrazine contamination in waterbodies.

Atrazine, an herbicide, is the second most used pesticide in the United States after glyphosate (Atwood and Paisley-Jones 2017) and is an essential additive in corn production as conducted through industrial agriculture practices. Relatedly, atrazine is also an important aquatic micropollutant; among the molecules commonly screened, it has long been one of the most frequently detected in North American waterbodies (Solomon et al. 1996; Baldwin et al. 2016). The relationship between the atrazine application on land and the downstream concentrations in inland lakes and reservoirs, however, is difficult to determine. Although herbicide application volumes can be estimated from the volumes purchased and the area of cultivation, once applied, the environmental fate of these additives can be quite variable among sites due to differences in soil, water, and air transport. Consequently, atrazine can be transported hundreds of miles from where it is applied depending on the transport route (Glotfelty et al. 1987; Chernyak et al. 1996). In soils, although its half-life can be as short as 5 d (Vryzas et al. 2012), the persistence of atrazine residues over several decades suggests that realized half-lives in the environment are significantly longer than predicted (Jablonowski et al. 2011; Vonberg et al. 2014a). This would explain why atrazine remains detected in German aquifers 21 yr after its ban in 1991 (Vonberg et al. 2014b). Reported half-lives of atrazine in surface waters vary wide widely from 3.2 d (Kosinski 1984) to 168 d (Bacci et al. 1989) depending on the conditions. Because of its widespread use, distribution and potential persistence, atrazine contamination of freshwater ecosystems is likely to remain a concern for decades to come, and yet may be a difficult endpoint to predict.

Atrazine is a likely endocrine disruptor in humans, based on amphibian and mammalian models (Hayes et al. 2010;
Brekenridge et al. 2015) and has been associated with cytotoxicity in human and animals cells (Liu et al. 2006; Huang et al. 2014), as well as oxidative stress (Jin et al. 2014) and dopaminergic effects (Li et al. 2015) in animal models. Although the risks beget by the presence of low concentration of atrazine in waterbodies used as drinking water sources and for recreation will likely never be fully ascertained, monitoring and subsequent predictive modeling can identify areas of potential concern and factors most associated with the presence of atrazine. The cumulative impact of atrazine, other contaminants and their interactions in the environment, and other anthropogenic pressures such as eutrophication due to agricultural practices and climate change may represent a further risk to the health of these systems.

While atrazine contamination of streams, rivers and groundwater of the contiguous United States has been monitored through the United States Geological Survey (USGS) National water-quality Synthesis project (NAWQ) since 1992 (Toccalino et al. 2014; Ryberg and Gilliom 2015), a comparative survey in lake ecosystems was only conducted a decade later, during the second campaign of the U.S. Environmental Protection Agency’s (USEPA) National Lakes Assessment (NLA) in 2012 (USEPA 2016). This dataset presents an unprecedented opportunity to improve our understanding of the fate of atrazine within a structured landscape, quantifying the among-site variability in contamination levels and how this may relate to differences in soil, land use, water transport (changes in hydrology due to climate), and local waterbody parameters (e.g., physicochemical parameters). Furthermore, as environmental micropollutant sample extraction and processing can be a difficult and costly endeavor, notably with regards to complex matrixes (Souza-Silva et al. 2015), we will explore the extent of our ability to predict this important water contaminant, allowing us to extend our knowledge to sites that have not yet been sampled. This in turn, if successful, should allow managers and decision makers to assess the best approaches to reduce contamination in surface waters.

Within watersheds, hydrological models have typically been used to estimate the fate of atrazine (Vazquez-Amabile et al. 2006; Larose et al. 2007), whereas in aquatic environments the spatial distribution of atrazine has, to date, only been modeled by two studies. Stackelberg et al. (2012) modeled atrazine concentrations in shallow groundwater (concentrations ranged from < 0.004 to 4.7 μg L⁻¹), and found that atrazine concentrations are largely controlled by the history of atrazine use in relation to the groundwater age. Yun and Qian (2015) found that differences in temporal atrazine trends in streams and lakes among sites were best explained by crop cover, where atrazine concentrations increased over time with an expansion of cropland in the southern United States but decreased or remained stable in northern regions of the United States where atrazine use was greatest in the 1990s. The concentration they found in lakes between 1990 and 2010 varied widely, from 1 × 10⁻⁶ to 1 × 10² μg L⁻¹. While these studies were key in identifying spatial heterogeneity in atrazine trends in groundwater, streams or lakes, and its potential link to land-use practices, the approaches did not consider how lakes may filter local to regional processes, and how accounting for these catchment-lake interactions may improve predictions at the continental scale. To understand the main factors contributing to the variable concentrations in lakes and to allow the modeling of broad-scale patterns of atrazine in waterbodies at the continental scale we developed empirical models that consider both landscape and within-waterbody processes. To do this we will consider a suite of potentially important variables from existing datasets and a watershed specific land-use dataset generated in this study. We sought to identify predictors of atrazine contamination of lakes and reservoirs and assess the respective importance of driver variables in predicting atrazine concentrations in waterbodies through the proper modeling of such response variables.

Methods

The dataset and response variable

We analyzed data from 1135 lakes and reservoirs in 48 U.S. states sampled in 2012 through the NLA survey. The NLA’s objectives are to identify the current health status of lakes, impoundments and reservoirs in the United States and assess how this has changed through time, and, to this end, sites sampled are randomly selected to be representative of U.S. waterbodies (Pollard et al. 2018). The United States contains a large number of human-altered lake systems and it has been shown that these different types of systems can impact biotic and abiotic conditions (Beaulieu et al. 2013; Hayes et al. 2017). During this second survey (the first occurring in 2007) surface water atrazine concentrations were measured. In each site, integrated water samples were collected from the euphotic zone (calculated as twice the Secchi depth), up to depths of 2 m. Samples were conserved in 60 mL white HDPE bottles and stored on ice for up to 1 week before shipping for analysis. Quantification was performed using the Abraxis Atrazine ELISA kit (Abraxis LLC).

Predictor variables

We considered a number of variables in predicting atrazine concentrations in waterbodies (Fig. 1, Table 1 [see online Supporting Information Material Table S1 for details]). These variables were transformed to meet conditions of normality and heteroscedascity when possible. The variables considered varied in type and data origin and were grouped in the following broad categories.

Land use and atrazine use intensity

Due to the uniquely anthropogenic origin of atrazine, we assumed that the most likely variables to have an influence on atrazine concentrations in waterbodies would be related to its use in the watershed of a given site. Waterbodies within watersheds with a high density of corn culture should have higher atrazine concentrations as atrazine is used quasi-exclusively
on corn crops (USGS, https://water.usgs.gov/nawqa/pnsp/usage/maps/show_map.php?year=2016&map=ATRAZINE).

To calculate land use within a waterbody’s watershed we used the 2012 U.S. Department of Agriculture (USDA) CropScape data layer (https://nassgeodata.gmu.edu/CropScape/), using the NLA watershed polygons as the limits. We calculated a number of watershed specific land-use parameters (expressed as the fraction of the total land cover in the watershed) including total agriculture ($P_{Agr}$), developed land ($P_{Dev}$), natural landscape ($P_{nat}$), which included forested areas, wetlands, shrublands and switchgrasses and water/perennial ice ($P_{Wat}$). As atrazine is used quasi-exclusively on corn crops, we also considered the area of corn culture relative to the watershed area ($P_{C}$). However, given that the dominant cropping system in the

![Distribution of atrazine and selected variables across the contiguous United States.](image)

Fig 1. Distribution of atrazine and selected variables across the contiguous United States. (A) Atrazine concentrations from the NLA 2012 campaign, (B) Estimated atrazine application surface density, (C) Waterbody type, (D–F) Lake properties, (G and H) Land use in the lake’s watershed, and (I) Lake morphology. To help visualization, for continuous variables (all variables except lake type) the study area was split into 2500 bins (50 vertically and 50 horizontally) the average value within each bin is displayed.
Table 1. Variables considered in models (see Table S1 for units, data ranges, and data transformation).

| Category (dataset) | Variables |
|--------------------|-----------|
| Atrazine (NLA 2012) | Atrazine [below/above detection limit] (ATR_PA), Atrazine concentration [detection limit (dl) = 0.046 μg L⁻¹] (logATR) |
| Ecoregion (NLA 2012) | Eastern Highlands [NAP = Northern Appalachians; SAP = Southern Appalachians], Plains and Lowlands |
| Water quality (NLA 2012) | Total phosphorus (TP), Total nitrogen (TN), pH, silicate (Si), sodium (Na), potassium (K), sulfate (SO₄²⁻), turbidity (Turb), chloride (Cl), calcium (Ca), aluminum (Al), acid neutralizing capacity (ANC), ammonium (NH₄), conductivity (Cond), color (col), dissolved organic carbon (DOC) |
| Physical/spatial (NLA 2012) | Lake type, watershed area (AW), surface water temperature (TS), day of the year (DOY), depth (D), day of the year (DOY), latitude (LAT), longitude (LON), urban (Lake intersects urban cluster or urban area), elevation (Elev), lake area (A_L), lake area : watershed area (A_L : A_W), brun-Vaïsälä buoyancy frequency (N) |
| Hydrology (Hydrolake 1.0) | Total volume (Vol), discharge average (Dis), slope [average within 100 m of lake polygon] (slope), shoreline development [ratio between shoreline length and the circumference of a circle of the same area, increases with increasing shoreline complexity] (SDev), volume reservoir [reported reservoir volume] (Vol_R), average depth [ratio between total lake volume and lake area] (Depth_AV), residence time [ratio between total lake volume and average discharge] (ResTime) |
| Land use (2012 USDA Cropscape raster with NLA2012 watershed polygons) | Corn (P_C), corn or soy (P_CS), natural landscape [barren, forest, shrubland, wetlands] (P_LNat), agriculture [all crops] (P_Ag), developed land [open space to high intensity] (P_Dev), perennial ice/snow and open water (P_PWat) |
| Estimated atrazine application (2012 USGS NAWQA County-level scale atrazine use estimate data with NLA2012 watershed polygons) | Estimated area density of atrazine application at lake (EA_L), average estimated area density of atrazine application in watershed (EA_W), estimated mass of atrazine applied in the watershed (EA_M) |

U.S. Corn Belt, where over 85% of U.S. corn is produced, consists of 2-yr corn-soybean rotation (Grassini et al. 2015), we likewise considered the sum of soybeans and corn cultures (corn + soy [P_CS]) as the presence of soy on any given year can be indicative of the presence of corn on a previous year.

The 2012 USGS NAWQA county-level scale atrazine application estimate data (https://water.usgs.gov/nawqa/pnsp/usage/maps/index.php) was used to determine the estimated area density of atrazine application in direct vicinity of the waterbody (EA_L [kg km⁻²]) considering the density of application of the county in which the waterbody is located, the estimated average area density of atrazine application in the watershed (EA_W [kg km⁻²]) and the mass of atrazine applied in the watershed (EA_M [kg]). This data is obtained by calculating the median pesticide-by-crop use rate at the county level using farm surveys of pesticide use and estimates of harvested crop acres. These rates are then applied to the harvested acreage of each crop within the county (for detailed methods see Thelin and Stone 2013, Baker and Stone 2015).

Physical parameters (HydrolAKES data)

Lakes and reservoirs may act as retention basins for atrazine, with subsequent immobilization of the herbicide in the sediments (Qu et al. 2017). Because of this, water retention time might be important in predicting atrazine concentrations, with long retention times being associated with lower water-column concentrations, as atrazine is adsorbed to the sediments in addition to being degraded in the water column through microbial transformation (e.g., oxidative dealkylation and hydrolysis; Fenner et al. 2013).

To estimate the hydraulic residence time we used the hydroLAKES 1.0 database (http://www.hydrosheds.org/page/hydroLAKES) based on the global integrated water model WaterGAP (Messager et al. 2016). The hydroLAKES 1.0 database provides shoreline polygons for waterbodies around the world with a surface area of at least 10 ha using remote sensing (methods provided in Lehner and Messager 2016). In addition to residence time, the database provides information on shoreline length, average depth, water volume and elevation. We merged the NLA and hydroLAKES dataset for all polygons showing overlap. We further eliminated waterbodies for which we found differences greater than a factor of 50 for common variables (watershed area, waterbody area, and elevation).

Physical parameters (NLA 2012 data)

Physical parameters, notably temperature, may also be important predictors of atrazine concentration in lakes and reservoirs. In soils, warmer temperature increases atrazine
biodegradation up to 30°C (Strong et al. 2000; Krutz et al. 2008) but also favors atrazine sorption to soil surfaces (McGlamery and Slife 1966; Yue et al. 2017). Properties such as area, depth and water column stability could influence the amount of atrazine arriving in the waterbody and its distribution in the water column. We thus considered waterbody area \(A_w\), watershed area \(A_{sw}\), waterbody to watershed ratio \(A_{w:sw}\), waterbody depth \(D\), surface water temperature \(T_s\), and the Brunt-Väisälä water-column stability index \(N\) in our models.

We considered waterbody type (which consisted of three categories from the 2012 NLA dataset: lakes, impoundments, and reservoirs) as a possible predictor. Note that the lakes category was a merger of natural lakes, enhanced natural lakes and abandoned impoundments, due to the low occurrence of abandoned impoundments \((n = 4)\) and enhanced natural lakes \((n = 44)\). To help characterize the broad spatial distribution of waterbodies in the dataset, we selected spatial and landscape parameters including elevation \(\text{Elev}\), latitude \(\text{CLat}\), longitude \(\text{CLon}\), and a categorical variable grouping waterbodies according to their Wadeable Streams Assessment ECO9 ecoregion (Fig. 2). Lastly, to control for seasonality, we considered the Day of the year \(D\) of sampling, since punctual timing of atrazine applications during certain times of the year could influence the concentrations found in the aquatic environment.

**Water chemistry (NLA 2012)**

We also considered water chemical parameters that might have an influence on the fate of atrazine in the water column. While the co-occurrence of nutrient and herbicide pollution from agricultural sources are likely to make the former a predictor of the occurrence of the second, low water column pH could be associated with greater recalcitrance of atrazine, comparably to what is observed in soils (Houot et al. 2000; Krutz et al. 2010). Higher salinities have also been associated with decreased atrazine degradation (Solomon et al. 1996; Lin et al. 2008). We thus chose a number of chemical parameters with potential ecological effects on the atrazine concentrations in waterbodies (Table 1), including nutrient concentrations (total nitrogen \(\text{TN}\), total phosphorus \(\text{TP}\), ammonium \(\text{NH}_4\), and total nitrogen to total phosphorus ratio \(\text{TN:TP}\)), turbidity \(\text{Turb}\), conductivity \(\text{Cond}\), color \(\text{Color}\), dissolved organic carbon \(\text{DOC}\), acid neutralizing capacity \(\text{ANC}\) and \(\text{pH}\). We further considered the concentration of select ions (chloride \(\text{Cl}^-\), sodium \(\text{Na}^+\), calcium \(\text{Ca}^{2+}\), potassium \(\text{K}^+\), sulfate \(\text{SO}_4^{2-}\) and silica \(\text{SiO}_2\)).

**Statistical analyses**

A challenge to modeling environmental contaminant concentration data is that such data tends to be right-skewed and comprising a large number of samples that are below the detection values. Substitutions, imputation methods and non-parametric models have been used to develop regression models for distributions with below-detection values, but such methods introduce biases, notably when nondetects make up the bulk of the data (Helsel 2005). One approach to avoid the pitfalls of such methods is to analyze the data sequentially using a hurdle (or two-part) model, firstly seeking to explain which waterbodies are exceeding (yes or no) the detection limit, and secondly estimating concentrations above the detection limit. Although interest for such models for continuous data is not new (Aitchison 1955) they have not been heavily adopted for continuous data, in contrast to the literature analyzing count data (Mullahy 1986). Recent work by Taranu et al. (2017), in which microcystin models were developed, has demonstrated the value of the method for continuous environmental data.

**Hurdle models**

To identify the best predictors of atrazine in lakes and reservoirs, we used the modeling strategy adopted by Taranu...
et al. (2017) consisting of a binomial-gamma [ZAG] hurdle model (also known as a delta model). The hurdle model is comprised of two components: (1) A binomial generalized linear model (GLM) identifying whether or not atrazine was detected, and (2) a truncated gamma GLM model predicting the concentration of atrazine above the compound’s detection limit.

Prior to running the hurdle model (binomial and truncated gamma GLMs), data were transformed and scaled (see online Supporting Information Material Table S1). We first considered the presence/absence of atrazine above the detection limit through a binomial GLM; coding concentrations below the detection limit as zero and concentrations above as one. Secondly, we considered the log-transformed concentration values after the addition of an offset corresponding to the log-transformed value of the detection limit (0.046 μg L⁻¹) through a gamma GLM (sensu Brillem et al. 2016); gamma models are defined for a threshold value of zero and this transformation allowed us to consider the detection limit as the hurdle.

To identify significant predictor variables for both the binomial and gamma GLMs (i.e., parts one and two of the hurdle model), we used the dredge function of the MuMIn R package (Bartoń 2018) to generate a subset of 0 to 5 variables from all variables in each category (water quality, physical/spatial, land use; see Table 1), leading to a group of competing models for each category. From the models generated, we retained those that were within three Bayesian information criterion (BIC) units from the best model. While the Akaike information criterion (AIC) is more widely used and also introduces a penalty term for the number of parameters in a model, BIC offers a greater penalty, further limiting the risk of model over fitting, which can be of considerable concern when using model dredging for model selection. For each set of variables retained in each category, we retained the best model through model averaging, selecting variables with a relative importance greater than 50% based on the sum of “Akaike weights” over all models in a given category. Colinear variables (Pearson’s r > 0.5) were eliminated and the variance inflation factors (VIF) of the final models were tested to be inferior to 2.

The validation test of the final model residuals could not be conducted by standard methods given our use of GLMs. That is, GLM residuals are not as readily interpreted as for ordinary least-square linear models as the expected distribution of the data changes with fitted values, and standard residual plots cannot aid in determining whether the model is correctly specified. Thus, to test for patterns in the residuals, we used the testResiduals() and simulateResiduals() functions from the DHARMa R package (Hartig 2019) as they simulate scaled residuals from the fitted model. We also computed a pseudo $R^2$ value considering the proportion of variance explained by the model when compared to the null model.

**General model.** From the variables selected for each category of variables we constructed a general model predicting atrazine occurrence and concentrations across the continental United States. We felt it necessary to proceed in this manner (first selecting variables within each category, and then building a model joining all) as (1) we were interested in exploring the predictive power of each category, and (2) the number of combinations for model dredging grows exponentially with the number of predictors. Thus, including all variables in a single step was not computationally feasible.

**Continental and Plains-Lowlands models.** The area sampled by the NLA is divided in National Aquatic Resource Survey (NARS), reporting three regions, which are based on aggregated Wadeable Streams Assessment (WSA), reporting nine regions (Herlihy et al. 2008) (Table 1; Fig. 2: Eastern Highlands [NAP, SAP]; Plains and Lowlands [CPL, NPL, SPL, TPL, UMW]; Western Mountains and Xeric [WMT, XER]). However, based on patterns of atrazine use in North America, we expected atrazine concentrations to be significantly greater in the Plains and Lowlands ecoregions (five of the nine WSA ecoregions), which comprises the U.S. Corn Belt. Thus, in addition to modeling the continental dataset, we also modeled exclusively the Plains and Lowlands region, as we hypothesized that variable selection at the continental scale might be confounded by broad regional differences of the predictive variables considered.

**LASSO models**

To test whether the two hurdle models (Continental and Plains-Lowlands) were over-fitted to the region used in each analysis, we compared the hurdle model results with a least absolute shrinkage and selection operator (LASSO) regression analysis. LASSO regression is a regularization technique designed to aid generalizing models with complex relationships, such as multicollinearity, through the addition of a penalty to model parameters (Tibshirani 1996). We opted for LASSO regression as it performs variable selection by shrinking coefficients of less important variables to zero (Tibshirani 1996). This however comes at the cost of retaining only a single variable from a group of highly correlated predictors (in contrast, ridge regression retains all parameters, shrinking some close to, but not at, zero). In general, ridge regression works well if there are many variables believed to be of equal importance (when most predictors impact the response), whereas LASSO tends to do well if there are a small number of important variables (when only a few predictors actually influence the response). We used the R glmnet (Friedman et al. 2010) and HDtweedie (Qian et al. 2013) packages for the binomial and gamma parts of the LASSO models, respectively, using a 10-fold cross-validation to fit the models with scaled predictors, extracting the largest lambda value that remained within one standard error of the specified minimum mean cross-validated error (misclassification error for the binomial model and mean square error for the gamma model). We repeated this process 100 times, extracting the mean and standard deviation from the distribution of lambda values. This method allowed us to extract the average relative importance of the
variables selected while remaining parsimonious in our selection.

**Modeling additional atrazine thresholds.** We further used LASSO regression to compare the predictors associated with surpassing thresholds other than the limit of detection, as the conditions associated with high atrazine concentrations, for which there is a known environmental concern, might differ from those selected through our gamma models. We considered the 1.8 μg L⁻¹ guideline for the protection of aquatic life (CCME 1999) as well as threshold values of 0.1 and 0.2 μg L⁻¹ atrazine. The latter two values were derived from recent work studying the effects of atrazine and other herbicides on the photosynthetic efficiency of freshwater phytoplankton using fast repetition-rate fluorescence (Beaulieu et al. 2020). The atrazine concentration of 0.2 μg L⁻¹ was the threshold concentration for no observed effect concentration (NOEC), while 0.1 μg L⁻¹ was the guideline value derived from the maximal allowable toxicant concentration (MATC) of 1.1 μg L⁻¹.

**Results**

**Atrazine detection and concentration (NLA dataset)**

Atrazine data in the 2012 NLA dataset was right-skewed and consisted mainly of values below the detection limit (Fig. 2). Atrazine was detected in 32% of the lakes and reservoirs sampled, mainly in the Plains and Lowlands region where the frequency of detection reached 50%. In the Western Mountains and Xeric and Eastern Highlands, atrazine was detected in only 9% of the waterbodies sampled. When the molecule was detected, concentrations in the Western Mountains and Xeric and Eastern Highlands were also lower than in the Plains and Lowlands region (\( \bar{x} = 0.18 \mu g \) L⁻¹; median = 0.08 μg L⁻¹ vs. \( \bar{x} = 0.47 \mu g \) L⁻¹; median = 0.15 μg L⁻¹, respectively). Across all waterbodies sampled by the NLA in 2012, the maximum atrazine concentration of 9.7 μg L⁻¹ was found in the Temperate Plains ecoregion.

**Physical/spatial parameters (HydroLAKES vs. NLA data)**

A total of 761 lakes and reservoirs out of 1135 were used to test for the influence of hydrological variables provided in the HydroLAKES (HL) dataset (see online Supporting Information Material Fig. S2). A considerable number of sites (374) did not show overlap of the HL polygons and the NLA sampling sites (which was verified by seeking overlap among waterbody polygons from both datasets). While this was in large part due to the 10 ha size cut-off value in the HL dataset, 173 of these sites were established as larger than 10 ha in the NLA data. A further 32 sites were rejected due to large discrepancies between variables common to both datasets. We chose to discard sites where differences of over 1.7 orders of magnitude were observed (watershed area, waterbody area, and elevation). However, as we did not wish to discard the NLA sites that were not included in the HL dataset but we were interested in hydrological effects, we built two sets of hurdle models, one with the full NLA dataset (1135 sites) that did not consider hydrological variables and a second (761 sites) with the reduced dataset considering all predictor categories.

**Hurdle model—Continental and Plains-Lowlands data**

Given that the variable selection method (dredging) used to construct our hurdle models does not allow for missing data, we constructed the hurdle models using a subset of the data (949 NLA; 568 HL). The total variance explained by the hurdle models (binomial plus gamma parts) ranged from 60% (NLA) to 75% (HL) for the continental scale dataset, and from 46% (NLA) to 66% (HL) for the Plain-Lowlands region (Table 2; Fig. 3). Generalized variance inflation factors (gVIF) never surpassed 1.61 for any of the retained variables. Although reliance on this threshold for variance inflation factors has been criticized (O’Brien 2007), we found that values from our models were well below the value of 10 generally prescribed in the literature (Dormann et al. 2012). This suggests that multicollinearity is not of great concern for our models. The deviance explained by models based on data limited to the Plains and Lowlands region was generally smaller than that of models considering all ecoregions although the variables retained were generally similar, comprising markers of eutrophication (TN, TP, Turbidity, and pH), water temperature, land-use variables and the estimated atrazine application rates. Of the HL variables considered, only Shoreline development (ShD) was found to be a predictor of atrazine concentrations above the detection limit. While the proportion of corn and soy culture in the watershed (\( P_{CS} \)) was the most important predictor for the detection of atrazine, the estimated surface density of atrazine application (\( E_{A} \)) best explained the concentrations, when detected. The presence of natural landscape was only important in predicting atrazine concentrations, with no effect on atrazine detection. While \( E_{A} \) was a better predictor of atrazine concentrations than \( P_{CS} \) (as well as presence-absence in HL models), interchanging these variables in competing models generally yielded comparable results as assessed by BIC (Table 2). In the HL dataset however, \( E_{A} \) performed noticeably better. Analysis of the residuals suggests that our models fit the data satisfactorily (see online Supporting Information Material Fig. S3).

**General model—Continental data**

Based on our continental hurdle model, we constructed a general model predicting atrazine concentrations across the contiguous United States (Fig. 4A). We found the model to fit the data adequately, although atrazine concentrations in the northern part of the Temperate Plains were consistently underestimated while those in the southern part of the region were underestimated, at times severely (> 2 μg L⁻¹; Fig. 4B).

Among the water quality variables retained by the general hurdle model, most are indexes of eutrophication resulting from agricultural practices. Increased pH, turbidity, total nitrogen, potassium and silica were found to be predictors in our models, though they were better at explaining the presence
Table 2. Atrazine hurdle model of transformed data. Model parameter coefficients with standard errors in parenthesis are presented. Parameters in bold or identified with a star (☆) were also selected by LASSO regression. See Table 1 for subscript/variable definition and transformation used.

| ECO3 region          | Model component (predicted variable) | df | Intercept | P_{CS} | P_{Nat} | P_{Wat} | EAL | Natural lake | TN | TP | Turb | Na | Si | pH | T_{s} | SH_{dp} | Max GVIF | Pseudo | R^2 | BIC      |
|----------------------|--------------------------------------|----|-----------|--------|---------|---------|-----|-------------|----|----|------|----|----|-----|-------|---------|----------|--------|-----|---------|
| NLA dataset (n=949)  | All Binomial (ATR_PA)                | 945| -5.09     | 4.27   | *       |         |     | 1.74        | *  |     |       |    |    |     |       |         | 1.07     | 0.31   | 843.6   |
|                      |                                       |    | (0.57)    | (0.42) | †       |         |     | (0.23)      |    |     |       |    |    |     |       |         |          |        |       |
|                      | Gamma (logATR-logDL)                | 297| -2.19     | -0.85  | *       | 0.24    | *   | 0.28        | *  |     |       |    |    |     |       |         | 0.05     | 1.46   | 0.29   | 213.9 |
|                      |                                       |    | (0.36)    | (0.18) | †       | (0.04)†|     | (0.09)      |    |     |       |    |    |     |       |         |          |        |       |
|                      | PlnLow Binomial (ATR_PA)            | 529| -7.62     | 3.49   |         |         |     | 0.73        | 0.46| 0.10|       |    |    |     |       |         | 1.18     | 0.2    | 621.6  |
|                      |                                       |    | (1.57)    | (0.44)†| †       |         |     | (0.18)      |(0.15)| (0.03)|       |    |    |     |       |         |          |        |       |
|                      | Gamma (logATR-logDL)                | 260| †         | -0.80  | -0.40   |         |     | -0.89       | *   |     |       |    |    |     |       |         | 1.1      | 0.26   | 227.0  |
|                      |                                       |    | (1.56)    | (0.17) | †       | (0.04)†|     | (0.18)      |(0.15)| (0.03)|       |    |    |     |       |         |          |        |       |
| HL dataset (n=569)   | All Binomial (ATR_PA)                | 565| -0.79     |        |         |         |     | 1.13        | 1.87| *   |       |    |    |     |       |         | 2.11     | 1.03   | 0.35   | 496.9 |
|                      |                                       |    | (0.20)    | (0.12)†| †       |         |     | (0.29)      |(0.66)|         |       |    |    |     |       |         |          |        |       |
|                      | Gamma (logATR-logDL)                | 181| †         | -0.89  | -1.31   | 0.28    | -0.43| -0.86       | *   |     |       |    |    |     |       |         | 1.39     | 0.37   | 139    |
|                      |                                       |    | (0.23)    | (0.20) | †       | (0.05)†|     | (0.05)      |(0.13)|         |       |    |    |     |       |         |          |        |       |
|                      | PlnLow Binomial (ATR_PA)            | 324| -6.96     | 0.90   | *       |         |     | 0.95        | -0.67| 0.86|       |    |    |     |       |         | 1.48     | 0.25   | 371.3  |
|                      |                                       |    | (1.93)    | (0.14)†| †       |         |     | (0.26)      |(0.21)| (0.24)|       |    |    |     |       |         |          |        |       |
|                      | Gamma (logATR-logDL)                | 160| †         | -0.56  | -1.13   | 0.28    | -0.52| -0.56       | *   |     |       |    |    |     |       |         | 1.61     | 0.43   | 136    |
|                      |                                       |    | (1.96)    | (0.57) | †       | (0.05)†|     | (0.11)      |(0.10)|         |       |    |    |     |       |         |          |        |       |
|                      |                                       | 162| 0.84      | -0.85  | -0.87   | -0.77  |     |            |     |     |       |    |    |     |       |         | 1.37     | 0.38   | 138.6  |

* 0.01 > variable p-value > 0.05.
† Variable retained by dredging but not in the final model.
‡ Colinear variables.
than the concentration of atrazine. Among the physical variables tested, only surface water temperature and waterbody type were selected. Water temperature, especially surface water temperature, can be influenced by eutrophication as increased biomass at the surface of the water column absorbs more heat (Paerl and Paul 2012). Replacing surface water temperature by average temperature in the water-column still led to the retention of water temperature in the models, although the effects of this variable became weaker. While unimportant in predicting the detection of atrazine, waterbody type was consistently selected as predictor for atrazine concentration, with lower concentrations in lakes than in impoundments and reservoirs, where the latter two types showed no difference.

**Fig 3.** Hurdle models fitted curves and confidence intervals. For each panel, the response variable (atrazine presence-absence or concentration) is shown on the y axis, while the most important predictor is shown on the x axis, with fixed levels of other variables selected by the models represented as panels and colors. (A) Binomial component of the model for all ecoregions. (B) Binomial component of the model for the Plains and Lowlands ecoregion. (C) Gamma model component for all ecoregions. (D) Gamma model component for the Plains and Lowlands ecoregion. Atrazine detection limit is shown as a dotted line in C and D.

**Hurdle vs. LASSO models—Continental and Plains-Lowlands data**

The hurdle and LASSO models selected in large part the same variables, although LASSO selected a greater variety of
predictors of interest (Fig. 5). For the larger NLA dataset (Fig. 5A), the estimated atrazine applied, and the proportion of corn and soy remained the best predictors of the detection of atrazine, while the proportion of natural landscape was more important than that of crops in predicting atrazine concentrations once above detection limit. While ecoregion was not found to be a significant predictor of atrazine concentrations in the LASSO models, the detection of atrazine was more probable in the Temperate Plains region (ECO-TPL) for the LASSO models run both at the continental scale and on the subset of lakes and reservoirs within the Plains and Lowlands region. Within the Plains and Lowlands regions, waterbodies were significantly less likely to have detectable atrazine concentrations in the Northern Plains (ECO-NPL) than in the region as a whole.

For the smaller HL dataset (Fig. 5B), only two HL variables were selected, with greater shoreline lengths being associated with greater probabilities of detecting atrazine and shorter residence times being associated with higher atrazine concentrations. A notable difference between the LASSO models of this HL dataset and those of the NLA dataset was the greater importance of the proportion of water in the watershed (P[Wat]) in the HL dataset. P[Wat] was negatively correlated with the concentrations of atrazine observed.

**LASSO—Additional atrazine threshold models**

Seventeen sites had atrazine concentrations greater than the guideline value of 1.8 μg L\(^{-1}\). We chose instead to use a series of smaller thresholds to determine the best predictors in waterbodies surpassing these limits (DL [302 sites], 0.1 μg L\(^{-1}\) [189 sites], 0.2 μg L\(^{-1}\) [111 sites], 0.5 μg L\(^{-1}\) [63 sites], and 1 μg L\(^{-1}\) [37 sites]). As the concentration threshold increased, the deviance explained by our models also increased up to a threshold of 0.5 μg L\(^{-1}\), beyond which the number of sites above the detection limit was likely insufficient (Table 3).

While being located in the Temperate Plains was always among the best two predictors selected, the importance of corn + soy crops was replaced with the type of waterbody at increasing threshold values, while the importance of the absence of natural landscapes also tended to diminish (Fig. 6).

**Discussion**

**Atrazine and watershed land use**

Atrazine is an important water contaminant, with highly variable concentrations across U.S. lakes and reservoirs, and observed peak concentrations in the Plains and Lowlands ecoregions. Identifying the main drivers of this heterogeneity is key to helping curtail future increases of and exposure to this contaminant. We show that while the density of surface application of atrazine and land use within the watershed are the most important determinants of the continued increased concentration of atrazine in waterbodies, the probability of first observing atrazine above the detection limit was strongly dependent on waterbodies being located in the Temperate Plains, with this category gaining in importance when considering higher thresholds. This region has been shown to be especially vulnerable to herbicide transport due to the runoff potential of its soils characterized by the presence of weakly permeable clayspans (Lerch and Blanchard 2003; Ghidey et al. 2010). Atrazine contamination of surfaces waters is greatest in this region with detection in 73% of the sites sampled and concentrations when detected far superior to dataset averages (\(\bar{x} = 0.79\ \mu g\) L\(^{-1}\); \(M = 0.28\ \mu g\) L\(^{-1}\)). While previous studies have suggested that atrazine concentrations in U.S. Midwest lakes have seen the greatest decreases since the 1990s comparatively
Fig 5. Importance of predictor variables as selected by LASSO regression based on coefficients estimated for scaled parameters. (A) Models based on all ecoregions and on the Plains and Lowlands ecoregion of the NLA dataset. (B) Models based on all ecoregions and on the Plains and Lowlands ecoregion of the HL dataset, where hydroLAKES parameters are highlighted in pink text. The bar color identifies whether the value of the coefficient is positive or negative.

Table 3. LASSO threshold models.

| Threshold value | DL (0.046 µg L\(^{-1}\)) | 0.1 µg L\(^{-1}\) | 0.2 µg L\(^{-1}\) | 0.5 µg L\(^{-1}\) | 1.0 µg L\(^{-1}\) |
|-----------------|---------------------|----------------|----------------|----------------|----------------|
| NLA dataset (n=949) | | | | | |
| Number sites exceeding threshold | 302 | 189 | 111 | 63 | 37 |
| Optimized lambda | 0.05 | 0.02 | 0.01 | 0.01 | 0.04 |
| Number passes | 31 | 159 | 237 | 159 | 34 |
| Deviance explained | 0.30 | 0.40 | 0.46 | 0.48 | 0.23 |
| Null deviance | 1187.25 | 947.54 | 684.86 | 463.47 | 312.63 |
| HL dataset (n=569) | | | | | |
| Number sites exceeding threshold | 187 | 116 | 76 | 44 | 25 |
| Optimized lambda | 0.04 | 0.02 | 0.03 | 0.01 | 0.06 |
| Number passes | 64 | 173 | 302 | 126 | 25 |
| Deviance explained | 0.34 | 0.49 | 0.45 | 0.53 | 0.20 |
| Null deviance | 720.60 | 575.50 | 447.36 | 309.76 | 205.14 |
to increases in southern U.S. regions, the concentrations found in the aquatic environment of the former region remain higher (Yun and Qian 2015). Our results show that waterbodies of this region remain heavily polluted by atrazine and are sensitive to its application upstream.

In terms of types of land-use practices and predictive capabilities, our results demonstrate that knowledge of the area of certain crops within a waterbody’s watershed is a suitable replacement for the surface density of application of atrazine, which was found to be the best predictor of atrazine concentrations in lakes and reservoirs. This is of interest as the former data on crop area are far more widely available. The improvement of the models when considering both soy and corn as predictors of atrazine rather than just corn alone suggests that the legacy of atrazine use is important in evaluating atrazine contamination of lakes and reservoirs. While soya crops are often rotated with corn crops (Grassini et al. 2015), soya crops themselves are not treated with atrazine and the importance of this crop in our models is likely reflective of atrazine’s persistence in the environment across years, either in soil or in freshwater.
Similarly, natural landscapes were the only land cover types consistently associated with the absence of atrazine in waterbodies, which given the widespread use and persistence of atrazine in the environment, echo the importance of land-use activities on atrazine contamination in these systems. Natural landscapes in a waterbody’s watershed may be efficiently retaining and/or removing atrazine from the aquatic environment. While we grouped a number of natural landscapes, it may be of interest to focus on the specific effect of wetlands in the watershed as these landscape features, both natural and constructed, are recognized for their ability to remove pesticides from agricultural runoff (Vymazal and Nezínková 2015). Removal occurs mainly through microbial activity, which is dependent on water residence time (Tournebize et al. 2017). In the Plains and Lowlands region, although natural landscape was not a predictor of atrazine presence-absence, it was an important predictor (negative relationship) of the concentrations found. Thus, despite the substantial atrazine contamination in this region, natural remediation may be effective in decreasing concentrations, however, not to levels below the detection limit.

**Atrazine and water quality**

Although land use was the most important factor in predicting the detection of atrazine, a significant amount of the variability in the detection of atrazine was accounted for by parameters routinely used to monitor water quality. Unsurprisingly, water quality variables associated with agriculture, erosion and eutrophication, such as total nitrogen, phosphorus, turbidity and pH, were important predictors of the detection of atrazine (Carpenter et al. 1998). While we did not quantify soil erosion specifically, we failed to find an effect of the slope of the landscape adjacent (100 m) to the waterbody. Relationships between atrazine and water column chemistry were stronger at the continental scale compared to the ecoregion scale (Plains and Lowlands), most likely because of lake water chemistry differences between ecoregions. Interestingly, in contrast to the importance of nitrogen in predicting the detection of atrazine, increased potassium and decreased silica were associated with increasing atrazine concentrations at the continental scale. Potassium is supplemented in intensive agriculture, and silica loss is generally associated with eutrophication through the uptake of silica and subsequent sinking of diatom frustules. It has also been demonstrated that agricultural harvesting is an anthropogenic sink of biogenic silica, which is highly soluble comparatively to mineral silicates (Vandevenne et al. 2012). This is particularly relevant to the culture of corn and other species from the order Poales as they accumulate substantially more silica in their shoots than other angiosperms (Hodson et al. 2005).

**Atrazine and physical variables**

In terms of physical variables, surface water temperature was consistently an important predictor at the scales considered, with increased detection of atrazine, and greater concentrations found at higher water temperatures. The solubility of atrazine increases with temperature (Curren and King 2001), which could explain the importance of this physical variable in our models. As deep lakes tend to be cooler, this variable may also represent to an extent waterbody morphology.

Waterbody type was an increasingly important variable in our models at higher concentration thresholds where lakes were associated with smaller concentrations of atrazine compared to reservoirs and impoundments. Subsequent analysis did not reveal substantial difference in the distributions of the water chemistry or physical predictive variables between the waterbody types considered. At the scale of the contiguous U.S. reservoirs were associated with greater proportions of corn + soy cultures. However, these differences were not significant in the Plains and Lowlands region. Our analysis did not, however, focus on the proximity of corn and soy culture in the watershed to the waterbody. Lakes may be more distal to the cultivated land or better buffered from them.

Despite not finding significant effects of hydrological variables in our models, hydrological difference between waterbody types might explain differences between these categories. Reservoirs in the U.S. drain catchment areas on average 6.5 times larger than lakes with residence times half as long (Hayes et al. 2017). Shorter residence times make reservoirs intermediary on the river–lake continuum and, in these systems, increased flow could lead to greater concentrations as the potential of compound transformation decreases. Interestingly, ANOVA of differences in HydroLAKES variables by lake type followed by Tukey’s Honest difference test showed that natural lakes were characterized by smaller slopes ($F$-value = 9.19, $p$-value < 0.001) and shoreline development ratios ($F$-value = 23.23, $p$-value < 0.001) in the watershed. These results, supported by the inclusion of shoreline development ratios in our models, suggest that the smaller atrazine contamination of U.S. lakes may be caused partly by reduced erosion and influence of the surrounding landscape in these systems when compared to reservoirs and impoundments.

**Strength of multicategory approach**

As agricultural practices are responsible for the contamination of atrazine in aquatic ecosystems, it is not surprising that variables related to agriculture strongly predicted the occurrence of atrazine in lakes and reservoirs. However, physico-chemical properties of waterbodies as well as catchment hydrology explained an additional and independent portion of the atrazine variability across this broad limnoscape. Remaining, unexplained variability might be due to methodological and fine-scale factors not quantified here. For instance, limitation of analytical methods may fail to detect low atrazine in some waterbodies, and thus misclassify the sites as atrazine absent in our binomial models. Furthermore, site-specific differences in atrazine degradation rates may increase.
noise of the gamma truncated portion of our models, notably since a high degree of variability in atrazine degradation rates has been observed in the literature as a function of land-use legacies and the evolution of atrazine-degrading capabilities in the environment (Udiković-Kolić et al. 2012). It has been demonstrated that pre-exposure of soils to atrazine increases the degradation of the molecule over time, as microbial consortia transitioned to organisms capable of transforming the molecule (Houot et al. 2000; Krutz et al. 2010). This suggests that weighing the importance of corn in the watershed by the number of years it has been extensively cultivated in the watershed might improve our ability to predict the presence of atrazine. Our results also support the inclusion of the main metabolites of atrazine in future studies, as this would better characterize the level of contamination in the environment, especially given the known toxicity of atrazine’s main metabolites desethylatrazine and desisopropylationazine (Ralston-Hooper et al. 2009; Magnusson et al. 2012; Alberto et al. 2017). Our models cannot account for the variability in atrazine volumes usage on individual farms or their adherence to best management practices, which may prove important. The mechanisms by which atrazine reaches waterbodies were likewise not considered in our models, with the exception of watershed slope within 100 m of the site from the HydroLAKES dataset. The inclusion of precipitation-related variables might provide further improvements, as atrazine concentrations in surface waters follow seasonal peaks driven by rain events (Wittmer et al. 2010). Our analysis also overlooked the importance of soil properties; it has been shown that soil runoff potential is a critical factor affecting herbicide transport, with the Temperate Plains region of the U.S. Corn Belt being particularly vulnerable (Lerch and Blanchard 2003). However, fluctuations of the water table within a given field can have great effects on overall losses. Freitas et al. (2008) demonstrated that avoiding atrazine applications on a region corresponding to 1% of a given field could decrease losses by 30%, which suggests that small-scale effects can drive effects at much larger scales, which may never be fully accounted by large continental models. Yet, despite these limitations, the four categories of variables examined in the present study explained as much as 75% of the total variability of occurrence and concentration of atrazine across the contiguous United States, with promising findings in terms of avenues for remediation (e.g., natural landscapes).

**Conclusions**

Atrazine contamination of U.S. lakes and reservoirs is widespread and was detected during the U.S. EPA’s 2012 National Lake Assessment at concentrations greater than 0.046 μg L⁻¹ (detection limit) in a third of sites. Its detection and concentrations were well modeled and predicted by a binomial-gamma hurdle model, and best explained by the density of atrazine application on land, though the proportion of corn + soy in the watershed performed comparatively well as a predictor. Water quality variables related to eutrophication (phosphorus, nitrogen, potassium, silica, turbidity and pH) were likewise associated with atrazine contamination, as were surface water and average water-column temperature.

While the probability of detecting atrazine did not differ between lakes and impoundments/reservoirs, observed concentrations were significantly lower in lakes. This may be influenced by the, on average, smaller shoreline development at these sites, which may be indicative of reduced erosion and influence of the surrounding landscape. Similarly, the presence of natural landscapes in the watershed was associated with decreased levels of contamination.

The ability to predict the detection of atrazine increased at higher threshold values, suggesting that it is easier to predict the presence of atrazine at concentrations representing a greater risk. Sites from the Temperate Plains region, likely due to soil properties, appear particularly vulnerable to atrazine contamination at higher concentrations.

These results (see Fig. 3C) quantify the extent to which reducing the amounts applied to the crop or reducing fraction of the lands dedicated to corn production will reduce the concentration of atrazine in lakes and reservoirs. Barring a full ban on atrazine (which, we note, might benefit farmers [see Ackerman et al. 2014]), reducing application could be achieved using alternative herbicides or changes in agricultural practices (Ponisio et al. 2014; Hunt et al. 2017). Furthermore, reducing land dedicated to corn production could be achieved by diversifying cultures, longer rotations between corn production years or reduction of the fraction of land dedicated to agriculture. Ethanol production from corn, which account for about 35% to 40% of the U.S. annual production (Mumm et al. 2014) and its controversial mandated use in fuel is a key policy that should be examined in this context.

**References**

Ackerman, F., M. Whited, and P. Knight. 2014. Would banning atrazine benefit farmers? Int. J. Occup. Environ. Health **20**: 61–70. doi:http://doi.org/10.1073/pnas.0812835106

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. J. Am. Stat. Assoc. **50**: 901–908. doi:http://doi.org/10.1111/j.1468-0084.1953.mp15001001.x

Alberto, D., I. Couée, C. Sulmon, and G. Gouesbet. 2017. Root-level exposure reveals multiple physiological toxicity of triazine xenobiotics in *Arabidopsis thaliana*. J. Plant Physiol. **212**: 105–114. doi:http://doi.org/10.1016/j.jplph.2017.01.013

Atwood, D., and C. Paisley-Jones. 2017. Pesticides industry sales and usage. U.S. Environmental Protection Agency, p. 1–32.
Baldwin, A. K., S. R. Corsi, L. A. De Cicco, P. L. Lenaker, M. A. Beaulieu et al. Modeling atrazine in U.S. waterbodies. CCME. 1999. Canadian water quality guidelines for the protection of aquatic life—atrazine, p. 1–4. In Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Baker, N. T., and W. W. Stone. 2015. Estimated Annual Agricultural Pesticide Use for Counties of the Conterminous United States, 2008–12. U.S. Geological Survey Data Series 907. 9.

Baldwin, A. K., S. R. Corsi, L. A. De Cicco, P. L. Lenaker, M. A. Lutz, D. J. Sullivan, and K. D. Richards. 2016. Organic contaminants in Great Lakes tributaries: Prevalence and potential aquatic toxicity. Sci. Total Environ. 554–555: 42–52. doi:http://dx.doi.org/10.1016/j.scitotenv.2016.02.137

Bartoń, K. 2018. MuMIn: Multi-Model Inference. R package version 1.42.1 [accessed 2019 July 01]. Available from https://CRAN.R-project.org/package=MuMIn

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. 58: 1736–1746. doi:http://dx.doi.org/10.4319/lo.2013.58.5.1736

Beaulieu, M., H. Cabana, and Y. Huot. 2020. Adverse effects of atrazine, DCMU and metolachlor on phytoplankton cultures and communities at environmentally relevant concentrations using fast repetition rate fluorescence. Sci. Total Environ. 712: 136239. doi:http://dx.doi.org/10.1016/j.scitotenv.2019.136239

Breckenridge, C. B., P. Sawhney Coder, M. O. Tisdal, J. W. Simpkins, K. D. Yi, C. D. Foradori, and R. J. Handa. 2015. Effect of age, duration of exposure, and dose of atrazine on sexual maturation and the luteinizing hormone surge in the female Sprague-Dawley rat. Birth Defects Res. B Dev. Reprod. Toxicol. 104: 204–217. doi:10.1011/j.1439-0264.2006.00743.x

Brillemann, S. L., M. J. Crowther, M. T. May, M. Gompels, and K. R. Abrams. 2016. Joint longitudinal hurdle and time-to-event models: An application related to viral load and duration of the first treatment regimen in patients with HIV initiating therapy. Stat. Med. 35: 3583–3594. doi:http://dx.doi.org/10.1002/imd.2064763.2012.725468

Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. 8: 559–568. doi:10.1890/1051-0761(1998)008[559:NPOSSW]2.0.CO;2

CCME. 1999. Canadian water quality guidelines for the protection of aquatic life—atrazine, p. 1–4. In Canadian environmental quality guidelines, 1999, Canadian Council of Ministers of the Environment, Winnipeg.

Chernyak, S. M., C. P. Rice, and L. L. McConnell. 1996. Evidence of currently-used pesticides in air, ice, fog, seawater and surface microlayer in the Bering and Chukchi seas. Mar. Pollut. Bull. 32: 410–419. doi:http://dx.doi.org/10.1016/0025-326X(95)00216-A

Curren, M. S. S., and J. W. King. 2001. Solubility of Triazine pesticides in pure and modified subcritical water. Anal. Chem. 73: 740–745. doi:http://dx.doi.org/10.1021/ac000906n

Dormann, C. F., et al. 2012. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. Ecography 36: 27–46. doi:http://dx.doi.org/10.1111/j.2041-210X.2009.00001.x

Fenner, K., S. Canonica, L. P. Wackett, and M. Elsner. 2013. Evaluating pesticide degradation in the environment: Blind spots and emerging opportunities. Science 341: 752–758. doi:http://dx.doi.org/10.1126/science.1236281

Freitas, L. G., H. Singer, S. R. Müller, R. P. Schwarzenbach, and C. Stamm. 2008. Source area effects on herbicide losses to surface waters—a case study in the Swiss plateau. Agrc. Ecosyst. Environ. 128: 177–184. doi:http://dx.doi.org/10.1016/j.agee.2008.06.014

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33: 1–22 [accessed 2019 July 01]. Available from http://www.jstatsoft.org/v33/i01/

Ghidery, F., C. Baffaut, R. N. Lerch, N. R. Kitchen, E. J. Sadler, and K. A. Suduth. 2010. Herbicide transport to surface runoff from a claypan soil: Scaling from plots to fields. J Soil Water Conserv 65: 168–179. doi:https://doi.org/10.2489/jswc.65.3.168

Glottelfty, D. E., J. N. Seiler, and L. A. Liljedahl. 1987. Pesticides in fog. Nature 325: 602–605. doi:10.1038/325602a0

Grassini, P., J. E. Specht, M. Tollenaar, I. Ciampitti, and K. G. Cassman. 2015. High-yield maize-soybean cropping systems in the US Corn Belt, p. 17–41. ISBN 978-0-12-417104-6. In V. O. Sadas and D. Calderini [eds.], Crop physiology. London: Academic Press/Elsevier. doi:10.1016/B978-0-12-417104-6.00002-9

Hartg, F. 2019. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level/Mixed) Regression Models. R package version 0.2.3 [accessed 2019 July 01]. Available from https://CRAN.R-project.org/package=DHARMa

Hayes, T. B., and others. 2010. Atrazine induces complete feminization and chemical castration in male African clawed frogs (Xenopus laevis). Proc. Natl. Acad. Sci. 107: 4612–4617. doi:http://dx.doi.org/10.1073/pnas.2005.11.032

Hayes, N. M., B. R. Deemer, J. R. Corman, N. R. Razavi, and K. E. Strock. 2017. Key differences between lakes and reservoirs modify climate signals: A case for a new conceptual model. Limnol. Oceanogr. Lett. 2: 47–62. doi:http://doi.org/10.1002/lol.1.1.406

Helsel, D. R. 2005. More than obvious: Better methods for interpreting nondetect data. Environ. Sci. Technol. 39: 419A–423A. doi:http://dx.doi.org/10.1021/es053368a

Herlihy, A. T., S. G. Paulsen, J. V. Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: The challenges of applying a
Lerch, R. N., and P. E. Blanchard. 2003. Watershed vulnerability to herbicide transport in northern Missouri and southern Iowa streams. Environ. Sci. Technol. 37: 5518–5527. doi:http://doi.org/10.1021/es030431s
Li, Y.-S., X. He, K. Ma, Y.-P. Wu, and B.-X. Li. 2015. The effect of exposure to atrazine on dopaminergic development in pubertal male SD rats. Birth Defects Res. B Dev. Reprod. Toxicol. 104: 184–189. doi:http://doi.org/10.1126/science.276.5310.248
Lin, T., Y. Wen, L. Jiang, J. Li, S. Yang, and Q. Zhou. 2008. Study of atrazine degradation in subsurface flow constructed wetland under different salinity. Chemosphere 72: 122–128. doi:http://doi.org/10.1016/j.chemosphere.2007.11.059
Liu, X.-M., J.-Z. Shao, L.-X. Xiang, and X.-Y. Chen. 2006. Cytotoxic effects and apoptosis induction of atrazine in a grass carp (Ctenopharyngodon idellus) cell line. Environ. Toxicol. 21: 80–89. doi:http://doi.org/10.1086/515999
Magnusson, M. M., K. K. Heimann, M. M. Ridd, and A. P. A. Negri. 2012. Chronic herbicide exposures affect the sensitivity and community structure of tropical benthic microalgae. Mar. Pollut. Bull. 65: 363–372. doi:http://doi.org/10.1016/j.marpolbul.2011.09.029
McGlamery, M. D., and F. W. Slife. 1966. The adsorption and desorption of atrazine as affected by pH, temperature, and concentration. Weeds 14: 237–239. doi:http://doi.org/10.2307/4040921
Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. 7: 13603. doi:http://doi.org/10.1038/ncomms13603
Mullahy, J. 1986. Specification and testing of some modified count data models. J. Econom. 33: 341–365. doi:http://doi.org/10.1016/0304-4076(86)90002-3
Mumm, R. H., P. D. Goldsmith, K. D. Rausch, and H. H. Stein. 2014. Land usage attributed to corn ethanol production in the United States: Sensitivity to technological advances in corn grain yield, ethanol conversion, and co-product utilization. Biotechnol. Biofuels 7: 1–17. doi:http://doi.org/10.1186/1754-6834-7-61
O’Brien, R. M. 2007. A caution regarding rules of thumb for variance inflation factors. Qual. Quant. 41: 673–690. doi:http://doi.org/10.2307/1267205
Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. 46: 1349–1363. doi:http://doi.org/10.1016/j.watres.2011.08.002
Pollard, A. I., S. E. Hampton, and D. M. Leech. 2018. The promise and potential of continental-scale limnology using the U.S. Environmental Protection Agency’s National Lakes Assessment. Limnol. Oceanogr. Bull. 27: 36–41. doi:http://doi.org/10.1002/lob.10238
Ponisio, L. C., L. K. M’gonigle, K. C. Mace, J. Palomino, P. de Valpine, and C. Kremen. 2014. Diversification practices reduce organic to conventional yield gap. Proc. R. Soc. B: Biol. Sci. 282: 20141396. doi:http://doi.org/10.1126/science.1185383
Qian, W., Y. Yang, and H. Zou. 2013. HDtweedie: The Lasso for the Tweedie’s Compound Poisson Model Using an IRLS-BMD Algorithm. R package version 1.1 [accessed 2019 July 01]. Available from https://CRAN.R-project.org/package=HDtweedie

Qu, M., H. Li, N. Li, G. Liu, J. Zhao, Y. Hua, and D. Zhu. 2017. Distribution of atrazine and its phyto remediation by submerged macrophytes in lake sediments. Chemosphere 168: 1515–1522. doi:http://doi.org/10.1016/j.chemosphere.2016.11.164

Ralston-Hooper, K., J. Hardy, L. Hahn, H. Ochoa-Acuña, L. S. Lee, R. Mollenhauer, and M. S. Sepúlveda. 2009. Acute and chronic toxicity of atrazine and its metabolites deethylatrazine and deisopropylatrazine on aquatic organisms. Ecotoxicology 18: 899–905. doi:http://doi.org/10.1016/j.ecoleng.2016.02.014

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. B. Methodol. 58: 267–288. doi:http://doi.org/10.2307/2346178?refreqid=search-gateway:bc3ab8e8952347f96b97c3ae19fc3a5e

Toccalino, P. L., R. J. Gilliom, B. D. Lindsey, and M. G. Rupert. 2014. Pesticides in groundwater of the United States: Decadal-scale changes, 1993–2011. Groundwater 52: 112–125. doi:http://doi.org/10.1016/j.jconhyd.2007.05.007

Tournebize, J., C. Chaumont, and U. Mander. 2017. Implications for constructed wetlands to mitigate nitrate and pesticide pollution in agricultural drained watersheds. Ecol. Eng. 103: 415–425. doi:http://doi.org/10.1016/j.ecoleng.2016.02.014

U.S. Environmental Protection Agency. 2016. National Aquatic Resource Surveys, National Lakes Assessment 2012 (data and metadata files) [accessed 2019 July 1]. Available from http://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

Usman, K. R., and others. 1996. Ecological risk assessment of atrazine and metolachlor within soil pro

Vandevenne, F., E. Struyf, W. Clymans, and P. Meire. 2012. Evolution of atrazine-degrading capabilities in the environment. Appl. Microbiol. Biotechnol. 96: 1175–1189. doi:http://doi.org/10.1007/jbiod.2011.08.010

Vazquez-Amabile, G., B. A. Engel, and D. C. Flanagan. 2006. Modeling and risk analysis of nonpoint-source pollution caused by atrazine using SWAT. Trans. Am. Soc. Agric. Biol. Eng. 49: 667–678. doi:http://doi.org/10.13031/2013.20486

Vonberg, D., J. Vanderborgh, A. Lelickens, S. Köppchen, T. Pütz, P. Burauel, and H. Vereecken. 2014a. Atrazine soil core residue analysis from an agricultural field 21 years after its ban. J. Environ. Qual. 43: 1450–1459. doi:http://doi.org/10.2134/jeq2013.12.0497

Vonberg, D., J. Vanderborgh, N. Cremer, T. Pütz, M. Herbst, and H. Vereecken. 2014b. 20 years of long-term atrazine monitoring in a shallow aquifer in western Germany. Water Res. 50: 294–306. doi:http://doi.org/10.1016/j.watres.2013.10.032

Vryzas, Z., E. N. Papadakis, K. Oriakli, T. P. Moysiadis, and E. Papadopoulou-Mourkidou. 2012. Biotransformation of atrazine and metolachlor within soil profile and changes in microbial communities. Chemosphere 89: 1330–1338. doi:http://doi.org/10.1016/j.chemosphere.2012.05.087

Vymazal, J., and T. Baležnínová. 2015. The use of constructed wetlands for removal of pesticides from agricultural runoff and drainage: A review. Environ. Int. 75: 11–20. doi:http://doi.org/10.1016/j.envint.2014.10.026

Wittmer, I. K., H. P. Bader, R. Scheidegger, H. Singer, A. Lück, I. Hanke, C. Carlsson, and C. Stamm. 2010. Significance of urban and agricultural land use for biocide and pesticide dynamics in surface waters. Water Res. 44: 2850–2862. doi:http://doi.org/10.1016/j.watres.2010.01.030

Yue, L., C. Ge, D. Feng, H. Yu, H. Deng, and B. Fu. 2017. Adsorption–desorption behavior of atrazine on agricultural soils in China. J. Environ. Sci. 57: 180–189. doi:http://doi.org/10.1016/j.jes.2016.11.002
Yun, J., and S. S. Qian. 2015. A hierarchical model for estimating long-term trend of atrazine concentration in the surface water of the contiguous U.S. JAWRA J. Am. Water Resour. Assoc. 51: 1128–1137. doi: http://doi.org/10.1021/es1038563

Acknowledgments
We thank the reviewers for their constructive comments and criticism. We are grateful to all the people invested in making scientific investigation accessible to all. The National Lakes Assessment 2012 data were a result of the collective efforts of dedicated field crews, laboratory staff, data management and quality control staff, analysts and many others from EPA, states, tribes, federal agencies, universities, and other organizations. Please contact nars-hq@epa.gov with any questions. HydroLAKES is publicly available for download at http://www.hydrosheds.org and is free for scientific, educational, and other uses. We would like to thank the Natural Sciences and Engineering Research Council of Canada (ESD34595492014) and the Fonds de Recherche du Québec—Nature et Technologies (175372) for funding.

Conflict of Interest
None declared.

Submitted 30 July 2019
Revised 08 January 2020
Accepted 05 July 2020

Associate editor: Thomas Kristo froeb