Assessing statistical and spatial validity of sediment survey design and sampling densities: examples from Lake Erie

Danielle E. Mitchell, K. Wayne Forsythe, Chris H. Marvin and Debbie A. Burniston

ABSTRACT

Spatial interpolation methods translate sediment contamination point data into informative area-based visualizations. Lake Erie was first sampled in 1971 based on a survey grid of 263 locations. Due to procedural costs, the 2014 survey was reduced to 34 sampling locations mostly located in deep offshore regions of the lake. Using the 1971 dataset, this study identifies the minimum sampling density at which statistically valid, and spatially accurate predictions can be made using ordinary kriging. Randomly down-sampled subsets at 10% intervals of the 1971 survey were created to include at least one set of data points with a smaller sample size than that of the 2014 dataset. Regression analyses of predicted contamination values assessed spatial autocorrelation between kriged surfaces created from the down-sampled subsets and the original dataset. Subsets at 10% and 20% of the original data density accurately predicted 51% and 75% (respectively) of the original dataset's predictions. Subsets representing 70%, 80% and 90% of the original data density accurately predicted 88%, 90% and 97% of the original dataset's predictions. Although all subsets proved to be statistically valid, sampling densities below 0.002 locations/km^2 are likely to create very generalized contamination maps from which environmental decisions might not be justified.

Key words | geostatistics, Great Lakes, kriging, sediment contamination, spatial analysis, visualization

Danielle E. Mitchell (corresponding author)
K. Wayne Forsythe
Department of Geography and Environmental Studies,
Ryerson University,
350 Victoria Street, Toronto, ON M5B 2K3,
Canada
E-mail: *dmitch@ryerson.ca*

Chris H. Marvin
Aquatic Contaminants Research Division, Water Science and Technology Directorate,
Environment and Climate Change Canada,
867 Lakeshore Road, Burlington, ON L7S 1A1,
Canada

Debbie A. Burniston
Water Quality Monitoring and Surveillance Division, Science and Technology Branch,
Environment and Climate Change Canada,
867 Lakeshore Road, Burlington, ON L7S 1A1,
Canada

INTRODUCTION

Harmful organic and inorganic pollutants from industrial, agricultural and urban sources have contaminated the water, sediment and ecosystems of Lake Erie since the mid-1800s (Evers *et al.* 2011). Non-essential heavy metals including lead (Pb), cadmium (Cd) and mercury (Hg) are highly toxic and persistent throughout these environments. Heavy metal poisoning can cause severe health risks to both human and wildlife populations through bioaccumulation, or direct contact with polluted water and sediments (Wuana & Okieimen 2011).

Lake Erie is the most densely populated and highly industrialized Great Lake while also being the smallest by volume at 484 km^3 (GLIN 2017). Along its shoreline are several major cities including Toledo, Ohio, Cleveland, Ohio, Erie, Pennsylvania and Buffalo, New York. Since the enactment of the Great Lakes Water Quality Agreement in 1972, efforts have been made by both the Canadian and American governments, and environmental agencies to remediate historical and contemporary ecosystem pollution throughout the Great Lakes basin (IJC 1972, 2002). Toxic substances are more persistent in the deepest parts of Lake Erie, which continue to contaminate the basin's ecosystems (Forsythe & Marvin 2005; Forsythe *et al.* 2010). Historic and long-term monitoring and

doi: 10.2166/wqrj.2018.029

mapping of these contaminants is crucial for ongoing remediation efforts.

Data collection

Since 1968, the Water Science and Technology Directorate (previously the National Water Research Institute) and the Great Lakes Surface Water Surveillance Program have collected and analysed sediment samples for organic and inorganic pollutants throughout the Great Lakes (Marvin *et al.* 2003, 2004; Rukavina *et al.* 2013; Forsythe *et al.* 2016a). In 1971, sediment samples were collected from Lake Erie using a square survey grid. The survey grid was composed of 263 sampling points spaced approximately 10 km apart. The most recent Lake Erie sediment survey was conducted in 2014, where the sampling density was significantly reduced due to the increased cost of sampling procedures (Forsythe *et al.* 2010). Sampling points did not follow the original sampling grid used in 1971; instead, a restricted number of sampling points was randomly selected that were mostly located in deep offshore regions of Lake Erie, placing emphasis on non-point source pollution. Only 34 sampling points make up the 2014 survey, almost eight times smaller than the initial dataset from 1971.

Geovisualization of sediment contamination

Dot maps are a traditional medium used to display sediment contamination (Largueche 2006; Forsythe & Marvin 2009; Gawedzki & Forsythe 2012). Contamination patterns can be observed in point form; however, trends and underlying factors for contaminant distribution (i.e., bathymetry and lake circulation patterns) are not recognizable or available to assist with interpretation (Forsythe *et al.* 2016c). Due to advances in geospatial computation, spatial interpolation methods, such as kriging, are now used to model estimated contamination in a continuous data layer (Goovaerts 1999; Lark & Ferguson 2004; Forsythe *et al.* 2016b, 2016c). Not only are interpolated surfaces cost- and time-effective (Simpson & Wu 2014), they also work to translate disjunctive point data into communicable visualizations of spatial patterns (Hessl *et al.* 2007). Doing so also expands the communication of scientific research to a much broader audience. For example, geovisualizations of trends, patterns and processes are significant pieces of information in policy development (Lark & Ferguson 2004; Forsythe *et al.* 2016a, 2016b), pollution control and management (Wu *et al.* 2011; Khosh Eghbal 2014), environmental science (Forsythe *et al.* 2016b, 2016c) and risk assessment (Markus & McBratney 2001).

Kriging

Developed in 1951, kriging was initially used to predict ore reserves in mines (Largueche 2006; Gawedzki & Forsythe 2012). Since then, the interpolation method has been successfully applied to other fields of study including soil studies (Gotway *et al.* 1996; Yan *et al.* 2007; Simpson & Wu 2014), hydrology (Olea 1984; De Solla *et al.* 2012), environmental management (Chang *et al.* 1998; Lui *et al.* 2006), zoology (Villard & Maurer 1996; McKenney *et al.* 1998), forestry (Chen *et al.* 2012), horticulture (Dille *et al.* 2002) and those specific to contamination in the Great Lakes (Forsythe & Marvin 2005; Forsythe *et al.* 2016c). There is however some disagreement among researchers as to the minimum survey samples required for kriging to produce statistically valid and meaningful interpolated surfaces (Chang *et al.* 1998; Li & Heap 2008; Simpson & Wu 2014).

According to Webster & Oliver (1993), a minimum of 150 and up to 225 sampling points are required to conduct reliable spatial interpolations, whereas Burrough & McDonnell (1998) suggest between 50 and 100 sampling points. Kriging with less than 50 sampling points was determined to create 'erratic' predictions (Webster & Oliver 1993) and at 19, kriging predicts 'results [which] may only be an artefact' of the actual process (Chang *et al.* 1998). All things considered, interpolation accuracies are not guaranteed to improve with an exhaustive sample density, and remind us of the cost- and time-effectiveness of collecting limited survey samples (Simpson & Wu 2014). Producing a statistically valid and meaningful interpolated surface is not only influenced by the number of sample points available (Webster & Oliver 1993; Painter *et al.* 2001), but also the sampling density (Gotway *et al.* 1996; Ramsey 1997; Chang *et al.* 1998; McKenney *et al.* 1998; Li & Heap 2008), sampling pattern (Villard & Maurer 1996; Goovaerts 1999; Yan *et al.* 2007) and sampling direction (Clark 1982; McBratney & Webster 1986; Ouyang *et al.* 2003).

Sampling surveys in other fields of research have also been affected by the increased cost of conducting field

work (Chang *et al.* 1998; McKenney *et al.* 1998; Lui *et al.* 2006; Yan *et al.* 2007; De Solla *et al.* 2012; Simpson & Wu 2014). In the interest of prolonging current and future monitoring efforts, researchers have attempted to determine the smallest possible sampling densities (from which to conduct kriging interpolation) by down-sampling their original datasets into randomly, or systematically, selected subsets (Chang *et al.* 1998; McKenney *et al.* 1998; Yan *et al.* 2007). Studies by Chang *et al.* (1998), McKenney *et al.* (1998) and Lui *et al.* (2006) were not all successful in producing reliable predictions in their respective experiments. Decreasing sampling density to 25% or less of the original data in McKenney *et al.* (1998) resulted in low prediction accuracies, whereas Chang *et al.* (1998) reported 'sufficient' results from point densities reduced by 20% and 40% of the original dataset. There is not an absolute minimum number of samples from which to perform kriging. Expert knowledge of the study site and phenomena acting within its boundaries is also essential for determining the representational accuracies of an interpolated surface (Goovaerts 1999; Forsythe *et al.* 2016b, 2016c).

Using the complete contamination dataset of Lake Erie from 1971, this study aims to identify the minimum sampling density at which statistically valid and spatially accurate predictions can be generated. The sampling pattern and density of the 2014 Lake Erie dataset will be scrutinized to provide insight into whether it is adequate for performing statistical and spatial analyses. This experiment intends to identify the threshold below which the sampling density of toxic contaminants in Lake Erie will produce very generalized prediction surfaces from which policy and environmental decisions might not be justified. In doing so, contemporary sediment surveys could be better developed alongside historic survey designs to accommodate evolving research objectives over time, while maintaining the spatial reliability of sampling densities required to produce statistically valid and meaningful interpolated maps.

MATERIALS AND METHODS

Sediment sampling surveys of Lake Erie were conducted by Environment Canada in 1971, and the Great Lakes Surface Water Surveillance Program in 2014. Contemporary sampling locations were randomly selected, mostly within deep offshore regions of Lake Erie. These were identified as having the greatest threats for sediment contamination. When sampling locations were randomly selected close to historic sample sites, they were relocated to that location to allow for temporal comparisons between sites. Descriptive statistics for Hg from the 1971 and 2014 surveys are presented in Table 1.

Sediment samples were collected as mini-box cores; for a complete outline of this sampling procedure, see Painter *et al.* (2001) and Marvin *et al.* (2002). Using this method, the top 3 cm of sediment was collected, stored in prewashed jars and frozen for transportation to the lab for analysis (Painter *et al.* 2001). The top 3 cm of the lake bed is considered the active layer, where the greatest amount of chemical and biological activity occurs (Dalia *et al.* 2014). As the upper limits of the lake bottom, it is also a 'substance exchange' zone between the lake sediment and the lake water (Dalia *et al.* 2014).

Sediment samples from 1971 measured Hg contamination between 0.008 μg/g and 7.488 μg/g. The maximum contamination recorded is an outlier of the dataset but was not removed since it would defy the principles of randomized sampling. Extreme sediment contamination values also identify highly polluted areas and should not be discarded or ignored when creating estimated contamination surfaces (Lui *et al.* 2006; Li & Heap 2008; Wu *et al.* 2011).

Ordinary kriging is the most commonly used technique among kriging interpolators, namely, for its valuable output of prediction errors and surfaces (Ouyang *et al.* 2003; Forsythe & Marvin 2009). This method was used to estimate the amount of Hg sediment contamination at unsampled locations throughout Lake Erie. The 1971 dataset was best modelled using a Gaussian distribution with a major range of 100,000 m, a minor range of 50,000 m and an anisotropic direction of 90°; the data were assumed to be anisotropic in

Table 1 | Descriptive statistics of Hg sediment contamination samples from Lake Erie

Year	Sample count	Minimum (μg/g)	Maximum (μg/g)	Average (μg/g)	Standard deviation
1971	263	0.008	7.488	0.610	0.706
2014	34	0.004	0.777	0.159	0.164

nature due to the influence of lake currents and circulation patterns on contamination distribution (Forsythe & Marvin 2005). These parameters were determined through experimentation using various models and evaluating the cross-validation statistics against the conventions for statistically significant distributions. Geographic weight was applied to a maximum of five and minimum of one neighbours. The same parameters were used on all down-sampled subsets to ensure consistency of the modelling technique.

Cross-validation statistics assess the accuracy, or amount of error, between an estimated value and the actual value at a sampled location (Ouyang *et al.* 2003; Li & Heap 2008). The resulting estimation errors indicate the strength of the model and semivariogram to predict all unsampled locations (McBratney & Webster 1986; Forsythe & Marvin 2005; Forsythe *et al.* 2010). The fit of a semivariogram model is accepted as accurate and unbiased when the mean prediction error (MPE) is close to 0, the average standard error (ASE) is as small as possible (<20), the standardized root-mean-squared prediction error (SRMSPE) is close to 1, and there is minimal difference between the root mean square prediction error (RMSPE) and ASE (Johnston *et al.* 2003; Simpson & Wu 2014; Forsythe *et al.* 2016a, 2016b, 2016c). When the MPE is close to 0, there is a marginal average difference between measured and predicted values at the same location; a MPE greater than zero indicates underestimation of the predicted values, whereas prediction errors less than zero show that values were estimated to be higher than they actually are (Osburn 2000; Forsythe & Marvin 2005). Additionally, a SRMSPE greater than or less than 1 represents under- and overestimation of variability, respectively (Johnston *et al.* 2003; Forsythe *et al.* 2016b, 2016c).

Down-sampled survey density

The 1971 dataset is the most extensive survey of Lake Erie available for analysis, and for that reason acts as the ideal sampling density in this research (hereinafter referred to as the 'complete dataset'). Since the 2014 dataset only includes 34 locations (13% of the 1971 survey), experimental subsets were down-sampled at 10% intervals to include at least one set of data points with a smaller sample size than the 2014 dataset. Down-sampled subsets (Figure 1) were selected by a stratified random sampling method using SPSS (IBM Corporation 2013).

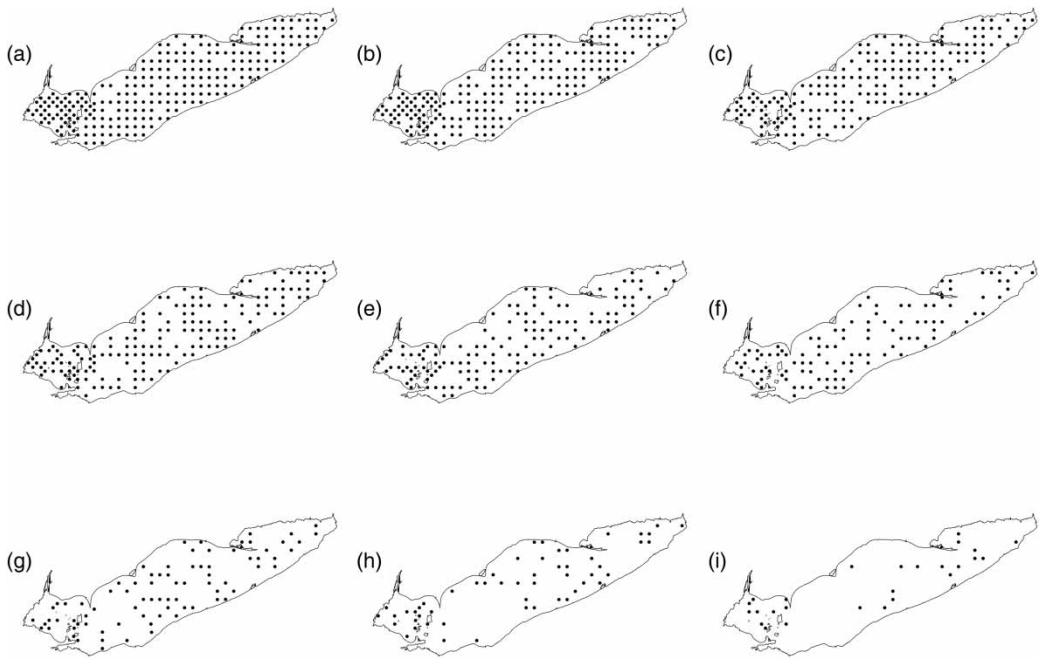

Figure 1 | Point density maps of (a) 90%, (b) 80%, (c) 70%, (d) 60%, (e) 50%, (f) 40%, (g) 30%, (h) 20% and (i) 10% randomly down-sampled subsets from the complete 1971 dataset.

Random selection of sampling locations was intentional to replicate (as closely as possible) the 2014 survey design. Descriptive and spatial statistics of the complete 1971 dataset and nine down-sampled subsets are presented in Table 2. The smallest down-sampled subset at 10% of the complete dataset includes 26 points, eight locations less than the 2014 dataset and is used as the lower sampling limit in this research.

Cross-validation statistics were analysed for statistical validity of the kriging model and parameters used for each subset (Table 3). Predicted sampling values from each subset were compared against predicted sample values of the complete dataset using a linear regression analysis (Table 3). This process was performed in McKenney *et al.* (1998), where slopes and correlation coefficients of 1 identify identical predictions between two datasets. Slopes greater than 1 identify underestimation of the subset prediction to the complete dataset, and slopes less than 1 show that the down-sampled subsets overpredict values in comparison to the complete dataset (McKenney *et al.* 1998).

Density change detection

Contamination maps utilizing the threshold effect level (TEL) and probable effect level (PEL) were created from each down-sampled subset. The TEL refers to the concentration below which adverse biological effects are expected to occur rarely, while the PEL defines the level above which adverse effects are expected to occur frequently (CCME 1999). The TEL and PEL for Hg are 0.17 μg/g and

Table 2 | Descriptive statistics of Hg sediment contamination from the complete 1971 and down-sampled subsets

Subset (%)	No. of samples	Min. (μg/g)	Max. (μg/g)	Range (μg/g)	Average (μg/g)	Standard deviation	Variance
100	263	0.008	7.488	7.48	0.610	0.706	0.498
90	237	0.008	7.488	7.48	0.620	0.722	0.521
80	210	0.010	2.929	2.919	0.594	0.579	0.335
70	184	0.008	2.929	2.921	0.559	0.558	0.311
60	158	0.008	7.488	7.48	0.616	0.775	0.601
50	132	0.008	2.929	2.921	0.632	0.564	0.318
40	105	0.008	7.488	7.48	0.714	0.879	0.773
30	79	0.008	2.929	2.921	0.552	0.563	0.317
20	53	0.008	1.881	1.873	0.520	0.476	0.227
10	26	0.056	2.817	2.761	0.873	0.779	0.607

Table 3 | Kriging cross-validation statistics of the complete 1971 dataset and all down-sampled subsets

% Data	MPE	RMSPE	SRMSPE	ASE	RMSPE-ASE	Adjusted r^2	Slope
100	−0.001	0.580	1.216	0.479	0.101	n/a	n/a
90	0.001	0.611	1.255	0.489	0.122	0.996	0.988
80	0.001	0.384	0.909	0.423	−0.039	0.924	0.971
70	0.002	0.354	1.019	0.350	0.004	0.926	0.943
60	0.003	0.703	1.188	0.596	0.106	0.950	1.049
50	0.002	0.448	0.998	0.449	−0.001	0.873	1.046
40	0.000	0.907	3.138	0.292	0.616	0.726	0.602
30	−0.006	0.391	0.963	0.403	−0.011	0.799	0.866
20	0.005	0.367	1.058	0.364	0.002	0.847	1.202
10	−0.007	0.339	0.794	0.482	−0.105	0.908	0.754

0.486 μg/g, respectively. The maps were reclassified to represent the spatial extent of Hg values in three categories: <TEL, ≥TEL to <PEL and ≥PEL (Table 4).

Each map was transformed into a raster dataset (500 m cell size, 0.05% of the 10 km grid) to perform calculations between the complete and subset datasets. Contamination categories of the complete dataset were renamed as 1 (<TEL), 2 (≥TEL to <PEL) and 3 (≥PEL). Subset categories were renamed as 100 (<TEL), 200 (≥TEL to <PEL) and 300 (≥PEL). The Raster Calculator tool in ArcGIS (Esri 2017) was used to add the reclassified contamination categories of the complete dataset to each subset individually. This resulted in the creation of new numeric categories which identify the degree and location of change between the complete dataset and each down-sampled subset. Change categories 101, 202 and 303 represent no change in contamination classifications between the complete (original 1971) and subset datasets. Change categories 102, 103 and 203 identify classification discrepancies where subset datasets predicted lower Hg contamination values than was predicted by the complete dataset. In contrast, change categories 201, 301 and 302 identify classification discrepancies where subset datasets predicted higher Hg contamination values than was predicted by the complete dataset.

RESULTS AND DISCUSSION

All subsets had a MPE close to 0, SRMSPE close to 1, and low RMSPE and ASE. MPE values close to 0 showed a strong agreement between actual and predicted Hg contamination values, meaning the kriging model made reasonable estimations at unsampled locations. The highest SRMSPE values were present in subsets which randomly included the 7.488 μg/g outlier (100%, 90%, 60% and 40%). With SRMSPE greater than 1, down-sampled subsets at 90%, 60% and 40% of the complete dataset had underestimated the variability of predictions made by the kriging model; not uncommon considering that kriging commonly underestimates predictions of high values (Webster & Oliver 2007).

Estimated values from each subset were individually plotted against the estimated values from the complete dataset to further assess the correlation between prediction densities (slope and r^2 values presented in Table 3). Subset predictions were moderately (0.726) to strongly (0.996) correlated with predictions made using the complete dataset. At 60%, 50% and 20% of the complete dataset, slopes >1 identified underestimation of the subset's prediction to estimates made from the complete data. Despite the differences in data density, locations, and pattern, predicted values of each down-sampled subset were statistically valid according to cross-validation statistics and correlation coefficients.

Predicted Hg sediment contamination using the complete 1971 dataset is presented in Figure 2. Isolines demarcate the spatial boundaries of threshold effect levels (0.17 μg/g) and probable effect levels (0.486 μg/g) of Hg contamination (CCME 1999). Most sediment contamination <TEL (2,679.72 km^2, or 10.39% of the Lake Erie analysis area) is located in the central basin, along the Ontario shoreline extending out into the centre of the lake. Sediment contamination ≥TEL and <PEL is located throughout the central and eastern basins representing 9,366.5 km^2, or 36.3% of the analysis area. Contamination ≥PEL represents the greatest spatial extent of sediment contamination (12,941.95 km^2, or 50.16% of the Lake Erie analysis area).

Subsets at 90%, 80% and 70% of the complete dataset produced very similar Hg contamination patterns throughout Lake Erie (Figure 3(a)–3(c)); exact spatial extents for all contamination categories are presented in Table 5. Contamination <TEL remains located along the Ontario shoreline near the eastern extent of the central basin. Sediment contamination ≥TEL to <PEL is consistently found across the central basin into the eastern basin, although the extent of this category grows towards the Niagara region at 80% and 70% subsets. The spatial extent of sediment contamination ≥PEL has consistent coverage in the western basin, as well as near the Ohio and New York shorelines. Overall, subset densities of 90%, 80% and 70%

Table 4 | Hg contamination intervals, and respective thresholds, used for the density change detection analysis

Contamination intervals <TEL (μg/g)	Contamination intervals ≥TEL to <PEL (μg/g)	Contamination intervals ≥PEL (μg/g)
0.000–<0.057	0.170–<0.275	0.486–<0.591
0.057–<0.113	0.275–<0.381	0.591–<0.697
0.113–<0.170	0.381–<0.486	≥0.697

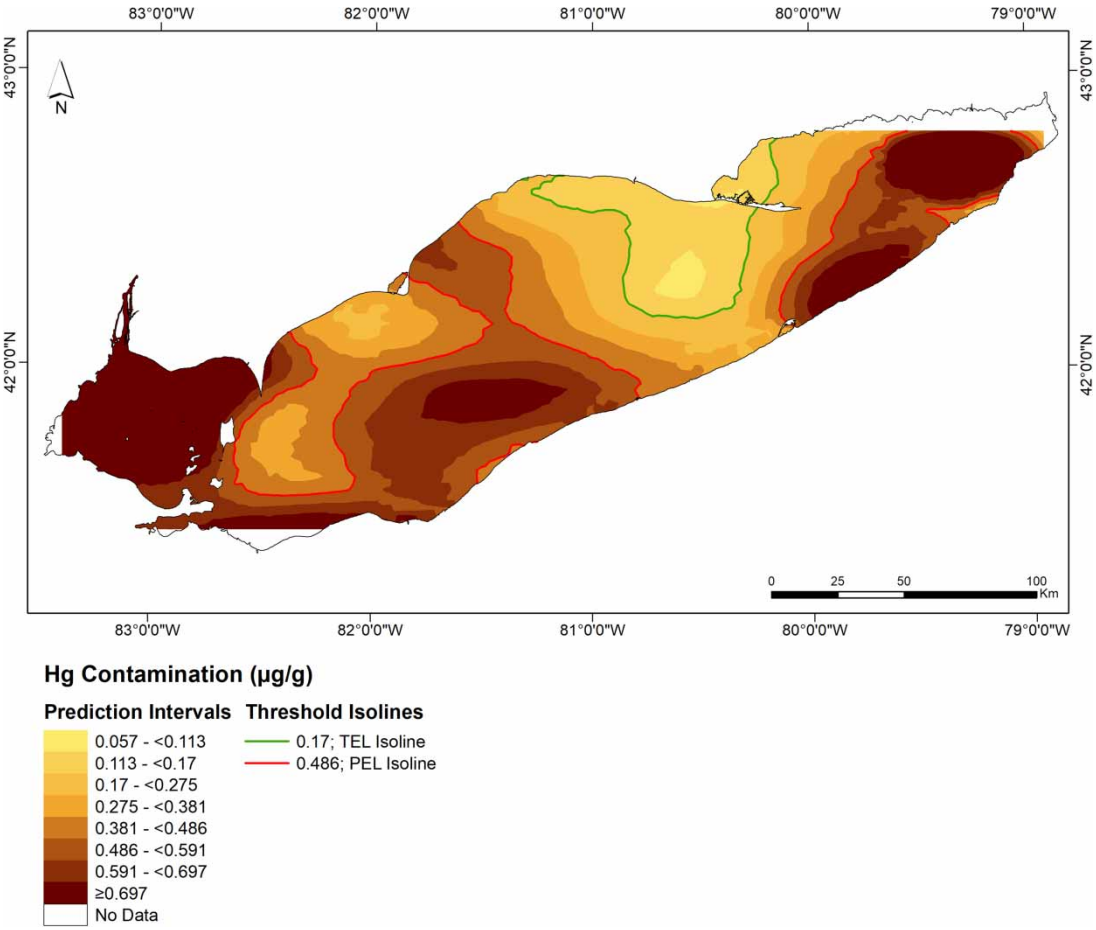

Figure 2 | Kriging predictions of Hg sediment contamination in Lake Erie based on the complete 1971 dataset.

of the complete dataset represent highly comparable patterns of Hg contamination throughout Lake Erie.

Predictions made using 60% and 40% of the complete data density also produced contamination patterns similar to those predicted using the complete dataset as seen in Figure 4(d) and 4(f) and Table 5. It is important to note, however, the unusually high SRMSPE of 3.138 produced using the 40% subset that highlights an underestimation of the predicted dataset. Isolines predicted using the 50% subset (Figure 4(e)) are especially abnormal for contamination <TEL, as well as areas ≥TEL to <PEL. These patterns are likely the result of a small sampling density which randomly omitted the 7.488 µg/g outlier. Overall, the range of Hg contamination expands to include minimum contamination values of <0.057 µg/g at these densities. The spatial extent of sediment contamination <TEL closely resembles the same extent of <TEL contamination as predicted by the complete dataset, with the exception of the 50% subset. Conversely, contamination predictions throughout the western basin remain consistent with predictions made with subsets at greater sampling densities.

The overall spatial extent of prediction maps made from 30%, 20% and 10% subsets are noticeably smaller (Figure 5(g)–5(i) and Table 5). At 30% of the complete data density, the area of contamination ≥PEL is noticeably different from the complete dataset. Contamination ≥TEL to <PEL represents most of the eastern basin and encroaches on the expected ≥PEL monopoly in the western basin. At 20% of the complete PEL isolines demarcate increasingly more generalized contamination patterns in comparison to those created using the complete dataset. Due to the small sampling density in the 20% and 10% subsets, sediment

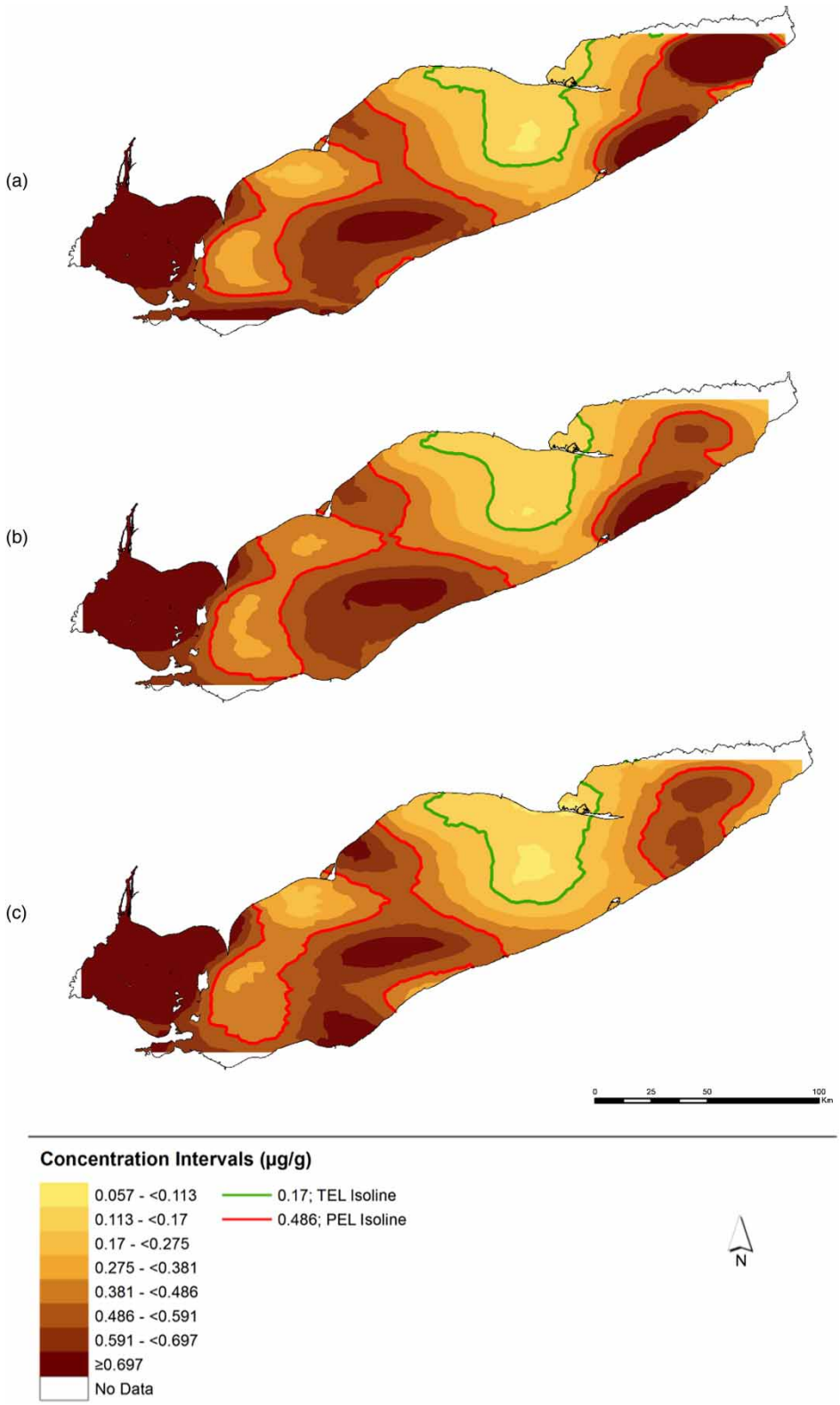

Figure 3 | Kriging predictions of Hg sediment contamination in Lake Erie based on (a) 90%, (b) 80% and (c) 70% subsets of the complete 1971 dataset.

Table 5 | Spatial extent (km^2) of predicted Hg contamination by subset

Subset	<TEL	≥TEL to <PEL	≥PEL	Total analysis area	No data
90%	2,497.62	9,489.27	13,001.47	24,988.36	811.63
80%	2,423.21	10,285.14	12,186.13	24,894.48	905.52
70%	2,881.99	10,295.58	11,810.79	24,988.36	811.63
60%	2,137.80	10,983.36	11,772.25	24,893.41	906.59
50%	760.12	10,883.36	13,250.99	24,894.47	905.52
40%	3,292.16	8,833.13	12,727.37	24,852.66	947.34
30%	2,413.04	11,426.09	9,447.82	23,286.95	2,513.06
20%	3,255.45	12,071.81	8,815.36	24,142.62	1,657.38
10%	n/a	8,841.77	11,340.78	20,182.55	5,617.45

samples were not located near the Ohio shoreline. This leaves the Sandusky and Cleveland areas of Lake Erie void of sediment contamination predictions. The 10% subset least resembles the contamination patterns created from kriging the complete dataset.

Categorical change detection analysis

Kriging of Hg contamination was best predicted using the 90%, 80% and 70% subsets. Less than 8% of pixels represented over- or underestimation of contamination thresholds (Table 6). Change detection using the 90% subset reported no change in 97% of the pixels. Over- (1%) and underestimations (2%) were typically found at the boundary of two contamination thresholds; this pattern was common at the boundaries between <TEL, and ≥TEL to <PEL using 80% and 70% subsets as well. Predictions made using the 80% and 70% subsets reported no change in 90% and 88% of pixels, respectively. Underestimations of ≥TEL to <PEL, and ≥PEL were common along the PEL isoline in the eastern basin, and nearest the Ohio shore. The locations and extent of over- and underestimations made by the 90%, 80% and 70% subsets are likely representative of the transition zones between contamination thresholds and the size of the raster cell used to detect categorical change.

Greater discrepancies in predictions were made using the 60%, 50% and 40% subsets than those at higher sampling densities; however, over- and underpredictions remained largely along the boundaries between contamination thresholds. Using the 60% subset, 86% of pixels reported seeing no change from predictions made using the complete dataset. Over- and underestimation of predictions were recorded in 8% and 6% of pixels, respectively, most of which densely surrounded the contamination threshold boundaries. Predictions made using the 50% and 40% subsets predicted 75% and 79% agreement between predictions made using the complete dataset, respectively. The 50% subset overpredicted (≥TEL to <PEL) nearly the entire region identified as <TEL by the complete dataset. Using this subset, over- and underestimations are no longer confined to contamination threshold boundaries as seen previously at higher sampling densities. Severe underestimation occurs using the 40% subset where a small section (147.5 km^2) of Hg contamination is predicted to be <TEL when previously identified as ≥PEL. An even smaller region (2.5 km^2) of Lake Erie was classified to represent severe overestimation where the 40% subset predicted contamination ≥PEL when previously reported as <TEL in predictions made by the complete dataset.

Interpolation of the 30%, 20% and 10% subsets resulted in the greatest instances of over- and underestimation of contamination versus the complete dataset. Predictions made using the 30% subset predicted 78% agreement between predictions made using the complete dataset. Agreements between datasets decreased to 75% using the 20% subset, and 51% using the 10% subset. Severe over- (269.75 km^2) and underestimations (81 km^2) were recorded in predictions made using the 10% subset. Over- and underestimations continued to be estimated below the TEL isoline

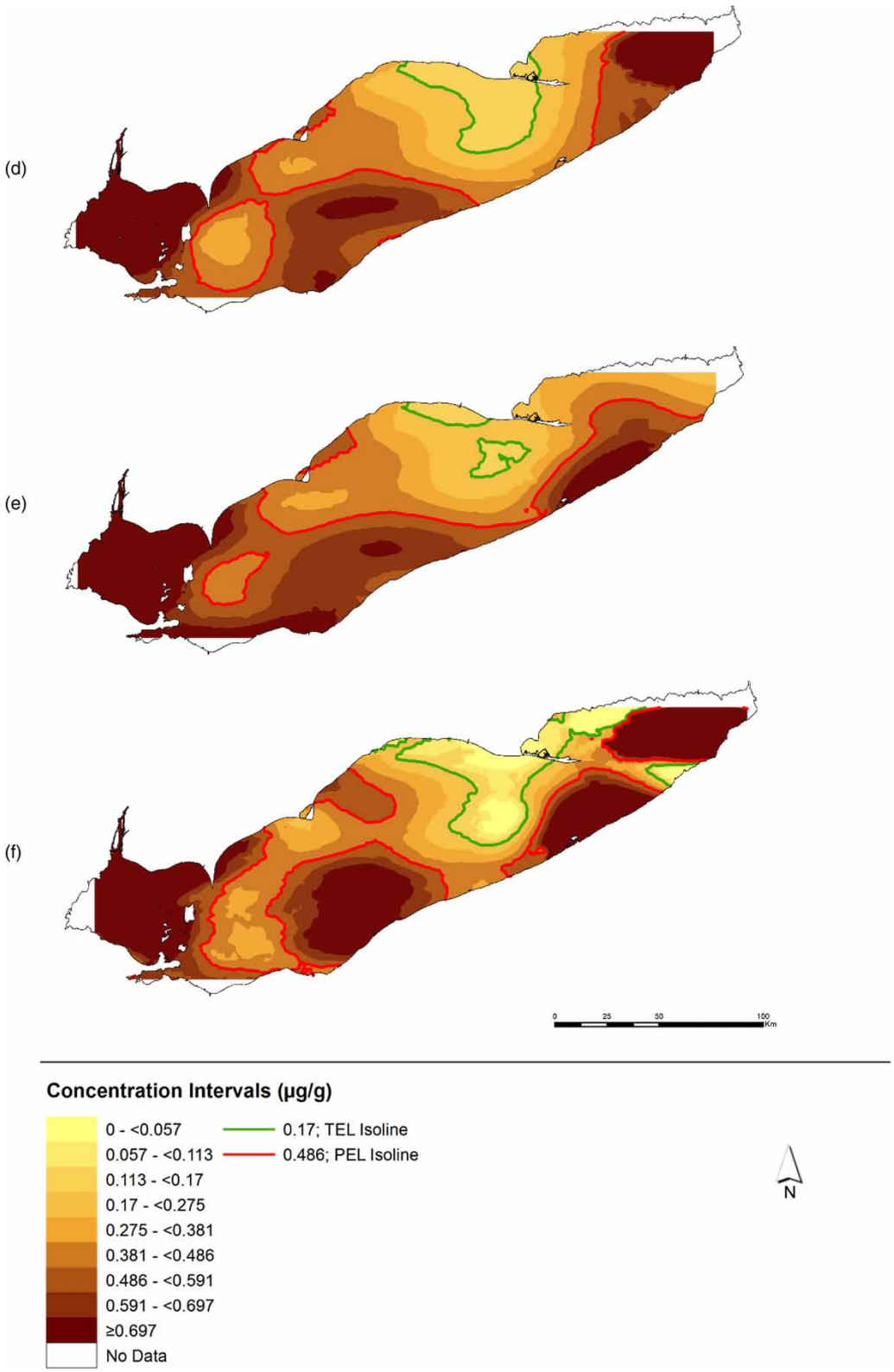

Figure 4 | Kriging predictions of Hg sediment contamination in Lake Erie based on (d) 60%, (e) 50% and (f) 40% subsets of the complete 1971 dataset.

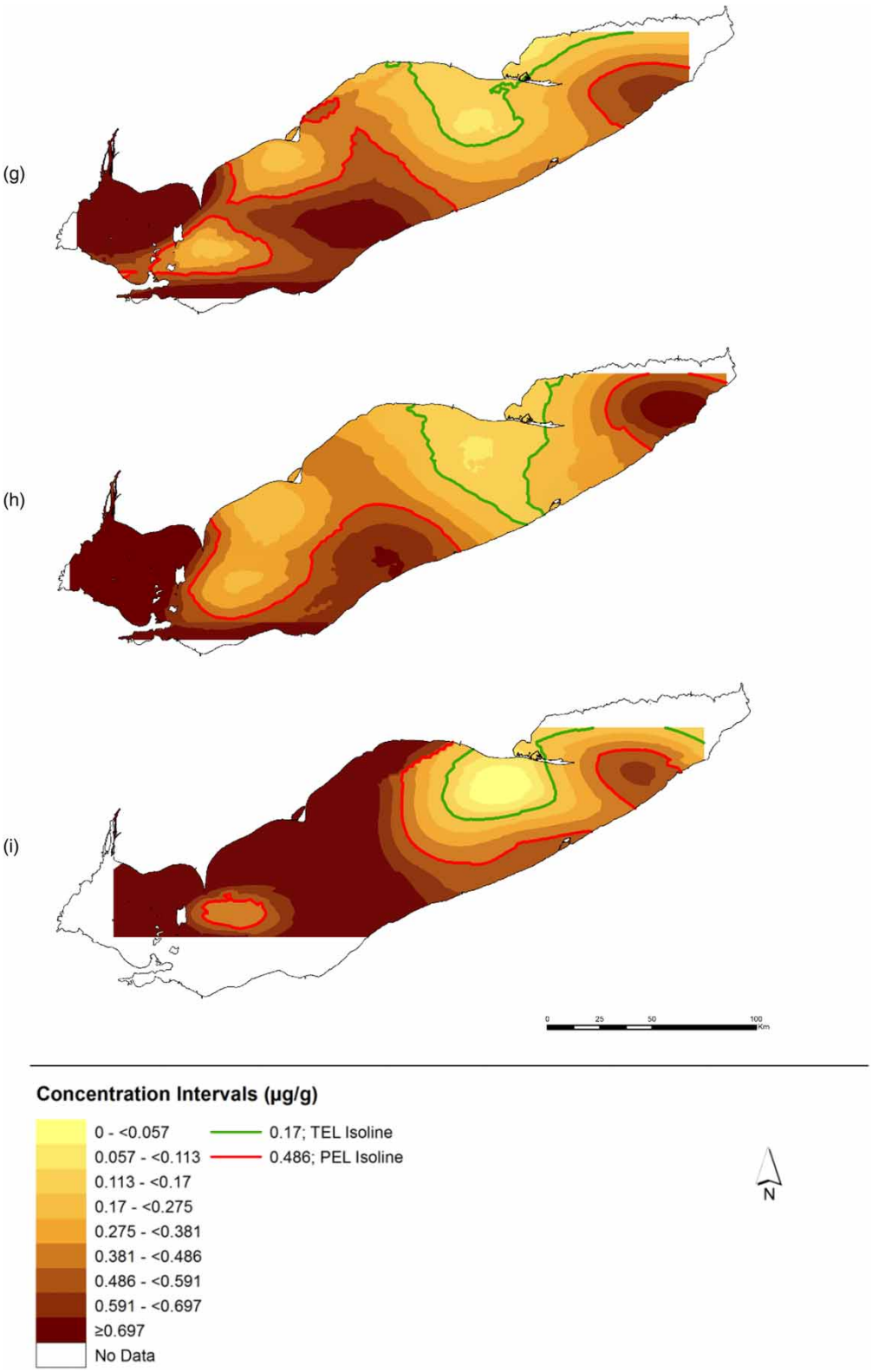

Figure 5 | Kriging predictions of Hg sediment contamination in Lake Erie based on (g) 30%, (h) 20% and (i) 10% subsets of the complete 1971 dataset.

Table 6 | Percent change in pixels classified to represent no change, over- and underestimations of Hg sediment contamination from kriging predictions based on 90%, 80%, 70%, 60%, 50%, 40%, 30%, 20% and 10% subsets of the complete 1971 dataset

% Subset	% No change	% Overestimation	% Underestimation	% No data
90	97	1	2	0
80	90	6	4	0
70	88	8	3	1
60	86	8	6	0
50	75	8	17	0
40	79	11	9	1
30	78	13	7	2
20	75	18	3	4
10	51	7	22	20

in the central basin, below the PEL isoline in the central basin above the PEL isoline in the eastern basin.

Smaller sampling densities and their spatial patterns can influence the coverage of predictions made throughout the lake. This change detection analysis also accounted for the proportion of pixels representing 'no data' as predicted surfaces made from subsets became increasingly smaller than that made from the complete dataset. The spatial extent of sampling points from the 90%, 80%, 70%, 60%, 50% and 40% subsets recorded 'no data' in ≤1% of pixels. Prediction surfaces made using the 30%, 20% and 10% subsets recorded 'no data' in 2%, 4% and 20% of the study area, respectively.

CONCLUSION

Prediction errors measure the legitimacy of interpolated surfaces. Although each down-sampled subset was deemed statistically valid by cross-validation statistics, spatial analysis of each predicted surface identified significant deviations from contamination patterns of the complete dataset, which is considered to be the most accurate representation of Hg contamination in Lake Erie. In some cases, statistical validity may not translate into spatially reliable contamination maps. For example, while the cross-validation statistics of the 2014 Hg dataset are satisfactory (MPE of −0.009, RMSPE of 0.166, SRMSPE of 1.446 and ASE of 0.106), predictive contamination maps at sampling densities of 0.001 samples/km^2 may not produce reliable or meaningful representations of contamination patterns throughout Lake Erie (Figure 6).

Herein lies some evidence regarding the potential confidence given by cross-validation statistics. Returning to the foundation of kriging, locations closer together are expected to record similar levels of contamination than those further apart (Verly *et al.* 1983). Sampling densities and location patterns are highly influential to the underlying data structure from which kriging is performed (Li & Heap 2008). As the sampling density decreases in smaller subsets, the search neighbourhood expands much wider to find a maximum of five and minimum of one nearby sampling locations. The average distance between nearest neighbours in the 90% subset is <1 km difference from the average nearest neighbours in the complete dataset. Nearest neighbours in the 10% subset are, on average, 6 km further apart than those in the complete dataset. Sparse data density is known to produce more generalized interpolated surfaces (Li & Heap 2008), which is noticeable in kriged maps created with subsets <20% of the complete dataset. Isolines demarcating the boundaries of sediment contamination <TEL, ≥TEL to <PEL and ≥PEL become smoother, altering their shape, size and locations throughout the lake. Visualizations of Hg contamination in Lake Erie convey very different stories at subsets <20% of the complete dataset. From a decision-making perspective, the representations of sediment contamination, especially from the 10% subset could lead to them being of limited usefulness regarding the state of contamination in Lake Erie.

Exploratory data analysis revealed an outlier in the 1971 Hg contamination dataset (7.488 μg/g) which was not removed from the dataset when randomly down-sampling subsets were created. This is simply due to that fact that extreme outliers are particularly important pieces of information in contamination datasets (Li & Heap 2008). The 7.488 μg/g outlier, located in the north-east corner of the eastern basin, was randomly selected to remain in the 90%, 60% and 40% subsets. The relatively high sample density of the 90% subset did not allow the outlier to drastically influence the interpolation method nor the resultant contamination map. Cross-validation statistics for the 60% and 40% subset indicate slight underestimation of

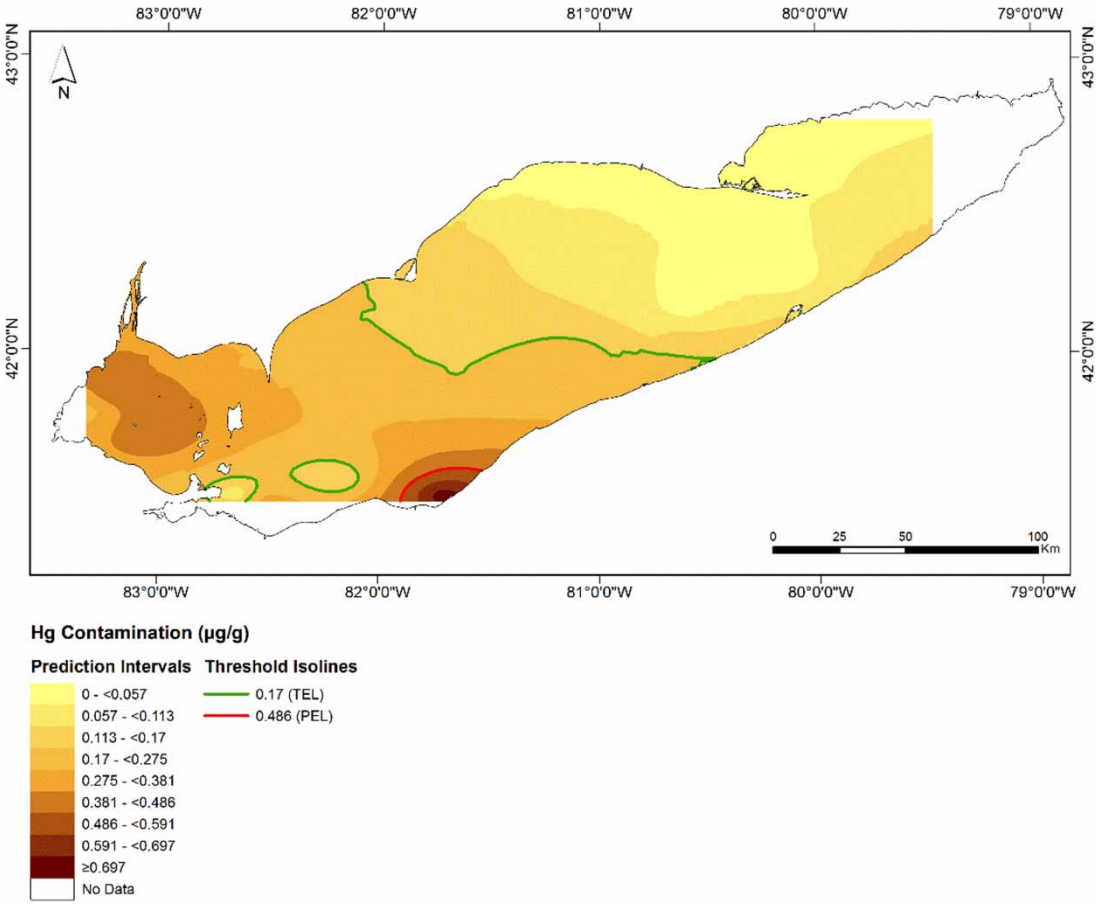

Figure 6 | Kriging predictions of Hg sediment contamination in Lake Erie based on the 2014 dataset.

variability in their respective datasets (Johnston *et al.* 2003; Forsythe *et al.* 2016a) relative to the rest of the subsets. The categorical change detection analysis identified both severe underestimation (103; subset (<TEL) underestimated complete dataset (≥PEL)) and severe overestimation (301; subset (≥PEL) overestimated complete dataset (<TEL)) of predicted contamination using the 40% subset in the eastern basin. The outlier appeared to have no extraneous influence on predictions made using the 60% subset, producing similar contamination values as were made by the complete dataset. In subsets without the outlier, substantial portions of the eastern basin are underestimated (≥TEL to <PEL) in comparison to the complete dataset (≥PEL).

One goal of these experiments was to determine if the 34 samples collected in the 2014 survey are sufficient to produce statistically valid and spatially meaningful contamination maps. It is, however, not the number of samples but rather the density and spatial distribution of sampling locations which should have been questioned. The sampling density of the 2014 survey is approximately 0.001 samples/km^2, similar (albeit slightly higher) to that of the 10% subset of the 1971 dataset. Predicted contamination levels at this density likely created very generalized contamination maps in comparison to estimations made from the complete 1971 dataset.

This experiment with randomly down-sampled subsets showed how interpolated surfaces became increasingly generalized with smaller sampling densities. Increased procedural expenses have influenced survey design methodology for contamination sampling in Lake Erie. The 34-point survey from 2014 focused its resources on offshore regions of the lake in an effort to analyse non-point source pollution patterns. Small data densities created from biased data collection will likely produce exponential bias through data

interpolation. Lake-wide spatio-temporal analyses were affected by the small sampling density of the most recent survey and it perhaps does not provide accurate interpolated sediment contamination patterns for the whole of Lake Erie from which environmental and policy decisions might be justified. Moving forward, a method to improve the survey design of sediment sampling could include smaller defined study areas within Lake Erie, in which kriging could be performed at smaller sampling densities and surface areas. In addition, resampling at as many points as possible from the original 1971 survey grid may improve the reliability of contamination maps over time while applying greater emphasis on non-point source pollution analyses.

REFERENCES

Burrough, P. A. & McDonnell, R. A. 1998 *Principles of Geographical Information Systems*. Oxford University Press, Oxford, p. 333.

Canadian Council of Ministers of the Environment (CCME) 1999 *Canadian Environmental Quality Guidelines*. Canadian Council of Ministers of the Environment, Winnipeg, MB.

Chang, Y. H., Scrimshaw, M. D., Emmerson, R. H. C. & Lester, J. N. 1998 Geostatistical analysis of sampling uncertainty at the Tollesbury Managed Retreat Site in Blackwater Estuary, Essex, UK: kriging and cokriging approach to minimise sampling density. *Science of the Total Environment* **221** (1), 43–57.

Chen, G., Zhao, K., McDermid, G. J. & Hay, G. J. 2012 The influence of sampling density on geographically weighted regression: a case study using forest canopy height and optical data. *International Journal of Remote Sensing* **33** (9), 2909–2924.

Clark, I. 1982 *Practical Geostatistics*. Applied Science Publishers Ltd, Barking, p. 129.

Dalia, M. S., Salem, A., Khaled, A., El Nemr, A. & El-Sikaily, A. 2014 Comprehensive risk assessment of heavy metals in surface sediments along the Egyptian Red Sea coast. *The Egyptian Journal of Aquatic Research* **40** (4), 349–362.

De Solla, S. R., Struger, J. & McDaniel, T. V. 2012 Detection limits can influence the interpretation of pesticide monitoring data in Canadian surface waters. *Chemosphere* **86** (6), 565–571.

Dille, J. A., Milner, M., Groeteke, J. J., Mortensen, D. A. & Williams II, M. M. 2002 How good is your weed map? A comparison of spatial interpolators. *Weed Science* **51** (1), 44–55.

Esri 2017 *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands, CA.

Evers, D. C., Wiener, J. G., Driscoll, C. T., Gay, D. A., Basu, N., Monson, B. A., Lambert, K. F., Morrison, H. A., Morgan, J. T., Williams, K. A. & Soehl, A. G. 2011 *Great Lakes Mercury Connections: The Extent and Effects of Mercury Pollution in the Great Lakes Region*. Report BRI 2011–18. Biodiversity Research Institute, Gorham, ME, p. 44.

Forsythe, K. W. & Marvin, C. H. 2005 Analyzing the spatial distribution of sediment contamination in the Lower Great Lakes. *Water Quality Research Journal of Canada* **40** (4), 389–401.

Forsythe, K. W. & Marvin, C. H. 2009 Assessing historical versus contemporary mercury and lead contamination in Lake Huron sediments. *Aquatic Ecosystem Health and Management* **12** (1), 101–109.

Forsythe, K. W., Paudel, K. & Marvin, C. H. 2010 Geospatial analysis of zinc contamination in Lake Ontario sediments. *Journal of Environmental Informatics* **16** (1), 1–10.

Forsythe, K. W., Marvin, C. H., Valancius, C. J., Watt, J. P., Swales, S. J., Aversa, J. M. & Jakubek, D. J. 2016a Using geovisualization to assess lead sediment contamination in Lake St. *Clair. The Canadian Geographer* **60** (1), 149–158.

Forsythe, K. W., Marvin, C. H., Valancius, C. J., Watt, J. P., Aversa, J. M., Swales, S. J., Jakubek, D. J. & Shaker, R. R. 2016b Geovisualization of mercury contamination in Lake St. Clair sediments. *Journal of Marine Science and Engineering* **4** (19), 1–10.

Forsythe, K. W., Marvin, C. H., Mitchell, D. E., Aversa, J. M., Swales, S. J., Burniston, D. A., Watt, J. P., Jakubek, D. J., McHenry, M. H. & Shaker, R. R. 2016c Utilization of bathymetry data to examine lead sediment contamination distributions in Lake Ontario. *AIMS Environmental Science* **3** (3), 347–361.

Gawedzki, A. & Forsythe, K. W. 2012 Assessing anthracene and arsenic contamination within Buffalo River sediments. *International Journal of Ecology* **2012**, 1–9.

Goovaerts, P. 1999 Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* **89** (1–2), 1–45.

Gotway, C. A., Ferguson, R. B., Hergert, G. W. & Peterson, T. A. 1996 Comparison of kriging and inverse-distance method for mapping soil parameters. *Soil Science Society of America Journal* **60** (4), 1237–1247.

Great Lakes Information Network (GLIN) 2017 *About the Lakes: Lake Erie*. www.glc.org/lakes/lake-erie (accessed 14 March 2018).

Hessl, A., Miller, J., Kernan, J., Keenum, D. & McKenzie, D. 2007 Mapping paleo-fire boundaries from binary point data: comparing interpolation methods. *The Professional Geographer* **59** (1), 87–104.

IBM Corporation 2013 *IBM SPSS Statistics for Windows, Version 22.0*. IBM Corporation, Armonk, NY.

International Joint Commission (IJC) 1972 *Great Lakes Water Quality Agreement with Annexes and Texts and Terms of References, Between the United States of America and Canada*. www.ijc.org/files/publications/C23.pdf (accessed 14 March 2018).

International Joint Commission (IJC) 2002 *Eleventh Biennial Report on Great Lakes Water Quality: The Challenge to Restore and Protect the Largest Body of Fresh Water in the World*. www.ijc.org/files/publications/C89.pdf (accessed 14 March 2018).

Johnston, K., Ver Hoef, J. M., Krivoruchko, K. & Lucas, N. 2003 *Using ArcGIS Geostatistical Analyst*. Esri, Redlands, USA. www.dusk2.geo.orst.edu/gis/geostat_analyst.pdf (accessed 14 March 2018).

Khosh Eghbal, M. Z. 2014 Spatial distribution of sediment pollution in the Khajeh Kory River using kriging and GIS. *Earth Sciences Research Journal* **18** (2), 173–179.

Largueche, F.-Z. B. 2006 Estimating soil contamination with kriging interpolation method. *American Journal of Applied Sciences* **3** (6), 1894–1898.

Lark, R. M. & Ferguson, R. B. 2004 Mapping risk of soil nutrient deficiency or excess by disjunctive indicator kriging. *Geoderma* **118** (1–2), 39–53.

Li, J. & Heap, A. D. 2008 *A Review of Spatial Interpolation Methods for Environmental Scientists*. Geoscience Australia, Canberra, Australia, p. 137.

Lui, X., Wu, J. & Xu, J. 2006 Characterizing the risk assessment of heavy metals and sampling uncertainty analysis in paddy field by geostatistics and GIS. *Environmental Pollution* **141** (2), 257–264.

Markus, J. & McBratney, A. B. 2001 A review of the contamination of soil with lead. 2. Spatial distribution and risk assessment of soil lead. *Environment International* **27** (5), 399–411.

Marvin, C. H., Charlton, M. N., Reiner, E. J., Kolic, T., MacPherson, K., Stern, G. A., Braekevelt, E., Estenik, J. F., Thiessen, L. & Painter, S. 2002 Surficial sediment contamination in Lake Erie and Ontario: a comparative analysis. *Journal of Great Lakes Research* **28** (3), 437–450.

Marvin, C. H., Charlton, M. N., Stern, G. A., Braekevelt, E., Reiner, E. J. & Painter, S. 2003 Spatial and temporal trends in sediment contamination in Lake Ontario. *Journal of Great Lakes Research* **29** (2), 317–331.

Marvin, C. H., Painter, S. & Rossmann, R. 2004 Spatial and temporal patterns in mercury contamination in sediments of the Laurentian Great Lakes. *Environmental Research* **95** (3), 351–362.

McBratney, A. B. & Webster, R. 1986 Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* **37** (4), 617–639.

McKenney, D. W., Rempel, R. S., Venier, L. A., Wang, Y. & Bisset, A. R. 1998 Development and application of a spatially explicit moose population model. *Canadian Journal of Zoology* **76** (10), 1922–1931.

Olea, R. A. 1984 Sampling design optimization for spatial functions. *Mathematical Geology* **16** (4), 369–392.

Osburn, W. L. 2000 *Geostatistical Analysis: Potentiometric Network for the Upper Floridian Aquifer in the St. Johns River Water Management District*. www.secure.sjrwmd.com/technicalreports/TP/SJ2002-1.pdf (accessed 14 March 2018).

Ouyang, Y., Nkedi-Kizza, P., Mansell, R. S. & Ren, J. Y. 2003 Spatial distribution of DDT in sediments from estuarine rivers of central Florida. *Journal of Environmental Quality* **32** (5), 1710–1716.

Painter, S., Marvin, C., Rosa, F., Reynoldson, T. B., Charlton, M. N., Fox, M., Lina Thiessen, P. A. & Estenik, J. F. 2001 Sediment contamination in Lake Erie: a 25-year retrospective analysis. *Journal of Great Lakes Research* **27** (4), 434–448.

Ramsey, M. H. 1997 Sampling and sample preparation. In: *Modern Analytical Geochemistry: An Introduction to Quantitative Chemical Analysis for Earth, Environmental and Material Scientists* (R. Gill, ed.). Routledge, New York, p. 342.

Rukavina, N., Dunnett, M. & Prokopec, C. 2013 *Great Lakes Sediment Database*. Environment and Climate Change Canada. Environment and Climate Change Canada, Government of Canada. Available online: www.ec.gc.ca/inre-nwri/default.asp?lang=En&n=9890771E-1 (accessed 14 March 2018).

Simpson, G. & Wu, Y. H. 2014 Accuracy and effort of interpolation and sampling: can GIS help lower field costs? *ISPRS International Journal of Geo-Information* **3** (4), 1317–1333.

Verly, G., David, M., Journel, A. G. & Marechal, A. 1983 Geostatistics for Natural Resource Characterization, Part 1. In: *Proceedings of the NATO Advanced Study Institute on Geostatistics for Natural Resource Characterization*, South Lake Tahoe, California, 6–17 September, 1983.

Villard, M.-A. & Maurer, B. A. 1996 Geostatistics as a tool for examining hypothesized declines in migratory songbirds. *Ecology* **77** (1), 59–68.

Webster, R. & Oliver, M. A. 1993 How large a sample is needed to estimate the regional variogram adequately?. *Geostatistics Troia '92* **1**, 155–166.

Webster, R. & Oliver, M. A. 2007 *Geostatistics for Environmental Scientists*, 2nd edn. John Wiley & Sons, Chichester, p. 330.

Wu, C., Wu, J., Luo, Y., Zhang, H., Teng, Y. & DeGloria, S. D. 2011 Spatial interpolation of severely skewed data with several peak values by the approach integrating kriging and triangular irregular network interpolation. *Environmental Earth Sciences* **63** (5), 1093–1103.

Wuana, R. A. & Okieimen, F. E. 2011 Heavy metals in contaminated soils: a review of sources, chemistry, risks and best available strategies for remediation. *International Scholarly Research Network: Ecology* **2011**, 1–20.

Yan, L., Zhou, S., Wu, C.-F., Li, H.-Y. & Li, F. 2007 Improved prediction and reduction of sampling density for soil salinity by different geostatistical methods. *Agricultural Sciences in China* **6** (7), 832–841.

First received 14 September 2017; accepted in revised form 12 April 2018. Available online 11 May 2018