Comparing Deterministic and Stochastic Methods in Geospatial Analysis of Groundwater Fluoride Concentration

K. Brindha [1,*] **, Majid Taie Semiromi** [2], **Lamine Boumaiza** [3] **and Subham Mukherjee** [4]

1. Hydrogeology Group, Institute of Geological Sciences, Department of Earth Sciences, Freie Universität Berlin, 12249 Berlin, Germany
2. Working Group Lowland Hydrology and Water Management, Leibniz Centre for Agricultural Landscape Research (ZALF), 15374 Müncheberg, Germany
3. Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON N2T 0A4, Canada
4. Physical Geography, Department of Earth Sciences, Institute of Geographical Sciences, Freie Universität Berlin, 12249 Berlin, Germany
* Correspondence: brindha.karthikeyan@fu-berlin.de or brindhakarthikeyan@gmail.com

Abstract: Dental and skeletal fluorosis caused by consuming high-fluoride groundwater has been reported over several decades globally. Prediction maps to estimate the fluoride contaminated area rely on interpolation methods. This study presents a comparison of the accuracy of nine spatial interpolation methods in predicting the fluoride in groundwater. Leave-one-out cross-validation (LOOCV), hold-out validation and validation with an independent dataset were used to assess the precision of the interpolation methods. This is the first study on fluoride with a large dataset (N = 13,585) applied at the regional level in India. Our findings showed that the inverse distance weighted (IDW) algorithm outperformed other methods in terms of less discrepancy between measured and predicted fluoride. IDW and local polynomial interpolation (LPI) were the only methods to predict contaminated areas (fluoride > 1.5 mg/L). However, the area estimated by the typical assessment of the percentage of unsuitable samples was much higher (6.1%) compared to that estimated by IDW (0.2%) and LPI (0.2%). LOOCV provided viable results than the other two validation methods. Interpolation methods are accompanied with uncertainty which are regulated by the sample size, sample density, sample distribution, minimum and maximum measured concentrations, smoothing and border effects. Drawing a comparison among variegated interpolation methods capturing a wide range of prediction uncertainty is suggested rather than relying on one method exclusively. The high-fluoride areas identified in this study can be used by the Government in planning remediation actions.

Keywords: fluoride; drinking water; health risk; geology; aquifer; interpolation analysis; Tamil Nadu; India

1. Introduction

Fluoride is a trace element occurring in groundwaters. Groundwaters with high levels of fluoride are commonly reported in several aquifers around the world. Many of these cases occur in developing and underdeveloped nations, where groundwater forms an important source of drinking water for the population. An estimated 200 million people worldwide are exposed to drinking water with fluoride concentrations above the World Health Organization's (WHO) guideline value (>1.5 mg/L) [1]. Exposure to high levels of fluoride is associated with dental and skeletal fluorosis, depending on the concentration and exposure period [2].

About 120 million people in India are at risk due to consumption of fluoride-contaminated groundwater [3], and an estimated 66 million people, including 6 million children, suffer from fluorosis [4]. Several studies have reported on the health risk of high-fluoride groundwater in India [5–11]. To overcome the fluoride problem, the extent of contamination should be first identified. Regular spatiotemporal monitoring and assessment of fluoride from

densely located sampling network is time-consuming, requires extensive manpower, is expensive, and is not always feasible. Therefore, mapping contaminated areas using spatial interpolation methods has come to the fore and has proved useful in geosciences [12–17].

Interpolation methods can be classified as deterministic and stochastic (geostatistical) methods. Deterministic methods are directly based on measured data, and they create surfaces based on the similarity extent or the extent of smoothing. Geostatistical methods do not directly use the data, but use the statistical properties of the measured data. These are based on a stochastic model that allows the derivation of optimal prediction at random points in the studied area [18]. Accuracy of these methods depends on the type and nature of input data, its quality (e.g., measurement errors, missing values, etc.), its distribution across the prediction area, and the boundaries. All interpolation methods have a smoothing effect that overestimates the lower values and underestimates the higher values. Thus, interpolated results are accompanied with some degree of uncertainty, thus leading to bias in pollution assessment and subsequent planning of remediation measures. Inconsistency in the results from interpolation methods and argument over the superiority of one method over another, or a universal method suitable to all studies is questionable.

A comparison of interpolation methods for predicting groundwater levels showed kriging method to perform with highest accuracy and minimum error [19]. Adhikary et al. [20] compared ordinary kriging (OK) and probability kriging methods for the estimation of heavy metal concentrations in groundwater and reported that probability kriging performs better than OK. In another water quality evaluation study, universal kriging (UK) outperformed OK and inverse distance weighting (IDW) [21]. Water quality index and leachate pollution index were better predicted by cokriging than kriging by Farzaneh et al. [22]. OK was able to identify the groundwater pollution extent in a coastal area more accurately, while IDW over-estimated the polluted areas [23]. Arsenic concentrations were estimated with least error by IDW than Gaussian kriging, spherical kriging and cokriging methods, but the results also varied when data from sub-areas were compared independently [24]. The only reported study on fluoride estimation has identified empirical Bayesian kriging (EBK) to be much superior than IDW [25].

Although thus far several studies have taken advantages of interpolation algorithms for providing a distribution map of contamination across studied areas, a comparison among the deterministic and stochastic interpolation methods has not been conducted. This holds especially true for fluoride contamination in groundwater. In the light of this comparison, the process underlying the interpolation methods and the sources of prediction uncertainty can be explored. In the present study, we compare three deterministic approaches (IDW, radial basis function (RBF), and local polynomial interpolation (LPI)) and six geostatistical methods (OK, spherical kriging, Gaussian kriging, simple kriging, UK, and EBK). Considering the importance of assessing the extent of fluoride contamination in groundwater to support the remediation of the problem, this study was carried out with the objective to assess the prediction accuracy of nine interpolation techniques. The ambiguity in the form of the extent and degree of contamination from the different methods are compared through three validation approaches. The present study is applied to Tamil Nadu State in India with a large fluoride database and is one of the known fluoride endemic regions in the world.

2. Methodology

2.1. Dataset Description

Fluoride concentration in groundwater was provided by the Public Works Department (PWD), Government of Tamil Nadu. PWD has monitoring wells that are spread throughout Tamil Nadu State, where systematic sampling and analysis are carried out twice a year; one in January representing post-monsoon and the other in July representing pre-monsoon (Figure 1). Data were available for 13,768 groundwater samples collected between 2011 and 2015. Initial data analysis included data cleaning and pre-processing such as checking the coordinates, identifying duplicates in well numbers, inaccurate recordings, number of

recordings per monitoring well, deleting inaccurate recordings, and identifying outliners. After data cleaning, 13,585 fluoride values, collected from 2735 monitoring wells, spanning a time period from 2011 to 2015 were used for further analysis. Monitoring wells were also classified based on the aquifer type information from the Central Ground Water Board (CGWB) [26] (Figure 1).
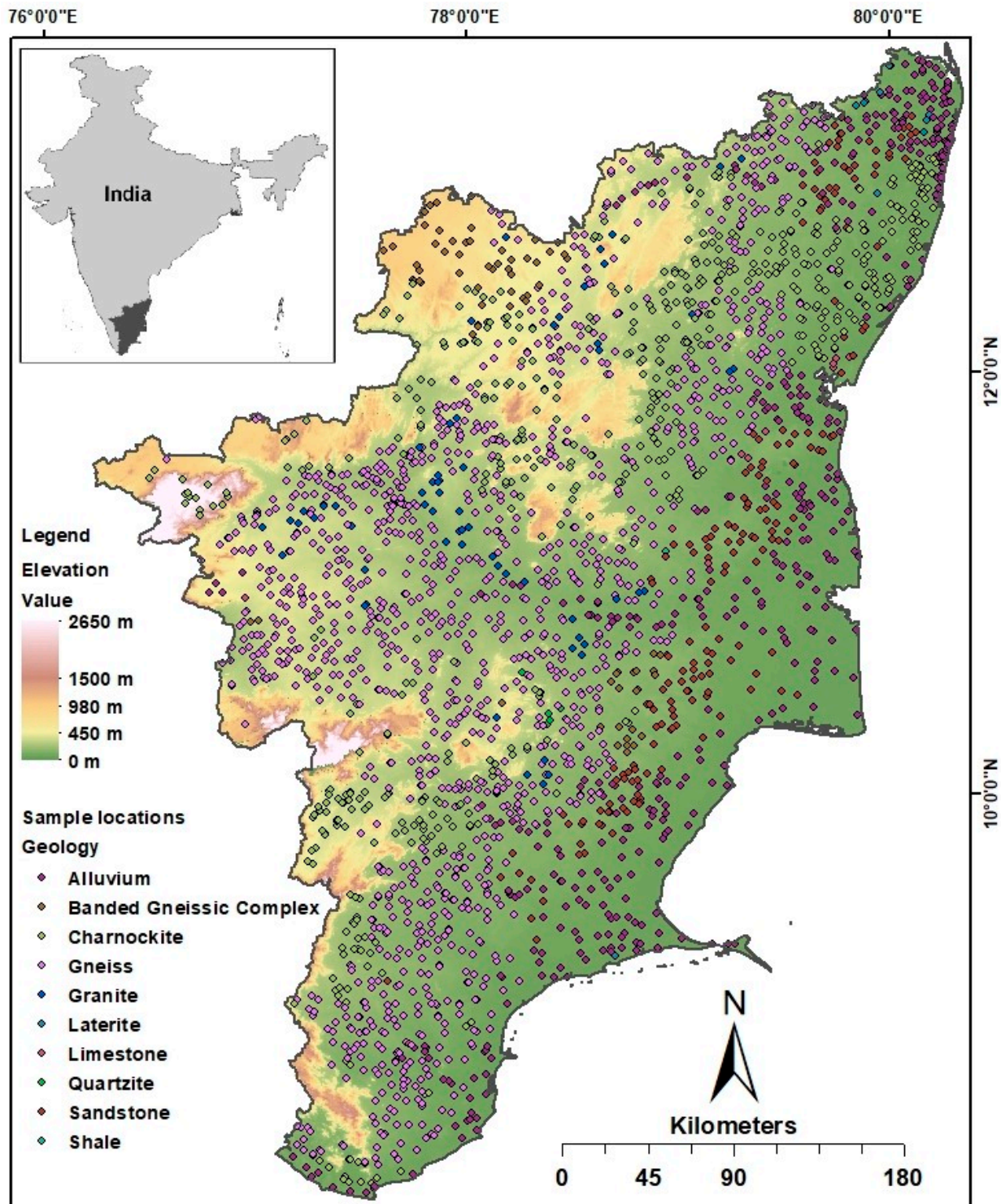

Figure 1. Study area with locations of fluoride measurements.

2.2. Interpolation Methods

Deterministic and geostatistical methods were compared in this study. These included the widely adopted interpolation methods in groundwater studies, i.e., IDW, RBF, LPI, OK, spherical kriging, Gaussian kriging, simple kriging, UK, and EBK. Brief explanations of the methods are provided below.

2.2.1. Inverse Distance Weighting

IDW predicts based on a linear combination of closely located data points. This method assigns weights to the data points such that points closer to the prediction location have higher weights and the weight decreases as the distance from the data point increases [27]. Predicted values are thus highly influenced by the assigned weights.

$$z = \frac{\sum_{i=1}^{n} w_i z_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

$$w_i = d_i^{-u} \tag{2}$$

where z is the predicted value, z_i is the known value, n is the total number of known values used in the interpolation process, w_i is the assigned weight to the known value, d_i is the distance between known and predicted values, u is the power parameter where weight decreases as distance increases from the prediction location. In this study, we used the commonly used inverse of the distance raised to the 2nd power [24].

2.2.2. Radial Basis Functions

RBFs are exact interpolation methods, i.e., the surface must pass through each measured data point. This method uses a basic equation, which is dependent on the distance between the interpolated point and the sampling points [28]. The predicted value at a given point is expressed as a sum of the following two components [29].

$$Z(x) = \sum_{i=1}^{m} a_i f_i(x) + \sum_{j=1}^{n} b_j \psi(d_j) \tag{3}$$

where $\psi(d_j)$ refers to the RBFs, d_j is the distance between the measured value and the predicted value x, $f_i(x)$ is a trend function considered as a member of a basis for the space of polynomials of degree <m, coefficients a_i and b_i are calculated by means of the resolution of the following system of n + m linear equations where n is the number of measured values used in the interpolation of the surface Z(x) [28].

$$Z(x_k) = \sum_{i=1}^{m} a_i f_i(x_k) + \sum_{j=1}^{n} b_j \psi(d_{jk}) \text{ for } k = 1, 2, \ldots, n \tag{4}$$

$$\sum_{j=1}^{n} b_j f_k(x_j) = 0 \text{ for } k = 1, 2, \ldots, m \tag{5}$$

In this study, we compare the completely regularized spline (CRS) of RBF, given by

$$\psi(d) = \ln\left(\frac{cd}{2}\right)^2 + E_1(cd)^2 + \gamma \tag{6}$$

where d is the distance between the location of measured and predicted values, c is the smoothing factor and γ is Euler's constant [30].

When compared with the other exact interpolation methods, IDW will never predict values above or below the maximum and minimum measured values, respectively. However, RBFs can predict values above the maximum and below the minimum measured values.

2.2.3. Local Polynomial Interpolation

Polynomial interpolation is classified as global and local polynomial interpolation (GPI/LPI). GPI fits a single polynomial to the entire surface, while LPI fits many polynomials using measured values from a specific neighborhood. The neighborhoods overlap and the value used for predicting each neighborhood is the value of the fitted polynomial at the center of the neighborhood. GPI uses the entire dataset to make the predictions, thus a change in one of the input values will change the entire map. LPI predicts based on small sets of measured data, i.e., predictions are made for smaller areas in a large area map. Change in any of the input values will only lead to change the results in the small area. GPI can be used to create smooth surfaces and identify long-range trends in the dataset. However, the data used in this study have short-range variations, and hence LPI, which is the appropriate method for interpolation of the available data, is included in this study.

2.2.4. Kriging Methods

Kriging assumes that the distance or direction between two measured data reflects a spatial correlation that can be used to explain the variation in the surface. The variogram is the geostatistical method for analyzing the spatial data, and forms the basis for kriging [31]. This is calculated using the formula below.

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (7)$$

where $\gamma(h)$ is the semivariogram, semivariance or variogram value at a distance interval of h, $N(h)$ is the number of sample pairs within the distance interval h, $z(x_i)$ and $z(x_i + h)$ are the sample values at two points separated by the distance interval h [31]. Kriging fits this mathematical function to a defined set of data points, or to all points within a specified radius, to predict the values for the unknown location.

OK is a weighted linear combination of the measured values [32], which is defined as:

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \quad (8)$$

where $\hat{Z}(x_0)$ is the predicted value at location x_0, $Z(x_i)$ is the measured value at the ith location, λ_i is the weight assigned to the measured value at the ith location and n is the number of measured values. The sum of weights in the above equation is equal to unity, i.e., $\sum_{i=1}^{n} \lambda_i = 1$.

The kriging option offers various functions based on the semivariogram model. The semivariogram model can be fitted to different mathematical models such as linear, exponential, circular, spherical, and Gaussian. The present study includes the most widely used spherical and Gaussian models.

Simple kriging is based on the following formula:

$$\hat{Z}(x_0) = \mu + \varepsilon(x_0) \quad (9)$$

where μ is a known constant. In OK, it is assumed that the expected value of the underlying process is the same over the entire domain studied. Hence, this constant was not necessary. In simple kriging, which is similar to OK, the mean (μ) is known in the entire domain [33], whereas in OK, μ is estimated.

UK is calculated based on the following formula:

$$\hat{Z}(x_0) = \mu(x_0) + \varepsilon(x_0) \quad (10)$$

where $\mu(x_0)$ is a deterministic function and $\varepsilon(x_0)$ is a random variation called microscale variation and mean of this is 0 [14].

Kriging methods usually require manual adjustment of parameters to arrive at accurate predictions. However, EBK automates most steps in the kriging model development. EBK automatically calculates the parameters through subsetting and simulations, and considers the errors estimated by using several semivariogram models. This is in contrast to other kriging models in which only one semivariogram from the observed data is normally calculated, which is used to predict the unknown values [34].

2.3. Validation of the Interpolation Methods

2.3.1. Leave-One-Out Cross-Validation

Leave-one-out cross-validation (LOOCV), commonly called cross-validation, is a widely used method for testing the accuracy of the interpolation. This method involves removing the fluoride data at one location and this value is predicted based on the neighboring data. This is repeated until all the data points are interpolated. Measured and predicted values of all sampling locations can be compared using this validation.

2.3.2. Hold-Out Validation

The data set is divided into two sets, namely, the training set and the test set. Normally, 80% of the data are used as the training set and 20% as test set. Eighty percent of the data spread throughout the study area were chosen at random, and predictions were made using this training set. The predicted values at the test set locations are then compared with the measured values. This is a quick and simple method for cross-validation, but may not be suitable for studies with limited sample size [30]. The accuracy of this method depends largely on the dataset that is classified as the training set and test set and the error may be significantly different based on the classified data sets. Here, a subset of 2191 wells were used as the training set and 544 wells were used as the test set.

2.3.3. Validation with an Independent Dataset

Two independent studies carried out in parts of Tamil Nadu were used to validate the predictions. The area covered in these independent studies include the Pambar and Vaniyar river basin [35,36]. Collection and analysis of groundwater samples were carried out in 78 locations. Multiple sampling campaigns were conducted between 2011 and 2013. Detailed information on the sampling and analysis are provided elsewhere [37,38]. Average fluoride concentration in these locations were compared with the predicted values by the interpolation methods based on the PWD data.

2.4. Comparison of the Interpolation Methods

The accuracies of the interpolation methods were assessed based on the mean relative error (MRE) and root mean square error (RMSE) from the measured and predicted fluoride concentrations. Equations for predicting these parameters are given below.

$$MRE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{z_o(x_i) - z_p(x_i)}{z_o x_i} \right| \pi r^2 \qquad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left| z_p(x_i) - z_o(x_i) \right|^2} \qquad (12)$$

where $z_o(x_i)$ and $z_p(x_i)$ are the observed and predicted values at location 'i', 'n' is the sample size. The smaller the MRE and RMSE values, the better the predictive power of the methods.

Coefficient of variance (CV) is calculated by,

$$CV = \left(\frac{SD}{\bar{X}} \right) \cdot 100 \qquad (13)$$

where SD is the standard deviation and $\bar{X}$ is the mean. SD is calculated using the following formula.

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}} \tag{14}$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{15}$$

where x_i is the observed value and, *n* is the number of values.

Pearson's correlation coefficient (r) is used to determine the strength of a linear relationship between two variables. This was used to measure the relationship between the measured and predicted values. The value of r = +1 indicates a perfect positive correlation, while r = −1 indicates a perfect negative correlation.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{Y})^2}} \tag{16}$$

where x_i and y_i are the measured and predicted values, and $\bar{X}$ and $\bar{Y}$ are the mean of the measured and predicted values.

3. Results

3.1. Measured Fluoride Concentration

Fluoride in groundwater during 2011–2015 in Tamil Nadu state, India ranged from 0.01 to 5 mg/L, with an average of 0.7 mg/L. Both WHO [2] and the Bureau of Indian Standards [39] have recommended 1 to 1.5 mg/L of fluoride in drinking water to avoid possible adverse dental and skeletal effects. As groundwater provides directly for domestic consumption including drinking water in the rural parts of India [40,41], we used this range for classifying the groundwater based on fluoride measurements. Of the groundwater fluoride samples (N = 13,585), 6.1% were above 1.5 mg/L, 14.7% were between 1 to 1.5 mg/L, and 79.2% were below 1 mg/L. Based on the 2735 sampling locations, the mean fluoride concentration ranges from 0.03 to 2.35 mg/L with an average of 0.7 mg/L. About 15.6% were within the Bureau of Indian Standards (BIS) drinking water specification range and, 2.8% lay above 1.5 mg/L.

3.2. Variation Based on Aquifer Type

Monitoring locations indicated that the highest fluoride concentration can be witnessed in gneissic areas followed by charnockite and alluvium aquifers. Range, mean and SD of fluoride concentrations in different aquifers are listed in Table 1. It is worth mentioning that there is a large variation in the number of measured fluoride samples which is associated with the size of each aquifer type. The area covered by each aquifer type is provide in Table S1 [26] (Supplementary material). From this table, it is evident that alluvium, gneiss and charnockite aquifers together cover up to 80% of the study area, and therefore account for 86% (N = 11,747) of the measured fluoride concentrations. The highest fluoride concentrations were present in gneissic, charnockite and sandstone aquifers. About 10%, 9% and 5% of the samples exceeded 1.5 mg/L of fluoride in granitic, gneissic and charnockite areas, respectively. Main source of fluoride in this region is ensued from weathering of fluoride rich rocks and rock–water interaction, which is supported by earlier studies [42–45].

Table 1. Distribution of fluoride in groundwater.

Aquifer Type	Number of Measured Fluoride Samples	Range (mg/L)	Mean (mg/L)	SD	Number of Samples above 1.5 mg/L of Fluoride
Alluvium	2368	0.01–2.77	0.47	0.38	48
Banded Gneissic Complex	380	0.05–1.77	0.71	0.35	8
Charnockite	2900	0.01–5.00	0.62	0.48	146
Gneiss	6479	0.01–5.00	0.77	0.51	563
Granite	307	0.01–2.50	0.88	0.49	32
Laterite	38	0.05–1.95	0.41	0.39	1
Limestone	67	0.05–1.50	0.59	0.36	0
Quartzite	27	0.15–1.68	0.93	0.48	5
Sandstone	1013	0.01–4.90	0.40	0.40	18
Shale	6	0.55–1.35	0.92	0.29	0
Total	13,585	0.01–5.00	0.66	0.49	821

3.3. Statistical Accuracy of Various Methods

Spatial prediction of groundwater fluoride by various methods is illustrated in Figure 2. The accuracy of the interpolation methods was assessed through evaluation metrics calculated for different validation methods (Table 2). The fluoride range and average concentration do not vary for the input in (1) LOOCV and (2) validation with independent datasets. This is because part of the dataset is not set aside for validation. However, in hold-out validation, the range and mean fluoride of the training set and the test set vary. For the training set, minimum, maximum, and mean fluoride were 0.03, 2.4 and 0.7 mg/L, respectively. In the test set, fluoride ranged from 0.05 to 2 mg/L, with a mean of 0.7 mg/L. Fluoride content for the independent data differed from 0.2 to 5.9 mg/L with a mean of 1.8 mg/L [37,38].

Table 2. Prediction accuracy of the different interpolation methods.

| Measure	IDW	RBF	LPI	OK	Gaussian Kriging	Spherical Kriging	Simple Kriging	UK	EBK
					LOOCV				
Predicted mean	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67	0.67
MRE	0.53	0.54	0.54	0.53	0.53	0.54	0.56	0.53	0.54
RMSE	0.32	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
CV, predicted (%)	42	37	36	37	38	38	32	37	37
r, measured vs. predicted	0.32	0.35	0.34	0.35	0.34	0.34	0.34	0.35	0.35
r, measured vs. error	0.47	0.58	0.60	0.58	0.55	0.55	0.68	0.58	0.58
					Hold-out validation				
Predicted mean	0.67	0.67	0.68	0.67	0.68	0.68	0.65	0.67	0.68
MRE	0.56	0.58	0.58	0.58	0.57	0.57	0.55	0.58	0.58
RMSE	0.32	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.31
CV, predicted (%)	41	36	34	36	37	37	34	36	35
r, measured vs. predicted	0.28	0.31	0.30	0.31	0.31	0.31	0.30	0.31	0.30
r, measured vs. error	0.49	0.58	0.62	0.58	0.57	0.57	0.64	0.58	0.60
					Validation with an independent dataset				
Predicted mean	0.86	0.86	0.87	0.86	0.86	0.85	0.82	0.86	0.86
MRE	0.50	0.51	0.50	0.50	0.50	0.50	0.53	0.50	0.50
RMSE	1.48	1.48	1.48	1.48	1.48	1.48	1.52	1.48	1.49
CV, predicted (%)	8	9	5	8	11	11	6	8	8
r, measured vs. predicted	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.02
r, measured vs. error	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00
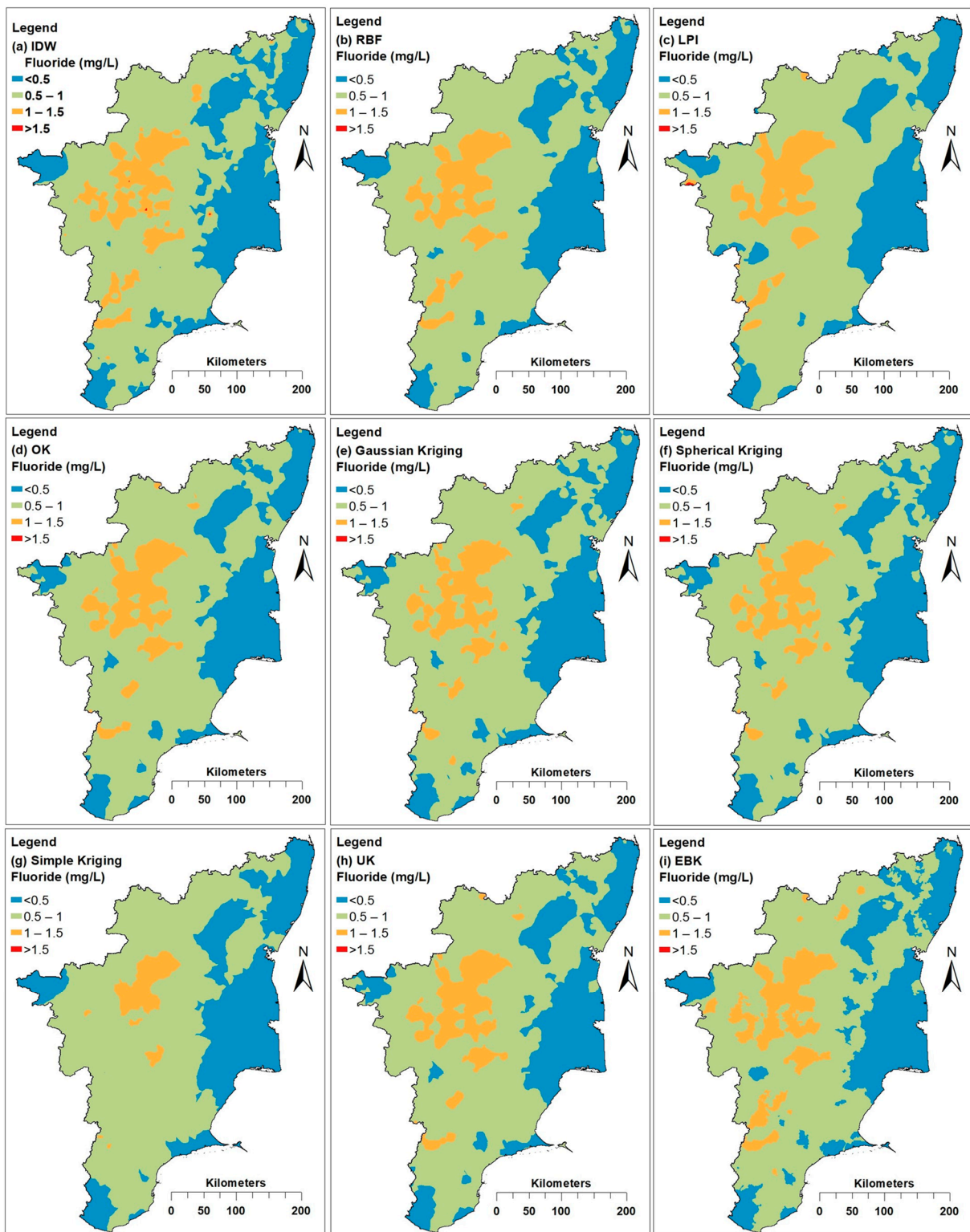
Figure 2. Predicted groundwater fluoride concentration by various interpolation methods.

Interpolation methods showed that there is a minor discrepancy between the predicted and measured means (Table 2). The lower the values of MRE and RMSE, the smaller the errors when using the methods. Simple kriging has the highest MRE, whereas IDW, OK, Gaussian kriging and UK delivered the smallest MRE by LOOCV (Table 2). Conversely, simple kriging was more accurate than the other interpolation methods, with the smallest MRE in the hold-out validation method. Moreover, OK, UK, EBK, RBF and LPI had larger

MRE values than the other methods. Most of the interpolation methods have smaller MRE for the validation subsets with independent dataset except for RBF and simple kriging (Table 2).

IDW interpolation has the maximum RMSE, while the other methods were slightly lower by LOOCV and hold-out validation (Table 2). Both validation methods resulted in similar values for all interpolation methods. RMSE was highest for all methods in the validation step with independent dataset as compared to LOOCV and hold-out validation. Among the interpolation methods, simple kriging had larger error followed by EBK and all other methods had the same RMSE. Of the three validation methods, validation with independent dataset had the lowest MRE but high RMSE.

CV for measured values is 56% in LOOCV and for the training-dataset. The CV of the measured values for the independent dataset was 61%. CV for predicted values by all interpolation methods (Table 2) were lower than the CV for the measured values. A significant reduction in CV was observed for the validation method with an independent dataset. IDW has the highest CV in LOOCV and hold-out validation. Gaussian and spherical kriging had the largest CV in the validation method with an independent dataset (Table 2).

3.4. Correlation and Prediction Error

The values of Pearson's correlation coefficient (r) calculated between the measured and predicted values are given in Table 2 and represented in Figures S1–S3. A slightly stronger correlation in comparison to other methods with r amounting to 0.35 was observed by OK, UK, EBK and RBF in LOOCV. Using the hold-out validation method, the kriging methods yielded a slightly higher r, amounting to 0.31, for OK, Gaussian, Spherical and UK. Of the deterministic methods, RBF has a higher r of 0.31 compared to r values of 0.28 for IDW and 0.30 for LPI when using the hold-out validation method.

The prediction error indicates the uncertainty associated with the predicted values at each location. The prediction error reflected in the interpolation maps is shown in Figures 3–5. The IDW interpolation method resulted in the lowest r between the measured fluoride and error in predicting fluoride, which amounted to 0.47 and 0.49 for LOOCV and hold-out validation, respectively (Table 2). Poor correlation (r) and the highest prediction error were observed for the validation method with independent dataset, which should be ascribed to the fact that the sampling wells of the independent dataset are densely located in the vicinity of each other, while only a few of them from the secondary dataset fall in this region (Table 2, Figure 6).

3.5. Prediction of Contaminated Areas Using Various Methods

The geospatial analysis of contaminated areas with fluoride concentration or the areas that require attention is performed using various interpolation methods (Figure 2). Comparison of the areas predicted by the interpolation methods are given in Table 3. IDW and LPI were the only methods that predicted areas with >1.5 mg/L fluoride. The remainder of the methods had predicted that the entire region had fluoride < 1.5 mg/L. This can be misleading, especially in studies such as the present one, where samples are collected over a large area with considerable distance between the samples. Simple kriging and UK predicted the lowest fluoride concentration, which is in the range between 1 and 1.5 mg/L. In comparison with the percentage of samples (~6%) and the predicted area in different fluoride range, the IDW and LPI methods provided closer predictions. Spatial interpolation in all methods showed similar patterns across areas in the central part and few patches in the southern part of the study area (Figure 2). Northern and eastern parts were found to not be prone to fluoride contamination. Most of central and western Tamil Nadu, and several small patches in the southern parts, recorded high fluoride.
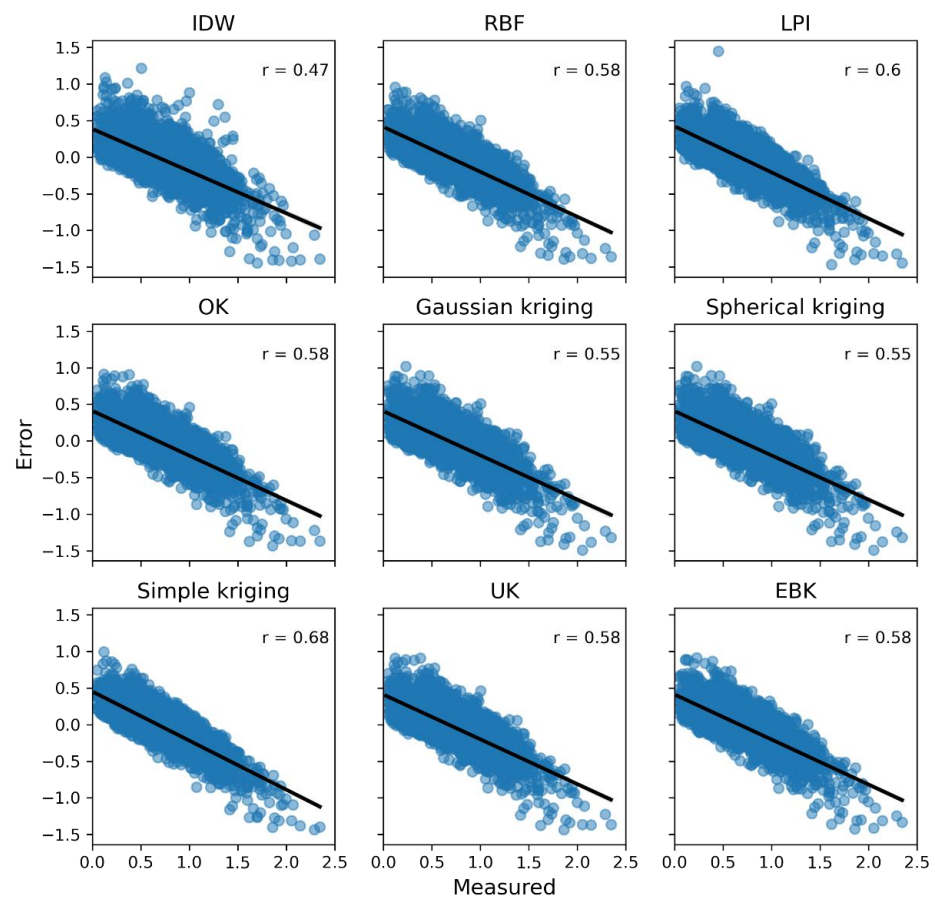
Figure 3. Measured fluoride concentration versus the calculated error in the LOOCV method.
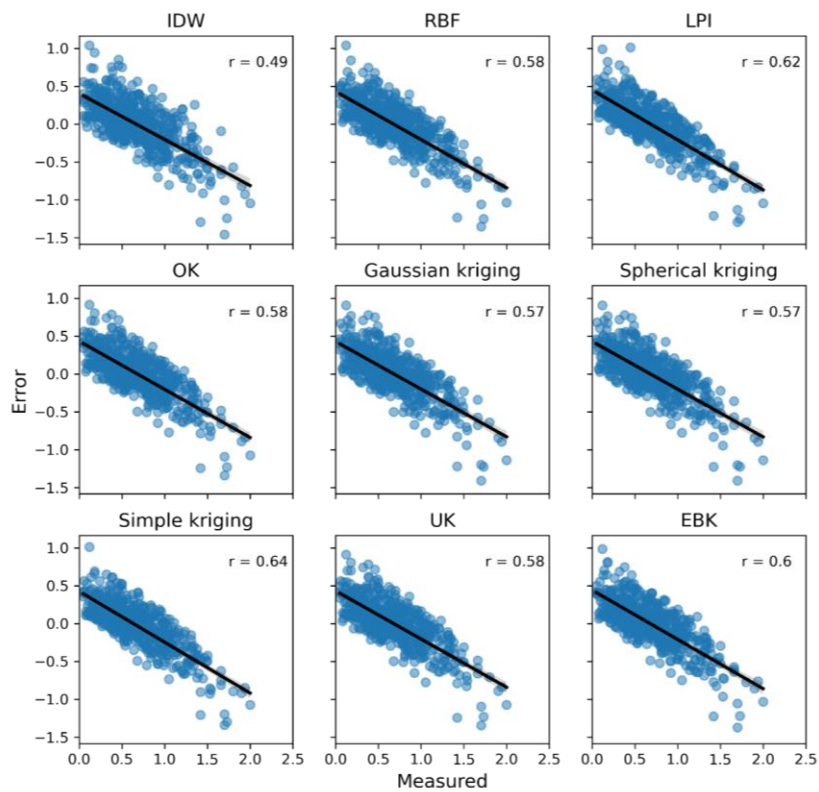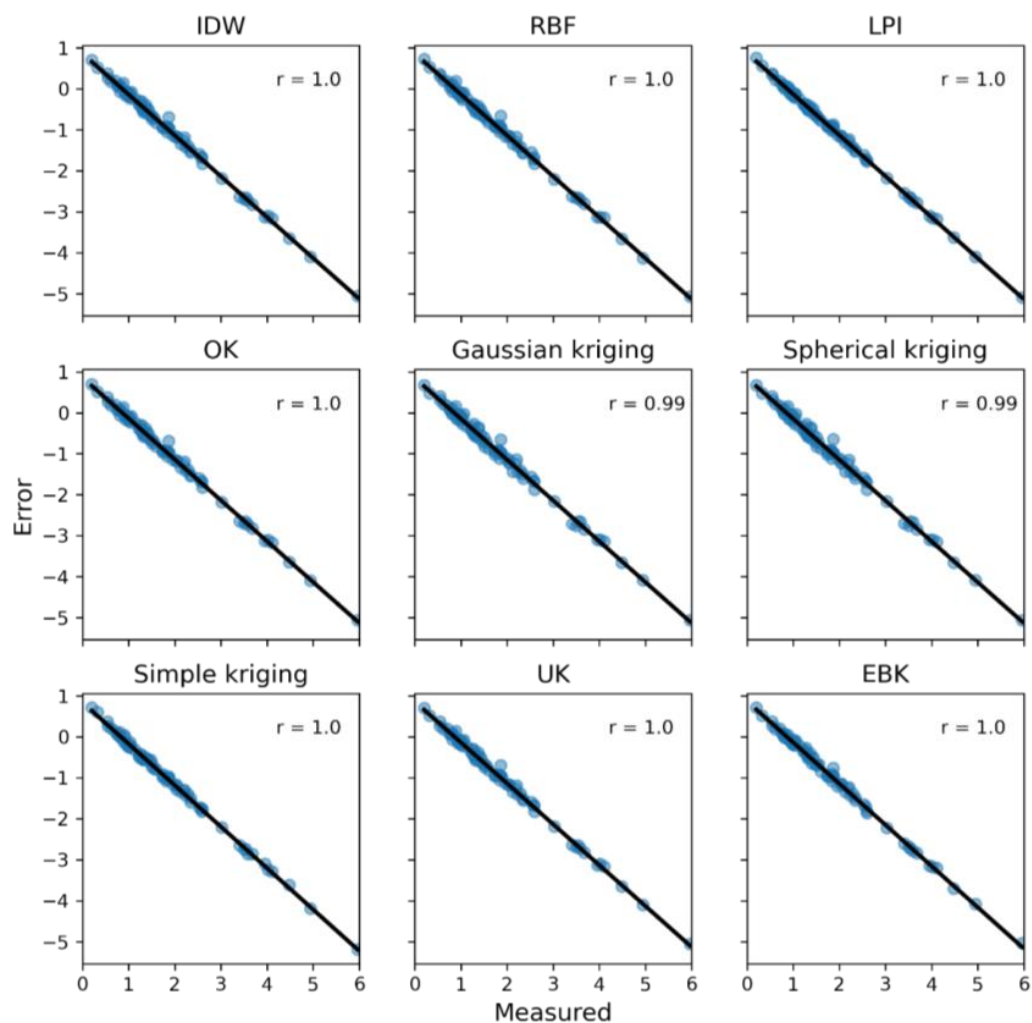

Figure 4. Measured fluoride concentration versus the calculated error in the hold-out validation method.

Figure 5. Measured fluoride concentration versus the calculated error in the validation method with an independent dataset.

Table 3. Fluoride contaminated area calculated by different interpolation methods.

| Method | Fluoride Range and Area in % | | | |
| | <0.5 | 0.5 to 1 | 1 to 1.5 | >1.5 |
	Very Low Fluoride	Low Fluoride	Suitable Range	Unsuitable
IDW	28.0	61.2	10.6	0.2
RBF	25.4	64.4	10.2	-
LPI	26.0	63.2	10.6	0.2
OK	26.5	63.6	9.9	-
Gaussian kriging	26.6	63.2	10.2	-
Spherical kriging	26.5	63.2	10.3	-
Simple kriging	29.6	65.9	4.5	-
UK	26.5	63.6	9.9	-
EBK	27.1	61.5	11.4	-
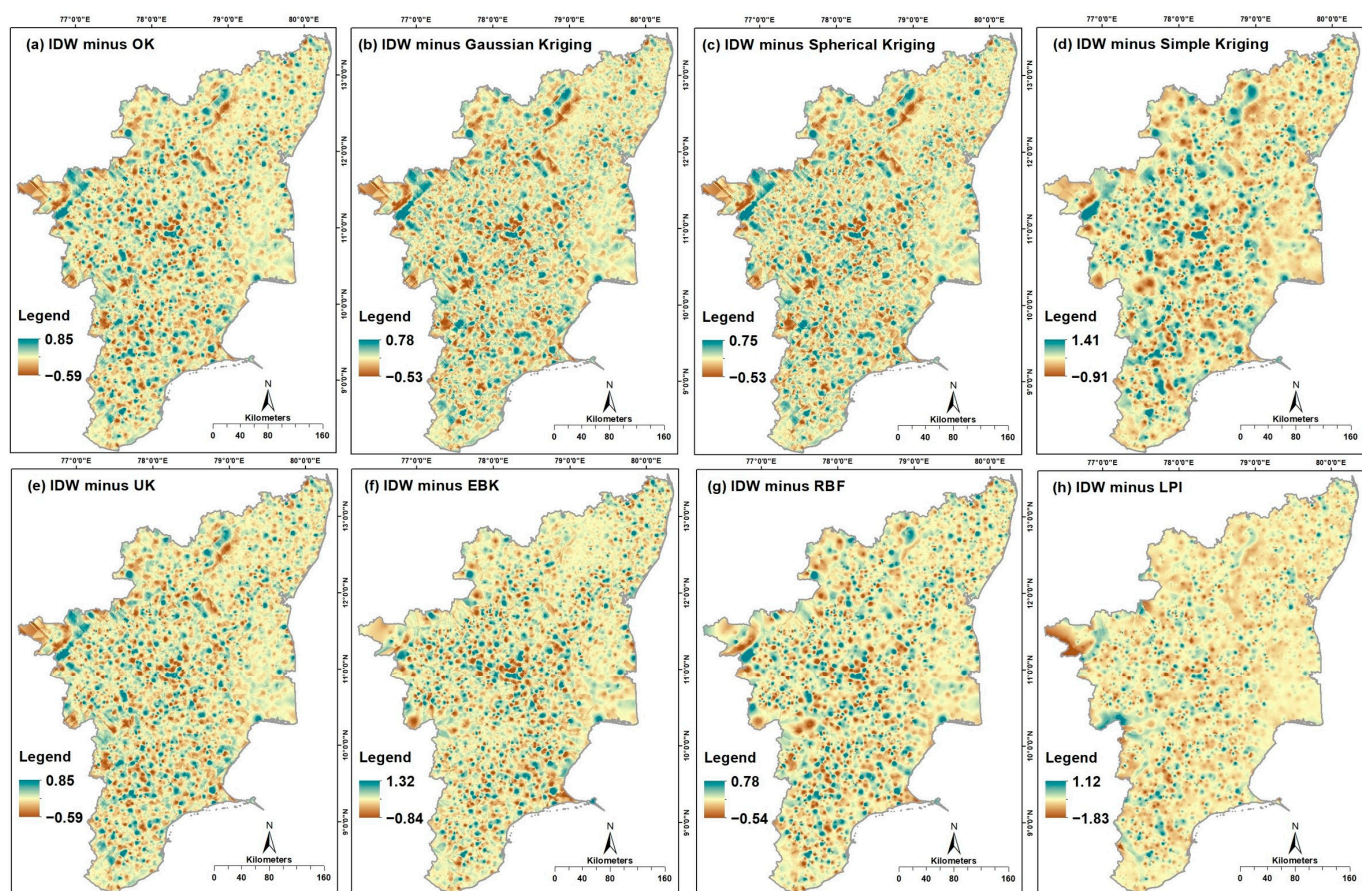
Figure 6. Over- and under-estimation of fluoride concentration obtained using all methods as compared with IDW (note that negative values indicate under-estimation and positive values indicate over-estimation).

*3.6. Over- and Under-Estimation of Contaminated Areas*

The fluoride concentration predicted by the interpolation methods were subtracted from one another to analyze the similarity or dissimilarity in prediction in the form of residuals (Figure 6, Table 4). Positive values indicate over-estimation of measured values by the predicted values, and negative values indicate under-estimation of measured values by the predicted values. Based on the previous results, IDW was found to be the most suitable method for the prediction of fluoride concentration. Hence, the under-estimation or over-estimation of the other interpolation methods was assessed in comparison to IDW. Figure 6 indicates the similarity and under-/over-estimation in the areas of prediction by IDW relative to other methods. RBF demonstrated the highest similarity to IDW in predicting the fluoride concentrations (Table 4). Simple kriging correlated poorly with IDW, owing to an over-estimation of fluoride. Gaussian kriging and LPI under-estimated the fluoride content than IDW. Residual histograms for the validation methods show more under-estimation in the fluoride values than over-estimation (Figures 7–9).

Table 4. Comparison of IDW method with other interpolation methods (area in %).

Comparison	Under-Estimated	Equal	Over-Estimated
IDW minus OK	11.2	77.9	10.9
IDW minus Gaussian Kriging	12.9	74.4	12.7
IDW minus Spherical Kriging	13.2	73.9	12.9
IDW minus Simple Kriging	11.5	63.3	25.2
IDW minus UK	11.2	77.9	10.9
IDW minus EBK	11	78.2	10.8
IDW minus RBF	7.2	85.1	7.7
IDW minus LPI	12.3	74.8	12.9
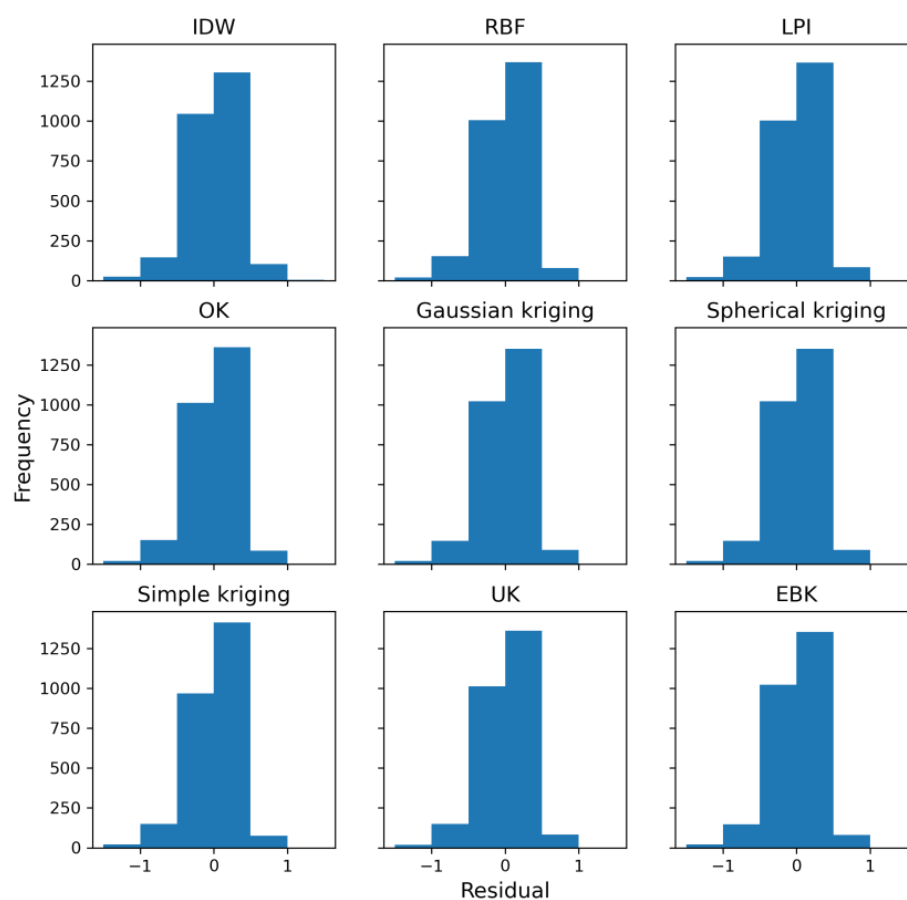
Figure 7. Residual histogram showing under-estimation and over-estimation in predicted values in the LOOCV method.

The estimation capability was further analyzed based on the geology (Table 1). Here, only IDW and LPI were considered, as they were the only methods to predict fluoride above the WHO [2] and BIS [39] standards and align with the direct assessment of the data given in Table 1. Additionally, the geology where the highest numbers of fluoride measurements were available, was further assessed (see Table 1 for number of measured fluoride concentrations). For example, shale, with six fluoride measurements, was not included in the analysis. A general trend of over-estimation, especially by 0.5 to 1 mg/L, were seen in both IDW and LPI in all aquifer types (Figures S4 and S5). However, the IDW prediction for granitic areas slightly underestimated fluoride by 0.5–1 mg/L. This variation was exceeded by only two data points.
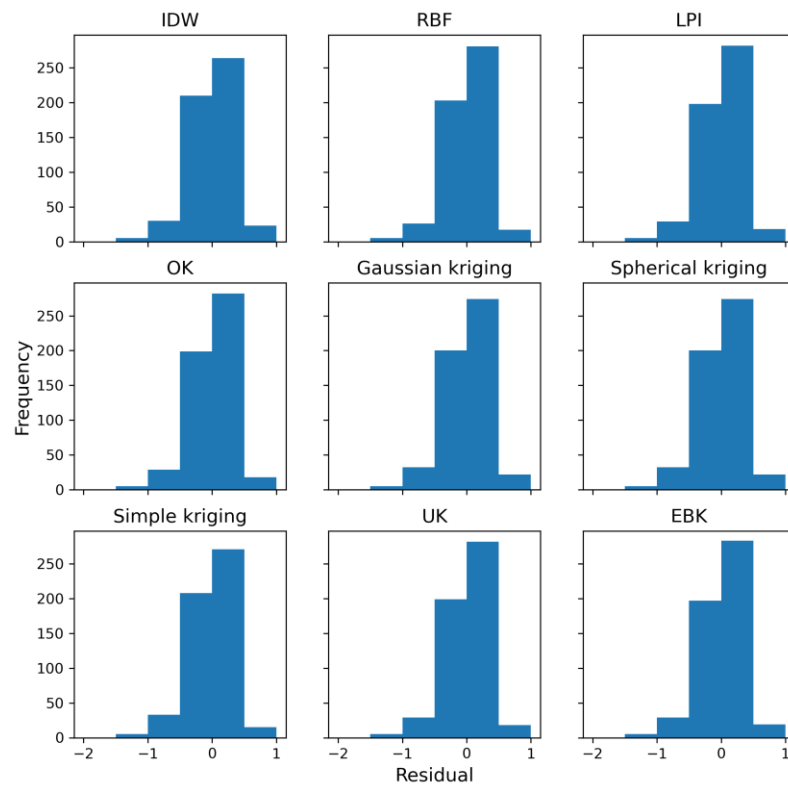
Figure 8. Residual histogram showing under-estimation and over-estimation in predicted values in the hold-out validation method.
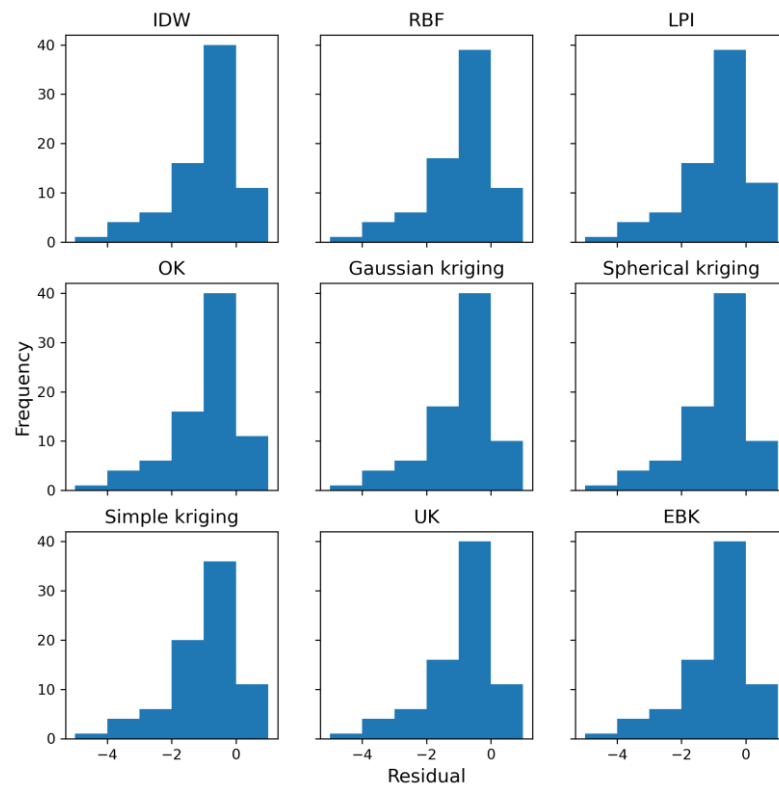
Figure 9. Residual histogram showing under-estimation and over-estimation in predicted values in the validation method with an independent dataset.

4. Discussion

Fluoride concentration in groundwater varies in time and space. In the present study, it was indicated that IDW and LPI methods outperformed other methods. These were the only methods that predicted vulnerable areas with high levels of fluoride up to 2.3 mg/L. Other methods predicted a minuscule area with the fluoride concentration above 1.5 mg/L in comparison with the total samples. The interpolation techniques smooth the values to minimize the estimated error of the global mean, while targeting to predict the values as accurately as possible in the spatial dimension [30]. This results in under-estimation of the local maximum values and over-estimation of the surrounding minimum values. Based on the level of the smoothing effects, the degree of under- and over-estimation varies. Low smoothing effect preserves as much measured data as possible, which is observed more in deterministic methods than stochastic ones. An optimal interpolation method should not rely on high smoothing to produce the results [46,47]. Studies have suggested that OK and EBK methods are superior to other methods, especially compared to IDW, as they have smoother effects [25,48]. A robust interpolation analysis should not rely on a great degree of smoothing effects.

The discrepancy among the interpolation methods was found to be minimal, although the evaluation metrics used in this study including MRE, RMSE and CV were lower for IDW than those of the other methods. Though IDW and LPI could estimate areas with high fluoride risk, these areas are not spatially identical in terms of the predicted fluoride concentration (Figure 2). Compared to the measured fluoride data in these locations, IDW could predict the spatial variation better than LPI. In national- and state-level observations, the monitoring wells are not often densely and evenly distributed across the region; rather, they are poorly and unevenly distributed with respect to each other. This is one of the reasons why high concentrations above 2 mg/L were not reported often in the secondary data collected (Table 1). However, many investigations conducted at local scales with high-density water sampling have reported concentrations above 2 mg/L [42,49,50], including the independent dataset used in the present study for validation [35,51]. Thus, the spatial distribution of wells play a significant role in forming the interpolation patterns [48]. To deal with this uncertainty, sampling sites/wells should be expanded at areas prone to high fluoride. Through this, we can confirm whether the presence of high concentration of fluoride in groundwater is a transient phenomenon, which is witnessed occasionally, or it is a long-standing and stable phenomenon, which requires considerable attention. Analyzing the sources of fluoride contamination in these areas will be of benefit in deciding on the distribution of the sampling network.

Border effects can also influence the interpolation patterns. Uncertainty was mostly identified along the boundary of the study area, where the coverage of the sampling network is not enough. This is the reason the prediction uncertainty reduces from the boundary to the inner zone [13,52]. In our study, we could not detect a considerable variation along the boundary. Under certain circumstances, when the nearest sampling site existing for interpolation is distant, the predicted geospatial maps for all interpolation methods showed a subtle difference and thus fell within a narrow range [13]. Uncertainty also increased with the weighing power used in some methods. For example, in IDW, the weighing power used during the interpolation influenced the CV. Normally, higher weighing power increases the CV, the associated error and the predicted extent of contaminated area [30,48]. In the present study, we used the most commonly used weighing power 2 that has provided a good fit.

Geostatistical methods, i.e., the various kriging methods, indicated more or less similar results, and their efficiency was not drastically different from the deterministic methods, i.e., IDW, LPI and RBF. This was well supported by the evaluation metrics (Table 2), but not for spatial interpretation (Table 3). Other studies have shown drastic variations in the interpolation methods for the targeted parameter with high MRE and RMSE. Such variations were not witnessed in this study as the concentration range did not show a wide difference such as that reported in former studies [24,30,48]. Comparison of the

concentrations predicted by the methods in this study showed closer prediction by IDW and RBF followed by EBK, OK and UK (Table 4).

Sample size also dictates the choice of validation method. Most studies work with smaller numbers of samples, which might not be statistically significant/sufficient for either hold-out validation or validation with an independent dataset. Hence, LOOCV is often employed [25,30]. This study, with a large dataset, provided the suitable opportunity to test three validation methods and evaluate the suitability/superiority of one interpolation method over another. Simple kriging produced the largest MRE when LOOCV and validation with an independent dataset were used, but the smallest MRE was obtained when using the hold-out validation method (Table 2). It should be noted that the training set and test set were chosen at random, therefore for a different set of training-data and test-data, the prediction by the methods may be different.

Results from the independent dataset showed poor correlation with the measured data from PWD wells. There were 78 sample data for the independent dataset and 22 samples for PWD over the same area, i.e., about one-third of the data was used to predict fluoride. Descriptive statistics for the independent data were as follows: range 0.2 to 5.9 mg/L; and mean was 1.8 mg/L. The minimum, maximum and average of the measured fluoride were 0.4, 1.6 and 0.9 mg/L, respectively. SDs for the PWD wells were smaller (0.3 mg/L) than the independent dataset (1.1 mg/L), indicating that the independent dataset comprised a higher variance. This proved that 68% of the data fell in the range between 0.6 and 1.2 mg/L for the PWD data and between 0.7 and 2.9 mg/L for the independent data for normally distributed data. In that case, the PWD data was much closer to the measured range than the independent dataset, thus suggesting that special circumstances leading to high concentrations were averaged-out in the PWD data, which is collected only twice a year. The shortest distance/space between PWD wells was 0.4 km and the highest was 12 km. Thus, the wells are not installed at equal intervals but are placed in a way to cover the large area except for the hilly regions. Nonetheless, due to the smaller study area in the independent study, the sample locations were distributed evenly and closely throughout the area. Here, one sample was collected within a radius of every 1 to 3 km^2, depending on the topography and geology, except for areas that were not accessible or there were no existing wells to monitor. Nevertheless, the fluoride concentration patterns were spatially correlated. Even though the *r* values were low, this did not mean that the values predicted by the interpolation methods were not useful, but rather that they can give a spatial orientation in the high or low values in the region.

In summary, interpolation techniques are often used to estimate contaminated areas or areas with potential risk of contamination. Ideal conditions for monitoring and collection of data are not always feasible. However, in light of the recent availability of open-source programs to create various maps, these interpolation methods have proved to be promising tools. Though the application of these methods is ample, procuring estimations/predictions reflect uncertainty. Effective prediction of the interpolation methods depends on several factors including (but not limited to) spatial location of samples, density of samples, number of samples, minimum and maximum concentration of the contaminant. Identifying a contaminated region is not a straightforward process and thus should not be undertaken via only predicted maps. Under-estimation of contamination may result in only short-term remediation, and the problem will re-emerge, while over-estimation may increase the burden on the government in the form of financial costs and manpower spent in resolving the problem. Hence, along with the interpolation methods, the natural background concentrations, land use and local experts' opinion should be included during the decision phase.

5. Conclusions

This study compares for the first time the accuracy and uncertainty of nine interpolation methods for predicting fluoride concentration over a large area on the basis of more than 13,000 observations. About 6% of the groundwater samples exceeded the guideline threshold of >1.5 mg/L of fluoride. Findings demonstrated that the prediction of contaminated areas obtained from the interpolation methods was lower than the traditional comparison of the number of samples above the desired limits. Deterministic methods provided closer prediction of contaminated areas than the geostatistical methods. To address the prediction uncertainty, one interpolation method should not be universally applied, while variegated interpolation algorithms should be employed and thus their resultant predictions should be compared for the purpose of the study. Interpolation methods would always commit errors to a certain extent. For stakeholders and decision makers who are already involved in regional- or national-level surveillance of water quality, these interpolation methods help to identify the most vulnerable areas to contamination. Additional samples should be collected from highly vulnerable areas to ascertain the seriousness of the problem before implementing possible remediation measures. From this study, it can be clearly concluded that relying on interpolation methods is not trustworthy and may result in biased remediation measures. IDW is appropriate for identifying fluoride risk zones in Tamil Nadu. Care should especially be given to over-estimated and under-estimated areas, because over-estimating the contaminated areas may increase the cost of the management practices, while under-estimation may not solve the problem thoroughly and the issue may resurface after a short time.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/w15091707/s1, Table S1: Area covered by different aquifer types in Tamil Nadu; Figure S1: Scatter plots of measured and predicted fluoride by LOOCV; Figure S2: Scatter plots of measured and predicted fluoride by hold-out validation; Figure S3: Scatter plots of measured and predicted fluoride by validation with an independent dataset; Figure S4: Residual histograms for various aquifer types by IDW interpolation; Figure S5: Residual histograms for various aquifer types by LPI.

References

1. Amini, M.; Mueller, K.; Abbaspour, K.C.; Rosenberg, T.; Afyuni, M.; Møller, K.N.; Sarr, M.; Johnson, C.A. Statistical Modeling of Global Geogenic Fluoride Contamination in Groundwaters. *Environ. Sci. Technol.* **2008**, *42*, 3662–3668. [CrossRef] [PubMed]
2. WHO. *Guidelines for Drinking Water Quality*, 4th ed.; World Health Organization: Geneva, Switzerland, 2011.
3. Podgorski, J.E.; Labhasetwar, P.; Saha, D.; Berg, M. Prediction Modeling and Mapping of Groundwater Fluoride Contamination throughout India. *Environ. Sci. Technol.* **2018**, *52*, 9889–9898. [CrossRef] [PubMed]
4. Chakraborti, D.; Rahman, M.M.; Chatterjee, A.; Das, D.; Das, B.; Nayak, B.; Pal, A.; Chowdhury, U.K.; Ahmed, S.; Biswas, B.K.; et al. Fate of over 480 million inhabitants living in arsenic and fluoride endemic Indian districts: Magnitude, health, socio-economic effects and mitigation approaches. *J. Trace Elem. Med. Biol.* **2016**, *38*, 33–45. [CrossRef]
5. Raju, N.J. Prevalence of fluorosis in the fluoride enriched groundwater in semi-arid parts of eastern India: Geochemistry and health implications. *Quat. Int.* **2017**, *443*, 265–278. [CrossRef]
6. Nayak, B.; Roy, M.M.; Chakraborti, D. Dental fluorosis. *Clin. Toxicol.* **2009**, *47*, 355. [CrossRef]

7.   Yadugiri, V.T. Fluorosis: A persistent problem. *Curr. Sci.* **2011**, *100*, 1475–1477.

8.   Bhowmik, A.D.; Shaw, P.; Mondal, P.; Munshi, C.; Chatterjee, S.; Bhattacharya, S.; Chattopadhyay, A. Incidence of fluorosis and urinary fluoride concentration are not always positively correlated with drinking water fluoride level. *Curr. Sci.* **2019**, *116*, 1551–1554. [CrossRef]

9.   Chakraborti, D.; Chanda, C.; Samanta, G.; Chowdhury, U.; Mukherjee, S.; Pal, A.; Sharma, B.; Mahanta, K.; Ahmed, H.; Sing, B. Fluorosis in Assam, India. *Curr. Sci.* **2000**, *78*, 1421–1423.

10.   Sahoo, P.K.; Ray, S.B.; Kerketta, A.; Behera, P.; Neogi, G.; Sahoo, H.B. Geogenic enrichment of fluoride in groundwater of hard rock aquifer in fluorosis prevalent area of Balangir district, Odisha, India. *Groundw. Sustain. Dev.* **2022**, *19*, 100830. [CrossRef]

11.   Jaydhar, A.K.; Chandra Pal, S.; Saha, A.; Islam, A.R.M.T.; Ruidas, D. Hydrogeochemical evaluation and corresponding health risk from elevated arsenic and fluoride contamination in recurrent coastal multi-aquifers of eastern India. *J. Clean. Prod.* **2022**, *369*, 133150. [CrossRef]

12.   Gotway, C.A.; Ferguson, R.B.; Hergert, G.W.; Peterson, T.A. Comparison of Kriging and Inverse-Distance Methods for Mapping Soil Parameters. *Soil Sci. Soc. Am. J.* **1996**, *60*, 1237–1247. [CrossRef]

13.   Liu, R.; Chen, Y.; Sun, C.; Zhang, P.; Wang, J.; Yu, W.; Shen, Z. Uncertainty analysis of total phosphorus spatial–temporal variations in the Yangtze River Estuary using different interpolation methods. *Mar. Pollut. Bull.* **2014**, *86*, 68–75. [CrossRef]

14.   Adhikary, P.P.; Dash, C.J. Comparison of deterministic and stochastic methods to predict spatial variation of groundwater depth. *Appl. Water Sci.* **2017**, *7*, 339–348. [CrossRef]

15.   Amini, M.A.; Torkan, G.; Eslamian, S.; Zareian, M.J.; Adamowski, J.F. Analysis of deterministic and geostatistical interpolation techniques for mapping meteorological variables at large watershed scales. *Acta Geophys.* **2019**, *67*, 191–203. [CrossRef]

16.   Ahmad, A.Y.; Saleh, I.A.; Balakrishnan, P.; Al-Ghouti, M.A. Comparison GIS-Based interpolation methods for mapping ground-water quality in the state of Qatar. *Groundw. Sustain. Dev.* **2021**, *13*, 100573. [CrossRef]

17.   Bronowicka-Mielniczuk, U.; Mielniczuk, J.; Obroślak, R.; Przystupa, W. A Comparison of Some Interpolation Techniques for Determining Spatial Distribution of Nitrogen Compounds in Groundwater. *Int. J. Environ. Res.* **2019**, *13*, 679–687. [CrossRef]

18.   Wameling, A. Accuracy of geostatistical prediction of yearly precipitation in Lower Saxony. *Environmetrics* **2003**, *14*, 699–709. [CrossRef]

19.   Shahmohammadi-Kalalagh, S.; Taran, F. Evaluation of the classical statistical, deterministic and geostatistical interpolation methods for estimating the groundwater level. *Int. J. Energy Water Resour.* **2021**, *5*, 33–42. [CrossRef]

20.   Adhikary, P.P.; Dash, C.J.; Bej, R.; Chandrasekharan, H. Indicator and probability kriging methods for delineating Cu, Fe, and Mn contamination in groundwater of Najafgarh Block, Delhi, India. *Environ. Monit. Assess.* **2011**, *176*, 663–676. [CrossRef]

21.   Murphy, R.R.; Curriero, F.C.; Ball, W.P. Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay. *J. Environ. Eng.* **2010**, *136*, 160–171. [CrossRef]

22.   Farzaneh, G.; Khorasani, N.; Ghodousi, J.; Panahi, M. Application of geostatistical models to identify spatial distribution of groundwater quality parameters. *Environ. Sci. Pollut. Res.* **2022**, *29*, 36512–36532. [CrossRef] [PubMed]

23.   Elumalai, V.; Brindha, K.; Sithole, B.; Lakshmanan, E. Spatial interpolation methods and geostatistics for mapping groundwater contamination in a coastal area. *Environ. Sci. Pollut. Res.* **2017**, *24*, 11601–11617. [CrossRef] [PubMed]

24.   Gong, G.; Mattevada, S.; O'Bryant, S.E. Comparison of the accuracy of kriging and IDW interpolations in estimating groundwater arsenic concentrations in Texas. *Environ. Res.* **2014**, *130*, 59–69. [CrossRef]

25.   Magesh, N.S.; Elango, L. Spatio-Temporal Variations of Fluoride in the Groundwater of Dindigul District, Tamil Nadu, India: A Comparative Assessment Using Two Interpolation Techniques. In *GIS and Geostatistical Techniques for Groundwater Science*; Senapathi, V., Viswanathan, P.M., Chung, S.Y., Eds.; Elsevier: Amsterdam, The Netherlands, 2019; pp. 283–296.

26.   CGWB. *Aquifer Systems of Tamilnadu and Puducherry, Central Ground Water Board, South Eastern Coastaö Region, Chennai*; Ministry of Water Resources, Government of India: New Delhi, India, 2012.

27.   Shepard, D. A two-dimensional interpolation function for irregularly-spaced data. In Proceedings of the 1968 ACM National Conference, New York, NY, USA, 27–29 August 1968; pp. 517–524.

28.   Aguilar, F.; Agüera, F.; Aguilar, M.; Carvajal, F. Effects of Terrain Morphology, Sampling Density, and Interpolation Methods on Grid DEM Accuracy. *Photogramm. Eng. Remote Sens.* **2005**, *71*, 805–816. [CrossRef]

29.   Mitášová, H.; Mitáš, L. Interpolation by regularized spline with tension: I. Theory and implementation. *Math. Geol.* **1993**, *25*, 641–655. [CrossRef]

30.   Xie, Y.; Chen, T.-B.; Lei, M.; Yang, J.; Guo, Q.-J.; Song, B.; Zhou, X.-Y. Spatial distribution of soil heavy metal pollution estimated by different interpolation methods: Accuracy and uncertainty analysis. *Chemosphere* **2011**, *82*, 468–476. [CrossRef] [PubMed]

31.   Webster, R.; Oliver, M.A. *Geostatistics for Environmental Scientists*, 2nd ed.; Statistics in Practice; John Wiley & Sons, Ltd.: Chichester, UK, 2007.

32.   Ryu, J.-S.; Kim, M.; Cha, K.-J.; Lee, T.H.; Choi, D. Kriging interpolation methods in geostatistics and DACE model. *KSME Int. J.* **2002**, *16*, 619–632. [CrossRef]

33.   Lichtenstern, A. Kriging Methods in Spatial Statistics. Bachelor's Thesis, Technische Universität Müunchen, Munich, Germany, 2013; p. 97.

34.   Krivoruchko, K. *Empirical Bayesian Kriging*; ArcUser Fall; Esri Press: Redlands, CA, USA, 2012; pp. 6–10.

35.   Kalpana, L.; Brindha, K.; Elango, L. FIMAR: A new Fluoride Index to mitigate geogenic contamination by Managed Aquifer Recharge. *Chemosphere* **2018**, *220*, 381–390. [CrossRef]

36. Brindha, K.; Jagadeshan, G.; Kalpana, L.; Elango, L. Fluoride in weathered rock aquifers of southern India: Managed Aquifer Recharge for mitigation. *Environ. Sci. Pollut. Res. Int.* **2016**, *23*, 8302–8316. [CrossRef]

37. Kalpana, L. Groundwater Quality with Special Reference to Fluoride and Groundwater Modelling for Simulating the Effect of Managed Aquifer Recharge in Pambar Basin, India. Unpublished. Ph.D. Thesis, Anna University, Chennai, India, 2014.

38. Jagadeshan, G. Geochemical Reactions Responsible for Fluoride Rich Groundwater and Remediation by Induced Recharge in Vaniyar River Basin, Tamil Nadu, India. Unpublished. Ph.D. Thesis, Anna University, Chennai, India, 2015.

39. *IS10500*; Indian Standard Drinking Water Specification. Bureau of Indian Standards (BIS): New Delhi, India, 2012.

40. WHO; UNICEF. *Progress on Drinking Water, Sanitation and Hygiene: 2017 Update and SDG Baselines. Geneva: World Health Organization (WHO) and the United Nations Children's Fund (UNICEF)*; Licence: CC BY-NC-SA 3.0 IGO; WHO: Geneva, Switzerland, 2017; p. 108.

41. Brindha, K.; Elango, L. Hydrochemical characteristics of groundwater for domestic and irrigation purposes in Madhuranthakam, Tamil Nadu, India. *Earth Sci. Res. J.* **2011**, *15*, 101–108.

42. Karthikeyan, K.; Nanthakumar, K.; Velmurugan, P.; Tamilarasi, S.; Lakshmanaperumalsamy, P. Prevalence of certain inorganic constituents in groundwater samples of Erode district, Tamilnadu, India, with special emphasis on fluoride, fluorosis and its remedial measures. *Environ. Monit. Assess.* **2010**, *160*, 141–155. [CrossRef]

43. Jagadeshan, G.; Kalpana, L.; Elango, L. Hydrogeochemistry of high fluoride groundwater in hard rock aquifer in a part of Dharmapuri district, Tamil Nadu, India. *Geochem. Int.* **2015**, *53*, 554–564. [CrossRef]

44. Nair, I.S.; Brindha, K.; Elango, L. Identification of salinization by bromide and fluoride concentration in coastal aquifers near Chennai, southern India. *Water Sci.* **2016**, *30*, 41–50. [CrossRef]

45. Thivya, C.; Chidambaram, S.; Rao, M.S.; Thilagavathi, R.; Prasanna, M.V.; Manikandan, S. Assessment of fluoride contaminations in groundwater of hard rock aquifers in Madurai district, Tamil Nadu (India). *Appl. Water Sci.* **2017**, *7*, 1011–1023. [CrossRef]

46. Falivene, O.; Cabrera, L.; Tolosana-Delgado, R.; Sáez, A. Interpolation algorithm ranking using cross-validation and the role of smoothing effect. A coal zone example. *Comput. Geosci.* **2010**, *36*, 512–519. [CrossRef]

47. Falivene, O.; Cabrera, L.; Sáez, A. Optimum and robust 3D facies interpolation strategies in a heterogeneous coal zone (Tertiary As Pontes basin, NW Spain). *Int. J. Coal Geol.* **2007**, *71*, 185–208. [CrossRef]

48. Mirzaei, R.; Sakizadeh, M. Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. *Environ. Sci. Pollut. Res. Int.* **2016**, *23*, 2758–2769. [CrossRef]

49. Singaraja, C.; Chidambaram, S.; Jacob, N.; Johnson Babu, G.; Selvam, S.; Anandhan, P.; Rajeevkumar, E.; Balamurugan, K.; Tamizharasan, K. Origin of high fluoride in groundwater of the Tuticorin district, Tamil Nadu, India. *Appl. Water Sci.* **2018**, *8*, 54. [CrossRef]

50. Manikandan, S.; Chidambaram, S.; Ramanathan, A.L.; Prasanna, M.V.; Karmegam, U.; Singaraja, C.; Paramaguru, P.; Jainab, I. A study on the high fluoride concentration in the magnesium-rich waters of hard rock aquifer in Krishnagiri district, Tamilnadu, India. *Arab. J. Geosci.* **2014**, *7*, 273–285. [CrossRef]

51. Jagadeshan, G.; Kalpana, L.; Elango, L. Major ion signatures for identification of geochemical reactions responsible for release of fluoride from geogenic sources to groundwater and associated risk in Vaniyar River basin, Dharmapuri district, Tamil Nadu, India. *Environ. Earth Sci.* **2015**, *74*, 2439–2450. [CrossRef]

52. Pan, H.; Huang, W.-Q. Influence of uncertainty in delimitation of seismic statistical zone on results of PSHA. *Acta Seismol. Sin.* **2003**, *16*, 213–218. [CrossRef]