An Evaluation of Interpolation Techniques for Reconstructing Ionospheric TEC Maps

Matthew P. Foster, *Student Member, IEEE*, and Adrian N. Evans, *Member, IEEE*

***Abstract*—Maps of the total electron content (TEC) of the ionosphere can be reconstructed using data extracted from global positioning system (GPS) signals. For historic and other sparse data sets, the reconstruction of TEC images is often performed using a multivariate interpolation technique. Although there are many interpolation methods available, only a limited number, for example kriging, have been applied to TEC data. This paper presents a quantitative comparison of various commonly used algorithms for scattered-data interpolation over a range of sparsities. Techniques evaluated include a relatively new approach called Adaptive Normalized Convolution (ANC) that has not previously been applied to ionospheric reconstruction. The proposed evaluation scheme employs a quantitative methodology applied to both simulated and real TEC data. Results show that, although the performance of kriging is good in many cases, it is several times worse than the best performing techniques at some sparsities. Natural-neighbor interpolation has a better overall performance than kriging for both simulated and TEC data. Although its performance is a few percent worse than other methods for the simulated data, ANC produces the best performance for the TEC reconstructions.**

***Index Terms*—Image reconstruction, interpolation, ionosphere, remote sensing.**

I. Introduction

THE DISTRIBUTION of electrons in the ionosphere is of interest to scientists and also to engineers working on applications such as earth–space communication systems, which must transmit through the ionosphere, and skywave systems, which make use of ionospheric refraction. Electron content is commonly examined using total electron content (TEC) mapping. This mapping finds use in other applications, such as studying the evolution of magnetic storms, which have, in the past, had profound effects on satellite communication systems and on other critical ground-based systems, such as the U.S. power grid. Information on the electron content of the ionosphere can be collected using the global positioning system (GPS) and by examining the phase and amplitude changes which occur in paths between transmitting satellites and ground-based receivers. These data can then be processed in order to create maps of the ionospheric TEC.

As the number of paths between GPS ground stations and satellites is relatively low, producing TEC maps is an exercise in reconstruction from sparse data. Recent research has mainly focused on methods, such as tomography, that provide time-dependent volumetric reconstructions [1], [2]. However, when the data points are too sparsely distributed, these techniques are underconstrained and do not produce meaningful results. In ionospheric studies, problems relating to sparsity are particularly prevalent in historic data sets. For example, in 1992, there were only 25 receiver sites operated by the International GPS Service (IGS) in the U.S. [3], by 1996, there were over 75, and now, there are over 500. Therefore, while the issues due to undersampling have largely disappeared for TEC imaging systems utilizing modern GPS data, they still remain for older data and regularly arise in other geoscience applications [4], [5]. Consequently, interpolation methods still have an important role to play in ionospheric studies. The most commonly used interpolation technique for TEC-mapping studies is kriging [6]–[8]. Although these studies have generally found kriging to perform satisfactorily, in other geophysical applications, numerous problems with the kriging method have been reported [9]. In addition, there are many other interpolation methods for geophysical data that have received little recent attention from the ionospheric imaging community.

The overall aim of this paper is to assess the performance of currently available multivariate interpolation techniques for ionospheric TEC mapping. The need to establish the relative performance of scattered-data-interpolation schemes has been recognized and partly addressed in the past (see for example [10]). However, only very specific case studies exist involving more up-to-date methods [6], [11], [12]. This paper considers, for the first time, the specific application of such techniques to TEC mapping and presents the results of a comprehensive quantitative evaluation, using both simulated data and real ionospheric electron-content measurements. Schemes evaluated include those previously used for TEC mapping (e.g., kriging), interpolation methods commonly used in other fields [such as interpolation based on Voronoi tessellations and radial basis functions (RBFs)], and schemes in use for other geophysical applications (such as natural-neighbor interpolation). We further propose the application of Adaptive Normalized Convolution (ANC) to the problem of TEC mapping and quantify its performance in comparison with extant techniques. ANC is a recently proposed interpolation scheme that has found application to the reconstruction of data with varying spatial frequency content, orientation, and anisotropy. As these properties are also found in TEC images, their reconstruction using ANC appears as an attractive proposition. The proposed objective evaluation scheme enables the benefits conferred by ANC to be quantitatively assessed.

Manuscript received October 2, 2007; revised December 18, 2007.

The authors are with the Department of Electronic and Electrical Engineering, University of Bath, BA2 7AY Bath, U.K. (e-mail: m.p.foster@bath.ac.uk; a.n.evans@bath.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2008.916642

The organization of this paper is as follows. Section II describes various commonly used state-of-the-art interpolation schemes. Normalized convolution (NC) and ANC are introduced and explained in Section III. The quantitative evaluation methodology is outlined in Section IV and used to determine the performance of the interpolation schemes described in Sections II and III. Finally, discussion and conclusions are presented in Section V.

II. Interpolation Schemes for Scattered Data

Scattered-data interpolation has been studied for many years and in many fields—because of this, it has been given many names. The term scattered, for example, is also referred to as "spatial," or "multivariate," and the term interpolation is often called "reconstruction" or, less formally, "approximation." An interesting history of interpolation from ancient times is provided by Meijering [13]. Although its fundamental concepts do not differ, multivariate interpolation is a more recent development. In their study of the mathematical development of multivariate interpolation up to the second half of the 20th century, Gasca and Sauer [14] cite the first modern literature on multivariate interpolation as the work of Borchardt and Kronecker, which appeared in 1860 and 1865, respectively.

Sparsity is a term that is often associated with scattered data but which can only be defined relative to the desired resolution of reconstruction. Empirically, if there are not enough data points to fully reconstruct every point in the field at all desired spatial frequencies, then the data are undersampled or sparse. Altering the resolution of the reconstruction changes the relative sparsity by changing the number of spatial frequencies which are harmonics of the field size. Although no single definition of sparsity exists, in this paper, we consider data to be sparse if values exist at fewer than 5% of the discrete elements present in the reconstructed field.

Interpolation methods can be divided into two categories, local and global, depending upon the locality of the points which are used to derive a given output point. Local techniques make use of a definition of locality to compute output values; only data which fall within a given point's local neighborhood are used to calculate output values. Global techniques use a weighted sum of all data to compute output values, and for large numbers of input points, an approximation is generally used. When a new datum is added to a globally interpolated field, the whole field must be recalculated, whereas for a locally interpolated field, only those positions within the neighborhood of the added datum need to be recalculated. These two points tend to favor the use of local techniques.

The interpolation schemes evaluated in this paper represent a broad cross section of those in common use. Specifically, they are as follows:

1) triangulation based (nearest neighbor [15], linear [16], and cubic [17]), Section II-A;
2) natural neighbor [18], Section II-B;
3) RBF [19]–[21], Section II-C;
4) biharmonic spline [22], Section II-C1;
5) ordinary kriging [23], [24], Section II-D.

Of the list earlier, only ordinary kriging, RBF interpolation, and biharmonic spline interpolation (BSI) are considered truly global techniques. Natural neighbor, nearest neighbor, and triangulation-based interpolations all use a neighborhood defined by the Delaunay triangulation of the input data coordinates.

A. Triangulation-Based Interpolation

Triangulations are often used as the basis for interpolation of irregular data—the Delaunay triangulation (described as follows) has some properties which make it particularly useful, and so, practically, most triangulations can be assumed to be of this type. Once a suitable triangulation has been obtained, each point in the data set will be connected to several others by triangle vertices. Given the values of a triangle's nodes (f_i, where $i = 1, 2, 3$), the interpolated value of any point within the triangle, can be found using

$$f(x, y) = \sum_{i=1}^{3} \phi_i(x, y) f_i \tag{1}$$

where $\phi_i(\mathbf{x})$ is the interpolating basis function, which weights the contributions of the inputs. For a simple case, linear interpolation, the basis function can be replaced by a simple first-order polynomial

$$f(x, y) = c_1 x + c_2 y + c_3. \tag{2}$$

The coefficients $\mathbf{c} = (c_1, c_2, c_3)$ can then be found by solving $\mathbf{Ac} = \mathbf{f}$, where $\mathbf{f} = (f_1, f_2, f_3)^{\mathrm{T}}$ and $\mathbf{A}$ is a 3×3 matrix of rows with the form $(x_i, y_i, 1)$, where i is the row number.

Higher (and lower) order basis functions can also be used but require larger numbers of input samples. Zeroth-order triangulation-based interpolation is known as nearest neighbor interpolation. Other commonly used schemes include quadratic and cubic interpolation. This process can also be generalized to higher dimensions.

B. Natural-Neighbor Interpolation

Watson defines natural neighbors as "points which share a common interface, or region, that is equally close to each of the pair, and all other neighbors are no closer" [25]. This means that if circles (or n-spheres in n dimensions) are drawn such that their circumferences pass through $n + 1$ or more data points, no data points will be within any of the n-spheres. This is related to the Delaunay triangulation, which can be formed by linking the data at points which are on the circumference of each n-sphere. The Delaunay triangulation is not unique when more than $n + 1$ points lie on a sphere edge (i.e., when points are coplanar).

Once the natural neighbors have been established, the interpolated output value at any point can be determined using a weighted sum of the values at its neighbors. The way in which the weights are determined is best described in terms of Voronoi tessellations, the geometric dual of the Delaunay triangulation (see Fig. 1).

For each point where a value is required, the following operations are performed.

1) Assume that the data are already tessellated.

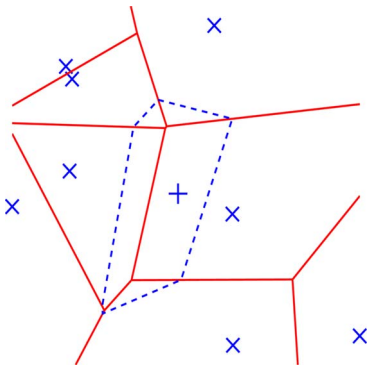

Fig. 1. Voronoi diagram showing (dotted line) new cell overlapping cells from the original tesselation.

TABLE I
EXAMPLE RBFS

Name	Equation
Linear	$\phi(r) = r$
Thin-plate Spline	$\phi(r) = r^2 \log r$
Multiquadratic	$\phi(r) = (r^2 + c^2)^{0.5}$
Inverse Multiquadratic	$\phi(r) = (r^2 + c^2)^{-0.5}$
Gaussian	$\phi(r) = e^{-ar^2}$
Biharmonic Spline	$\phi(r) = \lvert r\rvert^2(\log \lvert r\rvert - 1)$

2) Retessellate the data to include the output point. This adds a new Voronoi cell which overlaps the cells of the natural neighbors of the output point.
3) The contribution from each neighbor is given by the ratio of the area of overlap to the total area of the new cell. These ratios form the basis function $\phi_i(x, y)$ in (1).

In terms of the Delaunay triangulation, the basis function $\phi_i(x, y)$ is only nonzero within the circum-circles, which pass through the natural neighbor nodes. This means that the operation is local, in the sense that only neighboring values are used in the interpolation.

C. RBF Interpolation

RBF interpolation approximates a field of data using a weighted sum of radially symmetrical functions, known as basis functions [19]. One basis function is centered on each input sample, so that any given output point is composed of contributions from each input point. RBF interpolation is therefore considered as a global technique. The output at any given point $\mathbf{x}$ is given by

$$f(\mathbf{x}) = p_m(\mathbf{x}) + \sum_{j=1}^{N} \lambda_j \phi\left(\|\mathbf{x} - \mathbf{x}_j\|\right) \tag{3}$$

where p_m is a low-order polynomial surface with coefficients $c_0, c_1, \ldots, c_n$, which has been fitted to the data and is only used during linear and thin-plate spline interpolation, ϕ is the basis function whose form is fixed across the field, and λ_i is the weight for input $\mathbf{x}_i$. Many different basis functions can be used, with differing performance and order of continuity. Some common functions are shown in Table I [20], [21].

To find values for the weights λ and coefficients c, the linear system

$$\begin{bmatrix} A & P \\ P^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ c \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} \tag{4}$$

must be solved, where A is a matrix composed of evaluated basis-function values for every possible pair of input values, P is a matrix of input coordinates with leading ones

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix} \tag{5}$$

and $[\lambda\ c]^{\mathrm{T}}$ and $f = (f_1, f_2, \ldots, f_n)^{\mathrm{T}}$ are column vectors of weights and input values, respectively.

When a polynomial is not being fitted, the output system reduces to

$$A\lambda = f. \tag{6}$$

The calculation of the matrix A and solving the linear system described by (4) are computationally expensive operations, and this has motivated work aimed at decreasing the overall complexity, using techniques such as domain decomposition [26]. In addition to the basis functions given in Table I, some other interpolation methods can be formulated in terms of RBF interpolation. A well-known example is BSI, and this is described in more detail as follows.

1) BSI: BSI is a method of finding the minimum-curvature surface, which passes through a set of points [22]. It is fairly similar to cubic spline interpolation but simpler to compute and can be expressed as a basis function for use in RBF interpolation. The fitted surface is a linear combination of Green functions, and the BSI basis function is given in the last line of Table I. Comparison with other basis functions in the table shows the biharmonic spline to have a similar form to the thin-plate spline. The equations for other dimensions are given by Sandwell [22]. BSI can be numerically unstable for large numbers of points and, like cubic spline interpolation, has a tendency to drastically overshoot when points are close together. This problem occurs because of the imposed continuity in the surface's derivatives, which makes smoothly varying curves preferable around data points. Therefore, BSI is better suited to the interpolation of highly sparse data.

D. Kriging

Kriging was first suggested and developed in the 1960s by D. G. Krige, a South African mining engineer. It was originally developed as a technique for estimating yields of ore deposits from sparsely distributed core samples but has now been widely applied to many different fields and scenarios (for example, mining, mathematics, and classification [27]), as discussed in [28]. One of the main attractions of kriging is its ability to provide a variance estimate for each output point. Kriging and all geostatistical methods operate under the assumption that a process being interpolated or analyzed consists of a stochastic part and an underlying trend [29]. The trend may consist of

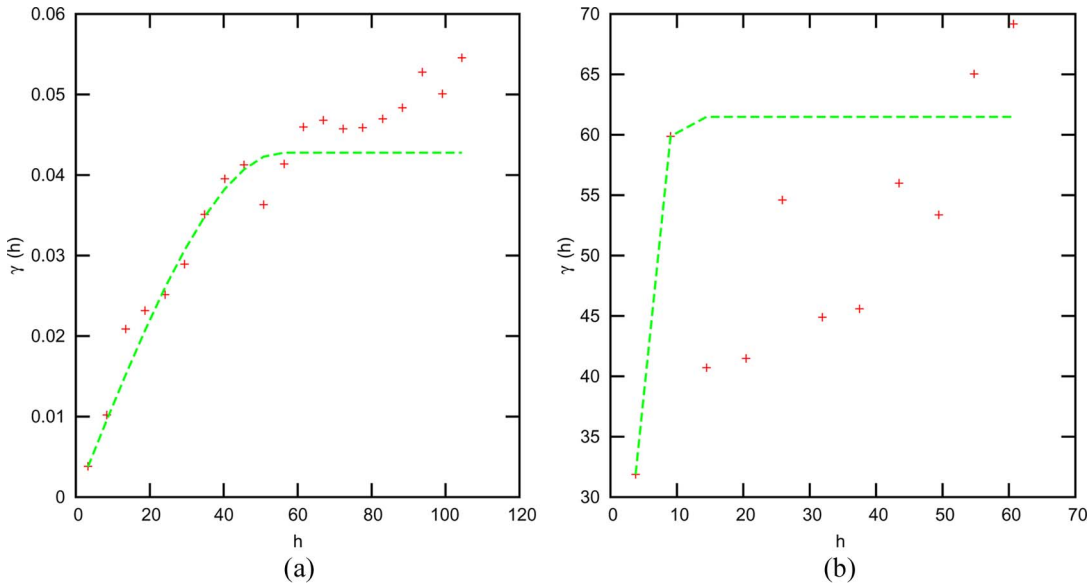

Fig. 2. Semivariograms $(+)$ and (dashed lines) automatically fitted spherical model of (8). (a) Simulated correlated data demonstrating a good fit and (b) degenerate TEC data showing a poor fit.

both local and global components. This is Matheron's "Theory of Regionalized Variables" [30]. The stochastic component is comprised of both random and autocorrelated parts, where the degree of autocorrelation is a function of distance. This means that points in close proximity are more closely correlated than distant ones.

The first step in interpolation using kriging is the formation of a semivariogram [23], [31]. This is a diagram of the spatial dependence of samples and is a function of all possible separations (or lags) and semivariance. The semivariance is defined by

$$\gamma(\mathbf{h}) = 0.5\left(f(\mathbf{x}) - f(\mathbf{x}+\mathbf{h})\right)^2 \tag{7}$$

where $f(\mathbf{x})$ contains the point values at a given location $(\mathbf{x})$ and $f(\mathbf{x}+\mathbf{h})$ is the point value at a point separated from $\mathbf{x}$ by the lag vector $\mathbf{h}$. For the isotropic 2-D case, it is simplest to calculate lags by using a distance metric between points. However, the number of combinations of sample positions which must be compared quickly becomes very large. For this reason, input coordinates are often binned to reduce the total number of lags.

After the semivariogram has been created, a model must be fitted to it. One example, the spherical semivariogram model is a curve of the form

$$\gamma_{\text{sph}} = \begin{cases} c\left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right), & \text{for } 0 \le h \le a \\ c, & \text{for } h > a \end{cases} \tag{8}$$

where, in geostatistical parlance, c is the "sill," the value that the semivariogram reaches after its initial rise, and a is the "range" or length of spatial dependence. This sill is generally close to the variance of the input values (details of this and other commonly used models can be found in [23] and [24]). Fig. 2(a) shows a typical semivariogram with a fitted spherical model.

The next, and final, step is the actual kriging process. Kriging uses a weighted average of input points to estimate the value of any given output point. The weights are found by minimizing the kriging variance—the difference between the estimate and the actual input value. An output variance is also directly calculable. As it minimizes the variance of the output, kriging is often called the "best linear unbiased estimator." However, the variance, which is minimized, is relative to the semivariogram model, so the results will only ever be as good as the model and, therefore, the semivariogram (for a good concise description of the kriging process, see [6]). It should also be noted that other authors have heavily criticized some of the underlying assumptions behind geostatistics [9], and these issues should be borne in mind when using kriging and related techniques.

III. Interpolation Using Adaptive NC

Interpolation techniques based on NC make use of the available data and confidence (or certainty) metadata to distinguish between positions where data samples are available, absent, or have a zero value. This helps improve the reconstructed output. NC techniques were originally proposed in 1993 and have been steadily increasing in popularity [32]. They have been applied to medical imaging [33], regularization of tensor fields [34], and motion compensation [35]. However, to date, there are no known applications of NC techniques to the reconstruction of geophysical data.

The most basic type of NC, known as zeroth-order NC, is defined by

$$f = \frac{f_i * g}{c_i * g} \tag{9}$$

where f_i is the input data, c_i is the input data confidence map, and g is the kernel function. These variables are functions of input data coordinates, for example, (x, y) in the 2-D case.

The term "normalized" comes from the denominator, which normalizes the output, converting it from what is essentially a filtered version of the input data to an interpolated output field. NC produces a local model of the input data using projections onto a set of basis functions. The locality comes from the kernel at each pixel, and the basis function is a polynomial whose order is generally less than two and is usually zero. When combined with kernel adaptations (described as follows), this is normally sufficient to reconstruct most high-frequency detail. However, a higher order set of basis functions could be used if desired—the downside is the increased computational cost [36]. The kernel function g can take various forms, and initial studies used a raised cosine [32]. More recently, the use of Gaussians has been popular [37], [38].

For irregularly spaced data, the kernel function must be sufficiently large to ensure that, at every output position, at least one input point is included. This can be achieved by adapting the size of the kernel at each output point. However, this provides very little in the way of extra performance when compared to simply selecting a large kernel size from the whole data set [37]. A more rewarding approach is to modify the size, orientation, and aspect ratio of the kernel, an approach known as ANC. Adaptation can be achieved by using a form of the gradient covariance matrix, the gradient square tensor (GST) [39]. However, before the GST can be constructed, the gradient of the input data must be estimated. This poses its own unique problems, since standard techniques for estimating gradient are not designed for sparse data.

There are two methods which can be used to calculate the necessary gradients without first performing an initial interpolation of the field. These are both NC based and are known as normalized differential convolution (NDC) [32] and differential of normalized convolution (DoNC) [40]. NDC involves filtering the input data with filters which have been "tilted" in various directions and, then, weighting the contributions from each using a least squares fit. DoNC is derived by differentiating (9) in orthogonal directions. NDC is less computationally expensive, although it produces results with slightly lower quality than DoNC (for more details on the mechanisms involved in these two processes, see [32] and [40]).

Once the gradients have been computed, the GST can be constructed. Taking a 2-D field as an example, the GST of a given point is a two-by-two matrix formed using inner products of gradients which have been presmoothed by a Gaussian filter, such that

$$\text{GST} = \begin{bmatrix} g_x^2 & g_{xy} \\ g_{xy} & g_y^2 \end{bmatrix} \tag{10}$$

where g_x and g_y are the gradients in the x- and y-directions, respectively, and multiple subscripts indicate the directions included in the gradient products. Computing the eigenvalues of the GST allows one to evaluate the various properties pertaining to the local area. Of these, the most important are the direction associated with the smallest eigenvalue, which corresponds to the local orientation (the direction along the local gradient)

$$\varphi_2 = \tan^{-1}\left(\frac{g_{xy}}{\lambda_2 - g_y^2}\right) \tag{11}$$

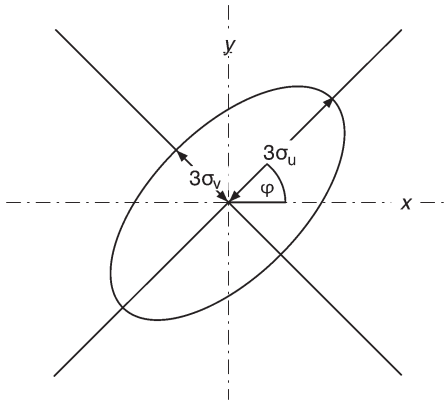

Fig. 3. Example of 2-D Gaussian filtering, as used in ANC, with major and minor axes (of size $3\sigma_u$ and $3\sigma_v$, given by (13) and (14) respectively) rotated by the angle φ, given by (11).

and the local anisotropy, which is given by

$$A = 1 - \frac{\lambda_2}{\lambda_1} \tag{12}$$

where λ_1 and λ_2 are the largest and smallest eigenvalues, respectively. The anisotropy can be used to control the size of the kernel. For the case of a 2-D Gaussian kernel, the standard deviations in the u- and v-directions, σ_u and σ_v, are given by

$$\sigma_u = C(1 - A)^\alpha \sigma_a \tag{13}$$

$$\sigma_v = C(1 + A)^\alpha \sigma_a \tag{14}$$

respectively, where σ_a is the distance to the closest input sample and c and α are constants which are generally $\approx$1. The filter is sized so that it sits within a box approximately $3\sigma_u$ wide by $3\sigma_v$ high, ensuring that the values at the edges are close enough to zero to minimize artifacts, after it has been rotated by φ_2, as determined by (11). This process is shown in Fig. 3. The final step is the filtering and normalizing process, as given by (9).

IV. Quantitative Evaluation and Experimental Results

The interpolation methods tested and examined in this paper are a mixture of commonly available implementations and custom-written code. From the triangulation-based class of techniques, the nearest neighbor, linear, and cubic interpolations available in MATLAB 2007a's `griddata` function were selected [41]. The natural-neighbor interpolation code was that available from [42]. The RBF interpolation of [19] was implemented using both linear and multiquadratic bases. No domain decomposition was used due to its complexity, and the fact that the fields being interpolated were relatively small. The biharmonic spline algorithm used was the `v4` algorithm in MATLAB's `griddata`, which uses the algorithm of [22]. The kriging method used is known as ordinary kriging and works with isotropic normally distributed data. The implementation evaluated was based on code given in [24], with some modifications. In particular, a spherical model was chosen as it represents a good tradeoff between the complexity associated

with models with a high degree of freedom and the poor performance exhibited by simpler functions such as the linear model. In tests, the spherical model was found to perform well with both simulated and TEC data. To enable the unsupervised reconstruction of TEC fields, the spherical model was automatically fitted to the semivariogram using a least squares method. The ANC interpolation technique implemented was the zeroth-order scheme of [36]. The kernel used was a 2-D Gaussian, whose size and orientation were set using (11)–(14). GSTs were constructed using gradients obtained from NDC. To reduce the complexity, the efficient decomposition technique that provides a close approximation for rotated Gaussians proposed by [43] was used in the final filtering stage. The Euclidean distance was used for all techniques requiring a distance metric.

The techniques to be evaluated were applied to both simulated and real TEC data. As simulated data provides ground-truth values, it has the advantage of allowing analysis of residual errors to be calculated at every point in the output field. In addition, parameters such as of the input sparsity can be carefully controlled. It should be noted, however, that the performance of any interpolation method can vary considerably with the statistics of the input data, and therefore, the results gained through simulation are not necessarily indicative of the general performance. Therefore, the ultimate test of reconstruction techniques remains their application to real data. To this end, the interpolation methods are applied to TEC data from the much studied October 2003 ionospheric storm.

A. Simulated-Data Results

Following the methodology detailed by Omre [31], two kinds of simulated-test data were generated. The first types were produced by generating fields of normally distributed random data, which were then filtered using a pillbox. The filtering process introduces autocorrelation with a lag distance dependent on the filter radius. These data have a multivariate Gaussian distribution, which is considered ideal for ordinary kriging. The second method generates univariate data with an approximately log-normal distribution by filtering fields of uniformly distributed random data. The multinormality is then removed by examining 5×5 neighborhoods, around each point, and randomly selecting from the ten highest values. Finally, the natural logarithm of each data point is calculated. Histograms illustrating typical distributions generated by these two methods are shown in Fig. 4.

The interpolation techniques were used to reconstruct each type of the simulated data from sparsities ranging from 95% to 99%, in steps of 1%, and 99% to 99.9%, in steps of 0.2%. The sampling was carried out by thresholding uniform pseudo-random numbers, so the percentage of remaining samples is not necessarily the same as the requested value. Plotted results show the actual sparsity obtained. Each generated data field was sampled $30\times$ at each sparsity and, then, reconstructed with each interpolation method. The number of reconstructions was set to 30 to minimize computation time, while ensuring the statistical significance of the results.

The rmse between the original and reconstructed data outputs were calculated and averaged over the 30 reconstructions (see Figs. 5 and 6). In both of these figures, the rmse values were

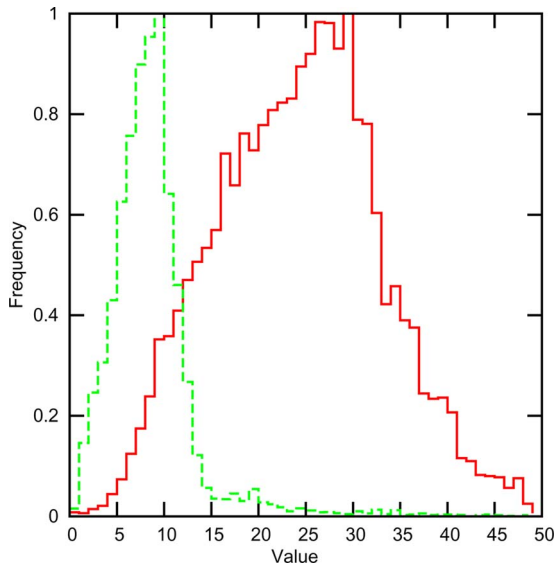

Fig. 4. Normalized histograms of the two types of simulated correlated data described in Section IV-A. (Solid) Multivariate and (dashed) univariate.

normalized by dividing by the average value of the data being interpolated to give the rmse as a proportion in which, for example, a value of 0.1 corresponds to a 10% error. For clarity, the results for the two worst performing techniques, linear and nearest neighbor interpolation, were removed.

The rmse performance of all interpolation techniques increases with sparsity for both types of simulated data. Overall, the rmse for the multivariate data increases from around 10% at a sparsity of 95% to 25%–30% at a sparsity of 99.6%. The performance for the univariate data at the corresponding sparsities is better, increasing from approximately 7% to 15% at sparsities of 95% and 99.6%, respectively.

BSI is the worst performing technique for both the univariate and multivariate simulated data, with a consistently higher rmse than the other schemes. Although kriging is the best performer for many sparsities, its error is dramatically increased at sparsities $> 99.3\%$ and at certain other lower sparsities, probably as a result of failing to correctly fit to the semivariogram. This is significant as the errors in these cases are up to four times those of the other techniques. The ANC performance at all sparsities is 1%–2% worse than the best performing techniques. Cubic interpolation generally performs well, and the overall best performer is natural-neighbor interpolation, which exhibits no anomalous behavior while maintaining good performance throughout.

This process described earlier was repeated while altering the size of the pillbox used to impose autocorrelation of the simulated data from 10 to 50 in steps of ten. Overall, this has little effect on the rmse of the reconstructions, with the exception of kriging, whose implicit assumptions about data autocorrelation are violated when the lag distance is small. In both the univariate and multivariate cases, natural-neighbor interpolation performed best with respect to changing radius of correlation. This is because its performance is based on data position rather than value.

In addition to providing overall error values, simulated data allow for analysis of residual errors at every point in the field.

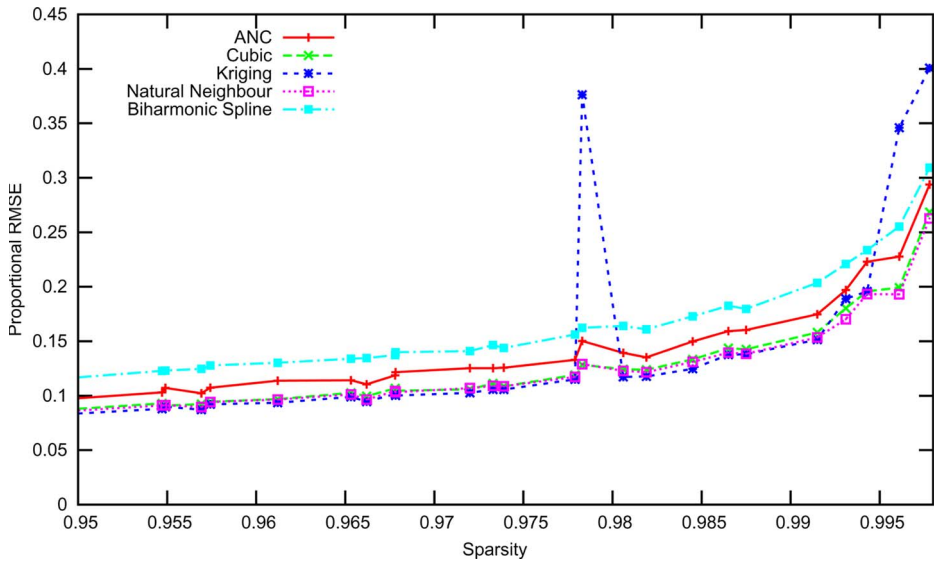

Fig. 5. Proportional rmse as a function of sparsity for simulated multivariate correlated data reconstruction.

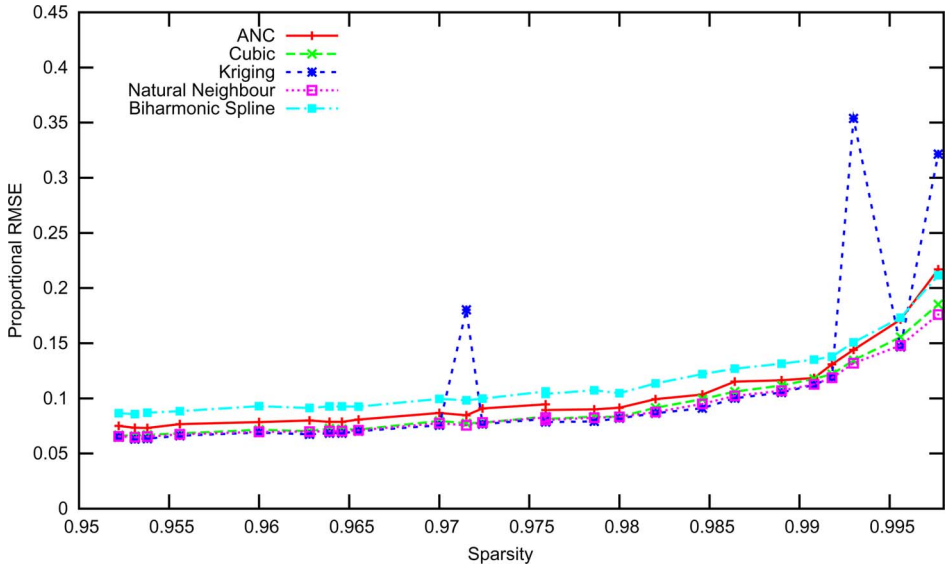

Fig. 6. Proportional rmse as a function of sparsity for simulated univariate correlated data reconstruction.

In all cases, the residual errors exhibited Gaussian distributions with means very close to zero, showing that the interpolation techniques have negligible bias.

B. TEC Data Results

The TEC is defined as the line integral of the electron content over a path between two points, usually a satellite and a receiver. Various methods have been developed for extracting TEC information from the amplitude and phase of GPS signals, e.g., [44] and [45]. The data used in this paper were processed and calibrated using the MIDAS tomographic inversion software from the University of Bath [1]. MIDAS calculates the TEC biases by analyzing the differences in length between the measured and inverted paths [46]. While other methods for removing biases are available [47], [48], the aim here was to provide representative TEC data for the evaluation of the interpolation techniques.

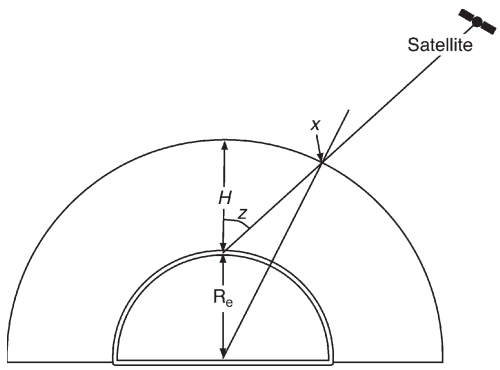

Fig. 7. Thin-shell ionosphere model. R_e is the radius of the Earth, z is the elevation angle from the ground station to the satellite, x is the point at which the path between the satellite and ground station intersects the shell, and H is the height of the shell.

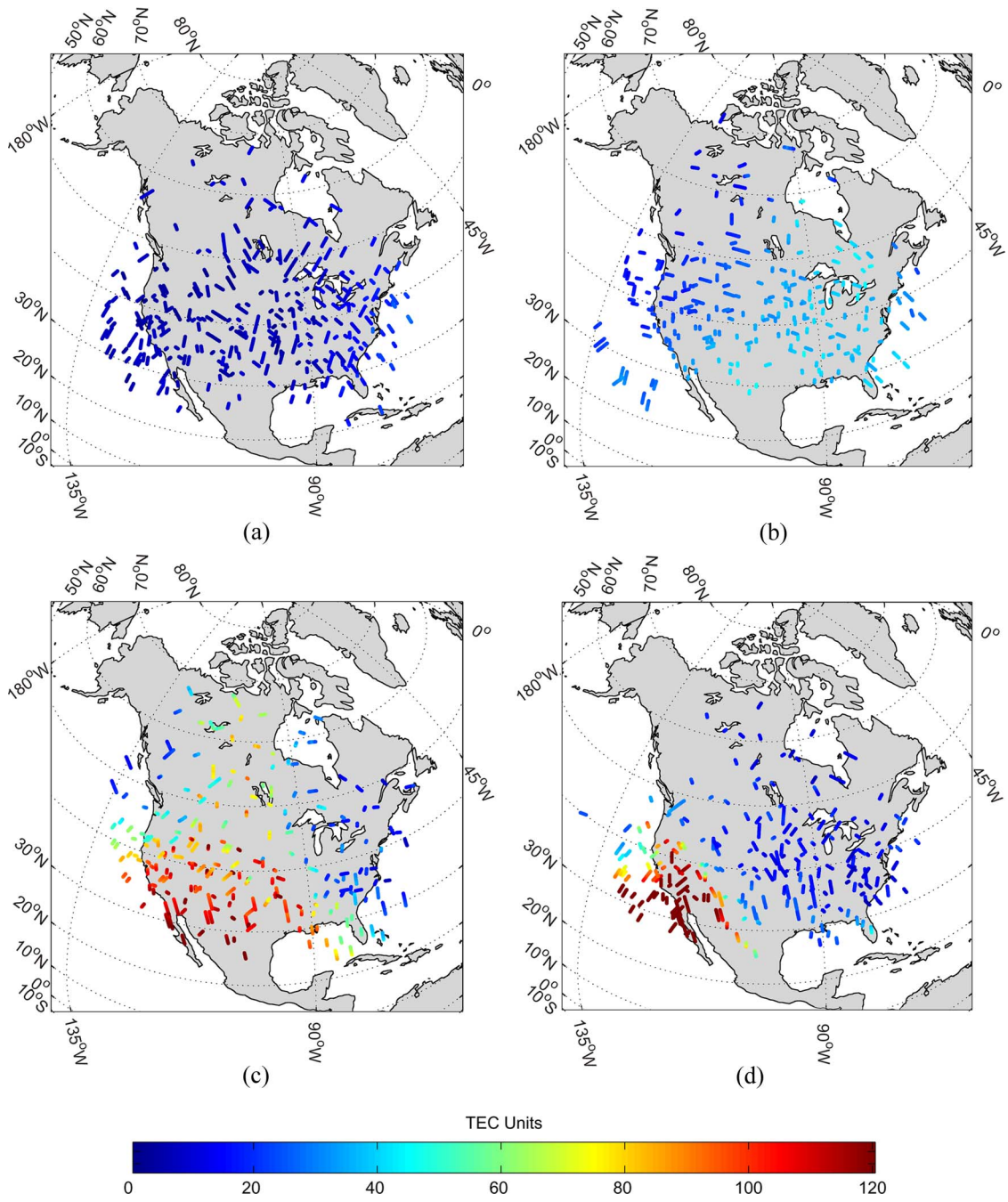

Fig. 8. Example TEC input data for the cross-validation test in Section IV-B. All data are from October 30, 2003 for the time periods (a) 1200–1215 UT, (b) 1600-1615 UT, (c) 2000–2015 UT, and (d) 0000–0015 UT in October 31.

The sources of data used to test TEC reconstruction are approximately 80 GPS measuring stations lying within 20–70° N, and 70–130° W. This corresponds to a coverage of most of North America. While more sites were available at that time, not all sites were used, as the main aim of this paper is to examine interpolation during high-sparsity cases. The time period over which data were drawn was from noon to midnight on October 30, 2003—the peak of the much studied "Halloween Storm" [49]. Data were considered stationary within 15-min intervals and projected onto a "thin shell" for reconstruction [47]. The thin shell used covered the same area as the ground stations and had latitudinal and longitudinal resolutions of 0.5°, giving rise to fields of size 101 × 121 pixels. As each ground-based receiver station can see approximately six satellites at any one time, there are around 500 paths associated with the 80 measuring stations. Each path's TEC values were projected onto an infinitesimal shell at a fixed height by calculating the ray's intersection with the shell. With reference to Fig. 7, the function which maps from slant to vertical TEC is then given by

$$F(z) = \left(1 - \frac{R_e \cos(90 - z)}{R_e + H}\right)^{-0.5} \tag{15}$$

where z is the satellite elevation angle (in degrees), H is the height of the shell (400 km in this case), and R_e is the radius

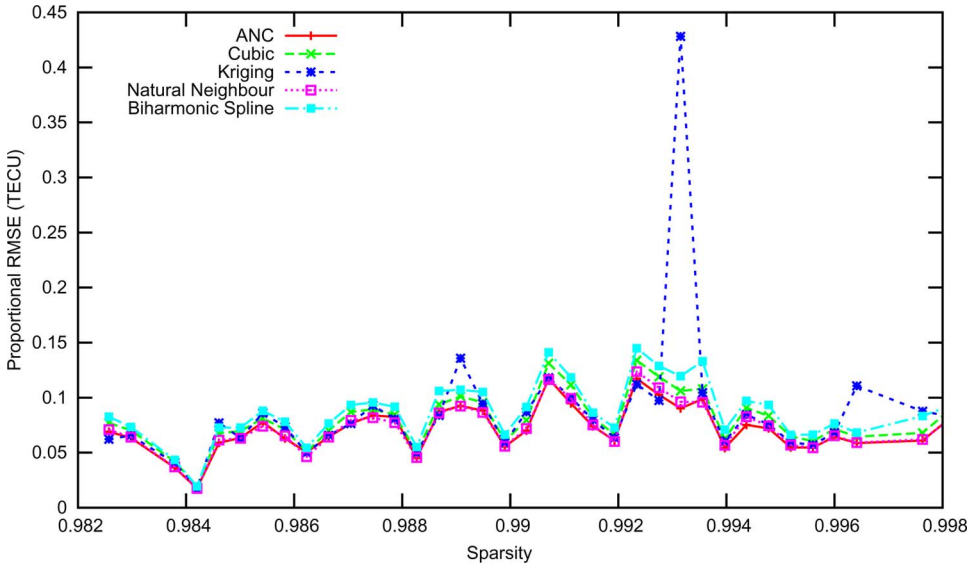

Fig. 9. Proportional rmse as a function of sparsity for TEC data from GPS path measurements. Errors were calculated using the cross-validation method described in the text (Section IV-B).

of the Earth. Paths with elevation angles $< 20^\circ$ were removed because of high levels of error associated with low angles [47]. The projected TEC values at three hourly intervals over the 12-h storm period are shown in Fig. 8.

As there is no ground-truth data against which interpolated TEC fields can be evaluated, accurate testing is more difficult. Previous approaches have used models as a basis for comparison (e.g., [50], used the CCIR model to compare f_0F_2 results[1]). However, in many cases, modeled data are far smoother than the actual phenomena being modeled, which leads to anomalous results. In particular, results tend to be biased toward favoring techniques which produce artificially smooth outputs. Alternatively, partitioning the data set into two classes, the testing and reconstruction data, allows the testing of output against real data which were not used in the reconstruction. This technique is known as cross-validation and is often used for testing the accuracy of classification systems where no ground-truth data are available [51].

The specific method used was k-fold cross-validation, in which the list of all ray paths is randomized and then partitioned into several blocks. Values corresponding to the first block are interpolated using the schemes under test, and the values which belong to the unused blocks are used to compute differences with the output value at corresponding positions. These can then be used to calculate the rmse and other difference metrics. Both the first and second blocks are then used for the reconstruction, and the errors for ray paths in the remaining blocks are found. The process is repeated until only one block remains. This approach has the added advantage of being able to produce input data fields with varying degrees of sparsity. To ensure the significance of the results, validation results should only be used where a reasonable number of validation positions are available. In this paper, only cases with upward of 30 positions were used. Care was also taken to ensure that the average TEC values of the points used for validation were similar in magnitude to those being used for the reconstructions, to avoid biasing the output values.

Fig. 9 shows the electron content results binned into 40 sections and averaged across each section for the interpolation techniques evaluated in the previous section. As before, the rmse have been divided by the average field value to produce proportional rmse results that are more directly comparable with the simulated results. The lower sparsity limit of 0.9825% is higher than for the simulated data, as this is the greatest density that can be achieved with the available data points. Compared to the simulated data, the increase in error with sparsity is less marked for all techniques.

Once again, BSI was the overall worst performing technique. The inconsistency of the errors across the range of sparsities associated with kriging that was exhibited in the simulated results is also evident. For example, at a sparsity of 99.3%, its rmse is over four times that of the best performing technique. Cubic interpolation is performed consistently with approximately average results.

Natural-neighbor interpolation again performed well but, unlike the results for simulated data, its performance is matched by ANC. In fact, over the range of sparsities, the proportional rmse produced by ANC is, on average, 0.08% less than the equivalent natural-neighbor results and also showed a slightly lower variance. In comparison, the average rmse produced by cubic and kriging interpolation were 0.91% and 1.57% worse than ANC, respectively. The inconsistency of kriging is reflected by a variance approximately ten times higher than the other techniques.

To illustrate the results produced by some of the different interpolation techniques, Fig. 10 shows example reconstructions produced by ANC and kriging for two sets of input data from Fig. 8(c). The input data in each case was 25% of the available GPS path signals, giving a sparsity of approximately 98.8%. For both sets of input data, kriging and ANC have produced slightly

[1]The CCIR model is now part of the International Reference Ionosphere (IRI) model.

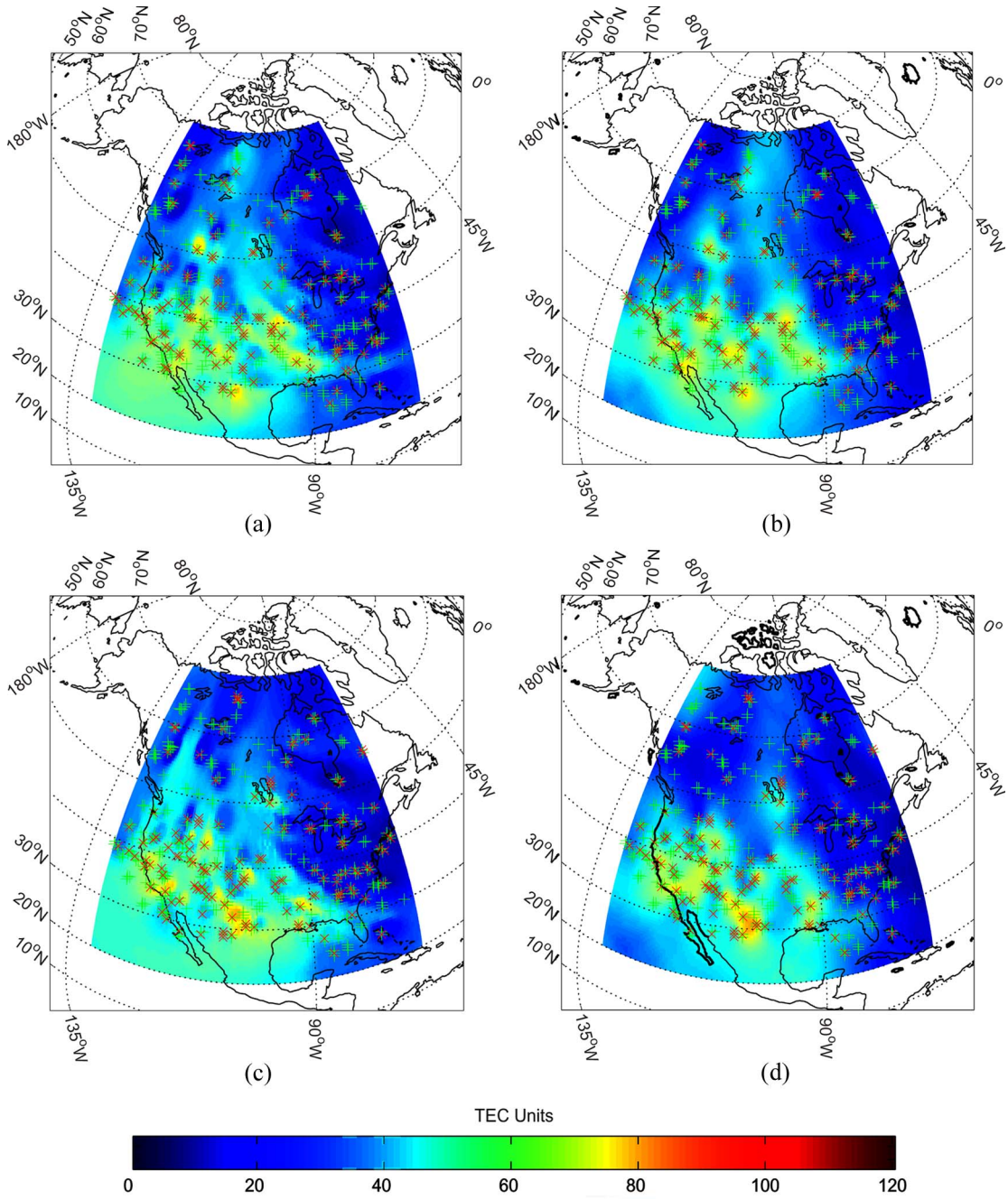

Fig. 10. Example output images produced by ANC and kriging using two sets of the input data from Fig. 8(c). Each set consisted of 25% of the available data $(\times)$ and the remaining data $(+)$ used to calculated the proportional rmse. (a) and (b) Set 1 results (sparsity of 98.94%) produced by ANC and kriging, respectively. (c) and (d) Set 2 results (sparsity of 98.79%) produced by ANC and kriging, respectively. The proportional rmse values are (a) 0.0640, (b) 0.0533, (c) 0.0503, and (d) 0.0421 TEC units.

different results. The proportional rmse for each reconstruction can be found using the remaining 75% of the data. For this case, the average proportional rmse is 0.0524 TEC units, and the difference between the kriging and ANC errors is less than 1%. Fig. 10 also shows that the set of input data used produces more significant differences in the output fields than the choice of interpolation method. This observation underlies the benefit of the cross-validation evaluation procedure, which removes any sensitivity to choice of input data by averaging the results within a given range of sparsities.

V. Discussion and Conclusion

In the literature, kriging has been the interpolation method of choice for producing TEC maps of the ionosphere. However, to date, there has been little in the way of evidence to support its adoption over other interpolation schemes. This paper has sought to address this issue by performing a comprehensive quantitative evaluation of kriging and a selection of other interpolation methods currently in use. To this end, an evaluation methodology that uses both simulated and TEC data has been proposed. With simulated data, error values can

be calculated at all output points. For TEC reconstructions, this is not possible and, instead, an evaluation using cross-validation was performed. Considering the overall performance for both simulated and TEC data, the following remarks about the individual interpolation techniques can be made.

Triangulation-based techniques are widely used in computer graphics applications. The best performing of these, cubic interpolation, has a relatively low complexity, and its performance is, in many cases, reasonably close to that of the best performing more complex methods. Therefore, for ionospheric applications where a small loss in accuracy can be sacrificed for a faster runtime, it is a reasonable choice of technique.

Although the kriging scheme used in this investigation performs well at many sparsities, it exhibits a very large variance for both the simulated and TEC data. This variance is due to spikes where the proportional rmse is excessively high. Two main stages of kriging-based interpolation are the construction of the semivariogram and the fitting of a suitable model, and both of these are sensitive to the settings of their various parameters. This is one of the main reasons why it is often recommended that kriging is implemented as an interactive process, as opposed to an automatic one. When the semivarogram model being used fails to accurately fit the experimental semivarigram, the output of the interpolation will be poor. Fig. 2(b) shows an example, degenerate semivariogram where the data have a high variance at all lags, a breach of the fundamental assumption of high autocorrelation at low lags. In these cases, the fitted model poorly matches the actual data, resulting in an output field with a high error. Although, in theory, it may be possible to automatically detect degenerate semivariograms and try to find a more suitable one to fit to, the procedure is very complex, and for the tests performed here, the level of sophistication required would be far beyond that required by the other techniques being evaluated.

Natural-neighbor interpolation performs well across all data types and sparsities. It is the best performing method for both types of simulated data and is only surpassed by ANC on the TEC data. The main drawback of natural-neighbor interpolation is that it is complicated to implement and there are few modern reference implementations available. However, if it were more widely known and its performance recognized, this situation could change.

While there are many performance features that are common for both the simulated and TEC reconstructions, there are also some significant differences, such as the change in rmse with sparsity and the relative performance of individual interpolation techniques. This suggests that, in testing interpolation methods, simulations should only be used if they are demonstrably very similar to the real data to be interpolated. If this is not the case, the data-driven cross-validation methodology demonstrated here is ideal for testing the performance of interpolation schemes using only real data. The major difference between the simulated and TEC data is that of anisotropy, and this helps explain why the relative performance of ANC, a technique that copes well with anisotropy, was dramatically improved for the TEC reconstruction. Indeed, given that ANC was the best performing technique for the TEC data, these results suggest that ANC and natural-neighbor interpolation should be the methods of choice for ionospheric reconstructions, as they offer an error performance that is better and more consistent than kriging.

REFERENCES

[1] C. N. Mitchell and P. S. J. Spencer, "A three-dimensional time-dependent algorithm for ionospheric imaging using GPS," *Ann. Geophys.*, vol. 46, no. 4, pp. 687–696, 2003.

[2] J. M. Pallares, G. Ruffini, and L. Ruffini, "Ionospheric tomography using GNSS reflections," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 321–326, Feb. 2005.

[3] E. Brockmann and W. Gurtner, "Combination of GPS solutions for densification of the European network: Concepts and results derived from 5 European associated analysis centers of the IGS," in *Proc. EUREF Workshop*, Ankara, Turkey, May 1996.

[4] M. Liao, T. Wang, L. Lu, W. Zhouzhou, and D. Li, "Reconstruction of DEMS from ERS-1/2 TANDEM data in mountainous area facilitated by SRTM data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 7, pp. 2325–2335, Jul. 2007.

[5] M. Gianinetto and P. Villa, "Rapid response flood assessment using minimum noise fraction and composed spline interpolation," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3204–3211, Oct. 2007.

[6] J. Blanch, T. Walter, and P. Enge, "Application of spatial statistics to ionosphere estimation for WAAS," in *Proc. ION NTM*, 2002, pp. 719–724.

[7] I. Stanislawska, G. Juchnikowski, L. R. Cander, L. Ciraolo, P. A. Bradley, Z. Zbyszynski, and A. Swiatek, "The kriging method of TEC instantaneous mapping," *Adv. Space Res.*, vol. 29, no. 6, pp. 945–948, Mar. 2002.

[8] P. Wielgosz, D. A. Grejner-Brzezinska, and I. Kashani, "Regional ionosphere mapping with kriging and multiquadric methods," *J. Glob. Position. Syst.*, vol. 2, no. 1, pp. 48–55, 2003.

[9] G. Philip and D. Watson, "Matheronian geostatistics—Quo vadis?" *Math. Geol.*, vol. 18, no. 1, pp. 93–117, 1986.

[10] R. Franke, "Scattered data interpolation: Tests of some methods," *Math. Comput.*, vol. 38, no. 157, pp. 181–200, Jan. 1982.

[11] M. H. Mahdian, E. Hosseini, and M. Matin, "Investigation of spatial interpolation methods to determine the minimum error of estimation: Case study, temperature and evaporation," in *Proc. 6th Int. Conf. GeoComputation*, Sep. 24–26, 2001. CD-ROM.

[12] M. Rauth, "Gridding of geophysical potential fields from noisy scattered data," Ph.D. dissertation, Univ. Vienna, Vienna, Austria, May 1998.

[13] E. Meijering, "A chronology of interpolation: From ancient astronomy to modern signal and image processing," *Proc. IEEE*, vol. 90, no. 3, pp. 319–342, Mar. 2002.

[14] M. Gasca and T. Sauer, "On the history of multivariate polynomial interpolation," *J. Comput. Appl. Math.*, vol. 122, no. 1/2, pp. 23–35, Oct. 2000.

[15] K. Sugihara, A. Okabe, and B. Boots, "Spatial tessellations: Concepts and applications of Voronoi diagrams," in *Probability and Statistics*. Hoboken, NJ: Wiley, 2000.

[16] D. F. Watson and G. M. Philip, "Triangle based interpolation," *Math. Geol.*, vol. 16, no. 8, pp. 779–795, Nov. 1984.

[17] D. F. Watson, *Contouring: A Guide to the Analysis and Display of Spatial Data*. New York: Pergamon, 1992.

[18] R. Sibson, "A brief description of natural neighbour interpolation," in *Interpreting Multivariate Data*. Hoboken, NJ: Wiley, 1981, pp. 21–36.

[19] J. Carr, W. Fright, and R. Beatson, "Surface interpolation with radial basis functions for medical imaging," *IEEE Trans. Med. Imag.*, vol. 16, no. 1, pp. 96–107, Feb. 1997.

[20] W. A. Light, "Some aspects of radial basis function approximation," in *Approximation Theory, Spline Functions and Applications*, Norwell, MA: Kluwer, 1992, pp. 163–190.

[21] M. Powell, "The theory of radial basis function approximation in 1990: Advances in numerical analysis," in *Proc. 4th Summer Sch.*, Lancaster, U.K., 1990, vol. 2, pp. 105–210.

[22] D. T. Sandwell, "Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data," *Geophys. Res. Lett.*, vol. 14, no. 2, pp. 139–142, Feb. 1987.

[23] N. A. C. Cressie, *Statistics for Spatial Data*. Hoboken, NJ: Wiley, 1991.

[24] M. H. Trauth, *MATLAB Recipes for Earth Sciences*. New York: Springer-Verlag, 2006.

[25] D. F. Watson, "Natural neighbour sorting," *Aust. Comput. J.*, vol. 17, no. 4, pp. 189–193, Nov. 1985.

[26] R. K. Beatson, W. A. Light, and S. Billings, "Fast solution of the radial basis function interpolation equations: Domain decomposition methods," *SIAM J. Sci. Comput.*, vol. 22, no. 5, pp. 1717–1740, 2001.

[27] A. Boucher, K. Seto, and A. Journel, "A novel method for mapping land cover changes: Incorporating time and space with geostatistics," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3427–3435, Nov. 2006.
[28] N. Cressie, "The origins of kriging," *Math. Geol.*, vol. 22, no. 3, pp. 239–252, 1990.
[29] G. Matheron, "The intrinsic random functions and their applications," *Adv. Appl. Probab.*, vol. 5, no. 3, pp. 439–468, Dec. 1973.
[30] G. Matheron, *The Theory of Regionalized Variables, and Its Applications*. Paris, France: Centre de Geostatistique, 1971.
[31] H. Omre, "The variogram and its estimation," in *Geostatistics for Natural Resources Characterization*. Amsterdam, The Netherlands: Reidel, 1984, pt. 1, pp. 107–125.
[32] H. Knutsson and C.-F. Westin, "Normalized and differential convolution: Methods for interpolation and filtering of incomplete and uncertain data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, New York, Jun. 1993, pp. 515–523.
[33] R. S. J. Estepar, M. Martin-Fernandez, C. Alberola-Lopez, J. Ellsmere, R. Kikinis, and C.-F. Westin, "Freehand ultrasound reconstruction based on roi prior modeling and normalized convolution," in *Proc. MICCAI*, 2003, vol. 2879, pp. 382–390.
[34] C.-F. Westin and H. Knutsson, "Tensor field regularization using normalized convolution," in *Proc. 9th Int. Conf. EUROCAST*, Feb. 2003, vol. 2809, pp. 564–572.
[35] G. Farneback, "Polynomial expansion for orientation and motion estimation," Ph.D. dissertation, Linköping Univ., Linköping, Sweden.
[36] T. Pham, L. van Vliet, and K. Schutte, "Robust fusion of irregularly sampled data using adaptive normalized convolution," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 10, pp. 1–12, 2006.
[37] T. Q. Pham and L. J. van Vliet, "Normalized averaging using adaptive applicability functions with applications in image reconstruction from sparsely and randomly sampled data," in *Proc. 13th Scand. Conf. Image Anal.*, 2003, vol. 2749, pp. 485–492.
[38] T. Pham, "Spatiotonal adaptivity in super-resolution of undersampled image sequences," Ph.D. dissertation, Quantitative Imag. Group, Delft Univ. Technol., Delft, The Netherlands, 2006.
[39] M. van Ginkel, J. van de Weijer, L. J. van Vliet, and P. W. Verbeek, "Curvature estimation from orientation fields," in *Proc. 11th Scand. Conf. Image Anal.*, 1999, pp. 545–551.
[40] R. Piroddi and M. Petrou, "Analysis of irregularly sampled data: A review," in *Advances in Imaging and Electron Physics*, vol. 132, New York: Elsevier, 2004, pp. 109–165.
[41] The Mathworks Inc., *Griddata*. [Online]. Available: http://www.mathworks.com/access/helpdesk/help/techdoc/ref/griddata.html
[42] P. Sakov, *Natural Neighbour Interpolation Software*. [Online]. Available: http://www.marine.csiro.au/~sakov
[43] J. Geusebroek, A. W. M. Smeulders, and J. van de Weijer, "Fast anisotropic Gauss filtering," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 938–943, Aug. 2002.
[44] R. Warnant and E. Pottiaux, "The increase of the ionospheric activity as measured by GPS," *Earth Planets Space*, vol. 52, no. 11, pp. 1055–1060, 2000.
[45] F. Arikan, O. Arikan, and C. B. Erol, "Regularized estimation of TEC from GPS data for certain midlatitude stations and comparison with the IRI model," *Adv. Space Res.*, vol. 39, no. 5, pp. 867–874, 2007.
[46] R. Meggs and C. Mitchell, "A study into the errors in vertical total electron content mapping using GPS data," *Radio Sci.*, vol. 41, no. 1, p. RS1 008, Feb. 2006.
[47] A. Mannucci, B. Iijima, U. Lindqwister, X. Pi, L. Sparks, and B. Wilson, "GPS and ionosphere," in *Review of Radio Science 1996–1999*. London, U.K.: Oxford Univ. Press, 1999.
[48] "A.I.U.B. CODE Center for Orbit Determination in Europe," *Differential GPS Code Biases (DCBs)*. [Online]. Available: http://cmslive2.unibe.ch/unibe/philnat/aiub/content/research/gnss/code__research/index_eng.html
[49] M. Hernandez-Pajares, J. M. J. Zornoza, J. S. Subirana, R. Farnworth, and S. Soley, "EGNOS test bed ionospheric corrections under the October and November 2003 storms," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2283–2293, Oct. 2005.
[50] T. Samardjiev, P. A. Bradley, L. R. Cander, and M. I. Dick, "Ionospheric mapping by computer contouring techniques," *Electron. Lett.*, vol. 29, no. 20, pp. 1794–1795, Sep. 1993.
[51] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. 14th Int. Joint Conf. Artif. Intell.*, 1995, vol. 2, pp. 1137–1143.

Matthew P. Foster (S'06) received the M.Eng. degree in computers electronics and communications from the University of Bath, Bath, U.K., in 2005. He is currently working toward the Ph.D. degree on reconstruction and motion estimation using sparse remotely sensed data in the Centre for Space, Atmospheric and Oceanic Science, Department of Electronic and Electrical Engineering, University of Bath.

Adrian N. Evans (M'02) received the B.Eng. degree from the Loughborough University of Technology, Leicestershire, U.K., in 1990 and the Ph.D. degree in medical image processing from the University of Southampton, Southampton, U.K., in 1994.

From 1993 to 1994, he was a Research Associate with the University of York, York, U.K. From 1994 to 1997, he was a Lecturer of information engineering with Massey University, Palmerston North, New Zealand. Since September 1997, he has been a Lecturer with the Department of Electronic and Electrical Engineering, University of Bath, Bath, U.K. He has published over 50 papers in journals and refereed conference proceedings. His current research interests include nonlinear image processing, feature extraction and motion estimation in application to remotely sensed, color and multispectral images, and video compression.