Modeling Locally Varying Anisotropy of CO2 Emissions in the United STATES

J.B. Boisvert

In a property modeling context, the variables of interest to be modeled often display complex nonlinear features. Techniques to incorporate these nonlinear features, such as multiple point statistics or cummulants, are often complex with input parameters that are difficult to infer. The methodology proposed in this paper uses a classical vector based definition of locally varying anisotropy to characterize nonlinear features and incorporate locally varying anisotropy into numerical property models. The required input is an exhaustive field of anisotropy orientation and magnitude. The methodology consists of (1) using the shortest path distance between locations to define the covariance between points in space (2) multidimensional scaling of the domain to ensure positive definite kriging equations and (3) estimation or simulation with kriging or sequential Gaussian simulation. The only additional parameter required when kriging or simulating with locally varying anisotropy is the number of dimensions to retain in multidimensional scaling. The methodology is demonstrated on a CO_2 emissions data set for the United States in 2002.

1. Introduction

Incorporating anisotropy into numerical models enhances the predictive power of the resulting models. Consider a simple pollution example, adding a slight northerly wind (Figure 1 left), the pollution is more continuous in a single direction (constant anisotropy). If we are given the knowledge that the pollution source is located in a mountainous valley (Figure 1 right), the wind is more erratic and the pollution spread is complex, thus, the direction of anisotropy varies and the pollution is said to be continuous in locally varying directions (locally varying anisotropy). The proposed methodology can be used to model any continuous or discrete variable that displays known locally varying anisotropy (LVA).

2. Past Work

Some methodologies exist for incorporating LVA because variables such as pollution spread, rain fall patterns and animal migration can be extremely nonlinear. Stream distances have been used to model pollution or fish migration in streams (Cressie and Majure 1997; Little, Edward and Porter 1997; Rathburn 1998; Loland and Host 2003; ver Hoef et al. 2006) where small 1D grids orientated along the stream are sufficient. Generating 3D models containing millions of cells with these methods would be infeasible. Moreover, the methodology presented in this paper is used to incorporate any form of LVA rather than being limited to physical boundaries and barriers (i.e. rivers) as is common in the literature.

Incorporating the stream distance into geostatistical modeling is not straightforward as there is no guarantee of positive definiteness of the normal equations when nonEuclidian distances are used; this issue has been addressed by Boisvert and Deutsch (2010) where MDS is used to eliminate indefinite matrices. MDS is applied in this paper in a way similar to Boisvert and Deutsch (2010) rather than Sampson and Guttorp (1992) who also use MDS to incorporate LVA but are more concerned with multiple measurements of the same variable at monitoring stations.

3. Methodology

At the heart of geostatistics is the spatial prediction of variables from sparse sample data. From these spatial estimates of concentrations, grades or porosities, volumetric calculations can be made and site classifications determined. When the spatial phenomenon of interest displays LVA, better spatial predictions can be made by incorporating LVA into the numerical modeling. In the proposed methodology, LVA is incorporated by using the optimized shortest path distance (SPD) between locations. The SPD accounts for the LVA, thus the resulting models display the desired nonlinear features. Kriging and sequential Gaussian simulation are tools often applied in geostatistical studies as an alternative to inverse distance weighting and are modified to incorporate LVA.

Multidimensional scaling (MDS) techniques are implemented to guarantee the positive definiteness of this system of equations.

The methodology, given in detail in Boisvert and Deutsch (2010), is applied to the 2002 total CO_2 output of the United States (Figure 2). Much of the LVA in this data set is due to high population centers and major traffic ways. The data is available on a 10km square grid. The exhaustive grid is resampled at a 100km resolution to provide the test data. Kriging with LVA is performed and compared to the known truth at a 10km resolution as well as to a traditional, second order stationary kriging. Kriging with LVA (Boisvert and Deutsch, 2010) consists of three main steps which are expanded upon below:

Step 1: Generate the LVA field

Step 2: Calculate the SPD

Step 3: Krige or Simulate with LVA

3.1 Step 1: Generate the LVA field

The LVA field is a vector field describing the orientation and magnitude of the anisotropy for each cell of the model. In GSLIB (Deutsch and Journel, 1998) convention, the LVA field is described by specifying the major direction of continuity with a strike angle (clockwise positive from North) and an anisotropy ratio (minor range / maximum range). In three dimensions three angles are required (strike, dip and plunge) and two ratios.

LVA field generation is an integral step of the methodology. A number of data sources could be used to infer the LVA field, including hard/soft data, seismic, knowledge of the underling physical laws acting on the variable of interest, direct measurement of the anisotropy parameters at discreet locations or expert interpretation. The technique used in this case study is based on the moment of inertia technique (Hassanpour, 2007). A covariance map (related to a variogram volume map) is created for a local region of the domain. The moment of inertia is used to determine the major direction of anisotropy, i.e. the direction in which the moment of inertia for the covariance (treated as mass) is minimum. This results in the LVA directions shown on Figure 2 right. In general the LVA field follows the major traffic ways as desired. In areas where there is no anisotropy (all high values or all low values) the orientation of the LVA field is erratic, but the anisotropy ratio is ~1.0 (isotropic), thus the LVA has little effect in these areas.

A second LVA field is generated by hand that follows only the major roadways on the map and assumes the CO_2 distribution is preferential along the directions of the roadways. This LVA field was created by manually digitizing the directions of the roads and kriging the sin/cos of the resulting vectors. Recombining the components of the vectors yields the exhaustive LVA field (Figure 2).

3.2 Step 2: Calculate the SPD

With the definition of the LVA field (step 1) the distance between any two locations in space can be determined by following along the directions of anisotropy and defining the minimum anisotropic path between locations. The distance between points is the path that results in the shortest anisotropic distance. For example, if the anisotropy is assumed to follow along the roadways because CO_2 emissions are higher in traffic, the SPD between points A and B (Figure 2) is not linear. In the case of second order stationarity, the shortest path is the straightline path; however, when LVA is introduced, nonlinear paths that follow the major directions of anisotropy are shorter. Using this distance in kriging or simulation, rather than the straightline path, is the mechanism through which the desired LVA is incorporated into modeling.

The SPD has been calculated by redefining the model space as a graph with edges defined between adjacent model cells (Boisvert and Deutsch, 2010). The length of each linear edge is calculated using the local anisotropy of the cell containing the edge. With the length of each edge defined, the SPD is calculated using the Dijkstra algorithm (Dijkstra 1959); the Dijkstra algorithm requires the user to supply the number of offsets to be used to define the graph. The offset is the number of adjacent cells that are connected with edges. If adjacent edges are connected only (offset=1) paths are restricted to following 45˚ increments. With certain directions of anisotropy artifacts can occur (Figure 3) as the shortest path is not found due to the restriction that the paths must

follow 45˚ increments. The solution to this is to connect model cells to cells beyond the adjacent cells, creating paths that follow at more arbitrary angles. For example, offset=2 connects cells separated by one cell. The limitation to increasing the number of offsets to an arbitrarily high value to allow for flexible paths is the associated increase in RAM required to store the graph in memory (Figure 4). Note that the Boost Graph library implementation (Siek and Lumsdaine 2001) of the Dijkstra algorithm was used.

We propose to use fast marching methods (FMM) as an alternative to the RAM intensive Dijkstra algorithm. FMM is similar in principle to the Dijkstra algorithm without the restriction to a graph. Consider the propagation of a wave front starting at point A (Figure 5) the wave propagates at a velocity (*F*) proportional to the LVA. As this front reaches other locations in the grid, the SPD distance is determined from the time required for the front to reach this location. The arrival times can be determined from the viscosity solution to the Eikonal equation:

$$\left|\nabla T\right| F = 1$$

where *F* is the speed perpendicular to the advancing wave front and *T* is the arrival time. There are a number of implementation details that are important; the interested reader is referred to Sethian (1999) for more detail. Essentially, the front is propagated through the model, distances/times to locations inside the front are fixed (black nodes in Figure 5) and arrival times to locations adjacent to the front are updated within the narrow band indicated in Figure 5.

The public implementation of FMM with LVA from Peyre (2008) was used to generate the distances to be used in the proposed methodology (Figure 6). Note that the distances calculated with both algorithms are quite similar; however, the memory requirements of the FMM algorithm are significantly lower (Figure 4) which is critical when applying this methodology to large 3D grids.

The FMM algorithm is similar to the Dijkstra algorithm with one offset but with significantly lower RAM requirements. The accuracy of the FMM is controlled by the grid size; increasing the density of cells or reducing the grid size results in more accurate paths. In a typical geostatistical analysis the model cell size is smaller than the desired LVA scale so that the nonlinear features of interest can be resolved at the selected cell resolution. It has been the authors' experience that considering grids smaller than the point scale definition with FMM does not produce noticeably better paths (Figure 3).

3.3 Step 3: Krige or simulate with LVA

Kriging or sequential Gaussian simulation incorporating LVA is identical to traditional geostatistics with the SPD substituted for the straightline distance and an additional step to ensure positive definiteness of the kriging equations. The use of the SPDs in kriging does not generate systems of equations that are guaranteed positive definite (Curriero 2005; Boisvert and Deutsch 2010). To obtain a positive definite system of equations, the entire grid is embedded in a *k*-dimensional Euclidean space with multidimensional scaling; specifically, ISOMAP-L (Tenenbaum et al. 2000) is used to perform the embedding with a regular grid of landmark points. The SPD is required between the landmark points and all cells in the model. In this case, *k*=225 evenly spaced landmark points were used in ISOMAP-L. After performing the embedding, each grid cell is located in a 224 dimensional (*k*-1) Euclidian space. Because the space is Euclidian, positive definiteness is guaranteed by using a variogram structure that is known to be positive definite in *k*-1 dimensions; the exponential variogram is recommended (Figure 8). The covariance used in the normal equations is calculated using a variogram model that is fit to the experimental variogram calculated from the available data after the application of ISOMAP-L (Figure 8). This variogram is isotropic in as the anisotropy has been considered with the SPD calculation and the MDS transformation. The experimental variogram is often well behaved because it is isotropic.

4. Discussion

Local prediction is improved when LVA is considered, measured by an increase of 9% in the covariance between the truth and estimates (Figure 8) in a cross validation assessment. While increasing the covariance in cross validation is a beneficial result, the true benefit of the methodology is the reproduction of important nonlinear features; in this example, local features due to traffic (highlights on Figure 7) are seen in the resulting estimate

map when LVA is considered, these features are not apparent using traditional kriging. For convenience, comparisons are made in normal score units; gains in covariance and local features are similar in original units.

CO_2 emission data from Gurney (2010) is available on an hourly basis for 2002. We are currently exploring techniques to extend the proposed methodology to the time domain as well. The LVA field changes throughout the year due to wind patterns, traffic concentrations, etc. Incorporating a changing LVA field to account for the time dimension would further increase the predictive nature and flexibility of the proposed methodology. Moreover, generation of the LVA field is an issue when time is considered as manual methods begin to become impractical when every time step must be analyzed as well as the added difficulty of maintaining consistent LVA fields across each time step.

The proposed methodology is only as effective as the LVA field used. LVA fields that are not representative of the spatial structure of the variable of interest can generate poor models. Care is required when determining an appropriate LVA field, some methodologies for LVA field generation can be found in Boisvert (2010) and Mohammadhassanpour (2007).

Once the LVA field has been established, the procedure requires only a single additional parameter: the number of landmark points to use with MDS. This parameter should be set as high as possible given the amount of time the practitioner is willing to wait (Figure 4 right). Using a 10x10 pattern of landmark points in 2D and a 5x5x5 pattern in 3D is a reasonable starting point, with a minimum of 50 landmark points as recommended by Izenman (2008). The number of landmark points required for convergence of ISOMAP-L to traditional MDS depends on the LVA field.

5. Conclusions

The proposed methodology has been used to reproduce complex nonlinear features in a CO_2 emission data set. While the samples are synthetic, the benefits of the methodology are clear with an increase in cross validation results as well as a closer visual match with nonlinear features such as high concentrations along roadways.

The FMM was introduced as an alternative to the Dijkstra algorithm which can have high CPU requirements for large grids. The RAM requirements of the Dijkstra algorithm are intractable for very large grids >20M cells and are even difficult to manage for moderately sized grids >10M. Practitioners looking to apply the LVA methodology should use the FMM for distance calculations; if there are significant CPU resources available the Dijkstra algorithm can be implemented and compared to FMM. Preference is given to the algorithm that finds the shorter path.

6. References

Boisvert J, Deutsch C (2010) Programs for Kriging and Sequential Gaussian Simulation with Locally Varying Anisotropy Using Non-Euclidean Distances. Accepted to Computers & Geosciences, June 2010

Boisvert J (2010) Geostatistics with Locally Varying Anisotropy. Dissertation, University of Alberta

Cressie N, Majure J (1997) Spatio-temporal statistical modeling of livestock waste in streams. Journal of Agricultural, Biological and Environmental Statistics 2:24-47

Curriero F (2005) On the use of non-Euclidean distance measures in geostatistics. Mathematical Geology 38:907-926

Deutsch CV, Journel A (1998) GSLIB: Geostatistical Software Library and User's Guide. Oxford University Press, New York

Dijkstra E (1959) A note on two problems in connection with graphs. In Numerische Mathematik 1:269-271

Gurney K (2010) Vulcan Project. http://www.purdue.edu/eas/carbon/vulcan/index.php . Accessed March 2010

Hassanpour R (2007) Tools for multivariate modeling of permeability tensors and geometric parameters for unstructured grids. Dissertation, University of Alberta

Izenman A (2008) Modern Multivariate Statistical Techniques: Regression, Classification and Manifold Learning. Springer, New York

Loland A, Host G (2003) Spatial covariance modelling in a complex costal domain by multidimensional scaling. Environmetrics 14:307-321

Peyre G (2008) Toolbox Fast Marching - A toolbox for Fast Marching and level sets computations. http://www.mathworks.com/matlabcentral/fileexchange/6110. Accessed April 2010.

Sampson P, Guttorp P (1992) Nonparametric estimation of nonstationary spatial covariance structure. Journal of the American Statistical Association 87:108-119

Sethian JA (1999) Level Set Methods and Fast Marching Methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision and materials science. Cambridge University Press, New York

Siek J, Lee L, Lumsdaine A (2001) The Boost Graph Library: User Guide and Reference Manual. Addison-Wesley

Tenenbaum J, Silva V, Langford J (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290:2319–2323

ver Hoef J, Peterson E, Theobald D (2006). Spatial statistical models that use flow and stream distance. Environmental Ecological Statistics 13:449-464

7. Figures

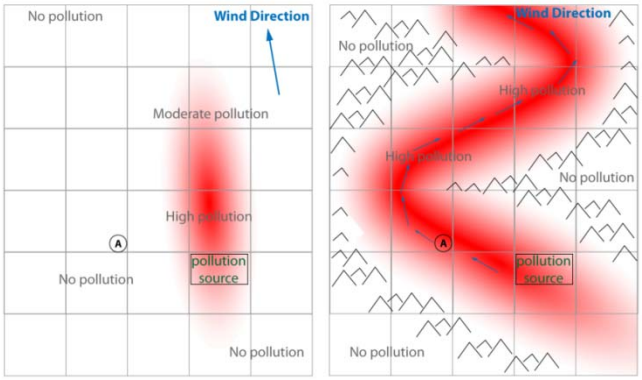

Figure 1: Cloud indicates increased concentration. Left: North wind, constant anisotropy. Right: Erratic wind, LVA.

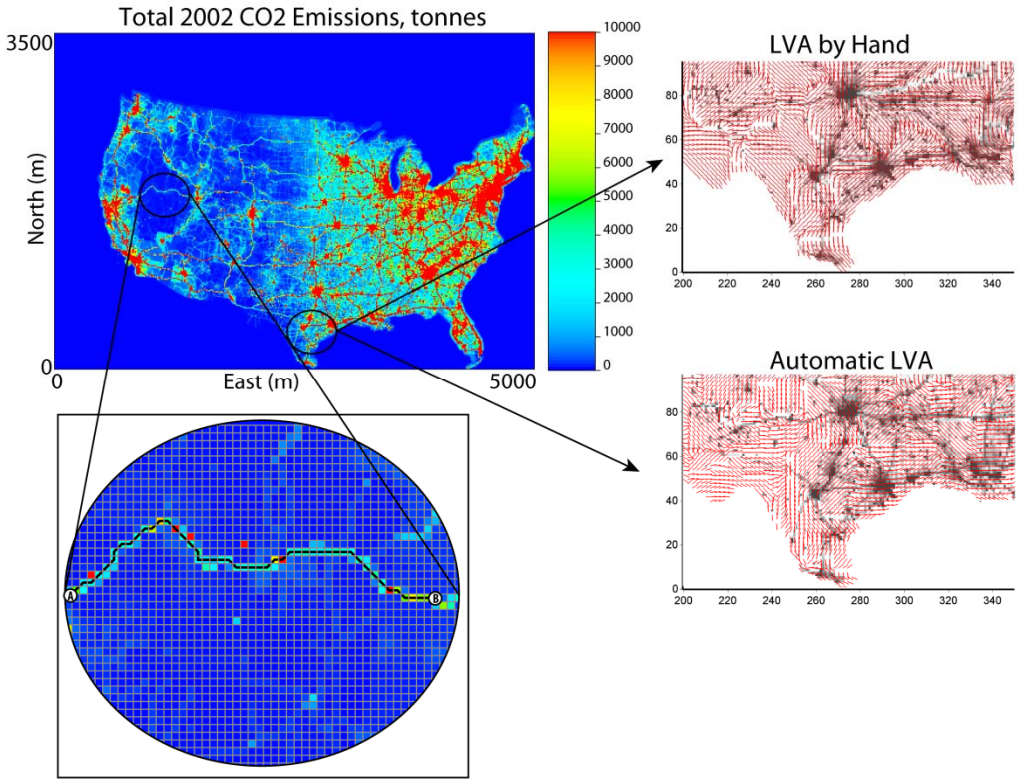

Figure 2: Upper Left: Tonnes of CO_2 emissions in the United States, 2002, 10km grid (source: Vulcan Project, Gurney 2010, data available online), graph with shortest path between A B is highlighted. Right: Sample area of the LVA fields.

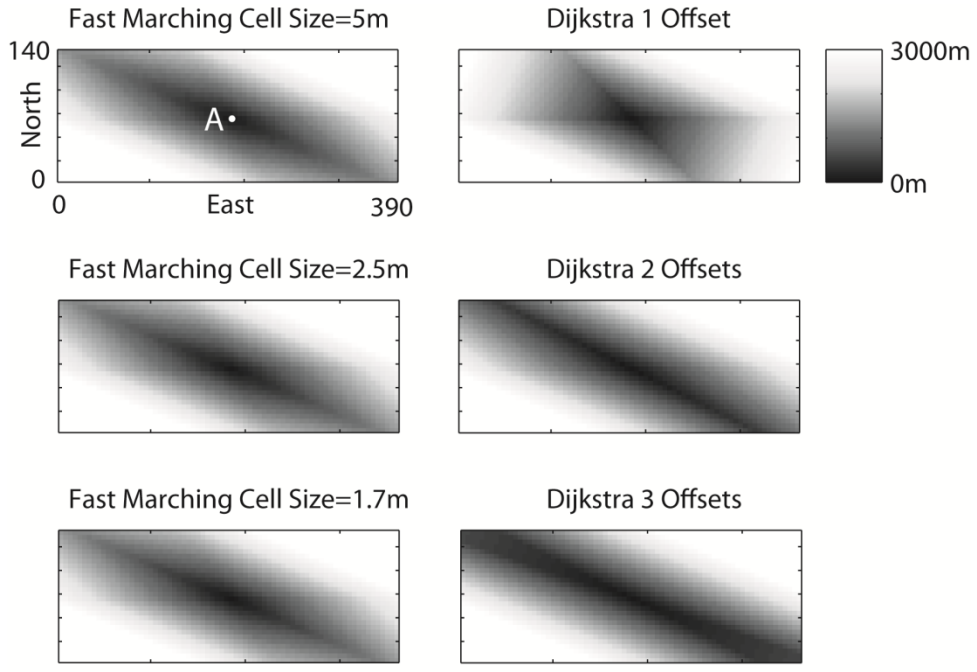

Figure 3: Distances calculated with the Dijkstra and fast marching algorithms. The section shows the shortest anisotropic distance from location A to all cells assuming a constant 20:1 anisotropy in the 112.5˚ direction.

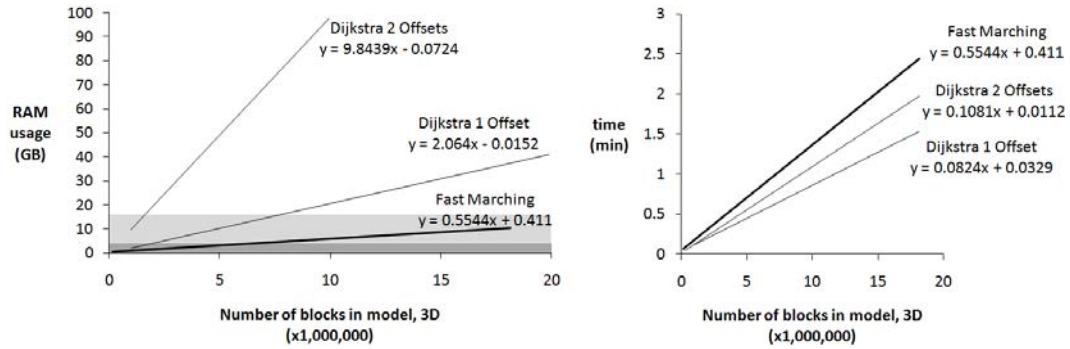

Figure 4: Left: RAM requirements of the shortest path algorithms. Highlights: A 32-bit machine limited to 4GB of RAM (dark gray region) and a 64-bit machine with 16GB of RAM (light gray region). Right: CPU time to calculate the distance from a landmark point to all cells in a model.

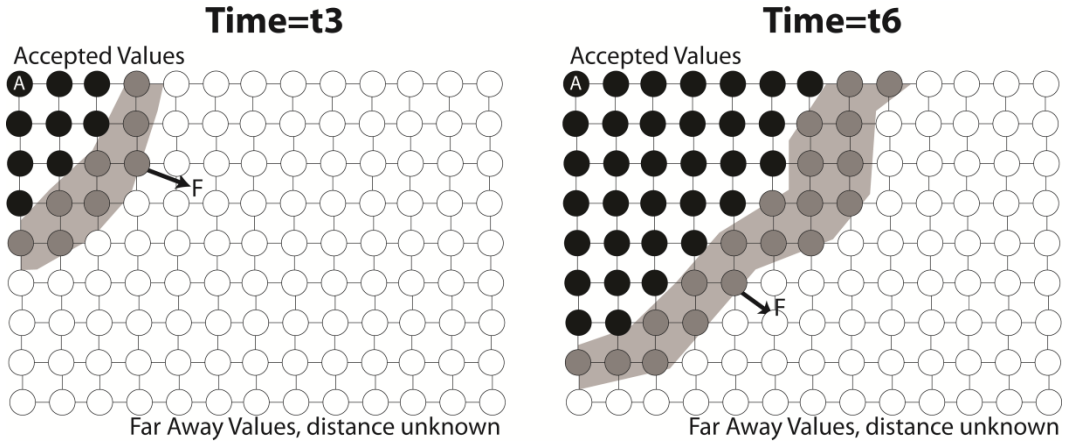

Figure 5: Progress of a front from location A, using FMM, modified from Sethian(1999).

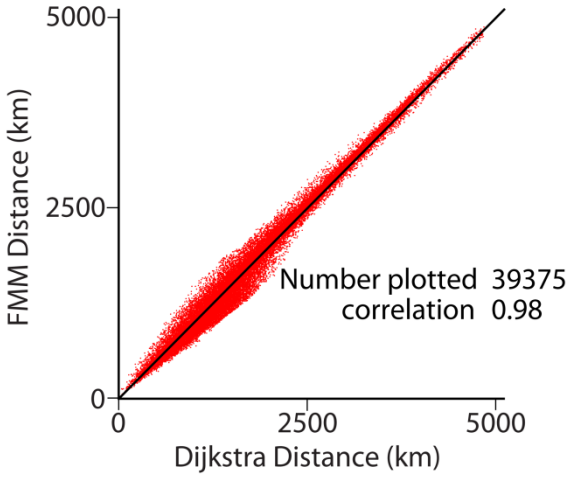

Figure 6: Comparison of the SPDs generated with the FMM and Dijkstra algorithms.

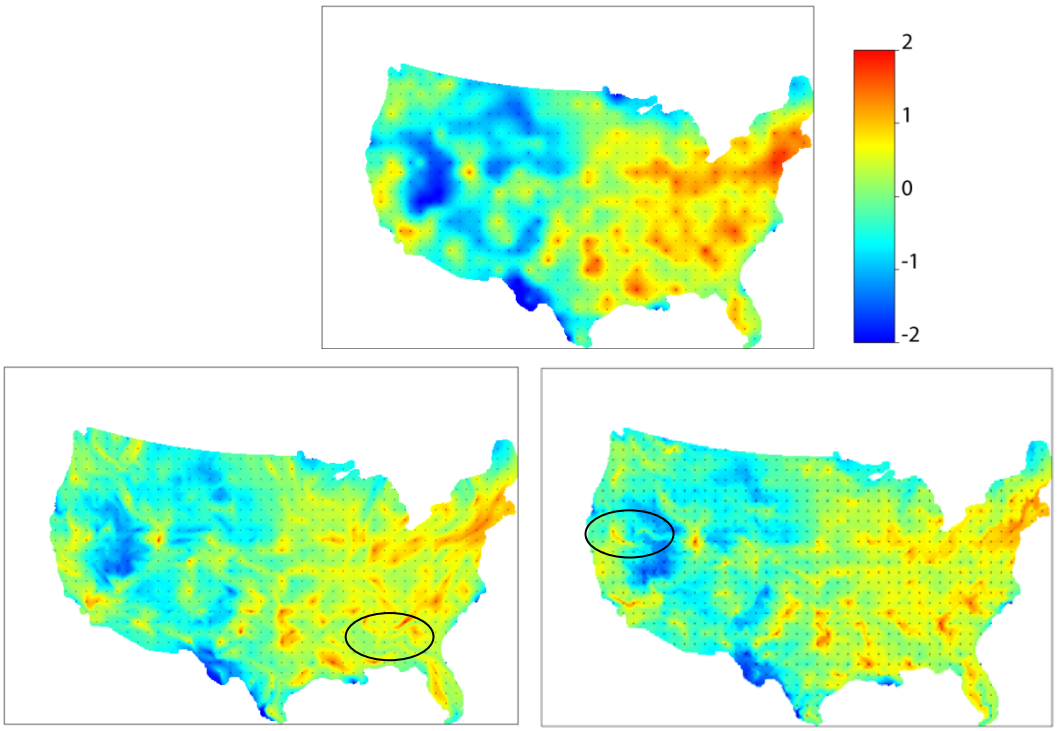

Figure 7: Above: Traditional kriging, estimate mean map (showing the locations of the 834 samples used). Lower Left: Kriging with the manual LVA field. Lower Right: Kriging with the automatic LVA field. Some areas where the LVA features are clearly visible are highlighted. Plot dimensions are 3500km x 5000km.

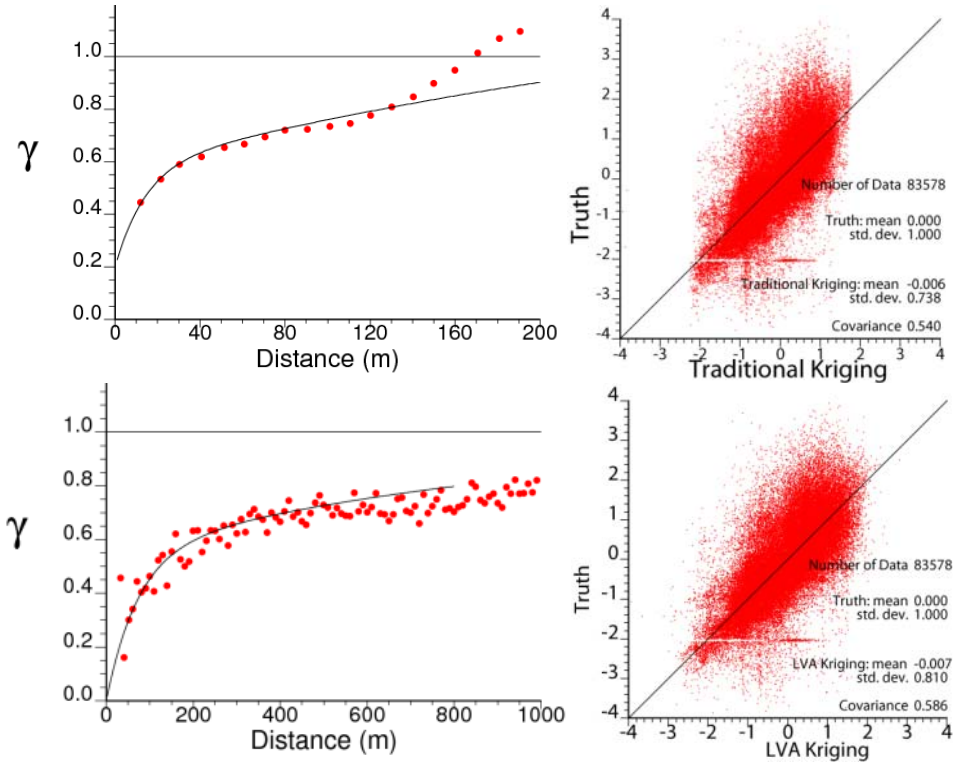

Figure 8: Left: Traditional variogram model (above) and isotropic LVA variogram (below). Right: Traditional kriging cross validation results (above) and LVA kriging cross validation results (below) from the automatic LVA field. The manual LVA field has a similar covariance of 0.583.