Advanced Space-Time Predictive Analysis With STAR-BME

Hwa-Lung Yu
Dept of Bioenvironmental Systems Engineering
National Taiwan University
Taipei, 10617, Taiwan
+886-2-3366-3454
hlyu@ntu.edu.tw

Shang-Jen Ku
Dept of Bioenvironmental Systems Engineering
National Taiwan University
Taipei, 10617, Taiwan
+886-2-3366-3454
ksj74208@hotmail.com

Alexander Kolovos
SpaceTimeWorks, LLC
255 G St #105
San Diego, CA 92101
+1-619-559-2497
alexander.kolovos@spacetimeworks.com

ABSTRACT
Stochastic analysis and prediction is an important component of space-time data processing for a broad spectrum of Geographic Information Systems scientists and end users. For this task, a variety of numerical tools are available that are based on established statistical techniques. We present an original software tool that implements stochastic data analysis and prediction based on the Bayesian Maximum Entropy methodology, which has attractive advanced analytical features and has been known to address shortcomings of common mainstream techniques. The proposed tool contains a library of Bayesian Maximum Entropy analytical functions, and is available in the form of a plugin for the Quantum GIS open source Geographic Information System software.

Categories and Subject Descriptors
D.2.2 [**Software and Engineering**]: Design Tools and Techniques – *modules and interfaces, software libraries*; G.3 [**Probability and Statistics**]: – *stochastic processes*

General Terms
Algorithms, Design, Theory

Keywords
Spatiotemporal analysis, stochastic processes, prediction, BME, modeling.

1. INTRODUCTION
The study of attributes in space-time can involve a variety of analytical computations. Such computations are integral components within Geographic Information Systems (GIS) modeling tools. In this scope, statistical techniques are used for such tasks as attribute prediction or simulation, tracking an attribute behavior, explaining geospatial phenomena, providing expert assessment ([1]-[3]). In particular, statistical techniques are employed to deal effectively with the element of uncertainty that exists inherently in many natural processes and/or might be introduced in the analysis by means of inaccurate measurements, lack of our exact knowledge about the attribute behavior, human error, etc. (e.g., [4]).

ACM SIGSPATIAL GIS' 12, November 6-9, 2012. Redondo Beach, CA, USA.

For geostatistical tasks such as attribute prediction, interpolation and simulation with spatial or space-time data, a number of software tools exist that provide solutions based on well-known stochastic methodologies in the literature; see, e.g., [5]-[9]. Such solutions can be found both in the open source community (GSLIB [10], modules in the R language, e.g., [11]), as well as in commercial packages ([12]-[13]).

Yet, the foundations of many common stochastic methodologies are built on restrictive assumptions and limitations. For example, linear models are used to describe phenomena that are nonlinear in nature; data are often assumed to follow Gaussian distributions due to internal requirements of linear interpolators; techniques are unable to account for uncertainty in interval or probabilistic data, and such measurements are either skipped or being used by reducing their informational content to single values.

As an alternative to classical geostatistical approaches that are burdened by the aforementioned shortcomings, Knowledge Synthesis (KS) was introduced as a cognitive framework to address these issues. Within this framework, the goal is to embrace all relevant information about an attribute in a rigorous, epistemic manner to improve the analysis. A well-tested KS methodology is the Bayesian Maximum Entropy (BME) technique; see, e.g., [14]. Numerical implementations of the BME theory have been previously available by means of advanced functions for space-time analysis and temporal GIS ([15]). To-date, two different software tools have been developed on the basis of this library of functions; namely, the Spatiotemporal Epistemic Knowledge Synthesis Graphical User Interface (SEKS-GUI; [16]) that is a Matlab toolbox, and the BME Graphical User Interface (BMEGUI) that is a Python-based application.

In this paper, we provide a brief overview of the KS principles and the BME methodology. We then propose a new tool oriented towards seamless GIS integration that offers key features of the BME analysis. We introduce the STAR-BME module (an acronym for Space-Time Analysis Rendering with BME) that is currently available as a plugin for the Quantum GIS (QGIS) open source software. We illustrate its features, functionality and ease of use by presenting the main steps through the different stages of spatiotemporal analysis in a simple modeling example.

2. KNOWLEDGE SYNTHESIS
2.1 A Framework for Space-Time Analysis
Predictive space-time analysis typically involves combining two discrete categories of information in attribute studies. First is the epistemic component that reflects our level of general knowledge regarding main attribute characteristics and scientific facts that apply in the analysis context. Second is the ontologic component that refers to the attribute data measurements at selected instances

in space and time. Mainstream spatiotemporal analysis techniques are conceptually limited with respect to the generality of resources they can employ and by their modeling assumptions.

The KS framework reviews these reasoning basics to extend the analysis characteristics, enable assimilation of multiple different knowledge sources, and thus achieve higher informativeness in the analysis input and output; see [17]-[18] for extended discussions and illustrative examples. To succeed in this goal, methodologies that implement KS are founded on a new perspective that extends common perception of data as single values. Under this viewpoint, soft data—for example, uncertain measurements in the form of intervals or probabilistic distributions—can be scientifically integrated in the stochastic analysis to retain the informational content of their uncertainty. Moreover, epistemic information assumes a role that is as important as the role of observed data; this is accomplished with the inclusion of precious knowledge about background processes—for example, applicable laws, models or principles—that contribute to the attribute behavior in space-time, and which would be otherwise ignored.

The following subsection expands on the specific KS framework BME methodology that is used by STAR-BME.

2.2 Bayesian Maximum Entropy

The BME methodology describes an operational Bayesian technique for prediction of unknown attribute values at selected space-time locations on a grid or individually specified. BME integrates different types of information in consecutive stages. The prior stage considers general knowledge bases (*G*-KB) that are derived from epistemic principles, such as physical laws and conceptual models. In the second (posterior) stage, the prior information is updated with information from the attribute data that BME terms as specificatory knowledge (*S*-KB). Blending the two types of knowledge bases yields the total given information K about the stochastic process.

Within a stochastic representation, an attribute is considered as a spatiotemporal random field $X_{\boldsymbol{p}}=X_{s,t}$ (S/TRF; [19]) at each location $\boldsymbol{p}=(\boldsymbol{s},t)$ in the unified spatial and temporal continuum with spatial coordinates $\boldsymbol{s}=(s_1,s_2)$ and temporal coordinate t. Specification of the $X_{\boldsymbol{p}}$ values at all points of the continuum determines a realization of the S/TRF. Randomness manifests itself as an ensemble of possible realizations of the $X_{\boldsymbol{p}}$ distribution. For a given vector $\boldsymbol{p}_k$ of prediction locations where attribute values are sought, BME starts with the given information K to compute the probability distribution f_K of the attribute values at each point $\boldsymbol{p}_k$. Given the f_K at $\boldsymbol{p}_k$, the prediction PDF yields a variety of attribute characteristics at the prediction locations, such as the attribute mean, its most probable value (mode), etc. See [14] for a detailed theoretical presentation of BME. Additional perspectives on KS and BME can be found at www.spacetimeworks.com.

The unique BME features have been illustrated in numerous studies across disciplines; see, e.g., [20]-[22]. In the following, we present the STAR-BME software module that brings to GIS the advanced features of spatiotemporal BME prediction as a dedicated QGIS plugin component.

3. INTRODUCING STAR-BME

In the past several years, GIS applications have been very successful in a variety of research fields. At the same time, a rapidly growing need has been experienced to implement temporal functionality in conventional GIS environments. Integration of spatiotemporal analysis tools and existing GIS software can lead to easier interaction between these tools and GIS built-in functions; in turn, this can facilitate GIS users in accessing and analyzing space-time data.

Based on this premise, the STAR-BME toolbox has been designed and developed so that it seamlessly integrates advanced functions of space-time analysis and modeling under conventional GIS environments. The result is the STAR-BME toolbox, which provides user-friendly access with graphical user interface for space-time analysis, and powerful, advanced modeling under the KS framework and BME in the QGIS environment. At the present stage, the major advantages of the STAR-BME toolbox for QGIS environment are two-fold as follows:

1) Implementation of functionalities for space-time data under the GIS environment
2) Development of KS-based prediction method for spatial and spatiotemporal data

The major features of the STAR-BME toolbox can be summarized as follows:

1) Practical display of space-time data in the GIS environment
2) Integration of multi-sourced space-time data in different formats
3) Display and incorporation of space-time data of multi-sourced uncertainties in probabilistic forms
4) Analysis of space-time dependence, by using empirical and mathematical spatiotemporal models
5) Prediction and mapping in space and time
6) Data export and display in multiple formats, including vector, raster and ASCII files

The STAR-BME module is free software that is publicly available at http://homepage.ntu.edu.tw/~hlyu/software/STAR.

4. STAR-BME ANALYSIS: A CASE STUDY OF PM10 IN TAIPEI

The STAR-BME toolbox is applied to spatiotemporal mapping of monthly particulate matter (PM10) concentration across the Taipei area in Taiwan during a 48-month period in 2004-2007 [20]. The PM10 data are observed from 26 stations across the area. For the purposes of software demonstration, the original dataset is divided into two parts, i.e. hard data and soft data. In particular, the hard data are the space-time observations directly obtained from the monitoring stations and assumed to be exact values; the soft data is data with uncertainty. We assume uncertainty to stem from known measuring issues at corresponding monitoring stations. The uncertainty is expressed by considering data from those stations as interval data within the ranges $[0.9m_{s,t};1.1m_{s,t}]$, where $m_{s,t}$ is the original observed

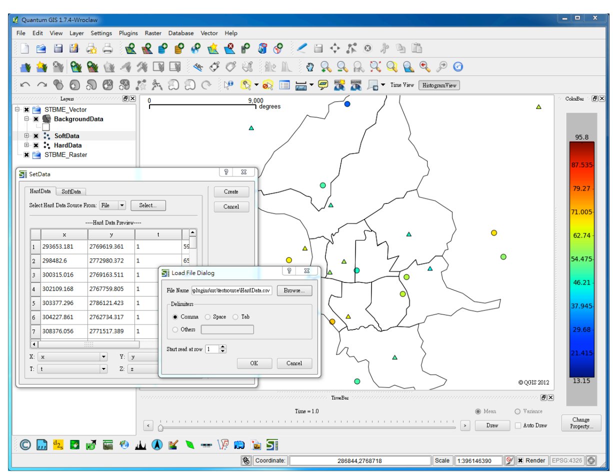

Figure 1. Map of PM10 hard and soft data in Taipei.

concentration levels. STAR-BME imports efficiently hard and soft data into QGIS for BME analysis. Figure 1 shows the data at instance t=1, where the layer of hard data is represented by circles, and the layer of soft data is shown as triangles. These symbols are filled with color that shows the value of the hard data or the mean of the soft data, according to the generated colorbar on the right side of the display window. The user can examine the data spatial distribution and values at any instance by sliding as desired the time bar at the bottom of the display window.

Additional perspectives of the space-time data are also available by means of the "time" and "histogram" views, in which the STAR-BME toolbox produces data views as time series and data histogram, respectively, at a certain spatial location. One such example is displayed in Figure 2.

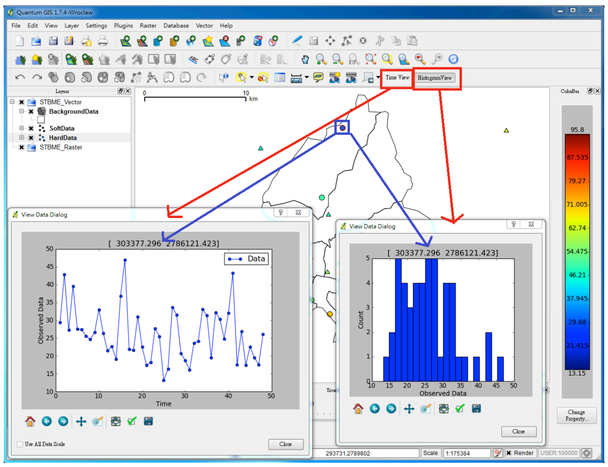

Figure 2. Time series and histogram views of data across time that were collected at a specific spatial location.

One aspect of the exploratory analysis in spatiotemporal modeling is investigating the data nonstationarity. This task is performed by means of spatiotemporal trend modeling. Estimation of a space-time trend removes large-scale variations from the original data, and yields stationary residuals for the purpose of prediction. STAR-BME provides the kernel smoothing and generalized

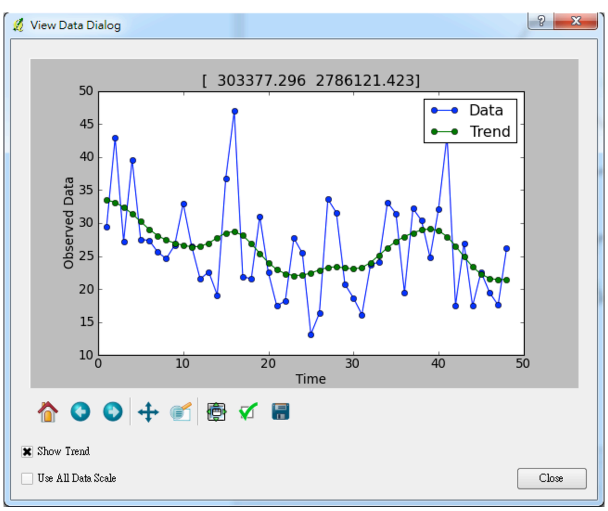

Figure 3. PM10 data time series and estimated trend at a selected monitoring location.

additive model options to estimate large-scale variation across space and time in the study domain. As an example of this process, Figure 3 shows the time series of original data values and the estimated trend at a certain monitoring station.

At a consecutive step, one needs to characterize the stochastic relationship between the space-time attributes. For this task, a suitable spatiotemporal dependence function should be employed, commonly known as a covariance (or a semivariance) model. STAR-BME computes the empirical space-time covariance at user-specified lags in both spatial and time domains. Then, a suitable model is fitted to the empirical function by using either manual or automatic algorithms. Figure 4 shows the parameters of the spatiotemporal covariance model that was fitted to the empirical one from the PM10 observations. The model is selected on the basis of the statistical Akaike Information Criterion (AIC). STAR-BME offers a variety of views for the selected space-time covariance model. For example, Figure 4 shows the marginal covariances at temporal lag τ=0 and spatial lag σ=0.

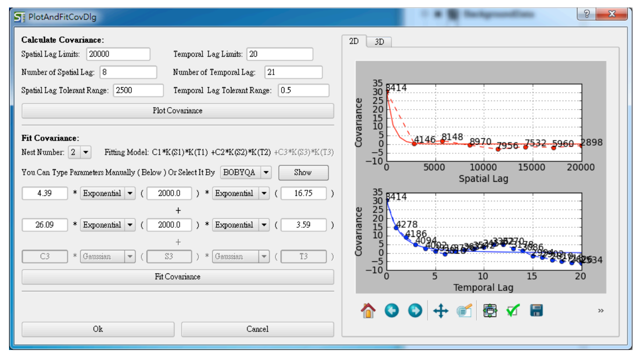

Figure 4. Space-Time covariance analysis in STAR-BME.

Eventually, STAR-BME uses the BME methodology to incorporate all hard and soft data for spatiotemporal prediction. Prediction can be made at user-specified points, grid or shapefiles. STAR-BME enables users to display the prediction results in an array of different spatial and temporal mapping perspectives.

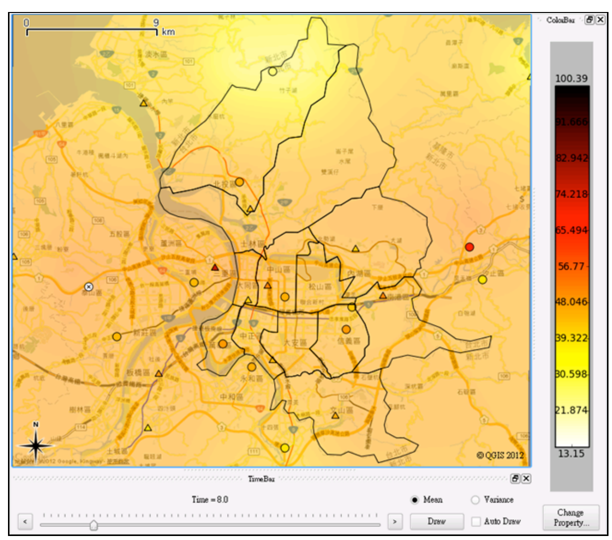

Figure 5. BME PM10 prediction means across the Taipei study area at a selected temporal instance (t=12). The prediction map overlays a Web Map Service area map.

For example, Figure 5 shows predicted PM10 values on a specified grid around Taipei at instance *t*=12. Also, Figure 6 shows a time series of the predicted values and their associated prediction error at a selected spatial location and at prediction instances from *t*=8 to *t*=15. Prediction results can be reused by specifying to store them in vector, raster, or ASCII text files.

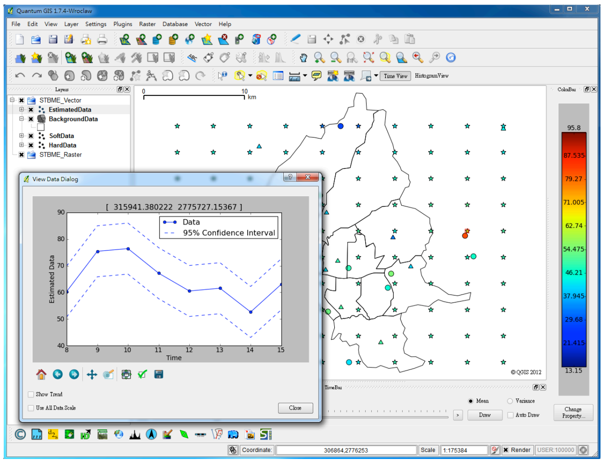

Figure 6. Time series of BME PM10 prediction means and errors at a selected spatial location for *t* in [8,15].

5. REFERENCES

[1] Alex Beutel, Thomas Mølhave, Pankaj K. Agarwal, Arnold P. Boedihardjo, and James A. Shine. 2011. TerraNNI: natural neighbor interpolation on a 3D grid using a GPU. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '11)*, Divyakant Agrawal, Isabel Cruz, Christian S. Jensen, Eyal Ofek, and Egemen Tanin (Eds.). ACM, New York, NY, USA, 64-74. DOI=10.1145/2093973.2093984 http://doi.acm.org/10.1145/2093973.2093984.

[2] Alexei Pozdnoukhov and Christian Kaiser. 2011. Scalable local regression for spatial analytics. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '11)*, Divyakant Agrawal, Isabel Cruz, Christian S. Jensen, Eyal Ofek, and Egemen Tanin (Eds.). ACM, New York, NY, USA, 361-364. DOI=10.1145/2093973.2094023 http://doi.acm.org/10.1145/2093973.2094023

[3] Guofeng Cao, Phaedon Kyriakidis, and Michael Goodchild. 2009. Prediction and simulation in categorical fields: a transition probability combination approach. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (GIS '09). ACM, New York, NY, USA, 496-499. DOI=10.1145/1653771.1653853 http://doi.acm.org/10.1145/1653771.1653853

[4] Dmitri V. Kalashnikov, Yiming Ma, Sharad Mehrotra, and Ramaswamy Hariharan. 2006. Index for fast retrieval of uncertain spatial point data. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems (GIS '06)*. ACM, New York, NY, USA, 195-202. DOI=10.1145/1183471.1183504 http://doi.acm.org/10.1145/1183471.1183504

[5] Matheron, G. 1963. Principles of Geostatistics. *Economic Geology,* 58, 1246–1266.

[6] Christakos, G. 1992. *Random field models in earth sciences*. San Diego: Academic Press.

[7] Chilès, J. P. and Delfiner, P. 1999. *Geostatistics-Modeling Spatial Uncertainty,* New York: John Wiley & Sons.

[8] Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation,* New York: Oxford University Press.

[9] Banerjee, S., Carlin, B. P., & Gelfand, A. E. 2004. *Hierarchical modeling and analysis for spatial data (monographs on statistics and applied probability)*. Boca Raton: Chapman & Hall/CRC Press.

[10] Deutsch, C. V., and A.G. Journel. 1992. *Geostatistical Software Library and User's Guide*. Oxford: Oxf Univ Press.

[11] Bivand, R.S., Pebesma, E.J., and V. Gómez-Rubio. 2008. *Applied spatial analysis with R*. New York: Springer.

[12] Johnston K., Ver Hoef J.M., Krivoruchko K., and N. Lucas. 2003. *ArcGIS 9: Using ArcGIS Geostatistical Analyst*. ESRI, Redlands, CA.

[13] SAS/STAT® User's Guide. *Available online at*: http://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm

[14] Christakos, G. 2000. *Modern spatiotemporal geostatistics*. New York: Oxford University Press.

[15] Christakos G., Bogaert P., and M.L. Serre. 2002. *Temporal GIS*. New York: Springer.

[16] Kolovos, A., Yu, H.-L., and G. Christakos. 2006. SEKS-GUI v.0.6 User Manual. San Diego: Dept of Geography, San Diego State University.

[17] Christakos G., Olea R.A., Serre M.L., Yu H.-L., and L.-L. Wang. 2005. Interdisciplinary Public Health Reasoning and Epidemic Modelling: The Case of Black Death. Springer-Verlag, New York, NY.

[18] Christakos, G. (2010). *Integrative problem-solving in a time of decadence*. New York: Springer.

[19] Christakos, G., and D.T. Hristopulos. 1998. *Spatiotemporal environmental health modelling: A Tractatus Stochasticus*. Boston: Kluwer Acad Publ.

[20] Yu H.-L., Ku S.-C., Yang C.-H., Cheng T.-J. and L. Chen. 2011. Assessment of Areal Average Air Quality Level over Irregular Areas: a Case Study of PM10 Exposure Estimation in Taipei (Taiwan), Advanced Air Pollution, Farhad Nejadkoorki (Ed.), ISBN: 978-953-307-511-2, InTech: http://www.intechopen.com/books/advanced-air-pollution/assessment-of-areal-average-air-quality-level-over-irregular-areas-a-case-study-of-pm10-exposure-est

[21] Yu H.-L., Yang S.-J., Yeh H.-J, and G. Christakos. 2011. A Spatiotemporal Climate-Based Model of Early Dengue Fever Warning in Southern Taiwan. *Stoch Envir Res and Risk Assessment*, 25(4), 485-494.

[22] Kolovos A., Skupin A., Jerrett M., and G. Christakos. 2010. Multi-Perspective Analysis and Spatiotemporal Mapping of Air Pollution Monitoring Data. *Envir Sci and Tech*, 44(17), pp.6738-6744, doi: http://dx.doi.org/10.1021/es1013328