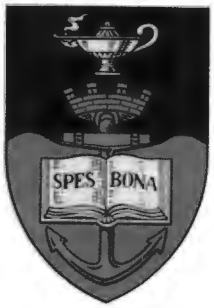
Contributions to Spatial Uncertainty Modelling in GIS: Small Sample Data



Danni Guo

Department of Statistical Sciences
University of Cape Town

June 2007

PhD Thesis Submitted in Fulfillment of the Degree of Doctor of Philosophy, in the Department of Statistical Sciences, University of Cape Town, South Africa.

Supervisor: Associate Professor Christien Thiart

Contributions to Spatial Uncertainty Modelling in GIS: Small Sample Data

Danni Guo

Department of Statistical Sciences
University of Cape Town

June 2007

PhD Thesis Submitted in Fulfillment of the Degree of Doctor of Philosophy, in the Department of Statistical Sciences, University of Cape Town, South Africa.

Supervisor: Associate Professor Christien Thiart

Declaration

I, Danni Guo, hereby,

1. grant the University of Cape Town free licence to reproduce the above thesis in whole or in part, for the purpose of research;
2. declare that:
    i. the above thesis is my own unaided work, both in concept and execution,
    ii. neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

I am now presenting the thesis for examination for the degree of PhD.

Danni Guo

Abstract

The thesis title is "*Contributions to Spatial Uncertainty Modelling in GIS: Small Sample Data*". Environmental data is very costly and difficult to collect and are often vague (subjective) or imprecise in nature (e.g. hazard level of pollutants are classified as "harmful for human beings"). These realities in practise (fuzziness and small datasets) leads to uncertainty, which is addressed by my research objective: "*To model spatial environmental data with fuzzy uncertainty, and to explore the use of small sample data in spatial modelling predictions, within Geographic Information System (GIS).*"

The methodologies underlying the theoretical foundations for spatial modelling are examined, such as geostatistics, fuzzy mathematics Grey System Theory, and $(\vee,\cdot)$ Credibility Measure Theory. *Fifteen papers* including *three journal papers* were written in contribution to the developments of spatial fuzzy and grey uncertainty modelling, in which I have a contributed portion of 50 to 65%. The methods and theories have been merged together in these papers, and they are applied to two datasets, PM_{10} air pollution data and soil dioxin data. The papers can be classified into two broad categories: fuzzy spatial GIS modelling and grey spatial GIS modelling.

In fuzzy spatial GIS modelling, the fuzzy uncertainty (Zadeh, 1965) in environmental data is addressed. The thesis developed a *fuzzy membership grades kriging* approach by converting fuzzy subsets spatial modelling into membership grade spatial modelling. As this method develops, the fuzzy membership grades kriging is put into the foundation of the credibility measure theory, and approached a full data-assimilated membership function in terms of maximum fuzzy entropy principle. The variable modelling method in dealing with fuzzy data is a unique contribution to the fuzzy spatial GIS modelling literature.

In grey spatial GIS modelling, spatial predictions using small sample data is addressed. The thesis developed a *Grey GIS* modelling approach, and two-dimensional order-less spatially observations are converted into two one-dimensional ordered data sequences. The thesis papers also explored foundational problems within the grey differential equation models (Deng, 1985). It is discovered the coupling feature of grey differential equations together with the help of ε-similarity measure, generalise the classical GM(1,1) model into more classes of extended GM(1,1) models, in order to fully assimilate with sample data information. The development of grey spatial GIS modelling is a creative contribution to handling small sample data.

Acknowledgements

My sincere thanks to my supervisor, Associate Professor Christien Thiart, for her valuable advice and directions, and for her continuous support and help in my thesis.

I am grateful to the California Air Resources Board for providing me with the air quality data used in this thesis.

I would like to thank Associate Professor Tonny Oyana, for providing me with the soil dioxin data used in this thesis.

And lastly, many thanks to my parents for their encouragements and patience.

Table of Contents

Declaration i

Abstract ii

Acknowledgements iii

Table of Contents iv

List of Figures vi

List of Tables ix

Chapter 1. Introduction **1-1**

1.1 Spatial GIS Modelling 1-1

1.2 Aims and Objectives 1-2

1.3 Overview of Thesis 1-5

Chapter 2. An Introduction to Methodology and Datasets **2-1**

2.1 Impreciseness and Uncertainty 2-1

2.2 Geostatistics 2-6

2.3 Fuzzy Mathematics 2-8

2.4 Grey Systems Theory 2-10

2.5 $(\vee,\cdot)$ Credibility Measure Theory 2-14

2.6 Geographic Information System (GIS) 2-15

2.7 Datasets Used in the Thesis 2-15

Chapter 3. Collection of Papers **3-1**

3.1 Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM_{10} for California, Using Fuzzy Kriging 3-3

3.2 Grey GIS 3-12

3.3 Grey Spatial Prediction of Fuzzy Air Quality Index 3-21

3.4 A New Very Small Sample-Based Non-Linear Statistical Estimation Method 3-31

3.5 Optimal Data Transformations in Grey Analysis 3-45

3.6 Grey Reliability Analysis of Complex system 3-54

3.7 GM(1,1)-Kriging Prediction of Soil Dioxin Pattern 3-63

3.8 Predicting Air Pollution Using Fuzzy Membership Grade Kriging 3-75

3.9 Rationale of GM(1,1) modelling – A Variation Approach 3-97
3.10 Generalizations to Standard GM(1,1) Model 3-105
3.11 Rationale Exploration of First-Order One Variable Grey Differential Equation Model via Variational Approach 3-115
3.12 The Coupling of regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models 3-123
3.13 Bivariate Credibility-Copulas 3-139
3.14 Kernel Estimation for Bivariate Credibility Copulas 3-150
3.15 Credibility Measure-Based Fuzzy Membership Grade Kriging 3-162
Chapter 4. Discussion **4-1**
4.1 Summary and Assessment 4-7
4.2 Future Developments 4-7
Chapter 5. Conclusion **5-1**

References R-1
Appendix A: Grey Temporal-Spatial GIS A-1
Appendix B: Summary of Contributed Papers B-1

List of Figures

Figure 1.3.1	Overview of Thesis Papers within the Research Process.	1-5
Figure 2.3.1	A Set and Its Indicator Function.	2-10
Figure 2.3.2	A Fuzzy Set and Its Membership Function.	2-10
Figure 2.7.1	PM_{10} Sample Locations in California.	2-17
Figure 2.7.2	Air Quality Monitoring Sites in Santa Barbara County, California.	2-17
Figure 2.7.3	PM_{10} Sample Locations for 2002 in Santa Barbara County, California.	2-17
Figure 2.7.4	Soil Dioxin Samples in Midland County, Michigan.	2-19
Figure 3.1.1	PM_{10} Sample Locations in California.	3-4
Figure 3.1.2	Linear Sample Membership Function of PM_{10}.	3-6
Figure 3.1.3	Fuzzy Variogram of Linear Sample Membership Function.	3-7
Figure 3.1.4	Standardized Error Plot of Linear Sample Membership Function.	3-7
Figure 3.1.5	Prediction Map of Linear Sample Membership Function.	3-8
Figure 3.1.6	Prediction Standard Error Map of Linear Sample Membership Function.	3-8
Figure 3.1.7	PM_{10} Concentrations Predicted by Linear Membership Function.	3-10
Figure 3.2.1	Air Quality Monitoring Sites in Santa Barbara County.	3-14
Figure 3.2.2	PM_{10} Sample Locations for 2002 in Santa Barbara County.	3-14
Figure 3.2.3	Grey Prediction for PM_{10} in 2002 in Santa Barbara County.	3-19
Figure 3.2.4	Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.	3-20
Figure 3.3.1	Air Quality Monitoring Sites in Santa Barbara County, California.	3-24
Figure 3.3.2	PM_{10} Sample Locations for 2002 in Santa Barbara County, California.	3-24
Figure 3.3.3	Illustration of the Projection of z(x,y)-value into the x-axis and y-axis respectively.	3-25
Figure 3.3.4	Illustration of the Projection of x(x,y)-value into x-axis and y-axis for Santa Barbara.	3-25
Figure 3.3.5	Grey Prediction for PM_{10} in 2002 in Santa Barbara County.	3-27
Figure 3.3.6	Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.	3-27

Figure 3.3.7	Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.	3-29
Figure 3.4.1	Air Quality Monitoring Sites in Santa Barbara County.	3-35
Figure 3.4.2	Figure 3.4.2 PM_{10} Sample Locations for 2002 in Santa Barbara County.	3-35
Figure 3.4.3	Illustration of the Projection of z(x,y)-value into the x-axis and y-axis respectively.	3-36
Figure 3.4.4	Illustration of the Projection of x(x,y)-value into x-axis and y-axis for Santa Barbara	3-36
Figure 3.4.5	Grey Prediction for PM_{10} in 2002 in Santa Barbara County.	3-39
Figure 3.4.6	Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.	3-39
Figure 3.7.1	Soil Dioxin Samples in Midland County.	3-65
Figure 3.7.2	Kriging Map of Soil Dioxin in Midland County.	3-66
Figure 3.7.3	Area Division in Midland County.	3-71
Figure 3.7.4	Grid Creation in Area 3.	3-71
Figure 3.7.5	Grey Dioxin Predictions for Midland County.	3-73
Figure 3.7.6	GM(1,1) Ordinary Kriging Map for Midland County.	3-73
Figure 3.8.1	PM_{10} Sample Locations in California.	3-76
Figure 3.8.2	Linear Sample Membership Function of PM_{10}.	3-84
Figure 3.8.3	QQ Plot of Linear Sample Membership Function.	3-85
Figure 3.8.4	Predicted Fuzzy Membership Grades Using a Linear Sample Membership Function.	3-85
Figure 3.8.5	Standard Error Map Using the Linear Sample Membership Function.	3-85
Figure 3.8.6	Quadratic Sample Membership Function of PM_{10}.	3-86
Figure 3.8.7	QQ Plot of Quadratic Sample Membership Function.	3-87
Figure 3.8.8	Predicted Fuzzy Membership Grades Using a Quadratic Sample Membership Function.	3-87
Figure 3.8.9	Standard Error Map Using the Quadratic Sample Membership Function.	3-88
Figure 3.8.10	Hyperbolic Tangent Sample Membership Function of PM_{10}.	3-89
Figure 3.8.11	QQ Plot of Hyperbolic Tangent Sample Membership Function.	3-89
Figure 3.8.12	Predicted Fuzzy Membership Grades Using a Hyperbolic Tangent Sample Membership Function.	3-90

Figure 3.8.13	Standard Error Map Using the Hyperbolic Tangent Sample Membership Function.	3-90
Figure 3.8.14	PM_{10} Concentrations Predicted by a Linear Membership Function.	3-92
Figure 3.8.15	PM_{10} Concentrations Predicted by a Quadratic Membership Function.	3-93
Figure 3.8.16	PM_{10} Concentrations Predicted by a Hyperbolic Tangent Membership Function.	3-94
Figure 3.15.1	PM_{10} Sample Locations in California	3-170
Figure 3.15.2	Hyperbolic Tangent Membership Function with Data-Assimilated Parameters $(\hat{\delta},\hat{\eta}) = (28.00, 6.004)$.	3-172
Figure 3.15.3	QQ Plot of Hyperbolic Tangent Membership Function.	3-172
Figure 3.15.4	Standard Error Map Using the Hyperbolic Tangent Membership Function.	3-173
Figure 3.15.5	PM_{10} Concentrations Predicted by a Membership Grade Kriging in terms of a Fully Data-Assimilated Hyperbolic Tangent Membership Function.	3-174
Figure 3.15.6	Ordinary Kriging Map Predicted Using Original Unadjusted PM_{10} Data.	3-175
Figure A.1	Air Quality Monitoring Sites in Santa Barbara County.	A-2
Figure A.2	PM_{10} Sample Locations for 2002 in Santa Barbara County.	A-2
Figure A.3	Grey Prediction for PM_{10} in 2002 in Santa Barbara County.	A-3
Figure A.4	Three Samples Sites with Historical PM_{10} Data in Santa Barbara County.	A-4
Figure A.5	Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.	A-8
Figure A.6	Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.	A-9

List of Tables

Table 2.4.1	Three System Dynamics on Information (in control theory context).	2-12
Table 2.4.2	Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.	2-13
Table 3.1.1	Conversion from Linear Membership to PM_{10} Concentrations.	3-9
Table 3.2.1	Projected Santa Barbara County PM_{10} Data Sequences and Coordinates.	3-16
Table 3.3.1	Fitted Probability Distributions from California State Data (2002).	3-22
Table 3.3.2	Projected Santa Barbara County PM_{10} Data Sequence and Coordinates.	3-26
Table 3.3.3	PM_{10} Time Sequences from Year 1988 to 1999 at Monitoring Site 17, 38 and 57.	3-27
Table 3.3.4	Information of the Six Monitoring Sites for Grey Prediction of Remaining 17 Sites.	3-28
Table 3.3.5	Grey Temporal Predictions of PM_{10} from 1992 to 2002 at Site 17, 38 and 57.	3-28
Table 3.3.6	Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.	3-28
Table 3.4.1	Pm_{10} Sample Information in Four Sites, 2002.	3-35
Table 3.4.2	Coordinate-Directional Observed PM_{10}-Valued Vectors.	3-35
Table 3.4.3	Projected Santa Barbara County PM_{10} Data Sequence and Coordinates.	3-36
Table 3.4.4	Stopping times and related covariates (normalized).	3-42
Table 3.5.1	The Range of the Grey Development Coefficient.	3-48
Table 3.7.1	Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.	3-67
Table 3.7.2	Equal-spaced Coordinated Points with Kriging Dioxin values.	3-72
Table 3.10.1	Extended GM(1,1) Models.	3-113
Table 3.12.1	Extended GM(1,1) Models.	3-138
Table A.1	PM_{10} Time Sequences from year 1988 to 1999 at Monitoring Site 17, 38 and 57.	A-3
Table A.2	Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.	A-4
Table A.3	Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.	A-5

Table A.4	Grey Temporal Predictions of PM_{10} from 1992 to 2002 at Site 17, 38 and 57.	A-7
Table A.5	Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.	A-8
Table B.1	Summary of Thesis Papers.	B-1

Chapter 1. Introduction

1.1 Spatial GIS Modelling

Generally speaking, events that can not be described by precise deterministic laws, is of uncertainty. Uncertainty is considered an insufficient information state of intrinsic nature. The focal research objective of this thesis is to model spatial environmental data with fuzzy uncertainty, and to explore the use of small sample data in spatial modelling predictions, within Geographic Information System (GIS).

The methodologies underlying the theoretical foundations for spatial modelling are examined, such as geostatistics, fuzzy mathematics, Grey System Theory, and $(\vee,\cdot)$ Credibility Measure Theory. Fifteen papers were written in contribution to the developments of spatial fuzzy and grey uncertainty modelling. The methods and theories have been merged together in these papers, and they are applied to two datasets, PM_{10} air pollution data of California and soil dioxin data of Midland County in Michigan. ArcGIS and Excel are used in the modelling process. The thesis papers can be classified into two broad categories: fuzzy spatial GIS modelling and grey spatial GIS modelling.

In fuzzy spatial GIS modelling, the fuzzy uncertainty in environmental data is addressed. Spatially distributed fuzzy data is in nature in the form of fuzzy subsets (Zadeh, 1965), and therefore is difficult to perform spatial modelling and presentation using GIS. The thesis papers aim at creating a convenient fuzzy kriging methodology. This led to the development of fuzzy membership grades kriging approach by converting fuzzy subsets spatial modelling into membership grade spatial modelling, which is easy to handle and implement into GIS. Although the early papers (chapter 1, 3, 8) used classical fuzzy mathematics concepts and methodology, it emphasized the data assimilated membership function by determining parameters of membership function in semi-statistical manner. In the later papers (chapter 13, 14, 15), the fuzzy membership grades kriging is put into the foundation of the credibility measure theory, and approached a full data-assimilated membership function in terms of maximum fuzzy entropy principle. The variable modelling method in dealing with fuzzy data is a unique contribution to the fuzzy spatial GIS modelling literature.

In grey spatial GIS modelling, spatial predictions using small sample data is addressed. The papers started with a projection approach (chapter 2, 4), however, the two-dimensional order-less spatially observations are converted into two one-dimensional ordered data sequences, and therefore the grey differential equation models can be successful applied. After the earlier Grey GIS modelling applications, it is found that there are a series of fundamental issues affecting the creation of grey spatial GIS modelling with small sample data. Therefore, the later papers (chapter 5, 6, 9, 10, 11, 12) explored foundational problems such as the transformations, rationale underling GM(1,1) model via variational calculus, and ε-similarity measure between two geometric curves (i.e., functions) etc. It is discovered the coupling feature of grey differential equations together with the help of ε-similarity measure, generalise the classical GM(1,1) model proposed by Deng (1985) into more classes of extended GM(1,1) models, in order to fully assimilate with sample data information. The development of grey spatial GIS modelling is a creative contribution to handling small sample data.

1.2 Aims and Objectives

The overall objective of this thesis is:

- To model spatial environmental data with fuzzy uncertainty, and to explore the use of small sample data in spatial modelling predictions, within GIS environment.

The aims of this thesis can be classified into two broad categories: fuzzy spatial GIS modelling and grey spatial GIS modelling. In fuzzy spatial GIS modelling, we explored spatial environmental data with fuzzy uncertainty, and in grey spatial GIS modelling, we explored the use of small sample data in spatial modelling predictions.

Fuzzy spatial GIS aims:

- To develop and explore a new spatial analysis technique, fuzzy membership grade kriging with semi-statistical membership, in order to address fuzzy spatial data recorded as crisp numbers (chapter 3.1, 3.8).
- To predict a spatial distribution of air fuzzy quality index, and to develop a small-sample based fuzzy spatial data analysis methodology based on grey differential equation theory, and convert order-less 2-dimensional spatial data into ordered data sequences (chapter 3.3).

- To develop the concept of fuzzy copula on the ground of $(\vee,\cdot)$-credibility measure theory, for the characterization of the full relationship among fuzzy variables (chapter 3.13).
- To develop a kernel-estimation based on the maximum entropy principle under the framework of $(\vee,\cdot)$-credibility measure theory (chapter 3.14).
- To establish a parameter estimation of the membership function in terms of maximum entropy principle on the ground of self-dual credibility measure theory (chapter 3.15).

Grey spatial GIS aims:

- To develop and explore a new small-sample based spatial data analysis methodology based on Grey System Theory (chapter 3.2).
- To investigate a new least-square theory based estimation technique under very small sample circumstances (chapter 3.4).
- To explore a new approach which addresses ill-designed spatial data, and to provide a better ordinary kriging results, and develop a mixed approach by combining grey differential equation models (chapter 3.7).
- To explore the transformation of the discrete data sequence, in Grey System Theory, and to develop two new transformations (chapter 3.5).
- To investigate a way to obtain the dynamic version of grey system reliability with sparse data availability, in terms of grey differential equation models (chapter 3.6).
- To investigate the underlying mechanism of GM(1,1) model, particularly, the lifetime partition into average functioning time, (average) repair-improvement and random error via GM(1,1), in terms of a variational viewpoint (chapter 3.9).
- To examine GM(1,1) modelling mechanism from its component-level models, and to propose families of extended GM(1,1) models via the coupling principle in GM(1,1) model (chapter 3.10).
- To develop a consistence measure in terms of the concept of grey relational analysis, and explore the evolving behaviour of a set of constraint functionals defined by ordinary differential equations in order to select the suitable functional satisfying the consistence measure during GM(1,1) modelling process (chapter 3.11).

- To examine the GM(1,1) model from its component-level and therefore identify the interactive coupling nature of differential equation model and corresponding regression model constituting of a GM(1,1) model (chapter 3.12).

1.3 Overview of Thesis

The thesis is divided into five broad chapters: chapter 1 - Introduction, chapter 2 - An Introduction to Methodology and Datasets, chapter 3 - Collection of Papers, chapter 4 - Discussion, and chapter 5 - Conclusion.

In chapter 1, spatial GIS modelling is discussed, the aims and objectives for the thesis are stated, and a general overview of the thesis is given.

In chapter 2, concepts and methodologies that are used to handle uncertainty in small datasets are introduced. The methods are: Impreciseness and Uncertainty, Geostatistics, Fuzzy, Logic, Grey System Theory, $(\vee,\cdot)$ Credibility Measure Theory, and Geographic Information System (GIS); and lastly, the air and soil pollution datasets used in the papers are introduced.

In chapter 3, fifteen journal and conference papers were written in contribution to the developments of spatial fuzzy and grey uncertainty modelling. Figure 1.3.1 provides an overview of the thesis papers within the research process. The numbers in the figure refers to the time sequence of the papers, the first paper being 1, and the last paper being 15. In the fifteen thesis papers, fourteen of the peer-reviewed papers were already published, and one paper is under-print. There are a total of three journal papers and twelve conference papers. In this chapter, the methods and datasets that are introduced in chapter 2 are applied and used in the papers.

In chapter 4, the papers are actively examined and criticized, and an overall view of the current development in the research is discussed, and the future developments and new directions for further research are outlined.

In chapter 5, a small summary of the thesis is given, stating the contribution of this thesis to spatial GIS literature, and examine the achievement of aims stated in chapter 1.

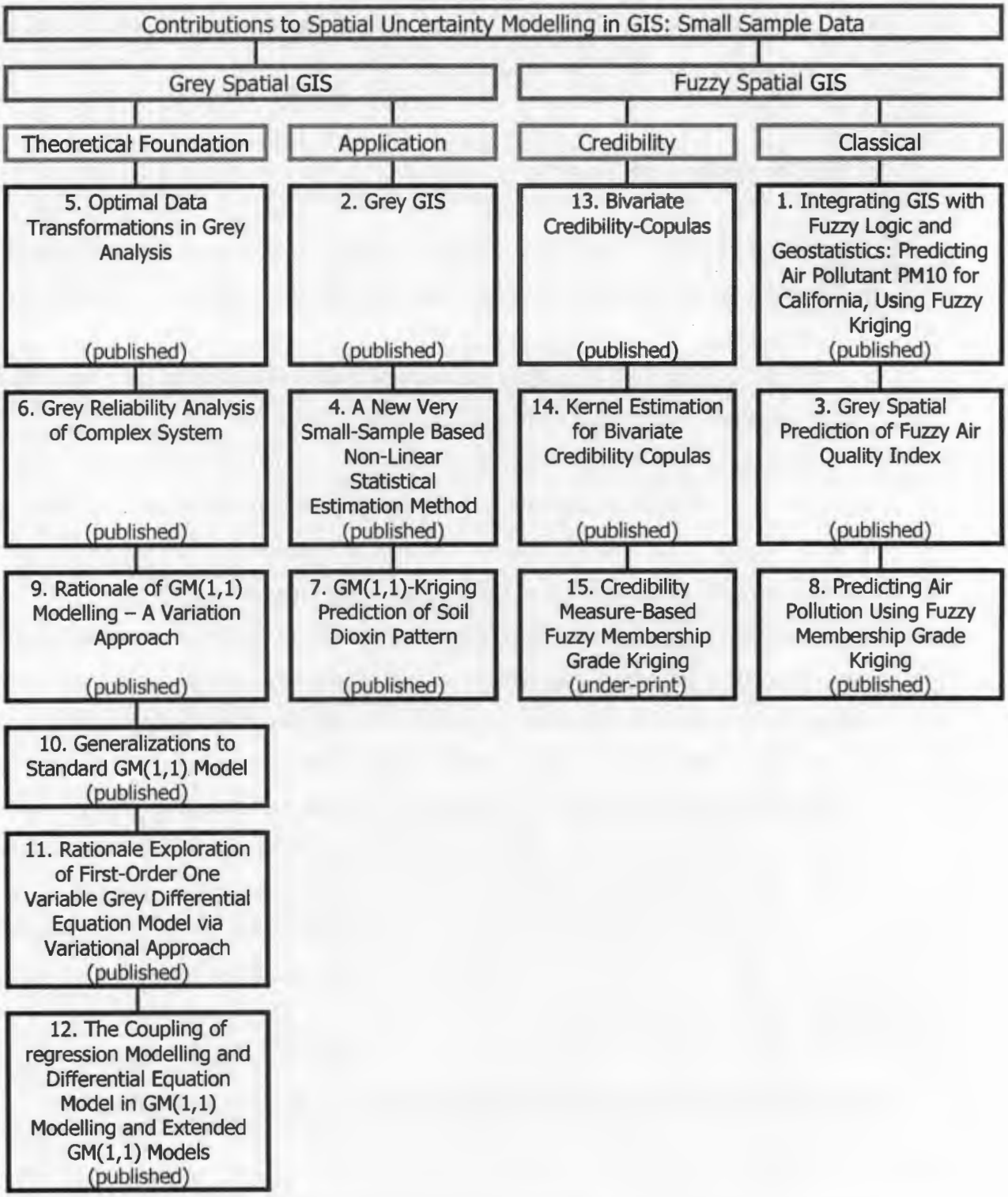


Figure 1.3.1 Overview of Thesis Papers within the Research Process.

Chapter 2. An Introduction to Methodology and Datasets

2.1 Impreciseness and Uncertainty

In the recent years, it seems that geostatistical developments do not involve sufficiently the characteristics of spatial information which reflects the underlying physical mechanism, and consequently geostatistical models do not approach the realities well enough. One of the reasons why geostatistical models are moving away from geostatistical applications may root deeply in the modern scientific methodology itself.

Just as Alvin Toffler pointed out sharply in the Forward- Science and Change of the book "Order out of Chaos" (Prigogine and Stengers, 1984), "One of the most highly developed skills in contemporary Western civilization is dissection: the split-up of problems into their smallest possible components. We are good at it. So good, we often forget to put the pieces back together again." In other words, it is often the practice in geostatistical modelling to isolate the focus point from its environment, and therefore to ignore the complex interactions of the problem and the rest of the real world.

However, two aspects need to be addressed in order to deal with the fatal weakness of the modern dissection methodology. Firstly, in the modelling process it is critical to catch up the fundamental characteristic of the problem studied, because such a kind of fundamental characteristic is common to the problem itself and its environment as well. Secondly, it is an essential task to account for the environmental impacts, that is, the interactions between the problem studied and its environment.

Uncertainty and certainty are one of the fundamental characteristics of our world surrounding us. In the human history, certainty is the first form of perception of the living world and often thought as a basic feature of existing world and an elemental human thinking pattern. Certainty is in nature a phenomenon resulting from all the necessary and sufficient conditions for the occurrence of an event are satisfied.

Only in the last few decades, uncertainty, particularly random uncertainty resulting from a partial of the conditions of an event's occurrence, was understood and received by scientific community. Probability calculus and statistical methodology are the mathematical foundation for

the understanding and modelling random uncertainty. Since the 1960s, other forms of uncertainty, especially fuzzy uncertainty (Zadeh, 1965), which reflect the unclear boundary of events, received notice from the engineering community.

Philosophically speaking, both randomness and fuzziness are objectively existing, inherent, and inseparable characteristics of the world reality.

1. Firstly, every event occurred in realistic world, no matter how complicated it might be, always enjoys its own self-existence and self-specification and therefore it is objective and certain.
2. The self-specification associated with each existing object does not often enjoy a clear and crisp boundary, although the pattern of change can be either evolving or sudden jump. However, it is not in general sudden jump from one level to another, but more possibly evolving from one stratum to another. In other words, between two different strata, there exist some middle layers. At each layer in-between, the event possesses features in various degrees from both strata. Therefore, vague boundaries and specifications are an objective existing form of our world reality.
3. The occurrence of any event ought to satisfy its generating conditions. If all the related generating conditions satisfied, it would be a certainty event. If partial generating conditions satisfied, then the event will be a chance event. So the cause-effect relationship is also an objectively existing form of changes.
4. The true states of existing real world are never completely grasped by human beings. What is available about the real world is the collected information. However, information collection itself is inevitably realized with the brand of times. The instruments and the depth of knowledge about our objective existing world are limited, by the level of science and technology achieved at the time. Therefore, the information collected at any time will have observational errors (random in general), and observational biases (fuzzy in nature).

Fuzzy and random uncertainties are intrinsically associated with geostatistical information data. In order to correctly utilize data information in the modelling process, we have to understand the fundamental and intrinsic feature of information, uncertainty. Logically speaking, randomness and fuzziness are two different types of uncertainties. Randomness is logically the break down of the law of causality, because of the lack of some conditions under which the event occurrence is

inevitable. It is the traditionally well-received formality of uncertainty in terms of the usage of probability calculus by science and engineering.

However, just as Zadeh (1965) pointed out, "It has become increasingly clear that there are some important facets of uncertainty which do not lend themselves to analysis by classical probability-based methods". Fuzziness, which is logically the break down of the law of excluding the middle, is neither well known nor largely ignored in the community of reliability engineering and management.

However, in the real world, fuzziness is sometimes inherent to the object observed, and particularly related to the evolution characteristics. In industrial environments, factors (or covariates) associated with system operation always change and evolve. Quite often we face such a situation where the outcome of a statistical experiment is not a real number, but described in a linguistic manner. Such a kind of fuzziness exists objectively whenever information extracted as a reflection, from the interaction between human brains, and the operating machine systems. Therefore, the existing certainty and fuzzy uncertainty are mutually coherent facets of the objective world.

Once the spatial information is collected, how to extract the useful part and to what extend to summarize it for further analysis will inevitably introduce another kind of fuzziness, subjective fuzziness. This is because information extraction depends heavily upon the experiences and knowledge of the modeller. Even for some events with certainty and clear specifications, the information or image in human brains is fuzzy in formality.

Furthermore, in geostatistical practices, a certain amount of factors and information may be neither random nor fuzzy, which can be called unclear information. The modeller could not grasp the underlying mechanism (facts), and therefore have no clear picture, because of the limitation on information collection or lack of the experiences in it. In other words, modellers have not touched the state of nature and reached any qualification or quantification on it yet. Such purely subjective and unclear information might be manipulated, in terms of subjective membership function, or subjective probability.

Uncertain information can be therefore classified into two categories: strong form and weak form. Information with randomness and fuzziness, which is typically objective uncertainty, belongs to strong form of uncertainty while information with unclearness and vagueness, which is typically subjective uncertainty, belongs to weak form of uncertainty. The weakness of unclearness can be understood in two aspects: firstly whenever unclearness coexists with

randomness and fuzziness it can be shaded or covered by the later; secondly when unclearness exist alone it can be represented by subjective membership function or subjective probability.

It is a commonly accepted principle in science that diversification of the world requires different methodologies. The two forms of uncertainty therefore need different mathematical treatments. It is true that the objectiveness of membership quantifications and usages is still a debatable issue, due to the fact that currently membership functions in applications are largely subjectively specified. However, it is wrong to claim simply the equivalence between membership function in fuzzy mathematics, and the prior distribution in Bayesian statistics.

There exists a duality between membership function and probability mathematically, but they should not be regarded as the same research object. Duality does not mean equivalence. Only in the case of weak form uncertainty, it might be true that there is no difference in nature, between subjective membership function and subjective probability. Therefore it is necessary to reflect both random and fuzzy uncertainties in our modelling efforts with geostatistical information.

There is an increasing awareness that an estimate is of little value in the absence of a measure of the associated uncertainty. This is specially the case of prediction of environmental variables where the prediction uncertainty is required to support spatial modelling. Over the last 20 years, geostatistical methods, like kriging, have been used successfully to investigate the spatial variability of continuously varying environmental variables and to incorporate this information into mapping. However, the kriging variance has often been misused as a measure of reliability of the kriging estimate.

The main limitation of the kriging method is that, when it is used to calculate the probability, it relies on the assumptions of normality of the distribution of prediction errors, and the variance of the errors is independent from the data values (Cressie, 1991). These conditions are rarely met for environmental attributes, which typically display highly skewed histograms. An alternative is to use indicator kriging, to derive, at each unsampled location, the conditional cumulative distribution function (CCDF) which models the uncertainty about the unknown value.

The indicator kriging approach does not rely on an analytical parametric modelling of the shape of the error histogram, hence it is referred to as nonparametric. Furthermore, it can account for measurement errors through a soft indicator coding of observations, which contrasts with most studies on heavy metals where the measurement errors were assumed to be negligible. Also, the CCDFs can be used to analyse how the uncertainty propagates when several variables are

combined. This uncertainty propagation can be conducted numerically by sampling the CCDFs of these variables many times to consider all possible combinations.

Uncertainty assessment is not a goal per se, but it is a preliminary step in the modelling process, such as delineation of hazardous areas. In the process of site characterization and remediation, multistage, or phased, sampling is often conducted so as to validate the result of prior sampling or to improve the cost-effectiveness of a sampling campaign. Phased sampling involves an interruption of the sampling process until the data are available for estimating contaminant concentrations at unsampled locations, which will guide the selection of locations where additional data are needed.

Chen (1998) found that two-stage sampling led to a smaller proportion of locations that were wrongly classified. Different criteria can be used to locate these additional samples. A common approach consists of designing a sampling scheme that minimizes the kriging variance. This approach is very convenient for multistage sampling, because as long as the variogram model is unchanged, the impact of sampling on the kriging variance can be assessed.

Kriging based spatial statistical data analysis methodologies are almost the standard exercises in GIS. However, kriging needs to be calculated from large sample sizes, and this poses a difficulty for environmental data. Environmental data are often sparse and their original purpose was usually not for kriging anlysis, and therefore the data are often inadequate in some way. In statistics, large samples such as 50 might be needed for kriging, and in GIS software, although as little as ten observations can be used for kriging, but the results are not good.

In this thesis, we present an alternative approach, *fuzzy membership grade kriging*, which is based on the analysis of the CCDFs, and so it is better suited to the presence of heteroscedasticity, i.e. the variance of the estimation errors depends on the actual data values. In this way, the sampling method is suited for the specific objective of improving the model instead of improving the accuracy of the prediction itself. The papers present a nonparametric methodology to assess, and combine the uncertainty arising from measurement errors and spatial predictions. We also present a methodology, *Grey GIS*, based on Grey System Theory, for dealing with very small datasets.

2.2 Geostatistics

Geostatistics is a branch of applied statistics developed by George Matheron in the 1970s (Gallagher, 1997). In applied statistical modelling, least squares or linear estimation is the most widely used approach. Matheron advanced an adaptation of such methods that is well suited to the solution of estimation problems, involving quantities that vary in space. This approach is known as the Theory of Regionalized Variables or Geostatistics. It can be used where information is fragmentary, and there is a need to estimate the values of properties at unsampled locations. The geostatistical approach is possible whenever the property of interest behaves as a spatially correlated variable. It treats geographic attributes as mathematical variables that depend on their positions on the Earth's surface (Kitanidis, 1997).

The central tool of geostatistics is the variogram or sometimes referred to as the semivariogram, it is a function which relates half the average squared difference between paired data values to the distance by which they are separated. *Semivariance* is a measure of the degree of spatial dependence between samples. The magnitude of the semivariance between points depends on the distance between the points. A smaller distance results in a smaller semivariance, and a larger distance results in a larger semivariance. The plot of the semivariances as a function of distance from a point is known as a *semivariogram* (Gallagher, 1997).

A mathematical model may be fitted to the variogram and the coefficients of the model may be used to assign optimal weights for interpolation using kriging. *Kriging* is an interpolation procedure used in geostatistics, using known values and a semivariogram to determine unknown values. The procedures involved in kriging incorporate measures of error and uncertainty when determining estimations. Based on the semivariogram used, optimal weights are assigned to unknown values in order to calculate unknown ones. Since the variogram changes with distance, the weights depend on the known sample distribution (Gallagher, 1997).

Matheron (1963) named the kriging method after D.G. Krige, a South African mining engineer. In the 1950s, Krige developed empirical methods for determining the true ore-grade distributions from the distributions based on sampled ore grades. As geostatistics was developing in mining engineering under Matheron in France, the very same ideas developed in meteorology under Gandin (1963) in Soviet Union. The original and simultaneous contribution of these authors was to put optimal linear prediction, in terms of variograms, into a spatial setting. Gandin's name

for this approach is Objective Analysis, and he used the term optimum interpolation instead of kriging (Cressie, 1991).

Geostatistical methods of interpolation, known as kriging, attempt to optimize interpolation by dividing spatial variation into three components: deterministic variation that can be treated as useful soft information, spatially autocorrelated but physically difficult to explain variations, and uncorrelated noise. The character of the spatially correlated variation is encapsulated in functions such as the autovariogram and semivariogram, and these provide the information for optimizing interpolation weights and search radii. Experimental variograms are computed from sample data in one, two, or three spatial dimensions. These experimental data are fitted by one of a limited number of variogram models, which serve to provide data, for computing interpolation weights. Geostatistical methods provide great flexibility for interpolation, providing different methods of interpolation. These methods of interpolation yield smoothly varying surfaces, accompanied by an estimation variance surface. The information in the variogram can be used to optimize sampling schemes, for mapping from point data (Burrough and McDonnell, 1998).

Ordinary Kriging (OK) is a type of kriging which assumes that local means are not necessarily closely related to the population mean, and which therefore uses only the samples in the local neighbourhood for the estimate. It relies on the spatial correlation structure of the data to determine the weighting values, and correlation between data points determines the estimated value at an unsampled point. Usually, ordinary kriging imposes Gaussian distribution assumptions on data structure. Linear kriging is a well established method for predicting variables at unsampled locations. Ordinary kriging is the most widely used variant of kriging. However, when data are sparse, the assumptions made about the underlying variation that has been sampled and the choice of method and its parameters can be critical, if one is to avoid misleading results (Burrough and McDonnell, 1998).

The nonparametric method does not impose distributional assumptions on data. Journel in 1983 prosposed a nonparametric method: indicator variogram and indicator kriging. The non-linear kriging methods provide estimates of the conditional distribution of the target quantity given available information. This non-distributional assumption provide advantages but also weaknesses, such as less efficiency and requiring large data samples. *Indicator Kriging* (IK) is a non-linear form of ordinary kriging in which the original data are transformed form a continuous scale to a binary scale (Burrough and McDonnell, 1998). It relies on the spatial correlation structure of the data to determine the weighting values, and correlation between data points determines the

estimated value at an unsampled point. It is unlike ordinary kriging in that it makes no assumption of normality, and is a nonparametric method, counterpart to the ordinary kriging (Spatial Analysis and Decision Assistance, 1996).

Indicator kriging is the application of kriging to indicator functions of the data. There are no assumptions made about the underlying invariant distribution, and the *0-1* indicator transformations of the data make the predictor robust to outliers. There are some disadvantages as well. Theoretically, indicator kriging requires estimation and modelling of many indicator variograms, and the resulting system of kriging equations is very large. However, indicator kriging has found applications in estimations of recoverable reserves, classification schemes for map analysis, estimation of spatiotemporal distributions of hydrogen-ion deposition, and risk assessment in environmental applications (Cressie, 1991).

In order to analyze spatially distributed imprecise data, it is inevitable to merge kriging and fuzzy mathematics together for creating a different approach to fuzzy kriging. Fuzzy kriging can be traced back as early as Omre's Bayesian kriging paper (1987), and since then various fuzzy kriging methodologies for spatial data directly have being developed, e.g. Bardossy et al. (1988, 1989, 1990a, 1990b), Diamond, (1989), Lee (2000), Bandemer and Gebhardet (2000) etc. It is necessary to point out that these fuzzy kriging approaches, we will call them as *direct fuzzy kriging*, because the prediction is realized by directly kriging on spatially distributed fuzzy observations (i.e., fuzzy numbers). Direct fuzzy kriging is inevitably involved with complicated mathematical operations on fuzzy sets, fuzzy statistical estimation and hypothesis testing.

2.3 Fuzzy Mathematics

There are some common characteristics of environmental data such as, large datasets, heterogeneity, and uncertainty. The fuzzy approach is to present a way to process vague or imprecise environmental data (Salski, 1996). The *fuzzy set theory* was originally introduced by Zadeh in 1965. Fuzzy logic is a type of logic for dealing with imprecise data. It was specifically designed to mathematically represent uncertainty and vagueness in data, and provide a set of tools for dealing with imprecise data intrinsic to many problems.

In fuzzy Logic, elements may have infinite graduation between True and False (Answermath, 2002). Fuzzy logic can be understood as a philosophy based on the fuzzy approach, and as a mathematical methodology (Hajek and Kramosil, 2002). Fuzzy theory holds that all things are

matters of degree, and also reduces black-white logic and mathematics to special limiting cases of grey relationships. Mathematically fuzziness means multi-valence so that multi-valued fuzziness corresponds to degrees of indeterminacy or ambiguity, partial occurrence of events or relations (Pujol, 2001).

Fuzzy set theory implements classes of data with boundaries that are not sharply defined. Fuzzy logic differs from classical Boolean logic in that, it implements soft linguistic variables on a continuous range of true values which allows intermediate values to be defined between conventional binary. The benefit of extending classical set theory methods into fuzzy techniques is the strength in solving real world problems, which inevitably contain some degree of imprecision and noise in the variables and parameters involved (Pacific Northwest National Laboratory, 2001). Additional benefits of fuzzy logic include its simplicity and its flexibility. Fuzzy logic can handle problems with imprecise and incomplete data, and it can model non-linear functions of arbitrary complexity (Surveys and Presentations in Information Systems Engineering, 1996).

Many modelling and problem solving tasks are too complex to be understood quantitatively. However, people can succeed by using knowledge that is imprecise rather than precise. Fuzzy logic resembles human reasoning in its use of approximate information and uncertainty, to generate models (Pacific Northwest National Laboratory, 2001). Fuzzy logic is a powerful problem-solving methodology with many applications embedded in control and information processing. It provides a simple way to draw definite conclusions from vague, ambiguous or imprecise information. Fuzzy logic supports modelling by its ability to takes in approximate data and finds precise solutions (Aptronix, 2000).

Fuzziness and randomness are two distinct types of uncertainty: vagueness and randomness co-exist objectively; in environmental science, imprecise, qualitative, even linguistic data are frequent; modelling efforts should reflect the fuzziness-randomness coexistence reality. The fuzzy approach presents a way to process vague or imprecise environmental data.

Definition of a membership function Zadeh (1965): The membership function of a fuzzy set $\tilde{A}$ of U is a mapping from U onto [0,1]. The fuzzy set is a generalized set:

$$\begin{array}{l} \vartheta_A : U \to \{0,1\} \quad (\textit{crisp set}) \\ \qquad \downarrow \\ \mu_{\tilde{A}} : U \to [0,1] \quad (\textit{fuzzy set}) \end{array} \tag{1}$$

A fuzzy set is characterized by membership function, but not uniquely.

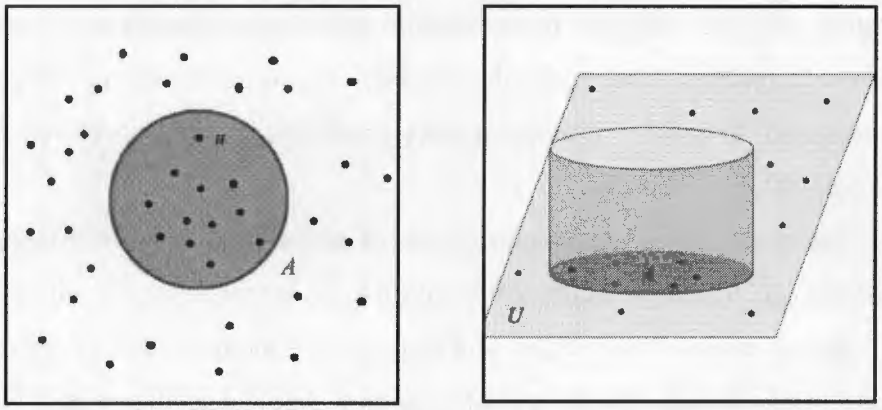


Figure 2.3.1 A Set and Its Indicator Function.

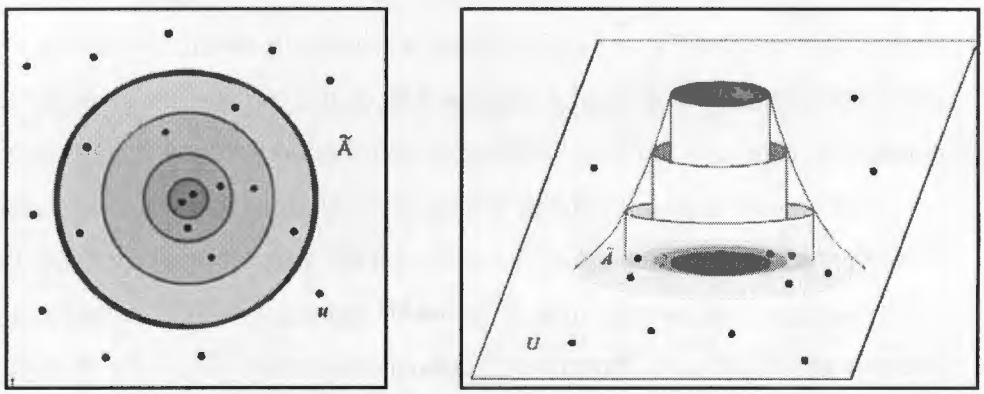


Figure 2.3.2 A Fuzzy Set and Its Membership Function.

In figure 2.3.1, in red areas, sample point u belongs to set A, and in white areas, sample u does not belong to A's indicator function. In figure 2.3.2, in red areas, membership of u=1, and in white areas, membership of u=0. The three pink areas showed grade evolution. Dark pink shows the membership of u to $\tilde{A}$ is high, and the light pink areas shows the membership of u to $\tilde{A}$ is low.

2.4 Grey System Theory

Today, we often deal with the information neither of fully observed (white information) nor of not observed at all (black information). In other words, the information is grey. The Grey System Theory was first proposed by Deng in 1982. The Grey System Theory avoids the inherent problems of conventional statistical methods and only requires a small data set to estimate the behaviour of an uncertain system. The Grey System Theory is best used in relational analysis

model construction; or used in prediction and modelling in situations where there is: uncertainty, multi-data input, discrete data, or insufficient data (Wen, 2004b).

Grey System Theory can be summarized as (Wen, 2004b):

- Grey generating, including: grey relational generating operation (GRGO), accumulated generating operation (AGO), inverse accumulative generating operation (IAGO), interpolation generating.
- Grey Relational Analysis is an impacting measurement model which takes the measurements of relations that change in between two systems or between two elements in the system in time.
- Grey model uses dummy concepts to transform different equations into differential equations. The three types of models are: GM(1,1) model, GM(1,N) model, and GM(0,N) model.
- Grey prediction is based on the GM(1,1) model.
- Grey decision-making by combining other techniques with GM(1,1) model to solve decision problems.
- Grey control use data to find the pattern of system behaviour, then combine with the prediction method in order to make the control system more optimal.

The methodology to solve random uncertainty is probability and statistics, i.e., to treat data in terms of statistical laws and prior (probability) laws. Since statistical laws were established via large samples, the more data we use the better will be the inference. Therefore, straightforward statistical modelling is often difficult to carry on and even not approachable, due to cost consideration for large samples. If we change our attitude and look at the real world from a different angle, say, system dynamics can be treated from the viewpoint of the degree of information availability, we would walk out from the shadow of large sample statistics.

In modern control theory, system dynamics are classified by the degree of information completeness, and accordingly different methodologies have been developed for each of them respectively. Colour is often used to indicate the degree of information availability. If the system information is fully known, the system is called a *white* system; if the system information is totally unknown, it is called a *black* system; and a system with partial information known and partial information unknown is called a *grey* system. Table 2.4.1 offers a general picture of the three systems classified by information criterion (modified from Liu et al., 2004).

Table 2.4.1 Three System Dynamics on Information (in control theory context).

	Dark	*Grey*	*White*
Information	unknown	incomplete	complete
Surface looking	Dark	Indistinguishable	Bright and clear
Process	New	replacement between new and old	Old and known
Property	Chaos and ignorant	Multi-level knowledge	pure
Methodology	negation	sublate	confirmation
Attitude	indulge	tolerant	rigorous
Result	No solution	No unique solution	Unique solution

A critical feature of grey system is information incompleteness. The task of establishing a model under the guidance of grey system theory is inevitably to seek a model based on *small sample data*. Its target is establishment of differential model and emphasizes the exploration, utilization and processing dynamic information containing in data.

It is commonly believed that differential equation is only suitable to continuous differentiable function. The characteristic behavior of a grey system is in general represented by discrete time-series sequences. In order to build a differential equation model, the modeller extracts the relevant variable through relational analysis. And further based on the investigation on the properties of the discrete functions, a dynamic model with differential equation is established on the base of discrete data, i.e., the grey model.

Researchers often believe that time series of a single variable can only provide very limit information. And even some researchers believe there will be high limitations to treat data containing large relational system based on a "1-dimensional" view of point. However, researchers with grey system theory thinking, abbreviated as *grey thinking*, observed that time series sequence contains extremely rich information and it accumulates the traces of all other variables that participate in the system dynamic process. For example, let $\{X_0(t)\}$ be a set taken from a time series of a variable. In real world, it is hard to assume a variable playing its role alone and rather some other p variables, say, $\{Z_k(t); k = 1, 2, \cdots, p\}$ also actively participate in a specific dynamic process. The common exercise in the past was selecting j principal control factors for building a mathematical model while the exercise based on grey system theory is trying to rebuild the dynamic process according the information contained in $\{X_0(t)\}$.

For the purpose of better understanding grey thinking, let us state the two basic principles in grey system theory. Grey concept is the one that its connotation is haze and vague but has a clear

boundary and extension. Therefore, the basic feature of grey system is information-incomplete. Logically, the consequence of the cause information incompleteness is non-uniqueness. The basic principles in grey system are thus derived from:

- Principle of incomplete information. The applications of principle of incomplete information are the dialectical unity of "small" and "large" (amount of information) and also the transformation between "local" and "whole" information.
- Principle of process non-uniqueness. Because the object under investigation in grey system theory does not have complete information, criterion is often multiple folds and it may often a multiple-to-multiple mapping from the causes to effects, therefore the representation of the process is in general not unique. In details, using grey system theory for solving real world problems, it displays often a form of non-unique solutions, non-unique parameter to be identifiable, non-unique model, non-unique decision methodology and non-unique result etc.

For a deeper understanding, we make brief comparisons among the three systems in table 2.4.2 (modified from Liu and Lin, 2006).

Table 2.4.2 Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.

Aspect	*Grey System*	*Probabilistic System*	*Fuzzy Set System*
Set foundation	Haze sets	Cantor sets	Fuzzy sets
Connotation and extension	Connotation haze with clear boundary and extension	Random event with connotation and extension well-defined	cognitive uncertainty (clear connotation but vague extension)
Core concept	Grey derivative and differential equations	Probability distribution	membership function
Data treatment	(inverse) accumulative generating operation	Sampling statistics & asymptotic distribution	Membership grade, λ-cut set & extension principle
Data requirements	Small sample size	Large sample size	empirical (+ sampling data)
Priori	Discrete smoothness	Typical probability distributions	Membership functions (or) possibility distribution
Thinking mode	Multi-angle	Repeated occurrences	Quantification of extension
Final presentation	Imprecise (interval)	$100(1-\alpha)\%$ Confidence interval	λ-level credibility, $0 \leq \lambda \leq 1$

2.5 $(\vee,\cdot)$ Credibility Measure Theory

Credibility Theory is a branch of mathematics which studies the behaviour of the fuzzy phenomena. In order to measure a fuzzy event, Zadeh proposed possibility measure in 1978. Although the possibility measure has been used widely, however, it does not have self-duality property. Therefore, a self-dual measure is needed (Liu, 2006). To define a self-dual measure, Liu and Liu (2002) proposed the concept of *credibility measure*. Liu (2004) has given an axiomatic foundation of credibility theory.

Let Θ be a nonempty set, and $\mathfrak{P}(\Theta)=2^{\Theta}$ the power set on Θ. Each element, let us say, $A\subset\Theta, A\in\mathfrak{P}(\Theta)$ is called an event. A number denoted as $\mathrm{C\tilde{r}}(A)$, $0\le \mathrm{C\tilde{r}}(A)\le 1$, is assigned to event $A\in\mathfrak{P}(\Theta)$, which indicates the credibility that event $A\in\mathfrak{P}(\Theta)$ occurs. $\mathrm{C\tilde{r}}(A)$ satisfies following axioms (Liu, 2004, 2006):

Axiom 1: $\mathrm{C\tilde{r}}(\Theta)=1$.

Axiom 2: $\mathrm{C\tilde{r}}(\cdot)$ non-decreasing, i.e., $\mathrm{C\tilde{r}}\{A\}\le\mathrm{C\tilde{r}}\{B\}$ whenever $A\subset B$.

Axiom 3: $\mathrm{C\tilde{r}}(\cdot)$ is self-dual, i.e., $\mathrm{C\tilde{r}}\{A\}+\mathrm{C\tilde{r}}\{A^c\}=1$ for any $A\in\mathfrak{P}(\Theta)$.

Axiom 4: $\mathrm{C\tilde{r}}\{\bigcup_i A_i\}\wedge 0.5=\sup_i\left[\mathrm{C\tilde{r}}\{A_i\}\right]$ for any $\{A_i\}$ with $\mathrm{C\tilde{r}}(A_i)\le 0.5$.

Axiom 5: Let set functions $\mathrm{C\tilde{r}}_k(\cdot):2^{\Theta_k}\to[0,1]$ satisfy **Axiom 1-4**, and $\Theta=\Theta_1\times\Theta_2\times\cdots\times\Theta_p$, then:

$$\mathrm{C\tilde{r}}(\theta_1,\theta_2,\cdots,\theta_p)=\begin{cases}\dfrac{1}{2}\displaystyle\prod_{k=1}^{p}\left(2\mathrm{C\tilde{r}}(\theta_k)\wedge 1\right) & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}<0.5\\ \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\} & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}\ge 0.5\end{cases} \tag{1}$$

for each $(\theta_1,\theta_2,\cdots,\theta_p)\in\Theta$. In this case, we write $\mathrm{C\tilde{r}}=\mathrm{C\tilde{r}}_1\times\mathrm{C\tilde{r}}_2\times\cdots\times\mathrm{C\tilde{r}}_p$.

Definition 1: Any set function $\mathrm{C\tilde{r}}:\mathfrak{P}(\Theta)\to[0,1]$ satisfies **Axiom 1-5** is called a $(\vee,\cdot)$-credibility measure. The triple $(\Theta,2^{\Theta},\mathrm{C\tilde{r}})$ is called the $(\vee,\cdot)$-credibility measure space.

2.6 Geographic Information System (GIS)

Geographic Information System (GIS) has evolved out of a long history of map making. Like all models, maps are simplified representations of reality. Modern GIS dramatically increase the amount of information that can be contained and manipulated in a map. It could present multiple information themes upon one map (James, 2001). *Geographic Information System (GIS)* is a system of hardware, software, data, people, organizations, and institutional arrangements, for collecting, storing, analyzing, and disseminating information, about areas of the earth (Dueker and Kjerne, 1989).

GIS references real-world spatial data elements into a coordinate system. These features can be separated into different layers (map themes or coverages). GIS stores attribute data, which is descriptive information of the map features. This attribute information is put into a database file separate from the graphics data file but is linked together. A GIS allows the exploration of both the spatial data and the attribute data at the same time. Statistics is a set of mathematical methods used to collect and analyze sample data. The sampled data often has a spatial component, and so it can be incorporated into a GIS (Wisconsin State Cartographer's Office, 2001).

The GIS technology evolved through multiple parallel but separate applications across numerous disciplines. Environmental Systems Research Institute (ESRI) is a leading vendor of GIS software (Environmental Systems Research Institute, 2001). In this thesis, ArcGIS software from ESRI is used in the applications. GIS is more than just a tool, it has consequences for how spatial data are collected, analyzed, and used. When spatial models are used to in modelling, GIS has the potential to influence the outcomes (James, 2001).

GIS is a powerful tool for spatial analysis and representation of data, and it can transform data into valuable information for the modellers (Dyke et al., 2003). Pollution is spatial in nature, and therefore GIS can help in environmental assessment and modelling by, presenting and analysing the air pollution data spatially. It can also provide objective, reliable, and comparable information outputs (Jaroslav, 2003).

2.7 Datasets Used in the Thesis

In this thesis, two datasets are used for applications, PM_{10} air pollution data of California, and soil dioxin data of Midland County in Michigan, U.S.A. These datasets are introduced in this chapter.

2.7.1 Air Pollution: PM_{10} Data

The Environmental Protection Agency (EPA) sets National Ambient Air Quality Standards for pollutants considered harmful to public health and the environment. PM_{10} is one of the seven air pollutants the EPA regulates, under the National Ambient Air Quality Standards. The seven air pollutants include: Carbon Monoxide, Lead, Nitrogen Dioxide, Particulate Matter (PM_{10}), Particulate Matter ($PM_{2.5}$), Ozone, and Sulphur Dioxide (SO_2).

Particulate Matter (PM) is a type of air pollutants that include dust, dirt, soot, smoke, and liquid droplets. Particulates are directly emitted into the air by sources such as factories, power plants, transportation sources, construction activity, fires, and windblown dust. Particulates are also formed in the atmosphere by condensation or transformation of emitted gases such as sulphur dioxide, nitrogen oxides, and volatile organic compounds into tiny droplets (Environmental Protection Agency, 2003).

PM_{10} is particular matter with a mass median aerodynamic diameter less than 10 micrometers. PM_{10} is therefore particulate matter that is very small, and remains suspended in the air for long periods of time. They usually consist of smoke, dirt, and dust particles, as well as mold, spores, and pollen (Rocky Mountain Center, 1996).

PM_{10} is easily inhaled into the deep lung. Exposure to high outdoor PM_{10} concentrations causes increased disease and death. PM_{10} have adverse effects on human heath, because the particles are so small that they can reach the thoracic or lower regions of the respiratory tract. The concerns for human health include effects on breathing and respiratory symptoms, aggravation of existing respiratory and cardiovascular disease, alterations in the body's defence systems against foreign materials, damage to lung tissue, carcinogenesis, and premature death. The people who are most vulnerable to these effects include people with cardiovascular disease, influenza, asthmatics, the elderly, and children.

In this thesis, we used the PM_{10} air pollution data of California, U.S.A., for applications. The current EPA standard for PM_{10} has set an annual allowable arithmetic average of PM_{10} not to exceed 50 $\mu g/m^3$ i.e. micrograms per cubic meter of air (Environmental Protection Agency, 2003). The PM_{10} samples were collected in 111 locations in California (figure 2.7.1).

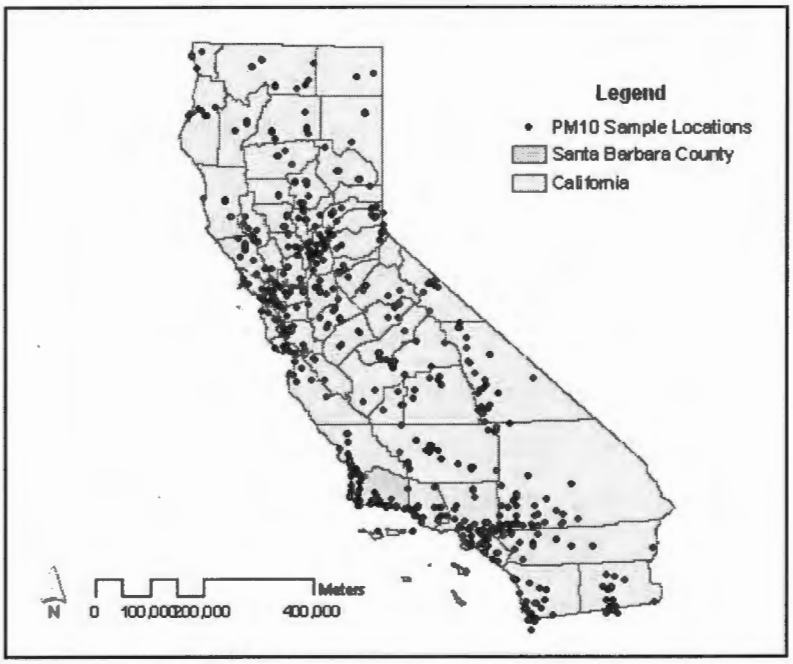


Figure 2.7.1 PM_{10} Sample Locations in California.

The California Air Resources Board has 55 air quality monitoring sites located within the Santa Barbara County (figure 2.7.2). However, very few samples are actually collected from the monitoring sites each year. In 2002, only 4 PM_{10} samples were collected within Santa Barbara County (figure 2.7.3). This is because of the cost and manpower involved in the sampling process.

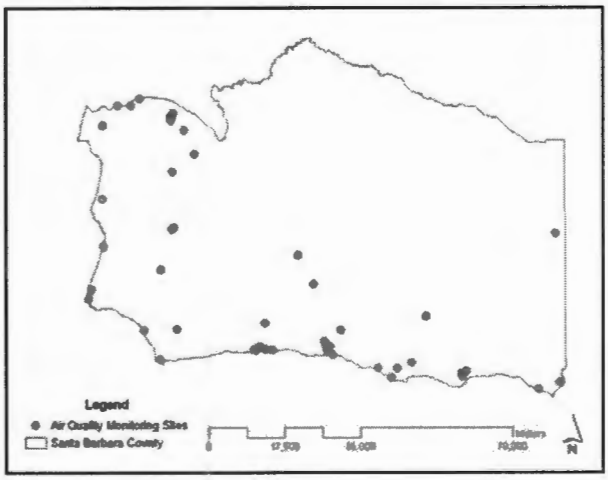


Figure 2.7.2 Air Quality Monitoring Sites in Santa Barbara County, California.

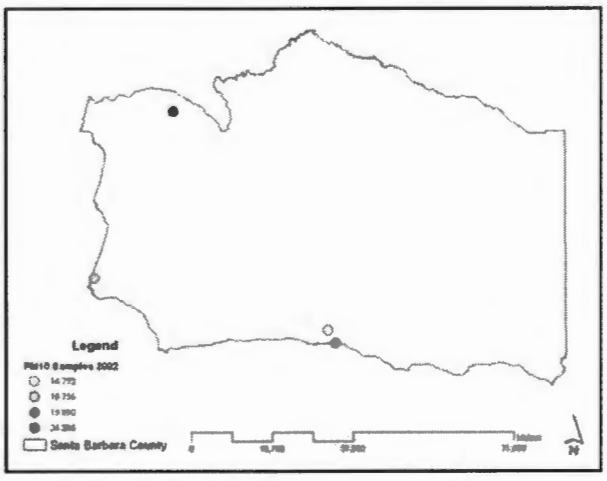


Figure 2.7.3 PM_{10} Sample Locations for 2002 in Santa Barbara County, California.

2.7.2 Soil Pollution: Dioxin Data

Dioxin is a complicated family of chemicals that includes dioxins, furans, and PCBs that have related properties and toxicity. Dioxins are man-made chemical compounds that enter the air through fuel and waste emissions, including motor vehicle exhaust fumes and garbage incineration. It is not deliberately manufactured, but rather an unintended by-product of industrial processes that use or burn chlorine (Alliance for Safe Alternatives, 2005).

Dioxin is one of the most studied chemicals on the planet. It can be found in air, water and soil, as well as in food. Dioxin emissions from incinerators reach people. Dioxin goes into the air and people breathe in the particles. But a bigger problem is that the particles settle on grazing land where cows eat the grass and the dioxin gets concentrated in the fat in their meat and milk. It also gets concentrated in cattle and hogs that are fed dioxin-tainted grain. Dioxins may also be carried in rain and contaminate soil and watersheds. The dioxin particles can fall directly into rivers, streams, and other bodies of water or reach these waterways in surface water runoff. Dioxin settles on the bottom where fish and shellfish ingest small particles of sediment (Alliance for Safe Alternatives, 2005).

People accumulate dioxins in their bodies mostly by eating dioxin contaminated food. Dioxin is stored in human adipose tissue. Dioxin causes a wide range of adverse health effects including cancer, birth defects, diabetes, learning and developmental delays, endometriosis, and immune system abnormalities. It is the most potent animal carcinogen ever tested (Alliance for Safe Alternatives, 2005). Most human exposure to dioxins occurs through the consumption of contaminated foods, especially animal fats. Certain population groups are more at risk, such as nursing babies and people who more animal fat or contaminated foods, because of their nearness to dioxin release sites are exposed to higher levels of dioxin (Environmental Protection Agency, 2005b).

Fortunately, dioxin contamination is getting better controlled. From the available data it seems that the exposure to dioxin has decreased over the last 10 years. However, it is still important for us to monitor dioxin level (Environmental Protection Agency, 2005b).

In this thesis, we used the soil dioxin data of Midland County, Michigan State, U.S.A., for applications. Soil dioxin samples are collected in Midland City and along the Tittabawassee River (figure 2.7.4). It is clear from the figure, that the soil dioxin samples are all concentrated in a small area of the Midland County.

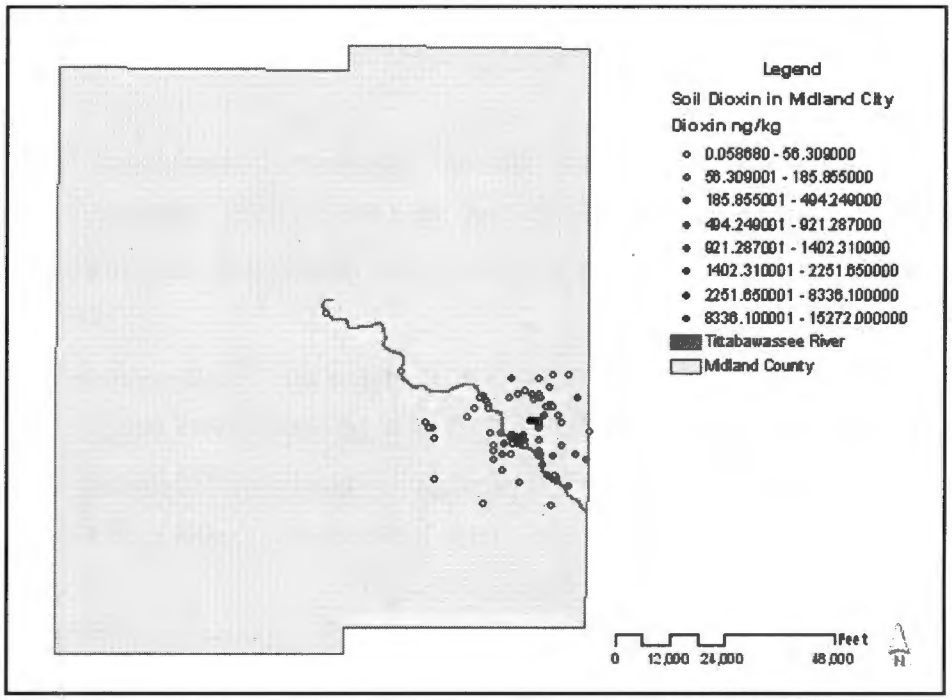


Figure 2.7.4 Soil Dioxin Samples in Midland County, Michigan.

The reported dioxin concentration is referred by "TEQ". Toxic Equivalents, or TEQs, are used to report the toxicity-weighted masses of mixtures of PCDD/Fs (TRIfacts, 2005). The measurement units are in ng/kg (nanograms per kilogram) which is part per trillion (ppt).

Chapter 3. Collection of Papers

This chapter consists of fifteen papers written in contribution to the developments of spatial fuzzy and grey uncertainty modelling. There are a total of three journal papers and twelve conference papers. Fourteen of the peer-reviewed papers were already published, and one paper is under-print.

List of peer-reviewed journals and conference publications, in time sequence of publication:

1. Guo, D., Guo, R., and Thiart, C. (2004). Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM10 for California, Using Fuzzy Kriging. Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2004), University of East Anglia, Norwich, U.K., pp. 375-380.
2. Guo, D., Guo, R., and Thiart, C. (2005). Grey GIS. Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2005), Glasgow, Scotland, pp. 379-387.
3. Guo, D., Guo, R., and Thiart, C. (2005). Grey Spatial Prediction of Fuzzy Air Quality Index. Proceedings of the 4th International Conference on Quality and Reliability (ICQR 2005), Beijing, China, pp. 71-78.
4. Guo, D., Guo, R., and Thiart, C. (2005). A New Very Small-Sample Based Non-Linear Statistical Estimation Method. Proceedings of the International Workshop on Recent Advances Stochastic Operations Research, Canmore, Alberta, Canada, pp. 50-57.
5. Guo, D., Guo, R., Thiart, C., and Cui, Y.H. (2005). Optimal Data Transformations in Grey Analysis. Proceedings of the 10th International Conference on Industrial Engineering – Theory, Applications and Practice, Clearwater, Florida, U.S.A., pp. 426-431.
6. Guo, R. and Guo, D. (2005). Grey Reliability Analysis of Complex System. Proceedings of the 10th International Conference on Industrial Engineering – Theory, Applications and Practice, Clearwater, Florida, U.S.A., pp. 432-437.
7. Guo, D., Guo, R., Thiart, C., Oyana, T., Dai, D., and Hession, S. (2006). GM(1,1)-Kriging Prediction of Soil Dioxin Pattern. Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2006), University of Nottingham, U.K., pp.69-77.

8. Guo, D., Guo, R., and Thiart, C. (2007). Predicting Air Pollution Using Fuzzy Membership Grade Kriging. Editors: Andy P Jones and Iain Lake. Journal of Computers, Environment and Urban Systems, Special Editions, Elsevier, 31(1): 33-51.
9. Guo, D., Guo, R., Thiart, C., and Cui, Y.H. (2006). Rationale of GM(1,1) Modelling – A Variation Approach. Proceedings of the International Conference on Maintenance Engineering (ICME 2006), October 15-18, 2006, Cheng Du, China, pp. 603-699.
10. Guo, D., Guo, R., and Thiart, C. (2006). Generalizations to Standard GM(1,1) Model. Proceedings of the International Conference on Maintenance Engineering (ICME 2006), October 15-18, 2006, Cheng Du, China, pp. 587-593.
11. Guo, R., Guo, D., and Cui, Y.H. (2006). Rationale Exploration of First-Order One Variable Grey Differential Equation Model via Variational Approach. Proceedings of 11th International Conference on Industrial Engineering, Theory, Applications and Practice, October 24-27, 2006, Nagoya, Japan, pp. 1302-1307.
12. Guo, D., Guo, R., and Thiart, C. (2006). The Coupling of Regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models. Editors: Kun-Li Wen and Cha'o-Kuang Chen. Journal of Grey System, 9(2): 143-154.
13. Guo, R., Guo, D., and Li, X. (2006). Bivariate Credibility-Copulas. Proceedings of the 1st International Conference on System Integration and Reliability Improvements (SIRI 2006), December 13-15, 2006, Hanoi, Vietnam, pp. 119-123.
14. Li, X., Guo, D., Guo, R., and Thiart, C. (2006). Kernel Estimation for Bivariate Credibility Copulas. Proceedings of the 1st International Conference on System Integration and Reliability Improvements (SIRI 2006), December 13-15, 2006, Hanoi, Vietnam, pp. 52-56.
15. Guo, D., Guo, R., and Thiart, C. (2007). Credibility Measure-Based Fuzzy Membership Grade Kriging. Editors: B.D. Liu. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 15. (*under-print*)

3.1 Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM_{10} for California, Using Fuzzy Kriging

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2004), April 28-30, 2004, University of East Anglia, Norwich, U.K., pp. 375-380.

3.1.1 Introduction

PM_{10} is one of the seven air pollutants the Environmental Protection Agency regulates, and exposure to high outdoor PM_{10} concentrations causes increased disease and death (Environmental Protection Agency, 2003). PM_{10} concentrations are sampled and measured in various places in California, U.S.A. (figure 3.1.1), however, it is too costly in terms of time, finance, and manpower to sample the entire state. As air pollution is spatial in nature and the measurement and the degree of pollution is fuzzy (the boundaries between negligible, moderate and severe is not clear cut), Geographic information systems (GIS) and the fuzzy approach to geostatistics can help in environmental assessment and modelling. The fuzzy approach (Zadeh, 1965) presents a way to process vague or imprecise environmental data, and geostatistics can be used where information is fragmentary, and there is a need to predict the values of properties at unsampled locations (Vega, 2002).

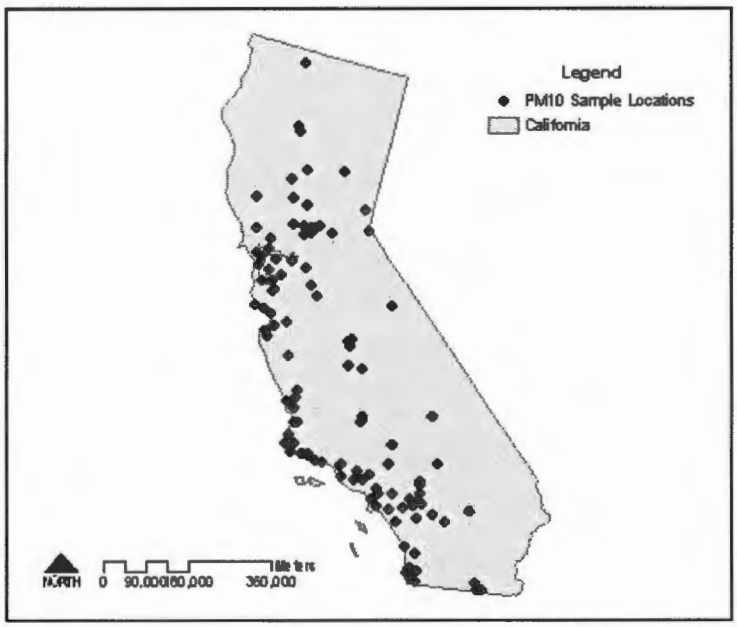


Figure 3.1.1 PM_{10} Sample Locations in California.

In order to analyze spatially distributed imprecise data, it is inevitable to merge kriging and fuzzy mathematics together for creating a different approach to fuzzy kriging. Fuzzy kriging can be traced back as early as Omre's Bayesian kriging paper (1987), and since then various fuzzy kriging methodologies for spatial data directly have being developed, e.g. Bardossy et al. (1988,

1989, 1990a, 1990b), Diamond, (1989), Lee (2000), Bandemer and Gebhardet (2000) etc. It is necessary to point out that these fuzzy kriging approaches, the prediction is realized by directly kriging on spatially distributed fuzzy observations (i.e., fuzzy numbers). These fuzzy kriging is inevitably involved with complicated mathematical operations on fuzzy sets, fuzzy statistical estimation and hypothesis testing, and are unsuitable for combining with GIS.

The aim of this paper is to integrate GIS with fuzzy logic and geostatistics, in order to predict PM_{10} concentrations for the entire California State based on the existing sample data. A methodology of fuzzy variogram and fuzzy kriging is proposed based on the indicator vaiogram and kriging developments in the literature, and it is easy to implement into GIS. Furthermore, unlike previous studies that used assumed membership functions, the sample membership functions are extracted from the data itself. The predicted membership grades are also transformed back into PM_{10} concentrations by using inverse functions. Using the converted prediction maps, we could clearly identify areas that are higher than the threshold, which are dangerous to human health.

3.1.2 Fuzzy Variogram

Kriging is an estimation procedure used in geostatistics, using known values and a semivariogram to determine unknown values (Krige, 1951). The procedure incorporates measures of error and uncertainty when determining estimation (Gallagher, 1997). Classical kriging estimator is a linear predictor based on the spatial observations directly, while the fuzzy membership grade kriging estimator we developed is a linear predictor based on the membership grades of spatial observations. The fuzzy kriging approach proposed in this paper is a direct extension to Journel's (1983) indicator kriging.

The experimental indicator semivariogram is defined as:

$$\hat{\gamma}_{\lambda}(\mathrm{h}) = \frac{1}{n_{\lambda}(\mathrm{h})} \sum_{i=1}^{n(\mathrm{h})} \left[\delta_{\lambda}(\mathrm{s}+\mathrm{h}) - \delta_{\lambda}(\mathrm{s})\right]^2 \tag{1}$$

Correspondingly, the fuzzy semivariogram, for fuzzy event $\tilde{A}$ is defined as:

$$\gamma_{\tilde{A}}(\mathrm{h}) = \int_{R^d} \left[\mu_{\tilde{A}}(\mathrm{s}+\mathrm{h}) - \mu_{\tilde{A}}(\mathrm{s})\right]^2 dP \tag{2}$$

Therefore the experimental fuzzy variogram is defined as:

$$\gamma_{\tilde{A}}(\mathrm{h}) = \frac{1}{n(\mathrm{h})} \sum_{j=1}^{n(\mathrm{h})} \left[\mu_{\tilde{A}}(\mathrm{s}_j + \mathrm{h}) - \mu_{\tilde{A}}(\mathrm{s}_j) \right]^2 \tag{3}$$

3.1.3 Computations and Analysis

The data preparations for the fuzzy variogram and kriging are very essential to the fuzzy geostatistical modelling. Therefore, similar to the indicator kriging, the fuzzy kriging can be done by simply replacing the spatial $z(\mathbf{s}_j), j=1,\ldots,m$ observations by the corresponding membership grades $\mu_{\tilde{A}}(\mathrm{s}_j + \mathrm{h}), j=1,\ldots,m$ respectively.

The modeller's interest is in whether living in California is safe or not, in terms of air pollution. Therefore fuzzy event $\tilde{A}$={High hazard level of PM_{10} in California} is worthwhile to try to investigate. It is meaningful to argue that the higher of the PM_{10} content in the air, the higher the degree of belongingness to fuzzy set $\tilde{A}$ should be. It is reasonable to assume that the linear sample membership function of the fuzzy event $\tilde{A}$ is linear. It has the feature of PM_{10} level at x=9.0 being assigned a 0 membership, and at x=90.3 being assigned a 1 membership. That is:

$$\hat{\mu}_{\tilde{A}}(x) = \begin{cases} 0 & \text{if} \quad 0 \le x < 9.0 \\ \dfrac{x-9}{81.3} & \text{if} \quad 9.0 \le x < 90.3 \\ 1 & \text{if} \quad x \ge 90.3 \end{cases} \tag{4}$$

The linear sample membership function $\mu'_{\tilde{A}}(x)$ (figure 3.1.2) is in general simple and mathematically easily manipulated.

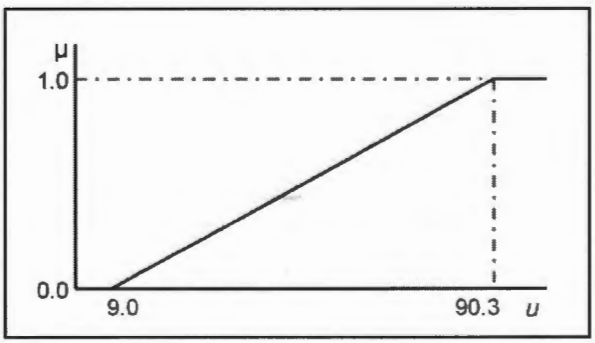


Figure 3.1.2 Linear Sample Membership Function of PM_{10}.

Using the linear sample membership function, a set of new membership values is calculated and then the Geostatistical Analyst Extension of ArcGIS is used to perform ordinary kriging on the new fuzzy data set. The rational quadratic model is used to fit the variogram model (figure 3.1.3).

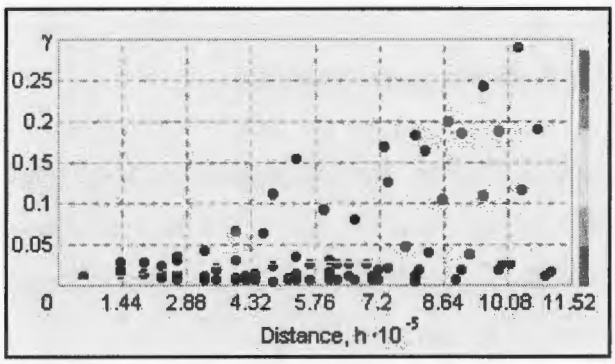


Figure 3.1.3 Fuzzy Variogram of Linear Sample Membership Function.

The QQ plot (figure 3.1.4) shows how well the model fits. A good fit is when the sample quantile distribution is very similar to the normal quantile line. In this plot, the middle sample values (dots), are very similar to the dash straight line. However, the extreme upper and lower values do not fit well.

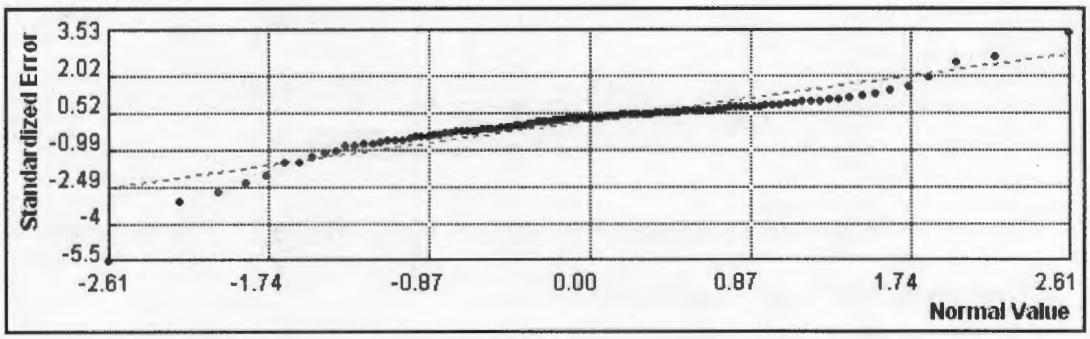


Figure 3.1.4 Standardized Error Plot of Linear Sample Membership Function.

The prediction map (figure 3.1.5) shows the predicted membership grades. The dark brown colours represent higher membership grades, near 1, which are of high PM_{10} concentrations. The light yellow areas represent low membership grades, near 0, which are of low PM_{10} concentrations. Therefore, the three dark brown areas on the lower half of California, show high PM_{10} concentrations. The upper areas of California are light yellow, and it shows areas of low PM_{10} concentrations.

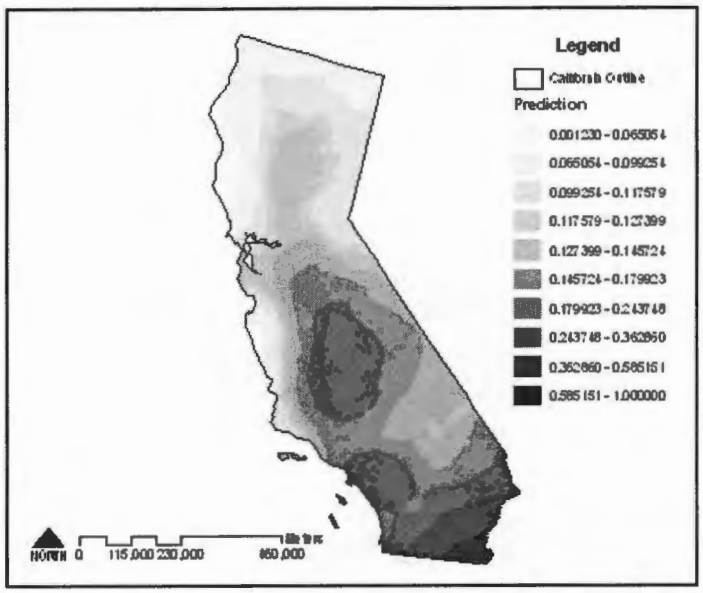


Figure 3.1.5 Prediction Map of Linear Sample Membership Function.

The predicted error map (figure 3.1.6) shows the standard error (uncertainty of the predicted value). The light yellow areas are areas of low error, and the dark brown areas are areas of high error. When one compares this map to the map of sample locations, one can understand the errors much better. Since most of the sample locations are near the coast, it is obvious that the errors are less in the sampled locations, and more errors in the unsampled locations.

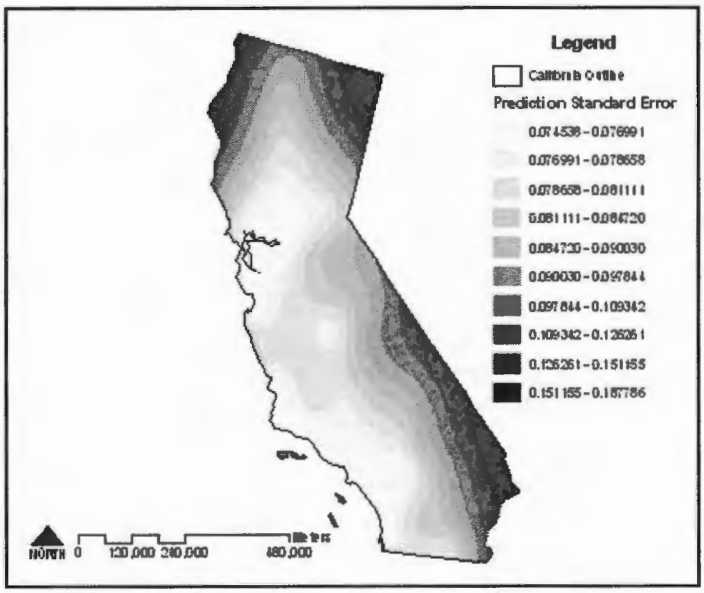


Figure 3.1.6 Prediction Standard Error Map of Linear Sample Membership Function.

The fuzzy membership is defined in terms of PM_{10} hazard levels. The form of membership is monotone-increasing continuous function, and so the inverse function exists.

The linear sample membership function takes the form:

$$\mu(x) = \frac{x-a}{b-a}, \quad a \le x \le b \tag{5}$$

For any given membership value $\mu_0 \in [0,1]$:

$$\mu_0 = \frac{x-a}{b-a} \tag{6}$$

Then:

$$x^{linear}(\mu_0) = \mu_0(b-a) + a \tag{7}$$

Conversion based on the membership prediction map class interval limits. In the prediction map, it is typically classified into ten classes, say, class 0, class 1, … class 9. Assume the class interval limits are $[\mu_{l_0}, \mu_{u_0}), [\mu_{l_1}, \mu_{u_1}), \ldots, [\mu_{l_9}, \mu_{u_9})$.

The class limits for the PM_{10} predicted value with linear sample membership function is:

$$[x_l, x_u) = \left[x^{linear}(\mu_l), x^{linear}(\mu_u)\right) \tag{8}$$

This makes it easy to convert the linear sample membership function back to the original units of measurement, PM_{10} concentration $\mu g/m^3$. These values are given in table 3.1.1 and the map using the concentration values as the legend is given in figure 3.1.7.

Table 3.1.1 Conversion from Linear Membership to PM_{10} Concentrations.

Class	Linear Membership	PM_{10} ($\mu g/m^3$)
0	0.001230 - 0.065054	09.1 - 14.3
1	0.065054 - 0.099254	14.3 - 17.1
2	0.099254 - 0.117579	17.1 - 18.6
3	0.117579 - 0.127399	18.6 - 19.4
4	0.127399 - 0.145724	19.3 - 20.8
5	0.145724 - 0.179923	20.8 - 23.6
6	0.179923 - 0.243748	23.6 - 28.8
7	0.243748 - 0.362860	28.8 - 38.5
8	0.362860 - 0.585151	38.5 - 56.6
9	0.585151 - 1.000000	56.4 - 90.3

The current EPA standard for PM_{10} has set an annual allowable arithmetic average of PM_{10} not to exceed 50 $\mu g/m^3$. However, we defined the hazard level differently in this case. Once PM_{10} concentrations are over 38 $\mu g/m^3$, it is classified a hazard area.

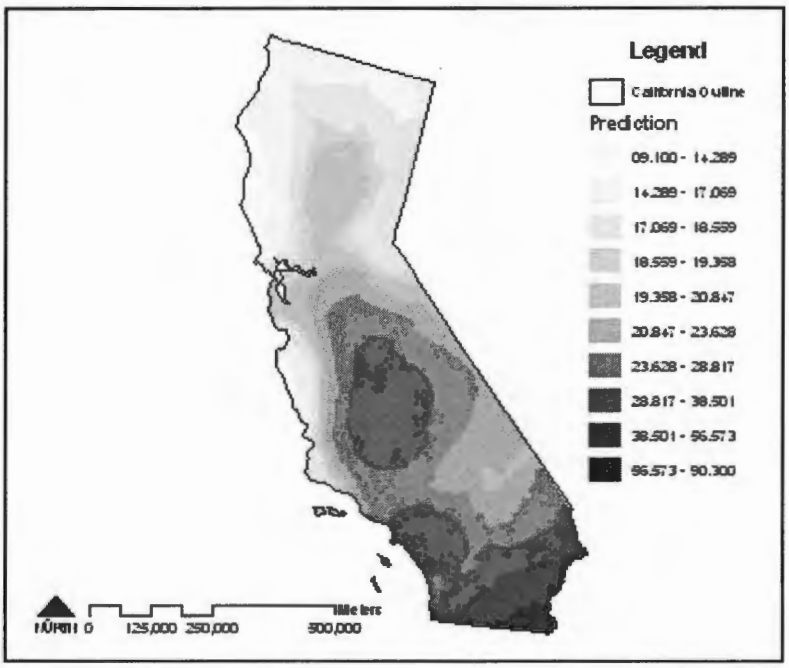


Figure 3.1.7 PM_{10} Concentrations Predicted by Linear Membership Function.

In figure 3.1.7, the last two classes, very dark brown colours, show areas of high levels of PM_{10} concentrations. The level of safeness regarding PM_{10} concentrations is shown clearly here, by the class divisions of concentration level. The upper area of California and parts of the coastline are shown to have very low PM_{10} concentrations, and are the safest areas. The linear membership show much finer divisions of areas of low PM_{10} concentrations.

3.1.4 Conclusion

In this paper, a theoretical integration of fuzzy logic and geostatistics is explored, named fuzzy variogram and fuzzy kriging. The fuzzy kriging is developed as a natural extension of indicator kriging, and then ordinary kriging is performed on the membership results. This way fuzzy kriging would have the advantages of both indicator and ordinary kriging. A new set of methodology is developed for this paper.

The fuzzy methodology developed here is different from previous work on GIS and fuzzy. Unlike previous studies that used assumed membership functions, in this paper, a sample membership function is extracted from the data itself. Therefore, the fuzzy methodology used in this paper is more solid and objective. In this paper, we developed a set of methodology for fuzzy logic and geostatistics to be integrated together in GIS.

3.2 Grey GIS

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2005), April 6-8, 2005, Glasgow, Scotland, U.K., pp. 379-387.

3.2.1 Introduction

Environmental data is very costly and difficult to collect, and quite often the sampled data are insufficient for any practical analysis. In this paper, we propose a small-sample based spatial data analysis methodology in terms of Grey System Theory, and we will call it *Grey Spatial Analysis.* The new methodology can be implemented and combined with Geographic Information Systems (GIS), and we will call it as *Grey GIS.* Grey GIS is a system dealing with grey geographical data or analyzing spatial data in terms of Grey System Theory. As an illustration, we will use PM_{10} data of Santa Barbara County, California, U.S.A.

Kriging based spatial statistical data analysis methodologies are almost the standard exercises in GIS. However, three fatal problems are always appearing as intrinsic companions tied with kriging methodologies. The first issue is sample size, the second problem is the isotropy, and the third one is the spatially stationary assumptions. Obviously the root cause of these problems is in kriging methodology which came from the traditional statistical thinking which is built under probabilistic theory and statistical inference and estimation theory (Guo, 2005a, 2005b).

Facing the challenge of data shortage, we will walk out from the umbrella of large-sample based statistical theory and develop a small-sample based spatial data analysis methodology based on the spirit of Grey System Theory (Deng, 1982). The newly established Grey Spatial Analysis methodology is particularly suitable for small area spatial predictions with a few observations. The Grey Spatial Analysis results are imprecise (grey numbers) in nature. However, practically the grey predictions are unexpectedly consistent with local sample information available.

3.2.2 PM_{10} Sample in Santa Barbara County

PM_{10} is one of the seven air pollutants the Environmental Protection Agency (EPA) regulates, and exposure to high outdoor PM_{10} concentrations causes increased disease and death (Environmental Protection Agency, 2003). The California Air Resources Board has 55 air quality monitoring sites located within the Santa Barbara County (figure 3.2.1).

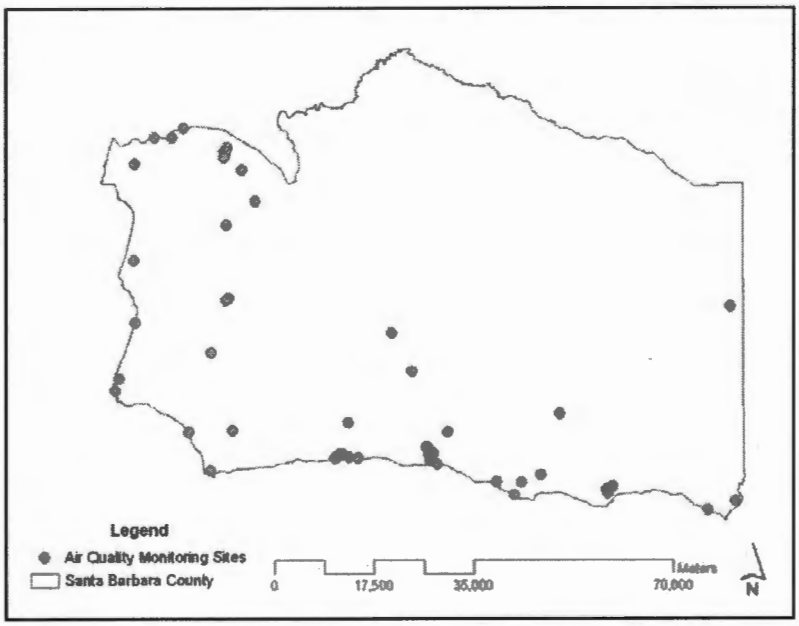


Figure 3.2.1 Air Quality Monitoring Sites in Santa Barbara County.

However, very few samples are actually collected from the monitoring sites each year. In 2002, only 4 PM_{10} samples were collected within Santa Barbara County (figure 3.2.2). This is because of the cost and manpower involved in the sampling process.

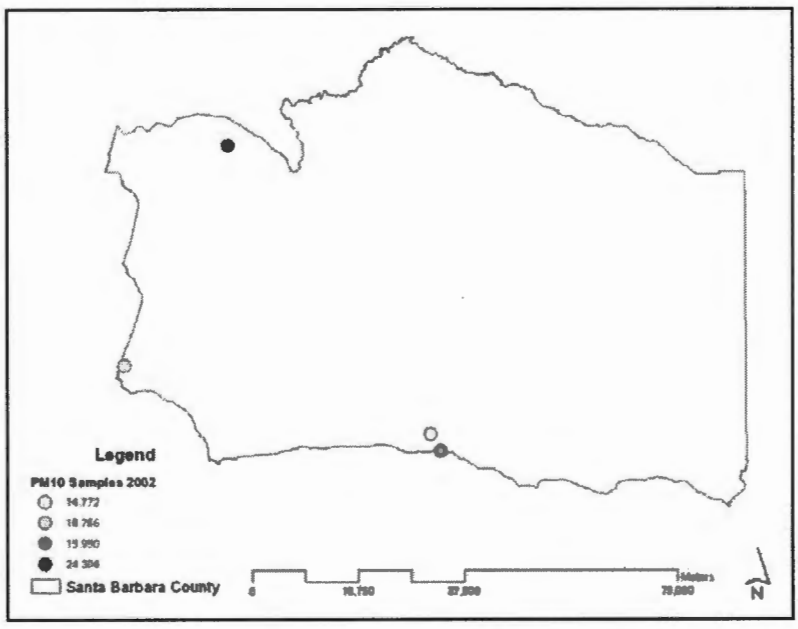


Figure 3.2.2 PM_{10} Sample Locations for 2002 in Santa Barbara County.

What can we do with *4* sample points? In GIS software, kriging prediction maps can only be made under the condition there must be *10* or more sample points (Environmental Systems Research Institute, 2001). Using Grey Spatial Analysis, we will be able to estimate the PM_{10} concentration in different locations within Santa Barbara. For convenience, in this case, we estimated the PM_{10} concentrations in the other 51 monitoring sites within Santa Barbara.

3.2.3 Grey System Theory

A system with partial information known and partial information unknown is called a *grey* system. In control theory, colour is often used to indicate the degree of information availability. For a given cognitive level, if the system information is fully known, the system is called a *white* system; if the system information is totally unknown, it is called a *black* system; accordingly any system lies between them is called a grey one (Fu, 1991).

Grey System Theory, first proposed by Deng in 1982, is a theory studying the analysis, model-establishment, prediction, decision and control. It avoids the inherent defects of conventional, large-sample statistical methods, and only requires a limited amount of discrete data sequence to estimate the behaviour of a system with insufficient information (Wen, 2004a).

The critical task of grey system theory is the establishment of grey differential equation utilizing small-sample discrete data sequence and emphasizes the exploration, utilization and processing dynamic information containing in the data sequence (Wen, 2004a).

3.2.4 Grey Spatial Analysis

3.2.4.1 Discrete Data Sequence Creation

Grey system theory intrinsically deals with discrete data sequence having natural order in either temporal or spatial form (Fu, 1991). However, the problem here is that the spatial data $\{Z(x,y), x, y \in \mathbb{R}\}$ has no order at all. Therefore, the fundamental idea in grey spatial prediction is to convert the order-less spatially distributed data into ordered data sequences.

The conversion is following the following four steps:

(i) Pick up a pair of coordinate (x_0, y_0) as the origin to establish a Cartesian 2-dimensional (localized) coordinate system $\{(x', y'), x' = x - x_0, y' = y - y_0\}$. (x_0, y_0) should be located at

the reasonably nearest lower-left corner of Santa Barbara County. In Santa Barbara County case, $(x_0, y_0) = (-120.7000, 34.3000)$.

(ii) Project $\{Z(x,y), x, y \in \mathbb{R}\}$ along x'-axis and y'-axis respectively. Denote them as data set $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and data set $\{Z_y(y'), y' \in \mathbb{R}^+\}$ where

$$\begin{aligned} Z_x(x') &= Z(x,y) \times x' \Big/ \sqrt{(x')^2 + (y')^2} \\ Z_y(y') &= Z(x,y) \times y' \Big/ \sqrt{(x')^2 + (y')^2} \end{aligned} \tag{1}$$

(iii) Order the numbers in data sets $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y(y'), y' \in \mathbb{R}^+\}$ according to the sequence of coordinate x' and the sequence of coordinate y' respectively and obtain two data ordered sequences, denoted as $\{Z_x^*(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y^*(y'), y' \in \mathbb{R}^+\}$ respectively.

(iv) Tabulate the two projection data sequences:

Table 3.2.1 Projected Santa Barbara County PM_{10} Data Sequences and Coordinates.

x'-coordinate	0.067300 (x_2)	0.264500 (x_4)	0.654200 (x_3)	0.674200 (x_1)
x-projection sequence	4.156992393	9.246591929	14.18756366	19.43545485
y'-coordinate	0.162200 (y_1)	0.189700 (y_3)	0.296100 (y_2)	0.643000 (y_4)
y-projection sequence	4.675809518	4.114003097	18.28953117	22.47848246

Notice that original coordinate values (x_i, y_i), $i = 1,2,3,4$ are used to keep track of the original data compositions for later prediction purpose.

3.2.4.2 Establishment of Grey Spatial Prediction Equation

It is observed that the distance $x_i' - x_{i-1}'$ $(y_i' - y_{i-1}') \neq$ constant. Therefore, the way to obtain the (grey) response equation is different from that with equal-distance spaced grey differential equation. The methodology here is utilizing two-stage least-square estimation procedure (Fu, 1991).

Let the s-coordinate sequence $Z_s^{(0)} = \{Z_s^{(0)}(s_i),\ s_i \in \mathbb{R}^+,\ i = 1,2,\cdots,N\}$ satisfying the following grey differential equations respectively,

$$Z_s^{(0)}(s_i)+\beta_s\overline{Z}_s^{(1)}(s_i)=\alpha_s,\ i=2,\cdots,N \tag{2}$$

where

$$\overline{Z}_s^{(1)}(s_i)=\frac{1}{2}\left(Z_s^{(0)}(s_i)+Z_s^{(0)}(s_{i-1})\right),\ i=2,\cdots,N \tag{3}$$

Accordingly, the first-stage least-square parameter estimators are

$$\left(\hat{\alpha}_s,\hat{\beta}_s\right)=\left(\mathrm{X}^T\mathrm{X}\right)^{-1}\mathrm{X}^T\mathrm{y}_N^s \tag{4}$$

where

$$\mathrm{X}=\begin{bmatrix} 1 & -\frac{1}{2}\left(Z_s^{(0)}(s_2)+Z_s^{(0)}(s_1)\right) \\ 1 & -\frac{1}{2}\left(Z_s^{(0)}(s_3)+Z_s^{(0)}(s_2)\right) \\ \vdots & \vdots \\ 1 & -\frac{1}{2}\left(Z_s^{(0)}(s_N)+Z_s^{(0)}(s_{N-1})\right) \end{bmatrix} \tag{5}$$

And the approximate derivatives are:

$$y_N^s=\begin{bmatrix} \dfrac{Z_s^{(0)}(s_2)-Z_s^{(0)}(s_1)}{s_2-s_1} \\ \dfrac{Z_s^{(0)}(s_3)-Z_s^{(0)}(s_2)}{s_3-s_2} \\ \vdots \\ \dfrac{Z_s^{(0)}(s_N)-Z_s^{(0)}(s_{N-1})}{s_N-s_{N-1}} \end{bmatrix} \tag{6}$$

Thus the first-stage response equation is:

$$Z_s^{(0)}(s)=\alpha_s+\gamma_s\exp\left(-\hat{\beta}_s s\right) \tag{7}$$

Then we enter the second-stage least-square estimation,

$$\left(\hat{\alpha}_s,\hat{\gamma}_s\right)^T=\left(\mathrm{D}_s{}^T\mathrm{D}_s\right)^{-1}\mathrm{D}_s{}^T\mathrm{z}_N^s \tag{8}$$

where

$$D_s = \begin{bmatrix} 1 & \exp\left(-\hat{\beta}_s s_1\right) \\ 1 & \exp\left(-\hat{\beta}_s s_2\right) \\ \vdots & \vdots \\ 1 & \exp\left(-\hat{\beta}_s s_N\right) \end{bmatrix} \tag{9}$$

and

$$z_N^s = \begin{bmatrix} Z_s^{(0)}(s_1) \\ Z_x^{(0)}(s_2) \\ \vdots \\ Z_s^{(0)}(s_N) \end{bmatrix} \tag{10}$$

Therefore after two-stage least-square fitting, the estimated response function is:

$$\hat{Z}_s^{(0)}(s) = \hat{\alpha}_s + \hat{\gamma}_s \exp\left(-\hat{\beta}_s s\right) \tag{11}$$

Then for $\forall(x, y)$, the grey spatial prediction equation will be given by:

$$\hat{Z}(x, y) = \sqrt{\left(\hat{\alpha}_x + \hat{\gamma}_x e^{-\hat{\beta}_x x}\right)^2 + \left(\hat{\alpha}_y + \hat{\gamma}_y e^{-\hat{\beta}_y y}\right)^2} \tag{12}$$

3.2.5 Grey Prediction and Comparison to Kriging Prediction Map

Using Grey Spatial Analysis, the PM_{10} concentrations are estimated for the other 51 monitoring sites, from the original 4 PM_{10} samples, and 34 monitoring sites can be estimated, as can be seen in figure 3.2.3. This is because the grey system theory method is good for interpolation, and not very suitable for extrapolation.

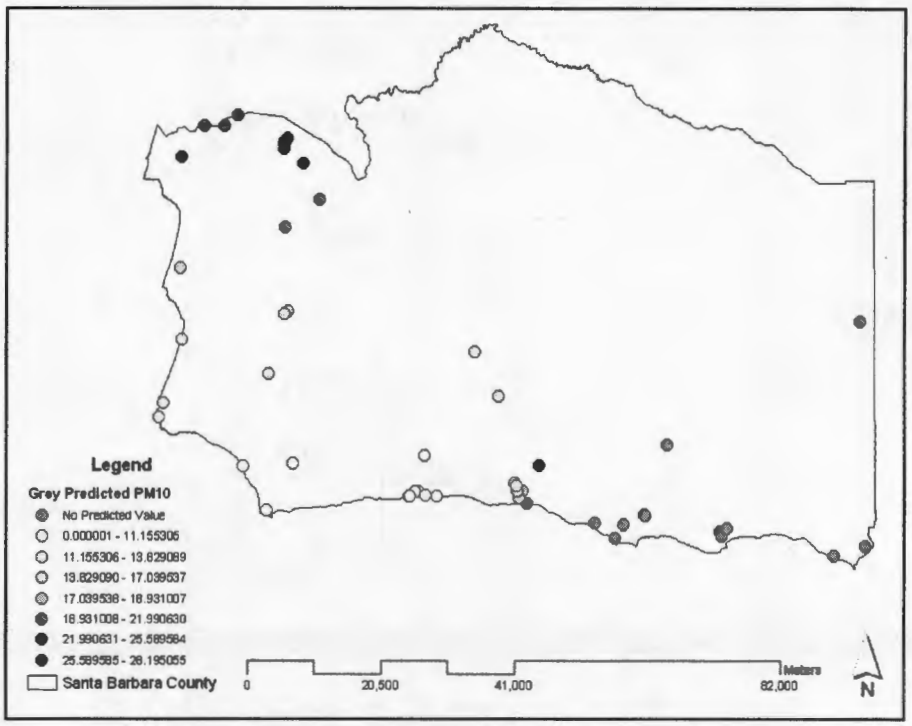


Figure 3.2.3 Grey Prediction for PM_{10} in 2002 in Santa Barbara County.

The Grey spatial predictions are unexpectedly consistent with the original 4 PM_{10} samples and reveal fairly detailed spatial distribution pattern. The monitoring sites with high PM_{10} concentration can be clearly seen, which is on the upper regions of Santa Barbara. As one can see from the figure 3.2.3, in places that the Grey Spatial Analysis can not predict well, it is reflected very clearly in the Grey prediction maps, and in this case shown in brown dots (17 sites). This is because, these "unpredictable" locations are far from the area covered by the 4 sampled locations and the Grey Spatial Analysis is in nature good in interpolations but not good in extrapolations when the predicted location "far away" from sample covered area. However, this problem can be solved, using our newly developed Grey Temporal-Spatial Analysis, which is explained in appendix A.

As a source of reference, an Ordinary Kriging prediction map is generated from 148 sample points from the entire California State. This is because Kriging prediction maps can only be made under the condition there must be 10 or more sample points, and therefore a kriging prediction map can not be produced here, so 148 samples from the entire California is used in order to generate this map. Kriging prediction map of the Santa Barbara County is shown in figure 3.2.4.

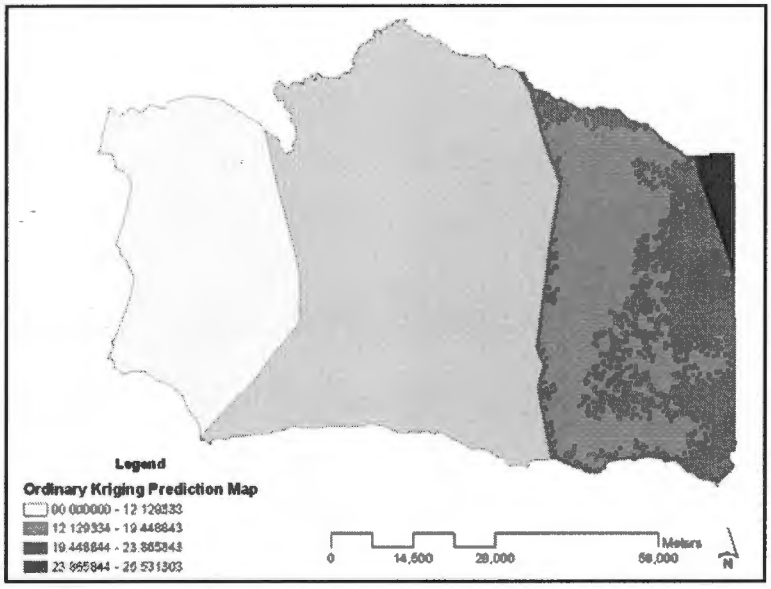


Figure 3.2.4 Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.

While kriging maps produced from large samples are good at predicting an overall trend for the state, it is however, not so good at local PM_{10} pattern predictions. The Grey PM_{10} predictions are shown to be very sensitive to local observations, and can show the approximate PM_{10} concentrations in 34 sites in Santa Barbara, based on only 4 sample data. In places that the Grey Spatial Analysis can not predict well, it is reflected very clearly in the Grey prediction maps.

3.2.6 Conclusion

In this paper, we have developed a small-sample based spatial data analysis methodology based on Grey System Theory, called G*rey Spatial Analysis*. The newly established Grey Spatial Analysis methodology is implemented and combined with GIS, and the grey prediction results have shown to be practical and reflect well the local reality.

3.3 Grey Spatial Prediction of Fuzzy Air Quality Index

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the 4th International Conference on Quality and Reliability (ICQR 2005), August 9-11, 2005, Beijing, China, pp. 71-78.

3.3.1 Introduction

Air quality is a vague concept and its hazard impacts on health are in an evolving manner. Thus air quality should be specified by a fuzzy event which is specified by its membership function. It is obvious that an overall air quality fuzzy index would give the modeller a convenient and simple description on air quality. A membership function of a fuzzy event, say, $\tilde{Q}$, denoted by $\mu_{\tilde{Q}}(x)$ is specified subjectively. Guo et al. (2004) proposed a semi-statistical approach for specifying membership on the base of percentiles from sample data. Jiang and Chen (2006) investigate an approach to define the membership function from probability density function by a transformation in the case of monotonic type of fuzzy event. It is noticed that any individual air quality index is related to the measurement of the air pollutant and it is logical to specify the individual air quality index by monotonic fuzzy membership function. We observe that an estimated probability density function can be obtained statistically in terms of historical massive air pollutant data collections. Table 3.3.1 gives the fitted probability density distributions based on year 2002 data collection from California State, U.S.A.

Table 3.3.1 Fitted Probability Distributions from California State Data (2002).

Pollutant	*Distribution*	*Scale parameter α*	*Shape parameter β*	*Threshold of hazard level* x_ε
Carbon Monoxide	gamma	0.8800347	3.376373	9 ppm
Lead	gamma	0.0267839	3.903671	1.5 $\mu g/m^3$
Nitrogen Dioxide	gamma	0.0056164	2.879439	0.053 ppm
Particulate Matter (PM_{10})	gamma	21.6422190	1.539542	50 $\mu g/m^3$
Particulate Matter ($PM_{2.5}$)	gamma	2.6047873	6.297649	15.0 $\mu g/m^3$
Ozone	gamma	0.0055217	16.131398	0.08 ppm
Sulfur Oxides	gamma	0.0009793	1.613868	0.03 ppm

The probability density function for a gamma random variable is:

$$f(x)=\frac{1}{\alpha\Gamma(\beta)}\left(\frac{x}{\alpha}\right)^{\beta-1}\exp\left(-\frac{x}{\alpha}\right),\ x>0 \tag{1}$$

Then the membership function for the hazard level from any individual air pollutant is:

$$\mu_{\tilde{M}}(x)=\begin{cases}0 & x\le 0\\ \int_0^x f_{M_*}(m_*)dm_* & 0<x<x_\varepsilon\\ 1 & x\ge x_\varepsilon\end{cases} \tag{2}$$

where

$$f_{M_*}(m_*)=\frac{1}{\alpha\Gamma(\beta)}\left(\frac{-m_*}{\alpha}\right)^{\beta-1}\exp\left(\frac{m_*}{\alpha}\right),\ m_*\in(-\infty,x_\varepsilon] \tag{3}$$

which implies $m_*=-x+x_\varepsilon\in(-\infty,x_\varepsilon]$.

As to the overall fuzzy air quality index, it can be created in terms of fuzzy logical function to reflect all (important) seven air pollutants.

$$\mu_{\tilde{Q}_o}(u)=\bigvee_{i=1}^{7}\mu_{\tilde{Q}_i}(u_i) \tag{4}$$

where $\tilde{Q}_o$ and $\tilde{Q}_i$, $i=1,2,\cdots,7$ are overall air quality index and the seven individual air quality index respectively.

It is obvious that finding the fuzzy overall air quality index is reduced into a problem to predict the spatial distribution of individual air quality index and thus the spatial distribution of an individual pollutant.

In spatial data analysis, kriging based spatial statistical data analysis methodologies are the standard application, and used widely for predictions. However, three fatal problems are always appearing as intrinsic companions tied with kriging methodologies. The first issue is sample size, the second problem is the isotropy, and the third one is the spatially stationary assumptions. Obviously the root cause of these problems is because kriging methodology came from the traditional statistical thinking, since statistical laws were established via large samples, the more data the better the analysis. Therefore, straightforward statistical modelling is often difficult to carry on and even not approachable due to sample size.

Facing the challenge of data shortage, we have to walk out from the umbrella of large-sample based statistical theory by changing our attitude and looking at the real world from a different angle. Because system dynamics can be treated from the viewpoint of the degree of information availability, it is possible for us to walk out from the umbrella of large sample statistics and develop a small-sample based spatial data analysis methodology based on the spirit of grey differential equation theory.

3.3.2 PM_{10} Samples in Santa Barbara County

PM_{10} is one of the seven air pollutants the Environmental Protection Agency (EPA) regulates, and exposure to high outdoor PM_{10} concentrations causes increased disease and death (Environmental Protection Agency, 2003). The California Air Resources Board has 55 air quality monitoring sites located within the Santa Barbara County (figure 3.3.1). However, the monitoring sites are not monitored every year, and therefore the air quality data are not collected every year. In 2002, only 4 sites reported PM_{10} samples within Santa Barbara County (figure 3.3.2).

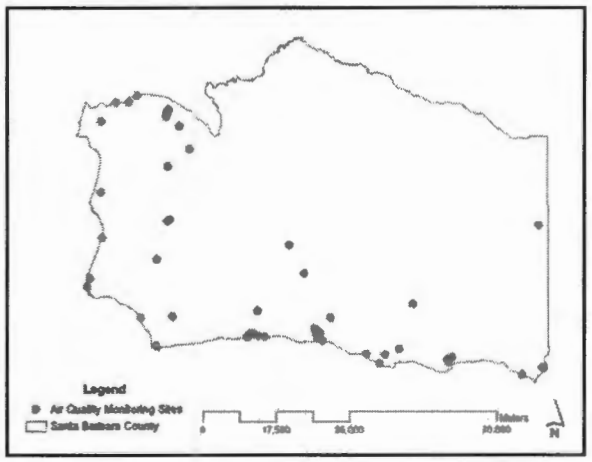


Figure 3.3.1 Air Quality Monitoring Sites in Santa Barbara County, California.

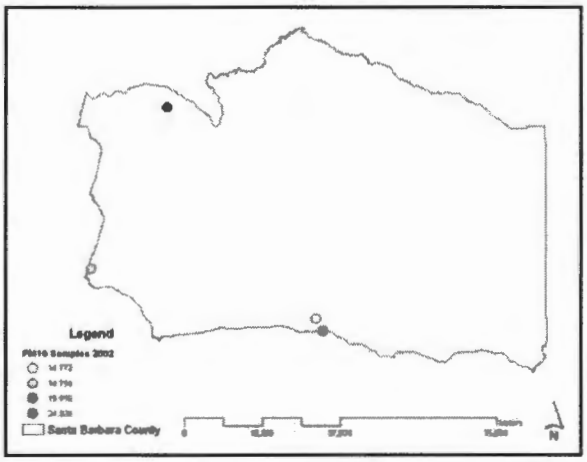


Figure 3.3.2 PM_{10} Sample Locations for 2002 in Santa Barbara County, California.

What can we do with *4* sample points? In GIS software, kriging prediction maps can only be made under the condition there must be *10* or more sample points. Using Grey Spatial Analysis, we will be able to estimate the PM_{10} concentration in different locations within Santa Barbara. For convenience, in this case, we estimated the PM_{10} concentrations in the other 51 monitoring sites within Santa Barbara County, based on the 4 sampled sites in 2002.

3.3.3 Grey Spatial Prediction Equation

The challengeable question is that we are dealing with spatial observations $\{Z(x,y), x, y \in \mathbb{R}\}$ having no ordering while grey differential equation approach intrinsically deals with discrete time sequence data (Fu, 1991). Therefore, the key point is to convert the order-less spatial data into ordered data sequences.

1. Pick up a pair of coordinate (x_0, y_0) as the origin to establish a Cartesian 2-dimensional (localized) coordinate system $\{(x', y'), x' = x - x_0, y' = y - y_0\}$. x_0, y_0 should be located at the reasonably nearest lower-left corner of Santa Barbara County. In Santa Barbara County case, $(x_0, y_0) = (-120.7000, 34.3000)$.
2. Project $\{Z(x,y), x, y \in \mathbb{R}\}$ along x'-axis and y'-axis respectively. Denote them as data set $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and data set $\{Z_y(y'), y' \in \mathbb{R}^+\}$ where

$$Z_x(x') = Z(x,y) \times x' \Big/ \sqrt{(x')^2 + (y')^2}$$
$$Z_y(y') = Z(x,y) \times y' \Big/ \sqrt{(x')^2 + (y')^2} \quad (5)$$

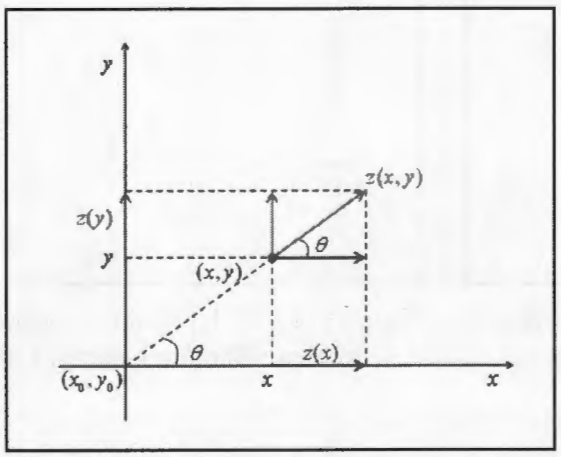


Figure 3.3.3 Illustration of the Projection of z(x,y)-value into the x-axis and y-axis respectively.

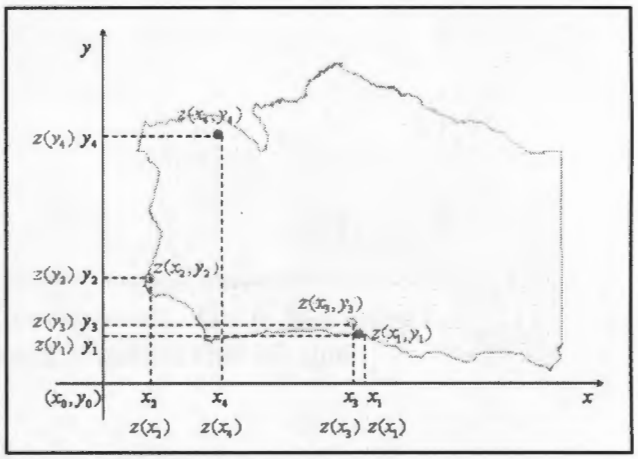


Figure 3.3.4 Illustration of the Projection of x(x,y)-value into x-axis and y-axis for Santa Barbara.

3. Order the numbers in data sets $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y(y'), y' \in \mathbb{R}^+\}$ according to the sequence of coordinate x' and the sequence of coordinate y' and respectively and obtain two data sequences, denoted as $\{Z_x^*(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y^*(y'), y' \in \mathbb{R}^+\}$ respectively.
4. Tabulate the two projection data sequences:

Table 3.3.2 Projected Santa Barbara County PM_{10} Data Sequence and Coordinates.

x'-coordinate	0.067300 (x_2)	0.264500 (x_4)	0.654200 (x_3)	0.674200 (x_1)
x-projection sequence	4.156992393	9.246591929	14.18756366	19.43545485
y'-coordinate	0.162200 (y_1)	0.189700 (y_3)	0.296100 (y_2)	0.643000 (y_4)
y-projection sequence	4.675809518	4.114003097	18.28953117	22.47848246

Notice that original coordinate values (x_i, y_i), $i = 1,2,3,4$ are used to keep track of the original data compositions for later prediction purpose.

It is observed that the distance $x_i' - x_{i-1}'$ $(y_i' - y_{i-1}') \neq$ constant. Therefore, the way to obtain the (grey) response equation is different from that with equal-distance spaced grey differential equation. The methodology here is utilizing two-stage least-square estimation procedure (Fu, 1991).

Let the s-coordinate sequence $Z_s^{(0)} = \left\{ Z_s^{(0)}(s_i),\ s_i \in \mathbb{R}^+,\ i = 1,2,\cdots,N \right\}$ (s is a dummy variable which can represent x' and y' respectively) satisfying the following grey differential equations respectively,

$$Z_s^{(0)}(s_i) + \beta_s \bar{Z}_s^{(1)}(s_i) = \alpha_s,\ i = 2,\cdots,N \tag{6}$$

where

$$\bar{Z}_s^{(1)}(s_i) = \frac{1}{2}\left(Z_s^{(0)}(s_i) + Z_s^{(0)}(s_{i-1}) \right),\ i = 2,\cdots,N \tag{7}$$

Therefore after two-stage least-square fitting, the estimated response function is

$$Z_s^{(0)}(s) = \hat{\alpha}_s + \hat{\gamma}_s \exp\left(-\hat{\beta}_s s\right) \tag{8}$$

Then for $\forall (x, y)$, the grey predicted two-dimensional spatial value will be given by,

$$\hat{Z}(x,y) = \sqrt{\left(\hat{\alpha}_x + \hat{\gamma}_x e^{-\hat{\beta}_x x}\right)^2 + \left(\hat{\alpha}_y + \hat{\gamma}_y e^{-\hat{\beta}_y y}\right)^2} \tag{9}$$

Guo et al. (2005) contains the details on the two-stage least-square estimation.

Using grey spatial prediction, the PM_{10} concentrations are estimated for the other 51 monitoring sites, from the original *4* PM_{10} samples, and 34 monitoring sites were estimated, with

17 sites that can not be predicted, as can be seen in figure 3.3.5. This is because the grey system theory method is good for interpolation, and not very suitable for extrapolation.

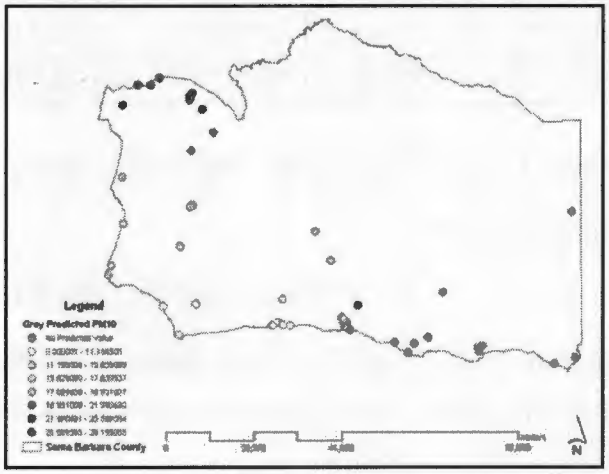


Figure 3.3.5 Grey Prediction for PM_{10} in 2002 in Santa Barbara County.

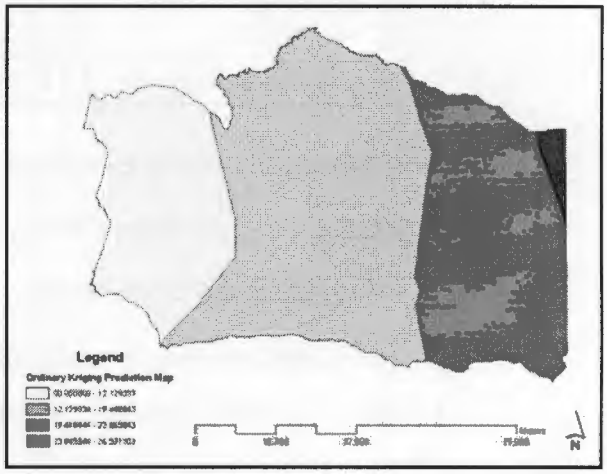


Figure 3.3.6 Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.

The Grey predictions are very consistent with the original *4* PM_{10} samples (relative errors range from 0.15% to 32%) and reveal fairly detailed spatial distribution pattern. As a source of reference, an Ordinary Kriging prediction map is generated from 148 sample points from the entire California State. This is because Kriging prediction maps can only be made under the condition there must be 10 or more sample points in GIS software, and therefore a kriging prediction map can not be produced here, so 148 samples from the entire California is used in order to generate this map. Kriging prediction map of the Santa Barbara County is shown in figure 3.3.6.

3.3.4 Using Historical Time-Series Data to Enhance Information

We examined the air quality data of Santa Barbara County further, and realized that there were two monitoring sites, we will call them site 17 and 38, which contains historical PM_{10} data but was not sampled in 2002 (table 3.3.3).

Table 3.3.3 PM_{10} Time Sequences from Year 1988 to 1999 at Monitoring Site 17, 38 and 57.

Site	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*
17	*	*	36.871	*	30.058	*	*	*	*	29.747	26.079	29.49
38	*	27.573	26.452	*	25.267	26.131	26.42	24.971	25	27.42	*	*
57	33.712	36.073	34.157	32.417	*	26.539	24.799	22.082	*	*	*	*

As a record, site 17 and 18 are within the Santa Barbara County, but Site 57 is the nearest one outside of the county. The purpose for adding site 57 is to strengthen the interpolation area coverage. With the addition of site 17, 38 and 57 in 2002 being available, we choose grey-predicted PM_{10} values at site 2, 23 and 45 respectively. Then, with these 6 PM_{10} values a *Grey Spatial Analysis* can be performed for the prediction of the remaining PM_{10} concentrations at the other unpredicted 17 sites.

Table 3.3.4 Information of the Six Monitoring Sites for Grey Prediction of Remaining 17 Sites.

Site	*Latitude*	*Longitude*	*PM_{10} Value 2002*
2	34.4622	-120.0258	19.23121493
17	34.4166	-119.6999	*
23	34.5136	-120.0055	26.91565579
38	34.4147	-119.8788	*
45	34.4691	-120.0394	15.79061345
57	34.4480	-119.2420	*

By examining the three data sequences in table 3.3.3, the site 38 has 8 years of interrupted data, and site 17 has 5 years of interrupted data. Therefore, classical time-series analysis methodology will not properly work here because the data sequence are interrupted, and even not equal-time spaced, i.e., some "missing" values between them.

Table 3.3.5 Grey Temporal Predictions of PM_{10} from 1992 to 2002 at Site 17, 38 and 57.

Site	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*
17	30.058	*29.606*	*28.971*	*28.621*	*28.428*	29.747	26.079	29.49	*28.214*	*28.204*	*28.199*
38	25.267	26.131	26.42	24.971	25	27.42	*25.705*	*25.630*	*25.558*	*25.490*	*25.424*
57	*30.056*	26.539	24.799	22.082	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*

3.3.5 Grey Spatial-Temporal Prediction Results

Table 3.3.6 Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.

Location	*Latitude*	*Longitude*	*PM_{10} Value 2002*
2	34.4622	-120.0258	19.23121493
17	34.4166	-119.6999	28.19866000
23	34.5136	-120.0055	26.91565579
38	34.4147	-119.8788	25.42448000
45	34.4691	-120.0394	15.79061345
57	34.4480	-119.2420	17.35317800

The table 3.3.6 shows the latest information for grey spatial predictions. With these data information, we use newly developed grey temporal spatial prediction methodology to make the predictions for the remaining 17 sites. As a record, we choose (-120.040, 34.4000) as the new origin for setting up a 2-dimensional Cartesian Coordinate system. The prediction results are shown in figure 3.3.7.

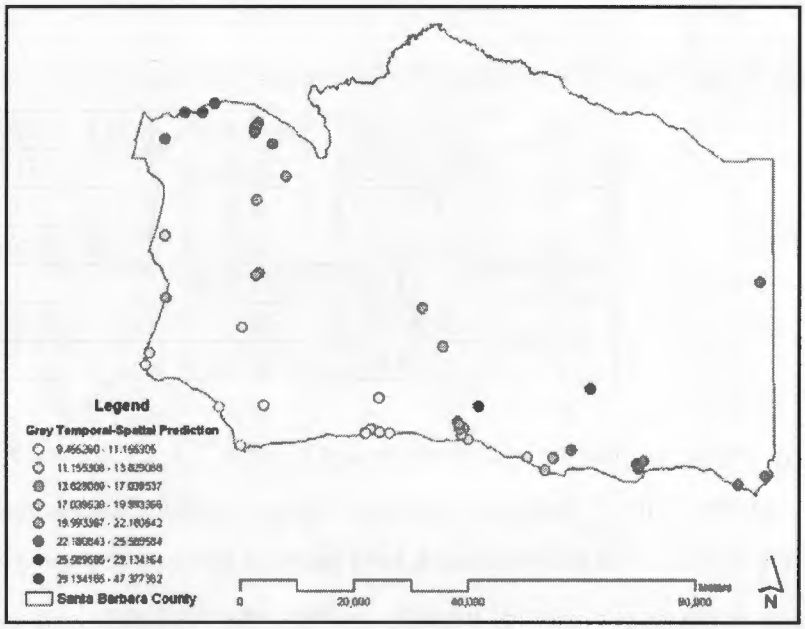


Figure 3.3.7 Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.

The grey predictions and kriging predictions are consistent in the general trend. However, grey predictions reveal better local properties. The Kriging prediction map using 148 sample data (figure 3.3.6) are shown to be relatively rough when compared to the Grey prediction maps using only 4 sample data. The Grey PM_{10} predictions are shown to be very sensitive to local observations.

3.3.6 Conclusion

In this paper, we have developed a small-sample based spatial data analysis methodology based on the spirit of grey differential equation theory. The key contribution is the conversion from order-less 2-dimensional spatial data into ordered data sequences. Furthermore, we follow the idea behind the grey system theory – modelling process is the one of enhancing the degree of

information sufficiency. The way to add grey temporal prediction on three sites is just increasing information. It is clear that the grey spatial prediction results reflect well the local reality. We must point out that the grey spatial prediction developments are not perfect and still needing to be improved. We will use the differential geometry to further solidify the mathematical foundation.

3.4 A New Very Small Sample-Based Non-Linear Statistical Estimation Method

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the 2005 International Workshop on Recent Advances Stochastic Operations Research (RASOR 2005), August 25-26, 2005, Canmore, Alberta, Canada, pp. 50-57.

3.4.1 Introduction

General speaking, the task of a mathematical modelling is to explore the dynamic laws on the state of space. Different from the modelling exercise in operations research, statistical methods is one in which for the given data collected from real world, according to related literature or one's own experience, researchers estimate the parameters according to a class of distributions, and then the fitted distribution facilitates the revelation of the dynamic rule underlying data collected. Therefore, a critical requirement to facilitate statistical data analysis is that the sample size must be large enough. Such a prerequisite makes classical statistics powerless whenever facing small sample data or a sparse sample.

It is obvious that statistical thinking, drawing conclusion from data is correct, but such a correct intention is often haunted by sparse data which is often regarded as inadequate information. In this thinking pattern, only information contained in a large sample is adequate. However, what is critical here is the dynamic law hidden behind the data collected, and not the underlying distribution. We could say that the large-sample barrier is created by a statistician's distributional thinking. Such a thinking logic dominated statisticians for more than a century and created today's large-sample statistical theoretical trend. A realistic data analyst should be concerned more about the dynamics underlying the data, and not the route of thinking. Therefore, there is a necessity to pursue new models based on sparse data.

3.4.2 Foundation of Grey Differential Equation Models

The core of the grey system theory is the grey models being composed of a large variation of grey differential equations (Fu, 1991). General speaking, differential equations are powerful tools in modern science because they reveal deeply the evolving dynamics of real world events. The common sense or the limitation of differential equation is that it can be only utilized to describe continuous differentiable functions. Mathematicians and statisticians are often feeling quite helpless when facing discrete data sequence. Grey system established the grey differential equation in order to solve an ordinary differential equation, using least-square approach for a given discrete data sequence.

Grey models can be called grey differential equation models. The commonly used grey models are GM(1,1), GM(2,1), GM(1,h), GM(0,h) and the Verhulst model (Wen, 2004a). The basic one is GM(1,1), and it will be discussed in detail.

Equation,

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

is called a one-variable first order grey differential equation with respect to time series sequence $X=(x(1),x(2),...,x(n))$, where

$$z^{(1)}(k)=\left(x^{(1)}(k)+x^{(1)}(k-1)\right)/2, \quad k=2,\cdots,n \tag{2}$$

and β is the *developing coefficient*, α is the *grey input*, and $x^{(0)}$ is a *grey derivative* which maximizes the information density for a given series to be modeled. This model is called GM(1,1).

Furthermore, the differential equation $dx^{(1)}/dt+\beta x^{(1)}=\alpha$ is called the whitened differential equation or the shadow equation of the grey differential equation 1. The *t* is continious counterpart of index *k*. The unknown parameter values (α,β) can be determined in terms of the classical least-square method.

$$\begin{bmatrix} a \\ b \end{bmatrix}=\left(X^T X\right)^{-1} X^T y \tag{3}$$

where

$$X=\begin{bmatrix} 1 & -z^{(1)}(2) \\ 1 & -z^{(1)}(3) \\ \vdots & \vdots \\ 1 & -z^{(1)}(n) \end{bmatrix} \quad \text{and} \quad y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{4}$$

Based on the estimated vector $[a,b]^T$ for parameter (vector) $[\alpha,\beta]^T$ and differential equation theory, the predicted equation (i.e., the response equation) is,

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{a}{b}\right]e^{-bk}+\frac{a}{b} \tag{5}$$

which corresponds to the GM(1,1) differential equation, $dx^{(1)}/dt+\beta x^{(1)}=\alpha$.

We would like to emphasize the following critical points:

(1) From equation 1 to 5, it is very clear that GM(1,1) model is established with respect to a sampled discrete data sequence. It is in nature a delicate application of the least-square theory at the first-order differential equation level.

(2) Based on equation 5, it should be noticed that the solution to the grey differential equation is obtained in terms of the corresponding shadow (differential) equation.

(3) The elegancy in grey differential equation models lies in the approximations of the derivatives and integrals of a given function. Deng (1982) named them as inverse accumulative generating operators (I-AGO) and accumulative generating operators (AGO) respectively.

(4) It should be further strongly emphasized that the sampled discrete data sequence itself implies that the data is naturally ordered in the sequence. The ordering index could be time or distance, from a reference point.

3.4.3 State Model with Natural Coordinate System

It is often the case that we are required to seek a state function involving multiple independent variables. Particularly these independent variables are natural coordinates. In some cases, the data sets have no natural order at all and the direct application of grey differential equation is not possible. And therefore it is critical to take necessary steps for converting the order-less data sequence into ordered data sequence(s). The following two subsections propose two ways in terms of PM_{10} spatial pattern prediction in Santa Barbra County, California, U.S.A. as illustrations (Guo et al., 2005)

3.4.3.1 Coordinate-Directional Data-Valued Approach

PM_{10} is one of the seven air pollutants the Environmental Protection Agency (EPA) regulates, and exposure to high outdoor PM_{10} concentrations causes increased disease and death (Environmental Protection Agency, 2005a). The California Air Resources Board has 55 air quality monitoring sites located within the Santa Barbara County (figure 3.4.1).

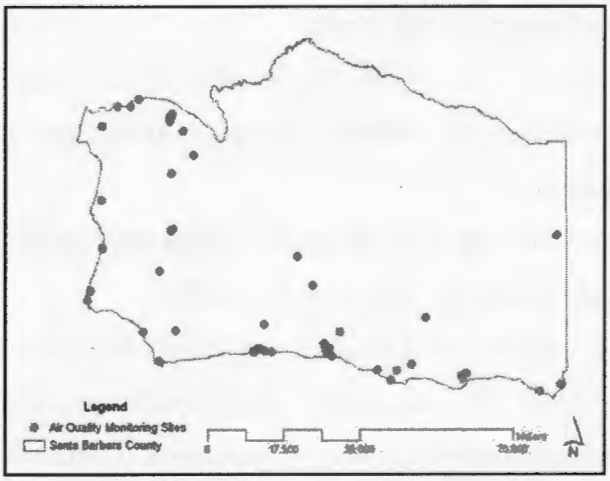


Figure 3.4.1 Air Quality Monitoring Sites in Santa Barbara County.

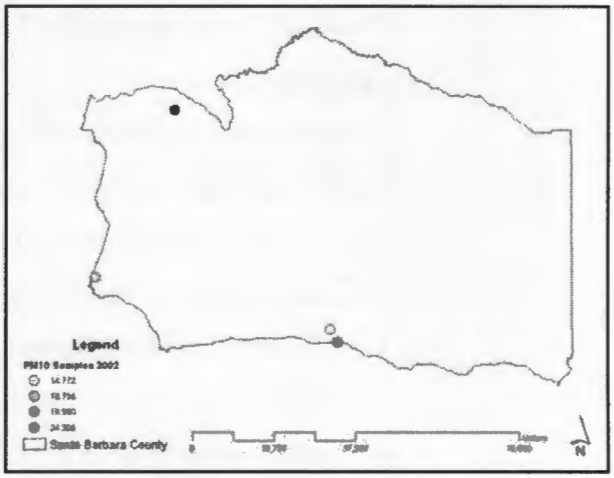


Figure 3.4.2 PM_{10} Sample Locations for 2002 in Santa Barbara County.

In 2002, PM_{10} data were collected from only 4 monitoring sites within the Santa Barbara County (figure 3.4.2). Out of 55 sites only 4 sites were monitored, due to the cost and manpower involved in the sampling process. The site code is the original location code in the data, but for our convenience, we use the site number 1, 2, 3, 4, for clear indication.

Table 3.4.1 PM_{10} Sample Information in Four Sites, 2002.

Sites Number	*Site code*	*Latitude* (y)	*Longitude* (x)	PM_{10}
1	2008	34.4622	-120.026	19.990
2	3023	34.5961	-120.633	18.756
3	3101	34.4897	-120.046	14.722
4	3486	34.9430	-120.436	24.306

(i) We pick up a pair of coordinate $(x_0, y_0) = (-120.70, 34.30)$ as the origin to establish a Cartesian 2-dimensional (localized) coordinate system $\{(x', y'), x' = x - x_0, y' = y - y_0\}$. The first column in table 3.4.2 lists the transformed coordinate values.

Table 3.4.2 Coordinate-Directional Observed PM_{10}-Valued Vectors.

Site Number	*Site Coordinate* (x', y')	$\cos\theta$	$\sin\theta$	$\|\vec{z}\|$
1	(0.6742, 0.1622)	0.97226	0.23391	19.990
2	(0.0673, 0.2961)	0.22164	0.97513	18.756
3	(0.6542, 0.1897)	0.96044	0.27850	14.722
4	(0.2645, 0.6430)	0.38042	0.92481	24.306

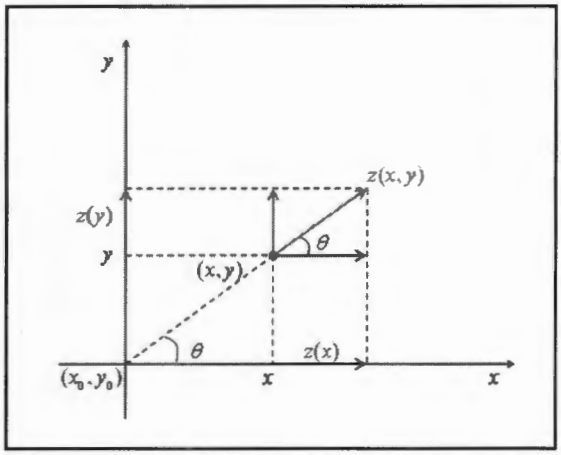


Figure 3.4.3 Illustration of the Projection of z(x,y)-value into the x-axis and y-axis respectively.

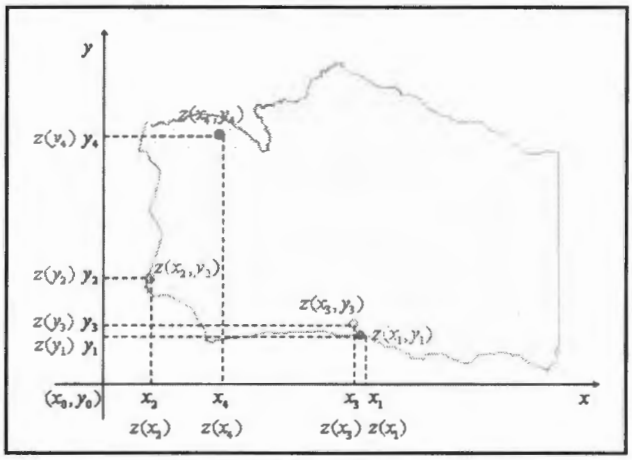


Figure 3.4.4 Illustration of the Projection of x(x,y)-value into x-axis and y-axis for Santa Barbara.

(ii) Project $\{Z(x,y), x, y \in \mathbb{R}\}$ along x'-axis and y'-axis respectively. Denote them as data set $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and data set $\{Z_y(y'), y' \in \mathbb{R}^+\}$ where

$$\begin{aligned} Z_x(x') &= \left\|\vec{Z}(x,y)\right\| \cos\theta \\ Z_y(y') &= \left\|\vec{Z}(x,y)\right\| \sin\theta \end{aligned} \tag{6}$$

(iii) Order the numbers in data sets $\{Z_x(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y(y'), y' \in \mathbb{R}^+\}$ according to the sequence of coordinate x' and the sequence of coordinate y' and respectively and obtain two data sequences, denoted as $\{Z_x^*(x'), x' \in \mathbb{R}^+\}$ and $\{Z_y^*(y'), y' \in \mathbb{R}^+\}$ respectively.

(iv) Tabulate the two projection data sequences:

Table 3.4.3 Projected Santa Barbara County PM_{10} Data Sequence and Coordinates.

x'-coordinate	0.067300 (x_2)	0.264500 (x_4)	0.654200 (x_3)	0.674200 (x_1)
x-projection sequence	4.156992393	9.246591929	14.18756366	19.43545485
y'-coordinate	0.162200 (y_1)	0.189700 (y_3)	0.296100 (y_2)	0.643000 (y_4)
y-projection sequence	4.675809518	4.114003097	18.28953117	22.47848246

We record the original data "order" in parenthesis after the ordered x' coordinate and y' coordinate respectively to keep track of the original data compositions for later prediction purpose.

It is observed that distance $x_i' - x_{i-1}'$ $(y_i' - y_{i-1}') \neq$ constant. Therefore, the way to obtain the (grey) response equation is different from that with equal-distance spaced grey differential equation. The methodology here is utilizing two-stage least-square estimation procedure (Fu, 1991, Guo et al., 2005).

Let s-coordinate sequence $Z_s^{(0)} = \left\{ Z_s^{(0)}(s_i),\ s_i \in \mathbb{R}^+,\ i = 1,2,\cdots,N \right\}$ (s is a dummy variable which can represent x' and y' respectively) satisfying the following grey differential equations respectively,

$$Z_x(x') = \left\| \vec{Z}(x,y) \right\| \cos\theta\ ,\quad Z_y(y') = \left\| \vec{Z}(x,y) \right\| \sin\theta \tag{7}$$

Then:

$$Z_s^{(0)}(s_i) + \beta_s \bar{Z}_s^{(1)}(s_{i+1}) = \alpha_s,\ i = 2,\cdots,N \tag{8}$$

where

$$\bar{Z}_s^{(1)}(s_{i+1}) = \frac{1}{2}\left(Z_s^{(0)}(s_{i+1}) + Z_s^{(0)}(s_i) \right),\ i = 2,\cdots,N \tag{9}$$

with

$$Z_s^{(0)}(s_{k+1}) = \frac{r_1^{(k+1)} - r_1^{(k+1)}}{s_{k+1} - s_k},\ k = 2,\cdots,6 \tag{10}$$

Accordingly, the first-stage least-square parameter estimators are estimated in terms of equation 1 to 7. Due to the basic exponential core part, we can re-write the first-stage response equation as:

$$Z_s^{(0)}(s) = \alpha_s + \gamma_s \exp\left(-\hat{\beta}_s s \right) \tag{11}$$

Then we enter the second-stage least-square estimation

$$\left(\hat{\alpha}_s, \hat{\gamma}_s \right)^T = \left(\mathrm{D}_s{}^T \mathrm{D}_s \right)^{-1} \mathrm{D}_s{}^T \mathrm{z}_N^s \tag{12}$$

where

$$D_s = \begin{bmatrix} 1 & \exp\left(-\hat{\beta}_s s_1\right) \\ 1 & \exp\left(-\hat{\beta}_s s_2\right) \\ \vdots & \vdots \\ 1 & \exp\left(-\hat{\beta}_s s_N\right) \end{bmatrix} \tag{13}$$

And

$$\hat{z}_N^s = \begin{bmatrix} Z_s^{(0)}(s_1) \\ Z_x^{(0)}(s_2) \\ \vdots \\ Z_s^{(0)}(s_N) \end{bmatrix} \tag{14}$$

Therefore after two-stage least-square fitting, the estimated response function is:

$$\hat{Z}_s^{(0)}(s) = \hat{\alpha}_s + \hat{\gamma}_s \exp\left(-\hat{\beta}_s s\right) \tag{15}$$

And hence for $\forall(x, y)$, the grey predicted two-dimensional spatial value will be given by

$$\hat{Z}(x,y) = \sqrt{\left(\hat{\alpha}_x + \hat{\gamma}_x e^{-\hat{\beta}_x x}\right)^2 + \left(\hat{\alpha}_y + \hat{\gamma}_y e^{-\hat{\beta}_y y}\right)^2} \tag{16}$$

Using grey spatial prediction Equation 16, PM_{10} concentrations are estimated for the other 51 monitoring sites from the original 4 PM_{10} samples, and 34 monitoring sites were estimated, with 17 sites unpredicted, as can be seen in figure 3.4.5. This is because the grey system theory method is good for interpolation, and not very suitable for extrapolation. Furthermore, in terms of two incomplete data on other sites, we used the GM(1,1) for the prediction of the 2002 PM_{10} values, and this enhanced the total sample data into six. With the augmented data set we estimated all the remaining 17 sites using grey temporal-spatial prediction method (appendix A), resulting in the new grey temporal-spatial prediction map in figure 3.4.6.

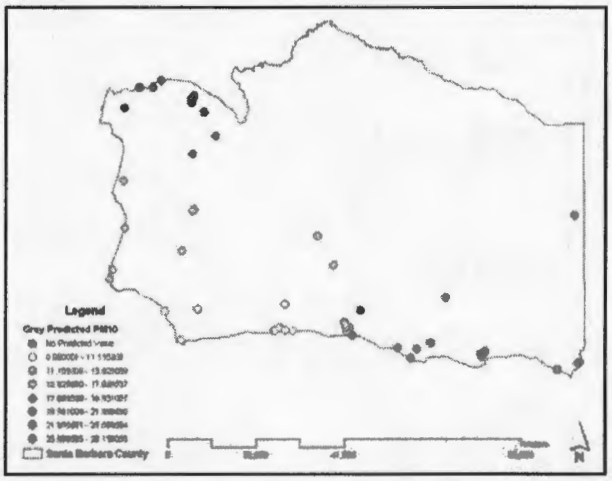


Figure 3.4.5 Grey Prediction for PM_{10} in 2002 in Santa Barbara County.

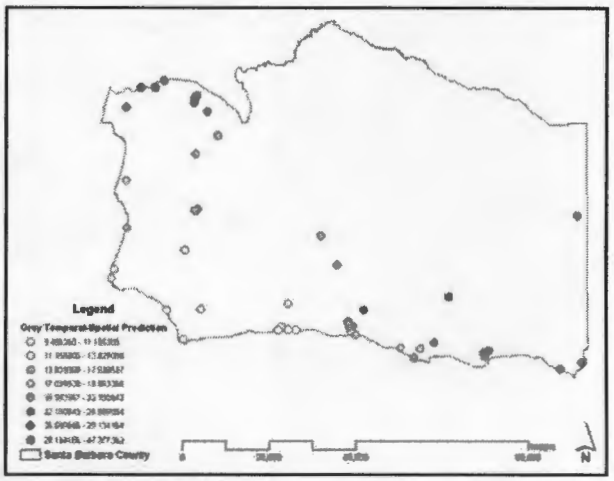


Figure 3.4.6 Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.

The error of the prediction using the relative error is e =(predicted – observed)/observed. If the relative error e is less that 5%, then we accept the grey prediction although the state prediction should be read as a grey number which is imprecise. The resulting grey predictions are consistent with the original PM_{10} samples (relative errors range from 0.15% to 32%), and they reveal a fairly detailed spatial distribution pattern.

3.4.3.2 Gradient Approach

However, the projection approach may give rise to a question that the original information recorded are spatial observations which constitute a scalar field in nature. In order to obtain the "projected values" along x_1, x_2, and x_3 axes respectively the generating of vector $\overrightarrow{OZ}$ and the related vector filed is *artificially* created and therefore the final prediction of the scalar field $z(x_1, x_2, x_3)$ is logically confusing although the computation results may produce very good results. Furthermore, it should be pointed out if only the coordinate-oriented observation-valued generated vector were value order-preserved along all axes, then the computation results are logical.

Therefore it is logical to search a methodology in order to generate *natural* vector field from the spatial observations and therefore the projections associated with the natural vector filed are also (value) order-preserved. From the theory of differential geometry, it is a well-known fact that there is a natural vector filed associated with a scalar field. In other words, given a scalar field

$\{z(x_1,x_2,x_3)\}$, the gradient vector field $\{\text{grad}\, z(x_1,x_2,x_3)\}$ is the natural one having natural projections along x_1, x_2, and x_3 axes respectively.

$$\text{grad}\, z(x_1,x_2,x_3)=\frac{\partial z}{\partial x_1}\vec{i}+\frac{\partial z}{\partial x_2}\vec{j}+\frac{\partial z}{\partial x_3}\vec{k} \tag{17}$$

where $\vec{i}, \vec{j},$ and $\vec{k}$ are unit vectors along x_1, x_2, and x_3 axes respectively.

To obtain the values of the three partial derivatives $\partial\varphi/\partial x_1$, $\partial\varphi/\partial x_2$, and $\partial\varphi/\partial x_3$ in terms of the spatial observational data, the directional derivative concept can be utilized. For any given two points in the three dimensional coordinate system, say, (x_1^1,x_2^1,x_3^1) and (x_1^2,x_2^2,x_3^2), a direction $\vec{l}$ can be defined by

$$\vec{l}=\cos(\alpha)\vec{i}+\cos(\beta)\vec{j}+\cos(\gamma)\vec{k} \tag{18}$$

where

$$\cos(\alpha)=\frac{x_1^2-x_1^1}{\rho}\quad \cos(\beta)=\frac{x_2^2-x_2^1}{\rho}\quad \cos(\gamma)=\frac{x_3^2-x_3^1}{\rho} \tag{19}$$

And

$$\rho=\sqrt{\left(x_1^2-x_1^1\right)^2+\left(x_2^2-x_2^1\right)^2+\left(x_3^2-x_3^1\right)^2} \tag{20}$$

Then the directional derivative is:

$$\frac{\partial\varphi}{\partial l}=\vec{l}\bullet\text{grad}\varphi=\frac{\partial\varphi}{\partial x_1}\cos(\alpha)+\frac{\partial\varphi}{\partial x_2}\cos(\beta)+\frac{\partial\varphi}{\partial x_3}\cos(\gamma) \tag{21}$$

Now, it is quite obvious that although the natural vector field $\text{grad}\,\varphi$ can not be directly found the directional derivative from point (x_1^1,x_2^1,x_3^1) to point (x_1^2,x_2^2,x_3^2) can be estimated. Therefore the approach to obtain these partial derivatives $\partial\varphi/\partial x_1$, $\partial\varphi/\partial x_2$, and $\partial\varphi/\partial x_3$ can be developed in terms of the directional derivative approximation.

$$\widehat{\frac{\partial\varphi}{\partial l}}=\frac{\varphi\left(x_1^2,x_2^2,x_3^2\right)-\varphi\left(x_1^1,x_2^1,x_3^1\right)}{\sqrt{\left(x_1^2-x_1^1\right)^2+\left(x_2^2-x_2^1\right)^2+\left(x_3^2-x_3^1\right)^2}} \tag{22}$$

Then for given observation set $\{\varphi(x_1^i,x_2^i,x_3^i),i=1,2,\cdots,m\}$, $m!/(2!(m-2)!)$ directional derivatives can be estimated and therefore at each given point (x_1^i,x_2^i,x_3^i) a linear equation system

having $m-1$ linear equations with the three partial derivatives $\partial\varphi/\partial x_1^i$, $\partial\varphi/\partial x_2^i$, and $\partial\varphi/\partial x_3^i$ can be established. This linear equation system for the three partial derivatives at given point (x_1^i, x_2^i, x_3^i) is in general over-specified but the least-square estimate for the partial derivatives can be found. In this manner, three sets of estimated partial derivatives $\{\partial\varphi/\partial x_j^i, i=1,2,\cdots,m\}$, $j=1,2,3$ can be obtained and ordered. They provide the natural projections and the application of GM(1,1) modelling to each individual partial derivative sequence and the state dynamics at partial derivative level can be determined.

However, our basic concern is to find the state dynamics of the scalar field itself. In other words, what we are interested is the scalar filed $\{\varphi(x_1,x_2,x_3),(x_1,x_2,x_3)\in D\subset\mathbb{R}^3\}$. By noticing the relation between the three partial derivatives $\partial\varphi/\partial x_1$, $\partial\varphi/\partial x_2$, and $\partial\varphi/\partial x_3$ and φ (Mathematics Handbook Group, 1979, pp. 439):

$$\begin{aligned}\varphi(x_1,x_2,x_3)=\varphi\left(x_1^{(1)},x_2^{(1)},x_3^{(1)}\right)+\int_{x_1^{(1)}}^{x_1}\frac{\partial\varphi}{\partial x_1}\left(u,x_2^{(1)},x_3^{(1)}\right)du\\+\int_{x_2^{(1)}}^{x_2}\frac{\partial\varphi}{\partial x_2}\left(x_1,v,x_3^{(1)}\right)dv+\int_{x_3^{(1)}}^{x_3}\frac{\partial\varphi}{\partial x_3}(x_1,x_2,w)dw\end{aligned}\tag{23}$$

3.4.4 State Model without Natural Coordinate System

In modelling exercises, it occurs often that we need to seek the functional relationship between a state variable interested and its covariates. The standard exercise is to seek the correlation matrix. However, it merely reflects the linear relation and also requires large data. Cox (1972) proposed proportional hazards models (PH) and gained great successful applications in medical and industrial research. However, if the data is sparse, PH models will no longer be applicable because of the large sample requirements.

On the other hand, in the spirit of the grey system theory a GM(1,*K*) model was proposed for describing the grey relation between state variable X_0 and other *K* state variables, which reveals their geometric relationship and thus non-linear in nature. It is unfortunate that the relationship is qualitative and emphasizing the ordering of the impacts from covariates although GM(*1,K*) model does work with reasonably small sample size (Wen, 2004a). For example, we are interested in

studying the impacts from repair *cost*, covariate *D* and *W*. Table 3.4.4 lists six data points where the covariates are normalized.

Table 3.4.4 Stopping times and related covariates (normalized).

Time from Last PM	*Cost*	*D*	*W*
54	0.0468986	0.7500	0.6153846
133	0.0716086	0.8125	0.9230769
147	0.1512859	0.9375	0.7692307
72	0.1195159	0.7500	0.8461538
105	0	0.8125	0.9230769
115	0.2647504	0.6875	0.6923076

It is obvious that we can pick *Cost* (denoted by *c*), covariates *D* and *W* (denoted by *d* and *w* respectively), i.e., the triple (*c,d,w*) as an affine coordinate system. However, it would involve more affine differential geometry concepts to handle the grey state estimation problem.

An intuitive approach is proposed as following. Put the six (*c,d,w*) observations as 6x3 matrix, denoted by *X*. Then seek an orthogonal matrix *O* such that $O(X^T X) = (X^T X)O = \mathrm{I}$ and thus three linearly independent vectors are identified, denoted by $x_1\vec{i}, x_2\vec{j}, x_3\vec{k}$ where,

$$\begin{aligned} x_1 &= a_{11}c + a_{12}d + a_{13}w \\ x_2 &= a_{21}c + a_{22}d + a_{23}w \\ x_3 &= a_{31}c + a_{32}d + a_{33}w \end{aligned} \tag{24}$$

Once the orthogonal coordinate system (x_1, x_2, x_3) is established, it will be feasible to utilize the gradient approach proposed for seeking the state dynamics in the coordinate system (x_1, x_2, x_3), and therefore establish the functional relationship between stopping time *t* and x_1, x_2, x_3. Once $t(x_1, x_2, x_3)$ is estimated, then by changing variables according to equation 24, the direct functional relationship $t^*(c, d, w)$ can be established.

3.4.5 System of Grey Differential Equations

For a complex system containing multiple mutually related factors and multiple independent control variables, any single model will not reflect the changes and developments of the whole

system. It is necessary to establish a system of grey differential equations in order to reveal the system state dynamic changes (Fu, 1991).

Assume that $X_1^{(0)}, X_2^{(0)}, \ldots, X_m^{(0)}$ are the data sequences of the system state variables and $U_1^{(0)}, U_2^{(0)}, \ldots, U_l^{(0)}$ are the data sequences of the system control variables, then the following grey differential equations constitute of the grey state equation system.

$$\begin{cases} dx_j^{(1)}/dt = \sum_{p=1}^{m} \alpha_{jp} x_p^{(1)} + \sum_{q=1}^{l} \beta_{jq} u_q^{(1)}, & j = 1,2,\cdots,m \\ du_v^{(1)}/dt = \gamma_v u_v^{(1)} + \delta_v, & v = 1,2,\cdots,l \end{cases} \tag{25}$$

As a matter of fact the state equation system is composed of m GM(1,m+l) and l GM(1,1) grey differential equations. It can be written as a matrix form:

$$\begin{cases} dX/dt = AX + BU \\ dU/dt = CU + D \end{cases} \tag{26}$$

where

$$\begin{aligned} A &= (\alpha_{st})_{m\times m} & B &= (\alpha_{uv})_{m\times l} \\ C &= diag(\gamma_p)_{l\times l} & D &= (d_q)_{l\times 1} \end{aligned} \tag{27}$$

X is called state vector, U the control vector, A the state matrix, B the control matrix, C the development matrix and D the grey input vector. The matrices, A, B, and C and vector D can be estimated in terms of least-square approach by utilizing the observations of $X^{(0)}$ and $U^{(0)}$. Let:

$$E_j(k+1) = \frac{1}{a_{jj}} \left[\sum_{p=2, p\neq j}^{m} \alpha_{jp} x_p^{(1)}(k+1) + \sum_{q=1}^{l} b_{jq} u_q^{(1)}(k+1) \right], \quad j = 1,2,\cdots,m \tag{28}$$

Then the approximate state response functions can be expressed by:

$$\begin{cases} \hat{x}_j^{(1)}(k+1) = \left\{ x_j^{(0)}(0) + E_j(k+1) \right\} e^{a_{jj}k} - E_j(k+1), & j = 1,2,\cdots,m \\ \hat{u}_v^{(1)}(k+1) = \left(u_v^{(0)}(0) + d_v/c_v \right) e^{c_v k} - d_v/c_v, & v = 1,2,\cdots,l \end{cases} \tag{29}$$

3.4.6 Concluding Remarks

In this paper we investigated a new least-square theory based estimation technique under very small sample circumstances. We call it a grey state model because the approach does not involve

the probability distribution underlying and directly seek the state law in terms of grey differential equation GM(1,1). The term grey indicates the estimated non-linear function with exponential functional as core part is imprecise. The error is often evaluated by relative error and the error structure can use the well-known δ-method.

3.5 Optimal Data Transformations In Grey Analysis

Danni Guo, Renkuan Guo, Christien Thiart, and Yanhong Cui

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7707, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za ycui@stats.uct.ac.za

This paper has been published in the Proceedings of the 10th International Conference on Industrial Engineering – Theory, Applications and Practice (IJIE 2005), December 4-7, 2005, Clearwater, Florida, U.S.A., pp. 426-431.

3.5.1 Introduction

Statistical data analysis is a mathematical optimization problem in nature. For example, the most popular statistical analysis model, linear regression analysis, is just a L_2 optimization problem. With the great efforts in mathematical research, more and more statistical models with different metrics as optimization criteria were studied, for example, L_1 norm, L_∞ and others. However, since the beginning of 1980, the concept of grey system, which only provides incomplete system information, was investigated and accordingly the data analysis of grey system was paid great attention. Particularly, it is worth to mention that the first order one variable grey differential model (Deng, 1982), abbreviated as GM(1,1), is the best researched one since it possesses the best predictable power.

Similar to statistical data analysis, transformation is regarded as an important approach to achieve better modelling for the revelation of the underlying mechanism governing the data. Grey data analysis is carried under the small sample constraint and thus possesses its own characteristic. Such unique feature is inevitably determining the characteristic of transformation and the optimality. The optimal criterion is established in terms of the class ratio concept which specifies the smoothness of a discrete data sequence and possible to grey exponential law underlying the discrete data sequence. The aim of this paper is to extend grey theory transformations, and explore the transformation of the discrete data sequence.

3.5.2 Admissible interval for class ratio $\sigma(k)$

The key step in GM(1,1) modelling (Deng, 1982) is the accumulation generating operation, (AGO), which generates a monotone-increasing discrete sequence not necessarily possessing exponential changing rate. As a matter of fact the real role player is class ratio $\sigma_x(k)$ (Deng, 2002a).

Definition 1: Given a discrete positive real-valued data sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, then $\{\sigma_x(k), k=2,\ldots,n\}$:

$$\sigma(k) = x^{(0)}(k) / x^{(0)}(k-1), k = 2,3,\cdots,n \tag{1}$$

is called *the class ratio* of the original data sequence $X^{(0)}$.

Definition 2: (GM(1,1) model). Equation:

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{2}$$

is called a one-variable first order grey differential equation (GM(1,1)) with respect to time series sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, where

$$\begin{aligned} z^{(1)}(k) &= \frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right] \\ x^{(1)}(k) &= \sum_{i=1}^{k} x^{(0)}(i) \\ k &= 2,\cdots,n \end{aligned} \tag{3}$$

and α is called the developing coefficient, β is the grey input, and $x^{(0)}$ is a *grey derivative* which maximizes the information density for a given series to be modeled. This model is called GM(1,1) model with equal-gap.

Furthermore, the differential equation $dx^{(1)}/dt + \beta x^{(1)} = \alpha$ is called the whitenization differential equation or the *shadow* equation of the grey differential equation 1. The unknown parameter values (α,β) can be estimated in terms of the classical least-square approach.

$$\begin{aligned} \alpha &= \frac{CD-(n-1)E}{(n-1)F-C^2}, \quad \beta = \frac{DF-CE}{(n-1)F-C^2} \\ C &= \sum_{k=2}^{n} z^{(1)}(k), \quad D = \sum_{k=2}^{n} x^{(0)}(k) \\ E &= \sum_{k=2}^{n} z^{(1)}(k)kx^{(0)}(k), \quad F = \sum_{k=2}^{n}\left[z^{(1)}(k)\right]^2 \end{aligned} \tag{4}$$

The grey filtering-prediction equation is

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{\hat{\alpha}}{\hat{\beta}}\right]e^{-bk}+\frac{\hat{\alpha}}{\hat{\beta}} \tag{5}$$

Lemma 3: The admissible interval for fitting a GM(1,1) model from the data sequence $X^{(0)}$ is that the class ratio has to satisfy inequality $e^{-2} \le \sigma(k) \le e^{2}$. Furthermore, for a given $X^{(0)}$ with dimensionality n, the admissible interval of the class ratio is $\left[e^{-2/(n+1)}, e^{2/(n+1)}\right]$.

Corollary 4: The admissible interval for parameter α (the development coefficient of GM(1,1) model) is $\alpha\in(-2,2)$. Furthermore for a given $X^{(0)}$ with dimensionality n, the admissible interval for the development coefficient of GM(1,1) model (Deng 2002a), α, is $[-2/(n+1),\ 2/(n+1)]$.

According to a simulation study by Liu et al. (2004), if the GM(1,1) model were expected to perform mid-long term prediction, then the admissible interval for α should be further tightened. We state this fact as a conjecture.

Table 3.5.1 The Range of the Grey Development Coefficient.

$-\alpha$	0.1	0.2	0.3	0.4	0.5	0.6	0.8	1.0	1.5	1.8
1-step error	0.129%	0.701%	1.998%	4.317%	7.988%	13.405%	31.595%	65.117%	-	-
2-step error	0.137%	0.768%	2.226%	4.865%	9.091%	15.392%	36.979%	78.113%	-	-
5-step error	0.160%	0.967%	2.912%	6.529%	12.468%	21.566%	54.491%	-	-	-
10-step error	0.855%	1.301%	4.067%	9.362%	18.330%	32.599%	88.790%	-	-	-

Conjecture 5: If α, the development coefficient of GM(1,1) model, satisfies condition: $\alpha\in[-0.3,0)$, then the GM(1,1) model can be used for long-tem or mid-term predictions; If α satisfies condition: $\alpha\in[-0.5,-0.3)$, then the GM(1,1) model can be used for short-tem predictions; If $\alpha\in[-0.8,-0.5)$, then using GM(1,1) model for short-term prediction needs to be cautious.

Corollary 6: If a GM(1,1) model is expected to be admissible for mid-long term prediction, then the admissible interval for the class ratio $\sigma(k)\in(e^{-0.3},1)$.

3.5.3 Deng's Transformations

Transformations are needed in order to convert the original data sequence $X^{(0)}$ which has a inadmissible or poorly admissible class ratio $\sigma_x(k)$, k=2,3,…,n, into a new data sequence $Y^{(0)}$, which has a better admissible class ratio $\sigma_y(k)$, k=2,3,…,n, such that the resulting grey differential

equation models can achieve a better modelling efficiency (suitable for mid-long term prediction purposes).

In general, transformation can be represented by the following functional:

$$y^{(0)}(k)=\varphi\left(x^{(0)}(k)\right),\ k=1,2,\cdots,n \tag{6}$$

Deng (2002b) proposed three data transformations: log-transformation ($[\ln(X^{(0)})]^{\lambda}$), root-transformation $\sqrt[\lambda]{X^{(0)}}$ and horizontal shift transformation ($\lambda+X^{(0)}$, $\lambda>0$). For log-transformation the new class ratio,

$$\sigma_y(k)=\left[\ln\left(x^{(0)}(k)\right)/\ln\left(x^{(0)}(k-1)\right)\right]^{\lambda} \tag{7}$$

Denote $H_k=\ln\left(\ln\left(x^{(0)}(k)\right)/\ln\left(x^{(0)}(k-1)\right)\right)$ and $c=2/\left[(n+1)H_k\right]$, then the problem is to select λ such that

$$\begin{cases} -c\le\lambda_k\le c & \text{if } H_k>0 \\ c\le\lambda_k\le -c & \text{if } H_k<0 \end{cases} \tag{8}$$

Denote the solution set by Λ_k, then the set $\Lambda=\bigcap_{k=2}^{n}\Lambda_k$ gives the range of λ if set $\Lambda\neq\varnothing$.

For root-transformation, it is notice that $\sigma_y(k)=\sqrt[\lambda]{x^{(0)}(k)/x^{(0)}(k-1)}=\sqrt[\lambda]{\sigma_x(k)}$. Denote $J_k=\ln\left(\sigma_x(k)\right)$, then for each given k, therefore the problem is to select λ such that:

$$\begin{cases} \lambda_k\le-\dfrac{(n+1)J_k}{2} \text{ or } \lambda_k\ge-\dfrac{(n+1)J_k}{2} & \text{if } J_k>0 \\ \lambda_k\le\dfrac{(n+1)J_k}{2} \text{ or } \lambda_k\ge-\dfrac{(n+1)J_k}{2} & \text{if } J_k<0 \end{cases} \tag{9}$$

Denote the solution set by Π_k, then the set $\Pi=\bigcap_{k=2}^{n}\Pi_k$ gives the range of λ if set $\Pi\neq\varnothing$.

Finally, for the horizontal shift transformation, notice that $\sigma_y(k)=\left(\lambda+x^{(0)}(k)\right)/\left(\lambda+x^{(0)}(k-1)\right)$, then select a λ such that:

$$\lambda_k = \min\left\{\frac{\exp(-2/(n+1))x^{(0)}(k-1)-x^{(0)}(k)}{1-\exp(-2/(n+1))}, \frac{x^{(0)}(k)-\exp(2/(n+1))x^{(0)}(k-1)}{\exp(2/(n+1))-1}\right\} \quad (10)$$

Denote the solution set by Υ_k for given k, then the set $\Upsilon = \bigcap_{k=2}^{n} \Upsilon_k$ gives the range of λ if set $\Upsilon \neq \varnothing$.

3.5.4 Guo's Type I Transformation

It is assumed that $X^{(0)}$ is (strictly) positive real-valued discrete data sequence where $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, then Type I transformation is:

$$y^{(0)}(k) = \exp\left(-\lambda x^{(0)}(k)\right), \ \lambda > 0 \quad (11)$$

Then the new class ratio $\sigma_y(k) = \exp\left(-\lambda\left[x^{(0)}(k)-x^{(0)}(k-1)\right]\right)$, then the Type I transformation becomes a problem of selecting a proper λ such that:

$$\begin{cases} -\dfrac{2}{(n+1)M_k} \le \lambda_k \le \dfrac{2}{(n+1)M_k} & \text{if } M_k > 0 \\ \dfrac{2}{(n+1)M_k} \le \lambda_k \le -\dfrac{2}{(n+1)M_k} & \text{if } M_k < 0 \end{cases} \quad (12)$$

where $M_k = x^{(0)}(k-1) - x^{(0)}(k)$. Denote the solution set by Δ_k for given k, then the set $\Delta = \bigcap_{k=2}^{n} \Delta_k$ gives the range of λ if set $\Delta \neq \varnothing$. Furthermore, if such $\lambda > 0$ exists, then it should be investigated that whether $-\alpha_\lambda$ could be limited within (0, 0.3) for securing mid-long term predictions.

3.5.5 Guo's Type II Transformation

It is a standard exercise that $X^{(0)}$ is assumed to be a strictly positive real-valued discrete data sequence where $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, $n \geq 4$. In grey theory, the minimum sample size for grey differential equation modelling is 4. In the grey methodology developments, $X^{(0)}$ is assumed to be strictly positive otherwise the key treatment in grey theory, AGO will not be

applicable and all the related methodologies, say, grey differential equation models, will be nullified. Deng's (rightward) horizontal shift transformation is one way to sequence the (transformed) discrete data sequence $Y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \ldots, y^{(0)}(n))$, $n \geq 4$ to be positive strictly.

However, the rightward horizontal shift could not secure the transformed discrete data sequence possessing exponential changing rate in the 1-AGO sequence $Y^{(1)} = (y^{(1)}(1), y^{(1)}(2), \ldots, y^{(1)}(n)$, $n \geq 4$, where,

$$y^{(1)}(k) = \sum_{i=1}^{k} y^{(0)}(i), \; k = 1, 2, \cdots, n \tag{13}$$

The motivation of Type II transformation is to extend the GM(m,N) model from (strictly) positive real-valued discrete data sequence assumptions into *arbitrary real-valued* discrete data sequence so that the applicability of real-valued discrete data sequence will be enlarged substantially. However, it is worth to point out that the transformation must be ordering preserved and trending (of data values) preserved.

There are various exponential family of the functions available with the trend and order preservation, say, double exponential-wise and normal density-wise functions. A double exponential transformation can be defined by:

$$\psi\left(x^{(0)}(k)\right) = \exp\left[-\lambda\left|x^{(0)}(k)\right|\right], \; \lambda > 0; \; k = 1, 2, \cdots, n \tag{14}$$

and a normal density-wise transformation can be defined by:

$$\psi\left(x^{(0)}(k); \lambda\right) = \exp\left[-\frac{\lambda\left(x^{(0)}(k)\right)^2}{2}\right], \; \lambda > 0; \; k = 1, 2, \cdots, n \tag{15}$$

As a Type II transformation example, let us examine the normal density-wise transformation defined in equation 15. It is noticed that the standard normal density is expressed by:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \tag{16}$$

which is related to ψ transformation,

$$\psi(x;\lambda) = e^{-\frac{\lambda x^2}{2}} = \sqrt{\frac{2\pi}{\lambda}}\phi\left(\sqrt{\lambda}x\right) \tag{17}$$

and the accumulated distribution for normal variate is:

$$\int_{-\infty}^{x^{(0)}(k)} \psi(x;\lambda) = \frac{\sqrt{2\pi}}{\lambda}\int_{-\infty}^{x^{(0)}(k)} \phi\left(\sqrt{\lambda}u\right)d\left(\sqrt{\lambda}u\right) = \frac{\sqrt{2\pi}}{\lambda}\Phi\left(\sqrt{\lambda}x^{(0)}(k)\right) \tag{18}$$

which implies that:

$$y^{(1)}(k) = \sum_{i=1}^{k} y^{(0)}(i) \approx \frac{\sqrt{2\pi}}{\lambda}\,\Phi\left(\sqrt{\lambda}y^{(0)}(k)\right), k = 1,2,\cdots,n \tag{19}$$

It is noticed that $\phi(x)$ is a normal cumulative distribution function. Then the remaining task of Type II transformation is to select the parameter $\lambda>0$ such that the transformed discrete data sequence $Y(0)$ possesses class ratio:

$$\sigma_y(k) = \frac{\psi\left(x^{(0)}(k);\lambda\right)}{\psi\left(x^{(0)}(k-1);\lambda\right)} = \frac{\phi\left(\sqrt{\lambda}x^{(0)}(k)\right)}{\phi\left(\sqrt{\lambda}x^{(0)}(k-1)\right)} \tag{20}$$

It is noticed that

$$\sigma_y(k) = \exp\left(-\frac{\lambda}{2}\left[\left(x^{(0)}(k)\right)^2 - \left(x^{(0)}(k-1)\right)^2\right]\right) \tag{21}$$

Denote $\Gamma_k = \left(x^{(0)}(k-1) + x^{(0)}(k)\right)\left(x^{(0)}(k-1) - x^{(0)}(k)\right)/2$ and it is noticed that the sign of Γ_k depends upon whether $x^{(0)}(k-1) > x^{(0)}(k)$ or $x^{(0)}(k-1) < x^{(0)}(k)$. Then the problem becomes the selection of $\lambda_k > 0$ for any given k such that:

$$\begin{cases} -\dfrac{2}{(n+1)\Gamma_k} \le \lambda_k \le \dfrac{2}{(n+1)\Gamma_k} & \text{if } \Gamma_k > 0 \\ \dfrac{2}{(n+1)\Gamma_k} \le \lambda_k \le -\dfrac{2}{(n+1)\Gamma_k} & \text{if } \Gamma_k < 0 \end{cases} \tag{22}$$

Denote the solution set by Θ_k for given k, then the set $\Theta = \bigcap_{k=2}^{n}\Theta_k$ gives the range of λ if set $\Theta \neq \varnothing$.

3.5.6 Concluding Remarks

In this paper, the transformation of the discrete data sequence is explored. The transformation, different from other data analysis methodology, must trend and order preserved and at the same time, it is expected the transformed data sequence can have a class ratio approaching 1. Deng's (2002b) three transformations are reviewed and two new transformations are developed. If Type II transformation can be applied universally and obtain satisfactory GM(m,N) models at transformed data sequence $Y^{(0)}$ level, then it is would be a major contribution to grey data analysis, particularly, residual analysis. The new transformation development may expand the grey data analysis to any number system, say real number system and complex number system and even further into vector space, matrix space and tensor space in general. However, so far it is still a theoretical frame and whether or not a transformation is successful is in general not guaranteed although the solution sets for them are given.

3.6 Grey Reliability Analysis of Complex System

Renkuan Guo and Danni Guo

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7707, Cape Town, South Africa
rguo@stats.uct.ac.za dannig@stats.uct.ac.za

This paper has been published in the Proceedings of the 10th International Conference on Industrial Engineering – Theory, Applications and Practice (IJIE 2005), December 4-7, 2005, Clearwater, Florida, U.S.A., pp. 432-437.

3.6.1 Introduction

Today's dazzling and rapid stride in technology have buried the traditional way of designing, manufacturing , operating and analyzing a single-part (or single-subsystem) and brought the 21th century into a system engineering age. Today's design and analysis, due to the complexity of the system designing, manufacturing and operating, has to investigate the operating behaviour of the complex system as a whole, i.e., its operating mechanism underlying the whole system and the interactions between subsystems within the system and the environmental impacts on the whole system operating behaviour as well, i.e., the interactions between the whole system studied and other systems, say, social systems, business systems, climate system and others. Inevitably, the reliability of complex system, a quality index of the operating behaviour of the system, will be one of the focus points of quality and productivity improvement.

Any complex system is an element of our world that exists objectively. The real world including its every element within it is revealed or represented by the information generated by the world. In this sense, research on any system, no matter how complex it is a matter of information processing. Thus research of a complex system is in nature the collection, analysis and feedback of the information on the system and its operating environments. Definitely the reliability of complex system is a treatment and analysis on the information of system reliability. Furthermore, the research of any system, no matter it is a simple one or a complex one, is no longer a research of the behaviour of any individual subsystem or part of the system although it is a necessary step, a typical part analysis route.

System is constituted by hard subsystems including parts and components and soft subsystems including management team and operators. Therefore, system reliability research has to be carried on the information or knowledge of the composition of the system because the system reliability is the comprehensive response index of the reliability of every individual subsystem.

The reliability of system is referred to as the quality index that the constituting subsystems, parts and components of the given system complete the specified functions (or missions) accurately in mutually harmonious manner under the specified conditions (including hard conditions: say, operating environments, input materials and soft conditions: say, system management and operator team, and computerized-monitoring-atomization subsystems etc), within specified time. It is necessary to point out that the reliability of a complex system is not a simple addition of these of the constituting subsystems. The subsystems composing the system may be

independent and may be mutually related (correlated). A key element of complex reliability research is the revelation of the mutual relationship between subsystems. The term *relationship* between subsystems is used for emphasizing the general or nonlinear (including linear) relation.

It is obvious that the analysis of complex system reliability is nothing but the analysis of the information extracted from the complex system, particularly the information from the dynamic operating behaviour of the system. As a general exercise in modern control theory, a system with complete and definite information is called a *white* system; a system with no information at all is called a black system; a system with partial certain information and partial uncertainty information is a grey system (Deng, 1982, 1997, 2002a). Therefore, the complex system reliability analysis is in nature an analysis of a system with partial information.

3.6.2 The Grey Characters of Complex System Reliability

System reliability, as a quality index, is the capability to complete the specified functions accurately in mutually harmonious manner under the specified conditions within specified time period. When we use probability as a measure for describing such a system capability is called reliability. However, it should be fully aware that the concept of system reliability itself possesses a certain grey characteristic. If we take the capability (i.e., reliability) of the movement of the operating system as an example, it can be reasonable to argue that when the operating system is carrying on its specified task for the specified time interval $[t_1, t_2]$ the accurate characteristic of each movement's behavior in the operating system is not clear. More specifically, let us use a train station ticket selling robot as an illustrative example.

The right "hand" of the robot is required to move from point a_1 (ticket drawer) to point a_2. All we know that each ticket is put at a point $\tilde{a}_2$ within the table area under ticket window in front of customer but we do not know the accurate error between design-specified point a_2 and the actual point $\tilde{a}_2$. If the movement error, denoted by e, of the robot's right hand is described by a probability density $\varphi(e)$, if the movement error satisfies system parameter-design constraint $e_1 \le e \le e_2$ then the robot's right hand can handle ticket selling task. Then there liability of the ticket selling robot's right hand can be expressed by:

$$R = \int_{e_1}^{e_2} \varphi(e)de \Big/ \int_{-\infty}^{\infty} \varphi(e)de \tag{1}$$

As a matter of fact the movement error e is a sequence discrete sequence $E = (e_1, e_2, \cdots, e_n)$. Define an indicator:

$$\vartheta_{[e_1,e_2]}(e) = \begin{cases} 1 & e_1 \le e \le e_2 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Then the system reliability can be calculated accordingly

$$R = \sum_{i=1}^{n} \vartheta_{[e_1,e_2]}(e_i) e_i \Big/ \sum_{i=1}^{n} e_i \tag{3}$$

Based on the analysis of ticket selling robot example, it is not difficult to reveal that the investigation process of system reliability is just a process of investigation of a series of grey problems with respect to a grey system. Therefore the grey problems existing in the reliability investigation, analysis, and computation may be explored in the following aspects. Because a complex system is constituted by relevant hard subsystems and soft subsystems, the content of reliability analysis and computation of the system operating behaviour inevitably involves the reliability of every subsystem (no matter soft or hard one) and relationships between subsystems. In this context, the grey problem comes from the uncertainty and non-describable knowledge on the operating reliability of each individual subsystem and the operating coordination of the subsystems.

1. Time impacts on reliability of individual subsystems. Time fact affect subsystem reliability can be analyzed from two angles: materials constituting of the subsystems are wearing our and downgrading according to specifications in long term and the shape and strength of materials are changed associated with movements in short term.
2. Operating environment impacts on reliability of individual subsystem. Operation environment involves hard side, say, temperature, humidity, dust, light etc and soft side, say, work floor culture and in general company and local social culture environment. It is worth to stress here, the environmental factors interact with time factors and such relationships are difficult to evaluate and therefore a grey issue appears here for consideration.
3. Human behavior impacts on reliability of individual subsystem. In today's globalization environment, more and more complex systems are international-made. Inevitably the human

factors affect system reliability directly and indirectly during the system design, manufacturing, shipping and the end-usage operating. Even making the focus narrow to system operating, it is obvious the human and system (machine) interaction is often too complicated to describe. This will be a grey problem again. System reliability, denoted by R_S is dependent upon the system (machine) operating reliability, denoted by R_M and human operating reliability, denoted by R_H, which means the capability for the human (operator) finishes the specified job accurately within specified time. If using probability for expressing such capability, it is called operating reliability. A simplified description is treating human and machine as a series system and then the system reliability is simply $R_S = R_M \times R_H$. However, more and more complex system with automatic control subsystems and human (operator) restricted overriding function can not use simple series or parallel system to model them.

4. System design impacts on reliability of individual system. Today's quality starts at design stage. The allocation of reliability to individual subsystem is not known completely. As a matter of fact, the system operating behavior is unknown in principle before the system being manufacturing and putting into functioning. Therefore, the investigation and analysis of the complex system reliability is enabling the system information from grey state evolving into a whiter state. The methodology of the study is to analyzing the isomorphic characteristics of the systems falling in the same category and to find the key factors and the relationships between the constituting subsystems. According to the prior information from the isomorphic systems the optimal design of the system reliability and the reliability allocation to individual subsystem.

In summary, the grey problems appeared in complex system reliability analysis lie in shorting of system structural clarity, shorting of the underlying mechanism of the interaction between subsystems and shorting of overall information of the system as a whole. Accordingly, the methodology to solve the grey reliability of the complex system should be developed in terms of the basic concept of grey mathematics, grey relation analysis and grey differential equations because grey theory offers a rich structure for dynamic modelling with sparse data availability (Duan et al., 1998).

3.6.3 (Generalized) Grey Load-Strength Reliability Analysis

The so-called grey reliability of a complex system can be virtually treated as effect resulting from the interaction between grey load and grey strength of the system. If the complex system is a structure, say, a bridge, a roof, etc, the terms load and strength directly reflect the physical property of the structure. In general context, load and strength terms will represent virtual states of the complex system. In the concept of reliability, if the system grey load is less than system grey strength the system is safe and reliable otherwise, the system is unsafe and unreliable. In the viewpoint of grey theory, system grey reliability will be a two-fold grey event, denoted by $\otimes^2(M)=\otimes_x\otimes_r(M)$, and accordingly, the grey reliability of a complex system can be expressed as the event (Xin and Liu, 1996):

$$\otimes^2(M)=\left\{\otimes_x(L)\underset{\otimes_r}{<}\otimes_x(S)\right\} \tag{4}$$

where $\otimes_r$ represents grey symbol of information type; $\otimes_x(L)$ and $\otimes_x(S)$ represent the (generalized) grey load and grey strength respectively.

In grey mathematics, the grey event is described by the whitenization weight function since this weight grade of the whitenization function represents the degree of the preference of a grey number with respect to its domain (from which the grey number takes values). Intuitively, the witenization weight expressed the information availability. A typical withenization function is defined by,

$$w(x)=\begin{cases}L(x) & x\in[a_1,b_1)\\ 1 & x\in[b_1,b_2]\\ R(x) & x\in(b_2,a_2]\end{cases} \tag{5}$$

where $L(x)$ is monotone-increasing with $L(a_1)=0$ and $L(b_1)=1$ and $R(x)$ is a monotone-decreasing function with $R(b_2)=1$ and $R(a_2)=0$.

Accordingly, the (generalized) grey load $\otimes_x(L)\in(L_l,L_c,L_u)$ with whitenization weight function $w(L_l)=w(L_u)=0$ and $w(L_c)=1$, $L_c=(L_l+L_u)/2$ in addition to the features defined in (5). Similarly, the (generalized) grey strength $\otimes_x(S)\in(S_l,S_c,S_u)$ with whitenization weight function $w(S_l)=w(S_u)=0$ and $w(S_c)=1$, $S_c=(S_l+S_u)/2$.

Therefore, the degree of reliability of a complex system can be defined by the sum of whitenization weights of all possible whitenization events corresponding to the two-fold grey event $\left\{\otimes_x(L) \underset{\otimes_r}{<} \otimes_x(S)\right\}$, i.e.,

$$\mu = \sum_{i=1}^{n} w\left(\otimes^2(M)\right) \tag{6}$$

Since we are required to deal with the whitenization of a two-fold grey event, some discussions are necessary. In general a whitenization of any grey event must be carried by two operations. Denote $W^{(1)}$ as the operation that generates possible whitenization event and $W^{(2)}$ the operation to calculate the whitenization weight function, then the operation of whitenization of a grey event can be expressed as:

$$\otimes(M) = \left[W^{(2)}W^{(1)}\right]\otimes(M) \tag{7}$$

Recall that the whitenization of a two-fold grey event requires two whitenization operations and thus the whitenization operator of a two-fold grey event is:

$$W_t = W_r \bullet W_x \tag{8}$$

where W_x represents the whitenization operator of information type and W_r represents the whitenization operator of inequality.

As to the whitenization of information type for a grey number (or event), a discretization approach will be discussed. Given a set of n level values, denoted by $\{\alpha_i : 0 \le \alpha_i \le 1, i = 1,2,\cdots,n\}$, then for each level α_i, two α_i-cut intervals, denoted by $\left[L_l^{\alpha_i}, L_u^{\alpha_i}\right]$ and $\left[S_l^{\alpha_i}, S_u^{\alpha_i}\right]$ respectively, are obtained. Then the two α_i-cut intervals are two whitenization subintervals, which can be treated as two interval grey numbers at level α_i, i.e.,

$$\begin{cases} \otimes_x(L,\alpha_i) \in \left[L_l^{\alpha_i}, L_u^{\alpha_i}\right] \\ \otimes_x(S,\alpha_i) \in \left[S_l^{\alpha_i}, S_u^{\alpha_i}\right] \end{cases} \tag{9}$$

Then n whitenization events of information type are obtained

$$\tilde{\otimes}_x\left[\otimes_r(M,\alpha_i)\right] = \otimes_x(L,\alpha_i) \underset{\otimes_r}{<} \otimes_x(S,\alpha_i) \tag{10}$$

It is reasonable to define the discretized whitenization weight function as

$$W_x\left[\otimes_x\left(\otimes_r(M),\alpha_i\right)\right]=\alpha_i, i=1,2,\cdots,n \tag{11}$$

As to the whitenization of inequality type, a typical treatment is to represent a grey inequality by a grey interval number. Let

$$\otimes(\mathrm{I},\alpha_i)=\otimes_x(S,\alpha_i)-\otimes_x(L,\alpha_i) \tag{12}$$

Then based on equation system (9), a set of interval grey numbers can be obtained:

$$\left\{\otimes(\mathrm{I},\alpha_i)\in\left[\mathrm{I}_l^{\alpha_i},\mathrm{I}_u^{\alpha_i}\right], i=1,2,\cdots,n\right\} \tag{13}$$

where

$$\begin{cases}\mathrm{I}_l^{\alpha_i}=\min\left\{\otimes_x(S,\alpha_i)-\otimes_x(L,\alpha_i)\right\}\\ \mathrm{I}_u^{\alpha_i}=\max\left\{\otimes_x(S,\alpha_i)-\otimes_x(L,\alpha_i)\right\}\end{cases} \tag{14}$$

Performing normalization for grey interval I^{α_i}, it is obtained that

$$\otimes(\delta,\alpha_i)\in\left[\delta_l^{\alpha_i},\delta_u^{\alpha_i}\right]\subseteq[0,1] \tag{15}$$

where $\mathrm{I}_l^M=\max\left\{I_l^{\alpha_i}\right\}$ and $\mathrm{I}_u^M=\max\left\{I_u^{\alpha_i}\right\}$,

$$\begin{cases}\delta_l^{\alpha_i}=\mathrm{I}_l^{\alpha_i}/\mathrm{I}_l^M\\ \delta_u^{\alpha_i}=\mathrm{I}_u^{\alpha_i}/\mathrm{I}_u^M\end{cases} \tag{16}$$

A further whitenization treatment can be pursued by

$$\gamma^{\alpha_i}=\pi\delta_l^{\alpha_i}+(1-\pi)\delta_u^{\alpha_i}, \pi\in[0,1] \tag{17}$$

It can be seen that γ^{α_i} is dependent upon the whitenization level α_i, the whitenization weight functions of grey load and grey strength and their relative position. Furthermore, γ^{α_i} is a new independent quantity and plays a role of weights on the grey reliability. Define:

$$\gamma_i=\begin{cases}\gamma^{\alpha_i} & \gamma^{\alpha_i}\geq 0\\ 0 & \gamma^{\alpha_i}<0\end{cases} \tag{18}$$

and linear weights are assumed then the two-fold grey event can be expressed by

$$W\left(\otimes^2(M)\right)=\left\{W_x\left(\otimes_r\left(M^{\alpha_i}\right)\right)\bullet W_r\left(M^{\alpha_i}\right)\right\}=\{\alpha_i\gamma_i\} \tag{19}$$

Then the system grey reliability expressed in terms of equation 6,

$$\mu=\sum_{i=1}^{n}W\left(\otimes^2\left(M^{\alpha_i}\right)\right)=\sum_{i=1}^{n}\alpha_i\gamma_i \tag{20}$$

By taking $n\to\infty$, i.e., accelerating the step sizes of the discretization treatment, the continuous version of system grey reliability is obtained.

$$\mu=\int_0^1\gamma(\alpha)d\alpha \tag{21}$$

Finally, we must emphasize that the developments from equation 4 up to equation 21 should be regarded as a static grey reliability within a given time interval $[0,t_1]$, and should be denoted as $\mu(t_1)$. For a dynamic grey system reliability investigation, it is necessary to seek a sequence of μ, denoted by $\{\mu(t_k),k=1,2,\cdots,N\}$ for given time interval sequence $\{[0,t_k],t_k<t_{k+1},k=1,2,\cdots,N\}$, then the grey differential equation models, particularly, GM(1,1) model can be used to establish the grey system reliability function $\mu(t),t>0$ based on the grey reliability discrete sequence. As stressed in the Introduction, the dimension N of the discrete data sequence can be as small as 4.

3.6.4 Concluding Remarks

In this paper we briefly discussed the grey characteristics of complex system reliability. We further investigate evaluation approach in terms of the generalized load-strength concept and the grey characterization of load-strength as two-fold whitenization event. A detailed theoretical framework is established for the mathematical treatments of the two-fold whitenization event and the static version of system grey reliability is obtained. Also, we briefly point the way to obtain the dynamic version of system grey reliability with sparse data availability in terms of powerful grey differential equation models. As to how to extract the generalized load-strength functional and applications will be our future research topics.

3.7 GM(1,1)-Kriging Prediction of Soil Dioxin Pattern

Danni Guo[1], Renkuan Guo[1], Christien Thiart[1], Tonny Oyana[2], Dajun Dai[2], and Sarah Hession[3]

[1]Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

[2]Department of Geography, Southern Illinois Unversity Carbondale
Carbondale, IL 62901-4514, U.S.A.
tjoyana@siu.edu dljt@siu.edu

[3]State of Michigan, Department of Environmental Quality,
Remediation and Redevelopment Division, Lansing, MI 48909-7926, U.S.A.
hessions@michigan.gov

This paper has been published in the Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2006), April 5-7, 2006, University of Nottingham, Nottingham, U.K., pp. 69-77.

3.7.1 Introduction

Environmental data is very costly and difficult to collect, and quite often the sampled data are insufficient for further analysis. Kriging is a commonly accepted spatial interpolation method. The feasibility of kriging analysis depends on data quality and quantity. The quality of data is referred to as the adequacy of the data spreading for the assumed spatial prediction task. The quantity means whether the sample size is large enough. Today, we often face the circumstances where a set of data is collected already, although from the viewpoint of kriging analysis the data is insufficient, but re-sampling is impossible because of the cost and time limits. Therefore, a solution must be found. In this paper, we propose a mixed approach by combining grey differential equation models, particularly, GM(1,1) model and ordinary kriging approach together, GM(1,1)-kriging. We will expand the existing limited sample data available to produce a GM(1,1)-kriging map. The new approach addresses ill-designed spatial data sampling design and provides a better spatial analysis results. This approach is illustrated using a small sample of soil dioxin data collected from Midland County, Michigan State, U.S.A.

3.7.2 Example: Soil Dioxin in Midland County

Dioxin is a complicated family of chemicals that includes dioxins, furans, and PCBs that have related properties and toxicity (Alliance for Safe Alternatives, 2005). Dioxins are man-made chemical compounds that enter the air through fuel and waste emissions, including motor vehicle exhaust fumes and garbage incineration (Environmental Protection Agency, 2005b). It is not deliberately manufactured, but rather an unintended by-product of industrial processes that use or burn chlorine (Alliance for Safe Alternatives, 2005).

Dioxin is one of the most studied chemicals on the planet. It is found throughout the environment and in our food supply. It causes a wide range of adverse health effects including cancer, birth defects, diabetes, learning and developmental delays, endometriosis, and immune system abnormalities. It is the most potent animal carcinogen ever tested (World Health Organization, 1998; Alliance for Safe Alternatives, 2005). Most human exposure to dioxins occurs through the consumption of contaminated foods, especially animal fats (World Health Organization, 1998).

Dioxin emissions from incinerators reach people. Dioxin goes into the air and people breathe in the particles. But a bigger problem is that the particles settle on grazing land where cows eat the grass and the dioxin gets concentrated in the fat in their meat and milk. It also gets concentrated in cattle and hogs that are fed dioxin-tainted grain. It can also be found in soils at different places. Dioxins may also be carried in rain and watersheds. The dioxin particles can fall directly into rivers, streams, and other bodies of water or reach these waterways in surface water runoff. Dioxin settles on the bottom where fish and shellfish ingest small particles of sediment. Dioxin then builds up in their fat or organs (World Health Organization, 1998; Alliance for Safe Alternatives, 2005).

In this paper, we have used the soil dioxin data of Midland County, Michigan State, U.S.A. The reported dioxin concentration is referred by "TEQ". Toxic Equivalents, or TEQs, are used to report the toxicity-weighted masses of mixtures of PCDD/Fs (TRIfacts, 2005). The measurement units are in ng/kg (nanograms per kilogram) which is part per trillion (ppt). Soil dioxin samples are collected in Midland City and along the Tittabawassee River (figure 3.7.1).

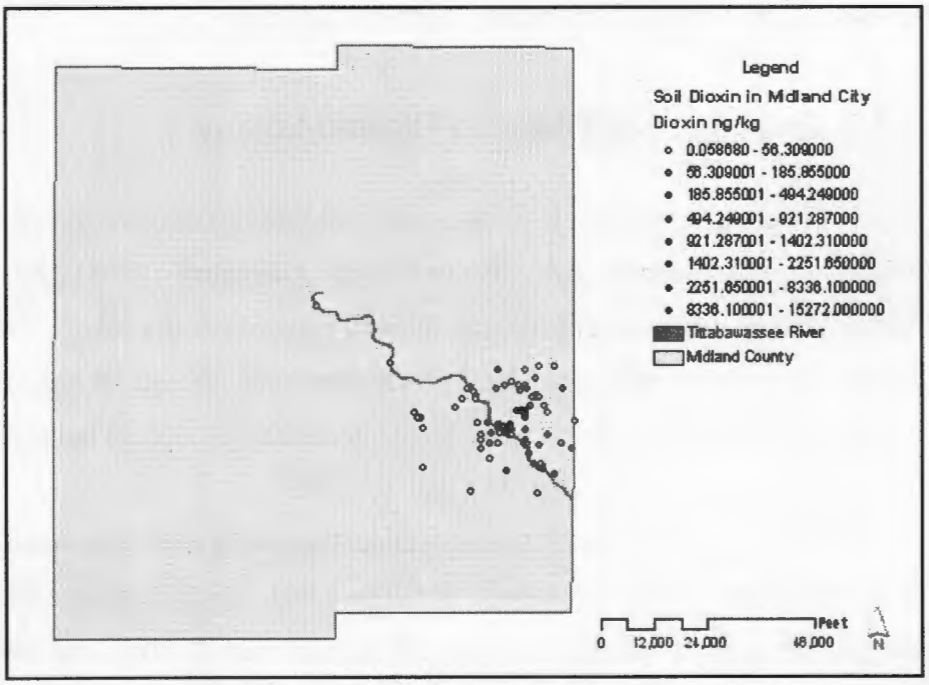


Figure 3.7.1 Soil Dioxin Samples in Midland County.

However, notice that these soil dioxin samples are concentrated in a small area of the Midland County. In this particular case, due to the spread of the sample data, a kriging map can be produced; however, it can only cover a small area around the samples (figure 3.7.2).

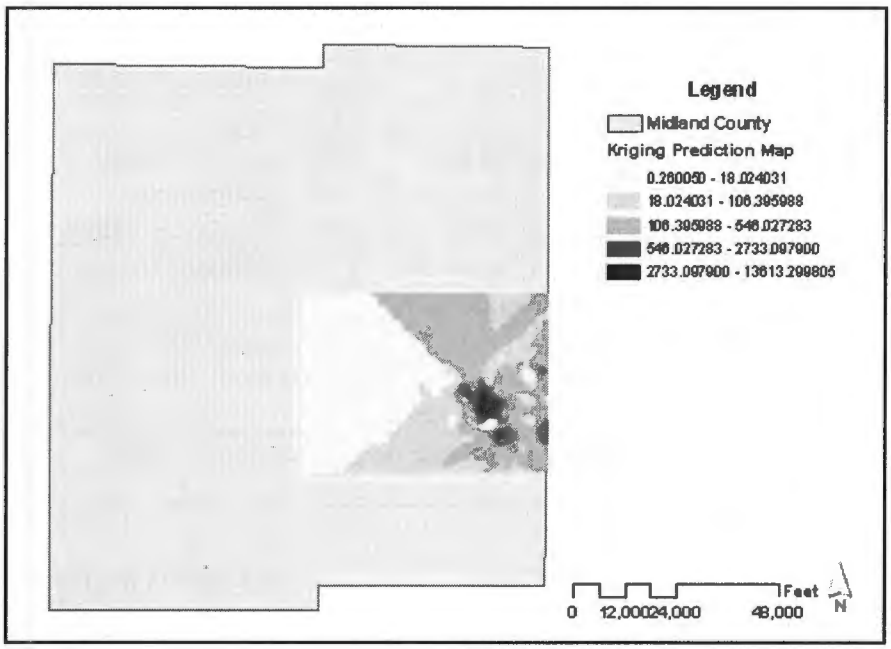


Figure 3.7.2 Kriging Map of Soil Dioxin in Midland County.

As you can see from above, in order to predict dioxin levels for the entire Midland County, more samples are need. Since sampling is a lengthy and expensive as well as time-consuming process, other options can be considered. In this case, grey predictions can be used to add to the sample data.

3.7.3 A Review of GM(1,1) Model in Grey System Theory

The Grey System Theory (Deng, 1982) was rooted in modern control theory, where system dynamics are classified by the degree of information availability, and accordingly the different methodologies are developed for each of them respectively. For a general picture of commonly faced three uncertain systems, we make a brief comparison in table 3.7.1.

A critical feature of the grey system is information incompleteness or more specifically, sparse information availability. The task of establishing a model a under the guidance of grey system theory is inevitably to build a model based on data of *small sample size*. Its target is the establishment of a grey differential equation and it emphasizes the exploration, utilization and processing dynamic information contained in the data (Guo, 2004; Guo and Love, 2005a, 2005b).

Table 3.7.1 Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.

Aspect	*Grey System*	*Probabilistic System*	*Fuzzy Set System*
Set foundation	Haze sets	Cantor sets	Fuzzy sets
Connotation and extension	Connotation haze with clear boundary and extension	Random event with connotation and extension well-defined	cognitive uncertainty (clear connotation but vague extension)
Core concept	Grey derivative and differential equations	Probability distribution	membership function
Data treatment	(inverse) accumulative generating operation	Sampling statistics & asymptotic distribution	Membership grade, λ-cut set & extension principle
Data requirements	Small sample size	Large sample size	empirical (+ sampling data)

Grey differential equation models play the core function in grey theory (Deng, 1982) and its modelling developments. In quality control context, GM(1,1), GM(2,1), and GM(1,N) are of fundamental importance. The basic one is GM(1,1) model.

Definition 1: Equation

$$x^{(0)}(k)+\alpha z^{(1)}(k)=\beta, \quad k=2,\cdots,n \tag{1}$$

is called a one-variable first order grey differential equation (GM(1,1)) with respect to time series sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)$, where

$$\begin{aligned} z^{(1)}(k) &= \frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right] \\ x^{(1)}(k) &= \sum_{i=1}^{k} x^{(0)}(i) \\ k &= 2,\cdots,n \end{aligned} \tag{2}$$

And α is called the developing coefficient, β is the grey input, and $x^{(0)}$ is a *grey derivative* which maximizes the information density for a given series to be modeled. This model is called GM(1,1) model with equal-gap.

Furthermore, the differential equation $dx^{(1)}/dt + \beta x^{(1)} = \alpha$ is called the whitenization differential equation or the *shadow* equation of the grey differential equation 1. The unknown parameter values (α,β) can be determined in terms of the classical least-square approach. Writing equation 1 as:

$$\beta+\alpha\left(-z^{(1)}(k)\right)=x^{(0)}(k),\ k=2,3,\cdots,n \tag{3}$$

Then a standard matrix form of the equation can be formed in terms of least-square theory,

$$X\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=y \tag{4}$$

where

$$X=\begin{bmatrix}-z^{(1)}(2) & 1\\ -z^{(1)}(3) & 1\\ \vdots & \vdots\\ -z^{(1)}(n) & 1\end{bmatrix}\ \text{and}\ y=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix} \tag{5}$$

which leads to the estimate for parameter (α,β).

$$\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=\left(X^TX\right)^{-1}X^Ty \tag{6}$$

Based on the estimates parameter (α,β) and differential equation theory, the predicted equation (i.e., the response or the filtering function) is,

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{\hat{\beta}}{\hat{\alpha}}\right]e^{-\hat{\alpha}k}+\frac{\hat{\beta}}{\hat{\alpha}} \tag{7}$$

which corresponds to the GM(1,1) differential equation,

$$\frac{dX^{(1)}}{dt}+\alpha X^{(1)}=\beta \tag{8}$$

As to the grey derivative sequence $X^{(0)}=\left\{x^{(0)}(i),\ i=1,2,\cdots,n\right\}$, it can be obtained in terms of the inverse accumulative generating operation:

$$\hat{x}^{(0)}(k)=\hat{x}^{(1)}(k)-\hat{x}^{(1)}(k-1),\ k=2,3,\cdots,n \tag{9}$$

For non-monotone data pattern, GM(2,1), the second order one variable grey differential equation, model may offer a better model-fitting. Therefore, let us briefly review the related developments (Liu et al., 2004).

Definition 2: Given the original discrete data sequence $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n))$, the second order one variable grey differential equation, abbreviated as GM(2,1), possesses the following form:

$$a^{(1)}x^{(0)}(k) + a_1 x^{(0)}(k) + a_2 z^{(1)}(k) = b,\ k = 2,3,\cdots,n \tag{10}$$

with the corresponding whitenization equation

$$\frac{d^2x^{(1)}}{dt^2} + a_1\frac{dx^{(1)}}{dt} + a_2 x^{(1)}(k) = b \tag{11}$$

where $a^{(1)}x^{(0)}(k) = x^{(0)}(k) - x^{(0)}(k-1)$, $k = 2,3,\cdots,n$ is the first order inverse AGO (i.e., 1-IAGO) and $Z^{(1)} = \{z^{(1)}(k),\ k = 2,3,\cdots,n\}$ are defined by equation 2.

The solution of GM(2,1) model is obtained by a typically two-step computation. The first step is to estimate the parameter vector $\mathrm{P}=(a_1, a_2, b)^T$. Let:

$$\mathrm{B} = \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ -x^{(0)}(4) & -z^{(1)}(4) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(n) & -z^{(1)}(n) & 1 \end{bmatrix} \quad \mathrm{Y} = \begin{bmatrix} a^{(1)}x^{(0)}(2) \\ a^{(1)}x^{(0)}(3) \\ a^{(1)}x^{(0)}(4) \\ \vdots \\ a^{(1)}x^{(0)}(n) \end{bmatrix} \tag{12}$$

Then the least-square estimator for the parameter vector $\mathrm{P}=(a_1, a_2, b)^T$ is:

$$\mathrm{P} = \left(\mathrm{B}^T\mathrm{B}\right)^{-1}\mathrm{B}^T\mathrm{Y} \tag{13}$$

The solution to the whitenization equation 14 which is a second order ordinary differential equation with constant coefficients, $x^{(1)} = x_P^{(1)} + x_H^{(1)}$ where $x_H^{(1)}$ is the general solution to the corresponding homogeneous second-order ordinary differential equation:

$$\frac{d^2x^{(1)}}{dt^2} + a_1\frac{dx^{(1)}}{dt} + a_2 x^{(1)}(k) = 0 \tag{14}$$

while $x_P^{(1)}$ is a particular solution to equation 11.

$$x_H^{(1)}(s) = \begin{cases} c_1 e^{r_1 s} + c_2 e^{r_2 s} \\ e^{\alpha s}\left(c_1 \cos(\beta s) + c_2 \sin(\beta s)\right) \end{cases} \tag{15}$$

which depends on form of the solutions r_1 and r_2 respectively. The particular solution is $x_P^{(1)} = b/a_2$ typically.

3.7.4 A Review of Ordinary Kriging Method

Ordinary kriging is a widely used geostatistical method for modelling spatial data (Cressie, 1991). It assumes that the local means are not necessarily closely related to the population mean, and therefore use only the samples in the local neighbourhood for estimate. Ordinary kriging relies on the spatial correlation structure of the data to determine the weighting values, and correlation between data points determines the estimated value at an unsampled point for inference, it makes the assumption of normality among the data points (Spatial Analysis and Decision Assistance, 2003).

In this study, we want to apply ordinary kriging to soil dioxin sample results generated from the GM(1,1) model. Once we have the GM(1,1) soil dioxin values, we could easily estimate additional values for the entire study region.

The basic mathematical idea behind ordinary krigingis to take N measurements $Z(r_1),\ldots,Z(r_N)$ of soil dioxin derived from the GM(1,1) model at known locations $r_1,\ldots r_N$, to obtain an estimate $\hat{Z}$ of Z at unsampled location r_0. We could use this linear equation to estimate neighbourly observations of soil dioxin.

$$\hat{Z}(r_0) = \sum_{i=1}^{N} \lambda_i Z(r_i) \tag{16}$$

Where λ_i is the weight assigned to each observation. We have only used closest observations within the searched neighbourhood to compute average weight and in the production of the estimate for soil dioxin values.

3.7.5 New Approach: Combining GM(1,1) Prediction at Equal-Spaced Grid with Ordinary Kriging

From figure 3.7.1, it is obvious that within Midland City and along the Tittabawassee River, we have rich data. However, the samples are limited to a small area making it difficult to perform an ordinary kriging analysis for the whole Midland County. First, we divide Midland County into 6 Areas (figure 3.7.3).

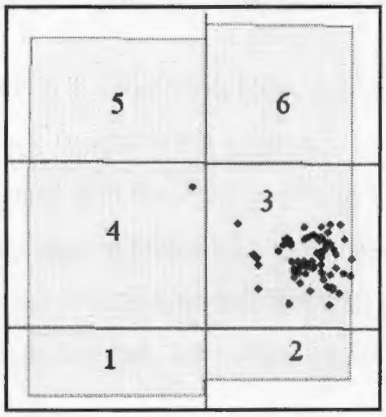


Figure 3.7.3 Area Division in Midland County.

An ordinary kriging map could only be produced for Area 3 (figure 3.7.2). Therefore, we divide Area 3 by equal-spaced vertical line V1-V5 and equal-spaced horizontal line H1 to H5 so that 25 intersection points can be recorded (figure 3.7.4).

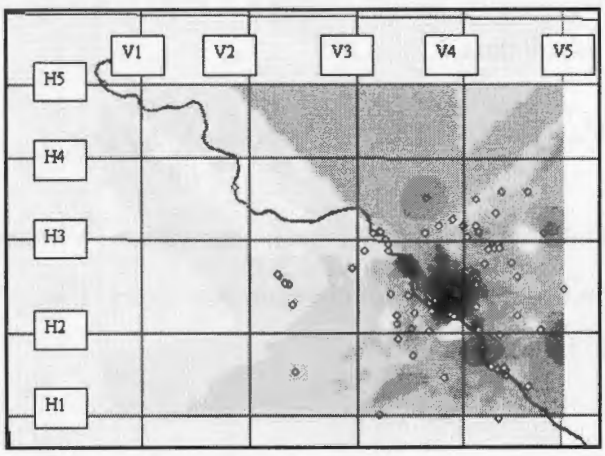


Figure 3.7.4 Grid Creation in Area 3.

Each intersection point will be recorded with northing coordinate, easting coordinate, and kriging soil dioxin values (table 3.7.2). Once the equal-spaced grids over Area 3 are created, we simply perform GM(1,1) modelling along vertical line V1 to V5 upward respectively with division points for covering Area 6. Similarly, we simply perform GM(1,1) modelling along vertical line V1 to V5 downward respectively with same spaced points for covering Area 2.

Table 3.7.2 Equal-spaced Coordinated Points with Kriging Dioxin values.

Point *ID*	*Easting*	*Northing*	*Kriging Dioxin Values*
1	13123134.239112	749177.692346	12.296460
2	13136803.482106	749313.031385	21.675454
3	13150202.047022	749177.692346	51.224597
4	13163600.611937	749177.692346	66.166684
5	13176051.803575	749177.692346	416.056757
6	13123269.578152	759598.798391	3.7297130
7	13136803.482106	759598.798391	13.725505
8	13150377.386061	759598.798391	36.230734
9	13163600.611937	759598.798391	0.0000000
10	13176187.142615	759734.137431	968.863086
11	13123269.578152	771237.955792	4.0243420
12	13136803.482106	771373.294832	5.0239170
13	13150202.047022	771237.955792	5.4372570
14	13163735.950976	771237.955792	79.625813
15	13176051.803575	771102.616753	398.105374
16	13123134.239112	781388.383758	16.7457250
17	13136938.821146	781388.383758	5.8539130
18	13150202.047022	781523.722798	161.311452
19	13163600.611937	781388.383758	176.095226
20	13176051.803575	781388.383758	265.932684
21	13123269.578152	790320.760368	5.8600940
22	13136803.482106	790320.760368	167.724327
23	13150337.386061	790320.760368	136.376999
24	13163735.950976	790320.760368	18.8408860
25	13176251.803575	790050.082289	205.140055

Area 4 can be GM(1,1)-predicted using horizontal lines H1 to H5 left-ward. Finally, we use Area 2 GM(1,1)-predicted points to GM(1,1)-predict Area 1. After Area 1, 2 and 6 are GM(1,1)-predicted, we essentially generate well-spread, equal-spaced grids over the unsampled areas. These GM(1,1)-predicted data are combined with originally sampled observations in Area 3 and forms an enlarged data set (figure 3.7.5).

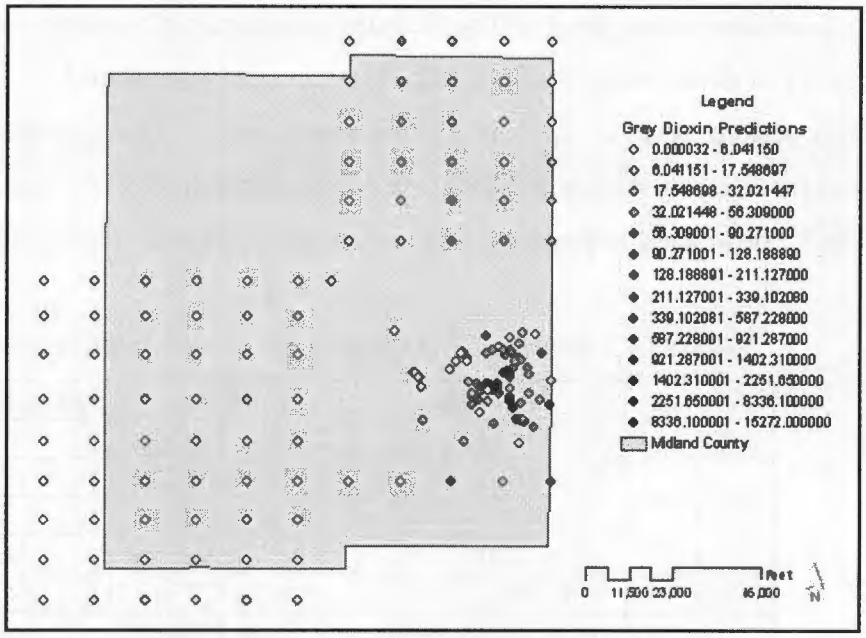


Figure 3.7.5 Grey Dioxin Predictions for Midland County.

The grey sample predictions appear as a grid of sample points (figure 3.7.5), expanding to cover much more of the Midland County. With these newly enlarged grey prediction dataset, an ordinary kriging map, called the GM(1,1)-kriging map, can now be produced for the entire Midland County (figure 3.7.6).

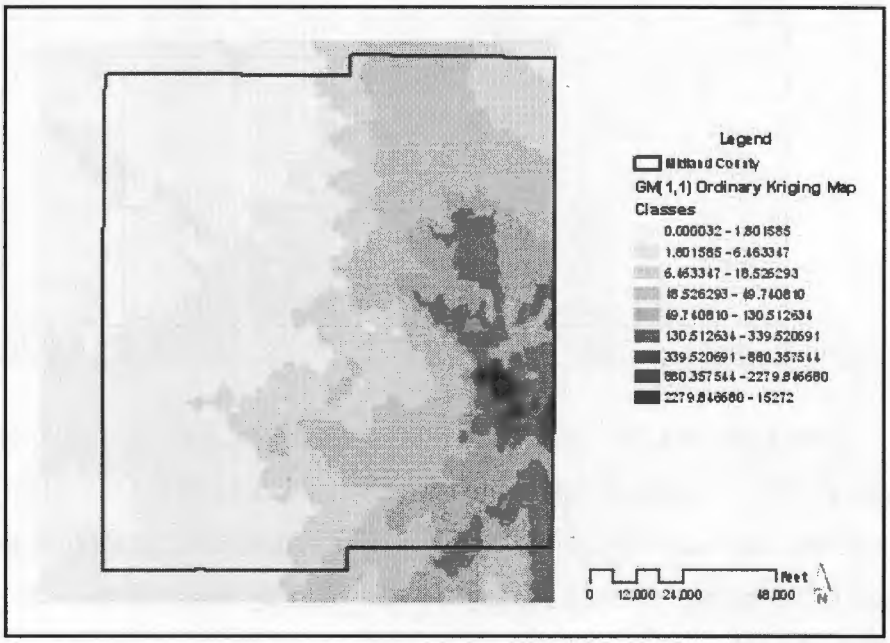


Figure 3.7.6 GM(1,1) Ordinary Kriging Map for Midland County.

Finally, looking at the GM(1,1) ordinary kriging map for the Midland County, one can clearly see that the GM(1,1)-kriging map can now cover the entire county. The predictions are clear and detailed, and consistent with the original soil dioxin samples.

3.7.6 Conclusion

In this paper, our purpose was two-fold: (1) to combine GM(1,1) model with ordinary kriging to account for small sample soil dioxin; and (2) then to solve the spatial prediction and analysis problem. GM(1,1)-ordinary method can better utilize any kind of data information because the new approach takes the advantages of GM(1,1) model: small sample size (as small as n=4), strong predictive ability and very simple numerical computations (Guo et al., 2005), which can be carried in Microsoft Excel. It is true that our work is just providing an illustration of the GM(1,1)-kriging and is still not refined. Further refinement of the GM(1,1)-ordinary kriging can be carried on by VBA programming. GM(1,1) modelling is less powerful for fluctuated data, therefore, GM(2,1) model may be engaged, that is the reason why we include the GM(2,1) model.

3.8 Predicting Air Pollution Using Fuzzy Membership Grade Kriging

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7707, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Journal of Computers, Environment and Urban Systems, Elsevier 2007, 31(1): 33-51. Editors: Andy P Jones and Iain Lake.

3.8.1 Introduction

Air pollution can affect our health in many ways, and PM_{10} (Particulate Matter) is one of the air pollutants that affect our health. PM_{10} is particulate matter with a mass median aerodynamic diameter less than 10 micrometers. PM_{10} is therefore very small and remains suspended in the air for long periods of time, it usually consists of smoke, dirt and dust particles, as well as mould, spores and pollen. PM_{10} is easily inhaled into the deep lung, and consequently the exposure to high outdoor PM_{10} concentrations will inevitably increase relevant diseases and cause death. The current standard for annual allowable arithmetic average of PM_{10} is not to exceed 50 micrograms per cubic meter of air (Environmental Protection Agency, 2005a).

In this paper, we will apply the fuzzy membership grade kriging technique to PM_{10} data of California, it is an extension of the previous we did on fuzzy kriging (Guo et al., 2004).

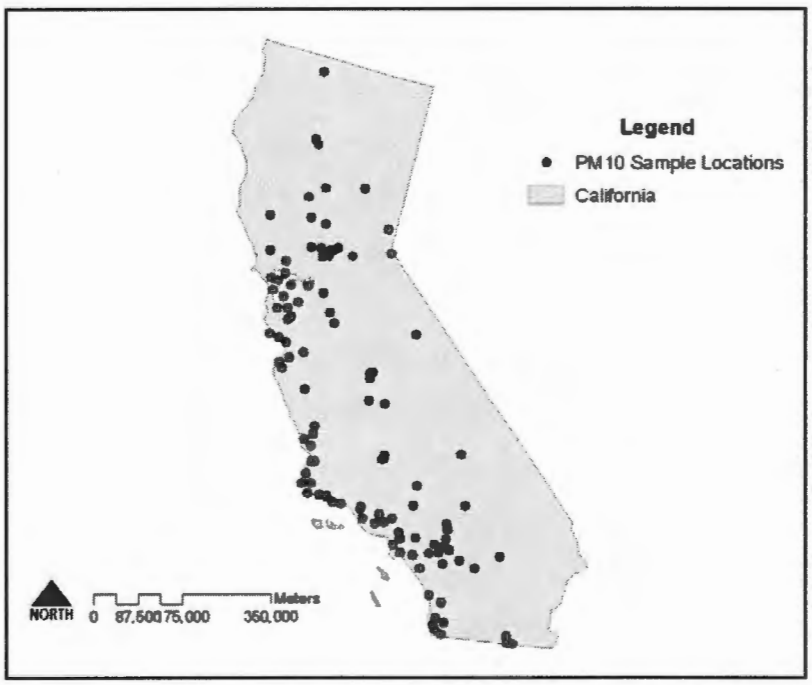


Figure 3.8.1 PM_{10} Sample Locations in California.

PM_{10} concentrations were collected in 111 locations in California (figure 3.8.1). However, it is too costly in terms of time, finance and manpower to collect PM_{10} samples over the entire state of California. Facing the problem of the lack of samples and lack of analysis of actual PM_{10} information, the modeller is keen to know the hazard level impacting their quality of life

regardless of their location being sampled or not. Furthermore it is noticed that the observed PM_{10} data were recorded as (positive real-valued) crisp numbers, but there exist impreciseness and vagueness with the recorded data. Therefore the correct way to analyze the PM_{10} data and predict the PM_{10} distribution over the whole California state by using the fuzzy membership grade kriging approach.

Kriging, initiated by Krige (1951), is the basic statistical methodology for predicting values at unsampled locations based on the indices sampled spatially surrounding the unsampled one. On the other hand, modelling a data set containing vagueness and impreciseness requires fuzzy mathematics, pioneered by Zadeh (1965). Therefore, in order to analyze spatially distributed imprecise data, it is inevitable to merge kriging and fuzzy mathematics together for creating a different approach-fuzzy kriging.

Fuzzy kriging can be traced back as early as Omre's Bayesian kriging paper (1987), and since then various fuzzy kriging methodologies for spatial data directly have being developed, e.g. Bardossy et al. (1988, 1989, 1990a, 1990b), Diamond, (1989), Lee (2000), Bandemer and Gebhardet (2000) etc. It is necessary to point out that in these fuzzy kriging approaches, referred as *direct fuzzy kriging*, the prediction is realized by directly kriging on spatially distributed fuzzy observations (i.e., fuzzy numbers). Direct fuzzy kriging is inevitably involved complicated mathematical operations on fuzzy sets and fuzzy statistical estimation and hypothesis testing issues and is difficult to implement in GIS.

For a simplified treatment, Guo (2004) proposed a *fuzzy membership grade kriging* methodology. The idea is actually very straightforward. Zadeh (1965) generalized the the {0,1}-two-valued indicator function which characterizes a crisp set into a [0,1]-infinite-valued membership function which defines a fuzzy set. Accordingly, Guo (2004) generalized Journel's (1983) threshold indicator coding, indicator variogram and indicator kriging into fuzzy membership grade, fuzzy membership grade variogram, and fuzzy membership grade kriging.

Following such simple route, we use membership transforms fuzzy data into membership grades and therefore the analysis is converted into kriging on spatial distributed *membership grades* (which are numbers from [0,1]). This approach is easy to understand and to implement in GIS (Guo et al., 2004).

3.8.2 Fuzzy Membership Grade Kriging

Kriging is an interpolation procedure used in geostatistics, using known values and a semivariogram to determine unknown values (Krige, 1951). The procedure incorporates measures of error and uncertainty when determining estimation (Gallagher, 1997). Classical kriging estimator is a linear predictor based on the spatial observations directly, while the fuzzy membership grade kriging estimator we developed is a linear predictor based on the membership grades of spatial observations. The fuzzy membership grade kriging approach proposed in this paper is a direct extension to Journel's (1983) indicator kriging. Therefore it is necessary to introduce the basics of indicator kriging.

3.8.2.1 Journel's Indicator Kriging

The usage of indicator function is a non-linear transformation for performing structural analysis emphasising the spatial distribution of grades at different threshold values. Journel (1983) proposed an indicator coding technique:

$$\vartheta_{\{z(\cdot)>z_k\}}\left(z(x_i)\right)=\begin{cases}1 & \text{if} \quad z(x_i)\in\{z(\cdot)>z_k\}\\ 0 & \text{if} \quad z(x_i)\notin\{z(\cdot)>z_k\}\end{cases} \tag{1}$$

where z_k is the predetermined threshold value and $z(x_i)$ is the sampled value at the i^{th} spatial location x_i. The indicator notation used here takes a set indicator form, $\vartheta_A(\cdot)$. It is a nonlinear transformation of the spatial data and therefore may improve the prediction substantially because of non-linear treatments. The indicator kriging utilizes an indicator variogram:

$$\gamma_z(h;z_k)=\frac{1}{2}\mathbb{E}\left[\left(\vartheta_{\{z(\cdot)>z_k\}}\left(z(x+h)\right)-\vartheta_{\{z(\cdot)>z_k\}}\left(z(x)\right)\right)^2\right] \tag{2}$$

where h represents the distance between two spatial locations x_i and x_i+h, z is the threshold value, and the sample values $z(x_i)$ and $z(x_i+h)$ are observed at x_i and x_i+h respectively. The indicator variogram can be expressed in terms of the related distributed functions:

$$\gamma_Z(h;z_k)=\frac{1}{2}\left[F_{Z(x)}(z_k)+F_{Z(x+h)}(z_k)-F_{Z(x+h),Z(x)}(z_k,z_k)\right] \tag{3}$$

In kriging analysis, the concept semivariogram uses covariance between two spatially distributed random observations for the revelation of the linear relationship between spatial observations. However, the *indicator semivariogram* is a probability, which reflects both linear and non-linear relationship between spatial observations.

The experimental indicator semivariogram is used in indicator kriging and defined by:

$$\hat{\gamma}_z(h;z_k)=\frac{1}{2N_h}\sum_{i=1}^{N_h}\left[\vartheta_{\{z(\cdot)>z_k\}}(z(x_i+h))-\vartheta_{\{z(\cdot)>z_k\}}(z(x_i))\right]^2 \tag{4}$$

N_h is the sample size within the diameter from the predicting point. Equation 5 represents the indicator kriging equation as an ordinary kriging on indicator values.

$$\begin{gathered}-\sum_{j=1}^{n}\lambda_j\gamma_z(x_i-x_j;z_k)+\gamma_z(x_0-x_i;z_k)-m=0,\quad i=1,\ldots,n \\ \sum_{i=1}^{n}\lambda_i=1\end{gathered} \tag{5}$$

And m is the lagrange multiplier.

Practically, we are required to replace (theoretical) $\gamma_z(\cdot;z_k)$ in equation 5 by (estimated) $\hat{\gamma}_z(\cdot;z_k)$ in equation 4 and find the estimated coefficients $\hat{\lambda}_i$, then the predicted indicator value (at the unsampled location x_0) is merely the linear function of these indicators with the estimated coefficients $\hat{\lambda}_i$ at the sampled locations:

$$\hat{\vartheta}_{\{z(\cdot)>z_k\}}(z(x_0))=\sum_{i=1}^{n}\hat{\lambda}_i\vartheta_{\{z(\cdot)>z_k\}}(z(x_i)) \tag{6}$$

It is obvious that Journel's kriging using a linear predictor of threshold indicator values of spatial observations. For details, see Cressie (1991) and Vega (2003).

3.8.2.2 Zadeh's Definition for Fuzzy Sets

Threshold indicator and indicator kriging opened a different way to perform spatial predictions. However, the threshold is a single value and therefore is lack of practical power, particularly facing a threshold range. For example, currently, PM_{10} hazardous level is identified as 50 (micrograms per cubic meter of air). As a matter of fact, the hazardous impacts for human body system should start at very low level (e.g. 30) and evolve to very high level (e.g. $\geq$50). For such an

unclear boundary (threshold) cases, indicator transformation will be less reflecting the true underlying mechanism because its {0,1}-two-value treatment. Facing the weakness, one of the remedies is using membership to replace indicator.

Recall that Zadeh (1965) define fuzzy set by extending the {0,1} two-valued indicator of a Cantor set into membership function taking values on interval [0,1]. In mathematical language, the indicator of a Cantor set A is:

$$\vartheta_A(x)=\begin{cases}1 & \text{if} \quad x\in A\\ 0 & \text{if} \quad x\notin A\end{cases} \tag{7}$$

A fuzzy set $\tilde{A}$ on a universe U is defined by the membership function:

$$\mu_{\tilde{A}}(u):U\rightarrow[0,1] \tag{8}$$

The membership grade $\mu_{\tilde{A}}(u)$ represents the degree of value u belonging to the fuzzy set $\tilde{A}$. If the grade takes value 1, u belongs to $\tilde{A}$ definitely; if the grade takes 0 value, u does not belong to fuzzy set $\tilde{A}$; otherwise u belongs to $\tilde{A}$ with grade $\mu_{\tilde{A}}(u)$ and does not belong $\mu_{\tilde{A}}(u)$ with grade 1- $\mu_{\tilde{A}}(u)$. Fuzzy set concept roots from some phenomena with clear connotation but unclear boundary. The link between a fuzzy set and a Cantor set is the so-called α -level cut set, A_α, $\alpha\in[0,1]$,

$$A_\alpha=\left\{u:\mu_{\tilde{A}}(u)\geq\alpha\right\} \tag{9}$$

It can be noticed that the indicator for the α -level cut set, A_α, $\alpha\in[0,1]$ is just:

$$\vartheta_{A_\alpha}(u)=\begin{cases}1 & \text{if} \quad \mu_{\tilde{A}}(u)\geq\alpha\\ 0 & \text{if} \quad \mu_{\tilde{A}}(u)<\alpha\end{cases} \tag{10}$$

For more details, see Vega (2003).

3.8.2.3 Fuzzy Membership Grade Kriging

Once we have the membership concept clarified and the relationship with indicator function being established. The proposed fuzzy membership grade kriging approach is very straightforward:

simply to substitute the threshold indicator function by fuzzy membership function (as its counterpart) from equation 1 to 6.

For clarity, we list the fuzzy membership grade kriging developments as equation 11 to 16. A typical fuzzy membership function can be defined by:

$$\mu_{\tilde{A}(z_1,z_2,z_3)}\left(z(\mathrm{x}_i)\right)=\begin{cases} l\left(z(\mathrm{x}_i)\right) & \text{if} \quad z_1 \le z(\mathrm{x}_i) < z_2 \\ r\left(z(\mathrm{x}_i)\right) & \text{if} \quad z_2 \le z(\mathrm{x}_i) < z_3 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where $l(\cdot)$ is a monotone-increasing function with $l(z_1)=0$, $l(z_2)=1$ and $r(\cdot)$ is a monotone-decreasing function with $r(z_2)=1, r(z_3)=0$. $\tilde{A}(z_1,z_2,z_3)$ represents the fuzzy set specified by equation 11, which dependents threshold values (z_1,z_2,z_3).

Definitely fuzzy membership function is not necessarily taking the form as equation 11. For example, for PM_{10} data analysis, the $r(\cdot)$ portion does not appear and the definition will be modified accordingly. $z(\mathrm{x}_i)$ represents the spatial observation taken at location x_i. More generally, we use notation $\tilde{A}(\{z_m\})$ to represent a fuzzy set specified by threshold sequence $\{z_m\}$.Then the fuzzy membership grade semivariogram is given by:

$$\gamma_z\left(h;\tilde{A}(\{z_m\})\right)=\frac{1}{2}\mathbb{E}\left[\left(\mu_{\tilde{A}(\{z_m\})}\left(\mathrm{z}(\mathrm{x}+h)\right)-\mu_{\tilde{A}(\{z_m\})}\left(\mathrm{z}(\mathrm{x})\right)\right)^2\right] \tag{12}$$

where h represents the distance between two spatial locations x_i and x_i+h, and z is the threshold value, and the sample values $z(\mathrm{x}_i)$ and $z(\mathrm{x}_i+h)$ are observed at x_i and x_i+h respectively. The fuzzy membership grade variogram can be expressed in terms of the related probabilities:

$$\begin{aligned}\gamma_z\left(h;\tilde{A}(\{z_m\})\right)=\frac{1}{2}\Big[&\Pr\left[Z(\mathrm{x}+h)\in\tilde{A}(\{z_m\})\right]+\Pr\left[Z(\mathrm{x})\in\tilde{A}(\{z_m\})\right]\Big]\\ &-\Pr\left[Z(x+\mathrm{h})\in\tilde{A},Z(\mathrm{x})\in\tilde{A}(\{z_m\})\right]\end{aligned} \tag{13}$$

The experimental fuzzy membership grade semivariogram is used in fuzzy membership grade kriging and is defined by:

$$\hat{\gamma}_z\left(h;\tilde{A}(\{z_m\})\right)=\frac{1}{2N_h}\sum_{i=1}^{N_h}\left[\mu_{\tilde{A}(\{z_m\})}\left(\mathrm{z}(\mathrm{x}+h)\right)-\mu_{\tilde{A}(\{z_m\})}\left(\mathrm{z}(\mathrm{x})\right)\right]^2 \tag{14}$$

Equation 15 represents the fuzzy membership grade kriging equation as an ordinary kriging on membership grades.

$$\begin{cases} -\sum_{j=1}^{n} \lambda_j \gamma_z\left(x_i - x_j; \tilde{A}\left(\{z_m\}\right)\right) + \gamma_z\left(x_0 - x_i; \tilde{A}\left(\{z_m\}\right)\right) - m = 0, i = 1,...,n \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{15}$$

The fuzzy membership grade predictor is then defined as:

$$p\left(\mu_{\tilde{A}(\{z_m\})}; B\right) = \sum_{i=1}^{n} \lambda_i \mu_{\tilde{A}(\{z_m\})}\left(z\left(x_i\right)\right), \quad \sum_{i=1}^{n} \lambda_i = 1 \tag{16}$$

Accordingly, the intrinsic hypothesis for $p\left(\mu_{\tilde{A}(\{z_m\})}; B\right)$ to be an unbiased estimator should be:

$$\mu_{\tilde{A}(\{z_m\})}\left(Z\left(x\right)\right) = \mu_0 + \varepsilon_0\left(x\right), \quad x \in D, \mu \in \mathbb{R} \tag{17}$$

For more mathematical details, see Guo (2003).

The core part of the fuzzy membership grade kriging is the specification of membership function. It is the common exercise that modeller assumes a function form for the fuzzy set to be investigated. If it generates good and valid modelling results, then we accept the assumed fuzzy membership function. Otherwise, some adjustments have to made for a more suitable fuzzy membership function. The adjusting and modelling should be carried on iteratively until model validation is confirmed.

Under some circumstances, modeller does not know the fuzzy membership function, however, the modeller has a sequence of α level cut sets $\left\{A_{\alpha_n}, n = 1,2,\cdots,K\right\}$ of the fuzzy set $\tilde{A}$. Then the membership function should be constructed by utilizing the information of α level cut sets $\left\{A_{\alpha_n}, n = 1,2,\cdots,K\right\}$.

3.8.2.4 An Algorithm for Fuzzy Membership Grade Variogram

From the previous discussions, it is clear that the fuzzy membership grade variogram plays a key role in fuzzy membership grade kriging. Therefore an algorithm is proposed for the theoretical evaluation of the fuzzy membership grade variogram. The algorithm can be carried out as follows:

Step 1: Generate a random sampling sequence $\{\alpha_k, k=1,2,...,N\}$ from a uniform distribution on [0,1].

Step 2: For any $i=n$, $n \in \{1,2,\cdots,N\}$, determine α_i-cut interval $[a_{\alpha_i}, b_{\alpha_i}]$ according to the membership function given by Equation 11, denoted as A_{α_i}.

Step 3: Transform sample data $\{z(\mathrm{x}_i), i=1,2,...,M\}$ into indicator data $\{\vartheta_{A_{\alpha_n}}(z(\mathrm{x}_i), i=1,2,...,M\}$. According to:

$$\vartheta_{A_{\alpha_n}}(z(\mathrm{x}_i)) = \begin{cases} 1 & \text{if} \quad a_{\alpha_n} \le z(\mathrm{x}_i) \le b_{\alpha_n} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$
$$\mathrm{x}_i \in D \subset \mathbb{R}^d, z \in \mathbb{R}, i=1,2,...,M$$

Step 4: Calculate the sample variogram at α_n-level:

$$2\hat{\gamma}_{A_{\alpha_n}}(h) = \frac{1}{N_n(h)} \sum_{j=1}^{N_n(h)} \left(\vartheta_{A_{\alpha_n}}\left(z\left(\mathrm{x}_j + \mathrm{h}\right)\right) - \vartheta_{A_{\alpha_n}}\left(z\left(\mathrm{x}_j\right)\right) \right) \tag{19}$$

Step 5: If $n=N$, then:

$$2\hat{\gamma}_z(h;\tilde{A}) = \frac{1}{N} \sum_{n=1}^{N} 2\hat{\gamma}_{A_{\alpha_n}}(\mathrm{h}) \tag{20}$$

Stop. Otherwise, $i=n+1$, go to Step 2. For details, see Guo (2003).

3.8.3 Applying the Fuzzy Membership Grade Kriging Technique to PM_{10} Data

We apply the fuzzy membership grade kriging technique to predict the PM_{10} spatial distribution over the entire California state. The modeller's interest is in whether living in California is safe or not, in terms of air pollution. Therefore fuzzy event $\tilde{A}$={High hazard level of PM_{10} in California} is worthwhile to try to investigate.

3.8.3.1 Linear Sample Membership Function

It is meaningful to argue that the higher the PM_{10} content in the air, the higher the degree of membership in fuzzy set $\tilde{A}$. It is reasonable to assume that the linear sample membership function

of the fuzzy event $\tilde{A}$ is linear. It has the feature of PM_{10} level at x=9.0 being assigned a 0 membership, and at x=90.3 being assigned a 1 membership. That is:

$$\hat{\mu}_{\tilde{A}(0,90.3,\infty)}(z)=\begin{cases}0 & \text{if} \quad 0\le z<9.0\\ (z-9)/81.3 & \text{if} \quad 9.0\le z<90.3\\ 1 & \text{if} \quad z\ge 90.3\end{cases} \tag{21}$$

The linear sample membership function $\mu_{\tilde{A}(0,90.3,\infty)}(x)$ is determined by threshold values $(0,90.3,\infty)$. It is simple and mathematically easily manipulated (figure 3.8.2).

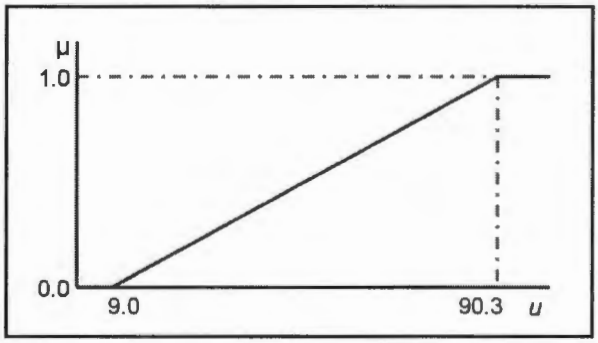


Figure 3.8.2 Linear Sample Membership Function of PM_{10}.

The QQ plot in figure 3.8.3 shows how well the model fits. A good fit is when the sample quantile distribution is very similar to the normal quantile line. In this plot, the middle sample values (dots) are very similar to the dashed straight line. However, the extreme upper and lower values fit less well. Figure 3.8.4 shows the predicted fuzzy membership grades. The dark coloured areas represent higher membership, which represent higher PM_{10} concentrations. The light coloured areas represent lower membership grades, which represent lower PM_{10} concentrations. Figure 3.8.5 shows the standard error. The light coloured areas represent lower error, and the dark coloured areas represent higher error. When one compares this map to the map of sample locations, one can understand the errors much better. Since most of the sample locations are near the coast, it is obvious that the errors are lower in the sampled locations, and higher in the unsampled locations.

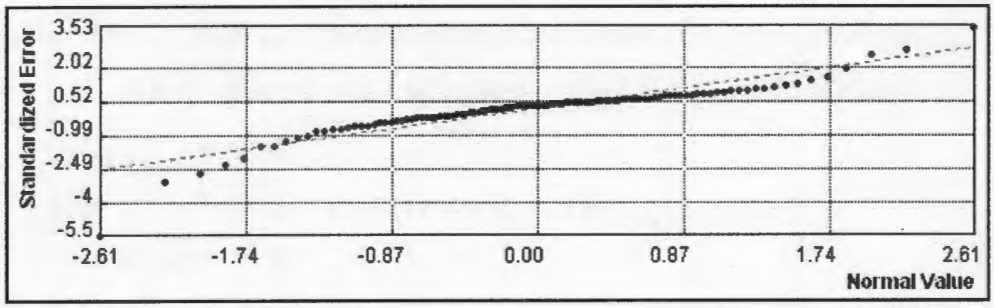


Figure 3.8.3 QQ Plot of Linear Sample Membership Function.

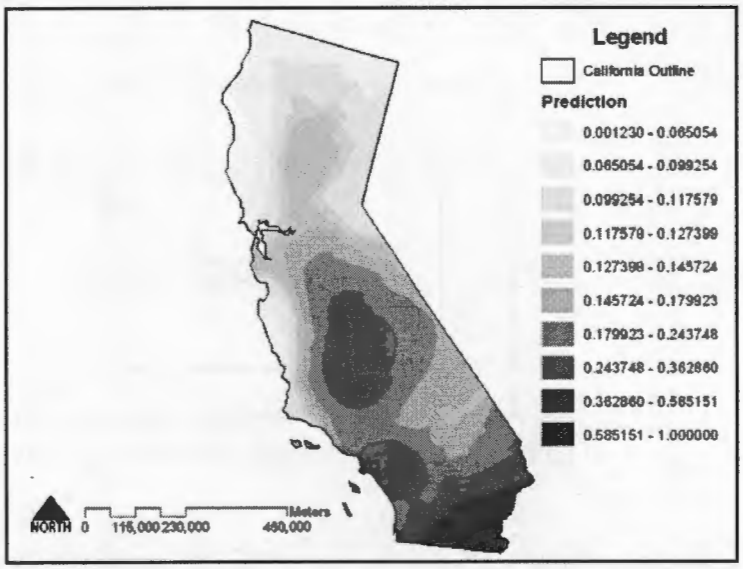


Figure 3.8.4 Predicted Fuzzy Membership Grades Using a Linear Sample Membership Function.

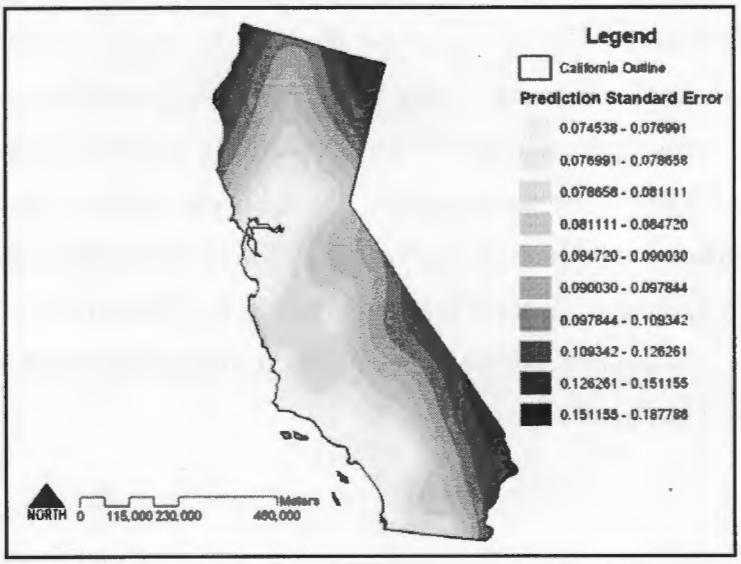


Figure 3.8.5 Standard Error Map Using the Linear Sample Membership Function.

3.8.3.2 Quadratic Sample Membership Function

Alternatively the fuzzy event $\tilde{A}$={High hazard level of in PM_{10} in California} might be a quadratic sample membership function, which offers the capacity to catch up the curvature of the evolution from stratum to stratum.

Again, the functional parameters of the quadratic sample membership function are data-determined. However, the curvature is up to the middle point selection since a quadratic equation has three parameters, *a*, *b*, and *c*.

$$\mu_{\tilde{A}}^{r}(z) = az^2 + bz + c \tag{22}$$

Solving the linear equation system will give the set of parameters $\left(\hat{a},\hat{b},\hat{c}\right)$, which is dependent on $\mu_{PM_{10}}$ value. $\mu_{PM_{10}}$=0.62421 is chosen for our problem, and deduced from the sample values.

$$\begin{cases} a(9.0)^2 + b(9.0) + c = 0 \\ a(41.1)^2 + b(41.1) + c = \mu_{PM_{10}} \\ a(90.3)^2 + b(90.3) + c = 1.0 \end{cases} \tag{23}$$

Solving these equations (23) for *a*, *b*, and *c*, result in the following membership function:

$$\mu_{\tilde{A}}^{r}(z) = \begin{cases} 0 & \text{if } 0 \le z < 9.0 \\ -1.4025\times10^{-4} z^2 + 2.6227\times10^{-2} z - 0.22468 & \text{if } 9.0 \le z < 90.3 \\ 1 & \text{if } z \ge 90.3 \end{cases} \tag{24}$$

The quadratic sample membership function is shown in figure 3.8.6.

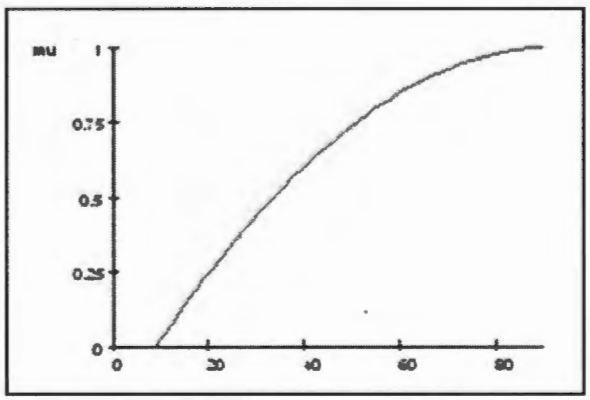


Figure 3.8.6 Quadratic Sample Membership Function of PM_{10}.

The QQ plot (figure 3.8.7) shows the middle sample values (dots) are very similar to the dashed straight line. The extreme upper and lower values fit the dash line much more closely, than in the linear sample membership. Figure 3.8.8 shows the predicted fuzzy membership grades. Notice that while this map is very similar to that of linear membership, however, the membership grades are very different. Figure 3.8.9 shows the standard error. Overall, this map does not seem too different from the linear error map, but certain inland areas do show a finer resolution.

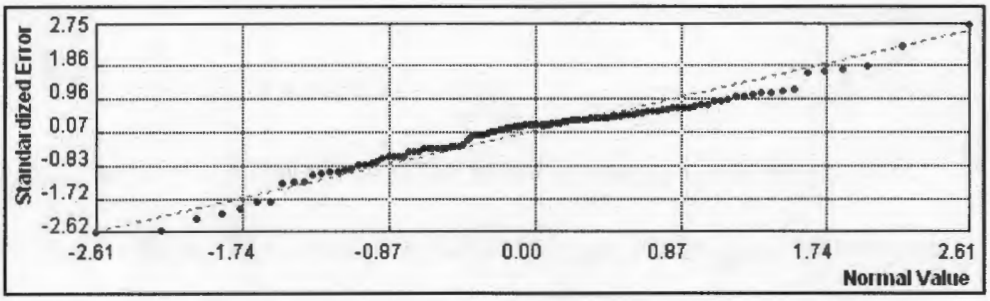


Figure 3.8.7 QQ Plot of Quadratic Sample Membership Function.

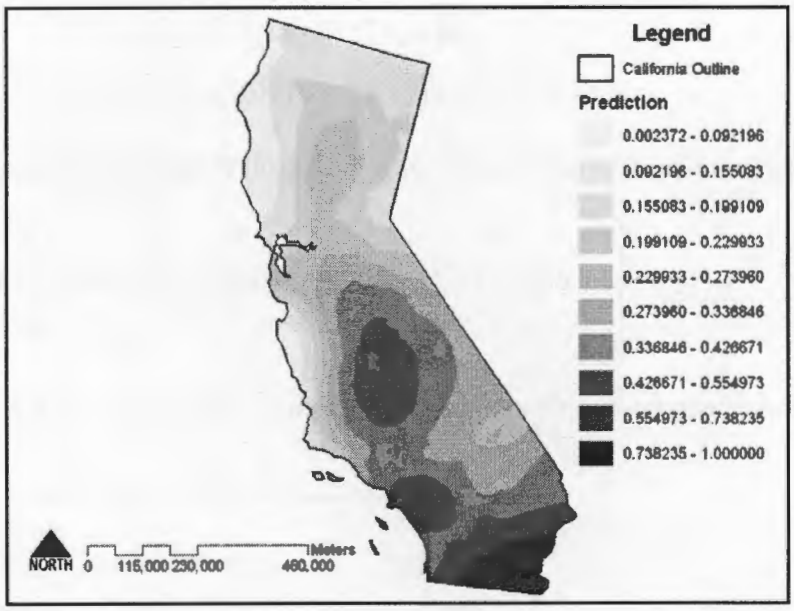


Figure 3.8.8 Predicted Fuzzy Membership Grades Using a Quadratic Sample Membership Function.

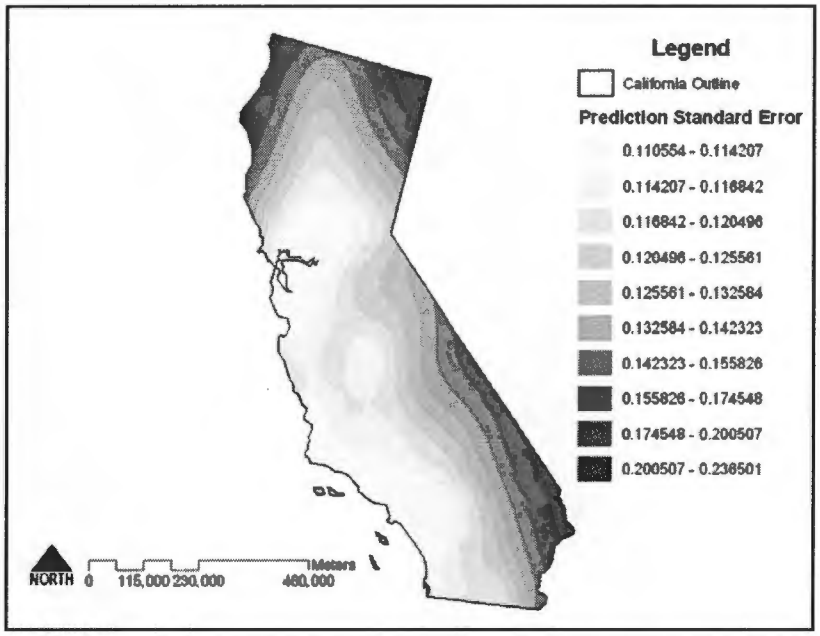


Figure 3.8.9 Standard Error Map Using the Quadratic Sample Membership Function.

3.8.3.3 Hyperbolic Tangent Sample Membership Function

The final membership function being considered for fuzzy event $\tilde{A}$={High hazard level of PM_{10} in California} is hyperbolic tangent sample membership function. The two unknown parameters (δ,η) specify the location and the scale of the sample membership function. It offers both the curvature catching capability of the evolution from stratum to stratum and smoothness over the whole range of PM_{10} value.

$$\frac{1}{2}\left(\tanh\left(\frac{z-\delta}{\eta}\right)+1\right) \tag{25}$$

The setting up of $\mu_{\tilde{A}}^{r}(x)$ is dependent upon the PM_{10} data. The hyperbolic tangent function tanh(x) ranges between -1and 1. Therefore, a vertical shift by 1 will secure a positive function. Furthermore, it should be normalized and ranges between [0,1] which is the requirement of a membership function.

$$\mu_{\tilde{A}}^{r}(z)=\frac{1}{2}\left(\tanh\left(\frac{z-22.1}{11}\right)+1\right) \tag{26}$$

We set $\left(\hat{\delta},\hat{\eta}\right)=(22.1,11)$ because 22.1 is the sample median of the PM_{10} observations, and $11 \approx q_{0.75} - q_{0.25} = 28.1 - 17 = 11.1$, the sample interquartile range. It is observed that when PM_{10} is 50 $\mu g/m^3$ or more, it is clearly indicated by membership grade value 1 (figure 3.8.10).

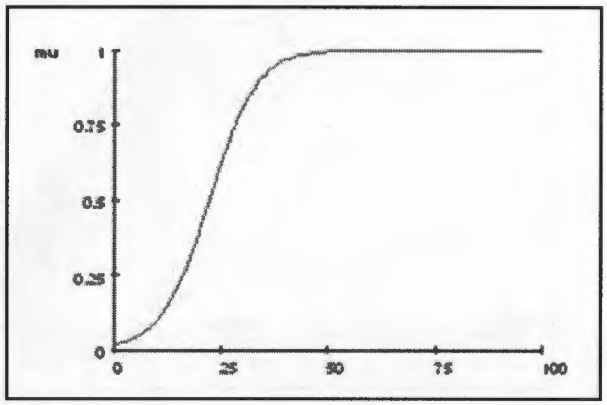


Figure 3.8.10 Hyperbolic Tangent Sample Membership Function of PM_{10}.

The QQ plot (figure 3.8.11) shows the middle sample values are very similar to the dashed straight line. The extreme upper and lower values do not fit the dash line well, but the middle values show improved correspondence. Figure 3.8.12 shows the predicted fuzzy membership grades. Notice that while this map is very similar to that of linear membership, and quadratic membership, however, the membership grades different in that it shows a finer resolution. Figure 3.8.13 shows the standard error. This map shows very different results from the linear and quadratic error maps. The class divisions are more refined and detailed.

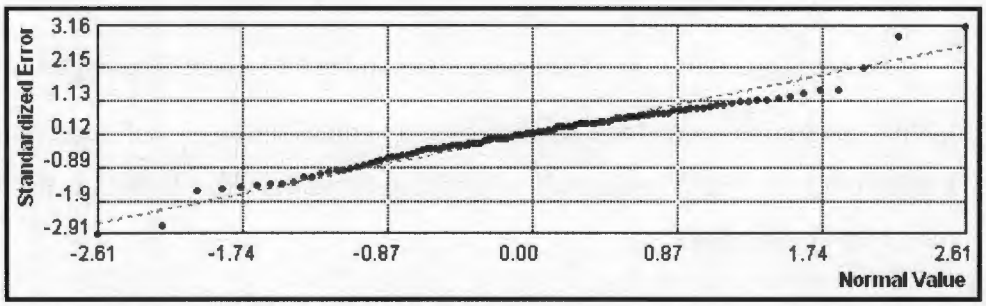


Figure 3.8.11 QQ Plot of Hyperbolic Tangent Sample Membership Function.

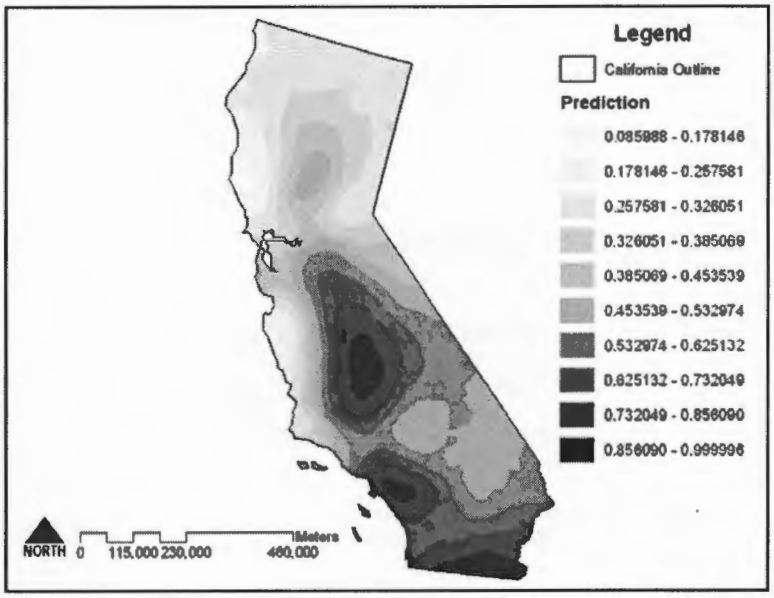


Figure 3.8.12 Predicted Fuzzy Membership Grades Using a Hyperbolic Tangent Sample Membership Function.

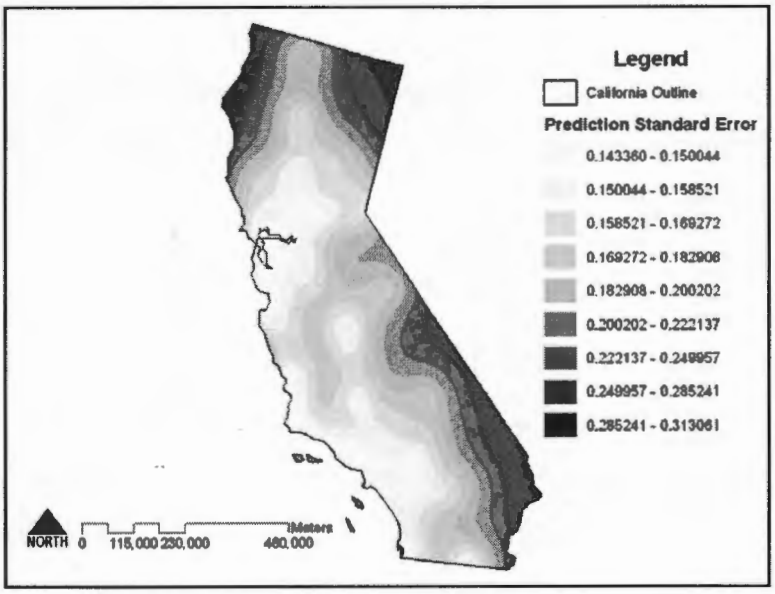


Figure 3.8.13 Standard Error Map Using the Hyperbolic Tangent Sample Membership Function.

3.8.4 GIS Analysis of PM_{10} Concentrations

Having produced the fuzzy membership grade kriging results, the interpretation of the fuzzy membership grade prediction map is a difficult task, particularly for those who do not have adequate background in fuzzy mathematics and statistics. Therefore, the predicted fuzzy

membership grades are converted back to PM_{10} concentrations to provide a more understandable format. The fuzzy membership grades are defined in terms of PM_{10} hazard levels. The membership functions are monotone-increasing continuous function, and so the inverse functions exist.

3.8.4.1 Linear Sample Membership

The linear sample membership function takes the form:

$$\mu(x) = \frac{x-a}{b-a}, a \le x \le b \tag{27}$$

For any given membership value $\mu_0 \in [0,1]$:

$$\mu_0 = \frac{x-a}{b-a} \tag{28}$$

then:

$$x^{linear}(\mu_0) = \mu_0(b-a)+a \tag{29}$$

Conversion is based on the fuzzy membership grade prediction map class interval limits. In the prediction map, it is typically classified into ten classes, e.g. class 0, 1,…,9. Assume the class interval limits are $[\mu_{l_0}, \mu_{u_0}), [\mu_{l_1}, \mu_{u_1}), ..., [\mu_{l_9}, \mu_{u_9})$.

The class limits for PM_{10} predicted value with linear sample membership function is:

$$[x_l, x_u) = [x^{linear}(\mu_l), x^{linear}(\mu_u)) \tag{30}$$

Looking at the converted prediction map in figure 3.8.14, one can analyse areas of high PM_{10} concentrations. In figure 3.8.14, the last two classes are considered as hazard areas. Once PM_{10} concentrations are over 38 $\mu g/m^3$, then it is classified as a hazard area. The level of safeness regarding PM_{10} concentrations is shown clearly here, by the class divisions of concentration level. The upper area of California and parts of the coastline are shown to have very low PM_{10} concentrations, and are the safest areas. The linear membership show much finer divisions of areas of low PM_{10} concentrations.

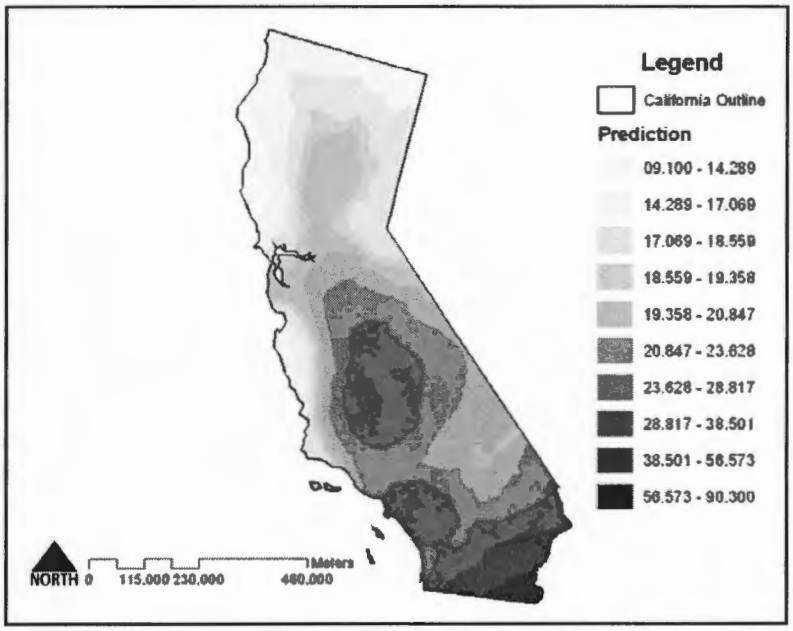


Figure 3.8.14 PM_{10} Concentrations Predicted by a Linear Membership Function.

3.8.4.2 Quadratic Sample Membership

The quadratic sample membership function takes the form:

$$\mu(x) = ax^2 + bx + c, c \le x \le d \tag{31}$$

For any given membership value $\mu_0 \in [0,1]$:

$$\mu_0 = ax^2 + bx + c \tag{32}$$

then:

$$ax^2 + bx + (c - \mu_0) = 0 \tag{33}$$

This leads to:

$$x^q(\mu_0) = \frac{1}{a}\left(-\frac{1}{2}b - \frac{1}{2}\sqrt{-4ac + 4a\mu_0 + b^2}\right) \tag{34}$$

or

$$x^q(\mu_0) = \frac{1}{a}\left(-\frac{1}{2}b + \frac{1}{2}\sqrt{-4ac + 4a\mu_0 + b^2}\right) \tag{35}$$

Depending on which x falls on $[c, d]$. The class limits for PM_{10} predicted value with quadratic sample membership function is:

$$\left[x_l, x_u\right) = \left[x^q(\mu_l), x^q(\mu_u)\right) \tag{36}$$

In figure 3.8.15, if the PM_{10} concentrations are over 37 $\mu g/m^3$, then it is classified as a hazard area. The upper area of California and parts of the coastline are shown to have very low PM_{10} concentrations, and are the safest areas. The lower parts and the middle regions of California are shown to have higher PM_{10} concentrations. The quadratic sample membership function show similar results to the sample linear membership function.

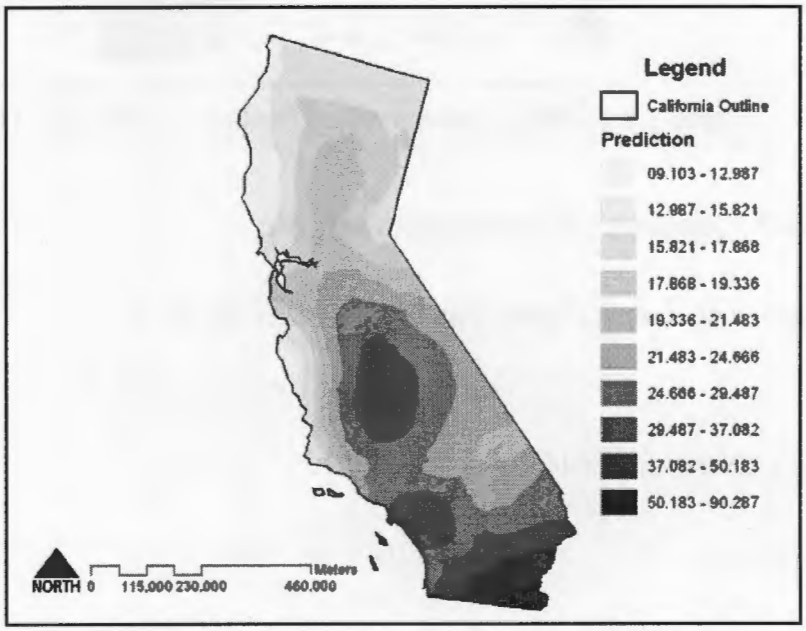


Figure 3.8.15 PM_{10} Concentrations Predicted by a Quadratic Membership Function.

3.8.4.3 Hyperbolic Tangent Sample Membership Function

The hyperbolic tangent sample membership function takes the form:

$$\mu(z) = \frac{1}{2}\left[\tanh\left(\frac{z-\delta}{\eta}\right)+1\right], \quad a \le z \le b \tag{37}$$

For any given membership value $\mu_0 \in [0,1]$:

$$\mu_0 = \frac{1}{2}\left[\tanh\left(\frac{x-\delta}{\eta}\right)+1\right] \tag{38}$$

then:

$$\frac{x-\delta}{\eta} = \frac{1}{2}\ln\left(\frac{\mu_0}{1-\mu_0}\right) \tag{39}$$

which leads to:

$$x^{\tanh}(\mu_0) = \frac{\eta}{2}\ln\left(\frac{\mu_0}{1-\mu_0}\right)+\delta \tag{40}$$

The class limits for PM_{10} predicted value with hyperbolic tangent sample membership function is:

$$[x_l, x_u) = \left[x^{\tanh}(\mu_l), x^{\tanh}(\mu_u)\right) \tag{41}$$

In figure 3.8.16, again the last two classes are hazard areas. Once PM_{10} concentrations are over 31 $\mu g/m^3$, it is classified as a hazard area. Areas of higher PM_{10} concentrations are in the middle regions and the bottom of California. The upper area of California and parts of the coastline are shown to have very low PM_{10} concentrations, and are the safest areas. The hyperbolic tangent membership show much finer divisions of areas of middle PM_{10} concentration levels.

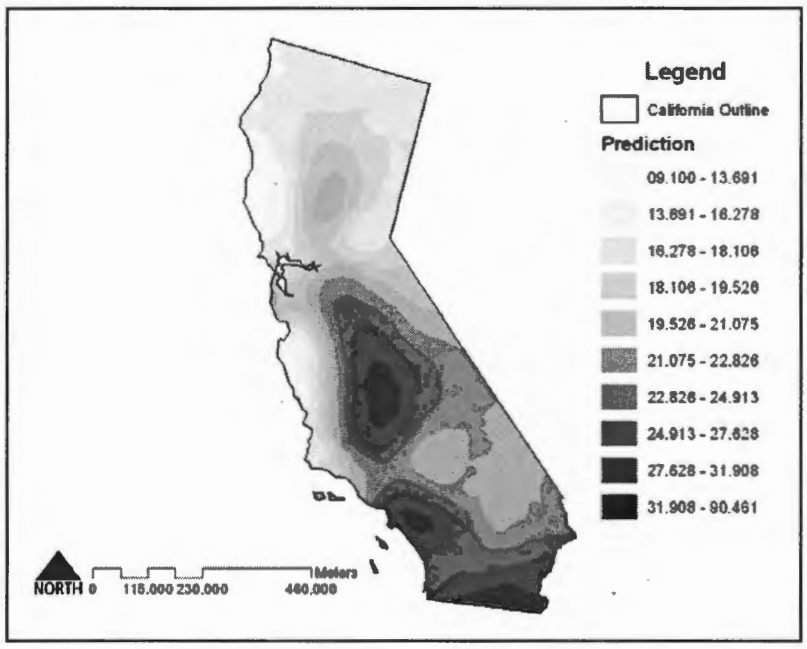


Figure 3.8.16 PM_{10} Concentrations Predicted by a Hyperbolic Tangent Membership Function.

3.8.5 Discussion

A new fuzzy membership function is developed, the hyperbolic tangent sample membership function. This sample membership function eliminated the piecewise or spline style definition of the sample membership function and it has fewer parameters. This will aid future prediction developments. The linear sample membership function has two parameters, and quadratric sample membership function has three parameters. Even though the hyperbolic tangent sample membership function has two parameters, the values are more refined.

In the traditional fuzzy mathematics literature, the membership function is assigned by the modeller. Thus the process is subjective, causing it to be criticised widely. Therefore in our modelling efforts, we follow a semi-statistical route by utilizing the sample membership function.

$$\mu_{\tilde{A}}^{s}: \begin{matrix} U \rightarrow & [0,1]^{ref} \\ u \rightarrow \mu_{\tilde{A}}^{s} & \in [0,1]^{ref} \end{matrix} \Bigg\} \tag{42}$$

Unlike previous studies that a membership function is hypothesized for the whole population, in this research, sample information was used to extract membership function following the work done by Chen (1998), Guo and Love (2003).

The sample membership functions are specified by our understanding of PM_{10} sample information by using descriptive statistics, histograms and fitted probability distributions. For example, the hyperbolic tangent sample membership function's δ location parameter simply equals the sample median, while the scale parameter η is from the sample's interquartile range. It is inevitable, that this technique contains some subjective factors. Although, in the current literature, almost all the membership function specifications are subjectively determined. However, the fuzzy membership grade kriging prediction part is fully statistical, and therefore objective.

In the model fitting evaluation, we rely heavily on graphic tools, i.e. the QQ plot of standardized error. However, statistically, we should rely more on hypothesis testing, particularly the Chi-Squared test.

3.8.6 Conclusion

In this paper, we proposed a *fuzzy membership grade kriging* approach for the spatial prediction of the PM_{10} concentrations in California. It is an exploratory data analysis. Three semi-statistical membership functions are produced and fuzzy membership grade kriging was performed. It is

noticed that the newly-introduced hyperbolic tangent membership function offers a more refining feature over the other two sample membership functions. The membership functions are semi-statistical because their sub-optimal nature. In order to seek an optimal statistical membership with solid mathematical foundation it is necessary to extract membership function from the data itself. Furthermore the fuzzy sets theory today can be established in terms of axiomatic foundation based on credibility (measure) theory (Liu et al., 2003). Our future work will provide more inside on fuzzy membership grade kriging based on the axiomatic foundations.

3.9 Rationale of GM(1,1) Modelling – A Variation Approach

Danni Guo, Renkuan Guo, Christien Thiart, and Yanhong Cui

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za ycui@stats.uct.ac.za

This paper has been published in the Proceedings of the International Conference on Maintenance Engineering (ICME 2006), October 15-18, 2006, Cheng Du, China, pp. 603-699.

3.9.1 Introduction

The first-order one-variable grey differential equation model, abbreviated as GM(1,1) model, requires very small sample size, as little as four sample points (Guo et al., 2005). Furthermore, the parameter estimation in GM(1,1) model is obtained via a simple regression, which is offered in any beginning statistics or data analysis course. A simple Excel spreadsheet can facilitate the necessary computations. Therefore the advantage for business managers and field or floor engineers to use GM(1,1) model in maintenance planning is very obvious. However, what is the rationale of GM(1,1) modelling on system lifetime is unclear and will become an obstacle, for reliability modellers with statistical analysis to accept GM(1,1) lifetime model and apply it in their practices. In this paper, we explore the underlying mechanism of GM(1,1) model, particularly, the lifetime partition into average functioning time, (average) repair-improvement and random error via GM(1,1).

3.9.2 A Brief Review on GM(1,1) model

Definition 1: (GM(1,1) model). Given a discrete positive real-valued (equal-spaced) data sequence $X^{(0)}=(x^{(0)}(1), x^{(0)}(2),\ldots, x^{(0)}(n))$, equation:

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

which is coupled with the differential equation,

$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha \tag{2}$$

is called a one-variable first order grey differential equation (GM(1,1)) with respect to time series sequence $X^{(0)}=(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, where:

$$\begin{aligned} z^{(1)}(k)&=\frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right] \\ x^{(1)}(k)&=\sum_{i=1}^{k} x^{(0)}(i) \\ k&=2,\cdots,n \end{aligned} \tag{3}$$

where β is called the developing coefficient, α is the grey input, and $x^{(0)}$ is called a *grey derivative* and $x^{(1)}(k)$ is called the k^{th} 1-AGO (accumulative generation operator) value.

Furthermore, the (coupled) differential equation $dx^{(1)}/dt + \beta x^{(1)} = \alpha$ (equation 2) is called the whitenization differential equation or the *shadow* equation of the grey differential equation in equation 1. The unknown parameter values (α,β) can be determined in terms of a standard L_2-optimization. A critical feature of (classical) GM(1,1) model proposed by Deng (1985) is performed with respect to a discrete positive data sequence.

3.9.3 Rationale exploration on GM(1,1)

We should emphasize here, the definition of GM(1,1) model couples a simple regression and a differential equation together. In grey theory literature, this coupling is explicitly implied by the 1-AGO term appeared in equation 1. However, here we state the definition of GM(1,1) in terms of coupled equations manner. Let us examine the logic underlying the coupling definition.

Mathematically, we can state equation 1 and 2 in the following optimization problem:

$$\begin{aligned} &\min_{\alpha,\beta,\gamma} \frac{1}{2}\sum_{i=1}^{n}\left(x^{(1)}(i)-x_{obs}^{(1)}(i)\right)^2 \\ &\text{s.t.} \\ &\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha \end{aligned} \tag{4}$$

This is a standard variational problem of searching a functional $x^{(1)}(t)$ minimizing $J=\left[\sum_{i=1}^{n}\left(x^{(1)}(i)-x_{obs}^{(1)}(i)\right)^2\right]\Big/2$ under a constant coefficient first-order linear ordinary differential equation $dx^{(1)}(t)/dt+\beta x^{(1)}(t)=\alpha$ as constraint, where $\gamma = x^{(0)}(0)$ is the initial value for the differential equation.

To obtain the parameter (α,β,γ) it is necessary to investigate the Lagrangian function $\mathfrak{L}$:

$$\mathfrak{L}=J+\mu\left(\gamma-x^{(1)}(0)\right)+\int_0^{T_f}\lambda(t)\left[\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)-\alpha\right]dt \tag{5}$$

where λ and μ are the Lagrangian multipliers. The necessary conditions for obtaining unconstrained minimum are:

$$\begin{cases} \dfrac{\partial \mathfrak{L}}{\partial x^{(1)}} = 0 \\ \dfrac{\partial \mathfrak{L}}{\partial M} = 0 \\ \dfrac{\partial \mathfrak{L}}{\partial \theta} = 0 \end{cases} \tag{6}$$

where $M^T = (\mu, \lambda)$ and $\theta^T = (\alpha, \beta, \gamma)$. The first equation of equation 14 leads to the so-called adjoint equation, the second equation recovers model equation (constraint) and the last equation gives the gradients with respect to (control parameter) $\theta^T = (\alpha, \beta, \gamma)$.

The Euler's Equation is:

$$\frac{\partial \mathfrak{L}}{\partial x^{(1)}} - \frac{d}{dt}\left[\frac{\partial \mathfrak{L}}{\partial \left(x^{(1)}\right)'}\right] = \left(x^{(1)} - x_{obs}^{(1)}\right) + \beta\lambda(t) - \lambda'(t) = 0 \tag{7}$$

And the adjoint equation is:

$$\begin{cases} \dfrac{d\lambda(t)}{dt} = \beta\lambda(t) + \left(x^{(1)} - x_{obs}^{(1)}\right) \\ \lambda(T_f) = 0 \end{cases} \tag{8}$$

The solution to the constraint is:

$$x^{(1)}(t) = \left(\gamma - \frac{\alpha}{\beta}\right)e^{-\beta t} + \frac{\alpha}{\beta} \tag{9}$$

Then we can obtain the expression for $\lambda(t)$ via multiple factor $e^{\beta t}$,

$$\begin{aligned} \lambda(t) = &\left[\frac{1}{2\beta}\left(\gamma - \frac{\alpha}{\beta}\right)e^{-2\beta T_f} + \frac{1}{\beta}\left(\frac{\alpha}{\beta} - x_{obs}^{(1)}(0)\right)e^{-\beta T_f}\right]e^{\beta t} \\ &- \frac{1}{2\beta}\left(\gamma - \frac{\alpha}{\beta}\right)e^{-2\beta t} - \frac{1}{\beta}\left(\frac{\alpha}{\beta} - x_{obs}^{(1)}(0)\right) \end{aligned} \tag{10}$$

The next step is to derive the first-order and second-order partial derivatives of J with respect to α, β, and γ respectively. Based on the equation $\partial \mathfrak{L}/\partial\theta = 0$,

$$\frac{\partial J}{\partial \theta} = -\int_0^{T_f} \left[\frac{\partial F}{\partial \theta}\right] \lambda(t)\, dt \tag{11}$$

where,

$$\begin{cases} F\left(x^{(1)}, \theta\right) = \dfrac{dx^{(1)}}{dt} = \alpha - \beta x^{(1)} \\ x^{(1)}(0) = \gamma \end{cases} \tag{12}$$

Then,

$$\begin{aligned} F\left(x^{(1)}, \theta\right) &= \alpha - \beta x^{(1)} \\ &= \alpha - \beta\left[\left(\gamma - \frac{\alpha}{\beta}\right)e^{-\beta t} + \frac{\alpha}{\beta}\right] \\ &= (\alpha - \beta\gamma)e^{-\beta t} \end{aligned} \tag{13}$$

which gives:

$$\begin{cases} \dfrac{\partial F}{\partial \alpha} = e^{-\beta t} \\ \dfrac{\partial F}{\partial \beta} = -\left[\gamma + (\alpha - \beta\gamma)t\right]e^{-\beta t} \\ \dfrac{\partial F}{\partial \gamma} = -\beta e^{-\beta t} \end{cases} \tag{14}$$

3.9.4 The Implementation of Control Parameter Searching

The searching of control parameter (α, β, γ) is just an application of Newton-Raphson method. Note that the gradient vector is:

$$\begin{pmatrix} \frac{\partial J}{\partial \alpha} \\ \frac{\partial J}{\partial \beta} \\ \frac{\partial J}{\partial \gamma} \end{pmatrix} = \begin{bmatrix} -\int_0^{T_f} e^{-\beta t}\ \lambda(t)dt \\ \int_0^{T_f} [\gamma + (\alpha - \beta\gamma)t] e^{-\beta t}\ \lambda(t)dt \\ \beta \int_0^{T_f} e^{-\beta t}\ \lambda(t)dt \end{bmatrix} \tag{15}$$

Recall that the searching parameters, i.e., (α, β, γ), is actually an assimilating process with data sampled and therefore, the integrals in equation 15 should be replaced by the summations respectively. In other words, equation 15 becomes,

$$\begin{pmatrix} \frac{\partial J}{\partial \alpha} \\ \frac{\partial J}{\partial \beta} \\ \frac{\partial J}{\partial \gamma} \end{pmatrix} = \begin{bmatrix} -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\lambda(i) \\ \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i] e^{-\beta i}\lambda(i) \\ \beta \frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\lambda(i) \end{bmatrix} \tag{16}$$

Then the second-order partial derivatives can be obtained:

$$\begin{aligned} \frac{\partial^2 J}{\partial \alpha^2} &= -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \alpha} \\ \frac{\partial^2 J}{\partial \alpha \partial \beta} &= \frac{1}{n}\sum_{i=1}^{n} i e^{-\beta i}\lambda(i) - \frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \beta} \\ \frac{\partial^2 J}{\partial \alpha \partial \gamma} &= -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \gamma} \\ \frac{\partial^2 J}{\partial \beta^2} &= \frac{1}{n}\sum_{i=1}^{n} (-\gamma i) e^{-\beta i}\lambda(i) + \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i](-i) e^{-\beta i}\lambda(i) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i] e^{-\beta i}\frac{\partial \lambda(i)}{\partial \beta} \end{aligned} \tag{17}$$

The iterations follows equation 18 until convergence criterion is satisfied.

$$\begin{pmatrix} \tilde{\alpha}^{(n+1)} \\ \tilde{\beta}^{(n+1)} \\ \tilde{\gamma}^{(n+1)} \end{pmatrix} \approx \begin{pmatrix} \tilde{\alpha}^{(n)} \\ \tilde{\beta}^{(n)} \\ \tilde{\gamma}^{(n)} \end{pmatrix} - \left(\frac{\partial^2 J(\theta^{(n)})}{\partial \theta_i^{(n)} \partial \theta_j^{(n)}} \right)^{-1} \begin{pmatrix} \partial J(\theta^{(n)}) / \partial \alpha \\ \partial J(\theta^{(n)}) / \partial \beta \\ \partial J(\theta^{(n)}) / \partial \gamma \end{pmatrix} \tag{18}$$

We need to point our that at the k^{th} iteration, the resulting estimate $\theta^{(k+1)} = \upsilon(\theta^{(k)}; X^{(0)})$.

3.9.5 The Small Sample Asymptotic Distribution of GM(1,1) Model

Although we stated that GM(1,1) model is a non-probabilistic modelling on a positive discrete data sequence it is still beneficial to examine the small sample asymptotic distribution of GM(1,1) model by assuming the sampled data is *i.i.d.* from a given distribution Ψ. An *M*-estimator for parameter vector $\theta = (\theta_1, \theta_2, \theta_3)^T = (\alpha, \beta, \gamma)^T$ satisfies the following equation system.

$$\sum_{i=1}^{n} \psi_j(x_i, t) = 0, j = 1,2,3 \tag{19}$$

The forms of $\psi_j, j = 1,2,3$ are defined equation system $\partial \mathfrak{L} / \partial \theta = 0$ defines the so-called *M*-estimator of parameter θ if we assume the data sequence is *i.i.d.* from a distribution Ψ. Under assumptions A4.1M-A4.5M stated by Field and Ronchetti (1990), an asymptotic expansion for the density T_n, the solutions of equation 19, denoted by $f_n(\cdot)$, is:

$$f_n(t_0) = \left(\frac{n}{2\pi}\right)^{3/2} c^{-n}(t_0) |\det A| |\det \Sigma|^{-1} \left(1 + o\left(\frac{1}{n}\right)\right) \tag{20}$$

where functions $\varpi_j(t_0)$,j=1,2,3 is the solution to equation 21:

$$\begin{aligned} &\int \psi_l(x, t_0) \exp\left(\sum_{j=1}^{3} \varpi_j(t_0) \psi_j(x, t_0)\right) f(x) dx, \; l = 1,2,3 \\ &c^{-1}(t_0) = \int \exp\left(\sum_{j=1}^{3} \varpi_j(t_0) \psi_j(x, t_0)\right) f(x) dx \\ &A = \left[E\left(\frac{\partial \psi(x,t)}{\partial t} \right) \right]_{t=t_0}, \; \Sigma = E\left(\psi_j(x, t_0) \psi_l(x, t_0)\right)_{1 \le j,l \le 3} \end{aligned} \tag{21}$$

And all expectations are taken with respect to the conjugate density,

$$h_{t_0}(x) = c(t_0) \exp\left(\sum_{j=1}^{3} \varpi_j(t_0) \psi_j(x, t_0)\right) f(x) \tag{22}$$

The error term holds uniformly for all t_0 in a compact set.

Once the approximate asymptotic joint density for $t_0 = (t_1^{(0)}, t_2^{(0)}, t_3^{(0)})$ is available, the small asymptotic densities for the two models: one for regression model and another for the solution to the constraint (shadow) differential equation will be ready for usages.

The regression model takes the form:

$$x_{reg}^{(0)}(k) = \alpha + \beta\left(-z^{(1)}(k)\right) \tag{23}$$

and the solution takes the form,

$$x_{shadow}^{(0)}(k) = \left(e^{-\beta} - 1\right)\left(\gamma - \frac{\alpha}{\beta}\right)e^{-\beta(k-1)} \tag{24}$$

In terms of transformations, we can obtain the two asymptotic densities, denoted by $f_{reg}(\cdot)$ and $f_{shadow}(\cdot)$ respectively. For a given error $\varepsilon > 0$, if $\left|f_{reg}(x) - f_{shadow}(x)\right| < \varepsilon$ we will accept the fact that the degree of assimilation is high. In other words, the coupling nature is satisfied.

3.9.6 Conclusion

In this paper, we investigate the nature of GM(1,1) model in terms of a variational viewpoint. Form the analysis performed in this paper, we conclude that standard GM(1,1) model is a variational problem with differential equation as its constraint. Furthermore, the efficiency of GM(1,1) model depends upon the choice of the constraint (shadow) differential equation. We also point out that a variational treatment of GM(1,1) model is already complicated and if we go further to investigate the asymptotic distributions as well as model efficiency from small asymptotic theory would not help us a better probabilistic modelling treatment.

The clear message from the variational analysis is GM(1,1) model proposed by Deng (1985) do enjoy the easiness and efficiency in modelling the discrete positive data sequence as long as the 1-AGO data sequence possesses grey exponential trend. For GM(1,1) modelling improvement, there are two ways: the first approach is the GM(1,1) modelling on residuals proposed by Guo and Cui (2006) and the second approach is using the extended GM(1,1) model based on the coupling principle (Guo et al., 2006a, 2006b).

3.10 Generalizations to Standard GM(1,1) Model

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the International Conference on Maintenance Engineering (ICME 2006), October 15-18, 2006, Cheng Du, China, pp. 587-593.

3.10.1 Introduction

The first-order one-variable grey differential equation model, abbreviated as GM(1,1) model, requires very small sample size, as little as four sample points (Guo et al., 2005). Furthermore, the parameter estimation in GM(1,1) model is obtained via a simple regression, which is offered in any beginning statistics or data analysis course. A simple Excel spreadsheet can facilitate the necessary computations. Therefore the advantage for modellers to use the GM(1,1) model in maintenance planning is very obvious. However, what is the rationale of GM(1,1) modelling on system lifetime is unclear and will become an obstacle, for modellers with statistical analysis to accept GM(1,1) lifetime model and apply it in their practices.

In this paper, we examine the three basic component models – simple regression model, differential equation model, and difference model in GM(1,1) model. The component-level model explorations will lead provide a clear picture of GM(1,1) modelling dynamics and we summarize it as coupling principle for the formation of GM(1,1) model. Furthermore, we explore the potential extensions to Deng's GM(1,1) model and expect more efficient extended GM(1,1) model could be developed in the future.

3.10.2 A Brief Review on the Standard GM(1,1) Model

The standard GM(1,1) model, proposed by Deng (1985), is defined by:

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

which is coupled with the differential equation:

$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha \tag{2}$$

These pair of equations are called a one-variable first order grey differential equation (GM(1,1) model) with respect to positive discrete data sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$ with time as index, where:

$$
\begin{aligned}
z^{(1)}(k) &= \frac{1}{2}\left[x^{(1)}(k) + x^{(1)}(k-1)\right] \\
x^{(1)}(k) &= \sum_{i=1}^{k} x^{(0)}(i) \\
k &= 2, \cdots, n
\end{aligned}
\tag{3}
$$

where β is called the developing coefficient, α is the grey input, and $x^{(0)}$ is called a *grey derivative* and $x^{(1)}(k)$ is called the k^{th} 1-AGO (accumulative generation operator) value. The differential equation $dx^{(1)}/dt + \beta\ x^{(1)} = \alpha$ (equation 2) in the GM(1,1) model is called the whitenization differential equation or the *shadow* equation of the grey differential equation 1.

The unknown parameter values (α, β) can be determined in terms of a standard simple regression modelling. The estimate (a, b) of parameter pair (α, β) is given by:

$$
(a,b)^T = \left(X^T X\right)^{-1} X^T Y \tag{4}
$$

where,

$$
X = \begin{bmatrix} 1 & -z^{(1)}(2) \\ 1 & -z^{(1)}(3) \\ \vdots & \vdots \\ 1 & -z^{(1)}(n) \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{5}
$$

The grey filtering-prediction equation is:

$$
\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \tag{6}
$$

where,

$$
\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{a}{b}\right] e^{-bk} + \frac{a}{b} \tag{7}
$$

The typical goodness-of-fit measure of GM(1,1) model is the relative error, i.e.

$$
e(k) = \frac{\left|x^{(0)}(k) - \hat{x}^{(0)}(k)\right|}{x^{(0)}(k)}, \quad k = 2, 3, \cdots, n \tag{8}
$$

and the model efficiency is defined as:

$$E = \frac{1}{n-1}\sum_{i=2}^{n} e(i) \tag{9}$$

3.10.3 The Analysis of Component-Level Models

The intention of the analysis on the component-level models is to examine and reveal the roles of each component-level model and therefore seek some way to improve modelling efficiency.

3.10.3.1 The Difference Equation in Standard GM(1,1) Model

It is obvious that if we rewrite equation 1, we will obtain a (nonhomogeneous) first-order difference equation.

$$\left(1+\frac{\beta}{2}\right)x^{(1)}(k)-\left(1-\frac{\beta}{2}\right)x^{(1)}(k-1)=\alpha, \quad k=2,\cdots,n \tag{10}$$

For notational convenience, denote $u_k = x^{(1)}(k)$, $u_{k-1} = x^{(1)}(k-1)$, $A = \alpha/(1+0.5\beta)$, and $B = -(1-0.5\beta)/(1+0.5\beta)$, then equation 10 will be in a standard form of first-order difference equation,

$$u_{l+1} + Bu_l = A, \quad l = k\text{-}1, k = 2,\cdots,n \tag{11}$$

Denote E as forward shift operator, i.e., $Eu_l = u_{l+1}$, then, the homogeneous equation of equation 11, is:

$$(E+B)u_l = 0, \quad l = k\text{-}1, k = 2,\cdots,n \tag{12}$$

With general solution:

$$u_l = \upsilon(-B)^l, \quad l = k\text{-}1, k = 2,\cdots,n \tag{13}$$

where υ is a constant determined by initial condition. Then the solution to equation 11 is:

$$u_k = x^{(1)}(k) = \frac{\alpha}{\beta} - \left(\frac{\alpha}{\beta} - x^{(0)}(1)\right)\left(\frac{1-0.5\beta}{1+0.5\beta}\right)^{k-1}, \quad k = 2,\cdots,n \tag{14}$$

Similar discussion will result in,

$$x^{(0)}(k)=\frac{\beta}{1+0.5\beta}\left(\frac{\alpha}{\beta}-x^{(0)}(1)\right)\left(\frac{1-0.5\beta}{1+0.5\beta}\right)^{k-2},\quad k=2,\cdots,n \tag{15}$$

The feature of difference equation embedded in GM(1,1) model is obvious but was not explored initially. The relationship between difference equation solution to equation 15 and differential equation solution form in equation 7 were investigated later (Deng, 2002a). The logical explanation was probably rooted in the intention for introducing grey differential equation concept in the grey theory.

3.10.3.2 The Differential Equation in Standard GM(1,1) Model

Equation 2 is an ordinary first-order (nonhomogeneous) differential equation of the form,

$$y'+p(x)y=q(x) \tag{16}$$

Therefore the solution to the constrained ordinary differential equation (ODE) in equation 1 is:

$$\begin{cases} y(t)=\left(y(0)-\dfrac{\alpha}{\beta}\right)\exp(-\beta t)+\dfrac{\alpha}{\beta} \\ y(0)=\gamma \end{cases} \tag{17}$$

What should be emphasized here is that the so-called whitenization equation (or shadow equation) in the coupled definition of GM(1,1) model is a *belief* reflecting modelers' experiences on certain social, business, engineering, or natural phenomena. It is the *priori* similar to the prior distribution in Bayesian modelling. But the prior in GM(1,1) model is in the form of dynamic rule rather than that in distribution. In Bayesian modelling, when sample size becomes large enough, the prior distribution the influence of prior becomes nullified. However, the constraint (i.e., the dynamics imposed) is not necessarily assimilated by actual data sampled. In other words, the constraint ordinary differential equation should be changed according to the data circumstances to avoid the functional misspecification of the dynamic rule.

3.10.3.3 The Regression Model in GM(1,1) Modelling

No matter the solution derived from the differential equation or from the shadow differential equation, the parameter pair (α,β) in the solution remains unknown. The quickest way to seek

(α,β) is to fit a simple regression model shown in equation 4 to 7. The simple regression modelling provides the "estimated" parameter values for the solution to the shadow differential equation. Without regression model, we merely stay at the theoretical level. However, once the regression model is performed, the empirical feature of GM(1,1) model reveals.

The common practices in currently GM(1,1) modelling are ignoring the goodness-of-fit of the simple regression itself. What will be extracted from simple regression modelling is merely to obtain the estimate of parameter pair, denoted by (a,b). The information on the estimated standard deviations is ignored because of the extremely small sample size. We are also aware that parameters from a regression with a goodness-of-fit measure will offer model efficiency for the GM(1,1) model.

As we pointed out after defining standard GM(1,1) model, Deng (1985), the coupling of the simple regression and difference equation is imposed. However, we must pointed out that the coupling of the simple regression and difference equation is intrinsic.

3.10.3.4 The Variational Coupling in GM(1,1) Model

We should emphasize here, the standard GM(1,1) model couples a simple regression and a differential equation together. In grey theory literature, this coupling is explicitly implied by the 1-AGO term appeared in equation 1. However, this is first time to state the definition of GM(1,1) in terms of coupled equations manner. Let us examine the logic underlying the coupling definition. Mathematically, we can state equation 1 and 2 in the following optimization problem,

$$\min_{\alpha,\beta,\gamma} \frac{1}{2}\sum_{i=1}^{n}\left(x^{(1)}(i)-x^{(1)}_{obs}(i)\right)^2$$

s.t. (18)

$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha$$

This is a standard variational problem of searching a functional $x^{(1)}(t)$ minimizing $J=\left[\sum_{i=1}^{n}\left(x^{(1)}(i)-x^{(1)}_{obs}(i)\right)^2\right]\Big/2$ under a constant coefficient first-order linear ordinary differential equation $dx^{(1)}(t)/dt+\beta x^{(1)}(t)=\alpha$ as constraint, where $\gamma=x^{(1)}(0)$ is the initial value for the

differential equation. Let λ be Lagrangian multiplier with a function form defined by the adjoint equation. For details, see Guo et al. (2006a, 2006b).

However, the variational formation can also be stated as:

$$\min_{\alpha,\beta,\gamma} \frac{1}{2}\sum_{i=1}^{n}\left(x^{(0)}(i)-x_{obs}^{(0)}(i)\right)^2$$
$$\text{s.t.} \tag{19}$$
$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha$$

3.10.4 Families of Shadow Differential Equations

The success of GM(1,1) model lies on the successful coupling of regression model with constraint differential equation model. In the standard GM(1,1) modelling exercises, appropriate data treatments, formation of regression model with differential equation as constraint, which is in nature a conversion of a constrained L_2 -optimization problem into a simple regression modelling problem. However, we notice that the standard GM(1,1) model utilizes a very simple differential equation as constraint which is based on the generalized energy law as a *priori*. The real world data may not follow generalized energy law at all, therefore it is necessary to explore new coupling in order to have a better data-assimilated model. In this section, we will explore some families suitable to meet certain coupling requirements imposed by the shadow differential equation. First let us discuss the general coupling principle.

3.10.4.1 The Coupling Principle in Formation of GM(1,1) Model

The parameters involved in GM(1,1) model are only two, which secures the feasibility of inference based on small sample size. Furthermore, the entries of regression model matrix, denoted by $X=(x_{ij})$, are calculated by AGO or IAGO operations from sampled data sequence $X^{(0)}=\left\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\}$ or assigned value "1". As to the form of the shadow differential equation, the requirements are also straightforward: only two parameters involved and the obtainable solution form.

The success of GM(1,1) model does not prevent us from an elementary fact that the efficiency of GM(1,1) model depends upon the sampled data. An important index of the potential degree of data assimilation is the class ratio.

$$\sigma^{(0)}(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},\ k=2,3,\cdots,n \tag{20}$$

If $\sigma^{(0)}(k)\in\left(e^{-2/(n+2)},e^{2/(n+2)}\right)$ for $k=2,3,\cdots,n$ then 1-AGO can secure the grey exponential trend of 1-AGO sequence $X^{(1)}=\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\}$ where $x^{(1)}(k)=\sum_{j=1}^{n}x^{(0)}(j)$ such that the degree of data assimilation will be high.

However, the behavior of sampling data from unknown system is not controllable or predictable and therefore we often face the situations that classical GM(1,1) model does not fit the data well. The diversity of real-world needs us to facilitate more potential grey differential equation model satisfying the following principle:

(i) The shadow (constraint) differential equation $dx^{(1)}/dt=F\left(t,x^{(1)};\theta\right)$ must be solvable with a *two* parameters, denoted by θ, and accordingly defines the form of regression model for parameter estimation.

(ii) The two-parameter regression model defined by the shadow differential equation must takes *linear form* in the two parameters with data sequence $X^{(0)}$ as observation vector Y and using mean 1-AGO sequence $Z^{(1)}$ or appropriate transformation $\eta\left(Z^{(1)}\right)$ for the construction of the regression model matrix.

One of the simplest forms of the regression models in accordance with the principle stated is:

$$x^{(0)}(k)+\beta\eta\left[z^{(1)}(k)\right]=\alpha,\ k=2,3,\cdots,n \tag{21}$$

Then the shadow differential equation takes the form:

$$\frac{dx^{(1)}}{dt}+\beta\eta\left(x^{(1)}\right)=\alpha,\ k=2,3,\cdots,n \tag{22}$$

i.e.

$$\frac{dx^{(1)}}{dt} = \alpha - \beta\eta\left(x^{(1)}\right) \tag{23}$$

Equation 23 gives an induced family of appropriate shadow differential equation satisfying Principle (i) and (ii). Hence, it will be worthwhile to investigate the appropriate choice of the couple of regression model and constraint differential equation model for potential extended GM(1,1) models in the next section.

3.10.4.2 The Families Coupled by Various Differential Equations

In the literature of first-order ordinary differential equations, there are certain solvable ones, i.e., the solution have a close form. A particular important issue is that these differential equations should be able to expressed as linearly with parameter α and β. We explored seven families: (1) the differential equation with separable variables; (2) the homogeneous differential equations; (3) the linear differential equations; (4) the differential equation of form $x = F(t, x')$; (5) The Lagrangian differential equation; (6) the Clairaut differential equation; and (7) the Raccati differential equation. For more detailed discussions, see Guo et al. (2006a, 2006b). Table 3.10.1 lists the potential extended GM(1,1) models in a coupling manner.

Table 3.10.1 Extended GM(1,1) Models.

Form	Constraint	Regression
1	$f_1(t)g_1(x)dt + f_2(t)g_2(x)dx = 0$	$x^{(0)}(k) = -f_1(k;\theta)g_1\left(z^{(1)}(k)\right)\Big/\left(f_2(k)g_2\left(z^{(1)}(k)\right)\right)$
2	$dx/dt = F(x/t)$	$x^{(0)}(k) = F\left(z^{(1)}(k)/k;\theta\right)$
3	$dx/dt + p(t)x = q(t)$	$x^{(0)}(k) + p(k)z^{(1)}(k) = q(k)$
4	$x = F(t, x')$	$\left(x^{(0)}(k) - \frac{\partial F}{\partial t}\left(k, x^{(0)}(k);\theta\right)\right) - \frac{\partial F}{\partial p}\left(k, x^{(0)}(k);\theta\right)x^{(-1)}(k) = 0$
5	$x = tf_1(x') + f_2(x')$	$\frac{1}{x^{(-1)}(k)} = \frac{f_1'\left(x^{(0)}(k);\alpha\right)}{x^{(0)}(k) - f_1\left(x^{(0)}(k);\alpha\right)} + \frac{f_2'\left(x^{(0)}(k);\beta\right)}{x^{(0)}(k) - f_2\left(x^{(0)}(k);\beta\right)}$
6	$x = tx' + F(x')$	$z^{(1)}(k) = c_0 k + F(c_0;\theta)$
7	$x' = p(t)x^2 + q(t)x + r(t)$	$u^{(0)}(k) + \left[q(k) + 2p(k)z_{x_1}^{(1)}(k)\right]z_u^{(1)}(k) + p(k) = 0$

In the formation 7, $u = 1 / (x - x_1)$, where x_1 is a particular solution to differential equation form 7. However, other forms of differential equations, for example, $t = F(x, x')$, or $F(t, x') = 0$, or $F(x, x') = 0$ are also good candidate to play the role of shadow differential equation.

3.10.5 Conclusion

In this paper, we examine Deng's (1985) standard GM(1,1) model from its component-level models: the shadow differential equation, the regression model and the intrinsic difference model. Furthermore, we examine the coupling nature of the standard GM(1,1) model, i.e., the GM(1,1) model is a variational problem with differential equation as its constraint. The "coupling" nature also restricts the form of the regression model in the object function and associated AGO and IAGO operations. Based on these examinations, we state the coupling principle for guiding the generalization to Deng's (1985) GM(1,1) model. Seven generalized families to the standard GM(1,1) model are proposed accordingly (Guo et al., 2006a, 2006b). We are expecting this foundational work will provide more flexible and feasible GM(1,1) models for fitting small sample data with high degree of data assimilation capability.

3.11 Rationale Exploration of First-Order One Variable Grey Differential Equation Model via Variational Approach

Renkuan Guo, Danni Guo, and Yanhong Cui

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
rguo@stats.uct.ac.za dannig@stats.uct.ac.za ycui@stats.uct.ac.za

This paper has been published in the Proceedings of 11th International Conference on Industrial Engineering, Theory, Applications and Practice, October 24-27, 2006, Nagoya, Japan, pp. 1302-1307.

3.11.1 Introduction

The first-order one-variable grey differential equation model, GM(1,1) model, requires very small sample size, as little as four sample points (Deng, 1985), but the modelling may result in very high predictability. Furthermore, the parameter estimation in GM(1,1) model is obtained via a simple regression, which is offered in any beginning statistics or data analysis course. A simple Excel spreadsheet can facilitate the necessary computations. Therefore the advantage for spatial analysts to use GM(1,1) model in GIS modelling is very obvious. However, what is the rationale of GM(1,1) modelling on system lifetime is unclear and will become an obstacle for modelling exercises with real world data to obtain model fully utilizing data information. In this paper, based on the fundamental fact that data-based mathematical modelling is in nature an optimization problem, we re-examine the rationale underlying GM(1,1) model in terms of variation calculus approach.

The investigation reveals that the classical GM(1,1) model (Deng, 1985) is just a variation problem under a special functional constraint which is a nonhomogeneous first-order linear ordinary differential equation with constant coefficient and constant term with a solution in exponential form. If the data information is consistent with the special exponential constraint functional form, the GM(1,1) model will be of high accuracy. In other words, the success of GM(1,1) model depends upon two factors: the behavior or evolving dynamic contained in sample of small size, called sample dynamic, and the dynamic of constraint functional defined by ordinary differential equation, called constraint functional dynamic. The closer the sample dynamic curve and the constraint functional dynamic curve are to each other, the better fit the GM(1,1). Based on this argument, we propose a consistence measure in terms of the concept of grey relational analysis (GRA), which was originally proposed by Deng (1985). Finally we explore the evolving behavior of a set of constraint functionals defined by ordinary differential equations in order to select the suitable functional satisfying the consistence measure during GM(1,1) modelling process.

3.11.2 A Brief Review on GM(1,1) Model

The success of GM(1,1) model lies on the following two aspects: data treatments and the conversion of a constrained L_2 -optimization problem into a simple regression modelling exercise.

Definition 1: (GM(1,1) model). Given a discrete positive real-valued (equal-spaced) data sequence $X^{(0)}=(x^{(0)}(1), x^{(0)}(2),\ldots, x^{(0)}(n))$, equation:

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

which is coupled with the differential equation,

$$dx^{(1)}(t)/dt+\beta x^{(1)}(t)=\alpha \tag{2}$$

is called a one-variable first order grey differential equation (GM(1,1)) with respect to time series sequence $X^{(0)}=(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, where

$$z^{(1)}(k)=\frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right], x^{(1)}(k)=\sum_{i=1}^{k} x^{(0)}(i), k=2,\cdots,n \tag{3}$$

where β is called the developing coefficient, α is the grey input, and $x^{(0)}$ is called a *grey derivative* and $x^{(1)}(k)$ is called the k^{th} 1-AGO (accumulative generation operator) value.

Furthermore, the (coupled) differential equation $dx^{(1)}/dt + \beta\ x^{(1)} = \alpha$ (in equation 2) is called the whitenization differential equation or the *shadow* equation of the grey differential equation in equation 1. The unknown parameter values (α,β) can be determined in terms of a standard L_2-optimization. Equation 1 can be re-written as in a simple regression equation:

$$y_k=\alpha+\beta x_k, \quad k=2,\cdots,n \tag{4}$$

with

$$y_k=x^{(0)}(k), \quad x_k=-z^{(1)}(k), \quad k=2,3,\cdots,n \tag{5}$$

The parameter estimate (a,b) can be obtained from equation 5,

$$(a,b)^T=\left(X^TX\right)^{-1}X^TY \tag{6}$$

where

$$X=\begin{bmatrix}1 & -z^{(1)}(2)\\ 1 & -z^{(1)}(3)\\ \vdots & \vdots\\ 1 & -z^{(1)}(n)\end{bmatrix}, \quad Y=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix} \tag{7}$$

The grey filtering-prediction equation is

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \tag{8}$$

where

$$\hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{a}{b} \right] e^{-bk} + \frac{a}{b} \tag{9}$$

The typical goodness-of-fit measure of GM(1,1) model is the relative error, i.e.

$$e(k) = \frac{\left| x^{(0)}(k) - \hat{x}^{(0)}(k) \right|}{x^{(0)}(k)}, \; k = 2,3,\cdots,n \tag{10}$$

and the model efficiency is defined as,

$$E = \frac{1}{n-1} \sum_{i=2}^{n} e(i) \tag{11}$$

A critical feature of (classical) GM(1,1) model proposed by Deng (1985) is performed with respect to a discrete positive data sequence.

3.11.3 Exploration on GM(1,1) via Calculus of Variation

We should emphasize here, the definition of GM(1,1) model couples a simple regression and a differential equation together. In grey theory literature, this coupling is explicitly implied by the 1-AGO term appeared in equation 3. Let us examine the logic underlying the coupling definition.

Mathematically, we can state equation 1 and 2 in the following optimization problem,

$$\begin{aligned} &\min_{\alpha,\beta,\gamma} \frac{1}{2} \sum_{i=1}^{n} \left(x^{(1)}(i) - x_{obs}^{(1)}(i) \right)^2 \\ &\text{s.t.} \\ &\frac{dx^{(1)}(t)}{dt} + \beta x^{(1)}(t) = \alpha \end{aligned} \tag{12}$$

This is a standard variational problem of searching a functional $x^{(1)}(t)$ minimizing $J = [\sum_{i=1}^{n} (x^{(1)}(i) - x_{obs}^{(1)}(i))^2]/2$ under a constant coefficient first-order linear ordinary differential

equation $dx^{(1)}(t)/dt + \beta x^{(1)}(t) = \alpha$ as constraint, where $\gamma = x^{(0)}(0)$ is the initial value for the differential equation.

To obtain the parameter (α, β, γ) it is necessary to investigate the Lagrangian function $\pounds$,

$$\pounds = J + \mu\left(\gamma - x^{(1)}(0)\right) + \int_0^{T_f} \lambda(t)\left[\frac{dx^{(1)}(t)}{dt} + \beta x^{(1)}(t) - \alpha\right] dt \tag{13}$$

where λ and μ are the Lagrangian multipliers. The necessary conditions for obtaining unconstrained minimum are:

$$\frac{\partial \pounds}{\partial x^{(1)}} = 0,\ \frac{\partial \pounds}{\partial M} = 0,\ \frac{\partial \pounds}{\partial \theta} = 0 \tag{14}$$

where $M^T = (\mu, \lambda)$ and $\theta^T = (\alpha, \beta, \gamma)$.

The first equation of equation 14 leads to the so-called adjoint equation, the second equation recovers model equation (constraint) and the last equation gives the gradients with respect to (control parameter) $\theta^T = (\alpha, \beta, \gamma)$.

The Euler's Equation is:

$$\frac{\partial \pounds}{\partial x^{(1)}} - \frac{d}{dt}\left[\frac{\partial \pounds}{\partial \left(x^{(1)}\right)'}\right] = \left(x^{(1)} - x_{obs}^{(1)}\right) + \beta\lambda(t) - \lambda'(t) = 0 \tag{15}$$

And the adjoint equation is:

$$\begin{cases} \dfrac{d\lambda(t)}{dt} = \beta\lambda(t) + \left(x^{(1)} - x_{obs}^{(1)}\right) \\ \lambda(T_f) = 0 \end{cases} \tag{16}$$

The solution to the constraint is:

$$x^{(1)}(t) = \left(\gamma - \frac{\alpha}{\beta}\right)e^{-\beta t} + \frac{\alpha}{\beta} \tag{17}$$

Then we can obtain the expression for $\lambda(t)$ via multiple factor $e^{\beta t}$, we obtain,

$$\lambda(t)=\left[\frac{1}{2\beta}\left(\gamma-\frac{\alpha}{\beta}\right)e^{-2\beta T_f}+\frac{1}{\beta}\left(\frac{\alpha}{\beta}-x_{obs}^{(1)}(0)\right)e^{-\beta T_f}\right]e^{\beta t}$$
$$-\frac{1}{2\beta}\left(\gamma-\frac{\alpha}{\beta}\right)e^{-2\beta t}-\frac{1}{\beta}\left(\frac{\alpha}{\beta}-x_{obs}^{(1)}(0)\right) \tag{18}$$

The next step is to derive the first-order and second-order partial derivatives of J with respect to α, β, and γ respectively. Based on equation $\partial\mathfrak{L}/\partial\theta=0$,

$$\frac{\partial J}{\partial \theta}=-\int_0^{T_f}\left[\frac{\partial F}{\partial \theta}\right]\lambda(t)\,dt \tag{19}$$

where

$$\begin{cases}F\left(x^{(1)},\theta\right)=\dfrac{dx^{(1)}}{dt}=\alpha-\beta x^{(1)}\\ x^{(1)}(0)=\gamma\end{cases} \tag{20}$$

Then

$$\begin{aligned}F\left(x^{(1)},\theta\right)&=\alpha-\beta x^{(1)}\\&=\alpha-\beta\left[\left(\gamma-\frac{\alpha}{\beta}\right)e^{-\beta t}+\frac{\alpha}{\beta}\right]\\&=(\alpha-\beta\gamma)e^{-\beta t}\end{aligned} \tag{21}$$

which gives:

$$\begin{cases}\dfrac{\partial F}{\partial\alpha}=e^{-\beta t}\\ \dfrac{\partial F}{\partial\beta}=-\left[\gamma+(\alpha-\beta\gamma)t\right]e^{-\beta t}\\ \dfrac{\partial F}{\partial\gamma}=-\beta e^{-\beta t}\end{cases} \tag{22}$$

3.11.4 The Data-Assimilation Measure

The investigation reveals that the key to have highly data-assimilated GM(1,1) modelling is the grey derivative sequence (i.e., the original observations), and it can be highly correlated by the 1-AGO sequence. Deng (2002a) stated that the model efficiency will be high, if the class ratio satisfies conditions:

$$\sigma^{(0)}(k)=x^{(0)}(k-1)/x^{(0)}(k)\in\left[e^{-2/(n+2)},e^{2/(n+2)}\right],\ k=2,3,\cdots,n \tag{23}$$

Standard GM(1,1) model imposed the constraint differential equation:

$$dx^{(1)}/dt=\alpha+\beta\left(-x^{(1)}\right) \tag{24}$$

Therefore, from the previous investigations, we know that the goodness-of-fit of the model will be high if the data fits the constraint exponential function well, otherwise the fit will be poor. However, the class ratio test is a passive measure of model accuracy but does not indicate the capacity of model extrapolations, which is a clear requirement in spatial prediction modelling. Therefore, we need to explore the concept of functional similarity, and define an appropriate measure for model accuracy as well as the model predictability power.

Definition 2: (Functional Similarity Measure). Given two functions f, g and a small positive number $\varepsilon>0$, if $\rho_\varepsilon[a,b]=\int_a^b|f-g|d\mu\Big/\sqrt{\int_a^b|f|d\mu\int_a^b|g|d\mu}<\varepsilon$, then we say that two functions f, g are ε-similar in interval $[a,b]$ with respect to measure μ.

Definition 3: (ε-Similarity Measure for Discrete Data Sequences) Let $X=\{x(k),\ k=1,2,\cdots,n\}$ and $Y=\{y(k),\ k=1,2,\cdots,n\}$ are two discrete data sequences with best-fitted polynomial spline functions of order $(n-1)$, denoted by $p_{n-1}(t)$ and $q_{n-1}(t)$ respectively. Then the ε-similar measure in interval $[1,n]$ is defined by:

$$\rho_\varepsilon[1,n]=\int_1^n|p_{n-1}(t)-q_{n-1}(t)|dt\Big/\sqrt{\int_1^n|p_{n-1}(t)|dt\int_1^n|q_{n-1}(t)|dt}<\varepsilon \tag{25}$$

It is very obvious the for the classical GM(1,1) model proposed by Deng (1985), the ε-similar measure in interval $[2,n]$ is:

$$\rho_\varepsilon[2,n]=\int_2^n\left|x_{n-1}^{(0)}(t)-\hat{x}_{n-1}^{(0)}(t)\right|dt\Big/\sqrt{\int_2^n\left|x_{n-1}^{(0)}(t)\right|dt\int_2^n\left|\hat{x}_{n-1}^{(0)}(t)\right|dt}<\varepsilon \tag{26}$$

where $x_{n-1}^{(0)}(t)$ is the best-fitted polynomial spline functions of order $(n-1)$ based on the original data sequence $\left\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\}$ and $\hat{x}_{n-1}^{(0)}(t)$ is the best-fitted polynomial spline functions of order $(n-1)$ based on the filtered data sequence $\left\{\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)\right\}$. As to

the predictability power in interval $[n+1, n+k]$, we just need to evaluate the ε-similarity measure on the interval $[n+1, n+k]$.

$$\rho_{\varepsilon}[n+1, n+k] = \int_{n+1}^{n+k} \left| x_{n-1}^{(0)}(t) - \hat{x}_{n-1}^{(0)}(t) \right| dt \Bigg/ \sqrt{\int_{n+1}^{n+k} \left| x_{n-1}^{(0)}(t) \right| dt \int_{n+1}^{n+k} \left| \hat{x}_{n-1}^{(0)}(t) \right| dt} < \varepsilon \qquad (27)$$

If the inequality holds, then the k-step extrapolation can be performed with high accuracy.

The ε-similarity measure proposed here is actual measuring the geometric similarity between two functions defined on the common domain. It will catches up the full functional relation including linear and non-linear parts. We expect that it will play an important roles in the generalization of classical GM(1,1) model as well as in the grey relational analysis although the complexity from the computing best-fitted polynomial function will be inevitable.

3.11.5 Conclusion

In this paper, we investigate the nature of GM(1,1) model in terms of a variational viewpoint. Form the analysis performed in this paper, we conclude that standard GM(1,1) model is a variational problem with differential equation as its constraint. Furthermore, the efficiency of GM(1,1) model depends upon the choice of the constraint (shadow) differential equation. The clear message from the variational analysis is GM(1,1) model proposed by Deng (1985) do enjoy the easiness and efficiency in modelling the discrete positive data sequence as long as the 1-AGO data sequence possesses grey exponential trend.

However, the classical GM(1,1) does not warrant the model accuracy as well the extrapolation capability, therefore it is necessary to use the best-fitted polynomial of order $(n-1)$ as constraint function in order to reach high data-assimilation requirement, although the convenience of GM(1,1) model will be totally lost. A possible remedy is to use the extended GM(1,1) model based on the coupling principle (Guo et al., 2006a, 2006b), in which the differential equation constraint function is chosen in terms of ε-similarity measure, between the constraint function and the best-fitted polynomial of order $(n-1)$.

3.12 The Coupling of Regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Journal of Grey System, 2006, 9(2): 143-154. Editors: Kun-Li Wen and Cha'o-Kuang Chen.

3.12.1 Introduction

The first-order one-variable grey differential equation model, abbreviated as GM(1,1) model, requires very small sample size, as little as four sample points (Guo et al., 2004), but the modelling may result in very high predictability. Furthermore, the parameter estimation in GM(1,1) model is obtained using a simple regression, which is offered in any beginning statistics or data analysis course. A simple Excel spreadsheet can facilitate the necessary computations. Therefore the advantage for modellers to use GM(1,1) model is very obvious. However, what is the rationale of GM(1,1) modelling on observational data is unclear and will become an obstacle for engineers and managers armed with statistical analysis to accept GM(1,1) modelling idea and apply it in their practices solidly. Also, it is observed that GM(1,1) model possesses a feature of data assimilation. The goodness-of-fit is not always convincing.

In other words, the standard GM(1,1) model, proposed by Deng (1985) is of limited applicability. It is necessary to further explore the nature of GM(1,1), identify the underlying mechanism leading to modelling success and extend it to wider modelling families. In this paper, we examine the three basic component models – simple regression model, differential equation model, and difference model in GM(1,1) model. The component-level model explorations will lead us to have a better picture of GM(1,1) modelling dynamics: it is a coupled differential equation and regression. The differential equation plays the role as constraint and defines the basic formation of regression model, while the appropriate form of regression model plays the role to estimate the two parameters assimilated by data. We state such coupling nature in GM(1,1) modelling as coupling principle, and also explore the potential extensions to Deng's GM(1,1) model.

3.12.2 A Brief Review on Standard GM(1,1) Model

The success of GM(1,1) model lies on the following two aspects: data treatments and the conversion of a constrained L_2 -optimization problem into a simple regression modelling exercise.

Definition 1: (GM(1,1) model). Given a discrete positive real-valued (equal-spaced) data sequence $X^{(0)}=(x^{(0)}(1), x^{(0)}(2),\ldots, x^{(0)}(n),)$ equation

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

which is coupled with the differential equation:

$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha \tag{2}$$

is called a one-variable first order grey differential equation (GM(1,1) model) with respect to time series sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$, where

$$\begin{aligned} z^{(1)}(k) &= \frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right] \\ x^{(1)}(k) &= \sum_{i=1}^{k} x^{(0)}(i) \\ k &= 2,\cdots,n \end{aligned} \tag{3}$$

where β is called the developing coefficient, α is the grey input, and $x^{(0)}$ is called a *grey derivative* and $x^{(1)}(k)$ is called the k^{th} 1-AGO (accumulative generation operator) value. Furthermore, the (coupled) differential equation $dx^{(1)}/dt + \beta x^{(1)} = \alpha$ (in equation 2) is called the whitenization differential equation or the *shadow* equation of the grey differential equation equation 1.

The unknown parameter values (α,β) can be determined in terms of a standard regression. Equation 1 can be re-written as in a simple regression form:

$$y_k = \alpha + \beta x_k, \quad k = 2,\cdots,n \tag{4}$$

with

$$y_k = x^{(0)}(k), \quad x_k = -z^{(1)}(k), \quad k = 2,3,\cdots,n \tag{5}$$

The parameter estimate (a,b) can be obtained from equation 6.

$$(a,b)^T = \left(X^T X\right)^{-1} X^T Y \tag{6}$$

where

$$X = \begin{bmatrix} 1 & -z^{(1)}(2) \\ 1 & -z^{(1)}(3) \\ \vdots & \vdots \\ 1 & -z^{(1)}(n) \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \tag{7}$$

The grey filtering-prediction equation is:

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) \tag{8}$$

where

$$\hat{x}^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{a}{b}\right]e^{-bk}+\frac{a}{b} \tag{9}$$

The typical goodness-of-fit measure of GM(1,1) model is the (absolute) relative error, i.e.

$$e(k)=\frac{\left|x^{(0)}(k)-\hat{x}^{(0)}(k)\right|}{x^{(0)}(k)},\ k=2,3,\cdots,n \tag{10}$$

and the model efficiency is defined as:

$$E=\frac{1}{n-1}\sum_{i=2}^{n}e(i) \tag{11}$$

A critical feature of (classical) GM(1,1) model proposed by Deng (1985) is performed with respect to a discrete positive data sequence.

3.12.3 The Three Component Models in GM(1,1) Modelling Mechanism

3.12.3.1 The Difference Equation in GM(1,1) Modelling

It is obvious that if we rewrite equation 1, we will obtain a (nonhomogeneous) first-order difference equation,

$$\left(1+\frac{\beta}{2}\right)x^{(1)}(k)-\left(1-\frac{\beta}{2}\right)x^{(1)}(k-1)=\alpha,\quad k=2,\cdots,n \tag{12}$$

For notational convenience, denote $u_k=x^{(1)}(k)$, $u_{k-1}=x^{(1)}(k-1)$, $A=\alpha/(1+0.5\beta)$, and $B=-(1-0.5\beta)/(1+0.5\beta)$, then equation 12 will be in a standard form of first-order difference equation,

$$u_{l+1}+Bu_l=A,\quad l=k\text{-}1,\ k=2,\cdots,n \tag{13}$$

Denote E as forward shift operator, i.e., $Eu_l=u_{l+1}$, then, the homogeneous equation of equation 13, is;

$$(E+B)u_l=0,\quad l=k\text{-}1,\ k=2,\cdots,n \tag{14}$$

With general solution,

$$u_l = \upsilon(-B)^l, \quad l = k\text{-}1, k = 2, \cdots, n \tag{15}$$

where υ is a constant determined by initial condition. Then the solution to equation 13 is,

$$u_k = x^{(1)}(k) = \frac{\alpha}{\beta} - \left(\frac{\alpha}{\beta} - x^{(0)}(1)\right)\left(\frac{1-0.5\beta}{1+0.5\beta}\right)^{k-1}, \quad k = 2, \cdots, n \tag{16}$$

Similar discussion will result in:

$$x^{(0)}(k) = \frac{\beta}{1+0.5\beta}\left(\frac{\alpha}{\beta} - x^{(0)}(1)\right)\left(\frac{1-0.5\beta}{1+0.5\beta}\right)^{k-2}, \quad k = 2, \cdots, n \tag{17}$$

The feature of difference equation embedded in GM(1,1) model is obvious but was not explored initially. The relationship between difference equation solution to equation 16 and differential equation solution form in equation 9 were investigated later (Deng, 2002a). The logical explanation was probably rooted in the intention for introducing grey differential equation concept in the grey theory.

3.12.3.2 The Differential Equation in GM(1,1) Modelling

Equation 2 is an ordinary first-order (nonhomogeneous) differential equation of the form:

$$y' + p(x)y = q(x) \tag{18}$$

Then the solution to equation 18 is:

$$y = \exp\left(-\int p(x)dx\right)\left[\int q(x)\exp\left(\int p(x)dx\right)dx + c\right] \tag{19}$$

Therefore the solution to the constrained ordinary differential equation (ODE) in equation 2 is:

$$\begin{cases} y(t) = \left(y(0) - \dfrac{\alpha}{\beta}\right)\exp(-\beta t) + \dfrac{\alpha}{\beta} \\ y(0) = y_0 \end{cases} \tag{20}$$

What should be emphasized here is that the so-called whitenization equation (or shadow equation) in the coupled definition of GM(1,1) model is a *belief* reflecting modellers' experiences on certain social, business, engineering, or natural phenomena. This belief is actually a *priori* in GM(1,1) modelling. The constraint is not necessarily assimilated by actual data sampled, i.e. the constraint ordinary differential equation should be changed according to the data structure.

3.12.3.3 The Regression Model in GM(1,1) Modelling

No matter the solution derived from the differential equation or from the shadow differential equation, the parameter pair (α, β) in the solution remains unknown. The quickest way to seek (α, β) is to fit a simple regression model shown in equation 4 to 7. The common practices in current GM(1,1) modelling are ignoring the goodness-of-fit of the simple regression itself. What will be extracted from simple regression modelling is merely to obtain the estimate of parameter pair, denoted by (a, b). The information on the estimated standard deviations is ignored because of the extremely small sample size. We are also aware that parameters from a regression with a goodness-of-fit measure will offer model efficiency for the GM(1,1) model.

As we pointed out after the definition of GM(1,1) model, the coupling of the simple regression and difference equation is imposed. However, we must pointed out that the coupling of the simple regression and difference equation is intrinsic.

3.12.4 A Variational Explanation on the Coupling

We should emphasize here, the definition of GM(1,1) model couples a simple regression and a differential equation together. In grey theory literature, this coupling is explicitly implied by the 1-AGO term appeared in equation 3. Let us examine the logic underlying the coupling definition.

Mathematically, we can state equation 1 and 2 in the following optimization problem:

$$\min_{\alpha,\beta,\gamma} \frac{1}{2}\sum_{i=1}^{n}\left(x^{(1)}(i)-x^{(1)}_{obs}(i)\right)^2$$

$$\text{s.t.} \qquad (21)$$

$$\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)=\alpha$$

This is a standard variational problem of searching a functional $x^{(1)}(t)$ minimizing $J=\left[\sum_{i=1}^{n}\left(x^{(1)}(i)-x^{(1)}_{obs}(i)\right)^2\right]\Big/2$ under a constant coefficient first-order linear ordinary differential equation $dx^{(1)}(t)/dt + \beta x^{(1)}(t) = \alpha$ as constraint, where $\gamma=x^{(1)}(0)$ is the initial value for the differential equation.

To obtain the parameter (α, β, γ), it is necessary to investigate the Lagrangian function $\mathfrak{L}$,

$$\mathfrak{L}=J+\mu\left(\gamma-x^{(1)}(0)\right)+\int_0^{T_f}\lambda(t)\left[\frac{dx^{(1)}(t)}{dt}+\beta x^{(1)}(t)-\alpha\right]dt \tag{22}$$

Where λ and μ are the Lagrangian multipliers. The necessary conditions for obtaining unconstrained minimum are,

$$\begin{cases}\dfrac{\partial\mathfrak{L}}{\partial x^{(1)}}=0\\ \dfrac{\partial\mathfrak{L}}{\partial M}=0\\ \dfrac{\partial\mathfrak{L}}{\partial\theta}=0\end{cases} \tag{23}$$

Where $M=(\mu,\lambda)^T$ and $\theta^T=(\alpha,\beta,\gamma)$. The first equation of equation 23 leads to the so-called adjoint equation, the second equation recovers model equation (constraint) and the last equation gives the gradients with respect to (control parameter) $\theta^T=(\alpha,\beta,\gamma)$.

The Euler's Equation is:

$$\frac{\partial\mathfrak{L}}{\partial x^{(1)}}-\frac{d}{dt}\left[\frac{\partial\mathfrak{L}}{\partial\left(x^{(1)}\right)'}\right]=\left(x^{(1)}-x^{(1)}_{obs}\right)+\beta\lambda(t)-\lambda'(t)=0 \tag{24}$$

And the adjoint equation is:

$$\begin{cases}\dfrac{d\lambda(t)}{dt}=\beta\lambda(t)+\left(x^{(1)}-x^{(1)}_{obs}\right)\\ \lambda(T_f)=0\end{cases} \tag{25}$$

The solution to the constraint is:

$$x^{(1)}(t)=\left(\gamma-\frac{\alpha}{\beta}\right)e^{-\beta t}+\frac{\alpha}{\beta} \tag{26}$$

Then we can obtain the expression for $\lambda(t)$ via multiple factor $e^{\beta t}$,

$$\lambda(t)=\left[\frac{1}{2\beta}\left(\gamma-\frac{\alpha}{\beta}\right)e^{-2\beta T_f}+\frac{1}{\beta}\left(\frac{\alpha}{\beta}-x_{obs}^{(1)}(0)\right)e^{-\beta T_f}\right]e^{\beta t} -\frac{1}{2\beta}\left(\gamma-\frac{\alpha}{\beta}\right)e^{-2\beta t}-\frac{1}{\beta}\left(\frac{\alpha}{\beta}-x_{obs}^{(1)}(0)\right) \tag{27}$$

The next step is to derive the first-order and second-order partial derivatives of J with respect to α, β, and γ respectively. Based on equation $\partial\mathfrak{L}/\partial\theta=0$,

$$\frac{\partial J}{\partial\theta}=-\int_0^{T_f}\left[\frac{\partial F}{\partial\theta}\right]\lambda(t)dt \tag{28}$$

where

$$\begin{cases}F\left(x^{(1)},\theta\right)=\dfrac{dx^{(1)}}{dt}=\alpha-\beta x^{(1)}\\ x^{(1)}(0)=\gamma\end{cases} \tag{29}$$

Then

$$\begin{aligned}F\left(x^{(1)},\theta\right)&=\alpha-\beta x^{(1)}\\&=\alpha-\beta\left[\left(\gamma-\frac{\alpha}{\beta}\right)e^{-\beta t}+\frac{\alpha}{\beta}\right]\\&=(\alpha-\beta\gamma)e^{-\beta t}\end{aligned} \tag{30}$$

Which gives:

$$\begin{cases}\dfrac{\partial F}{\partial\alpha}=e^{-\beta t}\\ \dfrac{\partial F}{\partial\beta}=-\left[\gamma+(\alpha-\beta\gamma)t\right]e^{-\beta t}\\ \dfrac{\partial F}{\partial\gamma}=-\beta e^{-\beta t}\end{cases} \tag{31}$$

3.12.5 The Implementation of Control Parameter Searching

The searching of control parameter (α, β, γ) is just an application of Newton-Raphson method.

Note that the gradient vector is:

$$\begin{pmatrix} \frac{\partial J}{\partial \alpha} \\ \frac{\partial J}{\partial \beta} \\ \frac{\partial J}{\partial \gamma} \end{pmatrix} = \begin{bmatrix} -\int_0^{T_f} e^{-\beta t}\ \lambda(t)dt \\ \int_0^{T_f} [\gamma + (\alpha - \beta\gamma)t] e^{-\beta t}\ \lambda(t)dt \\ \beta \int_0^{T_f} e^{-\beta t}\ \lambda(t)dt \end{bmatrix} \tag{32}$$

Recall that the searching parameters, i.e., (α, β, γ), is actually an assimilating process with data sampled and therefore, the integrals in equation 32 should be replaced by the summations respectively. In other words, equation 32 becomes,

$$\begin{pmatrix} \frac{\partial J}{\partial \alpha} \\ \frac{\partial J}{\partial \beta} \\ \frac{\partial J}{\partial \gamma} \end{pmatrix} = \begin{bmatrix} -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\lambda(i) \\ \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i] e^{-\beta i}\lambda(i) \\ \beta\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\lambda(i) \end{bmatrix} \tag{33}$$

Then the second-order partial derivatives can be obtained.

$$\begin{aligned}
\frac{\partial^2 J}{\partial \alpha^2} &= -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \alpha} \\
\frac{\partial^2 J}{\partial \alpha \partial \beta} &= \frac{1}{n}\sum_{i=1}^{n} i e^{-\beta i}\lambda(i) - \frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \beta} \\
\frac{\partial^2 J}{\partial \alpha \partial \gamma} &= -\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \gamma} \\
\frac{\partial^2 J}{\partial \beta^2} &= \frac{1}{n}\sum_{i=1}^{n} (-\gamma i) e^{-\beta i}\lambda(i) + \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i](-i) e^{-\beta i}\lambda(i) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i] e^{-\beta i}\frac{\partial \lambda(i)}{\partial \beta} \\
\frac{\partial^2 J}{\partial \beta \partial \gamma} &= \frac{1}{n}\sum_{i=1}^{n} [1 - \beta i] e^{-\beta i}\lambda(i) + \frac{1}{n}\sum_{i=1}^{n} [\gamma + (\alpha - \beta\gamma)i] e^{-\beta i}\frac{\partial \lambda(i)}{\partial \gamma} \\
\frac{\partial^2 J}{\partial \gamma^2} &= \beta\frac{1}{n}\sum_{i=1}^{n} e^{-\beta i}\frac{\partial \lambda(i)}{\partial \gamma}
\end{aligned} \tag{34}$$

In equation 34, it has three parameters, and therefore there are six equations. It is obvious that

the first-order partial derivatives of λ with respect to α, β and γ are required. Note that:

$$\begin{cases} \dfrac{\partial\lambda(t)}{\partial\alpha}=\left[\dfrac{1}{2\alpha^2}e^{-2\beta T_f}-\dfrac{1}{\alpha^2}e^{-\beta T_f}\right]e^{\beta t}+\dfrac{1}{2\alpha^2}e^{-2\beta t}+\dfrac{1}{\alpha^2} \\ \dfrac{\partial\lambda(t)}{\partial\beta}=\dfrac{-\gamma}{2\beta^2}e^{-\beta(2T_f-t)}+\dfrac{x_{obs}^{(1)}(0)}{\beta^2}e^{-\beta(T_f-t)}+\dfrac{\gamma}{2\beta^2}e^{-2\beta t}-\dfrac{x_{obs}^{(1)}(0)}{\beta^2} \\ \qquad\dfrac{t-2T_f}{2\beta}\left(\gamma-\dfrac{\beta}{\alpha}\right)e^{-\beta(2T_f-t)}+\dfrac{t-T_f}{\beta}\left(\dfrac{\beta}{\alpha}-x_{obs}^{(1)}(0)\right)e^{-\beta(T_f-t)}+\dfrac{t}{\beta}\left(\gamma-\dfrac{\beta}{\alpha}\right)e^{-2\beta t} \\ \dfrac{\partial\lambda(t)}{\partial\gamma}=\dfrac{1}{2\beta}e^{-2\beta T_f+\beta t}-\dfrac{1}{2\beta}e^{-2\beta t} \end{cases} \tag{35}$$

The searching is just an iterated scheme.

$$\begin{pmatrix}\tilde{\alpha}^{(n+1)}\\ \tilde{\beta}^{(n+1)}\\ \tilde{\gamma}^{(n+1)}\end{pmatrix}\approx\begin{pmatrix}\tilde{\alpha}^{(n)}\\ \tilde{\beta}^{(n)}\\ \tilde{\gamma}^{(n)}\end{pmatrix}-\left(\frac{\partial^2 J\left(\theta^{(n)}\right)}{\partial\theta_i^{(n)}\partial\theta_j^{(n)}}\right)^{-1}\begin{pmatrix}\partial J\left(\theta^{(n)}\right)/\partial\alpha\\ \partial J\left(\theta^{(n)}\right)/\partial\beta\\ \partial J\left(\theta^{(n)}\right)/\partial\gamma\end{pmatrix} \tag{36}$$

3.12.6 The Principle to Extend GM(1,1) Model

We observe that solution in the variational problem depends upon the formation of the constraint differential equation. The object function, particularly, the theoretical form of $x^{(1)}(\cdot)$ comes from constraint differential equation. In Deng's (1985) GM(1,1) model, the parameters involved in GM(1,1) model are only *two*, which reflects the small sample size restriction and secures the feasibility of inference based on small sample. Furthermore, the entries of regression model matrix, denoted by $X=(x_{ij})$, are calculated by AGO or IAGO operations from sampled data sequence $X^{(0)}=\left\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right\}$ or assigned value "1". As to the form of the shadow differential equation, the requirements are also straightforward: only two parameters involved in the obtainable solution form.

The success of GM(1,1) model does not prevent us learning from an elementary fact that the efficiency of GM(1,1) model depends upon the sampled data. An important index of the potential degree of data assimilation is the class ratio.

$$\sigma^{(0)}(k)=\frac{x^{(0)}(k-1)}{x^{(0)}(k)},\ k=2,3,\cdots,n \tag{37}$$

If $\sigma^{(0)}(k)\in\left(e^{-2/(n+2)},e^{2/(n+2)}\right)$ for k=2,3,...,n (Deng, 2002a) then 1-AGO can secure the grey exponential trend of 1-AGO sequence $X^{(1)}=\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\}$ where $x^{(1)}(k)=\sum_{j=1}^{n}x^{(0)}(j)$ such that the degree of data assimilation will be high.

However, the behavior of sampling data from unknown system is not controllable or predictable and therefore we often face the situations that standard GM(1,1) model does not fit the data well. The diversity of real-world needs us to facilitate more potential grey differential equation models satisfying the following principle:

(i) The shadow (constraint) differential equation $dx^{(1)}/dt=F(t,x^{(1)};\theta)$ must be solvable with the *two* parameters, denoted by θ, and accordingly defines the form of regression model for parameter estimation.

(ii) The two-parameter regression model defined by the shadow differential equation must takes *linear form* in the two parameters with data sequence $X^{(0)}$ as observation vector Y and using mean 1-AGO sequence $Z^{(1)}$ or appropriate transformation $\eta\left(Z^{(1)}\right)$ for the construction of the regression model matrix.

One of the simplest forms of the regression models in accordance with the principle stated is,

$$x^{(0)}(k)+\beta\eta\left[z^{(1)}(k)\right]=\alpha,\ k=2,3,\cdots,n \tag{38}$$

Then the shadow differential equation takes the form:

$$\frac{dx^{(1)}}{dt}+\beta\eta\left(x^{(1)}\right)=\alpha \tag{39}$$

i.e.,

$$\frac{dx^{(1)}}{dt}=\alpha-\beta\eta\left(x^{(1)}\right) \tag{40}$$

Equation 39 gives an induced family of appropriate shadow differential equation satisfying Principle (i) and (ii). Hence, it will be worthwhile to investigate the appropriate choice of the

couple of regression model and constraint differential equation model for potential extensions to Deng's (1985) GM(1,1) models in the next section.

3.12.7 Families of Shadow Differential Equations

In this section, we will explore some families suitable to meet the requirements as shadow differential equation.

3.12.7.1 The Differential Equation with Separable Variables

The basic form of separable-variable equation is:

$$f_1(t)\,g_1(x)\,dt + f_2(t)\,g_2(x)\,dx = 0 \tag{41}$$

i.e.,

$$\int \frac{f_1(t)}{f_2(t)}\,dt + \int \frac{g_2(x)}{g_1(x)}\,dx = c \text{ (constant)} \tag{42}$$

For example, $f_1 = \beta f_2$ and g_2=g_1/y, then a functional of form $y = \alpha e^{-\beta t}$ may obtained. The regression model coupled by the constraint is,

$$x^{(0)}(k) = -f_1(k;\theta)\,g_1\left(z^{(1)}(k)\right) \Big/ \left(f_2(k)\,g_2\left(z^{(1)}(k)\right)\right) \tag{43}$$

3.12.7.2 The Homogeneous Differential Equation

The basic form of the homogeneous differential equation is,

$$\frac{dx}{dt} = F\left(\frac{x}{t}\right) \tag{44}$$

Let u=x/t, then we obtain:

$$\ln t = \int \frac{du}{F(u) - u} + c \tag{45}$$

For example, $F(u) - u = \alpha \exp(-\beta u) + 1$, then,

$$x = \frac{t}{\beta}\ln\left(c_0 t^{\alpha\beta} - \alpha\right) \tag{46}$$

The regression model should take the form:

$$x^{(0)}(k) = F\left(\frac{z^{(1)}(k)}{k};\theta\right) \tag{47}$$

For model conveniences, $F(\cdot;\theta)$ should generate a linear additive form.

3.12.7.3 The Linear Differential Equation

The basic form of the linear differential equation is:

$$\frac{dx}{dt} + p(t)x = q(t) \tag{48}$$

Then the solution to equation 48 is,

$$x = \exp(-\int p(t)dt)\left[\int q(t)\exp(\int p(u)du)dt + c\right] \tag{49}$$

For example, let $p(t) = \beta$ and $q(t) = t/(1+\beta t)^2$, then

$$x = \frac{1}{\beta^2(1+\beta t)} + c\exp(-\beta t) \tag{50}$$

This is a direct generalization to GM(1,1) model, the regression model is:

$$x^{(0)}(k) + p(k)z^{(1)}(k) = q(k),\ k = 2,3,\cdots,n \tag{51}$$

3.12.7.4 Differential Equation of Form $x = F(t, dx/dt)$

Let $\varphi = dx/dt$. The equation can take the form,

$$\left(\varphi - \frac{\partial F}{\partial t}\right)dt - \frac{\partial F}{\partial \varphi}d\varphi = 0 \tag{52}$$

Assuming a general solution $\varphi = \rho(t, c_0)$ or $t = \psi(\varphi, c_0)$, then:

$$x = F\left(t, \varphi\left(t, c_0\right)\right)$$
or
$$\begin{cases} t = \psi\left(\varphi, c_0\right) \\ x = F\left(\psi\left(\varphi, c_0\right), \varphi\right) \end{cases} \tag{53}$$

For example, let $F(t,\varphi) = \beta\varphi t$, then $\varphi - \partial F/\partial t = \varphi(1-\beta)$ and $\partial F/\partial\varphi = \beta t$, therefore,

$$\frac{dx}{dt} = c_0 t^{1/\beta - 1} \tag{54}$$

And therefore, $x = c_0 \beta t^{1/\beta} + c_1$. The regression model associated with equation 52 takes the form:

$$\left(x^{(0)}(k) - \frac{\partial F}{\partial t}\left(k, x^{(0)}(k); \theta\right)\right) - \frac{\partial F}{\partial \varphi}\left(k, x^{(0)}(k); \theta\right) x^{(-1)}(k) = 0 \tag{55}$$

where $\partial F/\partial t$ and $\partial F/\partial p$ should generate linear functional form in t and φ.

3.12.7.5 Lagrangian Differential Equation

The basic form:

$$x = t f_1(\varphi) + f_2(\varphi) \tag{56}$$

where $\varphi = dx/dt$ .The Lagrangian equation can be changed to,

$$\frac{dt}{d\varphi} - \frac{f_1'(\varphi)}{\varphi - f_1(\varphi)} = \frac{f_2'(\varphi)}{\varphi - f_2(\varphi)} \tag{57}$$

Accordingly the regression model could be:

$$\frac{1}{x^{(-1)}(k)} = \frac{f_1'\left(x^{(0)}(k); \alpha\right)}{x^{(0)}(k) - f_1\left(x^{(0)}(k); \alpha\right)} + \frac{f_2'\left(x^{(0)}(k); \beta\right)}{x^{(0)}(k) - f_2\left(x^{(0)}(k); \beta\right)} \tag{58}$$

Notice that the regression equation should be linear in α and β or their functions.

3.12.7.6 Clairaut Differential Equation

The basic form:

$$x = tx' + F(x') \tag{59}$$

The solution form is,

$$x = c_0 t + F(c_0;\theta) \tag{60}$$

And the regression takes the form,

$$z^{(1)}(k) = c_0 k + F(c_0;\theta) \tag{61}$$

Where $F(\cdot,\theta)$ should be linear in θ (a scalar parameter).

3.12.7.7 Riccati Differential Equation

Let $\varphi = dx/dt$, then the basic form:

$$\varphi = p(t)x^2 + q(t)x + r(t) \tag{62}$$

Assuming a particular solution $x=x_1(t)$, perform the transformation $x=x_1+1/u$, the original equation can be converted into a linear differential equation:

$$\frac{du}{dt} + \left[q(t) + 2p(t)x_1(t)\right]u + p(t) = 0 \tag{63}$$

Then $u=1/(x-x_1)$, and $du/dt=-(dx/dt-dx_1/dt)/(x-x_1)^2$. Accordingly, the observational sequence $U^{(0)} = \left\{u^{(0)}(1), u^{(0)}(2), \cdots, u^{(0)}(n)\right\}$ with,

$$\begin{aligned} u^{(0)}(k) &= -\left(x^{(0)}(k) - x_1^{(0)}(k)\right) \Big/ \left(z^{(1)}(k) - z_1^{(1)}(k)\right)^2 \\ u^{(1)}(k) &= 1 \Big/ \left(z^{(1)}(k) - z_1^{(1)}(k)\right) \\ k &= 2, 3, \cdots, n \end{aligned} \tag{64}$$

Table 3.14.1 lists the above discussions on the extended GM(1,1) models.

Table 3.12.1 Extended GM(1,1) Models.

Form	Constraint	Regression
1	$f_1(t)g_1(x)dt+f_2(t)g_2(x)dx=0$	$x^{(0)}(k)=-f_1(k;\theta)g_1\left(z^{(1)}(k)\right)\Big/\left(f_2(k)g_2\left(z^{(1)}(k)\right)\right)$
2	$dx/dt=F(x/t)$	$x^{(0)}(k)=F\left(z^{(1)}(k)/k;\theta\right)$
3	$dx/dt+p(t)x=q(t)$	$x^{(0)}(k)+p(k)z^{(1)}(k)=q(k)$
4	$x=F(t,\varphi)$	$\left(x^{(0)}(k)-\frac{\partial F}{\partial t}\left(k,x^{(0)}(k);\theta\right)\right)-\frac{\partial F}{\partial \varphi}\left(k,x^{(0)}(k);\theta\right)x^{(-1)}(k)=0$
5	$x=tf_1(\varphi)+f_2(\varphi)$	$\frac{1}{x^{(-1)}(k)}=\frac{f_1'\left(x^{(0)}(k);\alpha\right)}{x^{(0)}(k)-f_1\left(x^{(0)}(k);\alpha\right)}+\frac{f_2'\left(x^{(0)}(k);\beta\right)}{x^{(0)}(k)-f_2\left(x^{(0)}(k);\beta\right)}$
6	$x=t\varphi+F(\varphi)$	$z^{(1)}(k)=c_0k+F(c_0;\theta)$
7	$\varphi=p(t)x^2+q(t)x+r(t)$	$u^{(0)}(k)+\left[q(k)+2p(k)z_{x_1}^{(1)}(k)\right]z_u^{(1)}(k)+p(k)=0$

3.12.7.8 Other Forms

Forms of differential equations, for example, $t=F(x,x')$, or $F(t,x')=0$, or $F(x,x')=0$ is also good candidate to play the role of shadow differential equation.

3.12.8 Conclusion

In this paper, we examine Deng's (1985) GM(1,1) model from its component-level models: the shadow differential equation, the regression model and the intrinsic difference model. Furthermore, we examine the coupling nature of the standard GM(1,1) model, i.e., the GM(1,1) model is a variational problem with differential equation as its constraint. The "coupling" nature also restricts the form of the regression model in the object function and associated AGO and IAGO operations whenever necessary. Based on these examinations, we state the coupling principle for guiding the generalization to Deng's (1985) GM(1,1) model. Seven extended GM(1,1) models are proposed accordingly, and they can be used when the standard GM(1,1) model does not work adequately. We are expecting this foundational work will provide more flexible and feasible GM(1,1) models for fitting small sample data with high degree of data assimilation capability.

3.13 Bivariate Credibility-Copulas

Renkuan Guo, Danni Guo, and Xiang Li

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
rguo@stats.uct.ac.za dannig@stats.uct.ac.za xli@stats.uct.ac.za

This paper has been published in the Proceedings of the 1st International Conference on System Integration and Reliability Improvements (SIRI 2006), December 13-15, 2006, Hanoi, Vietnam, pp. 119-123.

3.13.1 Introduction

Fuzzy uncertainty problem appeared in system reliability improvement process lies in shorting of system structural clarity, shorting of the underlying mechanism of the interaction between subsystems and shorting of overall information of the system as a whole. Accordingly, the methodology to solve the fuzzy reliability improvement of the complex system should be developed in terms of the basic concept of fuzzy mathematics.

Fuzzy mathematics initiated by Zadeh (1965) facilitated a foundation dealing with vague phenomena in system modelling. However, the fuzzy mathematical foundation initiated by Zadeh (1965, 1978) is membership function and possibility measure based and widely used. The possibility measure was originally expected to play the role of probability measure in probability theory, but did not because possibility measure does not possess self-duality property as that in probability theory. In standard probability theory, random variable and the distribution function play important roles for converting set-based arguments into variable-based arguments, which result in great conveniences in applications. Kaufmann (1975) first proposed the concept of fuzzy variable with the intention of creating the counterpart in probability theory. However, unfortunately, Kaufmann's fuzzy variable is another name of fuzzy subset and the mathematical operations are difficult to handle.

To resolve this dilemma, Liu (2004, 2006) proposed an axiomatic foundation for modelling fuzzy phenomena, named as credibility theory. The credibility measure possesses self-duality property and is able to play the role of that in probability theory. Furthermore, fuzzy variable concept and its distribution, which are parallel to these in probability theory, are developed. Furthermore, from practical point of view, the revelation of the full functional relation plays a critical role in system improvement analysis and therefore fuzzy correlation (i.e., standardized covariance) concept based on standard credibility theory, which describes a linear relation between two fuzzy variables, cannot facilitate a full relationship analysis between two fuzzy variables. However, the defining of full functional relation requires conditional credibility measure concept, which is impossible to handle based on Liu's standard credibility theory.

In this paper, based on Liu's non-classical credibility measure theory, i.e., $(\vee,\cdot)$-credibility measure theory, we explore the basic property of uniform-distributed fuzzy variable and explore joint uniform distribution for the fuzzy bivariate variables. Then, we propose the concept of fuzzy copula on the ground of $(\vee,\cdot)$-credibility measure theory, named as credibility-copula, for the

characterization of the full relationship among fuzzy variables. Finally, we explore a decomposition of credibility-copula function into product copula and an adjusted dependence function.

3.13.2 A Review of $(\vee,\cdot)$-Credibility Measure Theory

Let Θ be a nonempty set, and $\mathfrak{P}(\Theta)=2^{\Theta}$ the power set on Θ. Each element, let us say, $A\subset\Theta, A\in\mathfrak{P}(\Theta)$ is called an event. A number denoted as $\mathrm{C\tilde{r}}(A)$, $0\le \mathrm{C\tilde{r}}(A)\le 1$, is assigned to event $A\in\mathfrak{P}(\Theta)$, which indicates the credibility that event $A\in\mathfrak{P}(\Theta)$ occurs. $\mathrm{C\tilde{r}}(A)$ satisfies following axioms (Liu, 2004, 2006):

Axiom 1: $\mathrm{C\tilde{r}}(\Theta)=1$.

Axiom 2: $\mathrm{C\tilde{r}}(\cdot)$ non-decreasing, i.e., $\mathrm{C\tilde{r}}\{A\}\le\mathrm{C\tilde{r}}\{B\}$ whenever $A\subset B$.

Axiom 3: $\mathrm{C\tilde{r}}(\cdot)$ is self-dual, i.e., $\mathrm{C\tilde{r}}\{A\}+\mathrm{C\tilde{r}}\{A^c\}=1$ for any $A\in\mathfrak{P}(\Theta)$.

Axiom 4: $\mathrm{C\tilde{r}}\{\bigcup_i A_i\}\wedge 0.5=\sup_i\left[\mathrm{C\tilde{r}}\{A_i\}\right]$ for any $\{A_i\}$ with $\mathrm{C\tilde{r}}(A_i)\le 0.5$.

Axiom 5: Let set functions $\mathrm{C\tilde{r}}_k(\cdot):2^{\Theta_k}\to[0,1]$ satisfy **Axiom 1-4**, and $\Theta=\Theta_1\times\Theta_2\times\cdots\times\Theta_p$, then:

$$\mathrm{C\tilde{r}}(\theta_1,\theta_2,\cdots,\theta_p)=\begin{cases}\dfrac{1}{2}\displaystyle\prod_{k=1}^{p}\left(2\mathrm{C\tilde{r}}(\theta_k)\wedge 1\right) & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}<0.5\\ \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\} & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}\ge 0.5\end{cases} \tag{1}$$

for each $(\theta_1,\theta_2,\cdots,\theta_p)\in\Theta$. In this case, we write $\mathrm{C\tilde{r}}=\mathrm{C\tilde{r}}_1\times\mathrm{C\tilde{r}}_2\times\cdots\times\mathrm{C\tilde{r}}_p$.

Definition 1: Any set function $\mathrm{C\tilde{r}}:\mathfrak{P}(\Theta)\to[0,1]$ satisfies **Axiom 1-5** is called a $(\vee,\cdot)$-credibility measure. The triple $(\Theta,2^{\Theta},\mathrm{C\tilde{r}})$ is called the $(\vee,\cdot)$-credibility measure space.

Definition 2: The (induced) membership function of a fuzzy variable ξ on $(\Theta,\mathfrak{P}(\Theta),\mathrm{C\tilde{r}})$ is:

$$\mu(x)=\left(2\mathrm{C\tilde{r}}\{\xi=x\}\right)\wedge 1,\quad x\in\mathbb{R} \tag{2}$$

Conversely, for given membership function the credibility measure is determined by the credibility inversion theorem.

Theorem 3: (Liu, 2004, 2006) Let ξ be a fuzzy variable with membership function μ. Then for $\forall B \subset \mathbb{R}$,

$$\mathrm{C\tilde{r}}\{\xi \in B\} = \frac{1}{2}\left(\sup_{x \in B} \mu(x) + 1 - \sup_{x \in B^c} \mu(x)\right), \quad B \subset \mathbb{R} \tag{3}$$

As an example, if set B is degenerated into a point x, then

$$\mathrm{C\tilde{r}}\{\xi = x\} = \frac{1}{2}\left(\mu(x) + 1 - \sup_{y \neq x} \mu(y)\right), \quad \forall x \in \mathbb{R} \tag{4}$$

Definition 4: (Liu, 2004, 2006) The credibility distribution $\Phi : \mathbb{R} \to [0,1]$ of a fuzzy variable ξ on $(\Theta, \mathfrak{P}(\Theta), \mathrm{C\tilde{r}})$ is:

$$\Phi(x) = \mathrm{C\tilde{r}}\{\theta \in \Theta \,|\, \xi(\theta) \leq x\} \tag{5}$$

That is the credibility distribution $\Phi(x)$ is the accumulated credibility grade that the fuzzy variable ξ takes a value less than or equal to a real-number $x \in \mathbb{R}$. Generally speaking, the credibility distribution Φ is neither left-continuous nor right-continuous.

Theorem 5: Let ξ be a fuzzy variable on $(\Theta, \mathfrak{P}(\Theta), \mathrm{C\tilde{r}})$ with membership function μ. Then its credibility distribution,

$$\Phi(x) = \frac{1}{2}\left(\sup_{y \leq x} \mu(y) + 1 - \sup_{y > x} \mu(y)\right), \ \forall x \in \Re \tag{6}$$

Definition 6: Let Φ be the credibility distribution of the fuzzy variable ξ. Then function $\phi : \mathbb{R} \to [0, +\infty)$ of a fuzzy variable ξ is called a credibility density function such that,

$$\Phi(x) = \int_{-\infty}^{x} \phi(y)\mathrm{d}y, \quad \forall x \in \mathbb{R} \tag{7}$$

We now consider the credibility of an event A after it has been learned that some other event B has occurred. This new credibility of A is called the conditional credibility of the event A given B, denoted by $\mathrm{C\tilde{r}}(A \,|\, B)$.

Definition 7: Let $(\Theta, 2^{\Theta}, \mathrm{C\tilde{r}})$ be a credibility measure space and let $\forall A, B \in 2^{\Theta}$. Then the conditional credibility measure of A given B, $\mathrm{C\tilde{r}}(A \,|\, B)$,

$$\mathrm{C\tilde{r}}(A \mid B)=\frac{1}{2}\left[\frac{(2\mathrm{C\tilde{r}}(A\cap B))\wedge 1}{(2\mathrm{C\tilde{r}}(B))\wedge 1}+1-\frac{(2\mathrm{C\tilde{r}}(A^{c}\cap B))\wedge 1}{(2\mathrm{C\tilde{r}}(B))\wedge 1}\right] \tag{8}$$

Provided $\mathrm{C\tilde{r}}(B)>0$.

3.13.3 Concepts of Bivariate Credibility-Copula

In this section, we will explore the basic property of uniform distributed fuzzy variable and then the credibility distribution transformation on the ground of $(\vee,\cdot)$-credibility measure.

3.13.3.1 Uniform Distributed Fuzzy Variable

In probabilistic context, uniform distribution and the distribution transformation play very important roles. Let us explore whether similar developments exist on the ground of the credibility measure theory. We say a fuzzy variable υ to be (standard) uniform distributed if its credibility density takes the form:

$$\phi_{\upsilon}(x)=\begin{cases}1 & 0\le x\le 1\\ 0 & \text{otherwise}\end{cases} \tag{9}$$

The credibility distribution of fuzzy standard uniform variable υ is:

$$\Phi_{\upsilon}(x)=\begin{cases}0 & x<0\\ x & 0\le x<1\\ 1 & x\ge 1\end{cases} \tag{10}$$

And thus the induced membership function for the fuzzy of fuzzy standard uniform variable υ is:

$$\mu_{\upsilon}(x)=\begin{cases}2x & 0\le x<0.5\\ 2-2x & 0.5\le x\le 1\\ 0 & x<0 \text{ or } x>1\end{cases} \tag{11}$$

Conversely, the isosceles triangle membership located at (0,0), (0.5,1), and (1,0) characterizes a fuzzy standard uniform-distributed variable.

3.13.3.2 Credibility Distribution Transformation

Let ξ be a fuzzy variable with credibility distribution function $\Phi_\xi(\cdot)$. We are interested in the distribution for fuzzy variable $\Phi_\xi(\xi)$. Recall that the membership function for fuzzy variable ξ is,

$$\mu_\xi(\xi)=\begin{cases}2\Phi_\xi(\xi) & \text{if} \quad \Phi_\xi(\xi)<0.5 \\ 1 & \text{if} \quad \lim\limits_{y\uparrow\xi}\Phi_\xi(y)<0.5\le\Phi_\xi(\xi) \\ 2-2\Phi_\xi(\xi) & \text{if} \quad 0.5\le\lim\limits_{y\uparrow\xi}\Phi_\xi(y)\end{cases} \tag{12}$$

Then for any given value $x=\Phi_\xi(\xi)$, $\xi=\Phi^{-1}(x)$, thus we obtain the membership function for $X=\Phi_\xi(\xi)$,

$$\mu_{\Phi_\xi(\xi)}(x)=\begin{cases}2x & 0\le x<0.5 \\ 2-2x & 0.5\le x\le 1 \\ 0 & x<0 \text{ or } x>1\end{cases} \tag{13}$$

In other words, the transformation of a fuzzy variable by its credibility distribution will result in a fuzzy standard uniform distributed variable.

3.13.3.3 Definition of Bivariate Credibility-Copula

Similar to probabilistic copula theory, copula is a dependence index measuring a couple of two fuzzy variables. Let X and Y be two continuous fuzzy variable with credibility distributions Φ_X and Φ_Y respectively and $H_{XY}(\cdot,\cdot)$ be their joint credibility distribution. Let $\mathbb{I}=[0,1]$.

Definition 8: A bivariate credibility-copula is a function $C:\mathbb{I}\times\mathbb{I}\to\mathbb{I}$ such that:

(i) $C(0,x)=C(x,0)=0$ and $C(1,x)=C(x,1)=x$ for all $x\in\mathbb{I}$;

(ii) $C(\cdot,\cdot)$ is 2-*increasing*: for $\forall a,b,c,d\in\mathbb{I}$, $a<b$, and $c<d$,

$$v_C([a,b]\times[c,d])=C(b,d)-C(a,d)-C(b,c)+C(a,c)\ge 0 \tag{14}$$

The function v_C is called the C-volume of the rectangle $[a,b]\times[c,d]$.

In other words, a credibility-copula is the restriction to the unit square $\mathbb{I}\times\mathbb{I}=[0,1]\times[0,1]$ of a bivariate credibility distribution function whose margins are standard uniform. More formally, a copula C induces a credibility measure on $\mathbb{I}\times\mathbb{I}$ in terms of ν_C, i.e.,

$$\nu_C\left([0,u]\times[0,v]\right)=C(u,v) \tag{15}$$

Now let us state the credibility theoretical version of Sklar's theorem (Nelsen, 1999).

Theorem 9: Let H be a bivariate-credibility distribution function with marginal credibility distributions Φ_X and Φ_Y respectively. Then there exists a credibility-copula C such that $H(x,y)=C\left(\Phi_X(x),\Phi_Y(y)\right)$. Conversely, for any credibility distribution functions Φ_X and Φ_Y and any credibility-copula C, the function H defined above is a two-dimensional credibility distribution with margins Φ_X and Φ_Y respectively. Furthermore, if Φ_X and Φ_Y are continuous, the credibility-copula is unique.

Based on the credibility theoretical version of Sklar's theorem, given a two-dimensional credibility distribution $H(x,y)$ and marginal credibility distributions Φ_X and Φ_Y, the credibility-copula is,

$$C(u,v)=H\left(\Phi_X^{-1}(u),\Phi_Y^{-1}(v)\right) \tag{16}$$

Where the inverse of the credibility distribution denoted by $F^{-1}(\cdot)$ is defined as:

$$F^{-1}(u)=\sup\left\{x\mid F(x)\le u\right\} \tag{17}$$

By noticing the credibility distribution transformation arguments, $U=\Phi_X(X)$ and $V=\Phi_Y(Y)$ are fuzzy standard uniform-distributed variables on $[0,1]$, the credibility-copula is the joint uniform credibility distribution of two uniform-distributed fuzzy variable on $[0,1]\times[0,1]$. Therefore, credibility-copula fully describes the functional relation between two fuzzy variables.

3.13.3.4 Properties of Bivariate Credibility-Copulas

If $C(u,v)$, in nature a joint credibility uniform distribution, possesses a joint credibility density, denoted by $\partial C(u,v)/\partial u\partial v$, we say $C(u,v)$ is absolutely continuous. Given a two-dimensional

credibility distribution $H(x,y)$ and marginal credibility distributions Φ_X and Φ_Y, then it can be shown that:

$$\max\{\Phi_X(x)+\Phi_Y(y)-1,0\} \le H(x,y) \le \min\{\Phi_X(x),\Phi_Y(y)\} \tag{18}$$

Or equivalently,

$$\varpi(u,v)=\max\{u+v-1,0\} \le C(u,v) \le \min\{u,v\}=\omega(u,v) \tag{19}$$

It can also be shown that $\varpi(u,v)$, $\varpi(u,v)$, and $\chi(u,v)=uv$ are credibility-copulas. We also note that for continuous fuzzy variables X and Y:

(i) If and only if X and Y is an increasing function of the other, then $C(u,v)=\omega(u,v)$;

(ii) If and only if X and Y is a decreasing function of the other, then $C(u,v)=\varpi(u,v)$;

(iii) If and only if X and Y are independent;

Then $C(u,v)=\chi(u,v)=uv$ (which is referred to as *product copula*).

3.13.3.5 Archimedean Family of Bivariate Credibility-Copulas

Paralleling to probabilistic copula theory, the Archimedean credibility-copula is a function C: $\mathbf{I}\times\mathbf{I}\to\mathbf{I}$ defined by the generator ϕ such that,

$$C(u,v)=\phi^{[-1]}(\phi(u)+\phi(v)) \tag{20}$$

where $\phi^{[-1]}$ is the pseudo-inverse of ϕ:

$$\phi^{[-1]}(t)=\begin{cases}\phi^{-1}(t) & t\in[0,\phi(0)]\\ 0 & t>\phi(0)\end{cases} \tag{21}$$

The generator $\phi:\mathbf{I}\to[0,\infty]$, which is a strictly monotone decreasing and continuous convex function. An important property of Archimedean copula family is the associativity: i.e., $C(C(u,v),w)=C(u,C(v,w))$. A simple example of Archimedean copula is the generator takes the form $\ln\left((1-\theta(1-t))/t\right)$ and the generating copula is $C(u,v)=uv/(1-\theta(1-u)(1-v))$.

3.13.4 A Decomposition of Bivariate Copulas

Recently, Dos Anjos (2005) pointed out that in probabilistic context there exists a local dependence measure, which provides explicit and precise information of the underlying dependence structure and helps to reformulate bivariate distribution and associated copula. Now let us to explore the parallel developments on the ground of credibility theory.

Definition 10: (Expected value and variance) Let X be a fuzzy variable with credibility distribution Φ_X. If

$$\lim_{x\to-\infty}\Phi_X(x)=0,\quad \lim_{x\to\infty}\Phi_X(x)=1 \tag{22}$$

And the Lebesgue-Stieltjes integral $\int_{-\infty}^{\infty} x d\Phi_X(x)$ is finite, then we define $\int_{-\infty}^{\infty} x d\Phi_X(x)$ as the expected value of fuzzy variable X and denoted by $E[X]$. Furthermore, we define $E\left[(X-E[X])^2\right]$ as the variance of fuzzy variable and denoted by $V[X]$.

Definition 11: (Covariance) Let two fuzzy variables X and Y have a bivariate credibility distribution $H(x,y)$ and the two marginal credibility distributions Φ_X and Φ_Y with finite expected values $E[X]$ and $E[Y]$ respectively. Then $E\left[(X-E[X])(Y-E[Y])\right]$ and $E\left[(X-E[X])(Y-E[Y])\right]\Big/\left(\sqrt{V[X]}\sqrt{V[Y]}\right)$ are called the covariance and correlation of fuzzy variables X and Y respectively.

Remark 12: Under certain conditions, it can be shown that,

$$E\left[(X-E[X])(Y-E[Y])\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E[X])(y-E[Y])\,dH(x,y) \tag{23}$$

Definition 13: Let two fuzzy variables X and Y have a bivariate credibility distribution $H(x,y)$ and the two marginal credibility distributions Φ_X and Φ_Y. The underlying Spearman function ρ_H is the correlation coefficient between $\vartheta\{\Phi_X(X)\le\Phi_X(x)\}$ and $\vartheta\{\Phi_Y(Y)\le\Phi_Y(y)\}$ for each $(x,y)\in[-\infty,\infty]\times[-\infty,\infty]$,

$$\rho_H(x,y)=\frac{H(x,y)-\Phi_X(x)\Phi_Y(y)}{\sqrt{\Phi_X(x)\Phi_Y(y)(1-\Phi_X(x))(1-\Phi_Y(y))}} \tag{24}$$

Similarly, for copulas:

$$\rho_C(u,v)=\frac{C(u,v)-uv}{\sqrt{uv(1-u)(1-v)}} \tag{25}$$

Theorem 14: Let two fuzzy variables X and Y have a bivariate credibility distribution $H(x,y)$ and the two marginal credibility distributions Φ_X and Φ_Y with associated copula $C(u,v)$, then,

$$C(u,v)=uv+\rho_C(u,v)\sqrt{uv(1-u)(1-v)} \tag{26}$$

and

$$H(x,y)=\Phi_X(x)\Phi_Y(y)+\rho_H(x,y)\sqrt{\Phi_X(x)\Phi_Y(y)(1-\Phi_X(x))(1-\Phi_Y(y))} \tag{27}$$

Furthermore,

$$\rho_H\left(\Phi_X(x),\Phi_Y(y)\right)=\rho_C(u,v) \tag{28}$$

for all $(x,y)\in[-\infty,\infty]\times[-\infty,\infty]$ such that $u=\Phi_X(x)$ and $v=\Phi_Y(y)$.

Remark 15: As a common sense, the product copula, $\chi(u,v)$, represents non-dependence or independence. The value of product copula, $\chi(u,v)$, measures the degree of independence. The alternative representation of copula reveals a fundamental fact that the degree of dependence between two fuzzy variables is the sum of the product copula factor $\chi(u,v)=uv$ and an adjusted local dependence factor, $D(u,v)$. In this sense, copula is a quantity composed of two factors: independence measure and dependence measure. Therefore, the distance between copula, $C(u,v)$ and the product copula, $\chi(u,v)$, $D(u,v)=C(u,v)-\chi(u,v)$, measures the true degree of dependence, i.e., the degree apart from independence. While ρ_C is the standardized true degree of dependence in a local sense, i.e., at about given point $(u,v)\in\mathbb{I}\times\mathbb{I}$.

The decomposition of a copula help us a refined understanding of the measure of dependence. For example, let $C(u,v)=uv/(1-\theta(1-u)(1-v))$, then $D(u,v)=uv(1-1/(1-\theta(1-u)(1-v)))$

measures the pure bivariate dependence. Another is, $C(u,v)=uv+\theta uv(1-u)(1-v)$, $|\theta|\leq 1$, then $D(u,v)=uv\theta(1-u)(1-v)$ and $\rho_C(u,v)=\theta\sqrt{uv(1-u)(1-v)}$.

3.13.5 Conclusion

In this paper, we heavily reviewed the related work on $(\vee,\cdot)$-credibility measure theory. Based on the review, we propose the concept of copula for bivariate fuzzy variables, called as credibility-copula, which is a bivariate uniform credibility distribution in the sense of the credibility measure theory. Parallel to the basic understanding in probabilistic copula, we regard the credibility-copula to be a measure of the full bivariate dependence. Furthermore, we establish a decomposition of credibility-copula. The decomposition explicitly reveals that copula is a total measure of independence as well as dependence and only the removal of the product copula component from the copula could accurately describe the true dependence between the two fuzzy variables. We must point out that the mathematical formulation developments for credibility-copula and the decomposition are similar to those in the probabilistic copula theory. However, the underlying mechanism on the ground of credibility theory is different from that on the ground of probability theory because credibility measure theory and probability measure theory describe different phenomena: one is fuzzy uncertainty and other is random uncertainty although we did not explore the bivariate uniform credibility distribution in details.

3.14 Kernel Estimation for Bivariate Credibility Copulas

Xiang Li, Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
xli@stats.uct.ac.za dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper has been published in the Proceedings of the 1st International Conference on System Integration and Reliability Improvements (SIRI 2006), December 13-15, 2006, Hanoi, Vietnam, pp. 52-56.

3.14.1 Introduction

Fuzzy uncertainty problem appeared in system reliability improvement process lies in the shorting of system structural clarity, the shorting of the underlying mechanism of the interaction between subsystems, and the shorting of overall information of the system as a whole. Accordingly, the methodology to solve the fuzzy reliability improvement of the complex system should be developed, in terms of the basic concept of fuzzy mathematics.

Fuzzy mathematics initiated by Zadeh (1965) facilitated a foundation dealing with vague phenomena in system modelling. However, the fuzzy mathematical foundation initiated by Zadeh (1965, 1978) is based on a membership function and a possibility measure, and are widely used. The possibility measure was originally expected to play the role of probability measure in probability theory, but did not because possibility measure does not possess self-duality property as that in probability theory. In standard probability theory, random variable and the distribution function play important roles for converting set-based arguments into variable-based arguments, which result in great conveniences in applications. Kaufmann (1975) first proposed the concept of fuzzy variable with the intention of creating the counterpart in probability theory. However, unfortunately, Kaufmann's fuzzy variable is another name of fuzzy subset and the mathematical operations are difficult to handle.

To resolve this dilemma, Liu (2004, 2006) proposed an axiomatic foundation for modelling fuzzy phenomena, named as credibility theory. The credibility measure possesses self-duality property and is able to play the role of that in probability theory. Furthermore, fuzzy variable concept and its distribution, which are parallel to these in probability theory, are developed. Furthermore, from practical point of view, the revelation of the full functional relation plays a critical role in system improvement analysis and therefore fuzzy correlation (i.e., standardized covariance) concept based on standard credibility theory, which describes a linear relation between two fuzzy variables, cannot facilitate a full relationship analysis between two fuzzy variables. However, the defining of full functional relation requires conditional credibility measure concept, which is impossible to handle based on Liu's standard credibility theory.

In a recent paper, Guo et al. (2006a, 2006b) proposed the credibility-copula concept and explored its fundamental decomposition. However, unless the functional form of a copula is available, there will be no practical usage of credibility-copulas in system reliability

improvements. In this paper, we propose a kernel-estimation based on the maximum entropy principle under the framework of $(\vee,\cdot)$-credibility measure theory.

3.14.2 A Review of $(\vee,\cdot)$-Credibility Measure Theory

Let Θ be a nonempty set, and $\mathfrak{P}(\Theta)=2^{\Theta}$ the power set on Θ. Each element, let us say, $A\subset\Theta, A\in\mathfrak{P}(\Theta)$ is called an event. A number denoted as $\mathrm{C\tilde{r}}(A)$, $0\le \mathrm{C\tilde{r}}(A)\le 1$, is assigned to event $A\in\mathfrak{P}(\Theta)$, which indicates the credibility that event $A\in\mathfrak{P}(\Theta)$ occurs. $\mathrm{C\tilde{r}}(A)$ satisfies following axioms (Liu, 2004, 2006):

Axiom 1: $\mathrm{C\tilde{r}}(\Theta)=1$.

Axiom 2: $\mathrm{C\tilde{r}}(\cdot)$ non-decreasing, i.e., $\mathrm{C\tilde{r}}\{A\}\le\mathrm{C\tilde{r}}\{B\}$ whenever $A\subset B$.

Axiom 3: $\mathrm{C\tilde{r}}(\cdot)$ is self-dual, i.e., $\mathrm{C\tilde{r}}\{A\}+\mathrm{C\tilde{r}}\{A^c\}=1$ for any $A\in\mathfrak{P}(\Theta)$.

Axiom 4: $\mathrm{C\tilde{r}}\{\bigcup_i A_i\}\wedge 0.5=\sup_i\left[\mathrm{C\tilde{r}}\{A_i\}\right]$ for any $\{A_i\}$ with $\mathrm{C\tilde{r}}(A_i)\le 0.5$.

Axiom 5: Let set functions $\mathrm{C\tilde{r}}_k(\cdot):2^{\Theta_k}\to[0,1]$ satisfy **Axiom 1-4**, and $\Theta=\Theta_1\times\Theta_2\times\cdots\times\Theta_p$, then:

$$\mathrm{C\tilde{r}}(\theta_1,\theta_2,\cdots,\theta_p)=\begin{cases}\dfrac{1}{2}\displaystyle\prod_{k=1}^{p}\left(2\mathrm{C\tilde{r}}(\theta_k)\wedge 1\right) & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}<0.5\\ \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\} & \text{if } \min\limits_{1\le k\le p}\left\{\mathrm{C\tilde{r}}(\theta_k)\right\}\ge 0.5\end{cases} \tag{1}$$

for each $(\theta_1,\theta_2,\cdots,\theta_p)\in\Theta$. In this case, we write $\mathrm{C\tilde{r}}=\mathrm{C\tilde{r}}_1\times\mathrm{C\tilde{r}}_2\times\cdots\times\mathrm{C\tilde{r}}_p$.

Definition 1: Any set function $\mathrm{C\tilde{r}}:\mathfrak{P}(\Theta)\to[0,1]$ satisfies **Axiom 1-5** is called a $(\vee,\cdot)$-credibility measure. The triple $(\Theta,2^{\Theta},\mathrm{C\tilde{r}})$ is called the $(\vee,\cdot)$-credibility measure space.

Definition 2: The (induced) membership function of a fuzzy variable ξ on $(\Theta,\mathfrak{P}(\Theta),\mathrm{C\tilde{r}})$ is:

$$\mu(x)=\left(2\mathrm{C\tilde{r}}\{\xi=x\}\right)\wedge 1,\quad x\in\mathbb{R} \tag{2}$$

Conversely, for given membership function the credibility measure is determined by the credibility inversion theorem.

Theorem 3: (Liu, 2004, 2006) Let ξ be a fuzzy variable with membership function μ. Then for $\forall B \subset \mathbb{R}$,

$$\mathrm{C\tilde{r}}\{\xi \in B\} = \frac{1}{2}\left(\sup_{x \in B} \mu(x) + 1 - \sup_{x \in B^c} \mu(x)\right), \quad B \subset \mathbb{R} \tag{3}$$

As an example, if set B is degenerated into a point x, then

$$\mathrm{C\tilde{r}}\{\xi = x\} = \frac{1}{2}\left(\mu(x) + 1 - \sup_{y \neq x} \mu(y)\right), \quad \forall x \in \mathbb{R} \tag{4}$$

Definition 4: (Liu, 2004, 2006) The credibility distribution $\Phi : \mathbb{R} \to [0,1]$ of a fuzzy variable ξ on $(\Theta, \mathfrak{P}(\Theta), \mathrm{C\tilde{r}})$ is:

$$\Phi(x) = \mathrm{C\tilde{r}}\{\theta \in \Theta \,|\, \xi(\theta) \leq x\} \tag{5}$$

That is the credibility distribution $\Phi(x)$ is the accumulated credibility grade that the fuzzy variable ξ takes a value less than or equal to a real-number $x \in \mathbb{R}$. Generally speaking, the credibility distribution Φ is neither left-continuous nor right-continuous.

Theorem 5: Let ξ be a fuzzy variable on $(\Theta, \mathfrak{P}(\Theta), \mathrm{C\tilde{r}})$ with membership function μ. Then its credibility distribution,

$$\Phi(x) = \frac{1}{2}\left(\sup_{y \leq x} \mu(y) + 1 - \sup_{y > x} \mu(y)\right), \ \forall x \in \Re \tag{6}$$

Definition 6: Let Φ be the credibility distribution of the fuzzy variable ξ. Then function $\phi : \mathbb{R} \to [0, +\infty)$ of a fuzzy variable ξ is called a credibility density function such that,

$$\Phi(x) = \int_{-\infty}^{x} \phi(y)\,\mathrm{d}y, \quad \forall x \in \mathbb{R} \tag{7}$$

We now consider the credibility of an event A after it has been learned that some other event B has occurred. This new credibility of A is called the conditional credibility of the event A given B, denoted by $\mathrm{C\tilde{r}}(A \,|\, B)$.

Definition 7: Let $(\Theta, 2^{\Theta}, \mathrm{C\tilde{r}})$ be a credibility measure space and let $\forall A, B \in 2^{\Theta}$. Then the conditional credibility measure of A given B, $\mathrm{C\tilde{r}}(A \,|\, B)$,

$$\mathrm{C\tilde{r}}(A \mid B) = \frac{1}{2}\left(\frac{(2\mathrm{C\tilde{r}}(A \cap B)) \wedge 1}{(2\mathrm{C\tilde{r}}(B)) \wedge 1} + 1 - \frac{(2\mathrm{C\tilde{r}}(A^c \cap B)) \wedge 1}{(2\mathrm{C\tilde{r}}(B)) \wedge 1} \right) \tag{8}$$

Provided $\mathrm{C\tilde{r}}(B) > 0$.

3.14.3 Kernel Estimation of 1-Dimensional Credibility Distribution

Finding an estimated credibility distribution based on observed data from a fuzzy variable is very critical task in practices because a serious researcher has to defend an important principle, objectiveness. In other words, we should establish the credibility distribution in terms of data information collected objectively from the fuzzy variable itself. We will explore a nonparametric approach, kernel estimation under maximum entropy principle.

3.14.3.1 Kernel Estimation in Probability Theory

Definition 8: A kernel is a function $K(x) = c\kappa\left(\|x\|^2\right)$ mapping from $\mathbb{R}^d$ to $[0,\infty)$, where $\kappa(\cdot)$ is a piecewise nonnegative monotone decreasing function such that $\int_0^\infty \kappa(r)dr < \infty$ and c is a constant.

Two common kernels used in statistics theory are Gaussian kernel, $\left(1/\sqrt{2\pi}\right)\exp(-\|x\|^2/2)$, and Epanechikov kernel, $(3/4)\left(1-\|x\|^2\right)$. For finite support, both kernel reduced to $K(x/h) = 0$, if $\|x\| > h$, where parameter $h > 0$ is called the bandwidth for the kernel function K. For bivariate case, the Epanechikov product kernel takes the form,

$$K(x,y;h) = \begin{cases} \frac{9}{16}\left(1-\left(\frac{x}{h}\right)^2\right)\left(1-\left(\frac{y}{h}\right)^2\right) & \text{if } |x| < h,\ |y| < h \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

And the Epanechikov radial kernel takes the form:

$$K(x,y;h) = \frac{3}{4}\left(1 - \frac{x^2+y^2}{h^2}\right) \tag{10}$$

where x and y are coordinates.

For data assimilation purpose, for a one-dimensional data sample $\{x_1, x_2, \cdots, x_n\}$ the credibility kernel density takes a form:

$$f_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right) \tag{11}$$

While for the two-dimensional case, data sample $\{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$, the credibility kernel density takes the form:

$$f_h(x, y) = \frac{1}{nh}\sum_{i=1}^{n} K(x, y; h) \tag{12}$$

It is obvious that in order to have the highest degree of data assimilation an optimal bandwidth $h > 0$ has to be selected. Different from inference in the probabilistic sense, the inference on the ground of credibility theory must develop its own criterion. One of the criteria is the maximum entropy principle.

3.14.3.2 Maximum Entropy Principle

Entropy is a measure of uncertainty. The entropy of De Luca and Termini (1972) characterizes uncertainty resulting primarily from the linguistic vagueness rather than resulting from information deficiency, and vanishes when the fuzzy variable takes all the values with membership degree 1. However, we hope that the degree of uncertainty is 0 when the fuzzy variable degenerates to a crisp number, and is maximum when the fuzzy variable is an equipossible one, i.e., all values have the same possibility. In order to meet such a requirement, Li and Liu (2005) provided a new definition based on credibility measure.

Definition 9: (Fuzzy Entropy) Let ξ be a continuous fuzzy variable defined on credibility space $(\Theta, 2^{\Theta}, \mathrm{Cr})$, then,

$$H[\xi] = \int_{-\infty}^{\infty} S\left(\mathrm{Cr}\left(\{\theta : \xi(\theta) = u\}\right)\right) du \tag{13}$$

where

$$S(t) = -t \ln t - (1-t)\ln(1-t) \tag{14}$$

Similar to the probabilistic counterpart, given some constraints, there are possibly infinitely many members in the family of membership functions. The maximum entropy principle provides a guideline that it is possible to select the membership function that maximizes the value of entropy and satisfies the given constraints. The selection of membership function is in nature to maximize the entropy function by searching the appropriate parameter-values.

In mathematical language:

$$\begin{aligned} &\max_{\Pi} H[\xi] \\ &\text{s.t.} \\ &\zeta(\Pi)=\zeta_0 \end{aligned} \tag{15}$$

where Π is the parameter space and the parameter constraint is $\zeta(\Pi)=\zeta_0$.

3.14.3.3 Maximum Entropy Kernel Estimation of 1-Dimensional Credibility Distribution

Therefore, the arguments will follow the route, which starts from credibility kernel density to credibility distribution, then to the induced membership function and finally reaches the fuzzy entropy.

For one-dimensional case, given the sample $\{x_1, x_2, \cdots, x_n\}$, the credibility kernel density is $f_h(x)=(1/nh)\sum_{i=1}^{n} K\left((x_i-x)/h\right)$, then the credibility distribution takes the form:

$$\hat{\Phi}_X(x)=\frac{1}{nh}\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{x_i-u}{h}\right)du \tag{16}$$

Thus the membership function can be determined by,

$$\hat{\mu}_X(x)=\begin{cases} \dfrac{2}{nh}\displaystyle\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{x_i-u}{h}\right)du & \text{if } \hat{\Phi}_X(x)<0.5 \\ 1 & \text{if } \lim\limits_{y\uparrow x}\hat{\Phi}_X(y)<0.5\le\hat{\Phi}_X(x) \\ 2-\dfrac{2}{nh}\displaystyle\sum_{i=1}^{n}\int_{-\infty}^{x} K\left(\frac{x_i-u}{h}\right)du & \text{if } 0.5\le\lim\limits_{y\uparrow x}\hat{\Phi}_X(y) \end{cases} \tag{17}$$

Notice that,

$$\mathrm{Cr}\left(\{\theta:\xi(\theta)=x\}\right)=\frac{1}{2}\left(\mu(x)+1-\sup_{y<x}\mu(y)\right) \tag{18}$$

Then the fuzzy entropy will be bandwidth-dependent,

$$H[\xi;h]=\int_{-\infty}^{\infty} S_h\left(\mathrm{Cr}\left(\{\theta:\xi(\theta)=u\}\right)\right)du \tag{19}$$

However, we notice that the setting of constraints is extracting from data evidences and will be difficult for most of the engineering circumstances. Therefore, we suggest an empirical object function for parameter searching since the optimal value of the data-dependent object function has to reflect the constraints specified by system performance data implicitly. The data constrained object function is the average of entropies evaluated at $\{x_1,x_2,\cdots,x_n\}$, i.e.,

$$J\left(h\,|\,x_1,x_2,\cdots,x_n\right)=\frac{1}{n}\sum_{i=1}^{n}H\left[\xi=x_i;h\right] \tag{20}$$

Notice that as $n\to\infty$, $J_h\to H[\xi;h]$ asymptotically with parameter constrained by data structure. Finally, the seeking optimal bandwidth value is just a problem of:

$$\max_{h>0} J\left(h\,|\,x_1,x_2,\cdots,x_n\right) \tag{21}$$

Or alternatively, search the solution of the equation, denoted by $\hat{h}$:

$$\frac{dJ\left(h\,|\,x_1,x_2,\cdots,x_N\right)}{dh}=0 \tag{22}$$

Once the optimal bandwidth, $\hat{h}$, is obtained, the maximum entropy kernel credibility distribution based on sample data $\{x_1,x_2,\cdots,x_N\}$ will be obtained, denoted by, $\hat{\Phi}_{X,h}(x)$. Accordingly, the values of the kernel credibility distribution at the points $\{x_1,x_2,\cdots,x_N\}$ will be denoted by $\{\hat{u}_1,\hat{u}_2,\cdots,\hat{u}_n\}$, where $\hat{u}_i=\Phi_{X,h}(x_i)$, $i=1,2,\cdots,n$. Similarly, the kernel credibility distribution $\hat{\Phi}_{Y,h}(y)$ can be obtained and values of the kernel credibility distribution at the points $\{y_1,y_2,\cdots,y_n\}$ will be denoted by $\{\hat{v}_1,\hat{v}_2,\cdots,\hat{v}_n\}$, where $\hat{v}_i=\hat{\Phi}_{Y,h}(y_i)$, $i=1,2,\cdots,n$.

3.14.4 Kernel Estimation of Bivariate Credibility-Copula

Bivariate credibility-copula is in nature a bivariate joint uniform distribution. However, the copula representation has an advantage of its simple form, particularly, the one-parameter copula family. This feature may make the data assimilation process easier than that of the kernel estimation of the credibility distribution directly.

3.14.4.1 Bivariate Credibility-Copula and One-Parameter Copula Family

Similar to probabilistic copula theory, copula is a dependence index measuring a couple of two fuzzy variables. Let X and Y be two continuous fuzzy variable with credibility distributions Φ_X and Φ_Y respectively and $H_{XY}(\cdot,\cdot)$ be their joint credibility distribution. Let $\mathbb{I}=[0,1]$.

Definition 10: A bivariate credibility-copula is a function $C:\mathbb{I}\times\mathbb{I}\to\mathbb{I}$ such that:

(i) $C(0,x)=C(x,0)=0$ and $C(1,x)=C(x,1)=x$ for all $x\in\mathbf{I}$;

(ii) $C(\cdot,\cdot)$ is 2-*increasing*: for $\forall a,b,c,d\in\mathbf{I}$, $a<b$, and $c<d$,

$$\nu_C([a,b]\times[c,d])=C(b,d)-C(a,d)-C(b,c)+C(a,c)\geq 0 \tag{23}$$

The function ν_C is called the C-volume of the rectangle $[a,b]\times[c,d]$.

In other words, a credibility-copula is the restriction to the unit square $\mathbf{I}\times\mathbf{I}=[0,1]\times[0,1]$ of a bivariate credibility distribution function whose margins are standard uniform. More formally, a copula C induces a credibility measure on $\mathbf{I}\times\mathbf{I}$ in terms of ν_C, i.e.,

$$\nu_C([0,u]\times[0,v])=C(u,v) \tag{24}$$

Now let us state the credibility theoretical version of Sklar's theorem (Nelson, 1999).

Theorem 11: Let H be a bivariate-credibility distribution function with marginal credibility distributions Φ_X and Φ_Y respectively. Then there exists a credibility-copula C such that $H(x,y)=C(\Phi_X(x),\Phi_Y(y))$. Conversely, for any credibility distribution functions Φ_X and Φ_Y and any credibility-copula C, the function H defined above is a two-dimensional credibility distribution with margins Φ_X and Φ_Y respectively. Furthermore, if Φ_X and Φ_Y are continuous, the credibility-copula is unique.

Based on the credibility theoretical version of Sklar's theorem, given a two-dimensional credibility distribution $H(x,y)$ and marginal credibility distributions Φ_X and Φ_Y, the credibility-copula is,

$$C(u,v)=H\left(\Phi_X^{-1}(u),\Phi_Y^{-1}(v)\right) \tag{25}$$

Where the inverse of the credibility distribution denoted by $F^{-1}(\cdot)$ is defined as:

$$F^{-1}(u)=\sup\left\{x \mid F(x)\le u\right\} \tag{26}$$

By noticing the credibility distribution transformation, $U=\Phi_X(X)$ and $V=\Phi_Y(Y)$ are fuzzy standard uniform-distributed variables on $[0,1]$, the credibility-copula is the bivariate joint uniform credibility distribution of two uniform-distributed fuzzy variable on $[0,1]\times[0,1]$. Therefore, credibility-copula fully describes the functional relation between two fuzzy variables.

Paralleling to probabilistic copula theory, the Archimedean credibility-copula is a function C: $\mathbb{I}\times\mathbb{I}\to\mathbb{I}$ defined by the generator ϕ such that,

$$C(u,v)=\phi^{[-1]}\left(\phi(u)+\phi(v)\right) \tag{27}$$

where $\phi^{[-1]}$ is the pseudo-inverse of ϕ:

$$\phi^{[-1]}(t)=\begin{cases}\phi^{-1}(t) & t\in[0,\phi(0)]\\ 0 & t>\phi(0)\end{cases} \tag{28}$$

The generator $\phi:\mathbb{I}\to[0,\infty]$, which is a strictly monotone decreasing and continuous convex function. An important property of Archimedean copula family is the associativity: i.e., $C\left(C(u,v),w\right)=C\left(u,C(v,w)\right)$. A simple example of Archimedean copula is the generator takes the form $\ln\left(\left(1-\theta(1-t)\right)/t\right)$ and the generating copula is $C(u,v)=uv/\left(1-\alpha(1-u)(1-v)\right)$. One-parameter bivariate copula is not necessarily an Archimedean copula. A typical example is $C(u,v)=uv+\alpha uv(1-u)(1-v),\ |\alpha|\le 1$.

3.14.4.2 Entropy Kernel Estimation of One-Parameter Bivariate Credibility-Copula

In terms of the previous arguments, for bivariate sample data $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$, the entropy kernel estimated bivariate uniform values are $\{(\hat{u}_1,\hat{v}_1),(\hat{u}_2,\hat{v}_2),\cdots,(\hat{u}_n,\hat{v}_n)\}$.

Recall that a copula, $C(u,v)$, in nature a bivariate joint credibility uniform distribution, possesses a joint credibility density, denoted by $\partial^2 C(u,v)/\partial u\partial v$, we say $C(u,v)$ is absolutely continuous. Given a two-dimensional credibility distribution $H(x,y)$ and marginal credibility distributions Φ_X and Φ_Y, then its associated copula is $C(u,v)$. Conversely, given a copula $C(u,v)$ and marginal credibility distributions Φ_X and Φ_Y, then the bivariate joint distribution $H(x,y)$ can be found.

The joint bivariate membership function is defined by,

$$\mu(x,y)=\left(2\mathrm{Cr}\left((X,Y)=(x,y)\right)\right)\wedge 1 \tag{29}$$

In terms of the equality: $H(x,y)=C\left(\Phi_X(x),\Phi_Y(y)\right)$, we have:

$$\mu(x,y)=\begin{cases} C\left(\Phi_X(x),\Phi_Y(y)\right) & \text{if } C\left(\Phi_X(x),\Phi_Y(y)\right)<0.5 \\ 1 & \text{if } \lim\limits_{(s,t)\uparrow(x,y)} C\left(\Phi_X(s),\Phi_Y(t)\right)<0.5\le C\left(\Phi_X(x),\Phi_Y(y)\right) \\ 2-C\left(\Phi_X(x),\Phi_Y(y)\right) & \text{if } 0.5\le \lim\limits_{(s,t)\uparrow(x,y)} C\left(\Phi_X(s),\Phi_Y(t)\right) \end{cases} \tag{30}$$

Also,

$$\mathrm{Cr}\left(\{\theta:(X,Y)(\theta)=(x,y)\}\right)=\frac{1}{2}\left(\mu(x,y)+1-\sup_{(s,t)\uparrow(x,y)}\mu(s,t)\right) \tag{31}$$

Accordingly, the entropy takes the form:

$$H[(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S\left(\mathrm{Cr}\left(\{\theta:(X,Y)(\theta)=(s,t)\}\right)\right)dsdt \tag{32}$$

where

$$S(t)=-t\ln t-(1-t)\ln(1-t) \tag{33}$$

Recall that the kernel estimated bivariate joint uniform distribution has estimated sample data $\{(\hat{u}_1,\hat{v}_1),(\hat{u}_2,\hat{v}_2),\cdots,(\hat{u}_n,\hat{v}_n)\}$, thus $\{C(\hat{u}_1,\hat{v}_1),C(\hat{u}_2,\hat{v}_2),\cdots,C(\hat{u}_n,\hat{v}_n)\}$ are the observed copula values. Accordingly, the observed entropy can be calculated in the form,

$$J(\alpha \mid C(\hat{u}_i,\hat{v}_i), i=1,2,\cdots,n) = \frac{1}{n}\sum_{i=1}^{n} H[(X,Y)=(x_i,y_i);\alpha] \tag{34}$$

Solving equation $dJ/d\alpha = 0$ will give the maximum entropy estimation of the copula $C(u,v;\alpha)$.

3.14.5 Conclusion

In this paper, we heavily reviewed the related work on $(\vee,\cdot)$-credibility measure theory. Based on the review, we propose the concept of copula for bivariate fuzzy variables, called as credibility-copula, which is a bivariate uniform credibility distribution in the sense of the credibility measure theory. Parallel to the basic understanding in probabilistic copula, we regard the credibility-copula to be a measure of the full bivariate dependence. Furthermore, we establish a two-stage procedure for credibility-copula based on data-assimilation under the fuzzy maximum entropy principle.

3.15 Credibility Measure-Based Fuzzy Membership Grade Kriging

Danni Guo, Renkuan Guo, and Christien Thiart

Department of Statistical Sciences, University of Cape Town
Private Bag, Rhodes' Gift, Rondebosch 7701, Cape Town, South Africa
dannig@stats.uct.ac.za rguo@stats.uct.ac.za thiart@stats.uct.ac.za

This paper is currently under-print and will be published in the International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2007, vol. 15. Editors: B.D. Liu.

3.15.1 Introduction

Today, we often face the problem of fuzzy uncertainty when collecting environmental data and performing analysis. Accordingly, methodology to reveal the fuzzy uncertainty in the environmental data should be developed, in terms of the basic concept of fuzzy mathematics. However, whenever we use fuzzy mathematics for practical modelling, we will face a sequence of fundamental issues:

1. The first one is the self-duality in its theoretical foundation. Fuzzy mathematics initiated by Zadeh (1965) facilitated a foundation dealing with vague phenomena in fuzzy modelling. However, the fuzzy mathematical foundation initiated by Zadeh (1965, 1978) is membership function and possibility measure based and widely used. The possibility measure was originally expected to play the role of probability measure in probability theory, but could not because it does not possess self-duality property as that in probability theory.
2. The second issue is the variable-orientation issue. In standard probability theory, random variable and the distribution function play important roles for converting set-based arguments into variable-based arguments, which result in great conveniences in applications. Kaufmann (1975) first proposed the concept of fuzzy variable with the intention of creating its counterpart in probability theory. Unfortunately, Kaufmann's fuzzy variable is in fact another name for a fuzzy subset, and the mathematical operations are difficult to handle.
3. The third one is the membership specification issue. During almost four decades, fuzzy researchers had to specify membership function and set up the parameter values in terms of their own working experiences. Compared to the probabilistic counterpart, for random variable and its distribution, very rich (data-oriented) statistical estimation and hypothesis testing theory have been developed. The fuzzy statistical theory developed very slowly, and its applications are difficult due to the set-oriented foundation.

To resolve the first three dilemmas, Liu (2004, 2006) proposed an axiomatic foundation for modelling fuzzy phenomena, credibility theory. The credibility measure possesses self-duality property and is able to play the role of that in probability theory. Furthermore, fuzzy variable concept and its distribution, which are parallel to these in probability theory, are developed. However, in conditional credibility measure based on Liu's standard credibility theory faces

fundamental difficulty, and he proposed a non-classical credibility theory for dealing with the conditioning related modelling.

In this paper, we first review the $(\vee,\cdot)$-credibility measure theory. Then we propose an empirical average sample entropy criterion for estimating the parameters associated with membership function in terms of maximum entropy principle to achieve the aim of full data-assimilated membership function determination. Finally we propose membership grade kriging as an approach to model fuzzy uncertainty and spatially distributed random uncertainty simultaneously. PM_{10} data (2003) in California, U.S.A. is used as an example to illustrate our methodology.

3.15.2 A Review of $(\vee,\cdot)$-Credibility Measure Theory

Let Θ be a nonempty set, and $\mathfrak{P}(\Theta)=2^{\Theta}$ the power set on Θ. Each element, let us say, $A\subset\Theta, A\in\mathfrak{P}(\Theta)$ is called an event. A number denoted as $\tilde{\mathrm{Cr}}(A)$, $0\le\tilde{\mathrm{Cr}}(A)\le 1$, is assigned to event $A\in\mathfrak{P}(\Theta)$, which indicates the credibility that event $A\in\mathfrak{P}(\Theta)$ occurs. $\tilde{\mathrm{Cr}}(A)$ satisfies following axioms (Liu, 2004, 2006):

Axiom 1: $\tilde{\mathrm{Cr}}(\Theta)=1$.

Axiom 2: $\tilde{\mathrm{Cr}}(\cdot)$ non-decreasing, i.e., $\tilde{\mathrm{Cr}}\{A\}\le\tilde{\mathrm{Cr}}\{B\}$ whenever $A\subset B$.

Axiom 3: $\tilde{\mathrm{Cr}}(\cdot)$ is self-dual, i.e., $\tilde{\mathrm{Cr}}\{A\}+\tilde{\mathrm{Cr}}\{A^c\}=1$ for any $A\in\mathfrak{P}(\Theta)$.

Axiom 4: $\tilde{\mathrm{Cr}}\{\bigcup_i A_i\}\wedge 0.5=\sup_i\left[\tilde{\mathrm{Cr}}\{A_i\}\right]$ for any $\{A_i\}$ with $\tilde{\mathrm{Cr}}(A_i)\le 0.5$.

Axiom 5: Let set functions $\tilde{\mathrm{Cr}}_k(\cdot):2^{\Theta_k}\to[0,1]$ satisfy **Axiom 1-4**, and $\Theta=\Theta_1\times\Theta_2\times\cdots\times\Theta_p$, then:

$$\tilde{\mathrm{Cr}}(\theta_1,\theta_2,\cdots,\theta_p)=\begin{cases}\dfrac{1}{2}\displaystyle\prod_{k=1}^{p}\left(2\tilde{\mathrm{Cr}}(\theta_k)\wedge 1\right) & \text{if } \min\limits_{1\le k\le p}\left\{\tilde{\mathrm{Cr}}(\theta_k)\right\}<0.5\\ \min\limits_{1\le k\le p}\left\{\tilde{\mathrm{Cr}}(\theta_k)\right\} & \text{if } \min\limits_{1\le k\le p}\left\{\tilde{\mathrm{Cr}}(\theta_k)\right\}\ge 0.5\end{cases} \tag{1}$$

for each $(\theta_1,\theta_2,\cdots,\theta_p)\in\Theta$. In this case, we write $\tilde{\mathrm{Cr}}=\tilde{\mathrm{Cr}}_1\times\tilde{\mathrm{Cr}}_2\times\cdots\times\tilde{\mathrm{Cr}}_p$.

Definition 1: Any set function $C\tilde{r}:\mathfrak{P}(\Theta)\to[0,1]$ satisfies **Axiom 1-5** is called a $(\vee,\cdot)$-credibility measure. The triple $(\Theta,2^{\Theta},C\tilde{r})$ is called the $(\vee,\cdot)$-credibility measure space.

Definition 2: The (induced) membership function of a fuzzy variable ξ on $(\Theta,\mathfrak{P}(\Theta),C\tilde{r})$ is:

$$\mu(x)=\left(2C\tilde{r}\{\xi=x\}\right)\wedge 1,\quad x\in\mathbb{R} \tag{2}$$

Conversely, for given membership function the credibility measure is determined by the credibility inversion theorem.

Definition 3: (Liu, 2004, 2006) The credibility distribution $\Phi:\mathbb{R}\to[0,1]$ of a fuzzy variable ξ on $(\Theta,\mathfrak{P}(\Theta),C\tilde{r})$ is:

$$\Phi(x)=C\tilde{r}\left\{\theta\in\Theta\middle|\xi(\theta)\le x\right\} \tag{3}$$

That is the credibility distribution $\Phi(x)$ is the accumulated credibility grade that the fuzzy variable ξ takes a value less than or equal to a real-number $x\in\mathbb{R}$. Generally speaking, the credibility distribution Φ is neither left-continuous nor right-continuous.

Theorem 4: Let ξ be a fuzzy variable on $(\Theta,\mathfrak{P}(\Theta),C\tilde{r})$ with membership function μ. Then its credibility distribution,

$$\Phi(x)=\frac{1}{2}\left(\sup_{y\le x}\mu(y)+1-\sup_{y>x}\mu(y)\right),\ \forall x\in\Re \tag{4}$$

Definition 5: Let Φ be the credibility distribution of the fuzzy variable ξ. Then function $\phi:\mathbb{R}\to[0,+\infty)$ of a fuzzy variable ξ is called a credibility density function such that,

$$\Phi(x)=\int_{-\infty}^{x}\phi(y)\,dy,\quad \forall x\in\mathbb{R} \tag{5}$$

We now consider the credibility of an event A after it has been learned that some other event B has occurred. This new credibility of A is called the conditional credibility of the event A given B, denoted by $C\tilde{r}(A\mid B)$.

Definition 6: Let $(\Theta,2^{\Theta},C\tilde{r})$ be a credibility measure space and let $\forall A,B\in 2^{\Theta}$. Then the conditional credibility measure of A given B, $C\tilde{r}(A\mid B)$,

$$\mathrm{C\tilde{r}}(A \mid B) = \frac{1}{2}\left[\frac{(2\mathrm{C\tilde{r}}(A \cap B)) \wedge 1}{(2\mathrm{C\tilde{r}}(B)) \wedge 1} + 1 - \frac{(2\mathrm{C\tilde{r}}(A^c \cap B)) \wedge 1}{(2\mathrm{C\tilde{r}}(B)) \wedge 1}\right] \tag{6}$$

Provided $\mathrm{C\tilde{r}}(B) > 0$.

3.15.3 Maximum Entropy Data-assimilated Membership Function

A fundamental and thorny issue in fuzzy mathematical theoretical developments and applications is the determination of function form of fuzzy membership function and associated parameters. Different from its probabilistic counterpart, the data-assimilated parameter specification is very difficult to carry on. We use term estimation to refer to the parameter determination based sampled data in terms of statistical theory, and we use the term data-assimilation to refer to the membership function parameter specification based on observations in fuzzy mathematics.

3.15.3.1 Fuzzy Entropy

Entropy is a measure of uncertainty. The entropy of De Luca and Termini (1972) characterizes uncertainty resulting primarily from the linguistic vagueness rather than resulting from information deficiency, and vanishes when the fuzzy variable takes all the values with membership degree 1. However, we hope that the degree of uncertainty is 0 when the fuzzy variable degenerates to a crisp number, and is maximum when the fuzzy variable is an equipossible one, i.e., all values have the same possibility. In order to meet such a requirement, Li and Liu (2005) provided a new definition based on credibility measure.

Definition 7: (Fuzzy Entropy) Let ξ be a continuous fuzzy variable defined on credibility space $(\Theta, 2^{\Theta}, \mathrm{C\tilde{r}})$, then the fuzzy entropy, $H[\xi]$, is defined by:

$$H[\xi] = \int_{-\infty}^{\infty} S\left(\mathrm{Cr}\left(\{\theta : \xi(\theta) = u\}\right)\right) du \tag{7}$$

where

$$S(t) = -t \ln t - (1-t)\ln(1-t) \tag{8}$$

For convenience, we name $S(t)$ as entropy density at point t.

3.15.3.2 Maximum Entropy Principle

Similar to the probabilistic counterpart, given some constraints, there are possibly infinitely many members in the family of membership functions. The maximum entropy principle provides a guideline that it is possible to select the membership function that maximizes the value of entropy and satisfies the given constraints. The selection of membership function is in nature to maximize the entropy function by searching the appropriate parameter-values.

In mathematical language:

$$\begin{aligned} &\max_{\Pi} H[\xi] \\ &\text{s.t.} \\ &\zeta(\Pi)=\zeta_0 \end{aligned} \tag{9}$$

where Π is the parameter space and the parameter constraint is $\zeta(\Pi)=\zeta_0$.

3.15.3.3 Entropy Estimation

For example, if we take the hyperbolic tangent membership function proposed by Guo et al. (2003, 2006c), which takes a location-scale parametric form:

$$\mu(z)=\frac{1}{2}\left(\tanh\left(\frac{z-\delta}{\eta}\right)+1\right) \tag{10}$$

where parameter δ is the location and η is the scale parameter.

Recall that the entropy function is:

$$H[Z]=\int_{-\infty}^{\infty} S\left(\mathrm{Cr}\left(\{\theta : Z(\theta)=z\}\right)\right)dz \tag{11}$$

and that the credibility measure for event $\{\theta : Z(\theta)=z\}$ takes the form

$$\mathrm{C\tilde{r}}\left(\{\theta : Z(\theta)=z\}\right)=\frac{1}{2}\left(\mu(z)+1-\sup_{y\neq z}\mu(y)\right) \tag{12}$$

It is obvious that for the hyperbolic tangent membership function $\mu(\cdot)$ is monotone increasing and thus $\sup_{y\neq z}\mu(y)=1$ holds for any z value. Therefore we have the credibility measure for event $\{\theta : Z(\theta)=z\}$:

$$\mathrm{C\tilde{r}}\left(\left\{\theta:\xi(\theta)=z\right\}\right)=\frac{1}{2}\mu(z)=\frac{1}{4}\left(\tanh\left(\frac{z-\delta}{\eta}\right)+1\right) \tag{13}$$

The maximization of entropy function is reached by searching for (δ,η), which is the solution to the nonlinear equation system defined by the theoretical fuzzy entropy,

$$\begin{cases}\dfrac{\partial H[Z]}{\partial\delta}=0\\ \dfrac{\partial H[Z]}{\partial\eta}=0\end{cases} \tag{14}$$

with appropriate constraints.

However, what we aim at is not obtaining parameters from the theoretical entropy expression rather we must determine the parameters based on observations of the fuzzy variable *Z*. In other words, we need to develop a criterion to obtain data-assimilated membership function. Therefore, we suggest an empirical object function for parameter searching since the optimal value of the data-dependent object function has to reflect the constraints specified by observational data implicitly. The data assimilated object function is the average of entropy densities evaluated at $\{z_1,z_2,\cdots,z_n\}$ respectively, i.e.,

$$J[-L_1,L_2]=J\left((\delta,\eta)\,|\,\{z_1,z_2,\cdots,z_n\}\subseteq[-L_1,L_2]\right)=\frac{1}{n}\sum_{i=1}^{N}S\left(\mathrm{C\tilde{r}}\left\{Z(\theta)=z_i;(\delta,\eta)\right\}\right) \tag{15}$$

where a finite interval $[-L_1,L_2]$, $L_2>L_1\geq 0$ is defined for the domain of entropy. Note that with the finite entropy as $n\to\infty$, $J[-L_1,L_2]\to H[Z;(\delta,\eta)]$ asymptotically with parameter constrained by the data structure and $Z\in[-L_1,L_2]$, $L_2>L_1\geq 0$, which guarantees the theoretical entropy $H[Z]$ exists and finite in general.

Therefore, in environmental modelling circumstances, we will work with the problem of maximizing object function $J\left((\delta,\eta)\,|\,z_1,z_2,\cdots,z_n\right)$ subject to constraint imposed to membership function $\mu(\cdot)$. For example, for PM_{10}, at the pre-set threshold value of PM_{10}, $z_0=46$ with membership grade $\mu(z_0)=0.995$, because the Environmental Protection Agency regards 50 as the danger level for residents.

Thus the constraint on parameter pair (δ, η) is:

$$\delta + 2.998\eta = 46 \tag{16}$$

Then we need to solve a constrained maximum entropy problem:

$$\begin{aligned} &\max_{\delta,\eta} \{J\} \\ &\text{s.t.} \\ &\delta + 2.998\eta = 46 \\ &28 \le \delta \le 34 \end{aligned} \tag{17}$$

We use Matlab optimization toolbox for the parameter searching and obtain the data-assimilated parameters $(\hat{\delta}, \hat{\eta})$ and thus the full data assimilated hyperbolic tangent membership function takes the form:

$$\hat{\mu}_Z(z) = \frac{1}{2}\left(\tanh\left(\frac{z-28}{6.004}\right)+1\right) \tag{18}$$

3.15.4 Ordinary Kriging on Membership Grades of PM_{10} Values

The Environmental Protection Agency (EPA) sets National Ambient Air Quality Standards for pollutants considered harmful to public health and the environment. PM_{10} is one of the seven air pollutants the EPA regulates, under the National Ambient Air Quality Standards. *Particulate Matter* (PM) is a type of air pollutants that include dust, dirt, soot, smoke, and liquid droplets, Environmental Protection Agency[1]. PM_{10} is particulate matter that is very small, and remains suspended in the air for long periods of time. They usually consist of smoke, dirt, and dust particles, as well as mold, spores, and pollen, Rocky Mountain Center[9]. PM_{10} is easily inhaled into the deep lung. Exposure to high outdoor PM_{10} concentrations causes increased disease and death.

In this example, we used the PM_{10} air pollution data of California, U.S.A., for applications. The current EPA standard for PM_{10} has set an annual allowable arithmetic average of PM_{10} not to exceed 50 $\mu g/m^3$ i.e. micrograms per cubic meter of air, Environmental Protection Agency[1]. The PM_{10} samples were collected in 103 locations in California (figure 3.15.1).

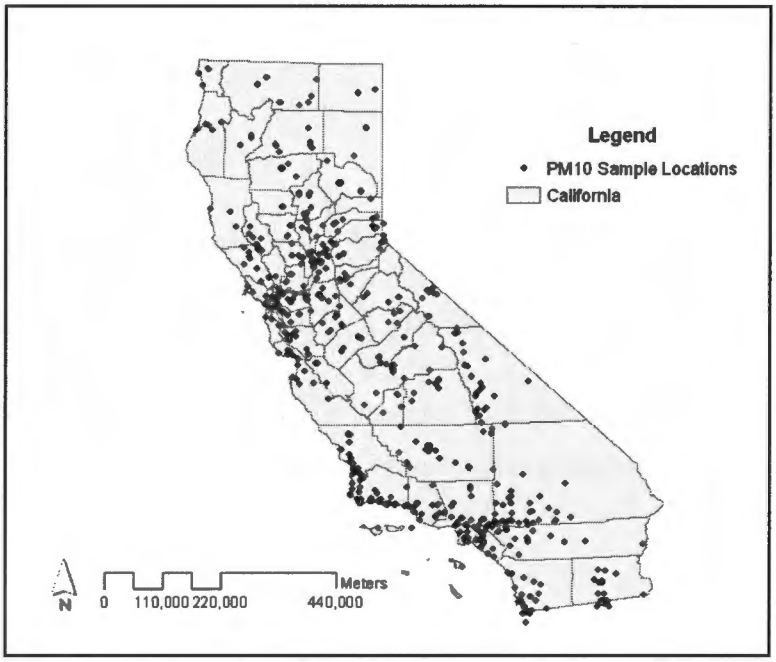


Figure 3.15.1 PM_{10} Sample Locations in California.

3.15.4.1 Membership Kriging

The fuzzy membership grade semivariogram (Guo et al., 2004) is given by:

$$\gamma_z(h;\{z_m\}) = \frac{1}{2}\mathbf{E}\left[\left(\mu(z(x+h)) - \mu(z(x))\right)^2\right] \tag{19}$$

where h represents the distance between two spatial locations x_i and $x_i + h$, and z is the threshold value, and the sample values $z(x_i)$ and $z(x_i + h)$ are observed at x_i and $x_i + h$ respectively. The experimental fuzzy membership grade semivariogram is used in fuzzy membership grade kriging and is defined by:

$$\hat{\gamma}_z(h;\{z_m\}) = \frac{1}{2N_h}\sum_{i=1}^{N_h}\left[\mu(z(x+h)) - \mu(z(x))\right]^2 \tag{20}$$

where N_h is the neighbourhood total sample number of diameter h.

Equation 21 represents the fuzzy membership grade kriging equation as an ordinary kriging on membership grades:

$$\begin{cases} -\sum_{j=1}^{n} \lambda_j \gamma_z\left(x_i - x_j;\{z_m\}\right) + \gamma_z\left(x_0 - x_i;\{z_m\}\right) - m = 0, i = 1,\ldots,n \\ \sum_{i=1}^{n} \lambda_i = 1 \end{cases} \tag{21}$$

where m is the Lagrange multiplier. The fuzzy membership grade predictor is then defined as:

$$p\left(\mu\left(z(x_i)\right);B\right) = \sum_{i=1}^{n} \lambda_i \mu\left(z(x_i)\right), \quad \sum_{i=1}^{n} \lambda_i = 1 \tag{22}$$

where B is the neighbourhood surrounding the location x. Accordingly, the intrinsic hypothesis for $p(\mu;B)$ to be an unbiased estimator should be:

$$\mu\left(Z(x)\right) = \mu_0 + \varepsilon_0(x), \quad x \in D, \mu \in \mathbb{R} \tag{23}$$

3.15.4.2 PM_{10} Air Pollution Example

The membership function being considered for fuzzy event $\tilde{A}$={High hazard level of PM_{10} in California} is assumed as hyperbolic tangent data-assimilated membership function. It offers both the curvature catching capability of the evolution from stratum to stratum, and smoothness over the whole range of PM_{10} value.

$$\hat{\mu}_Z(z) = \frac{1}{2}\left(\tanh\left(\frac{z-28}{6.004}\right)+1\right) \tag{24}$$

The two unknown parameters (δ,η) specify the location and the scale of the membership function. We use maximum entropy principle for obtain a data-assimilated membership function with parameters $(\hat{\delta},\hat{\eta}) = (28.00, 6.004)$, the maximum entropy is 0.3577.

It is observed that there are 20 out of 103 sampled PM_{10} observations with the data-assimilated membership grades being larger than 0.953014, while there are 9 of 103 sampled PM_{10} observations with the data-assimilated membership grades being less than 0.007013. When PM_{10} concentration is 50 $\mu g/m^3$ (hazard level) or more, the membership grade value will be 1.0 (figure 3.15.2).

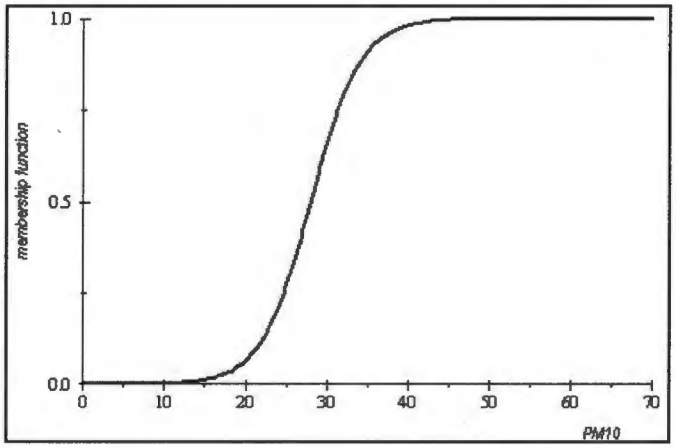


Figure 3.15.2 Hyperbolic Tangent Membership Function with Data-Assimilated Parameters $(\hat{\delta},\hat{\eta}) = (28.00, 6.004)$.

We perform the ordinary kriging on the 103 PM_{10} observations over California, which were collected in year 2003 by using ArcGIS. As an intuitive model goodness-of-fit measure, a QQ plot is generated (figure 3.15.3). The QQ plot shows the middle observational values are reasonably close to the dashed straight line although the extreme upper and lower values do not really fit the dash line well.

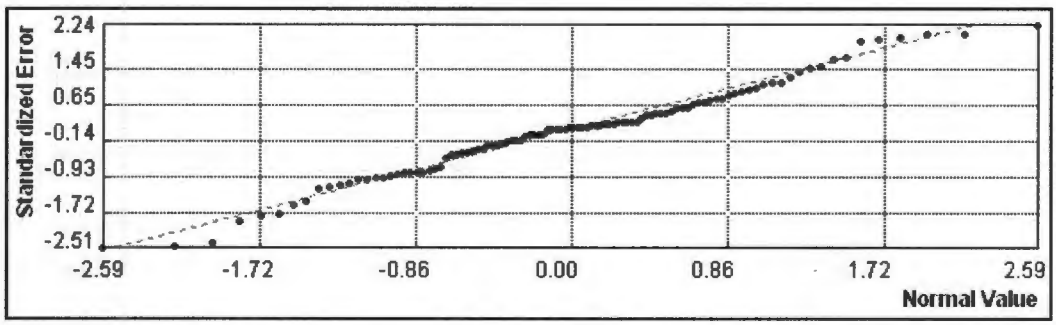


Figure 3.15.3 QQ Plot of Hyperbolic Tangent Membership Function.

In the membership grade kriging standard error map (figure 3.15.4), the class divisions are refined and detailed too, which is very consistent with PM_{10} observational site allocation pattern. The light yellow areas are areas of low error, and the dark brown areas are areas of high error. When one looks at the sample locations, one can understand the errors much better. Since most of the sample locations are near the coast, it is obvious that the errors are less in the sampled locations, and more errors in the unsampled locations.

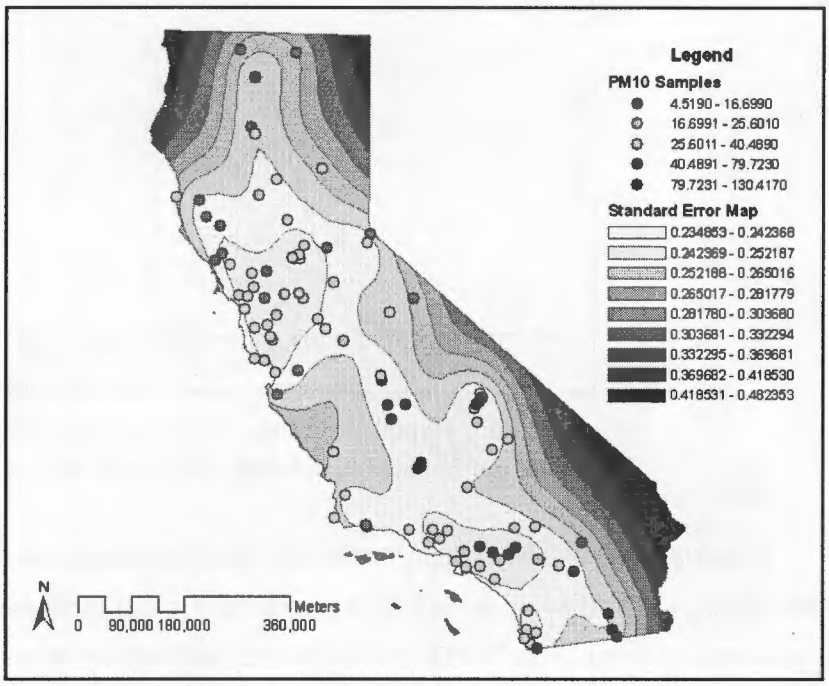


Figure 3.15.4 Standard Error Map Using the Hyperbolic Tangent Membership Function.

3.15.5 GIS Analysis of PM_{10} Concentrations and Comparisons

Having produced the fuzzy membership grade kriging results, the interpretation of the fuzzy membership grade prediction map is not an easy thing, particularly for those who do not have adequate background in fuzzy mathematics and statistics. Therefore, the predicted fuzzy membership grades are required to be converted back to PM_{10} concentrations to provide a more understandable format. The fuzzy membership grades are defined in terms of PM_{10} hazard levels. The membership functions are monotone-increasing continuous function, and so the inverse functions exist.

The hyperbolic tangent sample membership function takes the form:

$$\mu(x) = \frac{1}{2}\left[\tanh\left(\frac{x-\delta}{\eta}\right)+1\right], \quad a \le x \le b \tag{25}$$

For any given membership value $\mu_0 \in [0,1]$:

$$\mu_0 = \frac{1}{2}\left[\tanh\left(\frac{x-\delta}{\eta}\right)+1\right] \tag{26}$$

Then:

$$\frac{x-\delta}{\eta} = \frac{1}{2}\ln\left(\frac{1+(2\mu_0-1)}{1-(2\mu_0-1)}\right) = \frac{1}{2}\ln\left(\frac{2\mu_0}{2(1-\mu_0)}\right) = \frac{1}{2}\ln\left(\frac{\mu_0}{1-\mu_0}\right) \tag{27}$$

which leads to:

$$x^{\tanh}(\mu_0) = \frac{\eta}{2}\ln\left(\frac{\mu_0}{1-\mu_0}\right)+\delta \tag{28}$$

The class limits for PM_{10} predicted value with hyperbolic tangent sample membership function is:

$$[x_l, x_u) = \left[x^{\tanh}(\mu_l), x^{\tanh}(\mu_u)\right) \tag{29}$$

In terms of Equation 29, we convert the membership grade prediction map into PM_{10} concentration prediction map (figure 3.15.5).

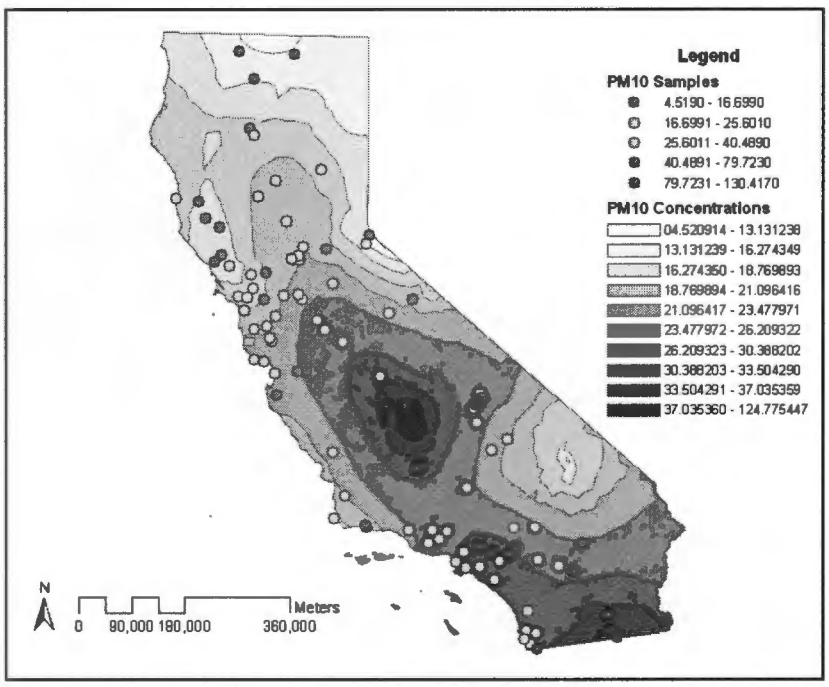


Figure 3.15.5 PM_{10} Concentrations Predicted by a Membership Grade Kriging in terms of a Fully Data-Assimilated Hyperbolic Tangent Membership Function.

In figure 3.15.5, the last three classes are high hazard areas since we believe once PM_{10} concentrations are over 30.35 $\mu g/m^3$, then it will pose a high hazard to human health. Areas of higher PM_{10} concentrations are in the middle regions of California and the bottom of California. The upper area of California and parts of the coastline are shown to have very low PM_{10} concentrations, and are the safest areas.

As a source of comparison, we have produced a kriging map of PM_{10} concentrations using the original unadjusted PM_{10} data. One can clearly see that the membership grade kriging predication map (figure 3.15.5) shows a finer division and contains hidden patterns that were not found in the normal kriging map (figure 3.15.6) .

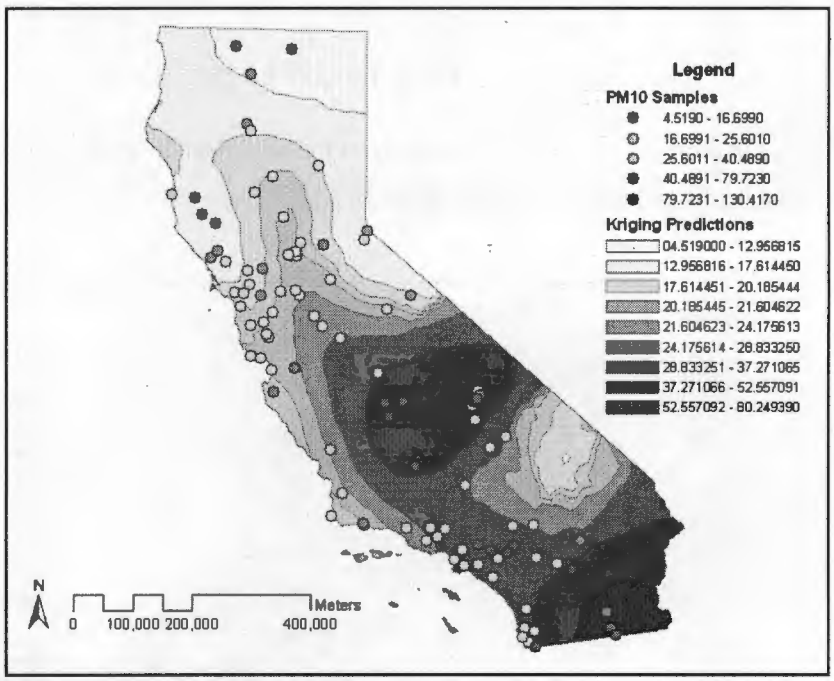


Figure 3.15.6 Ordinary Kriging Map Predicted Using Original Unadjusted PM_{10} Data.

3.15.6 Discussion

As a continuous efforts of the earlier developments by Guo et al. (2004, 2006c), in this paper, we do not only put the mathematical foundation in solid credibility measure theoretical framework (Liu, 2004, 2006) but also propose a full data-assimilation approach for obtaining membership function in terms of maximum entropy principle. In other words, the membership function form is

hypothesized but the parameters are determined in terms of observational data information. We emphasize again that different from statistical estimation based on sampled data, we use term data-assimilation for using observational information to specify a membership function.

We should notice that for the data-assimilation convenience, the functional form of membership function is a *priori*. Distributional priori is commonly accepted in statistics and therefore using a priori membership function form could be blamed as too subjective. As to the form of membership function should be continuous one and avoid piece-wise definition. It is obvious the hyperbolic tangent membership function with two parameters suits our purpose. Our computation experiences showed that the constraints are to be selected by the nature of the problem. However, the choice of initial values is very tricky and decided in a manner of trial and error.

The critical and practical importance of this paper lies on the development in full data-assimilated membership function determination. Therefore, fuzzy mathematics can totally walk out of the shadow of subjective image which is often attached by some researchers. Definitely, credibility measure theoretical foundation proposed by (Liu, 2004, 2006) offered an axiomatic mathematical foundation for fuzzy mathematics, which finally paves the way toward a rigorous fuzzy mathematical age.

3.15.7 Conclusion

In this paper, we used a credibility measure based fuzzy membership grade kriging approach for the spatial prediction of the PM_{10} concentrations in California. It is an exploratory data analysis. It is noticed that credibility measure theoretical foundation proposed by (Liu, 2004, 2006) put all the mathematical developments in a rigorous way, and therefore we have the chance to develop the maximum entropy data-assimilated membership function by assuming a hyperbolic tangent form. The membership function is no longer semi-statistical as Guo et al. proposed (2004, 2006c). In order to seek an optimal data-assimilated membership with solid mathematical foundation, it is necessary to extract membership function from the data itself. As we emphasized at the beginning of the membership grade kriging we aim at its convenience of application not the mathematical rigidity. A fully rigorous approach should use credibility-chance measure theoretical framework, proposed by (Liu, 2004, 2006) for dealing random fuzzy spatially distributed observational data.

Chapter 4. Discussion

4.1 Summary and Critical Assessment

This chapter summarised the developments and shortcomings of the fifteen thesis papers given in chapter 3.

1. In the paper, "Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM_{10} for California, Using Fuzzy Kriging", we explored the concept of using *Fuzzy Kriging* in GIS. This method is illustrated using PM_{10} air pollution data of California, U.S.A. We integrated GIS with fuzzy logic and geostatistics to predict PM_{10} concentrations for California from existing sample data. A methodology of fuzzy variogram and fuzzy kriging is proposed, based on the indicator vaiogram and kriging developments in the literature. Fuzzy kriging is developed as a natural extension of indicator kriging, and then ordinary kriging is performed on the membership results. This way fuzzy kriging would have the advantages of both indicator and ordinary kriging. Unlike previous studies that used assumed membership functions, in this paper, a sample membership function is extracted from the data itself. Therefore, the fuzzy methodology used in this paper is more solid and objective. However, the parameters for specifying membership are still not fully data-assimilated.
2. In the paper, "Grey GIS", we looked at the possibilities of *Grey Spatial Analysis*. We proposed a small-sample based spatial data analysis methodology in terms of Grey System Theory, and the new methodology can be implemented and combined with Geographic Information Systems (GIS), and we call it *Grey GIS*. Grey GIS is a system dealing with grey geographical data or analyzing spatial data in terms of Grey System Theory. Facing the challenge of data shortage, we developed a small-sample based spatial data analysis methodology based on Grey System Theory, called *Grey Spatial Analysis*. The newly established Grey Spatial Analysis methodology is particularly suitable for small area spatial predictions with a few observations. The Grey Spatial Analysis results are imprecise (grey numbers) in nature. The grey predictions prediction results have shown to reflect well the local reality, and consistent with local sample information available.
3. In the paper, "Grey Spatial Prediction of Fuzzy Air Quality Index", we explored the air quality index, and discussed the application of Grey Spatial Prediction. The main theme of this paper was to develop a small-sample based grey spatial prediction approach based on

order-less spatial observations. This methodology enable spatial analysis to be free from the shadow of traditional large-sample based semivariogram – kriging spatial prediction methodology. We have developed a small-sample based spatial data analysis methodology based on the spirit of grey differential equation theory. The key contribution is the conversion from order-less 2-dimensional spatial data into ordered data sequences. Furthermore, we follow the idea behind the grey system theory – modelling process is the one of enhancing the degree of information sufficiency. The way to add grey temporal prediction on three sites is just increasing information. It is clear that the grey spatial prediction results reflect well the local reality. However, the grey spatial prediction developments need to be improved. We will need to use differential geometry to further solidify the mathematical foundation. However, the applicability is limited because the fuzzy measure is not σ -additive while probability measure is in general.

4. In the paper, "A New Very Small Sample-Based Non-Linear Statistical Estimation Method", we looked at non-linear statistical estimation based on small samples. We developed a new approach to facilitate a statistical estimation of a scalar field function defined on an *n*-dimensional space with very small sample. The methodology combined concepts from differential geometry and data treatments from the least-square estimation at differential equation level. The estimated response function takes non-linear forms on the *n*-dimensional region where data information are available, and therefore it offers potential improvements over that of multivariate regression which is a super-plane on the same region. We call it a grey state model because the approach does not involve the underlying probability distribution, and directly seek the state law in terms of grey differential equation GM(1,1). The term grey indicates the estimated non-linear function with exponential functional as core part is imprecise.
5. In the paper, "Optimal Data Transformations in Grey Analysis", we explored the transformation of the discrete data sequence in Grey System Theory. The fact that the accuracy of GM(1,1) model is not automatically guaranteed causes concerns of the applicability in spatial modelling. One of the solutions is to explore the transformation of the discrete data sequence. The transformation, different from other data analysis methodology, must preserve trend and order, and at the same time, it is expected that the transformed data sequence can have a class ratio approaching 1. Deng's (2002a, 2002b) three transformations are reviewed and two new transformations are proposed. If Type II transformation can be

applied universally and obtain satisfactory GM(m,N) models at transformed data sequence $Y^{(0)}$ level, then it is would be a major contribution to grey data analysis, particularly, residual analysis. The new transformation development may expand the grey data analysis to any number system, say real number system and complex number system and even further into vector space, matrix space and tensor space in general. However, so far it is still a theoretical frame and whether or not a transformation is successful is in general not guaranteed although the solution sets for them are given.

6. In the paper, "Grey Reliability Analysis of Complex system", we investigated a way to obtain the dynamic version of grey system reliability with sparse data availability, in terms of grey differential equation models. We explored the grey characteristics of complex system reliability and investigated the evaluation approach in terms of the generalized load-strength concept, and the grey characterization of load-strength as two-fold whitenization event. A detailed theoretical framework is established for the mathematical treatments of the two-fold whitenization event and the static version of system grey reliability is obtained. We point the way to obtain the dynamic version of system grey reliability with sparse data availability in terms of grey differential equation models. As to how to extract the generalized load-strength functional and applications, it will be a future research topic.
7. In the paper, "GM(1,1)-Kriging Prediction of Soil Dioxin Pattern", we addressed the problem of ill-designed sample data. Today, we often face the circumstances where a set of data is collected already, although from the viewpoint of kriging analysis the data is insufficient, but re-sampling is impossible because of the cost and time limits. Therefore, a solution must be found to address ill-designed spatial data, and provides a better spatial analysis results. In this paper, we proposed a mixed approach by combining grey differential equation models, particularly, GM(1,1) model and ordinary kriging approach together and called GM(1,1)-kriging. We expanded the existing limited sample data available to produce GM(1,1)-kriging maps. This approach is illustrated using a small sample of soil dioxin collected from Midland County, Michigan State, U.S.A.
8. In the paper, "Predicting Air Pollution Using Fuzzy Membership Grade Kriging", we explored the spatial analysis technique – *fuzzy membership grade kriging* with semi-statistical membership, in order to address fuzzy spatial data recorded as crisp numbers. A practical situation often facing us is that fuzzy spatial data are recorded as crisp real-valued numbers, e.g., a PM_{10} record is 15.1, but we do know that it is an imprecise and vague observation. We

proposed a *fuzzy membership grade kriging* approach for the spatial prediction of the PM_{10} concentrations in California. It is an exploratory data analysis. Three semi-statistical membership functions are produced and fuzzy membership grade kriging was performed. It is noticed that the newly-introduced hyperbolic tangent membership function offers a refining feature over the other two sample membership functions. The membership functions are semi-statistical because their sub-optimal nature. In order to seek an optimal statistical membership with solid mathematical foundation it is necessary to extract membership function from the data itself. However, the parameters for specifying membership are still not fully data-assimilated.

9. In the paper, "Rationale of GM(1,1) Modelling – A Variation Approach", we investigated the underlying mechanism of GM(1,1) model, particularly, the lifetime partition into average functioning time, (average) repair-improvement and random error via GM(1,1), in terms of a variational viewpoint. From the analysis performed in this paper, we conclude that standard GM(1,1) model is a variational problem with differential equation as its constraint. Furthermore, the efficiency of GM(1,1) model depends upon the choice of the constraint (shadow) differential equation. We also point out that a variational treatment of GM(1,1) model is already complicated, and if we go further to investigate the asymptotic distributions as well as model efficiency from small asymptotic theory, it would not help us to get a better probabilistic modelling treatment. The clear message from the variational analysis is GM(1,1) model proposed by Deng (1985) do enjoy the easiness and efficiency in modelling the discrete positive data sequence as long as the 1-AGO data sequence possesses grey exponential trend. For GM(1,1) modelling improvement, there are two ways: the first approach is the GM(1,1) modelling on residuals proposed by Guo and Cui (2006) and the second approach is using the extended GM(1,1) model based on the coupling principle (Guo et al., 2006a, 2006b).
10. In the paper, "Generalizations to Standard GM(1,1) Model", we examined the GM(1,1) modelling mechanism from its component-level models, and proposed families of extended GM(1,1) models via the coupling principle in GM(1,1) model. We explored Deng's (1985) GM(1,1) model from its component-level models: the shadow differential equation, the regression model and the intrinsic difference model. Furthermore, we examined the coupling nature of the standard GM(1,1) model, i.e., the GM(1,1) model is a variational problem with differential equation as its constraint. The "coupling" nature also restricts the form of the

regression model in the object function and associated AGO and IAGO operations. Based on these examinations, we stated the coupling principle for guiding the generalization to Deng's (1985) GM(1,1) model. Seven extended GM(1,1) families are proposed accordingly. We are expecting this foundational work will provide more flexible and feasible GM(1,1) models for fitting small sample data with high degree of data assimilation capability.

11. In the paper, "Rationale Exploration of First-Order One Variable Grey Differential Equation Model via Variational Approach", we developed a data-assimilation measure in terms of the concept of ε-similarity measure. We examined the nature of GM(1,1) model in terms of a variational view. Form the analysis performed in this paper, we conclude that standard GM(1,1) model is a variational problem with differential equation as its constraint. Furthermore, the efficiency of GM(1,1) model depends upon the choice of the constraint (shadow) differential equation. The clear message from the variational analysis is GM(1,1) model proposed by Deng (1985) do enjoy the easiness and efficiency in modelling the discrete positive data sequence as long as the 1-AGO data sequence possesses grey exponential trend. However, the classical GM(1,1) does not warrant the model accuracy as well the extrapolation capability, therefore it is necessary to use the best-fitted polynomial of order $(n-1)$ as constraint function in order to reach high data-assimilation requirement, although the convenience of GM(1,1) model will be totally lost. A possible remedy is to use the extended GM(1,1) model based on the coupling principle (Guo et al., 2006b, 2006c), in which the differential equation constraint function is chosen in terms of ε-similarity measure between the constraint function and the best-fitted polynomial of order $(n-1)$.
12. In the paper, "The Coupling of Regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models", we examined the GM(1,1) model from its component-level. Furthermore, we explored the coupling nature of the standard GM(1,1) model, i.e., the GM(1,1) model is a variational problem with differential equation as its constraint. The "coupling" nature also restricts the form of the regression model in the object function and associated AGO and IAGO operations whenever necessary. Based on these examinations, we state the coupling principle for guiding the generalization to GM(1,1) model. Seven extended GM(1,1) models are proposed accordingly. We are expecting this foundational work will provide more flexible and feasible GM(1,1) models for fitting small sample data with high degree of data assimilation capability.

13. In the paper, "Bivariate Credibility-Copulas", we explored the concept of fuzzy copula on the ground of $(\vee,\cdot)$-credibility measure theory, called credibility-copula, for the characterization of the full relationship among fuzzy variables. The credibility-copula is a bivariate uniform credibility distribution in the sense of the credibility measure theory. Parallel to the basic understanding in probabilistic copula, we regard the credibility-copula to be a measure of the full bivariate dependence. Furthermore, we established a decomposition of credibility-copula. The decomposition explicitly reveals that copula is a total measure of independence as well as dependence and only the removal of the product copula component from the copula could accurately describe the true dependence between the two fuzzy variables. We must point out that the mathematical formulation developments for credibility-copula and the decomposition are similar to those in the probabilistic copula theory. However, the underlying mechanism on the ground of credibility theory is different from that on the ground of probability theory, because credibility measure theory and probability measure theory describe different phenomena: one is fuzzy uncertainty and other is random uncertainty.
14. In the paper, "Kernel Estimation for Bivariate Credibility Copulas", we developed a kernel-estimation based on the maximum entropy principle under the framework of $(\vee,\cdot)$-credibility measure theory. We extensively reviewed the related work on $(\vee,\cdot)$-credibility measure theory. Based on the review, we proposed the concept of copula for bivariate fuzzy variables, called credibility-copula, which is a bivariate uniform credibility distribution in the sense of the credibility measure theory. Parallel to the basic understanding in probabilistic copula, we regard the credibility-copula to be a measure of the full bivariate dependence. Furthermore, we established a two-stage procedure for credibility-copula based on data-assimilation under the fuzzy maximum entropy principle.
15. In the paper, "Credibility Measure-Based Fuzzy Membership Grade Kriging", we proposed an empirical average sample entropy criterion for estimating the parameter associated with membership function in terms of maximum entropy principle. It is a convenient spatial modelling with fully data-assimilated membership function. We used a credibility measure based fuzzy membership grade kriging approach for the spatial prediction of the PM_{10} concentrations in California. It is an exploratory data analysis. It is noticed that credibility measure theoretical foundation proposed by (Liu, 2004, 2006) put all the mathematical developments in a rigorous way, and therefore we have the chance to develop the maximum

entropy data-assimilated membership function by assuming a hyperbolic tangent form. The membership function is no longer semi-statistical as Guo et al. proposed (2004, 2006c). In order to seek an optimal data-assimilated membership with solid mathematical foundation, it is necessary to extract membership function from the data itself. As we emphasized at the beginning of the membership grade kriging we aim at its convenience of application not the mathematical rigidity. A fully rigorous approach should use credibility-chance measure theoretical framework, proposed by (Liu, 2004, 2006) for dealing random fuzzy spatially distributed observational data.

4.2 Future Developments

In this thesis, we explored fuzzy membership grade kriging and grey spatial prediction. Our future research will be more focused on variable modelling of spatially distributed data, with random and fuzzy uncertainty. It has the advantages of easy understanding, convenient implementation, and easy to combine with GIS analysis.

The spatially distributed observations may be recorded in the form of real-valued numbers, interval-valued numbers, or fuzzy numbers (i.e., fuzzy subsets) etc. Random uncertainty and fuzzy uncertainty coexist in spatial data intrinsically and inherently. Quite often, it causes modelling difficulties and non-suitability for the implementation with GIS. Therefore, it is still a long road to pursue variable modelling methodology within GIS.

There are some fundamental problems waiting for us to address:

1. How to present fuzzy variable and its credibility distribution on maps extracted from fuzzy subset data.
2. How to present random fuzzy variable and its chance distribution on maps extracted from random fuzzy subset data.
3. How to present bi-fuzzy variable and its chance distribution on maps extracted from random fuzzy subset data.

Facing the problem of sparse environmental data, we need to merge differential geometric theory, grey theory, differential equation theory, and other approximation theory together for spatial analysis and predictions of small sample data, and to present it on maps. In other words, Grey GIS will still be a crucial part of our future research direction.

Chapter 5. Conclusion

The research objective of this thesis is to model spatial environmental data with fuzzy uncertainty, and to explore the use of small sample data in Geographic Information System (GIS). This central objective was reached through the fifteen thesis papers, which were written in contribution, to the development of spatial uncertainty modelling. Fourteen of the peer-reviewed papers were already published, and one paper is under-print. There are a total of three journal papers and twelve conference papers.

Methodologies such as geostatistics, fuzzy mathematics, Grey System Theory, and $(\vee,\cdot)$ Credibility Measure Theory; have been merged together and applied to two datasets, PM_{10} air pollution data and soil dioxin data. The thesis papers were classified into two broad categories according to their aims: fuzzy spatial GIS modelling and grey spatial GIS modelling.

1. In fuzzy spatial GIS modelling, the fuzzy uncertainty in environmental data is addressed. Spatially distributed fuzzy data is in nature in the form of fuzzy subsets (Zadeh, 1965), and therefore is difficult to perform spatial modelling and presentation using GIS. The thesis papers aim at creating a convenient fuzzy kriging methodology. This led to the development of fuzzy membership grades kriging approach by converting fuzzy subsets spatial modelling into membership grade spatial modelling, which is easy to handle and implement into GIS. Although the early papers (chapter 1, 3, 8) used classical fuzzy mathematics concepts and methodology, it emphasized the data assimilated membership function by determining parameters of membership function in semi-statistical manner. In the later papers (chapter 13, 14, 15), the fuzzy membership grades kriging is put into the foundation of the credibility measure theory, and approached a full data-assimilated membership function in terms of maximum fuzzy entropy principle. The variable modelling method in dealing with fuzzy data is a unique contribution to the fuzzy spatial GIS modelling literature.
2. In grey spatial GIS modelling, spatial predictions using small sample data is addressed. The papers started with a projection approach (chapter 2, 4), however, the two-dimensional order-less spatially observations are converted into two one-dimensional ordered data sequences, and therefore the grey differential equation models can be successful applied. After the earlier Grey GIS modelling applications, it is found that there are a series of fundamental issues affecting the creation of grey spatial GIS modelling with small sample data. Therefore, the

later papers (chapter 5, 6, 9, 10, 11, 12) explored foundational problems such as the transformations, rationale underling GM(1,1) model via variational calculus, and ε-similarity measure between two geometric curves (i.e., functions) etc. It is discovered the coupling feature of grey differential equations together with the help of ε-similarity measure, generalise the classical GM(1,1) model proposed by Deng (1985) into more classes of extended GM(1,1) models, in order to fully assimilate with sample data information. The development of grey spatial GIS modelling is a creative contribution to handling small sample data.

Our future research will be more focused on variable modelling of spatially distributed data with random and fuzzy uncertainty. It has the advantages of easy understanding, convenient implementation, and easy to combine with GIS analysis. However, Grey GIS will still be a crucial part of our future research direction.

References

1. Alliance for Safe Alternatives. (2005). Dioxin – Common Questions and Answers. http://www.safealternatives.org/faq.html
2. Aptronix. (2000). What Is Fuzzy Logic? Aptronix Inc. http://www.aptronix.com/fide/whatfuzzy.htm
3. Answermath. (2002). Fuzzy Logic. Answermath.com. http://www.answermath.com/fuz.htm
4. Bandemer, H. and Gebhardet, A. (2000). Bayesian Fuzzy Kriging. Fuzzy Sets and Systems, 112: 405-418.
5. Bardossy, A., Bogardi, I., and Kelly, E. (1988). Imprecise (Fuzzy) Information in Geostatistics. Mathematical Geology, 20: 287-311.
6. Bardossy, A., Bogardi, I., and Kelly, E. (1989). Geostatistics Utilizing Imprecise (Fuzzy) Information. Fuzzy Sets and Systems, 31: 311-328.
7. Bardossy, A., Bogardi, I., and Kelly, E. (1990a). Kriging with Imprecise (Fuzzy) Variograms, I: Theory. Mathematical Geology, 22: 63-79.
8. Bardossy, A., Bogardi, I., and Kelly, E. (1990b). Kriging with Imprecise (Fuzzy) Variograms, II: Applications. Mathematical Geology, 22: 81-94.
9. Burrough, P.A. and McDonnell, R.A. (1998). Principles of Geographical Information Systems. New York: Oxford University Press.
10. Chen, S.Y. (1998). Engineering Fuzzy Set Theory and Application. Beijing: Defense Industry Publishing House.
11. Chrisman, N. (1997). Exploring Geographic Information Systems. New York: John Wiley & Sons.
12. Dueker, K.J. and Kjerne, D. (1989). Multi-Purpose Cadastre: Terms and Definitions. Falls Church, VA: ASPRS and ACSM.
13. Cox, D.R. (1972). Regression models and life tables. Journal of Royal Statistical Society, B26: 187-220.
14. Cressie, N.A.C. (1991). Statistics for Spatial Data. New York: John Wiley & Sons.
15. De Luca, A. and Termini S. (1972). A Definition of Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. Information and Control, 20: 301-312.

16. Deng, J.L. (1982). Control Problems of Grey Systems. Systems and Control Letters, 1: 288-294, March.
17. Deng, J.L. (1985). Grey Systems (Social · Economical). Beijing: The Publishing House of Defense Industry (in Chinese).
18. Deng, J.L. (1997). Grey Control System. Second Edition. Wuhan City: Huazhong University of Technology Press (in Chinese).
19. Deng, J.L. (2002a). The Foundation of Grey Theory. Second Edition. Wuhan City: Huazhong University of Technology Press (in Chinese).
20. Deng, J.L. (2002b). Grey Prediction and Decision. Second Edition. Wuhan City: Huazhong University of Technology Press (in Chinese).
21. Diamond, P. (1989). Fuzzy Kriging. Fuzzy Sets and Systems, 33: 315-332.
22. Dos Anjos, U. (2005). Representation of Bivariate Copulas via Local Measure of Dependence. Proceedings of the Second Brazilian Conference on Statistical Modelling in Isurance and Finance, August 28-September 3, 2005, Maresias, Brazil, pp. 264-267.
23. Duan, Q.J. and Zhang, F.X. (1998). Some Grey Problems in Reliability Analysis of Complicated System. Journal of Nanjing University, 12: 96-98 (in Chinese).
24. Dyke, H., Stamp, T., Chamberlain, H., and Ullyett, J. (2003). GIS and Policy Making: Beyond Aesthetically Pleasing Pictures. Environmental Systems Research Institute. gis.esri.com/library/userconf/europroc96/PAPERS/PN14/PN14F.htm
25. Environmental Protection Agency (EPA). (2003). Particulate Matter. U.S. Environmental Protection Agency. http://www.epa.gov/region09/air/aqreport/pm10.html
26. Environmental Protection Agency (EPA). (2005a). National Ambient Air Quality Standards, Air and Radiation. U.S. Environmental Protection Agency. http://www.epa.gov/
27. Environmental Protection Agency (EPA). (2005b). Dioxins. U.S. Environmental Protection Agency. http://www.epa.gov/
28. Environmental Systems Research Institute (ESRI). (2001). Getting to Know ArcGIS Desktop. Redlands: Environmental Systems Research Institute.
29. Field, C.A. and Ronchetti, E. (1990). Small Sample Asymptotics. Haywars (CA): Institute of Mathematical Statistics Monograph Series.
30. Fu, L. (1991). Grey System Theory and Applications. Beijing: The Publishing House of Science and Technology (in Chinese).

31. Gallagher, D. (1997). Geostatistics. Environmental Engineering Program, VirginiaTech. http://www.cee.vt.edu/program_areas/environmental/teach/smprimer/kriging/kriging/
32. Gandin, L.S. (1963). Objective Analysis of Meteorological Fields. Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965).
33. Guo, D. (2003). Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM_{10} for California, Using Fuzzy Kriging. MSc Thesis, Department of Geography, University of Durham, U.K.
34. Guo, D., Guo, R., and Thiart, C. (2004). Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM10 for California, Using Fuzzy Kriging. Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2004), University of East Anglia, Norwich, U.K., pp. 375-380.
35. Guo, D., Guo, R., and Thiart, C. (2005). Grey GIS. Proceedings of the Geographical Information Systems Research UK Conference (GISRUK 2005), Glasgow, Scotland, pp. 379-387.
36. Guo, D., Guo, R., and Thiart, C. (2006a). The Coupling of regression Modeling and Differential Equation Model in GM(1,1) Modeling and Extended GM(1,1) Models. A draft paper submitted to the Journal of Grey System. (Accepted).
37. Guo, D., Guo, R., Thiart, C., and Cui, Y.H. (2006b). Rationale of GM(1,1) Modeling – A Variation Approach. Proceedings of the International Conference on Maintenance Engineering (ICME 2006), October 15-18, 2006, Cheng Du, China. (Accepted).
38. Guo, R. (2004). A Fuzzy System Analysis on Repairable System Via Interval-Valued Fuzzy Set Approach. Proceedings of the 9th Annual International Conference on Industrial Engineering – Theory, Applications and Practice, November 27-30, 2004, the University of Auckland, Auckland, New Zealand, pp. 146-151.
39. Guo, R. (2005a). A Repairable System Modelling by Combining Grey System Theory with Interval-Valued Fuzzy Set Theory. International Journal of Reliability, Quality and Safety Engineering, 12 (3): 241-266. Kimura, M., Dohi, T., and Yun, W.Y. (Guest Editors).
40. Guo, R. (2005b). Repairable System Modelling Via Grey Differential Equations. Journal of Grey System, 8 (1): 69-91.

41. Guo, R. and Cui, Y.H. (2006). Foundations For GM(1,1) Modeling of System Lifetimes. A draft paper submitted to the Proceedings of the International Conference on Maintenance Engineering (ICME), October 15-18, 2006, Cheng Du, China. (Under review).
42. Guo, R. and Dunne, T. (2005). Stochastic Age Processes. Proceedings of the European Safety and Reliability Conference (ESREL 2005): Advances in Safety and Reliability, Tri City (Gdynia-Sopot-Gdansk), Poland, Vol. 1, pp. 745-752.
43. Guo, R. and Love, C.E. (2003). Reliability Modelling with Fuzzy Covariates. International Journal of Reliability and Safety Engineering, 10: 131-157.
44. Guo, R. and Love, C.E. (2004). A Virtual State Analysis Of Repairable System Via Logical Functions. Proceedings of the 9th Annual International Conference on Industrial Engineering – Theory, Applications and Practice, November 27-30, 2004, the University of Auckland, Auckland, New Zealand, pp. 140-145.
45. Guo, R. and Love, C.E. (2005a). Grey Repairable System Analysis. Invited paper as Plenary Lecture. Proceedings of the European Safety and Reliability Conference, June 27-30, 2005, Tri City, Poland, pp. 753-766.
46. Guo, R. and Love, C.E. (2005b). Fuzzy Set-Valued and Grey Filtering Statistical Inferences on a System Operation Data. International Journal of Quality in Maintenance Engineering – Advanced Reliability Modelling, 11(3): 267-278. Dohi, T., Kaio, N., and Yun, W.Y. (Guest Editors).
47. Guo, R., Guo, D., and Li, X. (2006c). Bivariate Credibility-Copulas. A draft paper submitted to the Proceedings of the 1st International Conference on System Integration and Reliability Improvements (SIRI 2006), December 13-15, 2006, Hanoi, Vietnam. (Under review).
48. Hajek, P. and Kramosil, I. (2002). Ten Claims About Fuzzy Logic. Institute of Computer Science, Academy of Science of the Czech Republic. http://new.euromise.org/english/material/95102023/hajekp.html
49. James, N. (2001). Mapping the Sense of Place. Undergraduate Thesis in Environmental Studies, Brown University, Spring 2001. http://envstudies.brown.edu/Thesis/2001/james/gishistory.html
50. Jaroslav, S. (2003). Special Achievement in GIS. Environmental Systems Research Institute. gis.esri.com/uc2003/sag/list/detail.cfm?SID=168
51. Jiang, T. and Chen, J.J. (2006). Linear Performance Function as a Basis for Assignment of Membership Function in Profust Reliability Context. Unpublished manuscript.

52. Journel, A.G. (1983). Nonparametric Estimation of Spatial Distributions. Journal of the International Association for Mathematical Geology, 15: 445-468.
53. Journel, A.G. and Huijbregts, C.J. (1978). Mining Geostatistics. London: Academic Press.
54. Kitanidis, P.K. (1997). Introduction to Geostatistics: Applications to Hydrogeology. Cambridge: Cambridge University Press.
55. Kosko, B. (1994) Fuzzy Thinking: The new Science of Fuzzy Logic. London: Flamingo.
56. Krige, D.G. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. Journal of the Chemical, Metallurgicall and Mining Society of South Africa, 52: 119-139.
57. Kaufmann, A. (1975). Introduction to the Theory of Fuzzy Subsets, Vol. I. New York: Academic Press.
58. Lee, E.S. (2000). Nero-Fuzzy Estimation in Spatial Statistics. Journal of Mathematical Analysis and Applications, 249: 221-231.
59. Li, X. and Liu, B. (2005). New Independence Definition of Fuzzy Random Variable and Random Fuzzy Variable. World Journal of Modelling and Simulation.
60. Liu, B. (2004). Uncertainty Theory: An Introduction to Its Axiomatic Foundations. Berlin: Springer-Verlag Heidelberg.
61. Liu, B. (2006). A Survey of Credibility Theory. Fuzzy Optimization and Decision Making, 5: 4.
62. Liu, B. and Liu, Y.K. (2002). Expected Value of Fuzzy Variable and Fuzzy Expected Value Models. IEEE Transactions on Fuzzy Systems, 10 (4): 445-450.
63. Liu, B., Zhao, R., and Wang, G. (2003). Uncertain Programming with Applications. Beijing: Tsinghua University Press (in Chinese).
64. Liu, S.F. and Lin Y. (2006). Grey Information – Theory and Practical Applications. London: Springer-Verlag.
65. Liu, S.F., Dang, Y.G., and Fang, Z.G. (2004). Grey System Theory and Applications. Beijing: The Scientific Publishing House (in Chinese).
66. Mathematics Handbook Group. (1979). Mathematics Handbook, pp. 439. Beijing: Higher Education Publisher (in Chinese).
67. Matheron, G. (1963). Principles of Geostatistics. Economic Geology, 58: 1246-1266.
68. Nelsen, R.B. (1999). An Introduction to Copulas. New York: Springer.

69. Omre, H. (1987). Bayesian Kriging – Merging Observations and Qualified Guesses in Kriging. Mathematical Geology, 19: 25-39.
70. Pacific Northwest National Laboratory (PNNL). (2001). What Is Fuzzy Logic? Pacific Northwest National Laboratory. http://www.emsl.pnl.gov:2080/proj/neuron/fuzzy/what.html
71. Prigogine, I and Stengers, I. (1984). Order Out of Chaos. London: Heinemann.
72. Pujol, A. (2001). Fuzzy Logic. Electronics, Electrics and Automation Engineering Department, Rovira i Virgili University. http://www.etse.urv.es/~aoller/fuzzy/fuzzy_logic.htm
73. Rocky Mountain Center. (1996). What is PM10? The Rocky Mountain Center. http://rocky.utah.edu/pm10/pm10what.html
74. Salski, A. (1996). Fuzzy Logic Approach to Data Analysis and Ecological Modelling. Ecology Center of the University of Kiel. http://www.erudit.de/erudit/events/esit99/12571_p.pdf
75. Spatial Analysis and Decision Assistance (SADA). (2003). Geospatial Methods. University of Tennessee Corporation. http://www.tiem.utk.edu/~sada/help/
76. Surveys and Presentations in Information Systems Engineering (SURPRISE). (1996). Fuzzy Logic and Its Uses. Surveys and Presentations in Information Systems Engineering, Department of Computing, Imperial College London. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol2/jp6/article2.html
77. TRIfacts. (2005). TEQ vs. TM-17 Method. TRIfacts.org. http://www.trifacts.org/teq_tm17/index.php
78. Vega, N.A. (2003). Potential Applicability of Fuzzy Logic in Geostatistics. Department of Mining and Minerals Engineering, VirginiaTech. http://filebox.vt.edu/users/anieto/web/fuzzy/fuzzy.htm
79. Wen, K.L. (2004a). Grey Systems: Modelling and Prediction. Tucson, U.S.A.: Yang's Scientific Research Institute.
80. Wen, K.L. (2004b). The Grey System Analysis and Its Application in Gas Breakdown and VAR Compensator Finding. International Journal of Computational Cognition, 2 (1): 21-44, March.
81. Wisconsin State Cartographer's Office. (2001). GIS Basics. Wisconsin State Cartographer's Office. http://www.geography.wisc.edu/sco/gis/history.html

82. World Health Organization (WHO). (1998). WHO Revises the Tolerable Daily Intake (TDI) for dioxins. World Health Organization European Centre for Environment and Health; International Programme on Chemical Safety. Organohalogen Compounds, 38: 295-298.
83. Xin, T.Y. and Liu, J.H. (1996). The Grey Reliability of Structure. Journal of Northwest Institute of Architecture Engineering, 1: 64-68, (in Chinese).
84. Zadeh, L.A. (1965). Fuzzy Sets. Information and Control, 8: 338-353.
85. Zadeh, L.A. (1973). Outline of a New Approach to the Analysis of Complex Systems and Decision Processes. IEEE Transactions on Systems, Man and Cybernetics, 3: 28-44.
86. Zadeh, L.A. (1978). Fuzzy Sets as a Basis for a Theory of Possibility. Fuzzy Sets and Systems, 1: 3-28.

Appendix A: Grey Temporal-Spatial GIS

This draft paper "Grey Temporal-Spatial GIS" is still under writing. The authors are Danni Guo, Renkuan Guo, and Christien Thiart.

A.1 Introduction

Collecting environmental data is a very costly and difficult process, and quite often the sampled data are insufficient for any practical analysis. In this paper, we will further extend *Grey Spatial Analysis* proposed by Guo et al. (2005) into "small-sample" based *Grey Temporal-Spatial Analysis*. PM_{10} data of Santa Barbara County, California, U.S.A. is used for illustration.

The newly proposed *Grey Temporal-Spatial Analysis* methodology will utilize available information on an area interested from three aspects: a few historical time sequences at specific site; predicted or directly observed values on locations surrounding interested area; and general trend over the study area. The results are imprecise (grey numbers) in nature. However, practically the grey predictions are unexpectedly consistent with local sample information available.

A.2 PM_{10} Distribution in Santa Barbara County

PM_{10} is one of the seven air pollutants the Environmental Protection Agency (EPA) regulates, and exposure to high outdoor PM_{10} concentrations causes increased disease and death (Environmental Protection Agency, 2005a). The California Air Resources Board has 55 air quality monitoring sites located within the Santa Barbara County (figure A.1).

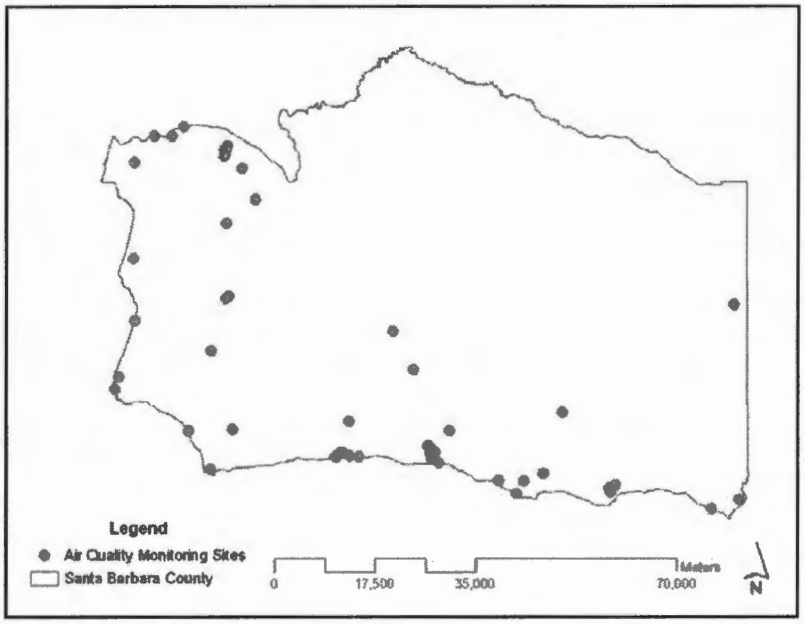


Figure A.1 Air Quality Monitoring Sites in Santa Barbara County.

However, due to the cost and manpower involved in the sampling process, very few samples are actually collected each year. In 2002, only 4 PM_{10} samples are collected within Santa Barbara County (figure A.2).

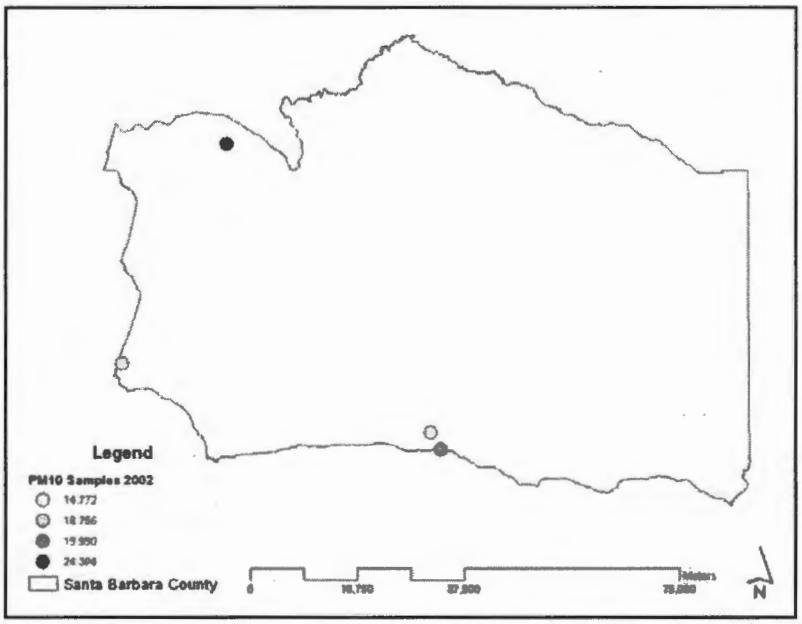


Figure A.2 PM_{10} Sample Locations for 2002 in Santa Barbara County.

In GIS software, kriging prediction maps can only be constructed under the condition that there must be *10* or more sample points. In this case, only *4* samples were collected in 2002, we utilize the newly developed *Grey Spatial Analysis* methodology to estimate the PM_{10} in different locations within Santa Barbara. For convenience, in this case, we estimated the PM_{10} concentrations in the other 51 monitoring sites within Santa Barbara County.

The *Grey Spatial Analysis* method only predicted 34 locations in Santa Barbara County, and there are *17* locations (brown dots) in the southeast of the county are unpredicted (figure A.3).

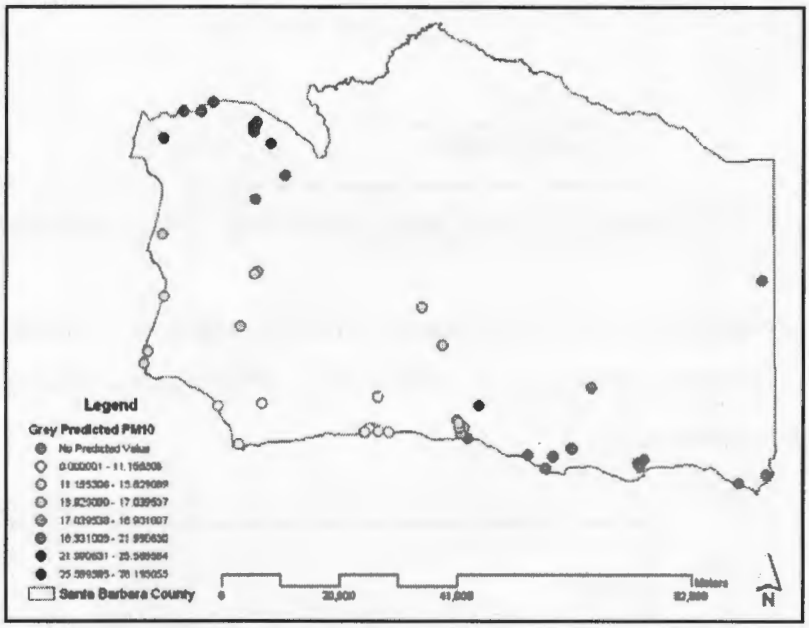


Figure A.3 Grey Prediction for PM_{10} in 2002 in Santa Barbara County.

We examined further the air quality data of Santa Barbara County further and realized that there were two sites, site 17 and 38, containing historical PM_{10} data time sequences, which only reached up to 1999 and 1997 respectively (table A.1).

Table A.1 PM_{10} Time Sequences from year 1988 to 1999 at Monitoring Site 17, 38 and 57.

Site	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*	*1995*	*1996*	*1997*	*1998*	*1999*
17	*		36.871	*	30.058	*	*	*	*	29.747	26.079	29.49
38	*	27.573	26.452	*	25.267	26.131	26.42	24.971	25	27.42	*	*
57	33.712	36.073	34.157	32.417	*	26.539	24.799	22.082	*	*	*	*

As a record, site 17 and 38, are within Santa Barbara County, but site 57 is the nearest one outside the county, and it has historical data up to 1995. The purpose for adding site 57 is to strengthen the interpolation area coverage (figure A.4).

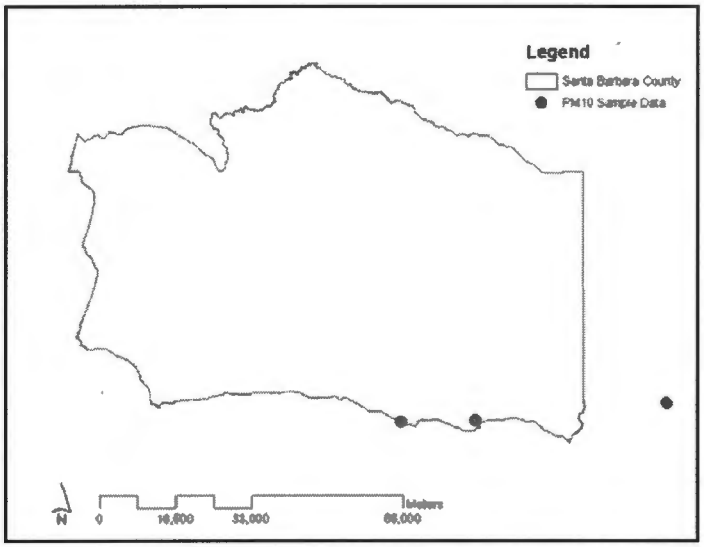


Figure A.4 Three Samples Sites with Historical PM_{10} Data in Santa Barbara County.

In addition of site 17, 38 and 57 in 2002 are available, we choose PM_{10} values (grey-predicted) at site 2, 23 and 45 respectively. Then, with these 6 PM_{10} values a *Grey Spatial Analysis* can be performed for the prediction of the remaining PM_{10} concentrations at the 17 sites.

Table A.2 Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.

Site	*Latitude*	*Longitude*	*PM_{10} Value 2002*
2	34.4622	-120.0258	19.23121493
17	34.4166	-119.6999	*
23	34.5136	-120.0055	26.91565579
38	34.4147	-119.8788	*
45	34.4691	-120.0394	15.79061345
57	34.4480	-119.2420	*

If we can predict PM_{10} 2002 values at site 17, 38 and 57 respectively, the remaining analysis is relatively easy. However, by examining the three data sequence in table A.1, the site 38 has 8 years of interrupted data, and site 17 has 5 years of interrupted data. Therefore, classical time-series analysis methodology will not properly work here because the data sequence is too short and even not equal-time spaced, i.e., some "missing" values between them.

A.3 Grey System Theory

The Grey System Theory (Deng, 1982) was rooted in modern control theory, where system dynamics are classified by the degree of information availability, and accordingly the different methodologies are developed for each of them respectively. For a general picture of commonly faced three uncertain systems, we give a brief comparison in table A.3 (modified from Liu and Lin, 2006).

Table A.3 Comparisons of "Grey", "Probabilistic" and "Fuzzy" Systems.

Aspect	*Grey System*	*Probabilistic System*	*Fuzzy Set System*
Set foundation	Haze sets	Cantor sets	Fuzzy sets
Connotation and extension	Connotation haze with clear boundary and extension	Random event with connotation and extension well-defined	cognitive uncertainty (clear connotation but vague extension)
Core concept	Grey derivative and differential equations	Probability distribution	membership function
Data treatment	(inverse) accumulative generating operation	Sampling statistics & asymptotic distribution	Membership grade, λ-cut set & extension principle
Data requirements	Small sample size	Large sample size	empirical (+ sampling data)

A critical feature of grey system is information incompleteness. The task of establishing model under the guidance of grey system theory is inevitably to seek model building based on data of *small sample size*. Its target is establishment of grey differential equation and emphasizes the exploration, utilization and processing dynamic information containing in data (Guo, 2004; Guo and Love, 2005a, 2005b).

A.4 Grey Temporal Analysis

Let us formally state that equation

$$x^{(0)}(k)+\beta z^{(1)}(k)=\alpha, \quad k=2,\cdots,n \tag{1}$$

is called a one-variable first order grey (temporal) differential equation with respect to time series sequence $X^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n))$, where

$$z^{(1)}(k)=\frac{1}{2}\left[x^{(1)}(k)+x^{(1)}(k-1)\right],\ \ k=2,\cdots,n$$
$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),\ \ k=2,\cdots,n \tag{2}$$

The unknown parameter values (α,β) can be determined in terms of the classical least-square method. We rewrite equation 1 as:

$$\alpha+\beta\left(-z^{(1)}(k)\right)=x^{(0)}(k),\ \ k=2,3,\cdots,n \tag{3}$$

Then a standard matrix form of the equation 1 can be formed in terms of least-square theory,

$$X\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=y \tag{4}$$

where

$$X=\begin{bmatrix}1 & -z^{(1)}(2)\\ 1 & -z^{(1)}(3)\\ \vdots & \vdots\\ 1 & -z^{(1)}(n)\end{bmatrix}\ \text{ and }\ y=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix} \tag{5}$$

which leads to the estimate for parameter (α,β),

$$\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}=\left(X^{T}X\right)^{-1}X^{T}y \tag{6}$$

But considering the three time sequences are not equal-time spaced (containing missing valued), it is necessary to take a second round of least-square to estimate the remaining parameters. The $\hat{\alpha}$ has no influence in the second round estimation, because the second round estimation works on derivative of $x^{(1)}(t)$ directly.

Let us assume that $x^{(0)}(t)$ takes the form

$$x^{(0)}(t)=\gamma\exp\left(-\hat{\beta}t\right)+\alpha \tag{7}$$

Then

$$(\hat{\alpha},\hat{\gamma})^{T}=\left(\mathrm{D}^{T}\mathrm{D}\right)^{-1}\mathrm{D}^{T}\mathrm{y} \tag{8}$$

where

$$D=\begin{bmatrix} 1 & \exp\left(-\hat{\beta}t_1\right) \\ 1 & \exp\left(-\hat{\beta}t_2\right) \\ \vdots & \vdots \\ 1 & \exp\left(-\hat{\beta}t_N\right) \end{bmatrix} \tag{9}$$

and

$$y=\begin{bmatrix} x^{(0)}(t_1) \\ x^{(0)}(t_2) \\ \vdots \\ x^{(0)}(t_N) \end{bmatrix} \tag{10}$$

Therefore after two-stage least-square fitting, the estimated response function is

$$x^{(0)}(t)=\hat{\alpha}+\hat{\gamma}\exp\left(-\hat{\beta}t\right) \tag{11}$$

which offers the grey temporal prediction at each site.

A.5 Grey Temporal Prediction Results

The predicted PM_{10} values from 1997 to 2002 follow a sharp decreasing trend because the observed values demonstrate such a local trend (table A.4).

Table A.4 Grey Temporal Predictions of PM_{10} from 1992 to 2002 at Site 17, 38 and 57.

Location	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
17	30.058	*29.606*	*28.971*	*28.621*	*28.428*	29.747	26.079	29.49	*28.214*	*28.204*	*28.199*
38	25.267	26.131	26.42	24.971	25	27.42	*25.705*	*25.630*	*25.558*	*25.490*	*25.424*
57	*30.056*	26.539	24.799	22.082	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*	*17.353*

As a reference, the Santa Barbara County's kriging map (figure A.5) shows a general trend of increasing values inland.

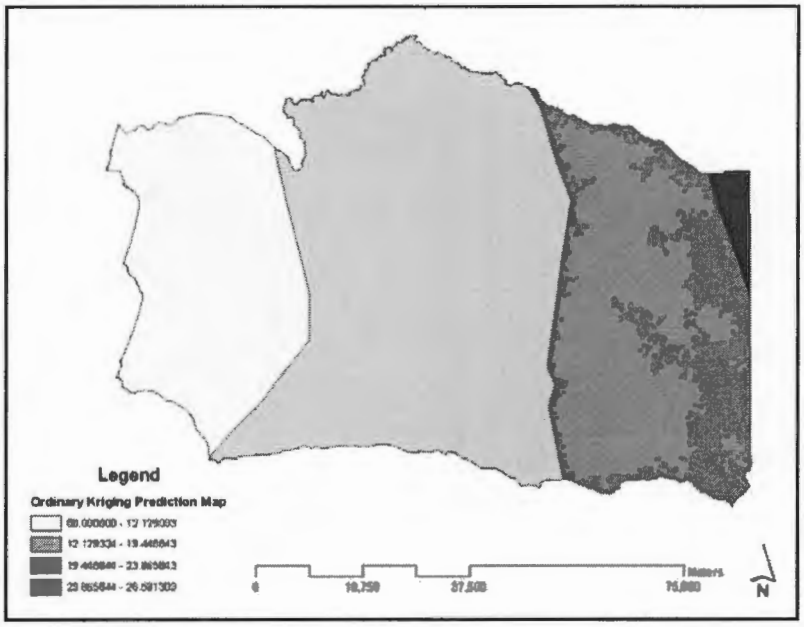


Figure A.5 Ordinary Kriging Prediction Map for PM_{10} in 2002 in Santa Barbara County.

A.6 Grey Temporal-Spatial Prediction Results

The latest information for grey spatial predictions is summarised (table A.5).

Table A.5 Information of the Six Sites for Next Grey Prediction of the Remaining 17 Sites.

Location	Latitude	Longitude	PM_{10} Value 2002
2	34.4622	-120.0258	19.23121493
17	34.4166	-119.6999	*28.1986600*
23	34.5136	-120.0055	26.91565579
38	34.4147	-119.8788	*25.4244800*
45	34.4691	-120.0394	15.79061345
57	34.4480	-119.2420	*17.3531780*

With this data information, we now use newly the developed Grey Spatial Analysis methodology (Guo et al., 2005) to make the predictions for the remaining 17 sites. As a record, we choose (-120.040, 34.4000) as the new origin for setting up a 2-dimensional Cartesian Coordinate system. The prediction results are shown in figure A.6.

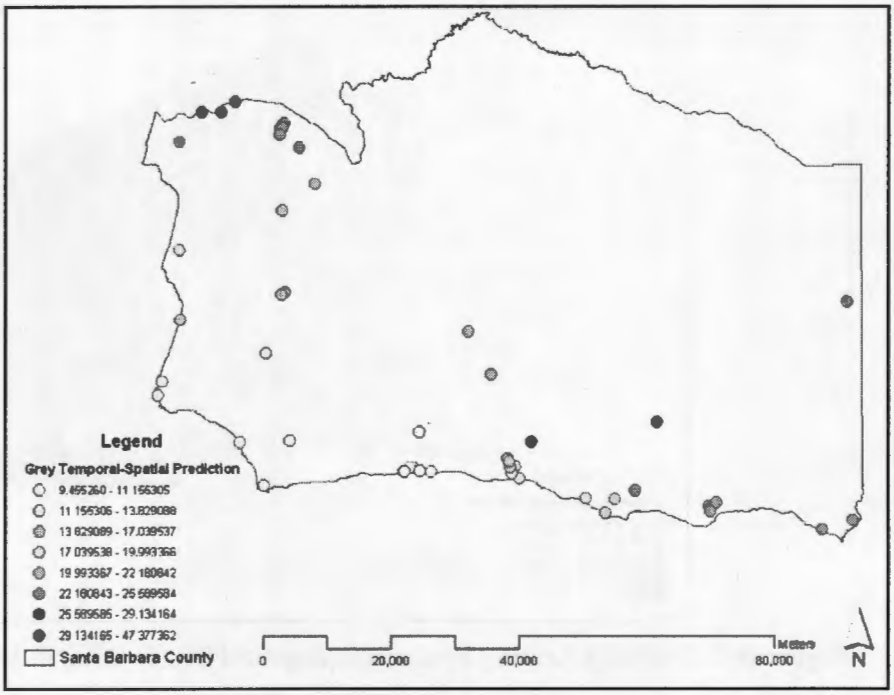


Figure A.6 Grey Temporal-Spatial Prediction for PM_{10} in 2002 in Santa Barbara County.

The grey predictions and kriging predictions are consistent in general trend. However, grey predictions reveal better local properties. The Grey PM_{10} predictions are shown to be very sensitive to local observations.

A.7 Conclusion

In this paper, we have developed a small-sample based temporal-spatial data analysis methodology, called *Grey Temporal-Spatial Analysis* and combined with GIS. The results have shown that the grey predictions reveal the both general tendency and locality reasonably well.

Appendix B: Summary of Contributed Papers

Table B.1 Summary of Thesis Papers.

Paper Title 1:	Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM10 for California, Using Fuzzy Kriging
Main Aims:	Develop a convenient mathematical approach for fuzzy observations spatial modelling and implement with GIS. The key ideas are using membership grades for kriging by avoiding complicated fuzzy subsets kriging and setting parameters for membership function by semi-statistical method in terms of data sampled. This paper is a logical extension of indicator kriging.
Methods used:	kriging, fuzzy mathematics, GIS
Contributions to literature:	Spatially distributed fuzzy observations (in nature, fuzzy subsets) are difficult to model and present with GIS. This paper creates a convenient methodology by converting fuzzy subset spatial modelling into membership grade spatial modelling, which is easily to handle and implement with GIS.
Contributing portion:	65%
Paper Title 2:	Grey GIS
Main Aims:	To develop a convenient mathematical method for modelling spatial data with very small sample size, in the case of where the traditional kriging is not applicable. The ideas are argued from three aspects: (i) The problem of modelling spatially distributed data with very small sample size is a practical one, (ii) Traditional statistical distribution-thinking is not suitable and we need to utilize small-sample-oriented highly predictive GM(1,1) model, suitable for ordered positive data sequence. (iii) Projecting two-dimensional order-less observations into one-dimensional ordered positive data sequence and use GM(1,1) model to establish two-dimensional prediction map, which is implemented and combined with Geographic Information Systems (GIS), and this method is *Grey GIS*.
Methods used:	GM(1,1) model, GIS
Contributions to literature:	This work creates a non-probabilistic Spatial-GIS modelling approach. Its nature is a delicate conversion of two-dimensional spatial problem into one-dimensional modelling and thus the approach is convenient to use.
Contributing portion:	60%
Paper Title 3:	Grey Spatial Prediction of Fuzzy Air Quality Index
Main Aims:	Develop a conveniently handled fully data-assimilated membership function. Air pollutant measurements are fuzzy quantities and then it is necessary to determine the membership functions in order to prepare the membership grade kriging. We use the link between

	distribution and membership function under certain assumptions for determining parameter settings statistically from data.
Methods used:	Grey System Theory, fuzzy mathematics
Contributions to literature:	The approach offers a full data-assimilated membership function determination solves the subjective exercises in fuzzy mathematics.
Contributing portion:	60%
Paper Title 4:	A New Very Small Sample-Based Non-Linear Statistical Estimation Method
Main Aims:	To investigate a new least-square theory based estimation technique under very small sample circumstances. We call it as a grey state model because the approach does not involve the probability distribution underlying and directly seek the state law in terms of grey differential equation models.
Methods used:	differential geometry, grey differential equations, GIS
Contributions to literature:	Raise the idea and propose various ways to merge grey differential equation with differential geometry theory together for creating a non-probabilistic state modelling.
Contributing portion:	60%
Paper Title 5:	Optimal Data Transformations In Grey Analysis
Main Aims:	The fact that the accuracy of GM(1,1) model is not automatically guaranteed causes concern of the applicability in spatial modelling. One of the route is to explore the transformation of the discrete data sequence, in Grey Theory. Two new transformations are developed.
Methods used:	Grey Theory, transformation, grey differential equation models
Contributions to literature:	Type II transformation can be applied universally and obtain satisfactory GM(m,N) models at transformed data sequence $Y^{(0)}$ level, then it is a major contribution to grey data analysis, particularly, residual analysis.
Contributing portion:	60%
Paper Title 6:	Grey Reliability Analysis of Complex system
Main Aims:	To investigate a way to obtain the dynamic version of grey system reliability with sparse data availability, in terms of powerful grey differential equation models.
Methods used:	Grey Theory, degree of whitenization, whitenization function
Contributions to literature:	The dynamic nature of grey events and the two-fold whitenization approach could be converted a Grey spatial modelling approach, for example, local environmental evaluation under damaging factors and regenerating factors.
Contributing portion:	50%
Paper Title 7:	GM(1,1)-Kriging Prediction of Soil Dioxin Pattern
Main Aims:	To develop a mixed approach by combining grey differential equation models, particularly, GM(1,1) model and ordinary kriging

	approach together, and call it GM(1,1)-kriging.
Methods used:	kriging, GM(1,1) Model, GIS
Contributions to literature:	A creative approach to spatial prediction under sparse data availability of the areas under investigation.
Contributing portion:	60%
Paper Title 8:	Predicting Air Pollution Using Fuzzy Membership Grade Kriging
Main Aims:	To develop a new spatial analysis technique – *fuzzy membership grade kriging* with semi-statistical membership, in order to address fuzzy spatial data recorded as crisp numbers.
Methods used:	kriging, fuzzy mathematics, GIS
Contributions to literature:	This is journal paper version of **paper 1** with substantial revisions according to referees' comments.
Contributing portion:	75%
Paper Title 9:	Rationale of GM(1,1) Modelling – A Variation Approach
Main Aims:	To investigate the underlying mechanism of GM(1,1) model, particularly, the lifetime partition into average functioning time, (average) repair-improvement and random error via GM(1,1), in terms of a variational viewpoint.
Methods used:	Grey System Theory, calculus of variation
Contributions to literature:	Solidify the theoretical foundation of GM(1,1) models.
Contributing portion:	60%
Paper Title 10:	Generalizations to Standard GM(1,1) Model
Main Aims:	To examine GM(1,1) modelling mechanism from its component-level models, and propose families of extended GM(1,1) models via the coupling principle in GM(1,1) model. We will state the coupling principle for guiding the generalization to Deng's (1985) GM(1,1) model. Seven extended GM(1,1) families will be proposed accordingly.
Methods used:	Grey System Theory, difference equation, differential equation
Contributions to literature:	Creative generalizing the standard GM(1,1) model to extended GM(1,1) models, and thus enhancing the data-assimilation capability with implications on GIS spatial modelling.
Contributing portion:	55%
Paper Title 11:	Rationale Exploration of First-Order One Variable Grey Differential Equation Model via Variational Approach
Main Aims:	To develop a data-assimilation measure in terms of the concept of ε-similarity measure, and explore the evolving behaviour of best fitted polynomial and a set of constraint functionals defined by ordinary differential equations in order to select the suitable functional satisfying the ε-similarity measure during GM(1,1) modelling process.
Methods used:	Grey System Theory, grey relation, polynomial spline curves

Contributions to literature:	Create a geometric curve based ε-similarity measure for the selection of high data-assimilated GM(1,1) model as well as grey relation analysis, which pays the way for spatially distributed covariate modelling.
Contributing portion:	55%
Paper Title 12:	The Coupling of Regression Modelling and Differential Equation Model in GM(1,1) Modelling and Extended GM(1,1) Models
Main Aims:	To examine the GM(1,1) model from its component-level and therefore identify the interactive coupling nature of differential equation model and corresponding regression model constituting of a GM(1,1) model. Based our analysis, we state a coupling principle for establishing an extended GM(1,1) model and further explore certain families of extended GM(1,1) models.
Methods used:	Grey System Theory, differential equations
Contributions to literature:	The first time identifying the coupling nature of GM(1,1) model and paying the way of generalizations
Contributing portion:	55%
Paper Title 13:	Bivariate Credibility-Copulas
Main Aims:	To develop the concept of fuzzy copula on the ground of $(\vee,\cdot)$-credibility measure theory, called credibility-copula, for the characterization of the full relationship among fuzzy variables. Then we explore a decomposition of credibility-copula function into product copula and an adjusted dependence function, because for direct fuzzy variable modelling, the relation between two fuzzy variables has to be facilitated.
Methods used:	fuzzy mathematics, credibility measure theory
Contributions to literature:	A creative contribution to credibility measure theoretical framework because credibility-copula concept is first proposed in this paper.
Contributing portion:	60%
Paper Title 14:	Kernel Estimation for Bivariate Credibility Copulas
Main Aims:	To develop a kernel-estimation based on the maximum entropy principle under the framework of $(\vee,\cdot)$-credibility measure theory. We propose the concept of copula for bivariate fuzzy variables, called credibility-copula, which is a bivariate uniform credibility distribution in the sense of the credibility measure theory. Furthermore, we establish a two-stage procedure for credibility-copula based on data-assimilation under the fuzzy maximum entropy principle.
Methods used:	Grey System Theory, credibility measure theory
Contributions to literature:	First time work on the data-assimilated credibility distribution.
Contributing portion:	55%