We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Modelling PM2.5 with Fuzzy Exponential Membership

Danni Guo

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/59617

1. Introduction

Air quality is a public and environmental issue that concerns people, whether in terms of global climate change or for health and the quality of life. The Environmental Protection Agency (EPA) regulates six primary air pollutants: Ozone, Particulate matter, Carbon Monoxide, Nitrogen Oxides, Sulfur Dioxide, and Lead [3]. Particulate matter (PM) refers to solid particles and liquid droplets found in air. Many manmade and natural sources produce PM directly, or produce pollutants that react in the atmosphere to form PM [9]. $PM_{2.5}$ are small particles or particulate matter, that are less than 2.5 micrometers in diameter. $PM_{2.5}$ can be produced by combustion from motor vehicles (esp. diesel powered buses and trucks), power plants, residential wood burning, forest fires, agricultural burning, and industrial processes [15]. They can also be formed in the air through when gases (air poullutants) and orgnaic compounds are transformed through chemical reactions.

These tiny particles $PM_{2.5}$ can cause health hazards for people and also for the environment. People with heart diesease and lung problems including asthma, and also the elderly and children, are particularily vulnerable and at high risk when exposed to high levels of $PM_{2.5}$. Due to the tiny size of these $PM_{2.5}$ particles, they can penetrate to the deepest parts of the lungs, which is very dangerous to the human health. The California Air Resources Board (CARB) scentifically conducts studies and reports on the impacts of air pollutant exposures on public health, and the studies shows the negative impacts of $PM_{2.5}$ which is known to cause premature death [2]. Scientific studies has repeatedly found links between particulate matter and many health problems of people who has been exposed to high levels of $PM_{2.5}$, including asthma, bronchitis, respiratory problems, including shortness of breath and painful breathing, and premature deaths [3].

One must also be aware of the effects of high levels of $PM_{2.5}$ on the environment as well. Since $PM_{2.5}$ are tiny, they can be carried by wind and travel great distances, so that it can cause

problems for areas downwind far from the actual source of air pollution. They have adverse effect on urban areas, agriculture, and the natural envrinonment. High levels of $PM_{2.5}$ can results in visibility problems, urban haze, and acid rain [3].

The U.S. Environmental Protection Agency has established standards requiring the annual average of the $PM_{2.5}$ to be not more than 15 micrograms per cubic meter [3]. The State of California monitors and reports on their air pollutants carefully, setting very high standards for their air quality ($\mu g/m^3$). From 1999-2011, there are 113 station locations monitoring $PM_{2.5}$. The site design originally planned was well spread statistically. See Figure 1. However, in reality, it is too costly in terms of time, finance, and manpower to keep all the 113 sites to be monitoring and recording every single year. Each year, only a part of the 113 sites were actually sampled, and each year at different locations.

Figure 1. Complete 113 $PM_{2.5}$ Observational Sites in the California State

Comparisons of $PM_{2.5}$ between the years are difficult, due to "missing data" at sample sites [6, 9]. A site that does not have a recorded $PM_{2.5}$ value is referred to as "missing value", and since there are no patterns so that serious problems would twist the kriging map constructions.

Observing the dataset in Figure 2, the worst (in 1999) only 11 sites (9.73% of 113 sites) were used and at the best (in 2009) 65 sites (57.52% of 113 sites) were used. Over 13 years, 1469 annual arithmatic means ($\mu g/m^3$) should be recorded, but actually, 556 annual arithmatic means ($\mu g/m^3$) were reported, which occupied 37.85%. Sitewise looking, only one site, Site 2596 (Placer County APCD), was collected data annually and had 13 recored annual arithmatic means ($\mu g/m^3$), while 16 sites had one annual arithmatic mean ($\mu g/m^3$) only. The comparisons of $PM_{2.5}$ annual arithmatic means ($\mu g/m^3$) between years for a given site or between sites for a given year, i.e., the investigations of $PM_{2.5}$ annual arithmatic means ($\mu g/m^3$) patterns will be an

Figure 1. Complete 113 PM$_{2.5}$ Observational Sites in the California State

Comparisons of PM$_{2.5}$ between the years are difficult, due to "missing data" at sample sites [6, 9]. A site that does not have a recorded PM$_{2.5}$ value is referred to as "missing value", and since there are no patterns so that serious problems would twist the kriging map constructions.

Figure 2. PM$_{2.5}$ Samples Collected in California in 1999 (11 sites) and 2011 (52 sites)

extremely difficult task due to data incompleteness. Therefore it is logical to engage fuzzy theory for treating the "missing" or scarce data.

Observing the dataset in Figure 2, the worst (in 1999) only 11 sites (9.73% of 113 sites) were used and at the best (in 2009) 65 sites (57.52% of 113 sites) were used. Over 13 years, 1469 annual arithmatic means (μg/m³) should be recorded, but actually, 556 annual arithmatic means (μg/m³) were reported, which occupied 37.85%. Sitewise looking, only one site, Site 2596 (Placer County APCD),

2. Literature review and methodology

2.1. Fuzzy theory and membership kriging approach

Zadeh's fuzzy theory [22, 23] poineered a new mathematical branch. His membership approaches were quickly spreading and merging into many other mathematical branches, for example, engineering, business, economics, etc. and generating huge impacts in mathematical theories and applications. But it is aware that associated with Zadeh's fuzzy mathematical achivements, researchers gradually discovered three fundamental issues: self-duality dilemma, variable dilemma, and membership dilemma. Guo et al. [8] discussed those dilemmas in detail and pointed out Liu's credibility measure theory [11] is a solid mathematical treatment to address fuzzy phenomenon modelling. The credibility measure, similar to probability measure, assumes self-duality. Consequently, parallel to probability theory, a fuzzy variable and its (credibility) distribution can be defined. Furthermore, the membership function of a fuzzy variable can also be specified by its credibility distribution. Without any doubts, credibility measure theory is applying to practical situations sucessfully, say, Peng and Liu [14] considered parallel machine scheduling problems with processing times, Zheng and Liu [24] studied a fuzzy vehicle routing optimization problem, Guo et al. [2008] proposed credibility distribution grade kriging for investigation California State PM$_{10}$ spatial patterns, Wang et al. [20, 21] investigated a fuzzy inventory model without backordering, Sampath and Deepa [16] developed sampling plans containing fuzziness and randomness, and others. Neverthe-less, the investigations in statistical estimation and hypothesis testing problems with fuzzy

credibility distribution are very slowly progressed, for example, Li et al. [10], Sampath and Ramya [17, 18] in worked with fuzzy normal distributions, and studied the exponential credibility distribution function [1].

With statistically well-designed scheme, the collected spatial data can be analyzed and presented by kriging maps. If we are facing spatial data with "missing" or scarce fuzziness, it is impossible to construct kriging maps. It is noticed that the air pollutants were measured in different sites each year, even the site design originally planned was well spread statistically. We call a site that does not have a recorded $PM_{2.5}$ concentration as "missing value" site, as continued from Guo's research [4, 5]. To address the basic requirements in constructing kriging maps, Guo first proposed membership kriging in Zadeh's sense, see the MSc thesis [4] in which the linear, quadratic and hyperbolic tangent membership functions were used. Later Guo [5, 7] had developed the membership kriging under credibility theory. Following the membership kriging route of treating uncertainty, Shada et al. [19] and Zoraghein et al. [25] made considerable contributions in their papers. In this chapter, we will integrate exponential membership and kriging, to fill in the "missing data" based on existing sample data, and making a comparison of $PM_{2.5}$ concentrations in California from 1999-2011.

2.2. Fuzzy exponential distribution

It is necessary to introduce the basics of Liu's fuzzy credibility theory [11]. Let Θ be a nonempty set and M a σ- algebra over Θ. Elements of M are called events. Cr{A} denotes a number or grade associated with event A, called credibility (measure or grade). Credibility measure Cr{$\cdot$} satisfies the axioms normality, monotonicity, self-duality and maximality:

$$\begin{aligned}
&\text{Cr}\{\Theta\} = 1 \\
&\text{Cr}\{A\} \le \text{Cr}\{B\} \text{ for } A \subset B \\
&\text{Cr}\{A\} + \text{Cr}\{A^c\} = 1 \text{ for any event } A \\
&\text{Cr}\left\{\bigcup_i A_i\right\} = \sup_i\left(\text{Cr}\{A_i\}\right) \text{ for any events } A_i \text{ with } \sup_i\left(\text{Cr}\{A_i\}\right) < 0.5
\end{aligned} \tag{1}$$

Definition 1: The set function Cr{} on M is called a credibility measure if it follows Axiom Normality, Axiom Monotonicity, Axiom Self-Duality and Axiom Maximality shown in Equation 1. The triplet (Θ, M ,Cr) is called a credibility space.

Definition 2: A measurable function,η, mapping from a credibility space (Θ, M,Cr) to a set of real numbers $\hat{\omega}$ $\mathbb{B}? \subset \mathbb{R}$.

Definition 3: The credibility distribution Ψ of a fuzzy variable η on the credibility measure space (Θ, M,Cr) is $\Psi: \mathbb{R} \to [0,1]$, where $\Psi(x) = \text{Cr}\{\theta \in \Theta \mid \eta(\theta)\, x\}$, $x \in \mathbb{R}$.

Definition 4: The function μ of a fuzzy variable η on the credibility measure space (Θ, M,Cr) is called as a membership function: $\mu(x) = (2\text{Cr}\{\eta = x\}) \wedge 1, \quad x \in \mathbb{R}$.

Theorem 5: (Credibility Inversion Theorem) Let η be a fuzzy variable on the credibility measure space (Θ, M, Cr) with membership function μ. Then for any set B of real numbers,

$$\mathrm{Cr}\{\eta \in B\} = \frac{1}{2}\left(\sup_{x \in B} \mu(x) + 1 - \sup_{x \notin B} \mu(x)\right). \tag{2}$$

Corollary 6: Let η be a fuzzy variable on a credibility measure space (Θ, M, Cr) with membership function μ. Then the credibility distribution Ψ is

$$\Psi(x) = \frac{1}{2}\left(\sup_{y \le x} \mu(y) + 1 - \sup_{y > x} \mu(y)\right), \text{ for } \forall x \in \mathbb{i} . \tag{3}$$

It is obvious the concept of credibility measure is very abstract although the credibility measure has normality, monotonicity, self-duality and maximality mathematical properties. The credibility measure loses the intuitive feature as Zadeh's membership orientation. Credibility Inversion Theorem and its corollary have just revealed the deep link between an abstract measure and intuitive membership. Such a link definitely paves the way for real-life applications. For example, the trapezoidal fuzzy variable has membership function μ:

$$\mu(x) = \begin{cases} 0 & x \le c - c_2 \\ \dfrac{c_2 + x - c}{c_2 - c_1} & c - c_2 < x \le c - c_1 \\ 1 & c - c_1 < x \le c + c_1 \\ \dfrac{c_2 - x + c}{c_2 - c_1} & c - c_1 < x \le c + c_1 \\ 0 & c + c_2 < x \end{cases} \tag{4}$$

where c, c_1, c_2 are all positive200, $c > c_1, c > c_2, c_2 > c_1$.

With the help of Equation 3, the fuzzy credibility distribution is thus,

$$\Psi(x) = \begin{cases} 0 & x \le c - c_2 \\ \dfrac{c_2 + x - c}{2(c_2 - c_1)} & c - c_2 < x \le c - c_1 \\ \dfrac{1}{2} & c - c_1 < x \le c + c_1 \\ \dfrac{2c_2 - c_1 + x - c}{2(c_2 - c_1)} & c - c_1 < x \le c + c_1 \\ 1 & c + c_2 < x \end{cases} . \tag{5}$$

Liu [12, 13], Wang and Tian [21] and [1] studied the exponential fuzzy distribution with a membership function, denoted as Exp(m):

$$\mu\left(x\right) = \frac{2}{1 + e^{\frac{\pi x}{\sqrt{6m}}}}, x \geq 0, m > 0 \tag{6}$$

The support of an exponential membership function is set [0,+∞), the nonnegative part of the real line, ℝ. The expected value and second moment of exponential fuzzy variable are

$$\mathrm{E}[\eta] = \frac{\sqrt{6}\ln 2}{\pi} m \tag{7}$$

and

$$\mathrm{E}\left[\eta^2\right] = m^2 \tag{8}$$

where the parameter $m > 0$ determining the mean and variance of the exponential fuzzy variable. Thus it is intuitive to reveal how the shape of membership curve affected by the possible values of the parameter $m > 0$.

x	$m=12$	$m=15$	$m=22$	$m=50$
0	1.0000	1.0000	1.0000	1.0000
1	0.9466	0.9573	0.9709	0.9872
2	0.8935	0.9147	0.9418	0.9744
3	0.8410	0.8724	0.9128	0.9615
4	0.7894	0.8306	0.8839	0.9487
5	0.7390	0.7894	0.8553	0.9360
6	0.6899	0.7490	0.8269	0.9232
7	0.6424	0.7094	0.7987	0.9105
8	0.5968	0.6707	0.7709	0.8978
9	0.5530	0.6332	0.7435	0.8851
10	0.5113	0.5968	0.7165	0.8724
11	0.4717	0.5616	0.6899	0.8598
12	0.4342	0.5277	0.6638	0.8473
13	0.3990	0.4952	0.6382	0.8348
14	0.3660	0.4640	0.6132	0.8223
15	0.3351	0.4342	0.5887	0.8100
16	0.3063	0.4059	0.5647	0.7976
17	0.2796	0.3789	0.5414	0.7854

x	$m=12$	$m=15$	$m=22$	$m=50$
18	0.2549	0.3533	0.5187	0.7732
19	0.2320	0.3292	0.4966	0.7610
20	0.2110	0.3063	0.4752	0.7490
21	0.1917	0.2848	0.4544	0.7370
22	0.1739	0.2645	0.4342	0.7251
23	0.1577	0.2455	0.4147	0.7133
24	0.1428	0.2277	0.3959	0.7016
25	0.1293	0.2110	0.3777	0.6899
26	0.1170	0.1954	0.3602	0.6784
27	0.1057	0.1808	0.3433	0.6669
28	0.0955	0.1672	0.3270	0.6556
29	0.0863	0.1546	0.3114	0.6443
30	0.0779	0.1428	0.2964	0.6332

Table 1. Impacts of m in the Shape of Exponential Membership Curve

The value choice of parameter m is not aimless. m = 22 corresponds to the $PM_{2.5}$ annual arithematic mean 11.85 ($\mu g/m^3$), while m = 50 corresponds to the $PM_{2.5}$ annual arithematic mean 25.89 ($\mu g/m^3$). Therefore, the one-parameter exponential fuzzy variable may well cope to the modelling requirements of the California $PM_{2.5}$ annual arithematic mean dataset. Furthermore, using the one-parameter exponential fuzzy variable it may develop a delicate scheme of testing credibility hypothesis.

2.3. Credibility hypothesis testing with exponential membership

Similiar to the popular Neyman-Pearson Lemma in probability theory, likelihood ratio L_0/L_1, constant k, and critical region C of size α, are involved in the testing hypothesis H_0: $\theta = \theta_0$ against alternative hypothesis H_1: $\theta = \theta_1$. The likelihood is defined as the product of the densities for given sampled population. Hence we can call Neyman-Pearson testing criterion as likelihood ratio criterion. Inevitably, Type I error and Type II error concepts are also engaged in describing testing procedure. For hypothesis testing under credibility theory, Sampath and Ramya [17] proposed a membership ratio criterion. The membership criterion applies to any forms of credibility distributions, but exponential credibility distribution has its unique advantage. Therefore, the remaining descriptions will be focused on credibility hypothesis testing under exponential membership function [1].

Definition 7: Credibility hypothesis is a statement describing the possible rejection of a null hypothesis, H_0: $\mu = \mu_0$ with the credibility distribution of a fuzzy variable against an alternative hypothesis H_1: $\mu = \mu_1$ with another credibility distribution of a fuzzy variable.

Definition 8: Credibility hypothesis testing is the rule describing reject or not reject a null hypothesis if the calculated values sampled from the fuzzy distribution defined by null hypothesis.

Definition 9: Credibility rejection region is the subset of the support under a fuzzy distribution, denoted as C, on which the null credibility hypothesis is rejected H_0: $\mu = \mu_0$, i.e., $C = \{\eta \in \Theta \mid H_0 \text{ is rejected}\}$

Definition 10: Type I error is the mistake by rejecting the null credibility hypothesis H_0: $\mu = \mu_0$ when it is true and Type II error the mistake by not rejecting the null credibility hypothesis H_0: $\mu = \mu_0$ when it is false.

Definition 11: Level of credibility significance is the maximal credibility to make a Type I error in testing a credibility hypothesis H_0: $\mu = \mu_0$, denoted as α.

Definition 12: The best credibility rejection region of credibility significance level α, C^*, if this region possesses the maximal power (measured by credibility) under alternative hypothesis K with all possible credibility rejection regions of level of credibility significance α, i.e.,

$$\text{Cr}\{\eta \in C^* \mid K\} \geq \text{Cr}\{\eta \in C \mid K\}, \tag{9}$$

where C is any region satisfying the condition $\text{Cr}\{\eta \in C \mid H\} \leq \alpha$. The power of the credibility hypothesis testing is $\text{Cr}\{\eta \in C^*\}$.

With the exponential membership function having single parameter $m > 0$, the best credibility rejection region of credibility significance level α, C^* should be an interval so that we name it as best credibility rejection interval of credibility significance level α.

Theorem 13: For testing the null credibility hypothesis H_0: $m = m_1$ against the alternative credibility hypothesis K: $m = m_2$, ($m_1 < m_2$, under exponential fuzzy distributions $\mu(x) = \text{Exp}(m)$ as Equation 6 specified, the membership ratio criterion is engaged. The criterion states that given credibility significance level $\alpha < 0.5$, the best credibility rejection interval C^*:

$$C^* = \left[\frac{\sqrt{6} m_1}{\pi} \ln\left(\frac{1-\alpha}{\alpha} \right), +\infty \right), \ \alpha \in (0, 0.5). \tag{10}$$

The credibility of credibility rejection interval C^* under alternative hypothesis is greater than α.

α	0.10	0.15	0.20	0.25	0.30
x_0	37.56361	29.65463	18.78181	14.48535	10.58305

Table 2. x_0 and α under H_0: $m_1 = 21.926$

Table 2 illustrates relationship between the best credibility rejection interval boundary x_0 for selected credibility significance level $\alpha < 0.5$. For example, let $\alpha = 0.20$, $m_1 = 21.93$, then the best credibility rejection interval $C^* = [18.78, +\infty)$. We have to point out that the choice of credibility significance level α in credibility hypothesis testing should not follow the "thumb rule" the

significance level α = 0.05 in probability hypothesis testing. Although the two significance levels have the same role in hypothesis testing, nevertheless, the practical meanings of credibility significance level α and the significance level α are quite different. From Table 2 and California PM$_{2.5}$ distribution patterns, it is is logical and practical selecting the credibility significance level α = 0.25, which gives x_0 =14.485.

3. Analysis and results

3.1. Exponential fuzzy membership kriging

Now having examined the methodology, we can now calculate the membership grades. But first let us have a quick overlook on California PM$_{2.5}$ 1999-2011 records.

Year	Number of Sites	Annual Maximum	Annual Minimum	Annual Average
1999	11	23.4190	8.3274	13.6146
2000	22	23.9229	7.7971	14.3465
2001	25	25.0446	4.1492	12.5643
2002	41	25.8332	6.3356	15.9160
2003	44	24.8348	4.3802	12.6993
2004	43	18.1962	4.4146	11.3249
2005	40	22.4178	4.8117	12.0750
2006	47	21.5991	5.1022	11.4211
2007	55	25.2073	3.2883	11.6240
2008	59	21.2357	6.7716	12.3861
2009	65	21.2423	3.3373	10.7292
2010	52	17.1908	2.9665	9.2255
2011	52	18.0950	3.3682	10.7934
Total	556			

Table 3. California PM$_{2.5}$ 1999-2011 Annual PM$_{2.5}$ Concentrations
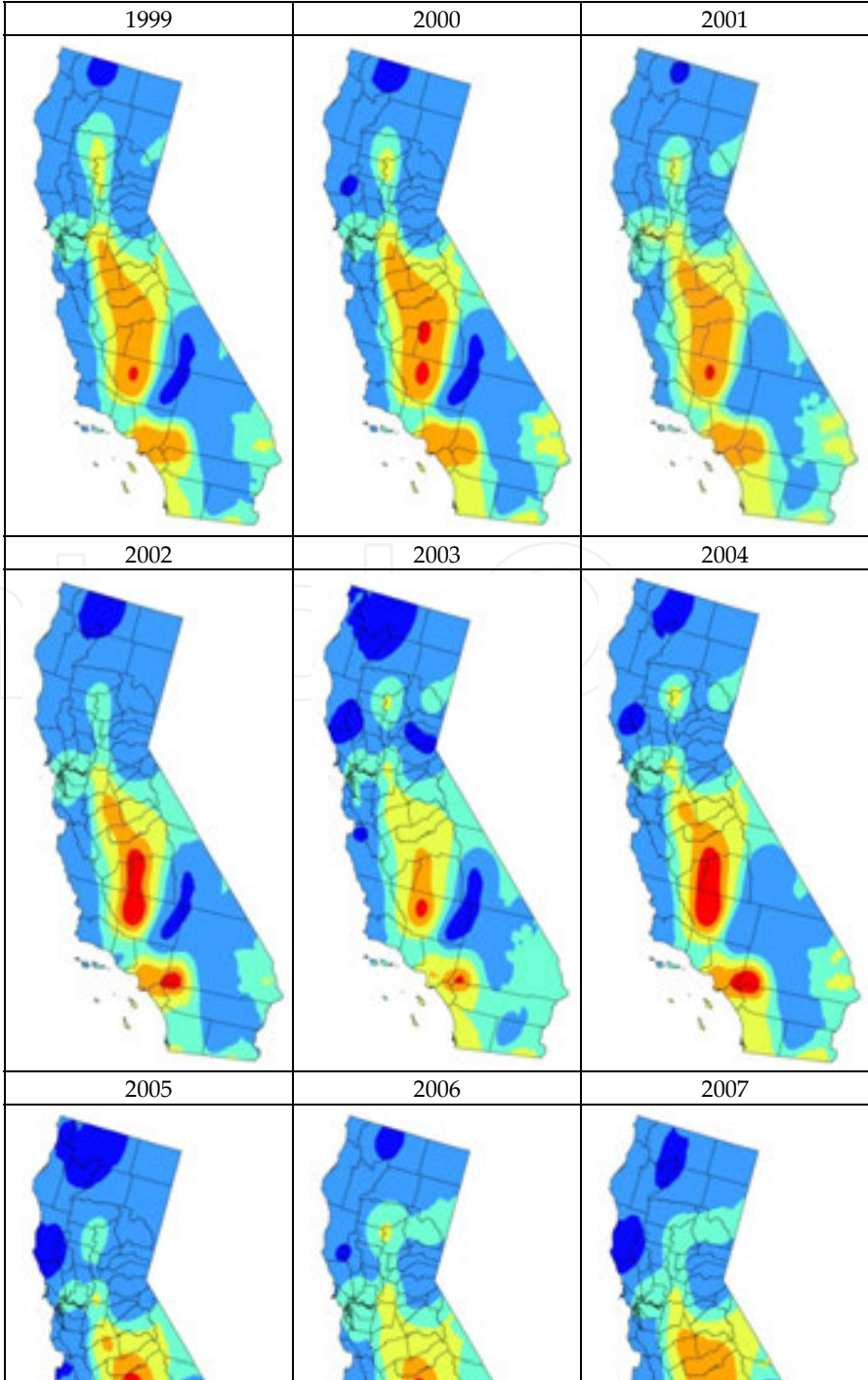
From Table 3, we can see that the annual average PM$_{2.5}$ range between 9.2255 and 15.9160. Most of the averages are wondering about 11.0 and 12.7. Therefore, it is very reasonable to estimate $\overline{X}$ =11.85.

Exponential membership grade kriging scheme:

1. Calculating overall exponential membership parameter m. Based on Equation 7, we can estimate the exponential membership parameter m. Then in this paper, m = 21.93 is used for membership grade smoothing, kriging, and hypothesis testing.

2. Calculating every exponential membership grade e_{ij} = 2.0/(1+exp($\pi x_{ij}/m\sqrt{6}$)) for site j given year i. For any site, if "missing" value, x_{ij} = 1, otherwise, x_{ij} = e_{ij} (> 0).

3. Performing membership grade interpolation. (i) Equal weights, for any given site j, if $x_{ij} = 1$, which is a "missing" value, let $x_{ij} \triangleq \hat{e}_{ij}$ where $\hat{e}_{ij} = (e_{i-1,j} + e_{i+1,j})/2$ (conditioning on the availability of the nearest neighbours $e_{i-1,j}$ and $e_{i+1,j} < 1$); (ii) Unequal weights, for a given site j, if the neighbour years are quite far, say, N between the gap, then, we may take linear interpolation for filling those "missing" value. that is given e_{ij} and $e_{i+N,j} < 1$, let $x_{i+l,j} \triangleq e_{ij} + l * (e_{i+N,j} - e_{ij})/N$, $l = 1, 2, ..., N$-1.

4. Performing membership grade extrapolation, including forward and backward extrapolation. Equal weights (1/3) are used mostly. Unequal weights are also used.

5. Carrying the filling "missing" cells task until thirteen years 1999 to 2011, each year 113 membership grades are all calculated, $\{\hat{e}_{ij}, i=1, 2, \cdots, 13, j=1, 2, \cdots, 113\}$.

Now, it is ready to construct exponential membership grades kriging maps and use these 13 maps for comparisons.
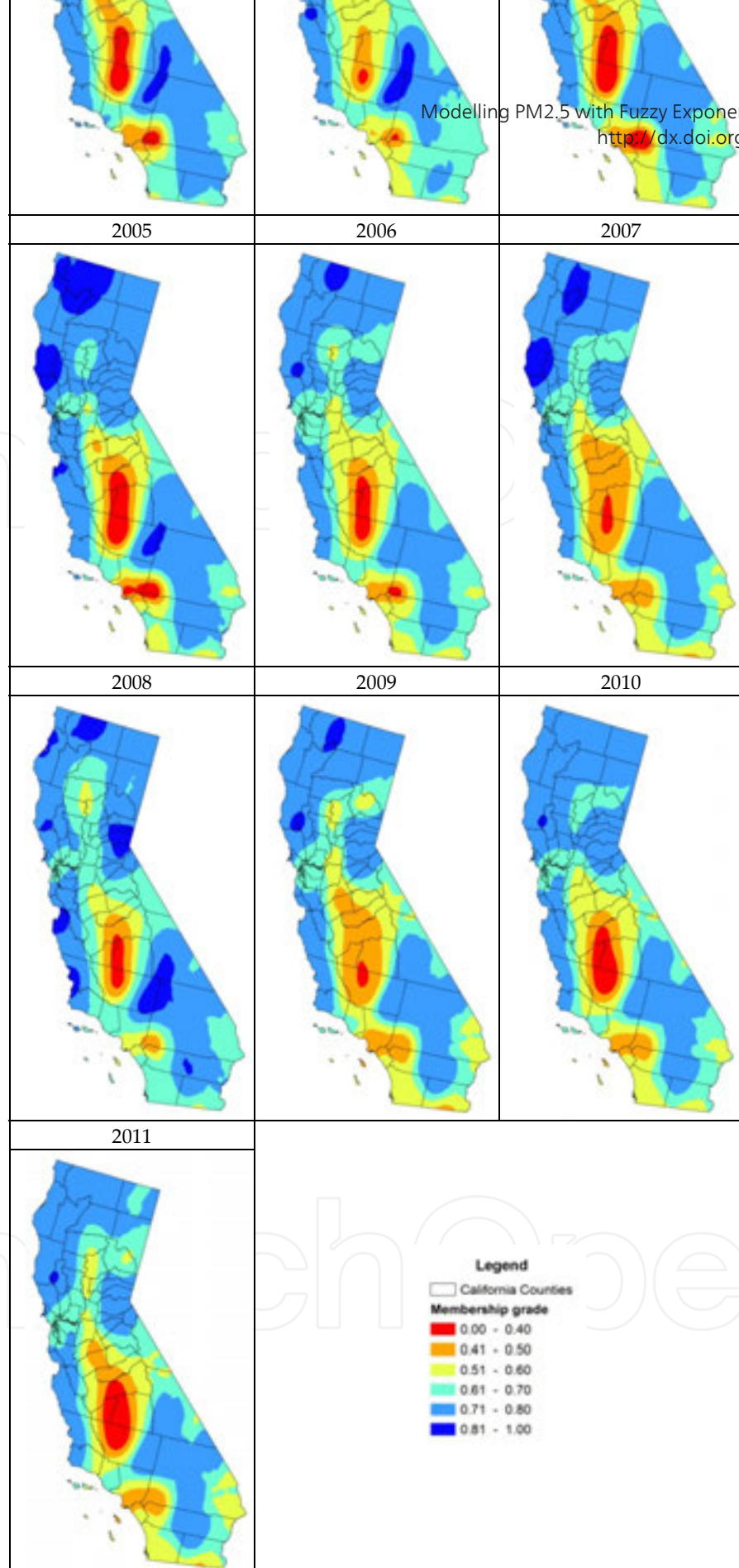
2005	2006	2007

2008	2009	2010

2011	

Legend
California Counties
Membership grade
- 0.00 - 0.40
- 0.41 - 0.50
- 0.51 - 0.60
- 0.61 - 0.70
- 0.71 - 0.80
- 0.81 - 1.00

Figure 3. Kriging maps of Exponential membership grades for California 1999-2011

Figure 3. Kriging maps of Exponential membership grades for California 1999-2011

As one can observe from Figure 3, central California regions have very low membership grades, and the rest of Calfornia have high membership grades.

As one can observe from Figure 3, central California regions have very low membership grades, and the rest of Calfornia have higher membership grades.

3.2. Calculated PM$_{2.5}$ concentrations

It is impossible for the public and governmental officers to accept the membership grade kriging maps. Therefore, kriging maps of every year PM$_{2.5}$ concentrations (collected and estimated together) must be constructed. The conversion formula is

$$\tilde{x}_{ij} = \frac{\sqrt{6}m}{\pi} \ln\left(\frac{2 - \hat{e}_{ij}}{\hat{e}_{ij}} \right). \tag{11}$$

Kriging maps with "completed" of every year PM$_{2.5}$ data $\{x_{ij}, i = 1,2,...,13, j = 1,2,...,113\}$. Comparison of kriging maps with the "completed" data can now be performed. See Figure 4. A complete dataset is now available, the 113 observation sites now have the full 13 year PM$_{2.5}$ concentration, containing 1469 data records. The 13 ordinary kriging prediction maps are generated, and one can clearly observe the changes in PM$_{2.5}$ year by year.
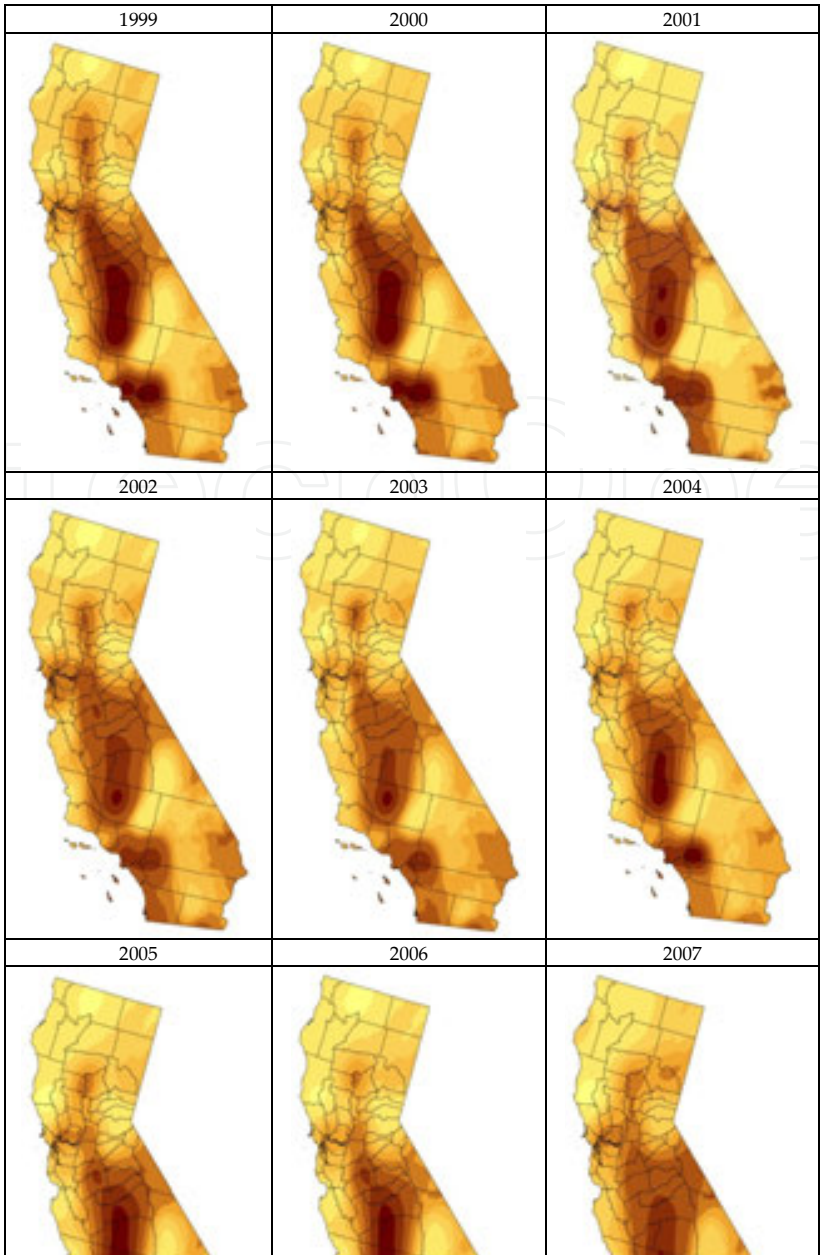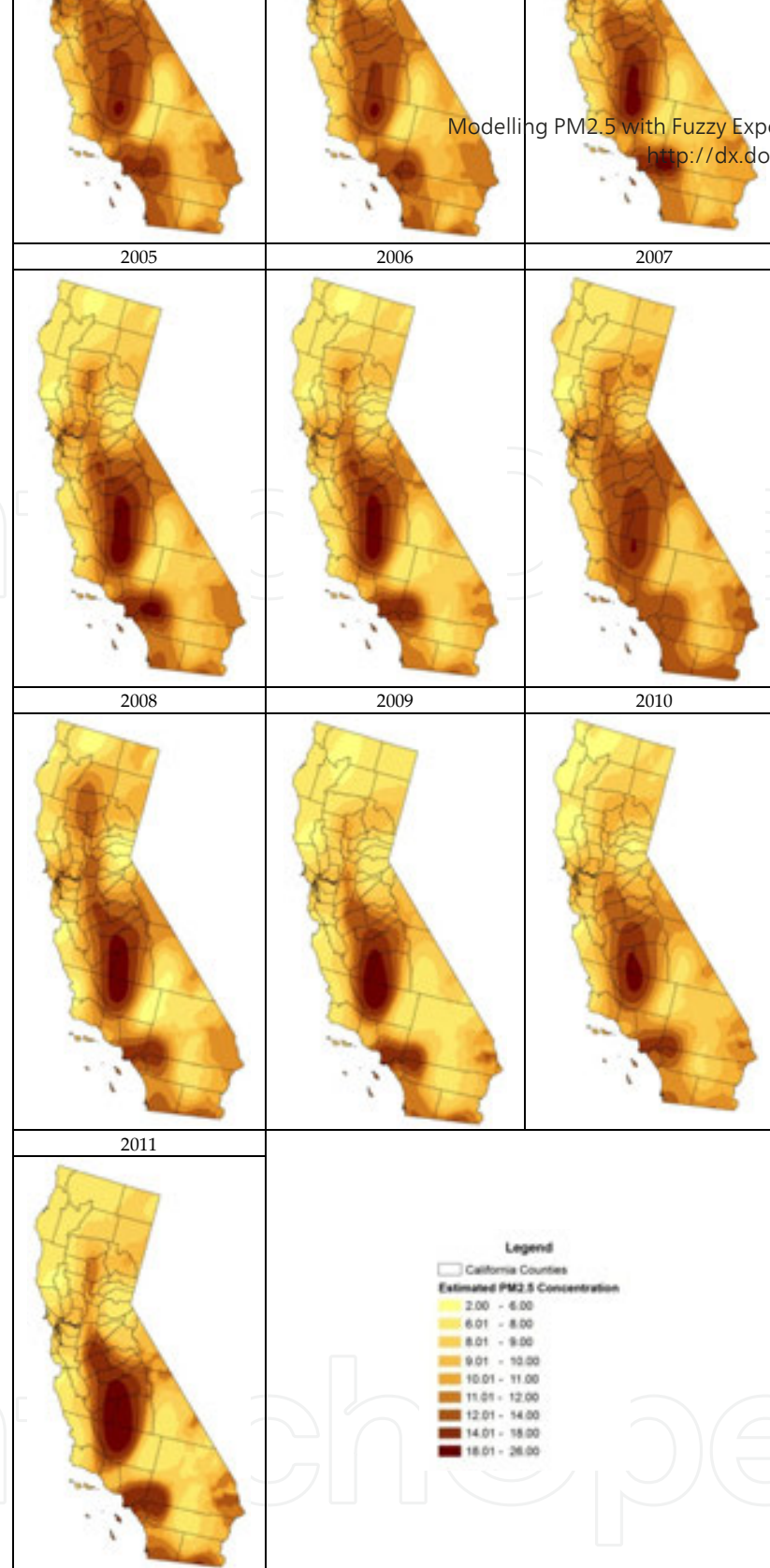
Figure 4. Kriging maps of completed estimated $PM_{2.5}$ concentrations for California 1999-2011

One can now observe and compare the year by year changes of $PM_{2.5}$ concentration, and note the regions of high and low $PM_{2.5}$ concentration. The dark brown colours represent high $PM_{2.5}$ concentrations, and light yellow colours represent areas with low $PM_{2.5}$ concentrations. Note that central California shows to have continual high levels of $PM_{2.5}$ concentration, year after year.

4. Interpretation and conclusion

In the results section, the dataset is now calculated and completed. However, it is now up to us to interpret the maps, and decide how to best make use of the calculated dataset, so that it provides us with easy to read information. And we can do this through a change map and 13 health safety maps.
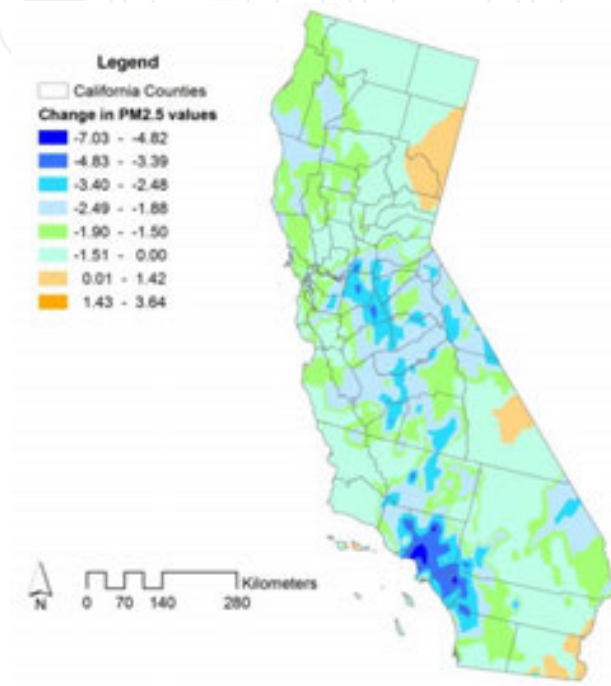

Figure 5. Changes in $PM_{2.5}$ Concentrations in California between 1999 and 2011

As one can clearly see from Figure 5, that $PM_{2.5}$ concentration has clearly decreased and air quality has improved remarkably over the years. The blue and green colours show negative changes, and orange shows positive changes. Counties such as Los Angeles and Orange show the highest decrease, and other counties such as Lassen, Plumas, Sierra, Inyo, and Imperial show some increase in $PM_{2.5}$ concentration. However, *a decrease in $PM_{2.5}$ concentration does not indicate safety in air quality*.

In terms of credibility hypothesis testing, say, with credibility significance level $\alpha = 0.25$, critical point for the best credibility rejection interval is $x_0 = 14.485$. The indictor λ_{ij} is defined as

$$\lambda_{ij} = \begin{cases} 1 & \tilde{x}_{ij} < 14.485 \\ 0 & \text{othewise} \end{cases}, \tag{12}$$

For comparisons of air quality safety, we generate $PM_{2.5}$ safety maps with two colours: blue colour if $\lambda_{ij} = 1$, orange colour otherwise, in total 13 safety maps. See Figure 6.
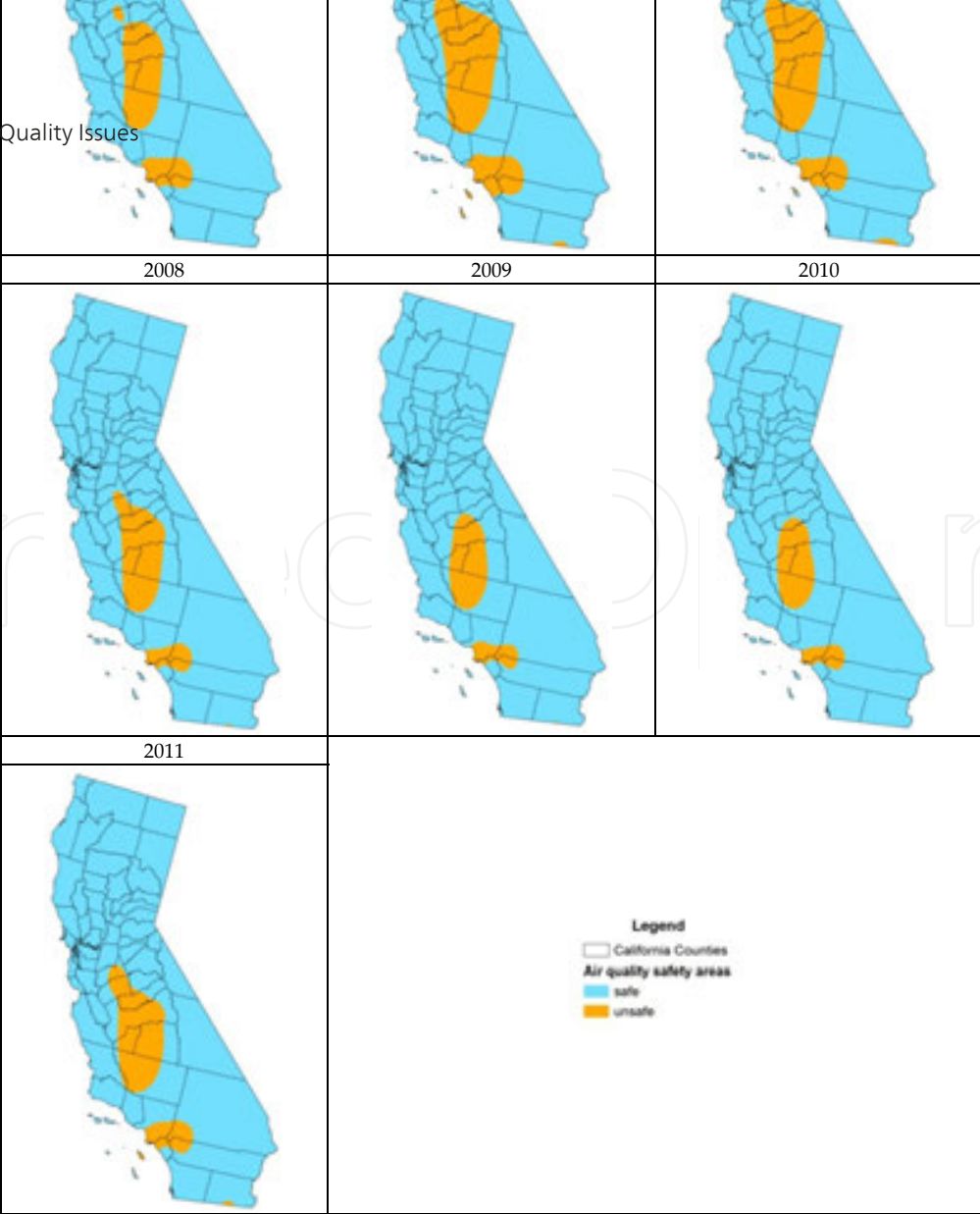
1999	2000	2001
2002	2003	2004
2005	2006	2007
2008	2009	2010
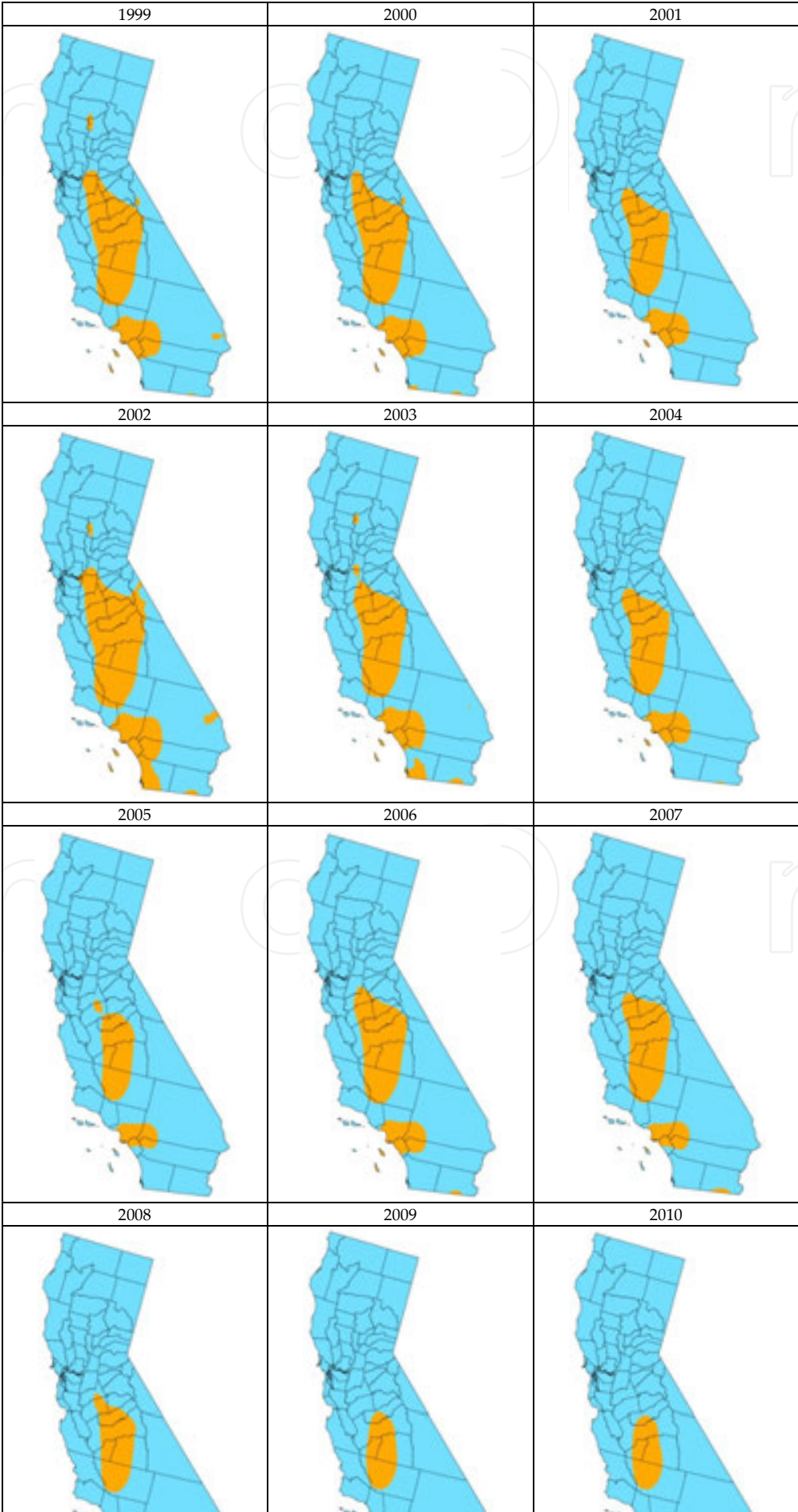
Figure 6. Air Quality PM$_{2.5}$ Safety areas in California 1999-2011

One can now observe that over the 13 years period 1999-2011, Stanislaus, Merced, Madera, Fresno, Kings, Tulare, Kern, Los Angeles, San Bernardino, Orange, and Riverside counties are the counties with the highest PM$_{2.5}$ safety problems. These areas have shown to be unsafe for public health safety, and especially for those with lung and heart problems, and for children and the elderly. These places are also sources of environmental and ecological concerns.

In conclusion, facing the difficult problem of "incomplete" PM$_{2.5}$ data in California from 1999-2011, we used the interpolation and extrapolation smoothing approaches for "filling" those "missing value" sites. For easy computation, the fuzzy exponential membership function is assumed. The treatment is based on an assumption that the smoothing is performed for a given site rather than over different sites for a given year. Such an assumption is emphasizing the fact: the data recorded are PM$_{2.5}$ concentration annual arithmetic means and they shouldn't change too dramatically over neighbour years. As to neighbour sites impacting, the member-ship grade kriging approach is adequate enough for generating smoothed maps. Furthermore, for utilizing credibility hypothesis testing theory, we perform parameter estimation of the fuzzy exponential membership function and in terms of membership ratio criterion for deriving the safety maps under 0.25 credibility significance level. Membership ratio criterion

is very similar to likelihood ratio criterion in theoretical development. By comparing those 13 PM$_{2.5}$ concentration safety maps to 1999-2011 change map, it is quite justifiable to say the safety maps under the credibility hypothesis testing procedure are very intuitive and convenient to the public. Finally, interpreting the 13 safety maps will provide the public with knowledge of air quality in California.

Acknowledgements

Author details

Danni Guo[*]

Address all correspondence to: d.guo@sanbi.org.za

Statistical Ecology Division, Biodiversity Research, Assessment and Monitoring, South African National Biodiversity Institute, Cape Town, South Africa

References

[1] Authors unknown. Credibility Hypothesis Testing of Fuzzy Exponential Distributions. International Journal of Fuzzy Computation and Modelling (IJFCM). (*accepted*)

[2] California Environmental Protection Agency Air Resources Board (CARB). Ambient Air Quality Standards (AAQS) for Particulate Matter. www.arb.ca.gov (accessed 3 July 2014)

[3] Environmental Protection Agency (EPA). National Ambient Air Quality Standards (NAAQS). U.S. Environmental Protection Agency. www.epa.gov (accessed 3 July 2014)

[4] Guo D. Integrating GIS with Fuzzy Logic and Geostatistics: Predicting Air Pollutant PM10 for California, Using Fuzzy Kriging. MSc thesis. University of Durham; 2003.

[5] Guo D. Contributions to Spatial Uncertainty Modelling in GIS: Small Sample Data. PhD Thesis. University of Cape Town; 2007.

[6] Guo D., Guo R., Cui YH., Midgley GF., Altwegg, R., Thiart, C. Climate Change Impact on Quiver Trees in Arid Namibia and South Africa. Blanco, Kheradmand (ed.)

Climate Change - Geophysical Foundations and Ecological Effects. InTech; 2011. p323-342.

[7]   Guo D, Guo R, Thiart C. Predicting Air Pollution Using Fuzzy Membership Grade Kriging. Computers, Environment and Urban Systems. 2007 31(1) 33-51.

[8]   Guo D, Guo R, Thiart C. Credibility Measure Based Membership Grade Kriging. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2007 15(2) 53-66.

[9]   Guo D., Guo R., Thiart C., Cui YH. Imprecise Uncertainty Modelling of Air Pollutant PM10. In: Nejadkoorki (ed.) Advanced Air Pollution. InTech; 2011. p193-212.

[10] Li X., Qin Z, Ralescu D. Credibilistic parameter estimation and its application in fuzzy portfolio selection. Iranian Journal of fuzzy systems. 2011 15(2) 57-65.

[11] Liu B. Uncertainty Theory: An Introduction to Its Axiomatic Foundations. Springer-Verlag Heidelberg, Berlin. 2004.

[12] Liu B. Uncertainty Theory. 3rd ed. Springer-Verlag Berlin. UTLAB; 2010.http:// orsc.edu.cn/liu/ut.pdf

[13] Liu B. Uncertainty Theory. 4th ed. Springer-Verlag Berlin. UTLAB; 2012. http:// orsc.edu.cn/liu/ut.pdf

[14] Peng J, Liu B. Parallel machine scheduling models with fuzzy processing times. Information Sciences. 2004 166(1-4) 49-66.

[15] Pittsburghtoday. Environment / PM2.5. Pittsburghtoday.org. www.pittsburghtoday.org/CurrentPM25US (accessed 3 July 2014)

[16] Sampath S, Deepa SP. Determination of Optimal Chance Double Sampling Plan using Genetic Algorithm. Model Assisted Statistics and Applications. IOS Press. 2013 8(4) 265-273.

[17] Sampath S, Ramya B. Credibility hypothesis testing of expectation of fuzzy normal distribution. Journal of Intelligent & Fuzzy Systems. 2014.

[18] Sampath S, Ramya B. Credibility hypothesis testing of fuzzy triangular distributions. Journal of Uncertain Systems. (*accepted*)

[19] Shada R, Mesgarib MS, Abkarc A, Shad A. Predicting air pollution using fuzzy genetic linear membership kriging in GIS. Computers, Environment and Urban Systems. 2009 33(6) 472-481.

[20] Wang X, Tang W, Zhao R. Fuzzy Economic Order Quantity Inventory Models without Backordering. Tsinghua Science and Technology. 2007 12(1) 91-96.

[21] Wang Z, Tian F. A Note of the Expected Value and Variance of Fuzzy Variables. International Journal of Nonlinear Science. 2010 9(4) 486-492.

[22] Zadeh LA. Fuzzy sets. Information and Control. 1965 8 338-353.

[23] Zadeh LA. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems. 1978 1 3-28.

[24] Zheng Y. Liu B. Fuzzy vehicle routing model with credibility measure and its hybrid intelligent algorithm. Applied Mathematics and Computation. 2006 176(2) 673-683.

[25] Zoraghein H, Alesheikh A A, Alimohammadi A, Vahidnia, M H. The utilization of soft transformation and genetic algorithm to model two sources of uncertainty of Indicator Kriging. Computers, Environment and Urban Systems 2012 36(6) 592-598.