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

3,800 Open access books available
116,000 International authors and editors
120M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

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
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 pollutants) and organic 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 disease and lung problems including asthma, and also the elderly and children, are particularly 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) scientifically 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 environment. 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 (μg/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 arithmetic means (μg/m$^3$) should be recorded, but actually, 556 annual arithmetic means (μg/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 recorded annual arithmetic means (μg/m$^3$), while 16 sites had one annual arithmetic mean (μg/m$^3$) only. The comparisons of PM$_{2.5}$ annual arithmetic means (μg/m$^3$) between years for a given site or between sites for a given year, i.e., the investigations of PM$_{2.5}$ annual arithmetic means (μg/m$^3$) patterns will be an
extremely difficult task due to data incompleteness. Therefore it is logical to engage fuzzy theory for treating the “missing” or scarce data.

2. Literature review and methodology

2.1. Fuzzy theory and membership kriging approach

Zadeh’s fuzzy theory [22, 23] pioneered 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 achievements, 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 successfully, 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. Nevertheless, the investigations in statistical estimation and hypothesis testing problems with fuzzy

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 (45 sites) and 2011 (52 sites)

Observing the dataset in Figure 2, the worst (in 1999) only 41 sites (9.73% of 413 sites) were used and at the best (in 2009) 65 sites (57.52% of 113 sites) were used. Over 13 years, 1469 annual arithmetic means ($\mu g/m^3$) should be recorded, but actually, 556 annual arithmetic means ($\mu g/m^3$) 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] pioneered 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 achievements, 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 successfully, 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. Nevertheless, 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 $\Theta$ be a nonempty set and $M$ a $\sigma$-algebra over $\Theta$. 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{align}
Cr\{\Theta\} &= 1 \\
Cr\{A\} &\leq Cr\{B\} \text{ for } A \subset B \\
Cr\{A\} + Cr\{A^c\} &= 1 \text{ for any event } A \\
Cr\bigg(\bigcup_{i} A_i\bigg) &= \sup\left\{Cr\{A_i\}\right\} \text{ for any events } A_i \text{ with } \sup\left\{Cr\{A_i\}\right\} < 0.5
\end{align}

\textbf{Definition 1:} The set function $Cr\{\cdot\}$ 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 $(\Theta, M, Cr)$ is called a credibility space.

\textbf{Definition 2:} A measurable function $\eta$, mapping from a credibility space $(\Theta, M, Cr)$ to a set of real numbers $\mathbb{R}$, is called a measurable function.

\textbf{Definition 3:} The credibility distribution $\Psi$ of a fuzzy variable $\eta$ on the credibility measure space $(\Theta, M, Cr)$ is $\Psi: \mathbb{R} \rightarrow [0,1]$, where $\Psi(x) = Cr\{\theta \in \Theta | \eta(\theta) \leq x\}$, $x \in \mathbb{R}$.

\textbf{Definition 4:} The function $\mu$ of a fuzzy variable $\eta$ on the credibility measure space $(\Theta, M, Cr)$ is called a membership function: $\mu(x) = (2Cr(\eta \leq x)) \wedge 1$, $x \in \mathbb{R}$.
Theorem 5: (Credibility Inversion Theorem) Let $\eta$ be a fuzzy variable on the credibility measure space $(\Theta, M, Cr)$ with membership function $\mu$. Then for any set $B$ of real numbers,

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

Corollary 6: Let $\eta$ be a fuzzy variable on a credibility measure space $(\Theta, M, Cr)$ with membership function $\mu$. Then the credibility distribution $\Psi$ is

$$\Psi(x) = \frac{1}{2} \left( \sup_{y \in x} \mu(y) + 1 - \sup_{y \in x} \mu(y) \right), \text{ for } \forall x \in \Theta. \hspace{1cm} (3)$$

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 $\mu$:

$$\mu(x) = \begin{cases} 0 & x \leq c - c_2 \\ \frac{c_2 + x - c}{c_2 - c_1} & c - c_2 < x \leq c - c_1 \\ 1 & c - c_1 < x \leq c + c_1 \\ \frac{c_2 - x + c}{c_2 - c_1} & c - c_1 < x \leq c + c_1 \\ 0 & c + c_2 < x \end{cases} \hspace{1cm} (4)$$

where $c, c_1, c_2$ are all positive $200, c > c_1, c > c_2, c_2 > c_1$.

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

$$\Psi(x) = \begin{cases} 0 & x \leq c - c_2 \\ \frac{c_2 + x - c}{2(c_2 - c_1)} & c - c_2 < x \leq c - c_1 \\ \frac{1}{2} & c - c_1 < x \leq c + c_1 \\ \frac{2c_2 - c_1 + x - c}{2(c_2 - c_1)} & c - c_1 < x \leq c + c_1 \\ 1 & c + c_2 < x \end{cases} \hspace{1cm} (5)$$
Liu [12, 13], Wang and Tian [21] and [1] studied the exponential fuzzy distribution with a membership function, denoted as \( \text{Exp}(m) \):

\[
\mu(x) = \frac{2}{1 + e^{-mx}} , x \geq 0, m > 0
\]

(6)

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

\[
E[\eta] = \frac{\sqrt{6} \ln 2}{\pi} m
\]

(7)

and

\[
E[\eta^2] = m^2
\]

(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 \( m > 0 \).

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

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

The value choice of parameter $m$ is not aimless. $m = 22$ corresponds to the PM$_{2.5}$ annual arithmetic mean $11.85$ (μg/m$^3$), while $m = 50$ corresponds to the PM$_{2.5}$ annual arithmetic mean $25.89$ (μg/m$^3$). Therefore, the one-parameter exponential fuzzy variable may well cope to the modelling requirements of the California PM$_{2.5}$ annual arithmetic 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

Similar to the popular Neyman-Pearson Lemma in probability theory, likelihood ratio $L_0/L_1$, constant $k$, and critical region $C$ of size $\alpha$, 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 \( \alpha \).

Definition 12: The best credibility rejection region of credibility significance level \( \alpha \), \( C^* \), 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 \( \alpha \), i.e.,

\[
\text{Cr}\{\eta \in C^* \mid K\} \geq \text{Cr}\{\eta \in C \mid K\},
\]

where \( C \) is any region satisfying the condition \( \text{Cr}\{\eta \in C \mid H_0\} \leq \alpha \). 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 \( \alpha \), \( C \), should be an interval so that we name it as best credibility rejection interval of credibility significance level \( \alpha \).

Theorem 13: For testing the null credibility hypothesis \( H_0: \mu = m_1 \) against the alternative credibility hypothesis \( K: \mu = m_2 \) \( (m_1 < m_2) \) 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 \( C^* \):

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

The credibility of credibility rejection interval \( C^* \) under alternative hypothesis is greater than \( \alpha \).

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>0.10</th>
<th>0.15</th>
<th>0.20</th>
<th>0.25</th>
<th>0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>37.56361</td>
<td>29.65463</td>
<td>18.78181</td>
<td>14.48535</td>
<td>10.58305</td>
</tr>
</tbody>
</table>

Table 2. \( x_0 \) and \( \alpha \) 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 \( \alpha \) in credibility hypothesis testing should not follow the “thumb rule” the
significance level \( \alpha = 0.05 \) in probability hypothesis testing. Although the two significance levels have the same role in hypothesis testing, nevertheless, the practical meanings of credibility significance level \( \alpha \) and the significance level \( \alpha \) are quite different. From Table 2 and California PM\(_{2.5}\) distribution patterns, it is logical and practical selecting the credibility significance level \( \alpha = 0.25 \), 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.

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

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 \( \bar{X} \) = 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 $x_{ij} \equiv \hat{x}_{ij}$ where $\hat{x}_{ij} = (e_{i-1,j} + e_{i+1,j}) / 2$ (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 $e_i$ and $e_{i+N} < 1$, let $x_{i+l,j} \equiv e_{ij} + l \times (e_{i+N,j} - e_{ij}) / N$, $l = 1, 2, ..., N-1$.

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, $\{\hat{e}_{ij}, i = 1, 2, ..., 13, j = 1, 2, ..., 113\}$.

Now, it is ready to construct exponential membership grades kriging maps and use these 13 maps for comparisons.
As one can observe from Figure 3, central California regions have very low membership grades, and the rest of California have higher membership grades.
As one can observe from Figure 3, central California regions have very low membership grades, and the rest of California 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

$$
\hat{x}_y = \frac{\sqrt{6m}}{\pi} \ln \left( \frac{2 - e_y}{e_y} \right).
$$

(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.
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.

Figure 4. Kriging maps of completed estimated PM$_{2.5}$ concentrations for California 1999-2011
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](image)

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 indicator $\lambda_{ij}$ is defined as

$$
\lambda_{ij} = \begin{cases} 
1 & \text{if } \bar{x}_i < 14.485 \\
0 & \text{otherwise}
\end{cases}
$$

(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.
<table>
<thead>
<tr>
<th>Year</th>
<th>Year</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>2000</td>
<td>2001</td>
</tr>
<tr>
<td>2002</td>
<td>2003</td>
<td>2004</td>
</tr>
<tr>
<td>2005</td>
<td>2006</td>
<td>2007</td>
</tr>
<tr>
<td>2008</td>
<td>2009</td>
<td>2010</td>
</tr>
</tbody>
</table>
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 membership 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

I would like to thank the California Air Resources Board for providing the air quality data used in this paper. This study is supported financially by the National Research Foundation of South Africa (Ref. No. IFR2009090800013).

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


