A Comprehensive Model for Analyzing the Opioid Crisis
Zhi-chen ZENG*, Qing-yang DAI and Run-ze LIN
College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
*Corresponding author

Keywords: Opioid crisis, Artificial Potential Field, ARIMA model, Particle Swarm Optimization.

Abstract. The trend of opioids abuse is spreading across the United States, causing a serious crisis in society. In this paper, we construct a mathematical model to forecast the opioid abuse in five states in the United States and propose several precautions. Based on the description of opioid abuse obtained through Artificial Potential Field method, we apply ARIMA model for forecast. To improve the model, we establish a measurement involving socio-economic factors through Analytic Hierarchy Process and optimize relevant parameters by Particle Swarm Optimization. Based on the model, we propose warning mechanism and several policies dealing with opioid abuse. Evaluation shows that proposed policies are effective, and the comprehensive model is beneficial to the management of opioid crisis.

Introduction
The United States is experiencing a nationwide crisis due to the abuse of opioids. Cross-sectional studies in the United States conducted by the Substance Abuse and Mental Health Services Administrations (SAMHSA) show a tripling in the rate of initiation on nonmedical prescription opioid use among young adults since 1990 and concomitant increases in attributable emergency abuse treatment admissions[1]. Studies have also shown increases over the past decade in the number of older adults seeking treatment for marijuana, cocaine, heroin, and other opioid[2]. The reason for opioid crisis is a combination of sociocultural factors, such as depressed economy, low employment, lack of education, and a high rate of prescribing[3].

In order to analyze the opioid abuse crisis, a mathematical model is built up based on opioid abuse data and census data provided by federal, state, and local forensic laboratories of five states (OH, KY, WV, VA and PA) in recent years, which helps describe and forecast the spread and distribution of opioid abuse. Relevant policies are proposed thereafter, and their effects are carefully evaluated.

Analysis of Opioid Abuse Distribution and Transmission
This section focuses on describing the distribution and transmission of opioid abuse. As shown in Fig. 1, the model uses Artificial Potential Field (APF) method to characterize the transmission of opioid abuse. Analytical Hierarchy Process (AHP) is applied to form a comprehensive indicator, and K-Means clustering based on the comprehensive indicator is applied, through which data points could be classified into three categories, and sources and end points of the potential field can be defined.

![Figure 1. Modeling scheme overview of drug spread.](image-url)
Generation of Artificial Potential Field

The main idea of APF method is to discover a field function representing the energy of the system, and the object in the field is applied a force either repulsive or attractive, minimizing the energy value of the entire system with the object[4].

APF basically consists of two components: the attractive potential ($U_{att}$) and the repulsive potential ($U_{rep}$), as shown in Eq. 1.

$$U_{att}(x) = \begin{cases} 
K_a |x - x_d|^2, & |x - x_d| \leq d_a \\
K_a (2d_a |x - x_d| - d_a^2)^2, & |x - x_d| > d_a
\end{cases}$$

$$U_{rep}(x) = \begin{cases} 
\frac{1}{2} K_r \left( \frac{1}{\rho} - \frac{1}{\rho_0} \right)^2, & \rho \leq \rho_0 \\
0, & \rho > \rho_0
\end{cases}$$

where $K_a$ is the attractive gain, $K_r$ is the repulsive gain, $x$ is the evaluated point, $x_d$ is the target point, $d_a$ is the distance threshold. $\rho = |x - x_d|$ is the distance between the evaluated point, and $\rho_0$ represents the predefined distance threshold.

By calculating the derivative of $U_{att}(x)$ and $U_{rep}(x)$, the attractive force and repulsive force could be represented as $F_{att}(x) = \frac{\partial U_{att}(x)}{\partial x}$ and $F_{rep}(x) = \frac{\partial U_{rep}(x)}{\partial x}$ respectively.

Afterwards, the resultant force can be calculated and the motion direction of the object can be determined by combining two coordinate forces, as shown in Fig. 2.

AHP Based K-Means Clustering & Results

To form the APF, $U_{att}$ and $U_{rep}$ are required. K-Means clustering helps to divide counties into three categories: sources, which provide repulsive forces; end points, which provide attractive force and unreachable areas, which do not affect the artificial potential field distribution. To better cluster the points, AHP[5] is applied to form a comprehensive indicator, which is a weighted linear combination of factors, such as Opioid Reports (OR), Total Opioid Reports of Counties (TORC) and Total Opioid Reports of States (TORS). Through AHP, weight vector is obtained as $W = [0.7854, 0.1488, 0.0658]$. Thus, the comprehensive indicator could be established as Eq. 2.

$$d(x, c_i) = \sqrt{\sum_{j=1}^{\text{len}(W)} W_j(x)(x_j^2 - c_{ij}^2)}.$$  

Further, K-Means clustering based on $d(x, c_i)$ is applied, and results are shown in Fig.3. Results show that PA and OH are sources; VA is the end point; while WV and KY are unreachable areas.

Spread and Distribution Description

Given that both repulsive and attractive forces of APF have been defined, artificial potential field space model of opioid distribution could be constructed. Take the result of synthetic opioids (SO) in year 2014 as an example, result is shown in Fig. 4, where the arrow represents the direction of the potential field gradient at the point, which is also the trend of the opioids spread.
Trend Forecasting & Warning Mechanism

This section focuses on describing the prediction of opioid abuse and establishing corresponding warning mechanism. Based on clustering results obtained in the previous section, two warning levels could be obtained through iterative threshold method. Afterwards, Gaussian Distribution Weighted Method (GDWM) is proposed, so that influencing factors of spatial neighboring points could be put into consideration. Moreover, by applying Auto-Regressive Integrated Moving Average model (ARIMA), the developing trend of SO and NSO cases can be predicted. Interpolating discrete-time sequence, the trend curve of opioid crime situation could be obtained, and the warning mechanism, which requires the prediction of time and location of opioid abuse outbreak, could be established.

Figure 4. Spread of SO in 2014.

Figure 5. Threshold iterating scheme.

Determination of Concerning Thresholds

To form specific concerns, thresholds for concerns need to be defined. Concerns are divided into two types, correspond to two thresholds to be determined. Fig. 5 shows the threshold iterating scheme, and the process could be describe as follow:

Take class 1 and class 2 for instance, firstly, select an initial threshold $T_1$, and calculate the mean value $\mu_1, \mu_2$ of class 1 and class 2 based on $T_1$; then, obtain the new threshold $T_2 = (\mu_1 + \mu_2)/2$. This process iteratively continues until $|T_2 - T_1| \leq \varepsilon$.

The two thresholds are determined to be 1462.3 (lower threshold) and 10118.6 (higher threshold).

Gaussian Distribution Weighted Method (GDWM)

Based on the fact that opioid transmission is closely related to the spatial distribution of different areas, spatial contribution from adjacent areas to the centroids should be concerned. From this aspect, GDWM is innovatively proposed.

Several cluster points are selected as geographical center points through K-Means clustering (Fig.6). The influential weights of adjacent counties decrease with the distance from corresponding geographical center points according to two-dimensional Gaussian distribution (Fig. 7), which means that closer counties have greater influence on corresponding center points. Therefore, spatial influence is included and the prediction model combining time-spatial factors could be constructed.

Figure 6. Result of clustering (100 points).

Figure 7. Gaussian distribution.
ARIMA Prediction for Warning Mechanism

ARIMA is a time series analysis model synthesized by Auto-Regressive model (AR), Integrated process (I) and Moving Average model (MA). It is designed to examine sequentially lagged relationships for relationships that may not be apparent in data collected periodically[6]. AR model quantifies relationship between current data and previous data, MA model solves the problem of random variables and I process deals with the nonstationary feature of time series.

For the p-order AR model, its autoregressive performance is to establish a linear regression model between its pre-p data \( \{Y_{t-p} \cdots Y_{t-1}\} \) and the current data \( Y_t \), which is expressed in Eq. 3.

For the q-order MA model, the moving average is expressed as a linear regression model between the white noise \( \{u_{t-q} \cdots u_t\} \) and the current time point data \( X_t \) of the previous q time points, which is expressed in Eq. 3[7].

\[
Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t, \quad X_t = u_t + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \cdots + \phi_q u_{t-q}.
\]

According to concerning thresholds obtained previously, counties may be divided into three classes, which are the source of transmission, endpoint of transmission and unreachable area. Four standards are established for warning mechanism, which can be referred to Table 1. Based on ARIMA model, example prediction result is shown in Fig. 8 and Fig. 9.

Figure 8. ARIMA prediction (Example result). Figure 9. Warning Mechanism of 2018.

<table>
<thead>
<tr>
<th>Standards</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Notice</td>
<td>upper confidence limit meets lower threshold</td>
</tr>
<tr>
<td>Serious Notice</td>
<td>predict value meets lower threshold</td>
</tr>
<tr>
<td>Common Warning</td>
<td>upper confidence limit meets higher threshold</td>
</tr>
<tr>
<td>Serious Warning</td>
<td>predict value meets higher threshold</td>
</tr>
</tbody>
</table>

Modified Model Involving Social-economic Factors

This section studies the influence of social-economic factors and provides policies to reduce opioid abuse. Based on census data, principal components are selected as evaluation index based on Principal Components Analysis (PCA)[8]. Afterwards, a linear combination of principal components is established as Social-Economic Evaluation Index (SEEI), and Particle Swarm Optimization method (PSO) [9] is applied to obtain the best weight \( W_i \) for the linear combination of principal components \( F_i \). Moreover, Comprehensive Evaluation Index (CEI) is established as a linear combination of OR and SEEI:

\[
CEI = a \cdot DR + b \cdot SEEI,
\]

which comprehensively considers the degree of opioid abuse and social-economic factors. And the optimal weights \( a, b \) for CEI could be calculated through Ordinary Least Square (OLS) method.

Principal Components Selection

Attributes of census data are divided into four categories \( C_i \). \( C_1 \) are relationship issues, \( C_2 \) are education issues, \( C_3 \) are society issues, \( C_4 \) are family history issues. Afterwards, census data is normalized, and PCA is performed, through which principal components \( F_i \) of different classes \( C_i \)
are obtained as \( F_i = \sum_j U_{ij}P_{ij} \). \( P_{ij} \) is corresponding factor property and \( U_{ij} = A_{ij}/\sqrt{\lambda_i} \), where \( \lambda_i \) is the principal component eigenvalue and \( A_{ij} \) is the factor load matrix.

**Particle Swarm Optimization**

SEEI, an index reflecting social-economic factors, is constructed as a linear combination of principal components \{\( F_1, F_2, F_3, F_4 \)\} with weight vector \( \omega = [\omega_1, \omega_2, \omega_3, \omega_4] \). For every county, its SEEI per year could be described as \( \text{SEEI} = [\omega_1 \omega_2 \omega_3 \omega_4][C_1 C_2 C_3 C_4]^T \).

To describe the relationship between \( \text{SEEI} \) and \( \text{OR} \), \( \text{SEEIM} \) and \( \text{ORM} \) are formed as Eq. 4.

\[
\text{ORM} = \begin{bmatrix}
  o_{r1} & \cdots & o_{rm}
\end{bmatrix}_{n \times m} \quad \text{SEEIM} = \begin{bmatrix}
  s_{s1} & \cdots & s_{sm}
\end{bmatrix}_{n \times m}
\]

\( dr_{ij} \) and \( ss_{ij} \) represent the value of \( \text{OR} \) and \( \text{SEEI} \) of county \( i \) in year \( j \) respectively. For better description, the aggregation of \( \text{ORs} \) and \( \text{SEEIs} \) is formed as a three-dimensional data cube, and weight is formed as a filter, as shown in Fig. 10, and \( \text{SEEIM} \) is the convolution of the two elements.

The optimal weight \( [\omega_1, \omega_2, \omega_3, \omega_4] \) best represents the absolute correlation between \( \text{SEEI} \) and \( \text{OR} \), or, minimizes the distance between \( \text{ORM} \) and \( \text{SEEIM} \). Thus, PSO could be a suitable technique for solution, and the optimal goal is \( \min u = f(\omega_1, \omega_2, \omega_3, \omega_4) = \|\text{ORM} - \text{SEEIM}\|_2 \).

The Scheme of PSO is shown in Fig. 11. Through PSO, the optimal weight vector is obtained and \( \text{SEEI} \) is established as \( \text{SEEI} = [-0.97585, 0.934164, 0.474319, -0.56624][C_1 C_2 C_3 C_4]^T \).

** Establishment of Comprehensive Evaluating Index **

To put both \( \text{SEEI} \) and \( \text{OR} \) into consideration, CEI is constructed as \( \text{CEI} = a \cdot \text{OR} + b \cdot \text{SEEI} \). OLS is applied to find optimal parameters. Independent variable \( X \), dependent variable \( Y \), weight parameter \( P \) and the residual matrix \( E \) are shwon in Eq. 5.

\[
X = \begin{bmatrix}
  \text{OR}_1 & \text{SEEI}_1 \\
  \text{OR}_2 & \text{SEEI}_2 \\
  \vdots & \vdots \\
  \text{OR}_n & \text{SEEI}_n
\end{bmatrix},
Y = \begin{bmatrix}
  \text{OR}_2 \\
  \text{OR}_3 \\
  \vdots \\
  \text{OR}_{n+1}
\end{bmatrix}, P = \begin{bmatrix}
  a \\
  b
\end{bmatrix}, E = \begin{bmatrix}
  e_1 \\
  e_2 \\
  \vdots \\
  e_n
\end{bmatrix}
\]

(5)

The goal of OLS is to find matrix \( P \) which minimizes the norm of the residual matrix \( L(P) = \|X P - Y\|_2 \), which could be achieved through calculating the derivative of \( L(P) \), that is \( \frac{\partial L(P)}{\partial P} = 0 \).

Corresponding solution is \( P = (X^T X)^{-1} X^T Y \), thus, \( \text{CEI} = 1.0154 \cdot \text{OR} - 0.0106 \cdot \text{SEEI} \). So far, Relationship, Education, Status and Family history are included to better construct our model.

**Strategy Formulation and Evaluation**

**Strategy Formulation**

Proposed policies mainly include four aspects as follows:
Suppress the birthplace of opioid trafficking;
Strengthen community construction and improve people’s livelihood;
Strengthen education, improve national quality of people and raise their awareness of opioid;
Strengthen customs control over imports and exports.

**Policy Scoring Mechanism Based on PMC Index**

*PMC index*[10] is an efficient tool to quantitatively evaluate a policy. To form *PMC index*, \( \omega_{ij} \in \{-3,-2,-1,0,1,2,3\} \) is used to describe the influential extent of how a policy will affect the \( j \)-th factor of category \( C_i \). The absolute value of \( \omega_{ij} \) reflects the strength of effect, where negative means inhibitory effect and positive means promoting effect. *PMC index* of a policy on \( C_i \) is \( S_i = \sum_j U_{ij} \omega_{ij} \), where \( U_{ij} \) is previously obtained through PCA. Normalizing PMC Indexes to the interval \([0.8,1.2]\), Impact Factors could be obtained. Sheme of policy scoring mechanism is shown in Fig. 12.

![Figure 12. Scheme of policy scoring mechanism.](image_url)

**Iterative Prediction Model**

To better represent indicators, an attribute vector \( Atr_{i,k} = [OR_{i,k} \ R_{i,k} \ E_{i,k} \ S_{i,k} \ F_{i,k}]^T \) is constructed to describe the condition of the \( i \)-th county in \( k \)-th year. Thus, Iterative prediction model could be applied for prediction.

\[
Atr_{i,k+1} = I \times Atr_{i,k}
\]

Assuming that changes of impact of policies on related factors are ignored in short-term forecasts, value of attributes could be updated annually by \( Atr_{i,k+1} = I \times Atr_{i,k} \), where \( I = \text{diag}[I_{OR} \ I_{R} \ I_{E} \ I_{S} \ I_{F}] \) is the matrix of Impact Factors. Further, according to \( CEI_{i,k} = [\alpha \omega_1 + \beta \omega_2 + \gamma \omega_3 + \delta \omega_4]Atr_{i,k} \), \( CEI \) under certain policies could be predicted. Example result is shown in Fig. 13.

Based on the results, Drug Index show a downward trend, while the *SEEI* show an upward trend after the implementation of policies, which indicates that policies proposed will be efficient.
Conclusions

This paper analyzes the cases of opioid crimes in five states, and obtains the opioid transmission model based on \textit{APF} and the space-time prediction model based on \textit{ARIMA} model. In addition, taking impact of socio-economic factors on opioid transmission into account, the relationship between socio-economic factors and the number of opioid cases are analyzed, and the space-time prediction model is improved, with more reasonable results obtained. Finally, on the basis of summarizing the previous models, policies are proposed to directly inhibit opioid abuse and indirectly affect social factors. Based on the \textit{PMC} index, consequences of direct and indirect suppression policies are evaluated, and iterative prediction method is utilized to verify the effectiveness of corresponding policies.

References


