Analysis of lung cancer morbidity and mortality based on particle swarm optimization

Naiwei Li¹,², Ziyi Cheng³, Youngmin Woo⁴, Jongmyung Ha¹ and Andre Kim¹

¹Department of pharmaceutical Engineering College of medical and Life Science, Silla University, Busan 46958, South Korea
²The Food and Biological Engineering, Wenjing College, Yantai University, Shandong Yantai 264005, China
³Key Laboratory of Emergency and Trauma, Ministry of Education, College of Emergency and Trauma, Hainan Medical University, Haikou 571199, China
⁴Natural Science Institute, Silla University, Busan 46958, South Korea

Corresponding author and e-mail: Andre Kim, adrlkim@163.com

Abstract. Based on the particle swarm algorithm to establish a mathematical model, this article counts the air quality index, tobacco sales, the number of cancer hospitals and the number of health personnel from 2008 to 2019, and quickly calculates the morbidity and mortality of lung cancer over the years, and uses statistics The relative error between the value and the calculated value is the evaluation standard. The result proves that the relative error between the calculated value and the statistical value of the morbidity and mortality of lung cancer is less than 10%, indicating that the calculation result of the particle swarm algorithm has good confidence.

1. Introduction
Lung cancer is one of the diseases with the fastest increase in morbidity and mortality and the greatest threat to people’s health and life. In the past 50 years, many countries have reported that the morbidity and mortality of lung cancer have increased significantly. The morbidity and mortality of lung cancer in men rank first among all malignant tumors, and women rank second. The etiology of lung cancer is related to many factors. A large amount of data shows that long-term heavy smoking leads to a significant increase in the incidence of lung cancer. Severe environmental particulate matter (PM2.5) can also induce an increase in the rate of lung cancer. Wang Guoping calculated and analyzed the number of lung cancer incidence and death cases in Wenjiang District of Chengdu, crude morbidity and mortality, China’s standardized morbidity and mortality, and world’s standardized morbidity and mortality. The results showed that men and middle-aged and elderly people, the incidence and death of lung cancer accounted for the highest [1]. Liao Jiang et al. conducted an analysis and research on the morbidity and mortality of malignant tumors and lung cancer in Chengdu, using the death-cause statistical analysis software Deathreg2002 for statistical analysis, and the results showed that the mortality rate of urban males was significantly higher than that of females, and rural females had the lowest mortality; The mortality rate in urban industrial areas is significantly higher than that in commercial and cultural areas, and the mortality rate of lung cancer increases with age [2]. Li Hongbao
and others analyzed the incidence and death data of Shanghai in the past 10 years and showed that with the increase of high-risk factors for cancer, the incidence of lung cancer is on the rise, and the age of onset tends to be younger, which seriously endangers the health of the population [3]. Lun Wenhui and others used the Joinpoint model to calculate the annual percentage change of lung cancer incidence and lung cancer mortality in the Guangzhou urban residents model, analyzed the change trend, and predicted lung cancer mortality from 2016 to 2025, providing a basis for Guangzhou’s lung cancer prevention and treatment strategy [4]. The morbidity and mortality of lung cancer increase with age, and reach their peak in the 80-84 years old. In recent years, some new evolutionary paradigms (such as swarm intelligence optimization, artificial immune system, distribution estimation algorithm, co-evolutionary algorithm, etc.) have been introduced into the field of multi-objective optimization and become an effective way to solve multi-objective optimization problems. As one of the typical optimization algorithms, particle swarm optimization uses velocity and position formulas to update the direction of particles. Due to its advantages such as easy understanding and implementation, it has received extensive attention in the fields of science and engineering.

2. Principles and algorithms

Particle Swarm Optimization (PSO) was originally derived from simulating the foraging behavior of birds and was proposed by Kenney and Eberhart in 1995 [5]. Each particle searches for the optimal solution according to its own flight direction and speed. At the same time, there is a strong competitive relationship between the particles, so each particle has its individual optimal value recorded. At the same time, the particles cooperate with each other to determine the global optimal value of the next iteration through communication, thus speeding up the algorithm convergence. Because the algorithm is easy to understand and has fewer adjustment parameters, it has become one of the fastest growing intelligent optimization algorithms and has been successfully applied in engineering practice, data mining, neural network training and other fields. Chang et al. used the concepts of individual optimal and global optimal in particle swarm optimization and multiple crossover operators to update the population, and proposed a novel hybrid multi-objective evolutionary algorithm [6].

Suppose that in a D-dimensional target search space, there is a particle community with a population size of $N$, that is, there are $N$ particles in the population, and the position of the $i$ particle is represented as a D-dimensional vector $X_i=(x_{i1}, x_{i2}, \ldots, x_{iD})(i=1,2,\ldots,N)$, its flight speed can be expressed as $V_i=(v_{i1}, v_{i2}, \ldots, v_{iD})(i=1,2,\ldots,N)$. The position of each particle represents a feasible solution to a problem in the target search space. At the beginning, the population is initialized as a set of random solutions, that is, randomly distributed throughout the search space. When the algorithm is executed, the state of the particle is updated by the following two formulas:

$$v_{id} = w v_{id} + c_1 r_{rand}(p_{id} - x_{id}) + c_2 r_{rand}(g_{id} - x_{id}) \quad (1)$$
$$x_{id} = x_{id} + v_{id} \quad (2)$$

In the above equation, $v_{id}$ represents the velocity of the $i$ particle in the $d$-dimension, and $X_{id}$ represents the position of the $i$ particle in the $d$-dimension. $W$ represents inertia weight. The larger the value is, the stronger the particle's exploration ability is, that is, the stronger the global search ability is. The smaller the particle is, the stronger the development ability is, i.e. the stronger the local search ability is. $c_1$ and $c_2$ are learning factors, which is also known as the acceleration constant. $c_1$ represents individual learning factor, $c_2$ represents social learning factor. $r_{rand1}$ and $r_{rand2}$ are independent random Uniform Numbers between $[0,1]$. $p_{id}$ represents the location of the optimal solution searched so far by the $i$ particle individual, and $g_{id}$ represents the location of the optimal solution searched so far by the entire population. The right side of the formula (1) can be divided into three parts. The first part is the inertia of the particles, which means that the particles have a tendency to maintain the flight speed of the previous generation; the second part is the individual's optimal position to guide the particles to move, reflecting the particles' progress toward their own experience Learning represents the tendency of particles to fly to the optimal position in their history; the third part is that the optimal particle of the population guides each particle to move, which reflects the information exchange between the populations and represents the tendency of the particle to fly to the optimal particle of the population.
In order to avoid the excessive flight range of a particle, it is necessary to set a maximum flight speed \( V_{d_{\text{max}}} \) for each dimension of the particle. When the velocity of the particle in the \( d \)-dimension is greater than or less than \( V_{d_{\text{max}}} \), the velocity of the particle in the \( d \)-dimension is set as \( V_{d_{\text{max}}} \) or \(-V_{d_{\text{max}}} \). The value of \( V_{d_{\text{max}}} \) is generally 10\%-20\% of the value range of decision variables on this dimension. The algorithm flow chart is shown in the figure below:

![Algorithm flow chart](image_url)

**Figure 1.** Algorithm flow chart.

### 3. Modeling and discussion

#### 3.1. Air quality Index

Tseng et al. reported that over 50\% of lung cancer patients in Taiwan have never smoked, and the change of PM2.5 level affects the morbidity of lung adenocarcinoma and the survival rate of patients [7]. In this paper, the annual average PM2.5, PM10, SO2, CO2, NO2 and other data related to air quality in China from 2008 to 2019 are statistically analyzed (see Table 1). PM2.5 and PM10 are selected to build an array as the search space of two dimensions in particle swarm optimization.

| Year | AQI | Range      | Quality grade      | PM2.5 | PM10 | SO2 | CO  | NO2 |
|------|-----|------------|--------------------|-------|------|-----|-----|-----|
| 2008 | 175 | 72–285     | Moderate pollution | 136   | 209  | 32  | 1.607 | 76  |
| 2009 | 222 | 74–462     | Heavy pollution    | 179   | 256  | 32  | 1.587 | 73  |
| 2010 | 149 | 43–234     | Light pollution    | 113   | 150  | 25  | 1.421 | 49  |
| 2011 | 114 | 40–211     | Light pollution    | 84    | 137  | 24  | 1.297 | 64  |
| 2012 | 90  | 43–160     | Good               | 60    | 100  | 16  | 1.037 | 51  |
| 2013 | 110 | 60–185     | Light pollution    | 76    | 131  | 15  | 0.819 | 52  |
| 2014 | 74  | 36–166     | Good               | 44    | 72   | 13  | 0.863 | 46  |
| 2015 | 75  | 27–122     | Good               | 40    | 71   | 13  | 0.868 | 47  |
3.2. **Tobacco sales**

The World Health Organization clearly stated that cigarettes are one of the important factors in the incidence of lung cancer. In 1890, cigarettes began to enter the Chinese commodity market, and factories were set up in 1893 to manufacture, and production and sales increased year by year. Since 1980, China's cigarette production has ranked first in the world, with 159.75 billion cigarettes produced in 1989. The third dimension space and the fourth dimension space in the particle swarm algorithm used in this paper represent China's cigarette production and sales volume in the past 12 years.

**Table 2.** Statistics of cigarette production and sales in China from 2008 to 2019.

| Year | Output (billion cigarettes) | Growth (%) | Cumulative sales value (billion cigarettes) | Growth (%) |
|------|-----------------------------|------------|-------------------------------------------|------------|
| 2008 | 2001.73                     | 7.5        | 4871.63                                   | 3.7        |
| 2009 | 2204.15                     | 10.1       | 5072.77                                   | 4.1        |
| 2010 | 2375.26                     | 7.8        | 5243.04                                   | 3.4        |
| 2011 | 2447.40                     | 3          | 5381.59                                   | 2.6        |
| 2012 | 2516.09                     | 2.8        | 5505.26                                   | 2.3        |
| 2013 | 2560.40                     | 1.8        | 5611.65                                   | 1.9        |
| 2014 | 2609.85                     | 1.9        | 5699.18                                   | 1.6        |
| 2015 | 2589.06                     | -0.8       | 5772.97                                   | 1.3        |
| 2016 | 2382.58                     | -8         | 5841.68                                   | 1.2        |
| 2017 | 2345.07                     | -1.6       | 5906.14                                   | 1.1        |
| 2018 | 2411.57                     | 2.8        | 5971.33                                   | 1.1        |
| 2019 | 2437.47                     | 1.1        | 6033.41                                   | 1          |

3.3. **Number of cancer hospitals and health personnel**

The number of Chinese cancer hospitals and the number of related health personnel will affect the morbidity and mortality of lung cancer to a certain extent, especially the mortality rate is highly correlated with medical resources. In theory, the more hospitals and employees there are, the lower the cancer mortality rate. But at the same time, the quality of practitioners and changes in treatment options will also affect the ultimate lung cancer mortality rate. This paper counts the number of cancer hospitals and the number of related practitioners from 2008 to 2019, as two different dimensional spaces of particle swarm algorithm.

**Table 3.** The number of cancer hospitals and health personnel in China from 2008 to 2019.

| Year | Number of Hospitals (homes) | Growth (%) | Health Number (person) | Growth (%) |
|------|------------------------------|------------|------------------------|------------|
| 2008 | 112                          | 2.22       | 58211                  | 4.52       |
| 2009 | 113                          | 0.89       | 61295                  | 5.30       |
| 2010 | 115                          | 1.77       | 63479                  | 3.56       |
2011 | 118 | 2.61 | 65263 | 2.81
2012 | 124 | 5.08 | 66647 | 2.12
2013 | 124 | 0   | 67531 | 1.33
2014 | 128 | 3.23 | 68257 | 1.08
2015 | 129 | 0.78 | 67215 | 1.53
2016 | 132 | 2.33 | 68825 | 2.40
2017 | 134 | 1.52 | 68667 | -0.23
2018 | 136 | 1.49 | 68951 | 0.41
2019 | 138 | 1.47 | 69338 | 0.56

### 3.4. Morbidity and mortality of Lung cancer

The PM2.5, PM10, cigarette production, cigarette sales, the number of cancer hospitals, and the number of health professionals in the past 12 years in China are integrated into a 6-dimensional target search space with a population size of 12, and a matrix is established as an input vector. The statistical data of lung cancer morbidity and mortality is selected as the theoretical value of the particle optimal solution, the iteration termination condition is set to 10% of the relative error of the optimal solution, and the particle algorithm iteration is performed through a 6×12 discrete matrix. It can be seen from Figure 2 that the morbidity and mortality of lung cancer show a downward trend as a whole. This is relatively related to air quality and medical conditions, but there is no obvious linear relationship with cigarette production and sales.

![Figure 2. Morbidity and mortality of Lung cancer from 2008 to 2019. (a) morbidity rate; (b) mortality rate.](image-url)

The PSO algorithm is used to establish a model to evaluate the calculation results, and the relative error between the statistical value and the calculated value is used as the evaluation standard (see Figure 3). The relative error between the calculated and statistical values of lung cancer morbidity and mortality is less than 10%, indicating that the results calculated by the particle swarm algorithm have better confidence. However, the relative error of more than 50% also exceeds 5%, indicating that the accuracy of the calculation results through the particle swarm algorithm is not enough, especially in the lung cancer incidence data from 2009 to 2012, and the lung cancer mortality data in 2010 and 2016. At the same time, this is also the data that has just passed the extreme point in the statistical data, which shows
that the particle swarm algorithm is used to calculate the morbidity and mortality of lung cancer, there will be "extreme convergence", which is also the need to improve in the iterative calculation section.

![Figure 3. Comparison of statistical and calculated values and their relative errors.](image)

4. Conclusions

Based on the typical mathematical model of particle swarm optimization algorithm, this paper quickly calculated the morbidity and mortality of lung cancer over the years through the statistical data of air quality index, tobacco sales, number of tumor hospitals and number of health personnel, etc. affecting the morbidity and mortality of lung cancer from 2008 to 2019. The results show that the simulated data are in good agreement with the measured data, and the relative error is less than 10%.

References

[1] Wang Guoping, Zeng Shouqun, He Ping et al. Investigation and Analysis of Morbidity and mortality of Lung Cancer in Wenjiang District, Chengdu City in 2008-2010[J]. Practical Preventive Medicine.2013,20(11):1314-1315.

[2] Liao Jiang, Du Qiyun, Jia Yong et al. Analysis of morbidity and mortality of malignant tumor and lung cancer in Chengdu[J]. Chinese Journal of Prevention and Control of Chronic Diseases.2006,14(4):286-287.

[3] Li Hongbao. Analysis of lung cancer incidence and mortality in Putuo District, Shanghai, 2002-2004[J]. Shanghai Journal of Preventive Medicine.2006,18(2):60-61.

[4] Lun Wen-hui, Li Tie-gang, MAO Chun et al. Trend analysis and prediction of lung cancer mortality and incidence in urban areas of Guangzhou[J]. Chinese Journal of Oncology Prevention and Treatment.2020,12(1):51-56.

[5] J. Kennedy, R. C. Eberhart. Particle swarm optimization[A]. Proc of the First IEEE International Conference on Neural Networks[C]. Perth, Australia: IEEE Press,1995:1942-1948.

[6] Tang L, Wang X. A Hybrid Multi objective Evolutionary Algorithm for Multi objective Optimization Problems[J]. IEEE Transactions on Evolutionary Computation,2013,17(1):20-45.

[7] Chien-Hua Tseng, MD, Ben-Jei Tsuang et al. The Relationship Between Air Pollution and Lung Cancer in Nonsmokers in Taiwan[J]. ORIGINAL ARTICLE.