For office use only	Team Control Number	For office use only
T1 _____	**1902081**	F1 _____
T2 _____		F2 _____
T3 _____	Problem Chosen	F3 _____
T4 _____	**C**	F4 _____

2019
MCM/ICM
Summary Sheet

Title

Summary

Keywords: keyword1; keyword2

Contents

1 Introduction

2 CA Model

To consider the relationship between adjacent counties, CA Model is introduced. Due to the specificity of this problem, we will modify the traditional CA Model to satisfy the limitations of this problem.

2.1 Properties

CA Model is consisted of multiple cells with state and the relartionship of adjacence. Here we present several properties regarding our revised CA Model.

- In our revised CA Model, a cell represents some county in the five states.
- The state of a cell denotes the number of predicted drug reports of some county.
- Any two cells are adjecent in CA Model if and only if the two counties they respectively represent are geographically adjecnt.
- In a new stage, some cell's state is determined by this cell's state and its neighbours' state in the previous stage.

Our revised CA Model differs from traditional CA Model in several aspects.

1. **Number of states**.
   As the state of cell is equivalent to the number of some county's drug reports, the cell in our revised CA Model has infinite kinds of state. However, the cell in traditional CA Model are restricted to finite kinds of state.
2. **Number of neighbours**
   Traditional CA model has to restrict the number of neighbours to be 4 or 8. In our CA Model, we shed the constraint by using arrays to store the neighbours, which implies that we can have theoretically infinite neighbours.

2.2 Symbols

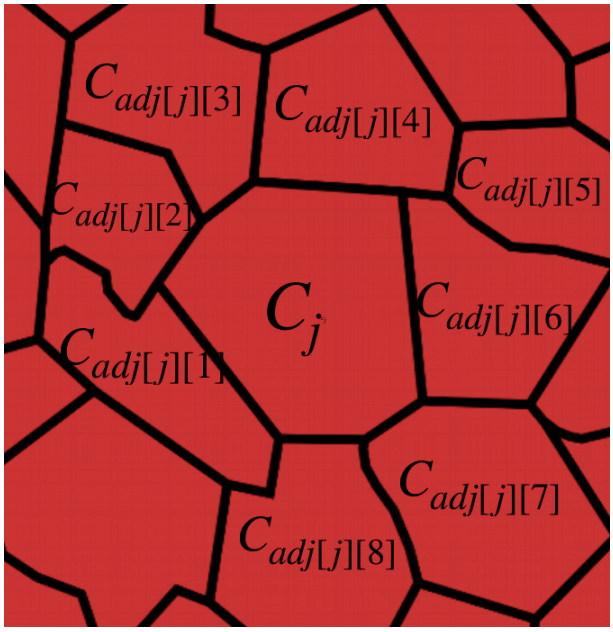

Figure 1: neighbours

Table 1: Main Symbols used in CA Model

Symbols	Definition
c_j	the j_{th} cell
s_{ij}	the j_{th} cell's state in stage i
s'_{ij}	the j_{th} cell's state in stage i (only for backstepping)
r_{ij}	the j_{th} cell's real data in stage i
n_j	the number of the j_{th} cell's neighbours
$adj[j]$	the array which stores the j_{th} cell's neighbours We can use $adj[j][k]$ to index the neighbours of the cell, where $k = 1, ..., n_j$
N	the number of counties in the five States
λ	a constant which describes the degree of adjecnt counties's contact

2.3 Prediction Rules

Any cell's state in the $(i+1)_{th}$ stage is determined by its state and its neighbours' state in the i_{th} stage. In other words, $s_{(i+1)j}$ is determined by s_{ij} and s_{ik} where c_k is a neighour of c_j.

With careful consideration, We choose the relationship to be linear. The Prediction formula is as follows.

$$s_{(i+1)j} = k_1 s_{ij} + k_2 \lambda \frac{\sum_{k=1}^{n_j} s_{i(adj[j][k])}}{n_j}$$

where k_1 denotes the influence of its own state, k_2 denotes the infulence of its neigh-

bours' state, λ is a constant which describes the degree of adjecnt counties's contact. After multiple tests and the consideration of reality, we choose λ to be 0.3.

This Prediction formula takes both the cell's influence and its neighbours' influence into consideraion.

2.4 Backstepping Rules

We also need to estimate the data in years before 2010 which is not provided. We simply rewrite the formulation above and make some approximation. To distinguish Prediction formula, we use s'_{ij} to replace s_{ij}. The formula is as follows:

$$s'_{ij} = \frac{s'_{(i+1)j}}{k_1} - \frac{k_2}{k_1}\lambda\frac{\sum_{k=1}^{n_j} s'_{(i+1)adj[j][k]}}{n_j}$$

where k_1 , k_2 and λ are identical to those in Prediction Rules.

2.5 Determine coefficients

To determine the coefficients k_1 and k_2, we use program to help us choose between 0 and 1 automatically. We have data about drug reports in years 2010-2017. The program will choose between 0 and 1 by step 0.001 to minimize the error. The formula to calculate error is as follows.

$$error = \frac{\sum_{i=2011}^{2017}\sum_{j=1}^{N}(r_{ij} - s_{ij})^2}{N * (2017 - 2011 + 1)}$$

where N is the number of counties in the five state.

Given $\lambda = 0.3$, we find the error is minimal when $k_1 = 0.999$ and $k_2 = 0.186$. Based on this, we can conclude that the cell's own influence is of most significance, but the influence of adjacent cells cannot be ignored.

2.6 Appliance of the CA Model

We have estimated the number of drug reports of every county in years from 2000 to 2030 using our revised CA model. Therefore we can estimate the data of every state by summing the data of its counties. Fig.2 shows the estimated data and real data of VA. And Fig.3 shows the estimated data and real data of PA. CA Model fits well with the State PA, where two lines are very close. As to the State VA, two lines have similar trends.

Using CA Model, we can get data of counties, which means we can get geographical distribution of data. This may provide insight into the data.

In Fig.4, there are six counties with estimated numbers greater than 1000 in the year 2000, namely Delware(OH) , Jefferson(KY), Montgomery(OH), Allegheny(PA), Hamilton(OH), Philadelphia(PA). In Fig.5, there are thirteen counties with estimated numbers

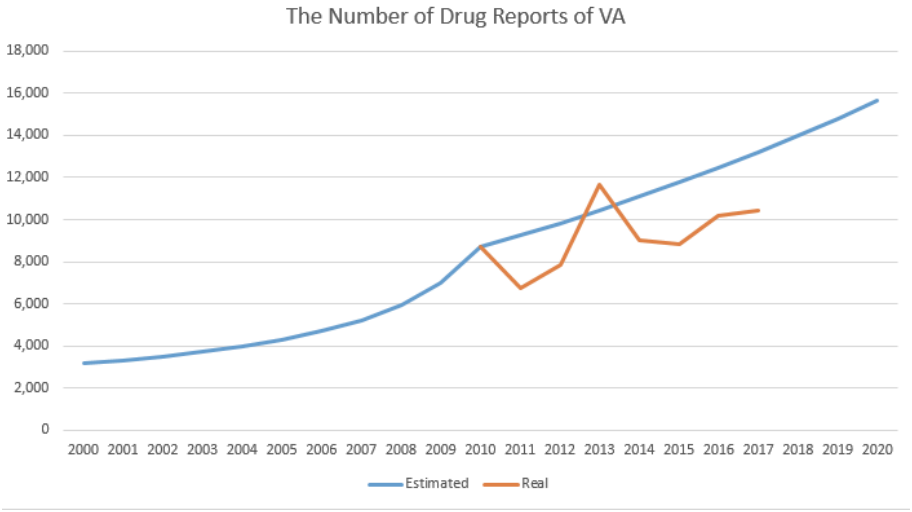

Figure 2: The number of drug reports of VA

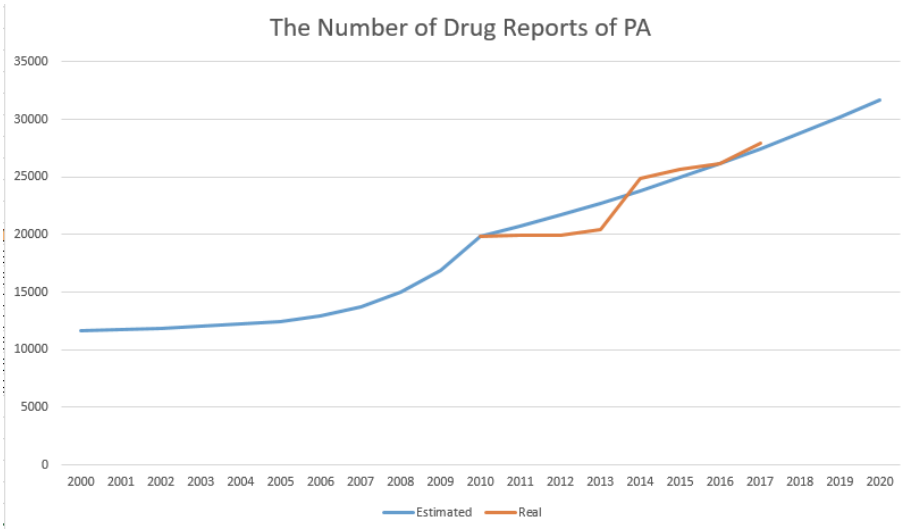

Figure 3: The number of drug reports of PA

greater than 1000 in the year 2020. We can notice that the growth in number spread from the red regions in Fig.4 Therefore we may conclude that the drug abuse might have started form the red regions in Fig.4.

2.7 Sensitivity Analysis

As the constant λ in CA Model may be hard to obtain or there might be some uncertainty, the choice of λ might affect the result of our CA Model. So in order to test the robustness of our revised CA Model , a sensitivity analysis is conducted by testing our CA Model with various λ.

Here we calculate the error by the formula in Section 2.5. We originally choose $\lambda = 0.3$. To test the robustness of our CA Model, we set $\lambda = 0.5$ and 0.8 respectively. The tests showed that our model is robust.

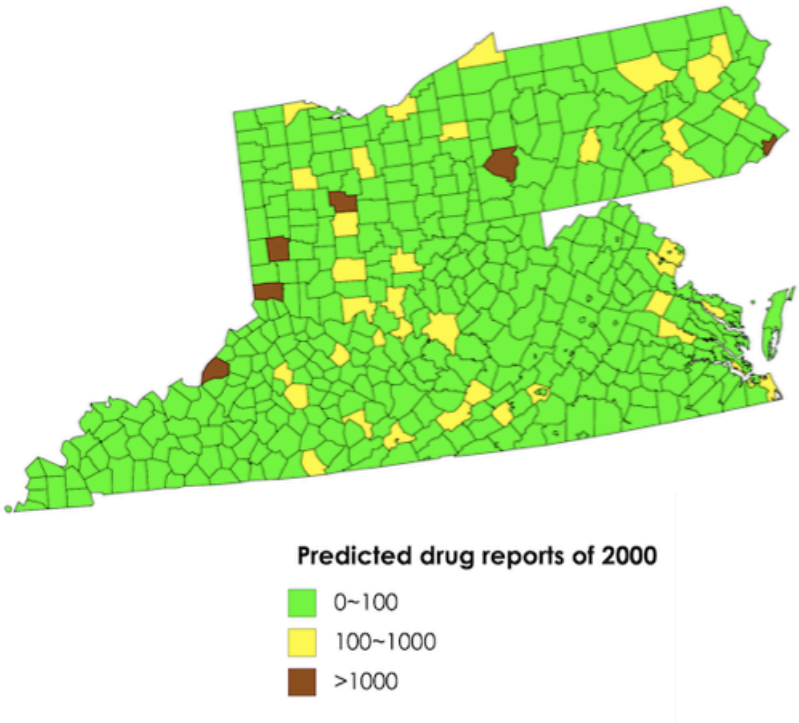

Figure 4: Estimated Geographical Distribution in the year 2000

Table 2: The influence of λ's change on error

λ	k_1	k_2	error
0.3	0.999	0.186	78691.4
0.5	0.999	0.112	78694.1
0.8	0.999	0.070	78694.1

References

Appendices

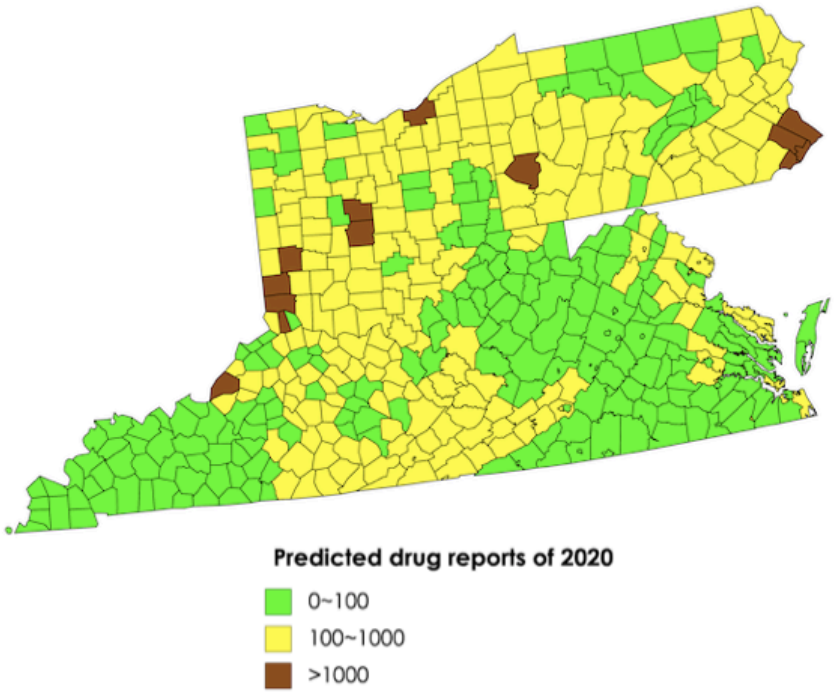

Figure 5: Estimated Geographical Distribution in the year 2020