Team Control Number

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

# 18434

Problem Chosen

**C**

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

---

**2013 Mathematical Contest in Modeling (MCM) Summary Sheet**
(Attach a copy of this page to each copy of your solution paper.)

Type a summary of your results on this page. Do not include
the name of your school, advisor, or team members on this page.

# Modeling Planet Air Health: Associating Global Bifurcation with Local Temporal Network


**Problem Clarification**: Ecological systems can experience a catastrophic regime shift if the forcing level goes beyond what the system resilience can handle. Among the various planet health measures, we focus on the modeling of a major global forcing, i.e. air pollution.

**Major Assumptions**: ⑴Major countries are perceived as entities representing local ecological systems. ⑵Air pollution produced by a country can be stored/resolved in local environments, released to upper atmosphere or transferred to other countries ⑶Upper atmosphere applies feedbacks to every country.

**Modeling Approach**: We construct a directed local temporal network with each node representing a country. Parameters associated with a node describe how air pollution is generated, resolved, and distributed. Links among the nodes represent possible pathways for pollution transfer. We then model a global node with bifurcation dynamic and connect it to the local network. Local forcings' influence on the global forcing and feedbacks from the global node are simulated by correlating the entropy of the local network and the chaos index in the global bifurcation. We set 2 measures at the local level and global level, reflected by local pollution accumulation and global bifurcation status respectively. The model parameters are estimated via analytical hierarchy processing.

**Results and Analysis**: We run the model and get the local & global health measure after 50 years. ⑴We get the status of air pollution accumulation for 41 major countries in the world. ⑵Our model predicts a global catastrophic regime shift at around 2045. ⑶Motivated by sensitivity analysis via Kendall's tau, we discovered the power law within our scale-free network and calculated each node's influence score via leader rank algorithm. ⑷We further propose a method for signaling the global regime shift by observing skewness of probability distribution of the sustainability value in the global bifurcation.

**Strengths and Weaknesses**: Our model accounts for air health both at the local and global level. We intricately connect the two levels by associating entropy of the local network and chaos index in the global bifurcation. The model successfully identifies most influential countries regarding air pollution, and gives predictions of global regime shift together with its warning signal. However, our model requires more direct and detailed data for a more accurate estimation. Since the model is based on intrinsic properties of air pollution, it's relatively difficult to generate it to fit other health measures.


## Key Concepts: Temporal network, Bifurcation, Power law, Scale-free

## Key Techniques: Analytical hierarchy processing, Leader rank, Skewness analysis

# Contents

# 1 Introduction

## 1.1 Catastrophic Shifts in Ecosystems

Biological systems are not resting on stationary levels. Rather, they are constantly changing within a defined range of deviations from a certain mean condition [5]. This behavior fits well to the notion of *Sustainability* or *Resilience*, with which a system can tolerate fluctuations and maintain the current status. However, an ecological system can sometimes collapse, or experience a catastrophic regime shift, if the destructive forces go beyond what the system resilience can handle.

Nowadays, major global-scale forcings are going far beyond the highest level we've ever experienced [14]. An abrupt and irreversible shift in the essential structure of the macro ecosystem can be expected, once a critical threshold is reached. While the local-scale regime shift effect has been carefully characterized in various literatures [6][9], we do not know much about the nature of possible planetary-scale state shifts, since existing theoretical models often fail to take into account the complex interactions and feedback loops [5].

Thus, our aim is to present a model that accounts for both regional and global factors with emphasises on temporal-changing parameters. Since a uniform measure for Earth's overall health condition is hard to set, we exclusively focus our attention on air/atmospheric pollution, which is a major forcing to the global environment and a suitable measurement for the ecosystem's quality [10].

## 1.2 Our Modeling Approach

A good global-scale ecosystem model should be able to accomplish several objectives:

- Effectively synthesizing various regional information, especially human-activity-related information. (population, industry size, political constrains, etc.)
- Tracking temporal changes within the model. For a network model, mechanisms governing nodal entities and connection status should be set to cope with time development.
- Being able to reflect the potential catastrophic regime shift in the global level, and offer adequate warning signals before the critical threshold is reached.

Based on the above considerations, we construct our model as follows. A schematic representation is given in figure 1.

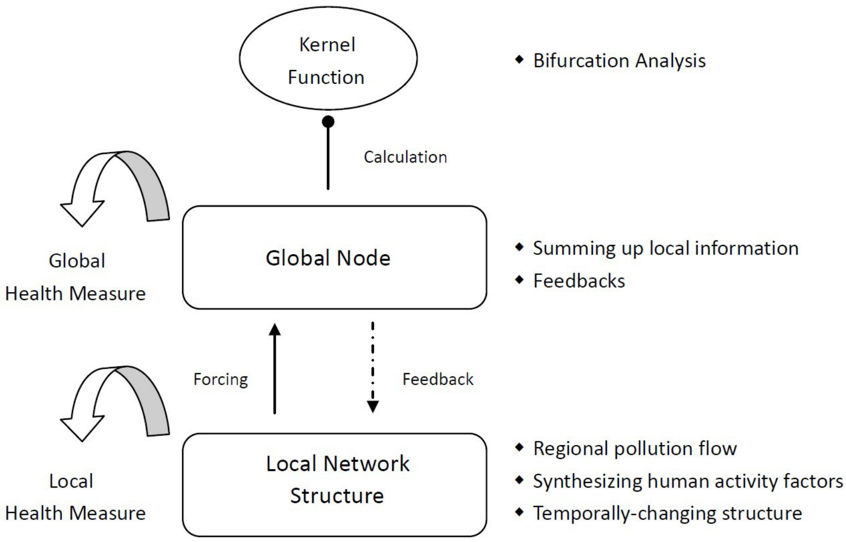


Figure 1: Schematic Representation of Our Model

Firstly, we construct a local-level network modeling regional air conditions. Every node represents a country. Parameters associated with a node describe how the air pollution is generated, resolved, and distributed. Links among the nodes represent possible pathways for pollution transfer. We set a time parameter for every node and link, thus reflecting changes in air pollution conditions over time. We define every node's pollution accumulation to be the local air health measure.

At the global level, we construct a special node called *Global Regime Shift Momentum*(GRSM), aiming at modeling the air pollution accumulated at the macro level and the overall tendency for the global ecological system to collapse. This node is connected to every local node and receives pollution transfer from them, in analog to the actual process of regional air pollution distribution to the atmosphere [10]. We also model feedbacks from the GPSM to local nodes, in an effort to reflect the global forcing's feedback loop to the local entities.

We associate a *kernel function* with the GRSM, whose parameters are directly affected by the local network properties. The kernel function itself is actually a differential equation describing the dynamical process of bifurcation, which will fall from a normal state to an alternative destruction state after a tipping point is reached. We define the kernel to be the global air health measure.

We perform simulations on the model to predict both the local and global air pollution status in the future, detect the tipping point, and establish a warning signal. Following that, we perform further analysis on the network properties and identify the most significant nodes in the network, thus providing information for possible ways to prevent the global ecological regime shift. Finally, we discuss strengths and weaknesses of our model.

# 2 Constructing the Temporal Network

In this section, we will describe definitions for the nodal entities and links of our network model, together with its functional mechanism and underlying significance.

## 2.1 Major Assumptions

Before formally introducing the model, we commence with the following assumptions:

- Major countries are perceived as entities representing local ecological systems. They are also viewed as the source of air pollution.
- Air pollution produced by a country can be stored/resolved in the local environment, released to the upper atmosphere or transferred to other countries through economic collaborations.
- The upper atmosphere applies a global forcing feedback to every individual country.
- A country's ability to produce/resolve air pollution is related to various local factors.

## 2.2 Basic Structure Description

Countries with divergent local statuses behave differently in air pollution producing/desolving. And they are entities that highlight human activity, with themselves being the major ecological systems on the Earth. As displayed in figure 2, countries around the world are marked with different levels of air pollution, which gives us the justification to define a basic node in the network to be a country. We

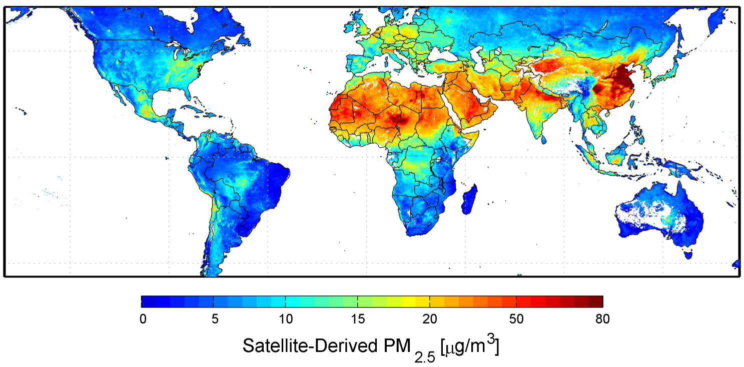


Figure 2: PM values for different countries in the world

model the local nodes to be countries around the world, together with a special node termed *Global Regime Shift Momentum*(GRSM), as shown in figure 3.

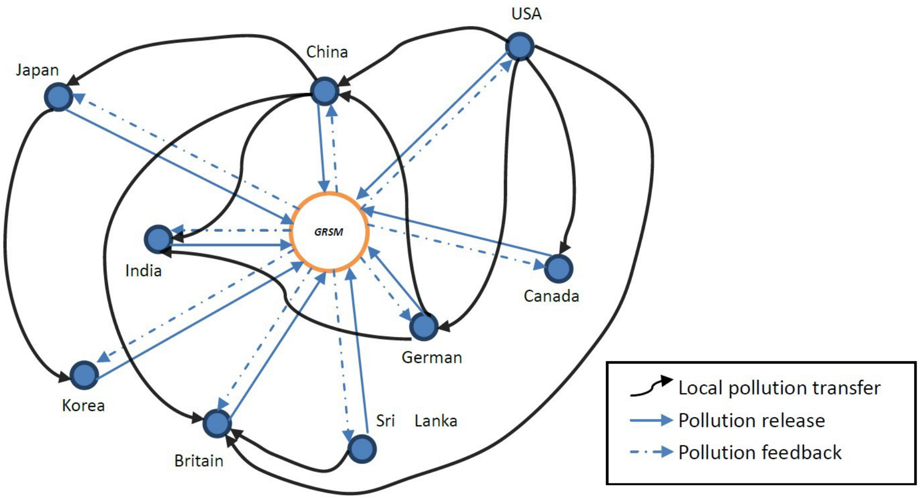


Figure 3: Basic network structure

### 2.2.1 Nodal Definition

We then describe parameters for every local node as follow:

- *Pollution Producing Capacity* (PPC) Amount of pollution produced by a country at a certain instant. The value of the current PPC for a certain country is calculated combining the local information of gross domestic product (GDP), population, land area, total energy consumption, since these values are observed to be variate with PPC [1].We denote PPC by $P_p(t)$.
- *Pollution Resolving Capacity* (PRC) Amount of pollution resolved by a country at a certain instant. This value is calculated combining the local information of domestic forest size, land area, technology level, and annual rainfall amount. We assume these values to be variate with PRC. We denote PRC by $P_r(t)$.
- *Pollution Accumulation* (PA) Amount of air pollution accumulated within the local node. We denote PA by $P_a(t)$. If we define the pollution released to the upper atmosphere by $R(t)$, the pollution transferred to other countries by $T(t)$, and the pollution feedback given by the upper atmosphere by $F(t)$ then PA is given by

$$P_a(t+1) - P_a(t) = P_p(t) - P_r(t) - R(t) - T(t) + F(t) \tag{1}$$

The time increment in the model is *month*. The property of $R(t)$ , $T(t)$ and $F(t)$ is discussed in the Link definition.

We still have a special node GRSM:

• The GRSM is a global node connected to every local node in the network. It receive pollution released from local nodes, and give feedbacks to them respectively. It has an PA value denoted by $P_A(t)$, but it does not have PPC and PRC. Detailed discussions are given later.

#### 2.2.2 Link Definition

As shown in figure 3, our network has 3 kinds of links:

• *Local Pollution Transfer* (LPT) We define a directed LPT from one country to another if practical data show there exist industrial collaboration between the two countries. Countries with a higher PPC typically have better economic status and tend to transfer their pollution to other countries; For countries with lower PPCs, vice versa. Thus, we denote LPT by $T_{\alpha\beta}(t)$, and a directed LPT from country $\alpha$ to country $\beta$ is given by

$$\mid T_{\alpha\beta}(t) \mid = k_1 \cdot \left[P_p^{\alpha}(t) - P_p^{\beta}(t)\right] \cdot P_a^{\alpha}(t) \tag{2}$$

Where $k_1$ is a scaling factor. We would expect a higher $T(t)$ if the discrepancy in PPC between the two countries is large, and if PA of the initiating country is large. For country $\alpha$, $T_{\alpha\beta}(t)$ is positive, while for country $\beta$ it's negative.

• *Pollution Release* (PR) The PRs are directed from local nodes to the GRSM. Since countries with larger PAs will generally release more air pollution to the upper atmosphere, we denote PR by $R(t)$ and define it as

$$R(t) = k_2 \cdot P_a(t) \tag{3}$$

Where $k_2$ is a scaling factor.

• *Pollution Feedback* (PF) We assume the upper atmosphere give pollution feedbacks to local nodes proportional to their country land areas. Countries with larger lands would receive more feedbacks. On the other hand, when the PA of GRSM is higher, we should also expect a higher PF. Thus, we denote the country land area by $S$ and the PF by $F(t)$. The value is given by

$$F(t) = k_3 \cdot S \cdot P_A(t) \tag{4}$$

Where $k_3$ is a scaling factor.

#### 2.2.3 Special Node: Global Regime Shift Momentum

Our atmospheric ecosystem is exposed to gradual changes from different regions in the world. In our model, These changes are modeled as PR from local nodes to the global node. Faced with those changes, the global system can maintain a stable state within a certain extent, but once a tipping point is reached, the system can experience an irreversible collapse to an alternative stable state.

Here we utilize a kernel function associated with the GRSM to simulate this system behavior. R.M.May has established a practical model accounting for grazing systems [9], and we observe an analogue between the biomass in grazing systems and Sustainability/Resilience of our atmospheric ecosystem. Thus, we employ May's model and define the kernel function as

$$\frac{dS}{dt} = rS\left(1 - \frac{S}{S_{max}}\right) - c\frac{S^2}{S^2 + S_0^2} + \sigma_S \eta_S(t) \tag{5}$$

Explanations for model parameters are given in table 1. We then describe the

| Model Parameters | Description and/or values |
|---|---|
| $S$ | Sustainability of the atmospheric ecosystem |
| $r$ | Sustainability recovery rate. Here $r = 1$ |
| $S_{max}$ | The maximum Sustainability the system can theoretically reach |
| $c$ | Chaos Index, a global forcing applied to the system |
| $S_0$ | The value of Sustainability in the alternative state |
| $\sigma_S$ | Standard Deviation of external noise applied to the system. |
| $\eta_S(t)$ | External noise applied to the system. i.e. Human activity |

Table 1: Parameter Description for the kernel function

rationale for such modeling: $S$ will go through a bifurcation dynamic, which is consistent with the behavior of an ecological system [5]. We set $r$ to be 1 since we assume sustainability of the air ecosystem can stay stable without forcing and noise. $\eta_S(t)$ is a Gaussian noise term, and is used to simulate human activity influence here. And most importantly, we use $c$ to model the direct forcing to the system.

Behavior of the kernel function is illustrated in figure 4. As can be observed, S value can go along two stable trajectories. The two trajectories are calculated by setting $dS/dt$ to be 0, aiming at obtaining the stable status. Any value that deviates from the trajectories will be attracted to one of them. While the upper trajectory in our model represent the current air ecosystem sustainability, the lower one stands for a destructive state where the global air quality becomes intolerable. We say the ecosystem experiences a regime shift if S value goes from the upper trajectory to the lower one.

We illustrate two possible ways for the regime shift to take place, respectively in subfigure A and subfigure B:

• Increasing chaos index. As the global forcing becomes increasingly intense, the tipping point is reached and a regime shift is unavoidable.

• Large standard deviation (SD) of the noise term, which means the regional human activity is utmost abrupt. (a nuclear war, etc.)

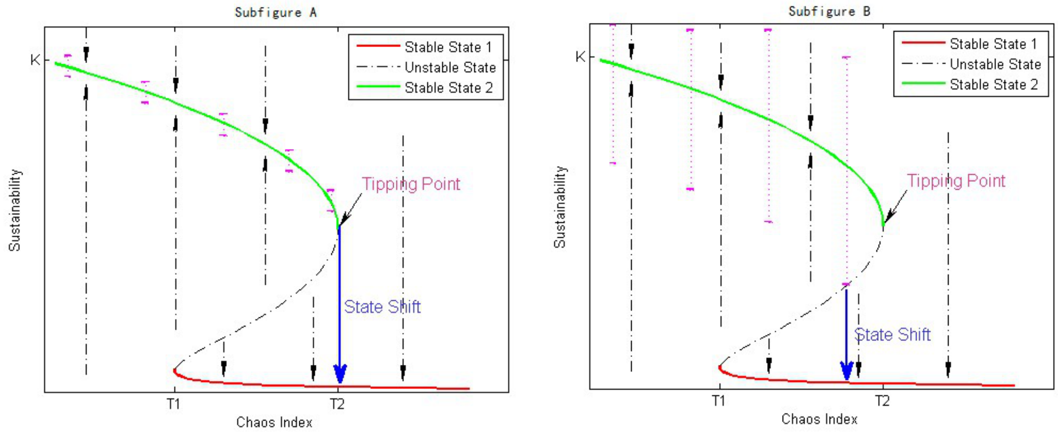


Figure 4: Illustration for the bifurcation behavior. *Subfigure A*: Regime shift caused by increasing chaos index; *Subfigure B*: Regime shift caused by large SD of the noise. i.e. Abrupt human activity.

As pointed out by [4], although diverse events can trigger such shifts, loss of sustainability is the key reason for a state switch. Hence, we focus our modeling exclusively on the global forcing, i.e. the chaos index.

We view global forcing as a factor that increases over time. Since the global forcing is implicitly modulated by regional factors, we associate the local network property with c by defining:

$$\frac{dc}{dt} = k_4 \cdot \frac{\text{Sum of local PAs}}{\text{Entropy of local network}} = k_4 \cdot \frac{\sum_{i=1}^{n} P_a^i}{-\sum_{i=1}^{n} \phi_i \ln \phi_i} \tag{6}$$

$$\text{With} \quad \phi_i = k_i^{out} / \sum k_i^{out} \tag{7}$$

Where $P_a^i$ is the PA for a single node, and $k_i^{out}$ is the out-degree of an individual node. $k_4$ is a scaling factor. Rationale for such a design is clear:

• Since the global forcing is driven by the sum of local forcings, a larger sum of local PA over the world implies a higher rate for global forcing increasing; and

• Since ecosystems are more vulnerable to pollution centralized at one region than distributed to multiple regions, a larger entropy of the local network would make a smaller rate for global forcing increasing: local pollution is more uniformly distributed in a network with a higher entropy.

By modeling the kernel function in this manner, we can expect a local-forcing-modulated increase in the global forcing (chaos index) over time, which can provide insights into forecasting the tipping point.

## 2.3 Metric for Health Measure

We propose metric at 2 different levels for the earth atmospheric health measure:

• Local measure: The local measure is directly reflected by the pollution accumulation (PA) for every individual node. A higher value of PA means a heavier air pollution in that region, hence a worse air health condition.

• Global measure: The global measure is directly reflected by the Sustainability (S) in the kernel function. If $S$ stays along the upper trajectory, we say it's acceptable and the air ecosystem is functioning normally; If $S$ falls to the lower trajectory, we say the system has collapsed and Earth's atmosphere is no longer healthy.

# 3 Running the Model

We include 41 major countries around the word as nodes in our model. By running the model for a certain period of time, we can forecast both local air health conditions and the global air health condition. Before the model simulation, we need to determine several parameters for the model:

• The pollution producing capacity (PPC) and pollution resolving capacity (PRC) for every local node. We use various human activity factors to estimate the parameter value.

• Directed connections (LPTs) among local nodes. We determine the existence of a link with practical information regarding industrial collaborations

• Scaling factors in equation (2), (3), (4), (5) and (6). We give a subjective raw estimation and discuss what data is needed in order to form an accurate one.

## 3.1 Parameter Estimation

### 3.1.1 PPC and PRC

We estimate the PPC and PRC for local nodes using accessible data. Since they have similar intrinsic property , we only illustrate the calculation of PPC here. As stated before, PPC for a chosen country is co-varying with local information including GDP, population, land area and total energy consumption. The 4 factors are then termed GDP, POP, ARE, and ENE respectively. We define the value of PPC ($P_p$) as a linear combination of these four factors:

$$P_p = \sum_{i=1}^{4} w_i I_i \tag{8}$$

Where $I_i$ are the normalized values for the 4 factors respectively. Applying the Analytical Hierarchy Process (AHP) [11], we first build a 4X4 matrix by pair comparison:

$$\begin{array}{c} \\ GDP \\ POP \\ ARE \\ ENE \end{array} \begin{array}{c} GDPPOPAREENE \\ \begin{pmatrix} 1 & 2 & 6 & 1/2 \\ 1/2 & 1 & 4 & 1/3 \\ 1/6 & 1/4 & 1 & 1/9 \\ 2 & 3 & 9 & 1 \end{pmatrix} \end{array} \tag{9}$$

The numbers' meanings are explained in [12], while the numbers themselves are determined by our subjective decisions.

The calculated weights are then given in table 2.

| Factor | GDP | POP | ARE | ENE |
|---|---|---|---|---|
| Weight | 0.2877 | 0.1661 | 0.0479 | 0.4983 |

Table 2: Calculated weights for PPC

To test the consistency of our weight estimation, we employ the method described in [2]. A good estimate requires:

• The principle eigenvalue $\lambda_{max}$ of the matrix should be close to number of alternatives. Here we have 4 alternatives and $\lambda_{max} = 4.02$.

• The consistency index (CI) should be close to 0. Here we have $CI = (\lambda_{max} - n) / (n - 1) = 0.0069$

• The consistency ratio (CR) should be less than 0.01. Here we have $CR = CI/RI = 0.0077$, where $RI$ is the average value for CI in terms of random matrices.

Thus, we conclude our estimate for PPC weights is reasonable and robust.

PRC is calculated in a manner identical to PPC, except for the fact that it's covarying with local information including domestic forest size, land area, technology level, and annual rainfall amount. We then term the 4 factors by FOR, ARE, TEC and RAI respectively. The calculated weights for PRC are given in table 3 Similar consistency analysis also qualifies the estimate for PRC as robust.

| Factor | FOR | ARE | TEC | RAI |
|---|---|---|---|---|
| Weight | 0.4959 | 0.1542 | 0.2672 | 0.0827 |

Table 3: Calculated weights for PRC

Next we get data for the multiple factors, i.e. the $I_i$ s. Data value for some major countries we consider to be representative and the data source are shown in table 4 and table 5 as an illustration.

The divergent factors have different Units in their measurements, which can cause confusions during the data processing. Thus, we perform a scale normalization before calculation. This normalizes values for all factors to a $[0, 1]$ interval. The calculated PPC and PRC for the listed countries are shown in figure 5.

| | GDP/ Billion USD | Population/ $10^4$ people | Surface Area/ $10^4$ sq.km | Energy consumption/ $10^4$ tons of standard oil |
|---|---|---|---|---|
| United States | 142969 | 31689 | 963.2 | 168640 |
| Great Britain | 26575 | 4820 | 24.4 | 15891 |
| China | 45218 | 136292 | 959.8 | 208494 |
| Australia | 10394 | 2167 | 774.1 | 31070 |
| Venezuela | 3111 | 2845 | 91.2 | 20353 |
| South Africa | 2753 | 4840 | 121.9 | 16064 |

Table 4: Values for estimating PPC. Source: http://data.worldbank.org/

| | Forest Area/ $10^4$ sq.km | Tech. Expenditure/ Billion USD | Rainfall/ Million CBM | Surface Area/ $10^4$ sq.km |
|---|---|---|---|---|
| United States | 319 | 3946 | 6440000 | 963.2 |
| Great Britain | 2.9 | 489 | 275029 | 24.4 |
| China | 211 | 674 | 6172800 | 959.8 |
| Australia | 166 | 168 | 3630635 | 774.1 |
| Venezuela | 49 | 11.8 | 1406154 | 91.2 |
| South Africa | 9.3 | 27.4 | 524600 | 121.9 |

Table 5: Values for estimating PRC. Source: http://data.worldbank.org/

We can observe from the figure: USA is in a near-balance between PPC and PRC, which implies a relatively stable atmospheric pollution condition. On the other hand, China (CHN)'s PPC clearly surpasses its PRC, forecasting a trend toward heavier air pollution. For Australia (AUS), PRC is far larger than PPC and implies the country has a potential for better air quality. This fits well to the reality [14].

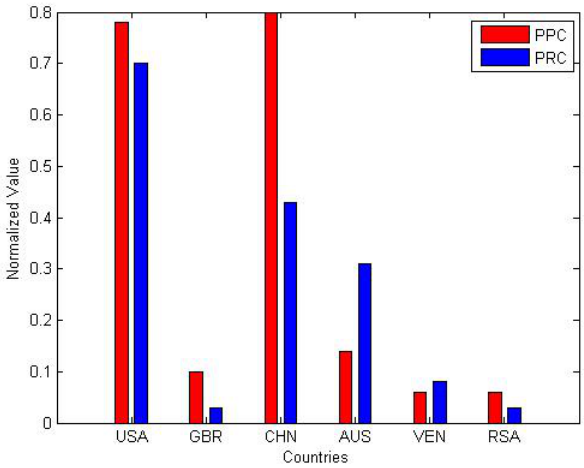


Figure 5: PPC and PRC values for the listed countries

### 3.1.2 Network Connection Establishment

We determine the existence of a link from one country to another by the following rule:

> *If country $\alpha$ and $\beta$ are engaged in a certain industrial relationship, through which $\alpha$ can potentially transfer air pollution burden to $\alpha$, then a link from node $\alpha$ to $\beta$ is identified.*

Such information is collected by searching past news. If no relevant information between a certain country pair is available, we consider the link to be non-existent. The connection status of the previously listed 6 countries is given in table 6 as an illustration.

| | United States | Great Britain | China | Australia | Venezuela | South Africa |
|---|---|---|---|---|---|---|
| United States | / | | ✠ | ✠ | ✠ | |
| Great Britain | | / | | | | |
| China | ✓ | | / | ✠ | | ✠ |
| Australia | ✓ | | ✓ | / | | ✠ |
| Venezuela | ✓ | | | | / | |
| South Africa | | | ✓ | ✓ | | / |

Table 6: ✓Link from row country to column country. ✠Link from column country to row country.

### 3.1.3 Scaling Factors

We next determine the scaling factors in our model. For an accurate estimation of the scaling factors, we must obtain more direct and detailed data regarding

- countries' economic/technological development; and
- portion of air pollution that can be transferred to another country and release to the upper atmosphere during a certain period of time; and
- an overall measure of the global air quality.

We give a raw estimation of all scaling factors, which are listed in table 7.

| Scaling factor | Which equation | Estimated Value |
|---|---|---|
| $k_1$ | (2) | $1/P_p^{\beta}(t_0)$for every country pair |
| $k_2$ | (3) | 0.618 |
| $k_3$ | (4) | $0.2 \cdot 10^{-4}$ |
| $S_{max}$ | (5) | 10 |
| $S_0$ | (5) | 1 |
| $\sigma_S$ | (5) | 0.5 |
| $k_4$ | (6) | 0.073 |

Table 7: Estimated scaling factors

## 3.2 Results

In running the model, we set the minimum time increment for $t$ as 1 month. We wish to predict Earth's air health after 50 years, hence the time interval is set to be 600 months. We now discuss the running results respectively for the local level and global level.

### 3.2.1 Local Air Health

The local health measure is reflected in the PA value of every local node. We normalize the PA values as percentages compared with the total amount of local PAs. Final PA distribution is given in figure 6. Previously selected countries are highlighted. Though the parameter settings are not ideally accurate, the distribu-

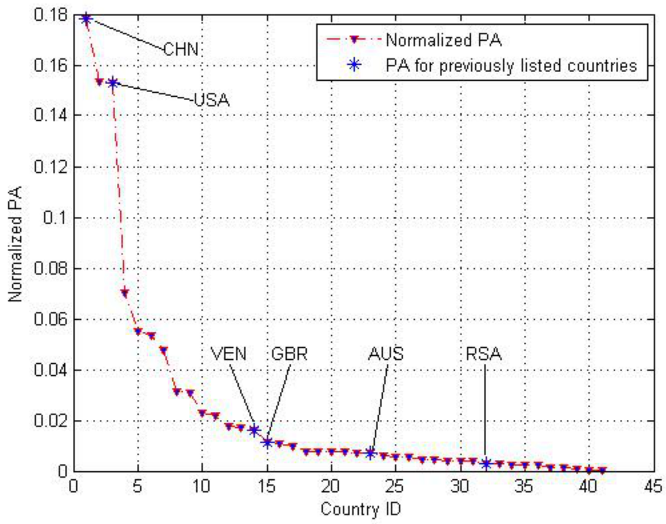


Figure 6: Final local PA distribution

tion is still reasonable to a certain magnitude. We can be more confident about the result once precise parameters are provided.

### 3.2.2 Global Air Health

As predicted, a global environmental collapse is observed at around 400 months later. This is due to the increase in $c$ (chaos index) over time. After the tipping point is reached, the regime shift is irreversible and $S$ (Sustainability) stays at a destructive low level. A prediction signaling the upcoming tipping point is highly desirable. The only available information aiding the prediction is the value of $S$, which can be quantified by a air quality measure. Hence we employ the method

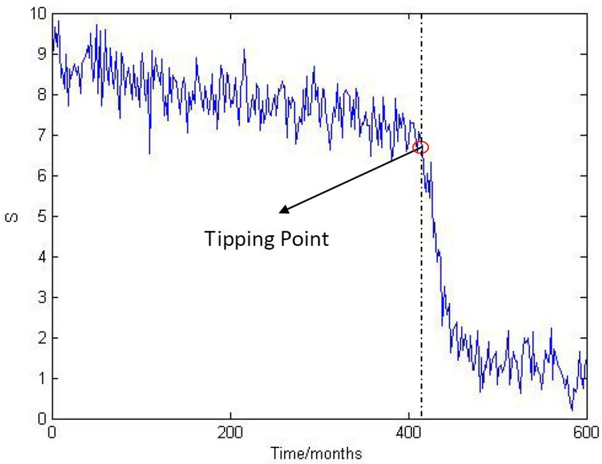


Figure 7: Modeling of global air health

described in [7]. By observing the probability distribution of $S$ value during a relatively short time interval (compared with the total simulation time), we can establish a correlation between the skewness of the probability distribution and the current value of chaos index.

For an illustration, we selected 2 time points with one close to the tipping point and the other far away from it. Then we choose a time interval of 300 days (10 months) centered at the 2 points and record the 2 probability distributions, which are given in figure 8. Clearly, a higher skewness level can be seen from the near point. Motivated by this fact, we then perform a correlation analysis between the skewness and the chaos index. We start by quantitatively defining the skewness as

$$\gamma = \frac{\int (x-\mu)^3 p(x)dx}{\sigma^3} \tag{10}$$

Where $p(x)$ is the density function of the distribution, and $\mu$ and $\sigma$ are the mean and standard deviation of the distribution respectively. Next, we choose time points along the overall interval and calculated their skewnesses together with corresponding chaos indexes, as shown in figure 9. As chaos index approaches the tipping point, the relative change in skewness gets increasingly abrupt. This provides a possible way for signaling the regime shift.

## 3.3 Sensitivity Analysis

We perform sensitivity analysis regarding the network structure. Since unexpected hazardous events (nuclear exploration, etc.) can happen without warning in local areas, it's necessary to see how the final PA distribution can change if local param-

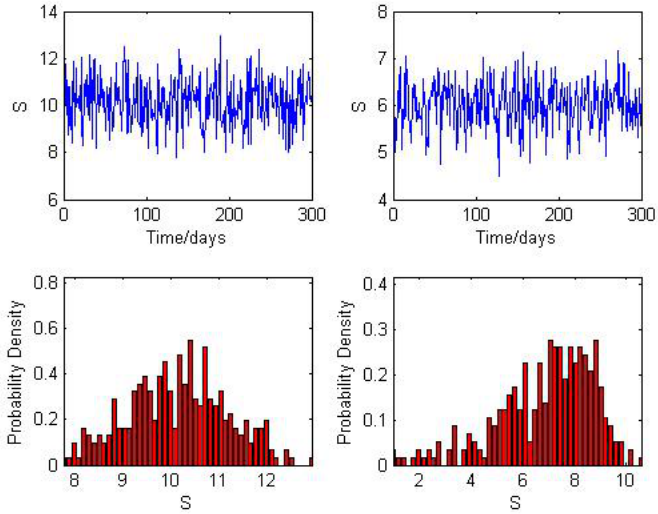


Figure 8: Probability distributions of $S$ for the 2 time intervals. *Left figure*: Time point close to the tipping point. *Right figure*: Time point far away from the tipping point

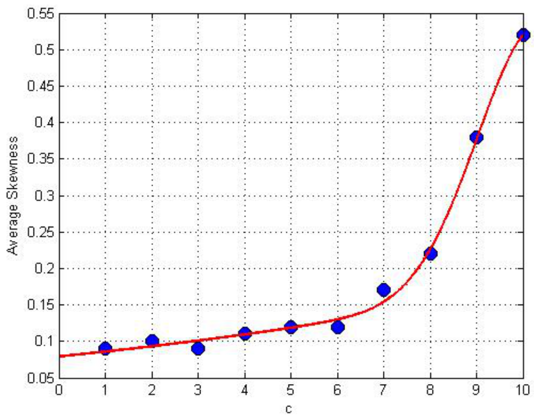


Figure 9: Skewness and chaos index

eters alter abruptly. We thus simulate unexpected local events in a certain node by doubling the initial value of PPC. An illustration is given in figure 10. Adjustment for China induces a relatively large fluctuation in the RA distribution, while the same doubling on PA applied to South Africa does not produce such a result. In order to measure the influence of an adjustment on a certain node, we utilize the concept of *Kendall's tau coefficient* [13] to analyze the PA ranking change among

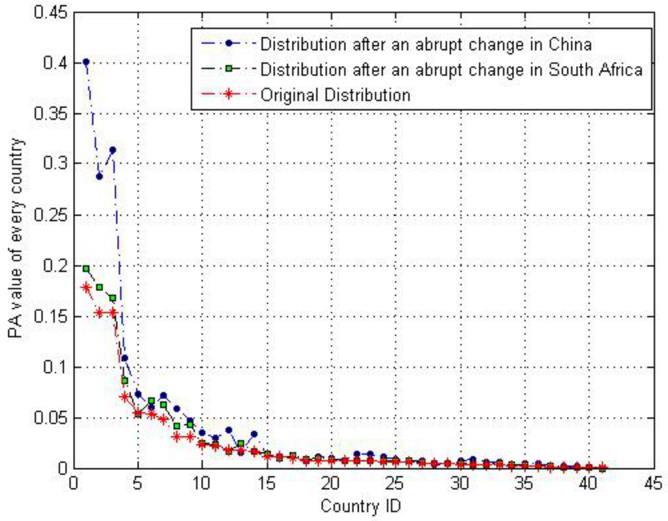


Figure 10: Final PA distribution after adjustment for China/South Africa

the 41 countries. The kendall's tau coefficient is formally defined as

$$\tau = \frac{\text{(number of concordant pairs)-(number of discordant pairs)}}{\frac{1}{2}n(n-1)} \tag{11}$$

Where $n$ is the number of IDs. In our model, $n = 41$. $\tau$ ranges within $[-1, 1]$, with 1 meaning no change in the ranking and $-1$ meaning the ranking is completely reversed. We compute kendall's taus for adjustments in all 41 countries. Results for the previously listed countries are shown in table 8 as an example.

| Country | CAN | USA | GBR | AUS | VEN | RSA |
|---|---|---|---|---|---|---|
| Kendall's tau | 0.46 | 0.32 | 0.88 | 0.82 | 0.96 | 0.97 |

Table 8: kendall's taus for the 6 countries

While China and USA have kendall's tau coefficients that approach 0, the other 4 countries' coefficients are near 1. This reflects a higher influence of local forcing from the former country group.

# 4 A Further Look into the Network Structure

In our sensitivity analysis, adjustments in some countries produce huge fluctuations in PA ranking, while PA ranking is not so sensitive to other countries. This motivates us to find the important nodes in the network. We start by considering the practical meaning of our network and how its connections are formulated.

#### 4.0.1 Power Law in the Network

A directed link from country $\alpha$ to $\beta$ implies the former are transferring air pollution to the latter through certain industrial activities. Thus, we consider how industrial collaborations are establish in the real world. Since less-developed countries always tend to establish connections with the most developed countries, our network has an interesting property: Some nodes have very large out-degrees (rich countries), while the others out-degree value is relatively small (poor countries).

For our network $G(V, E)$, where $V$ is the set of nodes and $E$ is the set of links, we simulate the process how our network is formulated: once a new node is inserted into the network, it establishes connections with $m$ nodes with a probability of $p_i = k_i^{out}/|E|$, where $k_i^{out}$ is the out-degree of the $ith$ node and $|E|$ is the total number of links. Thus, nodes with a large out-degree tend to "absorb" more "new comers", and consequently gain a even larger out-degree. On the other hand, nodes with a low According to [3], the degree distributions of such networks obey the *Power Law*, i.e.

$$P_k(\text{Degree Distribution}) \propto 2m^2k^{-3} \tag{12}$$

We then plot the out-degree distribution in our network, with the x-axis adjusted to logarithmic, as shown in figure 11. Though the distribution does not fits perfectly

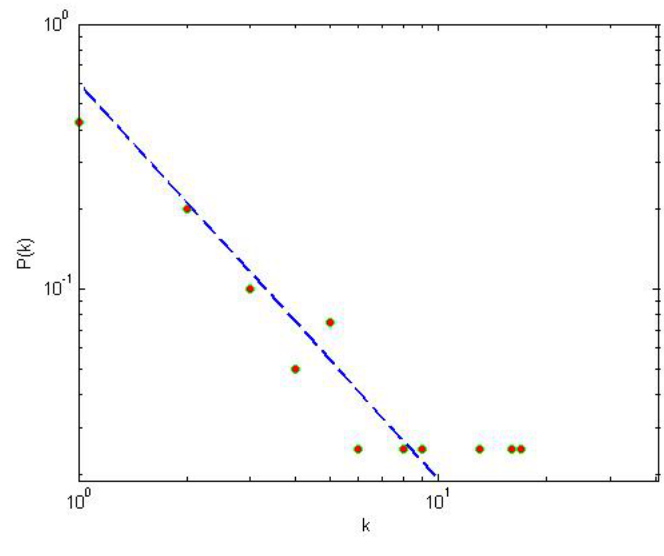


Figure 11: Out-degree distribution in our network

to the drawn line, such tendency is clear. Actually, our network can be viewed to obey power law, where $p_k$ decreases slowly even when $k$ is relatively large, compared with random networks.

### 4.0.2 Finding Important Nodes via Leader Rank

After mining out the network property, we are interested in identifying important nodes in our network. Since the out-degree value does not provide a good measure for influence of a node, we employ the *Leader Rank* method described in [8]. The method inserts a leading node into the network, which can aid the process of importance score assignment. This method outperforms others in its non-sensitivity to node initial values, which is exactly needed in our network. If we define node $i$'s importance score as $S_i(t)$, then the score evolves over time as

$$S_i(t+1) = \sum_{j=1}^{|V|+1} \frac{\alpha_{ij}}{k_j^{in}} S_j(t) \tag{13}$$

Where $k_j^{in}$ is the in-degree of a node. $\alpha_{ij} = 1$ if there exist a link from $i$ to $j$, and $\alpha_{ij} = 0$ if the link is non-existent. When the score-assignment converges, the final importance score for every node is given by

$$S_i = S_i(+\infty) + \frac{Sg(+\infty)}{|V|} \tag{14}$$

Where $Sg(+\infty)$ is the final score of the inserted leading node.

We then calculate the final scores of all nodes in our network. All initial scores are set to be 1 due to the parameter non-sensitivity of the ranking method. As a result, the top 5 most important nodes are identified in table 9.

| Country | USA | China | Russia | India | Canada |
|---|---|---|---|---|---|
| Leader Score | 6.28 | 5.33 | 4.58 | 4.05 | 2.69 |

Table 9: Leader score for the top 5

# 5 Avoid the Inevitable: Suggestions for Governments

Based on our analysis regarding the model, we propose 3 major suggestions for governors:

- At the local level, since we've identified the most influential nodes in the local network, those countries should pay special attention to air pollution control. Possible solutions include reducing pollution production (shutting down factories with heavy pollution, etc.) and increasing pollution resolving capacity (enlarging forests, etc.)

- At the global level, since the catastrophic shift can be slowed down or avoided by balancing air pollution accumulated in countries around the world (increasing the network entropy), we recommend the relatively developed countries stop transferring air pollution to other countries through industrial bounds. On the other hand, those technologically advanced countries should, if possible, provide others with techniques for air pollution cleaning.

- Our model predicts a global regime shift in air quality at around 2045 (400 months later). Time is short, and immediate action is highly desirable.

# 6 Strengths and Weaknesses

- **Strengths** Our model accounts for air health both at the local level and the global level, respectively with a temporal network and a bifurcation dynamic. We intricately simulate local forcings' effect on global forcing by associating entropy of the local network and chaos index in the global bifurcation. Feedbacks from the global forcing to local regions are also reflected by such a method. By running and analyzing the model, We successfully identify the most influential countries on air pollution, and give predictions of global regime shift together with its warning signal.

- **Weaknesses** Our model requires more direct and detailed data for a more accurate estimation. Since the model is based on intrinsic properties of air pollution, it's relatively difficult to generate it to fit other health measures.

# References

[1] International Energy Agency. Key World Energy Statistics. *www.iea.org*, 2011.

[2] Jose Antonio Alonso. Consistency im the analytical hierarchy process: a new approach . *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 14(445–459), 2005.

[3] Albert-Lszl Barabsi and Rka Albert. Emergence of scaling in random networks. *Science*, 286(509–512), 1999.

[4] Scheffer et. al. Catastrophic shifts in ecosystems. *Nature*, 413(591–596), 2001.

[5] Barnosky et.al. Approaching a state shift in Earth's biosphere. *Nature*, 486(52–58), 2012.

[6] V. Guttal and Jayaprakash. Impact of noise on bistable ecological systems. *Ecol. Model*, 201(420–428), 2007.

[7] Vishwesha Guttal and Ciriyam Jayaprakash. Changing skewness: an early warning signal of regime shifts in ecosystems. *Ecology Letters*, 11(450–460), 2008.

[8] Zhang Y.-C. Yeung C. H. Lü, L. and T. Zhou. Leaders in social networks, the Delicious case. *PloS one*, 6, 2011.

[9] R.M. May. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature*, 269(471–477), 1977.

[10] Mario J. Molina and Luisa T. Molina. Megacities and Atmospheric Pollution. *J.Air and Waste Manage.Assoc.*, 54(644–680), 2004.

[11] Thomas L. Saaty. The Analytic Hierarchy Process: A New Approach to Deal with Fuzziness in Architecture. *Architectural Science Review*, 25(64–69), 1982.

[12] Thomas L. Saaty. Axiomatic foundation of the analytical heirarchy process . *Management Science*, 32(841–855), 1986.

[13] K. P. Sen. Estimates of the regression coefficient based on Kendalls tau. *Journal of the American Statistical Association*, 1968.

[14] Robert Watson and A.Hamid Zakri. UN Millennium Ecosystem Assessment Synthesis Report. *United Nations Report*, 2005.