Article

Future Climate Data from RCP 4.5 and Occurrence of Malaria in Korea

**Jaewon Kwak [1], Huiseong Noh [2], Soojun Kim [3], Vijay P. Singh [4], Seung Jin Hong [2],
Duckgil Kim [5], Keonhaeng Lee [6], Narae Kang [2] and Hung Soo Kim [2,*]**

[1] Hydroclimatic Statistical Research Group, Centre Eau Terre Environnement, INRS, Québec, QC G1K 9A9, Canada; E-Mail: Jaewon.Kwak@ete.inrs.ca

[2] Department of Civil Engineering, Inha University, Incheon 402-751, Korea;
E-Mails: heesung80@hanmail.net (H.N.); hongsst81@gmail.com (S.J.H.);
naraeme@naver.com (N.K.)

[3] Columbia Water Center, Earth Institute, Columbia University, New York, NY 10027, USA;
E-Mail: soojun78@gmail.com

[4] Department of Biological and Agricultural Engineering, Zachry Department of Civil Engineering, Texas A & M University, College Station, TX 77843, USA; E-Mail: vsingh@tamu.edu

[5] Water Environment Research Department, Water Quality Assessment Research Division, National Institute of Environmental Research, Incheon 404-708, Korea; E-Mail: gilkim@korea.kr

[6] Water Resources Research Division, Water Resources and Environment Research Department, Korea Institute of Civil Engineering and Building Technology, Goyang-si, Gyeonggi-do 411-712, Korea; E-Mail: leeggun@kict.re.kr

* Author to whom correspondence should be addressed; E-Mail: sookim@inha.ac.kr;
Tel.: +82-32-860-7572; Fax: +82-32-876-9783.

External Editor: Paul B. Tchounwou

Abstract: Since its reappearance at the Military Demarcation Line in 1993, malaria has been occurring annually in Korea. Malaria is regarded as a third grade nationally notifiable disease susceptible to climate change. The objective of this study is to quantify the effect of climatic factors on the occurrence of malaria in Korea and construct a malaria occurrence model for predicting the future trend of malaria under the influence of climate change.

Using data from 2001–2011, the effect of time lag between malaria occurrence and mean temperature, relative humidity and total precipitation was investigated using spectral analysis. Also, a principal component regression model was constructed, considering multicollinearity. Future climate data, generated from RCP 4.5 climate change scenario and CNCM3 climate model, was applied to the constructed regression model to simulate future malaria occurrence and analyze the trend of occurrence. Results show an increase in the occurrence of malaria and the shortening of annual time of occurrence in the future.

Keywords: malaria; climate change; PCA-regression analysis; climate variable

1. Introduction

Malaria is an acute infectious disease caused by plasmodium parasite infection in red blood cells and liver cells. Approximately 300–500 million cases of malaria patients are reported annually, and over a million patients die from this parasitic infection. The World Health Organization (WHO) has designated malaria as one of the six major tropical diseases. Infection through plasmodium vivax and plasmodium falciparum is a cause of more than 95% of reported cases, and in Korea, plasmodium vivax infection is found in the majority of the reported cases. Generally, infectious diseases that use host as a means of infection are highly susceptible to climate which influences interactions within the ecosystem [1]. Mosquito-borne diseases are especially susceptible and it is known that they are greatly influenced by temperature, precipitation, humidity and other factors. Amongst these diseases, malaria is the most well addressed infectious disease and it is also most susceptible to climate change [2].

Many studies have been conducted to investigate how malaria is related to climate change and climate factors. Poveda *et al*. [3] conducted a study of malaria and time lag effect of El Nino climate in Columbia, U.S. and Craig *et al*. [4] investigated the relationship between malaria and the probability distribution of climate factors in the Sahara region. Paaijmans *et al*. [5] correlated malaria to mean temperature and analyzed the change in the influence of malaria with daily temperature change [6]. Olson *et al*. [7] analyzed the relationship between malaria and wetland and climate factors. Prediction of malaria using an impulse function computed from the climate factors has been made by Kuhn *et al*. [8], whereas Thomson *et al*. [9] predicted malaria infection through seasonal climate factors. Jhajharia *et al*. [10] investigated the influence of climate on incidences of malaria in the Thar desert of Northwest India and found that climatic variability and rise in temperature were key determinants to the transmission of malaria.

Current malaria and climate-related investigations have focused on determining the correlation between malaria and climate factors for predicting the occurrence of malaria [11–15]. Since climate change is now accepted as real, changes in the characteristics of malaria in relation to climate change have been investigated [16,17]. For example, malaria's vulnerability to trends of the past century have been analyzed [2,18,19]. Likewise, changes in the future characteristics of malaria in Africa, the most malaria-vulnerable continent, have been analyzed [20–22]. In order to prepare adequately for ongoing global warming, van Lieshout *et al*. [23] predicted fluctuations in malaria incidences according to the

SRES climate change scenario. These studies show future trends of malaria and climate change and how the occurrence of malaria infection will be affected by future climate change [16,24–31].

The objective of this study is therefore to investigate the occurrence of malaria and its correlation with climatic factors and construct a regression model for predicting malaria occurrences in Korea from climate factors. By simulating future malaria occurrences in Korea, the model can provide basic data for public health agencies. For this study, data on monthly malaria occurrences and climatic factors from 2001–2011 was collected. By applying the climate data generated from a climate change scenario and a climate model to the regression model, future malaria occurrences and their trends were analyzed.
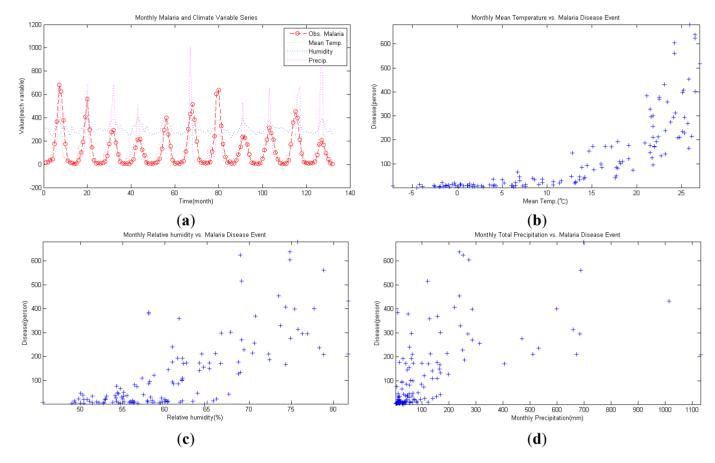
2. Malaria Occurrence in Korea

2.1. Trend of Malaria Occurrence

Although it is commonly believed that malaria seems to be decreasing in Korea since 2000 (active duty military service men included), facts point otherwise. For example, 1325 malaria patients were reported in 2005, 2021 cases in 2006, and 2192 cases in 2007, showing an 8.5% increase (Centers for Disease Control & Prevention, [32]). The communication of malaria is caused by female Anopehline [33]. Most mosquitoes that carry malaria show a much more vibrant activity as weather gets warmer [34]. Therefore, in Korea, an increase of malaria infection is strongly related to the rise in temperature caused by climate change [35].

2.2. Data Collection

Jang *et al.* [1] has reported that malaria has the highest infected domestic patients among the diseases showing relevance to climate factors and climate change. Since national infectious disease database has been established, quality data can now be acquired. Monthly data of the designated infectious diseases between 2001 and 2011 from the Center for Disease Control & Prevention was utilized for obtaining data on malaria occurrences. Data from 2001–2008 was used for calibration and data from 2009–2011 was used for verification. For the same period, climate data was obtained from the Seoul Regional Meteorological Observatory. Based on the data from Center for Disease Control & Prevention [32], the highest malaria occurrence in 2008 happened in Gyeonggi province (475 reported cases), followed by Seoul (180 reported cases), and Incheon (92 reported cases). The distribution of malaria was highly concentrated in the Seoul and Gyeonggi areas. Thus, climate variables from the Gyeonggi and Seoul Meteorological Observatories were averaged into one for use in this study. Also, data from weather stations at E.L. 200 m above the sea level may disrupt the average values of regional meteorological factors, so these were excluded from analysis. Meteorological factors included in the analysis were average temperature (°C), relative humidity (%), and precipitation (mm). The collected meteorological and malaria data are shown in Figure 1a, and the scatter diagram of malaria is shown in Figure 1b–d and for average temperature, relative humidity and precipitation, Figure 1d is referred to.

Figure 1. (**a**) Time series of each variable; (**b**) scatter diagram of mean temperature and malaria; (**c**) scatter diagram of relative humidity and malaria; (**d**) scatter diagram of monthly precipitation and malaria.

(**a**)

(**b**)

(**c**)

(**d**)

3. Methodology

This study performs a regression analysis for modeling the correlation between malaria which is the dependent variable and climate variables which are independent variables. While doing so, spectral analysis must be done to reflect the time lag between each variable and the BDS test was done to check the randomness of the time series. Also, the principal components regression should be applied and tested for considering multicollinearity between variables [36,37].

3.1. Spectral Analysis

Autospectrum or crossspectrum is a method to explain the distribution of variance of frequency drawn from a single or multiple data [38]. Covariance was estimated using autocovariance for data for time interval t and lag k:

$$Cov_{xx}(k) = \frac{1}{N-k-1}\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})$$

(1)

where N : Number of data series, t, and Δt : time interval and time step.

Autocorrelation function can be obtained by dividing the autocovariance by σ^2, variance of $x(t)$. Spectral function $X_{xx}(f)$ can be obtained by applying the Fourier transform to the autocorrelation

function. Spectral analysis would give periodic information of the malaria time series. Cross spectral analysis can also be applied to cross covariance of the $x(t)$, $y(t)$ times series and can be also used to differentiate the relevant covariance of two time series data set to have lag k. Cross covariance and cross spectrum (X_{xy}) of two times series with time interval t can be determined using Equations (2) and (3):

$$Cov_{xy}(k) = \frac{1}{N-k-1}\sum_{t=1}^{N-k}(x_t - \bar{x})(y_{t+k} - \bar{y})$$

(2)

$$|X_{XY}| = \sum_{k=0}^{Maxlag}\left(\frac{Cov_{xy}(k)}{\sigma_x\sigma_y}\right)(k)e^{\frac{i2\pi fk}{f_s}} \quad , \ f_s : \text{sampling frequency}$$

(3)

Also, using the cross spectrum on two time series, coherency can be calculated and information about the frequency between the two (see Equation (4)) can be obtained. Coherency between two time series is commonly represented in the $1 - C_{xy}$ form.

$$C_{xy} = \frac{|X_{xx}|^2}{X_{xx}X_{yy}}$$

(4)

where C_{xy}: coherency of time series, X_{xx}, and X_{yy}: autospectral density of each time series.

3.2. Brock-Dechert-Scheinkman(BDS) Statistic

The BDS statistics is a method to verify a hypothesis that times series complies with the random distribution based on the correlation integral. It is an effective tool to discern between random time series and nonlinear chaos or stochastic system [39–41]. Grassberger and Procaccia [42] suggested correlation integral as a method to measure the fractal dimension on deterministic data, and for embedding dimension m, computed as

$$C(m,N,r) = \frac{2}{M(M-1)}\sum_{1\leq i<j\leq M}\Theta(r - \|\vec{x}_i - \vec{x}_j\|), \quad r > 0$$

(5)

where, $\begin{aligned}\Theta(a) &= 0, \quad if \ a \leq 0 \\ \Theta(a) &= 1, \quad if \ a \geq 0\end{aligned}$.

N is the size of the data sets, $M = N – (m – 1)t$ is the number of embedded points in m-dimensional space, and $\|\cdot\|$ denotes the sup-norm. On target data $\vec{x}_i = (x_i, x_{i+t}, \cdots, x_{i+(m-1)t})$, $\vec{x}_i \in R^m$, $i = [1,2,\cdots,N]$, if the supremum norm is less than the r value, and $C(m,N,r)$ becomes a fractal value for the target data. If N expands to infinity for each r value, the fractal dimension within the r value can be defined as $C(m,r) = \lim_{N\to\infty} C(m,N,r)$. In the case where target data is a stochastic time series with stationary character and also possesses frequency and limit, the value of the corresponding limit can be represented as shown in Equation (6):

$$C(m,r) = \int\int_X \Theta(r - \|\vec{x} - \vec{y}\|)dF(\vec{x})dF(\vec{y}), \quad r > 0$$

(6)

Here, if data possesses independent identically distributed characteristics, $\Theta(r - \|\vec{x} - \vec{y}\|)$ can be represented as $\prod_{k=1}^{m} \Theta(r - |x_k - y_k|)$ and Eq. (6) becomes $C(m,r) = C^m(1,r)$. In this case, $C(m,r) = C^m(1,r)$ is an asymptotic normal distribution with an average of 0, and variance σ^2 as follows:

$$\sigma^2(m, M, r)/4 = m(m-1)C^{2(m-1)}(K - C^2) + K^m - C^{2m}$$
$$+ 2\sum_{i=1}^{m-1}[C^{2i}(K^{m-i} - C^{2(m-i)}) - mC^{2(m-i)}(K - C^2)] \tag{7}$$

Thus, using $C(1,r)$ and the K value, the value for coefficient C can be calculated as

$$K(m, M, r) = \frac{6}{M(M-1)(M-2)} \sum_{1 \le i < j \le M}[\Theta(r - \|\vec{x}_i - \vec{x}_j\|)\Theta(r - \|\vec{x}_i - \vec{x}_j\|)] \tag{8}$$

Thus, under the assumption that time series data is independent identically distributed, where $m > 1$, the BDS statistic can be represented as

$$BDS(m, M, r) = \frac{\sqrt{M}}{\sigma}[C(m,r) - C^m(1,r)] \tag{9}$$

The BDS statistic can be an effective tool to determine whether the given time series is random or is a nonlinear system (chaos or stochastic). To utilize the BDS statistic, selection of the m and *r* values is critical. In this study, *m* was defined as $2 \le m \le 5$ [39], and for r, $0.5\sigma \le r \le 1.5\sigma$ (σ = standard deviation) [43].

3.3. Principal Components Regression

If there is a correlation between selected independent variables that show multicollinearity during regression analysis, then some of the independent variables may be eliminated or new sets of observed values may be introduced. Also, depending on the situation in the field, correlation itself is also subject to elimination [44]. However, elimination of independent variables is not desirable, considering the information loss. As an alternative, principal component regression (PCR) is often utilized [45–48]. Principal component regression selects principal components (less than original variable in numbers) and combines them with the regression model through principal components analysis (PCA) by compressing the dimension as suggested by Morison [49]. By estimating the principle component score of independent variables and conducting regression analysis, information loss can be minimalized while avoiding multicollinearity [50].

In principal component regression, the eigenvalue (λ_i) matrix Λ of the correlation coefficient matrix and eigenvector V satisfy the following Equation (10):

$$V^T(Z^T Z)V = \Lambda, \; V^T V = VV^T = I \tag{10}$$

Eigenvector is an orthogonal matrix, so $VV^T = I$ is valid. When Equation (10) is substituted in the general form of regression model in the matrix form, it can be arranged into a regression model as:

$$y = XVV^T\beta + \varepsilon \tag{11}$$

where X : independent variables of ith, and β : coefficients matrix of independent variables, ε : error term.
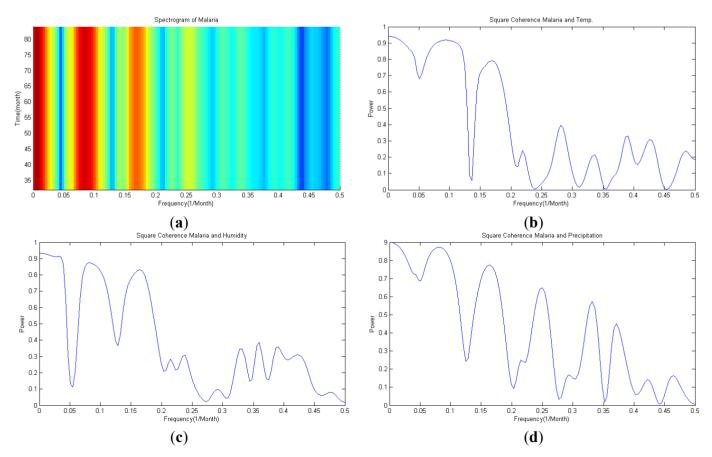
In Equation (11), XV is a linear combination of independent variables $X = [x_1, x_2, x_3, \ldots, x_n]$ and eigenvector X. It is a transformation of independent variable by multiplying the eigenvector so variables would be orthogonal and it is called principal components. Principal components are independent, so the multicollinearity issue is resolved and thus regression analysis can be conducted without loss of information.

4. Modeling of Malaria and Climate Variables

4.1. Nonlinear Regression Analysis

Regression analysis was conducted on monthly malaria occurrences and climate variables for the period between 2001 and 2008. Nonlinear regression was applied to climate variables and reported malaria occurrence counts to conduct regression analysis. Many recent studies [51,52] that studied malaria in Korea and climate variables considered the time lag effect so it was assumed that there is correlation in accordance with the time lag. So, autospectral analysis is in accordance with Figure 2a, and it also represented coherency analysis between each climate variable and malaria occurrences in Figure 2b–d.

Figure 2. (**a**) Spectrogram of malaria; (**b**) coherency of Temp. and malaria; (**c**) coherency of humidity and malaria; (**d**) coherency of precipitation and malaria.

As the autospectrum result of malaria shows in Figure 2a, the malaria occurrence time series seems to have a 12 month strong frequency (=0.083) and a 6 month weak frequency (=0.167). Figure 1a confirms that the malaria occurrence series possesses an annual period. Looking at the squared coherency from Figure 2b–d, the frequency of each climate variable and malaria can be defined. Coherency between malaria and average temperature shows a high coherency in 12 and 6 months (see Figure 2b), and the same can be said for humidity (see Figure 2c). Lastly, for malaria and precipitation, high coherency is shown in 12, 6 and 4 month periods. Considering that coherency explains the correlation level exhibited in every time series, each coherency seems to be the time lag effect for respective coherency. Thus, in this study, the time lag effect of respective period was considered when doing regression analysis.

Also, when data is analyzed, they are assumed random variables and it is assumed that each variable is independent when doing regression analysis. So, tests for randomness and nonlinearity were conducted for the malaria occurrence series. In this research, four commonly used nonparametric tests (Anderson Correlogram, Run Test, Spearman Rank, Correlation Coefficient and Turning point test) that Salas *et al.* [53] suggested and BDS statistics test were used.

The Anderson correlogram and Spearman test judged data as random data as shown in Table 1, but run test and turning point inferred otherwise. Statistics of BDS (m) showed that it exceeded the significance level in every dimension and this shows malaria occurrence series possessed *nonlinear chaotic* or *stochastic property*. Given this, when constructing the regression model for malaria occurrences, regression analysis was conducted with consideration for nonlinearity. Results of multiple regression analysis of malaria occurrences based on nonlinear regression analysis are shown in Figure 3a and Equation (12).

$$y = 1164 - 11.76 x_1 + 9.25 x_2 - 4.24 x_3 + 7.05 x_4 - 16.75 x_5 - 25.90 x_6 - 0.34 x_7$$
$$+ 0.0076 x_8 + 0.011 x_9 - 0.14 x_{10} + 1.03 x_1^2 - 0.39 x_2^2 + 0.063 x_3^2 - 0.038 x_4^2 \quad (12)$$
$$+ 0.11 x_5^2 + 0.21 x_6^2 + 0.0005 x_7^2 + 1.6 \times 10^{-5} x_8^2 + 1.2 \times 10^{-5} x_9^2 + 0.00019 x_{10}^2$$

where y : monthly malaria occurrence (person); x_1, x_2, x_3 : monthly mean temperature(℃) with 0, 6, 12 month lags; x_4, x_5, x_6 : monthly relative humidity(%) with 0, 6, 12 month lags; and x_7, x_8, x_9, x_{10} : monthly total precipitation(mm) with 0, 6, 12, 4 month lags.

Table 1. Test statistics of malaria time series (2001–2011).

Test method	Test statistic	95% C. I.	Randomness Check
Run Test	−7.3261	[−1.96, +1.96]	X
Anderson	0.1741	[−1.65. +1.65]	O
Spearman	0.5760	[−1.96, +1.96]	O
Turning Point	−11.9868	[−1.96, +1.96]	X
BDS(2)	10.5150	[−1.96, +1.96]	X
BDS(3)	9.7895	[−1.96, +1.96]	X
BDS(4)	9.2964	[−1.96, +1.96]	X
BDS(5)	8.9231	[−1.96, +1.96]	X

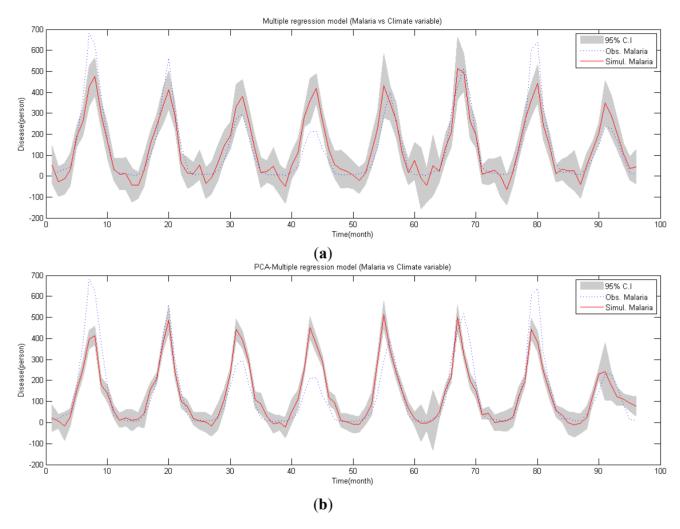**Figure 3.** Malaria simulation (2001–2008) (**a**) Malaria simulation by multiple regression; (**b**) malaria simulation by PCA-regression.

(a)

(b)

Results of multiple regression analysis show that the coefficient of determination R^2 of multiple regression model was 0.805, and the modified coefficient of determination R^2 was 0.753 and for the regression analysis model, F = 15.5. The significance probability was 4.57×10^{-19}, and statistical significance did exist (see Figure 3a). However, according to Figure 1a, it is highly likely that there is coherency between data, and if this is the case, regression model drawn from regression analysis may be under the influence of multicollinearity. Thus, analysis was conducted based on malaria occurrences and each of the Pearson, Kendall and Spearman correlation coefficients; it revealed that each variable had a high correlation, from 0.52 between average temperature and humidity to 0.78 between average temperature and malaria (see Figure 4). Thus, there is clearly a strong correlation between each climate variable and there could be an error when building a regression model for malaria through multiple regression analysis.

4.2. PCA-Regression Analysis

To conduct principal component regression analysis between malaria occurrences and climate variables, principal component regression was conducted on average temperature, average humidity and precipitation. Principal components scores up to principal components 1, 2, and 3 were given—

2.9698, 3.0490 and −0.6665, respectively. Principal components can be narrowed down to three with the condition of 80% cumulative contribution rate; the table of principal components analysis is shown in Figure 5.
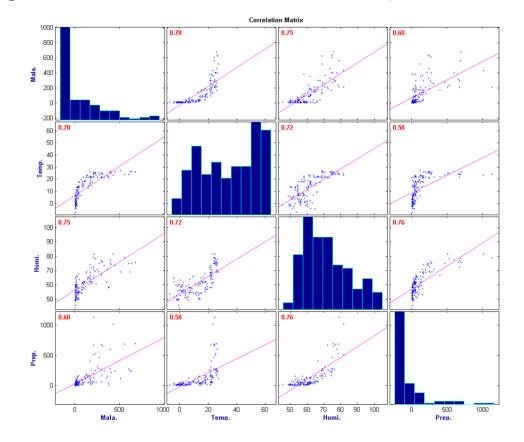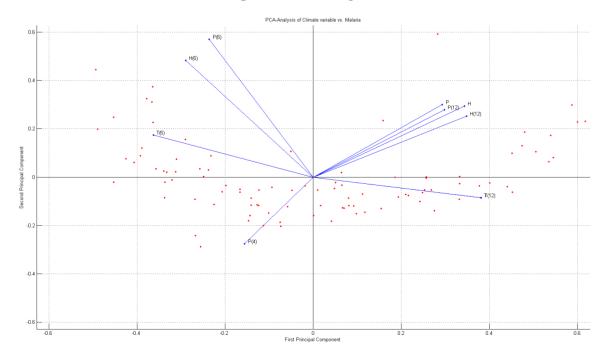
Figure 4. Correlation matrix of malaria and climate variable (Pearson's correlation).

Figure 5. PCA-diagram.

Principal components were categorized into the following: first principal components were average temperature (x_1) and average humidity (x_5, x_6); second principal components were average humidity (x_6) and precipitation (x_7); and third principal components were average temperature (x_1), average humidity (x_5, x_6) and precipitation (x_7). Principal component regression analysis can be conducted based on principal component analysis and principal component regression (PCA-regression) model of malaria occurrences is shown in Eqation (13):

$$y = 93.27 - 52.51 z_1 + 4.77 z_2 - 19.62 z_3 + 6.58 z_1^2 - 0.54 z_2^2 + 5.28 z_3^2$$

z_1 : 1st principal component,

$$z_1 = 0.38 x_1 - 0.08 x_2 + 0.05 x_3 - 0.02 x_4 + 0.46 x_5$$
$$+ 0.37 x_6 + 0.07 x_7 + 0.09 x_8 - 0.04 x_9 - 0.69 x_{10}$$

z_2 : 2nd principal component

$$z_2 = -0.36 x_1 + 0.17 x_2 + 0.08 x_3 - 0.01 x_4 - 0.14 x_5$$
$$+ 0.32 x_6 + 0.78 x_7 + 0.14 x_8 - 0.27 x_9 - 0.02 x_{10}$$

z_3 : 3rd principal component

$$z_3 = 0.38 x_1 - 0.08 x_2 + 0.06 x_3 - 0.001 x_4 - 0.46 x_5$$
$$+ 0.30 x_6 + 0.15 x_7 - 0.05 x_8 - 0.10 x_9 - 0.7 x_{10}$$

(13)

Results of malaria occurrence simulation based on the principal component regression model are compared in Figure 3. The coefficient of determination R^2 of principal components regression model was 0.743 and the modified coefficient of determination R^2 was 0.725. In the regression analysis model, F = 4.28 and significance probability was 3.4×10^{-24}, thus statistical significance was present (See Figure 3b). Also, since each principal component is independent, the regression model can be established without the multicollinearity problem. Thus, principal components regression analysis is quite plausible for simulation of malaria occurrences.

4.3. Validation of Malaria Model

To evaluate the suitability of malaria occurrence model calculated from PCA, evaluation of regression model was conducted using the malaria occurrence data from 2009–2011 and the results are shown in Figure 6.
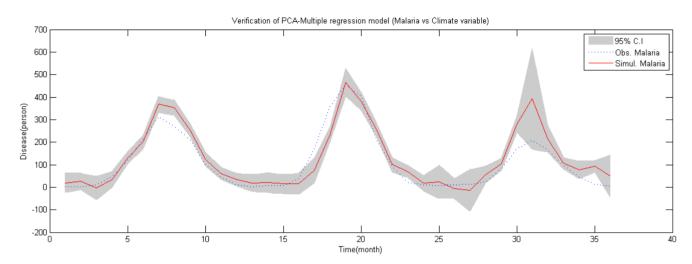
Results of evaluation using principal component regression analysis show that the coefficient of determination of PCA-regression model R^2 was 0.852 and the NRMSE (Normalized Root Mean Square Error) index, which shows the efficiency of the model suggested by Scott and Fred [54], was 0.117 and the RE (Relative Error; [36]) index was 0.026. It can be concluded that malaria occurrence time series simulated by the model does have significance with actual observation data.
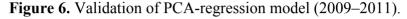
5. Malaria Occurrence under Climate Change

5.1. Climate Change Scenario

To simulate the condition of climate change using the malaria occurrence regression model as suggested in Section 4, future climate data is needed. In this study, future climate data was acquired

through a climate change scenario and a climate model and then we analyzed the trend of malaria occurrences. For climate change scenario, future carbon dioxide concentration scenario was used as the boundary condition of the climate model and many different situations were assumed and applied. RCP (Representative concentration pathways) scenario, which will appear in the IPCC 5th assessment report released in 2013/2014, is the most studied scenario [55–62], and in this research, the RCP scenario was applied.

Figure 6. Validation of PCA-regression model (2009–2011).

The RCP scenario is categorized into four different scenarios depending on the carbon dioxide reduction degree [57] (see Table 2). In this study, the SRES A1B scenario from the existing SRES AR4 scenario study [63], which counterparts RCP 4.5, and seems to conform with reality the most, was analyzed. Also, another RCP scenario study [64] presented that RCP 4.5 will most effectively reflect future climate in accordance with the present condition, so RCP 4.5 was applied.
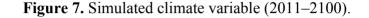
Table 2. RCP(Representative concentration pathways) scenario description.

Senarios	Description	CO density (ppm)
RCP 2.6	Peak in radiative forcing at ~3 W/m before 2100 year and then decline	490
RCP 4.5	Stabilization without overshoot pathway to ~4.5 W/m at stabilization after 2100 year	650
RCP 6.0	Stabilization without overshoot pathway to ~6 W/m at stabilization after 2100 year	850
RCP 8.5	Rising radiative forcing pathway leading to 8.5 W/m by 2100 year	1370

A numerical model which simulates the actual climate phenomenon using climate change scenario as the boundary condition was used for the climate model. The Korean government uses a local climate model HadGEM3-RA based on HadGEM2-AO model of Hadley Centre of British Met Office [65]. Monthly climate data from 2011–2100 according to RCP 4.5 climate change scenario and HasGEM3-RA climate model are shown in Figure 7.

5.2. Future Malaria Simulation and Analysis

The study simulated the future malaria occurrence using monthly climate data from 2011–2100 according to RCP 4.5 climate change scenario and the HadGEM3-RA climate model and malaria regression model. In the simulation data, 676 occurrences in July 2040 and 695 occurrences in July 2089 appeared to be the biggest incidents (Figure 8).
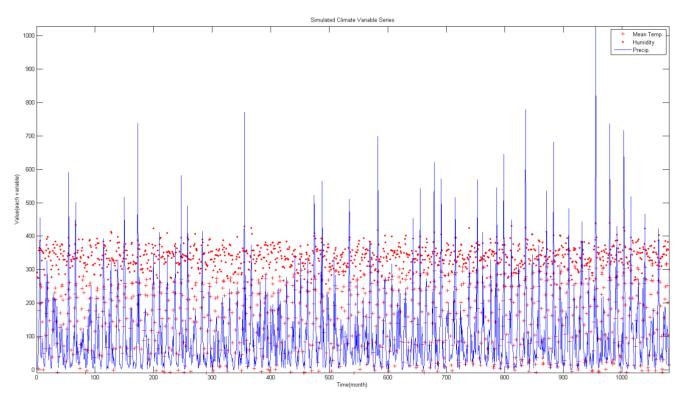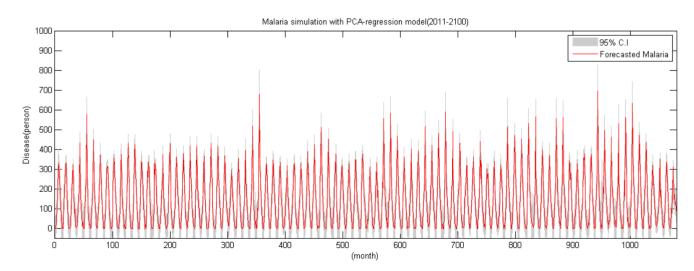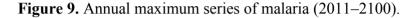
Figure 7. Simulated climate variable (2011–2100).

Figure 8. Simulated malaria time series (2011–2100).

There seems to be a general trend in accordance with the present state, but it shows a slight increase. When the annual maximum series is generated with the maximum malaria occurrence as

a criterion, the trend becomes highly visible. As shown in Figure 9, simulation shows that annual maximum occurrence per year increases as time passes.

   Also, there is some change in the annual occurrence trend. From 20010–2011, August was the month of maximum occurrence, but simulation data of 2011–2100 predicts July will be the month of maximum occurrence. In addition, compared to the spike of occurrence between June and August in the existing data, the occurrence will continually increase between April and July in the future (see the Figure 10).
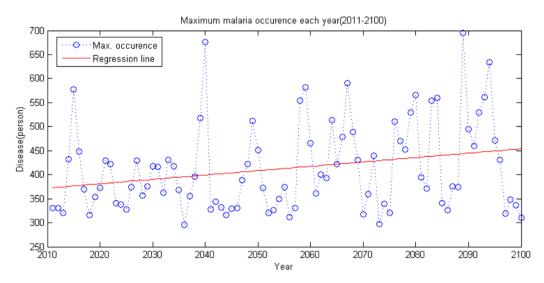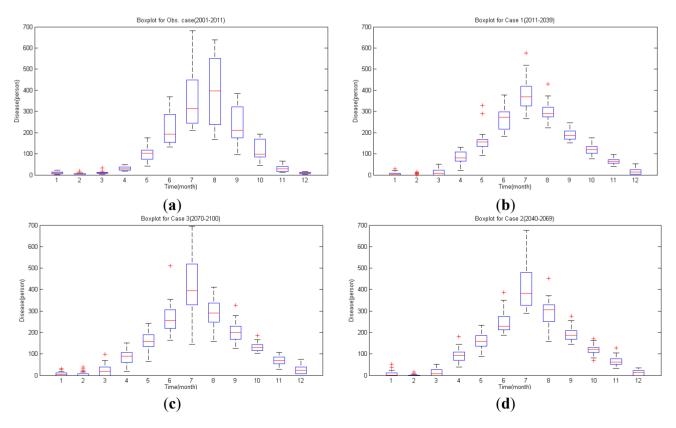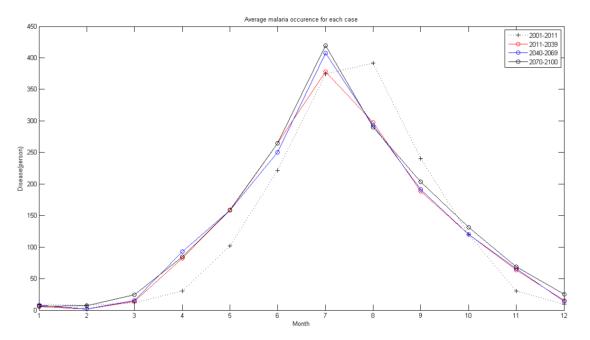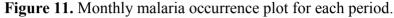
Figure 9. Annual maximum series of malaria (2011–2100).

Figure 10. Malaria occurrence boxplot for each period 2001–2011 year (**b**) 2011–2039 year (**c**) 2040–2069 year (**d**) 2070–2100 year.

(**a**)

(**b**)

(**c**)

(**d**)

In Figure 11, the monthly average malaria occurrence is shown conclusively. Compared to the concentrated occurrence between May and August, simulation shows that the occurrence will start in April and will reach its maximum in July in the future. Thus, in the future, preparation for malaria outbreak must be executed earlier than the present guideline in which preparation starts after the rainy season in summer.

Figure 11. Monthly malaria occurrence plot for each period.

6. Conclusions

The study clarifies correlation between monthly climate data and malaria infection occurrence and establishes a regression model. Also, the future malaria occurrence up to year 2100 is simulated using future climate data generated for a climate change scenario and climate model and analysis on malaria occurrence time series. Results drawn from this study can be concluded as below:

1. Correlation between malaria occurrence and monthly average temperature, relative humidity and precipitation data is analyzed with time lag effect between malaria occurrence and climate variables using spectral analysis between each variable. A strong coherency between each climate variable data is clear, thus regression model is analyzed under the influence of multicollinearity. To resolve this issue, principal component regression analysis based on PCA is used to establish a regression model. Using the regression model, malaria infection occurrences from 2009–2011 are tested and coefficient of determination R^2 is 0.852, NRSE is 0.117 and RE is 0.026, which clearly accounts for malaria infection.

2. By applying climate data between 2011 and 2100 using the RCP 4.5 climate change scenario and the CNCM3 climate model to the regression model, future malaria occurrence is simulated. Analysis of simulated data shows the malaria occurrence trend in general will gradually increase. Also, in the future, the occurrence period will diminish and it shows an increase of malaria occurrence before the rainy season in summer; thus, adaptation in the malaria occurrence response plan of Korea is needed.

Acknowledgments

Author Contributions

This research presented here was carried out in collaboration between all authors. Jaewon Kwak and Hung Soo Kim had the original idea for the study. Soojun Kim, Duckgil Kim and Keonhaeng Lee designed/conducted the research methods. Huiseong Noh, Seung Jin Hong and Narae Kang conducted data analysis. Vijay P. Singh and Hung Soo Kim contributed to the writing of the paper. All authors discussed the structure and comment on the manuscript at all stages.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Jang, J.Y.; Cho, S.H.; Kim, S.Y.; Cho, S.N.; Kim, M.S.; Baek, K.W. *Assessment of Climate Change Impact and Preparation of Adaptation Program in Korea*; Ministry of Environment: Seoul, Korea, 2003.
2. Patz, J.A.; Olson, S.H. Malaria risk and temperature: Influences from global climate change and local land use practices. *Proc. Natl. Acad. Sci.* **2006**, *103*, 5635–5636.
3. Poveda, G.; Rojas, W.; Qui.nones, M.L.; Velez, I.D.; Mantilla, R.I.; Ruiz, D.; Zuluaqa, J.S.; Rua, G.L. Coupling between annual and ENSO timescales in the malaria-climate association in Colombia. *Environmental health perspectives*, **2001**, *109*, 489–493.
4. Craig, M.H.; Snow, R.W.; Le Sueur, D. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology today*, **1999**, *15*, 105–110.
5. Paaijmans, K.P.; Read, A.F.; Thomas, M.B. Understanding the link between malaria risk and climate. *Proc. Natl. Acad. Sci.* **2009**, *106*, 13844–13849.
6. Paaijmans, K.P.; Blanford, S.; Bell, A.S.; Blanford, J.I.; Read, A.F.; Thomas, M.B. Influence of climate on malaria transmission depends on daily temperature variation. *Proc. Natl. Acad. Sci.* **2010**, *107*, 15135–15139.
7. Olson, S.H.; Gangnon, R.; Elguero, E.; Durieux, L.; Gu, L.; Foley, J.A.; Patz, J.A. Links between climate, malaria, and wetlands in the Amazon Basin. *Emerg. Infect. Dis.* **2009**, *15*, doi: 10.3201/eid1504.080822.
8. Kuhn, K.; Campbell-Lendrum, D.; Haines, A.; Cox, J.; CorvalHa, C.; Anker, M.; Malaria, R.B. *Using Climate to Predict Infectious Disease Epidemics*; World Health Organization: Geneva, Switzerland, 2005.
9. Thomson, M.C.; Doblas-Reyes, F.J.; Mason, S.J.; Hagedorn, R.; Connor, S.J., Phindela, T.; Palmer, T.N. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **2006**, *439*, 576–579.

10. Jhajharia, D.; Chattopadhyay, S.; Choudhary, R.; Dev, V.; Singh, V.P.; Lal, S. Influence of climate on incidences of malaria in the Thar desert, northwest India. *Int. J. Climatol.* **2013**, *33*, 312–325.

11. Martens, W.J.; Niessen, L.W.; Rotmans, J.; Jetten, T.H.; McMichael, A.J. Potential impact of global climate change on malaria risk. *Environ. Health Perspect.* **1995**, *103*, 458–464.

12. Kleinschmidt, I.; Bagayoko, M.; Clarke, G.P.Y.; Craig, M.; Le Sueur, D. A spatial statistical approach to malaria mapping. *Int. J. Epidemiol.* **2000**, *29*, 355–361.

13. Minakawa, N.; Sonye, G.; Mogi, M.; Githeko, A.; Yan, G. The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. *J. Med. Entomol.* **2002**, *39*, 833–841.

14. Small, J.; Goetz, S.J.; Hay, S.I. Climatic suitability for malaria transmission in Africa. *Proc. Natl. Acad. Sci.* **2003**, *100*, 15341–15345.

15. Zhou, G.; Minakawa, N.; Githeko, A.K.; Yan, G. Association between climate variability and malaria epidemics in the East African highlands. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2375–2380.

16. Martens, P.; Kovats, R.S.; Nijhof, S.; De Vries, P.; Livermore, M.T.J.; Bradley, D.J.; McMichael, A.J. Climate change and future populations at risk of malaria. *Global Environmental Change*, **1999**, *9*, S89–S107.

17. Hay, S.I.; Cox, J.; Rogers, D.J.; Randolph, S.E.; Stern, D.I.; Shanks, G.D.; Snow, R.W. Climate change and the resurgence of malaria in the East African highlands. *Nature* **2002**, *415*, 905–909.

18. Loevinsohn, M.E. Climatic warming and increased malaria incidence in Rwanda. *Lancet* **1994**, *343*, 714–718.

19. Patz, J.A.; Campbell-Lendrum, D.; Holloway, T.; Foley, J.A. Impact of regional climate change on human health. *Nature* **2005**, *438*, 310–317.

20. Rohr, J.R.; Dobson, A.P.; Johnson, P.T.; Kilpatrick, A.M.; Paull, S.H.; Raffel, T.R.; Thomas, M.B. Frontiers in climate change–disease research. *Trends Ecol. Evol.* **2011**, *26*, 270–277.

21. Ermert, V.; Fink, A.H.; Morse, A.P.; Jones, A.E.; Paeth, H.; Di Giuseppe, F.; Tompkins, A.M. Development of dynamical weather-disease models to project and forecast malaria in Africa. *Malar. J.* **2012a**, *11*, doi:10.1186/1475-2875-11-S1-P133.

22. Ermert, V.; Fink, A.H.; Morse, A.P.; Paeth, H. The impact of regional climate change on malaria risk due to greenhouse forcing and land-use changes in tropical Africa. *Environ. Health Perspect.* **2012b**, *120*, 77–84.

23. Van Lieshout, M.; Kovats, R.S.; Livermore, M.T.J.; Martens, P. Climate change and malaria: Analysis of the SRES climate and socio-economic scenarios. *Glob. Environ. Change* **2004**, *14*, 87–99.

24. Tanser, F.C.; Sharp, B.; Le Sueur, D. Potential effect of climate change on malaria transmission in Africa. *Lancet* **2003**, *362*, 1792–1798.

25. Bhattacharya, S.; Sharma, C.; Dhiman, R.C.; Mitra, A.P. Climate change and malaria in India. *Curr. Science-Bangalore* **2006**, *90*, 369–375.

26. Haines, A.; Kovats, R.S.; Campbell-Lendrum, D.; Corvalan, C. Climate change and human health: impacts, vulnerability and public health. *Public Health* **2006**, *120*, 585–596.

27. Gage, K.L.; Burkot, T.R.; Eisen, R.J.; Hayes, E.B. Climate and vectorborne diseases. *Am. J. Prev. Med.* **2008**, *35*, 436–450.

28. Costello, A.; Abbas, M.; Allen, A.; Ball, S.; Bell, S.; Bellamy, R.; Patterson, C. Managing the health effects of climate change. *Lancet* **2009**, *373*, 1693–1733.

29. Greer, A.; Ng, V.; Fisman, D. Climate change and infectious diseases in North America: the road ahead. *Can. Med. Assoc. J.* **2008**, *178*, 715–722.

30. Gething, P.W.; Smith, D.L.; Patil, A.P.; Tatem, A.J.; Snow, R.W.; Hay, S.I. Climate change and the global malaria recession. *Nature* **2010**, *465*, 342–345.

31. Parham, P.E.; Michael, E. Modeling the effects of weather and climate change on malaria transmission. *Environ. Health Perspect.* **2010**, *118*, 620–626

32. Korea Centers for Disease Control and Prevention. Available online: http://www.cdc.go.kr/ (accessed on 10 October 2014).

33. Kho, W.G. Reemergence of malaria in Korea. *J. Korean Med. Assoc.* **2007**, *50*, 959–966.

34. Epstein, P.R.; Diaz, H.F.; Elias, S.; Grabherr, G.; Graham, N.E.; Martens, W.J.; Susskind, J. Biological and physical signs of climate change: focus on mosquito-borne diseases. *Bull. Am. Meteorol. Soc.* **1998**, *79*, 409–417.

35. Chung, Y.S.; Yoon, M.B.; Kim, H.S. On climate variations and changes observed in South Korea. *Clim. Change* **2004**, *66*, 151–161.

36. Golub, G.H.; Van Loan, C.F. *Matrix Computation*, 3rd ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 1996.

37. Strang, G. *Linear Algebra and Its Applications*, 9th ed.; Thomsom Leaning: London, UK, 1998.

38. Dennis, W.R. *Echo Signal Processing*, 1st ed.; Springer: New York, NY, USA, 2003.

39. Brock, W.A.; Heish, D.A.; Lebaron, B. *Nonlinear Dynamics, Chaos, and Instability: Statiscal Theory and Economic Evidence*, 1st ed.; The MIT Press: Cambridge, MA, USA, 1991.

40. Brock, W.A.; Dechert, W.D.; Scheinkman, J.A.; Lebaron, B. A test for independence based on the correlation dimension. *Econom. Rev.* **1996**, *15*, 197–235.

41. Kim, H.S.; Kang, D.S.; Kim, J.H. The BDS statistic and residual test. *Stoch. Environ. Res. Risk Assess.* **2003**, *17*, 104–115.

42. Grassberger, P.; Procaccia, I. Measuring the strangeness of strange attractors. *Phys. D Nonlinear Phenom.* **1983**, *9*, 189–208.

43. Very Fast and Correctly Sized Estimation of the BDS Statistic. Available online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=151669 (accessed on 10 October 2014).

44. Gujarati, D. Multicollinearity: What happens if the regressors are correlated? In *Basic Econometrics*, 4th ed.; McGraw-Hill: New York, NY, USA, 2002.

45. Draper, N.; Smith, H. *Applied Regression Analysis*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1998.

46. Chatterjee, S.; Hadi, A.S.; Price, B. *Regression Analysis by Example*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2000.

47. Montgomery, D.C.; Peak, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2001.

48. Kutner, M.; Nachtsheim, C.; Neter, J.; Li, W. *Applied Linear Statistical Models*, 5th ed.; McGraw Hill: New York, NY, USA, 2005.

49. Morison, D. F. *Multivariate Statistical Methods*, 4th ed.; Thomsom Leaning: London, UK, 2005.

50. Lin, M.; Wei, L. The small sample properties of the principal components estimator for regression coefficients. *Commun. Statistic-Theory Methods* **2002**, *31*, 271–283.

51. Shin, H.S.; Yun, S.M.; Jung, K.H.; Lee, S.H. Climate change, food-borne disease prediction, and future impact. *Health Soc. Welf. Rev.* **2011**, *29*, 217–237.

52. Kim, S.H.; Jang, J.Y. Correlations between climate change-related infectious diseases and meteorological factors in Korea. *J. Prev. Med. Public Health* **2010**, *43*, 436–444.

53. Salas, J.D.; Smith, R.A.; Tabios, G.Q., III; Heo, J.H. Colorado State University, Fort Collins, CO, USA. Unpublished work, 1993.

54. Armstrong, J.S.; Collopy, F. Error measures For generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast.* **1992**, *8*, 69–80.

55. Moss, R.H.; Edmonds, J.A.; Hibbard, K.A.; Manning, M.R.; Rose, S.K.; van Vuuren, D.P.; Wilbanks, T.J. The next generation of scenarios for climate change research and assessment. *Nature* **2010**, *463*, 747–756.

56. Meinshausen, M.; Smith, S.J.; Calvin, K.; Daniel, J.S.; Kainuma, M.L.T.; Lamarque, J.F.; van Vuuren, D.P.P. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **2011**, *109*, 213–241.

57. Van Vuuren, D.P.; Edmonds, J.; Kainuma, M.; Riahi, K., Thomson, A.; Hibbard, K., Rose, S.K. The representative concentration pathways: An overview. *Clim. Change* **2011**, *109*, 5–31.

58. Arora, V.K.; Scinocca, J.F.; Boer, G.J.; Christian, J.R.; Denman, K.L.; Flato, G.M.; Merryfield, W.J. Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.* **2011**, *38*, doi:10.1029/2010GL046270.

59. Reisinger, A.; Meinshausen, M.; Manning, M. Future changes in global warming potentials under representative concentration pathways. *Environ. Res. Lett.* **2011**, *6*, 1–8.

60. Thomson, A.M.; Calvin, K.V.; Smith, S.J.; Kyle, G.P.; Volke, A.; Patel, P.; Edmonds, J.A. RCP4. 5: A pathway for stabilization of radiative forcing by 2100. *Clim. Change* **2011**, *109*, 77–94.

61. Emanuel, W.R.; Janetos, A.C. *Implications of Representative Concentration Pathway 4.5 Methane Emissions to Stabilize Radiative Forcing* (No. PNNL-22203); Pacific Northwest National Laboratory (PNNL): Richland, WA, USA, 2013.

62. Good, P.; Gregory, J.M.; Lowe, J.A.; Andrews, T. Abrupt CO_2 experiments as tools for predicting and understanding CMIP5 representative concentration pathway projections. *Clim. Dyn.* **2013**, *40*, 1041–1053.

63. Kwon, Y.; Kwon, W.; Boo, O. Future projections on the change of onset date and duration of natural seasons using SRES A1B data in South Korea. *J. Korean Geogr. Soc.* **2007**, *42*, 835–850.

64. So, B. J.; Kim, M. J.; Kwon, H. H. Evaluation of the next generation scenarios for climate change of KMA. *Mag. Korea Water Res. Assoc.* **2012**, *45*, 56–70.

65. Regional climate projection for East Asia and Korea using the HadGEM3-RA. Available online: http://meetingorganizer.copernicus.org/3ICESM/3ICESM-236-2.pdf (accessed on 10 October 2014).