

기후 인자를 이용한 한국의 말라리아 모형화 연구

곽 재 원

(웨이백대학교)

한 희 찬

(인하대학교)

이 종 소*

(인하대학교)

김 형 수

(인하대학교)

말라리아는 기후에 크게 영향을 받는 법정전염병 3군으로서 1993년 휴전선에서 재 발생한 이후에 매년 환자가 발생하고 있다. 본 연구의 목적은 말라리아 발생에 있어서 기후 인자에 따른 말라리아의 발생자수 예측 모형을 구축하는데 있다. 이를 위하여 2001년부터 2011년까지의 국내의 월별 말라리아 발생사례와 평균 기온, 습도, 강우량 등의 기후인자를 수집하였다. 수집된 자료에 대하여 스펙트럼 분석을 통하여 시간지연 효과를 고찰하고, 자료 간 다중공선성을 고려한 주성분 회귀분석을 통하여 회귀모형을 구축하였다. 2009년부터 2011년 사이의 말라리아 감염병 발생자료를 통하여 검증한 결과, 구축된 회귀 모형이 말라리아 발생자료에 대한 설명력이 있는 것으로 나타났다.

주요용어: 말라리아, 기후인자, 스펙트럼 분석, 주성분 회귀분석

이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2013R1A6A3A03027285)

* 교신저자: 이종소, 인하대학교(stynrehero@naver.com)

■ 투고일 2013.6.11 ■ 수정일: 2013.11.26 ■ 게재확정일: 2013.12.9

I. 서론

말라리아는 Plasmodium 속 원충이 적혈구와 간 세포 내에 기생함으로써 발병되는 급성 열성 감염증으로, 매년 약 3-5억의 말라리아 환자가 발생하며 그 중 100만 명이상이 사망하고 있는 중요한 기생충 감염으로 세계보건기구(WHO)가 선정한 6대 열대병 중에서도 가장 중요한 질환으로 인정되고 있다. 전 세계적으로는 삼일열원충과 열대열원충이 95% 이상이며, 우리나라의 말라리아는 삼일열원충(Plasmodium vivax) 감염이 대다수를 차지하고 있다. 일반적으로, 매개체를 통하여 발생하는 감염병은 생태계 내에서의 상호작용에 영향을 받기 때문 이에 영향을 미치는 기후 인자에도 큰 영향을 받는 것으로 알려져 있다(장재연 외, 2003). 특히, 모기를 매개로 하는 질병(Mosquito-borne disease)은 기후의 영향을 크게 받으며, 기온, 강수량, 습도 등이 중요한 영향을 미치는 것으로 알려져 있으며(박윤형 등, 2006), 그 중에서도 말라리아는 매개체에 의한 전염병 중 가장 중요하게 다루어지는 전염병인 동시에 기후변화에 가장 민감한 질병으로 알려져 있다(Patz et al., 2006).

이 때문에 말라리아와 기후 인자 및 기후변화를 연관 지으려는 연구가 다수 수행되어 왔다. Poveda et al(2001) 등은 미국 콜롬비아 지방에서 엘니뇨를 통하여 말라리아와 기후와의 시간적인 지연효과를 고찰한 바 있으며, 사하라 지방에서 기후 인자의 확률분포를 이용하여 말라리아와의 관계를 규명하려고 시도한 바 있다(Craig et al., 1999). 또한 평균 기온에 대하여 중점적으로 관심을 두고 상관관계와 민감도를 분석하기도 하였으며(Paaijmans et al., 2009), 일 단위의 기온 변화가 말라리아에 주는 영향을 세부적으로 분석하기도 하였다(Paaijmans et al., 2010). 특히, 최근에는 습지와 기후 인자와 말라리아의 상관관계를 분석하는 등의 다각적인 분석도 이루어지고 있으며(Olson et al., 2009), 기후 인자를 통한 충격함수에 따른 말라리아 예측을 하거나(Kuhn et al., 2005), 계절적 기후인자에 기반한 말라리아 감염 예측 연구도 수행되었다(Thomson et al., 2006). 이와 같이 현재까지의 말라리아와 기후 관련 연구는 주로 기후 인자와 말라리아와의 상관관계를 규명함으로써 말라리아의 발생에 대한 예측을 주로 수행하였다(Martens, et al., 1995; Kleinschmidt et al., 2000; Minakawa et al., 2002; Small et al., 2003; Zhou et al., 2004). 또한, 최근에 기후변화가 현실적인 문제로서 부각되면서 말라리아의 발생 문제에 대한 연구도 다각적으로 수행되고 있다. 지난 1900년대 초부터

기후의 변화에 따라서 말라리아의 변화 양상을 기후인자와 연관시켜서 분석한 연구도 수행된 바 있으며(Hay et al., 2002), 유사하게 동 아프리카의 지난 50년간의 온난화 경향과 말라리아 취약성을 분석한 연구(Patz et al., 2006), 지역적인 기후변화로 인하여 말라리아를 포함한 공중보건 측면에서 미래의 변화양상을 연구하기도 하였다(Patz et al., 2005). 기후온난화 측면에서도 아프리카 르완다 지역의 기후 온난화에 따른 말라리아의 증가 양상을 분석하는 연구나(Loevinsohn et al., 1994), SRES 기후변화 시나리오에 따른 말라리아의 증감 양상을 예측하기도 하였다(Van Lieshout et al., 2004). 이와 같은 기후변화와 말라리아에 대한 연구는 주로 향후 미래에 발생할 기후변화에 대하여 말라리아 감염병이 어떠한 발생 추이를 보일 것인가를 중점적으로 다루고 있다.(Martens et al., 1999; Tanser et al., 2003; Bhattacharya et al., 2006; Haines et al., 2006; Gage et al., 2008; Costello et al., 2009; Greer et al., 2009; Gething et al., 2010; Parham et al., 2010).

본 연구의 목적은 말라리아의 발생과 기후 인자와의 상관관계를 살펴보고, 이를 바탕으로 기후 인자에 따른 말라리아의 발생을 모의할 수 있는 모형을 구축하고, 실제 적용성을 검증하여 공중 보건의 기초자료로서 제시하는 데 있다. 이를 위하여 2001년부터 2011년 까지의 월간 말라리아 감염병 발생 현황과 같은 기간의 기후 인자를 수집하여 분석을 수행하고 회귀 모형화 하여 향후 기후변화에 의한 감염병과 사회적인 영향을 위한 기초자료로서 제시하고자 한다.

II. 적용 이론

1. 말라리아 감염병의 특성

말라리아는 매년 약 3~5억의 말라리아 환자가 발생하며 세계보건기구(WHO)가 선정한 6대 열대병 중에서도 가장 중요한 질환으로, 우리나라에서도 법정전염병 제3군으로 지정된 중요한 감염병이다. 국내에서는 지난 2000년 이후 감소 추세를 보이던 국내 말라리아 환자발생(현역군인 포함)이 2005년 1,324명, 2006년 2,021명, 2007년 2,192명으로 2007년 국내 말라리아 환자는 전년대비 8.5%의 증가를 보이고 있다(질병관리본

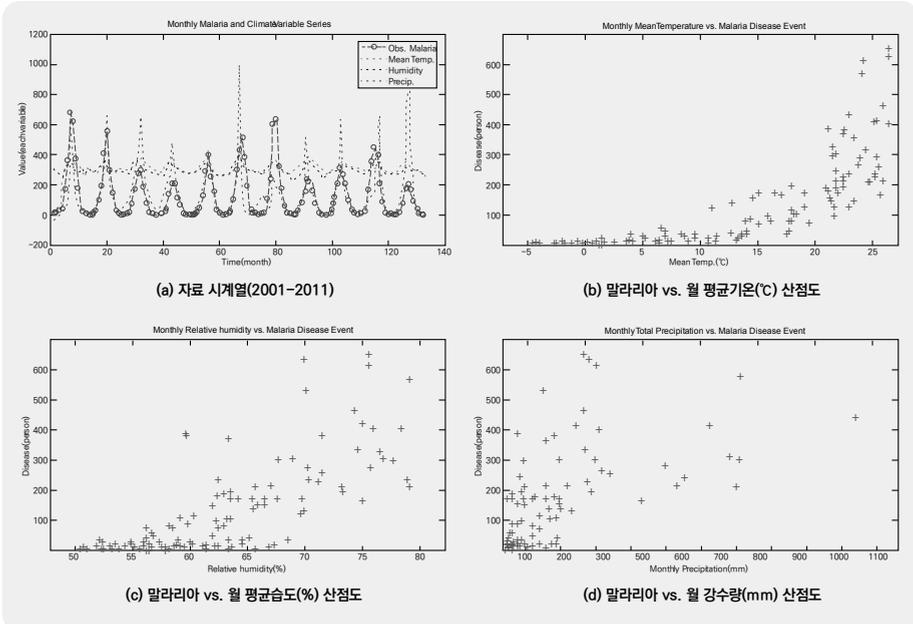
부, 2013). 이러한 말라리아의 전파는 얼룩날개모기속(Anopheline)에 속하는 암컷 모기를 통하여 감염되는 것으로 나타났다(고원규, 2007). 말라리아를 매개하는 모기는 일반적으로 따뜻한 기후로 갈수록 활동성이 증가하는 것으로 나타나고 있으므로(Epstein et al., 1998), 기후변화로 인하여 아열대 기후로 이동하고 있는 우리나라의 경우 지속적인 말라리아 감염병의 증가가 예상된다

2. 연구대상 및 자료

본 연구에서는 Jang et al(2003)의 연구에서 기후 인자 및 기후변화와 상관성이 높은 감염병으로 제시된 질병 중에 국내 발병자 수가 많아서 분석이 용이한 말라리아를 주요한 연구 감염병으로 선정하였다. 기존의 감염병 자료는 상대적으로 구축이 어려웠으나, 2000년도에 국가 전염병 데이터베이스가 구축된 이후로는 양질의 자료가 수집되고 있다. 말라리아의 발생 자료도 질병관리 본부의 2001년부터 2011년까지 신고된 월별 법정전염병 발병 자료를 사용하였으며, 자료 중에서 2001년부터 2008년까지의 자료를 보정 자료로, 2009년부터 2011년까지의 자료를 검증자료로서 사용하였다.

같은 기간의 기후자료는 서울 및 경기지역의 기상대에서 측정된 자료를 평균하여 사용하였다. 질병관리본부(2013)의 연구결과에 따르면, 2008년도 말라리아 환자 중 가장 많은 발생을 보인 지역은 경기(보고수 475명), 인천(보고수 92명), 서울(보고수 180명) 순으로 전체의 71.0%가 서울·경기지역에 집중되어 있었다. 따라서, 서울·경기 지역의 기상대의 기후인자를 평균한 값을 사용하였다. 말라리아와의 상관 분석을 위해서 포함된 기후인자는 평균기온(°C), 상대습도(%), 강수량(mm)을 사용 하였다. 또한, 기상 관측소 중에서 해발고도가 200m 이상인 고지대 지역의 기상측정소는 기후인자의 평균값을 왜곡시킬 수 있기 때문에 분석에서 제외하였으며, 기상청 무인자동관측소(Automatic Weather Station, AWS)도 자료 품질을 보장할 수 없으므로(기상청, 2013) 제외하였다. 수집된 2001년부터 2011년까지의 기상 및 말라리아 시계열은 [그림 1](a)와 같으며, 말라리아와 평균기온, 상대습도, 강수량과의 산점도 및 회귀분석 결과를 [그림 1](b)에서 [그림 1](d)까지 나타내었다. [그림 1]을 기준으로 판단할 때, 말라리아와 각각의 기후인자들은 양의 상관관계를 가지고 있는 것으로 판단된다.

그림 1. 자료시계열 및 말라리아-기후 인자 산점도(2001~2011년)



3. 분석방법론

가. 회귀분석(Regression analysis)

회귀분석은 독립변수와 종속변수 사이의 상관관계에 따른 선형적 관계식을 구하여 어떤 독립변수가 주어졌을 때 이에 따른 종속변수를 예측하거나, 수학적 모델이 얼마나 잘 설명하고 있는지를 판별하기 위한 적합도를 측정하는 분석 방법이다(Mogull, 2004). 일반적으로 회귀분석을 수행할 때 단변량으로 종속변수를 설명할 수 있을 경우에는 단순회귀분석, 다변량으로 설명할 수 있는 경우를 다중회귀분석이라고 한다. 따라서 한 개 이상의 독립변수를 이용하여 구축하는 회귀모형의 일반적인 형태는 식 (1)과 같다.

$$y = \beta X + \epsilon \quad (1)$$

β ; 독립변수의 집합

β_0 ; 회귀식의 y 절편
 β ; 독립변수의 계수 matrix
 ϵ ; 오차항

다중회귀분석에서 표본자료를 통한 회귀계수의 추정은 여러 가지 방법이 제시되어 있으나, 본 연구에서는 비선형 최소자승법을 사용하였다(Kelly, 1999). 또한, 회귀분석을 수행할 때에는 독립변수들 간에 강한 상관관계가 나타나는 다중공선성(Multicollinearity)을 고려하여야 한다. 이는 회귀분석이 독립변수 상호간에는 상관관계가 없다는 전제 가정을 위배하는 것이므로 적절한 회귀분석을 위해 변수의 선택을 통하여 해결하여야 한다(Golub & Loan, 1996; Strang, 1988).

나. 스펙트럼 분석(Spectral analysis)

자기 또는 교차 스펙트럼은 하나 또는 두 개의 자료의 도수(Frequency)가 가지는 분산의 분포를 설명하는 방법이다(Ricker, 2003). 어떠한 자료가 t 만큼의 시간 간격을 두고, k 만큼의 시차(lag)를 가질 때의 공분산 추정값은 자기공분산(autocovariance)를 사용하여 나타낼 수 있다.

$$cov_{xx}(k) = \frac{1}{N-k-1} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (2)$$

N ; 시계열의 개수

$t, \Delta t$; 시계열의 시간간격 및 시간

이러한 자기 공분산을 $x(t)$ 의 분산값 σ^2 으로 나누어 자기상관함수(autocorrelation function)을 얻을 수 있으며, 스펙트럼 함수 $X_{xx}(f)$ 는 자기상관함수를 복소 푸리에 변환시켜 얻는다. 이러한 스펙트럼 분석은 말라리아 시계열이 가지는 주기와 관련한 정보를 제공하여 준다. 교차 스펙트럼 분석의 경우에도 두 시계열 $x(t)$, $y(t)$ 에 대하여 교차 공분산을 대상으로 하며 k 만큼의 시차(lag)가 나도록 설정된 두 개의 시계열 자료 간의 상관된 분산을 판별하기 위한 방법이다. 일정 시간 간격 t 를 가지는 두 개의 시계열

$x(t)$, $y(t)$ 의 공분산 추정값 및 교차 스펙트럼(X_{xy})은 식 (3) 및 식 (4)와 같다.

$$cov_{xy}(k) = \frac{1}{N-k-1} \sum_{t=1}^{N-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) \quad (3)$$

$$|X_{xy}| = \sum_{k=0}^{Max\ Lag} \left(\frac{cov_{xy}(k)}{\sigma_x \sigma_y} \right) (k) e^{\frac{i2\pi f k}{f_s}} \quad (4)$$

또한, 두 개의 시계열 자료에 대하여 교차 스펙트럼을 이용하여 두 시계열의 간섭성(coherency)을 산정할 수 있으며 두 시계열 사이의 주기에 대한 정보를 취득할 수 있다(식 (5) 참조).

$$C_{xy} = \frac{|X_{xy}|^2}{X_{xx} X_{yy}} \quad (5)$$

C_{xy} ; 시계열의 간섭성

X_{xx}, X_{yy} ; 시계열의 자기상관밀도함수

일반적으로 두 시계열에 대한 간섭성은 $1 - C_{xy}$ 의 형태로 나타난다.

다. BDS 통계(Brock, Dechert, and Scheickman; BDS Statistics)

BDS 통계는 시계열 자료가 무작위한 분포를 따른다는 가설을 검증하기 위해 상관적분을 토대로 개발된 방법이다. 특히, 선형자료로 구분되는 무작위한 시계열 자료와 비선형 카오스 시스템 혹은 비선형 추계학적 시스템을 구분하는데 매우 유용한 통계기법이다(Brock et al., 1991, 1996; Kim et al., 2003). 시계열 자료가 무작위한 분포를 따른다는 가설 하에 $m > 1$ 인 경우의 BDS 통계는 다음과 같은 식(6)로 나타낼 수 있다.

$$BDS(m, N, r) = \frac{\sqrt{M}}{\sigma} [C(m, N, r) - C^m(1, M, r)] \quad (6)$$

$$\sigma^2(m, M, r) = 4m(m-1)C^{2(m-1)}(K-C^2) + K^m - c^{2m} \quad (7)$$

$$+ 2 \sum_{i=1}^{m-1} [C^{2i}(K^{m-i} - C^{2(m-i)}) - mC^{2(m-1)}(K-C^2)]$$

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

m ; 상관차원

M ; $N(m-1)$, 상관차원(m)에 대한 벡터

r ; 상관차원벡터의 반경

BDS 통계를 이용하여 시계열 자료를 무작위(Random) 자료인지 아니면 비선형 시스템(카오스, 비선형 추계학) 자료인지를 구분하기 위하여 유용하게 사용할 수 있다. BDS 통계를 사용하기 위해서는 m 및 r 값의 선택이 중요한데, 본 연구에서는 m 의 경우에는 $2 \leq m \leq 5$ 의 값과(Brook et al, 1991), r 의 경우에는 $0.5\sigma \leq r \leq 1.5\sigma$ (σ ; 표준편차)값을 사용하였다(Kanzler, 1999).

라. 주성분 회귀분석(Principal components regression)

회귀분석 과정에서 선택된 독립변수들 사이에 상관관계가 존재하여 다중공선성을 나타내는 경우, 일반적으로 독립변수의 일부를 제거하거나 새로운 관측치를 도입하기도 하며, 현장의 상황에 따라 상관관계를 제거하기도 한다(Gujarati, 2002). 그러나, 정보손실 측면에서 독립변수의 제거는 받아들이기 어렵기 때문에 대안으로서 주성분 회귀분석을 주로 사용하고 있다(Kutner et al., 2005; Montgomery et al., 2001; Chatterjee et al., 2000; Draper & Smith, 1998). 주성분 회귀분석은 원래의 변수보다 작은 개수의 잠재 주성분을 선택하여 차원을 축소함으로써 해석을 용이하게 하는 주성분분석(Morison, 2005)에 회귀분석을 결합한 개념으로서 독립변수의 주성분 점수를 산정하고 그에 따른 회귀분석을 수행함으로써 다중공선성 문제를 피하면서도 정보손실을 최소화하여 회귀분석을 수행하는 방법이다(Lin & Wei, 2002).

주성분 분석에서 상관계수 행렬의 고유치(λ_i) 행렬 A 와 고유벡터 V 는 다음의 식 (9)을 만족하게 된다.

$$V^T(Z^T Z) V = \Lambda, V^T V = V V^T = I \quad (9)$$

고유벡터는 직교행렬이기 때문에 $V V^T = I$ 가 성립하며 이를 식 (1)에 행렬 형태로 대입하면 식 (10)과 같은 주성분에 의한 회귀모형으로 정리될 수 있다.

$$y = X V V^T \beta + \epsilon \quad (10)$$

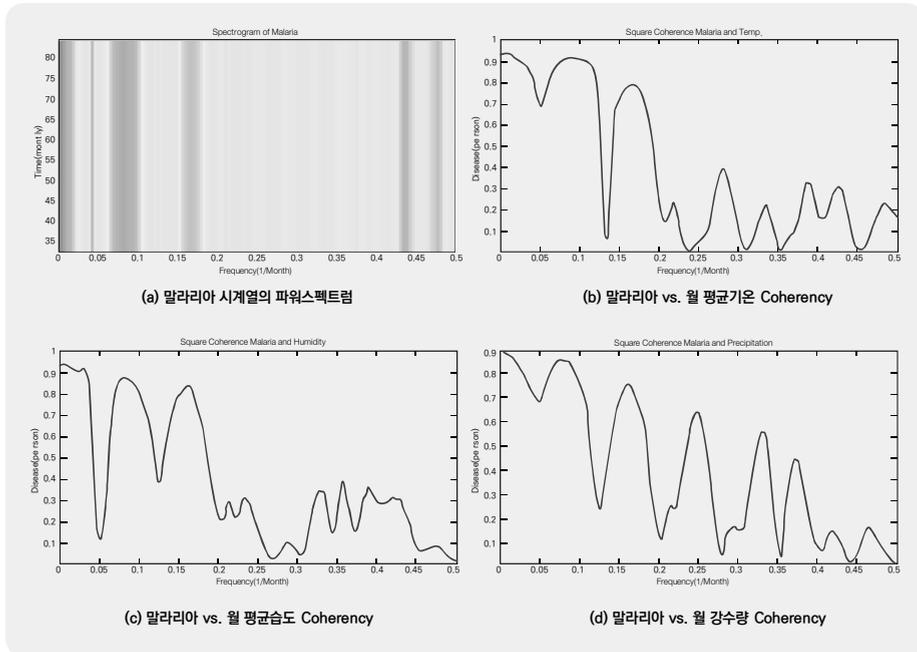
식 (10)에서 XV 는 독립변수 $X = [x_1, x_2, x_3, \dots, x_n]$ 과 고유벡터 V 의 선형 조합이며, 서로 연관성이 있는 독립변수 X 가 서로 직교(독립)하도록 고유벡터를 곱하여 변환한 것으로 주성분(principal components)으로 지칭한다. 주성분은 서로 독립이기 때문에 주성분을 통한 회귀모형인 식 (10)도 다중공선성에서 자유로우며 그에 따라서 정보의 손실이 없이 회귀분석을 수행할 수 있게 된다.

III. 적용 및 결과

1. 기후인자와 말라리아와의 상관관계 분석

수집한 2001년부터 2008년 까지의 월별 말라리아 발생환자수와 동 시간의 기후인자와의 회귀분석을 수행하였다. 각각의 기후인자와 말라리아 발생수와의 회귀분석은 비선형 회귀분석을 적용하였다. 회귀분석을 적용할 때에는 기존에 말라리아와 기후인자에 대하여 선행되어진 많은 연구(신호성 외, 2009; 김시현 외, 2010; 신호성, 2011; 김동진 외, 2012)에서 시간지연(k-lag effect)를 고려하고 있으므로 본 연구에서도 시간 지연에 따른 상관관계가 있을 것으로 가정하여 각 자료의 자기스펙트럼과 교차간섭성 분석을 수행하였다. 그림 2(a)는 말라리아 시계열에 대한 자기 스펙트럼 분석결과이며, 말라리아와 각 기후인자간에 교차간섭성 분석 수행 결과를 [그림 2](b) 부터 [그림 2](d) 까지 나타내었다. [그림 2](a)의 말라리아 자기스펙트럼 결과를 기준으로 판단하면, 말라리아 발생 시계열 자료는 12개월의 강한 주기($f=0.083$)와 6개월의 약한 주기($f=0.167$)를 가지는 시계열로 판단할 수 있으며, 그림 1(a)에서도 직관적으로 1년 단위의 주기를 가지는 것을 확인할 수 있다.

그림 2. 말라리아와 각 기후 인자 시계열 간의 Squared coherency(2001-2011)



또한, [그림 2](b) 에서부터 [그림 2](d)까지의 교차 간섭성을 통하여 각각의 기후인자와 말라리아가 상호간에 가지는 주기를 규명할 수 있다. 말라리아와 평균 기온(그림 2(b))의 경우에는 12개월($f=0.083$)과 6개월($f=0.167$)에서 높은 간섭성을 보이고 있으며, 말라리아와 습도(그림 2(c))의 경우에는 기온과 마찬가지로 12개월($f=0.083$)과 6개월($f=0.167$)에서 높은 간섭성을 보인다. 마지막으로 말라리아와 강수량의 경우에는 다른 기후인자와 다르게 12개월($f=0.083$), 6개월($f=0.167$), 4개월($f=0.25$) 주기로 간섭성이 나타나고 있다. 간섭성이 각 시계열 사이에 나타나는 상관평도를 설명한다는 점을 고려하면 각각의 간섭성은 말라리아에 대한 시간지연효과(time lag effect)로 판단된다. 따라서, 본 연구에서는 상관분석을 수행할 때 해당 주기에 대응하는 시간지연효과를 고려하였다.

또한, 기본적으로 자료의 통계적인 분석 시에 수집한 자료가 제한되어 있기 때문에 통계적인 분석을 위해서는 확률적인 변수(random variable)이라는 가정이 필요하며, 회

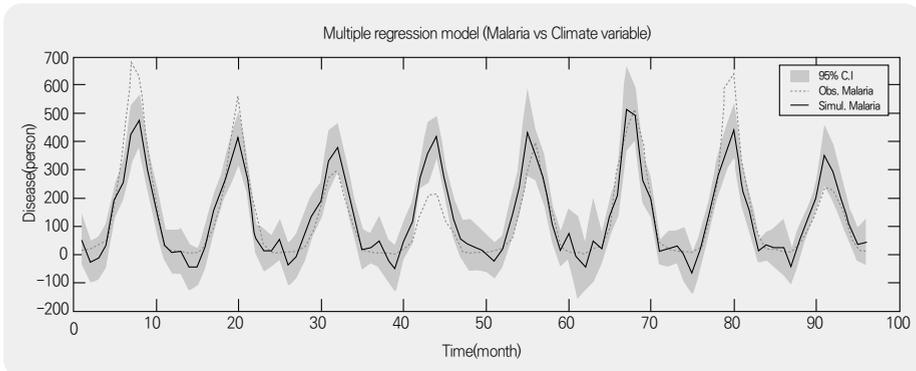
귀 분석을 수행할 때는 각각의 변수가 독립이라는 가정도 필요하다. 따라서, 수집된 말라리아 자료에 대하여 무작위성 및 비선형성에 대한 검정을 수행하였다. 자료의 독립성 또는 무작위성을 판단하기 위한 여러 통계적인 기법이 있는, 본 연구에서는 일반적으로 널리 이용되는 Anderson Correlogram, Run Test, Spearman Rank. Correlation Coefficient, Turning point test의 4가지 비모수 통계기법(Salas et al, 1993)과 BDS 통계검정을 수행하였으며, 각각의 검정통계량(Test statistic)과 95% 신뢰도 구간(95% C.I.), 검정 통계량에 따른 검정치를 표 1에 나타내었다.

표 1. 말라리아 시계열자료의 검정통계량(2001~2011년)

Test method	Test statistic	95% C. I.	Randomness Check
Run Test	-7.3261	[-1.96, +1.96]	X
Anderson's	0.1741	[-1.65, +1.65]	O
Spearman's	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

<표 1>의 검정통계량을 보면 Anderson Correlogram 과 Spearman 검정의 경우 해당 자료를 무작위 자료로서 판단하고 있으나, Run test 및 turning point 의 경우에는 무작위 자료가 아닌 것으로 판정하고 있다. BDS(m)통계치의 경우에 모든 차원에서 유의수준을 벗어나고 있으며, 이는 말라리아 자료가 비선형적인 특성을 가지고 있음을 나타내고 있다. 따라서, 말라리아에 대한 회귀모형 구축 시에는 비선형성을 고려한 회귀분석을 수행하여야 할 것으로 판단된다. 비선형 다중회귀분석을 바탕으로 하여 기후인자에 따른 말라리아 발생환자수의 다중회귀분석을 수행한 결과를 그림 3 및 식 (11) 에 나타내었다.

그림 3. 다중회귀분석을 통한 말라리아 시계열 자료의 모의(2001~2008년; 말라리아 관측치; Obs. Malaria, 말라리아 모의치; Simul. Malaria)



$$\begin{aligned}
 y &= 1164 - 11.76x_1 + 9.25x_2 - 4.24x_3 + 7.05x_4 - 16.75x_5 \\
 &\quad - 25.90x_6 - 0.34x_7 - 0.0076x_8 + 0.011x_9 - 0.14x_{10} + 1.03x_1^2 \\
 &\quad - 0.39x_2^2 + 0.063x_3^2 - 0.038x_4^2 + 0.11x_5^2 + 0.21x_6^2 + 0.0005x_7^2 \\
 &\quad - 1.6 \times 10^{-5}x_8^2 + 1.2 \times 10^{-5}x_9^2 + 0.00019x_{10}^2
 \end{aligned} \tag{11}$$

- y ; 말라리아 발병수(명/월)
- x_1 ; 월평균기온(°C)
- x_2 ; 월평균기온(°C) - 6개월 시차
- x_3 ; 월평균기온(°C) - 12개월 시차
- x_4 ; 월평균 상대습도(%)
- x_5 ; 월평균 상대습도(%) - 6개월 시차
- x_6 ; 월평균 상대습도(%) - 12개월 시차
- x_7 ; 월강수량(mm)
- x_8 ; 월강수량(mm) - 6개월 시차
- x_9 ; 월강수량(mm) - 12개월 시차
- x_{10} ; 월강수량(mm) - 4개월 시차

다중회귀분석 모의결과(그림 3)를 통한 다중회귀모형의 결정계수(R^2)는 0.805, 수정 결정계수(R^2)은 0.753 이며, 회귀분석 모형의 $F = 15.5$, 유의확률은 4.57×10^{-19}

이다. 따라서, 기후인자를 통한 말라리아 발생 모의는 통계적인 유의성을 가지는 것으로 판단된다. 그러나, 수집된 자료를 도시한 [그림 1](a)를 기준으로 판단하여 볼 때, 모든 자료가 동일한 주기를 가지고 있으며, 자간에 상관성이 있을 가능성이 높다. 이 경우 회귀분석으로 도출된 회귀모형도 다중공선성에 영향을 받을 수 있기 때문에, 말라리아 발생자 수와 각 기후인자간의 다중공선성에 대한 검정을 Pearson, Kendall 및 Spearman 상관계수를 통하여 수행하였다(표 2 참조). 상관계수에 의거한 분석 결과 평균 기온-습도의 0.52에서부터 평균 기온-말라리아의 0.92 까지 모든 기후인자들이 서로 높은 상관관계를 나타내는 것으로 도출되었다. 이러한 경우 각 기후인자들이 강한 상관성을 지니고 있는 것으로 판단되며 이 경우 다중회귀분석을 통하여 말라리아 회귀모형을 구축하기에는 다중공선성에 의한 오류가 있을 것으로 판단된다. 이 경우 변수를 제거하거나 현장 상황을 고려하여 해결하여야 하나(Golub & Loan, 1996; Strang, 1988) 정보의 감소 측면에서 받아들이기 어려우며, 본 연구에서는 이를 해결하기 위하여 주성분분석을 적용하였다.

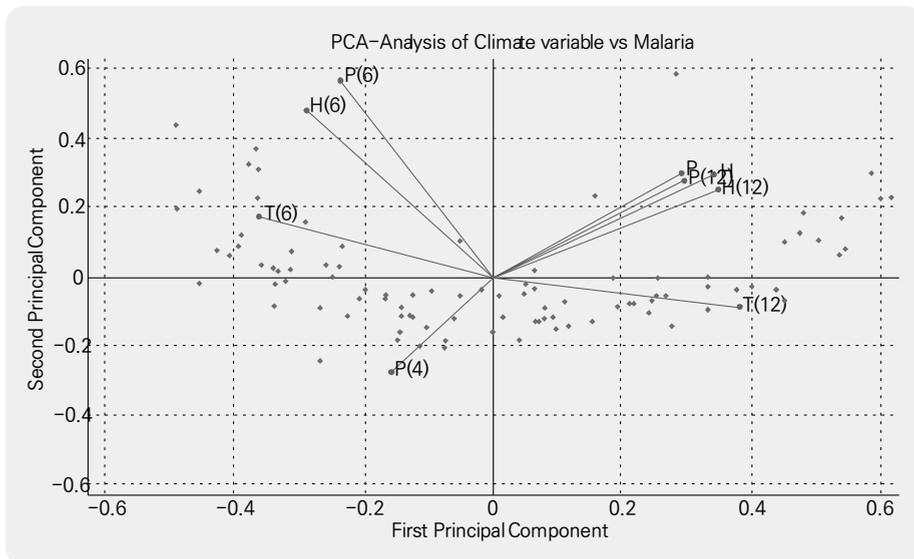
표 2. 말라리아 및 기후 인자간의 상관 Matrix

	Mean Temp.			Relative Humidity			Total Precip.			Malaria		
	Pearson	Kendall τ	Spearman ρ	Pearson	Kendall τ	Spearman ρ	Pearson	Kendall τ	Spearman ρ	Pearson	Kendall τ	Spearman ρ
Mean Temp.	1.00			0.72	0.52	0.72	0.58	0.60	0.80	0.78	0.74	0.92
Relative Humidity	0.72	0.52	0.72	1.00			0.76	0.58	0.76	0.75	0.54	0.73
Total Precip.	0.58	0.60	0.80	0.76	0.58	0.76	1.00			0.60	0.52	0.72
Malaria	0.78	0.74	0.92	0.75	0.54	0.73	0.60	0.52	0.72	1.00		

2. 주성분회귀분석을 통한 말라리아 회귀모형

말라리아와 기후 인자와의 주성분 회귀분석을 위하여 평균기온, 평균습도, 강수량에 대한 주성분 분석을 수행하였다. 본 연구에서는 누적 기여율 80%를 기준으로 기후인자를 3개의 주성분으로 요약하였다. 주성분 1, 2, 3까지의 주성분 점수는 각각 -2.9698, 3.0490, -0.6665 으로 산정되었으며, 해당 주성분 분석의 도표는 [그림 5]와 같다.

그림 4. 기후 인자의 PCA-Diagram



(P=Precip.; T=Temp.; H=Humidity; example P(4)= 4 month lag Precip.)

주성분은 평균기온(x_1), 평균습도(x_5, x_6)의 제1 주성분과 평균습도(x_6)와 강수량(x_7)의 제2 주성분, 평균기온(x_1)과 평균습도(x_5, x_6), 강수량(x_7)의 제3 주성분으로 요약된다. 주성분 분석을 바탕으로 주성분 회귀분석을 수행할 수 있으며, 각각의 주성분은 서로 독립이므로 다중공선성 문제에서 벗어나서 회귀분석을 수행할 수 있다. 말라리아 발생자 수의 주성분 회귀모형은 식 (12)와 같다.

$$y = 93.27 + 52.51z_1 + 4.77z_2 - 19.62z_3 + 6.58z_1^2 - 0.54z_2^2 + 5.28z_3^2 \quad (12)$$

z_1 ; 1st principal component,

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

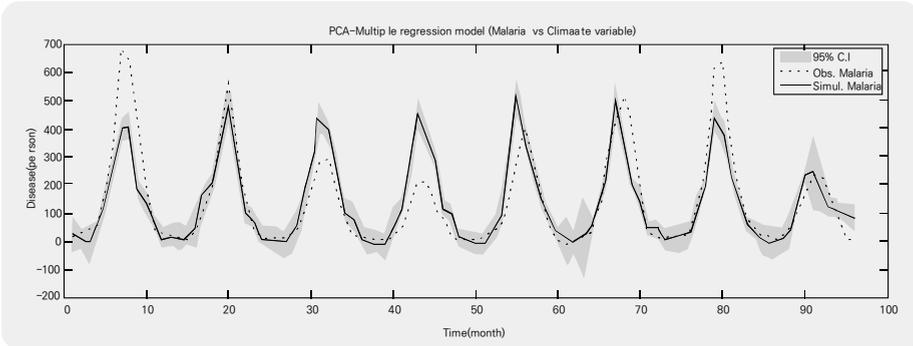
z_2 ; 2nd principal component

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

z_3 ; 3rd principal component

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

그림 5. 주성분 회귀분석을 통한 말라리아 시계열 자료의 모의(2001~2008년)

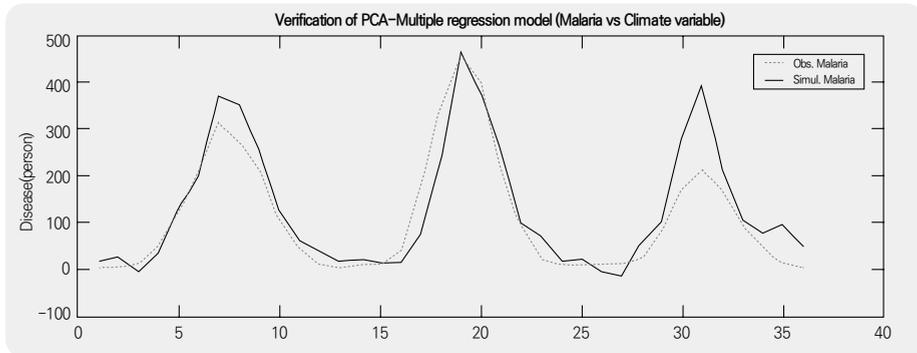


주성분 회귀모형에 따른 말라리아 발생 모의 결과는 [그림 6]와 같다. 기후인자를 이용한 말라리아의 주성분 회귀모형의 결정계수(R^2)는 0.743, 수정 결정계수(R^2) 0.725, 주성분 회귀분석 모형의 $F = 42.8$, 유의확률은 3.4×10^{-24} 로서 주성분 회귀분석을 통한 말라리아 모의는 통계적인 유의성을 가지는 것으로 판단된다. 또한, 다중공선성에 영향을 받지 않게 됨에 따라서 불확실성도 감소하는 경향을 보이며, 이는 [그림 3]의 결과와 [그림 6]에 표시된 95% 신뢰구간에서도 확인할 수 있다. 따라서, 상대적으로 주성분 회귀분석을 통한 말라리아 분석이 더 설득력을 가지는 것으로 판단된다.

3. 말라리아 회귀모형 검증

주성분 회귀분석을 통하여 산정한 말라리아 회귀모형의 적합성을 검증하기 위하여 2009년부터 2011년까지의 말라리아 발생자료와 기상대의 기상자료를 이용하여 회귀모형의 검증을 수행하였으며, 그 결과를 [그림 7]에 나타내었다.

그림 6. 주성분 회귀분석 모형의 검증(2009~2011년)



주성분 회귀분석을 통하여 검증한 결과, 주성분 회귀모형의 결정계수 R^2 은 0.852, 모형의 효율성을 나타내는 $NRMSE$ (Normalized Root Mean Square Error; Scott and Fred, 1992)지수가 0.117, RE (Relative Error; Golub and Loan, 1996) 지수가 0.026 으로서 모형을 통하여 모의한 말라리아 발생 시계열이 실제 관측자료와 유의성이 있는 것으로 판단된다. 따라서, 기후인자를 통하여 말라리아의 발생 추이를 예측하기 위하여 본 연구에서 제시한 모형을 적용할 수 있을 것으로 판단된다. 다만, 모형 구축 및 검증 시계열에서도 일부 최고발생점에서 오차가 발생하는 경우가 관찰되었다(2001, 2007년 8월 및 2011년 8월). 이는 말라리아가 단순 기후인자에만 영향을 받는 것이 아니라 생태계 내에서 상호작용 및 병리학적 영향도 받으므로 해당 영향에 대한 불확실성이 원인이 판단된다.

구축된 회귀모형은 일반적으로 얻을 수 있는 평균 기온과 습도, 강수량 자료를 통하여 말라리아의 대략적인 발생추이를 분석할 수를 예측할 수 있으므로 향후 공중보건 정책 및 예방 대응 측면에서 기초 자료로서 활용될 수 있을 것으로 판단된다. 다만, 본 연구에서 제시한 모형은 말라리아의 생태계 및 임상병리학적 특성을 제외하고 기후인자만을 이용하여 구축하였으며, 사용한 자료도 2001년부터 2011년 사이의 서울·경기지역의 자료만을 사용하였으므로 전체적인 말라리아의 경향성을 예측하기에는 한계가 존재한다. 그러나, 이러한 단점에도 불구하고 기후인자를 통하여 말라리아의 추이를 분석할 수 있다는 점에서 활용가능성이 높을 것으로 판단되며 특히, 근래와 같이 기후변화로 인한 변화가 주요한 사회적인 이슈로 떠오르는 시점에서는 기후변화 시나리오 및 기후

모형을 통하여 도출한 미래의 기후인자와 결합하여 미래의 말라리아 발생 현황을 예측할 수 있으므로 미래의 보건사회적인 연구 및 예측에 활용할 수 있을 것으로 판단된다. 그러나, 이 경우에도 모형에 적용된 기후인자를 제외한 요인의 변동에 따라서 불확실성이 발생할 것으로 판단되므로 향후 이러한 한계점을 명확히 인지하여야 할 것으로 판단된다.

IV. 결론

본 연구에서는 기후 자료와 말라리아 감염병 발생수와의 상관관계를 규명하기 위하여 월 평균기온, 평균습도, 강수량의 기후 자료를 이용하여 상관관계를 분석하고, 각 기후인자들 간의 스펙트럼 분석을 통하여 말라리아 발생수와 기후인자간의 시간 지연 효과를 고려하였다. 또한, 각각의 기후 자료에 대한 상관성 분석 결과, 기후 자료 간에 강한 상관성이 존재하는 것으로 나타나서 회귀모형이 다중공선성에 영향을 받는 것으로 분석되었으며, 이러한 문제를 해결하기 위하여 주성분 분석에 기반한 주성분 회귀분석을 통하여 회귀 모형을 구축하였다. 구축된 회귀 모형을 이용하여 2009년부터 2011년 까지의 말라리아 감염병 발생수를 검증한 결과 결정계수 R^2 은 0.852, $NRMSE$ 이 0.117, RE 가 0.026으로 말라리아 감염병에 대한 설명력이 있는 것으로 나타났다. 향후, 본 연구의 결과와 기후변화 시나리오의 결과를 결합하여 미래의 감염병 발생을 대응을 위한 기초자료로서 활용할 수 있을 것으로 판단된다.

참고문헌

- 기상청(2013). 기상청 홈페이지. <http://www.kma.go.kr/>. 2013.5.15.
- 고원규(2007). 국내 말라리아의 재유행. *대한의사협회지*, 50(11), pp.959-966.
- 김동진, 신호성, 채수미(2012). 기후변화가 설사병 발생에 미치는 영향 분석, *보건사회연구*, 32(1), pp.281-297.
- 김시현, 장재연(2010). 국내 기후변화 관련 감염병과 기상요인간의 상관성. *J. Prev Med Public Health*, 43(5), pp.436-444.
- 박윤형, 김호, 장원기, 김용배, 황보영, 기모란 외(2006). 기후변화에 의한 전염병 발생영향 통합관리체계 구축. 서울: 건강증진기금사업지원단.
- 신호성, 윤시몬, 정기혜, 이수형(2009). 기후변화와 식중독 발생 예측, *보건사회연구*, 29(1), pp.143-162.
- 신호성(2011). 기상요소와 지역 말라리아 발생자수의 상관관계. *보건사회연구*, 31(1), pp.217-237.
- 장재연, 박윤형(2008). 기후변화에 따른 건강피해 모니터링 및 위험인구 감소전략 개발 연구. 아주대학교 · 건강증진사업지원단.
- 보건복지부(2013). 질병관리본부 홈페이지. <http://www.cdc.go.kr/>. 2013.5.15.
- Bhattacharya, S., Sharma, C., Dhiman, R. C., Mitra, A. P. (2006). Climate change and malaria in India. *CURRENT SCIENCE-BANGALORE*, 90(3), pp.369.
- Brock, W. A., Heish, D. A., Lebaron, B. (1991). *Nonlinear dynamics, chaos, and instability: Statistical theory and economic evidence*. The MIT Press, Massachusetts (USA).
- Brock, W. A., Dechert W. D., Scheinkman, J. A., Lebaron, B. (1996). A test for independence based on the correlation dimension. *Econometric Reviews*, 15(3), pp.197-235.
- Chatterjee, S., Hadi, A. S., Price, B. (2000). *Regression analysis by example (3rd ed.)*. John Wiley & Sons.
- Costello, A., Abbas, M., Allen, A., Ball, S., Bell, S., Bellamy, R., Patterson, C. (2009). Managing the health effects of climate change. *The Lancet*, 373(9676), pp.1693-1733.

- Craig, M. H., Snow, R. W., Le Sueur, D. (1999). A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology today*, 15(3), pp.105-110.
- C. T. Kelley (1999). *Iterative Methods for Optimization*, SIAM Frontiers in Applied Mathematics.
- Dennis Ward Ricker (2003). *Echo Signal Processing*. Springer.
- Draper, N., Smith, H. (1998). *Applied regression analysis (3rd ed.)*. John Wiley & Sons.
- Epstein, P. R., Diaz, H. F., Elias, S., Grabherr, G., Graham, N. E., Martens, W. J., Susskind, J. (1998). Biological and physical signs of climate change: focus on mosquito-borne diseases. *Bulletin of the American Meteorological Society*, 79(3), pp.409-417.
- Gage, K. L., Burkot, T. R., Eisen, R. J., Hayes, E. B. (2008). Climate and vectorborne diseases. *American journal of preventive medicine*, 35(5), pp.436-450.
- Gething, P. W., Smith, D. L., Patil, A. P., Tatem, A. J., Snow, R. W., Hay, S. I. (2010). Climate change and the global malaria recession. *Nature*, 465(7296), pp.342-345.
- Golub, G.H., Van Loan, C.F. (1996). *Matrix computation (3rd ed.)*. The Johns Hopkins University press.
- Greer, A., Ng, V., Fisman, D. (2008). Climate change and infectious diseases in North America: the road ahead. *Canadian Medical Association Journal*, 178(6), pp. 715-722.
- Gujarati, Damodar. (2002). *Multicollinearity: what happens if the regressors are correlated?. Basic Econometrics (4th ed.)*. McGraw-Hill. pp.363-363.
- Haines, A., Kovats, R. S., Campbell-Lendrum, D., & Corvalan, C. (2006). Climate change and human health: impacts, vulnerability and public health. *Public health*, 120(7), pp.585-596.
- Hay, S. I., Cox, J., Rogers, D. J., Randolph, S. E., Stern, D. I., Shanks, G. D., Snow, R. W. (2002). Climate change and the resurgence of malaria in the East African highlands. *Nature*, 415(6874), pp.905-909.
- Jang J. Y., Cho S. H., Kim, S. Y., Cho, S. N., Kim, M. S., Baek, K. W. (2003). *Assessment of Climate Change Impact and Preparation of Adaptation Program in Korea*. Seoul: Ministry of Environment; 2003. (Korean)

- Kanzler, L. (1999). *Very fast and correctly sized estimation of the BDS statistic*. Manuscript . Christ Church and Department of Economics, University of Oxford .
- Kleinschmidt, I., Bagayoko, M., Clarke, G. P. Y., Craig, M., Le Sueur, D. (2000). A spatial statistical approach to malaria mapping. *International Journal of Epidemiology*, 29(2), pp.355-361.
- Kim, H. S., Kang, D. S., Kim, J. H. (2003). The BDS statistic and residual test. *Stoch. Environ. Res. Risk Assess*, 17, 1(2), pp.104-115.
- Kuhn, K., Campbell-Lendrum, D., Haines, A., Cox, J., Corvalán, C., Anker, M., Malaria, R. B. (2005). *Using climate to predict infectious disease epidemics*. Geneva: WHO.
- Kutner, M., Nachtsheim, C., Neter, J., Li, W. (2005). *Applied linear statistical models (5th ed.)*. McGraw Hill.
- Lin, M., Wei, L. (2002). The small sample properties of the principal components estimator for regression coefficients. *Communication in Statistic-Theory and Methods*, 31, pp.271-283.
- Loevinsohn, M. E. (1994). Climatic warming and increased malaria incidence in Rwanda. *The Lancet*, 343(8899), pp.714-718.
- Martens, P., Kovats, R. S., Nijhof, S., De Vries, P., Livermore, M. T. J., Bradley, D. J., ... McMichael, A. J. (1999). Climate change and future populations at risk of malaria. *Global Environmental Change*, 9, pp.S89-S107.
- Martens, W. J., Niessen, L. W., Rotmans, J., Jetten, T. H., McMichael, A. J. (1995). Potential impact of global climate change on malaria risk. *Environmental health perspectives*, 103(5), pp.458.
- Minakawa, N., Sonye, G., Mogi, M., Githeko, A., Yan, G. (2002). The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. *Journal of medical entomology*, 39(6), pp.833-841.
- Mogull, Robert G. (2004). *Second-Semester Applied Statistics*. Kendall/Hunt Publishing Company.
- Montgomery, D. C., Peak, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis (3rd ed.)*. John Wiley & Sons.

- Morison, D. F. (2005). *Multivariate statistical methods(4th ED.)*. Thomson Learning.
- Olson, S. H., Gangnon, R., Elguero, E., Durieux, L., Guégan, J. F., Foley, J. A., Patz, J. A. (2009). Links between climate, malaria, and wetlands in the Amazon Basin. *Emerging infectious diseases*, 15(4), pp.659.
- Paaijmans, K. P., Read, A. F., Thomas, M. B. (2009). Understanding the link between malaria risk and climate. *Proceedings of the National Academy of Sciences*, 106(33), pp.13844-13849.
- Paaijmans, K. P., Blanford, S., Bell, A. S., Blanford, J. I., Read, A. F., Thomas, M. B. (2010). Influence of climate on malaria transmission depends on daily temperature variation. *Proceedings of the National Academy of Sciences*, 107(34), pp.15135-15139.
- Parham, P. E., Michael, E. (2010). Modeling the effects of weather and climate change on malaria transmission. *Environmental health perspectives*, 118(5), pp.620.
- Patz, J. A., Campbell-Lendrum, D., Holloway, T., Foley, J. A. (2005). Impact of regional climate change on human health. *Nature*, 438(7066), pp.310-317.
- Patz, J. A., Olson, S. H. (2006). Malaria risk and temperature: Influences from global climate change and local land use practices. *Proceedings of the National Academy of Sciences*, 103(15), pp.5635-5636.
- Poveda, G., Rojas, W., Quiñones, M. L., Vélez, I. D., Mantilla, R. I., Ruiz, D., Rua, G. L. (2001). Coupling between annual and ENSO timescales in the malaria-climate association in Colombia. *Environmental health perspectives*, 109(5), pp.489.
- Salas, J. D., Smith, R. A., Tabios, G. Q. III and Heo, J. H. (1993). *Statistical Computer techniques in hydrology and water resources*. Unpublished text book un CE 622, Colorado State University.
- Scott, J. A., Collopy, F. (1992). Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, 8, pp.69-80.
- Small, J., Goetz, S. J., Hay, S. I. (2003). Climatic suitability for malaria transmission

in Africa, 1911 - 1995. *Proceedings of the National Academy of Sciences*, 100(26), pp.15341-15345.

Strang, G. (1998). *Linear algebra and its applications (3rd ed)*. Thomsom Learning.

Tanser, F. C., Sharp, B., Le Sueur, D. (2003). Potential effect of climate change on malaria transmission in Africa. *The Lancet*, 362(9398), pp.1792-1798.

Thomson, M. C., Doblas-Reyes, F. J., Mason, S. J., Hagedorn, R., Connor, S. J., Phindela, T., Palmer, T. N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, 439(7076), pp.576-579.

Van Lieshout, M., Kovats, R. S., Livermore, M. T. J., Martens, P. (2004). Climate change and malaria: analysis of the SRES climate and socio-economic scenarios. *Global Environmental Change*, 14(1), pp.87-99.

Zhou, G., Minakawa, N., Githeko, A. K., & Yan, G. (2004). Association between climate variability and malaria epidemics in the East African highlands. *Proceedings of the National Academy of Sciences of the United States of America*, 101(8), pp.2375-2380.

곽재원은 인하대학교에서 토목공학 박사학위를 받았으며, 현재 Quebec대학교 INRS연구소에서 박사후과정중이다. 주요 관심분야는 기후변화, 가뭄 및 사회/생태적 변화이며, 현재 기후변화 및 그 영향 등을 연구하고 있다.
(E-mail: firstsword@naver.com)

이종소는 인하대학교에서 토목공학 석사학위를 받았으며, 현재 인하대학교에서 박사과정을 수행 중이다. 주요 관심분야는 기후변화 및 수질이며, 현재 수질 영향 등을 연구하고 있다.
(E-mail: stynrehero@naver.com)

한희찬은 인하대학교에서 토목공학 학사학위를 받았으며, 현재 인하대학교에서 석사과정을 수행 중이다. 주요 관심분야는 기후자료 분석이며, 현재 기존의 기후자료 분석을 연구하고 있다.
(E-mail: hhch515@naver.com)

김형수는 미국 Colorado States Univ.에서 토목공학 박사학위를 받았으며, 현재 인하대학교에서 교수로 재직 중이다. 주요 관심분야는 기후변화이며, 현재 기후변화 및 그 영향 등을 연구하고 있다.
(E-mail: sookim@inha.ac.kr)

A Case Study: Malaria Modeling Based on Climate Variables in Korea

Kwak, Jaewon

(Quebec University)

Han, Heechan

(Inha University)

Lee, Jongso

(Inha University)

Kim, Hungsoo

(Inha University)

Malaria is influenced by climate variable and is the third grade infectious disease of Korea Government which has been slowly increasing patient in every year since its reoccurrence in the Military Demarcation Line in Korea in 1993. This study is focused on the influence analysis with climate variable and malaria occurrence and construct the model for malaria prediction based on climate factor. To do so, data of monthly malaria infection, mean temperature, relative humidity and total precipitation will be collected. To observe time lag effect, auto and cross spectral density analysis will be applied. To consider multicollinearity of each climate variable, model of malaria infection will be structured based on principal components regression method. Result shows that model explains and verifies malaria's infectivity using infection data from 2009 to 2011.

Keywords: Malaria, Climate Variable, Spectral Analysis, Principal Components Regression