

# Weather Generator를 이용한 APCC 계절예 측의 통계적 상세화 연구 -겨울철 낙동강 유역 중심으로-

Statistical Downscaling of APCC Seasonal Prediction using Weather Generator  
with Application to Nakdong River Basin in Boreal Winter

김무섭



# Weather Generator를 이용한 APCC 계절예 측의 통계적 상세화 연구 -겨울철 낙동강 유역 중심으로-

Statistical Downscaling of APCC Seasonal Prediction using Weather Generator  
with Application to Nakdong River Basin in Boreal Winter

김무섭



## 발간사

APEC 기후센터의 대표적인 기후서비스인 기후 계절예측 정보는 매월 전지구 규모로 생산되고 있습니다. 특히, APCC 계절예측은 세계 유수의 기후예측 기관으로부터 예측정보를 수집하여 다중모형앙상블기법(multi-model ensemble, MME)을 적용하여 기후예측정보를 생산하여 세계 최고 수준의 정확도와 신뢰도를 보여 주고 있습니다.

그러나 APCC 계절예측정보는 전지구 규모로 생산되기 때문에 실제 기후관련 응용분야에서 사용하는 데에는 시간공적 해상도가 걸림돌이 되고 있습니다. 기후관련 응용분야에서는 특정 대상지역에 특화된 기후자료를 필요로 하지만 계절예측 자체는 그러한 요구를 충족시키지 못하기 때문에 계절예측 정보를 실제 기후 응용분야에 사용하기 위해서는 예측정보의 상세화(downscaling)가 필수적인 과정입니다.

상세화기법은 역학적 방식과 통계적 방식으로 구분되는데 본 연구는 후자의 접근법을 사용합니다. 역학적 방식에 비해 통계적 방식은 통계적 편의(bias)가 작은 다량의 자료를 신속하게 생산할 수 있어 농업과 수문분야에서 널리 사용되어 왔습니다.

본 연구는 월 단위 계절예측 정보를 특정구역의 일 단위 정보로 상세화하는 것을 목표로 하고 있습니다. 이를 위해 우리는 weather generator를 사용하는데 weather generator는 농업이나 수문에서 기상자료를 생산하기 위해 이용되는 기상통계모형으로 주로 관측자료기간의 평균기후에 맞추어 추정을 합니다.

그러나 계절예측 상세화를 위해서는 통계모형이 경년변동성을 표현할 필요가 있으므로 본 연구에서는 weather generator에 경년변동성을 나타내는 모수를 도입하고 이를 계절예측과 연결하였습니다. 즉, 기존의 weather generator 연구에서 거의 고려되지 않았던 경년변동성 요소를 도입하고 이를 계절예측에 적용했다는 측면에서 통계적 상세화 연구로서 큰 의미를 지니고 있습니다.

끝으로 연구결과가 나오기 까지 본 연구를 수행해 주신 김무섭 박사의 노고에 감사드립니다. 또한 본 연구는 2015년 weather generator 연구를 모태로 하는데 2015년 연구에 조언과 도움을 주신 APCC의 여러 박사들과 자문위원단 분들께 감사드립니다. 본 연구를 기반으로 계절예측 상세화 연구가 앞으로도 더욱 발전해 나갈 수 있기를 바랍니다.

2017년 1월  
APEC 기후센터  
원장 정 홍 상

---

## ABSTRACT

The APEC Climate Center provides global-scale seasonal climate prediction every month. However, it is problematic to apply the prediction result directly to climate application studies due to its coarse resolution in space and time. In most application studies, climate data is required to capture local basin climate at least in daily time scales. Thus, it is necessary to reproduce detailed climate data from the seasonal prediction. This procedure is called downscaling.

There are two kinds of downscaling methods: dynamical and statistical downscaling. Dynamical downscaling is based on a regional climate model that simulates atmospheric circulation around a target basin better than general circulation models in virtue of the attention to detailed topography. However, it entails high computational cost, and the resulting data has a bias which must be appropriately corrected for usage in application studies. On the other hand, statistical downscaling can rapidly produce data with low computational burden, and the resulting data is less biased. However, it is based on an empirical relation that is not easy to understand dynamically, and requires reliable observation data collected for a long period.

In this study, we aim to develop APCC seasonal prediction downscaling scheme using a weather generator. A weather generator is a simulator that is based on a statistical model describing local basin climate and has been popularly used in the fields of hydrology and agriculture. In this study, it is employed as a temporal downscaling tool that produces daily weather scenarios from seasonal prediction. However, in general, weather generators are calibrated to climatology and thus

do not express inter-annual variability. Therefore, a new weather generator should be developed.

There are two kinds of approaches to construct the statistical models of weather generators: parametric and non-parametric approaches. The parametric approach assumes that weather variables follow a parametric distribution and is adopted by site-specific weather generators. On the other hand, since multisite weather generators have to describe the weather of several sites simultaneously, its statistical model can not avoid being complicated. Thus, the non-parametric approach has been popularly employed due to its flexibility. In this study, we consider a multisite weather generator for covering a local basin rather than a single site, but adopt parametric approach in order to effectively express inter-annual variability, i.e., inter-annually varying parameters are introduced into the statistical model. Moreover, they are linked with seasonal prediction through a prediction model. Thus, the values of the inter-annual varying parameters are determined by a given seasonal prediction and we then produce weather scenarios by running the calibrated weather generator.

The proposed downscaling scheme is applied to the Nakdong river basin during the boreal winter season. First, we estimate the inter-annual variation of the basin climate using observation data. For precipitation, precipitation intensity and mean of dry-spell length are chosen as inter-annually varying parameters. They are estimated by using a weighted average technique for overcoming lack of observation data. For temperature, the intercept terms in temperature regression model are taken, which are interpreted as monthly mean of maximum and minimum daily temperatures adjusted by precipitation effect. They are estimated by the least-square method as a usual regression analysis.

For linking the basin and global climates, we consider a regression equation as a prediction model. Especially, in order to mitigate overfitting, we employ

regularized canonical correlation analysis and variable selection technique as the fitting method. Geopotential height at 500 hPa (Z500) and sea-level pressure (SLP) Z500 and SLP are utilized as predictors that come from the APCC Seasonal Prediction issued in November, a month before the beginning of the season. Z500 appears to have predictability for monthly mean of maximum daily temperature adjusted precipitation effect in December and February, and SLP does for the parameters related to precipitation in December.



## 목 차

|                                       |    |
|---------------------------------------|----|
| 1. 서론                                 | 1  |
| 2. 연구 자료 및 방법                         | 4  |
| 2.1 연구 자료                             | 4  |
| 2.2 연구 방법                             | 5  |
| 2.2.1 유역기후의 경년변동성 표현을 위한 통계모형         | 6  |
| 2.2.1.1 유역강수모형                        | 6  |
| 2.2.1.2 유역기온모형                        | 9  |
| 2.2.2 경년변동성 예측모형                      | 11 |
| 2.2.2.1 predictand의 공변동성              | 11 |
| 2.2.2.2 예측인자의 탐색과 관측장 분석              | 11 |
| 2.2.2.3 GCM 예측장 분석                    | 12 |
| 2.2.2.4 회귀모형                          | 13 |
| 2.2.2.5 오차항 모형                        | 14 |
| 2.2.2.6 Weather Generator의 모수 조정 및 구동 | 14 |
| 3. 연구내용                               | 16 |

|                                   |    |
|-----------------------------------|----|
| 3.1 낙동강 유역에서 경년변동성 통계모형 적합결과----- | 16 |
| 3.1.1 유역강수모형 적합결과-----            | 16 |
| 3.1.2 유역기온모형 적합결과-----            | 19 |
| 3.2 낙동강 유역의 경년변동에 대한 예측모형-----    | 21 |
| 3.2.1 탐색적 사전분석-----               | 21 |
| 3.2.2 예측모형-----                   | 24 |
| 4. 결론 및 토론-----                   | 33 |
| Ⅱ REFERENCE-----                  | 35 |
| Ⅱ 부록-----                         | 36 |





## 1. 서론

APEC 기후센터는 매월 전지구규모 계절예측을 생산하여 공공에 제공하고 있다. 하지만 계절예측자료는  $2.5^{\circ} \times 2.5^{\circ}$  격자간격에서 월단위로 생산되어 시공간적 해상도가 낮고, 따라서 농업, 수자원, 방재 등과 같은 기후응용분야에 활용되는데 난관이 있다. 기후응용연구의 대상지역규모는 대체로 유역규모정도이고 시간규모는 하루 내지 시간 단위로 계절예측보다 시공간적으로 훨씬 더 세밀하다. 그래서 계절예측자료를 기반으로 고해상도의 자료를 재생산하는 과정이 필요한데, 이를 상세화(downscaling)라고 부른다.

상세화는 크게 지역기후모형(regional climate model)을 기반으로 하는 역학적 상세화(dynamical downscaling)와 통계모형(statistical model) 기반의 통계적 상세화(statistical downscaling)로 구분된다. 역학적 상세화는 지역기후모형에서 세밀하게 묘사된 지형 덕분에 대상유역 주변의 역학적 순환을 GCM(global climate model)보다 더 정확하게 모의할 수 있다. 하지만, 계산비용이 크며 결과자료에 편의(bias)가 있어 실제사용에서는 적절한 편의수정과정(bias-correction)이 따라와야 한다. 한편, 통계적 상세화는 자료생산에 큰 계산적 비용이 들지 않고, 대체로 자료의 편의가 덜하다. 하지만 정상성(stationarity)가정 하에서 추정된 경험적 관계(empirical relation)를 기반하기 때문에 역학적 이해가 어려운 경우가 많고, 또한 모형구축을 위해서 장기간 축적된 정확도 높은 관측자료가 필요하다. 이와 같이 두 방법은 장단점이 있으며 상호 보완이 가능한 측면이 있다. 상세화의 실제적용에 있어서는 상황에 따라 적절한 방법을 선택해야 할 것이다.

본 연구에서는 weather generator를 이용한 APCC 계절예측의 통계적 상세화 기법을 개발하고자 한다. Weather generator는 유역기후의 통계모형을 기반으로 일별 기상자료를 생성하는 모의체계(simulation system)로 농업과 수문 분야에서 널리 사용되고 있다. 본 연구에서는 weather generator를 APCC 계절예측의 시간적 상세화(temporal downscaling) 도구로 사용하고자 한다. 즉, 월단위로 주어진 계절예측에 상응하는 일별기상자료를 weather generator를 이용해 생산하는 것이다. 그런데, weather generator의 통계모형은 관측기간의 평균적인 기후에 맞추어 추정되기 때문에 본질적으로 경년변동성(inter-annual variability)을 표현하지 않는다. 그래서

APCC 계절예측의 시간적 상세화를 위해서는 통계모형이 경년변동성을 나타낼 수 있게 수정되어야 한다.

Weather generator의 통계모형구축법은 일반적으로 모수적(parametric) 접근방식과 비모수적(non-parametric) 그것으로 구분된다. 모수적 방식은 기후변수에 대해 모수적 분포(parametric distribution)를 가정하는 것으로 단일지점(site-specific) weather generator에서 주요하게 채택되었다, (cf. Richardson (1981), Racsco et al. (1991)). 한편, 유역전체의 기후를 모의하기 위한 다지점(multisite) weather generator는 기반통계모형의 복잡성 때문에 기존 연구에서는 주로 적용이 유연한 비모수적 방식이 쓰였다, (cf. Rajagopalan and Lall (1999), Apipattanavis et al. (2007), Eum et al. (2010)). 본 연구에서는 유역기후를 모의하기 위해 다지점 weather generator를 고려하되 기반통계모형구축에 있어서는 경년변동성을 효과적으로 표현하기 위해 모수적 방식을 쓴다. 즉, weather generator 통계모형에 경년변동성을 표현하는 모수를 도입하고, 나아가 그것의 경년변동을 설명하는 APCC 계절예측기반의 예측모형을 만듦으로써 APCC 계절예측 상세화 기법을 완성하고자 한다. 이러한 다지점 weather generator와 APCC 계절예측의 연결이 통계적 상세화의 체계가 된다. 그림 1은 상세화 체계를 도시하였다.

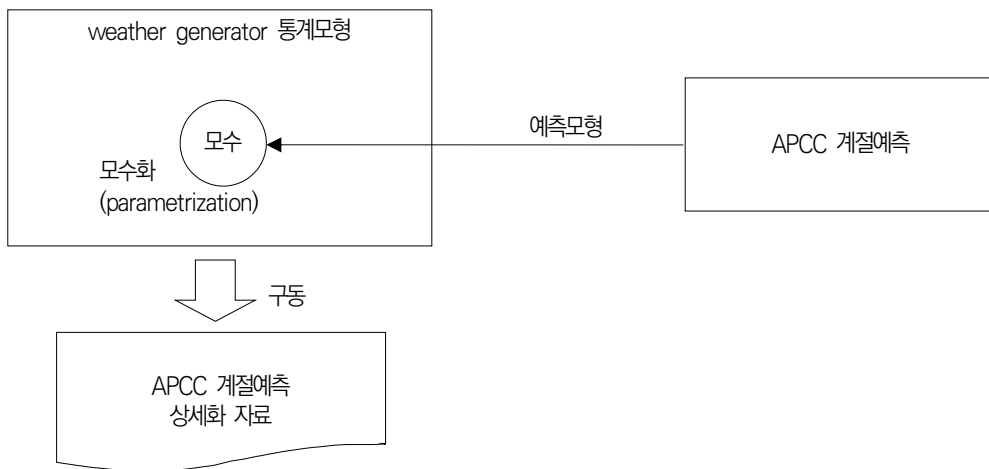


그림 1 APCC 계절예측 상세화 체계. weather generator의 통계모형을 경년변동성을 표현할 수 있도록 모수화(parametrization)시키고, 경년변동의 모수를 예측모형으로 APCC 계절예측과 연결한다. 계절예측으로 모수가 조정된 weather generator를 구동하여 상세화 자료를 얻는다.

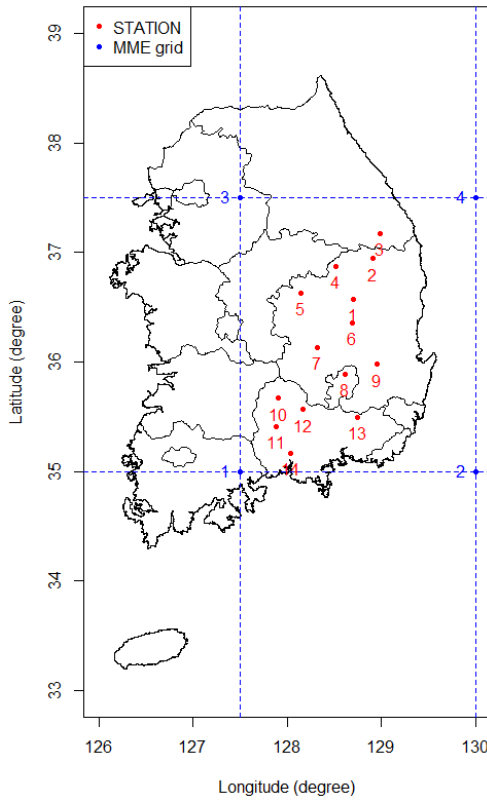
이후 본 연구보고서의 구성은 다음과 같다: 2절에서는 연구에서 쓰일 자료를 설명하고, 경년변동성을 나타내는 유역기후 통계모형을 도입한다. 그리고 경년변동을 설명하는 APCC 계절예측기반의 예측모형을 고려한다. 3절에서는 2절에서 소개된 방법을 실제 겨울철 낙동강 유역에 적용한 내용과 결과를 제시한다. 마지막으로 4절에서는 본 연구의 결론을 짓고 향후 연구에 위한 논의점을 언급한다. 부록은 본문에서 소개된 기법의 상세한 내용을 담는다.

## 2. 연구 자료 및 방법

### 2.1 연구 자료

본 연구를 수행하기 위해서는 3 종류의 기후자료가 필요하다. 첫 번째로 대상유역의 기상관측자료이다. 본 연구의 대상지역은 낙동강 유역이다. 현재, 가용한 기상자료는 크게 ASOS자료와 AWS자료가 있는데, 본 연구에서는 관측기간이 긴 ASOS자료를 쓴다. 낙동강 유역에 위치에 있는 14개 지점의 관측자료를 쓰며 기간은 1988년 1월 1일부터 2013년 12월 31일까지로 총 26년이다. 그림 2은 관측소의 위치를 나타내고 있다. 이 자료는 weather generator의 통계모형을 구축하고 유역기후의 경년변동을 추정하는데 주요하게 쓰인다.

두 번째 자료는 대규모 기후의 관측장(observation field) 자료이다. 본 연구에서는 관측장자료로 NCEP2 재분석 자료를 쓴다. 기간은 1979년 1월부터 2013년 12월까지이고 월평균자료이다. 이 자료는 예측인자(predictor) 탐색과 GCM 예측장 분석에서 주요한 역할을 한다. 마지막으로 예측모형의 예측인자를 제공할 예측장(prediction field)자료가 필요하다. 연구목적에 따라 APCC 계절예측자료를 쓴다. 예측자료는 강수량, 2m/850hPa 기온, Z500, SLP, SST, UV850을  $2.5^{\circ} \times 2.5^{\circ}$  격자간격으로 월별로 제공한다. APCC 계절예측자료는 총 6개의 GCM으로부터 얻어진 결과이다. 예측자료의 기간은 1983년 12월부터 2014년 2월까지로, 우리나라 겨울철 3개월 각각이 시작 한 달 전인 11월에 예측된 결과이다. 그래서 각 월에 대해서 예측선행시간이 다른데, 12월의 선행시간은 1달, 1월은 2달, 2월은 3달이다. 6개의 모형 중 5개는 앙상블크기(ensemble size)가 10개, 1개는 5개이며, 연구에서는 앙상블평균장(ensemble average)을 쓴다.



| No. | 이름 | No. | 이름 |
|-----|----|-----|----|
| 1   | 안동 | 8   | 대구 |
| 2   | 봉화 | 9   | 영천 |
| 3   | 태백 | 10  | 거창 |
| 4   | 영주 | 11  | 산천 |
| 5   | 문경 | 12  | 합천 |
| 6   | 의성 | 13  | 밀양 |
| 7   | 구미 | 14  | 진주 |

그림 2 낙동강 유역, ASOS 관측소(적색). 청색점은 APCC 계절예측 격자점을 나타낸다.

## 2.2 연구 방법

본 연구의 기술적 목표는 weather generator의 통계모형이 유역기후의 경년변동성(inter-annual variability)을 나타낼 수 있도록 하고, 경년변동의 원인을 APCC 계절예측에서 구하는 것이다. 이 목표에 따라 연구방법은 크게 두 부분으로 구분된다:

- weather generator 통계모형에서 경년변동성의 모수화(parameterization)와 추정<sup>1)</sup>,
- 경년변동을 설명하는 예측인자를 APCC 계절예측에서 탐색, 그리고 이를 기반으로 한 예측모형의 수립.

1) 통계학에서 말하는 추정(estimation)은 오차를 수반한다. 즉, 추정문제에서 대상인 모수의 참값이 있고, 추정량은 표본의 불확실성에 의해 참값과 차이가 있게 마련이다. 하지만 본 연구에서는 이를 고려하지 않는다. 즉, 추정량의 값이 바로 모수의 참값으로 간주한다.

## 2.2.1 유역기후의 경년변동성 표현을 위한 통계모형

유역기후의 통계모형은 강수와 기온을 대상변수로 한다. 두 변수에 대한 통계모형을 간단히 소개하고 이어 경년변동성을 나타내는 모수를 도입한다.

### 2.2.1.1 유역강수모형

유역강수모형은 습윤/건조기 교대 알고리즘(wet/dry-spell alternation algorithm)으로 기반으로 한다. 이는 습윤기(wet-spell)와 건조기(dry-spell)가 번갈아 가며 나타나는 자연기상현상을 본뜬 것으로 Semenov et al. (1998)의 연구에서 LARS-WG에 도입되었다. 이 모형에서 습윤기와 건조기는 번갈아 가며 독립적으로 생성된다. 건조기는 정의상 건조일<sup>2)</sup>의 지속으로, 그 생성은 지속시간을 일단위로 모의하면 된다. 지속시간에 대한 통계모형으로 음이항분포(negative binomial distribution)를 고려한다. 즉,  $L$ 이 지속을 나타내는 확률변수일 때,

$$\Pr(L = m + 1) = \frac{\Gamma(m + \theta)}{m! \Gamma(\theta)} \left( \frac{\mu^L - 1}{\theta + \mu^L - 1} \right)^m \left( \frac{\theta}{\theta + \mu^L - 1} \right)^\theta, \quad m = 0, 1, 2, \dots$$

여기서,  $\mu^L > 1$ 로서  $L$ 의 기댓값이고  $\theta > 0$ 으로 음이항분포의 형태모수(shape parameter)이다. 특히,  $\theta = 1$ 일 때, 음이항분포는 기하분포가 되며 이는 Markov 모형에 대응된다. 이러한 관점에서 음이항분포를 이용한 건조기 생성방식은 Markov 모형에 의한 것보다 더 일반적인 방식이라 말할 수 있다.

한편, 습윤기 생성은 더 복잡하다: 습윤기는 습윤일<sup>3)</sup>의 연속으로 정의되는데, 지속일 뿐만 아니라 강수량을 생성해야 하기 때문이다. 지점의 강수량 모형으로 감마분포를 고려한다:  $R$ 이 어느 지점의 일강수량을 나타내는 확률변수일 때,

$$\Pr(R \leq x | R > 0) = \frac{\alpha^\alpha}{\Gamma(\alpha) \{\mu^R\}^\alpha} \int_0^x z^{\alpha-1} \exp\left(-\frac{\alpha z}{\mu^R}\right) dz, \quad x > 0.$$

여기서,  $\mu^R = E(R | R > 0)$ 으로서 강수강도(precipitation intensity)라고 부르고,

2) 건조일(dry day): 유역의 모든 관측지점에서 강수가 없는 날

3) 습윤일(wet day): 대상유역의 어느 한 지점에서라도 강수가 있는 날

$\alpha > 0$ 는 감마분포의 형태모수이다. 여러 지점의 강수강도를 표현하기 위해 지점  $o$ 의 강수강도를  $\mu^R(o)$ 라고 표기한다.

유역강수의 경년변동을 표현하기 위해서 건조지속일 평균  $\mu^L$ 와 강수강도  $\mu^R(o)$ 의 경년변동을 고려한다. 이를 강조하기 위해서 각각을  $\mu_i^L$ 와  $\mu_i^R(o)$ 로 표기하며, 여기서 첨자  $i$ 는 해당년도를 지칭한다. 이들 모수를 각 연도에 대해 계절성을 감안하여 월별로 추정하는데, 이 추정에 있어 자료결핍문제가 부각된다. 예를 들어, 우리나라 겨울철을 기준으로 3달 동안 대략 16회 정도의 건조기가 나타나기 때문에 건조지속일의 표본평균은  $\mu_i^L$ 의 안정적인 추정량이 되지 못한다.

안정적인 추정을 위해 다음과 같은 가중평균(weighted average)기법을 고려한다: 월누적강수량의 유역평균  $A_i$ <sup>4)</sup>는 유역전체의 습윤정도(wetness)를 나타내는 대표적인 척도이다. 강수강도의 지점일강수량월표본평균<sup>5)</sup>과  $A_i$ 간 관계를 살펴보면 양의 상관관계가 확인된다. 그림 3는 그 예로서, 12월에 무작위로 선택된 4개 지점에서 뚜렷한 양의 상관관계를 드러낸다. 그래서,

$$\log \mu_i^R(o) = \beta_0(o) + \beta_1(o) \log A_i$$

을 연결함수(link function)로 하는 감마-일반화선형모형(generalized linear model)을 적합하여  $\mu_i^R(o)$ 의 적합값을 얻는다. 하지만 적합값은 유역전체의 습윤정도에 의해 설명되는 해당지점의 강수강도로 지점 고유의 경년변동이 드러나지 않는다. 그래서 최종적인 추정량은 지점일강수량월표본평균과 적합값의 가중평균으로 구한다. 단, 가중치는 교차검증을 통해 정한다. 교차검증을 통한 가중치 설정방법은 부록에서 다룬다.

4) 월누적강수량의 유역평균: 지점별 월누적강수량의 유역평균

5) 지점일강수량월평균: 지점습윤일의 일강수량 월표본평균. 지점습윤일: 해당지점에 강수량이 0.1mm 이상인 날

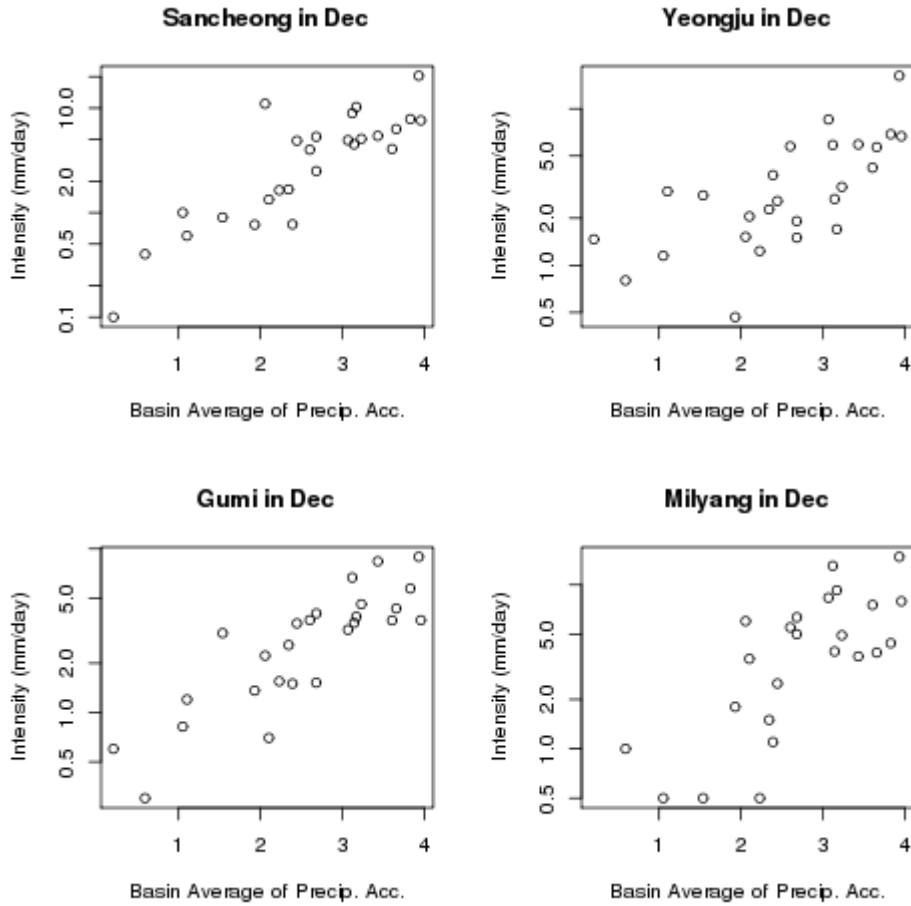


그림 3 12월 월누적강수량 유역평균( $x$ 축)과 지점 일강수량월표본평균( $y$ 축)의 산점도, ( $x, y$ 축은 log-scale). 뚜렷한 양의 상관관계를 보인다.

건조지속일 평균  $\mu_i^L$ 에 대해서 유사한 추정방식을 고려한다. 기초적인 추정량으로 건조지속일 표본평균을 고려할 수 있는데, 그림 4에서처럼  $A_i$ 와 음의 상관관계를 보인다. 그래서, 관계식

$$\log(\mu_i^L - 1) = \beta_0 + \beta_1 A_i$$

을 연결함수로 하는 음이항-일반화선형모형을 적합하여,  $\mu_i^L$ 의 적합값을 얻는다. 역시, 강수강도와 같은 방식으로 적합값과 건조지속일 표본평균의 교차검증을 통한 가중평균으로  $\mu_i^L$ 을 추정한다. 자세한 방식은 부록에서 설명한다.

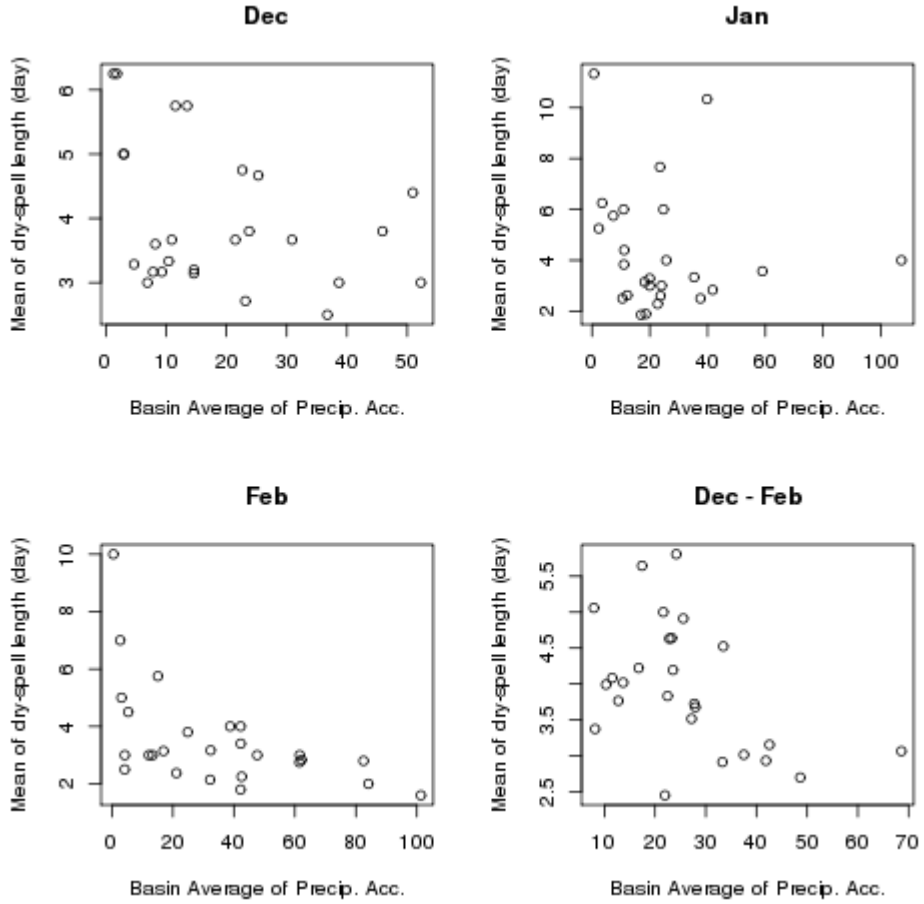


그림 4 각 월 그리고 3개월 누적강수량 유역평균( $x$ 축)과 건조기지속일 표본평균( $y$ 축)의 산점도. 음의 상관관계를 보인다.

### 2.2.1.2 유역기온모형

다음과 같은 유역기온모형을 고려한다:  $T(o, t)$ 와  $R(o, t)$ 가 각각  $o$ 지점,  $t$ 일의 일 최고기온( $T = T_{\max}$ ) 또는 일최저기온( $T = T_{\min}$ )을 그리고 강수량을 나타낸다고 할 때,

$$T(o, t) - \mu_T(0, t) = \beta_0^T(o) + \sum_{k=1}^2 \sum_{j=1}^2 \beta_{k,j}^T(o) U_{k,j}(t) + (\text{anomaly})$$

을 가정한다. 여기서,  $\mu_T(o, t)$ 는 기온의 연주기순환추세성분(annual cyclic trend),  $U_{k,1}(t)$ 와  $U_{k,2}(t)$ 는

$$R^{\gamma}(o_1, t-k+1), \dots, R^{\gamma}(o_d, t-k+1)$$

의 첫 번째와 두 번째의 EOF 계수를 나타낸다.<sup>6)</sup> 식 (1)는 강수에 따른 기온편차의 일평균변화를 표현한 회귀모형식으로,  $k=1$ 은 당일의 강수,  $k=2$ 은 이전일의 강수에 대응한다. 또한, (2)의  $\gamma > 0$ 은 회귀모형의 선형성을 위해 도입되었는데, 대체로  $\gamma = 0.1$  정도로 0에 가깝게 추정된다. 이는 강수량이 증가함에 따라 그 영향력이 체감함을 뜻하는데, 대기 중 수증기량이 포화상태에 이를수록 영향력증가가 누그러짐과 상통한다.

기온모형의 경년변동성을 부여하기 위해, 모형식 (1)의 절편항  $\beta_0^T(o)$ 의 경년변동을 가능하게 하고,  $\beta_{0,i}^T(o)$ 으로 표기한다. 즉, 식 (1)이

$$T(o, t) - \mu_T(0, t) = \beta_{0,i}^T(o) + \sum_{k=1}^2 \sum_{j=1}^2 \beta_{k,j}^T(o) U_{k,j}(t) + (\text{anomaly})$$

으로 바뀌며,  $\beta_{0,i}^T(o)$ 는 강수효과가 제거된 일기온의 월평균수준을 의미한다.  $\beta_{0,i}^T(o)$ 의 추정은 최소제곱법을 이용한다.

경년변동성을 표현하기 위해 도입된 모수를 표 1에서 정리하였다. 이 밖에 통계모형의 주요한 모수로 음이항분포와 감마분포의 형태모수, 강수효과를 나타내는 기온 회귀모형 (1)의 계수  $\beta_{k,j}^T(o)$  등이 있는데, 본 연구에서는 이들 모수가 유역의 특징에 의해 결정되는 측면이 강하다고 판단되어 경년변동성을 고려하지 않는다.

표 1            경년변동성을 나타내는 유역기후 통계모형의 모수

| 모수          | 표기                   | 설명                     |
|-------------|----------------------|------------------------|
| 강수강도        | $\mu_i^R(o)$         | 평균일강수량. 강수세기에 비례.      |
| 건조지속일수 평균   | $\mu_i^L$            | 평균건조지속일수. 건조정도에 비례.    |
| 일최고기온모형 절편항 | $\beta_i^{T\max}(o)$ | 강수효과가 제거된 일최고기온의 월평균수준 |
| 일최저기온모형 절편항 | $\beta_i^{T\min}(o)$ | 강수효과가 제거된 일최저기온의 월평균수준 |

6)  $o_1, \dots, o_d$ : 는 관측지점을 나타낸다.  $d$ 는 해당유역의 관측지점의 수.

## 2.2.2 경년변동성 예측모형

유역기후의 경년변동은 전지구적 기후변동과 관련이 있음은 분명하다. 그래서 표 1에서 제시된 모수를 대규모 기후와 연결하여 유역기후의 경년변동을 표현하고자 한다. 본 연구보고서에서는 이 연결에 통계적 방식을 적용한다.

### 2.2.2.1 predictand의 공변동성

본격적인 연결방식을 설명하기 앞서, 모수들의 공변동성(共變動性)을 고려할 필요가 있겠다. 강수강도와 기온모형의 절편항은 지점별로 계산되지만, 지점들이 한 유역에 위치하기 때문에 모수값들이 공변동한다. 실제로 EOF 분석에 따르면 첫 번째 공간모드에 의해 전체변동의 약 90%가 설명되며 그 모드는 유역 전반에 고른 형태를 보인다. 또한 그 외의 공간모드는 유역의 특이변동으로 이해하는 것이 합리적이라 판단되고, 따라서 대규모 기후와 연결시켜 설명하기에 적절하지 않아 보인다. 그래서 예측모형에서는 첫 번째 EOF 계수를 predictand로 한다.

### 2.2.2.2 예측인자의 탐색과 관측장 분석

예측모형수립의 첫 번째 단계는 예측인자의 탐색이다. 탐색을 위해 기존문헌을 참고하거나 탐색적 분석(exploratory analysis)을 통한다. 탐색을 통해 선별된 잠재예측인자의 예측력을 평가하기 위해 회귀모형

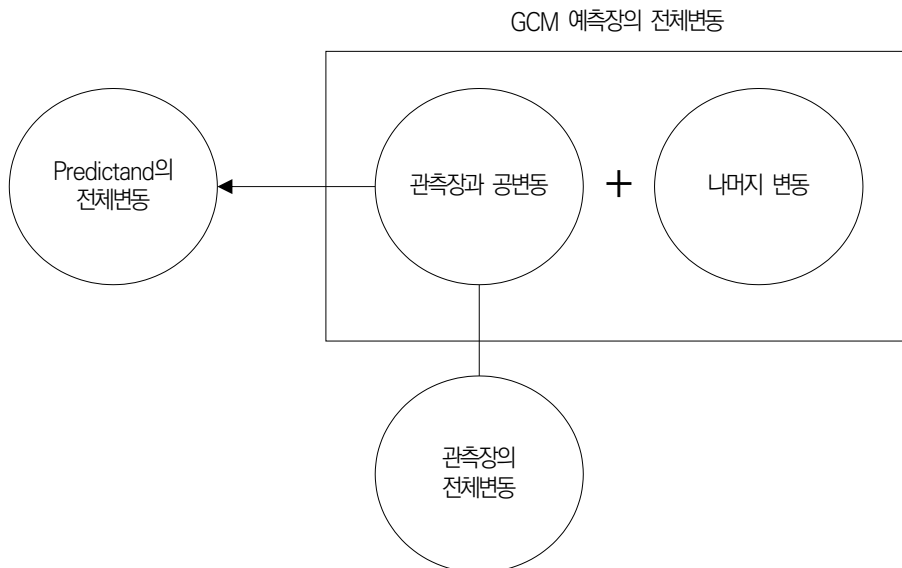
$$Y_i = \beta_0 + \sum_{j=1}^p \beta_j X_i(s_j) + \epsilon_i, \quad i = 1, \dots, n,$$

을 고려한다. 여기서,  $X_i(s)$ 는 잠재예측인자의 관측장(observation field)을 나타내고  $s = s_1, \dots, s_p$ 는 관측격자점을 뜻한다. 그리고,  $Y_i$ 은 predictand이며,  $i = 1, \dots, n$ 은 관측기간을 의미한다. 식 (4)은 관측장의 패턴에 따른 predictand의 반응을 나타내는데, 고려되는 격자영역이 대규모인 관계로 격자점의 수  $p$ 가  $n$ 보다 훨씬 더 크기 때문에 관측장의 복잡성(complexity)으로 인해 predictand의 전체변동이 완전히 설명된다. 그래서, 자료의 복잡성을 조절하는 추정법인 능형회귀방법(ridge regression

method)을 쓴다, (cf. Myers (1990)). 부록에서 능형회귀방법에 의해서 복잡성이 조절되는 원리를 설명한다. 기후변수가 식 (4)을 통해 측정된 예측력이 수용할 수 있는 수준이면 예측인자로 선택된다.

### 2.2.2.3 GCM 예측장 분석

두 번째 단계는 탐색된 예측인자를 기반한 예측회귀모형을 유도하는 것이다. 그런데 관측장은 예측대상시점 이후에 생산되기 때문에 실제적인 예측에는 활용될 수 없고, 이를 대신해 GCM의 예측장(prediction field)을 쓴다. 하지만, 관측장이 예측장으로 대체된 식 (4)을 예측모형으로 쓰면 GCM 고유의 편의(bias) 때문에 예측모형 해석이 어려워지고 과적합(overfitting)의 우려가 있다. 이러한 우려를 완화시키기 위해, 우선 GCM의 대규모 순환 예측에 대한 탐색적 분석을 시도한다. 그래서 GCM 예측장의 전체변동에서 관측장과 강한 상관이 있는 부분변동을 찾고, 이를 기반해 예측모형을 세움으로써 과적합문제를 완화하고자 한다. 그림 5는 예측모형의 개념을 도시하였다.



**그림 5** 예측모형의 개념도. GCM 예측장을 예측인자로 그대로 쓰는 경우 과적합이 발생할 수 있다. RCCA를 통해 GCM 예측장의 전체변동을 관측장과 강하게 공변동하는 부분과 나머지 변동으로 나누고, 이 중에 공변동 부분만을 predictand의 예측인자로 쓴다. 예측장을 그대로 쓰는 경우는 나머지 변동이 예측모형에 관여해 과적합을 일으키는 것으로 판단된다.

분석방법으로 RCCA (regularized canonical correlation analysis, regularized 정준상관분석)을 고려한다, (cf. Vinod (1976)). 이 분석법을 통해 관측장과 예측장의 상관관계가 큰 모드쌍을 얻을 수 있다: RCCA를 통해 관측장과 예측장의 계수벡터  $\phi_1, \dots, \phi_n$ 과  $\psi_1, \dots, \psi_n$ 를 얻는데, 각각 관측장과 예측장에서 서로 대응하는 공간모드로 이해할 수 있다. 또한,  $Z_i, i = 1, \dots, n$ 가 예측장을 나타낸다고 하면,  $(\langle X_i, \phi_j \rangle, \langle Z_i, \psi_j \rangle)$ ,  $i = 1, \dots, n$ 의 상관계수를  $j$ 번째 공간모드 간의 상관계수로 볼 수 있다. 한편, 예측문제 관점에서 교차검증을 통한 상관계수를 살펴보는 것이 필요하다:  $\phi_{1,i}, \dots, \phi_{n,i}$ 와  $\psi_{1,i}, \dots, \psi_{n,i}$ 를  $i$ 번째 관측년도를 전체자료에서 제외한 채 얻어진 계수벡터라고 할 때,  $(\langle X_i, \phi_{j,i} \rangle, \langle Z_i, \psi_{j,i} \rangle)$ ,  $i = 1, \dots, n$ 의 상관계수를 편의상 CV-상관계수라고 부른다. CV-상관계수가 0.9이상인 예측장의 공간모드만을 선별하여 예측모형에 사용한다.

#### 2.2.2.4 회귀모형

RCCA를 통해 선별된 예측장의 공간모드를 이용하여 회귀모형을 도출한다. 계절 예측에 참여하는 GCM이 여러 개이기에 구분을 위해서 예측장을  $Z_i^{(l)}, i = 1, \dots, n$ ,  $l = 1, \dots, L$ 로 표기한다.  $L$ 은 참여모형의 갯수이다. 또한  $\phi_1^{(l)}, \dots, \phi_n^{(l)}$ 과  $\psi_1^{(l)}, \dots, \psi_n^{(l)}$ 이 각 GCM에 대한 RCCA의 계수벡터를 나타낸다.  $l$ 번째 GCM 예측장에서 선별된 공간모드의 집합을  $J_l$ 이라고 할 때, 예측모형으로

$$Y_i = \beta_0 + \sum_{l=1}^L \sum_{j \in J_l} \beta_{l,j} \langle Z_i^{(l)}, \psi_j^{(l)} \rangle + \epsilon_i, \quad i = 1, \dots, n,$$

을 고려한다. 식 (5)에서 다시 predictand에 대한 예측력이 유의하지 않은 공간모드는 변수선택법<sup>7)</sup>에 의해 제외된다. 최종적으로 선별된 공간모드의 집합을  $J_l^*$ ,  $l = 1, \dots, L$ 라고 두면,

$$Y_i = \beta_0 + \sum_{l=1}^L \sum_{j \in J_l^*} \beta_{l,j} \langle Z_i^{(l)}, \psi_j^{(l)} \rangle + \epsilon_i, \quad i = 1, \dots, n,$$

얻는다.

7) 여기서는 LASSO 추정법을 썼다, (cf. Tibshirani (1996)).

### 2.2.2.5 오차항 모형

예측모형의 마지막 단계는 오차항 모형이다. 예측인자가 predictand의 전체변동을 설명하지 못하기 때문에, 예측의 완결을 위해서 나머지 변동을 나타낼 수 있도록 오차항  $\epsilon_i$ 의 생성이 필요하다. 더구나 여러 predictand의 오차항 사이에 상관관계가 있기 때문에 이에 대한 고려도 필수적이다. 이를 위해서, PRESS 잔차

$$e_i = Y_i - \hat{Y}_{i,(-i)}, \quad i = 1, \dots, n,$$

를 이용한다, (cf. Myers (1990)). 여기서,  $\hat{Y}_{i,(-i)}$ 는 전체자료에서  $i$ 번째 연도를 제외한 채 적합한 모형에서 얻어진  $Y_i$ 의 예측값을 나타낸다. 오차항의 비정규성을 감안하여, 우선 각 predictand마다 PRESS 잔차의 z-점수<sup>8)</sup>를 얻어 상관계수행렬을 계산하고, 그것을 공분산행렬로 하는 zero 평균벡터의 다변량정규분포로써 z-점수의 확률분포를 표현한다. 오차항생성에서는 다변량정규분포에서 새롭게 생성된 z-점수으로부터 z-점수화의 역변환을 통해 각 predictand의 새로운 오차항  $\epsilon_*$ 을 얻는다. 그래서 새로운 예측장  $Z_*^{(1)}, \dots, Z_*^{(L)}$ 이 주어졌을 때 대응하는 predictand의 값은

$$\beta_0 + \sum_{l=1}^L \sum_{j \in J_l^*} \beta_{l,j} \langle Z_*^{(l)}, \psi_j^{(l)} \rangle + \epsilon_*$$

이다.

### 2.2.2.6 Weather Generator의 모수 조정 및 구동

식 (7)으로 predictand의 값이 예측되었을 때 최종적으로 모수의 값을 정해야 할 것이다. 건조기지속일수평균은 그 자체가 predictand이기 때문에 더 이상의 과정이 필요 없지만 다른 모수의 경우는 공변동성으로 인해 첫 번째 EOF 계수를 예측하기 때문에 추가적인 절차를 거쳐야 한다. 예로 강수강도의 경우를 생각해 본다. EOF 분석의 고유값과 모드를  $\lambda_1, \dots, \lambda_d$ 와  $v_1, \dots, v_d$ 로 표현할 때, 예측강수강도

$$\mu_*^R = (\mu_*^R(o_1), \dots, \mu_*^R(o_d))$$

8) z-점수(z-score): 자료값의 percentile에 대응하는 표준정규분포의 분위수.

$$\mu_*^R = \langle \mu_*^R, v_1 \rangle v_1 + \dots + \langle \mu_*^R, v_d \rangle v_d,$$

으로 쓰여 지고, 여기서  $\langle \mu_*^R, v_k \rangle$  ( $k = 1, \dots, d$ )의 분산은  $\lambda_k$ 로 추정된다. 여기서  $\langle \mu_*^R, v_1 \rangle$ 은 예측모형으로 예측되지만 나머지 계수는 그렇지 않기 때문에 적절한 생성이 필요하다. 계수는 확률적으로 직교하기 때문에 정규성 가정 하에서 독립적으로 생성할 수 있다. 그래서 평균  $\frac{1}{n} \sum_{i=1}^n \langle \mu_i^R, v_k \rangle$ , 분산  $\lambda_k$ 인 정규분포로부터  $\langle \mu_*^R, v_k \rangle$  ( $k = 1, \dots, d$ )의 값을 생성하여 강수강도를 얻는다. 기온모형의 절편항도 같은 방식으로 처리할 수 있다. 이와 같이 계절예측을 통해 경년변동을 나타내는 모수의 값이 모두 정해지면 2015년 보고서에서 제시된 알고리즘에 따라 weather generator를 구동하여 상세화 자료를 생산한다, (cf. 김무섭 (2015)).

### 3. 연구내용

#### 3.1 낙동강 유역에서 경년변동성 통계모형 적합결과

본 절에서는 대상지역인 낙동강 유역에 앞서 설명한 통계모형을 적합하여 경년변동성을 살펴본다. 유역의 지점은 총 14개로 각각  $o_1, \dots, o_{14}$ 로 표현하며, 지칭하는 지점은 그림 2에서 찾을 수 있다.

##### 3.1.1 유역강수모형 적합결과

앞서 설명한 추정방식에 따라 강수강도  $\mu_i^R(o)$ 을 계산하고 경년변동성을 살펴본다. 그림 6에서처럼 강수강도는 겨울철 3개월마다 다른 경년변동성을 보이고 있다. 또한, 뚜렷한 공변동성을 보이고 있다. 그래서 EOF 분석을 통해 주요모드와 EOF 계수를 얻는다. 첫 번째 EOF 계수가 예측모형에서 predictand로 쓰인다. 그림 7는 EOF분석 결과인데, 첫 번째 모드가 전체변동의 90% 가까이 설명하고 있음을 알 수 있으며, 이는 공변동성을 뒷받침한다. 그림 8은 각 월에 대한 EOF분석에 얻어진 EOF 계수의 시계열 그림이다. 그림 6에서처럼 월마다 다른 경년변동성을 보이고 있는데, 12월이 상대적으로 약한 그리고 2월은 강한 경년변동성을 보이고 있다. (12월은 표준편차가 7.53 mm/day, 1월은 8.26 mm/day, 2월은 9.05 mm/day이다.)

건조기지속일수 평균  $\mu_i^L$ 의 경년변동성은 그림 9에서 보이고 있듯이 각 월마다 다른 변동성을 보이고 있는데, 12월은 표준편차가 0.34일 정도로 매우 약하며, 1월은 0.91이나 몇몇의 큰 변동으로 인한 것으로 보이며, 2월은 0.78일로 전 기간에 걸쳐 가장 뚜렷한 변동성을 나타낸다.

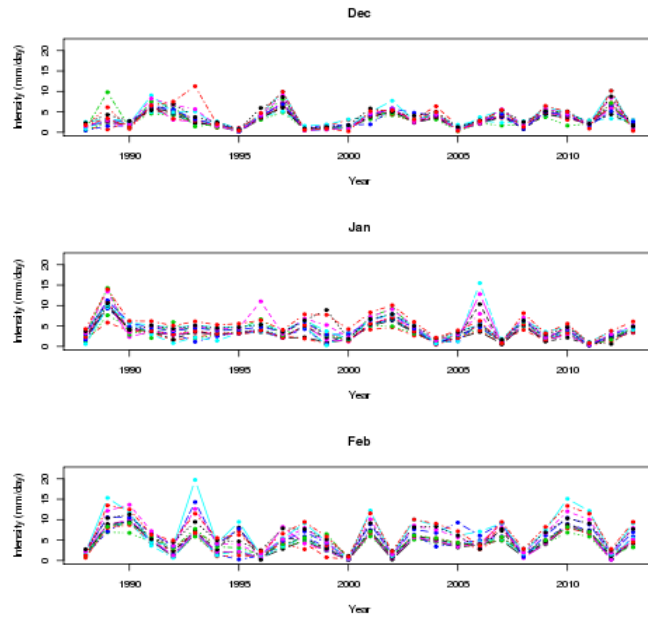


그림 6 낙동강유역 14개 지점 겨울철 강수량도  $\mu_i^R(o_j)$ ,  $j=1, \dots, 14$ 의 경년변동성. 월마다 다른 정도의 경년변동성을 보인다. 한편, 모든 월에서 공변동성이 뚜렷이 나타난다.

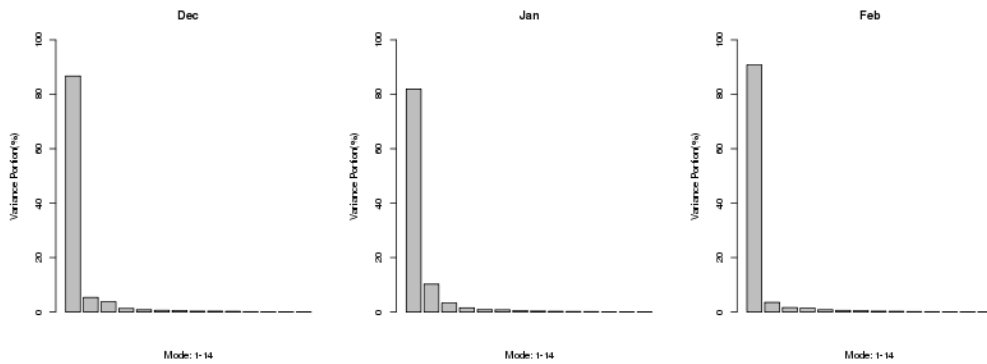


그림 7 EOF 분석결과-변동비율. 첫 번째 모드가 모든 월에서 전체변동의 90% 가까이를 설명하고 있다.

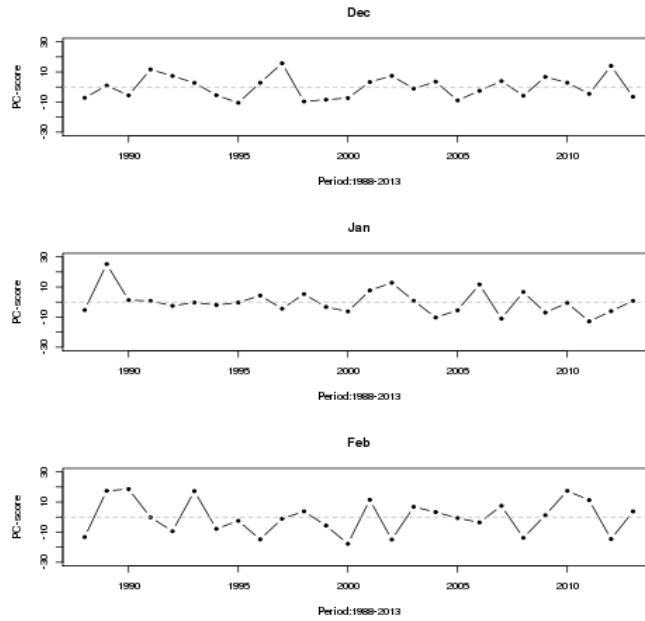


그림 8 강수강도에 대한 첫 번째 모드의 EOF 계수 시계열 그림. 12월에서 약한 경년변동성을 2월에서는 강한 경년변동성을 보인다. (표준편차, 12월: 7.53 mm/day, 1월: 8.26 mm/day, 2월: 9.05 mm/day)

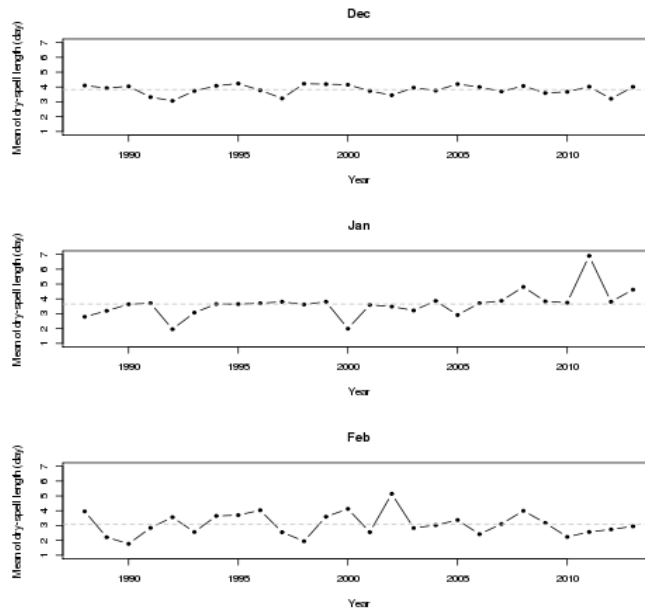


그림 9 각 월에서 건조지속일수 평균  $\mu_i^L$ 의 경년변동성. 월마다 다른 변동성을 나타내는데, 특히 12월의 변동성은 매우 약하다. 1월은 대체로 약한 변동을 나타내나 드물게 강한 변동을 보이고, 2월은 3개월 중 가장 두드러진 변동을 보인다. (표준편차, 12월: 0.34일, 1월: 0.91일, 2월: 0.78일.)

### 3.1.2 유역기온모형 적합결과

유역기온모형 (3)에서 절편항  $\beta_{0,i}^{Tmax}(o)$ 과  $\beta_{0,i}^{Tmin}(o)$ 의 경년변동성을 살펴보자. 그림 10-11에서처럼 절편항이 뚜렷한 경년변동성을 보이고 있다. 또한 명확한 공변동성이 관찰되는데, EOF 분석에 따르면 강수강도에서처럼 일최고기온의 경우에는 첫 번째 모드가 전체변동의 90% 이상을, 일최저기온의 경우는 약 87%를 설명한다. 그래서, 예측모형에서는 첫 번째 EOF 계수를 predictand로 정한다. 그림 12-13은 첫 번째 EOF 계수의 시계열 그림인데, 모든 월에서 대등한 경년변동을 보인다.

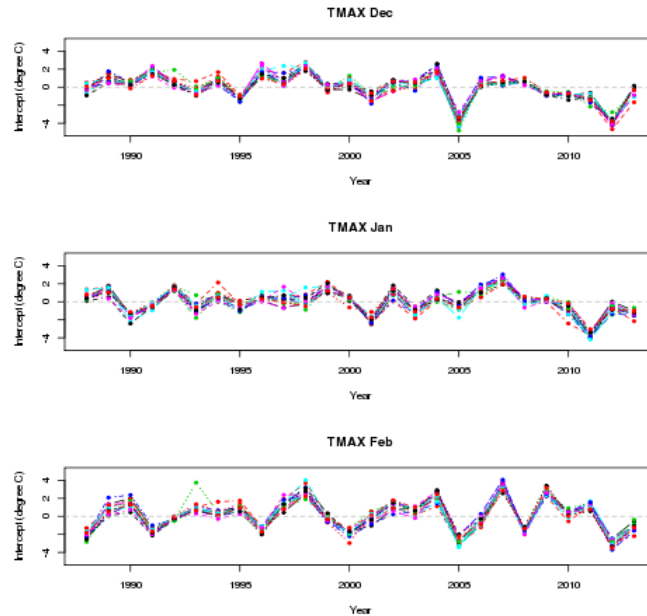


그림 10 유역일최고기온모형에서 절편항  $\beta_{0,i}^{Tmax}(o_j)$ ,  $j=1, \dots, 14$ 의 경년변동성. 모든 월에서 뚜렷한 경년변동성과 공변동성이 보인다.

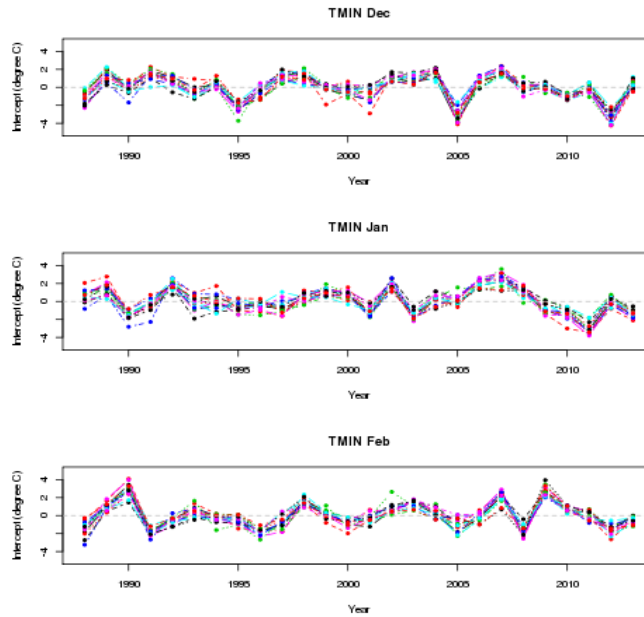


그림 11 유역일최저기온모형에서 절편항  $\beta_{0,i}^{Tmin}(o_j)$ ,  $j = 1, \dots, 14$ 의 경년변동성. 일최고기온의 경우와 같이 모든 월에서 뚜렷한 경년변동성과 공변동성이 보인다.

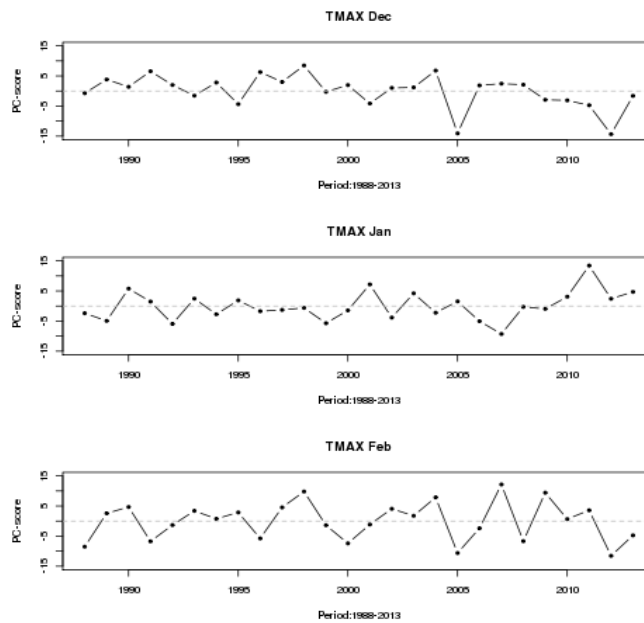


그림 12 일최고기온 절편항의 첫 번째 EOF 계수의 시계열 그림. 모든 월에서 대등한 경년변동성을 보이고 있다. 표준편차 12월: 섭씨 5.45도, 1월: 섭씨 4.81도, 2월: 섭씨 6.41도.

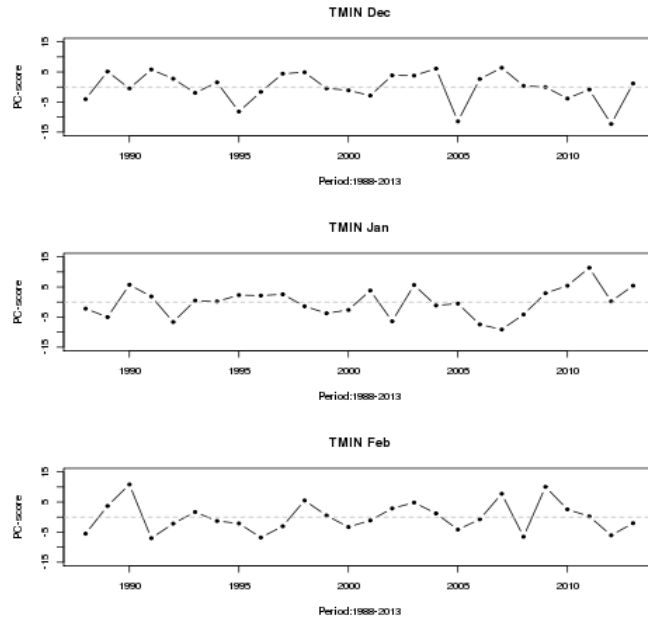


그림 13 일최저기온 절편항의 첫 번째 EOF 계수의 시계열 그림. 일최고기온의 경우처럼 모든 월에서 대등한 경년변동성을 보이고 있다. 표준편차 12월: 섭씨 4.99도, 1월: 섭씨 4.86도, 2월: 섭씨 5.00도.

## 3.2 낙동강 유역의 경년변동에 대한 예측모형

### 3.2.1 탐색적 사전분석

예측인자를 찾기 위한 사전분석으로서 겨울철 3개월 평균 관측장과 동기간에서 계산된 predictand 간의 상관계수그림을 그려보았다. 그림 14-15은 각각 일최고/최저 기온모형의 절편항의 첫 번째 EOF 계수와 Z500간의 상관계수를, 그림 16-17은 강수 강도의 첫 번째 EOF 계수/건조지속일수평균과 SLP 간의 상관계수를 격자점별로 나타낸다. 단, 그림에서는 5%수준에서 유의한 격자점에서만 양의 상관(적색) 또는 음의 상관(청색)을 표시한다. Z500의 경우 동아시아, 유럽, 인도양 부근에서 양의 상관 관계를, 극지방에서 음의 상관관계를 보인다. 한편, SLP는 한반도를 동서로 상반된 상관관계가 나타나고 있으며 유역과 거리가 먼 대서양에서도 상관관계를 보이는 영역이 드러난다. 이와 같은 사실을 바탕으로 Z500의 영역은 서경35도-동경200도, 북위30도-북위90도로, SLP는 서경180도-동경180도, 북위30도-북위60도로 설정한다. 그림 18에서 예측인자의 영역을 도시하였다.

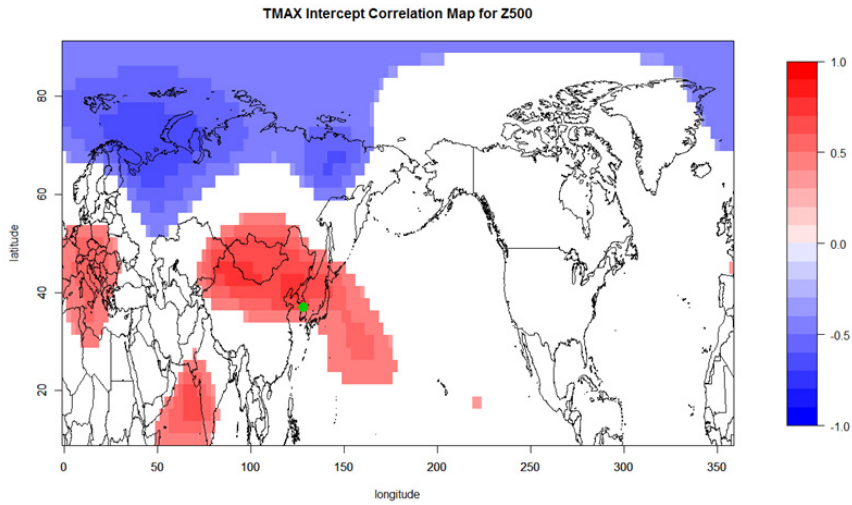


그림 14 일최고기온모형의 절편항과 Z500 관측장 간의 상관계수 그림. 동아시아, 인도양, 유럽대륙에서 양의 상관관계, 극지방을 중심으로 음의 상관관계가 유의하게 나타난다. (5% 수준에서 유의한 영역만을 표시)

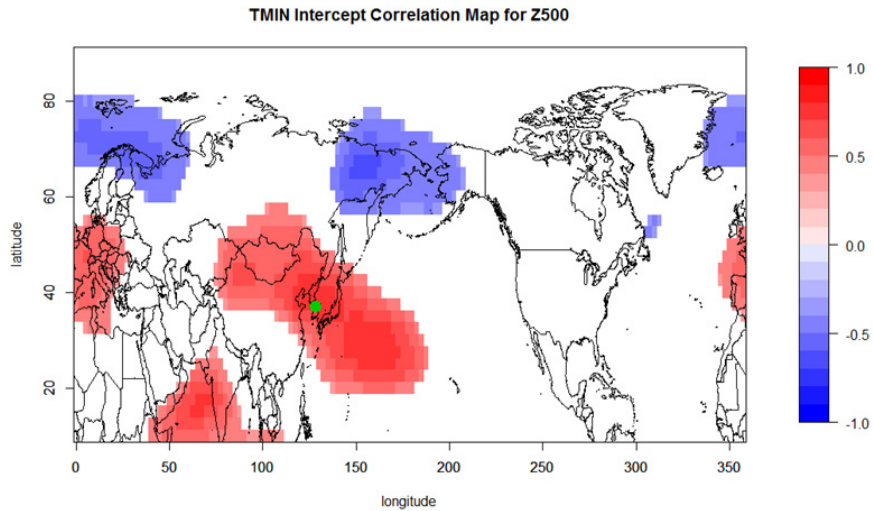


그림 15 일최저기온모형의 절편항과 Z500 관측장 간의 상관계수 그림. 일최고기온의 경우와 유사한 패턴을 나타낸다. (5% 수준에서 유의한 영역만을 표시)

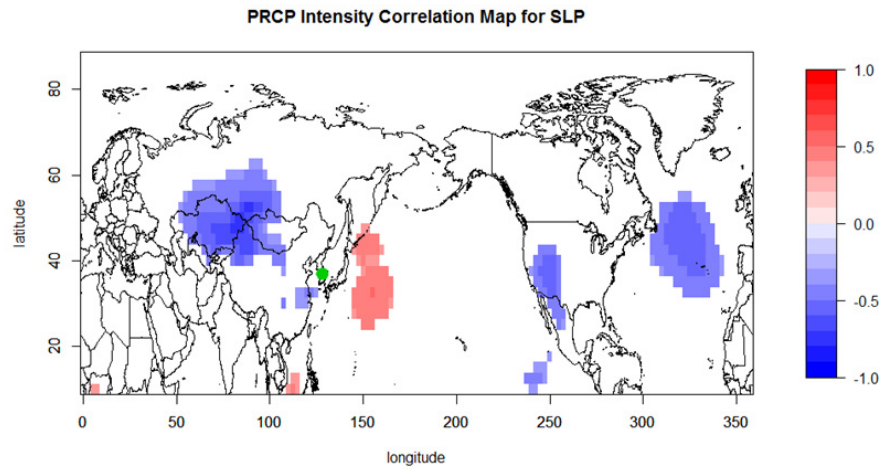


그림 16 강수강도와 SLP 간 상관계수 그림. 한반도 동서로 상반된 상관관계가, 북대서양에서 음의 상관관계가 유의하게 나타나고 있다. (5% 수준에서 유의한 영역만을 표시)

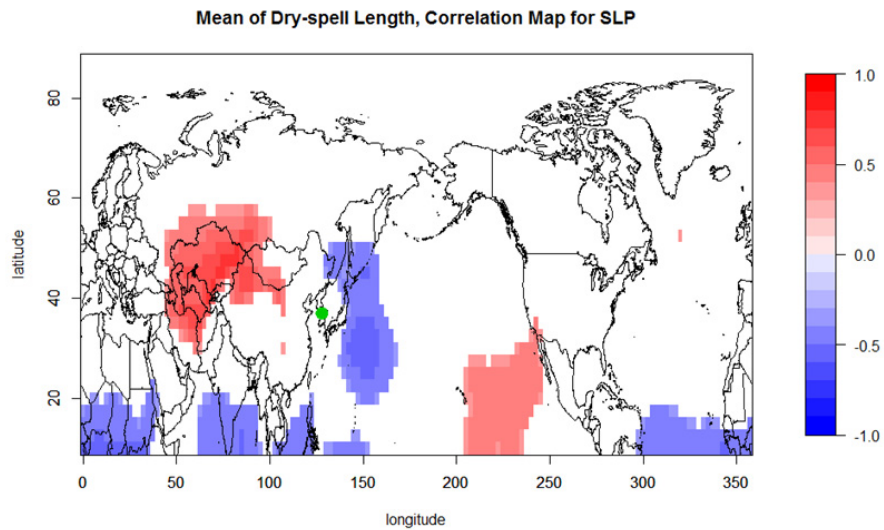


그림 17 건조기지속일수와 SLP 간 상관계수 그림. 한반도의 동서로 상반된 상관관계가 나타난다. 또한, 적도 인근에 유의한 상관관계가 나타나고 있다. (5% 수준에서 유의한 영역만을 표시)

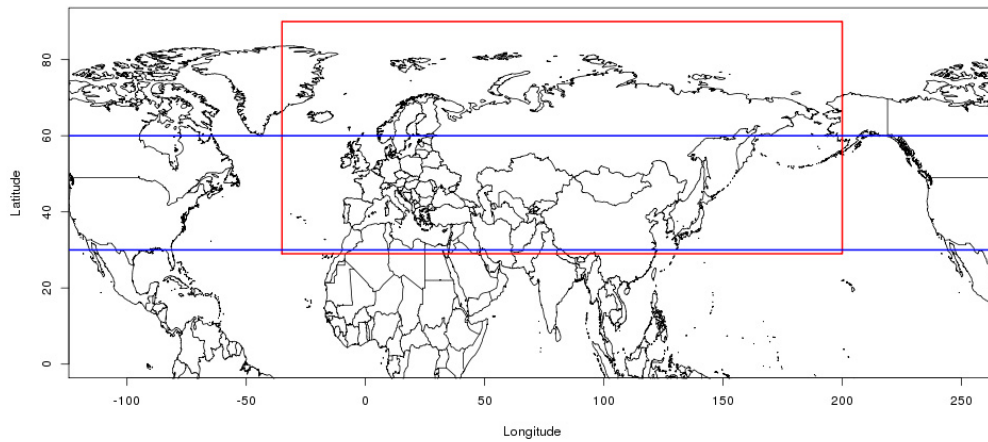


그림 18 Z500(적색)과 SLP(청색)의 영역

### 3.2.2 예측모형

각 월에 대해서 회귀모형 (4)을 통해 관측장에서 얻어진 예측인자의 예측력을 시험하였다. 예측력은 leave-one-out cross validation을 통해 얻어진 예측값의 MSE와 상관계수로 평가된다. 표 2은 예측인자의 성능을 평가한 결과이다. 단, 표에서 MSE ratio는 표본평균의 MSE에 대한 예측값의 MSE의 비율로 1보다 작을수록 예측력이 강하다는 것을 의미한다. 예측성능은 월별로 다른 특징이 나타나는데, 12월 경우에는 Z500과 SLP가, 2월은 Z500이 예측력을 보이고 있다. 한편 1월은 두 예측인자 모두 뚜렷한 예측성을 보이고 있지 못하다.

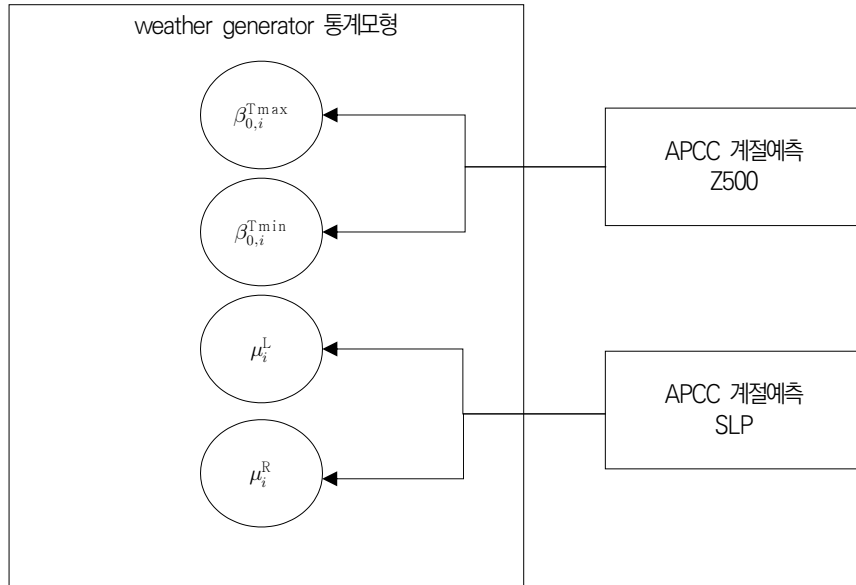


그림 19      경년변동성을 지니는 모수와 선택된 예측인자. 기온과 관계된 모수는 Z500, 강수와 관계된 것들은 SLP와 연결시킨다.

표 2            각 월별 관측장의 예측력 평가. 월마다 다른 예측성을 지니고 있다. 12월은 Z500과 SLP가 2월은 Z500이 예측성을 보이고 있다. 1월은 두 예측인자 모두 예측성을 보이고 있지 않다.  
음영표시: 예측성을 보이는 월과 predictand.

| 월   | predictand               | 예측인자 | MSE ratio | 상관계수 |
|-----|--------------------------|------|-----------|------|
| 12월 | 일최고기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.74      | 0.67 |
|     | 일최저기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 1.42      | 0.47 |
|     | 건조기지속일수 평균               | SLP  | 0.47      | 0.74 |
|     | 강수강도의 첫 번째 EOF 계수        | SLP  | 0.55      | 0.71 |
| 1월  | 일최고기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 1.39      | 0.22 |
|     | 일최저기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.98      | 0.39 |
|     | 건조기지속일수 평균               | SLP  | 1.35      | 0.31 |
|     | 강수강도의 첫 번째 EOF 계수        | SLP  | 2.10      | 0.12 |
| 2월  | 일최고기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.85      | 0.59 |
|     | 일최저기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.41      | 0.76 |
|     | 건조기지속일수 평균               | SLP  | 1.13      | 0.35 |
|     | 강수강도의 첫 번째 EOF 계수        | SLP  | 2.29      | 0.15 |

표 2에서 관측장이 예측성을 보이고 있는 월과 predictand의 경우(표 2에서 음영으로 표시)에 대해 예측모형 (6)을 도출하였다. 그림 20-24은 예측모형 (6)의 계수  $\beta_{l,j}$ 의 추정값을 도시하였다. 그림에서  $l = 1, \dots, 6$  (y축)은 GCM,  $j = 1, \dots, 10$  (x축)은 모드를 가리킨다. 대응하는 관측장과 예측장의 모드형태는 부록에서 주어진다. 12월의 경우에는 여러 예측장의 모드가 최종모형에 포함되었으나, 2월에 경우에는 매우 적거나 모두 제외되었다. 표 3의 예측모형의 성능을 평가한 결과인데, 12월에서는 관측장에 준하는 예측력이 보이지만 2월은 그렇지 못하다. (단, 12월은 선행시간이 1달전, 2월은 3달 전이다. 그래서 2월의 예측력이 낮은 것은 자연스러운 현상이다.) 그림 25-28은 predictand와 예측값의 산점도를 제시한다. 다음으로 예측모형의 오차항을 분석하였다. 그림 29-34를 보면, PRESS 잔차의 분포는 정규분포로 잘 근사가 되는 것으로 판단된다. 또한, 그림 35-37을 통해 다변량정규분포가 오차항 간의 상관관계를 적절히 표현하고 있음을 알 수 있다.

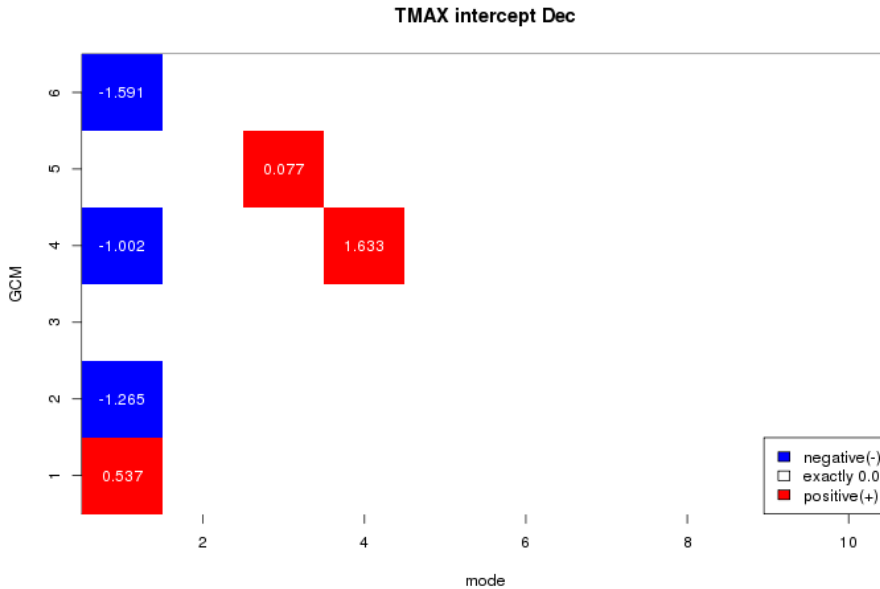


그림 20 12월 일최고기온모형 절편항의 첫 번째 EOF 계수에 대한 예측모형의 계수  $\beta_{l,j}$ ,  $l = 1, \dots, 6$ 은 GCM (y축),  $j = 1, \dots, 10$ 은 모드 (x축)를 뜻한다.  $\beta_0$ 의 추정값은 0.056.

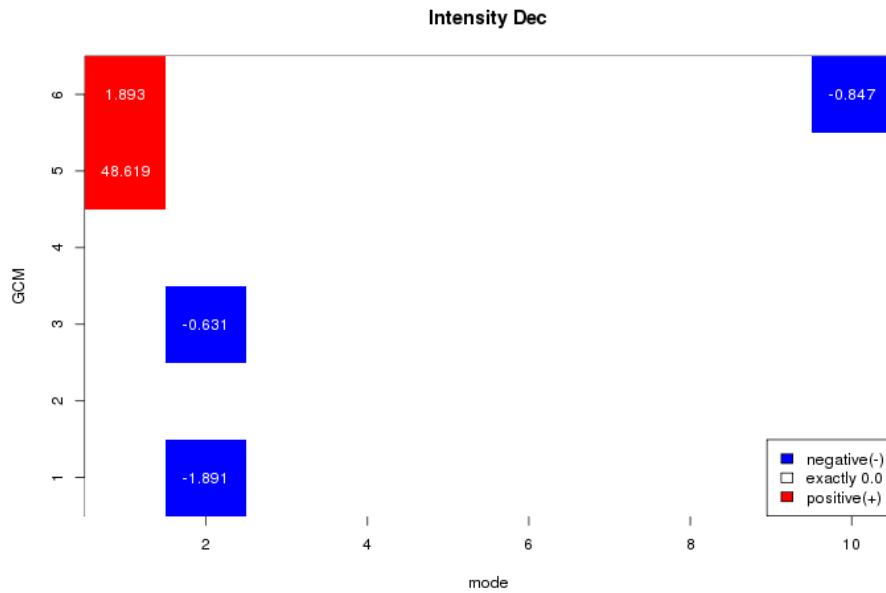


그림 21 12월 강수강도의 첫 번째 EOF 계수에 대한 예측모형의 계수  $\beta_{l,j}$ ,  $l=1,\dots,6$ 은 GCM (y축),  $j=1,\dots,10$ 은 모드 (x축)를 뜻한다.  $\beta_0$ 의 추정값은 0.115.

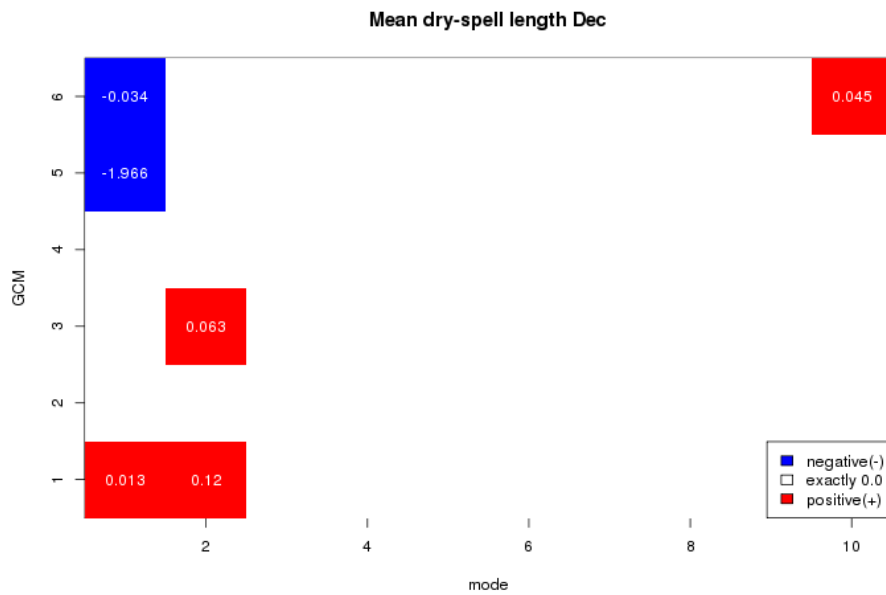


그림 22 12월 건조지속일수 평균에 대한 예측모형의 계수  $\beta_{l,j}$ ,  $l=1,\dots,6$ 은 GCM (y축),  $j=1,\dots,10$ 은 모드 (x축)를 뜻한다.  $\beta_0$ 의 추정값은 3.828

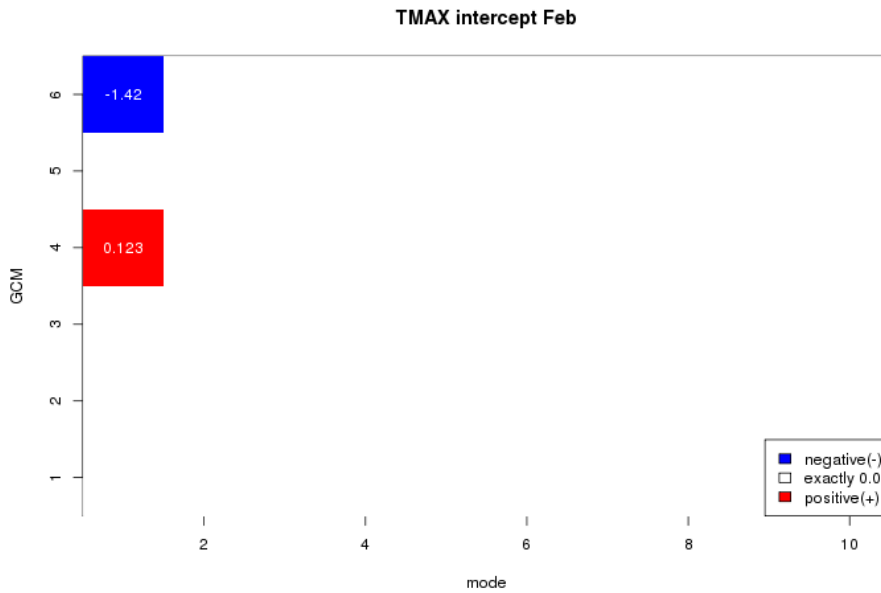


그림 23 2월 일최고기온 절편항의 첫 번째 EOF 계수에 대한 예측모형의 계수  $\beta_{l,j}$ ,  $l=1,\dots,6$ 은 GCM (y축),  $j=1,\dots,10$ 은 모드 (x축)를 뜻한다.  $\beta_0$ 의 추정값은 0.019. 총 2개의 모드가 최종모형에 남았다.

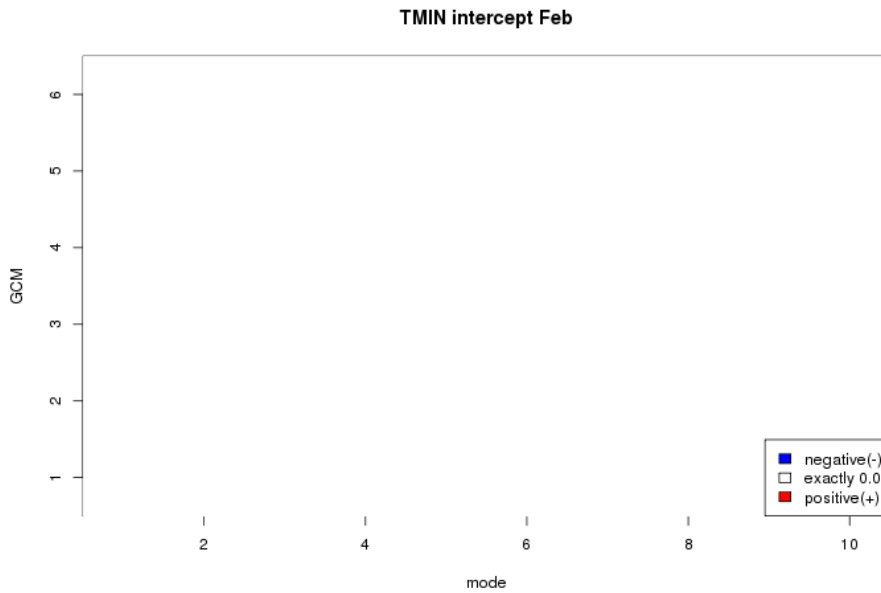


그림 24 2월 일최저기온 절편항의 첫 번째 EOF 계수에 대한 예측모형의 계수  $\beta_{l,j}$ ,  $l=1,\dots,6$ 은 GCM (y축),  $j=1,\dots,10$ 은 모드 (x축)를 뜻한다.  $\beta_0$ 의 추정값은 0.00. 예측장은 예측성이 지니고 있지 않다.

표 3 각 월별 예측장의 성능 평가. 12월에서는 Z500과 SLP가 예측성을 보이고 있으나, 2월은 예측력이 약하거나 없다.

| 월 (선행시간)     | predictand               | 예측인자 | MSE ratio | 상관계수 |
|--------------|--------------------------|------|-----------|------|
| 12월<br>(1달전) | 일최고기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.80      | 0.36 |
|              | 건조기지속일수 평균               | SLP  | 0.67      | 0.60 |
|              | 강수강도의 첫 번째 EOF 계수        | SLP  | 0.66      | 0.53 |
| 2월<br>(3달전)  | 일최고기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 0.89      | 0.18 |
|              | 일최저기온모형 절편항의 첫 번째 EOF 계수 | Z500 | 1.00      | 0.00 |

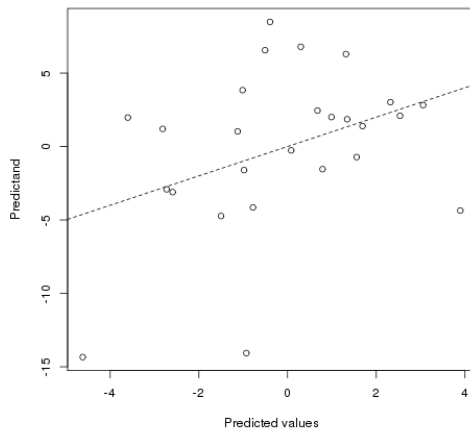


그림 25 12월 일최고기온 절편항 첫 번째 EOF 계수의 예측값과 predictand. 상관계수 0.36.

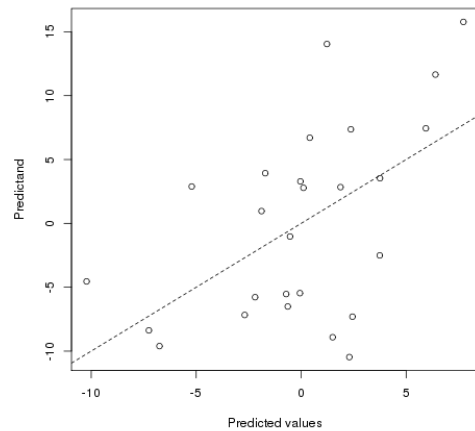


그림 26 12월 강수강도 첫 번째 EOF 계수의 예측값과 predictand. 상관계수 0.53.

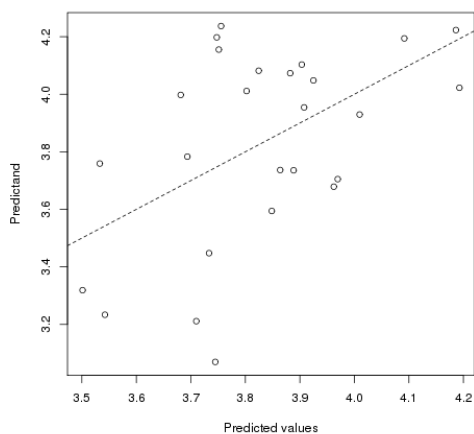


그림 27 12월 건조기지속일수평균의 예측값과 predictand. 상관계수 0.60.

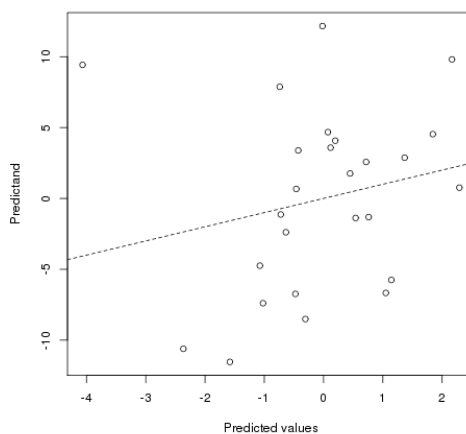


그림 28 2월 일최고기온 절편항 첫 번째 EOF 계수의 예측값과 predictand. 상관계수 0.18.

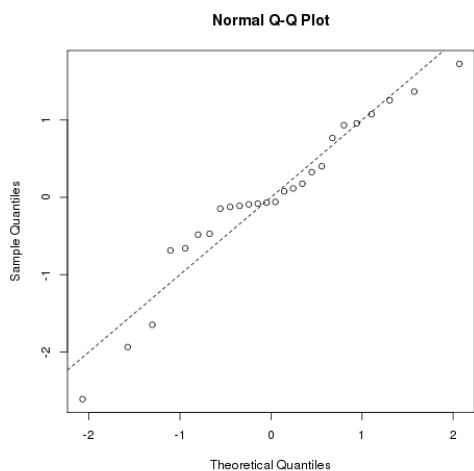


그림 29 12월 일최고기온 절편항 첫 번째 EOF 계수예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

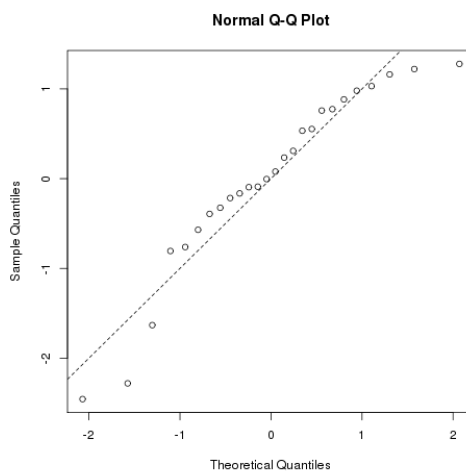


그림 30 12월 일최저기온 절편항 첫 번째 EOF 계수예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

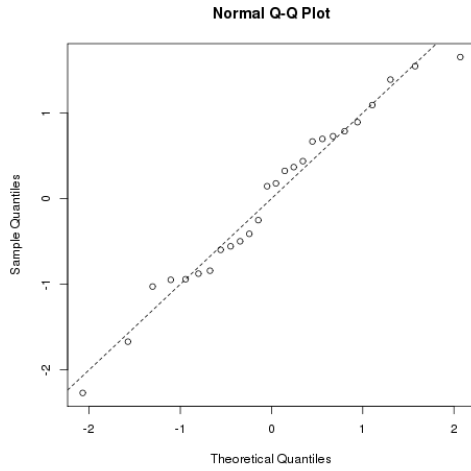


그림 31 12월 건조기지속일수평균예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

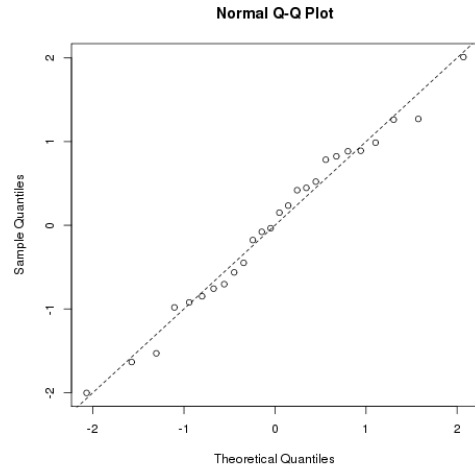


그림 32 12월 강수강도 첫 번째 EOF 계수예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

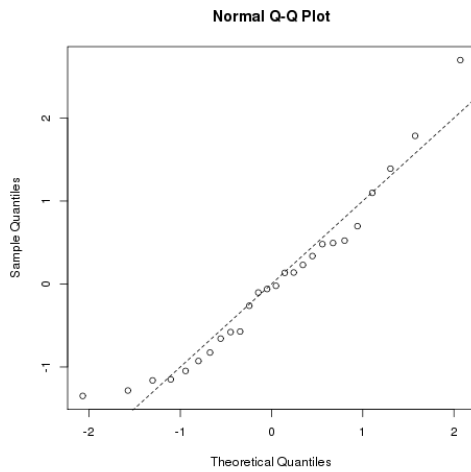


그림 33 2월 일최고기온 절편항 첫 번째 EOF 계수예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

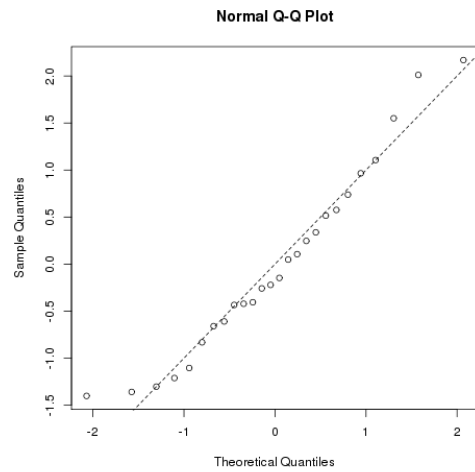


그림 34 2월 일최저기온 절편항 첫 번째 EOF 계수예측의 PRESS 잔차에 대한 정규성 확인. 정규분포의 형태에 가깝다.

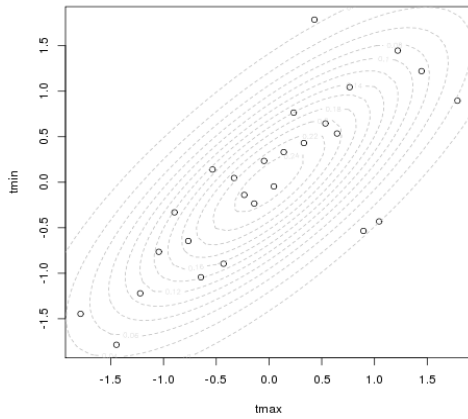


그림 35 12월 일최고/최저기온 첫 번째 EOF 계수예측 PRESS 잔차의 z-점수 산점도와 이변량 정규분포의 등고선그림. 대체로 이변량 정규분포가 적절한 것으로 보인다.

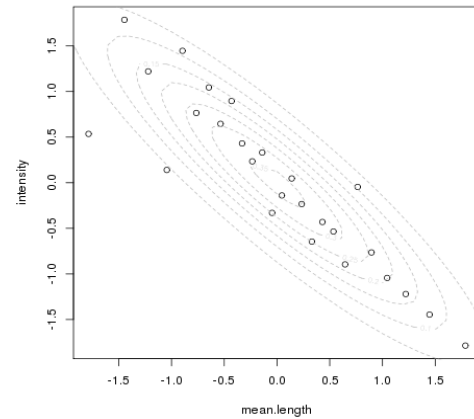


그림 36 12월 강수강도 첫 번째 EOF 계수와 건조기지속일수 평균예측 PRESS 잔차의 z-점수 산점도와 이변량 정규분포의 등고선그림. 대체로 이변량 정규분포가 적절한 것으로 보인다.

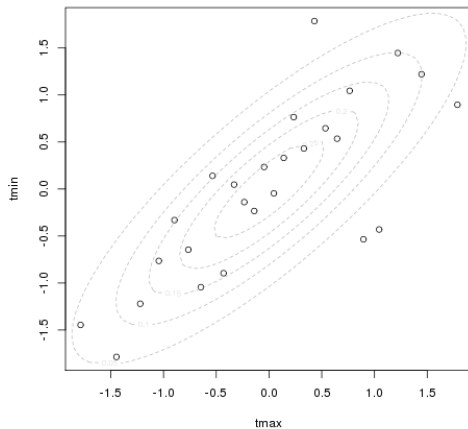


그림 37 2월 일최고/최저기온 첫 번째 EOF 계수예측 PRESS 잔차의 z-점수 산점도와 이변량 정규분포의 등고선그림. 대체로 이변량 정규분포가 적절한 것으로 보인다.

## 4. 결론 및 토론

본 연구에서 APCC 계절예측 상세화 기법을 겨울철 낙동강 유역을 시범대상으로 개발하고 시험하였다. 저해상도의 계절예측은 기후응용분야에서 활용성이 떨어지기 때문에 고해상도 자료 재생산을 위한 상세화 과정이 필요하다. 본 연구에서는 월단위 계절예측의 시간적 상세화를 위해 weather generator와 계절예측 간의 연결체계를 고려하였다. 연결체계는 weather generator의 통계모형을 경년변동성을 나타낼 수 있도록 모수화하는 것과 경년변동성을 표현하는 모수에 대해 APCC 계절예측기반의 예측모형을 구축하는 것으로 완성된다. 경년변동성을 표현하는 모수로 강수강도, 건조지속일 평균, 강수효과가 제거된 일최고/최저기온의 월평균을 선정하였다. 강수강도와 건조지속일 평균의 추정에 있어서, 자료부족으로 인한 불확실성이 불가피하는데, 본 연구에서는 교차검증을 통한 가중평균기법을 고려하였다. 한편, 예측모형의 수립 과정에서 과적합문제가 부각될 수 있다. 본 연구에서는, 이를 완화시키기 위해 예측장의 전체변동을 분해해 관측장과 공변동하는 부분을 얻고 이를 기반으로 한 회귀모형을 유도하는 것을 고려하였다. 낙동강 유역에 적용한 결과, 모수에 따라 정도차이는 있지만 경년변동이 나타나고, 기온과 관계한 모수는 Z500이 12월과 2월에, 강수에 관련해서는 SLP가 예측성을 보이고 있다. 본 연구에서는 11월에 생산된 향후 3개월 계절예측자료에서 Z500과 SLP에 중점을 두었다. 향후 연구에서 다른 계절에 대해서도 다양한 기후인자를 기반으로 기법을 적용할 계획이다.

APCC 계절예측 상세화의 성패는 예측모형의 성능과 계절내 변동(intra-seasonal variation) 표현의 적절성에 달려 있다. 예측모형은 계절기후를 특징짓는 모수의 값을 결정하기 때문에 그것의 성능은 상세화자료의 품질을 결정하는 핵심 요소이다. 또한 weather generator는 예측모형으로 특징지어진 계절기후의 계절내 변동을 표현하는데, 그 표현의 적절성이 또한 자료품질의 중요측도이다. 겨울철 낙동강 유역 연구에서 예측모형의 성능을 평가하였는데, 선행시간이 짧은 12월은 관측장에 준하는 성능을 보였으나 선행시간이 3달로 긴 2월은 그렇지 못하였다. 반면, 계절내 변동 표현의 적절성 평가는 본 연구에서 다루어지지 않았다. 계절내 변동에 관한 평가지표가 확립되지 않았고 제한된 관측자료 때문에 적절성 평가에 어려움이 있다.

본 연구에서 예측모형수립은 전적으로 통계적 방식을 통했다. 또한 예측인자 선정

에 있어 주관적 판단의 개입여지가 크다. 또한, 도출된 예측모형은 예측성이 있는 것으로 평가되었으나, 그것의 역학적 해석은 없는 상태이다. 한편, 기후자료가 고차원이고 축적시간이 짧아서, 통계적 방식은 과적합문제를 태생적으로 안고 있다. 이러한 문제를 근본적으로 해소하기 위해서는 역학적 이해와 지식을 기반으로 예측모형을 세워야 할 것이다. 현재 APCC의 타연구과제에서 통계-역학 계절예측모형 연구 중에 있으며, 향후 협력연구를 통해 상세화 예측모형의 고도화를 계획 중이다.

## REFERENCE

- 김무섭 2015: 통계적 상세화를 위한 Weather Generator 개발 APEC 기후센터 연구보고서 2015-04 ISBN: 979-11-5698-094-0
- Apipattanavis, S., Podestá, G., Rajagopalan, B., and Katz, R. W. 2007: A semiparametric multivariate and multisite weather generator. *Water Resources Research*, **43**(11). W11401, doi:10.1029/2006WR005714.
- Eum, H., Simonovic, S. P., and Y. Kim 2010: Climate change impact assessment using K-nearest neighbor weather generator: case study of the Nakdong river basin in Korea. *Journal of Hydrologic Engineer*. **15**(10) 772-785.
- Myers, R. H. 1990: *Classical and modern regression with applications*. PWS-KENT Publishing Compagny.
- Racsko, P., Szeidl, L., and Semenov, M. 1991: A serial approach to local stochastic weather models. *Ecological modelling*, **57**(1), 27-41.
- Rajagopalan, B. and Lall, U. 1999: A k-nearest neighbor simulator for daily precipitation and other variables. *Water Resour. Res.* **35**(10) 3089-3101
- Richardson, C. W. 1981: Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, **17**(1), 182-190.
- Semenov, M. A., Brooks, R. J., Barrow, E. M., and Richardson, C. W., 1998: Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates. *Climate research*, **10**(2), 95-107.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**(1) 267-288.
- Vinod, H. D. 1976: Canonical ridge and econometrics of joint production. *Journal of Econometrics* **6**, 129-137

## 부록

### RCCA의 예측장과 관측장 모드

RCCA를 통해 얻은 예측장과 관측장의 모드와 그것의 전체변동에서 차지하는 변동비율을 제시한다.

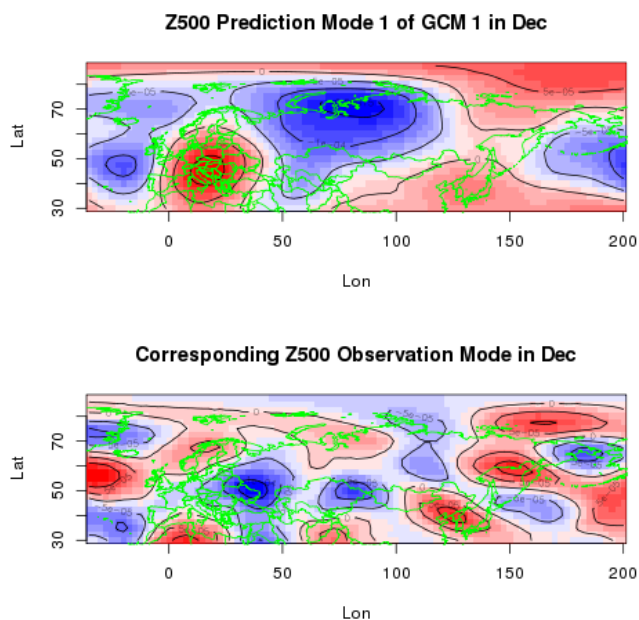


그림 38 12월 Z500 첫 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 5%, 관측장의 변동비율: 1.4%

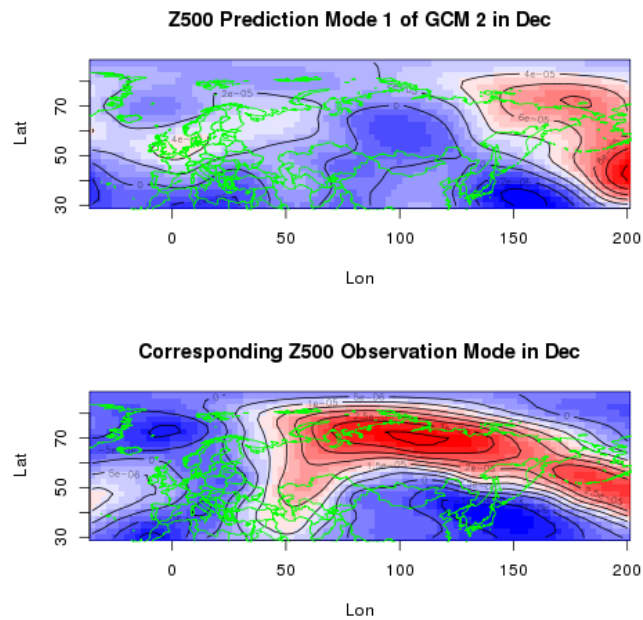


그림 39 12월 Z500 두 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 19%, 관측장의 변동비율: 12%

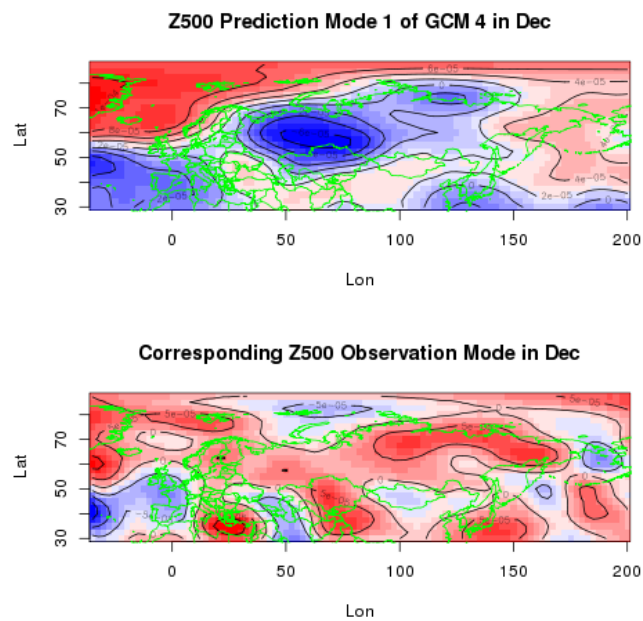


그림 40 12월 Z500 네 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 13%, 관측장의 변동비율: 2.2%

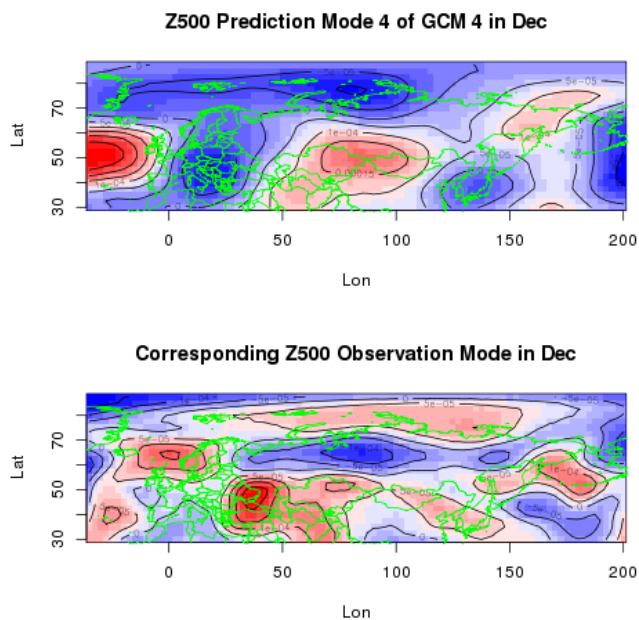


그림 41 12월 Z500 네 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 3%, 관측장의 변동비율: 0.8%

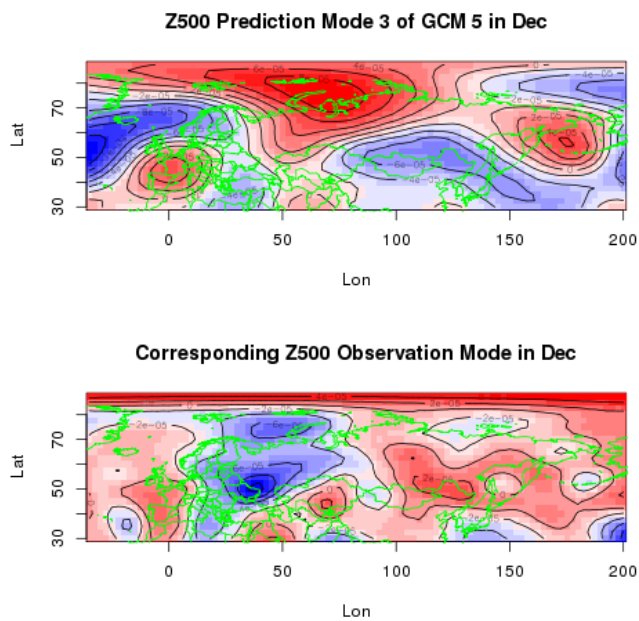


그림 42 12월 Z500 다섯 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 6%, 관측장의 변동비율: 3%

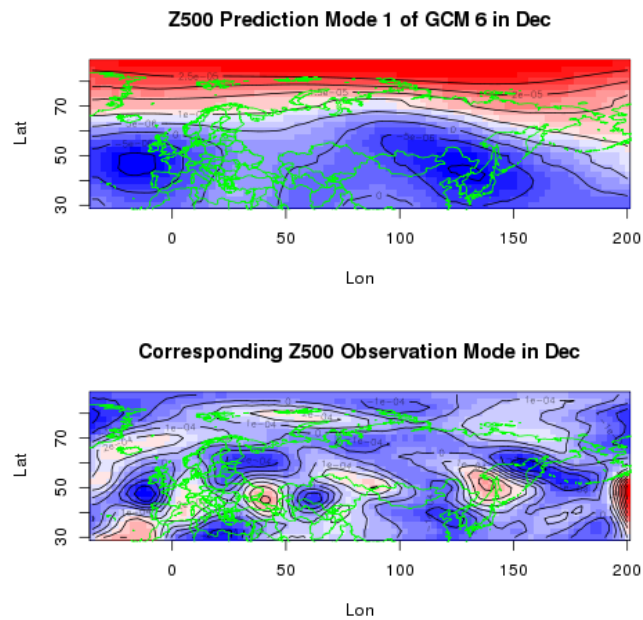


그림 43 12월 Z500 여섯 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 36%, 관측장의 변동비율: 0.1%

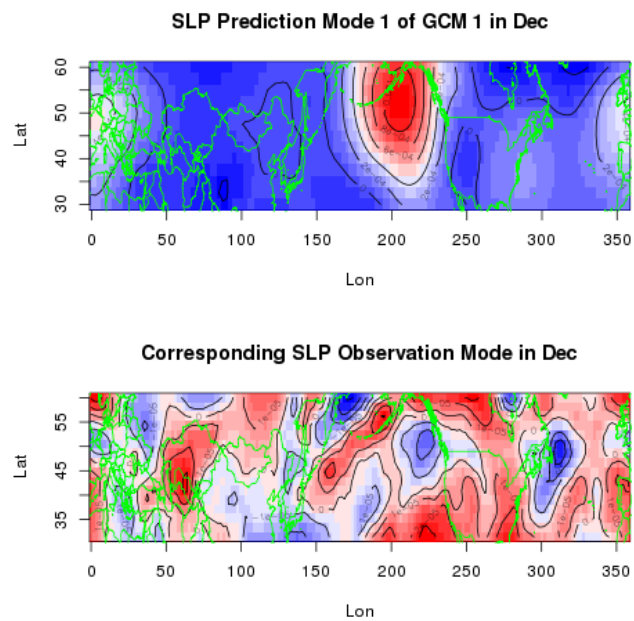


그림 44 12월 SLP 첫 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 54%, 관측장의 변동비율: 0.3%

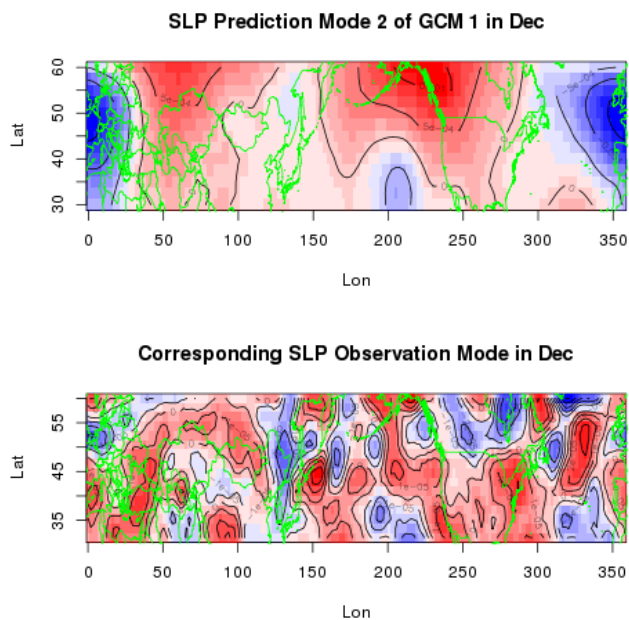


그림 45 12월 SLP 첫 번째 GCM의 예측장 모드2과 대응하는 관측장. 예측장 변동비율: 11%, 관측장의 변동비율: 0.2%

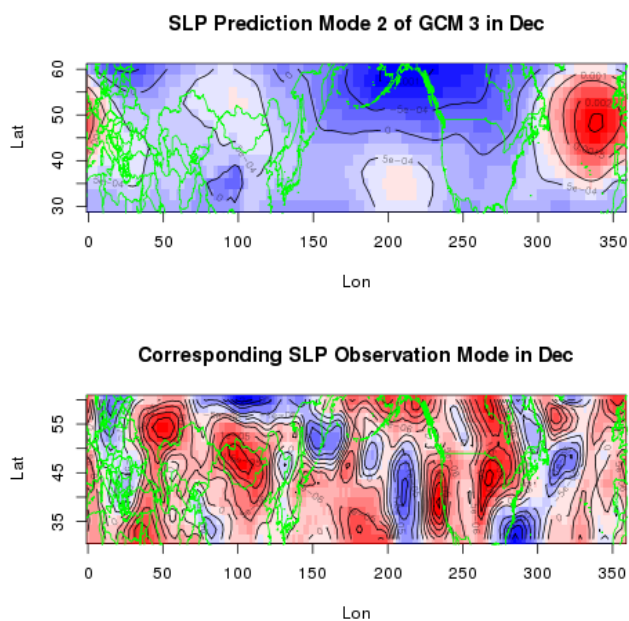


그림 46 12월 SLP 세 번째 GCM의 예측장 모드2과 대응하는 관측장. 예측장 변동비율: 14%, 관측장의 변동비율: 0.3%

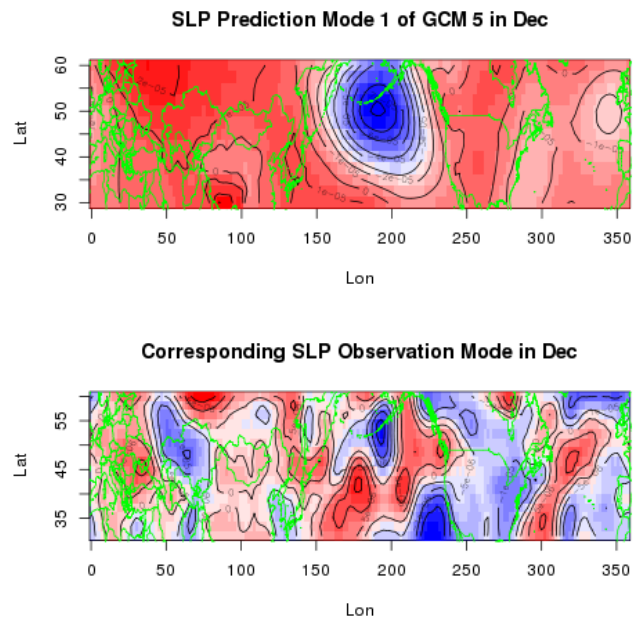


그림 47 12월 SLP 다섯 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 44%, 관측장의 변동비율: 0.8%

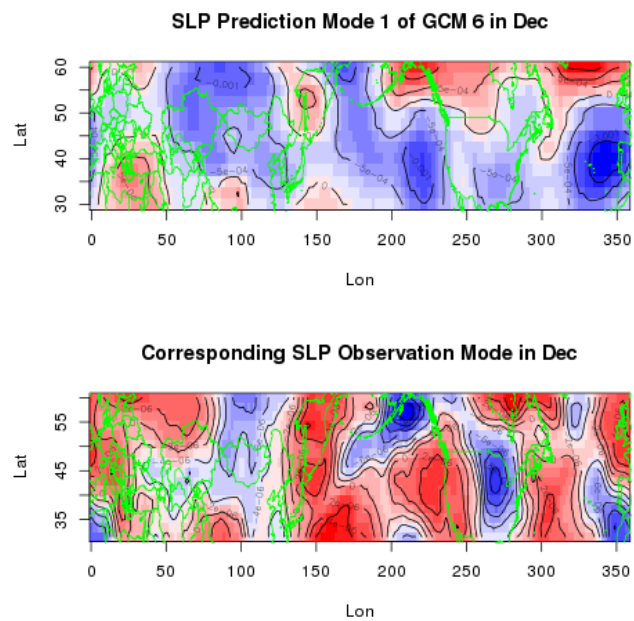


그림 48 12월 SLP 여섯 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 6%, 관측장의 변동비율: 0.2%

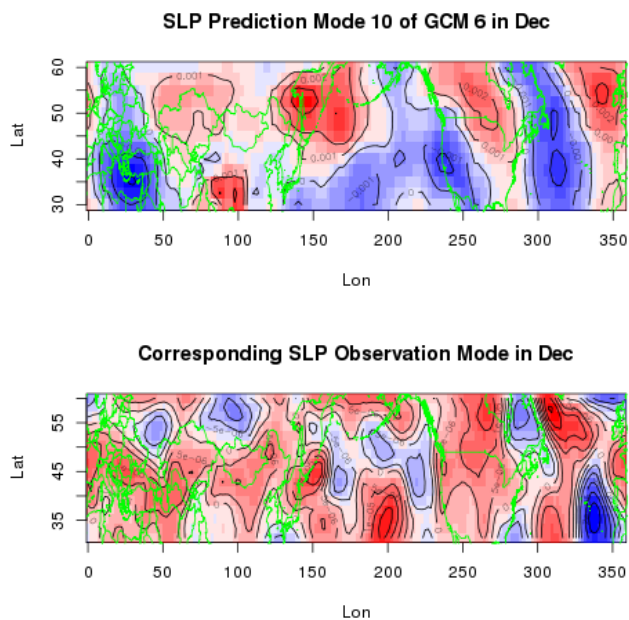


그림 49 12월 SLP 여섯 번째 GCM의 예측장 모드10과 대응하는 관측장. 예측장 변동비율: 1%, 관측장의 변동비율: 0.5%

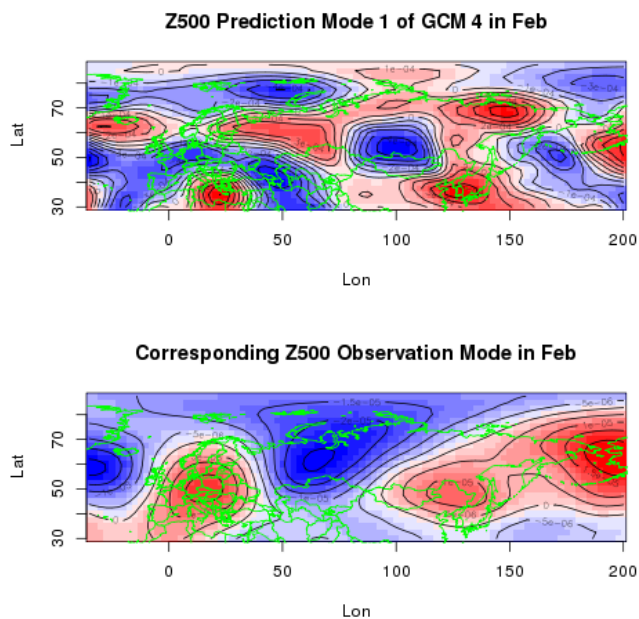


그림 50 2월 Z500 네 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 0.3%, 관측장의 변동비율: 24%

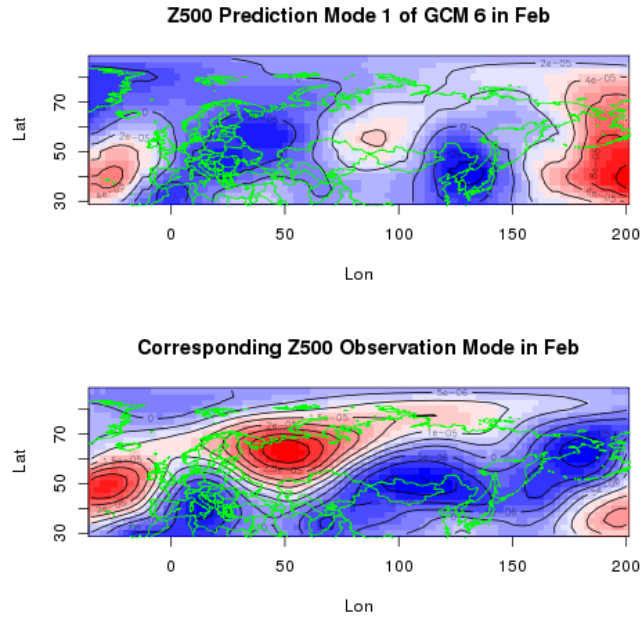


그림 51 2월 Z500 여섯 번째 GCM의 예측장 모드1과 대응하는 관측장. 예측장 변동비율: 9%, 관측장의 변동비율: 15%

### 교차검증을 통한 가중평균 계산

앞서 강수강도와 건조기지속일수의 평균의 추정하기 위해서 교차검증을 통한 가중평균을 언급하였다. 여기서 이에 대한 자세한 내용을 설명한다. 두 추정은 동일한 틀의 문제로 볼 수 있다. 우선 관측값  $X_1, \dots, X_n$ , ( $n > 1$ )이 있고 관측값의 모평균 (population mean)에 대한 참고값  $\mu_0$ 가 주어진 상황이다. 문제에서 참고값은 일반화선형모형을 통해 얻어진 적합값이다. 우리는 표본의 크기  $n$ 이 충분히 크지 않은 상태라서 참고값을 이용해 모평균 추정의 불확실성을 줄이고자 한다. 모평균추정량으로 표본평균  $\bar{X}$ 와  $\mu_0$ 의 가중평균  $w\bar{X} + (1-w)\mu_0$ ,  $0 < w < 1$ 를 고려하는데, 최적의 가중치  $w$ 를 구하기 위해 교차검증을 통한다.  $X_i$ 를 제외한 채 계산한 표본평균을  $\bar{X}_{-i}$ 라고 할 때,

$$\frac{1}{n} \sum_{i=1}^n \{w\bar{X}_{-i} + (1-w)\mu_0 - X_i\}^2$$

는 교차검증으로 얻은 평균제곱예측오차이다. 평균제곱예측오차를 최소로 하는  $w$ 의 값을 구하면,

$$w = \max \left\{ \frac{S_0^2 - S^2}{\frac{1}{n-1}S^2 + S_0^2 - S^2}, 0 \right\}, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad S_0^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_0)^2$$

이다. 특히, 표본의 크기  $n$ 이 커지거나  $\mu_0$ 가 참값에서 멀어질수록  $w \rightarrow 1$ , 즉, 추정량이 표본평균으로 수렴함을 알 수 있다.

가중평균의 성능을 실증적으로 검증하기 위해서 간단한 모의실험을 수행하였다. 평균이 1이고 형태모수가 0.6인 감마분포로부터 크기 5의 표본을 얻고, 참고값에 따른 가중평균의 MSE를 표본평균의 MSE와 비교하였다. 모의실험 10,000번을 통해 얻어진 MSE의 비(MSE ratio)가 그림 52에서 주어진다. 비의 값이 1보다 작을수록 가중평균이 나은 성능을 나타냄을 의미하는데, 참고값이 참값인 1의 근방에서 있을 때 가중평균이 좋은 성능을 보임을 알 수 있다.

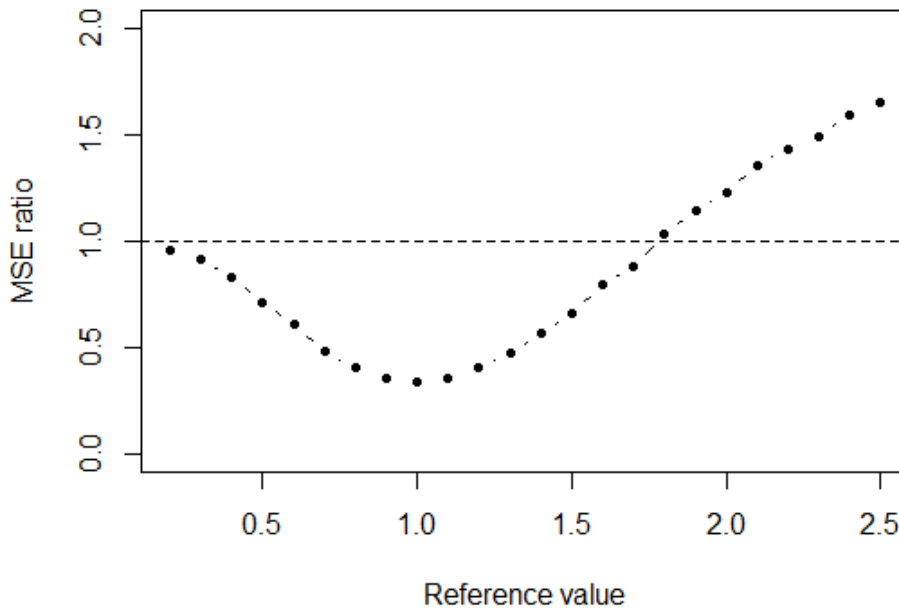


그림 52 가중평균의 성능. 참값의 근방에서 좋은 성능을 보인다.

### 관측장자료에 대한 능형회귀방법

회귀모형 (4)에 대한 능형회귀방법의 복잡성 조절방식의 원리를 살펴본다. 여기서는 능형회귀모형의 기본설정에서처럼, 식 (4)의  $Y_i$ 와  $X_i(s_j)$ 가 모두 평균 0로 조정되었다고 전제한다. 그래서 절편항이 0으로 처리될 수 있어 식은

$$Y_i = \sum_{j=1}^p \beta_j X_i(s_j) + \epsilon_i, \quad i = 1, \dots, n,$$

가 된다. 능형회귀방법의 예측은 다음과 같이 주어진다:  $X^*$ 가 새로운 예측인자일 때,

$$\sum_{j=1}^n \alpha_j \langle X^*, v_j \rangle \frac{n\lambda_j}{n\lambda_j + \delta}, \quad \alpha_j = \frac{1}{n\lambda_j} \sum_{i=1}^n Y_i \langle v_j, X_i \rangle,$$

인데, 여기서,  $\delta > 0$ 은 능형회귀방법의 조절모수이고,  $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ 과  $v_1, v_2, \dots, v_n$ 은  $X_1, \dots, X_n$ 을 이용한 관측장 EOF분석에서 얻어진 고유값과 EOF모드이다. 한편, 회귀모형 (8)은

$$Y_i = \sum_{j=1}^p \beta_j X_i(s_j) + \epsilon_i = \langle \beta, X_i \rangle + \epsilon_i = \sum_{j=1}^n \langle \beta, v_j \rangle \langle v_j, X_i \rangle + \epsilon_i, \quad \beta = (\beta_1, \dots, \beta_p),$$

라는 관점에서 그것의 절단형태인 회귀모형

$$Y_i = \sum_{j=1}^k \gamma_j \langle v_j, X_i \rangle + \epsilon_i, \quad k < n,$$

을 고려할 때, 최소제곱법에 의한 예측은

$$\sum_{j=1}^k \alpha_j \langle X^*, v_j \rangle$$

으로 주어진다. 두 예측 (9)과 (10)을 통합하여

$$\sum_{j=1}^n \alpha_j \langle X^*, v_j \rangle w_j$$

의 형태로 쓸 수 있으며, (9)와 (10)는 각각

$$w_j = \frac{n\lambda_j}{n\lambda_j + \delta}, \quad w_j = \begin{cases} 1, & j \leq k; \\ 0, & j > k, \end{cases}$$

인 경우이다. 여기서  $\delta \rightarrow \infty$  일 때  $n\lambda_j/(n\lambda_j + \delta) \rightarrow 0$ 이지만,  $\lambda_j$ 가 클수록 상대적으로 천천히 0으로 수렴한다는 점을 주목할 필요가 있다. 12월 Z500 관측장 (1983년에서 2013년)에 대해서  $n\lambda_j/(n\lambda_j + \delta)$ 의 값을 계산해 보았다. 상관행렬의 고유값이 기하급수적으로 감소하고,  $n\lambda_j/(n\lambda_j + \delta)$ 의 값도 감소속도가 확연히 다를 수 있다. 특히, 모드  $j$ 가 증가할수록 감소속도가 빠름을 확인할 수 있다. 그러므로 두 방법 모두가 주요 EOF 모드에 가중치를 주어 자료의 복잡성을 조절하는 방식인데, 능형회귀방법은 연속적이고 정량적인 조절을 가능하게 한다.

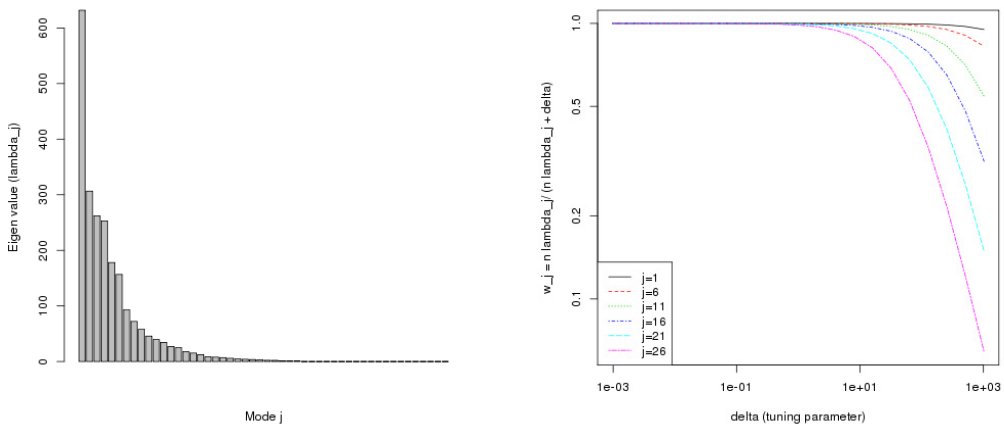


그림 53 12월 Z500의 상관행렬의 고유값의 분포와  $\delta$ 값의 증가에 따른  $n\lambda_j/(n\lambda_j + \delta)$ 의 변화. 고유값이 클수록 감소속도가 확연히 다를 수 있다.  $j=1$ 의 경우는 거의 감소가 없는 반면에  $j=26$ 은 거의 0으로 떨어졌다.

연구보고서 2016-08

---

Weather Generator를 이용한 APCC 계절예측의  
통계적 상세화 연구  
-겨울철 낙동강 유역 중심으로-

Statistical Downscaling of APCC Seasonal Prediction using Weather  
Generator with Application to Nakdong River Basin in Boreal Winter

김무섭



APEC 기후센터

48058 부산광역시 해운대구 센텀7로12

Tel: 051-745-3900 Fax: 051-745-3949

[www.apcc21.org](http://www.apcc21.org)



ISBN 979-11-5698-159-6