

## 서포트 벡터 머신을 이용한 NCAM-LAMP 고해상도 중기에측시스템 지점 시계열 자료의 통계적 보정

권수영<sup>1</sup>, 이승재<sup>1\*</sup>, 김만일<sup>2</sup>

<sup>1</sup>국가농림기상센터, <sup>2</sup>산림조합중앙회

(2021년 11월 10일 접수; 2021년 12월 8일 수정; 2021년 12월 14일 수락)

### A Statistical Correction of Point Time Series Data of the NCAM-LAMP Medium-range Prediction System Using Support Vector Machine

Su-Young Kwon<sup>1</sup>, Seung-Jae Lee<sup>1\*</sup>, Man-Il Kim<sup>2</sup>

<sup>1</sup>National Center for AgroMeteorology, Seoul 08826, Republic of Korea

<sup>2</sup>National Forestry Cooperative Federation, Seoul 05601, Republic of Korea

(Received November 10, 2021; Revised December 8, 2021; Accepted December 14, 2021)

#### ABSTRACT

Recently, an R-based point time series data validation system has been established for the statistical post processing and improvement of the National Center for AgroMeteorology-Land Atmosphere Modeling Package (NCAM-LAMP) medium-range prediction data. The time series verification system was used to compare the NCAM-LAMP with the AWS observations and GDAPS medium-range prediction model data operated by Korea Meteorological Administration. For this comparison, the model latitude and longitude data closest to the observation station were extracted and a total of nine points were selected. For each point, the characteristics of the model prediction error were obtained by comparing the daily average of the previous prediction data of air temperature, wind speed, and hourly precipitation, and then we tried to improve the next prediction data using Support Vector Machine(SVM) method. For three months from August to October 2017, the SVM method was used to calibrate the predicted time series data for each run. It was found that The SVM-based correction was promising and encouraging for wind speed and precipitation variables than for temperature variable. The correction effect was small in August but considerably increased in September and October. These results indicate that the SVM method can contribute to mitigate the gradual degradation of medium-range predictability as the model boundary data flows into the model interior.

**Key words:** Support Vector Machine, LAMP, Statistical correction



\* Corresponding Author : Seung-Jae Lee  
(sjlee@ncam.kr)

## I. 서 론

최근 국가농림기상센터에서는 NCAM-LAMP 중기 예측자료(Lee *et al.*, 2016)의 통계적 후처리와 개선을 위하여 R 기반의 지점 시계열 자료 검증 체계를 구축하였다. 이 시계열 검증체계를 이용하여 기상청 AWS 관측 자료와 NCAM-LAMP, KMA GDAPS 중기예측 모델 자료를 비교하여, 모델의 예측 성능을 평가하고 있다. 관측 지점에 가장 근접한 모델 위도 및 경도 자료를 추출하여 총 9개 지점에 대해 NCAM-LAMP, GDAPS 모델의 기온, 강수량, 풍속 일평균 예측 자료를 관측과 비교한 결과, 모델들은 풍속의 과대예측 경향을 뚜렷이 보였으며, 기온과 강수의 경우에는 두 모델의 예측 오차가 월별 및 변수별로 다르게 나타남을 보였다(Kwon and Lee, 2017). 이러한 모델의 예측 오차는 계통 오차와 무작위 오차의 합으로 볼 수 있으며, 여러 연구에서는 통계적 기법을 이용하여 수치모델이 가지고 있는 예측 오차를 줄이고자 노력하고 있다.

모델 오차를 줄이기 위해 일반적으로 쓰이는 통계적 후처리 방법은 모델 결과와 관측의 장기적인 상관관계를 통해 편차를 제거하는 MOS (Model Output Statistics)이다. MOS는 수치예보 모델의 계통오차를 설명할 뿐만 아니라, 기상 인자들과 모델 변수들 사이의 통계적 관계를 결정한다. 즉, 기존 모델에 의한 예측값에 보정값을 합하여 최종 예측값을 구하는 방식이다. MOS에서 사용하는 통계적 방법은 다중선형회귀(Sokol *et al.*, 2003; Ranaboldo *et al.*, 2011), 다항식(Antolik, 2005; Jacobsone *et al.*, 2009) 또는 로지스틱회귀(Sokol *et al.*, 2003; Wilks, 2009) 등이 가능하다.

근래에는 인공지능 기법 중 SVM (Support Vector Machine)에 기반하여 MOS를 연구하는 시도들이 주목을 끌고 있다. Wu *et al.*(2008)은 SVM과 회귀식을 결합한 수위예측 모형을 개발하여 수위 예측에 적용하였다. Moon *et al.*(2014)는 상세화된 공간해상도의 강우를 예측하기 위해 비선형 예측모형인 SVM을 활용하여, 상세화 이전의 25km 격자 강우에 비해 상관계수 향상과 평방제곱근오차(RMSE), Index of Agreement (IOA) 및 편차(bias)의 개선을 보였으며 전반적으로 높은 정확도를 지남을 보였다. Patil *et al.*(2015)는 풍속 일일 예측에 SVM을 사용한 결과, 평균절대오차(Mean Absolute Percentage Error, MAPE)는 약 7%이며, 상관계수는 1에 가까운 우수한 결과를 얻었다. Yoon *et al.*(2016)은 지하수위 변동패턴에 대한 오차지표 분석

결과 인공지능망(Artificial Neural Network, ANN) 모델이 SVM 모델보다 다소 높은 직접예측능력을 보여 주지만 반복예측모델에 대한 안정성은 SVM이 높음을 보였다. Bae *et al.*(2017)은 기상 입력변수와 관련하여 기후 예측 오차에 따른 태양광 출력 예측 정확도를 분석한 결과 SVM이 인공지능망 보다 높은 정확도를 보였다. Mokhtarzad *et al.*(2017)은 가뭄 예측에 SVM이 ANN 및 ANFIS (Adaptive Neuro-Fuzzy Inference System)보다 더 정확함을 보였다.

국내에서는 SVM 방법을 기상예측 자료에 적용한 연구는 드문 실정이며, 그 효용성 또한 잘 알려져 있지 않은 상황이다. 본 연구에서는 SVM의 유용성이 입증된 연구들을 바탕으로 하여, SVM 방식을 NCAM-LAMP 중기 기상예측 자료에 적용하고, 그 효용성을 살펴보고자 한다.

## II. 자료 및 방법

### 2.1. 예측 자료 및 관측 자료

#### 2.1.1. NCAM-LAMP 중기예측 자료

국가농림기상센터(NCAM)에서는 수요자 맞춤형 영농 영림을 지원하기 위하여 전용 수치모델링시스템인 지면대기모델링패키지(Land-Atmosphere Modeling Package, LAMP)를 구축하였다(Lee *et al.*, 2016). LAMP 패키지는 WRF/Noah-MP 결합시스템과 오픈라인 1차원 Noah-MP 지면모델로 구성되어 있으며, 전자는 7일 이상의 중기예측 자료를 1km 내외의 고해상도로 생산하는 일을 담당하고, 후자는 대표적인 농림생태계에 대하여 1년 모의 자료를 15분 간격으로 생산하는 일을 담당한다. LAMP의 예측자료는 현재 주 2회 정기적으로 생산되고 있으며 시간단위 및 일단위로 다양한 농림기상 변수들에 대해 데이터베이스화 되고 있다(So *et al.*, 2020; Lee *et al.*, 2020). 본 연구에서는 비교적 대표성 있는 농림지 기상 관측자료가 많고 LAMP 고해상도 운영 이력이 긴 경기도 지역에 대하여 810m 공간해상도 및 12일 예측 기간으로 생산된 자료를 이용하였다.

#### 2.1.2. 전지구예보모델 중기예측 자료

기상청에서 운영하는 수치예보 현업모델 중 전지구예보모델(Global Data Assimilation Prediction System, GDAPS)은 수평적으로 1024×769개의 격자 및 약 25 km의 수평 분해능을 가지고 있으며, 연직적으로는 약

80km 고도까지 70층으로 되어 있다. GDAPS는 일 2회 00, 12UTC에 288시간 예측과, 일 2회 06, 18UTC에 87시간 예측을 수행한다. 또한 4차원 변분법 기반의 자료동화가 접합되어 있기 때문에 6시간 주기 순환예측에 필요한 각종 배경장을 생산하기 위한 일 4회 15시간 예측을 수행한다. 일 1회 06UTC에는 해수면온도, 해빙자료 및 동서평균 오존량을 갱신하기 위한 배경장 갱신과정이 별도로 수행된다. GDAPS 자료는 기상청에서 국가농림기상센터 자료획득처리시스템 서버로 실시간 전송되고 있으며, 이 자료에 대한 최신 업데이트 내용은 Hwang *et al.*(2020)에 기술되어 있다.

**2.1.3. 자동기상관측 자료**

자동기상관측장비 AWS (Automated Weather

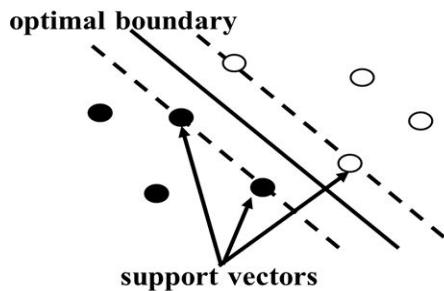
Station)는 주로 무인관측소에 설치하며, 기상 측정, 연산, 저장, 표출 등 모든 과정이 실시간 자동으로 처리된다. 기본 관측 요소는 기온, 풍향, 풍속, 강수량, 강수유무이다. 본 연구에서는 LAMP 모델의 중기예측 성능을 경기도 지역에서 지점별로 검증하기 위해 모델 최내부 도메인에 위치하는 기상청 9개의 관측소를 사용하였으며, 관측소별 정보는 Table 1과 같다.

**2.2. 통계적 후처리 방법**

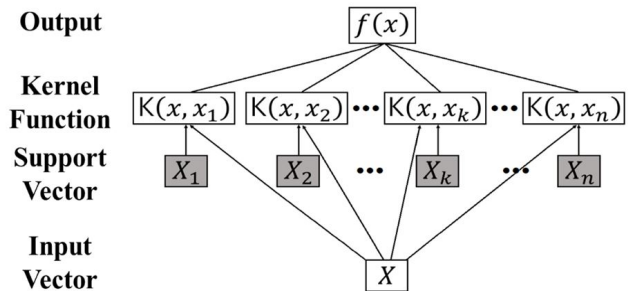
본 연구에서는 SVM을 이용하여 LAMP가 가지고 있는 예측 오차를 줄이고자 하였다. SVM은 훈련 또는 학습 자료(training data)를 통해 자료를 이분법으로 분리하는 기법으로서, 자료를 커널 매핑(kernel mapping)을 통해 고차원 공간으로 사상시킨 후, 최적화 기술을

**Table 1.** Information of the nine KMA observation stations

No.	Station name (Station number)	Address	Altitude (m)	Latitude (degree, N)	Longitude (degree, E)
1	Sinhyeon-dong (367)	Po-dong Siheung-si, Gyeonggi-do	26	37.4061	126.7844
2	Hyeondeok-myeon (358)	Ingwang-ri, Hyeondeok-myeon, Pyeongtaek-si, Gyeonggi-do	35	36.9672	126.9214
3	Gangbuk(424)	Suyu-dong, Gangbuk-gu, Seoul	56	37.6397	127.0257
4	Gangnam(400)	Samseong-dong, Gangnam-gu, Seoul	59	37.5134	127.0470
5	Hanam(444)	Sinjang-dong, Hanam-si, Gyeonggi-do	37	37.5359	127.2156
6	Seolbong(440)	Jungni-dong, Icheon-si, Gyeonggi-do	168	37.2750	127.4424
7	Okcheon (449)	Okcheon-ri, Okcheon-myeon, Yangpyeong-gun, Gyeonggi-do	44	37.5073	127.4522
8	Yongmunsan (326)	Sinjeom-ri, Yongmun-myeon, Yangpyeong-gun, Gyeonggi-do	197	37.5455	127.6112
9	Yeoju (548)	jeombong-ri, Yeoju-eup, Yeoju-gun, Gyeonggi-do	144	37.2688	127.6395



**Fig. 1.** Optimal boundary by SVM. Adapted from Asano(2004).



**Fig. 2.** Structure of SVM. Adapted from Yoon *et al.*(2016)

적용한 선형 회귀를 통해 최적의 초평면을 찾아 자료를 분류하는 방법이다(Fig. 1). SVM의 경우 모델에 적합한 것으로 판정된 자료가 벡터의 형태로 모델 입력층을 구성하는데 이를 지지벡터(support vector)라고 한다. SVM에서는 새로운 입력 벡터에 대해 지지벡터들과의 연산을 통해 출력값을 예측하도록 구성한다(Fig. 2).

SVM 모델 구성의 최종 목적은 주어진 자료에 대해 가장 적합한 지지벡터를 선정하고, 각 지지벡터들의 연결강도 및 편중값을 결정하는 것이다. 본 논문에서는 불용해가 나오는 것을 방지하고자  $\epsilon$ -SVM을 사용하였으며 최적화 식은 다음과 같다.

$$f(x) = \omega \cdot \Phi(X) + b$$

$$\begin{cases} y_i - \omega^T \Phi(X_i) - b \leq \epsilon + \xi_i \\ \omega^T \Phi(X_i) - b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

여기서  $x$ 는 입력 회귀벡터,  $y$ 는 출력벡터,  $\omega$ 는 계수벡터,  $b$ 는 편중값,  $\Phi$ 는 비선형 전이함수,  $\epsilon$ 는 오차 허용율,  $\xi_i$   $\xi_i^*$ 는 슬랙 변수(slack variable)를 나타낸다. 우선 선형함수  $f(x)$ 를 찾아서  $f(x)$ 와 모든 학습 자료의 최대편차가  $\epsilon$  이내가 되면서 동시에  $\omega$ (연결강도, 계수벡터)를 최소화하는 것이다. 위의 식에서  $\omega$ ,  $b$ ,  $\xi_i$ ,  $\xi_i^*$ 를 최소화하여 다음과 같은 식을 도출한다.

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

상수  $C$ 는  $\epsilon$ 보다 큰 오차를 얼마만큼 허용할 것인가와 함수  $f(x)$ 의 편평함(flatness) 사이의 타협점을 결정하는 것으로 비용계수(cost coefficient)라고 부른다. 즉 비용계수  $C$ 는 과적합을 막는 정도를 지정하는 파라미

터로 Fig. 1에서 자료를 잘못 분류하는 선을 갖게 될 경우 얼마만큼의 비용을 지불할 것인지를 지정한다. 비용계수  $C$ 가 상대적으로 큰 경우에는 학습 자료의 오차를 가능한 작게 만드는 것이고, 반대로  $C$ 가 상대적으로 작은 경우에는  $\omega$ 의 노름(norm)을 작게 만들기 때문에 함수  $f(x)$ 를 가능한 편평하게 만드는 것이다. SVM은 자료를 한 가운데로 얼마나 잘 나누는지, 잘못 구분한 점으로 인한 비용의 합을 최소화하는 선을 찾는다. 결과적으로 SVM은 비용을 사용해 과적합 정도를 조절하게 된다.

두 가지 자료 클래스를 분리하는 가장 쉬운 방법은 2차원 자료의 경우에는 선을, 3차원 자료의 경우에는 평면이다. 하지만 모든 자료를 항상 초평면으로 나눌 수 있는 건 아니며, 이러한 클래스를 분리하려면 비선형 영역이 필요하다. SVM은 이 문제를 커널 트릭이라는 기법으로 해결한다. 커널 트릭의 기본 아이디어는 주어진 자료를 적절한 고차원으로 옮긴 뒤 변환된 차원에서 SVM을 사용해 초평면을 찾는 것이다(Fig. 3). 본 논문에서는 이러한 커널 매핑에 가장 성능이 우수하다고 알려진 방사기저함수(radial basis function)을 사용하였다.

$$K(x,y) = \exp(-\gamma \|x - y\|^2)$$

여기서  $x$ 는 지지벡터,  $y$ 는 입력벡터,  $\gamma$ 는 방사기저함수 파라미터를 나타낸다.

또한 본 논문에서는 R 패키지를 사용하여 SVM함수를 튜닝(tuning)하였다. 튜닝은 성능을 최적화하거나 편향과 분산 사이의 균형을 맞출 때 알고리즘을 조절하기 위해 사용된다. 튜닝 단계에서  $\gamma$ 는 [0.1, 1, 10] 범위에서, 비용계수  $C$ 는 [0.1, 1, 10] 범위에서 모델 구성 단계에서의 오차를 최소화하는 파라미터를 선정했다.

예측의 정확도를 수치적으로 평가·비교하고자, 아

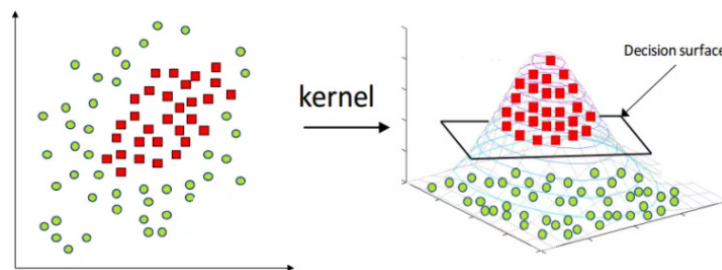


Fig. 3. Kernel Trick. Adapted from A. Statnikov et al.(2011).

**Table 2.** Time period in UTC (Universal Time Coordinated) for training data and test data

Month	Training data	Test data
July	12UTC 16 JUL 2017~28 JUL 2017	12UTC 30 JUL 2017~11 AUG 2017
August	12UTC 03 AUG 2017~15 AUG 2017	12UTC 17 AUG 2017~29 AUG 2017
September	00UTC 03 SEP 2017~15 SEP 2017	00UTC 17 SEP 2017~29 SEP 2017
October	00UTC 01 OCT 2017~13 OCT 2017	00UTC 15 OCT 2017~27 OCT 2017
November	00UTC 01 NOV 2017~13 NOV 2017	00UTC 15 NOV 2017~27 NOV 2017

래의 편차(Bias), 평균제곱근오차(RMSE) 통계량을 계산하였다.

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2}$$

여기서, N은 자료의 개수, F는 예측값, O는 관측값을 의미한다. 본 연구에 사용된 수치모델 자료의 경우, 1회의 LAMP 중기예측에서 변수별로 N=288개의 시간 단위 자료가 생성되며, 이는 통상 6시간 또는 3시간 간격의 자료를 검증이나 훈련에 이용하는데 비해 시간적으로 많은 자료를 이용하는 것이므로, 표본의 크기가 충분히 크다고 가정하였다.

### III. 결과 및 고찰

1시간 간격의 LAMP 모델 예측 값과 해당 시각의 관측 자료에서 기온(°C), 10m 풍속(m/s), 시간 강수량(mm) 변수에 대해 14일 간격으로 SVM 기법을 적용하였다. 1시간 간격의 관측 자료와 LAMP 자료를 훈련 세트(training set)로, 14일 후, 3시간 간격의 관측 자료와 GDAPS 자료를 시험 세트(test set)로 하여 교차검증법(cross-validation)을 수행하였다. 교차검증법은 주어진 데이터를 일부 나누어서 한 쪽 데이터로는 학습을 시켜 모델을 만들고, 나머지 학습시키지 않은 데이터로 그 모델에 대해 검증하는 방법을 말한다. 훈련 세트에서 관측 자료 및 LAMP 자료에 SVM 기법을 적용한 예측 자료와 LAMP 자료를 산술적 평균 방식의 편차 보정(bias correction)을 하였고, 최종 편차 보정된 LAMP\_SVM 자료를 시험 세트와 비교하였다. 1시간 간격의 훈련 세트와 3시간 간격의 시험 세트를 비교하기 위해 훈련 세트의 LAMP\_SVM 자료를 3시간 간격으로 평균을 낸 후, 시험 세트와 비교하였다. 훈련 자료

의 시간 간격과 시험 자료의 시간 간격은 Table 2와 같다. 7월부터 11월까지 5개월에 대하여 기온, 풍속, 강수량의 GDAPS 예측값과, SVM을 이용해 보정된 LAMP 예측값(LAMP\_SVM)을, 관측값과 비교하여 산출한 편차와 평방제곱근오차는 Table 3~Table 5에 정리되어 있다.

기온의 경우 GDAPS가 LAMP\_SVM보다 여전히 관측치에 가까운 값을 보였으며, SVM을 이용한 통계적 예측 보정의 효과가 GDAPS 기온 예측의 성능을 앞지르지 못하였다(Table 3). 이에 대한 한 원인은 GDAPS는 1.5m 높이에서, LAMP는 2m 높이에서 기온을 각각 예측하고 있으므로, 두 모델의 지상 기온 예측 높이의 차이로 여겨진다. 기상청의 지상기상관측지침(2016.7.25.개정)은 “기온은 대기의 온도를 말하지만, 같은 장소에서 같은 시각에도 높이에 따라서 기온이 다르기 때문에 일반적으로 기상관서에서 발표하는 기온은 지면 위(또는 설면 위) 1.2~1.5m 정도 높이의 온도를 말한다”고 명시하고 있다. 따라서 1.5m 높이에서 기온을 관측하는 AWS에 대하여, GDAPS가 LAMP보다 기온 예측에 유리한 측면이 있다고 판단되며, 추후 관측높이 기준에 대한 보완점을 찾을 필요가 있다고 판단된다.

그러나, 풍속의 경우 LAMP\_SVM은 GDAPS의 예측과 비슷하거나 관측치에 더 가까운 성능을 보여주었다(Table 4). 7~8월에는 현업 GDAPS의 예측력에 미치지 못했으나 9, 10, 11월로 가면서 LAMP\_SVM 예측력이 GDAPS를 뚜렷이 능가하여, SVM을 통한 통계적 보정 효과가 매우 긍정적으로 나타났다. 이는 SVM이 WRF 모델이 가지고 있는 풍속의 과대 모의 및 예측 현상(가령, Lee *et al.*, 2014)을 줄일 수 있는 좋은 대안이 될 수 있음을 제시한다.

강수량의 경우에도 전반적으로 LAMP\_SVM이 GDAPS와 비슷하거나 관측치에 더 가까운 값을 나타냈다(Table 5). 8, 11월에는 GDAPS의 예측이 더 우수

**Table 3.** Statistical scores (Bias, RMSE) between medium-range model prediction result and observation during about 12 days in July to November 2017 by air temperature

No.	Statistical score	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM
		July		August		September		October		November	
1	Bias	-0.46	-0.51	-0.39	2.06	-0.76	0.99	-0.89	1.92	-0.94	4.93
	RMSE	<b>2.45</b>	<b>2.42</b>	1.88	3.03	2.18	2.69	2.37	2.83	3.22	5.39
2	Bias	-0.79	-0.42	-0.67	1.05	-1.51	-0.19	-0.79	1.06	0.98	4.93
	RMSE	2.32	2.44	1.61	2.16	<b>2.58</b>	<b>2.52</b>	2.19	2.5	3.22	5.48
3	Bias	-1.74	0.63	-1.2	2.54	-1.92	1.23	-2.18	2.42	-0.83	5.75
	RMSE	<b>2.96</b>	<b>2.51</b>	2.22	3.47	<b>2.59</b>	<b>2.51</b>	2.64	3.08	2.91	6.08
4	Bias	-1.54	-0.34	-1.45	2.08	-1.8	1.09	-1.99	1.52	-0.73	5.35
	RMSE	<b>2.87</b>	<b>2.53</b>	2.46	3.16	2.4	2.4	<b>2.5</b>	<b>2.29</b>	3.02	5.7
5	Bias	1.01	4.01	0.78	3.44	0.12	2.46	0.18	3.45	0.38	5.91
	RMSE	2.4	4.54	1.83	4.01	1.99	3.49	2.08	4.12	2.92	6.39
6	Bias	-0.18	1.19	0.21	2.73	-0.8	1.66	-0.21	3.26	0.89	5.68
	RMSE	2.24	2.58	1.65	3.49	2.17	2.89	2.1	4.01	3.39	6.48
7	Bias	-0.73	-0.71	-0.43	1.41	-1.22	0.73	-0.87	1.91	0.07	4.94
	RMSE	2.06	2.48	1.85	2.77	2.62	3.33	2.53	3.88	3.01	5.74
8	Bias	1.25	-1.02	1.4	2.59	0.83	1.76	0.63	2.82	2.02	5.67
	RMSE	2.39	2.4	2.57	3.56	2.71	3.49	2.76	4.07	3.76	6.38
9	Bias	0.41	1.46	0.43	2.56	-0.29	1.65	-0.07	3.15	1.15	5.85
	RMSE	2.18	2.68	1.92	3.47	2.07	2.84	2.03	3.89	3.41	6.57

**Table 4.** Statistical scores (Bias, RMSE) between medium-range model prediction result and observation during about 12 days in July to November 2017 by 10-m wind

No.	Statistical score	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM
		July		August		September		October		November	
1	Bias	1.07	2.1	1.19	1.31	0.19	0.62	1.1	1.64	1.66	1.49
	RMSE	1.65	2.8	1.85	2	1.08	1.34	1.67	2.02	<b>2.47</b>	<b>1.92</b>
2	Bias	0.45	1.04	0.72	0.45	0.26	0.17	1.2	0.96	1.32	0.43
	RMSE	1.17	1.97	1.54	1.54	<b>1.31</b>	<b>1.24</b>	<b>1.96</b>	<b>1.59</b>	<b>2.56</b>	<b>1.73</b>
3	Bias	0.24	1.79	0.45	0.79	-0.43	0.17	0.59	1.32	0.5	0.37
	RMSE	1.27	2.71	1.53	1.76	<b>1.26</b>	<b>1.14</b>	1.35	1.72	<b>1.77</b>	<b>1.35</b>
4	Bias	0.68	2.09	0.92	1.45	-0.08	0.77	0.76	1.52	0.92	1.37
	RMSE	1.44	2.85	1.66	1.84	1.04	1.28	1.51	2.02	<b>1.99</b>	<b>1.97</b>
5	Bias	0.93	2.06	1.13	1.36	0.56	1.15	1.07	1.32	1.38	1.11
	RMSE	1.51	2.56	1.6	1.67	1.15	1.58	1.64	1.88	<b>2.27</b>	<b>1.64</b>
6	Bias	0.18	1.53	0.35	0.87	0	0.31	0.73	0.98	0.69	0.43
	RMSE	1.12	2.23	1.2	1.46	<b>1.18</b>	<b>1.08</b>	<b>1.51</b>	<b>1.4</b>	<b>2.01</b>	<b>1.38</b>
7	Bias	0.34	1.16	0.51	0.79	0.15	0.52	0.82	0.69	0.39	0.9
	RMSE	1.2	2.03	1.2	1.2	0.93	1.06	<b>1.57</b>	<b>1.04</b>	1.6	1.62
8	Bias	0.69	2.15	0.98	0.94	0.43	0.48	1.04	1.15	1.18	0.35
	RMSE	1.26	2.49	<b>1.38</b>	<b>1.2</b>	<b>1.13</b>	<b>0.91</b>	<b>1.6</b>	<b>1.41</b>	<b>1.97</b>	<b>1.00</b>
9	Bias	0.6	1.69	0.8	0.91	0.24	0.47	0.96	0.97	0.85	0.74
	RMSE	1.3	2.41	1.27	1.33	<b>0.95</b>	<b>0.79</b>	<b>1.61</b>	<b>1.33</b>	<b>2.06</b>	<b>1.42</b>

**Table 5.** Statistical scores (Bias, RMSE) between medium-range model prediction result and observation during about 12 days in July to November 2017 by hourly precipitation

No.	Statistical score	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM	GDAPS	LAMP_SVM
		July		August		September		October		November	
1	Bias	1.25	0.46	-0.07	-0.18	0.06	0.03	0.11	0.13	-0.15	-0.11
	RMSE	<b>5.74</b>	<b>3.44</b>	4.89	6.64	<b>0.41</b>	<b>0.1</b>	<b>0.58</b>	<b>0.42</b>	1.52	1.89
2	Bias	0.08	-0.14	0.28	-0.17	0	0.01	0.07	0.09	0.06	0.04
	RMSE	<b>3.39</b>	<b>3.31</b>	<b>1.62</b>	<b>1.5</b>	<b>0.56</b>	<b>0.54</b>	<b>0.41</b>	<b>0.28</b>	<b>0.37</b>	<b>0.26</b>
3	Bias	1.32	0.54	0.67	0.28	0.04	0.11	0.13	0.14	0.07	0.08
	RMSE	<b>6.28</b>	<b>3.3</b>	2.84	4.44	0.25	0.46	<b>0.67</b>	<b>0.58</b>	0.47	0.53
4	Bias	1.13	0.56	0.47	0.27	0.07	0.07	0.11	0.12	-0.02	0.07
	RMSE	<b>5.65</b>	<b>3.92</b>	2.78	4.74	<b>0.41</b>	<b>0.24</b>	<b>0.58</b>	<b>0.37</b>	0.8	1.24
5	Bias	0.88	0.52	0.33	-0.02	0.07	0.08	0.13	0.14	0.07	0.15
	RMSE	<b>4.63</b>	<b>3.9</b>	4.02	5.52	<b>0.52</b>	<b>0.43</b>	<b>0.69</b>	<b>0.55</b>	0.36	0.58
6	Bias	0.37	0.34	0.6	-0.27	0.07	0.04	0.1	0.13	0.03	0.08
	RMSE	<b>2.85</b>	<b>2.49</b>	<b>2.73</b>	<b>2.05</b>	<b>0.53</b>	<b>0.18</b>	0.54	0.62	0.4	0.42
7	Bias	0.99	0.44	0.6	0.13	0.1	0.09	0.13	0.15	0	0.04
	RMSE	<b>4.83</b>	<b>3.06</b>	2.66	4.36	<b>0.55</b>	<b>0.52</b>	<b>0.68</b>	<b>0.61</b>	0.74	0.96
8	Bias	0.85	0.14	0.08	-0.38	0.1	0.1	0.16	0.27	0.03	0.14
	RMSE	<b>4.98</b>	<b>1.77</b>	4.95	6.3	<b>0.56</b>	<b>0.55</b>	0.75	0.99	0.13	0.59
9	Bias	0.54	0.49	0.52	-0.39	-0.03	-0.05	0.1	0.09	0.01	0.05
	RMSE	2.64	2.99	<b>2.38</b>	<b>1.82</b>	1.05	1.06	<b>0.54</b>	<b>0.5</b>	0.41	0.59

했으나 7, 9, 10월은 LAMP\_SVM가 GDAPS의 예측력을 앞섰다. LAMP의 강수량 예측자료는 눈, 밭, 산림지역의 토양수분 모델에 고해상도 중기 입력자료로 제공되기 때문에(Hong *et al.*, 2015; 2020; Park *et al.*, 2018; Shin *et al.*, 2017), 이러한 강수량 예측의 통계적 개선은 농림 가뭄지수(Lee *et al.*, 2020; Lee *et al.*, 2021)와 산사태 예측(So *et al.*, 2020) 등의 정확도로 이어질 것으로 기대된다.

보정 후의 예측력이 보정 전보다 높아지는 것은 보정 전의 LAMP 모델예측값에 SVM을 튜닝(변수조정)한 예측값을 산술평균한 편의보정 값을 적용하여, 보정 전의 LAMP가 가지고 있던 오차를 줄여 주었기 때문이다. 특히, 9월과 10월에서 보여준 오차의 개선 수준은 Hyeon *et al.*(2015)에서 보인 풍속의 오차의 개선 수준과 비교할 때, 그보다 높은 수준이다. 이러한 결과는 지역 수치예보 모델이 시간 적분됨에 따라 경계 자료가 모델 영역 내부로 흘러 들어오면서 영역 내 예측가능성이 점점 저하되는 현상(Lee *et al.*, 2015)을 완화시키는

데 SVM 방법이 일정 부분 기여할 수 있음을 가리킨다. 양질의 입력 자료가 더욱 축적되고 훈련 기간이 증대되면 현업 표출 중인 NCAM AgroMeteogram (Lee *et al.*, 2017)의 예측력도 높아질 것으로 기대된다. 향후 물리역학적 및 정성적 분석을 병행하여 수치모델과 인공지능의 빅데이터 상호작용에 대한 이해도를 더욱 높이고, SVM 외에 다른 방법을 다중적으로 적용하는 방안도 시도할 계획이다.

## 적 요

NCAM-LAMP 중기예측 자료의 통계적 후처리와 개선을 위하여 R 기반의 지점 시계열 자료 검증 체계를 구축하였다. 이 시계열 검증체계를 이용하여 기상청 AWS 관측 자료와 NCAM-LAMP, KMA GDAPS 중기예측 모델 자료를 비교하였다. 이를 위해 관측 지점에 가장 근접한 모델 위도 및 경도 자료를 추출하여 총 9개 지점을 선정하였다. 각 지점에 대해 NCAM-LAMP,

GDAPS 모델의 기온, 강수량, 풍속 일평균 예측 자료를 관측과 비교한 결과, 모델들은 풍속의 과대예측 경향을 뚜렷이 보였으며, 기온과 강수의 경우에는 두 모델의 예측력이 월별 및 변수별로 다르게 나타났다. 이를 바탕으로 본 연구에서는 통계적 기법을 개발하여 NCAM-LAMP가 가지고 있는 오차를 줄이고자 하였다. 모델 오차를 줄이기 위해 일반적으로 쓰이는 MOS (Model Output Statistics) 기법 중에 인공지능 SVM (Support vector machine) 방식을 8~10월 기간에 적용한 결과, 8월에 비해서 10월이, 기온 변수에 비해서 바람과 강수 변수가 개선된 효과를 보여주었다. 이러한 결과는 풍속의 과대예측을 줄이고, 농림 가뭄지수와 산사태 예측 등을 개선시키며, 지역 수치예보 모델이 시간 적분됨에 따라 영역 내 예측가능성이 점점 저하되는 현상을 완화시키는데 SVM 방법이 일정 부분 기여할 수 있음을 가리키며, 현업 표출 중인 NCAM Agro-Meteogram 개선에도 도움을 줄 것으로 기대된다.

## 감사의 글

본 연구는 산림청(한국임업진흥원) 산림과학기술연구개발사업(2021341B10-2123-CD01)의 지원으로 수행되었습니다.

## REFERENCES

- Antolik, M. S., 2012: Model Output Statistics (MOS) - Objective Interpretation of NWP Model Output. Presented at the University of Maryland USA (2012).
- Asano, A., 2004: Support vector machine and kernel method Pattern information processing. *Pattern information Processing (2004 Autumn Semester) Session 12. (05. 1. 21)*
- Bae, K. Y., H. S. Jang, and D. K. Sung, 2017: Solar power prediction based on machine learning scheme and its error analysis. *Proceedings of Symposium of the Korean Institute of communications and Information Sciences*, 13-14.
- Hyeon, B., Y. Lee, and K. Seo, 2015: Evolutionary nonlinear regression based compensation technique for short-range prediction of wind speed using automatic weather station. *The Transactions of the Korean Institute of Electrical Engineers* **64**(1), 107-112.
- Hong, M.-K., S.-H. Lee, J.-Y. Choi, S.-H. Lee, and S.-J. Lee, 2015: Estimation of soil moisture and irrigation requirement of upland using soil moisture model applied WRF meteorological data. *Journal of the Korean Society of Agricultural Engineers* **57**(6), 173-183. (in Korean with English Abstract)
- Hong, M., S. Lee, S.-J. Lee, and J. Y. Choi, 2020: Application of high-resolution meteorological data from NCAM-WRF to characterize agricultural drought in small-scale farmlands based on soil moisture deficit. *Agricultural Water Management* **243**, 106493.
- Hwang, J., H.-O. Cho, Y. Lim, S.-W. Son, E.-J. Kim, J.-O. Lim, and K.-O. Boo, 2020: Extratropical prediction skill of KMA GDAPS in January 2019. *Atmosphere* **30**(2), 115-124. (in Korean with English Abstract)
- Jacobson, T., J. James, and N. C. Schwertman, 2009: An example of using linear regression of seasonal weather patterns to enhance undergraduate learning. *Journal of Statistics Education* **17**(2).
- Jo, N.-H., 2006: SVM load forecasting using cross-validation. *The Korean Institute of Electrical Engineers A*, **55**(11), 485-491.
- Kim, D., and K. Seo, 2015: Comparison of linear and nonlinear regressions and elements analysis for wind speed prediction. *Journal of Korean Institute of Intelligent Systems* **25**(5), 477-482.
- KMA (Korea Meteorological Administration), 2016: *Ground Weather Observation Guidelines*. (Available at <https://book.kma.go.kr/viewer/MediaViewer.ax?cid=33393&rid=5&moi=5241> accessed on 13 December 2021)
- Kum, D., C. Jang, J. Lee, and K. J. Lim, 2014: Analysis of rainfall according to the bias correction of climate change scenarios. *Proceedings of the Korea Water Resources Association Conference (2014-05)*, 470pp.
- Kwon, S., and S.-J. Lee, 2017: A statistical verification and improvement tool of point time series data of the NCAM-LAMP mid-term prediction system. *Poster session presented at: 18th Conference on Agricultural and Forest Meteorology*, 2017 November 10, Seoul National University.
- Lee, S., S.-J. Lee, J. H. Kang, and E. S. Jang, 2021: Spatial and temporal variations in atmospheric ventilation index coupled with particulate matter concentration in South Korea. *Sustainability* **13**(16), 8954.
- Lee, S.-J., J. Kim, M. Kang, and B. Malla-Thakuri, 2014: Numerical simulation of local atmospheric



- circulations in the valley of Gwangneung KoFlux sites. *Korean Journal of Agricultural and Forest Meteorology* **16**, 244-258.
- Lee, S.-J., J. S. Kang, and H. Yoo, 2015: *Atmospheric Modeling, Data Assimilation, and Predictability (Korean Version)*. SigmaPress.
- Lee, S.-J., J. Song, and Y.-J. Kim, 2016: The NCAM Land-Atmosphere Modeling Package (LAMP) Version 1: Implementation and evaluation. *Korean Journal of Agricultural and Forest Meteorology* **18**(4), 307-319.
- Lee, S.-J., Y. Kim, J. Song, and J. Kim, 2017: An Agrometeorological application of the meteogram: "Agrometeogram". American Meteorological Society 97th Annual Meeting, 22-26 January 2017, Seattle, WA, U.S.A.
- Lee, S., S.-J. Lee, and J. S. Koo, 2020: Database construction of high-resolution daily meteorological and climatological data using NCAM-LAMP: Sunshine hour data. *Korean Journal of Agricultural Forest Meteorology* **22**(3), 135-143. (in Korean with English Abstract)
- Mokhtarzad, M., F. Eskandari, N. J. Vanjani, and A. Arabasadi, 2017: Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environmental Earth Science* **76**, 729.
- Moon, H., J. Baik, S. Hwang, and M. Choi, 2014: Spatial downscaling of grid precipitation using support vector machine regression. *Journal of Korea Water Resources Association* **47**(11), 1095-1105.
- Park, J., H.-S. Kim, S.-J. Lee, and T. Ha, 2018: Numerical evaluation of JULES surface tiling scheme with high-resolution atmospheric forcing and land cover data. *SOLA* **14**, 19-24.
- Shin, Y. H., J. Y. Choi, S.-J. Lee, and S. H. Lee, 2017: Estimation of irrigation requirements for red pepper using soil moisture model with high resolution meteorological data. *Journal of the Korean Society of Agricultural Engineers* **59**(5), 31-40.
- So, Y. Y., S. J. Lee, S. W. Choi, and S.-J. Lee, 2020: Construction of NCAM-LAMP precipitation and soil moisture database to support landslide prediction. *Korean Journal of Agricultural and Forest Meteorology* **22**(3), 152-163. (in Korean with English Abstract)
- Statnikov, A., C. F. Aliferis, D. P. Hardin, and I. Guyon, 2013: *A Gentle Introduction to Support Vector Machines in Biomedicine: Case Studies*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1st edition, 2011.
- Wu, C. L., K. W. Chau, and Y. S. Li, 2008: River stage prediction based on a distributed support vector regression. *Journal of Hydrology* **358**, 96-111.
- Wilks, D. S., 2009: Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications* **16**(3), 361-368.
- Yoon, H., P. Yoon, E. Lee, and G.-B. Kim, and S.-H. Moon, 2016: Application of machine learning technique-based time series models for prediction of groundwater level fluctuation to national groundwater monitoring network data. *Journal of the Geological Society of Korea* **52**(3), 187-199.