

## 제주 감귤 과수원에서의 이슬지속시간 예측 모델 평가

박준상 · 서윤암 · 김규랑\* · 하종철

국립기상과학원 응용기상연구과

(2018년 8월 23일 접수; 2018년 9월 12일 수정; 2018년 9월 13일 수락)

## Evaluating the prediction models of leaf wetness duration for citrus orchards in Jeju, South Korea

Jun Sang Park, Yun Am Seo, Kyu Rang Kim\* and Jong-Chul Ha

Applied Meteorology Research Division, National Institute of Meteorological Sciences, Jeju 63568, Korea

(Received August 23, 2018; Revised September 12, 2018; Accepted September 13, 2018)

### ABSTRACT

Models to predict Leaf Wetness Duration (LWD) were evaluated using the observed meteorological and dew data at the 11 citrus orchards in Jeju, South Korea from 2016 to 2017. The sensitivity and the prediction accuracy were evaluated with four models (i.e., Number of Hours of Relative Humidity (NHRH), Classification And Regression Tree/Stepwise Linear Discriminant (CART/SLD), Penman-Monteith (PM), Deep-learning Neural Network (DNN)). The sensitivity of models was evaluated with rainfall and seasonal changes. When the data in rainy days were excluded from the whole data set, the LWD models had smaller average error (Root Mean Square Error (RMSE) about 1.5hours). The seasonal error of the DNN model had the similar magnitude (RMSE about 3 hours) among all seasons excluding winter. The other models had the greatest error in summer (RMSE about 9.6 hours) and the lowest error in winter (RMSE about 3.3 hours). These models were also evaluated by the statistical error analysis method and the regression analysis method of mean squared deviation. The DNN model had the best performance by statistical error whereas the CART/SLD model had the worst prediction accuracy. The Mean Square Deviation (MSD) is a method of analyzing the linearity of a model with three components: squared bias (SB), nonunity slope (NU), and lack of correlation (LC). Better model performance was determined by lower SB and LC and higher NU. The results of MSD analysis indicated that the DNN model would provide the best performance and followed by the PM, the NHRH and the CART/SLD in order. This result suggested that the machine learning model would be useful to improve the accuracy of agricultural information using meteorological data.

**Key words:** Leaf Wetness Duration, Mean Square Deviation, Number of Hours of Relative Humidity, Classification And Regression Tree, Penman-Monteith, Deep-learning Neural Network



\* Corresponding Author : Kyu Rang Kim  
(krk9@kma.go.kr)

## I. 서 론

한국의 농가수익은 과실 작물이 대부분을 차지하고 있으며, 제주도 역시 감귤이 2015년 기준으로 제주도 전체 농경지면적 중 약 32.8%를 차지할 정도로 주력 산업으로서 재배되고 있다. 감귤은 생육과 저장 단계에서 페니실리움에 의한 감귤 부패병, 검은썩음병 등의 곰팡이류 병해 발생 확률이 높고 온도와 습도 변화에 민감하다. 감귤 생산은 많은 부분이 노지에서 이루어지고 있으며 생육 단계에서 감귤 수확량 및 병해충 영향은 기상요소에 매우 민감하게 반응한다. 다양한 기상 요소 중 엽면습윤지속시간 혹은 이슬지속시간(Leaf Wetness Duration, LWD<sup>1)</sup>)은 그 정도에 따라 농업 환경에서 병충해를 발생시키는 주요 원인이 될 수 있으며, 작물의 성장 및 생산량을 감소시키는 간접적인 원인이 되기도 한다. 일반적으로 식물은 아침에 일사로 인해 기온이 상승하면서 식물동화 작용이 활발하게 되는데 식물 잎에 이슬이 맺혀있다면 이슬의 증발에 잠열과 같은 많은 에너지를 소비하게 된다. 따라서, 식물 기온은 하강하게 되고 식물의 광합성 작용을 방해 받게 된다. 또한, 식물 잎에 맺힌 이슬은 기상상태에 따라서 서리로 발전하여 작물 피해가 발생되고 생산량을 저하시킨다. 이러한 이슬 발생 및 지속 현상은 강우나 안개와 같은 대기 상태나 잎 표면의 성질에 의해서 결정된다(Klemm et al., 2002). 또한, 이슬은 대기와 지표 상태에 따라 발생 및 유지 시간이 다르게 나타난다. 보통 이슬은 야간에 지표 장파복사로 인해 기온이 이슬점까지 하강하여 대기의 수증기가 잎 표면에 물방울 형태로 나타나거나 강우 발생 시 잎 표면에 맺히는 것으로 정의한다. 이슬은 바람이 강하지 않고 습도가 높은 상태에서 온도가 급격하게 하강 할 때 발생하는 매커니즘과 일출 후 일사에 의해 소멸되는 매커니즘으로 구분 할 수 있다. 이처럼 이슬지속시간은 온도와 함께 식물 잎과 과실의 질병 발생을 결정하는 핵심 매개 변수이며, 최근 미국의 대규모 농업 단지에서는 딸기의 질병경고 시스템(Strawberry Advisory System, SAS)을 개발하여 이슬지속시간에 의해 발생할 수 있는 질병의 위험을 예측하여 살균제 살포를 결정하는 시스템을 운영 중이다(Montone et al., 2016).

이슬지속시간은 농업에서 매우 중요한 역할을 하지만 기상학적 변수로는 중요도가 낮게 여겨져 이를 측정하거나 추정하는 표준 프로토콜이 없다(Gleason, 2007). 또한 대기에서 수분의 과포화가 발생하게 되면 과실 표면에는 이슬이 발생하게 되고 관측장비 중에서도 이를 측정하는 센서가 있지만 실제 재배지에서 사용되기보다, 연구용으로 국한되어 이용되고 있다. 이슬지속시간 자료는 이슬 여부를 판별하는 센서를 이용하여 해당 시간의 이슬 여부를 관측하고 이슬로 판별된 시간의 합을 계산하여 생산한다. 이슬 센서는 정확한 관측 기준이 없고 센서 종류가 다양하기 때문에 반드시 보정 작업을 거쳐서 사용해야 한다(Rowlandson et al., 2015). 따라서 이슬지속시간은 현실적으로 정확하게 관측하는 것이 어려운 기상요소이며 주로 이슬지속시간 측정의 대안으로 ‘이슬 여부’를 추정하는 다양한 모델로 대체되었다(Beysens, 2016; Beysens et al., 2005; Cosh et al., 2009; Luo and Goudriaan, 2004; Maestre-Valero et al., 2012). 하지만 이마저도 모델 정확도의 문제로 영농현장에서 널리 사용되지 못하고 연구용으로만 사용된다는 한계가 있다.

기존에 개발된 이슬지속시간 모델은 입력 유형에 따라 물리적 모델과 경험적 모델 두 가지로 분류한다(Kim et al., 2002). 또한, 최근에는 인공지능 기법 중 기계학습 기법(Support Vector Machine, Random Forest, Artificial Neural Network)을 이용하여 예측 모델 개발이 이루어지고 있다. 물리적 모델은 에너지 평형을 기초로 식물 표면과 대기 사이의 이슬 형성과 증발 과정을 기반으로 개발되었고, 경험적 모델들은 대부분, 잎 표면의 습기와 기상 자료 사이의 관계를 통한 회귀 분석 및 기타 통계 기법을 사용하여 개발되었다(Sentelhas et al., 2008). 경험적 모델은 입력변수가 적다는 장점이 있으나, 특정 지역에 특화되었기 때문에 다른 지역에 모델을 적용할 때는 모델을 새로이 구축해야하는 단점이 있다. 물리적 모델은 다른 지역에 적용하기 쉽고 오류 발생이 적지만, 습도 및 태양 순복사 등의 기상자료가 정밀하지 않다면 정확성이 감소하는 문제점이 있다. 기계학습 기법은 최근 여러 사례에서 우수한 성능을 보여주고 있는 모델이지만(Lee and Nang, 2016; Park et al., 2017; Weiland et al.,

1) Leaf Wetness Duration은 엽면습윤지속시간으로 강수와 이슬로 인해 습윤이 지속되는 정도를 의미한다. 이때, 강수로 인한 습윤 조건을 제외하고 이슬만으로 습윤이 지속되는 정도는 이슬지속시간이라고 표현 할 수 있으며, 본 연구에서는 강수일을 제외한 사례를 논의하기 때문에 이후 LWD는 이슬지속시간으로 정의한다.

2011), 결과가 도출되기까지의 명확한 알고리즘을 설명할 수 없는 블랙박스 영역이 존재하며, 훈련된 범위를 벗어난 경우가 관측되었을 때는 예측 정확도가 떨어지는 단점이 있다. 이러한 모델 특성을 고려하여 기존에 가장 많이 연구된 4가지 이슬지속시간 예측 모델 중 제주 지역에 적용 가능한 이슬지속시간 모델을 알아보고자 하였다.

제주는 사면이 바다로 둘러싸인 섬이자, 남한에서 가장 높은 한라산이 있으며, 그리고 한반도에서 가장 남쪽으로 치우쳐진 지리적 환경으로 인해 제주만의 기후 특징이 있다. 제주의 연평균 상대습도는 70-80%로 월별 습도 변화가 거의 없으며, 연평균 풍속은 3-6 m/s로 내륙보다 강하게 나타나는 특성을 가지고 있으며, 최근에는 지구온난화로 인해 최저기온이 점차 상승하여 열대성 기후와 유사한 기후 분포를 나타낸다(Kim *et al.*, 2006). 기온, 풍속, 그리고 습도는 일사량과 더불어 이슬지속시간 예측 모델에서 가장 핵심 기상인자로 활용되는데 제주만의 독특한 기후 특징으로 인해 기존에 개발된 이슬지속시간 예측 모델은 기존과는 다른 예측 정확도가 나타날 것으로 예상된다. 본 연구는 제주의 기후 특성에 맞는 정확한 이슬지속시간을 예측하기 위한 모델을 선정하고자 서로 다른 특성을 지닌 4가지 모형을 통계적 오차 분석을 통해 성능을 평가하였다.

## II. 자료 및 방법

### 2.1. 자료 수집

본 연구에 사용된 기상 및 이슬 자료는 제주 전역 11개 기상관측 지점에서 수집하였다. 관측 지점들은 제주도 농업기술원(agri.jeju.go.kr)에서 2014년 7월부터 운영되어 오고 있다(Fig. 1). 관측 자료는 11개 지점에서 동일하게 결측이 없는 구간인 2016년 1월부터 2017년 12월까지 기간을 수집하였다. 각 지점은 기온, 습도, 풍향 및 풍속, 일사, 강우량 등의 기상 센서와 농업 기상 특성 분석을 위해 토양수분 및 토양온도와 초상온도 그리고 이슬 습윤 센서가 설치되어 있다. 관측 자료는 10분 간격으로 수집되었고 본 연구에서는 NHRH 모델에서 10분 단위의 상대습도 자료를 사용한 경우를 제외하고 1시간으로 누적평균하여 사용하였다. 또한, 이슬지속시간을 예측하는 물리 모델에서는 순복사와 기압 자료가 필요한데, 이를 위해 기압은 인접한 기상청 종관 관측 자료(지점번호: 184, 185, 189, 188)를 사용하였고 순복사는 Allen(1998)이 제시한 순복사 에너지 계산식을 이용하여 계산한 값을 사용하였다. 이때, 감굴나무는 1.5 m 정도의 높이로 겨울에도 나뭇잎이 거의 떨어지지 않기 때문에 1 m 이상 초지의 알베도인 0.16을 모든 계절에 적용하였다. 관측 지점은 주로 고도 150 m 이하이며, 밭 지역 3곳을 제외하고 모두 감굴 과수원 내부에 위치하고 있다(Table 1). 제주도에 재배되는 감굴 나무의 높이는 대부분 2 m 내외로 열매와 잎은 아래쪽에 주로 분포하고 있고 11개 지점에서 설치된 이슬 센서는 지면고도

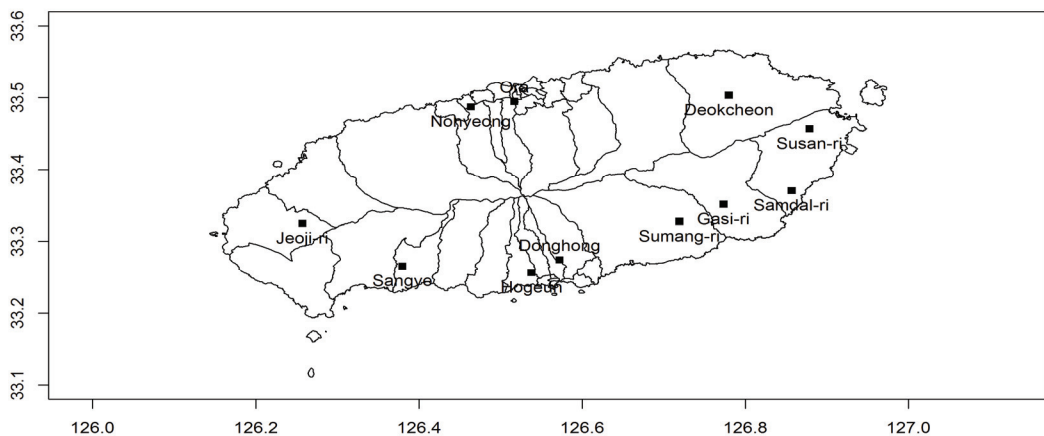


Fig. 1. Observation sites of Jeju agricultural research.

**Table 1.** Agro-meteorological network of observation in Jeju

	Site	Latitude	Longitude	Altitude (m)	Land cover
1	Deokcheon-dong	33.3	126.46	130	Field
2	Donghong-dong	33.16	126.34	143	Orchard (citrus)
3	Gasi-ri	33.21	126.46	88	Orchard (citrus)
4	Hogeun-dong	33.15	126.32	116	Orchard (citrus)
5	Jeoji-ri	33.19	126.15	114	Field (corn)
6	Ora-dong	33.29	126.31	70	Orchard (citrus)
7	Samdal-ri	33.22	126.51	29	Orchard (citrus)
8	Sangye-dong	33.15	126.22	161	Orchard (citrus)
9	Sumang-ri	33.19	126.43	145	Orchard (citrus)
10	Susan-ri	33.27	126.52	66	Field
11	Nohyeong-dong	33.29	126.27	52	Orchard (citrus)

2-2.5 m 사이에 설치되어 있다. 이슬은 미기상학적 특성에 따라 나무의 상부와 하부에 생성시간이 다르게 나타난다. Veronica *et al.*(2010)는 미국 플로리다에서 감귤 나무 높이에 따라 이슬지속시간을 관측한 결과 나무 상부에서 이슬지속시간이 가장 길게 나타났음을 밝힌바 있다. 따라서 제주도 귤나무 상부에서 측정된 이슬지속시간은 감귤 나무의 이슬을 대표하는데 적절하다고 판단하였다.

**2.2. 이슬지속시간 예측 모델**

본 연구에서 사용한 경험 모델은 Number of Hours of Relative Humidity (NHRH)와 Classification And Regression Tree/Stepwise Linear Discriminant (CART/SLD) 모델이며(Sentelhas, 2008; Gleason, 1994), 물리 모델은 증발량을 계산하는데 가장 뛰어난 성능을 보이는 Penman-Monteith 방법을 응용한 PM 모델을 사용하였다(Sentelhas *et al.*, 2006). 그리고 기계학습 모형은 단층 인공신경망을 다층으로 구성한 Deep-learning Neural Network (DNN) 모형을 사용하였다(LeCun *et al.*, 2015). 이들 4가지 모델의 물리적 또는 구조적인 특징은 다음과 같다.

**2.2.1. Number of Hours of Relative Humidity (NHRH):** 이 모델은 이슬지속시간을 추정하는 가장 간단한 모델로서 상대 습도가 90%의 특정 임계값을 초과하면 잎 표면에 이슬이 존재한다는 가정으로 개발되었다. Sentelhas *et al.*(2008)은 서로 다른 기후지역 4곳을 선정하여 실험한 결과, 지역적인 보정 후에

PM 모델 오차와 매우 유사한 오차가 나타남을 밝힌 바 있다. 본 연구에서는 상대습도 90% 이상의 상태가 20분 이상 지속될 때 이슬이 발생한 것으로 판단하였다.

**2.2.2. Classification and Regression Tree/Stepwise Linear Discriminant (CART/SLD):** Gleason *et al.*(1994)는 이슬지속시간을 추정하기 위해 상대습도, 이슬점온도, 그리고 풍속의 임계값을 기반으로 하는 계층적 결정 트리를 사용하는 경험적 CART/SLD 모델을 개발하였다. 임계값을 만족하지 않을 때는 이슬이 발생하지 않음으로 분류하고 임계값을 만족한다면 단계적 선형 판별 분석에 따라서 이슬 발생 여부가 결정된다. CART/SLD 모델의 입력 변수 중에는 표준 기상 관측소에서 일반적으로 관측 되지 않는 이슬점온도가 있기 때문에 식 (1)을 이용해 계산하고, 상대습도, 풍속, 온도는 관측을 통해 생성된다. 이렇게 관측된 변수를 이용하여 결정 트리를 적용하고 식 (2)와 식 (3) 경험식을 통해 최종적으로 이슬 발생여부가 결정된다. 상대적으로 물리적 모델 보다 입력변수가 적기 때문에 간단하면서도 구현하기 쉬운 모델이다.

$$T_{dew} = \left(\frac{RH}{100}\right)^{\frac{1}{8}}(112 + 0.9T_{air}) + 0.1T_{air} - 112 \quad (1)$$

$$\left[ \left( (1.6064 \sqrt{T_{air}}) + (0.0036 T_{air}^2) + (0.1531 RH) - (0.4599 W D) - (0.0035 T_{air} RH) \right) > 14.4674 \right] \quad (2)$$

$$[(0.7921 \sqrt{T_{air}}) + (0.0046 RH) - (23889 W) - (0.039 T_{air}W) + (1.0613 W D)] > 37 \quad (3)$$

여기서,  $T_{air}$ 는 기온, RH는 상대습도, W는 풍속, D는 기온과 이슬점 온도의 차이( $T_{air} - T_{dew}$ )이다.

**2.2.3. Penman-Monteith (PM):** PM 모델은 증발량을 계산하는 Penman-Monteith 방정식을 변형한 것으로 대기에서 잎 표면 방향의 음의 잠열 플럭스를 계산하여 이슬 유무를 결정하는 물리적 모델이다(식 (4)). 다른 물리 모델에 비해 PM모델은 작물관련 입력 변수가 필요하지 않으며, 비교적 적은 입력변수만을 필요로 하기 때문에 간단히 계산 할 수 있다는 장점이 있어, 본 연구의 대표적인 물리모델로 사용하였다. 하지만 그렇다 하더라도 경험적 모델보다는 기본적으로 많은 입력변수가 필요하며, 특히, 입력변수 중 순복사 에너지가 사용되는데 이 값의 정확도에 따라 예측 정확도 차이가 발생한다. PM 모델은 다양한 기후 조건 하에서 가장 낮은 공간 변동성을 가지며 평균 1.33 시간의 오차로 정확도와 활용도가 매우 우수하게 나타난 모델이다(Sentelhas *et al.*, 2006).

$$LE = -\frac{sR_n + [1200(e_s - e_a)/(r_a + r_b)]}{s + \gamma} \quad (4)$$

여기서, LE는 잠열플럭스(W/m<sup>2</sup>),  $s$ 는 포화수증기압의 기울기(hPa),  $e_a$ 는 실제수증기압(hPa),  $e_s$ 는 포화수증기압(hPa),  $\gamma$ 는 물리계수(0.64 kPa/K),  $R_n$ 는 케노피 순복사(J/min/cm<sup>2</sup>),  $r_a$ 는 바람 저항 계수이며,  $r_b$ 는 열전달 저항계수이다.

$$r_a = \frac{\ln[(Z_s - d)/(Z_0)]}{0.4u^*} \quad (5)$$

$$r_b = 307\left(\frac{D}{u}\right)^{1/2} \quad (6)$$

여기서,  $Z_s$ 는 센서의 높이(m),  $d$ 는 이동 높이(0.65  $Z_C$ ),  $Z_C$ 는 작물의 높이(m),  $Z_0$ 는 거칠기 길이(0.13  $Z_C$ ),  $u^*$ 는 마찰계수(m/s),  $D$ 는 잎의 유효 수치(0.07 m),  $u$ 는 풍속(m/s)이다.

**2.2.4. Deep-learning Neural Network (DNN):** 심층 신경망 모형은 다양한 연구에서 고전 통계 접근법보다 더 높은 정확도와 정밀도를 달성하는 것이 입증되어 최근에 가장 많이 사용하고 있는 기술이다. 본 연구에서 사용된 DNN 모델은 통계프로그램 R(3.4.4)에서 H2O 패키지를 사용하여 모델을 구성하였다. DNN 모델의 일반적인 구조는 입력층(input layer), 은

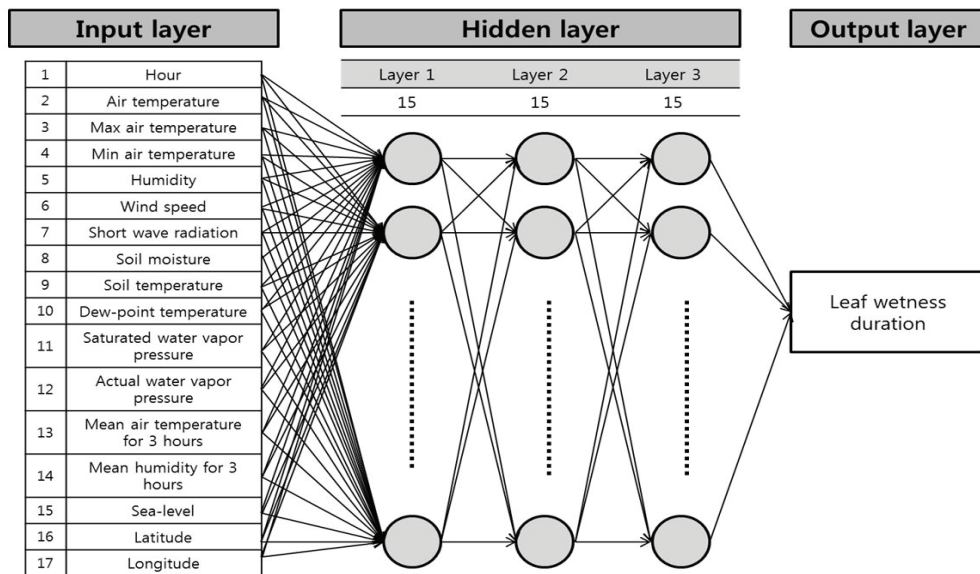


Fig. 2. Structure of the Deep-learning Neural Network (DNN) model for leaf wetness duration.

닉층(hidden layer), 출력층(output layer) 3가지로 구분된다(Fig. 2). DNN 모델 입력층에 사용된 기상인자는 기온, 1시간 동안의 최고 및 최저 온도, 상대습도, 풍속, 일사량, 토양수분, 토양온도, 그리고 관측 시간과 관측 지점의 위경도와 고도를 사용하였다. 또한, 이슬 형성에 관련이 있다고 판단되는 이슬점 온도, 포화수증기압, 현재수증기압, 과거 3시간 평균 기온 및 상대습도를 계산하여 입력층에 추가하였고 총 17개의 입력 노드를 사용하였다. 은닉층 수는 3개로 각 층의 노드는 15개로 설정하였고 활성화 함수(activation function)는 오류 발생에 적절히 대응하는 Rectified Linear Unit (ReLU)를 사용하여 모델을 구성하였다 (Shin and Kim, 2018). 입력층의 입력 변수는 앞서 언급한 기상인자를 사용하였으며, 출력층은 시간단위로 이슬 발생 유무를 확인 할 수 있도록 1개의 노드로 구성하였다. 2016년 1월 1일부터 2016년 12월까지 자료를 훈련 기간으로 구성하였고 모델 예측 정확도 테스트를 위해 2017년 이슬을 예측하여 실제 이슬 관측과 비교하였다. 은닉층과 노드 수는 Hinton *et al.*(2006)이 제안한 방법을 이용하여 경험적으로 가장 최소의 오차를 가지는 수로 결정하였으며, 구성된 모델을 훈련기간으로 검증한 결과는 상관계수가 0.79, 평균제곱근오차가 2.12 시간으로 나타났다.

**2.3. 모델 평가 방법**

보통 모델은 예측 정확도를 평가하기 위해 예측 자료와 관측 자료 사이의 통계적 방법을 사용하여 비교한다. 주요 평가 방법은 예측과 관측의 상관계수 (correlation coefficient, R)를 구하는 것이며 평균 절대 오차(Mean Absolute Error, MAE)와 평균 제곱근 오차(Root Mean Square Error, RMSE), 그리고 모델의 편의(Bias)를 분석하였고 주요 식은 아래와 같다.

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i - Y_i| \tag{7}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2} \tag{8}$$

$$Bias = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i) \tag{9}$$

여기서, N은 자료 개수, X는 예측값, Y는 관측값을 의미한다.

Wallach and Goffinet(1989)은 많은 농업 연구자들이 모델 평가를 위해 상관계수와 선형 회귀 방법을 자주 사용했지만, 예측 정확도를 직접적으로 분석하지 못함을 언급하였다. 또한, Gauch *et al.*(2003)은 상관계수와 선형 회귀 방법만으로는 모델 평가가 전반적으로 만족스럽지 않기에 Mean Squared Deviation (MSD)의 3 가지 구성요소를 분할하여 해석하는 방법을 제시하였다. 식 (10)으로 표현되는 MSD는 두 변수 사이의 제공된 편차 합을 단순히 관측 수 N으로 나눈 값이며, 두 변수 관계가 X=Y로 일치하는 정도를 알 수 있고, 이러한 기준에서 모델 예측과 관측 사이의 모델 예측 정확도를 판별하는 통계 분석 방법 중 하나로 활용되고 있다.

$$MSD = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \tag{10}$$

첫 번째 구성 요소 SB (squared bias)는 두 변수 평균의 불일치 정도를 판단하는 요소로 기하학적으로 X 방향으로 추세선 이동 정도를 의미하며 식 (11)과 같이 계산한다.

$$SB = (\bar{X} - \bar{Y})^2 \tag{11}$$

여기서,  $\bar{X}$ 는 X의 평균,  $\bar{Y}$ 는 Y의 평균이다.

두 번째 구성 요소는 NU (nonunity slope)로 표현되며 두 변수의 상관관계가 한쪽으로 치우쳐짐으로 인해 발생하는 기울기 정도를 표현한 것으로 식 (12)와 같다.

$$NU = (1 - b)^2 (\sum X_i^2 / N) \tag{12}$$

여기서, b는 X에 대한 Y의 최소 제곱 회귀의 기울기로  $\sum X_i Y_i / \sum X_i^2$ 로 나타낼 수 있다.

세 번째 구성 요소 LC (lack of correlation)는 식 (13)과 같으며 두 변수의 분포가 X=Y의 관계에서 얼마나 멀리 떨어져 있는지에 대한 분산 정도를 판단할 수 있다.

$$LC = (1 - r^2) (\sum Y_i^2 / N) \tag{13}$$

여기서,  $r^2$ 은 상관계수의 제곱으로  $(\sum X_i Y_i)^2 / \sum X_i^2 \sum Y_i^2$ 로 나타낼 수 있다.

세 가지 구성요소의 합은 MSD와 일치하기 때문에 각 구성요소의 영향력을 바탕으로 모델 성능에 대한 구체적인 분석이 가능하고 상관관계 및 선형 회귀 분석을 훌륭하게 보완할 수 있다. 본 연구에서는 4가지 모델을 MSD 방법을 이용하여 평가하고, 각 구성요소의 결과를 고려하여 모델 성능의 순위를 매겨보았다. 모델 평가를 위한 MSD의 값이 0이라면  $X=Y$ 라는 완전한 등가성 관계가 성립되고 0에 가까울수록 모델 성능은 우수하다고 판단할 수 있다.

### III. 결 과

#### 3.1. 강우 유무에 따른 모델 민감도

이슬의 발생과 소멸은 강우에 의한 매커니즘과 복사 에 의한 매커니즘으로 나눌 수 있는데, 강우에 의해 발생한 이슬의 발생 및 소멸 매커니즘은 기온 및 풍속 등 주변 기상인자의 영향이 복잡하게 작용된다. 따라서 본 연구에서는 강우일을 포함하는 날(all-day, AD)과 강우일을 포함하지 않은 날(clear-day, CD)로 구분하여 각 모델의 민감도를 테스트 하였다. 이슬지속시간 모델은 강우 현상을 표현하는 기상인자가 따로 없고 강우일 유무에 따라 이슬지속시간의 오차가 발생할 수 있다. DNN 모델은 특성상 훈련기간이 필요하며, 강우일을 포함하여 훈련한 AD 모델과 강우일을 제외하고 훈련한 CD 모델로 나누어 2개의 사례로 구분하였다. 두 사례의 DNN 모델은 2016년 기간에 대해 훈련 과정을 거쳤고, 2017년을 분석 기간으로 설정하였다. 동일한 기간의 비교를 위해 NHRH, CART/SLD, PM 모델 역시 2017년 기간에 대해 분석하였다. 11개 지점에서 강우 유무에 따라 이슬지속시간 예측 모델의 민감도 실험 결과는 Fig. 3과 같이 나타났다.

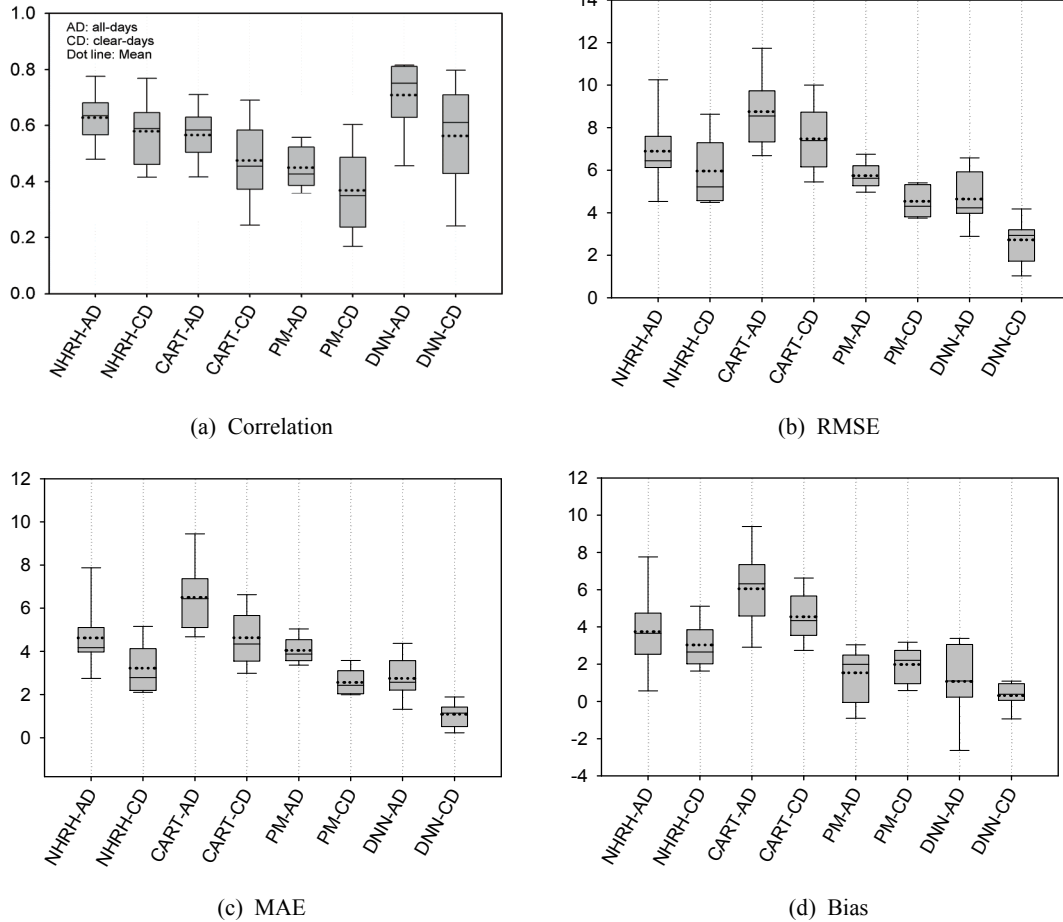
CD의 경우 4가지 모델 모두 상관계수가 감소하였는데 이는 총 이슬 사례일 중 강우일의 이슬 사례가 제외되어 나타난 현상으로 판단된다. 또한 각 모델 상관계수는 지점별로 편차가 더욱 커진 것을 확인할 수 있었다. 이는 종관규모로 전 지점 공통으로 나타난 강우일이 제외되어 지점별 기상인자의 특성이 더욱 강하게 작용된 것으로 판단된다. 특히 제주도는 일반적으로 한라산 때문에 동서남북의 기온과 습도가 다르게 나타나는 특징을 가지고 있기 때문에 지점 공통으로 나타난 특성을 제외한다면 지점별로 기온과 습도의 특성이 더욱 강화될 수 있음을 알 수 있다.

AD의 상관계수 평균은  $0.63 \pm 0.09$  (NHRH),  $0.57 \pm 0.09$  (CART/SLD),  $0.45 \pm 0.07$  (PM), 그리고  $0.71 \pm 0.11$  (DNN)로 나타났고, CD의 상관계수 평균은  $0.58 \pm 0.11$  (NHRH),  $0.48 \pm 0.14$  (CART/SLD),  $0.37 \pm 0.14$  (PM), 그리고  $0.56 \pm 0.16$  (DNN)으로 나타났다. 모든 모델의 AD와 CD에서는  $0.5\%$  ( $p < 0.005$ ) 수준에서 통계적으로 유의한 상관을 보였고 CART/SLD와 PM 모델의 CD 평균 상관계수는 다른 모델보다 낮은 수치로 나타났다. 이는 NHRH, CART/SLD, PM 모델 모두 제주도 남쪽 지역에 해당하는 2번(동홍)과 4번(호근)에서 가장 낮은 상관계수( $0.18 \sim 0.42$ )가 나타났기 때문이다. 제주도 남쪽 지점은 타 지점에 비해 비교적 습도가 높은 지역이며, 기존 연구들에 사용된 NHRH, CART/SLD, 그리고 PM 모델로는 이들 지역의 이슬 예측 정확도가 떨어지는 결과가 나타났다. 이에 반해 남쪽 2번과 4번 지점의 DNN 모델은 AD ( $0.61, 0.81$ )와 CD ( $0.43, 0.61$ ) 모두 상관계수가 높았다.

이슬지속시간 모델들은 AD에 비해 CD에서 평균 MAE와 RMSE 모두 약 1.5 시간 정도의 오차가 감소하였다. Bias는 PM 모델을 제외한 3개 모델에서 AD에 비해 CD의 평균값이 모두 감소하였고, DNN 모델은 1번(덕천)과 11번(노형) 지점을 제외하고 모두 양(+)의 Bias로 나타났다. CD 경우 4가지 모델 모두 Bias의 지점별 편차가 감소하였고 특히 DNN 모델은 편차가 크게 감소하여 모델 안정성이 향상되었다. PM 모델은 2번, 4번, 그리고 11번 지점을 제외하고 강우일 제외 시 Bias가 증가하였는데 이는 강우일에 과소모의가 많이 발생했음을 의미한다. PM 모델은 구조 특성상 복사에너지에 민감하게 반응하며, 야간이나 흐린날 처럼 순복사에너지가 음의 값이나 0에 가까운 값일 경우 과소모의하는 경향이 나타난다(Park *et al.*, 2017; Lee *et al.*, 2018). 이처럼 강우일을 제외하면 모델 오차는 감소하게 되며 이러한 결과는 강우 발생 시 이슬 형성과 소멸의 매커니즘을 따로 구분하여 모델을 개발하는 것이 필요함을 의미한다.

#### 3.2. 계절에 따른 모델 민감도

각 모델의 계절적 성능을 비교하고자 NHRH, CART/SLD, PM, 그리고 DNN 모델의 이슬 발생 사례를 계절별로 분류하여 분석하였다(Fig. 4) 각 모델 상관계수의 평균은 봄에 가장 높고 여름에 가장 낮게 나타났다. PM 모델 상관계수의 평균은 다른 모델과

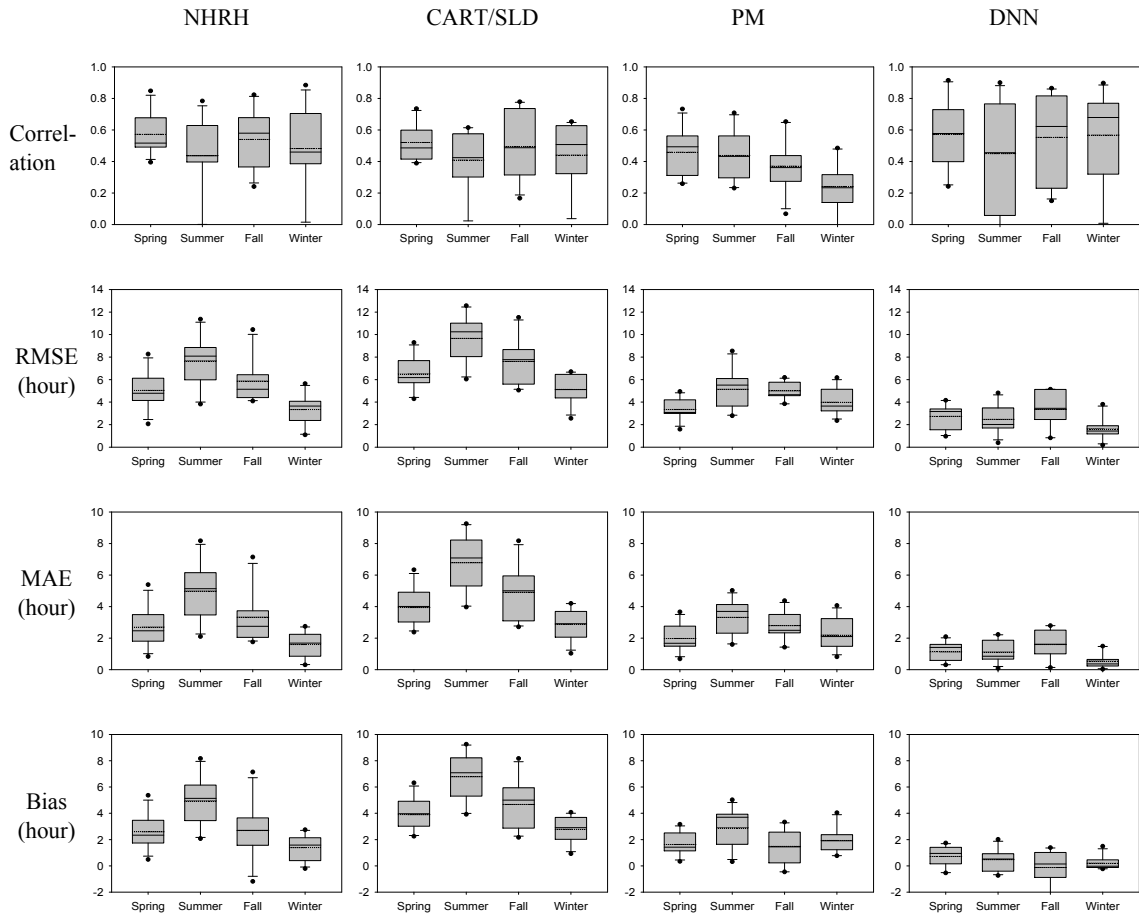


**Fig. 3.** Box plots of all-days (AD) and clear-days (CD) with NHRH, CART/SLD, PM, and DNN models at 11 sites. Dashed line is median and dot line is mean.

같이 봄에 가장 높았지만 가장 낮은 상관계수는 겨울에 나타났다. DNN 모델의 상관계수 평균은 여름을 제외한 나머지 계절에서는 0.5이상의 상관관계를 보였지만 모든 계절에서 지점별 편차가 크게 나타났다. 이는 훈련기간이 1년으로 비교적 짧아 학습된 국지 기후 경향과 다른 경향이 검증 기간에 나타났기 때문에 지점별로 편차가 크게 나타난 것으로 판단된다. 경험적 방법이 적용된 NHRH과 CART/SLD 모델의 상관계수는 여름과 겨울에 지점별 편차가 크게 나타났다. 이는 Hong *et al.*(2006)이 언급했듯이 한라산과 같은 높은 지형은 동서 또는 남북으로 지역별 기후가 다르게 나타날 수 있음에 기인한 것으로 판단된다. 따라서 각 지점들은 여름과 겨울에 발달하는 기단들로 인해 한라산을 지나면서 뒀 현상이 발생되고 이로 인

해 온도와 습도가 지점별로 다른 분포를 나타낸다. NHRH와 CART/SLD 모델의 RMSE와 MAE는 여름 오차가 가장 크고 겨울 오차가 가장 작은 경향으로 나타났다. 여름철 오차가 큰 이유는 제주도의 다습한 기후 특성으로 높은 습도 사례가 많이 나타나서 습도에 민감한 두 모델이 이슬로 오판한 사례가 많았기 때문이다. 그리고 PM과 DNN 모델의 RMSE와 MAE는 겨울을 제외하고 계절별 평균값의 차이가 다른 모델처럼 계절별 차이가 크지 않았다. PM 모델은 두 경험 모델과는 다르게 태양복사에너지가 가장 크게 영향을 미친다. 따라서, 태양복사에너지는 계절별로 차이가 크지 않기 때문에 계절별 차이가 NHRH, CART/SLD 모델 보다 편차가 작게 나타난 것이다. DNN 모델은 훈련기간에서 계절별 국지 기후 특성이





**Fig. 4.** Box plots of spring, summer, fall, winter with NHRH, CART/SLD, PM, and DNN models at 11 sites. Dashed line is median of data and dotted line is mean of data. The NHRH, CART/SLD, PM, DNN models were analyzed in 2017.

잘 적용된 결과라고 해석할 수 있다. 또한 겨울이 다른 계절에 비해 RMSE와 MAE 오차가 작은 것은 계절 특성상 이슬 사례가 많지 않았고 이슬이 발생하더라도 지속시간이 짧았기 때문이다. Bias는 모든 모델에서 양(+)의 방향으로 과대모의 하였고 NHRH와 CART/SLD 모델의 계절별 Bias는 RMSE와 MAE와 같은 경향으로 나타났다. PM 모델의 Bias는 여름에 가장 높고 봄에 가장 낮게 나타났다. DNN 모델의 Bias는 다른 모델에 비해 상대적으로 계절별 Bias가 큰 차이 없이 나타났으며 평균 2시간 미만으로 양의 Bias가 나타났다.

### 3.3. 지점별 모델 평가

11개 지점에서 관측한 이슬지속시간과 각 모델에서

예측한 이슬지속시간을 일단위로 비교한 결과는 Table 2와 같다. 4개 모델의 상관계수 평균은 PM 모델이 0.37로 가장 낮게 나타났고 DNN 모델은 0.57로 가장 높게 나타났다. NHRH, CART/SLD, 그리고 PM 모델은 2번과 4번 지점에서 가장 낮은 상관성을 보였다. 2번과 4번 지점은 제주도 남쪽에 위치하며 이 지역은 제주도에서 가장 다습한 지역이다. 습도 인자의 영향을 많이 받는 NHRH와 CART/SLD 모델은 제주 남쪽의 국지 기상 특성을 반영하지 못하였고, PM 모델은 다습한 지역의 예측 정확도가 상당히 낮은 것을 확인할 수 있었다. DNN 모델에서 0.5 미만의 상관계수 지점들은 2017년의 국지적 기상 인자들의 경향이

**Table 2.** Comparison among leaf wetness duration: observed and simulated with NHRH, CART/SLD, PM, and DNN models in 2017

Sites	NHRH			CART/SLD			PM			DNN						
	R	RMSE	MAE	Bias	R	RMSE	MAE	Bias	R	RMSE	MAE	Bias	R	RMSE	MAE	Bias
1 Deokcheon-dong	0.44	4.75	2.6	1.91	0.42	5.70	3.47	3.02	0.41	4.94	2.73	2.21	0.45	2.82	1.07	-0.64
2 Donghong-dong	0.37	5.13	2.63	2.61	0.33	6.16	3.55	3.54	0.16	5.32	2.51	2.43	0.39	1.38	0.34	0.12
3 Gasi-ri	0.74	5.47	3.01	2.98	0.69	6.97	4.28	4.27	0.61	3.8	2.18	0.96	0.78	3.33	1.54	1.05
4 Hogeun-dong	0.29	5.7	2.81	2.78	0.22	8.03	4.96	4.94	0.18	4.30	2.44	2.38	0.58	0.90	0.18	0.04
5 Jeoji-ri	0.60	5.47	2.8	2.23	0.58	7.01	4.13	3.92	0.58	3.75	2.04	0.93	0.69	3.05	1.47	0.28
6 Nohyeong-dong	0.48	5.53	3.20	3.0	0.44	7.67	5.19	5.04	0.23	5.09	3.33	2.78	0.40	3.30	1.29	0.74
7 Ora-dong	0.54	5.23	2.53	2.51	0.48	7.39	4.34	4.34	0.33	4.00	1.99	1.78	0.43	2.65	0.99	-0.03
8 Samdal-ri	0.51	6.92	3.99	3.95	0.46	8.73	5.69	5.69	0.30	5.36	3.65	3.28	0.77	1.75	0.58	0.48
9 Sangye-dong	0.57	9.21	5.88	5.87	0.55	10.29	6.86	6.86	0.41	5.42	3.11	2.74	0.61	3.25	1.47	1.24
10 Sumang-ri	0.72	3.14	1.32	0.30	0.69	5.38	2.87	2.67	0.49	3.81	2	0.49	0.42	3.75	1.63	0.68
11 Susan-ri	0.42	7.02	3.95	3.88	0.37	8.90	5.66	5.66	0.35	14.1	2.28	1.87	0.78	2.49	0.87	-0.63
average	0.52	5.78	3.16	2.91	0.48	7.48	4.64	4.54	0.37	5.44	2.57	1.99	0.57	2.61	1.04	0.30

\*R=Correlation, RMSE=Root Mean Squared Error (hour), MAE=Mean Absolute Error (hour)

2016년의 훈련된 기상인자의 경향을 벗어났기 때문으로 판단된다. 이는 충분한 훈련기간을 확보할 수 있다면 지점별 상관성은 더욱 향상 될 수 있을 것으로 판단된다. 보통 모델 성능은 상관계수와 함께 RMSE, MAE, 그리고 Bias를 종합적으로 고려하여 판단하는데, 특히, 상관계수는 0.5이상일 때 통계적 분석이 유의하다고 판단한다.

Kim *et al.*(2005)와 Luo and Goudriaan(2004)은 이슬지속시간 예측 모델의 RMSE가 2.5 이하로 나타날 때 정확도가 매우 높다고 평가하였다. 4개 모델 중 DNN 모델은 RMSE 2.61로 앞선 연구 결과와 가장 근접한 RMSE를 가지고 있고 성능이 가장 우수하였다. NHRH와 PM 모델의 RMSE 평균은 각각 5.78, 5.44로 나타났고 MAE와 Bias 오차도 PM 모델이 약간 작게 나타났다. 또한 지점별 RMSE 편차는 NHRH가 PM 모델 보다 크게 나타나 지점 전체에 대한 모델 안정성은 PM 모델이 우수하였다. 이와 같이 오차 통계를 통해 살펴본 모델 성능은 DNN 모델이 가장 우수하였고 CART/SLD 모델의 정확도가 가장 낮게 나타났다. 그리고 NHRH와 PM 모델은 비슷한 성능으로 나타났다. 지점별 성능은 전체적으로 앞선 결과와 비슷한 경향으로 나타났지만 9번(수망)과 10번(수산) 지점은 예외적인 결과가 나타났다. 9번 지점에서는 DNN 모델이 가장 우수하게 나타났고 다음으로 PM 모델이 좋은 성능을 보였고 NHRH 모델이 CART/SLD 모델과 비슷한 성능을 보였다. 9번 지점은 이슬 사례일이 196일로 지점들 중 가장 많은 사례가 관찰되었다. 즉, DNN 모델은 훈련 사례가 많을수록 모델 성능이 상승하기 때문에 DNN 모형이 가장 우수한 결과가 나타났다. NHRH 모델의 경우, 높은 습도 사례는 많았음에도 불구하고 실제 이슬이 발생하지 않았기 때문에 모델의 성능은 나쁘게 나타났다. 10번 지점은 NHRH 모델의 상관계수가 높고 RMSE가 가장 낮게 나타났고, 다른 지점에서 가장 좋은 성능을 보였던 DNN 모델이 낮은 상관계수와 가장 높은 RMSE가 나타났다. 10번 지점은 이슬 발생 사례일이 100일로 다른 지점에 비해 이슬 사례가 적게 나타났으며 타 지점에 비해 습도가 높게 나타난 사례가 많지 않았기 때문에 NHRH 모델의 적중률이 높게 나타났다.

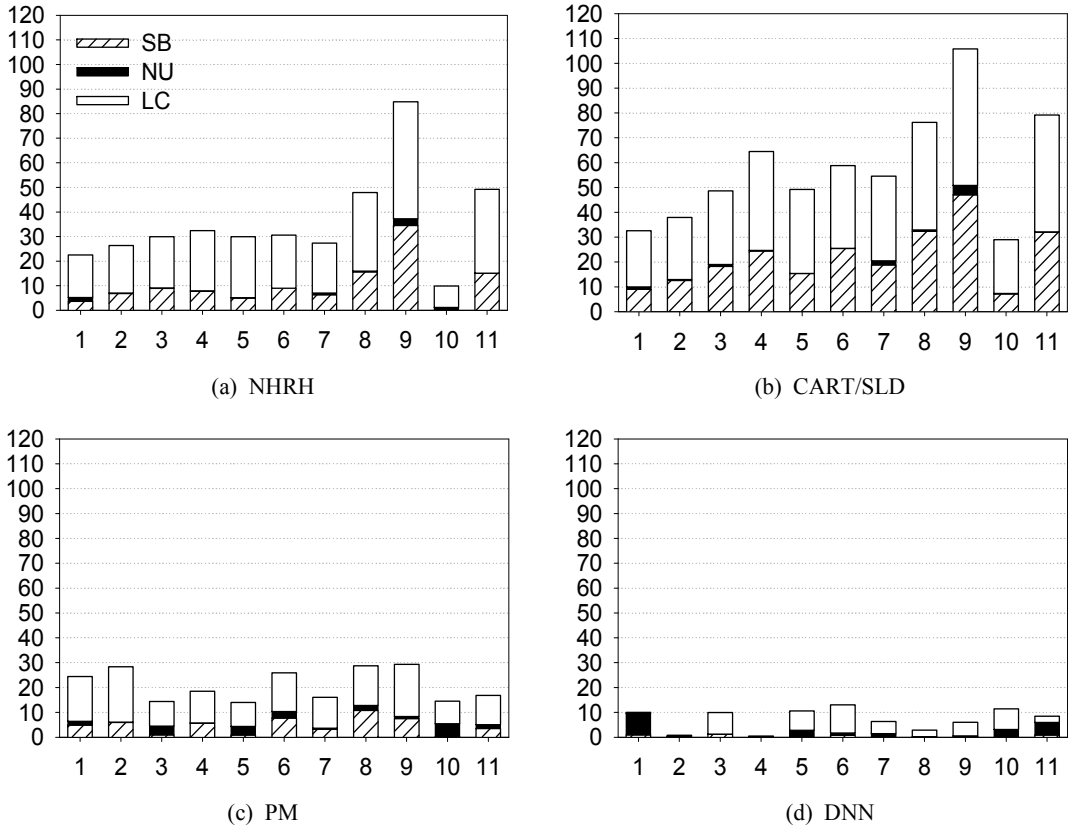
모델 오차를 통계적으로 분석하여 모델 예측 정확도를 평가하는 방법은 가장 흔히 사용되는 방법이지만 NHRH와 PM 모델처럼 비슷한 결과에서는 최종 모델

선택이 쉽지 않다. 이를 보완하기 위해 추가적으로 MSD 구성 성분 분석 방법을 통해 최종 모델 성능을 비교해 보았다.

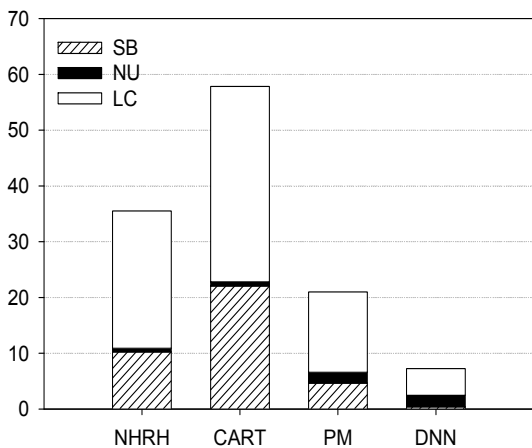
### 3.4. MSD 모델 평가

주로 모델과 관측의 정확도 평가에서는 1:1 대응과 같은 선형 회귀 통계 방법을 사용한다. 이슬은 기상 변수에 종속적인 관계를 가지고 있기 때문에 기상변수를 사용한 이슬 모델의 선형 회귀는 절편이 0이고 기울기가 1로 나타남을 기대한다. Fig. 5는 이슬지속시간 모델에 대한 지점별 MSD의 SB, NU, 그리고 LC 구성요소를 보여준다. 각 모델의 지점들은 대부분 LC가 가장 크고 다음으로 SB가 많은 비중을 차지하였으며 NU는 모델에 따라 차지하는 비중이 다르게 나타났다. 10번 지점을 제외하고 모든 지점의 성능은 DNN이 가장 우수하였고 다음으로 PM, NHRH, CART/SLD 순으로 나타났다. PM과 DNN 모델은 경험 모델 보다 NU의 비중이 높게 나타난다. 모델 성능 평가 시, NU와 SB는 예측과 관측 사이의 오류 형태(error type or false alarm) 정보를 포함한다. SB는 오류 형태가 긍정 오류(false positive)나 부정 오류(false negative)의 한쪽 방향으로만 몰려 있는 경우에 해당되고 NU는 긍정 오류와 부정 오류 양쪽 모두 존재하는 경우를 의미한다. 그리고 긍정 오류와 부정 오류의 사례가 많을수록 SB와 NU는 커지게 된다. NHRH와 CART/SLD 모델은 대부분의 지점에서 긍정 오류의 사례가 많이 나타났기 때문에 SB가 대부분 많은 비중을 차지한다. NU가 많은 비중을 차지하는 PM과 DNN 모델은 긍정 오류와 부정 오류 사례가 모두 많이 나타났기 때문에 NU가 크게 나타났다. 이때 DNN 모델은 이슬 발생의 학습 사례를 더 많이 추가할 수 있다면 NU 오차는 감소할 수 있다.

앞서 11개 지점별로 구한 MSD 구성요소를 최종적으로 평균하여 비교해 보았다(Fig. 6). MSD는 DNN이 가장 낮았고 PM, NHRH, CART/SLD 순으로 MSD 값이 크게 나타났다. 앞서 다소 애매 했던 NHRH와 PM 모델 성능은 MSD 분석 결과 성능이 명확하게 구별된다. 경험적 모델과 물리적 모델은 MSD에 기여하는 구성요소가 서로 다르게 나타났다. 경험 모델인 NHRH와 CART/SLD은 SB와 LC가 주로 나타난 반면, 물리 모델 PM과 기계학습 모델 DNN에서는 LC가 높게 나타났고 SB와 NU가 비슷한 수준으로 나타났다. LC가 4개 모델에서 가장 큰 부분을 차지하는 것은 모델이 본래 가지고 있는 구조적인 불확실성으로 판단



**Fig. 5.** Components of MSD in LC, NU, and SB for NHRH, CART/SLD, PM, and DNN models at 11 sites of Jeju in 2017.



**Fig. 6.** Components of MSD average in LC, NU, and SB for NHRH, CART/SLD, PM, and DNN models at Jeju in 2017.

된다. 모든 모델에서 3가지 구성 요소 중 LC가 차지하는 비중이 가장 크지만 크기는 CART/SLD, NHRH, PM, DNN으로 갈수록 낮아졌고 SB 역시 LC와 같은 경향으로 나타났다.

#### IV. 고 찰

본 연구는 제주 감귤 과수원에서 발생하는 최적의 이슬지속시간 예측 모델을 선정하고자 4가지(NHRH, CART/SLD, PM, DNN) 모델에 대한 성능을 비교 및 평가하였다. 평가 방법은 통계적 오차를 나타내는 Correlation, RMSE, MAE, Bias, 그리고 MSD의 구성 요소로 비교 분석하였다. 그 중 DNN 모델이 가장 낮은 오차 수치로 통계적으로 유의성이 높게 나타났으며, CART/SLD 모델이 가장 높은 오차 수치를 보여 통계적 유의성이 낮게 나타났다. NHRH와 PM 모델은 오차 값들 기준으로 PM 모델이 좋은 성능을 보였지만,

상관계수는 NHRH 모델이 더 좋게 나타났다. 향후 두 모델 성능의 개선 가능성을 살펴본다면 MSD의 NU 요소가 작은 NHRH 모델이 개선 여지가 크다. MSD의 LC와 NU 요소를 줄이기 위해서는 모델과 관측 자체적으로 가지는 오류를 수정하는 것이 필요하고 MSD의 SB 요소는 모델 내의 Bias 감소 요소를 추가하여 최적의 모델 상태로 조정할 수 있다. DNN 모델은 4가지 모델 중에서 최상의 성능을 가지며 학습 자료를 많이 확보할수록 모델의 정확성을 더욱 개선할 수 있다. 이때 기계학습 모델은 최적화 과정에서 많은 시간이 소모되고 최적 구조를 찾기 위해서 다양한 방법이 시도된다. 차후 추가적인 훈련 자료를 확보하고 다양한 내부 알고리즘 변화를 적용한다면 DNN 특성상 보다 정확도가 향상될 수 있을 것이다.

## V. 요약 및 결론

이슬은 크게 새벽에 발생하는 이슬과 강우로 인해 발생하는 이슬로 나눌 수 있다. 이 둘의 발생 매커니즘은 다르기 때문에 모델에서 강우일을 처리하는 방법이 중요하다. 강우일을 포함한 모델에 비해 강우일을 제외한 각 모델의 RMSE, MAE, 그리고 Bias는 대부분 감소하여 나타나 모델 정확도가 증가하였다. 이러한 결과로 강우와 이슬 매커니즘을 구분하여 모델 개발에 적용하는 것이 모델 정확도를 향상할 수 있다. 또한 NHRH, CART/SLD, 그리고 PM 모델은 여름에 가장 큰 오차가 발생하였고 겨울에 오차가 가장 작게 나타났다. 겨울은 이슬 사례가 많지 않았기 때문에 나타난 현상이며 여름 오차가 큰 것은 제주도 여름이 기후학적으로 다습한 환경을 가지고 있어 지점 별로 다른 기온과 습도가 나타났기 때문이다. DNN 모델을 제외한 3개의 모델은 이슬이 집중되는 봄과 가을에 예측 정확도가 비교적 비슷하게 나타났으나 여름의 낮은 예측 정확도가 모델 전체 예측 정확도를 낮게 하는 원인이 되었다.

모델에서 흔히 사용되는 RMSE, MAE, 그리고 Bias와 같은 통계적 오차 방법으로 살펴본 모델 성능은 DNN이 가장 우수하였다. CART/SLD의 예측 정확도가 가장 낮게 나타났고 NHRH와 PM 모델은 비슷한 성능을 보였다. 지점별로 다른 경향을 보인 경우도 있었지만 전체적으로 성능 순위는 DNN, NHRH 또는 PM, CART/SLD로 나타났다. 모델의 성능을 더욱 세세히 분류하고 싶다면 MSD 방법을 이용할 수 있다.

MSD는 SB, NU, 그리고 LC 구성요소로 구분되며 이를 이용한 모델 성능은 DNN이 가장 우수하였으며, 다음으로 PM, NHRH, CART/SLD 순으로 성능이 나타났다. 모델 성능이 우수할수록 SB와 LC는 감소하였지만 NU는 증가하는 경향이 나타났다. 이때 NU는 긍정 오류와 부정 오류가 동시에 나타날 때 증가하였고 한쪽 방향의 긍정 오류나 부정 오류가 발생한다면 SB 성분이 증가하는 것으로 나타났다.

이처럼 기존에 개발된 경험적 또는 물리적 모델 보다 기계학습 모델은 지역적 특징을 최적으로 적용하여 가장 우수한 성능을 나타낸다. 하지만 최근 DNN 모델은 Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short Term Memory network (LSTM) 등의 다양한 알고리즘이 빠르게 발전하고 있기 때문에 목적에 맞는 학습 방법 선택이 중요하다. 따라서 목적에 맞게 기계학습 모델을 선택하고 다양한 방법을 비교하는 실험이 중요하다. 추후 추가연구를 통해 제주 지역에 최적화된 이슬지속시간 예측 모델을 개발한다면 향후 병충해 예방과 농업 생산량 지원 시스템에서 유용하게 활용할 수 있을 것으로 기대된다.

## 적 요

2016년부터 2017년까지 제주 감귤과수원 11개 지점에서 관측한 기상 및 이슬 자료를 이용하여 이슬지속시간 예측 모델을 평가하였다. 이슬지속시간 모델의 민감도와 예측 정확도 평가에는 4 가지 모델(Number of Hours of Relative Humidity, Classification And Regression Tree/Stepwise Linear Discriminant, Penman-Monteith, Deep-learning Neural Network)이 사용되었다. 모델의 민감도는 강우와 계절 변화에 따라 평가하였다. 전체 자료에서 강우일 자료를 제외하면 이슬지속시간 모델들은 평균 오차(평균제곱근오차 약 1.5 hours)가 적게 나타났다. 기계학습 모델은 겨울을 제외한 계절별 오차가 비슷한 크기(평균제곱근오차 약 3 hours)로 나타났다. 나머지 모델들은 여름에 오차(평균제곱근오차 약 9.6 hours)가 가장 크고 겨울에 가장 작은 것(평균제곱근오차 약 3.3 hours)으로 나타났다. 모델 예측 정확도 평가 방법은 통계적 오차 분석 방법과 평균 제곱 편차 회귀 분석 방법을 사용하였다. 통계 오차를 통한 모델 성능은 DNN 모델이 가장 우수한 반면에 CART/SLD 모델은 예측 정확도가 가장 낮게

나타났다. 평균제곱 편차(MSD)는 모델의 선형성을 세 가지(제곱 바이어스(SB), 비균일성 기율기(NU), 상관관계 부족(LC)) 구성요소로 구분하여 분석하는 방법이다. 모델 성능이 우수할수록 SB와 LC는 감소하였고 NU는 증가하는 경향이 나타났다. MSD 분석 결과 DNN 모델이 가장 우수하였으며 다음으로 PM, NHRH, CART/SLD 순으로 나타났다. 본 연구에서 활용된 기계학습 모델은 기상 정보를 이용한 다른 농업 정보 생산의 정확도 개선에 크게 기여할 것으로 판단된다.

### 감사의 글

본 논문은 기상청 국립기상과학원 『기상업무지원 기술개발연구』 “생명산업기상기술개발(KMA2018-00621)”의 지원으로 수행되었습니다. 이슬 자료 수집에 협조해주신 제주특별자치도 농업기술원에 감사드립니다.

### REFERENCES

- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith, 1998: Crop evapotranspiration: Guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56, Rome, Italy: United Nations FAO.
- Beysens, D., 2016: Estimating dew yield worldwide from a few meteo data. *Atmospheric Research* **167**, 146-155.
- Beysens, D., M. Muselli, V. Nikolayev, R. Narhe, and I. Milimouk, 2005: Measurement and modeling of dew in island coastal and alpine areas. *Atmospheric Research* **73**, 1-22.
- Cosh, M. H., E. D. Kabela, B. Hornbuckle, M. L. Gleason, T. J. Jackson, and J. H. Prueger, 2009: Observation of dew amount using in situ and satellite measurements in an agricultural landscape. *Agricultural and Forest Meteorology* **149**, 1082-1086.
- Gauch, H. G., J. T. Hwang, and G. W. Fick, 2003: Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal* **95**, 1442-1446.
- Gleason, M. L., S. E. Taylor, T. M. Loughin, and K. J. Koehler, 1994: Development and validation of an empirical model to estimate the duration of dew periods. *Plant disease* **78**, 1011-1016.
- Gleason, M. L., 2007: Validation of weather inputs for disease warning systems. *American Phytopathological Society* **97**(7), 147.
- Hinton, G. E., S. Osindero, and Y. W. Teh, 2006: A fast learning algorithm for deep belief nets. *Neural computation* **18**(7), 1527-1554.
- Hong, K. O., M. S. Suh, and D. K. Rha, 2006: Temporal and spatial variations of precipitation in South Korea for recent 30 years (1976-2005) and geographic environments. *Journal of Korean Earth Sciences Society* **27**(4), 433-449.
- Kim, K. S., S. E. Taylor, M. L. Gleason, R. Villalobos, and L. F. Arauz, 2005: Estimation of leaf wetness duration using empirical models in northwestern Costa Rica. *Agricultural and Forest Meteorology* **129**, 53-67.
- Kim, K. S., S. E. Taylor, M. L. Gleason, and K. J. Koehler, 2002: Model to enhance site-specific estimation of leaf wetness duration. *Plant disease* **86**, 179-185.
- Kim, S. S., S. M. Jang, H. J. Baek, H. Y. Choi, and W. T. Kwon, 2006: Climatological variability of temperature and precipitation in Jeju. *Journal of Korean Earth Sciences Society* **27**(2), 188-197.
- Klemm, O., C. Milford, M. A. Sutton, G. Spindler, and E. Van Putten, 2002: A climatology of leaf surface wetness. *Theoretical and Applied Climatology* **71**, 107-117.
- LeCun, Y., Y. Bengio, and G. Hinton, 2015: Deep learning. *nature* **521**(7553), 436-444.
- Lee, H., J. B. Jee, J. S. Min, S. Kim, and J. H. Chae, 2018: Analysis of meteorological and radiation characteristics using WISE observation data. *Journal of Korean Earth Sciences Society* **39**(1), 89-102.
- Lee, Y. J., and J. Nang, 2016: A personal video event classification method based on multi-modalities by DNN-Learning. *Journal of KIISE* **43**(11), 1281-1297.
- Luo, W., and J. Goudriaan, 2004: Estimating dew formation in rice, using seasonally averaged diel patterns of weather variables. *NJAS wageningen journal of life sciences* **51**, 391-406.
- Maestre-Valero, J. F., R. Ragab, V. Martinez-Alvarez, and A. Baille, 2012: Estimation of dew yield from radiative condensers by means of an energy balance model. *Journal of Hydrology* **460**, 103-109.
- Park, J. H., D. H. Shin, and C. B. Kim, 2017: Deep learning model for electric power demand prediction using special day separation and prediction elements extension. *Journal of Advanced Navigation Technology* **21**(4), 365-370.

- Park, J. S., K. R. Kim, M. Kang, and B. J. Kim, 2017: The influence of evaporation from a stream on fog events in the middle Nakdong River. *Journal of Korean Earth Sciences Society* **38**(6), 395-404.
- Rowlandson, T., M. Gleason, P. Sentelhas, T. Gillespie, C. Thomas, and B. Hornbuckle, 2015: Reconsidering leaf wetness duration determination for plant disease management. *Plant Disease* **99**, 310-319.
- Sentelhas, P. C., T. J. Gillespie, M. L. Gleason, J. E. B. Monteiro, J. R. M. Pezzopane, and M. J. Pedro, 2006: Evaluation of a Penman-Monteith approach to provide "reference" and crop canopy leaf wetness duration estimates. *Agricultural and Forest Meteorology* **141**, 105-117.
- Sentelhas, P. C., A. Dalla Marta, S. Orlandini, E. A. Santos, T. J. Gillespie, and M. L. Gleason, 2008: Suitability of relative humidity as an estimator of leaf wetness duration. *Agricultural and forest meteorology* **148**, 392-400.
- Shin, D. H., and C. B. Kim, 2018: Short term forecast model for solar power generation using RNN-LSTM. *Journal of Advanced Navigation Technology* **22**(3), 233-239. (in Korean)
- Montone, V. O., C. W. Fraisse, N. A. Peres, P. C. Sentelhas, M. Gleason, M. Ellis, and G. Schnabel, 2016: Evaluation of leaf wetness duration models for operational use in strawberry disease-warning systems in four US states. *International journal of biometeorology* **60**, 1761-1774.
- Veronica, N. S., W. F. Clyde, A. P. Natalia, C. C. James, and C. Amy, 2010: Spatial variability of leaf wetness duration in citrus canopies. *Proceedings of the Florida State Horticultural Society* **123**, 49-55.
- Wallach, D., and B. Goffinet, 1989: Mean squared error of prediction as a criterion for evaluating and comparing system models. *Ecological Modeling* **44**, 299-306.
- Wieland, R., W. Mirschel, K. Groth, A. Pechenick, and K. Fukuda, 2011: A new method for semi-automatic fuzzy training and its application in environmental modeling. *Environmental modeling & software* **26**, 1568-1573.