

Hybrid SARIMAX-LSTM Forecasting of Hourly Water Reduce Ratio with ROC-Based Risk Threshold Optimization

Joon-Yong Kim*

*Professor, Dept. of AI Convergence, Seoul Theological University, Gyeonggi-Do, Korea

[Abstract]

Hourly water reduce ratio (WRR), defined here as the hourly rate of change in relative humidity, is used as an environmental proxy for rapid moisture-condition change relevant to fresh-concrete quality management. This study proposes a one-hour-ahead SARIMAX-LSTM hybrid forecasting framework in which SARIMAX models seasonal-linear dynamics and lagged weather effects, and LSTM corrects nonlinear residual components. The analysis uses 1,488 hourly observations from station ID 1189 between July 1 and September 1, 2021. Timestamp alignment, duplicate handling, missing-data treatment, and outlier auditing were explicitly specified. Wind speed was evaluated as a candidate exogenous variable, but it was not retained because it did not improve validation RMSE and MAE consistently. The final hybrid model achieved RMSE=4.413 and MAE=3.117 on the test set, representing numerical reductions of 1.66-5.99% in RMSE and 2.42-2.59% in MAE relative to the baselines. However, Diebold-Mariano tests with Holm correction and moving-block bootstrap confidence intervals did not establish statistical significance at the 0.05 level. ROC-based threshold optimization was conducted on the validation set using Youden's J and F1-score. Under $WRR \leq -5\%$, the selected alarm threshold was $-0.479\%/h$ and yielded Recall=0.926, Precision=0.210, F1=0.342, and ROC-AUC=0.788 on the test set. The framework is therefore suitable for early warning and secondary inspection support, not for direct diagnosis of concrete strength loss.

► **Key words:** Water Reduce Ratio, Time-Series Forecasting, SARIMAX, LSTM, Hybrid Model, ROC, Youden's J, Threshold Optimization

[요 약]

본 연구는 시간당 상대습도 변화율로 정의한 감수율(WRR)을 초기 재령 콘크리트의 수분 환경 급변을 간접적으로 반영하는 환경 위험지표로 활용한다. 제안 모형은 계절성과 지연 기상변수의 선형 효과를 설명하는 SARIMAX와 비선형 잔차를 보정하는 LSTM을 결합한 1시간 선행 예측 구조이다. 2021년 7월 1일부터 9월 1일까지 관측소 ID 1189에서 수집된 시간 단위 자료 1,488건을 분석하였으며, 시간 정렬, 중복 제거, 결측치 처리, 극단값 점검 절차를 명시하였다. 풍속은 후보 외생변수로 평가하였으나 검증 RMSE와 MAE를 일관되게 개선하지 못하여 최종 입력에서 제외하였다. 최종 하이브리드 모형은 테스트셋에서 RMSE=4.413, MAE=3.117을 기록하여 비교모형 대비 RMSE를 1.66~5.99%, MAE를 2.42~2.59% 수치적으로 감소시켰다. 다만 Diebold-Mariano 검정, Holm 보정 및 moving-block bootstrap 결과 0.05 수준의 통계적 유의성은 확인되지 않아 성능 우위는 탐색적 결과로 해석하였다. ROC 기반 Youden's J와 F1 최적화를 통해 WRR $\leq -5\%$ 기준의 경보 임계값을 $-0.479\%/h$ 로 선정하였으며, 테스트셋에서 Recall=0.926, Precision=0.210, F1=0.342, ROC-AUC=0.788을 보였다. 따라서 본 모형은 압축강도 저하의 직접 진단이 아니라 추가 현장 점검과 품질시험을 유도하는 조기경보 도구로 활용될 수 있다.

▶ **주제어:** 감수율, 시계열 예측, SARIMAX, LSTM, 하이브리드 모델, ROC, Youden's J, 임계값 최적화

I. Introduction

콘크리트의 초기 재령 단계에서 온도, 상대습도, 풍속과 같은 환경 요인은 표면 수분 증발과 양생 조건에 영향을 미친다[1-3]. 그러나 현장에서는 압축강도를 시간 단위로 직접 측정하기 어려워 현장에서 제작한 모델을 실험실에서 시간경과에 따라 시험하는 방식에 의존하는 경우가 많다. 그러므로 실제 현장에서 이뤄지는 양생환경의 위험을 조기에 탐지하기 위해 기상 기반 지표를 활용한 예측-경보 체계가 필요하다.

기존 연구는 시간대별 상대습도 변화율을 감수율(WRR)로 정의하고, 기온과 습도 변화에 따른 위험 구간을 계절형 시계열 모형으로 분석하였다[1]. 다만 원시자료의 시간 정렬, 중복 및 결측 처리, 풍속 변수의 사용 여부, 비선형 잔차 보정, 경보 임계값의 최적화 절차가 충분히 기술되지 않았다.

본 연구는 SARIMAX가 시간대별 계절성, 자기상관 및 지연 기상변수의 선형 효과를 먼저 설명하고, LSTM이 남은 잔치의 비선형 변화를 보정하는 잔차 학습형 하이브리드 구조를 제안한다[4-7]. 또한 최근 하이브리드 시계열 연구의 구조적 장점을 반영하되[8-10], 데이터 품질관리, 외생변수 제거(ablation), 시계열 의존성을 고려한 성능 비교 및 ROC 기반 임계값 최적화를 함께 제시한다.

본 연구의 기여는 다음과 같다. 첫째, WRR의 산출식과 시간 정렬, 결측치, 중복 및 이상치 처리 절차를 재현 가능하게 명시하였다. 둘째, SARIMAX 차수 및 LSTM 하이퍼

파라미터를 시계열 순차 검증으로 선정하였다. 셋째, 풍속의 물리적 관련성을 인정하되, 관측소 풍속의 추가 예측 기여도를 ablation 분석으로 검증하였다. 넷째, 모델 간 수치적 성능 차이와 통계적 유의성을 구분하기 위해 Diebold-Mariano 검정과 moving-block bootstrap을 적용하였다. 다섯째, 위험 이벤트 기준과 운영 경보 임계값을 분리하고 Youden's J 및 F1-score로 경보 기준을 보정하였다.

II. Materials and Methods

1. Data Source, WRR Definition, and Quality Control

원시자료는 기상관측소 ID 1189의 2021년 7월 시간 단위 자료 746건과 2021년 8월 5분 단위 자료 8,927건으로 구성된다. 분석 패널은 2021년 7월 1일 01:00부터 2021년 9월 1일 00:00까지의 1,488개 시간 단위 WRR 관측값이다. WRR 산출을 위해 2021년 7월 1일 00:00의 상대습도를 기준시점으로 추가 사용하였다.

모든 관측시각은 한국표준시(KST)로 통일한 뒤, 가장 가까운 정각으로 반올림하지 않고 해당 시각이 포함된 [h:00:00, h:59:59] 시간 구간에 배정하였다. 7월 자료는 8월 1일 00:00 이전에, 8월 5분 자료는 8월 1일 00:00 이후

에 우선 적용하였다. 동일 시간 구간에 복수의 기록이 존재하면 기온, 상대습도 및 풍속의 중앙값을 대표값으로 사용하였다. 이 절차는 관측값이 인접 시간대로 이동하는 반올림 왜곡과 순간 센서값의 영향을 줄이기 위한 것이다.

$$WRR_t(\%) = \left(\frac{RH_t - RH_{t-1}}{RH_{t-1}} \right) \times 100 \quad (1)$$

위의 식 (1)에서 RH_t 는 시점 t 의 상대습도(%), RH_{t-1} 은 직전 시간대의 상대습도(%)이다. 따라서 WRR의 단위는 시간당 상대습도 변화율(%/h)이다. WRR은 콘크리트 내부 수분량이나 압축강도의 직접 측정값이 아니라, 환경 변화에 대한 대리 지표로 해석한다. 학습-검증-테스트 자료는 미래 정보의 유입을 방지하기 위해 시간 순서를 유지하여 분할하였으며, 시계열 예측에서는 무작위 교차검증보다 시간 순서를 보존하는 검증 절차가 타당하므로 expanding-window기반의 검증 원칙을 적용하였다[11].

다음의 표 1은 원시자료가 최종 분석 패널로 변환되는 절차를 요약한 것이다. 주요 이슈는 비정각 관측값을 반올림하지 않고, 해당 시간 구간에 배정된 점, 동시간대의 복수 기록을 중앙값으로 집계한 점, 그리고 WRR 자체는 보간하지 않았다는 점이다. 이 절차를 통해 시간 누수와 인위적 WRR 변형을 최소화 하였다.

다음의 표 2는 전처리 후 사용한 1,488개의 시간 단위 자료 분포이다. WRR의 중앙값은 0.065%/h로 0에 가깝지만, 최소값 -25.019와 최대값 61.977%/h의 범위가 넓어 단순 평균 기반 모형만으로는 급격한 변화 구간을 충분히 설명하기 어려움을 보여준다. 기온과 상대습도는 여름철 관측소 자료에 부합하는 범위를 보이며, 풍속은 상대적으로 좁은 범위에 분포한다.

그림 1은 원시 기상자료에서 시작하여 시간 단위 정렬 및 품질관리, SARIMAX의 계절-선형 예측, LSTM의 잔차보정, 최종 WRR 예측, ROC 기반 경보 임계값 최적화로

Table 1. Data quality-control protocol

Stage	Procedure	Result
Analysis panel	2021-07-01 01:00 to 2021-09-01 00:00	1,488 hourly WRR observations
Temporal alignment	Floor assignment to hourly bins	No nearest-hour rounding
Duplicate handling	Remove exact duplicate; median for multiple values	2 duplicated July bins resolved
Missing exogenous data	Linear interpolation only for gaps ≤ 2 h	4 hourly bins; 14 variable values
Long gaps / incomplete windows	Exclude from model-window generation	No gap > 2 h after screening
Data split	Chronological 70:15:15 split	Train=1,041; Val=223; Test=224
Scaling	Fit scaler on training data only	No validation/test leakage

Table 2. Descriptive statistics after reprocessing

Variable	Unit	N	Mean	Std.	Min.	25%	Median	75%	Max.
WRR	%/h	1,488	0.247	7.089	-25.019	-3.276	0.065	2.839	61.977
Temperature	°C	1,488	27.649	3.566	18.350	25.100	27.500	30.063	37.400
Relative humidity	%	1,488	67.802	14.386	30.900	56.800	67.100	79.200	97.500
Wind speed	m/s	1,488	1.342	0.380	0.600	1.100	1.300	1.600	3.000

Overall Framework of the Proposed SARIMAX-LSTM Forecasting and Warning System

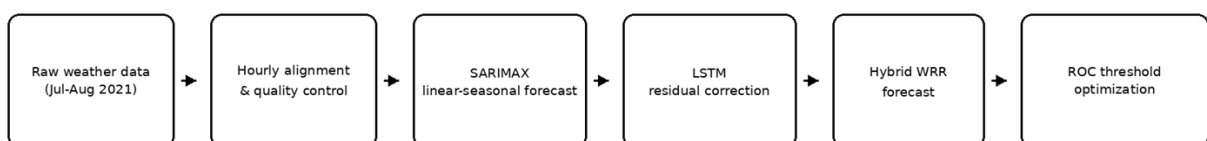


Fig. 1. Overall framework of the proposed hybrid forecasting and warning system

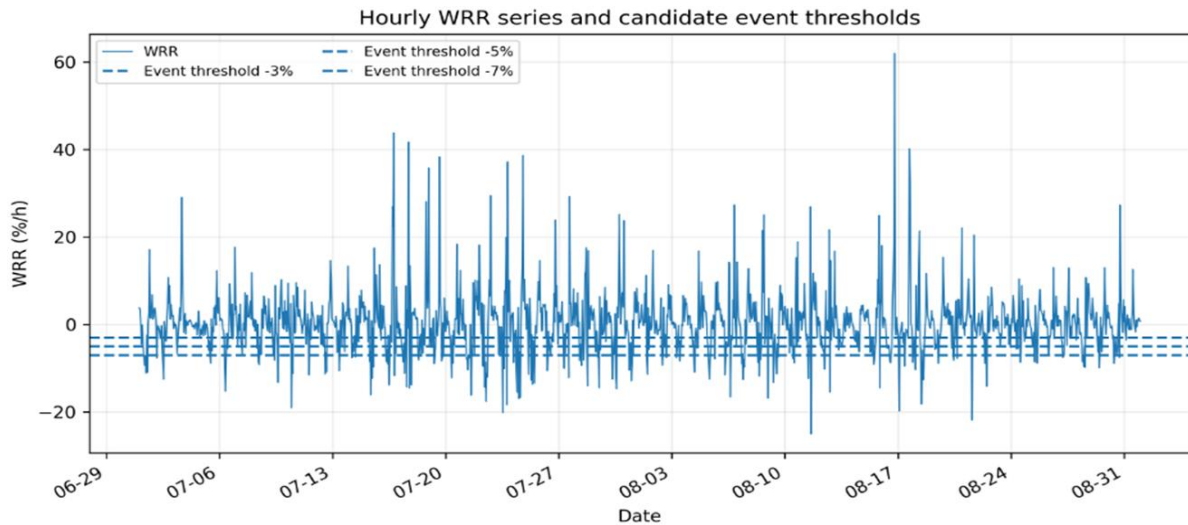


Fig. 2. Hourly WRR series and candidate event threshold

Table 3. Pearson correlation coefficients

Variable	WRR	Temperature	Relative humidity	Wind speed
WRR	1.000	-0.079	0.137	-0.083
Temperature	-0.079	1.000	-0.832	0.320
Relative humidity	0.137	-0.832	1.000	-0.486
Wind speed	-0.083	0.320	-0.486	1.000

이어지는 전체 분석 흐름을 나타낸다. 이 구조는 예측 성능 평가와 경보 운영 기준의 결정을 분리함으로써 데이터 누수를 줄이고 현장 적용성을 높인다.

그림 2는 전체 분석 기간의 시간당 WRR 변동과 세 가지 후보 이벤트 기준 (-3%, -5%, -7%)을 함께 표시한 것이다. WRR은 대부분 0 부근에서 변동하지만 특정 시간대에 큰 양과 음의 급변이 나타난다. 수평 기준선은 이벤트 정의의 민감도를 비교하기 위한 것으로, -3%는 상대적으로 민감한 조기경보, -7%는 보수적인 고위험 기준, -5%는 중간 수준의 기본적 기준을 의미한다.

WRR의 최대값 61.977%/h는 2021년 8월 16일 19:00에 발생하였다. 직전 시간대 상대습도는 52.6%, 해당 시간대 상대습도는 85.2%였으며, 두 시간대는 각각 12개의 5분 원시기록으로 구성되었다. 원시자료는 별도의 품질 플래그 열이 없었으나, 상대습도는 0-100%의 물리적 범위를 벗어나지 않았고 결측중복 오류도 확인되지 않아 해당 관측값을 유지하였다.

2. Variable Assessment and Wind-Speed Ablation

위의 표 3은 변수간 동시점 Pearson 상관계수를 나타낸 것으로 WRR과 기온(-0.079), 상대습도(0.137), 풍속(-0.083)이 모두 약한 수준으로 확인되었다. 이는 WRR 예측에 시간지연, 계절성 및 비선형 상호작용을 고려할 필요가 있음을 시사하며 풍속은 콘크리트 표면 증발과 건조 조건에 물리적으로 중요함을 의미한다[2].

그러나 본 연구의 풍속은 일반 기상관측소의 시간 평균 값으로, 실제 타설면 풍속을 완전하게 대변하지 못할 수 있다. 따라서 풍속을 선형적으로 배제하지 않고, 기온-상대습도 모형과 기온-상대습도-풍속 모형을 동일한 검증 절차에서 비교하였다.

다음의 표 4는 풍속을 외생변수에 추가했을 때의 ablation분석결과이다. 풍속을 추가하면 MAE는 0.002 감소했으나 RMSE가 0.019 증가하였다. 두 지표를 동시에 개선하지 못했으므로, 본 자료에서는 기온과 상대습도만을 최종 외생 변수로 사용하였다. 이는 풍속의 물리적 비관련

Table 4. Validation-set wind-speed ablation analysis for the hybrid model

Exogenous variables	Validation RMSE	Validation MAE	Decision
Temperature + Relative humidity	7.190	4.138	Selected
Temperature + Relative humidity + Wind speed	7.209	4.136	Not retained: no consistent improvement

Table 5. SARIMAX order selection on the training set

Candidate SARIMAX order	BIC	Ljung-Box Q(24)	p-value	Decision
(0,0,0)(0,0,0)24	7064.337	671.451	<0.001	Rejected
(0,0,0)(1,0,1)24	6710.724	48.948	0.002	Rejected
(1,0,1)(1,0,1)24	6718.002	44.530	0.007	Rejected
(1,0,3)(1,0,1)24	6701.281	33.323	0.097	Selected
(2,0,0)(1,0,1)24	6707.414	33.844	0.087	Rejected

성을 의미하는 것이 아니라, 본 관측소 풍속의 추가 예측 기여도가 제한적이었음을 의미한다.

3. SARIMAX-LSTM Hybrid Model and Hyperparameter Selection

본 연구는 1시간 선행 예측을 수행한다. 즉, 시점 t 의 WRR을 예측할 때 시점 $t-1$ 까지 이용 가능한 WRR 및 기상변수만 사용하였다. 따라서 외생변수는 x_{t-1} 로 지연시켜 입력하였으며, 관측시점의 상대습도 RH_t를 직접 사용하는 시간적 누수를 방지하였다. 실제 운영에서는 x_{t-1} 대신 기상예보 자료를 입력할 수 있다.

$$\Phi_p(L^{24})\phi_p(L)(1-L)^d(1-L^{24})^D(y_t - \beta^T x_{t-1}) = \Theta_Q(L^{24})\theta_q(L)\epsilon_t \quad (2)$$

위의 식 (2)에서 L 은 시차 연산자, (p,d,q) 와 $(P,D,Q,24)$ 는 각각 비계절 및 계절 차수이며, x_{t-1} 은 시점 $t-1$ 에서 관측된 기온과 상대습도이다. SARIMAX 후보모형은 적합도와 복잡도를 함께 고려하는 정보기준으로 비교하였다. AIC는 Akaike가 제안한 정보 기준이며[12], BIC는 표본 크기가 다

른 추가 복잡도 패널티를 포함하는 Schwarz 기준이다[13]. 본 연구에서는 BIC가 낮은 후보를 우선 검토하고, Box-Pierce 및 Ljung-Box 검정을 통해 잔차에 남아 있는 자기상관을 진단하여 최종 차수를 선정하였다[14],[15].

표 5는 SARIMAX 차수의 BIC와 24 시차 Ljung-Box의 검정결과로 (1,0,3)(1,0,1)24 모형은 표의 후보 중 가장 낮은 BIC(6701.281)을 보였고, 잔차 자기상관 검정의 p-value가 0.097로 0.05보다 커 잔차에 유의한 자기상관이 남지 않는다고 판단하였다. 따라서 적합도와 잔차 진단을 함께 만족하는 최종 기준모형으로 선정하였다.

$$r_t = y_t - \hat{y}_t^S \quad (3)$$

$$\hat{y}_t^H = \hat{y}_t^S + \hat{r}_t \quad (4)$$

식 (3)에서 $\hat{y}_t^S$ 는 SARIMAX 예측값, r_t 는 SARIMAX 잔차, $\hat{r}_t$ 는 LSTM이 예측한 잔차, $\hat{y}_t^H$ 는 최종 하이브리드 예측값이다. LSTM은 전체 WRR을 직접 학습하는 대신 24시간의 잔차, SARIMAX 예측값 및 지연 기상변수 시퀀스를 입력받아 비선형 잔차를 보정한다.

그림 3은 SARIMAX 예측값과 실제 WRR의 차이인 잔차를 LSTM이 학습하고, 이 잔차 예측값을 SARIMAX 예

Residual-learning architecture of the proposed hybrid model

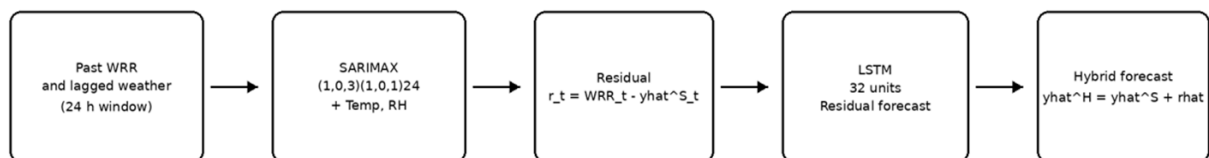


Fig. 3. Hourly WRR series and candidate event threshold:

yhat_S: SARIMAX forecast; r_t: SARIMAX residual; rhat: LSTM residual forecast; yhat_H: final hybrid forecast

Table 6. Hyperparameter search and final settings

Parameter	Candidate values	Validation criterion	Final setting
Input window	12, 24, 48 h	RMSE/MAE and parsimony	24 h
LSTM hidden units	8, 16, 32	Lowest validation RMSE	32
Learning rate	0.0005, 0.001, 0.003	Lowest validation RMSE	0.003
Dropout	0.0, 0.2, 0.3	Validation loss	0.20
Batch size	64, 128, 256	Training stability	256

Table 7. Window-length sensitivity of the hybrid model

Input window	Validation RMSE	Validation MAE	Interpretation
12 h	7.192	4.123	Shorter history: modestly higher RMSE
24 h	7.190	4.138	Selected: daily cycle and parsimonious input
48 h	7.190	4.137	Marginal difference: higher input dimension

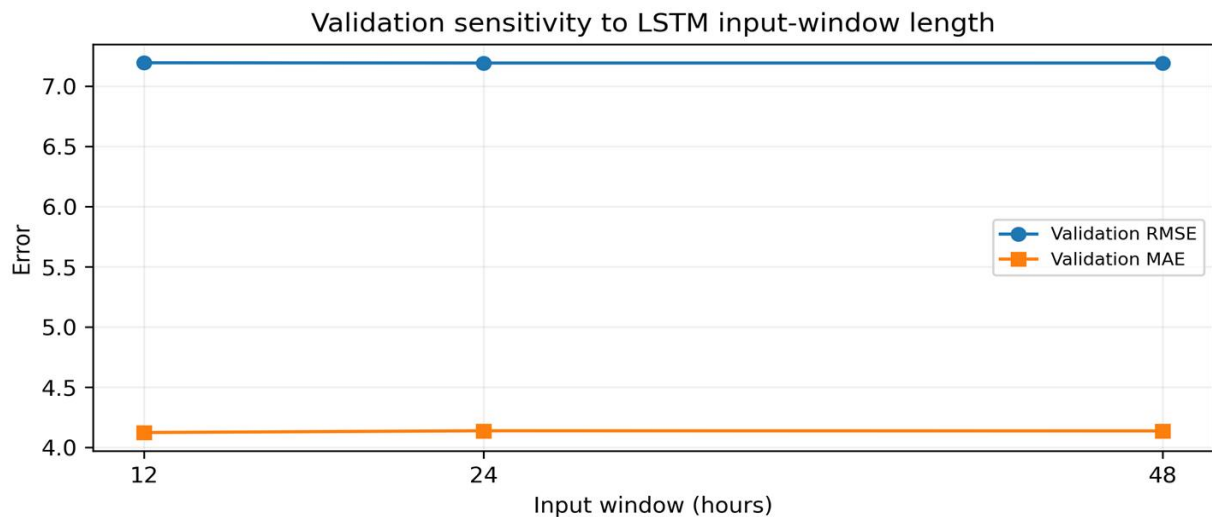


Fig. 4. Validation sensitivity to LSTM input-window length

측값에 더하여 최종 예측값을 구하는 프로세스를 나타낸 것이다.

위의 표 6은 시계열 순차 검증을 통해 탐색한 하이퍼파라미터 후보와 최종 선택값을 제시한 것이다. 입력 윈도우, 은닉 노드, 학습률은 검증 RMSE를 우선 기준으로 선택하였고, dropout은 검증손실, batch size는 학습 안정성을 고려하였다. 최종 설정은 성능만이 아니라 제한된 과적합 위험과 모형 복잡도도 함께 고려하였다.

표 7은 입력 윈도우 길이에 따른 검증 성능 민감도 분석이다. 12, 24, 48시간 윈도우 성능 차이는 매우 작아, 본 자료에서 윈도우 길이에 대한 민감도가 크지 않음을 보여준다. 그럼에도 24시간은 일주기 변동을 직접 포함하면서 48시간보다 입력 차원과 계산 부담이 작으므로 성능과 단순성의 균형을 고려해 최종 선택하였다.

24시간과 48시간의 검증 성능 차이는 0.01 미만으로 매우 작았다. 이에 일주기 패턴을 포함하면서도 입력 차원과 계산량이 작은 24시간 윈도우를 선택하였다. 32개 은닉 노드와 학습률 0.003은 검증 RMSE를 가장 낮게 하였고, 단일 LSTM layer, tanh 상태 활성화, sigmoid gate, dropout 0.20, Adam(beta1=0.9, beta2=0.999, epsilon=1e-7), 최대 200 epochs, patience=20 early stopping을 적용하였다.

위의 그림 4는 표 7의 수치를 시각화한 것으로, 12·24·48 시간 입력 길이에서 검증 RMSE와 MAE가 거의 수평에 가깝게 유지됨을 보여준다. 이는 24시간 선택은 압도적인 성능 우위 때문이 아니라 일주기성 반영과 입력 차원 축소라는 근거에 기반함을 보여준다.

4. Evaluation Metrics, Statistical Tests, and Threshold Calibration

$$RSME = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (5)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

RMSE는 큰 예측오차에 더 큰 가중치를 부여하며, MAE는 평균적인 절대오차를 직관적으로 제시한다[16]. 예측성능의 통계 비교는 절대오차식은 다음과 같다.

$$d_i = |y_i - \hat{y}_i^B| - |y_i - \hat{y}_i^H| \quad (7)$$

절대오차는 위의 식 (7)을 이용한 양측 Diebold-Mariano 검정으로 수행하였다[17]. 일주기 의존성을 고려하기 위해 Newey-West 장기분산의 최대 시차와 moving-block bootstrap의 블록 길이는 모두 24시간으로 설정하였고, bootstrap은 5,000회 반복하였다. 다중 비교 p-value는 Holm 방법으로 보정하였다.

Table 8. One-hour-ahead test-set regression performance

Model	RMSE	MAE	RMSE reduction vs. Hybrid	MAE reduction vs. Hybrid
SARIMA	4.503	3.194	2.01%	2.42%
SARIMAX	4.694	3.198	5.99%	2.54%
LSTM	4.487	3.200	1.66%	2.59%
Proposed Hybrid	4.413	3.117	-	-

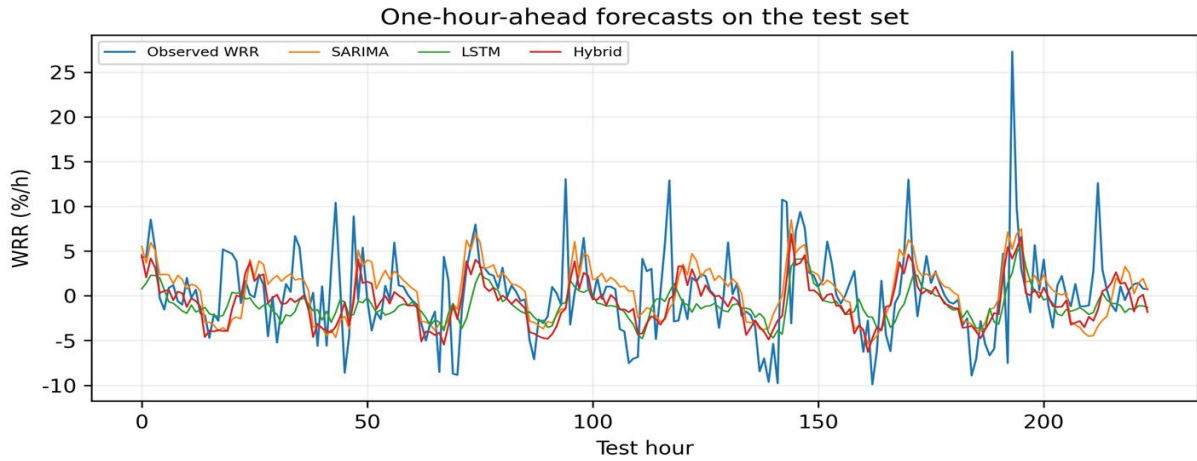


Fig. 5. One-hour-ahead forecasts of the observed WRR on the test set

$$E_t(c) = \Pi(y_t \leq c), c \in \{-3, -5, -7\} \quad (8)$$

$$A_t(\tau) = \Pi(\hat{y}_t^H \leq \tau) \quad (9)$$

식 (8)에서 c 는 관측 WRR로 정의한 이벤트 기준이며, 식 (9)의 τ 는 예측값으로 경보를 발생시키는 운영 임계값이다. 두 값은 서로 다른 개념이므로 분리하여 결정하였다.

$$s_t = -\hat{y}_t^H \quad (10)$$

ROC 계산에서는 낮은 WRR이 높은 위험을 의미하도록 식 (10)의 위험점수 s_t 를 사용하였다. 이에 따라 예측 WRR이 낮을수록 위험점수는 커지며, 일반적인 ROC해석과 같이 높은 점수가 높은 위험 가능성을 의미한다.

$$J(\tau) = \text{Sensitivity}(\tau) + \text{Ppccificity}(\tau) - 1 \quad (11)$$

$$F1(\tau) = \frac{2TP(\tau)}{2TP(\tau) + FP(\tau) + FN(\tau)} \quad (12)$$

Youden's J는 민감도와 특이도의 균형을 제공하므로 균형형 경보 임계값을 선택하는 데 사용하였다[18-20]. F1-score는 이벤트 비율이 낮을 때 정밀도와 재현율을 함께 평가하기 위해 사용하였다[15]. 모든 임계값은 검증셋에서 산출하고, 테스트셋에는 고정 적용하여 데이터 누수를 방지하였다.

III. Results and Discussion

1. Regression Performance and Comparative Interpretation

위의 표 8은 224개 테스트 시점의 1 시간 선행 예측오차를 비교한 것이다. 하이브리드 모형은 RMSE=4.413, MAE=3.117로 모든 비교모형보다 낮은 수치적 오차를 보였다. 가장 큰 차이는 SARIMAX 대비 RMSE 5.99%, MAE 2.54% 감소이며, 이는 외생 변수의 선형효과만으로 남은 비선형 잔차를 LSTM이 보정한 결과로 해석할 수 있다.

이상과 같이 본 연구에서 제안한 하이브리드 모형은 모든 비교모형보다 낮은 RMSE와 MAE를 보였다. SARIMA는 일주기 계절성과 자기상관을 반영하지만 외생 기상효과와 비선형 급변을 설명하지 못한다. SARIMAX는 지연기온·상대습도의 선형 효과를 추가하지만, 본 자료에서 그 효과가 단독으로 성능 향상으로 연결되지는 않았다. 직접 LSTM은 비선형 패턴을 학습하지만 제한된 표본에서 계절 구조와 잔차 변동을 동시에 학습해야 한다. 반면 하이브리드 모형은 SARIMAX가 구조적 성분을 먼저 추정하고 LSTM이 잔차 보정에 집중하므로 수치적으로 가장 낮은 오차를 기록하였다.

그림 5는 테스트 구간에서 실제 WRR과 하이브리드 예측값의 시간적 추세를 비교한 것이다. 예측선은 대부분의 일반 변동구간을 추적하지만, 급격한 양과 음의 스파이크

Table 9. Time-series-aware comparison of absolute forecast errors

Event criterion	Train rate	Validation rate	Test rate
WRR $\leq$ -3%	0.277	0.283	0.201
WRR $\leq$ -5%	0.187	0.175	0.121
WRR $\leq$ -7%	0.119	0.085	0.067

Table 10. Event-label distribution by criterion

Comparison	Mean Δ MAE	DM statistic	DM p-value	Holm p-value	95% block-bootstrap CI
SARIMA vs. Hybrid	0.077	0.692	0.489	0.978	[-0.154, 0.288]
SARIMAX vs. Hybrid	0.081	0.656	0.512	0.978	[-0.159, 0.276]
LSTM vs. Hybrid	0.083	1.084	0.278	0.835	[-0.074, 0.213]

Table 11. Validation-set ROC threshold optimization

Event criterion	Validation ROC-AUC	τ_J	Youden's J	τ_{F1}	Maximum F1	Selected threshold
WRR $\leq$ -3%	0.838	-0.479	0.621	-0.753	0.683	-0.479 (J)
WRR $\leq$ -5%	0.761	-0.479	0.536	-0.479	0.487	-0.479
WRR $\leq$ -7%	0.746	-2.083	0.490	-2.083	0.316	-2.083

에서는 오차가 커질 수 있다. 이는 표 8의 RMSE가 극단 변화의 예측실패에 민감한 이유와 연관되며, 환경 급변 구간에서 추가 현장정보의 필요성을 시사한다.

다음의 표 9에서 평균 Δ MAE가 양수인 것은 하이브리드 모형의 평균절대오차가 더 작았음을 의미한다. 그러나 DM 검정, Holm 보정 및 moving-block bootstrap 신뢰 구간은 0.05 수준의 통계적 유의성을 지지하지 않았다. 따라서 본 연구는 하이브리드 구조의 수치적 성능 우위를 제시하되, 제한된 단일 관측소 테스트셋에서는 일반화된 통계적 우위가 확정되지 않았다고 해석한다.

2. ROC-Based Event Detection and Operational Interpretation

표 10은 이벤트 기준이 엄격해질수록 위험 이벤트 비율이 감소함을 보여준다. -3% 기준은 약 20% 이상의 이벤트 비율을 가지는 민감한 조기경보 시나리오이고, -7% 기준은 7% 미만의 희소 이벤트를 다루는 보수적인 시나리오이다. 이벤트 비율이 낮아질수록 Precision, Recall, F1의 균형을 별도로 평가할 필요가 있다.

-3%, -5%, -7%는 실제 압축강도 저하의 확정 기준이 아니라, WRR 위험 이벤트의 민감도를 비교하기 위한 운영 시나리오이다. -3%는 조기 경보 성격의 민감한 기준, -7%는 보다 보수적인 고위험 기준, -5%는 그 중간 수준의 기본 분석 기준으로 해석하였다.

표 11은 검증셋에서 각 이벤트 기준별 ROC-AUC와 최적 경보 임계값을 제시한다. 기본 기준 WRR $\leq$ -5%에서는 Youden's J와 F1-score가 동일하게 -0.479/h에서 최대가 되어 균형형 탐지 기능과 동일한 경보 수준을 나타냈다. 반면 -3% 기준에서는 두 최적 임계값이 달라, 현장의 미탐·오경보 비용에 따라 선택이 달라질 수 있음을 보여준다.

그림 6은 본 기준 세 이벤트 정의에 대한 테스트셋 ROC곡선을 나타낸다. 곡선이 좌상단에 가까울수록 높은 민감도와 특이도의 조합을 제공한다. -5% 기준의 ROC-AUC는 0.788로 예측 WRR이 위험 이벤트를 구분하는 데 일정한 수준의 판별력을 가지지만, 단독으로 확정 진단에 사용하기에는 제한적임을 보여준다.

표 12는 검증셋에서 고정된 임계값을 테스트셋에 적용한 결과이다. 기본 기준 WRR $\leq$ -5%에서는 실제 이벤트 27건 중 25건을 탐지하여 Recall=0.926을 보였지만, FP=94로 Precision=0.210에 머물렀다. 따라서 이 임계값은 압축강도 저하를 확정하는 진단 기준이 아니라, 현장 점검 또는 추가 품질시험을 유도하는 조기경보 기준으로 해석해야 한다.

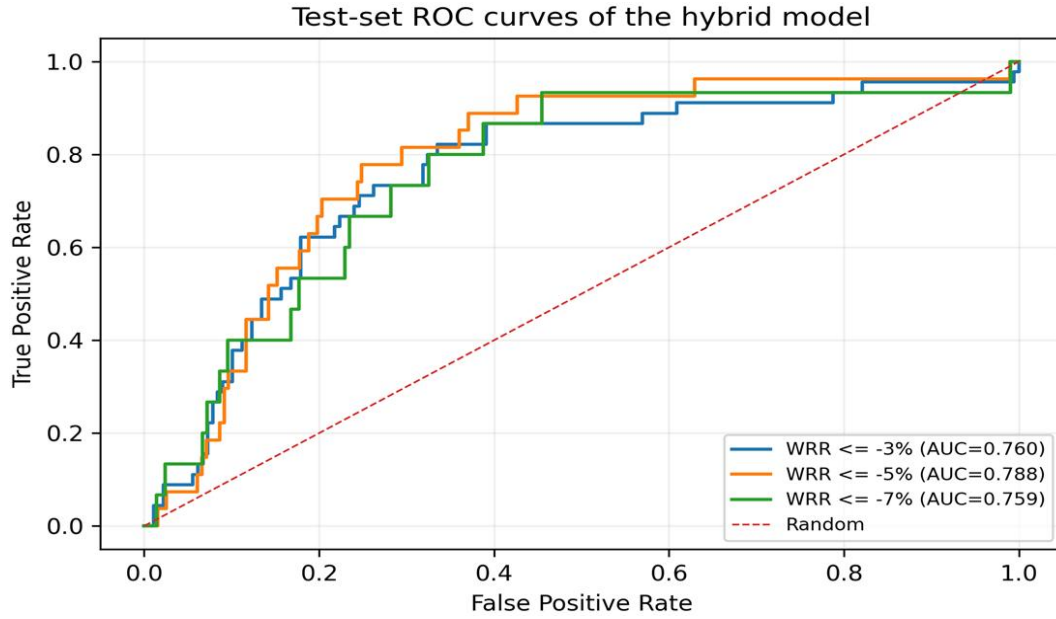


Fig. 6. Test-set ROC curves of the proposed hybrid model three event definitions

Table 12. Test-set event-detection performance

Event criterion	Applied tau	TN	FP	FN	TP	Precision	Recall	F1	ROC-AUC
WRR ≤ -3%	-0.479	99	80	6	39	0.328	0.867	0.476	0.760
WRR ≤ -5%	-0.479	103	94	2	25	0.210	0.926	0.342	0.788
WRR ≤ -7%	-2.083	154	55	5	10	0.154	0.667	0.250	0.759

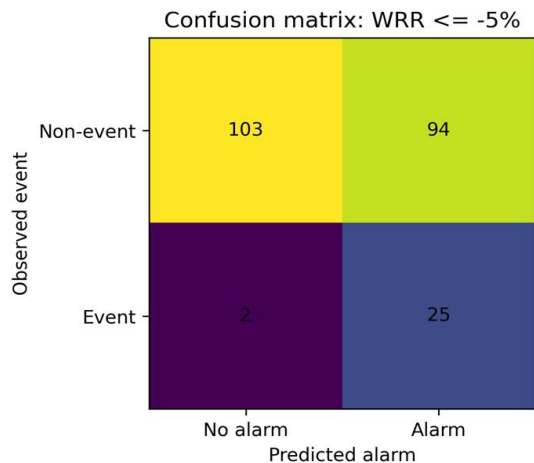
Fig. 7. Confusion matrix on the test set under WRR ≤ -5% and $\tau = -0.479\%/h$

그림 7은 기본 이벤트 기준에서의 혼동행렬을 시각화한 것이다. True Positive는 25건, False Negative는 2건으로 미탐은 적지만, False Positive가 94건으로 상대적으로 많이 나타났다. 이는 안전 우선 경보에서 흔히 나타나는 민감도-정밀도 상충관계를 보여주며, 현장 비용구조에 따라 임계값을 재조정할 필요가 있음을 시사한다.

IV. Conclusion

본 연구는 시간당 상대습도 변화율로 산출된 WRR을 기상 기반 환경 위험지표로 정의하고, 이를 1시간 선행 예측하기 위한 SARIMAX-LSTM 하이브리드 모델을 제안하였다. 원시자료를 시간 구간 floor 방식으로 정렬하고, 중복 값은 중앙값으로 집계하며, 단기 외생변수 예측은 제한적으로 보간하고 WRR 자체는 보간하지 않는 절차를 제시하였다. 또한 기존 원고의 WRR 최대값과 상관계수 오류를 원시자료 기반으로 수정하였다.

풍속은 물리적으로 중요한 후보변수이지만, 본 관측소 자료에서는 기온과 상대습도에 추가하여도 검증 RMSE와 MAE를 동시에 개선하지 못하였다. 이에 최종 모형은 지연 기온과 상대습도를 외생 변수로 사용하였다. SARIMAX 차수는 BIC와 Ljung-Box 잔차 진단을 통해 (1,0,3)(1,0,1)24로 선정하였고, LSTM은 24시간 윈도우와 32개 은닉 노드로 구성하였다.

테스트셋에서 제안 하이브리드 모형은 RMSE=4.413, MAE=3.117로 비교모형보다 낮은 수치적 오차를 보였다. 그러나 Diebold-Mariano 검정과 moving-block

bootstrap에서는 통계적 유의성이 확인되지 않았으므로, 하이브리드 구조의 우위는 현 자료에서 탐색적 성능 개선으로 해석해야 한다. 이는 단일 관측소, 약 2개월의 자료, 224개 테스트 시점이라는 제한된 검증 범위와도 관련된다.

ROC 기반 임계값 최적화에서는 $WRR \leq -5\%$ 의 경우 $\tau = -0.479\%/h$ 가 검증셋에서 Youden's J와 F1을 동시에 최대화하였다. 테스트셋에서 높은 재현율을 보였지만 정밀도는 낮아, 본 결과는 확정 진단보다는 조기 점검을 위한 보조 경보 체계로 활용하는 것이 타당하다.

향후 연구에서는 다지역·다계절 자료, 실제 타설면 풍속, 표면온도, 증발량, 균열 및 압축강도 측정치를 함께 수집하여 WRR과 실제 품질저하의 정량적 관계를 외부 검증하고, 오경보 비용과 미탐 비용을 반영한 민감 임계값 최적화 및 온라인 학습 기반 모델 갱신 구조로 확장할 계획이다.

ACKNOWLEDGEMENT

This study was supported by the Research Program funded by the Seoul Theologica University.

REFERENCES

- [1] J.-Y. Kim, "Analysis of Time-Series Data According to Water Reduce Ratio and Temperature and Humidity Changes Affecting the Decrease in Compressive Strength of Concrete Using the SARIMA Model," *Journal of The Korea Society of Computer and Information*, vol. 27, no. 10, pp. 123-130, 2022. DOI: <https://doi.org/10.9708/jksci.2022.27.10.123>
- [2] ACI Committee 305, ACI PRC-305-20: Guide to Hot Weather Concreting, American Concrete Institute, 2020.
- [3] I.-C. Yeh, "Modeling of strength of high-performance concrete using artificial neural networks," *Cement and Concrete Research*, vol. 28, no. 12, pp. 1797-1808, 1998. DOI: 10.1016/S0008-8846(98)00165-3.
- [4] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.
- [5] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, 2017. DOI: 10.1109/TNNLS.2016.2582924.
- [6] G. P. Zhang, "Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model," *Neurocomputing*, vol. 50, pp. 159-175, 2003. DOI: 10.1016/S0925-2312(01)00702-0.
- [7] M. Khashei and M. Bijari, "A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting," *Applied Soft Computing*, vol. 11, no. 2, pp. 2664-2675, 2011. DOI: 10.1016/j.asoc.2010.10.015.
- [8] Z. Zhao et al., "Study on the Prediction Effect of a Combined Model of SARIMA and LSTM Based on SSA for Influenza in Shanxi Province, China," *BMC Infectious Diseases*, vol. 23, art. 46, 2023. DOI: 10.1186/s12879-023-08025-1.
- [9] A. Atesogun and M. Gulsen, "A Hybrid Forecasting Structure Based on ARIMA and Artificial Neural Network Models," *Applied Sciences*, vol. 14, no. 16, art. 7122, 2024. DOI: 10.3390/app14167122.
- [10] D. Ashtar, S. S. Mohammadi Ziabari, and A. M. M. Alsahag, "Hybrid Forecasting for Sustainable Electricity Demand in The Netherlands Using SARIMAX, SARIMAX-LSTM, and Sequence-to-Sequence Deep Learning Models," *Sustainability*, vol. 17, no. 16, art. 7192, 2025. DOI: 10.3390/su17167192.
- [11] C. Bergmeir, R. J. Hyndman, and B. Koo, "A Note on the Validity of Cross-Validation for Evaluating Autoregressive Time Series Prediction," *Computational Statistics & Data Analysis*, vol. 120, pp. 70-83, 2018. DOI: 10.1016/j.csda.2017.11.003.
- [12] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, 1974. DOI: 10.1109/TAC.1974.1100705.
- [13] G. Schwarz, "Estimating the Dimension of a Model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978. DOI: 10.1214/aos/1176344136.
- [14] G. E. P. Box and D. A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, vol. 65, no. 332, pp. 1509-1526, 1970. DOI: 10.1080/01621459.1970.10481180.
- [15] G. M. Ljung and G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, vol. 65, no. 2, pp. 297-303, 1978. DOI: 10.1093/biomet/65.2.297.
- [16] T. Chai and R. R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)?," *Geoscientific Model Development*, vol. 7, pp. 1247-1250, 2014. DOI: 10.5194/gmd-7-1247-2014.
- [17] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253-263, 1995. DOI: 10.1080/07350015.1995.10524599.
- [18] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006. DOI: 10.1016/j.patrec.2005.10.010.
- [19] W. J. Youden, "Index for Rating Diagnostic Tests," *Cancer*, vol. 3, no. 1, pp. 32-35, 1950. DOI: 10.1002/1097-0142(1950)3:1<32::

AID-CNCR2820030106>3.0.CO;2-3.

- [20] R. Fluss, D. Faraggi, and B. Reiser, "Estimation of the Youden Index and Its Associated Cutoff Point," *Biometrical Journal*, vol. 47, no. 4, pp. 458-472, 2005. DOI: 10.1002/bimj.200410135.
- [21] J. Davis and M. Goadrich, "The Relationship between Precision-Recall and ROC Curves," *Proceedings of ICML*, pp. 233-240, 2006. DOI: 10.1145/1143844.1143874.

Authors



Joon-Yong Kim received the B.S., degrees in Civil Engineering from SungKyunKwan University, Korea, in 1985. Then received the M.S. and Ph.D. degrees in Computer Science and Engineering from KongJu National

University, Korea, in 2013 and 2018, respectively Dr. Kim joined the faculty of the Department of IT Convergence Software at Seoul Theological University, Gyeonggi-do, Korea, in 2020. He is currently a Professor in the Department of AI Convergence, Seoul Theological University. He is interested in Machine Learning, Auto ML, and AI.