Master Thesis Defense Presentation

Spatial and Temporal Data Analysis for Estimating Surface Rainfall Using Radar

Chair: Professor. Cho, Hwan-Gue
Reviewer 1: Professor. Choi, Yoon-Ho
Reviewer 2: Professor. Hong, Bonghee (Supervisor)

Student: Oudomseila PHOK

Date: *2019 / 05 / 30 (UPDATE)*
Location: 자연대연구실험동 404호

Content

Radar Weather Station (RWS)

- gives the current rainfalls of the range of RWS with radar image pixel values.

Description:

- record Rain 1.5 km from ground
- Current Rain Estimation Method

Data Output Format:

- Pixel in Color code
- 32 categories of Rainfalls (mm / hour)
- Update every 10 minutes

Range :

- up to 240 km / Radar (Main Weather Images are combined by multiple radar)
- 1 pixel is approximately 2 square kilometer

Aerial Rainfall Data

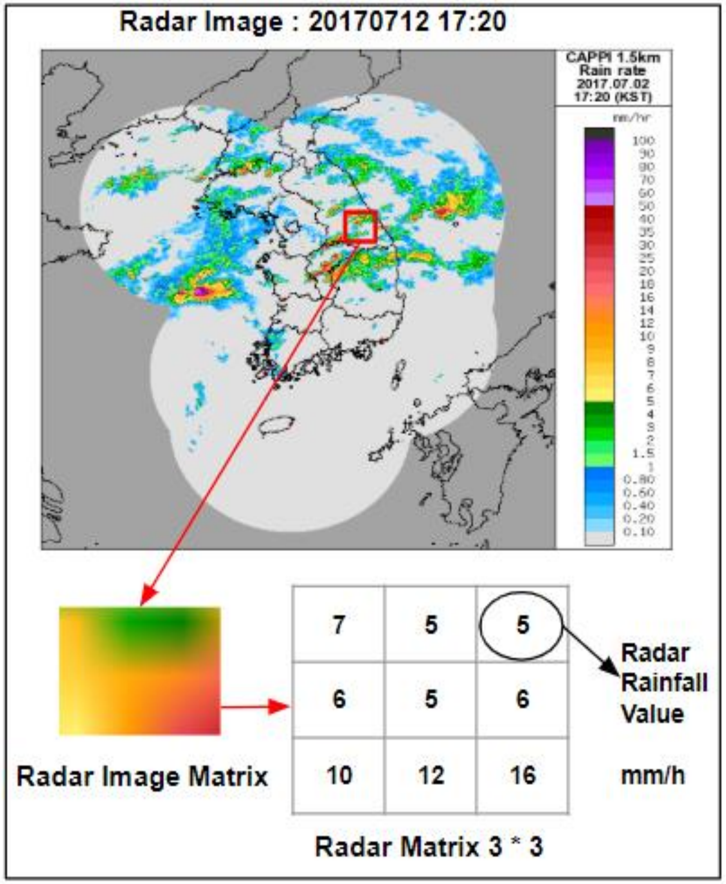


Figure 1. Example of Radar Image Data

Automatic Weather Station (AWS)

- give the rainfall of the AWS's ground location

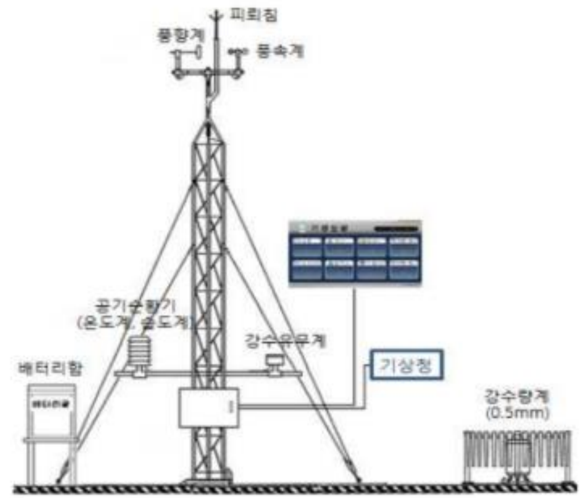
Description:

- Located on Surface
- Attached with multiple sensors to record surface rainfall, temperature

Data Output Format:

- numeric data output into csv
- record every minute

Surface Rainfall Data



Station No	Date	Temperature(C)	Precipitation(mm)	Wind Direction(deg)	Wind speed(m/s
160	2017-07-01 0:01	21.2	0	277.2	5.6
160	2017-07-01 0:02	21.2	0	275.7	4.5
160	2017-07-01 0:03	21.2	0	274	4.8
160	2017-07-01 0:04	21.2	0	279.1	4.6
160	2017-07-01 0:05	21.2	0	270.6	5.1

Figure 2. Example of Ground Weather Data (AWS)

1. Introduction – AWS & RWS Mismatch

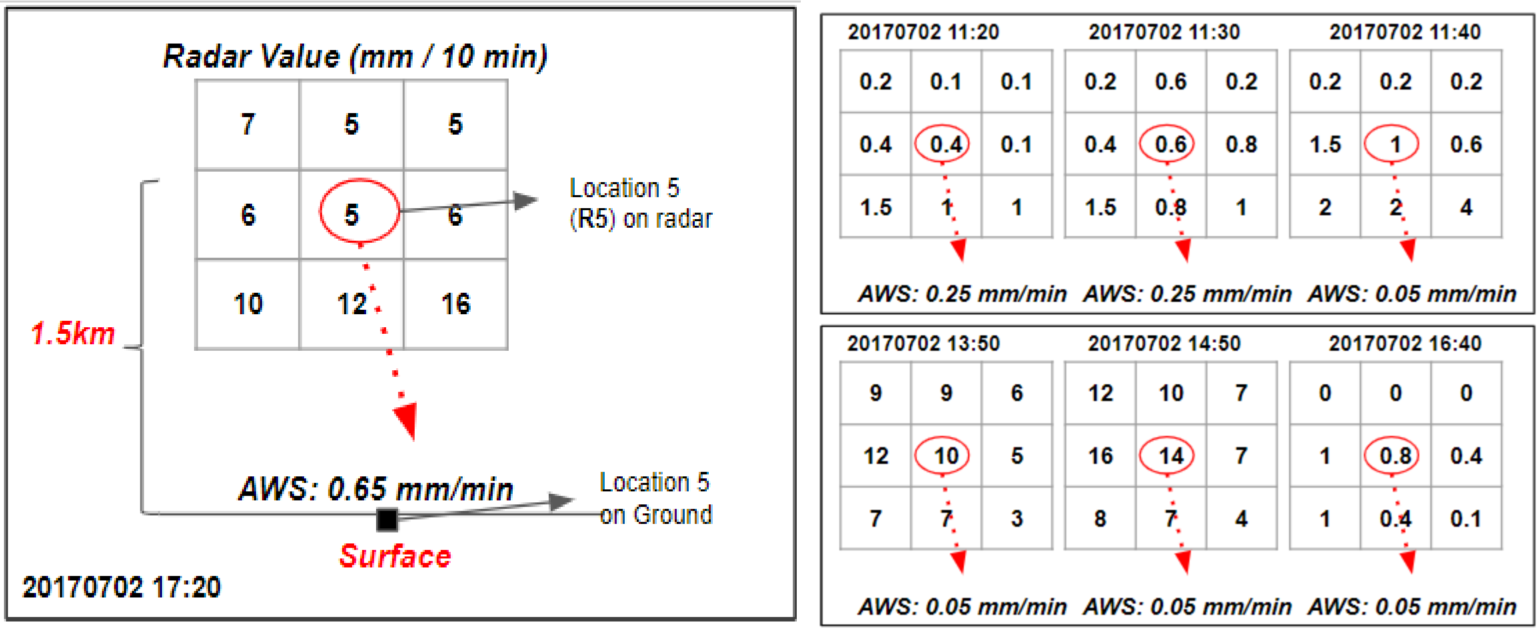


Figure 3. Mismatch Scenario of RWS and AWS Data

Description

- The RWS and AWS data does not match with each other.
- Therefore, the RWS data cannot directly predict the surface rainfall data

2. Problem Definition

- **Problem Definition**
- **Key Technical Issues**
- **ARIMA Model Explanation**
- **Research Contributions**
- **Model Processing Overview**

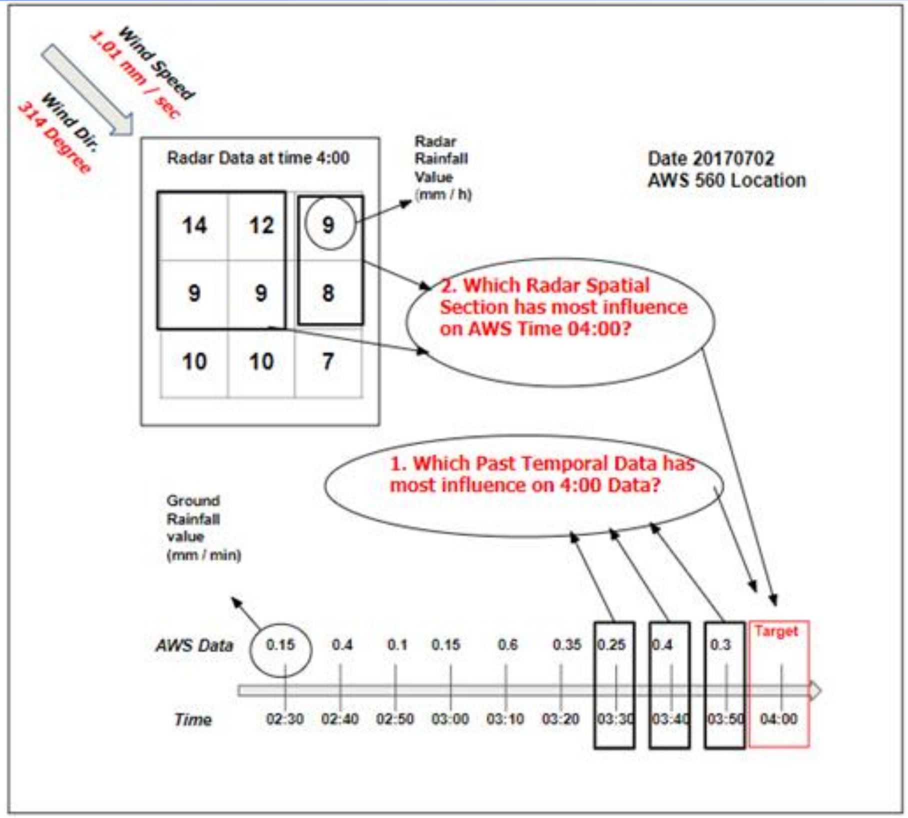


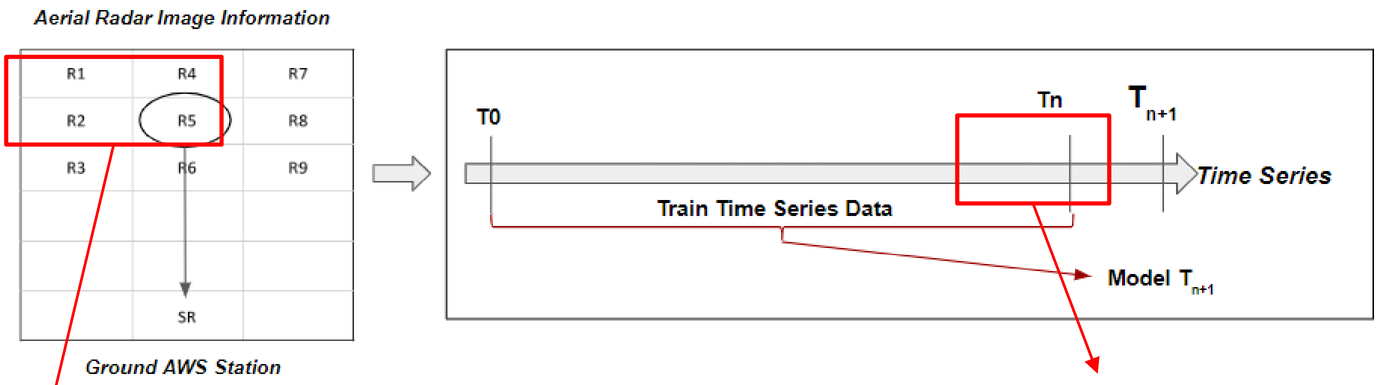
Figure 4. Surface Rainfall Prediction Model with Time and Spatial Data

Problem Definition :

"Generate a time series prediction model for surface rainfall using different temporal and spatial data"

3. Two Main Key Technical Issues

- Problem Definition
- **Key Technical Issues**
- ARIMA Model Explanation
- Research Contributions
- Model Processing Overview



1. How to select Influential Spatial Data?

2. how to select Influential Temporal Data?

Proposed Solution

Define Radar-Ground Correlation Similarity

↓

Select Spatial Data Based on High Correlation and Wind Criteria

Proposed Solution

Define Current-Past Correlation Similarity

↓

Select Temporal Data Based on High Correlation

4. ARIMA Model : Autoregressive Integrated Moving Average

Problem Definition

Key Technical Issues

ARIMA Model Explanation

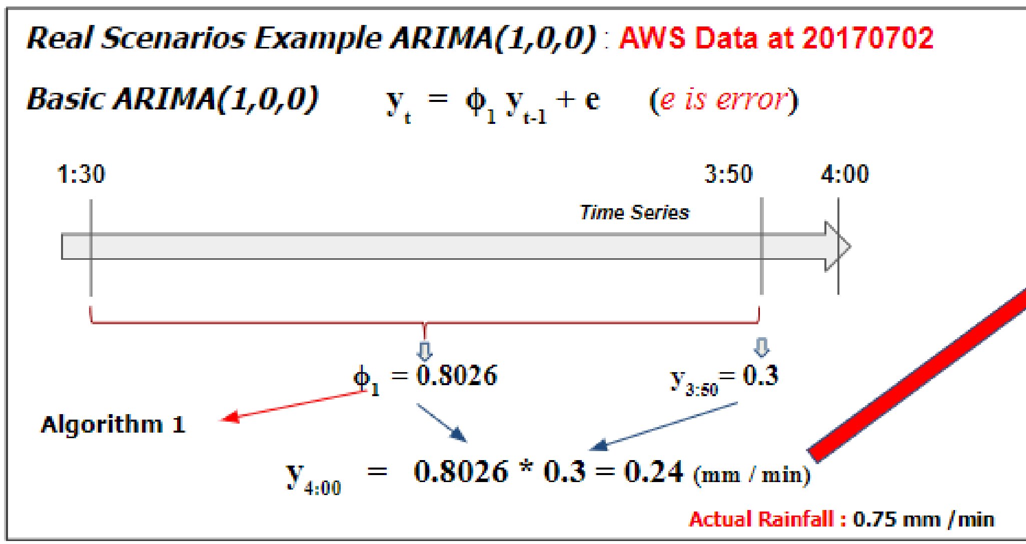
Research Contributions

Model Processing Overview

1. What is ARIMA Model?

"ARIMA is simply a time series prediction model that uses past data to predict future data"

Example:



Drawbacks

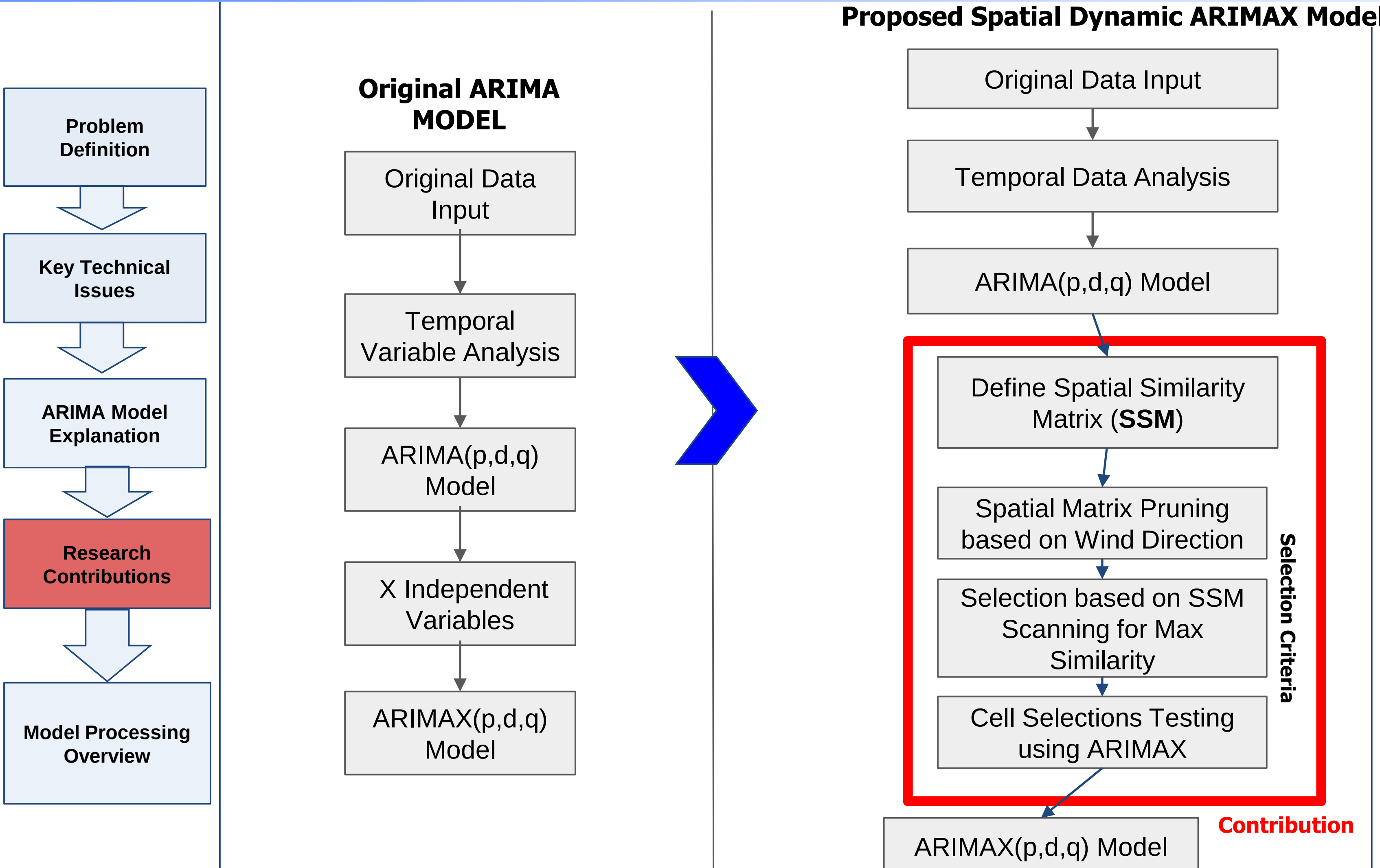
1./ Prediction only use Temporal Data

2./ Inclusion of other variable is possible, but without clear approach, better accuracy cannot be guaranteed

Figure 5. Example of ARIMA Model Prediction

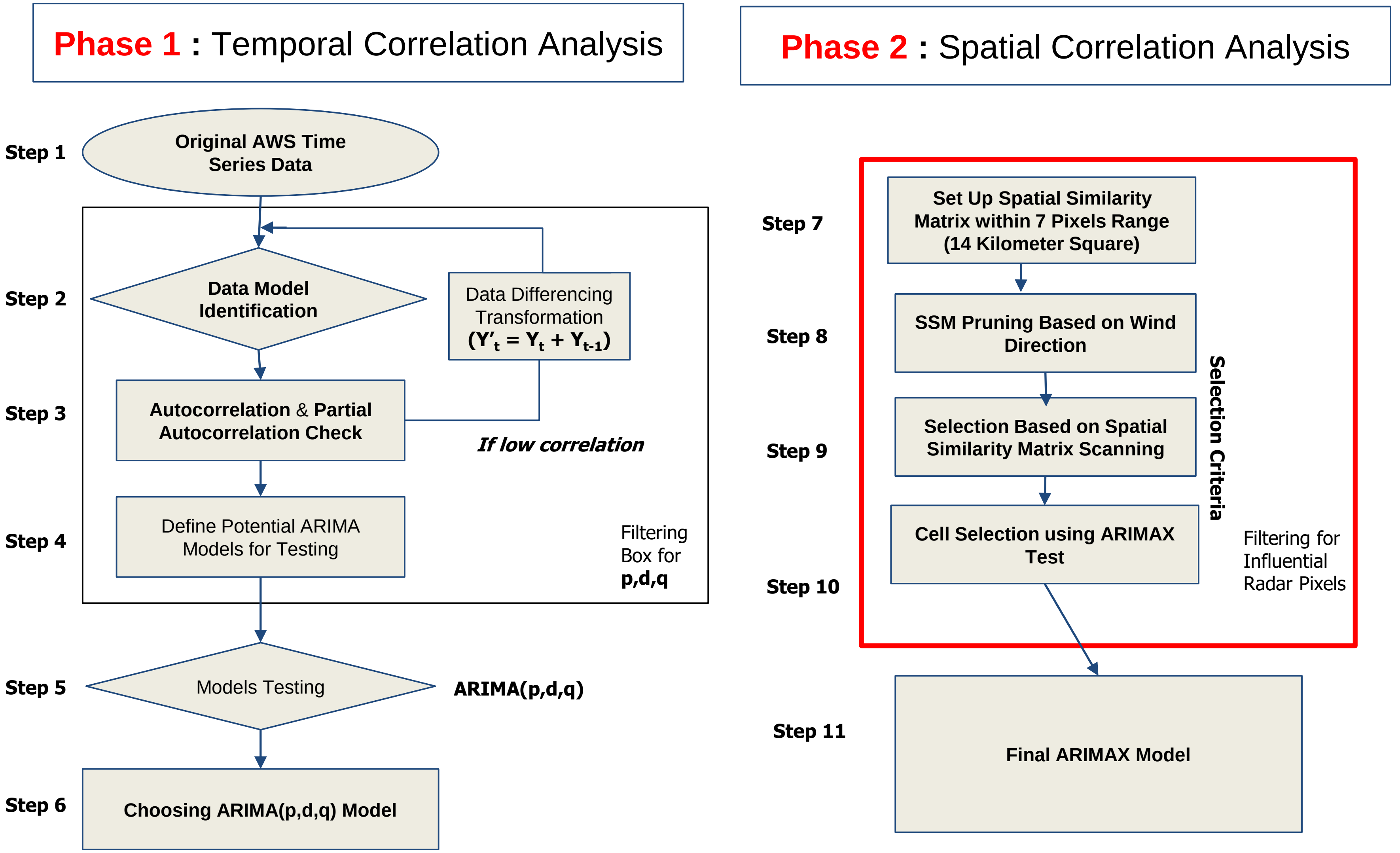
Reference:
https://people.duke.edu/~rnau/411arim.htm

5. Spatial Dynamic ARIMAX Model (SDAM)



Limitation: Spatially Correlated Variables must be existed

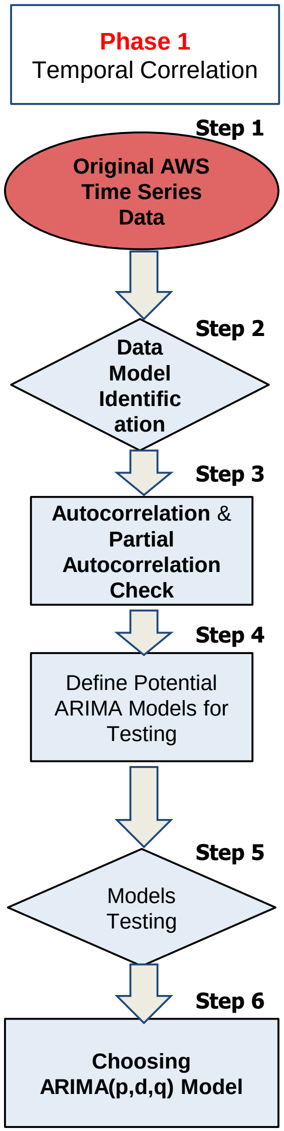
5. Proposed Model Process Flow



6. Implementation Details

Phase 1: Temporal Correlation Analysis

Step 1: Original AWS Data Extraction



Surface Rainfall in **CSV file** format

지점	일시	기온(°C)	1분 강수량(mm)	풍향(deg)	풍속(m/s)
560	2017-07-01 0:01	22.3	0	255.9	1.6
560	2017-07-01 0:02	22.3	0	243.1	2.2
560	2017-07-01 0:03	22.4	0	240.5	2.1
560	2017-07-01 0:04	22.4	0	236.5	2.2
560	2017-07-01 0:05	22.3	0	233.8	2.2

Surface Rainfall Data in Graph Pattern

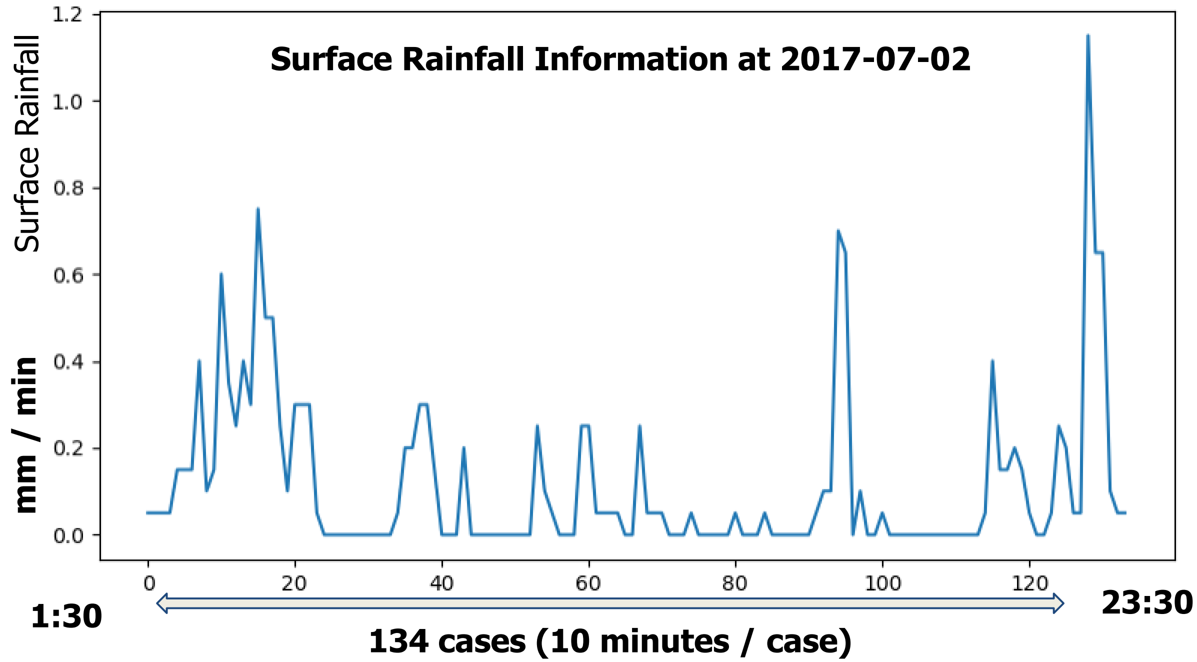
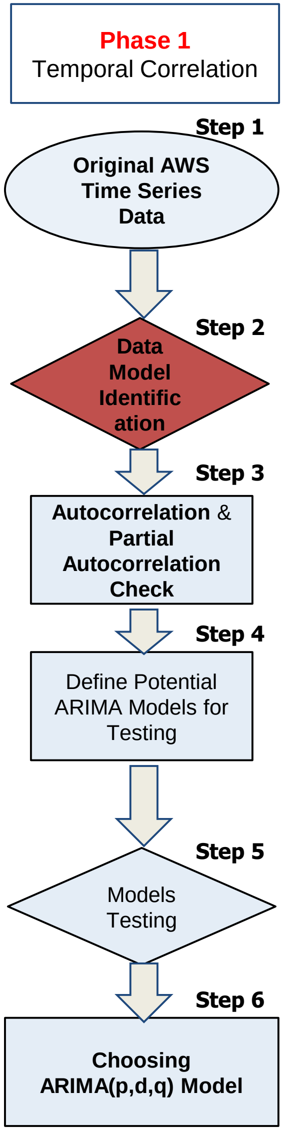


Figure 6 : AWS Rainfall Data in Time Series

Step 2: Different Data Models Extraction



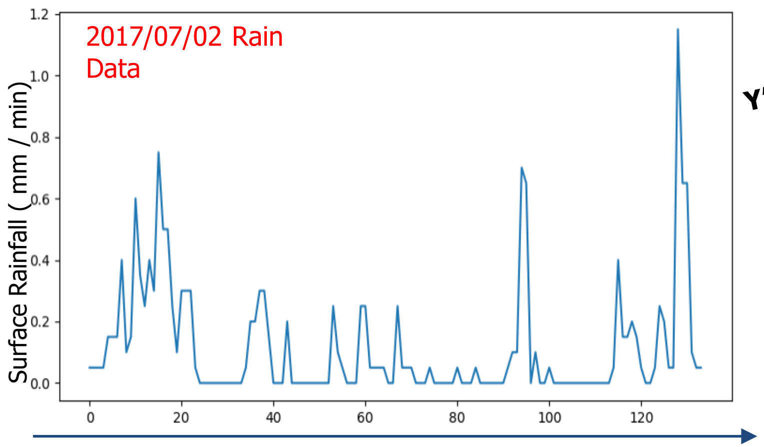


Figure 7: Original **Data Model**

$Y'_t = Y_t + Y_{t-1}$

1 time differencing

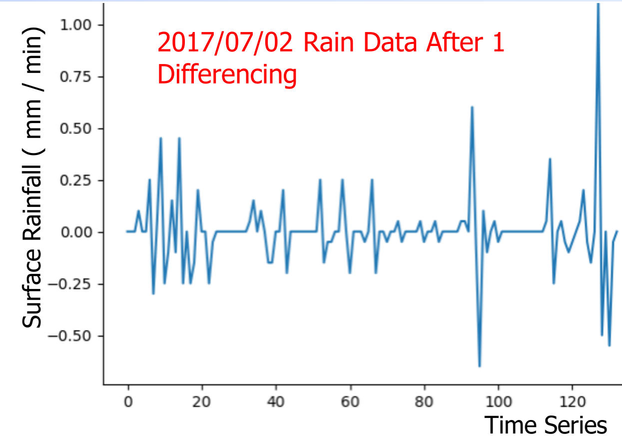


Figure 8: Potential Data **Model 1**

$Y''_t = Y'_t + Y'_{t-1}$

2 times differencing

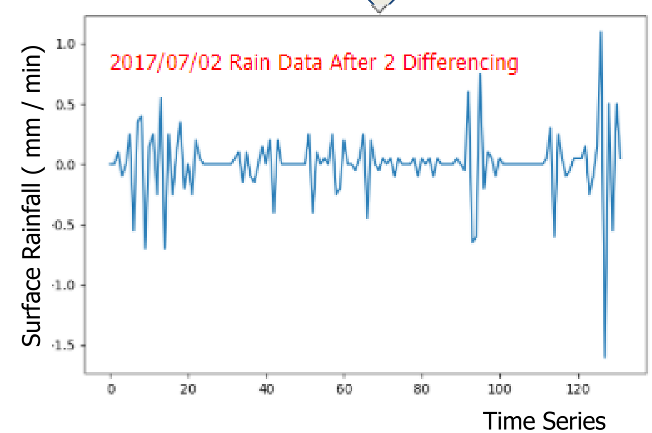


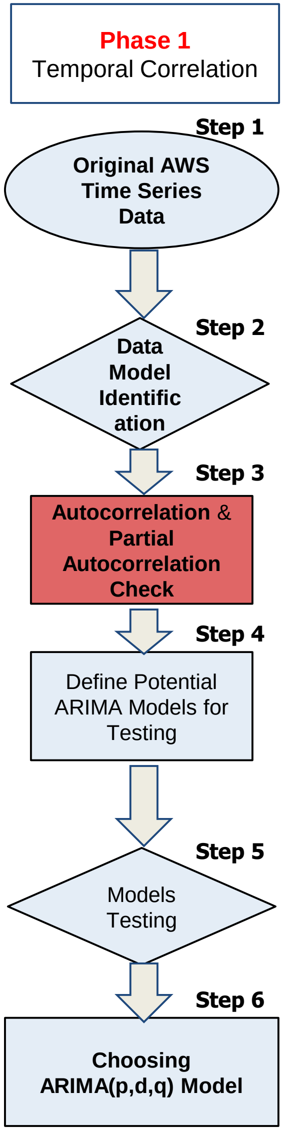
Figure 9: Potential Data **Model 2**

Technical Question: Is Original Data Model provides good correlation value?

Purpose:

- **Identify different potential Data Models** through Data Differencing transformation.
- **Compare Temporal Correlation Result**

Step 3 : Temporal Correlation Result (Concept)



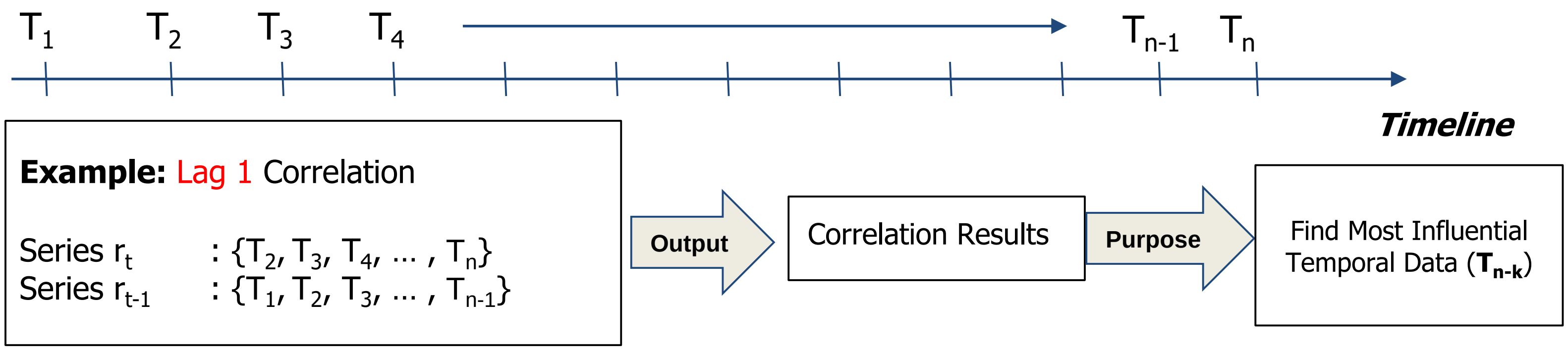


Figure 10: Concept of Correlation Result Calculation

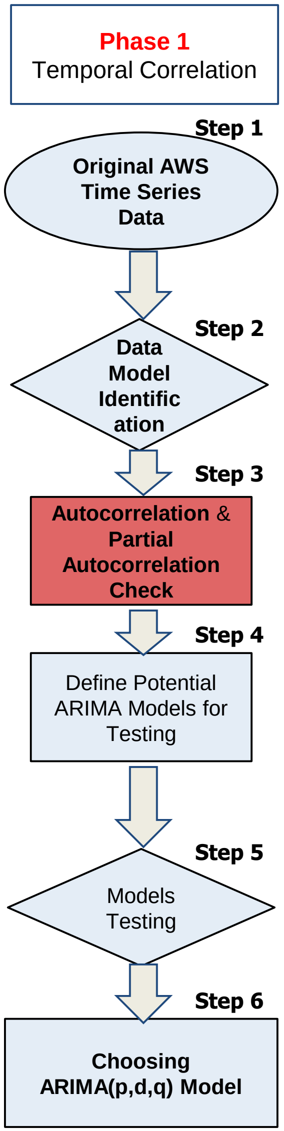
Details:

- Correlation Results refers to **Autocorrelation** and **Partial Autocorrelation results.**

- In ARIMA (p, d, q):

○ **d** is the data model ranking (d = 0 is original data model, refer to previous slide)
○ **p** is selected by the highest autocorrelation result within each **d**
○ **q** is select by the highest partial autocorrelation result within each **q**

Step 3: Original Data Model Correlation Analysis (d = 0)



Example of Original Data Temporal Correlation:

- In ARIMA (p, d, q), this is when d = 0 (original data)

⇒ p = 1 and q = 1 has high correlation result

⇒ Potential Test Models are ARIMA(1,0,1), ARIMA (1,0,0) and ARIMA(0,0,1)

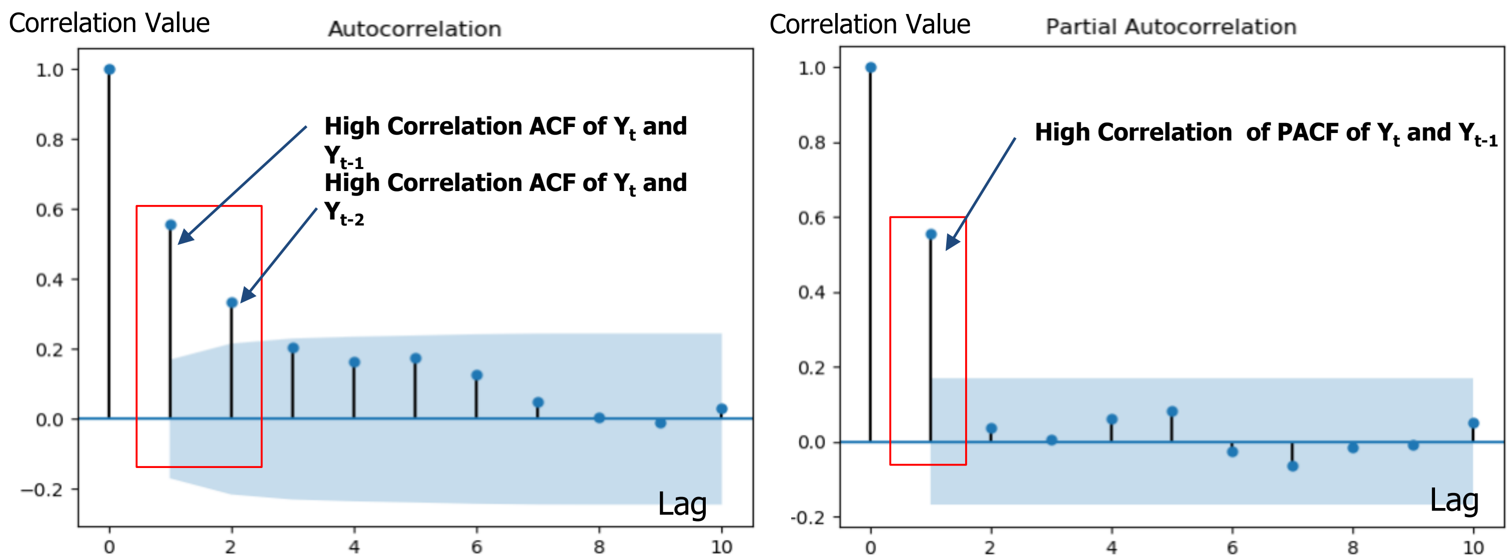
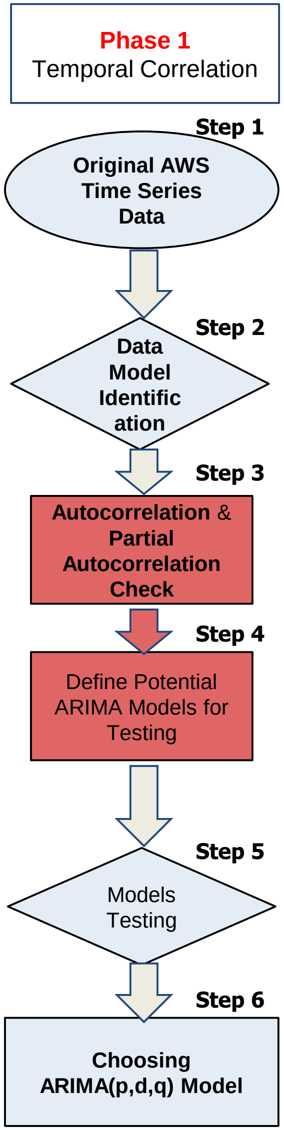


Figure 11: Temporal Correlation Results of Original Data Model

Step 4: Temporal Correlation Result and Potential Models



Summary Temporal Correlation Check when d = 0, d = 1 and d = 2

Purpose: Defining Potential Combination of p,d,q for ARIMA(p,d,q) before testing.

Original Data Model Result

d = 0 when no differencing.

Correlation Result:
p = 1 and q = 1

Potential Model:

ARIMA (1,0,1)
ARIMA (0,0,1)
ARIMA (1,0,0)

Data Model 1 Result

d = 1 when 1 differencing.

Correlation Result:
no high correlation result

Potential Model:

No Potential Models

Data Model 2 Result

d = 2 when 2 differencing.

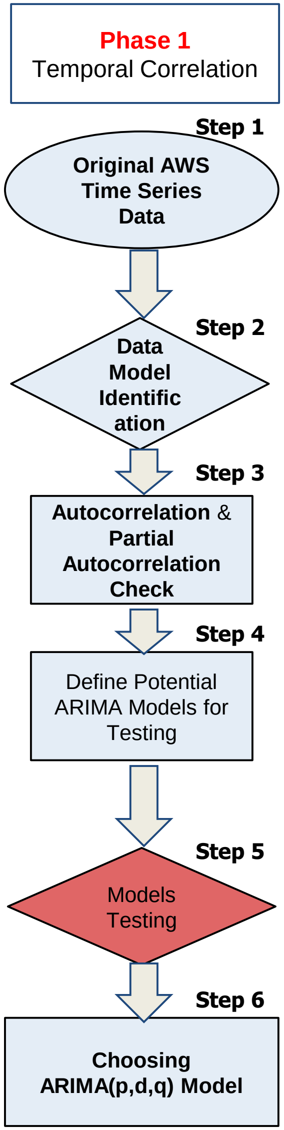
Correlation Result:
p = 1 and q = 1

Potential Model:

ARIMA (0,2,1)
ARIMA (1,2,0)

Model Testing (Next Step)

Step 5: Defining Possible Models for Testing Day 1 (2017-07-02)



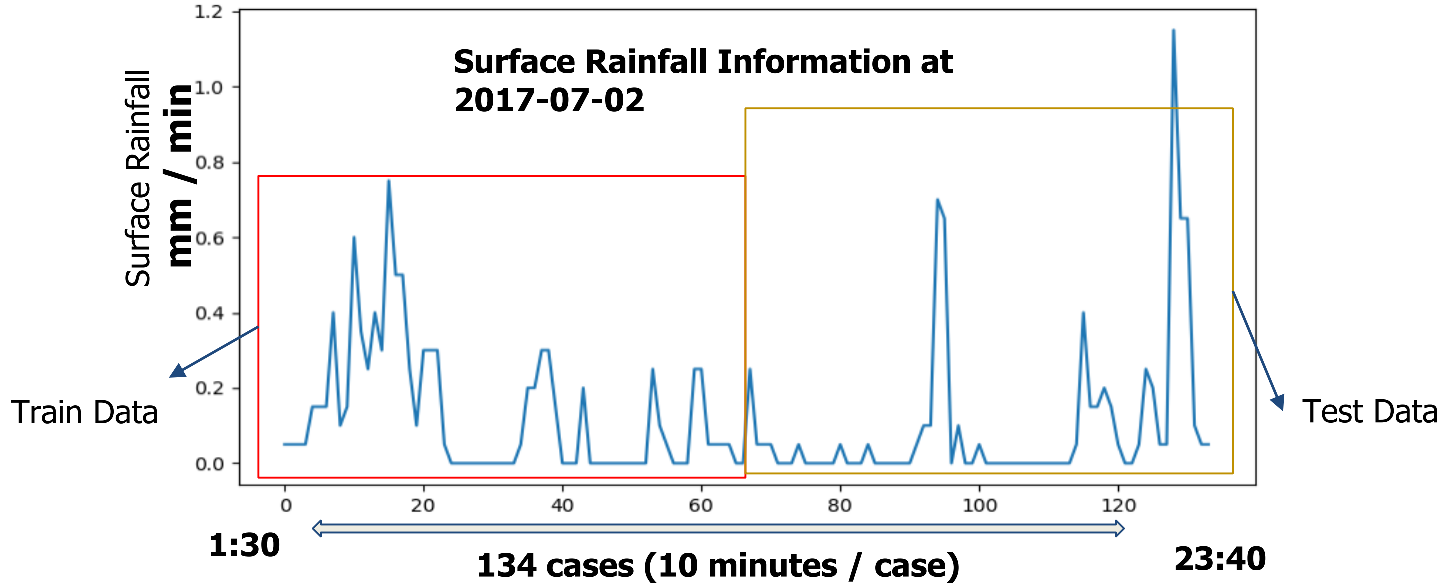
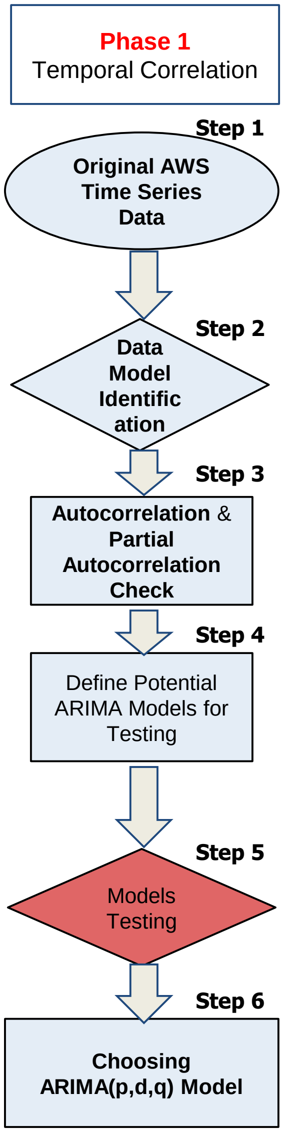


Figure 12: Train and Test Data Information

Table 1: Testing and Accuracy Table

ARIMA(p,d,q)	Equation (without constant)	Mean Square Error	Mean Absolute Error
ARIMA(1,0,0)	$\hat{y}_t = \phi_1 y_{t-1}$	0.037 mm / min	0.081 mm / min
ARIMA(1,0,1)	$\hat{y}_t = \phi_1 y_{t-1} + \theta_1 e_{t-1}$	0.038 mm / min	0.084 mm / min
ARIMA(0,0,1)	$\hat{y}_t = \theta_1 e_{t-1}$	0.042 mm / min	0.094 mm / min
ARIMA(0,2,1)	$\hat{y}_t = \theta_1 e_{t-1}$ // with 2 times differencing	0.044 mm / min	0.090 mm / min
ARIMA(1,2,0)	$\hat{y}_t = \phi_1 y_{t-1}$ // with 2 times differencing	0.082 mm / min	0.147 mm / min

Step 5: Defining Possible Models for Testing Day 2 (2017-07-10)



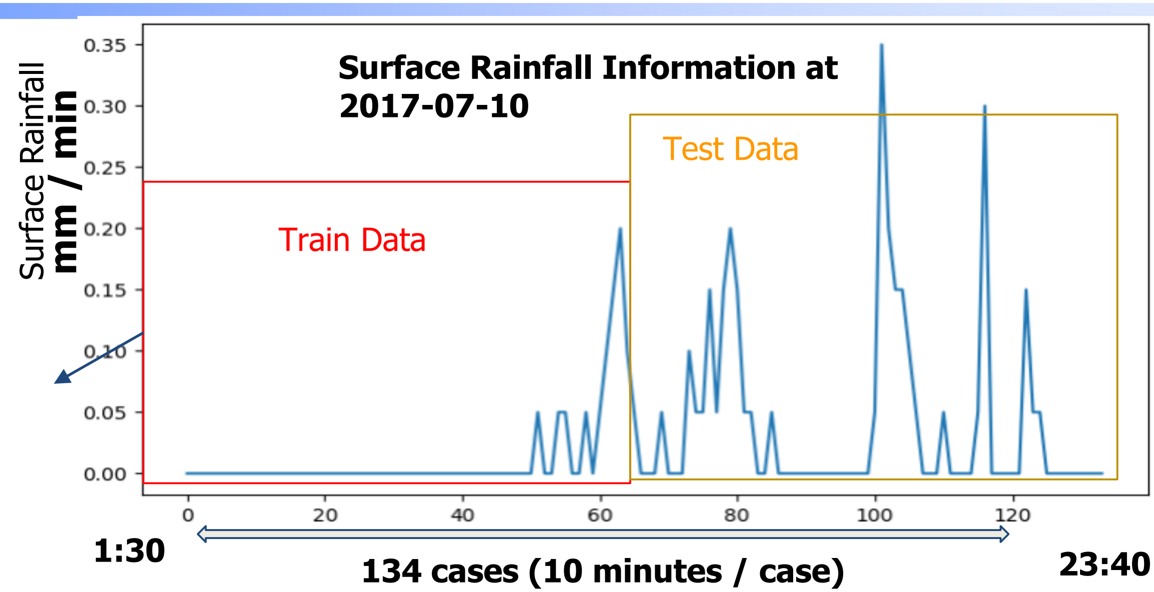
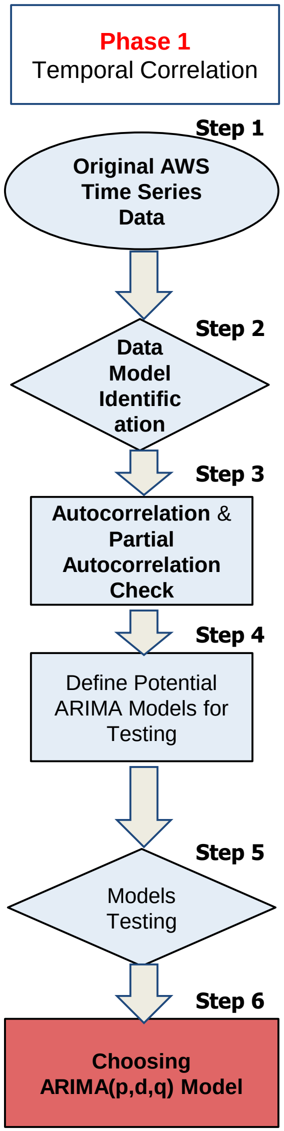


Figure 13: Train and Test Data Information

Table 2: Testing and Accuracy Table

ARIMA(p,d,q)	Equation (without constant)	Mean Square Error	Mean Absolute Error
ARIMA(1,0,0)	$\hat{y}_t = \phi_1 y_{t-1}$	0.005 mm / min	0.034 mm / min
ARIMA(1,0,1)	$\hat{y}_t = \phi_1 y_{t-1} + \theta_1 e_{t-1}$	0.005 mm / min	0.036 mm / min
ARIMA(0,0,1)	$\hat{y}_t = \theta_1 e_{t-1}$	0.006 mm / min	0.044 mm / min
ARIMA(0,2,1)	$\hat{y}_t = \theta_1 e_{t-1}$ // with 2 times differencing	0.006 mm / min	0.039 mm / min
ARIMA(1,2,0)	$\hat{y}_t = \phi_1 y_{t-1}$ // with 2 times differencing	0.011 mm / min	0.061 mm / min



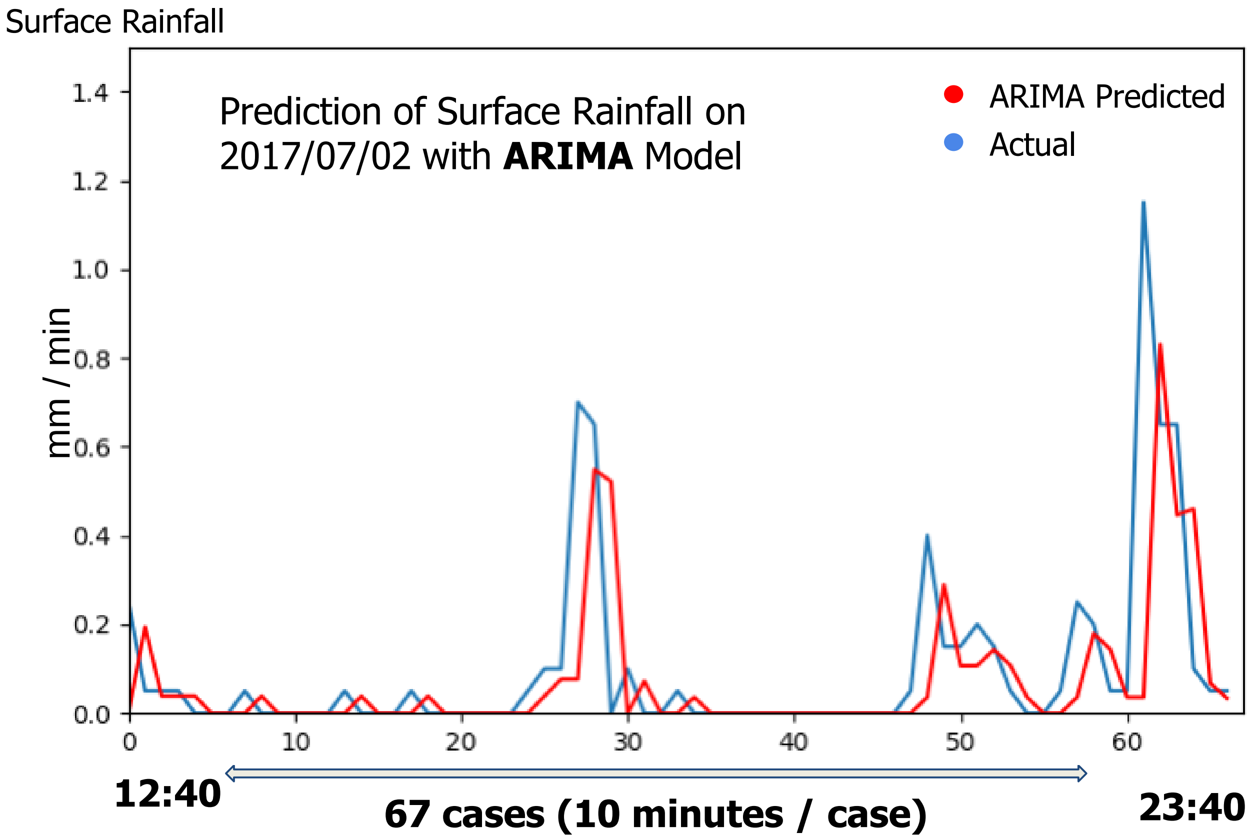


Figure 14: Prediction Result of AWS 560

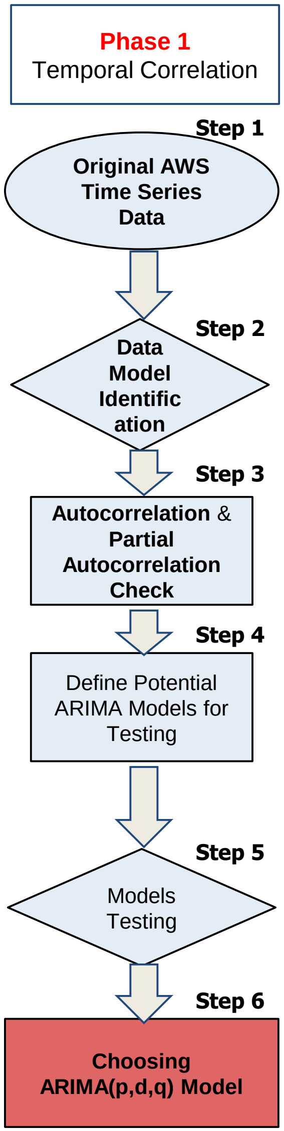
Best Fitted Model

$$\hat{y}_t = \phi_1 y_{t-1} + e$$

ARIMA Model depends on its past value y_{t-1} and estimator ϕ_1

ARIMA(p,d,q)	Equation (without constant)	Mean Square Error	Mean Absolute Error
ARIMA(1,0,0)	$\hat{y}_t = \phi_1 y_{t-1} + e$	0.037 mm / min	0.081 mm / min

Step 6 : Selected Model Testing Result ARIMA (1,0,0) Day 2 (2017-07-10)



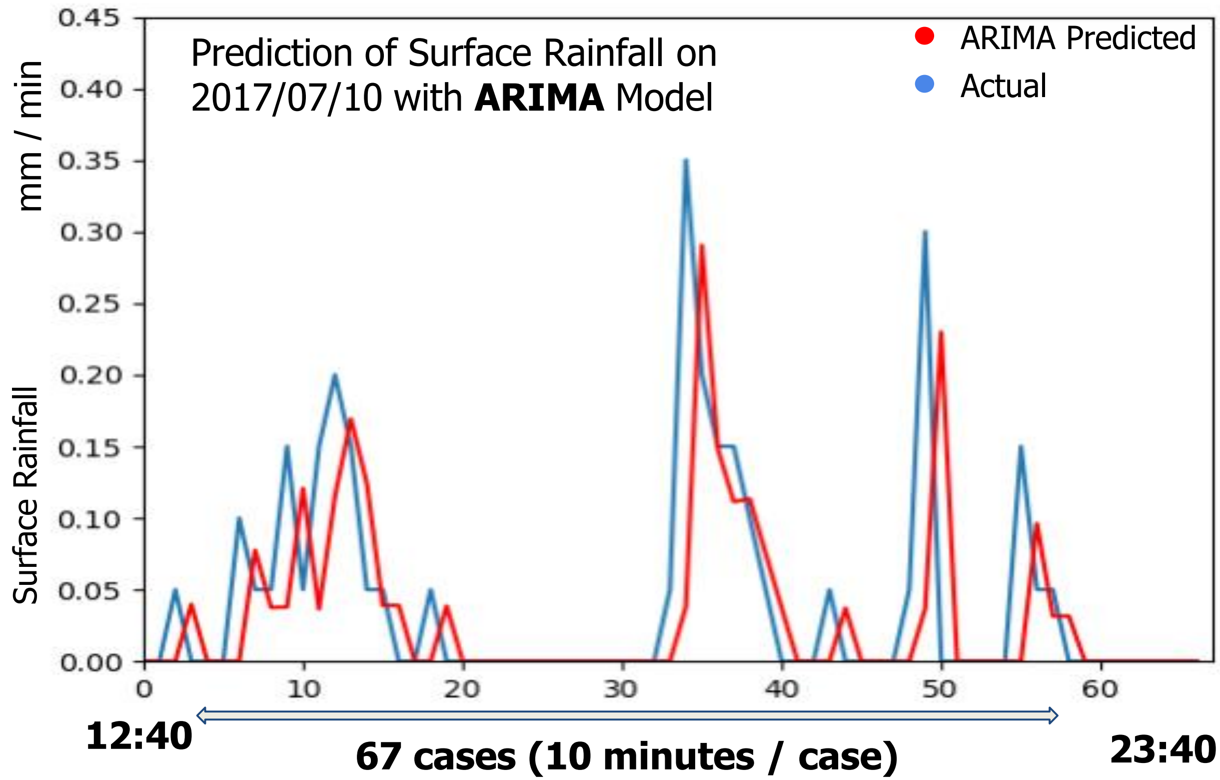


Figure 15: Prediction Result of AWS 560

Best Fitted Model

$$\hat{y}_t = \phi_1 y_{t-1} + e$$

ARIMA Model depends on its past value y_{t-1} and estimator ϕ_1

ARIMA(p,d,q)	Equation (without constant)	Mean Square Error	Mean Absolute Error
ARIMA(1,0,0)	$\hat{y}_t = \phi_1 y_{t-1} + e$	0.005 mm / min	0.034 mm / min

Phase 1: Summary of Temporal Correlation Analysis

◆ **Summary**

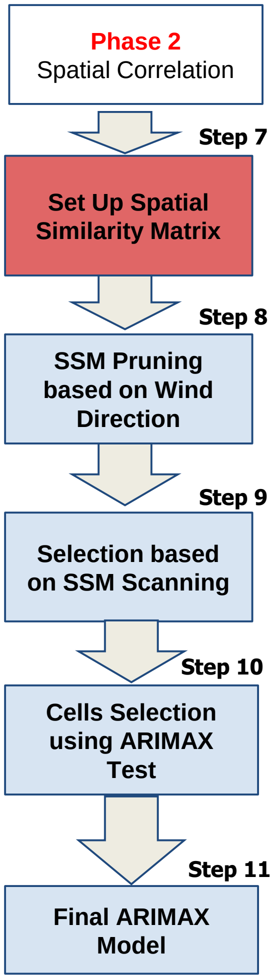
➔ Based on Table 1 and Table 2, ARIMA (1,0,0) is our best temporal model.

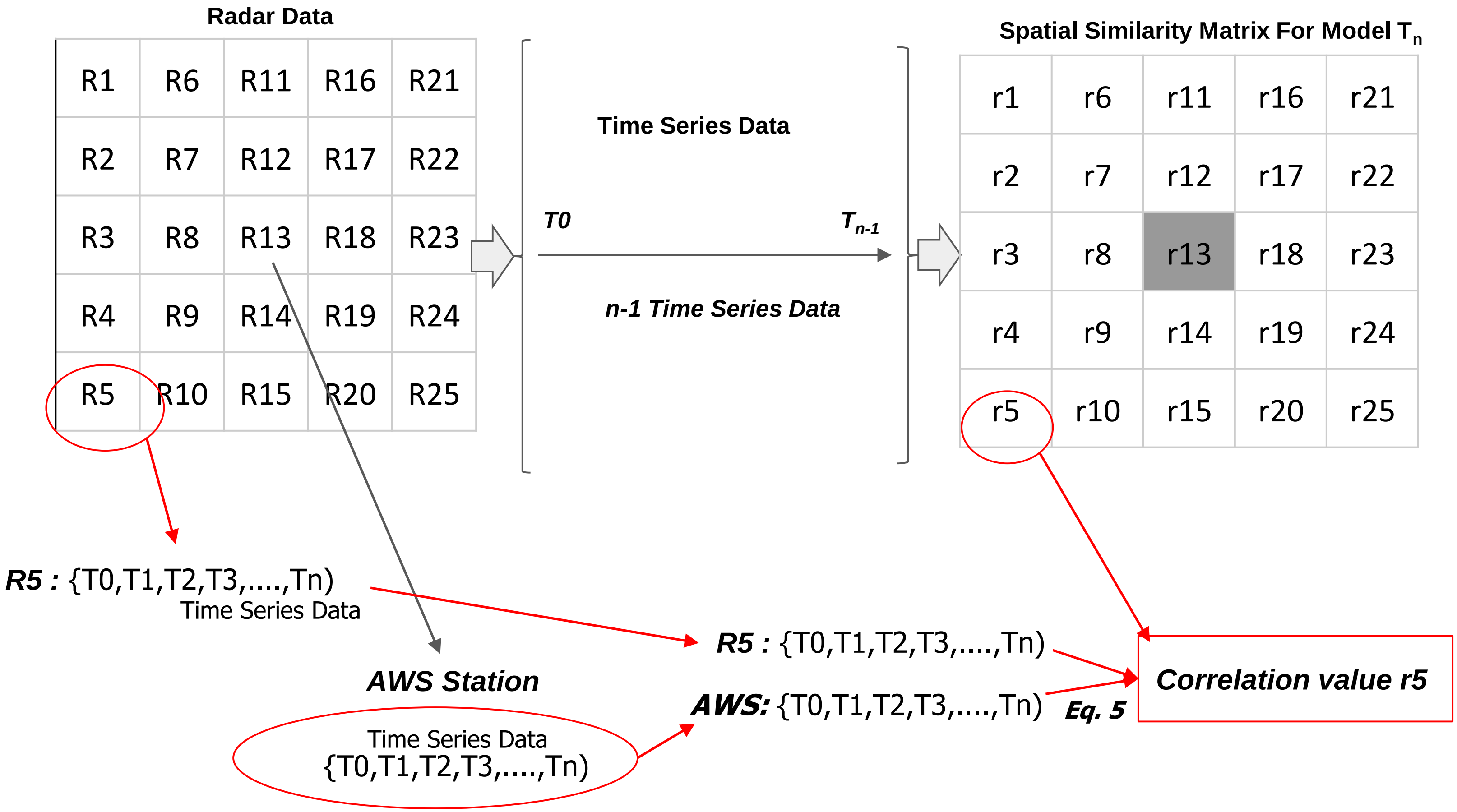
➔ ARIMA(1,0,0) will be the based temporal model for proceeding to Phase 2.

Next Phase

Phase 2: Spatial Correlation Analysis

Step 7: Define Spatial Similarity Matrix (SSM)





Pattern Similarity Check : (***Pearson Correlation Equation 5***)

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

r : Autocorrelation Value
X_i : One of Radar Pixel Data in our case
Y_i : AWS Data in our case
$\bar{x}$: Mean of X
$\bar{y}$: Mean of Y

Step 7: Example of 7 * 7 Pixels SSM (14 kilometer Square)

Phase 2
Spatial Correlation

Step 7

Set Up Spatial Similarity Matrix

Step 8

SSM Pruning based on Wind Direction

Step 9

Selection based on SSM Scanning

Step 10

Cells Selection using ARIMAX Test

Step 11

Final ARIMAX Model

SSM of 7 * 7 Pixel 20170702 [1:30 - 3:50] - Sampling Example

0.23	0.44	0.52	0.60	0.64	0.80	0.70
0.4	0.35	0.41	0.53	0.60	0.70	0.61
0.5	0.42	0.57	0.55	0.56	0.58	0.48
0.55	0.59	0.77	0.61	0.50	0.48	0.42
0.80	0.85	0.79	0.65	0.60	0.13	0.23
0.71	0.72	0.39	0.54	0.64	0.26	0.04
0.58	0.40	0.23	0.34	0.58	0.28	-0.17

How to Select the most influential cells with SSM?

Selection Criteria 1

SSM Pruning based on Wind Direction
Step 8

Selection Criteria 2

Selection based on SSM Scanning
Step 9

Selection Criteria 3

Selection based on ARIMAX Testing
Step 10

Step 8: SSM Pruning based on Wind Direction (Concept)

Phase 2
Spatial Correlation

Step 7

Set Up Spatial Similarity Matrix

Step 8

SSM Pruning based on Wind Direction

Step 9

Selection based on SSM Scanning

Step 10

Cells Selection using ARIMAX Test

Step 11

Final ARIMAX Model

SSM Before Pruning

Wind Directions

r1	r8	r15	r22	r29	r36	r43
r2	r9	r16	r23	r30	r37	r44
r3	r10	r17	r24	r31	r38	r45
r4	r11	r18	r25	r32	r39	r46
r5	r12	r19	r26	r33	r40	r47
r6	r13	r20	r27	r34	r41	r48
r7	r14	r21	r28	r35	r42	r49

SSM After Pruning (Case 8 Pruning)

Wind Directions

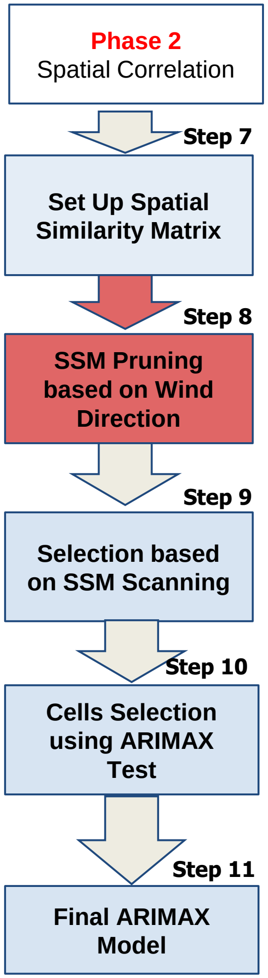
r1	r8	r15	r22	r29
r2	r9	r16	r23	r30
r3	r10	r17	r24	r31
r4	r11	r18	r25	r32
r5	r12	r19	r26	r33

Pruned Cells

Case No.	1 (N)	2 (NE)	3 (E)	4 (SE)	5 (S)	6 (SW)	7 (W)	8 (NW)
Direction Range	337.5 - 22.5 (deg)	22.5 - 67.5 (deg)	67.5 - 112.5 (deg)	112.5 - 157.5 (deg)	157.5 - 202.5 (deg)	202.5 - 247.5 (deg)	247.5 - 292.5 (deg)	292.5 - 337.5 (deg)

Goal: To make sure that are used is **consistent** with the **wind direction**.

Step 8: SSM Pruning based on Wind Direction (Example)



SSM of 7 * 7 Pixel 20170702 [1:30 - 3:50]

Wind Directions

314 Degree Angle

Remaining Area

Pruned Area

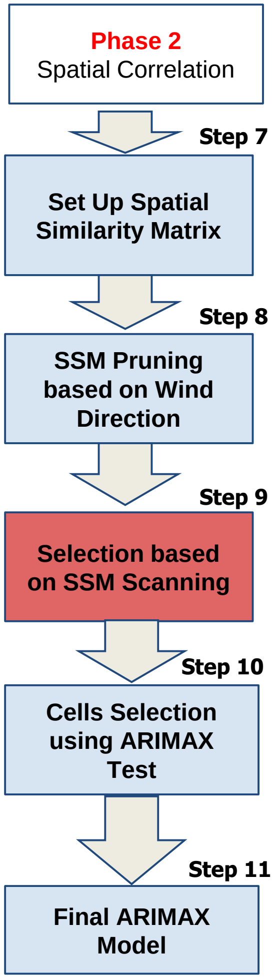
0.23	0.44	0.52	0.60	0.64	0.80	0.70
0.4	0.35	0.41	0.53	0.60	0.70	0.61
0.5	0.42	0.57	0.55	0.56	0.58	0.48
0.55	0.59	0.77	0.61	0.50	0.48	0.42
0.80	0.85	0.79	0.65	0.60	0.13	0.23
0.71	0.72	0.39	0.54	0.64	0.26	0.04
0.58	0.40	0.23	0.34	0.58	0.28	-0.17

SSM After Pruning

0.23	0.44	0.52	0.60	0.64
0.4	0.35	0.41	0.53	0.60
0.5	0.42	0.57	0.55	0.56
0.55	0.59	0.77	0.61	0.50
0.80	0.85	0.79	0.65	0.60

Example of Pruning SSM

Step 9: Defining Scanning Range for SSM (Spatial Similarity Matrix)



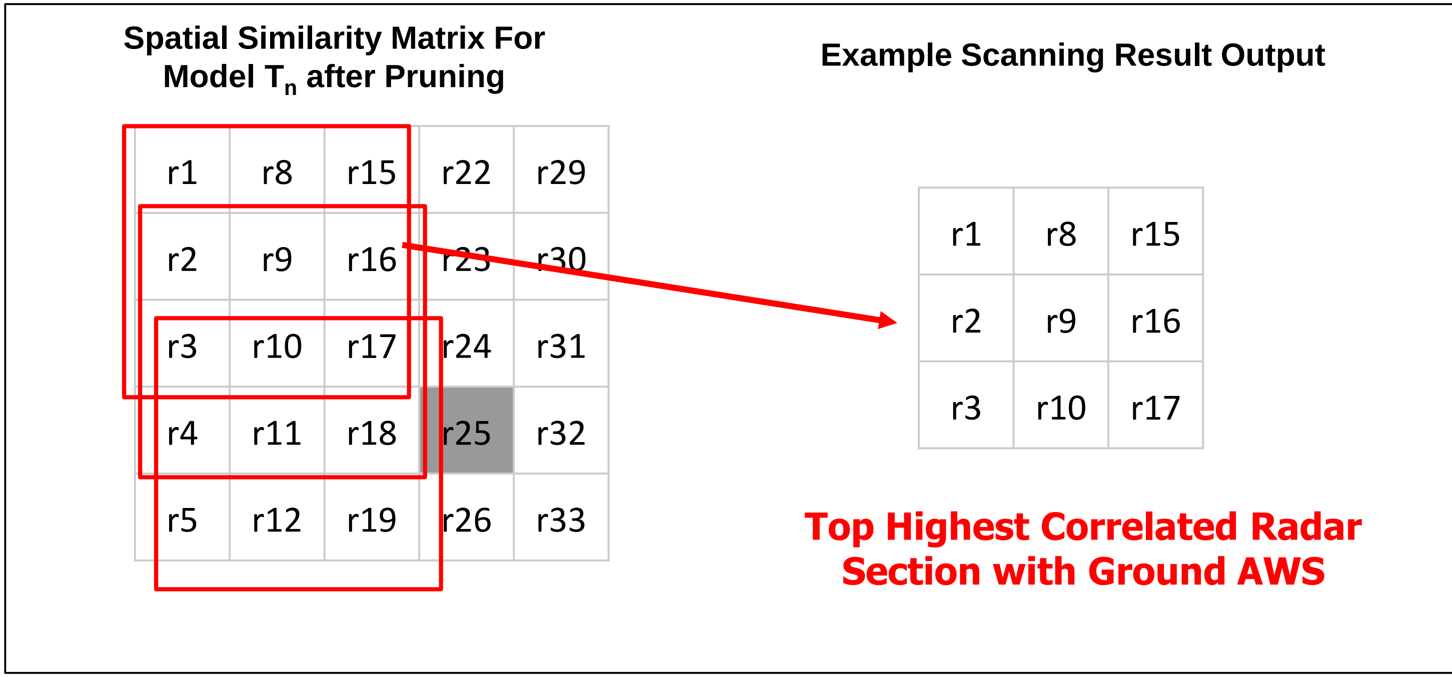
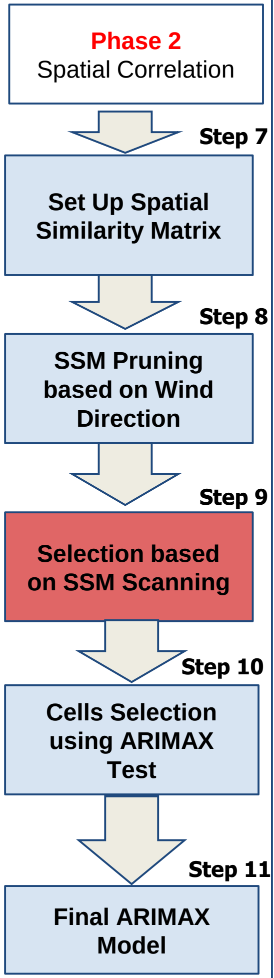


Figure 16: Matrix Scanning Concept

Description:

- Select out 3 * 3 Matrix from the 5 * 5 result
- Criteria of selection is selecting **the top cumulative correlation value of all cell.**

Step 9: Example of Scanning SSM after Pruning



Spatial Similarity Matrix For Model 4:00 from Step 7

0.23	0.44	0.52	0.60	0.64
0.4	0.35	0.41	0.53	0.60
2 0.5	9 0.42	16 0.57	0.55	0.56
3 0.55	10 0.59	17 0.77	0.61	0.50
4 0.80	11 0.85	18 0.79	0.65	0.60

0.5	0.42	0.57
0.55	0.59	0.77
0.80	0.85	0.79

SSM Scan Result

Description:

- SSM Scan Result is the top cumulative 3*3 matrix from the origin matrix.

Step 10: Cells Selection Using ARIMAX Test (Concept)

Phase 2
Spatial Correlation

Step 7

Set Up Spatial Similarity Matrix

Step 8

SSM Pruning based on Wind Direction

Step 9

Selection based on SSM Scanning

Step 10

Cells Selection using ARIMAX Test

Step 11

Final ARIMAX Model

Scanning Output from Step 9

r1	r8	r15
r2	r9	r16
r3	r10	r17

Top Highest Correlated Radar Section with Ground AWS

1. Should all cells be used for modelling?
2. Should we just use some cells?

Comparison Test Based on 2 days and 3 stations

Scenario 1

Use only 3 top variables

Scenario 2

Use only 6 top variables

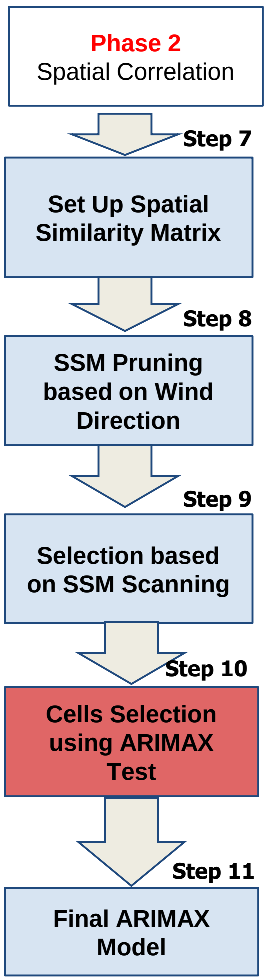
Scenario 3

Use only 9 top variables

Description:

- From the previous Step 9 results, we have 3 * 3 matrix with 9 cells
- We test from 3 variables as predictor variable to 9 variables to compare the accuracy

Step 10: Prediction Experiments Explanation



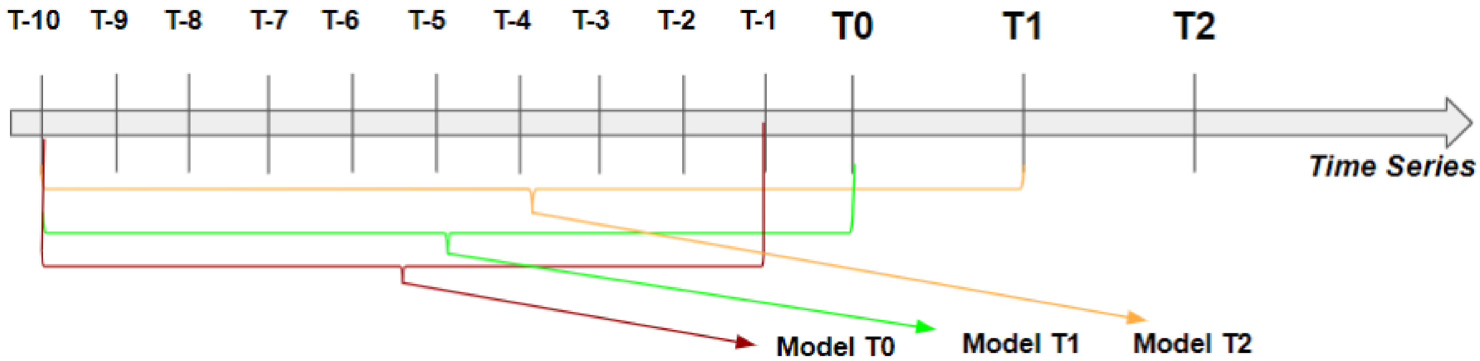


Figure 17: Model Generating Concept Using Past Data

Key Data Point Explanation:

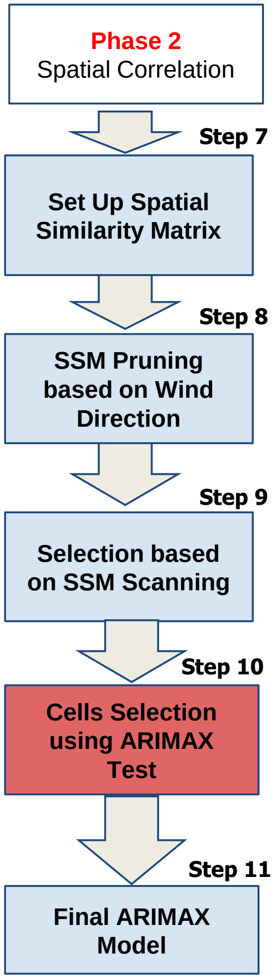
- **T_{-10}** to **T_{-1}** is the train data and the first prediction model is T_0 Model.
- **T_{-10}** to **T_0** is the train data for prediction model of T_1 Model.

Note:

Example: 3 Cells Selection means using 3 spatial cells:

$$\hat{Y}_t = \phi_1 Y_{t-1} + \boxed{\beta(R1_t - \phi_1 R1_{t-1}) + \beta(R2_t - \phi_1 R2_{t-1}) + \beta(R3_t - \phi_1 R3_{t-1})}$$

Step 10: Testing Result and Summary of Results



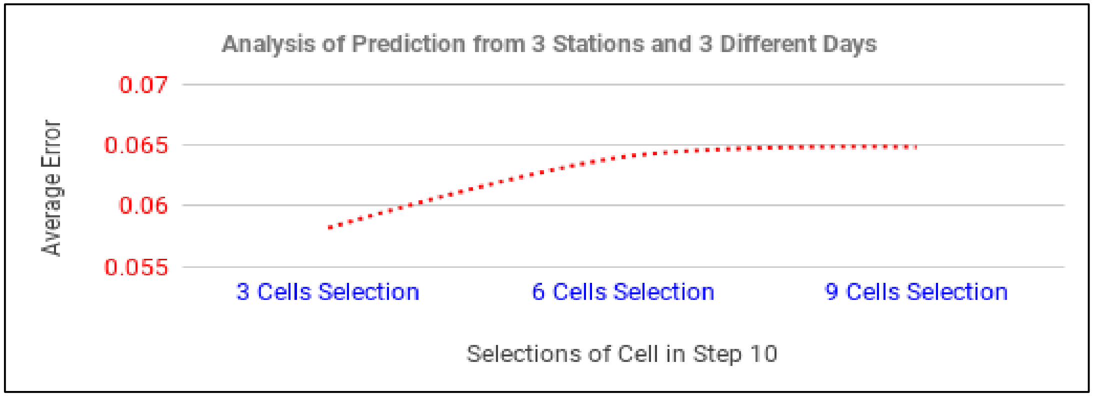


Figure 18: Result of Using Different Amount of Radar Cells

Analysis:

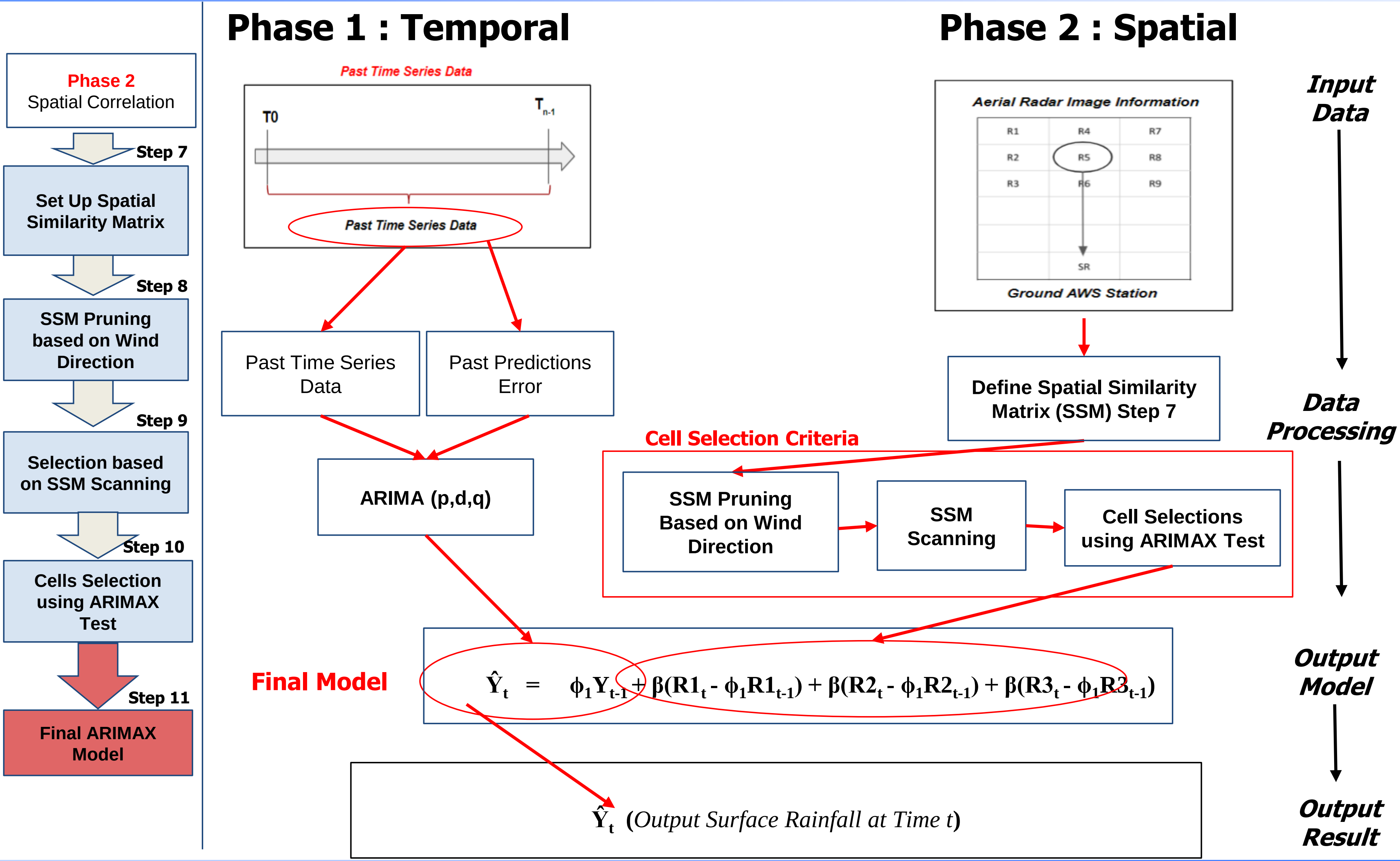
3 Cells Selections provide the lowest error rates and will be chosen for the final model.

Table 3: Testing and Accuracy of Figure 18 in Detail

Station ID	*Day*	*3 Cells Selections*	*6 Cells Selections*	*9 Cells Selections*
556	2017-07-02 **(1)**	0.026	0.027	0.028
556	2017-07-10 **(2)**	0.070	0.080	0.086
560	2017-07-02 **(1)**	0.067	0.064	0.070
560	2017-07-10 **(2)**	0.033	0.039	0.043
561	2017-07-02 **(1)**	0.106	0.126	0.115
561	2017-07-10 **(2)**	0.047	0.048	0.047
Average Error		0.0581 mm / min	0.064 mm / min	0.0648 mm / min

Cases from Different Stations and Different Days

Summary of Phase 1 and Phase 2 Process (Flow and Concept)

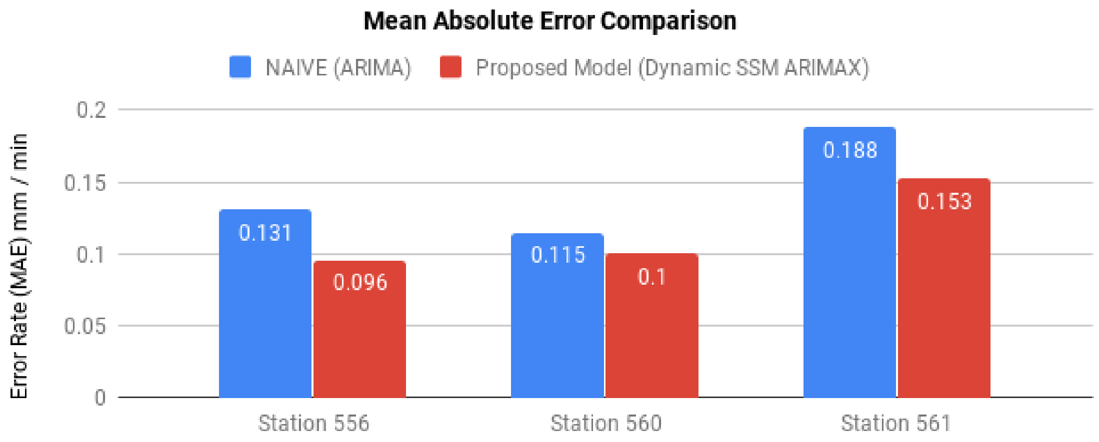


7. Discussion of Experimental Result – Error Ratio

Analysis:

- Experiment on Different Location on Two Different Days (2017-07-02 and 2017-07-10)

- Our Proposed Model provides less error rate compare to the Naive Model



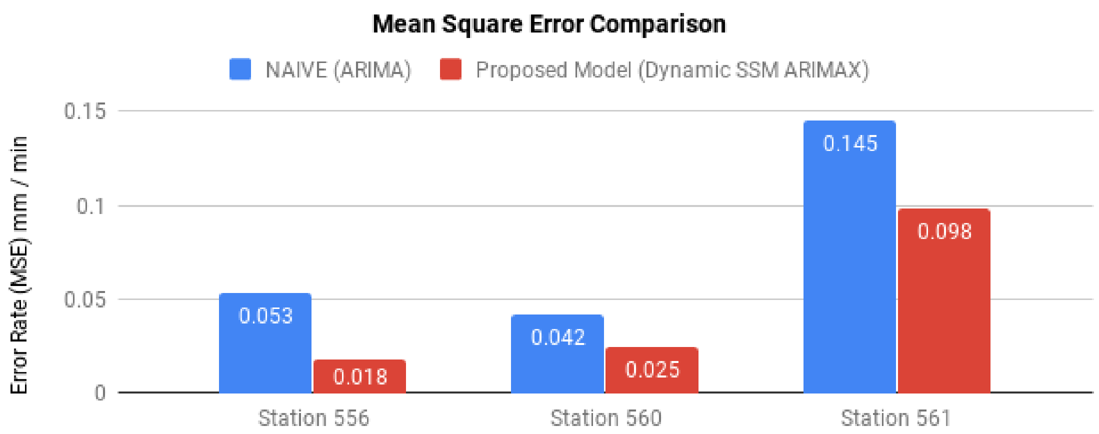


Figure 19: Absolute Error Rate Comparison Experiment

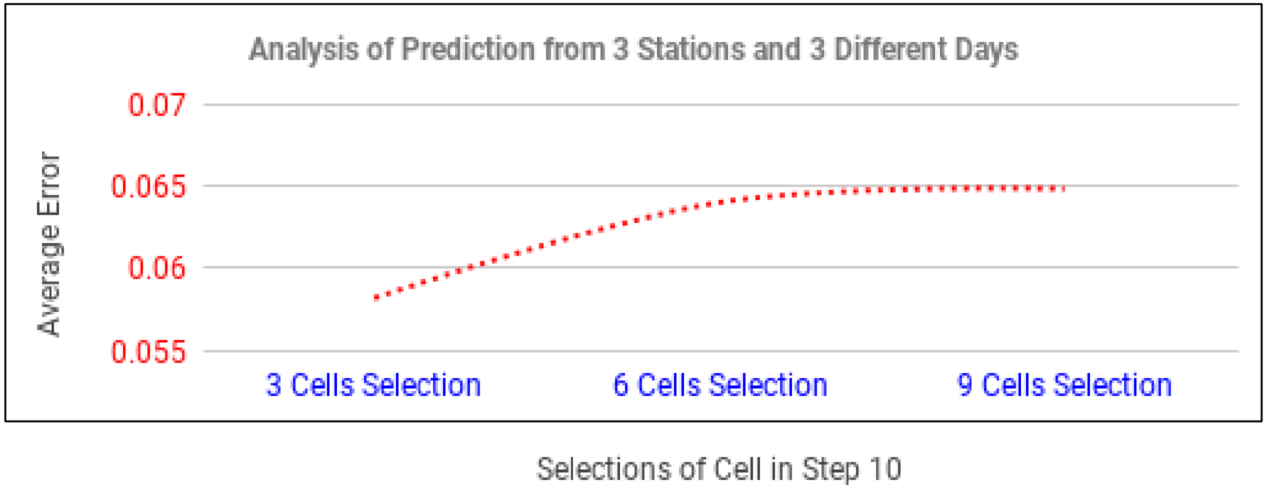


Figure 20: Increasing Spatial Variables Experiments

Analysis:

- Using **3 radar spatial variables** to predict surface rainfall provides less error
- Using many variables that has low correlation with ground rainfall could increase the error.

8. Conclusion & Future Works

Final Model:

Targeted SR   Temporal Data   Dynamic Spatial Data (Radar)

$$\boxed{\hat{Y}_t} = \boxed{\phi_1 Y_{t-1} \; +} \; \boxed{\beta(R1_t - \phi_1 R1_{t-1}) \; + \; \beta(R2_t - \phi_1 R2_{t-1}) \; + \; \beta(R3_t - \phi_1 R3_{t-1})}$$

- Surface Rainfall can be estimated using its past rainfall value by using **ARIMA (1,0,0) as the prediction model**

- **Integration of Radar Images Information** will increase the accuracy.

- **Spatial Similarity Matrix Scanning** for high similarity section is necessary to improve the accuracy of the prediction.

- **Future Works:**
  - Compare the prediction model with (Recurrent Neural Network)RNNs Model.

Reference

1. Korea Meteorological Administration: 2017 Radar Image, Radar Footage Open Portal, url: data.kma.go.kr,2017
2. Korea Meteorological Administration: 2017 AWS Data, AWS Observation, url: data.kma.go.kr, 2017
3. Jiwan L, Yongdoek S, Bonghee H, "Extraction of Weather Information on Road using CCTV video" IEEE2016 International Conference on Big Data and Smart Computing, Jan 2016
4. Oudomseila P, Jiwan L, Bonghee H, "Rainfall Prediction Model based on Radar Image Analysis Processing", 3rd International Conference o n Internet of Things, Big Data and Security 2018
5. Oudomseila P, Jiwan L, Bonghee H, "Surface Rainfall Estimation Based on Radar Image Analysis and Fully Connected Neural Network", Ko rea Computer Congress 2018
6. Robert N. "ARIMA Models for Time Series Forecasting" https://people.duke.edu/~rnau/411arim.htm last accessed on 2019/03/22
7. Seoul National Unversity Lecture Note, "ARIMA 모형 (ARIMA Procedure)" http://stat.snu.ac.kr/time/download/%EC%8B%A4%EC%8A%B5%EA%B0%95%EC%9D%983.pdf , last accessed on 2019/03/22
8. Kwon S, Jwae O, Hand Y, "Rainfall Forecasting using Data Mining and Deep Learning", Graduation Final Report of Pusan National Universi ty, 2017
9. Jiwan L, Yongdeok S., Bonghee H, "Extraction of Weather Information on Road using CCTV Video", IEEE 2016 International Conference o n Big Data and Smart Computing.
10. Leo E., Leot L., "The Relation Between Pearson's Correlation Coefficient r and Salton's Cosine Measure", Journal of the American Society f or Information Science & Technology (forthcoming)
11. Udom P., Phumchusri N., "A Comparison study between time series molde and ARIMA model for sales forecasting of distributor in plastic Industry", https://pdfs.semanticscholar.org/23db/07fb90ad62e8b53fac74bfbf855ba205d1cd.pdf last accessed on 2019/04/14
12. Lan P. S., "Statistics: Regression and Time Series Analysis", https://sites.psu.edu/movingpsychology/2012/11/29/statistics-regression-and-time-series-analysis/ last accessed on 2019/04/17
13. Oudomseila P, Jiwan L, Bonghee H, "Analysis of Train Data Range, Time Interval Gap, Radar Spatial Range Impacts on Surface Rainfall E stimation Model", Korean Database Conference 2018
14. Oudomseila P, Jiwan L, Bonghee H, "Time Series Radar Matrix Based Data Analaysis for Surface Rainfall Estimation", Korean Database So ciety Journal, Volume 35 Number 1, April 2019