Flood in Jakarta

ADI WIJAYA SUCHIANA

Outline

- Background – Define the problem
  - Data availability
- Methodology and Result
  - Gathering data
  - Data cleansing
  - Exploratory data analysis
  - Modeling
  - Implementation
- Conclusion and Future works

Background

Flood prediction???

PETA JAKARTA

smart infrastructure facility

UNIVERSITY OF WOLLONGONG

JAKARTA SMART CITY

DATA SCIENCE INDONESIA

Data Availability

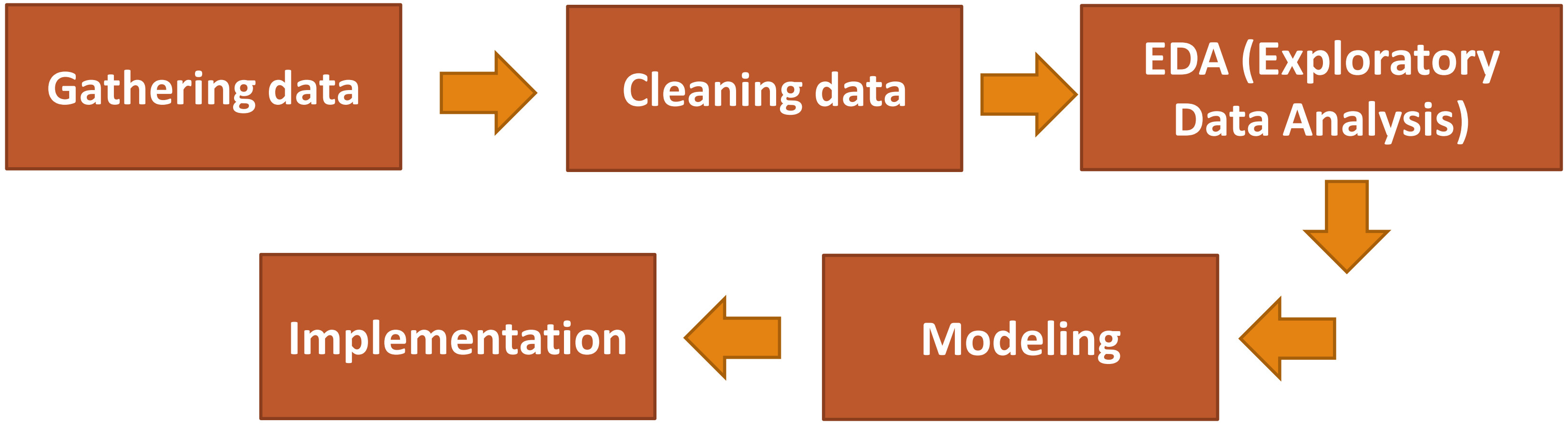
- Water level data di 11 pintu air (December 2014 – February 2015)
- Data kejadian banjir Jakarta (2013-2016)

Bendung Katulampa	Pos Angke Hulu
Pos Depok	Pasar Ikan
PA Manggarai	Pos Cipinang Hulu
PA Karet	Pos Sunter Hulu
Pos Krukut Hulu	PA Pulo Gadung
Pos Pesanggrahan	

How to predict flood using this data???

Methodology

Workflow


Gathering data
Cleaning data
EDA (Exploratory Data Analysis)
Modeling
Implementation

Gathering and Cleaning data

- Data yang berhasil didapatkan
  - Data ketinggian air (Januari 2014 – September 2016) (scraping from http://bpbd.jakarta.go.id/waterlevel)
  - Data category cuaca di lokasi pintu air (Terang, Mendung tipis, Mendung, Gerimis, dan Hujan)
  - Data average precipitation TRMM di semua pintu air(Januari 2014 – September 2016)
  - Data laporan banjir (Qlue, Twitter, detik)

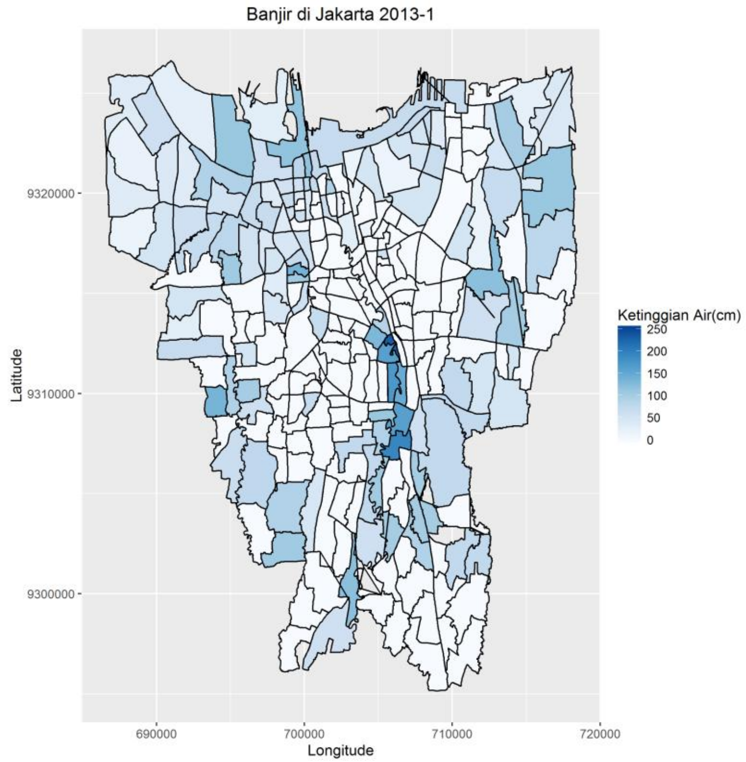
Data kejadian banjir di Jakarta (2013 – 2016)

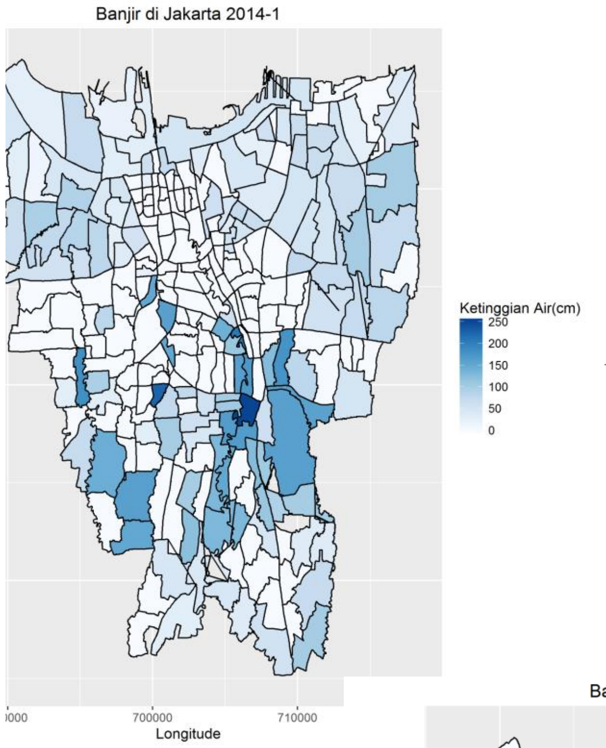
- Cleaning data
  - Pada data ketinggian air masih terdapat NA. Untuk mengatasi hal ini, dilakukan averaging pada data-data yang kosong dengan pertimbangan sifat data ketinggian air meningkat/menurun secara gradual dan tergantung dari nilai data jam – 1 dan data pada jam + 1

Exploratory Data Analysis

1. Data Kejadian Banjir (Tanggal, Rt, Rw, ketinggian air)

- Geocode daerah-daerah yang terkena banjir
- Dipasangkan sesuai region dengan data polygon Jakarta (sampai dengan tingkat kelurahan)
- Diambil ketinggian air rata-rata di daerah yang bersangkutan

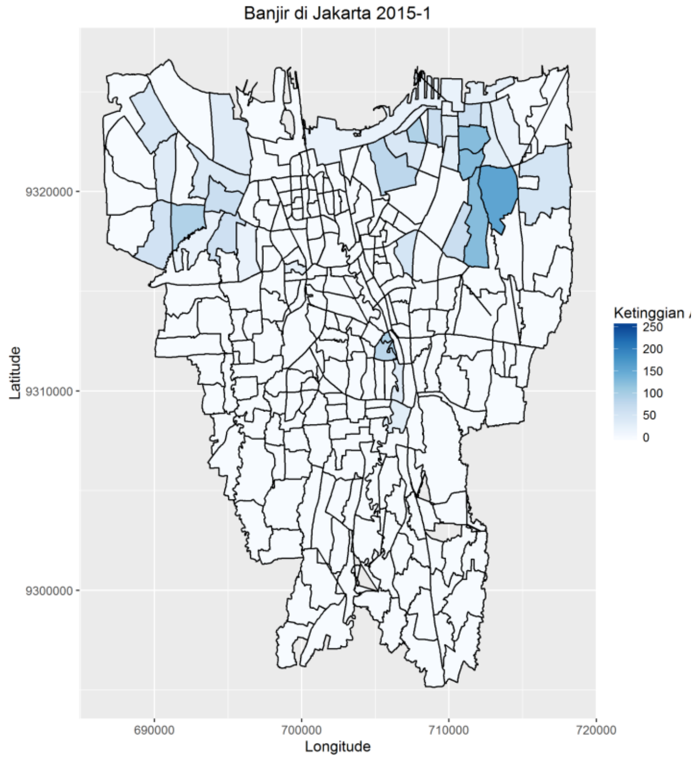


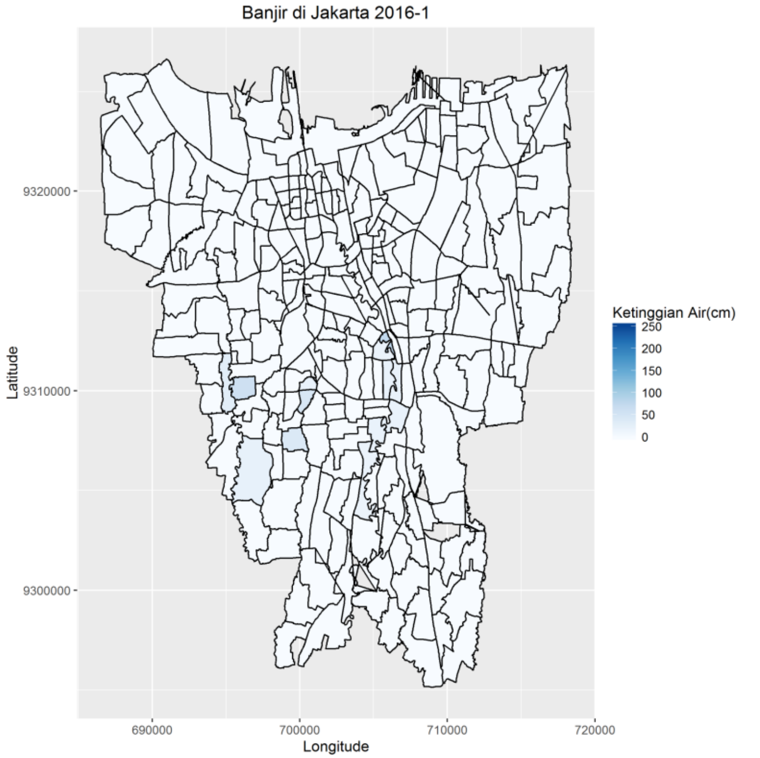


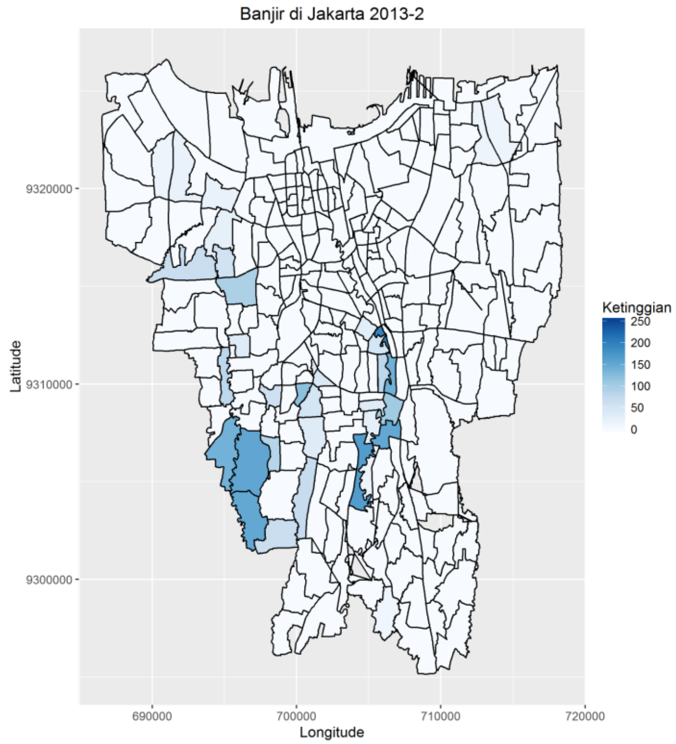
January

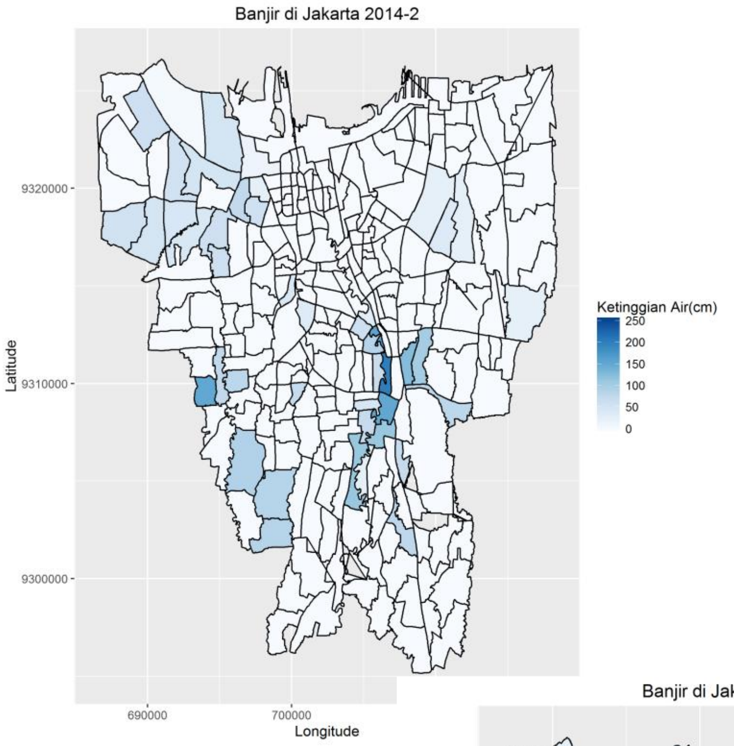
Exploratory Data Analysis.
Jakarta flood map from 2013 – 2016 .

No	Kelurahan	Average.Ketinggian.Banjir(cm)
1	BIDARA CINA	160
2	KAMPUNG MELAYU	190
3	CIPINANG MUARA	125
4	CIPINANG BESAR SELATAN	100
5	CAWANG	225
...	...	...



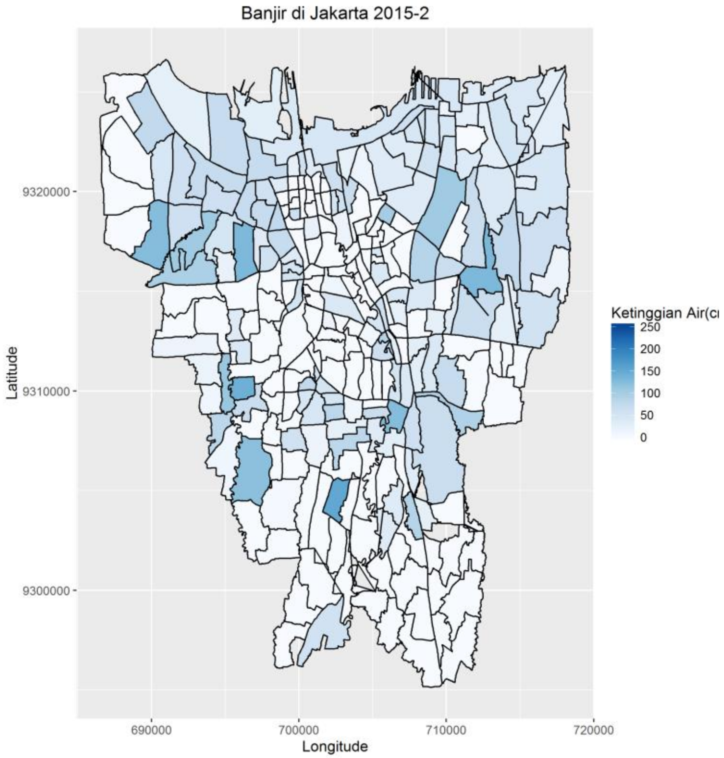


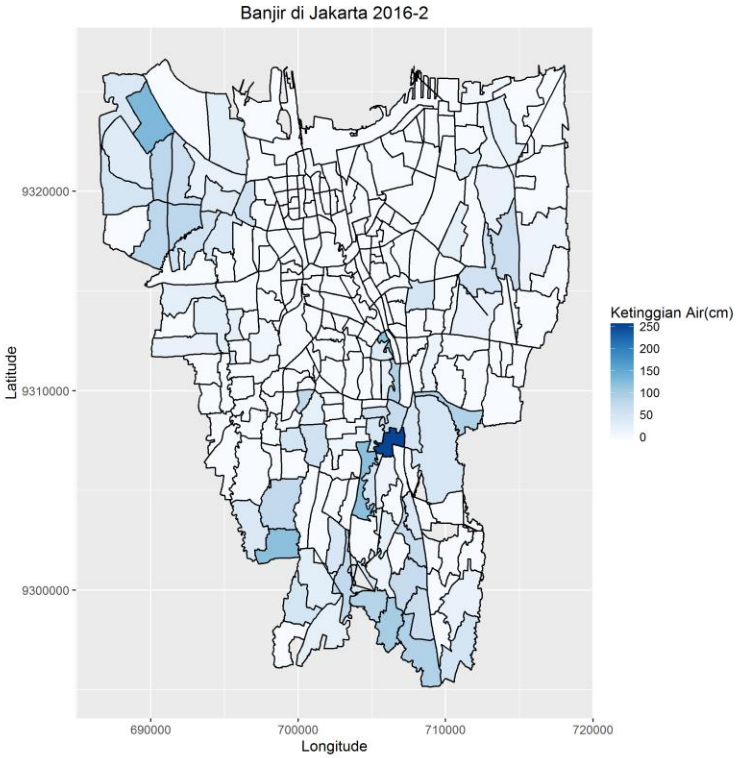




February

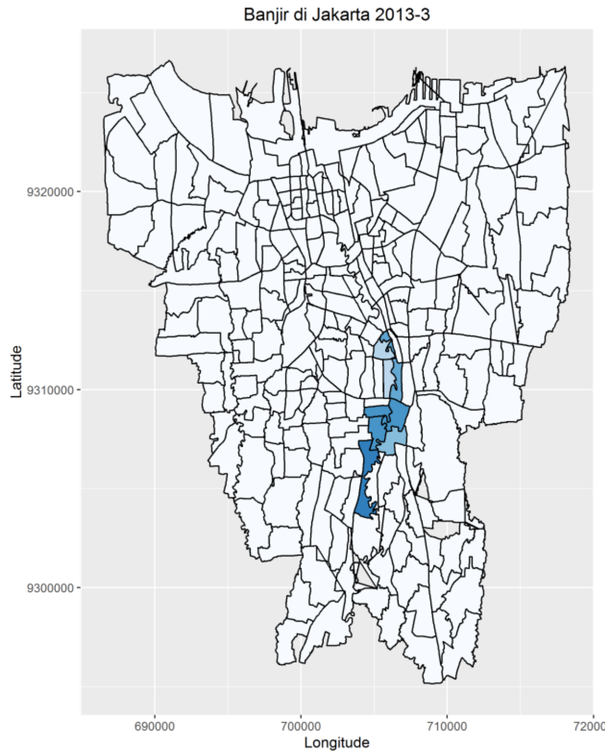
Exploratory Data Analysis.
Flood map shows that in 2013-2014 heavy flood occurred in January, while in 2015-2016 it was shifting to February.

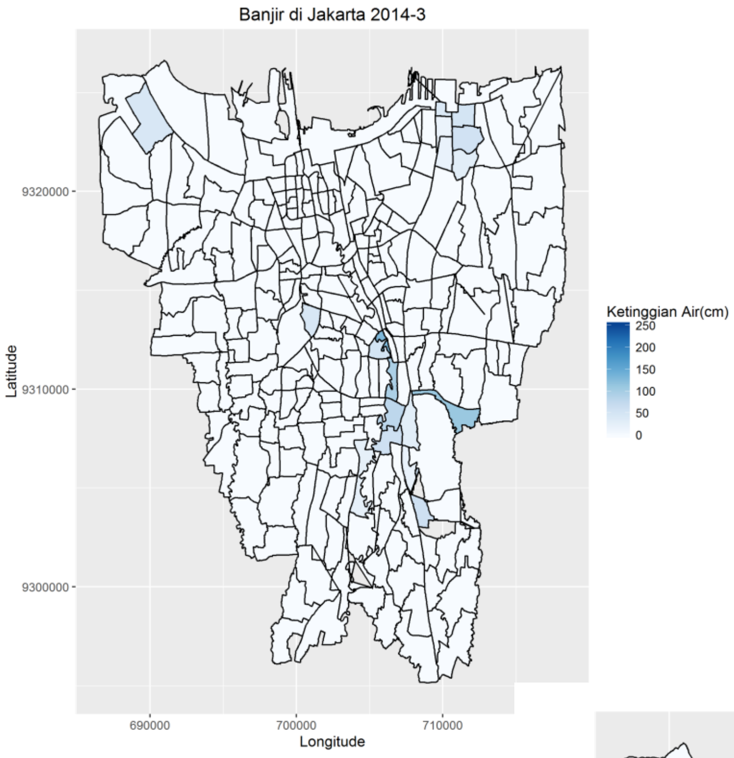


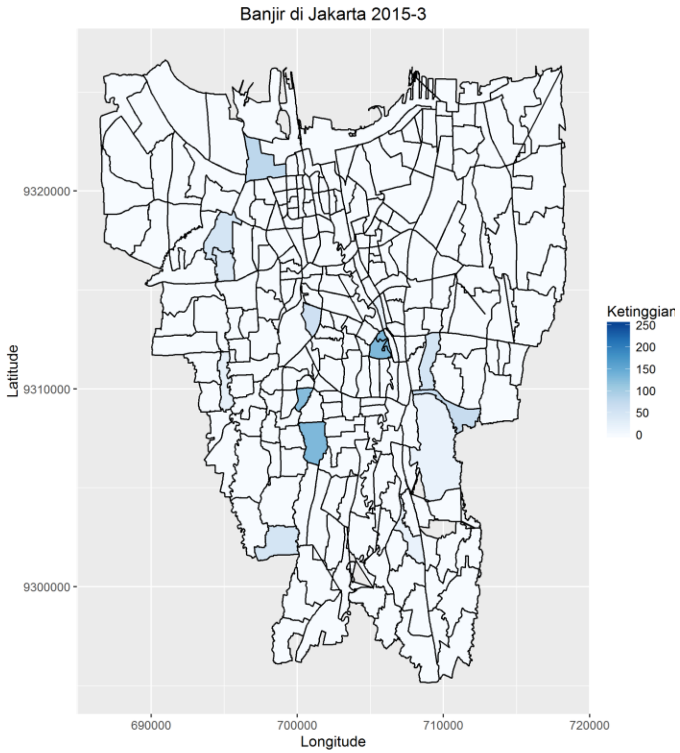


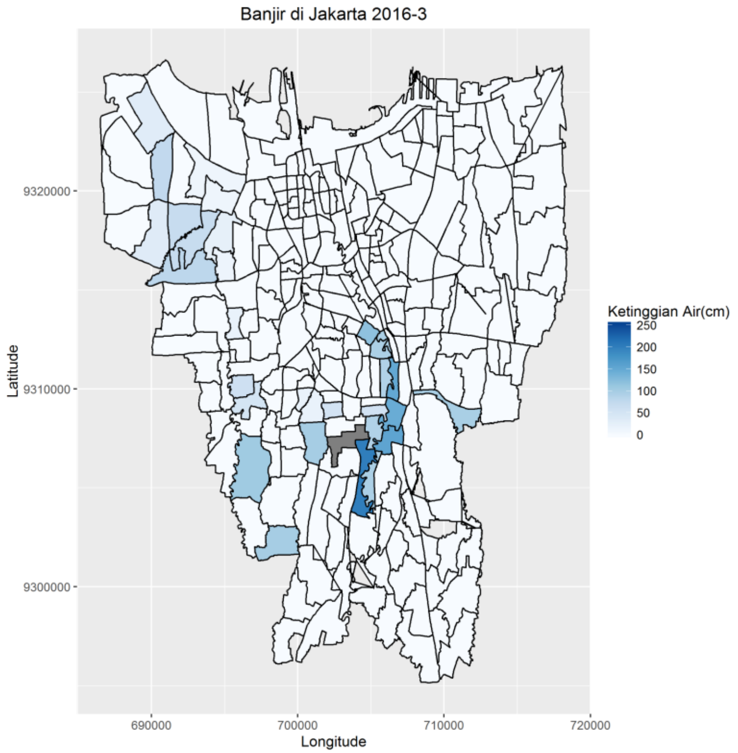
JAKARTA SMART CITY
DATA SCIENCE INDONESIA

March

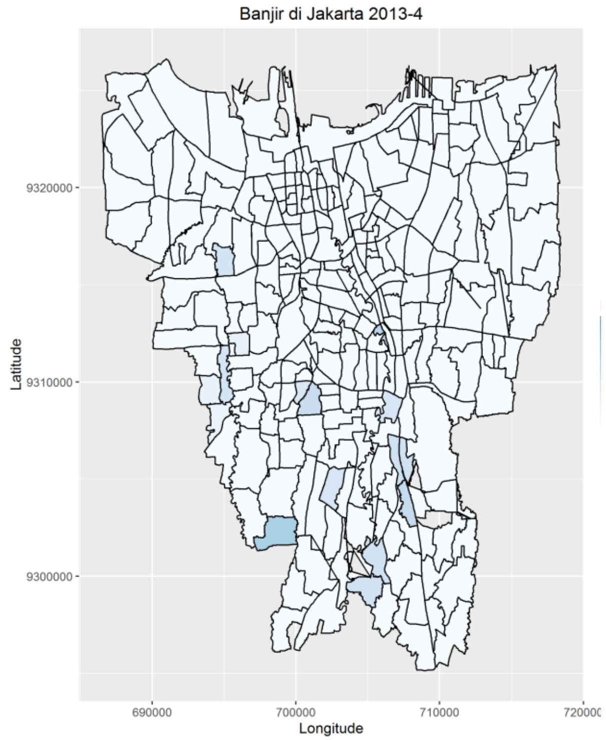

Banjir di Jakarta 2013-3
Latitude
9320000
9310000
9300000
690000
700000
710000
72000
Longitude


Banjir di Jakarta 2014-3
Latitude
9320000
9310000
9300000
690000
700000
710000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

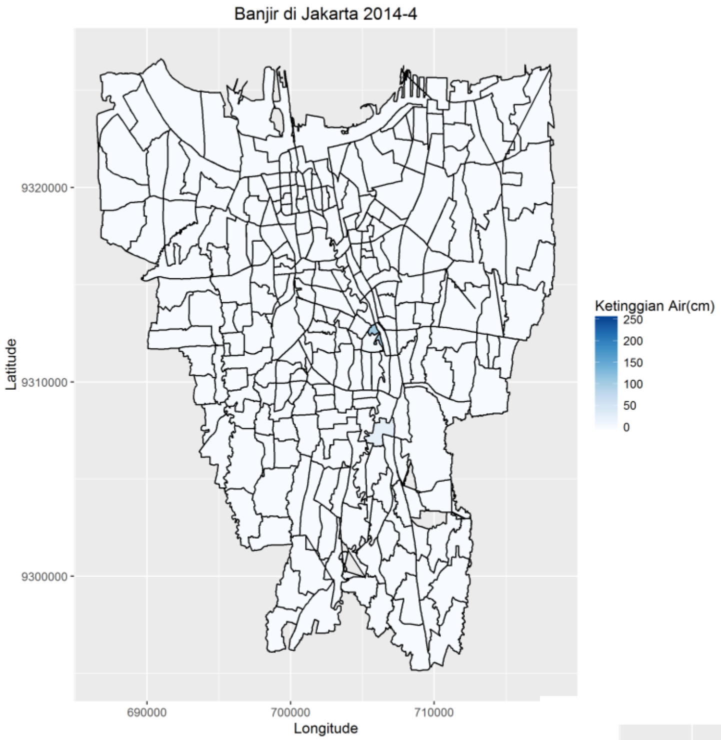

Banjir di Jakarta 2015-3
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude
Ketinggiar
250
200
150
100
50
0


Banjir di Jakarta 2016-3
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

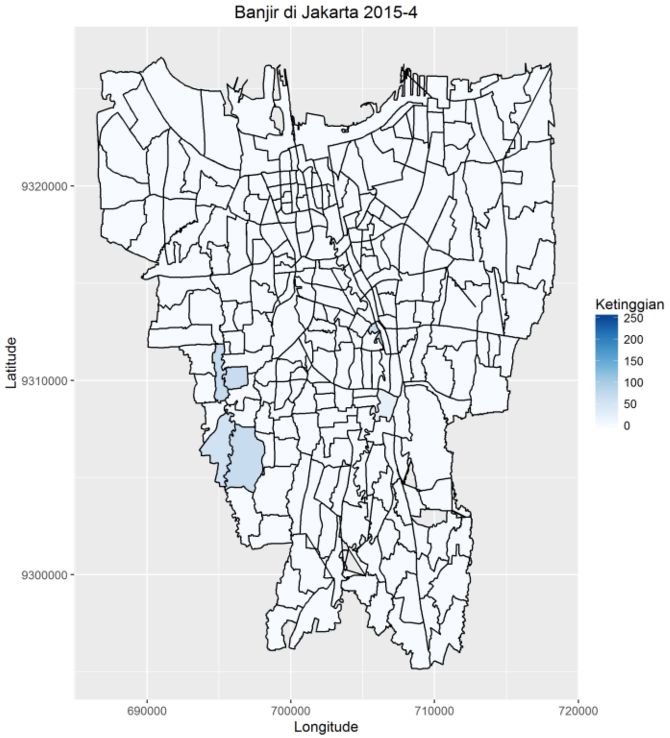
April



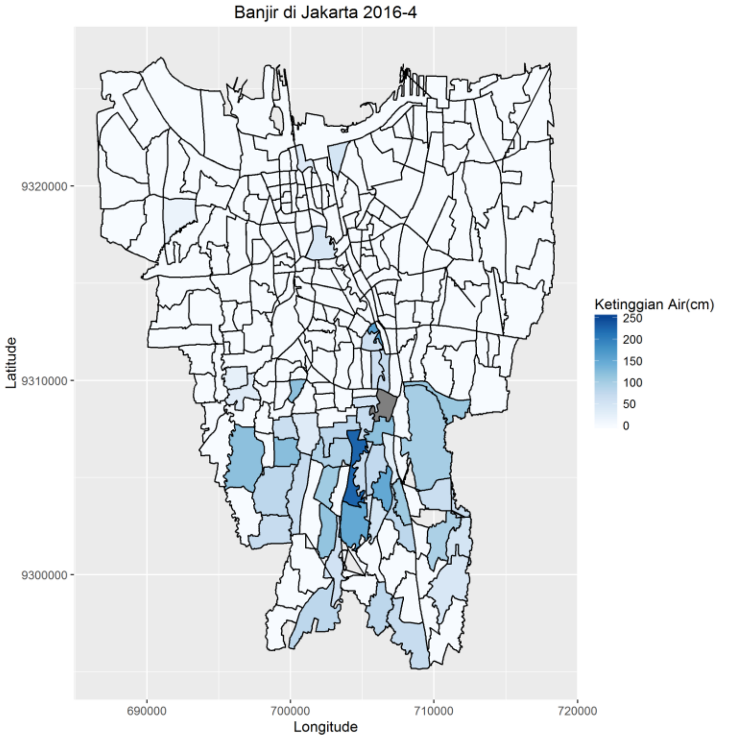
Banjir di Jakarta 2013-4
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
720000



Banjir di Jakarta 2014-4
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
Ketinggian Air(cm)
250
200
150
100
50
0

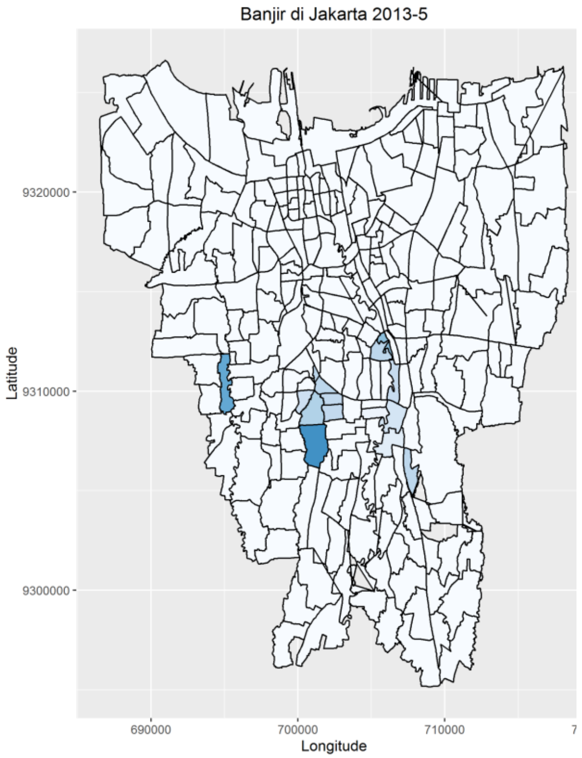


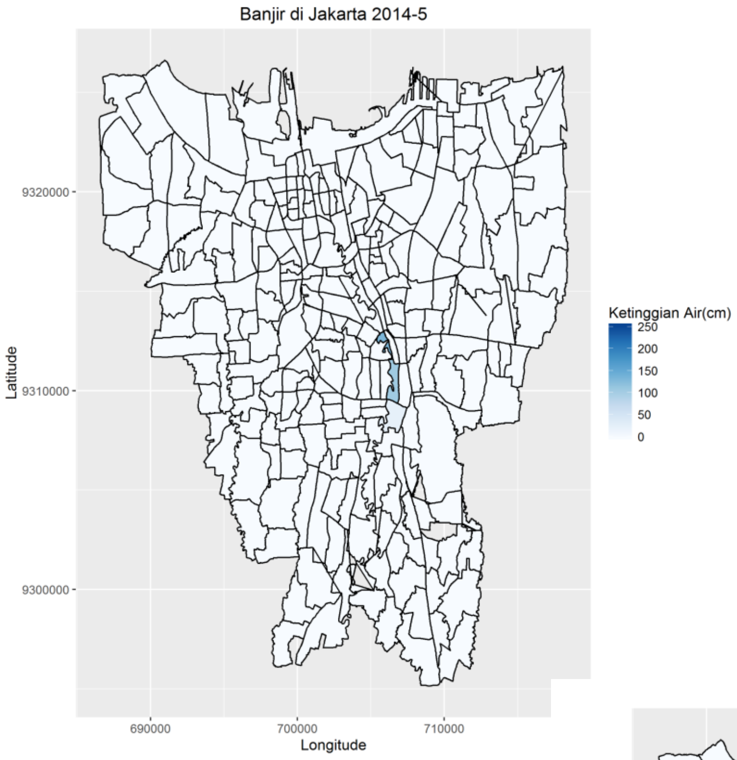
Banjir di Jakarta 2015-4
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
720000
Ketinggian
250
200
150
100
50
0

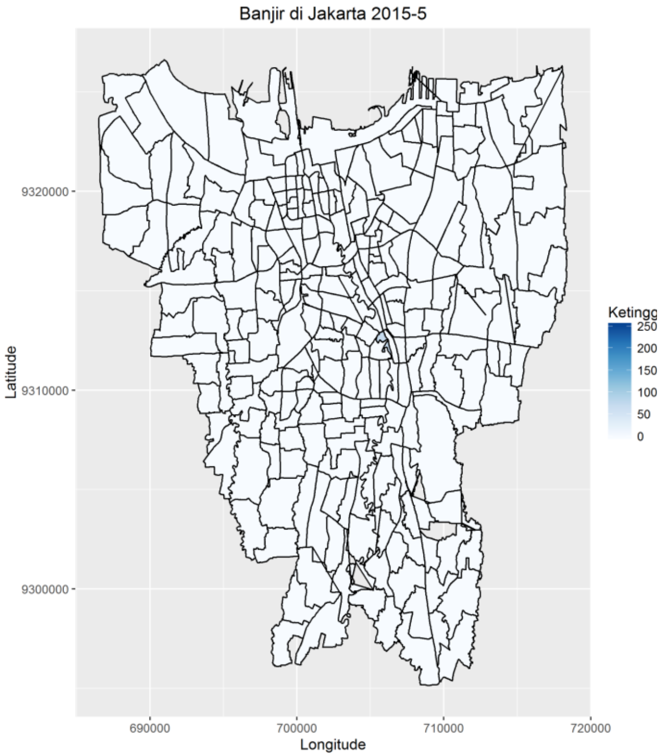


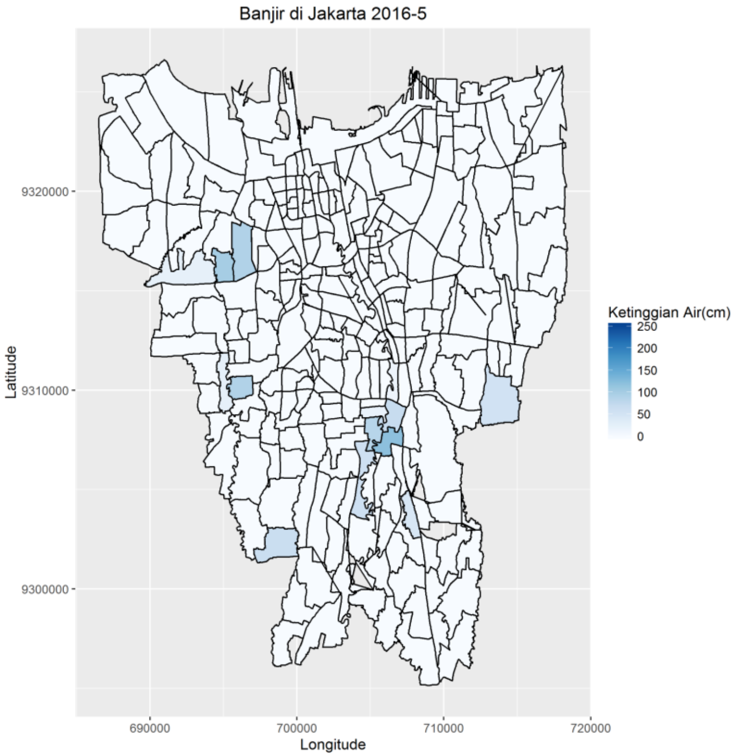
Banjir di Jakarta 2016-4
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
720000
Ketinggian Air(cm)
250
200
150
100
50
0

Mei



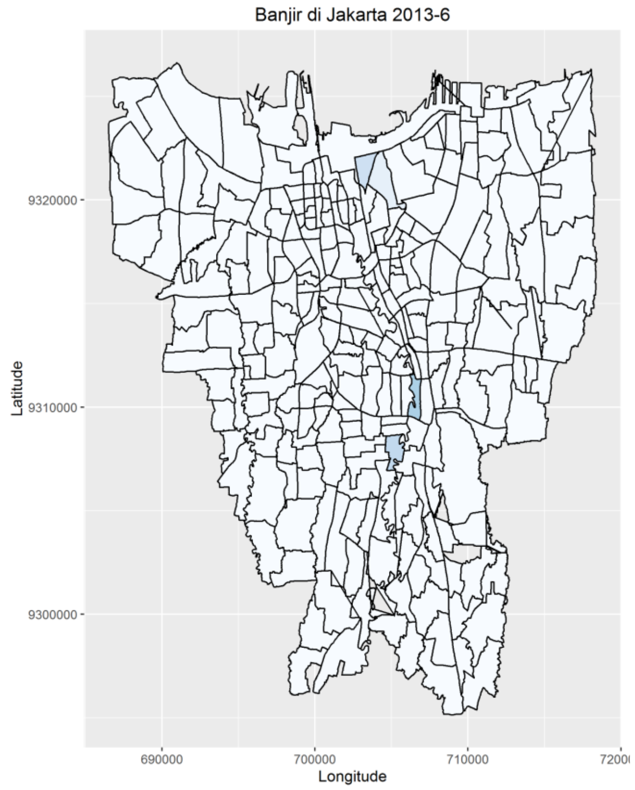


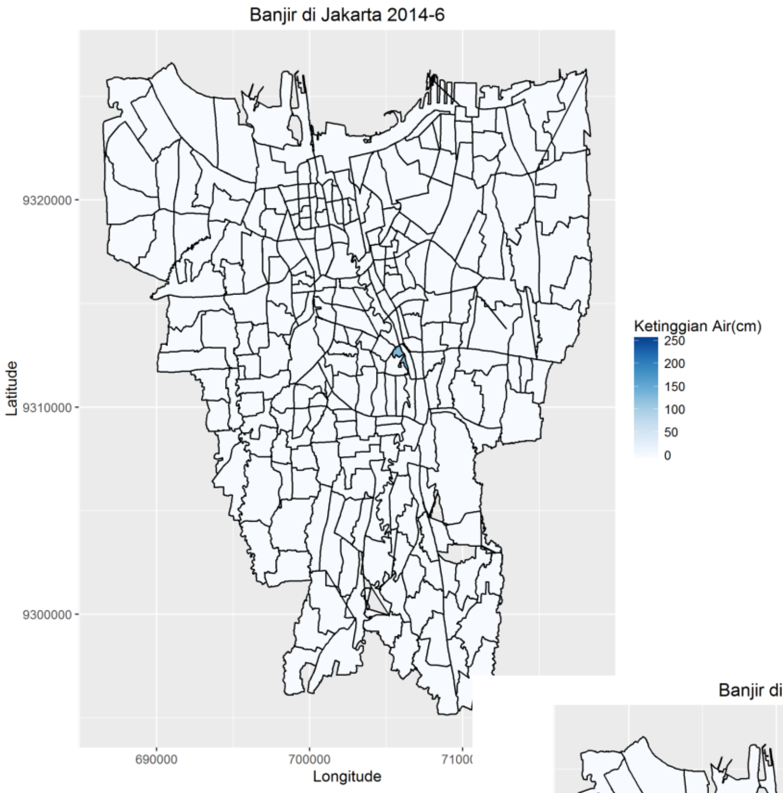




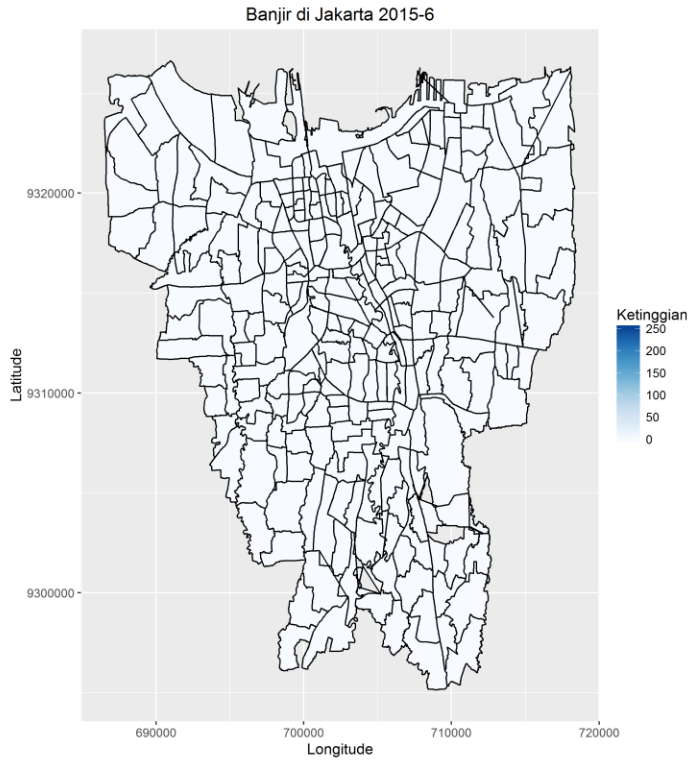
JAKARTA SMART CITY

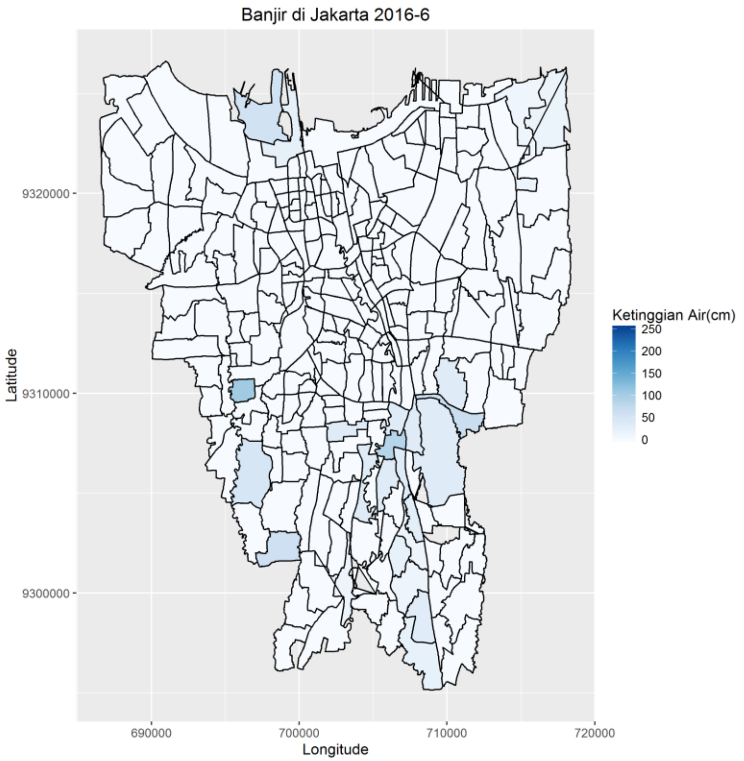
DATA SCIENCE INDONESIA

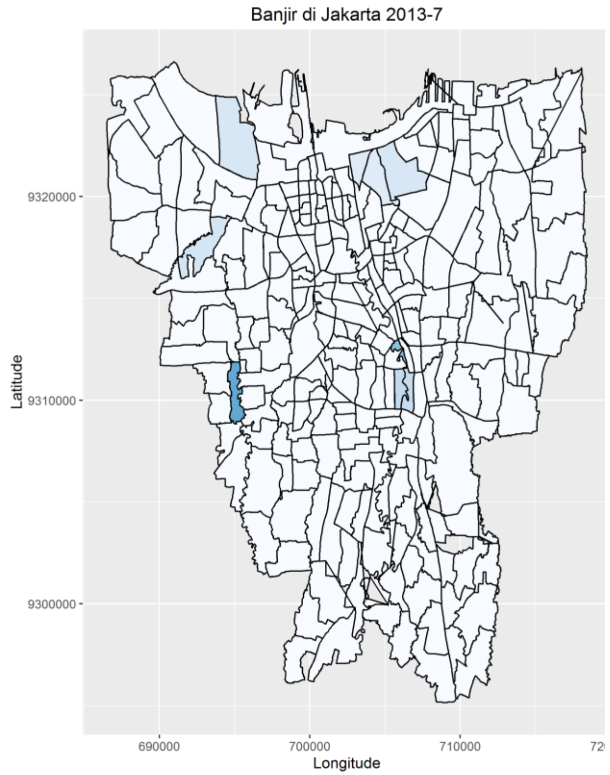

Banjir di Jakarta 2013-6
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000

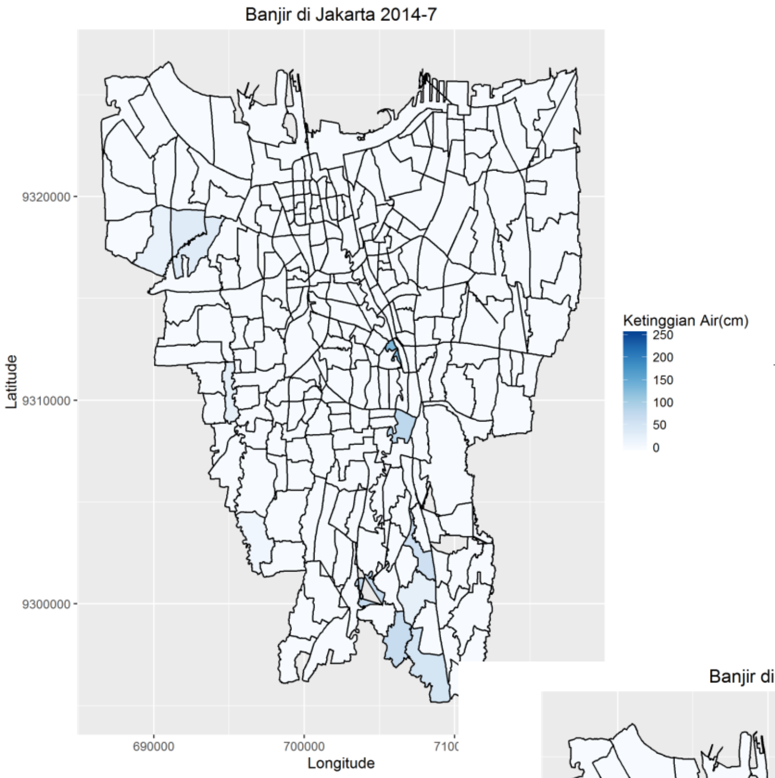

Banjir di Jakarta 2014-6
Latitude
Longitude
9320000
9310000
9300000
690000
700000
Ketinggian Air(cm)
250
200
150
100
50
0

June

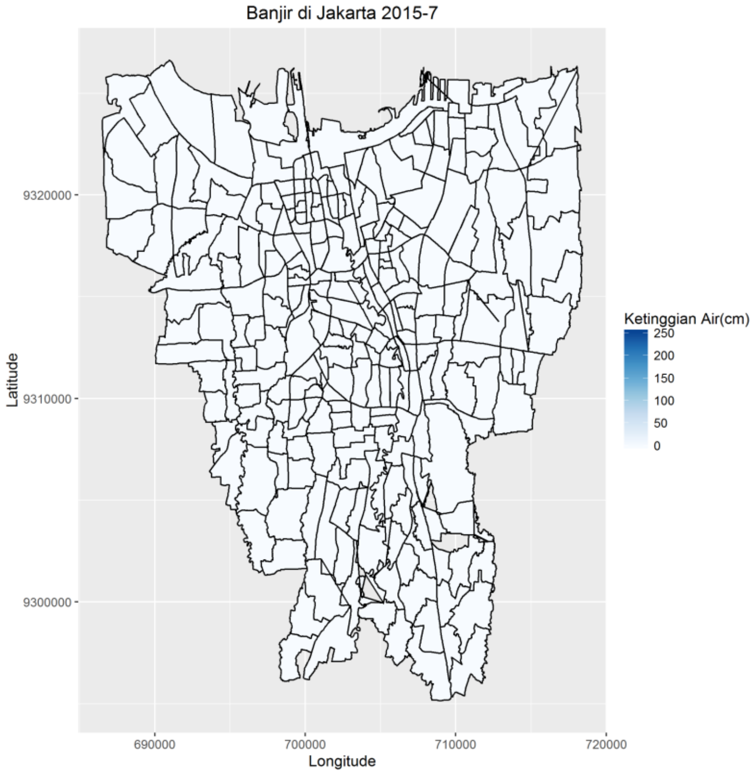

Banjir di Jakarta 2015-6
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
720000
Ketinggian
250
200
150
100
50
0

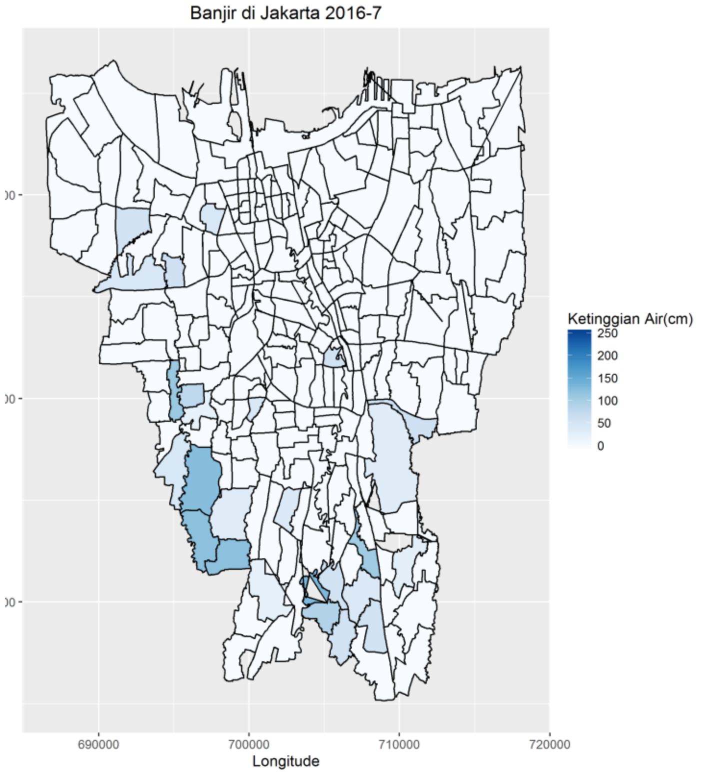

Banjir di Jakarta 2016-6
Latitude
Longitude
9320000
9310000
9300000
690000
700000
710000
720000
Ketinggian Air(cm)
250
200
150
100
50
0


Banjir di Jakarta 2013-7
9320000
9310000
9300000
Latitude
690000
700000
710000
Longitude

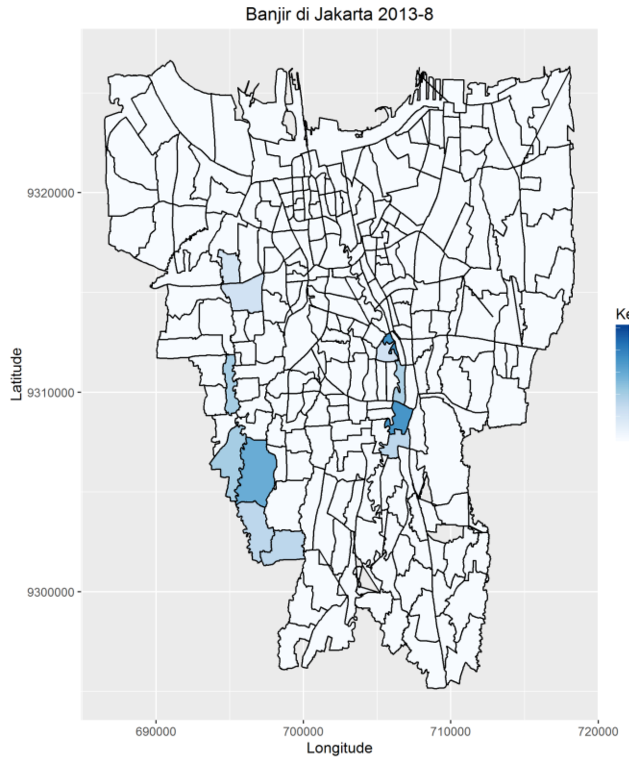

Banjir di Jakarta 2014-7
9320000
9310000
9300000
Latitude
690000
700000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

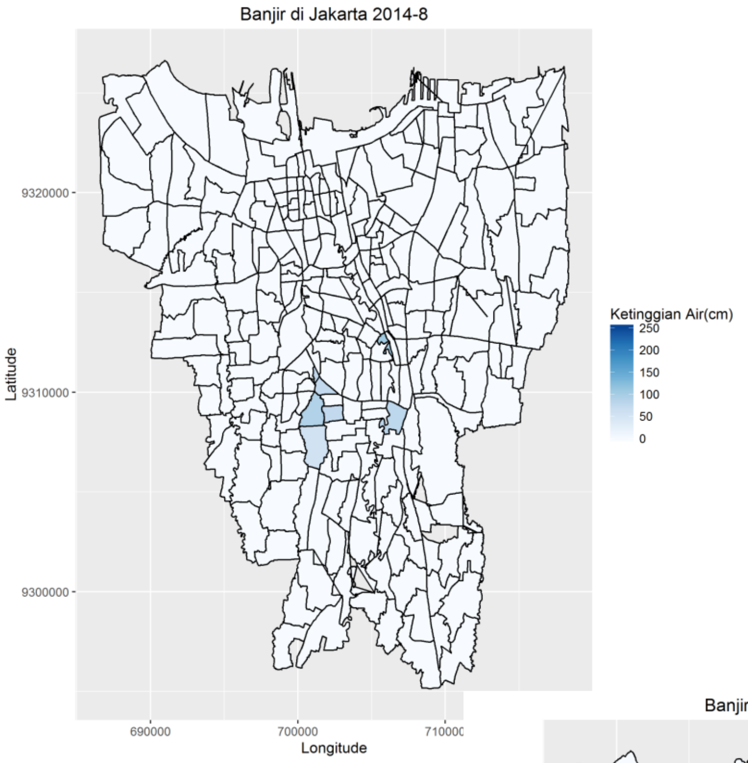
July

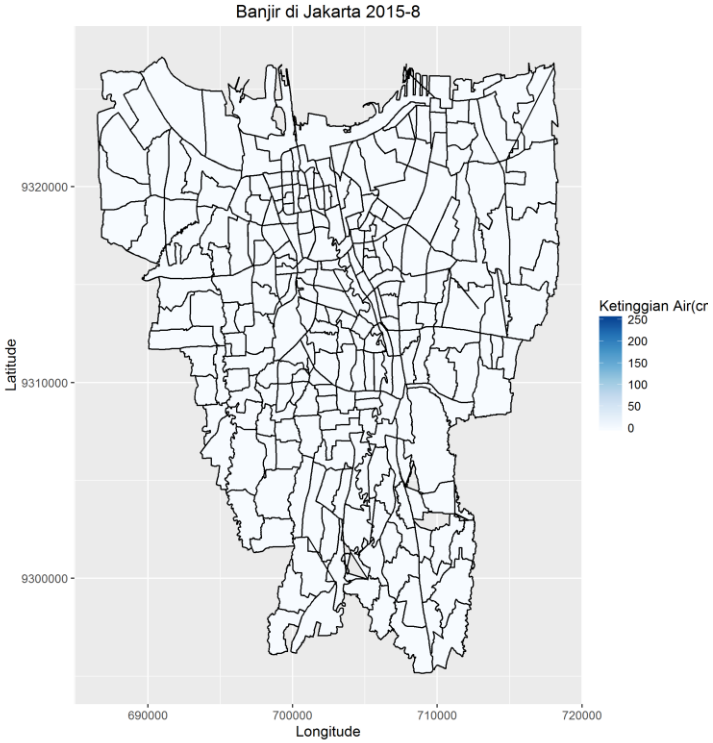

Banjir di Jakarta 2015-7
9320000
9310000
9300000
Latitude
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

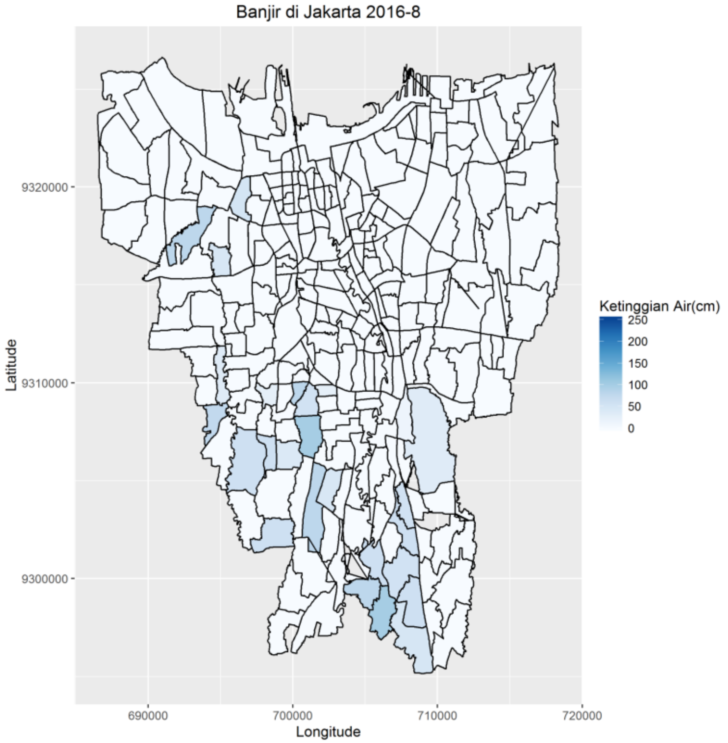

Banjir di Jakarta 2016-7
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

August

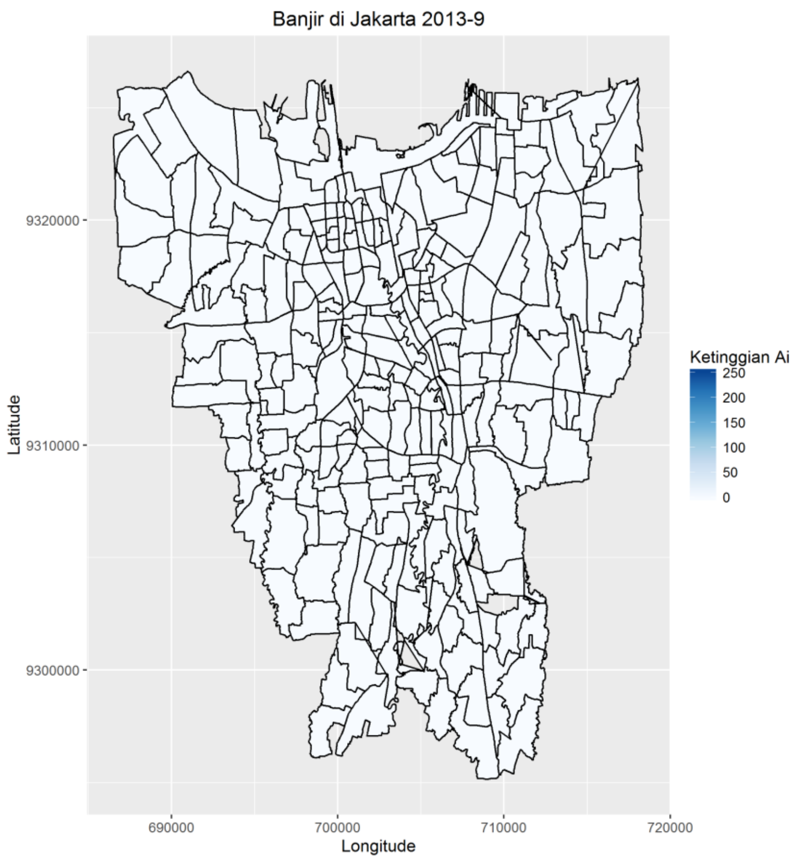

Banjir di Jakarta 2013-8
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000
Ke

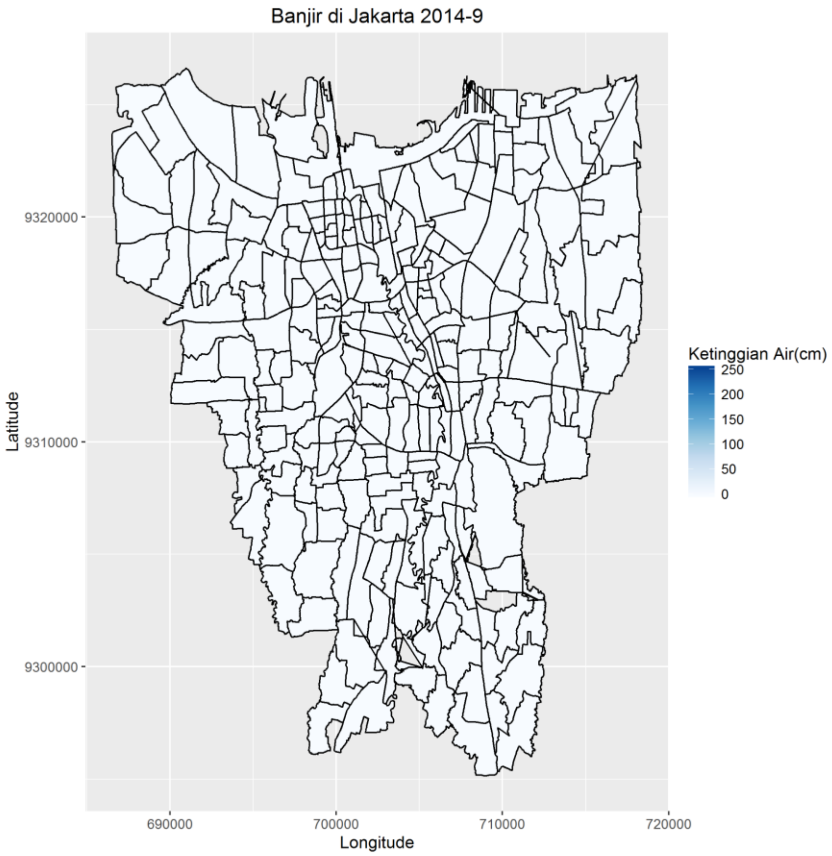

Banjir di Jakarta 2014-8
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
Ketinggian Air(cm)
250
200
150
100
50
0

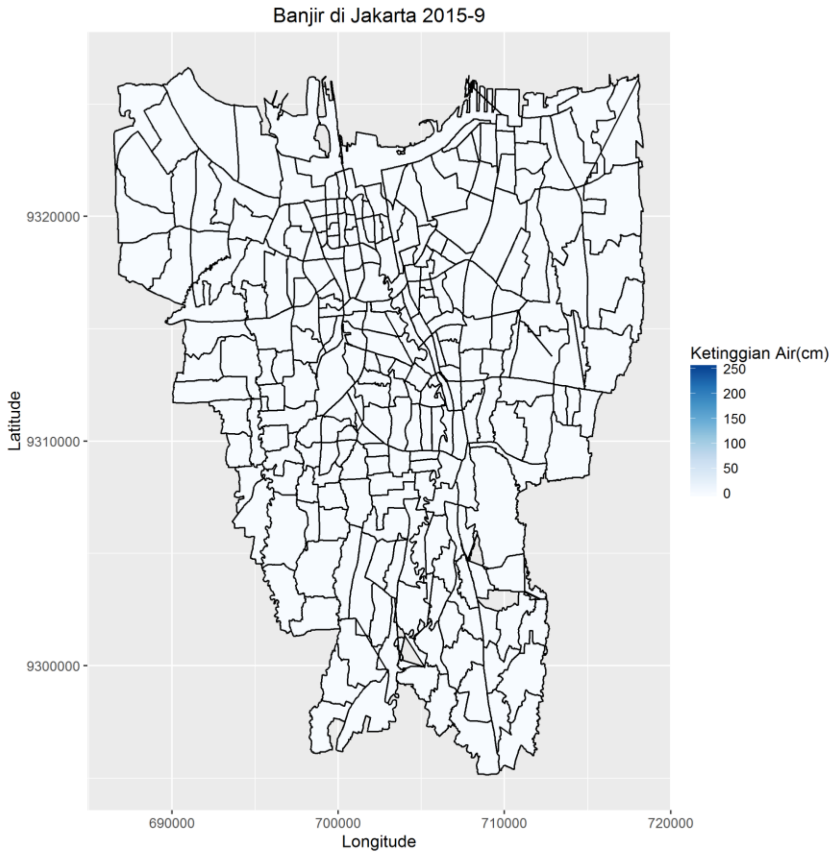

Banjir di Jakarta 2015-8
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000
Ketinggian Air(cr
250
200
150
100
50
0


Banjir di Jakarta 2016-8
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000
Ketinggian Air(cm)
250
200
150
100
50
0

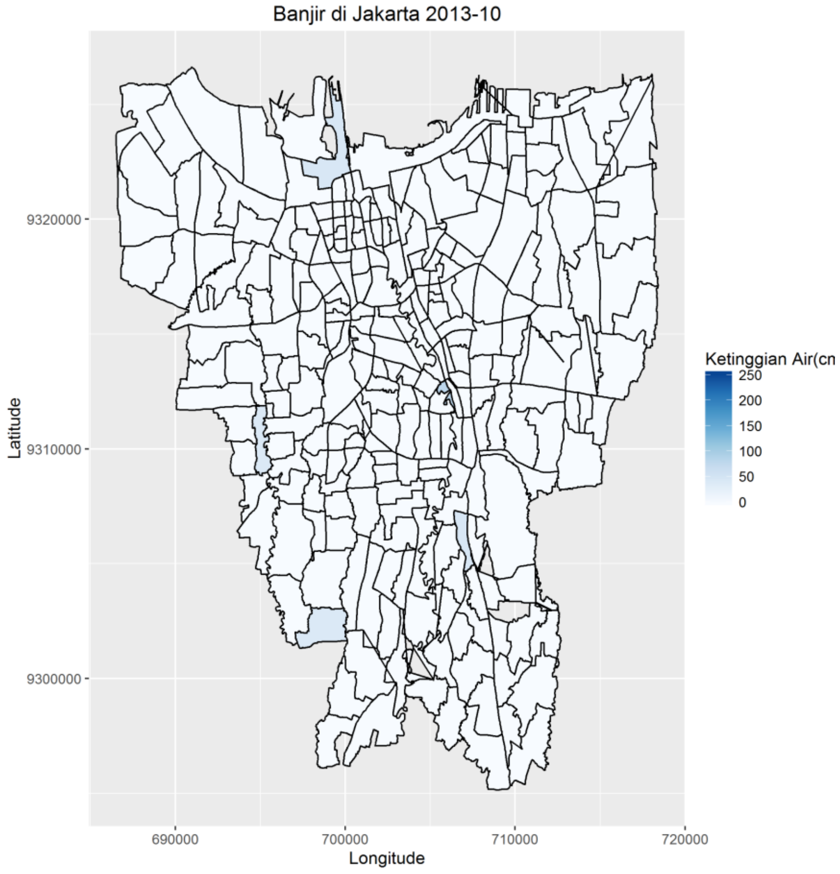
September

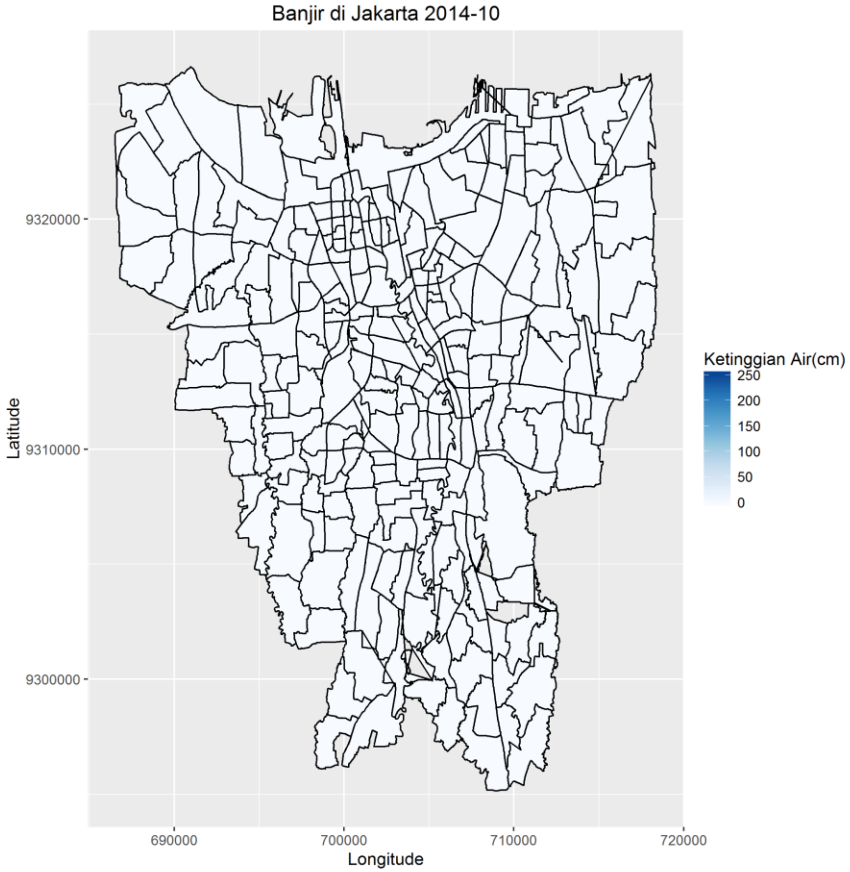

Banjir di Jakarta 2013-9
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude
Ketinggian Ai
250
200
150
100
50
0

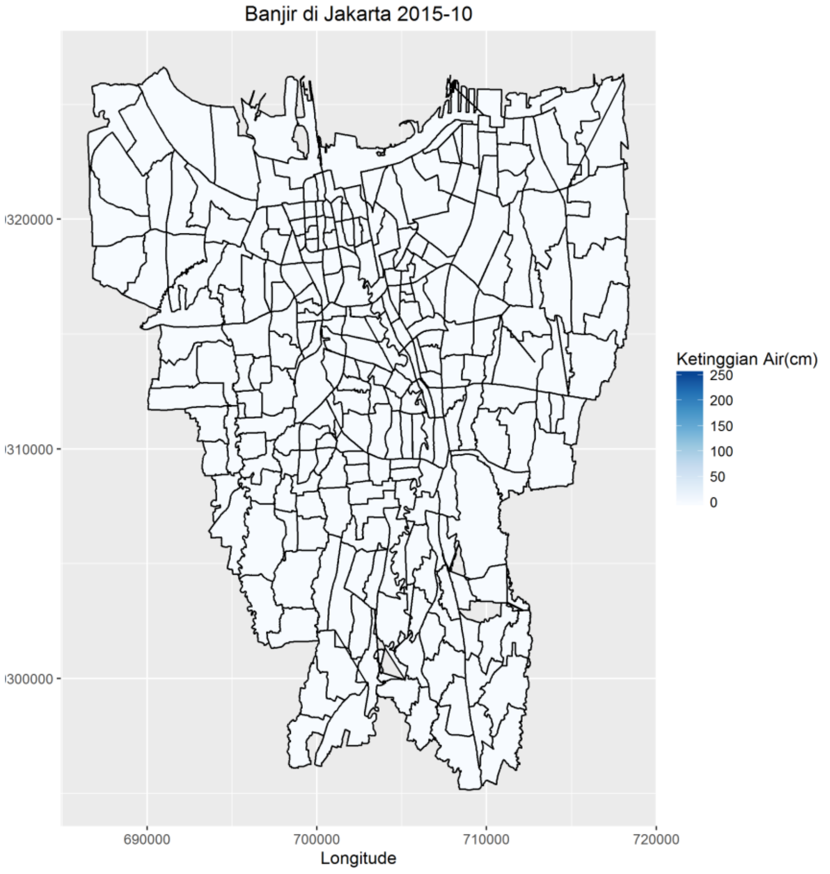

Banjir di Jakarta 2014-9
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0


Banjir di Jakarta 2015-9
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

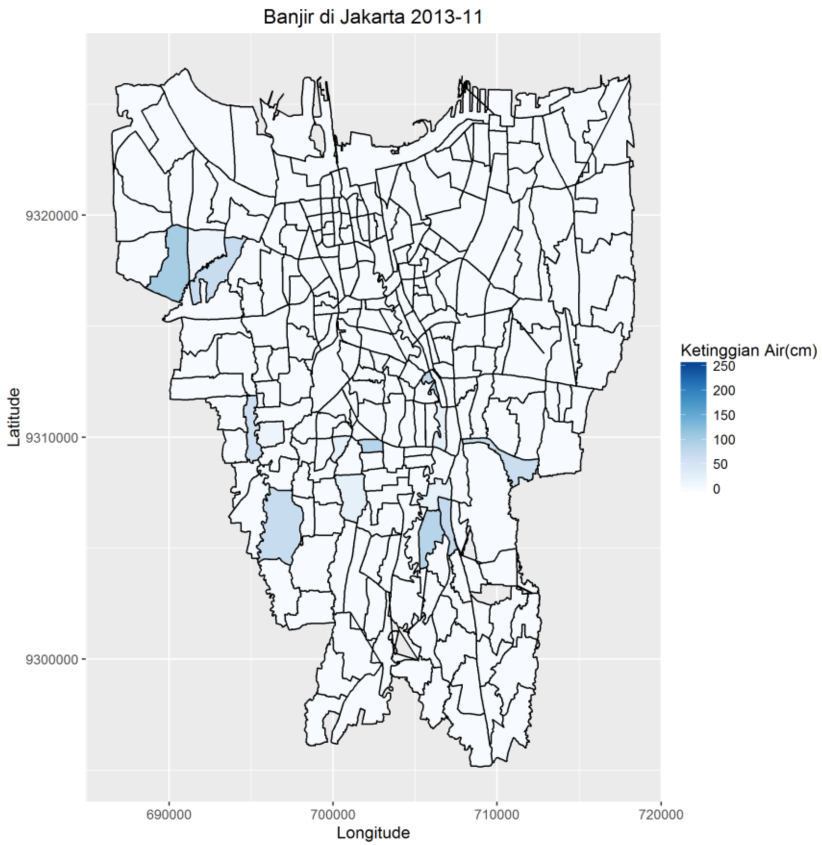
October

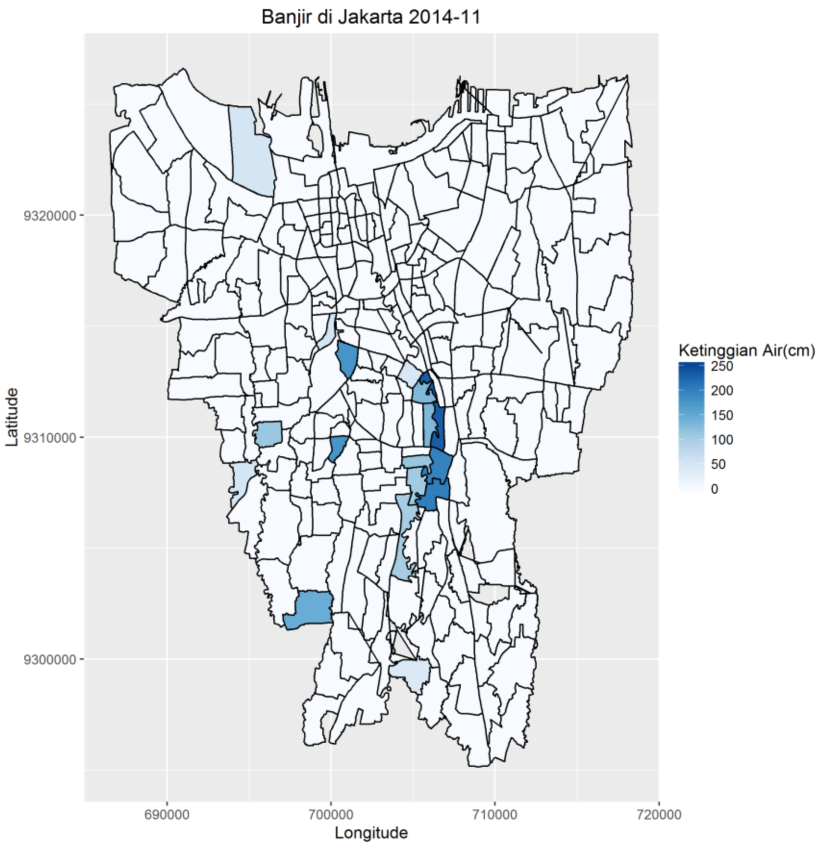

Banjir di Jakarta 2013-10
9320000
9310000
9300000
Latitude
690000
700000
710000
720000
Longitude
Ketinggian Air(cn
250
200
150
100
50
0

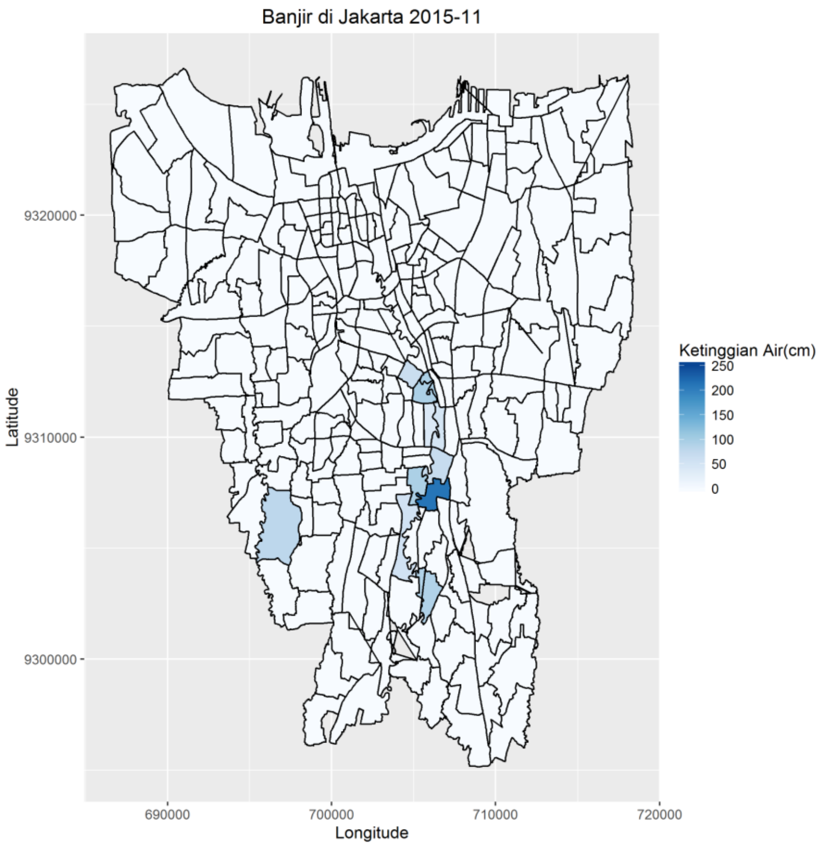

Banjir di Jakarta 2014-10
9320000
9310000
9300000
Latitude
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0


Banjir di Jakarta 2015-10
320000
310000
300000
690000
700000
710000
720000
Longitude
Ketinggian Air(cm)
250
200
150
100
50
0

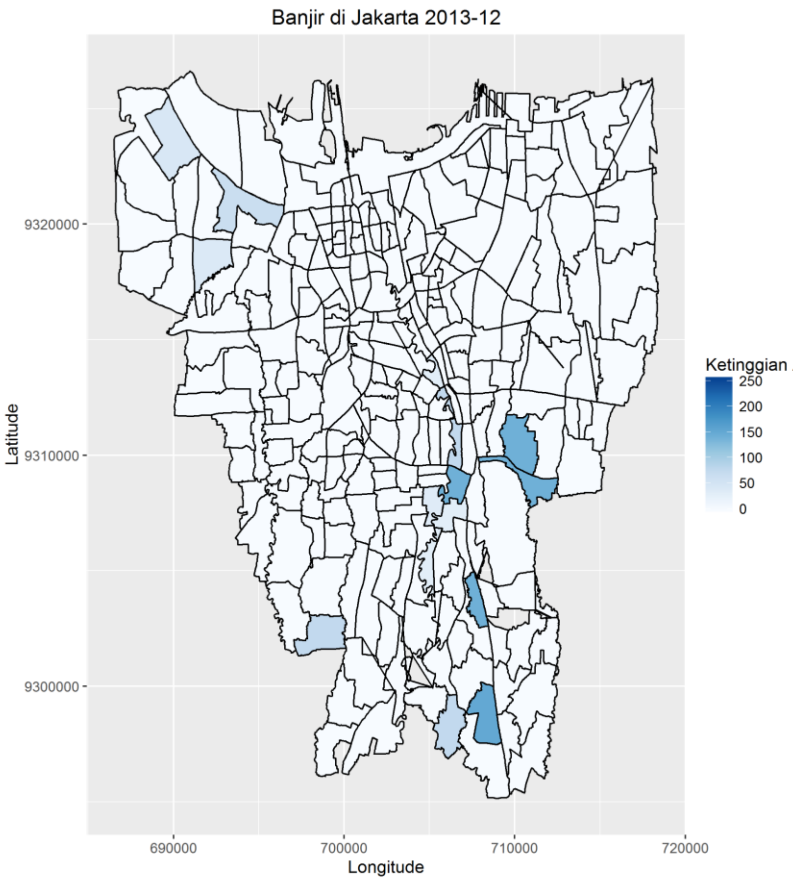
November


Banjir di Jakarta 2013-11
Ketinggian Air(cm)
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000

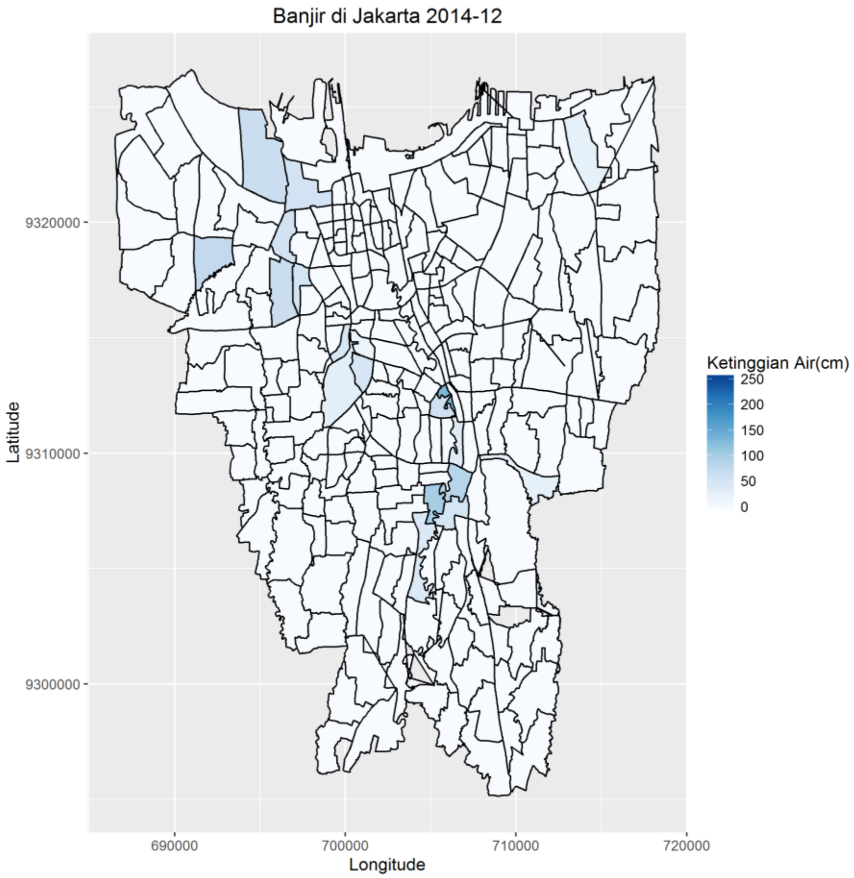

Banjir di Jakarta 2014-11
Ketinggian Air(cm)
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000


Banjir di Jakarta 2015-11
Ketinggian Air(cm)
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
Longitude
690000
700000
710000
720000

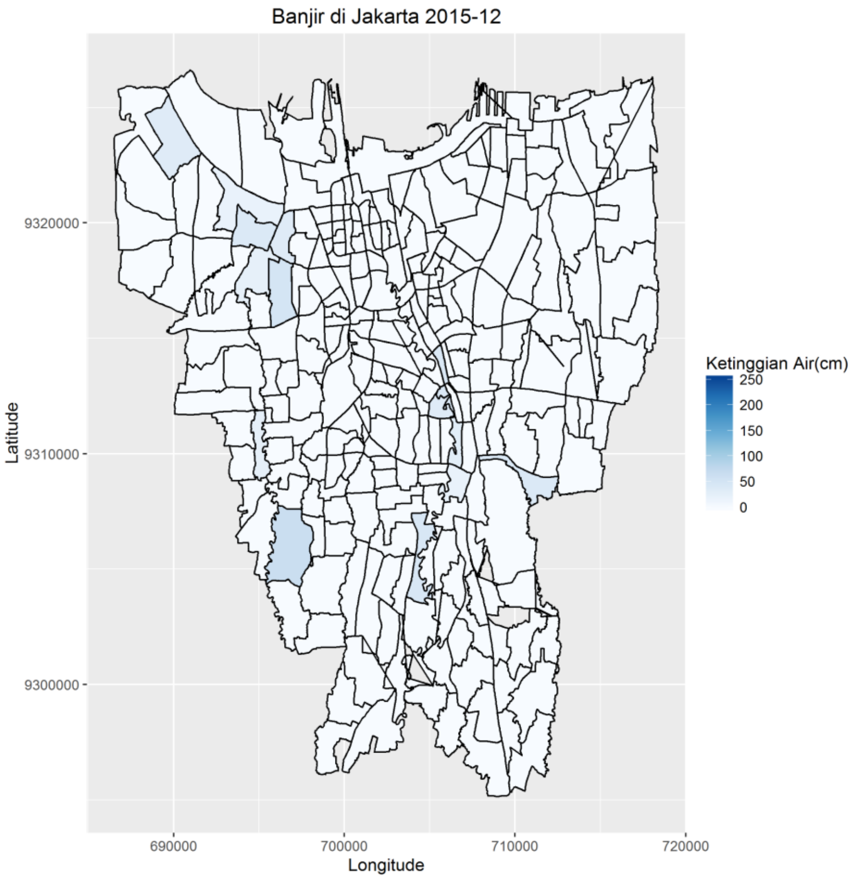
December



Banjir di Jakarta 2013-12
Ketinggian
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude



Banjir di Jakarta 2014-12
Ketinggian Air(cm)
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude

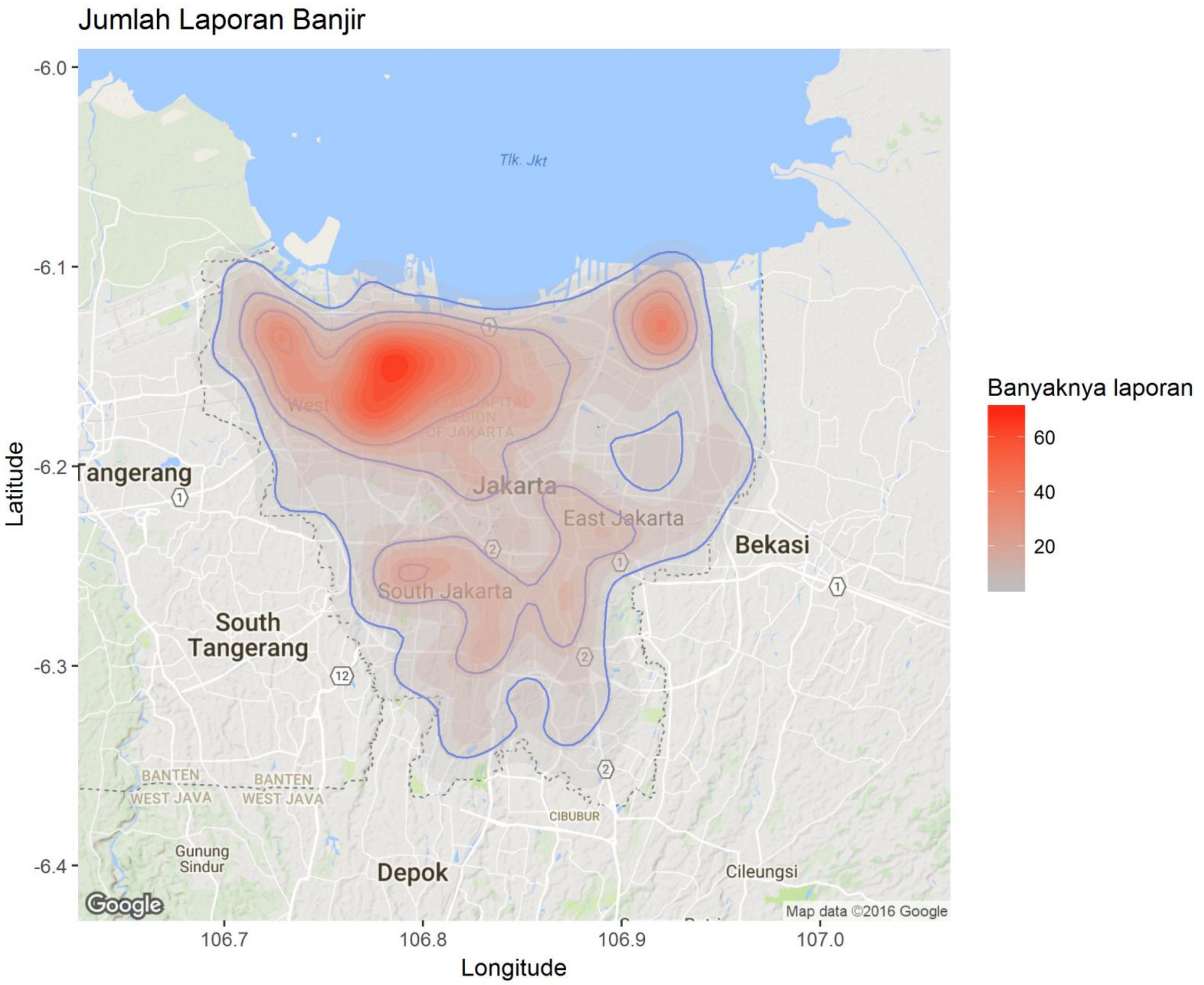


Banjir di Jakarta 2015-12
Ketinggian Air(cm)
250
200
150
100
50
0
Latitude
9320000
9310000
9300000
690000
700000
710000
720000
Longitude

Exploratory Data Analysis

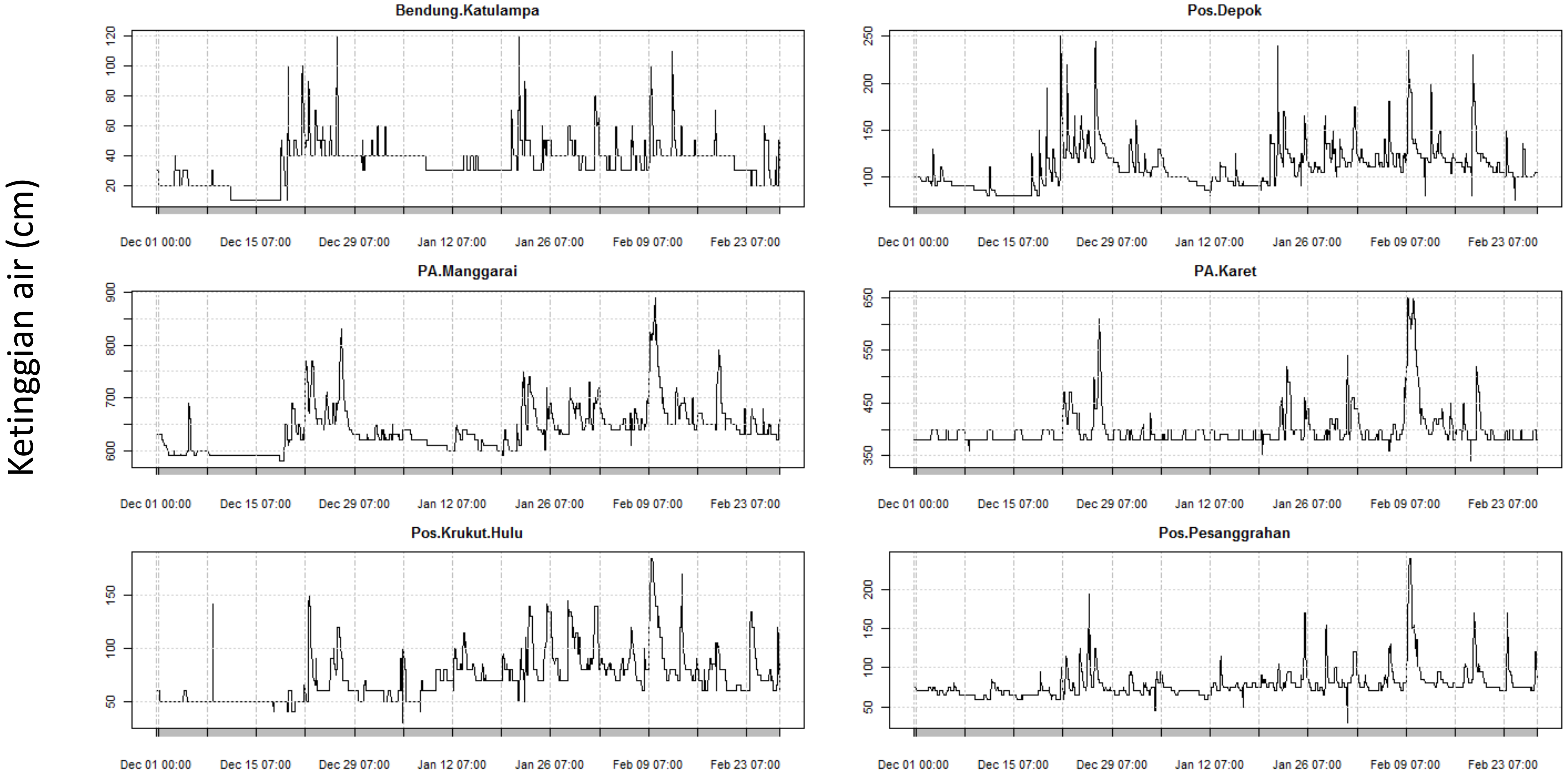
2. Data Laporan Banjir

We also collected flood report data from social media (Qlue, Twitter, and Detik) to monitor flood area in 2016 as shown in the heat map ‘Jumlah Laporan Banjir’. North and West Jakarta are the area with higher number of flood reports.



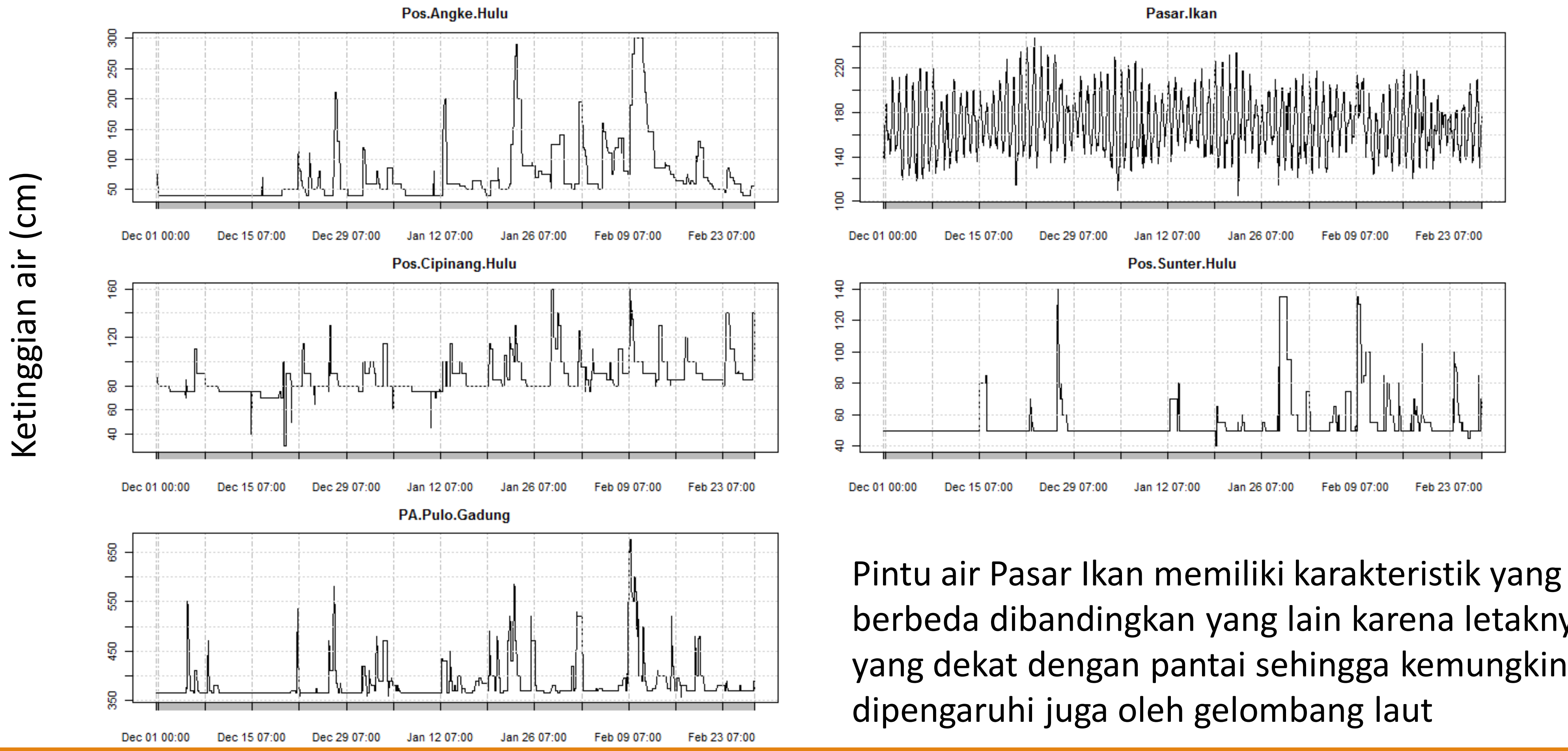
Exploratory Data Analysis

3. Water Level Data



Exploratory Data Analysis

3. Water Level Data

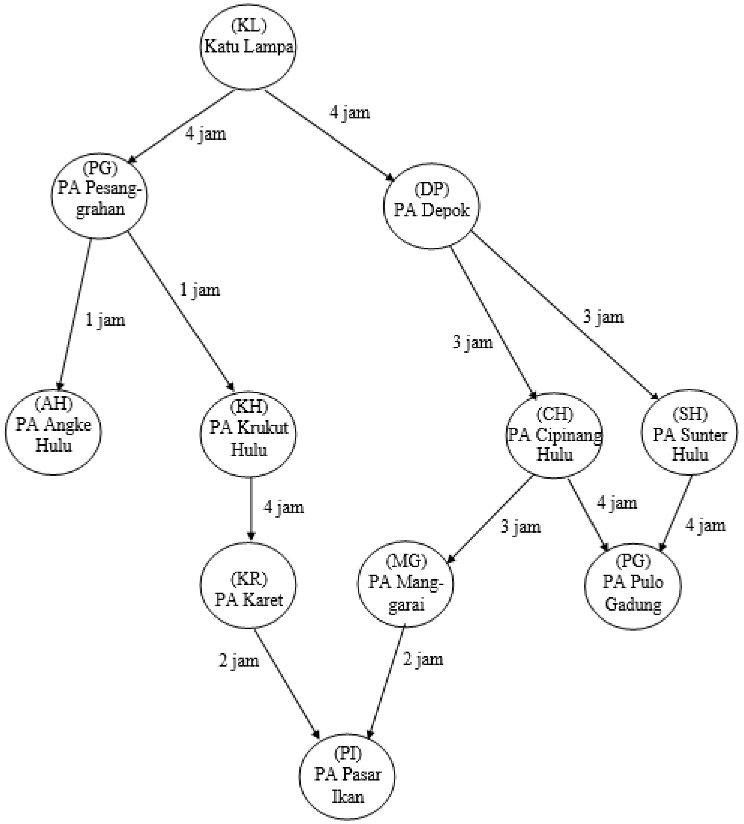


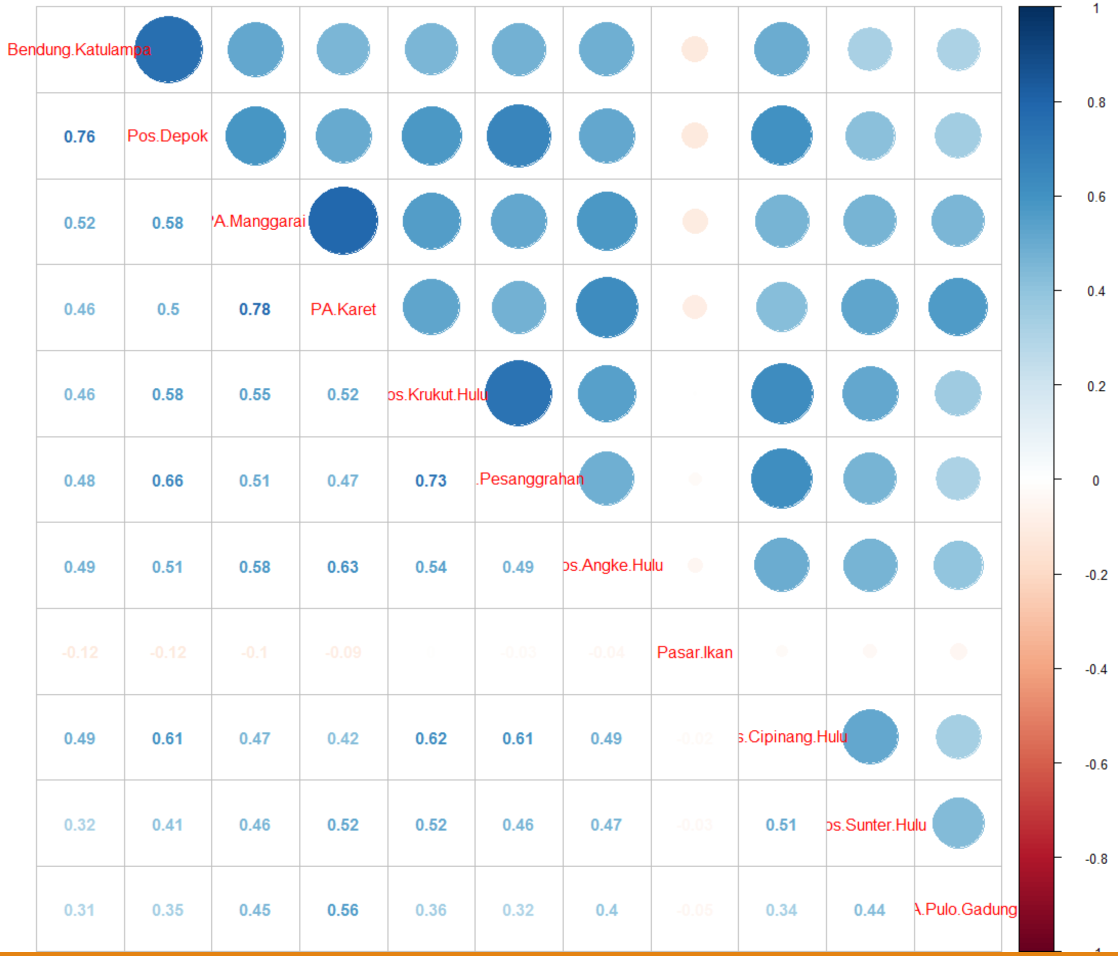
Pintu air Pasar Ikan memiliki karakteristik yang berbeda dibandingkan yang lain karena letaknya yang dekat dengan pantai sehingga kemungkinan dipengaruhi juga oleh gelombang laut

Exploratory Data Analysis

3. Water Level Data

Hubungan antara pintu air



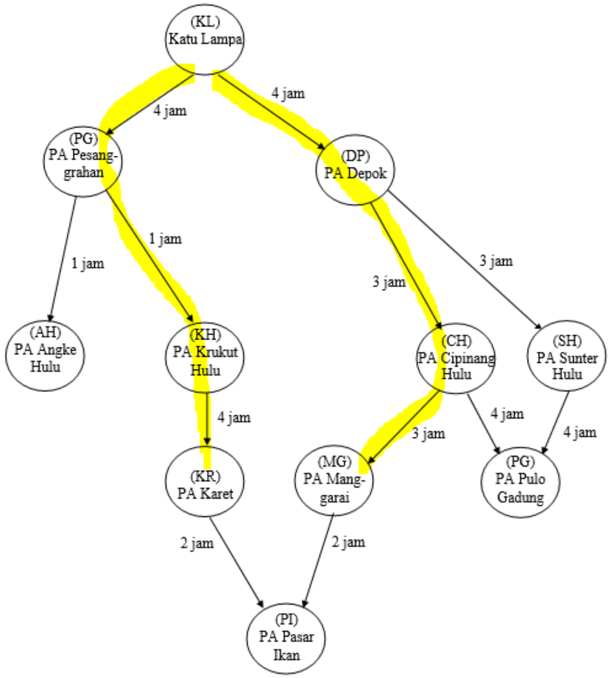


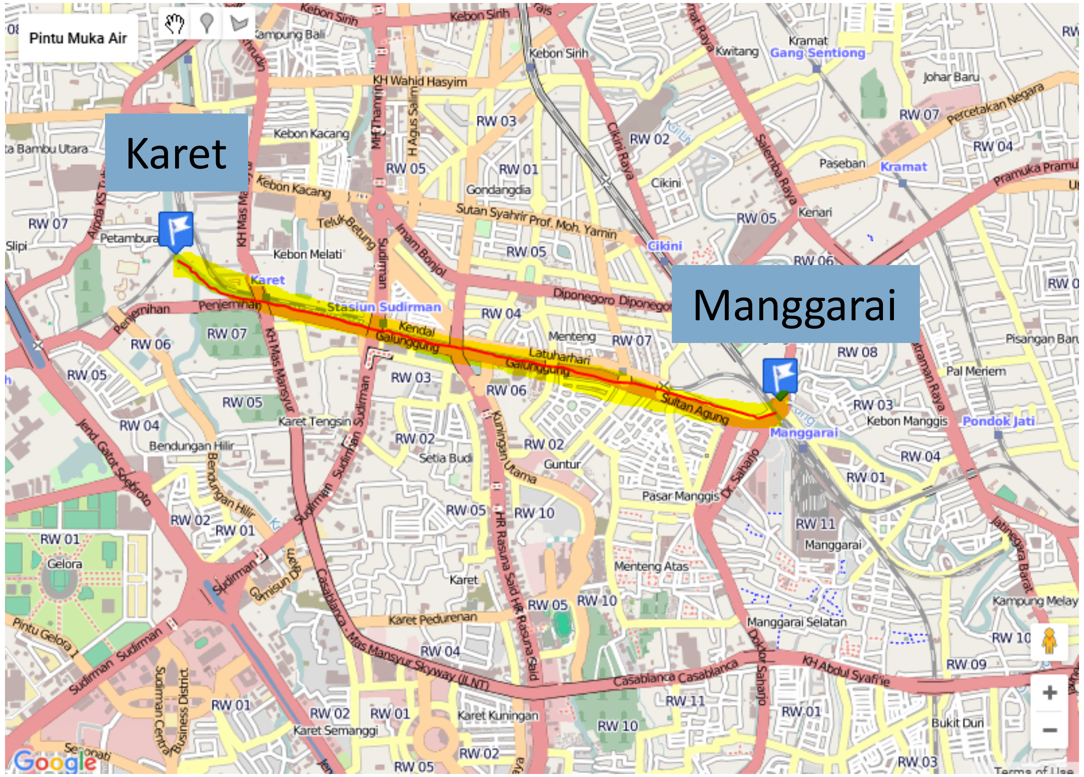
Cross Correlation between river gauge time series

Exploratory Data Analysis

3. Water Level Data

- Korelasi antara pintu air Manggarai dan Karet merupakan yang paling tinggi yaitu sebesar 0.78. Hal ini sangat rasional mengingat jarak antara pintu air Karet dan Manggarai hanya ± 5 km dan berasal dari satu aliran air yang sama yaitu sungai Ciliwung.



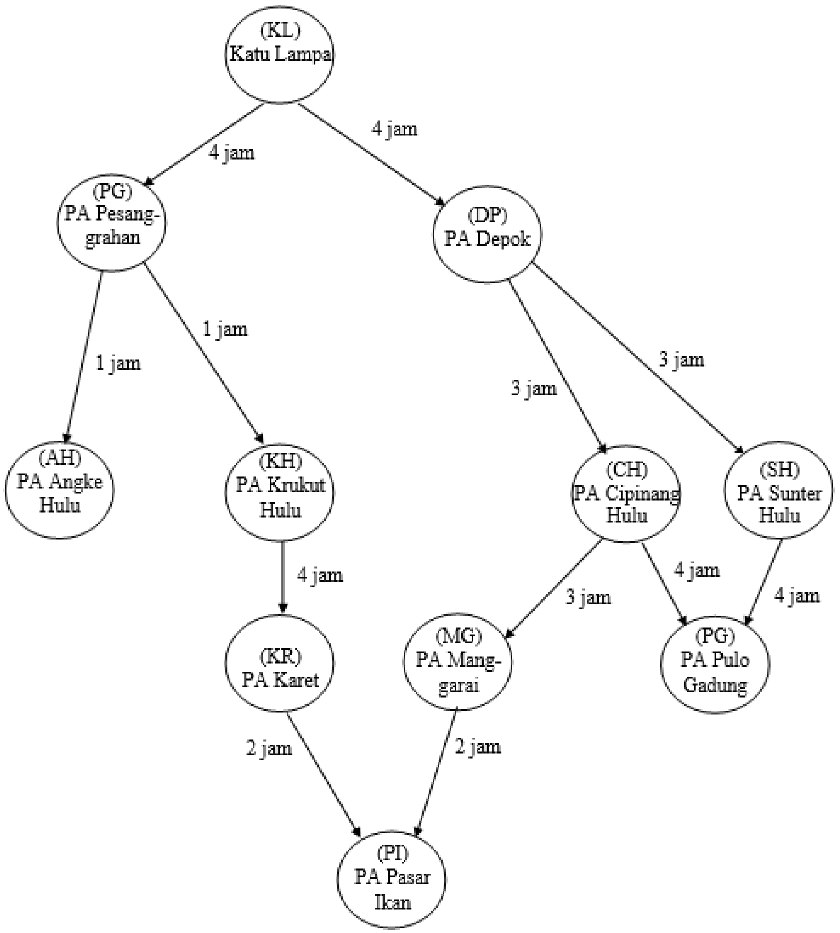


Water Level Model Jakarta

To predict flood in Jakarta, we are using indirect approach. Instead of directly creating flood model, we created model to predict water level in each river gauge across Jakarta (11 river gauges)

We can rely on this model because based on experiences. If we know the water level, we can determine flood area.

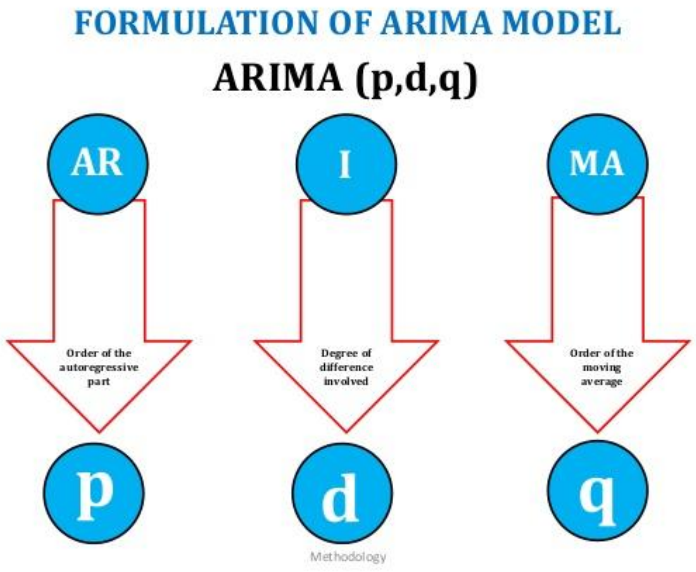
Data: Times series data recorded hourly from January 2014 – September 2016 with relationship shown in the diagram.

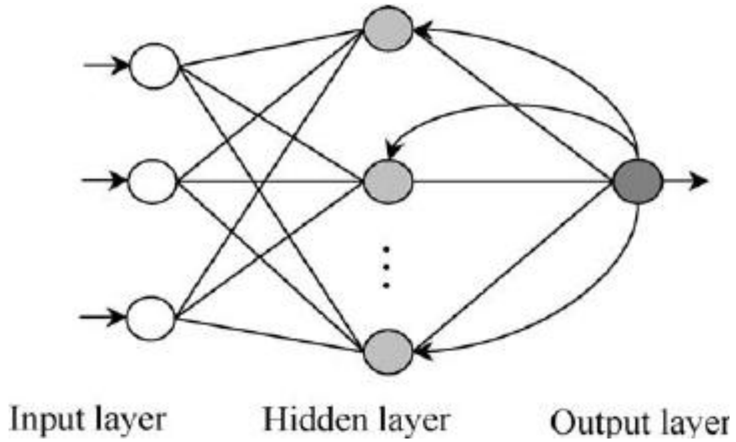


Result and Analysis

Water Level Model

- Time series data dari 11 pintu air
  - ARIMA
  - LSTM - Recurent Neural Network





https://www.slideshare.net/SaifurRahmanShohel/modeling-volatility-and-forecasting-of-stock-pricea-case-study-on-two-private-commercial-banks-in-bangladesh-58090202

http://jonamjar.com/2016/12/22/recurrent-neural-networks/

ARIMA

- Stationarity

- To test whether the model is stationary or not, I used two tests : Augmented Dickey–Fuller (adf) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test

	P-Value		Stationary	
Pintu Air	ADF	KPSS	ADF	KPSS
Bendung Katulampa	<0.01	0.1	V	V
Pos Depok	<0.01	0.01	V	X
PA Manggarai	<0.01	0.01	V	X
PA Karet	<0.01	0.01	V	X
Pos Krukut Hulu	<0.01	0.01	V	X
Pos Pesanggrahan	<0.01	0.1	V	V
Pos Angke Hulu	<0.01	0.1	V	V
Pasar Ikan	<0.01	0.024	V	X
Pos Cipinang Hulu	<0.01	0.1	V	V
Pos Sunter Hulu	<0.01	0.01	V	X
PA Pulo Gadung	<0.01	0.01	V	X

Jika mengacu pada ADF test maka dapat disimpulkan semua data pintu air stationer. Akan tetapi, kelemahan dari test ini adalah asumsi bahwa errornya homogen dan independen. Sehingga untuk memperkuat hipotesis dilakukan tes dengan metode KPSS. Didapatkan bahwa beberapa pintu air seperti, Depok, Manggarai, Karet, Krukut Hulu, Pasar Ikan, Sunter Hulu dan Pulo Gadung sudah stasioner.

ARIMA

- Analyze ACF and PACF Plot

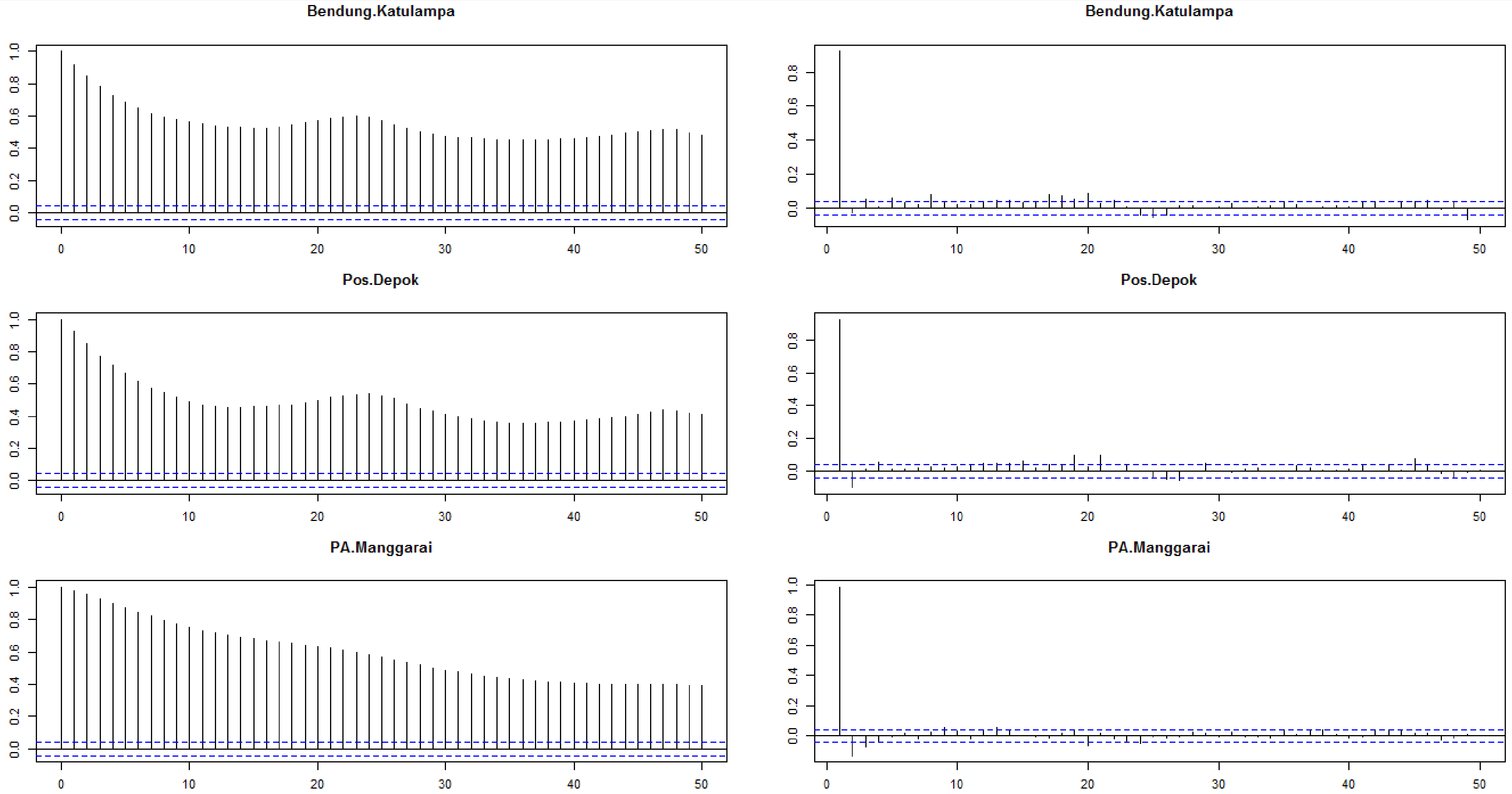
- Penentuan parameter manual dibandingkan dengan Auto ARIMA (Hyndman, 2008)* yang menggunakan pemilihan model dengan parameter p, d, q yang meminimalkan AICc (meminimalkan loss of information)

	Parameter			Auto ARIMA		
Pintu Air	p	d	q	p	d	q
Bendung Katulampa	1	0	0	1	1	2
Pos Depok	2	0	0	2	1	1
PA Manggarai	3	0	0	1	1	1
PA Karet	3	0	0	3	1	3
Pos Krukut Hulu	4	0	0	2	1	2
Pos Pesanggrahan	4	0	0	2	1	2
Pos Angke Hulu	1	0	0	1	1	1
Pasar Ikan	1	0	1	3	0	1
Pos Cipinang Hulu	1	0	0	1	1	1
Pos Sunter Hulu	2	0	0	0	1	2
PA Pulo Gadung	2	0	0	1	1	1

*Hyndman, R. J. and Y. Khandakar. 2008. Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software. Vol 27.

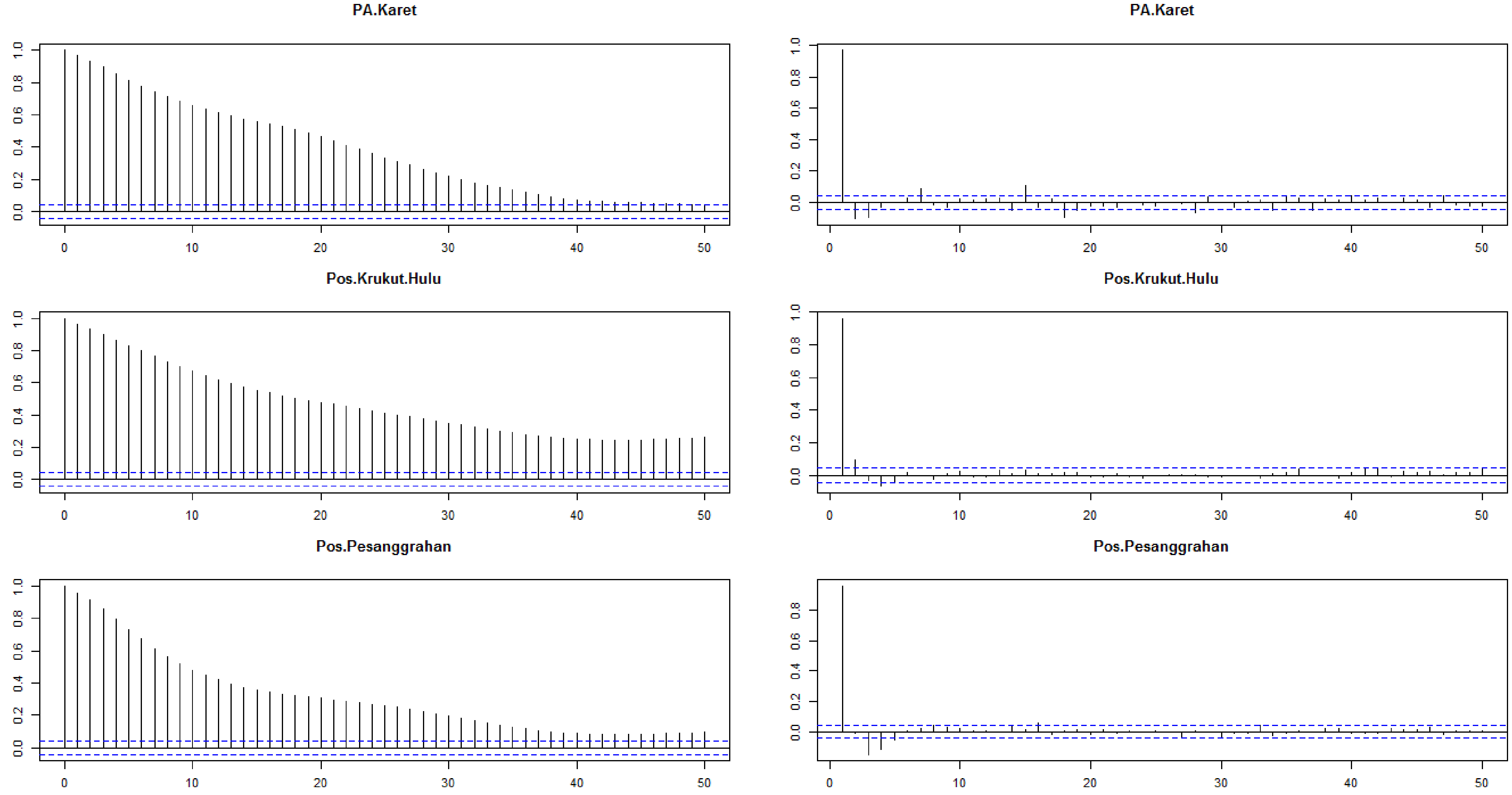
ARIMA

- Analyze ACF and PACF Plot



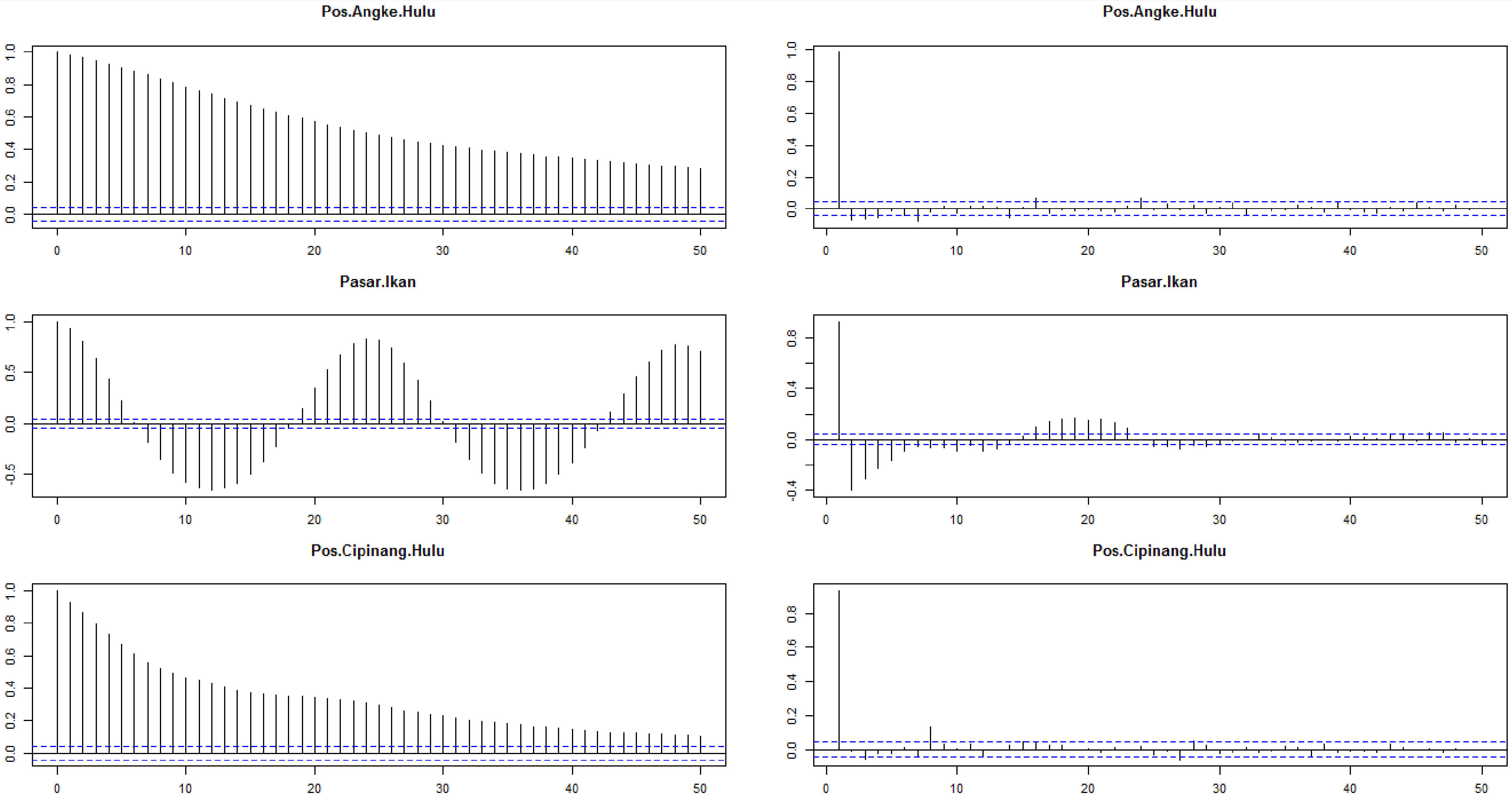
ARIMA

- Analyze ACF and PACF Plot



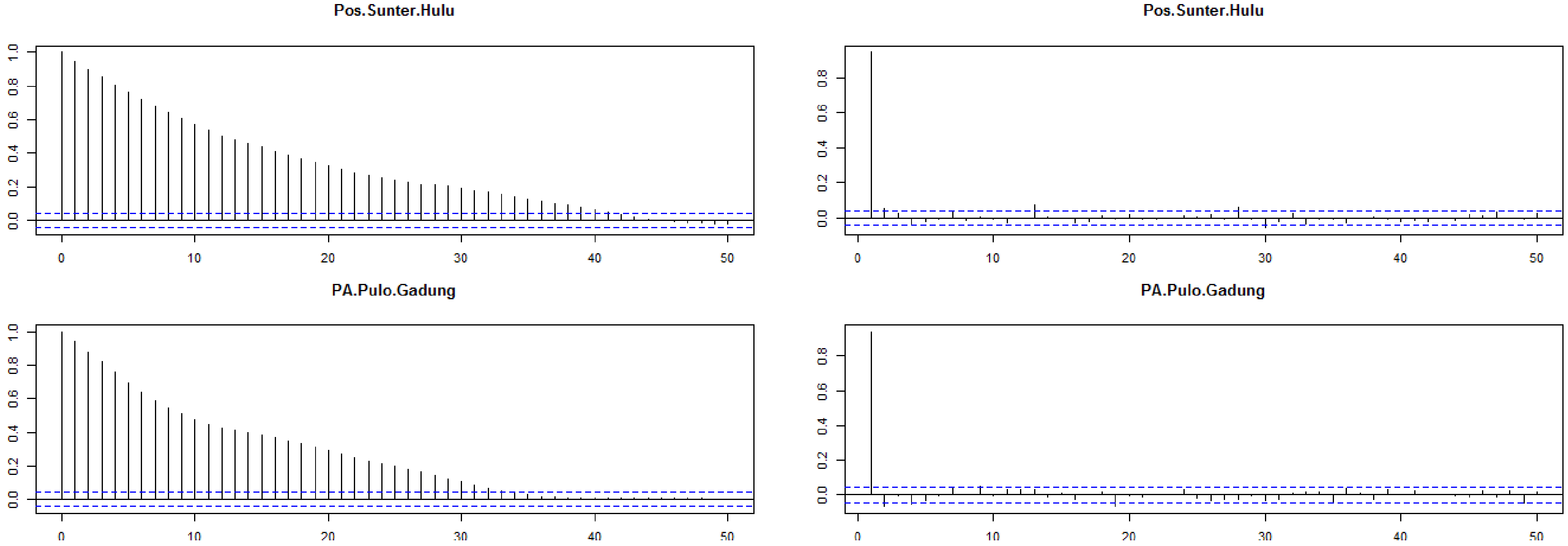
ARIMA

- Analyze ACF and PACF Plot



ARIMA

- Analyze ACF and PACF Plot



ARIMA

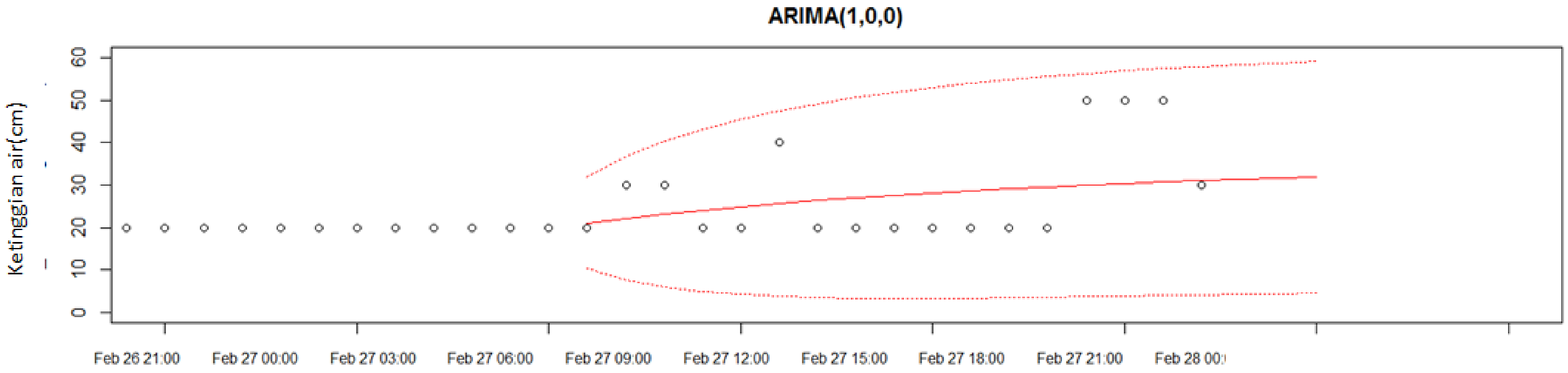
- Predict for the next 17 hours

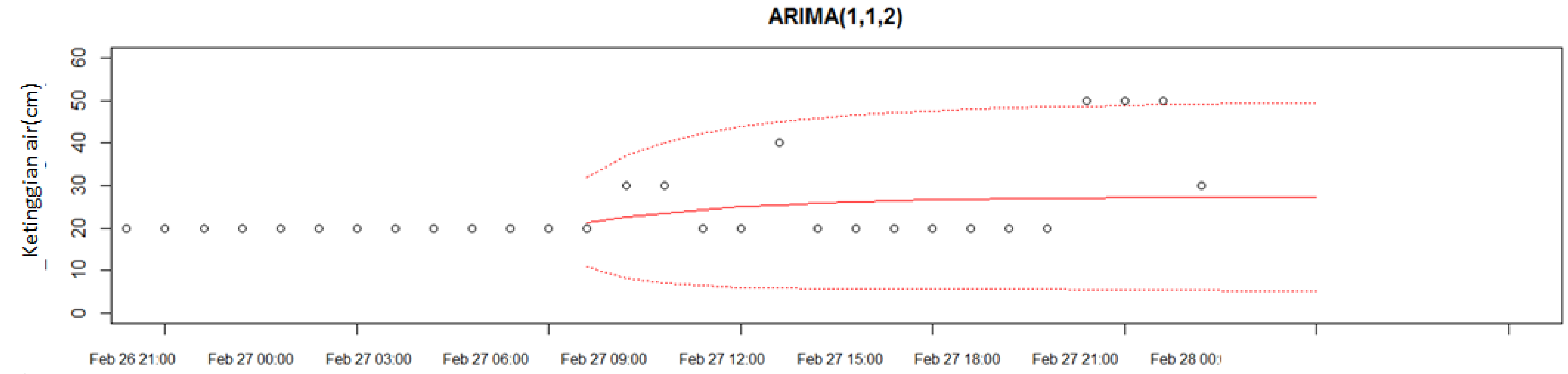
	RMSE (cm)	
Pintu Air	Manual Model	AutoArima
Bendung Katulampa	10.79	11.45
Pos Depok	2.71	1.75
PA Manggarai	12.41	12.37
PA Karet	13.42	16.94
Pos Krukut Hulu	26.80	29.64
Pos Pesanggrahan	22.45	25.86
Pos Angke Hulu	9.43	13.77
Pasar Ikan	29.02	13.67
Pos Cipinang Hulu	34.79	35.49
Pos Sunter Hulu	11.11	12.19
PA Pulo Gadung	7.38	8.40

Dengan menggunakan 2 model yang sudah dibuat untuk masing-masing pintu air. Model pun diuji untuk memprediksi nilai ketinggian air pada 17 jam kedepan. Manual model lebih baik di mayoritas pintu air dengan nilai RMSE(Root Mean Square Error) lebih kecil dibandingkan hasil Auto-Arima, bahkan di pintu air yang menurut tes KPPS tidak stationer.

ARIMA

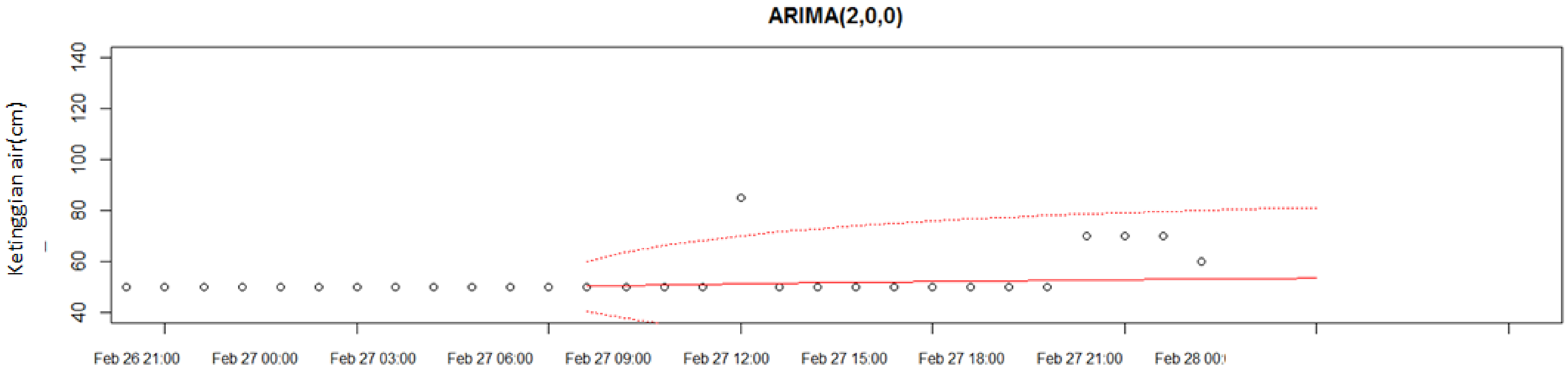
- Bendungan Katulampa (Desember 2014 – Februari 2015)

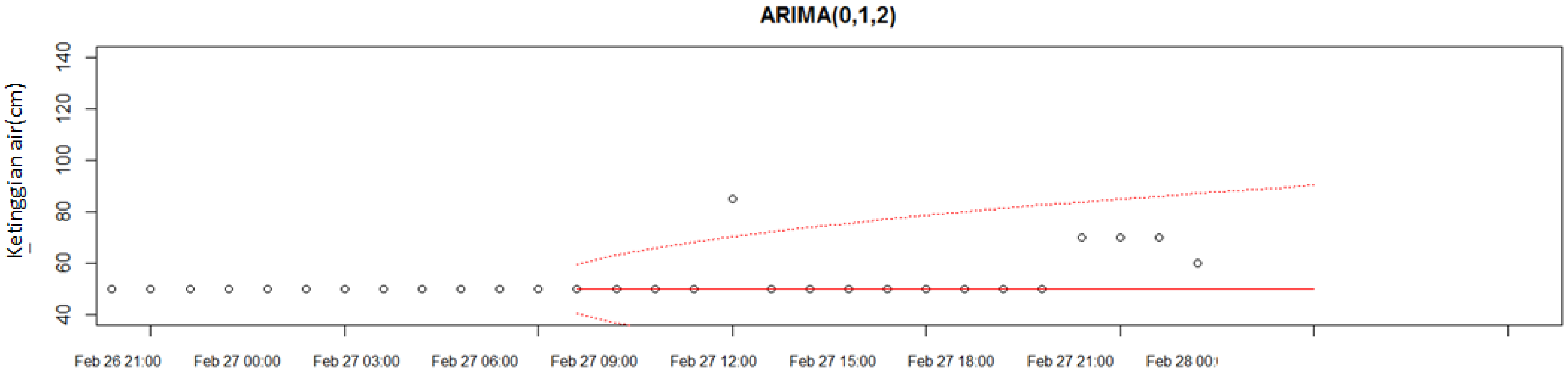




ARIMA

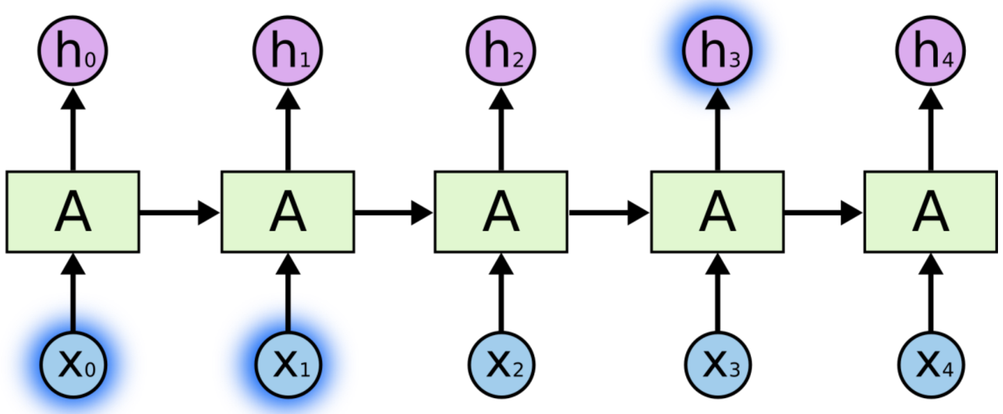
- Pos Sunter Hulu(Desember 2014 – Februari 2015)





LSTM Recurrent Neural Network

- Long Short-Term Memory
  - Capable of learning long-term dependencies
  - Can learn to use the past information



T = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]

Atau

Feature 1	Feature 2	Feature 3	Response
1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7
5	6	7	8
6	7	8	9
7	8	9	10

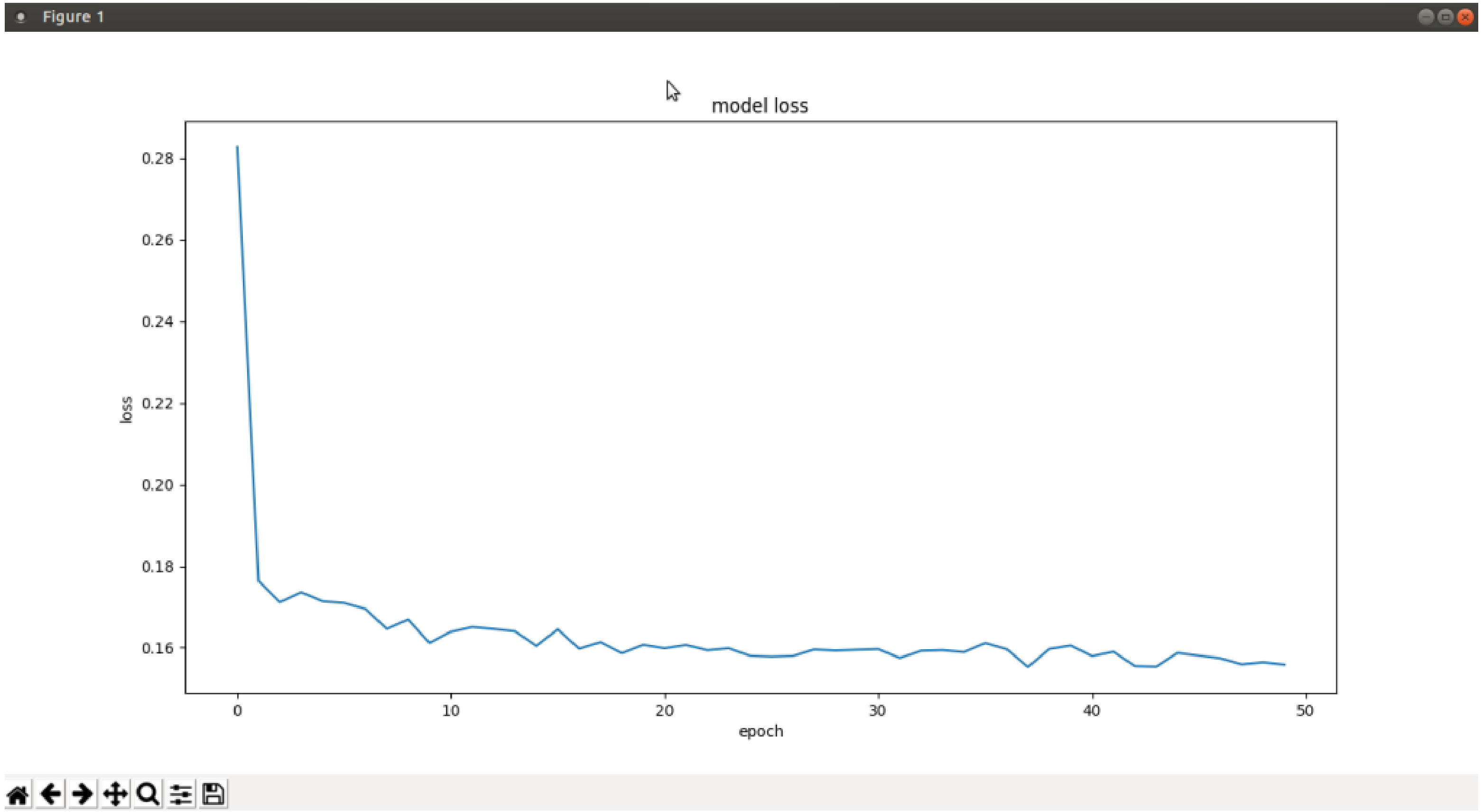
Feature 1 t-2	Feature 1 t-1	Feature 1 t	Response t+1
1	2	3	4
2	3	4	5
3	4	5	6
4	5	6	7
5	6	7	8
6	7	8	9
7	8	9	10

LSTM Recurrent Neural Network

- Parameter and layer network
  - Epoch = 10
  - Batch_size = 12, 100
  - Optimizer = Adam
  - Dropout layer
  - Layer = 4 layer : [timestep, 100, 50, 0] & 5 layer : [timestep, 30, 100, 50, 0]
  - Timestep

LSTM Recurrent Neural Network

- Epoch



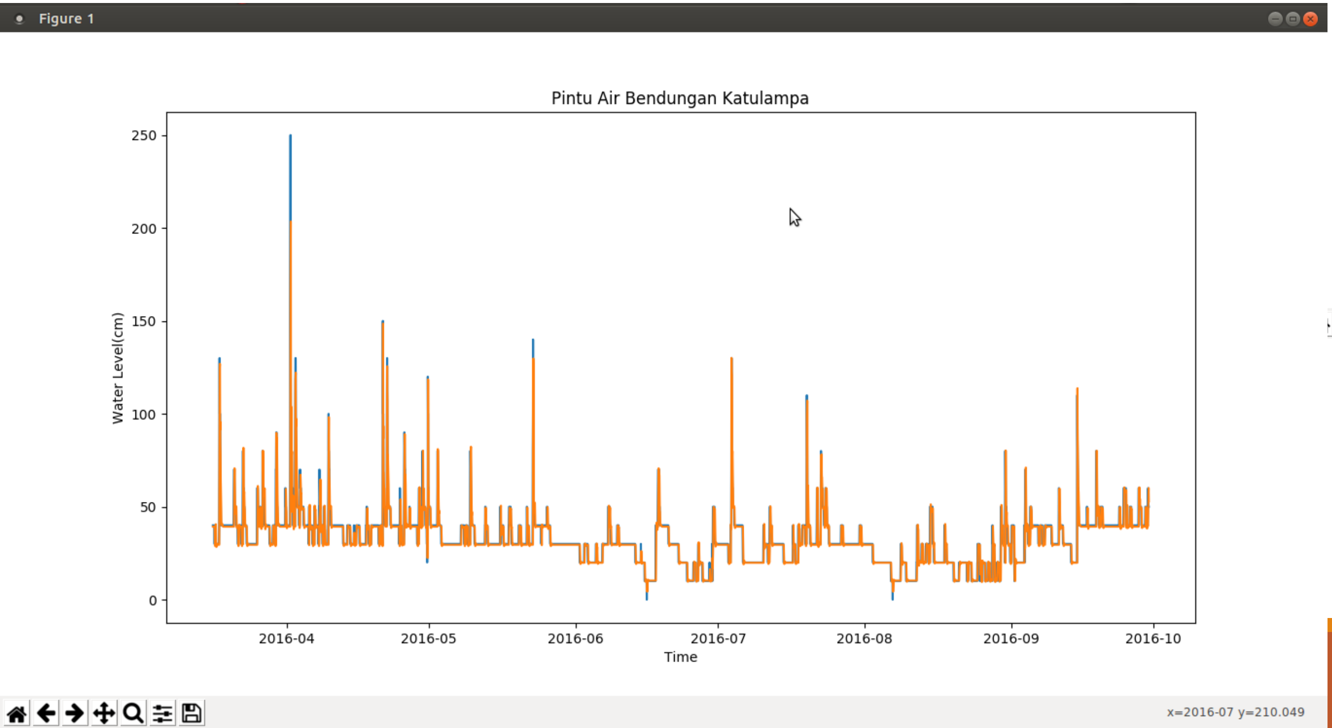
LSTM RNN vs ARIMA

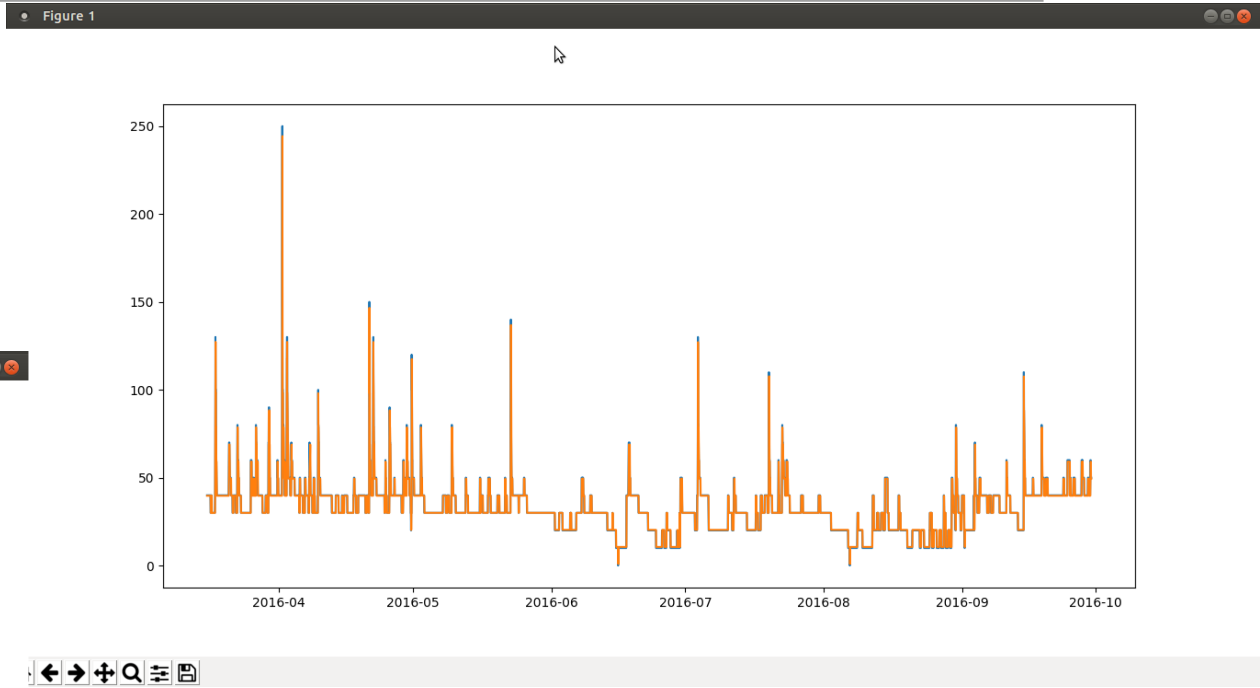
Notes:

Left LSTM – Right ARIMA

Orange → prediction line

Blue → Real data

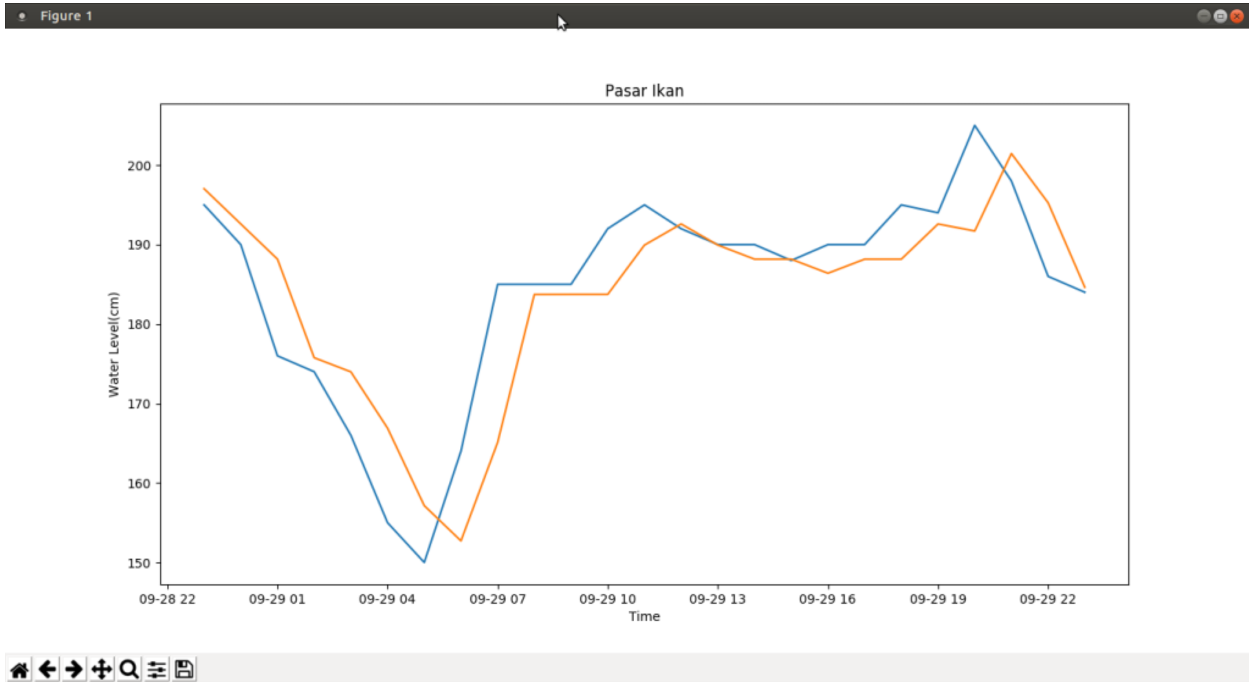


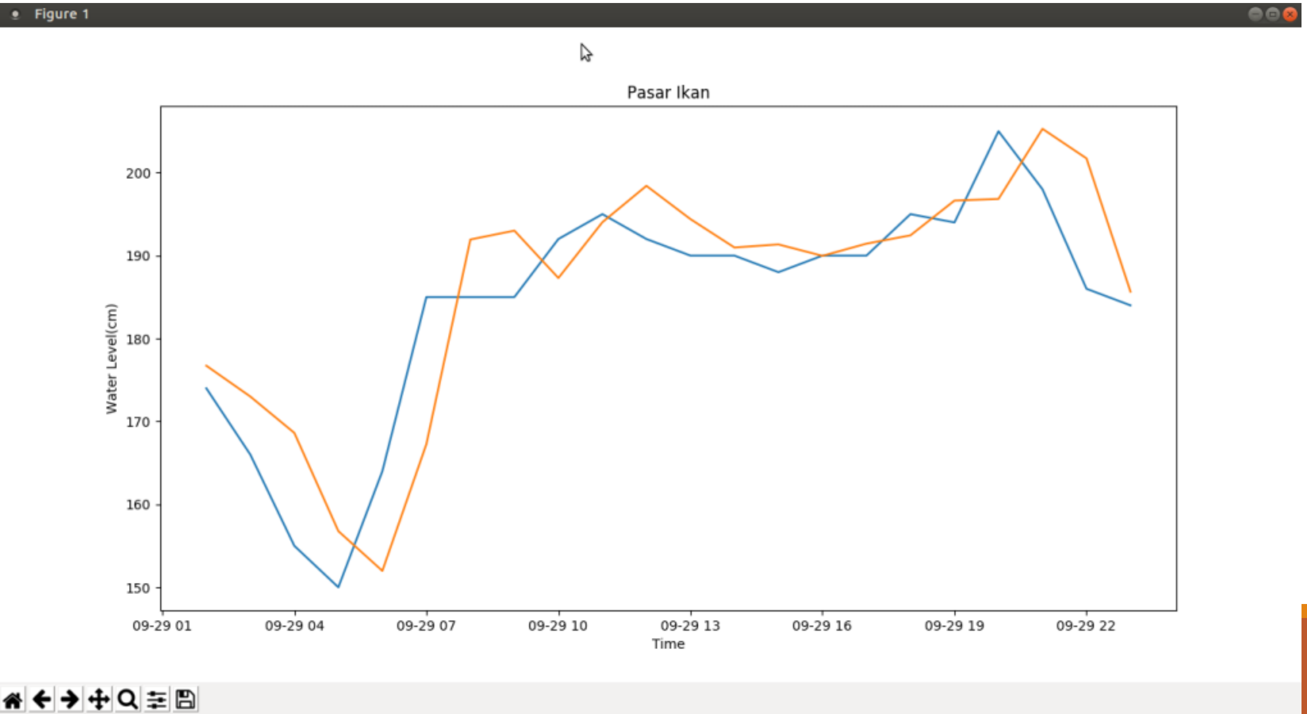


the prediction line is made up of singular prediction points that have had the whole prior true history window behind them

LSTM RNN vs ARIMA

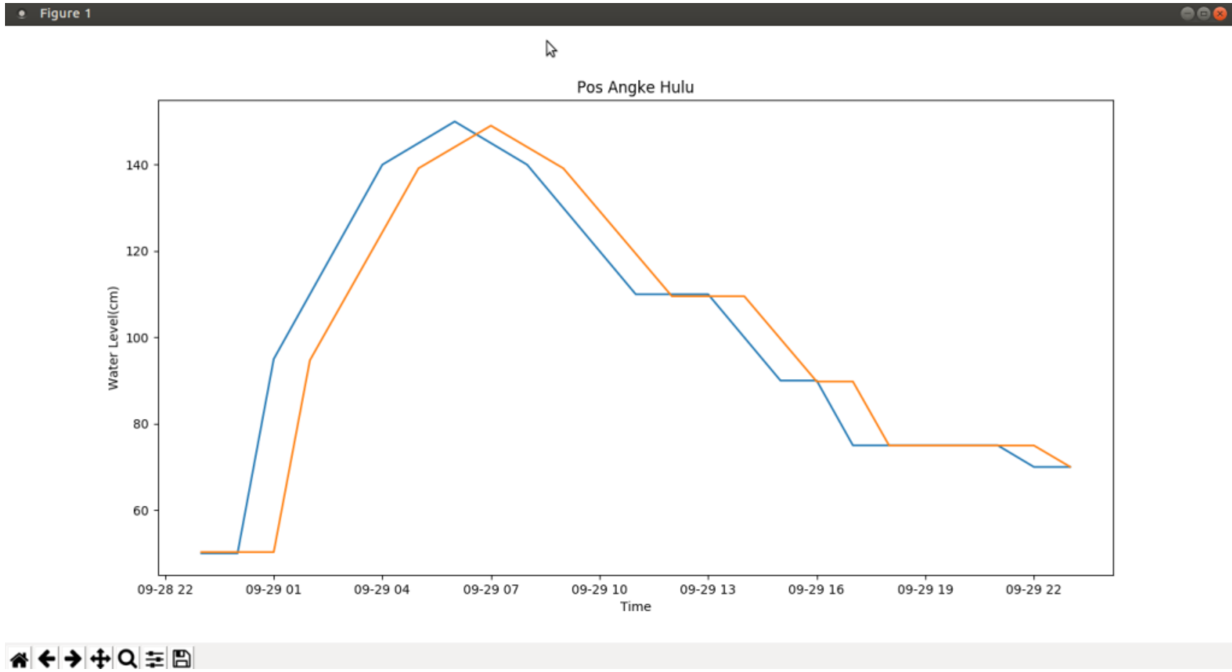
Notes:
Left LSTM – Right ARIMA
Orange → prediction line
Blue → Real data

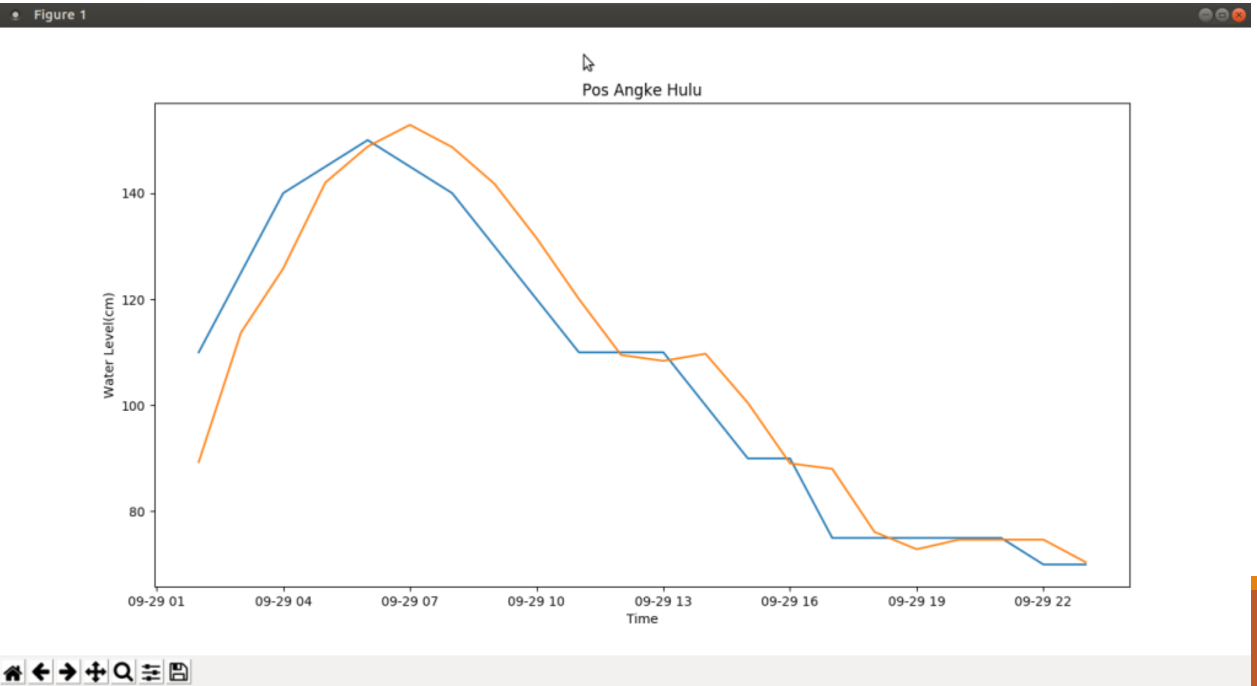




LSTM RNN vs ARIMA

Notes:
Left LSTM – Right ARIMA
Orange → prediction line
Blue → Real data



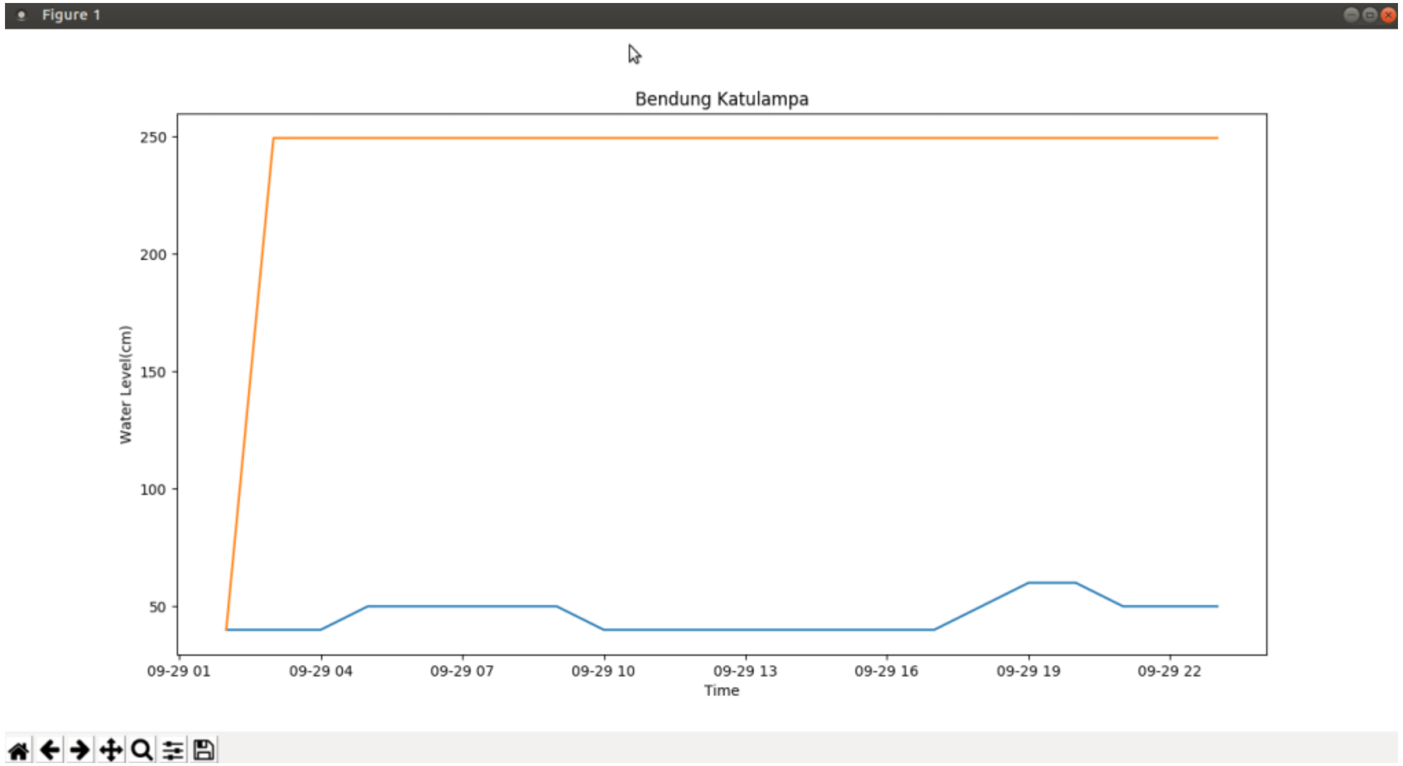


LSTM Recurrent Neural Network

Pintu Air	RMSE (cm)	
	ARIMA	LSTM RNN
Bendung Katulampa	4.47	4.69
Pos Depok	33.09	30.917
PA Manggarai	4.48	5.83
PA Karet	5.59	626
Pos Krukut Hulu	7.49	8.12
Pos Pesanggrahan	2.729	4.2
Pos Angke Hulu	11.84	8.76
Pos Waduk Pluit	4.65	5.16
Pasar Ikan	7.44	7.8
Pos Cipinang Hulu	6.31	2.86
Pos Sunter Hulu	0.24	2.51
PA Pulo Gadung	5.68	8.02

Permasalahan lainnya ketika menggunakan LSTM RNN adalah ketika ingin melalukan prediksi beberapa jam kedepan.

LSTM Recurrent Neural Network



Permasalahan lainnya ketika menggunakan LSTM RNN adalah ketika ingin melalukan prediksi beberapa jam kedepan.

Implementation

There are 33 models representing 11 river gauges (3 model for each).
The models are:

1. Auto-regressive with adjustment (adding delta variable based on river gauge’s relationship)
2. Auto-Arima (Hyndman, 2008)*
3. LSTM Recurrent Neural Network (Keras)

Notes:

LSTM RNN can be powerful using the right parameter but we use model 1(Auto-Regressive) due to its simplicity in the implementation

*Hyndman, R. J. and Y. Khandakar. 2008. Automatic Time Series Forecasting: The forecast
Package for R. Journal of Statistical Software. Vol 27.

https://www-01.ibm.com/software/industry/images

The model is implemented inside IOC-IBM for early warning system. Unfortunately, the system is not run automatically to predict flood in the future, still need input from user.

Conclusions and Future Works

- Conclusions
  - Heavy flood in Jakarta is shifting to February
  - North and West Jakarta are the area with higher number of flood reports
  - In production phase, simple model is more reliable due to its simplicity
  - LSTM → It is difficult to tune up the parameter
  - ARIMA → Simple, powerful, valid to be used in the system

- Future works
  - Create multivariate water level model
  - Create direct flood model using hourly water level and also rainfall data
  - Create real-time system that run automatically to predict water level and flood area