

Weather trend prediction and forecasting based on time-series model and deep neural network

3/24/19

Presented by: Muhan Hu, Han Weng, Fangrui Zhu

ECE657A Data and Knowledge Modeling and Analysis (Winter 2019)



UNIVERSITY OF
WATERLOO

| FACULTY OF
ENGINEERING

Outline

Part I. Target and data cleaning

Part II. Weather trend prediction

Part III. Weather forecasting

Part I. Target and data cleaning

Dataset

Goal

Spark/Hadoop data pre-processing

Data collection

- Raw Data:
 - 37 cities: Canada, United States, Israel
 - 6 years: 2012-2017
 - Hourly data
 - Temperature, humidity, pressure, wind direction, wind speed
 - ~9 million data points
- Goal
 - Weather trend (each feature) prediction – time-series regression
 - Weather forecasting - classification

Get data ready

Temperature
Wind direction
Wind speed
Humidity
Pressure

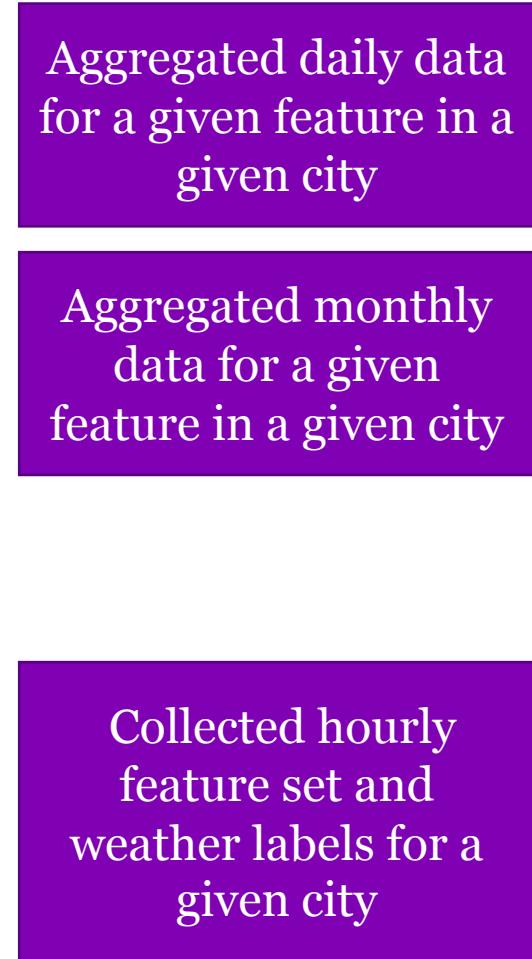
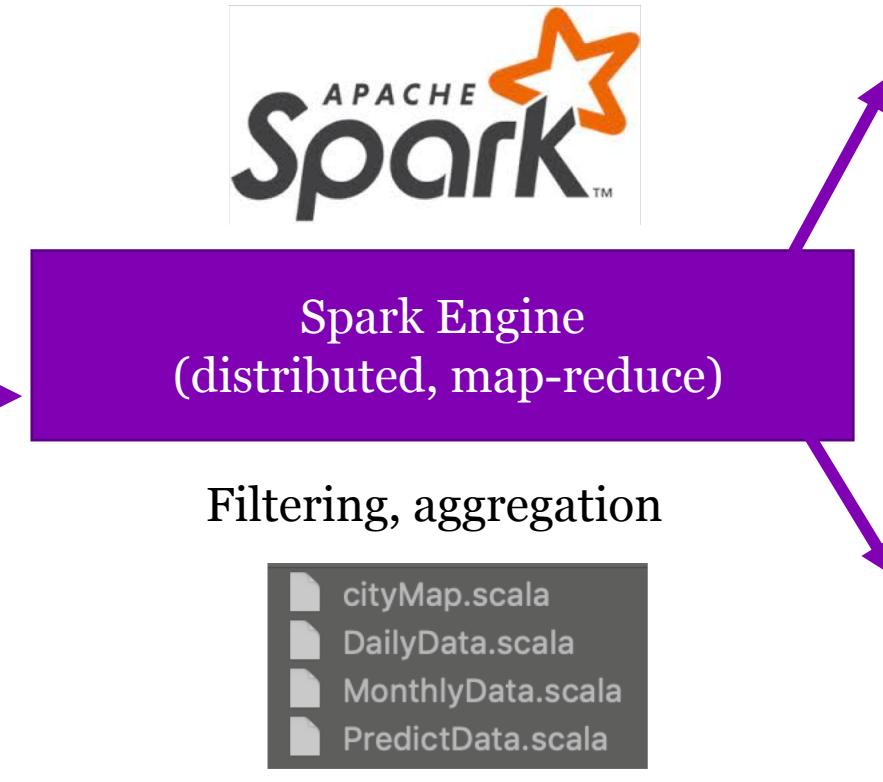
Get data ready



Get data ready



Get data ready



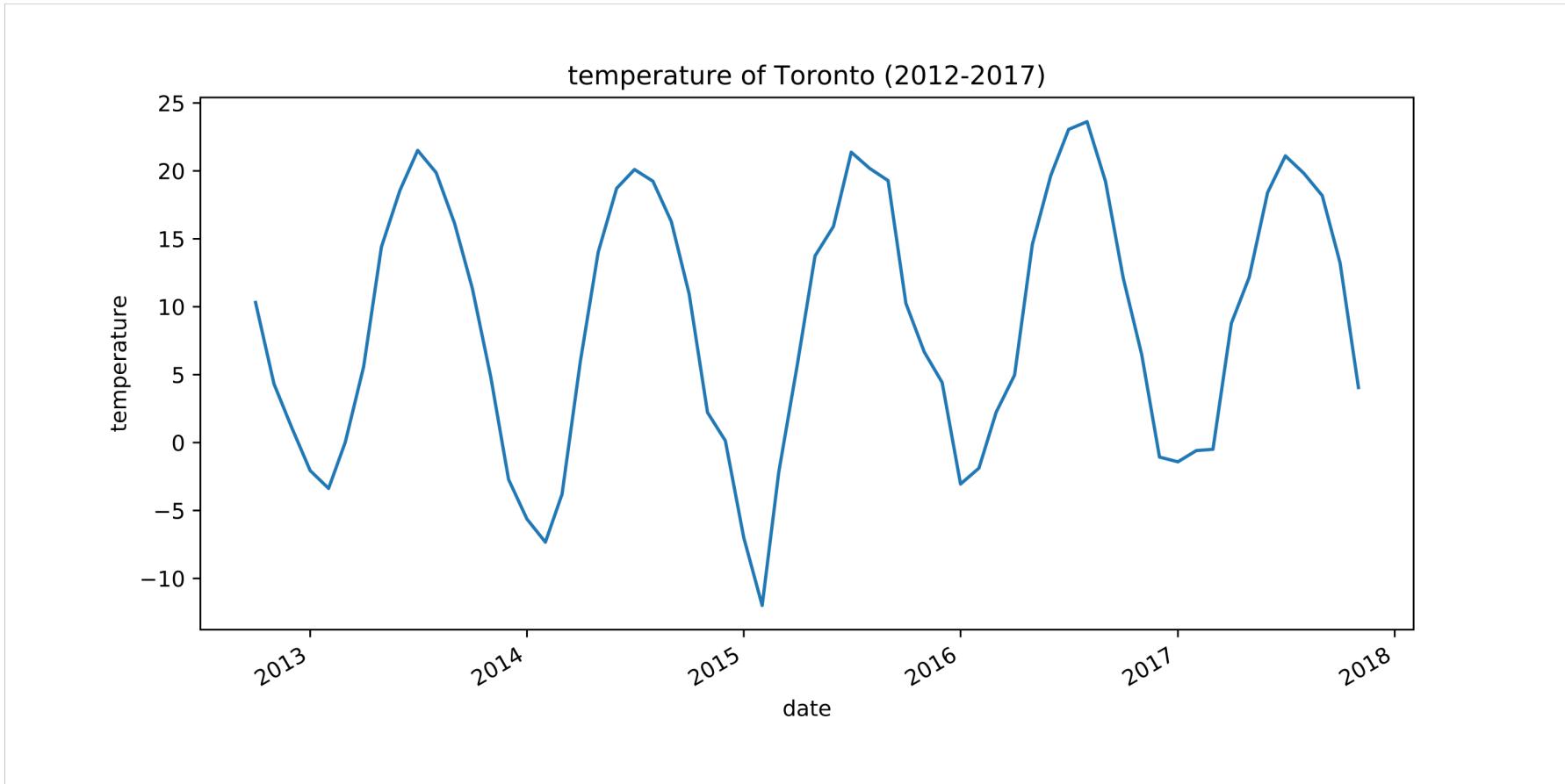
Part II. Weather trend prediction

Sample stationarity check

Time-series Modeling: AR, MA, ARIMA

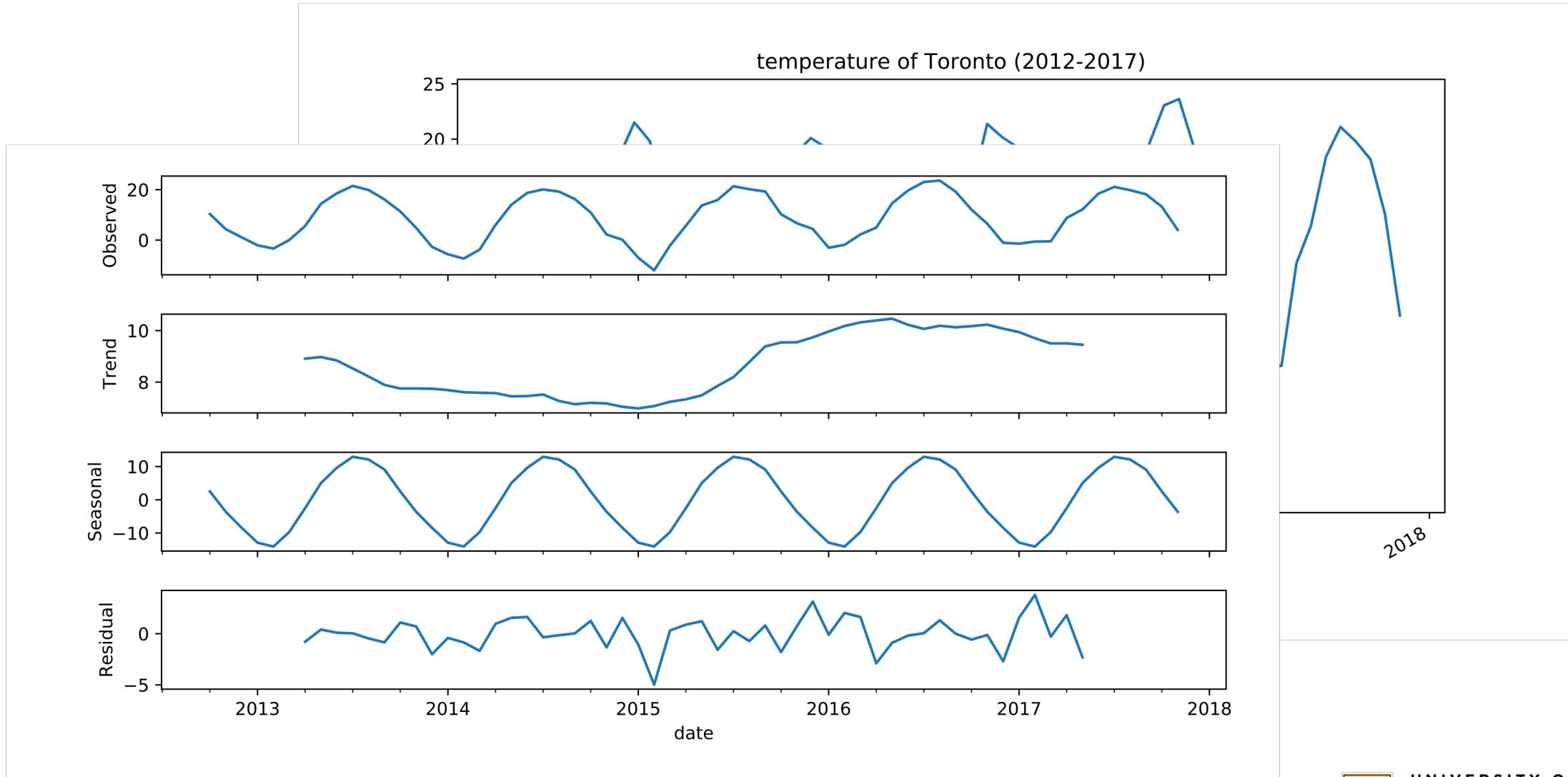
Trend prediction

Data stationarity – Original data

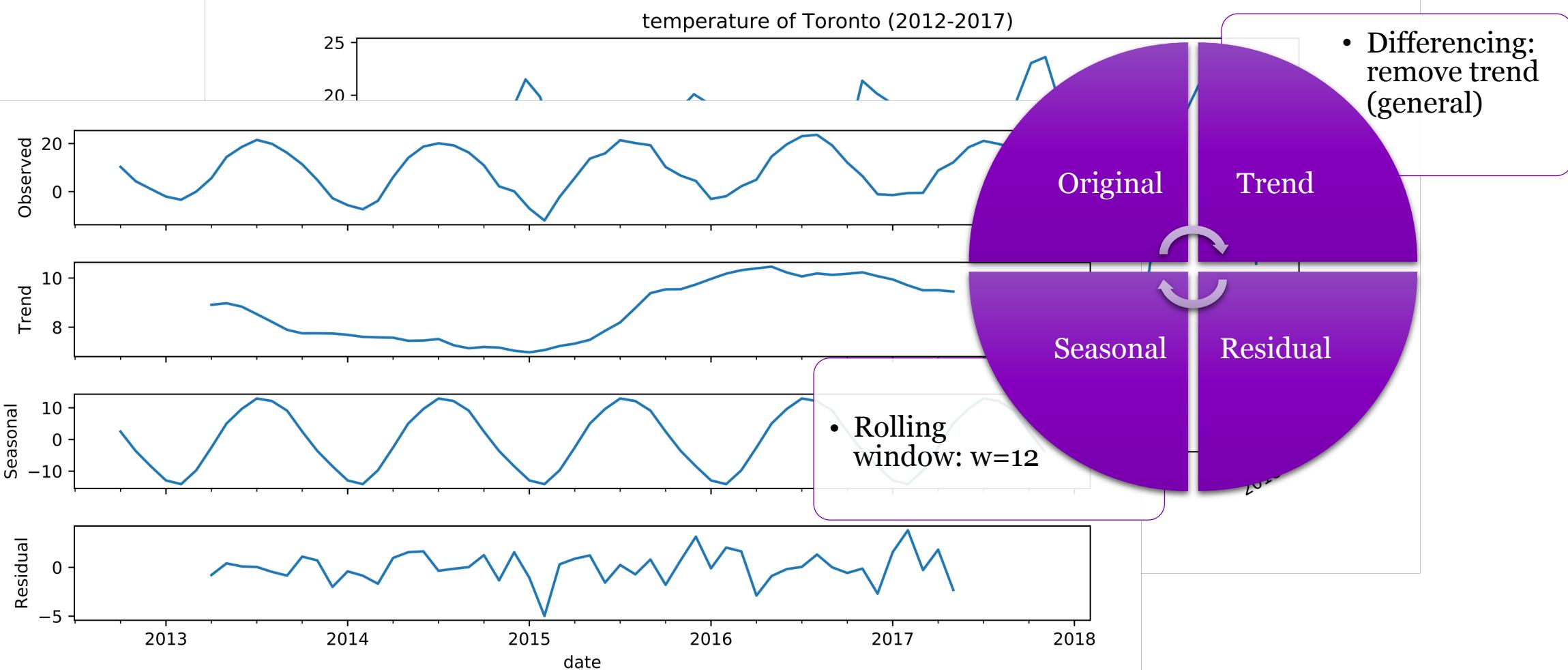


Original monthly temperature data for Toronto

Data stationarity – Improve



Data stationarity – Improve



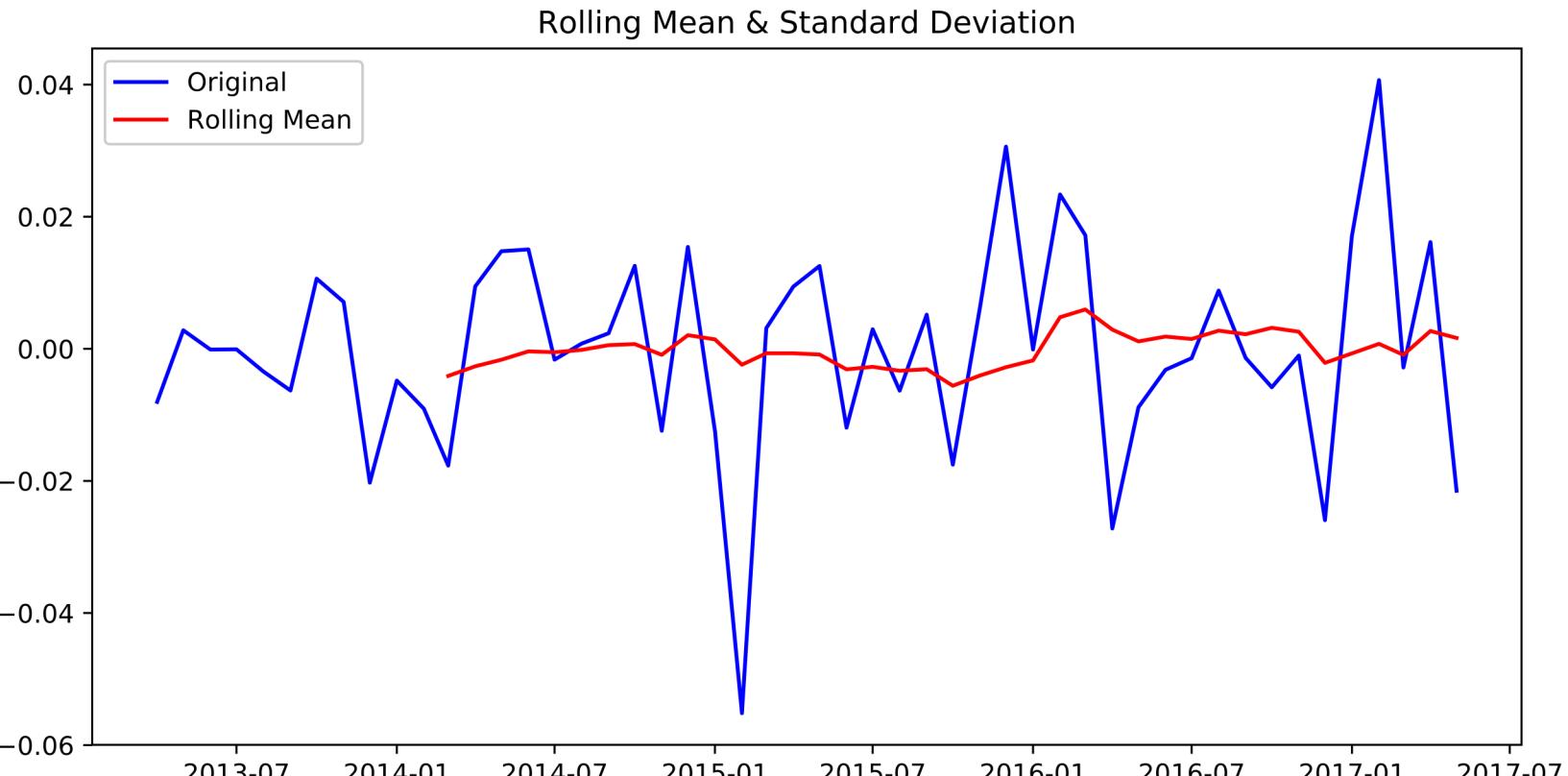
Data stationarity – Result check

Results of Dickey Fuller Test	Original	Log Scale Transformation	Time Shift Transformation
Test Statistic	-0.800512	-1.254808	-8.879796e+00
p-value	0.818972	0.649586	1.327455e-14
#Lags Used	10.000000	10.000000	9.000000e+00
Number of Observations Used	51.000000	40.000000	5.100000e+01
Critical Value (1%)	-3.565624	-3.605565	-3.565624e+00
Critical Value (5%)	-2.920142	-2.937069	-2.920142e+00
Critical Value (10%)	-2.598015	-2.606986	-2.598015e+00

Data stationarity – Result check

Results of Dickey Fuller Test	Original	Log Scale Transformation	Time Shift Transformation
Test Statistic	-0.800512	-1.254808	-8.879796e+00
p-value	0.818972	0.649586	1.327455e-14
#Lags Used	10.000000	10.000000	9.000000e+00
Number of Observations Used	51.000000	40.000000	5.100000e+01
Critical Value (1%)	-3.565624	-3.605565	-3.565624e+00
Critical Value (5%)	-2.920142	-2.937069	-2.920142e+00
Critical Value (10%)	-2.598015	-2.606986	-2.598015e+00

Data stationarity – Result check

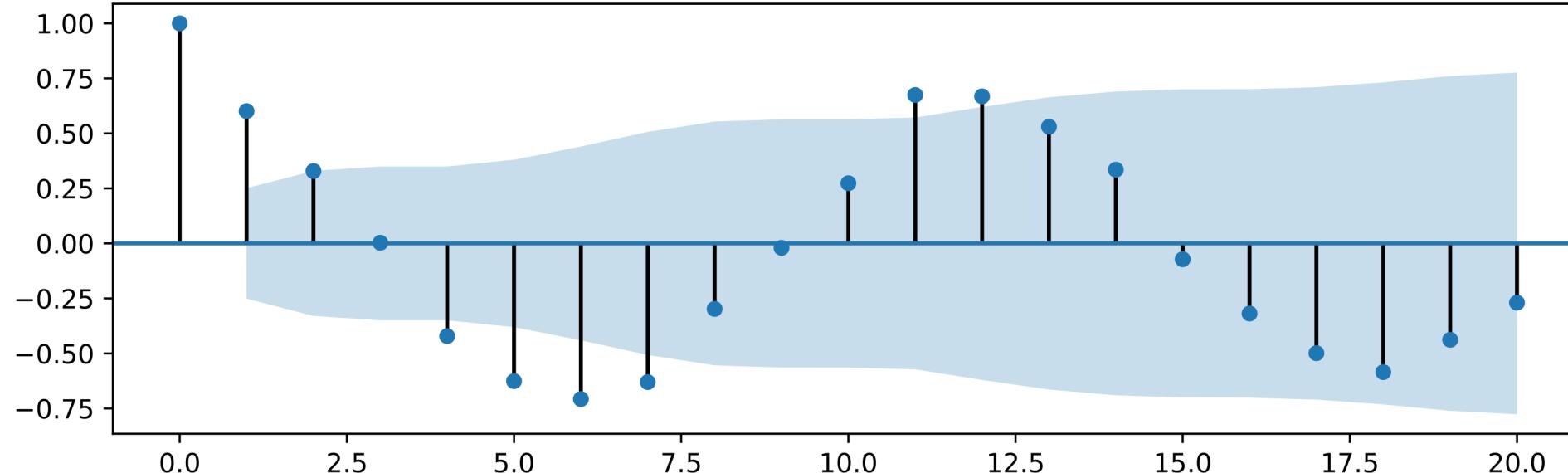


Results of Dickey Fuller Test	
Test Statistic	Original -3.298062
p-value	0.014965
#Lags Used	11.000000
Number of Observations Used	38.000000
Critical Value (1%)	-3.615509
Critical Value (5%)	-2.941262
Critical Value (10%)	-2.609200

Modelling – finding p, q value

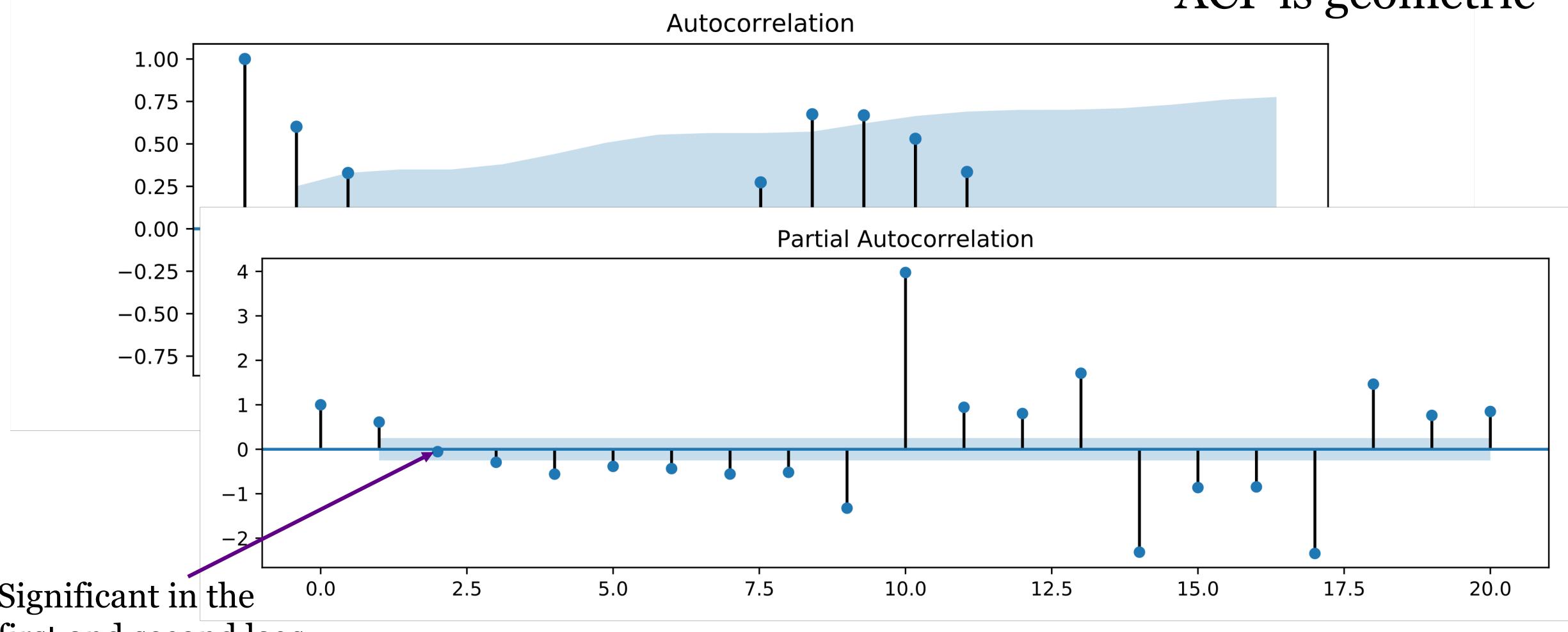
ACF is geometric

Autocorrelation



Modelling – finding p, q value

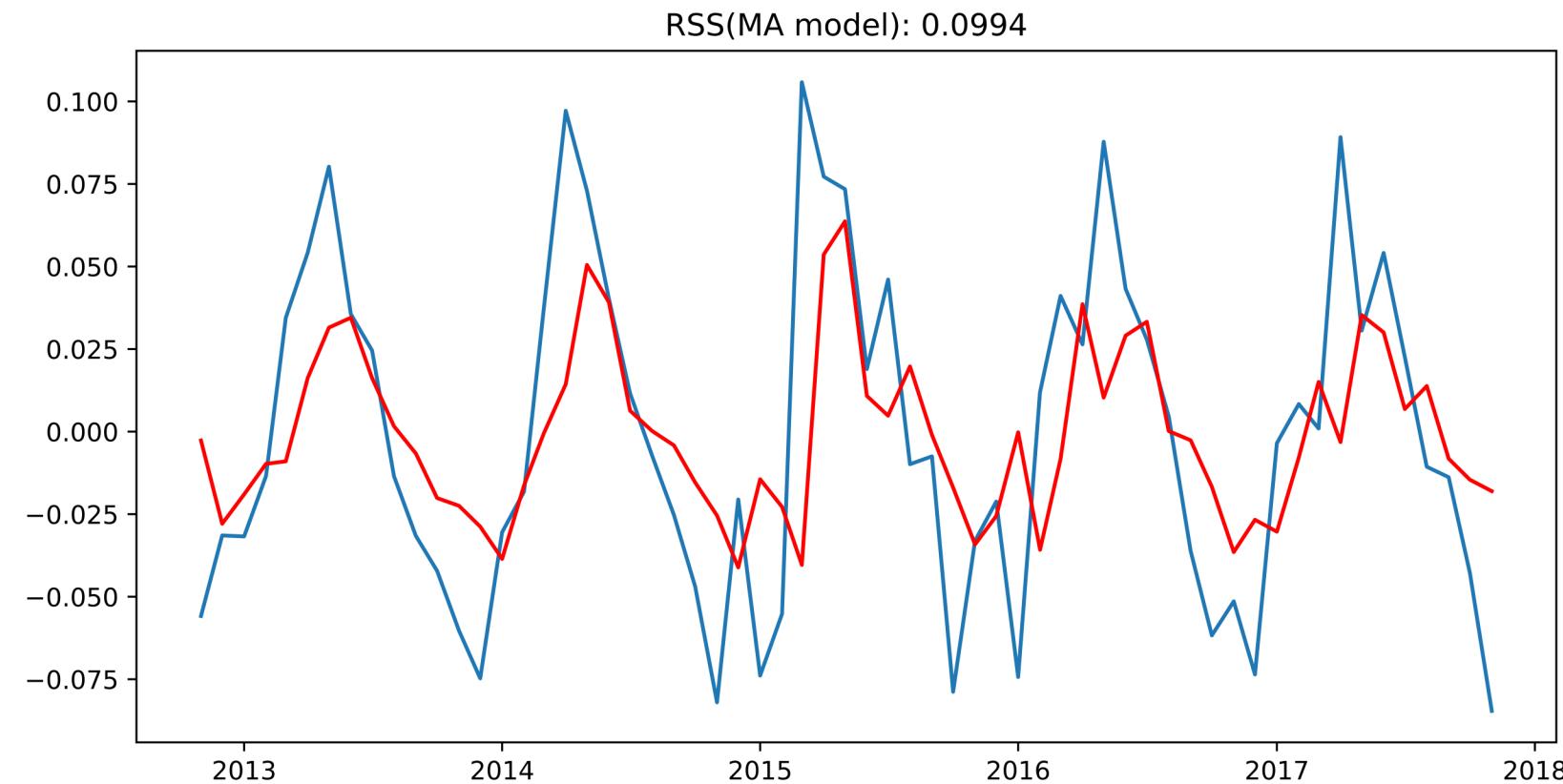
ACF is geometric



Models

- AR Model: consider p, d
- MA Model: consider q, d
- ARIMA: consider p, q and d

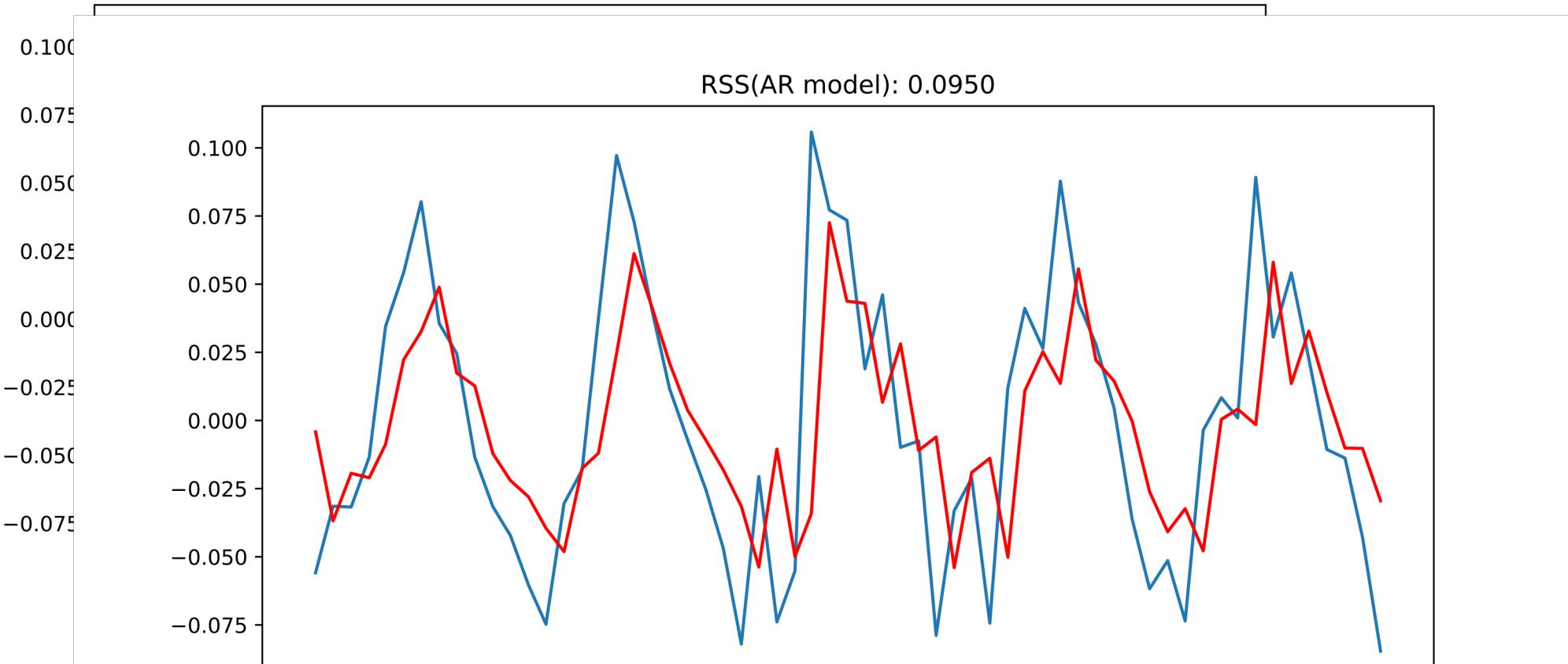
Models



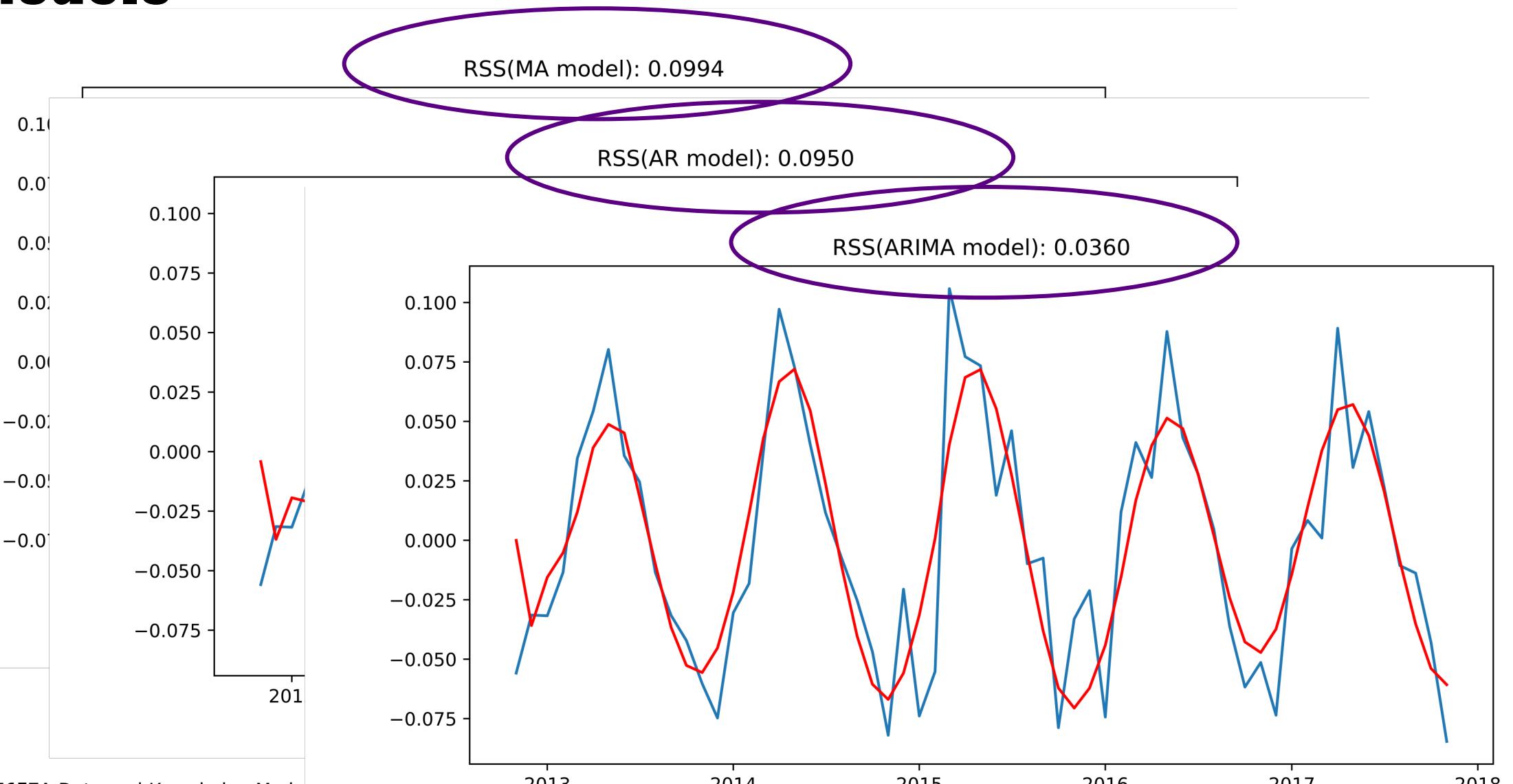
Models

RSS(MA model): 0.0994

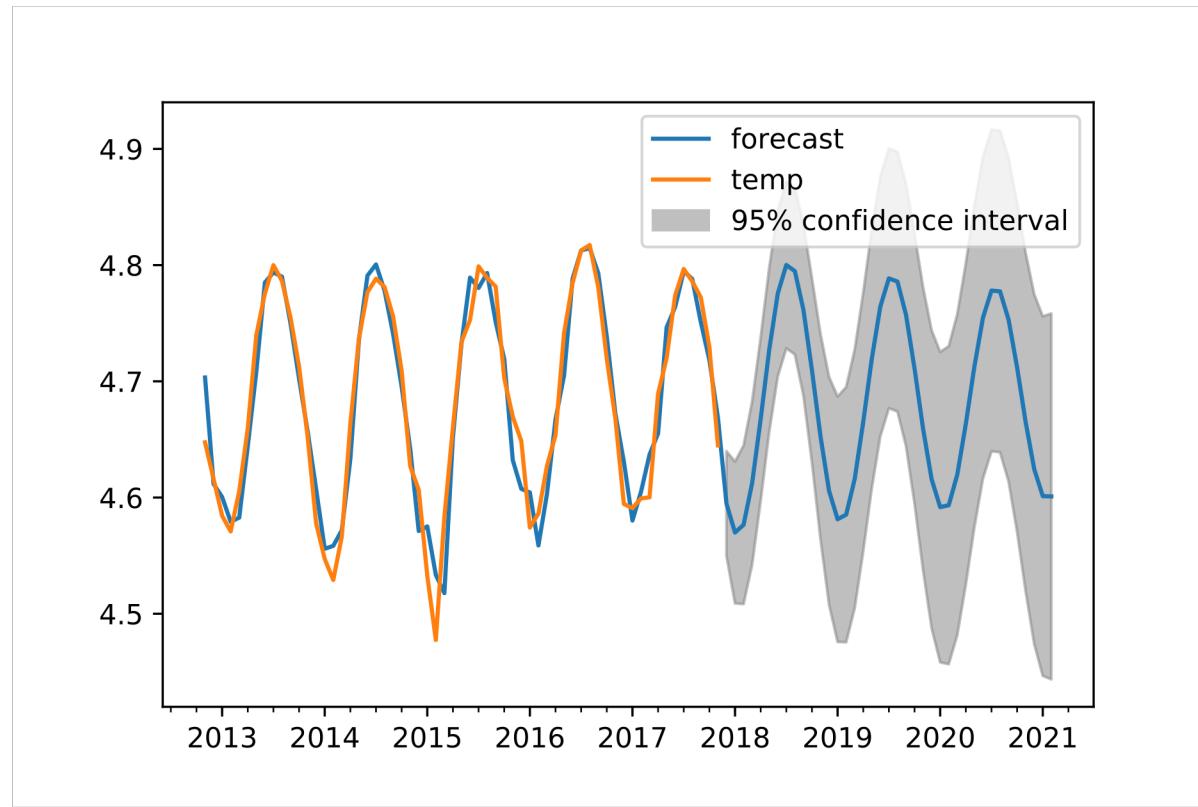
RSS(AR model): 0.0950



Models



Prediction results

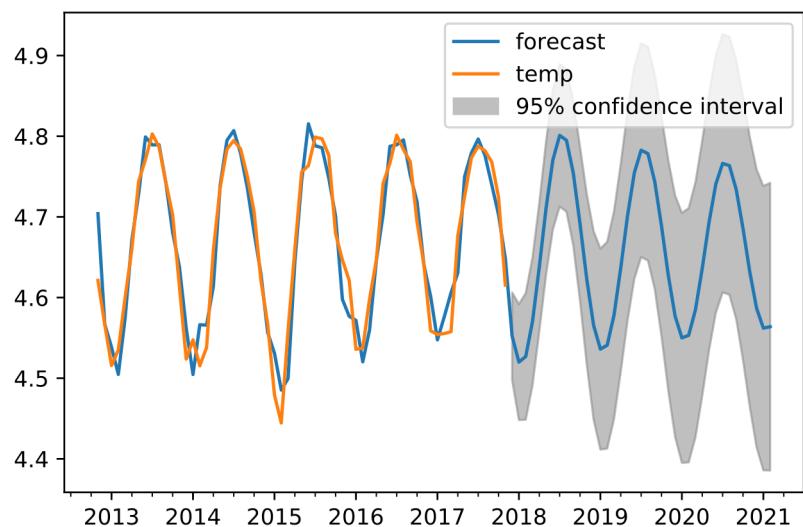


Temperature prediction (after log transformation)

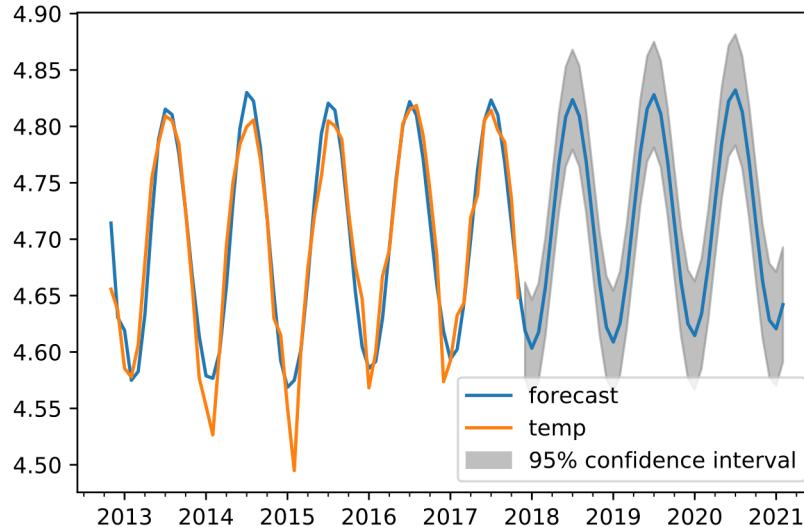
Date	Actual Data	Predicted Data
2017-07-01	21.106953	22.404260
2017-08-01	19.825096	21.353791
2017-09-01	18.181847	17.158485
2017-10-01	13.234041	11.012681
2017-11-01	4.049218	4.471090

Comparison between real and predicted data
(recovered to Celsius)

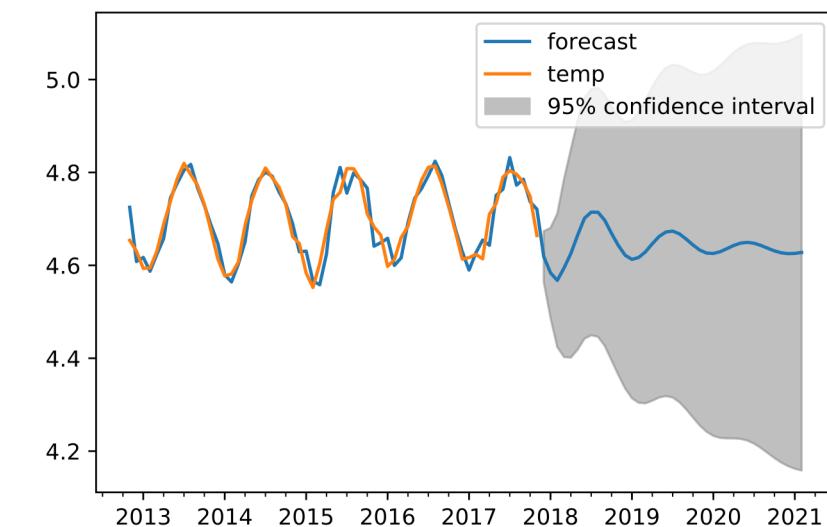
Prediction results for more cities



Montreal



Chicago



Boston

Prediction for other features (pressure, humidity, etc) and other cities also performed (not shown here)

Part III. Weather forecasting

Model training: DNN, Keras (TensorFlow backend)

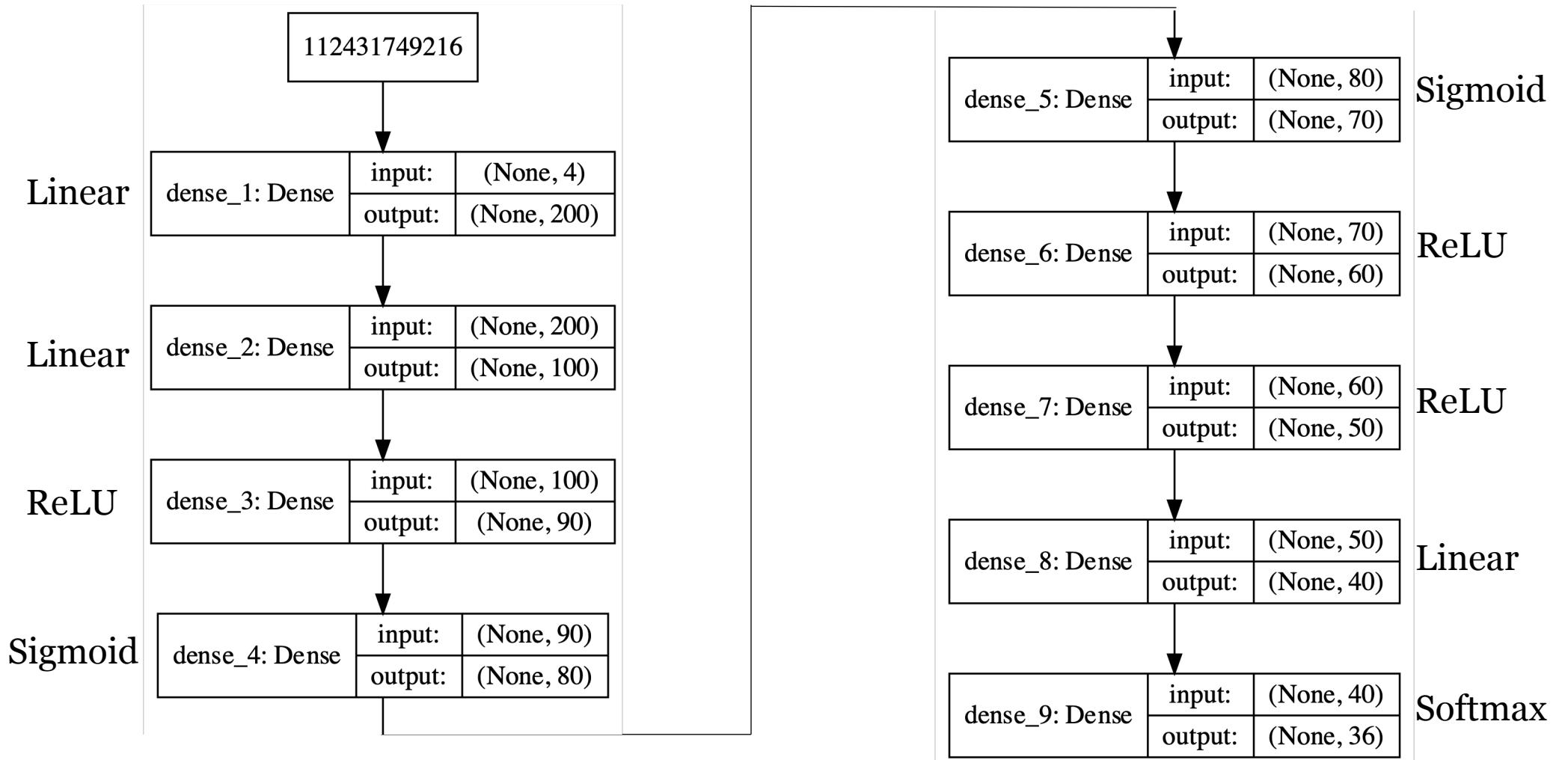
Result evaluation



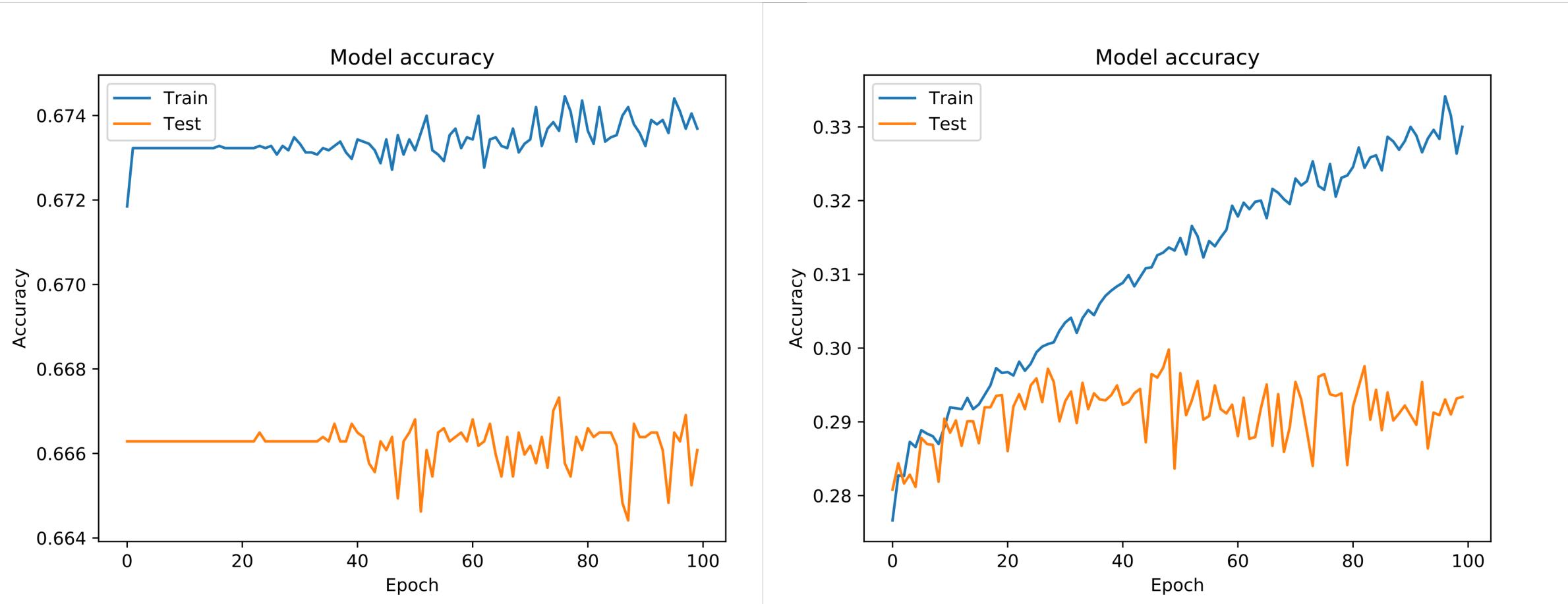
UNIVERSITY OF
WATERLOO

| FACULTY OF
ENGINEERING

DNN model

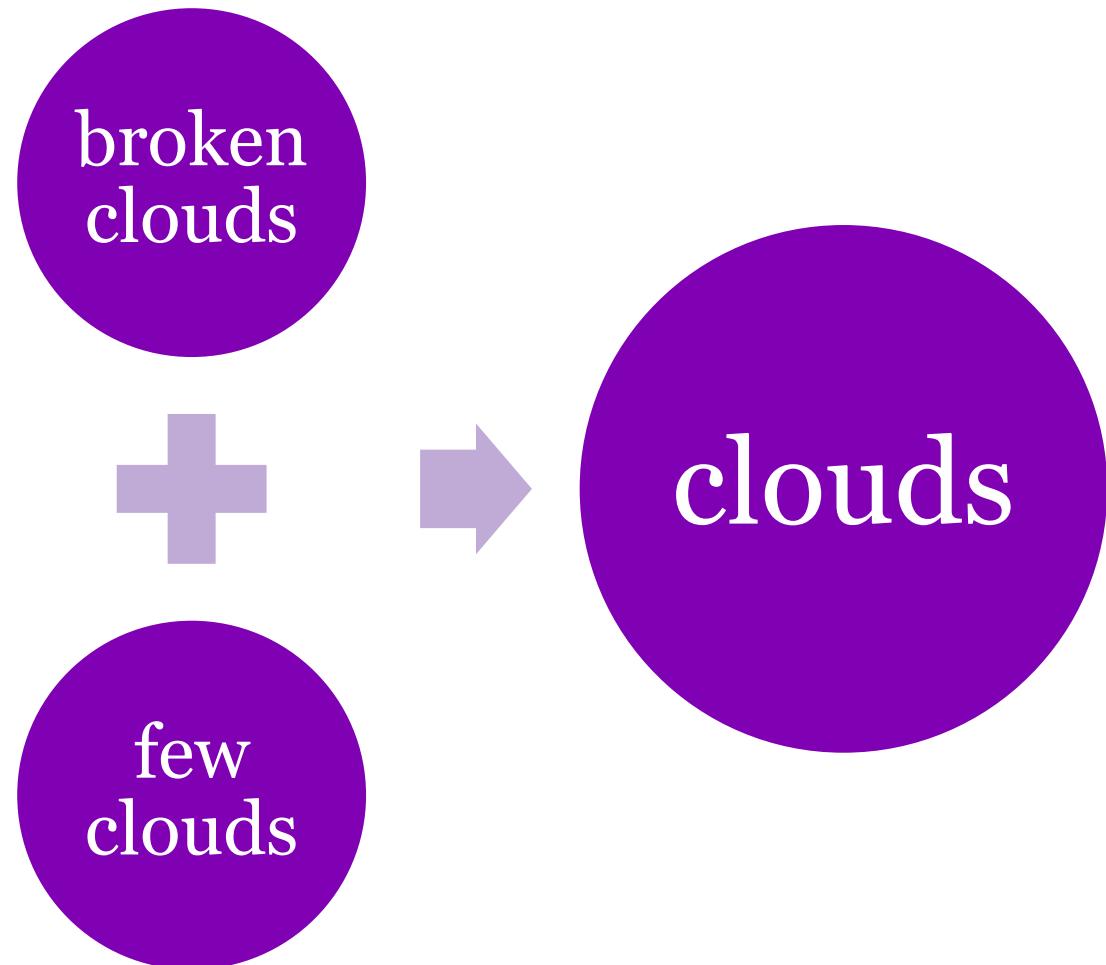


Original results (detailed labels with 35 classes)

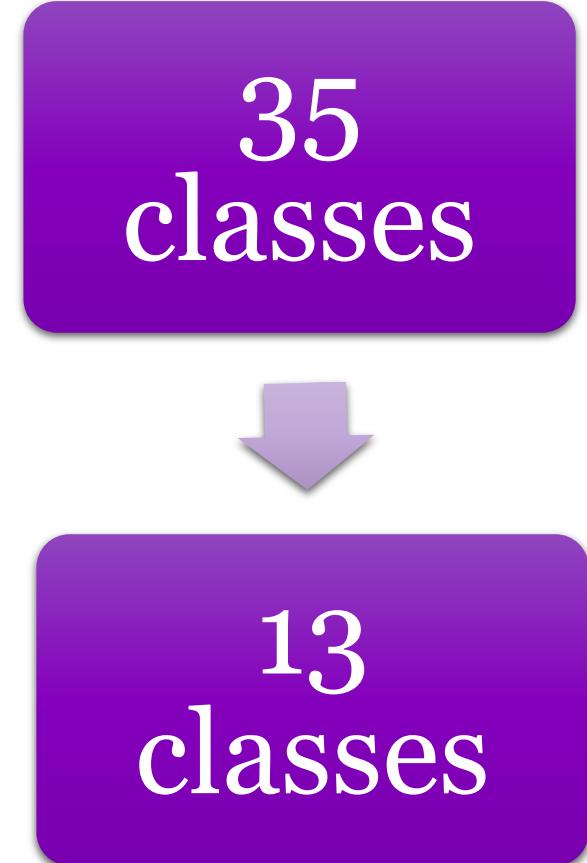


Classification accuracy in Phoenix and Vancouver (highest and lowest among all cities, respectively)

Accuracy improvement: reduce class numbers

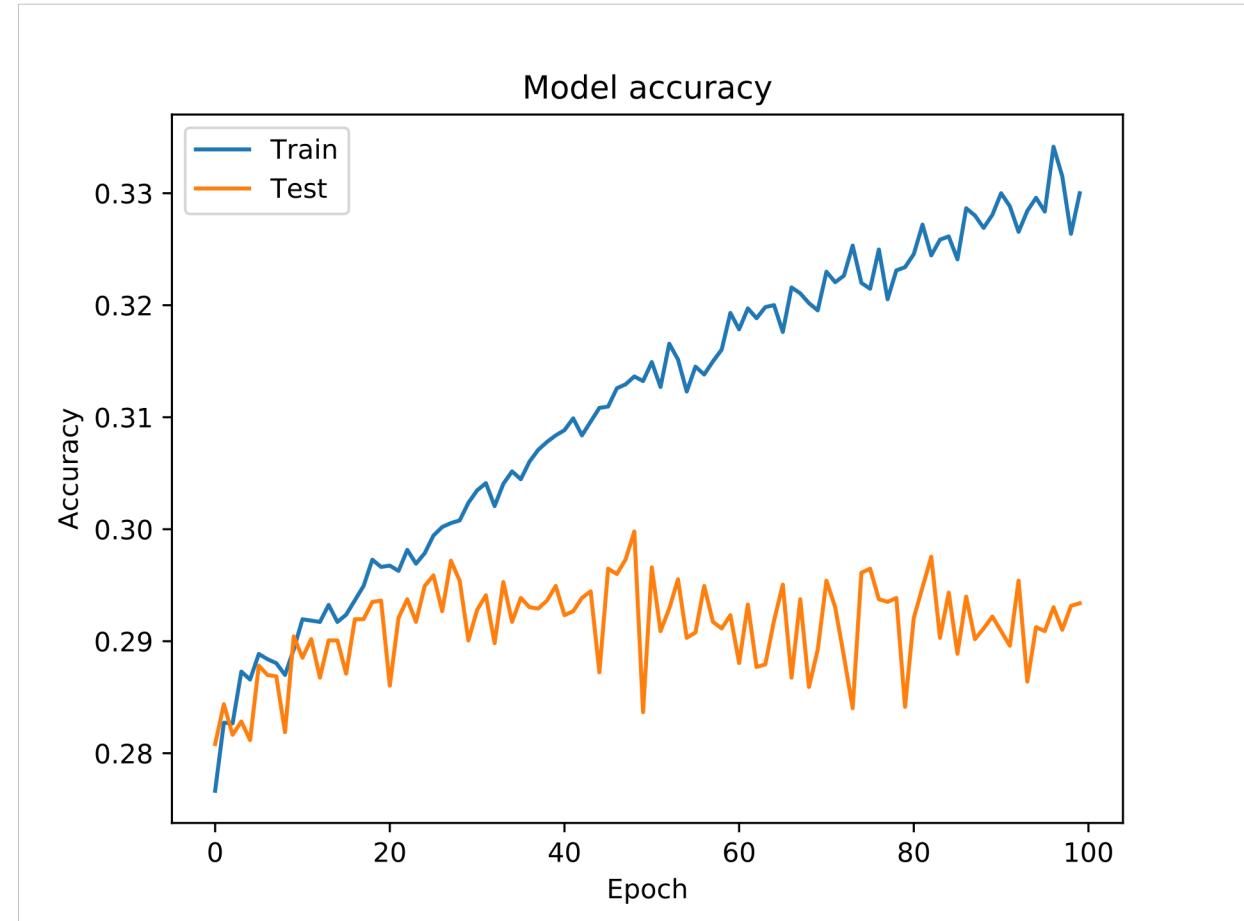


Again!



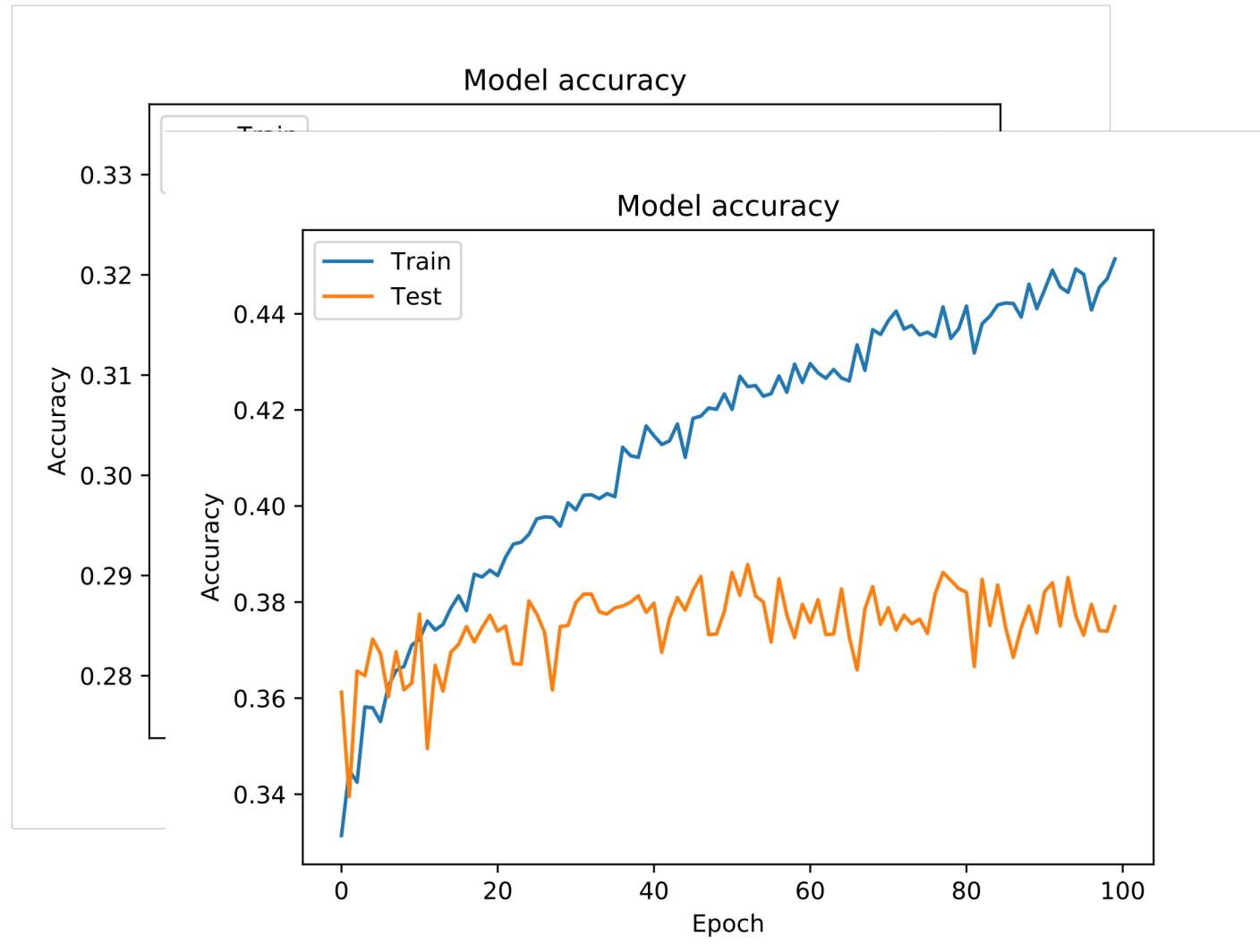
Accuracy improvement

- Phoenix: no obvious change
- Vancouver: 7.94% increase
- Toronto: 12.14% increase
- Portland: 15.45% increase
- Seattle: 8.4% increase

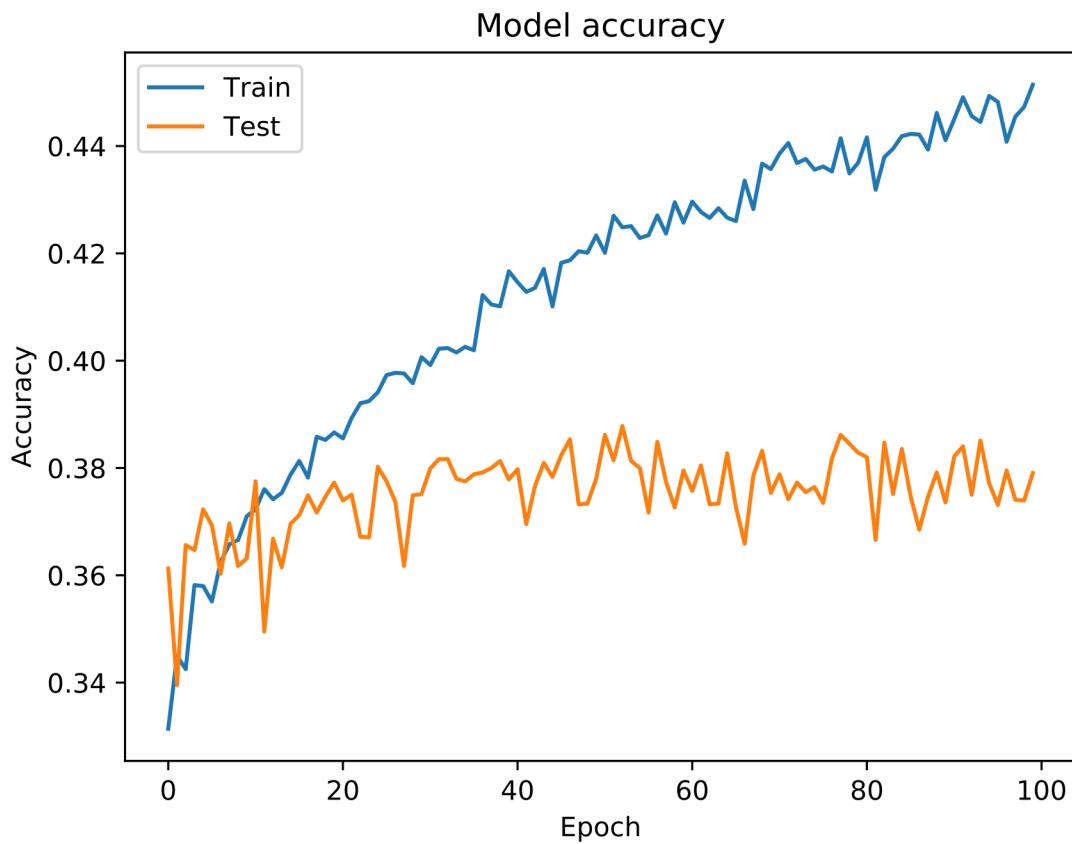


Accuracy improvement

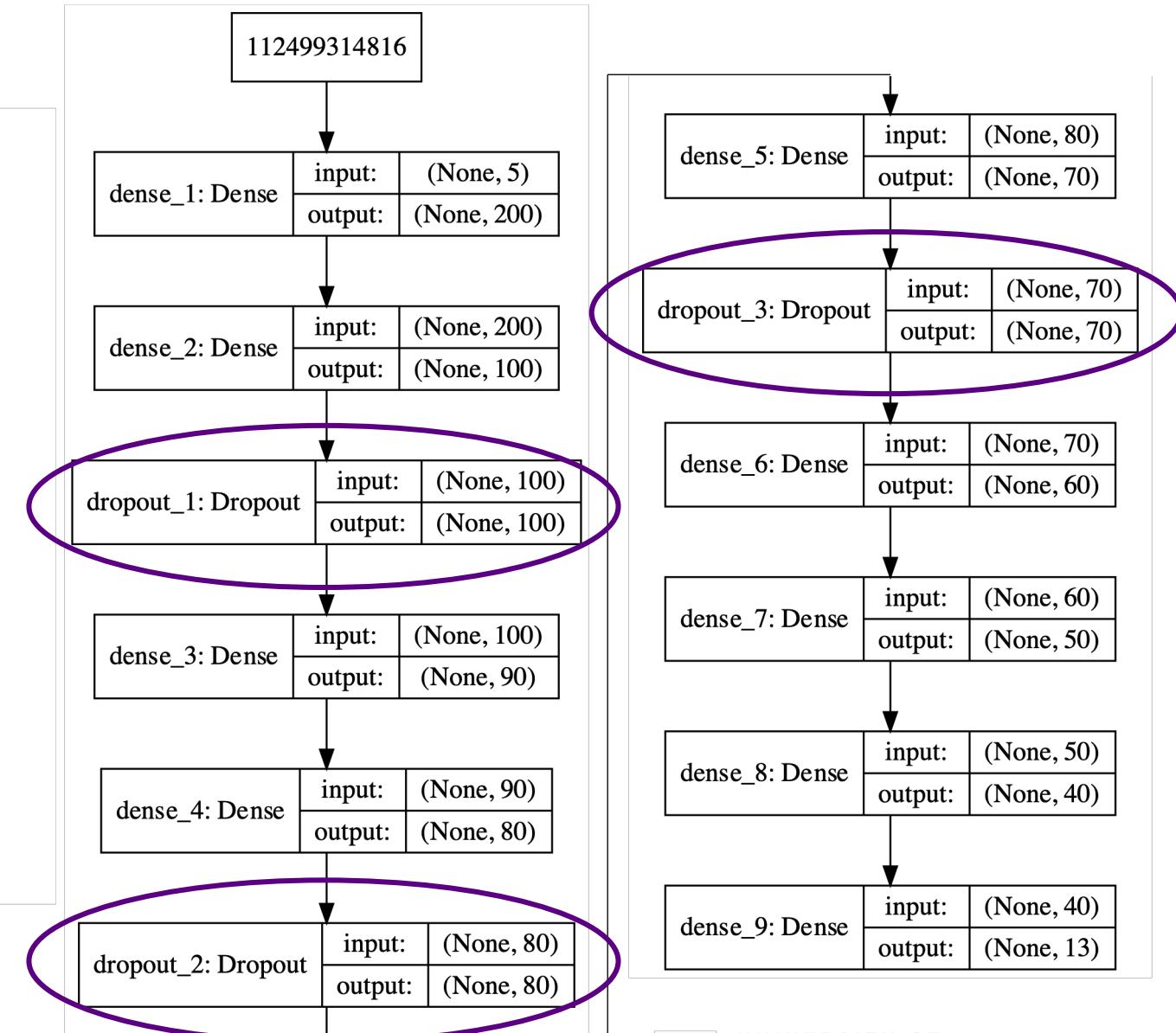
- Phoenix: no obvious change
- Vancouver: 7.94% increase
- Toronto: 12.14% increase
- Portland: 15.45% increase
- Seattle: 8.4% increase



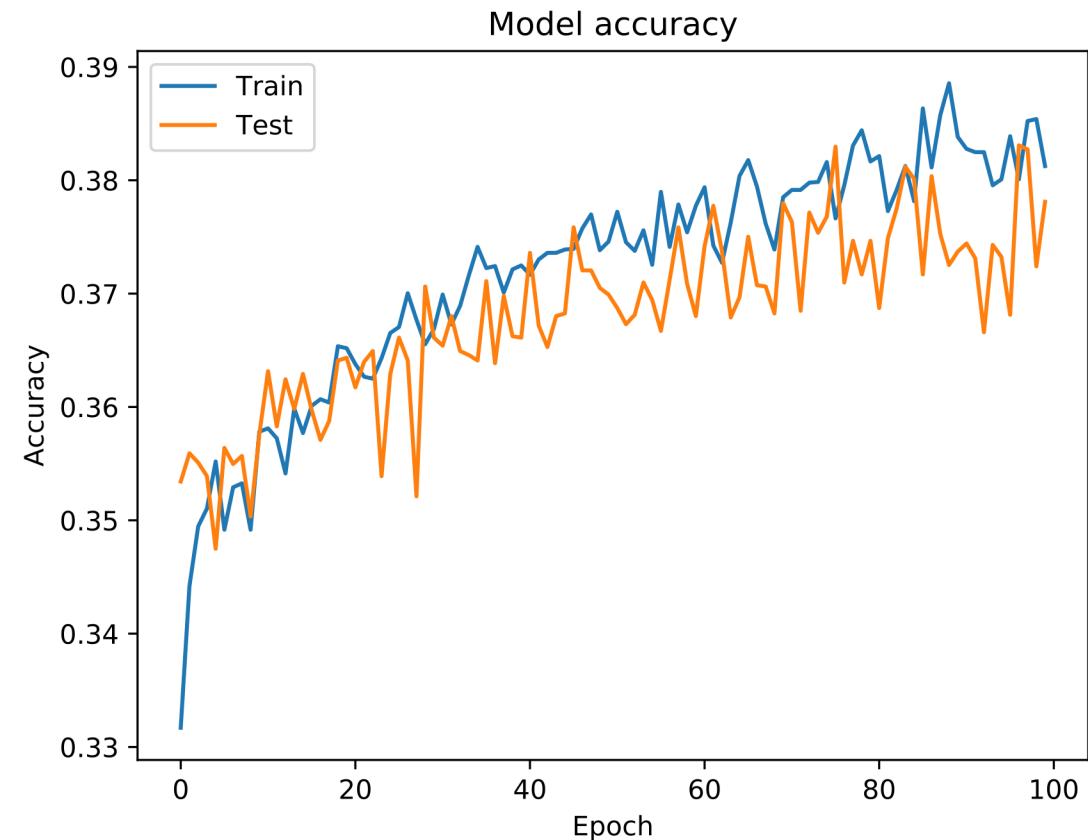
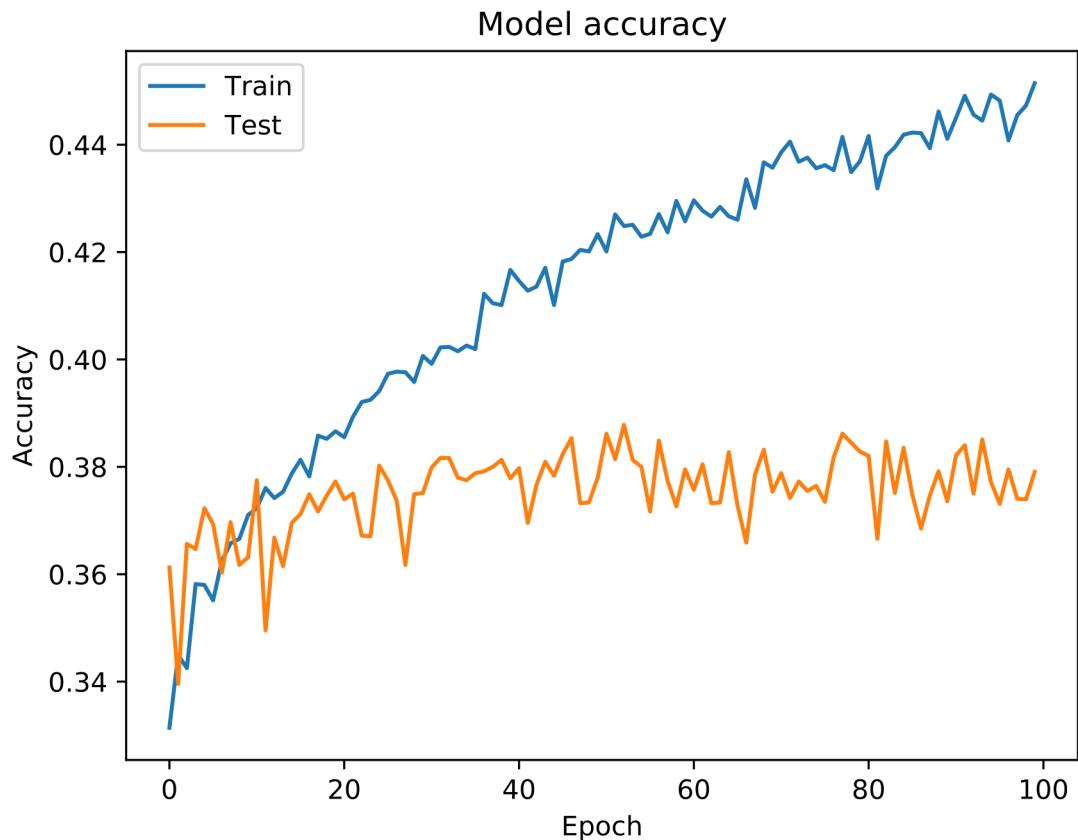
Overfitting check



Model accuracy result of Vancouver (overfit)

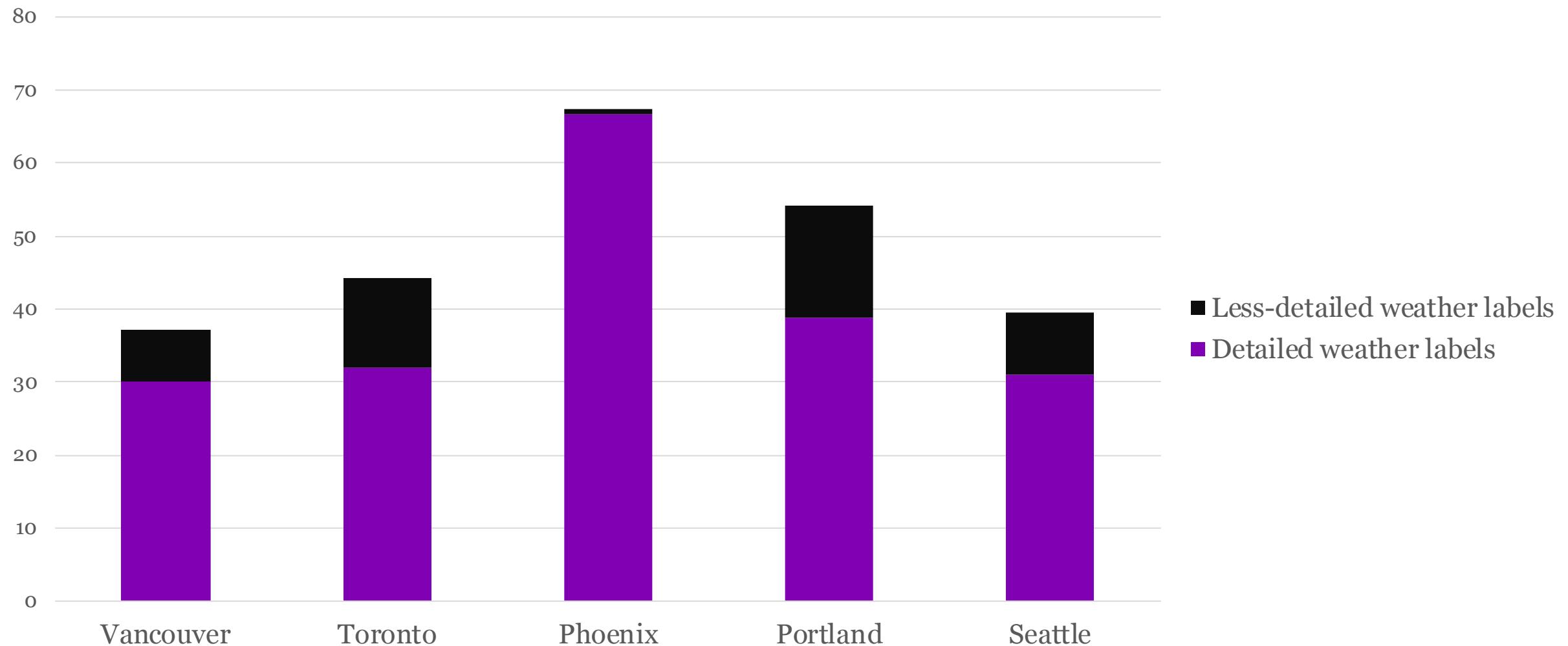


Overfitting check - optimized



Model accuracy result of Vancouver, left: before adding dropout, right: after adding dropout.

Forecasting Accuracy for selected North American cities





UNIVERSITY OF
WATERLOO



FACULTY OF ENGINEERING

Thanks for your attention!

Backup slides: Toronto time series result, using daily data

