

# Tuning a Weather Model

Paulo Jauregui, Anjali Golechha, Sooyeon Kim

Mentor: Dr. Wendell Nuss

Why might weather forecasting  
be important?

# Secret Sauce



**INGREDIENTS:** COOKED RICE (WATER, LONG GRAIN BROWN RICE, RED RICE), BROCCOLI, SHRIMP (SHRIMP, SALT), WATER, RED BELL PEPPERS, BABY CORN, EDAMAME, SRIRACHA PASTE (CHILI PEPPER PUREE, DISTILLED WHITE VINEGAR, CANE SUGAR, SEA SALT, GARLIC PUREE, EXPELLER PRESSED CANOLA OIL, GARLIC POWDER, XANTHAN GUM), SOY SAUCE (WATER, SOYBEANS, RICE, SALT), DICED ONION, EXPELLER PRESSED CANOLA OIL, GARLIC PUREE (GARLIC, WATER), CORNSTARCH, PAPRIKA OLEORESIN (FOR COLOR), SHRIMP EXTRACT, SALT.

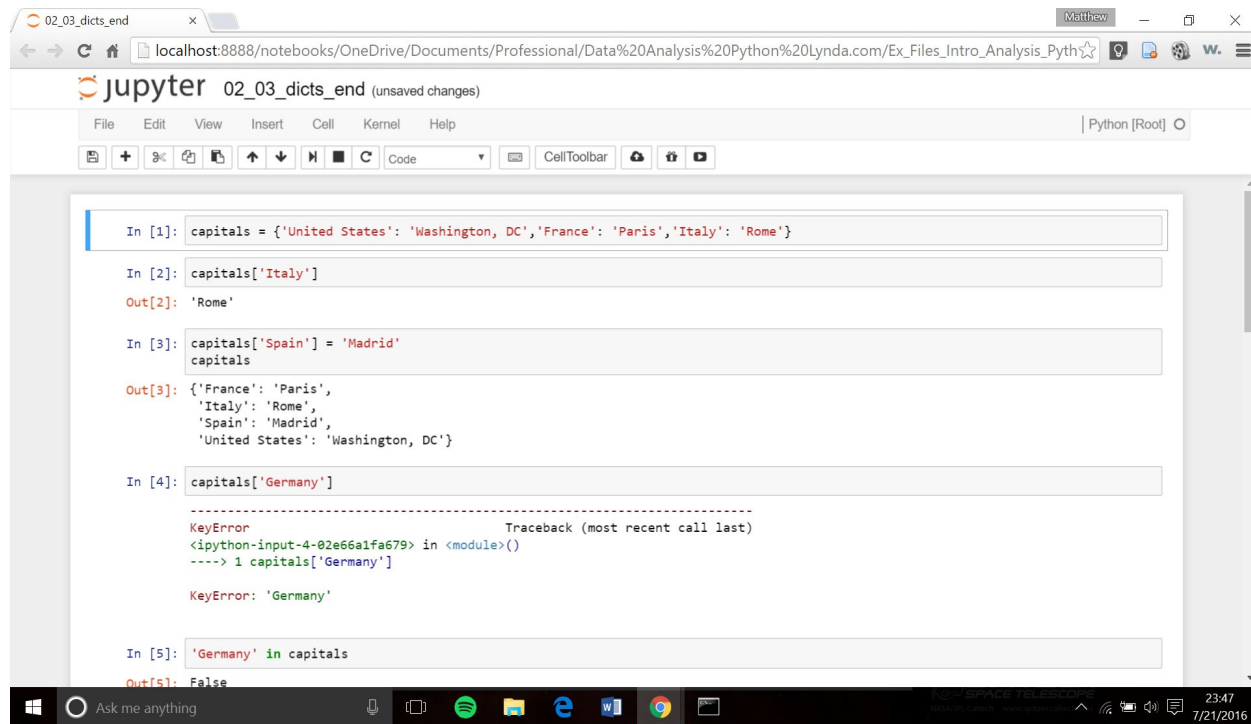
**CONTAINS SOY, SHRIMP.**

DIST. & SOLD EXCLUSIVELY BY:  
TRADER JOE'S, MONROVIA, CA 91016

PRODUCT OF CANADA

# Project Overview

- Jupyter Notebook
- Python



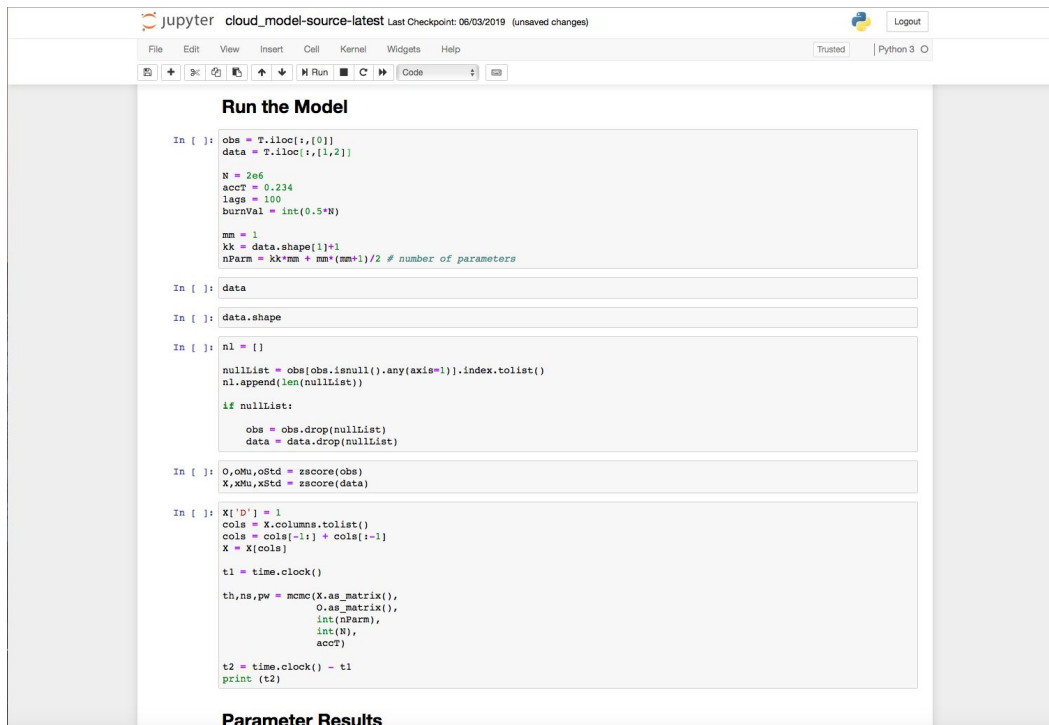
The screenshot displays a Jupyter Notebook window titled "02\_03\_dicts\_end" with unsaved changes. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Help) and a toolbar. The notebook content shows five input cells and their corresponding outputs:

```
In [1]: capitals = {'United States': 'Washington, DC', 'France': 'Paris', 'Italy': 'Rome'}  
  
In [2]: capitals['Italy']  
Out[2]: 'Rome'  
  
In [3]: capitals['Spain'] = 'Madrid'  
         capitals  
Out[3]: {'France': 'Paris',  
         'Italy': 'Rome',  
         'Spain': 'Madrid',  
         'United States': 'Washington, DC'}  
  
In [4]: capitals['Germany']  
-----  
KeyError                                Traceback (most recent call last)  
<ipython-input-4-02e66a1fa679> in <module>()  
----> 1 capitals['Germany']  
  
KeyError: 'Germany'  
  
In [5]: 'Germany' in capitals  
Out[5]: False
```

The Windows taskbar at the bottom shows the time as 23:47 on 7/21/2016.

# Model Explanation

- Evaluating learning machine
- Using pre-existing data points to determine specific test dates
- The more training data used, the more accurate the test data will be (Hypothesis)



```
jupyter cloud_model-source-latest Last Checkpoint: 06/03/2019 (unsaved changes)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
Run the Model
In [ ]: obs = T.iloc[:,0]
data = T.iloc[:,1,2]

N = 2e6
acct = 0.234
lags = 100
burnVal = int(0.5*N)

mm = 1
kk = data.shape[1]+1
nParm = kk*mm + mm*(mm+1)/2 # number of parameters

In [ ]: data

In [ ]: data.shape

In [ ]: nl = []

nullList = obs[obs.isnull().any(axis=1)].index.tolist()
nl.append(len(nullList))

if nullList:
    obs = obs.drop(nullList)
    data = data.drop(nullList)

In [ ]: O,OMu,sStd = zscore(obs)
X,xMu,xStd = zscore(data)

In [ ]: X['D'] = 1
cols = X.columns.tolist()
cols = cols[-1:] + cols[:-1]
X = X[cols]

t1 = time.clock()

th,ns,pw = mcmc(X.as_matrix(),
               O.as_matrix(),
               int(nParm),
               int(N),
               acct)

t2 = time.clock() - t1
print (t2)

Parameter Results
```

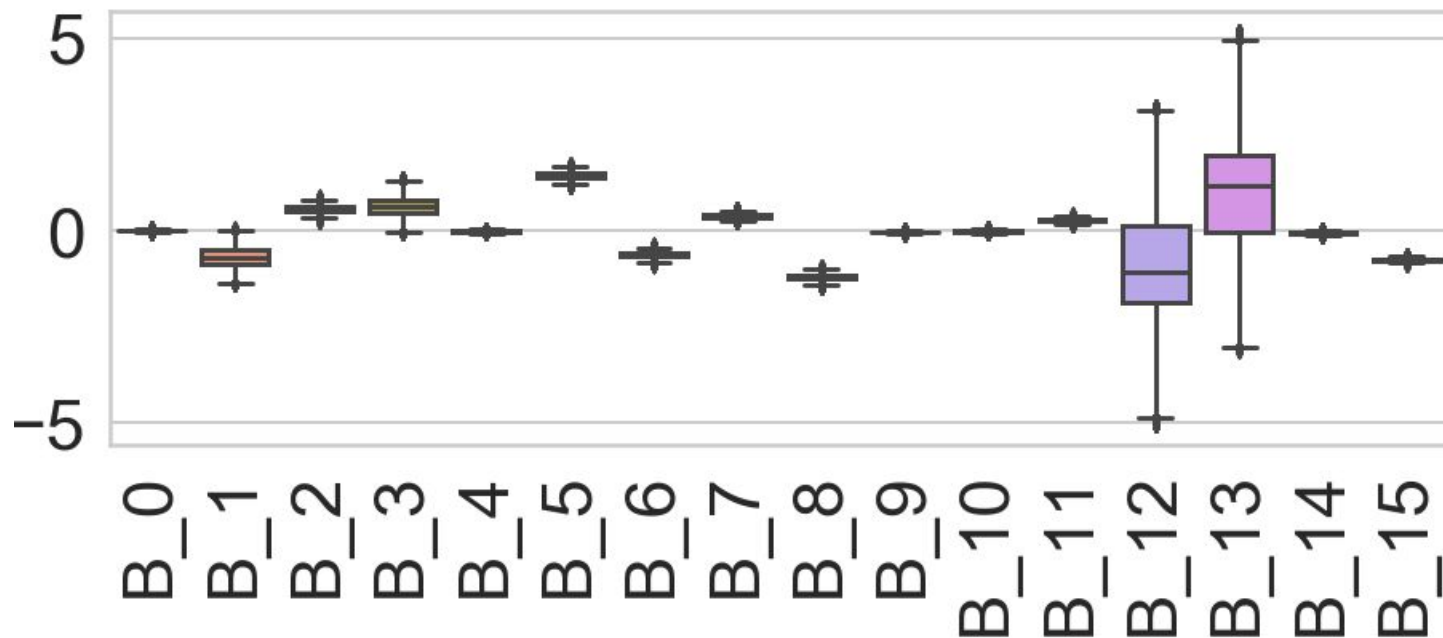
# Breaking down the Data Sets

File Name Format: Year/Month/Day/Time - ex. 1606081500Z.pts

- Cloud data sets have a total of 15 columns
- Wind data sets have a total of 10 columns

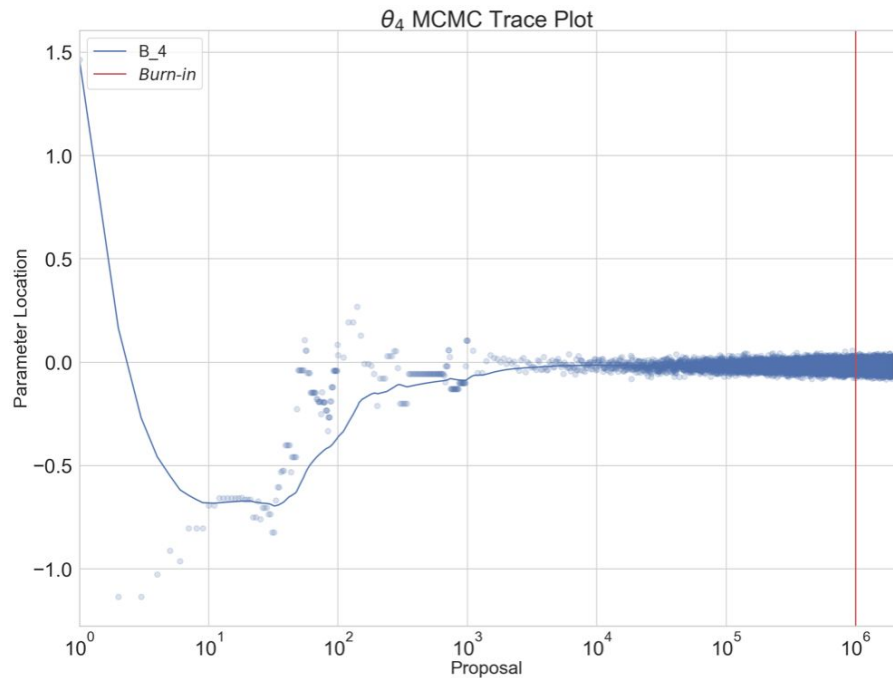
# Predictors

# Box Plot

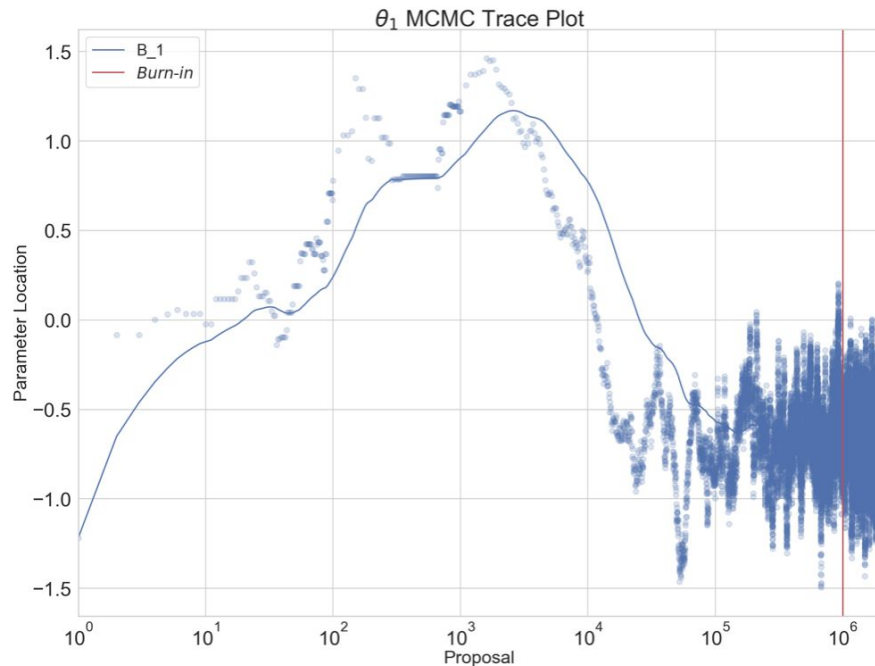




# Trace Plot

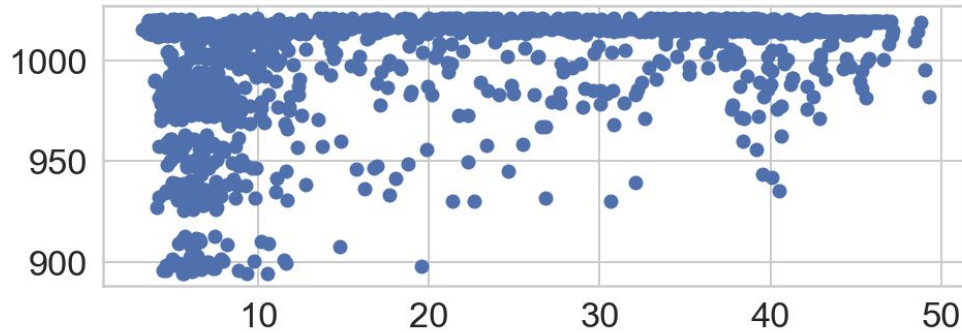


column 4: maximum vertical motion

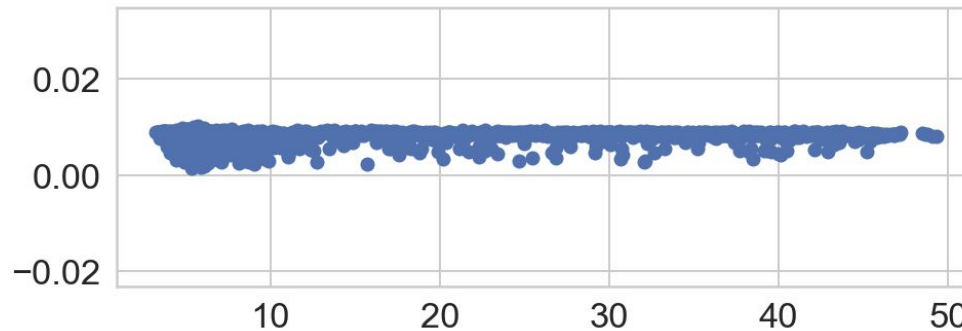


column 1: average cloud water

# Scatter Plot



column 11: surface pressure



column 8: maximum specific humidity

# Most effective/Promising

Cloud - maximum vertical motion, maximum relative humidity, average relative humidity, sea level pressure, surface pressure

Wind - N/A (all columns were needed to run the model)

# Understanding the Results

tt: verification

bz: prediction

de: difference between tt and bz

\* these arrays were used to generate heat maps,  
scatterplots, stream plots, and vector plots \*

```
de = []
bz = []
tt = []

for idx in range(t.shape[0]):

    pv = array(t.iloc[idx,[1,2,3,4,5]].tolist())
    truth = array(t.iloc[idx,[0]].tolist())
    fs = ppdNorm(bM,cm,pv,
                xMu,
                xStd,
                oMu,
                oStd,
                mm, kk)

    fm = np.median(fs,axis=0)
    bz.append(fm)
    tt.append(truth)
    de.append((truth - fm).tolist())

FS = pd.DataFrame(fs,columns=['I'])
# FS['day'] = idx
# FS = FS.set_index(['day'])

if (idx == 0):

    dfFS = FS.copy()

else:

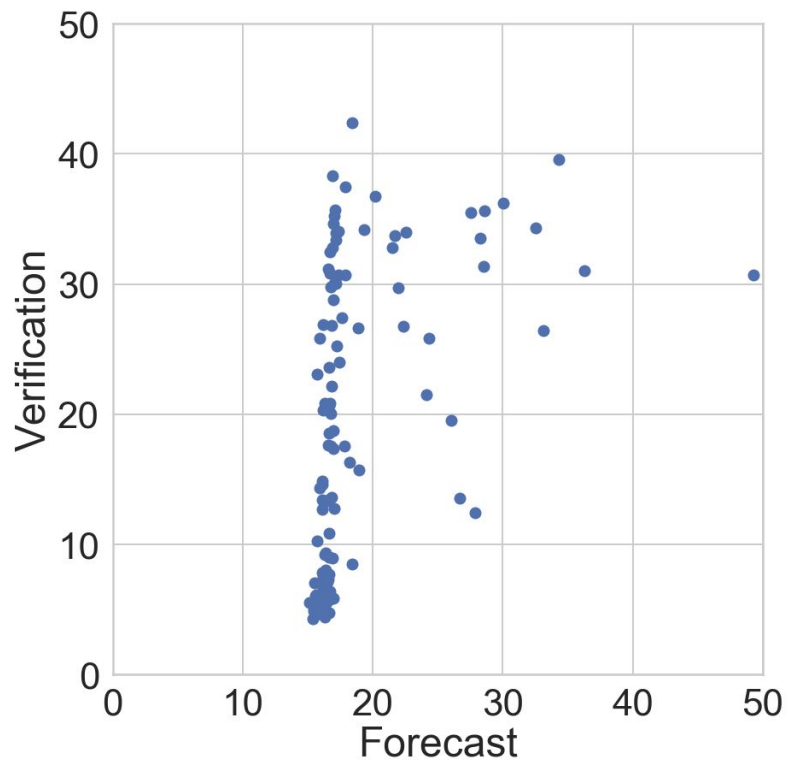
    dfFS = pd.concat([dfFS,FS], ignore_index=False)

#del FS
```

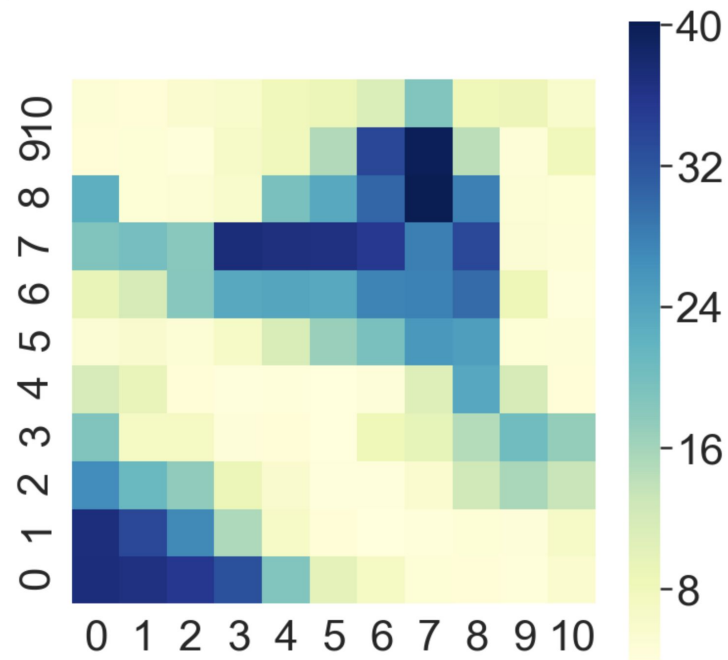
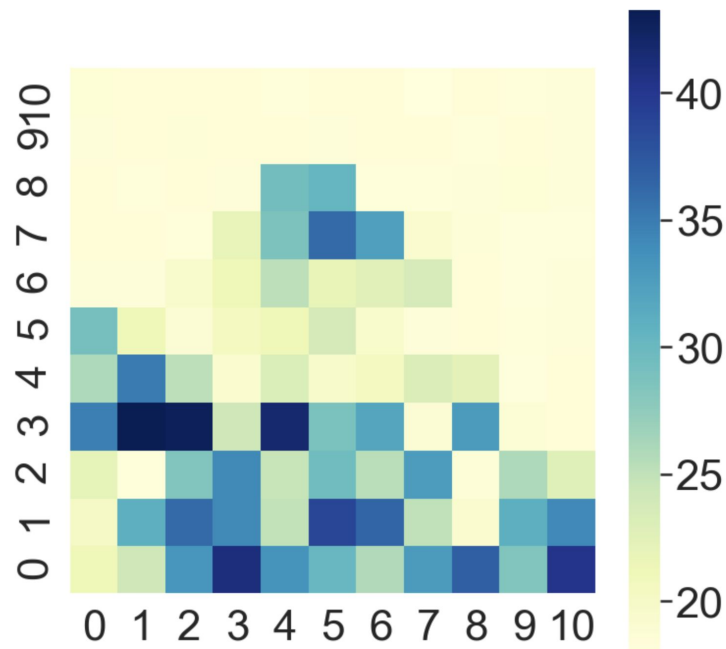
More Parameters  
More Problems

# Cloud Model

# Scatter Plot: bz vs. tt



# Predicted vs Truth Heat Map



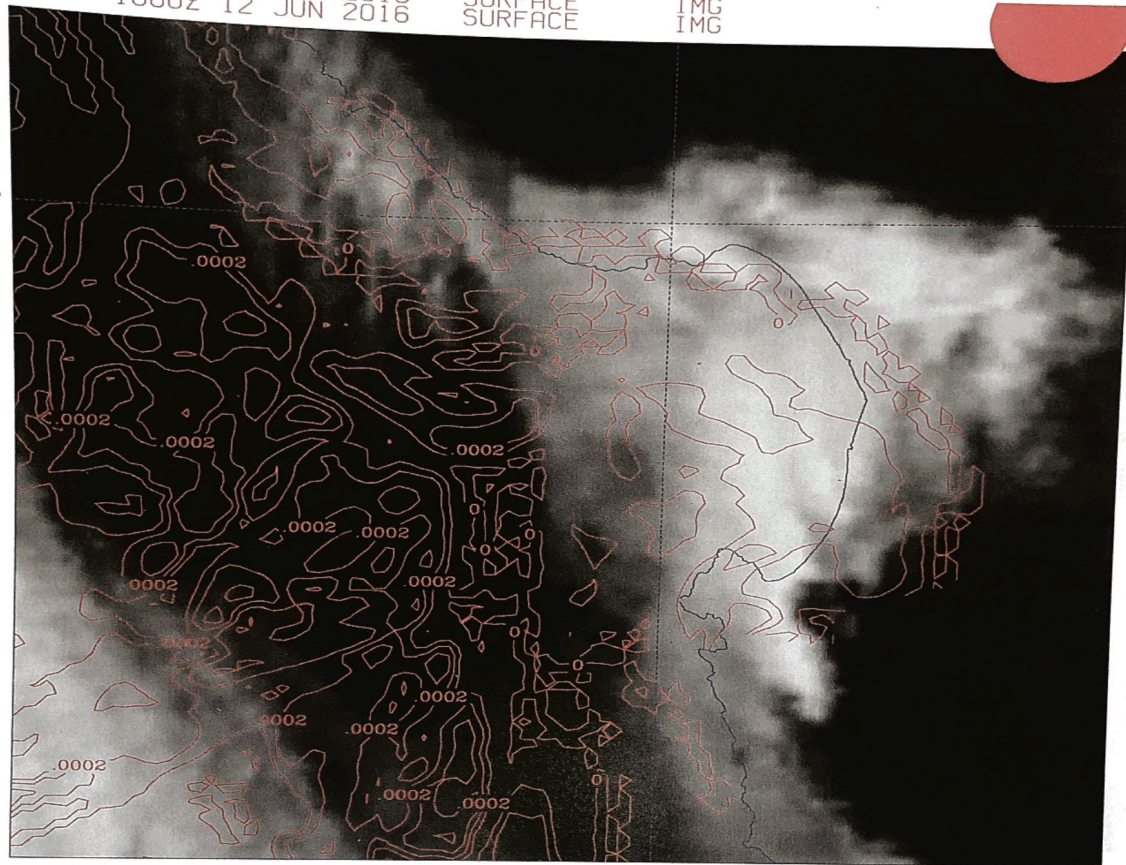


1800Z 12 JUN 2016  
1800Z 12 JUN 2016

SURFACE  
SURFACE

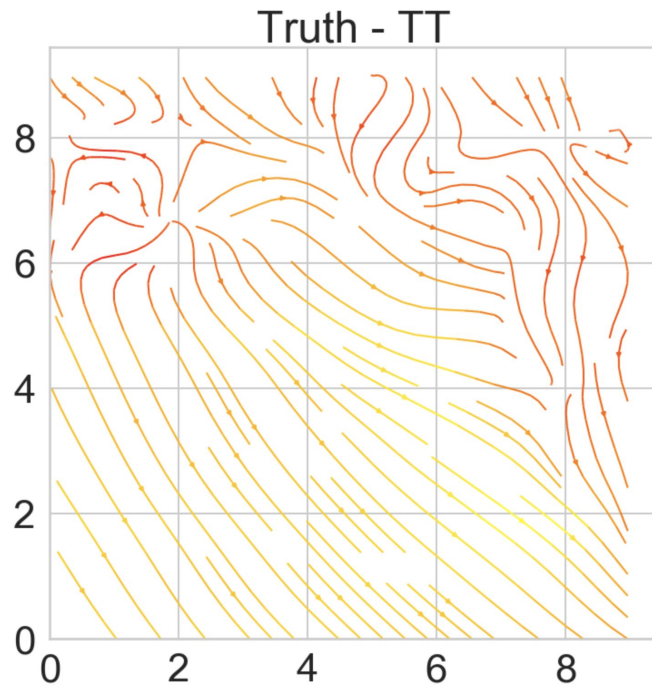
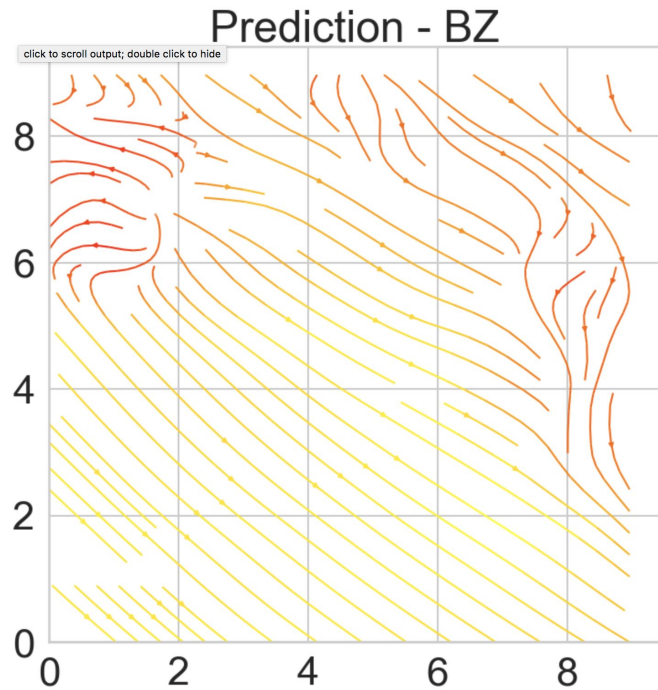
IMG  
IMG

37

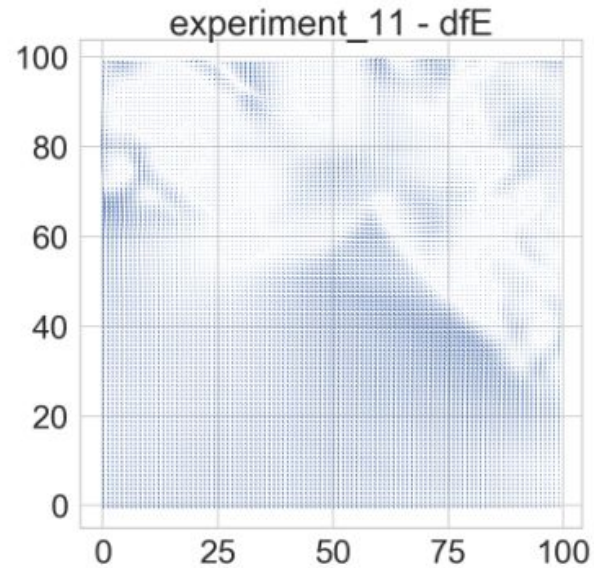
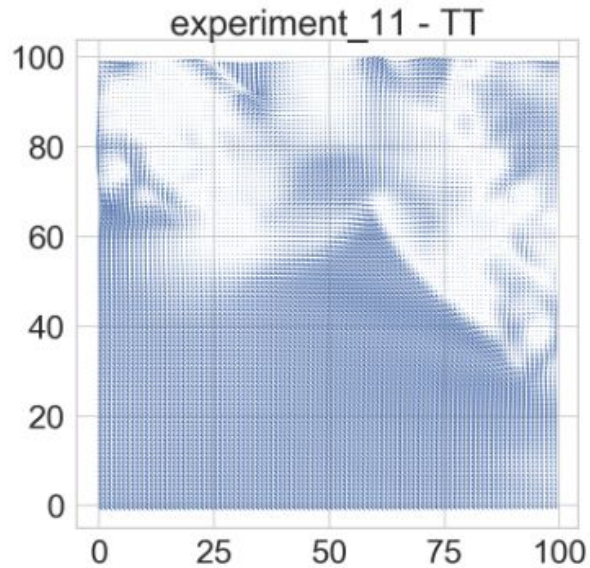
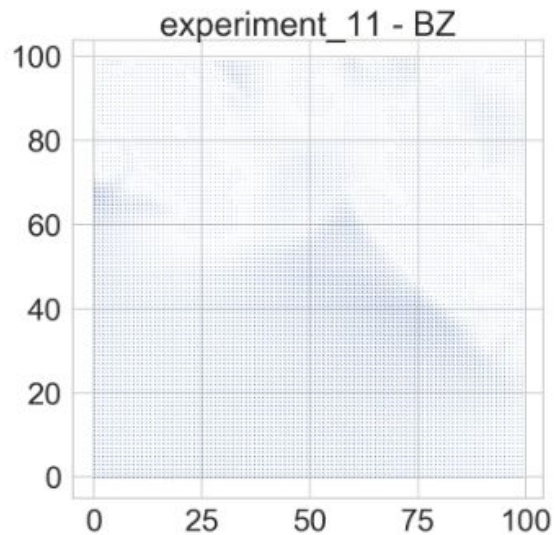


# Wind Model

# Stream Plot



# Vector Plot



# -Model Evaluation (and where it breaks)

# Integrated Cloud Water and Surface Pressure

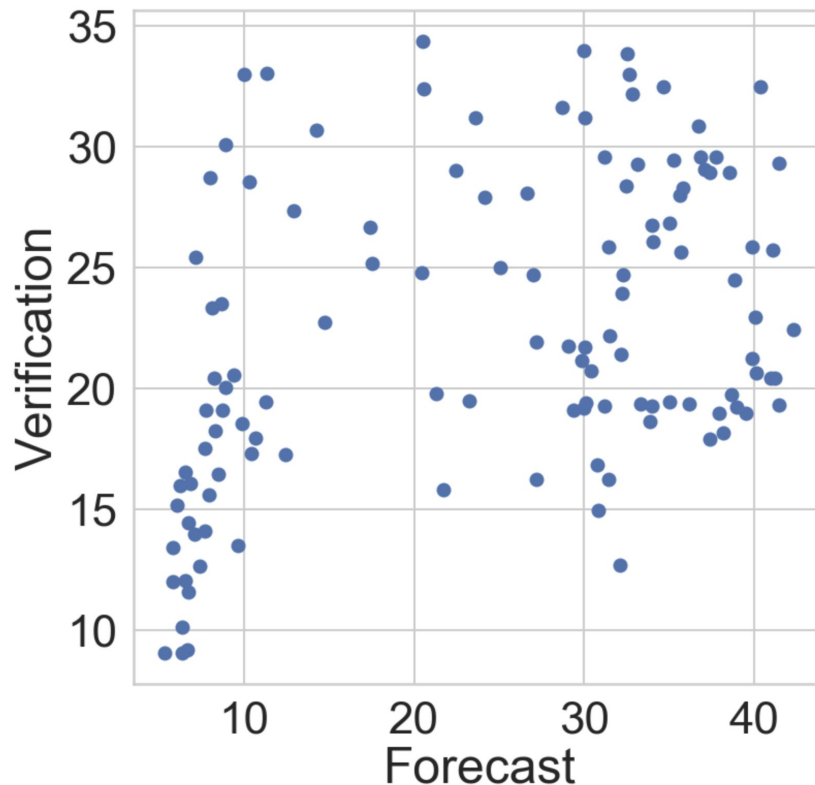
Training data:

160608-11 (4 days)

1500 hrs

Test:

16-06-12-1500



# Integrated Cloud Water and Surface Pressure

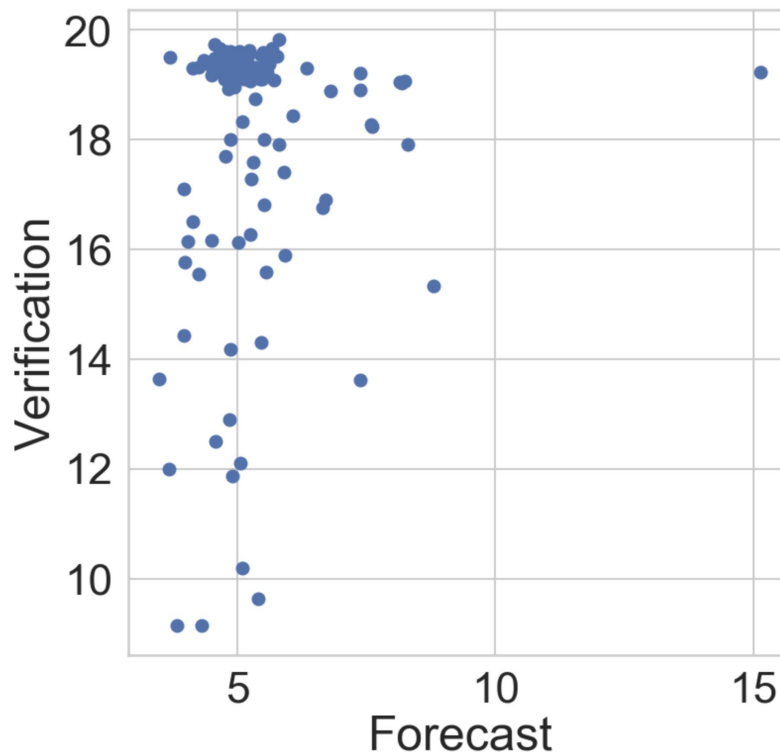
Training data:

160608-11 (4 days)

1500 hrs

Test:

16-06-24-1500



# Integrated Cloud Water and Surface Pressure

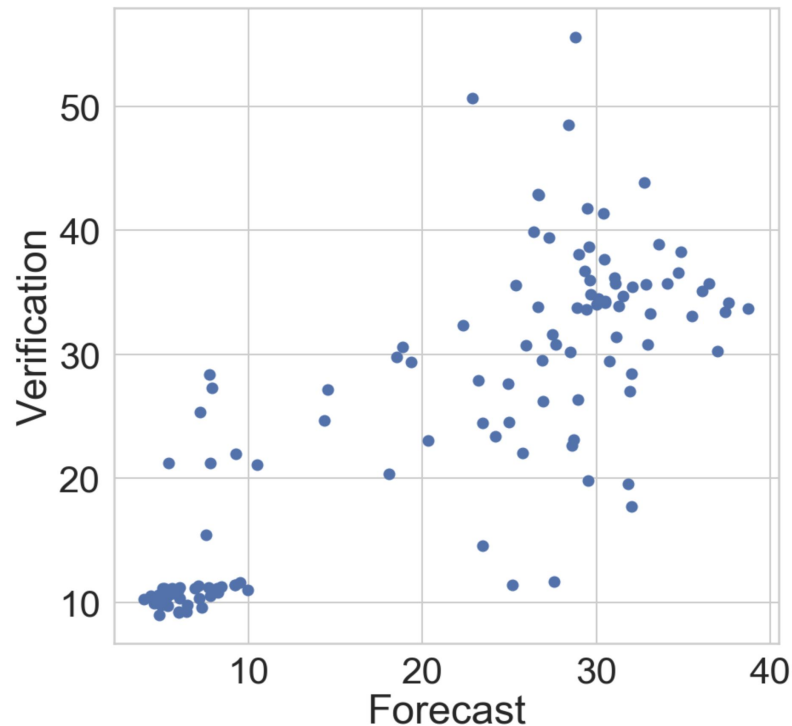
Training data:

160608-160623 (16 days)

1500 hrs

Test:

16-06-24-1500





# 8 June 2016 1500hrs

## OLS Regression Results

```
=====
Dep. Variable:      Observed cloud reflectivity    R-squared:                0.704
Model:              OLS                          Adj. R-squared:           0.704
Method:             Least Squares                 F-statistic:             2375.
Date:               Tue, 18 Jun 2019               Prob (F-statistic):      0.00
Time:              11:44:10                       Log-Likelihood:          -32989.
No. Observations:   10000                         AIC:                    6.600e+04
Df Residuals:       9989                         BIC:                    6.608e+04
Df Model:           10
Covariance Type:    nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	-7628.1427	342.949	-22.243	0.000	-8300.393	-6955.893
Average cloud water	7294.7560	3.08e+04	0.237	0.813	-5.31e+04	6.77e+04
Maximum cloud water	-1041.7062	1809.909	-0.576	0.565	-4589.492	2506.080
Integrated cloud water	-1091.6274	2764.131	-0.395	0.693	-6509.881	4326.627
Maximum Vertical Motion	-0.0161	0.778	-0.021	0.984	-1.541	1.509
Maximum Relative Humidity	-0.1906	0.020	-9.559	0.000	-0.230	-0.151
Average Relative Humidity	0.2363	0.016	15.200	0.000	0.206	0.267
Average Specific Humidity	-5730.5433	211.593	-27.083	0.000	-6145.309	-5315.778
Max Specific Humidity	6051.7509	305.580	19.804	0.000	5452.753	6650.749
Sea Level Pressure	7.4645	0.341	21.884	0.000	6.796	8.133
Surface Pressure	0.0811	0.007	11.165	0.000	0.067	0.095
Omnibus:	159.815	Durbin-Watson:	0.180			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	237.960			
Skew:	-0.178	Prob(JB):	2.13e-52			
Kurtosis:	3.666	Cond. No.	6.74e+08			

```
=====
```

### Warnings:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 6.74e+08. This might indicate that there are strong multicollinearity or other numerical problems.