Indian Institute of Technology Kharagpur

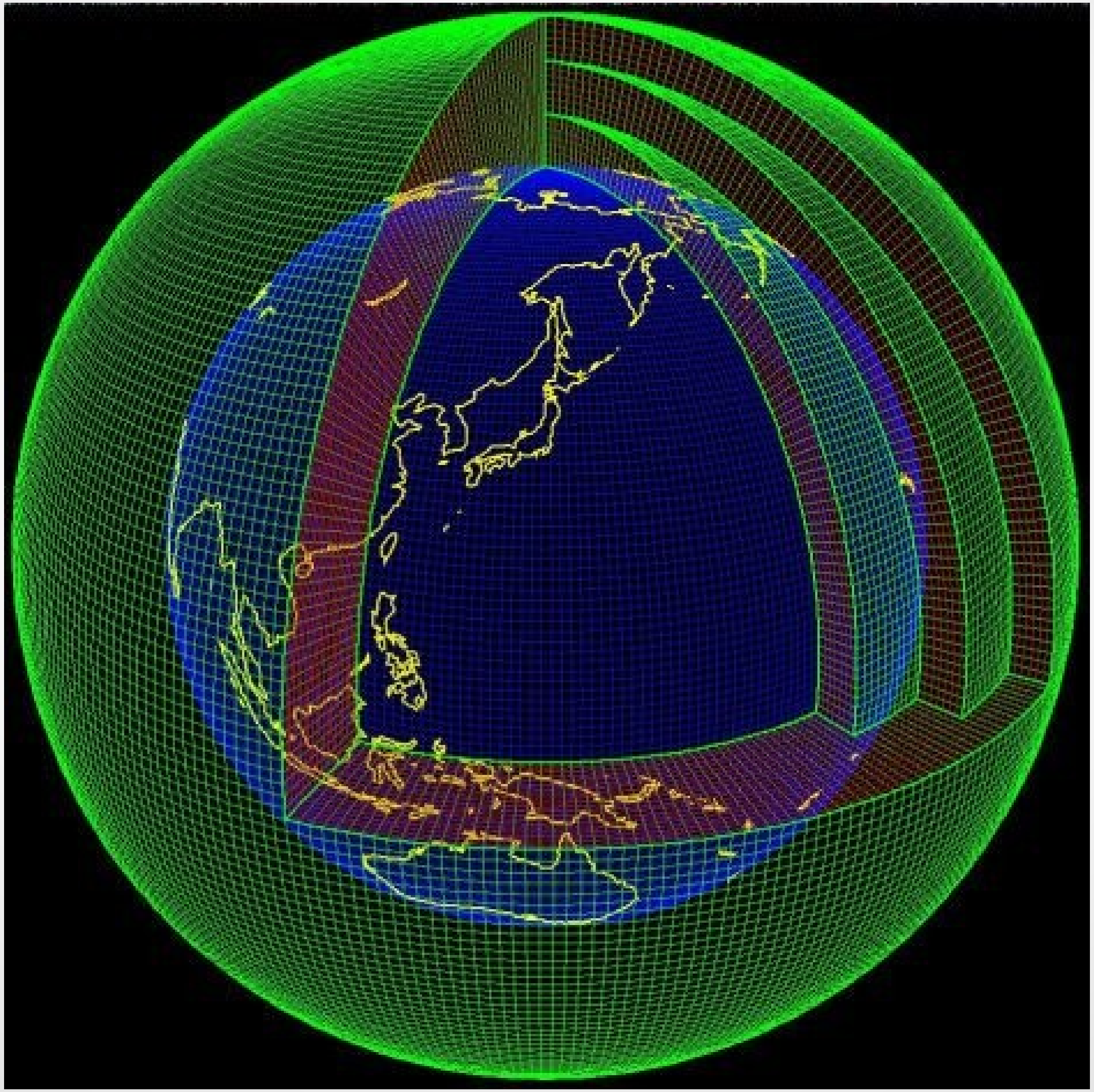
DATA ASSIMILATIN FOR NUMERICAL WEATHER PREDICTION

Master of Technology
2014-2016
Computer Science and Engineering

Shashi Kushwaha
14CS60R35

Under the guidance of:
Prof. Sourangshu Bhattacharya

Numerical Weather Prediction



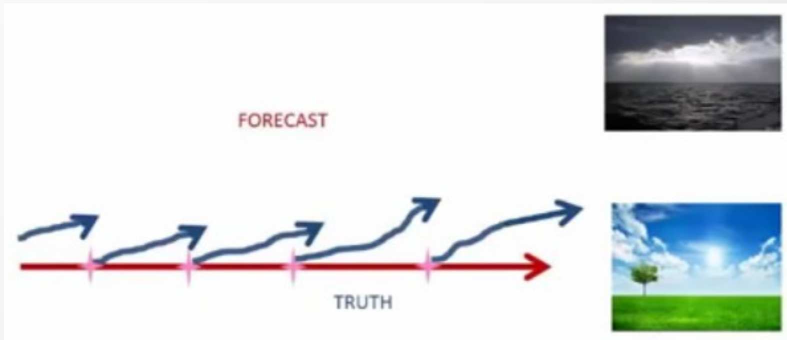
Components of NWP Model

- **Governing Equations**
  - F=ma, conservation of mass, moisture, and thermodynamic eqn. gas law
- **Numerical procedures**
  - Approximations used to estimate each term (especially important for advection terms)
  - Approximations used to integrate model forward in time Boundary conditions
- **Approximations of physical processes**
- **Initial conditions**
  - Observing systems, objective analysis, initialization, and data assimilation

But the forecase is not perfect

Thus, we need Data Assimilation

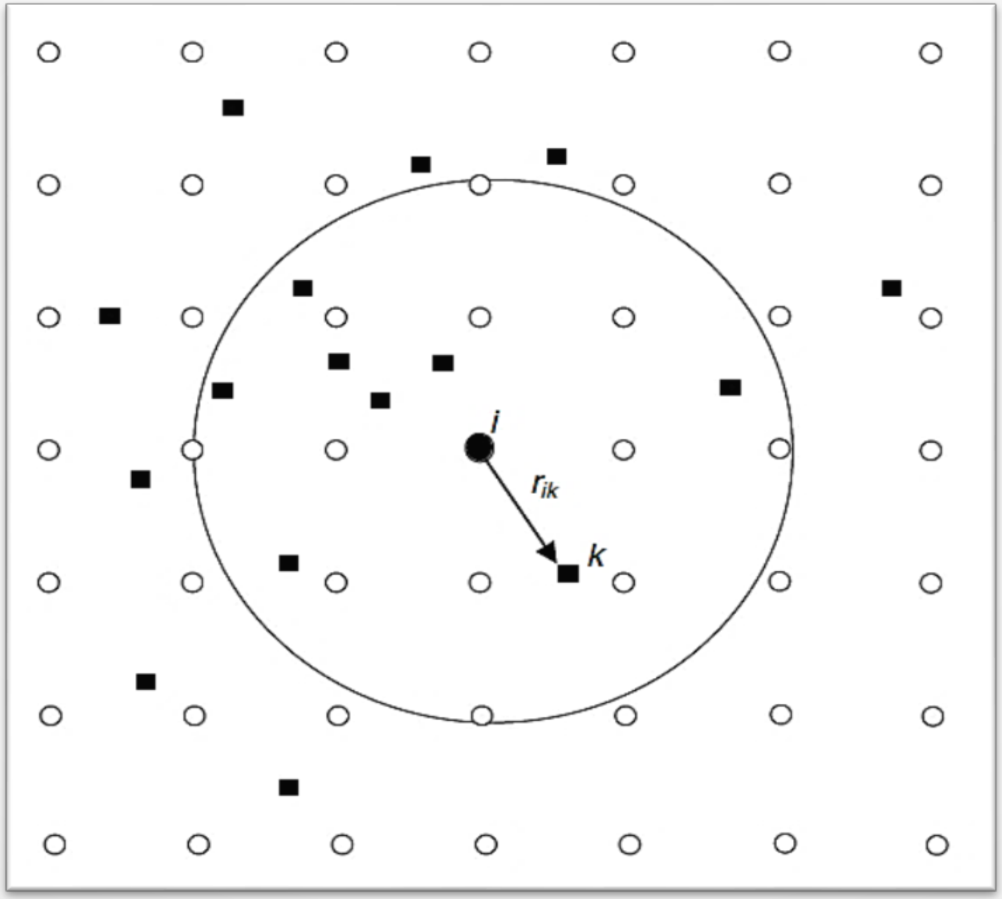
Taking observations at regular interval



Assimilating to reduce error in prediction

Data Assimilation

- Determine as accurately as possible the state of the atmospheric (or oceanic) flow.



Schematic of grid points (circles), irregularly distributed observations (squares), and a radius of influence around a grid point i marked with a black circle. In 4D Data Assimilation, the grid-point analysis is a combination of the forecast at the grid point (first guess)and the observational increments (observation minus first guess) computed at the observational points k .In certain analysis schemes, like SCM, only observations within the radius of influence, indicated by a circle, affect the analysis at the black grid point

Background Values

- First guess or prior information
- The Background is outcome of stabilized models, these models consider some initial conditions, and integrating governing equations on these conditions.
- Objective Analysis (OA) is example of such Model.
- The output of these models is in gridded form.

Observation Data

- Weather Prediction needs a diagnosis based on observation.

Terrestrial based observing

- Weather observations have for many years been made from networks of stations over land and from ships on passage.
- The conventional method of measuring wind, temperature and humidity above ground level is the Radio-sonde.

Space Based systems (Sensors)

- Passive sensors.
- Active sensors.

Data Assimilation Methods

The methods can be divided into classes:

- **Empirical methods**

– Successive Correction Method (SCM)

– Nudging

- **Constant statistical methods**

– Optimal interpolation (OI)

– 3D-Var

We are going to discuss Optimal Interpolation.

Optimal Interpolation

- Assimilate the new observational data in order to advance in time the “background” state that the weather forecasting numerical code has predicted.

- let X_b be the background vector characterizing the current state of the model
- H the observational operator
- Y_o the observational data to be assimilated in the model.
- then the analysis X_a is

$$X_a = X_b + W\left(Y_o - H\left(X_b\right)\right)$$

- With W the weights determined from the estimated statistical error covariance’s of the forecast and the observations.

Optimal Interpolation

The Weight matrix can be calculated as:

$$W = BH^T \left(R + HBH^T\right)^{-1}$$

The analysis error covariance:

$$P_a = \left(I_n - WH\right)B$$

And the background error covariance:

$$B \approx \alpha E\{[x_f(48h) - x_f(24h)][x_f(48h) - x_f(24h)]^T\}$$

Where:

P_a: Analysis error covariance

B: Background error Covariance

H: Interpolation matrix

I_n: Idendity matrix of size n x n

Extended Kalman filtering

- The forecast or background error covariance $P^f(t_i)$ is advanced using the model itself, rather than estimating it as a constant covariance matrix B.

- Kalman filtering consists of a "forecast step" that advances the forecast and the forecast error covariance:

$$X^f(t_i) = M_{i-1}[X^a(t_{i-1})]$$

- Calculation of P^f (forecast error covariance) will be explained later.

Extended Kalman filtering

- The “analysis” or update step:

$$X^a(t_i) = X^f(t_b) + K_i d_i$$

$$P^a(t_i) = (I - K_i H_i) P^f(t_i)$$

where

$$d_i = y_i^o - H[X^f(t_i)]$$

is the observational increment or innovation.

- Kalman gain matrix is calculated as in OI, and this matrix is used in the analysis step:

$$K_i = P^f(t_i) H_i^T [R_i + H_i P^f(t_i) H^T]^{-1}$$

Ensemble Kalman filtering

- K assimilation cycles are carried out at.
- All cycles with same observation data with some random perturbations added.
- Every cycle calculates a X^f
- After completing k cycles, we can calculate:

$$P^f \approx \frac{1}{K-1} \sum_{k=1}^{K} \left(X_k^f - \bar{X}^f \right) \left(X_k^f - \bar{X}^f \right)^T$$

- Where $\bar{X}$ is the mean.

- Also, hybrid between 3D-Var and ensemble Kalman filtering:

$$P^{f(hybrid)} = (1 - \alpha)P + \alpha B_{3D\text{-}Var}$$

Dataset:

- **Background Vector** (X^f) : Background values at grid points, N-Dimensional vector.
  - Downloaded from http://apdrc.soest.hawaii.edu
  - It's a matrix of size [longitude x latitude x depth].
- **Background Error Covariance** (P^f): Background (or forecast) error covariance values.
  - Downloaded from http://apdrc.soest.hawaii.edu
  - It's a matrix of size N x N
- **Observation vector** (y_o): Observational Values at several locations.
  - P-Dimensional Vector
  - Downloaded from: http://www.nodc.gov
- **Observation error covariance** (R) : R is the error covariance of the observations.
  - P x P size matrix, taken as alpha times I. Alpha is tunable.

Algorithm

1) Calculate Interpolation matrix (H):

- H is of size [P x N]
- Computing Interpolation matrix:

If (dist(i, j) < radius of influence(R)) {

$$dist(i,j) / \sum k \in all\ points\ within\ distance\ R : d(i,k)$$

}

Else {

dist(i,j) = 0

}

2) Using matrix H, calculate:

$$d_i = y_i^o - H[X^f(t_i)]$$

Algorithm

3) Calculate matrix K as:

$$K_i = P^f(t_i) H_i^T [R_i + H_i P^f(t_i) H^T]^{-1}$$

4) Using above matrices, calcualte analysis variable x^a as:

$$X^a = X^f + Kd$$

5) And background error covariance (to be used to calculate forecast error covariance in the next forecast cycle as:

$$P^a = (I - KH) P^f$$

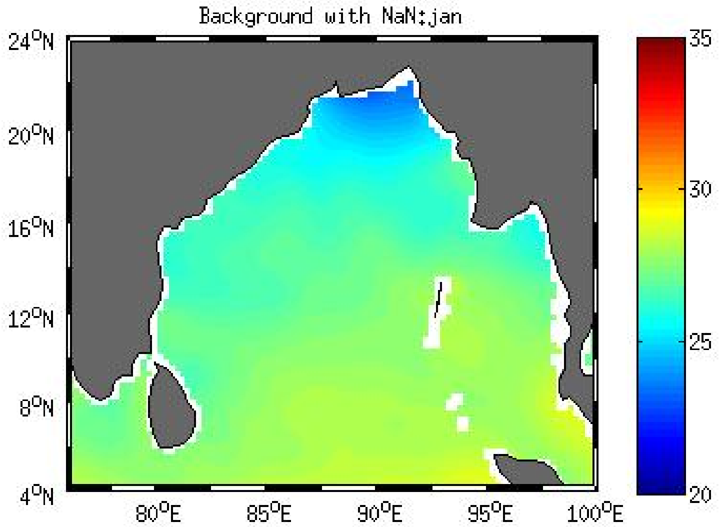
Output

- The variable X^a is the grid point values after assimilation cycle
- The variable P^a is the error covariance among the grid points.

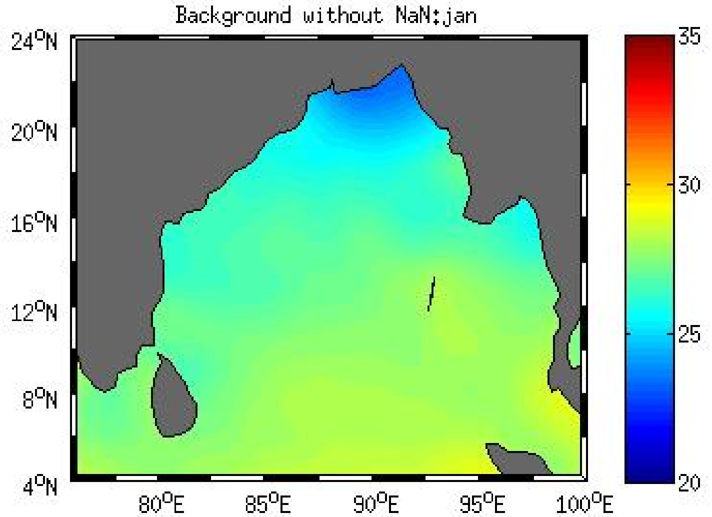
- The input data we downloaded from the websites had few entries that were “Not a Number”.
- We used Barnes objective analysis function to get rid of them.
- We demonstrate below the effect of the analysis on one month’s datats:

Working on Jan 15 data:

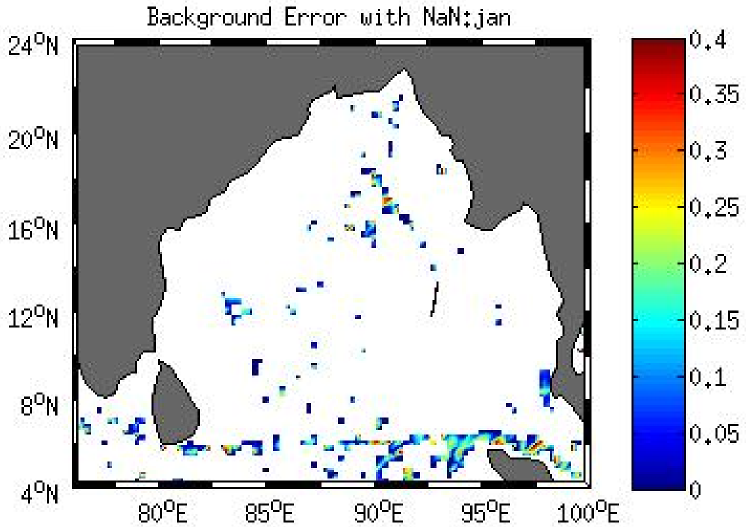
Coordinates are: Lon: 76E – 100E   Lat: 4N – 24N, Dimensions of grid: 96 x 80



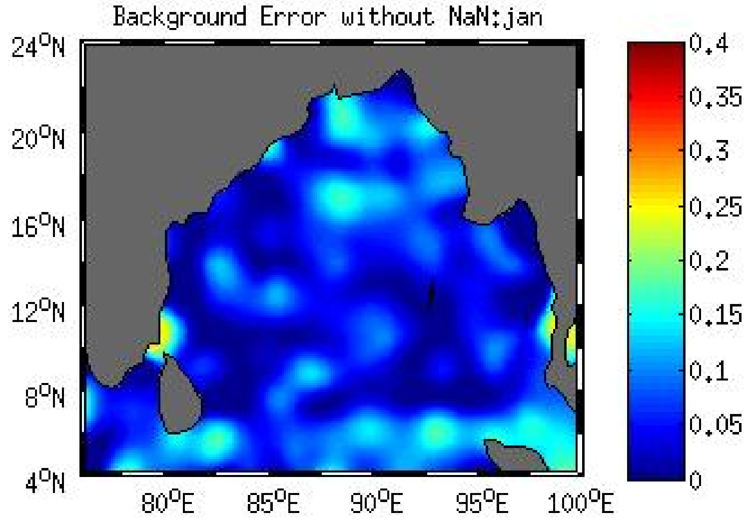
Background with NaN entries



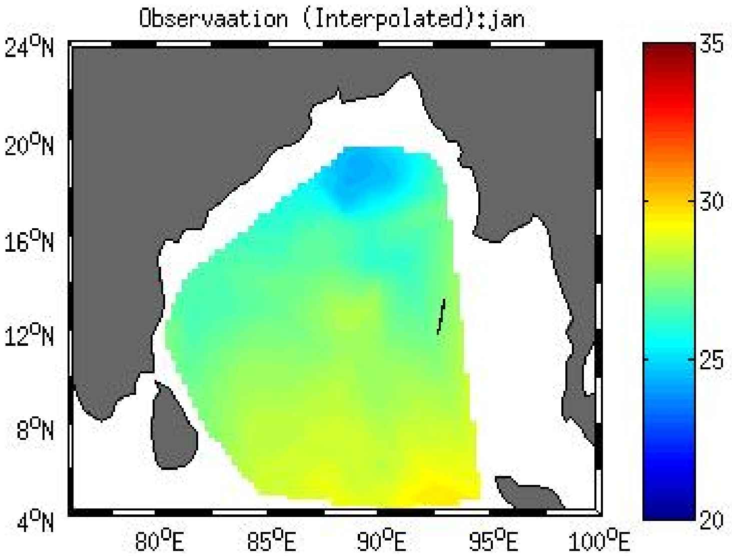
Background with NaN etries removed



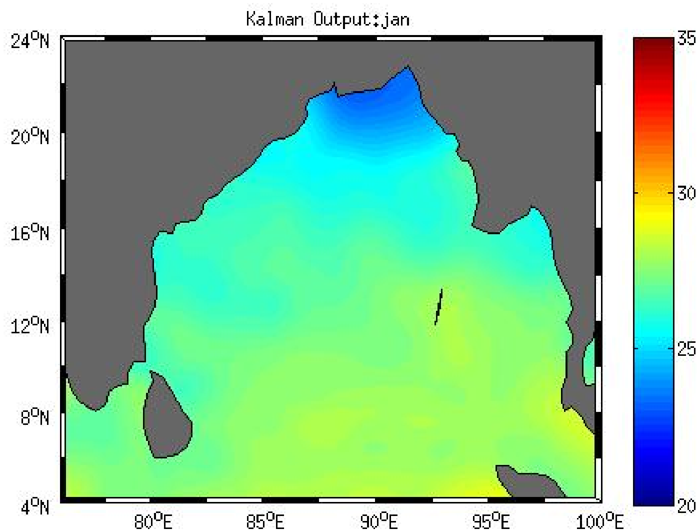
Background error with NaN entries



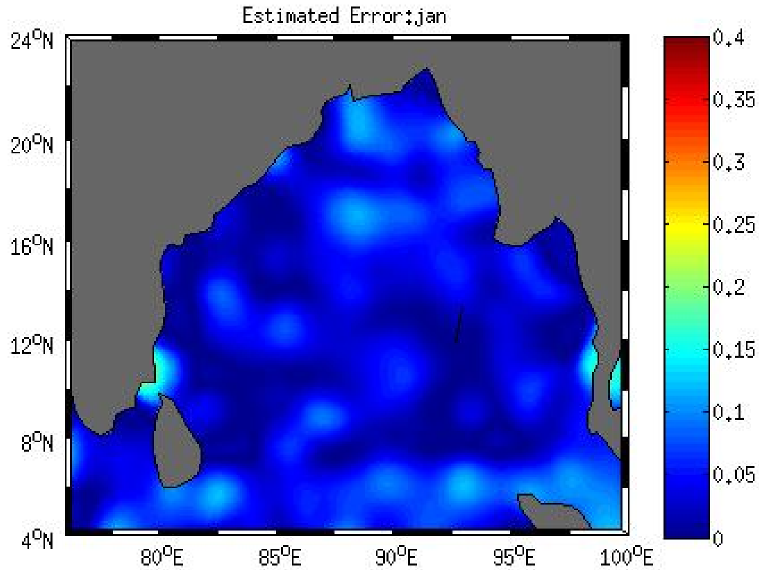
Background error with NaN entries removed



Observation (Interpolated)



Kalman Filtertring output

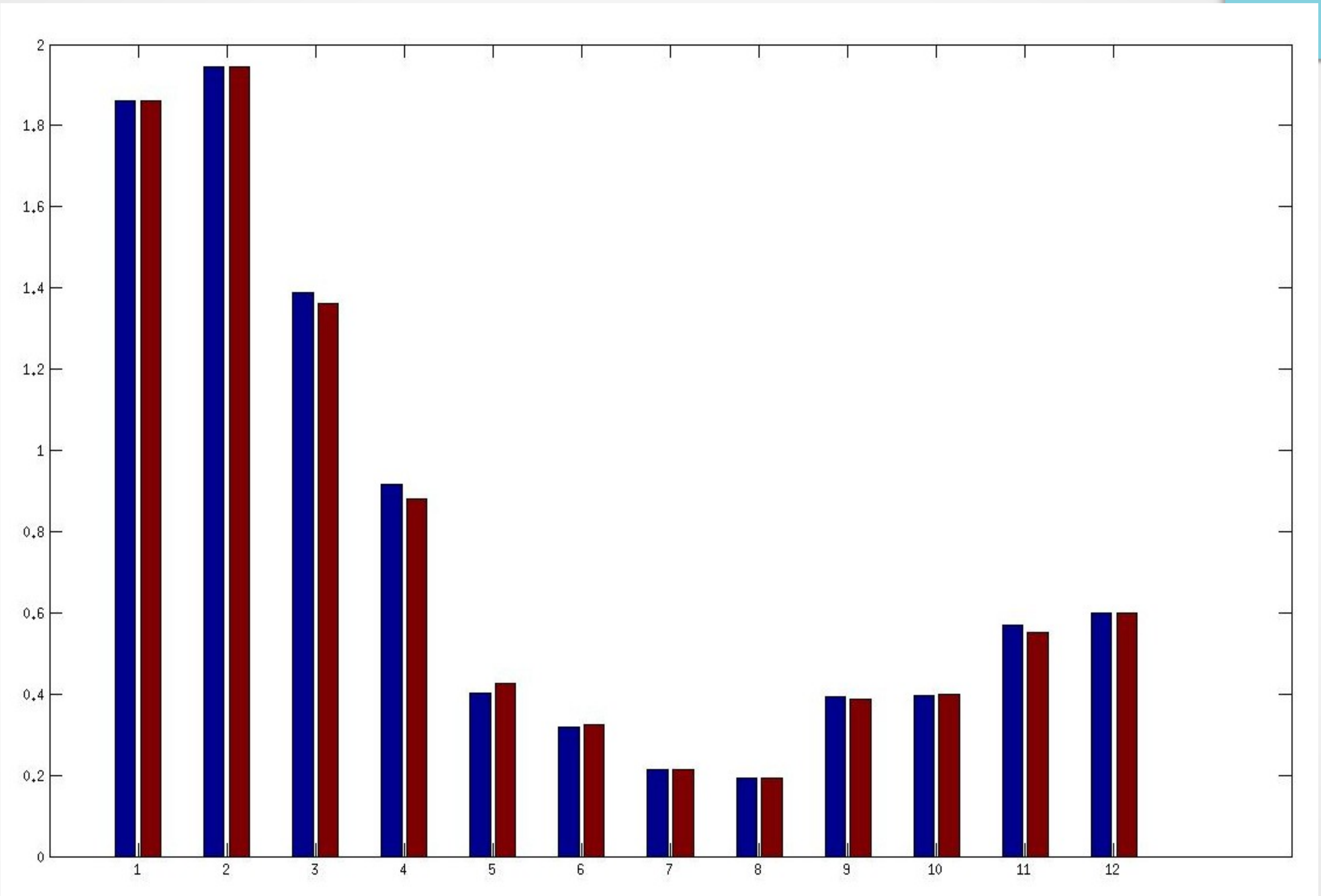


Estimated Error

Comparison with Optimal Interpolation

- To have an estimate of performance we interpolated the Assimilation output to observation space in following ways:
  - **Setting 1:** We used all of the observation space to perform assimilation as well as evaluate performance
    - We then plotted error difference on every observation point for two months taken randomly.
  - **Setting 2:** We used 70% of observation space to perform assimilation and interpolated the output in the rest 30% of space for evaluation.
    - We then plotted error difference on every observation point in the 30% set for same months taken above.
- We performed assimilation and evaluation on monthly data and plotted the data as following:

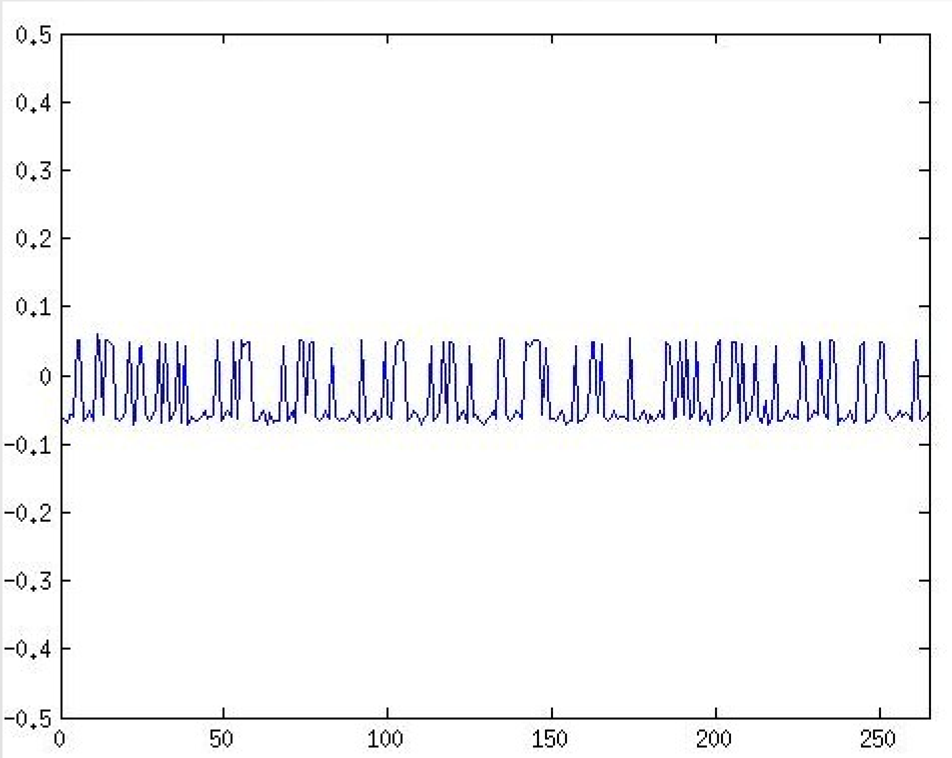
Setting 1



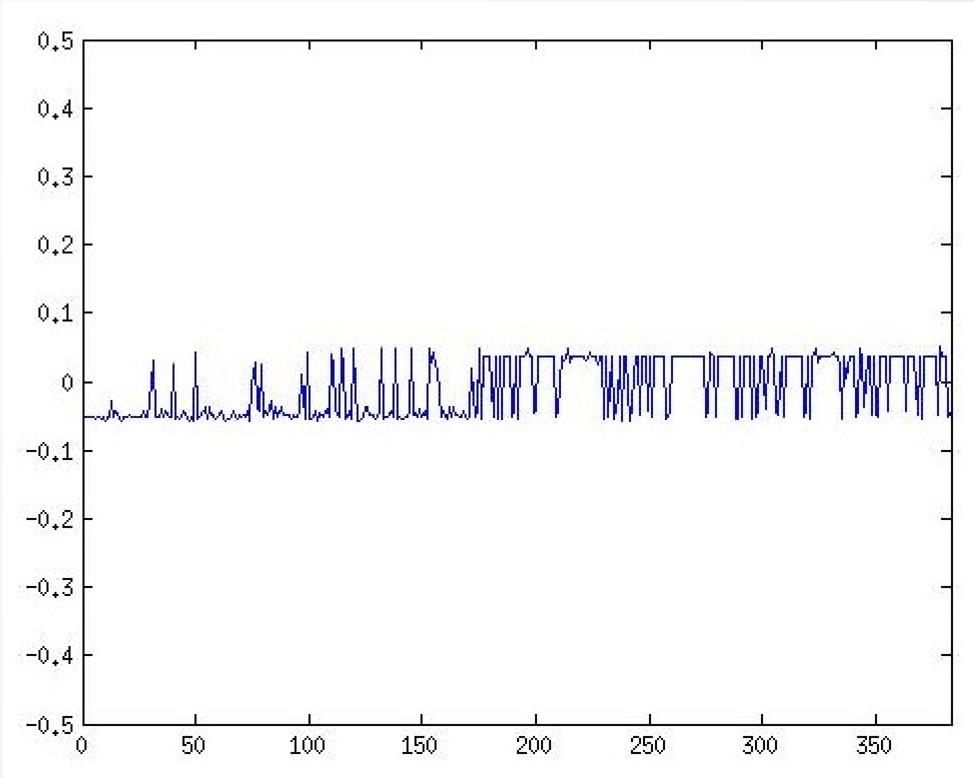
X-axis is months from January to December and Y-axis is the error. The blue bars represent Optimal Interpolation and red bars represent Kalman Filtering.

Error difference at each observation point.

- X-axis is the observation points and Y-axis is the difference between the error given by kalman filtering and that of Optimal Interpolation at the same point.

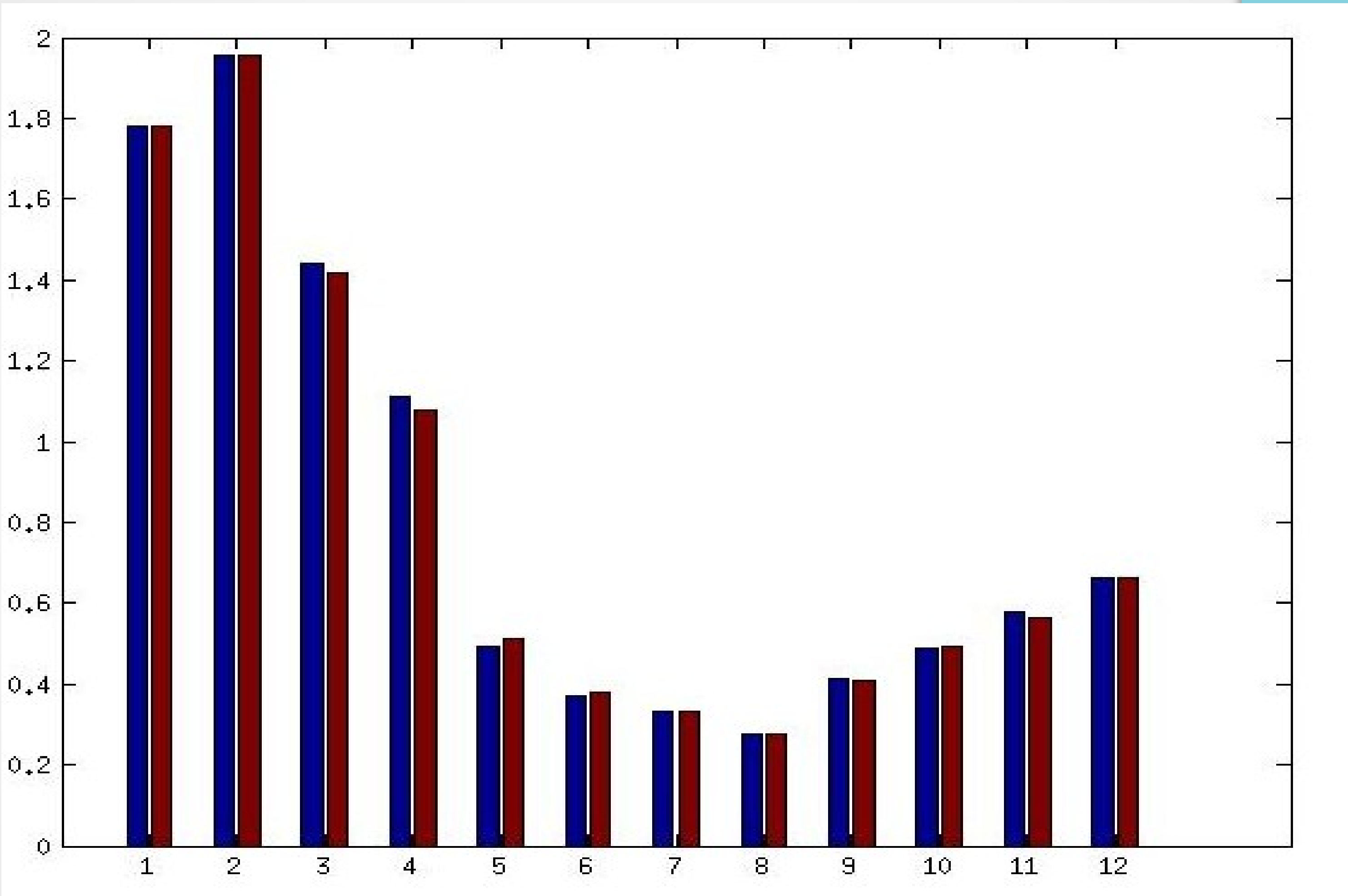


April



November

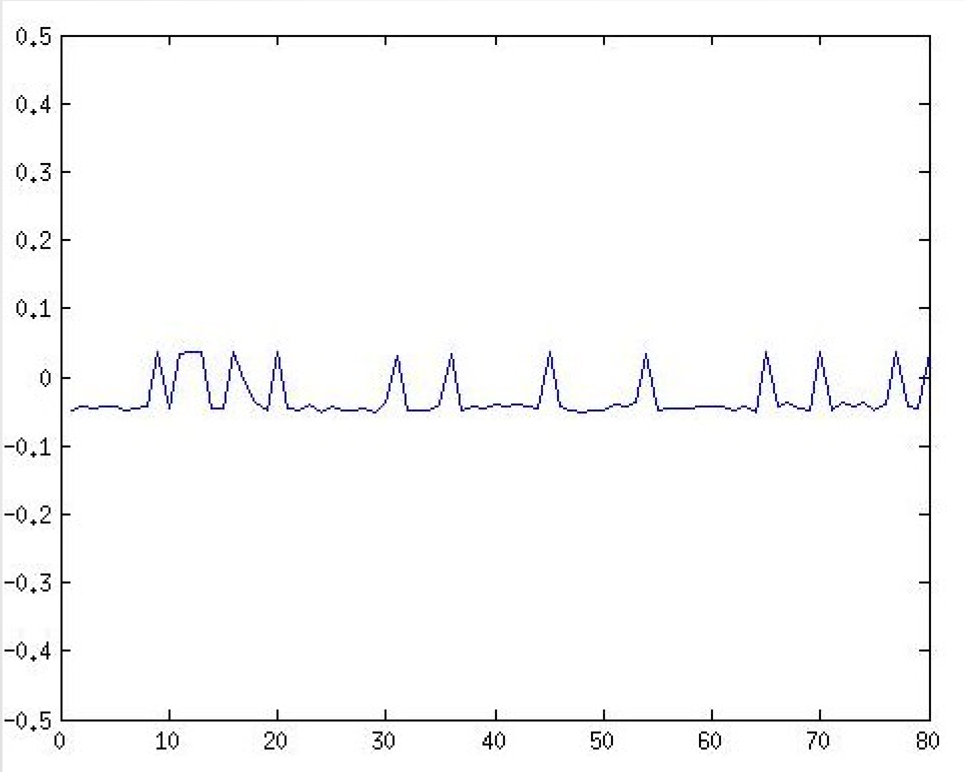
Setting 2



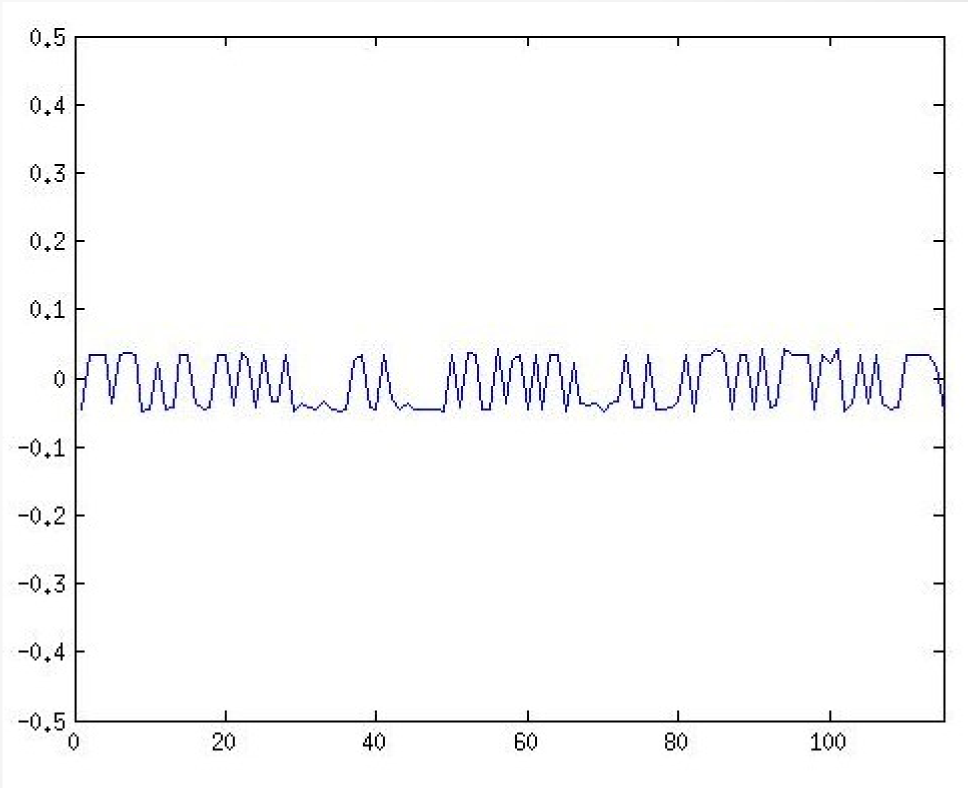
X-axis is months from January to December and Y-axis is the error. The blue bars represent Optimal Interpolation and Red bars represent Kalman Filtering.

Error difference at each observation point.

- X-axis is the observation points and Y-axis is the difference between the error given by kalman filtering and that of Optimal Interpolation at the same point.



April

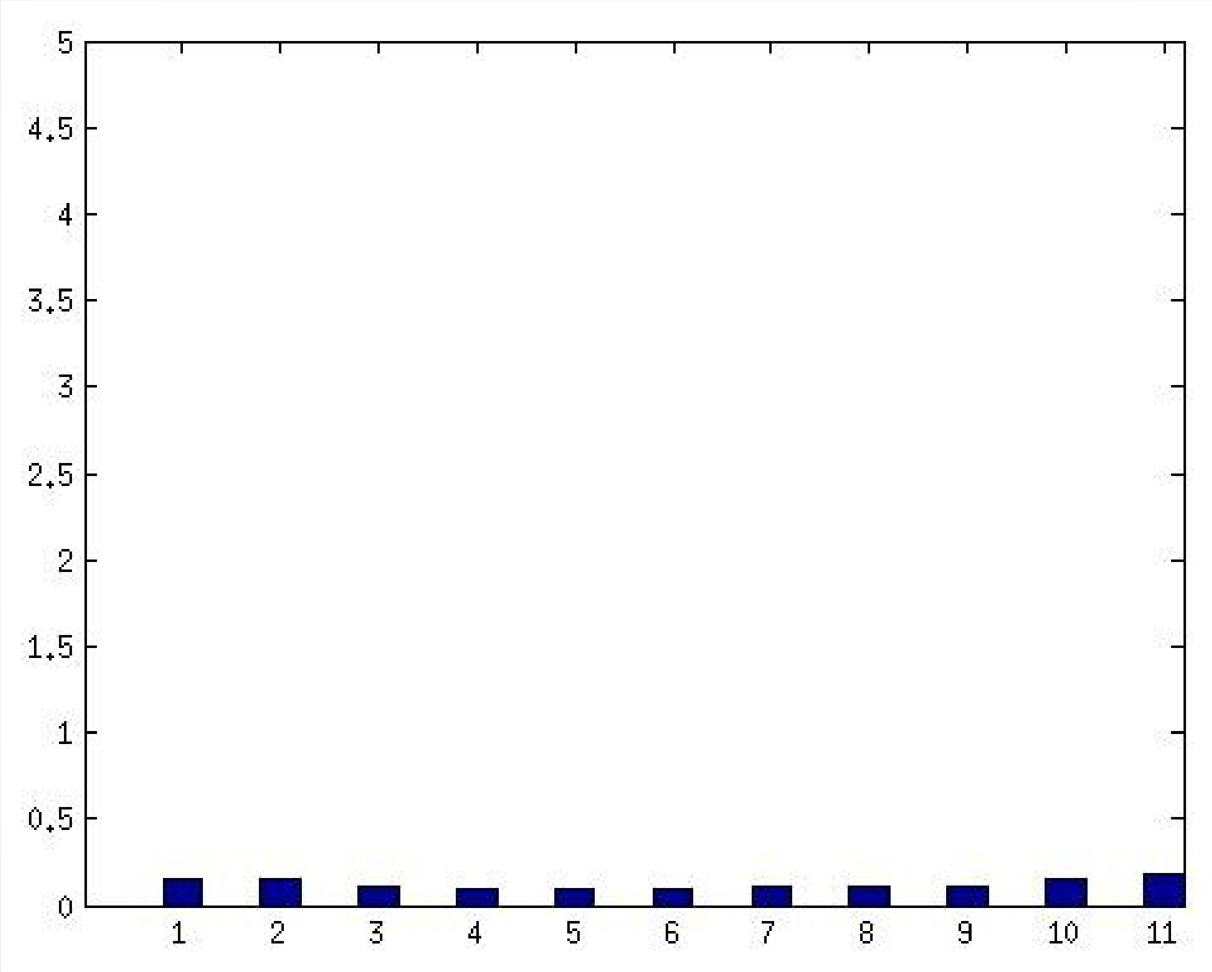


November

Why it did not work?

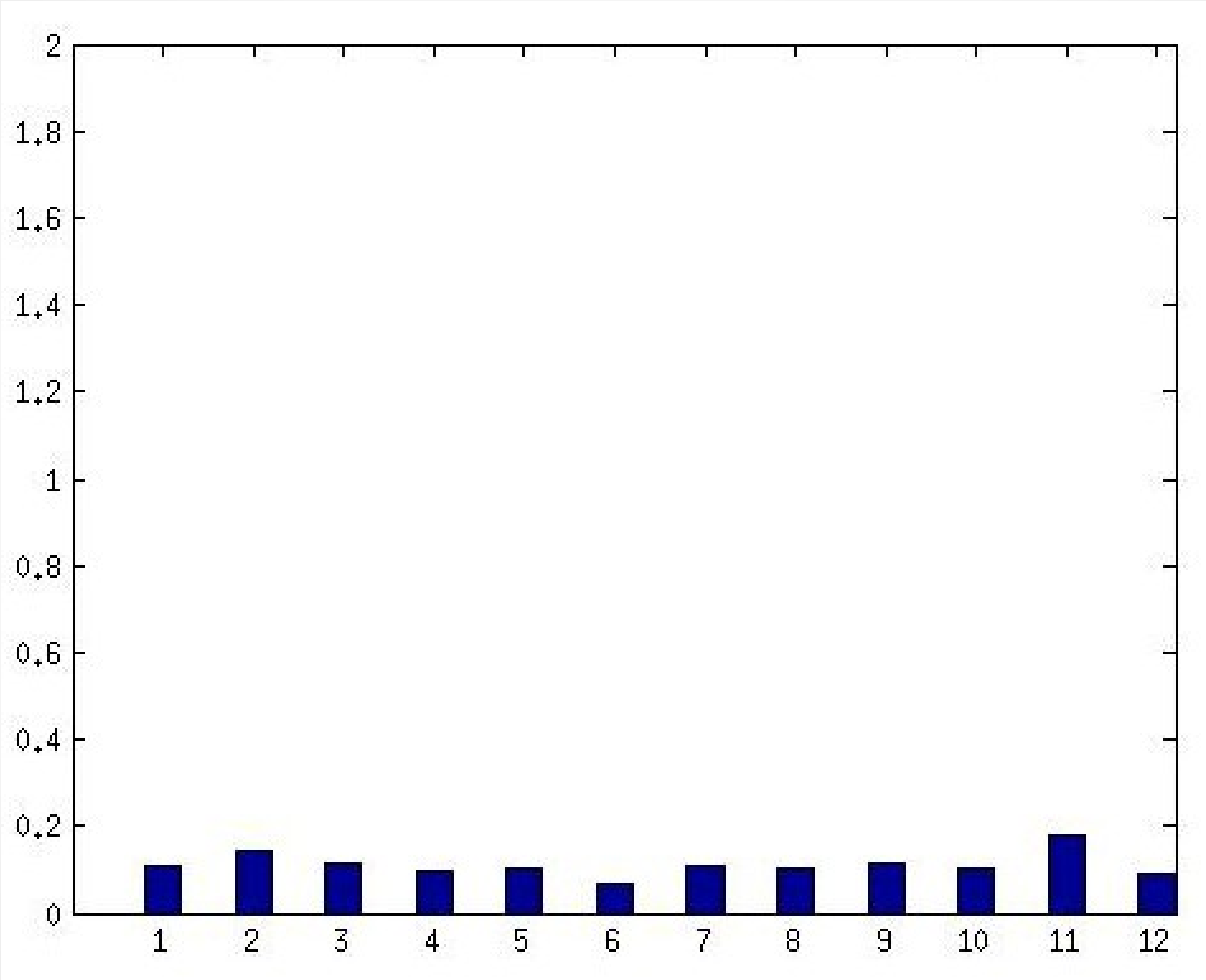
- Kalman Filtering is supposed to work better than Optimal Interpolation.
- It takes into account the changing background error covariance unlike Optimal Interpolation.
- We investigated the reasons the performance isn't better.
- We checked whether the background error we are getting is changing.
- And we checked how much is the background error in input.
- And last, we checked whether the observation differs much from the background.

RMS difference between successive background errors per month



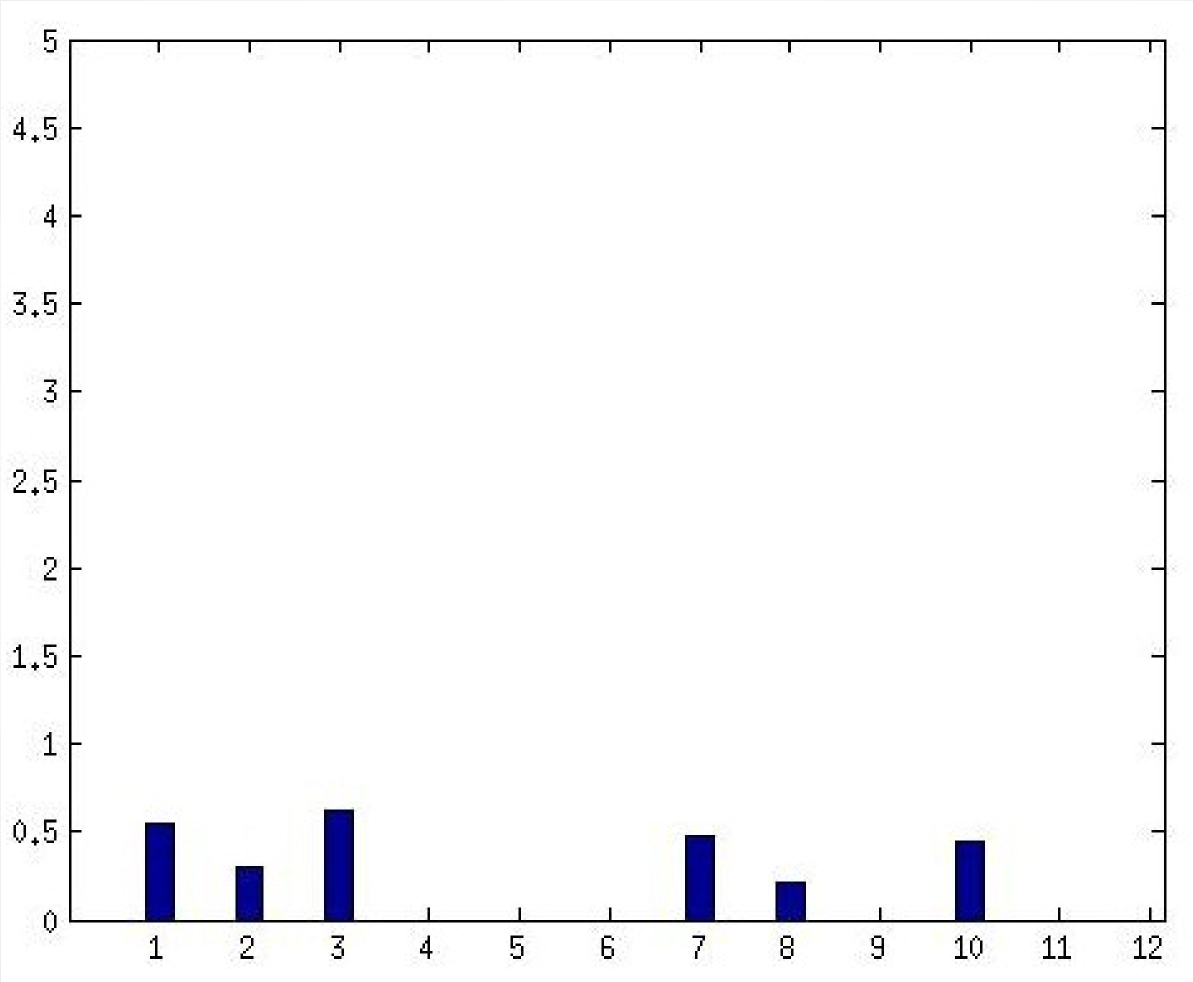
X-axis represents the i and (i+1)th month. Y-axis represents the RMS difference

Average background error per month



X-axis represents each month. Y-axis represents the average background error per month

RMS Difference between background and observation per month



X-axis represents each month. Y-axis represents the RMS difference between background and observation.

Conclusion and future Work

- Kalman Filtering has a significant advantage over Optimal Interpolation because it assumes that the background error covariance is subjected to change.
- In our study Kalman Filtering performed similar to the Optimal Interpolation for following reasong:
  - The background errors were similar over the period of time of study. In other words, we couldn't simulate the changing background error covariance.
  - There was not much background error.
  - The difference between background values and observation wasn't much. Hence, we weren't getting much new information.
  - The data we studied was for one year. Kalman Filtering needs more time to start showing real advantage.
- In future, we can try studying more data and also try simulating the change in background error covariance for better results.

Bibliography

[1] J.G. Charney. 1951. Dynamic forecasting by numerical process, In Compendium of Meteorology , American Meteorological Society, Boston, 1951, pp. 470–482

[2] R. A. Panofsky 1949 OBJECTIVE WEATHER-MAP ANALYSIS Journal of Meteorology 1949 6:6, 386-392

[3] Gilchrist and Cressman (1954) 1954 An Experiment in Objective Analysis QUARTERLY JOURNAL OF GEOPHYSICS

[4] Bergthorsson and Doos 1955 Numerical Weather Map Analysis University of Stockholm

[5] Gandin,L.S. 1963 Objective Analysis of Meteorological Fields Gidro-meteorologicheskoe Izdatelstvo (GIMIZ), Leningrad

Bibliography

[6] Eugenia Kalnay 2003 Atmospheric modeling, data assimilation and predictability

University of Maryland

[7] Data Assimilation Techniques

[http://www.atmos.millersville.edu/~lead/Obs_Data_Assimilation.html]

[8] Asia Pacific Data Research Center, APDRC

[9] World Ocean Database Select and Search

[10] Weather Observations used in NWP [http://weather.mailasail.com/Franks-Weather/Weather-Observations-Nwp]

[11] Numerical Weather Prediction and Data Assimilation, David Schultz, Mohan Ramamurthy, Erik Gregow, John Horel

THANK YOU