Open Journal of Geology, 2018, 8, 1002-1018
http://www.scirp.org/journal/ojg
ISSN Online: 2161-7589
ISSN Print: 2161-7570

AQUA Satellite Data and Imputation of Geopotential Height: A Case Study for Pakistan

Usman Saleem[1,2], Mian Sohail Akram[1], Muhammad Fahad Ullah[3], Faisal Rehman[3*], Muhammad Riaz Khan[4]

[1]Institute of Geology, University of the Punjab, Lahore, Pakistan
[2]Provincial Disaster Management Authority, Project Implementation Unit Punjab, Lahore, Pakistan
[3]Departments of Earth Sciences, University of Sargodha, Sargodha, Pakistan
[4]Pakistan Meteorological Department, Flood Forecasting Division, Lahore, Pakistan
Email: *faisal.rehman@uos.edu.pk

How to cite this paper: Saleem, U., Akram, M.S., Ullah, M.F., Rehman, F. and Khan, M.R. (2018) AQUA Satellite Data and Imputation of Geopotential Height: A Case Study for Pakistan. *Open Journal of Geology*, **8**, 1002-1018.
https://doi.org/10.4236/ojg.2018.810060

Received: August 13, 2018
Accepted: September 23, 2018
Published: September 26, 2018

Abstract

In current study an attempt is carried out by filling missing data of geopotiential height over Pakistan and identifying the optimum method for interpolation. In last thirteen years geopotential height values over were missing over Pakistan. These gaps are tried to be filled by interpolation Techniques. The techniques for interpolations included Bilinear interpolations [BI], Nearest Neighbor [NN], Natural [NI] and Inverse distance weighting [IDW]. These imputations were judged on the basis of performance parameters which include Root Mean Square Error [RMSE], Mean Absolute Error [MAE], Correlation Coefficient [Corr] and Coefficient of Determination [R^2]. The NN and IDW interpolation Imputations were not precise and accurate. The Natural Neighbors and Bilinear interpolations immaculately fitted to the data set. A good correlation was found for Natural Neighbor interpolation imputations and perfectly fit to the surface of geopotential height. The root mean square error [maximum and minimum] values were ranges from ±5.10 to ±2.28 m respectively. However mean absolute error was near to 1. The validation of imputation revealed that NN interpolation produced more accurate results than BI. It can be concluded that Natural Interpolation was the best suited interpolation technique for filling missing data sets from AQUA satellite for geopotential height.

Keywords

AIRX3STML, Missing Data Imputations, Missing Climatic Data, Upper Air Temperature

1. Introduction

Missing data is a big problem encountered at a number of times during environmental research [1] [2] [3] [4]. A lot of causes such as routine maintenances, sampling errors in satellite sensor, failures of satellite sensor during observations, meteorological abnormalities and human errors are responsible for the discontinuity of data set [3] [4]. Geopotential height is the height of a pressure surface in the atmosphere above mean sea level [MSL]. The geopotential height data gathered from AQUA satellite contains incomplete data matrices in 24 standard pressures levels [5]. A research can become inaccurate if missing data sets are used [4] [6]. Geopotential height was the function of air temperature, pressure, winds, and topography of the area, which required a careful method for its imputations. One of the oldest and most suggested methods to fill this missing information was replacing mean values of neighbor samples [1] [2] [3].

Many different interpolation techniques have been developed [2] [6] [7] [8] [9]. The best method depends upon the spatial and temporal variations of geopotential height in the atmosphere. Shen, Reiter [10], applied different interpolations on geopotential height keeping in view its variations in the atmosphere. Knox, Higuchi [11] investigate secular variations, Shabbar, Higuchi [12] did regional analysis, Griesser, Brönnimann [13] reconstructed geopotential height for 850, 700, 500, 300, 200 and 100 hPa. White [14] calculated statistics and climatology for the Northern Hemisphere's geopotential height over 1000 and 500 hPa. Wallace, Zhang [15] investigated intera-decadal variability and teleconnections in the Northern hemisphere's geopotential height over 500 and 700 hPa respectively.

Pakistan is the central country of South Asia bordered with India to East, China in North, South to Arabian Sea and Afghanistan to West (**Figure 1** and **Figure 2**). It is arid to semi-arid country except in the north areas which received annual rainfall of 760 mm to 2000 mm annually. Pakistan has four provinces, of which Baluchistan is the driest and desert area facing 210 mm of rain averagely [16]. 3/4th area of the country is getting no more than 250 mm of rain annually. In summer season relative humidity remains between 20% and 50%. In winter average temperature varies from 4˚C to 20˚C in most areas, while an increasing temperature of 0.6˚C to 1.0˚C is found along the coastal areas [17].

The actual thrust of this research work is to devise a workable methodology for carrying out scientific observations of upper atmosphere meteorology in Pakistan in spite of lacking modern equipment and technological resources in relevant departments. The published literature is not available in Pakistan, however, Saleem and Ahmed [18]; Saleem [19]; Saleem [20] are few initiatives on upper-level atmospheric observations.

2. Material and Methods

2.1. Data Used

In this research, the monthly mean of geopotential height [in meters] for the past 13 years, obtained from Atmospheric Infrared Sounder [AIRS] level 3, was used. AIRS was the instrument on AQUA satellite, which launched in May 2002.

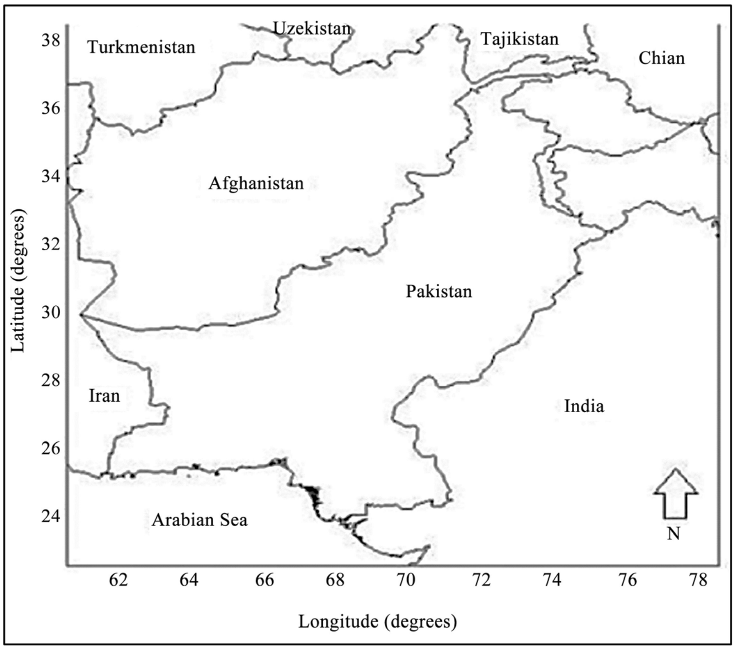


Figure 1. Location map of the Pakistan with it host regions (20).

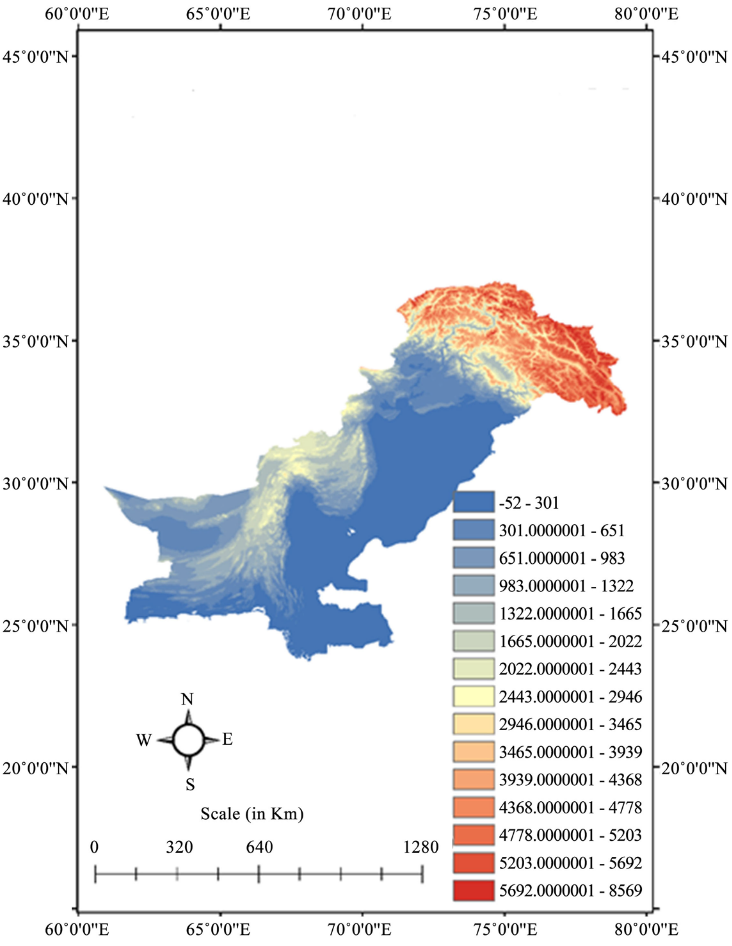


Figure 2. Altitude map of Pakistan showing elevation [in meters] depicted in different color scales (20).

This satellite has very high spectral resolutions: e.g., it captures climate data through nearly 2382 bands in the electromagnetic spectrum and its geopotential height product is very high resolution 0.5˚ × 0.5˚ grid cell. Version 6 of its product contains fewer biases in geopotential height [5]. Besides good quality of climate data, GESDISC[1] provides geopotential height data for the whole global.

2.2. Spatial Interpolations of Missing Geopotential Height

Randomly 30% of the 324 samples were missing data which were then estimated from the 70% known data using different interpolation techniques like IDW, NN, BI and NI [2] [7]. Robeson [21]; Price, McKenney [22]; Perry and Hollis [7]; Yozgatligil, Aslan [3] considered these performance parameters like, Mean Absolute Error [MAE], Root Mean Square Error [RMSE], Coefficient of Determinations [R^2] and Correlation Coefficient [Corr], to find out the best interpolation technique for missing climatic data set.

1) INVERSE DISTANCE WEIGHTING

This imputation resembles to Tobler's first law of geography in which the weight of the known samples will be determined based on the distances from the imputed sample [Robeson, 1994]. More will be the distance of neighbors from a predicted sample less will be their weight in interpolation. Ferrari and Ozaki [9] used Equation (1) which is given below:

$$z_{ij} = \frac{\sum_{a=1}^{n} z_{oa} d_{aj}^{-r}}{\sum_{a=1}^{n} d_{aj}^{-r}} \tag{1}$$

where d_{aj}^{-r} is the weighting factor of distance between the a^{th} original neighbor sample z_{oi}, z_{ij} is j^{th} the point to be estimated, n is the total number of the sample used, and r weighting factor. Langella [23], formula for IDW was used in the missing data imputations.

2) NEAREST NEIGHBORS INTERPOLATION [NN]

Missing values were directly imputed with a most suitable neighbor around the missing sample [24] [25] in this interpolation technique.

3) BILINEAR INTERPOLATION [BI]

Junninen *et al.* [2004] used Equations (2) and (3) for Bilinear Interpolations

$$z_i = z_{i1} + m\left(z_o + z_{o1}\right) \tag{2}$$

$$m = \frac{z_{i2} - z_{i1}}{z_{o2} - z_{o1}} \tag{3}$$

$$z_{o1} < z_o < z_{o2} \text{ and } z_{i1} < z_i < z_{i2}$$

It was a linear equation with $\left(z_{o1}, z_{I1}\right)$ and $\left(z_{o2}, z_{i2}\right)$ sample values, m being a gradient of this line.

4) NATURAL NEIGHBORS INTERPOLATION [NI]

[1]GESDISC stands for Goddard Earth Sciences Data Information Services Center.

This spatial interpolation gives the nearest neighbor value of the sample to the missing geopotential height. D. and Boissonnat and Cazals [25] explain the selection of such natural neighbors for randomly missing data being on Delaunay triangulation.

2.3. Performance Indicators for Interpolations

These following performance parameters have been frequently used by Robeson [21]; Price, McKenney [22]; Junninen, Niska [2]; Perry and Hollis [7]; Stahl, Moore [24]; Norazian [4]; Ferrari and Ozaki [9]; Saleem and Ahmed [18] for imputation of missing climate data set.

1) ROOT MEAN SQUARE ERROR [RMSE]

Root Mean Square was calculated by dividing the sum of the square of the difference between imputed geopotential heights and actual value with the total number of samples, and then finally taking the square root of this term [4]. Smaller values indicate a perfect estimation of missing data set. Equation (4) was its mathematical formula used in this research.

$$RMSE = \left(\frac{1}{n} \sum_{a=1}^{n} \left[z_{oa} - z_{ia} \right]^2 \right)^{\frac{1}{2}} \tag{4}$$

This parameter calculates the total difference [±] between original and interpolated geopotential height.

2) MEAN ABSOLUTE ERROR [MAE]

This provides more information about the residual error as compared with RMSE. Junninen, Niska [2] and Norazian [4] provided Equation (5) for MAE.

$$MAE = \frac{1}{n} \sum_{a=1}^{n} \left| z_{ai} - z_{ia} \right| \tag{5}$$

MAE value range from 0 to ∞. Its value close to 1 indicates more accurate and perfect imputation of missing data set.

3) CORRELATION COEFFICIENT [Corr]

Its value of +1 indicates very strong correlation and near to 0 signifies a bad correlation between actual and predicted geopotential height. Equation (6) was used for the correlation coefficient in this research.

$$corr = \frac{\operatorname{cov}\left(z_i, z_o\right)}{\partial_{zi} \partial_{zO}} \tag{6}$$

In Equation (6) nominator represents covariance while denominator represents the product of their standard deviations in the data set.

4) COEFFICIENT OF DETERMINATION [R^2]

This parameter provides a degree of correlation between the actual and predicted sample geopotential height [1] which varies between 0 and 1. Noor, Abdullah [4], suggested values closer to 1 indicate a perfect fit for the data set. Rahman and Islam, [2011] used the following formula for R^2.

$$R^2 = \left[\frac{1}{n} \frac{\sum_{a=1}^{n} \left(z_{ia} - A_i \right) \left(z_{oa} - A_o \right)}{\partial_{zi} \partial_{zo}} \right] \tag{7}$$

In Equation (7), A_i was the average of predicted samples and A_o is the average of sample values before prediction.

3. Results

These were the results of the performance parameter for each interpolation technique.

3.1. Performance Parameters from IDW

On all pressure level IDW showed very biased results. IDW produced highest RMSE ± 14.45 m over 1 hPa while lowest value of this error was ±3.66 m at 925 hPa. Actual and predicted values indicating low quality of interpolation for missing values of geopotential height with IDW as correlation coefficient was very low (Table 1).

3.2. Performance Parameters from Nearest Neighbor Interpolation

RMSE value remains between ±4.925 and ±11.369 m with Nearest Neighbor Interpolations. Such a large RMSE, poor correlation, and poor fit to the surface indicated bad refilling of data with this interpolation technique (Table 2).

3.3. Performance Parameters from Bilinear Interpolation

Bilinear Interpolation appeared to be relatively better as compared to the above mentioned two interpolations. RMSE was ±2.461 to ±5.241 m in refilling of gaps in data up to 1000 hPa. MAE remains less than 1 and strong correlation (0.98) was found in the imputation of geopotential height. Coefficient of Determination was close to 0.98 for imputation over 1, 1.5, 2, 3, 5, 7, 10, 15, 70, 100, 150, 200, 250, 300 hPa (Table 3).

3.4. Performance Parameters from Natural Neighbor Interpolation

Reasonable low RMSE come in refilling of geopotential height over 2, 3, 5, 7, 30, 50, 70, 200, 250, 400, 500, 600 hPa. Largest RMSE was ±5.10 m at 10 hPa and lowest RMSE ±2.2 m for refilling of gaps in data at 850, 925, 1000 hPa. A good correlation coefficient [near to 0.99] was come in the refilling of geopotential height. R^2 was near to 1 concluding a good line of fit between actual and predicted data set (Table 4).

4. Discussions

Refilling of geopotential height over 24 pressure levels was good with Bilinear and Natural Neighbor Imputations (Tables 1-4). In order to nominate optimum interpolation from both of them, scatter plots of original and estimated geopotential heights were investigated. Poor data refilling was come in February and March (Figure 3(a)).

Table 1. Results indicating poor performance parameters with Inverse Distance Weighting Interpolation.

Pressure Level	RMSE	AME	Correlation	R^2
1 hPa	5.241294	1.673678	0.994263	0.982482
1.5 hPa	4.727445	1.538098	0.992346	0.978878
2 hPa	5.144201	1.546157	0.992445	0.978924
3 hPa	4.216421	1.403774	0.992625	0.979272
5 hPa	4.258006	1.303426	0.988246	0.970826
7 hPa	3.712901	1.191135	0.991712	0.977557
10 hPa	4.406922	1.25975	0.988496	0.971253
15 hPa	4.131028	1.198271	0.989455	0.973121
20 hPa	4.807294	1.359347	0.985551	0.965473
30 hPa	4.615745	1.272488	0.975812	0.948054
50 hPa	4.481513	1.24537	0.978242	0.952734
70 hPa	4.001382	1.207328	0.989019	0.97236
100 hPa	3.815742	1.206655	0.996643	0.987169
150 hPa	3.96526	1.236372	0.99677	0.987438
200 hPa	4.67426	1.333385	0.998104	0.990057
250 hPa	4.395547	1.319586	0.996534	0.986976
300 hPa	4.593847	1.33011	0.995398	0.98474
400 hPa	4.281792	1.23829	0.991174	0.976463
500 hPa	4.30804	1.212168	0.983423	0.961512
600 hPa	4.509835	1.3308	0.980018	0.954683
700 hPa	4.140531	1.158596	0.97152	0.938137
850 hPa	2.461281	0.770069	0.981623	0.957836
925 hPa	2.722891	0.795279	0.988701	0.971507
1000 hPa	2.52145	0.669249	0.986646	0.968239

Table 2. Results indicating poor performance indicators from Nearest Neighbor Interpolation.

Pressure Level	RMSE	AME	Correlation	R^2
1 hPa	11.36953	4.694365	0.984868	0.964135
1.5 hPa	10.47272	4.227424	0.981804	0.95849
2 hPa	10.04097	4.013124	0.978183	0.952061
3 hPa	9.314465	3.667933	0.97511	0.945647

Continued

5 hPa	8.37878	3.189218	0.974254	0.943877
7 hPa	8.25621	3.144038	0.975983	0.94716
10 hPa	8.307455	3.156482	0.972696	0.940833
15 hPa	7.815421	2.909444	0.976439	0.947893
20 hPa	7.607512	2.753404	0.969492	0.935232
30 hPa	7.006525	2.610512	0.969131	0.93451
50 hPa	6.488704	2.500665	0.9731	0.942042
70 hPa	7.553283	2.904254	0.97597	0.947148
100 hPa	9.827749	4.04727	0.983055	0.96073
150 hPa	12.15678	5.267926	0.988911	0.972037
200 hPa	12.76332	5.485841	0.988962	0.972149
250 hPa	11.93628	5.083785	0.986951	0.968318
300 hPa	10.19128	4.385436	0.988056	0.970371
400 hPa	8.383855	3.392052	0.977312	0.949918
500 hPa	7.497696	2.884847	0.967422	0.930971
600 hPa	6.502538	2.370197	0.958783	0.91434
700 hPa	5.504249	1.981821	0.950236	0.897891
850 hPa	4.930112	1.53327	0.940682	0.879706
925 hPa	4.925109	1.46762	0.96189	0.919613
1000 hPa	5.786711	1.375848	0.981378	0.95728

Table 3. Results indicating good performance parameters for refilling of gaps in data with Bilinear Interpolation.

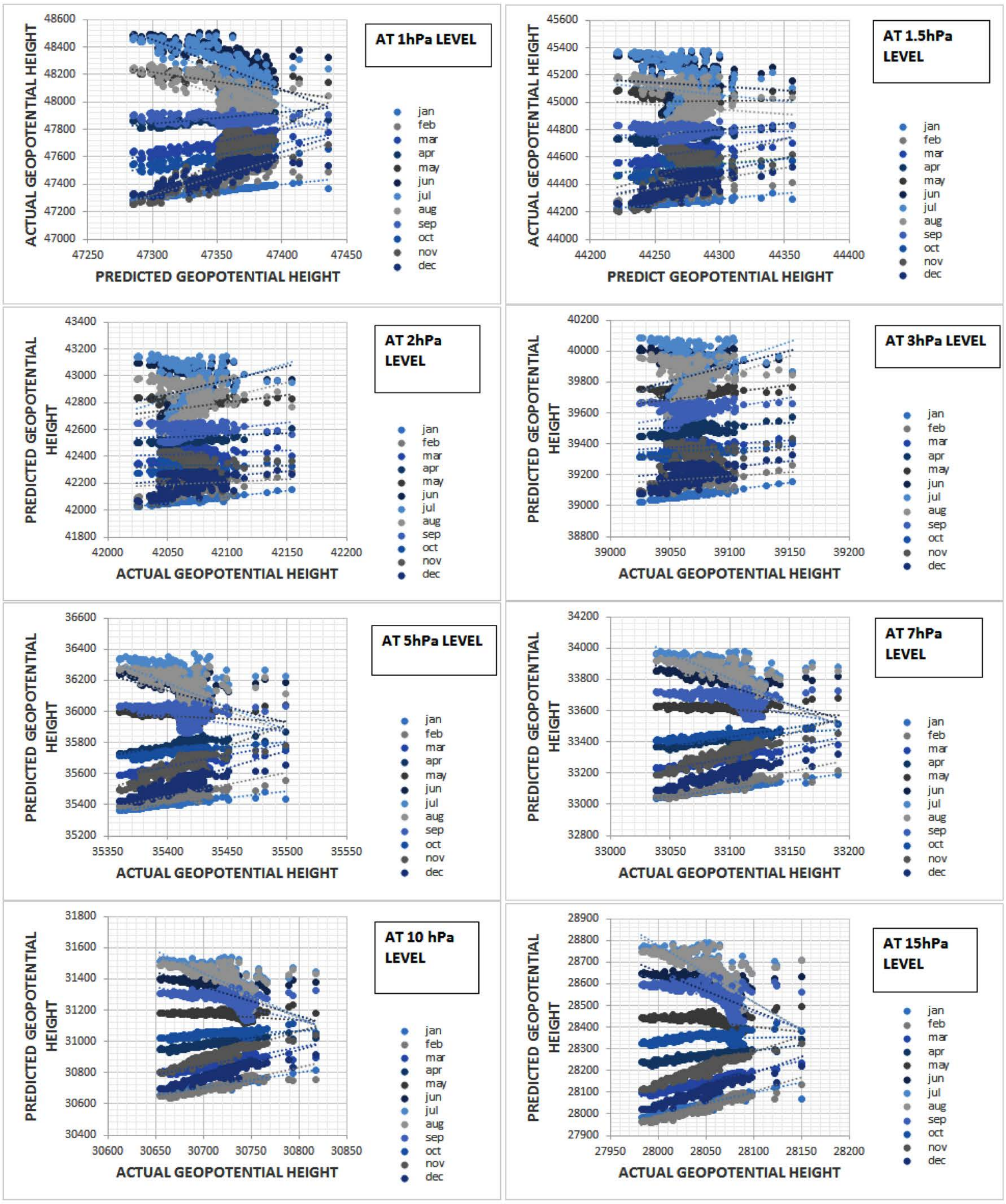
Pressure Level	RMSE	AME	Correlation	R^2
1 hPa	5.241294	1.673678	0.994263	0.982482
1.5 hPa	4.727445	1.538098	0.992346	0.978878
2 hPa	5.144201	1.546157	0.992445	0.978924
3 hPa	4.216421	1.403774	0.992625	0.979272
5 hPa	4.258006	1.303426	0.988246	0.970826
7 hPa	3.712901	1.191135	0.991712	0.977557
10 hPa	4.406922	1.25975	0.988496	0.971253
15 hPa	4.131028	1.198271	0.989455	0.973121
20 hPa	4.807294	1.359347	0.985551	0.965473
30 hPa	4.615745	1.272488	0.975812	0.948054
50 hPa	4.481513	1.24537	0.978242	0.952734
70 hPa	4.001382	1.207328	0.989019	0.97236
100 hPa	3.815742	1.206655	0.996643	0.987169
150 hPa	3.96526	1.236372	0.99677	0.987438

Continued

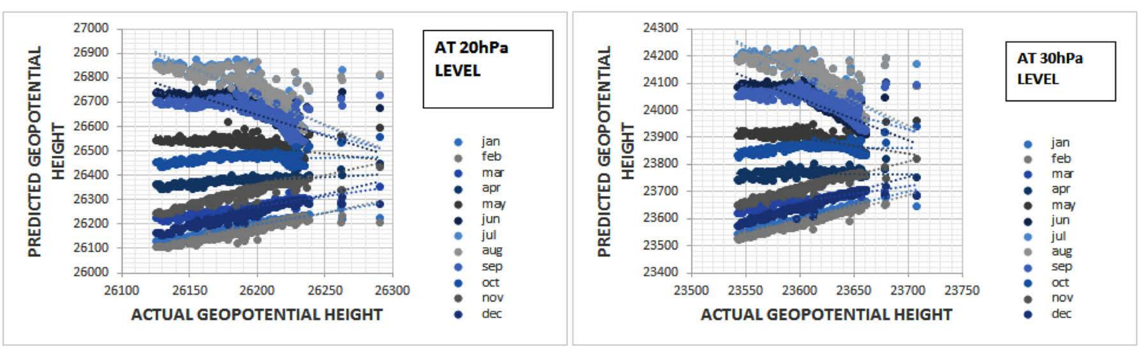
200 hPa	4.67426	1.333385	0.998104	0.990057
250 hPa	4.395547	1.319586	0.996534	0.986976
300 hPa	4.593847	1.33011	0.995398	0.98474
400 hPa	4.281792	1.23829	0.991174	0.976463
500 hPa	4.30804	1.212168	0.983423	0.961512
600 hPa	4.509835	1.3308	0.980018	0.954683
700 hPa	4.140531	1.158596	0.97152	0.938137
850 hPa	2.461281	0.770069	0.981623	0.957836
925 hPa	2.722891	0.795279	0.988701	0.971507
1000 hPa	2.52145	0.669249	0.986646	0.968239

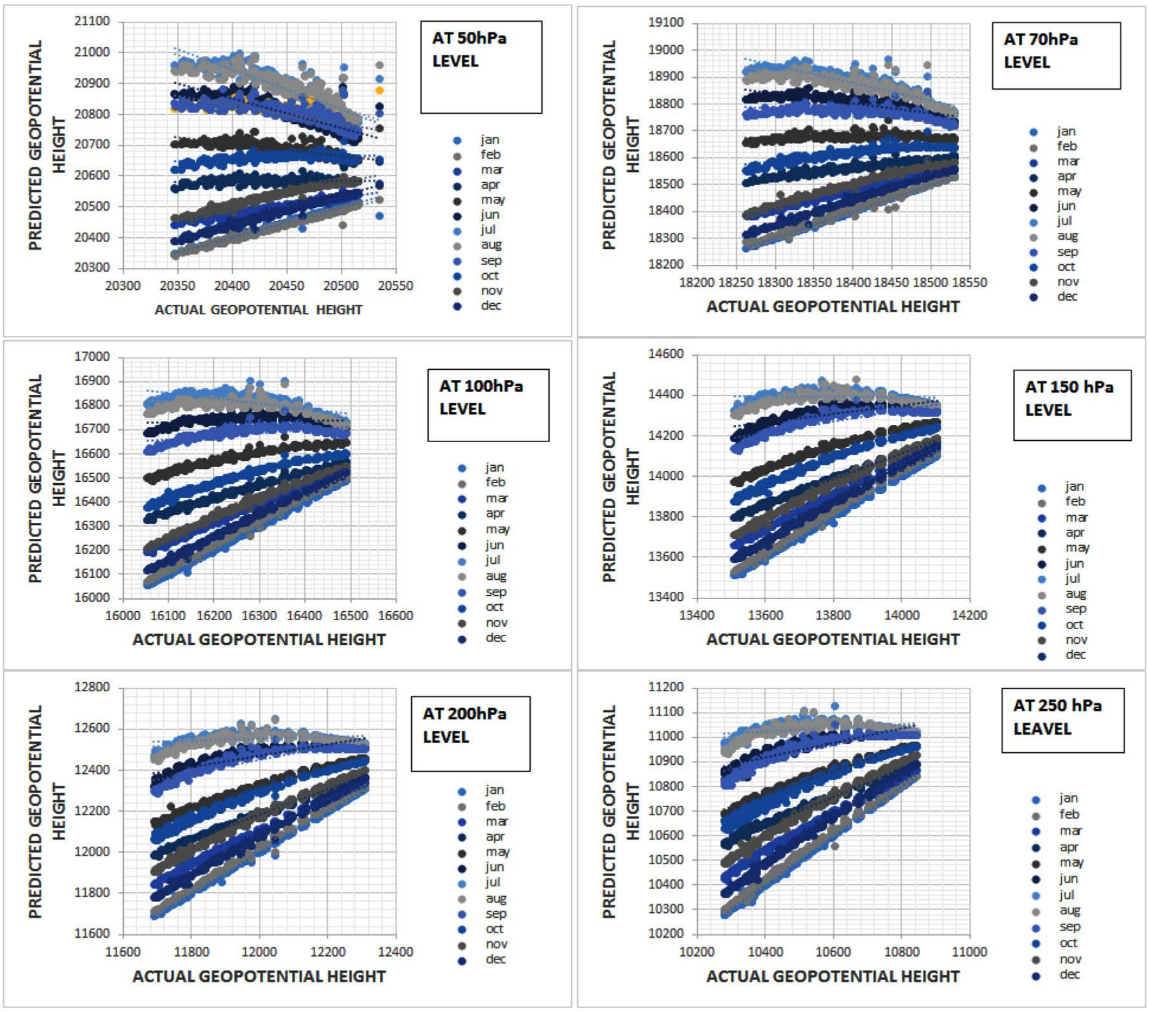
Table 4. Good results of performance indicators with Natural Neighbor Interpolation.

Pressure Level	RMSE	AME	Correlation	R^2
1 hPa	4.900104	1.64291	0.995605	0.98512
1.5 hPa	4.997688	1.620178	0.993656	0.981326
2 hPa	4.323708	1.466286	0.99451	0.982965
3 hPa	4.154958	1.328781	0.990859	0.975959
5 hPa	4.32224	1.388351	0.990687	0.975515
7 hPa	4.122477	1.27077	0.991668	0.977389
10 hPa	5.101086	1.41484	0.986257	0.96693
15 hPa	3.737039	1.146604	0.992122	0.978282
20 hPa	4.537878	1.274444	0.977021	0.950391
30 hPa	4.0766	1.160489	0.988907	0.972017
50 hPa	4.311496	1.22233	0.984238	0.963129
70 hPa	4.096588	1.184038	0.985229	0.965527
100 hPa	4.589851	1.356172	0.991379	0.976963
150 hPa	3.846614	1.223615	0.997187	0.988257
200 hPa	3.941453	1.285059	0.997698	0.989262
250 hPa	4.263726	1.286571	0.997583	0.989028
300 hPa	4.500221	1.31535	0.995976	0.985874
400 hPa	4.340019	1.27953	0.991552	0.977225
500 hPa	4.370527	1.287437	0.984613	0.963743
600 hPa	3.963257	1.203221	0.983525	0.961471
700 hPa	3.71123	1.096937	0.976069	0.947018
850 hPa	2.398759	0.70676	0.983242	0.960856
925 hPa	2.287255	0.683341	0.991601	0.977208
1000 hPa	2.484222	0.706179	0.987709	0.970155


AT 1hPa LEVEL
ACTUAL GEOPOTENTIAL HEIGHT
PREDICTED GEOPOTENTIAL HEIGHT
AT 1.5hPa LEVEL
ACTUAL GEOPOTENTIAL HEIGHT
PREDICT GEOPOTENTIAL HEIGHT
AT 2hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 3hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 5hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 7hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 10 hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 15hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT

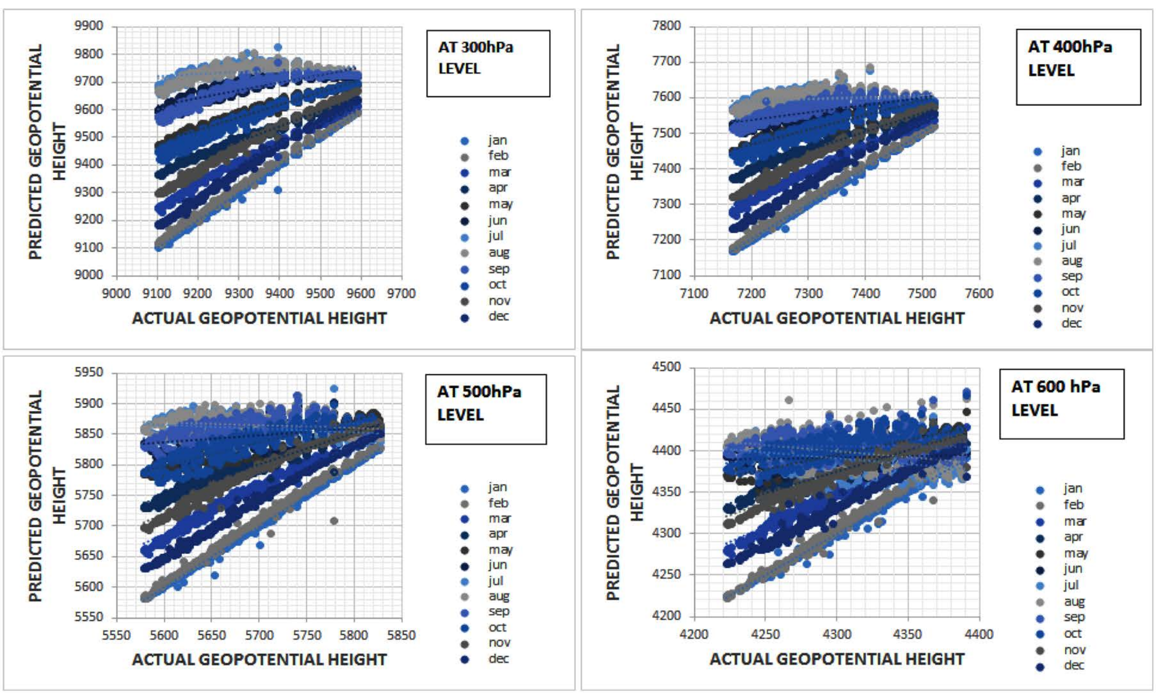
(a)


AT 20hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
AT 30hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT

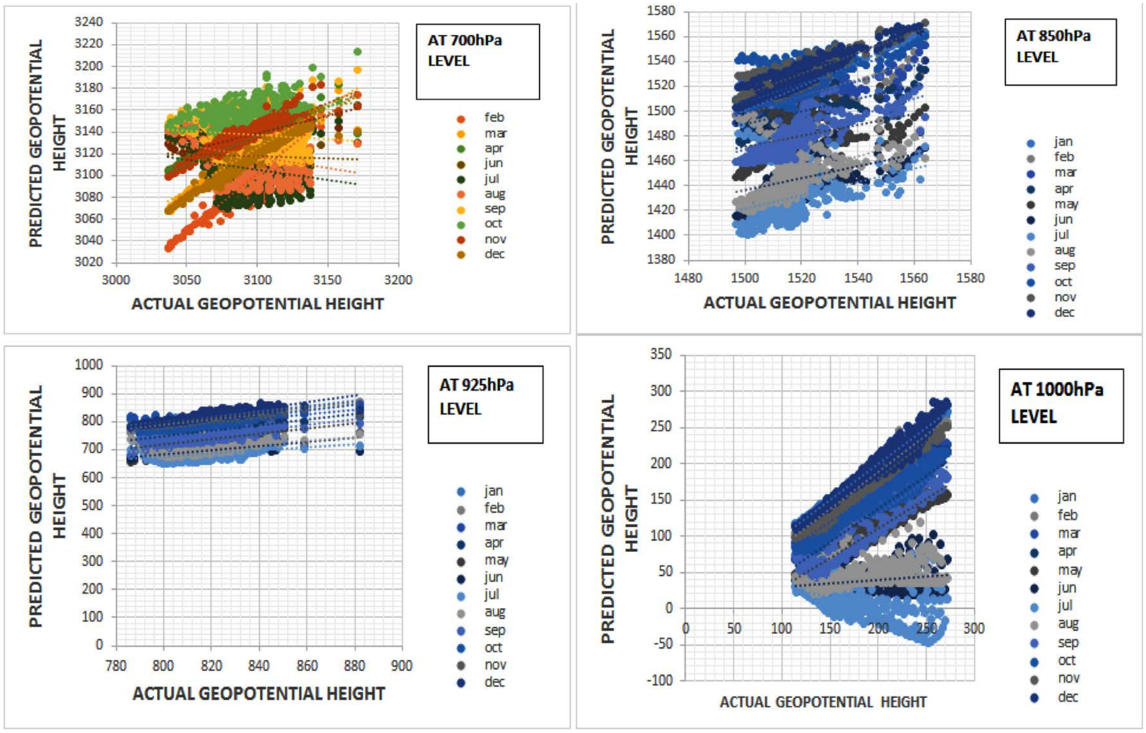


AT 50hPa LEVEL
AT 70hPa LEVEL
AT 100hPa LEVEL
AT 150 hPa LEVEL
AT 200hPa LEVEL
AT 250 hPa LEAVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec

(b)



AT 300hPa LEVEL
AT 400hPa LEVEL
AT 500hPa LEVEL
AT 600 hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec



(c)

Figure 3. (a) Results of interpolation of relative humidity [1 hPa to 15 hPa] with Bilinear Interpolation; (b) Bilinear Interpolation for relative humidity imputation from 20 hPa to 250 hPa; (c) Imputation of relative humidity from 300 hPa to 1000 hPa with Bilinear Interpolation.

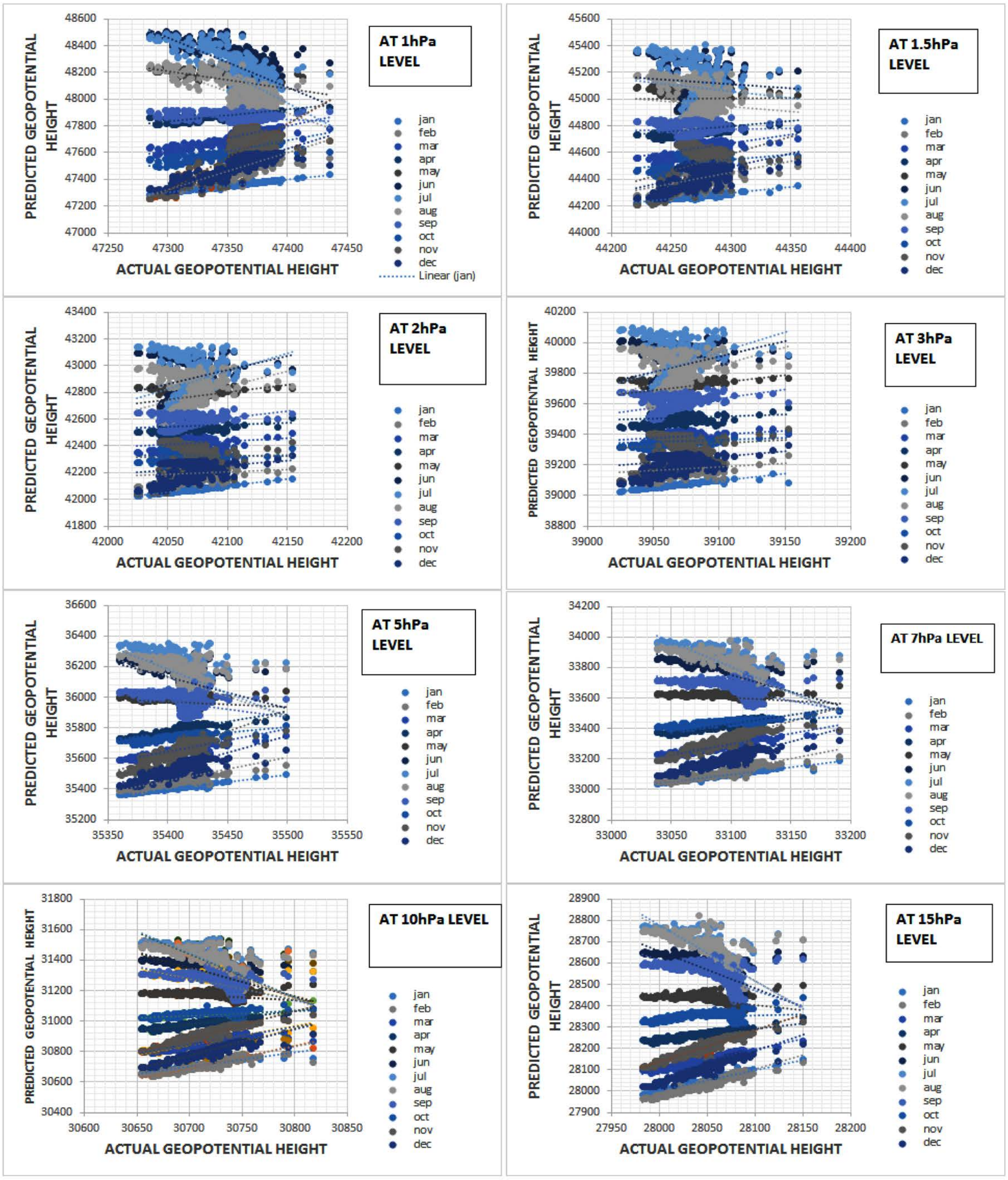
The Imputations for months of January, February and March were not precise (over 20, 30, 50, 70, 100, 150, 200, 250 hPa) with Bilinear Interpolation. Bilinear Interpolation for remaining pressure levels accurately filled the gaps in the Geopotential height (Figure 3(b)).

The original sample and imputed sample for each months were plotted together to create theses scatter plots. However (over 500, 600, 850, 1000 hPa) Bilinear Interpolation poorly filled months of February, March and April (Figure 3(c)).

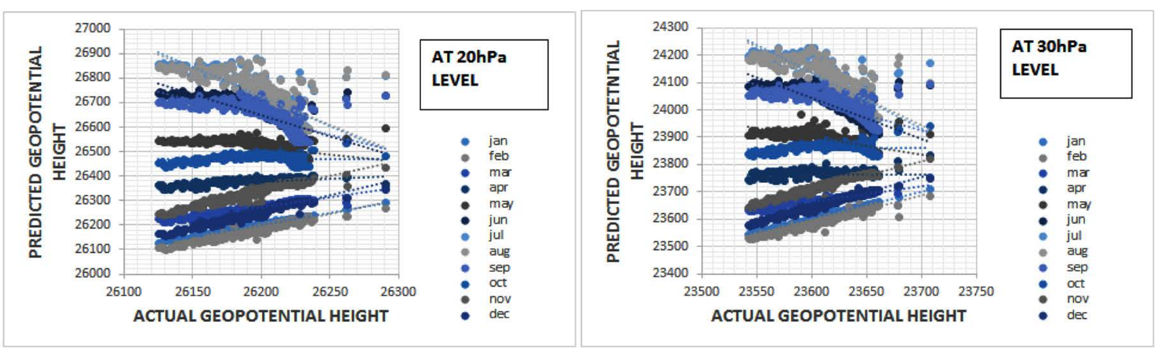
The similar technique of plotting original samples with imputed samples was used to create scatter plot of each month. Natural Neighbor Interpolation Imputations were more precise than Bilinear Interpolation. Only month of February was not good by Natural Neighbor Interpolation. Natural Neighbor Interpolation precisely filled the Geopotential height (over 1, 1.5, 2, 3, 5, 7, 10 and 15 hPa) (Figure 4(a)). The imputation with NNI for 20 hPa to 250 hPa and 300 hPa to 1000 hPa are illustrated in Figure 4(b) and Figure 4(c) respectively.

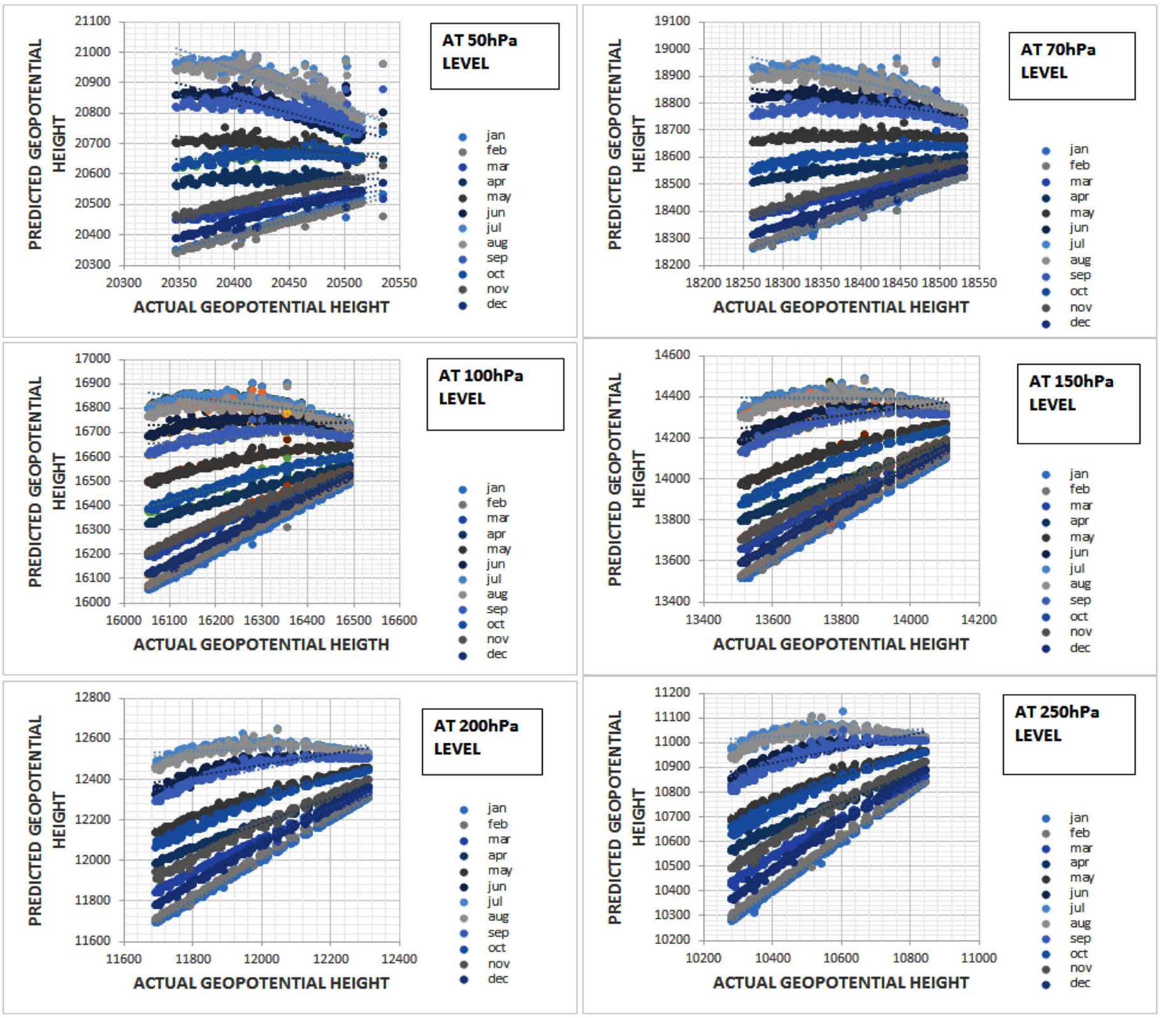
5. Conclusion

AQUA Satellite data was interpolated for Missing Data of Geopotential height. Based on critical checks and evaluation of interpolations regarding their product, it concluded that the NN and IDW interpolations for filling of missing

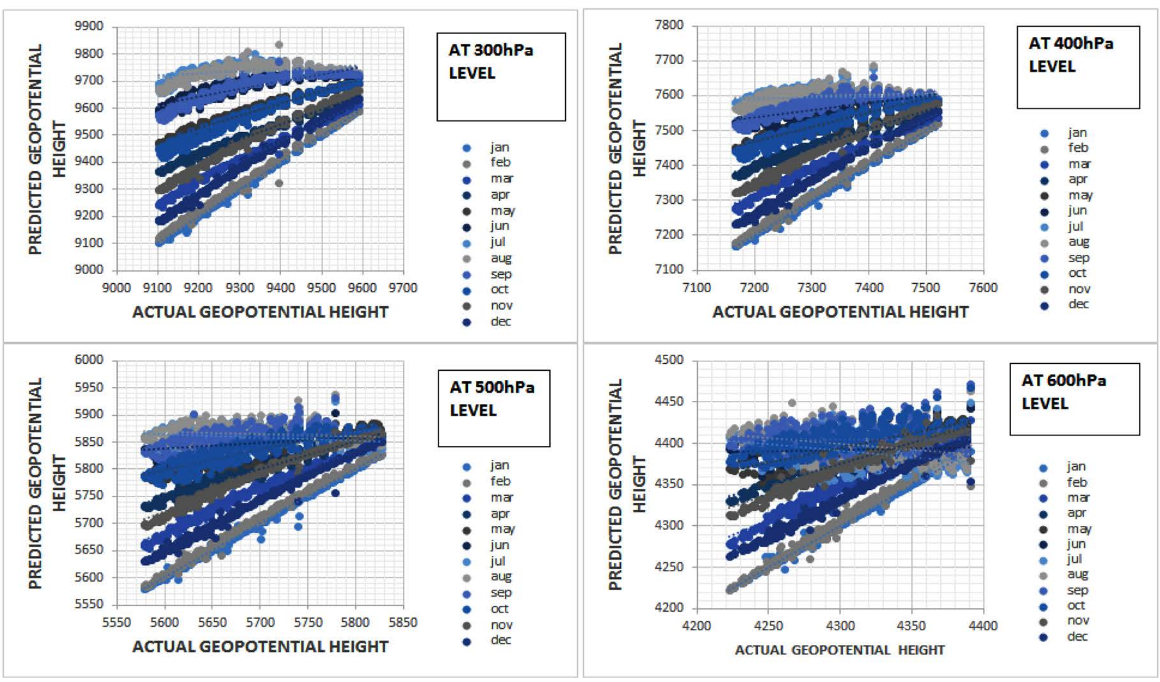

AT 1hPa LEVEL
AT 1.5hPa LEVEL
AT 2hPa LEVEL
AT 3hPa LEVEL
AT 5hPa LEVEL
AT 7hPa LEVEL
AT 10hPa LEVEL
AT 15hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec
Linear (jan)

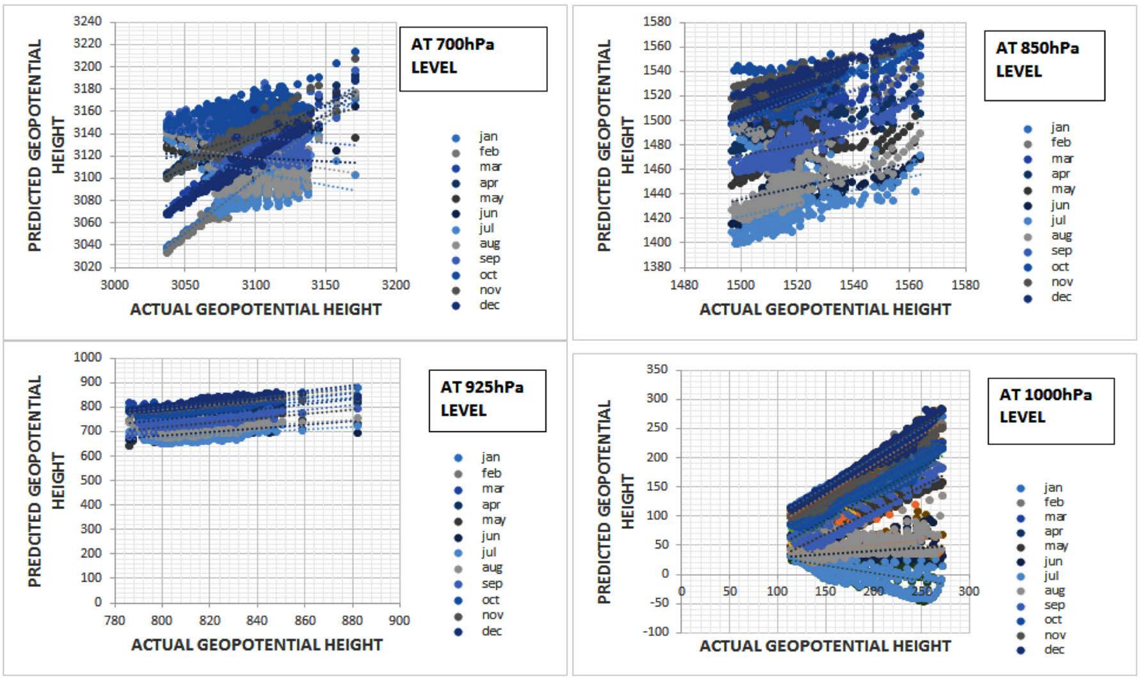
(a)


AT 20hPa LEVEL
AT 30hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec


AT 50hPa LEVEL
AT 70hPa LEVEL
AT 100hPa LEVEL
AT 150hPa LEVEL
AT 200hPa LEVEL
AT 250hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGTH
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec

(b)


AT 300hPa LEVEL
AT 400hPa LEVEL
AT 500hPa LEVEL
AT 600hPa LEVEL
PREDICTED GEOPOTENTIAL HEIGHT
ACTUAL GEOPOTENTIAL HEIGHT
jan
feb
mar
apr
may
jun
jul
aug
sep
oct
nov
dec



(c)

Figure 4. (a) Imputation of relative humidity (1 hPa to 15 hPa) with Natural Neighbor Interpolation; (b) Imputation of relative humidity (20 hPa to 250 hPa) with Natural Neighbor Interpolation; (c) Imputation of relative humidity (300 hPa to 1000 hPa) with Natural Neighbors Interpolation.

geopotential height data were proved not to be best and perfect (Table 1 and Table 2). Good results were found between BI and NI. However, after examining scatter plots of each month, it was found that NI was more accurate and reliable for missing data of geopotential height over 24 hPa levels.

Acknowledgements

The authors wish to acknowledge valuable guidance provided by Mr. Thomas Hearty and Mr. Edward T Olsen to refill gaps in AIRS relative humidity data set. The valuable suggestions are appreciated by Mr. Alessio Martion, University of the Rome, La Sapienza Italy which helped to improve this research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Schneider, T. (2001) Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values. *Journal of Climate*, **14**, 853-871.
https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2

[2] Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. and Kolehmainen, M. (2004) Methods for Imputation of Missing Values in Air Quality Data Sets. *Atmos-*

pheric Environment, **38**, 2895-2907. https://doi.org/10.1016/j.atmosenv.2004.02.026

[3] Yozgatligil, C., Aslan, S., Iyigun, C. and Batmaz, I. (2013) Comparison of Missing Value Imputation Methods in Time Series: The Case of Turkish Meteorological Data. *Theoretical and Applied Climatology*, **112**, 143-167. https://doi.org/10.1007/s00704-012-0723-x

[4] Noor, N.M., Abdullah, M.M.A.B., Yahaya, A.S. and Ramli, N.A. (2015) Comparison of Linear Interpolation Method and Mean Method to Replace the Missing Values in Environmental Data Set. *Materials Science Forum*, **803**, 278-281.

[5] Tian, B., Manning, E., Fetzer, E., Olsen, E., Wong, S., Susskind, J., *et al.* (2013) AIRS/AMSU/HSB Version 6 Level 3 Product User Guide. Jet Propulsion Laboratory, Pasadena, CA.

[6] Rahman, G. and Islam, Z. (2011) A Decision Tree-Based Missing Value Imputation Technique for Data Pre-Processing. *Proceedings of the Ninth Australasian Data Mining Conference*, Ballarat, Australia, 1-2 December 2011, 41-50.

[7] Perry, M. and Hollis, D. (2005) The Generation of Monthly Gridded Datasets for a Range of Climatic Variables over the UK. *International Journal of Climatology*: *A Journal of the Royal Meteorological Society*, **25**, 1041-1054. https://doi.org/10.1002/joc.1161

[8] Hofstra, N., Haylock, M., New, M., Jones, P. and Frei, C. (2008) Comparison of Six Methods for the Interpolation of Daily, European Climate Data. *Journal of Geophysical Research*: *Atmospheres*, **113**, 1-19. https://doi.org/10.1029/2008JD010100

[9] Ferrari, G.T. and Ozaki, V. (2014) Missing Data imputation of Climate Datasets: Implications to Modeling Extreme Drought Events. *Revista Brasileira de Meteorologia*, **29**, 21-28. https://doi.org/10.1590/S0102-77862014000100003

[10] Shen, R., Reiter, E.R. and Bresch, J.F. (1986) Vertical Interpolation of Meterological Variables. *Monthly Weather Review*, **114**, 123-134. https://doi.org/10.1175/1520-0493(1986)114<0123:VIOMV>2.0.CO;2

[11] Knox, J.L., Higuchi, K., Shabbar, A. and Sargent, N.E. (1988) Secular Variation of Northern Hemisphere 50 kPa Geopotential Height. *Journal of Climate*, **1**, 500-511. https://doi.org/10.1175/1520-0442(1988)001<0500:SVONHK>2.0.CO;2

[12] Shabbar, A., Higuchi, K. and Knox, J.L. (1990) Regional Analysis of Northern Hemisphere 50 kPa Geopotential Heights from 1946 to 1985. *Journal of Climate*, **3**, 543-557. https://doi.org/10.1175/1520-0442(1990)003<0543:RAONHK>2.0.CO;2

[13] Griesser, T., Brönnimann, S., Grant, A., Ewen, T., Stickler, A. and Comeaux, J. (2010) Reconstruction of Global Monthly Upper-Level Temperature and Geopotential Height Fields Back to 1880. *Journal of Climate*, **23**, 5590-5609. https://doi.org/10.1175/2010JCLI3056.1

[14] White, G.H. (1980) Skewness, Kurtosis and Extreme Values of Northern Hemisphere Geopotential Heights. *Monthly Weather Review*, **108**, 1446-1455. https://doi.org/10.1175/1520-0493(1980)108<1446:SKAEVO>2.0.CO;2

[15] Wallace, J.M., Zhang, Y. and Lau, K.-H. (1993) Structure and Seasonality of Interannual and Interdecadal Variability of the Geopotential Height and Temperature Fields in the Northern Hemisphere Troposphere. *Journal of Climate*, **6**, 2063-2082. https://doi.org/10.1175/1520-0442(1993)006<2063:SASOIA>2.0.CO;2

[16] Farooqi, A.B., Khan, A.H. and Mir, H. (2005) Climate Change Perspective in Pakistan. *Pakistan Journal of Meteorology*, **2**, 11-21.

[17] Iqbal, M.J. and Quamar, J. (2011) Measuring Temperature Variability of Five Major Cities of Pakistan. *Arabian Journal of Geosciences*, **4**, 595-606.

https://doi.org/10.1007/s12517-010-0224-0

[18] Saleem, M.U. and Ahmed, S.R. (2016) Missing Data Imputations for Upper Air Temperature at 24 Standard Pressure Levels over Pakistan Collected from Aqua Satellite. *Journal of Data Analysis and Information Processing*, **4**, 132-146. https://doi.org/10.4236/jdaip.2016.43012

[19] Saleem, M.U. (2016) Statistical Investigation and Mapping of Monthly Modified Refractivity Gradient over Pakistan at the 700 Hectopascal Level. *Open Journal of Antennas and Propagation*, **4**, 46-63. https://doi.org/10.4236/ojapr.2016.42005

[20] Saleem, M.U. (2015) Atmospheric Ducts: Their Applications in Radio Frequency Propagation Using Satellite Remote Sensing Techniques. LAP Lambert Academic Publishing, Germany, 56 p.

[21] Robeson, S.M. (1994) Influence of Spatial Sampling and Interpolation on Estimates of Air Temperature Change. *Climate Research*, **4**, 119-126. https://doi.org/10.3354/cr004119

[22] Price, D.T., McKenney, D.W., Nalder, I.A., Hutchinson, M.F. and Kesteven, J.L. (2000) A Comparison of Two Statistical Methods for Spatial Interpolation of Canadian Monthly Mean Climate Data. *Agricultural and Forest Meteorology*, **101**, 81-94. https://doi.org/10.1016/S0168-1923(99)00169-0

[23] Langella, G. (2010) Inverse Distance Weighted (IDW) or Simple Moving Average (SMA) INTERPOLATION. File Exchange Version 1.0.0.0.

[24] Stahl, K., Moore, R., Floyer, J., Asplin, M. and McKendry, I. (2006) Comparison of Approaches for Spatial Interpolation of Daily Air Temperature in a Large Region with Complex Topography and Highly Variable Station Density. *Agricultural and Forest Meteorology*, **139**, 224-236. https://doi.org/10.1016/j.agrformet.2006.07.004

[25] Boissonnat, J.-D. and Cazals, F. (2002) Smooth Surface Reconstruction via Natural Neighbour Interpolation of Distance Functions. *Computational Geometry*, **22**, 185-203. https://doi.org/10.1016/S0925-7721(01)00048-7