IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

Analysis of Modified SSA for Interpolation of Missing Values in Geo-Physical Data

Devarapalli Siva Subrahmanyeswara Rao

M.Tech, Dept of ECE, VR Siddhartha Engineering College, A.P, India

Abstract: To learn about the natural geophysical process first we need regular observations of various parameters such as precipitation, cloud water, temperature, and soil moisture. For example, the spatiotemporal data set of soil moisture is collected National and Snow Ice data centre. This spatiotemporal data set size is{ 44 × 102 × 92}.In this dataset, there are several inherent gaps due to various reasons such as sensor error, failure of data recording devices, lack of coverage and so on. The goal of an interpolation process is to estimate the unknown values in the gaps based on knowledge from existing values in the spatio-temporal neighbourhood. Here we plan to study an efficient spatio-temporal interpolation scheme based on singular spectral analysis (SSA). It was complicated because multiple hydrologic processes are there in soil moisture interpolation, this method uses to drive the content the modes of variation in soil moisture pattern to estimate missing values. The reconstructed data includes the estimates of missing values. The main goal of this research is to know the efficiency of the SSA based spatio-temporal data filling method.

Key words: NSIDC, AMSRE Data, Empirical Orthogonal Function (EOF), Singular Spectral Analysis (SSA).

I. INTRODUCTION

Earth sciences obey rules that are related to planet Earth. Geo-Scientists observe and gather data. They have uncertain in modelling the forces that have operated in a certain region. To learn about the natural geophysical process first we need regular observations of various parameters such as precipitation, cloud water, temperature, and soil moisture The Geo-Physical data is available in Advanced Microwave Scanning Radiometer Earth (AMSRE) observation system and we have collected the data from this system. By using this system we can measure the earth conditions like soil moisture, surface temperature. Soil moisture is one of the elements for controlling the swap over of water and heat energy between the land and plant transpiration and atmosphere through evaporation. As a result, soil moisture plays a very important role in the pattern development and the production of precipitation.

In this process some data will be lost due to the vegetation, UV rays or other climate conditions, to address this problem a new method has to be implemented. The main goal of this paper is implementation of singular spectral analysis based interpolation method. This method extracts singular spectrum from an incomplete data set. In this work we use SSA to approximate missing values from the covariance of known values in the given data plays a key role in SSA.

Earlier researchers (Beckers and Rixen in 2003) focused on the same problem [1] by applying empirical orthogonal function method in oceanographic data. This method was an partial reconstruction of the missing values, the error estimation is optimal for the number of spatial EOFs but recently [2] have developed a modified singular spectral analysis (MSSA) method to interpolate the gaps to fulfil these soil moisture data. The spatiotemporal empirical orthogonal function (EOFs) is reckoned in the spatio temporal data, and the SSA parameters are interpolated using the experimental readings of a data set. By using SSA method, the data reconstruction process was clearly explained.

In the spatiotemporal SSA and the spatial EOF methods, the whole spatial grid is considered for the calculation of lag covariance. The method [2] detains the best size of 3-D spatial blocks for the estimation of lag covariance.

This can be described from the following 1) local spatiotemporal variations to estimate covariance of a given values in the data set and 2) this method requires number of calculations for building the covariance matrix and the corresponding Eigen analysis is reduced. This memo is organized the following lines 1) The mathematical report in proposed method; 2) Experiment on a few sample data sets and 3) conclusion and scope of upcoming work

II. METHODOLOGY

AMSRE system in the Aqua Satellite records radiation temperature data from the Earth. That signal passes through the ground station. The data collected by the ground station. The below figure (Fig 1) shows the representation of spatial data in SSA method.

MW Signal: Microwave Signal
NSIDC: National Snow and Ice Data Centre
SM: Soil Moisture
SSA: Singular Spectral Analysis

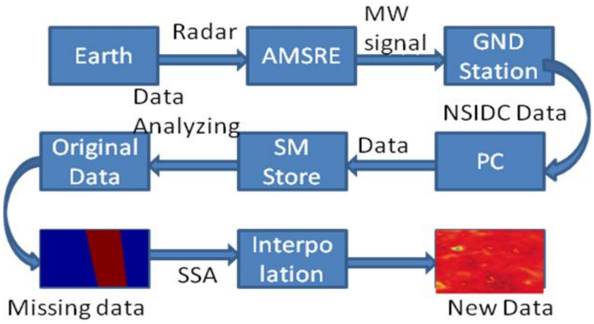

Figure.1 Spatial data representation in SSA method

From NSIDC data centre the soil moisture data is collected. In this soil moisture contains aqua level-3 data, with spatial resolution 25 km that soil moisture data has daily coverage at best. The drawback is that the data has missing value points. For reducing that problem first step is to find out the missing points in the actual data. The data is filled by using the SSA method. This method evolves the singular value decomposition and reconstruction processes. The missing values found from the actual data set can be filling by the reconstructed data set. In this order to find each and every value, we need do eigenanalysis, SSA is depends on the co-variance matrix, which decomposes eigenvectors and eigenvalues. First consider the spatio temporal data set E. The size of E is [M1 × M2 × M], where M1 and M2 are spatial data block dimensions and M is the number of days. The main intend of this paper is to reconcile the finest values M1 and M2 for M_{lat} and M_{lon} chosen data set shown in below figure 2.

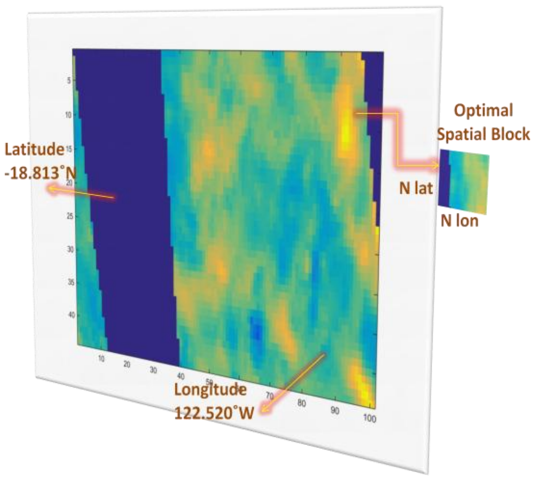

Figure 2: spatial grid explanation of missing values and certainty of the optimal subset.

The hugest hypothesis of this sort of investigation is there ways out a crucial structure in the information that can be unfaltering and used to entomb the information properties. In the event that the gaps found in the information set should be situated at fairly visit interims in contrast with the huge close-by wedges of missing information. On the off chance that the Trajectory network of size X (b, n), it is developed in L_w delay adaptations of E it was clarified in [4], where L_w is the length of the window under thought.

To ascertain the covariance E of the grid X, the vectors X are identified and the related covariance of the information set is figured, i.e.

$$E = \frac{1}{M_L}\left(X-\overline{X}\right)^T\left(X-\overline{X}\right). \qquad (1)$$

The measurements of lattice E is (Nmodes×Nmodes). For instance, a little information piece is not required for the enough covariance to estimated the missing qualities. The covariance figured from a greatly enormous information piece is not reasonable in light of the fact that its holes are likewise vast. Therefore, there is a need to choose a most positive size of an information piece, which requires a reliable covariance development. The Eigen disintegration is connected to that of the covariance grid of (1). In this strategy it can be composed as E=U*K*UX in this U includes the eigenvectors and the eigenvalues, K includes the corner to corner grid kr. These qualities are masterminded in the lower level request in K [3], [6]. In this framework documentation if the projection grid can be computed as B = XU. In the following stride by utilizing these estimations and to dole out specific SSA modes, the first information set is not completely recreated. On the off chance that the missing quality got from the real dataset, this information set is filled by the reproduced dataset. Keeping in mind the end goal to fill these misfortune values in the genuine information set, and it can be accomplished by figuring the reproductions G (d_i, v, s_j) of the real information set, i.e.,

$$G(d,v,s) = \frac{1}{N_S}\sum_{p=V_t}^{R_t} A(d,s-p+1)\,U(d,v,p) \quad (2)$$

Above condition (2), where N is the quantity of SSA modes chose from the reproduction forms in [3], [6]. Here d, v and s are given by its part d is in spatial, v is in mode, and s demonstrates for the fleeting files. At that point N_s is the variable and if the limits V_t and R_t could be clarified by the given conditions in [6]. At that point this halfway recreation of Gd information is given by

$$G_d(w,s) = \sum_{l=1}^{M_{SSA}} G(w,v,s) + \overline{X}. \qquad (3)$$

The quality in Gd at the areas dmissing are then embedded information set in G. The noteworthy technique for this calculation is to find the most ideal parameters for the MSSA calculation, to be specific in this strategy to consider the piece size, if the width is consider as the fleeting window and the methods of SSA is {M1, M2, N_{win}, MSSA} ideal. In this enhancement technique we present the efficient and engineered crevices into the information set. The examination of these crevices must be identified with those of the genuine holes. By utilizing the altered SSA strategy to insert these missing qualities in the information set. The parameters for {M1, M2, N_{win}, and MSSA} ideal are chosen doing the correlation of added qualities with that of the first values. We decide the ideal benefits of utilizing a lattice hunt down the obscure

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

parameters to minimize the blunder between the genuine interjected values.

III. EXPERIMENT

Validation: Expansion of real data with missing values by considering the spatiotemporal optimal validation data set. In this data set size is {44 × 102 × 10}. The first step in this process is to manually eliminate some random data points from the obtained spatiotemporal data. This new set of data point is interpolated to obtain the missing values and compared to the actual data. In that given data set first to calculate the missing values. The interpolation will be done using the SSA method. The correlation is calculated between the original data set of soil moisture values of the manually given gaps and matching values. From this data set after completion of interpolation process the outputs are original data, missing data and the interpolated data it could briefly as shown in the below figure 1. Now, the reconstruction error is also calculated between the validation data set of soil moisture values and the corresponding interpolated values.

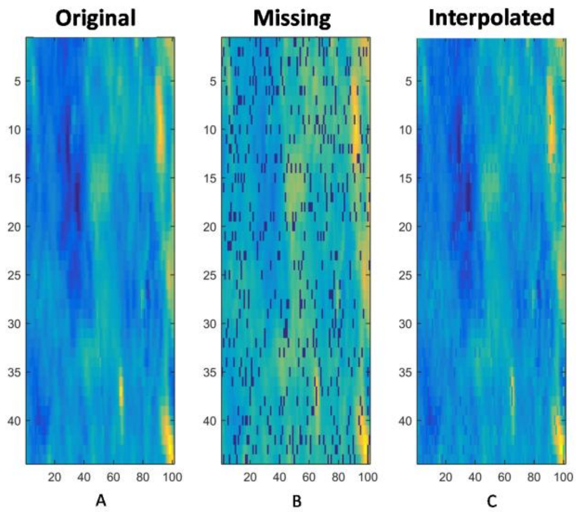

Figure-1: Interpolation from the validation data

However, the spatial-temporal block size considered for this experiment is {7×7}. It is observed that, using a spatial-temporal block gives better performance than using the entire data set. Another fact is that, an increase in the spatial-temporal block size also gives better results and good growth from the analysis pertained in the data set.

From this exploration, the information set of spatiotemporal soil moisture data is collected from the aqua satellite of Advanced Microwave Scanning Radiometer-Earth Observation System (AMSR-E) and this data presented in the National Snow and ice Data Centre (NSIDC). This data is available in the form of Aqua daily "AE_land3" level 3 product, and this data can be spread by the NSIDC. The release version of this data is "version 002." The AMSR-E system measures the 6.9-GHz, 10.7-GHz and up to 89.0 GHz brightness temperature in a few bands of the microwave frequencies. In this system, the data set has a 25-km spatial resolution and the structural errors are verified through this data set involving the corner coordinates. In this data set, the level 3 outcomes are obtained from advanced Level 2B soil moisture estimation [2].

Here, we consider the original spatiotemporal data set obtained from NSIDC data centre. This data set of size is (44 × 102 × 92). From the pool of data available, only the soil moisture data field is extracted for the desired region i.e. North-West Australia to South-East Australia consisting of the states of Western Australia, Queensland, South Australia, Northern Territory and New South Wales.

The data set was collected for the Winter 2004. The data collected it has several gaps due to various environmental and human factors. The environmental factors include the density of the vegetation and atmospheric interference. The tech factors include the intervals between consecutive visits, functional ability of the instruments, etc. In this data set have an in complete data set (missing values). Now we interpolate missing data using singular spectral analysis (SSA) method.

This region chosen for the study lies between -19˚N to -31˚N latitudes and 126˚W to 144˚W longitudes. The satellite image of the geographical location is as shown in Figure 2.

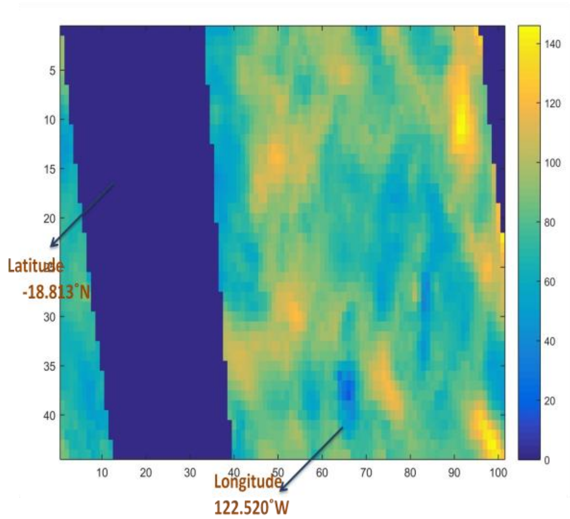

Figure.2 Missing Data over the spatiotemporal Data Set at -18.813˚N and 122.520˚W

The acceptable percentage of data points missing is usually around 35 percent. The missing data from the original data set is indicated in blue patches. These blue patches appear to be covering more than 35 percent of the data set. As part of validating the algorithm, about 15 percent of synthetic and systematic gaps are evolved in the data set. This type of data interpolate by using SSA method to fill the actual values from on the original data set. as shown in below figure 3.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

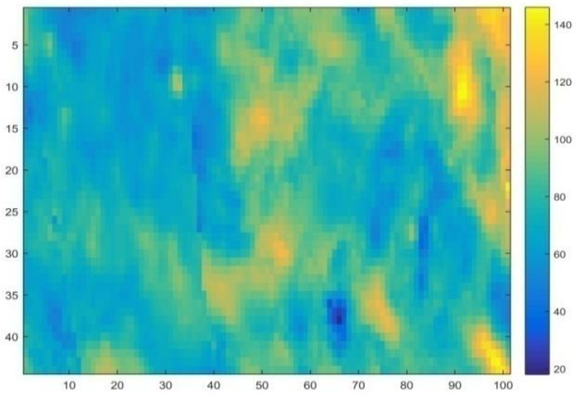

Figure .3: Implications of SSA method to the Original Data Set

If the below plot (Fig 4) clearly shows that the blue missing coloured 'o' are missing values and the red colour squares are interpolated values.

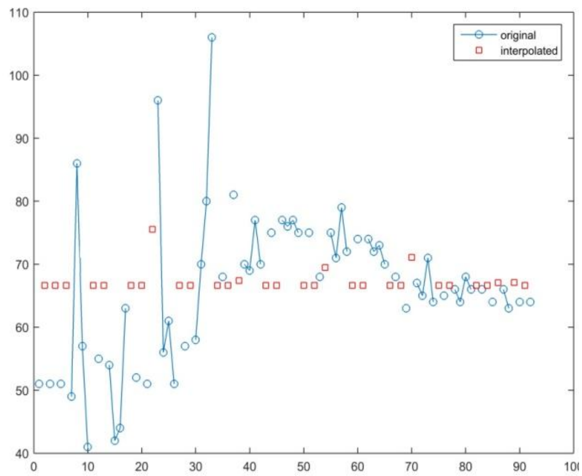

Figure-4: Modified SSA interpolation method of missing values and original values of data comparison

The Eigen analysis takes the major chunk of the calculation time in this algorithm as it provides the covariance matrix. These possessions required to perform the eigenanalysis usually the covariance matrix will depend on the size of the information set. The size of the block and the temporal window length are the two major factors influencing the computational time. The size of the covariance matrix to that of the optimal calculation time and least reconstruction inaccuracy is obtained when the parameter set {44 × 102 × 92} is used. A comparative study of original data interpolation in SSA method and customized approach at -18.813˚N and 122.520˚W is obtainable in Fig. 4. The blue circles and the red squares represent the synthetic gaps. After seeing the result it's clearly visible in the place of blue patches to do the SSA of interpolation the missing values replaces the new values these values are more accurate and closer at the most of the data points in the original data set. In this a small amount of interruptions in the original data. This data set is the result of real gaps.

This indicates the significance of local correlations in the evaluation of missing values. The parameters obtained are used to fill in the actual gaps at -18.813˚N and 122.520˚W and the resulting series has been shown in Fig 5.

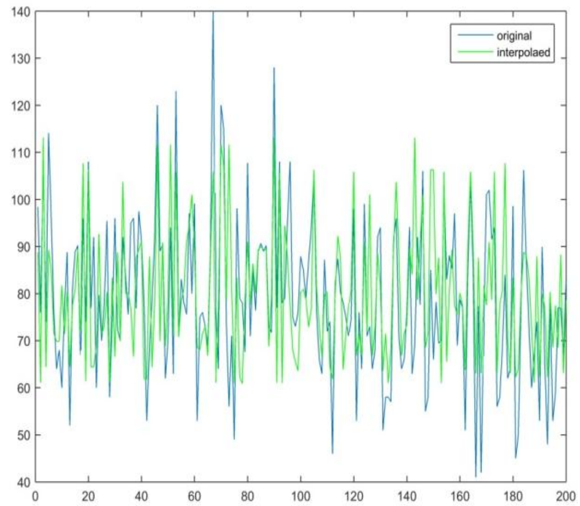

Figure.5 Spatio temporal data original and interpolated plot gap filled at -18.813˚N and 122.520˚W

From the time series, we can observe that the missing points occur in a systematic way and also at regular intervals. The blue graph indicates the time series with real gaps and the green graph indicates the time series with the interpolated values. The mean square error between validation data and the SSA interpolated data is 0.0192 so this interpolation method gives better performance compare to other algorithms

IV. CONCLUSION

A modified SSA-based interpolation has been applied from the soil moisture data. The main concept of this paper is to finding the actual data from the missing values, because missing values also have the information. In this problem to overcome by the better analysis and to do the researches for the further studies from the spatiotemporal data sets. From this spatiotemporal soil moisture data includes the synthetic and systematic gaps, by applying the modified SSA method to interpolate these synthetic and systemic problems and fixes the new values. After completion of interpolation process the interpolated data is matched by the actual data. This process is applicable to fill the real inherent gaps in the same spatiotemporal soil moisture data. In this exploration to enhances vast geophysical information sets, for example, ocean surface temperature and precipitation rate by deciding a comparative arrangement of ideal parameters.

ACKNOWLEDGEMENT

We wish to acknowledge the soil moisture data. It will be taken from AMSRE system in aqua satellite and this data consists of National Snow and Ice Data centre.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 9, September 2016

REFERENCES

[1] J.M.Beckers and M.Rixen "EOF calculations and data filling from incomplete oceanographic data sets" Atoms.Ocean.Technol. vol. 20, no.12, pp. 1839-1856. Dec 2003.
[2] Anish C.Turlapaty "IEEE Geosciences and Remote Sensing Letters", vol.8, No.2, March 2011.
[3] G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd ed. Baltimore, MD: The Johns Hopkins Univ. Press, 1996.
[4] M. Ghil, M. R. Allen, M. D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, and P. Yiou, "Advanced spectral methods for climate time series," Rev. Geophys.,vol. 40, no. 1, p. 141, Sep. 2001.
[5] S. Esakkirajan, T. Veerakumar, and P. Navaneethan, "Best basis selectionusing singular value decomposition," in Proc. 7th ICAPR, Kolkata, India, 2009, pp. 65–68.
[6] G. Strang, 'Eigenvalues and Eigenvectors,' in introducing to Linear Algebra, 3rd edition. Cambridge, MA: Wellesley Cambridge, 1998, pp. 274-352.

BIOGRAPHY

D. Siva Subrahmanyeswara Rao was born in Vijayawada in the year 1991. He received B.Tech degree in electronics and communication engineering from the Nova college of engineering and technology, Vijayawada, India, in 2012, and present pursuing M.Tech in communication engineering and signal processing from the VR. Siddhartha engineering college, kanuru, India, respectively.

 DOI 10.17148/IJARCCE.2016.5920