

Interpolation of Geophysical Data using Spatio-temporal (3D) Block Singular Value Decomposition

Anish C. Turlapaty^{1, 2}, Nicolas H. Younan¹, and Valentine G. Anantharaj²

¹Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762

²Geosystems Research Institute, Mississippi State University, Mississippi State, MS 39762, USA

Background: Spectral analysis of geophysical data is a popular tool for understanding the spatio-temporal signals underlying the data. These frequency domain methods need complete data with uniform sampling for a thorough analysis. In general, geophysical data obtained through satellites have several gaps. Several alternative techniques have been developed to analyze such datasets. However, interpolation of the missing values is the best solution to facilitate an effective analysis of geophysical data. For instance, in a given region over a period of time the soil moisture products available from the Advanced Microwave Scanning Radiometer-Earth Observation System (AMSR-E) onboard the National Aeronautic and Space Administration's (NASA) Aqua satellite has many inherent gaps due to the orbital coverage of the satellite. For a region in the Southeast United States, data is collected for years 2005 and 2006. This dataset has nearly 30% missing data due to sampling patterns, radio interference, retrieval issues and instrument errors. Interpolation of these missing values can be invaluable in hydrological applications.

Proposed Method: To address this issue, we have developed and implemented an iterative block SVD interpolation scheme. Our methodology is a generalization of Becker and Rixen's SVD method [1]. This revised technique approaches the SVD method when the spatio-temporal block equals the entire dataset and it is similar to the M-SSA (Multi channel Singular spectral analysis) method with $M = 1$ [2]. Our SVD-based interpolation scheme consists of two stages: decomposition and reconstruction. In step one, SVD analysis is performed on the full covariance matrix of a spatio-temporal data (3D) block to generate a set of eigenvectors and eigenvalues. From the set of eigenvectors, the dominant modes that contribute to a major portion of the variance of the dataset are selected. Then, projections of the dataset on each of the mode or eigenvector are computed. In step two, using these projections and the selected SVD modes, the

dataset is partially reconstructed. This method is similar to the reconstruction of a time series from Fourier coefficients or any other types of spectral coefficients. The missing values of the original dataset are filled with the values from the reconstructed dataset. These two steps are repeated recursively until the mean square error between two consecutive datasets converges to a minimum possible value. Once the gaps in the initial data block are filled; the process is repeated on the remaining blocks in a raster scan fashion.

Validation: The algorithm is validated using a sea surface temperature (SST) dataset. The SST data used here is level-4 analysis from the Global Ocean Data Assimilation Experiment high resolution SST pilot project (GHRSSST-PP). The level-4 data is a fusion of four sets of microwave observations retrieved from the following instruments: AMSR-E, Advanced very high resolution radiometer onboard the Television infrared observation satellite, Spinning Enhanced Visible and Infrared Imager onboard Meteosat Second Generation satellite, and Advanced Along Track Scanning Radiometer onboard the European space agency's environmental satellite. This dataset is a daily SST with a 25km spatial resolution and global coverage [3]. A set of systematic gaps are introduced in the SST dataset; and as a result 29% of the data is dropped. In the experiments with this datasets, the performance of the algorithm is similar to that of the standard interpolation schemes (Figure 1). The MSE comparison performed among the different methods, as illustrated in Figure 1, shows that the performance of our SVD method generally improves with the number of time steps of the data in comparison to other standard methods.

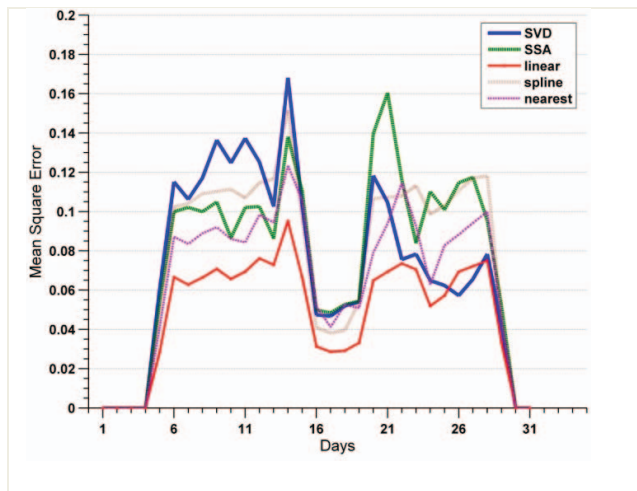


Figure 1. MSE Comparison between the actual SST versus interpolated SST, on a daily basis, computed from different interpolation algorithms.

Interpolation of soil moisture data and discussion: In a second validation experiment, we used the NASA AMSR-E Level 3 (AE_Land3 – Beta 03 release of Version 001) soil moisture product from the, distributed by the National Snow and Ice Data Center (NSIDC). The dataset has a 25km spatial resolution [4]. Soil moisture data fields are extracted and data sets are generated for a region consisting of the states of Mississippi, Arkansas, and a part of Louisiana. The dataset is collected over a period of two years from Jan 2005 to Dec 2006. The percentage of missing data points in AMSR-E soil moisture retrievals is usually under 35 with some exceptions. For fall 2005, a spatial distribution of the number of missing data points as a fraction of the total number of data points is shown in Figure 2(a). The method is tested on data from spring 2005 to fall 2006. For comparison purposes, The SSA algorithm is also applied to the same seasonal data sets. A temporal illustration of soil moisture retrievals at a grid point is presented in Figure 2(b). The plots in blue and red are the time series obtained from the new SVD and SSA methods. It is evident that the interpolation process is successful in retaining the temporal structure of the original data as there are no sudden jumps in the time series. The soil moisture image for October 16, 2005 with data filled using the 3D block SVD method is shown in Figure 2(c) and the result from the SSA method is shown in Figure 2(d). It is evident that both methods yield similar spatial structures.

An important characteristic of this method is that spatio-temporal neighborhood of a missing value can have significant information about that value. The method is validated on a dataset of particular importance for understanding interactions among climate processes. A spatio-temporal block size of $7 \times 7 \times 90$ is found to be optimal for interpolation used in the validation dataset. The interpolated data matched well with the actual data. The method is later applied to soil moisture dataset with many intermittent gaps. The interpolated time series is compared with the results obtained from the 1-SSA gap filling method. The two approaches agreed well even when a large percentage of data was missing.

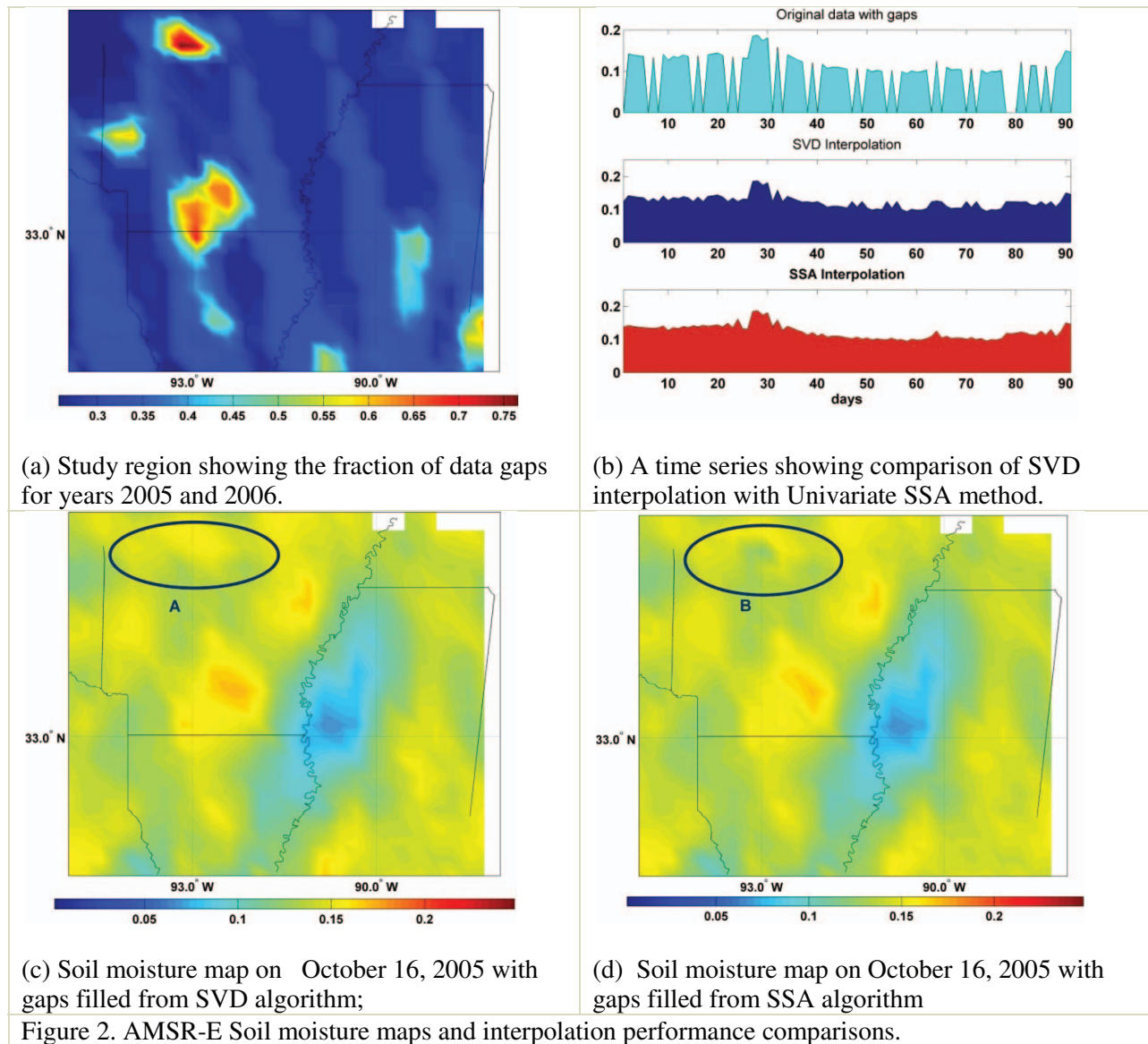


Figure 2. AMSR-E Soil moisture maps and interpolation performance comparisons.

References:

- [1] J.M. Beckers, M. Rixen, "EOF calculations and Data Filling from Incomplete Oceanographic Datasets," *J Atmos. Ocean Tech.*, vol. 20, no. 12, pp. 1839 – 1856, Dec. 2003.
- [2] D. Kondrashov, M. Ghil, "Spatio-temporal filling of missing points in geophysical data sets," *Nonlinear Proc. Geoph.*, vol. 13, pp 151- 159, May. 2006.
- [3] K.S. Casey, "GODAE high resolution SST products from the GDAC and LTSRF", Long term stewardship and reanalysis facility NOAA national oceanographic data center, USA, 2007.
- [4] E.G. Njoku, "AMSR Land Surface Parameters, Algorithm Theoretical Basis Document," Version 3.0. NASA Jet Propulsion Laboratory. Pasadena, CA, 1999.