5th ANKARA INTERNATIONAL AEROSPACE CONFERENCE **AIAC-2009-135**
17-19 August, 2009 - METU, Ankara

DEVELOPMENT OF A POD BASED DATA ASSIMILATION METHOD FOR METEOROLOGICAL FLOWFIELDS

Mustafa Kaya*, Gökhan Ahmet†, Sezgi Kumru‡ and Ismail H. Tuncer§
Middle East Technical University
Ankara, Turkey

ABSTRACT

Numerical weather prediction software requires assimilation of observation data obtained from weather observation stations for an accurate prediction. MM5 is an open-source software widely used for weather prediction over the world, including Turkish State Meteorological Service (DMİ). The accompanying software for MM5 such as 3DVAR and 4DVAR are used for assimilating observation data into numerical predictions. In this study a new data assimilation method based on Proper Orthogonal Decomposition is developed. The 3DVAR based data assimilation methodology for Turkey is first adapted into MM5 solutions. Then, the data assimilation using the developed method is presented and compared against the 3DVAR based weather predictions over Turkey.

INTRODUCTION

The Pennsylvania State University / National Center for Atmospheric Research Mesoscale Model, which is known as MM5, is a limited-area, nonhydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale atmospheric circulation[2]. MM5 solves the unsteady Navier-Stokes equations using a finite difference method. The solution may be obtained with the use of nested grids of various scales. MM5 is able to compute the flowfields in a large resolution range (2-200 km) which consists of the meso-beta and meso-gamma scales. The MM5 program is supported by several pre- and post-processing programs, which are mostly written in Fortran, and has been developed at The Pennsylvania State University and the National Center for Atmospheric Research as a community mesoscale model with contributions from users worldwide.

The numerical weather prediction (NWP) based on flowfield computations is an initial/boundary value problem. The initial and boundary conditions are provided from the solution of another numerical model with a larger scale, such as from a global atmospheric flow state. The quality of the prediction is directly related to how close the initial conditions are to the real flowfield. NWP centers produce initial conditions using generally a mathematical/statistical combination of the larger scale solutions and the meteorological data obtained from the local observation stations. The use of observation data within the flow solutions is known as data assimilation. In the last 50 years, there are a large number of studies on the data assimilation of the local observational data. Successive Corrections Method (SCM), Optimal Interpolation (OI), Nudging, 3-Dimensional Variational Approach (3DVAR), 4-Dimensional Variational Approach (4DVAR) and Kalman Filtering (KF) methods are widely used data assimilation techniques[5].

The SCM and the OI are interpolation and statistics based assimilation methods. The Nudging, 3DVAR, 4DVAR and KF methods are based on both the solution of the flow equations and the statistical data. The

*Ph.D. Graduate from Aerospace Engineering Department, Email: mkaya@ae.metu.edu.tr
†Graduate Research Assistant in Aerospace Engineering Department, Email: gahmet@ae.metu.edu.tr
‡Graduate Student in Aerospace Engineering Department, Email: skumru@ae.metu.edu.tr
§Prof. Dr. in Aerospace Engineering Department, Email: tuncer@ae.metu.edu.tr

SCM, which is an empirical approach, was firstly used in 1955[17]. Using the empirical interpolation weights for a specified domain, successive corrections are made to approach to the observed data. The OI is based on the least squares method and firstly used in 1963[15]. It is based on a linear combination of the observed data and the numerical forecasts in such a way that the true flowfield is statistically best approached. In the Nudging method[13], the flow equations are modified using a Newtonian relaxation for a convergence to the observed data along an empirical time scale. The 3DVAR[14] and 4DVAR[12] methods are solving an optimization problem to determine an initial flowfield that is closest to the three and four dimensional observation data. In this optimization problem, the objective function is the statistically scaled difference between the initial flowfield and the observation data. The optimization variables are all the flow variables on the nodes of the computational domain. Although the 3DVAR and 4DVAR methods are very successful in the assimilation of the observation data into the numerical solution, their computational cost is significantly high. The KF method[16], known also as sequential assimilation, takes the uncertainty of the weather forecast model into account so that the assimilation is more accurate. KF equations are in fact the same with ones used in the OI analysis. However, the forecast error covariance matrix is predicted in the OI, whereas it is computed in the KF by solving the flow equations. The 4DVAR and KF methods providing four-dimensional data assimilation, are highly accepted since they are very accurate, however both method requires long computation times[7].

Turkish State Meteorological Service (DMİ) employs MM5 for weather predictions over Turkey. A 9-km-resolution grid is used for the flowfield computation over Turkey, while 2-km-resolution nested grids are used for Istanbul and Antalya cities. DMİ obtains the initial conditions for Turkey from the larger scale data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). This data is based on a global solution for atmospheric computations performed using a grid resolution of 20 km. DMİ then employs MM5 for weather prediction for Turkey at higher grid resolutions. DMİ currently aims at gaining an ability for data assimilation of their local observational data into the ECMWF's numerical flowfields.

In this study, the 3DVAR based open-source data assimilation software, WRF-VAR, is first integrated to MM5 to run correctly with meteorological data taken in the observation stations located over Turkey. The 3DVAR method provides a reasonable data assimilation at the expense of high computation duration including pre-processing. In order to achieve data assimilation processes which do not require long computation time, a new method is developed as another objective of this study. The new data assimilation method is based on the Proper Orthogonal Decomposition (POD).

PROPER ORTHOGONAL DECOMPOSITION

The POD method gives the empirical modes representing the coherent structures of data obtained in experimental or numerical simulations[6]. Once the POD modes are determined, an approximate reconstruction of the original simulation, $\vec{x}$, is possible by the truncated summation of the modes with proper coefficients:

$$\vec{x} \approx w_1\vec{\phi}_1 + w_2\vec{\phi}_2 + \cdots + w_i\vec{\phi}_i + \cdots + w_n\vec{\phi}_n \tag{1}$$

where $\vec{\phi}_i$ is a POD mode, and w_i is the coefficient of the corresponding mode. The coefficients, w_i's, are determined so that $\vec{x}$, is as close as a specified distribution. The truncation order, n, is often "surprisingly" much less than the dimension of $\vec{x}$[6]. The POD theory and the evaluation of the POD modes are explained in detail in various studies[6, 11, 3, 4]. The summary from Reference [6] is as follows.

Using an arbitrary orthonormal basis, $\Phi = [\vec{\phi}_1, \vec{\phi}_2, \cdots, \vec{\phi}_m]$, an m-dimensional column vector, $\vec{x}$, can be expressed as

$$\vec{x} = w_1\vec{\phi}_1 + w_2\vec{\phi}_2 + \cdots + w_m\vec{\phi}_m \tag{2}$$

The aim of the POD analysis is to find the orthonormal basis, Φ, which provides $w_i \approx 0$, $i = n+1, n+2, \cdots, m-1, m$ with the minimum n value. Therefore, the solution of the following optimization problem is considered for the determination of the POD modes.

$$\begin{aligned} &\min_{\phi_i} \; E\{\|\vec{x} - \vec{\chi}(n)\|\} \\ &s.t. \quad \vec{\phi}_i \cdot \vec{\phi}_j = \delta_{ij} \qquad i, j = 1, 2, \cdots, m \end{aligned} \tag{3}$$

where $\vec{\chi}(n)$ is the approximate expression for $\vec{x}$, using the first n vectors from the orthonormal basis. $E\{\}$ is the averaging operator over the m-dimensional vectors. δ_{ij} is the Kronecker's Delta being equal to the unity if $i = j$, and zero otherwise.

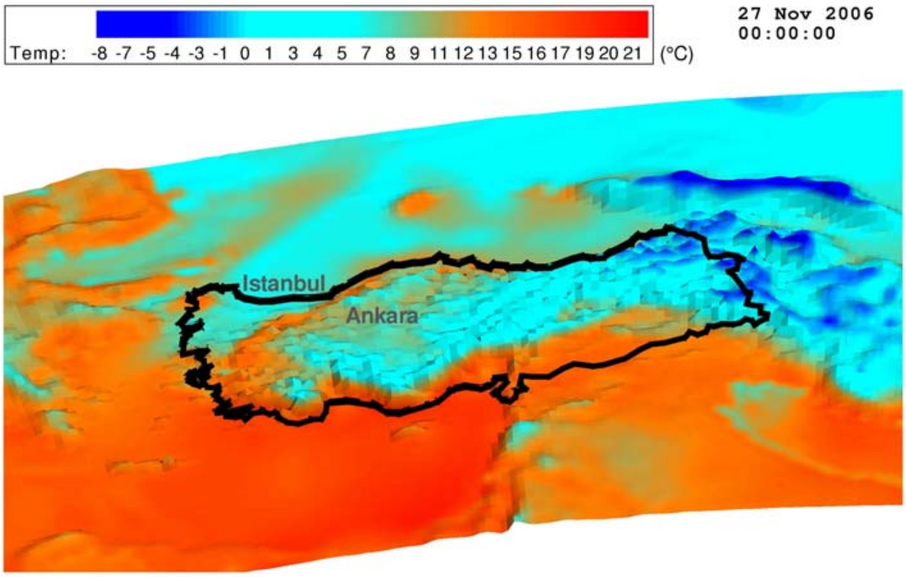

Figure 1: The temperature flowfield over Turkey

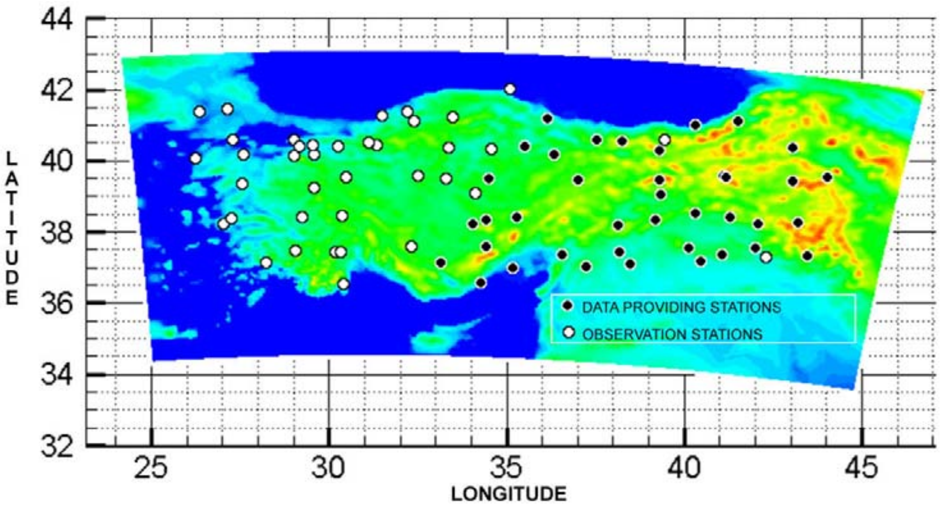

Figure 2: DMİ's observation stations over Turkey

The solution of Eqn.3 is easily obtained if the vectors are limited in a set consisting of a finite number of m-dimensional vectors. Let this set be a matrix;

$$X = [\vec{x_1} \quad \vec{x_2} \quad \cdots \quad \vec{x_l}] \tag{4}$$

where l is much less than m. Then, the POD modes which form the orthonormal basis are the eigenvectors of the matrix, $X * X^T$, where $*$ denotes the formal matrix multiplication, and X^T is the transpose of X. The corresponding eigenvalues, $\vec{s}$, in a descending order, is a measure for the importance of the POD modes[3].

POD is successfully applied in most aerodynamics and Computational Fluid Dynamics (CFD) problems for data reconstruction or for filling a set with missing data[3]. The idea of data assimilation based on POD is a recent research topic still waiting to be developed and to be applied[1].

RESULTS

As an initial study, a sample flowfield over Turkey is computed using the initial and boundary conditions provided by DMİ, and the performance of computations is assessed. The computed instantaneous temperature flowfield over Turkey is given in Figure 1. The computation for a 24-hour weather forecast takes about 140 minutes using 8 processors in parallel. If 16 processors are used, the computation takes about 95 minutes. It drops

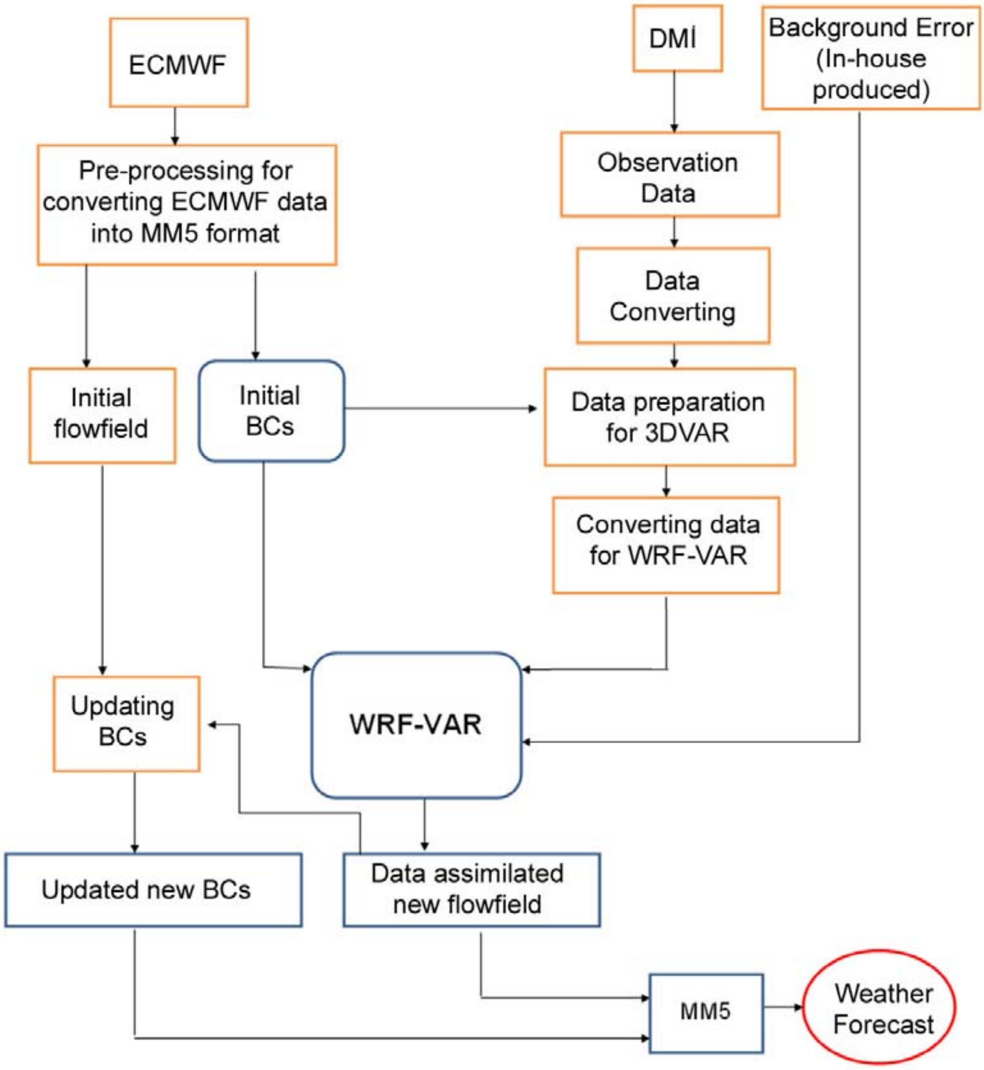

Figure 3: Data assimilation algorithm for the 3DVAR based WRF-VAR program

to 65 minutes for 24 processors. The same computation in DMİ takes 80 minutes on their high performance computer; IBM P690 with 16 processors running AIX5000 operating system.

3DVAR Based Data Assimilation using WRF-VAR

In the second stage of this study, the WRF-VAR program is adapted for the MM5 flow solutions over Turkey, and it is used for the 3DVAR based data assimilation. The observation data is provided by DMİ. DMİ's observation stations over Turkey is shown in Figure 2. As seen from the figure, only the observation data in Central and Eastern Anatolia are provided, and used for assimilating into the whole flowfield over Turkey. The observation data is assumed to be measured at ground stations.

In addition to the observation data, the WRF-VAR program requires an initial guess for the flowfield, the corresponding boundary conditions and the background error information. The background error is an accuracy measure of the numerical model, here the MM5 program, over the computational domain[5]. Parrish and Derber[8], Daley[9] and Thiebaux and Pedder[10] study the background error in detail. The background error for the computational domain used in this study is produced using the MM5 flow solutions from 1st March 2007 to 30th March 2007. The initial guess for the flowfield over Turkey and the corresponding boundary conditions are taken from ECMWF (European Centre for Medium-Range Weather Forecasts) via DMİ. Figure 3 summarizes the algorithm followed for the data assimilation process using WRF-VAR.

Observation data recorded on 31st March 2007 00:00:00 is used for the data assimilation by WRF-VAR. The temperature difference between the initial and data assimilated flowfields is given in Figure 4. The effect of the data assimilation is clearly seen in Central and Eastern Anatolia where the data providing observation stations are located. It is also seen in the figure that some non-zero temperature difference appears in Western Anatolia. The difference between the assimilated and initial flowfields is also given in terms of the wind velocity components, U and V, in Figure 5. Whereas the pressure difference turns out to be not significant, and is not presented.

POD Based Data Assimilation

For a POD based data assimilation, a 24-hour MM5 solution from 31st March 2007 00:00:00 to 1st April 2007 00:00:00 is first computed using the initial and the boundary conditions provided by ECMWF via DMİ. Then, the instantaneous flowfields computed at every half an hour are used to assemble Eqn. 4 to obtain the POD modes. That is, a total of 48 instantaneous flowfields at the ground level obtained from the unsteady flow solution are used. Each flowfield data set consists of 135x258=34830 nodal values for each flow variable, the temperature, the components of the wind velocity and the pressure, respectively: T, U, V, P.

The average temperature, pressure and wind velocity flowfields over the solution domain is given in Figure 6 for the time interval considered. As seen from the figure, the average temperature in Southern Turkey is 15-20 degrees higher than Eastern Turkey. In fact, it is observed that the elevation distribution over Turkey is almost inversely proportional to the temperature distribution. The similar behaviour is observed for the variation of pressure. A wind structure from south-west to north-east direction is also observed in the figure.

The POD modes may be considered as the deviations from the average flowfield. They give the coherent structures of the flowfields in an order of significance. The order and its significance are given by the magnitudes of the eigenvalues, $\vec{s}$. The 48 eigenvalues are given in Figure 7 in the descending order. As seen from the figure, the first 10 modes are quite significant and cover almost all the coherent structures in the predicted flowfields. Actually, it can be concluded that by using the first 15 modes, any flowfield data used in the evaluation of POD modes, X, can be reconstructed with 99% accuracy. For a 99.9% accuracy, 29 modes are required.

Figure 8 and 9 show the first two POD modes for the temperature, the pressure and the wind velocity distributions over the solution domain. The positive and negative correlations of the flow variables over the domain is observed in these figures.

The same observation data (31st March 2007 00:00:00) is now used for the POD based data assimilation. The temperature difference between the initial and data assimilated flowfields is given in Figure 10. Figures 11 and 12 show the pressure and velocity differences. Unlike the 3DVAR data assimilation results in which the local data around observation stations are altered significantly, the POD based data assimilation method leads to a variation of flow variables over the whole computational domain. This is expected due to the fact that POD modes reflect the correlation between the flow variables over the whole flowfield. It should be noted that the correlation and the resulting coherent data structures may involve both the terrain properties and the overall conservation laws for the flow variables.

The differences between the assimilated flow variables using the 3DVAR and the POD based methods are shown in Figures 13 and 14 at initial time (31st March 2007 00:00:00) and 3 hours later (31st March 2007 03:00:00). It is observed that the maximum absolute temperature difference between the flowfields assimilated by the 3DVAR and the POD based methods is about 1^oK at the initial time while the difference increases to about 1.5^oK after 3 hours. However, the overall effect over the whole domain diminishes in time. The same behavior is also observed for the pressure and and velocity distributions. In other words, the unsteady flow solutions obtained using the data assimilated initial conditions based on both the 3DVAR and the POD methods approach to each other.

CONCLUDING REMARKS

A 3DVAR based data assimilation method is first implemented for the MM5 simulations. An open source software, WRF-VAR is successfully used with the observation data taken over Turkey. It is observed that the 3DVAR method basically interpolates the observation data in the close proximity of the observation station and is not computationally efficient. In this study, a new data assimilation method, based on the Proper Orthogonal Decomposition (POD), is proposed and succesfully implemented. The preliminary results obtained show that the POD based data assimilation method has different characteristics than a 3DVAR based method. Further validation studies are in progress in order to establish the strengths and weaknesses of the method developed.

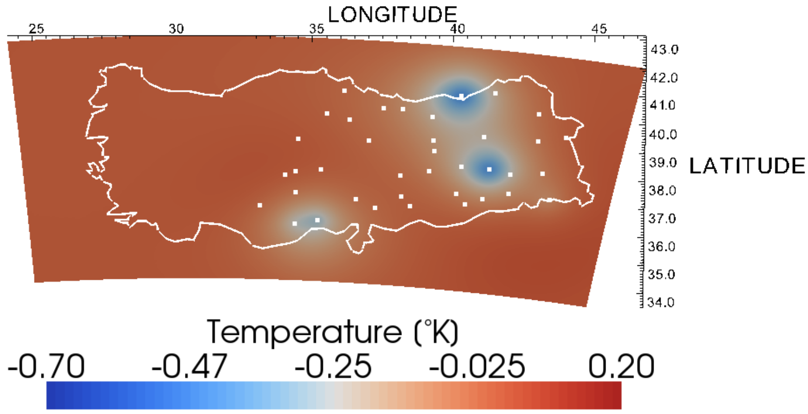

Figure 4: Temperature difference between the initial and the 3DVAR based data assimilated flowfields

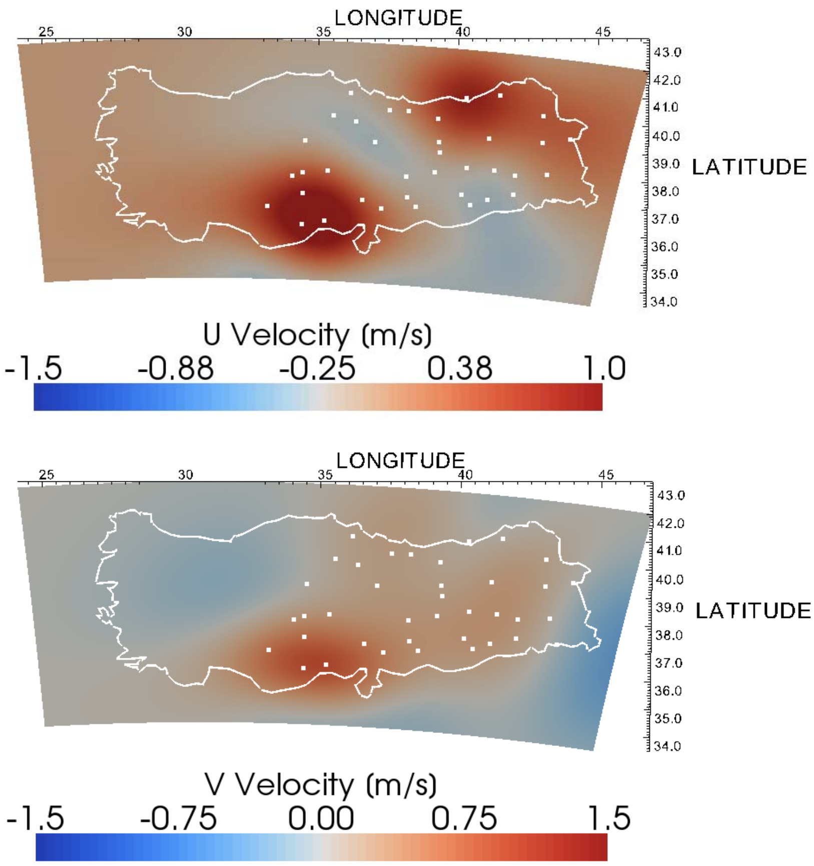

Figure 5: U and V velocities differences between the initial and the 3DVAR based data assimilated flowfields

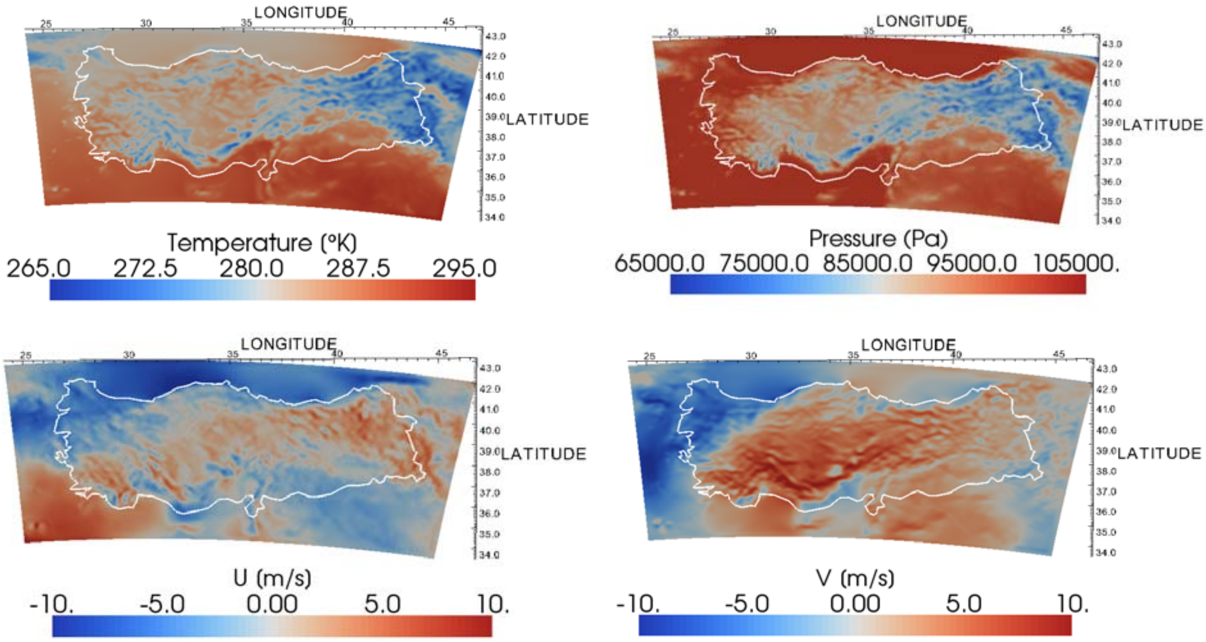

Figure 6: Average flowfields

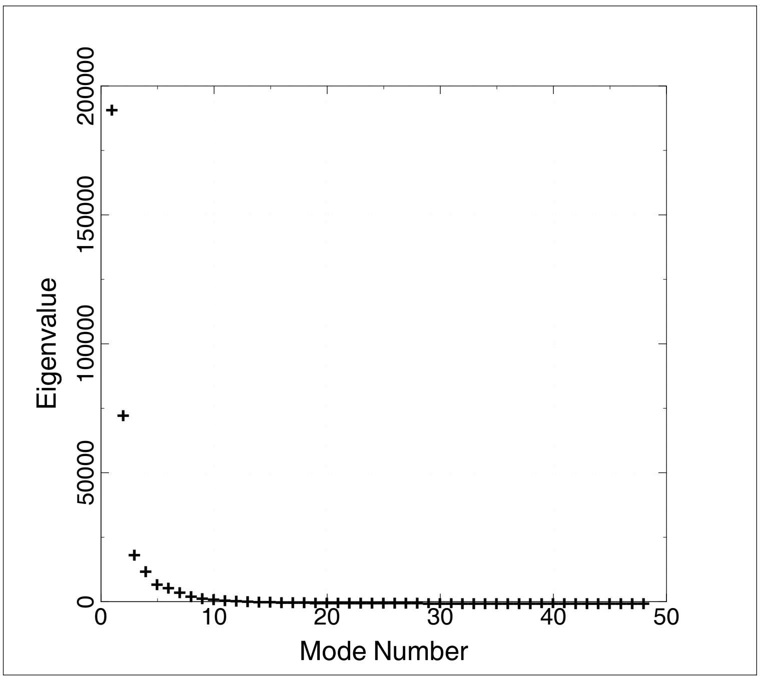

Figure 7: Eigenvalue magnitudes for the POD modes

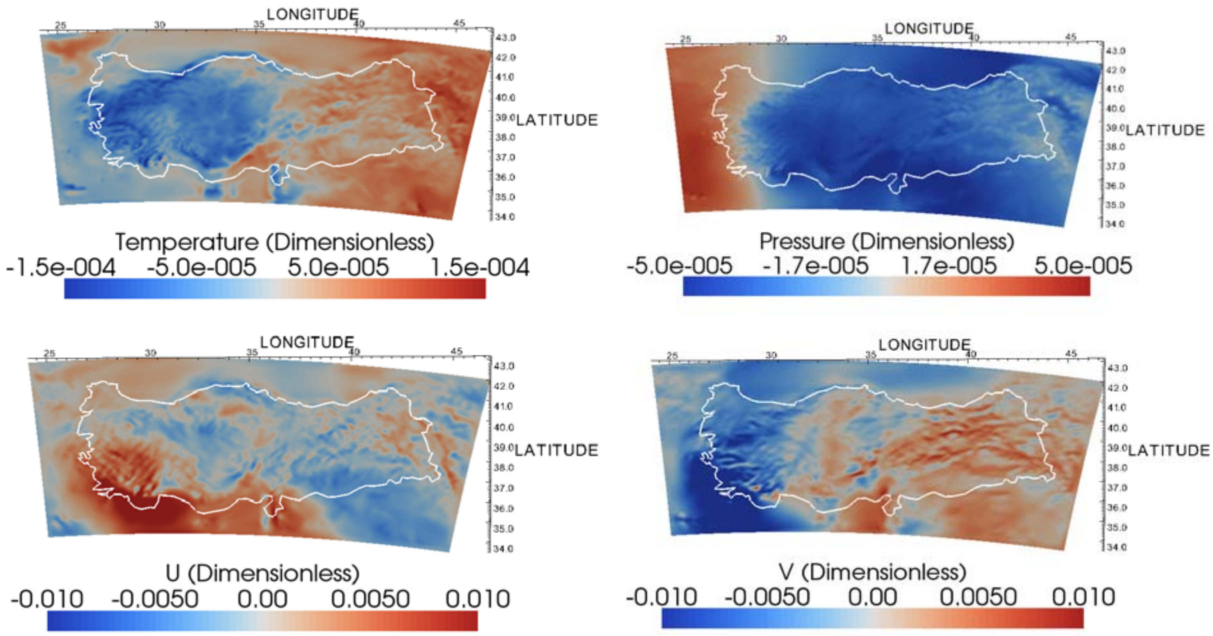

Figure 8: The first POD modes for flowfields

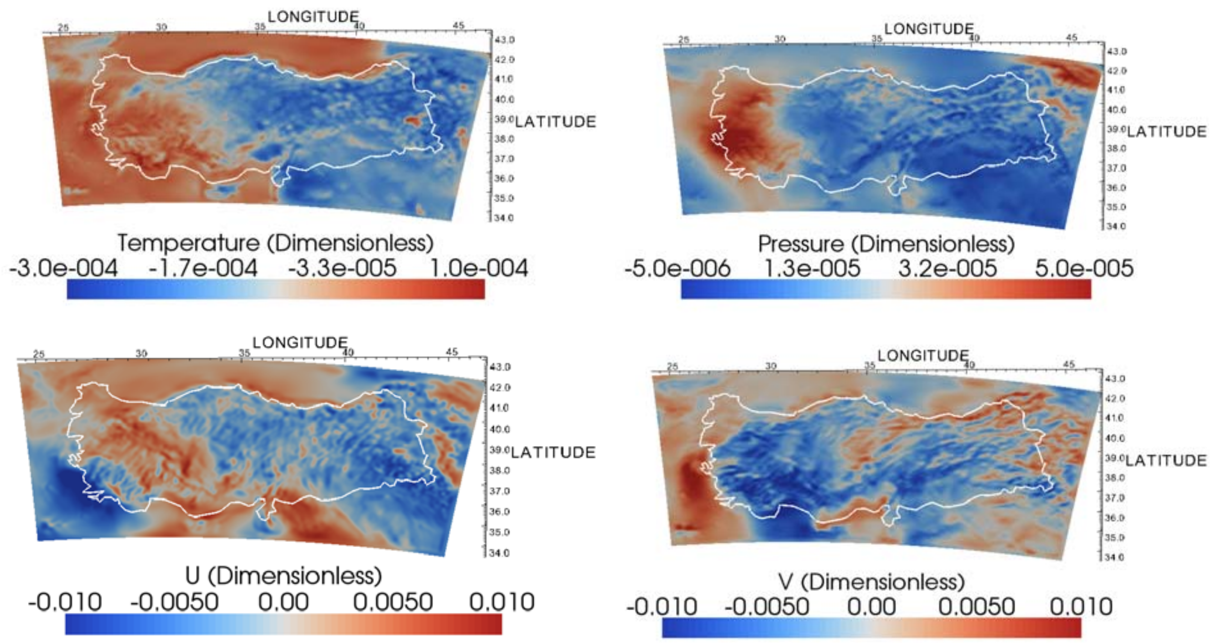

Figure 9: The second POD modes for flowfields

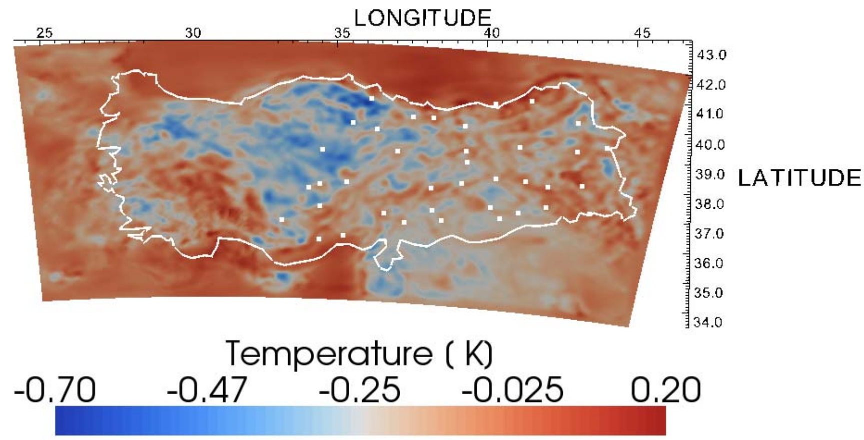

Figure 10: Temperature difference between the initial and POD based data assimilated flowfields

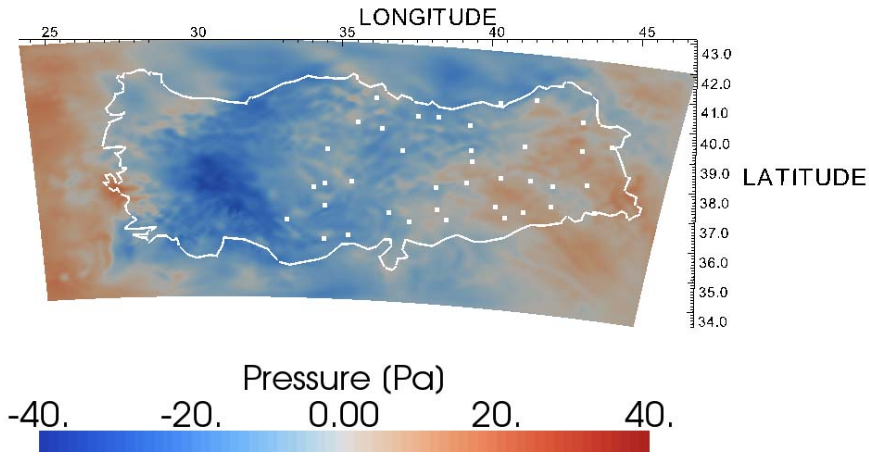

Figure 11: Pressure difference between the initial and POD based data assimilated flowfields

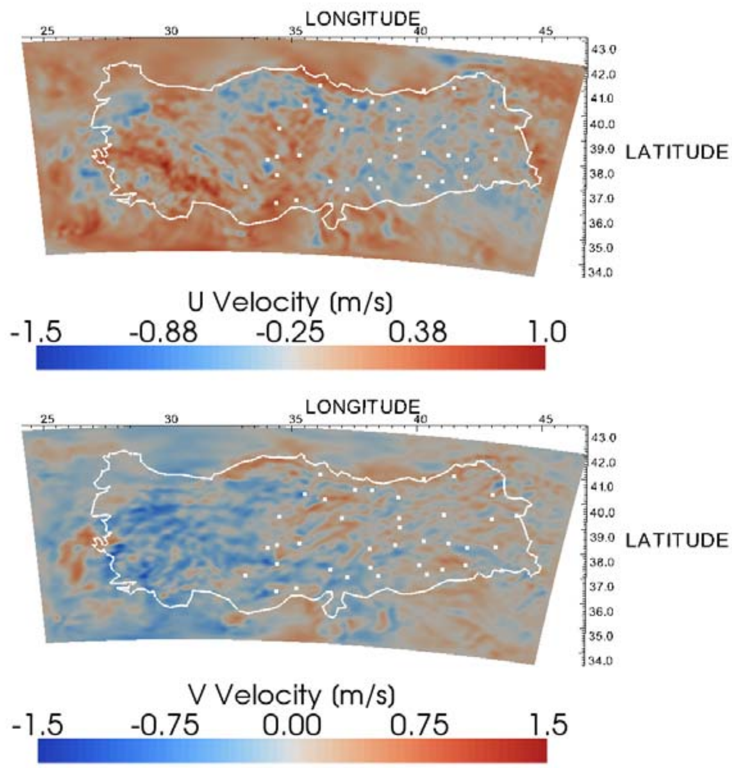

Figure 12: U and V velocities difference between the initial and POD based data assimilated flowfields

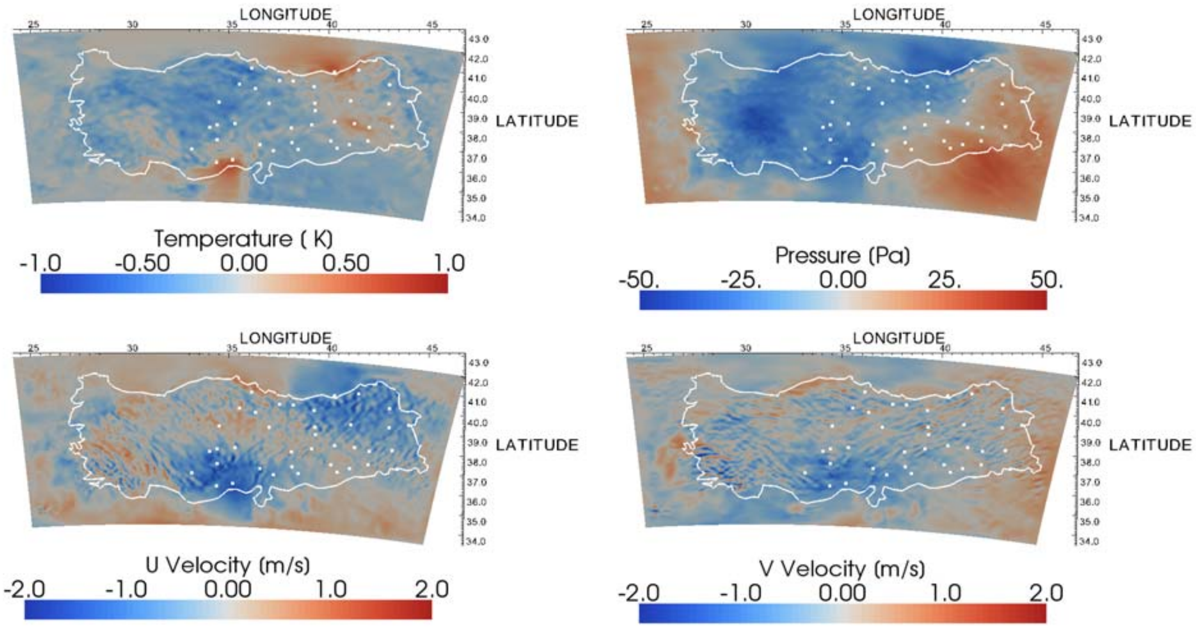

Figure 13: The difference between the POD and the 3DVAR based data assimilated flowfields at the initial time

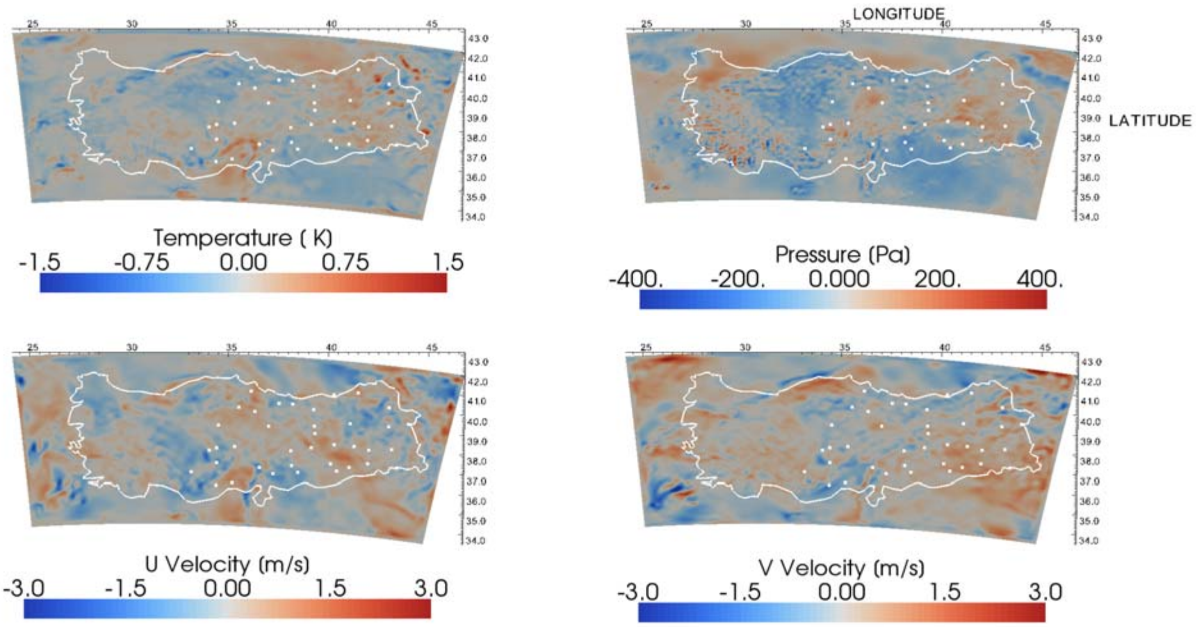

Figure 14: The difference between the POD and the 3DVAR based data assimilated flowfields 3 hours later

References

[1] Qiu, C. and Chou, J., *Four-dimensional data assimilation method based on SVD: Theoretical aspect*, Theor. Appl. Climatol. Vol. 83, 2006, p: 51-57.

[2] Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research, *PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and User's Guide*, NCAR, Penn State, Jan. 2005.

[3] Bui-Thanh, T., Damodaran, M. and Willcox, K., *Aerodynamic Data Reconstruction and Inverse Design Using Proper Orthogonal Decomposition*, AIAA Journal, Vol. 42, No:8, August 2004, p: 1505-1516.

[4] Lucia, D.J. and Beran, P.S., *Reduced-Order Model Development Using Proper Orthogonal Decomposition and Volterra Theory*, AIAA Journal, Vol. 42, No:6, June 2004, p: 1182-1190.

[5] Kalnay, E., *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge Univ. Pres, 2003.

[6] Liang, Y.C., Lee, H.P., Lim, S.P., Lin, W.Z., Lee, K.H. and Wu, C.G., *Proper Orthogonal Decomposition and Its Applications - Part I: Theory*, Journal of Sound and Vibration, Vol. 252, No:3, 2002, p:527-544.

[7] Chai, T., *Four-Dimensional Variational Data Assimilation Using Lidar Data*, PhD Thesis, Univ. of Iowa, August 2002, p:26-34.

[8] Parrish, D.F. and Derber, J.D, *The National Meteorological Center Spectral Statistical Interpolation Analysis System*, Monthly Weather Review, Vol. 120, 1992, p:1747-1763.

[9] Daley, R., *Atmospheric Data Analysis*, Cambridge Univ. Pres, 1991.

[10] Thiebaux, H.J. and Pedder, M.A., *Spatial Objective Analysis*, Academic Pres, 1987.

[11] Sirovich, L. and Kirby, M., *Low-Dimensional Procedure for the Characterization of Human Face*, Journal of the Optical Society for America, Vol. 4, No:3, 1987, p:519-524.

[12] Talagrand, O., *A Study of the Dynamics of Four Dimensional Data Assimilation*, Tellus, Vol. 33, 1981, p:43-60.

[13] Hoke, J. And Anthes, R., *The Initialization of Numerical Models by a Dynamic Relaxation Technique*, Montly Weather Review., Vol. 104, 1976, p:1551-1556.

[14] Sasaki, Y., *Some Basic Formalisms in Numerical Variational Analysis*, Montly Weather Review, Vol. 98, 1970, p:875-883.

[15] Gandin, L. S., *Objective Analysis of Meteorological Fields*, Gidrometerologicheskoe Izdatelstvo, Leningrad, 1963.

[16] Kalman, R., *A New Approach to Linear Filetring and Prediction Problems*, Trans. ASME, Ser. D, J. Basic Eng., Vol. 82, 1960, p: 35-45.

[17] Bergthorsson, P. and Ds, B., *Numerical Weather Map Analysis*, Tellus, Vol. 7, 1955, p:329-340.