

ASSIMILATION SYSTEM AT DHMZ: DEVELOPMENT AND FIRST VERIFICATION RESULTS

Asimilacijski sustav u DHMZ-u: Razvoj i prvi rezultati verifikacije

ANTONIO STANEŠIĆ

Croatian Meteorological and Hydrological Service
Grič 3, 10 000 Zagreb, Croatia
antonio.stanesic@cirus.dhz.hr

Prihvaćeno: 3.3.2011. u konačnom obliku: 29.4.2011.

Abstract: In this paper, a description of the setup for a local assimilation system for a limited area model, ALADIN (Aire Limitée Adaptation Dynamique développement InterNational), is given with a comprehensive description of the assimilation techniques used. The assimilation system at DHMZ (Meteorological and Hydrological Service of Croatia) consisted of two parts: the surface assimilation, which was used to change the state of a model land surface variables, and the upper air assimilation, which changed the upper air model fields. The surface assimilation was performed by the optimal interpolation (OI) technique, while the upper air assimilation was conducted using the 3D variational technique (3DVAR). In a previous research study, surface assimilation and upper air assimilation were used independently and were determined to be beneficial to forecast quality. Currently, they can be combined to improve forecast quality for both the low-level and upper-level fields. A basic verification for a period of 10 months was performed for a forecast starting from the initial state given by the assimilation system and the operational forecast. The verification results showed a positive impact of assimilation on forecast for the upper airfields and for screen-level variables.

Key words: assimilation, variational approach, 3Dvar, optimal interpolation, ALADIN

Sažetak: U ovom radu opisana je struktura lokalnog asimilacijskog sustava za model ograničene domene Aire Limitée Adaptation Dynamique développement InterNational (ALADIN) te je dan nešto opširniji pregled korištenih tehnika asimilacije. Asimilacijski sistem u Državnom hidrometeorološkom zavodu (DHMZ) sastoji se od dva dijela: površinske asimilacije koja mijenja stanje tla te visinske asimilacije koja mijenja visinska polja modela. Površinska asimilacija se radi pomoću tehnike optimalne interpolacije (OI) dok se za visinsku asimilaciju koristi 3D varijacijska tehnika (3DVAR). U prethodnim istraživanjima pronađen je pozitivan doprinos kvaliteti prognoze površinske asimilacije i visinske asimilacije kada su one korištene zasebno. U asimilacijskom sustavu na DHMZ-u oba tipa asimilacije korištena su zajedno u cilju dobivanja pozitivnog utjecaja, kako za površinska, tako i za visinska polja. Napravljena je verifikacija za prognozu inicijaliziranu iz asimilacijskog ciklusa i operativnu prognozu na periodu od otprilike 10 mjeseci. Rezultati verifikacije pokazuju pozitivan utjecaj asimilacije na kvalitetu prognoze kako prizemnih tako i visinskih polja.

Ključne riječi: asimilacija, varijacijski pristup, 3Dvar, optimalna interpolacija, ALADIN

1. INTRODUCTION

Numerical weather prediction (NWP) can be seen as an initial problem in mathematics where, if the initial state of the atmosphere at a given time is known, geophysical systems equations can be solved to obtain values for variables at future time points (Bjerknes, 1904). Sensitivity to initial conditions can be

even greater if the nonlinearity of a geophysical system is taken into account, which requires “best possible” initial conditions for the NWP model. Observations are usually used to obtain the “best possible” initial state of the atmosphere as a source of information. However, observations are scattered in space and

intermittent in time; thus, the background state (i.e., climatology and short-range forecast) is used as a source of additional information. Combining both types of information allows the “true” state of the atmosphere at a given time to be approximated. This process is called analysis. One definition of data assimilation was given by Talagrand (Talagrand, 1967), who stated that data assimilation can be described “as the process through which all available information is used in order to estimate as accurately as possible the state of the atmospheric or oceanic flow.” Data assimilation provides an optimal estimator (analysis), which has a minimum error in the least-square representation; in addition, analysis error should be smaller than the error of any of the given information (i.e., observations and background) used in the assimilation process.

The fundamental equation of linear data analysis (e.g., Bouttier, F. and Courtier, P., 1999) is given by the following:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b]) \quad (1)$$

where \mathbf{x} is state vector, \mathbf{y} is observation vector, \mathbf{K} is gain or weight matrix, and H is non-linear observation operator. Subscript a stands for analysis and subscript b stands for background. Equation 1 represents the analysis as a linear combination of the background and correction that depends on the weight matrix and departures of the background model state from the observations (in observation space). If Best Linear Unbiased Estimator (BLUE) is searched and the assumptions of unbiased background and observational errors, no correlation between background and observation errors and linear observation operator in the vicinity of the background state (\mathbf{H}) are used, then the weight matrix can be expressed as the following:

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (2)$$

where \mathbf{B} and \mathbf{R} represent the covariance matrix of background errors and the covariance matrix of observation errors, respectively. The presented analysis is optimal, i.e., the analysis state is as close as possible to the “true” state in a root-mean-square representation.

In NWP, a practical number of observations (p) is on the order of 10^5 per analysis. Therefore, to calculate the gain matrix, an explicit inversion of the matrix with dimensions ($p \times p$) is required, which is computationally too demanding. Two approaches can be used to tackle this problem. The first approach is called optimal interpolation (OI), and it is based on the assumption that, for each model variable, only a few observations are important for analysis. The BLUE equation is solved by small sub-domains (box-by-box) from the entire model domain. This approach reduces the size of the gain matrix; thus, the explicit inversion can be performed. A problem with the local OI is that, for adjacent points, different observations can be used; thus, spurious noise can appear (i.e., the analysis is not continuous). Noisy fields can also be present for large-scale analysis because all long waves are left out due to the local (box-by-box) computations of BLUE. One more disadvantage of OI is that only observations with simple observation operators can be used; this is because background error correlations between observation points ($\mathbf{H}\mathbf{B}\mathbf{H}^T$) and between model and observation points ($\mathbf{B}\mathbf{H}^T$) are explicitly computed.

The second approach for dealing with problems of large matrix dimensions is the variational approach. In this approach, a quadratic scalar function J (cost function) that measures the distance of the control variable from the background and the observations is defined as follows:

$$J(\mathbf{x}) = J_b + J_o = 0.5 \cdot (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \quad (3) \\ + 0.5 \cdot (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}])$$

where J_b is the background term and J_o is the observation term.

Minimization of the cost function from Equation 3 provides an analysis that is closest in a root-mean-square sense to the “true” state. If background and observation error probability density functions are Gaussian, the analysis is also a maximum likelihood estimator of the “true” state. Lorenz (1986) showed the equivalence between a solution of BLUE analysis and the variational approach. The variational approach provides a global analysis and makes it possible to use observations with

more-complex observation operators. However, there is a problem with the large dimension of \mathbf{B} for practical implementation because its inversion is needed. One solution for this issue will be presented in Section 2.2. More details about different approaches in solving this analysis problem can be found in Hólm (2008).

The data assimilation was first used in global or hemispheric NWP models. Limited area NWP models (LAM) for initialization usually use the global model analysis, interpolated to the finer LAM grid. They provide a dynamical adaptation of large-scale meteorological fields. At Meteorological and Hydrological Service of Croatia (DHMZ), a Aire Limitée Adaptation Dynamique développement International (ALADIN) limited area model (ALADIN International Team, 1997) is installed (ALADIN HR) and is currently running as dynamical adaptation of a global model. In the framework of the ALADIN model, both previously mentioned approaches of solving analysis problem are utilized. At first, optimal interpolation (CANARI) was used for the analysis of upper airfields. Afterwards, the variational approach was developed through three-dimensional variational analysis (3DVAR). The latter approach was tested in the framework of the ALADIN/France model, which revealed an improvement in some aspects of precipitation forecast (Fischer et al., 2005). 3DVAR is also used for the analysis of upper airfields in the Hungarian version of the ALADIN model (ALADIN/HU), where it was found to be beneficial for the forecast of most upper air variables, temperature at 2 m from the surface and precipitation (Bölöni, 2006). These previous examples showed that the upper air assimilation with 3DVAR can be beneficial for forecast with the ALADIN model. The ALADIN forecast could also be improved by using assimilation for screen level variables and using the analysis increments for updating land surface variables. Mahfouf (1991) reported that 2 m analysis increments of temperature and humidity, computed with OI, could be used to update land surface variables. Giard and Bazile (2000) have shown that improvement in the land surface description and the implementation of the assimilation scheme for soil moisture, based on OI, provides a clear improvement in the forecast of low-level fields.

Motivated by these results, an assimilation system was set up at the DHMZ in which surface and upper air assimilation are combined. The implementation of an assimilation system requires significant technical resources and manpower. The assimilation cycle, which is a sequence of 6 h forecasts and analysis as well as production (72 h forecast), needed to be set up. To run the cycle on a daily basis, enough computer power and storage capacity were needed. The observation data used in the assimilation needed to be preprocessed, stored and monitored. Facilities for performing these tasks were not available at DHMZ. The local \mathbf{B} matrix for 3DVAR needed to be calculated, which could be performed in several ways. The right choice had to be made, and some alternatives (e.g., different methods of calculating the \mathbf{B} matrix, calculation of the seasonal \mathbf{B} matrix, and tuning the \mathbf{B} matrix) could be tested. In fact, numerous tests can be made to set up the assimilation system; however, all of them need computer power and storage. Finally, verification must be performed to validate results.

The purpose of this study was to elaborate on a setup of assimilation at the DHMZ in the framework of the ALADIN HR model where both surface assimilation (OI) and upper air assimilation (3DVAR) are combined to get the benefit of assimilation both for screen level and upper air fields.

Also, a more detailed description of the assimilation techniques used in ALADIN is given, and the results of the first basic validation of the forecast initialized by the assimilation system are presented. In the following section, a scheme of the assimilation setup is presented. Section 2 provides more details on the surface and upper air assimilation in ALADIN. The results of the verification are shown and discussed in Section 3.

2. LOCAL IMPLEMENTATION OF THE ASSIMILATION SYSTEM IN THE ALADIN HR MODEL

ALADIN HR is an operational local setup of the mesoscale limited-area model ALADIN. It is a hydrostatic model with a horizontal grid spacing of 8 km and 37 vertical model levels. Details about the setup of the ALADIN HR model can be found in Ivatek-Šahdan and Tudor (2004), or details on a more recent setup

can be found in Tudor and Ivatek-Šahdan (2010). Operationally, initial and boundary conditions are taken from the global model ARPÉGE and interpolated to the ALADIN HR grid (dynamical adaptation). To obtain better initial conditions, data assimilation can be used. To implement data assimilation, first the assimilation cycle needs to be set up. The assimilation cycle is a sequence of analysis and 6 h forecasts that is run on a regular basis. In an assimilation cycle, information coming from observations is accumulated into the model state. A surface analysis assimilation cycle is even more important because land surface need more time to be updated. As mentioned in the introduction, appropriate facilities, like the database of observation, tools for preprocessing raw data (e.g., satellites), are not present at DHMZ. However, DHMZ is part of the Regional Cooperation for Limited Area modeling in Central Europe (RC LACE; <http://www.rlace.eu/>), and the LACE common observation preprocessing unit (OPLACE) is available for use. There, observation data are collected, preprocessed and disseminated to LACE countries. Locally, geographical selection of data and quality control is performed. In addition, the LACE observation monitoring tool is provided for local installation. Because of the lack of computer resources, the assimilation cycle and production are run in quasi-operational mode, i.e., observational data are taken at the opera-

tional time, but analysis and model integration is done with some time delay (i.e., after the end of the operational model run).

Scheme of local setup of assimilation cycle at DHMZ is shown in Fig. 1.

The assimilation cycle consists of several steps. In the first step (BLEND SUR), a 6 h forecast from a previous assimilation cycle is taken, and the sea surface temperature SST is replaced with the SST coming from the long cut off analysis of the ARPÉGE model (the ARPÉGE model is run later, whereas in the assimilation, all available data are used). This is done because SST is not locally assimilated. In the second step, surface analysis is performed, during which temperature and relative humidity at 2 m are used for updating land surface variables (Section 2.1). In next step, the upper airfields are analyzed (Section 2.2), and the output is used for initiating the 6 hour forecast at the end of the assimilation cycle. The assimilation cycle was run with a time delay sufficient to enable the use of the ARPÉGE long cut off coupling files as the boundary conditions for the short range (6 h) forecast. Because the timing of assimilation cycle and production was quasi-operational, long cut off ARPÉGE files and long cut off data were not available for production from the assimilation cycle; thus, short cut off ARPÉGE files and data were used. Steps in production were the same as in the cycle; the

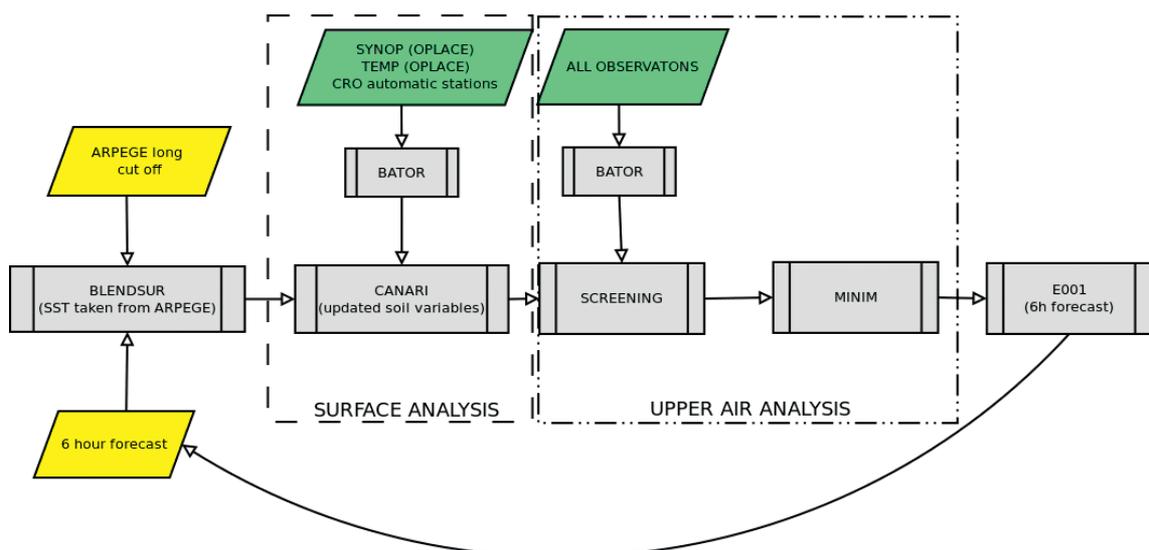


Figure 1. Scheme of the assimilation cycle implemented at DHMZ.

Slika 1. Shematski prikaz asimilacijskog ciklusa implementiranog u DHMZ-u.

only difference was that, at the end, the 72 h forecast was done. A digital filter initialization (DFI) is used for both the cycle and production before the integration of the model.

2.1 Surface assimilation

In the current operational setup of ALADIN HR, land surface variables are obtained by the interpolation of land surface variables from the global model. In the assimilation cycle, a surface assimilation was performed to change the state of the land surface in accordance with available observations. As a background, a 6 hour forecast with an updated SST was used (Figure 1). Data used in the analysis came from synoptic stations and upper air soundings. The quality control of the data for the surface analysis was performed via software named CANARI. It also solved the

BLUE analysis using the OI approach. CANARI was utilized at DHMZ to perform a mono-variate analysis of the boundary layer fields (i.e., 2 m relative humidity and temperature) to initialize the land surface. More about CANARI can be found in Taillefer (2002). Because there were no or very minimal observations of surface fields, the surface temperature and soil water contents could not be directly analyzed. CANARI software used the approach where analysis increments (the difference between the analysis and the background) that change the state of the land surface are derived from analysis increments of 2 m temperature and relative humidity. This conversion was performed via transfer coefficients. Transfer coefficients are fairly simple for temperature (e.g., linear functions) but are very complex for moisture. From eight prog-

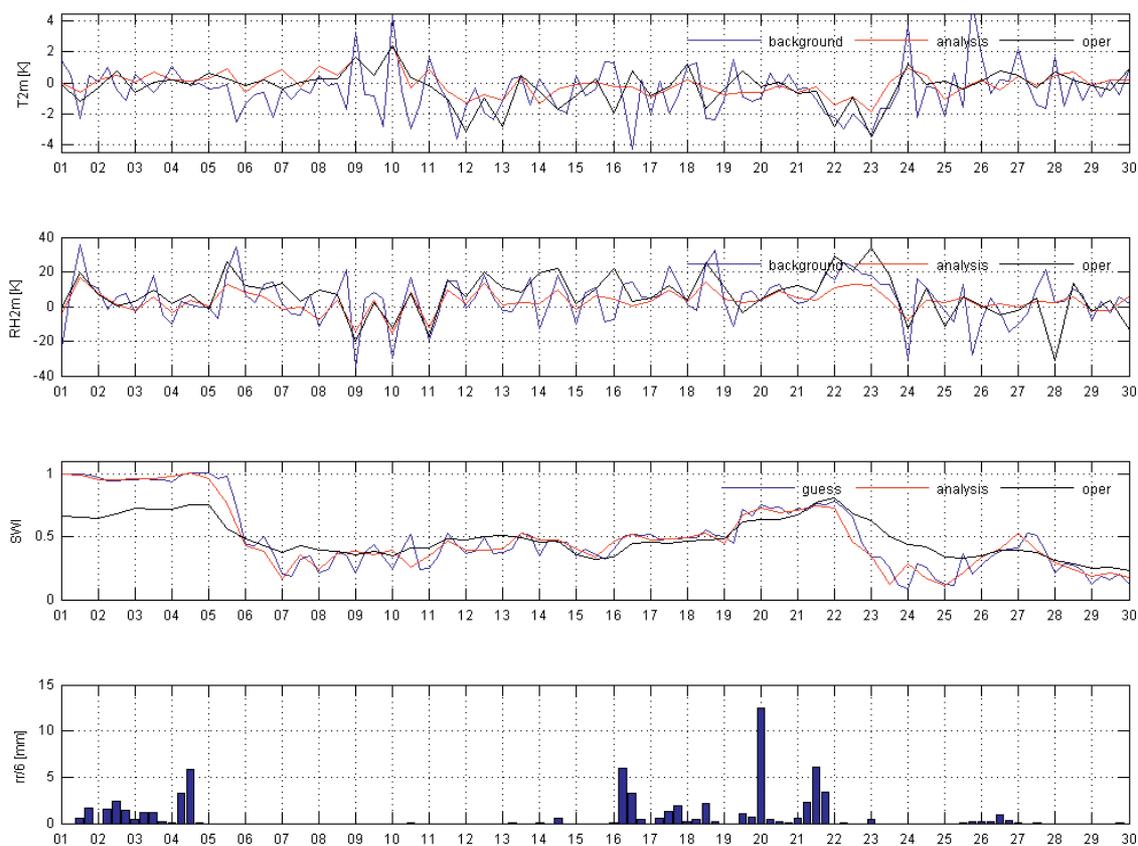


Figure 2. Bias of 2 m temperature (first row), 2 m relative humidity bias (second row) and SWI (third row) in June 2010 for 6 h forecast from assimilation cycle (blue)- every 6 h, analysis (red)-every 12 h, initial conditions for operational integration (black)-every 12 h. Last row: 6 h accumulated rain from 6 h forecast in assimilation cycle.

Slika 2. Srednje odstupanje temperature na 2m (prvi redak), srednje odstupanje relativne vlažnosti na 2m (drugi redak) i SWI (treći redak) u lipnju 2010. za: šestosatnu prognozu iz asimilacijskog ciklusa (plava) – svakih 6 sati, analizu (crvena) – svakih 12 sati, inicijalne uvjete za operativnu integraciju (crna) – svakih 12 sati.

nostic variables in the operational ISBA land surface scheme (Noilhan and Planton, 1989; Noilhan and Mahfouf, 1996; Giard and Bazile, 2000), four variables were updated in surface analysis procedure: the surface temperature (T_s), the mean surface temperature (T_p), the superficial water content (W_s) and the total water content (W_p). Readjustment of the model background was conducted according to the following expressions:

$$\begin{aligned}\Delta T_s &= \alpha_{T_s}^T \Delta T_{2m} \\ \Delta T_p &= \alpha_{T_p}^T \Delta T_{2m} \\ \Delta W_s &= \alpha_{W_s}^T \Delta T_{2m} + \alpha_{W_s}^{RH} \Delta RH_{2m} \\ \Delta W_p &= \alpha_{W_p}^T \Delta T_{2m} + \alpha_{W_p}^{RH} \Delta RH_{2m}\end{aligned}\quad (4)$$

where Δx stands for increment and α_x represents transfer coefficients. The effect of surface assimilation can be seen on Figure 2 (third row), where the evolution of soil wetness index (SWI) for the Zagreb Maksimir station and for the month of June 2010 is shown. The following relation defines SWI:

$$SWI = (W_p - W_{wilt}) \cdot (W_{fc} - W_{wilt})^{-1} \quad (5)$$

where it provides a fractional value between the wilting point (W_{wilt}) and field capacity (W_{fc}). The SWI will have value 1 when field capacity is reached (wet soil), and a value close to zero means that vegetation is unable to extract water from the root zone to the stomatal cells (dry soil).

The SWI of *analysis* and *background* are similar, but the SWI of *oper* is quite different in some periods. Comparing the *background* and *analysis*, it can be seen that, when a 2 m bias of background temperature is positive and the bias of background relative humidity is negative, SWI increases after analysis (e.g., 00 UTC 09. June). When the 2 m bias of background temperature is negative and the background humidity bias is positive; thus, the SWI decreases after analysis (e.g., 12 UTC 10. June). In addition, the SWI changes due to precipitation events (e.g., 01-05 June). Looking at the 00 UTC analysis and 12 UTC analysis separately, the *background* for 00UTC has a primarily positive bias for the 2 m temperature (negative for 2 m relative humidity) and 12 UTC *background* has a primarily negative

bias for the 2 m temperature (positive for 2 m relative humidity). These results were for one point in the domain; however, the characteristics remained similar for the domain average (Figure 3). The transfer coefficient for changing the analysis increment of 2 m temperature into an increment of surface temperature is empirically fixed and has a value 1. Thus, the increments of surface temperature were the same as increments obtained from analysis of the 2 m temperature. On Figure 3 a distinction between increments for 00 UTC analysis and increments for 12 UTC analysis was made to point out a clearly positive bias for the 6 hour forecast of the 2 m temperature at 00 UTC and negative bias at 12 UTC. For the mean soil temperature, the transfer coefficient has a value of $(2\pi)^{-1}$ but relaxation towards climatology is also used; thus, the relationship with the 2 m temperature increment was not so obvious. The transfer coefficients for relative humidity are much more complicated. They were obtained from a set of single-point simulations (Mahfouf, 1991) and afterwards were slightly reformulated for operational implementation (Bouttier et al., 1993 a, b). The relative humidity transfer coefficients depended on both the 2 m temperature and the 2 m relative humidity; thus, similar and straightforward conclusions as for the 2 m temperature could not be made. However, the correlation between total layer reservoir increments and 2 m temperature increments is noticeable in Figure 3. The calculations showed that the correlation coefficient between analysis increments of the surface temperature (i.e., 2 m temperature increments) and the total layer reservoir analysis increment had values of -0.76 and -0.66 for 00 UTC and 12 UTC, respectively. Part of this correlation also comes from the relative humidity because temperature and relative humidity at 2 m are dependent variables.

Based on these results, some conclusions about the local implementation of surface assimilation can be made. Land surface in the local assimilation cycle was different than in the operational setup, which can be seen from the values of the SWI. Changes in the SWI were smoother in the operational setup than in the assimilation cycle. This was due to the clear bias in the 6 hour forecast of the 2 m temperature and 2 m relative humidity; in addition, the sign of this bias changed depending on the analysis time. These biases need to be ad-

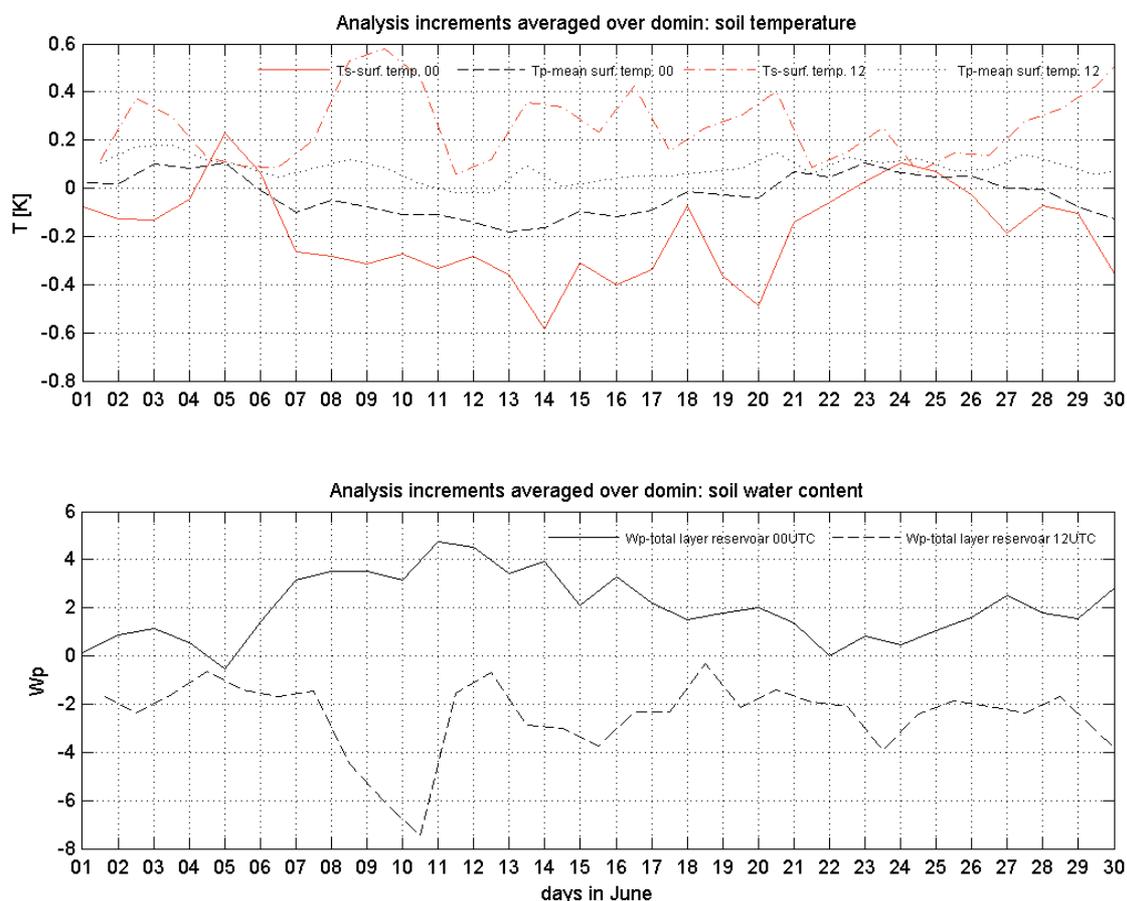


Figure 3. Analysis increments averaged over the domain for land surface variables for June 2010. First row: surface temperatures at 00 (full red line), at 12 UTC (dashed red line), mean surface temperature at 00 (dashed black line) and at 12 UTC (dotted black line). Second row: soil water content at 00 (full black line) and at 12 UTC (dashed black line).

Slika 3. Inkrementi analize parametara tla usrednjeni preko domene za mjesec lipanj 2010. Prvi redak: temperatura površine tla u 00 UTC (puna crvena linija), u 12 UTC (crtkana crvena linija), srednja temperatura površinskog sloja tla u 00 UTC (crtkana crna linija) i u 12 UTC (točkasta crna linija). Drugi redak: sadržaj vode u tlu u 00 UTC (puna crna linija) i u 12 UTC (crtkana crna linija).

dressed in the future because they can lead to unrealistic moistening (drying) of land surface.

2.2. Upper air assimilation

After updating the land surface variables, the upper airfields were analyzed (Figure 1). As mentioned in the introduction, data used for the upper air analysis were obtained from OPLACE. The observation type and variables assimilated at DHMZ are listed in Table 1.

The local preprocessing of data was performed in several steps. The OPLACE data was provided for the whole LACE domain, and the first step was to take a geographical

Table 1. Observation type and variables assimilated at DHMZ.

Tablica 1. Tipovi mjerenja i varijable asimilirane u DHMZ-u.

Observation type	Variable
SYNOP	Ps, T2m, RH2m
Aircraft	u, v
Atmospheric Motion Winds	u, v
TEMP	P, u, v, T, q
Wind profiler	u, v
Satellite radiances	radiance

selection of the data for the ALADIN HR domain (BATOR); afterwards, the data were stored in the ODB database. In the second step, quality control of data (SCREENING) and, for some observation types (e.g., satellite), bias correction were performed. In last step (MINIMIZATION), the upper air analysis was obtained. During all of the steps, the ODB database was updated. Afterwards, the ODB database could be used to monitor the number and type of data assimilated as well as to obtain departure values of the first guess and analysis. This was conducted with a common tool called the LACE observation monitor, which provides a daily overview of observational usage and it can also provide different statistics for some periods. Analysis increments were calculated using 3DVAR. The ALADIN implementation of 3DVAR closely followed the work of the ARPÉGE/IFS model (Courtier et al., 1998). The incremental version of Equation 3 can be written as follows:

$$J(\mathbf{x}) = J_b + J_o = 0.5 \cdot \delta \mathbf{x}^T \mathbf{B}^{-1} \delta \mathbf{x} + 0.5 \cdot (\mathbf{H} \delta \mathbf{x} - \mathbf{d})^T \mathbf{R}^{-1} (\mathbf{H} \delta \mathbf{x} - \mathbf{d}) \quad (6)$$

where $\delta \mathbf{x} = \mathbf{x} - \mathbf{x}_b$ is increment, $\mathbf{d} = \mathbf{y} - H(\mathbf{x}_b)$ represents innovation (i.e., the departure of the background model state from observations), H is the observation operator and \mathbf{H} is the linear approximation of H in the vicinity of the background. The analysis is determined by searching the minimum of the cost function. The gradient of the cost function is obtained by differencing J with respect to $\delta \mathbf{x}$:

$$\nabla_{\delta \mathbf{x}} J = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \delta \mathbf{x} - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{d} \quad (7)$$

The cost function (J) is then minimized using an iterative descent algorithm (quasi Newton method) to obtain the analysis. The problem for the practical implementation of the 3DVAR technique is that the \mathbf{B} matrix dimensions are too large. In order to reduce dimensions of the \mathbf{B} matrix, analysis is not conducted for all components of the model state vector, but only for a smaller number of them. In that way, the analysis is conducted in smaller, control variable space, where corrections to the background are allowed. In ALADIN, the implementation of 3DVAR analysis is per-

formed for the following components of the state vector: vorticity (ξ), divergence (η), temperature (T), logarithm of surface pressure (P_s) and specific humidity (q). However, the \mathbf{B} matrix is still too large to be explicitly inverted. Thus, the change of control variable is made, where the new control variable is defined as follows:

$$\begin{aligned} \delta \mathbf{x} &= L \chi \\ L^T \mathbf{B}^{-1} L &= I \end{aligned} \quad (8)$$

where χ represents the new variable and L is the operator that performs the conversion.

The background cost function and its gradient are then defined as follows:

$$\begin{aligned} J_b &= \chi^T \chi \\ \nabla_{\chi} J_b &= 2\chi \end{aligned} \quad (9)$$

After this transformation, no inversion of the \mathbf{B} matrix is needed, and it is only necessary to make a variable conversion. From equation (8), it can be deduced that $L = \mathbf{B}^{\frac{1}{2}}$. Its inverse $L^{-1} = \mathbf{B}^{-\frac{1}{2}}$ is defined as the sequence of operators that project the model state to the control variable space (χ), where its components are not correlated. Three types of correlations need to be accounted for during this projection: cross-covariance between model variables, horizontal and vertical auto-covariance for each model variable. Thus, the operator L^{-1} can be written as the sequence of operators

$$L^{-1} = \mathbf{B}^{-\frac{1}{2}} = (\mathbf{W}^{-1} \mathbf{V}^{-1})(\mathbf{D}^{-1})(\mathbf{K}^{-1}) \quad (10)$$

where separate steps in projection can be distinguished by parenthesis.

In the framework described by Berre (2000), errors of different analysis variables can be related using multiple linear regressions:

$$\begin{aligned} \zeta &= \zeta \\ \eta &= \mathbf{M} \mathbf{H} \zeta + \eta_u \\ (T, P_s) &= \mathbf{N} \mathbf{H} \zeta + \mathbf{P} \eta_u + (T, P_s)_u \\ q &= \mathbf{Q} \mathbf{H} \zeta + \mathbf{R} \eta_u + \mathbf{S} (T, P_s)_u + q_u \end{aligned} \quad (11)$$

Variable errors are decomposed into two parts: the balanced and unbalanced parts (subscript u). Because linear regression predictors

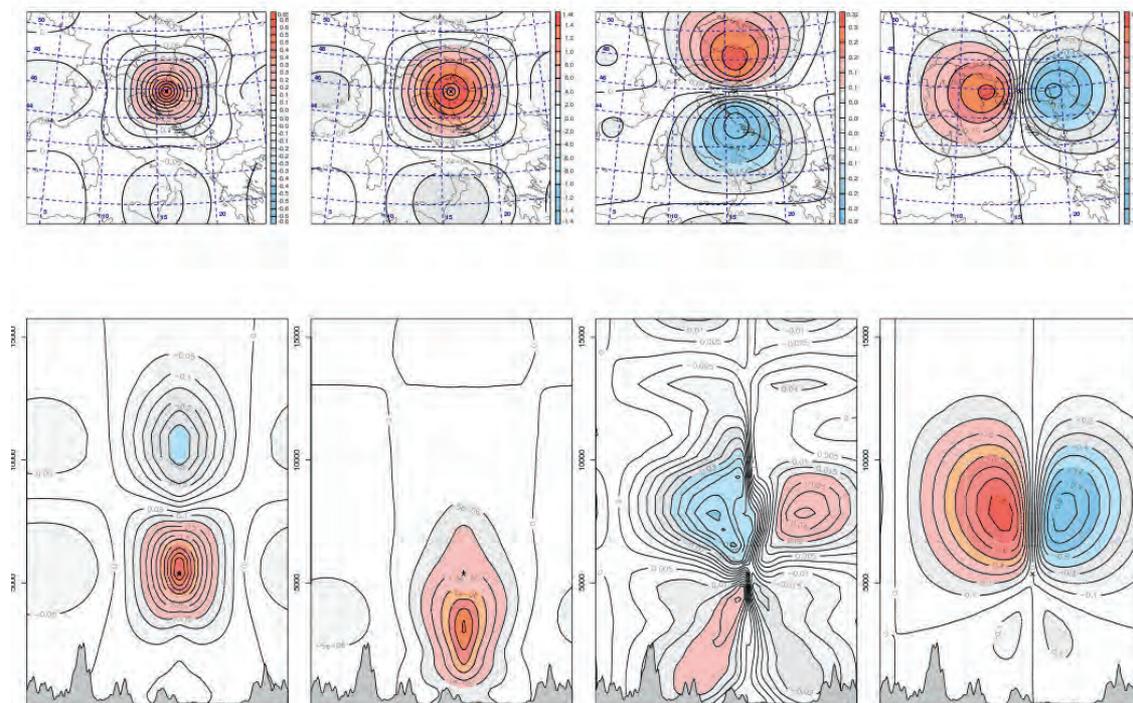


Figure 4. Single observation experiment. Impact of temperature innovation of 1K at 500 hPa at Zagreb. Horizontal analysis increment (first row) and vertical analysis increment (second row) at model level 18 (514 hPa) for temperature, specific humidity, zonal wind component, and meridional wind component.

Slika 4. Eksperiment sa jednim mjerenjem. Utjecaj inovacije temperature od 1K, na 500 hPa, na lokaciji Zagreba. Horizontalni inkrement analize (prvi redak) i vertikalni inkrement analize (drugi redak) na 18. nivou modela (514 hPa) za temperaturu, specifičnu vlažnost, zonalnu komponentu vjetra i meridionalnu komponentu vjetra.

are uncorrelated with residuals if the balanced part of the variable error is taken out, the new variables (unbalanced part) are mutually decorrelated, which is the first step in the transformation (K^{-1}).

Horizontal auto-covariance values are accounted for by making assumptions of horizontal homogeneity and isotropy. Thus, ALADIN spectral modes with different wavenumbers are independent, which makes each auto-covariance matrix block diagonal in spectral space (because of vertical correlations). To be closer to the identity matrix, a division by spectral standard deviations is applied (D^{-1}). Vertical auto-covariance values are accounted for by making projections on the eigenvectors of the vertical auto-correlation matrix of each wavenumber and each variable (V^{-1}). At the end, normalization by square-root of the eigenvalues is applied (W^{-1}).

To approximate unknown error statistics, different methods can be used. At DHMZ, the

standard NMC method (Parrish and Derber, 1992) is used. Error statistics were obtained by taking 100 forecast differences from the ALADIN HR model forecast for period 15.02.2008.-25.05.2008., where the model runs were initialized with a 24 hour time difference and forecasts were valid at the same time (36 h and 12 h forecast). The B matrix is important because it determines a way of spreading analysis increment. This can be seen in Figure 4, where analysis increments are shown for a single observation experiment with temperature innovation of 1K at Zagreb. Although having only temperature observations, balance equations (Eq. 11) between control variables produced analysis increments for all other control variables. Because specific humidity and temperature are control variables they had isotropic and homogenous analysis increment, which came from the assumptions in the B matrix decomposition. This was not the case for wind increments because wind is not a directly control

variable. Spurious noise in the increment at the edges of the domain came from the biperiodicization used in the spectral model ALADIN. One solution for this problem could be to make the E zone (i.e., the extension zone used to make the ALADIN model fields biperiodic) wider. Another way of dealing with this problem would be to limit the length-scale of the increment with compactly supported horizontal correlations (Gaspari and Cohn, 1999). However, none of these solutions were tested at DHMZ.

3. VERIFICATION RESULTS

The data assimilation setup at DHMZ, as described in the previous chapter, was running in quasi-operational mode from the end of February 2010. At approximately the same time, storage capacities were enhanced, which allowed the storing of 72 h forecasts initialized from the assimilation cycle. Stored data were used to evaluate the quality of the forecast initialized using the assimilation system (ASSIM) against the operational forecast (OPER), using a verification package VERAL (<http://old.chmi.cz/meteo/ov/aladin/docs/veral>). This was conducted over a time period of approximately 10 months and over the model

domain. The model results were interpolated to a location of observation and compared to synoptic and atmospheric sounding observations. Quality control of data was performed using CANARI internal data control algorithms. First, the ARPÉGE long cut off analysis was taken as the background to obtain a “neutral” observation selection. Selected observations were used for the computation of model departures from observations for both OPER and ASSIM. Departures were used for calculating some basic statistics like bias (BIAS), root mean square error (RMSE) and standard deviation (STD):

$$BIAS = N^{-1} \sum_{i=1}^N (F_i - O_i)$$

$$RMSE = \left(N^{-1} \sum_{i=1}^N (F_i - O_i)^2 \right)^{0.5} \quad (12)$$

$$STD = (RMSE^2 - BIAS^2)^{0.5}$$

where F_i represents the model value at the observation location, O_i represents the observed value and N is number of measurements.

The period used for verification was 02.03.2010.-04.12.2010., and comparisons with

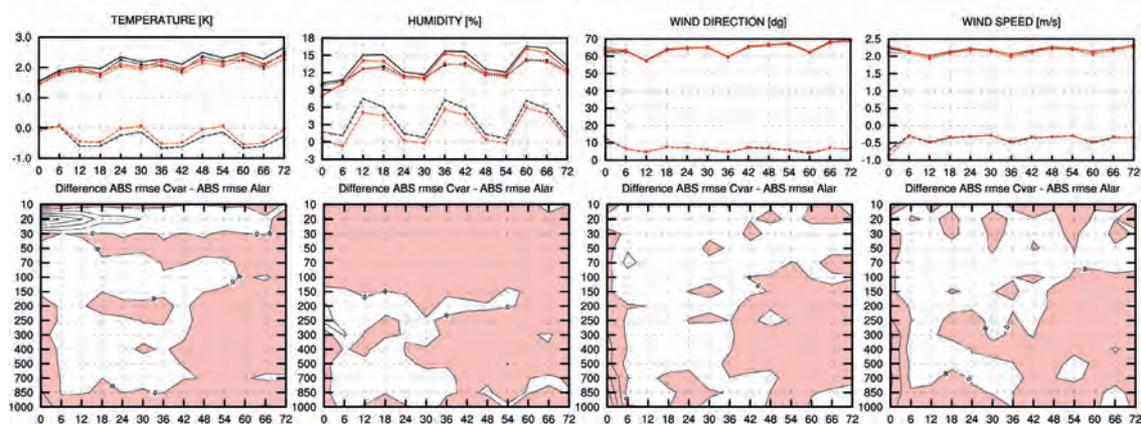


Figure 5. First row: verification scores vs. prognostic hour for screen level fields: temperature, humidity, wind direction and wind speed. BIAS-dashed lines, RMSE-full line, STD-dotted line. Red is ASSIM and Black OPER. Second row: difference of vertical profiles of absolute values of RMSE for ASSIM and absolute values of RMSE for OPER vs. prognostic hour. Red means that RMSE for ASSIM is smaller compared to the OPER.

Slika 5. Prvi redak: rezultati verifikacije u odnosu na prognostički sat za temperaturu na 2m, relativnu vlažnost na 2m, smjer i brzinu vjetra na 10m. Srednje odstupanje – crtkane linije, srednja kvadratna pogreška – puna linija, standardna devijacija – točkaste linije. ASSIM je crveno, a OPER crno. Drugi redak: razlika vertikalnih profila apsolutne vrijednosti srednje kvadratne pogreške za ASSIM i apsolutne vrijednosti srednje kvadratne pogreške za OPER u odnosu na prognostički sat. Crveno znači da je srednja kvadratna pogreška za ASSIM manja od one za OPER.

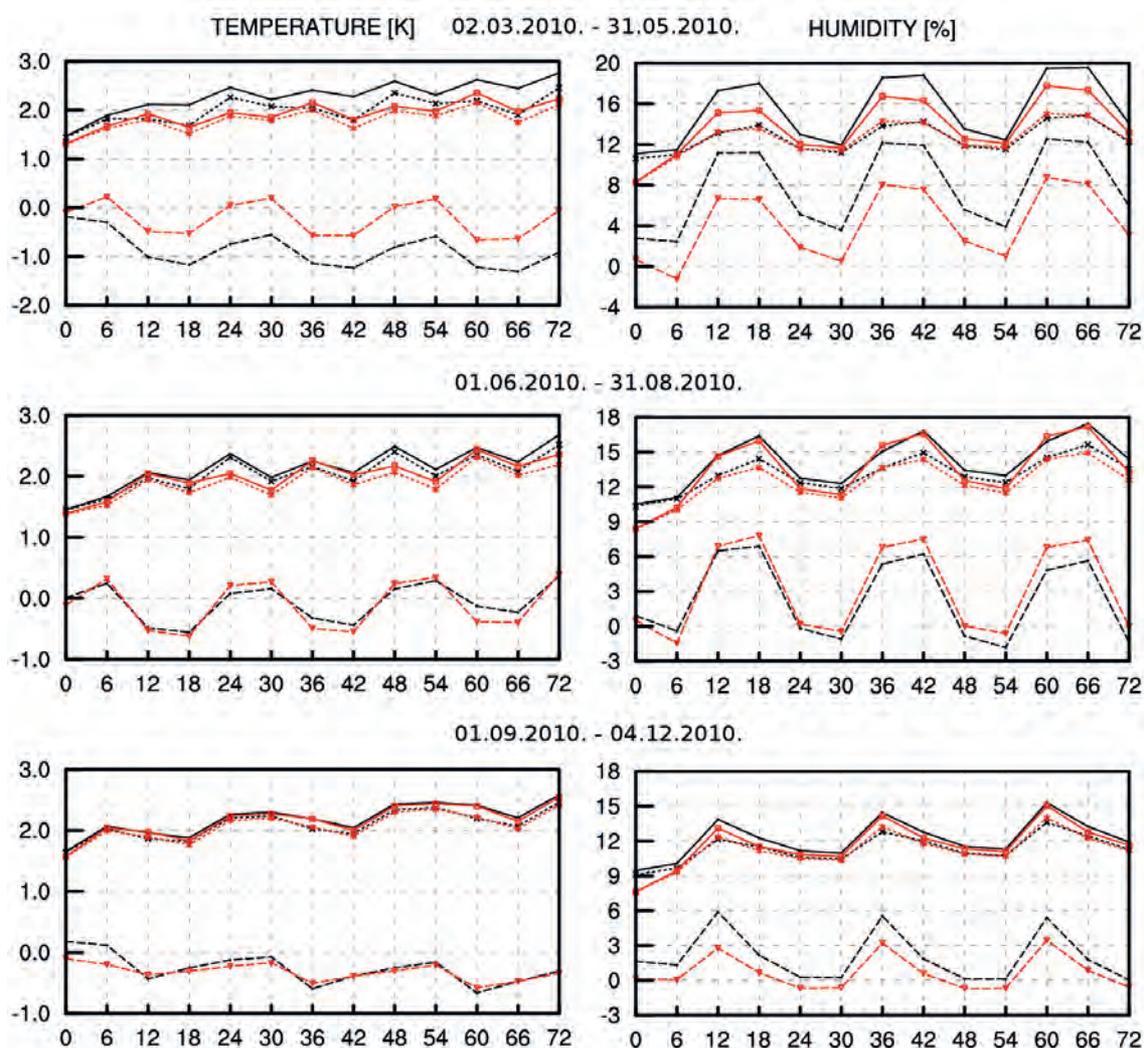


Figure 6 Seasonal verification scores for 2m temperature and humidity vs. prognostic hour. BIAS-dashed lines, RMSE-full line, STD-dotted line. Red is ASSIM and black represents the OPER.

Slika 6. Rezultati verifikacije za temperaturu na 2m i relativnu vlažnost na 2m izračunati po sezonama. Srednje odstupanje – crtkane linije, srednja kvadratna pogreška – puna linija, standardna devijacija – točkaste linije. ASSIM je crveno, a OPER crno.

measurements were taken every 6 hours. Statistics were computed for screen-level variables (i.e., t_2m , rh_2m , u_{10m} , v_{10m} , and geopotential) and for upper-air variables (i.e., t , rh , u , v , geopotential) for ALADIN HR domain. The results for the geopotential showed spurious noise that required further investigation; thus, they were left out of this study. The number of SYNOP observations per analysis time was approximately 700, while the number of TEMP observations per analysis time was up to 30 observations; however, this value can be significantly less.

The results of the verification for the whole period (Figure 5) for 2 m variables show that both the bias and root mean square error for the ASSIM were better for temperature and relative humidity and were mainly neutral for wind speed and direction. Thus, in this period impact of assimilation was positive for the 2 m variables. The results for the upper air showed that, for all variables, RMSE of ASSIM was mainly smaller than the OPER for the first 6 hours, which can be expected. The values of RMSE of ASSIM then became larger than the OPER and finally became smaller after 42

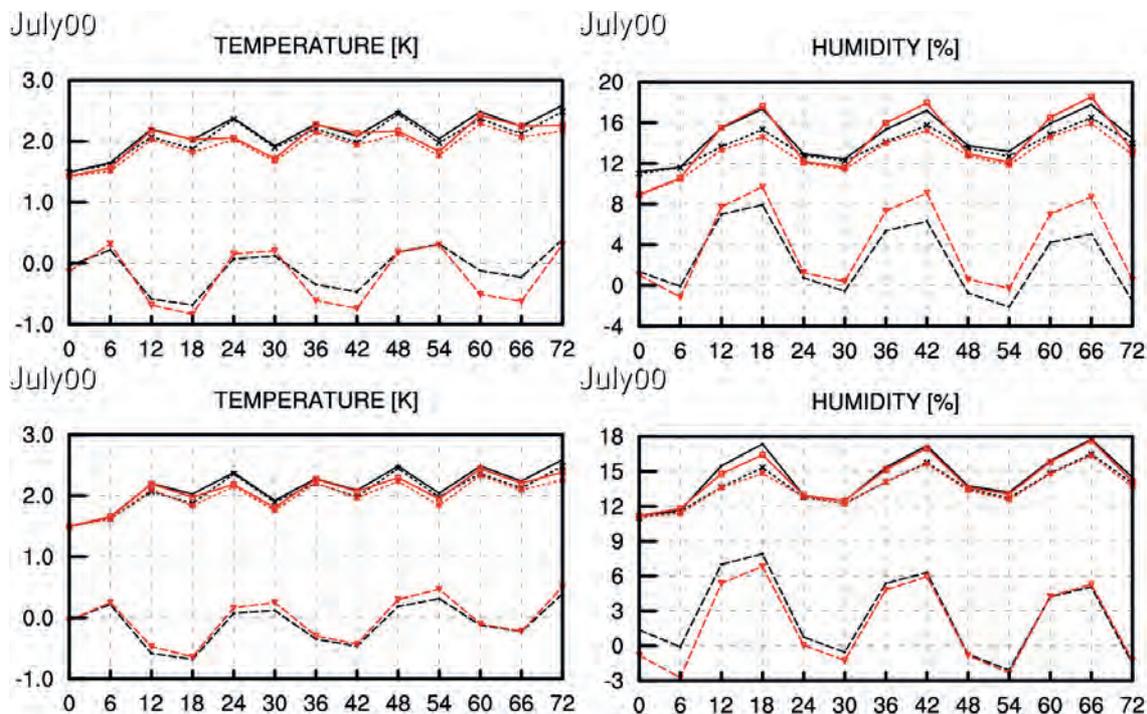


Figure 7. Verification scores for July 2010 vs. prognostic hour for 2 m temperature and 2 m relative humidity. First row: ASSIM (red) and OPER (black). Second row: ASSIM-noCANARI (red), OPER (black). BIAS-dashed lines, RMSE-full line, STD-dotted line.

Slika 7. Rezultati verifikacije za mjesec srpanj 2010. u odnosu na prognostički sat za temperaturu na 2m te relativnu vlažnost na 2m. Prvi redak: ASSIM (crveno) i OPER (crno). Drugi redak: ASSIM-noCANARI (crveno) i OPER (crno). Srednje odstupanje – crtkane linije, srednja kvadratna pogreška – puna linija, standardna devijacija – točkaste linije.

forecast hours. The RMSE of ASSIM for relative humidity was constantly smaller than the OPER above 150 hPa. The same applied for temperature RMSE below 850 hPa. The biggest and most positive impact of assimilation was found for the relative humidity. Although for some forecast hours and levels, OPER has smaller RMSE than ASSIM, the overall impression was that upper-air assimilation is beneficial for the forecast quality. Nevertheless, for upper-air results, it was hard to provide a general conclusion. One reason was that, because of the relatively small number of observations, more noise was present in the results of verification. Additionally, results for the upper-air were calculated for a relatively small number of geographical locations and only small part of domain diversity was covered. At some forecast time points, there was a very small number of atmospheric sounding observations present; thus, the statistical significance was questionable for the results at those times.

To investigate the behavior of forecast scores at different time points of the year, the whole period is divided into three parts: *spring* (02.03.2010.-31.05.2010.), *summer* (01.06.2010.-31.08.2010.) and *autumn* (01.09.2010.-04.12.2010.).

The results for the 2 m variables were rather neutral for wind speed and direction (not shown), whereas a bigger impact was seen with respect to temperature and relative humidity (Figure 6). For *spring*, the BIAS and RMSE of temperature and relative humidity were clearly smaller for ASSIM. Compared to other periods, this was not due to especially better statistics for ASSIM; rather, this was due to bad statistics for the OPER in this period. For *summer* data, the RMSE of temperature was smaller and the RMSE of relative humidity was just slightly smaller for ASSIM. However, the BIAS of temperature was greater after the first day of forecast for ASSIM, and the BIAS for the relative humidity for afternoon hours was also greater for AS-

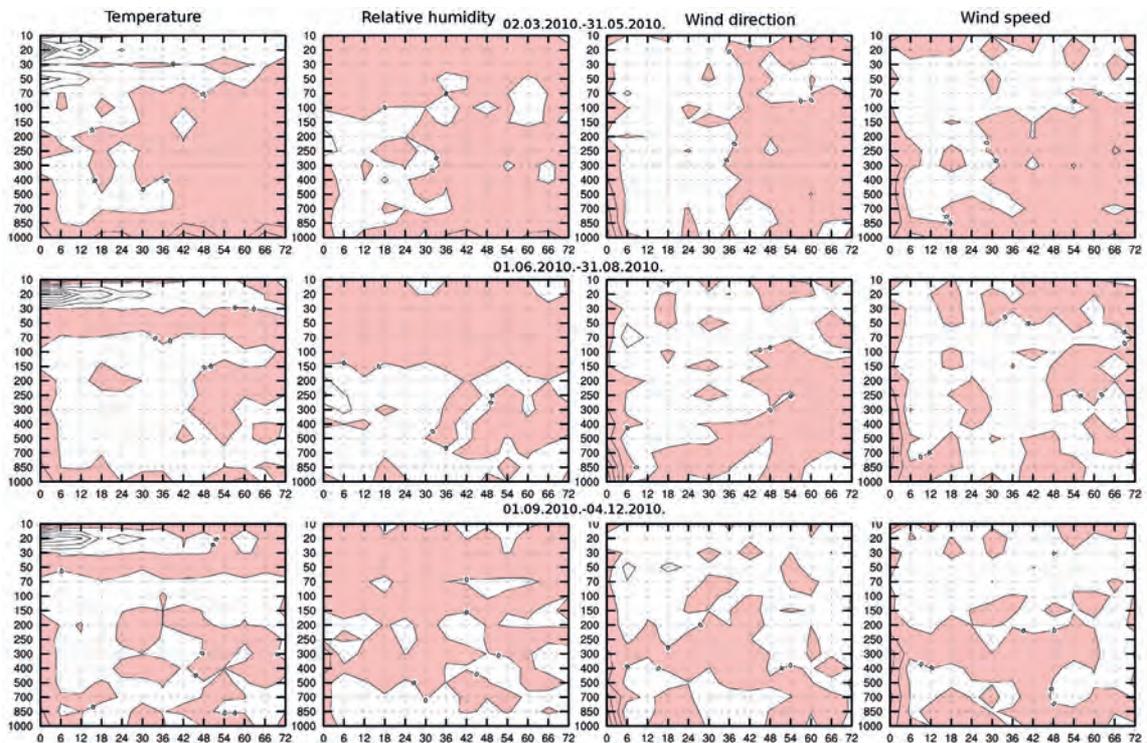


Figure 8. Difference in absolute value of RMSE for ASSIM and the absolute value of RMSE for OPER vs. the prognostic hour computed by seasons. Red means that the RMSE for ASSIM was smaller compared to that for the OPER.

Slika 8. Razlika vertikalnih profila apsolutne vrijednosti srednje kvadratne pogreške za ASSIM i apsolutne vrijednosti srednje kvadratne pogreške za OPER u odnosu na prognostički sat izračunata po sezonama. Crveno znači da je srednja kvadratna pogreška za ASSIM manja od one za OPER.

SIM. The *autumn* results for temperature were almost neutral, while for relative humidity, both the BIAS and RMSE were smaller for ASSIM. The bad results for the BIAS at 2 m for the ASSIM during the *summer* period most probably came from land surface characteristics, which were badly represented in the assimilation cycle for this period. To validate this assumption, an experiment was performed in which the assimilation cycle for July was run without CANARI land surface analysis and a production from this cycle was performed (ASSIM-noCANARI). Verification statistics (Figure 7) were compared with the ASSIM and OPER, and it showed that, for 2 m variables, scores of ASSIM-noCANARI were similar to those of OPER; thus, the scores were better than in ASSIM. Then again, some degradation was present in the upper air statistics (not shown). This “summer” problem is still under investigation.

Upper-air seasonal results (Figure 8) showed similar characteristic compared with results

for the whole period, and some seasonality was noticed, as well, but it was hard to provide a general conclusion on how the results changed throughout the year. In all seasons, the biggest and most positive impact of assimilation was on relative humidity. For other variables, the impact of assimilation was mainly positive for the first 6 hours, while afterwards the results were mixed. The most positive impact of assimilation was in the *spring* period, while the smallest positive impact of assimilation was in the *summer* period.

4. CONCLUSION

Presented above are the first steps towards the operational implementation of an assimilation system at DHMZ. An assimilation system setup is designed to update both upper air fields and land surface variables. This is achieved by combining land surface and upper-air analysis. Verification scores for a period of 10 months showed that the overall impact of the assimilation system on forecast was mainly positive for

screen level variables during all 72 hours of the forecast. For the upper-air fields, impact of assimilation was mainly positive for the first 6 hours, whereas later it was mixed. The impact of assimilation was not constant throughout the year. Further investigation is needed for the summer period to obtain better results with assimilation. An experiment that omitted the CANARI step showed that the BIAS of ASSIM at 2 m for July was most probably caused by a bad definition of land surface characteristics in the assimilation cycle. As seen in Section 2.1, the land surface characteristics are different in OPER and ASSIM due to surface assimilation. In addition, increments of land surface variables during surface assimilation were shown to be dependent on the BIAS of model 2 m temperature and 2 m relative humidity. In both Section 2.1 and Section 3, the afternoon hours showed a clear bias of some ALADIN HR screen level fields. For those hours, the bias was negative for 2 m temperature and positive for the 2 m relative humidity. Model biases can produce unrealistic land surface analysis increments and therefore produce inaccurate estimations of how moist or dry the land surface is. Afterwards, this can lead to false verification results for synoptic situations, when variables at 2 m are dominantly affected by land surface characteristics. Thus, more work is needed to reduce these biases (e.g., monitoring and blacklisting of observations, tuning of analysis, and ALADIN HR model tuning).

The upper air verification results were obtained from a much smaller sample of observations minus model differences. Therefore, further studies are needed to broaden the verification approach and include significance information. Only then can a clear conclusion be made about the benefit of assimilation. The current results showed that there are benefits in using the assimilation system and that it is most prominent in the *spring* period. One of the reasons for the benefits could be that the B matrix was computed for a similar period. Therefore, the seasonal B matrix computation should be considered in future.

Although some benefits of the assimilation system are shown, constant work is needed in maintenance and improvement of the system. The B matrix was never tuned, and there are ways to do a posteriori tuning (Desroziers et

al., 2005; Bölöni and Horvath, 2010). Different methods of computation of the B matrix can also be tried (ensemble B matrix was calculated but not tested). One other way that upper-air assimilation could be enhanced is bias correction of satellite data. At the end of December 2010, the variational bias correction (Auligné et al., 2007) was implemented instead of the static one (Harris and Kelly, 2001) that was used before. Using this approach, bias in the background was decreased, but the impact on forecast still must be validated. Work is also ongoing in implementing new observations (radar data), which will become very important when the ALADIN HR model goes to higher resolution. Except enhancing assimilation system, verification methods must also be enhanced. One of the enhancements would be to include verification of precipitation (e.g., SAL) but maybe the best way of demonstrating the impact of the assimilation system would be to use case studies, and that is something that has to be focused on in future research.

Acknowledgements: Many people participated in the efforts to implement the assimilation system at DHMZ, including office colleagues, Tomislav Kovačić, Kristian Horvath, Stjepan Ivatek-Šahdan, Martina Tudor, and international colleagues Gergely Bölöni, Alena Trojakova and many more. Their assistance and advice are gratefully acknowledged.

REFERENCES

- ALADIN International Team, 1997: The ALADIN project: Mesoscale modelling seen as a basic tool for weather forecasting and atmospheric research. *WMO Bull.*, **46**, 317–324.
- Auligné, T., A. P. McNally and D. P. Dee, 2007: Adaptive bias correction for satellite data in numerical weather prediction. *Quart. Jour. Roy. Meteor. Soc.*, **133**, 631–642.
- Berre, L., 2000: Estimation of Synoptic and Mesoscale Forecast Error Covariances in a Limited-Area Model. *Mon. Wea. Rev.*, **128**, 644–667.
- Bjerknes, V., 1904: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik. *Meteorologische Zeitschrift*, **21**, 1–7.

- Bouttier, F. and P. Courtier, 1999: Data assimilation concepts and methods, ECMWF lecture notes, European Centre for Medium-Range Weather Forecasts, Reading, England, 59 pp.
- Bouttier, F., J. F. Mahfouf, and J. Noilhan, 1993a: Sequential assimilation of soil moisture from atmospheric low-level variables. Part I: Sensitivity and calibration studies. *J. Appl. Meteor.*, **32**, 1335–1351.
- Bouttier, F., J. F. Mahfouf, and J. Noilhan, 1993b: Sequential assimilation of soil moisture from atmospheric low-level variables. Part II: Implementation in a mesoscale model. *J. Appl. Meteor.*, **32**, 1352–1364.
- Bölöni, G., 2006: Development of a variational data assimilation system for a limited area model at the Hungarian Meteorological Service. *Időjárás*, **110**, pp. 309–327
- Bölöni, G. and K. Horvath, 2010: Diagnosis and tuning of background error statistics in a variational data assimilation system. *Időjárás*, **114**, 1–19
- Courtier, P., E. Andersson, W. Heckley, J. Pailleux., D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier and M. Fisher, 1998: The ECMWF implementation of three-dimensional variational assimilation (3d-Var). I: Formulation. *Quart. Jour. Roy. Meteor. Soc.*, **124**, 1783–1807.
- Desroziers G., L. Berre, B. Chapnik and P. Poli, 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. Jour. Roy. Meteor. Soc.*, **131**, 3385–3396
- Fischer, C., T. Montmerle, L. Berre, L. Auger and S.E. tefönescu et al., 2005: An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quart. Jour. Roy. Meteor. Soc.*, **131**, 3477–3492
- Gaspari, G. and S. Cohn (1999): Construction of correlation functions in two and three dimensions. *Quart. Jour. Roy. Meteor. Soc.*, **125**, 723–757.
- Harris, B.A., and G. Kelly, 2001: A satellite bias correction scheme for radiance assimilation. *Quart. Jour. Roy. Meteor. Soc.*, **127**, 1453–1468
- Ivatek-Šahdan, S. and M. Tudor., 2004: Use of high-resolution dynamical adaptation in operational suite and research impact studies. *Meteorologische Zeitschrift*, **13**, 2; 99–108
- Giard, D. and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997–1015
- Hólm, E.V., 2008: Lecture notes on assimilation algorithms. ECMWF, European Centre for Medium-Range Weather Forecasts, Reading, England, 30 pp.
- Lorenz, A. C., 1986: Analysis method for numerical weather prediction. *Quart. Jour. Roy. Meteor. Soc.*, **112**, 1177–1194
- Mahfouf, J., F., 1991: Analysis of soil moisture from near-surface variables: A feasibility study. *J. Appl. Meteor.*, **30**, 1534–1547
- Noilhan, J. and J. F. Mahfouf, 1996: The ISBA land surface parameterization scheme. *Global Planet. Change*, **13**, 145–149
- Noilhan, J. and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **128**, 997–1015
- Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763
- Taillefer, F., 2002: CANARI – Technical Documentation – Based on ARPÉGE cycle CY25T1 (AL25T1 for ALADIN), available at <http://www.cnrm.meteo.fr/aladin/>
- Talagrand, O., 1967: Assimilation of observations, an introduction.. *J. Met. Soc. Japan*, Special Issue **75**, 1B, 191–209.
- Tudor, M. and S. Ivatek-Šahdan, 2010: The case study of bura of 1st and 3rd February 2007, *Meteorologische Zeitschrift*, **19**, 5; 453–466