Hrvatsko meteorološko društvo
Croatian Meteorological Society

HRVATSKI METEOROLOŠKI ČASOPIS
CROATIAN METEOROLOGICAL JOURNAL

Časopis je izdan povodom 20. godišnjice ALADIN projekta u Hrvatskoj

50

Hrv. meteor. časopis	Vol. 50	p. 1-144	ZAGREB	2015

Hrvatski meteorološki časopis – Croatian Meteorological Journal, 50, 2015., 47–70

Professional paper
Stručni rad

OPERATIONAL VALIDATION AND VERIFICATION OF ALADIN FORECAST IN METEOROLOGICAL AND HYDROLOGICAL SERVICE OF CROATIA

Operativna validacija i verifikacija prognoze modela ALADIN u Državnom hidrometeorološkom zavodu

MARTINA TUDOR, ANTONIO STANEŠIĆ, STJEPAN IVATEK-ŠAHDAN, MARIO HRASTINSKI, IRIS ODAK PLENKOVIĆ, KRISTIAN HORVATH, ALICA BAJIĆ AND TOMISLAV KOVAČIĆ

Meteorological and Hydrological Service, Grič 3, 10000 Zagreb, Croatia
Državni hidrometeorološki zavod, Grič 3, 10000 Zagreb, Hrvatska
tudor@cirus.dhz.hr

Recieved 8 May 2015, in final form 5 October 2015

Abstract: The numerical forecast using ALADIN model in Meteorological and Hydrological Service of Croatia is run operationally since July 2000. Over the years, various methods of validation and verification of the operational forecast have been applied. The classical methods using root mean square error and mean absolute error would often penalize the high resolution ALADIN when compared to a low resolution global model forecast due to double penalty paradigm. Therefore, the model was mostly evaluated by plotting the forecast and the measurements to allow subjective comparison, especially in weather situations that have high impact on the living and traffic conditions in Croatia. Here we show an overview of validation and verification products created operationally. These products intended for subjective validation in real time can help the forecaster in the decision if to rely on a particular forecast run more or less than to another. Statistical verification scores provide information on model bias and root mean square error but suffer from missing data due to automatic procedures used quality check and filtering of the measured data.

Keywords: ALADIN model, verification, validation, operational NWP

Sažetak: Operativna numerička prognoza vremena korištenjem modela ALADIN se na Državnom hidrometeorološkom zavodu provodi od srpnja 2000. Tijekom godina, primjenjivane su razne metode validacije i verifikacije operativne prognoze. Klasične metode srednje kvadratne i apsolutne pogreške često "kažnjavaju" model visoke rezolucije ALADIN u usporedbi s prognozom globalnog modela niske rezolucije zbog problema dvostrukog obračunavanja. Zbog toga je model ocjenjivan direktnom usporedbom s mjerenjima, posebno u vremenskim situacijama koje imaju značajan utjecaj na kvalitetu života i odvijanja prometa u Hrvatskoj. U ovom radu pokazujemo pregled operativnih produkata validacije i verifikacije. Operativni verifikacijski produkti namjenjeni za subjektivnu validaciju u realnom vremenu omogućuju prognostičarima u odluci o tome na koji prognostički produkt se mogu više osloniti. Statistička verifikacija daje informaciju o biasu modela i srednjoj kvadratnoj pogrešci modela, ali imaju nedostatke zbog automatskih procedura kojima se kontrolira kvaliteta i filtriraju mjereni podaci.

Ključne riječi: model ALADIN, verifikacija, validacija, operativna numerička prognoza vremena

1. INTRODUCTION

It is important to provide consistent verification information on the Numerical Weather Prediction (NWP) products produced by operational forecast models in order to inform the users, such as forecasters, on the reliability of the forecast fields. Furthermore this can help the research and development team comparing forecasts computed using different model tunings and so, to improve the forecast quality.

The value and skill of the forecast fields computed by NWP models has to be evaluated on a regular basis (Jolliffe and Stephenson 2011). New model versions are evaluated through comparisons with the operational forecasts. Optimal tunings for the operational forecasts are set up using different methods for verifying the forecast quality. Therefore, it is essential for every institution involved in operational NWP to do the forecast verification.

In order to compute a reliable assessment in the performance of the forecasting system, institutions usually rely on a statistical verification approach. As pointed by the reviewers, statistics of the modelled distribution of specific events and comparison with observed event distribution can give some hint on the ability of the model to predict extreme events, such as wind speed exceeding a certain threshold. The ability of dynamical adaptation to correctly forecast wind speed in severe bura wind has been documented and verified before. However, a statistically based forecast verification over a long time period seldom yields useful information on the system's ability to predict significant weather events (with an exception of specialized complex scores e.g. Ferro and Stephenson 2011), such as short events of severe wind that are a consequence of particular weather phenomena (see Tudor and Ivatek-Šahdan 2010, for an example) that pose a threat to the public safety and are considered rather rare (more in Odak Plenković et al. 2015). These events are associated to particular physical conditions that require a thorough analysis usually performed in a form of a case study (otherwise, in long term statistics, these extreme wind events are masked by long lasting episodes of severe wind). Such an approach shows the ability of the NWP system in predicting the weather extremes and is often used in the context of monitoring the operational performance and forecasting.

1.1. Predictability and impact of weather events

Some types of the weather systems are less predictable than other due to their intrinsic physical properties (Grazzini 2007). NWP models provide mostly correct early warning of highly predictable atmospheric processes such as Alpine lee cyclogenesis (Speranza et al. 1985) and other large scale synoptic disturbances. On the other hand, moist cyclogenesis (Romero 2011) in the presence of complex topographic features is much less predictable and NWP models predict them less correctly and not so long in advance. Although both of these events represent cyclogenesis producing high impact weather and similar measurable consequences on the ground, the ability of a NWP model to forecast one rather than the other is not a consequence of NWP model skill, but the intrinsic low predictability of certain types of events (e.g. Hally et al. 2014). The forecast could benefit from probabilistic forecast methods like ensemble modelling and the forecast of smaller scales might improve with higher resolution.

The key physical processes, that lead to the high impact weather events are investigated using detailed and often dedicated observations obtained during field campaigns pursued during special observing periods of various research programmes, such as Alpine Experiment (ALPEX, http://www.eol.ucar.edu/field_projects/alpex) and Mesoscale Alpine Programme (MAP, e.g. Tudor and Ivatek-Šahdan 2002) that focus mostly on cases of severe bura, while Convective and Orographically-induced Precipitation Study (COPS, Wulfmeyer et al. 2011), HyMeX (Drobinski et al. 2014) and MEDEX (Jansa et al. 2014) study high precipitation events, related to the presence of topography and moisture arriving from the sea surface.

Each component of the operational NWP system is usually evaluated through objective verification as the system is running by an increasing number of tools. Verification tools are developed in LACE (Limited Area for Central Europe, www.rclace.eu), ALADIN (Aire Limitée Adaptation Dynamique développement

InterNational, ALADIN International Team, 1997) and HIRLAM (High Resolution Limited Area Model, www.hirlam.org) communities and are available as common tools for intercomparison of operational model suites running in different services and for validation at home. Other consortia and forecasting centres (ECMWF, NCEP, WRF, COSMO, Unified Model) also develop validation tools.

The effect of changes in the operational NWP system is assessed through rigorous analysis of forecast verification results that should identify the weak points in the model performance. Unfortunately, while investigating the cause of problems, a researcher can face a situation where a change that benefits one aspect of forecast performance, can deteriorate the forecast of another. The set-up of an operational forecast model then results from compromises as one choice can be more successful in the prediction of low stratus and fog (Tudor 2010) and spoil the forecast in cases with shallow convection. In such a situation, more emphasis could be given to forecasting weather events that pose a threat to public safety and require warning (such as severe wind and high precipitation events) than on less significant weather. This strategy reflects itself in testing and tuning of a new model version on a period with severe weather (extreme precipitation or several successive storms). This choice has its drawbacks. Such a choice can lead to larger model errors and bias, if the high impact weather events (that are supposed to be relatively rare) are given more importance (when choosing the model set-up) than correct forecast in the cases of fine weather or less pleasant but harmless weather features. To summarize, a model developer and operational maintenance staff have often to decide, if it is more important to maintain good numbers obtained through statistical methods or a signal of developing high impact weather event that poses a threat to the public safety and requires timely warning although the exact location and time of the expected event cannot always be determined precisely. Of course, the model development aims and, in some cases, succeeds at the good performance in both, extreme events and regular statistics.

A model that is able to produce small scale and often less predictable weather patterns performs worse in terms of the statistical scores than a model that is not able to produce such phenomena in the forecast fields (unless a rather fortunate event when the deterministic forecast really hits the moment in time and space, there is one such example in Tudor and Ivatek-Šahdan 2010) since a slight misplacement of the event in space and time performs worse in terms of statistical scores based on point verification than if the event was not forecast at all (Rossa et al. 2008). Plotting the model together with observation data, allows the forecaster to recognize such situations in real time while the more complex methods for verifying the derived quantities yield information on the long term model performance. Statistical verification methods applied to standard output meteorological fields often do not point to specific errors in the model forecast due to numerical properties of the physics parameterisations (Tudor 2013). This would require additional output in each model time step, which is not practical in operational mode.

Another problem related to the operational forecast verification arises from the measured data itself. Measurements of meteorological variables also contain errors, especially those transferred in real time. These errors have to be quality checked and filtered from erroneous values by an automatic procedure. This procedure usually checks the measured data by comparing the values to some model result, such as the model analysis and discards the data as wrong if the difference is larger than some threshold value. Therefore the measurement error consist of the actual measurement error and a representativeness error since the measurement is point-wise while the model grid-point value is representative for the scale of the model grid. The model analysis used in this procedure is a result of combining a short range forecast (first guess) and measurements (possibly the same set as used for verification later). If the model is not able to forecast a certain phenomenon that exists in the measured data (and nature) the data are discarded (both in data assimilation and verification). On the other hand, if the measured data exhibit small but systematic error, these would be used for verification (and data assimilation). In consequence a model configuration that was successful in predicting a certain weather event might show poor performance in the statistical scores, if the system has thrown away the measured data for that particular event, since it did not exist in the model fields of the first guess. The short-cut verification procedure that computes the difference of model forecast to the analysis also suffers from this. As pointed by the reviewer, one should consider other methods of observation quality check that are independent of NWP such as cross correlation of observations. Hrastinski et al. 2015 in this issue describe spectral verification applied on operational forecast data.

Computing statistical scores for high resolution forecast is a particular challenge since the measured data are available in lower spatial resolution, while in-situ measured data are often representative for an area still smaller than the model grid cell that affects the results for coarser models. On the other hand, operational model output is usually written with a coarse time interval (1 or 3 hours) but still represent a temporal snapshot (except for accumulated quantities such as precipitation) on a regular spatial grid, while there are measured data available in higher temporal resolutions (e. g. 10 minutes) on irregular and less dense grid.
Depending on the type of measured data and model variables, one can choose from a number of choices:

1) to spatially and temporally average the model data and the measured data to the same, coarser horizontal grid before computing the statistical parameters,
2) to use values from grid points over a certain area and/or computed by different components of the operational suite as an ensemble of forecast values when comparing it to the in situ measured values,
3) more complex methods, such as SAL that evaluates the structure (S), amplitude (A), and location (L) of the precipitation field (Wernli et al. 2008) avoid double penalty problem,
4) verification of derived quantities such as energy and power spectra (see accompanying paper by Hrastinski et al in this issue).

2. VALIDATION AND VERIFICATION OF OPERATIONAL FORECAST

The operational model forecast in Meteorological and Hydrological Service (DHMZ hereafter) of Croatia is routinely evaluated, even using automated procedures, in order to provide essential feedback of forecast performance to users and model developers. The model evaluation relies on a range of products, from simple model to data comparison that allows subjective validation as well as detection of specific flaws in the model performance to verification statistics that evaluates the accuracy of the forecasts. Operational monitoring of forecast quality is complemented by in-depth case studies of forecast performance in cases of high impact weather, since prompt and correct forecast of such events has profound benefits for the traffic, safety and society in general. Model evaluation through validation and verification provides a comprehensive understanding of strengths and weaknesses of the forecasting system.

2.1. Tools

Plotting the direct model output forecast along with the measured data operationally provides forecasters the information, which components of the model forecast are close to the real values measured in nature. Such a comparison is plotted for all the stations and measured and observed data that are available operationally in real time from the territory of the Republic of Croatia as well as on about an equal number of stations from abroad. This approach allows to treat every forecast on each day as a case study but done in real time. Satellite estimates of precipitation are also used to evaluate forecast of precipitation combined with the measurements from the rain gauges. The plots of measured 24 hour precipitation are plotted using the same weight for measurements from rain gauges and satellite estimates. Only a fraction of Croatia is covered by meteorological radars and there is currently no measured data available operationally in a form of numbers (the software installed there produces only figures) so it is necessary to rely on satellite estimates for the areas where there is no other measured data. The satellite precipitation estimates are calibrated using in-situ data (Huffman et al. 2007) and have been used in numerous studies (a

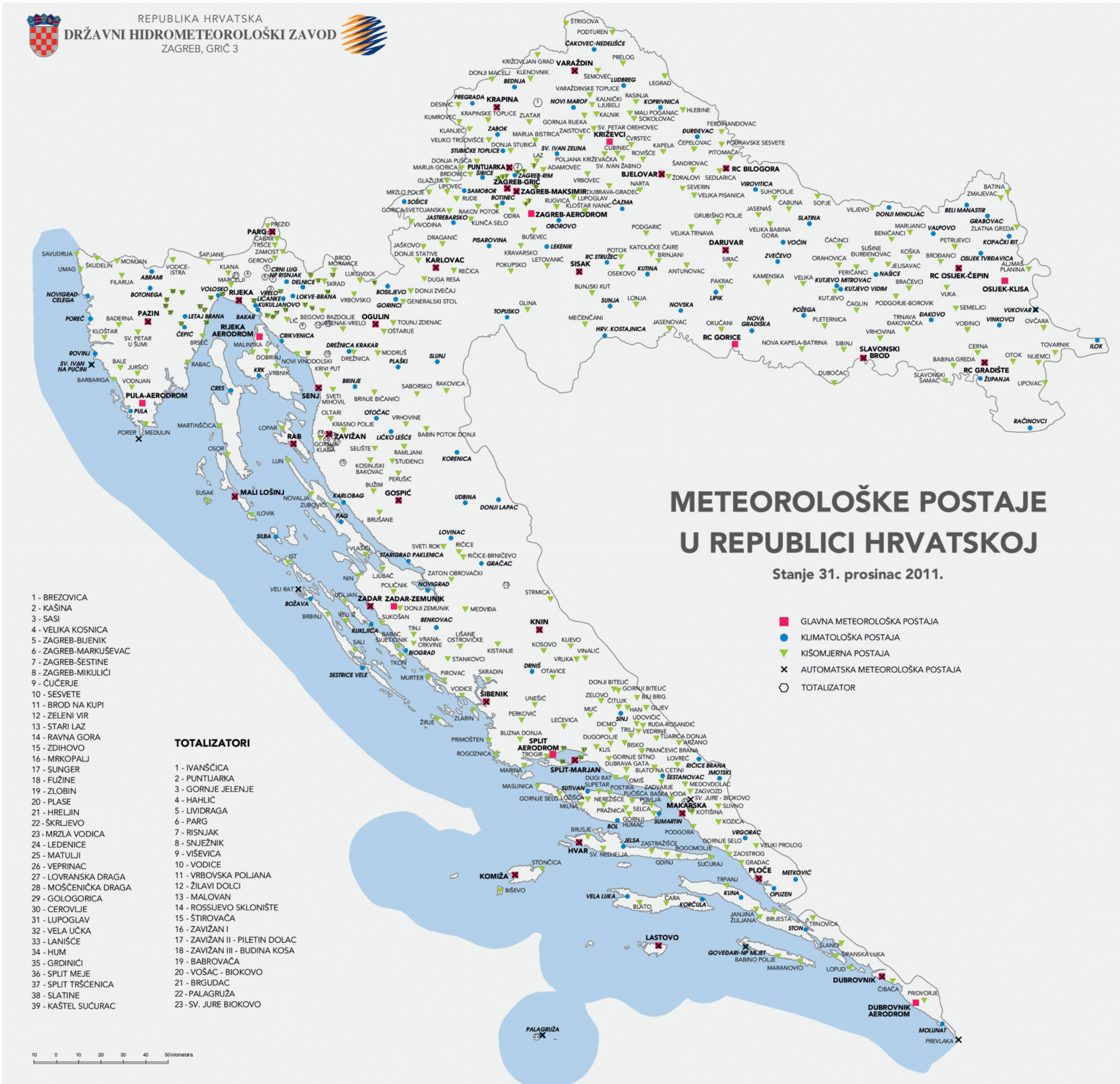

Figure 1. The meteorological observational network in Croatia (conditions on 31 December 2011) maintained by DHMZ, main stations that report measured and observed data with hourly interval (purple square), climatological stations that report measured and observed data three times a day (blue circle), rain gauge stations that report precipitation accumulated during previous 24 hours at 6 UTC (green triangle) and automatic stations that report measured data with an interval of 10 minutes (x sign).

Slika 1. Meteorološke postaje u Hrvatskoj (stanje 31. prosinca 2011. godine) u vlasništvu Državnog hidrometeorološkog zavoda, glavne meteorološke postaje sa mjerenjima i motrenjima sa satnim intervalom (ljubičasti kvadrat), klimatološke postaje sa mjerenjima i motrenjima 3 puta dnevno (plavi krug), kišomjerne postaje sa mjerenjima količine oborine u prethodna 24 sata u 6 UTC (zeleni trokut) i automatske meteorološke postaje sa izmjerenim podacima svakih 10 minuta (oznaka x).

separate list available for each year at http://trmm.gsfc.nasa.gov/publications_dir/publications.html).

There have been several evaluation methods developed within ALADIN and HIRLAM communities that are used for the verification of the operational forecast (a list of applications can be found at http://www.rclace.eu/?page=106). The operational forecast fields are sent to a joint LACE centre for verification in Ljubljana where the forecasts from the 7 LACE member countries, AT, CZ, HR, HU, RO, SI, SK, are evaluated and their performance is compared to other model forecasts in the framework of ALADIN verification project.

The methods for verification are further developed. A verification tool developed in ALADIN community, called VERAL (verification for ALADIN, http://www.rclace.eu/File/NWP_Utility_Inventory/ ALADIN/Czech_Republic/quest_CHMI_veral.doc) was installed in DHMZ and is used for operational verification as well as testing of new versions of model code available for operational forecast and various options and tunings before a new model version is introduced to the operational forecast suite. The VERAL program package produces standard deviation, bias and root mean square error of surface variables, such as 2m temperature and relative humidity, mean sea level pressure, wind components as well as wind speed and direction. It also computes the same statistical parameters in vertical-time cross sections for temperature, geopotential height, wind speed and direction and relative humidity on standard isobaric surfaces using vertical soundings available over the model domain.

Recently, HARMONIE verification package (Yang 2008) has been ported locally. The operational forecast data are also sent to the joint centre for verification using the same package, so this also enables inter-comparison of model forecasts from different services (and hence different models, model versions, options, resolutions etc.). This verification package is more advanced and computes more scores when compared to VERAL and offers more flexibility.

Verification is performed using standard GTS data available via global exchange and high resolution in-situ data from the national network (when possible). The measurements used for verification by the common tools such as the HARMONIE package mentioned previously are obtained from ECMWF, where the measured data are quality checked, flagged and discarded by an automatic procedure. Unfortunately, this means that some measured data during particular weather phenomena are discarded as wrong. It is still not possible to compute flow dependent scores (e.g. Rossa et al. 2008) with a tool available locally in DHMZ. However, forecasts of severe weather are being evaluated through diagnostic studies.

The performance of the operational forecast can be monitored on the intranet http://noa.gric.dhz.hr/~aladinhr/veral-oper-00-06-12-18/HTML-DIR/ where the recent and historic products of the VERAL verification package can be accessed. The scores are computed monthly, seasonally and separately for cold/warm part of the year. The scores are computed for both surface and upper air fields, regularly after the period of interest finishes. Further details and results are shown in section 4.1.

The verification products for the operational forecast computed using HARMONIE verification package are also available via intranet on the page http://noa.gric.dhz.hr/~aladinhr/HARMONIEverif/ where a practical user interface allows a choice in the variable and score to be displayed, section 4.2 shows several examples of scores.

2.2. Measured data

The measured data used for validation of model forecast comes from different sources. The in-situ measured data include the hourly SYNOP data, measured data from automatic stations from the operational network in Croatia, and 3 hourly SYNOP measurements from the international exchange. Precipitation is also evaluated a posteriori by comparison to rain gauge data from a dense observational network in Croatia (available with delay of several months) and data available from satellite estimates, such as the data available through Tropical Rainfall Measuring Mission (TRMM, Huffman et al. 2007).

The real time TRMM Multi-satellite Precipitation Analysis (TMPA) data is available through several products. The 3B40RT product contains precipitation estimate from combined measurements in the microwave spectrum. The 3B41RT product contains precipitation estimate computed from the data in the infrared part of the spectrum. Finally, the product used operationally is the 3B42RT that contains combined estimate of precipitation using both infrared and microwave data. Satellite derived precipitation data are used as provided by the Giovanni web server interface (Acker and Leptoukh 2007) on http://disc.sci.gsfc.nasa.gov, particularly, the 3 hourly accumulated precipitation data from the 3B42RT product were used to compute the 24 hourly accumulated rainfall for the period from 0600 UTC to 0600 UTC the next day in order to cover the same observation period as the rain gauge network. The data from all the stations in the national network and a choice of stations available through the international exchange are plotted against the forecast values, the examples are shown in further sections.

2.3. Operational forecast models run in DHMZ

The operational ALADIN forecast is run on 8 km resolution since 2000 (for details see Tudor et al. 2013). The domain size increased over the years, the operational model version, the methods to obtain initial conditions and the global model forecast that is needed for lateral boundary conditions have been changing. In 2015, a model set-up run in 4 km resolution has been introduced to run once per day. Parallel to the 8 km forecast, a high resolution dynamical adaptation of wind is run in 2 km resolution, adapted from Žagar and Rakovec (1999), using reduced physics package and fewer levels in the vertical (Ivatek-Šahdan and Tudor 2004). Since July 2011, high resolution ALADIN is run operationally in 2 km resolution using the complete set of physics parametrisations (including the deep convection) and non-hydrostatic dynamics on

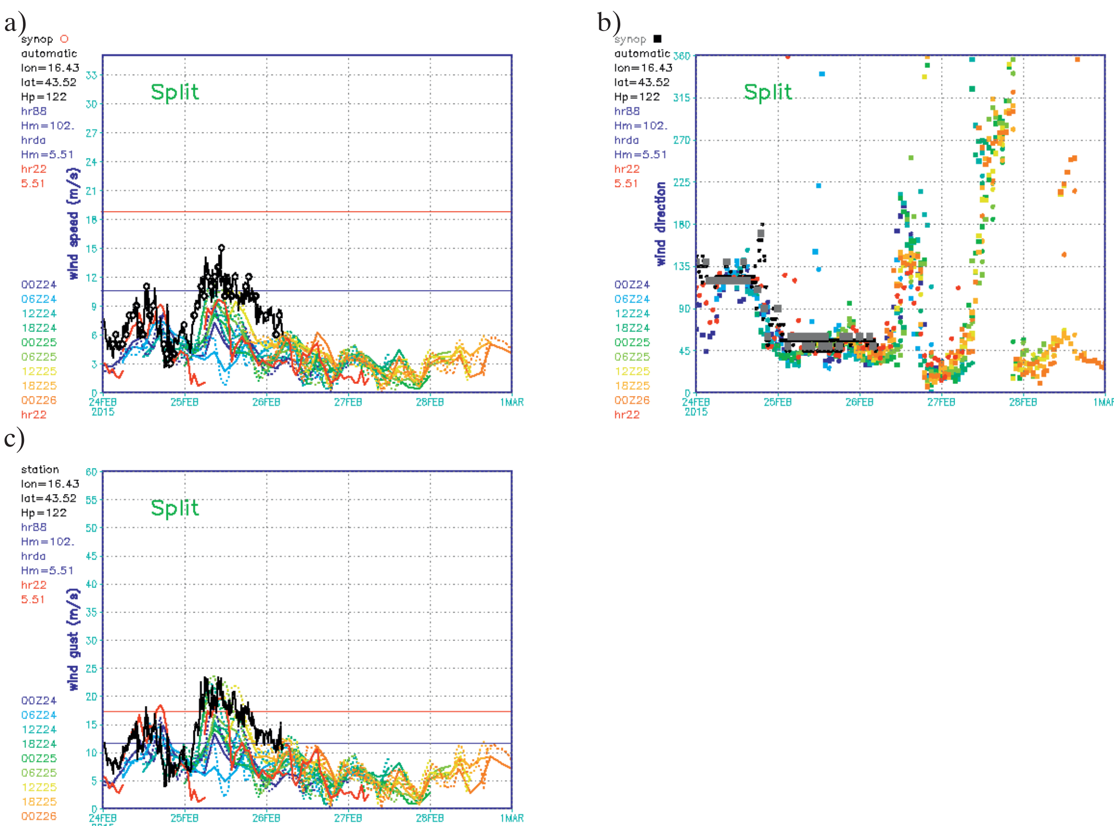

Figure 2. Wind speed on main station Split Marjan (a), direction (b) and gusts (c) as measured by automatic station (black), main station (circles for speed, grey for direction), operational forecasts using Aladin in 8 km resolution starting from different analyses are shown in different colours (full lines for speed, squares for direction), the analyses times are plotted in the same colour on the left, dynamic adaptation of wind field (dotted lines for speed, circles for direction) and 2 km resolution NH forecast (red lines for speed and red circles for direction).

Slika 2. Vjetar na glavnoj meteorološkoj postaji Split Marjan: srednja brzina (a), smjer vjetra (b) i udari vjetra (c) izmjereni na automatskoj postaji (crno), glavnoj postaji (krug za srednju brzinu i sivo za smjer vjetra), prognoze operativne verzije sa 8 km horizontalnom razlučivošću pokretane iz različitih analiza prikazane su različitim bojama (puna linija brzina vjetra, kvadratići smjer vjetra), termini analiza prikazani su na lijevoj strani, dinamička adaptacija polja vjetra (isprekidane linije za brzinu, krugovi za smjer vjetra) i nehidrostatska verzija sa 2 km horizontalnom razlučivošću (crvene linije za brzinu vjetra i crveni krugovi za smjer vjetra).

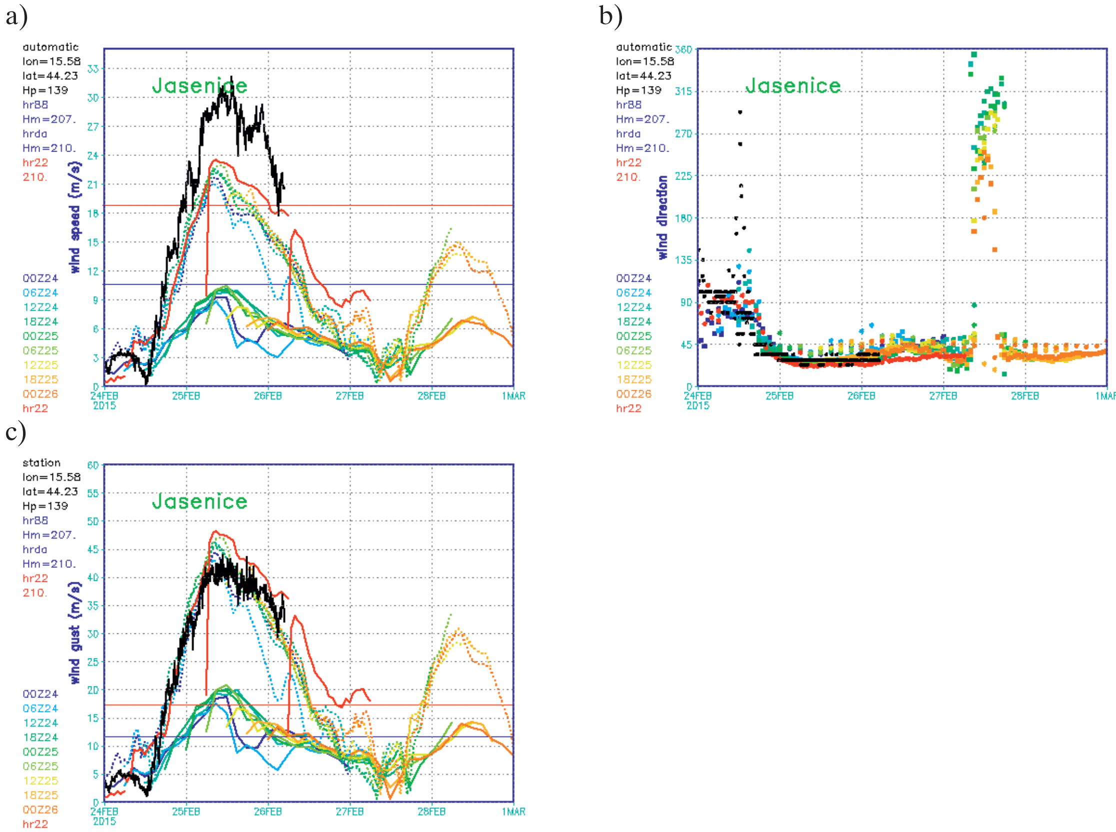

Figure 3. Wind speed on automatic station Jasenice (a), direction (b) and gusts (c) as measured by automatic station (black), operational forecasts using ALADIN in 8 km resolution starting from different analyses are shown in different colours (full lines for speed, squares for direction), the analyses times are plotted in the same colour on the left, dynamic adaptation of wind field (dotted lines for speed, circles for direction) and 2 km resolution NH forecast (red lines for speed and red circles for direction).

Slika 3. Srednja brzina vjetra (a), smjer vjetra (b) i udari vjetra (c) izmjerene na automatskoj postaji Jasenice (crno), prognoze ALADIN operativne verzije sa 8 km horizontalnom razlučivošću pokretane iz različitih analiza prikazane su različitim bojama (puna linija brzina vjetra, kvadrat smjer vjetra), termini analiza prikazani su na lijevoj strani, dinamička adaptacija polja vjetra (isprekidane linije za brzinu, krugovi za smjer vjetra) i nehidrostatska verzija sa 2 km horizontalnom razlučivošću (crvene linije za brzinu vjetra i crveni krugovi za smjer vjetra).

the same number of levels as the 8 km resolution forecast (Tudor and Ivatek-Šahdan 2010). These operational forecasts are compared to the measured data operationally through plots that enable subjective verification and information on the model performance to the forecaster on duty. The forecasts are also verified regularly using objective tools. Since wind speed direction and gusts are measured at 10 meters above ground and temperature and humidity on 2 meters (the heights of the standard meteorological measurement) the model forecast fields at these heights are computed by vertical interpolation from the lowest model level (about 17 meters above ground when 37 levels are used both in 8 and 2 km resolution forecasts) using a parametrised vertical profile dependent on the stability (Geleyn 1988). Wind gusts were computed using Brožkova et al. (2007).

3. OPERATIONAL PRODUCTS FOR SUBJECTIVE VALIDATION

Every hour, an operational procedure automatically collects raw measured data from SYNOP and automatic stations. The data from SYNOP stations are all aggregated into a single binary file for the whole day, while for each automatic station, a separate binary file is created. These files are used by the GrADS (Grid Analysis and Display System, Tsai and

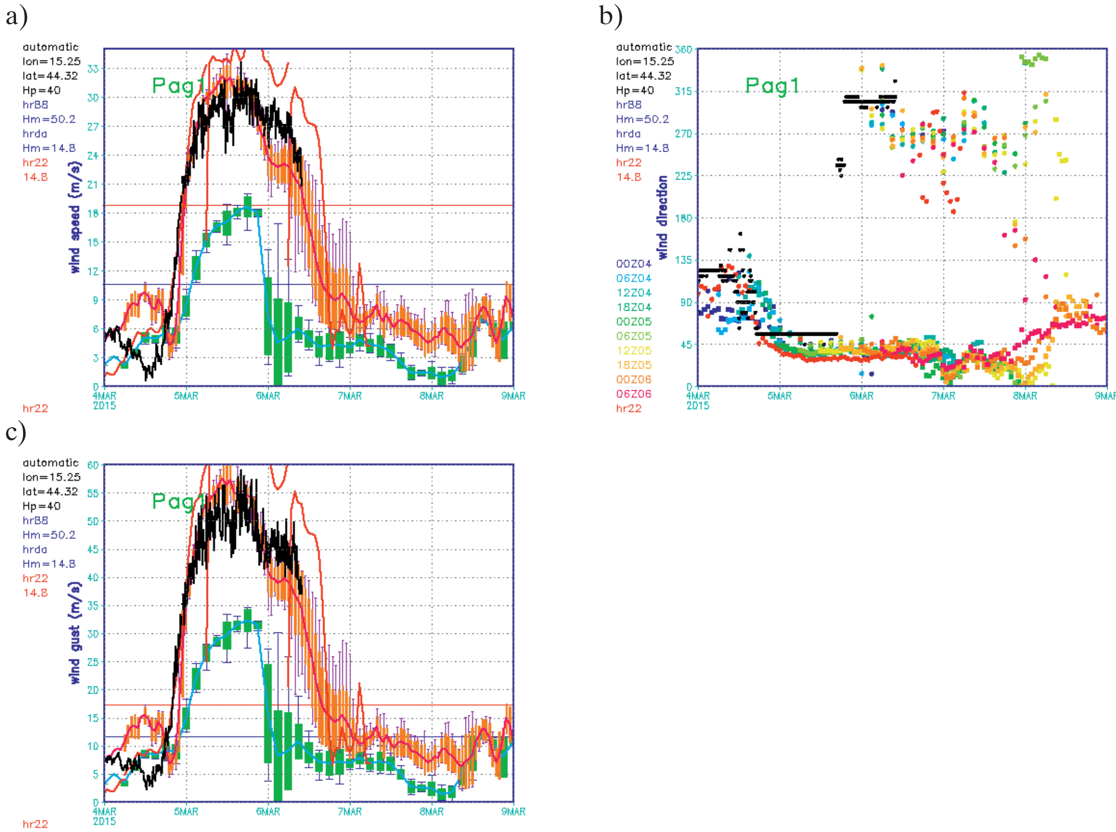

Figure 4. Wind speed on automatic station Pag1 (a), direction (b) and gusts (c) as measured by automatic station (black), operational forecasts using ALADIN in 8 km resolution starting from different analyses are shown as an ensemble forecast (full line is average, blue vertical lines minimum and maximum values and green square one standard deviation), dynamic adaptation of wind speed is also shown as an ensemble forecast (full cyan line is average, purple vertical lines minimum and maximum values and orange square one standard deviation) and 2 km resolution NH forecast (red lines for speed and red circles for direction).

Slika 4. Srednja brzina vjetra (a), smjer vjetra (b) i udari vjetra (c) izmjerene na automatskoj postaji Pag1 (crno), prognoze ALADIN operativne verzije sa 8 km horizontalnom razlučivošću pokretane iz različitih analiza prikazane su kao ansambl prognoza (puna linija je srednja vrijednost, plave vertikalne linije su minimalna i maksimalna vrijednost a zeleni kvadrat predstavlja jednu standardnu devijaciju oko srednje vrijednosti), dinamička adaptacija polja vjetra je također prikazana kao ansambl prognoza (puna modra linija je srednja vrijednost, ljubičaste vertikalne linije su minimalna i maksimalna vrijednost, a narančasti kvadrat predstavlja jednu standardnu devijaciju oko srednje vrijednosti) i nehidrostatska verzija sa 2 km horizontalnom razlučivošću (crvene linije za brzinu vjetra i crveni krugovi za smjer vjetra).

Doty, 1998) package to plot the evolution of model forecasts and measured values of meteorological parameters for each station in Croatia and chosen SYNOP stations from abroad.

Forecast values of wind speed, direction and gusts are interpolated bi-linearly to the point of measurement and their evolution in time is plotted and can be compared to forecasts from other operational runs (forecasts starting from previous and next analyses) and different configurations (8 km, 2 km dynamical adaptation, 2 km non-hydrostatic run). These plots are updated regularly every hour, when there is more measured data available. New forecast runs appear in the figure in the next hour after the forecast run is complete. The examples of these plots are shown in Figure 2 for the station Split Marjan from which there are available data for both SYNOP and automatic station reports and automatic station Jasenice in Figure 3. The upper left corner contains information on the station location (longitude, latitude and height) and types in black and grey letters. When there are data from both auto-

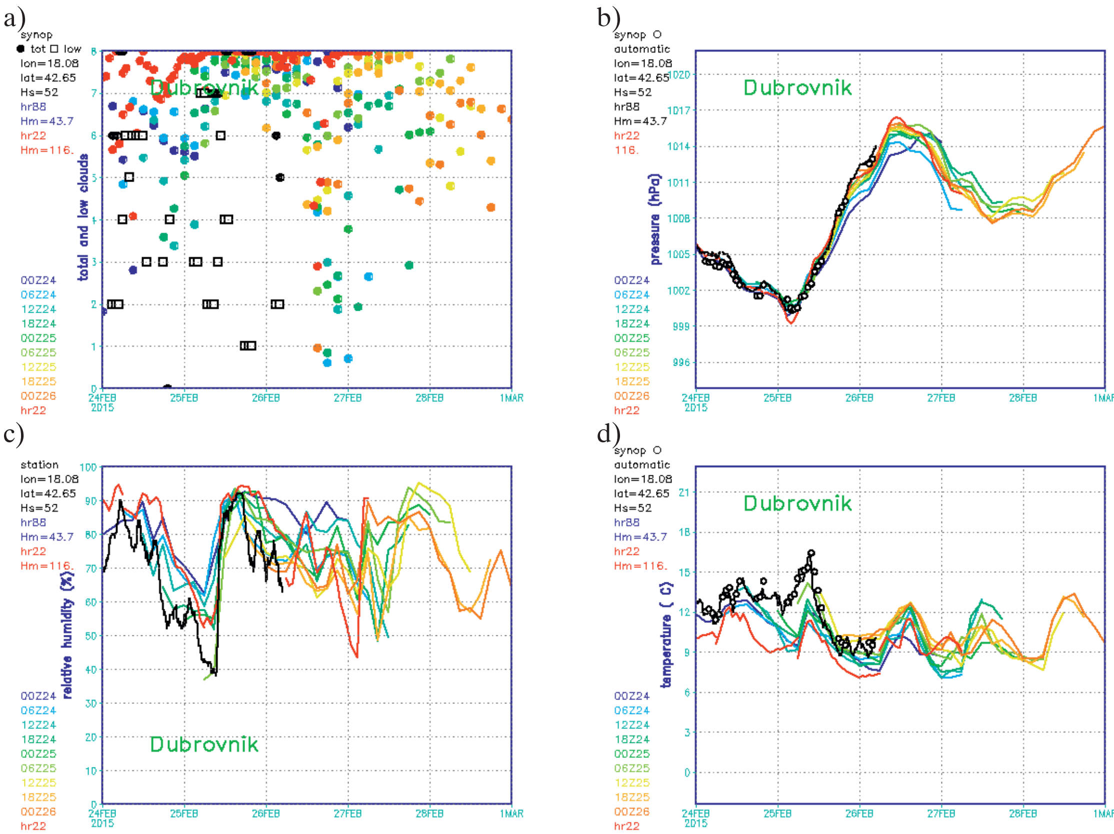

Figure 5. Observed cloudiness (full circle for total and empty square for low clouds) on main station Dubrovnik (a), measured pressure reduced to the mean sea level (b), relative humidity (c), and temperature at 2 meters (d), black circles for data reported in SYNOP, full line for measurements from automatic stations, operational forecasts using ALADIN in 8 km resolution starting from different analyses are shown in different colours (full lines for speed, squares for direction), 2 km resolution NH forecast (red circles for total cloudiness and red lines for other variables).

Slika 5. Glavna postaja Dubrovnik: motrena naoblaka (ispunjeni krugovi za ukupnu a prazni pravokutnici za nisku naoblaku) (a), mjereni tlak zraka reduciran na srednju razinu mora (b), relativna vlažnost zraka (c), i temperatura zraka na 2 m (d) crni krugovi su podaci dobiveni iz SINOP izvještaja, a puna linija su mjerenja automatske postaje, ALADIN operativna verzija sa 8 km horizontalnom razlučivošću pokretana iz različitih analiza prikazana je različitim bojama (puna linija za brzinu vjetra, kvadrati za smjer vjetra), termini analiza prikazani su na lijevoj strani, nehidrostatska verzija sa 2 km horizontalnom razlučivošću (crveni ispunjeni krugovi za ukupnu naoblaku a crvene linije za druge varijable).

matic and SYNOP reports available for the same station, the wind direction in SYNOP report is shown as grey squares and from automatic station as black circles. Below it, there is information on the terrain heights in 8 km resolution forecast (hr88) and 2 km resolution forecasts (hrda and hr22). The lower left side contains the time in UTC and day of the analysis from which the forecast runs have started in the same colour as the line or dots in the graph that plots the forecast data. The 8 km forecast and 2 km dynamical adaptation are run 4 times a day up to 72 hours, so there are forecasts from 9 to 12 runs shown in this figure. The 8 km forecasts are plotted as full lines for speed and squares for direction, the 2 km resolution dynamical adaptation forecasts are plotted as dotted lines for speed and circles for direction. The 2 km non-hydrostatic run is run only once per day (starting from 06 UTC) and up to 24 hours so there is only one result each time plotted as a full red line (as indicated by red hr22 mark in the lower left corner). The thresholds for strong and severe wind are shown as blue and red horizontal lines. Although the figures contain a lot of information and require some time to read them, they do show:

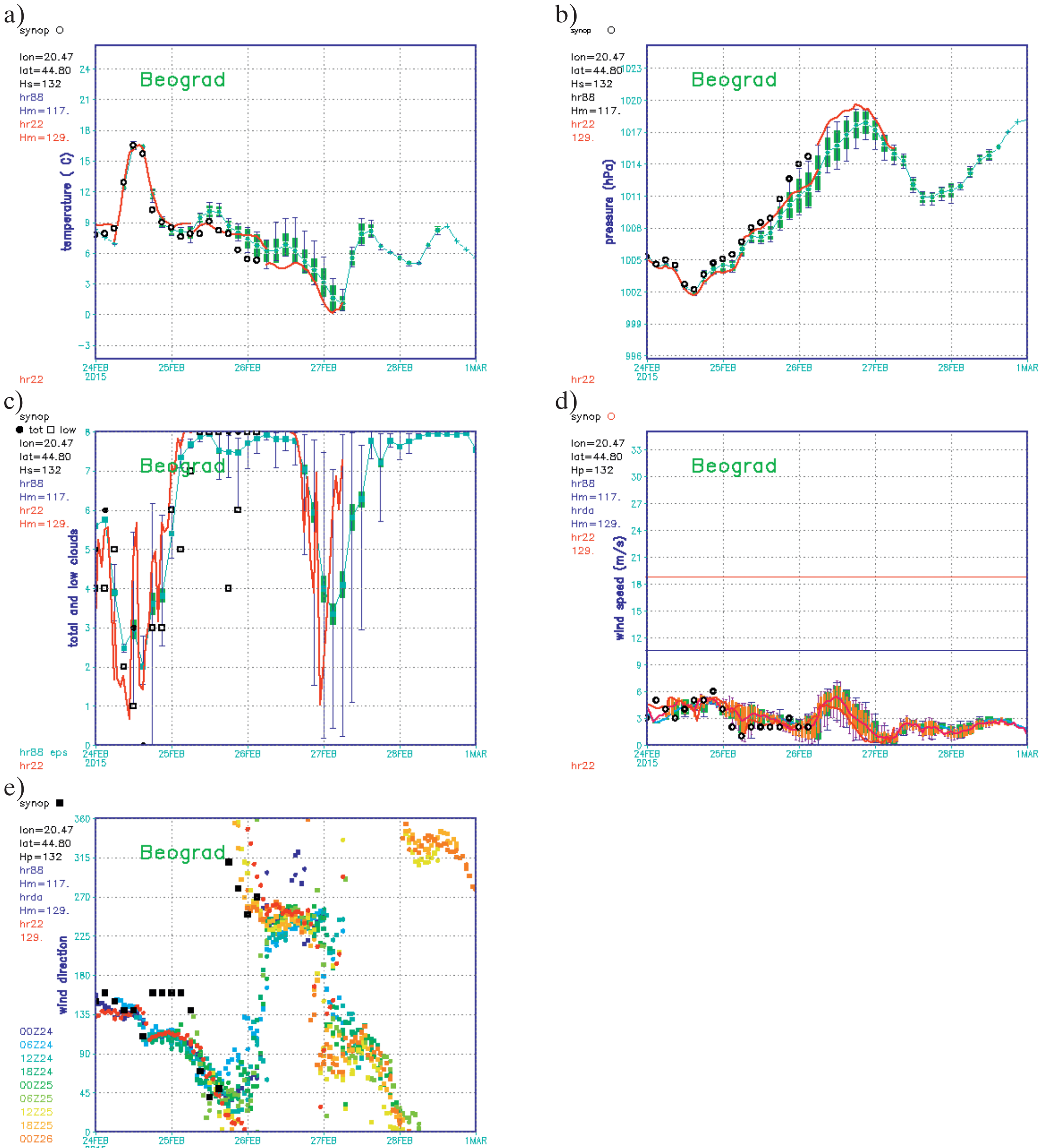

Figure 6. Measured temperature at 2m (black circles) on a station in Belgrade (a), pressure reduced to the mean sea level (b), observed cloudiness (c) full circle for total and empty square for low clouds, wind speed (d), and direction (e), operational forecasts in 8 km resolution starting from different analyses are shown as an ensemble forecast (full line is average, blue vertical lines minimum and maximum values and green square one standard deviation), dynamic adaptation of wind speed is also shown as an ensemble forecast (full cyan line is average, purple vertical lines minimum and maximum values and orange square one standard deviation) and 2 km resolution NH forecast (red lines for speed and red circles for direction).

Slika 6. Meteorološka postaja Beograd: mjerena temperature zraka na 2 m (a), mjereni tlak zraka reduciran na srednju razinu mora (b), motrena naoblaka (c) (ispunjeni krugovi za ukupnu a prazni pravokutnici za nisku naoblaku), brzina vjetra (d) i smjer vjetra (e), operativna verzija sa 8 km horizontalnom razlučivošću pokretana iz različitih analiza prikazane su kao ansambl prognoza (puna linija je srednja vrijednost, plave vertikalne linije su minimalna i maksimalna vrijednost a zeleni kvadrat predstavlja jednu standardnu devijaciju oko srednje vrijednosti), dinamička adaptacija polja vjetra je također prikazana kao ansambl prognoza (puna modra linija je srednja vrijednost, ljubičaste vertikalne linije su minimalna i maksimalna vrijednost, a narančasti kvadrat predstavlja jednu standardnu devijaciju oko srednje vrijednosti) i nehidrostatska verzija sa 2 km horizontalnom razlučivošću (crvene linije za brzinu vjetra i crveni krugovi za smjer vjetra).

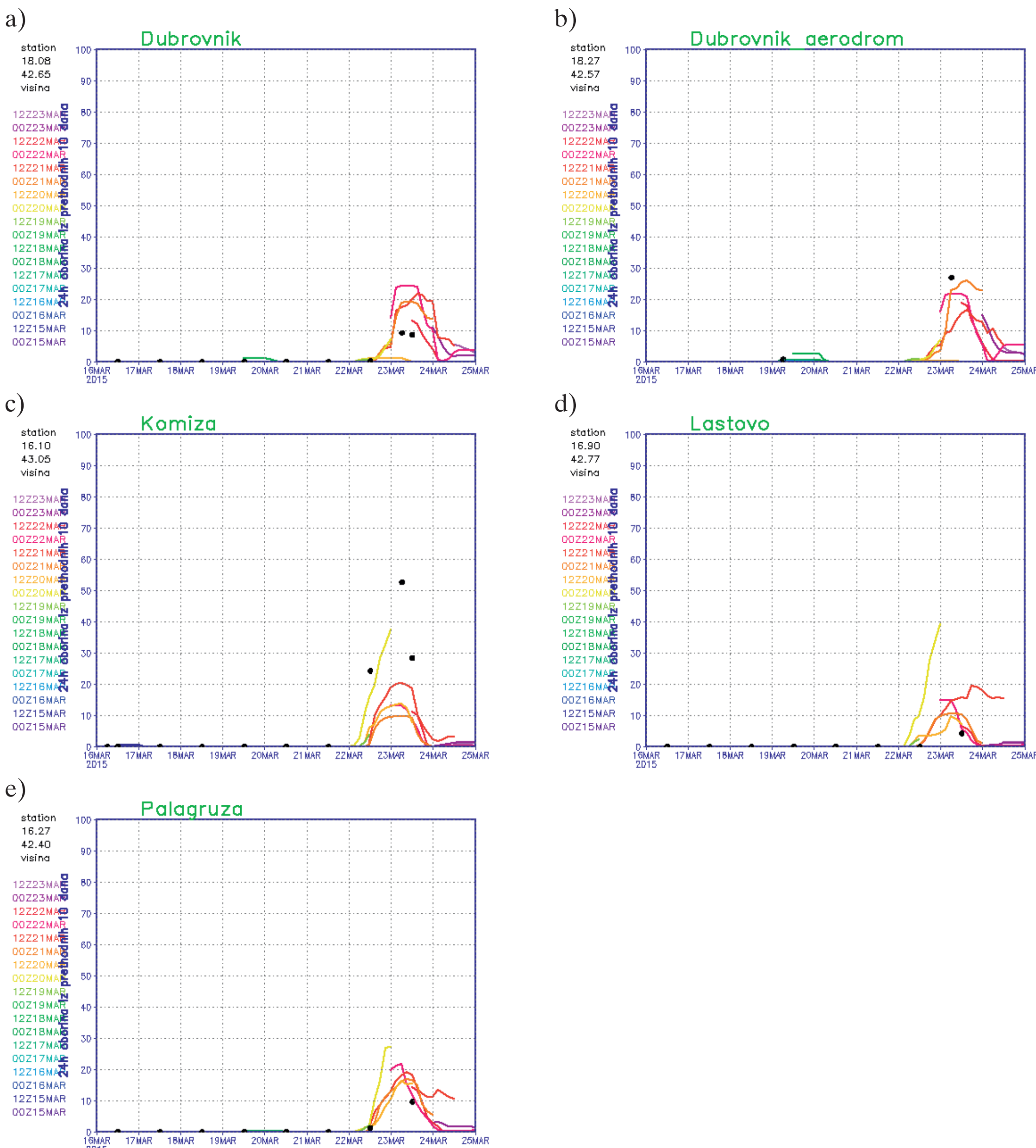

Figure 7. The predicted accumulated precipitation from previous 24 hours for previous 10 days on stations Dubrovnik (a), Dubrovnik airport (b), Komiža (c), Lastovo (d) and Palagruža (e), measured precipitation is shown in black dots (accumulated since the last shown dot), 8 km resolution ALADIN forecasts starting from different analyses are shown in different colours and the time and date of the analysis is written in the same colour on the left edge of each plot.

Slika 7. Prognoza 24 satne akumulirane oborine za prethodnih 10 dana za stanicu Dubrovnik (a), Dubrovnik Zračna luka (b), Komiža (c), Lastovo (d) i Palagruža (e), izmjerena oborina je prikazana crnim ispunjenim krugovima (akumulirana količina od prethodno prikazanog kruga), ALADIN prognoza sa 8 km horizontalnom razlučivosti pokretane iz različitih analiza prikazana je različitim bojama a vrijeme i datum analize nalazi se uz lijevi rub svake slike.

- if any of the forecasts have predicted the event,
- if the older forecast runs (shades of blue) are better or worse than the new forecast runs (shades of red),
- if the 2 km dynamical adaptation for the wind speed and 2 km resolution non-hydrostatic forecast for all parameters is better or worse than the 8 km resolution forecast,
- and even if there is a potential problem with the measured data e.g. wind speed in SYNOP report (circles) can be different than measured on the automatic station.

Figures 2 and 3 contain plots of each individual forecast for the particular location. Since there are 9 to 12 forecast runs available from both 8 km forecast and 2 km dynamical adaptation, these can be combined as two sets of ensemble forecasts containing 9 to 12 members each. The resulting ensemble forecast of wind speed and gusts is shown in Figure 4 for the automatic station PAG1 (showing wind direction as an ensemble forecast did not prove to be very informative, too much spread due to cyclical nature). Full blue line in Figure 4 is the average of the 8 km resolution ensemble, green bars show one standard deviation and blue vertical lines minimum and maximum values. The average of the ensemble of 2 km dynamical adaptation is in cyan, standard deviation is shown as orange bars and purple vertical lines show distance between minimum and maximum. The ensembles are shown for 3 hourly forecasts in 8 km resolution and hourly dynamical adaptation. The forecasts of the non-hydrostatic 2 km run are shown as the red lines since there is only one value available at the time. It is important to note that there are weather situations (such as bura, as shown in Figure 4) where the wind speed predicted by different operational set-ups is systematically different so the two ensembles could not be combined into one. Also, the 2 km non-hydrostatic forecast can be significantly different from the dynamical adaptation that it would be an outlier. These figures are clearer to read, but neither show if one forecast (perhaps starting from an older analysis) is better than the other nor if there are temporal variations in particular forecast runs also present in the measured data, if the variations occur at different times.

Model forecasts of cloudiness, mean sea level pressure, relative humidity and temperature at 2 meters are available only from the 8 km resolution forecast run and 2 km non-hydrostatic run. An example of forecast fields interpolated to the point of measurement for the station Dubrovnik is shown in Figure 5. Cloudiness is an observed parameter so it is only available from SYNOP stations (currently, there are no automatic measurements of cloudiness available at this time in Croatia). Measurements of relative humidity are used only from the automatic station data. The examples of plots where the cloudiness forecast is compared to the observed cloudiness is shown in Figure 5a for Dubrovnik and Figure 6c for Belgrade. The information on the station location, terrain height in the model and analyses of forecast runs shown in the figure are on the left side as for the wind field in Figures 2 and 3. The observed total cloudiness is shown as full black circles and low cloudiness as the empty squares. In Figure 5, individual forecasts are distinguished. Predicted cloudiness from forecasts starting from different analyses are shown in different colours, circles for cloudiness (only predicted total cloudiness is shown, predicted low clouds are not shown for clarity) lines for mean sea level pressure, temperature and relative humidity. Figure 6 shows an example of ensemble forecast product compared to the measurements at Belgrade SYNOP station for 2 m temperature, mean sea level pressure, cloudiness and wind speed (wind direction is again plotted only as individual forecasts). The correction to the temperature at 2 meters that accounts for the difference in altitude between the station and the model is not applied.

SYNOP reports contain data on measured accumulated rainfall from previous 6 or 12 hours and additionally accumulated rainfall from the previous 24 hours at 6 UTC every day. The automatic stations measure accumulated precipitation with a 5 minute interval, but there are only a few automatic stations that reported precipitation measurements. The model forecasts and measured accumulated precipitation from the previous 10 days for stations in south-east Croatia are shown in Figure 7. Forecasts starting from different analyses are shown in different colours and the time and date of the analysis are shown in the left (only forecasts starting from 00 and 12 UTC analyses are shown for clarity).

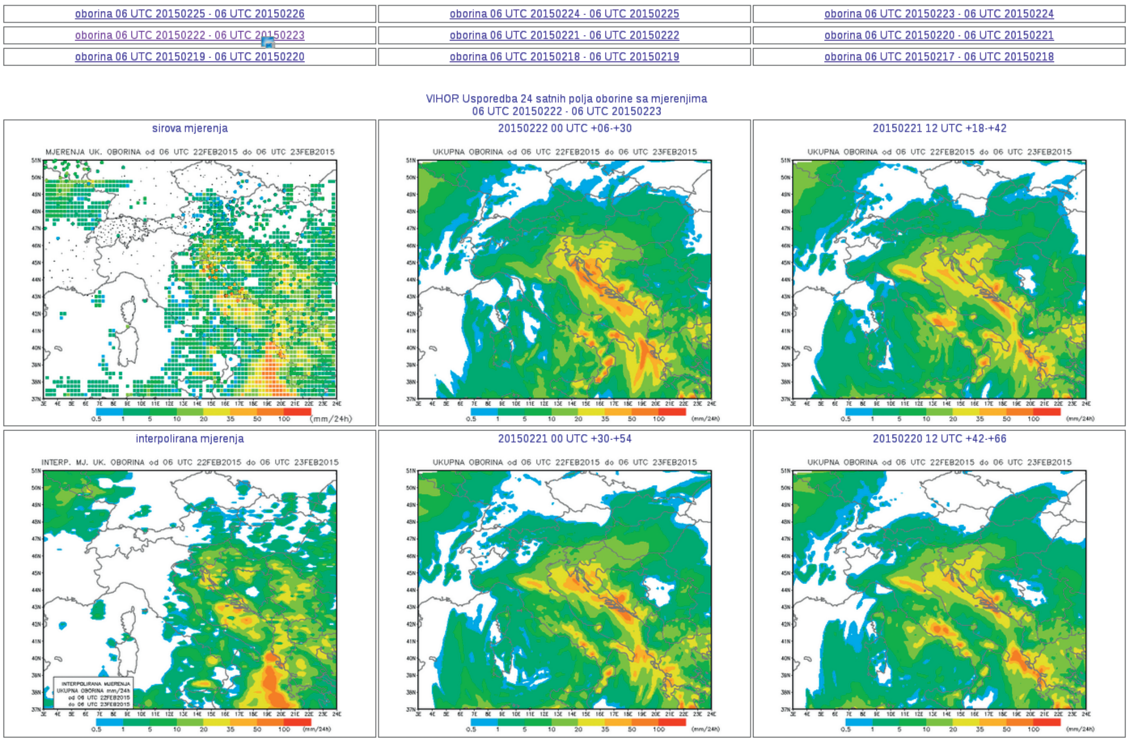

Figure 8. Screen-shot of comparison of precipitation field obtained using a combination of different measurements and precipitation forecast accumulated during 24 hours for different model runs and forecast times (initial time, start and end of accumulation forecast period are written above figures) for the period from 06 UTC 22 to 06 UTC 23 February 2015 for the whole 8 km resolution domain. Upper left figures shows measured precipitation on rain gauges as circles and satellite estimate as squares at the point of measurement. Lower left figure shows interpolated measured data.

Slika 8. Prikaz stranice sa usporedbom polja oborine dobivenog kombinacijom više vrsta mjerenja i 24 satne akumulirane prognoze za različite početne trenutke pokretanja modela i različita prognostička razdoblja (trenutak inicijalizacije modela kao i prognostičko razdoblje nalaze se iznad slika) za razdoblje 06 UTC 22. do 06 UTC 23. veljače 2015. godine za cijelu domenu sa 8 km horizontalnom razlučivošću. Gornja lijeva slika prikazuje mjerenja sa kišomjernih postaja kao ispunjeni krugovi a pravokutnici prikazuju satelitsku procjenu oborine za područje mjerenja, Donja lijeva slika prikazuje interpolirane vrijednosti mjerenja.

Since precipitation has a strong spatial signature that is not always determined by local topography, it is more informative to compare the precipitation forecast to the measured values on a map. The measurements of accumulated 24 hourly precipitation are available from a rather dense observational network of rain gauges (see green triangles in Figure 1) and many stations report this parameter in the SYNOP reports at 6 UTC. However, these measured data are available only over land and even there the spatial density of the available measurements is not uniform (see the distribution of dots in the upper left panel in Figure 8). Figure 8 shows a screen-shot of the web page interface where a map showing the measured data (upper left panel) of the accumulated 24 hour precipitation from rain gauges (as circles), estimate using the satellite data (squares) and interpolated measurements field (lower left panel) compared to products from different forecast runs (only 4 runs are shown here from available 9 covering the same 24 hourly period) for the whole 8 km resolution domain. There are similar interfaces showing the same measurements and fields zoomed over Croatia and for the 2 km resolution forecast over the whole domain (not shown). In the top of the Figure 8 there are links to other dates. The analysis times for the forecast in the figure are written above the figure as well as the forecast ranges when the precipitation was accumulated. All figures show data for the same period as the measurements.

Here we show the operational products as they are produced in order to present them and explain their function and value. The measured data are plotted as reported without quality check, filtering or flagging. It is up to the meteorologist that studies the figures to conclude if the measured data of a meteorological quantity are trustworthy for a particular station. Comparing the figures for stations nearby can help in determining this. Sometimes different sources of measured data show different results. For example, precipitation measured in-situ can be considerably larger or lower than the one estimated using satellite data. The reason is that the clouds with highest tops do not necessarily yield largest precipitation amounts at the surface (Hamada et al. 2015). This is especially obvious for the severe precipitation events when rain-gauges measure more than 100 mm of precipitation in 24 hours while the satellite estimate do not exceed 30 mm. On the contrary, there are days, when accumulated precipitation from the satellite estimate yields more than 100 mm in 24 hours over vast areas where sparse in-situ measurements show amounts even below 20 mm. This was noticed while following the performance of operational forecast (Figure 8 is an example of what is plotted every day). The amounts measured on rain gauges were not corrected for under-catchment due to wind speed (because few stations are equipped with anemometer) and satellite estimates are, as the name says, only estimates computed indirectly and represent values for a larger horizontal scale.

4. STATISTICAL PACKAGES FOR OBJECTIVE VERIFICATION

4.1. VERAL verification package

VERAL is verification package developed at the Czech Hydrometeorological Institute (CHMI, Janoušek 1999). It is used to compute standard deviation, bias and root mean square error of model variables compared to observations. Statistics are computed over the model domain (or a smaller area as defined by the user) and over a time period. The model results are interpolated both horizontally and vertically to observation location and compared to synoptic and atmospheric sounding observations. There was no correction applied to temperature at 2 meters for the difference between the station elevation in the model and reality.

When two experiments are compared, a quality control of measured data is performed in a first step. It compares measurements with global model analysis in order to obtain "neutral" observation selection. This means that the selection should not favour one of the experiments when choosing the data. Unfortunately, this also means that the measured data are discarded if they are much different from the global model analysis (that not necessarily detects some small scale and local weather phenomena that can be present both in high resolution forecast and in real weather outside) not because they are just plain wrong. Departures are computed similarly as in the surface data assimilation scheme that uses optimum interpolation (Giard and Bazile 2000). One could switch off the quality control and simply skip this step, but this means that the verification procedure would use measured data that are completely wrong.
In second step selected observations are used for computation of model departures from observations for both experiments. Departures are then used to calculate different statistical scores like bias (BIAS), root mean square error (RMSE) and standard deviation (STD):
where F_i represents model value at observa-

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

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

$$STD = \left(RMSE^2 - BIAS^2\right)^{0.5} \quad (3)$$

tion location, O_i represents the observed value and N is number of measurements.

The VERAL verification package also includes a graphics utility based on the NCAR graphics software that plots the statistical scores. The example shown in Figure 9 shows the evolution of BIAS and RMSE with the forecast range for a set of surface parameters, such as the 2 m temperature, mean sea level pressure (denoted as geopotential in the fig-

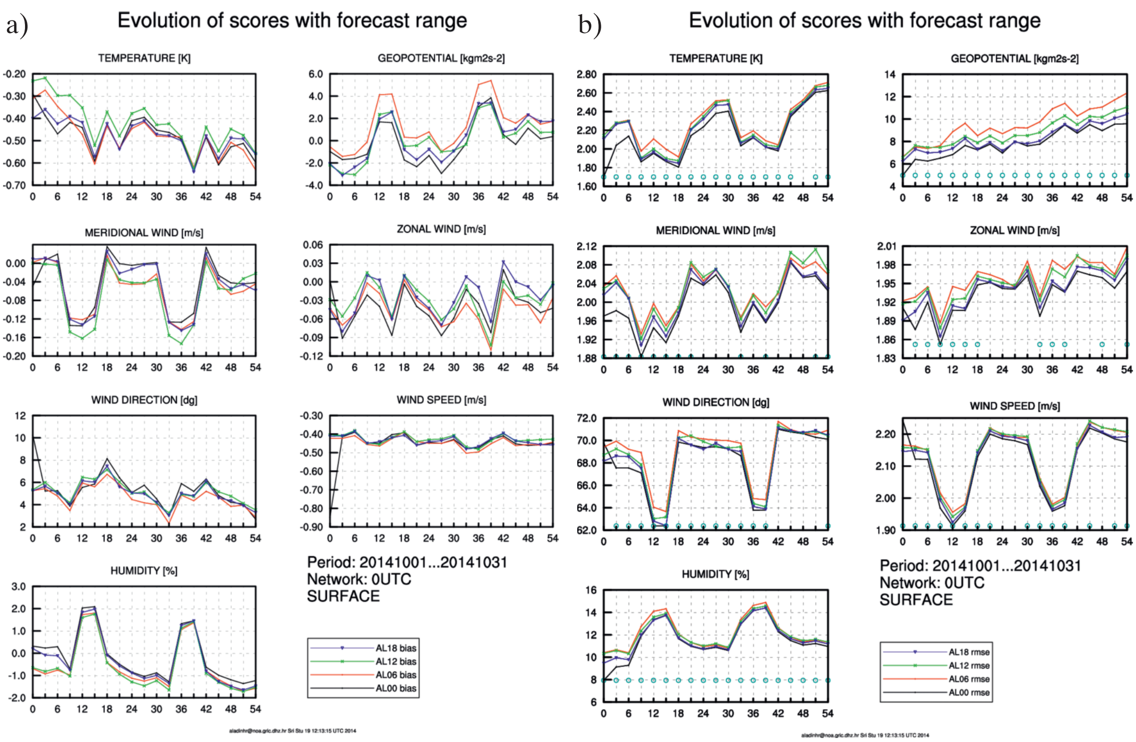

Figure 9. Bias (left) and root mean square error (right) for operational forecast in 8 km resolution initialized at different hours in the day, from 00 (black), 06 (red), 12 (green) and 18 (blue) UTC analyses for temperature (the scale is in Kelvin), mean sea level pressure (titled as geopotential), meridional and zonal wind components (in m/s), wind direction (in degrees) and speed (in m/s) and relative humidity (percent), the scores are computed for all operational forecasts run in for October 2014.

Slika 9. Pristranost srednjaka (lijevo) i korijen srednje kvadratične pogreške (desno) za operativnu prognozu vremena sa 8 km horizontalnom razlučivosti inicijalizirana u različitim trenucima u danu, za 00 (crno), 06 (crveno), 12 (zeleno) i 18 (plavo) UTC analize za temperaturu (skala u Kelvinima), tlak zraka reduciran na srednju razinu mora (označen kako "geopotential"), meridionalna i zonalna komponenta vjetra (m/s), smjer vjetra (stupnjevi) i brzina vjetra (m/s) i relativna vlažnost zraka (%), rezultati verifikacije izračunati su za sve operativne prognoze pokretane tijekom listopada 2014.

ure), meridional and zonal wind components, wind direction and speed and relative humidity (the title of each panel writes for which variable it is). The scores in Figure 9 are computed separately for runs starting at different times in a day, so BIAS and RMSE of 00 UTC forecasts are shown in black, 06 UTC forecasts in red, 12 UTC forecasts in green and 18 UTC in blue. A light blue circle in RMSE plots shows that the error is significant. The scores were computed for all forecasts that started during October 2014.

VERAL verification package also produces plots of vertical profiles of BIAS, RMSE and STD over the forecast range. Figure 10 shows BIAS and RMSE for 31 forecasts starting at 00 UTC during August 2013. The operational ALADIN forecast (panels with title AL00, and black lines in panels below) is compared to the ARPÉGE global model forecast that was used for the LBCs at the time (panels with title ARPE and red lines in panels below). The central panel shows the difference in biases and RMSEs, positive values (blue) mean that the BIAS and RMSE of ARPÉGE are larger than of the ALADIN forecast. ALADIN forecast has strong positive bias in relative humidity in the layer from 500 to 150 hPa (consequently RMSE is also rather large in that layer) but so does the ARPÉGE forecast. ALADIN uses data assimilation with 3D-var (Stanešić 2011). The positive bias is lower in ALADIN than in ARPÉGE at the beginning of the forecast run. This indicates that the excess moisture enters the domain through the lateral boundaries. Similarly, there is a negative bias in moisture in the layer below 500 hPa in the ARPÉGE forecasts that is reduced in the ALADIN forecast. Similarly to Figure 10, Figure 11 shows BIAS

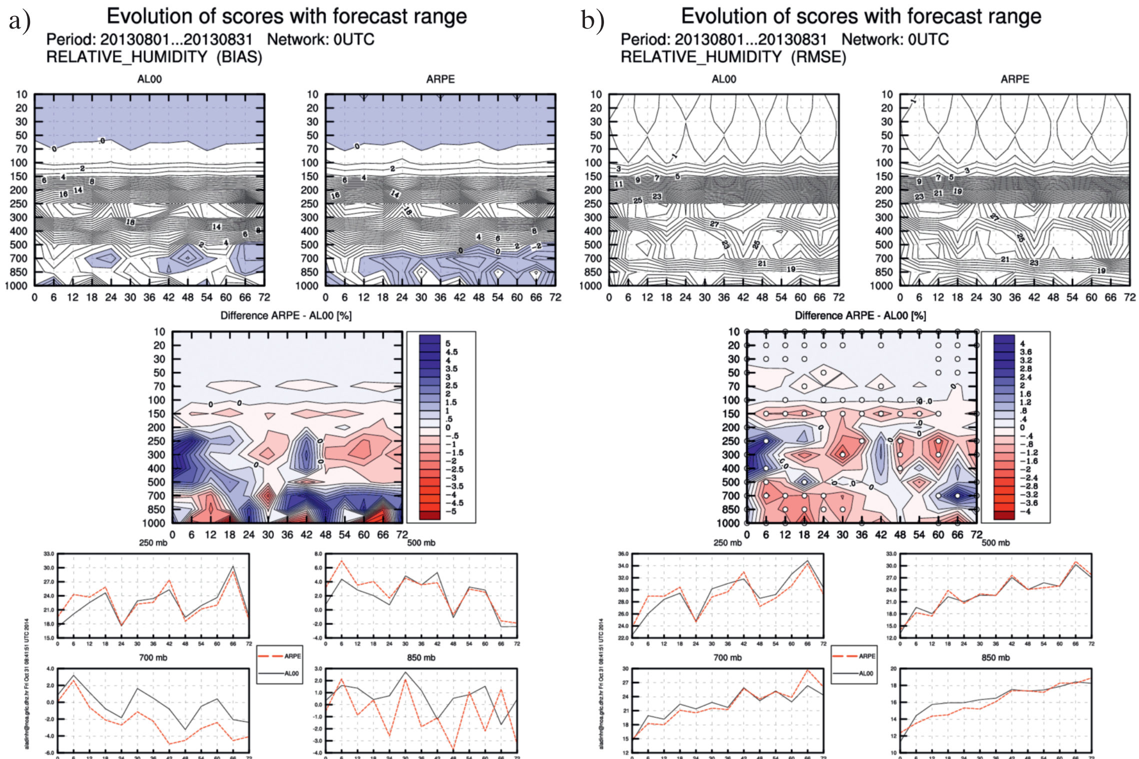

Figure 10. Bias (left) and root mean square error (right) for operational ALADIN forecast in 8 km resolution (top left panel and black lines in bottom panels) and ARPÉGE operational forecast (top right panel and red lines in bottom panels) initialized at 00 UTC, the scores are computed for all operational forecasts run in for August 2013 (when operational 8 km forecast was coupled to ARPÉGE). Top row shows vertical-time cross section of bias and root mean square error, the panel in the middle shows their difference, the dependence of bias and root mean square error on the forecast hour at different levels in the atmosphere is shown in bottom panels (the pressure on the level is written above the panel).

Slika 10. Pristranost srednjaka (lijevo) i korijen srednje kvadratične pogreške (desno) za ALADIN operativnu prognozu sa 8 km horizontalnom razlučivosti (gornji lijevi paneli te crna linija u donjim grafovima) i "ARPÉGE" operativna prognoza (gornji desni paneli te crvena linija u donjim grafovima) inicijalizacija modela u 00 UTC, rezultati verifikacije su izračunati za sve operativne prognoze pokretane u kolovozu 2013. godine (kada su za rubne uvjete korištene prognoze "ARPÉGE" modela). Najgornji red prikazuje vertikalne vremenske presjeke srednje pristranosti i korijena srednje kvadratične pogreške, panel u sredini prikazuje njihove razlike, razlike srednje pristranosti i korijena srednje kvadratične pogreške u ovisnosti o prognostičkom razdoblju i različitim visinama u atmosferi prikazan je u donjim grafovima (iznad grafa nalazi se tlak za koji je nivo prikazana usporedba).

and RMSE for 31 forecasts starting at 00 UTC during January 2014 computed for the operational ALADIN forecast (panels with title AL00, and black lines in panels below) and the IFS global model forecast run in ECMWF that was used for the LBCs at the time (panels with title IFSC and red lines in panels below). The central panel shows the difference in biases and RMSEs, positive values (blue) mean that the BIAS and RMSE of IFS are larger than of the ALADIN forecast. ALADIN forecast again has strong positive bias in relative humidity in even thicker layer from 850 to 150 hPa, but the BIAS of the IFS forecast keeps the same pattern. The positive bias is again lower in ALADIN than in IFS at the beginning of the forecast run showing that the excess moisture enters the domain through the lateral boundaries. Similarly, there is a negative bias in moisture in the layer below 850 hPa in the IFS forecasts. One should keep in mind that the observations were quality checked against the global model analyses of ARPÉGE. Measurements very far from the analysis were thrown away. The choice of ARPÉGE means that the data were checked against an independent source (not IFS nor ALADIN that are compared here).

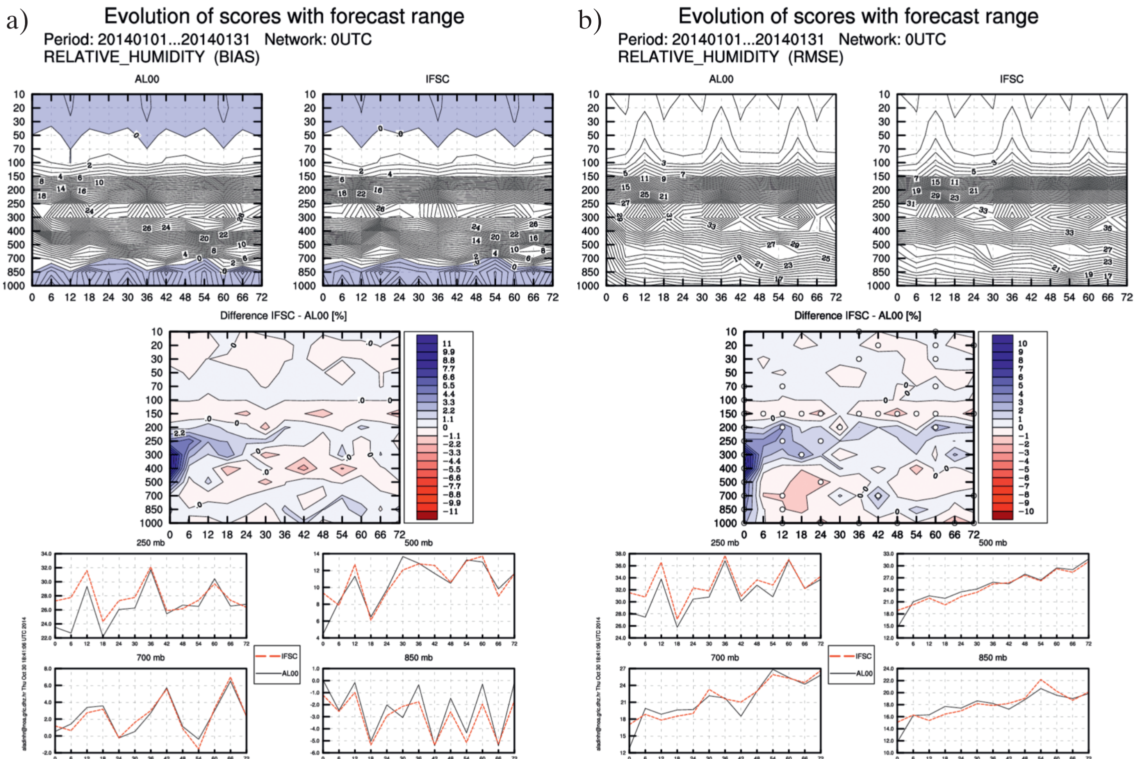

Figure 11. Bias (left) and root mean square error (right) for operational ALADIN forecast in 8 km resolution (top left panel and black lines in bottom panels) and IFS operational forecast (top right panel and red lines in bottom panels) initialized at 00 UTC, the scores are computed for all operational forecasts run in for January 2014 (when operational 8 km forecast was coupled to IFS). Top row shows vertical-time cross section of bias and root mean square error, the panel in the middle shows their difference, the dependence of bias and root mean square error on the forecast hour at different levels in the atmosphere is shown in bottom panels (the pressure on the level is written above the panel).

Slika 11. Pristranost srednjaka (lijevo) i korijen srednje kvadratične pogreške (desno) za ALADIN operativnu prognozu sa 8 km horizontalnom razlučivosti (gornji lijevi paneli te crna linija u donjim grafovima) i "IFS" operativna prognoza (gornji desni paneli te crvena linija u donjim grafovima) inicijalizacija modela u 00 UTC, rezultati verifikacije su izračunati za sve operativne prognoze pokretane u siječnju 2014. godine (kada su za rubne uvjete korištene prognoze "IFS" modela). Najgornji red prikazuje vertikalne vremenske presjeke srednje pristranosti i korijena srednje kvadratične pogreške, panel u sredini prikazuje njihove razlike, razlike srednje pristranosti i korijena srednje kvadratične pogreške u ovisnosti o prognostičkom razdoblju i različitim visinama u atmosferi prikazan je u donjim grafovima (iznad grafa nalazi se tlak za koji je nivo prikazana usporedba).

The quality of LBCs has a substantial impact on the quality of the LAM forecast that uses them, especially in situations, when their temporal resolution is insufficient to properly resolve features that rapidly enter the LAM domain through its lateral boundaries (Tudor and Termonia 2010). Such cases can lead to large errors especially in mean sea level pressure. LAM can be tuned using the available parameters in the physics to reduce and even remove excess moisture from the upper troposphere (that persists in verification scores for each month over the Croatian operational domain in 8 km resolution that covers the area shown in Figure 8, the number of radiosondes used varies from 2 to 30). But the excess moisture arrives from the lateral boundaries and LAM is then forced to remove an error that is caused somewhere else (in the global model fields used for LBCs). With such a tuning, LAM is forced to do something unnatural. Although the scores do look better, the forecast fields do not necessarily improve. There is more rainfall along lateral boundaries with inflow conditions (where wind brings moisture into the domain) with the new tuning than with the original one (these results are not shown here due to imposed length of the paper).

4.2. HARMONIE verification package

HARMONIE is an atmospheric modelling system used for numerical weather prediction and regional climate studies. The package includes a model code able to run at least three different NWP model configurations, such as HIRLAM, ALARO and AROME. It also includes a verification package that enables intercomparison of forecasts computed by different models and model versions. The verification package in HARMONIE (Yang, 2008, https://hirlam.org/trac/wiki/HarmonieSystemDocumentation/PostPP/Verification#no1) is designed to be a self contained package that computes verification scores using pre-extracted model and observational data. The observational data for the HARMONIE package is obtained through the ECMWF using the quality control, flagging and filtering procedure implemented there. The observations are checked against the global model analysis and discarded if the difference is larger than a prescribed threshold. This procedure again discards the data measured in some localized but intensive weather events not predicted by the global model. The package calculates a number of standard verification scores:

- Error as function of forecast lead time summarises the bias and root mean square error and their growth rate with the forecast range over a set of forecasts.
- Time sequences and vertical profiles show how the model data and the error characteristics are distributed in time or in the vertical.
- Error charts and tables show how a particular error is distributed in space, and stationwise linear correlation.
- Scatter plots show the correspondence between forecast and observed values.
- Mean diurnal cycles show how the mean model error changes in the course of the day.
- Histograms show the correspondence between the distributions of forecast and observed values.
- Student t-test shows how reliable are the differences between different experiments (model forecasts).

Additionally, there is a number of scores based on contingency tables such as: frequency bias (bias score), hit rate (probability of detection), false alarm ratio, false alarm rate, threat score, the equitable threat score, Hansen-Kuipers score, extreme dependency scores. The scores can be presented per station for the whole data set or filtered through different selection criteria based on e.g. a geographical domain, station elevation or properties of the data itself.

Model data are extracted and interpolated to observation location before being compared to the measured data. Only horizontal interpolation is performed for surface stations. For temperature, there is a choice to compare temperature or height adjusted temperature (with dry adiabatic lapse rate adjustment). Figure 12 shows errors computed with the

a) b)

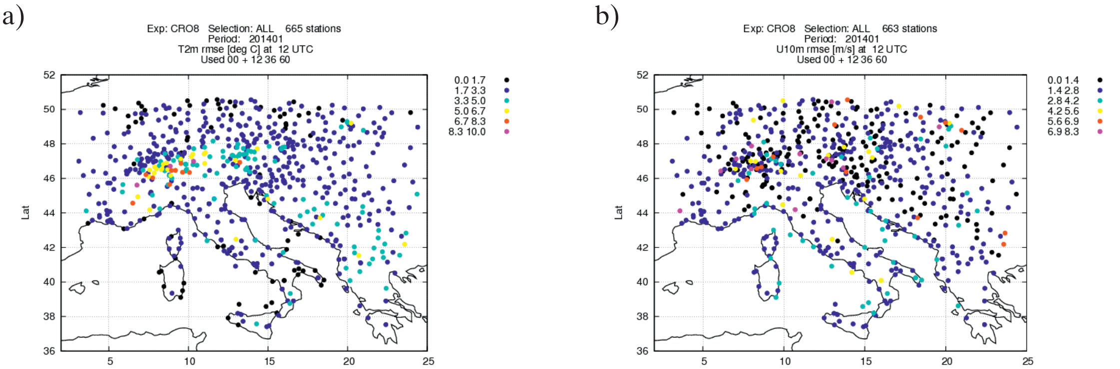

Figure 12. A map of the root mean square errors of 2 m temperature (a) and 10 m wind (b) at 12 UTC for January 2014, computed from 00 UTC forecasts (forecast ranges 12, 36 and 60 hours) 8 km resolution ALADIN model run in Croatia.

Slika 12. Karta srednje kvadratične pogreške za temperaturu zraka na 2 m (a) i brzinu vjetra na 10 m (b) za 12 UTC tijekom siječnja 2014. godine, za sva pokretanja modela u 00 UTC (za prognostička razoblja 12, 36 i 60 sati) sa hrvatskom verzijom ALADIN modela horizontalne razlučivosti 8 km.

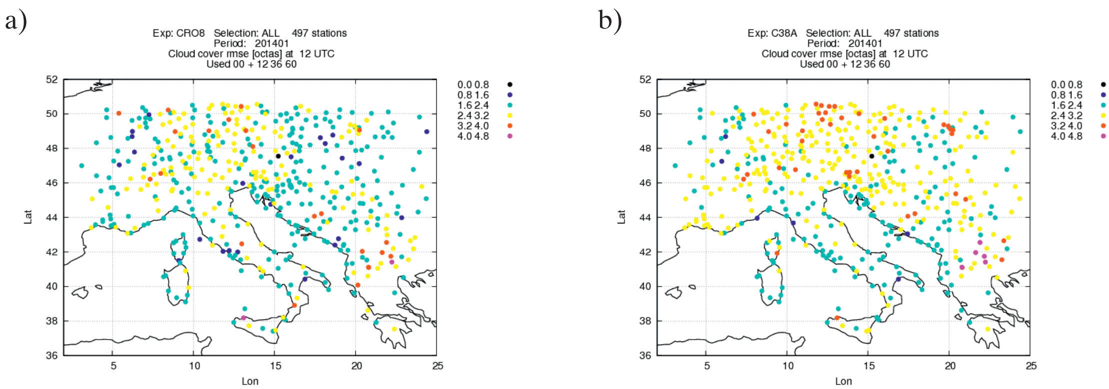

Figure 13. A map of the root mean square errors of cloud cover for the operational ALADIN forecast (a) and using the ALARO0 baseline physics package (b) at 12 UTC for January 2014, computed from 00 UTC forecasts (forecast ranges 12, 36 and 60 hours) 8 km resolution model run.

Slika 13. Karta srednje kvadratične pogreške ukupne naoblake za operativnu ALADIN prognozu (a) i prognozu gdje je korišten "ALARO0 baseline" paket fizikalnih parametrizacija (b) u 12 UTC za siječanj 2014. godine, izračunata za sva pokretanja modela u 00 UTC (prognostička razdoblja 12, 36 i 60 sati) sa modelom horizontalne razlučivosti 8 km.

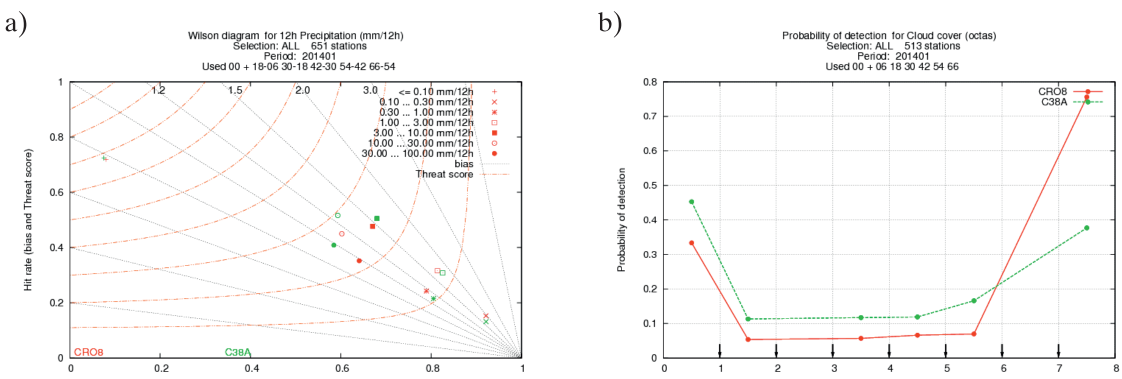

Figure 14. Wilson diagram for 12 hourly precipitation (a) and probability of detection for cloud cover (b) for January 2014, computed from 00 UTC operational forecasts in 8km resolution (CRO8, red) and experiments with ALARO0 baseline (C38A, green).

Slika 14. Wilsonov dijagram za 12 satnu akumuliranu oborinu (a) i za vjerojatnost detekcije naoblake (b) za siječanj 2014, izračunat za sva pokretanja operativnog modela u 00 UTC sa 8 km horizontalnom razlučivosti (CRO8, crveno) i eksperimentalne verzije modela sa ALARO0 baseline paketom fizikalnih parametrizacija (C38A, zeleno).

height adjustment still have larger values over mountainous areas. Measurements used in comparison are GTS measurements (SYNOP, SHIP, TEMP, AMDAR) obtained from ECMWF in a correct format that can be used by the HARMONIE verification software

HARMONIE verification package is installed locally in the national weather services in order to use it for comparison of different local model configurations. Additionally, a common HARMONIE verification for LACE countries has been established and it is used to compare various operational configurations of LACE countries (Austria, Croatia, Czech Republic, Hungary, Romaina, Slovakia, Slovenia) the products are available at http://www.rclace.eu/dynamic/extra/monitor/. Maps of RMSEs (Figure 12) of 2 m temperature and 10 m wind show unequal spatial density of observations and that the errors are the largest in the mountainous areas. Unfortunately, there are seldom any measurements from above the sea surface available for the verification of the surface parameters. Even those measurements that appear to be taken

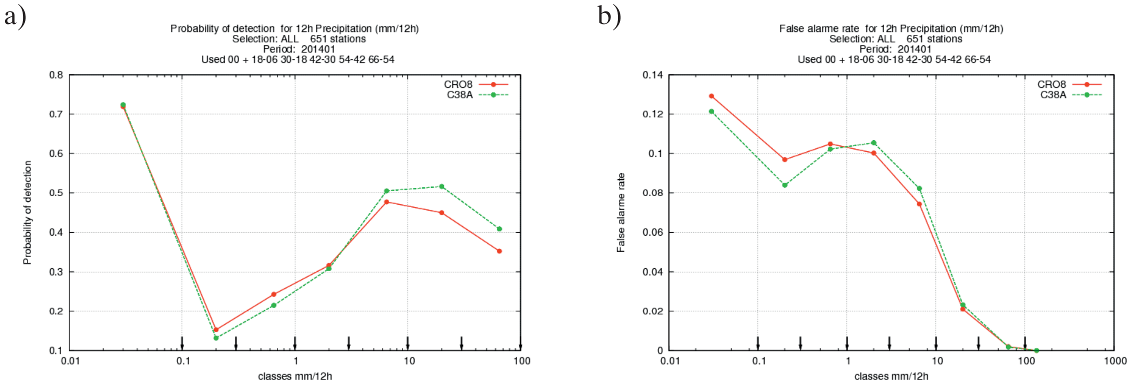

Figure 15. Probability of detection (a) and false alarm rate for 12 hourly precipitation (b) for January 2014, computed from 00 UTC operational forecasts in 8 km resolution (CRO8, red) and experiments with ALARO0 baseline (C38A, green).

Slika 15. Vjerojatnost detekcije naoblake (a) i učestalost lažnog alarma za 12 satnu oborinu (b) za siječanj 2014, izračunat za sva pokretanja operativnog modela u 00 UTC sa 8 km horizontalnom razlučivosti (CRO8, crveno) i eksperimentalnog modela sa ALARO0 baseline paketom fizikalnih parametrizacija (C38A, zeleno).

from above the sea, were actually taken on the islands situated there. If the NWP model contains a sea point at the same location, the forecast wind field is expected to have a larger error than it would have if this was a land point. HARMONIE verification package has been used to validate the ALARO0 baseline (Brožkova, private communication, http://www.rclace.eu/File/Physics/2012/ALARO_0_baseline_Dec2012.pdf) physics package in the framework of the operational forecast performance. Unfortunately, the standard setup of ALARO0 baseline in ALADIN model, code version CY38T1 yields larger errors than the ALADIN version operational at the time (Tudor et al. 2013) in the forecast of cloud cover (Figure 13) for the winter period (January 2014) when the stratiform cloudiness prevailed. This is due to a random maximum overlap assumption used for ALARO0 baseline while the operational version used random overlap assumption. Model forecast of clear and partly cloudy weather improves while the quality of the forecast of prevailing cloudy weather is deteriorated (Figure 14). The scores are better in the weather characterized by shallow convection clouds. This is because of the new tunings and developments diagnose the cloud cover better in shallow convection. The cloud cover due to the shallow convection is not diagnosed directly since their contribution to the evolution of the model variables is computed in the turbulence scheme (Geleyn 1987).

The precipitation forecasts of the operational suite and tests performed with the ALARO0 baseline have shown that the model has the lowest probability of detection (both in Wilson diagram in Figure 14a and in Figure 15a) and largest false alarm rate (Figure 15b) for events of weak precipitation (around 1mm in 12 hours) while the forecast reliability is larger for intensive precipitation events. A small error of 1 mm in 12 hours is insignificant in intensive precipitation events, but can be crucial when verifying weak precipitation. Similar effect is visible in performance of both old operational forecast and ALARO0 baseline (Figure 14a).

5. CONCLUSIONS

This article presents operational activities related to verification and validation of the operational forecast as well as pre-operational testing of the new model versions. The operational products contain figures that allow subjective evaluation of the forecast quality in the current weather situation. Maps of measured, satellite estimate and forecast precipitation are a good indicator of available high impact precipitation events that can be used in detailed studies especially through its web interface that allows rapid assessment of consecutive days.

The objective verification scores are computed regularly and verification data from the operational forecast are sent to centres where forecasts from different NMSs are compared. Regular and operational objective verification requires measured data that are available, quality checked and filtered in real time using an automated procedure. Common verification projects rely on measurements that are usually compared against a global model analysis and discards data that are too far from it. Local implementation of a validation software also requires an automated procedure for checking and filtering of measured data.

The statistical and spectral evaluation is computed for the operational versions of ALADIN meso-scale NWP model. For spectral evaluation, please see Hrastinski et al. in this issue. Statistical verification included several statistical measures, such as systematic error, mean squared error, mean absolute error and others. The decomposition of the mean squared error on its components: the bias of the mean, the bias of the standard deviation and dispersion or phase error (Hrastinski et al. in this issue) shows that the dispersion errors, i.e. phase errors are the main reason for the model error, especially in models of higher resolution. In other words, the analysis indicates that errors in the time of start or end of a certain process, including the creation or disappearance of wind flow, are the main reason for the error in the ALADIN model. Spectral verification is performed by spectral decomposition in wave-number and frequency domains. Such spectral decomposition provides information on the performance of the model in simulating the processes of certain time scales at specific location and is useful for the physical interpretation of the results.

There is a number of potentially useful tools to be implemented, especially for the verification of precipitation. One of them is the SAL method that evaluates the structure (S), amplitude (A), and location (L) of the precipitation field (Wernli et al. 2008). A new verification score, developed by ECMWF is Stable Equitable Error in Probability Space (SEEPS, Rodwel et al. 2010) in order to monitor the long-term trend in the quality of precipitation forecast. The tool evaluates 24 hourly accumulated precipitation against observed precipitation amounts. The precipitation amount is partitioned into three categories: "dry", "light" and "heavy" precipitation depending on the station climatology. That way, the SEEPS assesses salient features of the local weather and accounts for climate differences between stations (Haiden et al. 2012).

ACKNOWLEDGEMENTS

The authors thank two anonymous reviewers for their constructive comments that have significantly improved the manuscript.

LITERATURE

Acker, J.G. and G. Leptoukh, 2007. Online Analysis Enhances Use of NASA Earth Science Data, Eos, Trans. AGU, Vol. 88, No. 2.

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.

Brožková, R., M. Derková, M. Belluš, and A. Farda, 2006. Atmospheric forcing by ALADIN/MFSTEP and MFSTEP oriented tunings. *Ocean Sci.*, 2, 113-121.

Drobinski, P., V. Ducrocq, P. Alpert, E. Anagnostou, K. Béranger, M. Borga, I. Braud, A. Chanzy, S. Davolio, G. Delrieu, C. Estournel, N. Filali Boubrahmi, J. Font, V. Grubišić, S. Gualdi, V. Homar, B. Ivančan-Picek, C. Kottmeier, V. Kotroni, K. Lagouvardos, P. Lionello, M.C. Llasat, W. Ludwig, C. Lutoff, A. Mariotti, E. Richard, R. Romero, R. Rotunno, O. Roussot, I. Ruin, S. Somot, I. Taupier-Letage, J. Tintore, R. Uijlenhoet and H. Wernli, 2014. HyMeX: A 10-Year Multidisciplinary Program on the Mediterranean Water Cycle. *Bull. Amer. Meteor. Soc.*, 95, 1063-1082. doi: http://dx.doi.org/10.1175/BAMS-D-12-00242.1

Ferro, C.A.T. and D.B. Stephenson, 2011. Extremal Dependence Index: Improved verification measures for deterministic forecasts of rare binary events. *Wea. Forecasting*, 26, 699-713.

Geleyn, J.-F., 1987. Use of a modified Richardson number for parameterizing the effect of shallow convection. In: Matsuno Z. (ed)., Short and medium range weather prediction, Special volume of *Jour. Meteor. Soc. Japan*, 141-149.

Geleyn J.-F., 1988. Interpolation of wind, temperature and humidity values from model levels to the height of measurement. *Tellus A,* 40A, 347-351.

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.

Grazzini, F. 2007. Predictability of a large-scale flow conducive to extreme precipitation over the western Alps. *Meteorology and Atmospheric Physics*, 95, 3-4, 123-138 .

Hally, A., Richard, E. and Ducrocq, V., 2014. An ensemble study of HyMeX IOP6 and IOP7a: sensitivity to physical and initial and boundary condition uncertainties. *Nat. Hazards Earth Syst. Sci.*, 14, 1071-1084.

Haiden, T., M. Rodwell, D. Richardson, A. Okagaki, T. Robinson, and T. Hewson, 2012. Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Mon. Wea. Rev.*, 140, 2720-2733. (Also available as ECMWF Technical Memorandum 665)

Hamada, A., Y.N. Takaybu, C. Liu, and E.J. Zipser, 2015. Weak linkage between the heaviest rainfall and tallest storms. *Nature Communications.* DOI:10.1038/ncomms7213

Hrastinski, M., K. Horvath, I. Odak Plenković, S. Ivatek-Šahdan, and A. Bajić, 2015, Verification of the operational 10 m wind forecast obtained with the ALADIN mesoscale numerical weather prediction model. *Cro. Meteorol. J.* 50, 105-120.

Huffman, G.J., D.T. Bolvin, E.J. Nelkin, D.B. Wolff, R.F. Adler, B. Gu, Y. Hong, K.P. Bowman and E.F. Stocker, 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales, *J. Hydrometeorol.*, 8, 38-55.

Ivatek-Šahdan, S. and M. Tudor, 2004. Use of high-resolution dynamical adaptation in operational suite and research impact studies. *Meteorol Z* 13(2):1-10

Janoušek, M., 1999: Verification package VERAL (Description and User`s Guide). Technical documentation, RCLACE, 7 pp.

Jansa, A., P. Alpert, P. Arbogast, A. Buzzi, B. Ivančan-Picek, V. Kotroni, M.C. Llasat, C. Ramis, E. Richard, R. Romero and A. Speranza, 2014. MEDEX: a general overview, *Nat. Hazards Earth Syst. Sci.*, 14, 1965-1984.

Jolliffe, I.T. and D.B. Stephenson, (Eds) 2011. Forecast verification: A practitioner's guide in atmospheric science, 2nd edditions, John Wiley & sons Ltd., Chichester, 292 pp.

Odak Plenković, I., L. Delle Monache, K. Horvath, M. Hrastinski and A. Bajić, 2015. Post-processing of ALADIN wind Speed Predictions with an Analog-based Method. *Cro. Meteorol. J.* 50, 121-136.

Rodwell, M. J., D.S. Richardson, T.D. Hewson and T. Haiden, 2010. A new equitable score suitable for verifying precipitation in numerical weather prediction. *Q.J.R.Meteorol. Soc*, 136, 1344-1363. (Also available as ECMWF Technical Memorandum 615)

Romero, R., 2011. Application of Factor Separation to heavy rainfall and cyclogenesis: Mediterranean examples, in: *Factor Separation in the Atmosphere, Applications and Future Prospects*, edited by: Alpert, P. and Sholokhman, T., Cambridge University Press, Cambridge, UK, 87-119.

Rossa, A., P. Nurmi and E. Ebert, 2008. Overview of methods for the verification of quantitative precipitation forecasts. In. *Precipitation: Advances in Measurement, Estimation and Prediction*. Ed. Michaelides, S. Springer Berlin Heidelberg. pp. 419-452.

Speranza, A., Buzzi, A., Trevisan, A., and Malguzzi, P., 1985. A theory of deep cyclogenesis in the lee of the Alps. Part I: Modifications of baroclinic instability by localized topography, *J. Atmos. Sci.*, 42, 1521-1535.

Stanešić, A., 2011. Assimilation system at DHMZ: Development and first verification results, *Cro. Meteorol. J.* 44/45, 3-17.

Tsai, P. and B.E. Doty, 1998: A Prototype Java Interface for the Grid Analysis and Display System (GrADS). *Fourteenth International Conference on Interactive Information and Processing Systems*, Phoenix, AZ 11-16 January 1998.

Tudor M., 2010. Impact of horizontal diffusion, radiation and cloudiness parameterization schemes on fog forecasting in valleys. *Met. Atm. Phy.* Vol.108, 57-70.

Tudor, M., 2013. A test of numerical instability and stiffness in the parametrizations of the ARPÉGE and ALADIN models. *Geoscientific model development*, 6, 901-913.

Tudor, M. and S. Ivatek-Šahdan, 2002. The MAP-IOP 15 case study. *Cro. Meteorol. J.* 37, 1-14.

Tudor, M. and S. Ivatek-Šahdan, 2010. The case study of bura of 1st and 3rd February 2007, *Meteorol. Z.,* 19, 453-466.

Tudor, M. and P. Termonia, 2010. Alternative formulations for incorporating lateral boundary data into limited-area models. *Mon. Wea. Rev.* 138, 2867-2882.

Tudor, M., Ivatek-Šahdan, S., Stanešić, A., Horvath, K., and Bajić, A., 2013. Forecasting weather in Croatia using ALADIN numerical weather prediction model. In: *Climate Change and Regional/Local Responses* / Zhang, Yuanzhi ; Ray, Pallav (eds.). Rijeka: InTech, 59-88.

Wernli, H., M. Paulat, M. Hagen and C. Frei, 2008, SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Mon. Wea. Rev.*, 136, 4470-4487. doi: http://dx.doi.org/10.1175/2008MWR2415.1

Wulfmeyer, V., A. Behrendt, C. Kottmeier, U. Corsmeier, C. Barthlott, G.C. Craig, M. Hagen, D. Althausen, F. Aoshima, M. Arpagaus, H.-S. Bauer, L. Bennett, A. Blyth, C. Brandau, C. Champollion, S. Crewell, G. Dick, P. Di Girolamo, M. Dorninger, Y. Dufournet, R. Eigenmann, C. Flamant, T. Foken, T. Gorgas, M. Grzeschik, J. Handwerker, C. Hauck, H. Höller, W. Junkermann, N. Kalthoff, C. Kiemle, S. Klink, M. König, L. Krauss, C.N. Long, F. Madonna, S. Mobbs, B. Neininger, S. Pal, G. Peters, G. Pigeon, E. Richard, M.W. Rotach, H. Russchenberg, T. Schwitalla, V. Smith, R. Steinacker, J. Trentmann, D.D. Turner, J. van Baelen, S. Vogt, H. Volkert, T. Weckwerth, H. Wernli, A. Wieser and M. Wirth, 2011. The Convective and Orographically-induced Precipitation Study (COPS): the scientific strategy, the field phase, and research highlights. *Q.J.R. Meteorol. Soc.*, 137,3-30. doi:10.1002/qj.752

Yang, X. 2008, Development of Hirlam/Harmonie monitoring system. *HIRLAM Newsletter* no 54, 62-65.

Žagar, M., Rakovec, J. (1999): Small-scale surface wind prediciton using dynamical adaptation. *Tellus A*, 51, 489-504.