Geosci. Model Dev. Discuss., 8, 1029–1075, 2015
www.geosci-model-dev-discuss.net/8/1029/2015/
doi:10.5194/gmdd-8-1029-2015

Geoscientific Model Development Discussions
Open Access

Evaluation of the high resolution WRF-Chem air quality forecast and its comparison with statistical ozone predictions

R. Žabkar[1,2], L. Honzak[2,*], G. Skok[1,2], R. Forkel[3], J. Rakovec[1,2], A. Ceglar[4,2,], and N. Žagar[1,2]**

[1]University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia
[2]Center of Excellence SPACE-SI, Ljubljana, Slovenia
[3]Karlsruher Institut für Technologie, Institut für Meteorologie und Klimaforschung, Atmosphärische Umweltforschung, Garmisch-Partenkirchen, Germany
[4]University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia
[*]now at: BO-MO d.o.o., Ljubljana, Slovenia
[**]now at: Institute for Environment and Sustainability, Joint Research Centre, Ispra, Italy

Received: 15 January 2015 – Accepted: 17 January 2015 – Published: 4 February 2015

Correspondence to: R. Žabkar (rahela.zabkar@fmf.uni-lj.si)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

An integrated high resolution modelling system based on the regional on-line coupled meteorology-atmospheric chemistry WRF-Chem model has been applied for numerical weather prediction and for air quality forecast in Slovenia. In the study an evaluation of the air quality forecasting system has been performed for summer 2013. In the case of ozone (O_3) daily maxima the first day and second day model predictions have been also compared to the operational statistical O_3 forecast and to persistence. Results of discrete and categorical evaluations show that the WRF-Chem based forecasting system is able to produce reliable forecasts, which depending on monitoring site and the evaluation measure applied can outperform the statistical model. For example, correlation coefficient shows the highest skill for WRF-Chem model O_3 predictions, confirming the significance of the non-linear processes taken into account in an on-line coupled Eulerian model. For some stations and areas biases were relatively high due to highly complex terrain and unresolved local meteorological and emission dynamics, which contributed to somewhat lower WRF-Chem skill obtained in categorical model evaluations. Applying a bias-correction could further improve WRF-Chem model forecasting skill in these cases.

1 Introduction

Real-time air quality forecasting (RT-AQF) is a relatively new discipline in atmospheric sciences, which has evolved as a response to societal and economic needs, reflecting the progress in scientific understanding of physical processes and numerical and computational technologies (Zhang et al., 2012a). First RT-AQF systems, developed for forecasting air pollution in exposed urban regions, were either empirical methods based on persistence, climatology, human expertise and meteorological forecast (e.g. Wolff and Lioy, 1978), or statistical models taking advantage of existing links between pollutant concentrations, meteorological variables (wind speed and direction, tempera-

GMDD
8, 1029–1075, 2015

ture, cloudiness, moisture etc.) and physical (emissions) parameters (e.g. McCollister and Wilson, 1975; Cobourn, 2007; Vlachogianni et al., 2011). The next step in evolution of RT-AQF systems was the use of sophisticated chemical transport models that represent all major processes (meteorological and chemical) that lead to the formation and accumulation of air pollutants. Many of these RT-AQF systems consist of an offline coupled meteorological model and a chemical-transport model, where the meteorological model (e.g., ALADIN, ALADIN International Team, 1997; MM5, Grell et al., 1995; WRF, Skamarock et al., 2008) provides meteorological input for the chemical-transport model (e.g., EMEP, van Loon et al., 2004; CMAQ, Byun and Schere, 2006; CAMx, ENVIRON, 2011; CHIMERE, Menut et al., 2013) with an output time interval typically around 1 h. Examples are the EURAD (http://db.eurad.uni-koeln.de/index_e.html), SILAM (http://silam.fmi.fi/), ForeChem (http://atmoforum.aquila.infn.it/forechem/), CALIOPE (http://www.bsc.es/caliope/) forecast systems and others. The new generation of an online coupled models (e.g., MCCM, Grell et al., 2000; GATOR-GCMM, Jacobson, 2001; Meso-NH-C, Tulet et al., 2003; WRF-Chem, Grell et al., 2005; Enviro-HIRLAM, Baklanov et al., 2008; GEM-AQ, Kaminski et al., 2008; COSMO-ART, Vogel et al., 2009; WRF-Chem-MADRID, Zhang et al., 2010a) presents an alternative approach with one unified modelling system, in which meteorological and air quality variables are simulated together within the same model. The online approach permits the simulation of two-way interactions between different atmospheric processes including emissions, chemistry, clouds and radiation, and a better response of the simulated pollutant transport to changes of the wind field (Grell et al., 2004), and can thus provide a more realistic representation of the atmosphere. The use of online coupled models can be particularly important in regions with high aerosol loadings and cloud coverage (Otte et al., 2005; Eder et al., 2006), where physical processes in the atmosphere may be modified by the aerosol direct effect on radiation or by aerosol cloud interactions. Several reviews summarized the strengths and limitations of offline and online coupled models (e.g. Zhang, 2008; Klein, 2012; Baklanov et al., 2014).There is an increasing awareness that an integrated online approach is needed not only for assessment, forecasting

GMDD
8, 1029–1075, 2015

and communication of air quality, but also for weather forecasting (e.g. Baklanov, 2010; Grell and Baklanov, 2011; Klein et al., 2012; Zhang et al., 2012b; Baklanov et al., 2014). Nevertheless, there are several issues regarding the inclusion of chemistry into numerical weather prediction models. More evidence is required whether an integrated model can produce a good climatology of the most important chemical species, and if such a model is, considering many uncertainties, able to beat persistence forecasts of these species (Grell and Baklanov, 2011). These questions are calling for further research and studies exploring the performance of the models with an online coupled chemistry.

In recent years extensive efforts have been devoted to develop air quality (AQ) forecasting systems for Slovenia. In this study we explore the use of the state-of-the-science WRF-Chem model (Grell et al., 2005, 2011) with coupled meteorological, microphysical, chemical, and radiative processes for forecasting AQ in Slovenia during the summertime conditions. In last decade WRF-Chem has been increasingly applied to many areas worldwide (e.g., Misenis and Zhang, 2010; Fast et al., 2009; Zhang et al., 2010a, b; Li et al., 2011; Tie et al., 2009; Hu et al., 2012; Forkel et al., 2012, Žabkar et al., 2011, 2013). In most of these studies WRF-Chem model has been successfully used to simulate historical poor AQ conditions in hindcast approach. To our knowledge, only a few studies focused on using WRF-Chem for forecasting AQ, most of these have applied WRF-Chem forecast before and during field campaigns (McKeen et al., 2005, 2007, 2009; Yang et al., 2011). Takigawa et al. (2007) evaluated O_3 forecast for a 1 month time period from an one-way nested global-regional RT-AQF system with full chemistry based on the global CHASER (Sudo et al., 2002) and regional WRF-Chem models, while Saide et al. (2011) evaluated forecast system based on the WRF-Chem model for simulating carbon monoxide (CO) as a $PM_{10}/PM_{2.5}$ surrogate over Santiago de Chile for wintertime conditions. WRF-Chem-MADRID (Zhang et al., 2010a) with two additional gas-phase mechanisms, sectional representation for particle size distribution and more advanced model treatments compared to WRF-Chem, was applied by Chuang et al. (2011) and by Yahya et al. (2014) for forecasting AQ over the Southeastern US. In spite of a limited number of evalu-

GMDD
8, 1029–1075, 2015

ation studies published in the literature, an increasing number of real-time weather and air quality forecasting systems based on WRF-Chem is performed worldwide (http://ruc.noaa.gov/wrf/WG11/Real_time_forecasts.htm).

In our study we explore the forecasting skill of WRF-Chem model over topographically complex and geographically diverse area of Slovenia for three summer months (June–August 2013). Furthermore, in the case of O_3 we compare WRF-Chem predictions with statistical model for predicting O_3 daily maxima, currently used at Slovenian Environment Agency (SEA). Both first day (1 day) and second day (2 day) forecasts are considered, while a persistence model, which assumes that pollutant level today and tomorrow will be the same as yesterday, is used as a threshold for useful model prediction. Since the availability of accurate and reliable forecasting system could be useful to local authorities and could help to advise the public the proper preventive actions, we want to answer the question whether WRF-Chem model outperforms the statistical model or persistence. Namely, considering many uncertainties related to one unified model, it may not be easy for models with online chemistry to be able to perform well enough to meet the required standards, and more research and studies are needed to investigate that (Grell and Baklanov, 2011). Due to the limited number of previous studies focused on online coupled forecasting systems, the aim of our study is also to provide a greater insight into potential that lies in the approach based on an unified model for forecasting weather and air pollution. Finally, identified strengths, limitations and deficiencies of analyzed RT-AQFs, are expected to present the basis for further research.

2 Methodology

2.1 WRF-Chem forecast system

The RT-AQF system for Slovenia based on the WRF-Chem model version 3.4.1 is configured with two nested domains (Fig. 1) with horizontal resolution 11.1 and 3.7 km, and

GMDD
8, 1029–1075, 2015

151 × 100 and 181 × 145 grid points, respectively. The vertical structure of the atmosphere is resolved with 42 vertical levels, with highest near ground resolution of ~ 25 m. About 15 levels are located within the lowest 2 km to assure high vertical resolution of the daytime planetary boundary layer (PBL). To produce the 48 h forecast, the model is run every day, starting at 00:00 UTC, with meteorological initial (ICs) and lateral boundary conditions (BCs) taken from the Global Forecast System (GFS), a global numerical weather prediction system operated by the US National Weather Service (NWS). For chemical BCs forecasts from global MOZART-4/GEOS-5 (Emmons et al., 2010) RT-AQF system with temporal availability of 6 h are used. The instantaneous outputs at the 24th hour of the previous day forecast are used to initialize next day's forecasting simulation. An exception is the very first day of the first 48 h forecasting cycle, when global MOZART-4/GEOS-5 fields were used also to initialize chemistry. A three day spin-up ahead of the first analyzed forecast day is then taken into account to allow pollutants to accumulate in the air masses.

In the WRF-Chem model, several choices for parameterizations of physical and chemical processes are available (Grell et al., 2005; Skamarock et al., 2008; Peckham et al., 2012), and their choice can have a strong impact on the model predictions. We decided to apply the same schemes as were used in simulation SI1 for Phase-2 of the Air Quality Model Evaluation International Initiative (AQMEII) (e.g., Im et al., 2014a,b; Forkel et al., 2014; Kong et al., 2014; Balzarini et al., 2014; Baró et al., 2015). These include Yonsei University (YSU) PBL scheme (Hong et al., 2006), NOAH land-surface model (Chen and Dudhia, 2001), Rapid Radiative Transfer Method for Global (RRTMG) long-wave and short-wave radiation scheme (Iacono et al., 2008), Grell 3-D ensemble cumulus parameterization scheme (Grell and Devenyi, 2002) with radiative feedback, Morrison double-moment cloud microphysics (Morrison et al., 2008), RADM2 gas phase chemistry (Stockwell et al., 1990) and the MADE/SORGAM aerosol module (Ackermann et al., 1998; Schell et al., 2001). Current model implementation includes a modified RADM2 gas phase chemistry solver as described in Forkel et al. (2014), which avoids under-representation of nocturnal O_3 titration in areas with high NO emis-

GMDD
8, 1029–1075, 2015

sions. According to Forkel et al. (2014) the modified solver tends to over-estimate the low NO_2 concentration for pristine regions and in the free troposphere, which results in an overestimation of O_3. Due to the focus on polluted regions this deficiency was considered as less important than the advantage of better description of the titration. In addition, the comparatively small modelling domain (D1) ensures that the boundary conditions constrain the high bias of the modified solver for O_3 and NO_2 in the free troposphere. Also according to our sensitivity tests (results not shown) the modified solver showed better performance for O_3 daily maxima and O_3 nighttime minima than the QSSA RADM2 solver supplied originally with WRF-Chem model.

Among feedbacks only the aerosol direct effects on radiation according to Fast et al. (2006) and Chapman et al. (2009) are taken into account. As shown by Kong et al. (2014) for two air pollution episodes, this degree of aerosol-meteorology interactions in 3.4.1 version of the WRF-Chem improved model performance for high aerosol loads, while the representation of the indirect effects needs to be further improved to be able to outperform simulations with direct effects only.

Biogenic emissions are estimated using MEGAN (Model of Emissions of Gases and Aerosols from Nature; Guenther et al., 2006) online model calculations, while dust emissions are modelled according to Shaw et al. (2008) with an adjustment to avoid high dust fluxes from some Dalmatian islands in Croatia. A detailed anthropogenic inventory for pollutants CO, NH_3, NO_x, SO_2, and NMVOC, which has been for the purpose of AQ forecasting constructed for year 2009 by SEA, is used to estimate anthropogenic emissions in Slovenia. For areas outside Slovenia the recently updated anthropogenic emissions for the year 2009 based on the TNO-MACC-II (Netherlands Organization for Applied Scientific Research, Monitoring Atmospheric Composition and Climate – Interim Implementation), the same as prepared for phase-2 of the AQMEII exercise (Pouliot et al., 2014), are being used. Daily updates of the WRF-Chem based experimental AQ forecast are provided at http://meteo.fmf.uni-lj.si/onesnazenje.

GMDD
8, 1029–1075, 2015

2.2 Statistical ozone daily maximum forecast

The statistical O_3 model, currently used at SEA for forecasting O_3 daily maxima at 8 measuring sites in Slovenia (Fig. 3), is a multivariate regression tool combined with clustering algorithms to take into account measured data, weather forecast data, as well as the predicted backward trajectories of each monitoring site. As regards measurements, yesterday (at 12:00, 15:00, 18:00 and 21:00 LT, daily maximum, daily minim, daily average) and today early morning (07:00 LT) meteorological (pressure, relative humidity, direct and diffusive solar radiation, wind speed) and AQ data (O_3, NO_x, NO_2, CO, PM_{10}, SO_2) are used. For meteorological predictions the 24 h ECMWF forecast variables at 12:00 UTC of the forecast day at different vertical levels (1000, 925, 850, 500, 300 hPa) above the measuring sites are taken into account. Among all these variables by the use of stepwise technique, based on the F statistic only significant variables were selected to be included in multivariate regression equations for different monitoring sites (from 15 to 26 variables, depending on monitoring site).

The important part of the statistical forecast is calculation of 24 h backward trajectories on meteorological fields of ALADIN/SI forecast. The inclusion of 24 h predicted trajectories into statistical model is based on the study (Žabkar et al., 2008) which showed, that the highest O_3 daily maxima at monitoring sites in Slovenia are in general associated with short (slow-moving) backward trajectories with the southwestern origin, while the lowest measured daily maximum O_3 values for all the stations are associated with the clusters of long northwestern trajectories. Clusters of similar trajectories were for the purpose of statistical forecast calculated by *k* means clustering algorithms (Moody and Galloway, 1988; Žabkar et al., 2008) on 7 years (2004–2010) of data (ALADIN/SI trajectories). As an example, Fig. 2 is showing mean O_3 daily maxima for clusters of similar trajectories for one of the monitoring sites. The same 7 year time period of training data was used in the stepwise multiple regression procedure to determine the multiple regression prognostic equations associated with monitoring sites

GMDD
8, 1029–1075, 2015

and trajectory clusters, from measurements, ECMWF forecast data, average cluster O_3 daily maximum, and day-of-the-year variable.

The first step of the statistical O_3 prediction is the calculation of trajectories approaching the monitoring stations at 12:00 UTC of the forecast day. In the next step these backward trajectories of each monitoring site are associated to the nearest precalculated cluster of similar trajectories. Finally, the multiple regression equation of the associated group of trajectories is used to calculate the O_3 daily maximum prediction. It must also be noted, that the decision on declaring O_3 episodes is only partially based on the results from this statistical model; it also involves a decision made by AQ forecasters.

2.3 Evaluation methodology

We evaluate the 1 day and 2 day WRF-Chem meteorological and AQ forecasts during a 3 month period (June–August 2013).The main focus is on O_3 predictions. In the case of air pollutants, the instantaneous lowest model level mixing ratios are compared to the hourly averaged concentrations measured at monitoring stations from the national network and some other environmental information systems in Slovenia. Figure 3 shows locations of these AQ monitoring stations, and Table 1 lists the basic characteristics and pollutants with higher than 75 % availability of valid data during the analyzed time period for each of the AQ monitoring site. All AQ stations are background, 7 of them are measuring urban background, 1 suburban and 9 rural conditions. Valid O_3 measurements are for the analyzed time period available for 13 AQ stations. When studying the general model performance, data from additional 4 stations for two other pollutants (NO_2, PM_{10}) are also analyzed to get a better picture of model behavior over the domain, known for its large topographical and climate diversity. The coverage of three climate zones in Slovenia (Mediterranean, sub-alpine and mountainous) with monitoring stations is the following: NG, KOP and OTL are Mediterranean sites, KRV is mountainous station, and the remaining stations are sub-alpine. Besides elevated KRV, also ISK, OTL and VNA stations are measuring regional transport of pollutants.

GMDD
8, 1029–1075, 2015

For evaluation of predicted meteorological variables, data from SEA meteorological stations (MET, Fig. 3) for 2 m temperature (T2 m), 10 m wind speed (W10 m), relative humidity (RH), incoming shortwave radiation (SR) and precipitation (RR) are used. It must be noted, that MET stations with lower spatial representativeness (e.g. alpine stations) were not a priori excluded from the analyses, which needs to be taken into account when looking at evaluation results. The reason for not excluding these stations was that some interesting information for AQ forecast can be gained also by evaluation of meteorological forecast for these stations.

Basic statistical measures (correlation coefficient (CORR), mean error (ME), mean absolute error (MAE) and root mean square error (RMSE)) are used for evaluating model's forecasting skills of meteorological and AQ variables. In the case of O_3, correlation coefficients are presented also by Taylor diagrams (Taylor, 2001), which graphically summarize the similarity between model forecasts and observations not only in terms of their correlation, but also with their centered root-mean-square difference and the amplitude of their variations, represented by their SDs. Furthermore, some additional discrete statistical measures, including index if agreement (IOA), the mean normalized bias error (MNBE), and the mean normalized gross error (MNGE) are calculated for O_3 daily maximum concentrations predicted by the different models. Finally, to evaluate the model's ability to predict exceedances and non-exceedances also several categorical indices including accuracy (A), bias (B), false alarm ratio (FAR), probability of detection (POD) and critical success index (CSI) are calculated for different thresholds. Definitions of statistical measures are shown in Appendix A.

2.4 Meteorology and air quality of June–August 2013

The analyzed period was marked by three heat wave events, which contributed to the summer characterized by high temperatures, sunny weather and lack of precipitation in Slovenia. The first heat wave event with measured temperature daily maxima up to 35 °C occurred after a rather cold beginning of the month and lasted from 15–21 June. The event was terminated by a cold front passage and followed by the pronounced cold

GMDD
8, 1029–1075, 2015

episode during the end of June and the beginning of July. Another heat wave event with temperatures above 35 °C observed in the lowland, started on 26 July and was briefly interrupted on 29 July, when thunderstorms related to frontal passage were accompanied by exceptionally strong wind gusts. The most remarkable of three extraordinary hot episodes was recorded from 1–8 August. On the last day of this episode, 8 August, temperatures reached 40 °C at some measuring sites in Slovenia, and many of them observed their highest temperature ever recorded.

As expected for summertime conditions, measured concentrations of most air pollutants, including PM_{10}, were in general low during the analyzed time period. The only exception was O_3 with exceedances of 8 h target value (120 µg m^{-3}) measured at all AQ monitoring stations during the three heat wave events, which is the reason why the main focus of the present study is on this pollutant. During the second two events (in July and August) also threshold exceedances of 1 h daily maxima were recorded for O_3. In spite of the hot and sunny conditions during the first heat wave event in June 2013, measured daily O_3 maxima at the Slovenian stations did not exceed the 1 h information threshold value (1 h ITV; 180 µg m^{-3}), but reached 171 µg m^{-3} at the Mediterranean OTL and the elevated alpine KRV stations. During the second heat wave event 1 h daily maxima exceeded 180 µg m^{-3} at KRV, OTL, NG and KP (23–28 July), while the highest number of 1 h exceedances (20) has been in July measured at OTL station. Similarly, during the August heat wave event O_3 concentrations exceeded the 1 h ITV at LJ, MB, OTL, NG and KP from 2–7 August. To summarize, the Mediterranean stations (NG, OTL, KP) due to very high O_3 concentrations measured during the heat wave events (especially the second two events) exhibited the poorest AQ in Slovenia during the analyzed time period, while the legislation limit values have been exceeded only occasionally for the sub-alpine stations.

GMDD
8, 1029–1075, 2015

3 Results and discussion

3.1 Evaluation of meteorological variables

Table 2 shows conventional statistical scores evaluating the 1 day WRF-Chem forecast for the basic meteorological variables, 2 m temperature (T2 m; for hourly values and daily maxima), 10 m wind speed (W10 m), relative humidity (RH) and incoming solar radiation (SR). Results for three selected measuring sites (LJ, NG, MS) and overall result for all 24 MET monitoring sites (shown in Fig. 3) are presented separately.

In the case of T2 m 1 day (2 day) WRF-Chem meteorological forecast showed an overall correlation with measurements of 0.93 (0.94) for all 1 h values and 0.97 (0.96) for 1 h daily maxima. With an exception of three alpine stations with higher simulated bias, daily T2 m maxima were simulated with ME between −3.9 and −0.6 °C, depending on station spatial representativeness, while an average systematic underestimation of T2 m daily maxima was −2.1 °C both for 1 day and 2 day forecast. Nighttime T2 m minima showed lower systematic bias for 2 day forecast, which resulted in overall bias for all hourly T2 m values of −1.3 °C for 1 day and −0.8 °C for 2 day forecast. Predominant weak wind conditions with variable direction at stations located in complex topography were challenging to simulate. The general model tendency was to overestimate W10 m with overall ME of 0.8 m s^{-1} for 1 day and 2 day forecast, where for some stations bias can be very low (e.g. LJ; Table 2) and much higher for some other stations due to their local positioning in complex topography (e.g. HRA located in valley with ME of 1.9 m s^{-1}). For hourly values the correlation is lower (Table 2), but for mean daily W10 m values Pearson correlation coefficient between 0.4 and 0.9 has been simulated, depending on monitoring site. Relative humidity shows slightly better results for 1 day than for 2 day forecast with CORR of 0.77 and low overall ME of 2 % for 1 day forecast, which for particular stations can be positive (e.g. KRV) or negative (e.g. LJ, NG; Table 2).

Incoming solar radiation is the main energy source that drives all atmospheric processes, including PBL processes, and has a critical role also in atmospheric chemistry.

GMDD
8, 1029–1075, 2015

For almost all sites the mean SR was overestimated by the model, with an overall ME of 16 and 11 $W\,m^{-2}$ for 1 day and 2 day forecast, respectively. CORR was higher for 1 day (0.77) than for 2 day (0.71) forecast, with span from 0.64 to 0.90 for 1 day forecast at different stations.

Precipitation (RR) has an important role in cleansing of the atmosphere by wet deposition and scavenging. In average, the predicted precipitation underestimated the measured 3 month accumulations for −55 mm (1 day) or −8 mm (2 day forecast), where the station averaged predicted 3 month precipitation was 145 mm for 1 day, and 194 mm for 2 day forecast (results not shown). Although the WRF-Chem simulations sometimes failed to correctly predict the true amount and location of the more randomly spread summertime convective precipitation, the main precipitation events (e.g. those terminating three heat wave events) were well predicted and simulated.

3.2 Evaluation of air quality variables

In this section we evaluate WRF-Chem predictions for O_3, NO_2 and PM_{10}, as three of the most problematic pollutants in terms of harm to human health and compliance with EU limit values (EEA, 2012). Table 3 shows the domain wide performance statistics for 1 day and 2 day forecasts of these pollutants, where in the case of O_3 1 h and 8 h averages and daily maxima are analyzed separately. The comparison of 1 day and 2 day forecasts shows that concentrations of air pollutants were somewhat better forecasted 1 day than 2 days ahead by means of almost all of statistics shown in Table 3, with higher impact on O_3 predictions. Although the 2 day prediction was generally not worse for the majority of meteorological variables, the reason for better 1 day prediction in the case of O_3 could be somewhat stronger simulated winds on the second day of simulation. Stronger winds impact the transport and dispersion of pollutants, and have the greatest consequence for secondary pollutants (like O_3) which need time to be formed.

As shown in Table 3 the WRF-Chem simulations tend to overestimate the 1 h and 8 h O_3 values with ME of 13.7 and 13.8 $\mu g\,m^{-3}$, respectively. Looking at MAE, RMSE and CORR statistics, agreement with measurements is better for 8 h (22.9, 28.5 $\mu g\,m^{-3}$

GMDD
8, 1029–1075, 2015

and 0.69) than for 1 h O_3 values (25.5, 32.5 $\mu g\,m^{-3}$ and 0.65), which is in line with results of previous studies (e.g. Tong and Mauzerall, 2013) and suggests that the current modeling system has problem to simulate the small-scale fluctuations of O_3. On the other hand evaluations of predicted 8 h and daily O_3 maxima, which are of most concern, show a nice model performance (ME, MAE RMSE and CORR of −3.4, 13.7, 17.1 $\mu g\,m^{-3}$ and 0.81 for daily maxima, respectively), in line or even better than obtained in some previous studies (e.g. Tong and Mauzerall, 2006; Chuang et al., 2011; Yahya et al., 2014), which could be to some extent related to higher model resolution.

To understand results of the domain wide statistics (in Table 3) we further analyze spatial and temporal characteristics of model O_3 predictions. Figure 4 shows a spatial pattern of average simulated 1 day predictions for O_3, NO_2 and PM_{10} overlaid with measured averages, where in the case of O_3 results for all hourly values and for daily maxima are shown separately. Examples of forecasted and measured time series for O_3 at different stations are shown in Fig. 5. The elevated alpine KRV station is the only one with negative bias (−12 $\mu g\,m^{-3}$) in forecasted 1 h O_3 concentrations, which can be explained by the too low altitude of the KRV station in model topography, since the mean O_3 concentration increases with height. Beside KRV also Mediterranean KOP and OTL stations, as well as rural ZAV site, are stations with comparatively high measured nighttime O_3 levels, which results in low overall bias for all hourly O_3 values for these stations (from −3 to −7 $\mu g\,m^{-3}$). Namely, WRF-Chem model cannot capture well the profound nighttime O_3 reductions (shown also by Žabkar et al., 2013; Im et al., 2014a), which contributes to the overall over-prediction of hourly O_3 concentrations (from 10 to 36 $\mu g\,m^{-3}$) for stations with very low measured nighttime O_3 concentrations.

Looking at O_3 daily maxima (Fig. 4b), the under-predictions occur at alpine KRV (−16 $\mu g\,m^{-3}$) and at three Mediterranean stations (OTL, NG, KOP; from −14 to −11 $\mu g\,m^{-3}$). For Mediterranean stations the underestimations of daily maxima are most probably due to inaccurate representation of costal processes in model, which are crucial for PBL height evolution and accumulation of pollution in the near ground air layers. For TRB station located in narrow valley of the very complex terrain that can-

GMDD
8, 1029–1075, 2015

not be appropriately resolved in the current model topography, model over-predicts O_3 daily maxima for 14 µg m^{-3}. For other sub-alpine stations the bias of O_3 daily maxima predictions is lower.

To some extent the previously mentioned model over-predictions of nighttime O_3 minima could be explained by model incongruity in predicted NO_2 levels. When evaluating the primary pollutants one must be aware that in the model the instantaneous emissions are spread over an entire grid box, which results in underestimated emissions and concentrations close to the source regions and overestimated emissions and concentrations at rural locations adjacent to the source regions, and can thus cause a combined effect of negative and positive biases at urban and rural sites. Comparisons of WRF-Chem predicted NO_2 levels with measurements show that in spite of the high spatial resolution the concentrations of the small urban areas are insufficiently represented by the model (Fig. 4c). In Slovenia many towns are located in basins or very narrow valleys, usually poorly or even not resolved in model topography. Smoothed local emissions for these towns show significant underestimations of NO_2 concentrations (e.g. ZAG in Fig. 6). In combination with deficiently reproduced meteorological processes (calm and stable nighttime conditions in valleys and basins) this results in an underestimation of the O_3 loss by titration. This can explain the positive nighttime bias of O_3 found at these sites. The situation is better for bigger cities, located in wider basins, like LJ or CE (LJ; Fig. 6), while at rural sites NO_2 is either well simulated (e.g. MOH; Fig. 6), or slightly over-predicted due to increased emissions from adjacent urban area (e.g. ZAD; Fig. 6). The overall agreement of hourly NO_2 predictions with measurements was good for rural sites, while urban sites experienced under-predictions, which were highest for small cities, especially for NG (ME of −13 µg m^{-3}) and ZAG (ME of −14 µg m^{-3}).

Interesting to discuss are also results for predicted PM_{10} concentrations (Table 3 and Fig. 4d), showing slight over-prediction of daily PM_{10} levels at all stations which is somewhat surprising due to the fact that nearly all current off-line and on-line coupled chemical transport models show large systematic PM_{10} underestimations. For exam-

GMDD
8, 1029–1075, 2015

ple, within AQMEII exercise, where seventeen modeling groups from Europe and North America were brought together, running eight operational online-coupled air quality models over Europe and North America, the rural PM_{10} concentrations over Europe were underestimated by all models (model configurations) by up to 66 % while for the urban PM_{10} concentrations the underestimations were even much larger (up to 75 %) (Im et al., 2014b). The reason for slight over-prediction of PM_{10} levels could be to some extent attributed to the high model spatial resolution used in our study. Further, CORR for daily PM_{10} concentrations is rather low (0.34 and 0.37 for 1 day and 2 day forecasts, respectively; Table 3), which is partly due to the low temporal dynamics of measured daily PM_{10} concentrations during the analyzed time period (no recorded PM_{10} exceedances), and partly due to the simulated PM_{10} overestimations during the heat wave events. These over-predictions contributed also to the overall positive bias of predicted PM_{10} levels. As shown in Fig. 7 for two monitoring sites, there was a significant PM_{10} over-prediction simulated on 10 June (day 8 in Fig. 7), related to the pre-frontal advection of polluted air-masses coming from the north-western part of the domain D2 (coming from domain D1). The next significant PM_{10} over-prediction occurred during the first heat wave episode (17–22 June), when during the hot and low wind conditions (after 17 June) the PM_{10} levels started to build up in the PBL over entire domain D2 (and over southwestern parts of domain D1), and reached the maximum concentrations in Slovenia again with prefrontal advection of polluted air masses. Both over-predictions contributed to an overall positive bias in forecasted PM_{10} concentrations. Detailed analyses showed that high concentrations in domain D1 originated from boundary conditions, and appear to be a consequence of overestimated advection of Saharan dust in MOZART model predictions. The increase in PM_{10} concentrations over Slovenia was also simulated during the prefrontal advection related to the cold front which terminated the next two heat wave events in July and August (days 56–57 and days 67–68 in Fig. 7), but during these days predicted PM_{10} levels were close to the measured PM_{10} concentrations.

GMDD
8, 1029–1075, 2015

3.3 Evaluation and comparison of different methods for O_3 daily maximum predictions

In this section we want to answer the question how accurate is the 1 h O_3 daily maximum WRF-Chem forecast in comparison with the statistical model prediction or with persistence. According to Zhang et al. (2012a) statistical models are known to be generally more suitable for complex site-specific relations between concentrations of air pollutants and predictors. With appropriate and accurate predictors they have a higher accuracy as compared to deterministic models, which is beside the computational efficiency their main advantage (Zhang et al., 2012a). Among strengths of the deterministic models is that they give prognostic time- and spatially-resolved concentrations under typical and atypical scenarios, and can give scientific insights into pollutant formation processes (Zhang et al., 2012a). Furthermore, they allow forecast also for locations which are not monitored due to their complete spatial coverage. In spite of simplified description of physical and chemical processes in the deterministic models and inaccuracies and uncertainties in model inputs (in particular the emissions), some previous studies already suggested that deterministic models can also have skills close to statistical forecasting tools (e.g. Manders et al., 2009). In addition to evaluation and comparison of O_3 daily maxima predictions with WRF-Chem and the statistical model, we decided to add a persistence model as a threshold for useful model prediction. Persistence works well under stationary conditions, but because it cannot handle changes in weather and emissions, fails at the beginning and at the end of the episodes (Zhang et al., 2010a). Regarding the extremes, models of all types are known to have problem to accurately predict them, while persistence predicts extremes with a 1 day (2 day) time lag.

Figure 8 compares discrete statistics site by site for 1 day and 2 day model predictions of 1 h O_3 daily maxima. Similarly, Table 4 shows these statistics for all data with different thresholds applied (only for WRF-Chem and persistence, because statistical forecast is not available for all stations), and separately for different types of stations

GMDD
8, 1029–1075, 2015

(sub-alpine urban, rural, Mediterranean urban) with available statistical forecast. Looking at ME persistence gives results close to zero as long as no threshold is applied, while with threshold of 140 $\mu g\,m^{-3}$ (Table 4) ME of 1 day persistence ($-10.2\,\mu g\,m^{-3}$) is very close to the WRF-Chem model for 1 day predictions ($-11.9\,\mu g\,m^{-3}$), and for 2 day predictions WRF-Chem ($-14.6\,\mu g\,m^{-3}$) already beats the persistence ($-19.4\,\mu g\,m^{-3}$). Site-by-site comparison (Fig. 8) shows that for most stations the statistical forecast has a lower ME than WRF-Chem forecast, but there are also stations (ISK, HRA, LJ) with lower or equal ME for WRF-Chem than for statistical model, indicating the possible occurrence of atypical conditions not resolved by the statistical model. Looking at MAE and RMSE, at all stations except those with highest ME (KRV, TRB, KOP) WRF-Chem outperforms the persistence already in the 1 day forecast. Among sites with available statistical forecast there are only two (KRV, KOP) with WRF-Chem performing worse than statistical forecast. CORR is one of the parameters that suggest how much the model is able to follow the true nature of processes regardless the possible bias. For almost all stations WRF-Chem shows higher CORR than persistence for 1 day and 2 day forecasts. Only at the KRV station the 1 day statistical forecast (CORR = 0.80) somewhat outperforms WRF-Chem (0.76), and at NG and KOP CORR for WRF-Chem and statistical model is very similar.

The Taylor diagrams in Fig. 9 show CORR together with the centered root-mean-square difference (RMSD) between model forecasts and observations, and the amplitude of their variations (SD), where ideal forecast would lie in the right-bottom corner. WRF-Chem gives higher CORR and lower RMSD for all types of stations, while SD of WRF-Chem O_3 daily maxima predictions is underestimated and lower than for other model forecasts. The latter shows that the variability in WRF-Chem model predictions is not as large as that in observed values. MNBE in Fig. 8 has a course very similar to ME. For all forecasts except WRF-Chem for the TRB site (with MNBE of 16 %) which is located in a narrow valley that is not resolved in the current model resolution, MNBE is below the ±10–15 %, which is the US EPA (US EPA, 1991) recommended threshold for the models used for regulatory applications. For MNGE the US EPA recommendation

GMDD
8, 1029–1075, 2015

below 30–35 % for O_3 applications is met by all forecasts, even in the case of 2 day persistence model. With exception of the MS and KOP sites MNGE is lower for WRF-Chem than for statistical forecast, while for KOP and KRV sites with highest negative bias in WRF-Chem predictions, 1 day persistence gives best results, followed by the statistical forecast and WRF-Chem. Very similar are results for IOA with the range of 0–1, and score 1 indicating perfect model agreement with the observations. We can conclude that for most stations the WRF-Chem predictions are in line or even outperform the statistical model. With the exception of the stations with high bias due to very complex local topography (TRB), unresolved coastal processes (KOP) or alpine stations (KRV), the WRF-Chem forecasts are more accurate than persistence.

The key requirement for a forecast system is to be able to predict O_3 concentration levels greater than a given threshold. Thus, in addition to the discrete evaluation just presented, also the contingency-table-based statistics are an important metrics of forecast performance. Table 5 summarizes the categorical evaluation results for three different thresholds (120, 140, 160 $\mu g\,m^{-3}$). Namely, it should be taken into account that results of categorical statistics are very sensitive to the threshold chosen, as well as to the overall pollution levels during the analyzed months. Accuracy (A), which measures how often the forecasts are correct either above or below the threshold, increases with threshold level. Looking at 1 day forecast A is highest for statistical forecast at 120 $\mu g\,m^{-3}$ threshold, for WRF-Chem forecast at 140 $\mu g\,m^{-3}$ threshold, and in the case of 160 $\mu g\,m^{-3}$ threshold applied, for persistence. There is a tendency of the statistical model and of WRF-Chem to under-predict O_3 daily maxima. This shows as a bias (B) below 1 for these two models, where B determines whether the same fraction of events are both forecasted and observed. The false alarm ratio (FAR) that measures the percentage of forecast high O_3 events that turn out to be false alarms, gives highest skill for WRF-Chem, followed by statistical model and persistence. The probability of detection (POD) is a measure of how often a high threshold occurrence is actually predicted to occur, and is relatively low for WRF-Chem with respect to other models. Another useful measure, the critical success index (CSI), measures the percentage of

GMDD
8, 1029–1075, 2015

cases that are correctly forecasted out of those either forecasted or observed, and is for higher two thresholds best for persistence in the case of 1 day forecast, while on the second day WRF-Chem or the statistical model already perform better.

It must be noted, that in categorical evaluations systematic biases like those obtained with WRF-Chem for some stations (KOP, KRV), significantly impact the model performance. For example, if KOP and KRV stations were excluded from categorical evaluations, WRF-Chem performance improved by means of all statistical measures (results not shown). If correction techniques, based on observations and the previous day's forecast (e.g., McKeen et al., 2005, 2007; Kang et al., 2008) would be applied to correct the systematic biases, WRF-Chem forecasts might outperform other two models even in categorical evaluations.

4 Summary and conclusion

A high resolution modelling system based on an on-line coupled WRF-Chem has been applied for numerical weather prediction and for forecasting air quality in Slovenia. In the study the evaluation of the forecasting system has been conducted for three summer months. Both 1 day and 2 day predictions of meteorological and air quality variables have been analyzed. The focus has been on O_3 as the only pollutant with recorded exceedances of legislation limit values during the three heat wave events in June, July and August 2013. WRF-Chem daily O_3 maximum predictions have also been compared to the operational statistical model and persistence forecasts to answer the question how skillful are the WRF-Chem model predictions compared to these two models.

1 day and 2 day WRF-Chem PM_{10} forecasts showed a very low bias. Exceptions were two events with significantly over-predicted PM_{10} levels due to prefrontal advection of polluted air masses from neighboring regions. Knowing that majority of the current chemical transport models show large negative biases in simulated PM_{10} con-

GMDD
8, 1029–1075, 2015

centrations, these results present a good starting point for studying the importance of aerosol feedbacks with realistic model aerosol concentrations, left for future research.

The overall agreement of WRF-Chem NO_2 forecast with measurements was good for rural sites, while urban sites experienced model under-predictions, which were highest for small towns. One important reason is that many small towns are located in basins or very narrow valleys, usually poorly presented in model topography. Smoothed local emissions show as model underestimations of NO_2 concentrations for these towns. This in combination with insufficiently reproduced calm meteorological conditions in basins and valleys during the nighttime hours explains also WRF-Chem over-predictions of nighttime O_3 concentrations.

Evaluations of predicted 1 and 8 h daily O_3 maxima, which are in the case of this pollutant of the highest interest, show a nice WRF-Chem model performance. Nevertheless, there are also stations which experience high over- or under-predictions of O_3 daily maximum levels. For Mediterranean sites the under-predictions of the daily maxima are most probably due to inaccurate representation of costal processes in model, which are crucial for the PBL height evolution and accumulation of pollution in the near ground air layers. For some sub-alpine stations the reason for the higher bias in O_3 daily maximum predictions is their location either at elevated mountainous or coastal regions, or in narrow valleys which cannot be appropriately resolved in the current model resolution – that impacts how accurately model simulates the local processes responsible for the level of local pollution. Comparisons of WRF-Chem O_3 daily maximum forecasts with persistence and with statistical model predictions show that with respect to some statistical parameters the deterministic WRF-Chem forecast can outperform the other two for both 1 day and 2 day predictions. For example, correlation coefficient shows highest skill for WRF-Chem model, confirming the importance of complex processes as taken into account in an on-line coupled Eulerian model. Further improvement of WRF-Chem forecasting skill could be obtained by applying one of the bias-correction methods in order to account for unresolved topographical and coastal effects, as well as emission patterns. Chemical data assimilation, although currently still

GMDD
8, 1029–1075, 2015

in its infancy for online coupled meteorology-chemistry models (Bocquet et al., 2014), could in future also be used as an efficient method for improving prediction of chemical concentration fields.

Appendix A: Statistical measures

For ith observed (O_i) and the corresponding modelled (M_i) value of variable, discrete statistical measures are calculated as follows:

Mean error:

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)$$

Mean absolute error:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|M_i - O_i|$$

Root mean square error:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}$$

Correlation coefficient:

$$r = \frac{\sum_{i=1}^{N}\left(M_i - \overline{M}\right)\left(O_i - \overline{O}\right)}{\sqrt{\sum_{i=1}^{N}\left(M_i - \overline{M}\right)^2\left(O_i - \overline{O}\right)^2}}$$

GMDD
8, 1029–1075, 2015

Index of agreement:

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}\left(\left|M_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2}$$

Mean normalized bias error:

$$\mathrm{MNBE} = \frac{1}{N}\sum_{i=1}^{N}\frac{M_i - O_i}{O_i} \times 100$$

Mean normalized gross error:

$$\mathrm{MNGE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|M_i - O_i|}{O_i} \times 100$$

For categorical evaluation all model predictions are first classified into four groups (a, b, c and d):

a prediction is above, but observation is below the threshold

b prediction and observation are above the threshold

c prediction and observation are below the threshold

d prediction is below, but observation is above the threshold

Categorical statistics are calculated as follows:

Accuracy:
$A = \frac{b+c}{a+b+c+d}$

Bias:
$B = \frac{a+b}{b+d}$

GMDD
8, 1029–1075, 2015

False alarm ratio:
FAR = $\frac{a}{a+b}$

Probability of detection:
POD = $\frac{b}{b+d}$

Critical success index:
CSI = $\frac{b}{a+b+d}$

Acknowledgements. Centre of Excellence for Space Sciences and Technologies SPACE-SI (OP13.1.1.2.02.0004) is part financed by the European Union, European Regional Development Fund and Republic of Slovenia, Ministry of Education, Science and Sport. The authors thankfully acknowledge TNO for providing the TNO-MACC-II anthropogenic emissions. Statistical model predictions and measurement data used in the study were kindly provided by Slovenian Environmental Agency and Electroinstitute Milan Vidmar. The support through COST Action ES1004 EuMetChem is gratefully acknowledged.

References

Ackermann, I. J., Hass, H., Memmesheimer, M., Ziegenbein, C., and Ebel, A.: The parameterization of the sulfate-nitrate-ammonia aerosol system in the long-range transport model EURA D, Meteorol. Atmos. Phys., 57, 101–114, 1995.

ALADIN International Team: The ALADIN project: mesoscale modelling seen as a basic tool for weather forecasting and Atmospheric Research, WMO Bull., 46, 317–324, 1997.

Baklanov, A., Korsholm, U., Mahura, A., Petersen, C., and Gross, A.: Enviro-HIRLAM: online coupled modelling of urban meteorology and air pollution. Adv. Sci. Res., 2, 41–46, doi:10.5194/asr-2-41-2008, 2008.

Baklanov, A.: Chemical weather forecasting: a new concept of integrated modeling, Adv. Sci. Res., 4, 23–27, 2008, 2010.

Baklanov, A., Schlünzen, K., Suppan, P., Baldasano, J., Brunner, D., Aksoyoglu, S., Carmichael, G., Douros, J., Flemming, J., Forkel, R., Galmarini, S., Gauss, M., Grell, G., Hirtl, M., Joffre, S., Jorba, O., Kaas, E., Kaasik, M., Kallos, G., Kong, X., Korsholm, U., Kurganskiy, A., Kushta, J., Lohmann, U., Mahura, A., Manders-Groot, A., Maurizi, A., Mous-

GMDD
8, 1029–1075, 2015

siopoulos, N., Rao, S. T., Savage, N., Seigneur, C., Sokhi, R. S., Solazzo, E., Solomos, S., Sørensen, B., Tsegas, G., Vignati, E., Vogel, B., and Zhang, Y.: Online coupled regional meteorology chemistry models in Europe: current status and prospects, Atmos. Chem. Phys., 14, 317–398, 2014, http://www.atmos-chem-phys.net/14/317/2014/.

Balzarini, A., Pirovano, G., Honzak, L., Žabkar, R., Curci, G., Forkel, R., Hirtl, M., San José, R., Tuccella, P., and Grell, G.: WRF-Chem model sensitivity to chemical mechanisms choice in reconstructing aerosol optical properties, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2014.12.033, 2014.

Baró, R., Jiménez-Guerrero, P., Balzarini, A., Curci, G., Forkel, R., Hirtl, M., Honzak, L., Im, U., Lorenz, C., Pérez, J. L., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., and Žabkar, R.: Sensitivity analysis of the microphysics scheme in WRF-Chem contributions to AQMEII phase 2, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2015.01.047, 2015.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmos. Chem. Phys. Discuss., 14, 32233–32323, doi:10.5194/acpd-14-32233-2014, 2014.

Byun, D. W. and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, Appl. Mech. Rev., 59, 51–77, 2006.

Chapman, E. G., Gustafson Jr., W. I., Easter, R. C., Barnard, J. C., Ghan, S. J., Pekour, M. S., and Fast, J. D.: Coupling aerosol-cloud-radiative processes in the WRF-Chem model: investigating the radiative impact of elevated point sources, Atmos. Chem. Phys., 9, 945–964, 2009, http://www.atmos-chem-phys.net/9/945/2009/.

Chen, F. and Dudhia, J.: Coupling an advanced land surface–hydrology model with the Penn State NCAR MM5 Modeling System, Part I: model implementation and sensitivity, Mon. Weather Rev., 129, 569–585, 2001.

Chuang, M. T., Zhang, Y., and Kang, D. W.: Application of WRF-Chem-MADRID for real-time air quality forecasting over the southeastern United States, Atmos. Environ., 45, 6241–6250, 2011.

GMDD
8, 1029–1075, 2015

Cobourn, W. G.: Accuracy and reliability of an automated air quality forecast system for ozone in seven Kentucky metropolitan area, Atmos. Environ., 41, 5863–5875, 2007.
EC/2008/50: Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, Official Journal of the European Union, L152, 44 pp., 2008.
Eder, B. K., Kang, D., Mathur, R., Yu, S., and Schere, K.: An operational evaluation of the Eta-CMAQ air quality forecast model, Atmos. Environ., 40, 4894–4905, 2006.
EEA: The application of models under the European Union's Air Quality Directive: a technical reference guide, 72 pp., 2011.
EEA: Air Quality in Europe – 2012 Report, ISBN 978-92-9213-328-3, Office for Official Publications of the European Union, Luxembourg, 108 pp., 2012.
Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3, 43–67, doi:10.5194/gmd-3-43-2010, 2010.
ENVIRON: CAMx User's Guide, Comprehensive Air Quality Model With Extensions Version 5.40, ENVIRON International Corporation, Novato, California, 2011.
Fast, J., Aiken, A. C., Allan, J., Alexander, L., Campos, T., Canagaratna, M. R., Chapman, E., DeCarlo, P. F., de Foy, B., Gaffney, J., de Gouw, J., Doran, J. C., Emmons, L., Hodzic, A., Herndon, S. C., Huey, G., Jayne, J. T., Jimenez, J. L., Kleinman, L., Kuster, W., Marley, N., Russell, L., Ochoa, C., Onasch, T. B., Pekour, M., Song, C., Ulbrich, I. M., Warneke, C., Welsh-Bon, D., Wiedinmyer, C., Worsnop, D. R., Yu, X.-Y., and Zaveri, R.: Evaluating simulated primary anthropogenic and biomass burning organic aerosols during MILAGRO: implications for assessing treatments of secondary organic aerosols, Atmos. Chem. Phys., 9, 6191–6215, doi:10.5194/acp-9-6191-2009, 2009.
Fast, J. D., Gustafson Jr., W. I., Easter, R. C., Zaveri, R. A., Barnard, J. C., Chapman, E. G., Grell, G. A., and Peckham, S. E.: Evolution of ozone, particulates, and aerosol direct radiative forcing in the vicinity of Houston using a fully coupled meteorology-chemistry-aerosol model, J. Geophys. Res., 111, D21305, doi:10.1029/2005JD006721, 2006.
Forkel, R., Werhahn, J., Buus Hansen, A., McKeen, S., Peckham, S., Grell, G., and Suppan, P.: Effect of aerosol-radiation feedback on regional air quality – a case study with WFR/Chem, Atmos. Environ., 53, 202–211, 2012.

GMDD
8, 1029–1075, 2015

Forkel, R., Balzarini, A., Baró, R., Curci, G., Jiménez-Guerrero, P., Hirtl, M., Honzak, L., Im, U., Lorenz, C., Pérez, J. L., Pirovano, G., San José, R., Tuccella, P., Werhahn, J., and Žabkar, R.: Analysis of the WRF-Chem contributions to AQMEII phase2 with respect to aerosol radiative feedbacks on meteorology and pollutant distribution, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2014.10.056, 2014.

Grell, G. and Baklanov, A.: Integrated modeling for forecasting weather and air quality: a call for fully coupled approaches, Atmos. Environ., 45, 6845–6851, 2011.

Grell, G., Dudhia, J., and Stauffer, D.: A description of the fifth-generation Penn State/NCAR Mesoscale model (MM5), TN-398+STR, NCAR, Boulder, CO, 1995.

Grell, G. A. and Devenyi, D.: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, Geophys. Res. Lett. 29, 38.1–38.4, doi:10.1029/2002GL015311, 2002.

Grell, G. A., Emeis, S., Stockwell, W. R., Schoenemeyer, T., Forkel, R., Michalakes, J., Knoche, R., and Seidl, W.: Application of a multiscale, coupled MM5/chemistry model to the complex terrain of the VOTALP valley campaign, Atmos. Environ., 34, 1435–1453, 2000.

Grell, G. A., Knoche, R., Peckham, S. E., and McKeen, S. A.: Online vs. offline air quality modeling on cloud-resolving scales, Geophys, Res. Lett., 31, L16117, doi:10.1029/2004GL020175, 2004.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W., and Eder, B.: Fully coupled "online" chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, 2005.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181–3210, 2006, http://www.atmos-chem-phys.net/6/3181/2006/.

Hong, S., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Mon. Weather Rev., 134, 2318–2341, 2006.

Hu, X.-M., Doughty, D., Sanchez, K. J., Joseph, E., and Fuentes, J. D.: Ozone variability in the atmospheric boundary layer in Maryland and its implications for vertical transport model, Atmos. Environ., 46, 354–364, 2012.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., and Collins, W. D.: Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models, J. Geophys. Res., 113, D13103, doi:10.1029/2008JD009944, 2008.

GMDD
8, 1029–1075, 2015

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Kuenen, J. J. P., Makar, P. A., Manders-Groot, A., Neal, L., Perez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, K., Wolke, R., Yahya, K., Žabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2, Part I: Ozone, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2014.09.042, 2014a.

Im, U., Bianconi, R., Solazzo, E., Kioutsioukis, I., Badia, A., Balzarini, A., Baro, R., Bellasio, R., Brunner, D., Chemel, C., Curci, G., Denier van der Gon, H. A. C., Flemming, J., Forkel, R., Giordano, L., Jimenez-Guerrero, P., Hirtl, M., Hodzic, A., Honzak, L., Jorba, O., Knote, C., Makar, P. A., Manders-Groot, A., Neal, L., Perez, J. L., Pirovano, G., Pouliot, G., San Jose, R., Savage, N., Schroder, W., Sokhi, R. S., Syrakov, D., Torian, A., Tuccella, P., Werhahn, K., Wolke, R., Yahya, K., Žabkar, R., Zhang, Y., Zhang, J., Hogrefe, C., and Galmarini, S.: Evaluation of operational online-coupled regional air quality models over Europe and North America in the context of AQMEII phase 2, Part II: Particulate Matter, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2014.08.072, 2014b.

Jacobson, M. Z.: GATOR-GCMM: a global through urban scale air pollution and weather forecast model, 1, model design and treatment of subgrid soil, vegetation, roads, rooftops, water, sea ice, and snow, J. Geophys. Res., 106, 5385–5402, 2001.

Kaminski, J. W., Neary, L., Struzewska, J., McConnell, J. C., Lupu, A., Jarosz, J., Toyota, K., Gong, S. L., Côté, J., Liu, X., Chance, K., and Richter, A.: GEM-AQ, an on-line global multiscale chemical weather modelling system: model description and evaluation of gas phase chemistry processes, Atmos. Chem. Phys., 8, 3255–3281, doi:10.5194/acp-8-3255-2008, 2008.

Kang, D., Mathur, R., Rao, S. T., and Yu, S.: Bias adjustment techniques for improving ozone air quality forecasts, J. Geophys. Res., 113, D23308, doi:10.1029/2008JD010151, 2008.

Klein, T., Kukkonen, J., Dahl, A., Bossioli, E., Baklanov, A., Vik, A. F., Agnew, P., Karatzas, K. D., and Sofiev, M.: Interactions of physical, chemical, and biological weather calling for an integrated approach to assessment, forecasting, and communication of air quality, AMBIO, 41, 851–864, 2012.

GMDD
8, 1029–1075, 2015

Kong, X., Forkel, R., Sokhi, R., Suppan, P., Baklanovc, A., Gauss, M., Brunner, D., Baro Esteban, R., Balzarini, A., Chemel, C., Curci, G., Galmarini, S., Jiménez Guerrero, P., Hirtl, M., Honzak, L., Im, U., Pérez, J. L., Piravano, G., San Jose, R., Schlünzen, H., Tsegas, G., Tuccella, P., Werhahn, J., and Žabkar, R.: Investigation of meteorology and chemistry interactions and their representations in online coupled models with the supported case studies from AQMEII phase2, Atmos. Environ., online first, doi:10.1016/j.atmosenv.2014.09.020, 2014.

Li, G., Zavala, M., Lei, W., Tsimpidi, A. P., Karydis, V. A., Pandis, S. N., Canagaratna, M. R., and Molina, L. T.: Simulations of organic aerosol concentrations in Mexico City using the WRF-CHEM model during the MCMA-2006/MILAGRO campaign, Atmos. Chem. Phys., 11, 3789–3809, doi:10.5194/acp-11-3789-2011, 2011.

Manders, A. M. M., Schaap, M., and Hoogerbrugge, R.: Testing the capability of the chemistry transport model LOTOS-EUROS to forecast PM_{10} levels in the Netherlands, Atmos. Environ., 43, 4050–4059, 2009.

McCollister, G. and Wilson, K.: Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants, Atmos. Environ., 9, 417–423, 1975.

McKeen, S., Wilczak, J., Grell, G., Djalova, I., Peckham, S., Hsie, E.-Y., Gong, W., Bouchet, V., Ménard, S., Moffet, R., McHenry, J., McQueen, J., Tang, Y., Carmichael, G. R., Pagowski, M., Chan, A., Dye, T., Frost, G., Lee, P., and Mathur, R.: Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004, J. Geophys. Res., 110, D21307, doi:10.1029/2005JD005858, 2005.

McKeen, S., Chung, S. H., Wilczak, J., Grell, G., Djalalova, I., Peckham, S., Gong, W., Bouchet, V., Moffet, R., Tang, Y., Carmichael, G. R., Mathur, R., and Yu, S.: Evaluation of several $PM_{2.5}$ forecast models using data collected during the ICARTT/NEAQS 2004 field study, J. Geophys. Res., 112, D10S20, doi:10.1029/2006JD007608, 2007.

McKeen, S., Grell, G., Peckham, S., Wilczak, J., Djalalova, I., Hsie, E.-Y., Frost, G., Peischl, J., Schwarz, J., Spackman, R., Holloway, J., de Gouw, J., Warneke, C., Gong, W., Bouchet, V., Gaudreault, S., Racine, J., McHenry, J., McQueen, J., Lee, P., Tang, Y., Carmichael, G. R., and Mathur, R.: An evaluation of real-time air quality forecasts and their urban emissions over eastern Texas during the summer of 2006 Second Texas Air Quality Study field study, J. Geophys. Res., 114, D00F11, doi:10.1029/2008JD011697, 2009.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Tur-

GMDD
8, 1029–1075, 2015

quety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geosci. Model Dev., 6, 981–1028, doi:10.5194/gmd-6-981-2013, 2013.

Misenis, C. and Zhang, Z.: An examination of sensitivity of WRF-Chem predictions to physical parameterizations, horizontal grid spacing, and nesting options, Atmos. Res., 97, 315–334, 2010.

Moody, J. and Galloway, J.: Quantifying the relationship between atmospheric transport and the chemical composition of precipitation on Bermuda, Tellus B, 40, 436–479, 1988.

Morrison, H., Thompson, G., and Tatarskii, V.: Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one- and two-moment schemes, Mon. Weather Rev., 137, 991–1007, 2009.

Otte, T. L., Pouliot, G., Pleim, J. E., Young, J. O., Schere, K. L., Wong, D. C., Lee, P. C. S., Tsidulko, M., McQueen, J. T., Davidson, P., Mathur, R., Chuang, H.-Y., DiMego, G., and Seaman, N. L.: NCEP Notes: linking the Eta model with the community multiscale air quality (CMAQ) modeling system to build a national air quality forecasting system, Weather Forecast., 20, 367–384, 2005.

Peckham, S. E., Grell, G. A., McKeen, S. A., Barth, M., Pfister, G., Wiedinmyer, C., Fast, J. D., Gustafson, W. I., Ghan, S. J., Zaveri, R., Easter, R. C., Barnard, J., Chapman, E., Hewson, M., Schmitz, R., Salzman, M., and Freitas, S. R.: WRF-Chem Version 3.3 User's Guide, Boulder, Colo.: US Dept. of Commerce, National Oceanic and Atmospheric Administration, Oceanic and Atmospheric Research Laboratories, Global Systems Division, 2012.

Pouliot, G., Pierce, T., Denier van der Gon, H., Schaap, M., Moran, M., and Nopmongcol, U.: Comparing emissions inventories and model-ready emissions datasets between Europe and North America for the AQMEII Project, Atmos. Environ., 53, 75–92, 2012.

Saide, P. E., Carmichael, G. R., Spak, S. N., Gallardo, L., Osses, A. E., Mena-Carrasco, M. A., and Pagowski, M.: Forecasting urban PM_{10} and $PM_{2.5}$ pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model, Atmos. Environ., 45, 2769–2780, 2011.

Schell, B., Ackermann, I. J., Hass, H., Binkowski, F. S., and Ebel, A.: Modeling the formation of secondary organic aerosol within a comprehensive air quality model system, J. Geophys. Res., 106, 28275–28293, 2001.

Shaw, W. J., Allwine, K., Fritz, B. G., Rutz, F. C., Rishel, J. P., and Chapman, E. G.: An evaluation of the wind erosion module in DUSTRA N, Atmos. Environ., 42, 1907–1921, 2008.

GMDD
8, 1029–1075, 2015

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X. Y., Wang, W., and Powers, J. G.: A Description of the Advanced Research WRF Version 3, NCAR Technical Note, NCAR/TN-475þSTR, 113 pp., 2008.

Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The second generation regional acid deposition model chemical mechanism for regional air quality modeling, J. Geophys. Res., 95, 16343–16367, 1990.

Sudo, K., Takahashi, M., Kurokawa, J., and Akimoto, H.: CHASER: a global chemical model of the troposphere 1, model description, J. Geophys. Res., 107, ACH7.1–ACH7.20, doi:10.1029/2001JD001113, 2002.

Takigawa, M., Niwano, M., Akimoto, H., and Takahashi, M.: Development of a one-way nested global-regional air quality forecasting model, SOLA, 3, 081–084, 2007.

Tulet, P., Crassier, V., Solmon, F., Guedalia, D., and Rosset, R.: Description of the mesoscale nonhydrostatic chemistry model and application to a transboundary pollution episode between northern France and southern England, J. Geophys. Res., 108, 4021, doi:10.1029/2000JD000301, 2003.

US Environmental Protection Agency (US EPA): Guideline for regulatory application of the Urban Airshed Model, EPA-450/4–91-013, July 1991, United States Environmental Protection Agency, Research Triangle Park, NC, Office of Air Quality Planning and Standards, 1991.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Tie, X., Geng, F. H., Peng, L., Gao, W., and Zhao, C. S.: Measurement and modeling of O_3 variability in Shanghai, China; application of the WRF-Chem model, Atmos. Environ., 43, 4289–4302, 2009.

Tong, D. Q. and Mauzerall, D. L.: Spatial variability of summertime tropospheric ozone over the continental United States: implications of an evaluation of the CMAQ model, Atmos. Environ., 40, 3041–3056, 2006.

van Loon, M., Roemer, M. G. M., Builtjes, P. J. H., Bessagnet, B., Rouil, L., Christensen, J. H., Brandt, J., Fagerli, H., Tarrason, L., and Rodgers, I.: Model inter-comparison in the framework of the review of the unified EMEP model, Techical report R2004/282, TNO, 2004.

Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., and Kukkonen, J.: Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki, Sci. Total Environ., 409, 1559–1571, 2011.

GMDD
8, 1029–1075, 2015

Vogel, B., Vogel, H., Bäumer, D., Bangert, M., Lundgren, K., Rinke, R., and Stanelle, T.: The comprehensive model system COSMO-ART – Radiative impact of aerosol on the state of the atmosphere on the regional scale, Atmos. Chem. Phys., 9, 8661–8680, doi:10.5194/acp-9-8661-2009, 2009.

Wolff, G. T. and Lioy, P. J.: An empirical model for forecasting maximum daily ozone levels in the northeastern United States, J. Air Pollut. Control Assoc., 28, 1034–1038, 1978.

Yahya, K., Zhang, Y., and Vukovich, J. M.: Real-time air quality forecasting over the south-eastern United States using WRF/Chem-MADRID: multiple-year assessment and sensitivity studies, Atmos. Environ., 92, 318–338, 2014.

Yang, Q., W. I. Gustafson Jr., Fast, J. D., Wang, H., Easter, R. C., and Morrison, H.: Assessing regional scale predictions of aerosols, marine stratocumulus, and their interactions during VOCALS-REx using WRF-Chem, Atmos. Chem. Phys. Discuss., 11, 22663–22718, doi:10.5194/acpd-11-22663-2011, 2011.

Zhang, K., Wan, H., Wang, B., Zhang, M., Feichter, J., and Liu, X.: Tropospheric aerosol size distributions simulated by three online global aerosol models using the M7 microphysics module, Atmos. Chem. Phys., 10, 6409–6434, doi:10.5194/acp-10-6409-2010, 2010.

Zhang, Y.: Online-coupled meteorology and chemistry models: history, current status, and outlook, Atmos. Chem. Phys., 8, 2895–2932, doi:10.5194/acp-8-2895-2008, 2008.

Zhang, Y., Pan, Y., Wang, K., Fast, J. D., and Grell, G. A.: WRF-Chem-MADRID: incorporation of an aerosol module into WRF-Chem and its initial application to the TexAQS2000 episode, J. Geophys. Res., 115, D18202, doi:10.1029/2009JD013443, 2010a.

Zhang, Y., Wen, X.-Y., and Jang, C. J.: Simulating climate-chemistry-aerosol-cloud radiation feedbacks in continental US using online-coupled WRF-Chem, Atmos. Environ., 44, 3568–3582, 2010b.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, part I, history, techniques, and current status, Atmos. Environ., 60, 632–665, 2012a.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, part II, state of the science, current research needs, and future prospects, Atmos. Environ., 60, 656–676, 2012b.

Žabkar, R., Rakovec, J., and Gaberšek, S.: A trajectory analysis of summertime ozone pollution in Slovenia, Geofizika, 25, 179–202, 2008.

GMDD
8, 1029–1075, 2015

Žabkar, R., Rakovec, J., and Koračin, D.: The roles of regional accumulation and advection of ozone during high ozone episodes in Slovenia: a WRF-Chem modelling study, Atmos. Environ., 45, 1192–1202, 2011.

Žabkar, R., Koračin, D., and Rakovec, J.: A WRF-Chem sensitivity study using ensemble modelling for a high ozone episode in Slovenia and the Northern Adriatic area, Atmos. Environ., 77, 990–1004, 2013.

GMDD

8, 1029–1075, 2015

Table 1. AQ monitoring sites.

Monitoring site	Abbreviation	Type of zone	Altitude (m)	Pollutants	Statistical ozone forecast
Celje	CE	Urban	240	O_3, PM_{10}, NO_2	No
Hrastnik	HRA	Urban	290	O_3, SO_2	Yes
Iskrba	ISK	Rural	540	O_3, NO_2	Yes
Koper	KOP	Urban	56	O_3, PM_{10}	Yes
Kovk	KOV	Rural	608	NO_2	No
Krvavec	KRV	Rural	1740	O_3	Yes
Ljubljana	LJ	Urban	299	O_3, PM_{10}, NO_2	Yes
Murska Sobota	MS	Rural	188	O_3, PM_{10}, NO_2	Yes
Nova Gorica	NG	Urban	113	O_3, PM_{10}, NO_2	Yes
Otlica	OTL	Rural	918	O_3	Yes
Sv. Mohor	MOH	Rural	394	NO_2	No
Trbovlje	TRB	Suburban	250	O_3, PM_{10}, NO_2	No
Velenje	VEL	Urban	389	O_3, SO_2	No
Vnajnarje	VNA	Rural	630	NO_2	No
Zadobrova	ZAD	Rural	280	PM_{10}, NO_2	No
Zagorje	ZAG	Urban	241	O_3, PM_{10}, NO_2	No
Zavodnje	ZAV	Rural	765	O_3, NO_2	No

GMDD
8, 1029–1075, 2015

Table 2. Statistical scores for 1 h values of 2 m temperature (T2 m), 10 m wind speed (W10 m) and relative humidity (RH), and for daily average incoming solar radiation (SR). Shown are results for 1 day forecast, calculated separately for three measuring sites (LJ, NG, MS) and for 24 MET monitoring stations (ALL) during the 3 month period. In the case of temperature results for daily maxima are also shown.

Variable	Station	NoCases	Mean	ME	MAE	RMSE	CORR
T2 m 1 h (°C)	LJ	2129	20.3	−1.6	2.3	2.9	0.91
	NG	2184	21.8	−1.1	2.1	2.5	0.94
	MS	2184	19.2	−2	2.3	2.8	0.95
	ALL	47836	18.7	−1.3	2.3	2.9	0.93
T2 m max (°C)	LJ	89	26.5	−1.6	1.8	2.1	0.98
	NG	90	26.8	−3	3	3.3	0.96
	MS	90	26.2	−1.7	1.8	2	0.98
	ALL	1976	24.2	−2.1	2.7	3.2	0.97
W10 m ($m s^{-1}$)	LJ	2129	1.5	0	0.7	1	0.58
	NG	2183	2.7	1	1.4	1.9	0.35
	MS	2184	2.3	0.4	1.1	1.4	0.53
	ALL	43378	2.4	0.8	1.4	1.9	0.36
RH (%)	LJ	2066	62	−2	8	10	0.85
	NG	2121	62	−1	12	15	0.75
	MS	2121	69	3	8	11	0.88
	ALL	48556	68	2	11	14	0.77
SR ($W m^{-2}$)	LJ	90	276	19	31	43	0.84
	NG	90	278	4	32	43	0.77
	MS	90	273	15	26	37	0.9
	ALL	1710	273	16	35	49	0.77

GMDD
8, 1029–1075, 2015

Table 3. Domain wide performance statistics for 1 day and 2 day forecast in $\mu g\,m^{-3}$. For different pollutants statistics for all hourly (hour), 8 h averages (8 h), 8 h daily maximum (8 h max), daily maximum (max) or daily average (day) concentrations are shown.

		NoCases	Mean	ME	MAE	RMSE	CORR
O_3 (hour)	1 day	28391	94	13.7	25.5	32.5	0.65
	2 day	28391	94.2	13.8	25.8	32.9	0.64
O_3 (8 h)	1 day	28072	94.1	13.8	22.9	28.5	0.69
	2 day	28072	94.2	13.8	23.3	28.9	0.68
O_3 (8 h max)	1 day	1157	110.7	−0.7	13.6	17	0.77
	2 day	1157	110.9	−1	14.1	17.4	0.75
O_3 (max)	1 day	1170	115.8	−3.4	13.7	17.1	0.81
	2 day	1170	115.8	−3.9	14.4	17.9	0.79
NO_2 (hour)	1 day	26178	7.3	−5.1	7.5	10.8	0.3
	2 day	26178	7.5	−4.9	7.6	10.8	0.3
PM_{10} (day)	1 day	718	29	7.1	12	18.8	0.34
	2 day	718	29.1	7.2	12	19.1	0.37

GMDD
8, 1029–1075, 2015

Table 4. Discrete evaluation of 1 h daily maximum ozone predictions.

Stations	Threshold, NoCases	Forecast	Mean ($\mu g\,m^{-3}$)	ME ($\mu g\,m^{-3}$)	MAE ($\mu g\,m^{-3}$)	RMSE ($\mu g\,m^{-3}$)	CORR	MNBE (%)	MNGE (%)	IOA
All	> 0	F 1day	115.8	−3.4	13.7	17.1	0.81	−0.5	12.0	0.86
	1170	F 2day	115.8	−3.9	14.4	17.9	0.79	−0.7	12.6	0.84
		PER 1day	119.5	−0.4	15.8	21.1	0.65	1.6	14.5	0.81
		PER 2day	119.8	−0.4	21.7	27.7	0.39	2.8	19.6	0.65
	> 140	F 1day	143.3	−11.9	15.7	18.4	0.52	−7.4	9.9	0.57
	1102	F 2day	140.6	−14.6	17.1	20.0	0.41	−9.1	10.8	0.47
		PER 1day	145.0	−10.2	15.6	19.6	0.41	−6.5	10.0	0.52
		PER 2day	135.8	−19.4	24.76	29.2	0.31	−12.4	15.9	0.38
Sub-alpine urban with SF (LJ, HRA)	> 0	F 1day	115.3	1.1	10.7	14.0	0.84	3.4	11.1	0.91
	180	F 2day	115.4	0.8	12.0	15.2	0.80	3.5	12.2	0.88
		PER 1day	114.3	−0.3	16.7	21.7	0.64	2.2	16.5	0.80
		PER 2day	114.6	−0.3	21.9	27.8	0.41	3.9	21.6	0.65
		SF 1day	114.0	−0.5	11.9	15.7	0.81	1.6	11.2	0.88
		SF 2day	116.2	0.6	13.4	17.1	0.75	3.2	12.7	0.84
Rural with SF (MS, ISK, KRV, OTL)	> 0	F 1day	115.2	−8.1	14.6	17.6	0.80	−5.0	11.7	0.85
	360	F 2day	114.8	−8.8	15.5	18.9	0.77	−5.4	12.4	0.81
		PER 1day	123.6	−0.3	15.0	20.7	0.65	1.4	13.1	0.81
		PER 2day	124.1	−0.4	21.6	27.8	0.37	2.4	18.5	0.64
		SF 1day	121.5	−2.9	15.0	19.4	0.74	−0.7	12.2	0.83
		SF 2day	122.9	−1.8	15.8	20.5	0.67	0.5	13.2	0.79
Mediterranean urban with SF (KOP, NG)	> 0	F 1day	123.5	−11.8	17.4	22.5	0.76	−6.9	12.5	0.80
	179	F 2day	124.5	−11.2	17.2	21.8	0.77	−6.5	12.4	0.82
		PER 1day	135.9	−0.5	17.4	23.0	0.68	1.2	13.8	0.83
		PER 2day	136.0	−0.2	25.2	31.5	0.41	2.8	19.7	0.66
		SF 1day	129.3	−7.0	15.9	20.7	0.75	−3.6	11.6	0.83
		SF 2day	131.6	−4.5	15.6	20.4	0.74	−1.6	11.6	0.84

GMDD
8, 1029–1075, 2015

Table 5. Categorical evaluation of 1 h daily maximum ozone predictions for different thresholds, calculated for 8 monitoring sites with available statistical forecast.

Threshold	Forecast	A	B	FAR	POD	CSI	a	b	c	d
> 120	F 1day	0.78	0.76	0.13	0.66	0.61	35	239	317	121
	F 2day	0.77	0.75	0.14	0.64	0.59	37	233	307	127
	PER 1day	0.74	0.99	0.25	0.74	0.59	91	267	249	93
	PER 2day	0.64	1.00	0.34	0.65	0.49	123	235	209	124
	SF 1day	0.80	1.02	0.21	0.81	0.67	67	257	243	61
	SF 2day	0.77	1.03	0.23	0.80	0.65	77	264	225	66
> 140	F 1day	0.84	0.59	0.14	0.51	0.47	17	103	492	100
	F 2day	0.82	0.60	0.18	0.49	0.44	22	99	479	104
	PER 1day	0.82	1.00	0.31	0.69	0.53	62	141	435	62
	PER 2day	0.72	1.00	0.48	0.52	0.35	97	106	391	97
	SF 1day	0.79	0.73	0.29	0.52	0.43	40	99	398	91
	SF 2day	0.79	0.70	0.27	0.51	0.43	37	98	403	94
> 160	F 1day	0.91	0.37	0.32	0.25	0.22	9	19	627	57
	F 2day	0.91	0.30	0.26	0.22	0.21	6	17	622	59
	PER 1day	0.92	1.00	0.38	0.62	0.45	29	47	595	29
	PER 2day	0.88	1.00	0.56	0.43	0.28	43	33	572	43
	SF 1day	0.90	0.49	0.35	0.32	0.27	13	24	539	52
	SF 2day	0.90	0.63	0.41	0.37	0.29	19	27	540	46

GMDD
8, 1029–1075, 2015

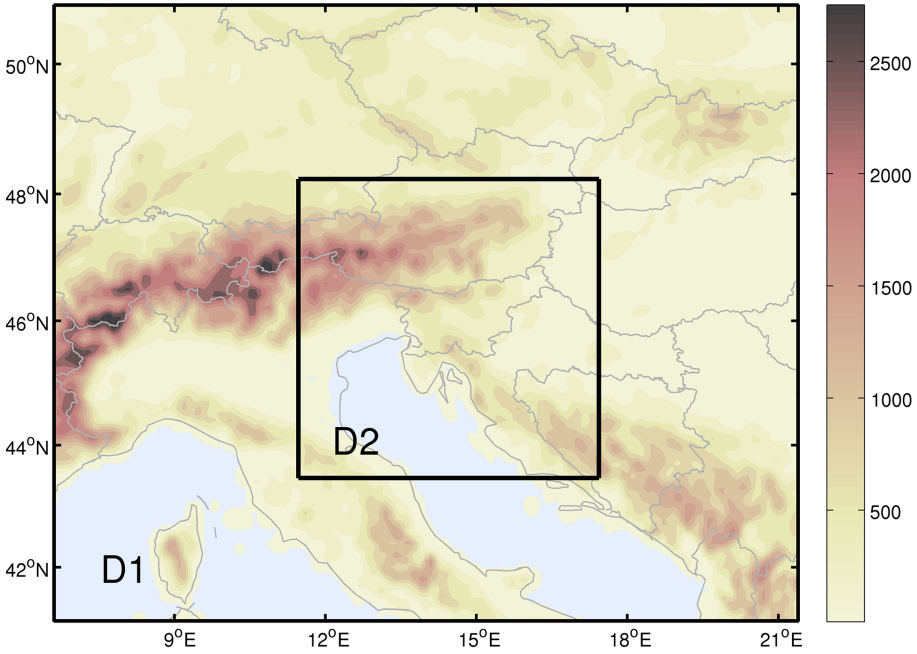

Figure 1. Modelling domains (D1, D2) used in WRF-Chem RT-AQF system. Orography (in meters) is shown in resolution of D1 domain (11.1 km).

GMDD
8, 1029–1075, 2015

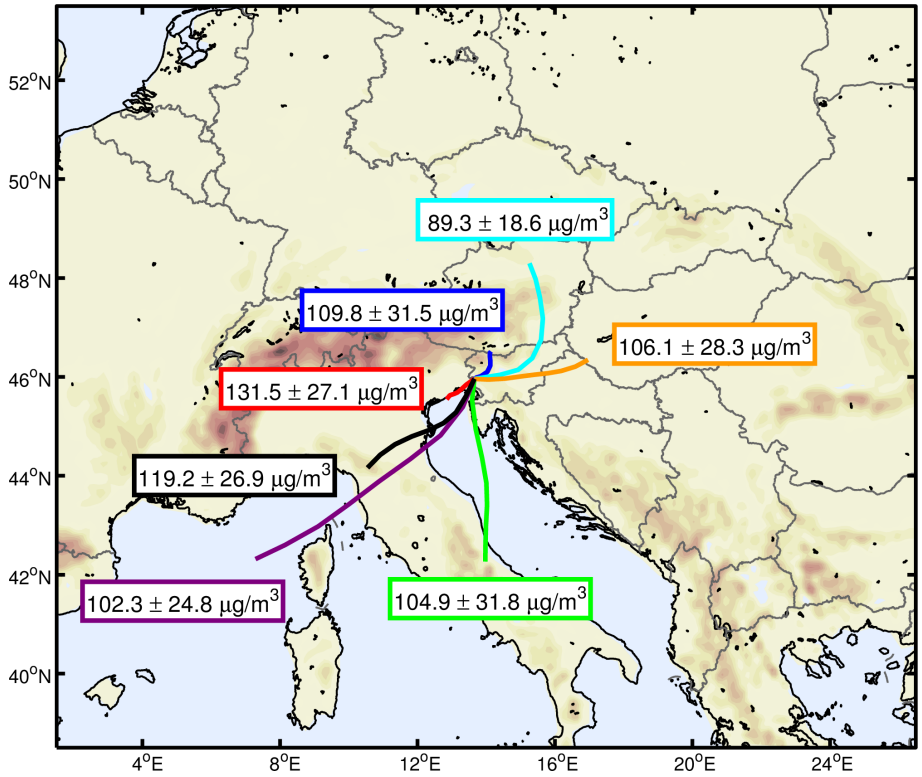

Figure 2. Example of ozone analysis for the Nova Gorica (NG) monitoring site (average daily maximum ± SD) for 7 clusters of similar trajectories, as used in the statistical ozone daily maximum forecast for the NG station.

GMDD
8, 1029–1075, 2015

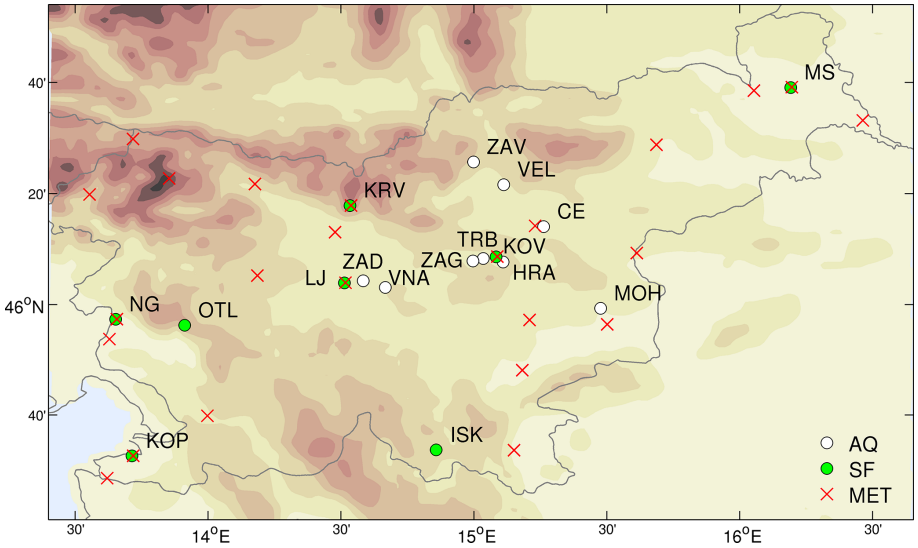

Figure 3. Locations of monitoring stations used in evaluation of air quality variables (AQ stations; shown are also station abbreviations) and meteorological variables (MET stations). Green dots indicate measuring sites with available ozone daily maximum statistical forecast (SF). For the meaning of abbreviations of AQ sites see Table 1.

GMDD
8, 1029–1075, 2015

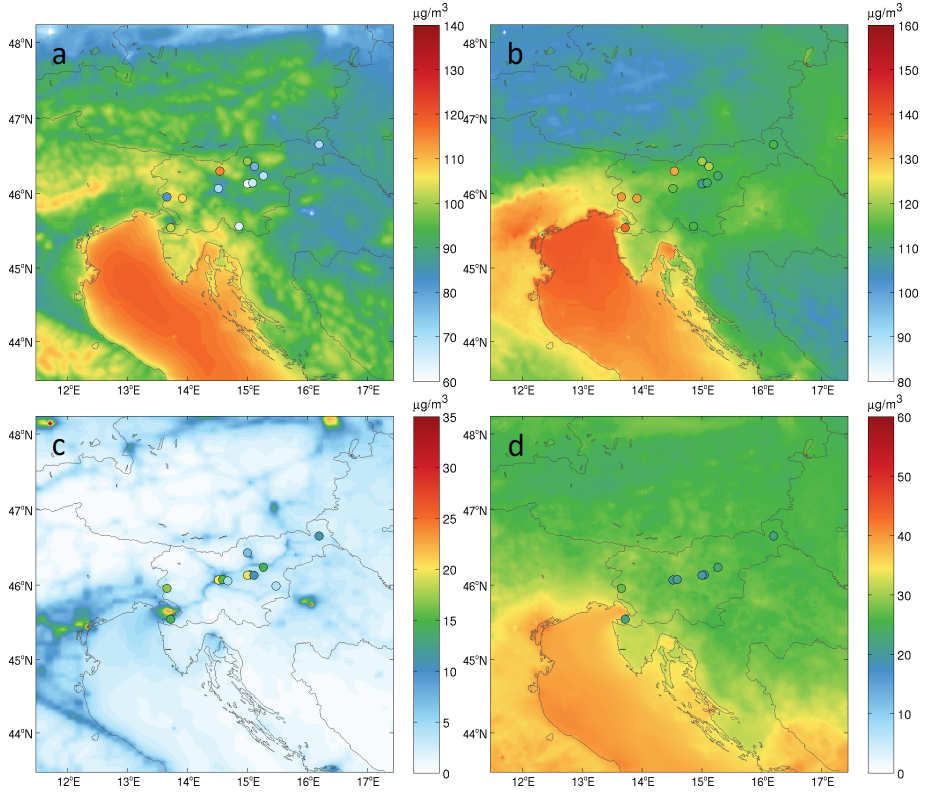

Figure 4. Three month average 1 day predictions of **(a)** hourly O_3, **(b)** O_3 daily maximum, **(c)** hourly NO_2, and **(d)** daily PM_{10}concentrations, overlaid with measurements.

GMDD
8, 1029–1075, 2015

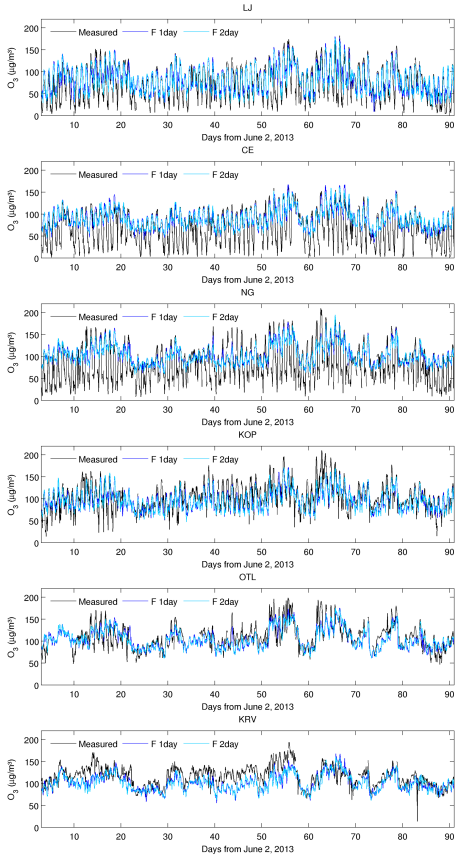

Figure 5. Time evolution of hourly ozone concentrations for 1 day (F 1day) and 2 day (F 2day) WRF-Chem predictions and measurements for some stations during the 3 month period.

GMDD
8, 1029–1075, 2015

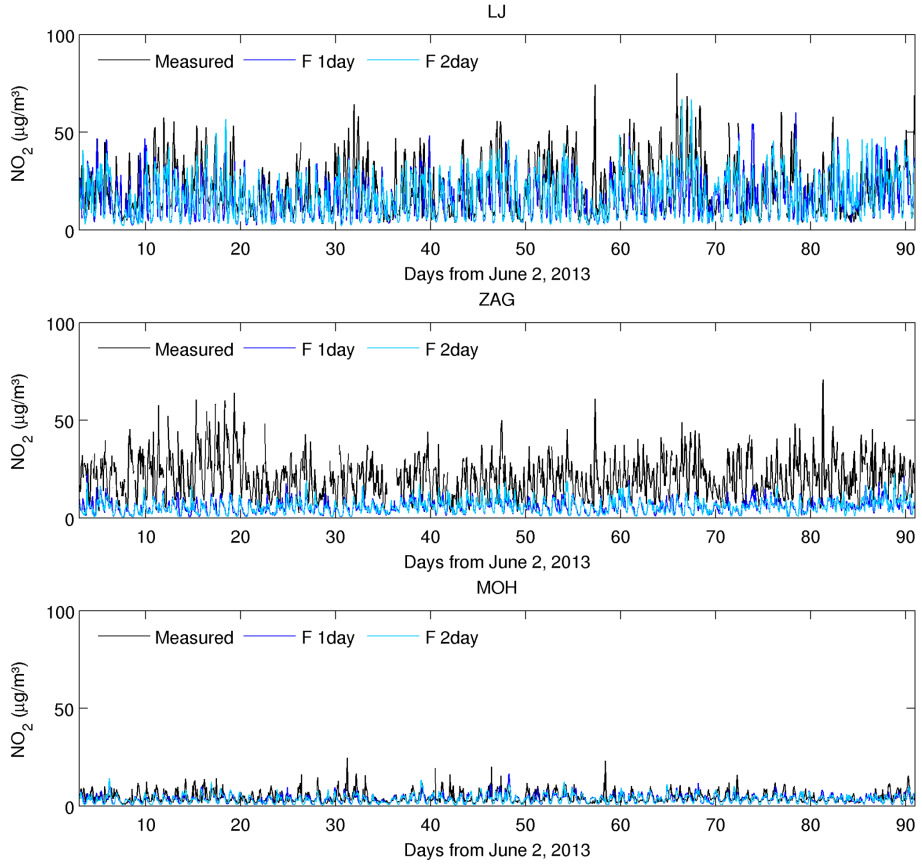

Figure 6. The same as Fig. 5 but for NO_2 at LJ, ZAG and MOH stations.

GMDD
8, 1029–1075, 2015

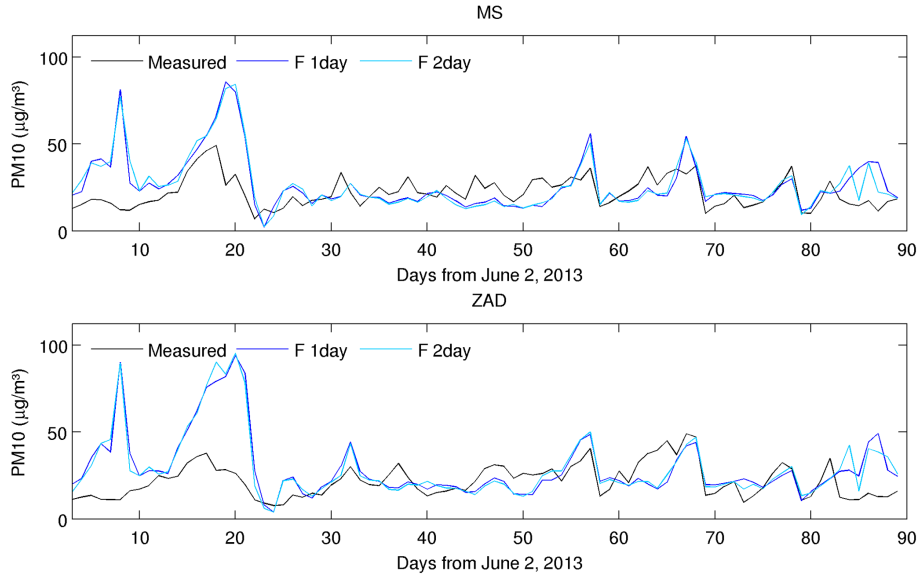

Figure 7. The same as Fig. 5, but for daily PM_{10} concentrations at MS and ZAD stations.

GMDD
8, 1029–1075, 2015

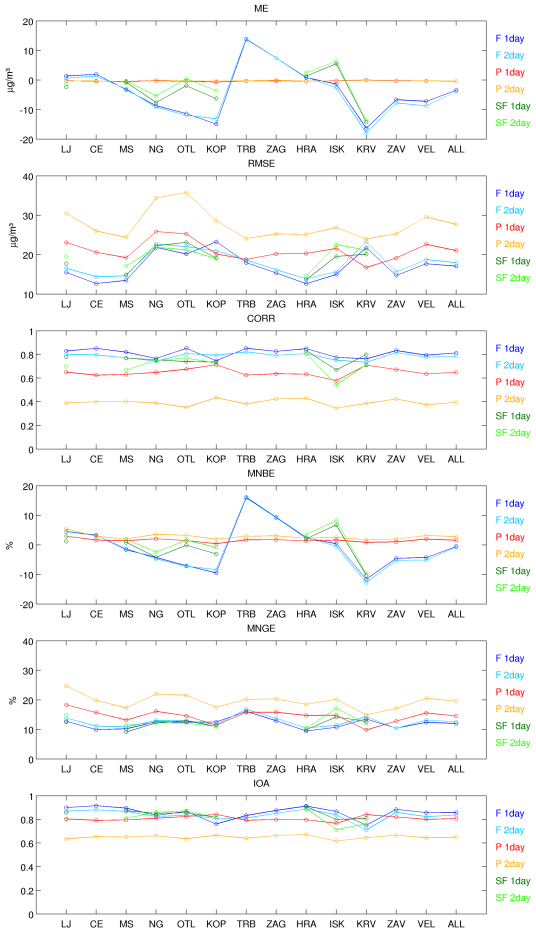

Figure 8. Site-by-site comparison of discrete statistics for 1 day and 2 day WRF-Chem (F 1day, F 2day), statistical (SF 1day, SF 2 day) and persistence model (P 1day, P 2day) predictions of ozone daily maxima during the 3 analyzed summer months.

GMDD
8, 1029–1075, 2015

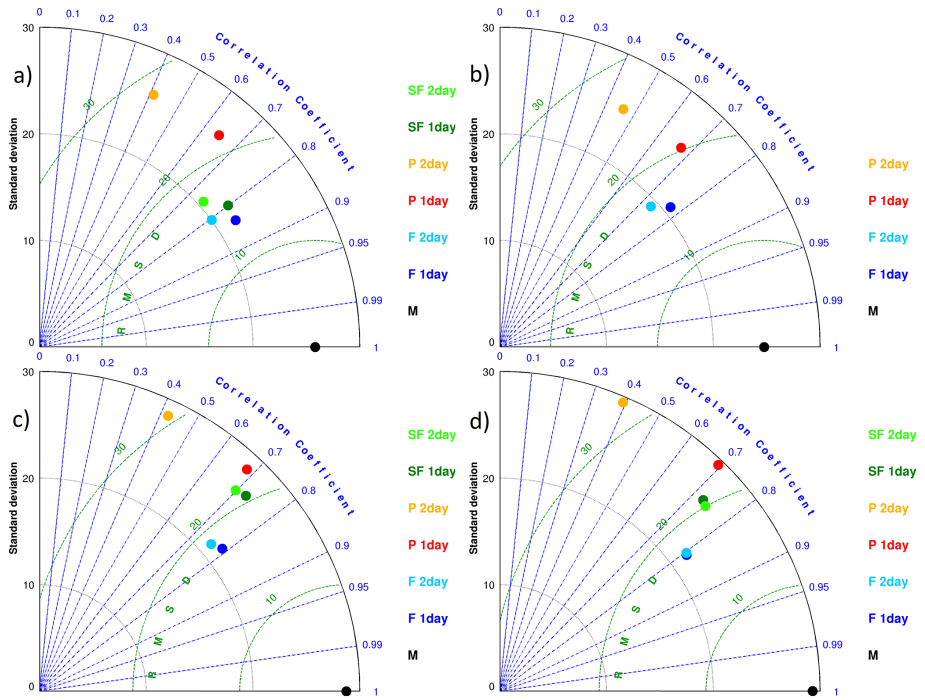

Figure 9. Taylor diagrams comparing 1 day and 2 day ozone daily maximum statistical forecast (SF), persistence (P) and WRF-Chem forecast (F) for **(a)** sub-alpine urban stations with SF (LJ, HRA), **(b)** sub-alpine urban stations without SF (CE, TRB, ZAG, VEL), **(c)** rural stations with SF (MS, ISK, KRV, OTL) and **(d)** Mediterranean urban stations (NG, KOP).

GMDD
8, 1029–1075, 2015