



# Evaluation of the high resolution WRF-Chem (v3.4.1) air quality forecast and its comparison with statistical ozone predictions

R. Žabkar<sup>1,2</sup>, L. Honzak<sup>2,a</sup>, G. Skok<sup>1,2</sup>, R. Forkel<sup>3</sup>, J. Rakovec<sup>1,2</sup>, A. Ceglar<sup>4,2,b</sup>, and N. Žagar<sup>1,2</sup>

<sup>1</sup>University of Ljubljana, Faculty of Mathematics and Physics, Ljubljana, Slovenia

<sup>2</sup>Center of Excellence SPACE-SI, Ljubljana, Slovenia

<sup>3</sup>Karlsruher Institut für Technologie, Institut für Meteorologie und Klimaforschung, Atmosphärische Umweltforschung, Garmisch-Partenkirchen, Germany

<sup>4</sup>University of Ljubljana, Biotechnical Faculty, Ljubljana, Slovenia

<sup>a</sup>now at: BO-MO d.o.o., Ljubljana, Slovenia

<sup>b</sup>now at: Institute for Environment and Sustainability, Joint Research Centre, Ispra, Italy

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

Received: 15 January 2015 – Published in Geosci. Model Dev. Discuss.: 4 February 2015

Revised: 15 June 2015 – Accepted: 19 June 2015 – Published: 16 July 2015

**Abstract.** An integrated modelling system based on the regional online coupled meteorology–atmospheric chemistry WRF-Chem model configured with two nested domains with horizontal resolutions of 11.1 and 3.7 km has been applied for numerical weather prediction and for air quality forecasts 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<sub>3</sub>) daily maxima, the first- and second-day model predictions have been also compared to the operational statistical O<sub>3</sub> forecast and to the 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, the correlation coefficient shows the highest skill for WRF-Chem model O<sub>3</sub> predictions, confirming the significance of the non-linear processes taken into account in an online 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). The 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 Liroy, 1978), or statistical models taking advantage of links between pollutant concentrations, meteorological variables (wind speed and direction, temperature, 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., 1994; 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 of around 1 h. Examples are the EURAD ([http://db.eurad.uni-koeln.de/index\\_e.html](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 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, thus, provides 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 et al., 2012; Baklanov et al., 2014). There is an increasing awareness that an integrated online approach is needed not only for assessment, forecasting 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 to determine 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) with coupled meteorological, microphysical, chemical, and radiative processes for forecasting AQ in Slovenia during 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 the WRF-Chem model has been successfully used to simulate historical poor AQ conditions with a hindcast approach. To our knowledge, only a few studies focused on using WRF-Chem for forecast-

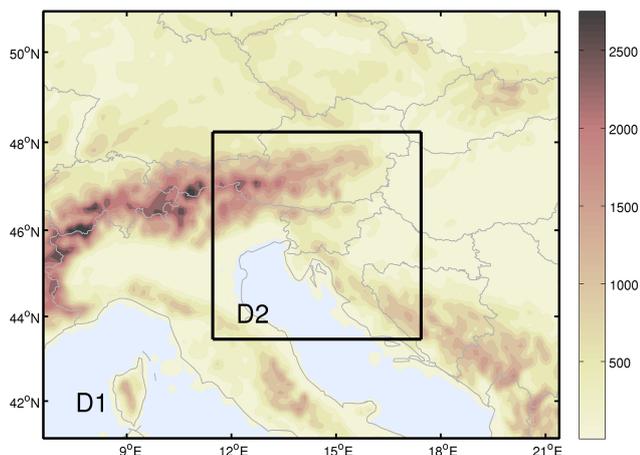
ing AQ, most of these applied the WRF-Chem forecast before and during field campaigns (McKeen et al., 2005, 2007, 2009; Yang et al., 2011). Takigawa et al. (2007) evaluated the O<sub>3</sub> forecast for a 1-month time period from a one-way nested global–regional RT-AQF system with full chemistry based on the global CHASER (Sudo et al., 2002) and regional WRF-Chem models, while Saide et al. (2011) evaluated a forecast system based on the WRF-Chem model for simulating carbon monoxide (CO) as a PM<sub>10</sub>/PM<sub>2.5</sub> 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 evaluation studies published in the literature, an increasing number of real-time weather and air quality forecasting systems based on WRF-Chem are implemented worldwide ([http://ruc.noaa.gov/wrf/WG11/Real\\_time\\_forecasts.htm](http://ruc.noaa.gov/wrf/WG11/Real_time_forecasts.htm)).

In our study we explore the forecasting skill of WRF-Chem model over the topographically complex and geographically diverse area of Slovenia for three summer months (June–August 2013). Furthermore, in the case of O<sub>3</sub> we compare WRF-Chem predictions with a statistical model for predicting O<sub>3</sub> daily maxima, currently used at the 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 an accurate and reliable forecasting system could be useful to the local authorities and could help to advise the public the proper preventive actions, we want to answer the question of 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 the potential that lies in the approach based on a unified model for forecasting weather and air pollution. Finally, identified strengths, limitations and deficiencies of analysed 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



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

(Fig. 1) with horizontal resolutions of 11.1 and 3.7 km, and  $151 \times 100$  and  $181 \times 145$  grid points, respectively. A one-way nesting is applied by two separate consecutive simulations, where outputs from the coarse grid integration are processed to provide boundary conditions for the nested run every 15 min. The vertical structure of the atmosphere is resolved with 42 vertical levels extending up to 50 hPa, with the highest resolution of  $\sim 25$  m near the ground. 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 conditions (ICs) and lateral boundary conditions (BCs) taken from the  $0.5^\circ$  data of the Global Forecast System (GFS) operated by the US National Weather Service (NWS). For chemical BCs, forecasts from the 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's 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 also used to initialize chemistry. A 3-day spin-up ahead of the first analysed 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 these can have a strong impact on the model predictions. In both domains we decided to apply the same schemes as were used in simulation S11 for phase-2 of the Air Quality Model Evaluation International Initiative (AQMEII) (e.g. Balzarini et al., 2015; Baró et al., 2015; Curci et al., 2015; Forkel et al., 2015; Im et al., 2015a, b; Kong et al., 2015; San José et al., 2015). These include the 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., 2009), Fast-J photolysis scheme (Wild et al., 2000), RADM2 gas-phase chemistry (Regional Acid Deposition Model, Stockwell et al., 1990) and the MADE/SORGAM aerosol module (Ackermann et al., 1995; Schell et al., 2001). Current model implementation includes a modified RADM2 gas-phase chemistry solver as described in Forkel et al. (2015), which avoids underrepresentation of nocturnal  $O_3$  titration in areas with high NO emissions. According to Forkel et al. (2015) the modified solver tends to overestimate 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 (quasi steady state approximation) RADM2 solver supplied originally with the 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. (2015) for two air pollution episodes, this degree of aerosol–meteorology interactions in the 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.

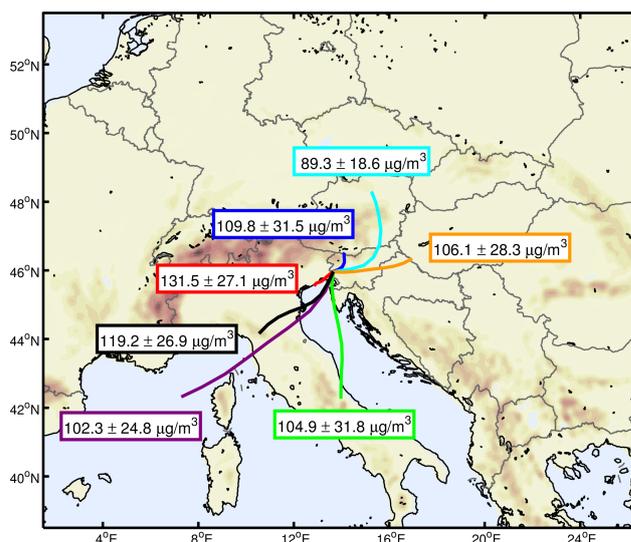
Biogenic emissions are estimated using MEGAN (Model of Emissions of Gases and Aerosols from Nature; Guenther et al., 2006) online model calculations, while dust emissions are modelled according to Shaw et al. (2008) with an adjustment to avoid high dust fluxes from some Dalmatian Islands in Croatia. A detailed anthropogenic inventory for pollutants CO,  $NH_3$ ,  $NO_x$ ,  $SO_2$ , and NMVOCs (non-methane volatile organic compounds), created for the purpose of AQ forecasting constructed for year 2009 by SEA (SEA, 2014), 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>.

## 2.2 Statistical ozone daily maximum forecast

The statistical O<sub>3</sub> model (Žabkar, 2011), currently used at SEA for forecasting O<sub>3</sub> daily maxima at eight measuring sites in Slovenia (Fig. 3), is a multivariate regression tool combined with clustering algorithms to take into account measured data, weather forecast data, and the predicted backward trajectories of each monitoring site. As regards measurements, the previous day's (at 12:00, 15:00, 18:00 and 21:00 local time, daily maximum, daily minimum, daily average) and present day's early morning (07:00 local time) meteorological (pressure, relative humidity, direct and diffusive solar radiation, wind speed) and AQ data (O<sub>3</sub>, NO<sub>x</sub>, NO<sub>2</sub>, CO, PM<sub>10</sub>, SO<sub>2</sub>) 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 using the stepwise technique based on *F* statistics, 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 the calculation of 24 h backward trajectories on meteorological fields of the ALADIN/SI forecast. The inclusion of 24 h predicted trajectories into a statistical model is based on the study of Žabkar et al. (2008) which showed that the highest O<sub>3</sub> daily maxima at monitoring sites in Slovenia are in general associated with short (slow-moving) backward trajectories with a southwestern origin, while the lowest measured daily maximum O<sub>3</sub> values for all the stations are associated with the clusters of long northwestern trajectories. Clusters of similar trajectories were for the purpose of the statistical forecast calculated by *k* means clustering algorithms (Moody and Galloway, 1988; Žabkar et al., 2008) on 6 years (2004–2010) of data (ALADIN/SI trajectories). As an example, Fig. 2 shows a mean O<sub>3</sub> daily maxima for clusters of similar trajectories for one of the monitoring sites. The same 6-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 and trajectory clusters, from measurements, ECMWF forecast data, average cluster O<sub>3</sub> daily maximum, and day-of-the-year variable.

The first step of the statistical O<sub>3</sub> prediction is the calculation of trajectories approaching the monitoring stations at 12:00 UTC of the forecast day. In the next step, the backward trajectories of each monitoring site are associated with the nearest pre-calculated cluster of similar trajectories. Finally, the multiple regression equation of the associated group of trajectories is used to calculate the O<sub>3</sub> daily maximum prediction. It must also be noted that the decision on declaring O<sub>3</sub> episodes is only partially based on the results from this statistical model, it also involves a decision made by AQ forecasters.



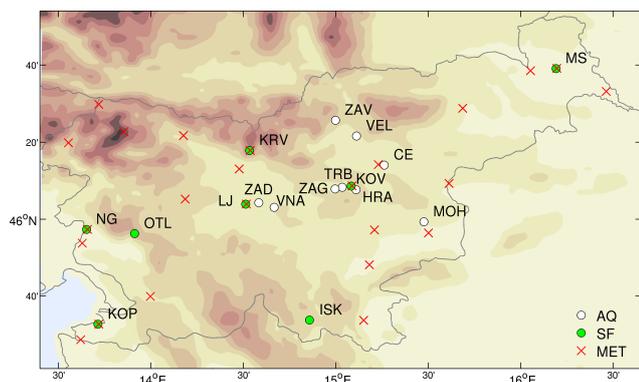
**Figure 2.** Example of ozone analysis for the Nova Gorica (NG) monitoring site (average daily maximum  $\pm$  standard deviation) for seven clusters of similar trajectories, as used in the statistical ozone daily maximum forecast for the NG station.

## 2.3 Evaluation methodology

We evaluate the 1-day and 2-day WRF-Chem meteorological and AQ forecasts on the high-resolution domain during a 3-month period (June–August 2013). The main focus is on O<sub>3</sub> predictions. In the case of air pollutants, the instantaneous lowest model level mixing ratios (with grid point centre at about 12 m above model orography – an exception is the KRV station as explained below) are compared to the hourly averaged concentrations measured at monitoring stations (which have a typical inlet height of 3 m) 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, including comparison of the station altitude, the height of model orography, model analysis height, and pollutants with higher than 75 % availability of valid data during the analysed time period for each of the AQ monitoring sites. In the case of the elevated alpine KRV station, AQ variables are evaluated for the fifth model layer instead of the first model layer. We made this exception for KRV since the height of the model topography was significantly underestimated there (Table 1) and the station is known to be strongly influenced by the conditions of the free troposphere. The selection of the fifth model layer for the KRV station is based on analyses performed for different model layers (results not shown) and was found to reduce the negative bias for O<sub>3</sub> due to too low WRF-Chem topography at this location. Although even for this model layer the location of the grid point representing the KRV station (1414 m) is still well below the true station altitude (1740 m), the O<sub>3</sub> bias for the KRV station is signifi-

**Table 1.** AQ monitoring sites.

Monitoring site	Abbreviation	Type of zone	Altitude (m)	Model orography (m)	Model analysis height (m)	Pollutants	Statistical ozone forecast
Celje	CE	Urban	240	300	313	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	No
Hrastnik	HRA	Urban	290	540	552	O <sub>3</sub> , SO <sub>2</sub>	Yes
Iskrba	ISK	Rural	540	579	591	O <sub>3</sub> , NO <sub>2</sub>	Yes
Koper	KOP	Urban	56	72	85	O <sub>3</sub> , PM <sub>10</sub>	Yes
Kovk	KOV	Rural	608	516	528	NO <sub>2</sub>	No
Krvavec	KRV	Rural	1740	1272	1414	O <sub>3</sub>	Yes
Ljubljana	LJ	Urban	299	287	300	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	Yes
Murska Sobota	MS	Rural	188	189	202	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	Yes
Nova Gorica	NG	Urban	113	150	163	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	Yes
Otlica	OTL	Rural	918	874	886	O <sub>3</sub>	Yes
Sv. Mohor	MOH	Rural	394	254	266	NO <sub>2</sub>	No
Trbovlje	TRB	Suburban	250	459	471	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	No
Velenje	VEL	Urban	389	461	474	O <sub>3</sub> , SO <sub>2</sub>	No
Vnajnarje	VNA	Rural	630	468	480	NO <sub>2</sub>	No
Zadobrova	ZAD	Rural	280	275	287	PM <sub>10</sub> , NO <sub>2</sub>	No
Zagorje	ZAG	Urban	241	431	443	O <sub>3</sub> , PM <sub>10</sub> , NO <sub>2</sub>	No
Zavodnje	ZAV	Rural	765	678	690	O <sub>3</sub> , NO <sub>2</sub>	No

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

cantly smaller than for the first layer, while the correlation coefficient between the measured and simulated O<sub>3</sub> levels remains similar in both cases (the fifth or the lowest model layer). Taking results from higher model layers would further decrease the negative model bias but would also worsen the correlation coefficient for O<sub>3</sub> at this station due to decreased impact of surface processes.

All AQ stations are background, seven are measuring urban background, one suburban and nine rural conditions. Valid O<sub>3</sub> measurements are for the analysed time period available for 13 AQ stations. When studying the general model performance, data from four additional stations for

two other pollutants (NO<sub>2</sub>, PM<sub>10</sub>) are also analysed to get a better picture of model behaviour over the domain, known for its large topographical and climate diversity. The coverage of three climate zones in Slovenia (Mediterranean, subalpine and mountainous) with monitoring stations is the following: NG, KOP and OTL are Mediterranean sites, KRV is a mountainous station, and the remaining stations are subalpine. As the elevated station KRV, the ISK, OTL and VNA stations are also influenced by regional transport of pollutants.

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 solar 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 information about the AQ forecast can also be gained by the evaluation of meteorological forecasts 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 the model's forecasting skills of meteorological and AQ variables. In the case of O<sub>3</sub>, 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 centred root mean square difference and the amplitude of their variations, represented by their standard deviations. Furthermore, some additional discrete statistical measures, including in-

dex of agreement (IOA), the mean normalized bias error (MNBE), and the mean normalized gross error (MNGE) are calculated for O<sub>3</sub> daily maximum concentrations predicted by the different models. Finally, to evaluate the model's ability to predict exceedances and non-exceedances, several categorical indices including equitable threat score (ETS), critical success index (CSI), bias (B), false alarm ratio (FAR) and probability of detection (POD) are calculated for different thresholds. Definitions of the statistical measures are shown in Appendix A.

## 2.4 Meteorology and air quality of June–August 2013

The analysed 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 a measured temperature daily maximum of up to 35 °C occurred after a rather cold beginning of the month and lasted from 15 to 21 June. The event was terminated by a cold front passage and followed by the pronounced cold 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 01 to 08 August. On the last day of this episode, 08 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<sub>10</sub>, were in general low during the analysed time period. The only exception was O<sub>3</sub> with exceedances of the 8 h target value (120 µg m<sup>-3</sup>) 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 following two events (in July and August) threshold exceedances of 1 h daily maxima were also recorded for O<sub>3</sub>. In spite of the hot and sunny conditions during the first heat wave event in June 2013, measured daily O<sub>3</sub> maxima at the Slovenian stations did not exceed the 1 h information threshold value (1 h ITV; 180 µg m<sup>-3</sup>) but reached 171 µg m<sup>-3</sup> at the Mediterranean OTL and the elevated alpine KRV stations. During the second heat wave event, the 1 h daily maxima exceeded 180 µg m<sup>-3</sup> at KRV, OTL, NG and KP (23–28 July), and the highest number of 1 h exceedances (20) was measured in July at the OTL station. Similarly, during the August heat wave event O<sub>3</sub> concentrations exceeded the 1 h ITV at LJ, MB, OTL, NG and KP from 02 to 07 August. To summarize, the Mediterranean stations (NG, OTL, KP), due to very high O<sub>3</sub> concentrations measured during the heat wave events (especially the second two events), exhibited the poorest AQ in Slovenia during the analysed time pe-

riod, while the legislation's limit values were exceeded only occasionally for the subalpine stations.

## 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), RH and incoming SR. Results for three selected measuring sites (LJ, NG, MS) and the overall result for all 24 MET monitoring sites (shown in Fig. 3) are presented separately.

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. For almost all sites the mean SR was overestimated by the model, with overall MEs of 16 and 11 W m<sup>-2</sup> for 1-day and 2-day forecasts, respectively. CORR was higher for 1-day (0.77) than for 2-day (0.71) forecasts, with a range of 0.64–0.90 for 1-day forecasts at different stations. The larger positive bias for the first day than for the second day can be attributed to less cloudy conditions during the first day of simulation.

In the case of T2 m 1-day (2-day), the 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 positive bias, daily T2 m maxima were simulated with MEs between -3.9 and -0.6 °C, depending on the station's spatial representativeness. All meteorological variables, including soil temperature and soil moisture, are always initialized with GFS data. This explains higher negative bias for T2 m during the first day of simulation in spite of the overestimated solar radiation. An average systematic underestimation of T2 m daily maxima was -2.1 °C both for 1-day and 2-day forecasts. Nighttime T2 m minima showed lower systematic bias for the 2-day forecast, which resulted in an overall bias for all hourly T2 m values of -1.3 °C for 1-day and -0.8 °C for 2-day forecasts. Predominantly 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 an overall ME of 0.8 m s<sup>-1</sup> for 1-day and 2-day forecasts, which for some stations bias can be very low (e.g. LJ; Table 2) and much higher for other stations due to their local positioning in complex topography (e.g. HRA located in valley with ME of 1.9 m s<sup>-1</sup>). For hourly values the correlation is lower (Table 2), but for mean daily W10 m values a Pearson correlation coefficient of between 0.4 and 0.9 has been simulated, depending on the monitoring site. Relative humidity shows slightly better results for 1-day than for 2-day forecasts with CORR of 0.77 and low overall ME of 2 % for the 1-day fore-

**Table 2.** Statistical scores for 1 h values of 2 m temperature (T2 m), 10 m wind speed (W10 m), 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	No. cases	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	47 836	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 <sup>−1</sup> )	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	43 378	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	48 556	68	2	11	14	0.77
SR (W m <sup>−2</sup> )	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

cast, which for particular stations can be positive (e.g. KRV) or negative (e.g. LJ and NG; Table 2).

Precipitation has an important role in cleansing of the atmosphere by wet deposition and scavenging. On average, the predicted precipitation underestimated the measured 3-month accumulations by −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 forecasts (results not shown). It must also be taken into account that the 3.4.1 model version does not allow including the information about hydrometeors at the boundaries of the nested domain (in the applied 1-way nesting procedure), which contributes to the negative simulated bias of precipitation. A large decrease in the precipitation bias from day 1 to day 2 suggests that different initialization methodology (e.g. using 1-day spin-up for meteorology) could improve the prediction of precipitation events.

### 3.2 Evaluation of air quality variables

In this section we evaluate WRF-Chem predictions for O<sub>3</sub>, NO<sub>2</sub> and PM<sub>10</sub>, 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<sub>3</sub>, 1 h and 8 h averages and daily maxima are analysed separately. The comparison of 1-day

and 2-day forecasts shows that concentrations of air pollutants were somewhat better forecasted 1-day ahead by means of almost all of statistics shown in Table 3, with higher impact on O<sub>3</sub> predictions. Although the 2-day prediction was generally not worse for the majority of meteorological variables, the reason for a better 1-day prediction in the case of O<sub>3</sub> could be the somewhat stronger simulated winds on the second day of simulation. Stronger winds impact the transport and dispersion of pollutants and have the greatest consequence on secondary pollutants (like O<sub>3</sub>) which need time to be formed.

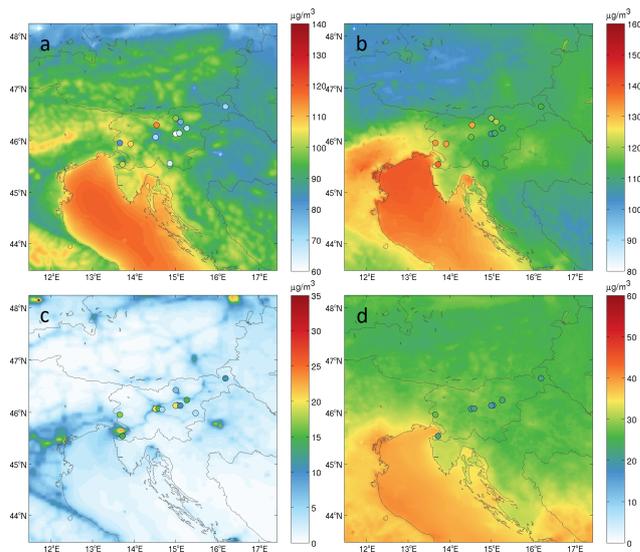
As shown in Table 3 the WRF-Chem simulations tend to overestimate the 1 and 8 h O<sub>3</sub> values with MEs of 14.5 and 14.6 μg m<sup>−3</sup>, respectively. Looking at MAE, RMSE and CORR statistics, agreement with measurements is better for 8 h (22.6, 28.1 μg m<sup>−3</sup> and 0.69) than for 1 h O<sub>3</sub> values (25.1, 32.1 μg m<sup>−3</sup> and 0.65), which is in line with results of previous studies (e.g. Tong and Mauzerall, 2006) and suggests that the current modelling system has problems simulating the small-scale fluctuations of O<sub>3</sub>. On the other hand, evaluations of predicted 8 h and daily O<sub>3</sub> maxima, which are of most concern, show good model performance (ME, MAE RMSE and CORR of −2.7, 13.3, 16.7 μg m<sup>−3</sup> and 0.81 for daily maxima, respectively) in line or even better than those obtained in some previous studies (e.g. Tong and Mauzerall,

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

		No. cases	Mean	ME	MAE	RMSE	CORR
O <sub>3</sub> (hour)	1 day	28 391	94.8	14.5	25.1	32.1	0.65
	2 day	28 391	95.0	14.5	25.5	32.5	0.64
O <sub>3</sub> (8 h)	1 day	28 072	94.8	14.6	22.6	28.1	0.69
	2 day	28 072	95.0	14.6	23.0	28.5	0.68
O <sub>3</sub> (8 h max)	1 day	1157	111.5	−0.1	13.2	16.5	0.77
	2 day	1157	111.6	−0.2	13.7	17.0	0.75
O <sub>3</sub> (max)	1 day	1170	116.5	−2.7	13.3	16.7	0.81
	2 day	1170	116.6	−3.1	14.0	17.5	0.78
NO <sub>2</sub> (hour)	1 day	26 178	7.3	−5.1	7.5	10.8	0.3
	2 day	26 178	7.5	−4.9	7.6	10.8	0.3
PM <sub>10</sub> (day)	1 day	718	29.0	7.1	12.0	18.8	0.34
	2 day	718	29.1	7.2	12.0	19.1	0.37

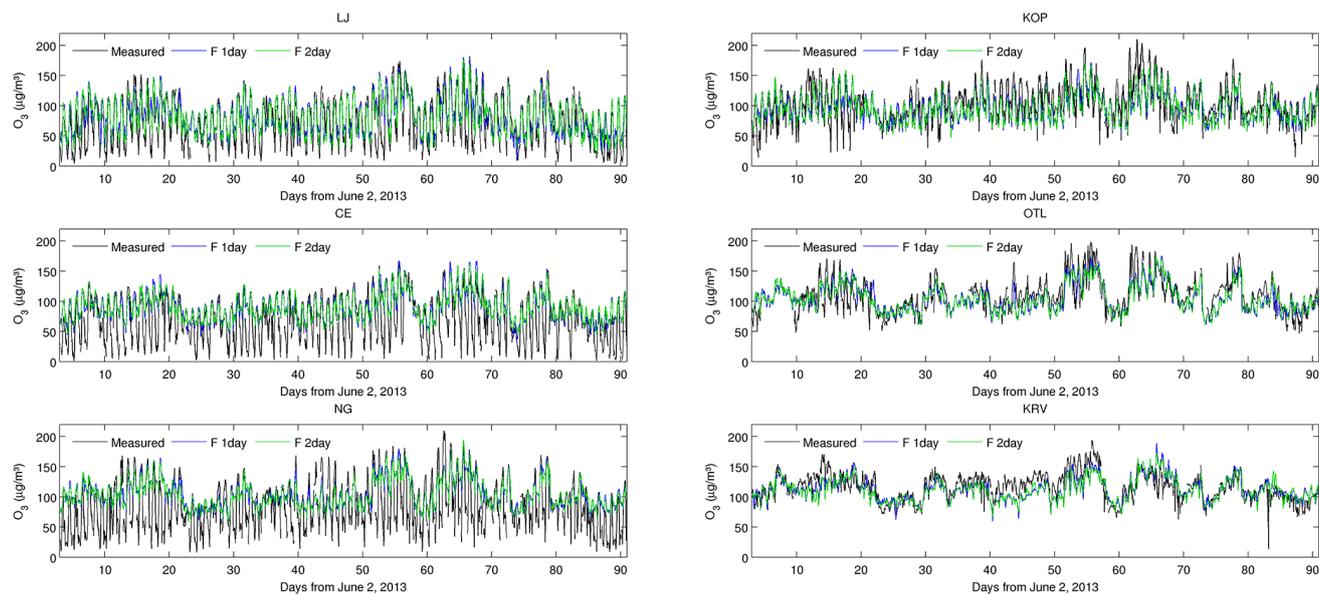
2006; Chuang et al., 2011; Yahya et al., 2014), which could be to some extent related to higher model resolution.

To understand results of the domain-wide statistics (in Table 3) we further analyse spatial and temporal characteristics of model O<sub>3</sub> predictions. Figure 4 shows a spatial pattern of average, simulated 1-day predictions for O<sub>3</sub>, NO<sub>2</sub> and PM<sub>10</sub> overlaid with measured averages where, in the case of O<sub>3</sub>, results for all hourly values and for daily maxima are shown separately. Examples of forecasted and measured time series for O<sub>3</sub> at different stations are shown in Fig. 5. In Fig. 4a the elevated alpine KRV station is the only one with a high negative bias ( $-12 \mu\text{g m}^{-3}$ ) in forecasted 1 h O<sub>3</sub> concentrations at the lowest model layer, which can be explained by the too low altitude of the KRV station in model topography. The high negative bias for hourly O<sub>3</sub> concentrations at the KRV station is reduced to a value of only  $-2 \mu\text{g m}^{-3}$  by using the fifth model layer concentrations as explained in Sect. 2.3. The fifth model level predictions will be used for KRV in all analyses that follow. Besides KRV, the Mediterranean KOP and OTL stations, as well as the rural ZAV site, are stations with comparatively high measured nighttime O<sub>3</sub> levels, which results in a low overall bias for all hourly O<sub>3</sub> values for these stations (from  $-2$  to  $-7 \mu\text{g m}^{-3}$ ). Namely, the WRF-Chem model cannot capture well the profound nighttime O<sub>3</sub> reductions (shown also by Žabkar et al., 2013; Im et al., 2015a), which contributes to the overall overprediction of hourly O<sub>3</sub> concentrations (from 10 to  $36 \mu\text{g m}^{-3}$ ) for stations with very low measured nighttime O<sub>3</sub> concentrations. For sites with the highest positive bias in 1 h O<sub>3</sub> concentrations (TRB, ZAG, HRA and ISK, with bias of 36, 31, 26 and  $32 \mu\text{g m}^{-3}$ , respectively), this can also be partly explained by the too high altitude of the stations in model orography (Table 1), since the mean O<sub>3</sub> concentration increases with height.



**Figure 4.** The 3-month average 1-day predictions of (a) hourly O<sub>3</sub>, (b) O<sub>3</sub> daily maximum, (c) hourly NO<sub>2</sub>, and (d) daily PM<sub>10</sub> concentrations for the first model layer, overlaid with measurements.

Looking at the O<sub>3</sub> daily maxima (Fig. 4b), the underpredictions occur at the alpine KRV station ( $-16 \mu\text{g m}^{-3}$  for the lowest model level shown in Fig. 4) and at three Mediterranean stations (OTL, NG, KOP; from  $-14$  to  $-11 \mu\text{g m}^{-3}$ ). For Mediterranean stations the underestimations of daily maxima are most probably due to inaccurate representation of coastal 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 a narrow valley of very complex terrain, which cannot be appropriately resolved in the current model topography, the model over-



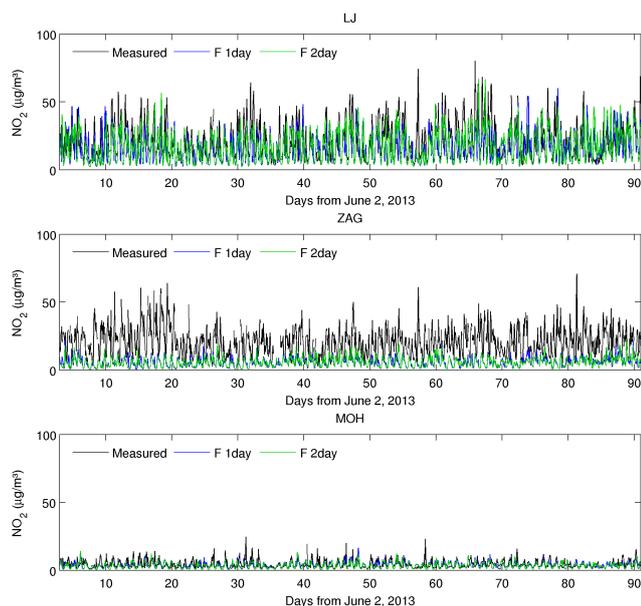
**Figure 5.** Time evolution of hourly ozone concentrations for 1-day (F 1 day) and 2-day (F 2 day) WRF-Chem predictions and measurements for some stations during the 3-month period.

predicts  $O_3$  daily maxima for  $14 \mu\text{g m}^{-3}$ . For other subalpine 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 error in predicted  $\text{NO}_2$  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  $\text{NO}_2$  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, in some cases, not resolved in model topography. Smoothed local emissions for these towns show significant underestimations of  $\text{NO}_2$  concentrations (e.g. ZAG in Fig. 6). In combination with poorly 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  $\text{NO}_2$  is either well simulated (e.g. MOH; Fig. 6) or slightly overpredicted due to increased emissions from adjacent urban areas (e.g. ZAD; Fig. 6). The overall agreement of hourly  $\text{NO}_2$  predictions with measurements was good for rural sites while urban sites experienced underpredictions, which were highest

for small cities, especially for NG (ME of  $-13 \mu\text{g m}^{-3}$ ) and ZAG (ME of  $-14 \mu\text{g m}^{-3}$ ).

Also interesting to discuss are the results for predicted  $\text{PM}_{10}$  concentrations (Table 3, Fig. 4d) showing slight over-prediction of daily  $\text{PM}_{10}$  levels at all stations, which is somewhat surprising due to the fact that nearly all current offline and online coupled chemical transport models show large systematic  $\text{PM}_{10}$  underestimations. For example, within the AQMEII exercise, where 17 modelling groups from Europe and North America were brought together, running eight operational online coupled air quality models over Europe and North America, the rural  $\text{PM}_{10}$  concentrations over Europe were underestimated by all models (model configurations) by up to 66 % while for the urban  $\text{PM}_{10}$  concentrations the underestimations were even much larger (up to 75 %) (Im et al., 2015b). The reason for slight overprediction of  $\text{PM}_{10}$  levels could be to some extent attributed to the high model spatial resolution used in our study. Furthermore, the CORR values for daily  $\text{PM}_{10}$  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  $\text{PM}_{10}$  concentrations during the analysed time period (no recorded  $\text{PM}_{10}$  exceeding), and partly due to the simulated  $\text{PM}_{10}$  overestimations during the heat wave events. These overpredictions contributed also to the overall positive bias of predicted  $\text{PM}_{10}$  levels. As shown in Fig. 7 for two monitoring sites, there was a significant  $\text{PM}_{10}$  overprediction simulated on 10 June (day 8 in Fig. 7) related to the pre-frontal advection of polluted air masses coming from the northwestern part of the domain D2 (coming from domain D1). The next significant  $\text{PM}_{10}$  overprediction occurred dur-

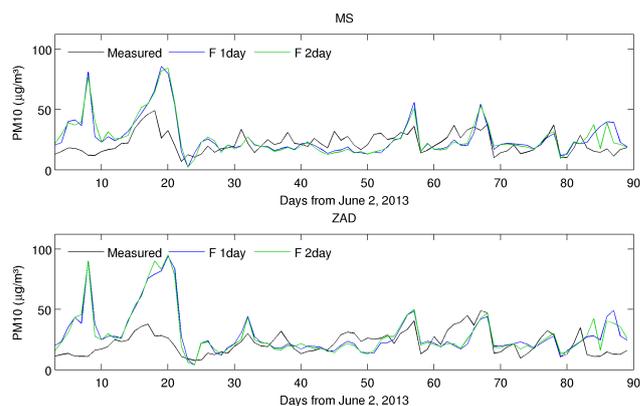


**Figure 6.** The same as Fig. 5 but for  $\text{NO}_2$  at the LJ, ZAG and MOH stations.

ing the first heat wave episode (17–22 June), when during the hot and low wind conditions (after 17 June) the  $\text{PM}_{10}$  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 overpredictions contributed to an overall positive bias in forecasted  $\text{PM}_{10}$  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  $\text{PM}_{10}$  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 and 57 and days 67 and 68 in Fig. 7); however, during these days the predicted  $\text{PM}_{10}$  levels were close to the measured  $\text{PM}_{10}$  concentrations.

### 3.3 Evaluation and comparison of different methods for $\text{O}_3$ daily maximum predictions

In this section we want to answer the question: how accurate is the 1 h  $\text{O}_3$  daily maximum WRF-Chem forecast in comparison to the statistical model prediction or to 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 along with their computational efficiency their main advantage (Zhang et al., 2012a). Among the strengths of the deterministic models are that they give prognostic time- and



**Figure 7.** The same as Fig. 5 but for daily  $\text{PM}_{10}$  concentrations at the MS and ZAD stations.

spatially resolved concentrations under typical and atypical scenarios and can give scientific insights into pollutant formation processes (Zhang et al., 2012a). Furthermore, they also allow for forecasts in locations which are not monitored due to their complete spatial coverage. In spite of simplified descriptions 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 similar to statistical forecasting tools (e.g. Manders et al., 2009). In addition to evaluation and comparison of  $\text{O}_3$  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 problems 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  $\text{O}_3$  daily maxima. Similarly, Table 4 shows these statistics for all data with different thresholds applied (only for WRF-Chem and persistence, because a statistical forecast is not available for all stations) and separately for different types of stations (subalpine urban, rural, Mediterranean urban) with an available statistical forecast. Looking at the ME, persistence gives results close to zero as long as no threshold is applied, while with a threshold of  $140 \mu\text{g m}^{-3}$  (Table 4) the ME of 1-day persistence ( $-10.2 \mu\text{g m}^{-3}$ ) is very close to the WRF-Chem model for 1-day predictions ( $-11.2 \mu\text{g m}^{-3}$ ), and for 2-day predictions WRF-Chem ( $-13.8 \mu\text{g m}^{-3}$ ) already beats persistence ( $-19.4 \mu\text{g m}^{-3}$ ). 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, KRV) with a lower or equal ME to WRF-Chem than for the statistical model, indicating the possible occurrence

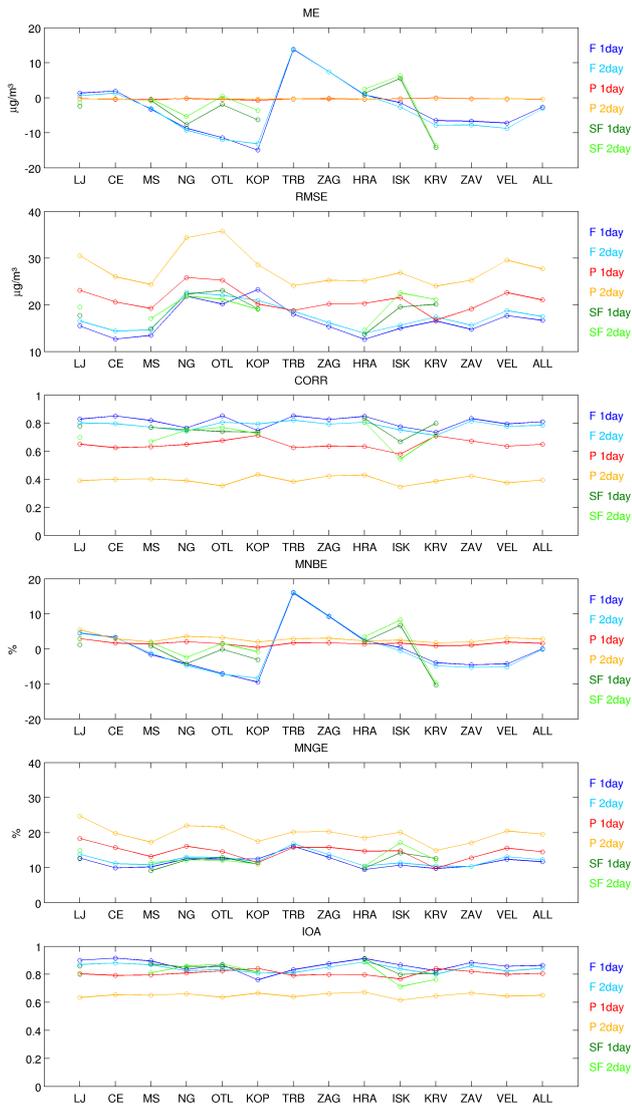
**Table 4.** Discrete evaluation of 1 h daily maximum ozone predictions (PER: persistence).

Stations	Threshold, no. cases	Forecast	Mean ( $\mu\text{g m}^{-3}$ )	ME ( $\mu\text{g m}^{-3}$ )	MAE ( $\mu\text{g m}^{-3}$ )	RMSE ( $\mu\text{g m}^{-3}$ )	CORR	MNBE (%)	MNGE (%)	IOA
All	> 0 1170	F 1 day	116.5	−2.6	13.3	16.7	0.81	−0.05	11.7	0.86
		F 2 day	116.6	−3.1	14.0	17.5	0.78	−0.1	12.3	0.84
		PER 1 day	119.5	−0.4	15.8	21.1	0.65	1.6	14.5	0.81
		PER 2 day	119.8	−0.4	21.7	27.7	0.39	2.8	19.6	0.65
	> 140 1102	F 1 day	144.1	−11.2	15.2	17.9	0.52	−6.8	9.5	0.57
		F 2 day	141.4	−13.8	16.5	19.4	0.42	−8.6	10.5	0.48
		PER 1 day	145.0	−10.2	15.6	19.6	0.41	−6.5	10.0	0.52
		PER 2 day	135.8	−19.4	24.76	29.2	0.31	−12.4	15.9	0.38
Subalpine urban with SF (LJ, HRA)	> 0 180	F 1 day	115.3	1.1	10.7	14.0	0.84	3.4	11.1	0.91
		F 2 day	115.4	0.8	12.0	15.2	0.80	3.5	12.2	0.88
	SF 1 day SF 2 day	PER 1 day	114.3	−0.3	16.7	21.7	0.64	2.2	16.5	0.80
		PER 2 day	114.6	−0.3	21.9	27.8	0.41	3.9	21.6	0.65
		SF 1 day	114.0	−0.5	11.9	15.7	0.81	1.6	11.2	0.88
		SF 2 day	116.2	0.6	13.4	17.1	0.75	3.2	12.7	0.84
Rural with SF (MS, ISK, KRV, OTL)	> 0 360	F 1 day	117.6	−5.6	13.3	16.3	0.80	−3.0	10.8	0.86
		F 2 day	117.4	−6.4	14.2	17.4	0.76	−3.4	11.4	0.84
		PER 1 day	123.6	−0.3	15.0	20.7	0.65	1.4	13.1	0.81
		PER 2 day	124.1	−0.4	21.6	27.8	0.37	2.4	18.5	0.64
	SF 1 day SF 2 day	SF 1 day	121.5	−2.9	15.0	19.4	0.74	−0.7	12.2	0.83
		SF 2 day	122.9	−1.8	15.8	20.5	0.67	0.5	13.2	0.79
Mediterranean urban with SF (KOP, NG)	> 0 179	F 1 day	123.5	−11.8	17.4	22.5	0.76	−6.9	12.5	0.80
		F 2 day	124.5	−11.2	17.2	21.8	0.77	−6.5	12.4	0.82
	PER 1 day PER 2 day SF 1 day SF 2 day	PER 1 day	135.9	−0.5	17.4	23.0	0.68	1.2	13.8	0.83
		PER 2 day	136.0	−0.2	25.2	31.5	0.41	2.8	19.7	0.66
		SF 1 day	129.3	−7.0	15.9	20.7	0.75	−3.6	11.6	0.83
		SF 2 day	131.6	−4.5	15.6	20.4	0.74	−1.6	11.6	0.84

of atypical conditions not resolved by the statistical model. Looking at MAE and RMSE for all stations, except those with the highest ME (TRB, KOP), the WRF-Chem outperforms the persistence already in the 1-day forecast. Among sites with available statistical forecasts there are only two (OTL, KOP) with WRF-Chem performing worse than the statistical forecast. CORR is one of the parameters that suggest how much the model is able to follow the true nature of processes regardless of 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 does the 1-day statistical forecast (CORR = 0.80) somewhat outperform WRF-Chem (0.74), and at NG and KOP the CORR values for WRF-Chem and the statistical model are very similar.

The Taylor diagrams in Fig. 9 show CORR together with the centred root-mean-square difference (RMSD) between model forecasts and observations, and the amplitude of their variations (standard deviation). The ideal model would have a correlation coefficient of 1 and a standard deviation equal to the observations, which means that it would be co-located with the black dot on the diagram. WRF-Chem gives higher CORR and lower RMSD for all types of stations, while the standard deviation of WRF-Chem O<sub>3</sub> daily maxima predictions is underestimated and lower than for other model fore-

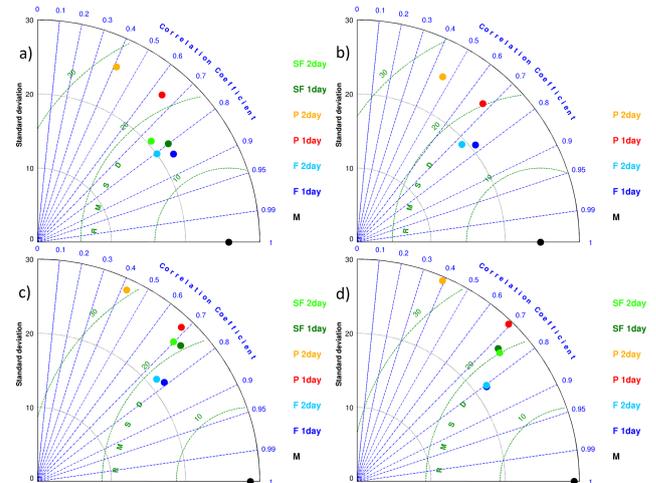
casts. The latter shows that the variability in WRF-Chem model predictions is not as large as that in observed values. The MNBE in Fig. 8 has very similar results to ME. For all forecasts except WRF-Chem for the TRB site (with MNBE of 16%), which is located in a narrow valley that is not resolved in the current model resolution, MNBE is below the  $\pm 10$ –15%, which is the US EPA (1991) recommended threshold for the models used for regulatory applications. For MNGE the US EPA recommendation below 30–35% for O<sub>3</sub> applications is met by all forecasts, even in the case of the 2-day persistence model. With the exception of the MS and KOP sites, the MNGE is lower for WRF-Chem than for the statistical forecast, while for the KOP and KRV sites 1-day persistence gives best results followed by the statistical forecast or WRF-Chem. The results are very similar for the IOA statistic with the range of 0–1, 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) or unresolved coastal processes (KOP), the WRF-Chem forecasts are more accurate than persistence. Here we recall that a high negative bias in the WRF-Chem forecast for the alpine KRV site, due to its low altitude



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

in model topography, was compensated by taking predictions from the fifth model level.

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, the contingency-table-based statistics are also an important metric of forecast performance. Table 5 summarizes the categorical evaluation results for three different thresholds (120, 140,  $160 \mu\text{g m}^{-3}$ ) of elevated  $O_3$  levels, which pose a greater risk to human health. Namely, it is important to take into account that results of categorical statistics are very sensitive to the threshold chosen, as well as to the overall pollution levels during the analysed months. Equitable threat score (ETS) measures the fraction of observed



**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) subalpine urban stations with SF (LJ, HRA), (b) subalpine urban stations without SF (CE, TRB, ZAG, VEL), (c) rural stations with SF (MS, ISK, KRV, OTL) and (d) Mediterranean urban stations (NG, KOP).

and/or correctly predicted events, adjusted for the frequency of hits that would be expected to occur by random chance. Although this score takes into account the climatology, it is not truly equitable. It ranges from  $-1/3$  to 1, where the minimum value depends on climatology (it is near 0 for rare events). Looking at Table 5, ETS shows equal skill for WRF-Chem and the statistical forecast, higher than persistence for the  $120 \mu\text{g m}^{-3}$  threshold (1-day and 2-day forecast). ETS decreases with increasing the threshold for both WRF-Chem and the statistical forecast, indicating the challenge that both models have to accurately predict the extremes. In the case of a  $140 \mu\text{g m}^{-3}$  threshold, WRF-Chem has the same ETS as persistence, higher than the statistical model for the 1-day forecast, while for the 2-day forecast WRF-Chem outperforms the statistical model, followed by persistence. In the case of a  $160 \mu\text{g m}^{-3}$  threshold, persistence has the highest ETS for a 1-day forecast, followed by the statistical model and WRF-Chem, while in the case of 2-day predictions the statistical model shows the highest skill and WRF-Chem the lowest. Another measure, the critical success index (CSI), is similar to ETS except that it does not take into account the climatology of the events and thus gives poorer scores for rarer events. It measures the percentage of cases that are correctly forecasted out of those either forecasted or observed, and ranges from 0 to 1 (1 indicating the perfect forecast). Similar to ETS, CSI gives higher scores for persistence in the case of the 1-day forecast for the two higher thresholds, while on the second day WRF-Chem or the statistical model already perform better. Bias ( $B$ ) determines whether the same fraction of events are both forecasted and observed. A tendency of the statistical model and of WRF-Chem to under-

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

Threshold	Forecast	ETS	CSI	B	FAR	POD	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
> 120	F 1 day	0.42	0.63	0.81	0.13	0.70	39	253	313	107
	F 2 day	0.39	0.61	0.79	0.14	0.68	41	245	303	115
	PER 1 day	0.31	0.59	0.99	0.25	0.74	91	267	249	93
	PER 2 day	0.17	0.49	1.00	0.34	0.65	123	235	209	124
	SF 1 day	0.42	0.67	1.02	0.21	0.81	67	257	243	61
	SF 2 day	0.38	0.65	1.03	0.23	0.80	77	264	225	66
> 140	F 1 day	0.40	0.50	0.64	0.15	0.551	19	111	490	92
	F 2 day	0.37	0.47	0.66	0.19	0.53	25	108	476	95
	PER 1 day	0.40	0.53	1.00	0.31	0.69	62	141	435	62
	PER 2 day	0.19	0.35	1.00	0.48	0.52	97	106	391	97
	SF 1 day	0.30	0.43	0.73	0.29	0.52	40	99	398	91
	SF 2 day	0.30	0.43	0.70	0.27	0.51	37	98	403	94
> 160	F 1 day	0.19	0.22	0.38	0.34	0.25	10	19	626	57
	F 2 day	0.17	0.20	0.34	0.35	0.22	9	17	619	59
	PER 1 day	0.40	0.45	1.00	0.38	0.62	29	47	595	29
	PER 2 day	0.22	0.28	1.00	0.56	0.43	43	33	572	43
	SF 1 day	0.23	0.27	0.49	0.35	0.32	13	24	539	52
	SF 2 day	0.25	0.29	0.63	0.41	0.37	19	27	540	46

predict O<sub>3</sub> threshold exceedances shows as a *B* below 1 for these two models. The false alarm ratio (FAR) that measures the percentage of forecasted high O<sub>3</sub> events that turn out to be false alarms gives the highest skill to WRF-Chem, followed by the 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.

It must be noted that in categorical evaluations systematic biases like those obtained with WRF-Chem for some stations (e.g. KOP) significantly impact the model performance. For example, if the KOP station was 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), were to be applied to correct the systematic biases, WRF-Chem forecasts might outperform the other two models even in categorical evaluations.

#### 4 Summary and conclusion

A high resolution modelling system based on an online 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 3 summer months. Since the selection of physical or chemical parameterization schemes influences and possibly changes the outcomes, we decided to apply schemes which are well documented and have previously been used

in other applications (e.g. AQMEII). Both 1-day and 2-day predictions of meteorological and air quality variables have been analysed. The focus has been on O<sub>3</sub> as it is the only pollutant with recorded exceedances of the legislation's limit values during the three heat wave events in June, July and August 2013. WRF-Chem daily O<sub>3</sub> maximum predictions have also been compared to the operational statistical model and persistence forecasts to answer the question of how skillful are the WRF-Chem model predictions in comparison to these two other models.

The 1-day and 2-day WRF-Chem PM<sub>10</sub> forecasts showed a very low bias. Exceptions were two events with significantly overpredicted PM<sub>10</sub> levels due to prefrontal advection of polluted air masses from neighbouring regions. Knowing that the majority of the current chemical transport models show large negative biases in simulated PM<sub>10</sub> concentrations, 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 the WRF-Chem NO<sub>2</sub> forecast with measurements was good for rural sites, while urban sites experienced model underpredictions 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 result in model underestimations of NO<sub>2</sub> 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 overpredictions of nighttime O<sub>3</sub> concentrations.

Evaluations of predicted 1 and 8 h daily O<sub>3</sub> maxima, which are in the case of this pollutant of the highest interest, show good WRF-Chem model performance. Nevertheless, there are also stations which experience high over- or underpredictions of O<sub>3</sub> daily maximum levels. For Mediterranean sites the underpredictions of the daily maxima are most probably due to inaccurate representation of coastal processes in the model, which are crucial for the PBL height evolution and accumulation of pollution in the near-ground air layers. For some subalpine stations the reason for the higher bias in O<sub>3</sub> 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 – which impacts how accurately a model simulates the local processes responsible for the level of local pollution. Comparisons of WRF-Chem O<sub>3</sub> daily maximum forecasts with persistence and with statistical model predic-

tions show that with respect to some statistical parameters the deterministic WRF-Chem forecast can outperform the other two for both 1- and 2-day predictions. For example, the correlation coefficient shows a higher skill for the WRF-Chem model, confirming the importance of complex processes as taken into account in an online 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 in its infancy for online coupled meteorology–chemistry models (Bocquet et al., 2015), could in future also be used as an efficient method for improving prediction of chemical concentration fields. For the WRF-Chem model, a technical note on the implementation of the aerosol assimilation and a guide for prospective users has been recently published by Pagowski et al. (2014).

**Appendix A: Statistical measures**

For  $i$ th observed ( $O_i$ ) and the corresponding modelled ( $M_i$ ) value of variable, discrete statistical measures are calculated as follows.

Mean error:

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

Mean absolute error:

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

Root mean square error:

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

Correlation coefficient:

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

Index of agreement:

$$IOA = 1 - \frac{\sum_{i=1}^N (M_i - O_i)^2}{\sum_{i=1}^N (|M_i - \bar{O}| + |O_i - \bar{O}|)^2}.$$

Mean normalized bias error:

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

Mean normalized gross error:

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

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

1. prediction is above, but observation is below the threshold;
2. prediction and observation are above the threshold;
3. prediction and observation are below the threshold;
4. prediction is below, but observation is above the threshold.

Categorical statistics are calculated as follows.

Equitable threat score:  $ETS = \frac{b - a_r}{a + b + d - a_r}$ , where  $a_r = \frac{(a+b)(b+d)}{a+b+c+d}$ .

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

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

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

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

*Acknowledgements.* Centre of Excellence for Space Sciences and Technologies SPACE-SI (OP13.1.1.2.02.0004) is in part financed by the European Union, European Regional Development Fund and the 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 the Slovenian Environmental Agency and Electroinstitute Milan Vidmar. The support through COST Action ES1004 EuMetChem is gratefully acknowledged.

Edited by: R. Sander

## 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 EURAD, *Meteorological Atmospheric Physics*, 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.: Chemical weather forecasting: a new concept of integrated modeling. *Adv. Sci. Res.*, 4, 23–27, 2010.
- Baklanov, A., Korsholm, U., Mahura, A., Petersen, C., and Gross, A.: Enviro-HIRLAM: on-line coupled modelling of urban meteorology and air pollution, *Adv. Sci. Res.*, 2, 41–46, 2008.
- 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., Moussiopoulos, 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, doi:10.5194/acp-14-317-2014, 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.*, 115, 604–619, doi:10.1016/j.atmosenv.2014.12.033, 2015.
- 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.*, 115, 620–629, 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.*, 15, 5325–5358, doi:10.5194/acp-15-5325-2015, 2015.
- 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, doi:10.5194/acp-9-945-2009, 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.
- 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.
- Curci, G., Hogrefe, C., Bianconi, R., Im, U., Balzarini, A., Baro, R., Brunner, D., Forkel, R., Giordano, L., Hirtl, M., Honzak, L., Jimenez-Guerrero, P., Knote, C., Langer, M., Makar, P. A., Pirovano, G., Perez, J. L., San Jose, R., Syrakov, D., Tuccella, P., Werhahn, J., Wolke, R., Zabkar, R., Zhang, J., and Galmarini, S.: Uncertainties of simulated aerosol optical properties induced by assumptions on aerosol physical and chemical properties: an AQMEII-2 perspective, *Atmos. Environ.*, 115, 541–552, doi:10.1016/j.atmosenv.2014.09.009, 2015.
- 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: Air Quality in Europe – 2012 Report, ISBN 978-92-9213-328-3, Luxembourg: Office for Official Publications of the European Union, 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 radia-

- tive 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 WRF/Chem, *Atmos. Environ.*, 53, 202–211, 2012.
- 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 phase 2 with respect to aerosol radiative feedbacks on meteorology and pollutant distribution, *Atmos. Environ.*, 115, 630–645, doi:10.1016/j.atmosenv.2014.10.056, 2015.
- Grell, G. A. 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. 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., Dudhia, J., and Stauffer, D.: A description of the fifth-generation Penn State/NCAR Mesoscale model (MM5), TN-398+STR, NCAR, Boulder, CO, 1994.
- 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.: On-line versus 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, doi:10.5194/acp-6-3181-2006, 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.
- 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.*, 115, 404–420, doi:10.1016/j.atmosenv.2014.09.042, 2015a.
- 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.*, 115, 404–420, doi:10.1016/j.atmosenv.2014.09.042, 2015b.
- 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.
- Kong, X., Forkel, R., Sokhi, R., Suppan, P., Baklanov, 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., Schllunzen, 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.*, 527–540, doi:10.1016/j.atmosenv.2014.09.020, 2015.
- 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<sub>10</sub> 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<sub>2.5</sub> 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., Turquety, 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.
- Pagowski, M., Liu, Z., Grell, G. A., Hu, M., Lin, H.-C., and Schwartz, C. S.: Implementation of aerosol assimilation in Grid-point Statistical Interpolation (v. 3.2) and WRF-Chem (v. 3.4.1), *Geosci. Model Dev.*, 7, 1621–1627, doi:10.5194/gmd-7-1621-2014, 2014.
- 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, 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. A. C., Kuenen, J., Zhang, J., Moran, M. D., and Makar, P. A.: Analysis of the emission inventories and model-ready emission datasets of Europe and North America for phase 2 of the AQMEII project, *Atmos. Environ.*, doi:10.1016/j.atmosenv.2014.10.061, online first, 2014.
- Saide, P. E., Carmichael, G. R., Spak, S. N., Gallardo, L., Osses, A. E., Mena-Carrasco, M. A., and Pagowski, M.: Forecasting urban PM<sub>10</sub> and PM<sub>2.5</sub> pollution episodes in very stable nocturnal conditions and complex terrain using WRF-Chem CO tracer model, *Atmos. Environ.*, 45, 2769–2780, 2011.
- San José, R., Pérez, J. L., Balzarini, A., Barò, R., Curci, G., Forkel, R., Galmarini, S., Grell, G., Hirtl, M., Honzak, L., Im, U., Jimenez-Guerrero, P., Langer, M., Pirovano, G., Tuccella, P., Werhahn, J., and Zabkar, R.: Sensitivity of feedback effects in CBMZ/MOSAIC chemical mechanism, *Atmos. Environ.*, 115, 646–656, doi:10.1016/j.atmosenv.2015.04.030, 2015.
- SEA: Upgrade of the system for monitoring air pollution, determining the causes of excessive burdening and analysis of the effects of improvement measures, Project presentation, Slovenian Environment Agency, available at: <http://www.arso.gov.si/en/>, last access: 12 June 2014.
- 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 DUSTRAN, *Atmos. Environ.*, 42, 1907–1921, 2008.
- 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-475thSTR, 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, 4339, doi:10.1029/2001JD001113, 2002.
- Tagigawa, 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.
- 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<sub>3</sub> 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.
- 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 north-

- ern France and southern England, *J. Geophys. Res.*, 108, 4021, doi:10.1029/2000JD000301, 2003.
- US EPA (US Environmental Protection Agency): Guidance for regulatory application of the Urban Airshed Model, EPA-450/4-91-013, July 1991, United States Environmental Protection Agency, Research Triangle Park, NC 27711, 1991.
- 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, Technical 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<sub>x</sub> and PM<sub>10</sub> in Athens and Helsinki, *Sci. Total Environ.*, 409, 1559–1571, 2011.
- 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.
- Wild, O., Zhu, X., and Prather, M. J.: Fast-J: Accurate Simulation of In- and Below-Cloud Photolysis in Tropospheric chemical Models, *J. Atmos. Chem.*, 37, 245–282, 2000.
- Wolff, G. T. and Liou, 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 southeastern 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., Morrison, H., Lee, Y.-N., Chapman, E. G., Spak, S. N., and Mena-Carrasco, M. A.: Assessing regional scale predictions of aerosols, marine stratocumulus, and their interactions during VOCALS-REx using WRF-Chem, *Atmos. Chem. Phys.*, 11, 11951–11975, doi:10.5194/acp-11-11951-2011, 2011.
- 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 U.S. 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.: Nadgradnja modela statističnega napovedovanja ozona s predhodnim razvrščanjem trajektorij v skupine, final report, available at: [http://www.arso.gov.si/zrak/kakovost%20zraka/poro%C4%8Dila%20in%20publikacije/poro%C4%8Dila%20o%20projektih/Porocilo\\_2011%20\\_napoved\\_ozona.pdf](http://www.arso.gov.si/zrak/kakovost%20zraka/poro%C4%8Dila%20in%20publikacije/poro%C4%8Dila%20o%20projektih/Porocilo_2011%20_napoved_ozona.pdf), 2011.
- Žabkar, R., Rakovec, J., and Gaberšek, S.: A trajectory analysis of summertime ozone pollution in Slovenia, *Geofizika*, 25, 179–202, 2008.
- 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.