



## Sensitivity Study of Cloud Cover and Ozone Modeling to Microphysics Parameterization

KINGA WALASZEK,<sup>1</sup> MACIEJ KRYZA,<sup>1</sup> MARIUSZ SZYMANOWSKI,<sup>2</sup> MAIGORZATA WERNER,<sup>1,3</sup> and HANNA OJRZYŃSKA<sup>1</sup>

**Abstract**—Cloud cover is a significant meteorological parameter influencing the amount of solar radiation reaching the ground surface, and therefore affecting the formation of photochemical pollutants, most of all tropospheric ozone (O<sub>3</sub>). Because cloud amount and type in meteorological models are resolved by microphysics schemes, adjusting this parameterization is a major factor determining the accuracy of the results. However, verification of cloud cover simulations based on surface data is difficult and yields significant errors. Current meteorological satellite programs provide many high-resolution cloud products, which can be used to verify numerical models. In this study, the Weather Research and Forecasting model (WRF) has been applied for the area of Poland for an episode of June 17th–July 4th, 2008, when high ground-level ozone concentrations were observed. Four simulations were performed, each with a different microphysics parameterization: Purdue Lin, Eta Ferrier, WRF Single-Moment 6-class, and Morrison Double-Moment scheme. The results were then evaluated based on cloud mask satellite images derived from SEVIRI data. Meteorological variables and O<sub>3</sub> concentrations were also evaluated. The results show that the simulation using Morrison Double-Moment microphysics provides the most and Purdue Lin the least accurate information on cloud cover and surface meteorological variables for the selected high ozone episode. Those two configurations were used for WRF-Chem runs, which showed significantly higher O<sub>3</sub> concentrations and better model-measurements agreement of the latter.

**Key words:** Cloud mask, meteorological modeling, ozone, WRF, Poland, model evaluation.

### 1. Introduction

Cloud cover plays important role in many atmospheric processes. Not only does it regulate Earth's

water cycle, but also its energy budget, and therefore radiative processes on the surface and atmospheric chemistry, and also interacts with aerosols in the atmosphere. Cloudiness affects ozone and other secondary pollutant formation by limiting incoming radiative fluxes to the surface layer. In meteorological and chemical transport models, e.g. WRF-Chem (GRELL *et al.* 2005; MADRONICH 1987; TIE *et al.* 2003; WILD *et al.* 2000), cloud cover information is passed on to photolysis schemes, thus influencing nitrogen dioxide (NO<sub>2</sub>) oxidation rates.

Cloud amount and cloud type are one of the most difficult meteorological parameters to predict. Cloud formation and dynamics depend on a wide variety of factors and processes, which are not accounted for in the model explicitly, simply because the atmospheric system is too complex and the current computational power is insufficient to resolve them. For these reasons, there is a need to apply approximations, which increase the uncertainty of cloud cover prediction (JOHNSON *et al.* 2015; VAN LIER-WALQUI *et al.* 2012). Since cloud microphysics interacts with many other elements of the weather system resolved by the model, those uncertainties are replicated and have an adverse effect on the overall forecast quality. In air quality modeling, it also affects estimation of pollutant concentrations, particularly ozone and other photochemical smog compounds, by regulating the amount of solar energy transferred to the surface.

There are many data types that cloud cover forecast verification can be based on (BRETHERTON *et al.* 1995). The most commonly used and longest data series that can be acquired are cloud fraction reports from ground-based weather stations (e.g. QIAN *et al.* 2012). Surface data are easily accessible in real time and widely used for verification of many other meteorological parameters, such as temperature, pressure or wind speed, but with cloud cover there are

<sup>1</sup> Department of Climatology and Atmosphere Protection, University of Wrocław, Wrocław, Poland. E-mail: kinga.walaszek@uni.wroc.pl; kinga.walaszek@uwr.edu.pl

<sup>2</sup> Department of Geoinformatics and Cartography, University of Wrocław, Wrocław, Poland.

<sup>3</sup> National Pollen and Aerobiology Research Unit, University of Worcester, Worcester, UK.

some setbacks. As the density of stations may be sufficient for other meteorological variables, cloudiness measuring network is very irregular and stations are located predominantly on land, so there is disproportion in data density over land and marine areas. There are also manual and automated stations, and the two different methods of gathering cloud fraction information may provide different outcomes (WMO 2008). Additionally, the number of synoptic stations worldwide has been decreasing (PETERSON and VOSE 1997; VOSE *et al.* 1992). Another issue is the frequency of the provided data—surface stations usually report at synoptic times, whereas regional meteorological models provide data at finer temporal resolution (1 h or less). Finally, there is more than one definition of cloud fraction and there are difficulties in transforming it into a variable that would be suitable for model verification.

One data source that solves the problem of irregular and sparse coverage of surface data are meteorological radars; however, they are designed to detect precipitation rather than cloud cover and are not commonly used for that purpose. Finally, there are satellite images, which not only have very large spatial extent, but also high spatial and temporal resolution and data are homogenous across the globe. Satellite data provide images in over a hundred spectral bands which allow the diagnosis of a variety of cloud products, from an unprocessed visible image to cloud mask, cloud top height, liquid water content, or brightness temperatures. Although these data are not always available in real time and go back only a few decades, it may serve a variety of applications related to model verification. There are two main types of satellites providing data for meteorological purposes: geostationary (e.g., the Meteosat series; FENSHOLT *et al.* 2011) and polar-orbiting (e.g. NASA's Terra and Aqua; KING *et al.* 2003). The main advantage of low Earth orbit satellites is their high spatial resolution, which may be even less than 1 km (down to 250 m at sub-satellite point in case of MODIS) and small distortions of the image. However, their orbit characteristics result in the data being available at irregular times, approximately 3–4 times a day. Geostationary satellites, on the other hand, which stay above a fixed point on the equator, have high temporal resolution (15 min for Meteosat

Second Generation), but spatial resolution is much lower than the polar-orbiting satellites. Meteosat MSG has 1 and 3 km resolution at sub-satellite point for High Resolution Visible (HRV) and infrared channels, respectively, and it decreases toward the edges of the image. The downside is that their coverage is limited by the satellite's field of view, so polar regions are either invisible or excluded because of large distortions.

Satellite imagery can be processed into a variety of products, and therefore enable various approaches to meteorological model verification (TUINDER *et al.* 2004). One of them is comparison of brightness temperatures (ZINGERLE and NURMI 2008; SÖHNE *et al.* 2008). It is usually not a parameter produced directly by meteorological models, but requires additional post-processing from other model output variables. Much more straightforward approach is to use cloud mask, which can be easily derived from cloud fractions at model levels (CROCKER and MITTERMAIER 2013). Satellite cloud mask is derived from multiple spectral channels, usually based on visible light and supported by infrared wavelengths, through a series of cloud detection tests. These data can then be compared with the modeled cloud mask to evaluate its results.

Meteorological model evaluation can also be based on various methods; one of them, referred to as categorical verification, uses grid-to-grid comparison, and another, object-based verification method, presents the features being verified as objects. In this study, we use both approaches to compare and quantify the differences between the cloud mask derived from the Weather Research and Forecasting (WRF) meteorological model simulation and satellite data. Four different microphysics parameterizations are tested for a selected period, favorable to formation of tropospheric ozone. Finally, for two parameterizations of microphysics, ozone concentrations are calculated with the WRF-Chem model, and the role of microphysics scheme on modeled O<sub>3</sub> is also described with the example of the episode of high ozone concentrations observed in central Europe.

There are two main aims of this study. The first aim is to evaluate the WRF model performance for cloud cover, using satellite data and objective

verification approach, and to test the model sensitivity to various microphysics schemes. The second aim is to examine the sensitivity of the WRF-Chem modeled ozone to the selected microphysics schemes. Simulation providing the highest model-measurements agreement will be used in further studies of tropospheric ozone in Poland.

## 2. Data and Methods

### 2.1. Study Area and Period

The analysis is performed for the area of Poland, which is characterized by transitional type of climate, with polar continental and polar maritime air masses being the two main drivers of weather conditions. This makes weather in Poland very changeable and difficult to predict. Episodes with stagnant anticyclone, providing many sunshine hours, high temperatures and low wind speeds, are not uncommon. This type of weather is very favorable for ground-level ozone formation, which is a major issue particularly for large cities and their peripheries. The EU Directive 2008/50/EC goal for 2010 has not been met and threshold values are still being exceeded (KRZYŚCIN *et al.* 2013; STASZEWSKI *et al.* 2012). Because one of the main aims of this study is to quantify the impact of selected microphysics parameterizations on air quality modeling, the test period is a high ozone episode of June 17th–July 4th, 2008. At that time, a vast anticyclone prevailed over Poland (Fig. 1), with low wind speed and high temperatures, which allowed photochemical smog to form in large cities and high concentrations of ground-level ozone were observed in Poland. The threshold value for 1-h average of  $180 \mu\text{g m}^{-3}$  set by aforementioned EU Directive was exceeded at four stations in Poland at least once.

### 2.2. The WRF Model

In this study, a multi-scale meteorological model, the Weather Research and Forecasting (WRF) version 3.5 (SKAMAROCK and KLEMP 2008) is used for the area of Poland. Simulations are performed for three one-way nested domains with grid size of  $45 \text{ km} \times 45 \text{ km}$  for the outermost,  $15 \text{ km} \times 15 \text{ km}$

for the intermediate, and  $5 \text{ km} \times 5 \text{ km}$  for the innermost domain, covering the area of interest. The model has 38 vertical layers with model top at 50 hPa. The domain configuration is presented in Fig. 2. Four simulations were run, each with a different microphysics parameterization—Purdue Lin (LIN *et al.* 1983), Eta Ferrier (ROGERS *et al.* 2005), WRF Single-Moment 6-class (HONG and LIM 2006), and Morrison 2-Moment (MORRISON *et al.* 2009), referred to as SIM1, SIM2, SIM3 and SIM4, respectively. Purdue Lin and Morrison schemes are currently the only two microphysics options that account for aerosol direct effects and are both widely used in WRF-Chem simulations (FORKEL *et al.* 2015; SAIDE *et al.* 2012; ZHANG *et al.* 2012). Eta Ferrier and WSM 6-class are also used in many applications, including model evaluation based on satellite data (GRASSO *et al.* 2014; OTKIN and GREENWALD 2008), studies of model sensitivity to microphysics for convective conditions (HONG *et al.* 2009) and heavy precipitation episodes (SEGELE *et al.* 2013). Other physics options remained the same for all model runs and include the Kain-Fritsch cumulus scheme, Yonsei University PBL scheme, unified Noah land-surface model, and RRTMG (IACONO *et al.* 2008) and RRTM (MLAWER *et al.* 1997) shortwave and longwave radiation, respectively. The model was initialized by the ERA-Interim data, available every 6 h with  $0.7^\circ \times 0.7^\circ$  horizontal resolution.

After evaluation of the cloud cover mask for the four WRF model simulations, the best and the worst configurations, in terms of the agreement with the satellite data, were used for the WRF-Chem model runs for the end of the study period—June 31st to July 4th. Details for the WRF and WRF-Chem model configurations are provided in Table 1. Because the differences between the two model runs are of interest here, the simple approach was applied, including restriction of the temporal variations in emissions from nature, while the TNO MACC II emissions (KUENEN *et al.* 2014) are assumed constant during the entire simulation. The chemical boundary conditions of trace gases consist of idealized, northern hemispheric, mid-latitude, clean environmental profiles based upon the results from the NOAA Aeronomy Lab Regional Oxidant Model (LIU *et al.* 1996). With all these simplifications it was

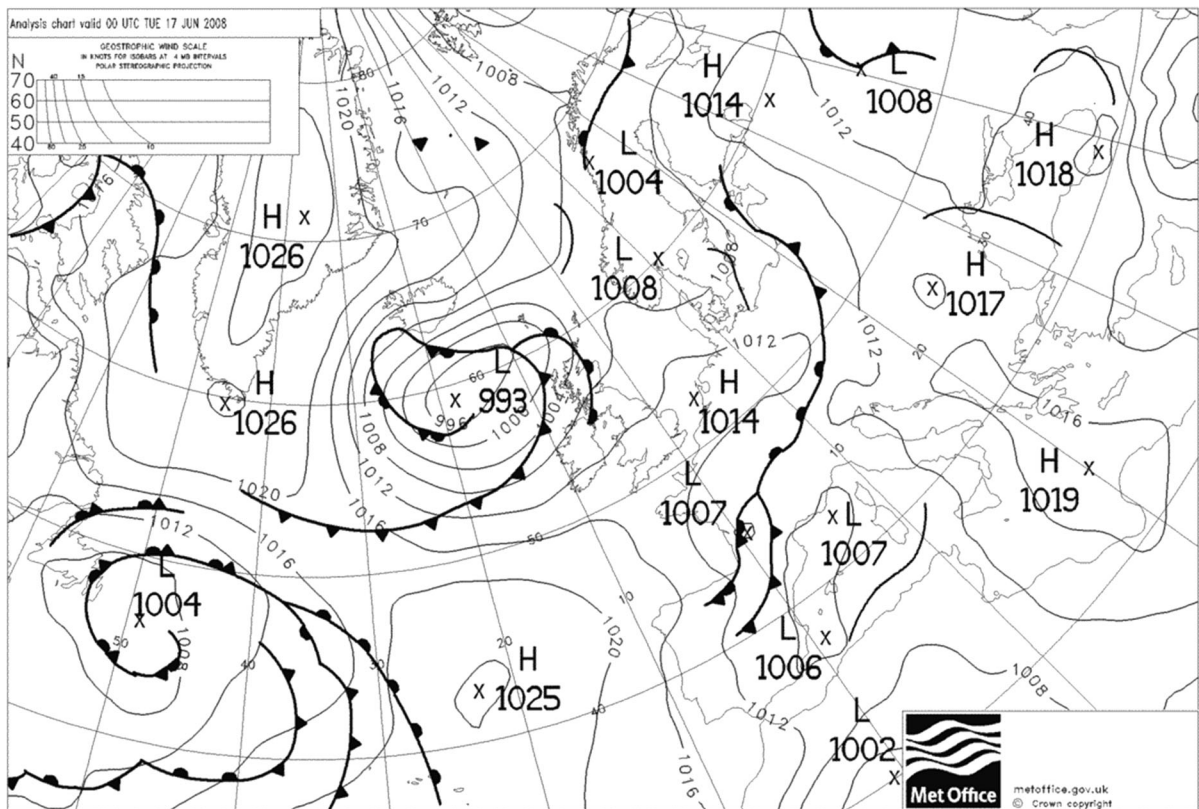


Figure 1

Synoptic situation for the first day of the study period (17.06.2008). Similar conditions prevailed throughout the whole period (17.06-04.07.2008)

computationally efficient to study the impact of microphysics parameterization on ozone concentrations, but it also influenced the chemistry model agreement with the measurements.

### 2.3. Measurements for Model Evaluation

The dataset used for evaluation of the model results is the cloud mask product, derived from the Meteosat Second Generation (MSG) SEVIRI instrument satellite imagery (DERRIEN and RAOUL 2010). This geostationary satellite offers high and constant temporal resolution, consistent with the WRF model output times (1 h), which is why it has been chosen over MODIS even despite its lower spatial resolution.

For generation of this product, a High Resolution Visible (HRV) channel and 11 infrared channels, particularly useful for nighttime hours and necessary

for distinction of clouds from e.g. snow cover, were used. Data are available every 15 min, but here the images at full hours were used to match the WRF model output. The final cloud mask product is obtained from Eumetsat, after a series of tests determining whether each grid cell is clear or cloudy. Cloud mask is a pessimistic field, which means that a grid cell can be classified as clear of clouds only if it passes every test. The full methodology of generation of the cloud mask product is described by DERRIEN and RAOUL (2010). HRV channel has a  $1 \text{ km} \times 1 \text{ km}$  resolution at sub-satellite point, whereas the remaining channels have  $3 \text{ km} \times 3 \text{ km}$  grid. The final product resolution is reduced to the lower grid resolution. Because of the curvature of the Earth, resolution decreases with distance to sub-satellite point and for Poland it drops to approximately 6–7 km. This is close to the spatial resolution of the



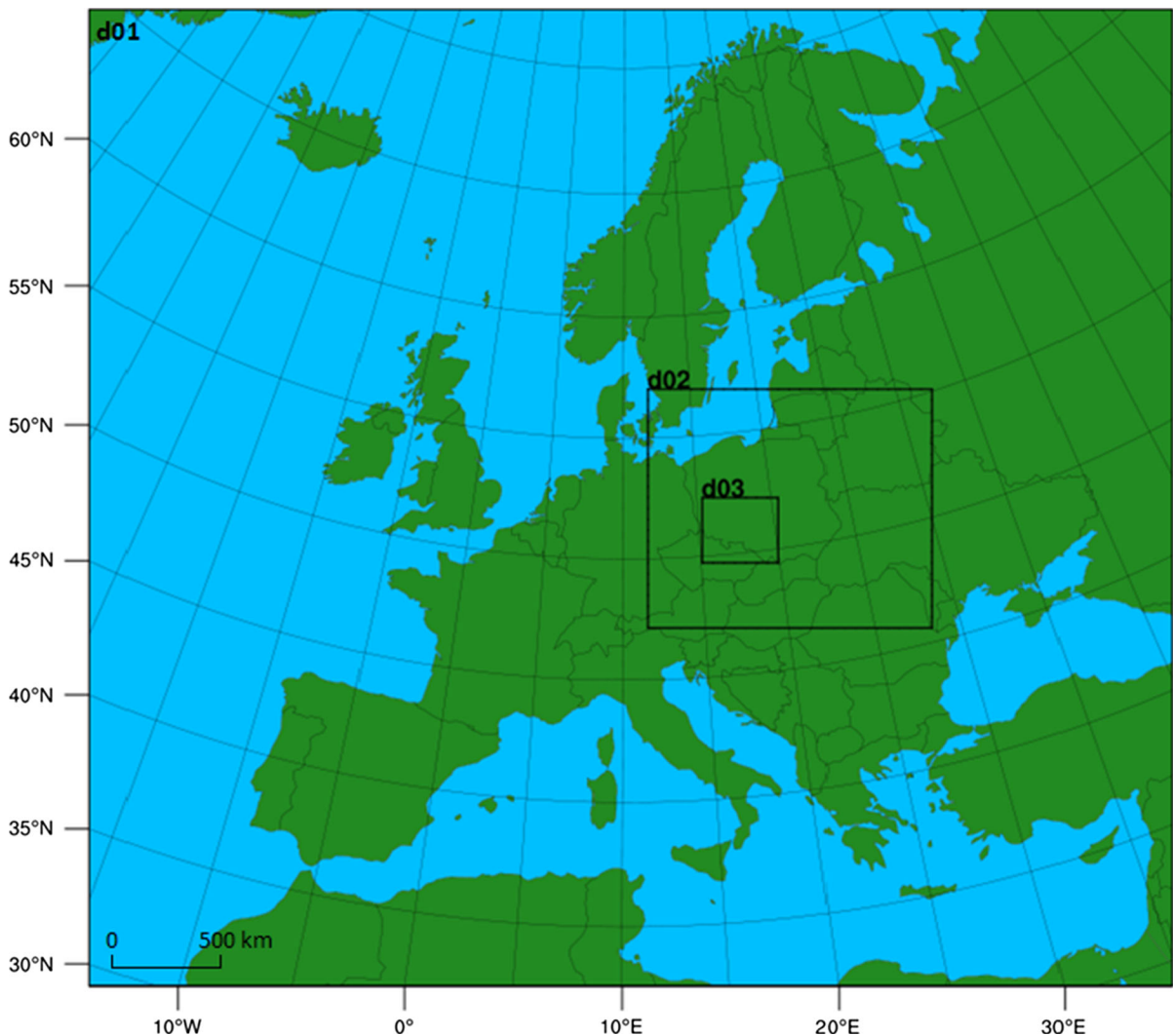


Figure 2

WRF model domain configuration. D01, d02 and d03 domains have spatial resolution of  $45 \text{ km} \times 45 \text{ km}$ ,  $15 \text{ km} \times 15 \text{ km}$ , and  $5 \text{ km} \times 5 \text{ km}$ , respectively. Results from domain d03 are analyzed

inner domain (d03) of the WRF model ( $5 \text{ km} \times 5 \text{ km}$ ) and the satellite data are resampled to the WRF grid for the spatial comparison.

For evaluation of other meteorological parameters, data from 57 synoptic stations in Poland were used. Ozone concentrations modeled with WRF-Chem were compared with hourly data derived from AirBase, from urban (Wrocław—Korzeniowskiego, WRK), suburban (Wrocław—Bartnicza, WRB), and regional background station (Śnieżka, SNI) in SW Poland.

#### 2.4. Evaluation of the Model Results Using the Cloud Cover Mask

There are multiple approaches that can be adopted to verification of cloud cover modeling. Here, two methods are used to evaluate the simulation results. First is categorical verification, which is probably the most widely used method. It is based on grid-to-grid comparison of measured and modeled values. Then, a contingency table is built, based on which various skill scores may be calculated. The main weakness of

Table 1

*Physics and chemistry parameterizations used in WRF and WRF-Chem model runs*

Parameter	
Cumulus scheme	Kain-Fritsch
Planetary boundary layer	Yonsei University
Land-surface model	NOAH LSM
Shortwave radiation	RRTMG
Longwave radiation	RRTM
Microphysics	Purdue Lin (SIM1) Eta Ferrier (SIM2) WSM6 (SIM3) Morrison 2-Moment (SIM4)
Gas-phase chemistry	RADM2
Aerosol model	MADE/SORGAM
Photolysis scheme	Fast-J
Wet deposition	Simplified parameterization for wet scavenging

this method is underestimation of model skill when analyzed phenomena are shifted in space. To overcome that weakness, an objective verification method can be used. This approach was initially developed for rainfall data and that is how it is commonly used, but it can be adopted to other applications, including cloudiness (CROCKER and MITTERMAIER 2013). In this approach, it is not grid cells, but objects, that are analyzed. An object is a continuous area that fulfills certain criterion, e.g. occurrence of precipitation or cloud cover. In this paper both approaches are used for evaluation of cloud cover simulations and the results are compared. A comparison of example maps, including percentage of area covered by clouds and number of cloud patches for satellite and WRF simulations, is also made.

### 2.5. Categorical Verification

Categorical verification involves a simple and intuitive approach that compares corresponding grid cells of observation and forecast. It can be applied to any phenomenon with values broken into categories; however, the most common use is for binary forecasts, e.g. occurrence of rainfall or cloud cover. In this case, a  $2 \times 2$  contingency table is built, presenting the count of grid cells falling into each of four categories: hits, misses, false alarms, and correct negatives (Table 2). A number of error measures can be calculated based on these data, four of which were

Table 2

*Contingency table used for categorical verification*

Forecast	Observed	
	Yes	No
Yes	a (hit)	b (false alarm)
No	c (miss)	d (correct negative)

Skill scores are calculated based on the number of grid cells falling into each category

selected for this study: Threat Score (TS), Probability of Detection (POD), False Alarm Ratio (FAR), and Frequency Bias Index (FBI; Table 3). Threat Score, also known as the Critical Success Index, measures the fraction of observed and forecast events that were correctly forecast (GILBERT 1884). The range of values is from 0 (no skill) to 1 (perfect score). It is sensitive to climatological frequency of the event and produces lower scores for rare events (SCHAEFER 1990). However, it allows to compare different model runs for the same domain and period of time, which is one of the aims of this study. Probability of Detection, also known as Hit Rate, measures the fraction of observed events that were correctly forecast. It also ranges from 0 (no skill) to 1 (all observed events were predicted). It is sensitive only to misses and hits and can be improved by overforecasting (JOLLIFFE and STEPHENSON 2003). Probability of Detection is usually used with False Alarm Ratio (probability of false detection), which measures the fraction of “yes” forecasts that were false alarms. The range of values is from 0 (no false alarms) to 1 (all “yes” forecasts were incorrect). Opposite to POD, it can be improved by underforecasting (WILKS 2006). Frequency Bias Index determines whether the model is under- or overforecasting the analyzed phenomenon. It ranges from 0 to infinity, with 1 as the perfect score. It should be noted that FBI is not a measure of model accuracy since it does not provide information on the magnitude of forecast errors (JOLLIFFE and STEPHENSON 2003). A summary of skill scores used in this study is provided in Table 3. Because in categorical verification only respective grid cells are compared, the so-called double penalty problem is an important issue. For example, when the forecast is even slightly shifted in space, the error may be counted twice—

once as a miss, and once as a false alarm. It may falsely reduce the score of model skill, as the event is, in fact, forecasted. In objective verification methods this issue is eliminated, because it is the objects, not individual grid cells, that are analyzed, and the distance of horizontal shift is also being accounted for as a part of the SAL measure.

## 2.6. Objective Verification

The Structure–Amplitude–Location (SAL) method was originally developed as a tool for verification of precipitation field forecasts (WERNLI *et al.* 2008). After simplification, the approach can be successfully applied also for other binary variables, such as cloud mask, which has been done previously, for example, by CROCKER and MITTERMAIER (2013) or ZINGERLE and NURMI (2008).

First, separate event fields need to be identified within a given domain. These objects are then compared to the respective observed fields, e.g. from Doppler radars or, in this case, satellite images. Afterward, geometric features of the objects, in this case—cloud cover ( $C_{\text{obs}}$ —cloud cover from satellite image,  $C_{\text{mod}}$ —from the model), are compared. The first parameter is structure ( $S$ ), which is defined as the average volume of objects, but because cloud mask field is uniform, it can be treated as a flat object and the structure component describes only its size (denoted as  $V$  in Eq. 1).  $S$  takes values from  $-2$  to  $2$ , where negative values mean that model underestimates average size of objects and positive values mean overestimation

$$S = \frac{V(C_{\text{mod}}) - V(C_{\text{obs}})}{0.5[V(C_{\text{mod}}) + V(C_{\text{obs}})]}. \quad (1)$$

The second component of the SAL measure is amplitude ( $A$ ), which calculates the domain-average cloud field. It can be interpreted as the degree to which the model is over- or underestimating the total amount of clouds in the domain. For data with continuous values, the size is understood as the total volume of objects, whereas for binary data it is the total area ( $D$  in Eq. 2).  $A$  takes values from  $-2$  to  $2$  as well, with negative values meaning underestimation of total cloud amount within the domain and positive values—overestimation. Please note that structure and amplitude components of SAL are nonlinear, for example  $S = -1$  means that model underestimates average cloud size three times, and similar statement is true for amplitude. In general,  $S$  and  $A$  values depend on observed total cloud amount and cloud size and therefore cannot be directly compared to studies for another region or episode. However, it allows to assess performance of different models for a fixed domain

$$A = \frac{D(C_{\text{mod}}) - D(C_{\text{obs}})}{0.5[D(C_{\text{mod}}) + D(C_{\text{obs}})]}. \quad (2)$$

For the location component, two parts of the measure are calculated: one parameter ( $L_1$ ) determines the distance between the observed and predicted domain-wide center of mass ( $X$  in Eq. 3), normalized by the use of the diagonal length of the domain ( $d$  in Eqs. 3 and 4). On the other hand, the  $L_2$  parameter measures the observed and predicted average distance between the objects center of mass and the domain

Table 3

*Skill scores calculated for cloud cover based on contingency table (above; a hit, b false alarm, c miss, d correct negative)*

Name	Definition	Interpretation	Range of values
Threat score or Critical Success Index	$TS = \frac{a}{a + b + c}$	Fraction of observed and/or forecast events that were correctly predicted	0–1 1: perfect score
Probability of detection or hit rate	$POD = \frac{a}{a + c}$	Fraction of observed events that were correctly forecast	0–1 1: perfect score
False alarm ratio	$FAR = \frac{b}{a + b}$	Fraction of forecast events that were false alarms	0–1 0: perfect score
Frequency Bias Index	$FBI = \frac{a + b}{a + c}$	Ratio of the frequency of forecast events to the frequency of observed events	0– $\infty$ 1: perfect score

Table 4  
*Error statistics calculated for temperature, relative humidity, wind speed and ozone concentrations*

Name	Definition	Range of values
Mean error	$ME = \frac{\sum (\text{sim-obs})}{n}$	$-\infty$ to $\infty$
Mean absolute error	$MAE = \frac{\sum  \text{sim-obs} }{n}$	0: perfect score 0 to $\infty$
Index of agreement	$IOA = \frac{\sum (\text{sim-obs})^2}{\sum ( \text{sim-obs}  +  \text{obs-obs} )}$	0: perfect score 0 to 1 1: perfect score

overall center of mass. For binary data, center of mass is simply the geometrical center (denoted as  $r$ ; Eq. 4). The  $L$  component is defined as the sum of  $L_1$  and  $L_2$  (WERNLI *et al.* 2008)

$$L_1 = \frac{|X(C_{\text{mod}}) - X(C_{\text{obs}})|}{d} \quad (3)$$

$$L_2 = 2 \left[ \frac{|r(C_{\text{mod}}) - r(C_{\text{obs}})|}{d} \right] \quad (4)$$

The results of object-based verification are then presented on SAL diagrams, which show the values of all components and relationship between them (Fig. 5). Because the values of  $S$  and  $A$  components have the same range of values, they are represented on the axes, whereas the value of  $L$  is represented by the color of the data points. Dotted lines denote mean values of  $S$  and  $A$  and the sides of the rectangle are the first and third quartiles. These elements facilitate interpretation of the diagram, as the closer the dotted lines are to the center of the diagram and the smaller the rectangle, the more accurate is the forecast.

### 2.7. Evaluation of Meteorological Variables and Ozone Concentration

Besides cloud cover, the impact of the microphysics scheme on three surface meteorological variables was analyzed: air temperature and relative humidity at 2 m, and wind speed at 10 m. Three statistical metrics were calculated for each parameter for all model runs based on observational data from synoptic stations: Mean Error (ME), Mean Absolute Error (MAE), and Index of Agreement (IOA). Mean Error was selected to show how much the model under- or overestimates measured values, whereas Mean Absolute Error shows the absolute value of

errors. Index of Agreement is a standardized measure of the overall model-measurement agreement (WILLMOTT 1981). The formulas and value range of the above statistics are presented in Table 4.

After the analysis of meteorological model simulations, the best and the worst simulations were selected for the WRF-Chem model runs. For these simulations, spatial distribution of mean  $O_3$  concentration and the differences between model runs are presented. For three air quality measurement stations representing different environments, temporal variability of measured and modeled 1-h average concentrations were compared.

## 3. Results and Discussion

### 3.1. Cloud Cover

Figures 3 and 4 present example cloud mask images from the satellite product and four WRF simulations for morning (9 AM UTC, 11 AM local time) and afternoon (3 PM UTC, 5 PM local time) hours. In both cases, the locations of modeled cloudy areas correspond to the satellite-derived image, but total cloud amount in the domain is smaller (39 % for SIM4 compared to 59 % on satellite image), particularly in the afternoon. Differences between simulations are much less pronounced than those between the model and satellite product, which suggest that the selection of the microphysics scheme has limited impact on the cloud mask results. The modeled clouds form patches of small cells rather than one vast cloudy area, like the satellite image—every simulation gives at least twice as many cloud cells as satellite. There are two reasons for this. It is related to the fact that cloud mask product



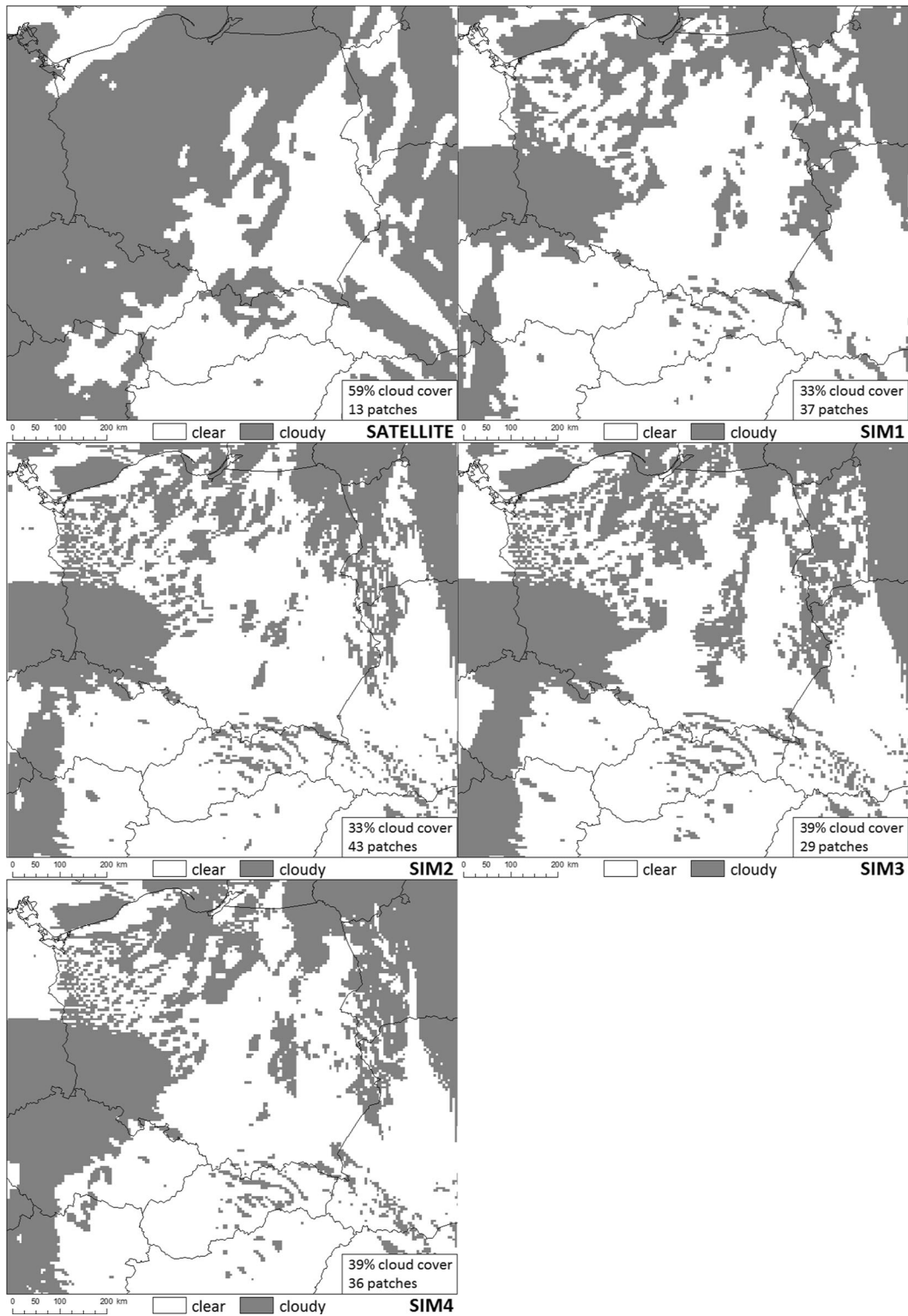


Figure 3

An example of MSG satellite cloud mask product and WRF simulation results for 20 June 2008, 9 AM UTC

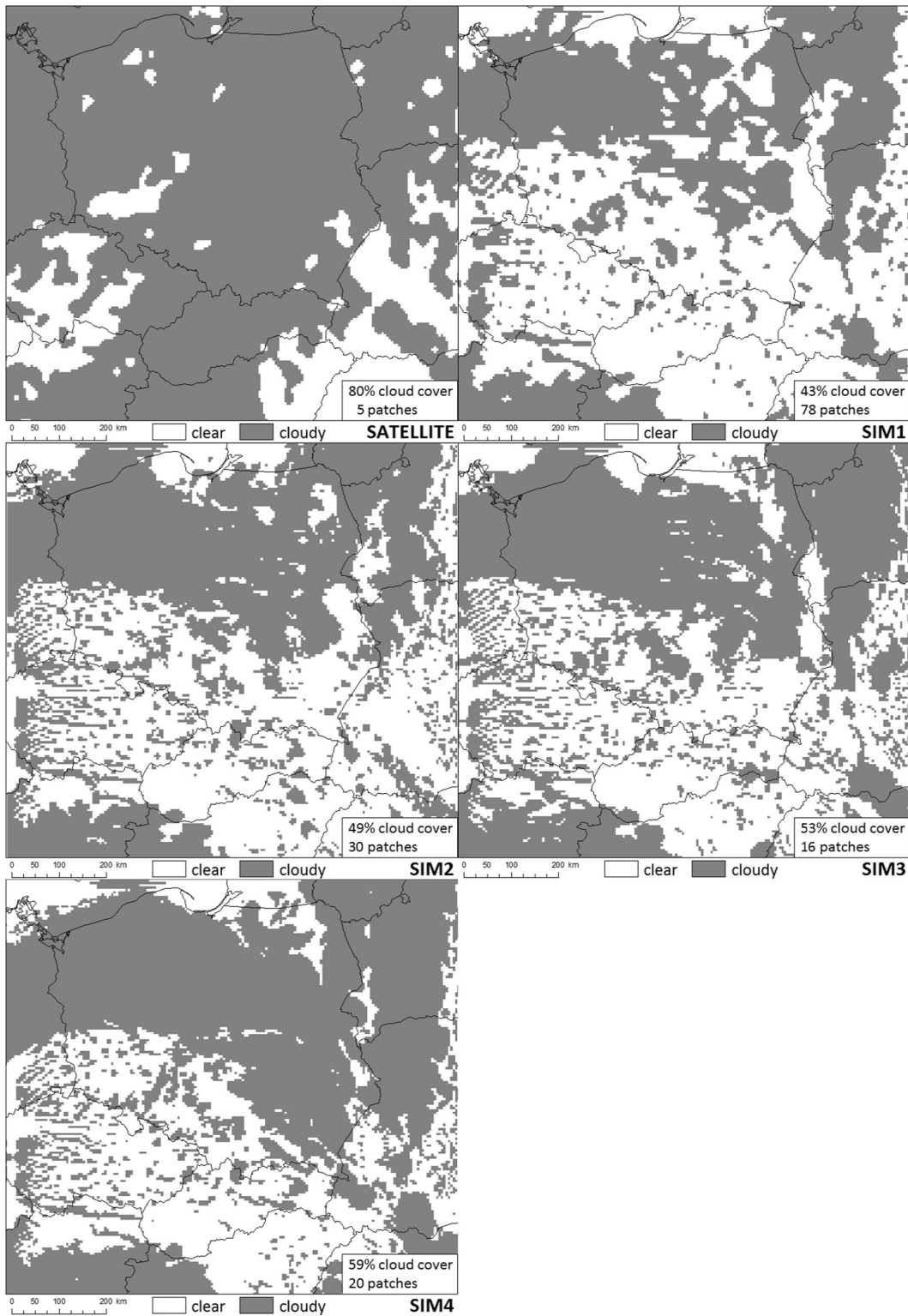


Figure 4  
An example of MSG satellite cloud mask product and WRF simulation results for 20 June 2008, 3 PM UTC

generated from Meteosat images has coarser spatial resolution over Poland than the WRF model domains. After resampling to the spatial resolution of the WRF model, the number of cells marked as cloudy might increase. It is also possibly the main reason why for all model simulations a set of orographic clouds in the Carpathians is visible in the morning, which is shown as a single cloud patch in the satellite image. The second reason is that the entire WRF model grid cell has to reach saturation level before it is marked as cloudy. Considering summer convective condition this might be unlikely, therefore the WRF model provides lower number of grid cells with clouds, if compared with satellite data. This is also supported by the larger differences between the WRF and satellite cloud mask for the afternoon hours, if compared to morning (Figs. 3, 4).

Considering the differences between simulations, they are much smaller than differences between any of the simulations and the satellite cloud mask. SIM4 produces the largest cloud amount and SIM1 the smallest. Another noticeable thing is a distinct quantitative difference between SIM4 and other simulations—cloud cells are larger and cover more area, which is supported by the value of FBI (Table 5).

### 3.2. The SAL Method

The results of the simulations evaluated with the SAL method are shown in Fig. 5. It shows that for all simulations both cloud size and total cloud amount, represented by  $S$  and  $A$  components, are underestimated by the model, as the majority of data points lie in lower left quadrants of the plots. The main cause is the fact that WRF does not account for subgrid-scale cumulus clouds in the cloud fraction output, which leads to underestimation of modeled cloud cover, as

the whole grid cell needs to be saturated to produce cloud. Satellite cloud mask, on the other hand, is a pessimistic field, which means that only the cells which pass all tests can be flagged as cloud-free, which increases the discrepancy between modeled and satellite-derived cloud cover. The best  $S$  and  $A$  values are for SIM4, as the rectangle limited by  $S$  and  $A$  first and third quartiles is small and located closest to the center of the diagram. It may be explained by the fact that Morrison Double-Moment is the most sophisticated of the selected microphysics options and the only double-moment scheme. SIM3 and SIM2 present similar performance, whereas SIM1 underestimates both cloud amount and size the most. For all simulations, the points with  $S$  and  $A$  components close to zero have generally also small  $L$  values; however, there are some exceptions—particularly in the lower right quadrant. There is a high density of data points with large location component and at the same time structure is significantly underestimated and amplitude is close to the median value. There are very few points with overestimated cloud amount and size, and most of them have small to moderate  $L$  component value. There are almost no data points with underestimated amount and overestimated cloud size at the same time. This is expected because grid cells on the edges of clouds are less likely to reach saturation, which causes decrease in both cloud size and total cloud cover. A study conducted by CROCKER and MITTERMAIER (2013) for the United Kingdom shows that UK4 and UKV models tend to overestimate cloud cover.

### 3.3. Categorical Verification

Table 5 shows four categorical verification measures. The results present poor model performance,

Table 5

*Categorical verification measures calculated for all WRF runs (TS Threat score, POD Probability of Detection, FAR false alarm ratio, FBI Frequency Bias Index)*

	TS	POD	FAR	FBI
SIM1	0.39	0.45	0.27	0.71
SIM2	0.40	0.47	0.28	0.74
SIM3	0.42	0.50	0.29	0.82
SIM4	0.42	0.53	0.30	0.94

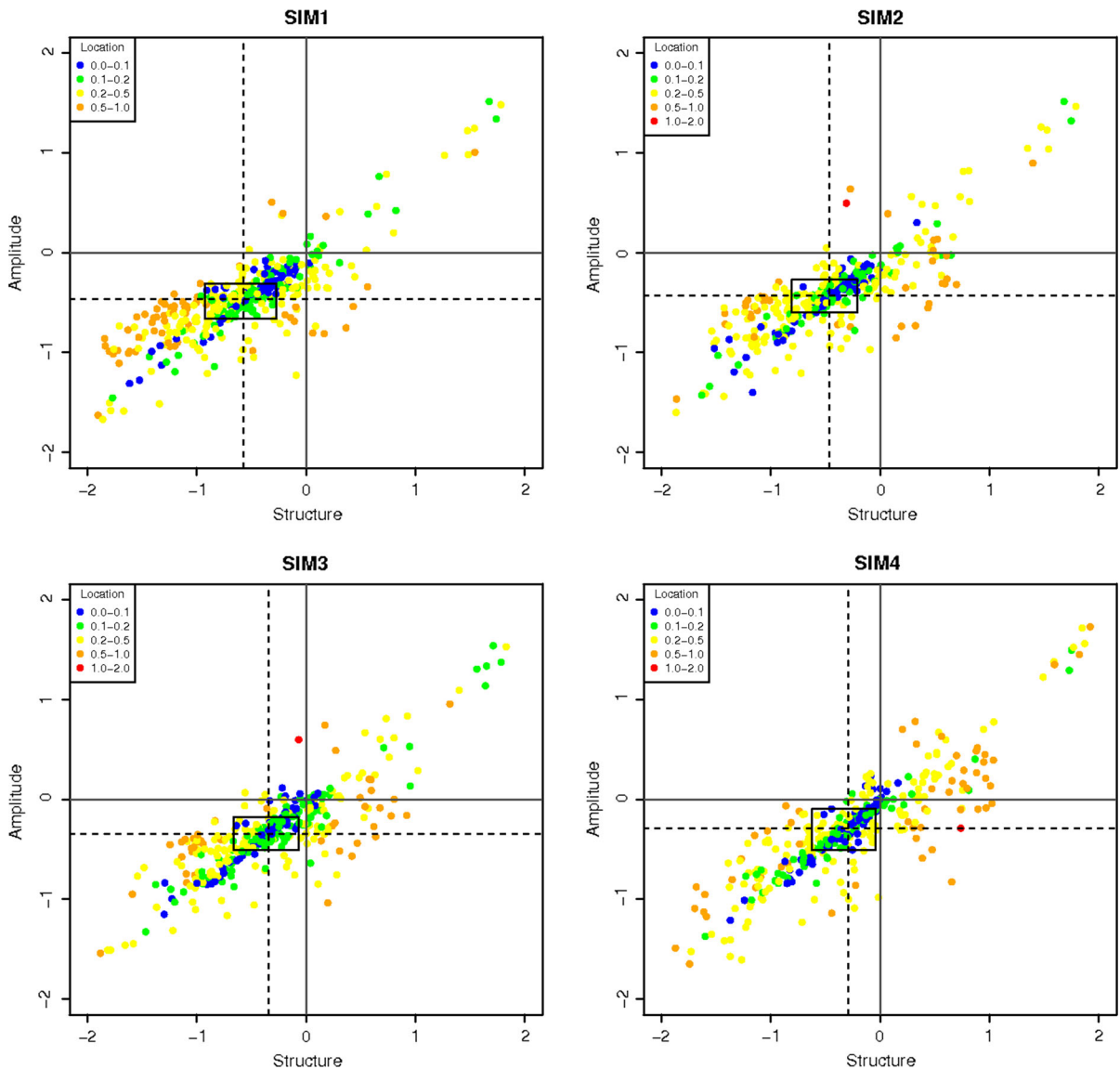


Figure 5

SAL diagrams for all WRF simulations, with Structure (Eq. 1) and Amplitude (Eq. 2) values are given by the position of the point on the diagram and Location (Eqs. 3 and 4) value is given by its color. Dotted lines indicate median values and the rectangles enclose points within 1st and 3rd quartiles of Structure and Amplitude

with TS not exceeding 0.5. As this measure is sensitive to both misses and false alarms, it is essential to examine which element had the most influence on the results. The values of POD are very low, which indicates large fraction of missed events, therefore it can be concluded that forecasting clear sky when cloud cover is present is a major issue, which is caused by subgrid-scale cloudiness not being resolved by microphysics schemes in WRF. It also

shows that nearly half of the observed cloudy grid points are not resolved by the model. SIM4 simulation gives the best result in terms of TS and FBI, which is very close to unity, but False Alarm Ratio is also higher here than for the remaining simulations. It suggests that the reason of high threat score is that this model run forecasts more cloud than other simulations, but otherwise it is not necessarily attributed to model skill.

The values of POD averaged for each hour of day are presented in Fig. 6. All simulations present a similar trend, with the lowest value shortly after sunrise and highest for late afternoon (above 0.6 for SIM4). It indicates that the WRF model is more skilled in resolving afternoon than morning cloudiness. However, one has to be careful in drawing direct conclusions, since most skill scores depend on total observed or modeled cloud amount. SIM4 has the highest values of all simulations for all but 1 h and the differences are the largest for 17:00–19:00 (up to 0.04). The results are much poorer for SIM1 and SIM2, where this parameter falls below 0.4. However, a better POD score is usually associated with larger FAR, because POD may be improved by

overforecasting, as the number of hits (to which POD is sensitive) is larger, but the number of false alarms, to which FAR is sensitive, also rises (Fig. 7).

### 3.4. Meteorological Variables

Modeled temperature, relative humidity and wind speed are evaluated based on hourly data from synoptic stations located in Poland. The results are summarized in Table 6. Temperature and humidity are overestimated and wind speed is underestimated by all model runs, which is shown by Mean Error. The differences in Mean Absolute Error between simulations are also small. Model-measurements agreement of wind speed, represented by IOA, shows

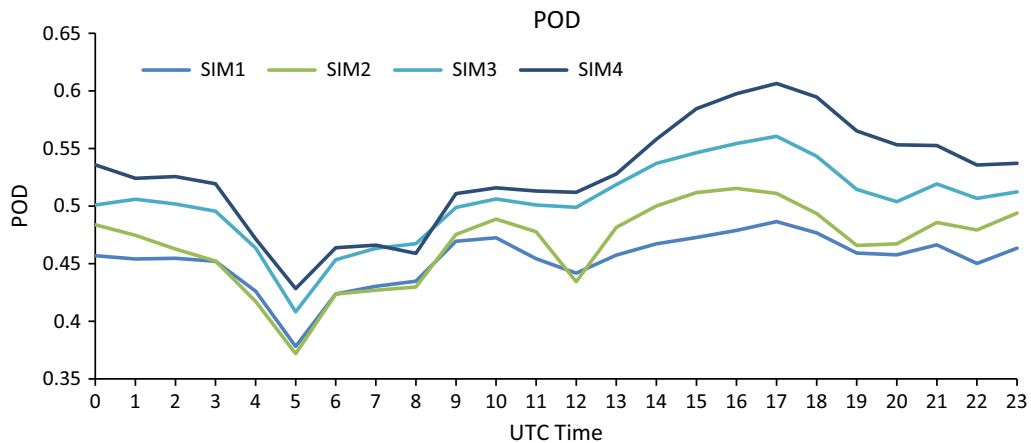


Figure 6  
Hourly values of Probability of Detection (POD) averaged for the study period for each of the four simulations

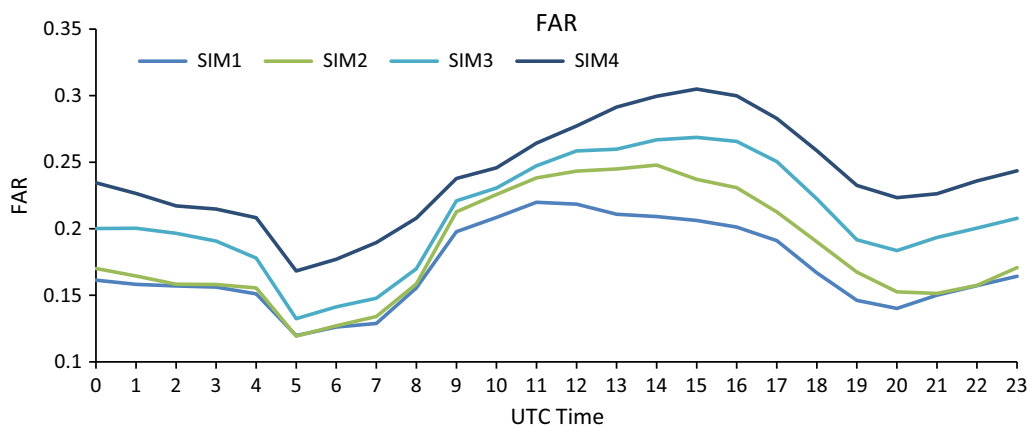


Figure 7  
Hourly values of False Alarm Ratio (FAR) averaged for the study period for each of the four simulations



Table 6

Error statistics calculated for WRF simulations of three meteorological variables: temperature (T2), relative humidity (RH2), and wind speed (WSPD)

	T2			RH2			WSPD		
	ME (°C)	MAE (°C)	IOA (-)	ME (%)	MAE (%)	IOA (-)	ME (ms <sup>-1</sup> )	MAE (ms <sup>-1</sup> )	IOA (-)
SIM1	0.135	1.694	0.947	1.998	10.438	0.869	-0.072	1.238	0.798
SIM2	<i>0.098</i>	<i>1.672</i>	<i>0.949</i>	1.664	<i>10.097</i>	<i>0.878</i>	-0.097	1.241	<i>0.979</i>
SIM3	0.146	1.692	0.947	2.209	10.491	0.868	-0.083	<i>1.233</i>	0.801
SIM4	0.266	1.695	0.947	<i>0.746</i>	10.323	0.871	-0.082	1.234	0.800

The numbers in italics denote simulations with lowest values of each statistic

a significant advantage of SIM2 over the other simulations. Generally, there are small differences between the WRF model running with different microphysics schemes, with SIM2 showing slightly better performance. Because the study period is dominated by stagnant anticyclone with low wind speed and no precipitation, the differences between model runs with different microphysics schemes are not pronounced. However, studies conducted for longer and more diverse periods show that Morrison Double-Moment scheme provides the most consistency with observations for meteorological variables and aerosol concentrations (BARÓ *et al.* 2015).

### 3.5. Ozone Concentrations

Figure 8 presents average 1-h ozone concentrations in the innermost WRF model domain for SIM1 and Fig. 9 for SIM4. Both maps show similar spatial pattern, with O<sub>3</sub> increasing toward the south-west, reaching 90 µg m<sup>-3</sup> in the Czech Republic. The concentrations modeled with SIM4 are generally higher than SIM1, particularly for areas with higher O<sub>3</sub> levels and over the Baltic Sea in the north, where the differences between SIM4 and SIM1 exceed 7 µg m<sup>-3</sup> (Fig. 10). The differences between model runs are confirmed by the time series charts in Fig. 11, which show that SIM4 produces higher O<sub>3</sub> levels for all sites. Better performance of the simulation running with Morrison microphysics may be a result of the fact that it is a double-moment scheme that takes into account aerosol direct effects. However, both simulations capture the daily ozone cycle in the urban environment, although the amplitude of changes is much lower than observed. This

could be linked to constant temporal emission profile applied, since it does not account for diurnal or weekly changes in anthropogenic emission, mainly from transport (e.g. morning and afternoon peaks in NO<sub>x</sub> emission). Another possible source of errors may be inadequate chemistry scheme, underestimating the rate of O<sub>3</sub> formation and destruction processes. Both of these reasons may be verified by changing emission input data or applying a different chemical mechanism. Model errors are on similar level to the study by FORKEL *et al.* (2015); however, it should be noted that the study period here is shorter. For rural station O<sub>3</sub> concentration is underestimated for the entire period by both simulations, which may be explained by underestimated background concentrations (default values used with WRF-Chem).

## 4. Conclusions

Although categorical verification of cloud cover forecast provides valuable information about model performance, it may falsely understate model skill in cases when clouds are even slightly dislocated. However, this type of verification can capture the model tendency to underestimate total cloud amount within the domain and enables the identification of possible sources of uncertainties. Objective verification methods may serve as a supplement to categorical approach, as it provides additional information on the structure of model-measurements discrepancies. The objective approach provides both direct information on whether the total cloudiness in the domain is over- or underestimated and to what extent, and also brings more detailed information on

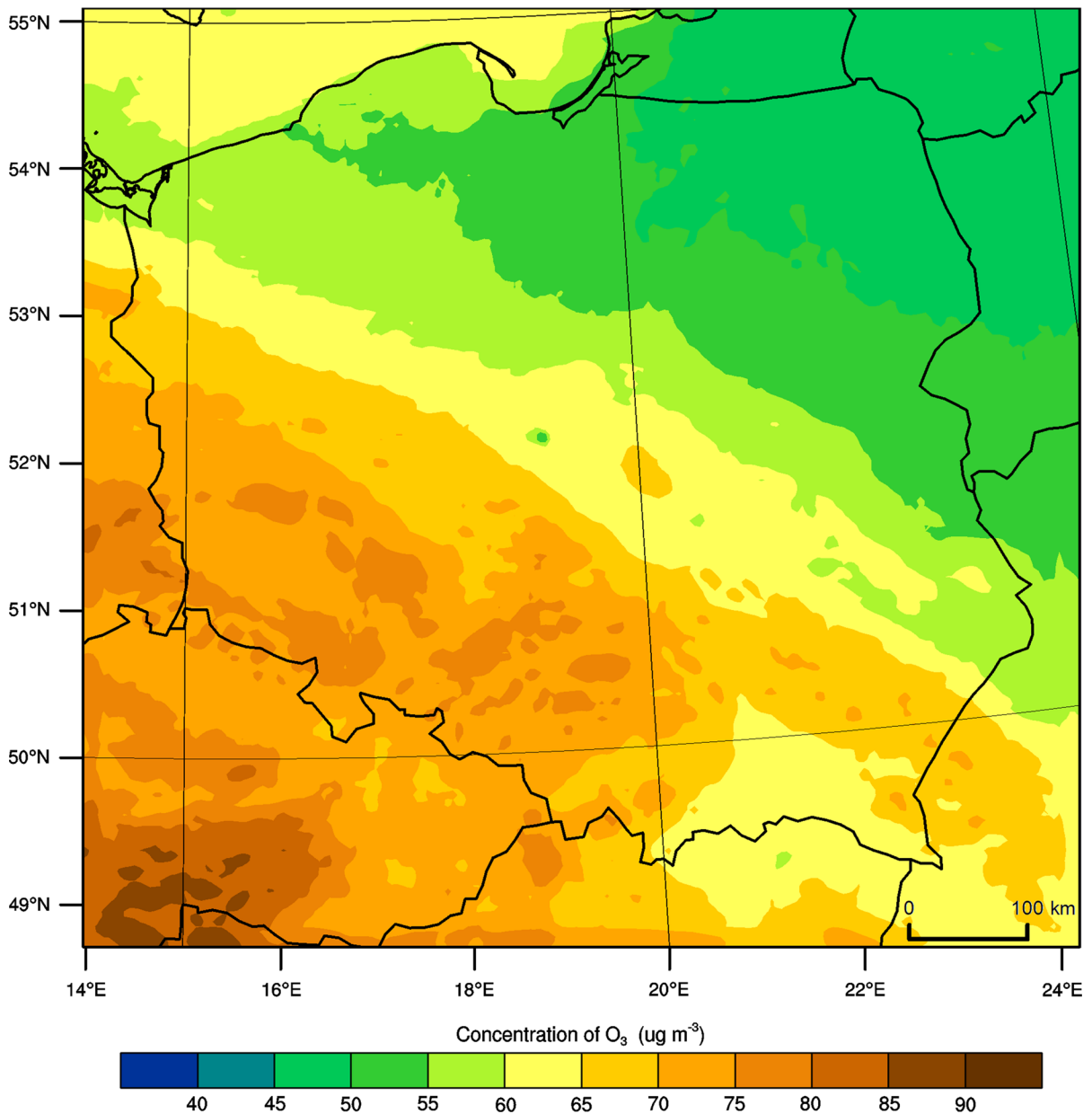


Figure 8  
Mean O<sub>3</sub> concentration for the episode of 30 June–4 July 2008 (SIM1, Purdue Lin)

the size and location of modeled cloud patches compared to the observed ones. By analyzing objects (i.e. cloudy areas) instead of individual grid points it also eliminates the double penalty problem, which becomes a large issue with high spatial resolution of meteorological models; therefore, model performance is not underestimated, as in the case of categorical verification method.

Both methods are consistent with the conclusion that all WRF simulations underestimate the amount of cloud cover. This may have further consequences on e.g. overestimation of the summer air temperature by the WRF model which was shown by KRYZA *et al.* (2015, this issue) for Central and Eastern Europe. One important factor is that satellite cloud mask is a pessimistic field, meaning that only a grid point that

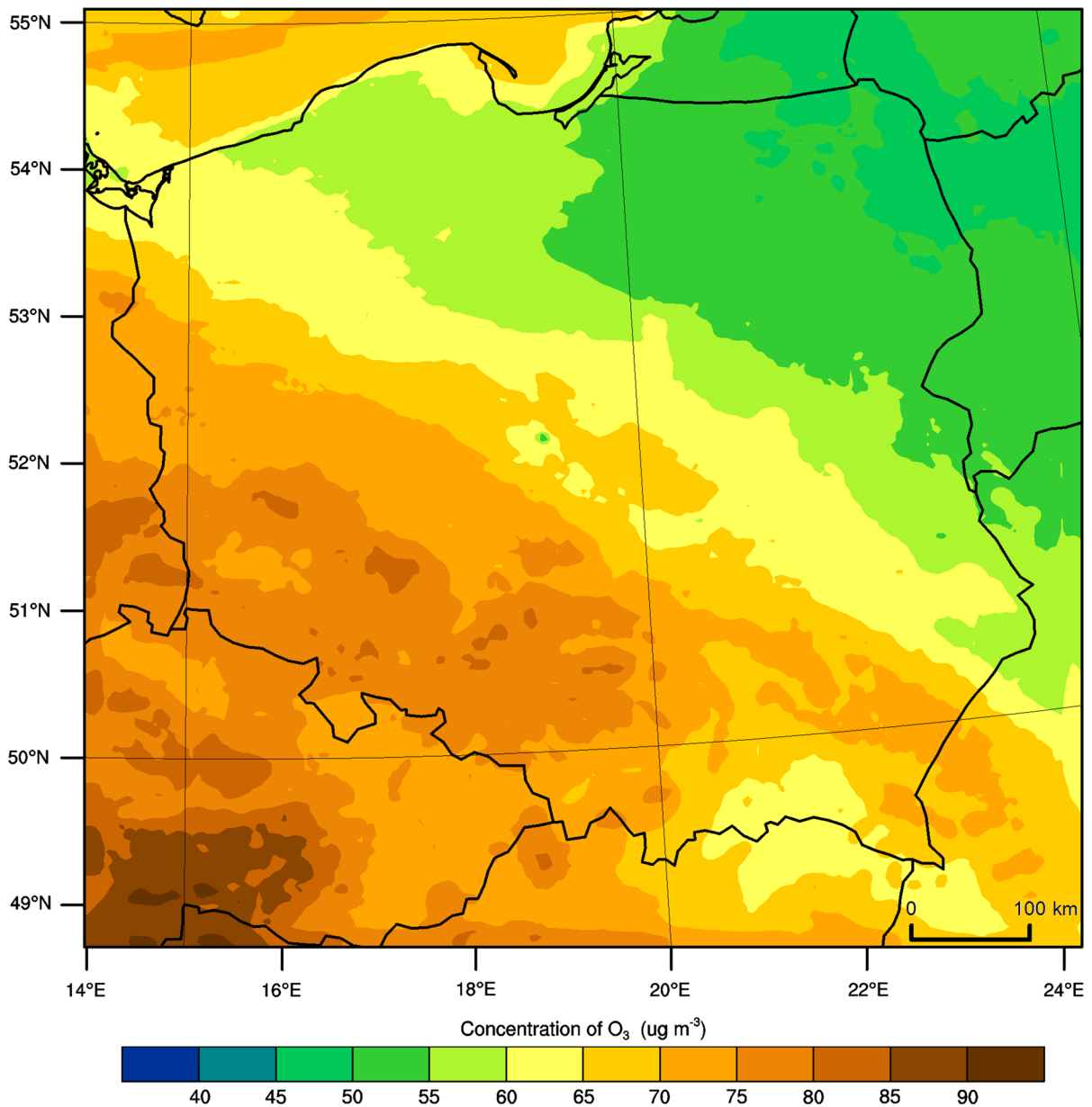


Figure 9  
Mean  $O_3$  concentration for the episode of 30 June–4 July 2008 (SIM4, Morrison 2-Moment)

passed all cloud detection tests can be classified as cloud-free. Although these data are consistent with MODIS and point surface observations, it will rather present more than less clouds (CROCKER and MITTERMAIER 2013). Another issue is the resolution of data—satellite cloud mask has similar, but not the same grid size as the model. Coarser resolution results in presenting a set of small cloud cells (e.g. *Alto cumulus*

*floccus*) as one wide patch, whereas the model resolves it differently. It may result in false underestimation of cloud cover and the average size of cloud cells, which may be the case here. Additionally, both methods are agreeable that SIM4 provides the best results of cloud cover and SIM1 presents significantly poorer performance. It refers to all analyzed cloud properties—SIM4 has the least underestimation of

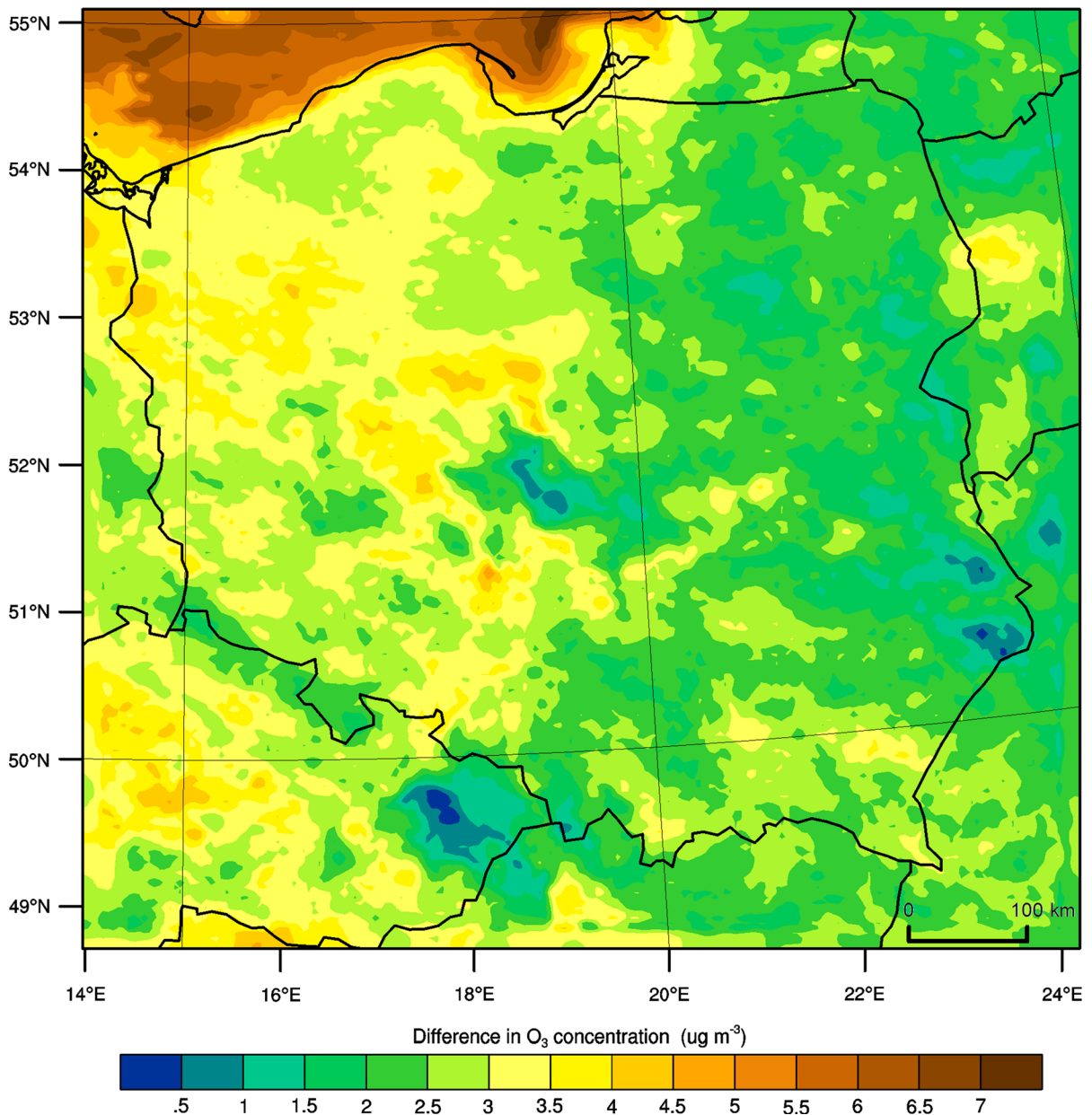


Figure 10  
Differences in mean O<sub>3</sub> concentration between SIM4 (Purdue Lin) and SIM1 (Morrison 2-Moment)

cloud size and total cloud amount, as well as its location within the domain. The difference is not as significant for surface meteorological variables, as only one performance measure for wind speed responds to the change in microphysics parameterization. However, the change of microphysics scheme has significant impact on WRF-Chem modeled ozone concentrations, particularly for high ozone

conditions. This could be attributed to the fact that cloud cover is used as input for photolysis schemes. It is important for risk assessment of critical ozone levels exceedance and its prediction. Therefore, the Morrison Double-Moment microphysics parameterization will be used in further research regarding the modeling of ozone concentrations during summer episodes in Poland and Central Europe.

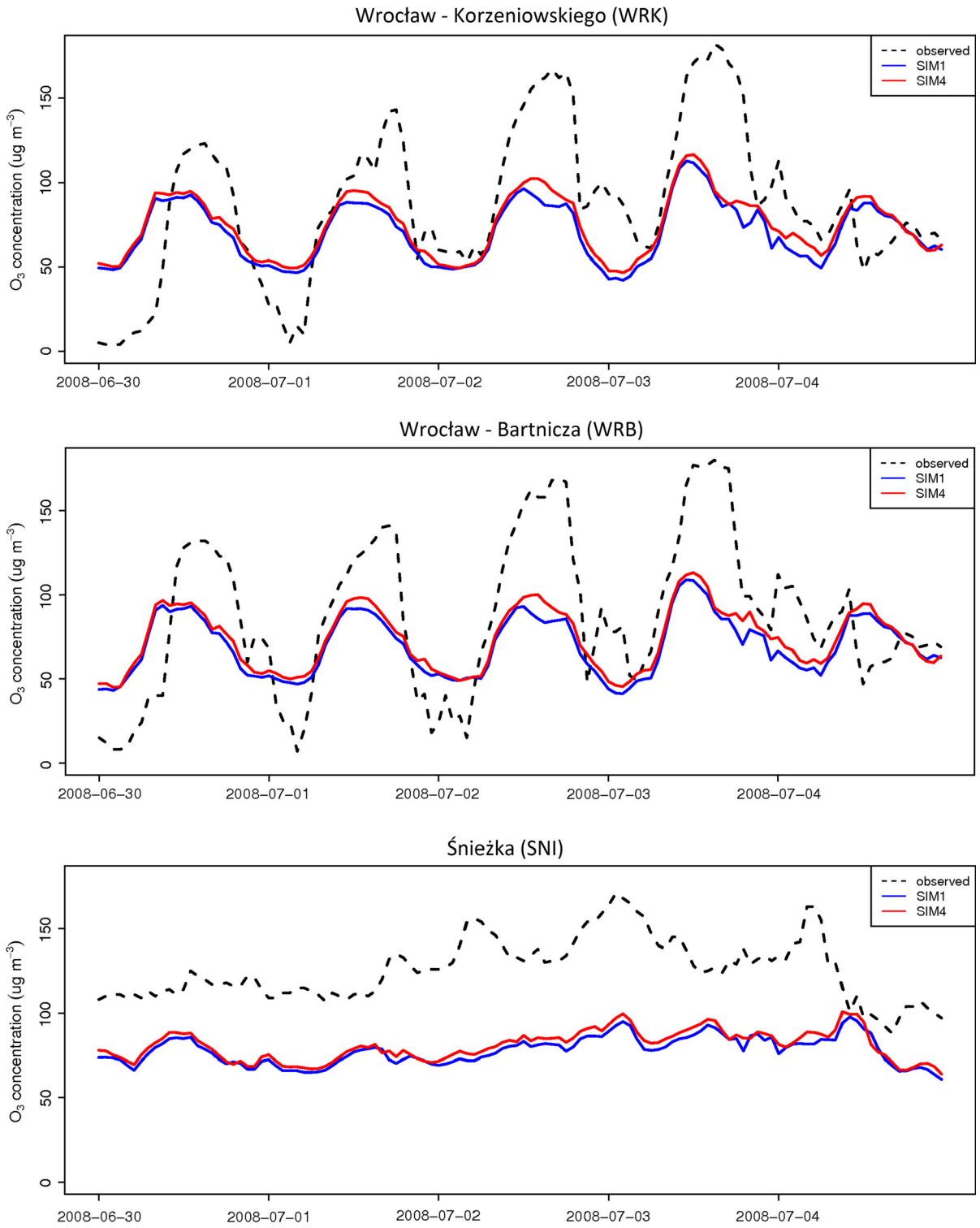


Figure 11  
Temporal variability of modeled and measured O<sub>3</sub> concentrations at WRK, WRB, and SNI station



### Acknowledgments

The study was supported by the Polish National Science Centre project no. UMO-2013/09/B/ST10/00594. The project was financed with means of the European Union under the Financial Instrument LIFE + and co-financed by the National Fund of Environmental Protection and Water Management. Calculations were carried out in the Wrocław Centre for Networking and Supercomputing (<http://www.wcss.wroc.pl>), Grant No. 170.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

### REFERENCES

- BARÓ, R., JIMÉNEZ-GUERRERO, P., BALZARINI, A., CURCI, G., FORKEL, R., GRELL, G., HIRT, M., HONZAK, L., LANGERF, M., PÉREZ, J. L., PIROVANO, G., SAN JOSÉ, R., TUCCCELLA, P., WERHAHN, J., and ŽABKAR, R. (2015). *Sensitivity analysis of the microphysics scheme in WRF-Chem contributions to AQMEII phase 2*. Atmospheric Environment, 115, 620–629. doi:10.1016/j.atmosenv.2015.01.047.
- BRETHERTON, C., KLINKER, E., BETTS, A., and COAKLEY, J. J. (1995). *Comparison of ceilometer, satellite, and synoptic measurements of boundary-layer cloudiness and the ECMWF diagnostic cloud parameterization scheme during*. Journal of the Atmospheric Sciences, 52(16), 2736–2751. Retrieved from <http://fir.library.oregonstate.edu/xmlui/handle/1957/26417>.
- CROCKER, R., and MITTERMAIER, M. (2013). *Exploratory use of a satellite cloud mask to verify NWP models*. Meteorological Applications, 20(2), 197–205. doi:10.1002/met.1384.
- DERRIEN, M., GLÉAU, H. LE, and RAOUL, M. (2010). The use of the high resolution visible in SAFNWC/MSG cloud mask. In Proceedings of the 2010 EUMETSAT Meteorological Satellite Conference, 20–24 September 2010, Córdoba, Spain (p. 57 pp.). Retrieved from <http://hal.archives-ouvertes.fr/meteo-00604325/>.
- FENSHOLT, R., ANYAMBA, A., HUBER, S., PROUD, S. R., TUCKER, C. J., SMALL, J., PAK, E., RASMUSSEN, M. O., SANDHOLT, I., and SHISANYA, C. (2011). *Analysing the advantages of high temporal resolution geostationary MSG SEVIRI data compared to Polar Operational Environmental Satellite data for land surface monitoring in Africa*. International Journal of Applied Earth Observation and Geoinformation 13(5), 721–729. doi:10.1016/j.jag.2011.05.009.
- FORKEL, R., BALZARINI, A., BARÓ, R., BIANCONI, R., CURCI, G., JIMÉNEZ-GUERRERO, P., HIRT, M., HONZAK, L., LORENZ, C., IM, U., PÉREZ, J. L., PIROVANO, G., SAN JOSÉ, R., TUCCCELLA, P., WERHAHN, J., and ŽABKAR, R. (2015). *Analysis of the WRF-Chem contributions to AQMEII phase 2 with respect to aerosol radiative feedbacks on meteorology and pollutant distributions*. Atmospheric Environment, 115, 630–645. doi:10.1016/j.atmosenv.2014.10.056.
- GILBERT, G. K. (1884). *Finley's tornado predictions*. American Meteorological Journal, 1, 166–172.
- GRASSO, L., LINDSEY, D. T., SUNNY LIM, K.-S., CLARK, A., BIKOS, D., and DEMBEK, S. R. (2014). *Evaluation of and Suggested Improvements to the WSM6 Microphysics in WRF-ARW Using Synthetic and Observed GOES-13 Imagery*. Monthly Weather Review, 142(10), 3635–3650. doi:10.1175/MWR-D-14-00005.1.
- GRELL, G., PECKHAM, S., and SCHMITZ, R. (2005). *Fully coupled "online" chemistry within the WRF model*. Atmospheric Environment, 39, 6957–6975. doi:10.1016/j.atmosenv.2005.04.027.
- HONG, S., and LIM, J. (2006). *The WRF single-moment 6-class microphysics scheme (WSM6)*. J. Korean Meteor. Soc., 42(2), 129–151. Retrieved from [http://www.mmm.ucar.edu/wrf/users/docs/WSM6-hong\\_and\\_lim\\_JKMS.pdf](http://www.mmm.ucar.edu/wrf/users/docs/WSM6-hong_and_lim_JKMS.pdf).
- HONG, S.-Y., SUNNY LIM, K.-S., KIM, J.-H., JADE LIM, J.-O., and DUDHIA, J. (2009). *Sensitivity Study of Cloud-Resolving Convective Simulations with WRF Using Two Bulk Microphysical Parameterizations: Ice-Phase Microphysics versus Sedimentation Effects*. Journal of Applied Meteorology and Climatology, 48(1), 61–76. doi:10.1175/2008JAMC1960.1.
- IACONO, M. J., DELAMERE, J. S., MLAWER, E. J., SHEPARD, M. W., CLOUGH, S. A., and COLLINS, W. D. (2008). *Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models*. Journal of Geophysical Research: Atmospheres, 113(D13), D13103. doi:10.1029/2008JD009944.
- JOHNSON, J. S., CUI, Z., LEE, L. A., GOSLING, J. P., BLYTH, A. M., and CARSLAW, K. S. (2015). *Evaluating uncertainty in convective cloud microphysics using statistical emulation*. Journal of Advances in Modeling Earth Systems, 7(1), 162–187. doi:10.1002/2014MS000383.
- JOLLIFFE, I. T., and STEPHENSON, D. B. (2003). *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons, Ltd.
- KING, M. D., MENZEL, W. P., KAUFMAN, Y. J., TANRÉ, D., GAO, B. C., PLATNICK, S., ACKERMAN, S. A., REMER, L. A., PINCUS, R., and HUBANKS, P. A. (2003). *Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS*. IEEE Transactions on Geoscience and Remote Sensing, 41(2 PART 1), 442–456.
- KRYZA, M., WAIASZEK, K., OJRZYŃSKA, H., SZYMANOWSKI, M., WERNER, M., and DORE, A. J. *High resolution dynamical downscaling of ERA-Interim using the WRF regional climate model. Part 1: model configuration and statistical evaluation for the 1981-2010 period*. Pure and Applied Geophysics (in revision).
- KRZYŚCIN, J. W., RAJEWSKA-WIĘCH, B., and JAROSIAWSKI, J. (2013). The long-term variability of atmospheric ozone from the 50-yr observations carried out at Belsk (51.84°N, 20.78°E), Poland. Tellus B; Vol. 65.
- KUENEN, J. J. P., VISSCHEDIJK, A. J. H., JOZWICKA, M., and DENIER VAN DER GON, H. A. C. (2014). *TNO-MACC\_II emission inventory; a multi-year (2003&ndash;2009) consistent high-resolution European emission inventory for air quality modelling*. Atmos. Chem. Phys., 14(20), 10963–10976. doi:10.5194/acp-14-10963-2014.
- LIN, Y.-L., FARLEY, R. D., and ORVILLE, H. D. (1983). *Bulk Parameterization of the Snow Field in a Cloud Model*. Journal of Climate and Applied Meteorology, 22(6), 1065–1092. doi:10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2.

- LIU, S. C., MCKEEN, S. A., HSIE, E.-Y., LIN, X., KELLY, K. K., BRADSHAW, J. D., SANDHOLM, S. T., BROWELL, E. V., GREGORY, G. L., SACHSE, G. W., BANDY, A. R., THORNTON, D. C., BLAKE, D. R., ROWLAND, F. S., NEWELL, R., HEIKES, B. G., SINGH, H., and TALBOT, R. W. (1996). *Model study of tropospheric trace species distributions during PEM-West A*. Journal of Geophysical Research, 101(D1), 2073. doi:10.1029/95JD02277.
- MADRONICH, S. (1987). *Photodissociation in the atmosphere: I. Actinic flux and the effects of ground reflections and clouds*. Journal of Geophysical Research: Atmospheres, 92(D8), 9740–9752. doi:10.1029/JD092iD08p09740.
- MLAWER, E. J., TAUBMAN, S. J., BROWN, P. D., IACONO, M. J., and CLOUGH, S. A. (1997). *Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave*. Journal of Geophysical Research: Atmospheres, 102(D14), 16663–16682. doi:10.1029/97JD00237.
- MORRISON, H., THOMPSON, G., and TATARSKII, V. (2009). *Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes*. Monthly Weather Review, 137(3), 991–1007. doi:10.1175/2008MWR2556.1.
- OTKIN, J.A., and GREENWALD, T. J. (2008). *Comparison of WRF Model-Simulated and MODIS-Derived Cloud Data*. Monthly Weather Review, 136(6), 1957–1970. doi:10.1175/2007MWR2293.1.
- PETERSON, T. C., and VOSE, R. S. (1997). *An Overview of the Global Historical Climatology Network Temperature Database*. Bulletin of the American Meteorological Society, 78(12), 2837–2849. doi:10.1175/1520-0477(1997)078<2837:A00TGH>2.0.CO;2.
- QIAN, Y., LONG, C. N., WANG, H., COMSTOCK, J. M., MCFARLANE, S. A., and XIE, S. (2012). *Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations*. Atmospheric Chemistry and Physics, 12(4), 1785–1810. doi:10.5194/acp-12-1785-2012.
- ROGERS, E., EK, M., FERRIER, B. S., GAYNO, G., LIN, Y., MITCHELL, K., PONDECA, M., PYLE, M., WONG, V. C. K., and WU, W. S. (2005). *The NCEP North American Mesoscale Modeling System: Final Eta model/analysis changes and preliminary experiments using the WRF-NMM*. In 21st Conference on Weather Analysis and Forecasting/17th Conference on Numerical Weather Prediction, Washington, D.C.
- SAIDE, P. E., SPAK, S. N., CARMICHAEL, G. R., MENA-CARRASCO, M. A., YANG, Q., HOWELL, S., LEON, D. C., SNIDER, J. R., BANDY, A. R., COLLETT, J. L., BENEDICT, K. B., de SZOEKE, S. P., HAWKINS, L. N., ALLEN, G., CRAWFORD, I., CROSIER, J., and SPRINGSTON, S. R. (2012). *Evaluating WRF-Chem aerosol indirect effects in Southeast Pacific marine stratocumulus during VOCALS-REx*. Atmospheric Chemistry and Physics, 12(6), 3045–3064. doi:10.5194/acp-12-3045-2012.
- SCHAEFER, J. T. (1990). *The Critical Success Index as an Indicator of Warning Skill*. Weather and Forecasting, 5(4), 570–575. doi:10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.
- SEGELE, Z. T., LESLIE, L. M., and LAMB, P. J. (2013). *Weather Research and Forecasting Model simulations of extended warm-season heavy precipitation episode over the US Southern Great Plains: data assimilation and microphysics sensitivity experiments*. Tellus A; Vol 65 (2013).
- SKAMAROCK, W., and KLEMP, J. (2008). *A time-split nonhydrostatic atmospheric model for weather research and forecasting applications*. Journal of Computational Physics, 227, 3465–3485. doi:10.1016/j.jcp.2007.01.037.
- SÖHNE, N., CHABOUREAU, J.-P., and GUICHARD, F. (2008). *Verification of Cloud Cover Forecast with Satellite Observation over West Africa*. Monthly Weather Review, 136(11), 4421–4434. doi:10.1175/2008MWR2432.1.
- STASZEWSKI, T., KUBIESA, P., and ŁUKASIK, W. (2012). *Response of spruce stands in national parks of southern Poland to air pollution in 1998–2005*. European Journal of Forest Research, 131(4), 1163–1173. doi:10.1007/s10342-011-0587-0.
- TIE, X., MADRONICH, S., WALTERS, S., ZHANG, R., RASCH, P., and COLLINS, W. (2003). *Effect of clouds on photolysis and oxidants in the troposphere*. Journal of Geophysical Research: Atmospheres, 108(D20), 4642. doi:10.1029/2003JD003659.
- TUINDER, O. N. E., de WINTER-SORKINA, R., and BUILTJES, P. J. H. (2004). *Retrieval methods of effective cloud cover from the GOME instrument: an intercomparison*. Atmospheric Chemistry and Physics, 4(1), 255–273. doi:10.5194/acp-4-255-2004.
- VAN LIER-WALQUI, M., VUKICEVIC, T., and POSSELT, D. J. (2012). *Quantification of Cloud Microphysical Parameterization Uncertainty Using Radar Reflectivity*. Monthly Weather Review, 140(11), 3442–3466. doi:10.1175/MWR-D-11-00216.1.
- VOSE, R. S., SCHMOYER, R. L., STEURER, P. M., PETERSON, T. C., HEIM, R., KARL, T. R., and EISCHEID, J. K. (1992). *The Global Historical Climatology Network: Long-Term Monthly Temperature, Precipitation, Sea Level Pressure, and Station Pressure Data*. Environmental Sciences Division Publication no. 3912, doi: 10.3334/CDIAC/clf.ndp041.
- WERNLI, H., PAULAT, M., HAGEN, M., and FREI, C. (2008). *SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts*. Monthly Weather Review, 136(11), 4470–4487. doi:10.1175/2008MWR2415.1.
- WILD, O., ZHU, X., and PRATHER, M. (2000). *Fast-J: Accurate Simulation of In- and Below-Cloud Photolysis in Tropospheric Chemical Models*. Journal of Atmospheric Chemistry, 37(3), 245–282. doi:10.1023/A:1006415919030.
- WILLMOTT, C. J. (1981). *On the Validation of Models*. Physical Geography, 2(2), 184–194.
- WILKS, D. S. (2006). *Statistical methods in the atmospheric sciences*, 704 p., Elsevier Inc., ISBN 9780123850225.
- World Meteorological Organization. (2008). *Guide to Meteorological Instruments and Methods of Observation*. WMO No. 8.
- ZHANG, Y., CHEN, Y., SARWAR, G., and SCHERE, K. (2012). *Impact of gas-phase mechanisms on Weather Research Forecasting Model with Chemistry (WRF/Chem) predictions: Mechanism implementation and comparative evaluation*. Journal of Geophysical Research, 117(D1), D01301. doi:10.1029/2011JD015775.
- ZINGERLE, C., and NURMI, P. (2008). *Monitoring and verifying cloud forecasts originating from operational numerical models*, Meteorological Applications, 15(3), 325–330. doi:10.1002/met.73.