

# The impact of initial and boundary conditions on severe weather event simulations using a high-resolution WRF model. Case study of the derecho event in Poland on 11 August 2017

60

**Mariusz Figurski<sup>1</sup>, Grzegorz Nykiel<sup>1,2</sup>, Adam Jaczewski<sup>1</sup>, Zofia Bałdysz<sup>1</sup>, Marcin Wdowikowski<sup>1</sup>**

<sup>1</sup> Institute of Meteorology and Water Management – National Research Institute, Poland

<sup>2</sup> Gdansk University of Technology, Faculty of Civil and Environmental Engineering, Poland

**DOI: 10.26491/mhwm/143877**

---

**ABSTRACT.** Precise simulations of severe weather events are a challenge in the era of changing climate. By performing simulations correctly and accurately, these phenomena can be studied and better understood. In this paper, we have verified how different initial and boundary conditions affect the quality of simulations performed using the Weather Research and Forecasting Model (WRF). For our analysis, we chose a derecho event that occurred in Poland on 11 August 2017, the most intense and devastating event in recent years. High-resolution simulations were conducted with initialization at 00 and 12 UTC (11 August 2017) using initial and boundary conditions derived from the four global models: Global Forecast System (GFS) from the National Centers for Environmental Prediction (NCEP), Integrated Forecast System (IFS) developed by the European Center for Medium-Range Weather Forecasts (ECMWF), Global Data Assimilation System (GDAS) and ERA5. For the last, we made separate calculations using data at the pressure and model levels. The results were evaluated against surface and radar data. We found that the simulations that used data from the GDAS and GFS models at 12 UTC were the more accurate, while ERA5 gave the worst predictions. However, all models were characterized by a low probability of detection and a high number of false alarms for simulations of extreme precipitation and wind gusts.

**KEYWORDS:** Simulations, severe weather, WRF, GFS, GDAS, ERA5, IFS, derecho, initial conditions, bow echo.

**SUBMITTED:** 20 February 2021 | **REVISED:** 27 October 2021 | **ACCEPTED:** 11 November 2021

## 1. INTRODUCTION

In the era of an intensely changing climate, improving our understanding of weather phenomena, along with their reliable description using numerical models, is crucial. Numerical weather prediction (NWP) models are essential for supporting decision-making especially as global warming intensifies weather-derived natural hazards (Cornwall 2016; Herring 2018). Such extreme weather phenomena, including mesoscale convective systems (MCS), which can be organized into bow echo structures with mesoscale vortices, pose a particular risk of life-threatening and economic losses (EEA 2020). Emergency management and mitigation efforts related to such phenomena strongly rely on weather forecasts, which have seen intense progress in accuracy during the last 40 years (Bauer et al. 2015).

Among many factors, initial conditions were recognized to have tremendous implications on weather prediction accuracy, since even small initial differences between two NWP solutions will grow significantly over time (Lorenz 1963; Krishnamurthy 2019). Consequently, assessing the impact of initial conditions on forecast skill has become an important task in developing reliable weather predictions. Reichler and Roads (2003) pointed out that initial conditions have a significant impact on short-term forecasts, which are dominant in extreme weather predictions, when the model error itself is less powerful. Sutton et al. (2006) used the Weather Research and Forecasting (WRF) model and verified that in the case of a high-resolution grid (5 km), the initial conditions (related to soil moisture) were comparable to the differences resulting from adopting various convective parameterizations to the lower resolution grid (20 km). From the point of view of extreme weather event predictions, Jankov et al. (2007) conducted a detailed study on the influence of various physical schemes on the mesoscale convective system for rainfall forecasting. They found that WRF rainfall forecasts modelled with various treatments of convection, microphysical schemes, and planetary boundary layers are sensitive to the datasets used for model initialization. The impact of the initial conditions on the predictability of heavy rainfall was also investigated by Bei and Zhang (2007), who pointed out that the error related to small disturbances in initial conditions leads to significant uncertainties in a mesoscale forecast. Since global models are recognized as the best source of a real-time atmospheric state that can be used as initial conditions for short-term mesoscale forecasts, their uncertainties may also degrade the final NWP accuracy (Wei et al. 2010). As a consequence, the impact of these models on the mesoscale model forecasts was investigated by, among others, Kumar et al. (2015). Using the WRF model and four different global models, namely, the Integrated Forecast System (IFS), developed by the European Center for Medium-Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS), the NCEP Global Forecasting System (GFS), and the National Center for Medium-Range Weather Forecasting (NCMRWF), they found that forecasts initialized using ECMWF/IFS produced the best solution for the Indian region. Some advantages from using IFS for the initialization of the WRF model were also found by Taszarek et al. (2019). Based on a derecho event, they found that, in contrast to the GFS, all IFS-based simulations correctly pointed out the possibility of extreme winds, although GFS-based simulations with a shorter lead time performed better for both the location and timing of extreme strong wind gusts (over  $40 \text{ m}\cdot\text{s}^{-1}$ ).

In this study, we present the effects of using varied initial conditions on simulations of severe weather events. Our analysis focused on a derecho event that occurred in Poland on 11 August 2017, one of the most intense and devastating events in recent years (Widawski, Pilorz 2018). We were

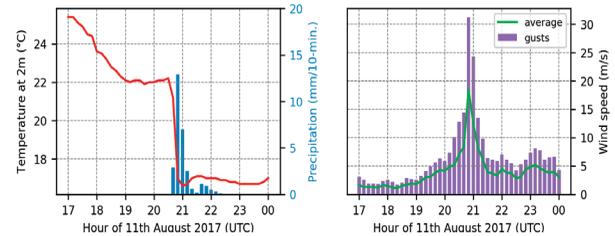


Fig. 1. Series of meteorological parameters observed at the Chojnice meteorological station. Air temperature (left, red line), precipitation (left, blue bars), wind speed (right, green line), and wind gusts (right, purple bars) are presented.

looking for answers as to which initial conditions would predict this phenomenon as accurately as possible. First, we provide information about the event. Second, we introduce the data and methodology, including a description of the WRF model domains and parameterization, information on all initial conditions tested, and the meteorological data used for quality evaluation. Next, we present results of the simulated meteorological parameters and validation using in situ ground measurements. Finally, we provide some conclusions and a discussion of the analysis.

## 2. DERECHO IN POLAND, 11 AUGUST 2017

Although the derecho analyzed in this study occurred on 11 August 2017, the weather conditions that contributed to its creation started to develop a few days earlier. They were mainly related to the formulation and movements of two pressure systems, namely a long-wave trough over western Europe and a wide ridge that was situated over central and eastern Europe. The hot and dry subtropical air masses associated with the ridge were separated from the much more humid, westerly located polar air masses by a wavy cold front that stretched along the western border of Poland. On 10 August, a slight change in the orientation of the low-pressure system enhanced the meridian-directed flow of tropical air from the Mediterranean Basin over central Europe. This dynamic triggered a mesoscale convective system over the Czech territory, which during the night passed over the territory of Poland. The storms associated with the system contributed to increased cloudiness, which effectively reduced the insolation over southwestern Poland until afternoon. At the same time, the air masses over the eastern part of Poland were constantly heating under cloudless conditions, contributing to the creation of a strong thermal gradient. The convergence line between these two pressure systems was moving northward, under the influence of the mid-level jet stream, which led to the formulation of a distinct vertical wind shear. At about 1800 UTC the convective cells embedded in the northerly propagating bow echo that was constantly growing and finally evolved into a mesoscale convective vortex (MCV). The development and transformation of convective cells into MCV, and as a consequence into a derecho, was supported by the appearance of a rear inflow jet (Sulik, Kejna 2020). The thunderstorm reached its strongest form between 20:00 and 2100 UTC, when the echo structure was about 150 km long. At that time, at the Chojnice synoptic station, precipitation exceeded  $13 \text{ mm}\cdot 10 \text{ min}^{-1}$  and was accompanied by a sharp drop in temperature (from  $22.2$  to  $16.5^\circ\text{C}$ ) (Fig. 1, left). At the same time, the average wind speed exceeded  $18.4 \text{ m}\cdot\text{s}^{-1}$  with a wind gust of up to  $31.2 \text{ m}\cdot\text{s}^{-1}$  (Fig. 1, right). Very high wind gusts were also recorded at synoptic stations in Elblag ( $42 \text{ m}\cdot\text{s}^{-1}$ ), Chrzastowo ( $36 \text{ m}\cdot\text{s}^{-1}$ ), and Gniezno ( $35 \text{ m}\cdot\text{s}^{-1}$ ). At 2230 UTC, the convective system reached the coast of the Baltic Sea and began to weaken visibly. The synoptic situation was described in detail by Wrona et al. (2022).

### 3. DATA AND METHODOLOGY

Here, we provide information on the WRF model used for simulations, together with domain characteristics, settings, and parameterization of physical processes. We also briefly introduce the initial conditions with some notes on their implementation in the WRF model. Lastly, we show meteorological data used for simulation verification: reflectivity and basic parameters from telemetry meteorological stations.

#### 3.1. WRF MODEL

We used the WRF model version 4.2.1 in nonhydrostatic mode, which has wide applications in atmospheric research and operational weather forecasting (Skamarock et al. 2019). Our version of the model was adapted to work on a high-powered Tryton computer at the Academic Computer Centre in Gdansk. For the simulation, 568 cores of the supercomputer were used, with 552 cores for the WRF model and 16 for I/O operations. This approach minimized the time it took to write the results by a factor of 10 compared to a standard run without I/O enabled.

The derecho event was simulated using various initial and boundary conditions derived from the four global models: NCEP/GFS, ECMWF/IFS, GDAS, and ERA5. All models were provided with a time resolution of 3 hours. The spatial resolution of the GFS, GDAS, and ERA5 models is  $0.25^\circ$ . Therefore, in our simulations, we designed three nested domains in one direction with a nesting ratio of 5. Horizontal resolutions of domains, in Lambert conformal projection, were 12.5, 2.5, and 0.5 km. The time steps were 60, 12, and 3 seconds, respectively, for each domain. The time resolution was 1 hour for the outermost domain and 10 minutes for the other two. Because the spatial resolution of the IFS model is  $0.125^\circ$ , the first domain in our simulations was 7.5 km. The areas covered by these three domains were the same for all models. The first domain (domain 1) covers most of Europe, the second one (domain 2), the area of Poland, while the third one (domain 3 – with the highest resolution) covers the area where the bow echo and the greatest damage occurred. In Figure 2, the coverages of domains 2 and 3 are presented in detail. Vertically, simulations were done for 50 levels up to 50 hPa. Since the ERA5 model levels are provided on the model levels (137) and pressure levels (38), we performed calculations using both of them, called ERA5M and ERA5P, respectively.

For all simulations, the same parameterization of physics and model dynamics was used. We applied a single-moment microphysics scheme with six hydrometeor classes (WSM6) (Zaidi, Gisen 2018), which is the most suitable for high-resolution simulations (see, e.g., Hong et al. 2010; Parodi et al. 2019). For domain 1, parameterization of convective processes was performed using the Grell-Freitas method (Grell, Freitas 2014), while for domains 2 and 3, explicit wet process physics was used. Moreover, we applied short-wave and long-wave radiation parameterizations according to the RRTMG radiation propagation scheme, which is a new version of the RRTM (Iacono et al. 2008). The Mellor Yamada Nakanishi Niino (MYNN) turbulence scheme with closure 2.5 was used to model boundary layer processes (Nakanishi, Niino 2009). The near-surface layer was parameterized according to the MYNN scheme (Nakanishi, Niino 2006). To remove numerical noise at the start of the simulations, digital filter initialization (DFI) was used (Peckham et al. 2016). The land topography, land use, and soil type datasets were included in the model at the WRF preprocessing stage. For domains 1 and 2, the standard data contained in the WRF model geographic database (LULC) of IGBP MODIS and USGS GMTED2010 (30 arc seconds) were used. However, these data are insufficient for high-resolution simulations, as indicated by other studies (De Meij, Vinuesa 2014; Jiménez-Esteve et al. 2018; Siewert, Kroszczyński, 2020). Therefore, for domain 3 with a 0.5 km spatial step, we prepared new geographic data based on CORINE Land Cover (CLC) 2018 with 100 m and terrain topography from the Shuttle Radar Topography Mission (SRTM) with 30 m resolution. Siewert and Kroszczyński (2020) showed

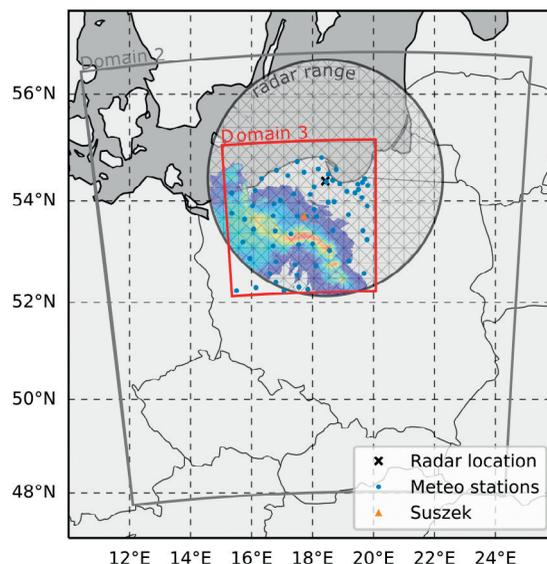


Fig. 2. WRF domain used in the study. The red rectangle represents a 500 m domain analyzed in this paper. The black cross shows the location of the meteorological radar in Gdansk (Poland) with a range equal to 250 km. The blue dots show the meteorological stations used for the validation of the results. The orange triangle identifies the location of Suszek, a village where the greatest damage was observed. In the background, the reflectivity derived from the meteorological radar in Gdansk is shown at 2000 UTC on 11 August 2017.

that CLC and SRTM data for microscale halves yield more accurate values for temperature and humidity at 2 m and wind at 10 m (speed and direction) compared to using default data from the WRF database.

In our analyses, we focused on the results derived from the innermost domain, which were stored with a 10-minute interval. Furthermore, we took into account initialized simulations at 0000 and 1200 UTC on 11 August 2017 to assess how the results depend on the starting time of the simulations.

#### 3.2. REFLECTIVITY DATA

In this study, the WRF reflectivity data were compared with data from the Meteor 1500C Doppler meteorological radar in Gdansk (Fig. 2, the black cross) included in the POLRAD network. This radar provides reflectivity scans in the range of 250 km, with 10-min interval. Also, Doppler velocity scans are available in the range of 125 km, at a 6 min interval. Each scan contains 9 slices for different elevation angles:  $0.5^\circ$ ,  $1.4^\circ$ ,  $2.4^\circ$ ,  $3.4^\circ$ ,  $5.3^\circ$ ,  $7.7^\circ$ ,  $10.6^\circ$ ,  $14.1^\circ$ ,  $18.5^\circ$ ,  $23.8^\circ$ . In the analysis, we used only reflectivity data after quality improvement using the RADVOL-QC system (Ośródko et al. 2014). In Figure 2 the radar range together with example maximum reflectivity extracted at 2000 UTC are presented.

#### 3.3. METEOROLOGICAL OBSERVATIONS

We used meteorological data from the basic measurement and observation network owned by the Institute of Meteorology and Water Management, National Research Institute. The meteorological situation on 11 August 2017 was represented based on data from 66 ground automated weather stations (AWS), located in the  $52\text{--}55^\circ\text{N}$  and  $16\text{--}20^\circ\text{E}$  domain, as shown in Figure 2 (blue dots). Data at 10-minute intervals of the following meteorological elements were analyzed: precipitation, air temperature, and wind speed, including wind gusts. Measurement data were subjected to the completeness verification procedure (looking for missing data caused by technical reasons) and data correctness

(analysis of differences in time series for stations that had more than one data source). The most frequent sources of data gaps from AWS are data transfer, sensor malfunctioning, exceptional equipment maintenance, and unreasonable recorded data (Storch et al. 1999; Lompar et al. 2019).

For synoptic stations with series of hourly data made by meteorological observers, the data gaps from AWS were supplemented with data from an observer. In the case of missing data, which could not be filled with the simple interpolation between adjacent time series records (up to 30 minutes), the analyzed time series data were supplemented with data from at least three neighboring stations using triangulation or inverse distance weighting methods, commonly used in climatic research (Storch et al. 1999; Daly et al. 2000; Claridge, Chen 2006; Henn et al. 2013). When verifying the completeness and correctness of the measurement data, it was found that data deficiencies were less than 1% of all analyzed cases.

### 3.4. SIMULATION VERIFICATION

To evaluate the quality of the simulated temperature and wind speed, we calculate the following statistical parameters:

- Bias or mean error (*ME*):

$$ME = \frac{\sum_{i=1}^n (F_i - O_i)}{n} \quad (1)$$

- Root mean square error (*RMSE*):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (F_i - O_i)^2}{n}} \quad (2)$$

- Unbiased *RMSE* (*uRMSE*):

$$uRMSE = \sqrt{\frac{\sum_{i=1}^n ((F_i - \bar{F}) - (O_i - \bar{O}))^2}{n}} \quad (3)$$

- Pearson correlation (*R*):

$$R = \frac{\sum_{i=1}^n ((F_i - \bar{F})(O_i - \bar{O}))}{\sigma_F \sigma_O} \quad (4)$$

Where  $F_i$ ,  $O_i$  are the forecast and observed values, respectively, indexed with  $i$ ; means the total number of observations and  $\sigma$  denotes the standard deviation. *ME* is the mean difference between the forecasted and observed values, and therefore a valuable metric to present the tendency of the model to being over- or underestimated. On the other hand, averaging of positive and negative *ME* values may result in misinterpretation. *RMSE* or *uRMSE* is the square root of the squared forecast deviations and is used to mark better or worse simulations. We provide *ME* and *R* calculated for all time steps between 1700 and 0000 UTC and present them as time series. These measures are also calculated for individual stations and presented on the map as scatter plots for hourly time steps. We utilize Taylor (Taylor 2001) and target (Jolliff et al. 2009) diagrams to show the relationship between statistical parameters and summarize model performance. The former is a polar coordinate diagram that assigns radial coordinates to the standard deviation and azimuth to the inverse cosine of the correlation coefficient (Eq. 4). The reference point (observation) is indicated for the polar coordinates  $(\sigma_O, 0)$ . The model-to-observation distance is proportional to *uRMSE* (Eq. 3) and provides a measure of the model uncertainty. In the target diagram, *uRMSE* is assigned to the X-axis, and *ME* (Eq. 1) is assigned to the Y-axis. The distance between origin and the model versus observation statistics is equal to *RMSE* (Eq. 2). As *uRMSE* is always positive,

Jolliff et al. (2009) proposed to utilize the positive region of the X-axis and multiply *uRMSE* by the sign of the standard deviation difference:

$$\sigma_d = \text{sign}(\sigma_F - \sigma_O) \quad (5)$$

Table 1. The 2x2 contingency table.

Forecast \ Observation	Yes	No
	Yes	A
No	C	D

Verification of forecasts for precipitation and wind gusts is based on the contingency table (Table 1). For dichotomous forecast verification, the following quality measures were calculated:

- Probability of detection (*POD*):

$$POD = \frac{A}{A+C} \quad (6)$$

- False alarm ratio (*FAR*):

$$FAR = \frac{B}{A+B} \quad (7)$$

- *bias*:

$$\text{bias} = \frac{A+B}{A+C} \quad (8)$$

- Critical success index (*CSI*):

$$CSI = \frac{A}{A+B+C} \quad (9)$$

The *POD*, *CSI* (also known as threat score) is represented by the so-called success ratio (*SR*) which ranges from 0 (worst) to 1 (best). For *FAR*,  $SR = 1 - FAR$ . A bias score equal to 1 indicates an unbiased model. Higher or lower values indicate that the forecast is overestimated or underestimated, respectively. Analyses were performed for selected levels. Values of 0.5 and 1.0 mm of accumulated precipitation were chosen over 10 minutes. For wind gusts, the values of 5 and 10 m·s<sup>-1</sup> were chosen. The *CSI* are presented in the same manner as for temperature and wind speed. The performance of individual models is summarized by means of a diagram introduced by Roebber (2009).

The Roebber performance diagram is utilized to summarize the dichotomous verification of precipitation and wind gust. With simple algebraic manipulations, one can relate *CSI* and *bias* to *SR* and *POD*:

$$\text{bias} = \frac{POD}{SR} \quad (10)$$

$$CSI = \frac{1}{\frac{1}{SR} + \frac{1}{POD} - 1} \quad (11)$$

If *SR* and *POD* are assigned to the X- and Y-axes, respectively, then isolines of *bias* and *CSI* can be drawn on a figure. A perfect forecast should be located on the upper right of the diagram.

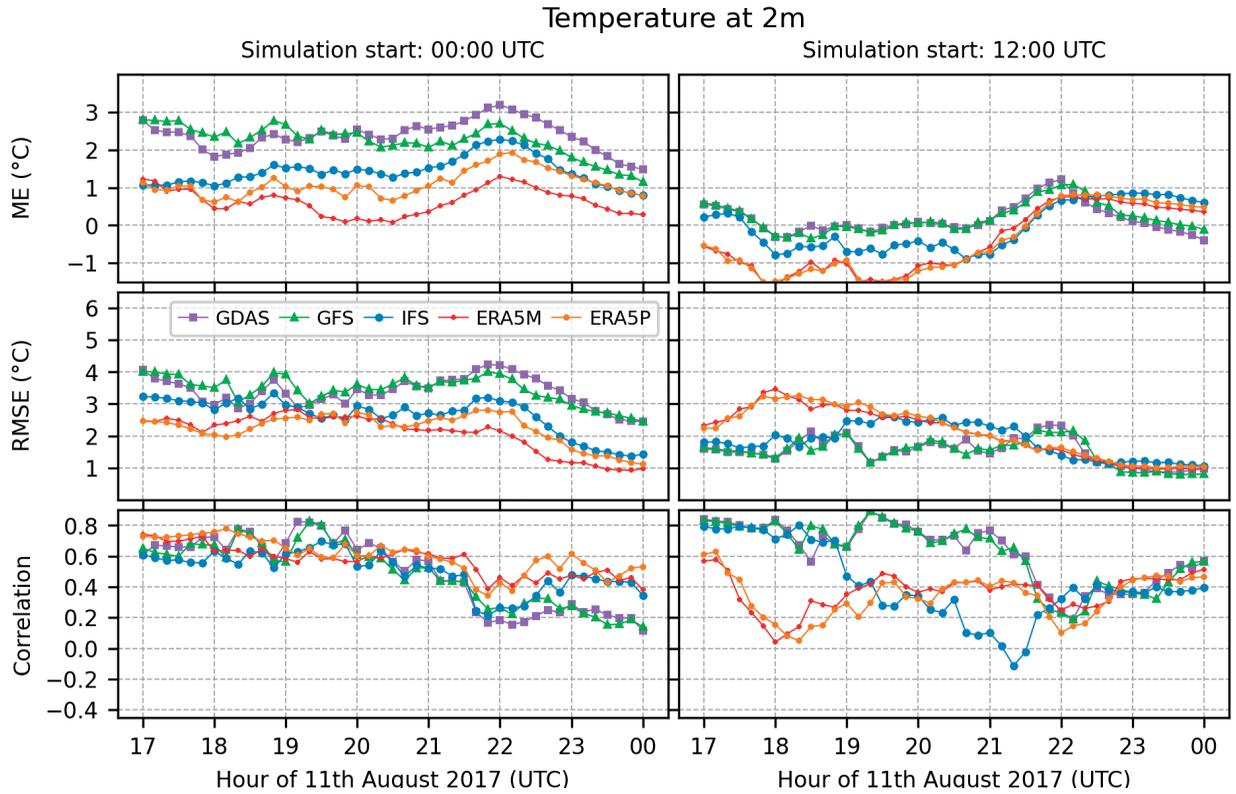


Fig. 3. Statistical parameters of the comparison temperature at 2 m derived from WRF simulations and in situ observations at meteorological stations. From the top: mean error (*ME*), root mean square error (*RMSE*), correlation coefficient.

64

Reflectivity was verified using maximum reflectivity fields derived from meteorological radar. To access which model gives more reliable results, we used the index of agreement proposed by Willmott and Wicks (1980) and Willmott (1981). This index is a dimensionless measurement of the accuracy of the model and is used in several meteorological and hydrological studies. We used a modified version of this index ( $d_{mod}$ ) proposed by Willmott (1984), which may be regarded as a more rigorous method than the original version (Pereira et al. 2018). As shown by Willmott et al. (2012),  $d_{mod}$  reaches its maximum value more slowly as the predicted values approach the observed values. The modified Willmott index is described by the following formula:

$$d_{mod} = 1 - \frac{\sum_{i=1}^n |F_i - O_i|}{\sum_{i=1}^n (|F_i - \bar{O}| + |O_i - \bar{O}|)} \quad (12)$$

$d_{mod}$  is bounded by 0 and 1: no agreement and a perfect fit, respectively.

#### 4. RESULTS

This section describes the results of the simulations, as well as their verification based on in situ measurements. We divided this section into subsections by meteorological parameters: temperature, wind speed, precipitation, wind gusts, and reflectivity. Some visualizations of the results are included in the Appendix. The analysis of temperature and wind speed is divided into two stages. In the first, model errors are analyzed in subsequent time steps, and in the second, at the location of each station. Results are presented separately for simulations initialized at 0000 and 1200 UTC. Evaluation of precipitation and wind gusts is based on dichotomous verification for two exceedance levels. The range and spatial distribution of *CSI* are presented. For reflectivity, the range of the Willmott index is shown.

#### 4.1. TEMPERATURE

In general, the predicted temperatures for 0000 UTC were noticeably higher than those obtained from the simulation at 1200 UTC. It is clearly seen in Figure 3 where *ME*, *RMSE* and the correlation of the temperature at 2 m between all simulations and in situ measurements are shown. For all calculations initialized at 0000 UTC, the temperature is overestimated. The largest *ME* are observed for GDAS-driven simulations with a maximum of 3°C at 2200 UTC. *ME* for runs initialized at 1200 UTC, in turn, are in the range of -1.5 to 1.5, except for ERA5-driven simulations that exceed -1°C. The most accurate results were obtained for simulations driven by GDAS and GFS. *ME* are slightly fluctuating for all runs until 2100 UTC, then the errors approach maximum values and decrease after 2200 UTC. The *RMSE* series for 1200 UTC runs follows the *ME* behavior. The highest values were found for GDAS and GFS with a maximum >4°C at about 2200 UTC. Following the coldest bias, ERA5-driven simulations initialized at 1200 UTC show the highest *RMSE* values with a peak at 1800 UTC. After 2200 UTC, the *RMSE* is visibly decreasing for all results. Correlation values calculated for every time step can be interpreted as quasi-spatial correlation of observation and model fields, except that the data are non-uniformly gridded. The correlation values for the 0000 UTC models are very close and quite high (0.5-0.8) until 2000 UTC when the *R* values decrease and clearly differ between models; the highest are for the ERA5 models (0.5-0.6) and the lowest for GDAS and GFS (0.2). For models initialized at 1200 UTC, the correlations are quite high for GDAS and GFS until 2100 and for IFS until 1900 UTC. Then, the *R* values decrease rapidly for the latter model. The correlations for the ERA5 models are the smallest, except for the period between 1900 and 2200 UTC where the IFS is the least correlated. The *R* values of all calculations initial-

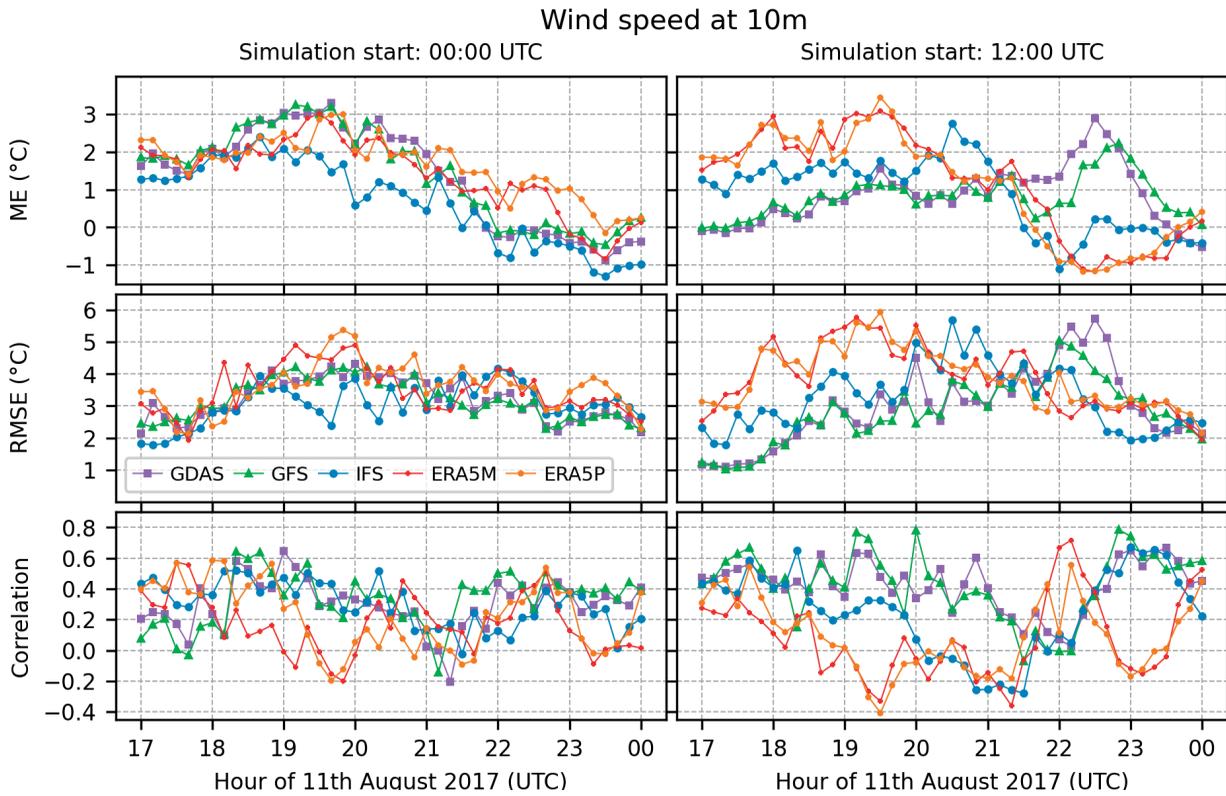


Fig. 4. Statistical parameters comparing wind speed at 10 m derived from the WRF model and in situ observations at meteorological stations. From the top: mean error (*ME*), root mean square error (*RMSE*), correlation.

ized at 1200 UTC behave similarly and increase slightly from 0.1-0.4 at 22 UTC to 0.4-0.6 at the end of the simulations.

More details about simulation differences are given in Figures A1 and A2 (Appendix), where selected temperature distribution maps between 0000 and 1200 UTC are shown. Among all results, those based on the ERA5 model stand out, mainly due to the differences in temperature between the 0000 and 1200 UTC simulations. Also, *ME* values for 0000 UTC GDAS- and GFS-driven WRF simulations are the highest for most of the stations (Fig. A3). The lowest values are found in the western part of the domain. The *ME* values for the IFS model are highest in the southern part and lowest in the northern part. The mean errors for the ERA5-driven simulations are the highest in the eastern part and the lowest in the western part of the domain. The *ME* values for the 1200 UTC models are quite similar in all domains except for the simulations driven by IFS and ERA5, which exhibit negative values in the eastern region. The *RMSE* (Fig. A4) are the largest for GDAS- and GFS- driven WRF calculations initialized at 0000 UTC, with maximum values in the eastern region, much like ERA5 driven runs initialized at 1200 UTC. The lowest *RMSE* are seen for GDAS and GFS from 1200 UTC. The correlation (Fig. A5) is  $>0.5$  in all domains for all results, with the lowest value for the station located on the Baltic coast.

#### 4.2. WIND SPEED AT 10 M

As in the case of temperature, the difference in wind speed was obtained for various initial conditions and the initialization of the simulation time. The mean error of the wind speed (Fig. 4) predicted from 0000 UTC increases to a maximum value of 2-3  $\text{m}\cdot\text{s}^{-1}$  at 1920 UTC and then decreases to negative values for all models except ERA5P. The lowest *ME* values are found until 2100 UTC for the GDAS and GFS predictions initialized at noon.

The values increase slightly from 0 to 1  $\text{m}\cdot\text{s}^{-1}$  at 2100 UTC. After this time, the *ME* for GDAS and GFS is rapidly increasing to 3  $\text{m}\cdot\text{s}^{-1}$  and 2  $\text{m}\cdot\text{s}^{-1}$ , respectively, and decreasing to the  $(-1, 0)$  range for other models. Maximum *RMSE* values are found at 2000 UTC, the largest for ERA5 driven simulations. The discrepancy for predictions initialized at 1200 UTC is much greater. The *RMSE* for ERA5 driven calculations exceeds 3  $\text{m}\cdot\text{s}^{-1}$ . Spatial correlations do not exceed 0.6 and decrease to  $-0.2$ . Positive correlation for the entire series is only found for IFS initialized at 0000 UTC and GDAS and GFS started at 1200 UTC. The results for the subsequent initialization time are more variable.

Detailed differences in the spatial distribution of the winds and their speed, obtained by adopting various initial conditions, can be observed in Figures A6 and A7, where wind gusts at 10 m between 1800 and 2200 UTC are presented. The spatial distribution of *ME* (Fig. A8) and *RMSE* (Fig. A9) is very variable and does not reveal any special pattern, unlike in the case of correlation (Fig. A10). Stations with high temporal correlations are found in the western and southern part of the GDAS and GFS WRF predictions initialized at 0000 UTC. The simulations of these models also have high *R* values for most locations except the southern region.

#### 4.3. PRECIPITATION

*CSI* values for events with precipitation sum  $>0.5 \text{ mm}\cdot 10 \text{ min}^{-1}$  (Fig. 5) are lower than 0.4 except for 2 cases for GDAS-driven simulations initialized at 0000 UTC. The cases are related to isolated precipitation events that appear from the southwestern direction (Fig. A11). Similar events are seen for the initialized predictions for IFS and ERA5P at 1200 UTC (Fig. A12). The *CSI* for runs initialized at 1200 UTC has larger values than those initialized at midnight (Fig. 6). Precipitation events of 1  $\text{mm}\cdot 10 \text{ min}^{-1}$  and greater have slightly lower predictability.

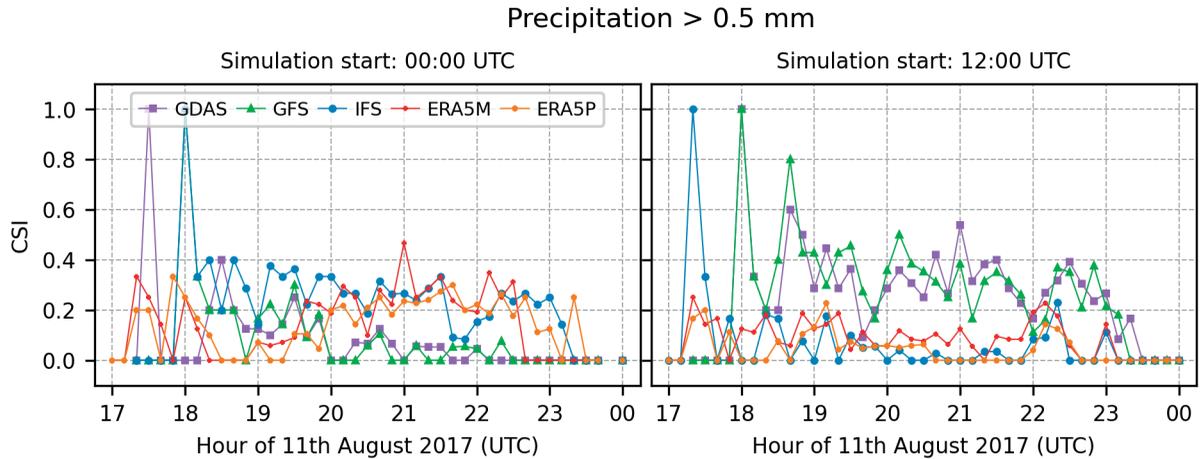


Fig. 5. Critical success index (*CSI*) of precipitation  $>0.5 \text{ mm}\cdot 10 \text{ min}^{-1}$ .

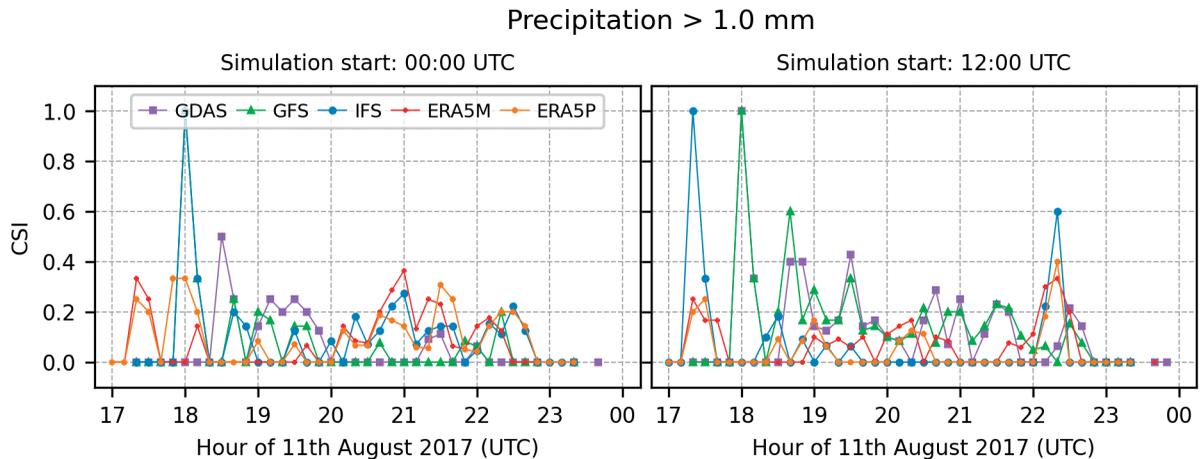


Fig. 6. Critical success index (*CSI*) of precipitation  $>1.0 \text{ mm}\cdot 10 \text{ min}^{-1}$ .

The spatial distributions of the *CSI* values for precipitation events  $>0.5$  and  $1 \text{ mm}\cdot\text{min}^{-1}$  are presented in Figures A13 and A14, respectively. The former events are slightly more predictable for IFS- and ERA5-driven simulations initialized at 0000 UTC. On the other hand, GDAS and GFS play the primary role among other models initialized at 1200 UTC. Events  $>1 \text{ mm}\cdot 10 \text{ min}^{-1}$  have worse predictability at particular locations.

#### 4.4. WIND GUSTS

During the subject derecho event, wind gusts were the most dangerous and caused numerous property losses. In this study, the simulated values differ significantly between each case. The largest damages were reported in the vicinity of the synoptic station in Chojnice. The maximum observed wind velocity from meteorological sensors at this station was  $31.2 \text{ m}\cdot\text{s}^{-1}$  at 2050 UTC. In general, all simulations significantly underestimated the maximum wind gust. The calculation based on the ERA5M and ERA5P model initialized at 0000 UTC was close to the actual value amounting to approximately  $19.5 \text{ m}\cdot\text{s}^{-1}$ . However, the simulated time of the maximum gusts was earlier than the actual time, 1830 and 1930 UTC, respectively. Simulations based on IFS (at 0000 UTC) and GDAS (at 0000 and 1200 UTC) were nearest in time to the occurrence of the phenomenon, taking place at 2040, 2020, and 2040 UTC, respectively. However, the predicted values were also underestimated:  $15.6$ ,  $15.5$ , and  $18.1 \text{ m}\cdot\text{s}^{-1}$ , respectively.

In Figure 7, the *CSI* series for wind gusts  $>5 \text{ m}\cdot\text{s}^{-1}$  are presented for models initialized at 0000 UTC. It is seen that the obtained values differ

between the models. Only the GDAS model has slightly better performance between 1900 and 2100 UTC for that exceedance level. Predictions initialized at 1200 UTC have lower predictability (Fig. 8), slightly more than 0.6, but touching the lowest possible level in some periods (GDA, GFS, and ERA5M). The *CSI* values vary much more than for predictions initialized at 0000 UTC. *CSI* for wind gusts  $>10 \text{ m}\cdot\text{s}^{-1}$  is much smaller and varies greatly over time.

The high predictability of wind gust events  $>5 \text{ m}\cdot\text{s}^{-1}$  (Fig. A15) is found for about half of the stations and simulations initialized at 0000 UTC. This is also the case for predictions driven by GDAS, GFS, and IFS initialized at 1200 UTC. Single locations are found to be more predictable in the southern part of the domain for the ERA5-driven computation of the 1200 UTC initialization time. For the  $10 \text{ m}\cdot\text{s}^{-1}$  exceedance level, higher *CSI* values (Fig. A16) are found for only a few stations and mainly for predictions driven by GDAS and GFS for both initialization times. Some single stations with higher *CSI* are also found for the IFS prediction initialized at 0000 UTC.

#### 4.5. REFLECTIVITY

The temporal evolution of the maximum reflectivity field for values higher than 45 dBZ with 30-min steps is presented in Figure 9 against measured values. The simulation of maximum reflectivity is presented with 1-h steps in Figures A17 and A18 for initialized runs of 0000 UTC and 1200 UTC,

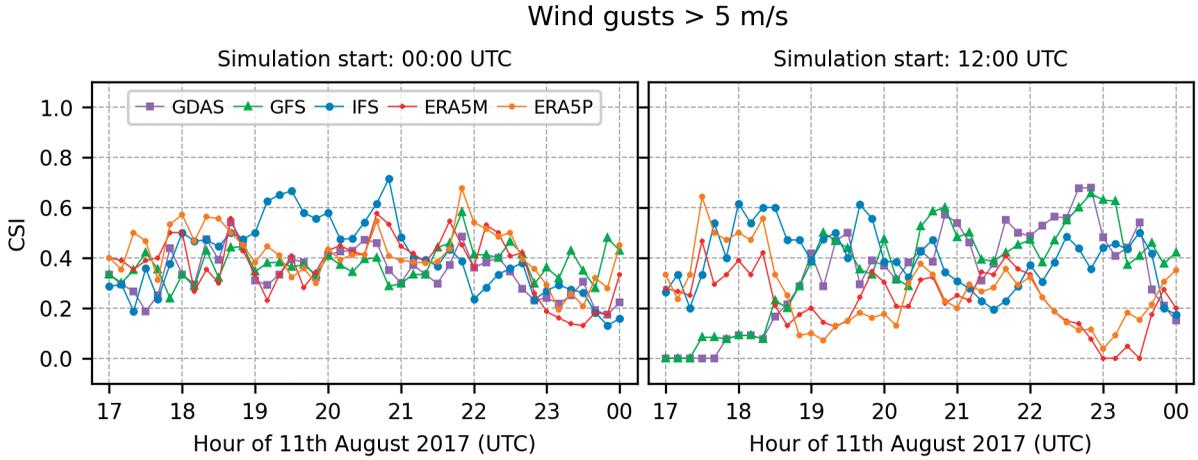


Fig. 7. Critical success index (CSI) of wind gusts >5 m·s<sup>-1</sup>.

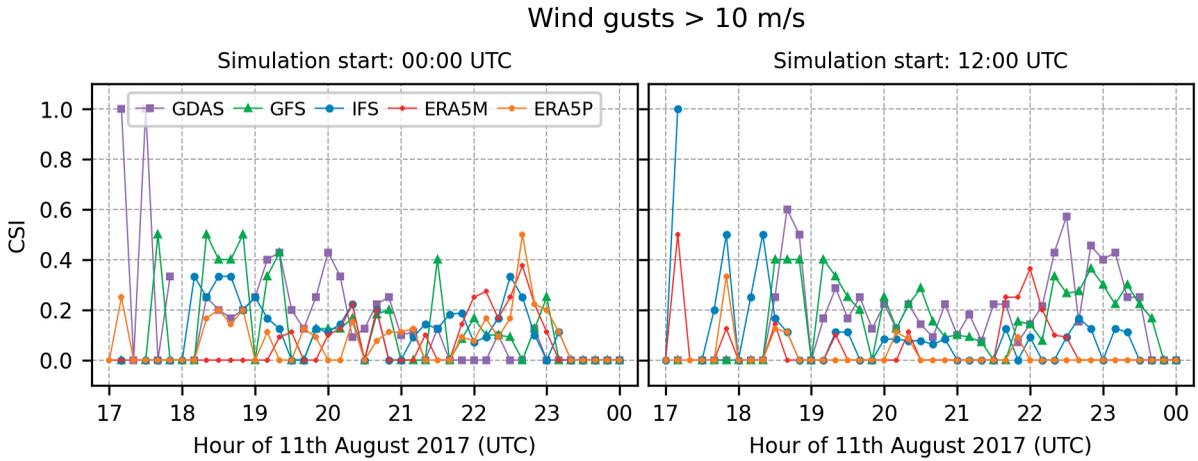


Fig. 8. Critical success index (CSI) of wind gusts >10 m·s<sup>-1</sup>.

respectively. It is clear (Fig. 9) that the modelled track of the convective line match was observed one only for GDAS and GFS predictions initialized at 1200 UTC. On the other hand, the shape and location of the thunderstorm are different from the observed one. Also, the modelled structure is not as sharp as detected by the radar. It has also been found that, in general, maximum reflectivity fields were simulated more westward for models initialized at 0000 UTC than at 1200 UTC. This is in line with the results shown by Tazsarek et al. (2019). This visual inspection is reflected in the Willmott index (Fig. 10). All models, except GDAS and GFS started at 0000 UTC, and IFS started at 1200 UTC are in advance of event. Afterward, the index is slightly increasing to about 0.4 at 2000 UTC. Thereafter, the values reduce to about 0.2 at about 2130 UTC and increase to about 0.5 afterward, mostly for GDAS and GFS runs initialized at 1200 UTC.

#### 4.6. SUMMARY OF THE EVALUATION OF STATISTICAL MODELS

The Taylor diagram for temperature (Fig. 11, left panel) shows that simulations of GDAS and GFS initialized at 1200 UTC performed best. The correlations are the highest, close to 0.9, and *uRMSE* is the lowest and closest to the observed value. It is also found that predictions initialized at 0000 UTC (indicated by blue markers) have larger deviations and those initialized at 1200 UTC (red marker) had smaller deviations or were nearly equal to the observed values. This pattern also can be

seen in the target diagram (Fig. 11, right panel), where blue markers are placed in the positive region of the X axis, and red markers (except for GDAS with slightly larger  $\sigma_F$  than  $\sigma_O$ ) are placed in the negative region. The mean errors are the lowest for 1200 UTC initialized runs and the largest for 0000 UTC runs that are warm-biased. According to the above diagrams, the ERA5 computations started at 0000 UTC seem to be the second best performing predictions.

The Taylor diagram for wind speed at 10 m (Fig. 12, left panel) shows the best performance of the GDAS and GFS simulations for both initialization times and the IFS initialized at 0000 UTC, but the skill is very low. *uRMSE* is about 3 m·s<sup>-1</sup>. The performance is the lowest for ERA5 predictions initialized at 1200 UTC with a correlation close to zero. All models are positively biased (Fig. 12, right panel) and the lowest mean errors are found for GDAS, GFS and GFS runs initialized at 1200 UTC and IFS initialized at 0000 UTC. The *ME* values differ slightly between models.

$$bias = \frac{POD}{SR} CSI = \frac{1}{\frac{1}{SR} + \frac{1}{POD} - 1} \quad (13)$$

In Figure 13, Roebber's performance diagram is presented for precipitation events >0.5 mm (left panel) and 1 mm (right panel) per 10-minutes. According to this figure, the best skill was found for the GDAS and GFS models initialized at 1200 UTC, but the *POD* and *SR* values

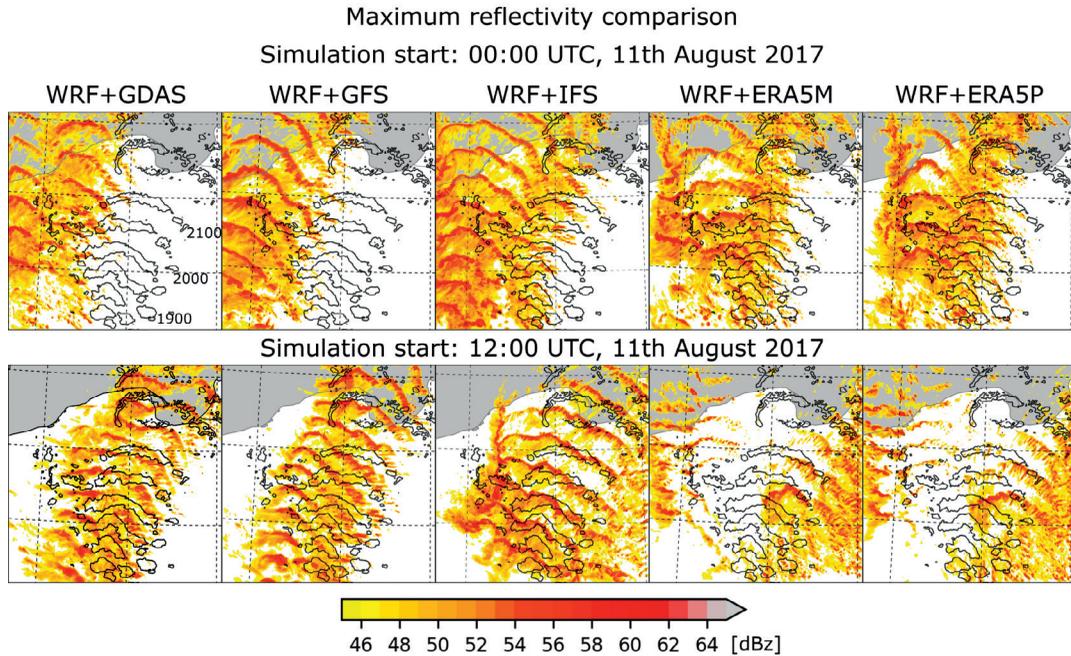


Fig. 9. Maximum reflectivity derived from the WRF model with various initial conditions and from meteorological radar. Data presented for hours between 1830 and 2300 UTC in 30-minute intervals. The black outlines represent the measured values. Reflectivity was plotted only for values higher than 45 dBz.

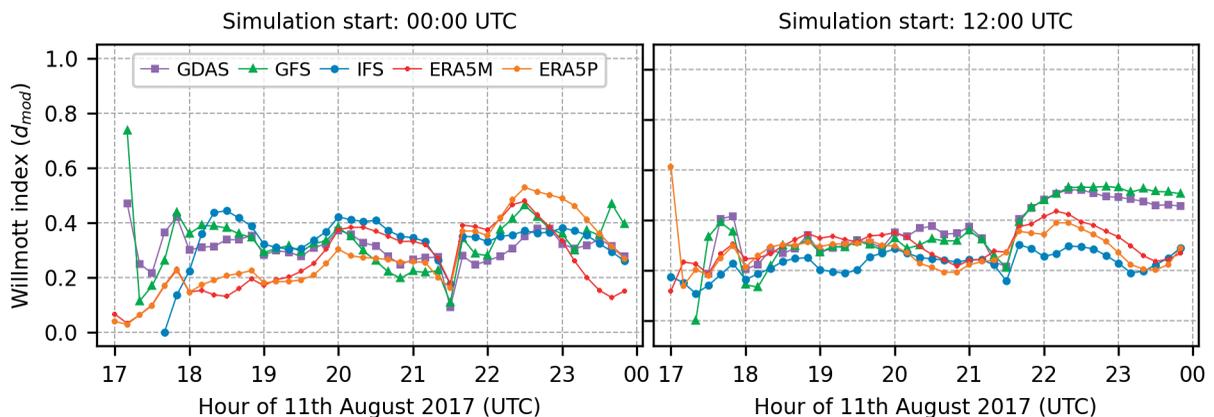


Fig. 10. Comparison of the Willmott index between WRF models with various initial and boundary conditions. The index was calculated on the basis of maximum reflectivity fields derived from the WRF model and meteorological radar.

do not exceed 0.5 and 0.3 for events  $>0.5$  and  $1 \text{ mm} \cdot 10 \text{ min}^{-1}$  respectively. Unlike these models, which are initialized at 0000 UTC, they suffer from very low prediction skills. The same applies to IFS and ERA5 runs initialized at 1200 UTC.

The performance diagrams for wind gust events are presented in Figure 14. All models, except ERA5 initialized at 1200 UTC, are located near  $CSI = 0.4$  isoline for weaker wind gust events, but GDAS and GFS have the lowest false alarm ratio. Stronger wind gust event predictions are worse, with the highest skill provided by the GDAS and GFS simulations initialized at 1200 UTC. The least skill is found for ERA5 predictions initialized at 1200 UTC and for both events.

## 5. CONCLUSIONS AND DISCUSSION

In this paper, we present a detailed verification of various WRF simulations that differ in initial and boundary conditions. As a case study, we chose the derecho event of 11 August 2017 that, was one of the most in-

tense and devastating hazardous phenomena in recent years in Poland. We performed high-temporal (10 minute) and spatial resolution (0.5 km for the third domain) simulations with initialization at 0000 and 1200 UTC. As boundary and initial conditions, we used a suite of global models. The bow echo feature was clearly seen in GFS- and IFS-based simulations initialized at both times, and GDAS initialized at 1200 UTC. Simulations performed using ERA5 data produce maximum reflectivity fields that are very different from those observed. The characteristic bow echo structure was visible in ERA5 simulations started at 0000 UTC, but the MCS was evolving too fast and only in the western part of the domain. The observed event was also reconstructed by other simulations initialized at that time. Simulations based on ERA5 at 1200 UTC show the field scattered throughout the domain. The closest to observed evolution of the bow echo structure was obtained from the results of GFS- and GDAS-based simulations started at a later initialization time. The results of all simulations were evaluated against surface observational and radar data. The best

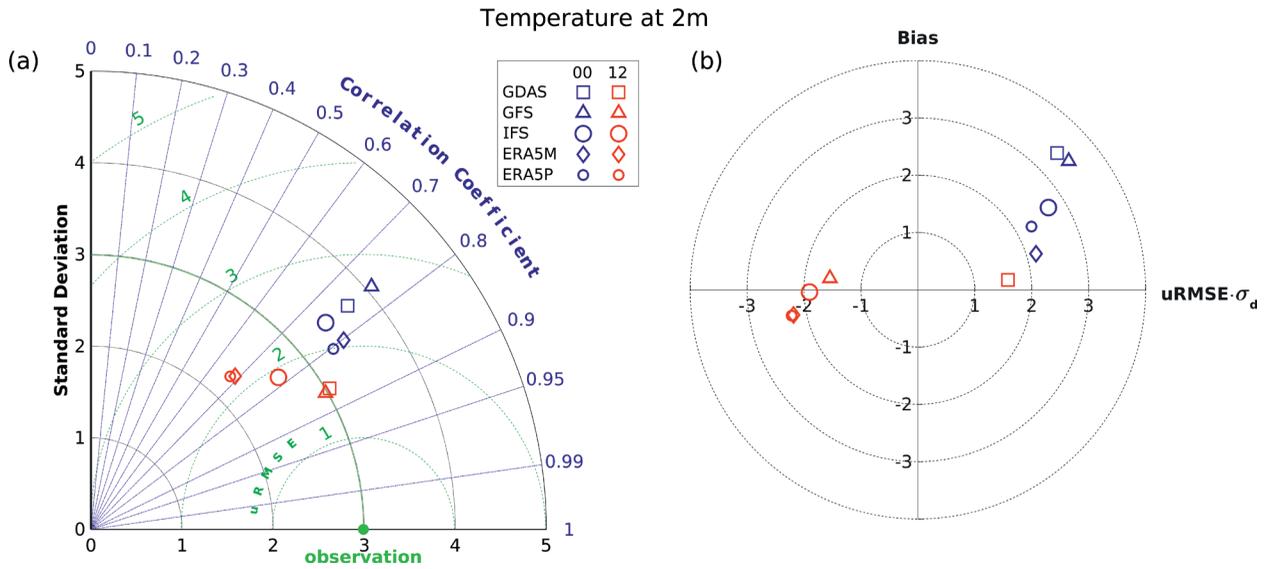


Fig. 11. Taylor (left) and target (right) diagrams showing model-to-observation statistics for temperature at 2 m.

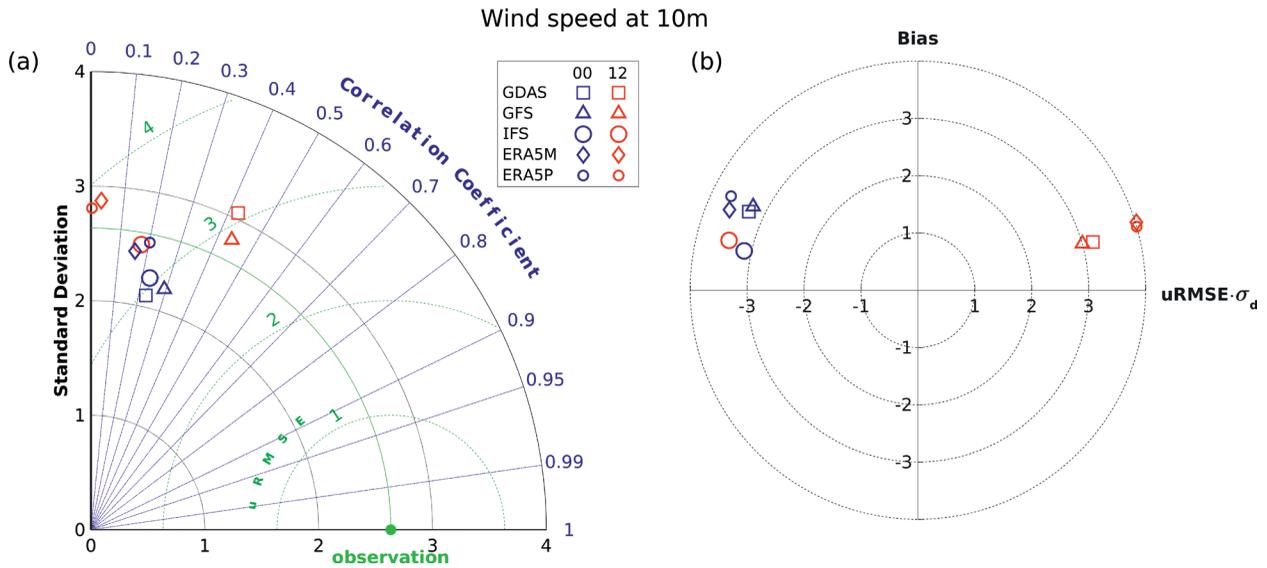


Fig. 12. Taylor (left) and target (right) diagrams showing model-to-observation statistics for wind speed at 10 m.

skill was found for GFS- and GDAS-driven simulations and shorter lead times. For the same time, ERA5-based results were characterized by the worst predictions. Middle-ranking IFS-driven runs show a similar dependence on lead time. This finding can be confirmed by the daily cycle for the *RMSE* signal found by Ylinen et al. (2020) for the temperature forecasts of the IFS ensemble prediction system for Europe. Goutham et al. (2021) found a similar correlation for the surface wind but a smaller *RMSE* for the IFS model. For the 0000 UTC initialization, all models predicted a more westward phenomenon than the actual case. The 1200 UTC simulations look better in this respect, which was also found by Tazarek et al. (2019). Although the GDAS and GFS models were the best predictors of the location of the derecho, all models were characterized by a low probability of detection and a high number of false alarms forecasting extreme precipitation and wind gusts. The study by Gevorgyan (2018) also shows that exact prediction of extreme convective events is a hard challenge and WRF simulations are very sensitive to chosen microphysics and forcing data.

#### DATA AVAILABILITY

Data from all simulations, in NetCDF format, are available on The Bridge of Knowledge platform of Gdansk University of Technology under the CC BY-NC-SA license (Figurski, Nykiel 2020a, b, c, d, e).

#### ACKNOWLEDGMENTS

Calculations were carried out at the Academic Computer Centre in Gdansk.

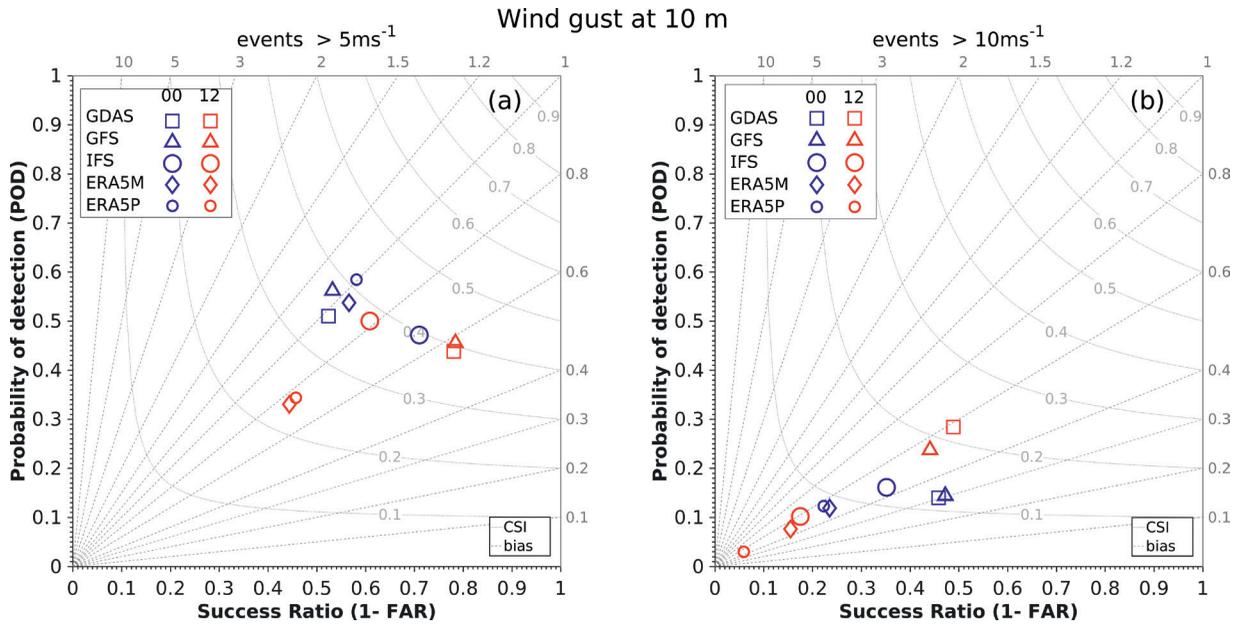


Fig. 13. Roebber performance diagram presenting results of dichotomous verification of precipitation events with a sum of 10 min >0.5 mm (left panel) and >1 mm (right panel).

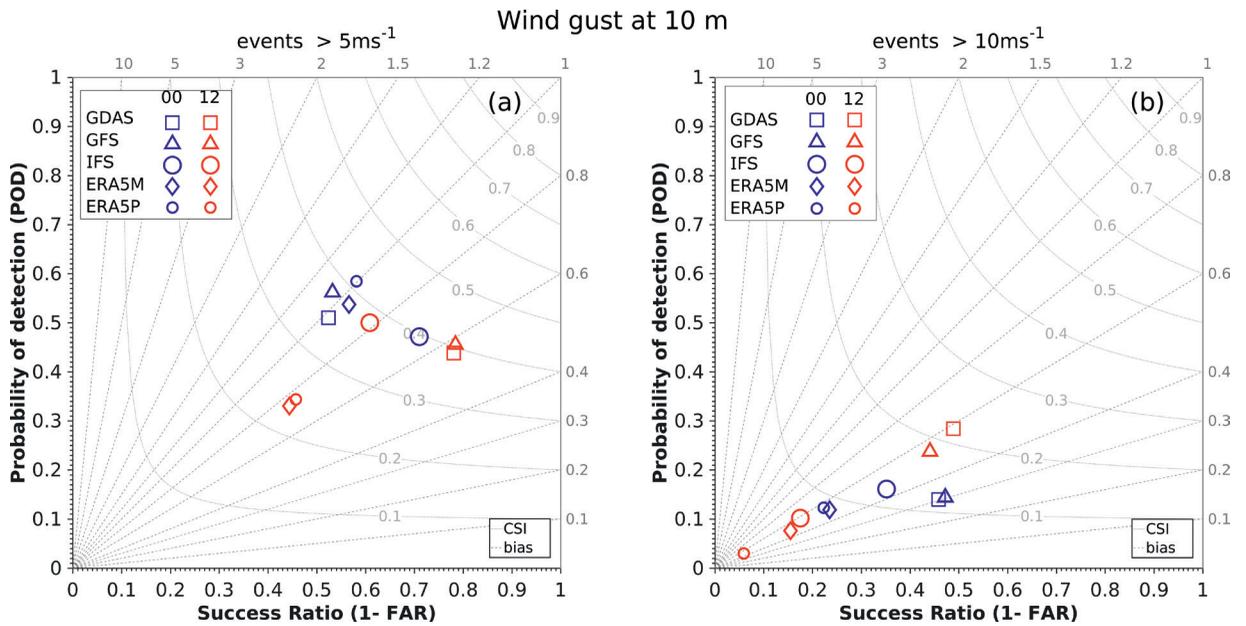


Fig. 14. Roebber's performance diagram presenting results of dichotomous verification of wind gust events with speeds >5 m·s<sup>-1</sup> (left panel) and >10 m·s<sup>-1</sup> (right panel).

## REFERENCES

- Bauer P, Thorpe A, Brunet G, 2015, The quiet revolution of numerical weather prediction, *Nature*, 525, 47-55, DOI: 10.1038/nature14956.
- Bei N, Zhang F, 2007, Impacts of initial condition errors on mesoscale predictability of heavy precipitation along the Mei-Yu front of China, *Quarterly Journal of Royal Meteorological Society*, 133 (622), 83-99, DOI: 10.1002/qj.20.
- Claridge D.E., Chen H., 2006, Missing data estimation for 1-6 h gaps in energy use and weather data using different statistical methods, *International Journal of Energy Research*, 30 (13), 1075-1091, DOI: 10.1002/er.1207.
- Cornwall W., 2016, Efforts to link climate change to severe weather gain ground, *Science*, 351 (6279), 1249-1250, DOI: 10.1126/science.351.6279.1249.
- Daly S.F., Davis R., Ochs E., Pangburn T., 2000, An approach to spatially distributed snow modelling of the Sacramento and San Joaquin basins, California, *Hydrological Processes*, 14 (18), 3257-3271, DOI: 10.1002/1099-1085(20001230)14:18<3257::aid-hyp199>3.0.co;2-z.
- De Meij A., Vinuesa J.F., 2014, Impact of SRTM and Corine Land Cover data on meteorological parameters using WRF, *Atmosphere Research*, 143, 351-370, DOI: 10.1016/j.atmosres.2014.03.004.
- EEA, 2020, Economic losses from climate-related extremes in Europe, CSI 042, CLIM 039, available online at [https://www.eea.europa.eu/ds\\_resolveuid/cf171dde4a5846f99ac-722b92807e9f1](https://www.eea.europa.eu/ds_resolveuid/cf171dde4a5846f99ac-722b92807e9f1) (data access 11.11.2021).
- Figurski M., Nykiel G., 2020a, Simulation of the derecho event in Poland of 11th August 2017 using the WRF model and ECMWF data as initial conditions [Data set], Gdańsk University of Technology, DOI: <https://doi.org/10.34808/kxnc-rc32>.
- Figurski M., Nykiel G., 2020b, Simulation of the derecho event in Poland of 11th August 2017 using the WRF model and ERA5 data on pressure levels as initial conditions [Data set], Gdańsk University of Technology, DOI: <https://doi.org/10.34808/ava3-yd08>.
- Figurski M., Nykiel G., 2020c, Simulation of the derecho event in Poland of 11th August 2017 using the WRF model and ERA5 data on model levels as initial conditions [Data set], Gdańsk University of Technology, DOI: <https://doi.org/10.34808/fkmj-dh09>.
- Figurski M., Nykiel G., 2020d, Simulation of the derecho event in Poland of 11th August 2017 using the WRF model and GFS data as initial conditions [Data set], Gdańsk University of Technology, DOI: <https://doi.org/10.34808/3s00-kn10>.
- Figurski M., Nykiel G., 2020e, Simulation of the derecho event in Poland of 11th August 2017 using the WRF model and GDAS data as initial conditions [Data set], Gdańsk University of Technology, DOI: <https://doi.org/10.34808/ech3-7z70>.
- Gevorgyan A., 2018, Convection-permitting simulation of a heavy rainfall event in Armenia using the WRF model, *Journal of Geophysical Research. Atmospheres*, 123 (19), DOI: 10.1029/2017jd028247.
- Goutham N., Alonzo B., Dupré A., Plougonven R., Doctors R., Liao L., Mougeot M., Fischer A., Drobinski P., 2021, Using machine-learning methods to improve surface wind speed from the outputs of a numerical weather prediction model, *Boundary-Layer Meteorology*, 179, 133-161, DOI: 10.1007/s10546-020-00586-x.
- Grell G.A., Freitas S.R., 2014, A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling, *Atmospheric Chemistry and Physics*, 14 (10), 5233-5250, DOI: 10.5194/acp-14-5233-2014.
- Henn B., Raleigh M.S., Fisher A., Lundquist J.D., 2013, A comparison of methods for filling gaps in hourly near-surface air temperature data, *Journal of Hydrometeorology*, 14 (3), 929-945, DOI: 10.1175/JHM-D-12-027.1.
- Herring S.C., Christidis N., Hoell A., Kossin J.P., Schreck C.J., Stott P.A., 2018, Explaining extreme events of 2016 from a climate perspective, *Bulletin of the American Meteorological Society*, 99 (1), S1-S157, DOI: 10.1175/BAMS-ExplainingExtremeEvents2016.1.
- Hong S.-Y., Limn J.-O.J., 2006, The WRF single-moment 6-class microphysics scheme (WSM6), *Journal of Korean Meteorological Society*, 42 (2), 129-151.
- Hong S.Y., Lim K.S.S., Lee Y.-H., Ha J.C., Kim H.W., Ham S.J., Dudhia J., 2010, Evaluation of the WRF Double-Moment 6-Class microphysics scheme for precipitating convection, *Advances in Meteorology*, 10, DOI: 10.1155/2010/707253.
- Iacono M.J., Delamere J.S., Mlawer E.J., Shephard M.W., Clough S.A., 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), DOI: 10.1029/2008JD009944.
- Jankov I., Gallus Jr. W.A., Segal M., Koch S.E., 2007, Influence of initial conditions on the WRF-ARW model QPF response to physical parameterization changes, *Weather and Forecasting*, 22 (3), 501-519, DOI: 10.1175/WAF998.1.
- Jiménez-Esteve E.B., Udina M., Soler M.R., Pepin N., Miró J.R., 2018, Land use and topography influence in complex terrain area: A high resolution mesoscale modeling study over the Eastern Pyrenees using the WRF model, *Atmospheric Research*, 202, 49-62, DOI: 10.1016/j.atmosres.2017.11.012.
- Jolliff J.K., Kindle J.C., Shulman I., Penta B., Friedrichs M.A.M., Helber R., Arnone R.A., 2009, Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment, *Journal of Marine Systems*, 76 (1), 64-82, DOI: 10.1016/j.jmarsys.2008.05.014.
- Krishnamurthy V., 2019, Predictability of weather and climate, *Earth and Space Science*, 6 (7), 1043-1056, DOI: 10.1029/2019EA000586.
- Kumar P., Kishitawal C.M., Pal P.K., 2017, Impact of ECMWF, NCEP, and NCMRWF global model analysis on the WRF model forecast over Indian Region, *Theoretical and Applied Climatology*, 127, 143-151, DOI: 10.1007/s00704-015-1629-1.
- Lompar M., Lalić B., Dekić L., Petrić M., 2019, Filling gaps in hourly air temperature data using debiased ERA5 data, *Atmosphere*, 10 (1), 13, DOI: 10.3390/atmos10010013.
- Lorenz E.N., 1963, The predictability of hydrodynamic flow, *Transactions of the New York Academy of Sciences*, 25 (4), 409-432, DOI: 10.1111/j.2164-0947.1963.tb01464.x.
- Los F.J., Blaas M., 2010, Complexity, accuracy and practical applicability of different biogeochemical model versions, *Journal of Marine Systems*, 81 (1), 44-74, DOI: 10.1016/j.jmarsys.2009.12.011.
- Nakanishi M., Niino H., 2006, An improved Mellor-Yamada level-3 model: its numerical stability and application to a regional prediction of advection fog, *Boundary-Layer Meteorology*, 119, 397-407, DOI: 10.1007/s10546-005-9030-8.
- Nakanishi M., Niino H., 2009, Development of an improved turbulence closure model for the atmospheric boundary layer, *Journal of the Meteorological Society of Japan. Serie II*, 87 (5), 895-912, DOI: 10.2151/jmsj.87.895.
- Ośródką K., Szturc J., Jurczyk A., 2014, Chain of data quality algorithms for 3-D single-polarization radar reflectivity (RADVOL-QC system), *Meteorological Applications*, 21 (2), 256-270, DOI: 10.1002/met.1323.
- Parodi A., Lagasio M., Maugeri M., Turato B., Gallus W., 2019, Observational and modelling study of a major downburst event in Liguria: The 14 October 2016 case, *Atmosphere*, 10 (12), 788, DOI: 10.3390/atmos10120788.
- Peckham S.E., Smirnova T.G., Benjamin S.G., Brown J.M., Kenyon J.S., 2016, Implementation of a Digital Filter Initialization in the WRF model and its application in the rapid refresh, *Monthly Weather Review*, 144 (1), 99-106, DOI: 10.1175/MWR-D-15-0219.1.
- Pereira H.R., Meschiatti M.C., Pires R.C. de M., Blain G.C., 2018, On the performance of three indices of agreement: an easy-to-use r-code for calculating the Willmott indices, *Bragantia*, 77, 394-403, DOI: 10.1590/1678-4499.2017054.
- Reichler T.J., Roads J.O., 2003, The role of boundary and initial conditions for dynamical seasonal predictability, *Nonlinear Processes in Geophysics*, 10 (3), 211-232, DOI: 10.5194/npg-10-211-2003.
- Roebber P.J., 2009, Visualizing multiple measures of forecast quality, *Weather and Forecasting*, 24 (2), 601-608, DOI: 10.1175/2008WAF2222159.1.
- Siewert J., Kroszczyński K., 2020, GIS Data as a valuable source of information for increasing resolution of the WRF model for Warsaw, *Remote Sensing*, 12 (11), 1881, DOI: 10.3390/rs12111881.
- Skamarock C., Klemp B., Dudhia J., Gill O., Liu Z., Berner J., Wang W., Powers G., Duda G., Barker D., Huang X., 2019, A description of the advanced research WRF model version 4, DOI: 10.5065/1dfh-6p97.
- Sulik S., Kejna M., 2020, The origin and course of severe thunderstorm outbreaks in Poland on 10 and 11 August, 2017, *Bulletin of Geography, Physical Geography Series*, 18 (1), 25-39, DOI: 10.2478/bgeo-2020-0003.

- Sutton C., Hamill T.M., Warner T.T., 2006, Will perturbing soil moisture improve warm-season ensemble forecasts? A proof of concept, *Monthly Weather Review*, 134 (11), 3174-3189, DOI: 10.1175/MWR3248.1.
- Storch H., Zwiers F., 1999, *Statistical analysis in climate research*, Cambridge University Press, DOI: 10.1017/CBO9780511612336.
- Tazarek M., Piłgus N., Orlikowski J., Surowiecki A., Walczakiewicz S., Pilorz W., Piasecki K., Pajurek Ł., Pótrolniczak M., 2019, Derecho evolving from a mesocyclone – a study of 11 August 2017 severe weather outbreak in Poland: Event analysis and high-resolution simulation, *Monthly Weather Review*, 147 (6), 2283-2306, DOI: 10.1175/MWR-D-18-0330.1.
- Taylor K.E., 2001, Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research. Atmospheres*, 106 (D7), 7183-7192, DOI: 10.1029/2000JD900719.
- Wei M., Toth Z., Zhu Y., 2010, Analysis differences and error variance estimates from multi-centre analysis data, *Australian Meteorological and Oceanographic Journal*, 59, 25-34, DOI: 10.22499/2.5901.005.
- Widawski A., Pilorz W., 2018, The mesoscale convective systems with bow echo radar signatures as an example of extremely severe and widespread geohazard in Poland, *Environmental and Socio-Economic Studies*, 6 (1), 10-16, DOI: 10.2478/enviro-2018-0002.
- Willmott C.J., Wicks D.E., 1980, An Empirical method for the spatial interpolation of monthly precipitation within California, *Physical Geography*, 1 (1), 59-73, DOI: 10.1080/02723646.1980.10642189.
- Willmott C.J., 1981, On the validation of models, *Physical Geography*, 2 (2), 184-194, DOI: 10.1080/02723646.1981.10642213.
- Willmott C.J., 1984, On the evaluation of model performance in physical geography, [in:] *Spatial statistics and models*, G.L. Gaile, C.J. Willmott (eds.), Springer, Dordrecht, 443-460, DOI: 10.1007/978-94-017-3048-8\_23.
- Willmott C.J., Robeson S.M., Matsuura K., 2012, A refined index of model performance, *International Journal of Climatology*, 32 (13), 2088-2094, DOI: 10.1002/joc.2419.
- Wrona B., Mańczak P., Wozniak A., Ogrodnik M., Folwarski M., 2022, Synoptic conditions of the derecho storm. Case study of the derecho event over Poland on August 11, 2017, *Meteorology Hydrology and Water Management*, DOI: 10.26491/mhwm/152798.
- Ylinen K., Rätty O., Laine M., 2020, Operational statistical postprocessing of temperature ensemble forecasts with station-specific predictors, *Meteorological Applications*, 27 (6), e1971, DOI: 10.1002/met.1971.
- Zaidi S.M., Gisen J.L.A., 2018, Evaluation of Weather Research and Forecasting (WRF) microphysics single moment class-3 and class-6 in precipitation forecast, [in:] *MATEC Web of Conferences*, 03007, DOI: 10.1051/mateconf/201815003007.

APPENDIX

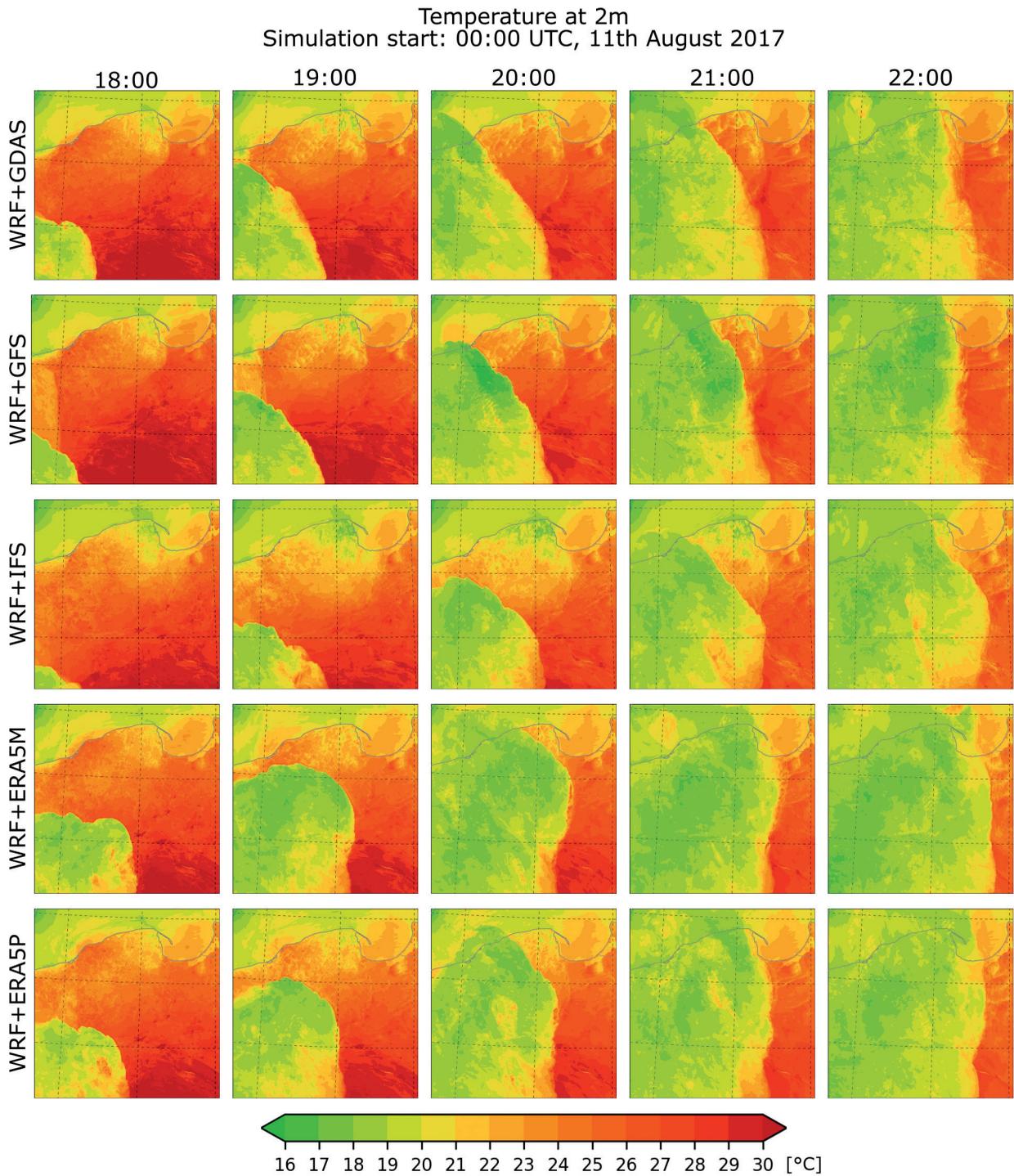


Fig. A1. Temperature at 2 m derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 0000 UTC, 11th August 2017.

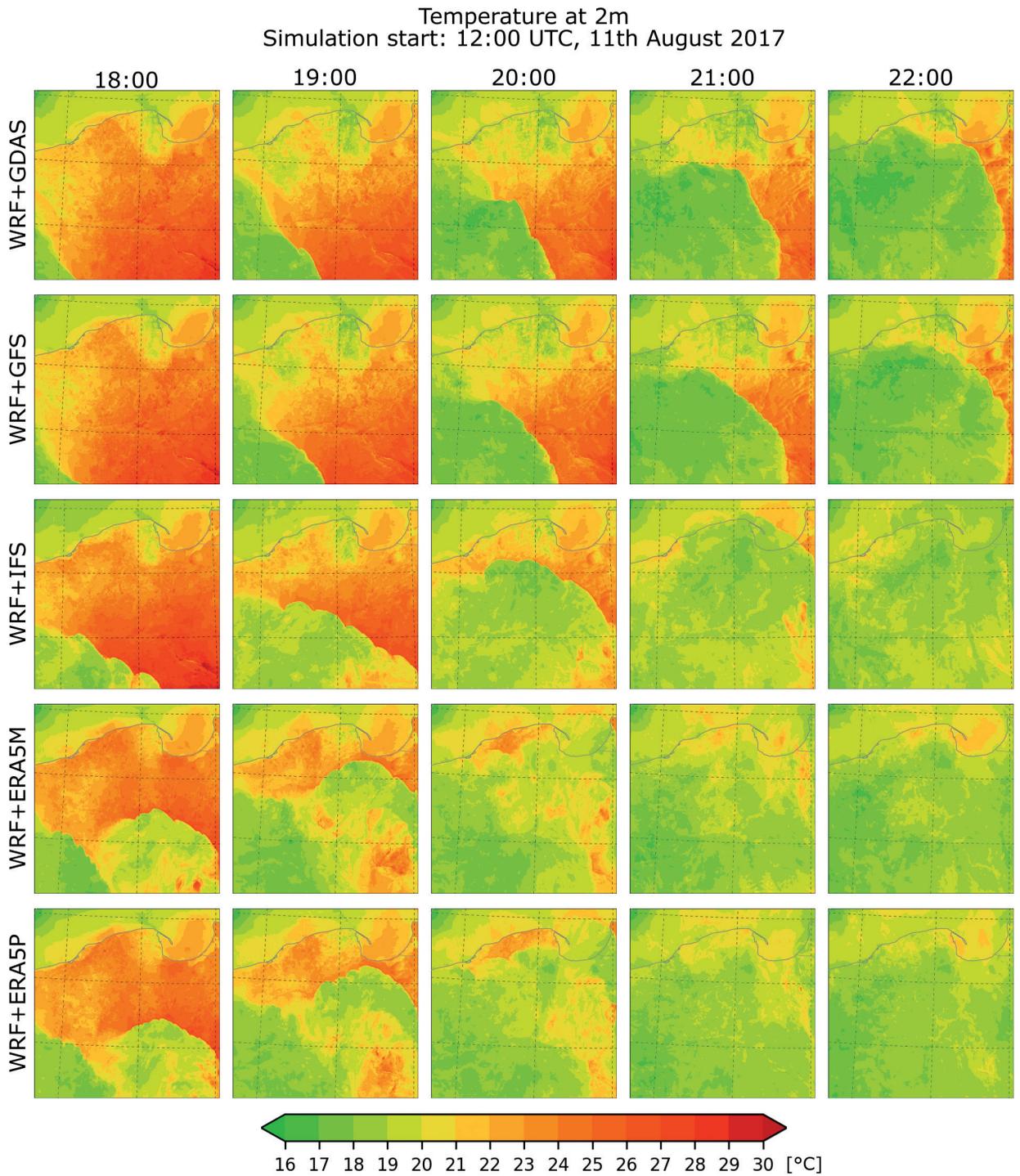


Fig. A2. Temperature at 2 m derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 1200 UTC, 11th August 2017.

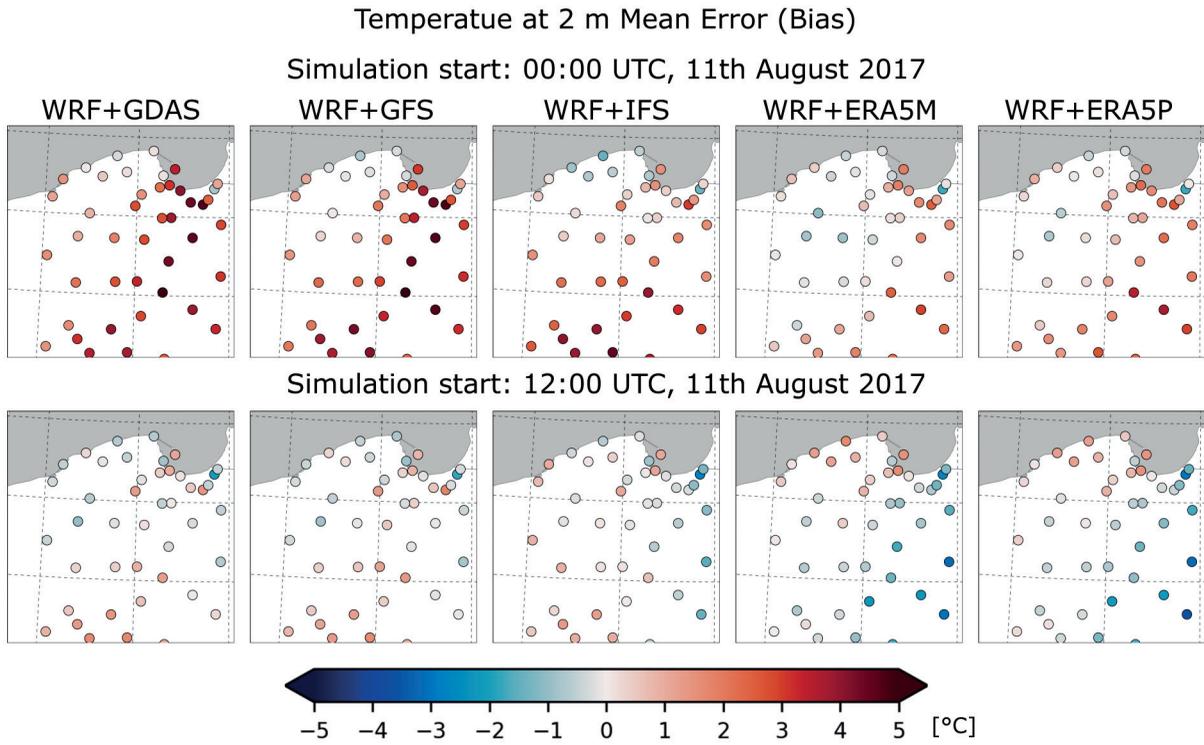


Fig. A3. Mean error (bias) of air temperature (at 2 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

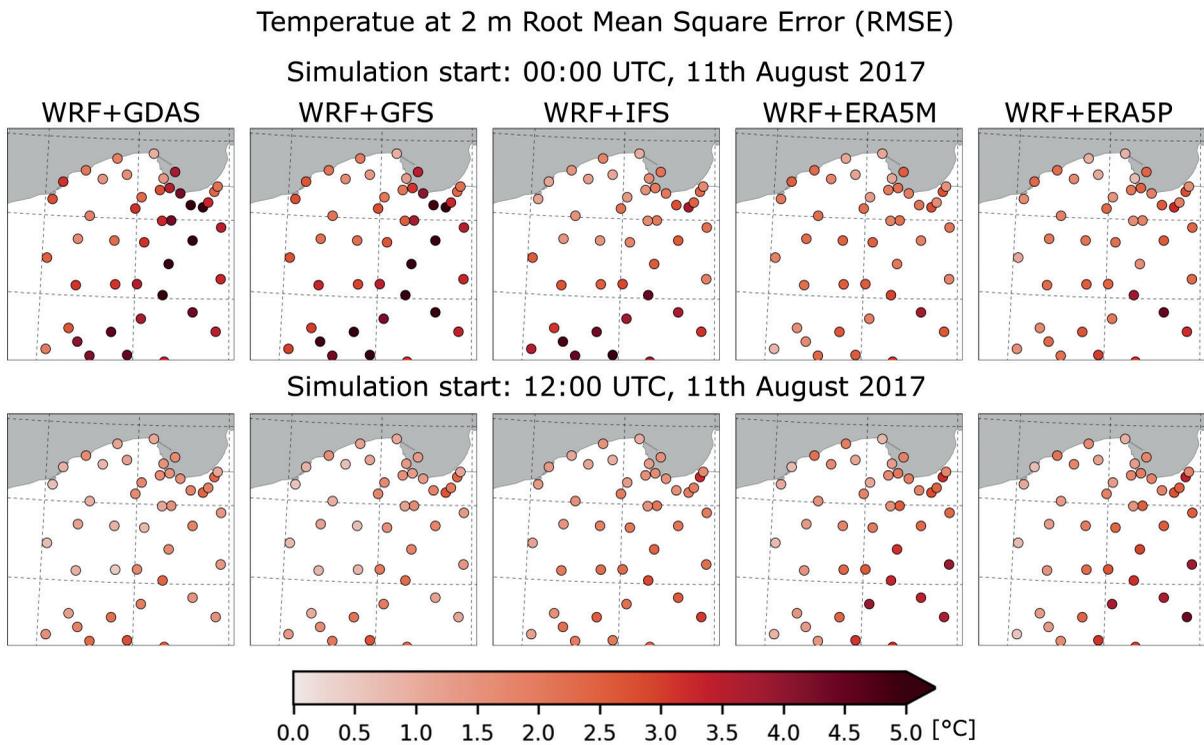


Fig. A4. Root mean square error (RMSE) of air temperature (at 2 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

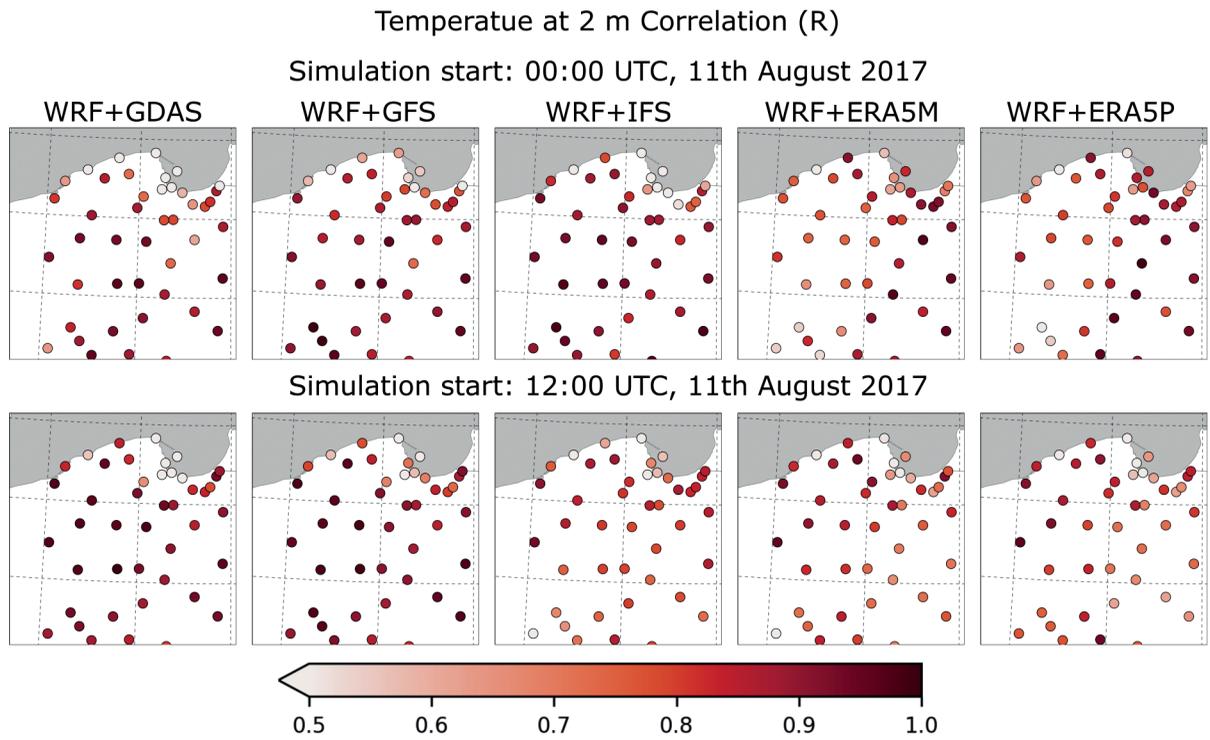


Fig. A5. Correlation of air temperature (at 2 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

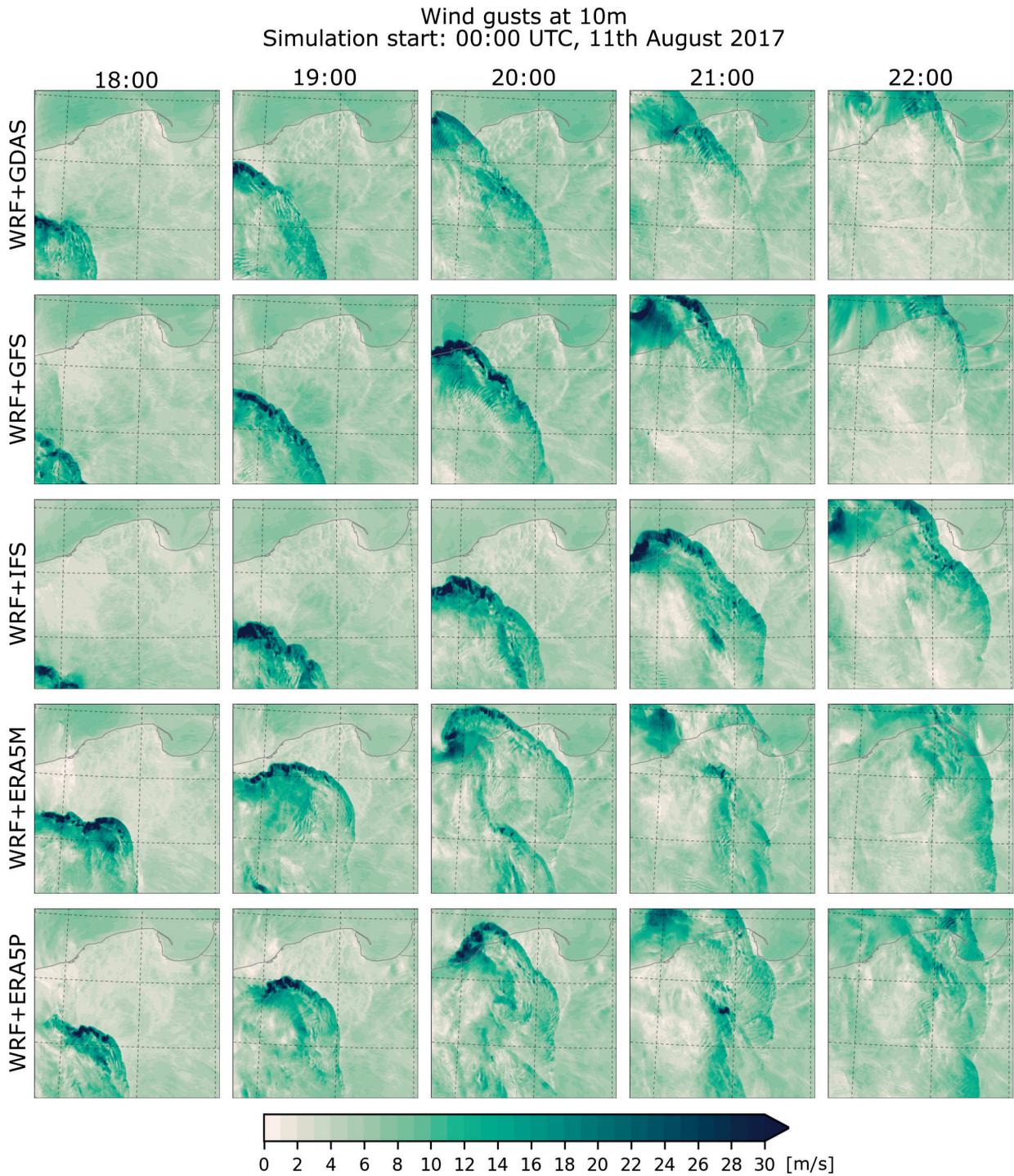


Fig. A6. Wind gusts at 10 m derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 0000 UTC, 11th August 2017.

Wind gusts at 10m  
Simulation start: 12:00 UTC, 11th August 2017

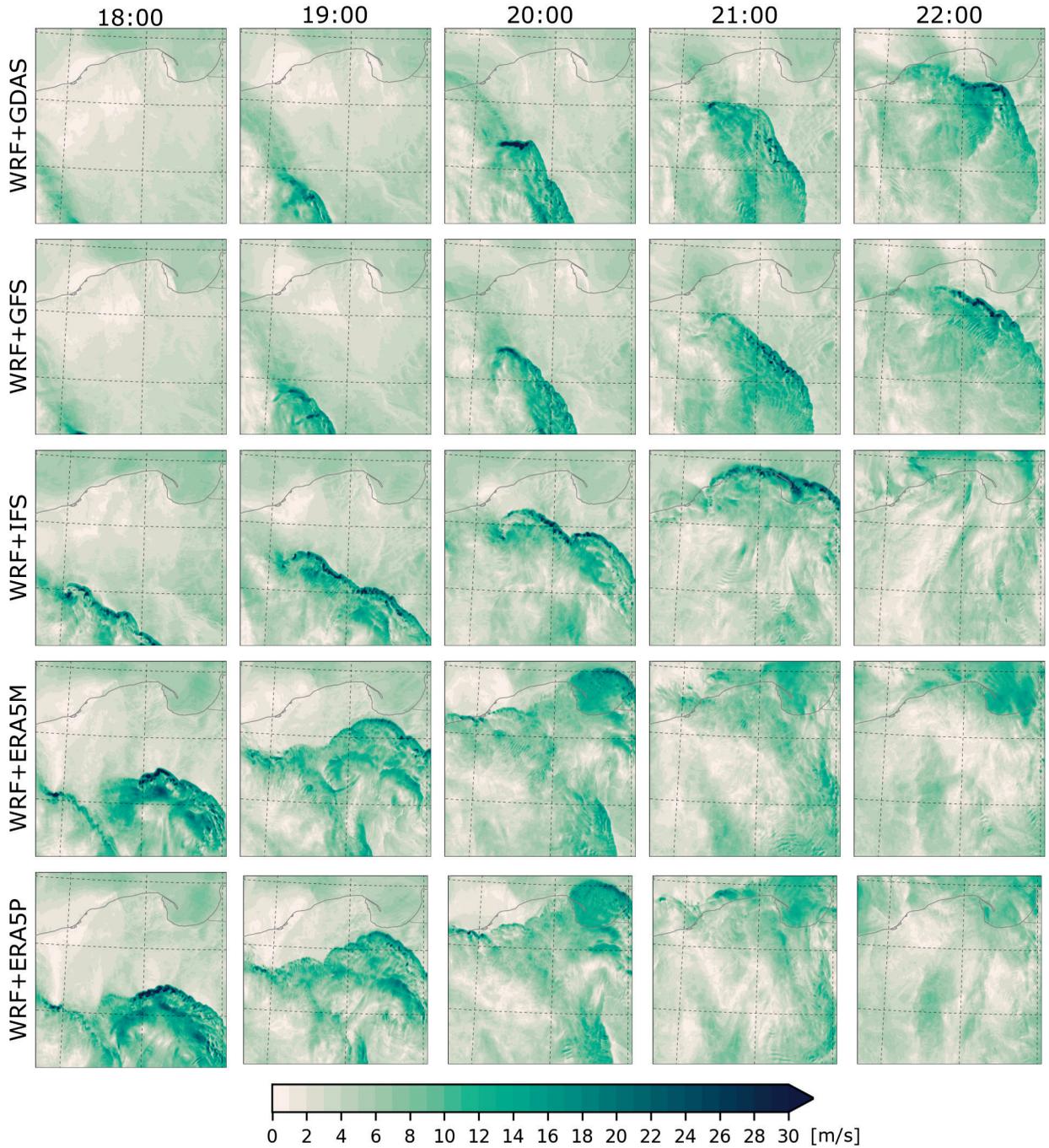


Fig. A7. Wind gusts at 10 m derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 1200 UTC, 11th August 2017.

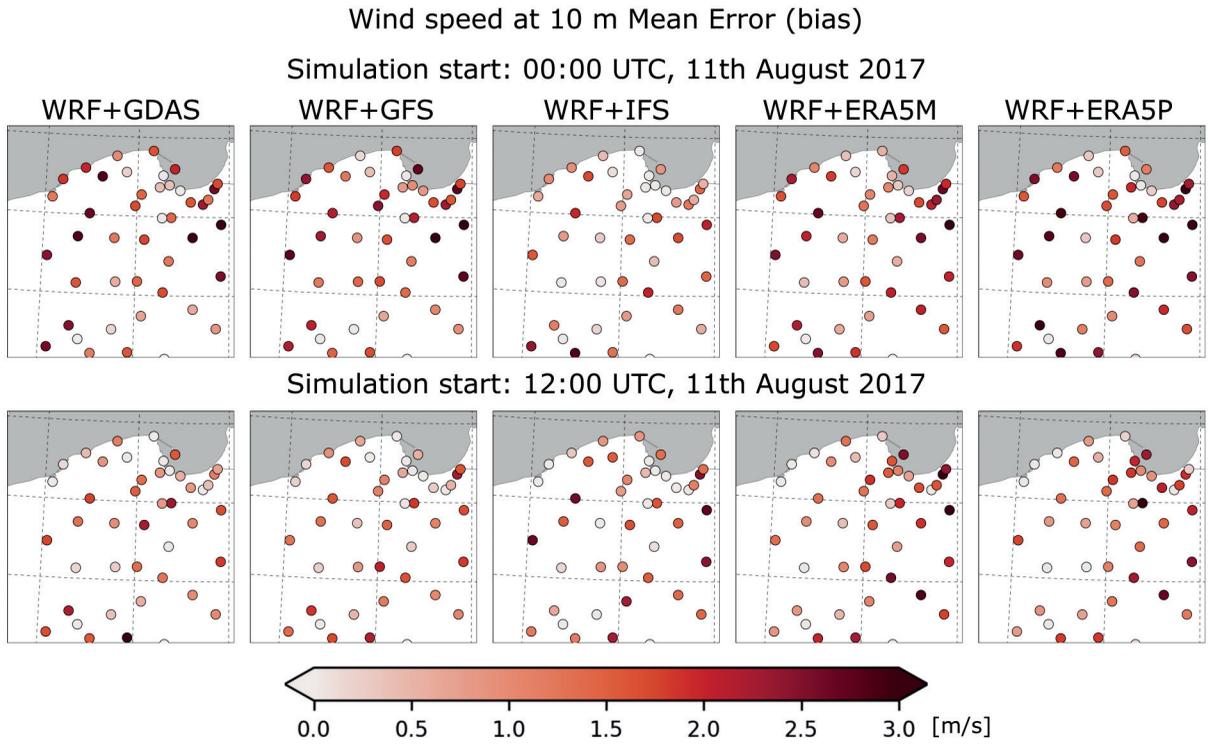


Fig. A8. Mean error (bias) of wind speed (at 10 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

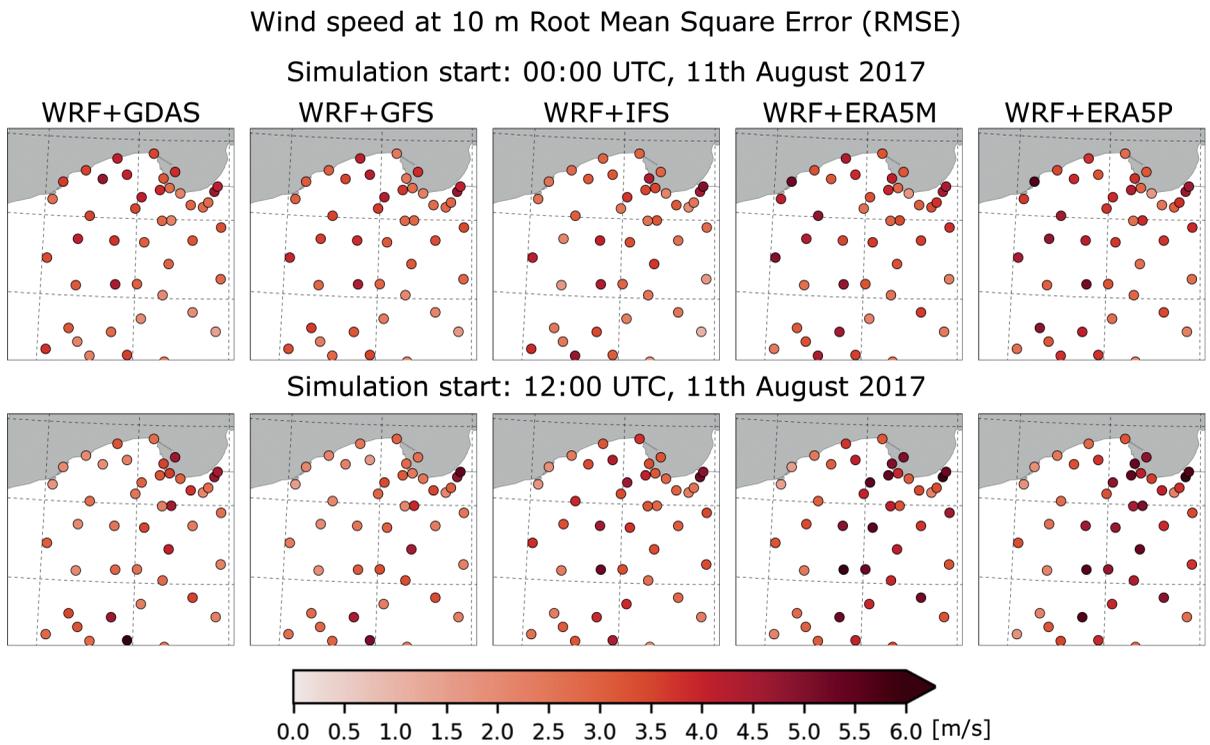


Fig. A9. Root mean square error (RMSE) of wind speed (at 10 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

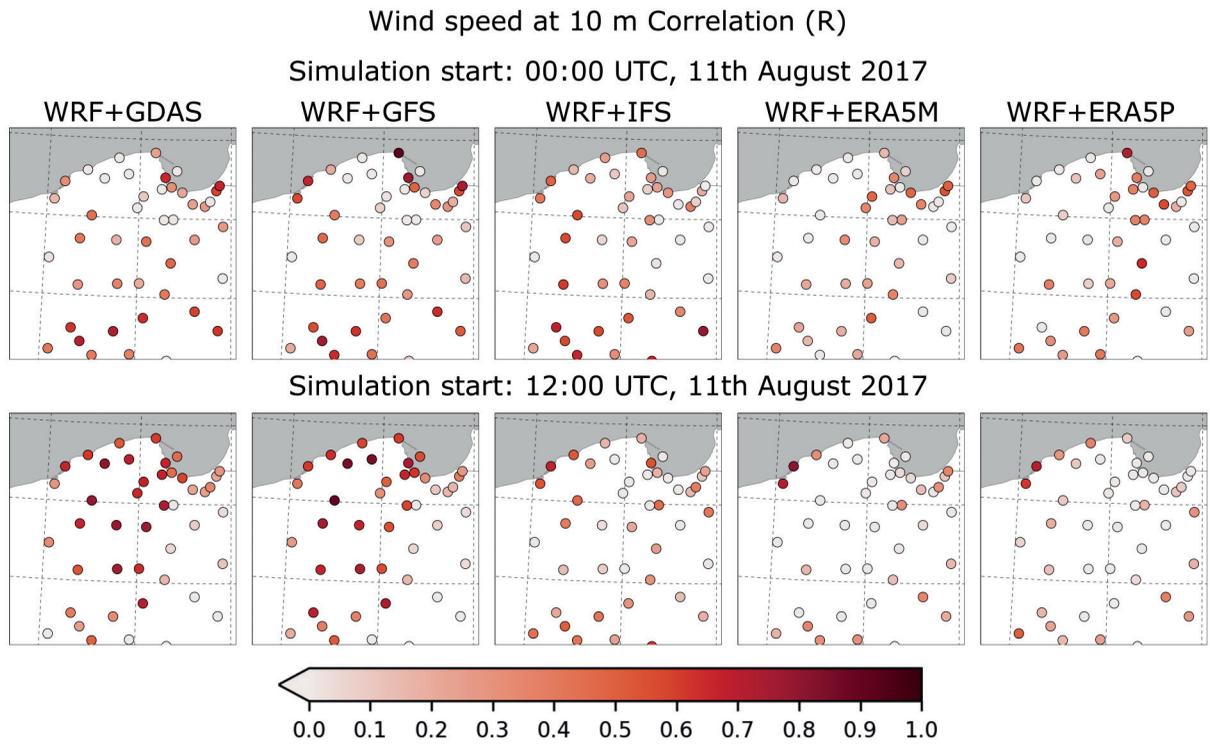


Fig. A10. Correlation of wind speed (at 10 m) derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

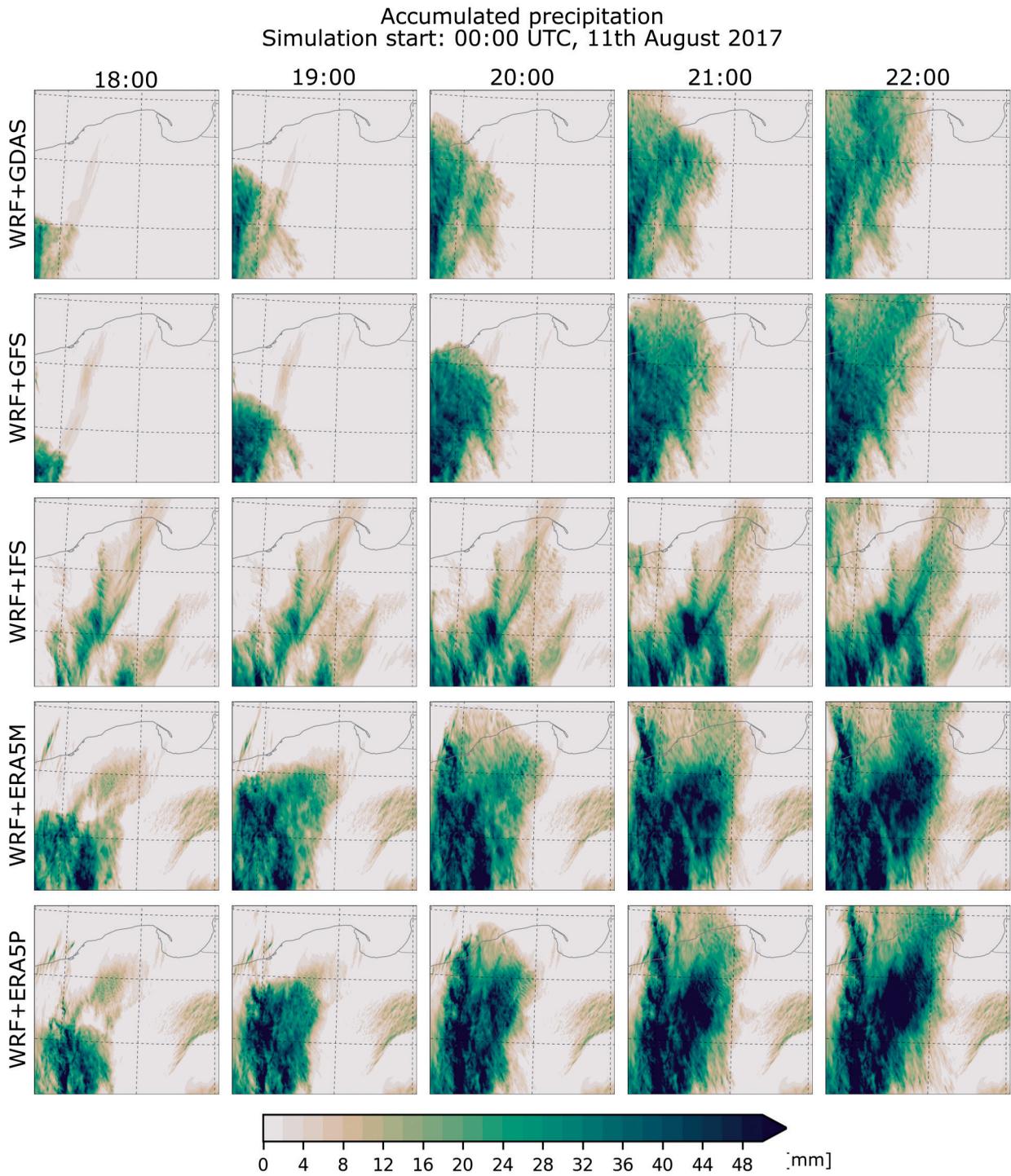


Fig. A11. Accumulated precipitation derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 0000 UTC, 11th August 2017.

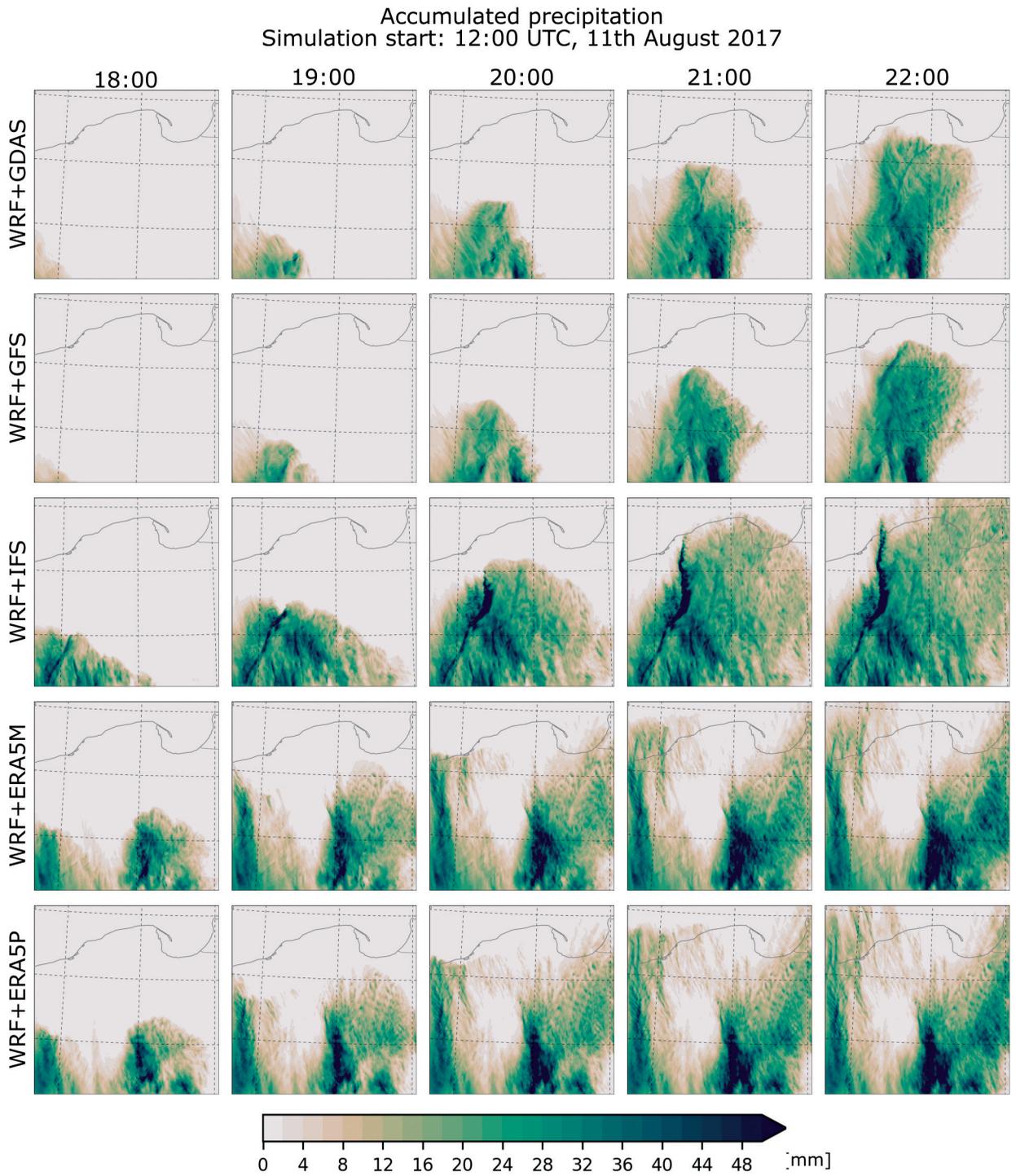


Fig. A12. Accumulated precipitation derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 1200 UTC, 11th August 2017.

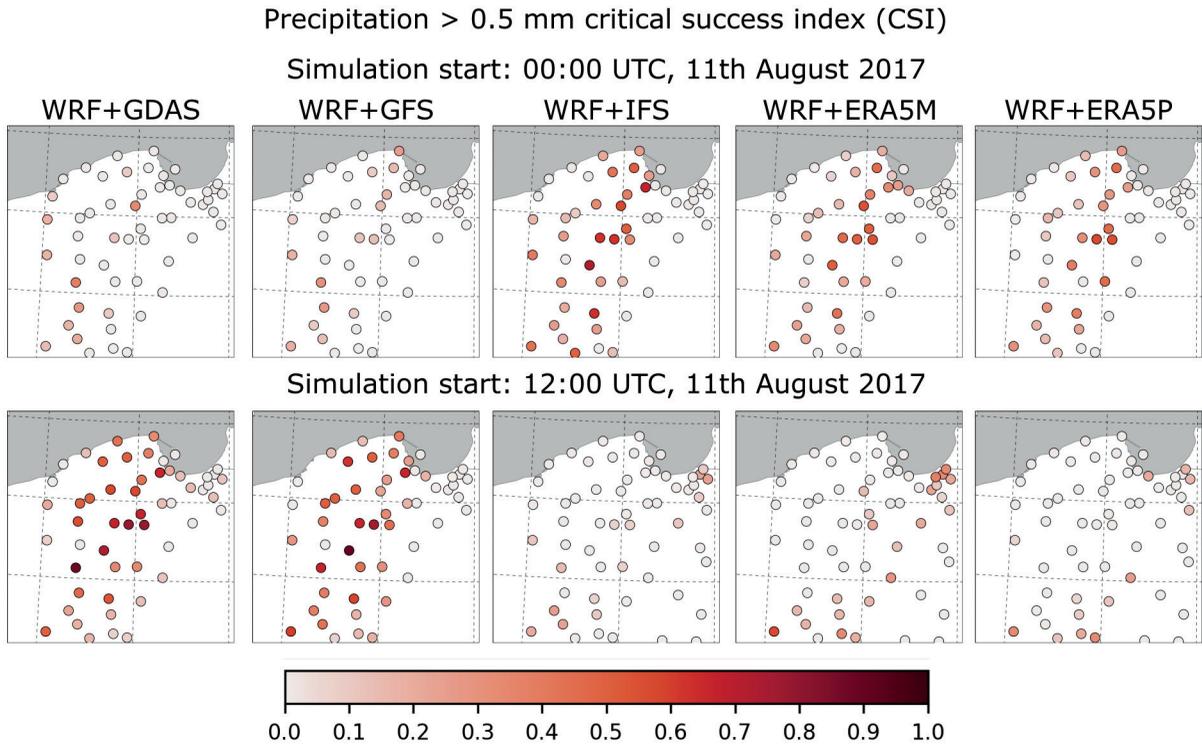


Fig. A13. Critical success index (CSI) of precipitation events > 0.5 mm-10min-1 derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

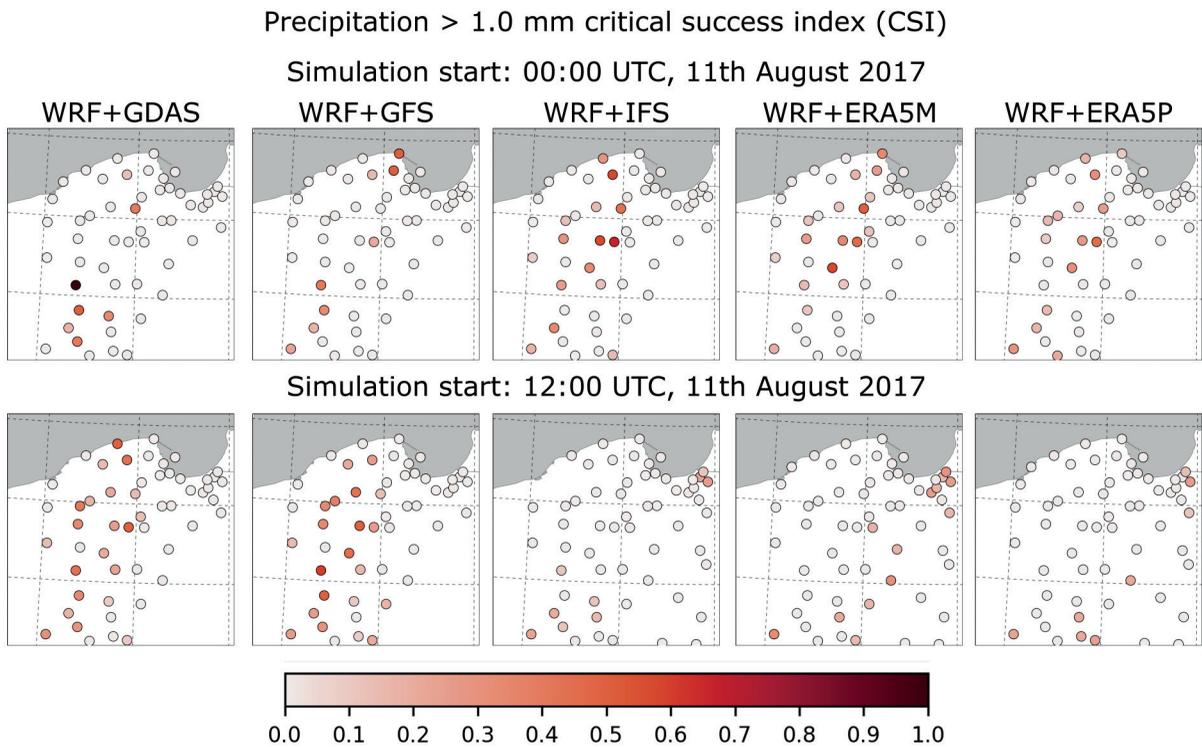


Fig. A14. Critical success index (CSI) of precipitation events > 1 mm-10 min-1 derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

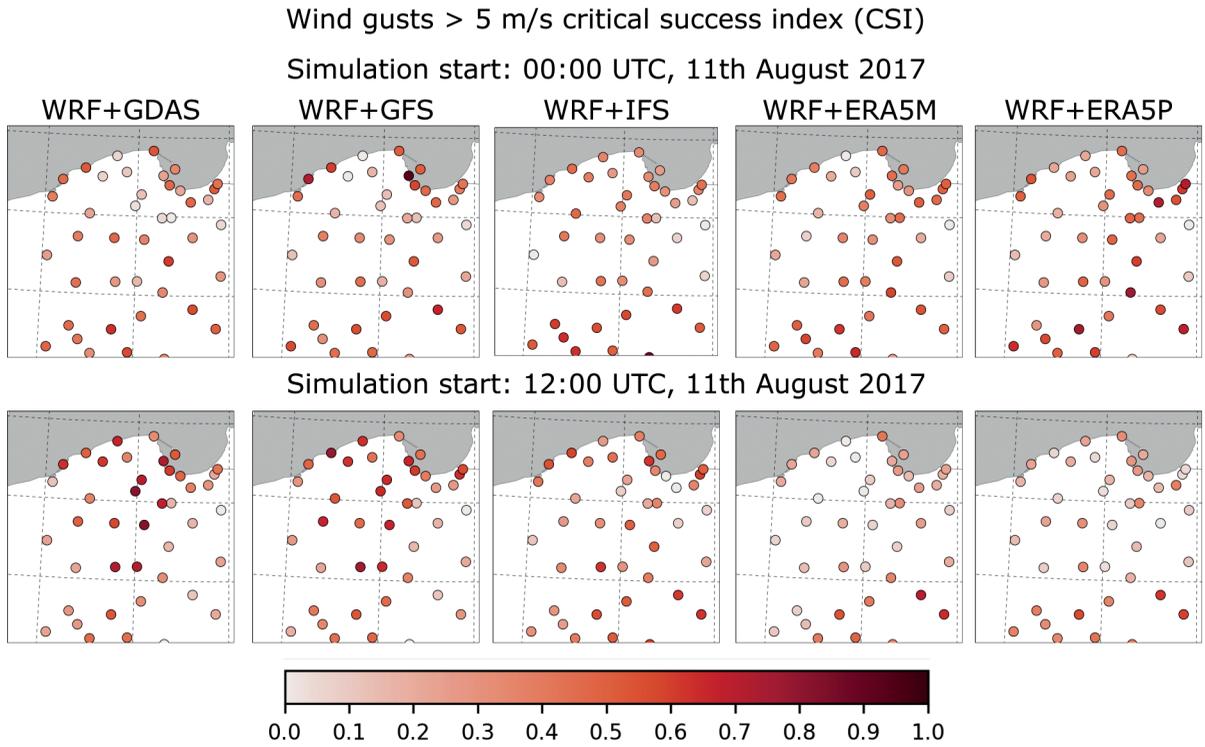


Fig. A15. Critical success index (CSI) of wind gust events > 5 m-s<sup>-1</sup> derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

84

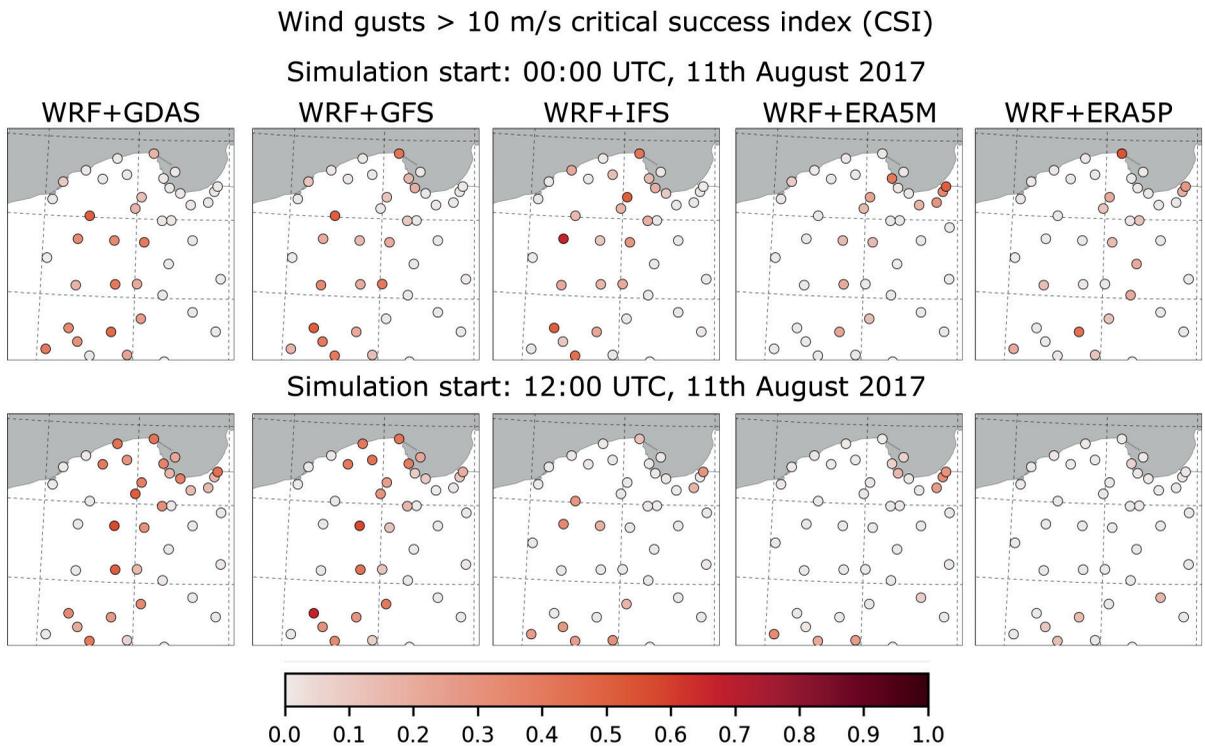


Fig. A16. Critical success index (CSI) of wind gust events > 10 m-s<sup>-1</sup> derived from the WRF model and from in situ observations at meteorological stations. Only data between 1700 and 2400 UTC were analyzed.

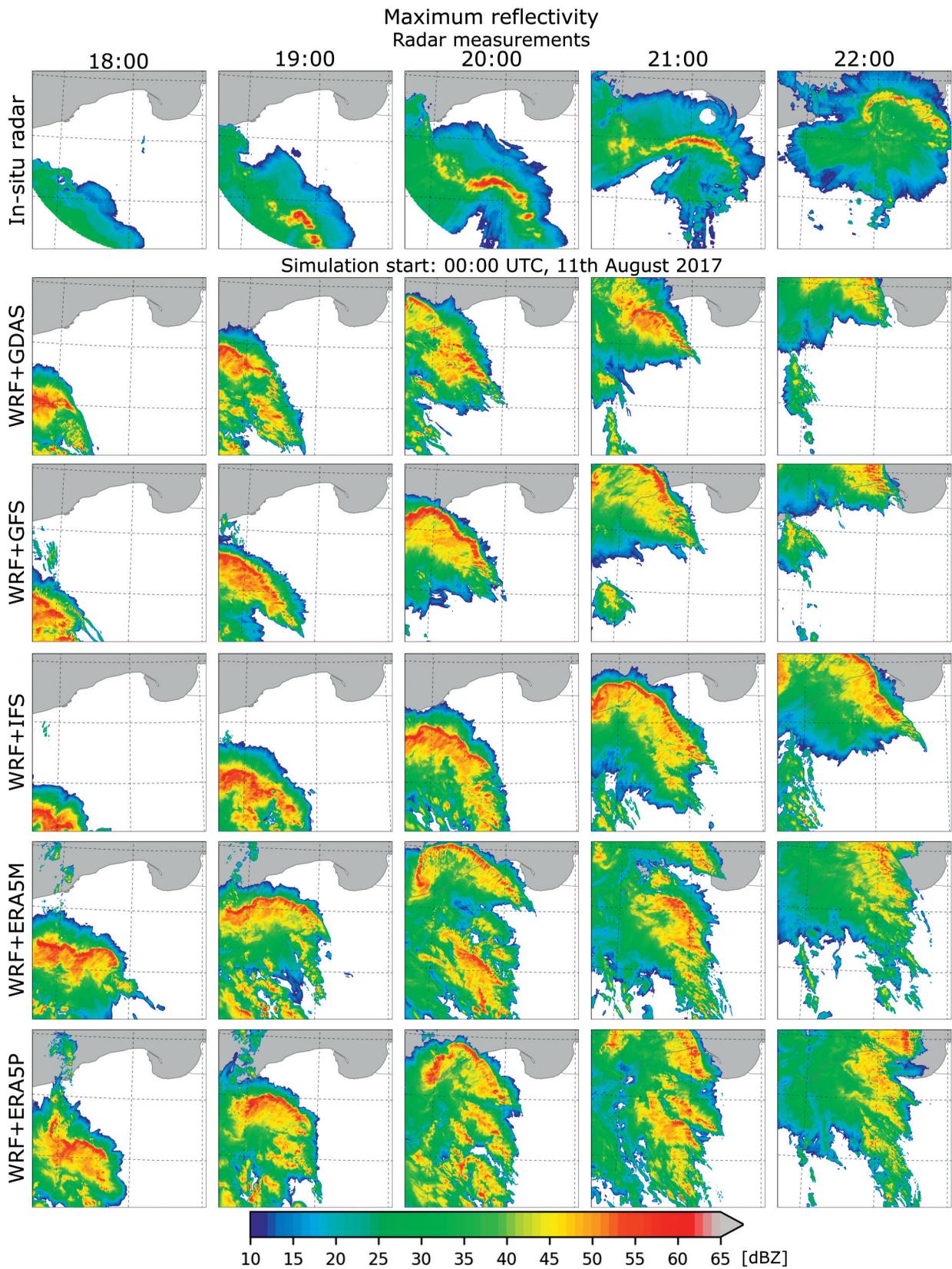


Fig. A17. Maximum reflectivity derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 0000 UTC, 11th August 2017. For comparison, data from meteorological radar are presented at the top.

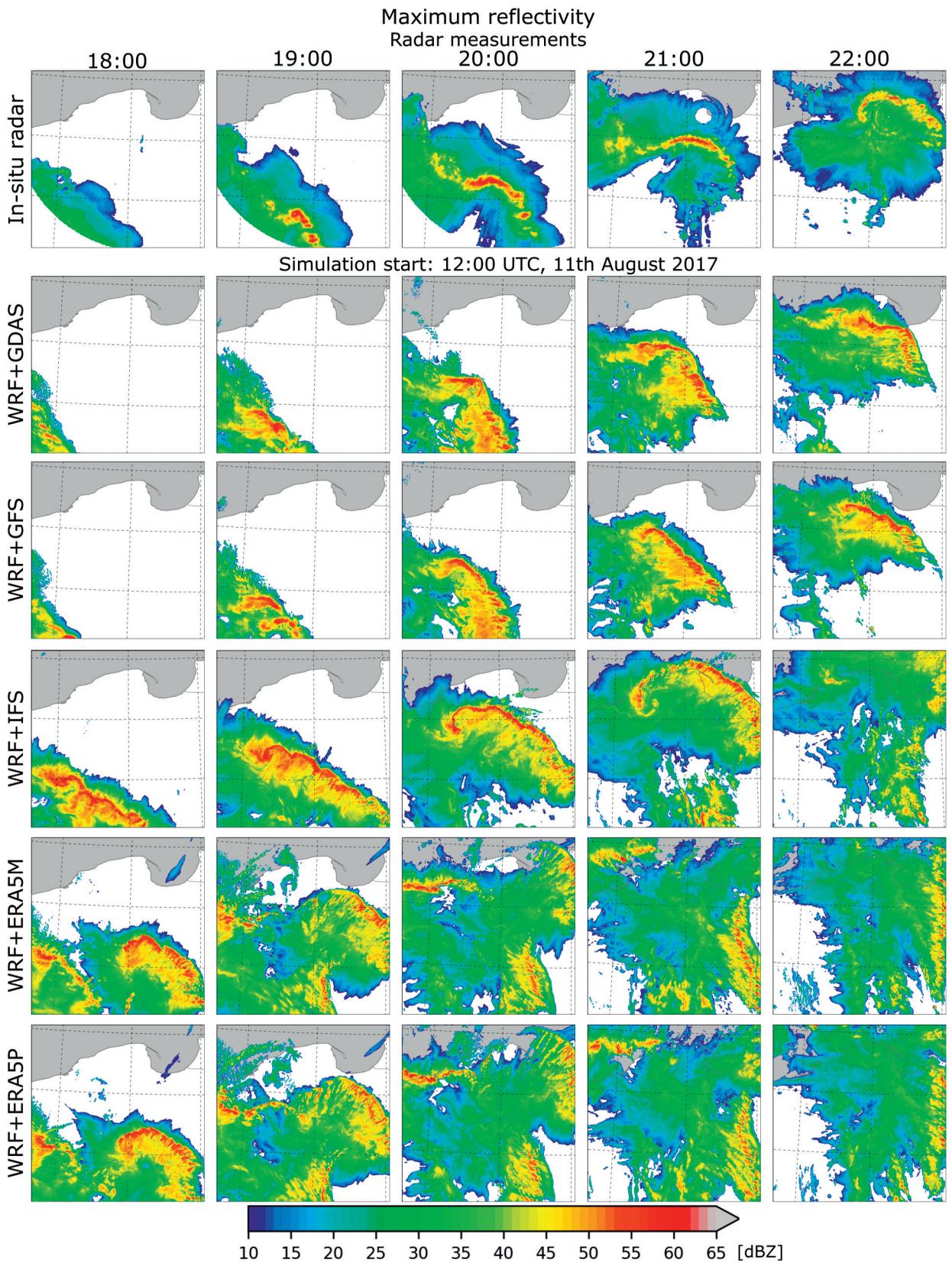


Fig. A18. Maximum reflectivity derived from the WRF between 1800 and 2200 UTC with various initial conditions: GDAS, GFS, IFS, ERA5M (model levels), ERA5P (pressure levels). Forecast start: 1200 UTC, 11th August 2017. For comparison, data from meteorological radar are presented at the top.

