DOI:10.28974/idojaras.2023.2.5

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 127, No. 2, April – June, 2023, pp. 233–251

Graupel mixing ratio forecast from a cloud resolving numerical weather prediction model as a tool for lightning activity prediction

Boryana Tsenova*, **Konstantin Mladenov**, and **Milen Tsankov**

National Institute of Meteorology and Hydrology,
Sofia, Bulgaria

**Corresponding author E-mail: boryana.tsenova@meteo.bg*

(Manuscript received in final form July 15, 2022)

Abstract— Graupel mixing ratio over Bulgaria for the warm half year of 2021 based on the AROME-BG numerical weather prediction (NWP) model, is evaluated and connected lightning data detected by the ATDnet lightning location network. Lightning data and forecasted graupel mixing ratios were considered on resolutions of 5×5 km and 10×10 km with flash rate for one and three hours, as well on a daily base using upscaling neighborhood method. Two daily model runs are considered – at 06 and 18 UTC. Commonly used skill-scores in meteorological forecasts are used as evaluation metrics – probability of detection (POD), false alarm rate (F), proportion correct index (PC), and frequency bias index (FBI). Lightning probability forecast (based on graupel mixing ratio) is evaluated at diurnal, monthly, and spatial bases. Results show that graupel mixing ratio taken from the cloud resolving NWP model AROME-BG could be used as a tool to forecast lightning probability with a relatively high performance. Decreases of forecast spatial resolution and time frequency lead to improvement of forecast probability of detection (POD) and frequency bias index (FBI) and to a slight deterioration of its false alarm rate (F) and its percent correct (PC), and the impact of forecast time frequency is more pronounced.

Key-words: graupel mixing ratio, lightning, AROME, ATDnet

1. Introduction

Laboratory studies by *Reynolds et al.* (1957), *Takahashi* (1978), and *Jayaratne et al.* (1983) indicated that particle charging in thunderstorms is associated with collisions between riming graupel and vapor grown ice crystals. *Bruning et al.* (2007) found that lightning initiated shortly after the detection of graupel in thunderclouds.

Dotzek et al. (2001) classified that the majority of lightning activity occurs in the cloud regions where graupel particles are present, followed by those with snow and hail. The conclusions of *Lund et al.* (2009) were consistent with Dotzek and Bruning, which show lightning initiation in or near the cloud regions that contain graupel. With the rapid development and improvement of the numerical weather prediction (NWP) models, the regional numerical forecast is usually based on cloud-resolving models that predict the atmospheric microphysics for a few days ahead. *McCaul et al.* (2009) proposed approaches for making time- and space- dependent quantitative short-term forecasts of lightning threats using the Weather Research and Forecasting (WRF) model. They considered upward fluxes of precipitating ice hydrometeors in the mixed-phase region at the -15 ºC level and vertically integrated amounts of ice hydrometeors in each model grid column. Their simulations of diverse cases over North Alabama showed that the vertically integrated amounts of ice hydrometeors are promising for depicting the areal coverage of lightning threat. In our study, the non-hydrostatic AROME model is used. AROME-BG is a small scale numerical prediction model, operational at the National Institute of Meteorology and Hydrology of Bulgaria since November 2017. It was designed to improve short range forecasts of severe events such as intense precipitations, severe storms, fog, and urban heat during heat waves. It is developed by Meteo-France in close collaboration with national and international institutes so as to benefit from the latest research in atmospheric modeling. The physical parameterizations of the model come mostly from the research Méso-NH model, whereas the dynamic core is the non-hydrostratic ALADIN one. In the present study, graupel mixing ratio for the warm half year of 2021 based on AROME-BG over Bulgaria is evaluated and connected to the detected lightning by using an upscaling neighborhood method. Lightning data are taken from the ATDnet (Arrival Time Differencing NETwork) lightning location network of the Met Office (*Lee*, 1986; *Gaffard et al*., 2008). The aim of the study is to evaluate the relationships between the graupel mixing ratio forecasted by AROME-BG and the detected lightning by considering the forecast accuracy in time and location for lightning activity prediction.

In the next sections of the paper, a more detailed description of the NWP model version (Section 1.1), the ATDnet (Section 1.2), and the evaluation metrics used for the study (Section 1.3) are presented. Results and summary are in Section 2 and Section 3, respectively followed by the conclusion in Section 4.

1.1. AROME-BG model

AROME is a non-hydrostatic limited area cloud-resolving model, used to improve the short range forecasts of severe events. It was developed by Meteo-France in close collaboration with national and international institutes so as to benefit from the latest research in atmospheric modeling. It uses mostly the physical parameterizations from the Méso-NH model (http://www.aero.obs-mip.fr/mesonh/) and the dynamic core of the ALADIN model (*Termonia et al*.,

2018). The microphysics scheme used in the present study (as in the operational version of the model running over Bulgaria – AROME-BG) is the three-class ice parameterization ICE3 scheme presented in *Pinty* and *Jabouille* (1998). It follows the approach of *Lin et al*. (1983), which is a three-class ice parameterization coupled to a Kessler scheme (*Kessler*, 1969) used for the warm processes. The ICE3 scheme follows the evolution of the mixing ratios of six water species: vapor, cloud and rain drops, pristine ice, snow, and graupel. The concentration of the precipitating particles is parameterized according to *Caniaux et al*. (1994). The pristine ice category is initiated by two heterogeneous nucleation processes: formation of ice embryos in a supersaturated environment over ice (deposition) following *Meyers et al.* (1992) and freezing of supercooled droplets. In the model, the secondary production of ice crystals or rime-splintering mechanism is following *Hallett* and *Mossop* (1974). The homogeneous nucleation of pristine ice starts at temperatures lower than −35 °C. Ice crystals grow by water vapor deposition. The snow phase is initiated by autoconversion of primary ice crystals, and it grows by deposition of water vapor, by aggregation through small crystal collection and by the riming produced by impaction of cloud droplets and of raindrops. Graupel particles are produced by the heavy riming of snow or by rain freezing, when supercooled raindrops come in contact with pristine ice crystals. According to the heat balance equation and the efficiency of their collecting capacity, graupel particles can grow in dry and wet modes (when riming is very intense and the excess of non-freezable liquid water at the surface of the graupel is shed and forms raindrops). At temperatures above 0 °C, ice particles melt into cloud and rain drops. The cloud droplet autoconversion, accretion, and rain evaporation follow the Kessler scheme. The diameter spectrum of each water species is assumed to follow a generalized gamma distribution. Power-law relationships are used to link the mass and the terminal speed velocity to the particle diameters. Microphysics prognostic variables are advected by the semi-Lagrangian scheme. They act on inertia and gravity terms in the momentum equation and with their thermal inertia in the thermodynamical computations. In addition, ICE3 has been upgraded by a subgrid condensation scheme (*Bougeault*, 1982; *Bechtold et al*. 1995) and a probability density function (PDF)- based sedimentation scheme (*Bouteloup et al*., 1995).

The operational AROME-BG model configuration (actually based on the model version cy43t2) at NIMH is the following: the integration domain is covering Bulgaria, with a horizontal resolution of 2.5 km, 60 vertical levels, a time step of 60 s, and a forecast range of 36 h. It runs four times daily – at 00, 06, 12, and 18 UTC – and uses the ALADIN-BG output for initial and boundary conditions. In the present study, the 06 and 18 UTC runs are used. For the aim of the study, the integrated graupel mixing ratio, $r_{g,}$ between model levels 35 and 15 (or between 2756 m and 10306 m) was considered and, as a first approach, cases with $r_g > 0$ were considered as probable lightning occurrence.

1.2. ATDnet

In the present study, lightning data are based on data from the ATDnet (Arrival Time Differencing NETwork) over the territory of Bulgaria for the period April-September 2021. The ATDnet is the most recent version of the VLF (very low frequency) lightning location network of the Met Office that operates since 1987 (*Lee*, 1986, *Gaffard et al*., 2008). It takes advantage of the long propagation paths of the VLF spherics emitted by lightning discharges, which propagate over the horizon via interactions with the ionosphere. The differences in the arrival times of these strokes at the outstations are used to calculate the lightning's location. The ATDnet predominantly detects spherics created by cloud to ground (CG) strokes, as the energy and polarization of spherics created by CG return strokes mean that they can travel more efficiently in the Earth-ionosphere waveguide, and so are more likely to be detected at longer ranges than the typical inter-/intracloud (IC) discharges (*Anderson* and *Klugmann*, 2014). Data are collected every minute and BUFR encoded using the universal BUFR template for lightning data with 15 minutes of data combined into one file, which is then sent by the UK Met Office on behalf of the World Meteorological Organization to member states through its global telecommunication system.

As mentioned above, the resolution of AROME-BG is 2.5×2.5 km. However, due to the uncertainty of ATDnet accuracy over the studied region (*Fig. 1a*), in the present study, lightning data and forecasted graupel mixing ratios were considered on resolutions of 5×5 km and 10×10 km with flash rate for one and three hours, as well on a daily base using the upscaling neighborhood method. This technique was used to evaluate the accuracy of the forecast for the precise location and time, respectively. Cases with lightning were considered as bins from the corresponding mesh with at least one detected flash. As an example, *Fig. 1b* presents the spatial distribution of lightning density on a grid with a resolution of 5 km for the period April-September 2021 over the considered region (left panel) as well the spatial distribution of the cases with at least one detected flash for the same period with the same spatial resolution (right panel). These last values are used to evaluate the ability of the graupel mixing ratio forecasted by AROME-BG to predict lightning activity. In *Fig. 2,* the diurnal distributions of the cases with at least one flash detected by ATDnet over the considered region on grids with a resolution of 5×5 km and 10×10 km, with frequencies of one and three hours, for the period April-September 2021 are shown. *Figs. 1* and *2* give and idea of the number of cases with lightning (as well their spatial distribution) that were compared with the corresponding forecast for the graupel mixing ratios integrated between the 35 and 15 model levels. From *Fig. 2a* it is visible, that a big majority of cases with lightning is between 09 and 21 UTC which is normal for our latitudes. *Fig. 2b* shows the monthly distribution of cases with lightning.

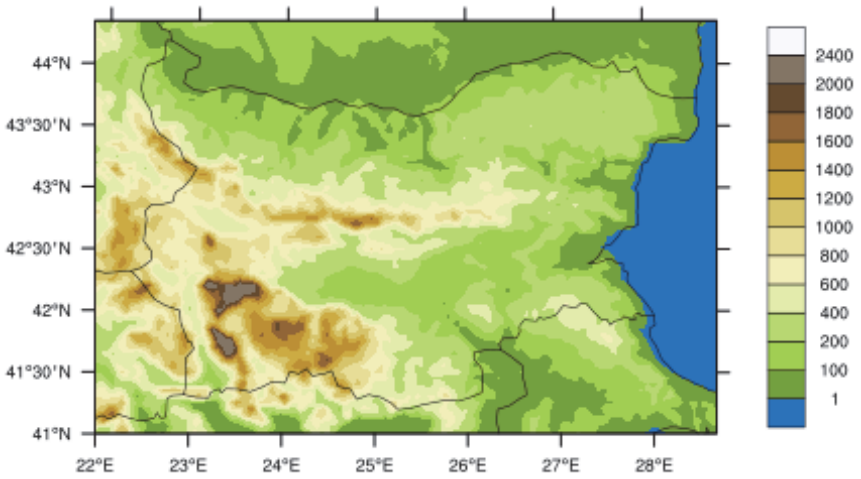

Fig. 1a. Orography of the domain covering Bulgaria and its surroundings, over which the lightning probability forecast performance is evaluated.

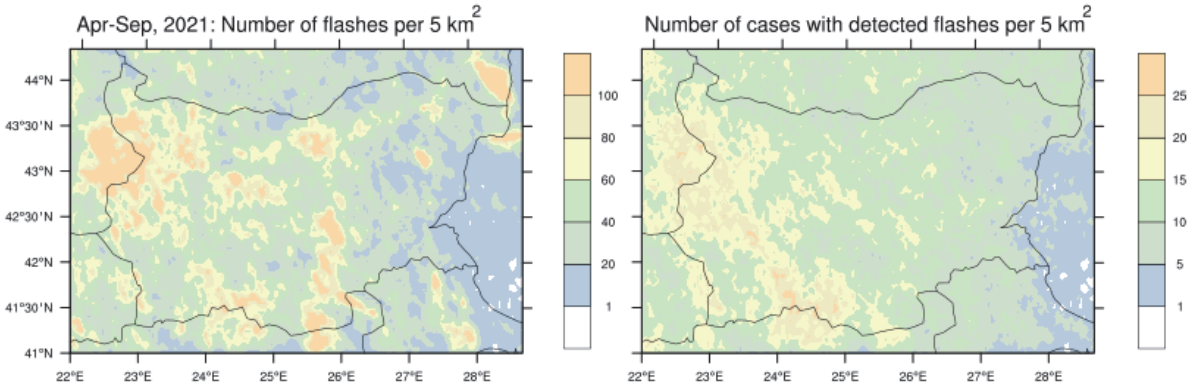

Fig. 1b. Number of flashes per 5x5 km detected between April and September 2021 (left panel) and the corresponding number of cases with at least one detected flash on the grid with a resolution of 5 km (right panel) for the same period.

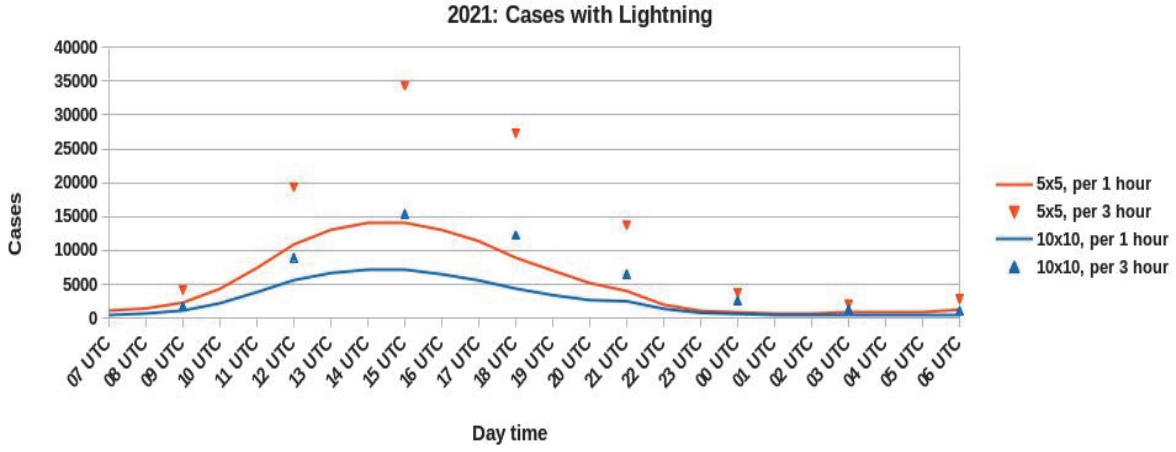

Fig. 2a. Diurnal distributions of the cases with at least one flash detected by ATDnet over the considered region on a grid with a resolution of 5×5 km (red) and 10×10 km (blue), with a frequency of one (line) and three hours (triangle), for the period April-September 2021.

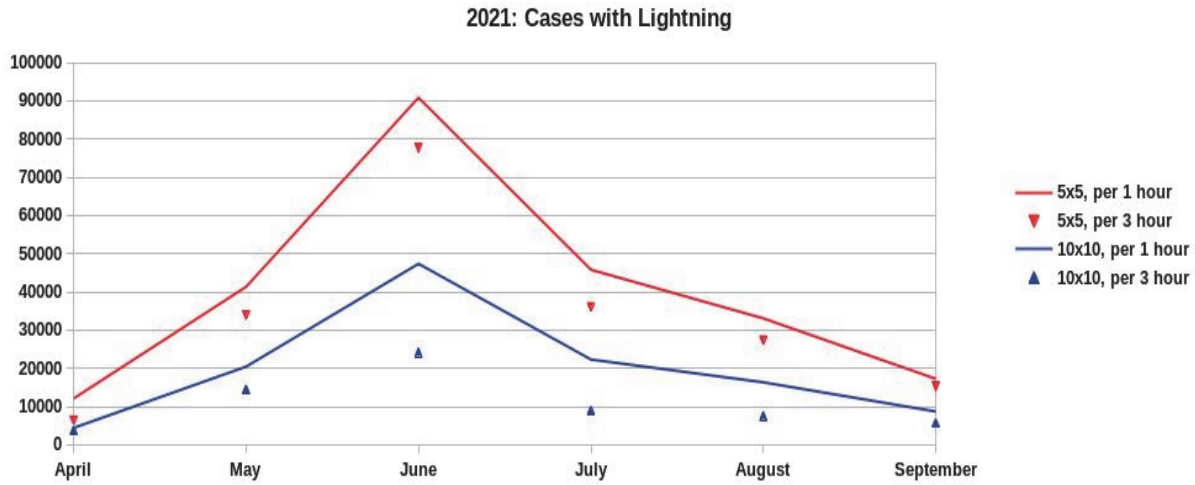

Fig. 2b. Monthly distributions of the cases with at least one flash detected by ATDnet over the considered region on a grid with a resolution of 5×5 km (red) and 10×10 km (blue), with a frequency of one (line) and three hours (triangle), for the period April-September 2021.

1.3. Evaluation metrics

The used here evaluation metrics include four types of commonly used skill-scores in meteorological forecasts: frequency bias index (FBI), proportion correct index (PC), probability of detection (POD) and false alarm rate (F).

Table 1. Parameters used for the evaluation of the forecast skill-scores

	Event observed	
Event forecast	Yes	No
Yes	a	b
No	c	d

Frequency bias index: $FBI = \frac{a+b}{a+c}$; $0 < FBI < \infty$; for a perfect score $FBI = 1$; if $FBI < 1$, there is under forecasting, while if $FBI > 1$ there is an over-forecasting.

Proportion correct index: $PC = \frac{a+b}{a+b+c+d}$; $0 < PC < 1$; for a perfect score $PC = 1$; yes and no forecasts are rewarded equally.

Probability of detection: $POD = \frac{a}{a+c}$; $0 < POD < 1$; for a perfect score $POD = 1$; it is sensitive to misses events and hits only.

False alarm rate: $F = \frac{b}{b+d}$; $0 < F < 1$; for a perfect score $F = 0$; it is sensitive to false alarms .

2. Results

2.1. Diurnal distribution of forecast skill-scores

Figs. 3–6 present the diurnal distributions of the probability detection (POD), false alarm rate (F), proportion correct index (PC), and frequency bias index (FBI), respectively, for the different forecast resolutions and time frequencies for the period between April and September 2021, for the two forecast runs at 06 and 18 UTC. From *Fig. 3* it is visible, that the probability of detection increases with the onset of daylight (when most lightning cases occur) and decreases with the onset of the dark part of the day. Decreasing the spatial resolution leads to a slight increase of POD, while decreasing the time frequency of the forecast leads to a more considerable increasing of POD. Values of POD for the two model runs at 06 and 18 UTC are similar for the respective day hours forecast, except for the three first hours of the 18 UTC run forecast, when the very low POD could be due to the model spin up. Considering daily forecasts of graupel mixing ratio (or forecast on a daily base), values of POD for lightning activity are similar for the two resolutions (5×5 km and 10×10 km) reaching values above 0.9 for daily hours and 0.7 for night hours.

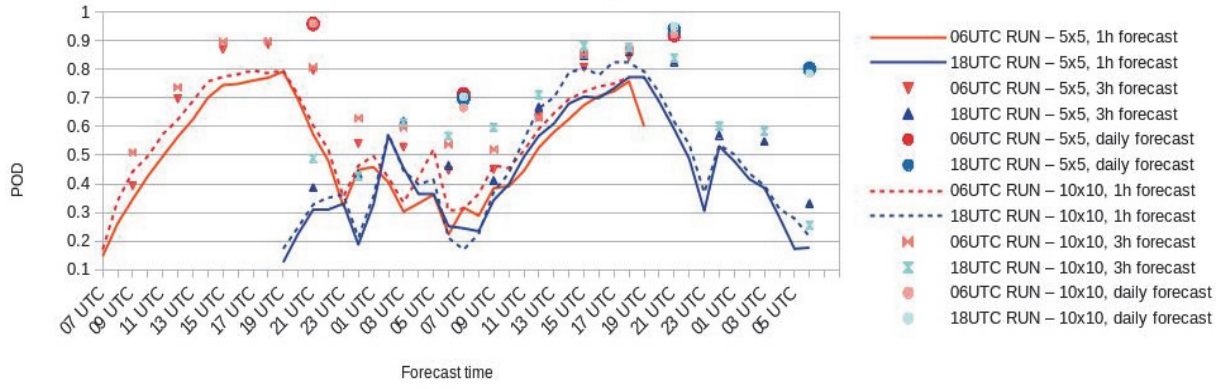

Fig. 3. Diurnal distribution of the probability of detection (POD) for the period April-Sepember 2021.

The decreases of forecast spatial resolution and time frequency may lead to an improvement of forecast probability of detection, but also to a slight deterioration of its false alarm rate (*Fig. 4*) and its proportion correct index (*Fig. 5*), that are slightly worse during the day hours in comparison to night hours. F is below 0.2 for the two resolutions (5×5 km and 10×10 km) and forecast frequencies (1h and 3h) for the night hours lightning activity and between 0.2 and 0.4 for the day hours. On a daily base, the forecasted false alarm rate is around 0.3 for the night and 0.4 for the day. The values of F and PC (*Fig. 5*) for the two

model runs are identical (excluding the spin up time of three hours for the model run at 18 UTC). The proportion correct index is above 0.7 for all forecast considerations except when it is considered on a daily base, where PC <0.6 for the first day forecast on a resolution of 10×10 km.

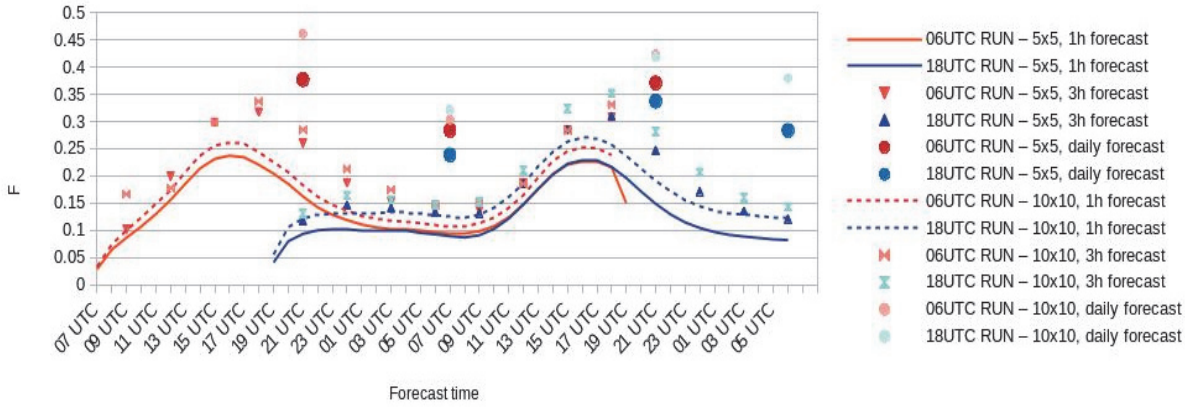

Fig. 4. Diurnal distribution of the false alarm rate (F) for the period April-September 2021.

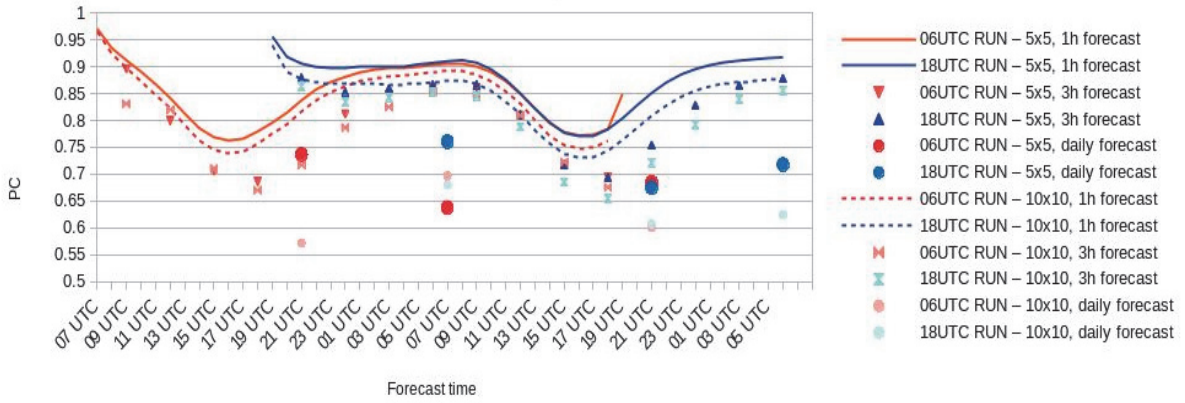

Fig. 5. Diurnal distribution of the proportion correct (PC) for the period April-September 2021.

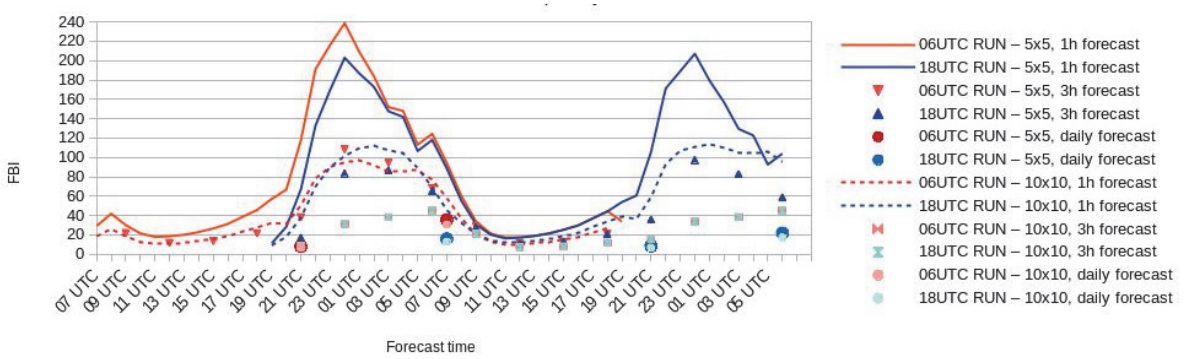

Fig. 6. Diurnal distribution of the frequency bias index (FBI) for the period April-September 2021.

The behavior of the frequency bias index (*Fig. 6*) is similar to behavior of POD with better results for day hours (although still overestimating the lightning activity probability), with improvements due to decreases of forecast resolution and/or its time frequency.

2.2. Monthly distribution of forecast skill-scores

The monthly values of POD, F, PC, and FBI are shown in *Figs. 7–10*. Due to the low number of lightning cases in April 2021 (see *Fig.2b*), for a better view of the results for May-September, April is omitted form these figures. Considering the monthly values of the different indices one can see, that they are different as for the different months, as for the two model runs, and for the considered forecast resolution and time frequency as well.

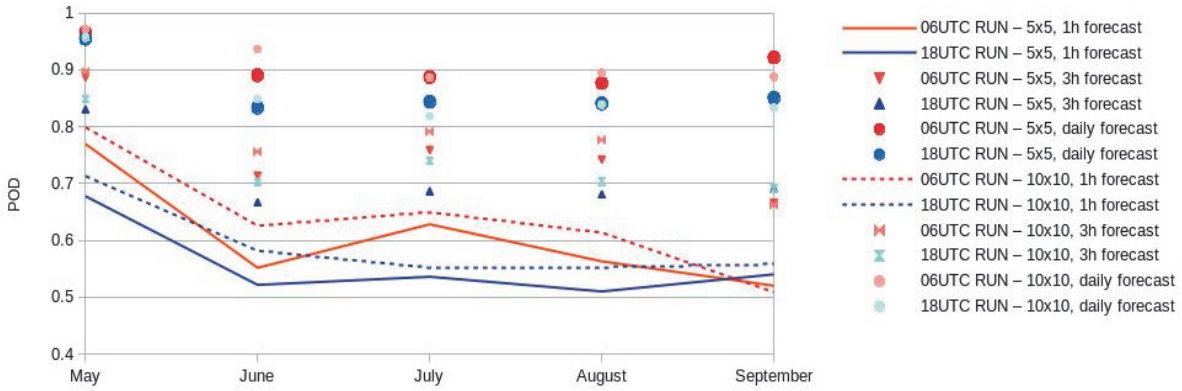

Fig. 7. 2021: Monthly distribution of the probability of detection (POD) for the period May-September 2021.

POD (*Fig. 7*) is higher for model run at 06 UTC in comparison to this for model run at 18 UTC. Probability of detection is highest in May (around and above 0.7) and decreases slightly for other months (remaining above 0.5).

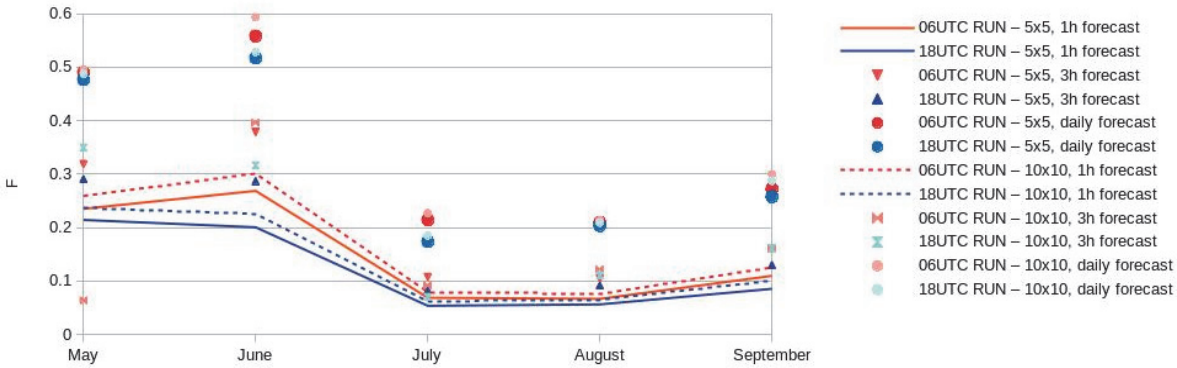

Fig. 8. 2021: Monthly distribution of the false alarm rate (F) for the period May-September 2021.

F (*Fig. 8*) and PC (*Fig. 9*) are considerably lower/higher, respectively, during July, August, and September in comparison to the other months, when the 18 UTC run seems to predict more correctly lightning activity in comparison to the 06 UTC run.

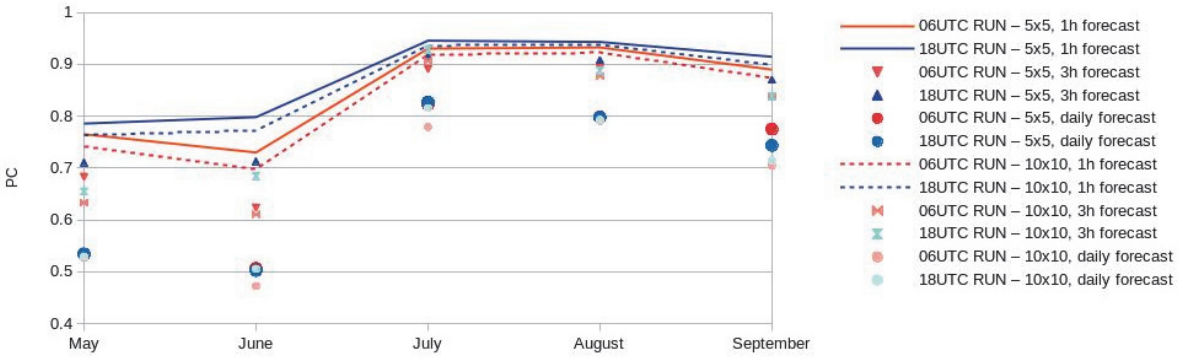

Fig. 9. 2021: Monthly distribution of the proportion correct (PC). for the period May-September 2021.

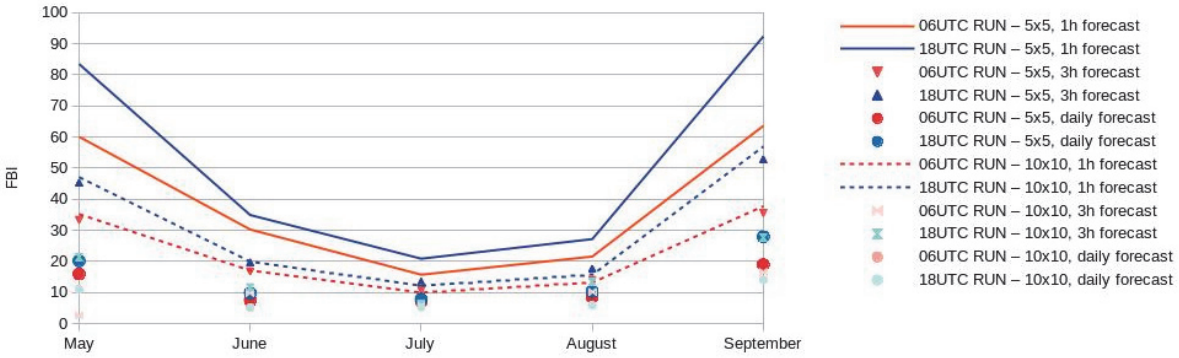

Fig. 10. 2021: Monthly distribution of the frequency bias index (FBI) for the period May-September 2021.

Monthly values of FBI (*Fig. 10*) show a considerable thunderstorm over-forecasting, especially in May and September. Results for the 06 UTC run are slightly better in comparison to those for the 18 UTC run, and the decrease of spatial and temporal resolution also improve the FBI slightly. However, the number of incorrectly forecasted "no lightning cases" is considerably higher than the number of incorrectly forecasted "lightning cases" leading to high values of FBI.

2.3. Spatial distribution of forecast skill-scores

Bulgaria is a relatively small country in the eastern part of the Balkan Peninsula, but with varying topography. Plains occupy about 35% of the territory, while plateaus and hills occupy 41%. The eastern boarder of the country is the Black Sea. The lowest point is the sea level, while the highest (Musala) is 2925 m, which is also the highest mount on the Balkan Peninsula. It is interesting to evaluate the performance of the lightning probability forecast based on the AROME-BG graupel mixing ratio over different parts of the country. *Figs. 11* and *12* show the monthly spatial distribution of POD and F (*Fig. 11*), and PC and FBI (*Fig. 12*) for the model runs at 06 and 18 UTC and the flash density detected by ATDnet for the months April to September 2021. For April, the spatial distribution of POD for the 06 and 18 UTC model runs are similarly with POD mainly above 0.9 at regions where flashes were detected (see the flash density in April), as for points from the grid without any detected flashes, POD is considered as 0. F is mainly between 0.3 and 0.4. The proportion correct index is between 0.5 and 0.7 for April (*Fig. 12*). The distribution of FBI for April shows that over the regions without any detected flashes, FBI is below 5, while where flashes were detected, FBI is between 20 and 60 for the 06 UTC run and reaching values above 100 for the 18 UTC run. For May, similarly to April, the spatial distribution of POD for the 06 and 18 UTC model runs are similar, with POD mainly above 0.9 at regions where flashes were detected. F and PC for May are mainly between 0.2 and 0.4 and between 0.6 and 0.8, with respectively lower or higher values (between 0.2 and 0.3 and between 0.7 and 0.8) at the southeastern (for the 06 UTC run) and eastern parts (for the 18 UTC run). FBI is between 5 and 20 over a big part of western Bulgaria for the 06 UTC model run, while for the 18 UTC run it is mostly above 20. Over the southeastern part for the 18 UTC run, FBI still reaches values above 100. The spatial distribution of POD in June and July differs more considerably for the two model runs. There are much more regions especially at the western part, where POD is above 0.9 for the 06 UTC model run in comparison to the 18 UTC run. In July, over a big part of eastern Bulgaria and over the Black Sea, POD is below 0.5, although there were detected flashes (see the flash density in July). F is below 0.2 in July over the whole considered region, while for June it even reaches values above 0.5. Similar is the distribution of PC following its score evaluation. FBI is mainly below 40 over the whole considered region during the months June, July, and August. The spatial distribution of POD in August is similar for the two model runs, with values mostly above 0.9. There are only some regions (at the northern and southeastern parst of Bulgaria), where The 18 UTC run shows lower scores, with values around 0.5. F and PC spatial distributions for August are similar to those for July. In September, spatial distribution of POD for the 18 UTC run shows more regions with POD>0.9 than for the 06 UTC run, as well more

regions with lower values of F and higher values of PC. However, values of FBI are higher over a big part of the domain for the 18 UTC run, showing higher over-forecast in comparison to the 06 UTC run.

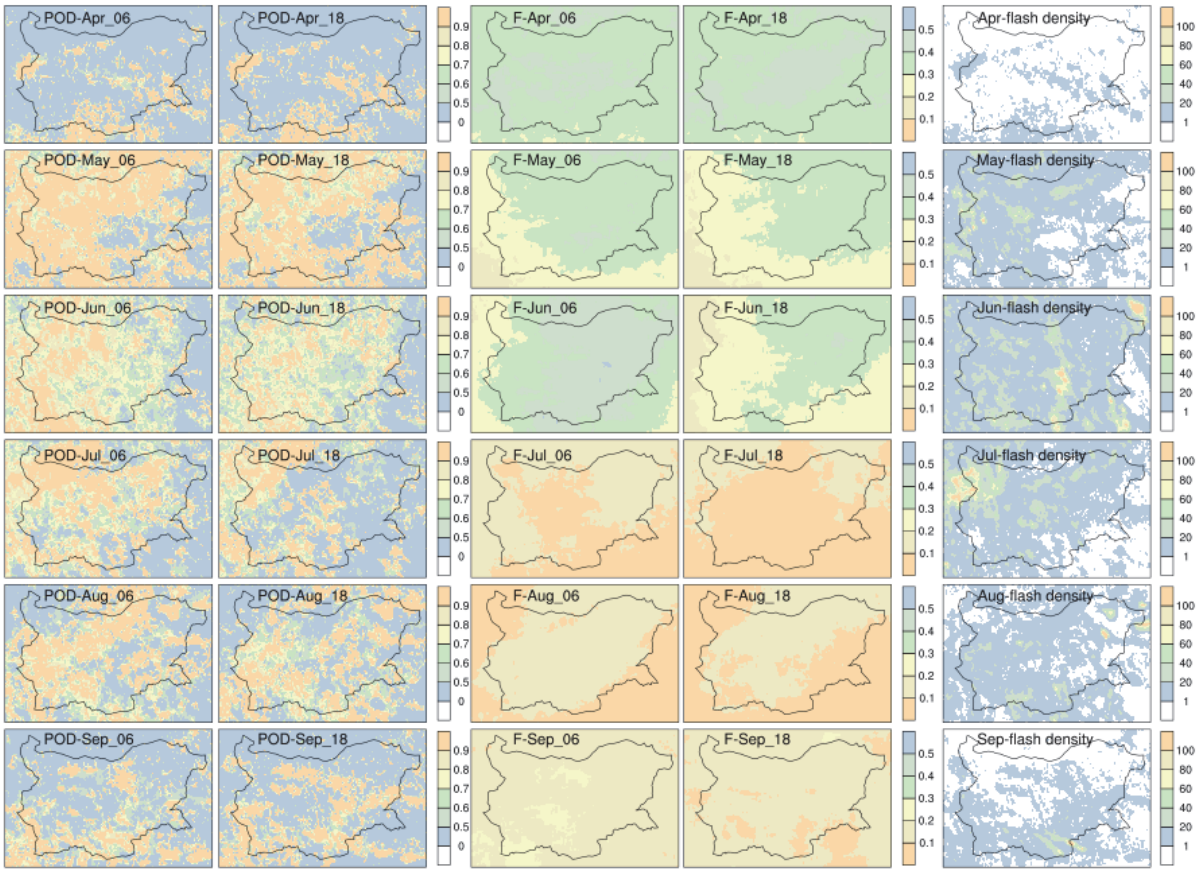

Fig. 11. 2021: Spatial distribution of probability of the detection (POD) and false alarm rate (F) for a grid of 5×5 km, for the model runs at 06 and 18 UTC (denoted respectively with "_06" and "_18"), and the flash density detected by ATDnet for the months April to September 2021 which is used to evaluate model forecast with a time frequency of 3 hours.

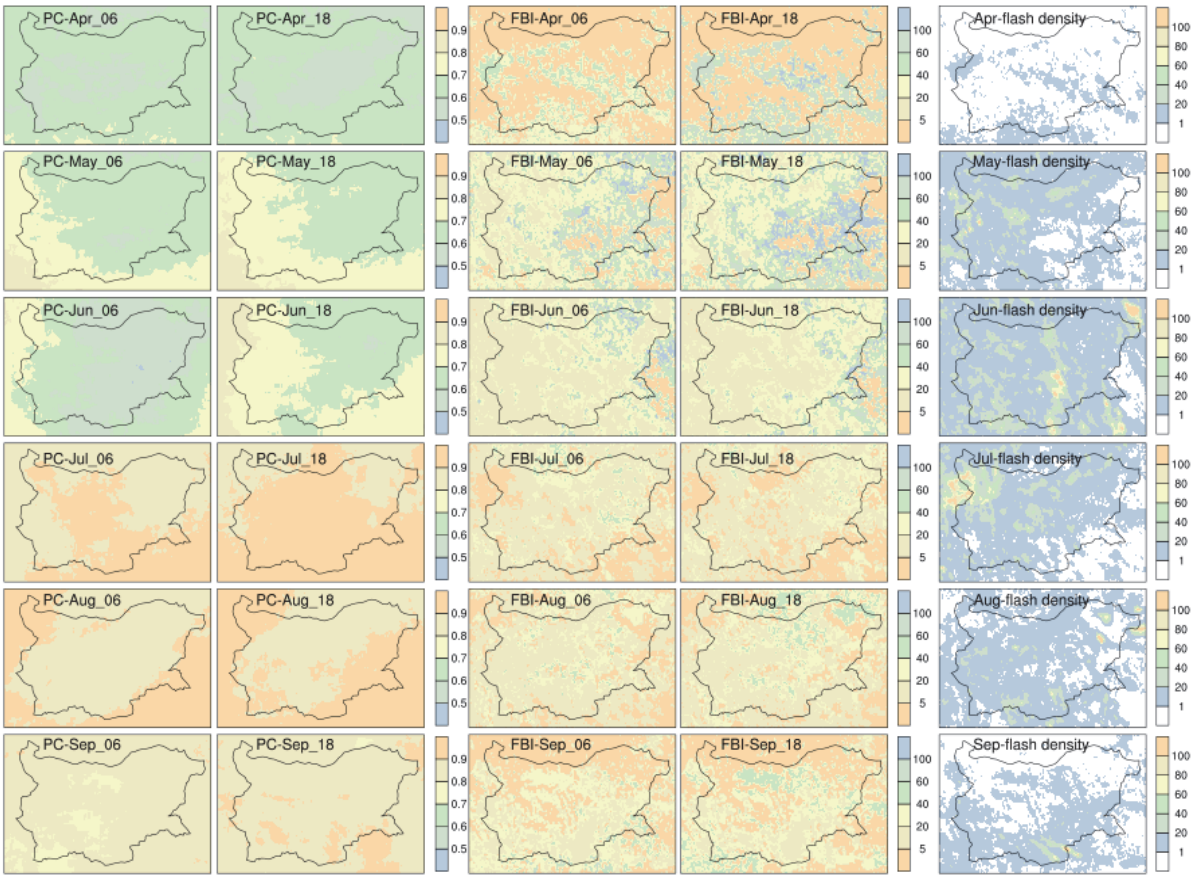

Fig. 12. 2021: Spatial distribution of proportion correct index (PC) and frequency bias index (FBI) for a grid of 5×5 km, for the model runs at 06 and 18 UTC (denoted respectively with "_06" and "_18"), and the flash density detected by ATDnet for the months April to September 2021 which is used to evaluate model forecast with a time frequency of 3 hours

Fig. 13 shows the boxplot of POD and F for lightning probability forecast (evaluated at a resolution of 5 km with a frequency of 3 hours) over different altitudes of the domain for the two model runs. As by default, the boxplot displays the 0th, 25th, 50th, 75th, and 100th percentiles of the corresponding data, or in other words, their minimum, first and third quartiles with median between, and maximum values. The altitude were separated in 8 groups: $z=0$ m, $0 \text{ m}<z\leq100$ m, $100 \text{ m}<z\leq500$ m, $500 \text{ m}<z\leq1000$ m, $1000 \text{ m}<z\leq1500$ m, $1500 \text{ m}<z\leq2000$ m, and $z>2000$ m. The width of boxes is proportional to the number of cases. It is visible that the probability of detection POD is the worst over the sea ($z=0$ m) with mean values about 0.3 for the two model runs. However, the POD medians equal to 0 may indicate that there are more points of the sea with no detected flashes than with not correctly forecasted lightning activity. Also, only over the sea, third quartiles of POD (0.5 and 0.67 for 06 UTC and 18 UTC model runs, respectively) differ from their maximum values that are 1. With the increase of the altitude,

POD scores get better. At the coastal zones (z<50 m and z<100 m), the median of POD is 0.5 (for the two model runs) with mean values also close to 0.5. At altitudes between 100 m and 500 m (z<500 m), which is the largest sample of data, the median of POD for the morning model run is 0.67, while for the evening run it is 0.5. Between 500 m and 2000 m, there is a significant difference between the minimum value (0) and first quartile of POD (that is between 0.4 and 0.5) for the two model runs, except for the 18 UTC run at altitudes between 500 and 1000m (z<1000). At these altitudes, the POD medians are around 0.8 and the mean values are between 0.6 and 0.7. The false alarm rate does not differ considerably for the different altitudes as for the two model runs, with median around 0.2.

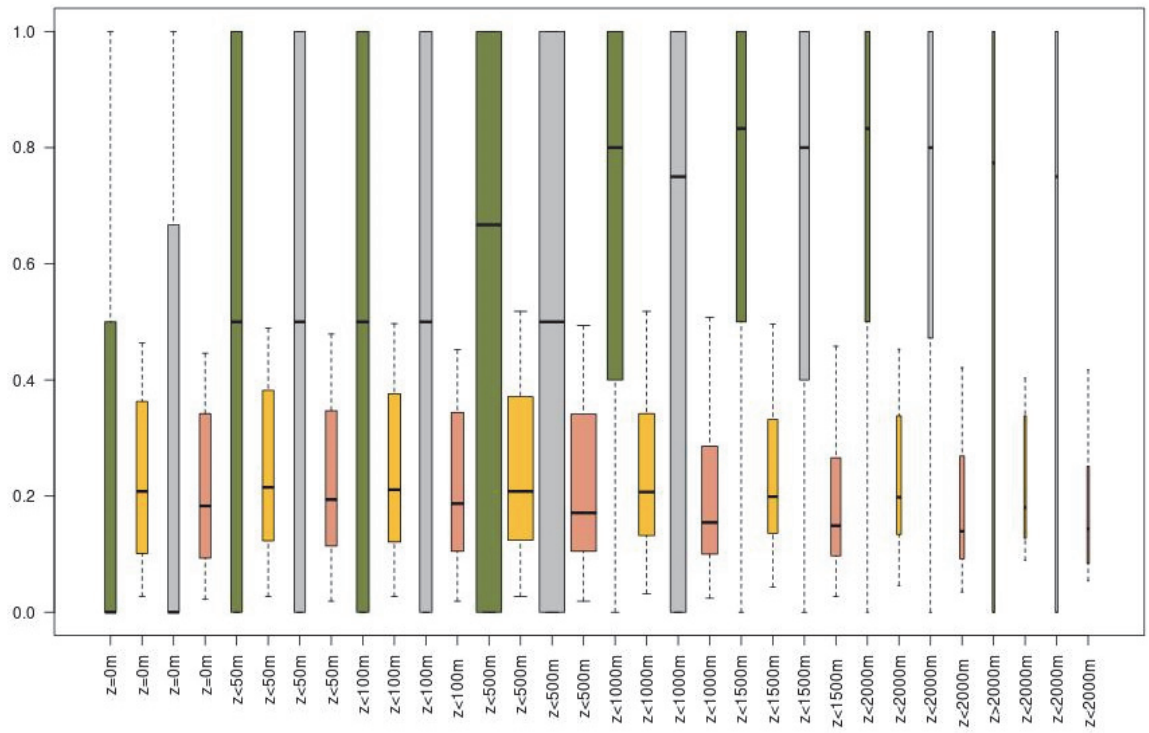

Fig. 13. Box plot of the probability of detection POD (green for the 06 UTC model run and grey for the 18 UTC model run) and false alarm rate F (yellow for the 06 UTC model run and orange for the 18 UTC model run) for different altitudes over the sea level. (The forecast is evaluated at a resolution of 5×5 km with a 3 hours frequency.)

2.4. Case study

Fig. 14 presents the hourly forecast (06 UTC run) for the graupel mixing ratio, r_g, integrated between 2756 m and 10306 m combined with the corresponding "lightning cases" according to the ATDnet flashes on a spatial grid of 5×5 km on July 1, 2021. This date was chosen as it is representing the different forecast skill-scores for the different hours. For this day, the thunderstorm probability forecast

based on r_g seems to be relatively satisfactory for the afternoon hours (as well as for the two days of the forecast), however with a visible overestimation of thunderstorm regions, while during the morning and evening hours, there are more lightning cases over regions, where no graupel mixing ratio was forecasted by AROME-BG. There were no detected lightning between 08 and 09 UTC on July 1 and between 02 and 07 UTC on July 2, and then POD and FBI have no values, while F and PC show very good skill-scores (with very low and high values, respectively), because regions with forecasted $r_g > 0$ are also a few. There are some hours when FBI is very high. At 08 UTC on July 2 for example, there are only three lightning cases on the eastern part of the domain (in the Black Sea) that were not predicted by the model. This results, as expected, in a null probability of detection, but also in a very high frequency bias index of 236 due to the low number of lightning cases. One hour later, at 09 UTC, lightning cases were still not predicted by the model, but their number is 29 (much higher in comparison to 08 UTC). However, as the number of non-lightning cases with forecasted $r_g > 0$ are not much more than those at 08 UTC (994 versus 709), it results in a significantly lower FBI of 34 (still very high and showing an over-forecasting of the thunderstorm probability over the whole domain). FBI is below 1 (0.19) only at the first hour of model integration, when there were no regions with forecasted $r_g > 0$, while lightning were detected. Results from the present case study also show that the increase of the number of lightning cases lead to a deterioration of the false alarm rate skill-scores almost independently of the forecast success rate, which suggests that F should not be considered independently.

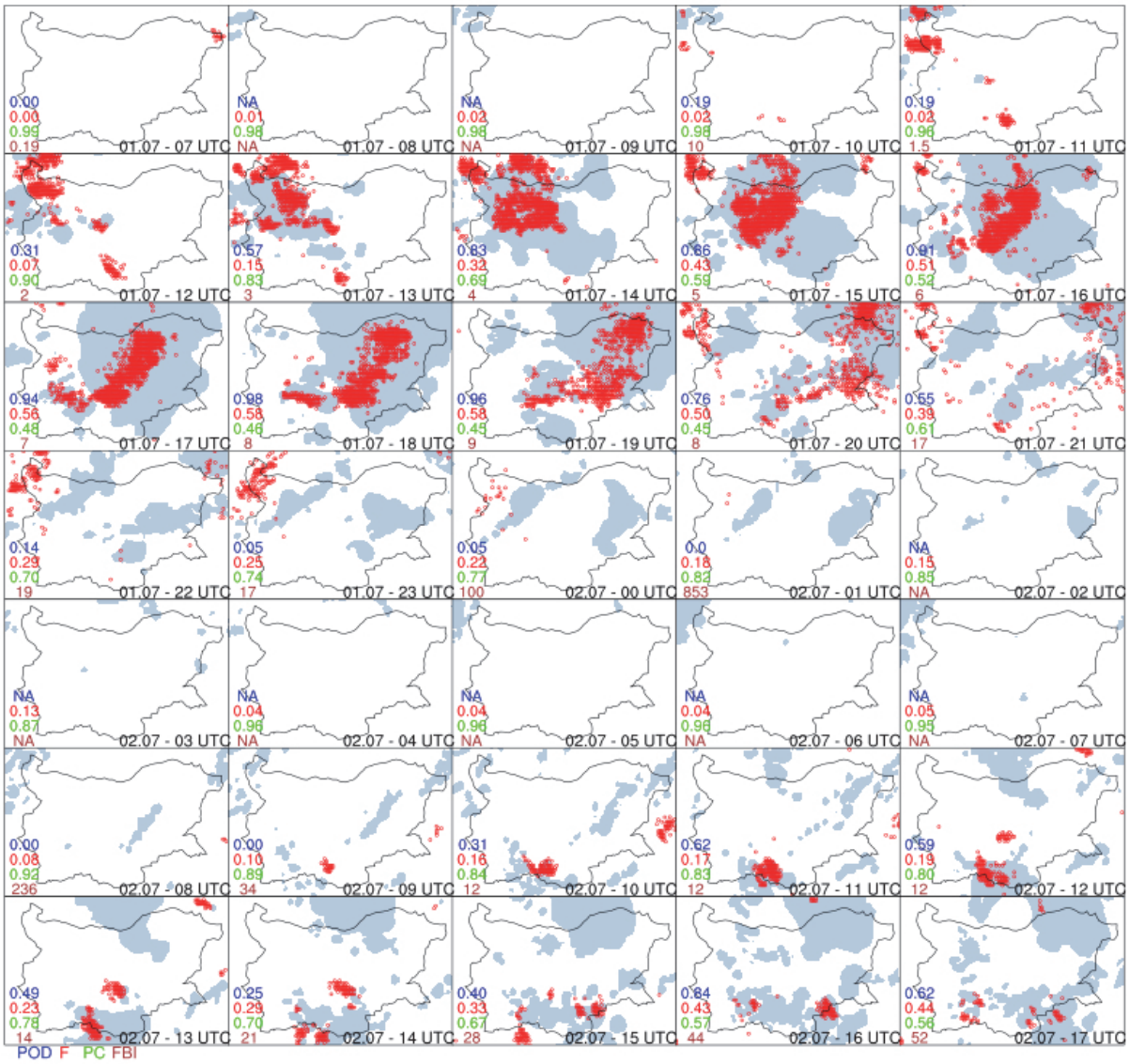

Fig. 14. AROME-BG forecast on July 1, 2021 (06 UTC run) for the graupel mixing ratio $r_g > 0$ integrated between 2756 m and 10306 m (grey) and cases with flashes detected by the ATDnet (red) with spatial resolution 5×5 km. Respective values of POD (in blue), F (in red), PC (in green), and FBI (in brown) are indicated for each hour of the forecast.

3. Summary

The results of this study show that:

– The decreases of the spatial resolution and time frequency of the forecast lead to an improvement of forecast's probability of detection (POD) and frequency bias index (FBI) and to a slight deterioration of its false alarm rate (F) and its proportion correct index (PC), and the impact of the time frequency is more pronounced.

Considering diurnal forecast distributions:

– The probability of detection (POD) and the frequency bias index (FBI) give considerably better scores of the forecast during the day hours in comparison

to the night hours, which shows that lightning activity is better forecasted and with less overestimations during the day, when in principle most flashes are detected;

- The false alarm rate (F) and the proportion correct index (PC) give slightly better scores of the forecast during the night hours. An explanation to these results could be the higher number of night cases that are without detected flashes, and then their proportion to all corresponding night cases is higher;
- There are no significant differences between POD, F, and PC for the two different model runs at 06 and 18 UTC, while FBI gives slightly better scores to the 18 UTC model run;

Considering monthly distributions:

- The robability of detection (POD) is higher for the first part of the warm half year of 2021, while the other considered skill-scores give better results in July, August, and September; the month with highest number of lightning cases, June has the lowest skill-scores;
- Regarding POD and FBI, there are differences in the monthly scores of the two model runs, and the 06 UTC run is with higher values of probability of detection and lower values of frequency bias index;

Considering the spatial forecast distributions:

- The false alarm rate (F) and the proportion correct index (PC) have similar spatial distribution of skill-scores for all months;
- In April, May, and September, the probability of detection (POD) is high over the regions with detected lightning activity for the two model runs, but the false alarm rate (F) have also high values (especially in April) over the whole considered domain; the score of the frequency bias index (FBI) differs for the two model runs during these months with higher values for the 18 UTC run;
- In June and July (the months with highest lightning activity over Bulgaria), skill-scores give different results for the two model runs, with higher POD and F for the 06 UTC run; better scores are obtained in the western part of Bulgaria, while the forecast for lightning activity over the Black Sea (eastern part of the domain) is worse;

Considering the height over the sea level forecast distribution

- Lightning probability forecast performance is similar for the two model runs at all considered altitudes, with slightly lower performance of the evening forecast; POD is improving with the increase of the terrain height, while F is not affected significantly.

4. Conclusions

In the present study, graupel mixing ratio for the warm half year 2021 based on AROME-BG model forecasts over Bulgaria was evaluated and connected to the detected lightning by using upscalling neighborhood method. Lightning data and forecasted graupel mixing ratios were considered on resolutions of 5×5 km and 10×10 km with flash rate for one and three hours, as well on a daily base using the upscaling neighborhood method. This technique was used to evaluate the accuracy of the forecast for the precise location and time, respectively. The main conclusion from the study is that the graupel mixing ratio taken from the cloud-resolving NWP model AROME-BG could be used as a tool to forecast lightning probability with a relatively high performance. The relatively low performance over the sea could be due to different reasons, like the model performance or the not appropriate considered model levels over the water pool. Such investigations have to be done in the future. Furthermore, it is expected that the inclusion of data assimilation in the operational numerical weather prediction at the National Institute of Meteorology and Hydrology of Bulgaria and especially the assimilation of lightning data will improve the lightning probability forecast in time and place.

References

Anderson, G. and *Klugmann, D.,* 2014: European lightning density using ATDnet data. *Nat. Hazards Earth Syst. Sci. 14*, 815–829. https://doi.org/10.5194/nhess-14-815-2014

Bechtold, P., J. Cuijpers, P. Mascart, and *P. Trouilhet*, 1995: Modelling of trade-wind cumuli with a low-order turbulence model–toward a unified description of Cu and Sc clouds in meteorological models. *J. Atmos. Sci. 52*, 455–463. https://doi.org/10.1175/1520-0469(1995)052<0455:MOTWCW>2.0.CO;2

Bougeault, P., 1982: Cloud-ensemble relations based on the gamma probability distribution for the higher-order models of the planetary boundary layer. *J. Atmos. Sci. 39*, 2691–2700. https://doi.org/10.1175/1520-0469(1982)039<2691:CERBOT>2.0.CO;2

Bouteloup, Y., 1995: Improvement of the spectral representation of the earth topography with a variational method. *Mon. Weather Rev. 123*, 1560–1573. https://doi.org/10.1175/1520-0493(1995)123<1560:IOTSRO>2.0.CO;2

Bruning E.C., Rust W.D., Schuur T.J., MacGorman D.R., Krehbiel P.R., and *Rison W.,* 2007: Electrical and polarimetric radar observations of a multicell storm in TELEX. *Mon. Weather Rev. 135*, 2525. https://doi.org/10.1175/MWR3421.1

Caniaux, G., Redelsperger, J.-L., and *Lafore, J.-P.,* 1994: A numerical study of the stratiform region of a fast-moving squall line. Part I: general description and water and heat budgets. *J. Atmos. Sci. 51*, 046–2074. https://doi.org/10.1175/1520-0469(1994)051<2046:ANSOTS>2.0.CO;2

Dotzek N., Holler H., Thery C., and *Fehr T.,* 2001: Lightning evolution related to radar-derived microphysics in the 21 July 1998 EULINOX supercell storm. *Atmos. Res. 56* (1-4), 335. https://doi.org/10.1016/S0169-8095(00)00085-5

Hallett, J. and *Mossop, S.C.*, 1974: Production of secondary ice particles during the riming process. *Nature 249*, 26–28. https://doi.org/10.1038/249026a0

Gaffard, C., Nash, J., Atkinson, N., Bennett, A., Callaghan, G., Hibbett, E., Taylor, P., Turp, M., and *Schulz, W.,* 2008: Observing lightning around the globe from the surface, in: the Preprints, 20th International Lightning Detection Conference, Tucson, Arizona, 21–23.

Jayaratne, E.R., Saunders, C.P.R., and *Hallett, J.,* 1983: Laboratory studies of the charging of soft-hail during icecrystal interactions. *Quart. J. Roy. Meteorol. Soc. 109*, 609–630. https://doi.org/10.1002/qj.49710946111

Kessler, E., 1969: On the distribution and continuity of water substances in atmosphericcirculations. Meteorol. Monogr. 10, 32 Bolton. https://doi.org/10.1007/978-1-935704-36-2

Lee, A.C., 1986: An operational system for the remote location of lightning flashes using a VLF arrival time difference technique. *J. Atmos. Oc. Technol. 3*, 630–642. https://doi.org/10.1175/1520-0426(1986)003<0630:AOSFTR>2.0.CO;2

Lin, Y.-L., Farley, R.D., and *Orville, H.D*., 1983: Bulk parameterization of the snow field in a cloud model. *J. Clim. Appl. Meteorol. 22*, 1065–1092 https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2

Lund N.R., MacGorman D.R., Schuur T.J., Biggerstaff M.I., and *Rust W.D.*, 2009: Relationships between lightning location and polarimetric radar signatures in a small mesoscale convective system. *Month. Weather Rev. 137*, 4151. https://doi.org/10.1175/2009MWR2860.1

McCaul, E.W., Goodman, S.J., LaCasse, K.M., and *Cecil, D.J*. 2009: Forecasting lightning threats using cloud-resolving model simulations, *Weather Forecast. 24*, 709–729. https://doi.org/10.1175/2008WAF2222152.1

Meyers, M.P., DeMott, P.J., and *Cotton, W.R.,* 1992: New primary ice-nucleation parameterizations in an explicit cloud model. *J. Appl. Meteorol. 31*, 708–721. https://doi.org/10.1175/1520-0450(1992)031<0708:NPINPI>2.0.CO;2

Pinty, J.-P. and *Jabouille, P.,* 1998: A mixed-phased cloud parameterization for use in a mesoscale non-hydrostatic model: Simulations of a squall line and of orographic precipitation. Preprints, Conf. on Cloud Physics, Everett, WA, Amer. Meteor. Soc., 217–220.

Reynolds, S., Brook, M., and *Gourley, M.F.,* 1957: Thunderstorm charge separation. *J. Meteor*. *14*, 426–436. https://doi.org/10.1175/1520-0469(1957)014<0426:TCS>2.0.CO;2

Takahashi, T., 1978: Riming electrification as a charge generation mechanism in thunderstorms. *J. Atmos. 35*, 1536–1548. https://doi.org/10.1175/1520-0469(1978)035<1536:REAACG>2.0.CO;2

Termonia, P., Fischer, C., Bazile, E., Bouyssel, F., Brožková, R., Bénard, P., Bochenek, B., Degrauwe, D., Derková, M., El Khatib, R., Hamdi, R., Mašek, J., Pottier, P., Pristov, N., Seity, Y., Smolíková, P., Španiel, O., Tudor, M., Wang, Y., Wittmann, C., and *Joly, A*.: The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1. *Geosci. Model Dev., 11*, 257–281. https://doi.org/10.5194/gmd-11-257-2018