# SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

| | |
|---|---|
| **Reporting year** | ……………2021/2022………………………………..…… |
| **Project Title:** | Improvement of very-short term forecast using lightning and radar data assimilation |
| **Computer Project Account:** | …………SPITFEDE……………………………………..…… |
| **Principal Investigator(s):** | ……Stefano Federico…………………………..…… |
| | ………………………………………………………………..…… |
| **Affiliation:** | ………… CNR-ISAC ……………………..…… |
| **Name of ECMWF scientist(s) collaborating to the project** (if applicable) | …Claudio Transerici (cmn)………………………..……<br>…Rosa Claudia Torcasio (it85) ……..…… |
| **Start date of the project:** | ………1 January 2021…………………………………..…… |
| **Expected end date:** | …… 31 December 2023…………………..…… |

## Computer resources allocated/used for the current year and the previous one

(if applicable)

Please answer for all project resources

| | | Previous year | | Current year | |
|---|---|---|---|---|---|
| | | Allocated | Used | Allocated | Used |
| **High Performance Computing Facility** | (units) | | | 20000000 | 3200000 |
| **Data storage capacity** | (Gbytes) | | | 35 TB | 53 TB |

## Summary of project objectives (10 lines max)

The general objective of the project is to evaluate the impact of data assimilation on the very-short term forecast over Italy, with emphasis on high precipitation events. This objective is large and must be focused on specific goals that can be tackled in the framework of this special project.
For this project,t the data assimilation is focused on lightning and radar reflectivity data, even if some additional data taken from SENTINEL satellites could be explored, while the verification is mainly on precipitation forecast. Some additional parameters, as surface winds, could be also considered for verification.

## Summary of problems encountered (10 lines max)

No significant problems were encountered working on the project. The disk quota of 35 TB is full and additional space was used on ECFS. Should this cause problems, we will transfer some of the datasets elsewhere. Considering the consumption of allocated resources (20.000.000 SBU), the data 3.200.000 SBU is not representative of what we did in the project. The availability of TEMS, first, and then of the Bull supercomputer gave us the possibility to run many simulations (for more than 15.000.000 of SBU) on the new machines, at no SBU cost. So, most of the work was done on the Bull supercomputer and the CRAY was not used. For this reason, the SBU consumption is low compared to the initial estimation. In addition, the possibility to use the licensed package IDL on the Bull supercomputer gave us the opportunity to do to post-processing directly on ECMWF supercomputers without transferring data and justify the high consumption of the disk space. The staff of ECMWF is warmly acknowledged for giving us an ideal system to develop the project.

## Summary of plans for the continuation of the project (10 lines max)

In the remainder of the project the focus will be on the assessment of the role of radar reflectivity data assimilation (RDA) and lightning data assimilation (LDA) on the improvement of the Very Short-Term precipitation forecast over Italy. In the first two years of the project, the best setting of the WRFDA was found for the assimilation of both observations: for radar the 3D-Var is applied by using WRFDA, while for lightning the nudging is preferred.

## List of publications/reports from the project with complete references

A first paper was published in the first year of the project. The reference to this paper is:
Federico, S.; Torcasio, R.C.; Puca, S.; Vulpiani, G.; Comellas Prat, A.; Dietrich, S.; Avolio, E. Impact of Radar Reflectivity and Lightning Data Assimilation on the Rainfall Forecast and Predictability of a Summer Convective Thunderstorm in Southern Italy. Atmosphere 2021, 12, 958. https://doi.org/10.3390/atmos12080958

A second paper was also submitted, in which lightning forecast over Italy using WRF model is considered. The paper was submitted to the special issue "Remote Sensing of Lightning and Its Applications in Atmospheric Electricity Studies" (https://www.mdpi.com/journal/remotesensing/special_issues/lightning_electricity) of the journal "Remote sensing" and the paper title is: "A year-long total lightning forecast over Italy with a Dynamic Lightning Scheme and WRF". This paper is currently under review.

## Summary of results

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

**WRF model**

The WRF numerical model is used in this project with Advanced WRF dynamic (WRF-ARW), version 4.1.3. The simulations use one domain, with 635*635 grid points and 50 unevenly spaced vertical levels with a model top at 50 hPa. The model domain (Figure 1) covers the Central Mediterranean and the whole Italian territory. It has a horizontal grid spacing of 3 km.

The physical schemes employed include the Thompson microphysics scheme, the Mellor-Yamada-Janjic scheme, using a one-dimensional prognostic turbulent kinetic energy scheme with local vertical mixing, the 5-layer thermal diffusion for land surface processes scheme, the Monin-Obukhov (Janjic Eta) scheme for surface layer physics, the Dudhia scheme for the short-wavelength radiative scheme, and the Rapid Radiative Transfer Model (RRTM) for the longwave radiative scheme. No cumulus parameterization was used as these are convection-allowing simulations.

Initial and boundary conditions for the simulations are provided by the Integrated Forecasting System (IFS) global model of the European Centre for Medium-Range Weather Forecasts (ECMWF). Specifically, we use the analysis-forecast cycle issued at 12 UTC on the day before the actual day to forecast. Data are downloaded at 0.25° horizontal resolution.

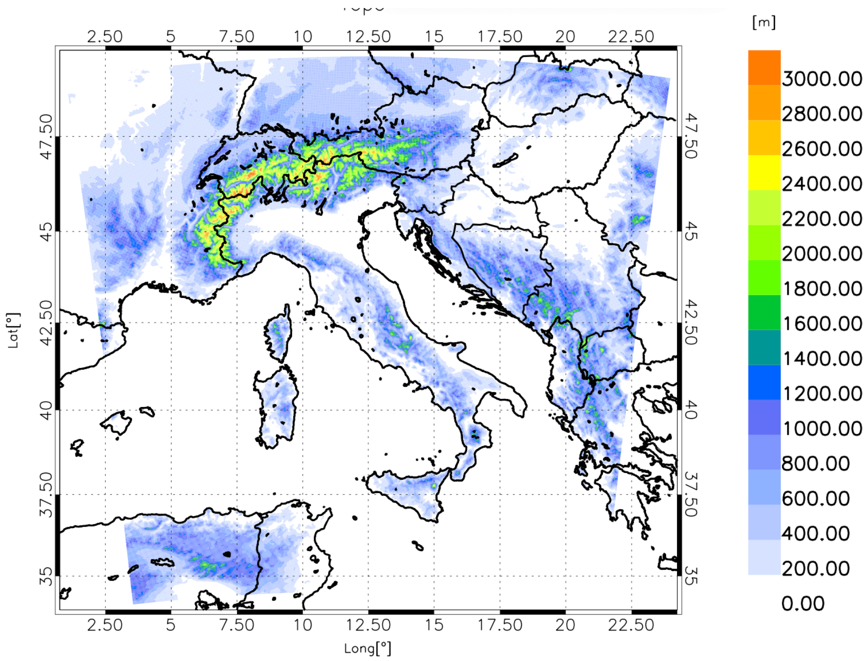



Figure 1: Orography of the WRF model for the numerical experiments.

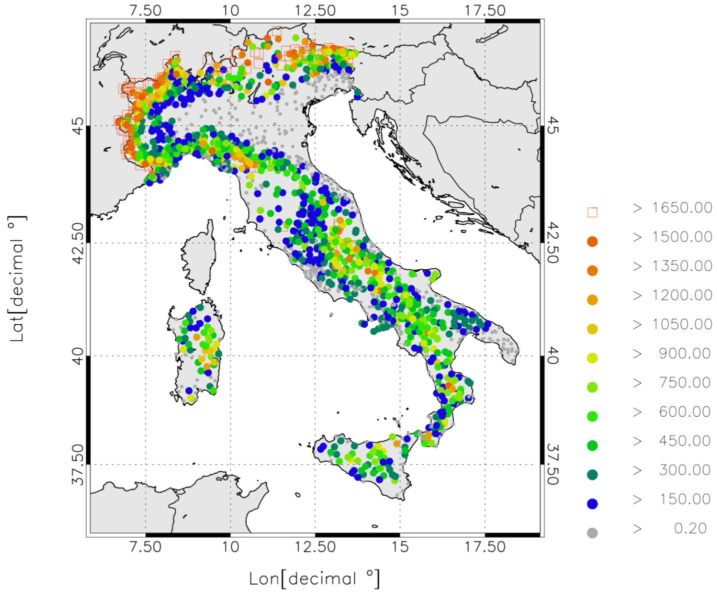



Figure 1: Raingauges used for verification. Colors are the heights in mtetres.

**Lightning and radar reflectivity data assimilation**

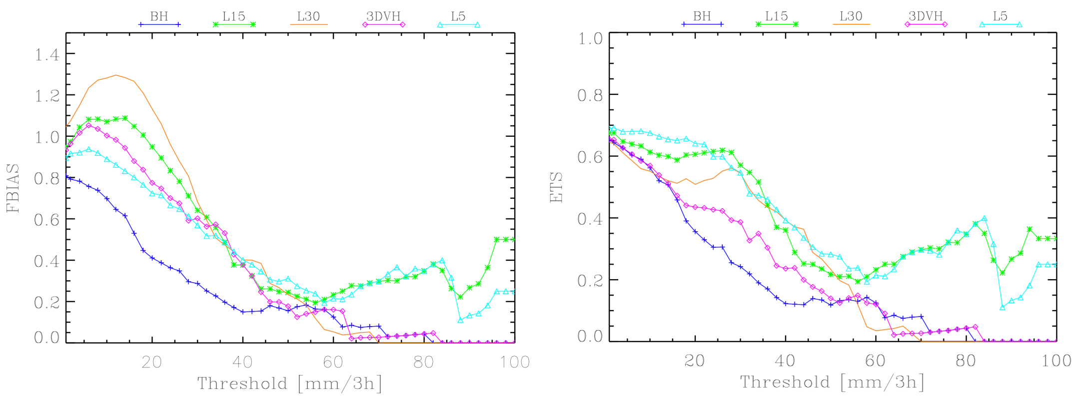
Lightning data assimilation was done in the project with two different techniques: nudging and 3D-Var. Radar reflectivity data assimilation is done with 3D-Var. The lightning data assimilation is made by nudging following the scheme of Fierro et al. (2012). Lightning data assimilation made by 3D-Var is performed through pseudo-profiles of saturated relative humidity between the Lifting Condensation Level (LCL) and the -25°C isotherm and then assimilated through WRFDA. The radar reflectivity data assimilation is done using the indirect method in WRFDA accounting for snow and graupel mixing ratio in addition to the rainwater mixing ratio. The saturated water vapor mixing ratio is assimilated inside the cloud.

**Forecast verification**

Forecast verification is done with two different methods. For all the applications, except for lightning prediction, we verify the precipitation forecast over Italy at the short-range. The setting is the following: we assimilate lightning or/and radar reflectivity for six hours, then we forecast the precipitation in the next three hours. The initial 6h are used for both data assimilation and for the spin-up of the model, while the next three hours are the free forecast, which is verified with raingauges. The rainfall is verified over all Italy with about 3000 raingauges spread over all the country (Figure 2). Not all raingauges report valid observation for the specific period, however the average number of reporting stations is around 2800. Lightning forecast is verified daily (i.e. for the next day total lightning forecast). The verification is performed over the area 6-19°E and 37-47.5°N accounting for Italy and the central Mediterranean basin.

**Comparison of LDA with different data assimilation methods**

In this section we analyze the results of numerical experiments with different settings of the lightning data assimilation into the WRF model.
Ten days of moderate-intense rainfall over Italy are considered for this experiment: form 13 to 22 October 2019. These days were characterized by localized and/or spread convection over Italy. Simulations were done using a short-term approach. Different settings of the LDA were considered: a) L5 is the configuration in which lightning are accumulated every 5 minutes and the scheme of Fierro et al. (2012) is updated every 5 minutes (note that while the scheme is updated every 5 minutes, nudging is applied every model time-step, which is 12 s); b) L15, which is like L5 but flashes are accumulated every 15 minutes in the scheme used for LDA and the scheme is updated every 15 minutes; c) L30, which is like L5 and L15, but flashes are accumulated over a 30min time interval and the LDA scheme is updated every 30 minutes; d) 3DVH, in which a pseudo profile of saturated water vapor mixing ratio is generated at every grid-point in which flashes are observed and then this pseudo-profile of relative humidity is assimilated by 3D-Var. For the lightning data assimilation with 3D-Var, flashes are accumulated over 30 minutes time intervals and LDA is applied every 30 minutes.
Figure 3 show the results for the computation of scores from dichotomous contingency tables accumulated over all the periods (about 120.000 pairs of observed and forecast rainfall). They are Frequency Bias (FBIAS), Equitable Threat Score (ETS), Probability of detection (POD) and False alarm Ratio (FAR). Precipitation threshold considered in the analysis are up to 100 mm/3h, which represent a heavy precipitation event.
The analysis of the FBIAS shows a clear improvement of the score when flashes are assimilated with different methods. In particular, the FBIAS was underestimated by the background for the period, and it is less than 0.1 for thresholds larger than 60 mm/3h. Assimilating lightning, except for the 3D-Var method, the FBIAS increases to 0.3-0.5 for thresholds larger than 60 mm/3h. The very important role of LDA to improve the short-term precipitation forecast over Italy is shown by the results of the ETS score. All configurations assimilating lightning have better scores compared to the background. Among the LDA forecasts, L5 and L15 have similar scores and have the best performance, then follows L30 and finally the LDA with 3D-Var approach.



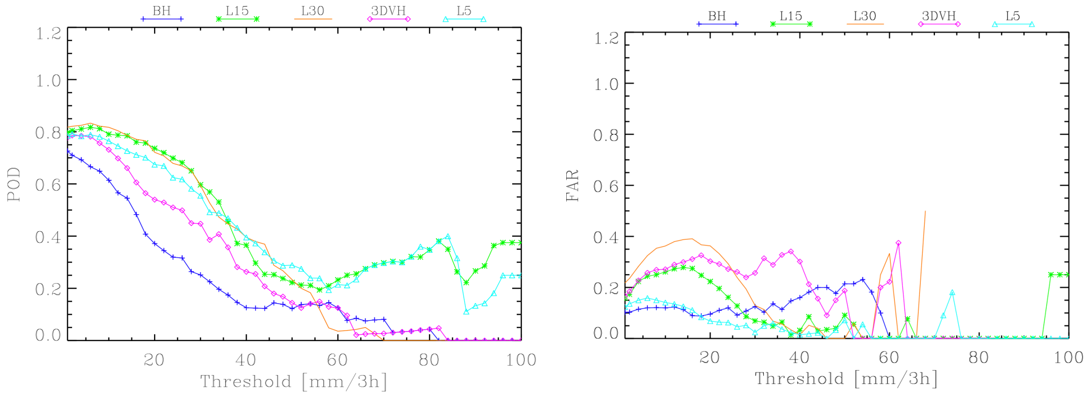



Figure 3: Scores of the short-term forecast for different LDA. BH is the background, L5 is the LDA updated every 5 minutes; L15 is the LDA updated every 15 minutes; L30 is the LDA updated every 30 minutes; 3DVH is the LDA so the 3D-Var. The upper-left panel is the Bias, the upper-right panel is the ETS, the lower-left panel il the POD and the lower-right panel is the FAR.

It is clearly shown the ability of LDA, using L5 and L15 configurations, to improve the forecast of deep convective events over Italy. Compared to L5 and L15, the configuration L30 and the LDA using the 3D-Var approach show worse performance. For these configurations, LDA is able to improve the performance for moderate to high precipitation (from 20 mm to 50 mm/3h), however there is no beneficial impact for larger thresholds.
These results are confirmed by the analysis of the POD score. Again, all configurations assimilating lightning improve the forecast, and the impact is notably high for the L5 and L15 configurations because the precipitation is improved for heavy rainfall (> 60 mm/3h). However, the 3DVH and L30 configurations improve the forecast up to 60 mm/3h rainfall threshold with no impact for higher rainfall.
The analysis of the FAR shows that LDA increases the FAR for thresholds below 20 mm/3h-30 mm/3h, depending on the configuration, except for the L5, which performs very well for all thresholds. However, for larger thresholds, the FAR decreases for L5, L15 and L30 settings. This clearly shows the key role of the LDA for the prediction of heavy rainfall. The POD increases while the FAR decreases, showing that LDA precisely predicts the position of deep convection.
All in all, the results of this experiment show that LDA improves substantially the precipitation forecast over Italy at the short-range. The configurations L5 and L15 have better performances, likely because they follow closely the time evolution of the storm, while L30 and 3DVH have worse performance. The analysis of the FAR shows that the 3DVar approach causes more false alarms than other configurations and a better tuning of the 3DVar is necessary to improve the short-term precipitation forecast.

**Comparison between LDA and RDA and their combined effect**

In this section we show the impact of lightning data assimilation (LDA) and radar data assimilation (RDA) on the precipitation forecast at the short-term. We evaluated the performances of LDA, RDA and of their combined effect. Radar reflectivity and lightning data were assimilated in the WRF model. Radar reflectivity data assimilation was performed by 3DVar, considering CAPPI on 7 vertical levels, namely one level every 1km from 2km to 8km. Lightning data are assimilated through nudging with the L15 setting.
The experiment was conducted considering 10 days with moderate to intense precipitation activity over Italy. The 10 selected days are the following: 03, 04, 05, 07, 11, 14, 15, 24, 26 and 27 October 2020. Four different model configurations were considered:

B, without lightning or radar reflectivity data assimilation;
L, with lightning data assimilation;
R, with radar reflectivity data assimilation;
RL, with both lightning and radar reflectivity data assimilation.

WRF model was initialized using initial and boundary conditions from European Centre for Medium range Weather Forecast (ECMWF) analysis/forecast available at a horizontal grid spacing of 0.25°. The analysis/forecast cycle issued at 12 UTC on the day before the actual day to forecast were used for initialization.
The forecasts were performed using a Very Short-term Forecast (VSF) approach. More in detail, we ran 4 simulations a day, to cover the whole day. Each run lasted 12 hours. The first 6 hours of the run were considered as spin-up time and were also used for data assimilation when DA is applied, while the last 6 hours were considered as forecast period.
In order to assess the model performance at different time ranges after the assimilation stage, we verified model precipitation capability both in the first three hours after assimilation and in the following three hours, i.e. from three to six hours after the end of the assimilation phase.

In this section we show the scores computed for the whole period for the forecast interval 0-3h. The scores (Figure 4) are computed for the threshold of 1mm/3h and from 2mm/3h to 60 mm/3h every 2 mm/3h by summing the elements of the dichotomous contingency tables for each time period.

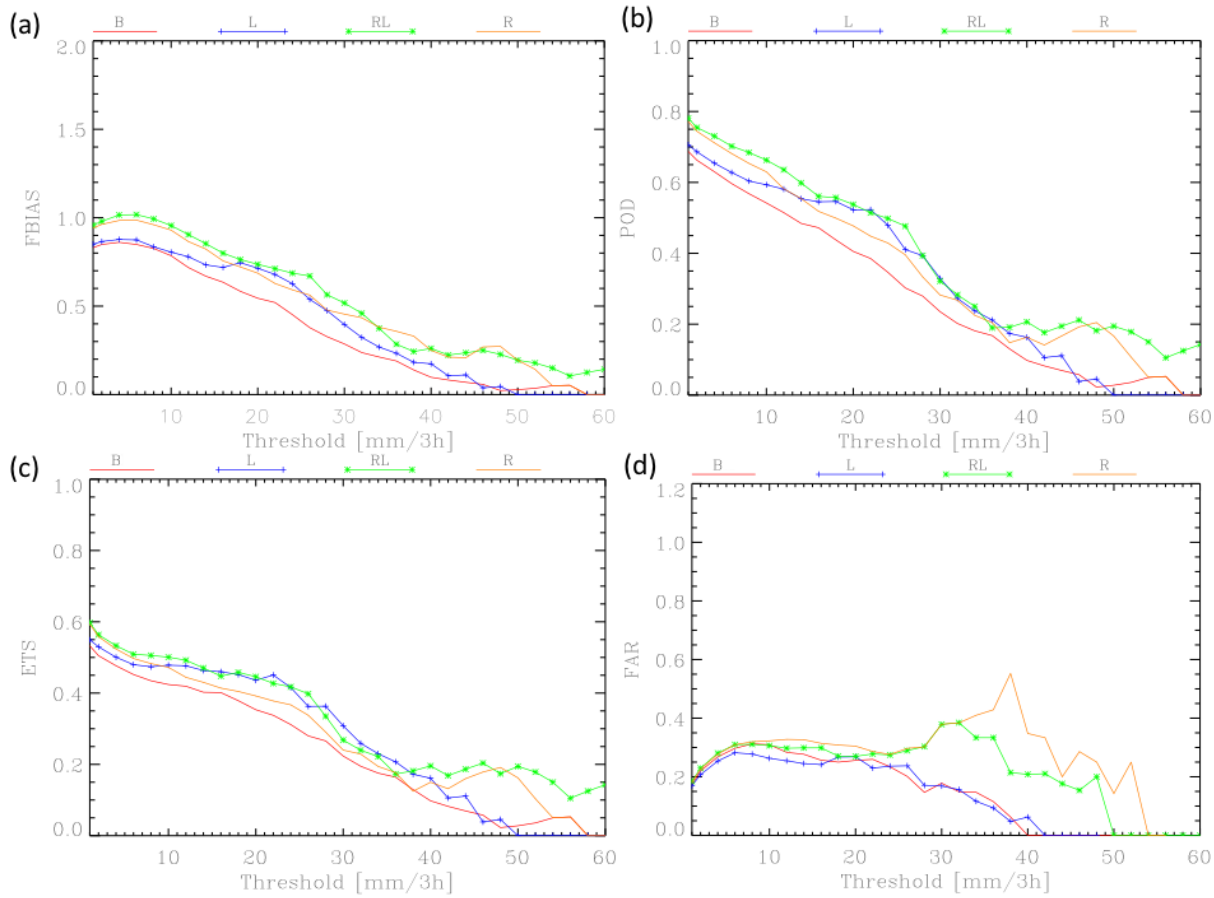



Figure 4: Precipitation scores for 10 days in October 2020 for the phase 0-3h after assimilation period: (a) FBIAS; (b) POD, (c) ETS and (d) FAR.

FBIAS is improved by all types of data assimilation considered: RDA, LDA, and their combined effect. The control forecast has a FBIAS lower than 1 for all thresholds; the forecast L, R and RL improve the score, which is closer to 1 for almost all thresholds, even if there is an underestimation for thresholds higher than 30 mm/3h. Considering the POD score, we note an improvement of the forecast with radar and/or lightning data assimilation. The improvement given by RDA is larger compared to LDA, especially for thresholds larger than 40 mm/3h. Interestingly, the POD for RL forecast is the best for almost all thresholds suggesting that RL forecast is improved by the contribution of both LDA and RDA.
Considering false alarms, it is noticed an increase of the FAR score for R and RL forecasts for thresholds larger than 26 mm/3h, while the FAR of L forecast is in line with the control forecast. The ETS score, which considers for both correct forecasts and false alarms, shows better forecast for RDA compared to LDA for threshold below 10 mm/3h and for thresholds above 40 mm/3h. For intermediate thresholds LDA performs better. The RL forecast has the best performance for most thresholds, confirming the usefulness of assimilating together radar reflectivity and lightning at the short-range.
Considering the forecasting tine between 3 and 6h, the improvement given by RDA and LDA to the precipitation forecast is much smaller than that shown in Figure 4, even if some improvements are apparent for thresholds larger than 40 mm/3h.
All in all, the results of this study show that: a) RDA and LDA have an important impact on the precipitation forecast over Italy; b) RDA and LDA acts synergistically for precipitation prediction, and the simulations assimilating both data sources have the best forecast between 0 and 3h; c) the impact between 3 and 6 h forecast is small and is apparent only for the largest thresholds.
In the following of this special project, we will explore better those findings for an extended period, to see also the impact of the two data sources, radar and lightning, on the precipitation forecast over Italy.

**Lightning prediction**

Lightning prediction is an important task for several applications. It is also a necessary step to implement a spurious convection suppression by lightning data assimilation.
In the second year of the project, the electrification scheme of Lynn et al. (2012) was tested over Italy for 162 cases occurred between 1 March 2020 and 28 February 2021. The cases span the four seasons and reflect the seasonal behavior of lightning: 69 cases occurred in summer, 46 in fall, 18 in winter and 29 in spring.
Three different configurations of the lightning dynamic scheme are considered: L50, L75 and L100. These configurations differ for the key parameter of assumed charge transferred in one second within the convective and stratiform clouds. They are, respectively, $0.5*10^{-4}$ C, $0.74*10^{-4}$ C and $1.0*10^{-4}$C.
For these simulations, the forecasts last 36 h. The first 12h are used to spin-up the model from a cold start , while the remaining 24h are the actual forecast.
Among the three dynamic lightning scheme configurations considered, L75 accurately forecasts the total number of strokes recorded for all the cases, L50 underestimates the strokes and L100 overestimates the strokes. However, the relative performance of L50, L75 and L100 for the number of strokes depends on the season.
We considered the performance of the lightning forecast in different seasons, and we found the best results for summer and fall, followed by winter and finally spring. We also considered the performance of lightning forecast over the land and the sea, and we get that the lightning prediction is better over the land than the sea. This may reflect the fact the land focus convection in preferential directions that are at least in part correctly simulated by the model.
The results of this quite large (about 10.000.000 SBU) numerical experiment have been synthetized to a scientific paper, which is under review in Remote Sensing.

**Conclusions**

Up to now several aspects of lightning and radar reflectivity data assimilation were explored with WRF over Italy. We found that the lightning data assimilation by nudging has a better performance compared to the 3D-Var, especially the L15 and L5 configurations. This result is likely caused by the fact that nudging can be applied more often that 3D-Var and the model is better able to follow the evolution of the convection, which is a fast-evolving phenomenon. We also did a preliminary assessment of the role of lightning and radar reflectivity data assimilation on the precipitation forecast at the short range. Results show that the assimilation of both parameters is useful to improve the precipitation forecast. The synergist use of both data can give added value to the simulation, nevertheless this point needs further assessment. This will be the main subject of the third year f research.

**References**

Fierro, A. O., E. R. Mansell, C. L. Ziegler, and D. R. MacGorman (2012). Application of a lightning data assimilation technique in the WRF-ARW Model at cloud-resolving scales for the tornado outbreak of 24 May 2011. Mon. Wea. Rev., 140, 2609–2627, doi:10.1175/MWR-D-11-00299.1.

Gao, J., and Stensrud, D. J. (2012). Assimilation of Reflectivity Data in a Convective-Scale, Cycled 3DVAR Framework with Hydrometeor Classification. *Journal of the Atmospheric Sciences* 69, 3, 1054-1065, available from: <https://doi.org/10.1175/JAS-D-11-0162.1>

Lynn, B. H., Y. Yair, C. Price, G. Kelman, and A. J. Clark (2012). Predicting cloud-to-ground and intracloud lightning in weather forecast models. Wea. Forecasting, 27, 1470–1488, doi:10.1175/WAF-D-11-00144.1.