FINAL REPORT PROGRAM LEFE

Program LEFE/MANU	Project Title **Advanced Statistical Methods for the Study of Atmospheric Deep Convection (ASMA)**	Years 2020 – 2020

PI: Olivier Caumont, olivier.caumont@meteo.fr, CNRM
Participating Laboratories: Laero, IGE

Contribution to
SOLID project (CNES)
Other funding sources: CNES, Météo-France

Context

While statistical learning is becoming more and more popular due to the availability of new and easily accessible algorithms, large computational resources and more and more data, and its potential application to the field of meteorology seems immense, its use is still in its infancy in the field of meteorology.

Objectives / scientific questions

The objective of this project was to apply and evaluate innovative statistical methods (mainly so-called 'machine-learning' algorithms) to the processing of big data in the field of atmospheric deep convection study.

Main results

- Design and validation of a simulator of lightning data seen from space using lightning data seen from the ground to mimic the observations of the upcoming Meteosat Third Generation (MTG) Lightning Imager (LI).
- Design and validation of an observation operator for space-based lightning observations in preparation for lightning data assimilation.
- Design and validation of automatic processing of connected personal weather station data.

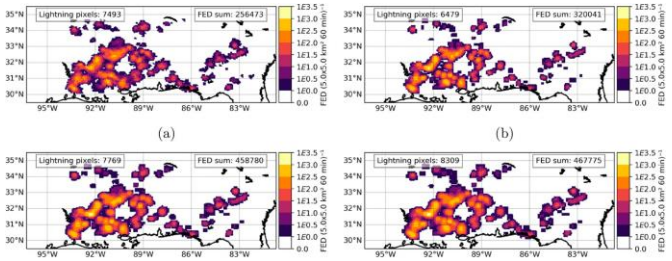

Fig. 1: (a) Observed and simulated hourly flash extent density (FED) using (b) a linear support vector regressor, (c) a multilayer perceptron, and (d) bagging with a *K*-nearest neighbour regressor for 2000–2100 UTC 26 May 2018 over the USA. The FED grid uses pixels of 5 km × 5 km. (from Erdmann et al. 2022)

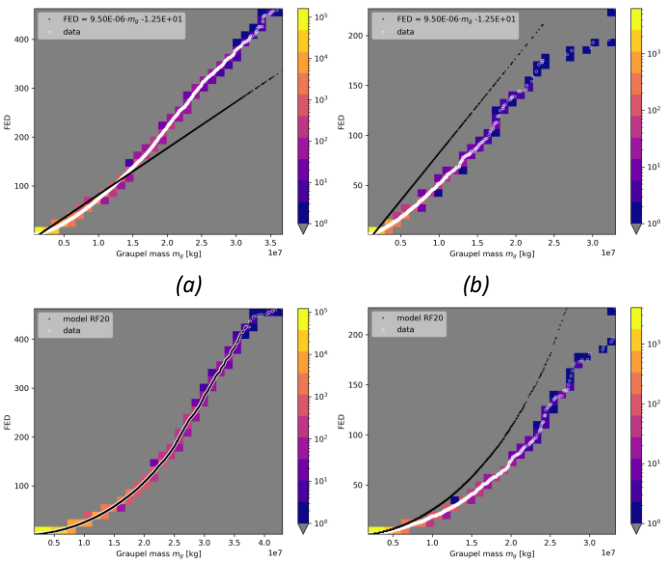

Fig. 2: The observed FED distribution versus the distribution of AROME-France graupel mass m_g at model grid points closest to each FED observation. Grid points with any of FED or m_g equal to zero are not considered. (a) shows the training of a linear regression for 24 days in 2018, and (b) shows the results of a validation for independent data of 2 additional days in 2018. Colours indicate the number of samples, the white points plot the data points, and black points applied the m_g values in the given linear regression. (c) and (d) show the same, but for a random forest regressor with 20 decision trees. (from Erdmann 2020)

Fig. 1 shows 1-h accumulated flash extent density (FED) maps observed by the Geostationary Lightning Mapper (GLM; a), while the other three panels show the corresponding FED as simulated from observations made by a ground-based lightning location system, namely the National Lightning Detection Network (NLDN) over the USA. The FED simulations are generated through a two-step approach. First, a machine-learning algorithm provides GLM flash characteristics from NLDN flash characteristics. Then, the simulated GLM flash characteristics are further processed to derive a realistic 2D distribution of pseudo-GLM optical pulses (also called events) and flash-scale products including FED on a regular grid. For the first step, a wide variety of machine-learning models have been tested. In addition, a multi-step approach has been designed, which allows correlations between features to be taken into account while keeping computational costs affordable.

Then an observation operator for FED has been designed based on a model proxy, namely the graupel mass above the −5 °C isotherm (m_g), which has been demonstrated in the literature to correlate well with total (intracloud and cloud-to-ground) lightning activity. To circumvent space-time shifts between observations and forecasts – a well-known problem especially for thunderstorms –, all proxy and observation values are first sorted and then matched, i.e., the distributions of values are matched. The results of this matching is presented in Fig. 2 for a linear regression model (1st row) and a random-forest model (2nd row), both for training (1st column) and validation (2nd column) data sets. The random-forest model clearly outperforms the linear regression model by yielding a coefficient of determination of 1.000 for the training data set (vs. 0.927 for the linear regression) and 0.719 for the validation data set (vs. 0.557 for the linear regression).

Future of the project:

A follow-up project has been funded by LEFE MANU for three additional years from 2021 to 2023. It allowed for the implementation after 2020 of actions that could not be carried out in 2020 and those results will be described in the final report of this follow-up project. This concerns in particular:

- Near-ground rain rate diagnostic from numerical weather prediction very-short-term forecasts and radar reflectivity (6-month master internship in 2021).
- Further development, sensitivity study, and validation of the observation operator for space-based lightning observations.
- Design and validation of hail diagnostic near the ground from relevant observations.

Number of publications, communications and PhD theses

Articles (2):

Erdmann, F., O. Caumont, and É. Defer, 2022: A geostationary lightning pseudo-observation generator utilizing low-frequency ground-based lightning observations. *Journal of Atmospheric and Oceanic Technology*, **39**(1) (Jan. 2022), 3–30. doi: 10.1175/jtech-d-20-0160.1.

Mandement, M. and O. Caumont, 2020: Contribution of personal weather stations to the observation of deep convection features near the ground. *Natural Hazards and Earth System Sciences*, **20**(1) (Jan. 2020), 299–322. doi: 10.5194/nhess-20-299-2020.

Conferences (2):

Erdmann, F., O. Caumont, and É. Defer, 2020: Preparation for the assimilation of the upcoming Meteosat Third Generation Lightning Imager data in AROME. *4th workshop on assimilating satellite cloud and precipitation observations for NWP* (Reading, United Kingdom, Feb. 3–6, 2020). ECMWF, JCSDA and EUMETSAT NWP-SAF.

Erdmann, F., E. Defer, O. Caumont, R. L. Holle, and S. Pédeboy, 2020: Utilizing Low-Frequency Ground-Based Lightning Locating Networks to Simulate Optical Lightning Observations of Geostationary Satellites. *100th AMS Annual Meeting* (Boston, MA, United States of America, Jan. 12–16, 2020). American Meteorological Society. url: https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/362218.

PhD theses (3):

Combarnous, P. (on-going since autumn 2020): *Maturation de l'assimilation des données de l'imageur de foudre de troisième génération de Meteosat (MTG LI) pour la prévision météorologique numérique de la convection profonde* (Maturation of Meteosat Third Generation Lightning Imager (MTG LI) data assimilation for the numerical weather prediction of deep convection).

Erdmann, F., 2020: *Préparation à l'utilisation des observations de l'imageur d'éclairs de Météosat Troisième Génération pour la prévision numérique à courte échéance* (Preparation for the use of Meteosat Third Generation Lightning Imager observations in short-term numerical weather prediction), 279 p., Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier), Toulouse, France. URL: http://thesesups.ups-tlse.fr/4947/

Mandement, M., 2020: *Apport des données d'objets connectées pour l'étude de la convection profonde à fine échelle* (Contribution of data from connected objects to the study of deep convection on a fine scale), 212 p., Institut national polytechnique de Toulouse (INP Toulouse), Toulouse, France. URL: https://oatao.univ-toulouse.fr/28318/