

Cloud Cover Nowcasting with Deep Learning

Léa Berthomier
AI Lab
METEO FRANCE
Toulouse, France
lea.berthomier@meteo.fr

Bruno Pradel
AI Lab
METEO FRANCE
Toulouse, France
bruno.pradel@meteo.fr

Lior Perez
AI Lab
METEO FRANCE
Toulouse, France
lior.perez@meteo.fr

Abstract—Nowcasting is a field of meteorology which aims at forecasting weather on a short term of up to a few hours. In the meteorology landscape, this field is rather specific as it requires particular techniques, such as data extrapolation, where conventional meteorology is generally based on physical modeling. In this paper, we focus on cloud cover nowcasting, which has various application areas such as satellite shots optimisation and photovoltaic energy production forecast.

Following recent deep learning successes on multiple imagery tasks, we applied Deep Convolutional Networks on Meteosat satellite images for cloud cover nowcasting. We present the results of several architectures specialized in image segmentation and time series prediction. We selected the best models according to machine learning metrics as well as meteorological metrics. All selected architectures showed significant improvements over persistence and the well-known U-Net surpasses AROME physical model.

Index Terms—nowcasting, cloud, meteorology, deep learning

I. INTRODUCTION

Cloud cover nowcasting aims at forecasting the position of clouds on a short time scale of up to 3 hours. Historically, the main task of weather services is to ensure the safety of life and property. While having multiple applications, cloud cover remains less critical for safety than other phenomena such as precipitation or thunderstorms which explain its relatively lower popularity in the literature. It remains a field of interest for meteorological organisations, for observation satellite management, to optimize their image shots, and for solar panels management, to forecast their production of electricity. Put simply, at METEO FRANCE, a nowcasting product is a fusion of two approaches: pure data extrapolation techniques, which perform better on the very short term (up to 1 hour), and a mix between extrapolation and classical physical models which remain the best option for further forecasting. Extrapolation methods use data observed in the near past: mainly satellite and radar images. In this paper, we will focus only on the data extrapolation part and the experiments exhibited in the paper only consider the next hour and a half. In order to study the cloud cover, we will use satellite images resulting from products of EUMETSAT, the European Organisation for the Exploitation of Meteorological Satellites.

Section II will cover the state of the art of time series prediction applied on images, which guided our choices of neural networks architectures. In Section III, we describe our methodology, metrics and choices of models. Section IV

presents a benchmark of several models on simple synthetic data. Finally, Sections V, VI, VII and VIII present our benchmark and results on satellite images, including a study on the influence of temporal depth as an input and a comparison with two meteorological models.

II. BACKGROUND AND RELATED WORK

Cloud nowcasting is mainly computed with techniques of extrapolation of pixels or objects from wind vectors. For instance [1] uses optical flow and ground-based sky imagery for short-term solar forecasting and recent work applies extrapolation of atmospheric motion vector for cloud classification forecasting using images from geostationary satellite [2]. However, these methods require long computation times when applied to large areas (around 10 minutes for a full scan disk), which, added to the data reception time, often makes the first forecasts useless. In addition, these methods only rely on the observed dynamics of the clouds, and therefore cannot predict evolution in terms of size and shape of clouds, as underlined in [3].

Deep Learning techniques, while they were only developed in the recent years, show impressive results in the field of image processing such as image classification, object detection or image segmentation, and they require very short computation times for inference tasks. Convolutional neural networks already yielded good results on optical flow computation [4] and rainfall nowcasting [5], [6], [7], [8], [9]. Both rainfall and cloud cover nowcasting aim at forecasting the future positions of moving objects and are considering time series of images as inputs and targets of computational models. While rainfall has gained a lot of attention in the recent nowcasting literature, to the best of our knowledge cloud cover nowcasting with deep learning has only been explored in the case of predictions of satellite images [10]. Consequently all rainfall deep learning nowcasting related works will be of interest.

Nevertheless, rainfall and cloud cover differ in that cumulative rainfall is a continuous variable computed on a pixel over a period of time, while cloud cover is a discrete class observed at a given moment on one pixel (see below for a description of the Cloud Type classification product by EUMETSAT). Consequently, cloud cover classification and forecasting is akin to a task of image segmentation applied to image time series. Thus, the output of a deep learning cloud cover nowcasting model will consist in several fields

of probabilities of having a cloud on each pixel, one for each time step, where rainfall nowcasting models output only the most probable value of cumulated rainfall on each pixel. This distribution of probability is an advantage of deep learning techniques over extrapolation methods, which output binary values, as it allows to take different kind of decisions depending on the problem at hand.

Finally, cumulated rainfall is measured by doppler radar every 5 minutes while cloud cover classification results from satellite images and are collected on longer time intervals, ranging from 5 to 15 minutes depending on the satellite, even if they offer larger views. For instance, Meteosat Second Generation 4 (MSG4) captures a full scan disk of the planet each 15 minutes. This difference in time step might have consequences on the performance of the models and raises the question of how much temporal depth should be given to the model as an input.

III. METHODOLOGY

Our goal was to forecast the position of clouds on a short time scale of 1h30' from satellite images using deep convolutional networks as our models. To this end, we used the following data and methodology.

A. Data

Our data come from the "Geostationary Nowcasting Cloud Type" classification product of EUMETSAT [11], classifying clouds in 16 classes depending on their height and type. This classification is computed with various visible and infrared channels images shot by Meteosat Second Generation (MSG), a geostationary satellite at longitude 0 degree. The result is a 3712 by 3712 pixels image computed every 15 minutes and indicating the cloud type of each pixel. Our dataset was composed of images from 2017 and 2018.

B. Experimental protocol

Our long term objective is to forecast the class, position and height of the clouds for each pixel over the whole globe. However, for the experiments presented in this paper, we focused on forecasting only the position of the clouds, by using binary images of cloud cover as our input data : the value 1 indicates a cloud on a pixel and the value 0 indicates a cloudless pixel. Due to computation issues, we also cropped and projected the satellite images to keep only squares of size 256 by 256 pixels showing France. The final resolution of the images was 4.5 kilometers and each couple of images was spaced by 15 minutes.

The input data of the model consisted of 4 binary images (spanning 1 hour) and our goal was to forecast the cloud cover for the next 1h30. Consequently, the output of the model consisted of 6 images of values ranging between 0 and 1 and representing the probability of having a cloud on each pixel. Figure 1 shows an example of a sequence from the dataset.

Therefore, in order to train our models, we constituted a dataset of sequences of 10 binary images, spanning 2h30. Data from 2017 and the first semester of 2018 were used as training

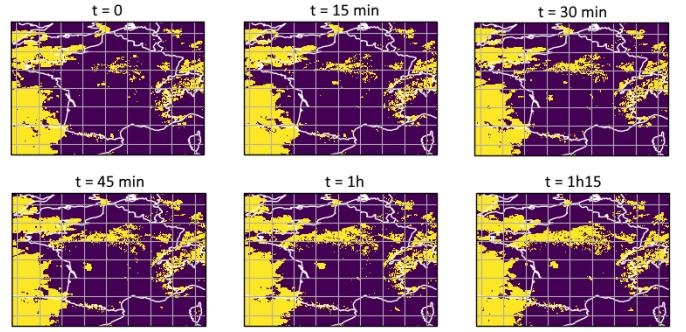


Fig. 1. Example of 6 images from a sequence of the test set.

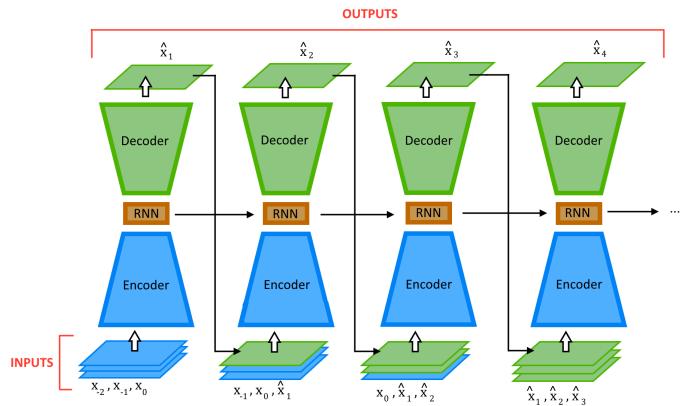


Fig. 2. Diagram of the recurrent network's principle : The x_i are the model's input and the \hat{x}_i are the desired output. The output of each decoder is used as input to compute the next time step's forecast. Information between time steps is communicated through the RNN's output.

set and the second semester of 2018 was used as validation set.

Each of the models was trained 3 times on the dataset during 20 epochs, meaning that each sample in the dataset was seen 20 times by the network. All the following metrics were computed as means of the 3 runs.

C. Metrics

Once the models were trained with the data, we computed the mean squared error (MSE) of the models for each of the 6 output time steps. We compared these errors to the MSE of the persistence, a common meteorological baseline used in nowcasting, which consists in using the last input as prediction for all the output time steps. Indeed, in short term forecasting, the last observation is often the best prediction.

In addition, we also computed the MSE over the binarized outputs of the model : each value of the output images was rounded to 0 if it was lower than 0.5 or 1 if it was greater than 0.5, before computing the MSE.

D. Neural networks architectures

We built 20 different models based on the following 4 types of architecture, with several variations like the number of

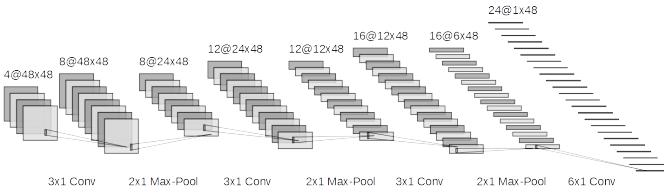


Fig. 3. Reduce-LSTM - Principle of the dimension reduction.

convolutional layers and adding residual layers or inception layers:

- **Convolutional networks** with stacked convolutional layers of varying kernel size.
- **U-Net networks**, implementing the work of [12] which specializes in image segmentation, as well as variations from this architecture (adding or removing convolutional layers, removing the residuals, changing the activation function, adding fully-connected layers, ...).
- **Recurrent networks** based on the diagram on Figure 2, with variations in the architecture of the encoders and decoders.
- **LSTM networks** adapted to our data type. As we deal with time-series of 2D data, we couldn't fit them directly into a LSTM which expects 1D multi-channel data as inputs. We therefore propose the following architecture that we call **Reduce-LSTM**. We consider time as a channel and reduce the 2D input tensors of shape $X * Y * 4$ to a 1D multi-channel tensors of size $Y * 256$ with an encoder composed of (3,1) and (2,1) kernels for the convolutional and max pooling layers. This dimension reduction is illustrated in Figure 3. We then repeat this process on the input after swapping the X and Y axes, we stack the two resulting 1D multi-channel tensors, and use this as an input for a traditional LSTM.

IV. BENCHMARK ON SIMPLE MOVING SHAPES

We started our study by focusing on synthetic images of simple moving shapes in order to check that our models could perform a simple task of time series prediction. The dataset was made of 5000 sequences of 10 images showing simple shapes like squares or circles following randomly generated rotations and linear movements. We made six datasets by varying the speed, transparency and size of the shapes, in order to test the limits of the models.

Each of the 20 models was run 3 times on each of the 6 datasets. Figure 4 shows the average mean squared error (MSE) for the best model of each type and for each time step. The MSE over the binarized outputs is also represented with dashed lines. We can see that each model performs better than the persistence and that the prediction gets worse as the prediction time step is further in time, which was to be expected. Our benchmark shows that the original U-Net model performed better than the other models, with a MSE at 17 percent of the persistence, and a MSE over binarized outputs at 21 percent of the persistence.

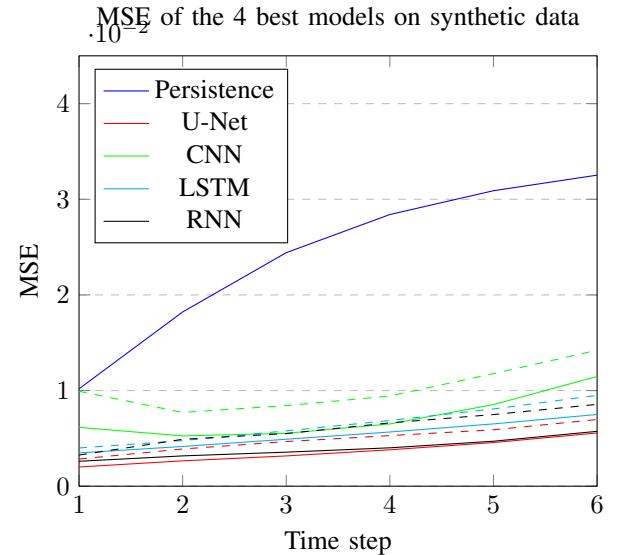


Fig. 4. Average MSE of each of the 4 best models and the persistence for each time step of the prediction. The dashed lines represent the MSE computed on binarized outputs.

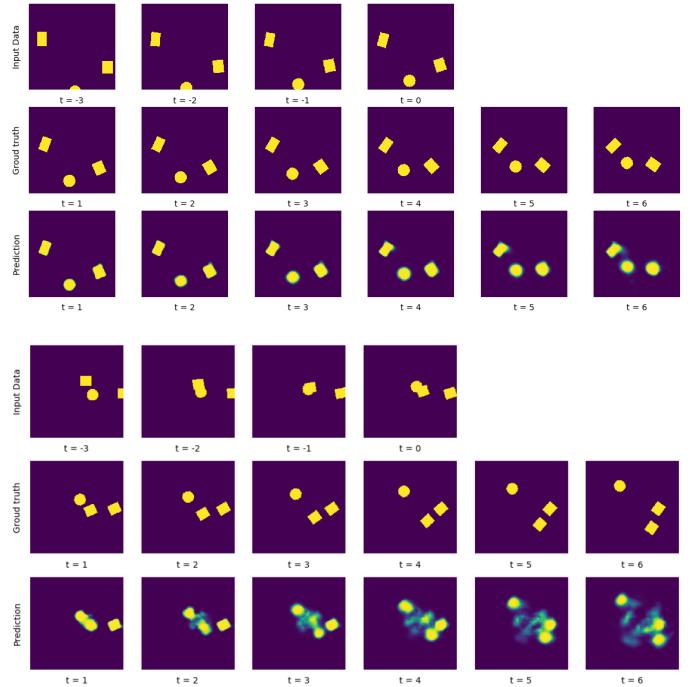


Fig. 5. Two examples of U-Net's predictions : the first line is the input data, the second line is the ground truth and the third line is the output of the model.

When examining the outputs of the models, this first study shows promising results : Figure 5 shows predictions made by the U-Net model, using input from the test set. The network manages to make good predictions of the future position of the object and its shape. It even does a fairly good job when objects overlap in the input images, making them difficult to distinguish.

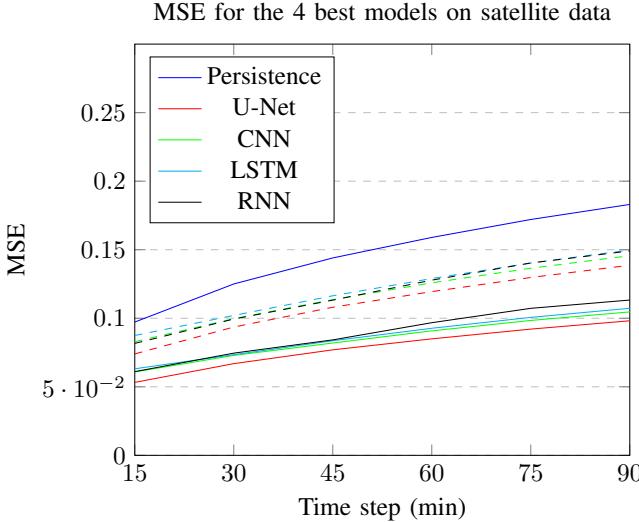


Fig. 6. Average MSE of each model and the persistence for each time step of the prediction. The dashed lines represent the MSE computed on binarized values.

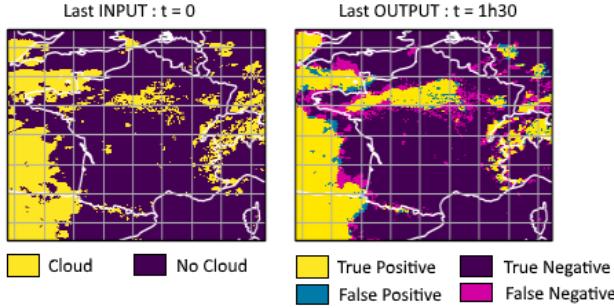


Fig. 7. Example of prediction made by the U-Net architecture. Left figure: last input image received by the model. Right figure: last output of the model.

This study on synthetic data also reveals that the main source of error of the predictions is the architecture of the neural networks. Indeed, the exact future position of the shapes can be theoretically computed from the last two input images, at least when the shapes don't overlap. The error cannot be attributed to a weak signal in the inputs.

V. BENCHMARK ON REAL SATELLITE DATA

Based on the first study on synthetic data, we selected four models, the best of each category, and tested them on the satellite data described in section III.

Figure 6 shows the average MSE at each time step for each model, as well as the MSE computed over binarized images, compared to the persistence. We can see that each model performed better than persistence, even on the first prediction, and that the U-Net got an error 53 percent smaller than persistence. In addition, the MSE on binarized outputs is 75 percent smaller than persistence. All models managed significant improvement of persistence and were able to learn to predict the movement of cloud covers. Figure 7 shows an

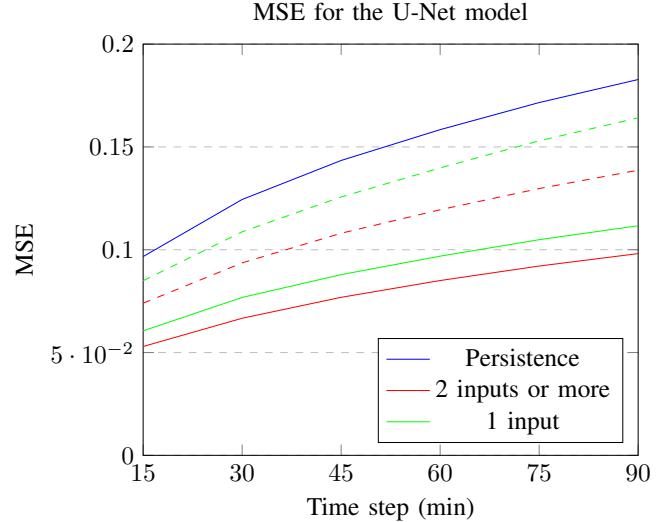


Fig. 8. Average MSE of the U-Net model and the persistence, with different number of input images, for each time step of the prediction.

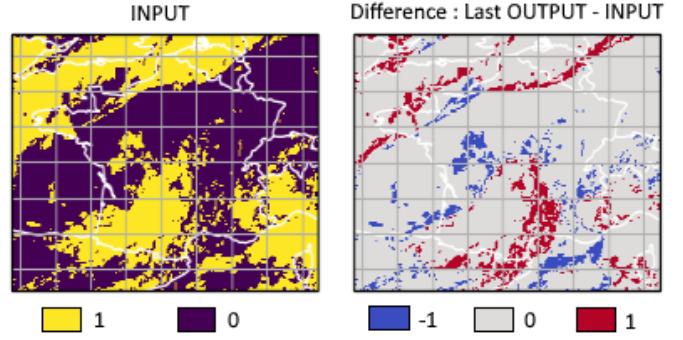


Fig. 9. Single input experiment: the left image represents the last input given to the model; the right image is the difference between the last output ($t+1h30$) and the input. Appearing clouds are represented in red and disappearing clouds are in blue.

example of prediction made by the U-Net model, comparing the last input of the model and its last prediction.

We can see that stationary clouds (East and South-West) are correctly predicted. In addition, the global movement of clouds is well represented : the North-West clouds are correctly moving to the North. Finally, the cloud appearing in the North is well approximated despite its blurry edges.

VI. TEMPORAL DEPTH OF INPUTS

Then, we studied the influence of the number of input images on the performance of the model. For this experiment, we only studied the U-Net model, as it performed better than the others on the previous experiments. The idea was to check if the model could use older input data to make better predictions, or if the last four images considered so far were enough to predict the future positions of clouds.

To this end, we varied the length of input data by changing the number of input images given to the model, which ranged between 1 and 6, corresponding to 15 minutes to 1h30 of

temporal depth. This experiment showed that using more than 2 input images did not offer any benefits compared to using only 2 inputs: the variations in the MSE and in the output images were not significant. This result supports the conclusions of [8] on rainfall nowcasting which shows that reducing the input's temporal context from 90 minutes to 30 minutes does not affect the performances significantly.

In addition, using only one input image still results in a better performance than persistence in terms of MSE, even when the model shouldn't have any indication of cloud movements. Figure 8 represents the evolution of MSE regarding the number of input images considered and figure 9 shows an example of output from the U-Net trained with only one input image. We can see that the model added clouds on the east of existing clouds (in red), which tends to make them move to the east. As the model was only trained on France area, our hypothesis is that the model learned a part of the local climatology of France where clouds are mostly moving from west to east. Therefore, the model can predict the most probable movement of cloud masses even with only one input image.

VII. SIZE AND SHAPE EVOLUTION

A well known weakness of data extrapolation techniques is their difficulty to model size and shape evolution of clouds. Extrapolation techniques shift existing clouds, tending to preserve initial dimensions. The ability of a model to take into account this attribute of clouds is very hard to measure accurately but simply observing the image outputted, we can get an idea. As shown in Figure 10, we extracted multiple sequences in which deep learning models increase or decrease the size of an existing cloud. However, new clouds are not created by our architecture (which would have been surprising).

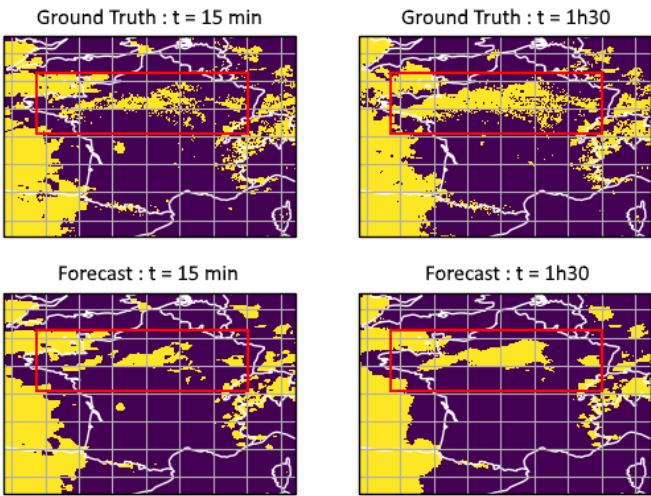


Fig. 10. Ground truth: in the area enclosed by a red line, we observe a cloud increasing in volume between +15min and +90min. Forecast: We exhibit U-net predictions at the corresponding times. The model correctly enlarges the surface of the cloud.

VIII. COMPARISON WITH METEOROLOGICAL MODELS

Finally, we compared the performance of the U-Net network and persistence with two other meteorological models :

- **EXIM** : Eumetsat image extrapolation tool [2] which uses high resolution atmospheric wind vectors to forecast the position and type of clouds up to 2h30 over the whole Meteosat disk, with a time step of 15 minutes.
- **AROME** : METEO FRANCE non-hydrostatic very-high-resolution model [13] [14], based on physical modelling, which forecasts several physical parameters up to 48 hours over the French territory, with a time step of 1 hour.

It should be noted that EXIM's forecasts are available 20 minutes after the beginning of the run, and AROME's are available after an hour. Thus, the first time steps of the forecasts are unusable. Meanwhile, the inference of one sample with the U-Net model can be computed in about 20 seconds on an average laptop, and persistence's forecast are available instantly, therefore all of their forecasts can be used in practice.

For this comparison, as we want to compare the performances of the models at equivalent runs, the forecasts are considered available immediately after the beginning of the run, which is favorable to EXIM and AROME.

In addition, the domain of this comparison is still the French territory but the time span is February to May 2019, as EXIM's data was not available on a larger span at the time of the study.

Figure 11 represents the evolution of the MSE for each of the 4 models. The MSE over the binarized outputs is also represented with a dashed line. Our comparison shows that the U-Net model performed better than the other models. In addition, we can notice that persistence surpasses EXIM after 45 minutes, which shows the limitations of classical extrapolation methods. As expected, AROME's MSE is worse

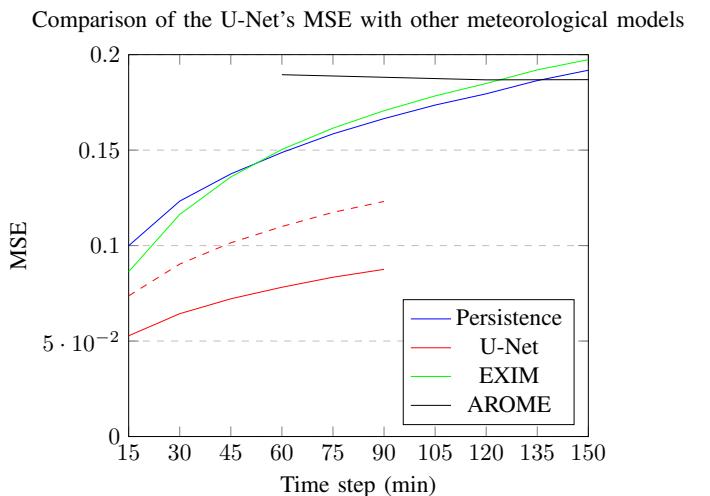


Fig. 11. Average MSE of the U-Net model, persistence, EXIM and AROME for each time step of the forecast. The dashed line represent the MSE computed on binarized values.

than the other models for the first time steps, but it's MSE is slowly decreasing and becomes better than EXIM and the persistence after 2 hours. Indeed, as we could expect from a physical model, AROME shows its strength on mid to long term forecasts, when extrapolation methods are no longer relevant.

IX. CONCLUSION AND FUTURE WORK

We provided results of multiple deep learning architectures on cloud cover nowcasting, more precisely on the data extrapolation task every 15 minutes and up to an hour and a half. All architectures proved to outperform significantly persistence at each step, in terms of MSE. So far, U-net architecture has proven to be the best for this task and we didn't find any custom architecture to surpass it. While persistence is a baseline of interest in meteorology, U-Net also surpasses AROME, METEO FRANCE non-hydrostatic very-high-resolution model, and EXIM, Eumetsat image extrapolation tool, while demanding significantly less computation time.

As of today, our neural network architectures are not capable of predicting the emergence of new clouds. This limitation is shared by all pure data extrapolation techniques to the best of our knowledge. Adding other data sources in the hope of overcoming this limitation will be the subject of another study. We also plan to use user-tailored metrics for satellite or solar panels applications.

Finally, it will be of interest to try extending the length of the forecasts and measure when AROME's physical modelling surpasses deep learning techniques, in the hope that deep learning could eventually perform the final bridge between very short-term and mid-term forecasting.

ACKNOWLEDGMENT

We would like to thank V. Chabot, M. Claudon, C. Jauffret, G. Larvor and M. Sorel from METEO FRANCE for their help and advice. We also thank EUMETSAT for the production and availability of satellite images and the *Centre de Météorologie Spatiale* for the Cloud Type product. Finally, we are deeply grateful to the *Fonds de Transformation de l'Action Publique* (French government program for public services modernization) whom funding allowed the genesis of the Artificial Intelligence Laboratory of METEO FRANCE.

REFERENCES

- [1] Wood-Bradley, P., Zapata, J., and Pye, J. (2012). Cloud tracking with optical flow for short-term solar forecasting. Solar Thermal Group, Australian National University, Canberra, Australia.
- [2] EUMETSAT NWCSAF (2019). Algorithm theoretical basis document for the extrapolated imagery processor of the NWC/GEO. http://www.nwcsaf.org/documents/27295/0/NWC-CDOP2-GEO-ZAMG-SW-ATBD-EXIM_v2.1.pdf
- [3] Papadakis, N., Héas, P., and Mémin, E. (2007). Image assimilation for motion estimation of atmospheric layers with shallow-water model. In 8th Asian Conference on Computer Vision (ACCV 2007).
- [4] Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., and Brox, T. (2017). Flownet 2.0: Evolution of optical flow estimation with deep networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2462-2470).
- [5] Xingjian, S. H. I., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Advances in neural information processing systems (pp. 802-810).
- [6] Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D. Y., Wong, W. K., and Woo, W. C. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In Advances in neural information processing systems (pp. 5617-5627).
- [7] Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C. and Hickey, J. (2019). Machine learning for precipitation nowcasting from radar images. Computer Vision and Pattern Recognition.
- [8] Sønderby, C. K., Espeloh, L., Heek, J., Dehghani, M., Oliver, A., Salimans, T., ... and Kalchbrenner, N. (2020). MetNet: A Neural Weather Model for Precipitation Forecasting. arXiv preprint arXiv:2003.12140.
- [9] Ayzel, G., Scheffer, T., and Heistermann, M. (2020). RainNet v1. 0: a convolutional neural network for radar-based precipitation nowcasting. Geoscientific Model Development, 13(6), 2631-2644.
- [10] Hong, S., Kim, S., Joh, M., and Song, S. K. (2017). Psique: Next sequence prediction of satellite images using a convolutional sequence-to-sequence network. arXiv preprint arXiv:1711.10644.
- [11] EUMETSAT NWCSAF (2019). Algorithm theoretical basis document for the cloud product processors of the NWC/GEO. http://www.nwcsaf.org/Downloads/GEO/2018/Documents/Scientific_Docs/NWC-CDOP2-GEO-MFL-SCI-ATBD-Cloud_v2.1.pdf
- [12] Ronneberger, O., Fischer, P., and Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
- [13] Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., ... and Masson, V. (2011). The AROME-France convective-scale operational model. Monthly Weather Review, 139(3), 976-991.
- [14] Brousseau, P., Seity, Y., Ricard, D., and Léger, J. (2016). Improvement of the forecast of convective activity from the AROME-France system. Quarterly Journal of the Royal Meteorological Society, 142(699), 2231-2243.