Data assimilation emulation for AROME using SEVIRI observations only

Ariadna Perelló Achfari, Vincent Chabot, Loïk Berre, Thomas Carrel-Billiard, Olivier Audouin

Groupe de Modélisation et d'Assimilation pour la Prévision (GMAP), Centre National de Recherches Météorologiques (CNRM), Météo-France, Toulouse, France

ariadna.perello-achfari@meteo.fr

Introduction

• A NWP model $\mathcal{M}$ needs an estimate of the atmospheric state to launch a forecast for the next time step. Data assimilation produces this estimate, referred to as the analysis vector $\boldsymbol{x^a}$ and defined by Eq. 1.

$$\boldsymbol{x^a} = \boldsymbol{x^b} + \boldsymbol{\delta x} \approx \boldsymbol{x^b} + \mathbf{K}(\boldsymbol{y^o} - \mathcal{H}(\boldsymbol{x^b})) \quad \text{with} \quad \mathbf{K} = \mathbf{BH^T}(\mathbf{HBH^T} + \mathbf{R})^{-1} \tag{1}$$

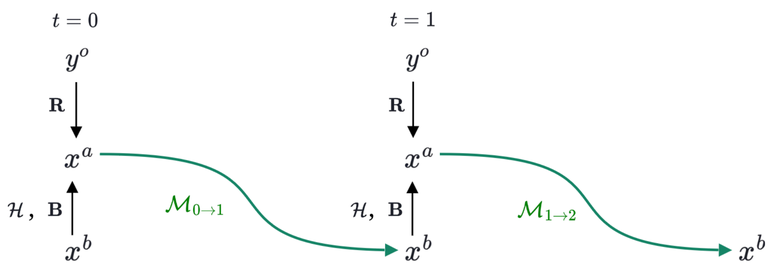


Figure 1 – Principle of data assimilation cycling

where :

- $\boldsymbol{x^b}$ background vector
- $\boldsymbol{y^o}$ observation vector
- $\boldsymbol{\delta x} = \boldsymbol{x^a} - \boldsymbol{x^b}$ analysis increment
- $\mathbf{H}$ tangent-linear of the observation operator $\mathcal{H}$
- $\mathbf{R}$ observation error covariance matrix
- $\mathbf{B}$ background error covariance matrix

• Deep learning techniques have shown some first promising results in emulating either whole data assimilation or some of its components, offering lower computational costs and a higher modeling flexibility than conventional approaches.

The purpose of this work is to emulate the data assimilation step in Météo France's regional model AROME using the SEVIRI imager observations.

Supervised learning approach

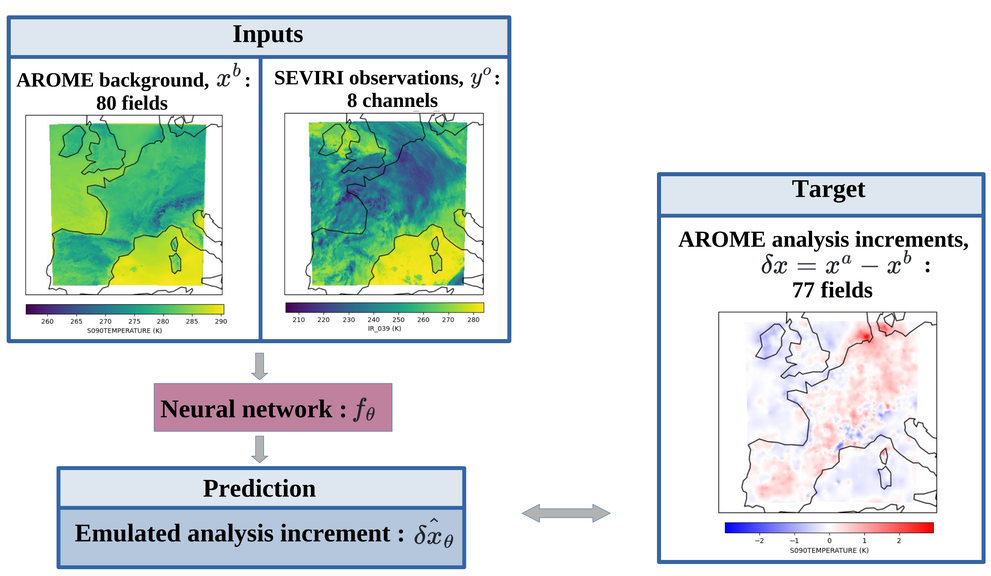


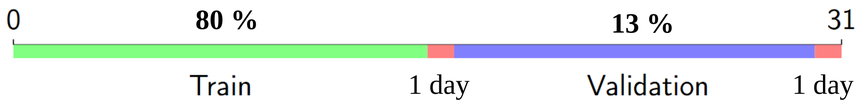
Figure 2 – Diagram of the data assimilation emulator, inspired by [1]. Given the background $\boldsymbol{x^b}$ and the observations $\boldsymbol{y^o}$, the model f_θ retrieves the predicted analysis increment $\boldsymbol{\delta\hat{x}_\theta} = f_\theta(\boldsymbol{x^b}, \boldsymbol{y^o})$ by minimizing the distance to the target $\boldsymbol{\delta x}$. θ is the set of neural network parameters.

Training data

	AROME background/analysis	SEVIRI observations
General features	▪ 3D-Var ▪ Obtained from all the observations assimilated in AROME	▪ SEVIRI : Imager on board a geostationary satellite ▪ Brightness temperatures (IR and visible domains)
Resolution	▪ Horizontal resolution : 5,2 km ▪ Vertical resolution : 19 model levels	Horizontal resolution : 5,2 km (after re-interpolating on AROME grid)
Variables/ channels	▪ Surface pressure (sp) ▪ Temperature (T) ▪ Zonal (u) and meridional (v) winds ▪ Specific humidity (q)	▪ 6 temperature channels ▪ 2 water vapor channels

• Training, validation and test datasets

- Period : from 15/02/2021 to 14/10/2024, with a one hour time step
- Training and validation datasets : years 2021, 2022, 2023. For each month :



Model architecture

The model used is the MFAI [2] library's **UNETR++** (Fig. 3), a Vision Transformer architecture adapted from [3]. The model has approximately 15 000 000 trainable parameters.

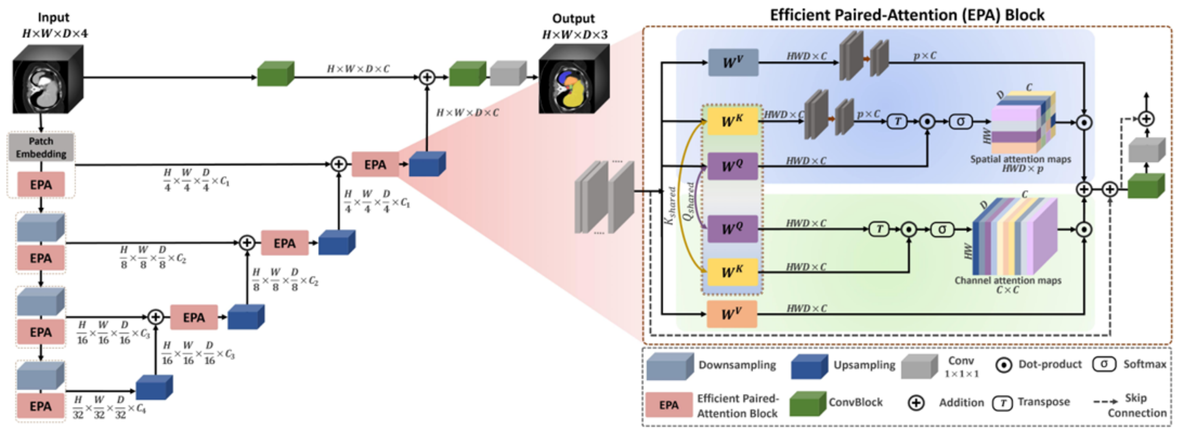


Figure 3 – UnetR++ architecture [3]

Experiments and results

• Experiments

Sensitivity experiments were conducted by varying the training hyperparameters, the input data and the neural network's architecture hyperparameters. The model was most sensitive to the hidden size, which was set to 512.

• Results

The resulting model was evaluated on the validation dataset. The network performs better at the pressure levels corresponding to SEVIRI channels sensitivity peaks (Fig. 4) : near the surface and between 500 and 300 hPa for specific humidity, and near the surface for temperature (Fig. 5). The model reproduces about 25% of the total analysis signal in the lowest levels for temperature.

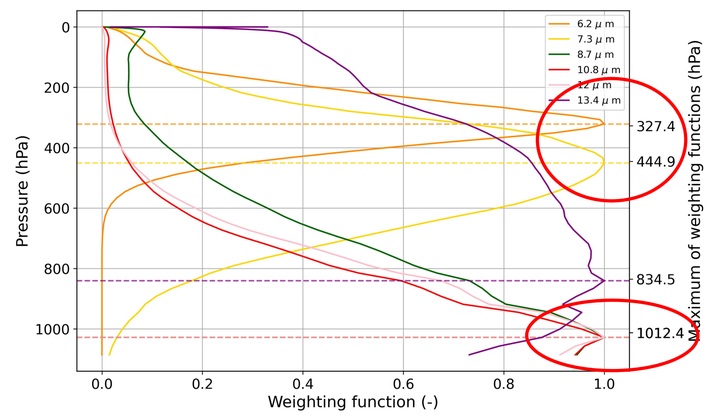


Figure 4 – SEVIRI channels weighting functions

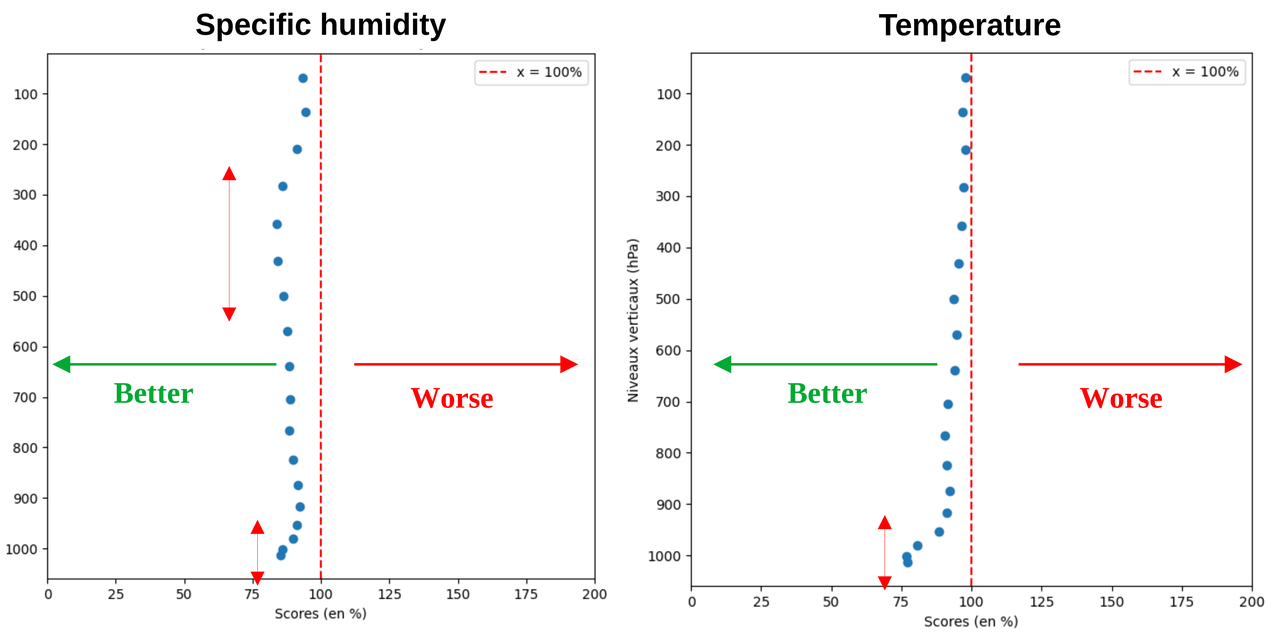


Figure 5 – Relative RMSE mean vertical profiles averaged over all the validation samples, for specific humidity (left) and temperature (right)

The temperature and specific humidity increments (Fig. 6) show that large scale structures are well represented, with strong spatial correlations between predictions and targets. However, fine scale features are smoothed.

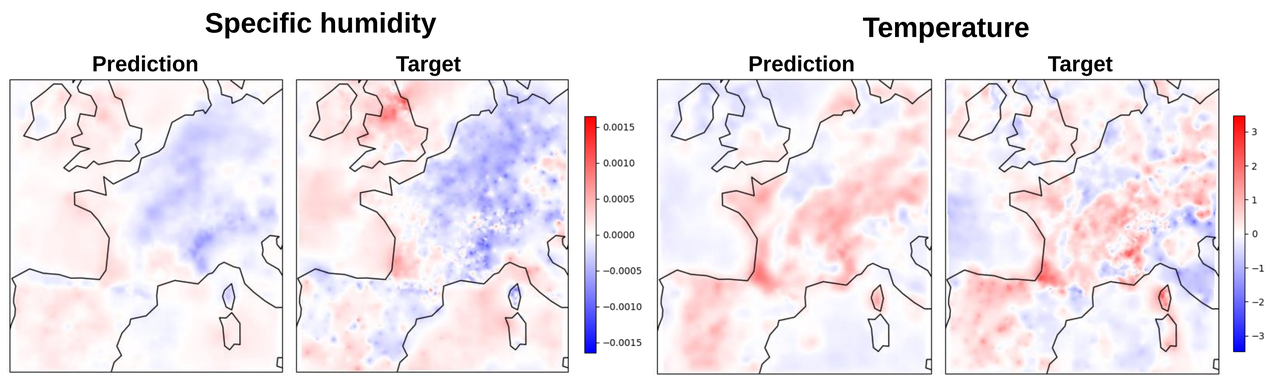


Figure 6 – Specific humidity (left) and temperature (right) increments at 1012 hPa valid on 23/02/2021 at 9 a.m.

Conclusions and prospects

• Conclusions

The model exhibits a strong physical consistency with the observations :

- it performs better at the layers of the atmosphere where SEVIRI channels are the most informative.
- temperature and specific humidity increments are better predicted than wind and surface pressure increments. Indeed, SEVIRI observations mostly provide information about temperature and water vapor variations.
- SEVIRI observations deliver large scale information, used by the network to retrieve most of the large scale signal.

• Future prospects

- Compare predictions with a SEVIRI-only AROME analysis.
- Feed more observations and/or consecutive time steps into the model.
- Increase the training dataset size.

We would like to thank the **Machine Learning Pilot Project** for supporting this work.

References

[1] Eric S. Maddy, Sid-Ahmed Boukabara et Flavio Iturbide-Sanchez. "Assessing the Feasibility of an NWP Satellite Data Assimilation System Entirely Based on AI Techniques". In : IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17 (2024), p. 9828-9845. url : https://api.semanticscholar.org/CorpusID:269620706.

[2] Météo-France et al. MFAI. url : https://github.com/meteofrance/mfai.

[3] Abdelrahman Shaker et al. UNETR++ : Delving into Efficient and Accurate 3D Medical Image Segmentation. 2024. arXiv : 2212.04497 [cs.CV]. url : https://arxiv.org/abs/2212.04497.