Deep Learning for regional ensemble forecasting : first results

Laure Raynaud and contributors

MAELSTROM workshop, 7 November 2023

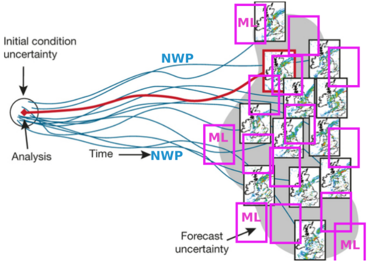
DL for ensembles

▷ Due to their high numerical cost, ensembles are the perfect target for DL application

▷ Size and resolution of operational ensemble forecasts are still constrained by computational resources

▷ We've started exploring DL potential for 2 tasks

- Oversampling of NWP distributions



- Statistical downscaling

▷ With application to the regional Arome forecasts

- ▷ Operated by Météo-France since 2016 over Western Europe
- ▷ 2.5km resolution from 2016 to 2022, 1.3km resolution since 06/2022
- ▷ 17 members (including control), 4 times per day

▷ **Objective** : design a hybrid physics/AI Arome-EPS with $\mathcal{O}(1000)$ members at hectometric resolution ($\sim$ 500m)

▷ Minimum requirements :

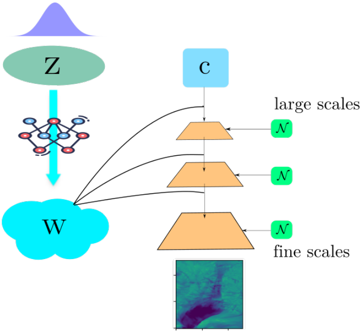
- AI-generated members are physically consistent
- AI-generated members improve probabilistic performances
- Ideally, hybrid physics/AI EPS comparable to a large purely physics-based EPS

Plan

1. DL for ensemble oversampling : methodology
2. DL for ensemble oversampling : performance results
3. DL-based statistical downscaling
4. Conclusions and future works

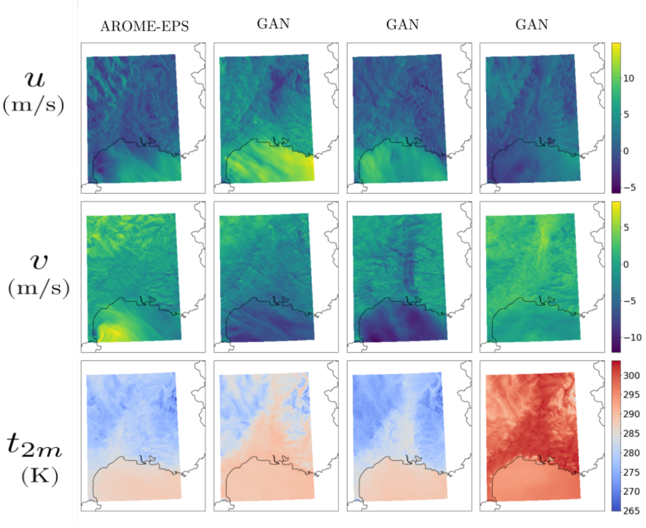
Step 1 : DL to generate NWP-like samples

- ▷ Generative DL is well-adapted : VAE, GAN, diffusion models
- ▷ The StyleGAN2 architecture is appealing (Karras *et al.*, 2019)
  - disentangled latent space
  - different style vectors injected at each layer, allowing for more control on the attributes of generated images



- ▷ Trained on a 18-month archive of Arome-EPS forecasts
- ▷ Start with generation of 2m-temperature and 10m-u/v wind.

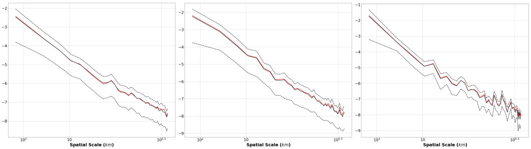
1 - Unconditional forecast generation : examples



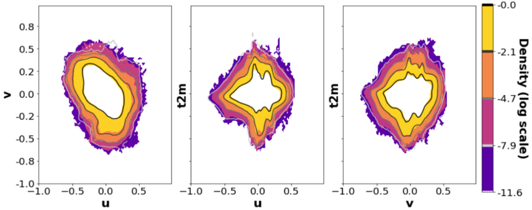
▷ Generated samples look great !

1 - Unconditional forecast generation : evaluation

▷ PSD Arome (black) vs GAN (red) for u10, v10 and T2m



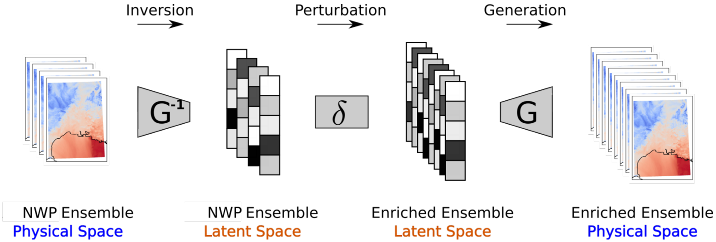
▷ Bivariate distributions Arome (contours) vs GAN (color)



- The GAN accurately learns the distribution of Arome forecasts
- The GAN produces samples with proper physical/spatial consistencies
- Detailed evaluation in Brochet *et al.* 2023 (AI4ES).

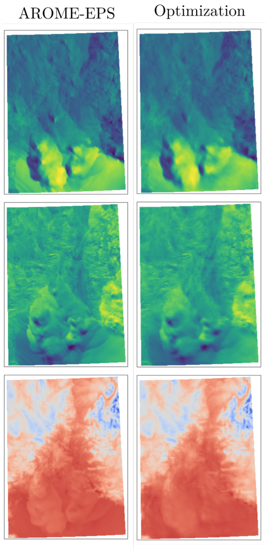
Step 2 : building on one or several NWP forecasts to produce new ones

▷ Easy to edit a sample generated by StyleGAN by modifying some of its style vectors
▷ To modify a real (NWP) sample we need to approximate its latent vector : this is called StyleGAN inversion
▷ Latent-space sampling : perturb 'latent' NWP forecasts



▷ Can work with only a single deterministic forecast as input
▷ Can generate as many members as we want

1 - Conditional forecast generation : StyleGAN inversion

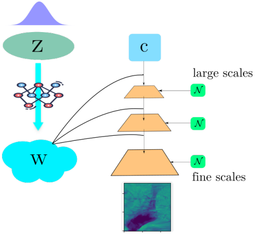


▷ StyleGAN inversion is achieved with an optimization method, that iteratively refines a latent code by minimizing the reconstruction error

▷ Mean-square regression with gradient descent using *frozen generator*

$$\mathbf{w}_* = \arg\min_{\mathbf{w}} \|\mathcal{G}(\mathbf{w}) - X\|^2$$

How to perturb the style vectors ?



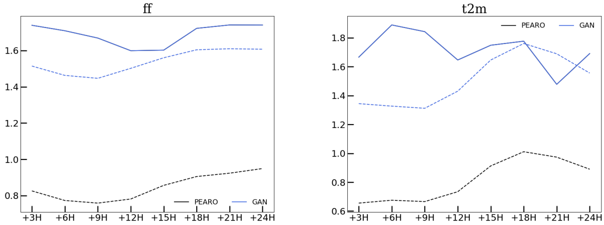
- ▷ Different style vectors are injected at each layer
- ▷ One can perturb style vectors on 'relevant' layers while keeping the others unchanged
- ▷ The set of 'relevant' layers to perturb can be optimized based on probabilistic scores.

Plan

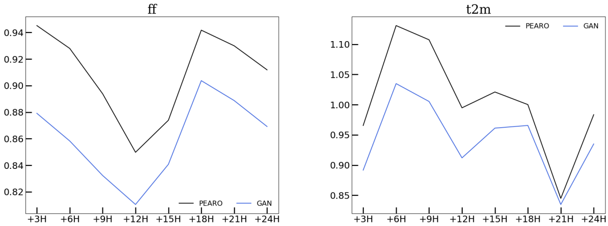
1. DL for ensemble oversampling : methodology
2. DL for ensemble oversampling : performance results
3. DL-based statistical downscaling
4. Conclusions and future works

2 - Evaluation of DL ensembles

▷ DL-enhanced ensemble (120 mb) significantly outperforms the operational Arome ensemble (Courtesy : G. Moldovan)

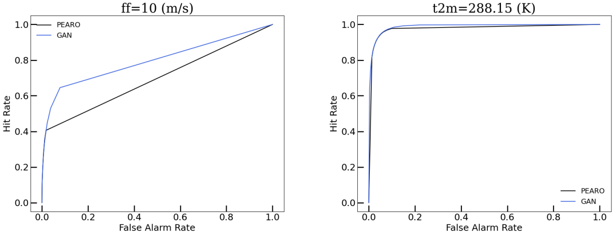


(a) Spread/skill

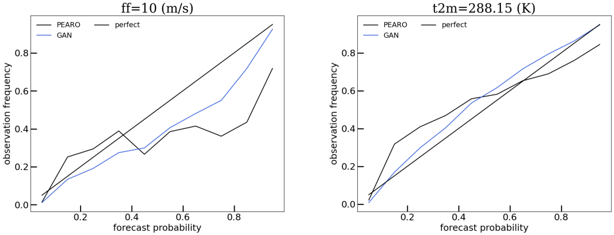


(b) CRPS

2 - Evaluation of DL ensembles



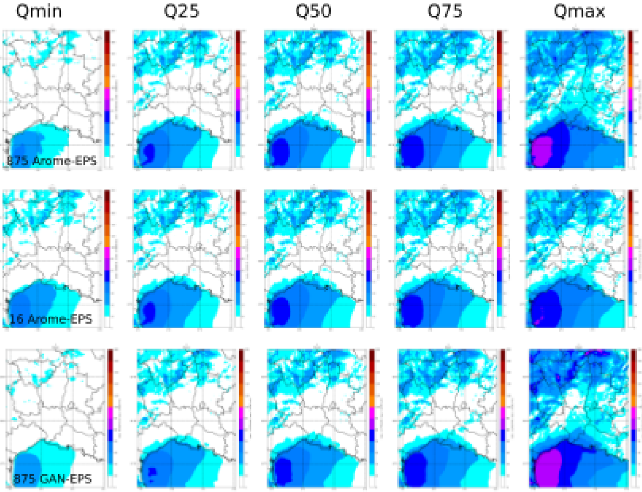
(a) ROC



(b) Reliability

2 - Evaluation of DL ensembles

▷ Comparison of DL-enhanced ens to a large 875-mb Arome ens



▷ The DL-enhanced ensemble properly extends the tails of the distribution while preserving the main part.

2 - Future works, challenges and open questions

▹ The latent-space sampling provides both **improved performances** and **physically-consistent members**

▹ The method is based on a latent representation of atmospheric fields, no need for an emulator of atmospheric dynamics

▹ It can be applied to all forecast types : deterministic, ensemble, physically-based, AI-based, ...

▹ Extension to other variables under investigation, e.g. precipitation

▹ Benefit of DL ensembles for high impact events to be explored

▹ Examine sensitivity to the number of conditioning forecasts

▹ The method does not address bias correction (no obs), but standard post-processing methods could be applied to the DL ensembles

▹ Other generative DL approaches could be used : we found **diffusion models as skillful as GANs** (for unconditional generation)

Metric	W_1	SWD	$PSD_{err,u}$	$PSD_{err,v}$	$PSD_{err,t2m}$
StyleGAN	5.3	8.0	**0.61**	**0.78**	0.75
DDPM	**3.5**	**7.5**	1.01	0.83	**0.4**

(Courtesy : J. Rabault)

▹ Next step will be **conditional diffusion**, following for instance Google paper Li *et al.*, 2023 : SEEDS Emulation of Weather Forecast Ensembles with Diffusion Models

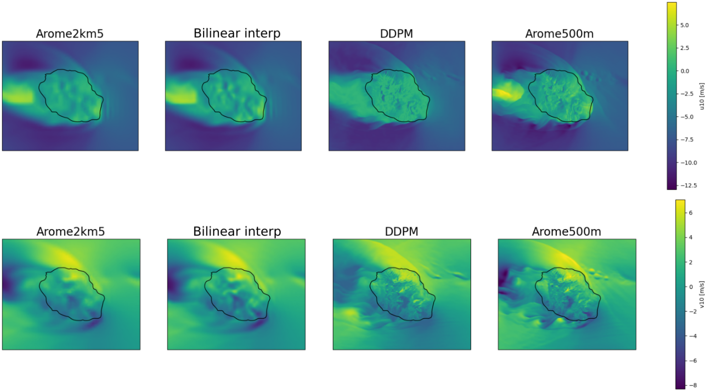
▹ Comparison with **probabilistic emulators** (when available)

Plan

1. DL for ensemble oversampling : methodology
2. DL for ensemble oversampling : performance results
3. DL-based statistical downscaling
4. Conclusions and future works

3 - DDPM for high-res wind forecasts

▷ Denoising Diffusion Probabilistic Models (DDPM) can also be used for super resolution
▷ Application to *downscale wind forecasts* from Arome 2.5km to Arome 500m (Courtesy : L. Danjou)

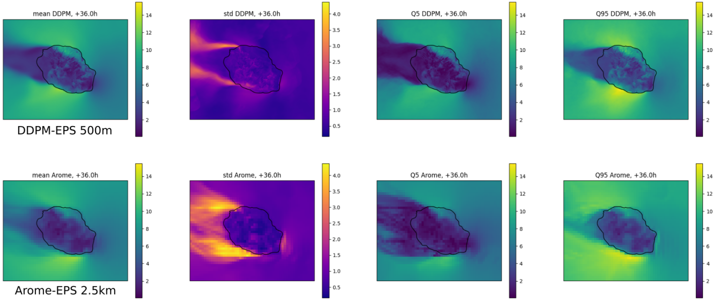


▷ DDPM better at capturing the spatial structure than the intensity
▷ DDPM better than standard CNNs (e.g., UNet)

3 - DDPM for high-res wind ensemble forecasts

▷ Ensembles can be easily generated with DDPM

▷ *Example* : generation of a 128-mb ensemble of 500m forecasts, conditioned only on the deterministic 2.5km forecast



▷ DDPM spread has some similarity with Arome spread, but it is smaller

▷ To be continued, including a comparison to GAN.

▷ We are still in an exploratory phase, but DL already confirms its potential to significantly enhance ensemble size and resolution in a realistic fashion

▷ Further steps will include :

- demonstrating the relevance of the approaches developed using other regional and global NWP forecasts
- further exploration of the potential of DDPM (including ways to reduce their cost)
- subjective evaluation of DL-enhanced ensembles by end-users

▷ The development of an Arome emulator is another big challenge