About Machine Learning to explore and enhance ensemble weather forecasts : an overview of research activities at CNRM

Laure RAYNAUD
Joint work with C. Brochet, L. Danjou, G. Moldovan and A. Mounier
Météo-France, National Center for Meteorological Research

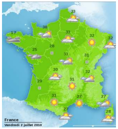
AI4AOC, 12 April 2023

1 - Context : operational weather forecasting

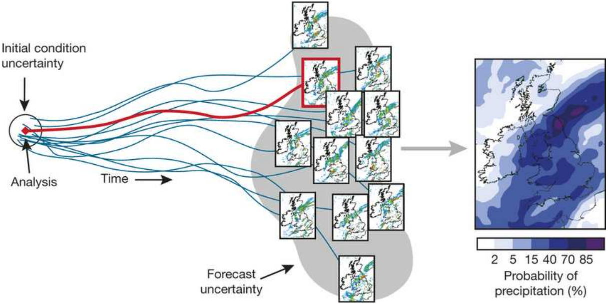
▷ Météo-France daily runs several tens of weather forecasts for the next days using **Numerical Weather Prediction (NWP) models**

Physical laws $\Rightarrow$
Initial state t_0 $\Rightarrow$

$\Rightarrow$ **Forecasts at** $t_0 + \delta t$



▷ Due to the chaotic nature of the atmosphere **ensemble forecasts** are used to predict the pdf of the future atmospheric state (instead of a single state)



▷ Current NWP forecasts are both :

- **too numerous to be rapidly and efficiently analysed by human brains**
  ⇒ this calls for tools that automatically extract the relevant information, targeted to end-users needs
- **not enough numerous and accurate to anticipate high-impact weather events whose predictability is low**
  ⇒ very large samples of high-resolution forecasts are needed to identify low-probability events, but producing more forecasts with NWP models is too expensive

▷ ML may be a solution to address these 2 problems.

Plan

1. Introduction
2. Exploring ensembles with ML
3. Enhancing ensemble forecasts with ML

2 - Exploring large sets of forecasts with ML

▷ Objective : analyze and summarize the large quantity of NWP information to facilitate the human expertise

▷ Two approaches considered *(PhD A. Mounier)*

- Automatic detection of coherent structures in NWP forecasts targeting specific high-impact events such as damaging thunderstorms, tropical cyclones
  ▷ Occurrence of such events in the forecasts ?
  ▷ Where and when ?

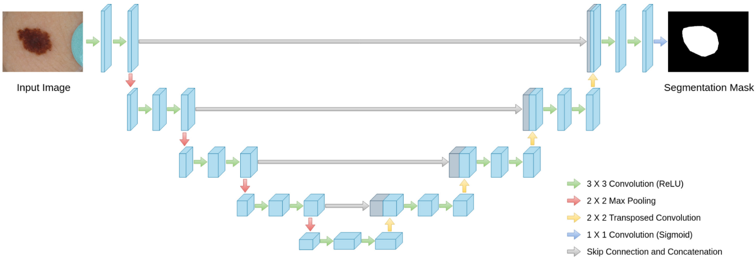
- Ensemble cluster analysis : summarize several tens of forecasts with a limited number of possible scenario that can easily be communicated to end-users.
  ▷ How similar are the forecasts ?
  ▷ What are the different scenario for the next hours and days ?

2 - Detection of coherent structures

▷ Convolutional Neural Networks (CNN) are very powerful for image segmentation

▷ Applications of CNN to the detection of weather objects are increasing (Racah et al. 2017, Lagerquist et al. 2019, Rasp et al. 2020)

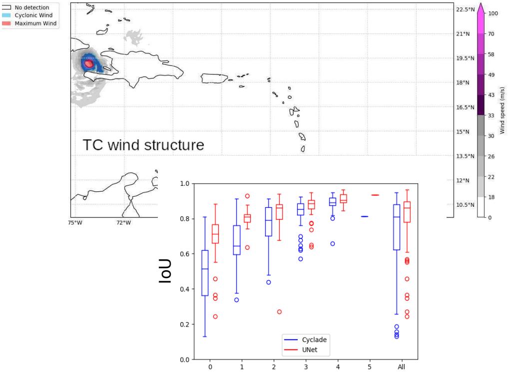
▷ The U-Net encoder-decoder (Ronneberger et. al 2015) is simple and efficient for this task

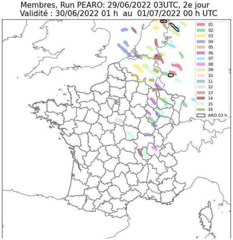


▷ Main difficulty : obtain the large labeled datasets required for training

2 - Detection of coherent structures

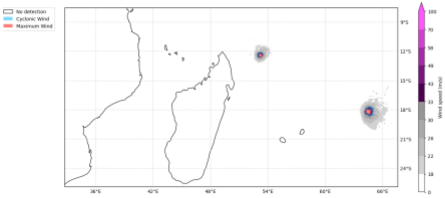
▷ Application to the detection of tropical cyclones and bow echoes (a specific type of thunderstorm with a shape of bow)
▷ Both are rare events : requires some pre-processing (data balancing) + weighted cross-entropy
▷ Datasets of several hundreds of forecasts have been entirely hand-labeled by experts

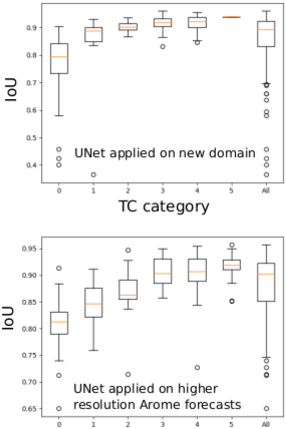




Bow echo detection *(PhD A. Mounier)*
https://journals.ametsoc.org/view/journals/aies/1/2/AIES-D-21-0010.1.xml

- Can we use the trained U-Net on slightly different data ?
- Tests domain generalization and resolution generalization





- Performance results are similar to those on the original data.

2 - What's next ?

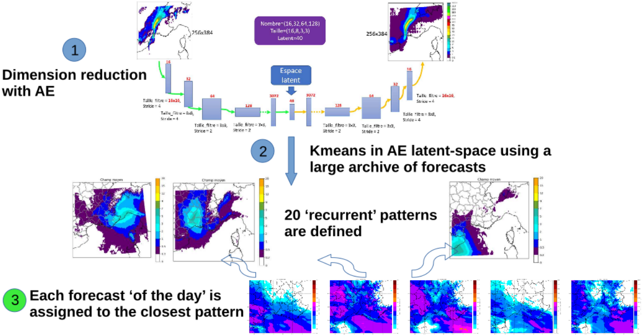
- Develop enhanced training datasets without extensive human labeling : active learning ?
- Take into account inhomogeneities in the labels (due to different experts for instance)
- Adapt the ML algorithms to slightly different sources of data (eg, new versions of NWP, observed data) : transfert learning / domain adaptation
- Explore other weather objects (e.g., supercells, weather fronts)

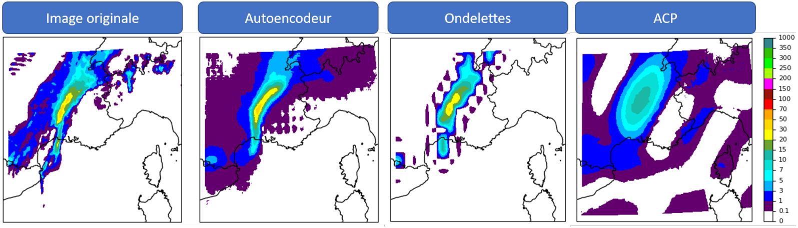
▷ Ensemble cluster analysis aims at regrouping similar forecasts in the same scenario

▷ How to measure the similarity between forecasts is a key and complex question, especially at high resolution

▷ PCA often used prior to clustering of meteorological fields

▷ We propose to use an Auto-Encoder and latent-space clustering





Courtesy : A. Mounier

▷ **What's next ?**

- Scenario based on multivariate (e.g. precipitation and wind gusts) and heterogeneous inputs (e.g. wind forecast and local wind energy production, PhD to start May 2023)
- Take into account the temporal consistency in the ML framework
- Joint learning of AE and clustering ?

Plan

1 Introduction

2 Exploring ensembles with ML

3 Enhancing ensemble forecasts with ML

3 - The POESY project (ANR, 2021-2024)

▷ Towards **hybrid prediction systems**, combining classical physical modelling with ML techniques, **to generate large samples of high-resolution forecasts**.

Challenge 1 : ML-based statistical downscaling

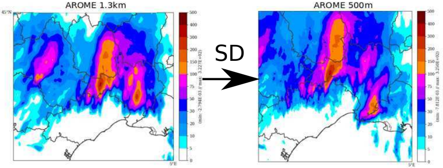
Improve the realism of predicted high-impact events with a learning-based statistical downscaling of NWP forecasts to sub-km resolutions

Challenge 2 : ML-based forecast generation

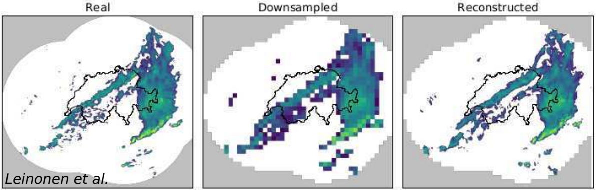
Improve the representation of forecast probability distributions by increasing the NWP sampling from O(10) to 0(1000) thanks to complementary ML-generated forecasts

▷ All works are done using AROME ensemble forecasts.

3 - ML-based downscaling

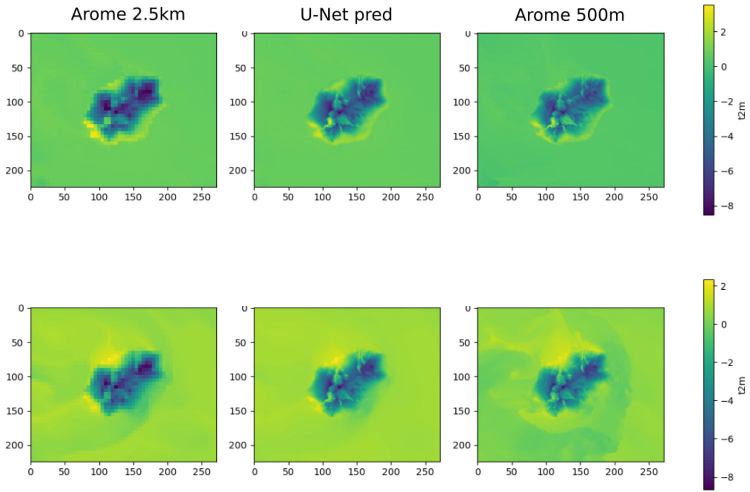


▷ Statistical downscaling (SD) learns a relationship between coarse-scale simulations and fine-scale simulations.
▷ Encouraging proof-of-concept studies applied CNN for downscaling weather/climate forecasts (Leinonen et al., 2020 ; Höhlein et al., 2020)



▷ **Learn SD from Arome 2.5/1.3km to Arome 500m** forecasts

- ▷ First promising results on temperature with a standard U-Net
- ▷ To be continued ...

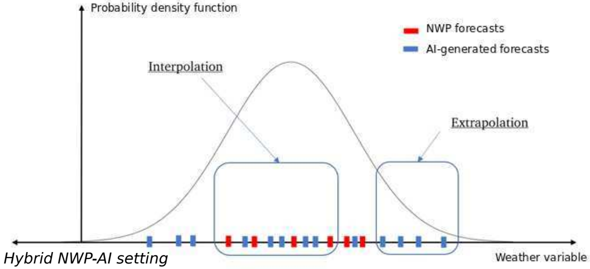


(Courtesy : L. Danjou)

3 - ML-based forecast generation

▹ NWP-based distributions are under-sampled with $\mathcal{O}(10)$ forecasts

▹ Are **deep generative models** such as **Generative Adversarial Networks (GANs)** able to produce "NWP-like" forecasts ?

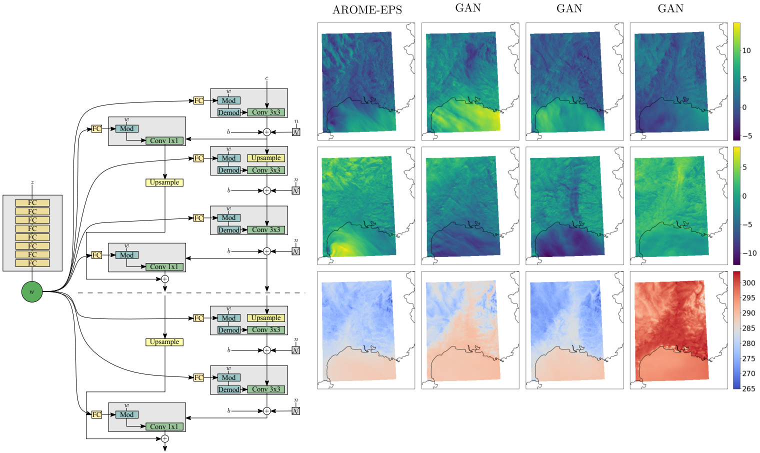


▹ Application of GANs to NWP and atmospheric state generation remains largely to be explored

▹ First promising results in works by Bathia et. al (2020), Besombes et al. (2021), Ravuri et al. (2021)

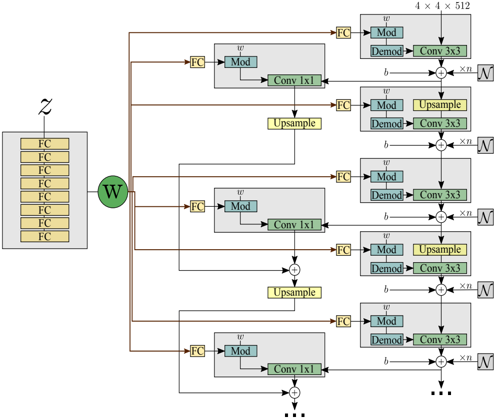
3 - ML-based forecast generation

▷ **Our first step : produce 'AROME-like' samples with unconditional GANs** (courtesy : C. Brochet, G. Moldovan)



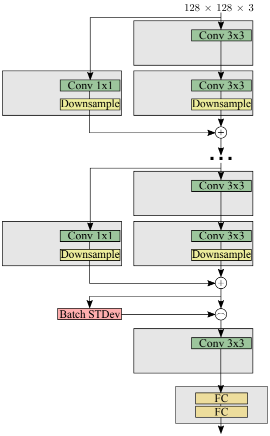
▷ Focus on 2m-temperature and 10-m wind fields

▷ The Generator



- proGAN, styleGAN, styleGAN2 (Karras et al. 2017, 2018, 2019)
- Introduction of intermediate latent space $w \in W$ to encourage disentanglement
- Control of output features at different scales via latent code (adjusting the style)
- Noise injection to encourage stochastic variation

▷ The Discriminator

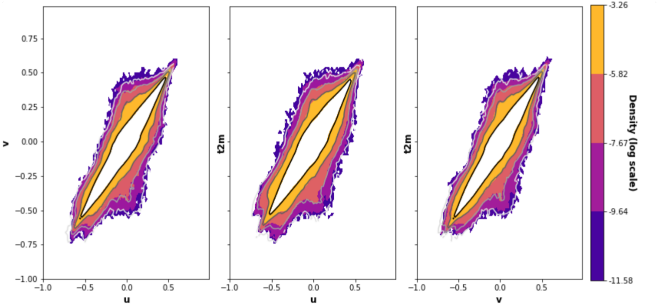


- Much simpler architecture
- Clearly inspired by residual networks
- Increased variation using minibatch standard deviation.
- Output is a classification score

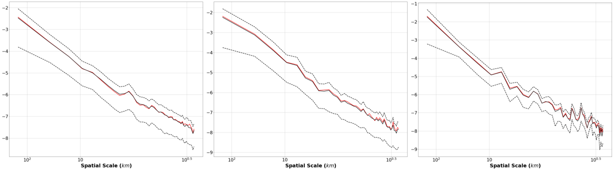
▷ **Evaluation of a StyleGAN trained over 17 months of Arome forecasts**

- Comparison of AROME and GAN distributions

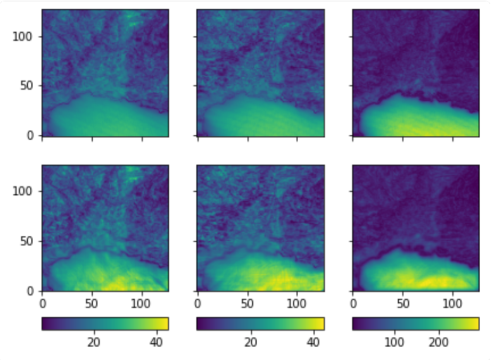
	SWD_{128}	SWD_{64}	SWD_{32}	SWD_{16}	W_1
ResGAN	5.7	7.3	12	39	12
StyleGAN	5.1	4.6	5.2	17.6	6.3



- Comparison of AROME and GAN spatial structures



u v t_{2m}



3 - ML-based forecast generation

▷ **Our second step : condition the generation on the current weather situation**

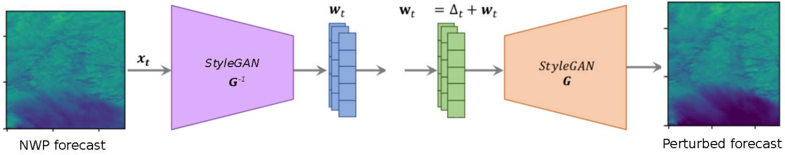
▷ Principle : apply perturbations in the latent-space of 'true' AROME forecasts to generate new ones

▷ Inverting an image onto the latent space of pre-trained generators is possible with StyleGAN

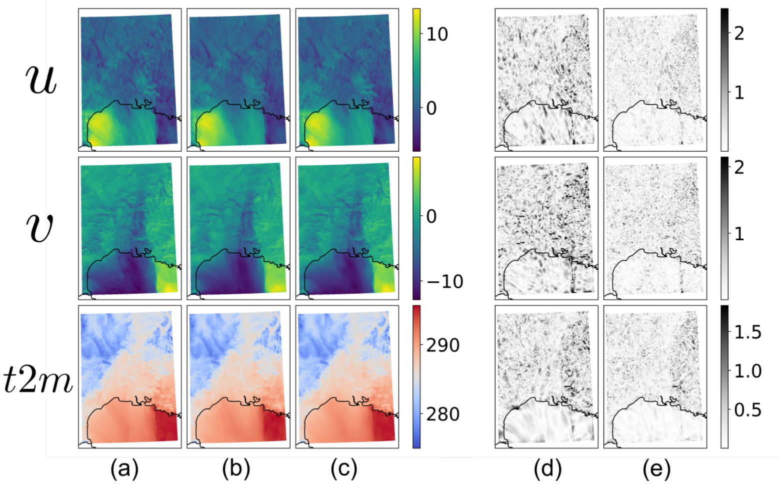
$$\mathcal{L}(\mathbf{w}) = \sum_{i=1}^{N} (x_i - \mathcal{G}(\mathbf{w})_i)^2$$

$$\mathbf{w}^* = \arg\min \mathcal{L}(\mathbf{w})$$

▷ How to design latent-space perturbations ?

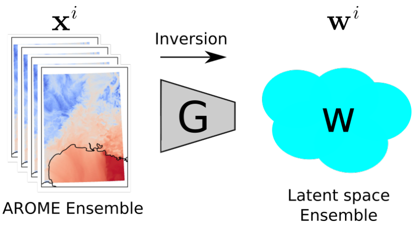


▷ **Inversion example**



(a) AROME forecast, (b) Inversion to W, (c) Inversion to $W+$,
(d)-(e) Reconstruction errors

▷ **Conditional generation of forecasts via perturbations in the latent space**



Generating forecasts conditioned to AROME ensembles :

- Inversion of a given ensemble forecast $\mathbf{x}^i$ to the latent space to obtain $\mathbf{w}^i$

$$\mathbf{w}^i = \psi\left(\mathbf{x}^i\right)$$

- Generating forecasts by applying perturbations to the latent space ensemble

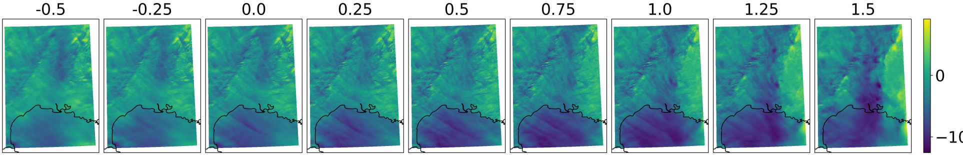
$$\tilde{\mathbf{w}}^i = \mathbf{w}^i + \delta\mathbf{w}^i$$

$$\mathbf{x}^{\mathrm{GAN},i'} = \mathcal{G}\left(\tilde{\mathbf{w}}^i\right)$$

▷ Interpolation/Extrapolation in the latent space

Generating forecast by performing interpolation/extrapolation operations between the members of the original ensemble in the latent space

$$\mathbf{w}'^{i} = \mathbf{w}^{i} + \alpha\left(\mathbf{w}^{i} - \mathbf{w}^{j}\right), \ i = 1, 2, ..., N_e, \ j = 1, 2, ..., N_e, \ i \neq j,$$

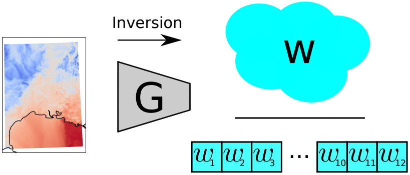


Examples obtained with different α

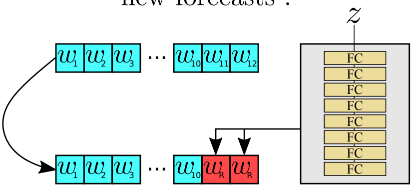
- *Pros* : builds on the Arome ensemble to design perturbations
$\Rightarrow$ secure and sound
- *Cons* : may struggle to generate sufficiently new samples.

▷ **Generating forecasts via style-mixing**

→ Deeper look into the inversion result : the latent code

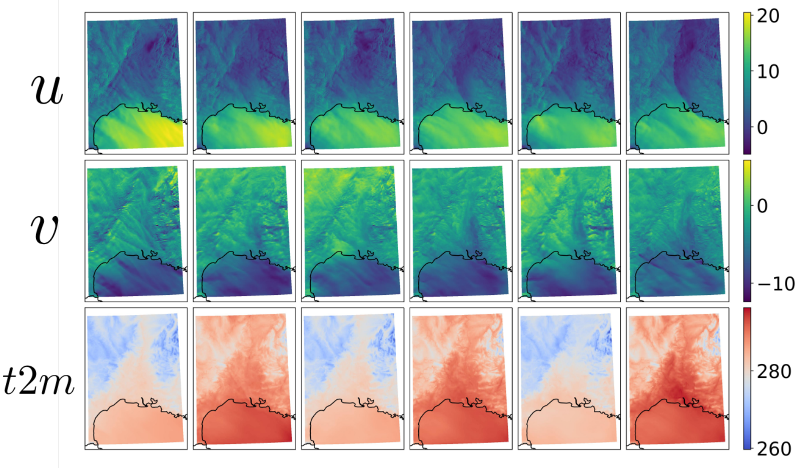


→ Mixing the style vectors of the forecast in the latent space with random latent codes to generate new forecasts :



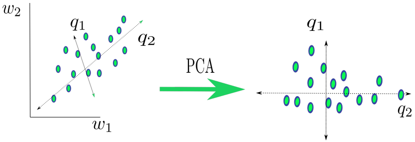
- *Pros* : may allow for generations that go beyond the Arome ensemble
- *Cons* : how to control what we are doing ?

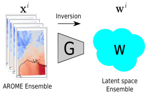
▷ **Example of generation of new forecasts via style-mixing**
(obtained by freezing the first 6 style vectors)



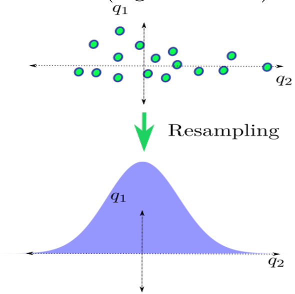
▷ Latent ensemble PCA

→ Using latent code distribution : finding ensemble principal directions



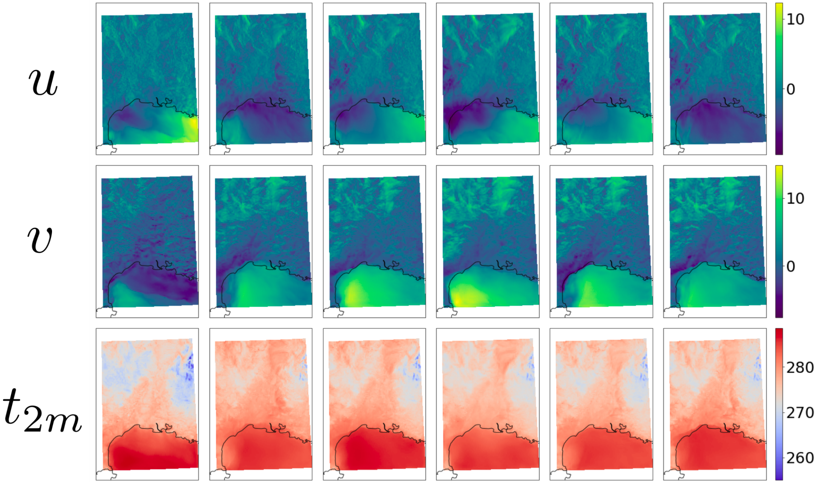


→ Sampling independent directions with the empirical variance (e.g Gaussian) :



- *Pros* : likely more control on the generated ensemble
- *Cons* : the 'searching space' for new features is reduced (low-rank covariance matrix).

▷ **Example of new forecasts via PCA-resampling**

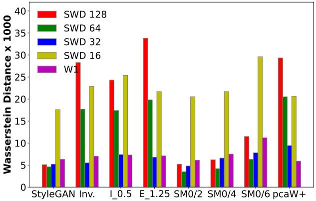


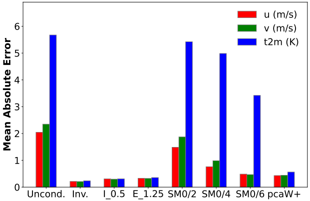
▷ **How to evaluate the conditioned generations ?**

- Are the generated forecasts realistic and physically sound ?
  - Wasserstein distance, SWD, power spectra ...
- Do the generated forecast enrich the ensemble ? Do they add new information ?
  - Entropy, ensemble dispersion
  - Forecast scores : Bias, RMSE, all standard probabilistic metrics (CRPS, reliability, resolution)
- Does an Arome-conditioned GAN ensemble fill the distribution as well as a large Arome ensemble would do ?
  - Compare with a very large Arome ensemble
  - Limited to very few cases due to computational cost

3 - ML-based forecast generation

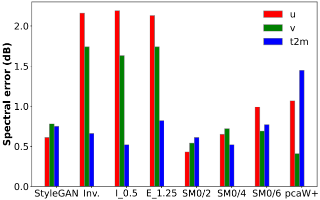
▷ **Early evaluation results** : ‘sanity’ check
74 days regularly sampled over $1_{1/2}$ year : a 16-member Arome ensemble is available each day for 8 lead times (each 3 hr up to 24h).

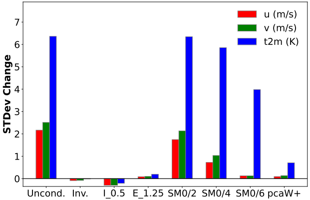




Wasserstein distance / spectral error evaluation of generated climates

Abs. Diff ens mean / STDev change in the generated daily ensembles





3 - What's next ?

▷ To be explored in the next few months :

- Continue the benchmarking of latent-space perturbation methods with a large set of objective metrics
- Comparison to a very large Arome ensemble ($\mathcal{O}(1000)$)
- Focus on extremes
- Generation of precipitation forecasts, a highly-intermittent, non-gaussian variable ...
- Large-scale forecast generation (going from a small domain to the full AROME domain)
- Generation of temporal sequences
- Investigate diffusion models ?

▷ ML is a very promising tool to explore forecasts and design new synthesis products targeted to the users needs

▷ Some applications, such as object detection, are sufficiently mature to consider an operational use in the coming months

▷ GAN as an ensemble emulator : we are only at the beginning of this challenging and exciting work !

▷ Preliminary results are quite encouraging but there are also a lot of unanswered questions and avenues to explore

▹ This work is just a glimpse of what could be done with ML to enhance weather forecasts

▹ What about ML for the learning of atmospheric dynamics ?

- ML for emulating model physics is being explored at CNRM (B. Balogh et al.)
- Several papers recently introduced purely ML-based forecasts (without any physical model) with performances as good as state-of-the-art NWP to some extent (see FourCastNet, Pangu Weather, GraphCast), is it the future of weather prediction ?
  See WG on Learning-based Emulators for OAC Processes
- We would like to explore such models for regional high-resolution forecasting, although it may be strongly challenging (in terms of data, ML complexity and GPU infrastructures)

▹ The optimal way to combine ML and NWP is largely unknown yet, but several avenues may be worth exploring.