Towards a data-driven emulator of AROME

PREDICTABILITY Team
With the help of many others

DESR/CNRM/GMAP

February 2026

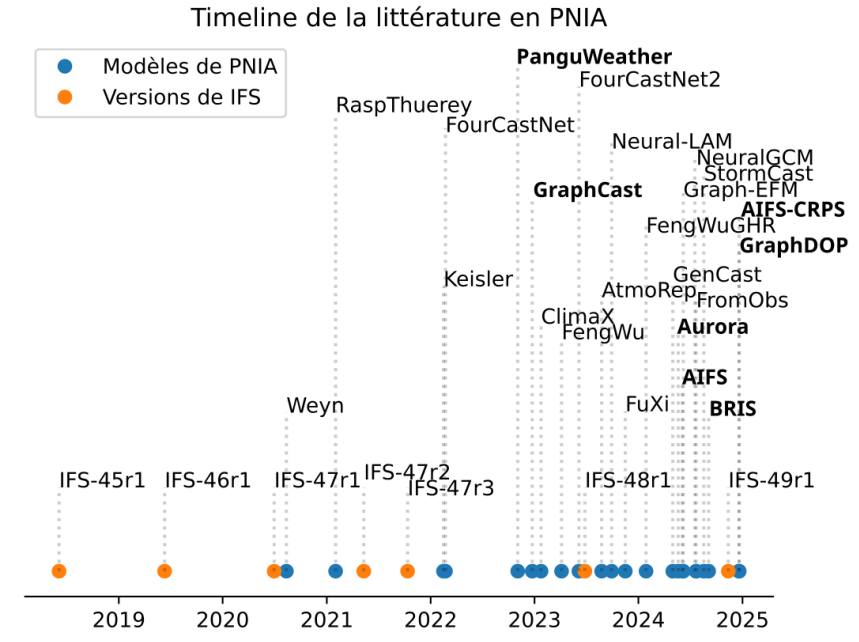
RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

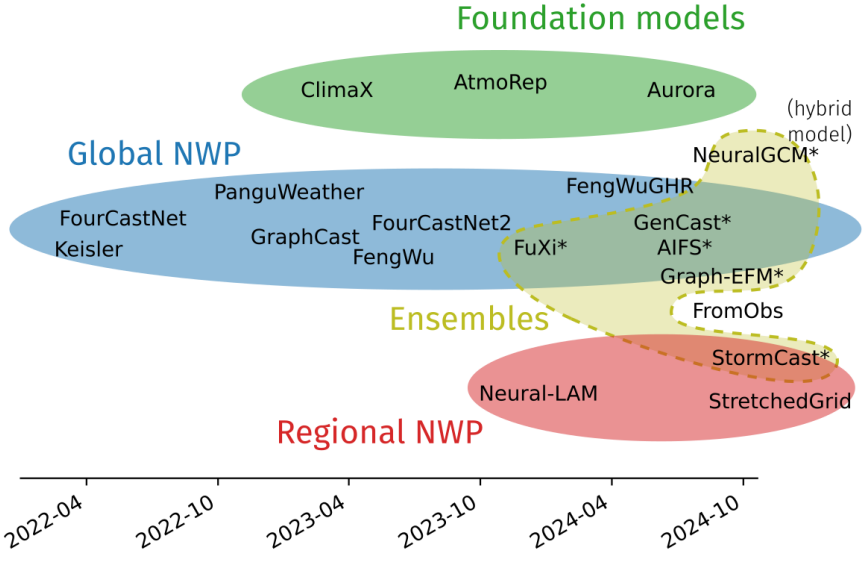
METEO FRANCE
À VOS CÔTÉS, DANS UN CLIMAT QUI CHANGE

Introduction

Machine Learning Weather Prediction (MLWP)

- Data-driven methods ("MLWP") have risen to **state-of-the-art**.
- From global forecasting to **pervasiveness**
- **Fast-paced** advances





What is our purpose ?

Can one do kilometer-scale MLWP with the skill of AROME ?

- ▷ Many National Weather Services have a similar goal.
- ▷ **Probably too soon to tell**, but already some results.
- ▷ Notion of **"skill"** is not univocal

Where do we stand today at Météo-France ?

- **What is the MLWP team even doing ?** 🤷
  - Data on a graph
  - Time leaps and auto-regression
- **What is the model worth ?** 🤔
  - Scores, errors and what they tell
  - Two interesting cases

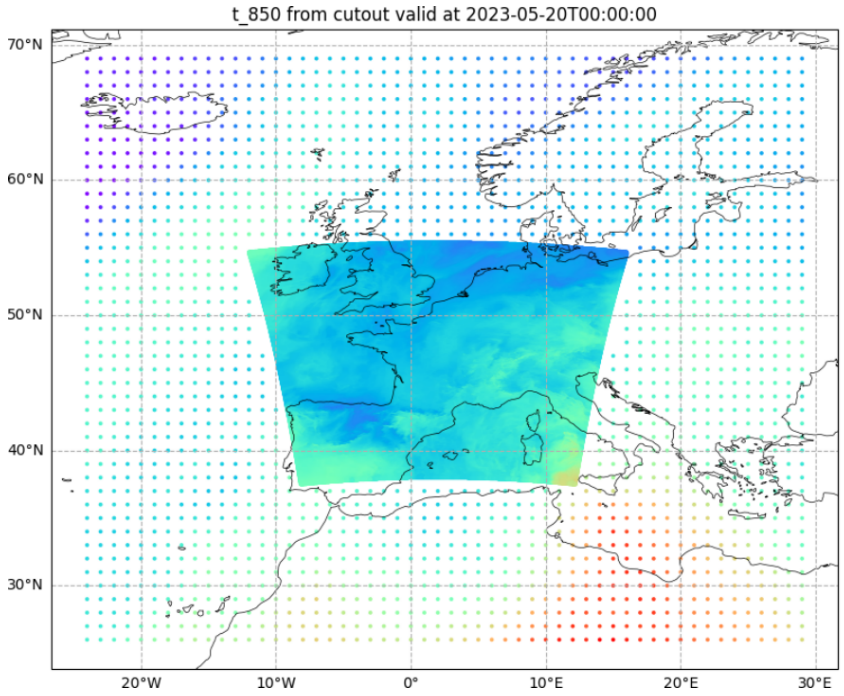
A primer on MLWP

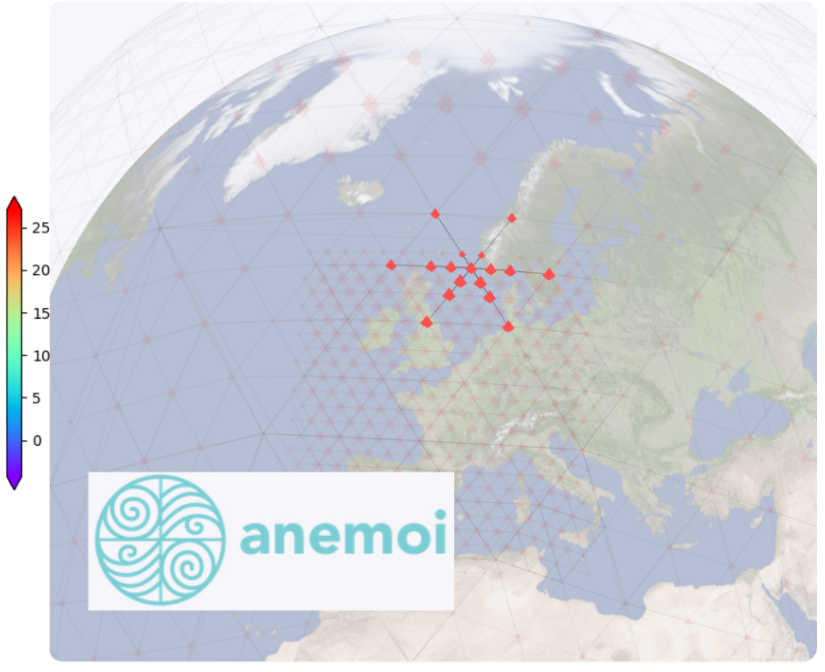
A primer on MLWP : a simple recipe

High-quality data (dubbed X_t):

- ERA5 ($\sim 30km$)
- AROME Analysis ($\sim 2.5km$)

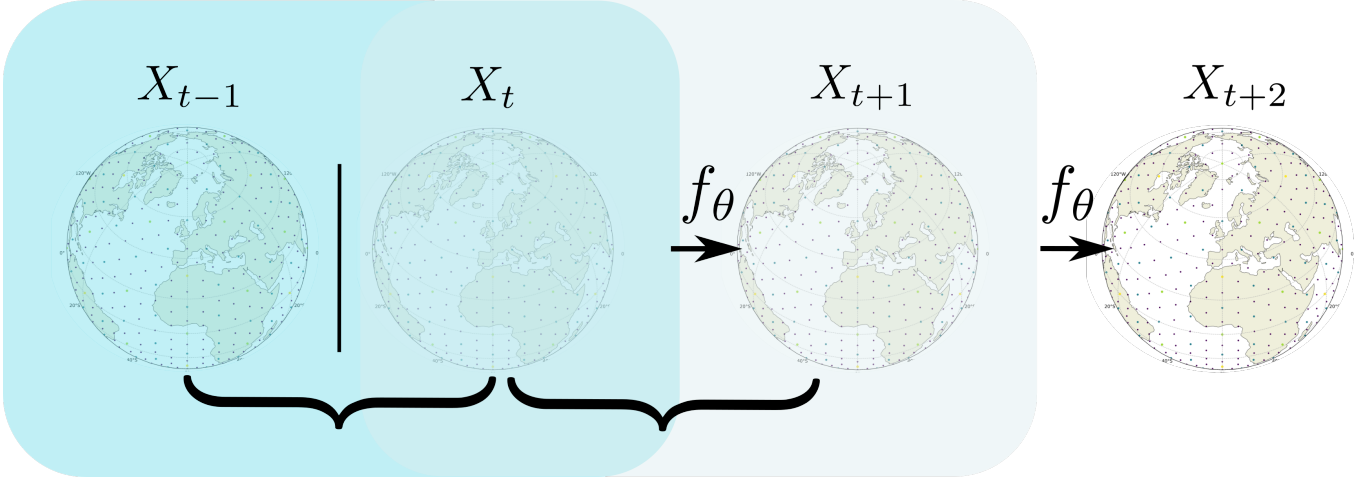
A computing grid : the "**graph**"





A primer on MLWP : a simple recipe

- A (neural-network) model : f_θ ; **independent from the graph**[1][2]
- A training algorithm : find the best model f_θ that predicts any data X.
- "Best" $\equiv$ minimize mean square error



Auto-regression

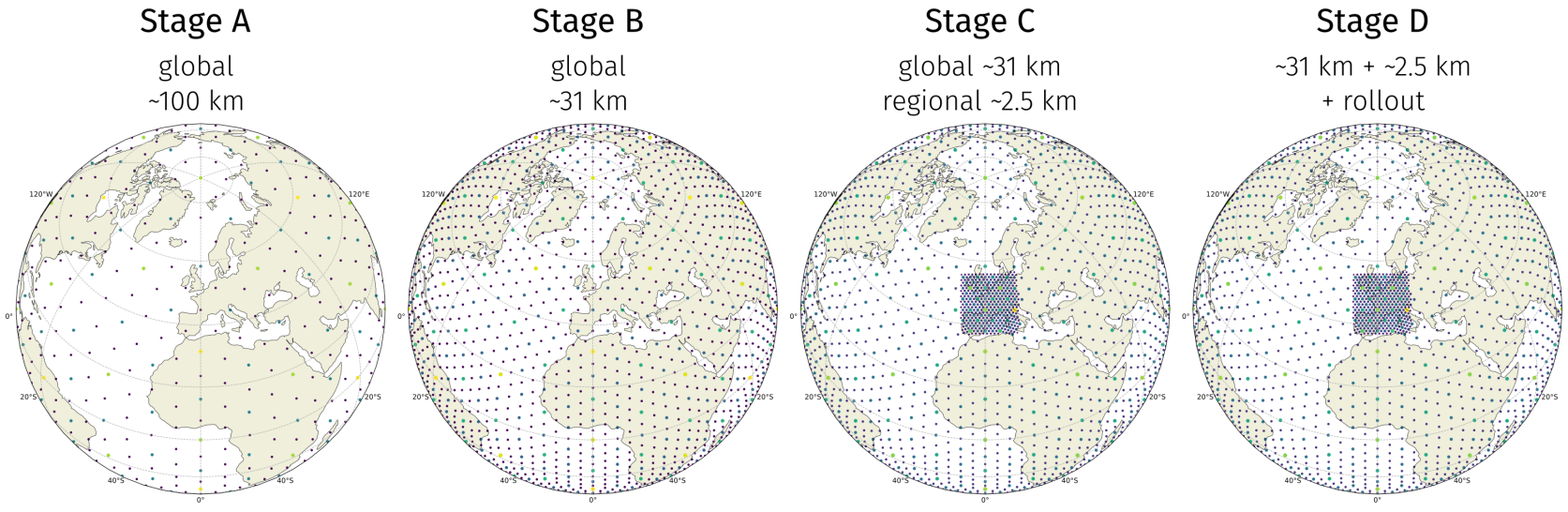
Using one's own output as next input

[1] θ : a set of **very many** parameters. 245M in our case. GPT-4 : 1.75T (?)

[2] The **graph** defines **where information flows**, f_θ computes **what flows**

A primer on MLWP : the training process 🔧

Training on several stages an ”AIFS-like” model on ”stretched-grid”, d’après Nipen et al. (2025)



- **Stage A :** global low-res *pre-training*
- **Stage B :** global high-res *training*
- **Stage C :** regional + global *fine-tuning*
- **Stage D :** increasing *rollout*

MLWP: Anemoi 💻

Anemoi is a framework mainly developed at ECMWF

▷ a handful of packages for datasets, training, inference

▷ many contributions from all over Europe (mainly)

- **What Anemoi is:** a powerful framework to MLWP, end-to-end ✅
- **What Anemoi is *not*:** your general AI toolbox to do everything ❌
- **At Météo-France:** Assimilation, Nowcasting, Chemistry, Climate, etc... ↗️

AROME-IA, a brief description

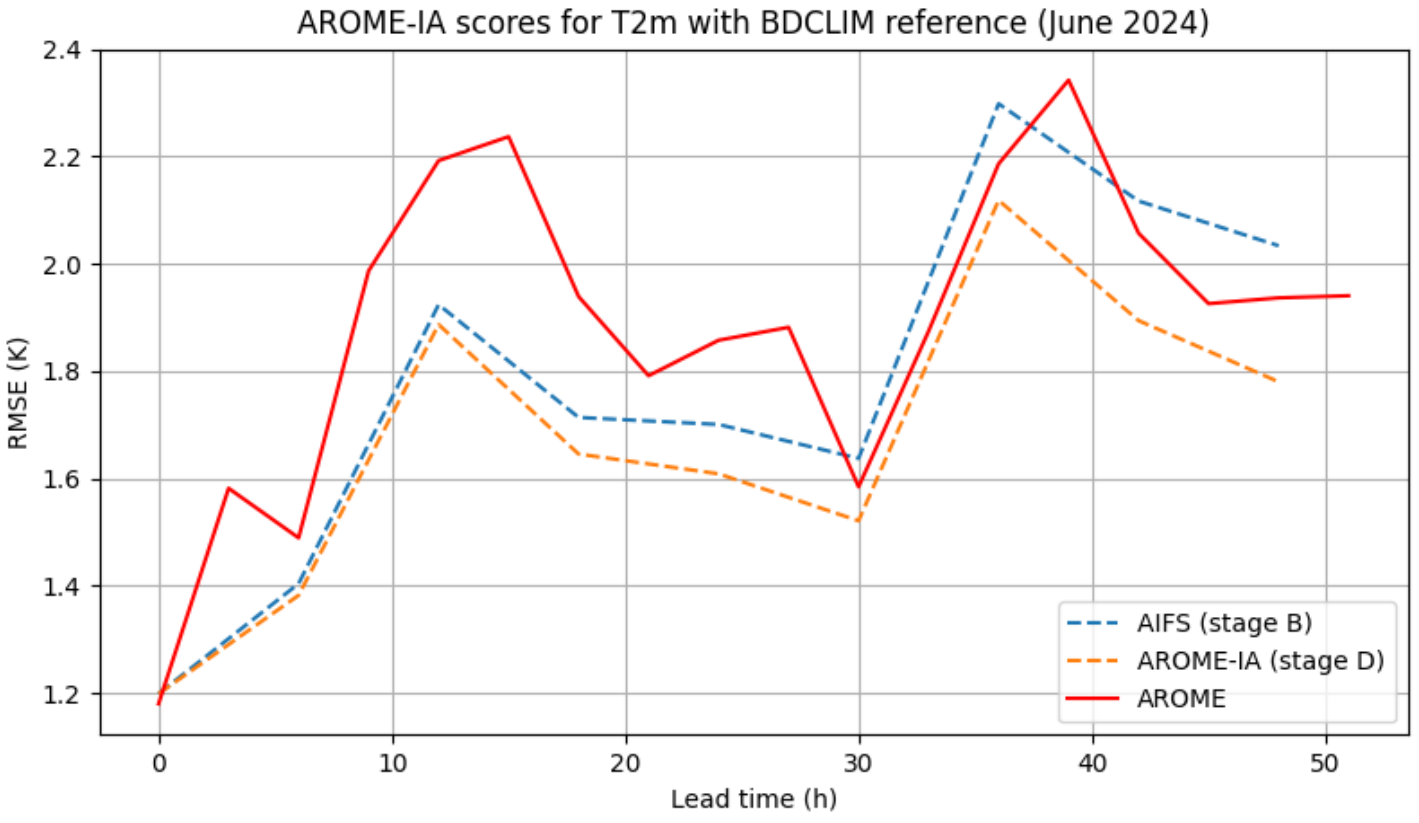
- **Purpose :** deterministic regional forecasting
- **Resolution :** 2.5 km (regional) / 0.25° (global), 13 levels, $\Delta t = 6$ h
- **Variables :** $d = 72$
  - Surface (7 prognostic) : $t_2, u_{10}, v_{10}, p_{mer}, p_{surf}, t_{surf}, td_2$
  - Surface (1 diagnostic) : total precipitation
  - Altitude (5) : t, u, v, z, q
- **Trained with :**
  1. ERA5, 1979–2022 (43 y.)
  2. AROME: BDAP, operational analysis, 2020–2024 (4.5 y.)

Main caveats

- Limited set of variables
- Limited timestep (6h only)
- Not gone very far in the tuning yet

What is the model worth ?

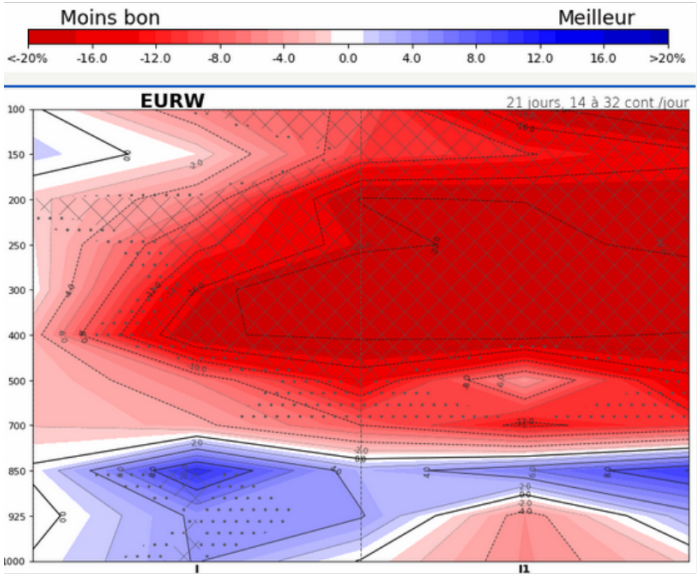
Scores, errors and what they tell



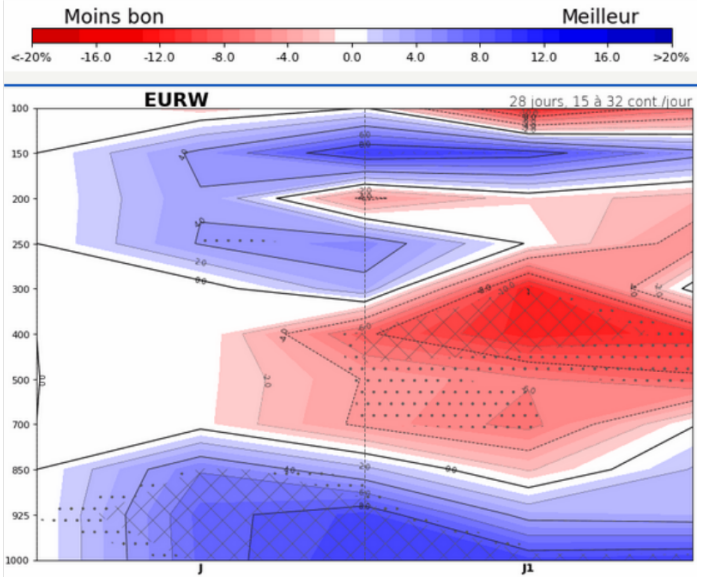
- **Score :** RMSE on T2m // BDCLIM. **Period:** june 2024
- Training on km-scale data is actually useful ! 🥳

Scores along the vertical 🌧️

October 2024



June 2024



- Relative RMSE gain between AROME and AROME-IA (stage D)
- Temperature scores against Radiosoundings
  - Coulours : AROME-IA **better**/**worse** than AROME.
- **Results :** AROME-IA

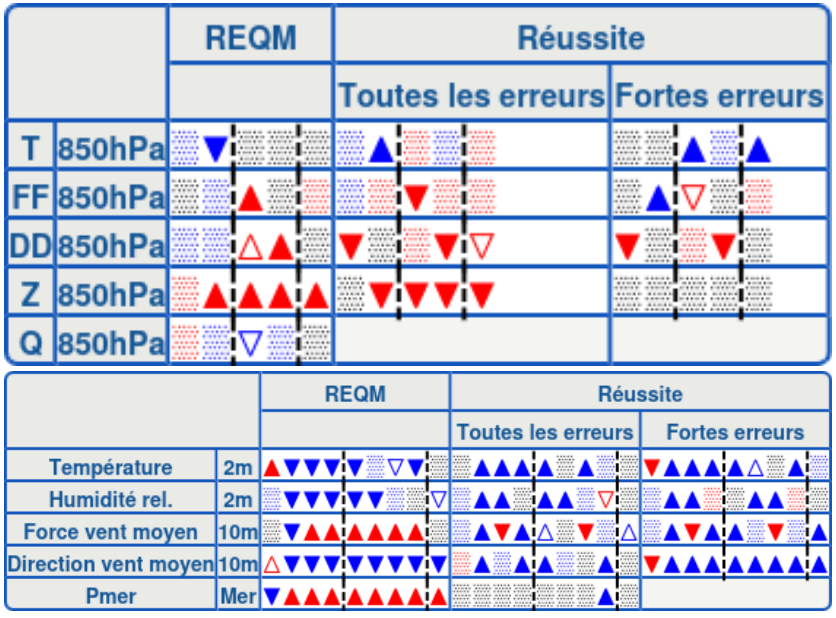
  🔵 near surface,

  🔴 in altitude,

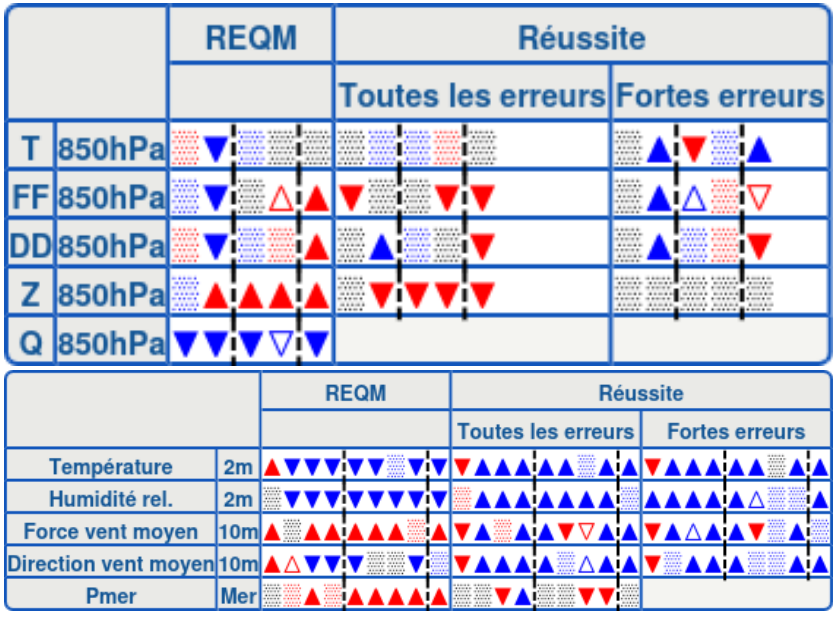
  ▷ **season-dependent**

Score cards

Sur octobre 2024



Sur juin 2024

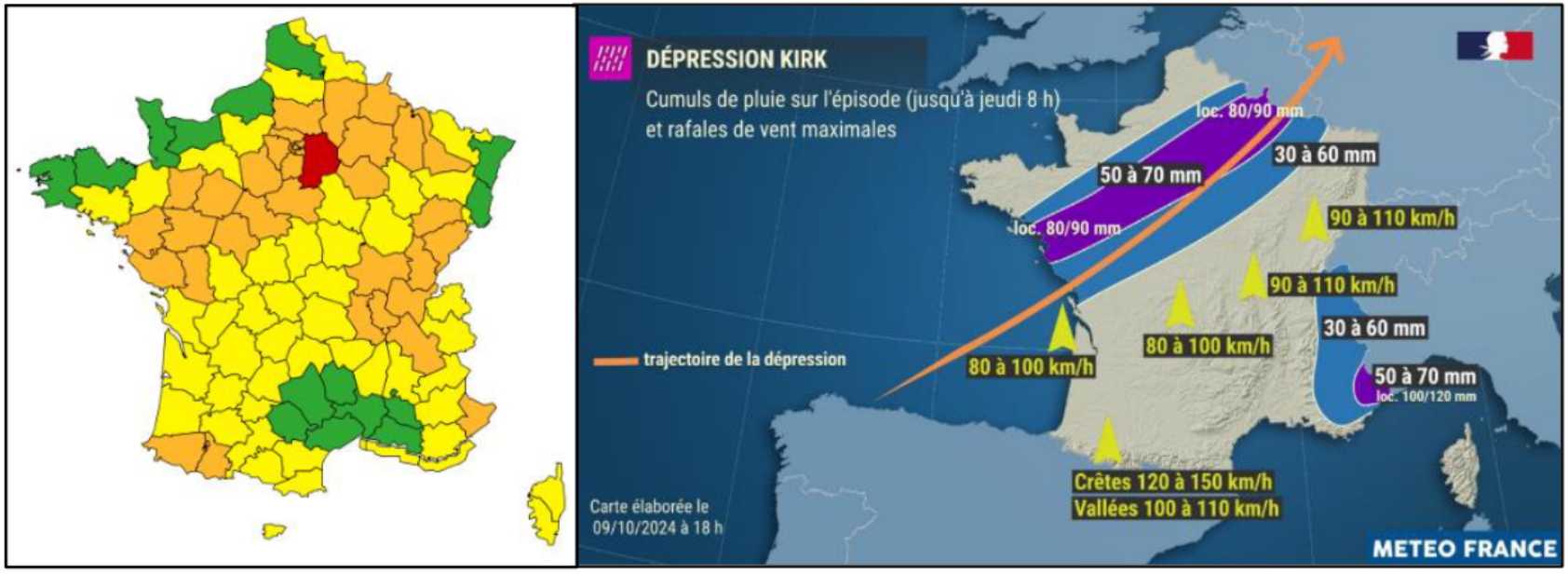


- Coulours : AROME-IA better/worse than AROME.
- AROME-IA is **better near surface** than in mid-troposphere
- Performances **highly** variable-dependent
  - Better with temperature, humidity, wind direction
  - Worse with pressure, geopotential, wind speed

Storm Kirk

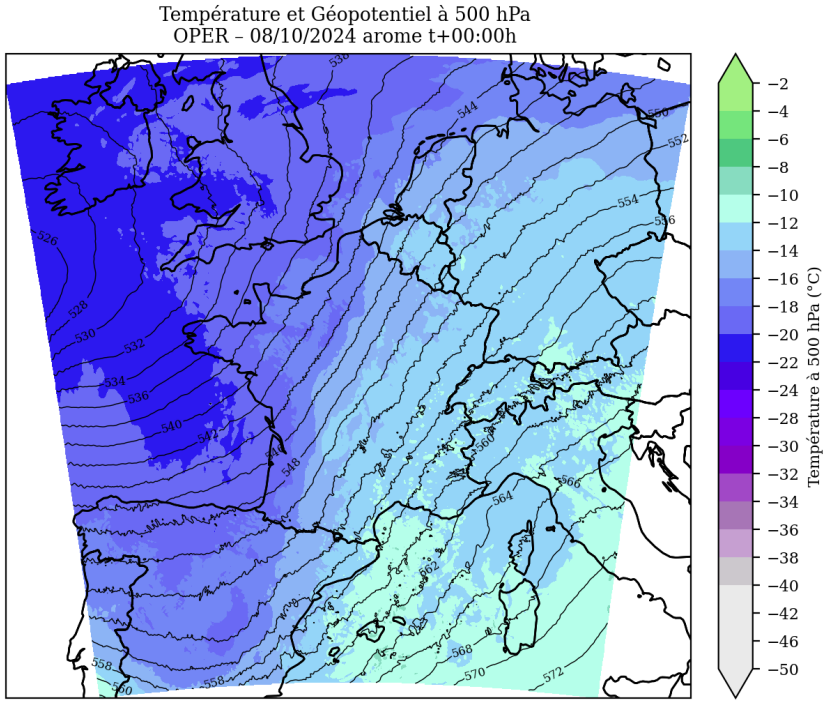
Kirk : Vigilance

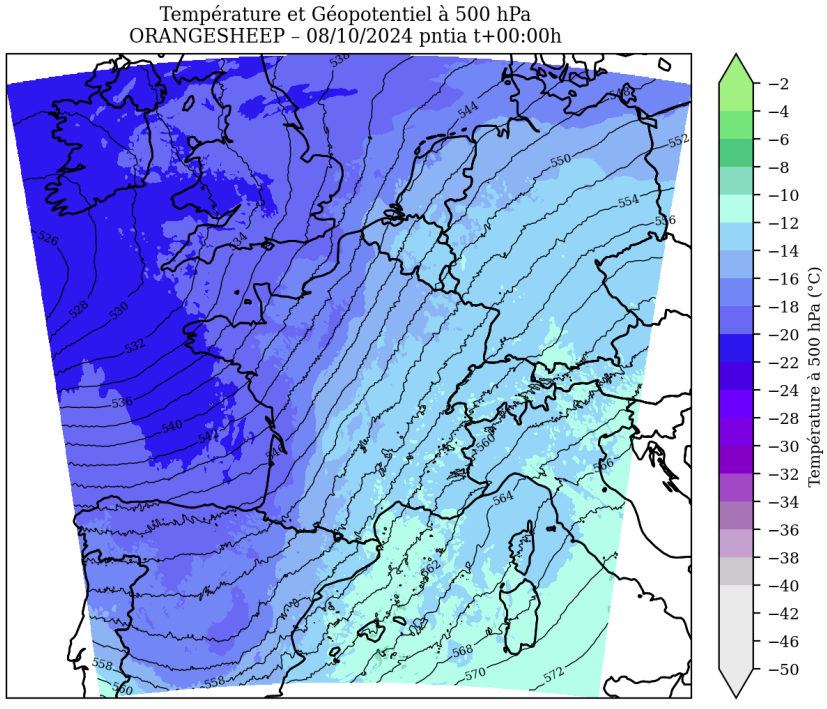
The **Kirk** depression, ex-hurricane, triggered sustained rainfall on the North-West between Wed., Oct. 9th and Thu., Oct. 10th, 2024 ; along with strong winds in the Pyrénées and Center-East.



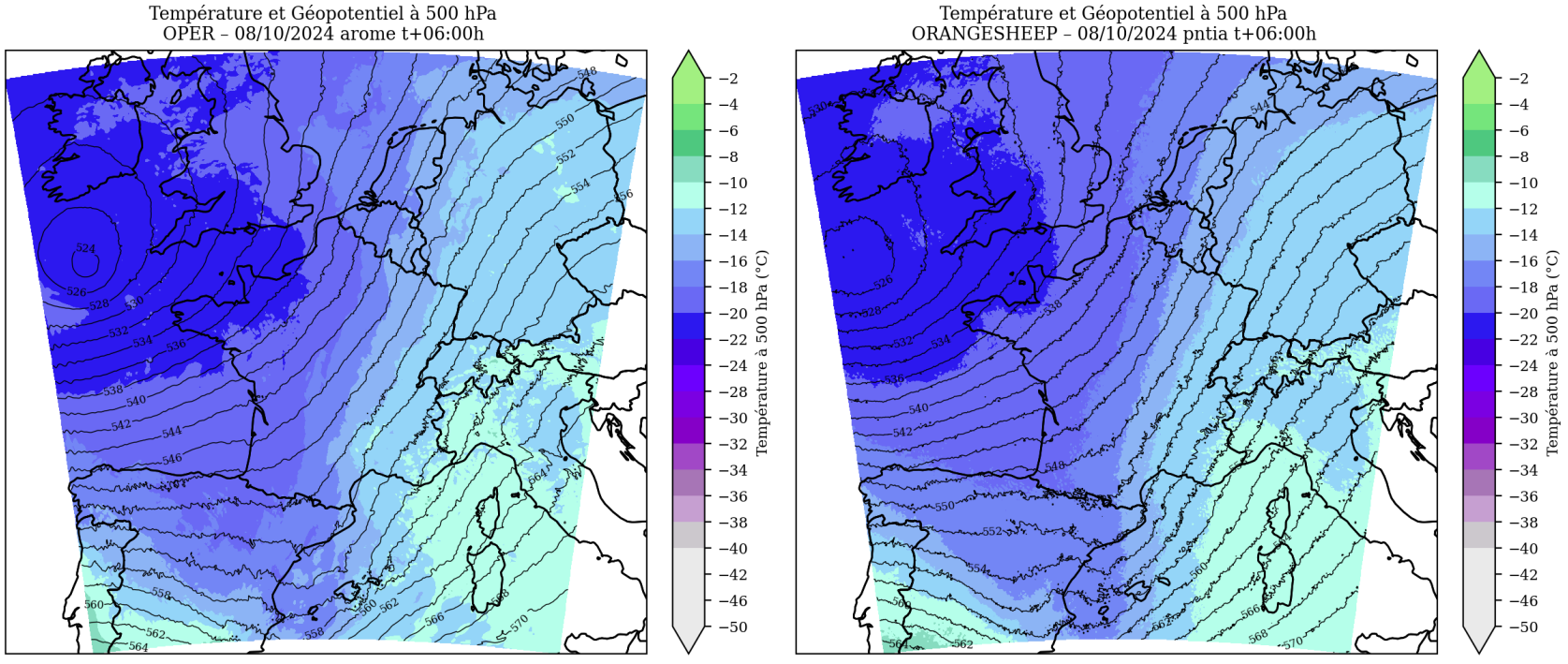
- **1 red warning** for river flooding at Grand-Morin river (Seine-et-Marne)
- **35 departments** in orange warning for wind and rain-flood.

Kirk : mid-troposhere

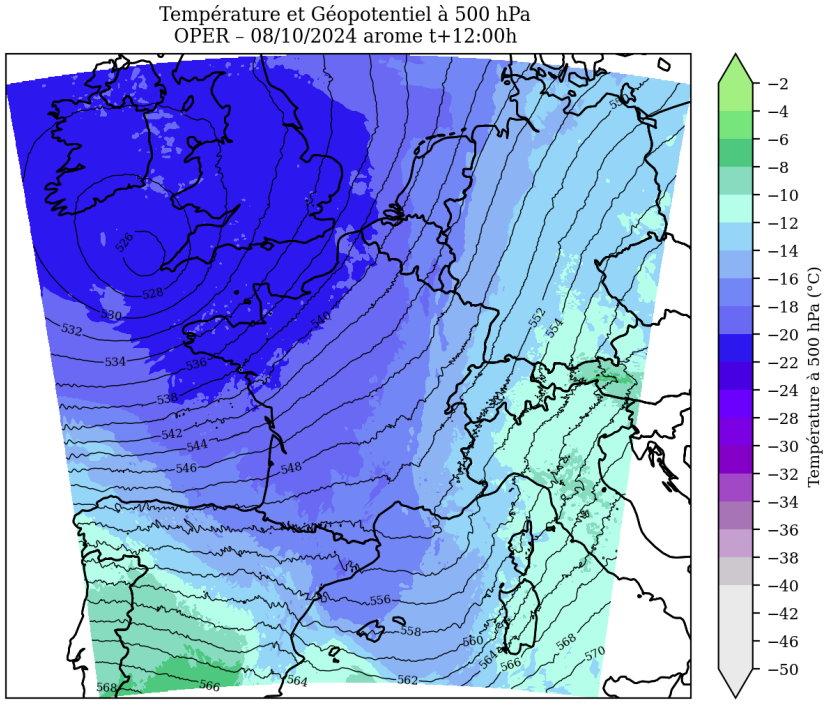


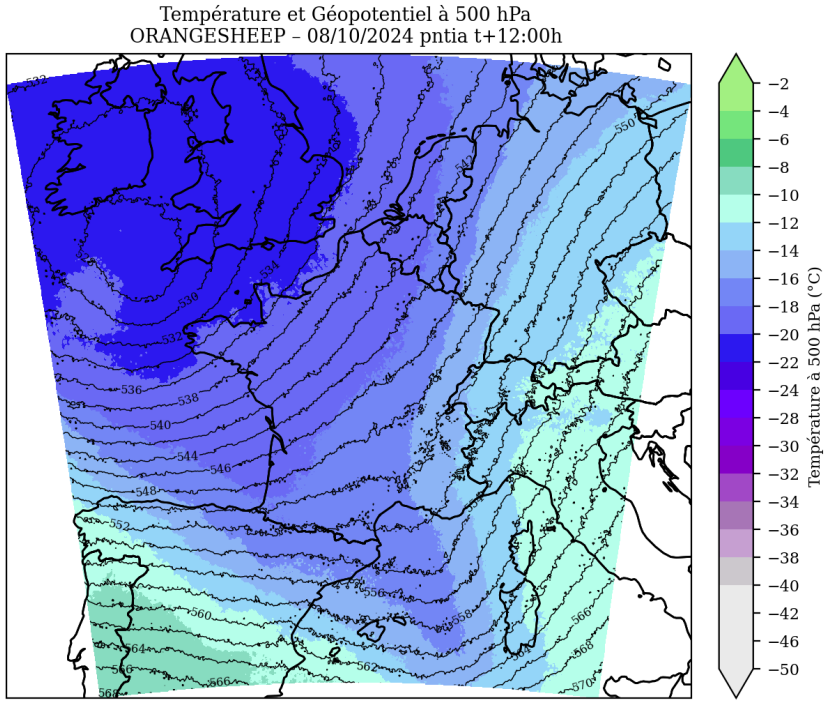


Kirk : mid-troposhere

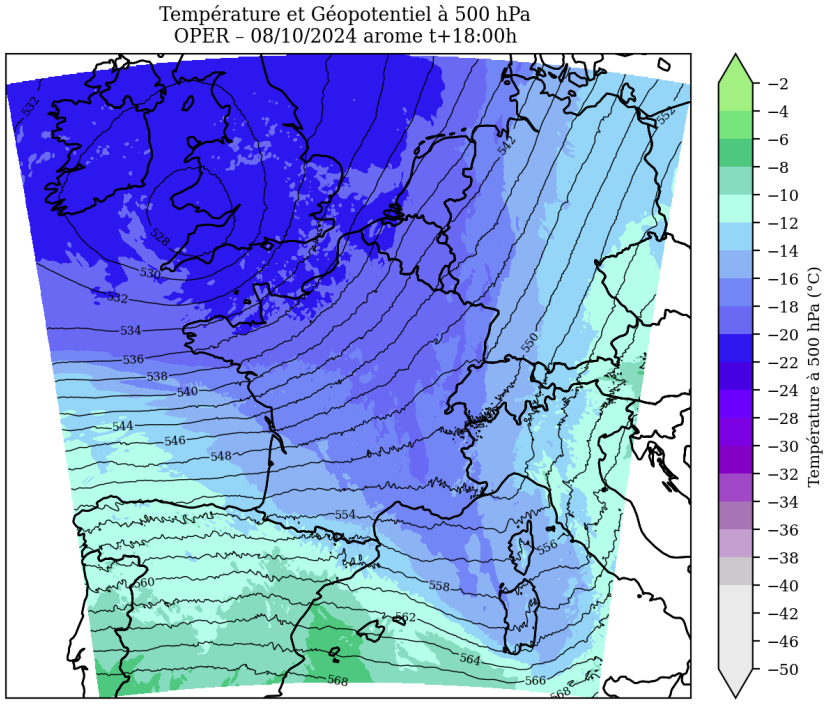


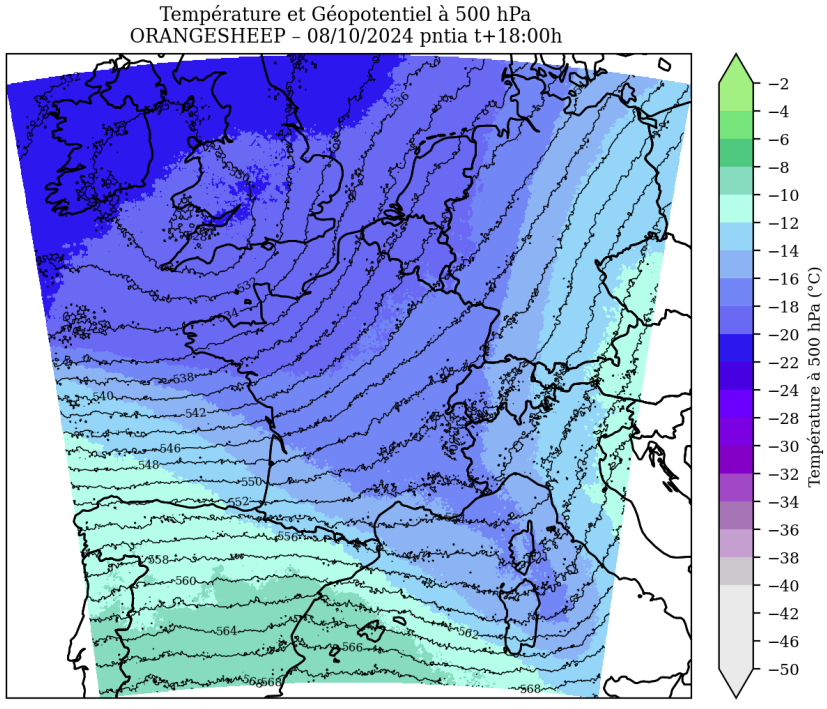
Kirk : mid-troposhere



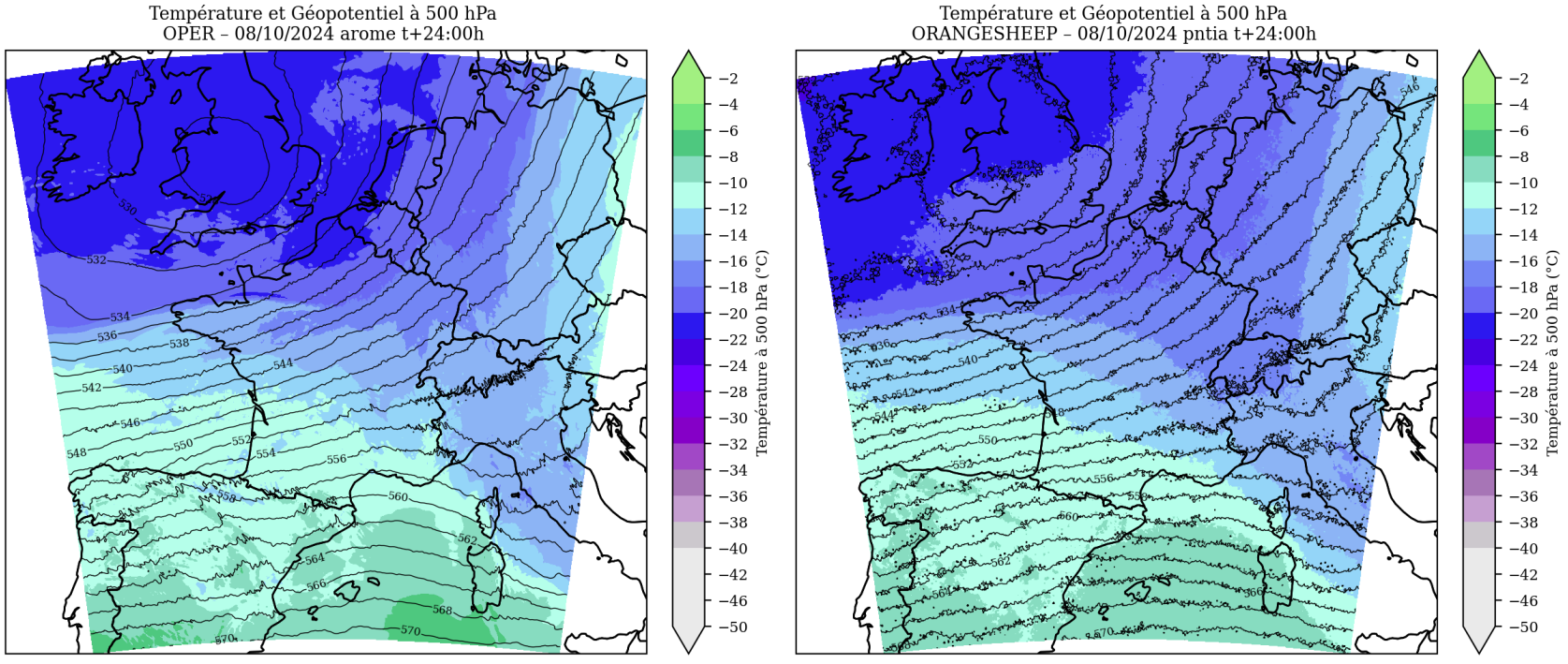


Kirk : mid-troposhere

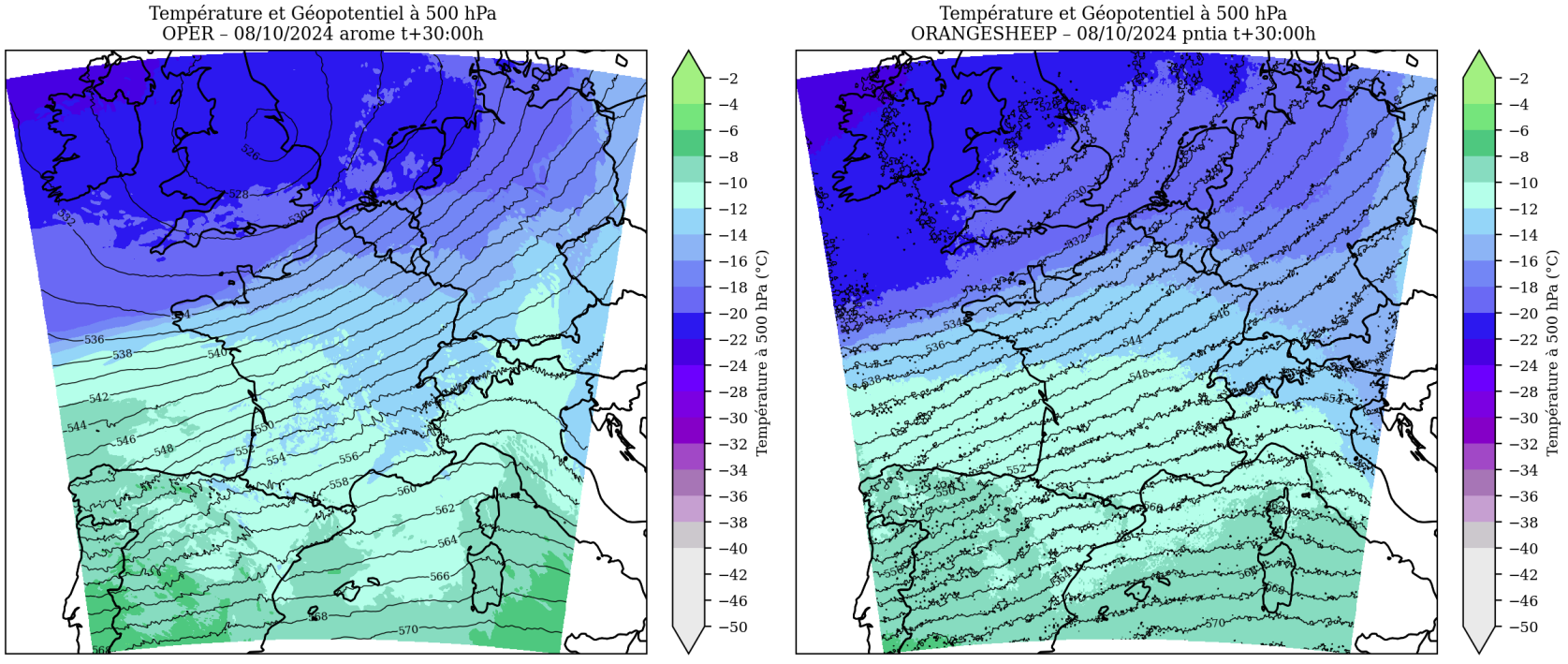




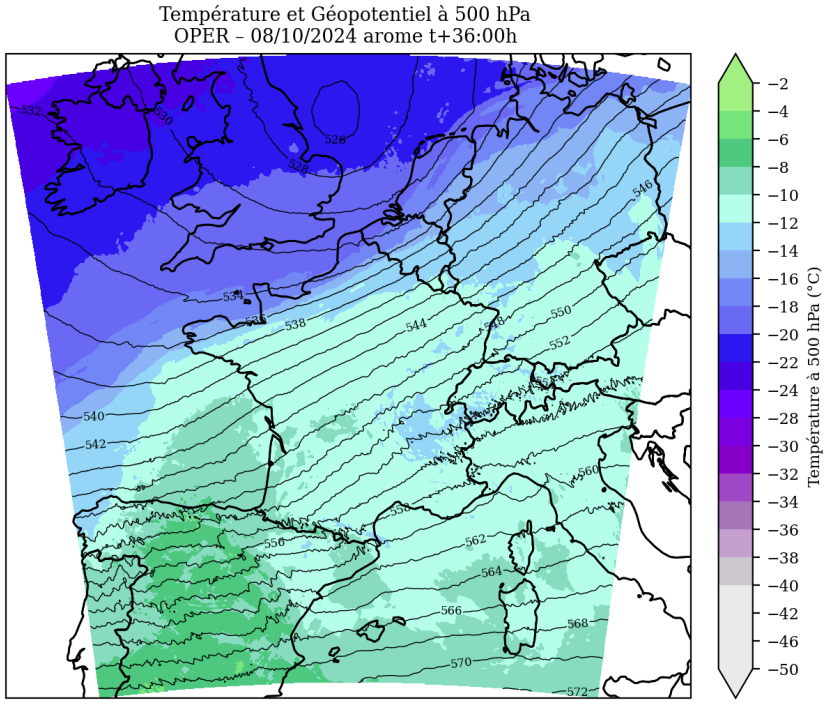
Kirk : mid-troposhere

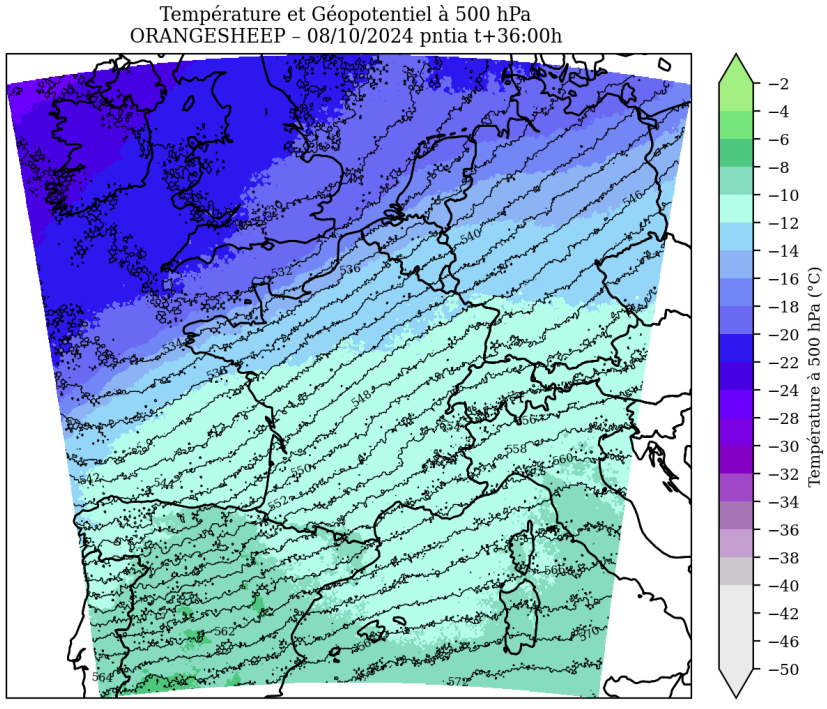


Kirk : mid-troposhere

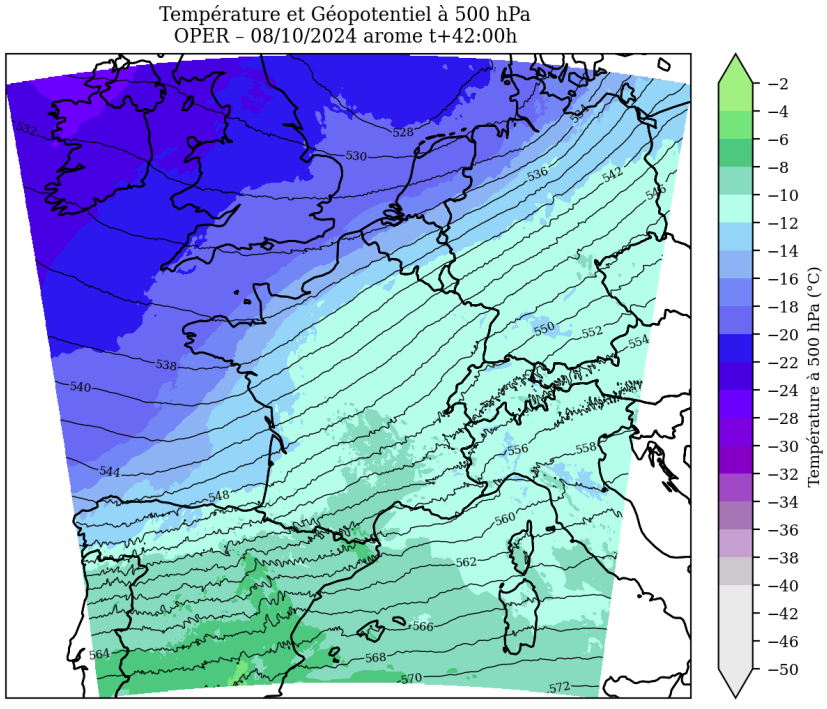


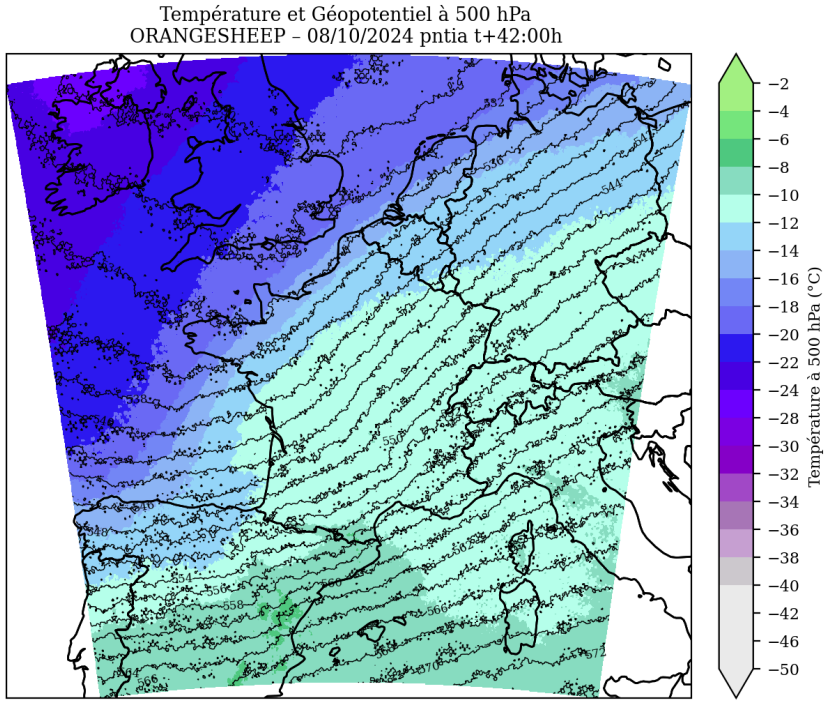
Kirk : mid-troposhere



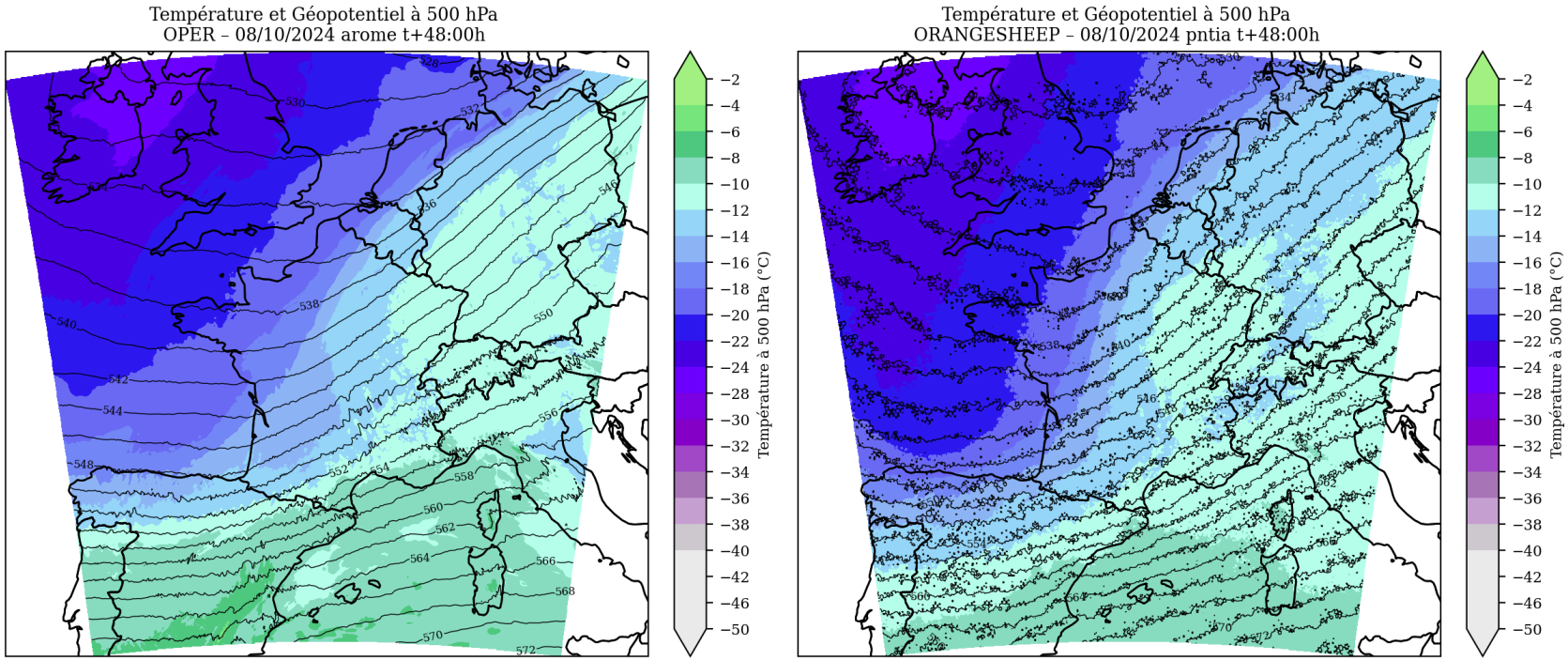


Kirk : mid-troposhere

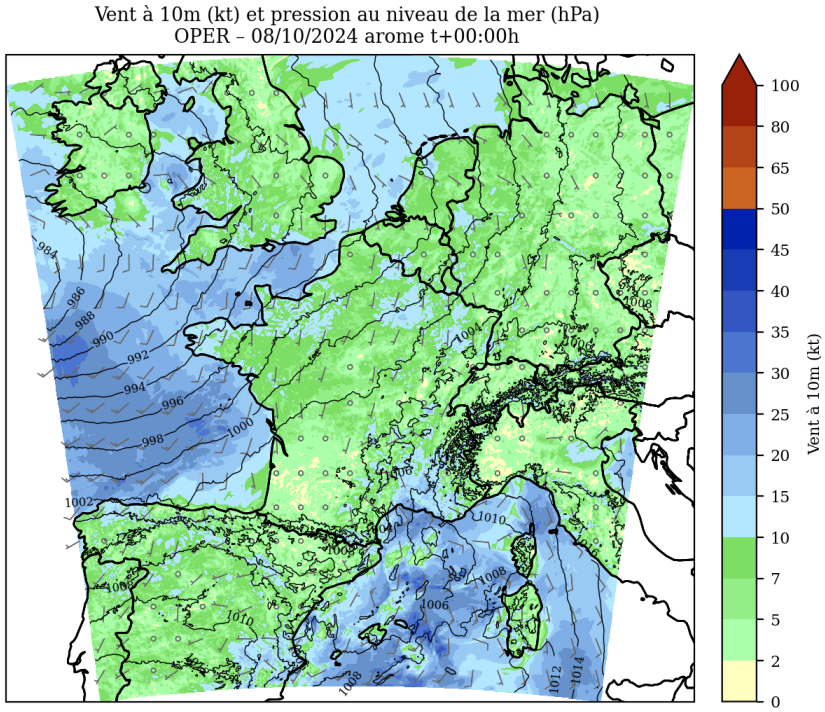


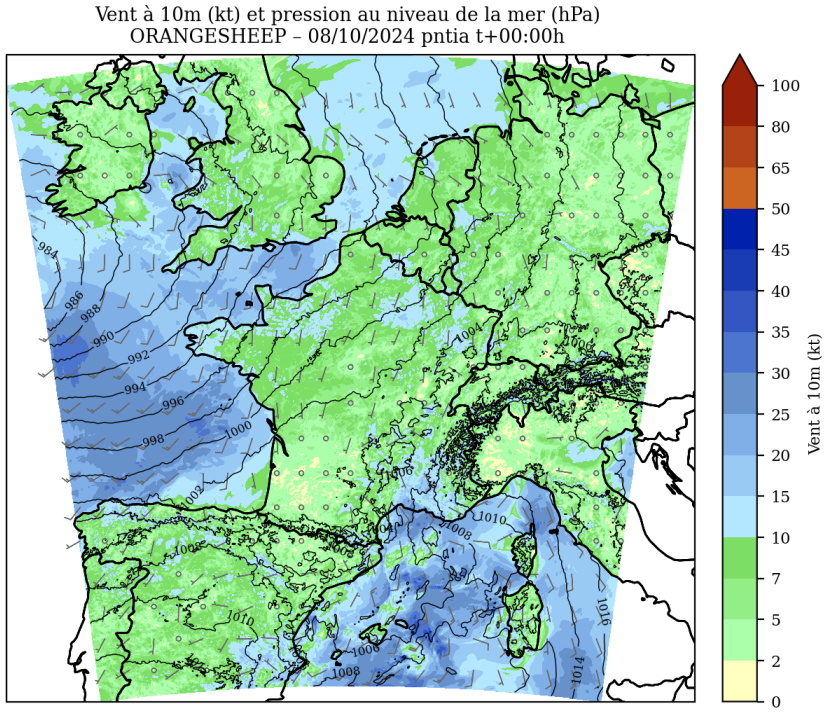


Kirk : mid-troposhere

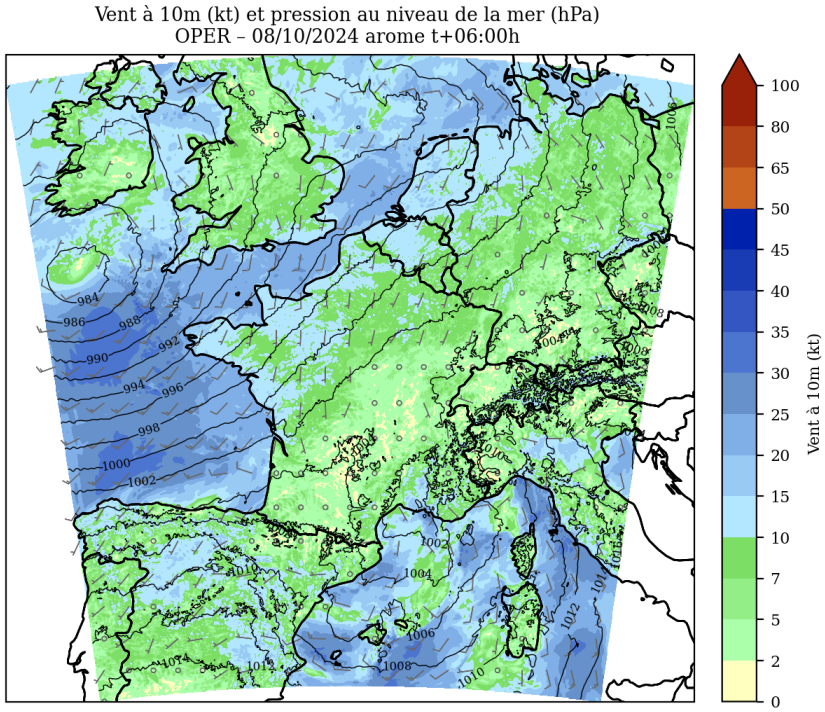


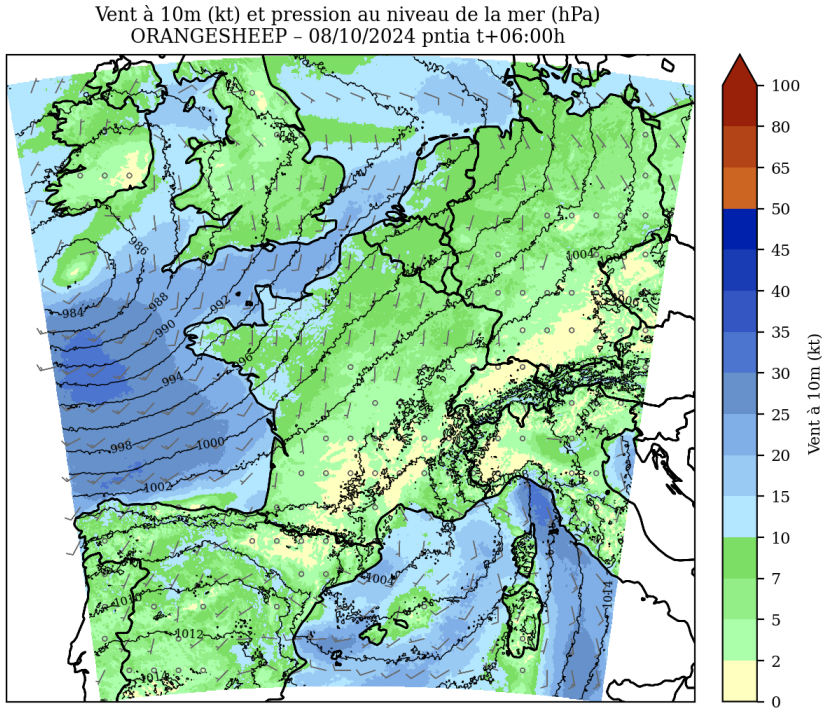
Kirk : surface wind



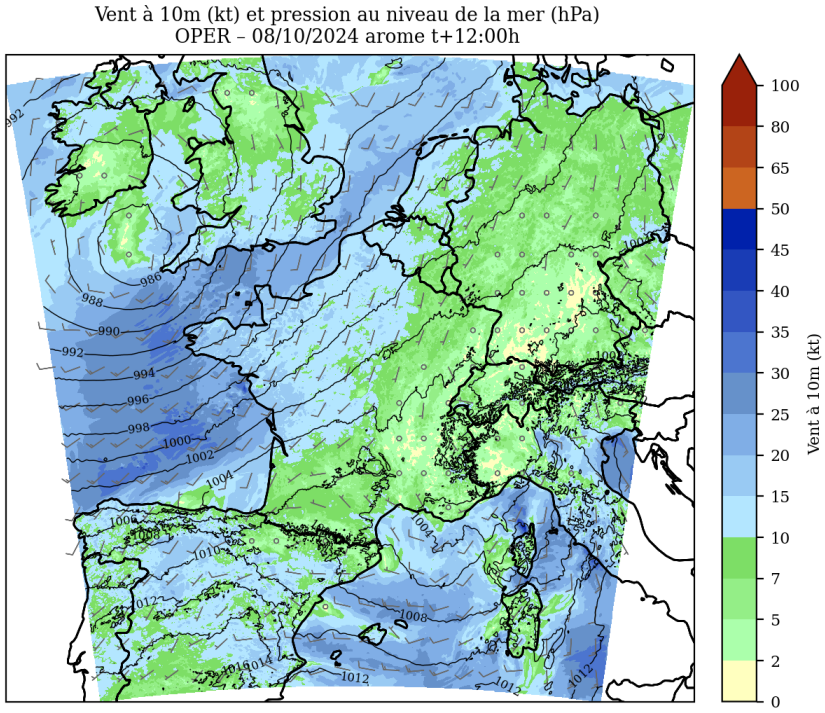


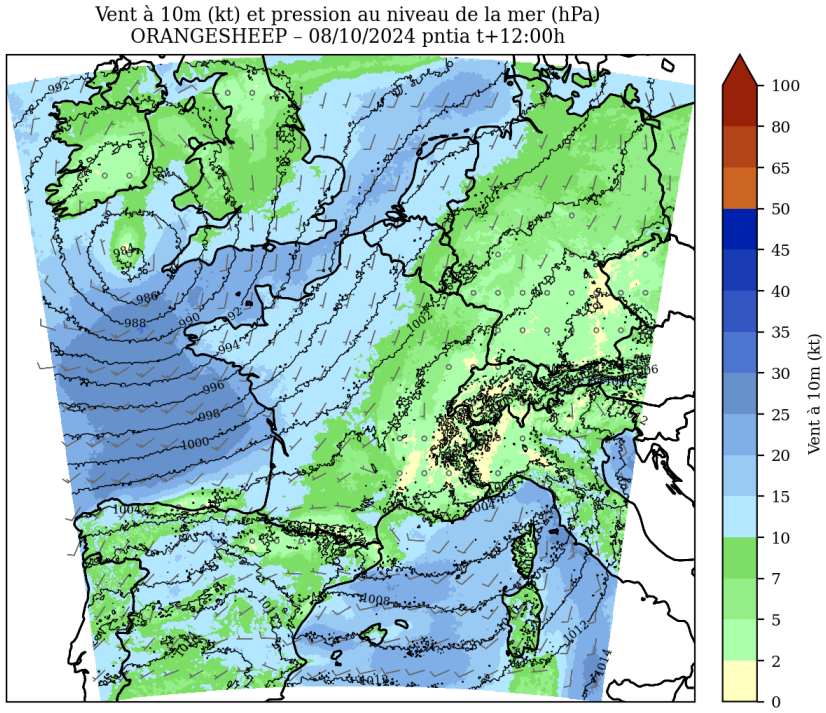
Kirk : surface wind



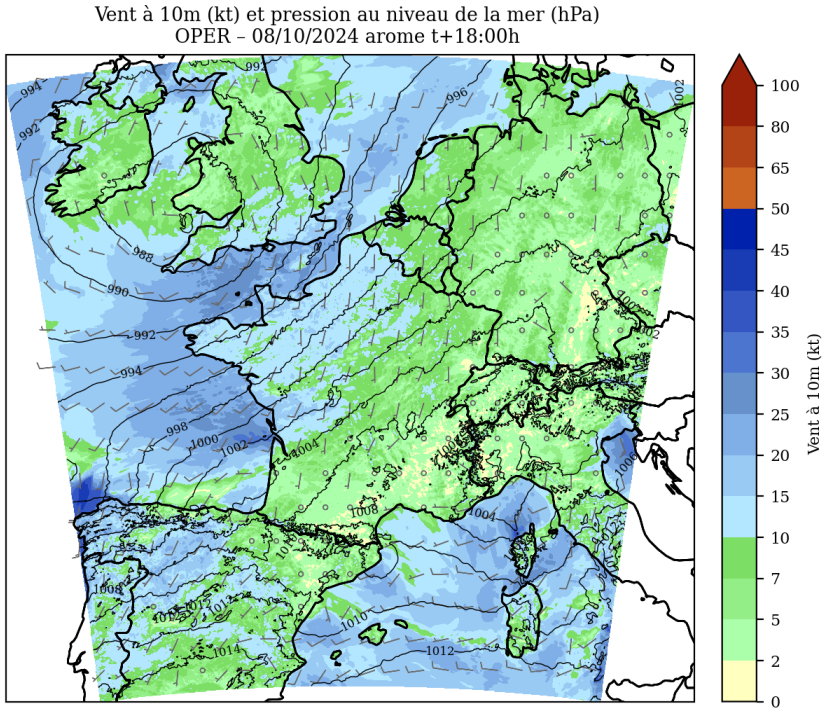


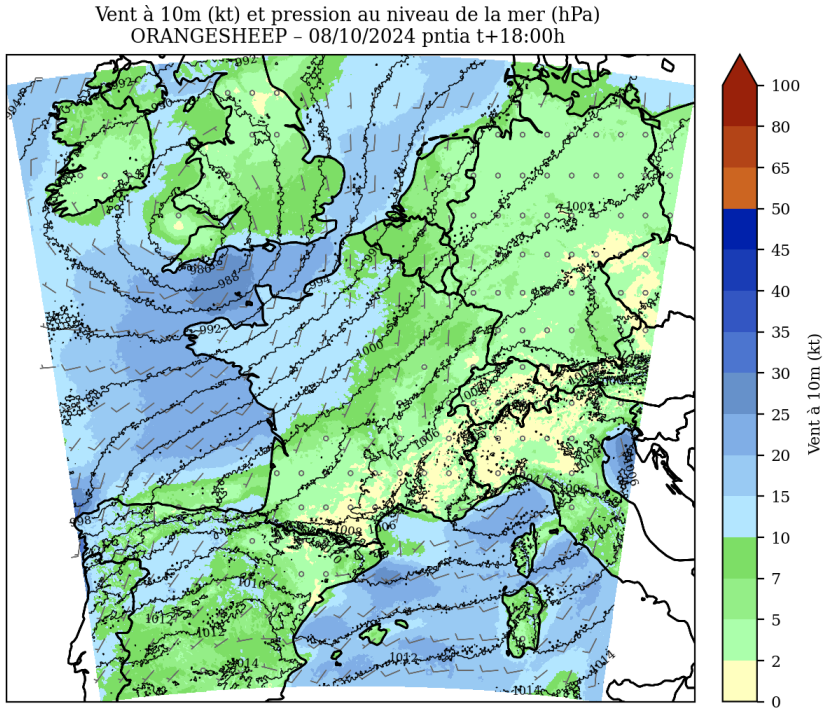
Kirk : surface wind



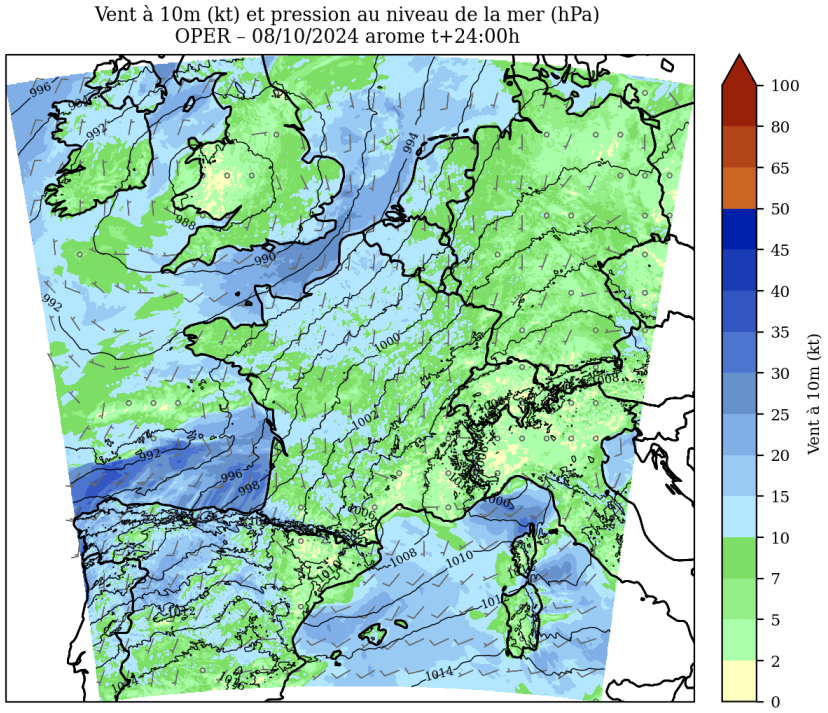


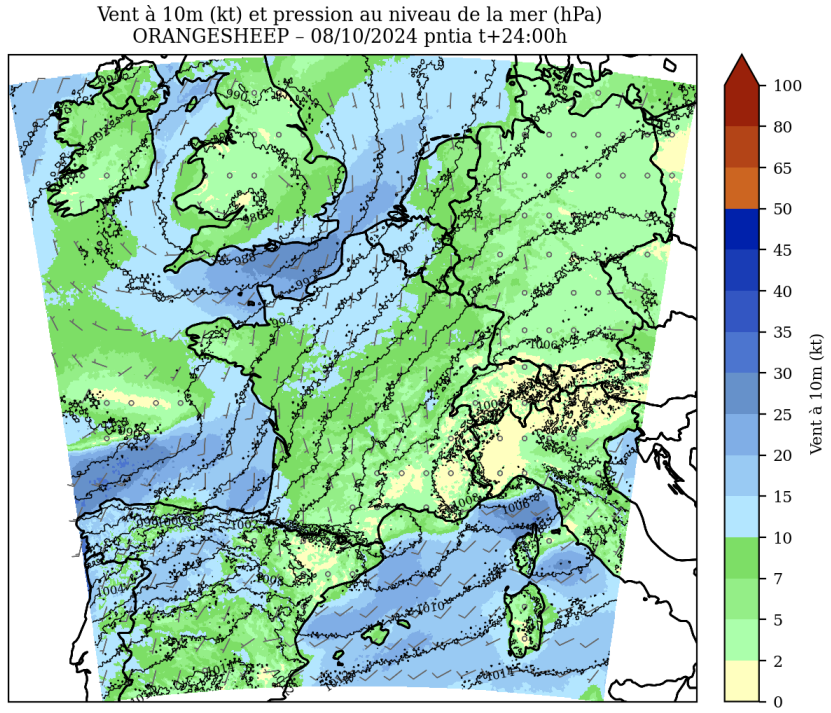
Kirk : surface wind



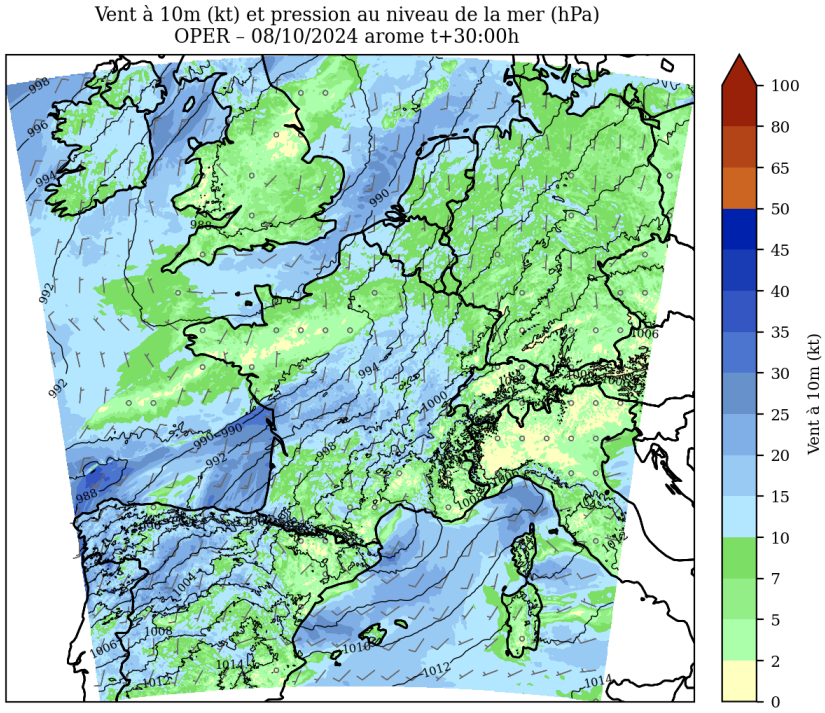


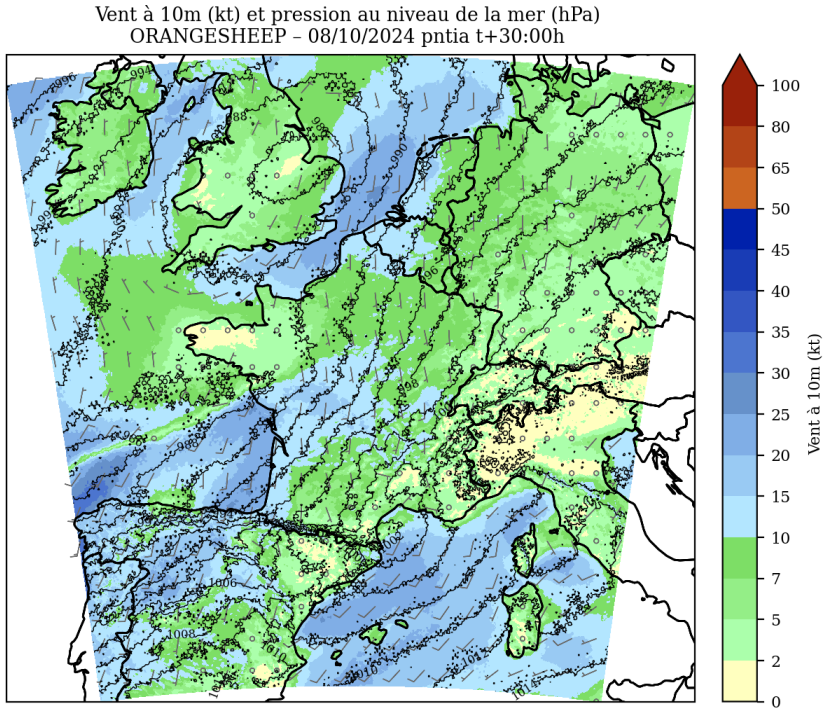
Kirk : surface wind



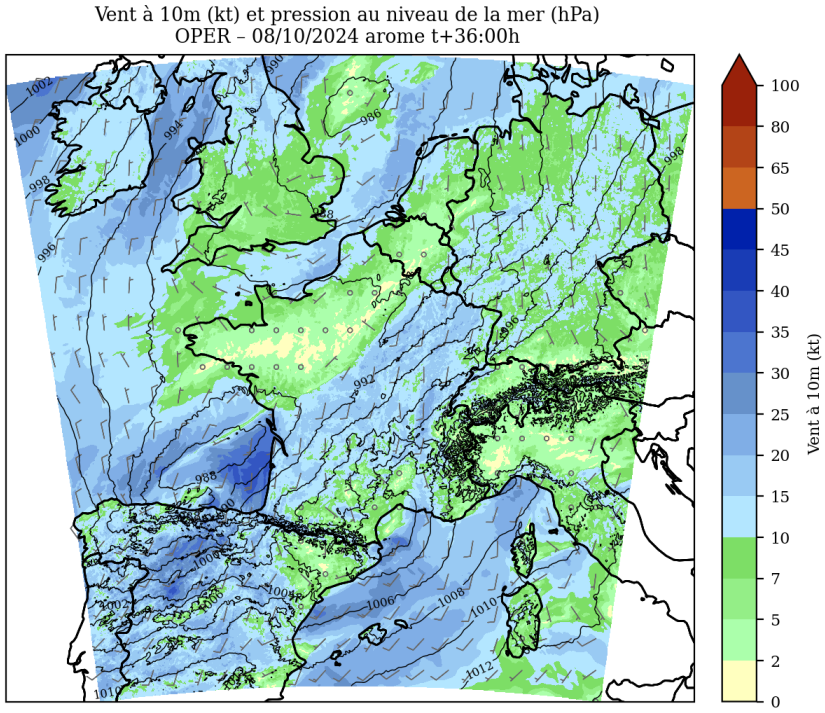


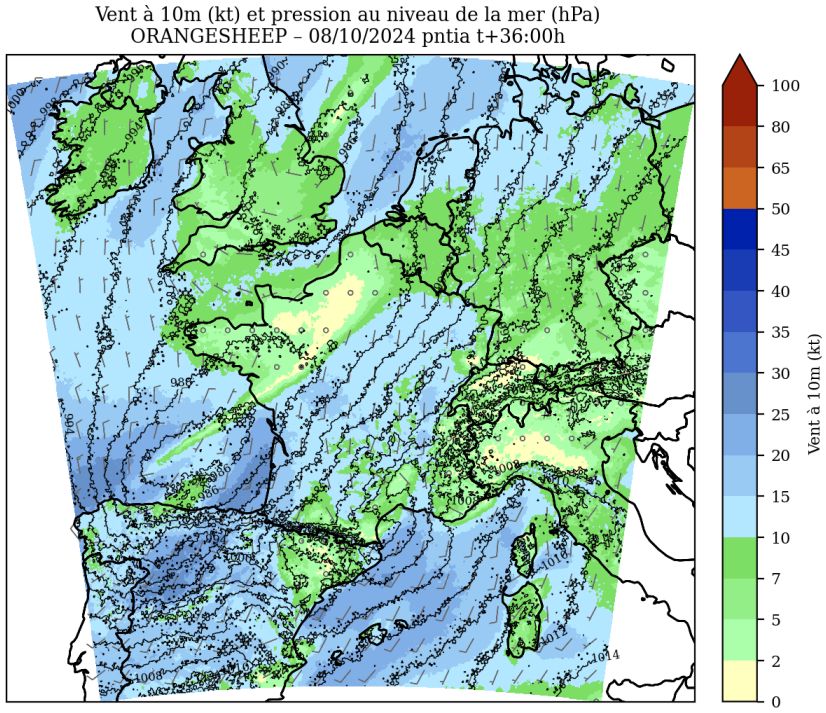
Kirk : surface wind



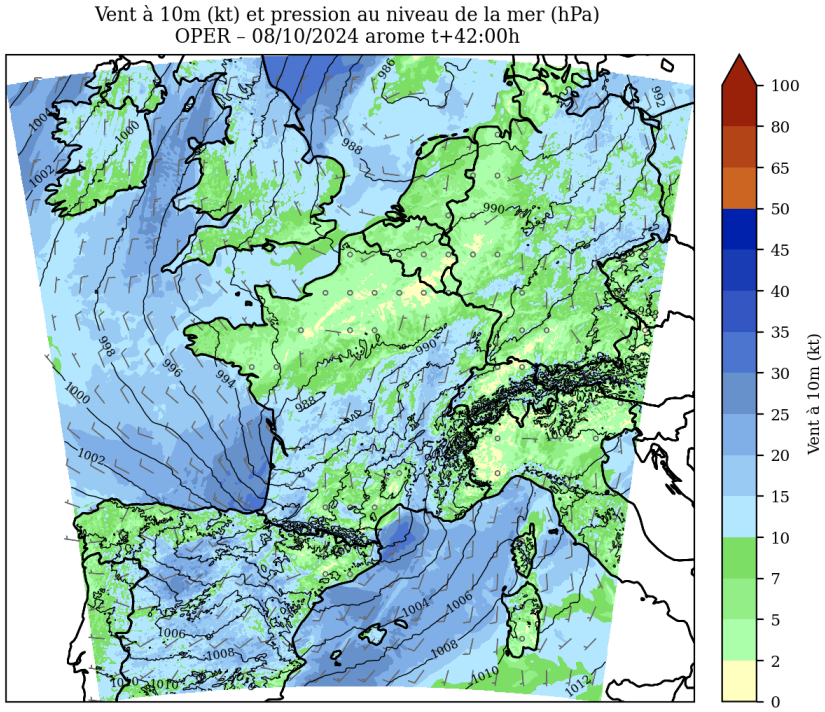


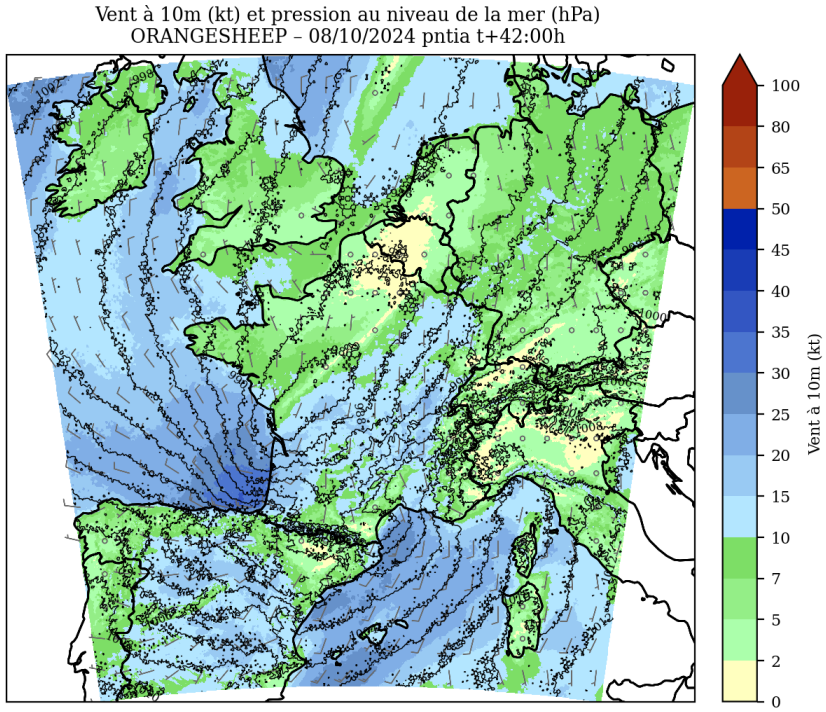
Kirk : surface wind



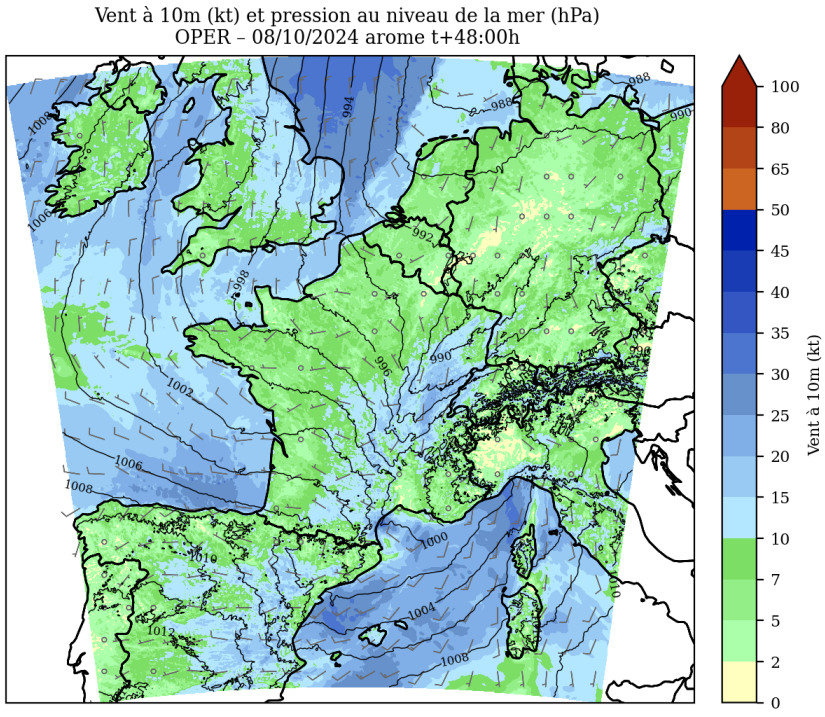


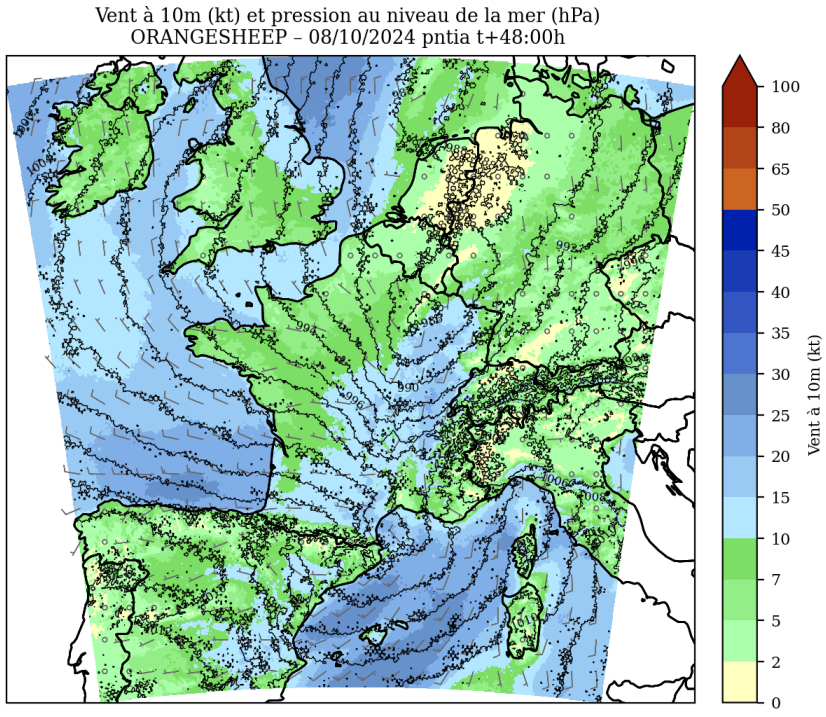
Kirk : surface wind



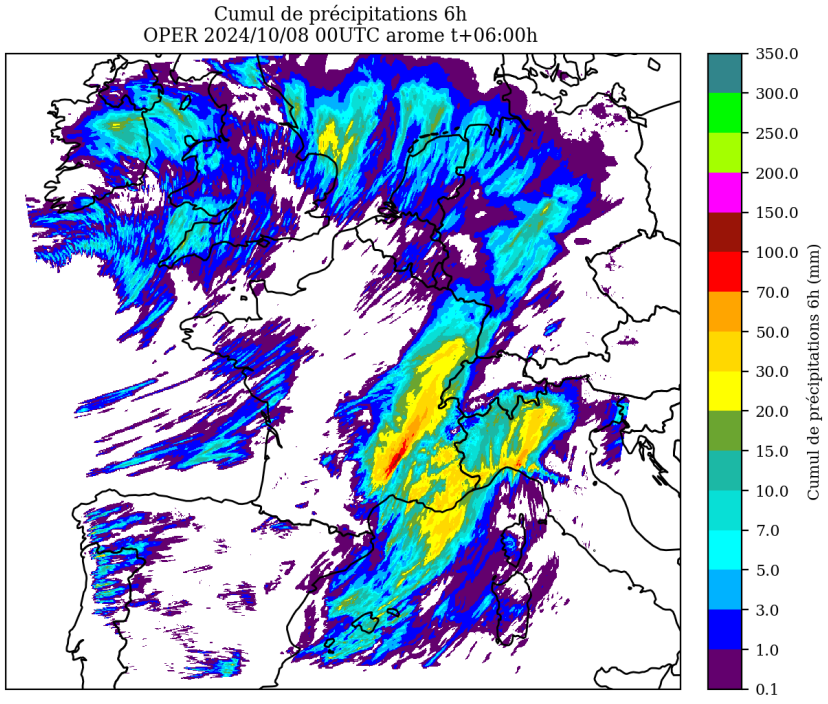


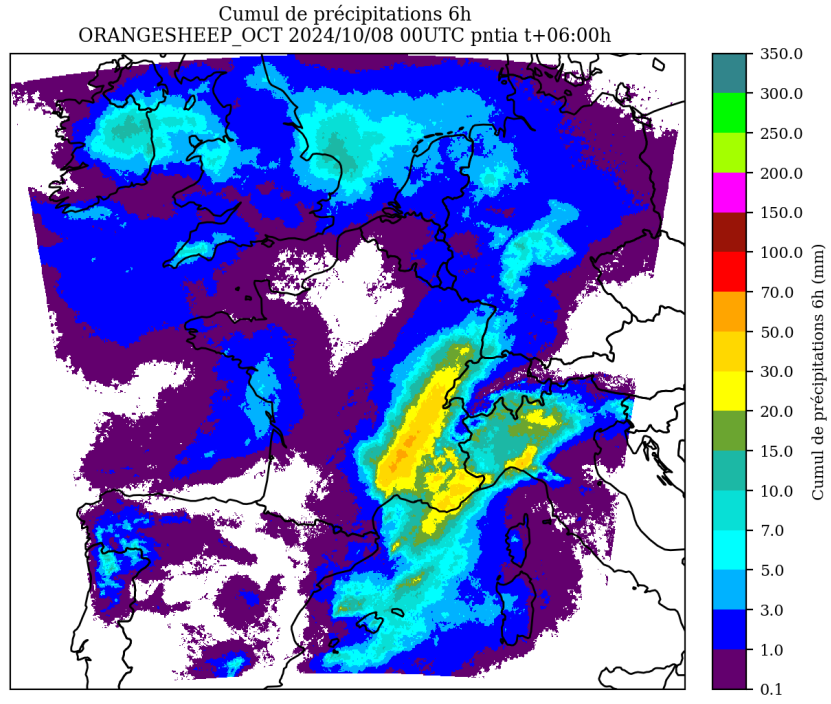
Kirk : surface wind



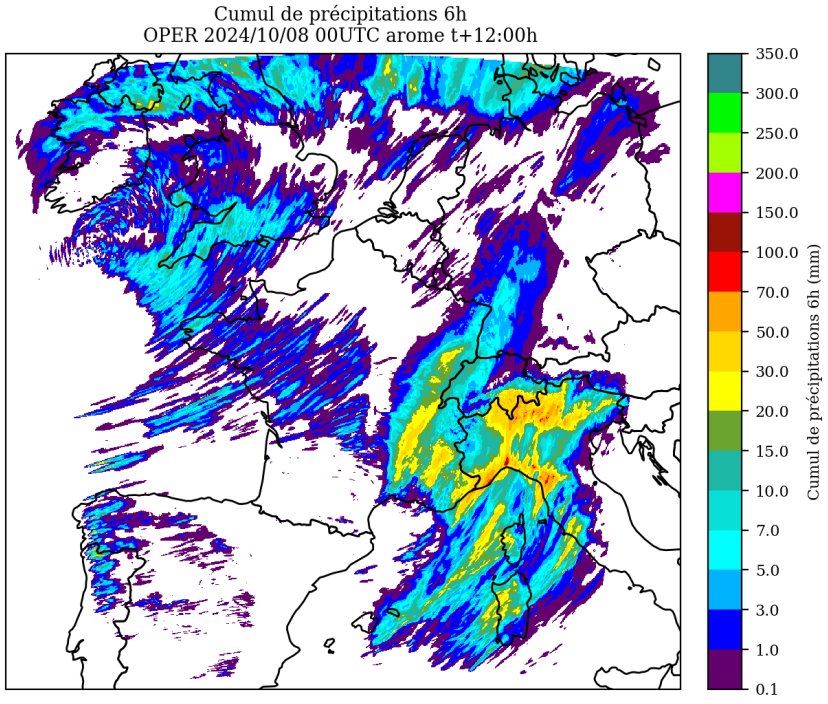


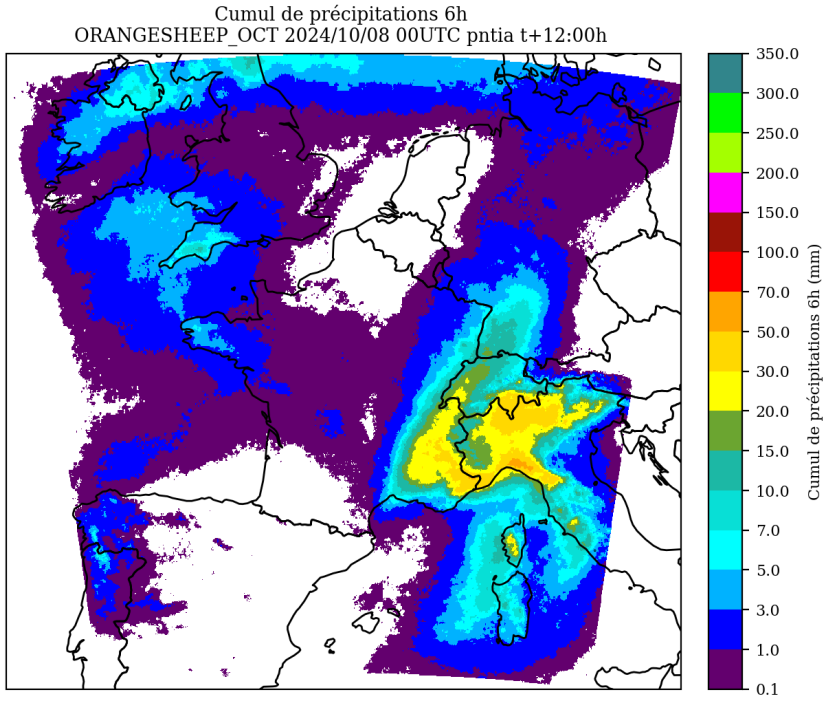
Kirk : rainfall



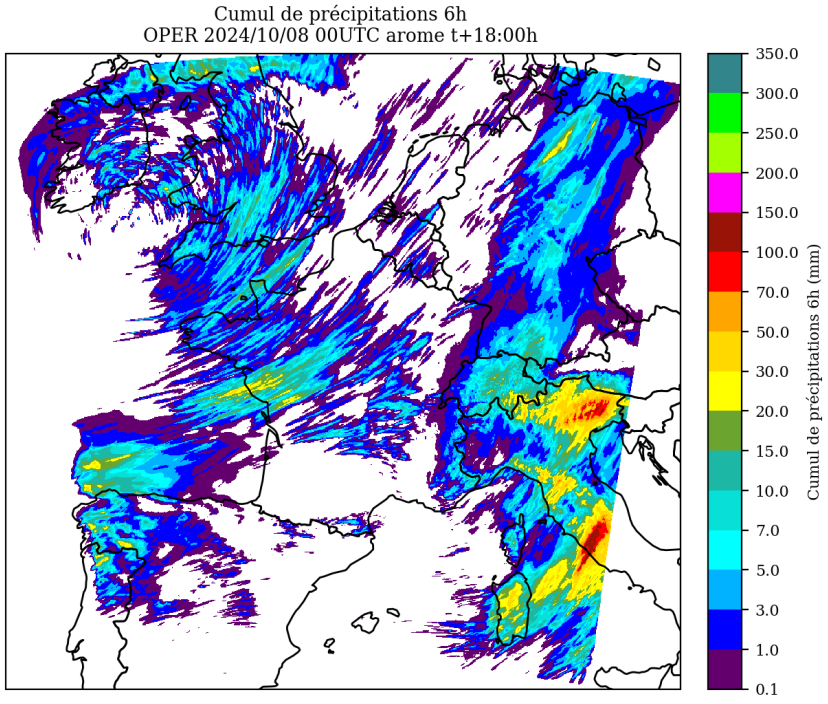


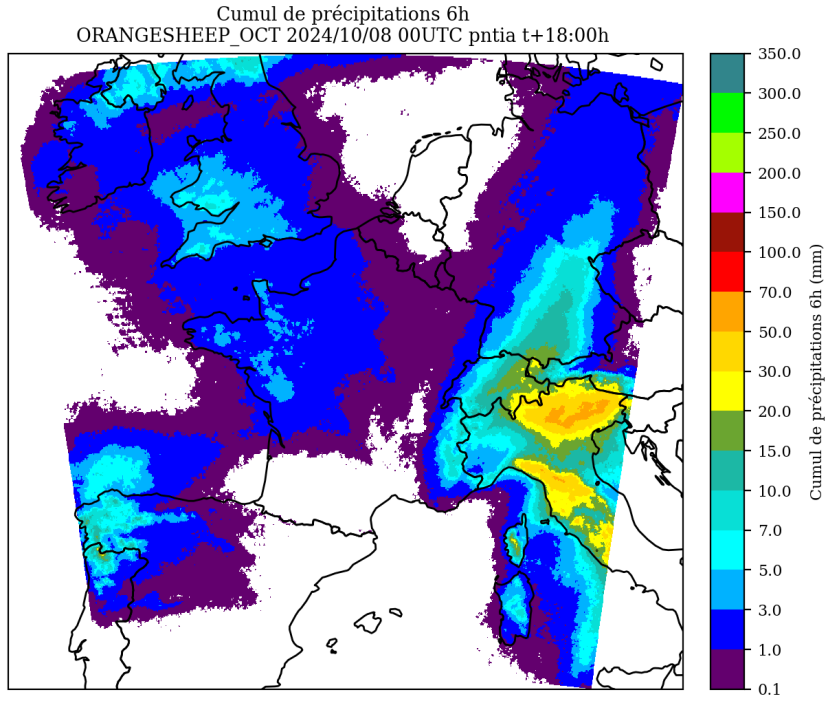
Kirk : rainfall



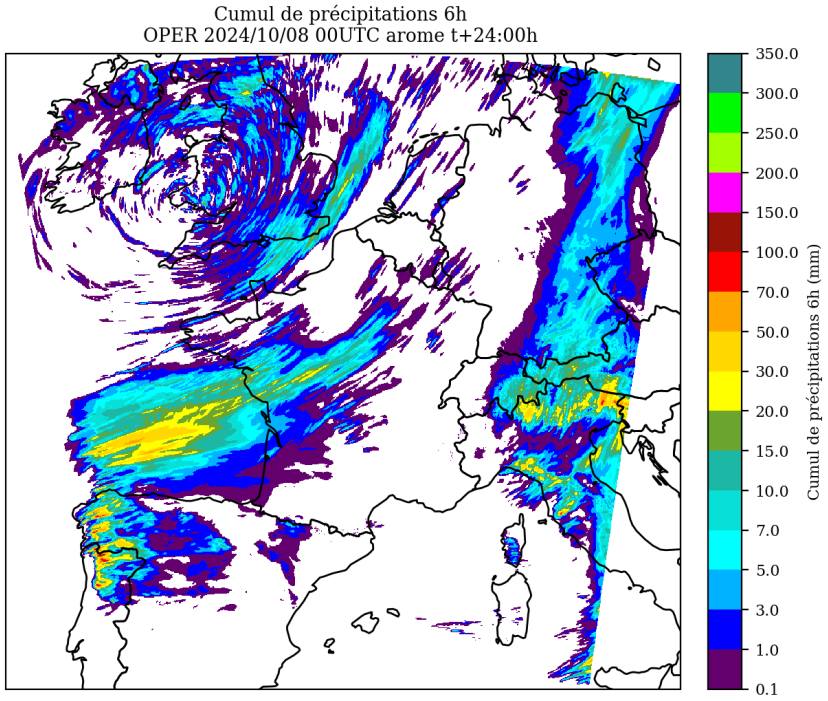


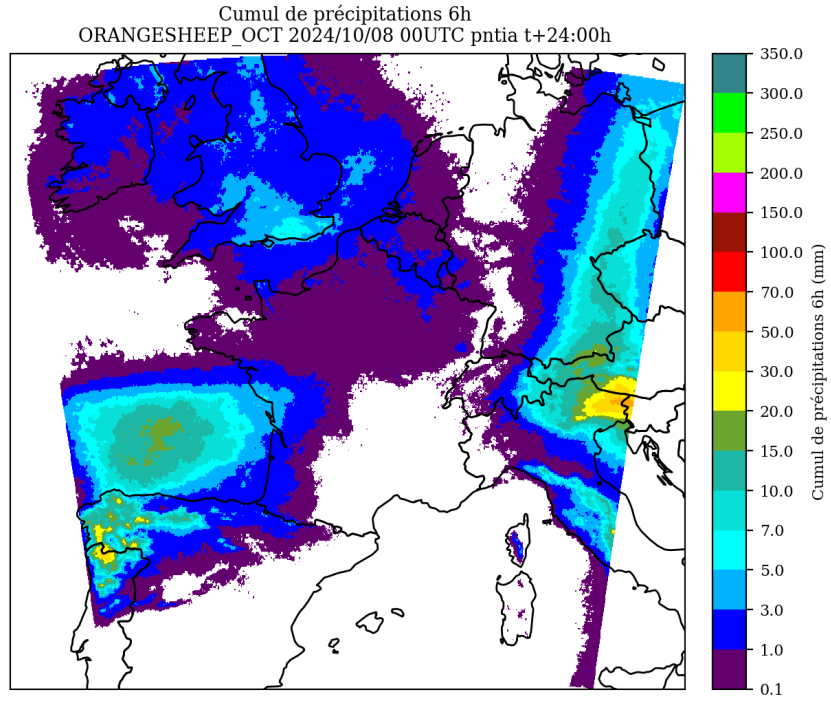
Kirk : rainfall



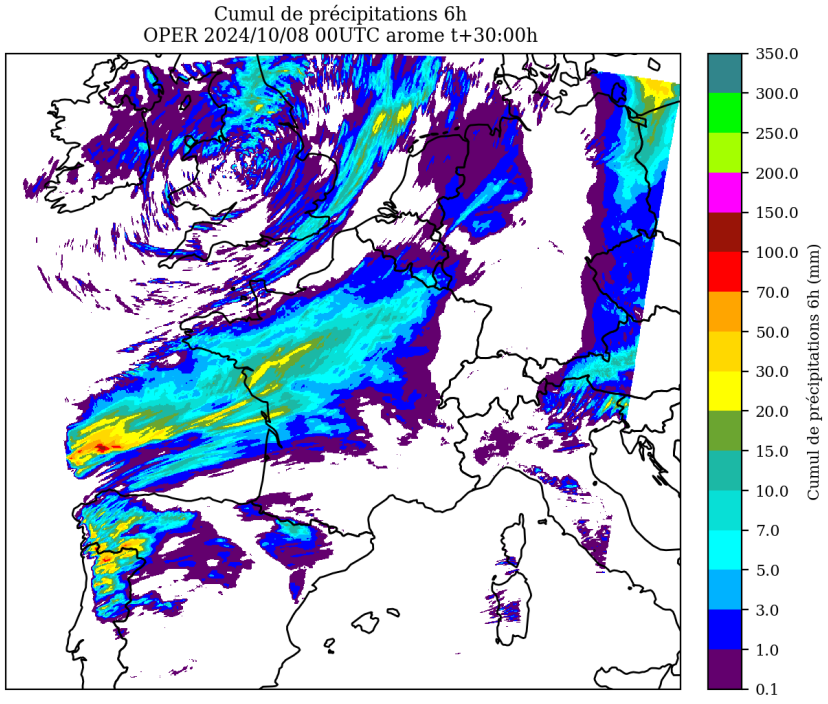


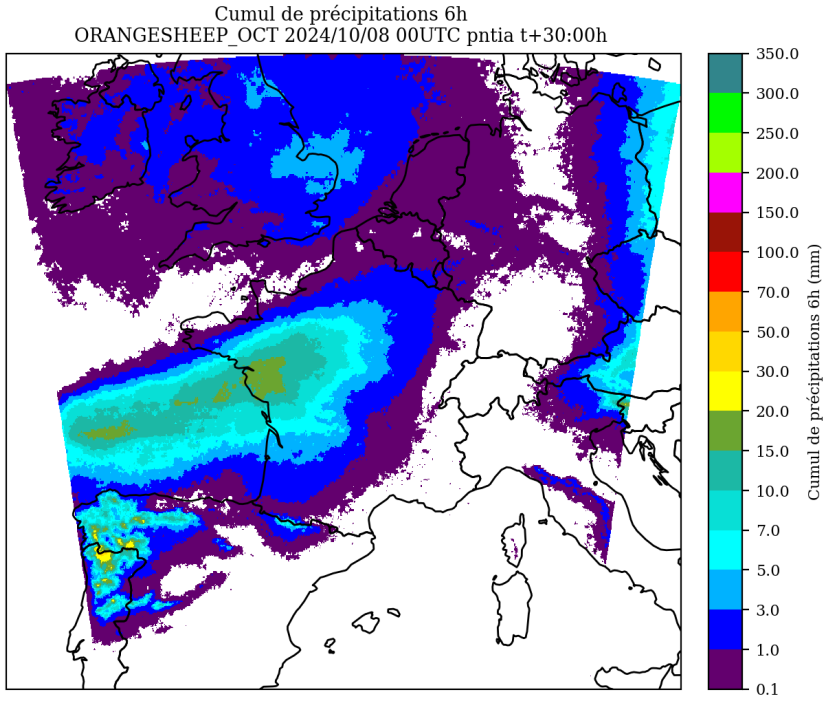
Kirk : rainfall



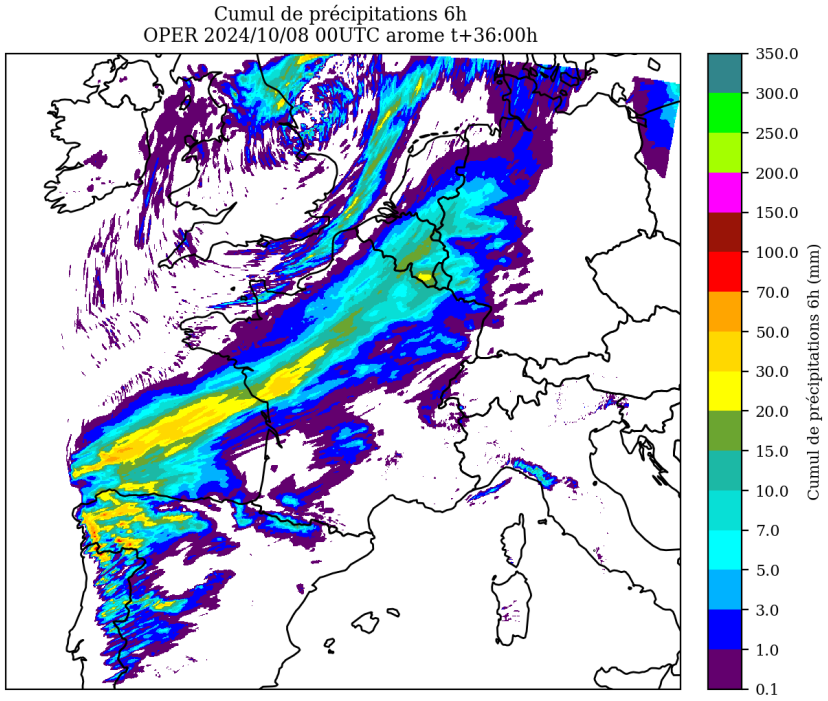


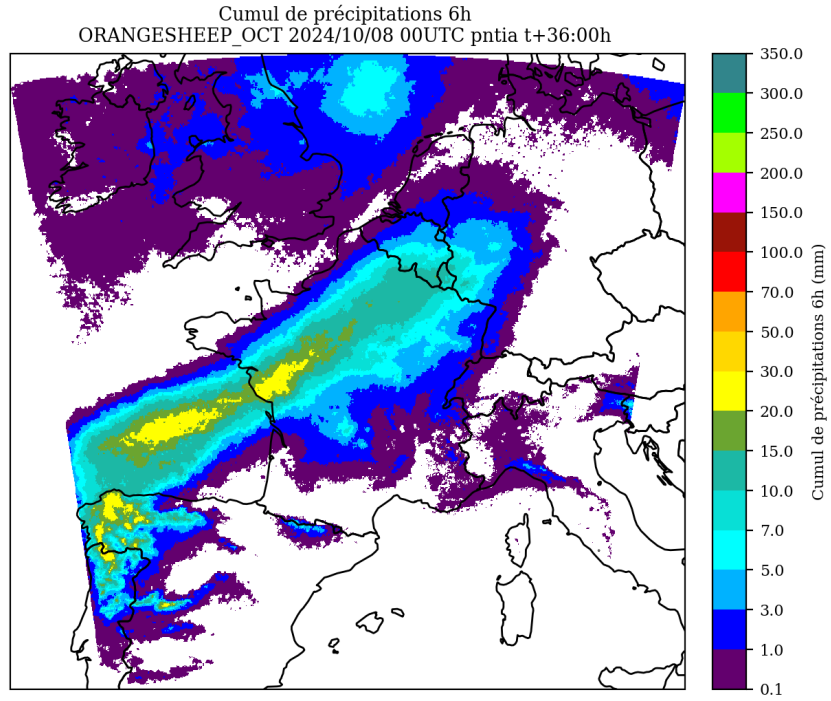
Kirk : rainfall



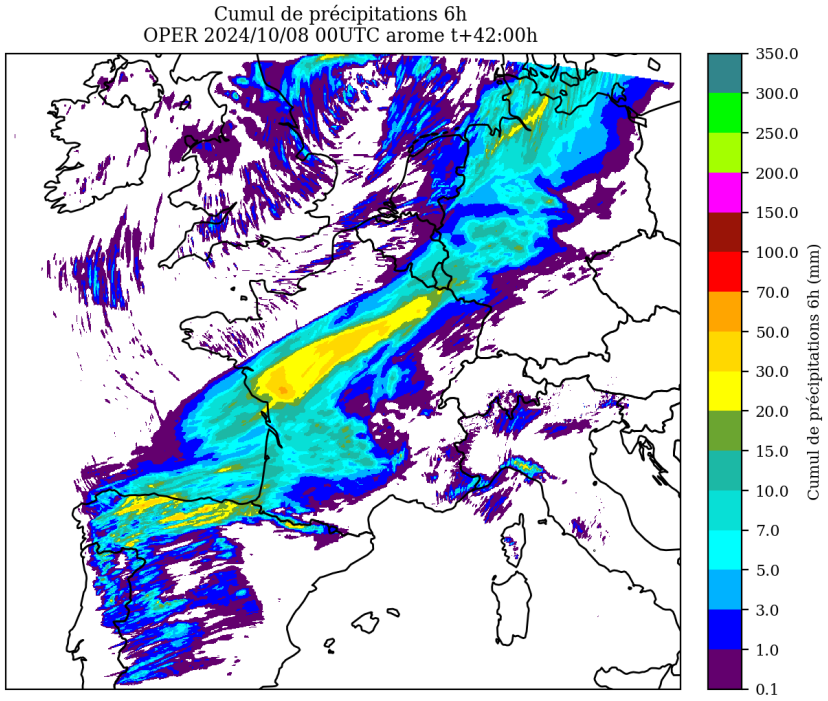


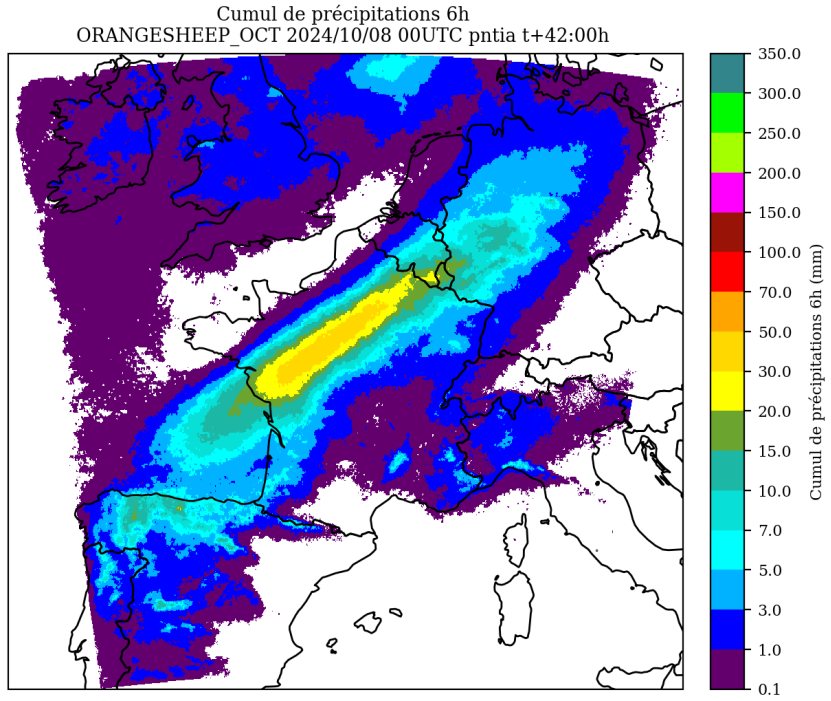
Kirk : rainfall



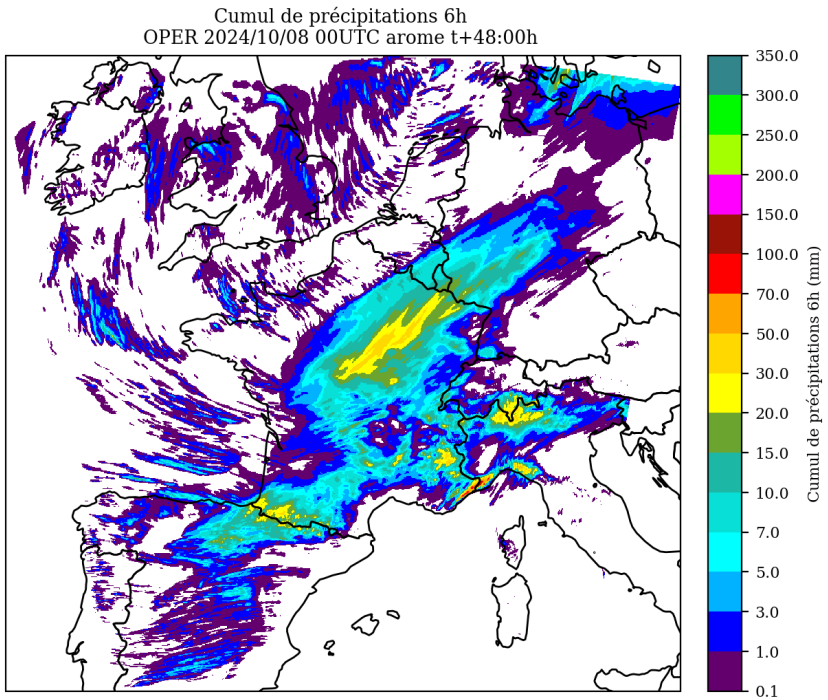


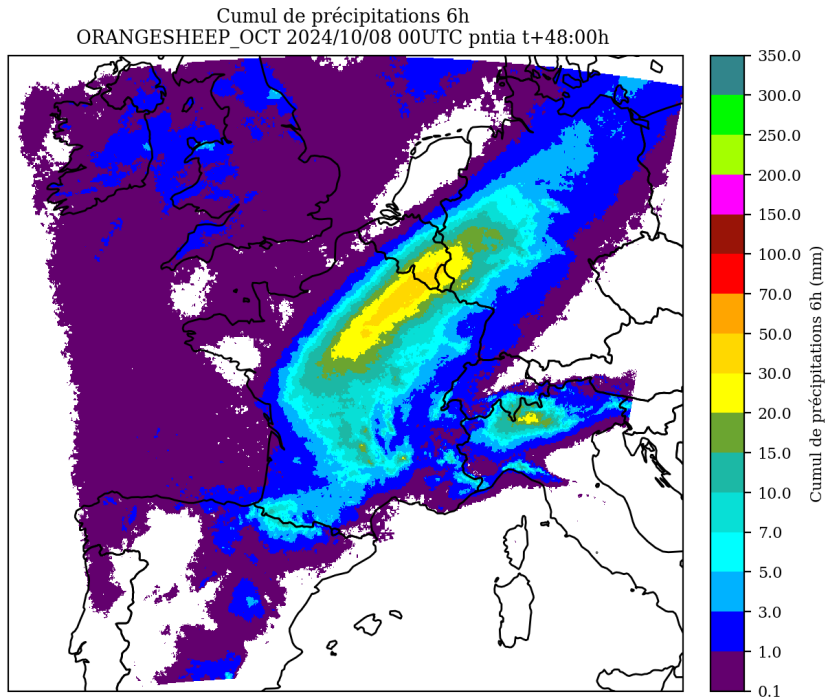
Kirk : rainfall



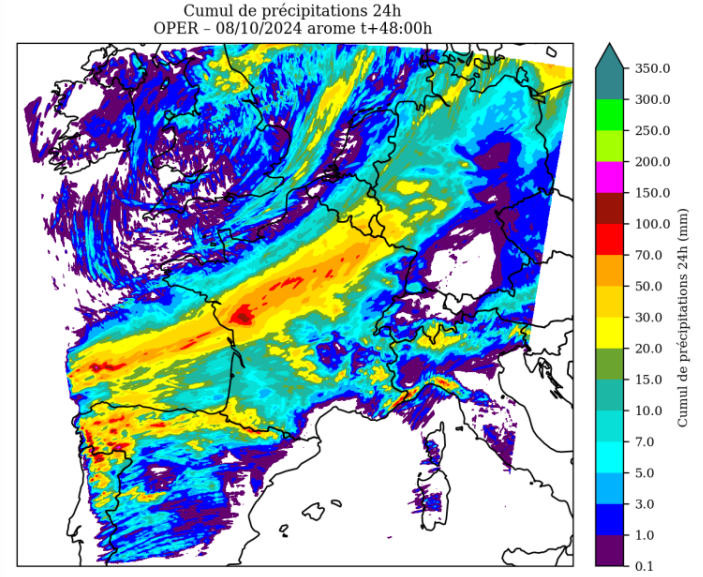


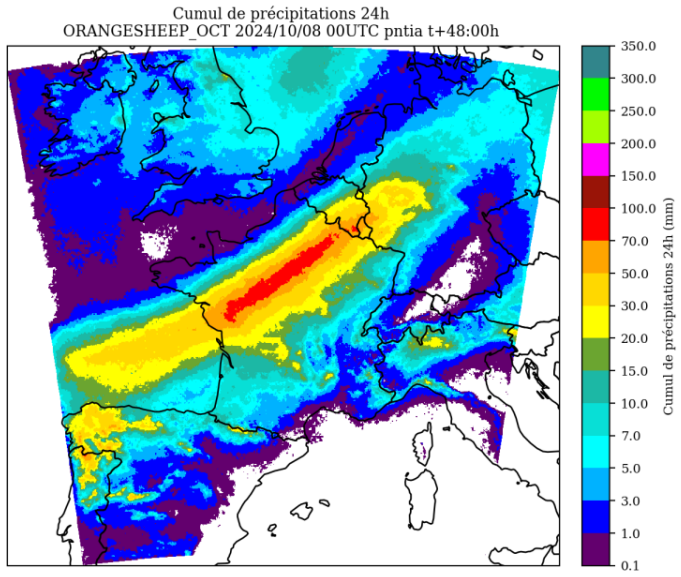
Kirk : rainfall

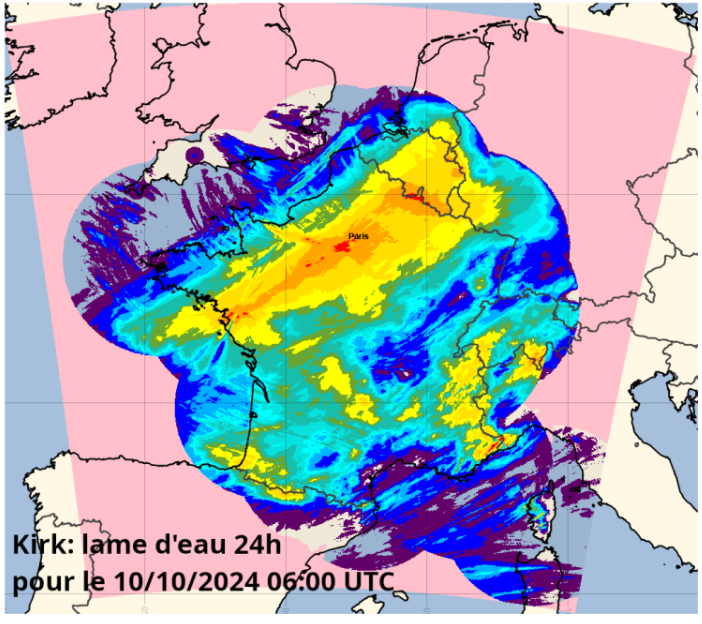




Kirk : event accumulated rainfall







Position and timeline of precipitation are correctly forecast

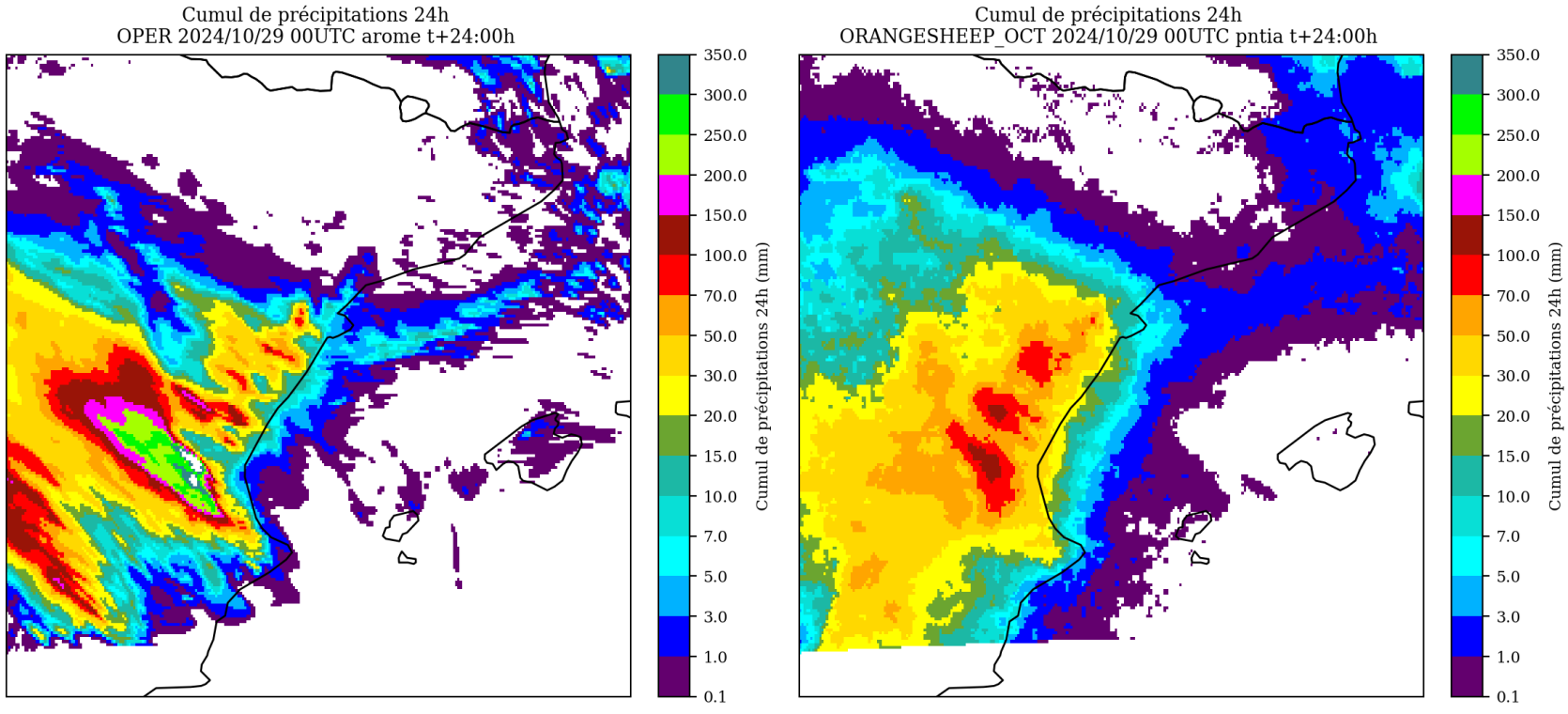
Intensity of rainfall inaccurate

- too strong on Center
- too weak on Alps and South West

Weird field texture,
both blurry and noisy

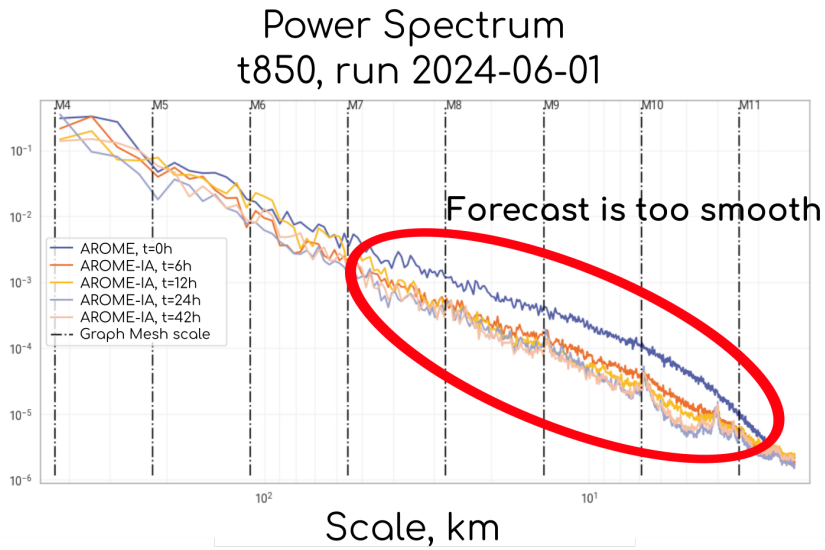
The Valence ”DANA” event

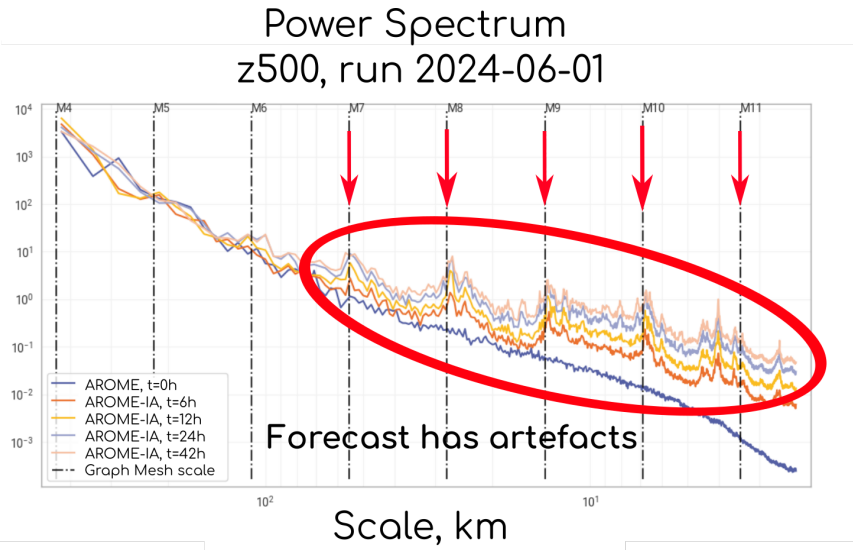
Valence : AROME and AROME-IA



- **Intensity** : at max, AROME-IA < 150 mm while AROME > 350 mm, obs. = 771 mm (Turís)
- **Position** grossly correct. **Something is wrong with the texture**

Spectra : quantifying texture





- Common features for many models (not only ours)
- Potential fixes :
  - ▷ longer training
  - ▷ different optimization objectives
  - ▷ more consistent data

Conclusion

Conclusions

- **MLWP scores are comparable to AROME**, for some variables ✔️
- **MLWP still struggles**:
  - 🌧️ for precipitations, **position is ok**, but **intensity is inaccurate**
  - 💥 on the fields' **activity**: smoothing + artefacts
  - ⛈️ for **below-meso-scale features**: e.g convection

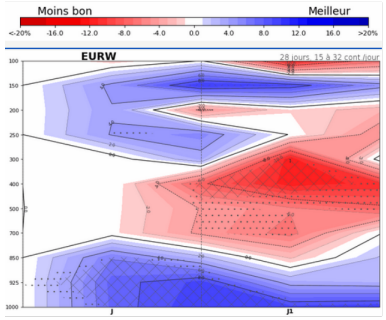
Conclusions and perspectives 🔍

Short term perspectives…

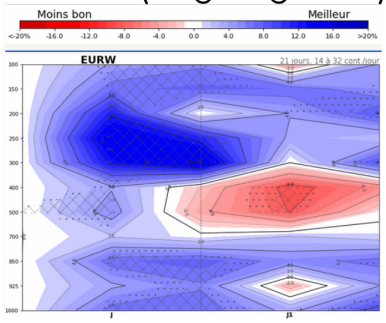
- **Adapt** the training strategy
- Use **ARPEGE**-analysis for coupling
- Exploit the **AROME-Reanalysis** (ARRA) dataset (1962-2020).

Training with a larger rate

Previous



Current (ongoing work)



…And beyond

- **Develop diagnostics** to quantify deviation from physics
- **Add observations** as inputs
- Develop **ensemble generation** methods

Glossaire

- **Auto-régression :** estimation de l'état suivant d'un système en fonction de son état présent.
- **Fine-tuning :** technique visant à continuer sur de nouvelles données ou de nouvelles tâches un entrainement déjà fait précédemment.
- **Learning rate :** paramètre, noté γ, contrôlant l'ampleur de la mis à jour des poids lors de l'apprentissage. Plus γ est grand, plus l'apprentissage va vite, mais il risque de diverger. Plus γ est petit, plus l'entrainement est stable, mais plus il est lent.
- **Transfer learning :** type d'apprentissage visant à transférer des connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles.
- **Rollout :** avancement dans le temps par auto-régression.