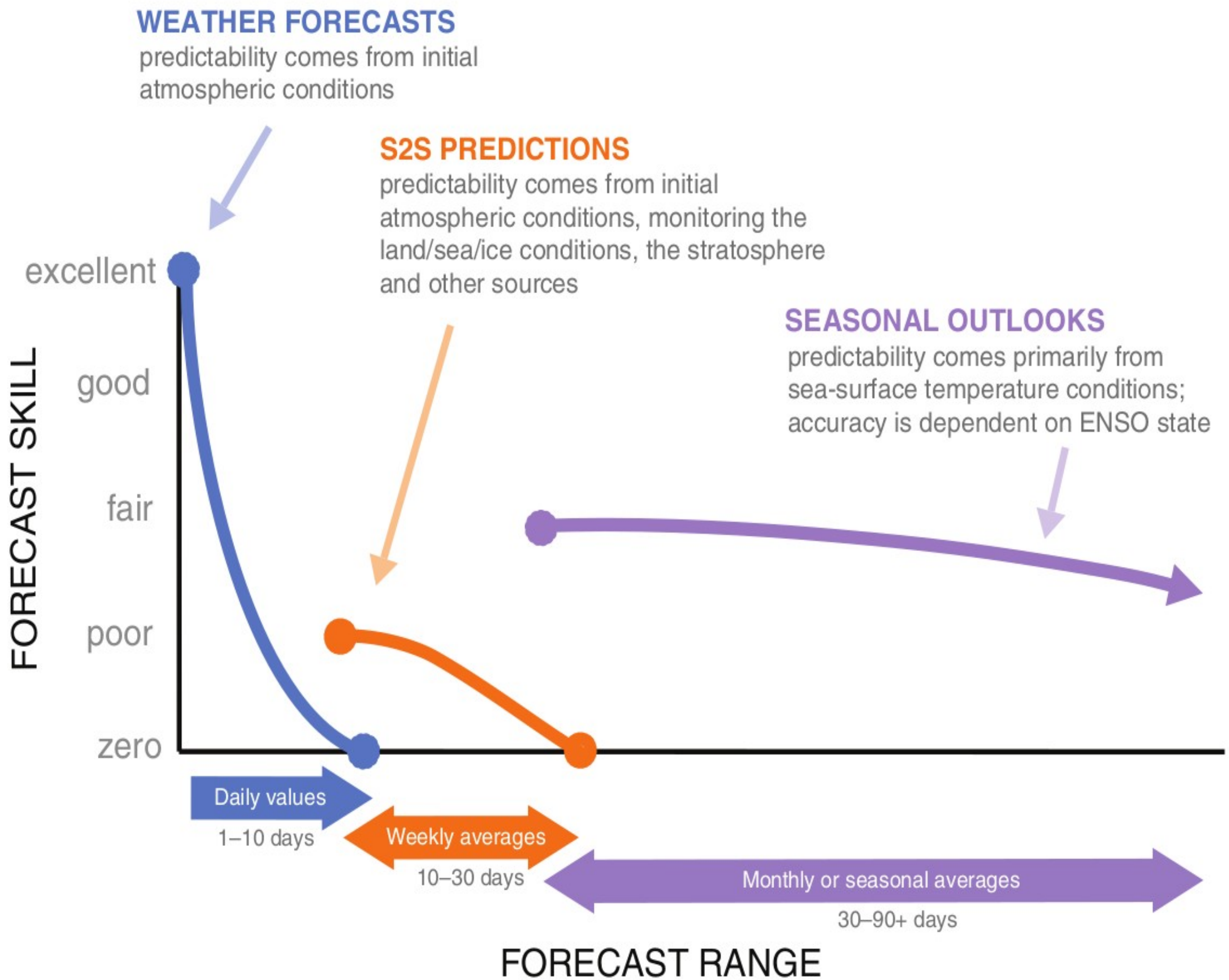
# Improving dynamical sub-seasonal forecasts with machine learning and ensemble barycenters

Anastase Charantonis, Rémi Flamary, **Naveen Goutham, Camille Le Coz,** Riwal Plougonven, Peter Tankov, Alexis Tantet, **Ganglin Tian**

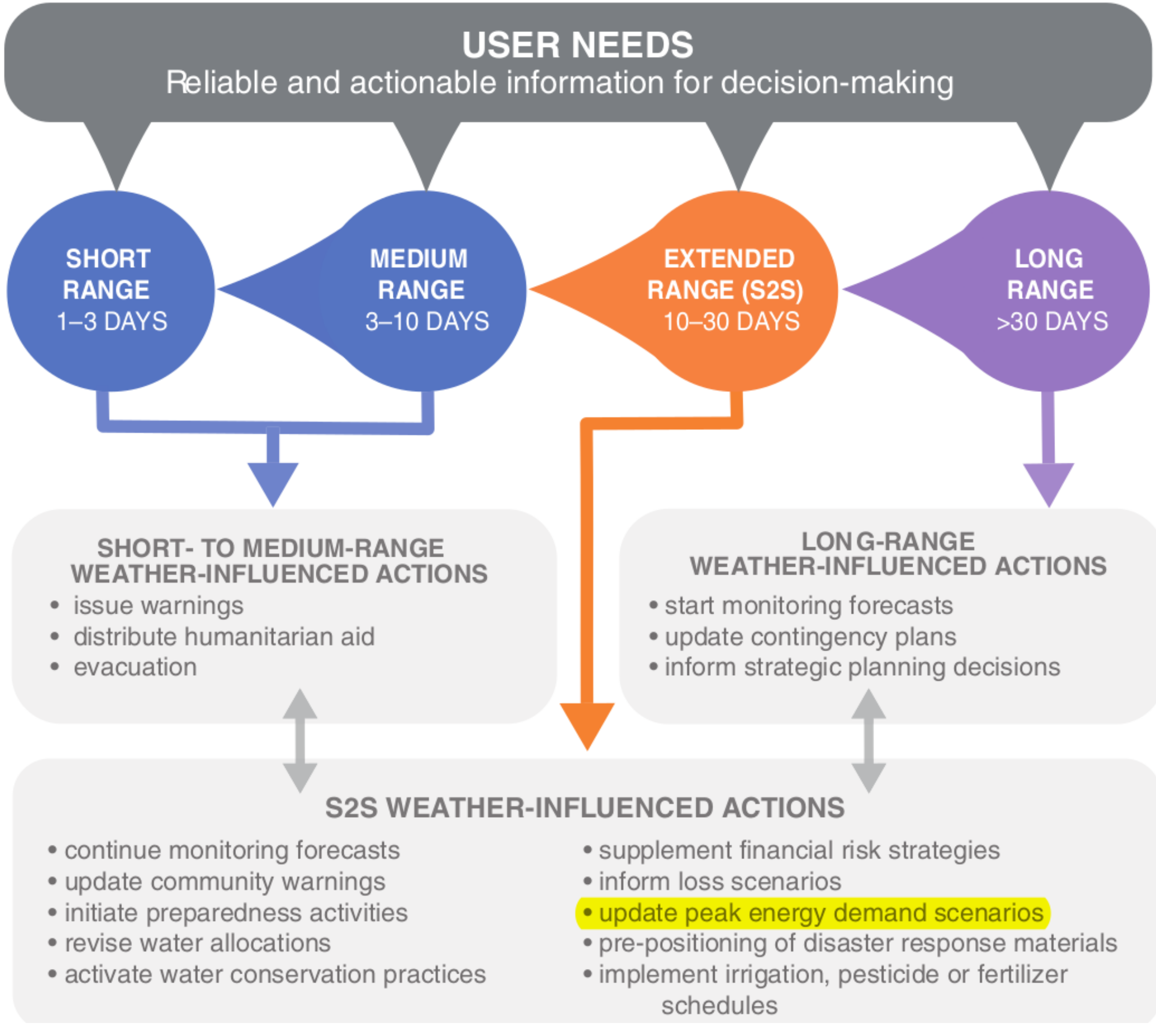
# What are sub-seasonal forecasts? (White *et al., 2017*)



## Why sub-seasonal forecasts? (White *et al., 2017)*

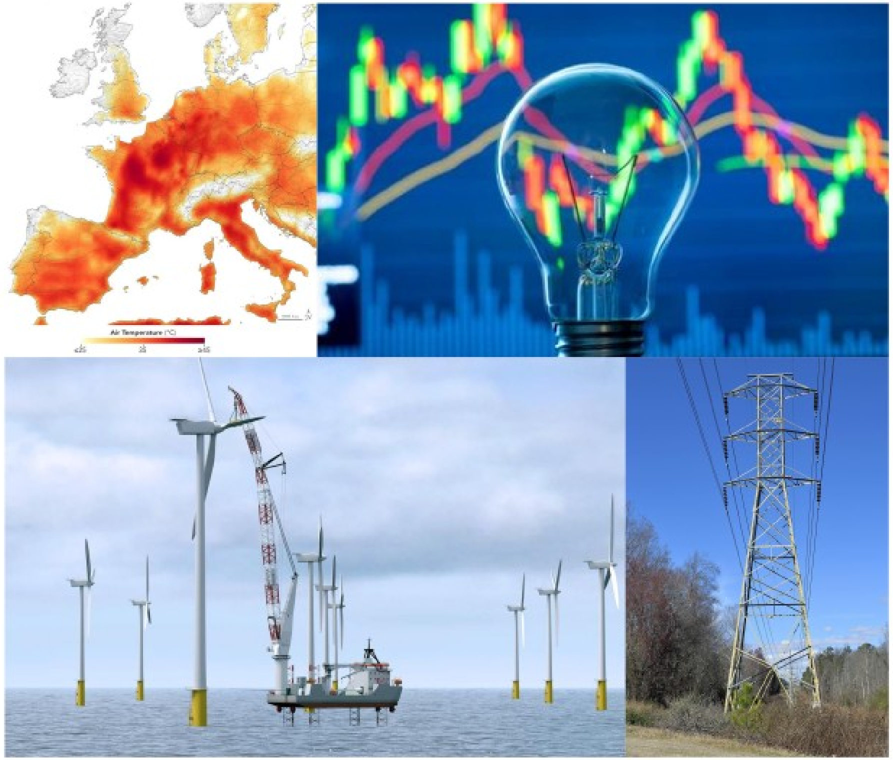
Interesting dynamical problem

+



# Potential use of S2S forecasts in energy sector

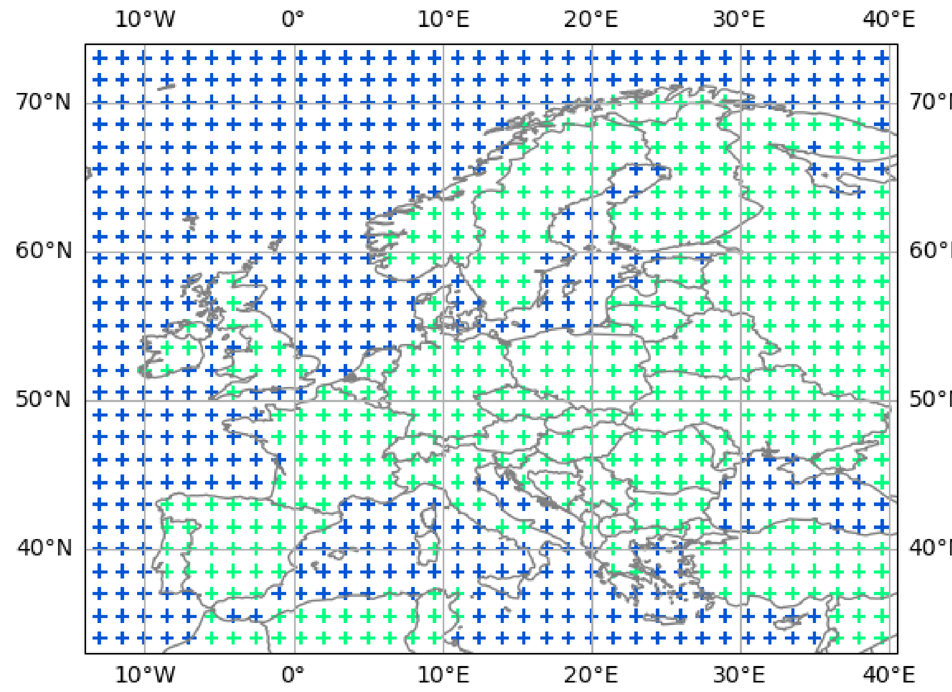
- ➔ Risk assessment
- ➔ Determine capacity reserve level
- ➔ Schedule maintenance
- ➔ Trading/hedging
- ➔ Estimate grid transmission capacity



E.g. Dubus (2014), White *et al.* (2017)

# Observables of interest here

- **Meteorological variables** rather than application-specific variables to provide baseline measure of skills:
    - 2m air temperature (**T2m**)
      → mind the **trend**!
    - 100m wind speed (**W100m**)
      → mind the **height**!
- **Winters** (DJF)
- **Weekly averages**
- Max lead time: **week 6**
- **Europe**
- **Grid-point scale** (0.9°/2.7°)



# Selected literature on direct NWP forecasts

- **Surface temperature**:
  Vitart (2014), Buizza & Leubecher (2015), Monhart *et al.* (2018), Büeler *et al.* (2020), Dorrington *et al.* (2020)

- **Wind speed**:
  Lynch *et al.* (2014), Lledó & Doblas-Reyes (2020), Büeler *et al.* (2020)

Ground-based stations
≠ Domain
≠ Metric
Need for updates

10m instead of 100m

# Why more than direct NWP forecasts?

## Why more than direct NWP forecasts?

- If there is a **strong dependence** between a surface field of interest and **another field that is forecast better**
  → Complement direct forecast with information transferred from dependent field using **supervised learning**

  Schepen *et al.* (2012, 2014, 2016), Orth & Seneviratne (2014), Alonzo *et al.* (2017), Kämäräinen *et al.* (2019), Strazzo *et al.* (2019), Ramon *et al.* (2021)

## Why more than direct NWP forecasts?

- If there is a **strong dependence** between a surface field of interest and **another field that is forecast better**
  → Complement direct forecast with information transferred from dependent field using **supervised learning**
  Schepen *et al.* (2012, 2014, 2016), Orth & Seneviratne (2014), Alonzo *et al.* (2017), Kämäräinen *et al.* (2019), Strazzo *et al.* (2019), Ramon *et al.* (2021)
- If the dependence is **nonlinear**
  → Add nonlinearities in statistical model using **Convolutional Neural Networks (CNN)**
  Höhlein *et al.* (2020)

## Why more than direct NWP forecasts?

- If there is a **strong dependence** between a surface field of interest and **another field that is forecast better**
  → Complement direct forecast with information transferred from dependent field using **supervised learning**
  Schepen *et al.* (2012, 2014, 2016), Orth & Seneviratne (2014), Alonzo *et al.* (2017), Kämäräinen *et al.* (2019), Strazzo *et al.* (2019), Ramon *et al.* (2021)
- If the dependence is **nonlinear**
  → Add nonlinearities in statistical model using **Convolutional Neural Networks (CNN)**
  Höhlein *et al.* (2020)
- If there is a **weak dependence** between **ensemble forecasts model errors from different NWP models**
  → Aggregate ensembles from multiple models using **ensemble barycenters**
  Ning *et al.* (2014), Robin *et al.* (2017, 2019), Papayiannis *et al.* (2018), Vissio & Lucarini (2018), Vissio *et al.* (2020)

# Research questions

I. Are the **dynamical European sub-seasonal predictions** of wind speed and temperature more **skillful** than climatology?

**NWP** T2m + W100m *N. Goutham et al., 2022*

## Research questions

I. Are the **dynamical European sub-seasonal predictions** of wind speed and temperature more **skillful** than climatology?

**NWP** T2m + W100m *N. Goutham et al., 2022*

II. Can these predictions be improved by using machine learning to **combine direct forecasts with information from other fields**?

**NWP** + **ML** T2m + W100m *N. Goutham et al., 2023*

# Research questions

I. Are the **dynamical European sub-seasonal predictions** of wind speed and temperature more **skillful** than climatology?

**NWP** T2m + W100m *N. Goutham et al., 2022*

II. Can these predictions be improved by using machine learning to **combine direct forecasts with information from other fields**?

**NWP** + **ML** T2m + W100m *N. Goutham et al., 2023*

III. Is the relationship between these fields **linear** at these horizons?

**NWP + ML-Deep** W100m *G. Tian, on-going*

## Research questions

I. Are the **dynamical European sub-seasonal predictions** of wind speed and temperature more **skillful** than climatology?

**NWP** — T2m + W100m — *N. Goutham et al., 2022*

II. Can these predictions be improved by using machine learning to **combine direct forecasts with information from other fields**?

**NWP + ML** — T2m + W100m — *N. Goutham et al., 2023*

III. Is the relationship between these fields **linear** at these horizons?

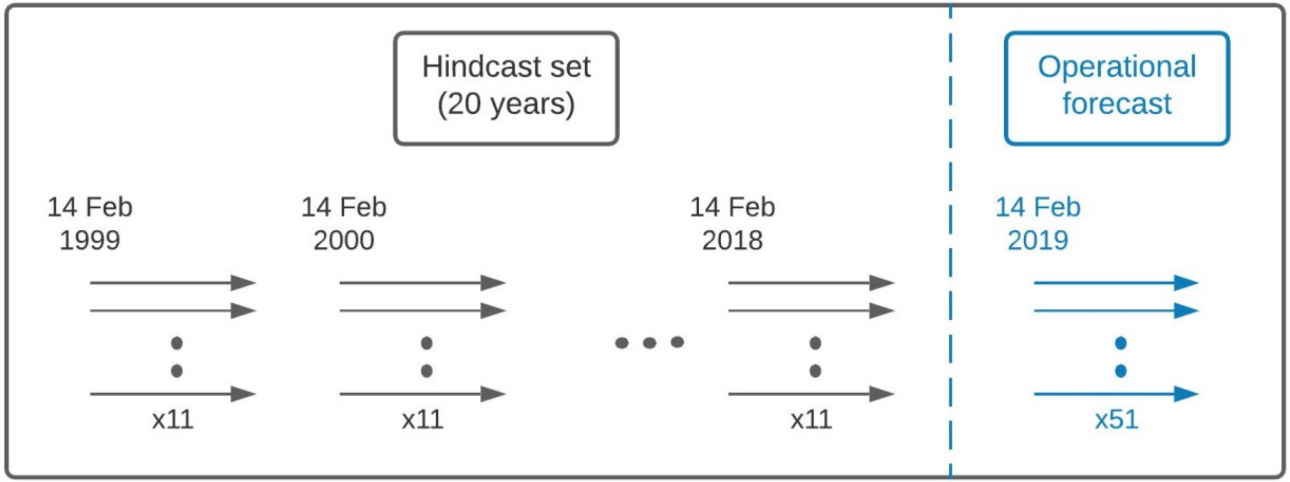
**NWP + ML-Deep** — W100m — *G. Tian, on-going*

IV. How to efficiently **combine ensemble** sub-seasonal forecasts **from different NWP models?**

**NWP (+ ML) + MME** — T2m — *C. Le Coz, on-going*

# Common methodology: data types

- Reforecasts (hindcast) from **S2S project** (ECMWF/NCEP) to:
  - Compute climatology
  - Compute trend
  - Calibrate
  - Train (Goutham: 1999-2016, Tian:1994-2014)
- Forecasts from **S2S project** (ECMWF/NCEP) as inputs to:
  - Train (Le Coz: 2015-2022)
  - Test (Goutham: 2016-2020, Tian: 2016-2021, Le Coz: 2015-2022)

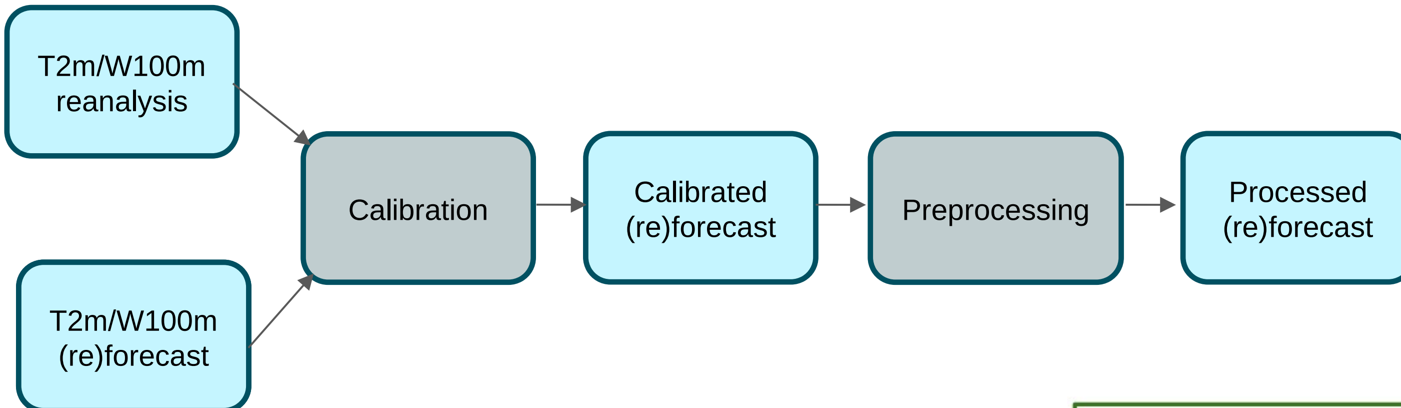



Vitart & Robertson (2018)

# Common methodology: data types

- Reforecasts (hindcast) from **S2S project** (ECMWF/NCEP) to:
  - Compute climatology
  - Compute trend
  - Calibrate
  - Train (Goutham: 1999-2016, Tian:1994-2014)
- Forecasts from **S2S project** (ECMWF/NCEP) as inputs to:
  - Train (Le Coz: 2015-2022)
  - Test (Goutham: 2016-2020, Tian: 2016-2021, Le Coz: 2015-2022)
- Reanalysis (ERA5/MERRA-2) as target (reference) to
  - Train
  - Test
- Continuous Rank Probability Score (CRPS) as
  - Loss function
  - Skill score

# Common methodology: calibration and preprocessing




- Calibration (statistics):
  - Mean-variance adjustment (Leung et al. 1999)
- Preprocessing:
  - Remove trend
  - Remove seasonal cycle

$$\boxed{x_k^* = (x_k - \bar{x}_{cli})\frac{\sigma_{ref}}{\sigma_{cli}} + \bar{o}_{ref}}$$

Where,
$x_k$ = raw member
$\bar{x}_{cli}$ = mean of reforecasts
$\sigma_{cli}$ = std. deviation of reforecasts
$\bar{o}_{ref}$ = mean of observed climatology
$\sigma_{ref}$ = std. deviation of observed climatology

# Common methodology: skill scores (accuracy)

Based on Continuous Rank **Probability** Score (CRPS)

$$\mathrm{CRPS} = \int_{-\infty}^{+\infty} [\mathrm{F(y)} - \mathrm{F_o(y)}]^2 \mathrm{dy}$$

Where,

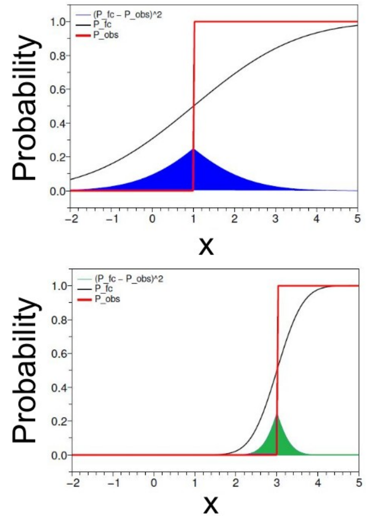
F(y) = empirical CDF of forecasts

$F_o(y)$ = CDF of observation

$$\mathrm{F_o(y)} = \begin{cases} 0, & \text{if } \mathrm{y} < \mathrm{o} \\ 1, & \text{if } \mathrm{y} \geq \mathrm{o} \end{cases}$$

o = observation

**Note**: Best = 0




Credits: ECMWF

# Common methodology: skill scores (accuracy)

Based on Continuous Rank **Probability** Score (CRPS)

- **Average** CRPS Skill Score (CRPSS):

$$CRPSS = 1 - \frac{\langle CRPS_{forecast} \rangle}{\langle CRPS_{climatology} \rangle}$$

$$\mathrm{CRPS} = \int_{-\infty}^{+\infty} [\mathrm{F(y)} - \mathrm{F_o(y)}]^2 \mathrm{dy}$$

Where,

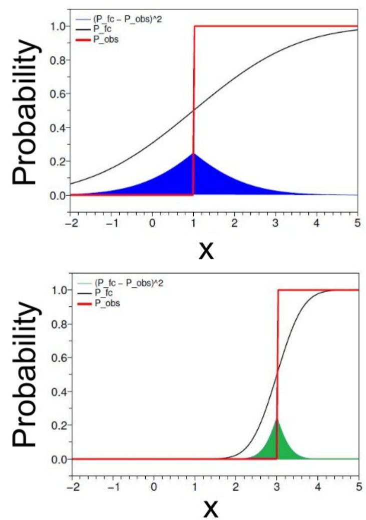
F(y) = empirical CDF of forecasts

$F_o$(y) = CDF of observation

$$\mathrm{F_o(y)} = \begin{cases} 0, & \text{if } \mathrm{y} < \mathrm{o} \\ 1, & \text{if } \mathrm{y} \geq \mathrm{o} \end{cases}$$

o = observation

**Note**: Best = 0




Credits: ECMWF

# Common methodology: skill scores (accuracy)

Based on Continuous Rank **Probability** Score (CRPS)

$$\mathrm{CRPS} = \int_{-\infty}^{+\infty} [\mathrm{F(y)} - \mathrm{F_o(y)}]^2 \mathrm{dy}$$

Where,

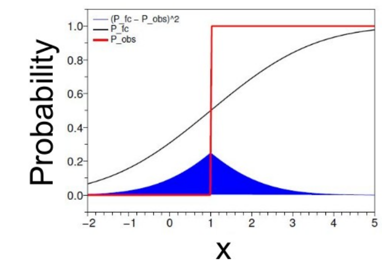
F(y) = empirical CDF of forecasts

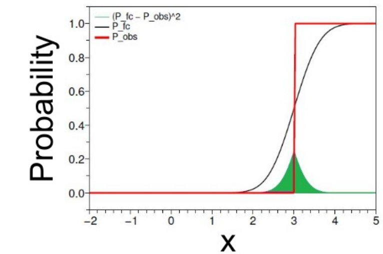
$F_o$(y) = CDF of observation

$$\mathrm{F_o(y)} = \begin{cases} 0, & \text{if } \mathrm{y} < \mathrm{o} \\ 1, & \text{if } \mathrm{y} \geq \mathrm{o} \end{cases}$$

o = observation

**Note**: Best = 0




Credits: ECMWF

- **Average** CRPS Skill Score (CRPSS):

$$CRPSS = 1 - \frac{\langle CRPS_{forecast} \rangle}{\langle CRPS_{climatology} \rangle}$$

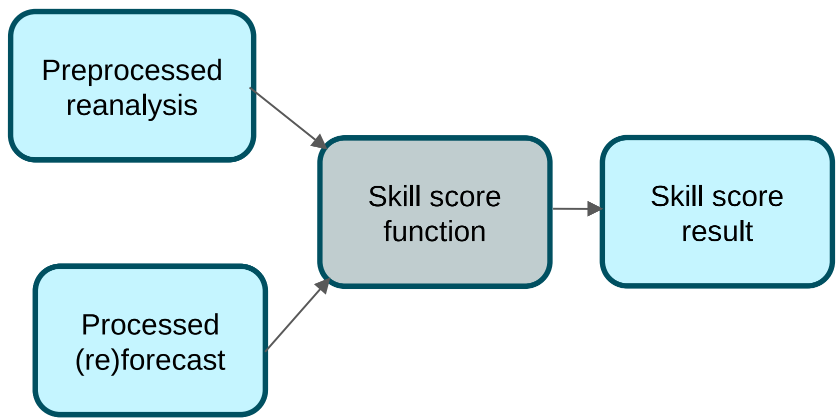
- **Proportion** of Skillful Forecasts (CRPSp):

$$CRPSp = \frac{\#\{CRPS_{forecast} > CRPS_{climatology}\}}{\#\{CRPS_{climatology}\}} \times 100$$

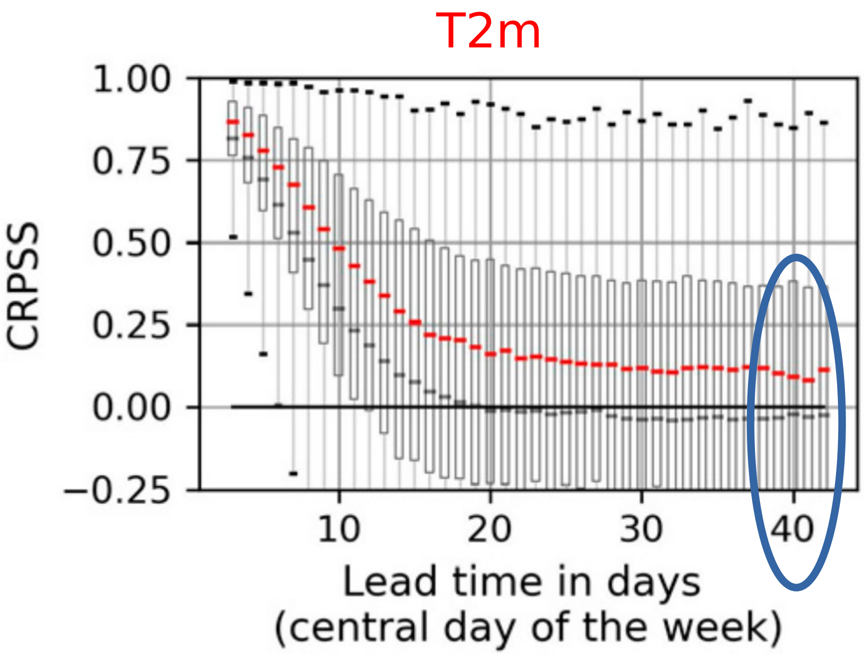
# I. Are the dynamical European sub-seasonal predictions of wind speed and temperature more skillful than baseline climatology?

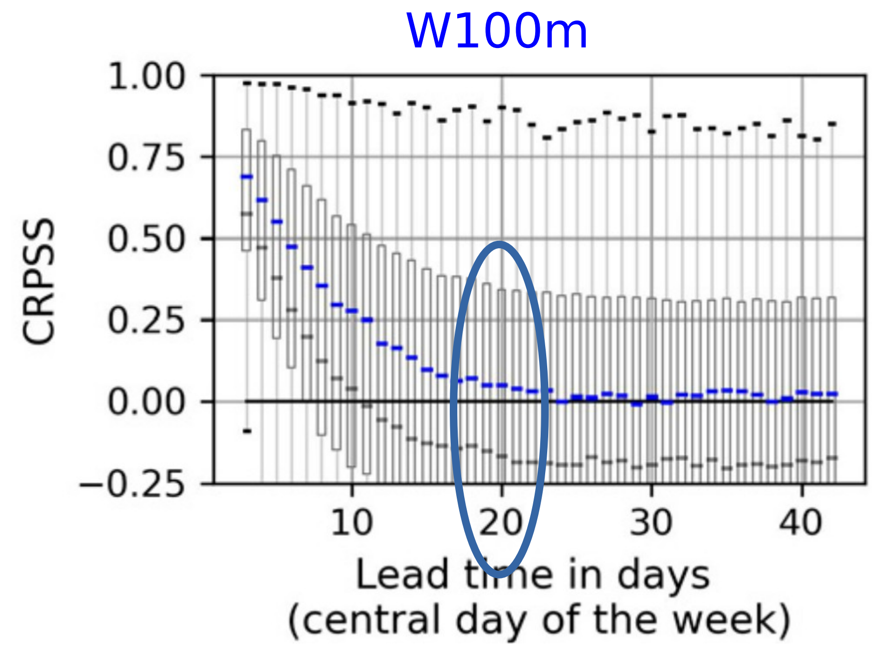
Goutham *et al*., 2022. How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature? *Mon Wea Rev* 150, 1621–1637.
https://doi.org/10.1175/MWR-D-21-0207.1

# I. Methodology: computing scores w.r.t. reanalysis



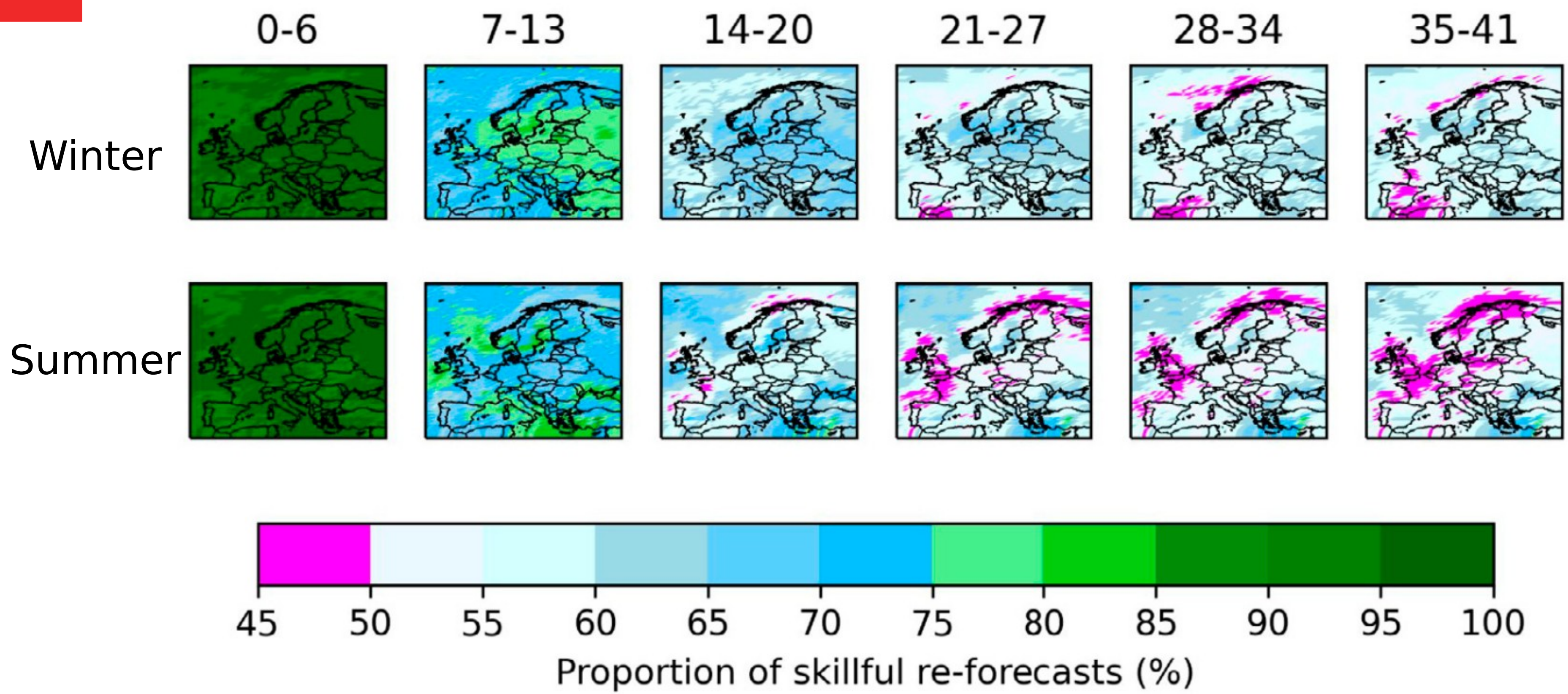
# I. Results: Europe average CRPSS




- **T2m:** significant skills up to **6 weeks** (p < 5%) on average
- **W100m:** significant skills up to **3 weeks** on average
- **T2m** skills generally **better than W100m**
- Also true for CRPSp

# I. Results: map of CRPSp for temperature




- **Skill** at the scale of **grid-point** (0.9°x0.9°) [more blue/green]
- **Better** skill over **eastern Europe** [more blue/green east]
- **Better** skill in **Winter than** in **Summer** [more blue in top]

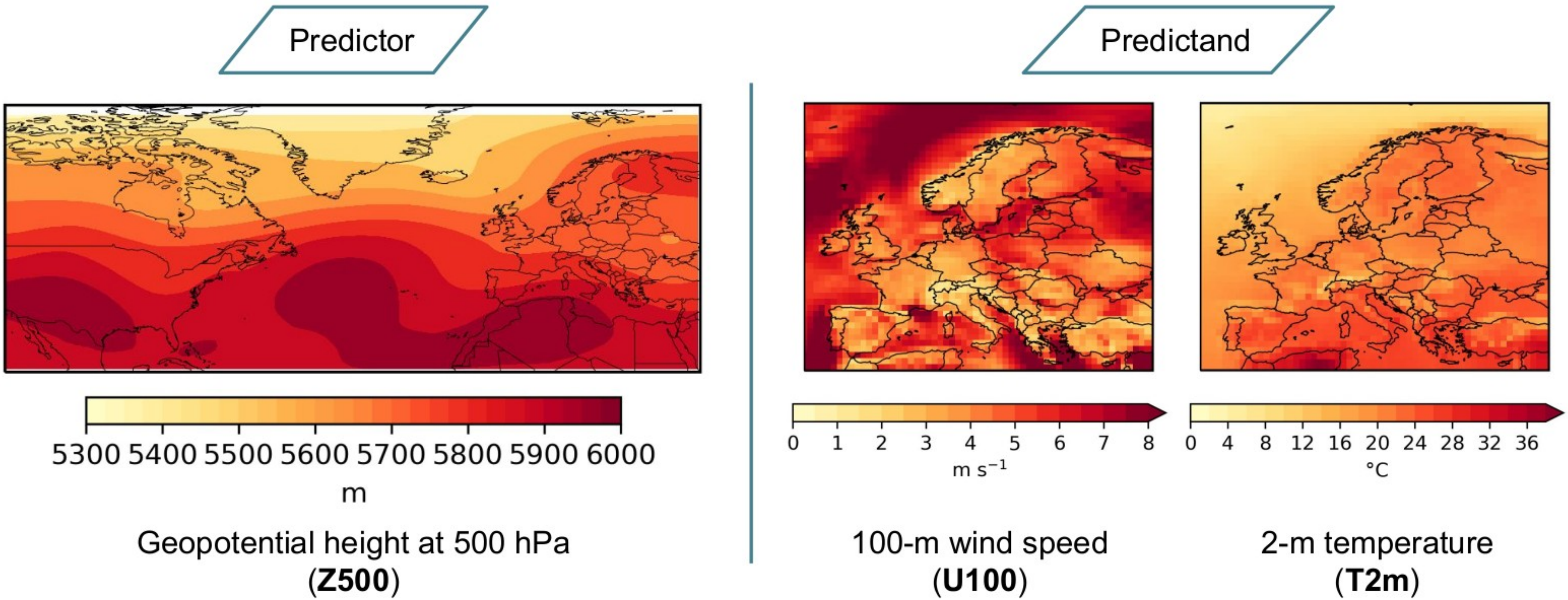
# I. Conclusions

- ➔IFS forecast skill can go up to 6 weeks encouraging applications
- ➔ Skill of 2m-temperature > 100m-wind speed
- ➔ Seasonal variations in relative skills:
  - ➔2m-temperature: winter > summer > spring/autumn
  - ➔100m-wind speed: winter > summer/autumn > spring
- ➔ Spatial pattern in skills:
  - ➔2m-temperature: eastern Europe > western Europe
  - ➔100m-wind speed: northern Europe > southern Europe
- ➔Forecasts > reforecasts because of larger ensemble
- ➔Skills beyond 2 weeks for lower and upper terciles
- ➔Limitation: forecast-reanalysis probabilistic dependence

# II. Can these predictions be improved by using machine learning to combine direct forecasts with information from other fields?

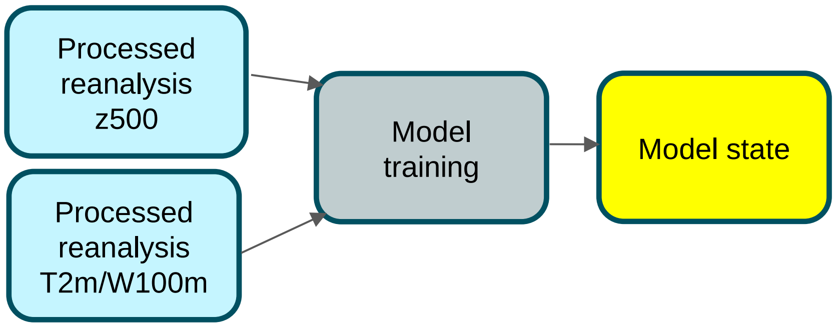
Goutham *et al.*, 2023. Statistical Downscaling to Improve the Subseasonal Predictions of Energy-Relevant Surface Variables. *Mon Wea Rev* 151, 275–296. https://doi.org/10.1175/MWR-D-22-0170.1

# II. Methodology: downscaling forecast information

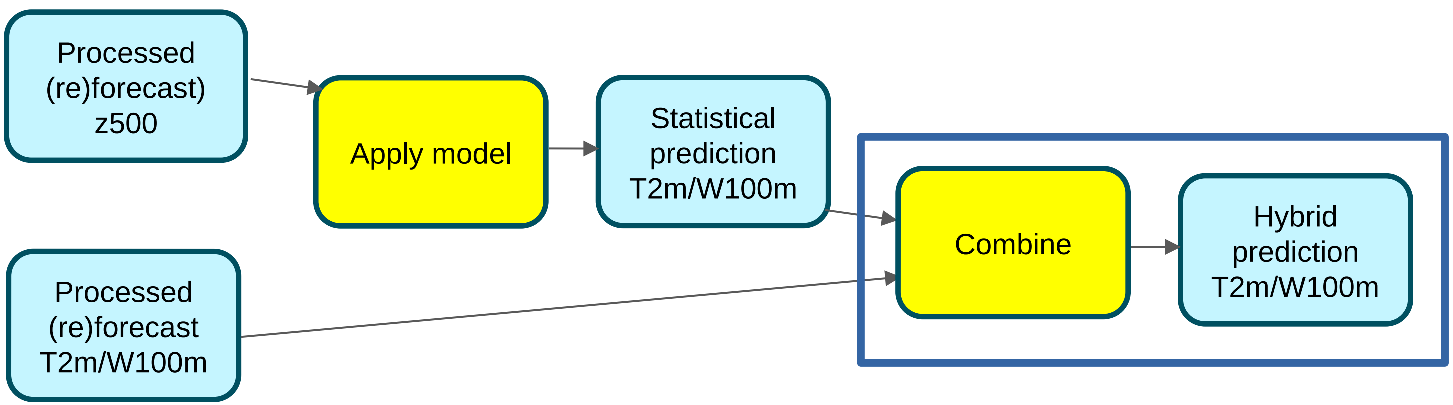


# II. Methodology: downscaling forecast information

- Training: learn relationship between z500 and W100m




- Statistical prediction: apply model to z500 and combine



# II. Methodology: redundancy analysis

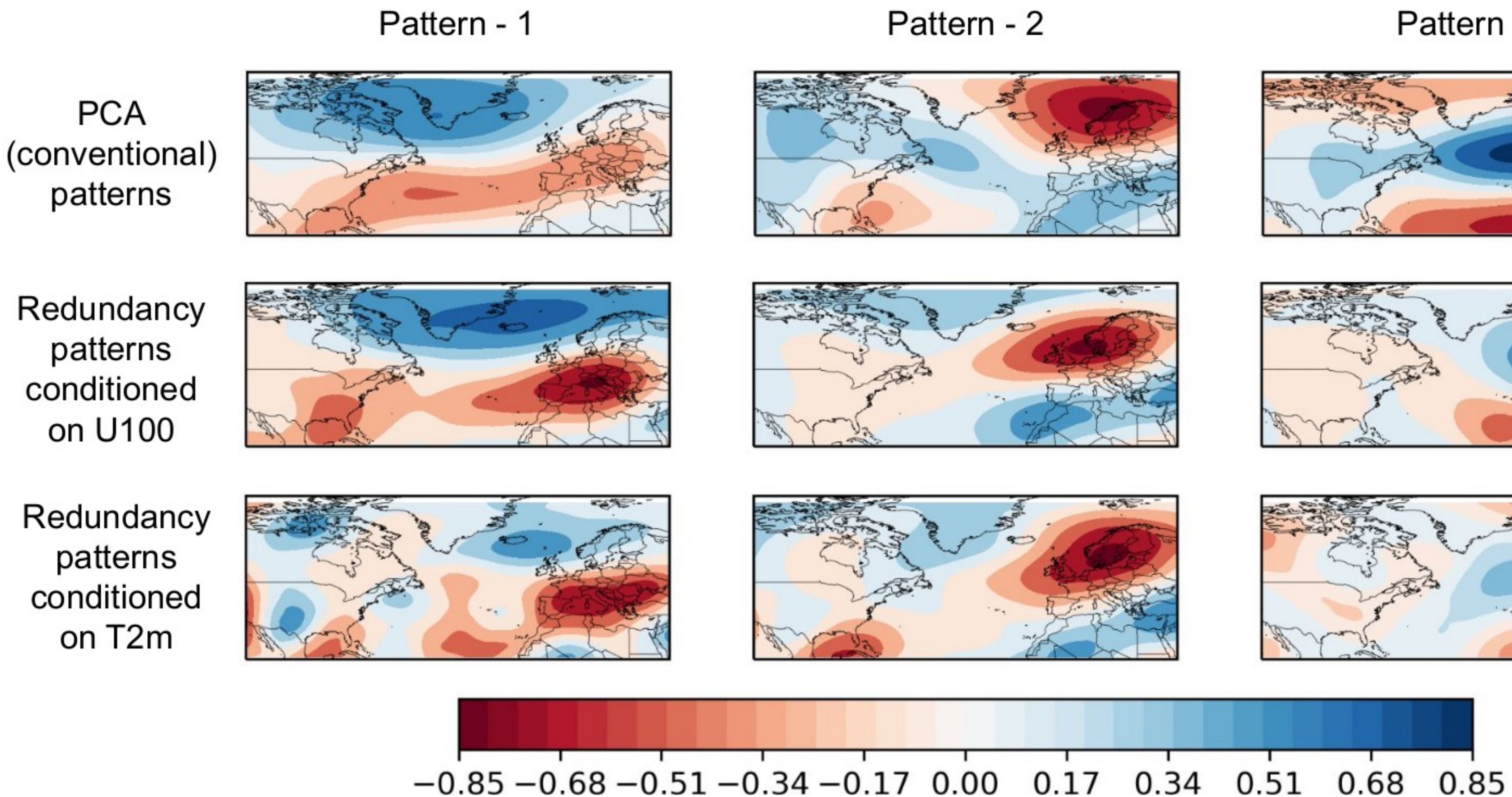
- Multi-output least squares on a linearly truncated basis
- Truncate to maximize average coefficient of determination
- Like canonical component analysis but different normalization

| | |
|---|---|
| **Predictor** | Z500 |
| **Predictand** | U100 and T2m |
| **Training period** | ERA5 reanalysis, DJF 1999-2016 (17 years) |
| **Testing period** | Operational forecasts, DJF 2016-2020 (4 years) |
| **Testing kind** | Leave-one-out cross validation |
| **Truncation criteria** | Maximizing the median of Fair-CRPSS (Ferro 2014) over the domain. |

Here, **fair** = infinite ensemble size

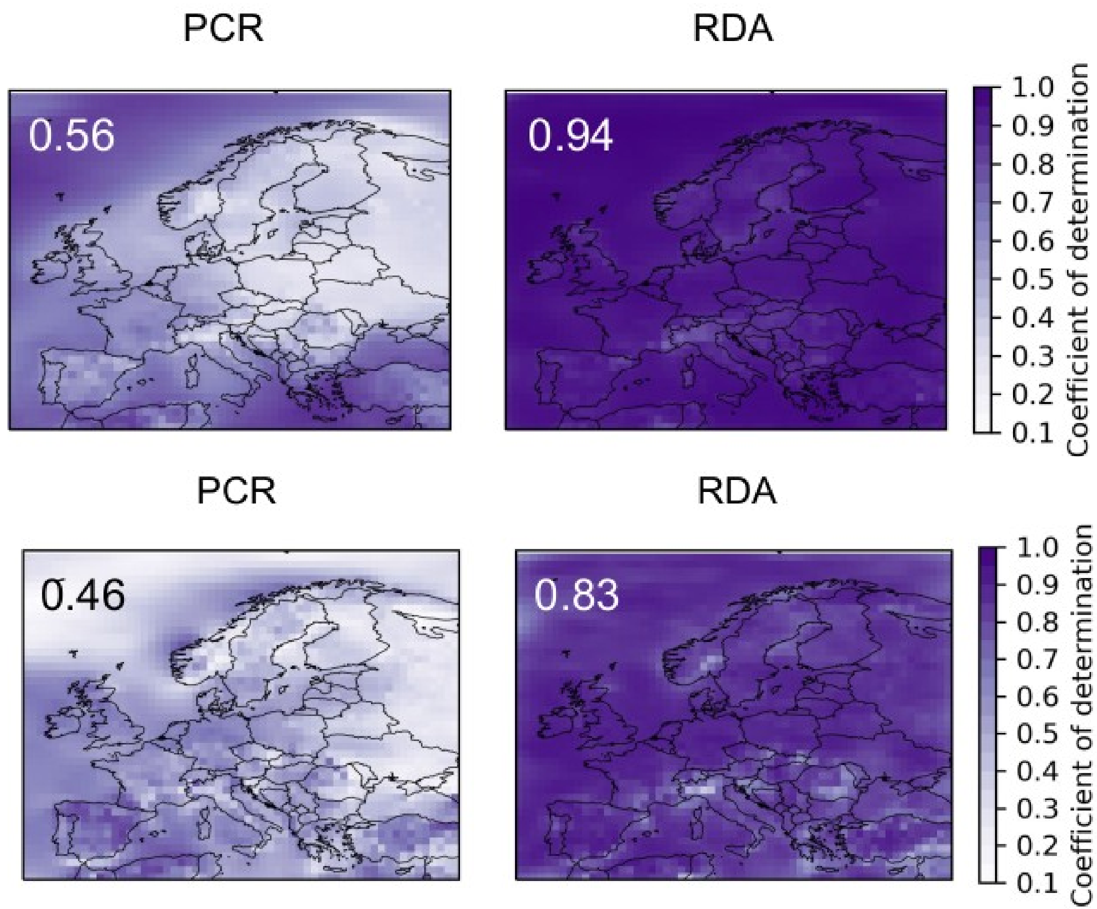
E.g. von Storch & Zwiers (1999)

# II. Methodology: comparing PCA and RDA patterns



# II. Methodology: comparing skills of PCA / RDA regression

- T2m
- W100m




→ RDA more skillfull than PCA

# II. Results: map of CRPSp for W100m

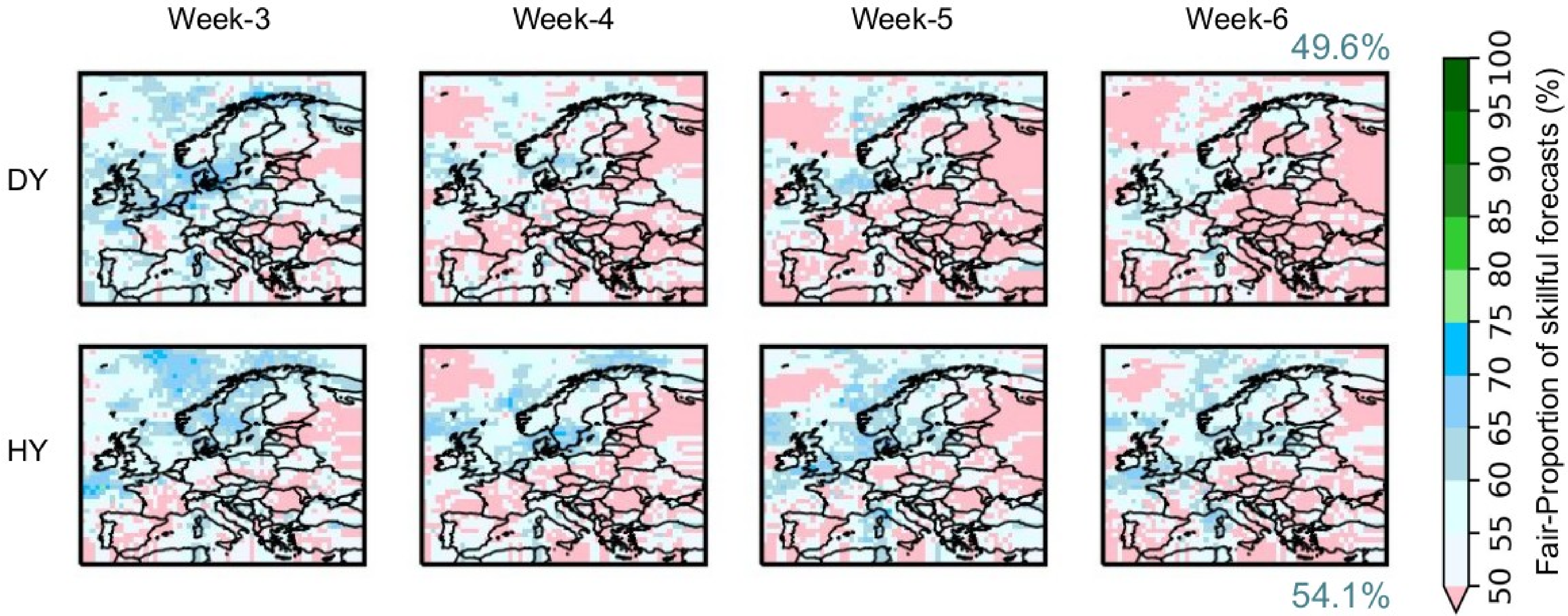



*Fig: Comparison of proportion of skillful forecasts between dynamical and hybrid predictions of U100. Forecasts in DJF 2016-2020.*

→ **Hybrid** > Dynamical > Statistical [more blue in HY than DY], especially at long lead time.

## II. Conclusions for both W100m and T2m

➔RDA patterns show a significantly higher explanatory power than the PCA counterparts

➔Hybrid predictions are significantly more skillful than either dynamical or statistical predictions alone.

➔The added value of hybrid predictions increases with lead time.

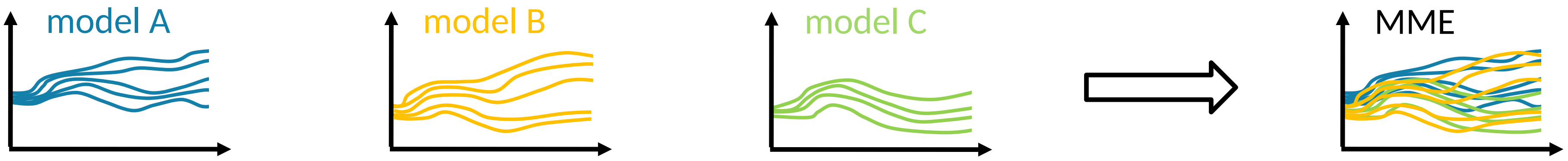
# III. Is the relationship between these fields linear at these horizons?

# IV. How to efficiently combine ensemble sub-seasonal forecasts from different NWP models?

# IV. Motivation of Multi-Model Ensembles (MME)

MME methods have been shown to improve forecast skill

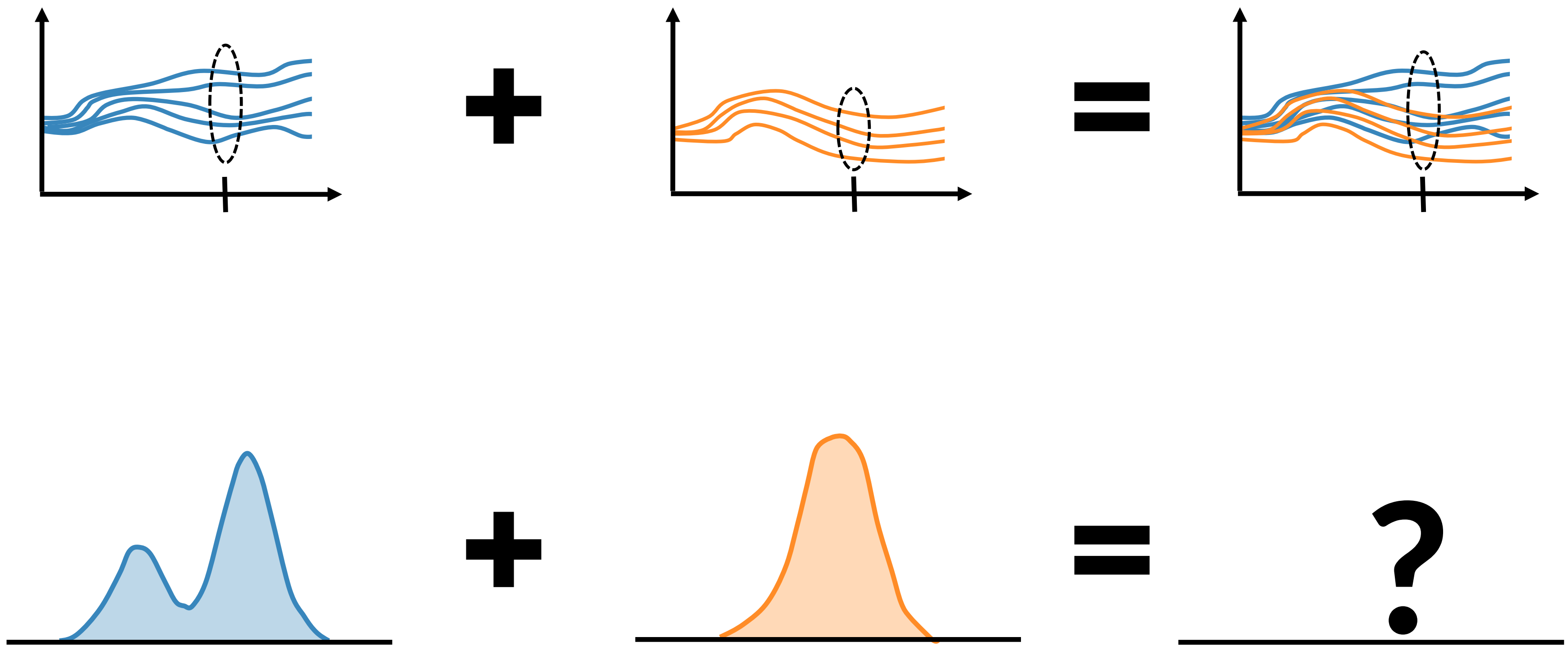
- Complementary skill
- Better estimate the forecast uncertainty
  - Larger ensemble
  - Take into account model uncertainties




- Hagedorn, R., Doblas-Reyes, F.J. and Palmer, T.N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – I. Basic concept. Tellus A: Dynamic Meteorology and Oceanography, 57(3), p.219–233.
- Casanova, S., and B. Ahrens, 2009: On the Weighting of Multimodel Ensembles in Seasonal and Short-Range Weather Forecasting. Mon. Wea. Rev., 137, 3811–3822,

# IV. Opportunities from probabilist framework

- Ensemble forecast = discrete probability distribution



# IV. Methodology: barycenter of distributions

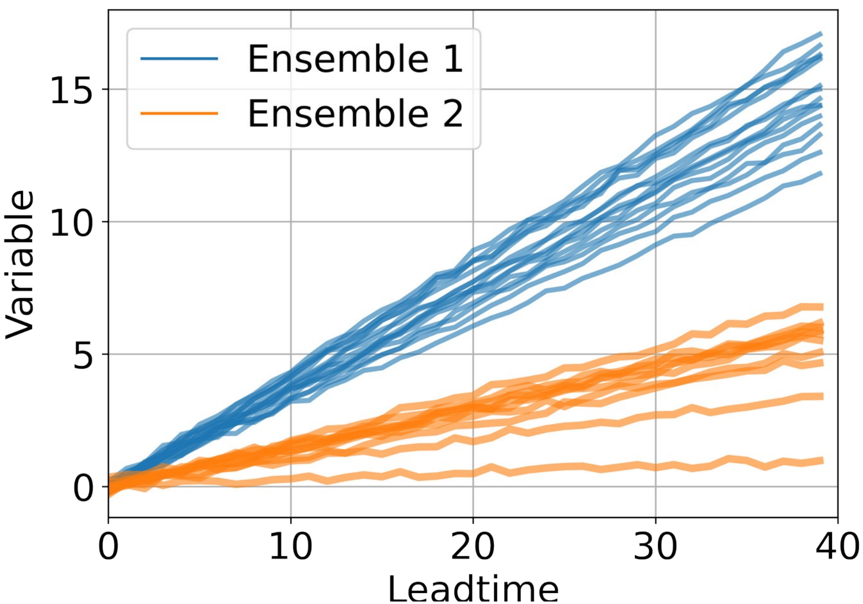
What is the barycenter of two distributions $\mu_1$ and $\mu_2$ ?

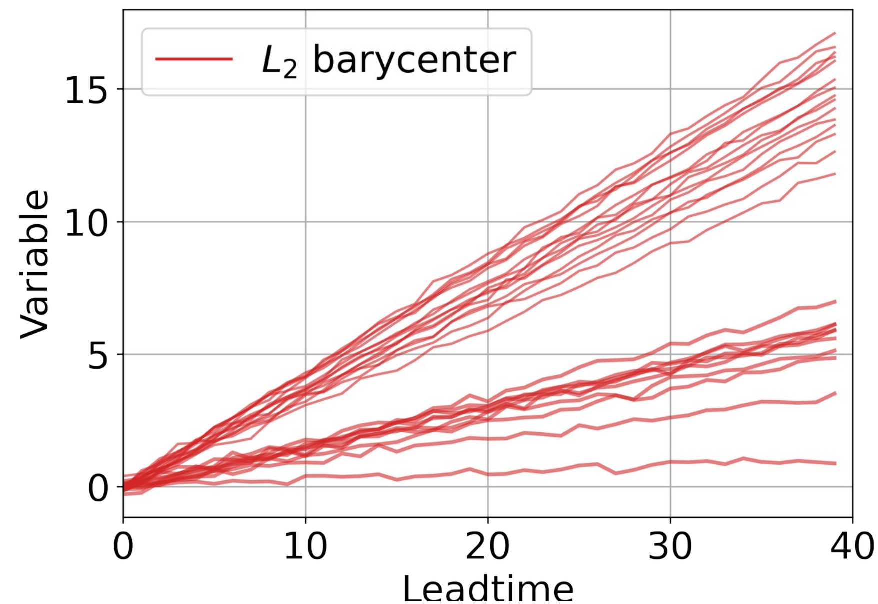
$$\mu_b = argmin[d(\mu, \mu_1)^2 + d(\mu, \mu_2)^2]$$

where $d$ is a distance.

# IV. Methodology: pooling barycenter

L2 distance: $$d(\mu_1, \mu_2) = \|\mu_1 - \mu_2\| = \left(\lim (\mu_1(x) - \mu_2(x))^2 dx\right)^{1/2}$$



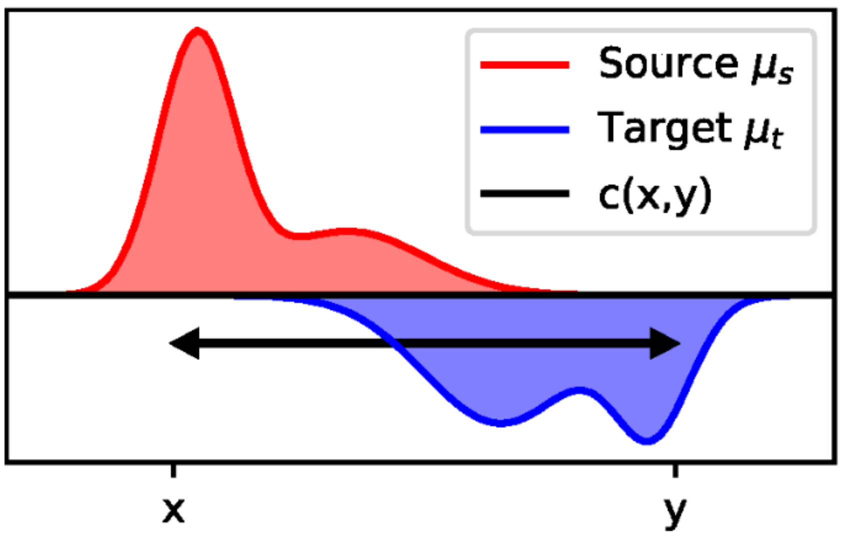



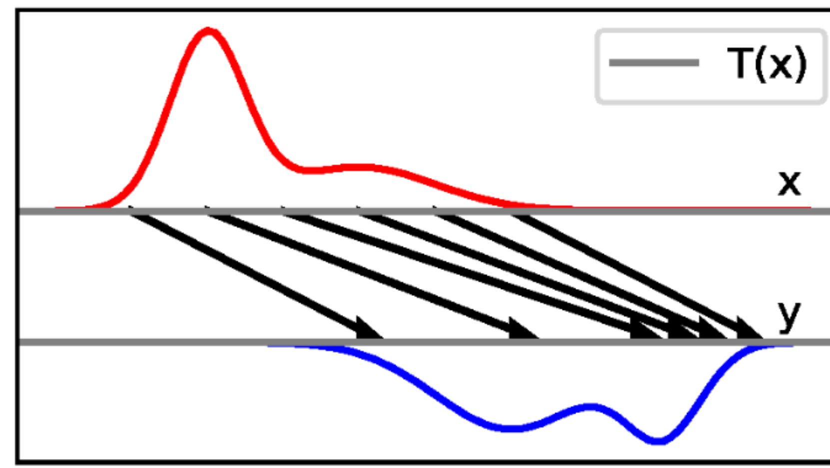
# IV. Methodology: Wasserstein barycenter

- Wasserstein distance: minimum average transport cost

$$d(\mu_1, \mu_2) = W_2(\mu_1, \mu_2)$$

Optimal transport (horizontally):

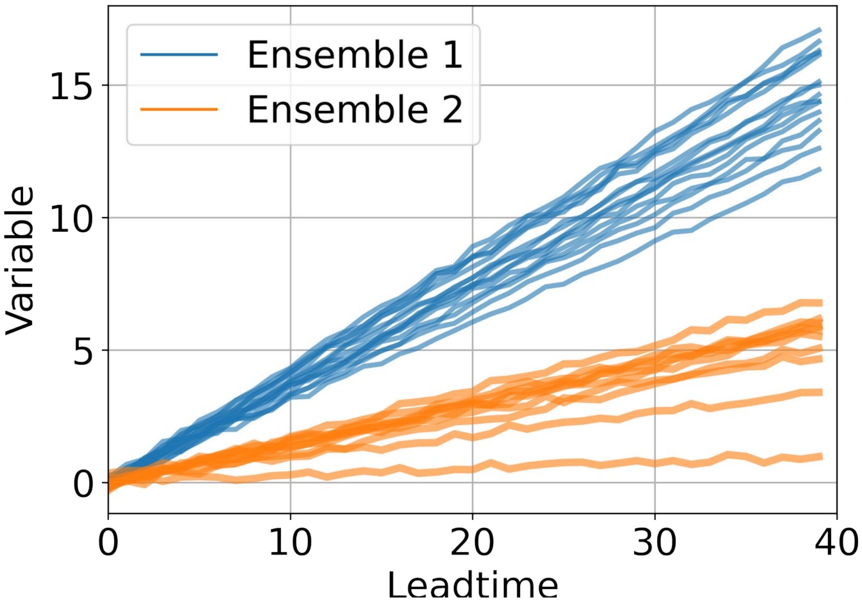


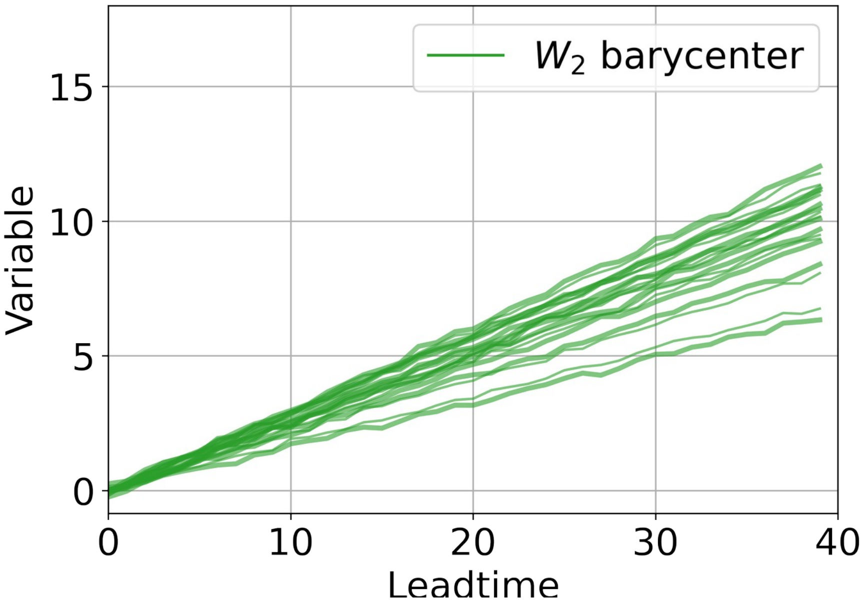




Source: R. Flamary lectures

# IV. Methodology: Wasserstein barycenter

➔ Wasserstein barycenter: $\mu_{W_2} = argmin\left(W_2^2(\mu, \mu_1) + W_2^2(\mu, \mu_2)\right)$



# IV. Methodology: weighted barycenter with machine-learned weights

What is the barycentre of two distributions $\mu_1$ and $\mu_2$ ?

$$\mu_b = argmin[\alpha . d(\mu, \mu_1)^2 + (1-\alpha) . d(\mu, \mu_2)^2]$$

where

- $d$ is a distance
- $0 \leq \alpha \leq 1$ is a constant weight

Learn weight(s) from a training dataset

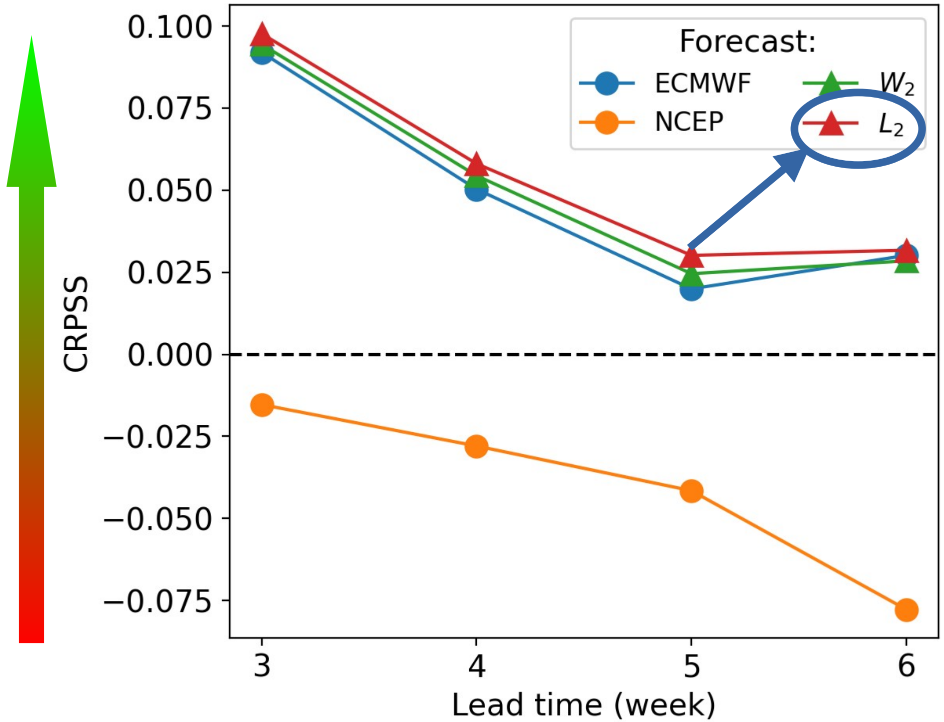
# IV. Methodology: combining ECMWF IFS and NCEP

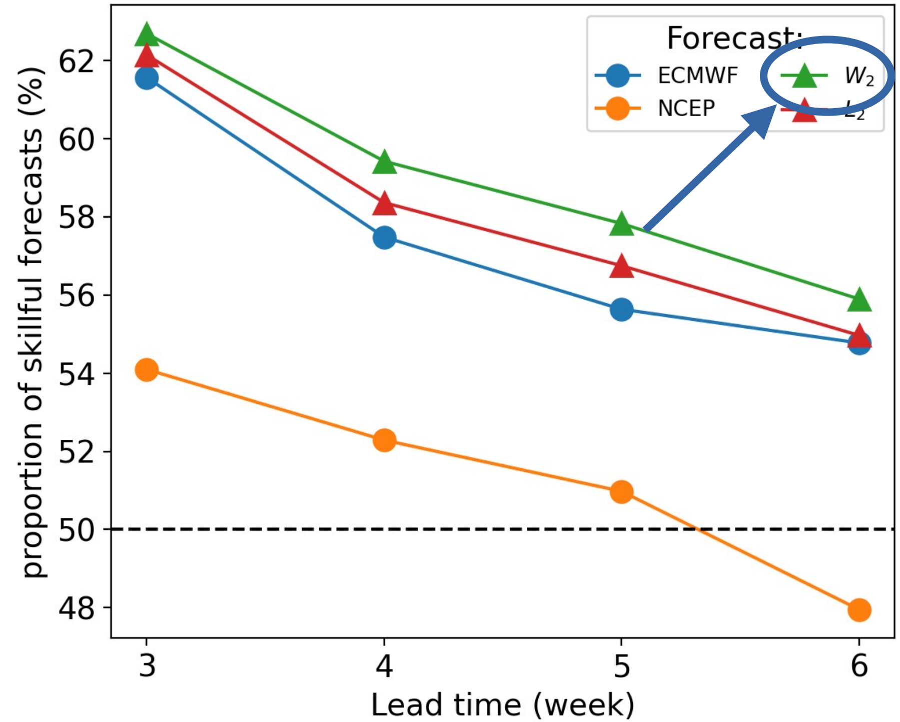
- The S2S database of sub-seasonal ensemble forecast from 11 centers

| | ECMWF | NCEP |
|---|---|---|
| Ensemble size | 50 | 15 |
| Time range | Days 0-46 | Days 0-44 |
| Frequency | 2/weeks | Daily |

- Reference: MERRA-2 (reanalysis)

# IV. Results: spatial average



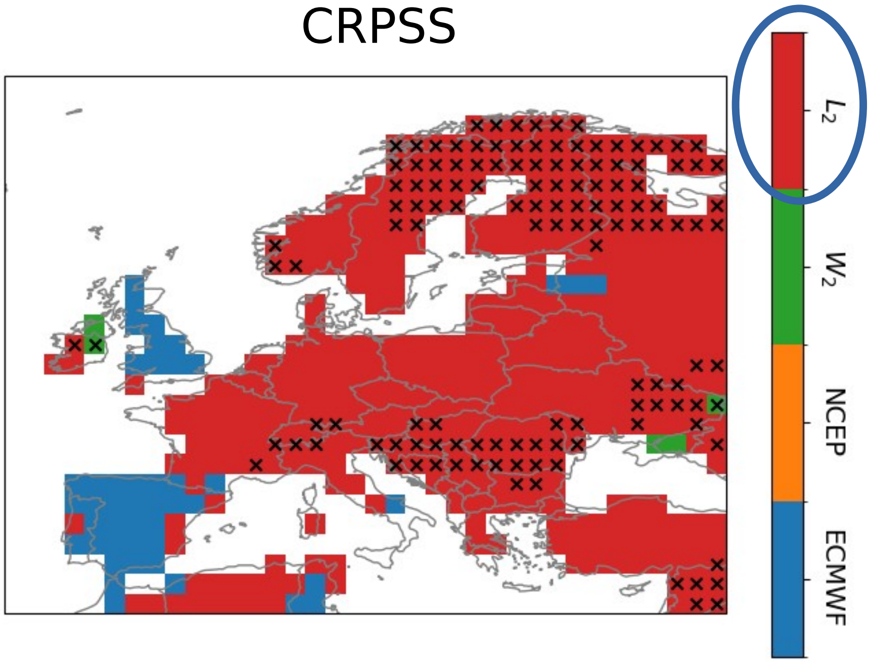




- $L_2$-barycenter best w.r.t. CRPSS
- $W_2$-barycenter best w.r.t. proportion of skillful forecasts

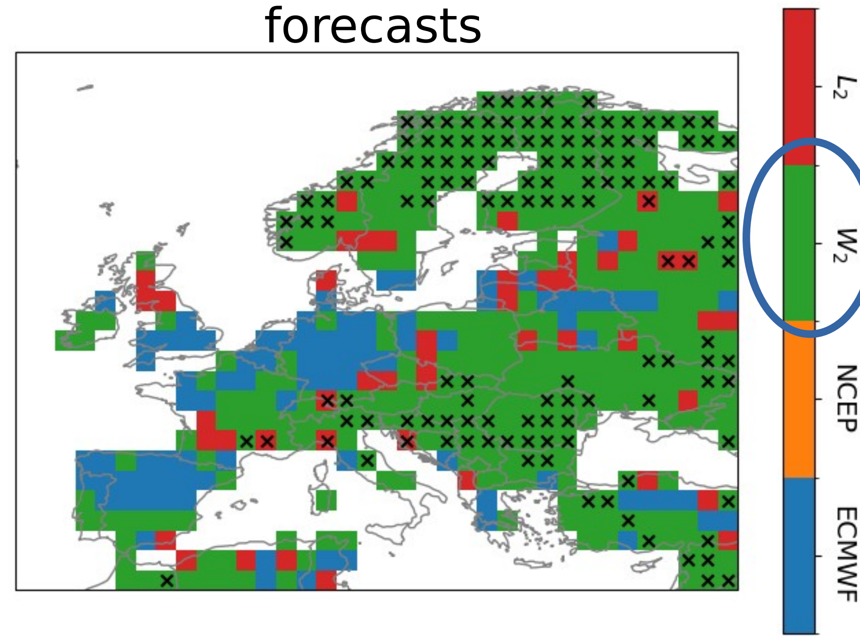
➔ Best model depends on the score, but improvement w.r.t. to single model

# IV. Results: grid-point scale

CRPSS




Proportion of skillful forecasts




➔ Best model depends on the score

➔ Best model also depends on the locations

# IV. Conclusions

➔ Multi-model ensemble generally improve on single-model forecast.

➔ Best combination method depends on the score (and location).

➔ The model's weight $\alpha$ is critical (large impact on the scores).

Next steps:

➔ Application of the barycenters to several models

➔ Temporal weights

➔ Barycenters to combine inputs/outputs in downscaling approach

# General perspectives for discussion

- **Sub-seasonal predictability potential confirmed** for presented observables and domain
- Many factors entering predictions (preprocessing, downscaling, input variables, training period…) depending on scores → **can we do better than trial-and-error?**
- **Compatibility issues between parameters of different model simulations** (different ensemble sizes in particular)
- **Relevance of skill improvement depend on application** (e.g. application to wind droughts in N. Goutham PhD thesis)
- What is the extra economic/ecological cost of machine learning?

  → **Balance between dynamics and statistics in a sufficient forecasting system ?**

# Thank you!

Contact: alexis.tantet@lmd.ipsl.fr

# Selected references

➔ Alonzo *et al.*, 2017. Modelling the variability of the wind energy resource on monthly and seasonal timescales. *Renew Energy* 113, 1434–1446. https://doi.org/10/gf3wf4

➔ Goutham *et al.*, 2022. How Skillful Are the European Subseasonal Predictions of Wind Speed and Surface Temperature? *Mon Wea Rev* 150, 1621–1637. https://doi.org/10.1175/MWR-D-21-0207.1

➔ Goutham *et al.*, 2023. Statistical Downscaling to Improve the Subseasonal Predictions of Energy-Relevant Surface Variables. *Mon Wea Rev* 151, 275–296. https://doi.org/10.1175/MWR-D-22-0170.1

➔ Höhlein *et al.*, 2020. A comparative study of convolutional neural network models for wind field downscaling. *Meteorol Appl* 27, e1961. https://doi.org/10.1002/met.1961

➔ Santambrogio, 2015. *Optimal Transport for Applied Mathematicians*. Birkhäuser, Cham.

➔ Vissio, G., Lembo, V., Lucarini, V., Ghil, M., 2020. Evaluating the Performance of Climate Models Based on Wasserstein Distance. *Geophys Res Lett* 47, e2020GL089385. https://doi.org/10/ghn76n

➔ Vitart, F., Robertson, A.W., 2018. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. NPJ Climate and Atmospheric Science. https://doi.org/10/gfsv2d

➔ Chap. 14 in von Storch, H., Zwiers, F.W., 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge.

➔ White *et al.*, 2017. Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol Appl* 24, 315–325. https://doi.org/10.1002/met.1654

# Appendix

# Redundancy analysis mathematical formulation

Von Storch et al. 1999, Tippett et al. 2008, Wilks 2014,2019

**1. Redundancy analysis**: maximize the explained variance

$$\mathrm{P} = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,t} \\ p_{2,1} & . & \cdots & p_{2,t} \\ . & . & \cdots & . \\ p_{m,1} & . & \cdots & p_{m,t} \end{pmatrix} \qquad \mathrm{Q} = \begin{pmatrix} q_{1,1} & q_{1,2} & \cdots & q_{1,t} \\ q_{2,1} & . & \cdots & q_{2,t} \\ . & . & \cdots & . \\ q_{n,1} & . & \cdots & q_{n,t} \end{pmatrix}$$

$$\boxed{\mathrm{X} = \mathrm{E}_{\mathrm{P}}^{\mathrm{T}}.\mathrm{P}' \quad \text{and} \quad \mathrm{Y} = \mathrm{E}_{\mathrm{Q}}^{\mathrm{T}}.\mathrm{Q}'}$$

The joint sample variance-covariance matrix of the leading predictor and predictand PCs is given by

$$\boxed{\mathrm{S} = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{YX} & S_{YY} \end{pmatrix}}$$

The eigen decomposition of the predictand covariance matrix conditioned on the predictor

$$\boxed{\mathrm{S}_{\hat{\mathrm{Y}}\hat{\mathrm{Y}}} = \mathrm{S}_{\mathrm{YX}}.\mathrm{S}_{\mathrm{XX}}^{-1}.\mathrm{S}_{\mathrm{XY}}}$$

yields predictand patterns B and eigen values $\Lambda$

The predictor patterns can be obtained as

$$\boxed{\mathrm{A} = \frac{1}{\sqrt{\Lambda}}\mathrm{S}_{\mathrm{XX}}^{-1}.\mathrm{S}_{\mathrm{XY}}.\mathrm{B}}$$

The predictor and predictand redundancy PCs can be computed as

$$\mathrm{V} = \mathrm{A}^{\mathrm{T}}.\mathrm{X} \quad \text{and} \quad \mathrm{W} = \mathrm{B}^{\mathrm{T}}.\mathrm{Y}$$

respectively.

**2. Redundancy predictions**:

Given a new predictor PC set $X_o$, one can compute the predictand redundancy PCs as

$$\boxed{\hat{\mathrm{W}} = \mathrm{R}^{\mathrm{T}}.\mathrm{V}_{\mathrm{o}} = \mathrm{R}^{\mathrm{T}}.\mathrm{A}^{\mathrm{T}}.\mathrm{X}_{\mathrm{o}}}$$

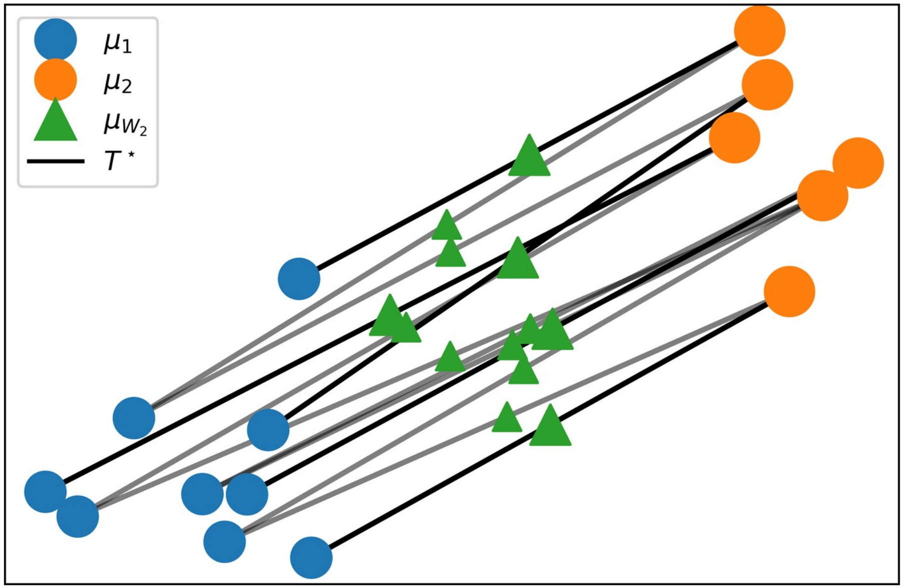
Where R = $\sqrt{\Lambda}$

The predictand vector can be reconstructed as

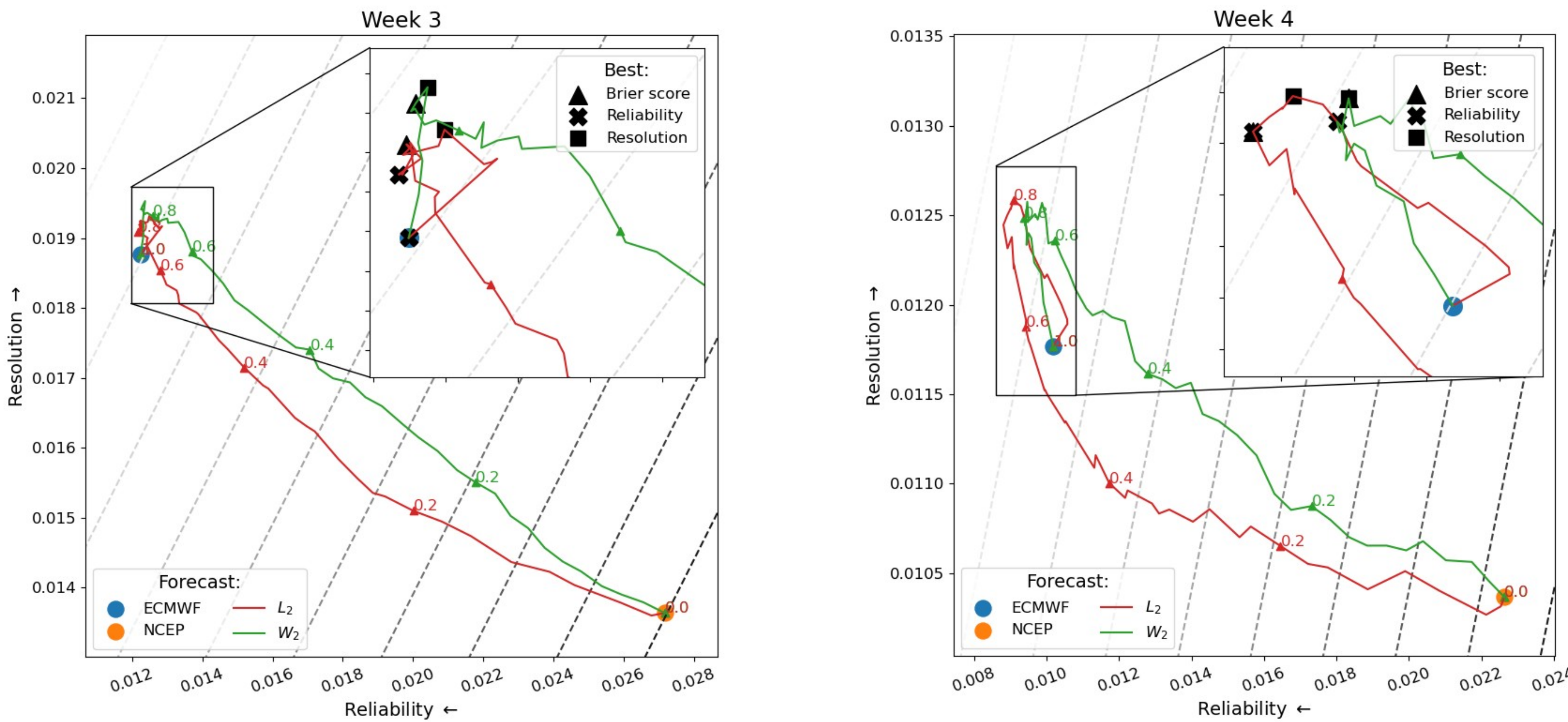
$$\boxed{\hat{\mathrm{Q}}_{\mathrm{o}}{}' = \mathrm{E}_{\mathrm{Q}}.(\mathrm{B}^{\mathrm{T}})^{-1}.\hat{\mathrm{W}} = \mathrm{E}_{\mathrm{Q}}.(\mathrm{B}^{\mathrm{T}})^{-1}.\mathrm{R}^{\mathrm{T}}.\mathrm{A}^{\mathrm{T}}.\mathrm{X}_{\mathrm{o}}}$$

# Barycentre with the Wasserstein distance

➔ Wasserstein barycentre: $\mu_{W_2} = argmin\left(W_2^2(\mu, \mu_1) + W_2^2(\mu, \mu_2)\right)$



# Results: Brier score decomposition for first tercile




- $L_2$-barycentre has a better reliability.
- $W_2$-barycentre has a better resolution.