Modelling the distributions of surface atmospheric variables and a parametric bias correction approach

Users of ECMWF's forecasts (UEF2026),
1-4 June, 2026, ECMWF, Reading

Mahmud Hasan Ghani, Francesco Trotta, Enrico Scoccimarro and Nadia Pinardi

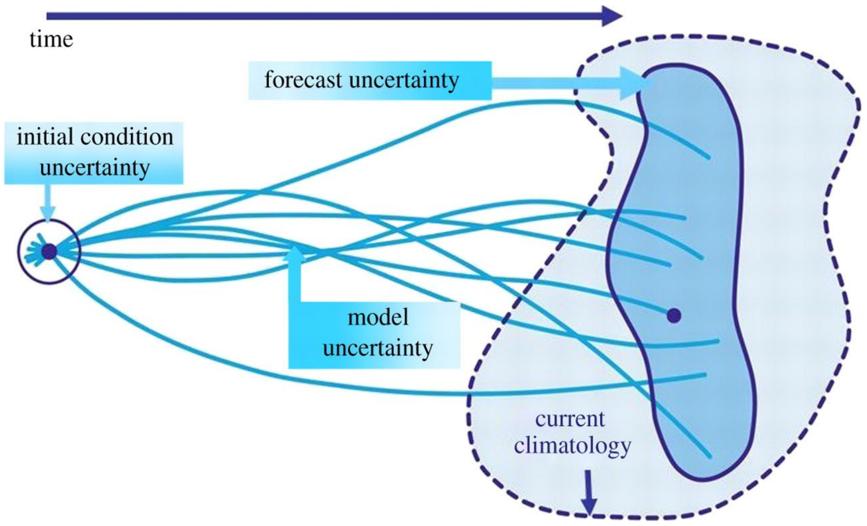
Department of Physics and Astronomy, University of Bologna
CMCC Foundation - Euro-Mediterranean Center on Climate Change

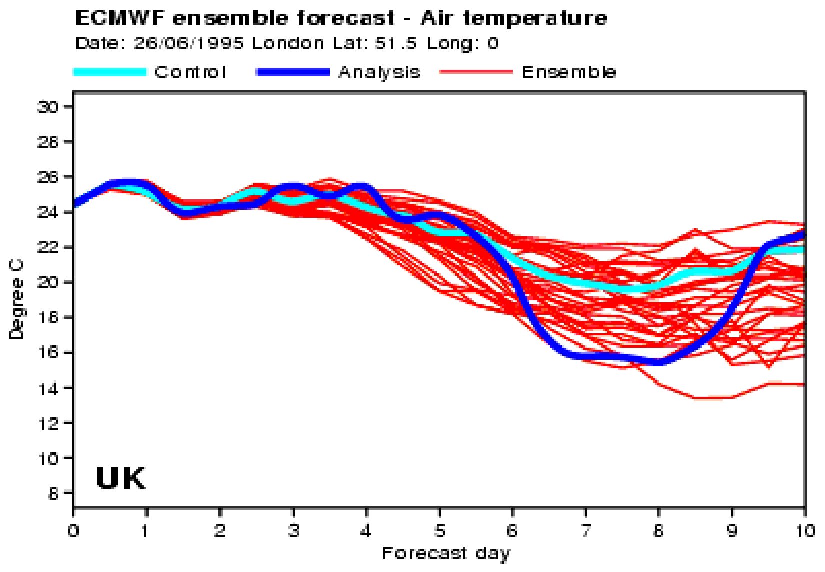
ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici

Uncertainty in numerical model predictions

- Addressing uncertainty is a fundamental approach in improving ocean or atmospheric forecast
- With the non-linear state of atmosphere, the surface variables are directly linked to ocean-atmosphere dynamics and play influential role in an ocean forecasting system
- For the quantification of uncertainties, perturbation is applied with different forcing components- initial conditions of oceanic fields, bathymetry and atmospheric fields
- Uncertainty in the atmospheric forcing can be quantified by the study of the statistical distributions of the atmospheric variables and their moments





Slingo and Palmer, 2011

Statistical post-processing approach

- Refining and improving skills of numerical weather prediction heavily rely on statistical methods (Hamil, 2011; Hemri et al., 2014; Vannitsem et al., 2021; Zanetta et al., 2022)
- Statistical correction technique is the fundamental approach in post processing to reduce forecast error using statistics which an integral part of operational forecast (Vannitsem et al., 2021)

Post-processing methods

Parametric/distributional approach

- Atmospehric bias correction(climate forecast) mostly rely on the two parameter Gaussian or Gamma distribution
- Quantile-mapping is a widely-used approach, but gaussian parametric corrections fail to mirror the asymmetry of non-stationary forecast(atmoshpheric) fields

Non-parametric approach

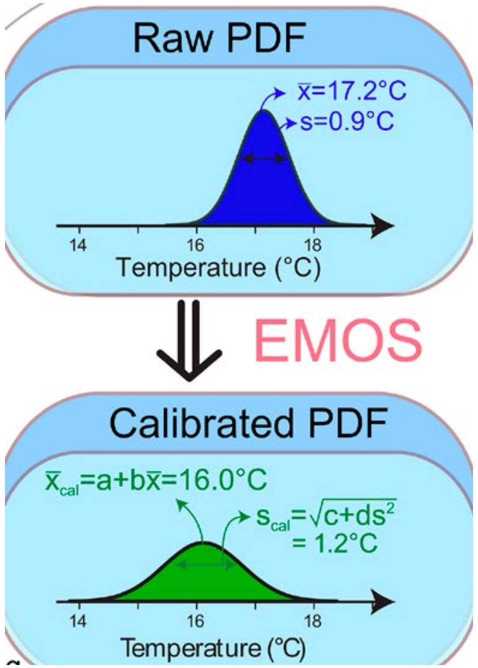
- Non parametric model without distributional assumption
- Empirical quantile mapping or CDF mapping does not depend on pre-defined distribution but uses the actual raw data points

A Parametric Skew-Normal Quantile Mapping is a specifically optimized statistical approach

Parametric approach for bias correction

- The statistical methods for bias correction are mostly related to quantile association and the widely used one is quantile mapping (QM) (Vrac 2015; Piani 2010; Wood et al 2004) and alternatively known as probablity mapping or distribution mapping
- QM is widely applied for climate model variables(i.e. air temperature, precipitation), mainly for correction of biases and statistical downscaling from coarse to finer scales (Piani et al, 2010; Themeßl, et al., 2011; Vrac 2014; Cannon et al 2015, Maciuas et al, 2016).
- Using empirical cumulative distribution function (CDF), QM is applicable to all possible meteorological variables (Themeßl, et al., (2011)

Vannitsem et al., 2021



Probability Density Function(PDF) for atm. surface variables

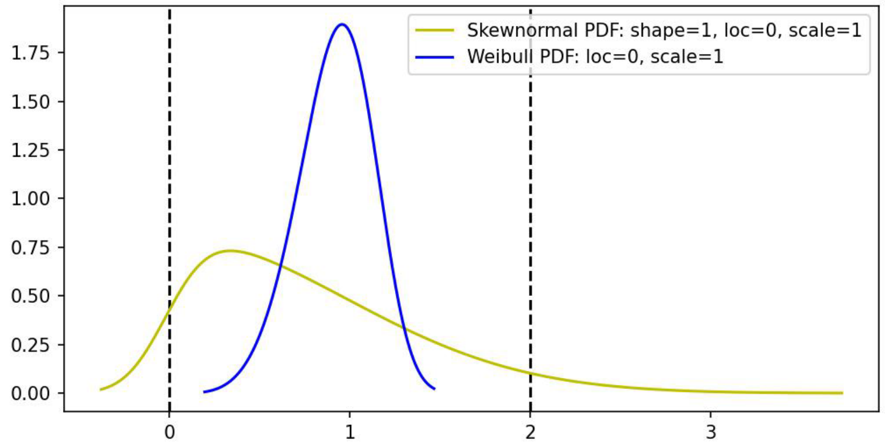
- We aim to utilize the PDF parameters derived from the 'historical' period as reference values to generate 'modelled' time series, which are then applied for a correction approach in the distributional biases of ECMWF forecast
- The parametric QM approach with the skew-normal PDF analysis of the atmospheric variables ensures the inherent asymmetry in atmospheric anomalies (Azzalini, 2005).

The PDF models:

The Weibull PDF (1) comprises shape and scale parameters, and the support is x > 0, +∞. The Skew-normal PDF (4) has shape parameter (α), location parameter (μ), and scale parameter (λ), and its support is $x \in (-\infty, +\infty)$.

$$f\ (x;\ k,\lambda) = \frac{k}{\lambda}\ \left(\frac{x}{\lambda}\right)^{k-1} exp\ ^{-(\frac{x}{\lambda})^{k}} \text{ .... (1)}$$

$$f\ (x;\alpha,\mu,\lambda) = \frac{2}{\lambda}\ \phi\left(\frac{x-\mu}{\lambda}\right)\Phi\ \left(\frac{\alpha(x-\mu)}{\lambda}\right) \text{ .... (2)}$$



Weibull vs Skew-normal PDF

Probability distributions of atmospheric variables

Climatology and seasonality adjustment

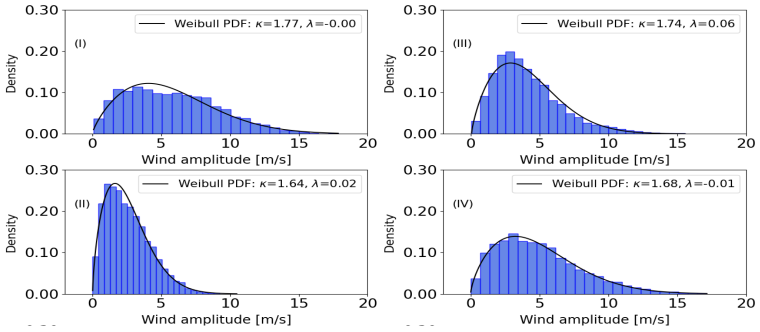
ECMWF analysis and Forecast dataset : spatial resolution 0.125-degree, daily mean resolution (converted from 6-hourly)
COSMO-CLM: (COSMO model in Climate Mode) model over the EURO-CORDEX area, 0.11 deg, time interval - 1 hr
Observed variables: D2M-2m dewpoint temperature, MSLP-Mean Sea Level , T2M-2m temperature, Wind components-U10M & V10M

Climatology of a variable X_t containing a daily time series ($x_{d1}, x_{d2}, x_{d3}, \dots\ x_{dn}$)

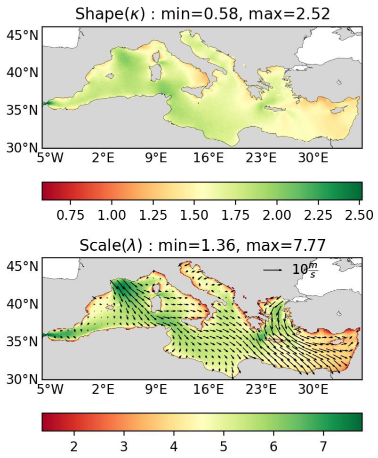
Climatology (Ct) = $\frac{1}{N}\sum_{j=1}^{N} X_{t,j}$ .... (3)

Anomlaies $(\widetilde{X})$ = $X_t - C_t$ ....(4)

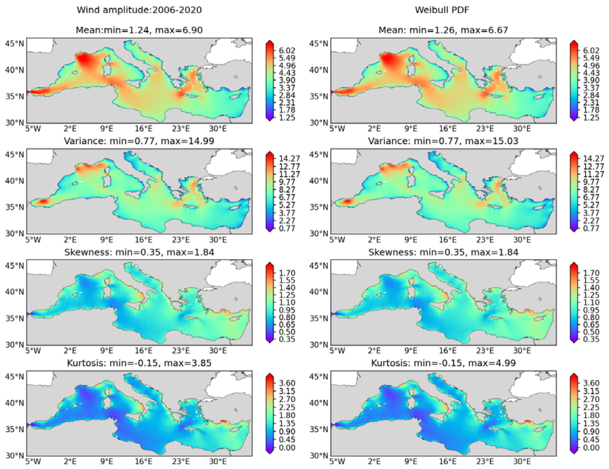
Single grid histograms with Weibull PDF fitting on Alboran Sea point (-1, 36.6), Gulf of lion point (4, 42.5), 36.5), Adriatic Sea= (16.5, 42, Ionian Sea= (17.5, 37),



Parameter distributions



Qualitative validation of the moments



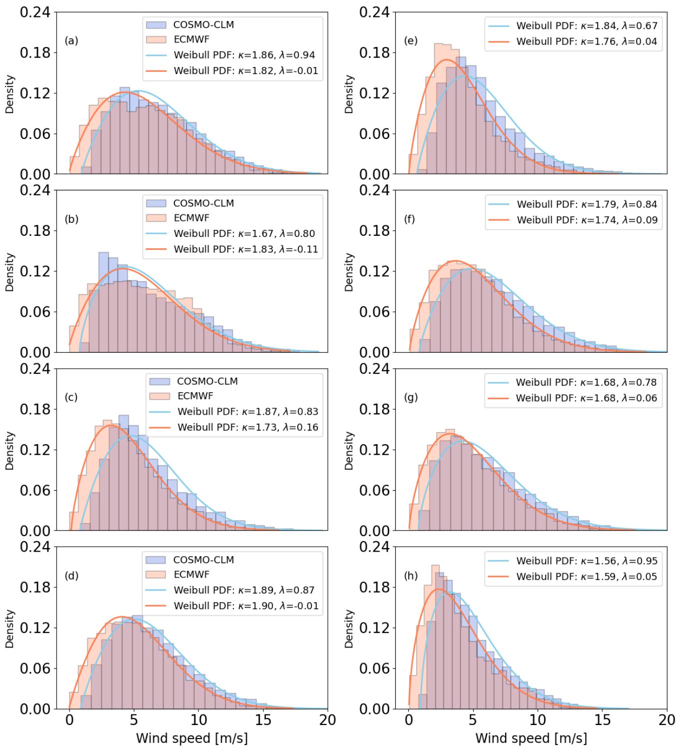
Probability distributions of atmospheric variables

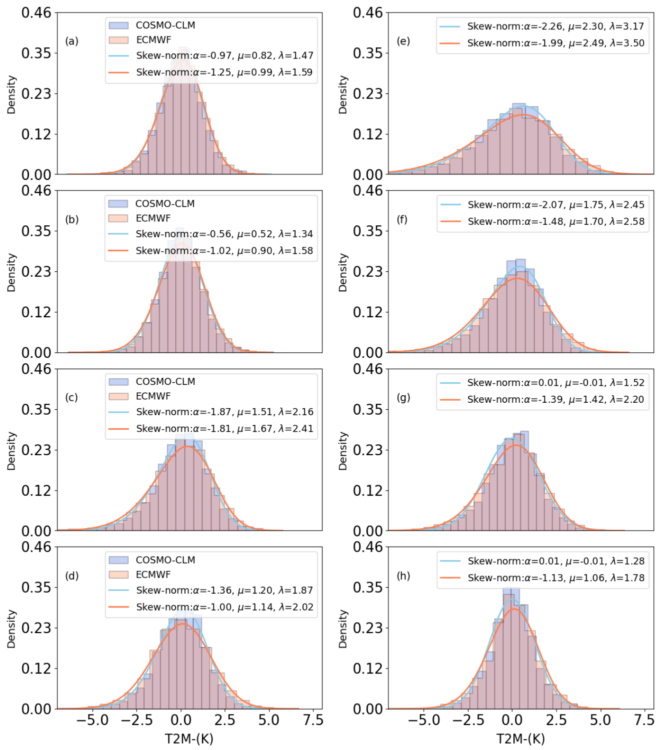
Probablity distribution for NWP and Climate model variables

$$W = \sqrt{(U - \bar{U})^2 + (V - \bar{V})^2}$$

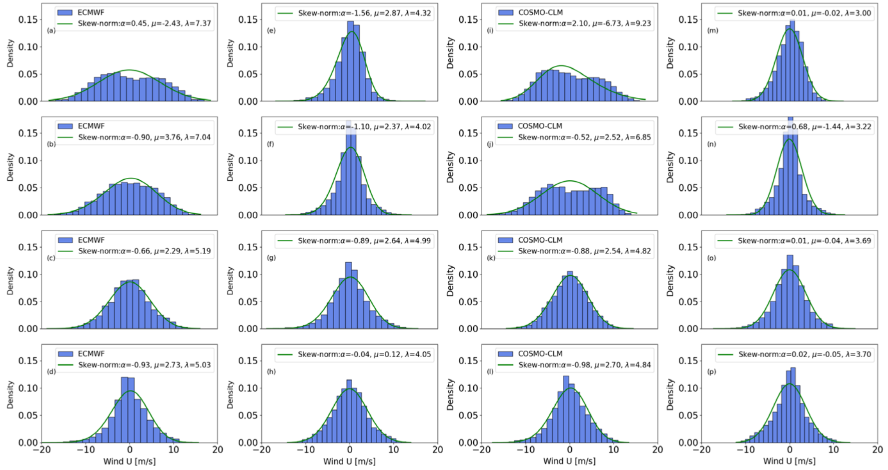
where W =wind speed anomaly, $\bar{U}$ =climatology, and U=U10m, V=V10m

Wind Speed

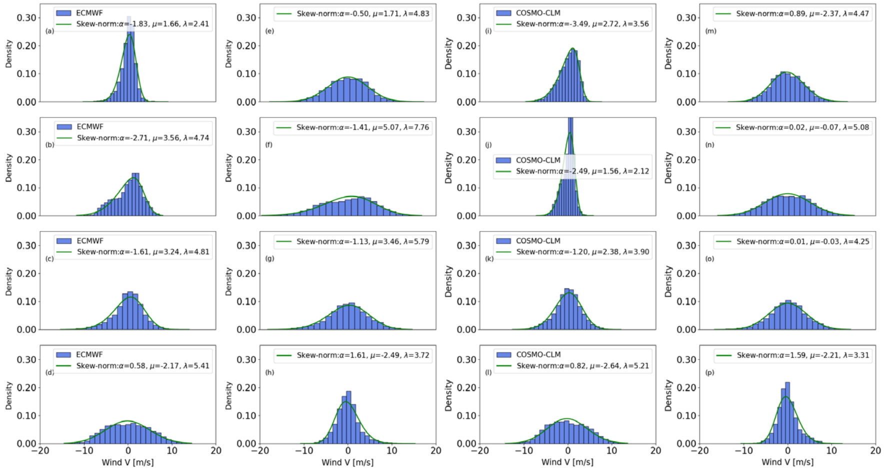




ECMWF analysis Wind U COSMO-CLM

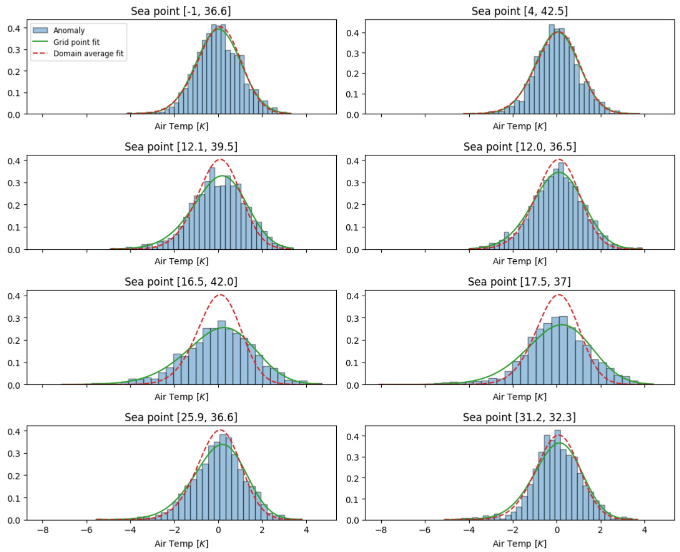


ECMWF analysis Wind V COSMO-CLM

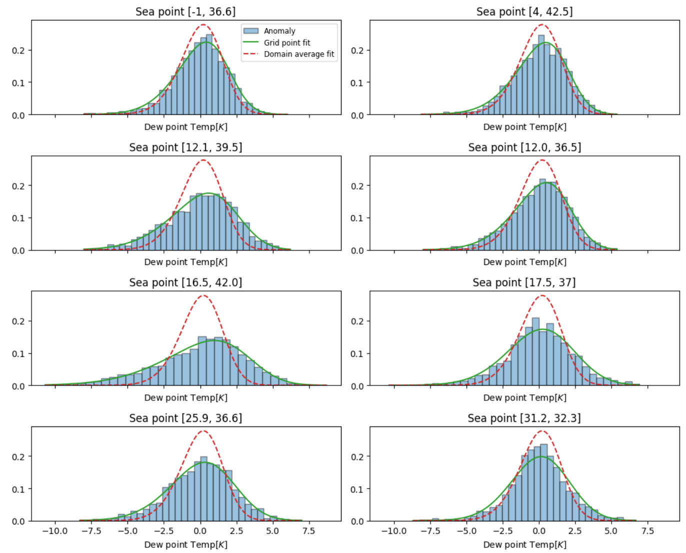


PDF fit variation: grid point and domain averaged

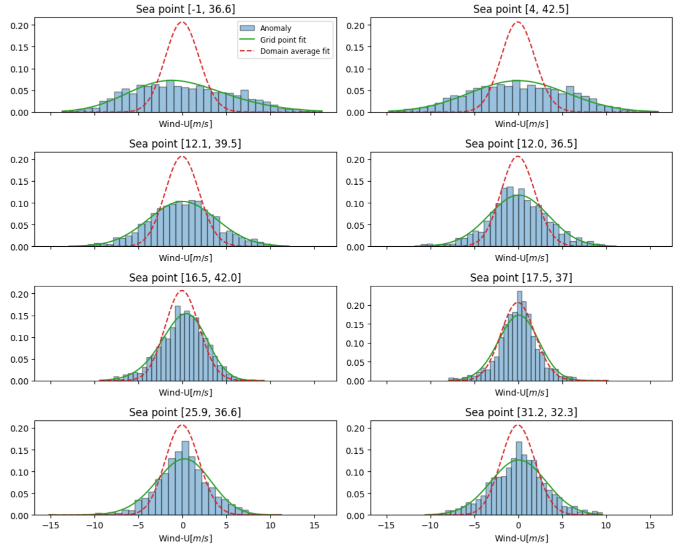
Air temperature



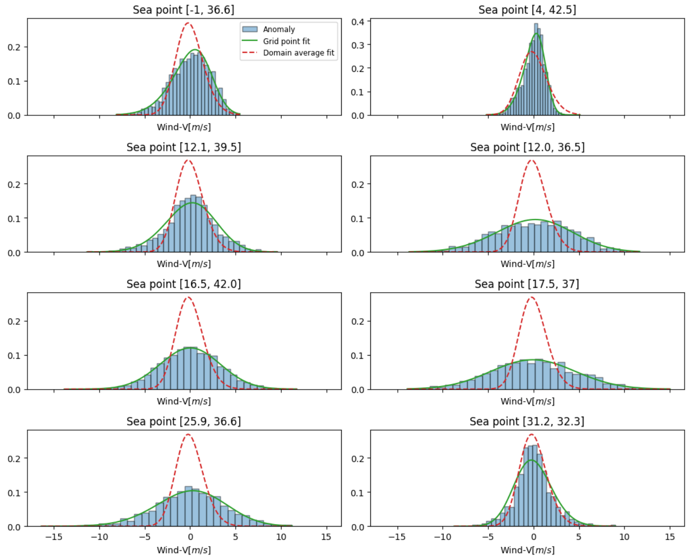
Dew point temperature



Wind U



Wind V



Sampling points: (I) Alboran Sea, (II) Gulf of Lion, (III) Tyrrhenian Sea, (IV) Sicily Strait, (V) Adriatic Sea, (VI) Ionian Sea, (VII) Aegean Sea, (VIII) Levantine Sea]

Parametric QM using skew-normal PDF

P = $\mathrm{CDF}_{\mathrm{model}}$(X_i, shape_fc, loc_fc, scale_fc)

$X_{\text{quantile mapped}}$ = F^{-1} $\mathrm{CDF}_{\mathrm{analysis}}$(P, ref_shape, ref_loc, ref_scale)

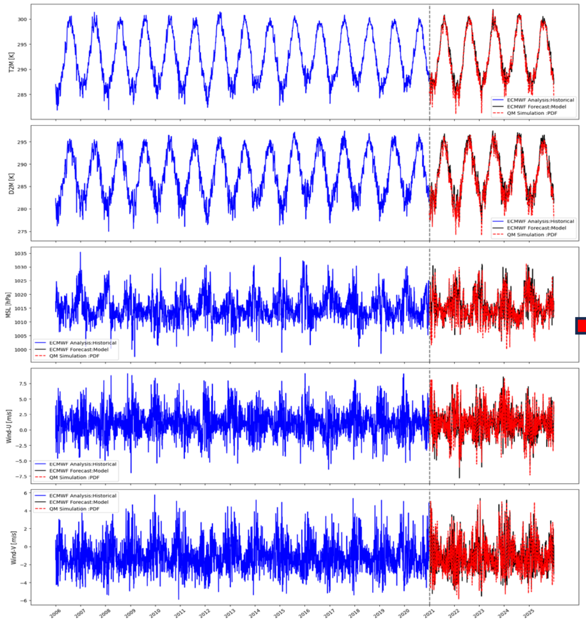
$$p_{(t,i,j)} = cdf_{skewnormal}\left(x_{forecast}, \alpha_{fc}, \mu_{fc}, \lambda_{fc}\right)$$

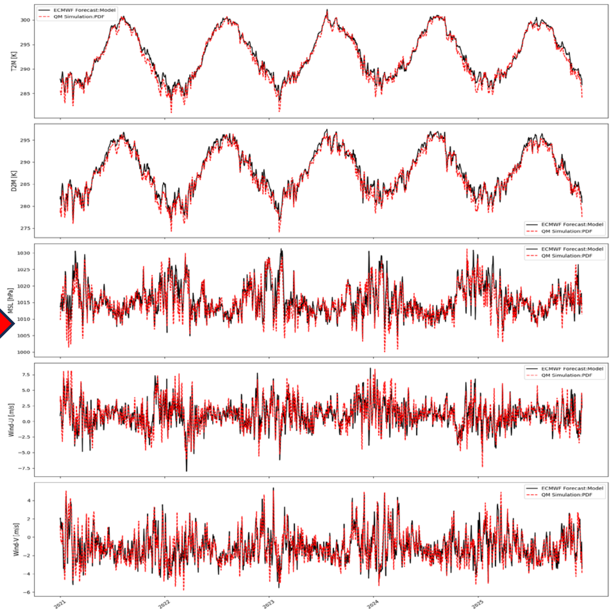
$$X_{corrected} = F^{-1}{}_{analysis}\ (P_{(t,i,j)})$$

Where,

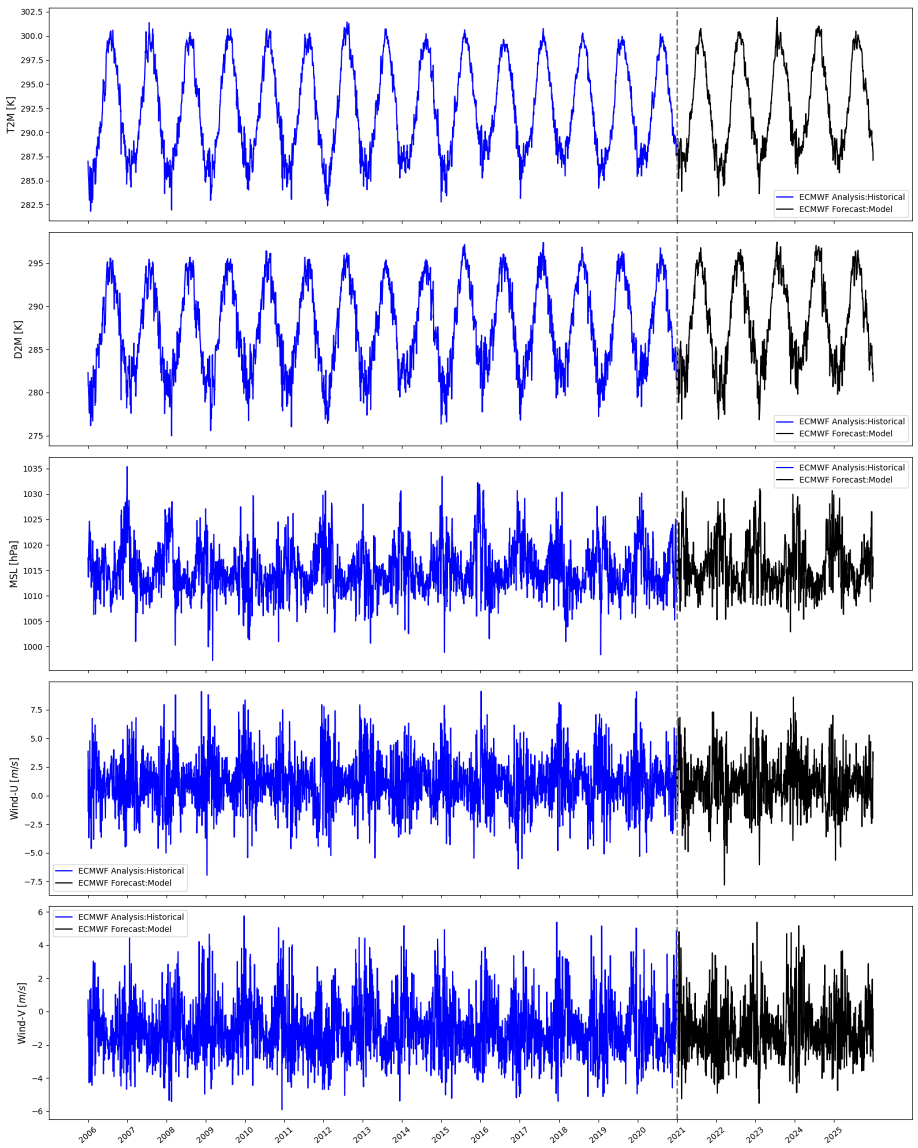
$$F^{-1}{}_{analysis} = PPF(\alpha_{an}, \mu_{an}, \lambda_{an})$$

Analysis 2006-2020; Forecast:2021-2025; QM simulation:2021-2025





Analysis 2006-2020; Forecast:2021-2025



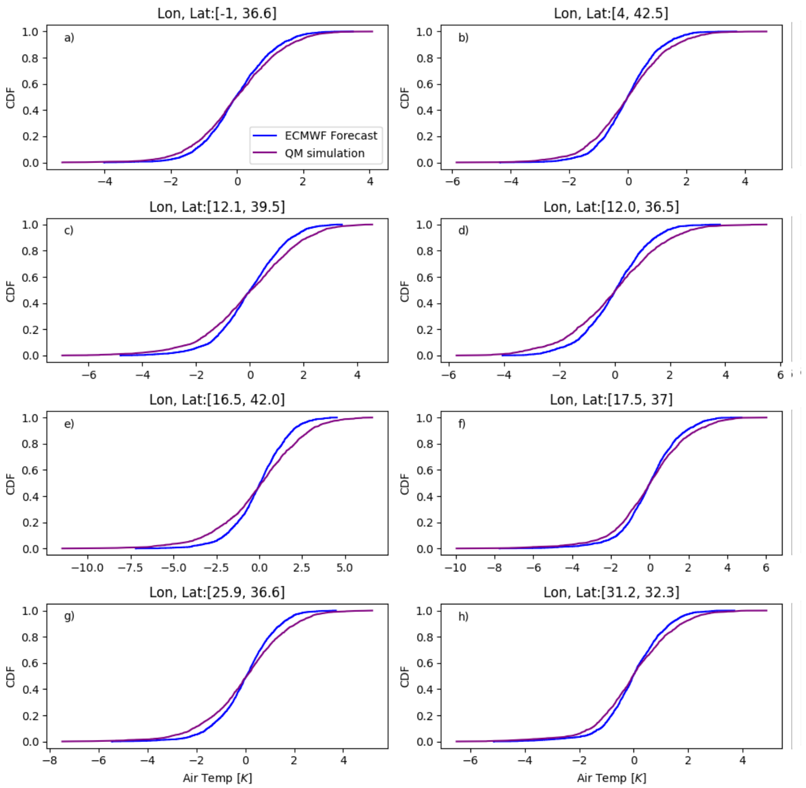
Extremes in the distributions

Percentile values are computed from the estimated PDFs using the inverse Cumulative Density Function (CDF).

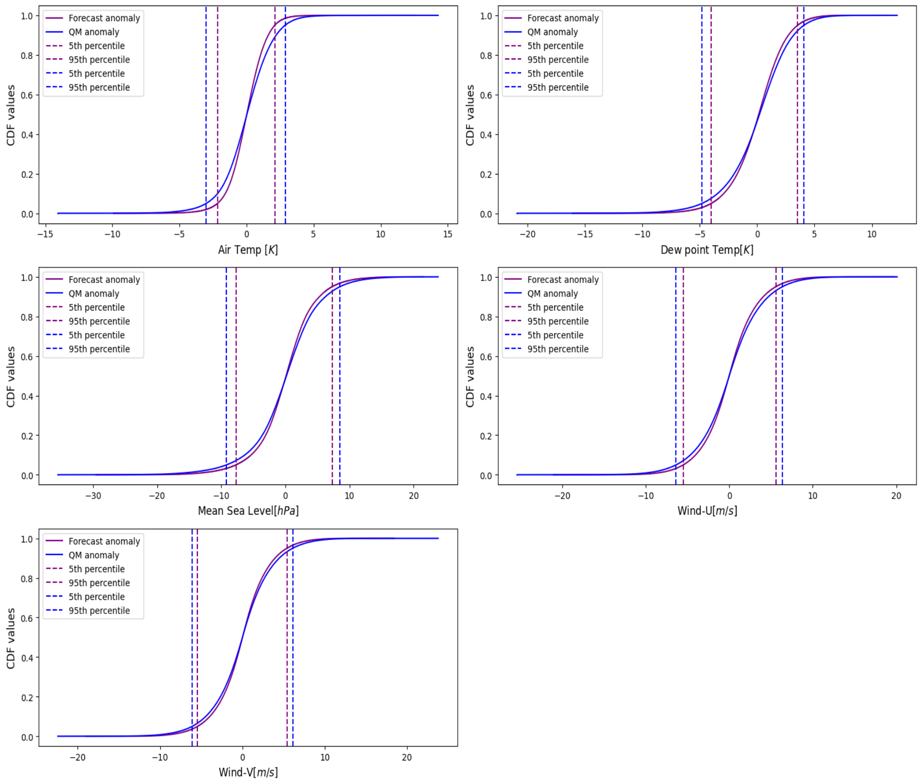
$$f\ (x; \alpha, \mu, \lambda) = \Phi\left(\frac{x-\mu}{\lambda}\right) - 2T\ \left(\frac{(x-\mu)}{\lambda}, \alpha\right)$$

Where, $\Phi(z)$ is the CDF of a normal distribution and $T\ (z; \alpha)$ auxiliary function related to skewness parameter α

CDF on the grid points



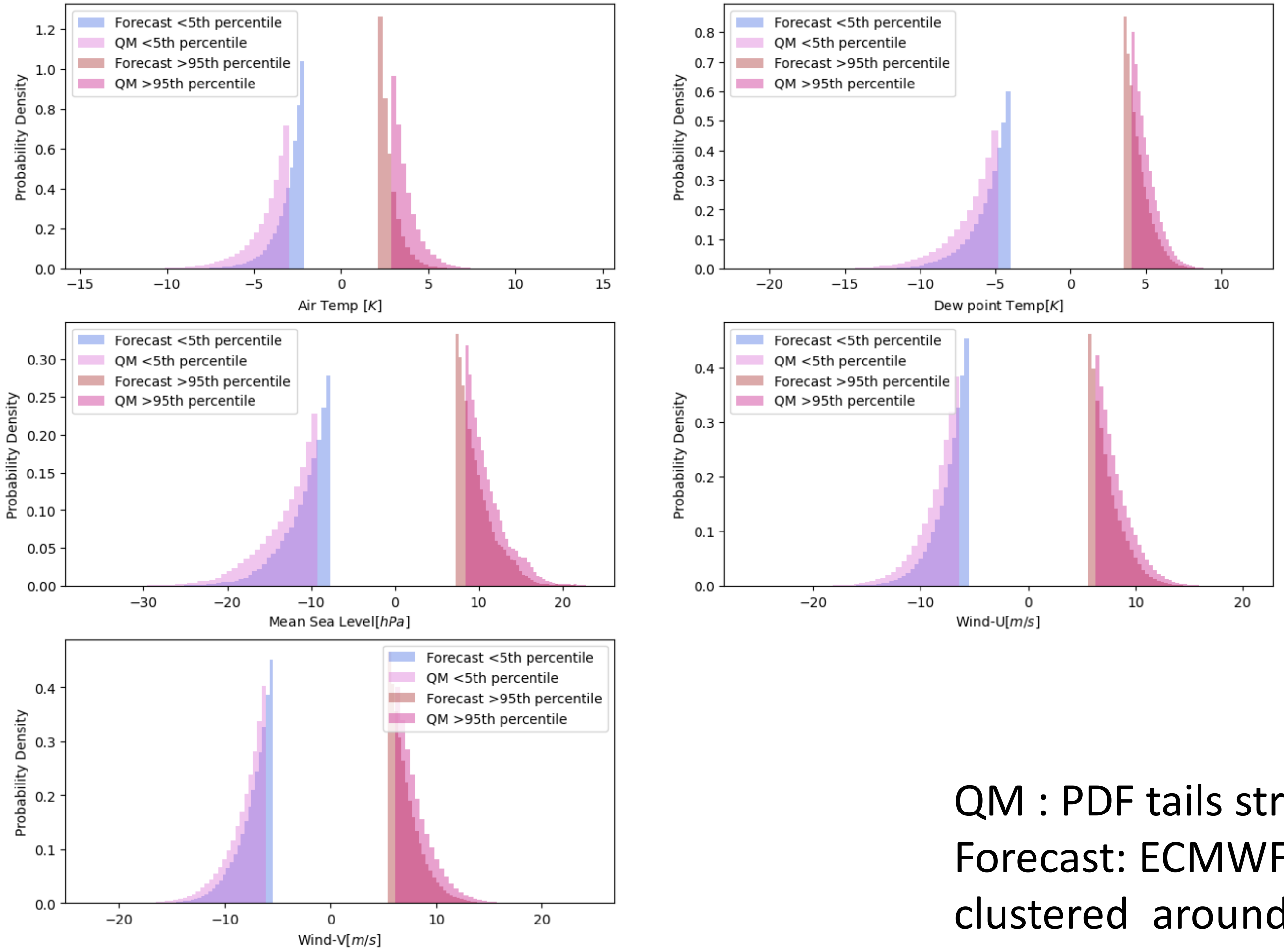
CDF for the Mediterranean Sea



Extremes in the distributions

Distributions of the 95th and 5th percentiles: Forecast and QM

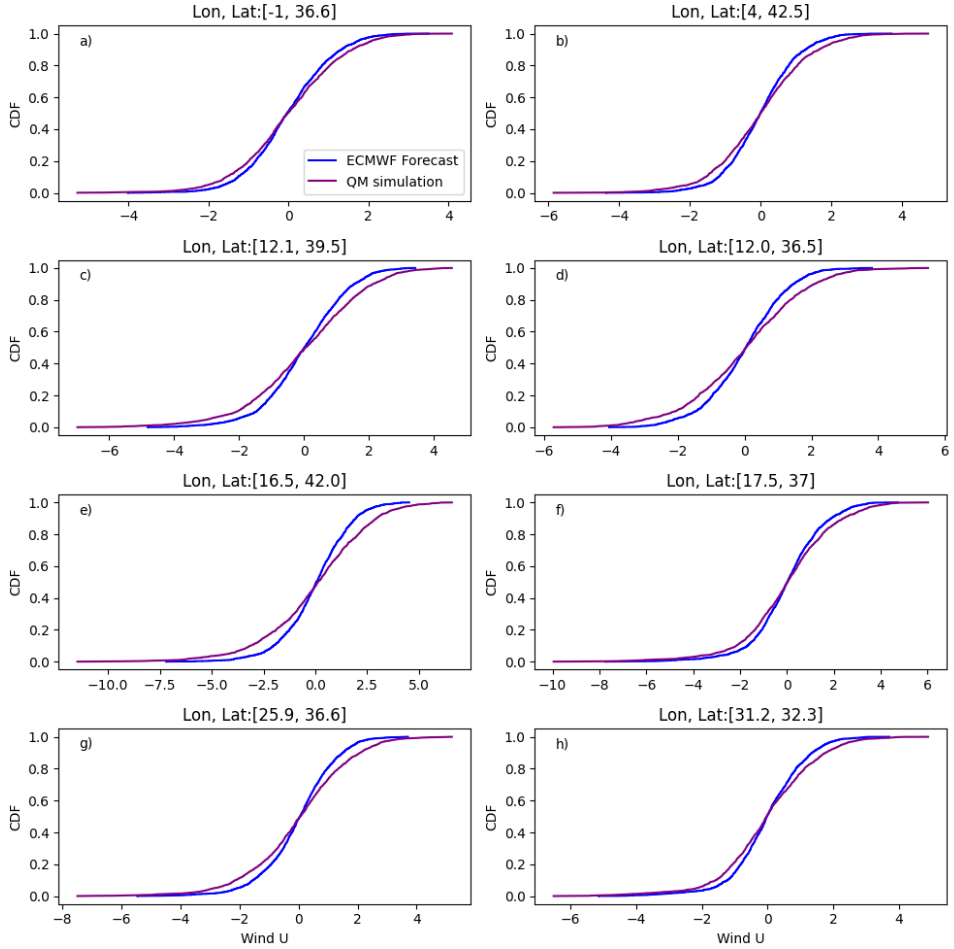
Histograms of combined 95th percentile and 5th percentile



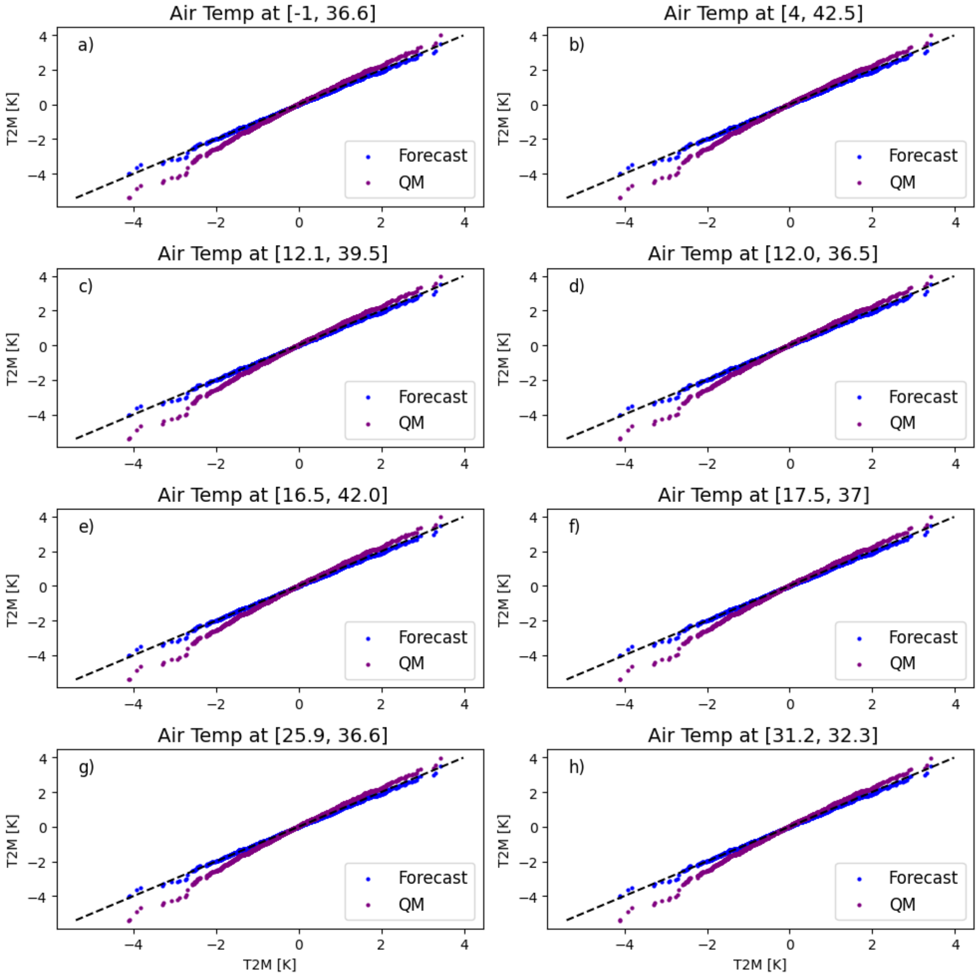
QM : PDF tails stretches out further left or right
Forecast: ECMWF forecast peaks exhibits tall, narrow and clustered around the mean

Extremes in the distributions

Distributions of CDF: Forecast and QM



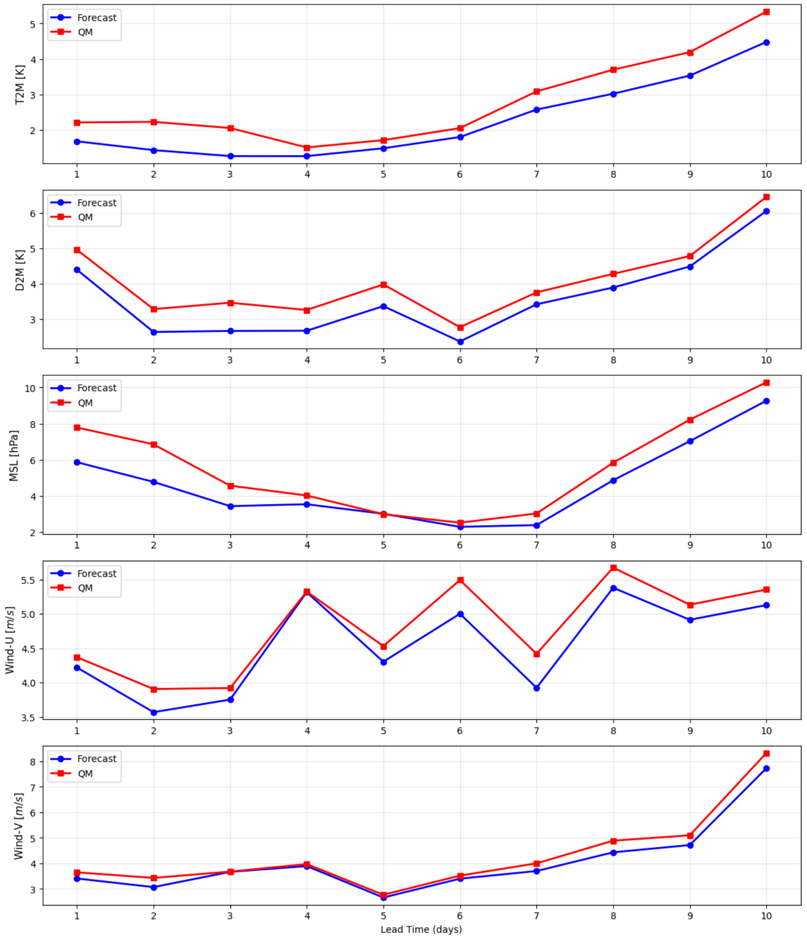
Q-Q plots show raw forecast misses historical tail extremes



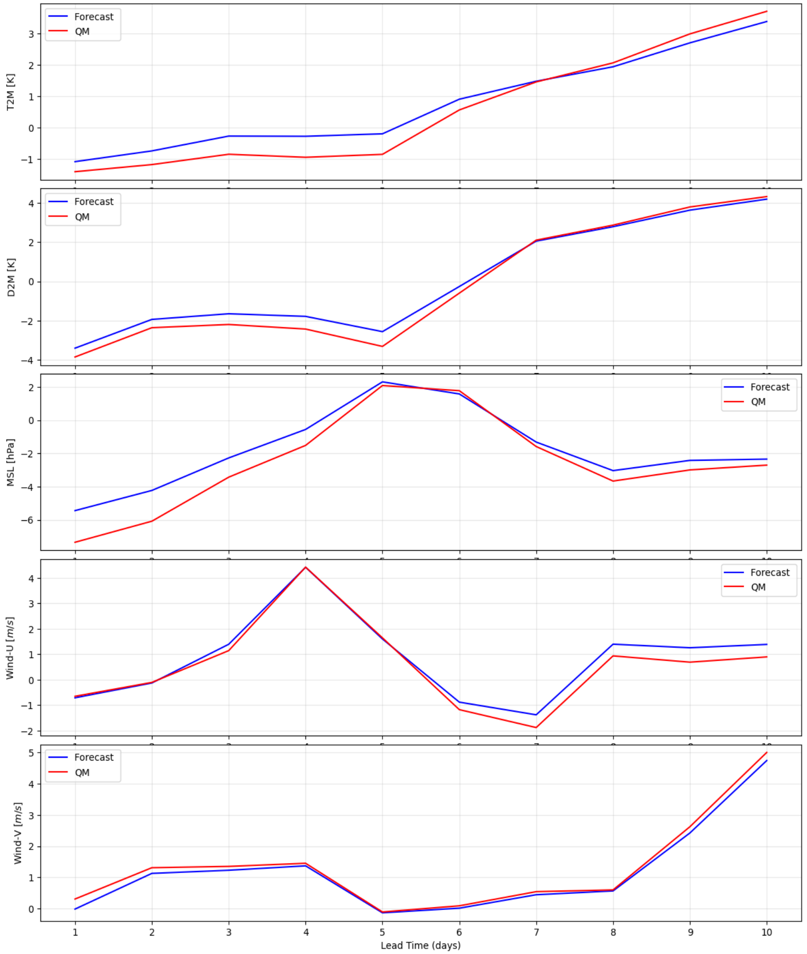
A 10-day ECMWF forecast

A 10-day forecast 01-01-2021 against QM simulation

RMSE

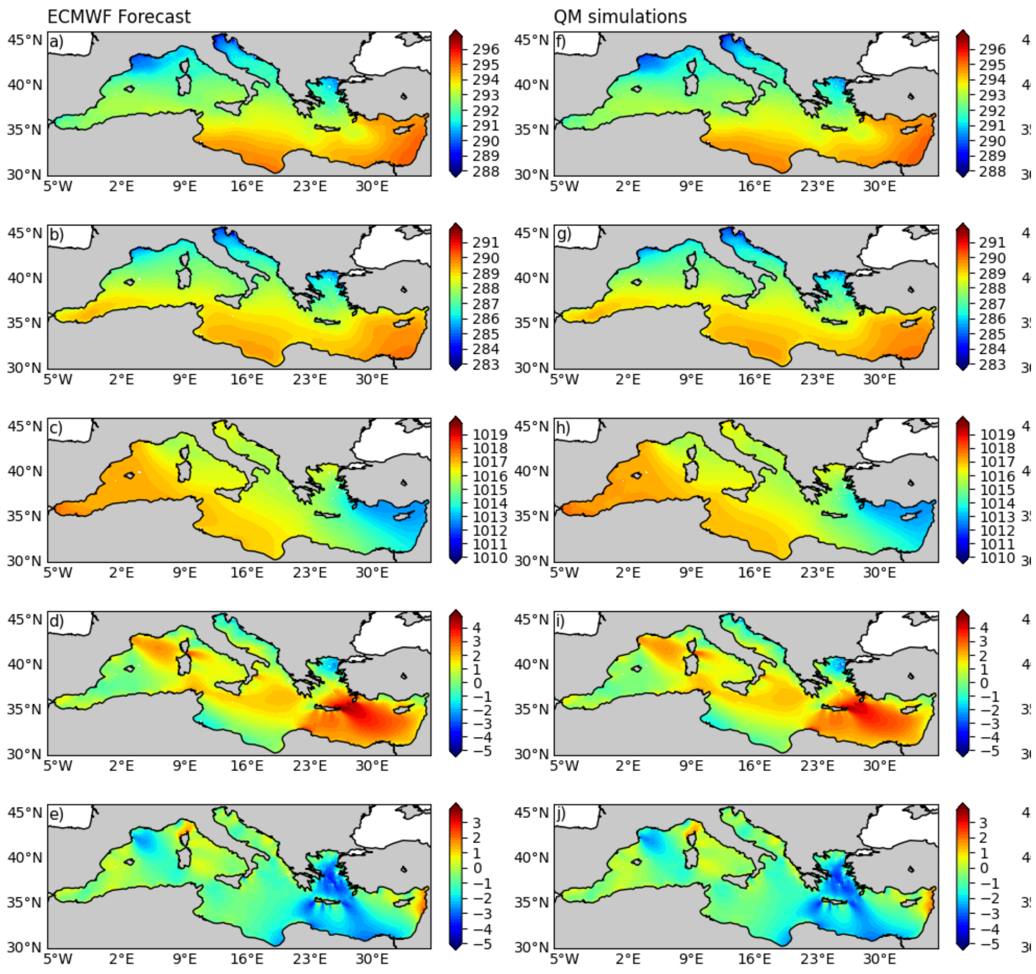


Bias

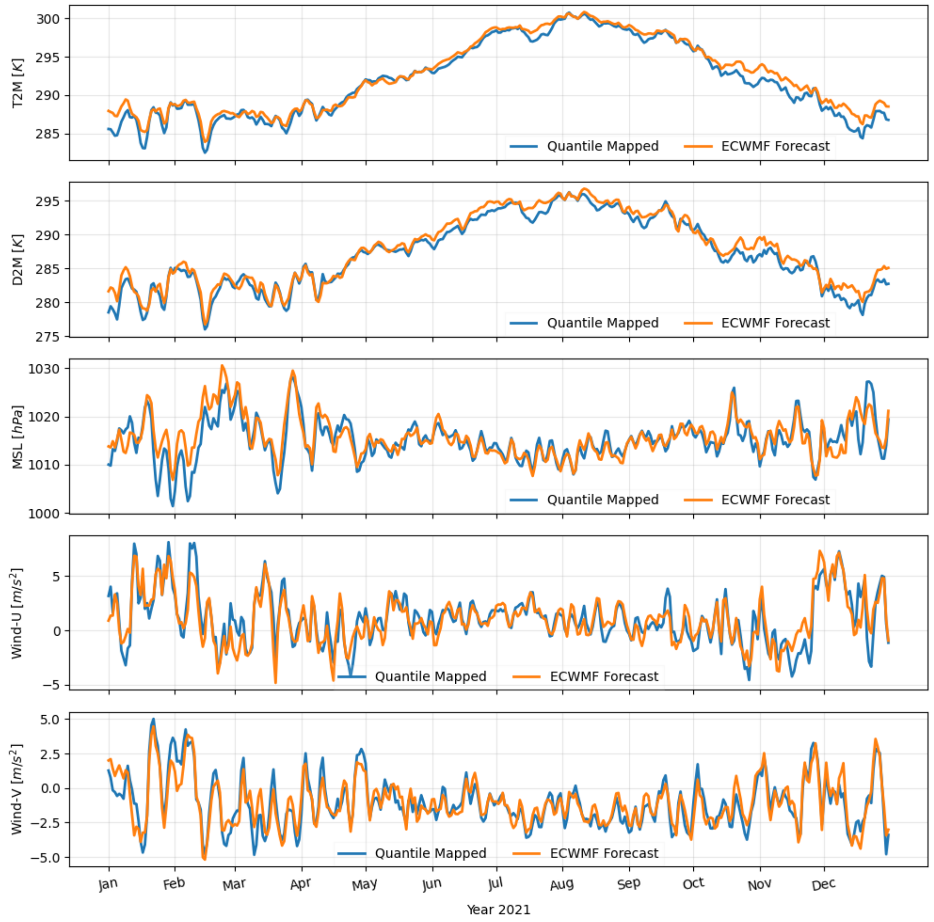


ECMWF forecast vs QM simulations

Atmospheric variables are mapped for the period 2021-2025 using historical skewnormal PDF parameters values computed from the ECMWF analysis (2006-2020)



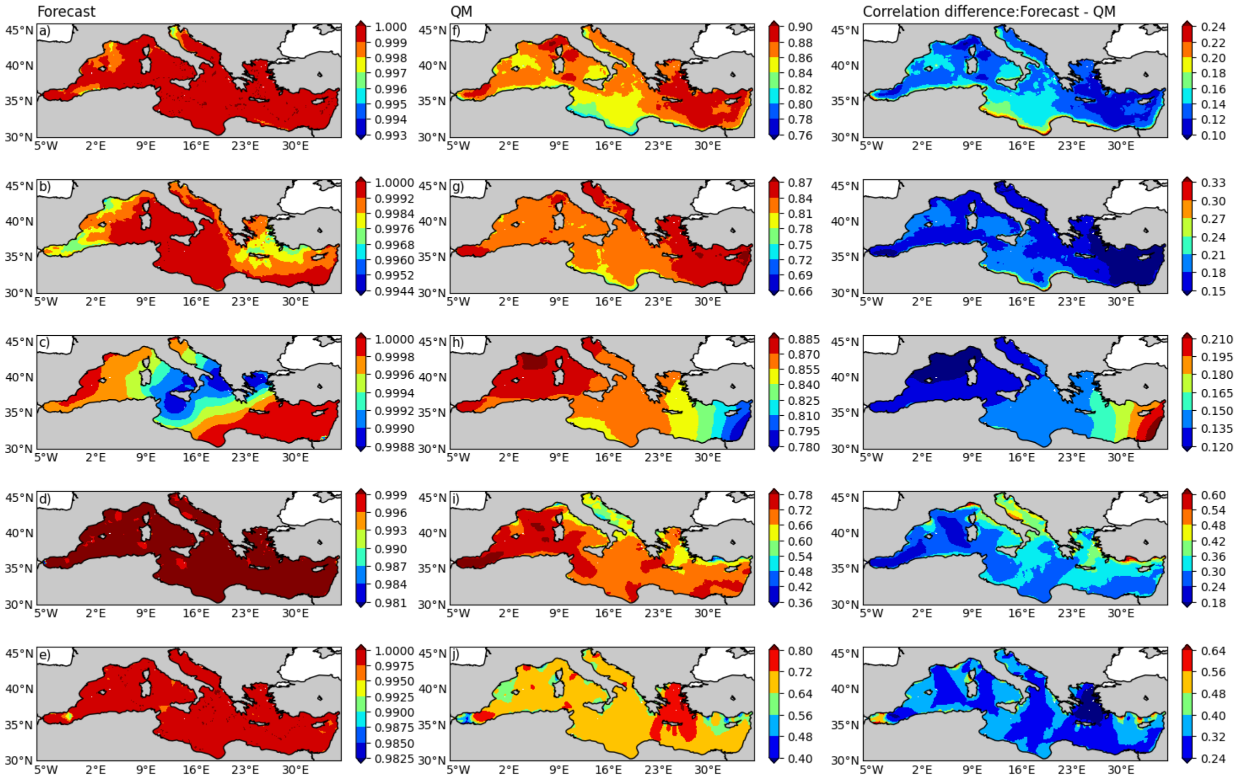
ECMWF forecast and QM simulation year 2021



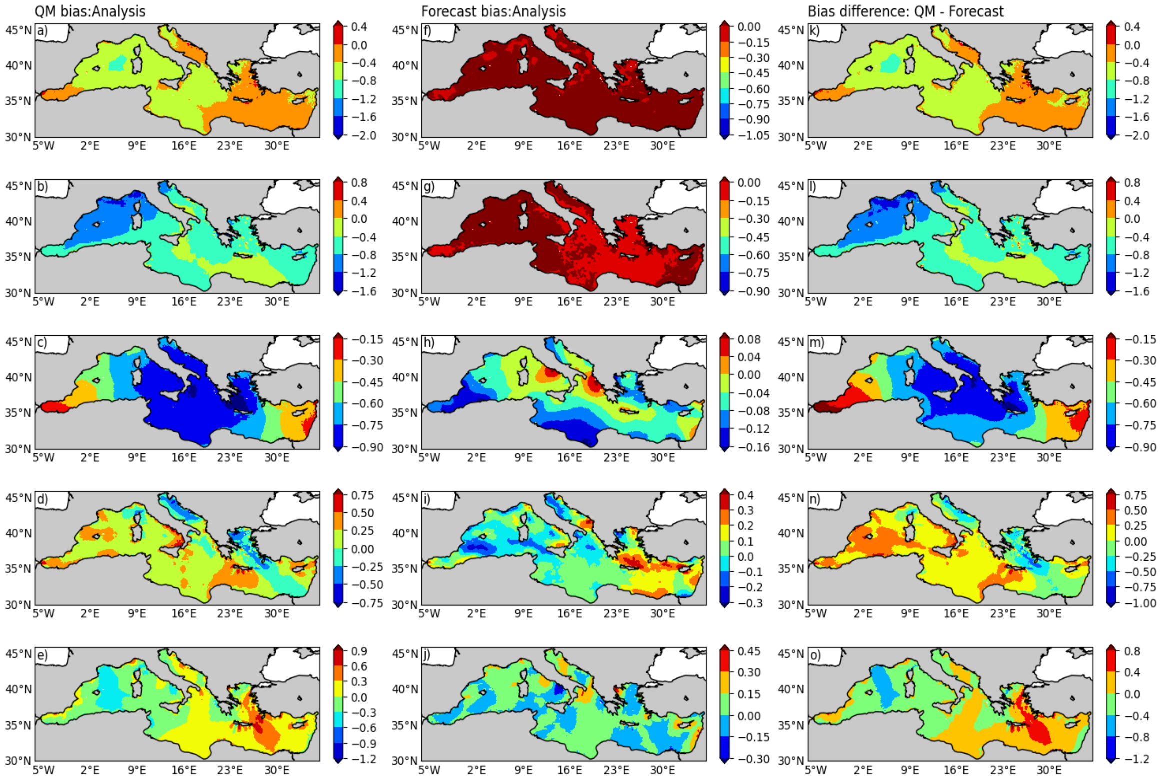
Spatial error statistics

- Raw ECMWF forecast is inherently ‘conservative’, systematically smoothing out extreme anomaly events
- QM disrupts artificial smoothness with realistic, sharp and asymmetric peaks
- A localized spatial correlation occurs within air-sea heat flux hot spots
- QM corrected fields exhibits negative dominance bias across the domain, raw forecast show low magnitude and mixed

Spatial correlation



Bias distribution



Summary

- The Weibull PDF is well suited for wind speed, while the skew-normal PDF effectively captures asymmetry in the tails of major atmospheric surface variables
- A shape parameter (three-parameter PDF) is necessary to describe asymmetry in distributions, whereas the location and scale parameters alone are insufficient
- Most atmospheric variables are non-Gaussian or multimodal; therefore, the skew-normal PDF is a useful alternative for representing asymmetric distributions
- Raw ECMWF forecast generally yields lower global bias and RMSE due to their smoother spatial and temporal structure.
- Skew-normal quantile mapping (QM) simulations better reproduce distribution tails, extremes, and higher-order moments

Next

- Machine learning (ML) techniques provide a newer approach for further investigations, where PDF parameters can be estimated directly using ML-based models.

Summary

Thanks for your attention

Questions! Suggestions!