Progress in ensemble forecasting and verification at ECMWF

Martin Leutbecher, Zied Ben Bouallègue, Thomas Haiden, Simon Lang and Sarah-Jane Lock

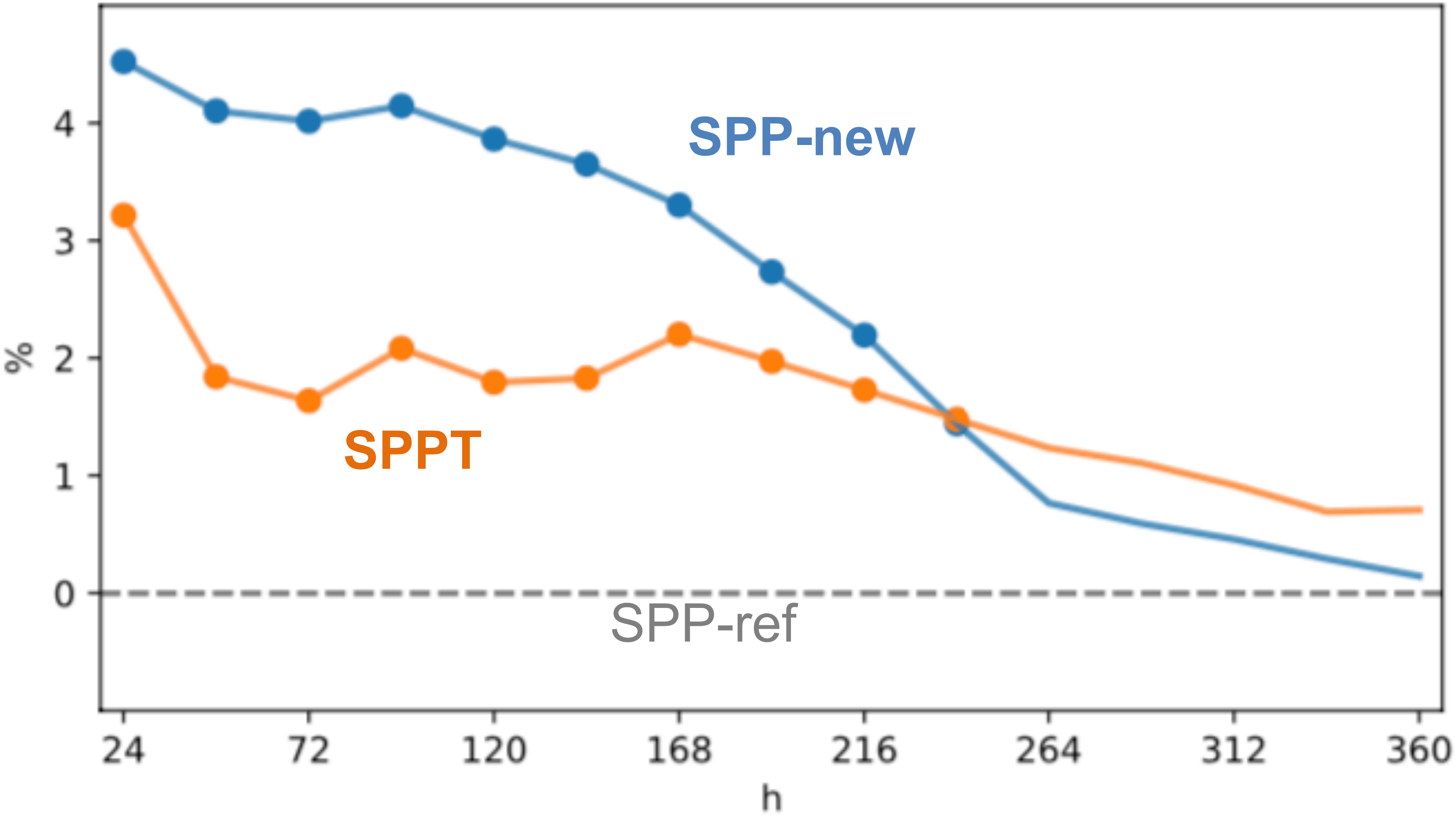
Outline

- Representation of model uncertainties
  - Progress with SPP
  - Development of stochDP
- Planned resolution upgrades
- Representing observation uncertainties in ensemble verification
- Understanding the CRPS through a simplified Gaussian joint distribution of forecast and observations

SPP revision: Summary

- **SPP** stands for **Stochastically Perturbed Parametrisations**; applied in the IFS physics parametrisations in radiation, vertical mixing, cloud and convection schemes
- Represents model uncertainties close to sources, improves physical consistency compared to SPPT, e.g. local conservation properties of energy and moisture
- Original version (**ref**) described by Ollinaho et al. 2017, https://doi.org/10.1002/qj.2931
  - 19 (20) quantities perturbed, 2000 km correlation scale for random fields
  - generates overall less spread than SPPT
- Revised version (**new**, consists of 7 stages) (Lang et al. 2020, submitted to QJ)
  - 27 quantities perturbed
  - 1000 km correlation scale for random fields
  - increased variance of random fields
  - generates slightly more spread than SPPT overall and is about as skilful as SPPT (Latest SPPT config., see Lock et al, 2019, https://doi.org/10.1002/qj.3570)

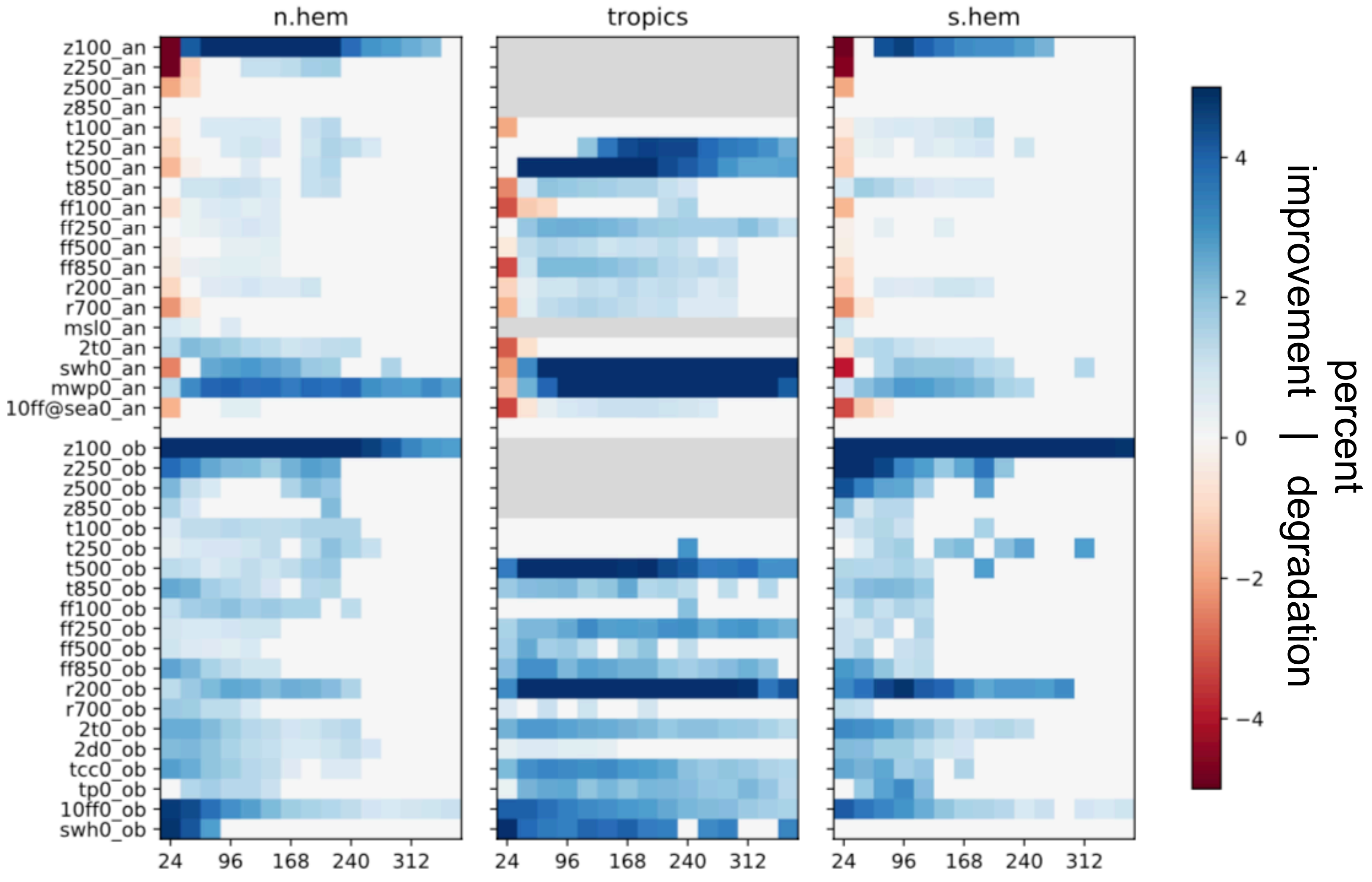
Relative Z500 ensemble spread increase wrt SPP-ref in N-Hem extra-tr.



SPP-new versus SPP-ref scorecard showing fCRPS changes

verified against analysis

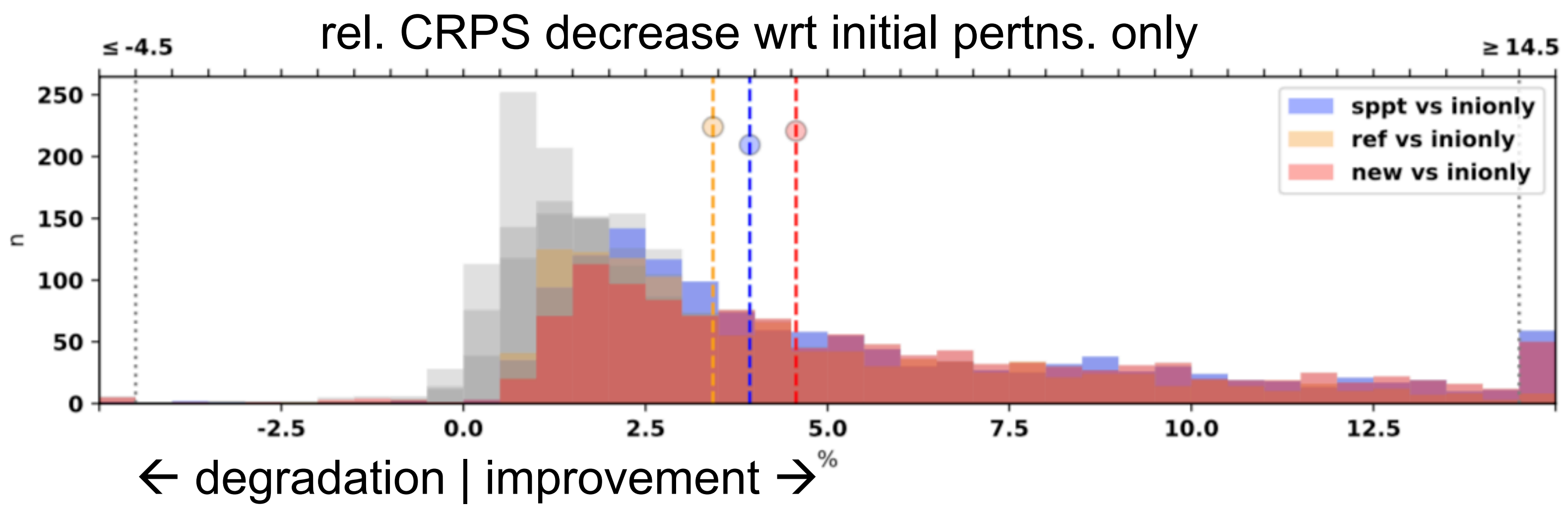
verified against observations



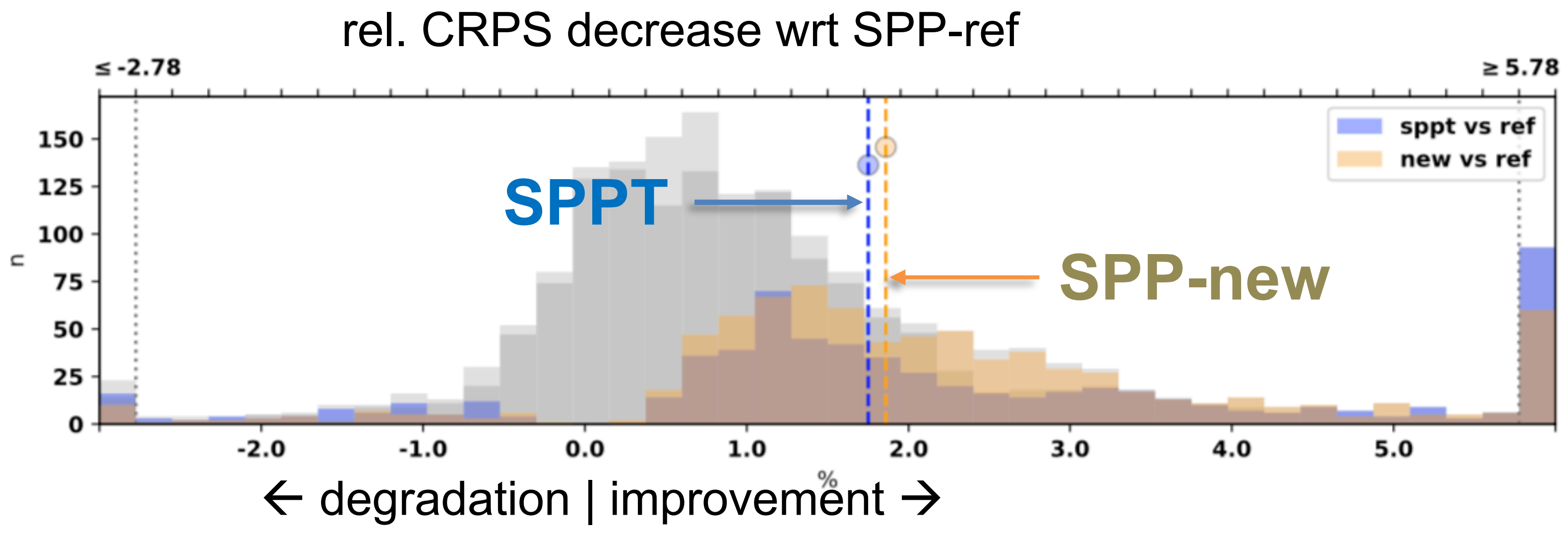
lead time (h)

Coloured: 99.7% sig. level

Histogram of relative CRPS changes



b)



lead times 24 h ... 360 h
combination of variables and levels evaluated in scorecard
3 regions: N-Hem, S-Hem, tropics

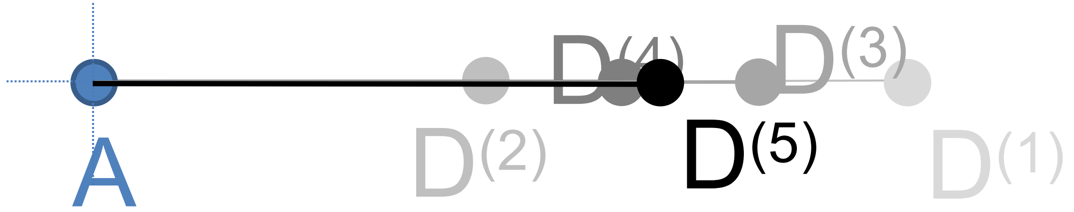
vertical lines: median change

based on 8 members and fair CRPS, boreal summer + boreal winter, 212 start dates, TCo399

colour: stat. significance 99.7%, grey otherwise

STOCHDP:

Stochastically perturbed semi-Lagrangian (SL) departure point (DP) estimates



Diamantakis & Magnusson (2016):

- Explored convergence rate of the iterative DP estimate
- Slowest convergence ←→ most complex flow (strong shear / curvature)
- e.g. Typhoon Neoguri:
  - HRES forecast: initialised: 2014-07-05, 00UTC

Fig. 1c: t+96h, 850hPa windspeeds

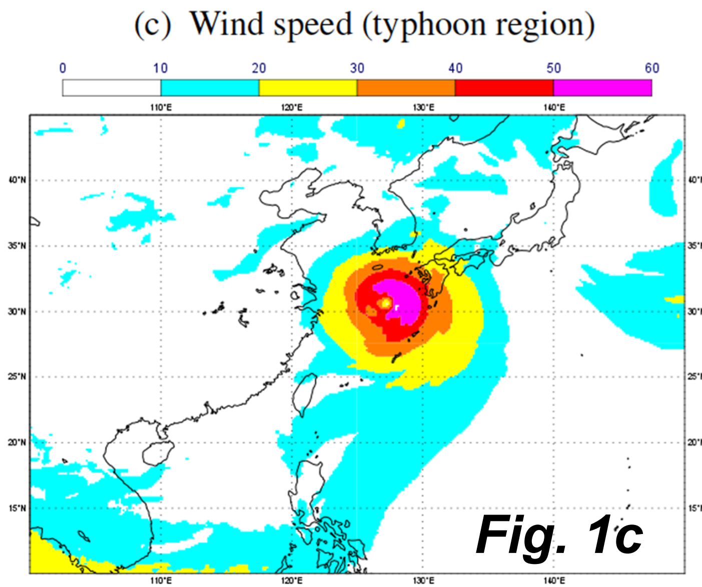
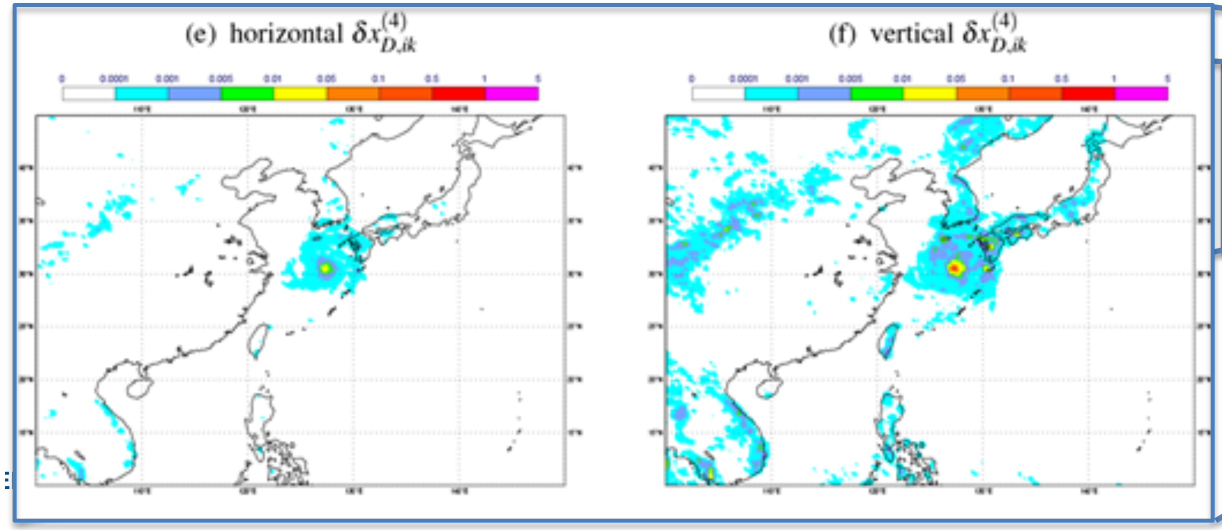
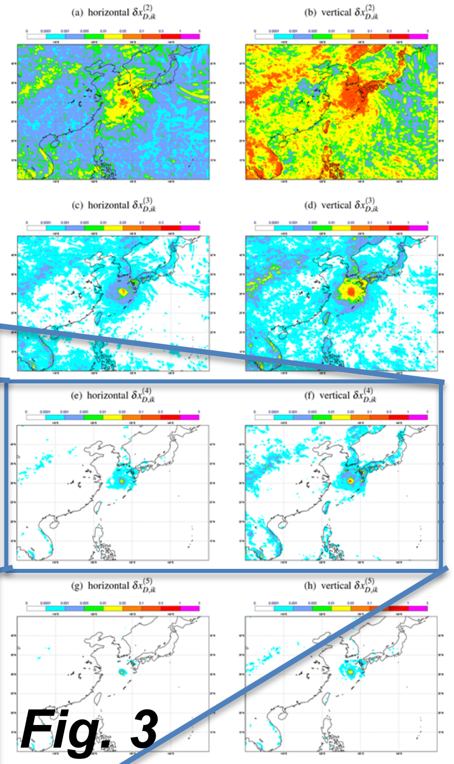


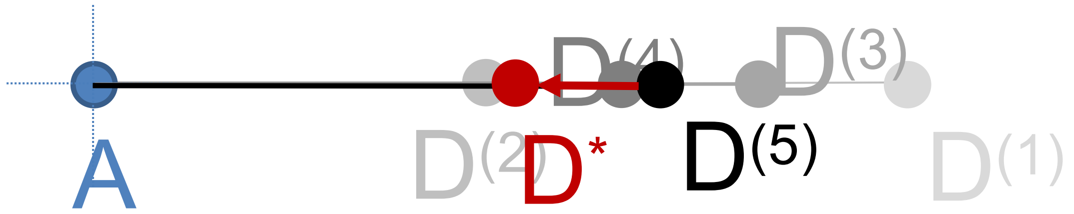
Figure 3: difference in DP estimate between consecutive iterations (scaled)





STOCHDP:

Stochastically perturbed semi-Lagrangian (SL) departure point (DP) estimates



Model uncertainty scheme, “STOCHDP”:

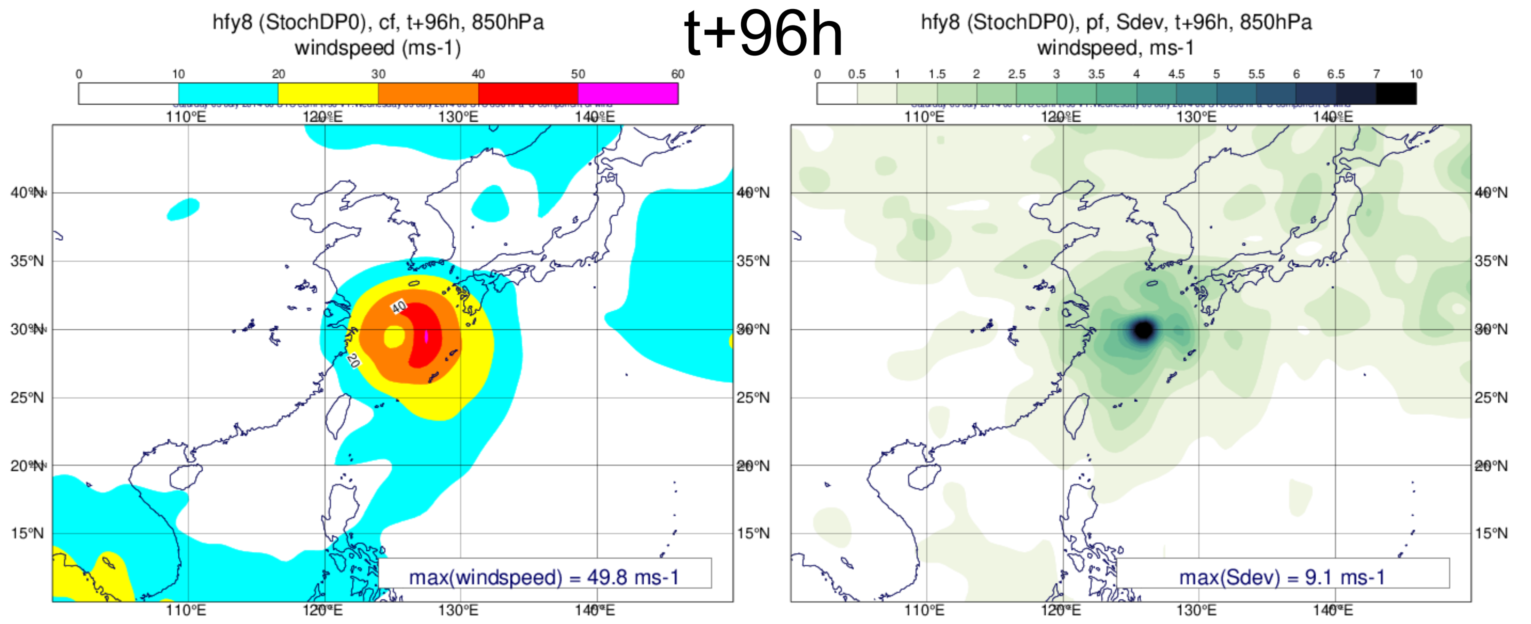
- use the DP estimate convergence rate to attribute MU:

$$D^* = D^{(5)} + r\left(D^{(5)} - D^{(5-i)}\right), i = 1..4$$

where $D*$ is the perturbed DP and r is a random number

➢ STOCHDP represents MU from SL advective winds

Early results, e.g.:
- Typhoon Neoguri case
- ENS: STOCHDP only
- TCo639L91, dt=720s
- 20+1 members
- Peak ENS stdev develops and tracks with TC



Control forecast

Ensemble stdev

Next resolution upgrades

Current medium-range resolution: TCo639L91 (18km), dt=720s

Candidates for medium range ensemble (15 days):

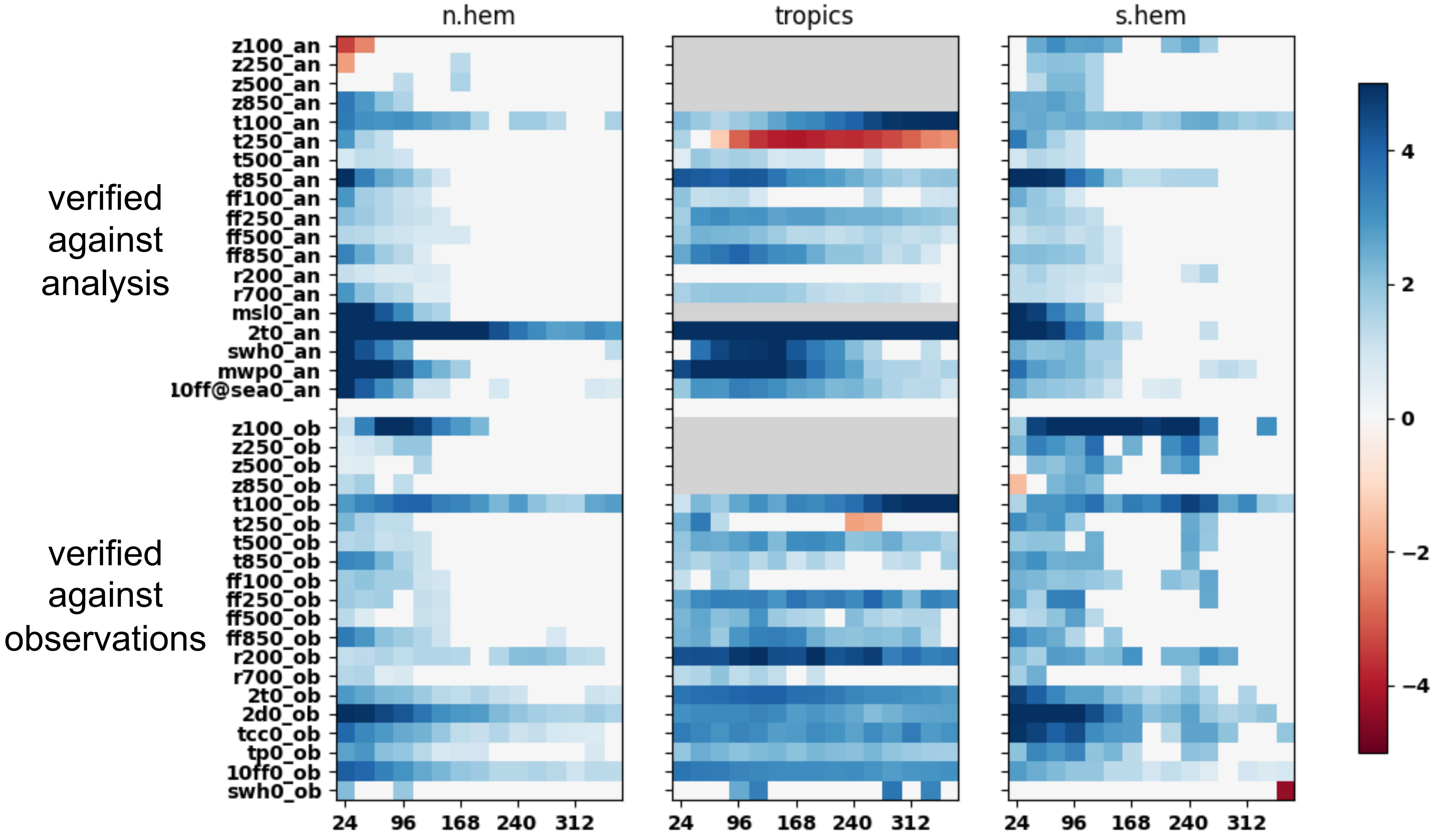
Resolution, timestep	Factor* (Single Precicion ~ x 0.7)
TCo911L137 (12.7 km), dt=600s	3.8 (2.7)
TCo1023L137 (11.3 km), dt=450s	6 (4.2)
TCo1279L137 (9 km), dt=450s	8.9 (6.2)

* Approximate cost increase (from RD experimentation, no operational output) relative to TCo639L91 (18 km), dt=720s:

Planned in two stages (pending performance tests on XC40)

- L91 (DP) → L137 (SP), Cycle 47r2, Cray XC40 in Shinfield Park, Reading, Q2 2021
- TCo639 → TCo1000+, Cycle 48r1, BullSequana XH2000 in Bologna, Q3/Q4 2022

TCo1279L137 vs TCo639L91, scorecard fCRPS



Coloured: 95% sig. level

summer+winter, 212 initial dates

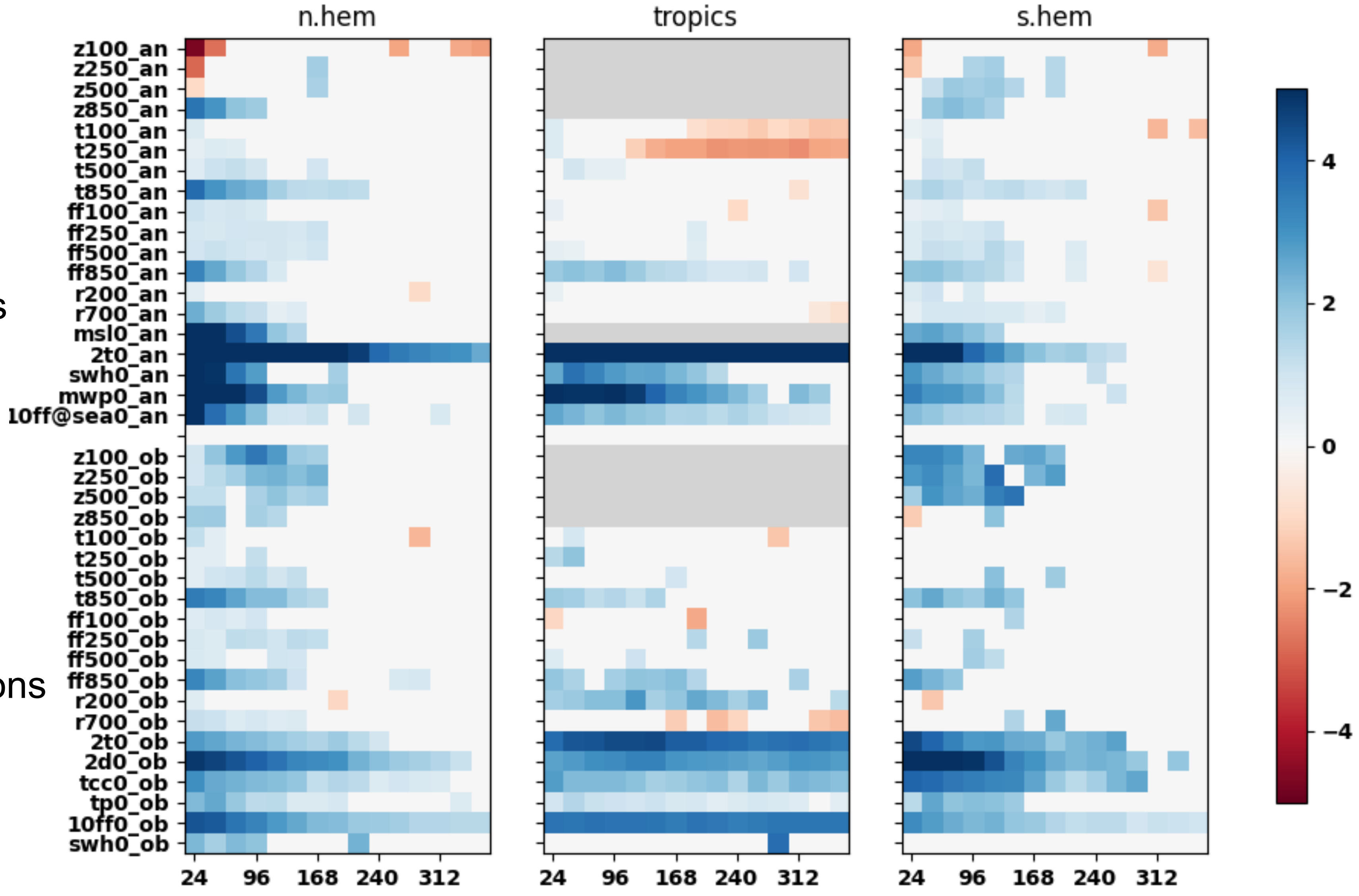
8 perturbed fc

ECMWF

TCo1279L137 vs TCo639L137, scorecard fCRPS

verified against analysis

verified against observations

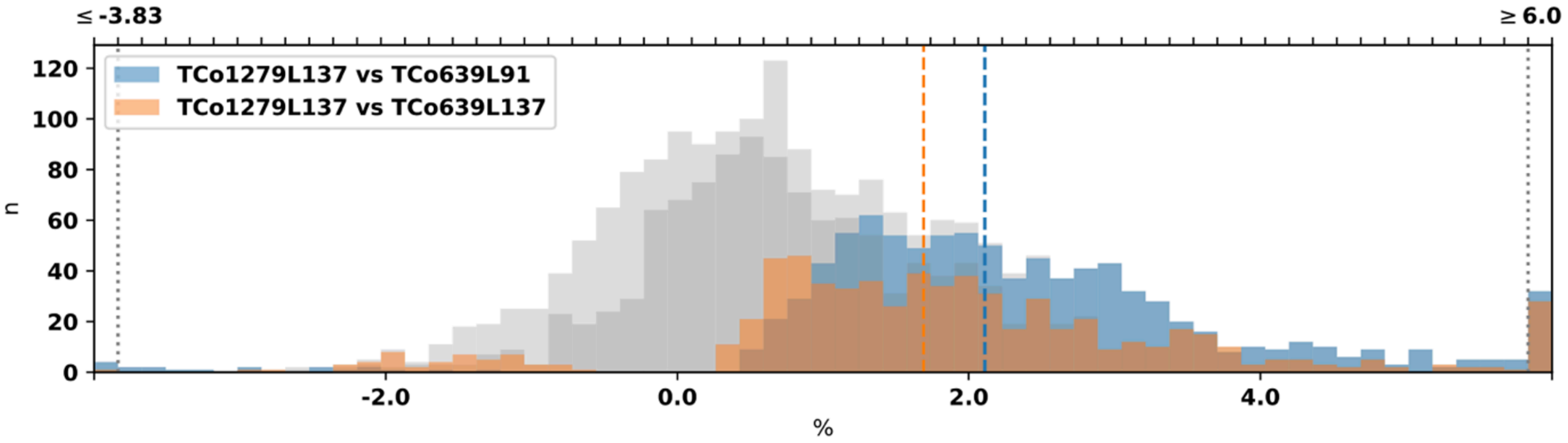


summer+winter,
212 initial dates

8 perturbed fc

Coloured: 95% sig. level

ECMWF



Histograms summarizing the relative fCRPS changes (%) for all variables and lead times from combined boreal summer and boreal winter period experimentation (212 initial dates). Shown are results for TCo1279, with 137 levels versus the TCo639 91-level control experiment and versus the TCo639 137-level control experiment. Statistical significance is indicated by colour and the vertical lines show the median score change.

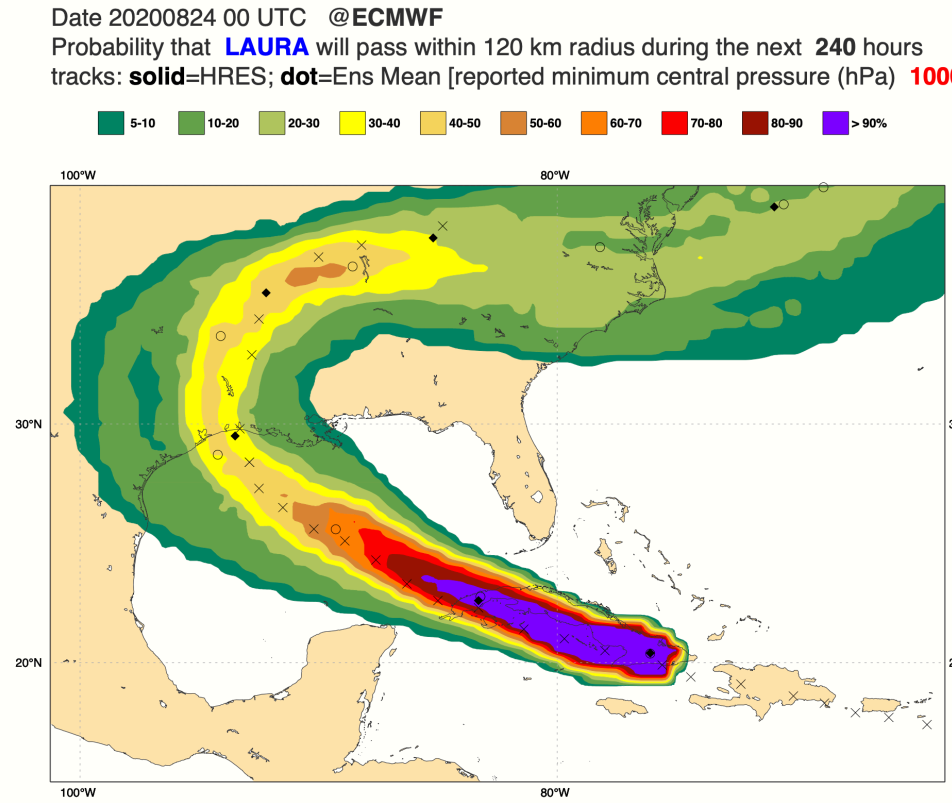
TC Laura 2020, resolution sensitivity
Oper ENS (TCo639L91) versus TCo1279L137 Experiment

Both started from same (oper) initial conditions, 50 perturbed members

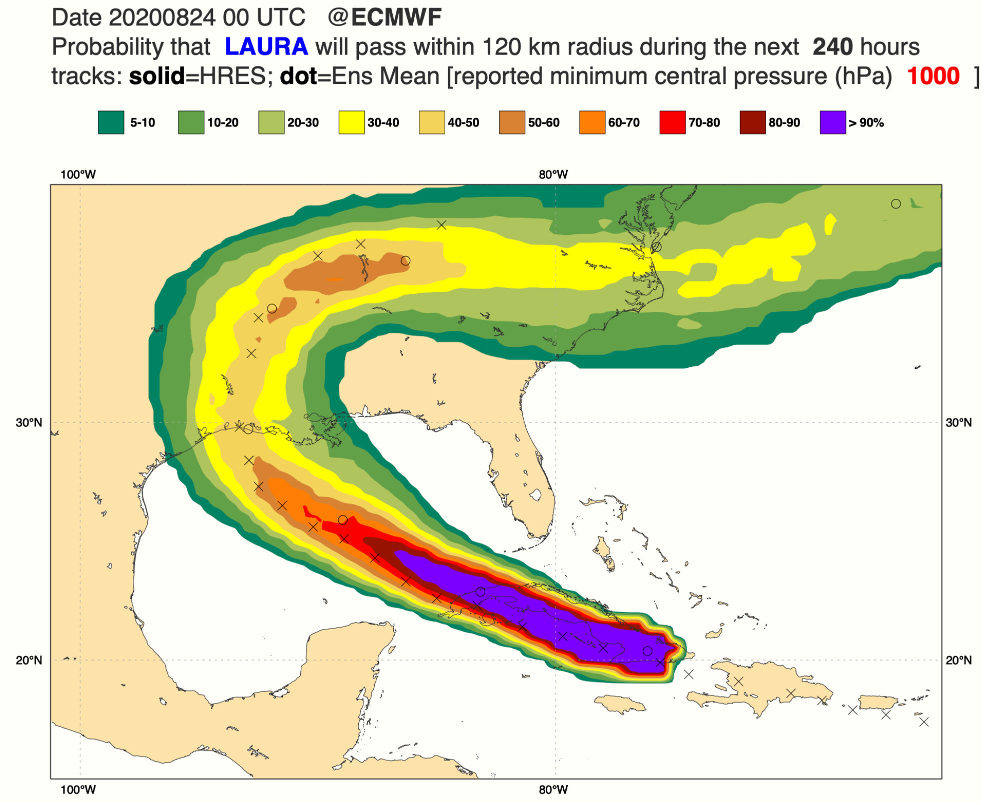
Subjective case assessment, TCo1279L137 vs TCo639L91:

Initial Date	Landfall location	Core Pressure at Landfall
2020082300	Significantly improved	Significantly improved
2020082312	≈ Neutral	Significantly improved
2020082400	Significantly improved	Significantly improved
2020082412	Significantly improved	Significantly improved
2020082500	Significantly improved	Significantly improved
2020082512	Improved	Significantly improved

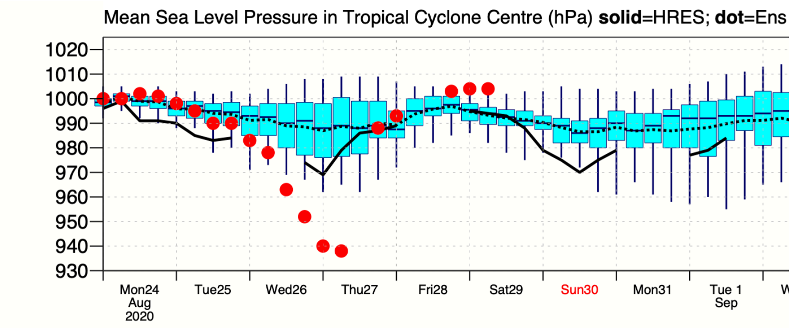
Oper, TCo639L91



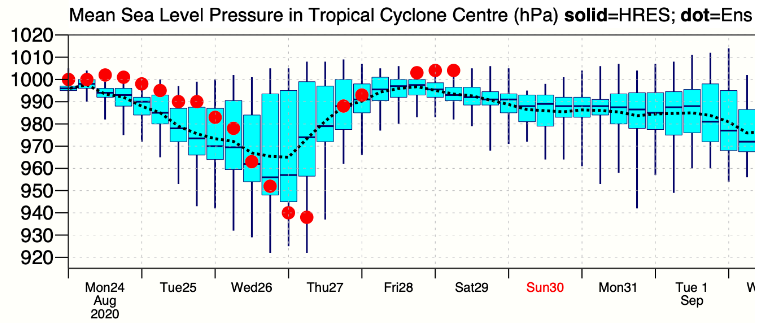
TCo1279L137



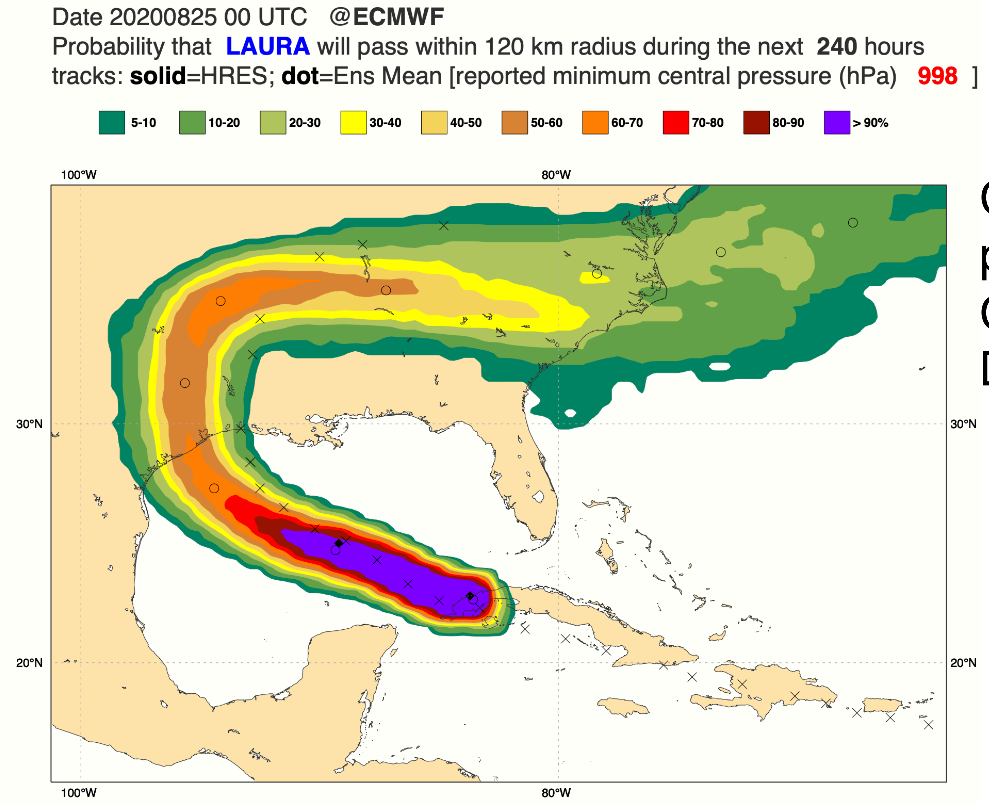
Crosses : observed position
Circle : ensemble mean
Diamonds : Oper HRES



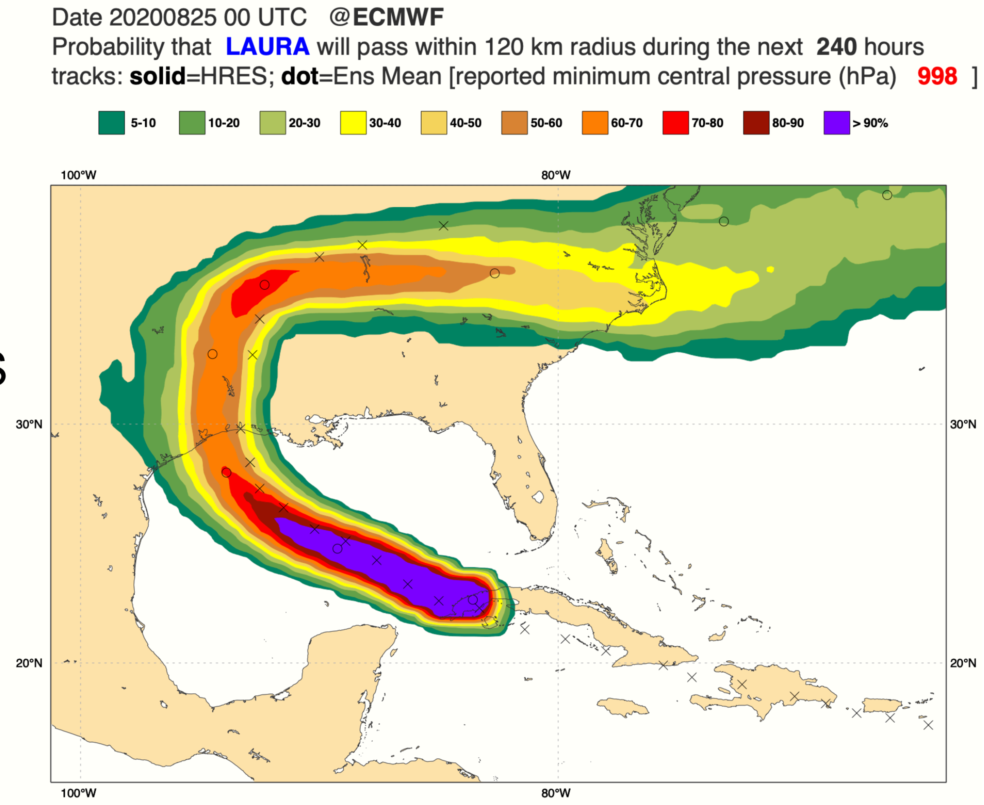
Red dot : observed core pressure
Solid black line : Oper HRES
Box plot : ensemble distribution



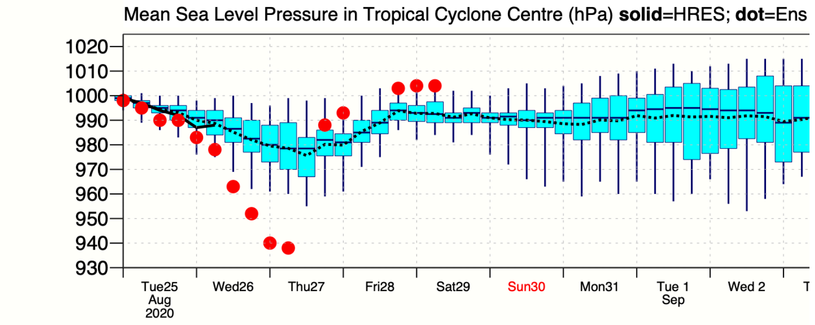
Oper, TCo639L91



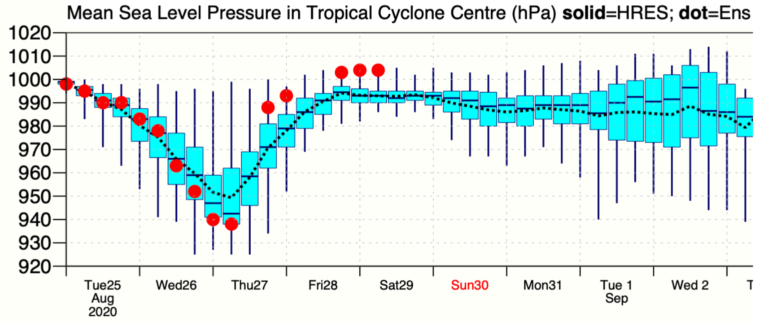
TCo1279L137



Crosses : observed position
Circle : ensemble mean
Diamonds : Oper HRES



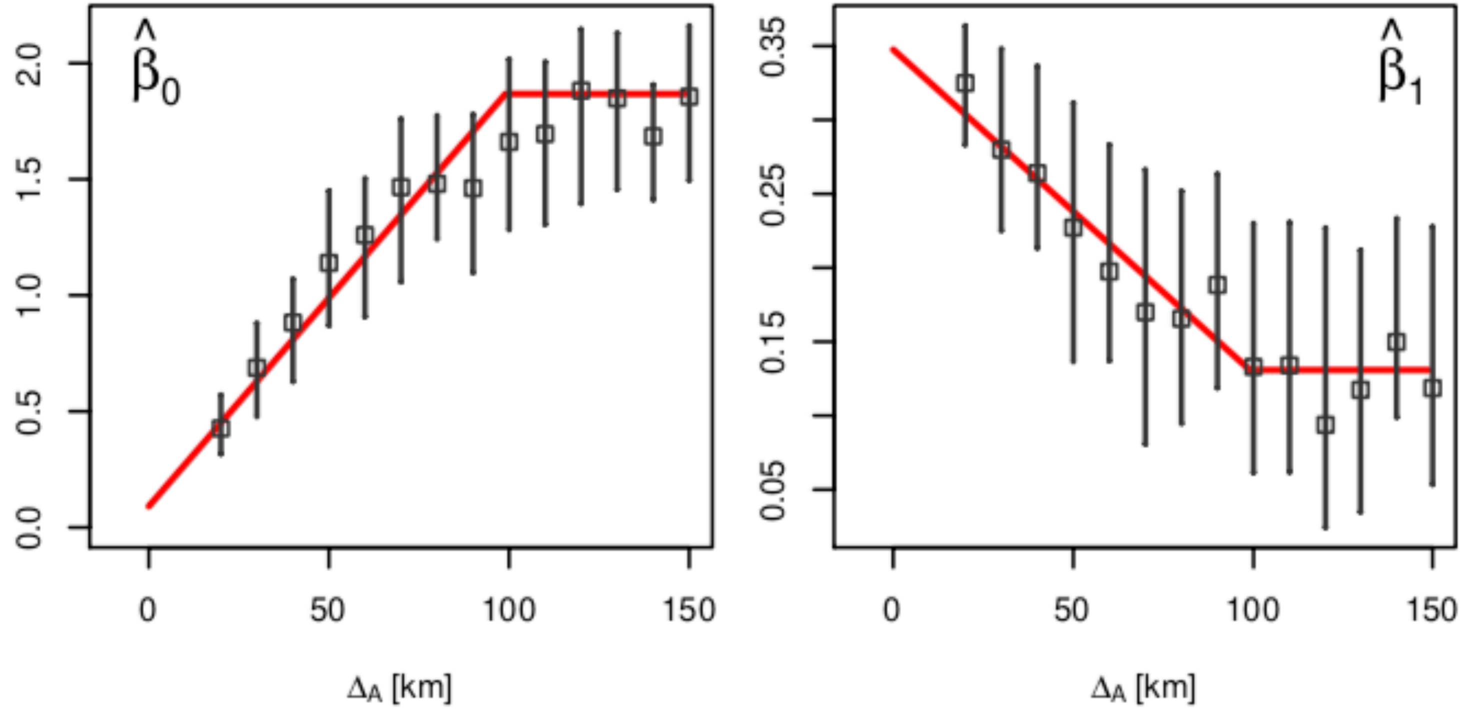
Red dot : observed core pressure
Solid black line : Oper HRES
Box plot : ensemble distribution



Accounting for representativeness error in ensemble verification

- Many references state why there is a need to account for representativeness errors (REs) in ensemble verification
- However, in practice most ensemble verification does not account for it.
- Progress has been made at ECMWF to account for REs routinely: Ben Bouallègue et al (2020, https://doi.org/10.1175/mwr-d-19-0323.1) and Ben Bouallègue (2020, ECMWF Tech Memo 865, https://doi.org/10.21957/5z6esc7wr)
- Parametric models have been estimated using a normal distribution, a truncated normal distribution and a censored shifted gamma distribution for 2-metre temperature, 10-metre wind speed and 24-hour precipitation, respectively.
- The statistical models describe the distribution of values at station locations given area-average values of the respective variables
- The models are given as function of the horizontal averaging scale and can be applied to any NWP model output
- The distributions are estimated from high-density station observations over Europe by minimising the CRPS

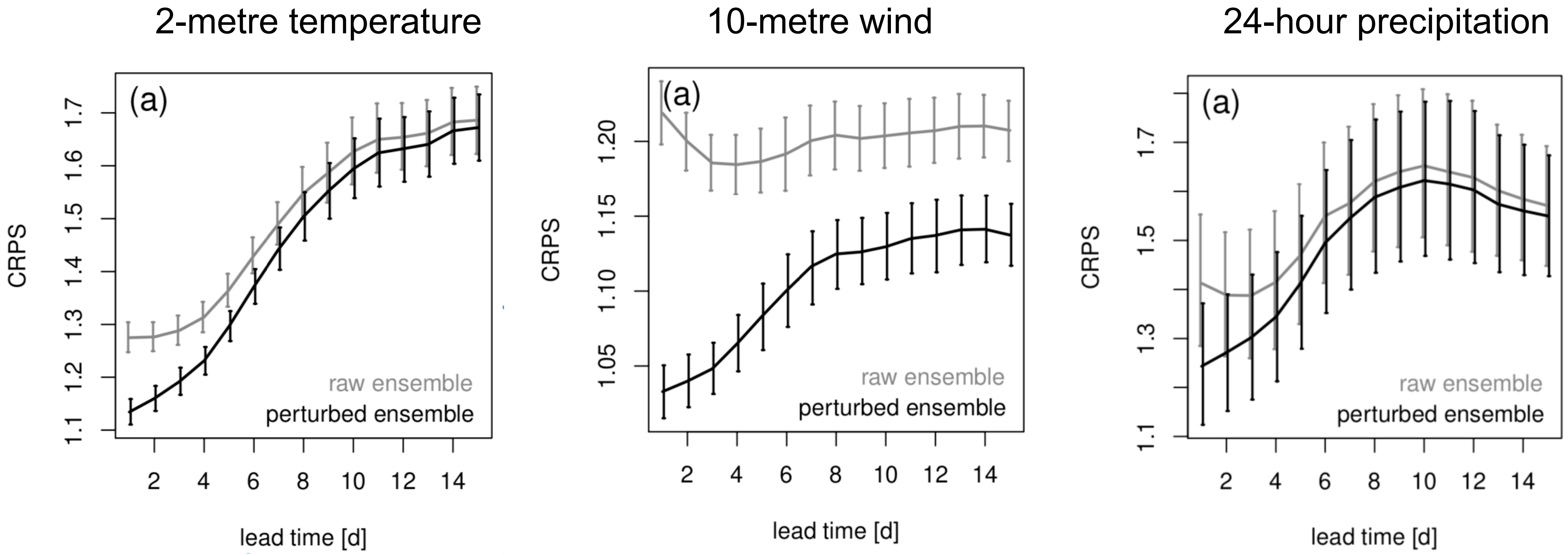
Statistical model for 2-metre temperature station observations



$$t^\star = t + 0.0065 D_e \quad \text{with} \quad D_e = e_\mathrm{m} - e_\mathrm{o} \quad , \tag{2}$$

$$\mu = t^\star, \quad \sigma = \beta_0 + \beta_1 \sqrt[4]{| D_e |} \quad . \tag{3}$$

Impact on CRPS from accounting for REs



Europe, summer 2018
see ECMWF Tech Memo 865

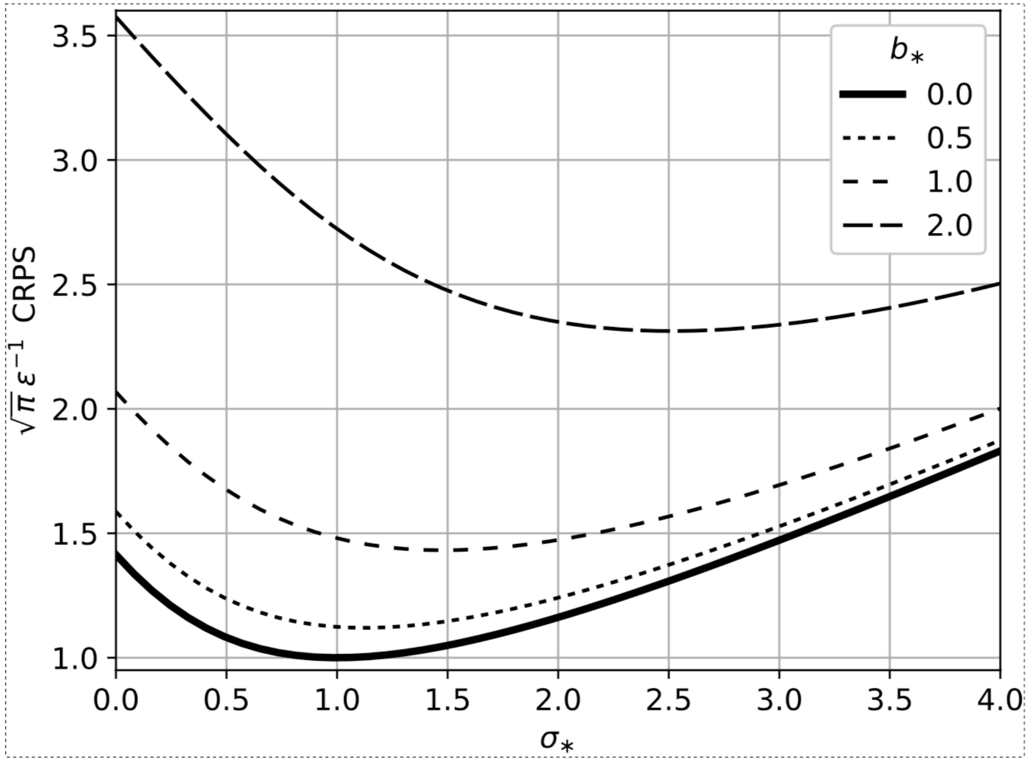
Understanding changes of the Continuous Ranked Probabilty Score

- Decisions about implementing changes depend on scorecards, i.e. Δ CRPS
- Can we explain quantitatively why the CRPS changes?
- Yes, with some simplifying assumptions
- Assume homogeneous Gaussian model (hoG) of forecast-observation distribution
- This permits to compute the expected CRPS as a closed form expression
- Details are in an article (Leutbecher &Haiden, QJ in review)

$$\mathbb{E}\,\mathrm{CRPS} = \frac{\epsilon}{\sqrt{\pi}}\left[\sqrt{2+2\sigma_*^2}\,\exp\left(-\frac{b_*^2}{2+2\sigma_*^2}\right) + \sqrt{\pi}\,b_*\,\mathrm{erf}\left(\frac{b_*}{\sqrt{2+2\sigma_*^2}}\right) - \sigma_*\right]$$

with ϵ^2 denoting the variance of the error of the ensemble mean,
b_* denoting the bias of the ensemble mean normalised with ϵ, and
σ_* denoting the spread-error ratio

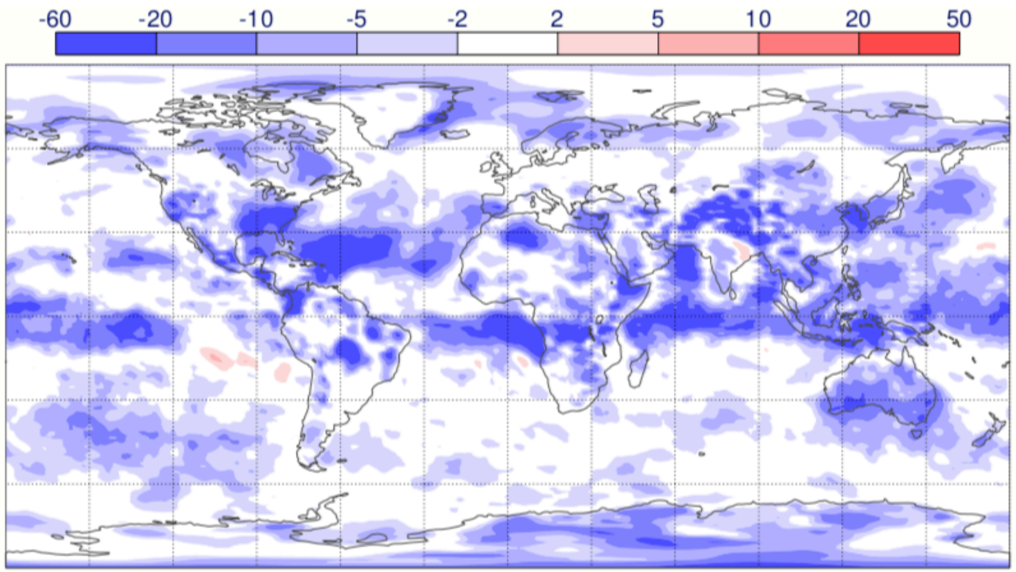
Expected CRPS as function of the spread-error ratio



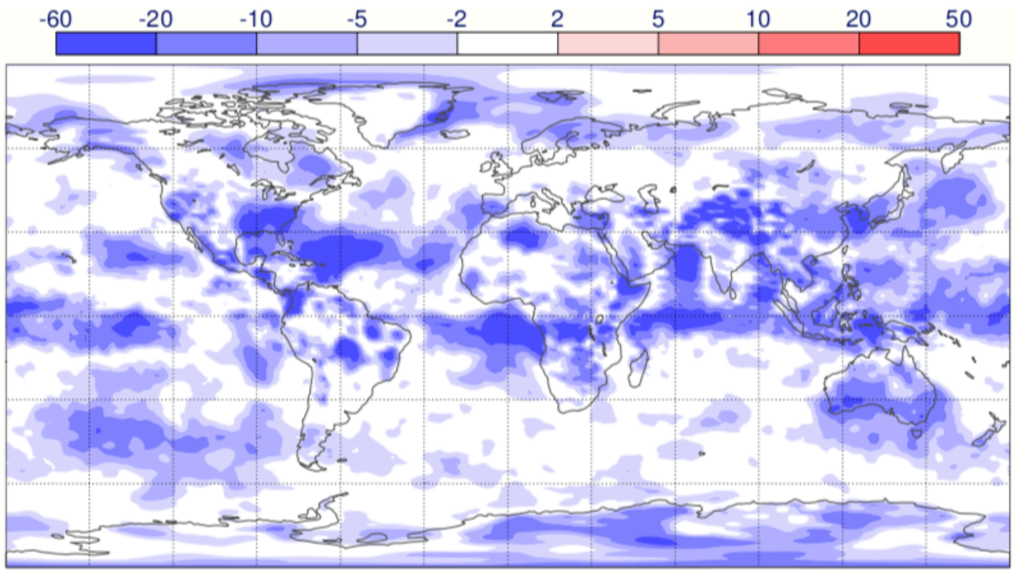
- With no forecast bias $b_* = 0$, the CRPS minimum is at a spread error ratio $\sigma_* = 1$.
- However, when the bias is significant, the CRPS minimum occurs at spread error ratios $\sigma_* > 1$.
- Should we still optimise the CRPS of raw ensemble forecasts?

CRPS change due to bias correction: T850 JJA2019

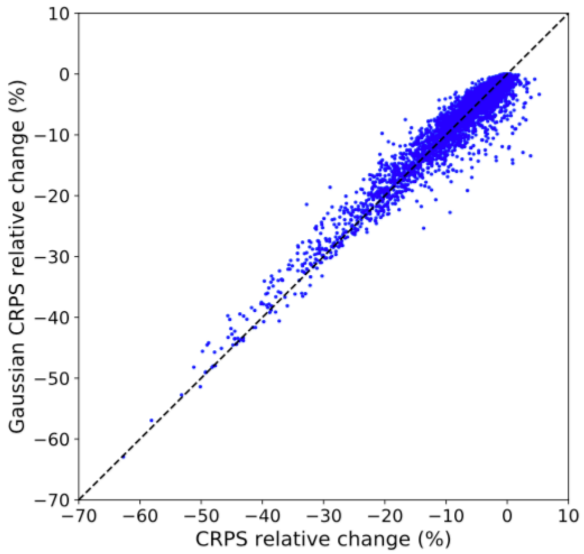
(a) ΔCRPS



(b) ΔC



(a) $r = 0.970$



The hoG approximation of the CRPS

- Provides decomposition of CRPS into reliability and resolution components
- Approximates actual CRPS well enough to yield useful diagnostics
- Diagnostics that augments information available in scorecards can be based on the hoG approximation
- Implications on development targets:
  - Maximise skill before or after bias correction?
  - Importance of accounting for obs/an uncertainties

Summary

- major revision of SPP with positive impact on ensemble skill
- development of model uncertainty representation in semi-Lagrangian advection (STOCHDP)
- resolution upgrades
  - 91 → 137 levels (pending performance tests), Reading Q2 2021
  - 18 km → 9–11 km, Bologna Q3/Q4 2022
- representation of observation uncertainties in ensemble verification
- understanding CRPS changes with a homogeneous Gaussian approximation