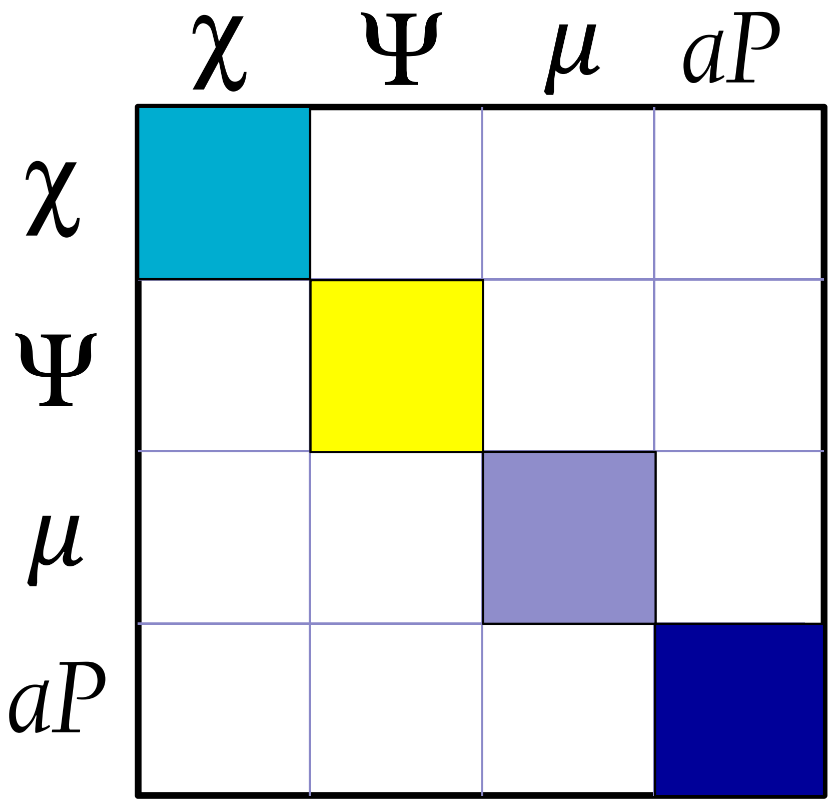
Met Office convective-scale 4DVAR system, tests and improvement

Marco Milan*, Marek Wlasak, Stefano Migliorini, Bruce Macpherson

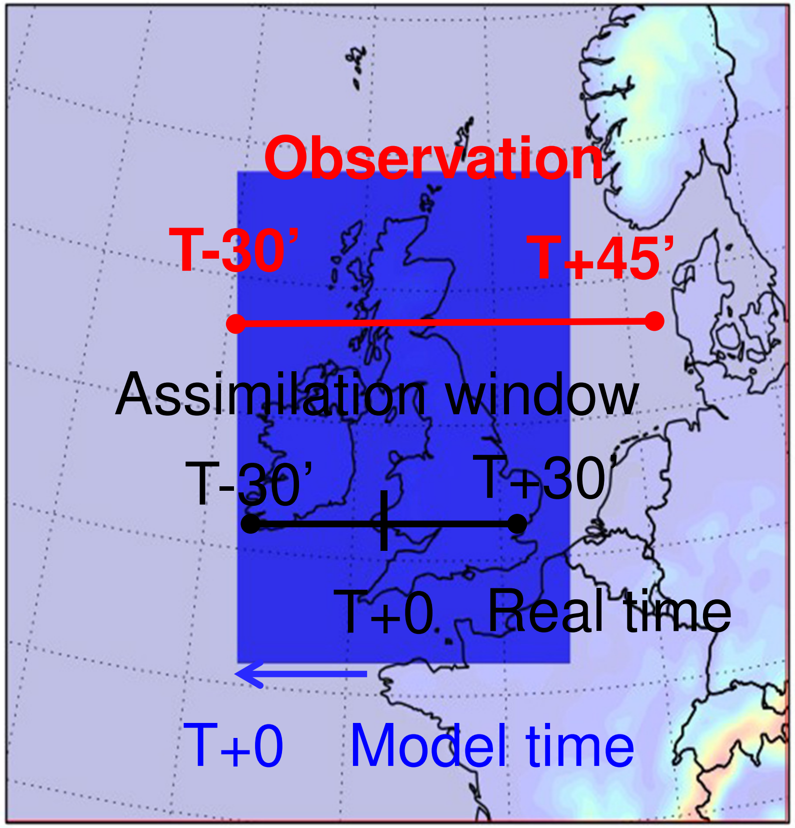
Acknowledgment: Inverarity Gordon, Gareth Dow, Mike Thurlow, Mike Cullen

OVERVIEW

- Current system, hourly 4D-Var
- The Background error covariance matrix
- NMC method
- High resolution tests
- New moisture incrementing operator
- Conclusions
- Future plans



UKV model



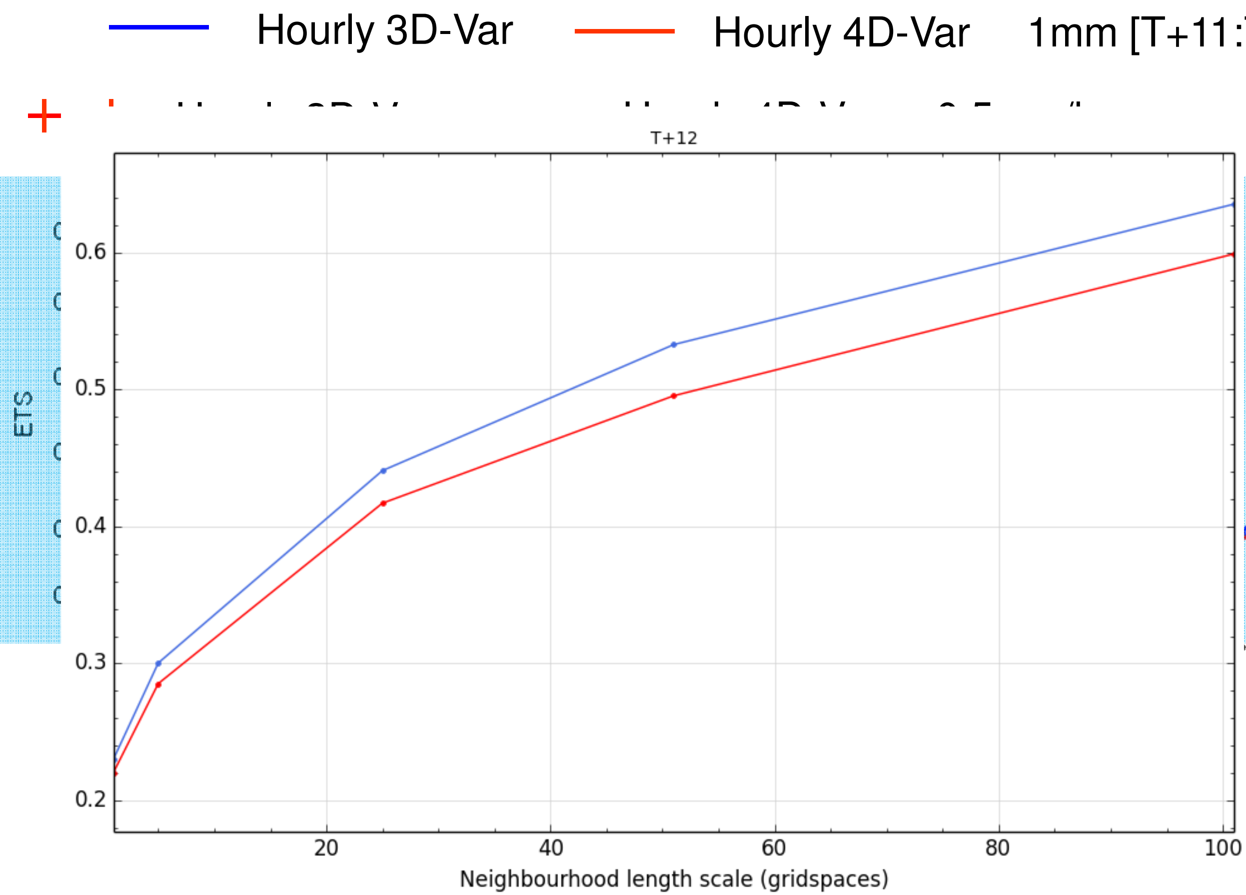
1.5km fixed resolution

1.5→4km variable resolution

- Hourly 4D-Var assimilation method, operational July 2017.
- Linear Perturbation Forecast (PF) model and DA, 4.5 km resolution.
- Variable resolution for UM model to 1.5km
- Lateral boundaries from 00, 06, 12, 18 UTC runs of 10km Global model
  - 'Age' of lbc data in UKV runs lies in range hh-3hr→hh-8hr
- Observation cut-off 45 mins.
- Vertically adaptive grid
- Initialisation by digital filter Jc term.
- Extra Observations in addition to Global.
- Introduction of VarBC.

4DVar Impact on precipitation (summer period)

3 hourly
3DVar
vs
Hourly
4DVar

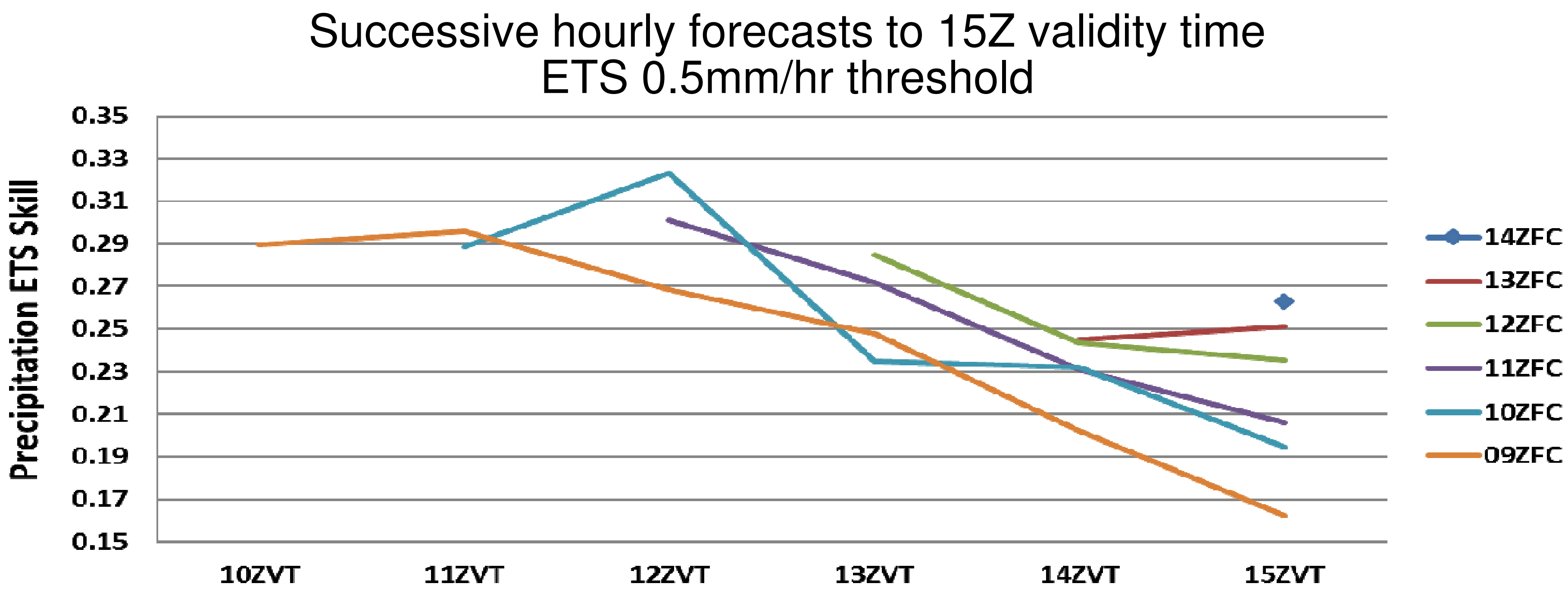


•ETS Skill

- **FSS vs Radar**
- ***Degraded rainfall by t+12, more OBS***

Courtesy of Gareth Dow

Objective Rainfall Nowcasting Benefit from hourly cycles



Meaned over ~ 6 week summer period

Courtesy of Gareth Dow

The Background error covariance matrix **B**

- In a variational data assimilation method the cost function can be written: $J=J_b+J_o+J_c$
- The J_b term in the cost function is a key factor in:
  - Driving the analysis “towards” the observation state or the background state.
  - Spatial spreading out of information from different observations, in 4D-Var also temporal.
  - Leading the analysis to maintain a state close to balance.
- **B** depends on forecast model, method of calculation and season.
- In LAM **B** is influenced also by the error coming from the boundary conditions.
- On the boundary, increments take into account the relationship between LAM and global model.

Computation of **B**

- Different methods can be used to define **B**:
  - Comparison of forecast with observations:
    - The forecast error is assumed to be uncorrelated with the observation errors.
  - Kalman Filter methods:
    - Use estimation of model error covariance matrix.
    - Propagate the forecast error using the model.
    - Issue for linear approximations.
  - NMC method:
    - The forecast error is approximated using lagged forecast differences valid at the same time.

Tests using NMC method

- Difference at same validity time (T+m)-(T+n)=Tmn
- Control run uses forecast differences from **3 hourly UKV-3DVar** T63
- First tests using T21 and T63 based on **hourly UKV-4DVar** data and T63+J_b scaling (variances as T21), gave discouraging results.
- Lesson learnt:
  - Error structure depends on the different forecast lead times used
- Start different tests using a larger sample (4 months):
  - T43, 1 hour time lag avoiding spin up problems
  - T63
  - T31, compromise between long and short time lag
  - T31, using data where the forecast starts at 00, 06, 12, 18. To have more information from OBS and large differences between forecasts

Some results

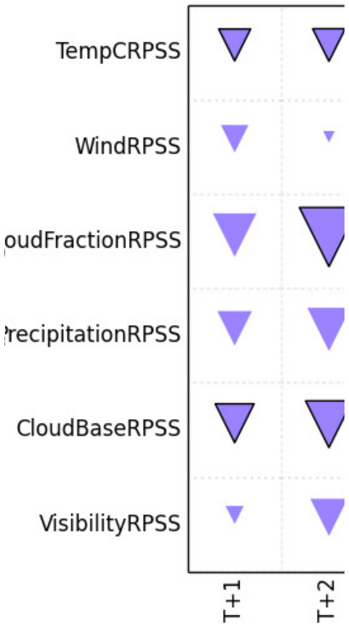
- Statistics for different Variables.
- Neighbourhood statistics over 7 x 7 points.

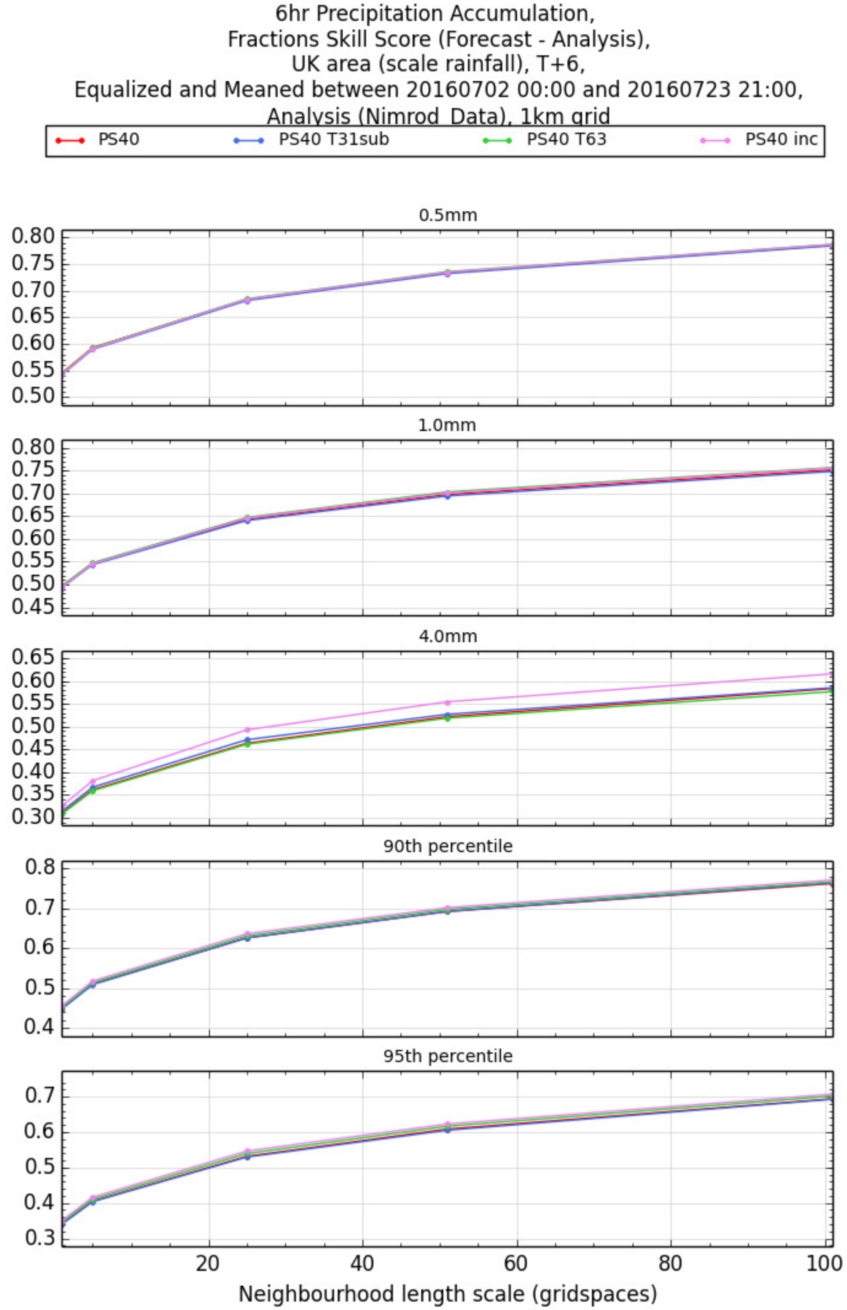
T31 Winter:

- Negative impact in many fields.
- Worst in the first 3-6 hours of forecast.
- FSS no strong differences.

T31 Summer:

- Still problem in the first 6 hours
- Cloud Fraction worst during the whole forecast.
- FSS no strong differences





A statistical based moisture incrementing operator

- Based on a linear relationship between increments in liquid cloud water content (q_{cl}'), increments in total moisture content (q_T') and temperature increments (T'). Using the background cloud fraction (C)

$$q'_{cl} = C_l(q'_T - q'_{cf} - q'_s) \cong C_l(q'_T - q'_{cf} - q_s \frac{\partial \ln e_s}{\partial T} T') \qquad q'_{cl} \cong \frac{C_l(1-C_f)}{1-C_l C_f}\left(q'_T - q_s \frac{d \ln(e_s(T))}{dT} T' \right)$$

- Similar for frozen cloud water content (q_{cf}')

$$q'_{cf} \cong \frac{C_f(1-C_l)}{1-C_l C_f}\left(q'_T - q_s \frac{d \ln(e_s(T))}{dT} T' \right)$$

- The linear regression was based on data from MOGREPS-UK. From the linear regression parameters depending on height and background relative humidity are found.

$$(q'_{cl})_i = a(\overline{rh_T}, z) \frac{\overline{C}_l(1-\overline{C}_f)}{1-\overline{C}_l \overline{C}_f}\left((q'_T)_i - \overline{q}_s \overline{\frac{\partial \ln e_s}{\partial T}} (T')_i \right)$$

Results

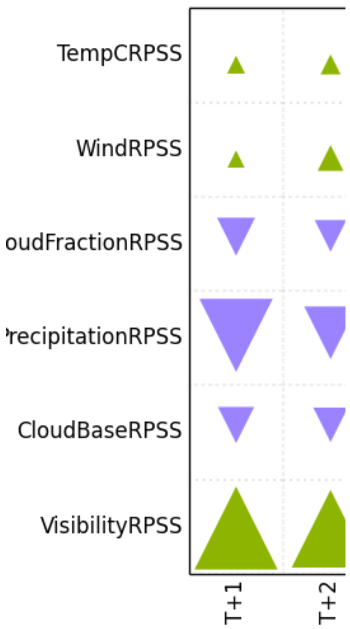
- Neighbourhood statistics over 7 x 7 points.

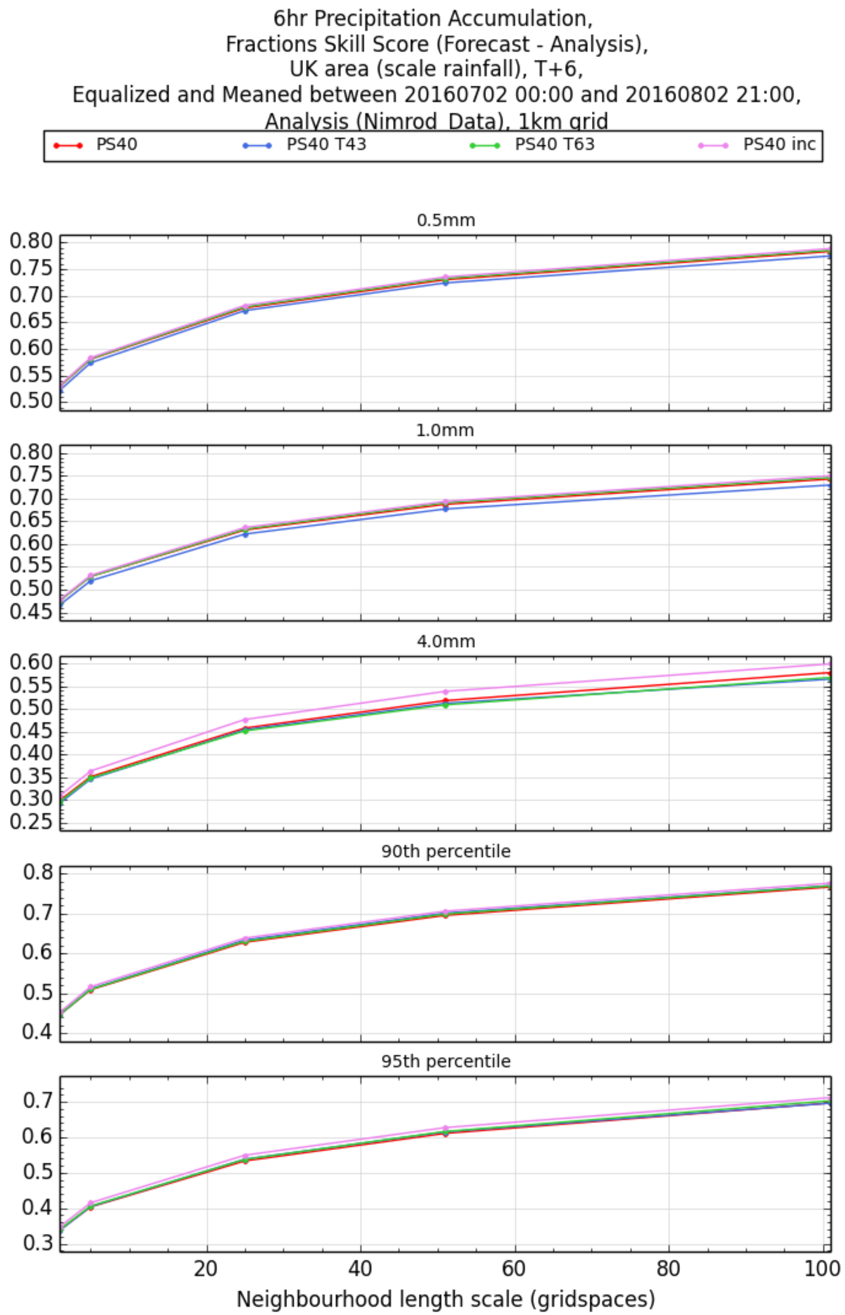
Winter:

- Positive impact for the first hours in temperature and wind.
- Some issues in Cloud.
- First 6 hours FSS comparable results.

Summer:

- Skill scores similar to winter
- Cloud forecast little better
- First 6 hours FSS comparable results





Conclusions and Future work

- UKV hourly 4D-Var operational since July 2017.
- Possible issues in the NMC method
  - In spectral space error coming from lower wavenumber depends on motions at larger scale.
  - Use of Hybrid approach related to global ensemble can estimate the error from larger motions.
  - Implement methods to change the influence of a single observation.
- We are developing a new moisture incrementing operator
  - We plan to compute the coefficients using a larger sample.
  - The Cloud fraction is now given from the background, an outer loop could be helpful.
  - Future work includes the addition of cloud fraction increments in the moisture incrementing operator

QUESTIONS?

References

- Migliorini, S., Lorenc, A., and Bell, W. (2017), A moisture incrementing operator for the assimilation of humidity- and cloud-sensitive observations. Q.J.R. Meteorol. Soc., under review.

- Rawlins, F., S. Ballard, K. Bovis, C. A.M., D. Li, I. G.W., A. Lorenc, and T. Payne (2007). The Met Office global four-dimensional variational data assimilation scheme. Q.J.R. Meteorol. Soc. 133, 347–362

Questions

- Score used in HIRA
- Incremental 4D-Var
- The NMC method assumptions and limits
- T43, T63, T63inc
- Increments in the new moisture incrementing operator

Score used in HIRA

- MAE: Average of absolute differences between forecasts and observations.

$$\langle |f - O| \rangle$$

An event occurred in the OBS is set to 1 and to 0 if not. In HIRA the probability is related to the frequency in the neighbour points which are lower than a threshold (neighbourhood probability).

- FSS: Fraction Skill Score, use the neighbourhood probabilities with *m* neighbourhood. Bsed on FBS (Fraction Brier Score)

$$FBS = \frac{1}{N}\sum_{i=1}^{N}\left[P_f - P_O\right]^2,\ FBS_{worst} = \frac{1}{N}\left[\sum_{i=1}^{N} P_f^2 + \sum_{i=1}^{N} P_O^2\right]$$

$$FSS = 1 - \frac{FBS}{FBS_{worst}}$$

N grid boxes

Score used in HIRA

In HIRA the probabilities are based on the neighbourhood probability.

- RPS: Ranked Probability Score, mean square error of probabilistic multi-category forecasts. K categories, using CDFs (P) with an Heaviside function for OBS:

$$RPS = \sum_{k=1}^{K} (P_f(k) - P_O(k))$$

- CRPS: Continuous Ranked Probability Score

$$\int_{-\infty}^{+\infty} \left(P_f(x) - P_O(x)\right)^2 dx$$

Incremental 4D-VAR

- Based on the formulation of Rawlins et al. 2007

$$
\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^g + \delta\mathbf{x} \\
\delta\mathbf{w} &= S(\mathbf{x}^g + \delta\mathbf{x}) - S(\mathbf{x}^g) \simeq \mathbf{S}\delta\mathbf{x} \\
\delta\mathbf{w}^b &= S(\mathbf{x}^b) - S(\mathbf{x}^g)
\end{aligned}
$$

- g, first guess; a, analysis; b, background; o observations
- *S* is a non-linear simplification operator with tangent linear approximation **S**
- 4DVAR Cost function, using the simplified increments (notation avoids sums)

$$J(\delta\mathbf{w}) = \frac{1}{2}(\delta\mathbf{w} - \delta\mathbf{w}^b)^T\mathbf{B}^{-1}(\delta\mathbf{w} - \delta\mathbf{w}^b) + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}^o})^T\underline{\mathbf{R}^{-1}}(\underline{\mathbf{y}} - \underline{\mathbf{y}^o})$$

Incremental 4D-VAR

- In the minimization a CVT (Control Variable Transform) is used.

- The **B** become an Identity

- New variable using CVT

$$\delta\mathbf{w} = \mathbf{U}\mathbf{v} = \mathbf{U}_p\mathbf{U}_a\mathbf{U}_h\mathbf{U}_v\mathbf{v}$$

$$J_b = \frac{1}{2}(\mathbf{v} - \mathbf{v}^b)^T(\mathbf{v} - \mathbf{v}^b)$$

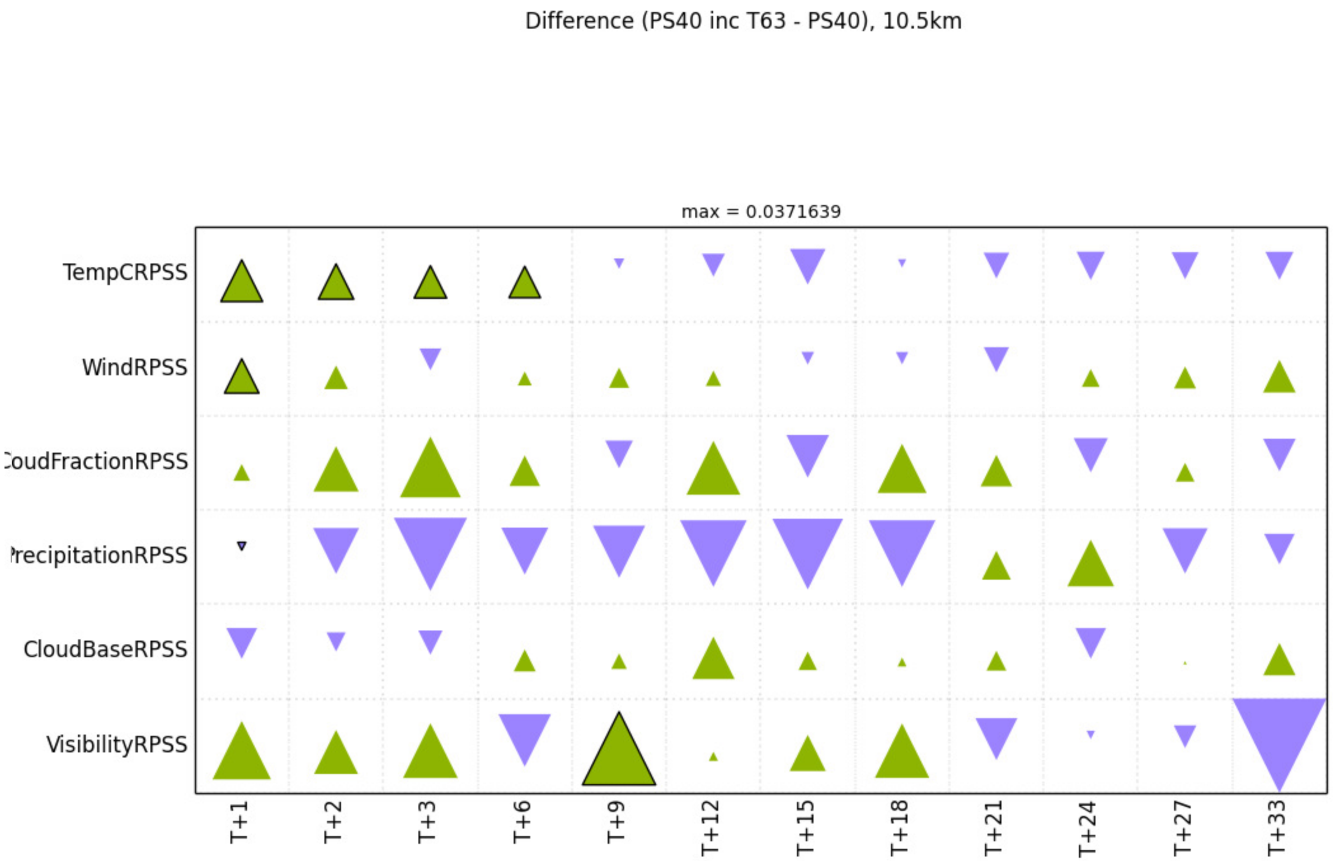
$$J_o = \frac{1}{2}(\underline{\mathbf{y}} - H(\mathbf{x}^g) - \mathbf{H}\mathbf{U}\underline{\mathbf{v}})^T\underline{\mathbf{R}^{-1}}(\underline{\mathbf{y}} - H(\mathbf{x}^g) - \mathbf{H}\mathbf{U}\underline{\mathbf{v}})$$

- $\mathbf{U}_a$, is the vertically adaptive grid transform (AG; Piccolo and Cullen, 2011)

The NMC method assumptions and limits

- When the differences between the forecasts are small the NMC method underestimate variances. The analysis will be less influenced from the observations.
- The forecasts used to compute the differences are assumed uncorrelated.
- Leads to a climatological approximation of the covariances. The error due to the synoptic case is not taken into account.
- Large scale atmospheric states evolve with LBC.
- For LAM to reduce the influence due to LBC, forecast differences are based on forecast using the same LBC.

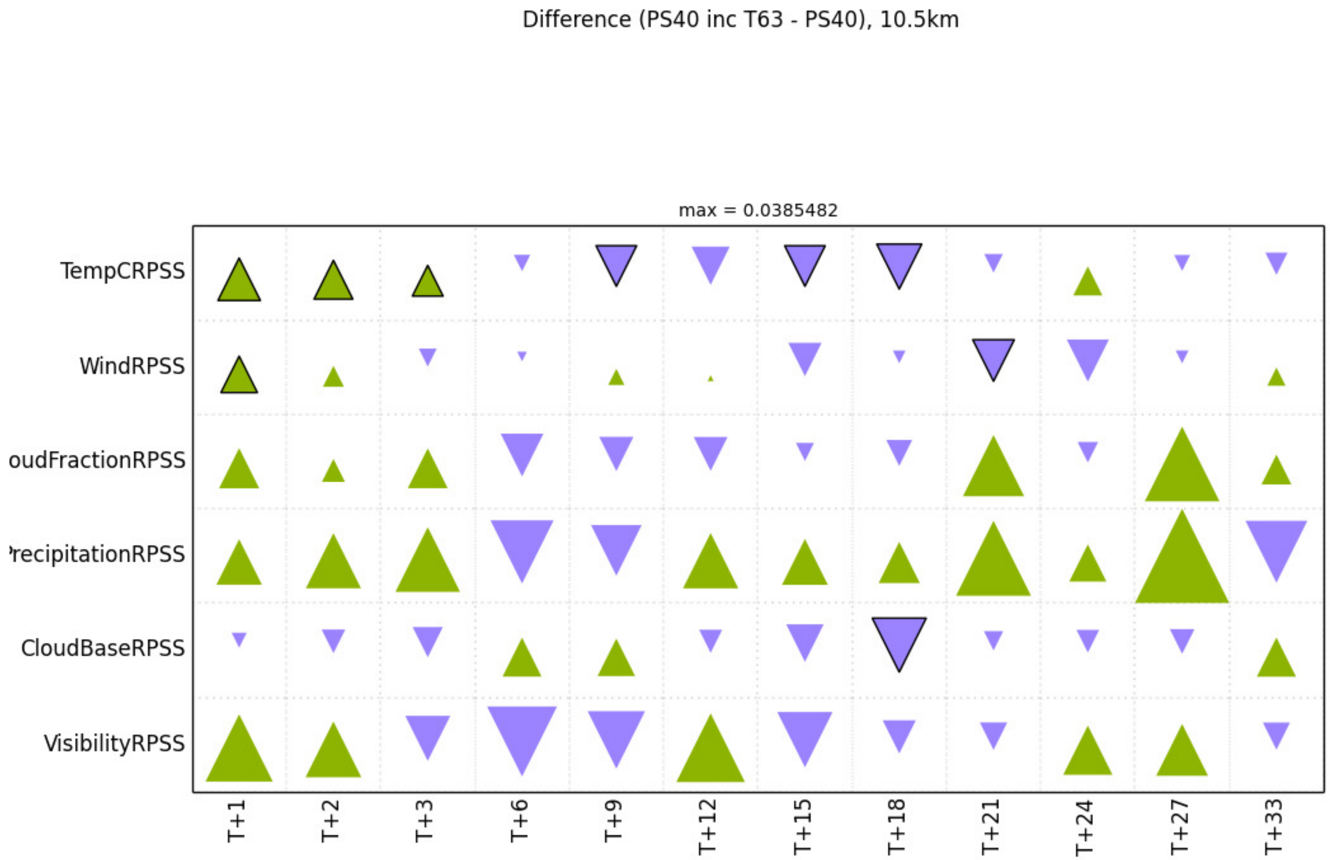
Further results (winter)



.

- T43: poor skill in all fields
- T63: improves many fields, especially surface temperature at beginning of forecast.
- T63inc: TempCRPSS positive at beginning of forecast. Precipitation RPSS at T+1 stat. significantly negative.

Further results (summer)

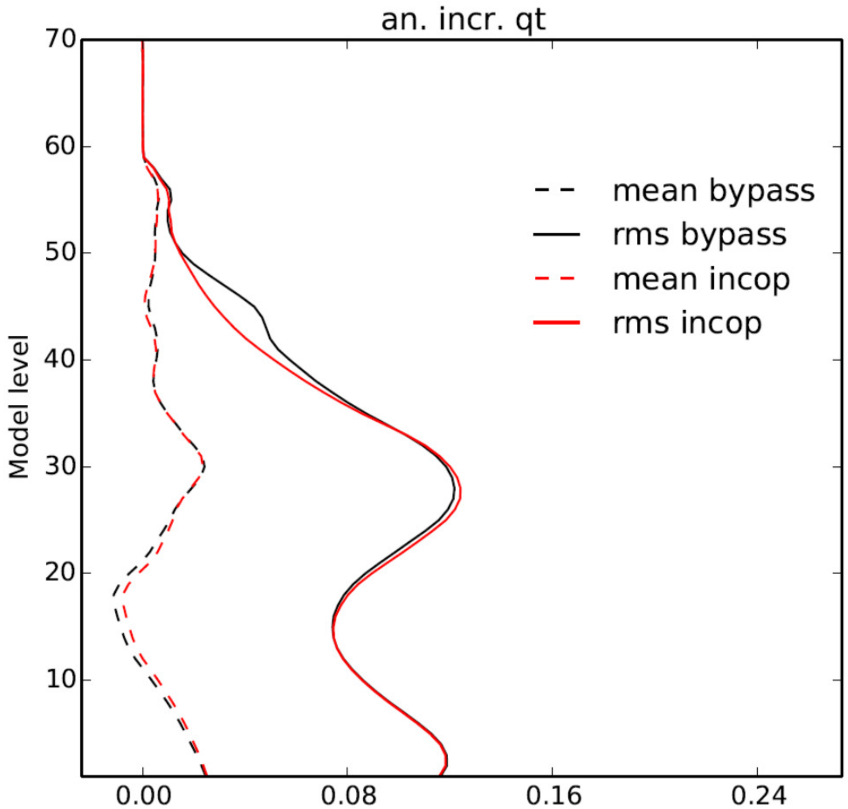


.

- T43: better results than winter. Still problems in surface temperature and clouds.
- T63: Generally positive impact.
- T63inc: Surface temperature better until T+6, after worst. Other variables quite neutral impact.

Increments in new moisture incrementing operator

summer



winter