

# Ensuring robust verification procedures for model change upgrades

- How can we be sure that a model change package is a positive upgrade ?
- Which tests should be done ?
- How should high resolution forecasts be assessed ?
  - What verification scores ?
  - How long trials ?
  - Subjective assessment role ?
- How to make sure the process is robust ?
- Area of concern and uncertainty

## Met Office **Current pre-operational tests**

*100 case studies*

*Chosen from forecaster assessment/classification by weather types*

*Objective scores – RMSE, HiRA, FSS*

*Subjective as time allows*

*2 seasons with full DA ( winter/summer)*

*Objective scores*

*Subjective as time allows*

*Parallel suite*

*Objective scores*

*Subjective as time allows*

Cases	
Precipitation frontal	19
Showers	22
Snow	7
Winds and Gales	6
Thunderstorms	3
Cloud (general)	9
Stratus	7
Stratocumulus	9
Temperature	6
Frost	4
Anticyclones and clear	15

# Work plan:

- Review process and interview test owners
- Consolidate references / work in progress on FSS (MM) and verification methods
- Simple synthetic cases to understand basic questions:
  - Ideal length of trials
  - Different scores/verification for strictly controlled change (same inputs)
- Basic properties/behaviour of scores
- Pseudo-station obs from radar in HiRA to contrast with FSS
- Alternative FSS formulations:
- Controlled perturbations to forecast/obs to highlight contrasting scores