Site-specific post-processing at the Met Office

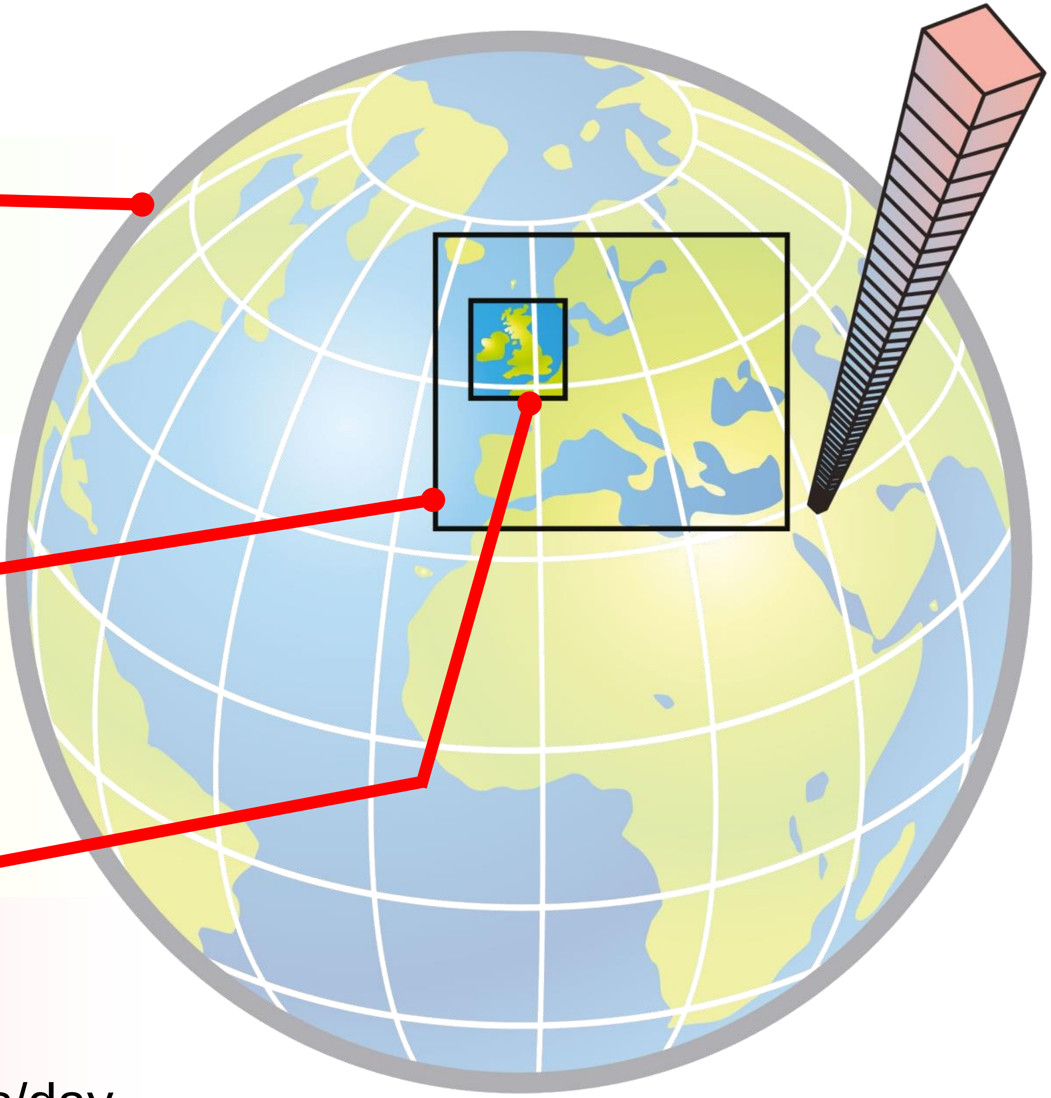
Nina Schuhen 12 February 2015

Outline

- The Met Office NWP model suite
- BestData and blended forecasts
- Calibration of MOGREPS-UK forecasts
- Outlook and further research

The Met Office NWP model suite

Deterministic NWP model suite



Global

- 17km 70 Levels
- 48 hour forecast twice/day
- 6 day forecast twice/day

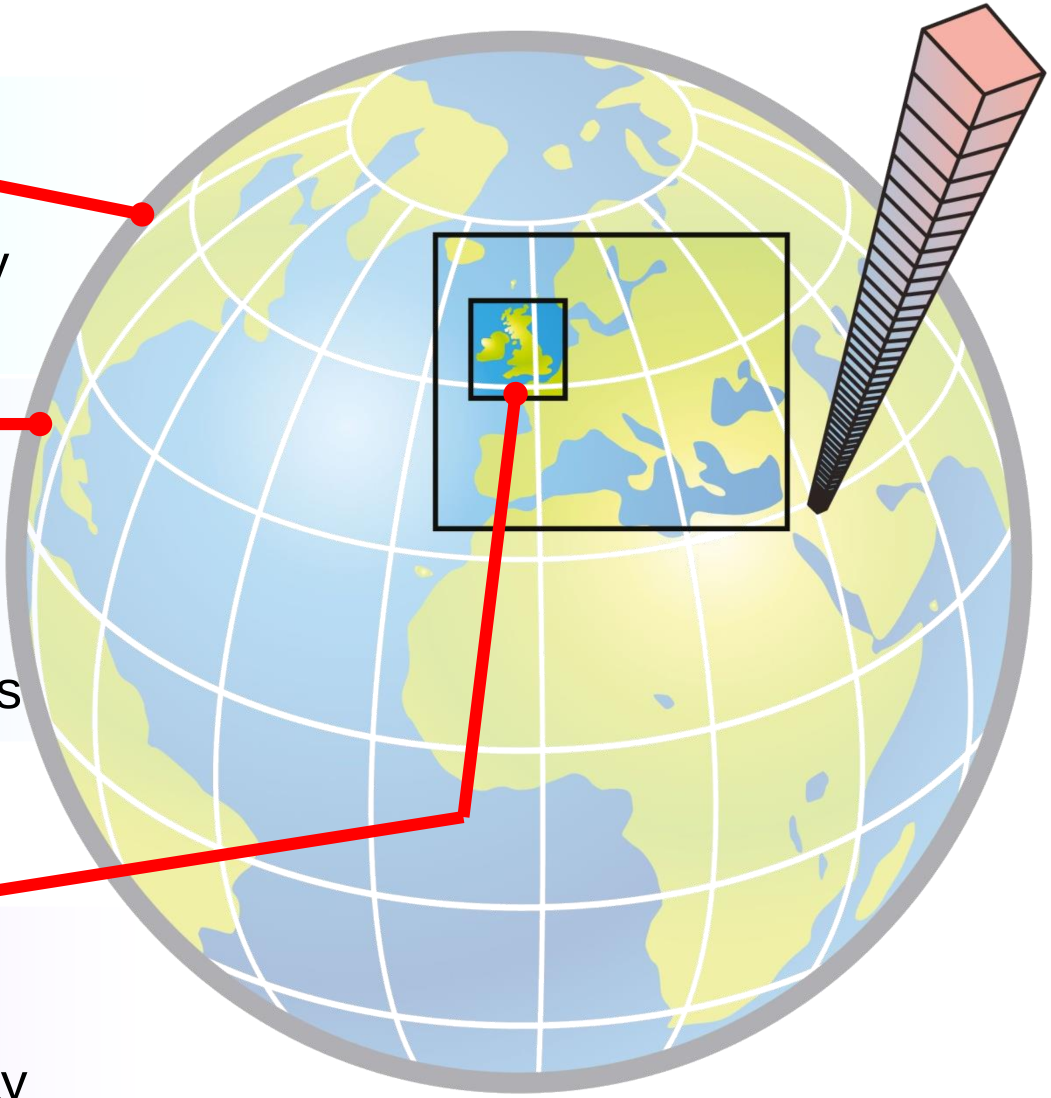
Euro4

- 4.4km 70 Levels
- 60 hour forecast twice/day
- 5 day forecast twice/day

UKV

- 1.5km 70 Levels
- 36 hour forecast eight times/day

Ensemble NWP model suite



MOGREPS-15

- 60km 70 Levels
- 15 day forecast 2 times/day
- 24 members

MOGREPS-G

- 33km 70 Levels
- 7 day forecast 4 times/day
- 12 members
- 24 member lagged products

MOGREPS-UK

- 2.2km 70 Levels
- 36 hour forecast 4 times/day
- 12 members

Met Office

Ensemble NWP model suite

~~MOGREPS-15~~

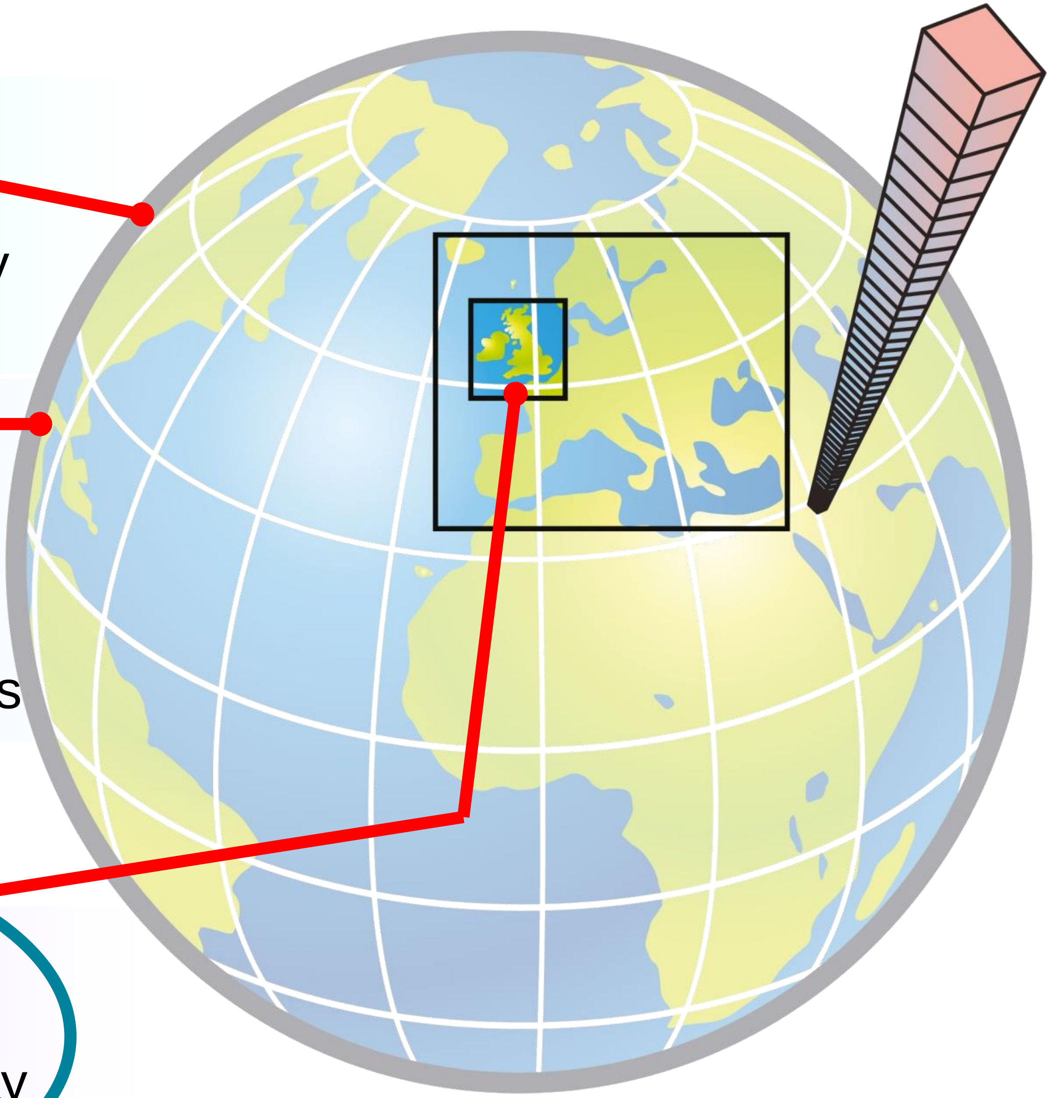
- ~~60k [illegible] els~~
- ~~15 [illegible] t 2 times/day~~
- ~~24 [illegible] embers~~

MOGREPS-G

- 33km 70 Levels
- 7 day forecast 4 times/day
- 12 members
- 24 member lagged products

MOGREPS-UK

- 2.2km 70 Levels
- 36 hour forecast 4 times/day
- 12 members



Blending

Making use of multiple models

BestData

- Many customer products, using many single model feeds
  - Products tailored to that feed
  - Hard to migrate or incorporate new models

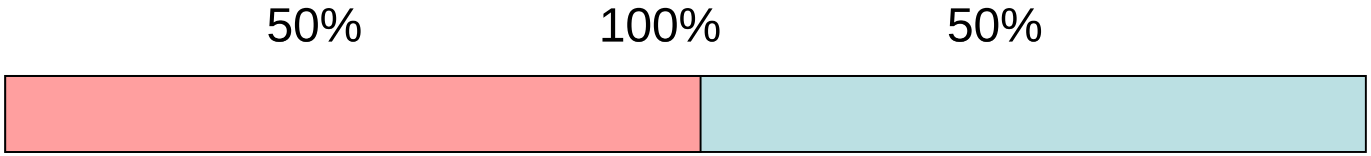
- Solution: Blend of all available models at a given time
  - Interface between models and products
  - Single data feed to downstream systems
  - Allows for model changes without changing infrastructure

- Forecasts are Kalman-filtered prior to blending

Blending

- At any given time and any given forecast horizon, we want to make effective use of all available models
- ~ 2500 single predictions available for a short-range forecast
- Includes the latest nowcast and ECMWF ensemble from 15 days ago
- Can’t create a true multi-model ensemble, as it is impossible to keep all model data on-line
- Solution: Recursively applied weighted average

Blending Example

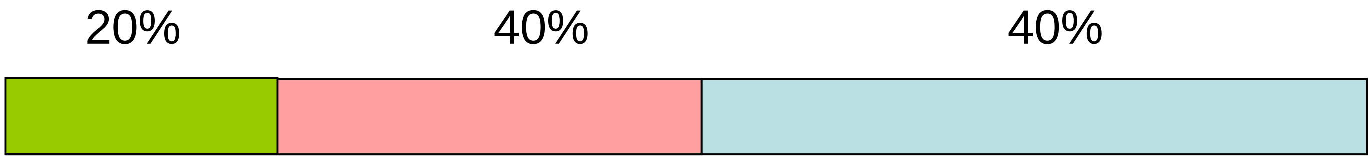


50%

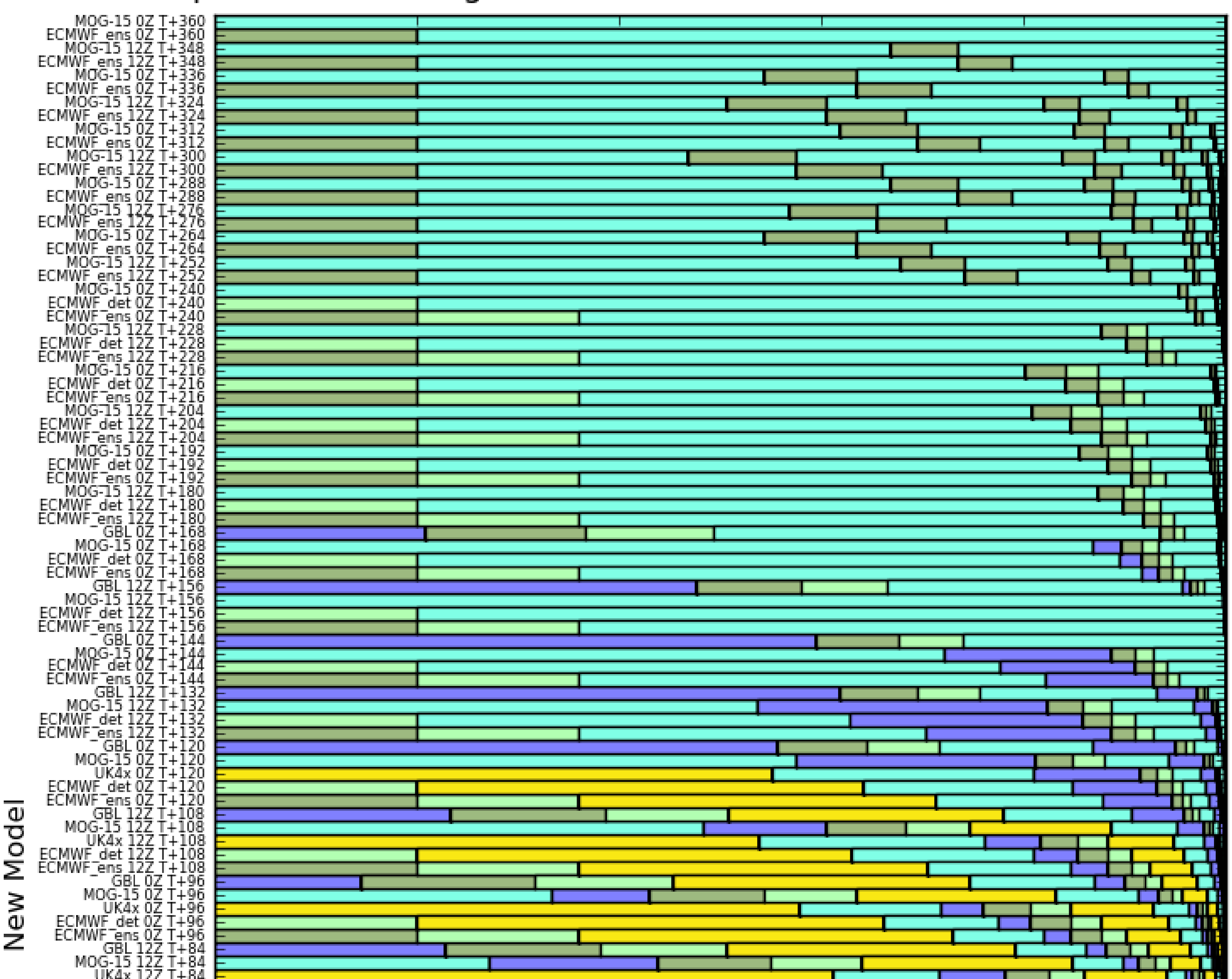
100%

50%

Blending Example



Blending Example


MOG-15 0Z T+360
ECMWF_ens 0Z T+360
MOG-15 12Z T+348
ECMWF_ens 12Z T+348
MOG-15 0Z T+336
ECMWF_ens 0Z T+336
MOG-15 12Z T+324
ECMWF_ens 12Z T+324
MOG-15 0Z T+312
ECMWF_ens 0Z T+312
MOG-15 12Z T+300
ECMWF_ens 12Z T+300
MOG-15 0Z T+288
ECMWF_ens 0Z T+288
MOG-15 12Z T+276
ECMWF_ens 12Z T+276
MOG-15 0Z T+264
ECMWF_ens 0Z T+264
MOG-15 12Z T+252
ECMWF_ens 12Z T+252
MOG-15 0Z T+240
ECMWF_det 0Z T+240
ECMWF_ens 0Z T+240
MOG-15 12Z T+228
ECMWF_det 12Z T+228
ECMWF_ens 12Z T+228
MOG-15 0Z T+216
ECMWF_det 0Z T+216
ECMWF_ens 0Z T+216
MOG-15 12Z T+204
ECMWF_det 12Z T+204
ECMWF_ens 12Z T+204
MOG-15 0Z T+192
ECMWF_det 0Z T+192
ECMWF_ens 0Z T+192
MOG-15 12Z T+180
ECMWF_det 12Z T+180
ECMWF_ens 12Z T+180
GBL 0Z T+168
MOG-15 0Z T+168
ECMWF_det 0Z T+168
ECMWF_ens 0Z T+168
GBL 12Z T+156
MOG-15 12Z T+156
ECMWF_det 12Z T+156
ECMWF_ens 12Z T+156
GBL 0Z T+144
MOG-15 0Z T+144
ECMWF_det 0Z T+144
ECMWF_ens 0Z T+144
GBL 12Z T+132
MOG-15 12Z T+132
ECMWF_det 12Z T+132
ECMWF_ens 12Z T+132
GBL 0Z T+120
MOG-15 0Z T+120
UK4x 0Z T+120
ECMWF_det 0Z T+120
ECMWF_ens 0Z T+120
GBL 12Z T+108
MOG-15 12Z T+108
UK4x 12Z T+108
ECMWF_det 12Z T+108
ECMWF_ens 12Z T+108
GBL 0Z T+96
MOG-15 0Z T+96
UK4x 0Z T+96
ECMWF_det 0Z T+96
ECMWF_ens 0Z T+96
GBL 12Z T+84
MOG-15 12Z T+84
UK4x 12Z T+84
New Model

Calibrating MOGREPS-UK

With selected results

Overview of our plans

- MOGREPS-UK initially, MOGREPS-G and ECMWF ensemble later
- Three main parameters:
  - surface temperature
  - 10m wind speed
  - hourly precipitation accumulation
- Calibrate and verify for 149 observation sites over the UK and Ireland
- Compare EMOS and BMA approaches

Surface temperature

- Ensemble Model Output Statistics:
  - Gaussian distribution with incorporated ensemble information

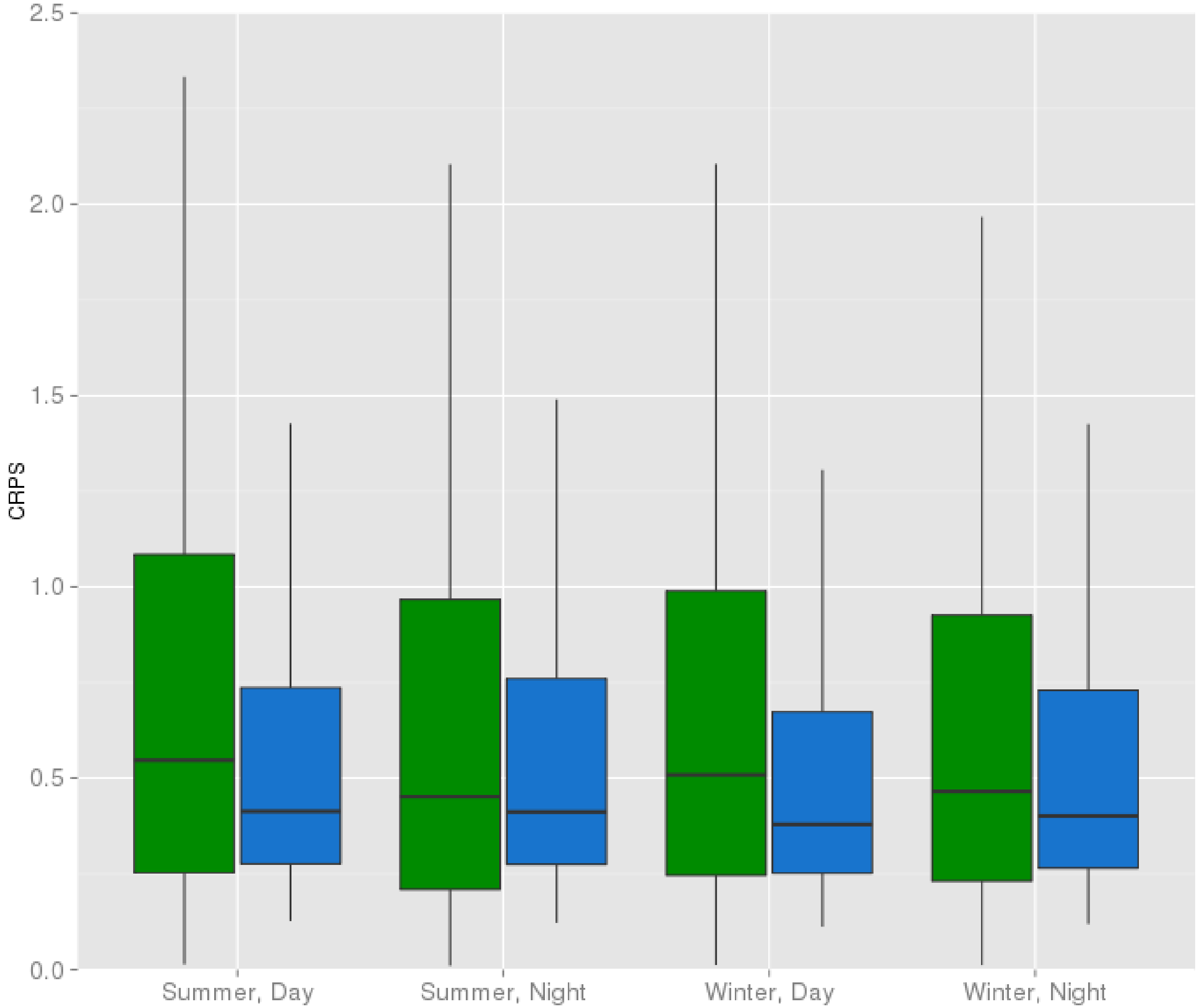
$$Y \mid X_1, \ldots, X_{12} \sim \mathcal{N}\left(a + \beta^2 \cdot \overline{X},\, \gamma^2 + \delta^2 \cdot S^2\right)$$

  - Minimum CRPS estimation
- Bayesian Model Averaging:
  - Weighted sum of Gaussian distributions centered around the bias-corrected ensemble members

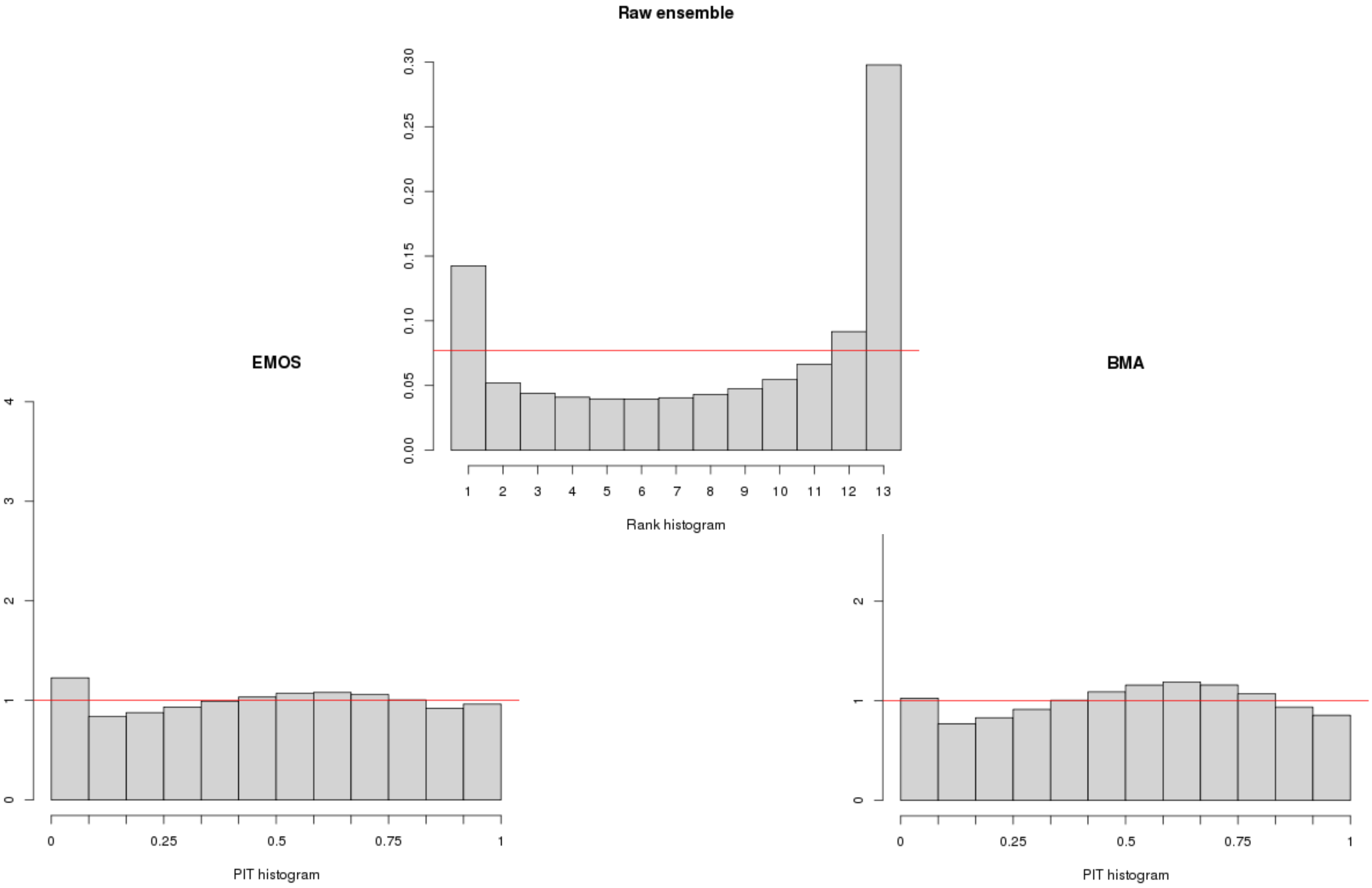
$$P\left(Y \mid X_1, \ldots, X_{12}\right) = \sum_{m=1}^{12} w_m \cdot \mathcal{N}\left(a_m + b_m \cdot X_m,\, \sigma^2\right)$$

  - Linear regression and maximum likelihood (EM algorithm)

T2M: CRPS distribution

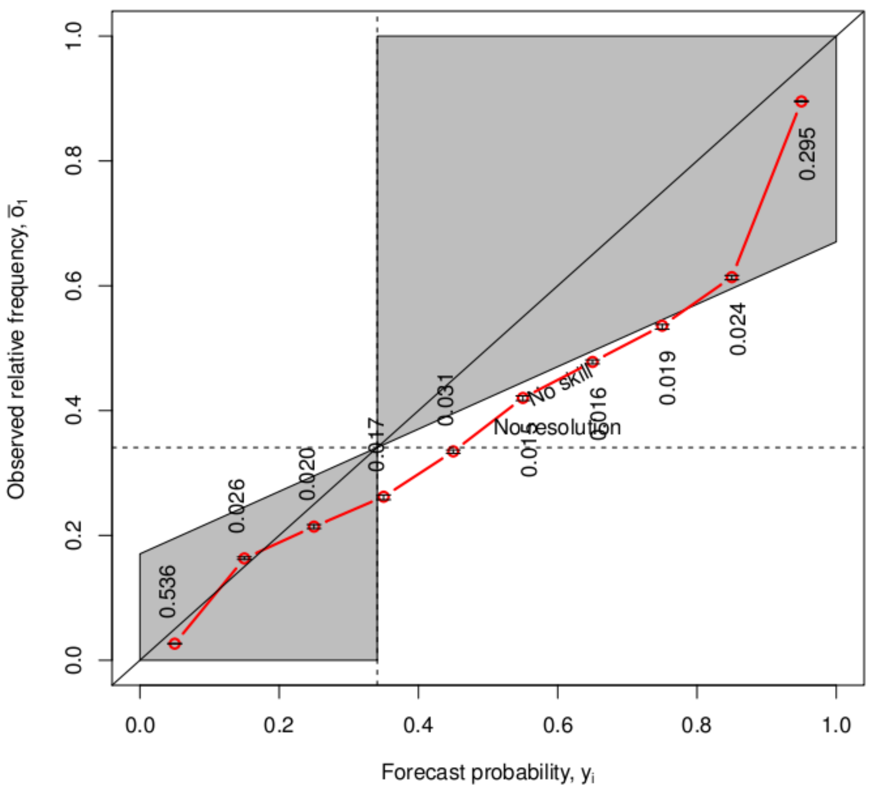


T2M: PIT / rank histograms

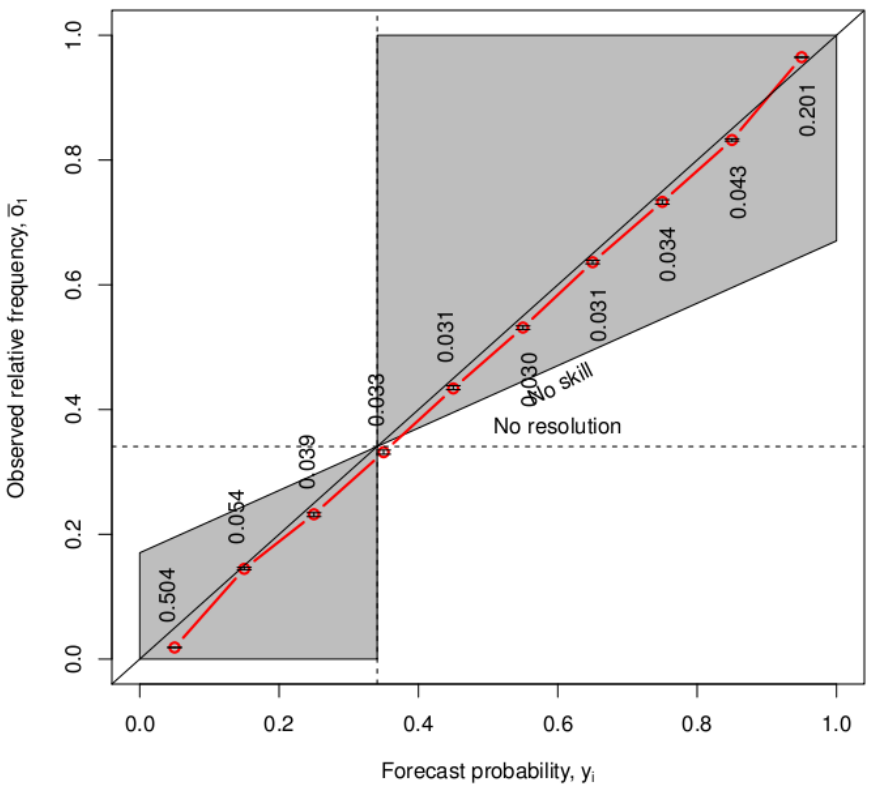


T2M: Cold winter nights

Ensemble, Threshold 5



Calibrated, Threshold 5



10m wind speed

- Ensemble Model Output Statistics:
  - Truncated Gaussian distribution (cut-off at 0) with incorporated ensemble information

$$Y \mid X_1, \ldots, X_{12} \sim \mathcal{N}^0 \left(a + \beta^2 \cdot \overline{X}, \, \gamma^2 + \delta^2 \cdot S^2\right)$$

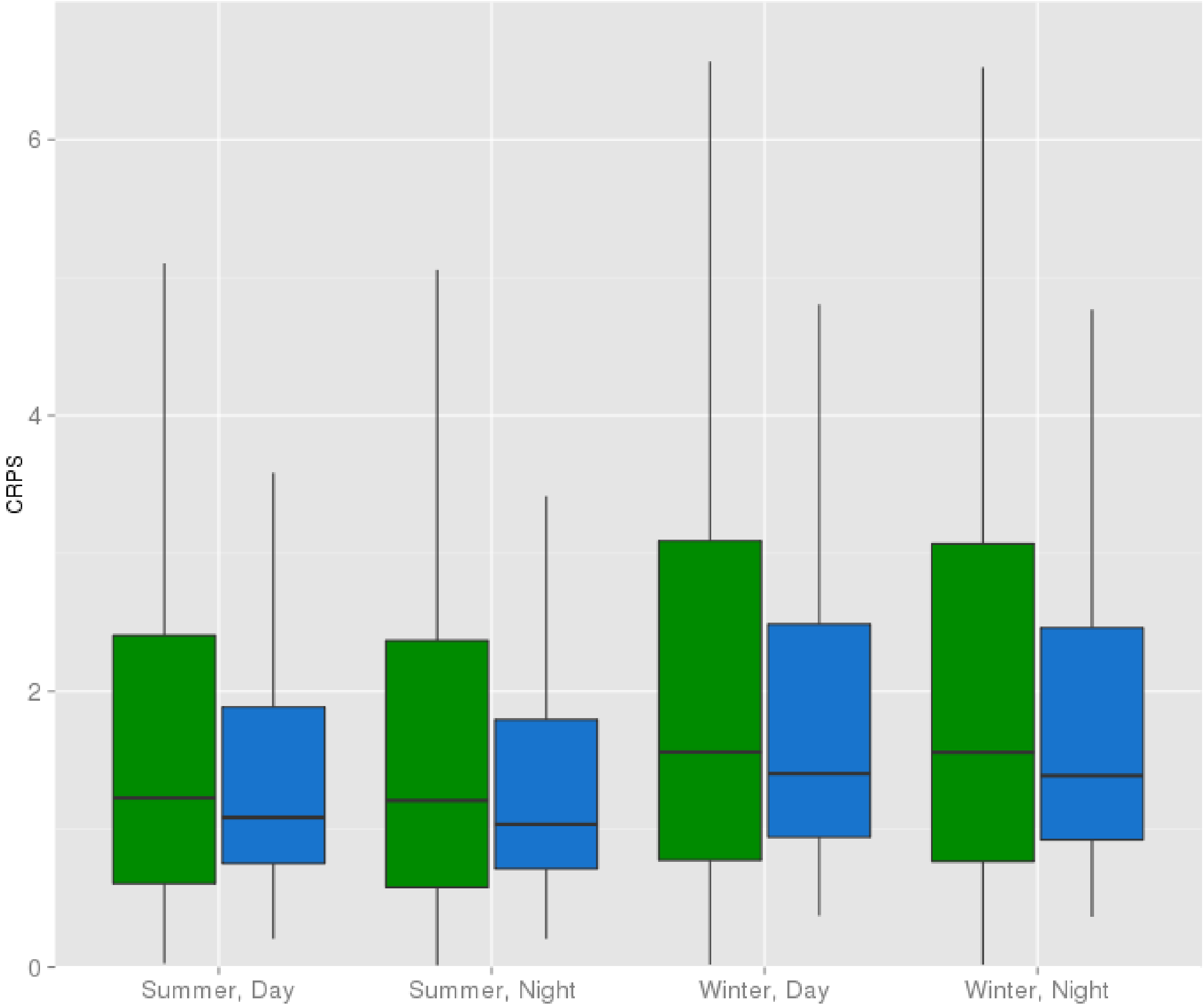
  - Minimum CRPS estimation
- Bayesian Model Averaging:
  - Weighted sum of gamma distributions centered around the bias-corrected ensemble members

$$P\left(Y \mid X_1, \ldots, X_{12}\right) = \sum_{m=1}^{12} w_m \cdot \gamma\left(\alpha_m, \, \beta_m\right)$$

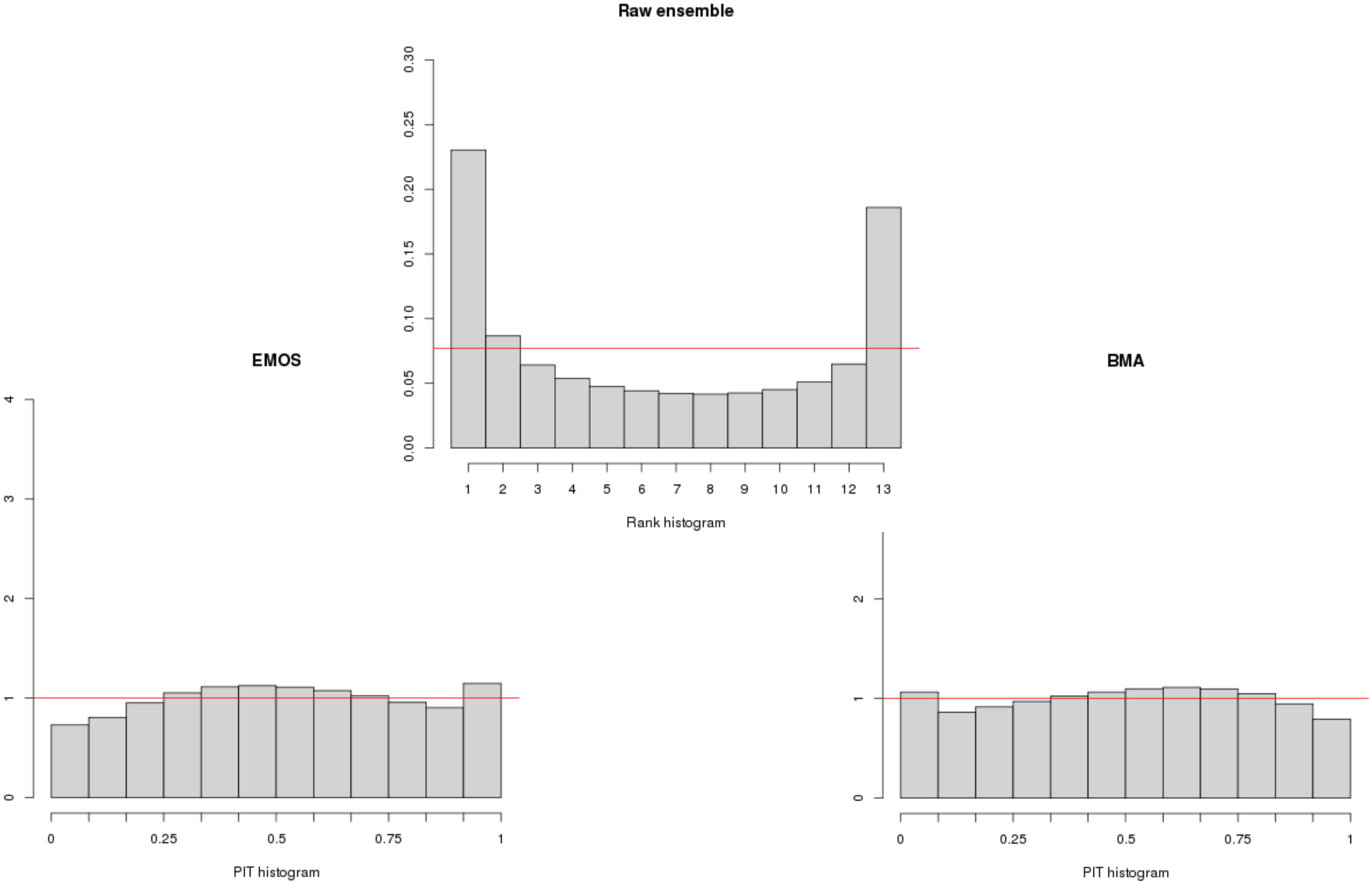
$$\mu_m = \frac{\alpha_m}{\beta_m} = b_{0m} + b_{1m} \cdot X_m \qquad \sigma^2 = \frac{\alpha_m}{\beta_m^2} = c_0 + c_1 \cdot X_m$$

  - Linear regression and maximum likelihood (EM algorithm)

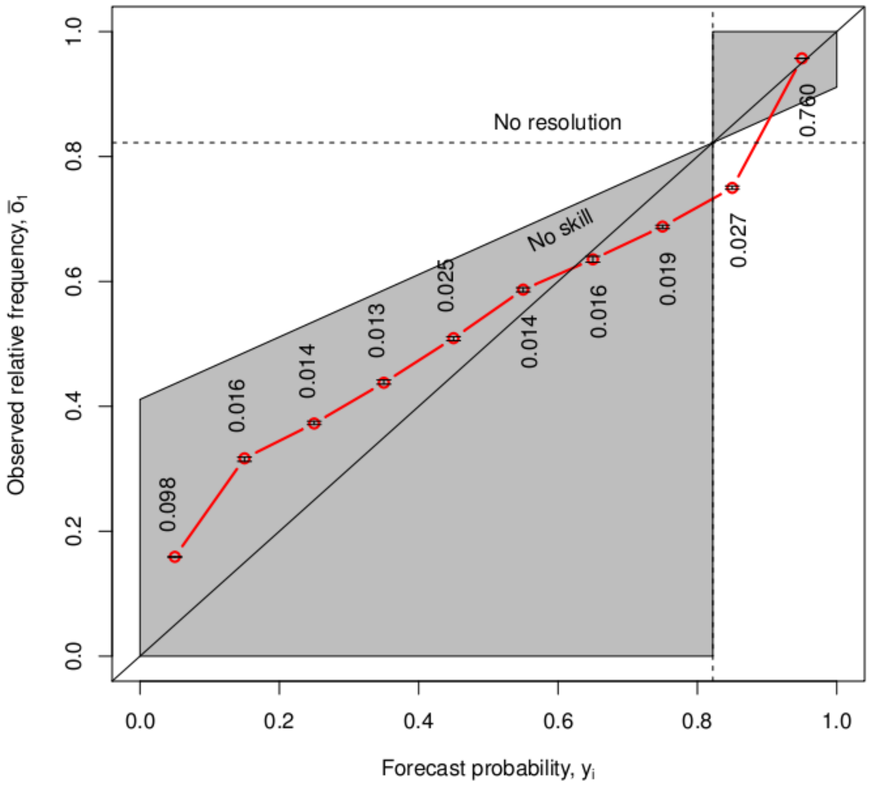
10mWS: CRPS distribution

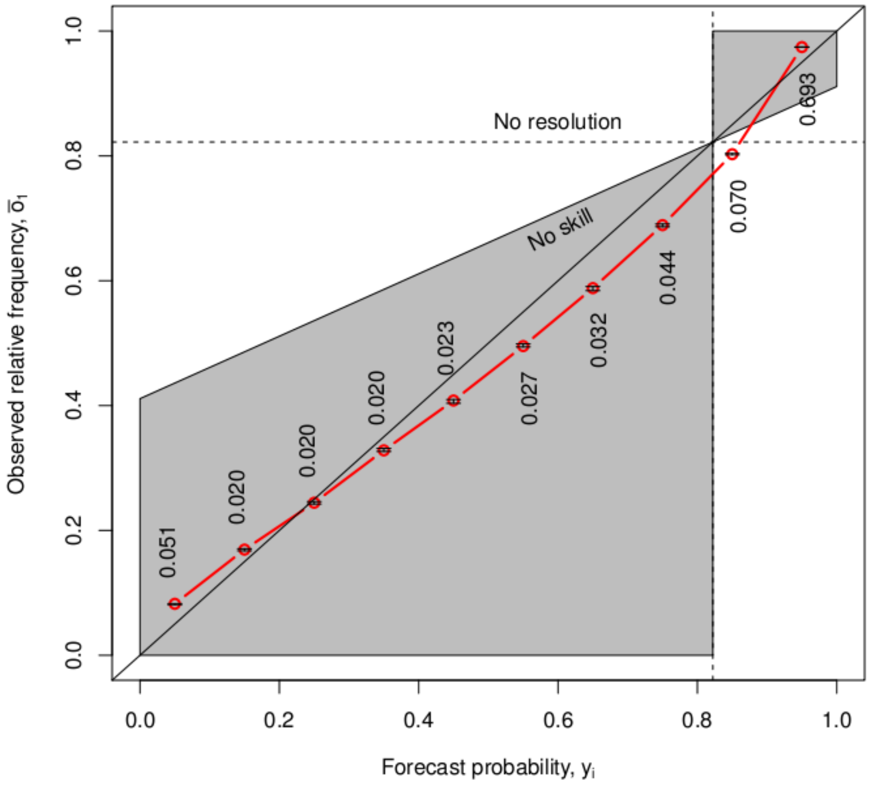


10mWS: PIT / rank histograms



10mWS: Daytime winds in winter


Ensemble, Threshold 19


Calibrated, Threshold 19

EMOS for precipitation: Variant 1

- Generalised extreme value distribution family:

$$Y \mid X_1, \ldots, X_{12} \sim \mathcal{GEV}\left(\mu,\, \sigma,\, \xi\right)$$

  - Location μ, scale σ and shape $\xi = 0.2$
  - Mean $m = a_0 + a_1 \cdot \overline{X} + a_2 \cdot \overline{\mathbb{1}_{\{X=0\}}}$
  - Standard deviation $\sigma = b_0 + b_1 \cdot \mathrm{MD}\left(X\right)$ with ensemble mean difference

$$\mathrm{MD}\left(X\right) = \frac{1}{M^2} \sum_{i,j=1}^{M} \left|X_i - X_i\right|$$

- Minimum CRPS estimation, but can be very unstable depending on the choice of ξ

Scheuerer, M. (2014) Probabilistic quantitative precipitation forecasting using Ensemble Model Output Statistics. Q. J. R. Meteorol. Soc., 140, 1086–1096

EMOS for precipitation: Variant 2

- Gamma distribution with a discrete component in zero

$$Y \mid X_1, \ldots, X_{12} \sim p_0 \cdot \mathbb{1}_{\{Y=0\}} + (1 - p_0) \cdot \Gamma(\alpha, \beta) \cdot \mathbb{1}_{\{Y>0\}}$$

  - PONP $$\log\left(\frac{p_0}{1 - p_0}\right) = a_1 + a_2 \cdot \overline{X}^{1/3} + a_3 \cdot \frac{1}{12} \sum_{k=1}^{12} \mathbb{1}_{\{X_k=0\}}$$
  - Mean $$\mu = \frac{\alpha}{\beta} = b_1 + b_2 \cdot \overline{X}^2$$
  - Variance $$\sigma^2 = \frac{\alpha}{\beta^2} = c + d \cdot \left(S^2\right)^2$$

- Hourly accumulation: power transformation 2

Further research

- Optimise temperature and wind speed methods in terms of predictors and training data
- Investigate grouping sites according to type
- Use Kalman-filtered data to calibrate
- Precipitation:
    - “Increase” discrete component?
    - Look into power transformation
    - Different view on verification?
- Apply to MOGREPS-G and ECMWF
- Trial BMA to blend models

Questions and discussion

What goes into BestData?

Currently

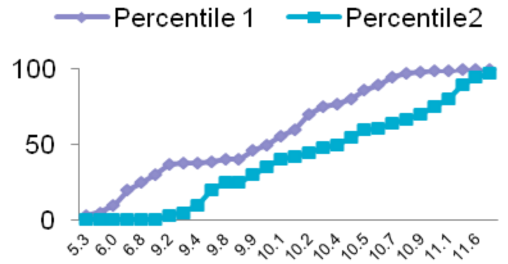
- ECMWF ensemble (00Z&12Z) to T+360
- ECMWF deterministic (00Z&12Z) to T+240
- MOGREPS-15 (00Z&12Z) to T+360
- Global Model (00Z & 12Z) to T+168
- MOGREPS-G (0,6,12 & 18Z) to T+168
- Euro4 (0,6,12&18Z) to T+120/T+60
- UKV, via UKPP (0,3,6,9,12,15,18,21Z) to T+36
- UKPP nowcasts (hourly) to T+6

BestData parameters include

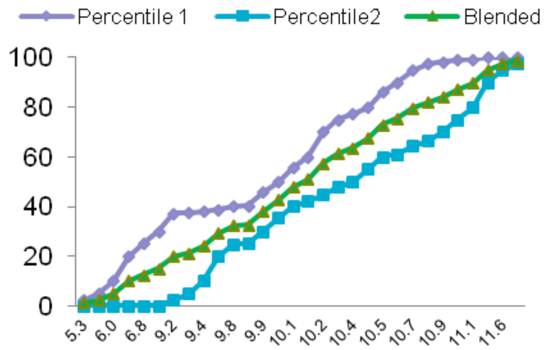
- Screen temperature
- ‘Feels like’ temperature
- Max & min temperatures
  - and temperature ranges
- 10m windspeed & direction
- Visibility
- Precipitation
- Weather type
- UV Index
- Available out to day 15 - Seamless

Blending Percentiles

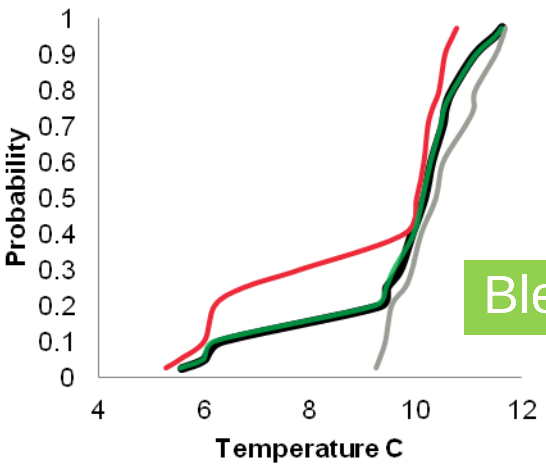
- We start by calculating where the percentile values from each ensemble would exist in the other probability distribution



- Probability spaces can then be combined with different weights depending on model skill.



We then calculate the percentile values in the new blended distribution to produce our new blended percentiles



Blended Percentile

Deterministic forecasts can be similarly blended in

Training data

- Trade-off:
  - Longer training sets make estimation more stable
  - Shorter training periods are adaptive to model and seasonal changes

⇨ Sliding training window of 25 days (40 for precip)

- Available data: 09/2013 – 09/2014
- Due to the set-up of Best Data, we calibrate each model run separately
- Lead times are grouped into 6-hourly chunks and then each chunk is calibrated separately

Gamma model: CRPS distribution