Statistical Postprocessing for Ensembles of Numerical Weather Prediction Models

Tilmann Gneiting

Institut für Angewandte Mathematik
Universität Heidelberg

Newton Institute, Cambridge
December 21, 2010

Probabilistic Weather Forecasting Using NWP Ensembles

Statistical Postprocessing

Bayesian Model Averaging (BMA)

Ensemble Model Output Statistics/Nonhomogeneous Gaussian Regression (EMOS/NGR)

Case Studies

BMA for Wind Direction

EMOS/NGR for Wind Speed

The Challenge: Physically Consistent Ensemble Forecasts of Spatio-Temporal Trajectories

Summary

Weather Prediction vs. Climate Prediction

Probabilistic Weather Forecasting Using NWP Ensembles

Numerical Weather Prediction (NWP) models are based on **deterministic simulation models** that represent the physics of the atmosphere (Bjerknes 1904)

system of **six partial differential equations** (conservation of momentum, mass, energy and entropy, and equation of state) **in six variables** (two velocity components, density, pressure, temperature, humidity)

equations are discretized and run forward in time to obtain **deterministic forecasts** of future states of the atmosphere

data assimilation systems provide **initial conditions** that describe the current state of the atmosphere on a 3d grid: on the order of 1–10 million inputs

Probabilistic forecasting

weather forecasting has traditionally been viewed as a **deterministic problem**, despite major sources of **uncertainty**

initial conditions: incomplete network of observations, measurement error, shortcomings in the data assimilation cycle, . . .

model formulation: incomplete knowledge of physical processes (e.g., inaccurate parameterizations of sub grid-scale processes), incomplete and inaccurate numerical schemes, . . .

major **shift of paradigms** since the early 1990s

Palmer (2000): "Although forecasters have traditionally viewed weather prediction as deterministic, a **culture change towards probabilistic forecasting** is in progress."

probabilistic forecasting aims at finding joint **predictive probability distributions** for **future quantities** or **events**, rather than just a single-valued point forecast

What is a good probabilistic forecast?
Maximizing sharpness subject to calibration

Gneiting, Balabdaoui and Raftery (2007) contend that the goal of probabilistic forecasting is to **maximize** the **sharpness** of the predictive distributions **subject to calibration**

calibration

refers to the statistical compatibility between the predictive distributions and the observations

joint property of the forecasts and the observations

sharpness

refers to the spread of the predictive distributions

property of the forecasts only

proper scoring rules such as the **logarithmic score** or the **continuous ranked probability score** (Gneiting and Raftery 2007) allow for a joint assessment of calibration and sharpness

Proper scoring rules

a **scoring rule** is a function

$$\mathsf{s}(F, x)$$

that assigns a numerical score to each pair (F, x), where F is the **predictive distribution** and x is the **verifying observation**

we consider scores to be **negatively oriented** penalties that forecasters aim to **minimize** on the average

a **proper** scoring rule s satisfies the expectation inequality

$$\mathbb{E}_G \, \mathsf{s}(G, X) \leq \mathbb{E}_G \, \mathsf{s}(F, X) \quad \text{for all} \quad F, G,$$

thereby encouraging **honest** and **careful assessments** by the forecaster

measure-theoretic characterizations relate to **convex analysis** and **information theory** (Gneiting and Raftery 2007)

Examples of proper scoring rules

the most popular example is the **logarithmic score**,

$$\mathrm{s}(f, x) = -\log f(x),$$

that is, the negative of the **predictive density**, f, evaluated at the verifying observation

my favorite score is the **continuous ranked probability score**,

$$\begin{aligned} \mathrm{crps}(F, x) &= \int_{-\infty}^{\infty} (F(y) - \mathbb{I}(y \geq x))^2 \, \mathrm{d}y \\ &= \mathbb{E}_F \, |X - x| - \frac{1}{2} \mathbb{E}_F \, |X - X'| \end{aligned}$$

where X and X' are independent random variables with cumulative distribution function F

the continuous ranked probability score is reported in the **same unit as the observations** and **generalizes** the **absolute error**, to which it reduces in the case of a point forecast

the **kernel score** representation allows for direct analogues on the **circle** and on **Euclidean spaces** (Gneiting and Raftery 2007)

NWP Ensembles

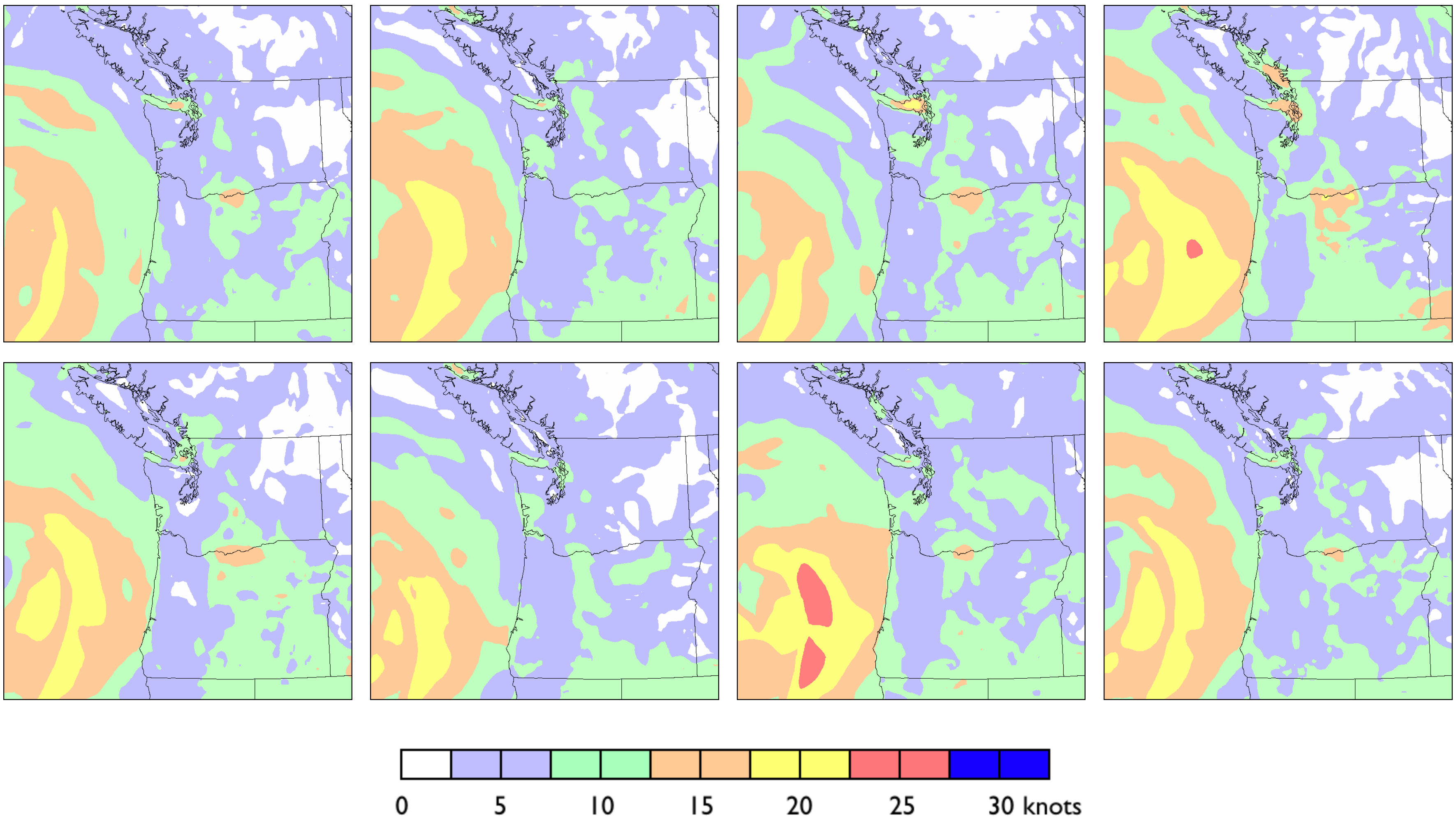
preferred method of **probabilistic weather prediction** is based on **ensembles** of **NWP forecasts**:

- each **ensemble member** is a single-valued, deterministic forecast from an NWP model
- but the forecasts differ from each other with respect to the two major sources of uncertainty: **initial conditions** and/or **model formulation**

global ensemble prediction systems have been operational at ECMWF and NCEP since December 1992

limited area systems such as the **University of Washington Mesoscale Ensemble (UWME)** (Grimit and Mass 2002; Eckel and Mass 2005) operate at lead times up to three days

University of Washington Mesoscale Ensemble (UWME)



48-hour ahead UWME forecast of maximum wind speed valid 7 August 2003

Spread-error correlation (the good), biases and lack of calibration (the bad), and the need for statistical postprocessing of NWP ensembles (the ugly)

spread-error correlation: there is a **positive association** between the **ensemble range** (known a priori) and the **forecast errors** (only known a posteriori)

model biases: systematic errors in NWP forecasts

lack of calibration: NWP ensembles tend to be **underdispersive**, in that the observed value falls far too often outside the ensemble range

model biases and the **lack of calibration** provide a strong and well recognized case for the **statistical postprocessing** of **NWP ensembles**

Statistical Postprocessing of NWP Ensembles

the goal of **statistical postprocessing** is to generate **calibrated** and **sharp** predictive distributions from the output of **NWP ensembles**

need to address **model biases** and **lack of calibration**

two general approaches to the **statistical postprocessing** of **forecast ensembles** have emerged, namely

- **Bayesian model averaging (BMA)**, where each ensemble member is associated with a kernel function, with a weight that reflects the member's relative skill
- **ensemble model output statistics (EMOS)** or **nonhomogeneous Gaussian regression (NGR)**, which fits a single, parametric predictive PDF using summary statistics from the ensemble

Example: BMA and EMOS for surface temperature

consider an **ensemble forecast**, $x_1, \ldots, x_m$, for **surface temperature**, y, at a given time and location

BMA employs Gaussian kernels with a linearly bias-corrected mean, that is, the BMA predictive PDF is the **Gaussian mixture**

$$p(y \mid x_1, \ldots, x_m) = \sum_{i=1}^{m} w_i \, \mathcal{N}(a_i + b_i x_i, \, \sigma^2)$$

with the BMA weights $w_1, \ldots, w_m$, bias parameters $a_1, \ldots, a_m$ and $b_1, \ldots, b_m$, and a common spread parameter σ^2

EMOS/NGR employs a **single Gaussian** predictive PDF, in that

$$p(y \mid x_1, \ldots, x_m) = \mathcal{N}(a + b_1 x_1 + \cdots + b_m x_m, \, c + d s^2)$$

with bias parameters a and $b_1, \ldots, b_m$, and spread parameters c and d, where s^2 is the ensemble variance

in our experience, the two approaches yield nearly equal predictive performance, with **BMA** being the more **flexible**, and **EMOS/NGR** the more **parsimonious** method

Bayesian model averaging (BMA)

Raftery et al. (2005) describe a BMA implementation for **temperature** and **pressure**, where the kernels are Gaussian

Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133, 1155–1174.

Sloughter et al. (2007) give a BMA implementation for **quantitative precipitation**, where the kernels are mixtures of a point mass at zero and a power-transformed gamma density

Sloughter, J. M., Raftery, A. E., Gneiting, T. and Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135, 3209–3220.

Fraley et al. (2010) show how to deal with **exchangeable** and **missing** ensemble **members**

Fraley, C., Raftery, A. E. and Gneiting, T. (2010). Calibrating multi-model forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Monthly Weather Review*, 138, 190–202.

Bayesian model averaging (BMA)

Sloughter et al. (2010) present a BMA implementation for **wind speed**, where the kernels are power-transformed gamma densities

Sloughter, J. M., Gneiting, T. and Raftery, A. E. (2010). Probabilistic wind forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, 105, 25–35.

Bao et al. (2010) present a BMA implementation for **wind direction**, where the kernels are von Mises densities on the circle

Bao, L., Gneiting, T., Grimit, E. P., Guttorp, P. and Raftery, A. E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, in press.

code is available in an **R package** named `ensembleBMA`, which is maintained by Chris Fraley at the University of Washington

BMA postprocessed probabilistic **temperature** and **quantitative precipitation** forecasts over the **Pacific Northwest** are available in **real time** at `http://probcast.washington.edu`

Ensemble model output statistics/Nonhomogeneous Gaussian regression (EMOS/NGR)

Gneiting et al. (2005) describe an EMOS/NGR implementation for **temperature** and **pressure**, where the predictive distribution is Gaussian

Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133, 1098–1118.

Thorarinsdottir and Gneiting (2010) present an EMOS/NGR implementation for **wind speed**, where the predictive distribution is truncated normal, to reflect the nonnegativity of wind speed

Thorarinsdottir, T. L. and Gneiting, T. (2010). Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroskedastic censored regression. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 173, 371–388.

an **R package** tentatively named `ensembleMOS` is under preparation (Bobby Yuen, University of Michigan)

Locally adaptive estimation

in **heterogeneous terrain** (for example, ocean versus lowlands versus mountains, upwind versus downwind, ...), it is critically important to develop BMA and EMOS/NGR techniques with **spatially varying** parameters

in the context of **temperature** forecasts with **BMA**, Kleiber et al. (2010) present methods for doing this:

local BMA seeks, at every grid point, training cases that are similar to the ensemble forecast at hand

geostatistical model averaging (GMA) fits, at every observation station, a BMA model, and interpolates the parameters thereof to the grid point at hand, using **Bayesian regularization**

Kleiber, W., Raftery, A. E., Baars, J., Gneiting, T., Mass, C. F. and Grimit, E. P. (2010). Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Monthly Weather Review*, in press.

these methods are being extended to **quantitative precipitation**

Case Study: BMA for Wind Direction

wind direction is an **angular variable** that takes values on the circle

bias correction via **circular-circular regression** using Möbius transforms in the complex plane (Downs and Mardia 2002; Kato, Shimizu and Shieh 2008)

ensemble calibration via the **BMA** predictive PDF

$$p(y \,|\, x_1, \ldots, x_m) = \sum_{i=1}^{m} w_i \, g(y \,|\, f_i, \kappa)$$

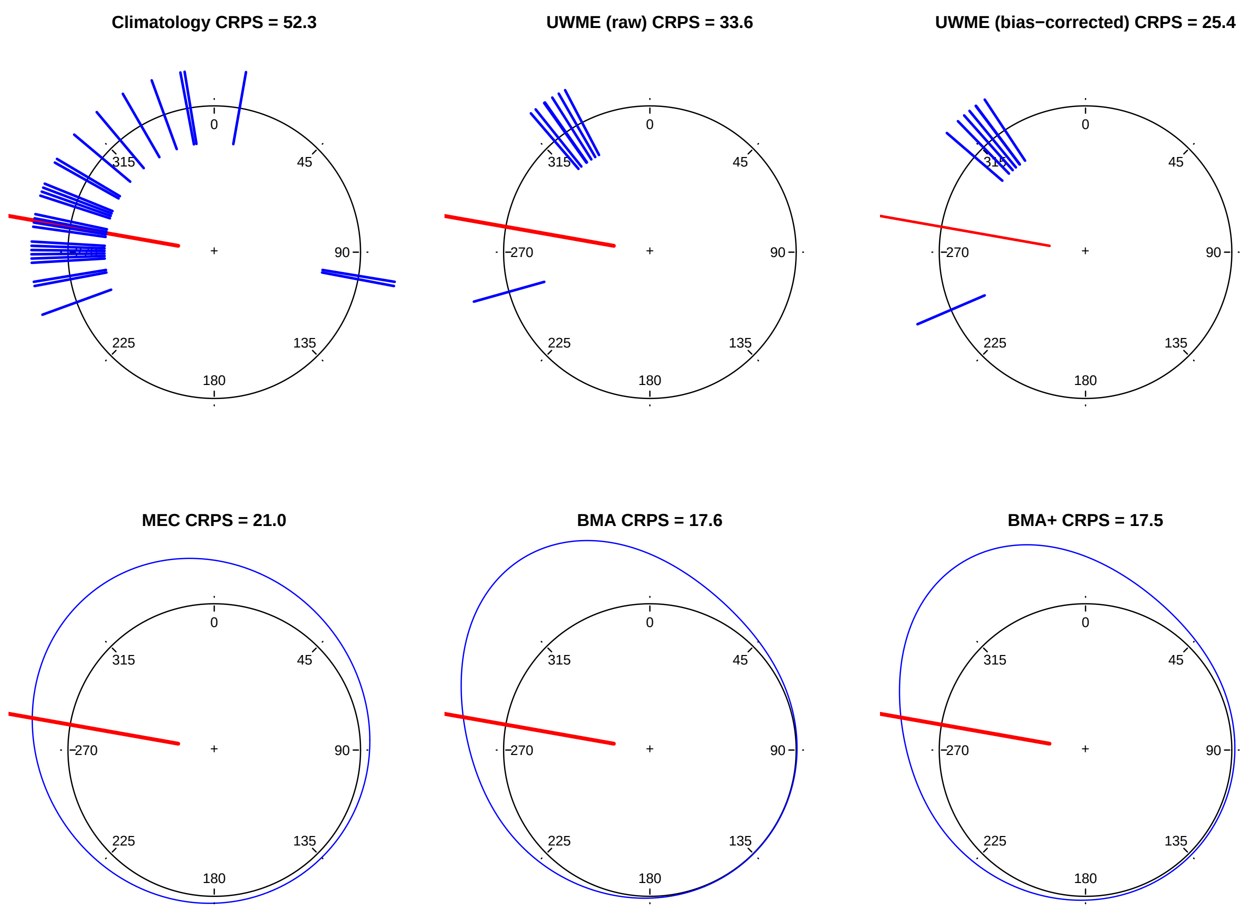
where $g(y \,|\, \mu, \kappa) \propto e^{\kappa \cos((y-\mu)\,\pi/180)}$ is a **von Mises** density with **mean direction** μ and **concentration parameter** κ, and f_i is a bias-corrected version of x_i, for $i = 1, \ldots, m$

the slightly more general **BMA**$^+$ predictive PDF includes an additional **uniform** component

applied to **48-hour forecasts** with the eight-member **UWME** in **2003**, using a sliding **28-day training period**

Example: Castlegar, British Columbia, valid 26 August 2003

	GFS	ETA	CMC	GSP	JMA	NGP	TCB	UKM	Unif
UWME	325	321	332	330	319	254	328	325	—
UWME (bc)	323	316	321	327	311	247	323	318	—
BMA	.11	.12	.11	.13	.11	.13	.12	.16	—
BMA^+	.10	.11	.10	.12	.10	.11	.11	.15	.10



Predictive performance results aggregated over the Pacific Northwest

	$\text{CRPS}_{\text{circ}}$	AE_{circ}
Climatology	35.9	56.9
UWME	35.0	42.8
UWME (bc)	31.2	39.3
MEC	28.8	39.3
BMA	27.8	39.4
BMA^+	27.6	39.3

Case Study: EMOS/NGR for Wind Speed

wind speed is a **nonnegative** and highly **skewed variable**

we perform **bias correction** and **ensemble calibration** simultaneously via the **EMOS** predictive PDF

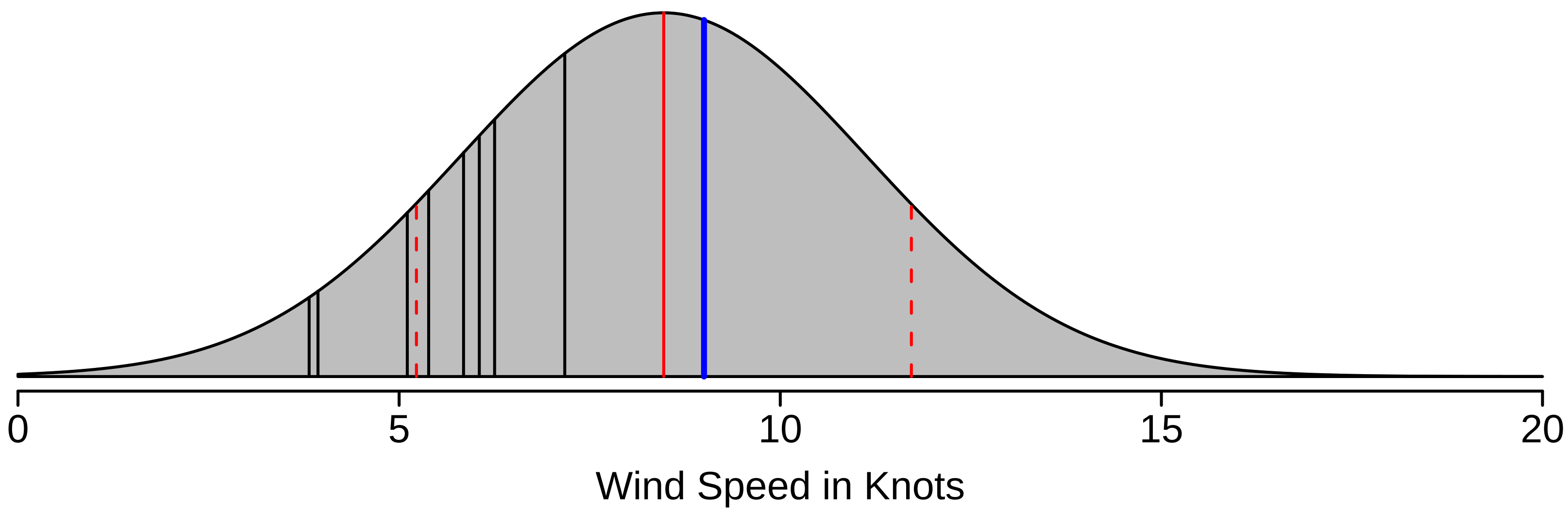
$$p(y \,|\, x_1, \ldots, x_m) = \mathcal{N}^0(a + b_1 x_1 + \cdots + b_m x_m,\, c + d s^2),$$

where the superscript denotes a **truncated normal density** with a **cutoff** at **zero**

applied to **48-hour forecasts** with the eight-member **UWME** in **2008**, using a sliding **40-day training period**

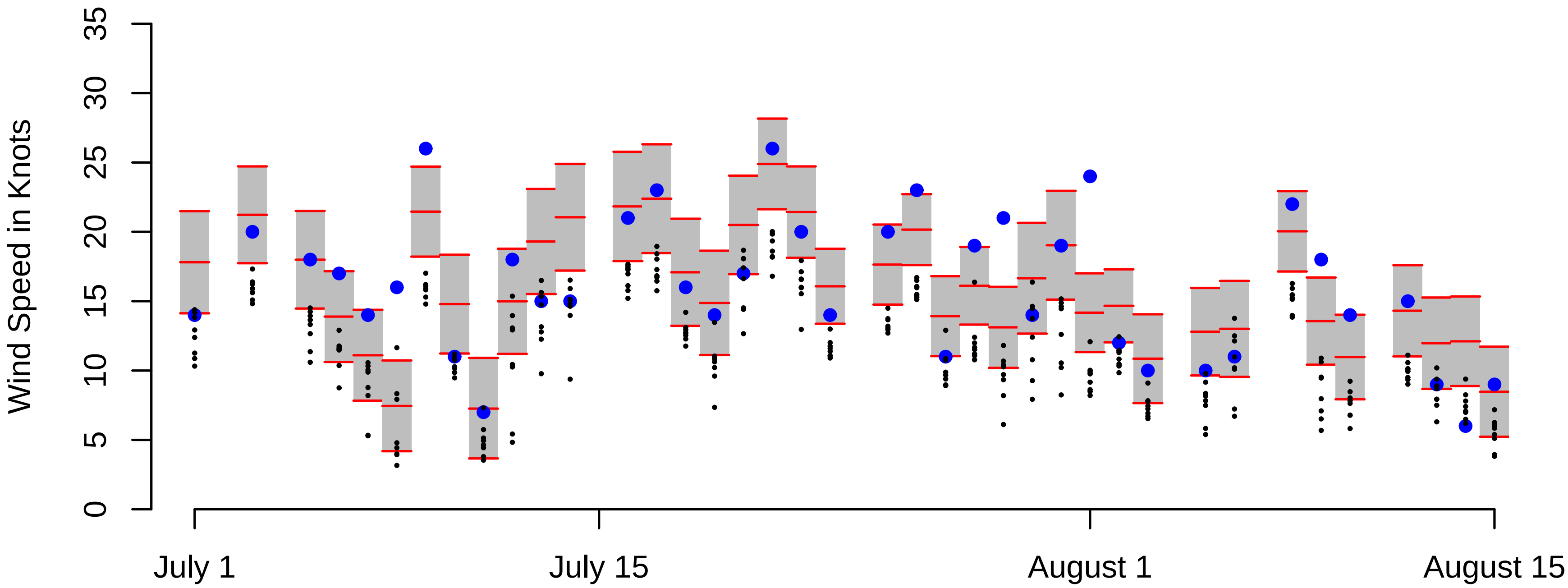
Example: The Dalles, Oregon, valid 15 August 2008

a	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	c	d
	AVN 5.11	CMC 5.85	ETA 6.25	GSP 3.82	JMA 3.94	NGP 6.05	TCW 7.17	UKM 5.39		
3.12	0.00	0.12	0.31	0.05	0.64	0.00	0.00	0.00	7.08	0.00



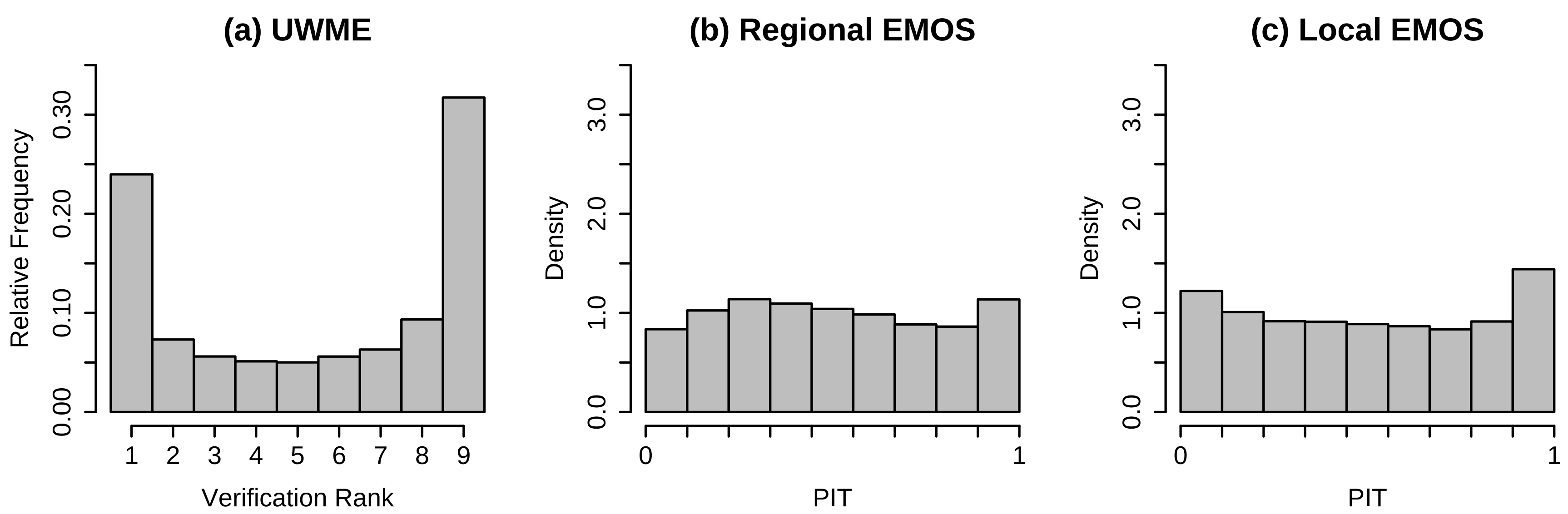
EMOS forecasts and performance at The Dalles

	CRPS	MAE
Climatology	3.53	5.12
UWME	3.53	4.42
Regional EMOS	2.85	3.96
Local EMOS	2.61	3.65



Predictive performance results aggregated over the Pacific Northwest

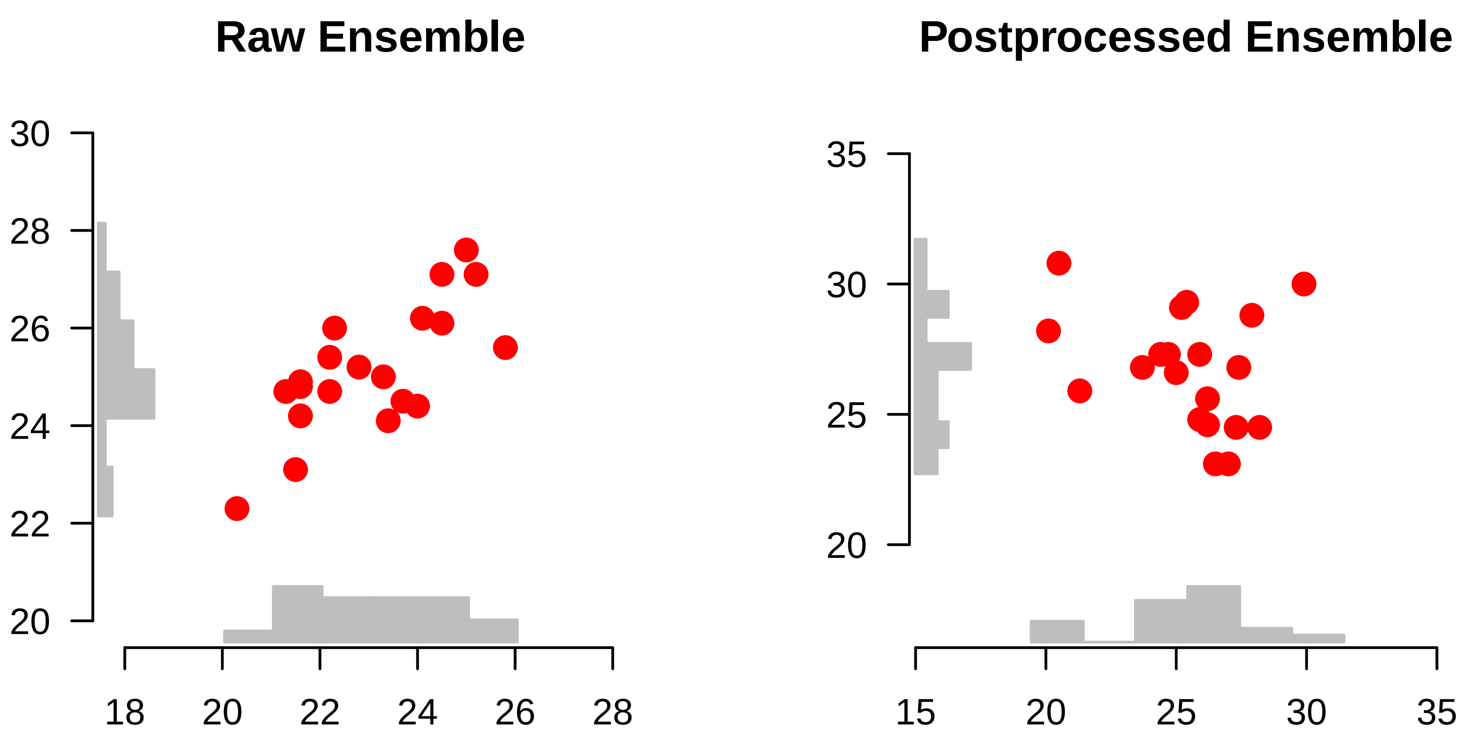
	CRPS	MAE
Climatology	2.45	3.50
UWME	2.47	3.15
Regional EMOS	2.16	3.00
Local EMOS	1.89	2.61



The Challenge: Physically Consistent Ensemble Forecasts of Spatio-Temporal Trajectories

the above methods apply to a **single weather variable** at a **single location** and a **single look-ahead time** only

unfortunate, because the postprocessed forecast fail to show **physically consistent** multivariate **dependence structures**



The challenge for the new decade

the most pressing need now is to develop methods that yield **physically realistic** probabilistic forecasts of **spatio-temporal trajectories**

for **single or multiple weather variables** at single or multiple **locations** and/or single or multiple **look-ahead times**

particularly important for **quantitative precipitation**, where probabilistic weather field forecasts provide crucial input for **flood management**

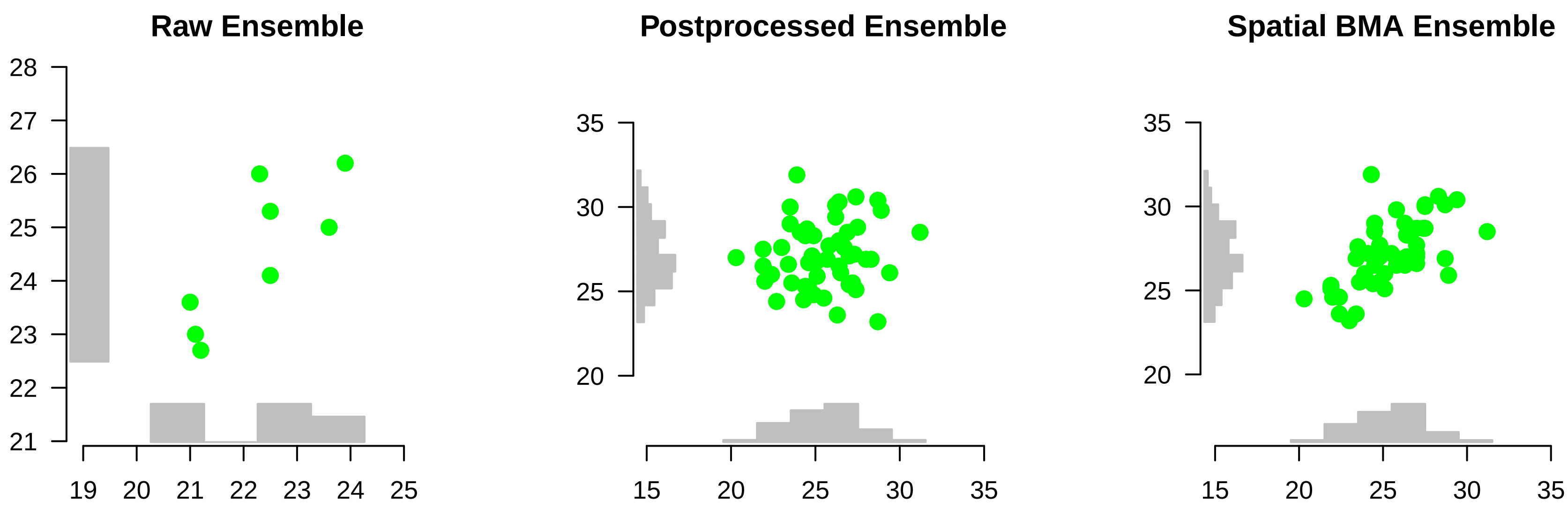
other **key applications** include **air traffic control** and **ship routeing**

at this time, we are investigating the use of parametric and non-parametric **copula approaches** for doing this, with **spatial BMA** being one such technique

Spatial BMA technique

basic idea:

- **dress** each ensemble member with physically realistic, **spatially** or **spatio-temporally correlated**, simulated **error fields**
- **combine** into a postprocessed **Spatial BMA** ensemble



Spatial BMA technique

developed and implemented successfully for **temperature**

Gel, Y., Raftery, A. E. and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The Geostatistical Output Perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, 99, 575–587.

Berrocal, V. J., Raftery, A. E. and Gneiting, T. (2007). Combining spatial statistical and ensemble information for probabilistic weather forecasting. *Monthly Weather Review*, 135, 1386–1402.

combines the strengths of **numerical** and **statistical modeling**

any ensemble size feasible

depends on **advanced** and/or novel statistical **methodology**, particularly for non-Gaussian weather variables, such as **quantitative precipitation**

Summary

the goal in **probabilistic forecasting** is to **maximize** the **sharpness** of the predictive distributions **subject to calibration**

probabilistic weather forecasting is done using **NWP ensembles**, but these tend to be **biased** and **lack calibration**, thus calling for **statistical postprocessing**

we have developed two general approaches, namely **Bayesian model averaging (BMA)**, and **ensemble model output statistics** or **nonhomogeneous Gaussian regression (EMOS/NGR)**

these apply to single weather variables at single locations and single look-ahead times

the grand challenge now is the quest for physically **consistent** probabilistic forecasts of **spatio-temporal** weather **trajectories**

Weather Prediction vs. Climate Prediction

weather prediction concerns **snapshots** of the state of the atmosphere

there is a **wealth** of **training data** that allows us to assess the performance of statistical techniques

climate prediction concerns the **probability** or **frequency distribution** of the state of the atmosphere

there is a **dearth** of **training data**

the emerging concept of **seamless prediction** forges weather forecasting and climate change studies into a single framework (Palmer et al., BAMS 2008; Hazeleger et al., BAMS 2010)

http://ecearth.knmi.nl