Space Weather with Quantified Uncertainty: Optimizing Ensembles for Probabilistic Predictions

Enrico Camporeale (CIRES, University of Colorado at Boulder & NOAA Space Weather Prediction Center)

https://ml-space-weather.github.io

Problem statement

Most of the models used in Space Weather (physics-based or empirical) are ***deterministic***, meaning that they provide single-point predictions.

It is important to be able to generate ***probabilistic*** forecasts, that is to associate ***uncertainties*** to single-point predictions.

Indeed, many would argue that
a forecast is not a forecast if it is not probabilistic!

How to generate a probabilistic forecast from a deterministic model?

The ACCRUE (Accurate and Reliable Uncertainty Estimate) method solves that problem

The standard way is by ensemble modeling, i.e. in a **Monte Carlo** fashion: one produces an ensemble of (single-point) forecasts by slightly changing the initial conditions or some other parameters of the model. The ensemble results can then be interpreted probabilistically.

Serious drawbacks of the Monte Carlo approach:

- Robust but extremely expensive: very slow convergence rate (square root of the number of samples);
- Assumes that one knows the correct probability distribution of inputs/parameters, which are often non-observables. To be done correctly, this would require a **Bayesian calibration** as in:


J. R. Statist. Soc. B (2001)
63, Part 3, pp. 425–464

Bayesian calibration of computer models

Marc C. Kennedy and Anthony O'Hagan

which itself requires Markov-Chain Monte Carlo samples!
Do not EVER run an ensemble simulation without having performed a Bayesian calibration of your parameters!

What is a probabilistic forecast anyway?

The general public does not understand the meaning of a probabilistic forecast


Risk Analysis, Vol. 25, No. 3, 2005 DOI: 10.1111/j.1539-6924.2005.00608.x

"A 30% Chance of Rain Tomorrow": How Does the Public Understand Probabilistic Weather Forecasts?

Gerd Gigerenzer,[1]* Ralph Hertwig,[2] Eva van den Broek,[1] Barbara Fasolo,[1] and Konstantinos V. Katsikopoulos[1]

<u>"30% chance of rain tomorrow"</u>

What does it mean?

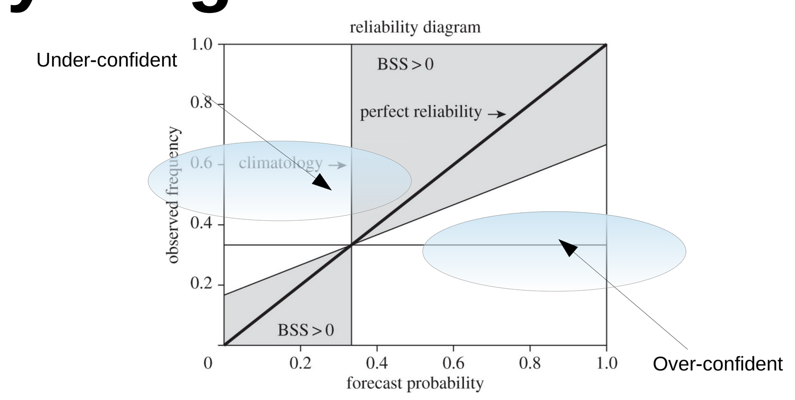
If you take a large enough sample of days for which it is predicted "30% chance of rain", it rains in about 30% of these. Tomorrow belongs to that sample.

Or, more simply (but you must really be a Bayesian): it will rain in 30% of all the instances of tomorrow.

This is the definition of Reliability of a probabilistic forecast

Reliability Diagram

A Reliability diagram measures how closely the forecast probabilities of an event correspond to the actual chance of observing the event.
It shows the **Observed frequency** of an event plotted against its **Forecast probability.**



Reliability is necessary but not sufficient. Why? Take a climatology model (prediction based on the long-term observed average). By definition Climatology has perfect reliability, but no skill!
We also need accuracy!

Accuracy: how do you measure it?

Many scores have been proposed. Here we make the two following working assumptions:

- The output of interest is a real continuous variable
- The forecast we want to generate is a <u>Gaussian distribution</u>

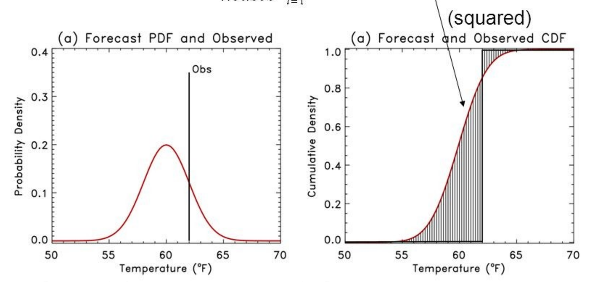
$$\text{error } \varepsilon = y^0 - \mu.$$

In this case, one appropriate score to measure the forecast accuracy is the

Continuous Ranked Probability Score

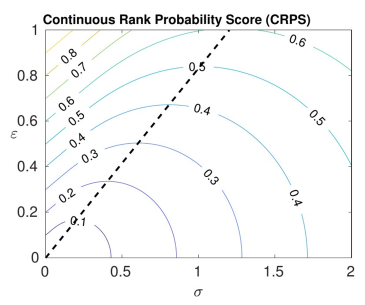
- Let $F_i^f(x)$ be the forecast probability CDF for the *i*th forecast case.
- Let $F_i^o(x)$ be the observed probability CDF (Heaviside function).

$$CRPS(forecast) = \frac{1}{ncases}\sum_{i=1}^{ncases}\int_{x=-\infty}^{x=\infty}\left(F_i^f(x) - F_i^o(x)\right)^2 dx$$



For a perfect (deterministic) forecast CRPS = 0

In the case of Gaussian forecast the CRPS has an analytical expression:



$$\mathrm{CRPS}(\mu,\sigma,y^o) = \sigma\left[\frac{y^o-\mu}{\sigma}\,\mathrm{erf}\left(\frac{y^o-\mu}{\sqrt{2}\sigma}\right) + \sqrt{\frac{2}{\pi}}\exp\left(-\frac{(y^o-\mu)^2}{2\sigma^2}\right) - \frac{1}{\sqrt{\pi}}\right]$$

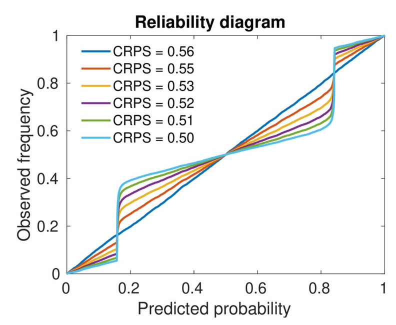
where μ is the mean, σ the standard deviation and y^o the observed value.

Problem re-statement

We have a deterministic model that produces an output μ, as a function of a multi-dimensional input x.

We want to use μ as the mean value of a Gaussian distribution that is interpreted as a probabilistic forecast.

What is the optimal value to choose for the standard deviation σ ?
(in general also function of x)



Example: for the same model, by decreasing CRPS one gets a worse reliability diagram

IDEA (that does not work)**:** Define σ as the one that minimizes the score CRPS.

This is easy. Simply differentiate $\frac{d\mathrm{CRPS}}{d\sigma} = \sqrt{\frac{2}{\pi}}\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) - \frac{1}{\sqrt{\pi}}$

The minimizer of CRPS is $\sigma^2_{\text{min, CRPS}} = \varepsilon^2/\log 2$ hence the variance is proportional to the error.

<u>Why does it not work?</u> The approach of simply minimizing CRPS <u>does not work</u>, because it generates forecasts that are not reliable (i.e. CRPS does not enforce reliability)

IDEA (that works!)
Define a cost function that enforces simultaneously reliability and accuracy.
We call it the **Accuracy-Reliability** (AR) cost function

Accuracy- Reliability cost function

First, we need to define a Reliability Score (RS), i.e. a measure of how reliable a model is.
Again we restrict to the case of Gaussian forecast.

The main idea is that for a reliable model the relative errors are also distributed normally.
We define the RS as:

$$\mathrm{RS} = \int_{-\infty}^{\infty}[\Phi(\eta) - C(\eta)]^2 d\eta$$

with $\eta_i = (y_i^o - \mu_i)/(\sqrt{2}\sigma_i)$ the relative errors, and the cdf $\Phi(\eta) = \frac{1}{2}(\mathrm{erf}(\eta)+1)$

where $C(\eta)$ is the empirical cumulative distribution of the relative errors η, that is

$$C(y) = \frac{1}{N}\sum_{i=1}^{N} H(y - \eta_i) \qquad (7)$$

A crucial detail of this work is that the RS as defined above can be calculated analytically, via expansion into a telescopic series:

$$\mathrm{RS} = \sum_{i=1}^{N}\left[\frac{\eta_i}{N}(\mathrm{erf}(\eta_i)+1) - \frac{\eta_i}{N^2}(2i-1) + \frac{\exp(-\eta_i^2)}{\sqrt{\pi}N}\right] - \frac{1}{2}\sqrt{\frac{2}{\pi}}$$

Keeping in mind that the CRPS and the RS are competing scores, this becomes a **two-objective minimization problem**.

We can finally define our <u>Accuracy-Reliability cost function</u> as a linear combination of the two:

$$\mathrm{AR} = \beta\cdot\overline{\mathrm{CRPS}} + (1-\beta)\mathrm{RS}.$$

We choose the scaling factor β as

$$\beta = \mathrm{RS}_{\min}/(\overline{\mathrm{CRPS}}_{\min} + \mathrm{RS}_{\min}).$$

Problem re-re-statement:
Given a number of observations y and model predictions μ we want:

1) to estimate the optimal standard deviations σ that minimizes the AR cost function;

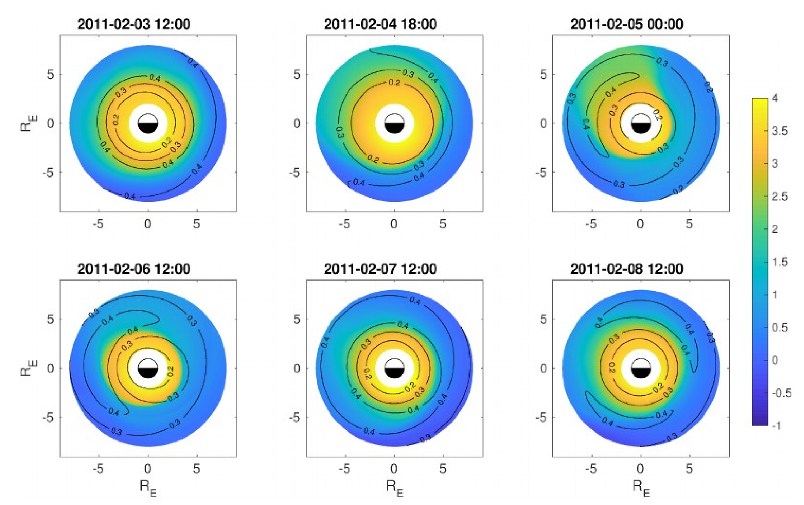
2) to have a mechanism that generates σ as function of the inputs x for unseen data (i.e. without having the observations y)

<u>**We use a deep neural network to achieve both goals**</u>

Take home message (and results)

We have designed a generative model that:

- Takes a set of deterministic forecasts and the corresponding ground truth values (the details of the deterministic model are not important: it can be empirical or physics-based)
- Transform the deterministic forecasts into probabilistic predictions, in the form of Gaussian distributions
- The probabilistic forecast is guaranteed to be both accurate and reliable (i.e. a trade-off between the two)
- By using a neural network the standard deviations are generated as function of the multidimensional input, for any (new) input.



Example: DEN2D is a data-driven, deterministic model that estimates the electron plasma density in the plasmasphere. We generate probabilistic predictions based on DEN2D. The colormap is the electron density mean, the isocontours the associated standard deviations

TABLE 1: Comparison between different methods on several multidimensional datasets. Median values are reported, calculated over 50 runs. Confidence intervals represent one standard deviation. Best values are in bold

Method / Score			CRPS	RECAL	KM	ACCRUE
			CRPS			
Dataset	**Size**	**Dim.**				
Boston Housing	506	13	0.25 ± 0.05	0.25 ± 0.04	0.25 ± 0.03	**0.23 ± 0.04**
Concrete	1030	8	0.22 ± 0.03	0.23 ± 0.13	0.26 ± 0.02	**0.21 ± 0.03**
Energy	768	8	0.059 ± 0.03	0.056 ± 0.03	0.087 ± 0.01	**0.052 ± 0.01**
Kin8nm	8192	8	0.17 ± 0.005	0.16 ± 0.01	0.24 ± 0.005	**0.16 ± 0.005**
Power plant	9568	4	0.13 ± 0.003	0.13 ± 0.05	0.15 ± 0.002	**0.12 ± 0.01**
Protein	45,730	9	0.38 ± 0.02	0.47 ± 0.13	0.40 ± 0.007	**0.37 ± 0.02**
Wine	1599	11	0.48 ± 0.03	0.50 ± 0.29	**0.46 ± 0.02**	0.48 ± 0.06
Yacht	308	6	0.06 ± 0.08	0.06 ± 0.02	0.19 ± 0.02	**0.06 ± 0.02**
Score			Cal. err. (%)			
Dataset	**Size**	**Dim.**				
Boston Housing	506	13	26.2 ± 7.9	20.6 ± 5.5	17.5 ± 3.7	**16.7 ± 5.9**
Concrete	1030	8	22.6 ± 5.8	14.4 ± 3.8	22.1 ± 3.0	**11.5 ± 3.9**
Energy	768	8	29.3 ± 8.9	29.2 ± 8.0	28.3 ± 2.8	**13.0 ± 6.5**
Kin8nm	8192	8	15.9 ± 1.28	8.3 ± 1.30	25.5 ± 0.5	**5.8 ± 1.28**
Power plant	9568	4	12.5 ± 1.4	3.4 ± 0.9	16.1 ± 0.8	**2.6 ± 0.8**
Protein	45,730	9	13.1 ± 0.8	**5.0 ± 0.9**	10.6 ± 0.9	5.4 ± 0.88
Wine	1599	11	16.0 ± 3.7	**7.9 ± 2.0**	8.0 ± 2.4	8.3 ± 2.4
Yacht	308	6	26.0 ± 9.4	24.3 ± 13.5	36.6 ± 3.0	**19.5 ± 8.5**

E-mail: enrico.camporeale@noaa.gov

Camporeale et al. (2019) On the generation of probabilistic forecasts from deterministic models, *Space Weather*, 17(3), 455
Camporeale, E., & Carè, A. (2021). ACCRUE: Accurate and Reliable Uncertainty Estimate in Deterministic models. *International Journal for Uncertainty Quantification*, 11(4).

University of Colorado Boulder