# On the generation of probabilistic forecasts from deterministic models

Enrico Camporeale

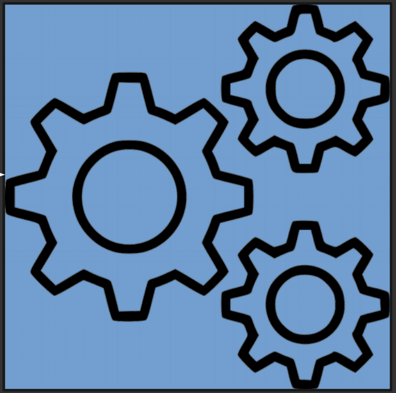
CIRES / CU Boulder & NOAA Space Weather Prediction Center
Center for Mathematics and Computer Science (CWI), Amsterdam, Netherlands

Collaborators: Xiangning Chu (U. Colorado), Oleksiy Agapitov (SSL, Berkeley)
Jacob Bortnik (UCLA), Algo Care' (U. Brescia, Italy)

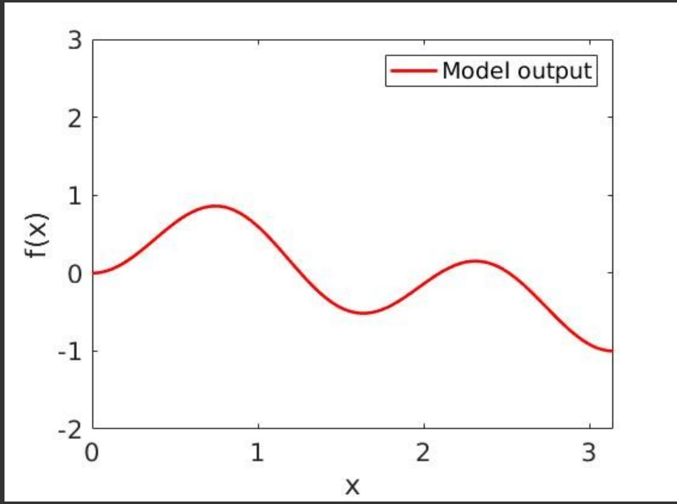
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776262 (AIDA)

# Problem statement

Set of Inputs,
Parameters,
etc.

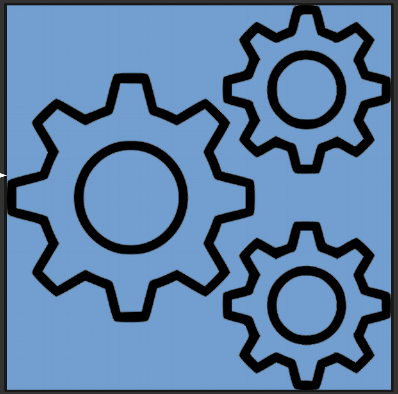


Deterministic model

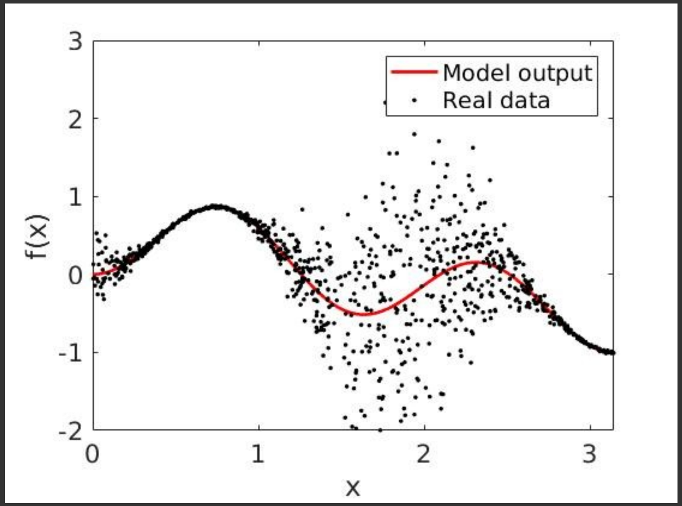


# Problem statement

Set of Inputs, Parameters, etc.



Deterministic model + Real Data

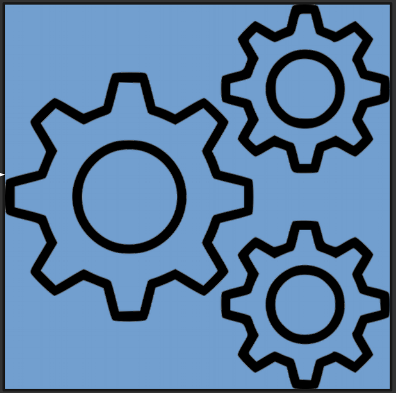



Uncertainties:

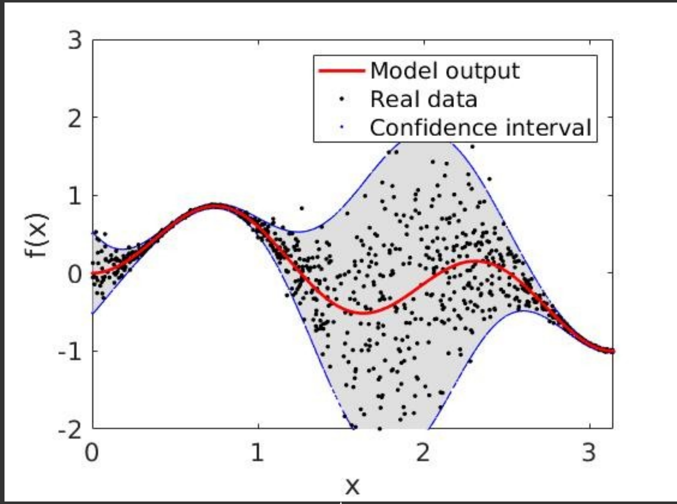
- Epistemic (do not know the parameters exactly)
- Aleatoric (physics that is not in the model)
- Algorithmic (numerical errors)

# Problem statement

Set of Inputs, Parameters, etc.



Deterministic model + Real Data




Uncertainties:

- Epistemic (do not know the parameters exactly)
- Aleatoric (physics that is not in the model)
- Algorithmic (numerical errors)

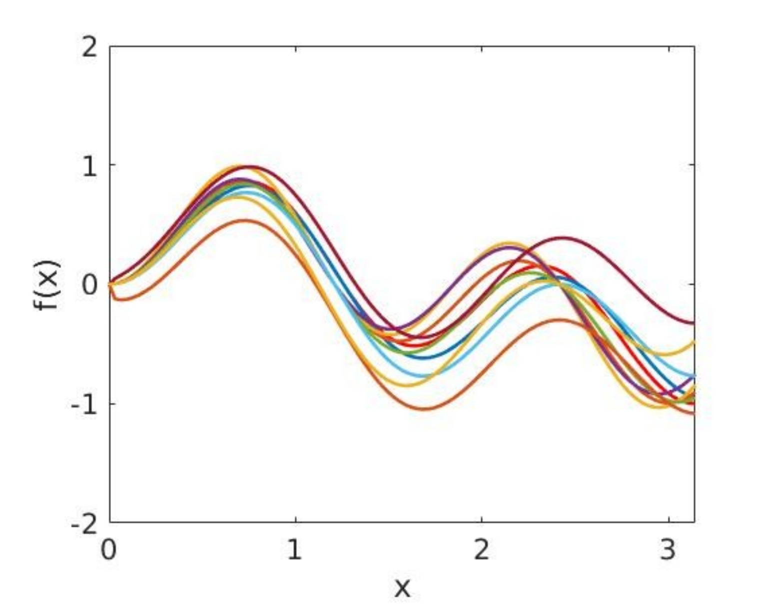
?

Generate heteroskedastic uncertainty

# Standard Approach

- The golden standard approach to estimate uncertainties based on a deterministic model is by running a

  **Monte Carlo ensemble**
  (e.g. by small perturbations of initial conditions)

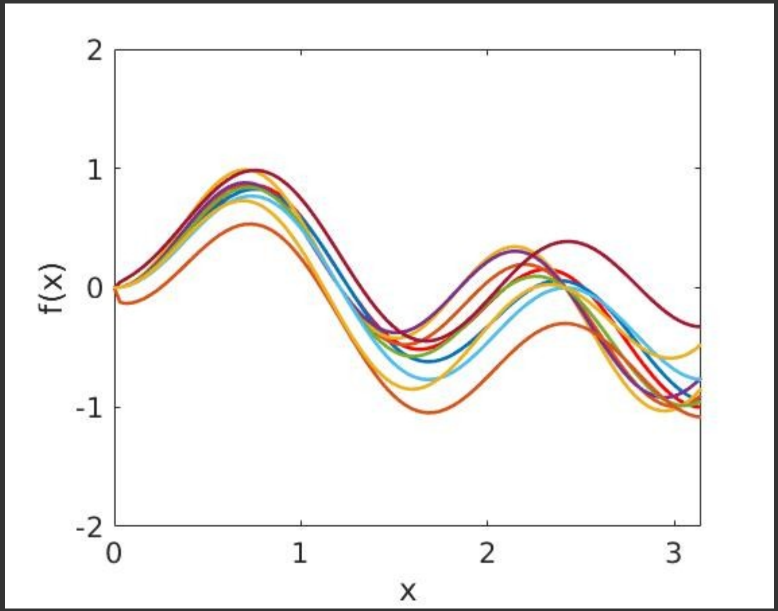



- This has two problems:
  –
  –

# Standard Approach

- The golden standard approach to estimate uncertainties based on a deterministic model is by running a

  **Monte Carlo ensemble**
  (e.g. by small perturbations of initial conditions)

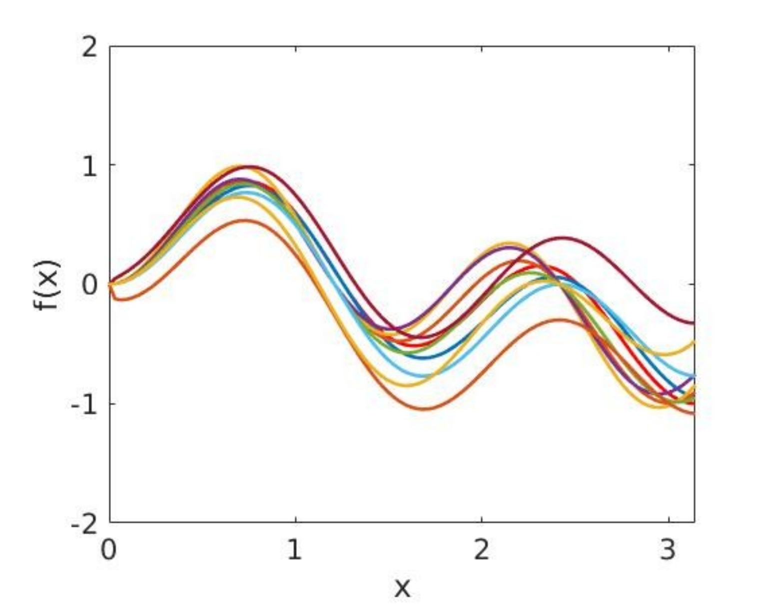



- This has two problems:
  - It requires many runs (expensive)
  -

# Standard Approach

- The golden standard approach to estimate uncertainties based on a deterministic model is by running a

  **Monte Carlo ensemble**
  (e.g. by small perturbations of initial conditions)




- This has two problems:
  - It requires many runs (expensive)
  - It requires to know what is the probability distribution of inputs
    (as in the work of R. Sarma , Thursday morning)

# Take home message

We have devised a method that:

- Estimates the uncertainties associated with single-point outputs generated by a deterministic model, in terms of Gaussian distributions;
- Ensures the optimal trade-off between accuracy and reliability;
- Does not need to run ensembles. It costs as much as training a neural network
- Code available: zenodo.1485608

Space Weather

RESEARCH ARTICLE

10.1029/2018SW002026

**Key Points:**

- We introduce a new method to estimate the uncertainties associated with single-point outputs generated by a deterministic model
- The method ensures a trade-off between accuracy and reliability of the generated probabilistic forecasts
- Computationally cheap model:

## On the Generation of Probabilistic Forecasts From Deterministic Models

E. Camporeale[1,2], X. Chu[3], O. V. Agapitov[4], and J. Bortnik[5]

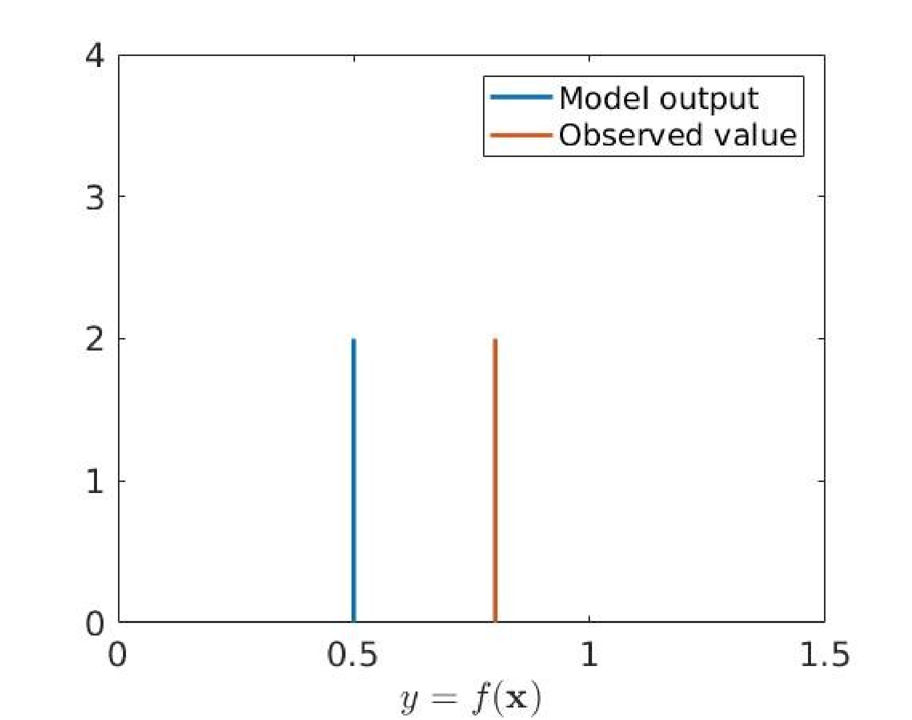
[1]Center for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [3]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA, [4]Space Sciences Laboratory, University of California Berkeley, Berkeley, CA, USA, [5]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA

# Utilitarian Approach

Let us assume that for a single (multidimensional) input **x**, our model predicts an output y = f(x).

Blue line → Model output
Red line → Real (observed value)

Working hypothesis:
We want to use the model output as the mean of a Gaussian distribution that is interpreted as a probabilistic forecast.



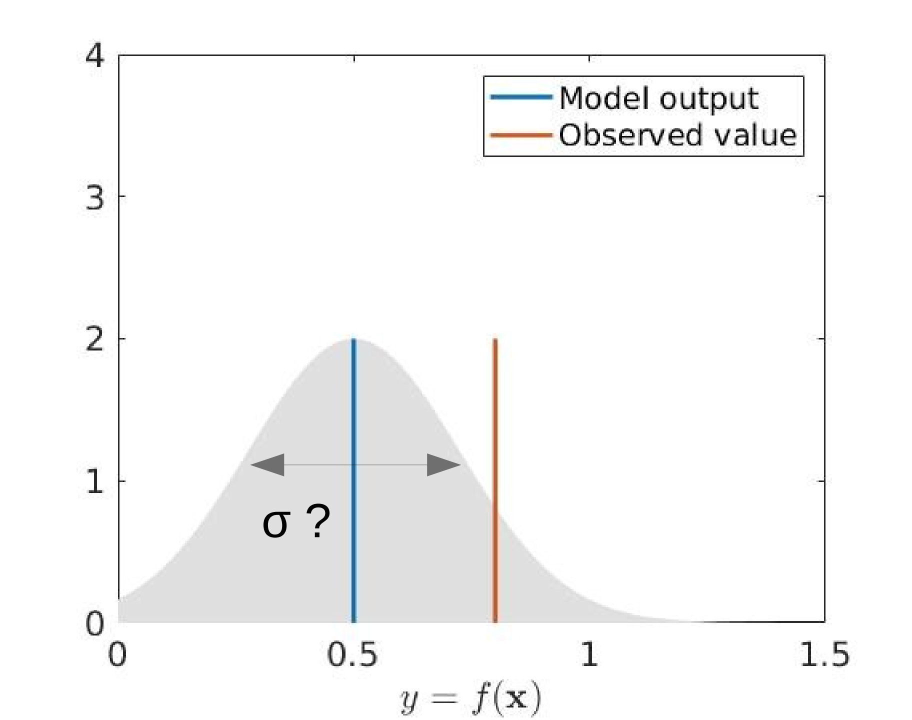
# Utilitarian Approach

Let us assume that for a single (multidimensional) input **x**, our model predicts an output y = f(x).

Blue line → Model output
Red line → Real (observed value)

<u>Working hypothesis:</u>
We want to use the model output as the mean of a Gaussian distribution that is interpreted as a probabilistic forecast.

**What is the optimal width of a Gaussian forecast?**



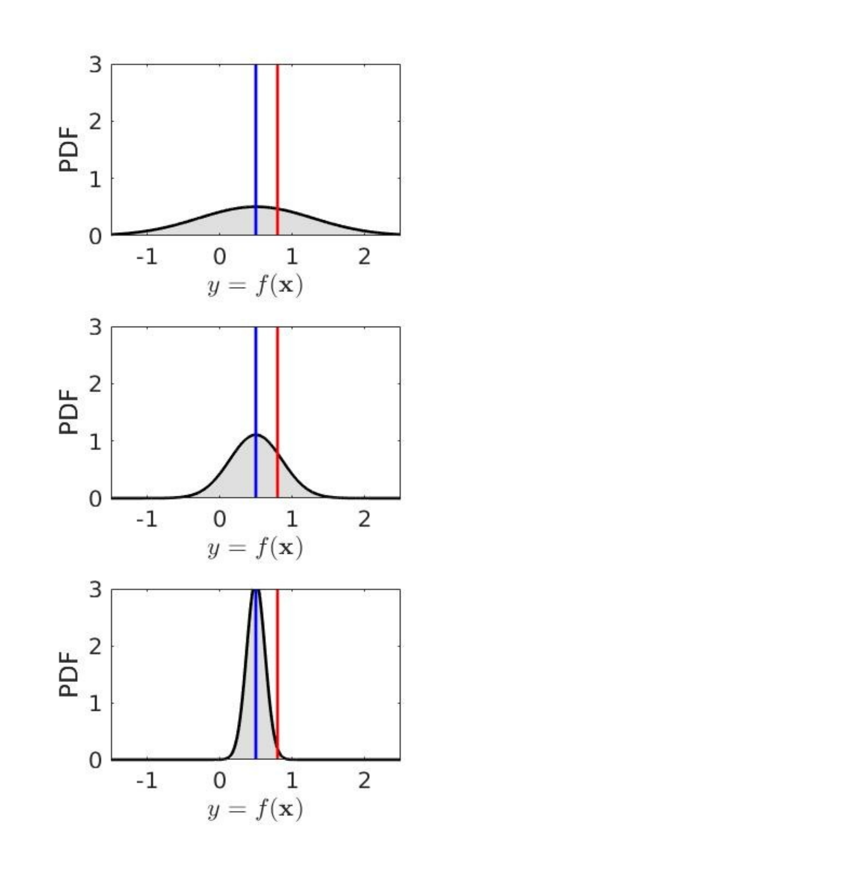
# Continuous Rank Probability Score

What is the optimal width of a Gaussian forecast?

"Optimal" with respect to what?

We use the CRPS as a score for "accuracy".

- CRPS is a generalization of Brier score
- It has a simple graphical interpretation
- CRPS = 0 for perfect forecast

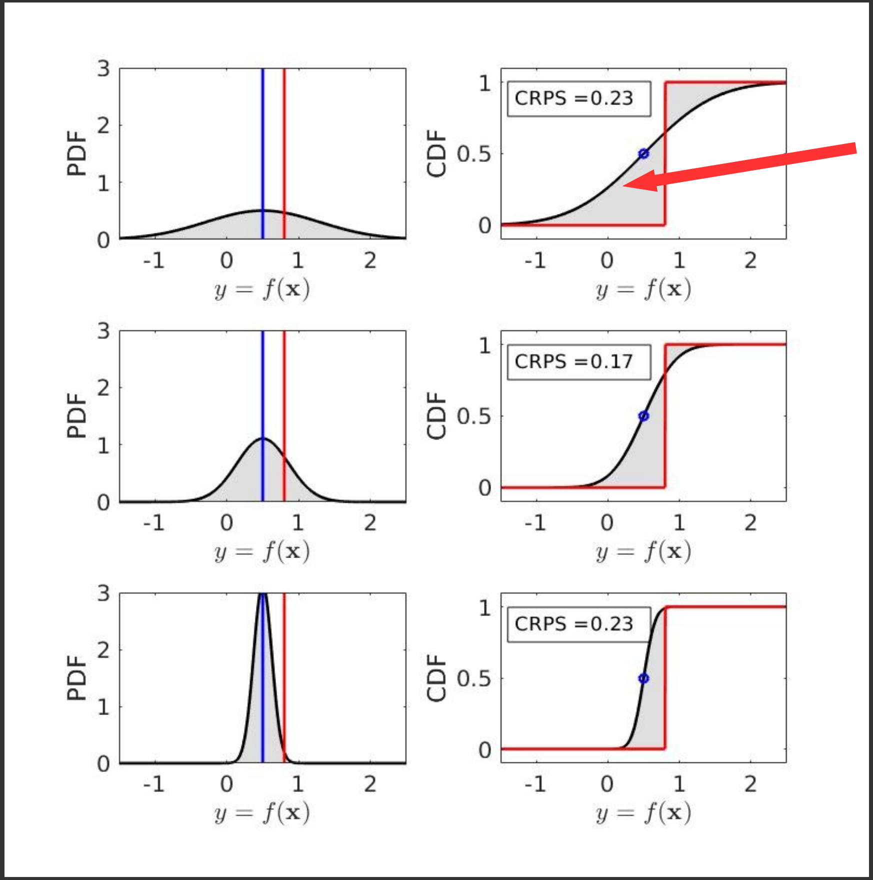


# Continuous Rank Probability Score

- CRPS is a generalization of Brier score
- It has a simple graphical interpretation
- CRPS = 0 for perfect forecast
- CRPS = $\int \left( C(y) - H(\hat{y}) \right)^2 dy$

CDF

Step function




CRPS proportional to this area

# Continuous Rank Probability Score

For a Gaussian distribution, CRPS has an analytical expression which is a function of:

- Error ε: difference between model output and observed value
- Standard deviation σ

If we define the optimal σ the one that minimizes CRPS

we obtain a probabilistic forecast that is not RELIABLE

# What is a probabilistic forecast anyway?

*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]**

TABLE 2. Responses to Q14a, the meaning of the forecast "There is a 60% chance of rain for tomorrow" ($N$ = 1330).

| | Percent of respondents |
|---|---|
| It will rain tomorrow in 60% of the region. | 16 |
| It will rain tomorrow for 60% of the time. | 10 |
| It will rain on 60% of the days like tomorrow.* | 19 |
| 60% of weather forecasters believe that it will rain tomorrow. | 22 |
| I don't know. | 9 |
| Other (please explain). | 24 |

* Technically correct interpretation, according to how PoP forecasts are verified, as interpreted by Gigerenzer et al. (2005).

974 WEATHER AND FORECASTING

**Communicating Uncertainty in Weather Forecasts: A Survey of the U.S. Public**

REBECCA E. MORSS, JULIE L. DEMUTH, AND JEFFREY K. LAZO

*National Center for Atmospheric Research,* Boulder, Colorado*

# What is a probabilistic forecast anyway?

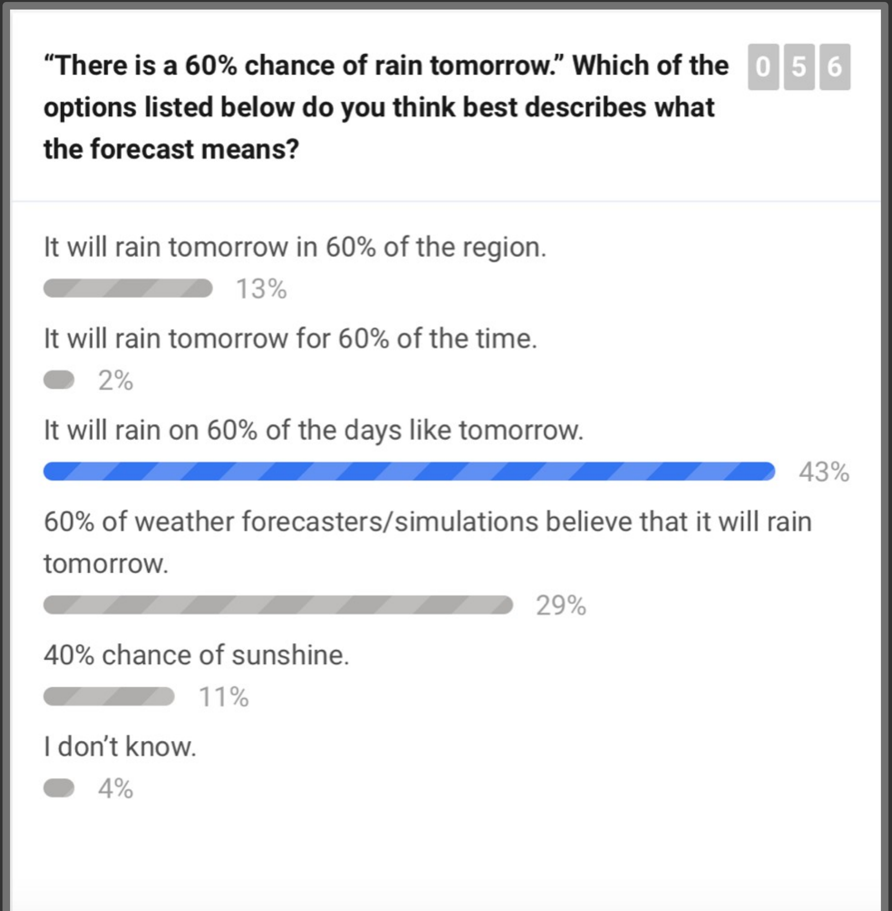



TABLE 2. Responses to Q14a, the meaning of the forecast "There is a 60% chance of rain for tomorrow" ($N = 1330$).

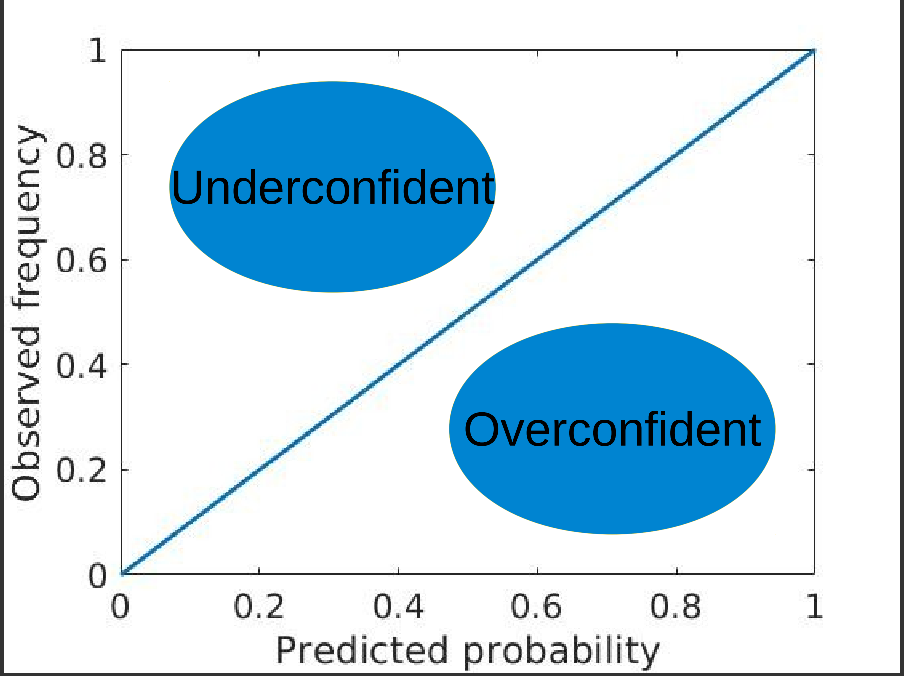
| | Percent of respondents |
|---|---|
| It will rain tomorrow in 60% of the region. | 16 |
| It will rain tomorrow for 60% of the time. | 10 |
| It will rain on 60% of the days like tomorrow.* | 19 |
| 60% of weather forecasters believe that it will rain tomorrow. | 22 |
| I don't know. | 9 |
| Other (please explain). | 24 |

* Technically correct interpretation, according to how PoP forecasts are verified, as interpreted by Gigerenzer et al. (2005).

# Reliability diagram

Reliability is the property of a probabilistic model that measures its **statistical consistency with observations.**

For example, for forecasts of discrete events, the reliability measures if an event occurs on average with frequency *p*, when it has been predicted to occur with probability *p*.



# Reliability diagram

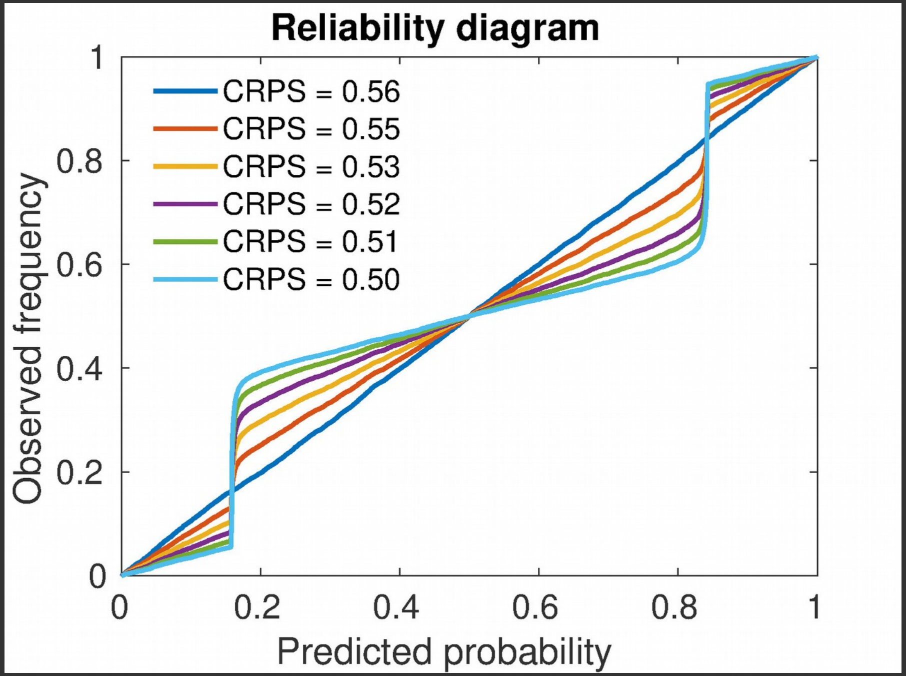
The values of σ that minimize CRPS can be derived analytically:

$$\sigma^2_{opt} \sim \varepsilon^2$$

The more we minimize CRPS, the worse reliability we get.

Mathematical proof is straightforward.

See arXiv:1811.12692



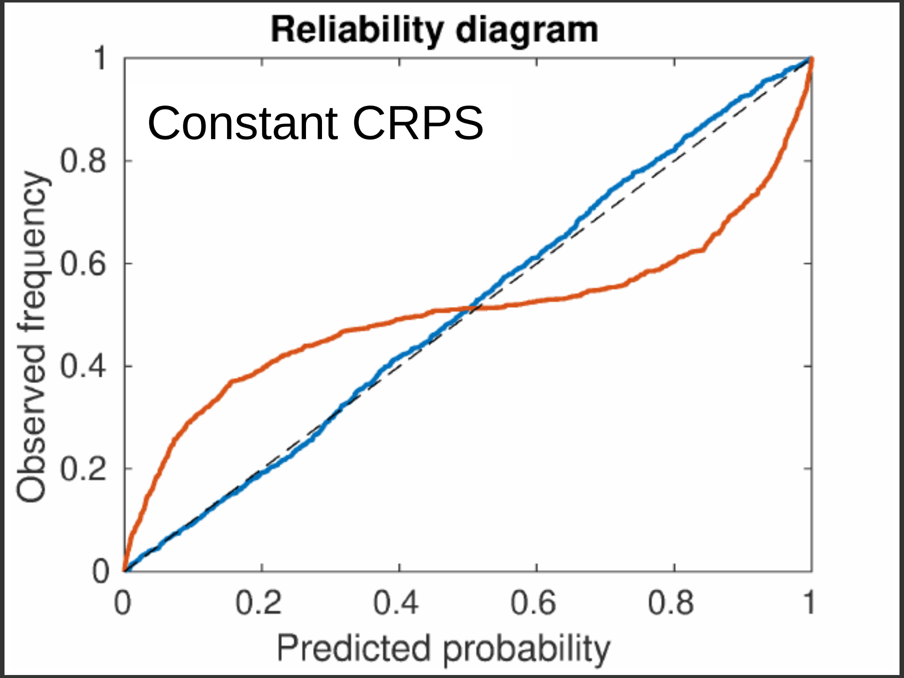
# Reliability diagram

The values of σ that minimize CRPS can be derived analytically:

$$\sigma^2_{opt} \sim \varepsilon^2$$

A given value of CRPS is not informative on the reliability

See arXiv:1811.12692



# Reliability Score

We define an analytical Reliability Score (RS)

- It measures how far the distribution of standardized errors ε/σ is from a Standard Normal distribution (other definition of reliability)

$$\eta_i = \varepsilon_i/(\sqrt{2}\sigma_i), \tag{5}$$

where the standard deviations $\sigma_i$ are determined by the input vector. If $\sigma(\mathbf{x})$ is not constant then this definition acts to both standardize and transform the error distribution. While the forecast errors $\varepsilon_i$ may not be Gaussian, in the case of a normally distributed forecast we expect $\eta$ calculated over a sample of N prediction-observation pairs to follow a standard normal distribution with CDF $\Phi(\eta) = \frac{1}{2}(\mathrm{erf}(\eta)+1)$. Hence, we define the Reliability Score (RS) as

$$\mathrm{RS} = \int_{-\infty}^{\infty} \left[\Phi(y) - C_\eta(y)\right]^2 dy, \tag{6}$$

where $C_\eta(y)$ is the empirical cumulative distribution of the standardized errors $\eta$, that is

$$C_\eta(y) = \frac{1}{N}\sum_{i=1}^{N} H(y - \eta_i) \tag{7}$$

with $\eta_i = (y_i^o - \mu_i)/(\sqrt{2}\sigma_i)$.

# Two-objective cost function

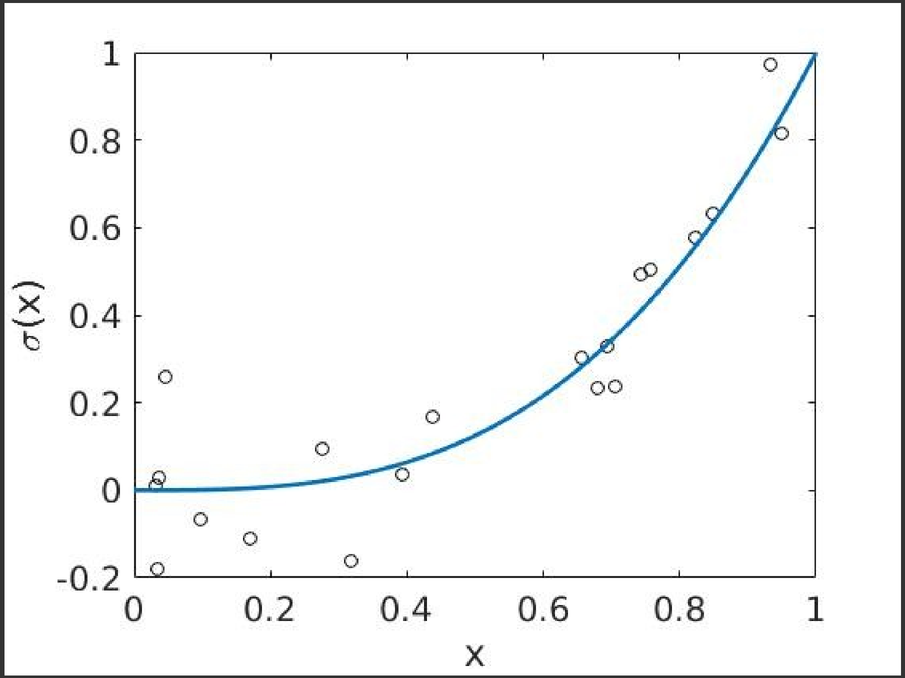
- This is a two-objective optimization problem, because reliability and accuracy are competing objectives.
- We define the Accuracy-Reliability (AR) cost function:

$$AR = \underbrace{CRPS}_{\text{Accuracy}} + \beta * \underbrace{RS}_{\text{Reliability}}$$

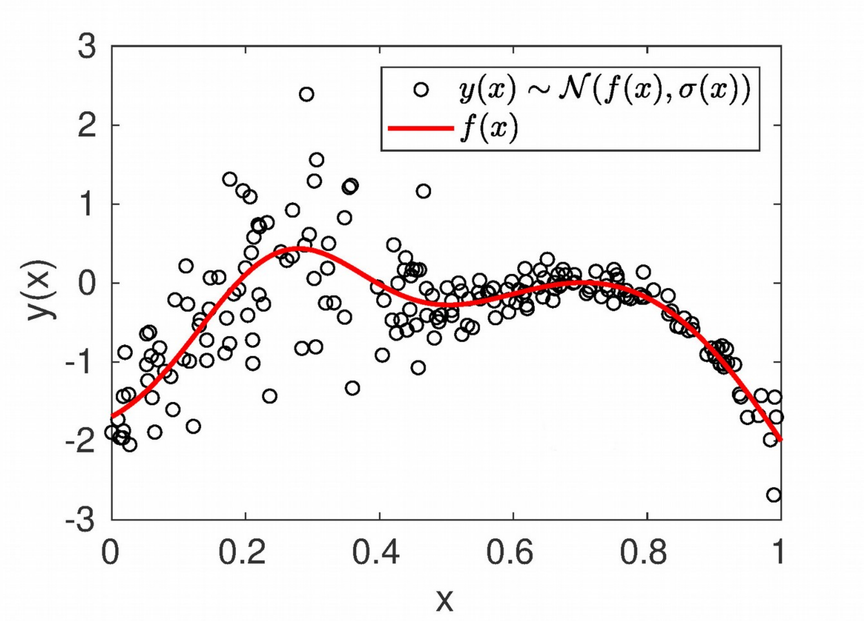
- Accuracy and Reliability cannot both be minimized simultaneously
- We have to find the best trade-off

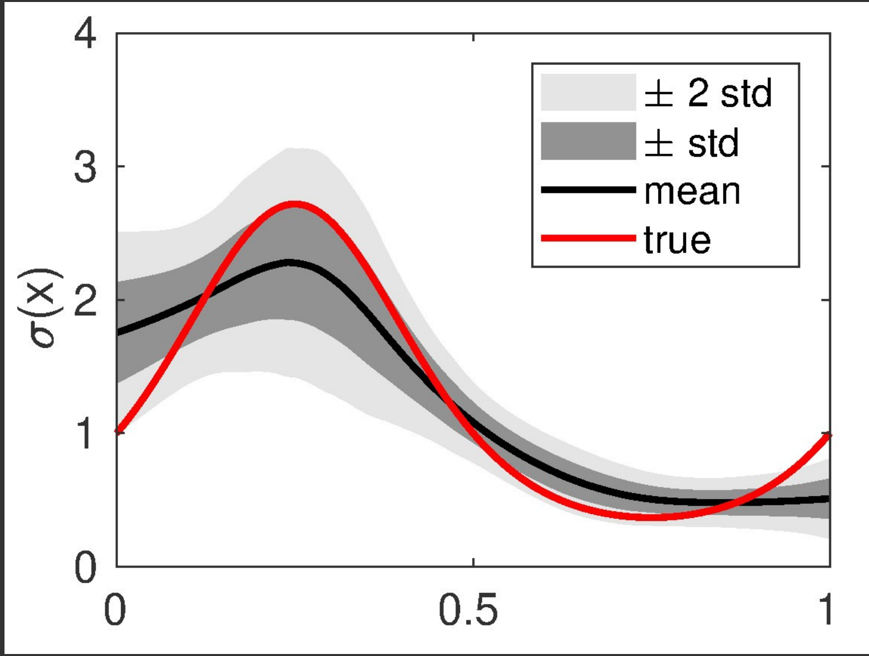
# The Method

- Take a sample of N errors ε (difference between model output and observed values) and the corresponding model inputs **x**
- We define as optimal standard deviation σ the one that optimizes the Accuracy-Reliability cost function (that has an analytical expression)
- We also want to have a way of generating a smooth function σ(**x**) for any value of **x**
- We use a neural network that takes **x** as input and produces σ(**x**) as output by minimizing AR cost function.



# Synthetic example



# DEN2D: model for plasmasphere electron density

Estimates the electron density based on history of geomagnetic indexes using a Neural Network.

X. Chu et al. *JGR* (2017)

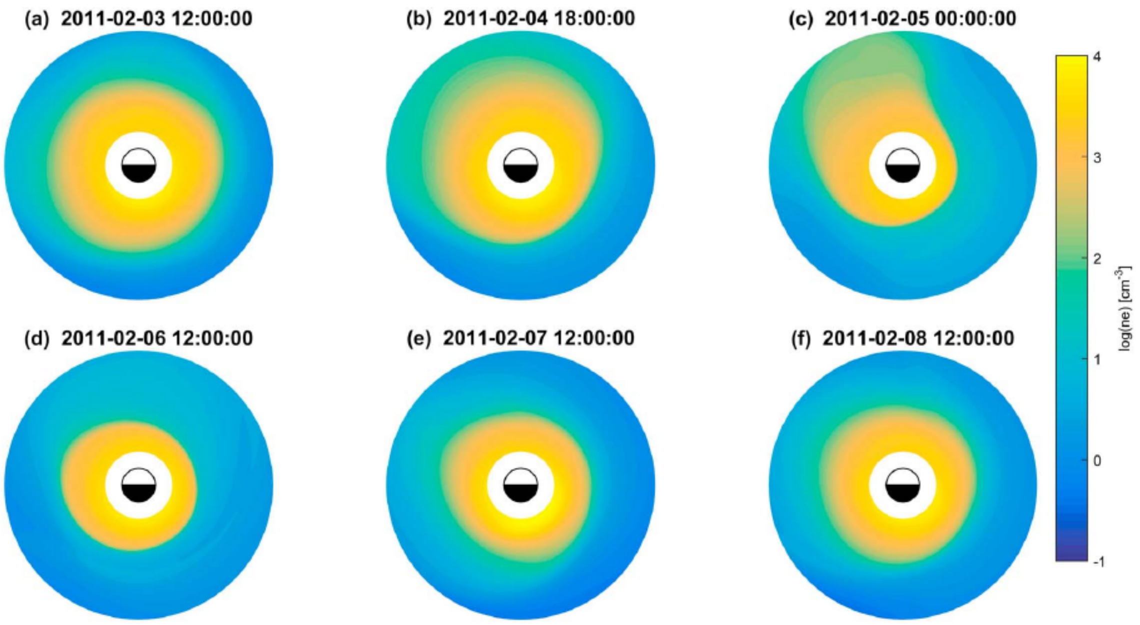
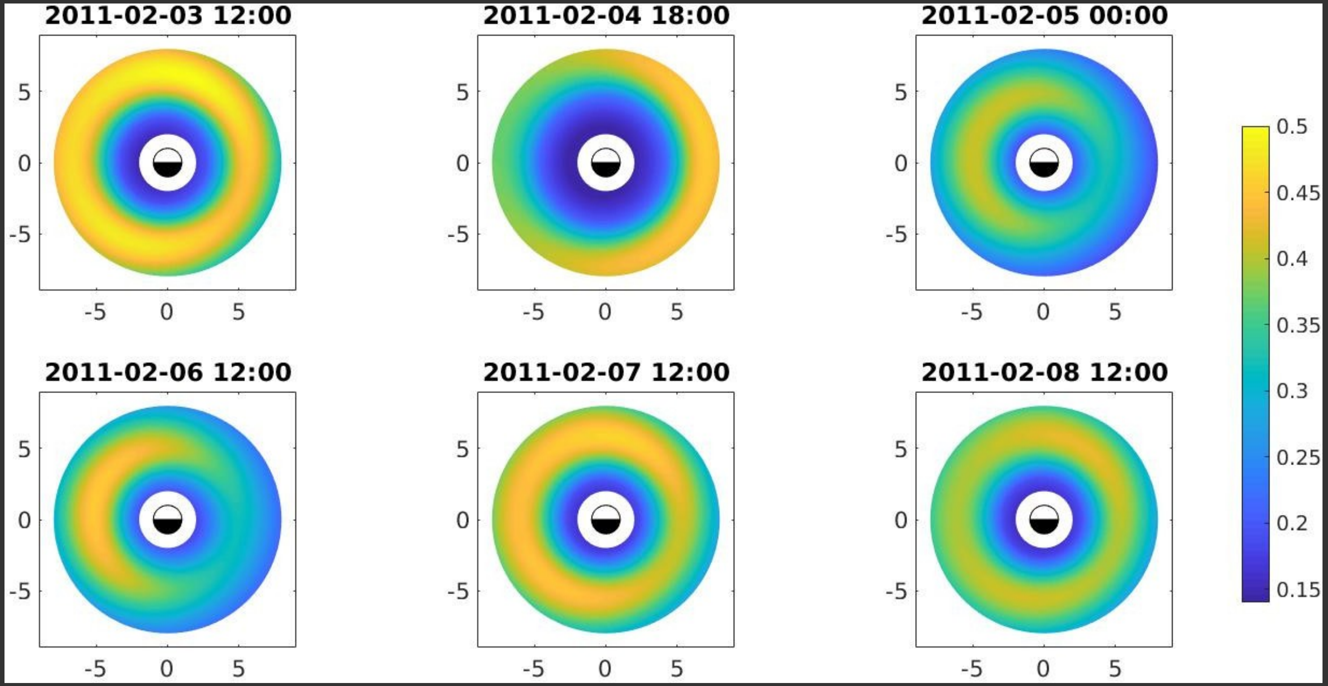



**Figure 6.** A series of panels showing the neural network reconstruction of the global plasma density in the equatorial plane as a function of *L* and MLT at the following times: (a) 2011–02-03/12:00:00 (quiet time before the storm); (b) 2011–02-04/18:00:00 (during the main phase); (c) 2011–02-05/00:00:00 (at the minimum of *SYM-H*); (d) 2011–02-06/12:00:00, (e) 2011–02-07/12:00:00, and (f) 2011–02-08/12:00:00 (recovery phase). The colorbar represents the logarithm of the electron number density in el/cc.

# DEN2D: model for plasmasphere electron density

Estimates the electron density based on history of geomagnetic indexes using a Neural Network.

X. Chu et al. *JGR* (2017)




Estimate of the standard deviation (uncertainty) with the AR method

# Machine Learning benchmark

Table 1: Comparison between different methods on several multidimensional dataset.

| Method | | | NLPD | KM | RECAL | AR | NLPD | KM | RECAL | AR | NLPD | KM | RECAL | AR |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Score | | | NLPD | | | | CRPS | | | | Cal. err. (%) | | | |
| Dataset | Size | Dim. | | | | | | | | | | | | |
| Boston Housing | 506 | 13 | 0.64 | 1.42 | **0.52** | 0.90 | 0.27 | 0.54 | **0.25** | 0.83 | 10.5 | 8.7 | 14.4 | **8.3** |
| Concrete | 1,030 | 8 | 0.55 | 0.84 | **0.50** | 0.62 | 0.26 | 0.28 | **0.22** | 0.26 | 6.0 | 19.7 | 10.6 | **5.3** |
| Energy | 768 | 8 | -0.31 | -0.30 | **-0.12** | -0.23 | 0.13 | **0.08** | 0.14 | 0.13 | 9.6 | 28.4 | 15.6 | **7.7** |
| Kin8nm | 8,192 | 8 | 0.32 | 0.85 | 0.57 | **0.31** | **0.2** | 0.24 | 0.24 | **0.2** | 2.8 | 24.7 | 6.7 | **2.0** |
| Naval propulsion | 11,934 | 15 | -1.52 | 0.84 | -0.82 | **-1.58** | 0.06 | 0.22 | **0.05** | 0.06 | 6.6 | 45 | **2.0** | 4.4 |
| Power plant | 9,568 | 4 | **0.04** | 0.25 | 0.28 | **0.04** | 0.15 | 0.15 | 0.15 | 0.15 | 3.1 | 15.8 | **2.6** | **2.6** |
| Protein | 45,730 | 9 | **1.11** | 1.18 | 1.18 | 1.18 | 0.43 | 0.43 | 0.45 | **0.42** | 7.4 | 8.3 | **1.0** | 7.4 |
| Wine | 1,599 | 11 | 1.24 | **1.16** | 1.20 | 1.29 | 0.47 | **0.43** | 0.46 | 0.47 | 11.9 | 11.3 | **5.5** | 10.6 |
| Yacht | 308 | 6 | -0.44 | 0.39 | **-0.64** | -0.25 | 0.12 | 0.20 | **0.09** | 0.12 | 16.2 | 27.8 | 20.4 | **12.0** |

Lesson learned from ML community: new methods are always tested against standard benchmarks and compared with ‘baseline’ methods

# Take home message

We have devised a method that:

- Estimates the uncertainties associated with single-point outputs generated by a deterministic model, in terms of Gaussian distributions;
- Ensures the optimal trade-off between accuracy and reliability;
- Does not need to run ensembles. It costs as much as training a neural network
- Code available: zenodo.1485608

Space Weather

RESEARCH ARTICLE

10.1029/2018SW002026

**Key Points:**
- We introduce a new method to estimate the uncertainties associated with single-point outputs generated by a deterministic model
- The method ensures a trade-off between accuracy and reliability of the generated probabilistic forecasts
- Computationally cheap model:

## On the Generation of Probabilistic Forecasts From Deterministic Models

E. Camporeale[1,2], X. Chu[3], O. V. Agapitov[4], and J. Bortnik[5]

[1]Center for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands, [2]Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA, [3]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA, [4]Space Sciences Laboratory, University of California Berkeley, Berkeley, CA, USA, [5]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA, USA