Information-theoretical evaluation of probabilistic forecasts

Steven V. Weijs [1,2]

[1] Dept. Water Resources Management, TU Delft, The Netherlands
[2] Now at Environmental Fluid Mechanics (EFLUM), EPFL, Lausanne, Switzerland

Introduction

Evaluation of probabilistic forecasts serves to answer questions related to the "goodness" of the forecasts. Murphy distinguished 3 different types of goodness.

Quality	$<-->$ correspondence observations and forecasts
Value	$<-->$ additional value to user
Consistency	$<-->$ correspondence beliefs and forecasts

The information-theoretical view

Quality can be equated with information. It is the amount of uncertainty about the outcome that is taken away by the forecast. Because logically it should not depend on the user, it is best to define it as a negatively oriented score, which measures the remaining uncertainty after the forecast.

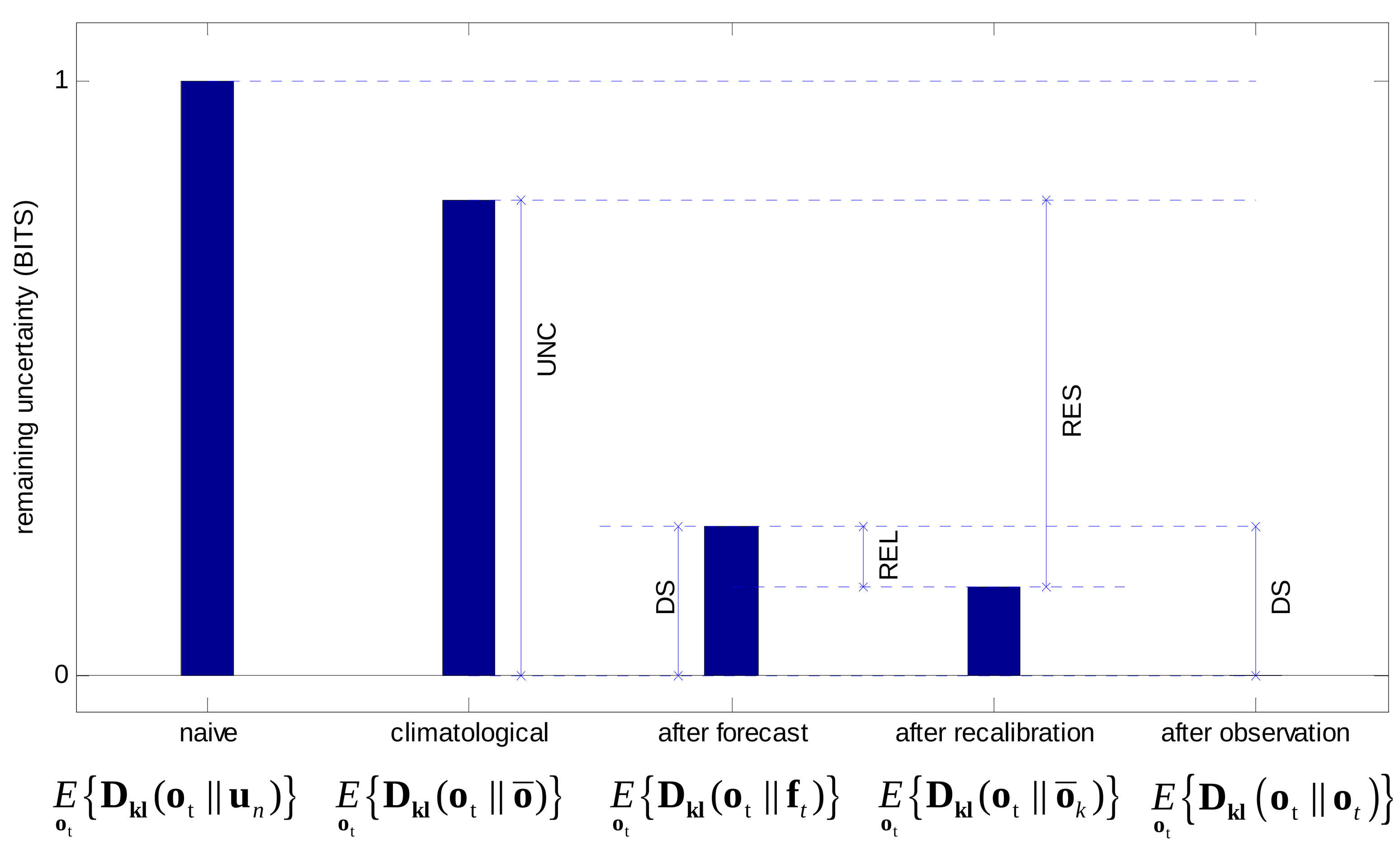
Value depends on the user. It is based on the quality of the forecasts and the user's utility function. We can see it as the useful information.

Consistency is good if the person holding the beliefs is rational.

Kullback-Leibler divergence and its decomposition

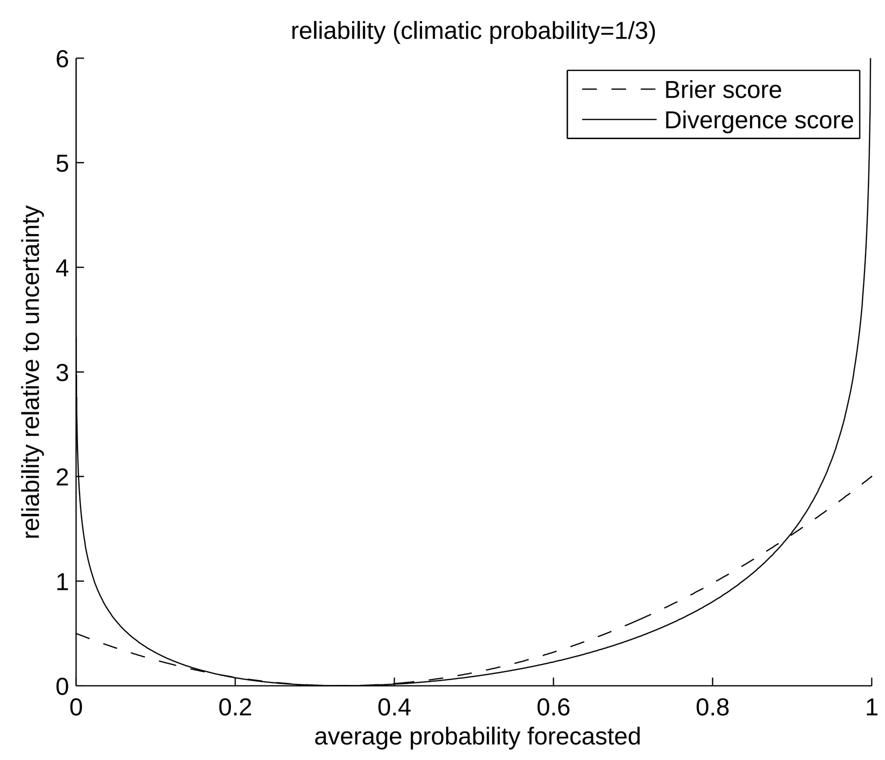
The Kullback-Leibler divergence from the observation to the forecast provides a natural measure for forecast quality. We call this the divergence score (DS), which can be decomposed as REL-RES+UNC (reliability, resolution, uncertainty).

$$\mathrm{DS} = \frac{1}{N}\sum_{t=1}^{N} \mathrm{D}_{KL}(\mathbf{o}_t\|\mathbf{f}_t) = \frac{1}{N}\sum_{k=1}^{K} n_k \mathrm{D}_{KL}(\bar{\mathbf{o}}_k\|\mathbf{f}_k) - \frac{1}{N}\sum_{k=1}^{K} n_k \mathrm{D}_{KL}(\bar{\mathbf{o}}_k\|\bar{\mathbf{o}}) + \mathrm{H}(\bar{\mathbf{o}}) \qquad (1)$$



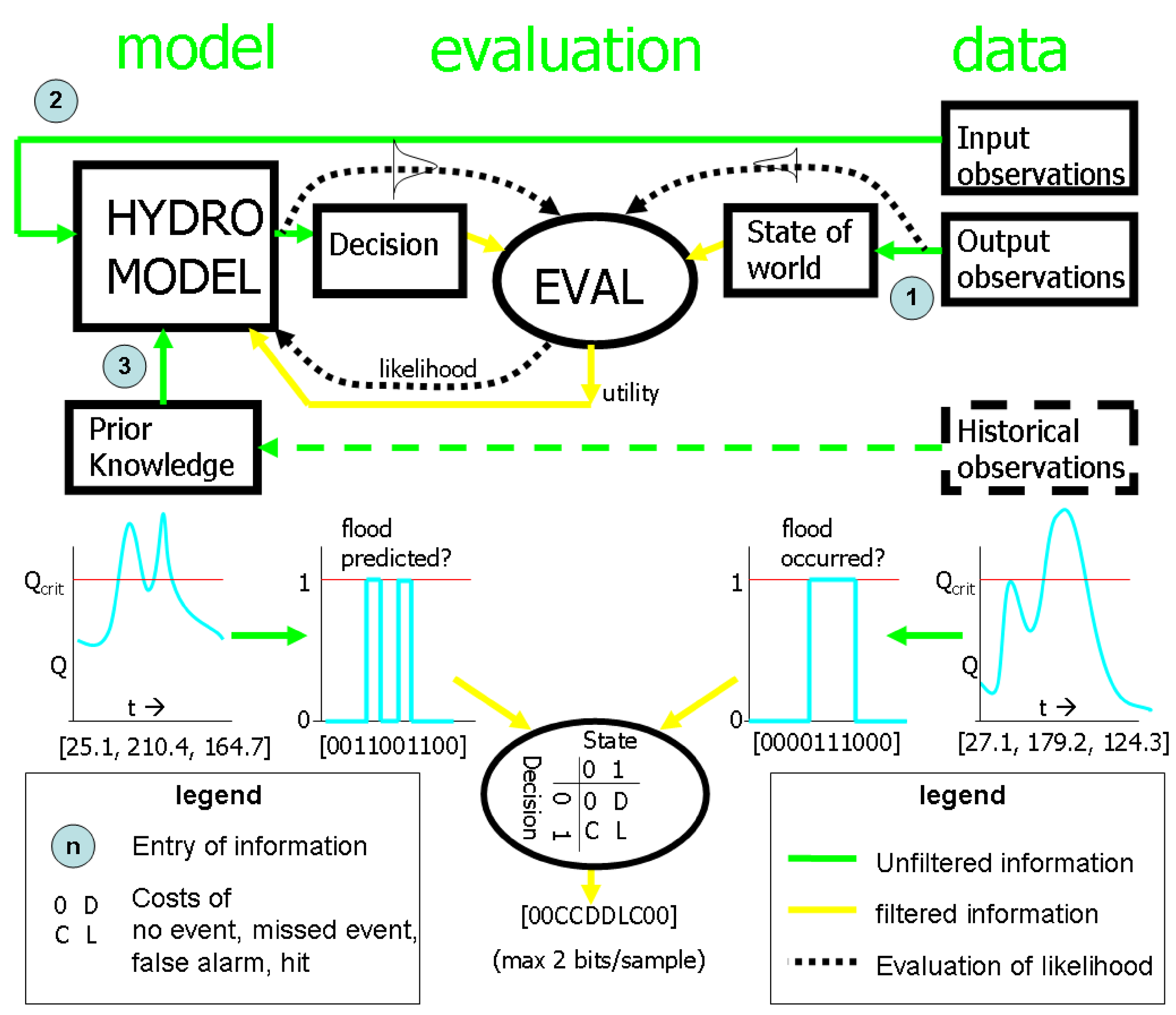
Desirable properties

The Divergence Score, is a **local** and **proper** score. Proper means that the expected score is only maximized if the forecast probabilities match the observed frequencies. Local means that the score does not depend on forecast probabilities assigned to events that did not occur. Whether locality is desirable is debated, e.g. (C)RPS is non-local, but we can now see that non-locality amounts to inference from non-existing information. In contrast to the Brier Score, the DS is not bounded for certain wrong forecasts. This is in line with high stakes one would be willing to risk in decisions that depend on near 0 or 1 probabilities.



Why not optimize useful information?

Many user utility functions are non local, so this would mean drawing conclusions from non-existing information. Furthermore, utility functions also act like a filter on the information that *is* in the observations.



current institution: ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

TU Delft