In the context of quick or flash flood forecasting, the use of rainfall-runoff models is far to be widespread. One possible explanation could be the inability of these models to give acceptable forecasts for operational stakes (are models monsters)?

We assume that it also could be explained, at least partially, by the inadequacy of model’s assessment methods, and especially by the inadequacy of the assessment criteria. First objective of this study is to investigate whether there is a link between various criteria used to assess quality of flood simulations/forecasts. An immediate following objective is to identify the “best” criteria for flood simulation assessment. Key questions are:

- Are all criteria in agreement?
- Does a good performance imply good flood simulations/forecasts? (Meaning of performances?)

**Nash & Sutcliffe Efficiency**

**Description**

- Widely used by modellers’ community
- Often used as a reference for model’s or catchments comparison (see eq. 2)

**Results and discussion**

- High variability with catchment: useful to test most models but not so clear discrimination between other models
- Problem of inter-model discrimination

**Synthesis**

<table>
<thead>
<tr>
<th>Variables (3% of F1 and F2)</th>
<th>71.84%</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>-0.34</td>
</tr>
<tr>
<td>NSE update</td>
<td>-0.63</td>
</tr>
<tr>
<td>Variations</td>
<td>0.22</td>
</tr>
<tr>
<td>CSI</td>
<td>0.36</td>
</tr>
<tr>
<td>Rel. errors</td>
<td>0.76</td>
</tr>
</tbody>
</table>

**Analysis**

On all available results of all 11 catchments and 6 models for 5 selected representative catchments:
- Mean NSE & Mean with update
- Criteria on discharge simulations (large evaluation) (cf. C)
- CSI with low threshold (cf. D)
- Relative errors = « tolerance » (cf. A)

**Main results** (Tab. 3 & Fig. 9)

- NSE, CSI & Variations: positively correlated
- NSE & NSE update: negatively correlated
- Relative errors: no dependence with another criterion

**Results of principal component analysis (Fig. 8)**

- Explains roughly 50% of variance
- Explains roughly 75% of variance
- Explains roughly 25% of variance

**Conclusions**

On an important dataset, 5 categories of criteria were tested (Moulin, 2007) to assess the quality of flood simulations. We were looking for:

- A relationship between these criteria (“are they all in agreement?”) which could lead:
  - either to the choice of a best criteria,
  - or to a better knowledge of the relationship between these criteria.
- A better understanding of the meaning of performance with specific criteria (“Does a good performance imply good flood simulations/forecasts?”)

This work leads to the following conclusions:

1/ There is a link between the 5 selected criteria.
2/ The fifth criteria (on relative errors), appears independent of the other ones.

As a consequence, one of these tested criteria used on its one may appear not as the ideal criterion for flood modelling assessment but as a criterion “not so bad” (a stopgap)? In a court of miracles of hydrologist, boggars can’t be choosers.

The Nash and Sutcliffe Efficiency (widely used) and the criterion on relative errors (easier to understand) may be considered as two one-eyed criteria in a land of blind criteria.