The PREMHyCE project: a comparative evaluation of hydrological models for low-flow forecasting on French catchments

P. NICOLLE, V. ANDREASSIAN
pierre.nicolle@irstea.fr; http://webgr.irstea.fr

Basel, 19 September 2017
CHR Symposium
"Low flows in the Rhine Catchment"
1. Introduction

Context

• In France in 2013: 83% of total withdrawals from rivers

• Water uses affected by water shortages in rivers

• Climate change: perspective of more severe summer low-flows

• Lack of forecasting tool at national scale

Study objectives

• Comparing hydrological models for low-flow ensemble forecasting in a common test protocol

• Assessing the ability of forecasting tools to anticipate low-flow situations (magnitude, maximum lead-time)

• Developing operational low-flow forecasting tool
2. Materials and methods

Catchment set
- 21 with no or limited influence (380 to 4300 km²)
- 11 influenced by dams or water withdrawals (120 to 44000 km²)

Data set
- Daily streamflow (HYDRO French database): 14 to 36 years (1974-2010)
- Daily P, PE, Temp (SAFRAN climate reanalysis): 51 years (1959-2010)
- Daily influences (dam volume, withdrawals): 11 to 25 years (1999-2010)
2. Material and methods

Hydrological models

- Five rainfall-runoff models already used in operational conditions in France
- Daily continuous functioning
- Different modelling approaches (model type, spatial resolution)
- Various number of free parameters
- Influences not systematically taken into account
- Various use of assimilation schemes or statistical correction procedures
2. Material and methods

Ensemble forecasting

- Future meteorological inputs (P, PE, Temp): climatic archive (50 scenarios) →
  Get general results, include severe drought conditions

![Graph showing ensemble forecasting for rainfall and discharge]
2. Material and methods

Evaluation method

- Split-sample test approach (incl. 3-year warm-up)

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>Validation</td>
</tr>
</tbody>
</table>

- Test in hindcasting mode: retrospective run at each time step of period, forecast as in real time
- Calibration method and objective function: choice of the modeller based on his experience with his model

The Seine river at Pont-sur-Seine

![Graph of the Seine river at Pont-sur-Seine showing observation, prevision, and controle](image)
Target variables for low-flows

- Moving average streamflow over 3 days
- Streamflow threshold: Q80 (80% of streamflows above the threshold)
- Low-flows characteristics:
  - Volume deficit
  - Low-flow duration

2. Material and methods
Benchmark and evaluation criteria

- Large selection of efficiency criteria: evaluation of different qualities of hydrological models for forecasting
  - Range of flows (all and low-flows)
  - Relative to the cross of threshold
  - Low-flow characteristics
  - Sharpness
  - Reliability

2. Material and methods
Benchmark and evaluation criteria

- Large selection of efficiency criteria: evaluation of different qualities of hydrological models for forecasting
  - Range of flows (all and low-flows)
  - Relative to the cross of threshold
  - Low-flow characteristics
  - Sharpness
  - Reliability
Benchmark and evaluation criteria

- Large selection of efficiency criteria: evaluation of different qualities of hydrological models for forecasting
  - Range of flows (all and low-flows)
  - Relative to the cross of threshold
  - Low-flow characteristics
  - Sharpness
  - Reliability
Benchmark and evaluation criteria

- Large selection of efficiency criteria: evaluation of different qualities of hydrological models for forecasting
  - Range of flows (all and low-flows)
  - Relative to the cross of threshold
  - Low-flow characteristics
  - Sharpness
  - Reliability

- Mean performances on all catchments
- Target lead times for model evaluation: 7 days and 30 days
Benchmark and evaluation criteria

- Large selection of efficiency criteria: evaluation of different qualities of hydrological models for simulation and forecasting
  - Range of all flow
  - Specific to low flows
  - Relative to the cross of threshold
  - Low flow characteristics
  - Sharpness and reliability (forecasting)

- Models to be compared to benchmark:
  - Natural variability of observed streamflow (Bench)
3. Forecasting results

Lead time: 7 days

- Differences for a few criteria, but difficult to identify a better model
- Significant gain compared to the benchmark
3. Forecasting results

Lead time: 7 days

- Differences for a few criteria, but difficult to identify a better model
- Significant gain compared to the benchmark
- Significant gain when using streamflow assimilation or post-correction methods

- No assimilation or post-correction method
- Use of assimilation or post-correction method
3. Forecasting results

Lead time: 30 days

- Assimilation or post-correction methods less useful with increasing lead-time
- Performance loss with increasing lead-time
- Closer than benchmark but still better
Evaluation of the Useful Forecasting Lead time (UFL)

- Definition:
  Lead time beyond which the model does not bring valuable information compared to the benchmark (natural variability of streamflow)
- Here valuable information if model efficiency at least 20% better than the benchmark efficiency

- UFL depends on efficiency criteria
- UFL varies between:
  - Catchments, but no relation between UFL and low-flow or catchment characteristics
  - Models

![Non-influenced](image1.png)  
![Influenced](image2.png)
4. Operational Forecasting tool

Operational implementation of forecasting tool

- Beta version with GR6J since July 2017
- Ensemble low-flow forecasting at 90 days lead-time
- Test on 70 catchments, 19 on the French part of the Rhine catchment
4. Operational Forecasting tool

Schematic representation of real-time functioning

**Observation**
SAFRAN (P, T°C)
Streamflow
Groundwater observation

**Model calibration**
Model parameters

**Hydrological Models**
GR6J
Mordor
SIM
Gardenia
PRESAGES

**Outputs**
Streamflow Forecasts
Synthesis plot

**Scenarios**
Climatic archive
Medium range meteorological forecast

**Users**
4. Operational Forecasting tool

ill at Didenheim

01-04-2017

Fiche de synthèse de résultats en prévision temps réel

Ill at Didenheim

Situation probability

Q (m³/s)

01 Apr 01 May 01 Jun 01 Jul 01 Aug 01 Sep 01 Oct 01 Nov 01 Dec

0.0 0.2 0.4 0.6 0.8 1.0

0.0 2.0 4.0 6.0 8.0 10.0 20.0 50.0 100.0

CHR Symposium 2017 - "Low flows in the Rhine Catchment"
4. Conclusion and perspectives

Conclusion

• Common protocol to compare and evaluate hydrological models for low-flow forecasting
• No superior model on all catchments or criteria, comparison with benchmark: quantification of the actual value of low-flow forecasting by hydrological models
• Using assimilation or post-correction method less interesting with increasing lead-time
• Simple method to determine Useful Forecasting Lead-time
• Performances quite good on influenced catchments, with various simple methods to account for influences

Perspectives

• Deployment in operational services in 2018
• Integrating other models in operational tool
• Multi-model approach
Thank you!

Further details in:

Contact:
pierre.nicolle@irstea.fr; http://webgr.irstea.fr

Acknowledgements:
Meteo-France, EDF, BRGM, University of Lorraine, AFB
<table>
<thead>
<tr>
<th>Models</th>
<th>GARD</th>
<th>GR6J</th>
<th>MORD</th>
<th>PRES</th>
<th>SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short name used here</td>
<td>GARDENIA</td>
<td>GR6J</td>
<td>MORDOR</td>
<td>PRESAGES</td>
<td>SIM</td>
</tr>
<tr>
<td>Type</td>
<td>Conceptual</td>
<td>Conceptual</td>
<td>Conceptual</td>
<td>Conceptual</td>
<td>Physically-based</td>
</tr>
<tr>
<td>Spatial distribution</td>
<td>Semi-distributed</td>
<td>Lumped</td>
<td>Lumped</td>
<td>Lumped</td>
<td>Distributed</td>
</tr>
<tr>
<td>Number of free-parameters</td>
<td>4 to 9 (+2 to 4 for snowmelt)</td>
<td>6 (+2 : snow routine)</td>
<td>11 (+4: snow routine)</td>
<td>7 (+3 : snow routine)</td>
<td>0</td>
</tr>
<tr>
<td>Calibration criteria</td>
<td>RMSE with ln(Q)</td>
<td>(KGE + KGE)/2</td>
<td>(KGE + KGE)/2</td>
<td>Nash–Sutcliffe with $Q^{0.2}$</td>
<td></td>
</tr>
<tr>
<td>Post-correction method (simulation)</td>
<td>Not used</td>
<td>Not used</td>
<td>Not used</td>
<td>Empirical method (Berthier, 2005)</td>
<td>Quantile/quantile post-treatment</td>
</tr>
<tr>
<td>Assimilation method (forecast)</td>
<td>When a flow discrepancy appears, the model tanks are updated proportionally to their variance</td>
<td>Correction based on error at first time step before forecast, with decreasing effect when lead time increases</td>
<td>Correction based on errors at previous time steps before forecast, with decreasing effect when lead time increases. No update of model stores.</td>
<td>Update of gravitary routing store</td>
<td>No assimilation method but a quantile/quantile post-treatment</td>
</tr>
<tr>
<td>Structure overview: production</td>
<td>Actual evapotranspiration is computed using a non-linear soil capacity. GW exchange is a proportion of the GW flow</td>
<td>A rainfall interception by PE, a non-linear SMA store, an intercatchment GW exchange function</td>
<td>A rainfall excess/soil moisture accounting store; an evaporating reservoir; an intermediate store and a deep store</td>
<td>A soil store, rainfall interception by PE</td>
<td></td>
</tr>
<tr>
<td>Structure overview: transfer</td>
<td>A non linear tank distributes the effective rainfall into runoff and GW recharge. The aquifer is represented by a linear tank.</td>
<td>Two unit hydrograph, two parallel nonlinear routing stores</td>
<td>Direct, indirect and baseflow components are routed using a unit hydrograph (Weibull law)</td>
<td>Two unit hydrographs, two linear routing stores: one for streamflow recession, one for interflow</td>
<td></td>
</tr>
<tr>
<td>References on simulation applications in France</td>
<td>800 to 1000 rivers simulated in France</td>
<td>Garavaglia (2011); Paquet et al. (2013)</td>
<td>Lang et al. (2006a, 2006b)</td>
<td>Vidal et al. (2010b)</td>
<td>Habets et al. (2008)</td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadratic criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KGE</td>
<td>Kling-Gupta Efficiency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2M</td>
<td>Nash-Sutcliffe Efficiency bounded in $[-1;1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-flow quadratic criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2M_i</td>
<td>Nash-Sutcliffe Efficiency calculated with $1/Q$ and bounded in $[-1;1]$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE_{ut}</td>
<td>Root mean square error calculated when observed streamflow is less than $Q_{80}$ threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume-based and temporal criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vdef</td>
<td>Ratio of observed and simulated cumulative annual volume deficits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFD</td>
<td>Ratio of observed and simulated cumulative low-flow duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFD</td>
<td>Ratio of observed and simulated cumulative low-flow duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFD</td>
<td>Ratio of observed and simulated cumulative low-flow duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DatSt</td>
<td>Relative difference between observed and simulated start of annual low-flow period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DatEn</td>
<td>Relative difference between observed and simulated end of annual low-flow period</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POD</td>
<td>Probability of detection, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAR</td>
<td>False alarm rate, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSI</td>
<td>Critical success index, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuous low-flow quadratic and probabilistic criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE_{ut}</td>
<td>Root mean square error calculated when observed streamflow is less than $Q_{80}$ threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DRPS</td>
<td>Discrete Ranked Probability Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume-based and temporal criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vdef</td>
<td>Ratio of observed and simulated cumulative annual volume deficits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LFD</td>
<td>Ratio of observed and simulated cumulative low-flow duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sharp</td>
<td>Mean width of interval defined by 10% and 90% percentiles of forecast distribution when observed streamflow is less than $Q_{80}$ threshold</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cont_ratio</td>
<td>Percentage of observation in the 80% forecasted confidence interval when observed streamflow is less than $Q_{80}$ threshold (80% of observed streamflow should be included in the interval)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Threshold criteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POD</td>
<td>Probability of detection, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAR</td>
<td>False alarm rate, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSI</td>
<td>Critical success index, based on contingency table</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BS_{vig} BS_{cri}</td>
<td>Brier Score with vigilance threshold ($Q_{80}$) or crisis threshold ($Q_{95}$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Modelisation of influences

Source: Payan et al., 2008
Simulation results

- Models similar on average, more difficulties for SIM
- Performances slightly better on non-influenced than influenced catchments
- Significant gain compared to the benchmark
Simulation results
Simulation results

- Mean variability of performances between models:
  - For each catchment, standard variation of performances of models (sdm)
  - Mean of sdm
- Mean variability of performances between catchments:
  - For each model, standard variation of performances on catchments (sdc)
  - Mean of sdc
3. Simulation results

The Meuse river at St-Mihiel

- Catchments where all models simulate overall well streamflows
Other catchments where performances depend more on the models

Performance depends on catchments and models
Simulation results

The Gapeau river at Hyères

~ similar on CSI

RESAGES on Volume deficit

→ Performances of catchments depends on criteria