Evaluation of AWRA-L, WaterDyn and CABLE

Andrew Frost

Water Information Services, Bureau of Meteorology

OzEWEX 1st National Workshop
28th October 2014
Session 3: How well can we trust our models, and how can we be sure?

Propositions

• The simplest model that explains the observations is necessarily the best model. ☑

• All models are wrong, but some are still useful. ☑

• The models are not the main problem, it is the quality of the data and assumptions that go into them. ☑

• Much more effort is needed to objectively assess the performance of alternative models. ☑ ☑ ☑

• We need to stop calibrating our models, it leads to a false sense of security. ✗

• In circumstances where calibration is essential for a model to be useful, we should just use an empirical model (for example, based on data mining or Bayesian methods). ☑ ✗

• We cannot know whether to trust our models. Therefore multi-model ensembles should be standard operational practice, not just a research endeavour. ✗

• In the absence of quantitative knowledge of model inter-dependence, ensemble methods are meaningless. ✗

• Inappropriate values for unconstrained parameters (through calibration or assumption) should remove any trust in predictive ability. ☑

• Talking about ‘physically-based’ models is meaningless when there is not enough data to construct an empirical model. ☑
Motivation: Which model(s) can we use for retrospective runoff, soil moisture, ET, recharge reporting?

BoM reports on water:

- availability
- Use: National Water Account
- current situation
## Model possibilities

<table>
<thead>
<tr>
<th></th>
<th>WaterDyn (AWAP)</th>
<th>CABLE-SLI</th>
<th>AWRA-L</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Developer</strong></td>
<td>CSIRO/BOM/BRS</td>
<td>CSIRO/BOM +</td>
<td>CSIRO/BOM (WIRADA)</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Monitoring terrestrial water balance</td>
<td>Landsurface scheme for ACCESS</td>
<td>Water reporting and monitoring</td>
</tr>
<tr>
<td><strong>Soil layers</strong></td>
<td>2 (depth, saturated volumetric water content)</td>
<td>10 (saturated hydraulic conductivity, field cap, etc etc)</td>
<td>3 (sat. hydraulic conductivity, % AWC)</td>
</tr>
<tr>
<td><strong>Calibration</strong></td>
<td>Parameter sensitivity to 6 catchments in Murrumbidgee</td>
<td>Calibration to derived ET (50 catchments) and flux tower data</td>
<td>Streamflow over ~300 catchments</td>
</tr>
</tbody>
</table>

**Others:** rainfall-runoff models, empirical methods, etc.
AWRA-L conceptual structure
What to do?

Much more effort is needed to objectively assess the performance of alternative models.

- **Catchment evaluation**
  - **Streamflow** – 780 unimpaired catchments (Zhang et al, 2013)

- **Point testing**
  - **Soil moisture** –
    - SASMAS (Rüdiger et al, 2007)
  - **Benchmarking system**: Warren (2012) ++
Streamflow: Unimpaired catchments

- AWRA-L performs better due to calibration and conceptual structure.
- Locally calibrated rainfall runoff models provide benchmark.
- AWRA-L is close to locally calibrated models in validation.

We need to stop calibrating our models, it leads to a false sense of security.

Soil moisture

Profile (0-90cm)

OzNET

SASMAS

AMSR-E

CosmOz

Monthly Pearson's r

Monthly Pearson's r (val.294)
In circumstances where calibration is essential for a model to be useful, we should just use an empirical model (for example, based on data mining or Bayesian methods).
How well can we trust our models?

- Hydrological assessment not undertaken routinely
  - Especially for variables other than streamflow
- Operational/community comparison and demonstration of models ....through PALS?
  - standard tests against published unimpaired catchment and point data
- Objective assessment is hard
  - different forcing data, calibration techniques, scales, soil store depths, purposes of models etc
- **First step:** set up system for ongoing benchmarking and use as analogue check that model is fit for intended purpose