There are no monsters, only models with huge uncertainties!

George Kuczera, Dmitri Kavetski, Mark Thyer, Benjamin Renard and Stewart Franks
Introduction

• Monster catchments are “deviants” whose behaviour is far from what the hydrologist expects.
• Should we shun them?

Absolutely not!

• Monster catchments expose the frailties of our science
• They should spur us to improve our hydrologic science

A few thoughts on two questions:
• What has uncertainty to do with monsters?
• What can we really learn?
Linking Reality and Models

Reality

True Input(s) → True processes → True response(s)

Input errors
- e.g. rainfall measurement + sampling errors

Structural errors
- e.g. lumped nature of CRR models
- wrong process hypothesis

Response errors
- e.g. rating curve errors

Observed Input(s) → Conceptualized processes → Observed response(s)

Model
Input (Forcing) Data Uncertainty

- Rainfall can be highly variable, dependent on climate and topography.
- Areally-averaged rainfall is subject to potentially very large errors.
- Inadequate sampling of the rainfall field by a limited number of raingauges.

Rainfall can be highly variable, dependent on climate and topography.
Areally-averaged rainfall is subject to potentially very large errors.
Inadequate sampling of the rainfall field by a limited number of raingauges.

From Linsley et al, 1958 (fig 3-7)
Output (Response) Data Uncertainty

- System responses can be very diverse (e.g., integrated response such as streamflow, point responses such as soil moisture).
- Errors in observing system responses can vary widely.
- In the case of streamflow, the accuracy of discharge estimates is affected by instrument and rating curve error. Can expect ~10-20% std errors.

System responses can be very diverse (e.g., integrated response such as streamflow, point responses such as soil moisture). Errors in observing system responses can vary widely. In the case of streamflow, the accuracy of discharge estimates is affected by instrument and rating curve error. Can expect ~10-20% std errors.
Model (Structural) Error

- All environmental models make simplifying assumptions
- Lumped approximation of distributed (heterogeneous) nonlinear system → Flux dynamics based on averaging cannot reproduce true response
- Sometimes assumptions are grossly wrong e.g. omit significant process
- Environmental models are effectively stochastic predictors (even if true forcings were known, the model would not be able to reproduce the true response)
- Model error is hard to characterize:
  - Additive random error on output or on state (classic approach)
  - Vary parameters stochastically over some characteristic time scales
Living with Errors

• If we don’t give due respect to errors we can make “monsters” out of catchments and fools of ourselves

• Significant errors demand a framework that:
  – proposes hypotheses about significant hydrologic processes and sources of error
  – provides an ‘honest’ inference conditioned on available data
  – vigorously scrutinizes hypotheses using all available evidence

• Bayesian approach can support such a framework:
  – Capable of dealing with a wide variety of error types
  – Posterior diagnostics to check hypotheses
  – We call it BAyesian Total Error Approach

• Bayesian analysis is not automatic → It demands creativity and vigour especially in the feedback loop
A Bayesian Hierarchical CRR Model

Posterior (parameters, latent | observations)
\[ \alpha \ p \ (\text{observed outputs | CRR parameters, latent, observed inputs}) \times p \ (\text{latent | hyperparameters}) \times \text{prior (hyperparameters, parameters)} \]
Role of Epochs in Hierarchical Structure

Input error epochs

$$\phi \leftarrow p(\phi \mid \Phi)$$

Model error epochs

$$\lambda \leftarrow p(\lambda \mid \Lambda)$$
A Tale of Two Monsters: Horton Catchment

Area: 1920 km²
Strong rainfall gradient
Avg. annual rainfall: 660-1270mm
Avg. annual runoff: 96mm
(Runoff coefficient ~ 0.11)
CRR Model – GR4J

- Simple, parsimonious model
- Evaluated on large sample of catchments: semi-arid, temperate, humid
- 4 parameters
  - $x_1$: Production Store Capacity
  - $x_2$: Water exchange coefficient
  - $x_3$: Routing Store Capacity
  - $x_4$: Routing timing parameters

Is Horton Catchment a Monster?

Fit GR4J maximizing Nash-Sutcliffe statistic (equivalent to minimizing sum of squares of errors SLS):

- Gross errors around peaks
- Recessions poorly matched
Is Horton Catchment a Monster?

- Estimation of GR4J parameter X1 is accurately wrong.
- Inference significantly different depending on which raingauge used → “Fitting to rainfall errors”
- Not much chance regionalizing CRR parameters if non stationary or sensitive to rainfall
Rainfall Errors: How Big?

Rainfall errors can be large when using single raingauge.
Input Error Model

- Rainfall errors characterized using rainfall multipliers $\phi$

- Rainfall: $r_t = \varphi_t \tilde{r}_t$, $\log(\varphi_t) \sim N(\mu_r, \sigma_r^2)$

- Daily or storm epochs
  - Single multipliers for days with significant rainfall
  - Single multiplier for each storm epoch
Heteroscedastic discharge rating error model

\[ \tilde{Y}_t = \hat{Y}_t + \gamma_t, \quad \gamma_t \sim N(0, \sigma^2_\gamma) \]

\[ \sigma_\gamma = 0.4 + 0.086 \cdot \hat{Y}_t \]
Methodology

• Standard Least Squares (SLS)
  – No input error/structural errors, homoscedastic output errors
• Weighted Least Squares (WLS)
  – No input error/structural errors, heteroscedastic output errors
• BATEA
  – Input error, no structural errors, heteroscedastic output errors
  – Daily Rainfall Multipliers (DAY)
  – Storm Rainfall Multipliers (STORM)
• Calibration Period
  – Different gauges and different time periods
  – Four gauges to 2 years
  – Two gauges to 5 years (054021, 054138)
• Validation Period
  – 13 years – Two gauges
Comparison of Parameter Estimates

BATEA parameters estimates are more consistent, albeit, more uncertain than WLS, SLS.
Posterior Diagnostics

- Posterior diagnostics for checking error model hypotheses (latent variables properties)
- Need care to avoid artifacts

- Standard validation schemes (e.g. split-sample test) based on a single ‘optimal’ simulated series
- Want to check model and error hypotheses using predictive distribution for data not used in calibration
- Counting the number of observations inside interval bounds is far from sufficient…
Posterior Diagnostics: Predictive Uncertainty in Calibration

- Majority of data inside probability limits,
- Probability limits reasonable
- Rainfall multipliers are estimated
Posterior Diagnostics: Predictive Uncertainty in Validation

SLS
- misses some peaks
- data outside probability limits

BATEA
- majority of data inside probability limits
- probability limits are wide
- rainfall multipliers are unknown so sampled from hyperdistribution
Posterior Diagnostics: Predictive QQ plots

- Counting the number of observations inside interval bounds is far from sufficient.
- Data \( (q_t), t=1\ldots n \) from a validation period should be consistent with the predictive distribution \( (F_t), t=1\ldots n \).
- The \( p \)-values \( p_t = F_t(q_t) \) should follow a uniform distribution on \([0,1]\).
- QQ-plot of observed \( (p_t), t=1\ldots n \) vs. theoretical quantiles from a \( U[0,1] \) should be close to the 1:1 line.

\[
\text{Quantile of observed } p\text{-values} \quad \begin{cases} 1:1 \\ \text{BATEA} \\ \text{SLS} \end{cases}
\]

- 13 years data
- Runoff > 2mm
- Underprediction
- Observed runoff has \( p \) value \( \sim 1 \)

\[
\text{Predictive distribution}
\]
Rainfall multipliers were assumed to be independent and log-normally distributed

\[ r_t = \varphi_t \tilde{r}_t, \quad \log(\varphi_t) \sim N(\mu_r, \sigma_r^2) \]
**Posterior Diagnostics: Residual Discharge**

- Lots of zero or near-zero daily runoff
- Assumed discharge error model with std dev = 0.4 mm at near-zero runoff
A Tale of Two Monsters: Vesonne Catchment

- Located in Rhone Region
- Area: 72.4 km²
- Rainfall-dominated hydrology (regime pondere)
- Use GR4J model

Can Vesonne be a ‘monster’?
SLS Calibration

NS=0.88
SLS Parameters

X2 negative → water being exported from catchment. Major groundwater transfers occur through the sandstone aquifer? (Nicolas Le Moine)
SLS Validation Predictive Uncertainty

Predictive uncertainty consistently overestimated

Red = contribution of parameter uncertainty
BATEA Calibration with Storm Epoch Rainfall Multipliers

NS=0.83 but 10% std error on discharge

X2 close to zero. Why?
BATEA Calibration with Storm Epoch Rainfall Multipliers

\[ r_t = \varphi_t \tilde{r}_t , \quad \log(\varphi_t) \sim N(\mu_r, \sigma_r^2) \]

Rainfall error multipliers approx log-normal
BATEA Calibration with Storm Epoch Rainfall Multipliers

\[ r_t = \varphi_t \tilde{r}_t , \quad \log(\varphi_t) \sim N(\mu_r, \sigma^2_r) \]

Mean of multiplier $\sim 0.57 \rightarrow$ Gauge overestimates rainfall by 43%?

Modal CV of 0.4 $\rightarrow$ rather big
BATEA Validation with Storm Epoch Rainfall Multipliers

![Graph showing Total Error over time steps]

---

**THE UNIVERSITY OF NEWCASTLE AUSTRALIA**
Beyond day 600 observed runoff is zero for extended period even though significant rainfall
Prediction limits are well above zero!
Inconsistent with calibration
Predictive QQ plots

**BATEA rainfall**

Consistent overprediction

**SLS**

Overestimated predictive uncertainty
BATEA Calibration with X2 Stochastic Parameter

No rainfall error multipliers

X2 parameter randomly sampled at start of each storm epoch

\[ \log_e (-X2) \sim N(m_{X2}, s_{X2}^2) \]

\( NS = 0.89 \)
BATEA Calibration with $X_2$ Stochastic Parameter

$$\log_e(-X_2) \sim N(m_{X_2}, s_{X_2}^2)$$
BATEA Validation with X2 Stochastic Parameter

Red = contribution of X2 uncertainty

Extended zero runoff failure
Predictive QQ plots

SLS

BATEA rainfall

BATEA X2

BATEA X2 > BATEA rainfall or SLS
BATEA Calibration: Rainfall Multipliers and Stochastic X2

Rainfall error multipliers + stochastic X2:
- Can mimic each other
- Need more information to break interactions

Image of absolute posterior correlation

Rainfall multiplier block

Stochastic CRR parameter block
A Reflection

• Horton and Vesonne are not really “monsters”:
  – Fitting to rainfall errors
  – Difficulty closing water balance

• There are no monsters, just hydrologists hampered by a lack of information:
  – Need enough observed fluxes to close water balance
  – Need enough data/information to get closure on errors

• Bayesian framework may be useful to process information about errors and scrutinize hypotheses
There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
</table>

There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer models</td>
<td>King GIGO (garbage in, garbage out) rules supreme</td>
</tr>
</tbody>
</table>
There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer models</td>
<td>King GIGO (garbage in, garbage out) rules supreme</td>
</tr>
<tr>
<td>+ Smart, curious hydrologist</td>
<td>Off with GIGO’s head. At least we can toss out the garbage</td>
</tr>
</tbody>
</table>
## There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer models</td>
<td>King GIGO (garbage in, garbage out) rules supreme</td>
</tr>
<tr>
<td>+ Smart, curious hydrologist</td>
<td>Off with GIGO’s head. At least we can toss out the garbage</td>
</tr>
<tr>
<td>+ Process and data closure on water balance</td>
<td>Beware of “monsters”</td>
</tr>
</tbody>
</table>
There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer models</td>
<td>King GIGO (garbage in, garbage out) rules supreme</td>
</tr>
<tr>
<td>+ Smart, curious hydrologist</td>
<td>Off with GIGO’s head. At least we can toss out the garbage</td>
</tr>
<tr>
<td>+ Process and data closure on water balance</td>
<td>Beware of “monsters”</td>
</tr>
<tr>
<td>+ Information on output errors</td>
<td>Cannot get closure on errors but insight on input/model error</td>
</tr>
</tbody>
</table>
## There is “No Free Lunch” in Hydrology

<table>
<thead>
<tr>
<th>Resources</th>
<th>Insights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer models</td>
<td>King GIGO (garbage in, garbage out) rules supreme</td>
</tr>
<tr>
<td>+ Smart, curious hydrologist</td>
<td>Off with GIGO’s head. At least we can toss out the garbage</td>
</tr>
<tr>
<td>+ Process and data closure on water balance</td>
<td>Beware of “monsters”</td>
</tr>
<tr>
<td>+ Information on output errors</td>
<td>Cannot get closure on errors but insight on input/model error</td>
</tr>
<tr>
<td>+ Information on input errors</td>
<td>Maybe get enough closure to “see” model errors and subject competing models to fiery scrutiny</td>
</tr>
</tbody>
</table>
Finito
BATEA Calibration with Storm Epoch Rainfall Multipliers

The diagram shows the total error over time steps, with rainfall intensity plotted against time. Observed data is represented by markers, and the 50% confidence interval is indicated by lines.