Gaussian Processes and Hydrologic Uncertainty

Omar Wani

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.
Uncertainty in Models

Sources of uncertainty – a general classification:

- Observational uncertainty
- Model structure uncertainty
- Parametric uncertainty
Uncertainty Estimation

Using complementary models to assign prediction intervals.
Post Processors – e.g. SKIBLUE
Post Processors – e.g. SKIBLUE
Post Processors – e.g. SKIBLUE

Figure 2.1 Development of SKIBLUE

$$v = \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix} \quad (\text{past data})$$

$$v_t (\text{prediction step})$$

$$knn \to \infty$$

$$r_p | r_p < r$$

$$r \to 0$$

Development of SKIBLUE
Post Processors – e.g. SKIBLUE

- Observed Water Level
- Modelled Water Level
- 50% Prediction Interval
- 90% Prediction Interval

Water Level (m)

Time [dd.mm.yyyy]


Separating Uncertainty - Using Gaussian Processes (GP)

\[ Y_{obs} = y_M + E + B \]

Mimics the effects of meas. errors

Gaussian, independent

Mimics the effects of Inp./Str. errors

Dietzel and Reichert 2012
Del Giudice et al. 2013
GP as Model Bias

\[ Y_{obs} = \mathbf{y}_M + \mathbf{E} + \mathbf{B} \]

Ornstein–Uhlenbeck process with \( \mu = 0 \)

Bias is autocorrelated

\[ db_M(t) = -\frac{B_M(t)}{\tau} dt + \sqrt{\frac{2}{\tau}} \left( \frac{\sigma^2}{B_{ct}} + (\kappa x(t-d))^2 \right) dW(t) \]

Dietzel and Reichert 2012
Del Giudice et al. 2013
Parameter Estimation

Likelihood function

\[ \mathcal{L}_M(y_o|\theta, \psi, x) = \frac{(2\pi)^{-\frac{n}{2}}}{\sqrt{\text{det}(\Sigma(\psi, x))}} \exp \left( -\frac{1}{2} \left[ y_o - y_M(\theta, x) \right]^T \Sigma(\psi, x)^{-1} \left[ y_o - y_M(\theta, x) \right] \right) \]

Inference using Bayes theorem

\[ f_{post}(\theta, \psi|y_o, x) = \frac{f(\theta, \psi) \mathcal{L}_M(y_o|\theta, \psi, x)}{\int \int f(\theta, \psi) \mathcal{L}_M(y_o|\theta, \psi, x) \, d\theta \, d\psi} \]
Application to Rawthey

Division into three sub catchments and then manuel calibration
Application to Rawthey
Application to Rawthey

Rawthey - Two Bucket Model, Calibration + Validation

Nash Efficiency (validation) = 0.83, Observational Coverage = 92%
- Model
- Observations
- Rainfall
- 90% Uncertainty Bands
- Calibration Validation divide

Cumecs vs Time Step (Hr)
Application to Rawthey

Posterior parameter distributions
\[
\text{Prob}(Z | \theta) = \text{Prob}(Z_{t_1}, \ldots, Z_{t_n} | \theta) = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(Y_{t_1}, \ldots, Y_{t_n} | \theta) \, dY_{t_1} \cdots dY_{t_n}
\]

\[Z_t = \begin{cases} \text{high,} & Y_t > y_{\text{threshold}} \\
\text{low,} & Y_t \leq y_{\text{threshold}} \end{cases}\]
GP as a Binary Likelihood Function

\[ \text{Prob}(Z \mid \theta) = \text{Prob}(Z_{t_1}, \ldots, Z_{t_n} \mid \theta) \]

\[ = \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(Y_{t_1}, \ldots, Y_{t_n} \mid \theta) \, dY_{t_1} \cdots dY_{t_n} \]

\[ Z_t = \begin{cases} \text{high,} & Y_t > y_{\text{threshold}} \\ \text{low,} & Y_t \leq y_{\text{threshold}} \end{cases} \]
GP as a Binary Likelihood Function

\[
\text{Prob}(\mathbf{Z} \mid \theta) = \text{Prob}(Z_{t_1}, \ldots, Z_{t_n} \mid \theta)
\]

\[
= \int_{l_1}^{u_1} \cdots \int_{l_n}^{u_n} p(Y_{t_1}, \ldots, Y_{t_n} \mid \theta) \, dY_{t_1} \cdots dY_{t_n}
\]

\[
Z_t = \begin{cases} 
\text{high,} & Y_t > y_{\text{threshold}} \\
\text{low,} & Y_t \leq y_{\text{threshold}}
\end{cases}
\]
GP as a Binary Likelihood Function – Adliswil

- 7.8 km²
- 18000 Einwohner
Results on synthetic observations

<table>
<thead>
<tr>
<th>Data</th>
<th>Nash–Sutcliffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0.513</td>
</tr>
<tr>
<td>Posterior continuous</td>
<td></td>
</tr>
<tr>
<td>Posterior binary</td>
<td></td>
</tr>
</tbody>
</table>

![Graph depicting discharge over time](image-url)
Results on synthetic observations

<table>
<thead>
<tr>
<th>Data</th>
<th>Nash–Sutcliffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0.513</td>
</tr>
<tr>
<td>Posterior continuous data</td>
<td>0.797</td>
</tr>
<tr>
<td>Posterior binary data</td>
<td></td>
</tr>
</tbody>
</table>
Results on synthetic observations

<table>
<thead>
<tr>
<th>Data</th>
<th>Nash–Sutcliffe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior</td>
<td>0.513</td>
</tr>
<tr>
<td>Posterior continuous</td>
<td>0.797</td>
</tr>
<tr>
<td>Posterior binary</td>
<td>0.779</td>
</tr>
</tbody>
</table>
Benefits and Challenges in using Gaussian Processes as Bias

- Captures the correlation structure rigorously
- Marginalization/conditioning analytically feasible
- Only two meta-parameters - Inferred in the Bayesian Framework
- Extension to limited data scenarios (Like binary observations)

- Heteroscedasticity necessitates transformation
- Computationally expensive
- Identifiability problem: Input and structural uncertainty
Thank you!

Wani, O., 2014
Extra slides
Herent - SKIBLUE

For 90% Prediction Interval:
- Quantile Regression
- SKIBLUE
- UNEEC
- Expected PICP

For 50% Prediction Interval:
- PICP

For 90% Prediction Interval:
- MPI (m)

For 50% Prediction Interval:
- MPI (m)

Yeaton 1998 1999 2000
Timestep

eawag aquatic research

ETH Zürich