Uncertainty, precision and reliability of eco-hydrological models

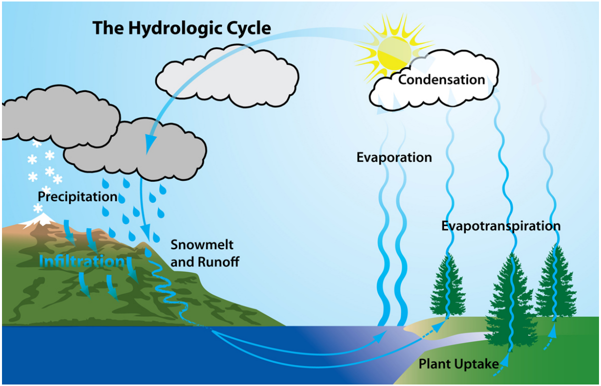
Damian Mingo Ndiwago[1,2]
Supervisors : Jack S. Hale[1]
Christophe Ley[2]
Remko Nijzink[3]
Stan Schymanski[3]

[1]University of Luxembourg, [2]Ghent University, [3]Luxembourg Institute of Science and Technology

May 21, 2021

Overview

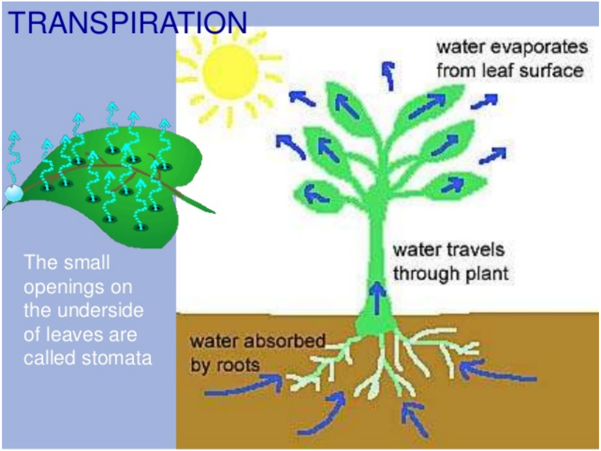
Introduction



Hydrological cycle (*Retrieved from: https://www.nj.gov/drbc/hydrological/*)

Ecohydrology studies the role and movement of water between plants and their surroundings.

Introduction



Water loss through the leaves

(Retrieved from:https://www.aplustopper.com/transpiration-icse-solutions-class-10-biology/)

Importance of eco-hydrology

Sustainable water management can abate drought.

- A sound understanding of eco-hydrology is essential for the sustainable management of water resources (Zalewski 2002).
- Good understanding of the role of plants across different scales is pivotal for designing sustainable water management strategies (Asbjornsen et al. 2011).

Research aims

- There are many models as there are eco-hydrologists.
  - Eco-hydrologists want to choose the most reliable model in predicting the current and future state of the environment.
- Scientific questions
  - How can we rigorously compare and choose between competing models?
  - How can we balance parsimony with goodness of fit?
  - How can we compare models of fundamentally different type?

Eco-hydrology models

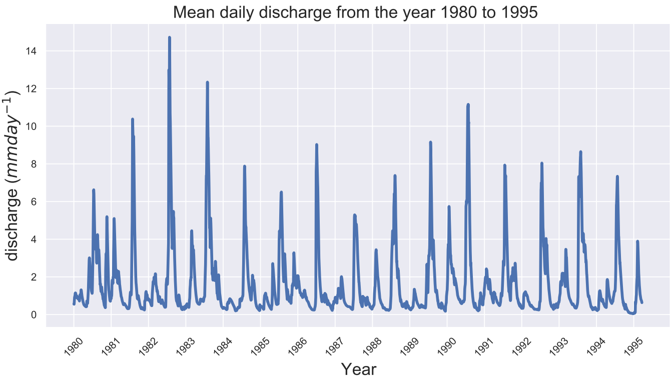
- Ecohydrological models are based on physical processes such as evapotranspiration.
- They can be physical, bucket, or optimality models.
  - Bucket models represent various layers of soil as buckets.
  - Optimality based models require data only for validation in contrast to bucket models that require data for parameter identification too.

Review of literature and challenges

Several approaches have been used to study eco-hydrological processes such as:

- Bayesian inference has been used extensively in eco-hydrological modelling for parameter identification.
  - Recent studies include: (Yang et al., 2016; Tang et al., 2018, 2019).
- There are few studies on Bayesian model comparison and selection in Eco-hydrology. Some studies are:
  - Marshall, Nott, and Sharma (2005).
  - Volpi et al., 2017, Brunetti et al., 2017, 2019.

Potential data sets | Camels data set



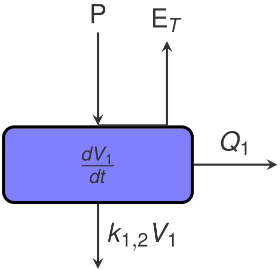
Mean discharge based on the data published by (Addor et al., 2017).

Example of a bucket model | Hydrologiska Byråns Vattenbalansavdelning model (HBV)

The HBV model describes the movement of water by various processes from one layer (bucket) to another.

- **P : precipitation**
- **E_T : Evapotranspiration**
- Q_1 : discharge
- k_1 : recession coefficient
- k_{12} : recession coefficient

$$\frac{dV_1}{dt} = P - E_T - K_{1,2}V_1 - Q_1$$
$$= P - E_T - K_{1,2}V_1 - K_1V_1$$



HBV as a system of Ordinary differential equations

$$V_1'(t) = P - E_T - k_1 V_1$$

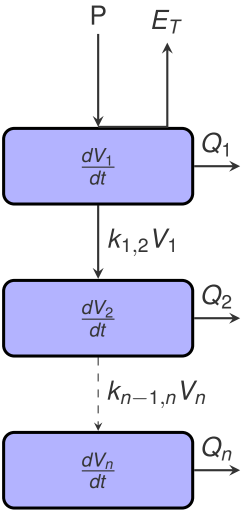
$$V_2'(t) = k_{1,2} V_1 - k_2 V_2$$

$$\boldsymbol{V_n'(t) = k_{n-1,n} V_{n-1} - k_n V_n}$$

$$\mathrm{V}_1'(0) = V_2'(0) = V_n'(t) = 0$$

n = 2,...,N

$$Q_1 = k_1 V_1, \quad Q_n = k_n V_n$$



$$\boldsymbol{Q_T = \sum_{i=1}^{n} k_i V_i}$$

$$i = 1, .., n$$

Research strands

1. Bayesian model comparison and selection.
2. Prior impact assessment.

Bayesian inference

- For a given model M_m, with parameter vector θ_m and y observed data.
- We include the model indicator M_m since we are concerned with model selection.

Theorem (Bayes theorem)

$$\underbrace{P(\theta_m|M_m,y)}_{\text{posterior}} = \frac{\overbrace{P(y|\theta_m,M_m)}^{\text{likelihood}}\,\overbrace{P(\theta_m|M_m)}^{\text{prior}}}{\underbrace{P(y|M_m)}_{\text{marginal(averaged)likelihood}}}$$

$$= \frac{P(y|\theta_m,M_m)P(\theta_m|M_m)}{\int P(y|\theta_m,M_m)P(\theta_m|M_m)d\theta_m}$$

Bayesian model comparison

1. Bayes factor.
   - Based on the marginal or averaged likelihood.
2. Information theoretic criteria.
   - Based on predictive accuracy.
   - Examples includes:
     - Deviance information criterion (DIC).
     - Widely applicable Akaike information criterion (WAIC).

Bayesian model comparison | Bayes factor (BF)

- The Bayes factor of model 1 (M_1) compared to model 2 (M_2) is:

$$\text{BF}_{12} = \frac{p(y|M_1)}{p(y|M_2)} = \frac{\int P(y|\theta_1, M_1)P(\theta_1|M_1)d\theta_1}{\int P(y|\theta_2, M_2)P(\theta_2|M_2)d\theta_2}$$

- The Bayes factor is obtained by taking the ratio of the posterior distributions of the models being compared.
- $\text{BF}_{12} > 1$ is in favour of model 1.
- There is table for interpretation of Bayes factors in use.

There is usually no analytic solution for the marginal likelihood. Thus, we use sampling based methods:

1. Naive Monte Carlo
2. Harmonic mean estimator (Newton & Raftery, 1994).
3. Generalized harmonic mean estimator (Gelfand & Dey, 1994).
4. Thermodynamic integration (Ogata, 1989; Gelman & Meng, 1998).
5. Bridge sampling (Meng Wong, 1996).

Bayesian model comparison | Bridge sampling estimator

The marginal likelihood is defined as based on the bridge sampling estimator by (Frühwirth-Schnatter 2004; Meng and Wong 1996).

$$p(y|M_m) = \frac{\mathbb{E}_{g(\theta)} h(\theta) p(\theta) p(y|\theta)}{\mathbb{E}_{p(\theta|y)} h(\theta) g(\theta)}$$

To compute the marginal likelihood based on sampled values, we use

$$\hat{p}(y|M_m) = \frac{\frac{1}{N_1} \sum_{n=1}^{N_1} h\left(\theta^{*(n)}\right) p\left(\theta^{*(n)}\right) p\left(y|\theta^{*(n)}\right)}{\frac{1}{N_2} \sum_{n=1}^{N_2} h\left(\theta^{(n)}\right) g\left(\theta^{(n)}\right)}$$

$$\theta^{*(n)} \sim g(\theta), \quad \theta^{(n)} \sim p(\theta|y)$$

Bayesian model comparison | Examples

1. Data was generated according to:

$$y_i = \alpha_0 + 0.95 x_{i_1} + 0.12 x_{i_4} + \varepsilon_i$$
$$\varepsilon_i \sim N(0, 2.0)$$
$$i = 1, ..., 30$$

2. α_0, x_{i_1} and x_{i_2} are orthogonal Legendre polynomials.
3. Three models were fit to the data and model comparison performed with BF.

$$y_i = \alpha_1 x_{i1} + \alpha_4 x_{i4} + \varepsilon_i \tag{M$_1$}$$

$$y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_4 x_{i4} + \varepsilon_i \tag{M$_2$}$$

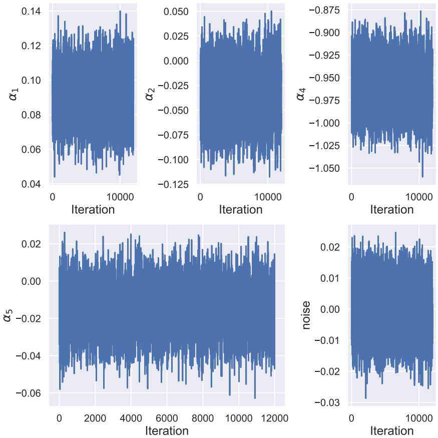
$$y_i = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_4 x_{i4} + \alpha_5 x_{i5} + \varepsilon_i \tag{M$_3$}$$

Priors

$\alpha_i \sim N(0, 0.3)$ where $i \in \{1, 2, 4, 5\}$

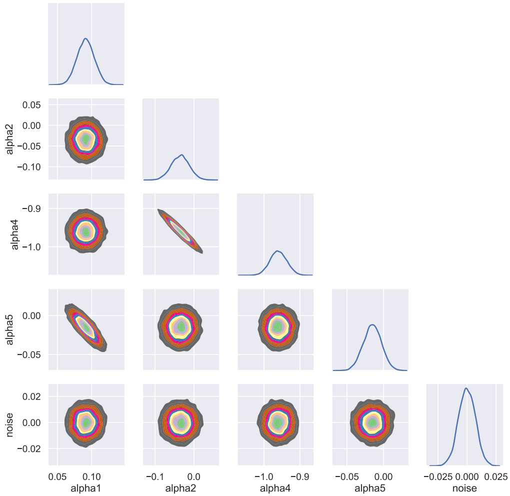
$\varepsilon_i \sim N(0, 0.08)$ where $i = 1, ..., 30$

1. Convergence was by inspection of trace plots.
2. There are several formal test of convergence.
3. The models were run for 150,000 iterations, with a burn-in of 30,000 thinning of 10.



Trace plots for model (M_3) with four parameters.

Bayesian model comparison | Results of examples



Posterior distributions of the parameters for M_3

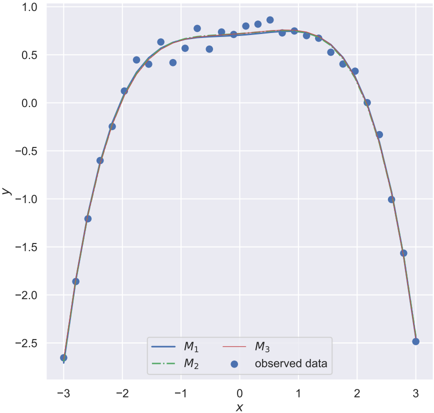
Summary statistics

Estimated model parameters

Parameter	Model		
	M_1	M_2	M_3
α_1	0.08215	0.08232	0.06691
α_2		-0.01634	-0.01678
α_4	-0.99151	-0.97607	-0.97561
α_5			0.01909
noise	0.00007	0.00009	-0.00014

DTU DRIVEN

UNIVERSITÉ DU LUXEMBOURG LISER



Fitted models, and observed data points.

.

Model comparison by Bayes factor

Model	ln-marginal likelihood	ln-Bayes factor	$\log_{10}$Bayes factor
M_1	-40737.55		
M_2	-139219.05	BF_{12} = 98481.505	BF_{12} = 42769.974
M_3	-1796271.2	BF_{13} = 1755533.656	BF_{13} = 762418.58

Bayes factor favours M_1 with 3 parameters. The Bayes factor is decisive in favour of M_1.

Current and future work

- Make model comparison using different Bayesian techniques with synthetic and real-world data for the HBV model.
- Extend the model comparison and selection to other eco-hydrological models like optimality based models.
- Make prior impact assessment for the various models.

Thank you

Acknowledgement

The Doctoral Training Unit **Data-driven computational modelling and applications** (DRIVEN) is funded by the Luxembourg National Research Fund under the PRIDE programme (PRIDE17/12252781).

https://driven.uni.lu

References

[1] Maciej Zalewski. "Ecohydrology—The use of ecological and hydrological processes for sustainable management of water resources/Ecohydrologie—La prise en compte de processus écologiques et hydrologiques pour la gestion durable des ressources en eau". In: *Hydrological Sciences Journal* 47.5 (2002), pp. 823–832.

[2] Sylvia Frühwirth-Schnatter. "Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques". In: *The Econometrics Journal* 7.1 (2004), pp. 143–167.

[3] Xiao-Li Meng and Wing Hung Wong. "Simulating ratios of normalizing constants via a simple identity: a theoretical exploration". In: *Statistica Sinica* (1996), pp. 831–860.