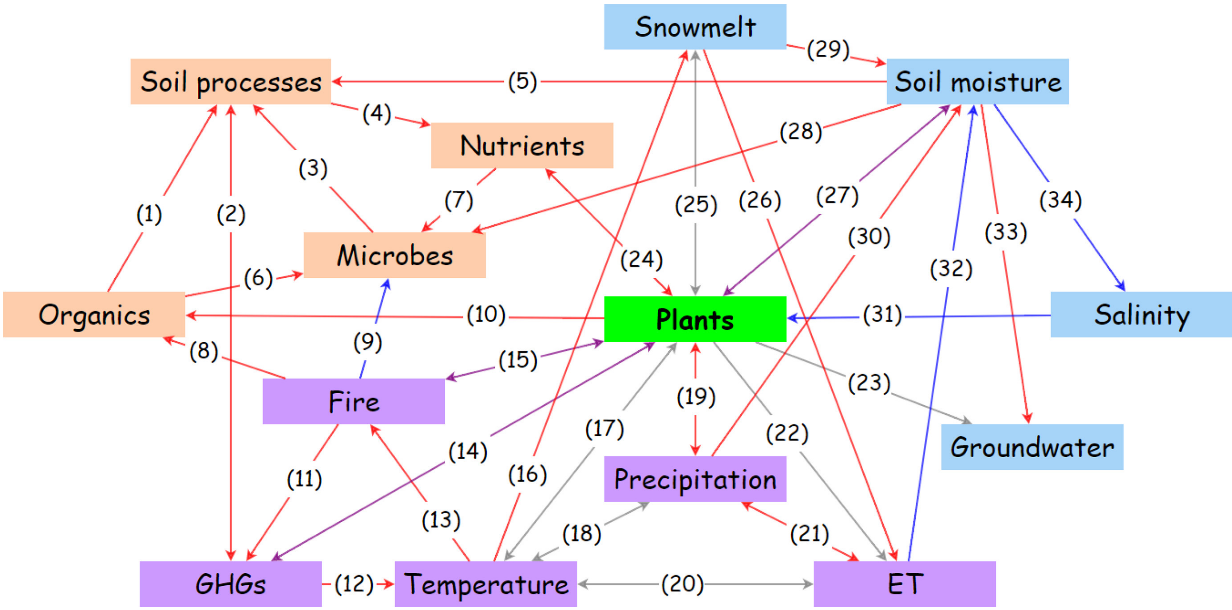
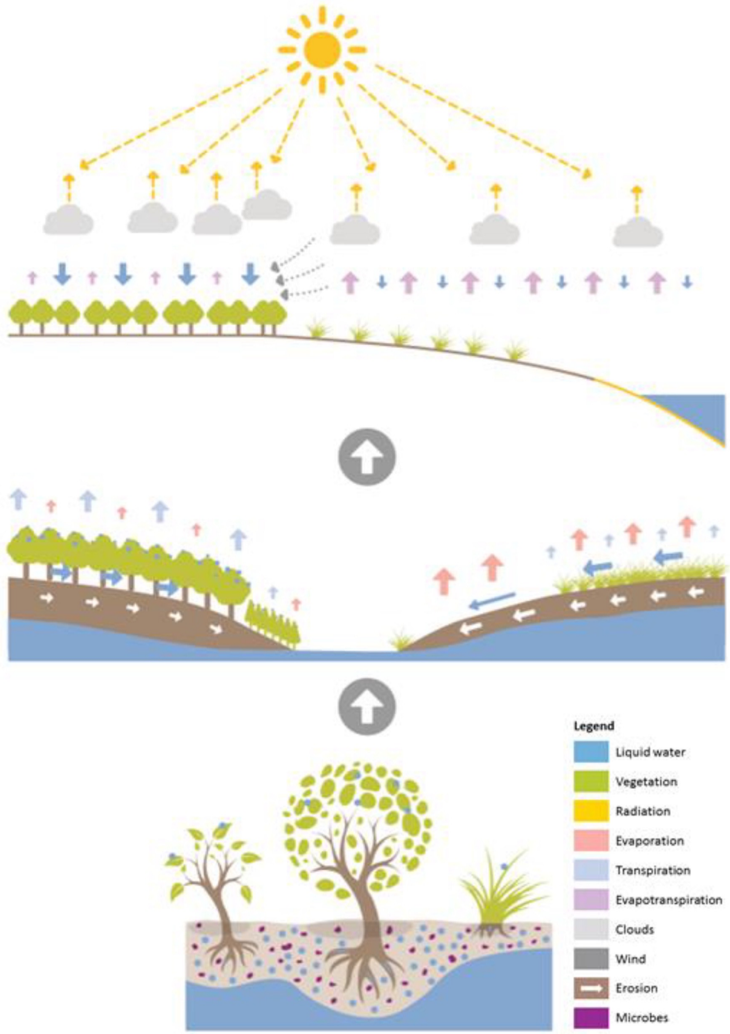
Building Trustworthy Environmental Models

LUCY MARSHALL
FACULTY OF SCIENCE AND ENGINEERING
MACQUARIE UNIVERSITY

With many thanks to:
Clare Stephens, Hae Na Yoon, Fiona Johnson, Dayang Li, Yating Tang, Xia Wu, Arpit Kapoor, Tyler Smith, Ashish Sharma

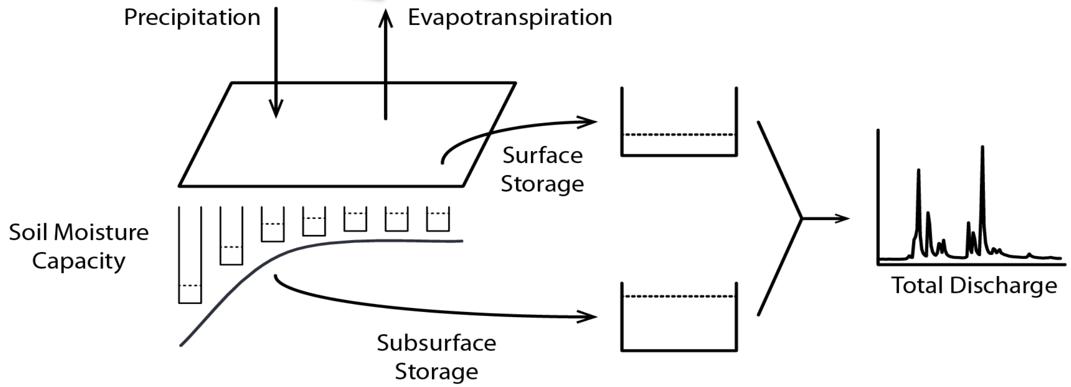
Dealing with complexity in environmental systems





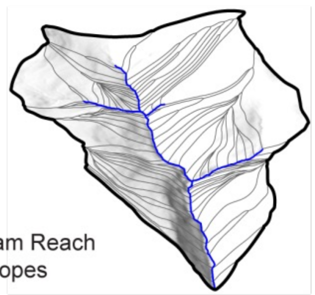
Stephens, C. M., Lall, U., Johnson, F. M., & Marshall, L. A. (2020). Landscape changes and their hydrologic effects: Interactions and feedbacks across scales. *Earth-Science Reviews*, 103466.

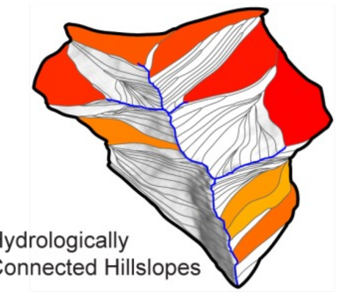
Dealing with complexity in environmental systems

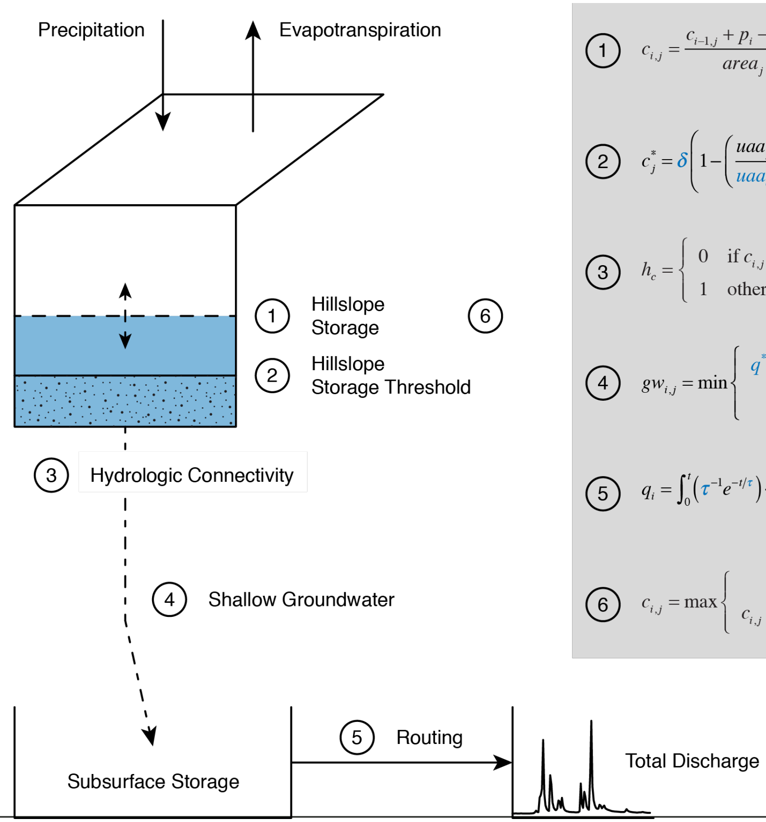


Simulating hydrologic processes

MACQUARIE University SYDNEY·AUSTRALIA







1 $$c_{i,j} = \frac{c_{i-1,j} + p_i - aet_i}{area_j}$$

2 $$c_j^* = \delta\left(1 - \left(\frac{uaa_j}{uaa_c}\right)^{\kappa}\right)$$

3 $$h_c = \begin{cases} 0 & \text{if } c_{i,j} \le c_j^* \\ 1 & \text{otherwise} \end{cases}$$

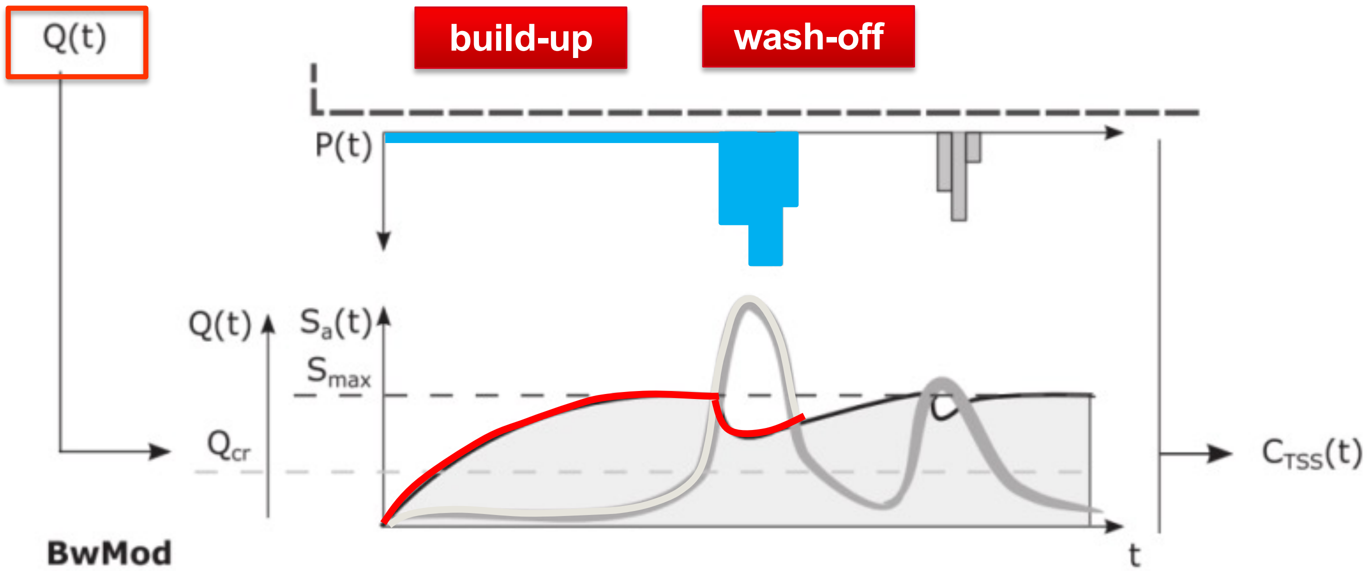
4 $$gw_{i,j} = \min\begin{cases} q^* \cdot \boldsymbol{\omega}_{f_j} \\ c_{i,j} \end{cases}$$

5 $$q_i = \int_0^t \left(\tau^{-1} e^{-t/\tau}\right) \cdot \sum_j gw_{i,j} \cdot (t - \tau) \cdot d\boldsymbol{\tau}$$

6 $$c_{i,j} = \max\begin{cases} 0 \\ c_{i,j} - gw_{i,j} \end{cases}$$

Smith, T., Marshall, L., McGlynn, B., & Jencso, K. (2013). Using field data to inform and evaluate a new model of catchment hydrologic connectivity. *Water Resources Research*, *49*(10), 6834-6846.

Integrating hydrologic and water quality variables

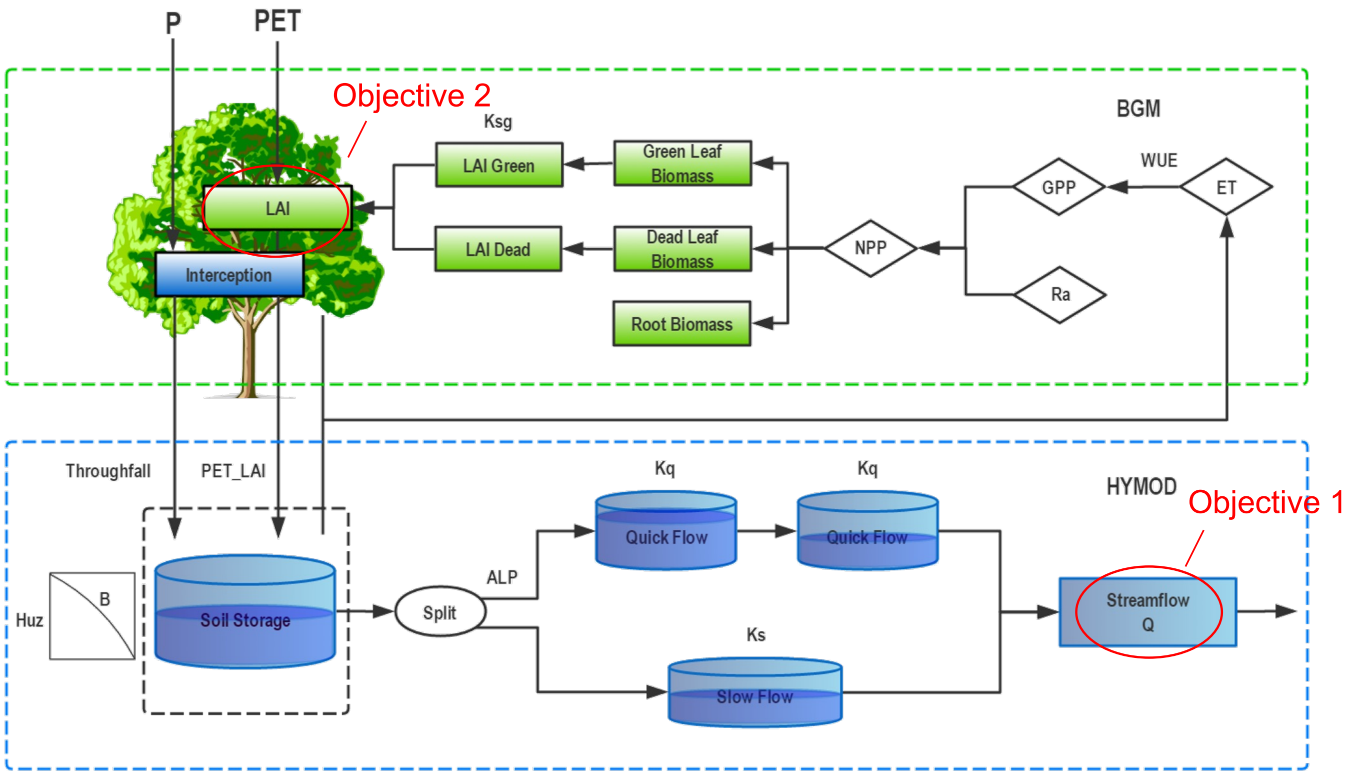


Build-up / wash-off model (BwMod)

Sikorska, A.E., et al., The value of streamflow data in improving TSS predictions - Bayesian multi-objective calibration. *Journal of Hydrology*. 530: p. 241-254, 2015.
Wu, X., L. Marshall, A. Sharma, Improving Total Suspended Solids (TSS) predictions with data transformations in the data domain and time domain.

Integrating hydrologic and vegetation dynamics

MACQUARIE University SYDNEY·AUSTRALIA



Tang, Marshall, et al. (2017). A Bayesian alternative for multi-objective ecohydrological model specification. *Journal of Hydrology* http://dx.doi.org/10.1016/j.jhydrol.2017.07.040

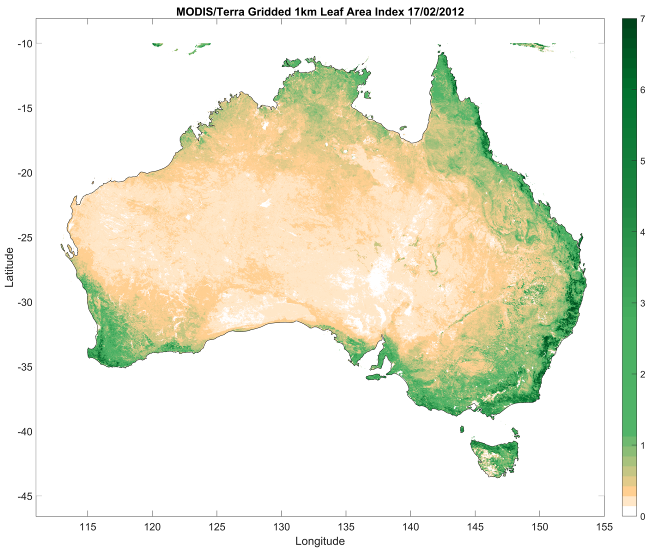
Model Inputs: P, PET
Model Outputs: Q, LAI

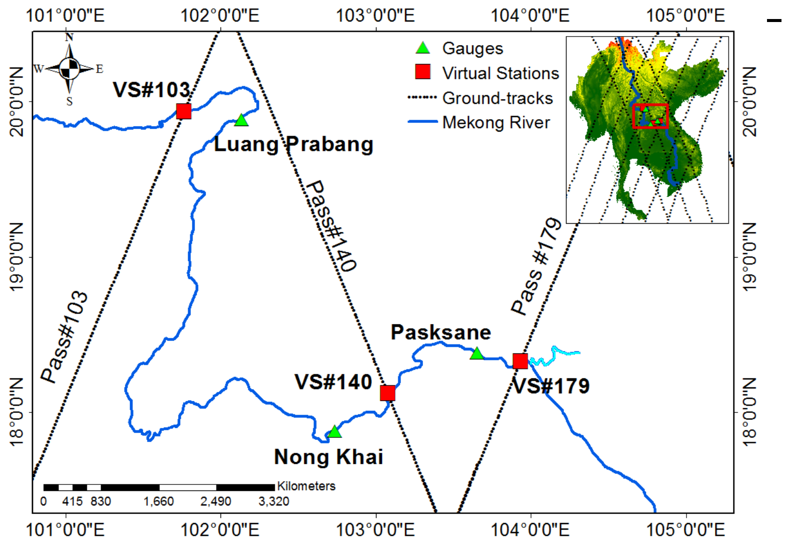
Calibrating Parameters:

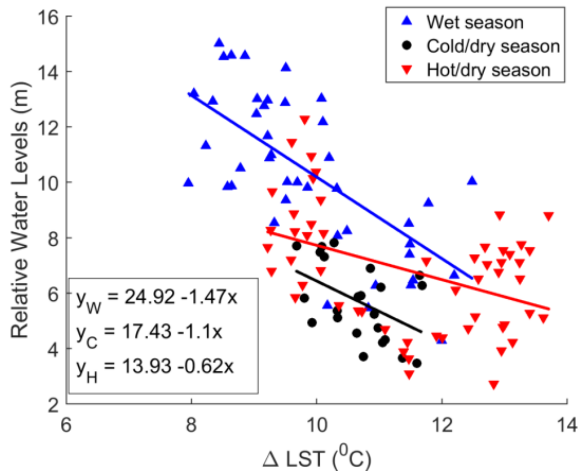
Huz: Height of soil moisture tank
B: Distribution function shape
Alp: Quick-slow split
Kq: Quickflow routing rate
Ks: Slowflow routing rate

WUE: Water use efficiency
Ksg: Natural decay factor for live/green biomass

Increasingly available information







LETTER • OPEN ACCESS

Beyond river discharge gauging: hydrologic predictions using remote sensing alone

Hae Na Yoon[1], Lucy Marshall[3,1,2] and Ashish Sharma[1]

Published 17 February 2023 • © 2023 The Author(s). Published by IOP Publishing Ltd

Environmental Research Letters, Volume 18, Number 3

Citation Hae Na Yoon *et al* 2023 *Environ. Res. Lett.* **18** 034015

DOI 10.1088/1748-9326/acb8cb

ELSEVIER

Remote Sensing of Environment

Volume 212, June 2018, Pages 31-46

Deriving daily water levels from satellite altimetry and land surface temperature for sparsely gauged catchments: A case study for the Mekong River

Hung T. Pham [a, b], Lucy Marshall [a], Fiona Johnson [a], Ashish Sharma [a]

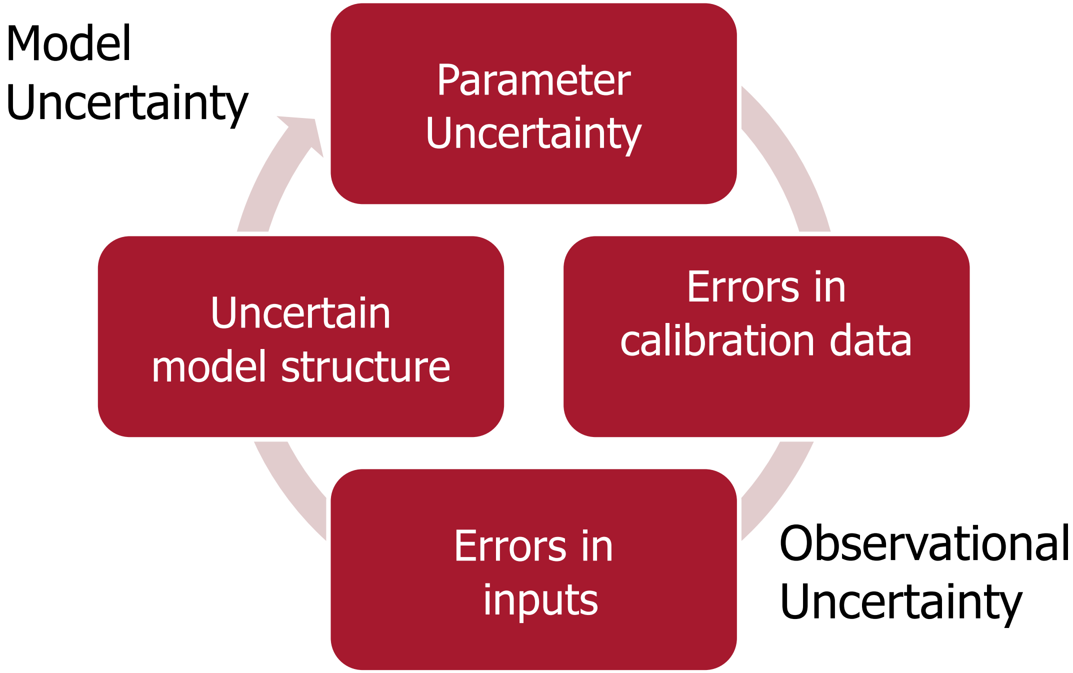
Show more

https://doi.org/10.1016/j.rse.2018.04.034

Get rights and content

Uncertainty in environmental analysis is innate

- Significant effort has been put into uncertainty frameworks that attempt to explicitly address potential model uncertainties and build trust in models
- Each is aimed at reconciling observations/information with model equations and assumptions



A Bayesian uncertainty framework

Adapted after Clark, Ecology Letters, 2005.

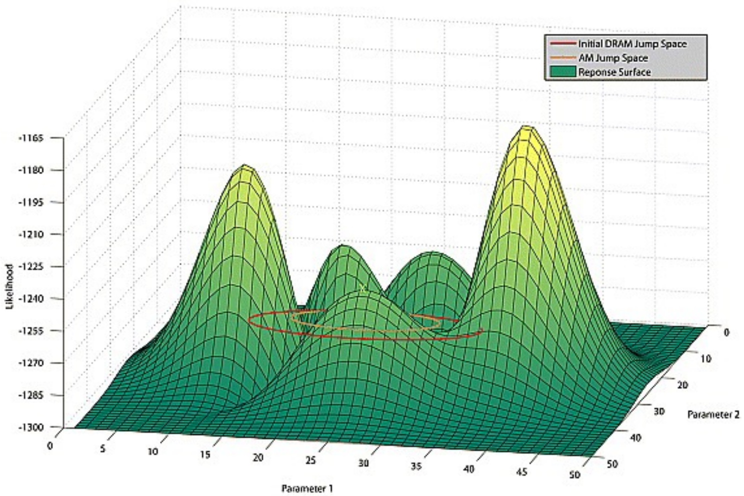
Standard Bayesian Model	Multiple Sources of Data	Hierarchical Model	Ensemble Model

What are recent opportunities and challenges for reconciling environmental models and data?

y ~ variable of interest
x ~ input data, climatological variables
θ ~ parameters

Model inference and optimisation

- ***Parameter and data uncertainty***



Smith, T. J., & Marshall, L. A. Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques. *Water Resources Research*, 44(12), 2008.

Jeremiah, Sisson, Sharma, Marshall, Efficient hydrological model parameter optimization with Sequential Monte Carlo sampling, *Environmental Modelling & Software*, Volume 38, 2012.

Step 1:

Assume Gaussian, homoscedastic, independent errors

Step 2:

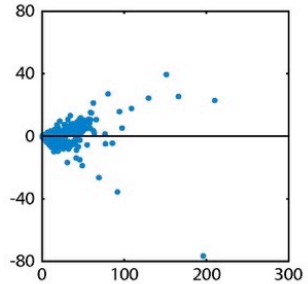
Check for zero-inflation
(>10% zeros in observed discharge record)

≈ 75% zero observations

Add zero-inflation component to likelihood function

Step 3:

Check for heteroscedastic residuals

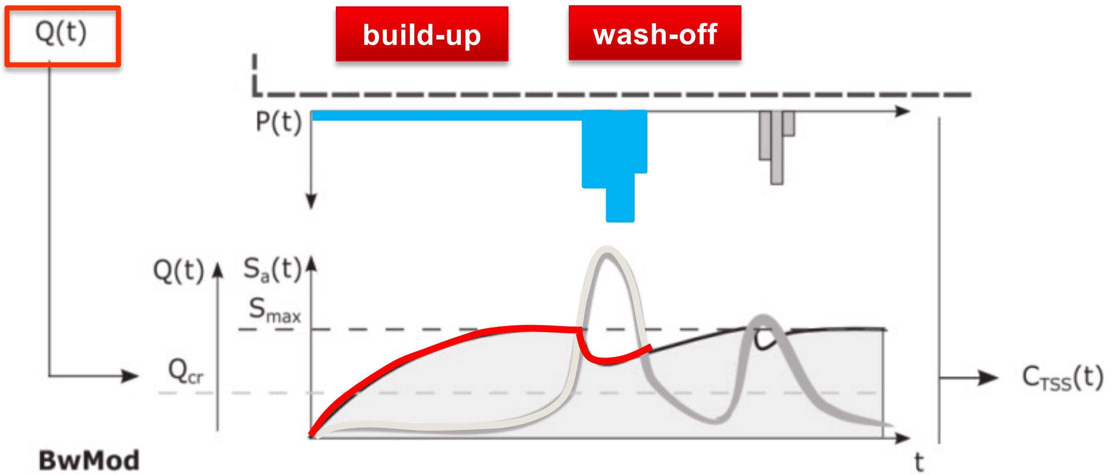


Add Box-Cox transform component to likelihood function

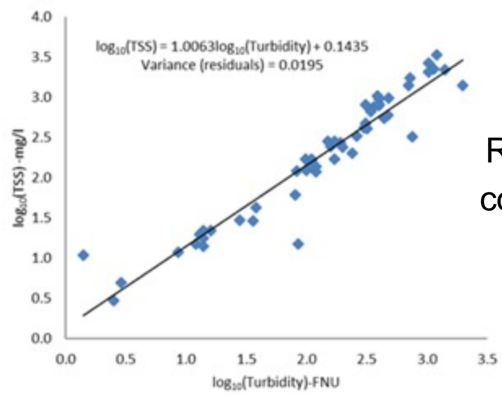
T. Smith, L Marshall, A Sharma, Modeling residual hydrologic errors with Bayesian inference, *Journal of Hydrology*, Volume 528, 2015

Wu, X., L. Marshall, A. Sharma, The influence of data transformations in simulating Total Suspended Solids using Bayesian inference, *Environmental Modelling and Software*, 2019.

Using proxy data with measurement error



Build-up / wash-off model (BwMod)



Regression model of TSS concentration and turbidity

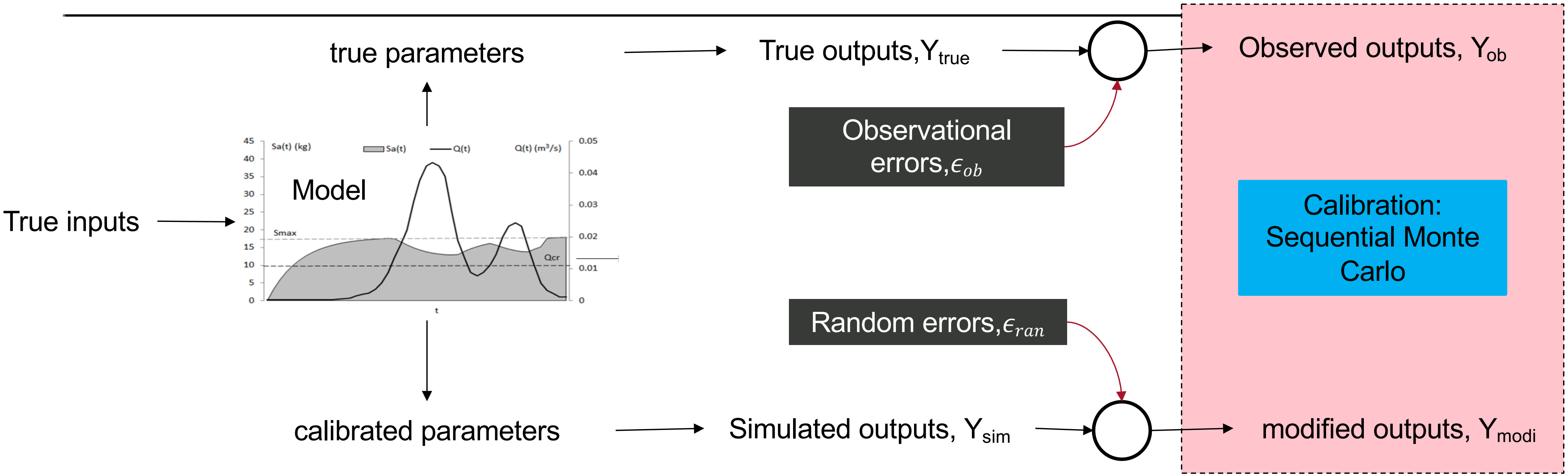
observational errors

- proxy measurement errors
- measurement errors
- representational errors

Wu, X., Marshall, L. and Sharma, A., 2022. Incorporating multiple observational uncertainties in water quality model calibration. *Hydrological Processes*, 36(1).

Wu, X., Marshall, L., & Sharma, A. (2022). Quantifying input uncertainty in the calibration of water quality models: reordering errors via the secant method. *Hydrology and Earth System Sciences*, *26*(5), 1203-1221.

Using proxy data with measurement error



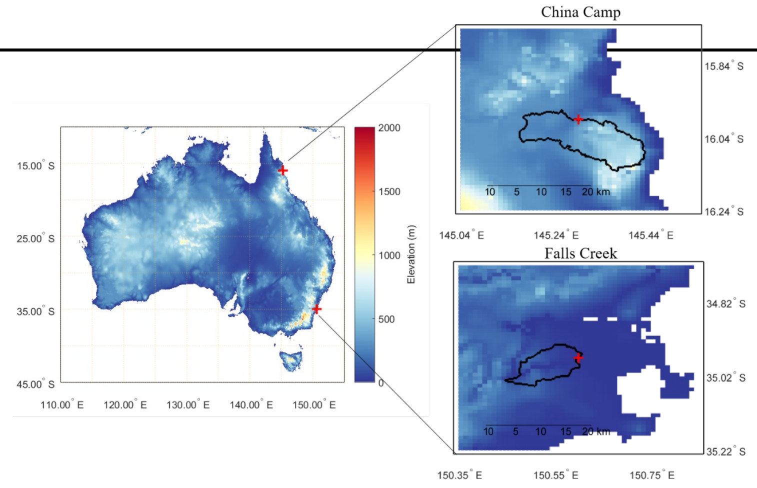
Wu, X., Marshall, L., & Sharma, A. (2022). Quantifying input uncertainty in the calibration of water quality models: reordering errors via the secant method. *Hydrology and Earth System Sciences*, *26*(5), 1203-1221.

Using multivariate data with measurement error

Define a new error model using MODIS satellite products:

(1) FparLai_QC expressing pixel quality information

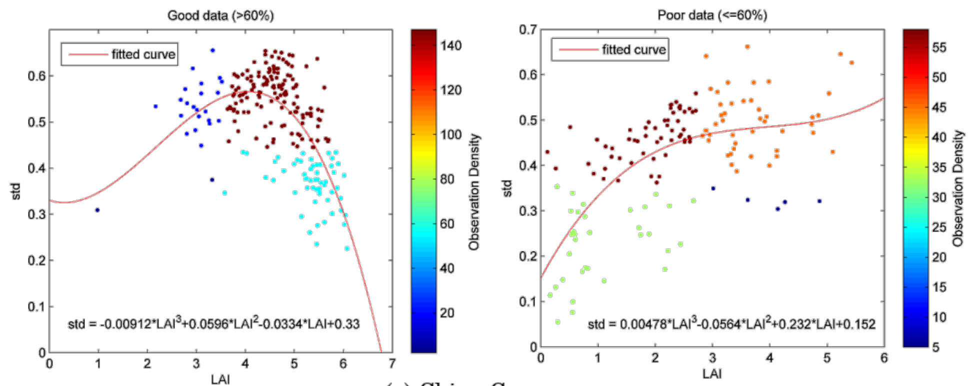
(2) LaiStdDev_1km representing retrieval uncertainty (STD LAI)



$$\widetilde{LAI_t} = LAI_t + \gamma_t = \widehat{LAI_t} + \varepsilon_t + \gamma_t = \widehat{LAI_t} + \lambda_t$$

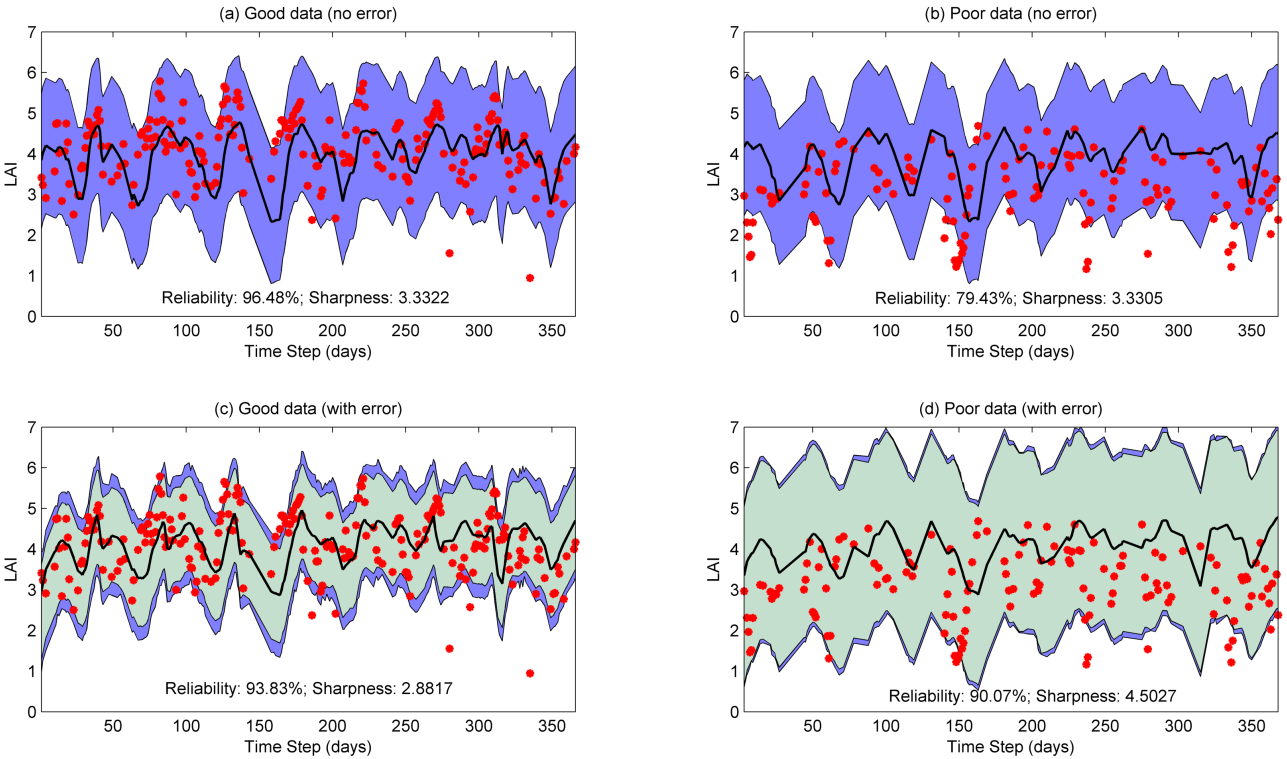
$$\lambda_t \sim N\left(0, \sigma^2_{LAI_t} + \sigma^2_M\right)$$

Tang, Marshall, et al. (2019). Ecohydrologic error models for improved Bayesian inference in remotely sensed catchments. *Water Resources Research.*



(a) China Camp

Using multivariate data with measurement error



- Intervals when simulating ‘good’ data become narrower
- Uncertainty estimates for ‘poor’ data are improved
- Issues with model structural errors remain

Figure: 90% Confidence limit of the total (dark blue) and residual (light blue) errors between measured and predicted LAI values for good/poor data for **China Camp**. In each plot, the red dots are the LAI observations and the black line is the predictions. The computed statistics (reliability and sharpness) is shown at the bottom.

Tang, Marshall, et al. (2019). Ecohydrologic error models for improved Bayesian inference in remotely sensed catchments. *Water Resources Research.*

Bayesian inference and environmental models

MACQUARIE University SYDNEY · AUSTRALIA

- Bayesian inference is an attractive framework for quantifying uncertainty in process-based models, allowing for complex descriptions of model error;
- Favoured in the environmental modelling community because:
  - ✓ Allows for expert knowledge
  - ✓ Expands the use of process based models
  - ✓ Based on the modeler’s domain knowledge

Rise of Machine Learning

- Unprecedented success of ML in certain disciplines has increased its momentum in fields like water resources

Shen, C. (2018), Deep learning: A next-generation big-data approach for hydrology, *Eos*, 99, https://doi.org/10.1029/2018EO095649. Published on 25 April 2018.

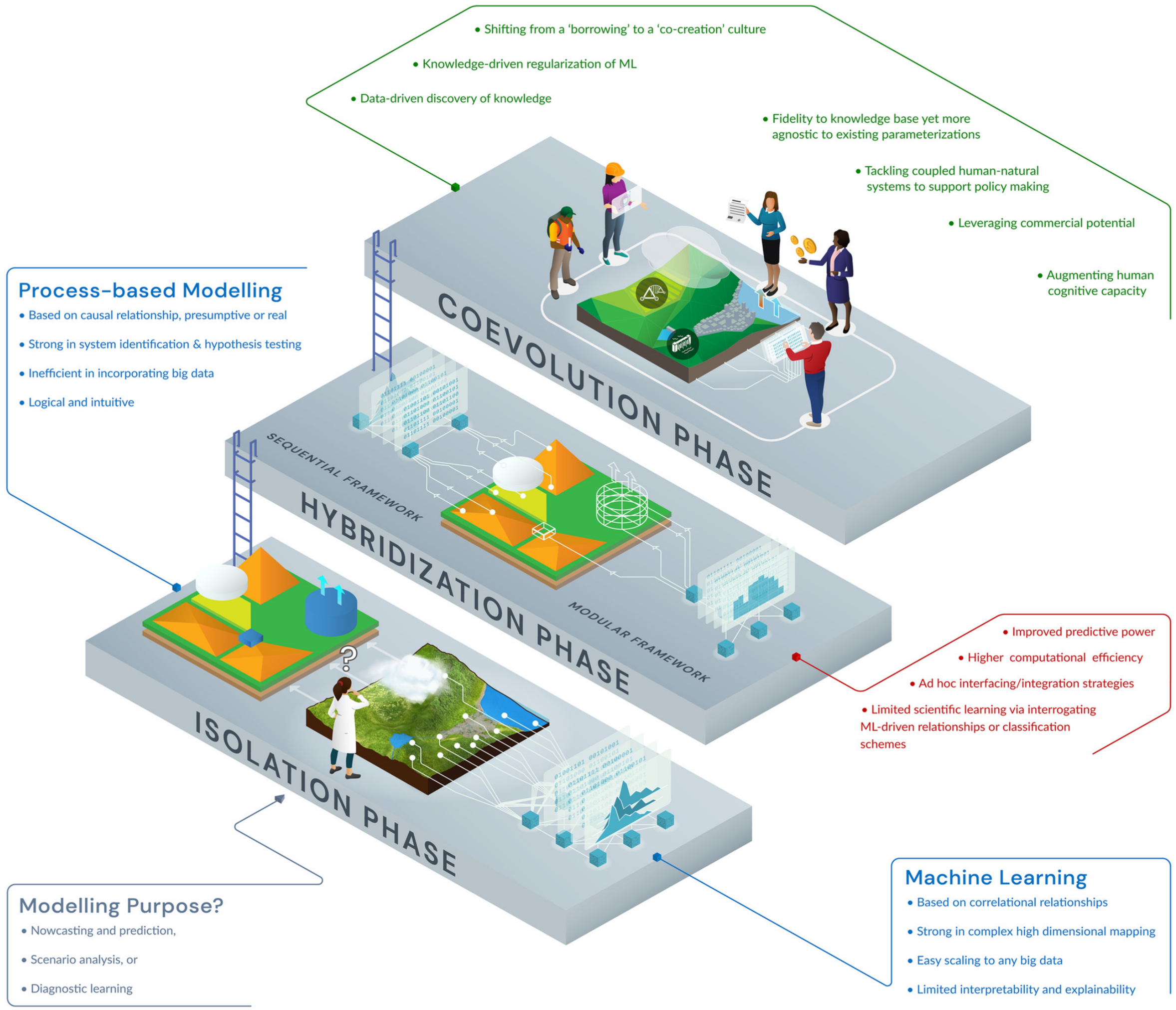
Two Grand Challenges in ML for Environmental Analysis:

- *Lack of explainability*
- *Divorce from the system knowledge-base*

Building trustworthy environmental models

Approaches to hybridization:

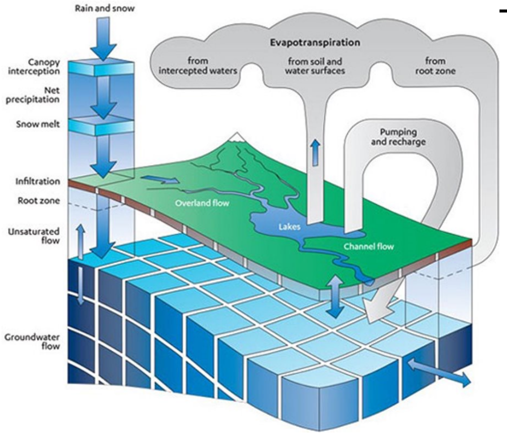
- Coupled models
- ML-informed process-based models
- Physics-informed ML models

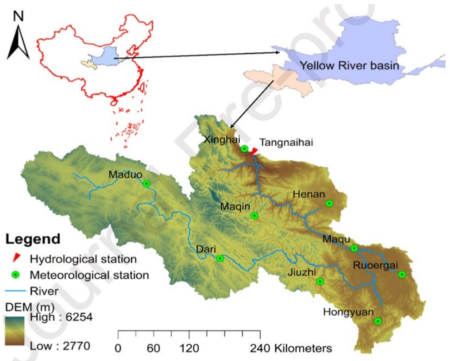


RIE

Razavi, S., Hannah, D.M., Elshorbagy, A., Kumar, S., Marshall, L., Solomatine, D.P., Dezfuli, A., Sadegh, M. and Famiglietti, J., (2022). Coevolution of Machine Learning and Process-based Modelling to Revolutionize Earth and Environmental Sciences: A Perspective. Hydrological Processes, p.e14596.

Hybridization of ML and Process Models



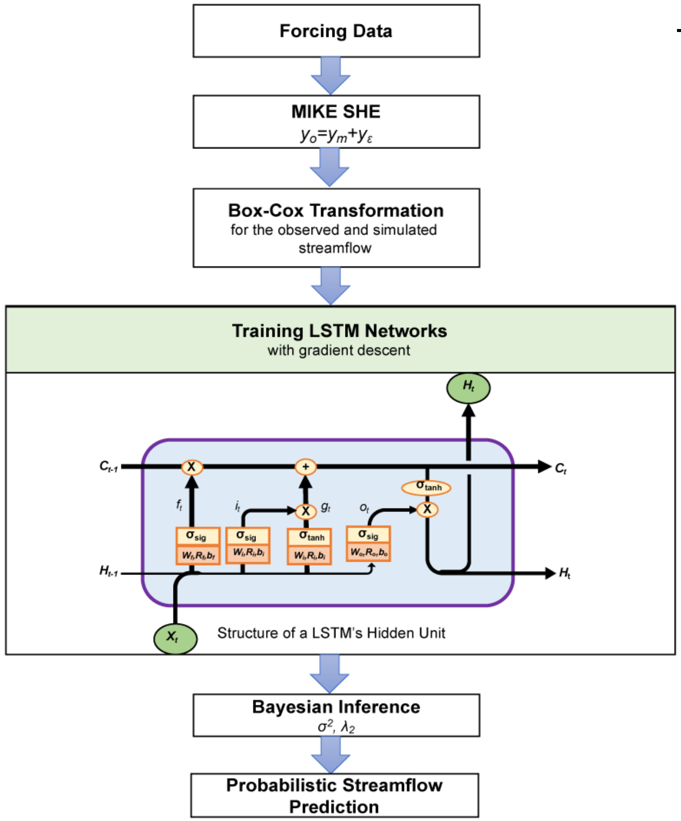


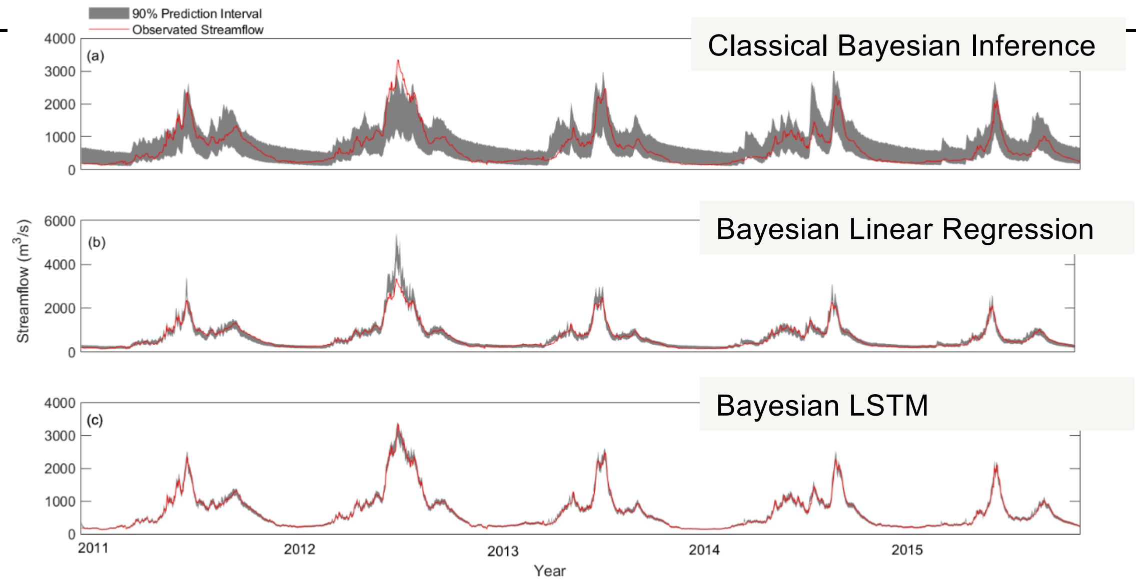
Coupled Models

- An underlying process model is used to simulate water fluxes and operational scenarios.
- A ML model (a Long Short-Term Memory (LSTM) network) is applied to the model errors to improve probabilistic predictions.
- Overall goal is to couple a deep learning approach with a hydrological model to characterize predictive uncertainty and system processes together.

Li, D., Marshall, L., Liang, Z., Sharma, A., & Zhou, Y. (2021). Characterizing distributed hydrological model residual errors using a probabilistic long short-term memory network. *Journal of Hydrology*, *603*, 126888.

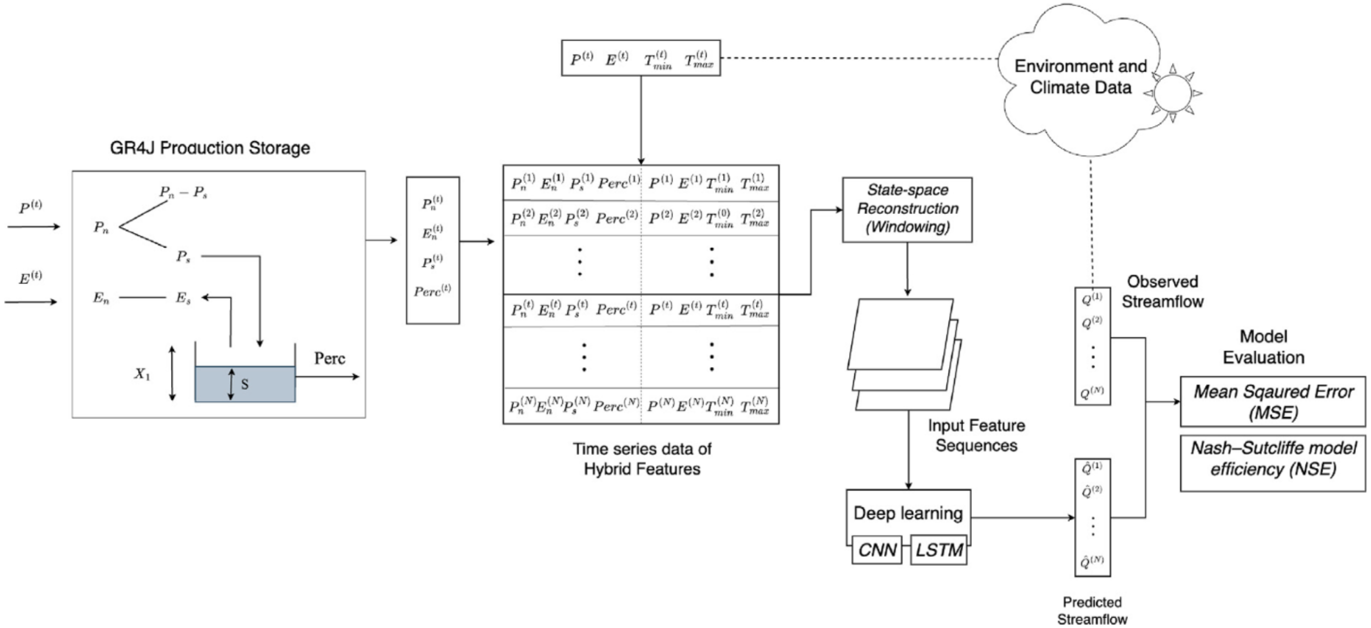
Hybridization of ML and Process Models





Li, D., Marshall, L., Liang, Z., Sharma, A., & Zhou, Y. (2021). Characterizing distributed hydrological model residual errors using a probabilistic long short-term memory network. *Journal of Hydrology*, *603*, 126888.

Hybridization of ML and Process Models

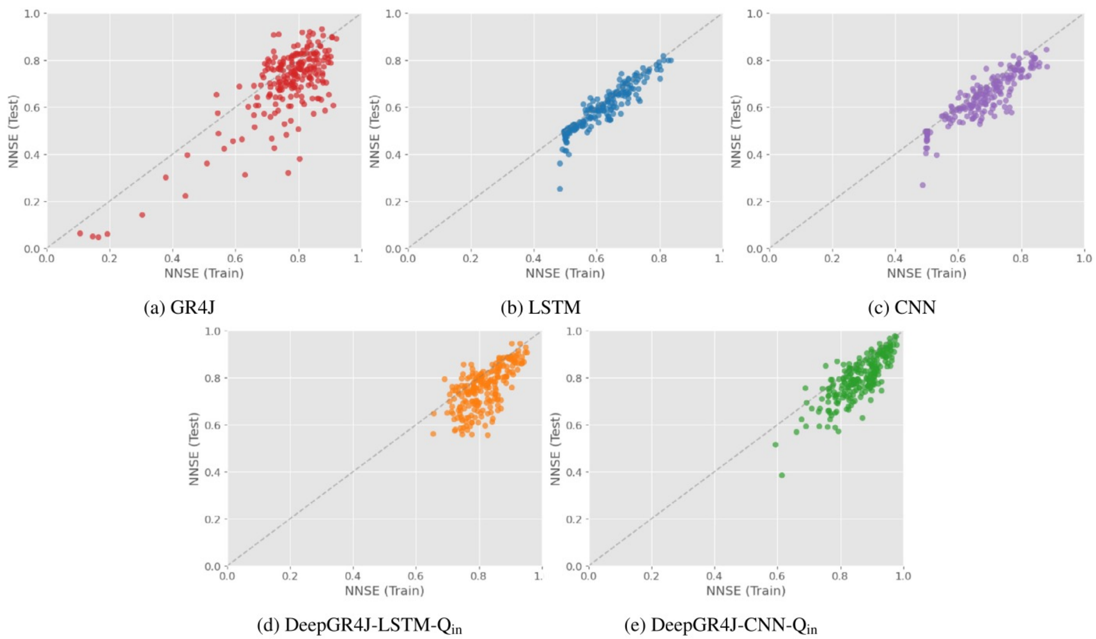


Hybridized Conceptual Models

- A component of a process model is replaced by a ML model to improve predictions
- Overall goal is to provide more flexibility to components of the model that aren't well defined by existing knowledge

Kapoor, A., Pathiraja, S., Marshall, L. and Chandra, R., 2023. DeepGR4J: A deep learning hybridization approach for conceptual rainfall-runoff modelling. *Environmental Modelling & Software*, *169*, p.105831.

Hybridisation of ML and Process Models



(a) GR4J (b) LSTM (c) CNN

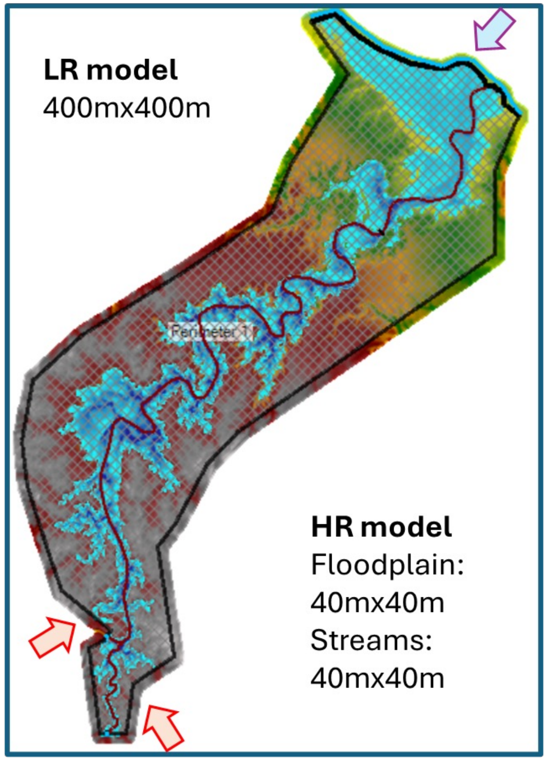
(d) DeepGR4J-LSTM-Q_{in} (e) DeepGR4J-CNN-Q_{in}

Hybridized Conceptual Models

- A component of a process model is replaced by a ML model to improve predictions
- Overall goal is to provide more flexibility to components of the model that aren't well defined by existing knowledge

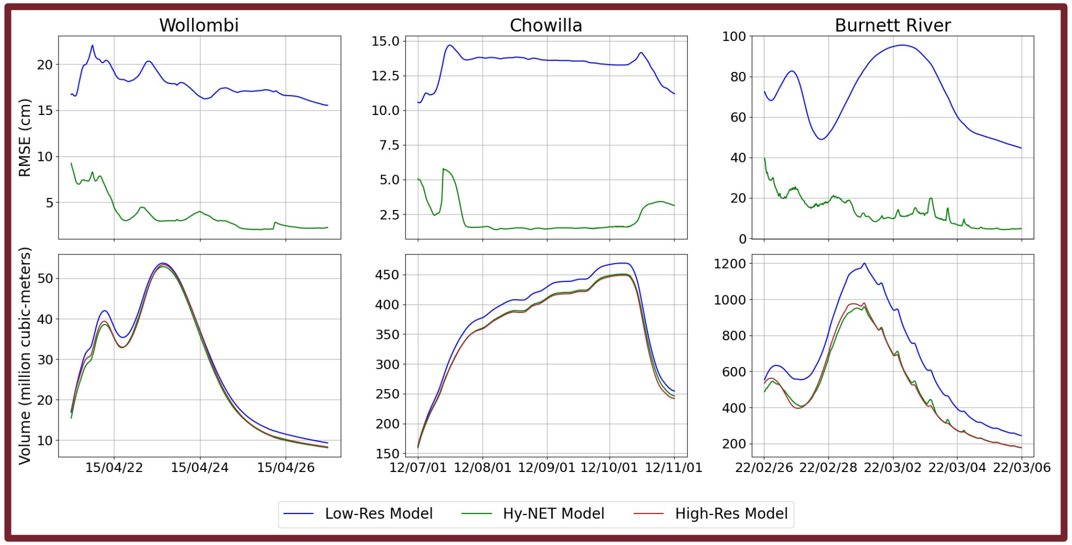
Kapoor, A., Pathiraja, S., Marshall, L. and Chandra, R., 2023. DeepGR4J: A deep learning hybridization approach for conceptual rainfall-runoff modelling. *Environmental Modelling & Software*, *169*, p.105831.

Hybridization of ML and Process Models



Physics Informed ML

- A low resolution hydrodynamic model is used as an input to a high-resolution Deep Learning model.
- Overall goal is to provide physically-derived information to guide and constrain predictions from the ML approach



What is the path forward?

- It is clear that AI/Machine Learning provides a powerful tool for environmental analysis
- How can we build trustworthy models that capitalise on the ideals of both mechanistic and machine learning methods?
- Requires strong collaboration and co-design between data scientists and domain researchers

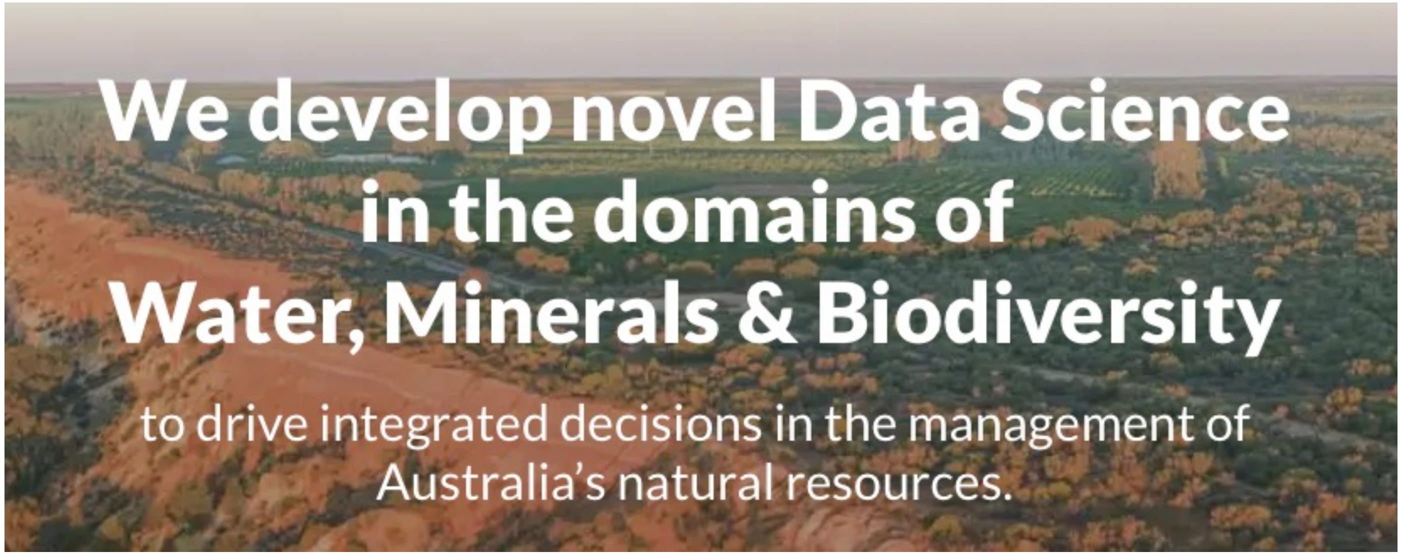
DARE

ARC Training Centre in
Data Analytics for Resources & Environments

Data Analytics in Resources and Environments (DARE)

A multidisciplinary ARC Training Centre that trains a new generation of world-leading data scientists with applied domain knowledge

Students are embedded with industry and government partners to improve informed decision making around Australian resources



Building Trustworthy Environmental Models

LUCY MARSHALL
FACULTY OF SCIENCE AND ENGINEERING
MACQUARIE UNIVERSITY

With many thanks to:
Clare Stephens, Hae Na Yoon, Fiona Johnson, Dayang Li, Yating Tang, Xia Wu, Arpit Kapoor, Tyler Smith, Ashish Sharma