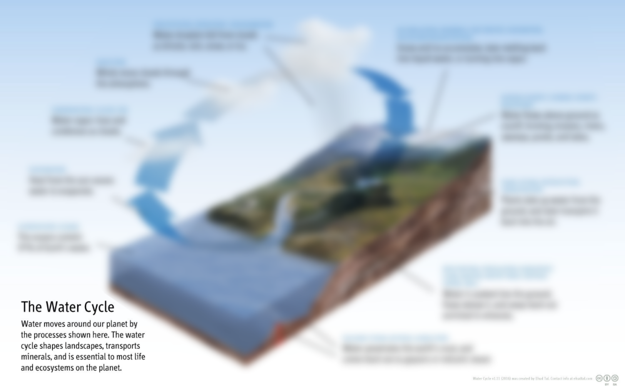
Probabilistic Models for the Uncertain Hydrologist

Benjamin Renard

Probabilistic Models for the Uncertain Hydrologist



Benjamin Renard
INRAE, UR RECOVER, Aix-Marseille University

Habilitation à Diriger des Recherches
Ecole Doctorale Sciences de l'Environnement / ED 251

Introduction


crédit: Diego Delso

Introduction

crédit: Diego Delso

crédit: Aleda12

Introduction

crédit: Diego Delso

crédit: Irstea

crédit: Aleda12

Introduction

crédit: Diego Delso

crédit: Irstea

crédit: Aleda12

crédit: Clicgauche

Introduction

crédit: Diego Delso

crédit: Irstea

crédit: João Cautela

crédit: Aleda12

crédit: Clicgauche

Introduction

crédit: Diego Delso

crédit: Irstea

crédit: João Cautela

crédit: Aleda12

crédit: Clicgauche

crédit: RuB

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?

photo: M. Lagouy

Introduction

Three Practical Questions

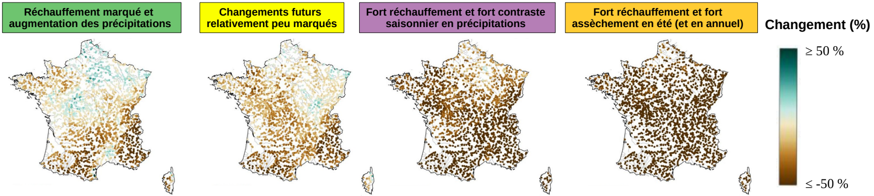
1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?

crédit: héliotypie Ernest Louis Désiré Le Deley

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?



crédit: EXPLORE 2

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Diversity of situations

- From flood to drought

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Diversity of situations

- From flood to drought
- Present, past, future

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Diversity of situations

- From flood to drought
- Present, past, future
- From one site to thousands of catchments

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Common Threads

- Water flowing in rivers

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Common Threads

- Water flowing in rivers
- Estimating unknown quantities affected by uncertainty

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Common Threads

- Water flowing in rivers
- Estimating unknown quantities affected by uncertainty
- Uncertainty quantification relies on probabilistic models

Introduction

Three Practical Questions

1. What is the discharge flowing in this river right now?
2. Was the great flood of 1910 in Paris a 100-year event?
3. How much water will flow in French rivers during the summers of the next 50 years, and how warm will it be?

Common Threads

- Water flowing in rivers
- Estimating unknown quantities affected by uncertainty
- Uncertainty quantification relies on probabilistic models

Methodological Motto

$\longrightarrow$ Developing **probabilistic models** for the **uncertain hydrologist**

Outline

1. Uncertainty in Streamflow Data

Outline

1. Uncertainty in Streamflow Data

2. Uncertainty in and around Hydrologic Models

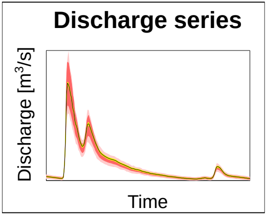
Outline

1. Uncertainty in Streamflow Data
2. Uncertainty in and around Hydrologic Models
3. Hydrologic Variability

Uncertainty in streamflow data

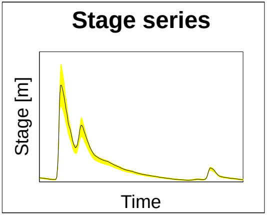
Production of River Discharge Series

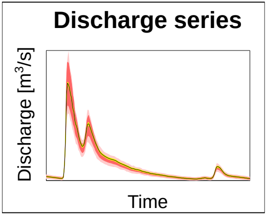
- River discharge: a key variable for hydrology... that can't be measured continuously



Production of River Discharge Series

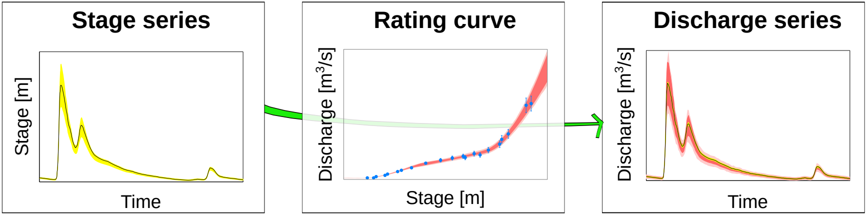
- River discharge: a key variable for hydrology... that can't be measured continuously
- But river stage can!





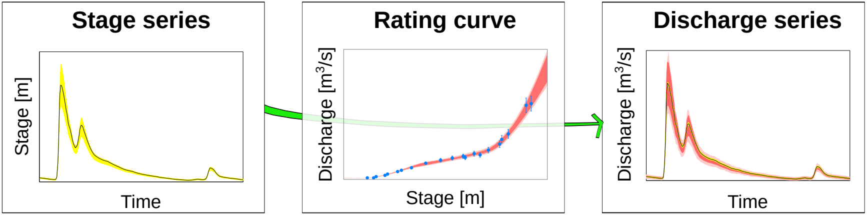
Production of River Discharge Series

- River discharge: a key variable for hydrology... that can't be measured continuously
- But river stage can!
- Rating curve: $Q = f(h)$



Production of River Discharge Series

- River discharge: a key variable for hydrology... that can't be measured continuously
- But river stage can!
- Rating curve: $Q = f(h)$

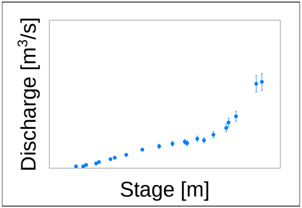


⟶ How to **formulate and estimate rating curves?**

⟶ How to **quantify and propagate uncertainties?**

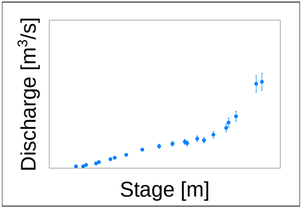
Rating Curve Formulation

Just fit some flexible curve!



Rating Curve Formulation

~~Just fit some flexible curve!~~

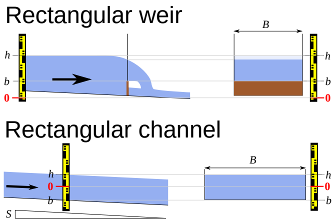


Rating Curve Formulation

~~Just fit some flexible curve!~~

1. Use hydraulics: $Q = a(h - b)^c$

   with a, b, c related to physical quantities that can be measured (albeit uncertainly)



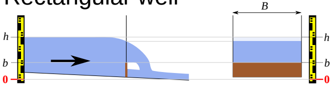
Rating Curve Formulation

~~Just fit some flexible curve!~~

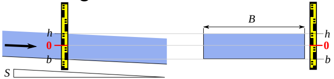
1. Use hydraulics: $Q = a(h-b)^c$

   with a, b, c related to physical quantities that can be measured (albeit uncertainly)

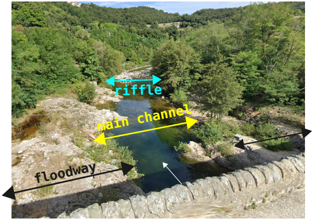
Rectangular weir



Rectangular channel



2. A ‘control matrix‘ to combine multiple controls



	riffle	channel	floodway	
				κ_1
low stage	1	0	0	$a_1(h-b_1)^{c_1}$
				κ_2
medium stage	0	1	0	$a_2(h-b_2)^{c_2}$
				κ_3
high stage	0	1	1	$a_2(h-b_2)^{c_2} + a_3(h-b_3)^{c_3}$

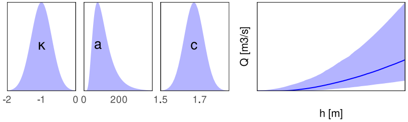
Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$

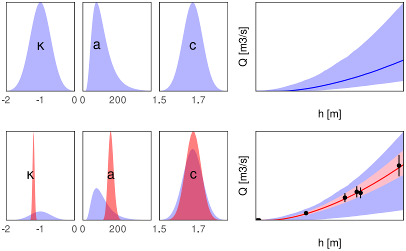


Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$

Posterior distribution $p(\boldsymbol{\theta}, \gamma | \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{Q}})$

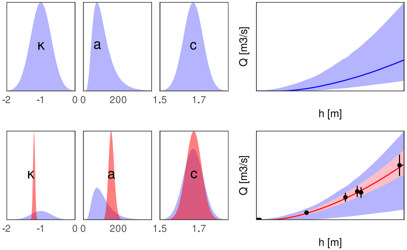


Rating Curve Estimation

$\longrightarrow$ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$

Posterior distribution $p(\boldsymbol{\theta}, \gamma | \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{Q}})$

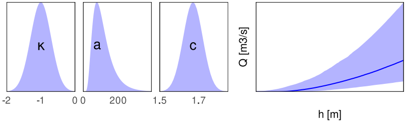


Probabilistic model linking gaugings $(\tilde{h}_i, \tilde{Q}_i)$ and RC $\hat{Q}_i = f_{RC}(h_i | \boldsymbol{\theta})$:

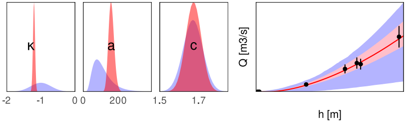
Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$



Posterior distribution $p(\boldsymbol{\theta}, \gamma | \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{Q}})$



Probabilistic model linking gaugings $(\tilde{h}_i, \tilde{Q}_i)$ and RC $\hat{Q}_i = f_{RC}(h_i|\boldsymbol{\theta})$:

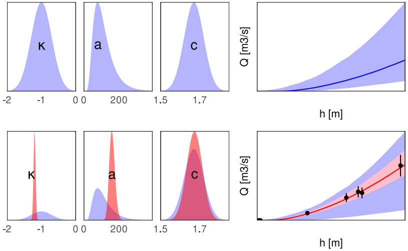
$$\tilde{Q}_i = f_{RC}(\tilde{h}_i|\boldsymbol{\theta}) + \delta_i + \varepsilon_i$$

Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$

Posterior distribution $p(\boldsymbol{\theta}, \gamma | \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{Q}})$



Probabilistic model linking gaugings $(\tilde{h}_i, \tilde{Q}_i)$ and RC $\hat{Q}_i = f_{RC}(h_i|\boldsymbol{\theta})$:

$$\tilde{Q}_i = f_{RC}(\tilde{h}_i|\boldsymbol{\theta}) + \delta_i + \varepsilon_i$$

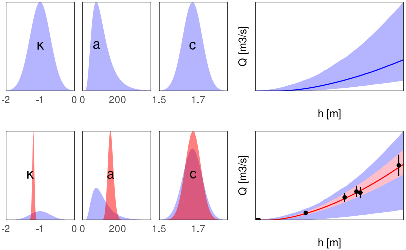
Measurement error $\delta_i \sim \mathcal{N}(0, u_i)$; u_i assumed known

Rating Curve Estimation

⟶ **Bayesian Rating Curve** estimation (BaRatin)

Prior distribution $p(\kappa, a, c) = p(\boldsymbol{\theta})$

Posterior distribution $p(\boldsymbol{\theta}, \gamma | \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{Q}})$



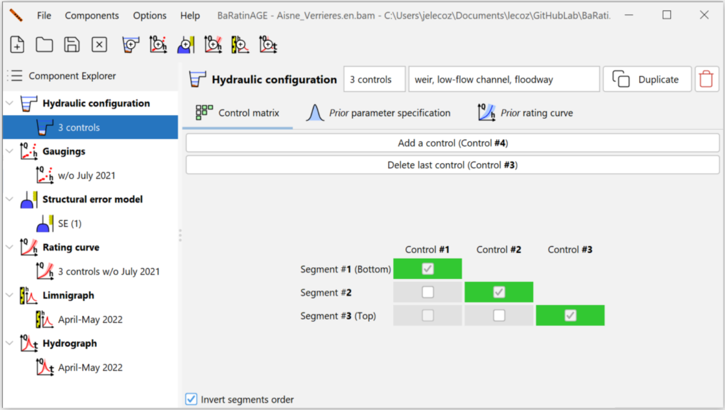
Probabilistic model linking gaugings $(\tilde{h}_i, \tilde{Q}_i)$ and RC $\hat{Q}_i = f_{RC}(h_i|\boldsymbol{\theta})$:

$$\tilde{Q}_i = f_{RC}(\tilde{h}_i|\boldsymbol{\theta}) + \delta_i + \varepsilon_i$$

Measurement error $\delta_i \sim \mathcal{N}(0, u_i)$; u_i assumed known

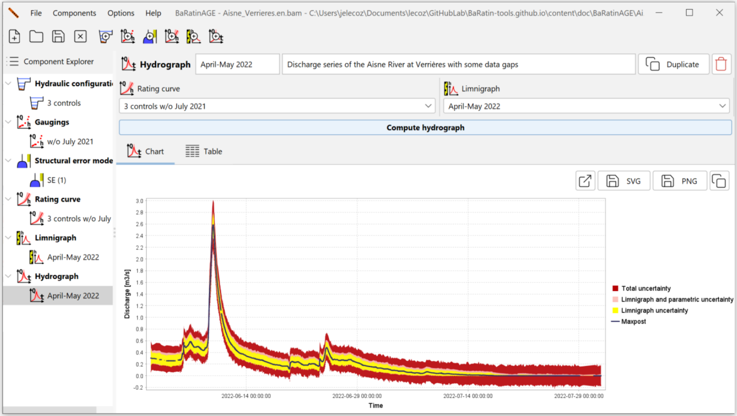
Structural error $\varepsilon_i \sim \mathcal{N}(0, \sigma_i)$; $\sigma_i = \gamma_1 + \gamma_2 \hat{Q}_i$ to be estimated

Operational Tool: BaRatinAGE

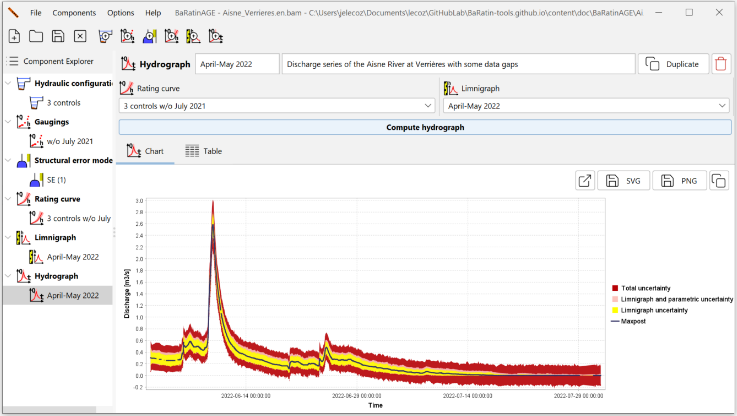


Operational Tool: BaRatinAGE

Operational Tool: BaRatinAGE

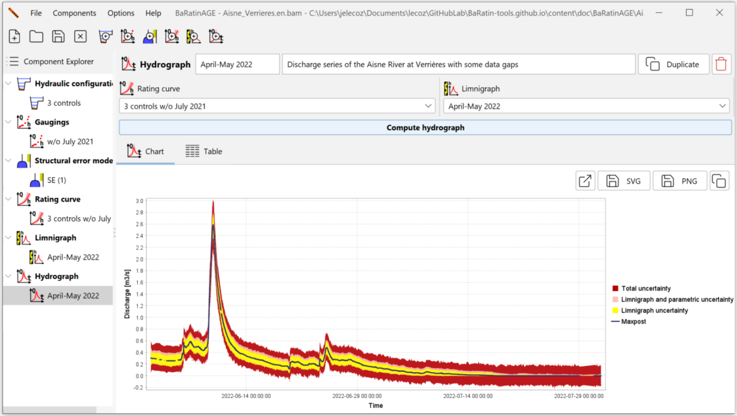


Operational Tool: BaRatinAGE



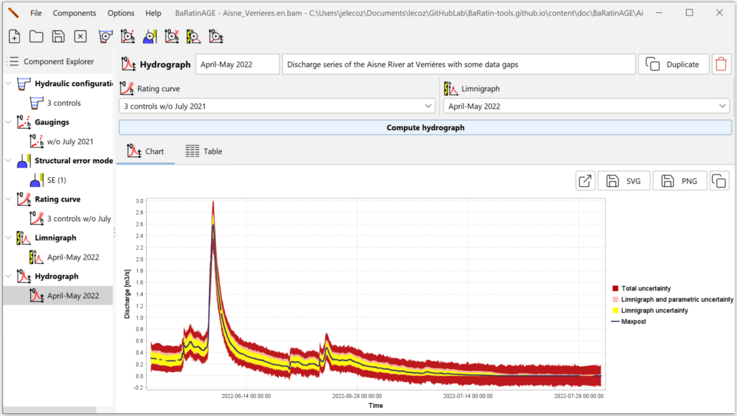
- A viable solution to estimate uncertain $Q = f(h)$ rating curves...

Operational Tool: BaRatinAGE



- A viable solution to estimate uncertain $Q = f(h)$ rating curves...
- ... which are the exception rather than the rule!

Operational Tool: BaRatinAGE



- A viable solution to estimate uncertain $Q = f(h)$ rating curves...
- ... which are the exception rather than the rule!
  1. rating shifts
  2. complex rating curves $Q = f(h, \text{others})$

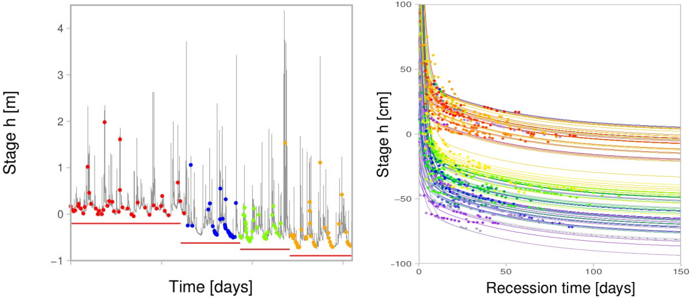
Rating Shift Happens!

⟶ Detecting rating shifts (M. Darienzo, F. Mendez)

Rating Shift Happens!

⟶ Detecting rating shifts (M. Darienzo, F. Mendez)

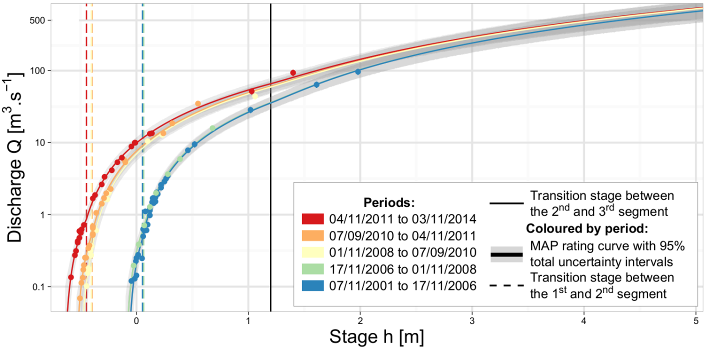
From the stage time series



Rating Shift Happens!

⟶ Detecting rating shifts (M. Darienzo, F. Mendez)

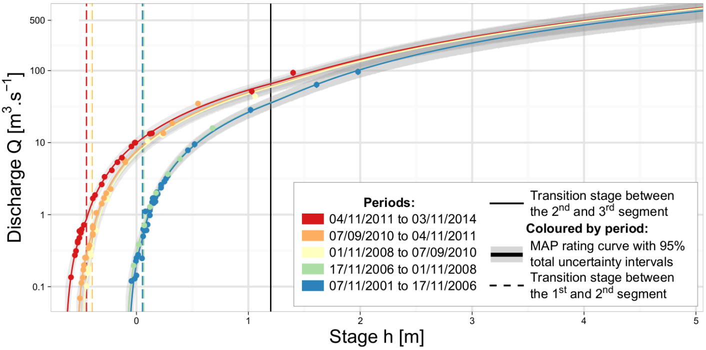
⟶ Estimating shifting curves (V. Mansanarez)



Rating Shift Happens!

⟶ Detecting rating shifts (M. Darienzo, F. Mendez)

⟶ Estimating shifting curves (V. Mansanarez)

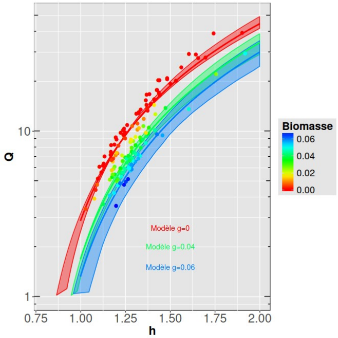


Hydraulic assumptions on what changes $\Longrightarrow$ stable vs. varying RC parameters

Complex Rating Curves $Q = f(h, \ldots)$

⟶ Vegetation model (E. Perret)

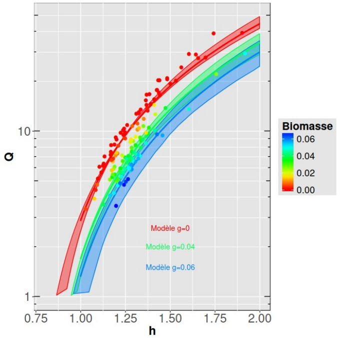
$Q = f(h, V)$ due to seasonal aquatic vegetation growth

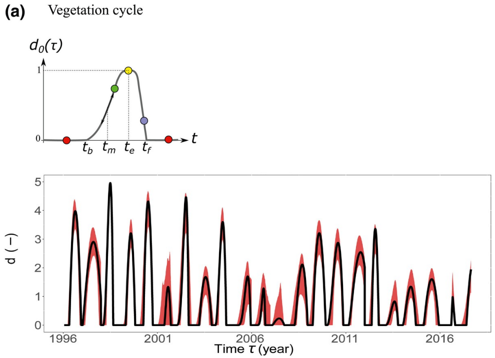


Complex Rating Curves $Q = f(h, \ldots)$

⟶ Vegetation model (E. Perret)

$Q = f(h, V)$ due to seasonal aquatic vegetation growth
As V is not measured, $Q = f(h, \cancel{V} t)$





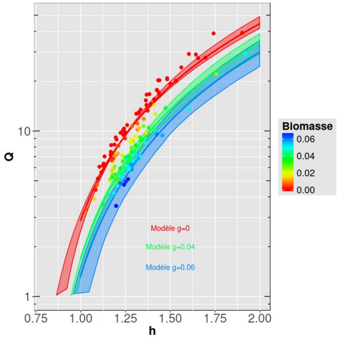
Complex Rating Curves $Q = f(h, \ldots)$

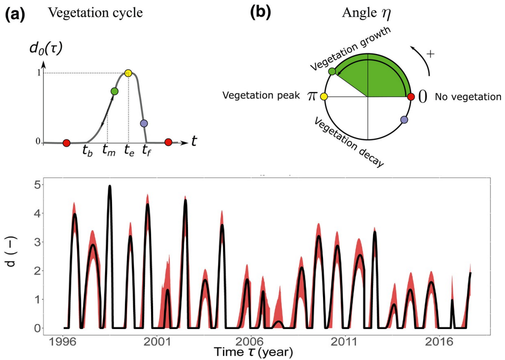
$\longrightarrow$ Vegetation model (E. Perret)

$Q = f(h, V)$ due to seasonal aquatic vegetation growth

As V is not measured, $Q = f(h, \cancel{V} t)$

Possible to incorporate qualitative information on the vegetation state.



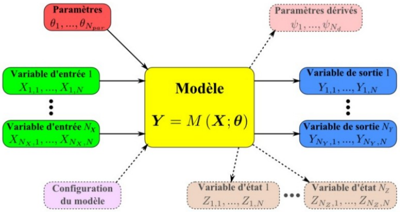


From BaRatin to BaM

Replacing $Q = f(h)$ by another model does not really change the statistical framework used in BaRatin.

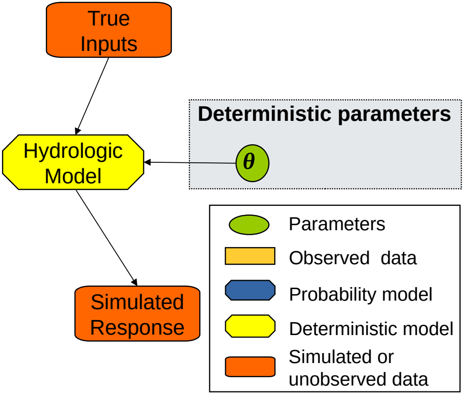
$\longrightarrow$ **Development of BaM** (Bayesian Modeling)

'Any' model can be plugged in. Examples:
sediment transport, optical camera model, chemistry, 1D hydraulics, hydrologic.



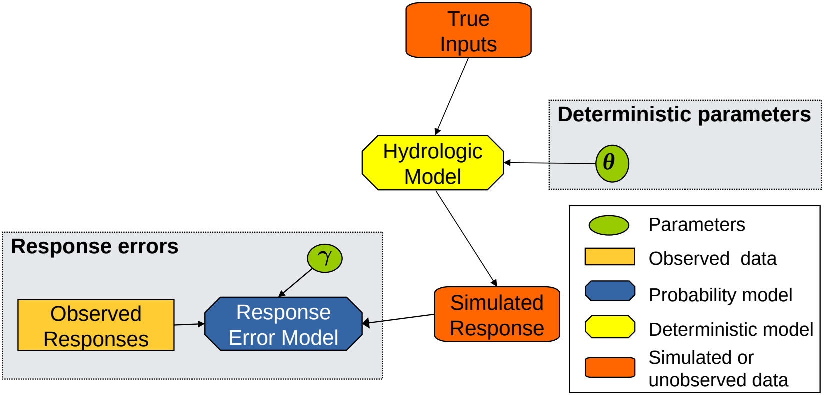
Uncertainty in & around hydrologic models

Bayesian Total Error Analysis



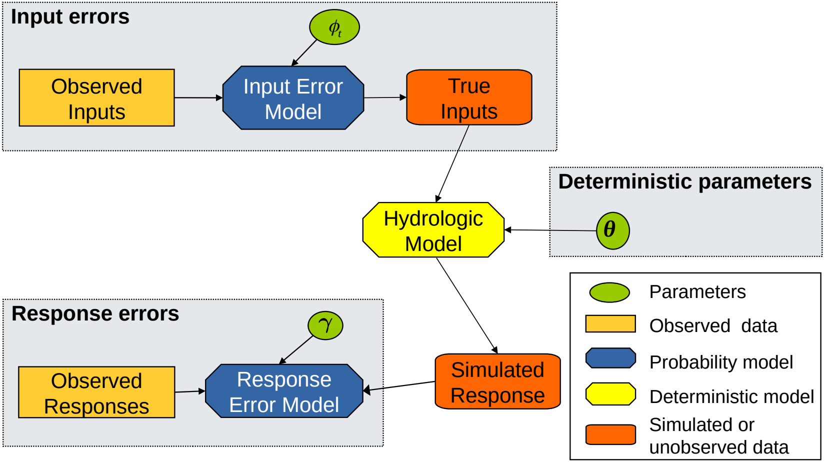
$$\hat{Q}_t = \mathcal{M}(\boldsymbol{r}_{1\to t}|\boldsymbol{\theta})$$

Bayesian Total Error Analysis



$$\tilde{Q}_t = \mathcal{M}(\boldsymbol{r}_{1\to t}|\boldsymbol{\theta}) + \delta_t + \varepsilon_t$$

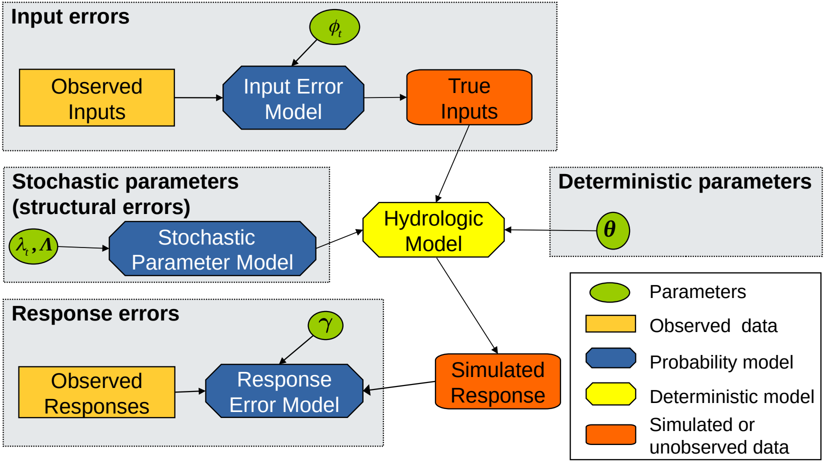
Bayesian Total Error Analysis



$$\tilde{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}) + \delta_t + \varepsilon_t$$

$$\phi_k \sim \mathcal{LN}(\mu, \sigma)$$

Bayesian Total Error Analysis

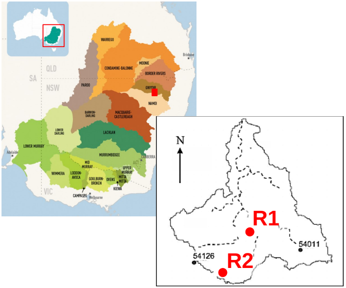


$$\tilde{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \delta_t + \varepsilon_t$$

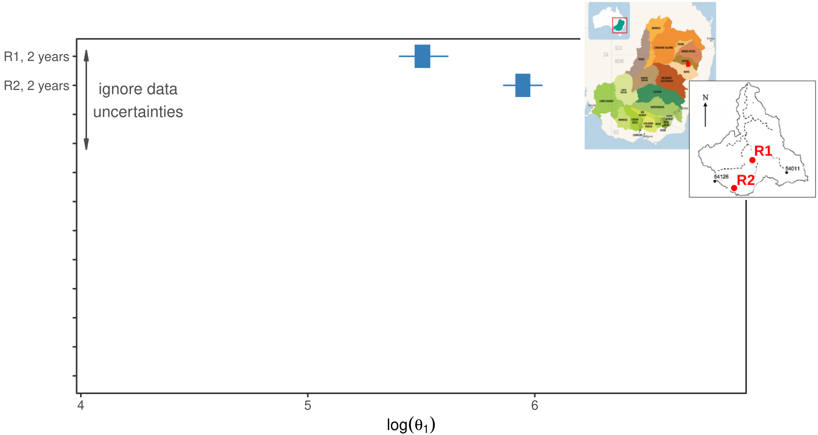
$$\lambda_k \sim \mathcal{D}(\Lambda)$$

Recognizing Data Uncertainty

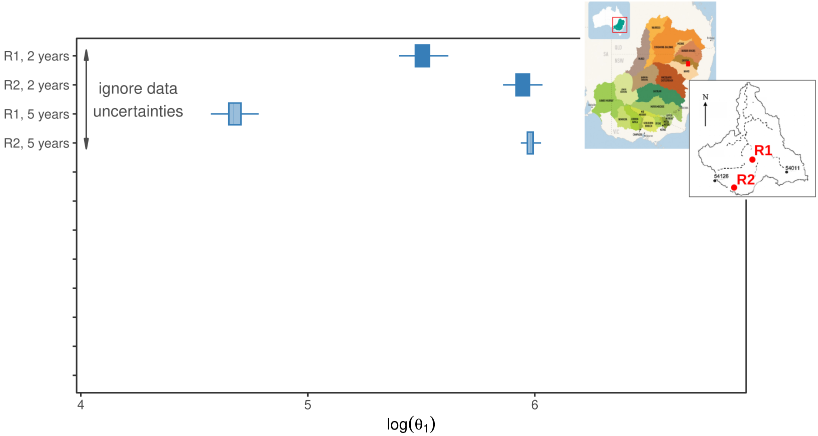
- Horton catchment (1920 km^2), Australia
- Calibrate $\mathcal{M}(\boldsymbol{\theta})$ using rainfall data from either R1 or R2
- Compare inference schemes ignoring or acknowledging streamflow and rainfall data uncertainties (uQ and uR)



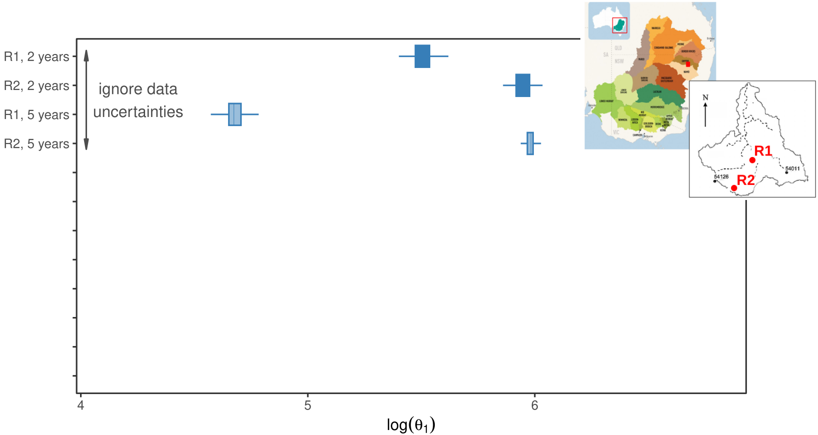
Recognizing Data Uncertainty



Recognizing Data Uncertainty



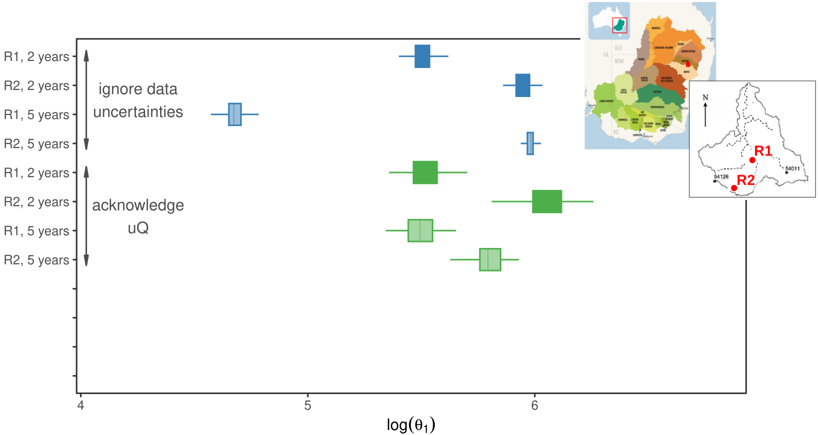
Recognizing Data Uncertainty



- Ignoring data uncertainties $\Longrightarrow$ parameters fit to data errors

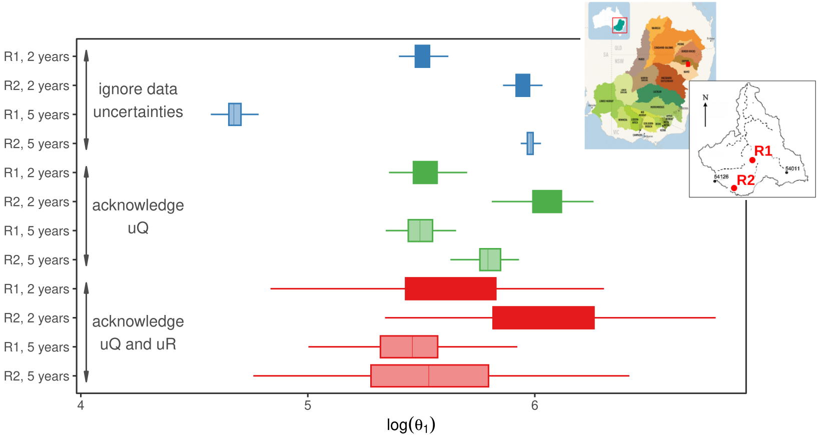
  reliability of streamflow prediction? parameter regionalization?

Recognizing Data Uncertainty



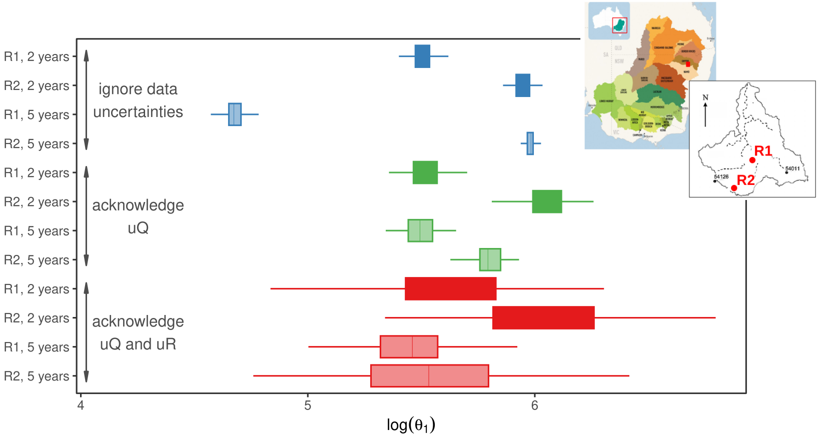
- Ignoring data uncertainties $\Longrightarrow$ parameters fit to data errors
  reliability of streamflow prediction? parameter regionalization?

Recognizing Data Uncertainty



- Ignoring data uncertainties $\Longrightarrow$ parameters fit to data errors
  reliability of streamflow prediction? parameter regionalization?

Recognizing Data Uncertainty



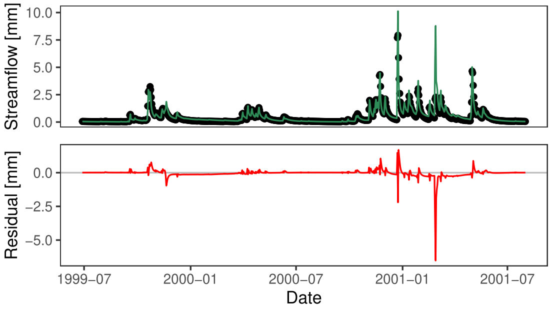
- Ignoring data uncertainties $\Longrightarrow$ parameters fit to data errors
  reliability of streamflow prediction? parameter regionalization?
- Acknowledging data uncertainties $\Longrightarrow$ parameters more stable
  vaguely correct rather than precisely wrong!

Structural Uncertainty through Stochastic Parameters

⟶ What do structural errors look like?

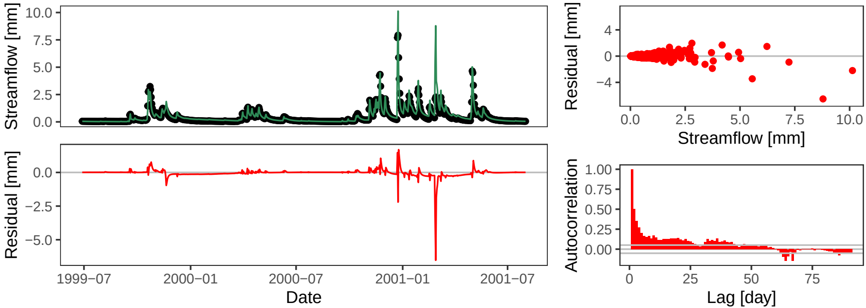
Structural Uncertainty through Stochastic Parameters

⟶ What do structural errors look like?



Structural Uncertainty through Stochastic Parameters

⟶ What do structural errors look like?



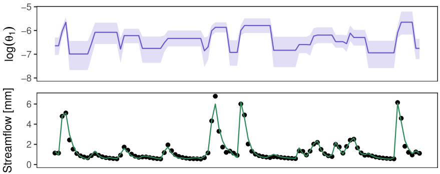
- High heteroscedasticity, complex dependence structure
- Structural error model may be as complex as the model itself...

Structural Uncertainty through Stochastic Parameters

- Allow θ_1 to vary in time according to : $log(\theta_{1,k}) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma)$

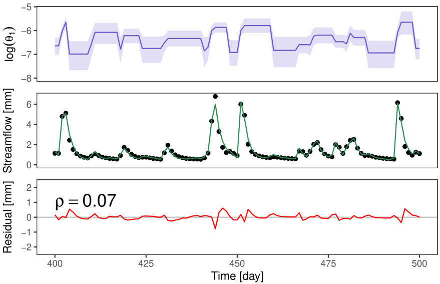
Structural Uncertainty through Stochastic Parameters

- Allow θ_1 to vary in time according to : $log(\theta_{1,k}) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma)$



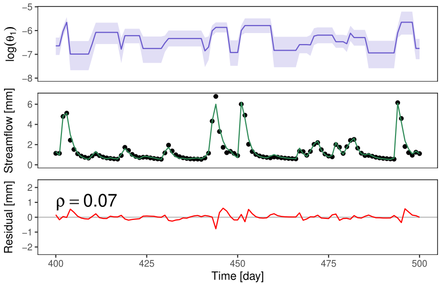
Structural Uncertainty through Stochastic Parameters

- Allow θ_1 to vary in time according to : $log(\theta_{1,k}) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma)$
- Residuals are smaller and much simpler



Structural Uncertainty through Stochastic Parameters

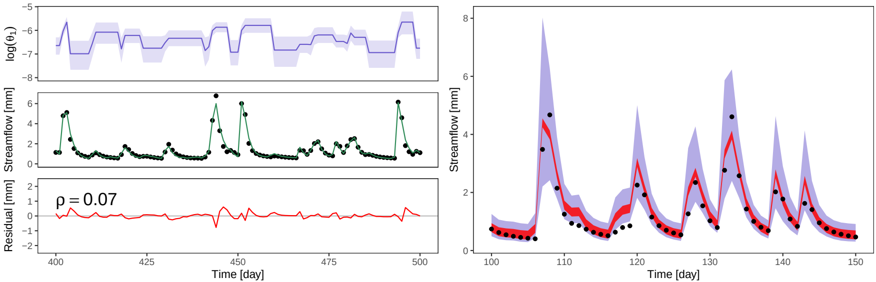
- Allow θ_1 to vary in time according to : $log(\theta_{1,k}) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma)$
- Residuals are smaller and much simpler



⟶ Stochastic parameters allows using simple statistical models for structural uncertainty, at the cost of a demanding inference

Structural Uncertainty through Stochastic Parameters

- Allow θ_1 to vary in time according to : $log(\theta_{1,k}) \overset{iid}{\sim} \mathcal{N}(\mu, \sigma)$
- Residuals are smaller and much simpler
- In prediction, randomly sample $log(\theta_1)$ values from $\mathcal{N}(\hat{\mu}, \hat{\sigma})$



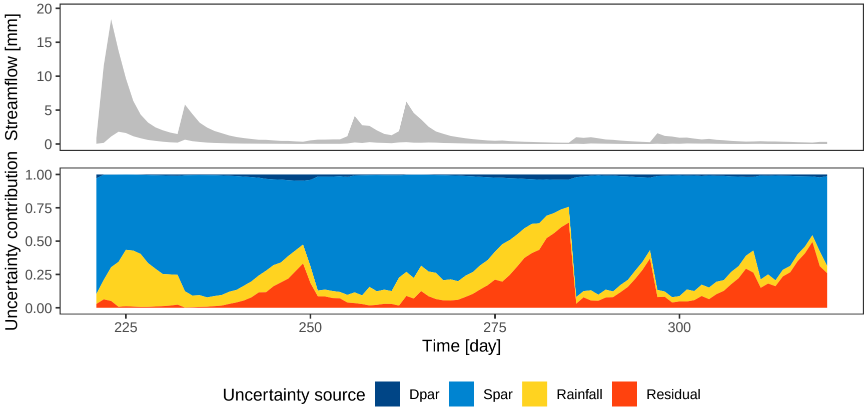
$\longrightarrow$ Stochastic parameters allows using simple statistical models for structural uncertainty, at the cost of a demanding inference

Decomposing Predictive Uncertainty

Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\rightarrow t} \times \phi_{1\rightarrow t}|\boldsymbol{\theta}, \boldsymbol{\lambda}_{1\rightarrow t}) + \varepsilon_t$

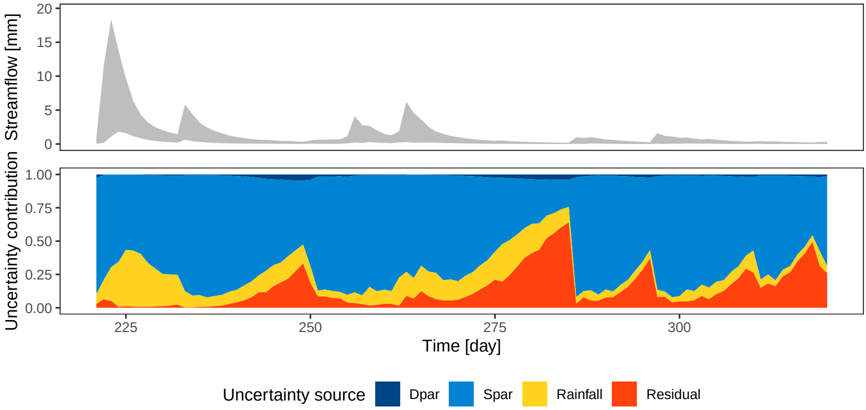
Decomposing Predictive Uncertainty

Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$



Decomposing Predictive Uncertainty

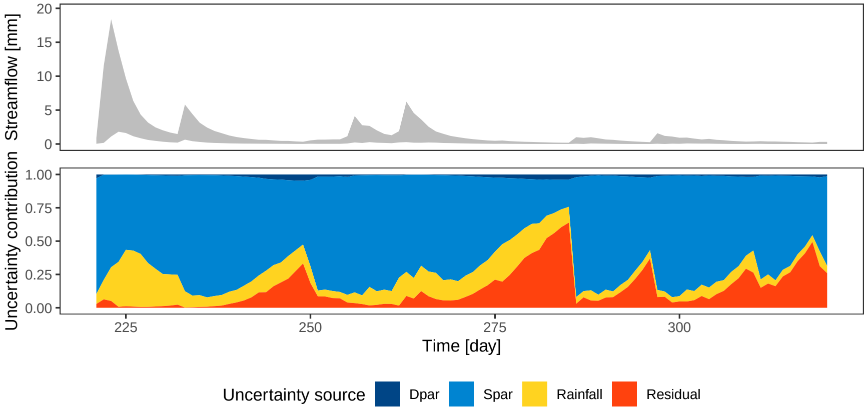
Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$



- main sources of predictive uncertainty $\implies$ ways to reduce it

Decomposing Predictive Uncertainty

Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$



- main sources of predictive uncertainty $\implies$ ways to reduce it
- specific sources may be turned off or replaced

Decomposing Predictive Uncertainty

Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$

⟶ **Requirements:**

1. data uncertainties are known beforehand
   If you don't know the quality of data, you won't know the quality of the model

Decomposing Predictive Uncertainty

Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$

⟶ **Requirements:**

1. data uncertainties are known beforehand

   If you don't know the quality of data, you won't know the quality of the model

2. data uncertainties are not too large

   A minimal data precision may be necessary to estimate structural uncertainty

Decomposing Predictive Uncertainty

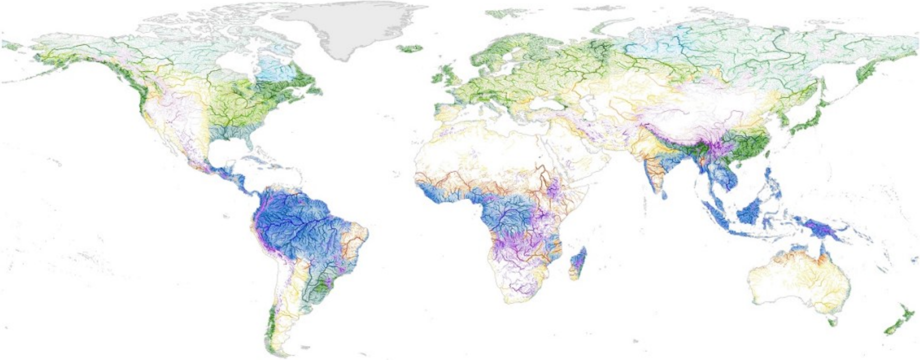
Putting everything together... $\hat{Q}_t = \mathcal{M}(\tilde{\boldsymbol{r}}_{1\to t} \times \phi_{1\to t} | \boldsymbol{\theta}, \boldsymbol{\lambda}_{1\to t}) + \varepsilon_t$

⟶ **Requirements:**

1. data uncertainties are known beforehand
   If you don't know the quality of data, you won't know the quality of the model
2. data uncertainties are not too large
   A minimal data precision may be necessary to estimate structural uncertainty

Otherwise, the problem is **ill-posed** due to **non-identifiable parameters**

Hydrologic Variability

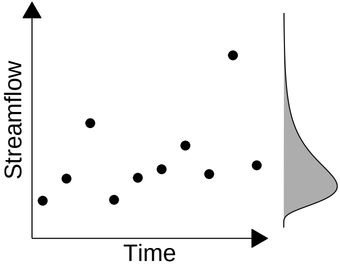


crédit: HydroSHEDS

Estimating Distributions

Managing water resources and risks often requires estimating distributions:

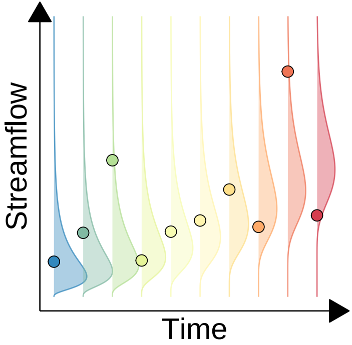
- Dam design: T-year quantile from the **marginal** distribution of streamflow



Estimating Distributions

Managing water resources and risks often requires estimating distributions:

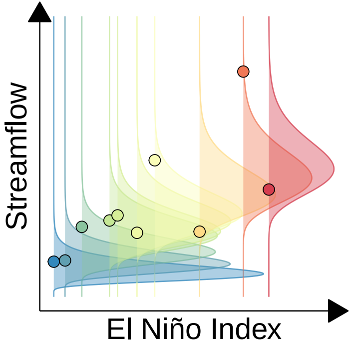
- Dam design: T-year quantile from the **marginal** distribution of streamflow
- Does climate change compromise dam design? **time-varying distribution**



Estimating Distributions

Managing water resources and risks often requires estimating distributions:

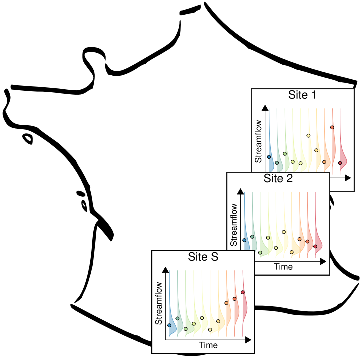
- Dam design: T-year quantile from the **marginal** distribution of streamflow
- Does climate change compromise dam design? **time-varying distribution**
- Adapt dam operation to climate state: **conditional distribution** of streamflow given some predictor.



Estimating Distributions

Managing water resources and risks often requires estimating distributions:

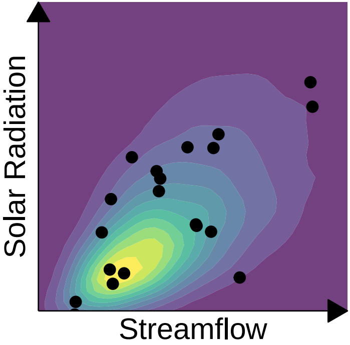
- Dam design: T-year quantile from the **marginal** distribution of streamflow
- Does climate change compromise dam design? **time-varying distribution**
- Adapt dam operation to climate state: **conditional distribution** of streamflow given some predictor.
- Large hydroelectricity companies ask this at the national scale: **space-and-time varying distribution**



Estimating Distributions

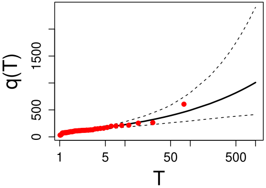
Managing water resources and risks often requires estimating distributions:

- Dam design: T-year quantile from the **marginal** distribution of streamflow
- Does climate change compromise dam design? **time-varying distribution**
- Adapt dam operation to climate state: **conditional distribution** of streamflow given some predictor.
- Large hydroelectricity companies ask this at the national scale: **space-and-time varying distribution**
- Renewable energy production depends on the **joint distribution** of wind, solar radiation and streamflow.

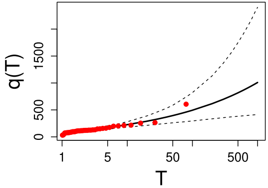


Estimating a Marginal Distribution

A simple statistical problem, plagued with sampling uncertainty.
Ex: estimate $GEV(\mu, \sigma, \xi)$ with 40 years of data.



Estimating a Marginal Distribution

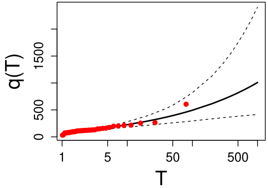


A simple statistical problem, plagued with sampling uncertainty.
Ex: estimate $GEV(\mu, \sigma, \xi)$ with 40 years of data.

A possible solution: use ancient data such as historical information, flood marks, old stage series, etc.

1. Hydraulics to translate this information in terms of streamflow.
2. Adapted statistical methods (censoring, data uncertainties, etc.)

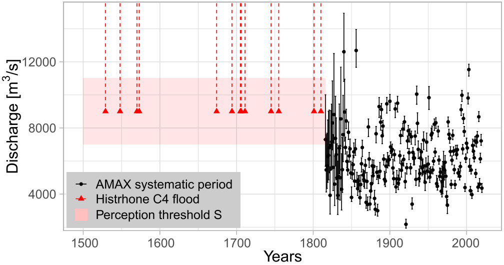
Estimating a Marginal Distribution

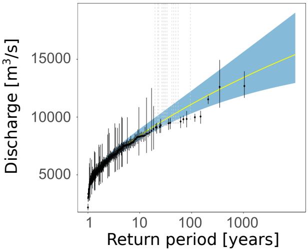


A simple statistical problem, plagued with sampling uncertainty.
Ex: estimate $GEV(\mu, \sigma, \xi)$ with 40 years of data.

A possible solution: use ancient data such as historical information, flood marks, old stage series, etc.

1. Hydraulics to translate this information in terms of streamflow.
2. Adapted statistical methods (censoring, data uncertainties, etc.)





(M. Lucas PhD)

Estimating a Time-Varying Distribution

$\longrightarrow$ Time-varying models

$$\mathcal{M}_0 : Y(t) \sim GEV(\mu, \sigma, \xi)$$

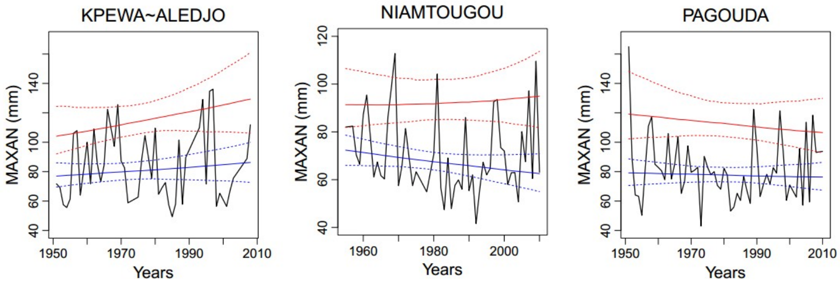
$$\mathcal{M}_1 : Y(t) \sim GEV\left(\mu \times (1 + \lambda x(t)), \sigma, \xi\right)$$

Estimating a Time-Varying Distribution

⟶ Time-varying models

$$\mathcal{M}_0 : Y(t) \sim GEV(\mu, \sigma, \xi)$$

$$\mathcal{M}_1 : Y(t) \sim GEV\left(\mu \times (1 + \lambda x(t)), \sigma, \xi\right)$$



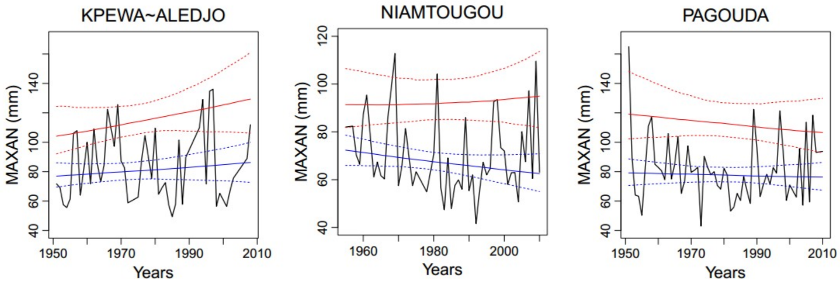
(M. Badjana PhD)

Estimating a Time-Varying Distribution

⟶ Time-varying models

$$\mathcal{M}_0 : Y(t) \sim GEV(\mu, \sigma, \xi)$$

$$\mathcal{M}_1 : Y(t) \sim GEV\left(\mu \times (1 + \lambda x(t)), \sigma, \xi\right)$$



(M. Badjana PhD)

A simple statistical problem, plagued with sampling uncertainty!

Estimating Space-and-Time-Varying Distributions

$\longrightarrow$ Space-and-time-varying models (X. Sun PhD)

$$\mathcal{M}_{loc} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda(s)x(t)), \sigma(s), \xi(s)\right)$$

Estimating Space-and-Time-Varying Distributions

$\longrightarrow$ Space-and-time-varying models (X. Sun PhD)

$$\mathcal{M}_{loc} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda(s)x(t)), \sigma(s), \xi(s)\right)$$

$$\mathcal{M}_{reg} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda \; x(t)), \sigma(s), \; \xi \;\right)$$

Estimating Space-and-Time-Varying Distributions

⟶ Space-and-time-varying models (X. Sun PhD)

$$\mathcal{M}_{loc} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda(s)x(t)), \sigma(s), \xi(s)\right)$$

$$\mathcal{M}_{reg} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda \; x(t)), \sigma(s), \; \xi \;\right)$$

- All sites contribute to the estimation of regional parameters
- Less stringent hypothesis possible (hierarchical models)
- Need to describe **spatial dependence** (elliptical copulas)

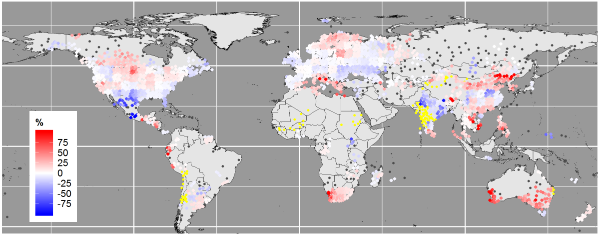
Estimating Space-and-Time-Varying Distributions

⟶ Space-and-time-varying models (X. Sun PhD)

$$\mathcal{M}_{loc} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda(s)x(t)), \sigma(s), \xi(s)\right)$$

$$\mathcal{M}_{reg} : Y(s,t) \sim GEV\left(\mu(s) \times (1 + \lambda \; x(t)), \sigma(s), \; \xi \;\right)$$

- All sites contribute to the estimation of regional parameters
- Less stringent hypothesis possible (hierarchical models)
- Need to describe **spatial dependence** (elliptical copulas)



Hidden Climate Indices Models

⟶ Standard vs. Hidden Climate Index model

$\mathcal{M}_{stand.} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda x(t))\right)$

Hidden Climate Indices Models

⟶ Standard vs. Hidden Climate Index model

$$\mathcal{M}_{stand.} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda x(t))\right)$$

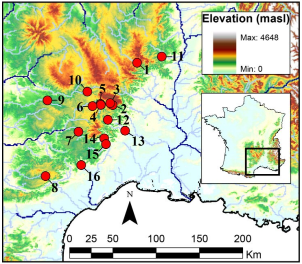
$$\mathcal{M}_{hidden} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda \tau(t))\right)$$

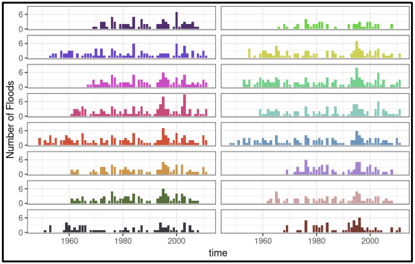
Hidden Climate Indices Models

⟶ Standard vs. Hidden Climate Index model

$$\mathcal{M}_{stand.} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda x(t))\right)$$

$$\mathcal{M}_{hidden} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda \tau(t))\right)$$



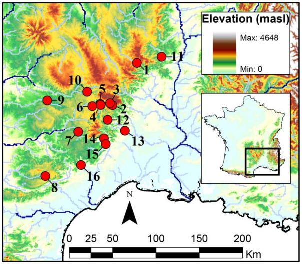


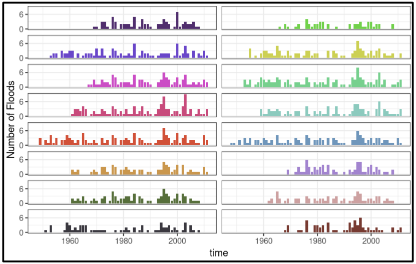
Hidden Climate Indices Models

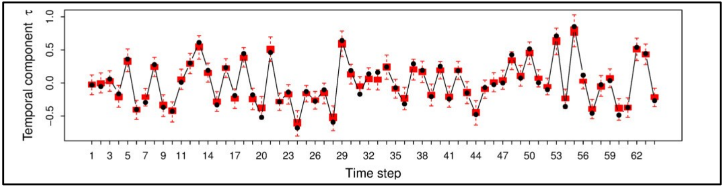
⟶ Standard vs. Hidden Climate Index model

$$\mathcal{M}_{stand.} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda x(t))\right)$$

$$\mathcal{M}_{hidden} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda \tau(t))\right)$$







Hidden Climate Indices Models

⟶ Generalization to non-homogeneous regions

$$\mathcal{M}_0 : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$$

Hidden Climate Indices Models

$\longrightarrow$ Generalization to non-homogeneous regions

$$\mathcal{M}_0 : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$$

$$\mathcal{M} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda_1(s)\tau_1(t) + \ldots + \lambda_K(s)\tau_K(t))\right)$$

Hidden Climate Indices Models

$\longrightarrow$ Generalization to non-homogeneous regions

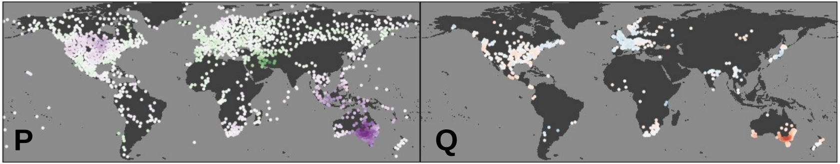
$$\mathcal{M}_0 : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$$

$$\mathcal{M} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda_1(s)\tau_1(t) + \ldots + \lambda_K(s)\tau_K(t))\right)$$

$\longrightarrow$ Multi-variate extension + large-scale scalability

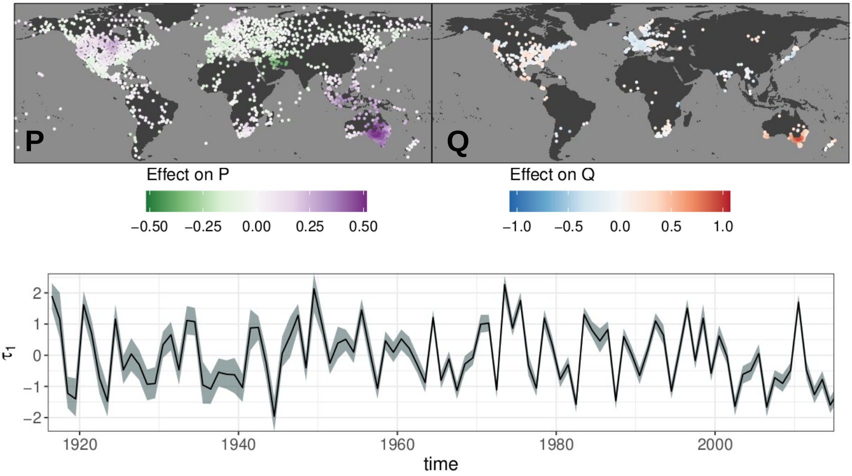
Hidden Climate Indices Models

⟶ A global-scale analysis of floods and heavy precipitation



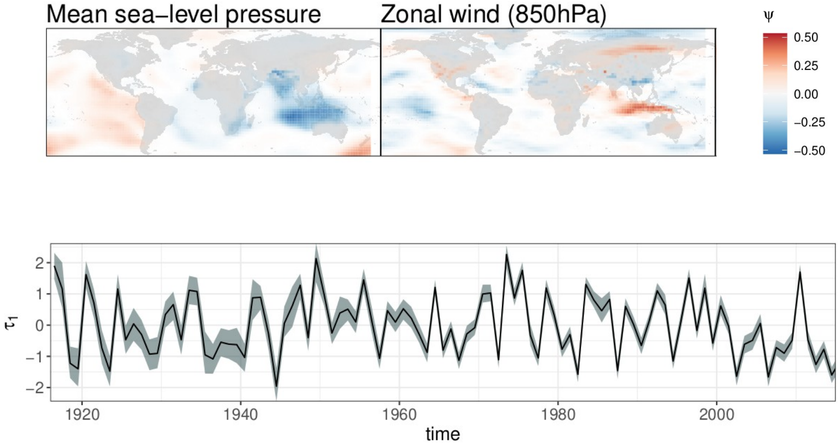
Hidden Climate Indices Models

⟶ A global-scale analysis of floods and heavy precipitation



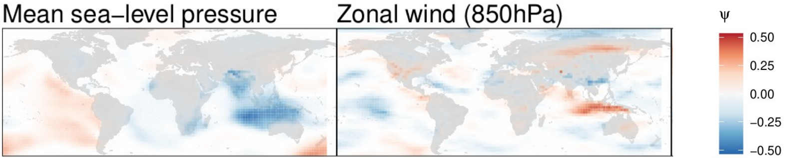
Hidden Climate Indices Models

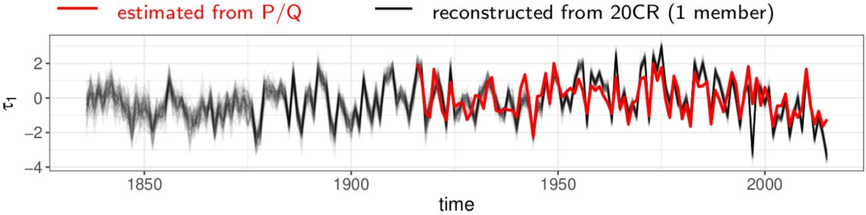
⟶ A global-scale analysis of floods and heavy precipitation



Hidden Climate Indices Models

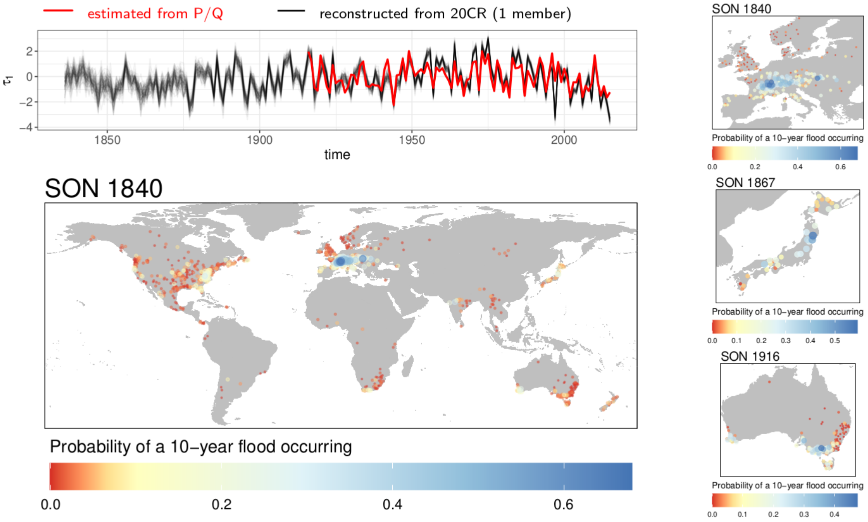
⟶ A global-scale analysis of floods and heavy precipitation



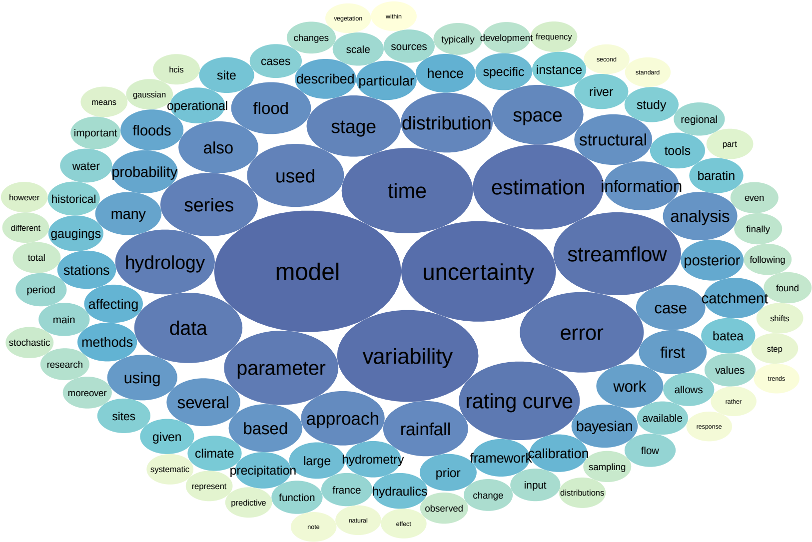


Hidden Climate Indices Models

⟶ A global-scale analysis of floods and heavy precipitation



Conclusion


vegetation
within
changes
scale
sources
typically
development
frequency
cases
hcis
site
described
particular
hence
specific
instance
second
standard
means
gaussian
operational
flood
stage
distribution
space
river
study
important
floods
also
structural
regional
part
water
probability
used
time
estimation
tools
information
baratin
however
historical
series
many
analysis
even
different
gaugings
model
uncertainty
streamflow
finally
total
hydrology
posterior
following
stations
period
affecting
case
catchment
found
main
data
error
shifts
stochastic
methods
batea
step
research
parameter
variability
first
values
using
trends
moreover
rating curve
work
allows
several
rather
sites
approach
rainfall
available
based
bayesian
response
given
climate
flow
systematic
precipitation
large
hydrometry
prior
framework
calibration
sampling
represent
france
hydraulics
input
change
distributions
predictive
function
observed
note
natural
effect

Summary

- Development of probabilistic models enabling the production of uncertain hydrologic predictions

Summary

- Development of probabilistic models enabling the production of uncertain hydrologic predictions
- Key contributions:
  1. Uncertainties affecting streamflow time series
  2. Uncertainties affecting hydrologic models
  3. Modeling hydrologic variability in space, time and between variables

Perspectives

⟶ **Uncertainties in Hydrologic Data and Models**

Perspectives

⟶ **Uncertainties in Hydrologic Data and Models**

1. **Treatment of structural errors**
   - Further assess the interest of stochastic parameters
   - Evaluate non-zero-mean residual models (conditional biases)

Perspectives

⟶ **Uncertainties in Hydrologic Data and Models**

1. **Treatment of structural errors**
   - Further assess the interest of stochastic parameters
   - Evaluate non-zero-mean residual models (conditional biases)

2. **Distributed models**
   - Hydrologic (SMASH) or Hydraulics
   - Define spatialized error models

Perspectives

⟶ **Uncertainties in Hydrologic Data and Models**

1. **Treatment of structural errors**
   - Further assess the interest of stochastic parameters
   - Evaluate non-zero-mean residual models (conditional biases)

2. **Distributed models**
   - Hydrologic (SMASH) or Hydraulics
   - Define spatialized error models

3. **Beyond streamflow series**
   - What other data types are valuable to constrain model calibration?
   - Sporadic gaugings, flood videos, flood marks, satellite (SWOT), etc.

Perspectives

⟶ **Hydrologic Variability**

Perspectives

⟶ **Hydrologic Variability**

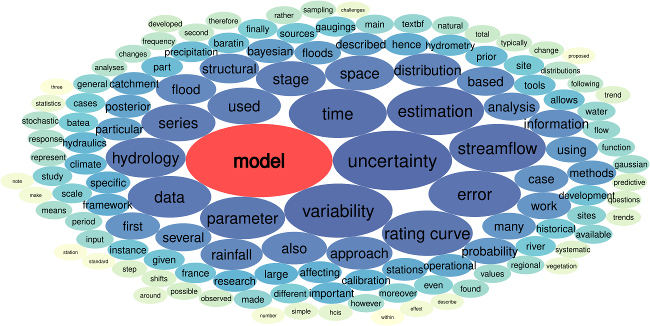
1. **Methodological questions**
   - Using Hidden Climate Indices to derive spatial extremes models?
   - Clarify the link between HCI models and Machine Learning approaches such as probabilistic PCA

Perspectives

⟶ **Hydrologic Variability**

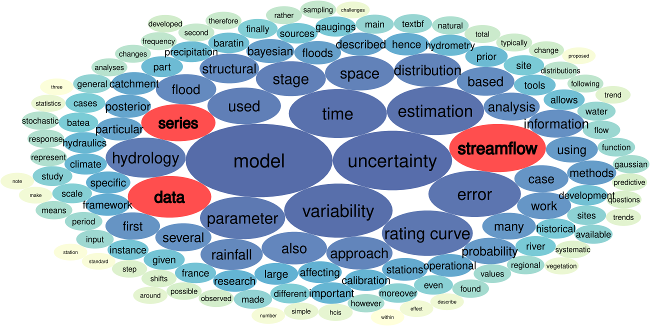
1. **Methodological questions**
   - Using Hidden Climate Indices to derive spatial extremes models?
   - Clarify the link between HCI models and Machine Learning approaches such as probabilistic PCA
2. **Applications**
   - Global droughts (low flows and precipitation deficit)
   - Relation with the wildfire hazard: a testbed for multi-risk assessment?

Key aspects



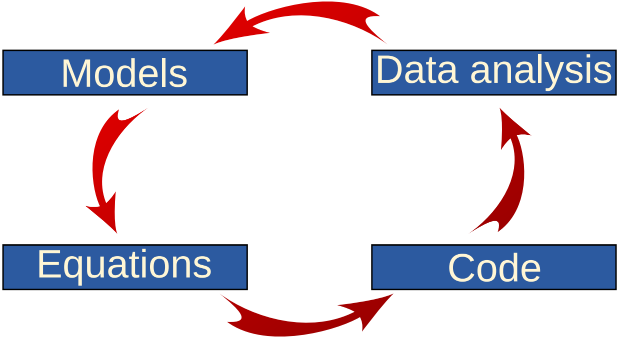
- Probabilistic modeling is central

Key aspects



- Probabilistic modeling is central
- Data are invaluable, and models should adapt to data

Key aspects



- Probabilistic modeling is central
- Data are invaluable, and models should adapt to data
- Coding is an important component

Key aspects

- Probabilistic modeling is central
- Data are invaluable, and models should adapt to data
- Coding is an important component
- Operational transfer as both an output and an input of research work

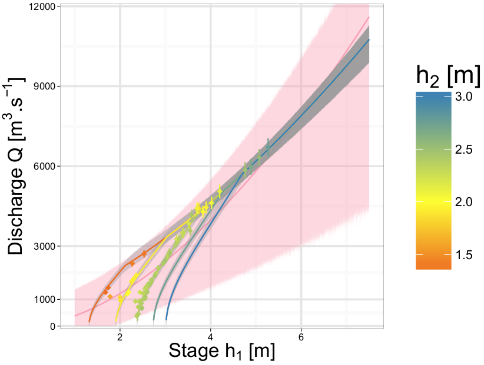
And most importantly... Thank You!

Xtra

Complex Rating Curves $Q = f(h, \ldots)$

$\longrightarrow$ Stage-Fall-Discharge model (V. Mansanarez)

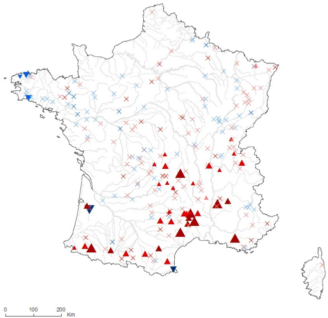
$Q = f(h_1, h_2)$ due to variable backwater (dam, tide)
Used operationally for dam-influenced stations
Work still in progress for tide-influenced stations



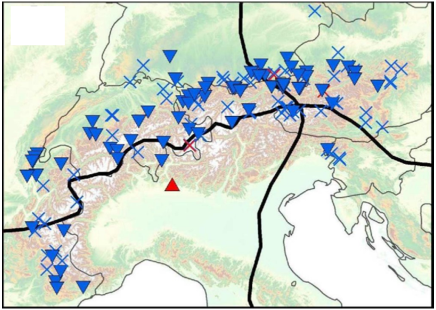
Estimating a Time-Varying Distribution

⟶ Trend detection studies

Drought duration, 1968-2008

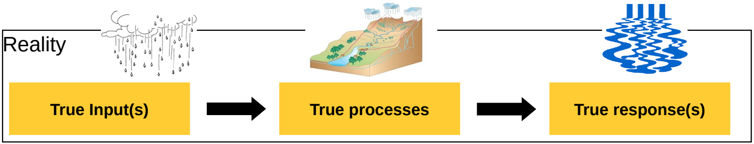


Snowmelt flows start, 1961-2006



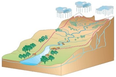
(I. Giuntoli, A. Bard)

Error Sources in Hydrologic Modeling



Error Sources in Hydrologic Modeling

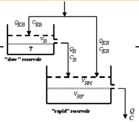
Reality



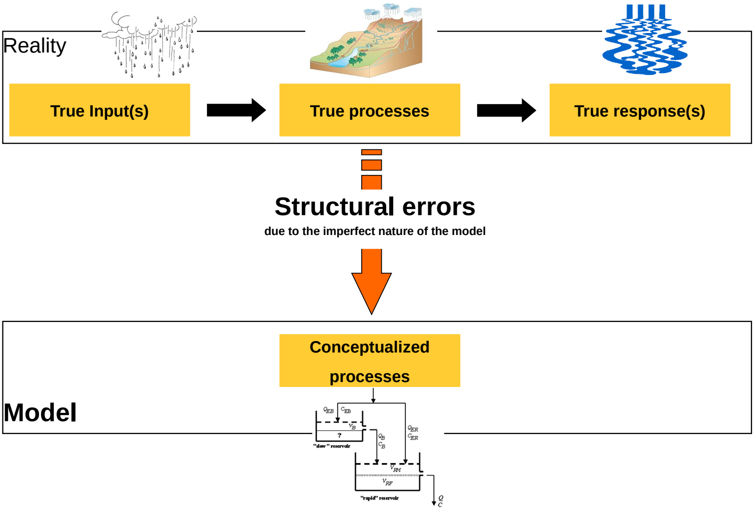
True Input(s) → **True processes** → **True response(s)**

Model

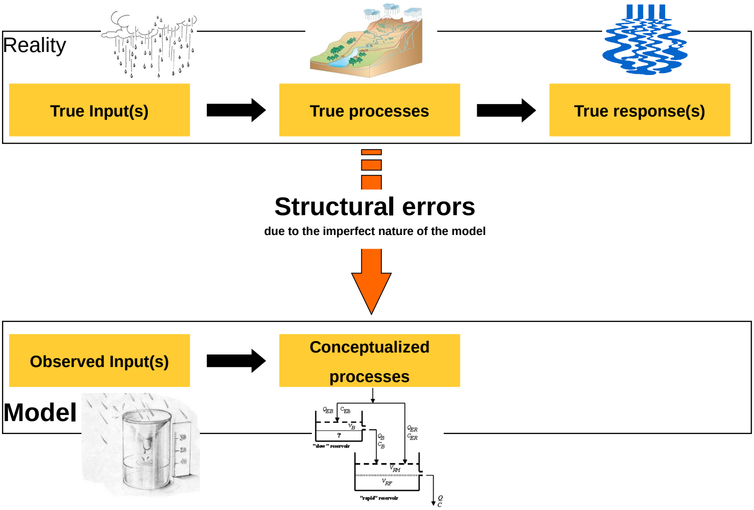
Conceptualized processes



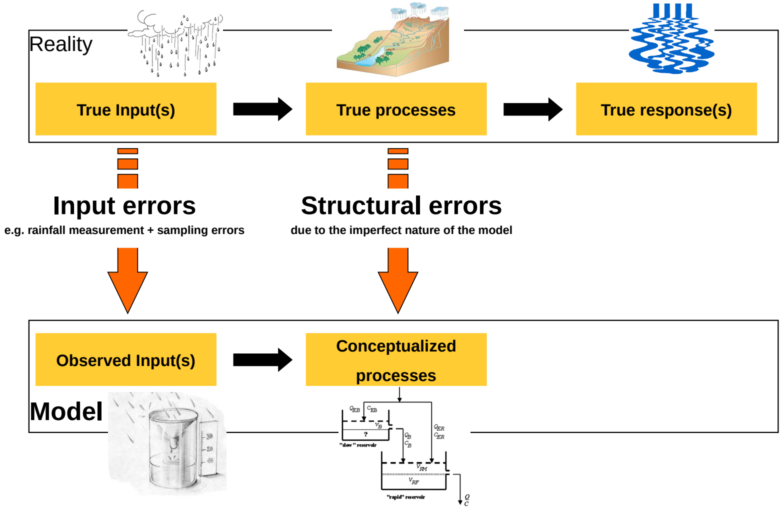
Error Sources in Hydrologic Modeling



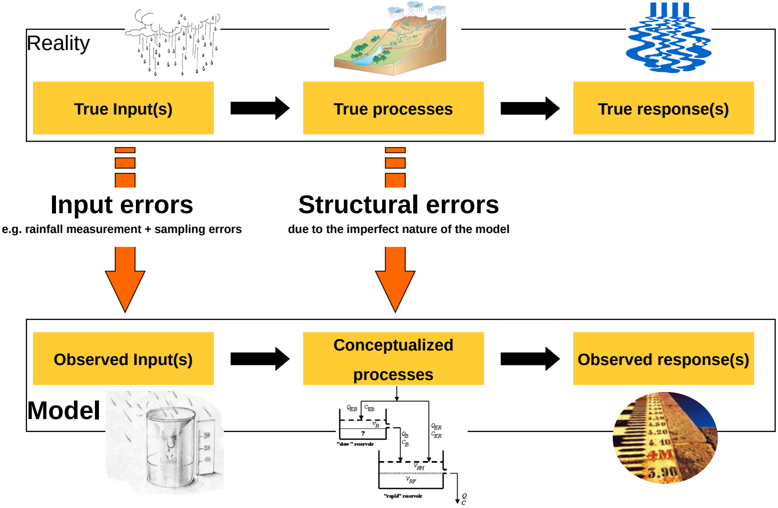
Error Sources in Hydrologic Modeling



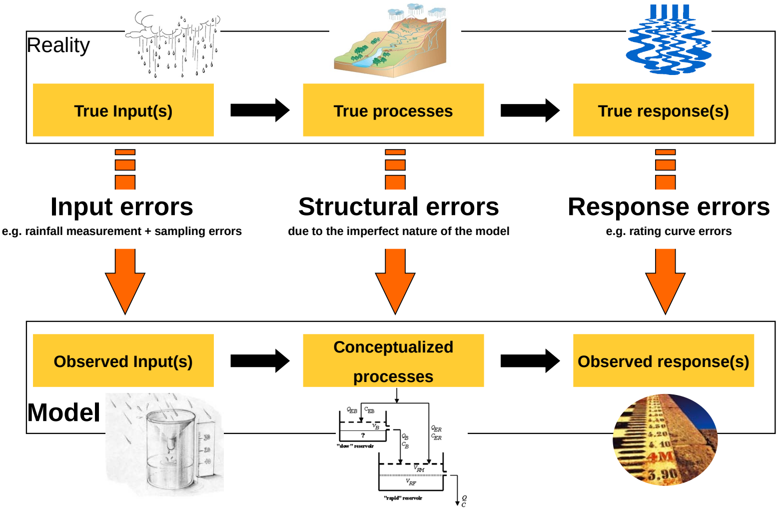
Error Sources in Hydrologic Modeling



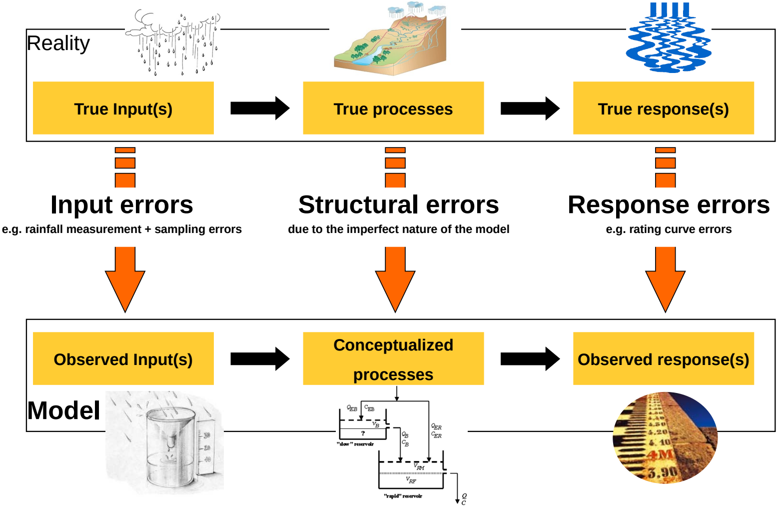
Error Sources in Hydrologic Modeling



Error Sources in Hydrologic Modeling

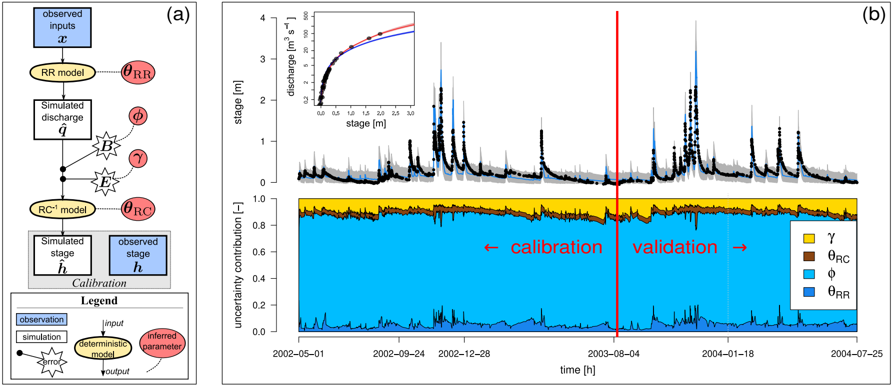


Error Sources in Hydrologic Modeling

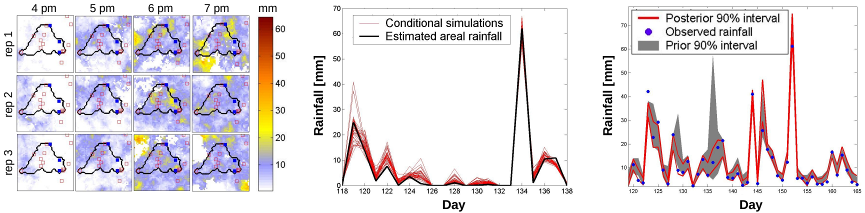


+ Parametric uncertainty

Systematic Rating Curve Errors



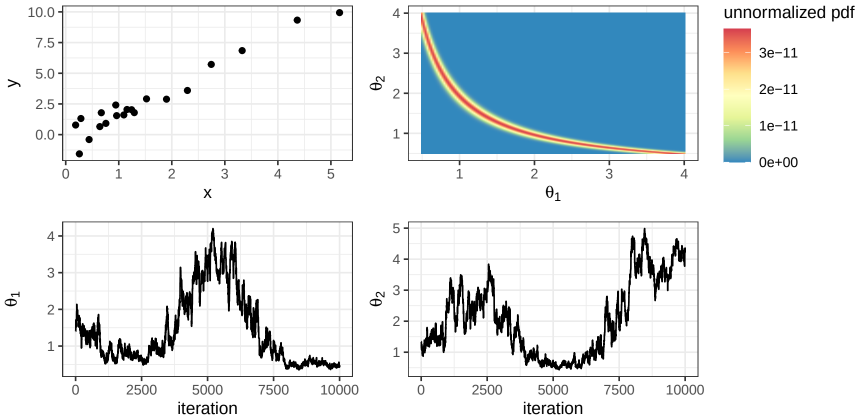
Rainfall Uncertainty



(a) (b) (c)

Non-identifiability

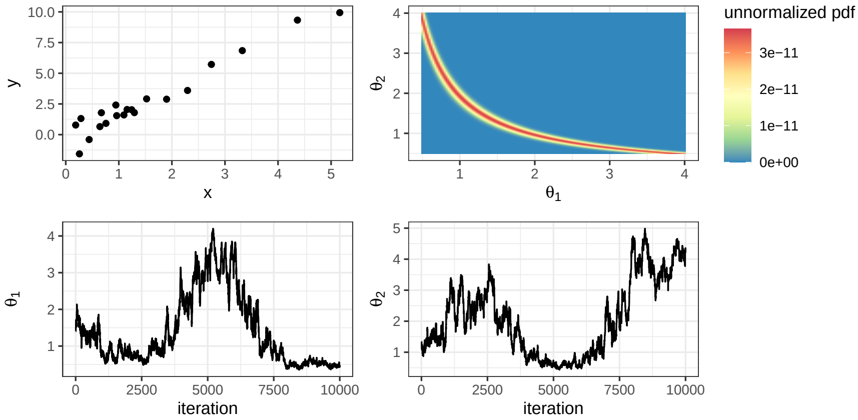
Model: $\hat{y} = \theta_1 \theta_2 x$



Identifying θ_1 and θ_2 based on this information is an ill-posed problem.

Non-identifiability

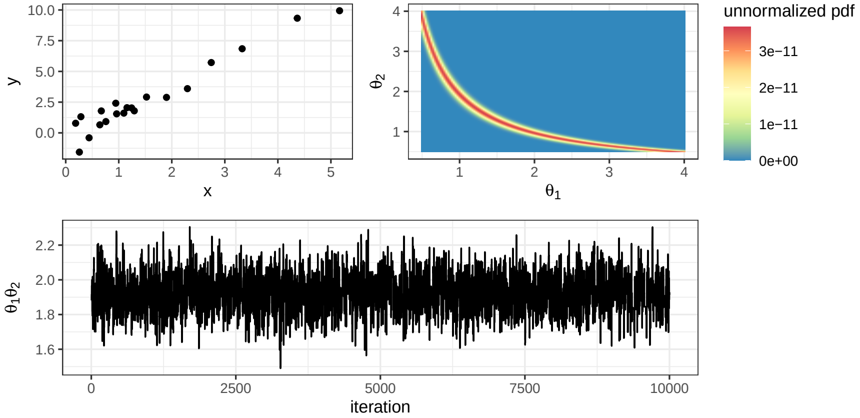
Model: $\hat{y} = \theta_1 \theta_2 x$



Identifying θ_1 and θ_2 based on this information is an ill-posed problem.
Non-identifiability and ill-posedness are frequently encountered in the calibration of hydrologic models

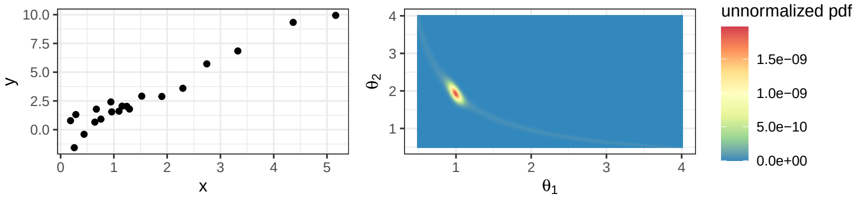
Non-identifiability

Solution 1: 'under-parameterize', $\hat{y} = \gamma x$ with $\gamma = \theta_1\theta_2$



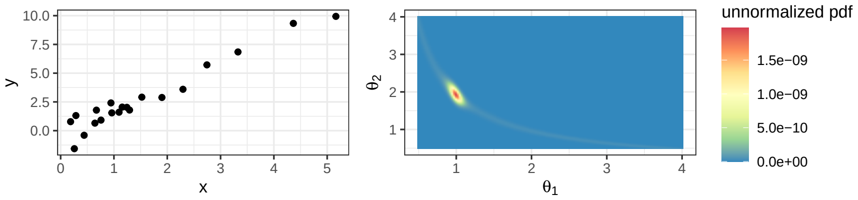
Non-identifiability

Solution 2: use prior, $\theta_1 \sim \mathcal{N}(1, 0.1^2)$

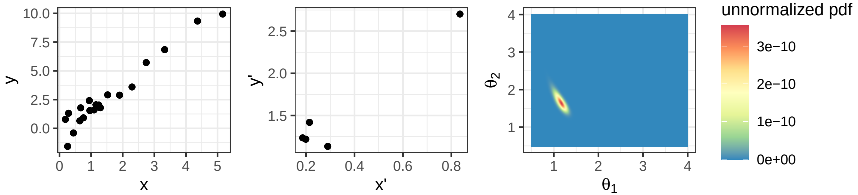


Non-identifiability

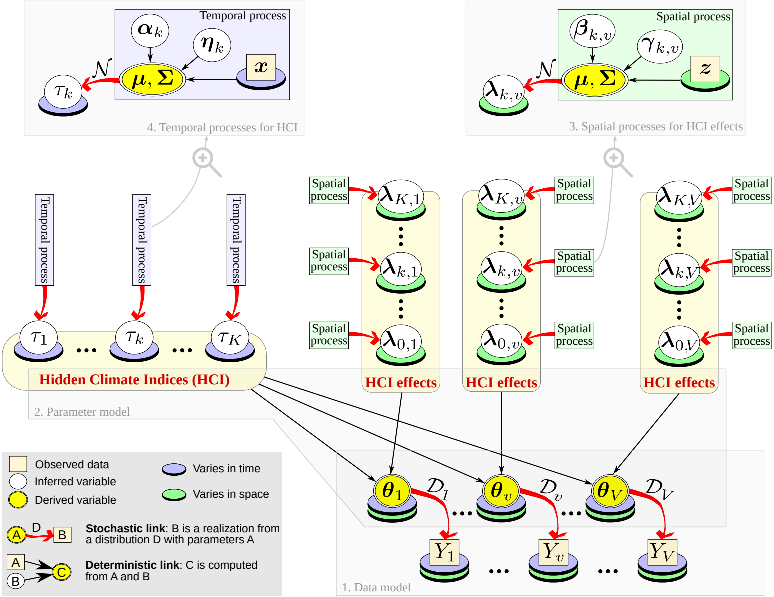
Solution 2: use prior, $\theta_1 \sim \mathcal{N}(1, 0.1^2)$



Solution 3: use other data (not more of the same!)
$\hat{y} = \theta_1\theta_2 x$ and $\hat{y}' = \exp(\theta_1 x')$



HCI framework



Hidden Climate Indices Models

⟶ Generalization to non-homogeneous regions

$$\mathcal{M}_0 : Y(s, t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$$

Hidden Climate Indices Models

⟶ Generalization to non-homogeneous regions

$\mathcal{M}_0 : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$

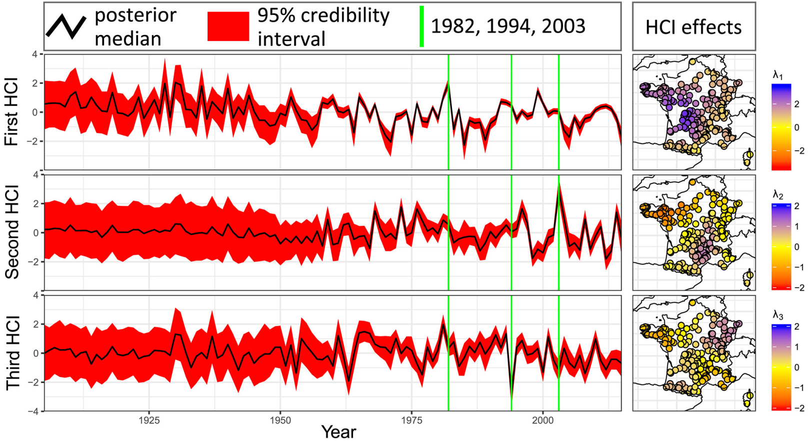
$\mathcal{M} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda_1(s)\tau_1(t) + \ldots + \lambda_K(s)\tau_K(t))\right)$

Hidden Climate Indices Models

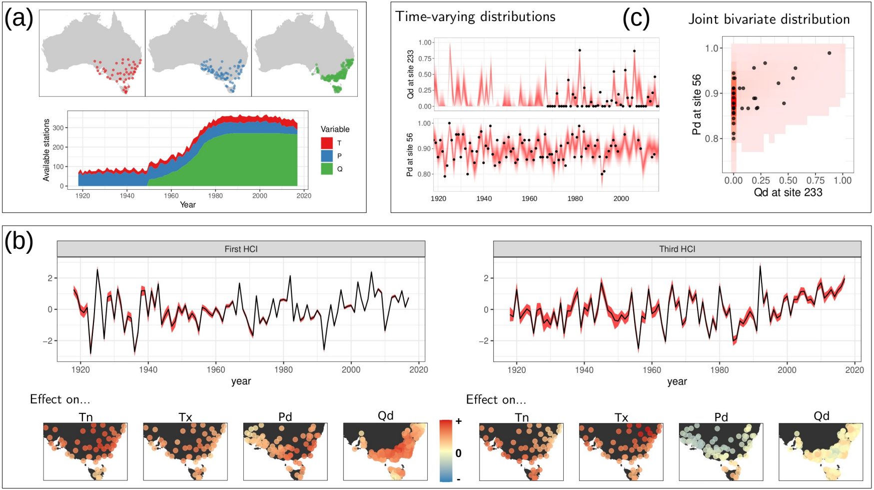
⟶ Generalization to non-homogeneous regions

$\mathcal{M}_0 : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda\tau(t))\right)$

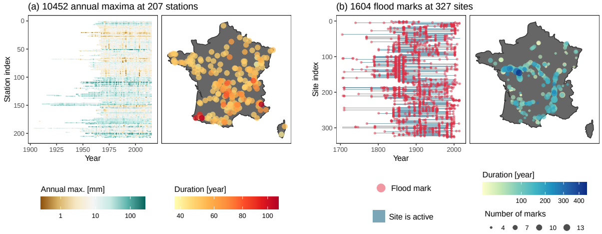
$\mathcal{M} : Y(s,t) \sim \mathcal{D}\left(\theta(s) \times (1 + \lambda_1(s)\tau_1(t) + \ldots + \lambda_K(s)\tau_K(t))\right)$



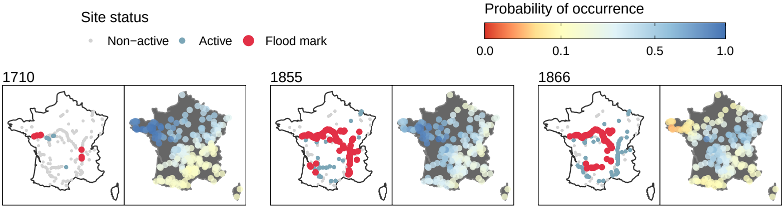
Hot-and-Dry Australian Summers



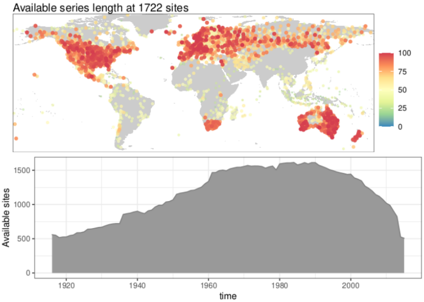
Flood Marks Case Study

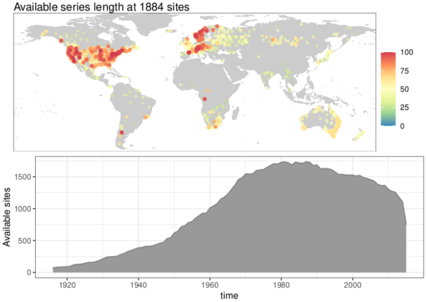


$$\begin{cases} Q(s,t) \sim GEV\left(e^{\mu_0(s)} \times \left(1+\sum_{k=1}^{K} \mu_k(s)\tau_k(t)\right), e^{\mu_0(s)} \times e^{\gamma(s)}, \xi(s)\right) \\ O(r,t) \sim \mathcal{B}\left(logit^{-1}\left(\lambda_0(r)+\sum_{k=1}^{K} \lambda_k(r)\tau_k(t)\right)\right) \end{cases}$$



Floods and Heavy Precipitation at the Global Scale





$$
\begin{cases}
P(s,t) \sim Beta\left(\mu_P(s,t), \nu_P(s,t)\right); Q(s,t) \sim Beta\left(\mu_Q(s,t), \nu_Q(s,t)\right) \\
logit\left(\mu_P(s,t)\right) = \zeta_{\mu_P}(s) + \sum_{k=1}^{K} \lambda_{k,P}(s)\tau_k(t) + \sum_{k=1}^{K} \theta_{k,P}(s)\delta_k(t) \\
logit\left(\mu_Q(s,t)\right) = \zeta_{\mu_Q}(s) + \sum_{k=1}^{K} \lambda_{k,Q}(s)\tau_k(t) + \sum_{k=1}^{K} \theta_{k,Q}(s)\omega_k(t)
\end{cases}
$$