Potsdam Institute for Climate Impact Research

How to Tailor my Process-based Hydrological Model? Dynamic Identifiability Analysis of Flexible Model Structures

Tobias Pilz, Till Francke, Gabriele Baroni, Axel Bronstert

EGU General Assembly 2020, Vienna, Austria

04 May 2020

Leibniz Association

Universität Potsdam

Alma Mater Studiorum Università di Bologna

How to tailor my model?

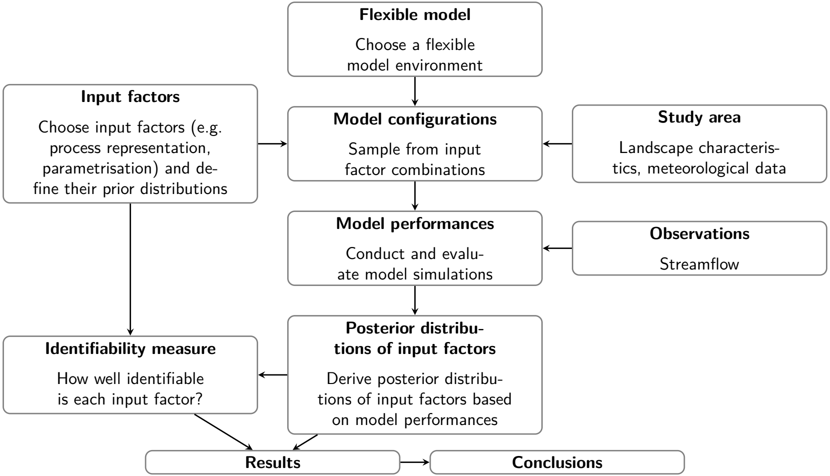
To tailor a model we can choose from alternative

- process representations
- solvers for the integration of the Ordinary Differential Equations (ODE) in time
- parametrisations

But how do we know

1. what is the optimal model configuration?
2. does the optimal configuration change over time?
3. how does it change in space?

Proposed framework *(based on the DYNIA framework by Wagener et al., 2003)*



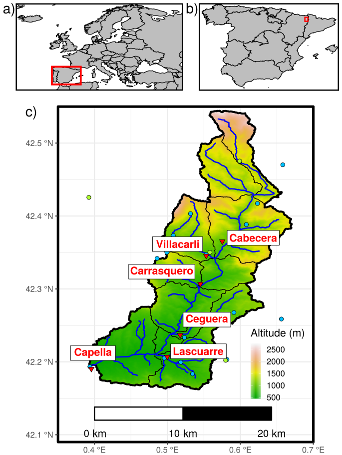
NOTE: This is an interactive presentation. Click on the boxes to learn more. Please view in presentation mode.

Study area and data: Isábena catchment, NE Spain

Back to overview

- Area: $425\,\text{km}^2$
- Mountainous topography: 500 m to 2700 m
- Rainfall spatially heterogeneous (annual sum 450 mm in lowlands, up to 1600 mm in mountains)
- Discharge at outlet $4.1\,\text{m}^3/\text{s}$ ($<1\,\text{m}^3/\text{s}$ to $370\,\text{m}^3/\text{s}$)
- Hydrological regime determined by natural factors
- Land cover: deciduous woodland, agriculture, pasture, and bushes in the valley bottoms with evergreen oaks and pines
- Many research projects, including intensive hydro-sedimentological monitoring
- Datasets: $15\,\text{m} \times 15\,\text{m}$ ASTER DEM, soil type and land-use maps, meteorological data, discharge data

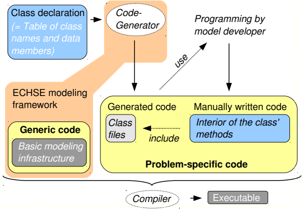
(Bronstert et al., 2014; Francke et al., 2018a; García-Ruiz et al., 2001)



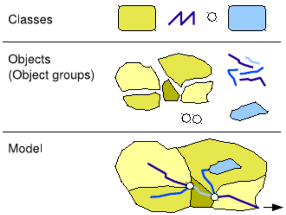
Thin black lines outline subbasins, red triangles mark the position of discharge gauges, blue and green points show gauges of rainfall and other meteorological variables, respectively.

- Free and open source: https://github.com/echse
- Two component framework
  - generic part
  - model engine (user code)
- Object-oriented programming concept (C++)
- User-friendly model implementation
- Pool of processes: simple exchange and extension of implementations
- Set of numerical integration schemes (ODE solvers)
- Arbitrary model conception (lumped vs. distributed, conceptual vs. process-oriented)

(Kneis, 2015)



ECHSE framework.



Model engine implementation.

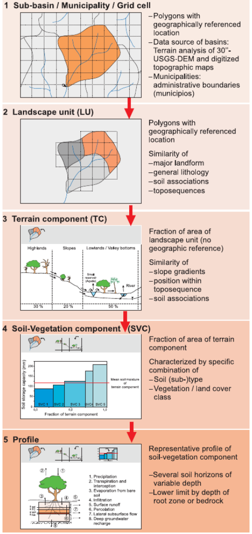
Conception of implemented engine based on the WASA-SED model

- Process-oriented hydrological model
- Complex hierarchical spatial discretisation scheme
  - Efficient simulation of hillslope-scale processes
- Adapted to semi-arid environments
  - Lateral runoff redistribution
  - Hortonian runoff
  - Ex- and re-infiltration

→ Particularly suited for environments with heterogeneous vegetation cover and considerable amounts of bare soil

→ Successfully applied in the Isábena and similar catchments

(Bronstert et al., 2014; Francke et al., 2018b; Güntner and Bronstert, 2004; Mueller et al., 2009, 2010)



Source: *Güntner (2002)*

Input factors and realisations

Input factors are adjustable elements of a model set-up. *(Pianosi et al., 2016)*
Specific values of the, in this case, discrete-valued input factors are denoted as *realisations*.

Input factor definitions for this case-study

Ia Evapotranspiration processes (32 realisations)

- Penman-Monteith, Shuttleworth & Wallace, alternatives for sub-processes (e.g. stomatal resistance)

Ib Soil water processes (8 realisations)

- Alternatives for infiltration, percolation, soil water retention

Ic Runoff concentration processes (2 realisations)

II ODE solvers (8 realisations)

III Parametrisations (1000 realisations)

- 7 parameters; realisations obtained by sampling from parameter distributions

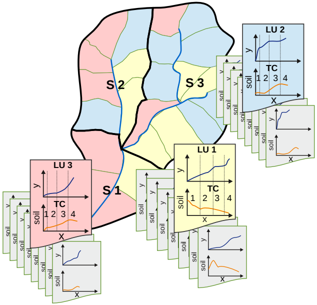
⇒ The realisations define the **prior distribution** for each input factor (equal weight for each realisation, i.e. uniform distribution assumed)

Model configurations

- $32 \times 8 \times 2 \times 8 \times 1000 = 4096000$ possible configurations from input factor combinations $\Rightarrow$ computationally not feasible
- 12 000 samples were randomly drawn

Model set-up for the case study (independent of specific configuration)

- Delineation of model units (subbasins, LUs, TCs, SVCs) using the lumpR software *(Pilz et al., 2017)*
- Derivation of soil and vegetation parameters from databases and pedotransfer functions
- Preprocessing of meteorological data (gap filling, spatial interpolation) with ECHSE tools *(Kneis, 2012)*



Model simulation and performance evaluation

Back to overview

Simulation settings

- Three years from 1 January 2013 to 31 December 2015
- Daily resolution
- Up to 20 iterations of warm-up years to bring model states into equilibrium

Performance evaluation

- Case-specific choice of performance metric: root mean square error (RMSE)
- For dynamic analysis computed over moving window ($w = 15$ resulting 31 days) for each simulation day d: $RMSE(d) = \sqrt{\frac{1}{2w+1} \sum_{i=d-w}^{d+w} \left(q_s(i) - q_o(i)\right)^2}$ with q_s simulated and q_o observed discharge

Determination of posterior distributions

Bayesian approach: $\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$

- Prior is defined by the realisations (and their weights) of each input factor
- Likelihood is derived by taking additional data into account $\Rightarrow$ many approaches exist

In this study employs an informal approach

- Oriented at the GLUE approach *(Beven and Binley, 1992)*
- Separate model configurations into behavioural and non-behavioural groups based on their performances
- Here: 10 % best performing configurations considered as behavioural

$\Rightarrow$ 1200 posterior model configurations

$\Rightarrow$ Remaining realisations and their frequencies define the posterior distribution of each input factor

Identifiability measure

Back to overview

For each input factor

- n_{prior} is the number of realisations in the prior distribution
- n_{post} is the remaining number of realisations in the posterior distribution
- $IM = 1 - \frac{n_{post} - 1}{n_{prior} - 1}$ with $n_{prior}, n_{post} \in \mathbb{N}$

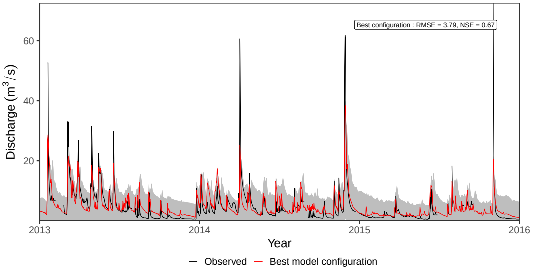
That means

- $IM = 0$ is obtained when $n_{post} = n_{prior}$
  - $\rightarrow$ all realisations of an input factor defined in its prior distribution are still present in the posterior distribution
  - $\rightarrow$ the input factor is not identifiable
- $IM = 1$ is obtained when $n_{post} = 1$
  - $\rightarrow$ only one realisation left in the posterior distribution
  - $\rightarrow$ the input factor is well identifiable

Results: discharge simulations

Next: static IM

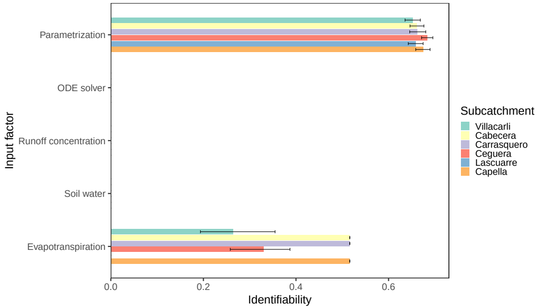
- Most observation values fall into 90 % probability range of model configurations
- Large peaks often underestimated
  - → Partly attributable to poorly detected heavy precipitation events
- Falling limbs of discharge events sometimes not well matched
  - → Measurement uncertainty?
  - → Missing calibration?
  - → Optimal model structure not yet included?



Gray area shows the 90 % probability range of all (prior, uncalibrated) model configurations.

Results: static identifiability measure

- Only Evapotranspiration and Parametrisation exhibit some degree of identifiability
- Consistently zero identifiability for ODE solver, Runoff concentration, and Soil water
  - → All implemented realisation can lead to acceptable model performance
- Relatively consistent results for different subcatchments (except Lascuarre)



Black errorbars represent the 95 % confidence interval estimated via bootstrapping.

Results: posterior distributions of input factors

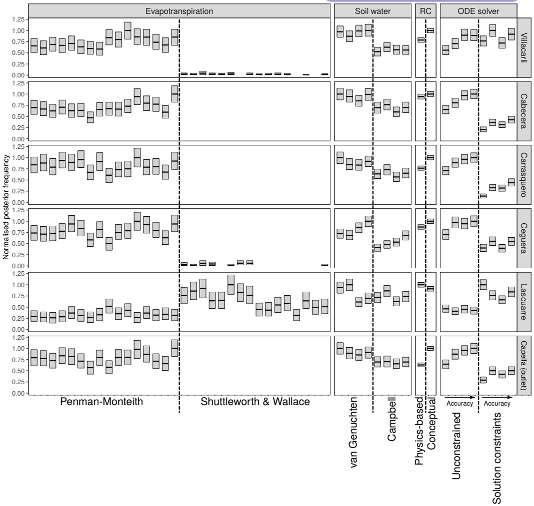
Next: posterior distributions II

Evapotranspiration

- Penman-Monteith clearly superior to Shuttleworth & Wallace except for Lascuarre
- For subprocesses no obvious pattern

Soil water

- Retention model of highest importance: van Genuchten (mostly) slightly superior to Campbell
- Realisations of other processes (infiltration and percolation approaches) equally plausible



Boxes represent the 95 % confidence interval and mean (black horizontal line) estimated via bootstrapping.

Results: posterior distributions of input factors

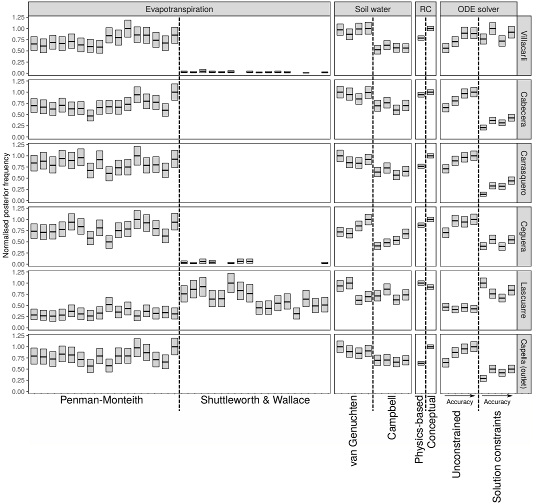
Next: dynamic posterior distributions

Runoff concentration (RC)

- Conceptual approach with delay factor (calibration parameter) superior (except Lascuarre)

ODE solver

- Unconstrained solvers mostly superior to solvers with solution constraints (physical limits)
  - → Model performance compensates for unrealistic model states?
- Solvers with higher accuracy (higher order) mostly achieve better performances
  - → Sometimes even simple Euler approach achieves high importance (Lascuarre)



Boxes represent the 95 % confidence interval and mean (black horizontal line) estimated via bootstrapping.

Results: dynamic posterior distributions

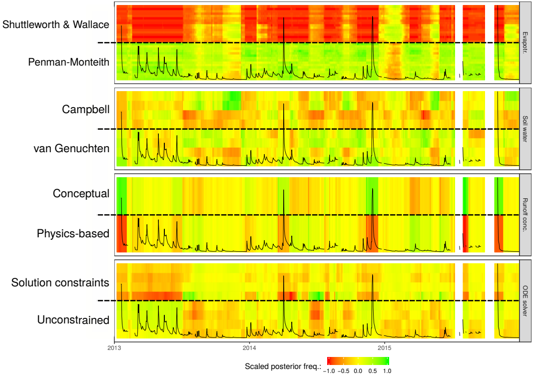
Next: dynamic posterior distributions II

Evapotranspiration

- Most of the time Penman-Monteith superior
- During dry periods Shuttleworth & Wallace gains importance

Soil water

- Highly diverse patterns → posterior distribution changes with flow / wetness conditions



Red: low posterior frequency, less plausible representations. Green: high posterior frequency, more plausible representations. Black lines: discharge hydrograph (gauge Capella = catchment outlet).

Results: dynamic posterior distributions

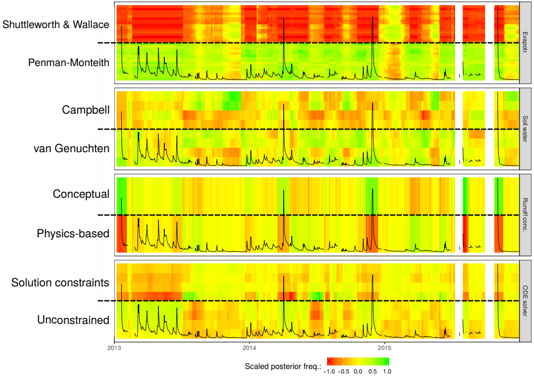
Back to overview

Runoff concentration

- During peak flows: conceptual approach favoured
- Most of the time no clear identifiability

ODE solver

- Highly diverse patterns $\rightarrow$ posterior distribution changes with flow / wetness conditions
- Unconstrained solvers slightly favoured during high flows / wet conditions



Red: low posterior frequency, less plausible representations. Green: high posterior frequency, more plausible representations. Black lines: discharge hydrograph (gauge Capella = catchment outlet).

Conclusions

Back to overview

The proposed framework

- consists of coupling a flexible model environment with dynamic identifiability analysis
- can be used to identify most plausible model configuration(s)
- is generic and leaves many options to the user in terms of software, definition and implementation of input factors, model evaluation etc.
- can provide valuable information about process behaviour in a catchment
  - → Which process representations / underlying theory explains observed dynamics best?

The case study shows that

- parametrisation and evapotranspiration are the best identifiable input factors
- model structure identifiability varies over time
- identifiability is influenced by wetness conditions and landscape characteristics
- there are unexpected results possibly due to complex interactions and compensations effects between ODE solver, process representation and parametrisation
  - → Unconstrained ODE solvers lead to unrealistic model states but better model performance

References I

Back to overview

Beven, K. and A. Binley (1992). "The future of distributed models: Model calibration and uncertainty prediction". In: *Hydrological Processes* 6.3, pp. 279–298. DOI: https://doi.org/10.1002/hyp.3360060305.

Bronstert, A., J.-C. de Araújo, R. J. Batalla, A. C. Costa, J. M. Delgado, T. Francke, S. Foerster, A. Guentner, J. A. López-Tarazón, G. L. Mamede, P. H. Medeiros, E. Mueller, and D. Vericat (2014). "Process-based modelling of erosion, sediment transport and reservoir siltation in mesoscale semi-arid catchments". In: *Journal of Soils and Sediments* 14.12, pp. 2001–2018. DOI: https://doi.org/10.1007/s11368-014-0994-1.

Francke, T., S. Foerster, A. Brosinsky, E. Sommerer, J. A. Lopez-Tarazon, A. Güntner, R. J. Batalla, and A. Bronstert (2018a). "Water and sediment fluxes in Mediterranean mountainous regions: comprehensive dataset for hydro-sedimentological analyses and modelling in a mesoscale catchment (River Isábena, NE Spain)". In: *Earth System Science Data* 10.2, pp. 1063–1075. DOI: https://doi.org/10.5194/essd-10-1063-2018.

Francke, T., G. Baroni, A. Brosinsky, S. Foerster, J. A. López-Tarazón, E. Sommerer, and A. Bronstert (2018b). "What did really improve our meso-scale hydrological model? A multi-dimensional analysis based on real observations". In: *Water Resources Research* 54.11, pp. 8594–8612. DOI: https://doi.org/10.1029/2018WR022813.

García-Ruiz, J. M., S. Beguería, J. I. López-Moreno, A. Lorente, and M. Seeger (2001). *Los recursos hídricos superficiales del Pirineo aragonés y su evolución reciente*. Logroño, Spain: Geoforma.

Güntner, A. (2002). *Large-scale hydrological modelling in the semi-arid North-East of Brazil*. PIK Report 77. Potsdam, Germany: Potsdam Institute for Climate Impact Research.

Güntner, A. and A. Bronstert (2004). "Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas". In: *Journal of Hydrology* 297.1–4, pp. 136–161. DOI: https://doi.org/10.1016/j.jhydrol.2004.04.008.

Kneis, D. (2012). *geostat: Utilities for spatial interpolation*. R package version 0.1. URL: https://github.com/echse/echse_tools/tree/master/R/packages/geostat.

References II

Kneis, D. (2015). “A lightweight framework for rapid development of object-based hydrological model engines”. In: *Environmental Modelling & Software* 68, pp. 110–121. DOI: https://doi.org/10.1016/j.envsoft.2015.02.009.

Mueller, E. N., T. Francke, R. J. Batalla, and A. Bronstert (2009). “Modelling the effects of land-use change on runoff and sediment yield for a meso-scale catchment in the Southern Pyrenees”. In: *CATENA* 79.3, pp. 288–296. DOI: https://doi.org/10.1016/j.catena.2009.06.007.

Mueller, E. N., A. Güntner, T. Francke, and G. Mamede (2010). “Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model”. In: *Geoscientific Model Development* 3.1, pp. 275–291. DOI: https://doi.org/10.5194/gmd-3-275-2010.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener (2016). “Sensitivity analysis of environmental models: A systematic review with practical workflow”. In: *Environmental Modelling & Software* 79, pp. 214–232. DOI: https://doi.org/10.1016/j.envsoft.2016.02.008.

Pilz, T., T. Francke, and A. Bronstert (2017). “lumpR 2.0.0: an R package facilitating landscape discretisation for hillslope-based hydrological models”. In: *Geoscientific Model Development* 10.8, pp. 3001–3023. DOI: https://doi.org/10.5194/gmd-10-3001-2017.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003). “Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis”. In: *Hydrological Processes* 17.2, pp. 455–476. DOI: https://doi.org/10.1002/hyp.1135.