1. Using flexible models within BATEA

This exercise session illustrates some of the insights into hydrological model calibration that can be obtained using Bayesian inference and a flexible framework for hypothesizing and testing hydrological models.

1.1. The SUPERFLEX and FUSE modelling frameworks

SUPERFLEX is a multi-component modelling toolkit that allows building hydrological models from generic components (building blocks) that can be customized and configured within a flow network configuration. The generic components are reservoirs, lag functions and connection elements (Figure 1). The flux functions describing storage-discharge relationships, shape of lag functions, etc, are selected from a library of functions and can be used to hypothesize and build multiple alternative model structures.

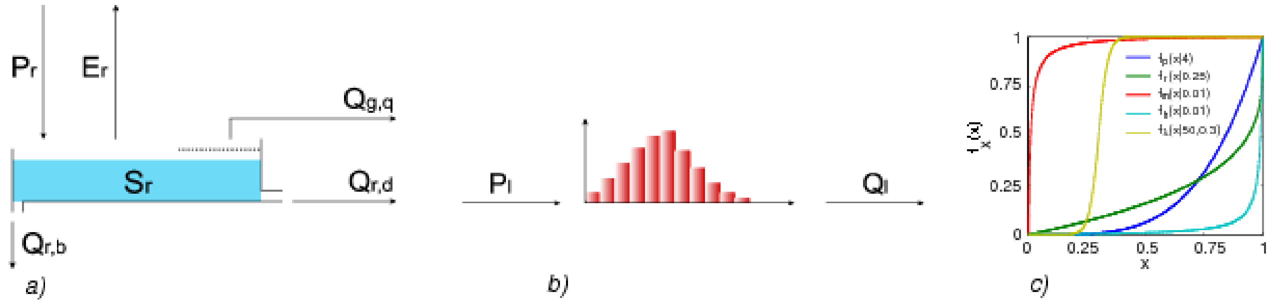


Figure 1. Generic building blocks of the flexible framework SUPERFLEX: (a) Generic reservoir, (b) Lag function. Representative shapes of constitutive functions are shown in panel (c).

SUPERFLEX is implemented as a standalone Fortran-95 code, and is also integrated into the built-in model library in *Petit Bateau*. Hence, installing *Petit Bateau* gives you immediate access to SUPERFLEX.

Another flexible model framework already connected to *Batea* is the Framework for Understanding Structural Errors (FUSE) [Clark et al., 2008; Clark and Kavetski, WRR2010]. FUSE allows creating alternative model hypotheses, using building blocks from models such as VIC, TOPMODEL, Sacramento, and others.

In both FUSE and SUPERFLEX, the model structures are formulated as systems of differential equations and solved using robust numerical methods (eg, Clark and Kavetski, WRR2010; Kavetski et al, WRR2011).

1.2. Configuring SUPERFLEX

SUPERFLEX is configured using an input file (typically named `flexConfigXX.dat`), which specifies model components, their connectivities and parameters. No source code needs to be recompiled.

Open the configuration file `userguide_tutorials\superflexDemos\common\flexConfig10.dat` to get an idea of the layout. For the purposes of this course, the authors already identified a few configurations of interest, but you may still wish to briefly examine the file layout for configuring a SUPERFLEX model.

A number of options are available to control different aspects of hydrological model-building:

1) The types of constitutive relationships for each SUPERFLEX hydrological unit.

2) The list of parameters describing the model configuration, including active and fixed parameters

3) Numerical solver options for the numerical solution of the equations.

Note the key distinction between the specification of the structure of the model ("governing equations") and the selection of mathematical/numerical techniques for their solution. Further discussion of the theoretical and practical significance of this distinction can be found in Clark and Kavetski (WRR, 2010).

1.3. *Summary of hydrological case studies*

In previous research, the SUPERFLEX model has been applied in several case studies based on Luxembourgish and New Zealand catchments.

Two case studies are considered in this exercise session.

1: Weierbach catchment, hourly data. This case study illustrates effect of excluding and including the lag function. The effect is not huge but nonetheless quite significant. It demonstrates that the lags in the system tend to become increasingly important when modelling at short time scales.

For this case study, the data `Weierbach_1h.dat` is used, in combination with `WeierbachH.INF_DAT` for selecting the time periods. The SUPERFLEX configuration files are `flexConfig10.dat` (no-lag) and `flexConfig11.dat` (lag).

2: Weierbach catchment (schist) and Huewelerbach catchment (sandstone/marls); daily data. This case study illustrates the effect of including an additional component (a groundwater reservoir) into the model structure. The effect is minor for the Weierbach, but major for the Huewelerbach. It suggests that different dominant processes may require different model representations, thus favouring a flexible framework.

For this case study, the data `Weierbach_1d.dat` and `Huewelerbach_1d.dat` are used, in combination with `WeierbachD.INF_DAT` and `HuewelerbachD.INF_DAT` for selecting the time periods. The SUPERFLEX configuration files are `flexConfig10.dat` (no groundwater) and `flexConfig20.dat` (with groundwater).

[For reference/checking only], the optimal parameter sets are given in files `OptParConf10WeierbachD.dat` and `OptParConf20WeierbachH.dat` for the Weierbach, and `OptParConf10HuewelerbachD.dat` and `OptParConf20HuewelerbachD.dat` for the Huewelerbach.

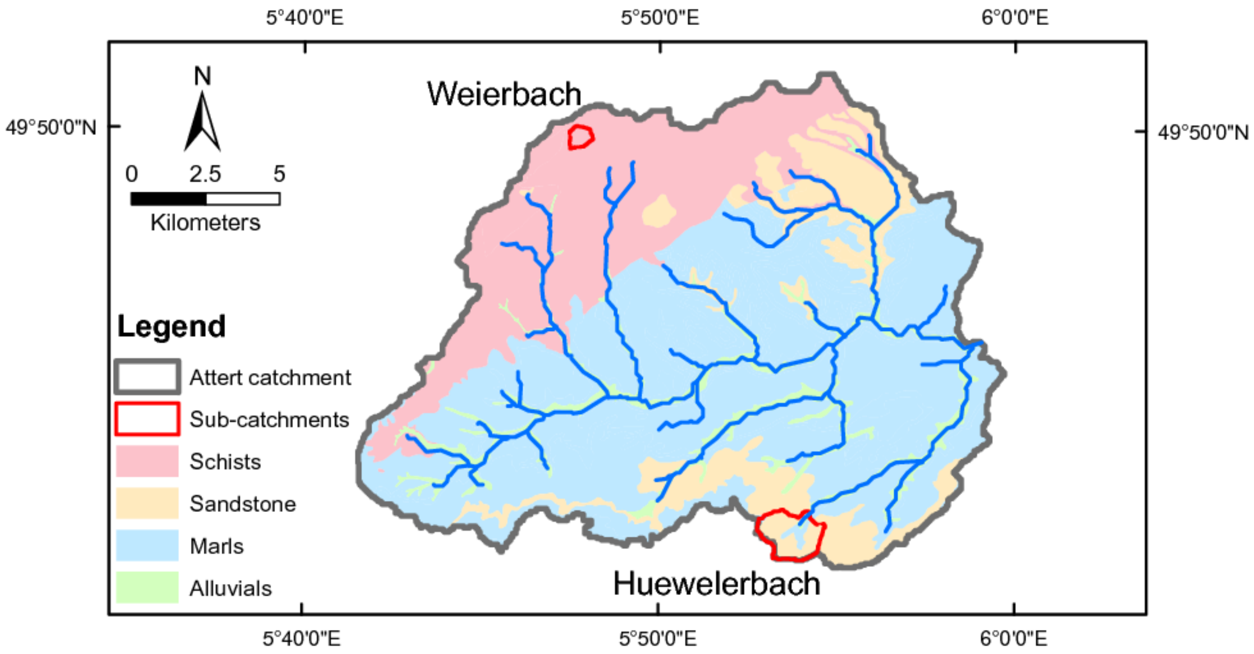


Figure 2: Location of the Huewelerbach and Weierbach case study catchments in Luxembourg

1.4. Specific SUPERFLEX configurations

1.4.1. General description

It does not take long to ascertain that attempting to infer the parameters of very complex model structures is a hopelessly non-identifiable problem [e.g., see the classic work of Jakeman and Hornberger, WRR1990]. Instead, SUPERFLEX is intended to be used to hypothesize, build and test much more parsimonious configurations. In some cases, independent insights into the catchment dynamics can guide the selection of the model structure, whereas in other cases empirical trial-and-error approaches may be necessary. The best ways to accomplish this are the subject of ongoing research.

For the purposes of this tutorial, we created 3 distinct model configurations M1, M2 and M3, specified via 3 `FlexConfig` files. Schematic representations of these models are shown in Figures 3, 4 and 5 below.

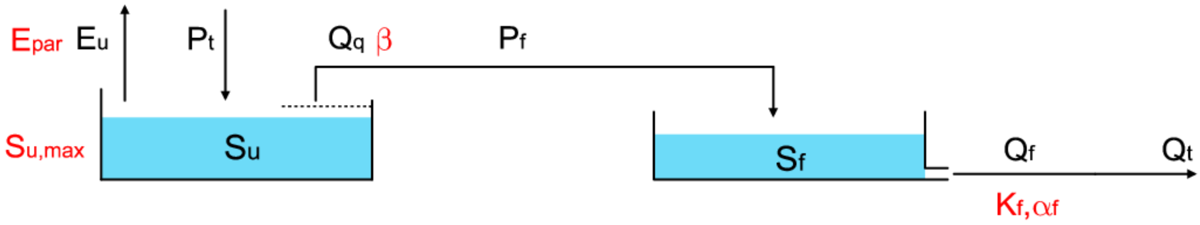


Figure 3. Schematic of model M1 (FlexConfig10.dat): 2-bucket model without a lag function.

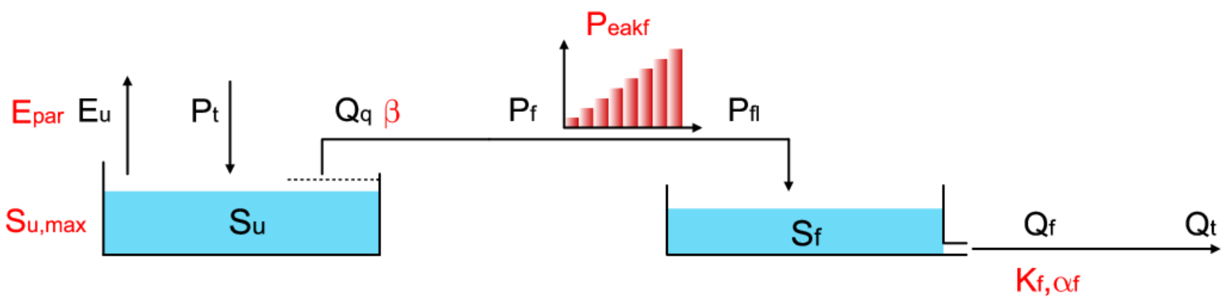


Figure 4. Schematic of model M2 (FlexConfig11.dat): 2-bucket model with a lag function.

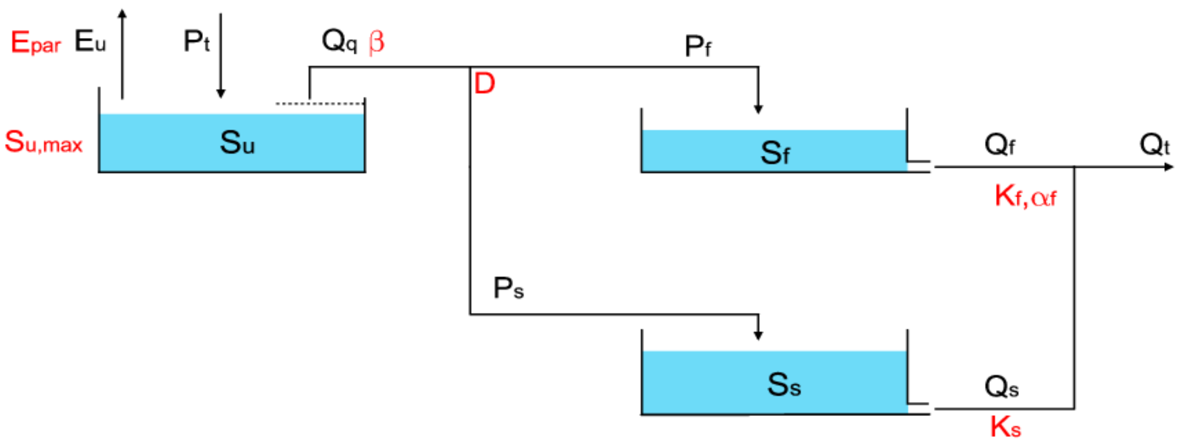


Figure 5. Schematic of model M3 (FlexConfig20.dat): 3-bucket model with a groundwater store.

1.4.2. Water balance equations and constitutive functions

The water balance equations and constitutive functions are described in Tables 1 and 2 below.

Table 1. Properties of model structures M1-M3. Nθ is the number of parameters and Ns is the number of states. UR, FR, and SR denote the unsaturated, fast, slow reservoirs. LF is the lag function.

Model			Components				Parameters							
	N_θ	N_s	UR	FR	SR	LF	E_{par} (-)	$S_{u,max}$ (mm)	β (-)	T_f (h)	K_f ($mm^{1-\alpha}$/h)	α (-)	D (-)	K_s (1/h)
M1	5	2	✓	✓	-	-	✓	✓	✓	-	✓	✓	-	-
M2	6	3	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	-	-
M3	8	3	✓	✓	✓	-	✓	✓	✓	-	✓	✓	✓	✓

Table 2. Water balance equations and constitutive functions of model structures M1-M3 (✓ and "-" indicate presence or absence respectively). The operator * denotes the convolution.

Water balance equations:	M1	M2	M3		Constitutive relationships	M1	M2	M3
$\mathrm{d}S_u/\mathrm{d}t = P_t - Q_q - E_u$	✓	✓	✓		$Q_q = P_t f_p(\bar{S}_u \mid \beta)$	✓	✓	✓
$\mathrm{d}S_f/\mathrm{d}t = P_f - Q_f$	✓	-	✓		$E_u = C_e E_p f_m(\bar{S}_u \mid E_{par})$	✓	✓	✓
$\mathrm{d}S_f/\mathrm{d}t = P_{fl} - Q_f$	-	✓	-		$P_{fl} = (P_f * h_f)(t)$	-	✓	-
$\mathrm{d}S_s/\mathrm{d}t = P_s - Q_s$	-	-	✓		$h_f = \begin{cases} t/T_f^2, & t < T_f \\ 0, & t > T_f \end{cases}$	-	✓	-
$Q_q = P_f + P_s$	-	-	✓		$P_s = DQ_q$	-	-	✓
$Q_t = Q_f$	✓	✓	-		$Q_f = k_f S_f^{\alpha}$	✓	✓	✓
$Q_t = Q_f + Q_s$	-	-	✓		$Q_s = k_s S_s$	-	-	✓

Table 3. Details of functions referred to in Table 2.

Function	Name
$\bar{S}_u = S_u / S_{u,max}$	Scaled storage (relative to a maximum capacity)
$f_p(x \mid m) = x^m$	Power function

$f_m(x\,\vert\,m) = (1+m)\dfrac{x}{x+m}$	Monod-type kinetics, adjusted so that $f_m(1\vert m)=1$

1.5. *Case study A: Effect of adding model components: lag function*

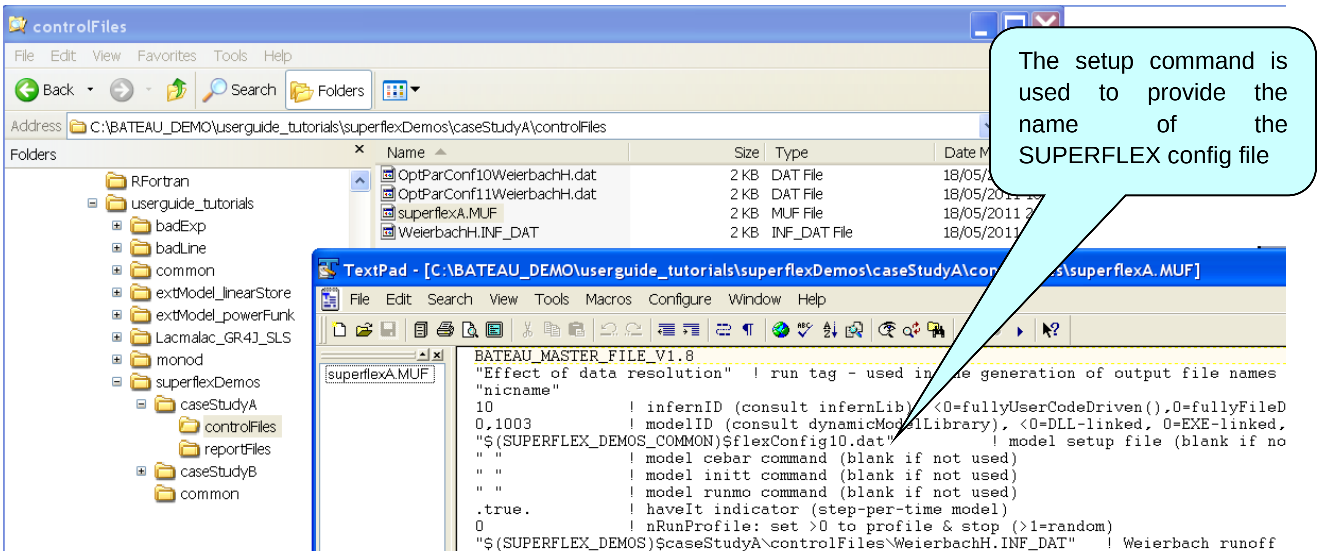
1.5.1. Weierbach catchment

Modify the file `exe\bateau_muf_director.txt` to point to `superflexA.MUF` by moving the corresponding pointer to the 2nd line position. Save and close the muf_director file.

Open `userguide_tutorials\superflexDemos\caseStudyA\controlFiles\superflexA.MUF` file. Recall that the `MUF` files are used for *Petit Bateau* case study specification. The files provided in this tutorial are already linked to the SUPERFLEX model (via `modelID=0,1003`), and also specify the data and inference settings. View and close the `MUF` file.

Model configuration without lag function

Ensure the SUPERFLEX configuration file is set to `flexConfig10.dat` in `superflexA.MUF`.



Run this case study as follows.

Press the DYNX button and move the slider bar to get a feeling on how model parameters affect predictions. Try to find a good fit by manual calibration. Check the NashSut (Nash and Sutcliffe) value (closer to 1 = better fit), and the other statistics.

Optimize the model (choose Optimize > Quasi-Newton – to save time in this simple example, just use 2 restarts). Save the optimum using ManualC > makeCurrent[arg0] and press CTRL-A to rescale the axes for better presentation. Check the goodness of fit (using the Nash-Sutcliffe statistic displayed in the interface next to the label NashSut).

Write the calibrated prediction to a file (File > Write_Pan1).

Inspect dynamics of the various model reservoirs (e.g. click Pan2, time, OK, Sf, OK)

Grid the objective functions of some parameters. For this, click Analyse > GridObjFunk in the context menu, then choose one or two parameters (e.g. alfaF and kF), click OK and select a metric (e.g. NashSut) to be analysed. In the next dialog, set the grid resolution (e.g., set nX and nY to 51), and click OK in the next dialog. You should now see the contour plot of the metric you selected (e.g., the Nash-Sutcliffe index). To rescale the z axis, click on GridConts > zMinCont and specify a lower bound (e.g. 0 for the Nash-Sutcliffe index). If necessary, you can export the grid data to a file (e.g. File > WriteFreqsTriplet).

Close *Petit Bateau*.

Model configuration with lag function

Switch to a SUPERFLEX model configuration without the lag function by specifying the configuration file `flexConfig11.dat` in `superflexA.MUF`.

Run this case study. Optimize the model (2 restarts) and inspect goodness of fit. Observe the moderate improvement in model performance when \ the lag function component is added.

Write the calibrated prediction to a file (File > Write_Pan1)

Close *Petit Bateau*.

In excel, open the 2 files containing the model predictions, and compare the hydrographs of the 2 models.

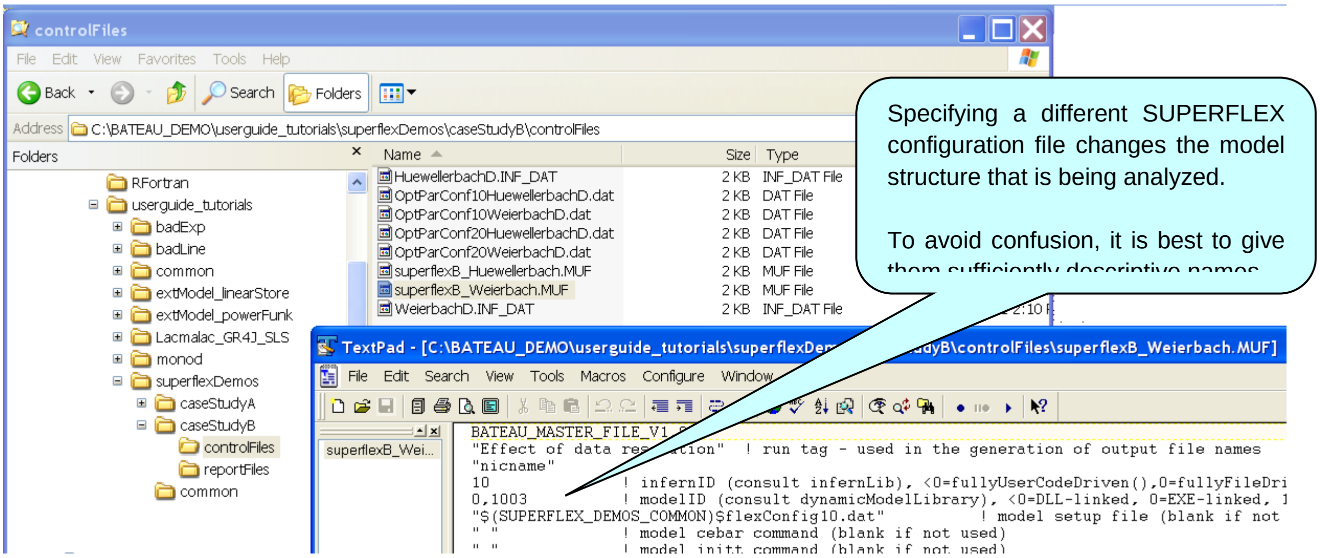
Case study B: Cross-catchment differences: grundwasser store

1.5.2. Weierbach catchment

Open and modify the file `exe\bateau_muf_director.txt` to point to `superflexB_Weierbuch.MUF` (by moving its pointer to the 2nd line of the file). Save and close the `muf_director`.

Model configuration without groundwater store

Open the file `superflexDemos\caseStudyB\controlFiles\superflexB_Weierbach.MUF`. Ensure the SUPERFLEX configuration file is `flexConfig10.dat` (on line 6, immediately below the `modelID=0,1003` statement). This corresponds to a model without a separate groundwater store.



Run this case study. Optimize the model (for speed, use 0 (zero) quasi-Newton restarts) and inspect the goodness of fit. Save the hydrograph to a file.

Model configuration with groundwater store

Open `superflexB_Weierbach.MUF` and specify the SUPERFLEX configuration file `flexConfig20.dat`. This adds a groundwater store to the model architecture.

Run this case study. Optimize the model (for speed, use 2 quasi-Newton restarts) and inspect the goodness of fit. Save the hydrograph to a file.

Observe that adding the grundwater store does not noticeably improve the goodness of fit.

In excel, compare the hydrographs for these 2 model configurations.

Inspect dynamics of the various model reservoirs (e.g. click Pan2, time, OK, Ss, OK)

Huewelerbach catchment

Open `exe\bateau_muf_director.txt` and point it to the case study file `superflexB_Huewelerbach.MUF`. Save and close the `muf_director`.

Model configuration without groundwater store

Open the file `superflexDemos\caseStudyB\controlFiles\superflexB_Huewelerbach.MUF`. Set the SUPERFLEX configuration file to `flexConfig10.dat`. Recall from the earlier section that this specifies a configuration without a separate groundwater store.

Run this case study. Optimize the model (for speed, use 2 quasi-Newton restarts) and inspect the goodness of fit. Save optimal model predictions to a file

Model configuration with groundwater store

Specify the SUPERFLEX configuration file `flexConfig20.dat` in `superflexB_Huewelerbach.MUF`. Recall from Section 1.5.2 that this adds a groundwater store to the model.

Run this case study. Optimize the model (for speed, use 2 quasi-Newton restarts) and inspect the goodness of fit. Save optimal model predictions to a file.

Unlike the Weierbach catchment, the goodness of fit in the Huewelerbach basin improved notably by adding an extra store. The model remains imperfect (eg, plot the model residuals vs response).

1.5.3. Hydrological insights and Bayesian analysis in Case Study A

This case study suggests the presence of lags in the Weierbach catchment system. While the existence of delays between the rainfall forcing and discharge response in most catchments is well known, the physics underlying it in specific catchments remains poorly understood.

Generally, lags occur due to the transfer of water across different catchment compartments (including low-permeable regions), and routing in the river channel system. However, in most cases it is difficult to quantitatively attribute the delays to different processes and flow paths. For this reason, lumped conceptual representations, such as the convolution-based lag function used in SUPERFLEX and many other models, are often more tractable than more complex models when estimated by calibration to observed data.

Even at a basic level, the Bayesian analysis is used here to formulate the objective function and, as such, provides a self-consistent framework for parameter calibration, estimation of uncertainties in the parameters, quantification of model error, generation of predictions and associated uncertainty estimates, etc.

Note that while the simplistic additive formulation is used in this example, its deficiencies can be explicitly diagnosed, and, in several important aspects already be overcome (see Section XXX on WLS).

1.5.4. Hydrological insights and Bayesian analysis in Case Study B

While the groundwater reservoir did not notably improve the model performance in the Weierbach catchment, it appeared to be quite critical for reproducing the streamflow dynamics of the Huewelerbach catchment. These findings are confirmed by experimental knowledge of these catchments: it is generally known that schist formations are relatively impermeable, while sandstone formations can hold considerable water storage and hence sustain streamflow during dry weather periods.

Being able to confirm such field insights through mathematical modeling is reassuring and substantiates the initial perceptions of how the catchment works. In addition, model comparison may tell us something about the relative importance of individual processes and compartments, and how best to represent them.

The Bayesian formulation is exploited to construct probabilistic prediction limits, which give us a better quantitative feel for the predictive uncertainty, but also facilitate a more meaningful evaluation of the effects of model modification. For any inference to be credible and verifiable, it is important to explicitly state the error assumptions used to derive the posterior distribution (objective function) and scrutinize these assumptions a posteriori (eg, see Thyer et al, WRR2009, for an illustration). Much work remains to be done to advance these aspects of hydrological modeling, especially in conditions of scarce data availability.

Finally, note the computational efficiency permitted by a combination of (i) numerically robust and smooth model implementation, (ii) modern Newton-type optimization, (iii) a computationally efficient programming language (here, Fortran-95), and (iv) a modern computer. Whereas in the “old days” of the 20th century optimization alone could take days, nowadays we can begin inferring rainfall-runoff models in a matter of minutes on a laptop!