Departament d'Ecologia
Universitat de Barcelona

Ter River influence on Sau Reservoir limnology
Empirical and watershed-scale modeling

La influència del riu Ter en la limnologia de l'embassament de Sau
Modelització empírica i a escala de conca

Rafael Marcé Romero

TESI DOCTORAL

Departament d'Ecologia
Universitat de Barcelona
Programa de doctorat: Ecologia Fonamental i Aplicada
Bienni 2000–2002

Ter River influence on Sau Reservoir limnology
Empirical and watershed-scale modeling

Memòria presentada per Rafael Marcé Romero per optar al grau de Doctor per la Universitat de Barcelona

Rafael Marcé Romero
Barcelona, a 26 de març de 2007

Vist i plau del director de la Tesi

Dr. Joan Armengol Bachero
Catedràtic del Departament d'Ecologia
Universitat de Barcelona

Part II

WATERSHED SCALE MODELING

Chapter 4

Using spatially distributed parameters and multi-response objective functions to solve parameterization of complex applications of semidistributed hydrological models

4.1 Introduction

Semidistributed models are the usual choice in assessment and prediction of water quality in rivers when the origin of the water quality constituents (e.g. point versus diffuse sources) is a major target (Singh and Woolhiser 2002). Although good empirical alternatives exist if the working time scale is large enough (e.g. Alexander et al. [2002] for annual time steps), dynamic models should be considered at shorter time scales. In this modeling framework, the quality and reliability of a water quality constituent calculation lies on a correct assessment or prediction of the quantity of water flowing through the river. Thus, the hydrological performance of a watershed model is determinant to the success of its water quality modules.

But to correctly assess the origin of the water quality constituents (and consequently of the water arriving at the river reach) not only the water flowing through the river channel should be modeled: a good estimation of the surface versus subsurface contributions to the runoff is needed. Unfortunately, in most water quality applications at the watershed scale surface flows in the channel of the rivers are the only hydrological data available to help the whole parameterization of the hydrological modules. Even if more data exists, problems related to incompleteness and inaccuracy usually restricts the amount of useful data (Singh and Woolhiser 2002). Despite of this, few works check the uncertainty associated to the different components of the water inflow to the channel. This could be problematic for water quality applications, especially if outcomes should support management decisions, since

the same flow record in the river channel can be obtained from very different combinations of subsurface inputs and surface runoff, which can have dramatic differences in water quality characteristics (Butturini and Sabater 2000).

This common lack of high quality field measurements of water flows across the watershed, summed to the high number of adjustable parameters usually present in semidistributed models, ends with hydrological simulations that reasonably fit observed river runoff, but with a high degree of uncertainty in parameter values, and consequently in the separation of the modeled response in surface and sub-surface inputs to the river. In complex modeling applications this situation is even worse: if the model-driving characteristics of the watershed (meteorology, land uses, soil properties, geomorphology) have high variability across the basin or along the time, the number of parameters needed to account for this variability should be even higher. In such a situation, the uncertainty in adjustable parameter values and routing calculations will be severely enhanced, unless good field data and prior information on parameter values are available.

During the last years, a considerable effort has been devoted to combine automatic calibration methods with proper data management during the calibration process, in an attempt to minimize the uncertainty in the modeled hydrological response. One approach is the use of multiple objective functions during the calibration processes, in order to constrain the parameter uncertainty forcing the system to fit different data measured in the field (Hunt et al. 2006), or different features of a data series (Madsen 2000). This can be solved combining the different objective functions in a single one (Madsen 2003), searching for the entire Pareto set of parameters that fits the data (Gupta et al. 1998; Vrugt et al. 2003), or a region of it (Khu and Madsen 2005). On the other hand, the use of proper regularization and prior information methodologies allow the inclusion of all the information available in our observed data and expertise into the calibration process (Doherty and Johnston 2003). This last methodology is very useful when a high degree of spatial or temporal heterogeneity in model parameters is present, because it allows the inclusion of many parameters in the calibration processes without converting the problem into an ill, underdetermined one.

In this study we show how a complex watershed scale hydrological model application can be successfully solved combining a regularized fitting methodology with a multiobjective function (OF) approach. The usefulness of the spatial discretization of adjustable parameters was tested with a standard calculation of parameter uncertainty, whereas the effect of the inclusion of different data in the calibration process on the output was tested with a stepwise procedure, in which different kind of data were sequentially added to the OF. It is a typical hydrological simulation for water quality purposes, in the sense that only limited river water flow information is available to parameterize the flow routing equations. Also, as the final aim of this model will be the simulation of water quality constituents, we pay special attention to the effect of the inclusion of different data types into the OF on the uncertainty of the modeled subsurface input to the river.

4.2 The study site and the model

The Ter River watershed is a 1380 km^2 populated area, mainly covered by woodland (78%) and agricultural land (16%). The headwaters are located in the north end of the basin (Fig. 4.1), at 2500 m altitude in the Spanish Pyrenees ranges, and for this study the watershed ends at the Sau Reservoir dam, at 360 m altitude, configuring a steeped basin with a mean slope of 0.27. However, the terrain slope, as well as other features in the watershed, shows a high spatial heterogeneity across the basin (Fig. 4.1). The headwaters flow over hard materials (igneous and metamorphic rocks) covered by a mixture of high altitude shrublands and conifer and deciduous forests over a steeped terrain (mean slope 0.38). Downstream, around the meeting point of the two main headwater courses, the forested land accounts for nearly all the terrain, the slopes have moderated (mean slope 0.28), and the river

flows over sedimentary rocks. Then the river enters in the populated agricultural plain (most of the slopes under 0.03) where the main human settlements are located. Here the alluvial deposits are abundant, and unirrigated crops cover most part of the land. Finally, the Southeast end of the basin is a mountainous terrain similar to that located in the north top section. Thus, the watershed includes a complex mixture of geomorphologic, lithologic, and land cover features. In addition, empirical evidence from twelve meteorological stations located inside the watershed (Fig. 4.1) shows that rain and other meteorological variables are also highly spatially variable, especially in the North-South direction, and at any time scale. All in all, the watershed under study constitutes a demanding problem for any hydrological model.

The Hydrological Simulation Program-Fortran (HSPF) is a lumped hydrological model well suited to deal with complex watersheds: the possibility to split the watershed into subbasins converts the model into a semidistributed application, allowing the inclusion of spatial heterogeneity of watershed features and meteorological inputs in the modeling framework. Also, the hydrological modules solve for the surface and subsurface inputs to the river flow (Bicknell et al. 2001). Nevertheless, HSPF is a robust, relatively user-friendly application that includes many useful tools and sources of help for modelers, including databases with parameter values from many previous applications (USEPA 1999). In addition, the USEPA supported BASINS expert system offers a number of preprocessing tools compatible with HSPF (USEPA 2001).

The basic modeling unit in HSPF is the Hydrological Response Unit (HRU), a piece of terrain that is intended to have homogeneous watershed and meteorological characteristics, and thus a unique hydrological response. For this study, the HRUs were defined overlaying three watershed features (terrain slope, lithology, and land use) and the meteorological input. The terrain slope was divided into three categories (S1: 0–0.06; S2: 0.06–0.15; S3: $>$0.15) following other HSPF applications (SWM 2002). The lithological layer was defined as a three items map, and the land uses layer included eight categories (Fig. 4.1). Twelve meteorological stations supplying hourly data were available, and the meteorological variables were spatially distributed assigning the input of each meteorological station to one or more subbasins, defining twelve meteorological zones. We discarded the use of Voronoi polygons or similar methodologies to draw the meteorological zones because the spatial extent of the meteorological events in this watershed is clearly limited by topography. The combination of the four layers gave a theoretical maximum of 864 HRUs, but due to the spatial covariance between layers and the assimilation of HRUs of less than 500 *ha* by HRUs with similar characteristics the actual value was 109 HRUs. The output of the HRUs is diverted to the river reaches included in the subbasins where the HRUs are present. The subbasin delineation was automatically performed with the ArcView algorithm implemented in the BASINS package, using a 100 *m* grid digital elevation model, giving 58 subbasins (Fig. 4.1).

Since each of the 109 HRUs generated acts as an independent hydrological model, and HSPF has a considerable amount of parameters (Table 4.1), we are facing a very complex problem. Fortunately, some of the parameter values could be estimated from GIS based information, inventories, or bibliographical research. Thus, a considerable degree of spatial heterogeneity (i.e. specific parameter values for each HRU) could be preserved. But many parameters represent abstract properties of the basin, or simply there was no field information available, and consequently a calibration process must optimize their values. Obviously, we cannot calibrate 109 different values for the eleven adjustable parameters considered in this study (Table 4.1). Instead, we tried to preserve some heterogeneity by considering different parameter values in relation to some watershed feature (lithology, terrain slope, or land use) considered as a key factor governing the parameter behavior (Table 4.1). Information in USEPA (1978; 1999; 2000) was used to define the watershed feature relevant to each adjustable parameter.

In addition, temporal heterogeneity was considered relevant for two parameters: *UZSN* and *LZETP*. HSPF is a flexible modeling environment,

Table 4.1 – Parameters of the Hydrological Simulation Program-Fortran (HSPF) model.

Parameters Optimized	Description	Zonation	Prior Value	Range [a]
Storages				
LZSN	Lower zone nominal storage (in)	3 by lithology		0.01-15
UZSN	Upper zone nominal storage (in)	3 by slope		0.01-10
UZSNΣ	Monthly variability for UZSN	Applies only to S1	0	0-1
Infiltration				
INFILT	Index to the infiltration capacity (in hr^{-1})	3 by lithology	0.08 [b]	0.001-0.5
INFEXP	Exponent in the infiltration equation	3 by lithology		0-10
Recession				
KVARY	Non-exponential groundwater recession (in^{-1})	3 by lithology	0	0-5
AGWRC	Base groundwater recession (day^{-1})	2 by land grups		0.85-1
IRC	Interflow recession (day^{-1})	3 by slope	0.7 [c]	0.3-0.85
Routing				
INTFW	Interflow inflow parameter	3 by lithology		1-10
LZETP	Lower zone evapotranspiration parameter	5 by land use		0.1-0.9
LZETPΣ	Monthly variability for LZETP	Applies to all	0	0-1
BASETP	Baseflow evapotranspiration by deciduous forest	Applies only to DC	0	0-0.2
DEEPFR	Groundwater fraction lost by deep percolation	3 by lithology	0	0-0.5

Parameters Estimated	Description	Source	Reference
Terrain Features			
COVIND	Snow pack that covers the entire HRU (in)	Function of slope	This study [d]
SLSUR	Terrain slope	GIS	e
LSUR	Length of the overland flow plane (ft)	Function of slope	This study [f]
NSUR	Manning's *n* for the overland flow plane	Function of land use	c
Vegetal Cover			
FOREST	Fraction of land covered by vegetation	Inventories	g
INTERCEPT	Monthly interception storage capacity (in)	Inventories + Literature	h
River Reaches			
FTABLE	Tables summarizing reaches morphometry	Measured	i

[a] Modified from USEPA 2000.
[b] USEPA 1978.
[c] USEPA 2000.
[d] COVIND = 1+10 · SLSUR.
[e] LANDSAT-TM 1997-1998 30m grid by Institut Cartogràfic de Catalunya (Spain).
[f] LSUR = 100 + 6358 · exp(-0.099 · SLSUR).
[g] CREAF 2001.
[h] van der Leeden et al. 1991; Llorens 1997; Llorens et al. 1997; Bellot and Escarré 1998; Rodà et al. 1999; Gallart et al. 2002.
[i] ACA 2001.

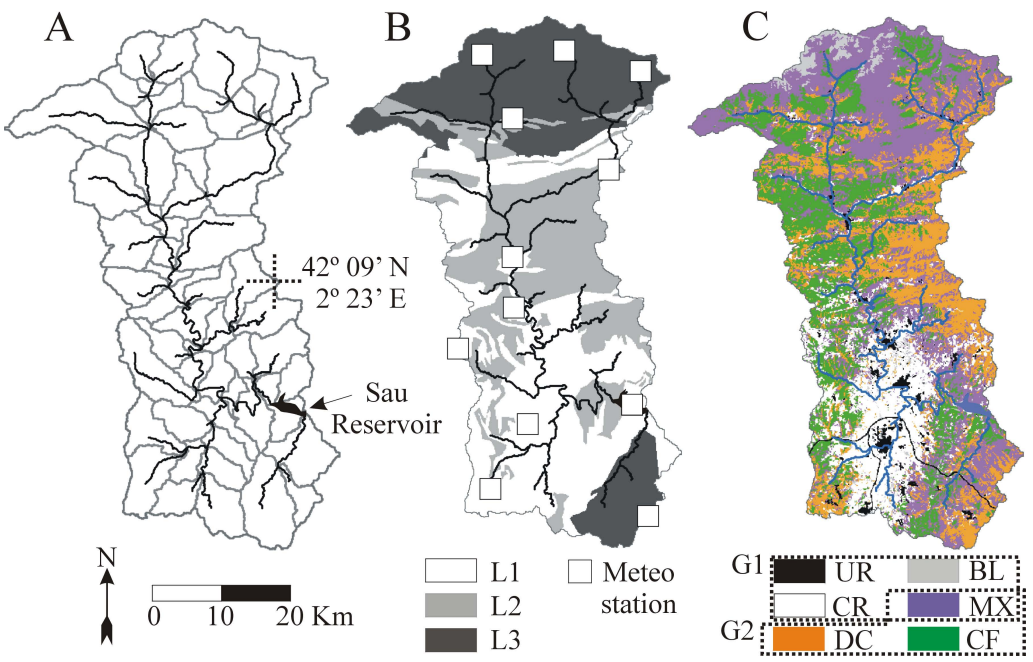

Figure 4.1 – (A) Ter watershed location and subbasins delineated for HSPF simulation. (B) Lithological zones (L1: alluvial deposits and soft sedimentary rocks; L2: consolidated sedimentary rocks; L3: igneous and metamorphic rocks) and location of meteorological stations. (C) Land uses in the catchment (UR: urban; CR: unirrigated crops; DC: deciduous forest; BL: barren land; MX: for clarity, meadows, shrublands, and few portions of oak forest are included here; CF: conifers forest). G1 groups the non-forested uses, while G2 includes the different forestlands. These are used to distribute the parameter *AGWRC* (*see* Table 4.1).

and allows the definition of monthly parameter values. Since defining twelve values for these two parameters would almost double the number of adjustable parameters, we introduced two new parameters, that we called $UZSN\Sigma$ and $LZETP\Sigma$ (Table 4.1), which represent the amplitude of variation of the corresponding quantity about its average value. The twelve monthly values were computed using:

$$UZSN_n = UZSN + (UZSN \times UZSN\Sigma) \times \sin\left((x_n - 270) \times \frac{2\pi}{365.25}\right) \quad (4.1)$$

and,

$$LZETP_n = LZETP + (LZETP \times LZETP\Sigma) \times \sin\left((x_n - 91) \times \frac{2\pi}{365.25}\right) \quad (4.2)$$

where $n = (1,2,\ldots,12)$ stands for the month index, and $x = (0,31,60,\ldots,335)$ represents the day of the year representative for each of the twelve months. Thus, the two equations define a sinus curve centered at a different day of the year, if the Σ parameters have a non-zero value. The parameter $UZSN\Sigma$ only affects the $UZSN$ value defined for HRU's included in the slope type 1 ($S1$), whereas $LZETP\Sigma$ affects all the $LZETP$ values (*see* Table 4.1).

4.3 Calibration strategy

4.3.1 Objective function

Gupta et al. (1998) proposed a new calibration paradigm based on the multi-objective function approach. Following Madsen (2003), in a multi-OF context calibration can be performed on the basis of multi-variable measurements (e.g. river runoff and groundwater levels), multi-site measurements (i.e. a variable measured in several sites within the watershed), and multi-response modes (i.e. OFs that includes several aspects of the hydrological response, but on the basis of the same measured variable). In a typical hydrological simulation for water quality purposes, rarely the multi-variable approach is accessible, because usually only river runoff is available. Thus, in these applications the multi-response approach is the most frequently applied, supplying multi-site measurements of river flow in the best situations. This study represents an extreme situation, where only the multi-response option was available. Consequently, we applied our efforts in defining a good multi-response OF. The responses included in the OF follow.

- WEEK. This response includes weekly mean flows in the Ter River during the period 1999–2002. Since the only recent flow data within this watershed comes from the daily water budgets in Sau Reservoir (Fig. 4.1), we investigate the relationship at different time scales between the water budget estimation and the river flow measures from a gauging station that operated just upstream the reservoir during the 1980s. We found that the best correlation was computed with the mean weekly data (r^2=0.89). The period from January 2003 to July 2004 was reserved as a validation data set.

- HOUR. Since the recommended HSPF working time step is the hour, and some of the driving parameters work at this time step (Bicknell et al. 2001), we considered important to include a series of hourly river flow into the OF. Again, the source was the water balance in Sau Reservoir. Since an hourly record of this kind is prompted to include a high level of measurement noise, we looked for a strong storm event with dramatic short-term changes in the river flow. Only one such event could be found in the database pertaining to the studied period (from 10 May 2002 to 14 May 2002, Figure 4.2). Thus, no validation data was available for this response.

- TRIM. We aggregated the flow response in 91 days periods. Although this response cannot be considered independent of the $WEEK$ series, the inclusion of series at coarse time resolution is common during calibration of semidistributed models (Shamir et al. 2005). The calibration and validation periods were as in $WEEK$.

- BASE. Although the knowledge of subsurface inputs to the river flow are of capital importance both in hydrological and water quality applications, direct measures of this flow are very difficult to obtain. Since one of the aims of this study was to test the effect of different kind of data on the uncertainty in the modeled subsurface inputs to the river, we considered the inclusion of some estimation of the baseflow in the calibration process. Thus, we used the Arnold and Allen (1999) numerical filter to estimate the baseflow hydrograph from the daily flow record in the Ter River. As a numerical estimate, this method carries some uncertainty, expressed in the fact that the result is not a single baseflow series, but a collection of three passes of the numerical filter. Although the most probable result is the first pass (Arnold and Allen 1999), we used all the available information defining a specific residual for the *BASE* data. Whereas for the other responses the residuals r between observed and simulated data were just the difference series, for *BASE* we used:

$$\begin{aligned} &\text{if } t\_pass_i < mod_i < f\_pass_i,\ r_i = 0 \\ &\text{if } mod_i > f\_pass_i,\ r_i = mod_i - f\_pass_i \\ &\text{if } mod_i < t\_pass_i,\ r_i = mod_i - t\_pass_i \end{aligned} \quad (4.3)$$

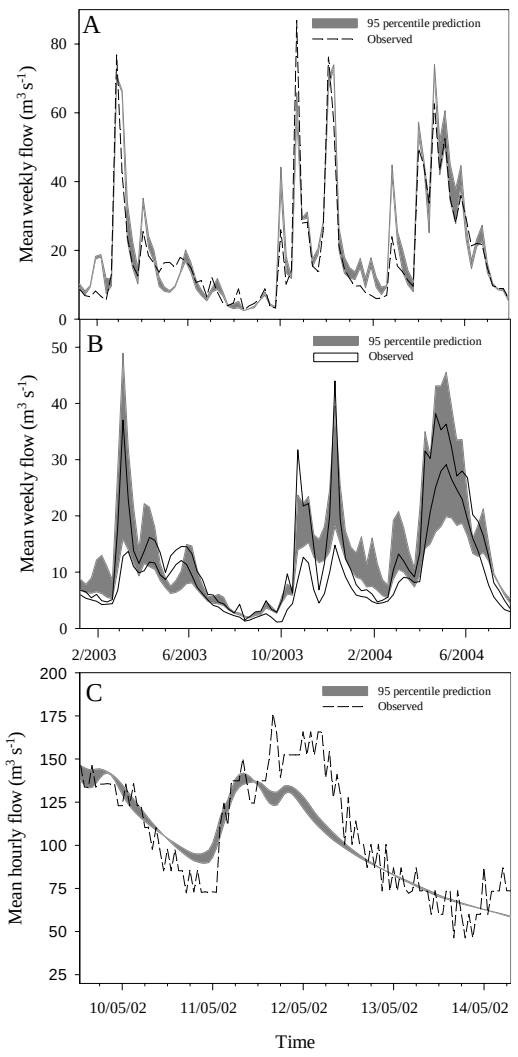

Figure 4.2 – (A) *WEEK* and (B) *BASE* observed values and prediction during the validation period. (C) *HOUR* observed values and prediction during the calibration period.

where mod_i stands for the i^{th} modeled subsurface input to the river, and f_pass_i and t_pass_i are the i^{th} estimated groundwater flux into the river calculated by the first and third baseflow algorithm pass, where most probably the actual baseflow value is located. Following the reasoning detailed for $WEEK$, we aggregated the series taking weekly mean values, and the calibration and validation periods also were as in $WEEK$.

- DURC. The inclusion of duration curves (i.e. the percent of time in which the river flow equaled or exceeded some given values) is a current practice in HSPF applications (Doherty and Johnston 2003). The definition of several values to which calculate exceedence times (in this study 1.7, 3.5, 7, 14, 28, 56, and 85 $m^3\ s^{-1}$) includes a frequency descriptor into the calibration process, which is relatively independent of the original flow series and can be compared to a nonparametric streamflow signature, i.e. its calculation does not depend on statistical assumptions (Shamir et al. 2005). To calculate $DURC$ we used the daily flow record, because for this nonparameric descriptor the small-scale noise should not have a deleterious effect. Calibration and validation periods were as in $WEEK$.

The next step is to define the way in which the parameter estimation method will use the above responses during the calibration process. Although the most powerful way would be to search for the entire Pareto set of parameter values that simultaneously fits the different responses (i.e. considering each response as a separate OF), previous runs with standard methods to apply these methodologies (Vrugt et al. 2003) in our problem revealed that the convergence of results was extremely slow (very poor convergence after 100 000 simulations of 25 *seconds* each one), therefore unsuitable for all practical purposes. We turned to the PEST package (Doherty 2004), which implements a robust variant of the Gauss-Marquardt-Levenberg method, and because it is a stable local OF minimum finder, that also gives an estimation of the uncertainty in op-

timized parameter values. Since PEST is not able to manage several OFs at the same time, we aggregated the different responses into a single OF, using the Euclidian distance function by Madsen (2003). This transformation function compensates for differences in the magnitudes of the different responses, giving to all the same influence on the aggregated OF near the optimum. This is equivalent to search a point in the front of the Pareto set (Khu and Madsen 2005). Finally, we used the PD_MS2 driver included in the PEST package, which applies PEST in a population-based inspired approach to efficiently find the global OF minimum (Skahill and Doherty, 2006). We compared the results obtained with the SCE algorithm (Duan et al. 1992) and PD_MS2 in several calibration runs with a simplified version of our model, and we observed that the difference between both methods were always less than 5% of the final value of the OF.

4.3.2 Regularization

The model structure presented in Section 4.2 determines that we had to calibrate 34 parameters, which is a considerable amount. In this context, the risk of overfitting and poor convergence due to numerical instability problems during the calibration process is elevated. Since the reliability of final parameters values and the resultant spatial and temporal heterogeneity is highly degraded if such overfitting exists (Moore and Doherty 2005), we should introduce some regularization constraint into the calibration process.

We applied three different regularization strategies. First, we imposed a smoothing constraint on parameter values. This was achieved by taking differences between the values of parameters pertaining to the same model parameter (e.g. between the three *LZSN* parameters defined by lithology), and requesting that each such difference be zero if possible. The second strategy was to impose a parsimony criterion. For this, we defined a preferred state for certain parameter values (Table 4.1), in order to include the effect caused by these parameters only if necessary. For instance, $UZSN\Sigma$ and $LZETP\Sigma$ were defined with a preferred value of 0, i.e. no intra-annual variability for $UZSN$ and $LZETP$ parameters. Thus, any departure from zero will be penalized during the calibration process. In a similar way, a zero preferred prior value was assigned to $KVARY$, $BASETP$, and $DEEPFR$ parameters. Finally, the third strategy was to assign a prior value for $INFEXP$ and IRC, using bibliographical information consistent with our watershed characteristics (Table 4.1). All these regularization constraints were implemented in PEST supplying the preferred values as observations in the form of prior information equations (Doherty 2004), which constitutes an additional observation group ($REGU$) added to the OF. This implies that the residuals pertaining to $REGU$ will depart from zero during the calibration process only if this is compensated by the residuals for the other OF components going to zero. The advanced regularization options included in PEST were not implemented, because they are incompatible with the PD_MS2 driver.

4.4 Numerical analyses

In order to elucidate if the final values of related parameters are different enough to justify the proposed zonation, we estimated the distribution of the final parameters values taking advantage of the confidence interval calculation implemented in PEST (Doherty 2004). At the end of the calibration run, PEST gives the optimized value (i.e. the 50 percentile), and the 2.5 and 97.5 percentiles for each parameter. We adjusted a Beta cumulative distribution function (CDF) to these three points, in order to obtain an approximation of the posterior parameter distribution. If deviation of normality was too high to prevent a Beta function to fit the percentiles, two Beta CDFs were overlaid to draw the distribution. Once the empirical distributions were obtained, we sampled 10 000 points from each of the fitted CDFs. These synthetic samples were then used to test paired differences between the values of the parameters pertaining to the same model parameter (e.g. between the three *LZSN* parameters defined by lithology). This was achieved evaluating the distributions overlapping with a 1000 boot-

strap resampling procedure, in which the probability of a point from one distribution lying inside the 95 percentile range of the other distribution was computed.

Although we first solved a calibration run including all the available information (i.e. defining an OF with all the responses, plus *REGU* as a regularization constraint. This run will be referred hereafter as $6f$), the fact that the PD_MS2 calibration runs demanded only a moderate amount of HSPF runs (around 15 000), allowed us to apply an stepwise procedure to assess the influence of the inclusion of different data on the calibration process. Taking the most simple regularized calibration run as a reference (i.e. including only the *WEEK* series and *REGU* in the OF. This run will be referred hereafter as $2f$), we performed PD_MS2 calibration runs including one additional response at a time (e.g. $2f + TRIM$, $2f + BASE$, etc.). Then, the best calibration run was taken as the new starting point (called $3f$), and the process was repeated until all responses were included in the OF (i.e. the run $6f$). To define the best runs we calculated the Nash-Sutcliffe Coefficient (NS) (Legates and McCabe 1999) for each component of the OF, and then we standardized the results by that obtained with $2f$:

$$z_c^r = \frac{y_c^r - y_c^{2f}}{s_c} \tag{4.4}$$

where z is the standardized NS calculated with the c component of the OF for the calibration run r, y is the NS for the same indices, s is the standard deviation of all the NSs calculated for the c component of the OF, and y_c^{2f} is the NS calculated with the c OF component for the run $2f$. We standardized the NSs because the variability across runs was very different between OF components, and we were interested in an aggregate measure of performance. Then, for each calibration run the mean of the standardized NSs was calculated (excluding *REGU* and the OF response included during the run). Therefore, this mean expresses the effect of the inclusion of a new response in the OF on the fit to the other modeled responses. The run with the highest mean was considered best, and was taken as the new starting point for the next step.

Finally, the effect of the inclusion of different data in the OF on the uncertainty in the modeled baseflow was assessed with the predictive analysis capability of the PEST package. The aim of predictive analysis is to maximize and minimize certain model prediction (in this case, the mean weekly baseflow during the validation period). To achieve this, the previously optimized parameter values should be modified, with the restriction that the OF cannot exceed certain threshold value (for a comprehensive treatment *see* Doherty [2004]). Fortunately, theory exists that relates this threshold to a specific prediction probability, using a first-order Taylor series expansion linearization (Vecchia and Coley 1987). Accordingly, we calculated 95% confidence intervals for the mean weekly baseflow during the validation period, for the different calibration runs stated above. The same approximation was used to draw 95% prediction intervals for the different responses using the results from $6f$.

4.5 Results

Table 4.2 shows the NSs calculated for all the calibration runs, and the NS for the prediction of $6f$ compared to validation data. Focusing on $6f$ results, both for calibration and validation experiments the NSs are in the range observed in many other studies. Only the NS for *HOUR* shows a value considerably lower than for the other groups (except the regularization group *REGU*). However, observing Figures 4.2 and 4.3, it is clear that the major features present in the observed data are represented in the model with enough detail, including the *HOUR* response. Modeled values uncertainty did not bracket many observed values, suggesting that error sources other than parameter uncertainty are present, as is usual in any complex hydrological application (Vrugt et al. 2005). Therefore, we considered that the model performance is acceptable, and can be compared to that obtained in similar studies.

Table 4.2 – Response-specific Nash-Sutcliffe Coefficients for the different calibration runs.

Objective Function Components	WEEK	HOUR	TRIM	BASE	DURC	REGU	
			Nash-Sutcliffe Coefficient (Calibration)				
WEEK + REGU (2f)	0.974	0.223	0.959	0.979	0.9998254	0.931	
2f + TRIM	0.972	0.261	0.979	0.978	0.9999784	0.733	
2f + DURC	0.968	0.641	0.957	0.968	0.9999997	0.739	
2f + BASE	0.972	0.139	0.939	0.993	0.9997346	0.803	
2f + HOUR (3f)	0.975	0.886	0.958	0.976	0.9999183	0.674	
3f + DURC	0.968	0.880	0.961	0.963	0.9999994	0.573	
3f + BASE	0.971	0.857	0.942	0.991	0.9999407	0.682	
3f + TRIM (4f)	0.972	0.872	0.980	0.950	0.9999674	0.711	
4f + DURC	0.971	0.852	0.975	0.938	0.9999989	0.072	
4f + BASE (5f)	0.976	0.866	0.975	0.981	0.9999760	0.539	
5f + DURC (6f)	0.975	0.835	0.976	0.977	0.9999989	0.081	
			Standard Deviation of Nash-Sutcliffe Coefficient				
	0.003	0.302	0.015	0.017	0.0000856		
			Nash-Sutcliffe Coefficient Standarized by 2f				*Mean* [a]
WEEK + REGU (2f)	0	0	0	0	0		
2f + TRIM	-0.50	0.12	1.33	-0.07	1.79		0.34
2f + DURC	-1.98	1.38	-0.13	-0.68	2.04		-0.35
2f + BASE	-0.75	-0.28	-1.42	0.85	-1.06		-0.88
2f + HOUR (3f)	0.64	2.19	-0.10	-0.17	1.08		0.36
3f + DURC	-2.19	2.17	0.11	-0.97	2.03		-1.02
3f + BASE	-0.94	2.09	-1.19	0.74	1.35		-0.26
3f + TRIM (4f)	-0.66	2.15	1.45	-1.74	1.66		-0.25
4f + DURC	-0.92	2.08	1.06	-2.46	2.03		-0.43
4f + BASE (5f)	0.81	2.12	1.09	0.13	1.76		1.56
5f + DURC (6f)	0.60	2.02	1.17	-0.11	2.03		0.60
			Nash-Sutcliffe Coefficient (Validation)				
6f	0.942	na	0.973	0.968	0.99977	0.081	

[a] Mean of the Nash-Sutcliffe Coefficients for the corresponding calibration run, excluding the coefficient corresponding to the response included during the run.

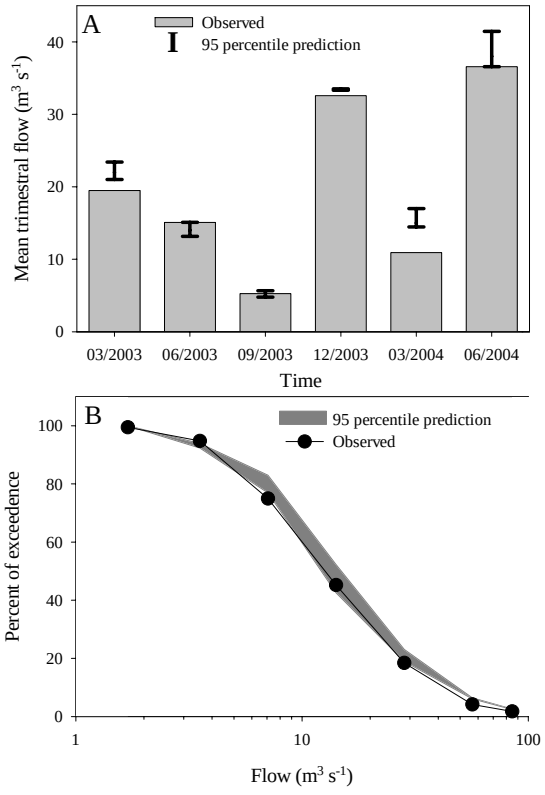

Figure 4.3 – (A) TRIM and (B) DURC observed values and prediction during the validation period.

4.5.1 Spatial distribution of adjustable parameters

Taking again the results obtained from the $6f$ calibration run, we tabulated the sensitivities of all the adjustable parameters with respect to the different components of the OF and the composite, the final optimized values, and their uncertainties as calculated by PEST (Table 4.3). From the composite sensitivity, we can observe that the most sensitive parameters are those related to the nominal storages (*LZSN* and *UZSN*), the losses of the system via different pathways (*LZETP* and *DEEPFR*), and the shape of the recession curves (*AGWRC*). By contrast, parameters related to the interflow had no effect on the OF. More interestingly, there were some parameters that showed a significant spatial variability. Whereas the storage parameters did not show any significant variation across the watershed feature considered (i.e. the distribution of final parameter values was highly overlapped between parameters pertaining to the same model parameter), for the infiltration driving parameters (*INFILT* and *INFEXP*), recession parameters, evapotranspiration, and deep percolation values very significant spatial variability was evident across different basin attributes. Remarkably, all but one of the sensitive parameters assigned to different lithological types showed significant heterogeneity, especially *L*3 compared to the other two lithologies. Similarly, parameters defined by land use classes (*LZETP*) or groups (*AGWRC*) showed significant differences between them.

Monthly variation in parameter values (expressed by $UZSN\Sigma$ and $LZETP\Sigma$) was only supported for the evapotranspiration parameters, which showed a strong seasonal cycle (Table 4.3). Percentiles for $UZSN\Sigma$ included zero, therefore no seasonal variation in this nominal storage is supported by the calibration data.

4.5.2 Relevance of data types on the calibration process

Table 4.2 shows that the first response added to the $2f$ calibration setting was *HOUR*, followed by *TRIM*, *BASE*, and finally *DURC*. Considering the standardized NSs for the $6f$ run, the most important differences with respect to the $2f$ run were that showed by *HOUR* and *DURC*, and only the fit with *BASE* was marginally inferior. Thus, it can be stated that the inclusion of different data types into the calibration process exerted a positive effect on the performance of the model. However, although the inclusion of a response in the OF motivated a positive effect in the model performance modeling this same response, no clear pattern can be depicted about the effect on the other responses. For instance, the effect of the inclusion of *DURC* into $2f$ over the fit to *TRIM* was negative (*see* Table 4.2), but the same response added to $3f$ (when *HOUR* was already included in the OF) exerted a positive effect on *TRIM*. Similarly, the effect of including addi-

Table 4.3 – Parameter sensitivity and uncertainty for the run 6*f*.

	Sensitivities [a]							*Percentiles*	
Parameter	WEEK	HOUR	TRIM	BASE	DURC	*Composite*	*Optimized Value* [b]	*2.5*	*97.5*
$LZSN^{L1}$	0.04	0.03	**0.13**	0.03	**0.20**	**0.087**	3.46	2.13	5.26
$LZSN^{L2}$	0.03	0.03	**0.11**	0.02	**0.19**	**0.077**	5.96	3.59	9.40
$LZSN^{L3}$	0.04	0.03	**0.14**	**0.05**	**0.20**	**0.093**	5.57	3.83	7.85
$UZSN^{S1}$	0.05	0.04	**0.11**	0.02	**0.36**	**0.115**	0.018	-0.117 [d]	0.172
$UZSN^{S2}$	0.03	0.02	**0.09**	-	**0.19**	**0.068**	0.040	-0.267 [d]	0.460
$UZSN^{S3}$	**0.32**	**0.11**	**0.71**	0.08	**1.84**	**0.612**	0.032	0.004 [d]	0.061
$UZSN\Sigma^{S1}$	-	-	-	-	-	-	0.004	-0.087 [d]	0.432
$INFILT^{L1}$	0.01	0.02	0.02	0.01	**0.09**	0.033	0.17	0.09	0.30
$INFILT^{L2}$	0.02	0.03	0.03	0.02	**0.13**	0.047	0.09	0.05	0.16
$INFILT^{L3}$	0.01	0.02	0.03	0.01	**0.09**	0.033	$*^{(L1)}$ $***^{(L2)}$ 0.38	0.23	0.60 [d]
$INFEXP^{L1}$	-	0.02	0.01	-	0.03	0.016	4.72	2.63	7.98
$INFEXP^{L2}$	0.01	0.01	0.02	-	0.04	0.019	3.78	2.28	5.95
$INFEXP^{L3}$	-	0.02	0.02	-	0.03	0.016	$**^{(L1)}$ $\S^{(L2)}$ 1.81	0.81	3.34
$KVARY^{L1}$	0.01	0.01	0.02	0.02	**0.05**	0.022	0 [c]	-0.19 [d]	0.29
$KVARY^{L2}$	0.01	0.01	0.02	0.02	0.04	0.019	0.18	-0.14 [d]	0.75
$KVARY^{L3}$	0.03	0.04	**0.06**	**0.05**	**0.10**	**0.056**	$\S^{(L1)}$ 0.38	0.10	0.77
$AGWRC^{G1}$	**0.31**	**0.16**	**0.61**	**0.55**	**0.63**	**0.452**	$**^{(all)}$ 0.980	0.975	0.986
$AGWRC^{G2}$	**0.11**	**0.08**	**0.20**	**0.18**	**0.34**	**0.182**	0.959	0.945	0.976
IRC^{S1}	-	-	-	-	-	-	0.53	0.39	0.87 [d]
IRC^{S2}	-	-	-	-	-	-	0.51	0.39	0.82
IRC^{S3}	0.01	0.04	-	-	0.03	0.016	0.50	0.43	0.60
$INTFW^{L1}$	-	-	-	-	-	-	7.04	0.89 [d]	55.4 [d]
$INTFW^{L2}$	-	-	-	-	-	-	7.23	0.95 [d]	54.8 [d]
$INTFW^{L3}$	-	-	-	-	-	-	6.88	0.84 [d]	56.3 [d]
$LZETP^{CR}$	0.02	0.01	**0.06**	0.03	**0.09**	0.041	0.61	0.43	0.89
$LZETP^{MX}$	0.04	0.03	**0.14**	**0.06**	**0.18**	**0.090**	$\S^{(CR, CF)}$ 0.89	0.68	1.17 [d]
$LZETP^{DC}$	**0.05**	0.02	**0.15**	0.04	**0.18**	**0.088**	$***^{(all)}$ 0.31	0.24	0.39
$LZETP^{CF}$	0.04	0.02	**0.11**	0.04	**0.19**	**0.080**	0.63	0.48	0.83
$LZETP^{BL}$	-	-	0.01	-	-	-	0.75	0.46	1.25 [d]
$LZETP\Sigma$	**0.05**	0.01	**0.13**	**0.05**	**0.07**	**0.063**	1.0 [c]	0.83	1.21 [d]
$BASETP^{DC}$	-	-	0.02	0.01	0.03	0.014	0.04	-0.003 [d]	0.17
$DEEPFR^{L1}$	-	-	0.01	-	0.02	-	0.05	-0.004 [d]	0.16
$DEEPFR^{L2}$	-	-	0.01	-	0.02	-	0.06	0.001	0.18
$DEEPFR^{L3}$	0.04	0.04	**0.12**	**0.05**	**0.21**	**0.092**	$***^{(all)}$ 0.34	0.24	0.47

[a] For clarity, sensitivities below 0.01 are not shown, those above 0.05 are in boldcase, and those above 0.1 are underlined.

[b] Significant paired differences between spatially distributed parameters are denoted by: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, and § $p<0.1$. Within brackets, the index for the significantly different parameter.

[c] Parameter optimized on a limit (*see* Table 4.1).

[d] Percentile outside the predifined range for the parameter (see Table 4.1).

tional responses into $3f$ was notoriously negative (*see* the *Mean field* in Table 4.2), but once $TRIM$ was added to the OF, the inclusion of *BASE* exerted a very positive effect on the fit to the other responses. Again, once *BASE* was inside the OF, the inclusion of $DURC$ was positive for the rest of OF components compared to the $2f$ run, despite the inclusion of this response was consistently negative before *BASE* was in the OF. All these indicates that the inclusion of additional responses in an aggregated, single OF has a profound effect in its multidimensional shape, and consequently in the whole calibration process.

The inclusion of data types in the OF also affected the regularization component of the OF ($REGU$). Table 4.2 shows that the NS for $REGU$ descended as more responses were added to the OF, which means that a moderate additional amount of spatial heterogeneity and non-parsimony in model parameters were needed to account for the new details to be modeled. However, the most striking effect is that showed by the inclusion of $DURC$ once $HOUR$ and $TRIM$ were already included in the OF. Then, the fit to $REGU$ descended by an order of magnitude, implying that a huge amount of spatial heterogeneity and non-parsimony in model parameters was necessary to maintain an acceptable level of fit. This again stresses the fact that the shape of the aggregated OF in the hypercube defined by the number of parameters undergoes very profound changes as new data are included into the calibration process.

The central role played by $DURC$ in the calibration process was also highlighted from the response-specific parameter sensitivities during run $6f$ (Table 4.3). This response was sensible to variations of several parameters, including nominal storages, infiltration drivers, recession curves, evapotranspiration losses, and deep percolation. $TRIM$ was also sensibly affected by the same parameters, excluding those related to infiltration processes. As expected, parameters related to recession and groundwater routing (*AGWRC*, *LZSN*, *LZETP*, and *DEEPFR*) showed high sensitivity for *BASE*. The short-term surface responses ($HOUR$ and $WEEK$) were sensitive to several parameters.

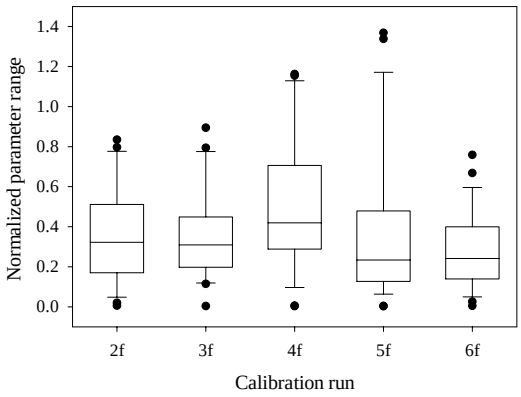

Figure 4.4 – Box-whisker plots of the normalized range of the 25 sensitive parameters after completion of five calibration runs (middle line, median; lower box line, first quartile; upper box line, third quartile; lower whisker line, 2.5 percentile; upper whisker line, 97.5 percentile; dots denote individual outliers).

4.5.3 Relevance of data types on the uncertainty of baseflow calculations

The inclusion of several responses into the OF is intended to constraint the possible values that the adjustable parameters can take. Thus, it was expected that the range of the estimated parameter uncertainty, as implemented in PEST, would reduce as more data were included in the OF. Consequently, we calculated the range of uncertainty for the 25 composite-sensitive parameters (Table 4.3), taking the 95% percentile limits calculated by PEST and normalizing by the predefined parameter limits (*see* Table 4.1). Figure 4.4 shows the results for the five best calibration runs expressed as box-plots that describe the distribution of the normalized ranges for the 25 parameters. The results during the first inclusions ($3f$ to $4f$) did not support the expectation, because indeed the parameter ranges globally rose. Only after the inclusion of $DURC$ (i.e. $6f$) the uncertainties were globally below the values obtained during the run $2f$.

However, the uncertainty in the modeled mean

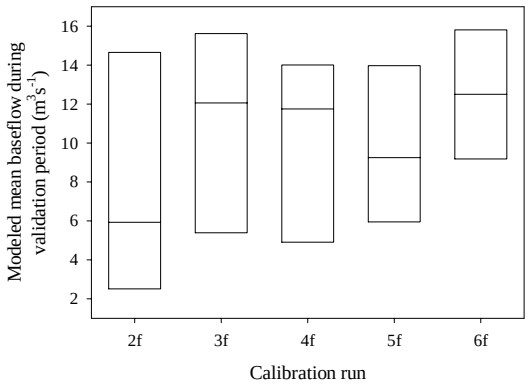

Figure 4.5 – Modelled mean baseflow and associated uncertainty during the validation period after completion of five calibration runs (middle line, median; lower box line, 2.5 percentile; upper box line, 97.5 percentile).

baseflow obtained during the validation period supports the idea that the inclusion of several responses in the OF had beneficial consequences (Fig. 4.5). The effect on baseflow uncertainty was conspicuous in the minimum values, especially in the steps that defined $3f$ and $6f$. Also interesting, the mean value for $2f$ is most probably biased, if it is compared to the mean values for the other runs, established at higher values. Thus, the inclusion of responses in the OF not only reduced the uncertainty of calculations, but also added accuracy to the results.

4.6 Discussion and conclusions

The final distribution of optimized parameters clearly showed that maintaining some spatial heterogeneity in the model parameterization was justified. Several parameters distributed by both lithology and land uses showed significant differences, although those linked to the slope classification did not show any spatial heterogeneity. This poses the question whether an a priori definition of intended homogeneous zones is the best way to impose some spatial resolution to parameter values. Despite the fact that this approach worked well in this study, a more convenient method will involve the definition of as many parameters as HRUs the model has. Of course, the application of such a method in a complex watershed-scale model demands appropriate numerical tools to deal with the hundreds, if not thousands, of parameters that will be present. One alternative is the hybrid regularized inversion by Tonkin and Doherty (2005), in which after a time consuming initial calibration run the eigenvectors representing principal orthogonal directions in parameter space are used to define a feasible number of super parameters. Then, these super parameters can be optimized using a standard fitting methodology. In principle, we could use this method to parameterize the system, and then searching for the relationships between parameter values and watershed features. This could be a powerful way to assess the feasibility of the final parameter values, and to understand the hydrological cycle in the basin. However, the hybrid regularized inversion is based on parameter sensitivities, which in a complex hydrological model can be dependent on the value around that sensitivities are calculated. Thus, a good procedure could be using an a priori definition of homogeneous zones to search for a suitable initial parameters set to feed the hybrid inversion. Although the results from this study could be a good starting point to test this possibility, this is beyond the scope of this work.

Independently of the parameter zonation methodology applied, a proper prior information scheme is of a paramount importance for the success of a regularized calibration process (Tonkin and Doherty 2005). Despite we used an extremely simple regularization methodology, the calibration runs were numerically stable, and the final parameter values and ranges (Table 4.3) were reasonable. For instance, although the prior information scheme imposed homogeneity and parsimony criteria, the inversion process assigned a positive $DEEPFR$ value for the hard rock, fractured area, while maintaining the other areas not significantly different from zero. Similarly, the $LZETP$ value for the deciduous forest was significantly lower than that for other forestlands, as is expected (Swift et al. 1975). The seasonality of this parameter, expressed by $LZETP\Sigma$,

was also very pronounced, indicating a strong seasonality in evapotranspiration losses, a usual result for Mediterranean areas (Bernal et al. 2004). It is also worthy to mention the significant differences found in the *INFEXP* parameters (Table 4.3). Since a value significantly greater than 2 prompt HSPF to switch from an infiltration-excess runoff generation to a saturation-excess one (Berris 1995), we can state that there is a very different hydrological behavior between the different lithological zones in our watershed. The high altitude mountainous zone is governed by a Hortonian overland flow, while zones dominated by sedimentary rocks and alluvial deposits show saturation overland flow. This result agrees with the view of Johnson et al. (2003), who argued against the common practice of HSPF modelers to maintain the default (i.e. 2) *INFEXP* value during simulations. All the above significant differences between spatially discretized parameters clearly shows that a rich variety of hydrological behaviors can be present in a complex watershed, and that a proper regularized inversion is a powerful methodology to help solving the calibration step.

Although the aggregation of several responses in the OF was helpful during the calibration process, it is certainly difficult to predict what kind of data will be most worthy to include into the OF. Apart from the main target (e.g. *WEEK* in this study), the inclusion of nonparametric descriptors like *DURC* is judged convenient (Shamir et al. 2005). However, as we demonstrated in this study, the beneficial effects of a response can arise only after the inclusion of other a priori less efficient data. Thus, we recommend to introduce in the OF as many as data are available, including by-products of sampled data (e.g. *TRIM* and *BASE*). Since the topological complexity of an aggregated OF will probably increase as more responses are included, algorithms that search for the Pareto set of parameters (Gupta et al. 1998; Vrugt et al. 2003; Khu and Madsen 2005) without aggregating responses should be preferred a priori. However, the applicability of these methodologies will depend on the computational burden necessary to reach convergence, which can be very demanding for complex watershed-scale hydrological models. In addition, the use of data derived from observed data is not free of risk. In our case, *BASE* was built with an algorithm that assumes a constant dynamics for the recession curve. Considering that HSPF includes a parameter (*KVARY*) to model time-varying recession curves, the use of a series like *BASE* could artificially impose zero values for *KVARY*. Although it seems that this was not the case for our calibration, since $KVARY^{L3}$ was optimized to a non-zero value (Table 4.3), the presence of this significant non-zero value for *KVARY* is not in accordance with the assumptions used to build the *BASE* series. Thus, caution should be taken when including non-observational data into the OF. In principle, temporal aggregation (e.g. *TRIM*) and non-parametric values (e.g. *DURC*) are best options (Shamir et al. 2005).

In any case, calibrating a complex watershed-scale model using only one response is a very dangerous procedure, even in the case a good prior information scheme is available. As depicted in Figure 4.4, the uncertainty associated to water routing was very high for the $2f$ run, and the estimated values were probably biased. For a water quality application this would be critical, especially if management decisions lay on the modeling outcomes, because this lack of precision and accuracy produces useless results despite the goodness of fit attained with river runoff.

In conclusion, combining regularized inversion and a multi-response OF useful results can be obtained even with complex applications including spatial and temporal heterogeneity in model parameters. Since good numerical methods appropriate for these purposes are now available (e.g. Tonkin and Doherty 2005), maintaining the observed basin heterogeneity in the model abstraction should always be considered. In fact, the inclusion of the inherent heterogeneity present in real catchments into hydrological models is an old concern that still is a matter of debate (e.g. Jakeman and Hornberger 1993; Beven 1996; Boyle et al. 2001; Schulz et al. 2006), and this concern should be transferred to water quality models despite hydrology is not the main target in these applications. The management of field observations has also focused the attention of hydrologists (Kuczera 1982; Raat et al.

2004), and the results from this work emphasizes that when an OF aggregating different data is the only practical option, caution should be taken when including and excluding responses. Until a new generation of poor parameterized ”gray box” models (in the sense of Kirchner [2006]) are available, we must work with the state-of-the-art ”mathematical marionettes” (again in the sense of Kirchner [2006]), which usually are not suitable to work with Pareto solvers if the model structure is complex. In these cases, the approximations outlined in this paper, combined with other tools developed to calibrate complex models, like prior sensitivity analysis (Zheng and Keller 2006) and inclusion of water quality constituents as chemical tracers (Bernal et al. 2004), can help obtaining reliable model outcomes even in complex, real world problems.

Chapter 5

Modeling river water temperature in semidistributed water quality applications using deterministic, empirical, and hybrid formulations

5.1 Introduction

The role of water temperature on a myriad of processes in the river ecosystem is an old issue in aquatic sciences, and its significant effects have been demonstrated at every scale, from molecular interactions to system-wide measures of lumped processes. Thus, this variable is a recurrent driver in deterministic or empirical models in which biotic and/or chemical activity is directly or indirectly involved (Chapra 1997). Nonetheless, water temperature greatly determines the behavior of other physical features like water density. Although traditionally little attention has been paid to river water density in studies focused on biological or chemical dynamics in the river itself, the situation changes when the river model outcomes are coupled to a lentic system. Since the river water enters the lake at a depth defined by the reciprocal differences between water densities (Alavian et al. 1992), a proper modeling framework coupling a river and a lake ecosystem should take into account river temperature as an important factor shaping lake hydrodynamics. This is particularly relevant in man-made reservoirs, where the riverine advective flow is a dominant component of hydrodynamics (Ford 1990).

Despite the coupling between river and lacustrine processes can be achieved by raw empirical approaches (e.g. Marcé et al. 2006), the use of process-based, spatially explicit models has become more popular as increasing computing power has been available to researchers and managers (Singh and Woolhiser 2002). In despite of the increasing complexity and associated uncertainty in the model outcomes, these approximations are tempting because their inherent causal approach, the possibility to include dynamic changes in the components of the system, and the spatial and temporal resolution they offer. If water quality concerns and their solutions are major targets of the application, the

river input to the lake (flow, heat, nutrients, etc.) is usually solved through a semidistributed watershed scale model. Although a priori the use of fully spatially distributed applications is possible, water quality modelers working at the watershed scale often face the problem of data scarcity, which makes fully distributed models far from practical. However, concerning the simulation of the river water temperature, the division between distributed and semidistributed models are of secondary importance, since the equations that solve for this variable are usually implemented at the level of the river reach, unit that is topologically and functionally equivalent in both modeling frameworks.

Although the role of land derived inputs of heat on river temperature dynamics is not irrelevant (Bogan et al. 2004), most of the work on river water temperature has been focused on the non-advective heat exchanges. From the point of view of modeling, the non-advective processes can be formulated explicitly including all the known heat exchanges between water, atmosphere, and riverbed. Alternatively, all these processes can be lumped in empirical relationships relating a defined lumped exchange process and some variable measured in the field. In the first case, a considerable amount of meteorological and geomorphologic data is needed to feed and fit the equations, but we gain detailed knowledge about the precise mechanisms controlling temperature in the reach. In the second case, we can obtain reliable results with just a series of meteorological data (e.g. air temperature), but we loss most of the details about heat exchange dynamics, an affordable drawback if our main interest is not the temperature dynamics itself.

The basic antecedent of our work is the aim of building a semidistributed watershed scale river model that will be linked to a reservoir dynamic model in the future, in order to implement remediation measures against eutrophication. Therefore, we were interested in the river water temperature not only as a driver of biological and chemical processes, but also as a determinant of the river water density. In this paper we explore the ability of a semidistributed model predicting river water temperature using a do-everything approach for the non-advective heat exchanges at the reach level, an empirical relationship for the same purposes, and finally a hybrid approximation. For the empirical models definition, we chose the equilibrium temperature approach, using the air temperature as a predictor variable. Although temperature gradients are not the main causal driver of heat exchanges in rivers (Johnson 2003), we do not consider this procedure a pure empirical, causal-void exercise, since in its theoretical formulation the equilibrium temperature is a physically based concept (Edinger et al. 1968).

One limitation of this work is that we tested the performance of the three approaches modeling water temperature in a single case study. However, we consider our scenario as a prototypical situation for temperature modeling in water quality applications at the watershed scale for several reasons: (1) as usual in water quality applications, we are facing a huge watershed, compared to the study sites of process-based research. (2) We have a lot of good quality and spatially distributed meteorological data, but the meteorological stations are usually located far from the river reaches, and usually do not represent the local riverine microclimate. And (3), only limited river temperature information is available to parameterize the equations.

5.2 Study site

The Ter River watershed at Sau Reservoir is 1380 km^2 (Fig. 5.1), mainly covered by woodland (78%) and agricultural land (16%). As usual in the Mediterranean region, precipitation is highly variable in both, the spatial and the temporal dimensions. Most part of the watershed has annual precipitations around 800 *mm*, although in the mountainous North end values rise to 1000 *mm*, and locally till 1200 *mm*. Precipitations fall mainly during April-May and September. Ter River daily mean temperature at Roda de Ter (Fig. 5.1) ranges from 3 $^{\circ}C$ to 30 $^{\circ}C$, whereas there is a marked variability in the air temperature range across the watershed. Whereas in the vicinity of the sampling point daily mean temperatures go from -3 to 30 $^{\circ}C$, in the

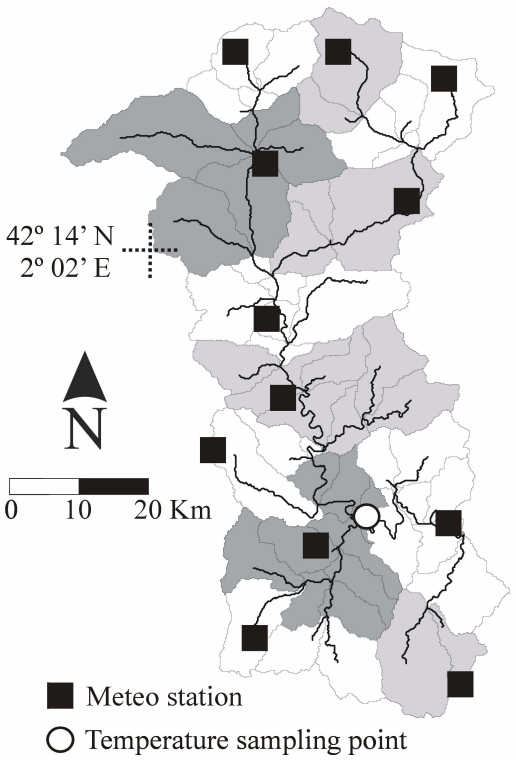

Figure 5.1 – Ter watershed map showing stream network (dark line), temperature sampling point, and meteorological stations used in this study. Areas included in gray lines correspond to the subbasins delineation with HSPF. Background colors (white to dark gray) define the meteorological information splitting arrangement.

mountainous North end of the watershed temperature ranges between -14 to 18 $^{\circ}C$. This variability should be considered during empirical modeling, because at water temperature below 0 $^{\circ}C$ and above 20 $^{\circ}C$ the relationship between air and water temperature losses its linear dependence (Mohseni and Stefan 1999).

The headwaters are located in the north end of the basin (Fig. 5.1), at 2500 *m.a.s.l.* in the Spanish Pyrenees ranges, and for this study the watershed ends in the sampling point near Roda de Ter, just upstream Sau Reservoir, at 443 *m.a.s.l.* The headwaters flow over hard materials (igneous and metamorphic rocks) covered by a mixture of high altitude shrublands, and conifer and deciduous forests over a steeped terrain. Downstream, around the meeting point of the two main headwater courses, the forested land accounts for nearly all the terrain, the slopes have moderated, and the river flows over sedimentary rocks. Then the river enters in the populated agricultural plain where the main human settlements are located. Here the alluvial deposits are abundant, and unirrigated crops cover most part of the land. Thus, the watershed includes a complex mixture of geomorphologic, lithologic, and land cover features.

The sampling station was located in a well mixed reach 50 *m* wide with 2 *m* of maximum depth. Mean streamflow was 17 $m^3\ s^{-1}$. River water was sampled in the thalweg ca 50 *cm* above the river bed, using a pump connected to an automatic sampling station including a temperature probe (aquaTest by Adasa Sistemas SA, Barcelona, 0.1$^{\circ}C$ resolution), run by the local water agency (Agència Catalana de l'Aigua, ACA). Although ACA has strict maintenance routines, periodic checks by the authors assured the performance of the sampling equipments during the study period. The sampling station collected hourly water temperature data from January 2001 to July 2004. Hourly meteorological data for this study (air temperature, relative humidity, solar radiation, wind velocity, and precipitation) comes from 12 meteorological stations (Fig. 5.1) run by the local meteorological agency (Servei Meteorològic de Catalunya, SMC).

5.3 Modeling framework and advective heat inputs

For water quality applications the most usual time step for modeling purposes is the day. Only rarely a shorter time step will be required, and as we move beyond the week the question of why using a complex dynamic model arises. Thus, the main modeling objective was the prediction of mean daily river water temperatures at the sampling point. A second objective was the modeling of the minimum daily river water temperature. This was motivated

by two facts. First, the minimum daily temperature defines the maximum water density achieved during the 24 hours period. This information could be critical during some periods of the year (e.g. January-March), in which the river can enter the reservoir as a density current or an overflow depending on slight variations of the river density. And second, we already had the model application ready to be used at the hourly time step, because a previous effort was made to validate the hydrological modules of the model, which should be run at the hourly step. Thus, we had the opportunity to test the ability of the model fitting the daily amplitude of river water temperature (expressed as the ability to fit both, mean and minimum water temperatures) without significant supplementary effort.

HSPF is a semidistributed hydrologic and water quality modeling environment that includes a variety of formulations for water routing through the watershed and other features of the flowing water and materials carried by it. As a semidistributed application, HSPF lumps calculations at the level of subbasin. Thus, it is necessary to split the watershed into subbasins, allowing the inclusion of spatial heterogeneity of watershed features and meteorological inputs to land and river reaches. We refer to *Chapter* 4 for the subbasins delineation procedure and other details of the semidistributed model construction. Regarding temperature modeling, it is especially relevant to consider that if more than one meteorological station is available, different inputs can be defined for the different subbasins. In our case, each subbasin was assigned to a nearby meteorological station considering topography. Figure 5.1 shows the meteorological information splitting arrangement used in the Ter watershed.

The modeling exercises presented throughout this work are based on a successful application of the hydrological modules of the Hydrological Simulation Program-Fortran (HSPF) in the Ter River (*see Chapter* 4). The Nash-Sutcliffe coefficient for the comparison between modeled and observed daily flow values at the sampling point was 0.93, and the median absolute error was 1.4 $m^3\ s^{-1}$. In addition, the model separated the surface and subsurface inputs to the river with a satisfactory associated uncertainty. Thus, we used these flow routing results to model the advective heat exchanges and reach water content in the temperature model.

The advective heat inputs to the river were modeled in an identical way in the different simulations presented in this work. Using built-in options inside the HSPF package (Bicknell et al. 2001), the temperature of the water entering the river as surface runoff from the surrounding land ($T_{adv,s}$) was computed using a linear regression with air temperature:

$$T_{adv,s} = a_s + b_s \times T_{air} \tag{5.1}$$

where a_s and b_s are optimized parameters, and T_{air} is the measured air temperature.

The temperature of the water entering the river as subsurface interflow from the surrounding land ($T_{adv,i}$) was computed using a mean departure from air temperature plus a smoothing factor:

$$T_{adv,i} = T'_{adv,i} + S_i \times (T_{air} + D_i - T'_{adv,i}) \tag{5.2}$$

where $T'_{adv,i}$ is the temperature of the interflow at the start of the interval, S_i is the smoothing factor, and D_i is the difference between air temperature and the mean temperature of the interflow. S_i and D_i are adjustable parameters.

Finally, the water entering the river as groundwater flow from the surrounding land ($T_{adv,g}$) was computed with a formulation equivalent to that shown for interflow:

$$T_{adv,g} = T'_{adv,g} + S_{g,m} \times (T_{air} + D_{g,m} - T'_{adv,g}) \tag{5.3}$$

where $T'_{adv,g}$ is the temperature of the groundwater at the start of the interval. $S_{g,m}$ and $D_{g,m}$ are the smoothing factor and the mean departure from air temperature, respectively. The $m = (1,\ 2, \ldots,\ 12)$ index stands for the month of the year. Since the groundwater versus air temperature relationship can suffer profound changes in watersheds where spring snow melting is significant, we decided to apply 12 different monthly values for the parameters controlling the groundwater temperature. However, since

this would add 24 adjustable parameters to the simulation, we solved the monthly values using a sinus function with adjustable amplitude and phase:

$$S_{g,m} = S'_g + \left(S'_g \times A_{S'g}\right) \times \sin\left((x_m - P_g) \times \frac{2\pi}{365.25}\right) \quad (5.4)$$

$$D_{g,m} = D'_g + \left(D'_g \times A_{D'g}\right) \times \sin\left((x_m - P_g) \times \frac{2\pi}{365.25}\right) \quad (5.5)$$

where $x = (0, 31, 60, \ldots, 335)$ represents the day of the year representative for the month m. S'_g and D'_g are adjustable mean values for the smoothing factor and the mean departure from air temperature, respectively. $A_{S'g}$ and $A_{D'g}$ are the adjustable amplitude of variation for the previous defined mean values, and P_g is the day of the year in which the sinus wave is centered. With this formulation, we can fit 24 adjustable values with just 5 adjustable parameters.

The upstream and downstream reach boundaries were solved by HSPF with the built-in hydraulic modules (Bicknell et al. 2001). The temperature of other water exchanges (i.e. waste water treatment plants inputs and water supply extractions) was also considered using information supplied by ACA. Precipitation on the river reach was also considered as a source of heat, assuming no difference between rain and air temperature during the event.

5.4 Non-advective heat exchanges

We calibrated and validated three models differing in the formulation of the non-advective heat exchanges between the river water and its surroundings. First, a deterministic approach was used taking advantage of the HSPF formulations for solar radiation (Q_S), net long-wave radiation (Q_L), evaporative heat flux (Q_{EV}), and convective heat transfer (Q_{CON}). This model will be referred hereafter as DET. Since HSPF uses common, well-known formulations for these processes (Raphael 1962; Morin and Couillard 1990; Sinokrot and Stefan 1993), we will not detail here all these equations. However, some of the fitted parameters coming from these processes will be described and discussed later. All the variables measured in the meteorological stations were needed to feed the DET model. The change in water temperature due to the non-advective processes is calculated by HSPF through a Taylor series expansion about the temperature at the start of the interval.

Second, we defined a model (referred hereafter as TEQ) modifying the HSPF Fortran code to substitute the non-advective terms outlined above with one lumped process, defined using the equilibrium temperature concept. Equilibrium temperature (T_e) is the water temperature at which the sum of the heat fluxes across the air/water interface is zero (Edinger et al. 1968). If such an equilibrium temperature can be calculated, and we assume that the total heat flux between the atmosphere and the river (Q_a, $kcal\ m^{-2}\ hr^{-1}$) is proportional to the temperature difference between the water temperature and the equilibrium temperature, this heat exchange would be reduced to a Newton's law of cooling (Edinger et al. 1968):

$$Q_a = K_e \times (T_e - T_w) \quad (5.6)$$

where K_e is a heat conduction coefficient ($kcal\ m^{-2}\ hr^{-1}\ {}^{\circ}C^{-1}$) and T_w is the water temperature. Obviously, to apply this formulation for modeling purposes values for both K_e and T_e should be calculated or adjusted. Variability in K_e figures has been related to meteorological variables, especially wind velocity (Thomann and Mueller 1987). However most attempts to model river water temperature using the equilibrium temperature concept have considered K_e as a constant, with no evident deleterious effects on model performance (Caissie et al. 2005). Therefore, we considered K_e as a single-valued adjustable parameter.

By contrast, the T_e cannot be considered as a constant, and its calculation is the trickiest step implementing the temperature equilibrium concept in

Table 5.1 – Parameters and prior ranges used in the definition of the three different model formulations for river temperature. Equation numbers refer to equations in the text.

Parameter	Description	Units	Range
	Advective components (common to all models)		
a_s	Intercept for surface runoff temperature simulation (Eq. 5.1)	ºC	-20 – 40
b_s	Slope for surface runoff temperature simulation (Eq. 5.1)	-	0 – 2
S_i	Smoothing factor for interflow (Eq. 5.2)	-	0 – 1
D_i	Mean departure from air temperature for interflow (Eq. 5.2)	ºC	0 – 30
S'_g	Mean smoothing factor for groundwater (Eq. 5.4)	-	0 – 1
$A_{S'g}$	Amplitude of variation of S_g during the year (Eq. 5.4)	-	0 – 1
D'_g	Mean departure from air temperature for groundwater (Eq. 5.5)	ºC	-3 – 30
$A_{D'g}$	Amplitude of variation of D_g during the year (Eq. 5.5)	-	0 – 1
P_g	Centre of the sinus wave for S_g and D_g monthly variation (Eq. 5.4, 5.5)	Day of the year	1 – 365
	Non-advective exchanges for models DET and DET+		
C_S	Shading factor for net short wave radiation	-	0 – 2
C_L	Proportionality constant for atmospheric longwave radiation emissivity	K^{-2}	$0 – 20 \cdot 10^{-6}$
C_C	Conductive-convective heat transport coefficient	kcal m^{-3} hr^{-1} $ºC^{-1}$	$0 – 20 \cdot 10^{-4}$
C_E	Evaporation coefficient	$mbar^{-1}$	$0 – 10 \cdot 10^{-9}$
	Boundary exchange for model DET+		
K_b	Heat conduction coefficient between water and boundary (Eq. 5.8)	kcal m^{-2} hr^{-1} $ºC^{-1}$	0 – 200
T'_{eb}	Mean equilibrium temperature for boundary compartment (Eq. 5.9)	ºC	-10 – 45
$A_{T'eb}$	Amplitude of variation of T_{eb} during the year (Eq. 5.9)	-	0 – 1
P_{Teb}	Centre of the sinus wave for T_{eb} monthly variation (Eq. 5.9)	Day of the year	1 – 365
	Non-advective exchanges for model TEQ		
K_e	Heat conduction coefficient (Eq. 5.6)	kcal m^{-2} hr^{-1} $ºC^{-1}$	0 – 20
a_{Te}	Intercept for equilibrium temperature simulation (Eq. 5.7)	ºC	-10 – 45
b_{Te}	Slope for equilibrium temperature simulation (Eq. 5.7)	-	0 – 2

river temperature models. Although a priori the equilibrium temperature can be calculated from meteorological data equaling to zero the sum of all non-advective heat exchanges (Bogan et al. 2003), we deliberately kept the model simple and independent of most of the meteorological data assuming a liner relationship between T_e and T_{air} (Caissie et al. 2005):

$$T_e = a_{Te} + b_{Te} \times T_{air} \tag{5.7}$$

where a_{Te} and b_{Te} are optimized parameters. The rationale of this linear rule is based on the close relationship between air temperature and dew point temperature in temperate regions (Mohseni and Stefan 1999). Thus, air temperature is the only meteorological input required to fit the TEQ model.

Finally, we calibrated a third model including an additional boundary heat exchange into the *DET* formulation (model referred hereafter as DET+). We modeled this exchange in an empirical way, using a HSPF built-in option that defines a boundary heat exchange (Q_b, $kcal\ m^{-2}\ hr^{-1}$) as the adaptation of the equilibrium temperature concept:

$$Q_b = K_b \times (T_{eb,m} - T_w) \tag{5.8}$$

where K_b is the adjustable heat conduction coefficient between water and the boundary compartment ($kcal\ m^{-2}\ hr^{-1}\ ^{\circ}C^{-1}$), and $T_{eb,m}$ is the adjustable equilibrium temperature for the boundary compartment during month m. Similarly to Equations 5.4 and 5.5, $T_{eb,m}$ values were computed fitting a sinus wave:

$$T_{eb,m} = T'_{eb} + \left(T'_{eb} \times A_{T'eb}\right) \times \sin\left((x_m - P_{Teb}) \times \frac{2\pi}{365.25}\right) \tag{5.9}$$

where T'_{eb}, AT'_{eb}, and P_{Teb} are adjustable parameters. It is worthy to mention that, defined in this way, this boundary compartment could represent heat exchanges between river water and river bed, but also an additional exchange with the atmospheric compartment.

All in all, the DET model for temperature simulation using a deterministic approach for the non-advective terms had 13 adjustable parameters, including 4 for the non-advective terms (Table 5.1). DET+ model had 17 adjustable parameters, eight of them coming from the equations of the non-advective terms. The TEQ model is a convenient simplification of the temperature model, since only 12 adjustable parameters are required, three of them involved in the non-advective heat exchanges (Table 5.1).

5.5 Calibration strategy

We calibrated the three models of river temperature using a Marcov Chain Monte Carlo (MCMC) sampler entitled the Shuffled Complex Evolution Metropolis algorithm (SCEM-UA), which was specifically developed to deal with calibration of complex, non-linear models (Vrugt et al. 2003b). SCEM-UA is an evolutionary MCMC sampler that works in a Bayesian statistics framework to find the joint probability distribution of the model parameter values in the light of observed data. Thus, SCEM-UA does not give a unique parameter set that fits observed data. Instead, it draws a probability density function for each parameter, which expresses the parameter uncertainty. SCEM-UA methodology starts from a prior parameter distribution, usually consisting in a constant valued function defined inside realistic upper and lower bounds. In our case, we assigned wide prior parameter ranges (Table 5.1) to avoid being disturbed by artificially imposed boundaries. However, the range of some parameters is already defined in HSPF, and these limits were not exceeded. Then the algorithm tries to converge to the posterior target distribution using the calibration data, calculating the likelihood of each parameter set generated. This frequently implies thousands of model runs. SCEM-UA uses the scale reduction factor by Gelman and Rubin (1992) as a convergence criterion. Scale reduction factor for all parameters falling below 1.2 indicates convergence. However, we extended the calibration procedure well beyond this limit to assure a good

convergence to the true posterior distribution. Some utilities from the PEST package (Doherty 2004) were used to assist linking the Matlab SCEM-UA code from Vrugt et al. (2003b) to the HSPF modeling environment.

We used the measured average (*ADT*) and minimum (*MDT*) daily river temperature from January 2001 to December 2002 to calibrate the model. Data for 2003 and 2004 were left for the validation check.

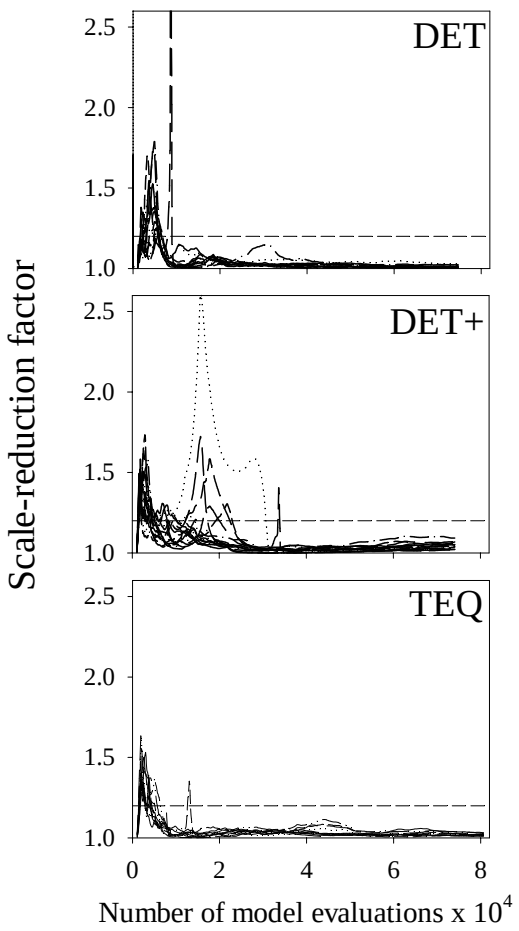

Figure 5.2 – Scale-reduction factor evolution for all parameters involved in the models tested during calibration with the SCEM-UA algorithm. The convergence criterion (i.e. 1.2) was added as a dashed line for reference.

In order to avoid a bias towards the high temperature summer values during the calibration step, we assigned weights for each residual to guarantee that all data had the same potential effect on the objective function during calibration. The weight for each residual was calculated as the inverse of the observed value. The objective function was defined as the sum of all weighted squared residuals between observed and modeled data.

Model performance was compared calculating the Root Mean Square Error (RMSE) and the Nash-Sutcliffe (NS) coefficient (Legates and McCabe 1999) for calibration and validation sets. We also calculated the Akaike's Information Criterion (AIC), a model selection procedure that accounts for model fitting performance but also for the number of adjustable parameters included in the model (Johnson and Omland 2004). AIC measures the information lost when assuming a model structure to explain a given data set, therefore the model with the smallest AIC is considered more convenient. Since AIC values are in a relative scale, the differences between AIC values and the smallest AIC value (ΔAIC) are usually reported. Thus, the larger the ΔAIC difference for a model, the less probable that it is the best model. Finally, Akaike (1983) suggested that the exp(-0.5 ΔAIC) approximates the relative likelihood of a model given the data, which are then normalized to obtain a positive set of Akaike weights (w) that sum 1.

5.6 Results

5.6.1 Models performance

The three calibration runs with the SCEM-UA algorithm satisfactorily converged to a posterior distribution for all parameters (Fig. 5.2). The scale-reduction factor for all parameters was well below the 1.2 criterion after 40 000 model runs. Since more than 70 000 model runs were performed during calibration of the three scenarios, we considered results from the final 10 000 model runs (i.e. the final 10 000 parameter combinations) a good approximation of the true posterior parameter distribution.

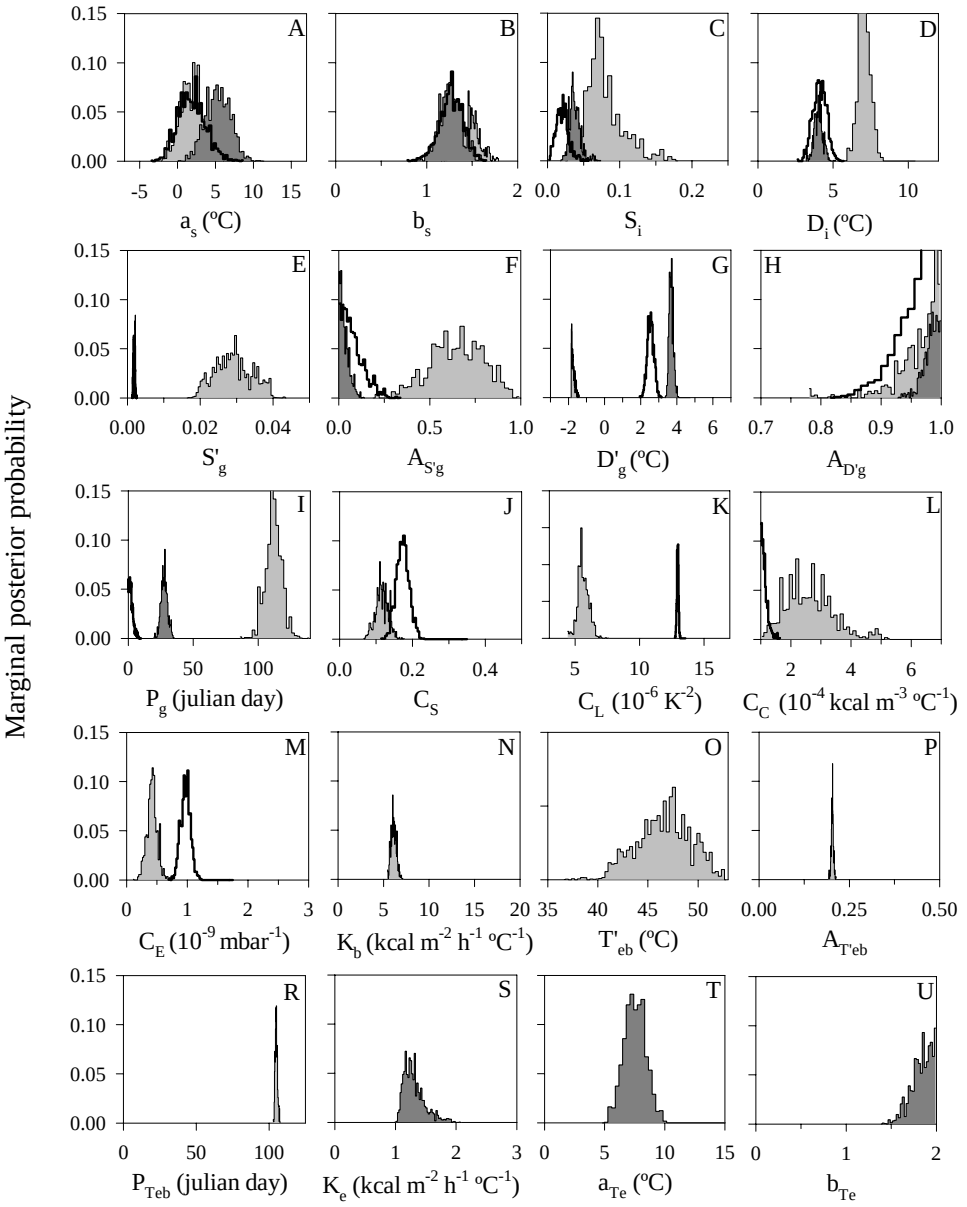

Figure 5.3 – Marginal posterior distribution for all parameters involved in the models tested during calibration with the SCEM-UA algorithm (TEQ: dark gray filled distributions; DET+: gray filled distributions; DET: no filled, bold lined). Distributions were constructed using results from the last 10 000 model runs. Panels A to I show distributions for parameters associated to advective heat exchanges (common to all models). Panels J to M show distributions for parameters associated to non-advective heat exchanges (common to DET and DET+ models). Panels N to R define parameter distributions for the boundary exchange process defined in model DET+. Finally, panels S to U refer to parameters associated to non-advective heat exchanges for model TEQ. Parameter definitions are as in Table 5.1.

Table 5.2 – Performance of tested models fitting average (*ADT*) and minimum (*MDT*) daily river water temperature during calibration and validation periods. The Akaike's information criterion and Akaike's weights calculated pooling both periods are also shown.

Model	Data series	Calibration period		Validation period		All data		
		RMSE (ºC)	NS	RMSE (ºC)	NS	AIC	ΔAIC	*w*
DET+	*ADT*	1.21	0.961	1.25	0.964	1088	0	1.0
	MDT	1.30	0.950	1.26	0.962			
TEQ	*ADT*	1.39	0.949	1.46	0.952	1744	657	0.0
	MDT	1.48	0.935	1.46	0.949			
DET	*ADT*	1.55	0.936	1.84	0.923	2329	1241	0.0
	MDT	1.74	0.910	1.69	0.931			

The distributions for all parameters involved in the three tested models are plotted in Figure 5.3. Comparing the range of posterior parameter distributions with the defined prior range (Table 5.1), it is clear that most parameters showed good sensitivity to the objective function, and most posterior parameter distributions were finally adjusted to a narrow part of the prior parameter range. Considering parameters involved in advective heat exchanges (Fig. 5.3, panels A to I), DET and TEQ models behave in a similar way, whereas the same parameters from DET+ model were adjusted to quite different values. Parameters shaping non-advecive heat exchanges shared by DET and DET+ (Fig. 5.3, panels J to M) were adjusted to very different values, despite the fact that these parameters control the same processes in both models.

Considering RMSE and NS values (Table 5.2) the three models were reasonably well calibrated and validated for both, minimum (*MDT*) and average (*ADT*) daily river water temperature. However, DET+ model performed best, and DET model gave the worst outcomes. This model performance ordination could not be related with the number of parameters defined in the different models. Despite being the richest parameterized model, DET+ had the best AIC, a criterion that accounts for and penalizes model complexity. The AIC value for DET+ results is so good compared to that of the other models that the associated Akaike weight (*w*) is one, suggesting that DET+ clearly outperforms the alternative model formulations even considering the high number of parameters involved in DET+.

Figure 5.4 shows daily average and minimum water temperatures against the range of values obtained with the last 1000 DET+ simulations, for both calibration and validation periods. The fit was very reasonable and no systematic biases could be detected. In addition, a comparison of seasonally detrended observations and results from TEQ and DET+ models (Fig. 5.5) showed that both temperature models explain a considerable amount of variance not related to the strong seasonal effect present in the original series. However, also in this case DET+ model gave better results.

The outperformance of DET+ respect TEQ model was also apparent when comparing the modeled and observed daily temperature range, expressed as monthly distributions of the daily range (Fig. 5.6). Noticeably, TEQ model did not adequately model the daily temperature range from May to September. Although not perfectly, DET+ model accounted for the seasonal trend in river temperature daily range, suggesting that DET+ outperforms TEQ not only at a seasonal or daily scale, but also at shorter time steps.

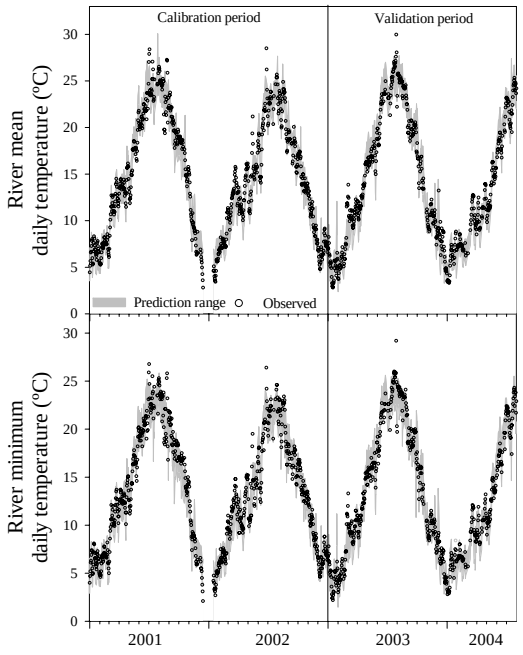

Figure 5.4 – Observed average and minimum daily water temperature values and prediction range for model DET+ (i.e. range of the last 1000 simulations with model DET+). Figure includes both calibration and validation periods.

Finally, we calculated the insertion depth of the river water in the downstream reservoir considering the reciprocal differences in water density (Armengol et al. 1999). Figure 5.7A shows the observed standardized insertion depth for dates with reservoir data available (approximately monthly frequency), and Figure 5.7B shows the hourly insertion depth for two consecutives days showing dramatic changes in insertion depth. Overlapped are the insertion depth calculated with DET+ and TEQ outcomes (i.e. the median value of the last 1000 simulations). Although daily calculations are quite similar (Figure 5.7A), DET+ gave better results (82% of explained variance against 63% with TEQ). For the hourly evolution (Figure 5.7B), DET+ is clearly superior. Although the initial and final conditions are similar, the timing of the major changes in river insertion was much better modeled with DET+.

5.6.2 Heat exchanges

Figure 5.8 details the dynamics of advective inputs temperature and of the non-advective heat exchanges for model TEQ. Temperature of the three advective inputs (Fig. 5.8A) showed a strong seasonal pattern, and as expected $T_{adv,g}$ evolution is much more smoothed than the more superficial fluxes.

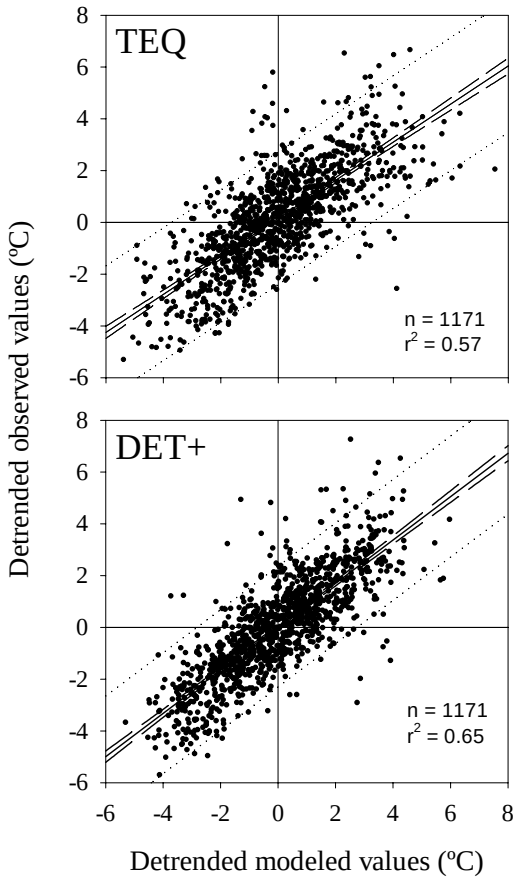

Figure 5.5 – Detrended modeled values against detrended observed values using results from TEQ and DET+ models. Median values from distributions in Figure 5.3 were used to generate model results. Regression lines and intervals are added for reference.

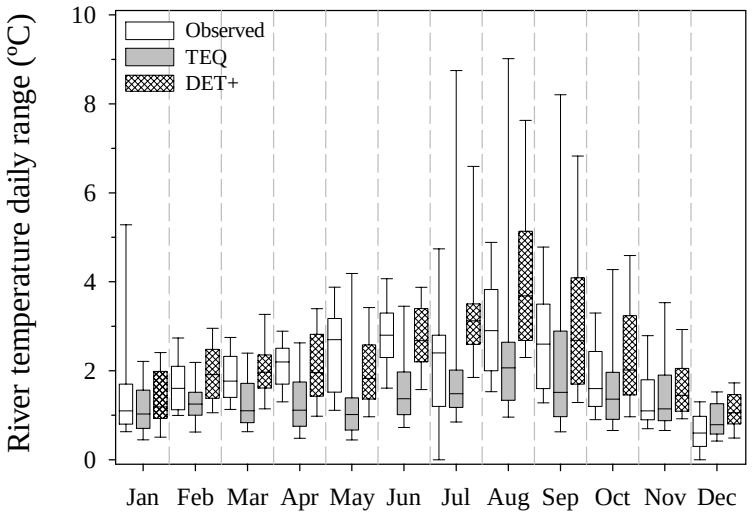

Figure 5.6 – Monthly box-whisker plot for observed daily river temperature range, and results obtained with DET+ and TEQ models. Median values from distributions in Figure 5.3 were used to generate model results.

However, a close inspection of Figure 5.8 reveals a problematic feature. $T_{adv,s}$ reaches remarkable high values during summer, mainly promoted by the high value for the intercept of the linear relationship with air temperature (a_S, Fig. 5.3A). Figure 5.8B shows the range of T_e calculated from the last 1000 model runs, and air and water temperature for reference. It is remarkable the high equilibrium temperatures reached during summer, caused by the high slope of the relationship with air temperature (Fig. 5.3U). The net non-advective heat exchange calculated with Equation 5.6 (Fig. 5.8C) ranges from near zero to 15 $MJ\ m^{-2}\ day^{-1}$, and shows the common seasonal pattern. However, maximum values during summer were relatively small considering that the study site is located in a Mediterranean region.

The different components of the non-advective heat exchange for the DET formulation are plotted in Figure 5.9A. The net long wave radiation flux seemed to be relatively high respect the short wave net radiation, mainly because the shading factor for solar radiation (C_S, Fig. 5.3J) removed around 80% of incident radiation measured in the meteorological stations, and because the proportionality constant for atmospheric long wave radiation (C_L, Fig. 5.3K) was optimized in a value higher than the typical figure (i.e. around $9 \times 10^{-6}\ K^{-2}$).

Figure 5.9B shows the very puzzling results obtained for the different components of the non-advective heat exchange for the DET+ model. If the physical meaning of results for heat exchanges in DET model were suspicious, there is no doubt that the different components of the DET+ model do not represent what was intended during model formulation. The heat exchange with the boundary compartment (Q_b) had non-realistic values that could not be attributed to a plausible bed exchange process. The same was true for the net long wave component, which toke very unrealistic negative values. The rest of the heat exchange components seemed to be small, especially Q_S and Q_E.

However, the total net exchange for both models (Figure 5.10) were quite comparable, suggesting that the different components of heat exchange in model DET+ should have some physical meaning. Remarkably, the sum of Q_S, Q_{COND}, and Q_{EV} components of DET+ depicted a rather stable seasonal cycle, while the sum of Q_b and Q_L accounted for most of the short-term variability, especially during

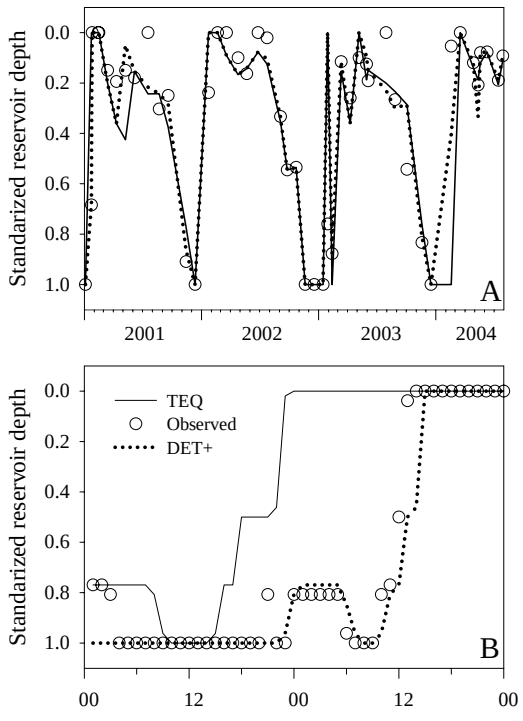

Figure 5.7 – Insertion depth of Ter River in Sau Reservoir, expressed as the standardized reservoir depth. (A) Monthly observed data and results using daily river temperature modeled with DET+ and TEQ. (B) Observed insertion depth from 23^{rd} January to 24^{th} January 2001, and results using hourly river temperature modeled with DET+ and TEQ.

summer months. This was stressed by the fact that the $Q_b + Q_L$ sum was linearly correlated with the total net heat exchange (n=2404, p-value < 0.0000, r^2=0.84).

5.7 Discussion

Results from the model performance comparison have an important implication for hydrological applications at the watershed scale. Even in the case that good, abundant, and spatially distributed meteorological data are available, pure deterministic approaches could not be the best option to accurately predict river water temperature. Actually, the only pure deterministic approach used in this study (DET model) showed the worst behavior. As we discuss below, we relate this misperformance with the fact that meteorological data collected at meteorological stations designed and placed to describe regional patterns could not adequately reflect the riverine meteorological conditions that ultimately drive heat exchange processes.

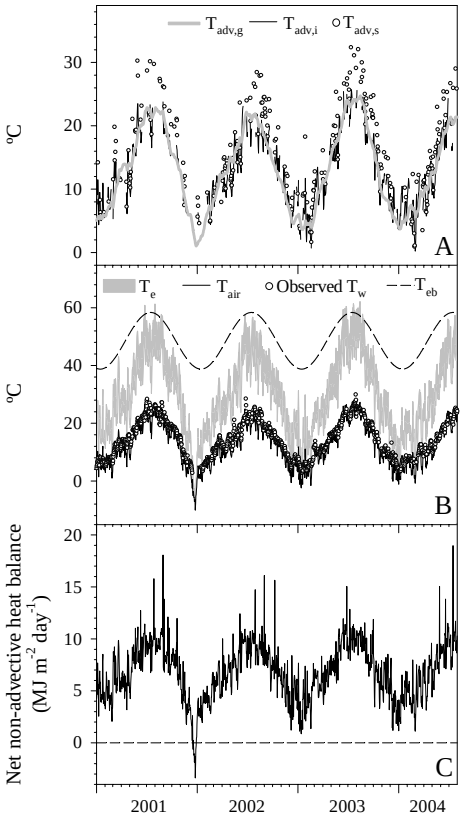

Figure 5.8 – Daily values during calibration and validation of the TEQ model for (A) temperature of advective inputs from groundwater, interflow, and surface runoff. (B) Air temperature measured near the sampling point, range of equilibrium temperature modeled, and observed water temperature. Additionally, this panel shows the evolution of the equilibrium temperature calculated for the boundary exchange process in DET+. (C) Net non-advective heat balance calculated with the TEQ model.

The empirical boundary heat exchange (Q_b) in DET+ is not working as an exchange processes with the streambed, but as an air-water exchange overlapped to the other defined atmospheric exchanges. In this sense, we can consider DET+ as a hybrid formulation, in which deterministic and empirical approximations combined to give the best outcomes. In fact, the evolution of T_{eb} (Fig. 5.8B) is quite coincident with that shown by T_e, especially during summer months. However we cannot relate this coincidence with a common physical interpretation of T_e and T_{eb}, because Q_b exchange defined through T_{eb} takes only sense in combination with Q_L. Although the empirical and deterministic formulations are inextricably linked in DET+, it seems that the calibration procedure has identified a model that depends on a seasonal defined equilibrium temperature modulated by processes fed with short term meteorological data (mainly Q_L, which strongly depends on air temperature).

On the other hand, the TEQ model works better than DET, suggesting that lumped processes dependent on air temperature are good options when true (i.e. measured over the river surface) riverine meteorological conditions are not available. However, a close inspection of TEQ results reveals some important limitations. The relationship between air temperature and equilibrium temperature showed very high slope (b_{Te} around 2, Fig. 5.3U), compared to other studies that found slopes around unity (Mohseni and Stefan 1999; Caissie et al. 2005). Although our values are not directly comparable because most studies relating air and equilibrium temperatures work with weekly mean temperatures, this high slope makes difficult to establish a direct link between TEQ formulation and the customary physical interpretation of T_e. In this sense, another fact argues against the physical meaning of the fitted T_e in TEQ. While the theoretical linear relationship between air and equilibrium temperatures strongly depends on the linear relationship between air and dew point temperature (Mohseni and Stefan 1999), dew point temperature in our watershed dramatically losses its linear dependence with air temperature at 19 $^{\circ}C$. Considering that air temperatures in our watershed are well above this value during summer, we

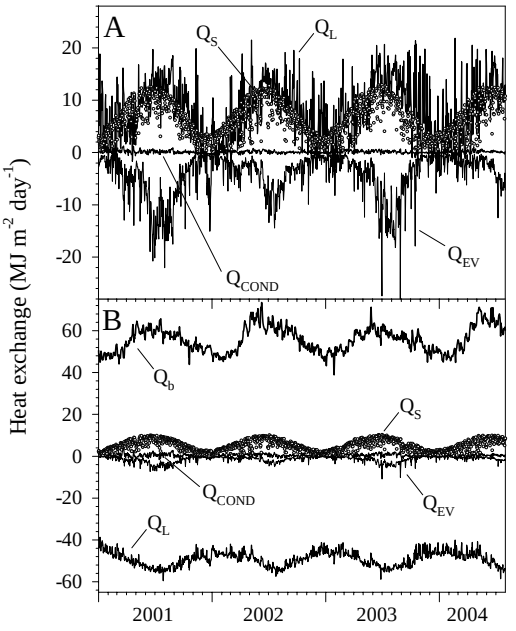

Figure 5.9 – Non-advective heat exchange components modeled with (A) DET model, and (B) DET+ model.

suspect that the high slope value obtained fitting the air versus equilibrium temperature relationship defined for TEQ model could be the result of this disagreement between model formulation assumptions and the real scenario in our watershed. In addition, the poor results obtained with the daily temperature range (Fig. 5.6) and the short term changes in river insertion depth (Fig. 5.7B), confirms that pure empirical approximations depending on air temperature cannot work properly at short time scales (Stefan and Preud'homme 1993). Therefore, despite the fact that TEQ model performance was superior to that for DET model, we cannot guarantee that the final fitted model is working in a physical-grounded way.

The hybrid DET+ formulation has proved to be the most successful, despite the fact that its physical meaning is uncertain. It combines deterministic formulations with wide prior parameter bounds and a seasonally defined equilibrium temperature concept. This combination seemed to avoid the prob-

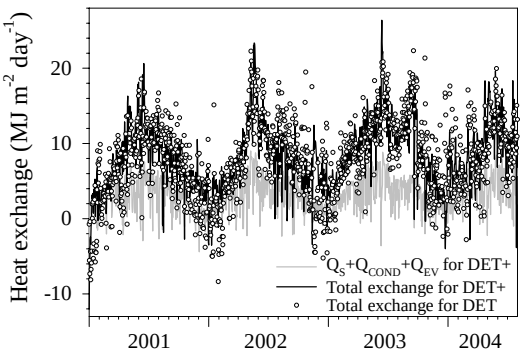

Figure 5.10 – Net non-advective heat exchange for models DET and DET+. A combination of several heat exchange components for model DET+ are also shown for reference.

lems encountered with the raw equilibrium temperature formulation in TEQ, especially for model results at time steps below one day. However, even recognizing the limitations of the TEQ formulation, this approximation could be used in case we are interested in a crude simulation of river water temperature at a daily scale. But in this case, care must be taken when interpreting results for the advective components of the model and the non-advective heat exchanges. The parameterization of the different formulations could be forced to adequately fit river water temperature (this could be the case of the advective temperature calculations for TEQ), but the reliability of the heat fluxes per se could be very low.

In a more general context, we interpret the results from this work as a warning against the imprudent use of meteorological data collected for regional characterization to feed temperature models at the watershed scale. Among the variables typically used to model river water temperature, air temperature seems to be the one most easily corrected for differences between data collected at the meteorological station and the conditions in the river. Thus, if only meteorological data from far locations are available, models including empirical or hybrid formulations driven by air temperature should be preferred.

Chapter 6

Scaling nutrient in-stream processes from the reach to the watershed using nutrient spiralling metrics

6.1 Introduction

Excess of man-induced nutrient loading into rivers has driven freshwater eutrophication (Vollenweider 1968; Heaney et al. 1992; Reynolds 1992) and degradation of coastal areas and resources at a global scale (Walsh 1991; Alexander et al. 2000; McIsaac et al. 2001). Thus, cultural eutrophication assessment and control are amongst the most important issues natural resource managers must face, markedly in watersheds under strong human impact. Although in most occasions control measures are based on bulk calculations of river nutrient loading (Marcé et al. 2004), the differentiation between point and non-point nutrient sources is of great importance in supporting water management decisions in polluted watersheds.

While the amount of nutrients loaded to rivers by point sources is easily calculated with effluent inventories, obtaining figures of non-point nutrient sources loading to running waters is a difficult topic. Non-point sources could be evaluated trough empirical approximations or with watershed-scale deterministic models. In the first case, and a part from crude mass-balance models (Howarth et al. 1996; Jaworski et al. 1992), information about land uses and farm and agricultural development in the watershed is combined with bibliographic or measured nutrient export coefficients to obtain annual or seasonal nutrient loads (Beaulac and Reckhow 1982). This method apply the reported yields (mass of nutrient per unit drainage area) from small, homogeneous watersheds to land types contained within larger heterogeneous watersheds. Thus, estimates are potentially imprecise and biased (Beaulac and Reckhow 1982). Note that some refinements have been developed, mainly for rivers in United Kingdom (Johnes 1996; Johnes et al. 1996; Johnes and Heathwaite 1997). Later, Smith et al. (1997) proposed a hybrid methodology (SPARROW) based on the empirical adjustment of the coefficients through non-linear regression with data measured in the watercourses. This approach has proven to be very efficient for studies at the annual time-step (Alexander et al. 2002).

By contrast, watershed-scale deterministic models can work at any time scale, and much effort has been paid to the design and development of this kind of models during the last decade (e.g. HSPF, Bicknell et al. 2001; SWAT, Srinivasan et al. 1993; INCA, Whitehead et al. 1998; AGNPS, Young et al. 1995). These models describe transport and loss processes in detail with mathematical formulations accounting for the spatial and temporal variations in sources and sinks in watersheds. The complexity of deterministic models often implies intensive data and calibration requirements, which generally limits their application in large watersheds. Deterministic models also often lack robust measures of uncertainty in model coefficients and predictions, although recent developments for hydrological applications are prone to be used in biogeochemical models as well (Raat et al. 2004).

However, the fundamental problem using watershed-scale models is the uncertainties involved in aggregating the components of fine-scale deterministic models in watershed applications (Rastetter et al. 1992) and in extrapolating the results of field-scale measurements to larger spatial scales. This is a very important topic, because in principle it is highly desirable using the knowledge gained through fine-scale studies (e.g. nutrient uptake rate for different river producers communities, nutrient fate in the food web, and so on) to predict coarse-scale phenomena (e.g. the overall nutrient export from watersheds). However, incorporating interactions between many components in a big-scale model could be cumbersome, simply because the number of possible interactions could be very large (Beven 1989). The usual strategy to avoid being seized by a model including precise formulations for each one interaction (and thus counting thousands parameters) is to lump components into aggregated units. But although lumping could efficiently reduce the number of parameters to a few tens, we still cannot guarantee that the information obtained during fine-scale research will apply to lumped categories. The behavior of an aggregate is not necessarily equivalent to the sum of the behaviors of the fine-scale components from which it is constituted (O'Neill and Rust 1979).

The working unit for the nutrient in-stream processes of most watershed-scale models is the reach. Within this topological unit, several formulations for biogeochemical reactions are included depending on the model complexity (e.g. adsorption mechanisms, algae nutrient uptake, benthic release, decomposition). However, if the quantification of these processes is not the main research target, a much more convenient in-stream model would consist in a reach-lumped formulation of stream nutrient consumption. If this consumption is empirically quantifiable at the reach scale, then we will be able to apply the field research to the model without the problems associated to upscaling results from fine-scale studies.

In the case of nutrients fate in streams, the *Nutrient Spiralling Concept* (Newbold et al. 1981) is a convenient simplification of the biogeochemical transformations involved, because the nutrient spiraling metrics are empirically evaluated at the reach scale (Stream Solute Workshop 1990). Within this framework, the fate of a molecule in a stream is described as an spiral length, that is the average distance a molecule travels to complete a cycle from the dissolved state in the water column, to a streambed compartment, and eventually back to the water column. The spiral length consist of two parts: the uptake length (S_w), which is the distance traveled in dissolved form, and the turnover length, which is the distance traveled as a particle. Usually, S_w is much longer than turnover length, and it focuses research based on the nutrient spiraling concept. S_w is evaluated at the reach scale, with nutrient enrichment experiments (Payn et al. 2005) or following nutrient decay downstream a point-source (Martí et al. 2004). Thus, if we include the formulation of nutrient spiralling metrics in the in-stream modules of a watershed-scale model, we will be able to use field-based research to parameterize the model.

In this chapter, we explore the possibility of using the information coming from empirical research based on the nutrient spiraling concept to feed the nutrient in-stream modules of a watershed-scale model. However, instead of directly incorporating the filed data on the model, we let a calibration al-

gorithm fit the model to observed data, and then we compared the adjusted parameters to equivalent information coming from field-based research. In this way, we are using the model as an heuristic platform working in two directions: we explore the possibility of using nutrient spiralling metrics in a watershed-scale model, but we also analyze some regularities observed in the nutrient spiraling metrics measured in impaired streams. Of course, this circular reasoning cannot validate any hypothesis, but considering that a model cannot be used for formal testing anyway (Oreskes et al. 1994), we considered this procedure a much more interesting exercise.

6.2 Materials and Methods

6.2.1 Study Site

We were interested in implementing a watershed-scale model for total phosphorus (TP) concentration in the Ter River watershed (Spain) from headwaters to Sau Reservoir (Fig. 6.1). We are considering 1380 km^2 of land with a mixture of land uses and vegetal covers. Headwaters are located in the Pyrenees ranges above 2000 *m.a.s.l.*, and run over igneous and metamorphic rocks covered by mountain shrubs communities and alpine meadows. Downstream, watercourses are surrounded by a mixture of conifer and deciduous forest, and sedimentary rocks become dominant. Then, Ter River enters the alluvial agricultural plain (400 *m.a.s.l.*) where non-irrigated crops dominate the landscape. Main Ter River tributaries are the Fresser River in the Pyrenees, the Gurri River in the agricultural plain, and Riera Major in the Sau Reservoir basin.

The Ter River watershed includes several urban settlements, specially in the agricultural plain (100 000 *inhabitants*). Industrial activity is also important, and this implies the existence on numerous phosphorus point-sources (Fig. 6.1A) coming from industrial spills and effluents from wastewater treatment plants (WWTP). Additionally, pig farming is an increasing activity, generating manures that are directly applied to the nearby crops as a fertilizer, at a rate of 200 $kg\ P\ ha^{-1}\ yr^{-1}$. Ter River long term median flow upstream Sau Reservoir at Roda de Ter (Fig. 6.1) is 10 $m^3\ s^{-1}$, and TP concentration frequently exceeds 0.2 $mg\ P\ L^{-1}$.

6.2.2 Modeling framework

The main target of the watershed-scale model was the prediction of daily TP river concentration at Roda de Ter (Fig. 6.1A). We used the Hydrological Simulation Program-Fortran (HSPF), a deterministic, semidistributed model that accounts for water routing in the watershed and water quality issues (Bicknell et al. 2001). Hydrology and river temperature have been already simulated and validated in the Ter River watershed using HSPF at the hourly time step (see *Chapter 4* and *5*). Figure 6.2 shows the simulated daily river streamflow and temperature against observations at Roda de Ter for sampling dates when river TP concentration were available. Note that not only streamflow, but also the contribution of surface and subsurface flows are simulated in HSPF. For simulations included in this *Chapter*, we will use the water routing and temperature river results obtained in *Chapter 4* and *5*. As a semidistributed application, HSPF splits the watershed in different subbasins, were equations are then solved. The split arrangement used in this work coincides with that shown in *Chapters 4* and *5*, and it is shown in Fig. 6.1A.

River TP data for this study comes from the Sau Reservoir long-term monitoring program, which includes a sampling point upstream the reservoir at Roda de Ter (Fig. 6.1A). Sampling was weekly to monthly, and spans from January 1999 to July 2004. In total 106 samples were analyzed with the alkaline persulfate oxidation method (Grasshoff et al. 1983). Data collected at Roda de Ter was the basic data used for calibration and validation of the river TP modules of HSPF. In addition, TP data coming from 14 sampling stations run by the local water agency (Agència Catalana de l'Aigua, ACA) were used as a supplementary set for model verification (Fig. 6.1A). The amount of data in those stations was highly variable, and the reliability of figures was dubious (e.g. precision only at the first deci-

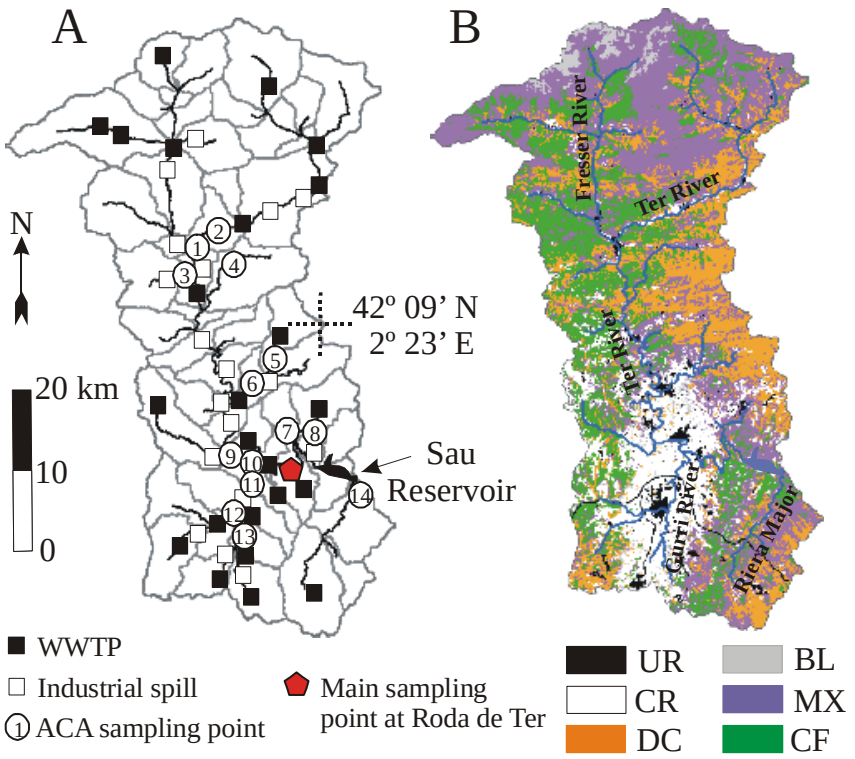

Figure 6.1 – (A) River TP sampling points and TP point sources in the Ter River watershed. Subbasins delineated for HSPF simulation are also shown. (B) Main watercourses and land uses in the catchment (UR: urban; CR: unirrigated crops; DC: deciduous forest; BL: barren land; MX: for clarity, meadows, shrublands, and few portions of oak forest are included here; CF: conifers forest).

mal in most occasions). Thus, we did not consider this information adequate for calibrating the model.

6.2.3 Point sources and diffuse inputs

TP concentration and water load information for point sources comes from ACA, and consist in a heterogeneous database with very detailed data for some spills, and crude annual values for others. For this, we decided to include in the model an adjustable multiplicative factor for WWTP inputs (C_w) and another for industrial spills (C_i), in order to correct for potential monotonous biases in the database (Table 6.1). Thus, daily TP load for one spill was the ACA value for the day times the correction factor.

Diffuse TP inputs to the watercourses were modeled using water routing results from *Chapter 4*. Since we were mainly interested in the in-stream processes, and in order to keep the model structure as simple as possible, we calibrated the model against river TP data collected during sampling dates with no surface runoff during at least seven days before sampling (i.e. the 106 samples mentioned above). Thus, we ignored TP transport in surface runoff. The TP concentration in interflow and groundwater flow was modeled assuming a power dilution dynamics. We modified the HSPF code to include the following formulations:

Table 6.1 – Prior ranges and final adjusted values during calibration of parameters used in the definition of the *TP* model. Equation numbers refer to equations in the text.

Description		Units	Range	SCE-UA value
In-stream TP decay				
V_f	Watershed scale uptake velocity (Eq. 6.4)	$m\ s^{-1}$	$2.8{\cdot}10^{-11} – 2.5{\cdot}10^{-5}$	$1.41{\cdot}10^{-6}$
TC	Temperature correction factor for TP decay (Eq. 6.4)	$^oC^{-1}$	1 – 2	1.06
Diffuse TP inputs				
b_i	Slope for TP concentration versus interflow discharge (Eq. 6.1)	mm^{-1}	0 – 1.8	0.56
a_i	Intercept for TP concentration versus interflow discharge (Eq. 6.1)	$mg\ P\ L^{-1}$	$3.5{\cdot}10^{-5} – 0.38$	0.002
b_g	Slope for TP concentration versus groundwater discharge (Eq. 6.1)	mm^{-1}	0 – 1.8	0.026
a_g	Intercept for TP concentration versus groundwater discharge (Eq. 6.1)	$mg\ P\ L^{-1}$	$3.5{\cdot}10^{-5} – 0.38$	0.05
Point-sources correction				
C_w	Correction factor for TP load from WWTP's	-	0 – 9	0.63
C_i	Correction factor for TP load from industrial spills	-	0 – 9	1.16

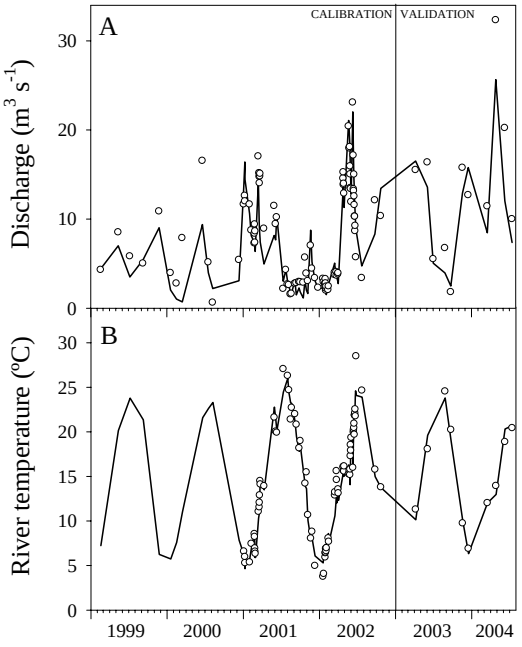

Figure 6.2 – (A) Observed (open circles) and modeled (line) discharge at Roda de Ter for *TP* sampling dates. (B) Observed (open circles) and modeled (line) mean daily river temperature at Roda de Ter for *TP* sampling dates. No observed data was available before 2001, see *Chapter 5*.

$$\begin{aligned} TP_i &= a_i \times Q_i^{\,b_i} \\ TP_g &= a_g \times Q_g^{\,b_g} \end{aligned} \qquad (6.1)$$

where TP_i and TP_g are *TP* concentration ($mg\ P\ L^{-1}$) in interflow and groundwater discharge, respectively. Q_i and Q_g are the interflow and groundwater discharge coming from the nearby land to the watercourse in *mm*. a_i, a_g, b_i, and b_g are adjustable parameters. Note that we did not consider spatial heterogeneity for these parameters. Thus, they should be considered as averages for the entire watershed. However, since river *TP* data for calibration comes from one sampling point, it is expectable that optimized parameter values will correspond more closely to the situation around this sampling point.

6.2.4 In-stream model definition

The in-stream *TP* fate was modeled as a first order decay following the Stream Solute Workshop (1990), and can be conceptualized as:

$$\begin{aligned}\frac{\partial TP}{\partial t} = &-\frac{Q}{A}\frac{\partial TP}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left[AD\frac{\partial TP}{\partial x}\right] \\ &+ \frac{Q_i}{A}(TP_i - TP) \\ &+ \frac{Q_g}{A}(TP_g - TP) \\ &- k_c\, TP\end{aligned} \tag{6.2}$$

where t is time (s), x is distance (m), Q is river discharge ($m^3\ s^{-1}$), A is river cross-sectional area (m^2), and k_c (s^{-1}) is an overall uptake rate coefficient. Q_i and Q_g are as in Equation 6.1 but expressed in $m^3\ s^{-1}$. The first term of the equation refers to advection, the second one to dispersion, and third and fourth to lateral subsurface inflows. All these terms are solved by the HSPF hydrologic and hydraulic modules.

The last term of Equation 6.2 simulates solute transfers between water column and benthic compartment. Of course, this represents an extreme simplified formulation, and must be interpreted as a net transport, because more complex settings accounts for independent dynamics of benthic release and concentration in one or more benthic compartments (Newbold et al. 1983). In fact, the formulation of solute fate including a first order decay formulated as in Equation 6.2 is one of the default options in the HSPF in-stream module. However, k_c describes solutes transfer in a volumetric basis, and applying a single value in a system with varying water depth could be very unrealistic. A much more convenient formulation considers solute transfers as a flux across the sediment/water interface, by means of a mass transfer coefficient (v_f, $m\ s^{-1}$):

$$\begin{aligned}\frac{\partial TP}{\partial t} = &-\frac{Q}{A}\frac{\partial TP}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left[AD\frac{\partial TP}{\partial x}\right] \\ &+ \frac{Q_i}{A}(TP_i - TP) \\ &+ \frac{Q_g}{A}(TP_g - TP) \\ &- \frac{v_f}{h}\, TP\end{aligned} \tag{6.3}$$

where h is river depth (m). Obviously, from this we can establish $v_f = h \times k_c$, which implies that v_f is a scale free parameter. As well, v_f more nearly describes abiotic transfers (Stream Solute Workshop 1990). We modified the HSPF code to incorporate this formulation in the in-stream modules, including also the built-in HSPF temperature correction factor. The final formulation was:

$$\begin{aligned}\frac{\partial TP}{\partial t} = &-\frac{Q}{A}\frac{\partial TP}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left[AD\frac{\partial TP}{\partial x}\right] \\ &+ \frac{Q_i}{A}(TP_i - TP) \\ &+ \frac{Q_g}{A}(TP_g - TP) \\ &- \frac{v_f\, TC^{(T_w-20)}}{h}\, TP\end{aligned} \tag{6.4}$$

where TC is the temperature correction factor and T_w ($^\circ C$) is river water temperature. Thus, the in-stream module of the watershed-scale model only included two adjustable parameters (Table 6.1).

v_f is related to the nutrient spiralling metric S_w trough the following relationship:

$$S_w = \frac{uh}{v_f} \tag{6.5}$$

where u is water velocity ($m\ s^{-1}$). Since nutrient uptake experiments in rivers and streams usually report S_w values for representative reaches, we can calibrate the watershed model with observed data and compare the obtained S_w with reported values from real systems (including data from Ter River watershed).

Some remarks concerning Equation 6.4. First, we are assuming that areal uptake rate ($U = v_f \times TP$) will be independent of TP concentration. Although a Monod function relating U and TP is usually applied for this purpose, high TP concentrations in Ter River watershed streams must be well established in the asymptotic section of the relationship (Mulholland et al. 1990). Thus, no conspicuous effect of TP on U was expected. Although this is not a realistic assumption for some pristine reaches in Ter

River headwaters and Riera Major, it most probably applies to reaches around the main sampling point, to which model calibration will be more sensitive. Finally, we are assuming a monotonous effect of temperature on solute transfer in the range of water temperatures measured in our streams.

6.2.5 Calibration strategy

We calibrated the 8 parameters model (Table 6.1) using TP data at Roda de Ter sampling point collected from 1999 to 2002. TP data for the period 2003-2004 were left for the validation check. However, since discharge in the river used during calibration was a modeled variable, we corrected the possible effect of errors in discharge simulation on modeled TP values. TP concentration in the river at Roda de Ter followed a dilution dynamics with discharge (Fig. 6.3), thus any mismatch between observed and modeled discharge will have a profound effect on the calibration process, especially at low discharge. To solve this problem, we performed calibration on a corrected TP observed series, using:

$$TP_c = TP \frac{TP'_{mod}}{TP'_{obs}} \qquad (6.6)$$

where TP_c is the corrected TP observed value. TP'_{mod} and TP'_{obs} are the TP values predicted by the power regression relating TP and discharge at Roda de Ter (Fig. 6.3), using the modeled and the observed discharge, respectively (Fig. 6.2A). The correcting quotient in Equation 6.6 averaged 1.09 for all TP data used during calibration.

Calibration was done automatically using the Shuffled Complex Evolution algorithm (SCE-UA), which was developed to deal with highly non-linear problems (Duan et al. 1992). From an initial population of random generated parameters, the algorithm uses shuffling, competitive evolution, and random search to efficiently find the parameter set that minimize an objective function. In this case, the sum of the squared errors between model outcomes and corresponding TP_c values. We performed the calibration run using SCE-UA as implemented in the PEST package (Doherty 2003), with parameter bounds detailed in Table 6.1. Convergence to an optimized parameter set was achieved after 7000 model runs.

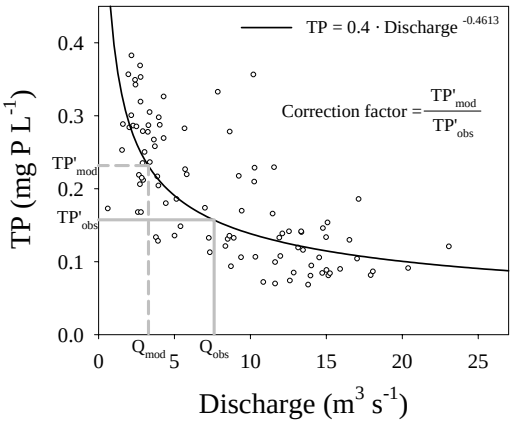

Figure 6.3 – Power relationship between TP concentration and discharge at Roda de Ter. A graphical example of calculation of the correction factor for TP is included for reference (*see* text for details). Q_{mod} and Q_{obs} represent arbitary modeled and observed discharges.

6.3 Results

Table 6.1 summarizes the optimized values for model parameters after calibration with the SCE-UA algorithm. Factors for point sources correction were different than one, suggesting that the available database had significant biases. The TP load from WWTP seemed to be overestimated in the database, while the industrial spills were slightly underestimated. Applying C_w and C_i for the mean annual TP loads we obtained 19 000 $kg\ P\ yr^{-1}$ from WWTP and 12 300 $kg\ P\ yr^{-1}$ from industrial spills. Considering the diffuse TP inputs, the power function fitted for the groundwater discharge TP concentration (Table 6.1) had a very mild slope, implying that TP_g was nearly a constant value in the range of Q_g measured in the Ter watershed (TP_g around 0.06 $mg\ P\ L^{-1}$). By contrast, the slope for the power relationship between TP_i and Q_i defined a clear dilution dynamics, with TP_i concentration

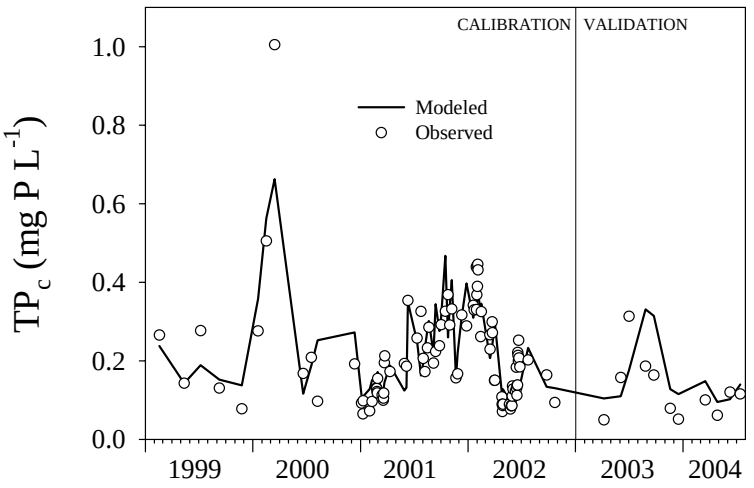

Figure 6.4 – Observed TP_c values and model outcomes at Roda de Ter during calibration and validation periods.

ranging from 0.6 to 0.04 $mg\ P\ L^{-1}$ depending on Q_i values. Using these power relationships with time series of Q_i and Q_g we obtained mean annual TP loads of 23 600 $kg\ P\ yr^{-1}$ from groundwater discharge and 12 800 $kg\ P\ yr^{-1}$ from interflow discharge.

The watershed mass transfer coefficient v_f was fitted to a very low value (Table 6.1), only comparable with values obtained in point-sources impaired streams (Doyle et al. 2003; Martí et al. 2004). Values from pristine streams usually falls between 10^{-3} and 10^{-5} $m\ s^{-1}$ (Doyle et al. 2003). Our low v_f defines a watershed with watercourses with very low phosphorus retention capacity. Of course, this would most probably hold in reaches around the sampling point at Roda de Ter, while in headwater streams the value will most probably be underestimated. Thus, we must take this v_f figure as a coarse-scale value. On the other hand, temperature correction factor (TC, Table 6.1) was adjusted to 1.06. Considering that mean daily river water temperature in the watershed ranges from 5 to 27 $^{\circ}C$ (Fig. 6.2), this implies that v_f values were multiplied by a factor (Equation 6.4) that ranged from 0.4 to 1.3. Thus, effective v_f values after temperature correction ganged between 5.6×10^{-7} and 1.8×10^{-6} $m\ s^{-1}$. This significant dependence on water temperature suggest that v_f for TP in this watershed is controlled to a some extent by biological activity.

The fit between observed data and model outcomes at Roda de Ter was very satisfactory (Fig. 6.4). The model explained the 72% of variance in river TP_c values during the calibration period (the contribution of the very high value during year 2000 was modest. Without this point the explained variance amounted 69%). However, the model performed worse during high flow conditions (or low TP concentrations), as Figure 6.5 clearly shows. This is most evident during the validation period, a very wet period (Fig. 6.2). Thus, it seems that the model is missing some significant effect at high flows, that could be attributed to physically-mediated higher retention during high flows not accounted for in our formulation, or to an overestimation of TP_g during very wet periods. Unfortunately, data from this study did not allow an accurate evaluation of this possibility.

In addition, the fit between median TP values coming from ACA stations and model results was good (Fig. 6.6), although ACA station 7 showed values that were considerably higher than model outcomes. Considering that adjusted v_f value represent a very low value, this most probably implies a

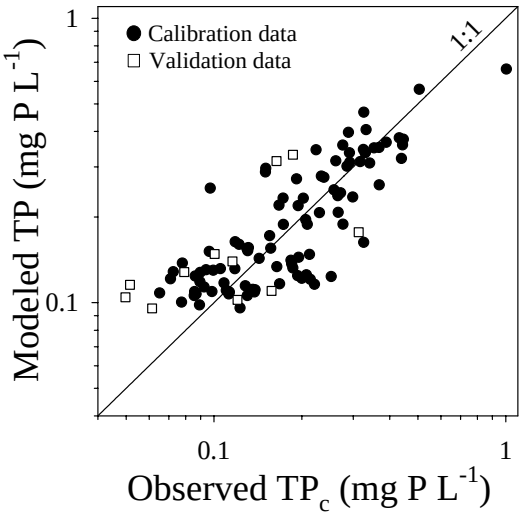

Figure 6.5 – Observed versus modeled TP_c values at Roda de Ter during calibration and validation periods.

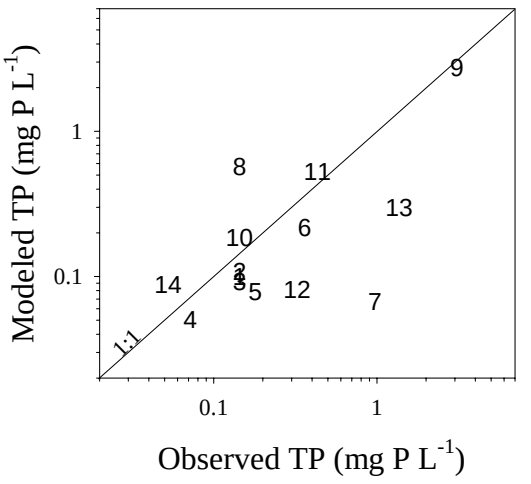

Figure 6.6 – Median TP values observed in the different ACA sampling stations against modeled values (numbers as in Fig. 6.1A).

missing point source in the database upstream this sampling point.

6.4 Discussion

Results from this study showed that the basic formulation from which the *Nutrient Spiralling Concept* research is based is a good alternative to model the nutrient in-stream processes in watershed-scale models. Even considering that we worked in a worst case scenario, in the sense that limited river TP concentration data were available to calibrate the model, model outcomes were satisfactory, and final parameter values realistic.

These results pose the following question: can we use field estimations of nutrient spiralling metrics to feed our model? Of course, the best method to test this possibility were to measure S_w (and calculating v_f with Equation 6.5) in several reaches in the Ter watershed, and then compare this with our estimate. But this is beyond the scope of this work, and published data in the Ter watershed reports nutrient spiralling metrics mostly for pristine streams (Martí and Sabater 1996; Butturini and Sabater 1998). However, mean v_f for two nutrient retention experiments in a reach in the impaired Riera de Tona (Gurri River tributary, Fig. 6.1B) were 4.6×10^{-6} ms^{-1} (Martí et al 2004), which is an astonishing similar figure compared to our adjusted reference value (Table 6.1). In fact, using Martí's empirical value in our model only promoted a modest deviation in the model results (66% of TP_c explained variance compared to 72% with the optimized parameter).

Is this result just a fortunate coincidence or there is some regularity for nutrient retention metrics in nutrient enriched streams? During recent years, researchers have accumulated data that suggest nutrient enriched streams have lower retention efficiency (i.e, lower v_f or higher S_w) than pristine streams (Doyle et al. 2003; Martí et al 2004; Haggard et al. 2005; Merseburger et al. 2005; Gücker and Pusch 2006; Ruggiero et al. 2006). To test how our model results fit into this picture, we collected S_w results for phosphorus (for many studies v_f results were not available) from pristine and nutrient enriched streams (Table 6.2). Note that these results come from very heterogeneous field procedures (nutrient additions, nutrient decay downstream a point source, isotopic tracers), and that

Table 6.2 – S_w and discharge for different nutrient retention experiments in pristine and impaired streams.

	System	Discharge (m^3 s^{-1})	S_w (m)	Source
	Pristine streams			
1	Riera Major (Spain)	0.0544	300	Butturini and Sabater 1998
2	Pine Stream (USA)	0.0021	49	D'Angelo and Webster 1991
3	Hardwood Stream (USA)	0.0025	31	D'Angelo and Webster 1991
4	Pioneer Creek (USA)	0.0856	370	Davis and Minshall 1999
5	Bear Brook (USA)	0.0145	49	Hall et al. 2002
6	Cone Pond outlet (USA)	0.0023	8	Hall et al. 2002
7	Hubbard Brook (USA)	0.0866	85	Hall et al. 2002
8	Paradise Brook (USA)	0.0067	29	Hall et al. 2002
9	W2 stream (USA)	0.0011	6	Hall et al. 2002
10	W3 stream (USA)	0.0069	22	Hall et al. 2002
11	W4 stream (USA)	0.0042	14	Hall et al. 2002
12	W5 stream (USA)	0.0016	19	Hall et al. 2002
13	W6 stream (USA)	0.0027	15	Hall et al. 2002
14	West Inlet to Mirror Lake (USA)	0.0010	12	Hall et al. 2002
15	Myrtle Creek (Australia)	0.0049	76	Hart et al. 1992
16	Montesina Stream (Spain)	0.0019	8	Maltchik et al. 1994
17	Riera Major (Spain)	0.0578	177	Martí and Sabater 1996
18	La Solana Stream (Spain)	0.0207	89	Martí and Sabater 1996
19	West Fork (USA)	0.0042	65	Mulholland et al. 1985
20	Walter Branch (USA)	0.0060	167	Mulholland et al. 1990
21	Watershed 2, Oregon (USA)	0.0010	697	Munn and Meyer 1990
22	Hugh White Creek (USA)	0.0040	85	Munn and Meyer 1990
23	Coweeta Stream (USA)	0.0022	9	Newbold 1987
24	Sturgeon River (USA)	1.2600	1400	Newbold 1987
25	West Fork, 1st order (USA)	0.0042	165	Newbold 1987
26	West Fork, 2nd order (USA)	0.0310	213	Newbold 1987
27	West Fork (USA)	0.0046	190	Newbold et al. 1983
28	Barbours Stream (New Zealand)	0.0450	289	Niyogi et al. 2004
29	Kye Burn Stream (New Zealand)	0.0240	388	Niyogi et al. 2004
30	Stony Stream (New Zealand)	0.0700	266	Niyogi et al. 2004
31	Sutton Stream (New Zealand)	0.0530	872	Niyogi et al. 2004
32	Lee Stream (New Zealand)	0.0710	240	Niyogi et al. 2004
33	Broad Stream (New Zealand)	0.1550	920	Niyogi et al. 2004
34	Dempsters Stream (New Zealand)	0.0290	669	Niyogi et al. 2004
35	Kuparuk River (Alaska)	1.3500	2955	Peterson et al. 1993
36	East Kye Burn (New Zealand)	0.0150	94	Simon et al. 2005
37	North Kye Burn (New Zealand)	0.0230	222	Simon et al. 2005
38	JK1-JK3 streams (USA)	0.0082	42	Valett et al. 2002
39	SR1-SR3 streams (USA)	0.0052	87	Valett et al. 2002
40	Cunningham Creek (USA)	0.0097	104	Wallace et al. 1995
41	Cunningham Creek after logging (USA)	0.0252	47	Wallace et al. 1995
42	Hugh White Creek (USA)	0.0190	30	Webster et al. 1991
43	Sawmill Branch (USA)	0.0025	32	Webster et al. 1991
44	Big Hurricane Branch (USA)	0.0177	31	Webster et al. 1991
	Point-source-enriched streams			
a	Koshkonong River with dam (USA)	6.2107	57449	Doyle et al. 2003
b	Koshkonong River without dam (USA)	12.7500	188115	Doyle et al. 2003
c	Demmitzer Mill Brook (Germany)	0.0220	4144	Gücker and Pusch 2006
d	Erpe Brook (Germany)	0.5110	5539	Gücker and Pusch 2006
e	Columbia Hollow (USA)	0.1183	8667	Haggard et al. 2005
f	Fosso Bagnatore (Italy)	0.0099	3480	Ruggiero et al. 2006
g	Daró Stream (Spain)	0.0460	3510	Martí et al. 2004
h	Riera de Tenes (Spain)	0.0045	2080	Martí et al. 2004
i	Riera de Berga (Spain)	0.0710	14250	Martí et al. 2004
j	Riera d'en Pujades (Spain)	0.0180	170	Martí et al. 2004
k	Riera de Tona (Spain)	0.0305	7550	Martí et al. 2004
l	Ondara Stream (Spain)	0.0600	2560	Martí et al. 2004
m	Verneda Stream (Spain)	0.0250	3200	Martí et al. 2004
n	Riera de Figueres (Spain)	0.1630	250	Martí et al. 2004
o	Passerell Stream (Spain)	0.0120	4790	Martí et al. 2004
p	Barrenys Stream (Spain)	0.1500	2490	Martí et al. 2004
q	Negre Stream (Spain)	0.0220	2120	Martí et al. 2004
r	Salat Stream (Spain)	0.0530	50	Martí et al. 2004
s	Riera d'Osor (Spain)	0.0310	2850	Martí et al. 2004
t	Llobregat de la Muga (Spain)	0.0470	3740	Martí et al. 2004

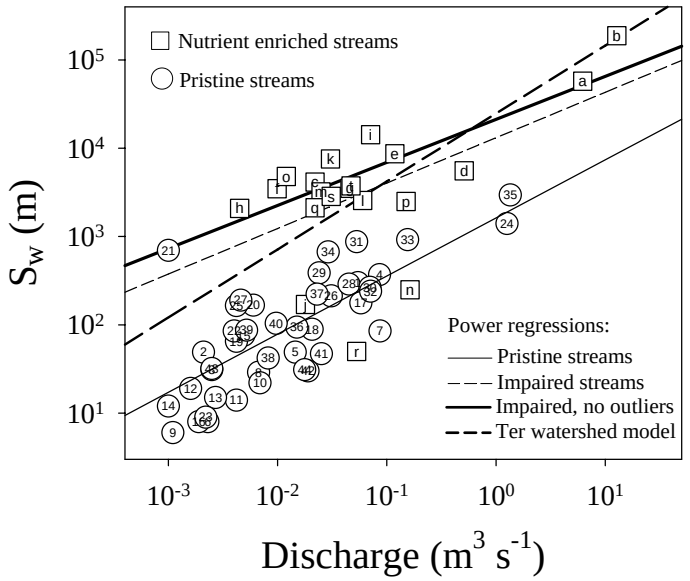

Figure 6.7 – Discharge versus S_w for pristine and nutrient enriched streams. Numers and letters as in Table 6.2. See the text for power regressions details.

lump seasonal studies with one-measure data, and habitat specific experiments with whole stream determinations. In despite of this, a clear relationship could be established between S_w values and discharge (Fig. 6.7), a result already reported for phosphorus (Butturini and Sabater 1998) and ammonia retention (Peterson et al. 2001) in pristine streams. Our fitted power relationship between S_w and discharge ($1622\ Q^{0.65}$, $n = 44$, p-value < 0.0001, $r^2 = 0.56$) differed slightly from equation reported by Butturini and Sabater (1998), because our database includes recent data. However, the most interesting fact in Figure 6.7 is that a significant power relationship was also fitted with data coming from nutrient-enriched streams ($13\ 163\ Q^{0.51}$, $n = 20$, p-value < 0.0097, $r^2 = 0.32$). Thus, the lack of relationship between phosphorus S_w and discharge reported in some studies (Martí et al. 2004) should be attributed to a narrow discharge range. In fact, the relationship between S_w and discharge is highly expectable from Equation 6.6 (Stream Solute Workshop 1990).

Now we can plot the power relationship obtained transforming the adjusted v_f value in the Ter watershed to S_w with Equation 6.6. This is the dashed bold line in Figure 6.7, and corresponds to the equation $24\ 742\ Q^{0.77}$. Note that for comparisons between the different power regressions, the adequate parameter is the intercept of the power regression, because the slope will depend on the geomorphologic traits of the rivers included in each relationship (Stream Solute Workshop 1990). Taking this in mind, power regressions for Ter River watershed and for impaired streams were very similar, specially is we do not consider points labeled as *j*, *r*, and *n* to calculate the regression for nutrient enriched rivers ($21\ 256\ Q^{0.49}$, $n = 17$, p-value < 0.0001, $r^2 = 0.73$, bold line in Fig. 6.7). In fact, it is quite usual to find very short or even negative values for phosphorus S_w in nutrient enriched streams (Martí et al 2004; Merseburger et al. 2005), and it is certainly difficult to assign this retention or release to actual in-stream processes or to lateral inflows of nutrients by seepage, because in impaired streams point sources are usually associated to diffuse inputs (Merseburger et al. 2005).

The results from Figure 6.7 could be interpreted in two ways. First, retention efficiency greatly diminishes in nutrient enriched streams, and the variability between impaired streams is not so high to prevent assigning a range of typical v_f values for these kind of systems. Then, the second interpretation follows: we can use data from empirical studies of nutrient retention to parameterize a watershed-scale model. Obviously, this is a circular reasoning, and the model results cannot be used to state that we demonstrated what the above interpretations imply. But this is a nice example of how models can work as heuristic tools to compare hypothesis and stimulate research (*see* Oreskes et al. 1994).

To conclude, we have demonstrated that a lumped, hardly parameterized formulation of the in-stream nutrient fate in rivers could be very efficient in a large-scale model, and that this opens the very interesting possibility of directly using data collected at the field into large-scale models. This avoids the exercise of upscaling fine-scale research results to parameterize do-everything models with many parameters, many of them finally adjusted to bibliographical values if no adequate field data are available. Of course, this is not a valid option if the detailed biogeochemical processes in the stream are research targets, or if we need explicit formulations of these processes to simulate complex biotic or abiotic interactions. However, the coarse-formulation approach should suffice in many modeling exercises.

General conclusions

GENERAL CONCLUSION 1. Some features of the eutrophication process in Sau Reservoir (i.e. hypolimnetic oxygen content and phosphorus internal load) are directly linked to the Ter River water quality. Despite the assumptions of classical empirical models, any epilimnetic process play a significant role. This is not a system-specific result, but a fundamental difference between advective-dominated systems and most lakes.

Concerning the neuro-fuzzy technique for load calculations. This new method for river load calculations performs better than classical methodologies, avoiding the major drawbacks associated with their use. Also, the new method has shown some interesting properties that make it a suitable exploratory data analysis tool for non-linear problems.

Concerning hypolimnetic chemistry in Sau Reservoir. The oxygen and phosphorus content in this layer depends on dissolved organic matter and nitrate carried by Ter River, with no significant intervention of epilimnetic processes.

Concerning hypolimnetic oxygen content in reservoirs. The oxygen content in the hypolimnion of reservoirs is greatly influenced by dissolved organic carbon carried by the river. Since residence time in reservoirs is short, the effects of allochthonous organic matter will be more conspicuous in systems receiving large amounts of labile, human-derived dissolved organic carbon. These results should motivate the formulation of a new paradigm for the empirical modeling of the oxygen content and resilience to eutrophy in reservoirs.

GENERAL CONCLUSION 2. The state-of-the-art and the available information for rivers and watersheds in Catalonia suffice to apply complex hydrological and biogeochemical watershed-scale models. The new calibration methodologies and data-management strategies allow the implementation of such complex applications even considering field data scarcity.

Concerning the hydrological simulation. A proper combination of model structure definition and multiobjective calibration is essential to implement semidistributed watershed-scale hydrological models. If calibration is carefully done, even some spatial heterogeneity could be saved for some parameters. However, in a complex modeling scenario it could be very difficult to establish a priori which field data will be more useful during the calibration process.

Concerning the river temperature simulation. When data from meteorological stations located far from the river reaches are the only input available, empirical or hybrid models including formulations relating air and water temperature should be preferred.

Concerning the total phosphorus simulation in the river. The empirical information collected at the reach scale following methodologies based in the nutrient spiraling concept is a very interesting option to feed models at the watershed scale. Although the modeling exercises presented in this dissertation are not susceptible to proof hypotheses, our results support the view that sewage-impacted rivers are less efficient retaining phosphorus than pristine streams.