

# PARAMETER AND STATE REGULARIZATION FOR PREDICTION OF DISTRIBUTED HYDROLOGIC SYSTEMS

E.E. van Loon\* K.J. Keesman\*\*,<sup>1</sup>

\* *Hydrology and Quantitative Water Management Group,  
Wageningen University*

\*\* *Systems and Control Group, Wageningen University*

Abstract: State and parameter regularization schemes are applied to a distributed hydrologic system for discharge prediction at different locations in the catchment. Experimental field data from a 200 ha catchment in Costa Rica have been used to evaluate the prediction method. It appears that, in general, state regularization leads to better predictions at finer resolutions than parameter regularization, and parameter regularization to better performance at coarse resolutions. ©IFAC 2003

Keywords: State/parameter estimation, regularization, hydrologic systems

## 1. INTRODUCTION

In general the identification problem for distributed hydrologic systems focuses on a single catchment and a limited observation period. For a hydrological model to be a useful tool in planning, it should be possible to apply that same model to other locations or time periods without redoing the full identification step or having to collect large amounts of observations. Another reason for not changing the model structure at different situations is to keep model results compatible with earlier results and therewith reasonably straightforward to interpret. On the other hand, it has become clear from numerous studies that a certain amount of re-calibration is always required in catchment modelling. The question remains however, which parameters in the model are best suited for this purpose, how to determine the parameter values from data and at which resolution to define the model. In view of previous work on model (re-)calibration, especially the last question is interesting since the choice for the resolution at

which a hydrological problem is defined is often not discussed nor explained. This is a remarkable situation, since a simple count of variables in any realistic hydrologic problem shows that the available observations alone cannot contain sufficient information to determine a distributed-parameter model to a reasonable degree, even not if the problem is limited to the re-calibration of a few parameters. Hence, the problem is *ill-posed*. When seeking a way to identify a model in face of ill-posedness, a common approach in many engineering disciplines is the development of an alternative model structure, with fewer degrees of freedom and a more suitable parameterization. However, as pointed out above, in hydrological problems the change of model structure is often undesirable since it would render the model incompatible with other applications or make the model results difficult to interpret. Therefore the problem is usually handled by using additional information, in the form of assumptions about parameter values and relations between parameters. Here the aim is now to provide such a follow-up by considering a *regularization* approach to combine a set of calibrated catchment scale models of over-

---

<sup>1</sup> Corresponding author.

E-mail address: karel.keesman@wur.nl

land flow with additional observations. Section 2 presents the regularization approach to solve the estimation problem for a general class of ill-posed time-varying linear models of overland flow. In Section 3 regularization is applied to predict overland flow in the study catchment. This is followed by an inter-comparison of various ways to re-calibrate models in Section 4. The implications of the results are discussed and conclusions are drawn in Section 5.

## 2. CALIBRATING A HYDROLOGICAL SYSTEM THROUGH REGULARIZATION

### *Conversion into standard linear time varying form*

It has been shown by van Loon and Keesman [2000] that a large class of hydrological systems, which vary in time and space and are often non-linear, can be represented by a linear time-varying state-space model of the following form

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{B}_k \mathbf{C}_k \mathbf{x}_{k-1} + \mathbf{A}_k \mathbf{u}_k \quad (1)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{e}_k \quad (2)$$

where  $\mathbf{x}_k$  contains  $M$  state variables for each of the  $L$  spatial units and the input vector  $\mathbf{u}_k$  contains at least  $L$  elements at time instant  $k$  (but  $\mathbf{u}_k$  may be larger, depending on the structure of  $\mathbf{A}_k$  and the amount of input variables). The structured matrices  $\mathbf{A}_k$ ,  $\mathbf{B}_k$  and  $\mathbf{C}_k$  (transition matrices) contain time-varying coefficients  $\theta_k$  which are stochastic functions of  $\mathbf{x}_{k-1}$  or  $\mathbf{u}_k$  and can only take values between zero and one ( $0 \leq \theta_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) \leq 1$ ). In what follows the states are the stored soil water ( $s_{k,l}$ ), the overland flow due to infiltration excess ( $r_{k,l}$ ), the overland flow due to saturation excess ( $t_{k,l}$ ) and the cumulative soil moisture content ( $w_{k,l}$ ) all in mm. The effective rainfall ( $p_{k,l}$ ), that is the difference between rain and evapotranspiration, is the input to the hydrologic system. The vector  $\mathbf{y}_k$  contains  $P$  observations and the matrix  $\mathbf{H}_k$  is an observation matrix which relates the state variables to the observations, and the output error vector  $\mathbf{e}_k$  contains modelling as well as observation errors. In the following, the focus is on either on-line estimating the unknown coefficients in the transition matrices or estimating the state variables at time instant  $k$  from observations available at time instant  $k$  (parameter and state estimation respectively).

To apply the estimation techniques the above equations are converted to the following standard regression form

$$\mathbf{y}_k = \mathbf{D}_k \mathbf{m}_k + \mathbf{e}_k \quad (3)$$

where  $\mathbf{y}_k$  is a vector with measured values,  $\mathbf{D}_k$  the design or data matrix with known values,  $\mathbf{m}_k$  a vector with values to be estimated, and  $\mathbf{e}_k$  a noise vector. The way in which (1) and (2) have to be transformed to obtain (3) depends on whether we consider *parameter* or *state estimation*.

For *parameter estimation* substitute (1) into (2), so that

$$\mathbf{y}_k = [\mathbf{H}_k \mathbf{A}_k \mathbf{B}_k \mathbf{C}_k] \mathbf{x}_{k-1} + [\mathbf{H}_k \mathbf{A}_k] \mathbf{u}_k + \mathbf{e}_k \quad (4)$$

where  $\mathbf{x}_{k-1}$  will be substituted by its estimate  $\hat{\mathbf{x}}_{k-1}$ , which is common practice in nonlinear parameter estimation problems with output error structure. In what follows  $\mathbf{x}_{k-1}$  is estimated using (1)

$$\hat{\mathbf{x}}_{k-1} = \mathbf{A}_{k-1} \mathbf{B}_{k-1} \mathbf{C}_{k-1} \hat{\mathbf{x}}_{k-2} + \mathbf{A}_{k-1} \mathbf{u}_{k-1} \quad (5)$$

and  $\mathbf{x}_0$  is assumed to be known.

Hence, the known vectors  $\hat{\mathbf{x}}_{k-1}$  and  $\mathbf{u}_k$  are put into the data matrix  $\mathbf{D}_k$ , and all the unknown parameters in  $\mathbf{A}_k$ ,  $\mathbf{B}_k$ ,  $\mathbf{C}_k$  and  $\mathbf{H}_k$  are put in the parameter vector  $\mathbf{m}_k$ . A detailed derivation is given in Appendix C of van Loon [2001]. Consequently, (4) can be rewritten as (3) with observation vector  $\mathbf{y}_k \doteq [y_k(1) \ y_k(2) \ \dots \ y_k(P)]^T$ ,  $\mathbf{D}$  the design matrix which embodies the geometry as well as inputs to the system, and  $\mathbf{m}_k \doteq [\theta_k(1) \ \theta_k(2) \ \dots \ \theta_k(Q)]^T$  with  $Q = MPL + PL$  the size of the model parameter vector. Besides the  $P$  observations there are some additional constraints. These arise from the fact that the columns of the transition matrices in the original model (1) sum to unity. These constraints form a set of additional linear equations

$$\mathbf{y}_{cons,k} = \mathbf{D}_{cons,k} \mathbf{m}_k \quad (6)$$

Even with these additional hard constraints there are, for any realistic distributed hydrological problem, relatively few observations available compared to the number of model parameters, so that information is lacking to determine uniquely all the model parameters, i.e. the problem has an almost singular regression matrix since  $\mathbf{D}_k$  has many more columns than rows.

When considering *state estimation* we propose to rewrite (1) and (2) as follows,

$$\mathbf{y}_k - [\mathbf{H}_k \mathbf{A}_k] \mathbf{u}_k = [\mathbf{H}_k \mathbf{A}_k \mathbf{B}_k \mathbf{C}_k] \mathbf{x}_{k-1} + \mathbf{e}_k \quad (7)$$

which again can be written into the regression form of (3), where now  $\mathbf{m}_k$  is defined as  $\mathbf{m}_k = [x_{k-1}(1) \ \dots \ x_{k-1}(4L)]$ . Similar to (6) equality constraints on the states can be formulated, e.g. derived from the global mass balance

$$\sum_{l=1}^L (s_{k,l} + r_{k,l} + t_{k,l} - p_{k,l}) = 0 \quad (8)$$

$\forall k = 1, \dots, K$  or after some manipulation (see Appendix A in van Loon and Keesman [2000])

$$\mathbf{v}^T [\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}] = 0 \quad (9)$$

with  $\mathbf{v} \doteq [1(1) \dots 1(3L) 0(1) \dots 0(L)]^T$  and  $\hat{\mathbf{x}}_{k-1} \doteq \mathbf{A}_{k-1}\mathbf{B}_{k-1}\mathbf{C}_{k-1}\hat{\mathbf{x}}_{k-2} - \mathbf{A}_{k-1}\mathbf{u}_{k-1}$ . Note that the vector  $\mathbf{v}$ , comprising a  $(1, 3L)$  sub-vector with ones and a  $(1, L)$  sub-vector with zeros, simply allows a summation of the first three variables (each defined at  $L$  spatial units) of  $\mathbf{x}$ , viz.  $\mathbf{s}$ ,  $\mathbf{r}$ , and  $\mathbf{t}$ . By defining  $\mathbf{D}_{cons,k} = \mathbf{v}^T$  and  $\mathbf{y}_{cons,k} = \mathbf{v}^T\hat{\mathbf{x}}_{k-1}$ , (9) can be written in the form of (6). Notice that in this approach the dynamic hydrologic system equations are added as equality constraints, see (9), to the regression-type model (3). A more conventional to the aforementioned problems would be based on observer-theory where design and updating of the gain matrix plays a central role. The reasons for not using conventional techniques for our problem is to avoid the tuning of the gain matrix, which is a laborious task for the set of models under study.

### Discrete inverse theory

The concept of the generalized or pseudo-inverse ( $\mathbf{G}$ , see [Rao and Mitra, 1971]) is used to find a solution to (3). The exact form of the generalized inverse depends on the problem at hand. Some frequently used generalized inverses are  $\mathbf{G} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T$  (least squares solution) or  $\mathbf{G} = \mathbf{D}^T(\mathbf{D}\mathbf{D}^T)^{-1}$  (minimum length solution). For a given generalized inverse, the unknowns at each time instant  $k$  can be estimated from

$$\hat{\mathbf{m}} = \mathbf{G}\mathbf{y} \quad (10)$$

where the subscript  $k$  is omitted for ease of notation.

The relation between the estimated and the true model parameters ( $\mathbf{m}_{true}$ ) follows from inserting (3) in (10), so that

$$\hat{\mathbf{m}} = \mathbf{m}_{true} + (\mathbf{G}\mathbf{D} - \mathbf{I})\mathbf{m}_{true} + \mathbf{G}\mathbf{e} \quad (11)$$

where the  $Q \times Q$  matrix  $\mathbf{G}\mathbf{D} (\equiv \mathbf{R})$  is an orthogonal projection matrix, which is often referred to as *model resolution matrix*. Since in general  $E(\hat{\mathbf{m}}) \neq \mathbf{m}_{true}$ , except when  $\hat{\mathbf{m}}$  is the least squares solution so that  $\mathbf{G}\mathbf{D} = \mathbf{I}$  with  $E(\mathbf{G}\mathbf{e}) = 0$ , it follows that it is likely to be biased. Notice that the estimated parameter vector is a function of the true parameter vector, the deviation of the model resolution matrix from the identity matrix, and some mapping of the output error.

Similarly, the estimated model parameters  $\hat{\mathbf{m}}$  may be used to evaluate how well predictions by the model correspond to the observed data through

$$\hat{\mathbf{y}} = \mathbf{D}\hat{\mathbf{m}} \quad (12)$$

where  $\hat{\mathbf{y}}$  is the predicted output  $\mathbf{y}$ . By substituting (10) into (12) and applying the same ordering as in (11) the following expression is obtained

$$\hat{\mathbf{y}} = \mathbf{y} + (\mathbf{D}\mathbf{G} - \mathbf{I})\mathbf{y} + \mathbf{e} \quad (13)$$

Here the  $P \times P$  matrix  $\mathbf{D}\mathbf{G} (\equiv \mathbf{N})$  is called the *data information matrix*. This matrix describes how well the predictions match the original data, apart from the observation errors in  $\mathbf{e}$ . The diagonal elements in the information matrix indicate how much weight a datum has in its own prediction.

Another measure for model quality is the covariance matrix of the estimated model parameters ( $\mathbf{C}_m^{\wedge}$ ). The resolution matrix, information matrix and model covariance matrix are useful to define the criteria of a good inverse and thus implicitly good measures of the model quality, e.g. [Backus and Gilbert, 1968]. On the basis of some norm of  $(\mathbf{R} - \mathbf{I})$ ,  $(\mathbf{N} - \mathbf{I})$ , or  $\mathbf{C}_m^{\wedge}$  trade-off curves can be constructed (see e.g. [Hansen, 1992, Menke, 1989]). A rather popular technique, due to its robustness and the avoidance of using the covariance matrix, is *generalized cross-validation* [Golub et al., 1979, Hansen, 1998, Wahba, 1977], which amounts to the minimization of the generalized cross validation function

$$f_{gcv} = \frac{\|\mathbf{y} - \mathbf{D}\hat{\mathbf{m}}\|_2^2}{(\text{trace}(\mathbf{D}\mathbf{G} - \mathbf{I}))^2} \quad (14)$$

where  $\mathbf{G}$  is the matrix to be chosen. It is based on the philosophy that if an arbitrary element  $y_i$  is left out, the corresponding solution should predict this observation well; and that the choice of the solution should be independent of an orthogonal transformation of  $\mathbf{y}$  [Hansen, 1998]. This function  $f_{gcv}$  will be used in what follows to find a proper generalized inverse  $\mathbf{G}$ .

### Solution method

The generalized inverse  $\mathbf{G}$  for our ill-posed problem can only be formed by including prior information to reconstruct unknown system properties and/or regularization (dampening) factors to reduce instabilities. Therefore a set of soft constraints is added to (3), that is

$$\begin{bmatrix} \mathbf{y} \\ \rho\mathbf{m}_{pri} \end{bmatrix} = \begin{bmatrix} \mathbf{D} \\ \rho\mathbf{I} \end{bmatrix} \mathbf{m} + \begin{bmatrix} \mathbf{e} \\ \mathbf{e}_{pri} \end{bmatrix} \quad (15)$$

where  $\mathbf{I}$  is a  $Q \times Q$  identity matrix and  $\rho$  is a so-called regularization parameter that will be explained later. Since

$$\text{rank}\left(\begin{bmatrix} \mathbf{D} & \mathbf{y} \\ \rho\mathbf{I} & \rho\mathbf{m}_{pri} \end{bmatrix}\right) = \text{rank}\left(\begin{bmatrix} \mathbf{D} & \mathbf{y} \\ \rho\mathbf{I} & \rho\mathbf{m}_{pri} \end{bmatrix}\right) \quad (16)$$

the problem can now be solved in a least-squares sense.

At this point differences arise between the parameter and state estimation solutions. In the case of parameter estimation we use the averages of the stochastic parameters obtained from previous experiments as prior information ( $\mathbf{m}_{pri}$ ), whereas in the case of state estimation we use the estimates at the previous time instant ( $\mathbf{m}_{k-1}$ ).

Depending on the regularization parameter  $\rho$ , the solution to (15) will vary between the minimum length solution (i.e. the solution based on prior information) or the least squares solution in the observation space. As noted above, generalized cross-validation is used here to determine the desired value of  $\rho$ . The generalized inverse therefore gets of the following form

$$\mathbf{G} = (\mathbf{D}^T\mathbf{D} + \rho\mathbf{I})^{-1}\mathbf{D}^T \quad (17)$$

The solution is found by applying a number of steps. The mathematical operations involved in each step at each time instant  $k$  are fully described in van Loon [2001], Appendix D. Since (4) varies over time, at each time instant ( $k$ ) a different solution is obtained. In this study  $\rho$  varied between  $3.4 \cdot 10^{-4}$  and  $4.7 \cdot 10^{-2}$ .

### 3. DESCRIPTION AND USE OF MODELS AND DATA SETS

The regularization algorithm described in Section 2 requires a prior model in combination with a data set. In this study a data set from the 200 ha Horizontes catchment in Costa Rica is used. The data set contains observations of rain, ground water levels, overland flow, soil properties, vegetation properties and topography (not discharge - since this is to be predicted). The data was collected from April 1996-August 1998 [van Loon, 2001]. In van Loon and Keesman [2000] a set of models (with different structures and parameter values) has been introduced to represent the hydrology of the catchment. This set with prior models is used here. The prediction is done for 15 individual rainfall events. Predictions are made in three different ways: 1) in open loop form (i.e. only using the required model inputs), 2) by applying parameter regularization, and 3) by applying state regularization. The prediction algorithm proceeds by first rewriting the prior models in the form

of (3) and subsequently applying the algorithms for open loop calculations, parameter and state regularization for discharge prediction per rainfall event. Not all the available data is used for regularization but rather subsets of various sizes. The largest subset only contains 75% of the data, while the remaining 25% is used for validation purposes. The nine subsets are established by a latin-hypercube sampling scheme, where 25, 50 and 75% of the observation time instants are combined with 25, 50 and 75% of the observation locations (thus yielding 9 combinations). In all cases 25% of the observation time instants and locations are used for validation. The naming of the subsets is shown in Table 1. The relative root mean squared error (RRMSE) of the discharge prediction is considered for models at various resolutions and for the various sub-sets.

Table 1. Naming of nine sub-sets.

coverage in time	coverage in space		
	25%	50%	75%
25%	1.1	1.2	1.3
50%	2.1	2.2	2.3
75%	3.1	3.2	3.3

## 4. RESULTS

### *Overall model performance*

The results for the open-loop prediction (without using observations) are shown in Figure 1. The figure shows the time-averaged RRMSE of the discharge prediction for Horizontes as function of upstream area, for different model resolutions. In the same figure, also the RRMSE of the discharge prediction for the calibration period is shown. In both the calibration and the prediction periods the RRMSE appears to decline for increasing catchment size. In addition, the RRMSE for the calibration period is considerably smaller than that in the prediction period. The fact that a larger RRMSE is found when considering a smaller area, implies that the heterogeneity of the small area is not represented well by the model; the model units are too coarse to capture the system heterogeneity at this scale.

### *Performance of state and parameter regularization*

In Figure 2 the time-averaged RRMSE of the discharge prediction is shown for both parameter and state regularization, when using sub-set 1.1 of Horizontes (see Table 1). It shows that the RRMSE for these situations is considerably lower than that in the case of an open-loop prediction,

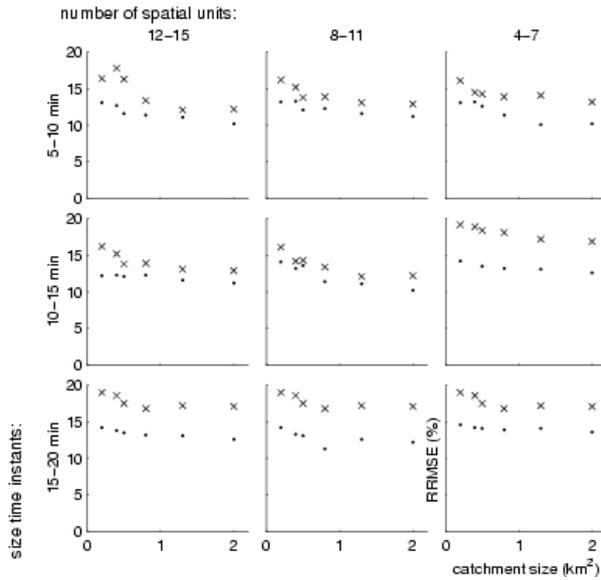


Fig. 1. Time-averaged RRMSE of discharge prediction in Horizontes at six different measurement locations for open-loop models: calibration (●) and prediction (×).

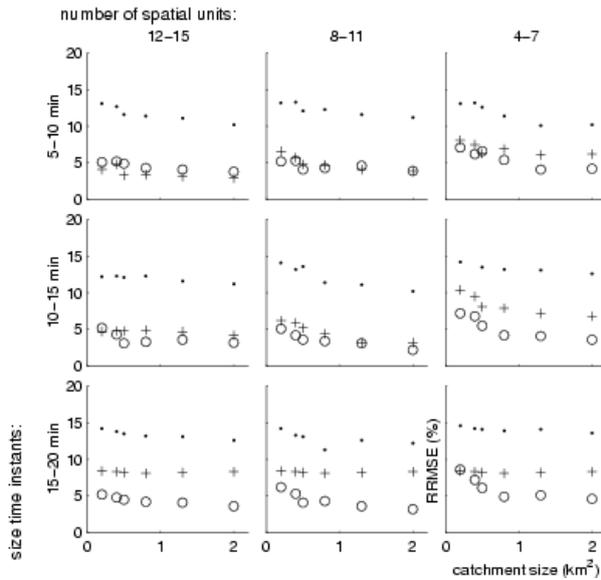


Fig. 2. Time-averaged RRMSE of discharge prediction in Horizontes at different measurement locations for models with parameter (+) and state (o) regularization and compared with calibration results (●).

at all resolutions but especially at the finer resolutions. State regularization generally leads to better predictions than parameter regularization (the o lay below the + symbols). For the subcatchment Horicajo similar results appear (not shown here).

In what follows the RRMSE of the predictions will be averaged over the six measurement locations. This allows to study the relation between the amount of data used for conditioning, the model resolution where the minimum RRMSE is found,

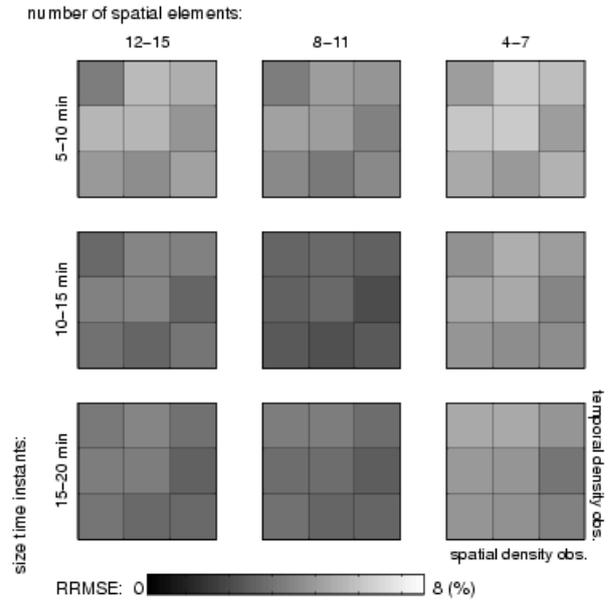


Fig. 3. RRMSE of discharge predictions by applying *parameter* regularization, for different resolutions and observation densities.

and the value of the RRMSE itself. In Figure 3 the RRMSE is shown for different observation densities (using the Horizontes data) when applying parameter regularization. The average RRMSE values of each plot in Figure 2 correspond to the values in the upper left corners of the nine plots in Figure 3 (+ in Figure 2) and Figure 4 (o in Figure 2). The cells in each of the nine sub-plots of the figures are defined by Table 1, e.g. lowest observation densities are in the upper left corner. It appears that for increasing observation densities in general the RRMSE decreases, and in addition the optimum RRMSE is found at the medium space-time resolutions. Figure 4 shows the same information when applying state estimation. By comparing Figures 3 and 4 it appears that the minimum RRMSE value for state regularization is slightly smaller than that for parameter regularization, but only at high observation densities and at finer resolutions. The plots of RRMSE when using the Horicajo observations show a very similar pattern. In general, state regularization leads to better predictions at finer resolutions (that is with a large number of spatial elements and small sampling time) than parameter regularization, and parameter regularization to better performance at coarse resolutions.

## 5. DISCUSSION AND CONCLUSIONS

Recall that in this study a so-called Tikhonov regularization is applied to the problem of combining model results with observations, in this way leading to predictions that are far better

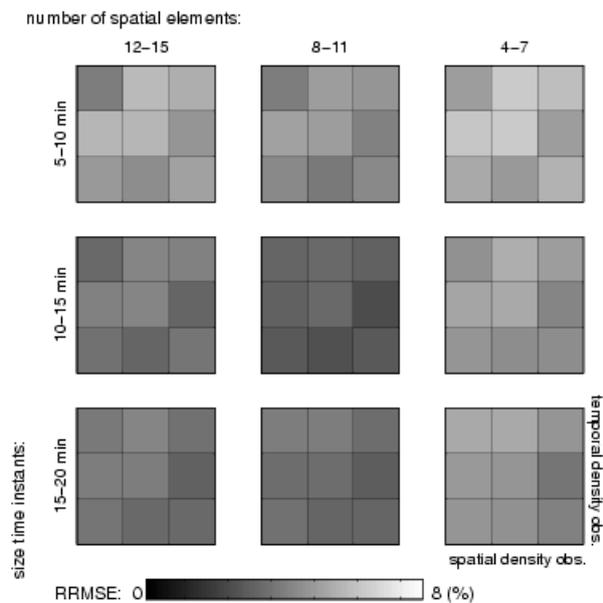


Fig. 4. RRMSE of discharge predictions by applying *state* regularization, for different resolutions and observation densities.

than the open-loop predictions with the same models. An objective and structured weighting of both components has been achieved through generalized cross-validation. Two regularization strategies were compared, i.e. parameter and state regularization. It was shown that both techniques lead to similar results, which differ in some details. Parameter regularization leads to better results at low data availability, whereas state regularization leads to better results at high data availability. This result is consistent with the fact that there are fewer parameters than states to be estimated. State regularization realizes its best predictions at finer resolutions than parameter regularization. This can be explained from the fact that parameter regularization maintains the *relative* structure of the kernel functions over different spatial units, and is in that way less flexible than state regularization. The comparative advantage of this flexibility appears at higher data availability and finer resolutions.

It is important to note that the methodology employed in this paper can relatively easily be adapted to different models or observations. If a hydrologic model can be written in state space form (according to (1) and (2)), the solution algorithm of Section 2 can be applied without any adjustments. This is an important asset since it means that different parameter distributed dynamic models can be evaluated with the same or different data for regularization relatively easily, without the requirement to include exactly the same state variables in the model dynamics. This point has been demonstrated by using a set of models instead of a single model to generate pre-

dictions. The technique used in this study only requires the solution of a series of independent constrained linear regression problems, for which there are numerous efficient solution algorithms available.

## REFERENCES

- G. Backus and F. Gilbert. The resolving power of gross earth data. *Geophys. J. R. Astr. Soc.*, 16: 169–205, 1968.
- G. H. Golub, M. T. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.
- P. C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Review*, 34: 561–580, 1992.
- P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, 1998.
- W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45 of *International Geophysics Series*. Academic Press, rev. edition, 1989.
- C. R. Rao and S. K. Mitra. *Generalized Inverse of Matrices and its Applications*. John Wiley & Sons, 1971.
- E. E. van Loon. *Overland Flow: Interfacing Models with Measurements*. PhD thesis, Wageningen University, 2001.
- E. E. van Loon and K. J. Keesman. Investigating scale-dependent models: the case of overland flow at the hillslope scale. *Wat. Resour. Res.*, 36(1):243–255, 2000.
- G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.