**Advances in
Geosciences**

Discharge assimilation in a distributed flood forecasting model

D. Rabuffetti

ARPA Piemonte – Forecasting and monitoring area, Corso Unione Sovietica, 216, 10134 Torino, Italia

Abstract. In the field of operational flood forecasting, uncertainties linked to hydrological forecast are often crucial. In this work, data assimilation techniques are employed to improve hydrological variable estimates coming from numerical simulations using all the available real-time water level measurements. The proposed assimilation scheme, a classical Kalman filter extension to non-linear systems, is applied in a rainfall-runoff distributed model based on the SCS-CN approach. The complex hydrological system of the Toce river basin is studied, a mountainous catchment of about 1500 km^2 in the Italian alps, through the development of a prototype available for operational use. For the considered flood event, the assimilation scheme is stable, even when available observations show gaps or outliers. It allows significant improvements in the simulation results, in particular when the focus is addressed to the peak.

1 Introduction

In the last years, local, national and international authorities showed an increasing awareness of flood and inundation hazard. Higher and higher damages and large costs in term of human lives, ask for new and more efficient strategies for risk management (WMO, 2004). Flood forecasting certainly plays a major role in these new approaches through the early warning to Civil Protection authorities allowing a more efficient organisation of mitigation and safety plans. The present work deals with the uncertainties in the hydrological forecasting models, often resulting into slight reliable discharge estimates especially when little and fast responding basins are focused on. The aim of the work is to provide a tool for reducing the errors and evaluating the uncertainties in the quantitative discharge forecasts (QDF), based on a simple scheme of discharge observations assimilation. A Kalman filter recursive technique (Jazwinski, 1970) is de-

Correspondence to: D. Rabuffetti
(d.rabuffetti@arpa.piemonte.it)

veloped to avoid heavy calculations that can increase computational time loosing one of the main requirements for a real-time system. Applications of Kalman filter can be found in hydrological modelling (Todini and Wallis, 1978). The key point here is the use of a distributed hydrological model and easily available real time data so to develop a useful tool simple to be extended to a large number of cases. Only rain gauge data and discharge observations are used, as they are usually available in survey networks. A prototype system is developed in the frame of the Piemonte hydrometeorological alert system, in the upper Po river catchment, northwestern Italy (Rabuffetti and Barbero, 2004). The case study focuses on the October 2000 flood of the Toce river.

2 The rainfall-runoff model

The distributed event-based model, FEST (Mancini, 1990), is used. The two main components of the hydrological cycle are represented into separated modules: run-off production on the hill slopes; flood wave formation and propagation along the river network. The first module implements the simple SCS-CN infiltration model (USDA, 1986) that calculates the run-off for each cell of the DTM. The second module is based on the Muskingum-Cunge routing model. In this application, the infiltration module is considered in the assimilation process. The CN method is based on the well known relation between total precipitation, P, and run-off, R, where $Ia = c \cdot S$ and $c = 0.2$ as usual:

$$R = \begin{cases} \frac{(P-Ia)^2}{P-Ia+S} & P > Ia \\ 0 & P \le Ia \end{cases} \tag{1}$$

S is an infiltration capacity index [mm] that can be evaluated as a function of the single coefficient CN, which depends on soil characteristics and land use. Though the method is quite rough, it offers the chance to describe all the soil properties with a single parameter that does not need calibrating. Several applications of the method have shown that S should not be taken as a constant but that it varies on the base of
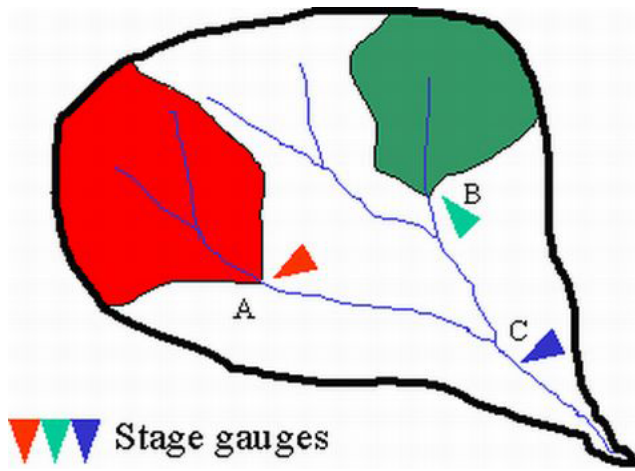
Fig. 1. Sketch of subcatchment definition. The subcatchment closed at section C cannot be considered because it would cover subcatchments A and B.

the saturation degree of the soil. Three wetness classes are defined based on antecedent precipitation. Recently Bingner and Theurer (2001) proposed a continuous variation of S during the rainfall event, evaluated by means of a simple water balance equation. In this way S becomes a sort of state variable describing the residual soil infiltration capacity.

The CN method shows a heavy sensitivity to the estimation of initial condition. From an operational point of view, this can create serious problems during real-time forecasting (Mancini et al., 2002). For this reason, it has been considered to avoid the initial condition assessment preferring an indirect evaluation based on the observed response of the catchment and its consequent assimilation in the system variable S.

3 The assimilation scheme

Discharge measurements at catchments outlet can be seen as a way to represent the run-off generation process integrating all the contributions of the single catchment portions. Focusing on the different subcatchments closed at each stage gauge in the river network, one can obtain a number of independent parts of the whole catchment of interest (Fig. 1). Each subcatchment is characterised by the state variable S, calculated every time step, t, as the mean on the cells belonging to it.

For every independent subcatchment, an assimilation procedure is implemented. For theoretical details see Cohn (1997) and Welch and Bishop (2001). The different components of the scheme: the dynamic system, the observation model and the errors are described in the following paragraphs. The calculation procedure is quite simple. First, at each computational time step the deterministic distributed model calculates the expected discharge for all the cross section of interest. Then the availability of discharge data for any of these sections is checked for. Where data are available, with no requirement to have all the observations at

specified time intervals (crucial feature in a real-time framework when data availability cannot be always guaranteed), observed cumulated flood volumes and total run-off coefficients are calculated. In the end, by means of Kalman filter, $\overline{S}_t$ is corrected and the correction used to update all the cell values. The question is to understand if and how much a mean subcatchment correction, does improve the results of the distributed model at cell scale.

3.1 The dynamic model

With respect to the CN approach, $\overline{S}_t$ should be kept constant during a rain event. So a simplified dynamic model can be set up enhancing the fact that the model is not perfect through the introduction of a noise, w_t, which represent model errors.

$$\overline{S}_{t+1} = \overline{S}_t + w_t \tag{2}$$

3.2 The model error

An estimate of the model error is based on the hypothesis that S can range between a dry and a wet condition value (USDA, 1972). This range is usually very high compared with S value itself, meaning that the 'a priori' estimate of the model error should be a very high value. In any case, once fixed the observation error, this value can be considered a calibration parameter of the Kalman filter: here, $\sigma_{\overline{S}}^2 = 100\,\text{mm}^2$ is used because a first sensitivity analysis showed that this value increased the assimilation stability.

3.3 The observation model

Rainfall data and cumulated flood volumes, $V_{\text{flood}}(t)$, easily calculated from discharge data, are analysed in terms of total run-off coefficient, $\phi_{\text{obs}}(t) = V_{\text{flood}}(t)/P(t-L)$. The time lag, L, is introduced to account for the time needed for total precipitation to become catchment run-off. This lag is calculated with the SCS lag equation (USDA, 1973).

In general, the observation model can be written as $\phi(t) = f(\overline{S}_t) + v_t$, highlighting the errors in the noise v_t. Equation (1) relates ϕ to the state by the expression below:

$$\phi(t) = \frac{R(t)}{P(t)} = \frac{(P(t) - c \cdot \overline{S}_t)^2}{P(t)\left(P(t) + (1-c)\overline{S}_t\right)} \quad \text{with} \quad P > c \cdot \overline{S}_t \tag{3}$$

When $P > c \cdot \overline{S}_t$, Eq. (3) can be solved as second-degree equation in $\cdot \overline{S}_t$. Equation (3) has two real and different solutions, $\hat{S}_{t,1} e \hat{S}_{t,2}$. By the Tartenville analysis method, it is easy to verify that $\hat{S}_{t,1} > (P/c) > \hat{S}_{t,2}$ and that $\widehat{S}_{t,1}$ is not acceptable because $\hat{S}_{t,1} > P/c$ in contradiction with the condition in Eq. (3).

3.4 Observation error

Observation error is represented by the estimated variance of ϕ_{obs}. In this case $\sigma_{\phi_{\text{oss}}}^2 = 0.04$ is assumed. That's to say that the observed run-off coefficient has a mean error of
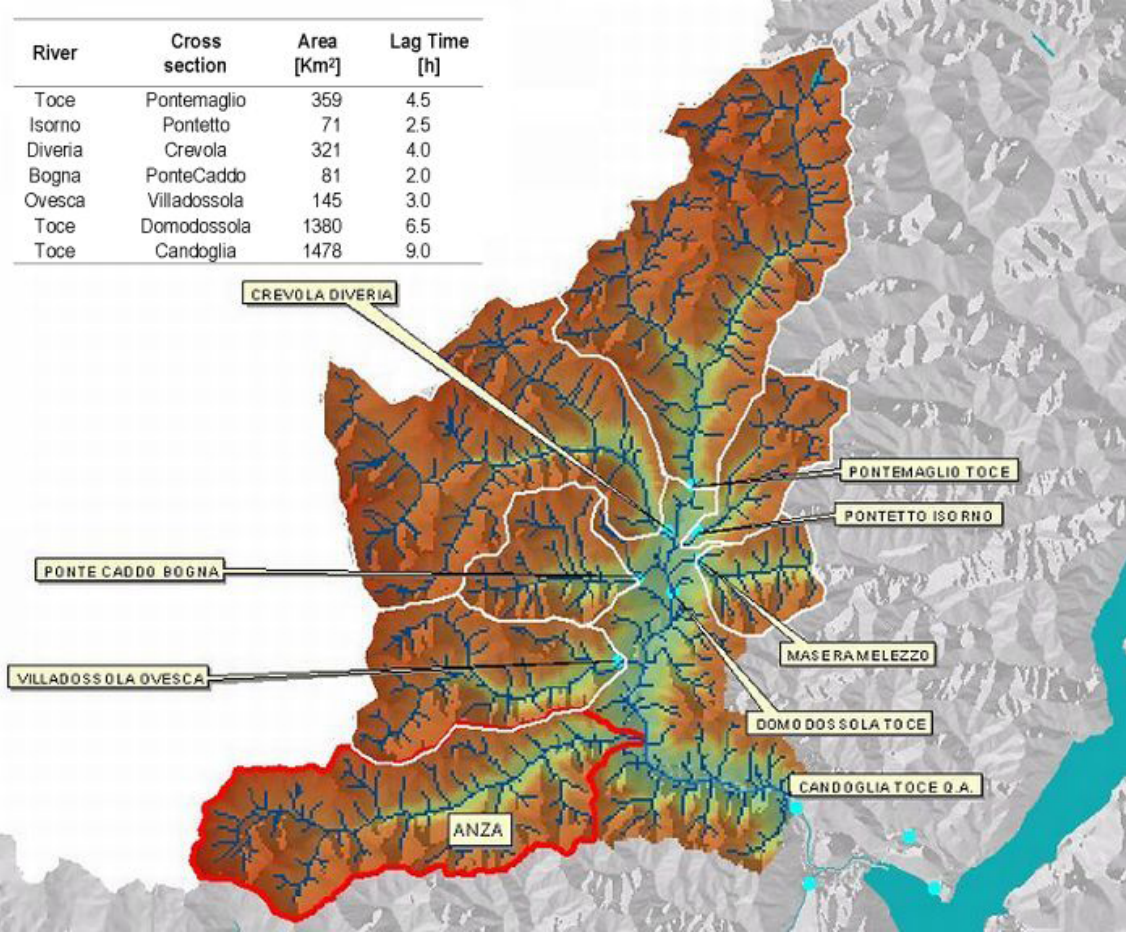
River	Cross section	Area [Km²]	Lag Time [h]
Toce	Pontemaglio	359	4.5
Isorno	Pontetto	71	2.5
Diveria	Crevola	321	4.0
Bogna	PonteCaddo	81	2.0
Ovesca	Villadossola	145	3.0
Toce	Domodossola	1380	6.5
Toce	Candoglia	1478	9.0

CREVOLA DIVERIA

PONTE CADDO BOGNA

VILLADOSSOLA OVESCA

PONTEMAGLIO TOCE

PONTETTO ISORNO

MASERA MELEZZO

DOMODOSSOLA TOCE

CANDOGLIA TOCE Q.A.

ANZA

Fig. 2. Toce catchment with monitored cross section. Subcatchments and river network derived from 200×200 m DEM are highlighted. Along the Anza river, there are no stage gauges.

± 0.2 which seems a reasonable hypothesis taking into account flood event observations as well as the uncertainty either in discharge measurements and in mean area rainfall estimates. An "a priori" estimation is here possible because in the kalman filtering approach what determines the results is the relative weight of model and observation errors. The sensitivity analysis on model error (Sect. 3.2) is then enough to understand the filter performance.

3.5 State updating

The Kalman filter works for linear systems. In this case, Eq. (3) can be approximated by the tangent linear approach around the solution (Bouttier and Courtier, 1999).

$$\frac{d\phi}{d\overline{S_t}}(t)\bigg|_{\hat{S}_{t,1}} = -2c\frac{(P(t) - c\hat{S}_{t,1})}{P(t)\left[P(t) + (1-c)\hat{S}_{t,1}\right]}$$

$$-(1-c)\frac{(P(t) - c\hat{S}_{t,1})^2}{P(t)\left[P(t) + (1-c)\hat{S}_{t,1}\right]^2} \quad (4)$$

One can now calculate the Kalman gain, the innovation to $\overline{S_t}$ and the new estimate variance:

$$K_t = \sigma^2_{\overline{S_t}} \cdot \frac{d\phi}{d\overline{S_t}}(t)\bigg|_{\hat{S}_t} \cdot \left[\sigma^2_{\overline{S_t}}\left(\frac{d\phi}{d\overline{S_t}}(t)\bigg|_{\hat{S}_t}\right)^2 + \sigma^2_{\phi_{oss}}\right]^{-1}, \quad (5)$$

$$\Delta\overline{S_t} = K_t(\phi_{oss}(t) - \phi(t)), \quad (6)$$

$$\sigma^2_{\overline{S_t}} = (1 - K_t \cdot \frac{d\phi}{d\overline{S_t}}(t)\bigg|_{\hat{S}_t})\sigma^2_{\overline{S_{t-1}}}. \quad (7)$$

3.6 Initial condition

The Kalman procedure is recursive and needs, for the first time step integration, an estimate of the initial condition and of its variance. The simple assumption made here is to start from normal conditions with an estimate variance $\sigma^2_{\overline{S}}$ as defined in Sect. 3.2.

4 Application to the Toce river basin

The Toce river is one of the main tributaries of the Lake Maggiore in northwestern Italy. It's a medium sized Alpine
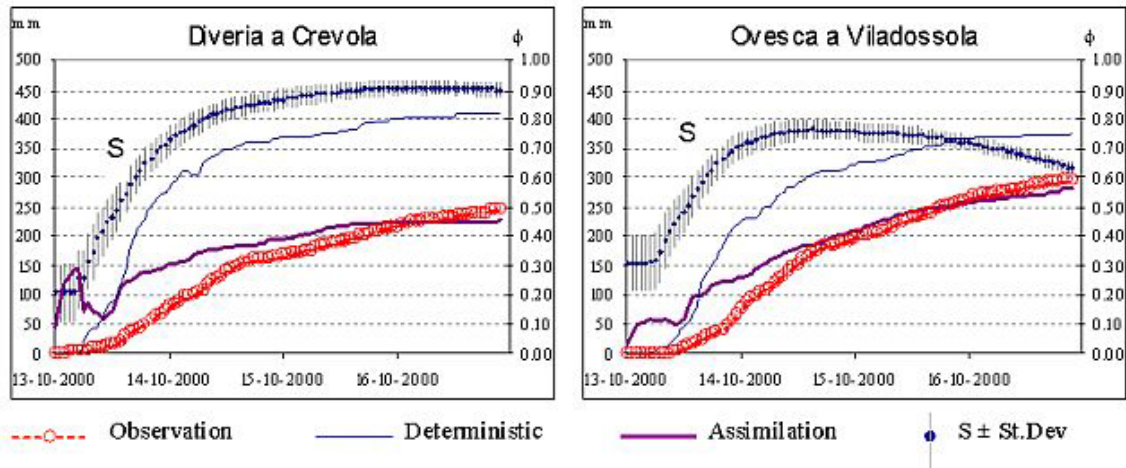
Fig. 3. Evolution of $\overline{S}$ and of its estimation variance. Comparison between $\phi(t)$ different evolutions.
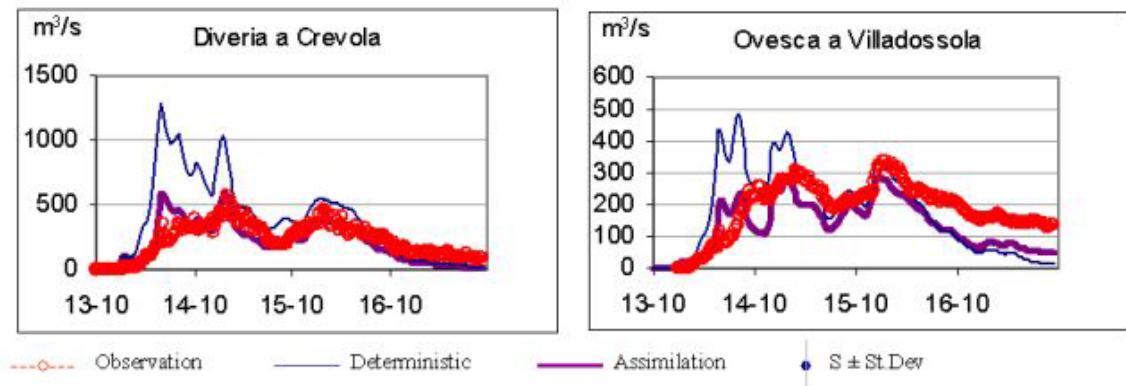

Fig. 4. Comparison between discharge hydrographs.

catchment with an area of about $1500\,\text{km}^2$. Nevertheless, it is a very important river in the Piemonte Region alert system because, beyond the damages it can cause in the developed Ossola valley, it has a crucial role in the lake floods. This catchment has been widely studied for its frequent floods as one can see in the Meteorological Alpine Program (MAP). The event here analysed is the heavy flood occurred in October 2000.

ARPA Piemonte manages the regional survey network that counts for more then 20 rain gauges in the area providing 10 min rainfall. Furthermore 8 water level gauges, providing 30 min level observations, are present: 7 are located at the confluence of all the main tributaries and one at the catchment closure (Fig. 2). In the model the hourly values are used.

4.1 Assimilation of tributaries flood volume observation

In the first experiment, flood volume observations along the tributaries are assimilated to correct subcatchments' $\overline{S}$ while discharge at Candoglia outlet is used only for verification.

In general, for the flood volumes of the tributaries in Fig. 3, one can notice an improvement, highlighting the fact that the assimilation procedure works in a proper way for the single subcatchments Anyway, for operational purposes, it is important to look at QDF to evaluate the effects of the assimilation. Therefore, in Fig. 4, forecasted hydrographs are presented: observed, calculated by the deterministic model and with assimilation. Improvements are quite evident for all the cross section. Table 1 synthetically reports the results: it is important to highlight the good results for Candoglia, which indirectly takes advantage of the assimilation.

4.2 Assimilation of flood volume observation at catchment outlet

The second experiment is complementary to the first. Flood volume at Candoglia is assimilated and data for tributaries are used for verification. In fact, from an operational point of view, it often happens that data on the tributaries, especially for little and mountain streams, are scarcely reliable and suffer a larger number of problems during heavy flood than measurements on the main rivers do. In this context, it

Table 1. Comparison of the results from the deterministic model and from the assimilation of flood volumes along tributaries. Higher values in bold. Model: A, with assimilation; D, Deterministic.

	Model	Diveria a Crevola	Isorno a Pontetto	Bogna a Pontecaddo	Melezzo a Masera	Ovesca a Villadossola	Toce a Candoglia	RMSE
Max absolute	A	323.6	53.4	**588.0**	127.2	145.2	1071	–
error [m^3/s]	D	**1062.2**	**75.7**	442.8	**129.8**	**356.8**	**1543**	–
Relative	A	−0.04	0.04	**0.58**	**−0.61**	0.10	0.17	0.35
peak error	D	**−1.23**	**0.13**	0.50	0.36	**−0.57**	**−0.30**	**0.62**
Relative	A	0.15	0.00	−0.08	**−0.39**	−0.50	**0.29**	0.29
volume error	D	**−0.50**	**0.19**	**−0.14**	0.17	**−1.02**	0.08	**0.48**

Table 2. Comparison of the results from the deterministic model and from the assimilation of flood volume at "Toce a Candoglia". Higher values in bold. Model: A, with assimilation; D, Deterministic.

	Model	Diveria a Crevola	Isorno a Pontetto	Bogna a Pontecaddo	Melezzo a Masera	Ovesca a Villadossola	Toce a Candoglia	RMSE
Max absolute	A	523.77	**111.86**	**560.95**	**137.35**	169.33	888.12	–
error [m^3/s]	D	**1062.21**	75.68	442.80	129.84	**354.59**	**1543**	–
Relative	A	−0.57	**0.23**	**0.67**	**0.45**	−0.22	0.09	0.43
peak error	D	**−1.23**	0.13	0.50	0.36	**−0.57**	**−0.30**	**0.63**
Relative	A	−0.28	**0.31**	0.01	**0.28**	−0.87	**0.23**	0.42
volume error	D	**−0.50**	0.19	**−0.14**	0.17	**−1.08**	0.09	**0.50**

is important to understand if discharge data on the main rivers can give useful information on the hydrological response of the minor subcatchments.

Details given in show that the assimilation of the basin outlet can have positive effect also on the tributaries QDF. The poorer performance with respect to the former experiment is justified taking into account that here much less data are assimilated. Furthermore, information retrieved by the catchment outlet is averaged on the whole and heterogeneous catchment. For the smallest catchments a worsen is noticed. Nevertheless, Diveria and Ovesca forecasts show significant improvements probably because their catchments are quite similar to the complete one and they are the main tributaries contributing to the Toce flood at Candoglia.

5 Conclusion

In this work, a simple scheme is proposed to exploit discharge data in the assimilation of the hydrological state of the catchment, an important feature for real-time operational hydrological forecasts for little and medium sized catchments.

The infiltration model is based on the SCS-CN method; the state variable $\overline{S}$ is updated with a kalman filter scheme assimilating rainfall and discharge observations, in terms of flood volume and run-off coefficient. The scheme proves to be stable from a numerical point of view with a very slight computational time request. The assimilation procedure allows a general and significant improvement in discharge forecasts.

The results show that discharge observations on the main tributaries can produce great benefits to the discharge forecasts in the tributaries themselves as well as for the whole catchment outlet.

In one experiment, only the catchment outlet was assimilated to simulate not well-surveyed catchments. Also in this case, an improvement can be noticed either related to the closure section and to some of the tributaries, in particular those mainly contributing to the flood.

In conclusion, its possible to state that the simple scheme proposed can have an effective implementation in a hydrometeorological forecasting system.

A further important result, which will be better analysed in future work, is the fact that the assimilation scheme evaluates also the uncertainties of the state variable estimate. This should allow building up scenario forecasts, which are very interesting for the decision makers who have to analyse model forecasts and issue alerts.

Edited by: V. Kotroni and K. Lagouvardos
Reviewed by: anonymous referee

References

Bingner, R. L. and Theurer, F. D.: AnnAGNPS TECHNICAL PRO-CESSES, Documentation, Version 2, 2001.

Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, Meteorological training course, lecture series, ECMWF, 1999.

Cohn, S. E.: An introduction to estimation theory, Data Assimilation Office, Note 97-01, 1997.

Jazwinski, A.: Stochastic Processes and filtering theory, Academic Press, New York, 1970.

Mancini M.: La modellazione della risposta idrologica: effetti della variabilità spaziale e della scala di rappresentazione del fenomeno dell'assorbimento, Phd Thesis, Politecnico di Milano, 1990.

Mancini, M., Montaldo, N., and Rosso, R.: Effetti di Laminazione di un sistema di invasi artificiali nel bacino del fiume Toce, L'Acqua, IV, 31–42, 2000.

Mancini, M., Mazzetti, P., Nativi, S., Rabuffetti, D., Ravazzani, G., Amadio, P.: Definizione di soglie pluviometriche di piena per la realizzazione di un sistema di allertamento in tempo reale per il bacino dell'Arno a monte di Firenze, 28° Convegno di Idraulica e Costruzioni idrauliche, Potenza 16–19 settembre 2002, II, 497–505, 2002.

Montaldo, N., Toninelli, V., Mancini, M., and Rosso, R.: Coupling Limited Area Models with Distributed Hydrologic Models for Flood Forecasting: the Toce Basin Study Case, IAHS, 274, 229–236, 2002.

Rabuffetti, D. and Barbero, S.: The Piemonte Region Meteo-hydrological Alert Procedure and the Real Time Flood Forecasting System, WMO, http://www.apfm.info/pdf/case_studies/cs_italy.pdf, 2004.

Welch, G. and Bishop, G.: An Introduction to the Kalman Filter, University of North Carolina at Chapel Hill, 2001.

Todini, E. and Wallis, J. R.: A real time rainfall run-off model for an on-line flood warning system, in: Application of Kalman filter to hydrology, hydraulics and water resources, edited by: Chaolin, C., 355–368, 1978.

USDA – U.S. Department of Agriculture, Soil Conservation Service: National Engineering Handbook, Section 4, Hydrology, Rev. ed., U.S. Department of Agriculture, 1972 and 1986.

USDA – U.S. Department of Agriculture, Soil Conservation Service: A Method for Estimating Volume and Rate of Runoff in Small Watersheds, TP-149, 1973.

WMO: Associated Program on Flood Management, 2004.