Reply to Dr. Uwe Ehret

I wish to sincerely thank Dr. Ehret for having analysed the paper with such attention: his thoughtful and helpful comments will certainly lead to an improvement in the paper. I hope he finds that his concerns have been adequately addressed in this response.

The Reviewer's Comments are in blue, boldface font, and the author's responses are in black, plain type.

1) ... the new and original parts of the presented work need to be pointed out more clearly, especially in delineation to the work that has already been done in this field (e.g. Jain and Srinivasulu, 2006, referenced in the work).

I will make clearer this aspect in the last paragraph of section 4.

The principal difference with all the cited works is that the present approach issues multi-step ahead forecasts: to do so it makes use, in the classification phase, not only of past streamflow data but also of past rainfall data; this is not done in many of the cited modular applications, as written in Section 4, but it is instead crucial for multi-step forecasts given the importance of past precipitation in the evolution of future further flows.

Also in the work by Jain and Srinavasulu (2006), a single step ahead forecast is issued and the alternative methods they propose for decomposing the data are always based on a separation of the hydrograph in different rising and falling limbs, that is, they are based on past streamflow data alone. I therefore believe that, even if from the text of their paper I understood that the rainfall values were included in the SOM mapping, the resulting classification is mainly seen and interpreted as a classification of the streamflow magnitude by the authors themselves who state:

"It may be noted that the SOM classifications correspond to different magnitude flows e.g. low, medium, and high based on the descriptive statistic in terms of mean."

This is not what is obtained in the present work, where instants with similar flow values may belong to different classes if they represent a different hydrometeorological condition. The reason of the interpreatation given by Jain and Srinavasulu may be due to the fact that the number of nodes of the SOM is small (3 and 4 respectively), but it is always difficult to interpret the data and analyses developed by other authors. It follows that in their work the potential of the SOM method is probably not thoroughly explored nor exploited and they obtained better results with the other modular approaches, based on the decomposition of the hydrograph. This is not surprising nor in disagreement with what obtained here, since for one-step ahead forecasts the role of past precipitation is certainly less important than for longer forecast horizons and the proposed decomposition methods might be more efficient than the SOM for interpreting the behaviour of the flow progress alone. On the contrary, for issuing forecasts over longer lead-times, the role of past precipitation is crucial and the algorithms based on the decomposition of the hydrograph cannot take it into account.

2) • Working with areal-averaged rainfall in 830 km2 catchment is a substantial loss of spatial information which influences shape and timing of the discharge hydrograph. The proposed model may perform better when rainfall input is provided in, say, 2 subregions (up- and downstream). The subdivision may be done applying a SOM on the individual raingauges.

The suggestion is interesting and it may be developed in a future study, even if I am not sure that the results will show an improvement in comparison to the areal average over the entire watershed. The averaging may, in fact, improve the efficiency of the forecasting since it produces a smoothing of the fluctuations recorded at each gauge, while preserving the general pattern of the storm event over the watershed.

A spatially-distributed representation of the input variables is worthy and may be important[1] in a model where there is also a

[1] *And even in distributed modelling, it may be more important a correct estimation of the total rainfall value over the watershed rather than its spatial distribution: in the work by Brath et al. (Journal of Hydrology, 2004, vol. 291, pp. 232–253), we developed a calibration/validation experiment with a spatially-distributed model and one of the analyses considered the influence of the spatial representation of the rainfall input. It was found that the model performances do not seem to noticeably deteriorate under the hypothesis of spatially uniform rainfall, provided that the mean areal rainfall intensity is reliably estimated, on the basis of a sufficiently extended number of raingauges. This result, in accordance with what*

spatially-distributed description of the properties of the watershed (and of the relevant processes), whereas in a systemic model there is no description of the internal physical phenomena and the concern is the overall response of the entire system (in our case of the entire watershed) to a given input.

In addition, even if it is true that a distributed input is a source of more detailed information on the precipitation field, and therefore it may be deemed to help the reproduction of the response of the system to such field, in a systemic model this also implies a multiplication of the input nodes and therefore of the parameters to be estimated (and when the rainfall input variable, as in the present case, is provided in input for a number of past steps, this number has to be multiplied for the additional input subregions in which the watershed may be divided, thus leading to a considerable increase of nodes). Such increase may lead to overfitting and therefore to poor generalisation ability: a network with too large a number of nodes may fit exactly the training set, but it may learn spurious relationships peculiar to the training data (in essence it fits also the noise), becoming lacking in generalisation capability. As a consequence, the training performance generally improves as the number of nodes increases, whereas the performance on an external validation set tends to deteriorate when the nodes become too numerous.

In a previous work (Toth and Brath, 2004) we analysed explicitly this point, testing the capability of an ANN for rainfall-runoff forecasting where the rainfall input was spatially distributed: the rainfall was not averaged over the watershed but each raingauge observation was connected through a separate input node (or better, more separate nodes for t-1, t-2, t-3, ..). The forecasts issued in validation with this network structure were sensibly worse than those obtained with the simpler structure with spatially averaged rainfall input, because of the increased model complexity and the following overfitting. The input dimension corresponding to the suggestion made by the Referee would be much more parsimonious (only two separate rainfall inputs are proposed and not one for each raingauge), and it might lead to an improvement in the modelling ability, but, as a rule, the increase of input variables in a data-driven model must be accurately pondered, since it is worthy only if the additional information that is fed to the system is so meaningful to compensate the increase of the parameters number.

A part from the opportunity of testing the Referee's suggestion in future research work, I would like to underline that lumped rainfall-runoff models, either of the conceptual or of the systemic type, have been and still are successfully applied (both in the literature and in the operational practice) over watersheds of much larger dimension than the Sieve river.

In particular, almost all the ANN applications for rainfall-runoff forecasting purposes adopt a lumped approach and such applications are often carried out on watersheds whose drainage areas are comparable or greater than that of the Sieve River watershed. Considering only, as an example, the papers cited in the present paper, the following works consider in input the mean areal precipitation and refer to case study watersheds of area larger (in some cases much larger!) than 700 km2: Cameron et al. (2002); Solomatine and Dulal (2003); Jain et al. (2004); Khan and Coulibaly (2006); Shamseldin et al. (2007), Parasuraman and Elshorbagy (2007), Gopakumar et al. (2007), Jain and Srinivasulu (2006), Corzo and Solomatine (2007).

3) • To a non-expert, the SOM technique is not completely clear from the explanation (P. 906 line 1 to p.907, line 12). Please explain in a more detailed, step-by-step manner.

As required also by the Referee Prof. Solomatine, I will explain better the process for the tuning of the network weights and the SOM use on respectively calibration and validation data, adding in the text the concepts below.

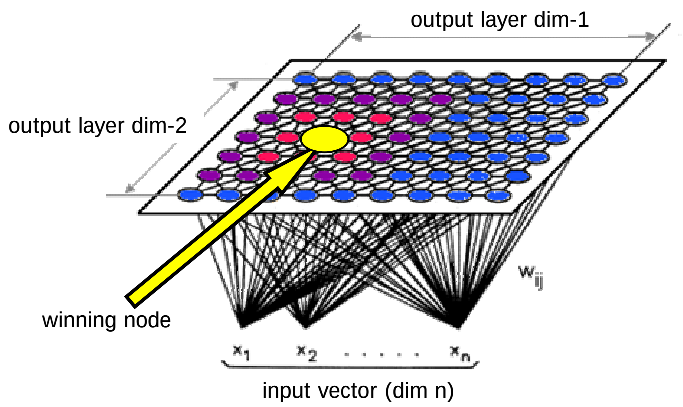
The weigths of the nodes of the SOM are adjusted, through the learning process, on the vectors of the calibration set. In the learning process all the calibration input vectors are processed through the SOM incrementally, one after the other, re-iteratively: for each sample input vector **x**, the weigths of the winner node and of the nodes in its neighborhood are changed closer to **x**. During the learning process, individual changes may be contradictory, but the net outcome in the process is that ordered values for the **W** emerge over the array.

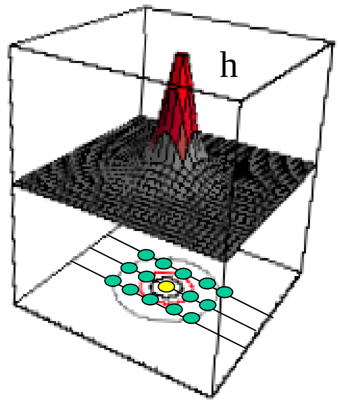
At the end of the learning phase, the weigths have reached a final, tuned value and the SOM may be used (without changing the weigths any more) to classify the calibration vectors: for each input vector,the best-matching unit, that is the closest, in the SOM output layer is identified. In exactly the same way, the tuned SOM may used to associate any new vector to one of the units of the SOM output layer, thus attributing the new data to the clusters identified before.

---

obtained in Obled et al. 1994, suggests that, at least as far as some case studies are concerned, a spatially uniform rainfall field may correspond to an adequate input representation also for a spatially-distributed rainfall-runoff model.

In addition, if the Referee suggests it, I may add a figure depicting the SOM network (like the one on the left):





and a figure (like the one on the right) for explaining the neighbourhood shape, used for decreasing the adjustment for increasing distance from the winning node on the output layer.

4) • For the sake of bridging the gap between system-theoretic and 'classical' (i.e. conceptual or process-based) modelers, a comparison of model performance with a simple conceptual r-r model will be helpful. Thereby, the number of free parameters subject to optimization/calibration in each model should be addressed and compared.

The assessment of the ability of neural network approaches in comparison with conceptual/process based models is certainly important, but it is not the focus of the present paper. I would refer the Referee to other publications where I described such comparisons for applications similar to the global model here presented: Toth and Brath, 2002; Toth and Brath, 2004, Toth and Brath, 2007.

5) The major part of the conclusions is comprehensible and substantial: a) the suitability of SOMs for automated classification of hydro-meteorological conditions, b) the benefit of selecting an adequate number of classes. However, further reasoning is necessary to explain why the global ANN, with respect to efficiency, equals or outperforms the multi-model ANNs for short lead times (one to three hours).

I believe that this is due to the fact that for shorter lead-times there is less influence of the hydro-meteorological situation of the forecast instant (which will instead control the evolution of the phenomena for longer time horizons). Therefore, for short lead-times, also the global model, not differentiating the hydro-meteo conditions, may allow satisfactory performances. I will add this consideration to the revised version (thank you for having raised the point).

6) Also, the conclusions should pick up and discuss findings of Jain and Srinivasulu, 2006 (referenced in the work) stating that decomposing a flow hydrograph based on physical concepts is better than using the SOM decomposition.

As said in the reply to comment 1), even if it is always difficult to interpret the data and analyses developed by other authors, the fact that Jain and Srinivasulu (2006) obtained better results for the one-step ahead forecasts with the other modular approaches, based on the decomposition of the hydrograph, is not in disagreement with what obtained here. In fact, for one-step ahead forecasts, the role of past precipitation is certainly less important than for longer forecast horizons and the decomposition methods proposed by Jain and Srinivasulu (2006) might be more efficient than the SOM for interpreting the behaviour of the flow progress. On the contrary, for issuing forecasts over longer lead-times, the role of past precipitation is crucial and the algorithms based on the decomposition of the hydrograph cannot take it into account. I would rather add this consideration only at the end of Session 4 and not repeat it in the Conclusions section.

Specific comments

7) • P. 902, line 8: include a short description of the hydrometeorological regime of the basin in general and of the calib/valid period (e.g. did singular floods occur, which can strongly influence the results etc.).

I will add a brief description of the morphology and of the climate of the watershed:

“The watershed is morphologically characterised by moderate to strong relief in the upper and lower sections and by a gently rolling plain in the central part. Except in the valleys, dedicated to agriculture, the terrain is forested and mountainous. The main stream follows the south-east direction and it is around 58 km long.

The fact that Mediterranean water is warmer than Atlantic water throughout the year and the presence of island barriers in the Mediterranean serve as preconditions for strong cyclogenesis causing most rainfall over the Sieve River between late Fall and early Spring, November being the wettest month. The summer months, especially July, are the driest, owing to the dominance of the Azores high-pressure cell.”

I will also add the main statistics of the streamflow observations for calibration and validation periods:

	N of data	Mean (m^3/s)	Std (m^3/s)	CV	Prc 75 (m^3/s)	Prc 95 (m^3/s)
Calibration period	26280	14.84	39.85	2.69	13.78	47.38
Validation period	11688	12.11	27.32	2.26	13.73	45.72

The mean and the percentiles are analogous, but the variability of the calibration streamflow data is more pronounced than that of the validation data: this is probably due to the fact that at the beginning of the calibration period, in Autumn 1992, occurred the major events (high but nor exceptional: two peaks with a return period corresponding to 5 years) of the observation period. This is not a drawback but quite the opposite: the calibration period in fact must have enough information contents, including a wide range of hydrological conditions and in particular it is useful that it includes the highest output values, due to the difficulties ANNs may experience in extrapolation (De Vos & Rientjes, WRR, 44, W08434, doi:10.1029/2007WR006734, 2008).

Last but not least, all the proposed forecasting models are calibrated on the very same set and the comparison, that is the focus of the work, is therefore fair.

8) • P. 901, line 23 to p. 902, line 3: If you mention this, clearly state the connection to your work, i.e. explain what your approach is and how you overcome the existing deficiency (missing consideration of the influence of the calib dataset).

I will add that

“The impact of the differences in the calibration data set is the core of the proposed approach, where different data sets are identified, to be used specifically for modelling the future evolution of similar data”.

9) • P.905, lines 5-10: As mentioned in the general evaluation, point out clearly the new and original parts of the presented work in comparison to the existing works presented in the section above. Delineate especially to work closely related to yours (e.g. Jain and Srinivasulu, 2006, referenced in this work).

See Reply to comment 1)

10) • P. 906, line 15: State that the Euclidean distance is the standard measure, otherwise explain why you have selected this specific measure.

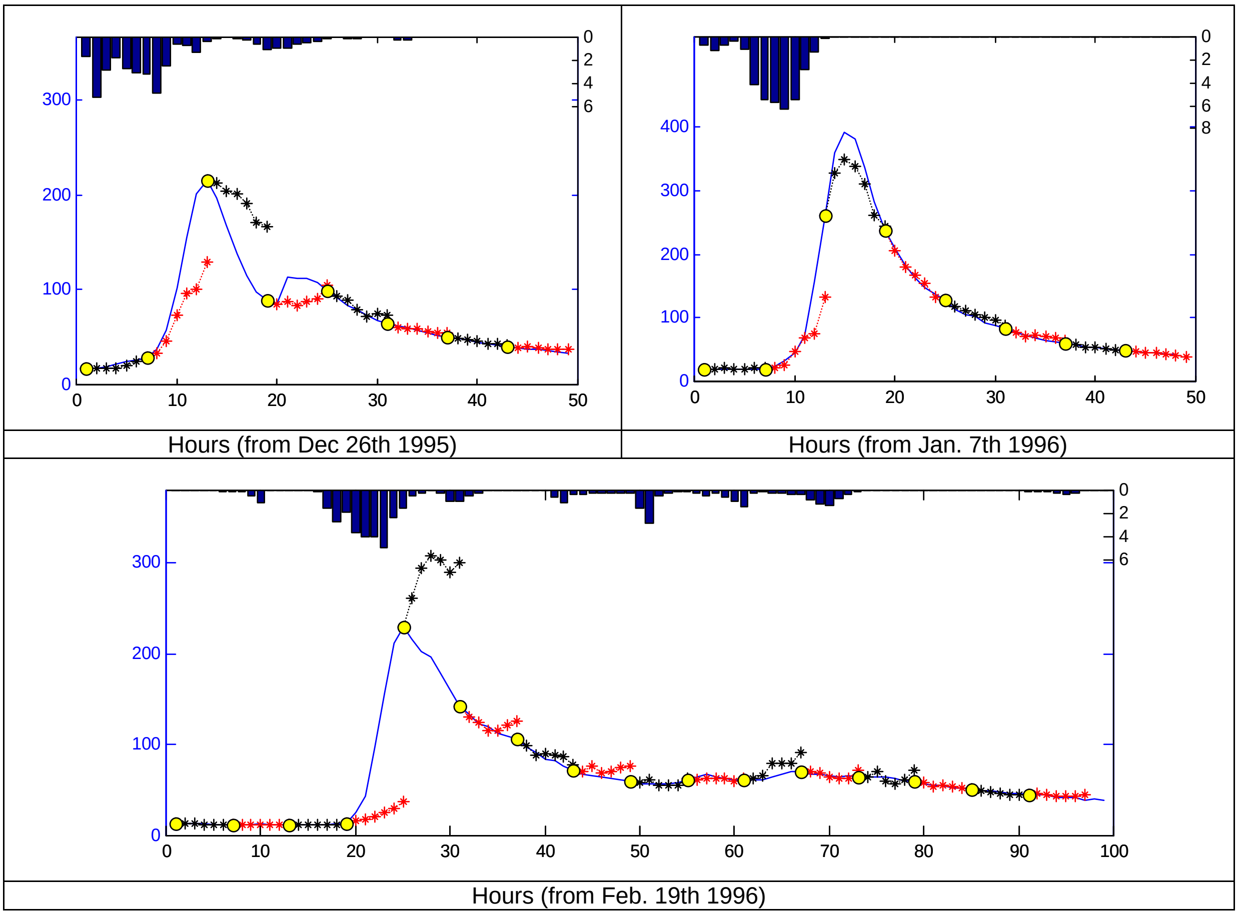
In the revised version I will add that the Euclidean is the most popular distance in SOM application. (Actually I carried out also some preliminary tests with a Manhattan distance, but the differences were negligible and I chose to use the distance used in the majority of SOM application).

11) • P. 908, line 17 / fig 1: Please add the textual explanation of the hydro-meteorological situations to figure 1. Especially in case of black/white printout, this increases understandability.

I will add to the caption of figure 1 a summary of the text of ll. 17-25, p. 908.

12) • P. 909 ff (chapter rainfall-runoff modeling): As the one to six hour forecast hydrograph in your work is based on six different models, it would be interesting to see a forecast hydrograph as combination of the six individual forecasts in order to asses its temporal autocorrelation and 'realistic looks'.

I will add the following hydrographs, showing the 1 to 6 hours ahead forecasts (alternatively red or black stars) issued in correspondence of different forecast instants (yellow circles) for three validation events.



The look, even if somewhere fluctuating (as expectable), is not too unrealistic; for the longest lead-times, the correlation with the past streamflows is not the only feature governing the trend of the forecasts, as it must be, since the forecast is based not only on past flows but also on past rainfall values

Finally, in the final version of the manuscript, I will take into account the valuable suggestions (listed below) on English revision.

• Replace 'numerousness' by 'occupancy'

• Replace 'precious' by 'valuable'

• Replace 'providing in input past inflows …' by 'using past inflows …. as input'

• P. 898, line12: name the information used: Streamflow and areal-averagad rainfall of the last three (two, respectively) hours.

• P. 899, line 12: explicitly

• P. 899, line20:In order to take into account …

• P. 903, line 6: replace 'resulted' by 'was'

• P. 903, line 25: Add: 'see below' to clarify to reader that some of the interesting applications are presented in the next paragraph.

• P. 904, line 22: … on past river flow only. However, information on the recent precipitation depths is valuable ….

• P. 905, lines 15-18: Maybe the point is clearer for the reader if the argumentation is turned around: 'It is important to underline that the combination of rainfall and runoff observations prior to the forecast contains valuable information about the systems' state of saturation and hence on its reaction to rainfall forcing in the forecast period'.

• P. 910, line 15: [0,oo[

• P. 911, line 16: '… problem, the opportunity to ….. was tested, so to ensure…'

• P. 911, line 25: recognized

• P. 912, lines 14-15: replace '…only information available in the forecast instant' by '… most relevant information available in the forecast instant: streamflow and rainfall observations.'

• P. 912, line 18: replace with: '… appear penalized by the partly low class occupancy'.

REFERENCES NOT CITED IN THE PAPER

Toth, E., Brath, A., Use of spatially-distributed or lumped precipitation inputs in conceptual and black-box models for runoff forecasting, in "Hydrological risk, Recent advances in peak river flow modelling, prediction and real-time forecasting - Assessment of the impacts of land-use and climate changes", Editors A. Brath, A. Montanari and E. Toth, pp. 247–261, Editoriale Bios, Cosenza, 2004.

Toth E., Brath A., Flood Forecasting Using Artificial Neural Networks in Black-Box and Conceptual Rainfall-Runoff Modelling, Atti dell'International Environmental Modelling and Software Society Conference 2002, Lugano (Switzerland), 24-27 June 2002, , iEMSs Publ., Como, Vol. 2, pp. 166-171, 2002.