

## ***Interactive comment on “Flood forecasting using transboundary data with the fuzzy inference system: The Maritza (Meriç) River” by A. Aydın et al.***

**Anonymous Referee #2**

Received and published: 3 August 2016

In this study the authors present, in a very synthesized way, the results of the application of a data-based method (the Fuzzy Inference System) for flow forecasting in the River Maritza at the Turkish city of Kirişhane. I understand that the model developed by the authors might have a relevance in the study region, but in order to fulfill the standards of a scientific publication, a considerable amount of additional work needs to be done by the authors.

The main weaknesses of the study are, in order of importance:

(1) The authors do not validate the forecast model. There are many strategies for validation, but probably the most straightforward one is the split-sample validation (see e.g. Klemes 1986). In this study, the authors use the same period (9-21 February 2010)

C1

for the calibration and for the validation of the model. Reporting the performance measures of the model(s) for this periods gives only a goodness-of-fit measure. Presenting them as validation measures (section 2.2 Model validation) is scientifically wrong, specially for a data-based method. Data-based methods are particularly characterized by performing well inside the range of values for calibration, and performing more poorly, even catastrophically outside these ranges. These connects another issue, related with the period used. Why just 9-21 February? Is this flood event representative in terms of processes [see also comment (6) below]? Is the reason for the short period chosen related with data availability/limitations? Note that in the literature, usual period lengths are 10yr calibration + 10yr (different, non-overlapping) validation; maybe the shortest that I could remember is 1yr calibration 2yr validation. I recommend strongly to the authors to use a longer dataset, and split it for calibration and validation.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, 31(1), 13-24.

(2) There is no presentation or discussion of the results in the text. The presentation of the results is limited to give a reference to Table 3, and Figures 6 and 7. One cannot expect a 15-lines section to include "Results and discussion". In the revised paper, the authors need to present quantitatively the performances, and comment (maybe theorize) about why do the methods perform differently? maybe related to routing/travel times? Also, it seems to me that the (exceptionally) high performance of the model is an artifact of the high number of rules used in the FIS, i.e. over-fitting (99 rules is of the same order of magnitude of the number of data points used in the study  $12 \times 12 = 144$ ). The authors need to present alternatives with a smaller set of rules and investigate the decrease of performance. Also, section 3 is titled "Results and discussion", but there is no discussion. The results need then to be discussed in context of similar studies from the literature.

(3) There is no description what-so-ever of the method used. From Page3 Line9 to Page4 Line4 is only textbook knowledge about fuzzy sets absolutely not relevant for

C2

this paper, the can be condensed in 3 sentences. The last paragraph should explain more in detail the Fuzzy Inference System, how the rules work, how they are defined, all in the context of this paper.

(4) In order to make provide some useful take-home message to a potential reader, it would be extremely useful a comparison to alternative methods. This does not need to be extreme long or complex, the authors could for example perform a regression (linear and/or non-linear) with the same predictors as inputs used in the FIS, and compare the performances of both models IN VALIDATION, i.e. in a period different to the one on which the models were calibrated.

(5) There are no conclusions. P6 L18 to P7 L2 is a summary (i.e., another abstract) and P7 L3-10 is a very brief and shallow outlook. There are no conclusions or take-home messages. In this sense, it also not clear what is the novelty of the paper.

(6) There are no references to the relevant river network structures of the region. Looking jointly at Figures 2 and 3, there must be a very important tributary to the Maritza between Svilengrad and Kirishane, as the discharge difference is large compared to the river reach length (at least I assume it, given that there is no scale in Fig. 2). Also, a brief discussion of the travel times [see comment (2)] would be needed to justify the inputs used in the FIS. Also, a hydrological description of the region is given in P5 L5-10 but, is this related to what happened between 9-21 Feb.2010 in terms of processes [see also comment (1) above] or in terms of flood waves travel times. Related to point (4), it would be useful to compare to results of data-based methods to very simple hydrodynamical (here simple routing, e.g. kinematic or diffusive wave would be enough). Of course, ideally, a comprehensive comparison of the FIS or the regression suggested in (4) could include a forecast rainfall-runoff model, but I understand if this is out of the scope of the present publication.

Please find below a (non-comprehensive) list of other issues that need to addressed in the resubmission:

### C3

P1 L7-9: The first 3 sentences of the abstract convey the same message, please condense in one sentence.

P1 L10: What is "short"? Please report length (e.g. in km)

P1 L10: "there is not adequate warning time for Turkey to alert the population against flash floods" I strongly disagree with the formulation of this sentence. It somehow transmits the message that "it is not adequate to alert the population against flash floods", which is disturbing. Please combine with the following sentence, just saying that discharge/water levels from Bulgaria are needed.

P1 L14: "Flow data from the Plovdiv, Svilengrad and Kirishane stations were gauged every two hours covering the period [...]" I would assume that the discharge data was gauged (measured) for a longer period. I would rephrase as "Discharge data from ... was used with a 2-hours resolution [...]"

P1 L15: I would add something like "The aim is to predict discharge for Maritza@Kirishane (if this is true) at a given time t. The four models differ "

P1 L20: determination coefficient -> coefficient OF determination P1 L20: Nash-Sutcliffe sufficiency -> Nash-Sutcliffe EFFICIENCY P1 L21: "correlation coefficient": which one? Linear correlation (i.e. Pearson) correlation coefficient?

P1 L22: "[...] models produced highly satisfactory results." - Please be quantitative, and avoid adjectives like "highly satisfactory" (in comparison to what?). Same for Line 24 ("satisfactory"). Please report the most relevant performance values in lines 22-24.

P2 L20: "[...] are capable of more accurate prediction [...]", more accurate than what?

P2 L24 to P3 L6: Please move this to the Data (catchment description) section.

P4 eq(4): The formula of the Coefficient of Determination is wrong. Equation (4) estimates the square of the linear (Pearson, product) correlation coefficient, or  $r^2$  (small "r" squared). Capital "R" squared,  $R^2$ , i.e. the Coefficient of Determination is defined

### C4

as  $1 - \text{SumSquaresResidual}/\text{SumSquaresTotal}$ . In this context this is equivalent to the definition of the Nash-Sutcliffe efficiency. Please note that  $r^2$  and  $R^2$  ONLY coincide in the case of a simple linear regression. One way to see that eq(4) is not correct is that the Coefficient of Determination may be negative (if the model performance is very poor), and this is impossible with the expression in eq(4).

P5 L12: again the discharge was gauged for a longer period, the authors USED only the mentioned period

P5 L24 "Mamdani model", please give reference

P5 L29 to P6 L1: "After determination of the degrees of the data in different regions in which numbers can be different, the fuzzy rules of the models were generated by using AND logical conjunction (Fig. 4)." I am not sure Figure 4 shows what the authors are describing in this sentence.

P6 L2: Figure 5, in it's current status in not needed

P6 L10 "A summary of the results of the performance criteria of the models is given in Table 3". Please comment the actual results in the section. Also, the name of the section is Results and Discussion and there is no discussion.

P7 L6-9: "Furthermore, long-term observation of the hydro-meteorological conditions of the floods is needed. Hence, developed models are crucial in the prediction of current flows. Doing nothing is not an option; thus, this study has shown that developed fuzzy rule-based models can satisfactorily predict the flow regime with high accuracy and therefore can be used to develop a reliable flow forecasting system." The authors here link the hydro-meteorological processes generating floods with the fuzzy forecast model presented. The data-based method discussed in this paper does not have any base what so ever on processes, it is just a blind fit of the FIS to a single flood event.

P12 Table3: Correlation values of 7.XX are wrong. They are limited, by definition in eq(7) to values between -1 and +1.

C5

---

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., doi:10.5194/nhess-2016-86, 2016.

C6