

Authors: We thank Liz Stephens very much for her review of the paper and will below address her comments point by point. C stands for comment and A for answer.

C: This is an interesting article that provides a valuable end-user viewpoint on how the EFAS system would like to see it improved. I have a few general comments that I think are required to help make the paper easier to understand and place in context. Firstly, I think the definitions of terms used such as 'forecaster', 'forecast', 'decisionmaker' need to be explained, since they could be misinterpreted. Here (I think) forecaster is used to describe the professional forecasters who are the end-users of EFAS. Some readers of this article might think that the forecaster refers to the scientists working at ECMWF or JRC. You also make reference to 'communication to decision-makers' (e.g. p2221, Line 1). Are these decision-makers the forecasters? This is confusing.

A: We are sorry for the confusion. Liz is right, here forecaster is used to describe the professional forecasters who are the end-users of EFAS as opposed to scientists working at ECMWF, the JRC or any of the other research focused institution. We will add a sentence on how we define forecaster and end-user. In the text on p221, Line 1 we changed the decision makers to "forecasters and end-users". Since the response to a forecast differs very much from country to country we cannot be very specific on how decisions are made.

C: Secondly, I would also like to see a bit more background on these forecasters /decision-makers. How do they use the forecast? Is it emailed to them? Do they access it online? Is it integrated into their own systems? What do they use it for? Do they issue early-warnings from it? Do they use it for management of flood defences or dams? How much contact do they have with those responsible for EFAS? Is it just the annual workshop? A bit more detail on this will set the context a bit better.

A: The idea behind EFAS-IS is to provide an early warning of high flows and to serve the European Emergency Response Centre (http://europa.eu/rapid/press-release_MEMO-13-427_en.htm) with an independent tool to assess floods in Europe and provide early warning information to national services in particular for trans boundary catchments. .

Early warning are issued and communicated in many different ways. If it is forecasted for 3 consecutive days that the warning level will be exceeded an alert is issued by email to the EFAS members in the affected basin and through publication in the restricted area of the web site - please note that all services also have access to the website at any time and can make their own assessment. A daily email giving an overview is also sent to the European Emergency Response Centre.

This is an early warning system, which is currently not designed to initiate any high cost action such as e.g. an evacuation. The information EFAS provides could be used in a variety of different ways by the national and regional organisations, for example, to amend a shift rota and increase readiness in flood incident management. In most cases EFAS is not yet fully integrated into national systems due to the challenge of making decisions under uncertainty and institutional barriers (for detailed discussion see Stephens et al., 2012, Demeritt et al., 2012, 2012, Norbert et al., 2010).

A recent study by the Environment Agency of England and Wales nicely illustrates what can be done (please see "Getting the best out of probabilistic flood forecasting – on the art and science of aiding decision making during incidents" <http://evidence.environment-agency.gov.uk/fcrm/en/Default/HomeAndLeisure/Floods/WhatWereDoing/IntoTheFuture/ScienceProgramme/ResearchAndDevelopment/FCRM/Project.aspx?ProjectID=c0899b01-6fcd-4775-9aeb-a2dcc8ec4d39&PageID=7ef5c014-c33b-4eb4-af59-ec6f0a09d80f>).

We will also add information on the level of contact. All EFAS partners are coming to the annual meeting. In addition there is a weekly phone conference, there are additional email and phone exchanges, there is a web forum, EFAS provides training (centrally or within partner institutions), there are side meetings at conferences and courtesy visits by the EFAS operational centres of individual partner institutions. EFAS also sponsors research activities in establishing the value of the EFAS system (<http://www.kcl.ac.uk/sspp/departments/geography/research/hrg/projects/EuropeanizingFloodForecasting.aspx>) which allows the partners to give feedback through a mediating, independent institution.

Stephens E, Edwards T, DEMERITT D. (2012) Communicating probabilistic information from climate model ensembles—lessons from numerical weather prediction. *WIREs Climate Change*. 3: 409-26 DOI: [10.1002/wcc.187](https://doi.org/10.1002/wcc.187)

Demeritt D, Nobert S, Cloke H, Pappenberger F. (2012) The European Flood Alert System (EFAS) and the communication, perception and use of ensemble predictions for operational flood risk management *Hydrological Processes*. DOI: [10.1002/hyp.9253](https://doi.org/10.1002/hyp.9253)

Demeritt D, Nobert S, Cloke H (2010) Challenges in communicating and using ensembles in operational flood forecasting *Meteorological Applications* 17: 17: 209-22 DOI: [10.1002/met.194](https://doi.org/10.1002/met.194)

Nobert S, Demeritt D, Cloke HL (2010) Informing operational flood management with ensemble predictions: lessons from Sweden *Journal of Flood Risk Management* 3: 72-79

C: Thirdly, I think some further explanation of EFAS is also required, since you refer to multi-model systems and so on later on in the article. It might be important to consider the intended audience of this paper, it appears that technical terms have been largely avoided, which may be fine, but some basic-level description of the science would be useful, e.g. for those outside of EFAS developing hydrological forecast systems for end-user applications.

A: We will add more information about the EFAS system. For the purpose of this discussion please see http://en.wikipedia.org/wiki/European_Flood_Alert_System and <http://www.efas.eu> for more details.

Specific Comments and Technical Changes

C: Title: This should either be “Forecaster priorities”, “Forecaster’s priorities” or “Forecasters’ priorities”.

I would go with the first, although I don’t know whether for a scientific journal it would be better to say ‘End-user priorities’, otherwise it sounds like this paper is about technical ECMWF or JRC opinions for forecast improvements. Equally, I think this paper deals with the improvement to the Early Warnings and not the forecasts themselves – e.g. it discusses better dissemination, decision-making tools. While the authors might think that the forecast is the entire product, dissemination included, readers of HESS might think that the forecast is just the raw output of the model. This needs defining in the article.

A: Is used in a similar way as "student resources" and there is no need in this grammatical sense for apostrophies. In the revised version we will carefully define the terms forecast and forecaster.

C: P2220 Section 2.3 I think this section would benefit, given the discussion later in the paper, by having more information about the flow of information in EFAS. E.g. Are both the

NWP and hydrological models ensemble-based? How many ensemble members in each? Are these multi-model or initial condition ensembles? (this is important due to the discussion about including a multi-model system). Who runs the forecasts operationally? How is this information passed on to the end-users / professional forecasters? Can you describe how the forecast information is currently used for decision-making?

A: We apologize for the shortness of the description which was deleted as most authors felt that the information is published multiple times elsewhere. We will rectify this and add it again. The forecasting system is currently only based on NWP ensembles and in total 156 ensembles are computed each day. Initial condition uncertainty is only partially addressed through the use of different high resolution forecasts to bridge the temporal gap between observations and forecast start time. Please also note that EFAS uses a Bayesian type post-processing at locations where real-time measurements are available accounting for multiple sources of uncertainty (see Bogner and Pappenberger 2011)

The system is executed by a consortium out of several partners (please see www.efas.eu for details). Communication and usage of information is discussed above. More in depth technical description can be found on <http://www.efas.eu>

Bogner, K., and F. Pappenberger (2011), Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system, *Water Resour. Res.*, 47, W07524, [doi:10.1029/2010WR009137](https://doi.org/10.1029/2010WR009137).

C: P2221 Line 7 Should be 'to start off' not 'to start of'

A: Corrected in the revised version

C: P2226 Section 4.2. You say that the category for improving existing decision-making tools was unpopular, and that this could imply that the tools available today are sufficient. Are the tools you refer to the ones available as part of the EFAS forecast, or those internal to the forecasters that EFAS is disseminated to? I think this point needs further discussion. Perhaps they could be unpopular because these tools are decision-maker specific, or the end-users of the EFAS forecasts think that these decision-making tools are their responsibility? Maybe they wouldn't want to admit that their decision-making tools are inadequate?

A: We apologise that this was not clear, the issue of improving the dissemination tool is deemed important, as it ranked among the top then, which is also mentioned in the paragraph. We will add further discussion. Regarding the existing tools, the EFAS system uses an interface which is available to all partners of EFAS. The partners are critical and constructive and we feel that they would openly voice their opinion. Since EFAS went operational the centres have taken a very active role in suggesting improvements to the system, which would imply that they take on the responsibility to improve the system

C: P2226 Line 21 It is not clear that the sentences from 'Priorities of a more technical nature' onwards belong in this section.

A: Agree, the priorities were mixed up, this will be corrected.

C: P2227 Line 11 change 'it virtually' to 'it is virtually'

A: Agree

C: P2227 Line 22. You state previously in the paper that the most important priority being the inclusion of more models “could be governed by the wish to include the forecaster’s hydrological model of choice, and not necessarily the idea of a full uncertainty system.” So do you think this outcome is as a result of a forecaster wanting their ‘model of choice’ or a desire for a better quantification of uncertainty? If the latter, does this not contradict the conclusions of Pappenberger et al. 2013. ‘Visualising probabilistic flood forecast information’? Do you think that the EFAS users have changed their minds in the course of a year, and what may have influenced this?

A: We do not believe that this is not a contradiction. The forecasters within the EFAS community embrace uncertainty and want it captured, which was clear in Pappenberger et al 2013 and further supported here. A multi-model system may be one way to address this, which is a popular hypothesis presented to the forecasters in many presentations. However, operational forecasters may also have other underlying reasons for multi-models for example, they would like to include a model as they use it in their national systems and have greater experience with it (the issue of *trust* and *familiarity* is an important in operational forecasting

C: P2228 ‘It is a more important priority for the forecaster to improve the forecast in the medium range’ Was this a result found in this study? Perhaps needs rephrasing to clarify.

A: Yes, this was found in the study since we asked both short-range <3 and medium range >3 days. This information is found in the appendix, but we will clarify in the text.

C: P2229 Line 8. “Although, the priority to standardise hydrological data format was selected as least popular out of the five topics pitched, one should not forget that it was actually a top five contender (thus seen as more important than other topics)” I’m not sure that it is strictly true to say that it was in the top 5 – other groups might have had a second favourite idea that they didn’t go with. I think “one should not forget it was actually a top five contender (thus seen as more important than other topics)” should be changed to “one should not forget it was chosen as a topic to be pitched in the first place”. Or something like that.

A: Agree, we will change the formulation.

C: Figure 2: I can’t read the text in this figure

A: Graphic layout will be improved.

C: P2229 Line 9 Change “this point to” to “this points to”

A: Agree

C: P2229 Line 13. Is this for political reasons?

A: Could be, but we can only speculate on the reason for this. Perhaps they are happy with having the separation between national forecasts and EFAS.

C: P2229 Line 20. What do you mean by generally improved forecast skill? Do you mean a general improvement in the forecast skill, or an improvement to the general forecast skill? I

think the two are subtly different. How is forecast skill currently calculated? How is skill currently communicated? Does this need to change?

A: We do not understand how they are subtly different - can you please clarify?

Skill is measured with respect to the issuing of warnings in a contingency format (hits, misses, false alarms) as well as through various numerical performance measures (see Pappenberger et al. 2011). Various skill scores are also published in Bimonthly EFAS bulletins (see <http://www.efas.eu/efas-bulletins.html>)

However, we do believe that there is a more fundamental discussion on the meaning of skill and its adequate representation, which cannot be addressed in this paper. There has been a detailed discussion already in terms of calibration/evaluation of hydrological models, see e.g. Schaefli and Gupta, 2007 or in meteorological forecasting see Murphy and Katz, 1985 (amongst many others). Indeed this is a focus of our recent research which will be presented at the 13th EMS Annual (09 – 13 September 2013 | Reading, United Kingdom) in the session by M. Zappa, B. Ahrens and H. Formayer on "*Interfacing hydrological and meteorological models in forecasting systems*" (details see <http://wp.me/p3tVcm-cu>) with the topic "*No Skill, Fake Skill and Real Skill*" - please come for a more detailed discussion or wait for the correspondent blog on <http://www.hepex.org>

Pappenberger, F., Thielen, J., del Medico, M., 2011, [The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the European Flood Alert System](#), *Hydrological Processes*, 25(7), DOI : 10.1002/hyp.7772, 2010

Schaefli, B. and Gupta, H. V. (2007), Do Nash values have value?. *Hydrol. Process.*, 21: 2075–2080. doi: 10.1002/hyp.6825

Murphy, A. H., and R. W. Katz, 1985: *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. Westview Press, Boulder, CO, Sherden, W., 1998: *The Fortune Sellers*. Wiley, New York,

C: P2229 Line 24. Is there much difference in the time taken for these priorities to be implemented too?

A: Yes, high complexity implies longer time scales

C: P2230 Line 15. Is it worth including links to the training activities of EFAS and HEPEX?

A: We will include references to the HEPEX activities. The issue of training is described above and will be included as clarification in the paper.

C: P2230 Line 15. Change to 'these training and collaboration efforts need to be'

A: Agree

C: P2250 Line 19 Change "Could be on way" to "Could be one way"

A: Agree

C: P2230 Section 5.2. Would forecast verification scores be understood by the EFAS end-users? Does there need to be work addressing what scores they are most interested in, or training them in forecast evaluation? Is there data for assessing long-term performance at specific points, particularly for flood events which are rare. Would this cause a discrepancy

between what the end-users want to know about forecast performance and what information can actually be provided to them? “Research activities are already initiated to address this issue” - Can you provide more detail on this?

A: We take extensive care in trying to compute scores which are mathematically sound, focus on our end-users needs and can be understood. This will be further explained in the revised version. For any type of score or performance measure there is a need to have them explained and support interpretation. Training needs in this area are extensive and there is a risk that they can be interpreted wrongly. There is the additional issue that flooding is an extreme event and hence sample size is extremely limited with all its implications on a robust evaluation. Many institutions go to great lengths in trying to present the evaluation of extremes in a proper and scientifically correct manner (e.g. ECMWF has established multiple working groups to this topic). Regarding the available data, the short answer is no, there is not enough information on flood events to calculate all the scores we would like or need. The EFAS system is at least on a continental scale and allows us to pool data from different locations (with all the pitfalls associated to it). Our research activities are trying to address this issue by combining the presentation of scores, with training at meetings or through the EFAS bimonthly bulletin and evaluation of case studies.

C: P2230 Line 6 should be NWP's not NWP's

A: Agree

C: P2230 Line 17. “A number of projects.” Can you give examples? Are these in Europe?

A: Will include references. One ongoing example is GEOWOW (<http://www.geowow.eu/project.html>)

C: P2231 Line 21 Since you refer to forecast dissemination, I think it would be important to state somewhere in the article how the forecast is currently disseminated.

A: Will be done in the more detailed description of EFAS

C: P2231 Line 22 “decisions” not “decision”

A: Agree

C: P2232 Line 2 Change “are directed” to “should be directed”

A: Agree

C: P2232 Line 5 Change “are concerning either to improve the physical representations in the used models and improve the forecast...” to “concern whether to improve the physical representations in the adopted models or to improve the forecast on lead times >3 days” Is there any indication of which processes should be better represented physically?

A: We believe both issues are important and they do not stand against each other. We will rephrase to make this clearer. On the wish list are for example ice jams and snow melt - both very difficult to represent. We will add some more examples of physical processes which need better representation.

C: P2232 Line 6 Is there any indication of what lead times would be useful to the decisionmakers? E.g would it be helpful for them to have have skillful forecasts at 5 days or 15 days?

A: Needs across Europe are incredibly diverse and the role/duties of one national agency is not equivalent to another. EFAS was perceived as a tool of to provide forecasts between day 3 and 10 - this is the skill base we focus on. Even within a single country or agencies different lead times maybe desired and value will be different. Operational procedures will adapt to the maximum lead time at which skill full forecasts can be provide. Studies are performed by individual countries to establish such value (see aforementioned Environment Agency study).

C: P2232 Line 11 change “both to the hydrological and meteorological” to “both to the hydrological and the meteorological”

A: Agree

C: P2232 Change “It was developed from the opinions of a large group of professional flood forecasters in which we consider the best way to improve existing operational flood warning systems” to “It was developed from the opinions of a large group of professional flood forecasters on the best ways to improve existing operational flood warning systems” or similar.

A: Agree

C: P2232 Line 20 I think the sentence beginning 'Other areas...' could be expanded to provide a better explanation of what is meant by past performance and uncertainty assessments (e.g. are these not the same thing in a way?) and what you mean by multi-model approaches, especially in relation to EFAS. (e.g. will you be including more NWP models or more hydrological models or more ensemble members?)

A: Will be clarified in revised version.

C: P2232 Line 25 Should this be changed to “and will be dealt with accordingly”

A: Agree

C: P2232 Line 8 I'm not sure what first point, second point, etc. refers to, not the numbered list at least?

A: The first second etc refers to the numbered list. This should be clear enough.

C: P2232 Line 11. Could you expand on 'include more NWPs'. Do you mean include models from different NWP agencies, or do you mean include more ensemble members?

A: We are using all ensemble members that are available for ECMWF and Cosmo-LEPS. We are also using the deterministic models from ECMWF and DWD. The question is regarding including more models, since the precipitation fields can differ significantly between different models - we will clarify this.