Supplementary material

S.1 Pitching your chosen priority in front of a jury

The first part of the exercise was designed to encourage the meeting participants attending the workshop to come up with areas that would need to be prioritised in order to improve flood forecasting and to present one area of priority in front of a scrutinising jury of "expert forecasters", modelled after the popular TV series *Dragon's Den*^{TM}. A total of 30 participants from 15 institutions attended the meeting, the representatives being operational forecasters, operational managers or researchers involved in developing forecast tools.

Each participant was given written directions one day in advance to define their research priorities for EFAS development. In the directions, the participants were asked to consider what their priorities were, including a brief description, and also to identify why this was their choice. The following day the participants were randomly divided into five groups and given the instructions that each group had 45 min in order to prepare a 5 min pitch for the one research priority which they all considered to be the most important. The first task for the group was to agree on which topic they would advocate in the pitch.

In the next stage, each group pitched their priority in front of the rest of the participants, including the panel of five "experts", each of whom asked one question following the pitch. After all of the groups had made their pitch, each participant was given 10 Swedish kronor to represent money available for investing in the presented priorities. They were then to reward the priorities that they thought most worthy of investment by putting a voluntary sum of money in boxes, which represented the five presented priorities. They were also given the option not to invest in any priority and keep the money for themselves. The group that had the best pitch based on the financial investment of the participants was crowned as finding the most important research priority, and was rewarded with a prize.

S2. Questionnaire

When analysing the priorities, five categories emerged: (i) cooperation, training and dissemination; (ii) improved tools for decision making; (iii) improved skill of forecast; (iv) new tools to evaluate and compare forecasts; and (v) data collection and processing (Table 5). The priorities were put into each category (for a full list see Table S1). The first question in the questionnaire was to rank the five categories listed above in importance from 1 to 5, where 1 was the most important. Respondents were then asked to rank a further 23 priorities according to their importance as “very important”, “important”, “neutral”, “not so important”, “unimportant” or “no opinion”. The categorisation was not visible for the respondents. The questions were asked in random order so as to not bias the results towards a certain category. The response frequency was 83 %.

S3. Results from the two exercises

The results from the first gauging of the forecasters priorities (blue line, Fig. S1) differ somewhat in comparison with the results from the individual questions when they are summarised according to category (black line, Fig. S1). In both cases improving the general performance of the forecast (category 3) is seen as most important, but “better tools to evaluate and compare forecasts” (category 4) and “improve data collection and processing” (category 5) both become more important when the individual questions are summarised in comparison to the initial ranking in the first question. Also, “more cooperation, training, workshops, etc.” (category 1) is seen as important in the individual questions, but not in the first question of the survey. The mean of category 1 suffers from the unpopularity of question 2 (dissemination and communication through social media). The dotted line (Fig. S1) shows that this category would rank as the most important with the results from this particular question omitted. The difference in results from the first ranking question in comparison with the others from part 2 could reflect the fact that the forecasters have a predetermined view that increasing forecast skill is the most important way to improve in the forecast chain. However, this is challenged when the results of the individual questions are ordered in their respective category. Here other areas emerge as more important, such as better communication and training, and the need for a tool to assess the general skill of the model.

We would note that the results from part 2 (survey) are not independent of the results in the part 1 (workshop exercise) since the participants already had the priorities presented to them, and they were for obvious reasons deemed important already. Also, there was some time to digest the discussions and results from the exercise, and the votes after the presentation in the *Dragon’s Den* should be considered a first guess.

In part 1 there was very limited time given to prepare the presentations (1 h), and this could have had an influence (although perhaps positive) on how each group selected their respective priority. In some groups there was a thorough discussion, followed by a vote of the most popular priority to be put forward as their pitch. This often led to time constraints in the preparation of the pitch. Other groups quite quickly settled on the most important issue, and had time to prepare the presentation. Furthermore, language barriers, the composition of each group, and particularly dominant individuals could have affected the choice of pitch from each group.

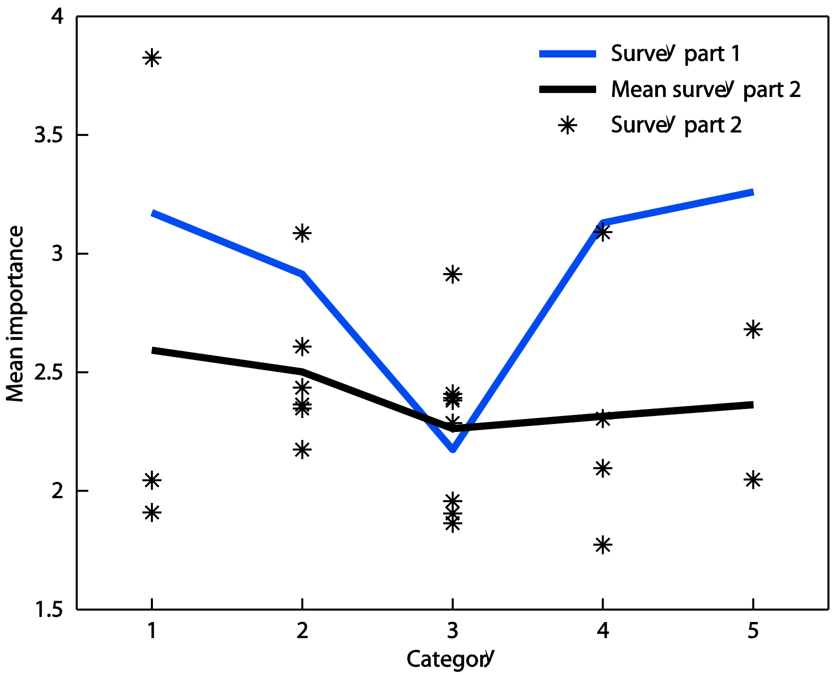

Figure S1. The black line portrays the results from the ranking of the different categories and the blue line the results of the second part of the survey with the questions ordered according to the categorisations. The dotted line denotes the mean of the questions in category 1, excluding question 2. The figure also shows the importance of the individual answers according to their categorisations.

Table S1: The priorities that were collected during the discussions and sent out in the questionnaire. They have been grouped according to their category (Cat). The last two columns denote the importance and their individual rank according to the survey results.

Number	Cat.	What	Brief Description	Why	Importance	Rank
1	1	Building a European Flood Forecasting infrastructure	Strengthen the operational European Flood Forecasting Community for example by fostering more knowledge exchange between the EFAS partners through organisation of workshops, staff exchange and other outreach activities.	Medium range ensemble flood forecasting is fairly novel in many countries and agencies. In particular operational decision making is in its infancy – better knowledge exchange will lead to good practise around Europe.	2.05	6
2	1	Replace/expand web forum by social networks	EFAS should be part of the social networks, such as twitter, LinkedIn, Facebook.	It would increase the exposure and make it easier for doing news groups, improve communications	3.83	23
3	1	Education and training of how to use and interpret forecasts	For example with dialogues, exercises and training courses for civil protection agencies, local authorities and forecasters. In particular train more young hydrologists.	More training will increase understanding and utility of medium range ensemble forecasts	1.91	4
4	2	Increase the frequency of forecasts	This option would make it possible to increase the number of forecasts for example from 2 to 4 times a day.	This would provide you with more up-to-date information during flood situations	3.09	21
5	2	Thresholds for warning levels (translation to	Homogenization of threshold definitions among basins (quantile differences) to real return periods	This would allow for comparable thresholds across Europe for all rivers and basins	2.35	12

		return periods)				
6	2	Rapid risk and hazard maps based on EFAS forecasts	Combining EFAS forecasts with local flood hazard maps and rapid risk maps could provide more information on the potential hazard associated with predicted floods.	This would allow a better decision on the priorities of actions to prevent potential hazards	2.61	18
7	2	Improve the forecast dissemination	Invest to improve the delivery of forecast to the partners. This could include smart phone applications or web technology to include EFAS forecast directly in partners standard forecast interfaces and software by distributing GIS shape files or other WMS services. For example, there are many “spatial” outputs that might be further used in more detailed evaluation on national level especially in an alert situation	Better forecast dissemination will allow partners to have access to EFAS forecast from within their own system and thus increase usage.	2.17	9
8	2	Improve the visualisation and product generation	This investment is to improve forecasts are presented (e.g. make the interface more configurable), include additional products (for example satellite images) or other derivative products (e.g. specific runoff, soil saturation or SWE forecasts)	Better visualisation will lead to better decision making. New products can be used as auxiliary information to national and international services. Should be georeferenced and downloadable	2.43	17
9	2	Flash flood guidance	Evaluation of soil-saturation to better estimate levels of dangerous heavy	Useful tool for flash-flood warnings	2.36	13

			precipitation (<6 hourly). To be used in combination with flash-flood warnings			
10	3	Increase the average skill of the medium range forecast (>3 days).	This would mean that a forecast for day 5 would become as skill-full as a forecast for day 4 with the present system	A more skilful forecast will be a more useful forecast	1.90	3
11	3	Increase the average skill of the short range forecast (<3 days).	Invest in flash flood capability. For example develop a better Flash Flood Guidance based on evaluated saturation of soil and computation of "dangerous precipitation" comparing rainfall forecasts with observations from national networks	This would mean that EFAS develops more capability in the flash flood forecasting	2.38	14
12	3	Increase the temporal resolution of the forecast	For example apply hourly (30 min) time steps in the forecast instead of 6-hourly time steps	Higher temporal resolution will give more detail on the time a threshold is exceeded	2.91	20
13	3	Increase the spatial resolution of the forecast	Give prediction on smaller areas and smaller river basins For example EFAs will provide information on a $1km^2$ grid rather than $5km^2$	Higher spatial resolution will give more information on smaller catchment scale	2.39	15
14	3	Improve physical model representations	Improve model representation of hydrological structures to improve model performance. For example, a better representation of snow water	Hydrological structures often dominate flood response and a better representation will be crucial for an improved forecast. Large parts of Europe suffer particularly under snow melt	2.29	10

			equivalent or evapotranspiration.	driven floods. Any improvement will be largely beneficial to the skill of these floodings.		
15	3	Include reservoir management	Build a model in order to capture reservoir management, defining output flows in case of the reservoir would be empty or full.	In Europe there are many rivers that are heavily regulated.	2.41	16
16	3	Introduce more NWP ensembles for meteo input	Scientific literature shows that a multi-model approach with a grand ensemble of NWP ensembles increases the scores	This will lead to a more robust modelling system with better estimation of the uncertainties, especially regarding the rainfall forecasts.	1.96	5
17	3	Introduce multi-model approach for hydrological modelling	Scientific literature shows that there are many challenges in hydrological modeling that cannot be solved by single models. Multi-modeling systems provide one efficient solution to for have better understanding of the spatio-temporal characteristics of catchments	This will lead to better forecasts and hence be beneficial to everybody who uses the forecast. It will capture variability and uncertainty better. Different models model processes differently.	1.86	2
18	4	Distinguish between different flood situations	Give information about the type of flood expected, whether it is a snow-melt driven spring flood, due to extreme weather (fluvial) or long-term raining related to ground water (pluvial)	Would make decisions easier on the action needed and the risk associated	3.09	22
19	4	Report past performance for	Displaying selected statistics of forecast system performance at station level	Information about the past (climatology, mean bias, anomaly RMSE and anomaly	1.77	1

		the hydrological and meteorological forecasts	performance at individual stations both “Hind cast skill” for particular area, cross-section, point, season, forecast lead time as well as the most recent performance to detect “forecast busts”	correlation, ensemble spread and signal to noise ratio) is crucial to establish trust in a system and understand weaknesses and strengths in certain situations		
20	4	Increase the historical time series	Collect data from larger time series to get better watches or alerts.	EFAS warnings are based on short time series.	2.30	11
21	4	Changing the way probabilities are calculated/presente d	What is presented today as probabilities are strictly speaking the modelled frequencies of predicted floods. The calculation of probabilities could be done more robustly	It would improve the estimation of the real flood probabilities.	2.10	8
22	5	Blending of national and EFAS forecasts	Creation of a seamless forecasting system in which national short range forecasts (1-2) days build an intrinsic part of the medium range system (EFAS).	Flood forecasting is a continuous process in terms of lead time. Although different type of systems will have different strengths and weaknesses a seamless merging approach would make it easier to make decisions on all available information. It would enable to further include EFAS to national system	2.68	19
23	5	Improve standardization of hydrological data	There is a plethora of data formats used in operational and real-time hydrology not only for measured and collected data, but also for model data. Data provision	Better access to data will lead to better forecasts as verification, calibration and updating tasks are made easier.	2.05	7

guidelines to standardize formats which will improve the forecasts, for example by promoting the INSPIRE standards.