Supplement of Hydrol. Earth Syst. Sci., 26, 2899–2922, 2022
https://doi.org/10.5194/hess-26-2899-2022-supplement

Hydrology and Earth System Sciences
Open Access

Supplement of

The role of multi-criteria decision analysis in a transdisciplinary process: co-developing a flood forecasting system in western Africa

Judit Lienert et al.

Correspondence to: Judit Lienert (judit.lienert@eawag.ch)

Contents

1 Supplement: Methods

1.1 Generating FANFAR system configurations

The FANFAR system configurations consisted of different possible configurations of the FANFAR flood forecasting and alert system. These were defined using several participatory approaches during the first FANFAR workshop (Niamey, Niger, Sept. 2018). We utilized the Strategy Generation Table (Gregory et al., 2012;Howard, 1988) to co-design possible system configurations for the two FANFAR platforms "Hydrology Thematic Exploitation Platform" (Hydrology-TEP or H-TEP) and the "Interactive Visualization Portal" (IVP) together with the workshop participants. The Hydrology-TEP is the ICT environment in which the forecast production system is deployed to produce new forecasts and alerts every day. In the FANFAR co-design workshops, this is primarily intended for experienced hydrologists and ICT experts. The IVP is the web visualization interface of FANFAR (displaying outputs from H-TEP), which is normally accessed by any user. All emergency managers and most of the participating national hydrologists interacted with the FANFAR system through the IVP in the co-design workshops and in between.

To prepare the workshop, we defined different system elements of the H-TEP and IVP (e.g., forecasted variables, observed variables, distribution channels, etc.). Possible characteristics of these elements were developed by experts of the FANFAR consortium (e.g., possible characteristics for the element "forecasted variables" were river discharge, water level, precipitation, etc., and different combinations). During the "Strategy Generation Table" session, we presented these elements and their possible characteristics to the participants. We pre-defined three strategies, for which we asked the participants to discuss and decide on suitable characteristics for each system element, one by one: 1) "The most easy-to-use system", 2) "The most attractive system for West Africa", and 3) "The most robust system that works in all sorts of current realities in West Africa". For instance, frequent electricity shortcuts may occur, or internet connection may be poor or unstable. For the H-TEP session specifically, we defined one additional strategy: 4) "The system that requires the least resources for West African end users" (e.g., skilled personnel, good internet connection, good / constant power supply).

Additionally, we performed a combination of "Brainwriting 635" (Litcanu et al., 2015;Paulus and Yang, 2000) with the "Cadavre Exquis" game. This consists of a collection of words, written on a paper by a participant. It is then folded and handed over to the next person. The aim of this session was thus to also interactively develop additional system configurations for the IVP using the same strategies as in the Strategy Generation Table sessions. The Brainwriting 635 method allowed the participants to extensively discuss the system characteristics before deciding. In contrast to the Strategy Generation Table session, the participants decided alone on the characteristics of one single system element, independently of the other elements, only knowing the strategy. These results were used to consolidate the configurations developed with the Strategy Generation Table.

After all three parallel sessions, results were discussed in a plenary session. As post-processing, the strategies from all sessions were combined to characterize the entire FANFAR system, i.e., combining both the IVP and H-TEP part. Two experts from SMHI and Eawag later discussed and defined further system configurations to cover technically interesting system configurations.

1.1.1 Strategy Generation Table

We adapted the "Strategy Generation Table" method from Gregory et al. (2012). Each session was moderated by two FANFAR consortium members. The moderators prepared big sticky notes with possible system configuration "elements" and smaller sticky notes with characteristics for each "element". The moderator and participants built a table with the big sticky notes (elements) as headers, and possible characteristics for each element in rows (example see Table S1). The system configuration elements were discussed independently from each other (i.e., characteristics of column A were created independently of column B).

Table S1. Example of Strategy Generation Table: system configuration elements (columns A – D) and corresponding characteristics (rows a – e belonging to each column) to generate suitable system configurations for the Interactive Visualization Portal. Note: an unequal number of characteristics can be used. This example is a shorter representation of the table actually generated in workshop 1.

	A Forecasted variables	B Observed variables	C Model performance / accuracy	D Data download	E Distribution channels	F Language
a	River discharge	None	No performance metrics shown	No option to download data	Website with interactive features	English
b	River discharge & water level	Water level from satellites	Display performance metric for forecasts	Tabular data for selected station in TXT format	Website with static images	English and French
c	River discharge, water level, & precipitation	Water level from in-situ measurements and satellites	Blank out areas where forecasting performance is too low	Tabular data for selected station in Excel format	Website with text summaries only	English, French, and Portuguese
d	River discharge, water level, precipitation, & evaporation	River discharge from in-situ measurements		Map of displayed variable(s), in PNG format	SMS alert notifications	English, French, Portuguese, Arabic
e	River discharge, water level, precipitation, evaporation, soil moisture storage	In-situ water level and river discharge, and water level from satellites		Map of displayed variable(s), in Shapefile format	Email alert notifications	

In the H-TEP session, the following system configuration elements were discussed:

- **Input data**: "diversity of observational data sources", "data processing options (input data)", and "meteorological input / forcing data
- **Related to the models:** "number of models used"
- **Outputs:** "type of information derived / forecast outputs from HTEP"
- **Distribution channels:** "distribution options from TEP"
- **Additional tool features:** "degree of automatization", "language", and "support system".

In the IVP session, the main elements were: "Forecasted variables"; "Observed variables"; "Model performance / accuracy"; and "Reference thresholds to compare with". Additional tool features were also discussed, namely "Data download"; "Distribution channels"; "Language"; and "Alert notification system".

Thereafter, workshop participants formulated strategies to define different FANFAR system configurations. In this context, a strategy is a logically consistent set of characteristics (of every element) that are combined to create a comprehensive system option (i.e., FANFAR system configuration) through a recognizable question or driver (Howard, 1988). Thus, for each strategy one characteristic from each column was selected. In both sessions, three strategies were used as guiding question (example see Figure S1).

Strategy 1: Most easy-to-use → What is the most easy-to-use system? (What could your neighbor use?).

Strategy 2: Most attractive → What is the most attractive system for West-Africa?

Strategy 3: Most robust → What is the most robust system that works in all sorts of current realities in West Africa?

In the H-TEP session, an additional strategy was created. This "Least resources" strategy refers to a system that requires the least resources for West-Africa (note: not the least resources for the FANFAR consortium or SMHI), e.g., skilled personnel, good internet connection, and good / constant power supply.

The one that can be implemented faster.

The most desired in an ideal world?

	A	B	C	D	E	F
	Forecasted variables	**Observed variables**	**Model performance / accuracy**	**Data download**	**Distribution channels**	**Language**
a	River discharge	None	No performance metrics shown	No option to download data	Website with interactive features	English
b	River discharge & water level	Water level from satellites	Display performance metric for forecasts	Tabular data for selected station in TXT format	Website with static images	English and French
c	River discharge, water level, & precipitation	Water level from in-situ measurements and satellites	Blank out areas where forecasting performance is too low	Tabular data for selected station in Excel format	SMS alert notifications	English, French, and Portuguese
d	River discharge, water level, precipitation, & evaporation	River discharge from in-situ measurements		Map of displayed variable(s), in PNG format	Email alert notifications	English, French, Portuguese, Arabic
e	River discharge, · level, itation, ıtion, soil moisture storage	In-situ water level and river discharge, and water level from satellites		Tabular data in TXT format and map displayed in PNG format	Website with interactive features, SMS and Email	

Figure S1. Example of creating strategies to generate system configurations for the FANFAR Interactive Visualization Portal. Examples of system configuration: "The strategy that can be implemented fastest" (red); "The most desired strategy in an ideal world" (green).

1.1.2 Brainwriting 635 combined with Cadavre Exquis

The session combining "Brainwriting 635" (Litcanu et al., 2015;Paulus and Yang, 2000) with the "Cadavre Exquis" game was also moderated by two people. There were 17 French speaking workshop participants. This group aimed at creating plausible FANFAR system configurations of the Interactive Visualization Portal (IVP). The approach followed the same structure as the "Strategy Generation Table" groups. Brainwriting 635 was used as interactive brainstorming method to develop characteristics for each element of possible FANFAR system configurations (see sect. S1.1.1 for details concerning elements and characteristics). Session participants were split into three smaller groups. Each group discussed six elements: "Observed variables", "Model performance / accuracy", "Forecasted variables", "Reference thresholds to compare with", "Representation of variables", and "Distribution channels". After this part, a speaker was elected in each group to highlight the results to the other groups.

Thereafter, the Cadavre Exquis game was used to generate system configurations for the IVP. The same three guiding questions were used as for the Strategy Generation Table sessions (sect. S1.1.1). Each group tackled only one question. The characteristics developed and discussed during the Brainwriting 635 part (with the exception of "Model performance / accuracy") were now selected in such a way that they were suitable for the guiding question, e.g., "What is the most attractive system for West-Africa?" For this, each participant selected the most suitable characteristic for one element, wrote it on a sheet of paper, and handed it to her neighbor. This person now selected the most suitable characteristic for the next element, and handed the folded paper to his neighbor. The Cadavre Exquis game is thus a structured way of asking each participant to choose a characteristic for the next system configuration element without knowing the previous selections (they only know the strategy).

1.2 Eliciting weights

We elicited weights in five groups in the second FANFAR workshop (Table S2). More information see main text.

Table S2. Overview of stakeholder groups and weight elicitation method used during the second FANFAR workshop. Group-ID abbreviation: G1 – G5 = group number, A / B = sub-group, EM = emergency managers, HY = hydrologists, F / E = language.

Group-ID	Stakeholder type	Language	No. of participants	Method
G1A_EM_F	EM	F	8	Swing
G2A_HY_F	HY	F	Sub-group of total 11	Swing
G2B_HY_F	HY	F	Sub-group of total 11	Swing
G3A_HY_E	HY	E	14	Simos card
G4A_EM_E	EM	E	3	Simos card
G5A_AGRHYMET	AGRHYMET	F and E	3	Simos card

1.2.1 Swing weight elicitation

We used a bottom-up hierarchical standard Swing weight elicitation method (Eisenführ et al., 2010) for the French speaking workshop participants. We describe the procedure here in some detail for readers not familiar with Decision Analysis methods. We first carried out the weight elicitation on the lower-level objectives of the objectives hierarchy, one branch after the other until all four branches were covered (i.e., two sub-objectives belonging to "High information accuracy and clarity", three from "Good information access", two from "Low costs", and three from "High sustainability"). In Swing, the stakeholders first order all objectives in question (i.e., those from one branch) in terms of importance, answering the question: "Imagine that all objectives are on their worst possible level. You can now choose to "swing" one to its best level. Which objective is most important to you to improve to its best possible level?" After setting all objectives back to the worst level, the second most important objective is moved to the best possible level, and so on. Note: it is very important that the stakeholders understand the objectives and attributes, which need to be explained, along with the ranges (worst and best possible case). After ranking objectives, they are rated. Per default, the hypothetical option, where the most important objective is on its best level, and all others on their worst levels, receives 100 points. The hypothetical option with all objectives being on their worst level receives 0 points. The stakeholders are asked to assign points to the hypothetical options in between, which reflect their preferences. For instance, if they assign 50 points to the hypothetical option where the second most important objective is on its best level, it is half as important to them to improve this objective to its best level, compared to the most important objective.

After having done this for each hierarchy branch, we repeated the procedure for the higher-level objectives *across* the hierarchy. We used the most important lower level objective of each branch identified in the first step. To check for consistency, we repeated the procedure across the hierarchy for the higher level objectives using the second most important lower level objectives.

Stakeholders can be uncertain about their preferences or in group sessions, there may be disagreement among group members about the exact numbers. Therefore, we allowed the participants to state a range of swing points representing their uncertainty. We took the mean of these ranges as main preference statement and considered the ranges in the later sensitivity analyses.

We transformed the swing points to local weights $\boldsymbol{w_r}$ (i.e., the weight of the lower level objective relative to the other lower level objectives within the same branch; eq. S1; note that the sum of all local weights in one branch equals 1), using following formula (Eisenführ et al., 2010):

$$w_r = \frac{t_r}{\sum_{i=1}^{m} t_i} \quad \text{(eq. S1)}$$

w_r = local weight of objective r; t_r = swing points of objective r; t_i = swing points of objectives within same branch; m = number of objectives within branch

Second, we calculated the global weight W_r (i.e., the weight of the lower level objective relative to all other lower level objectives; eq. S2) by multiplying the local weight of the lower-level objective by the local weight of the respective higher-level objective (Marttunen et al., 2018). Note that the sum of all global weights equals one.

$$W_r = w_{L1_r} \times w_{L2_r} \qquad \text{(eq. S2)}$$

W_r = global weight of objective r; w_{L1r} = local weight of lower-level objective r; t_{L2r} = local weight of higher-level objective which includes sub-objective r

We translated the ranges of uncertainty stated during the assignment of the swing points into weight ranges using the same formula. These were used to inform the later sensitivity analyses. Below, we describe our procedure in some more detail:

The minimum local weight resulted from considering the minimum Swing points for the given objective, and the maximum Swing points for all other objectives within the same branch. Likewise, the maximum local weight resulted from the maximum Swing points of the objective, and the minimum Swing points of the other objectives. A justification for this approach is that sensitivity analyses should exactly challenge the assumptions of the model, in this case the weight preference statements, by covering (plausible) extremes. We calculated the ranges of the global weight for each objective by multiplying the limits of the local weight of the lower-level objective by the limits of the local weight of the corresponding higher level objective.

The minimum global weight ($W_{r_{min}}$) was calculated by eq. S3:

$$W_{r_{min}} = w_{L1_{r_{min}}} \times w_{L2_{r_{min}}} \qquad \text{(eq. S3)}$$

Where the minimum local weight of the lower level objective $w_{L1_{r_{min}}}$ resulted from considering the lowest value of Swing points for that objective and the highest value of Swing points for the other objectives within that branch. Likewise, the minimum local weight of the higher level objective $w_{L2_{r_{min}}}$ resulted from considering the minimum value of Swing points assigned to this higher level objective and the maximum value of Swing points assigned to all other higher-level objectives.

Likewise, we calculated the upper limit of the range in weights for each objective by considering the scenario that results in the highest possible global weight for that objective. The maximum global weight ($W_{r_{max}}$) was calculated by eq. S4:

$$W_{r_{max}} = w_{L1_{r_{max}}} \times w_{L2_{r_{max}}} \qquad \text{(eq. S4)}$$

As stated above, this method results in the largest possible range of global weights given the preference statements of the participants. However, the most extreme global weights of the range ($W_{r_{min}}$ and $W_{r_{max}}$) are multiplications of the extremes of uncertainty of different individual preference statements and therefore are unlikely. The extremes should be interpreted as boundary of possible uncertainty and serve as qualitative indication of which objectives for each stakeholder group require a sensitivity analysis on weights.

1.2.2 Simos' revised card procedure

In three sessions, we used an adaptation of the revised Simos' procedure to elicit weights with cards (Figueira and Roy, 2002;Pictet and Bollinger, 2008). Again, we describe the procedure here for readers not so familiar with Decision Analysis methods. First, we presented the different objectives including worst/best cases (range) to the participants on sticky notes, and asked them to stick the notes onto a blank wall ordering them side-by-side according to their importance. Note that we did not introduce numbers or give any signals that could lead to an obvious bias, e.g., numbering the objectives in whatever predefined order. Whenever there were two or more objectives with the exact same importance, they were put in the same position.

After rank-ordering all objectives on the wall, the moderator assigned a default score of 100 to the most important objective and asked the group members for a score between 0–100 for the least important objective. This allowed to obtain the range (in points) between least important (chosen points) and most important objective (per default 100 points). Later, the ratio of points of the most important objective (100) to the points of the least important objective will be called Z (see below). The Z value thus indicates how many times the most important objective is more important than the least important objective. In this step, the moderator allowed for a range for the score on the least important objective.

Second, we tasked the group members to discuss the difference in importance of successively ranked objectives. They were asked to stick blank notes between objectives to represent the relative differences in importance between successive objectives (i.e., the more sticky notes between two objectives, the larger the difference in importance). The moderator asked questions to check consistency regarding the blank sticky notes, e.g., between two non-consecutive objectives, and the score given to the least important objective, e.g., if the most important objective was x times more important that the least important objective.

To translate the resulting composition of sticky notes on the wall into weights, we first calculated the Z value. This is the ratio of the score of the most important objective (100 per definition) by the score of the least important objective assigned by the group. Thereafter, we first calculated the initial weight O_r of each objective (eq. S5), and afterwards rescaled the weights to achieve the final global weights W_r (eq. S6), which sum up to one (Pictet and Bollinger, 2008):

$$O_r = r_{min} + (Z-1)\frac{r - r_{min}}{r_{max} - r_{min}} \quad \text{(eq. S5)}$$

O_r = initial weight of objective r; Z = ratio of most important objective and least important objective; r = inverse rank; r_{min} = def. 1; r_{max} = maximum rank

$$W_r = \frac{1}{\sum_{i=1}^{n} O_i} \times O_r \quad \text{(eq. S6)}$$

W_r= global weight of objective r; O_r = initial weight of objective r, n =number of objectives

As the groups were allowed to state a range for the score of the least important objective, we translated the min/max boundaries of that range into Z values as well. The moderator decided on a value within the range to focus the analysis on. In addition to the weight set resulting from the focus Z value, we calculated the individual weights using the min/max values of the Z-value. This resulted in a total of three different weight sets, in each set the weights sum up to one. The set resulting from the focus Z value was used for the main MCDA, while the other two sets were incorporated in sensitivity analyses (sect. S1.4).

1.2.3 Observations from weight elicitation informing sensitivity analyses

To examine for which groups and objectives sensitivity analyses should be carried out, we gathered all comments documented by the moderators (Table S3). This allows us to understand, where uncertainties in the elicited preferences are to be expected.

Table S3. Summary of remarks on preferences during weight elicitation and uncertainty for each stakeholder group (SH-Group).

SH-GROUP	SUMMARIZED REMARKS ON PREFERENCE STATEMENTS AND UNCERTAINTY
G1_EM_F (SWING)	**Weight elicitation on lower level objectives of hierarchy:** Consensus on Swing points, uncertainty ranges given on some objectives → sensitivity analysis on uncertainty ranges. **Weight elicitation on higher level objectives (comparison of most important objective of each branch):** There was a long discussion. Result: give equal weight to each top-level objective, because of connected chain in emergency response that needs all four main objectives. They are all equally important. → No uncertainty in this statement; no sensitivity analysis needed. **Consistency check on higher level objectives (comparison of second most important objective of each branch):** This consistency check did not seem to work properly, which was also an elicitation problem. It would have been necessary to discuss inconsistencies to initial weight elicitation, but there was no time. Proposal of moderator (JL): regard first set of weights from initial elicitation as "good set" and do sensitivity analysis with second set from consistency check. → sensitivity analysis with different weight set from consistency check. **Group comments and reasoning concerning preferences on objectives within branches:** ***High information accuracy and clarity***: From the point of view of the emergency managers it is more important to have clear information than highly precise information (this may be different for hydrologists). ***Good information access***: We have to receive the information BEFORE the flood. This is more important than the reliable distribution channels. ***Low costs***: It must possible that we can finance / pay / buy the product; we can wait for a good product. ***High sustainability***: If there are no people in West Africa that can use and maintain the system, the financing is of no use; the system does not help us in such a case.

SH-GROUP	SUMMARIZED REMARKS ON PREFERENCE STATEMENTS AND UNCERTAINTY
G2_HY_F (SWING)	**Weight elicitation on lower level objectives:** Two groups of different opinion for the two sets of objectives within branches "High information accuracy and clarity" and "High sustainability". Consensus on the other two sets of objectives within branches "Good information access" and "Low costs". Both groups gave a range of Swing points for most lower-level objectives. → Sensitivity analysis on uncertainty ranges. **Weight elicitation on higher level objectives (conducted for each preference set for the lower level objectives):** Two groups of different preferences for higher level objectives with different ranking of higher level objectives and different ratings (Swing points) including uncertainty ranges. Not sure if the participants were always the same ones in these groups with different preference sets. Both groups gave a range of Swing points for some higher level objectives. → Sensitivity analysis on uncertainty ranges. **Group comments and reasoning concerning preferences on objectives within branches:** ***High information accuracy and clarity:*** Group 1: the (un-) precise information is really the most important problem we have; Group 2 was a minority who found it more important to have clear risk information ***Good information access***: The ranking was very clear. For objective 2, there was disagreement about the importance (points). Some said: objective 2 is nearly as important as 1 because you really need the information on time; Weight of objective 3: everybody easily agreed. ***Low costs:*** They all agreed. They explained: if we can improve objective 1 (short development time), we can save lives; which is why this objective is more important. ***High sustainability:*** There was a relatively long discussion concerning the order (ranking) of the objectives, which is why the group was split up into two groups. Reasoning of some: we can have all the money we like (objective 1), but if we do not have qualified people, the money is of no use. Others said: with money, we can also employ qualified people; no money, then nothing works. There was no discussion about the third objective, but they wanted agreed on a range.
G3_HY_E (SIMOS CARD)	Some disagreement on certain values: solved by registration of all contributions and specific value assigned by majority rule. One specific participant from NIHSA had a slightly different preference set. The remaining participants had very similar preferences. Z-value range from 5 to 20, focus on 10 (decided by moderator) → sensitivity analysis on Z-value range.
G4_EM_E (SIMOS CARD)	Very small group. Easy to discuss in detail and explain objectives and their ranges well, while focusing on objectives that were difficult to understand. Z-value range from 3.33 to 5, focus on 3.33 (decided by moderator) → sensitivity analysis on Z-value range.
G5_AGRHY-MET (SIMOS CARD)	Bilingual session EN and FR. Small and very fast group. Some participants helped translate objectives and understood them fast. Participants work together and know each other well. Z-value range from 3.33 to 5, focus on 5 (decided by moderator) → sensitivity analysis on Z-value range.

The Swing weight elicitation resulted in many more remarks from the moderator on uncertainty and differing opinions within groups, than from the Simos card method. This is not a property of the respective method but due to the fact that two different moderators elicited the weights. In the Swing process, the moderator actively encouraged the participants to discuss and state their differing opinions, which then were separately documented. In Simos card method, the moderator requested the group to reach a common opinion, and only this consensus was documented.

Because of very strongly differing weight sets of one Swing group, we decided to split the group consisting of French hydrologists into two separate stakeholder groups for the analysis. One preference set was assigned to the group G2A_HY_F and the opposing opinion formed the group G2B_HY_F.

During the weight elicitation of the French speaking emergency managers (G1A_EM_F) there was a consensus of all group members on the ranking and weighting of the objectives on both levels of the hierarchy. However, the consistency check conducted on the higher-level objectives (using the second-most important objective) resulted in a preference statement that was inconsistent with the initial weight elicitation. The time restrictions during the workshop did not allow to discuss the inconsistence with the participants and resolve the issue. Therefore, we split this group into two sub-groups for the analysis: Sub-group 1 with the preference set that resulted from the initial weight elicitation, and sub-group 2 with the preference set for the higher-level objectives, which resulted from the consistency check. For the MCDA analysis, we used the weight set of

Sub-group 1 and conducted a sensitivity analysis with the weight set of Sub-group 2 to check whether the MCDA results differ for the two sub-groups.

In the English-speaking groups (G3A_HY_E, G4A_EM_E, and G5A_AGRHYMET) that used the Simos card method, the moderator requested the group members to reach a consensus within each group. Therefore, there were no differing preference sets elicited within these groups and consequently we did not form sub-groups.

1.3 MCDA aggregation model

We used an unconventional non-additive aggregation model to integrate the predictions and stakeholder preferences in the MCDA (see main text). The properties and advantages of the weighted power mean model are extensively discussed in (Haag et al., 2019). The implications are visualized for two dimensions in Figure S2, i.e., for two exemplary objectives only. Using the additive aggregation model (weighted arithmetic mean) would result in linear isolines (right plot), while a nonlinear model results in curved isolines (left plot). As a consequence, for the power mean aggregation model, if one attribute is on a very bad level (i.e., value is low), the total value after aggregation is also low (red coloring), even if the second attribute achieves a very good level (i.e., high value). At the other extreme, if both attributes each achieve a good level (i.e., high value), the aggregation model is close to linear (blue coloring). In additive aggregation, a poor achievement on one attribute can be fully compensated if the other attribute achieves a good level. The ValueDecisions app (Haag et al., 2022) has the advantage that different aggregation models can easily be implemented, which allows testing the sensitivity of the results to model assumptions.

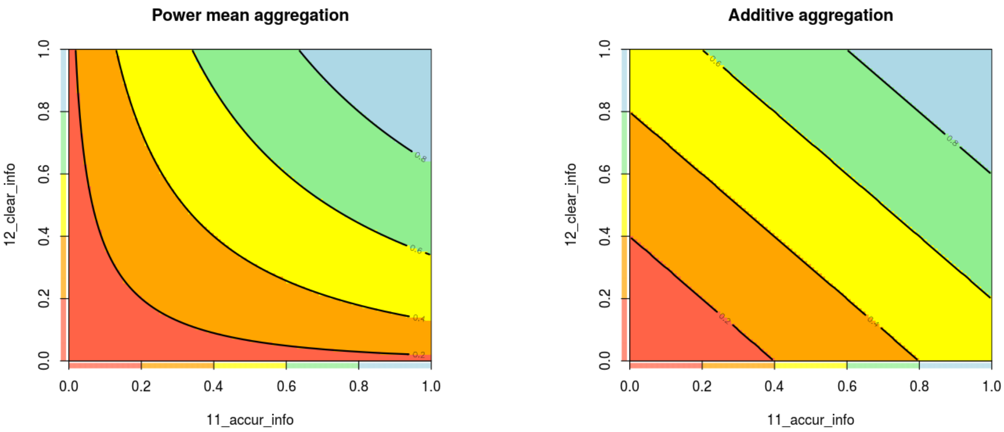


Figure S2. Visualization of the power mean aggregation model with γ = 0.2 as utilized in our MCDA (left) and the commonly used additive aggregation model, or weighted arithmetic mean (right; which can be inferred from the power mean model if γ = 1). Displayed here is a hypothetical case where the two objectives 11_accur_info, and 12_clear_info are aggregated with equal weights. The axes represent the option's (i.e., FANFAR system configurations') achieved value on the two objectives. The isolines (or any other x-y-point in the plot) indicate the option's total value after aggregation of the two objectives.

1.4 Sensitivity analyses on elicited weights and aggregation model

It is common to use local sensitivity analyses to check the sensitivity of the MCDA results to diverging preferences (e.g., Eisenführ et al., 2010;Zheng et al., 2016), or to model assumptions. We checked the sensitivity of our results to other aggregation models, and to other weights, since there was not always consensus concerning the exact numbers among group members of the weight elicitation workshops (see sect. S1.2.3, and Table S3). We give the settings of sensitivity analyses in the Results (**Error! Reference source not found.** main text), with setting S0 as default.

To test implications of the aggregation model (see main text, sect. 2.2.9), we recalculated the MCDA for other reasonable models (Haag et al., 2019), running a new MCDA for each. Applying alternative models requires one click in the ValueDecisions app (Haag et al., 2022).

During weight elicitation, consistency checks are carried out. For the group of French speaking emergency managers (G1A_EM_F), the consistency check revealed an inconsistency for the weighting of the higher-level objectives, which was not possible to resolve during the workshop (Table S3). The difference between the weight sets of these two sub-groups were large, especially for the higher level objectives 2_access, and 3_costs (Figure S3). Therefore, we examined the effect of the alternative weight set (sub-group 2) on the MCDA results in a sensitivity analysis (sensitivity analysis S21).

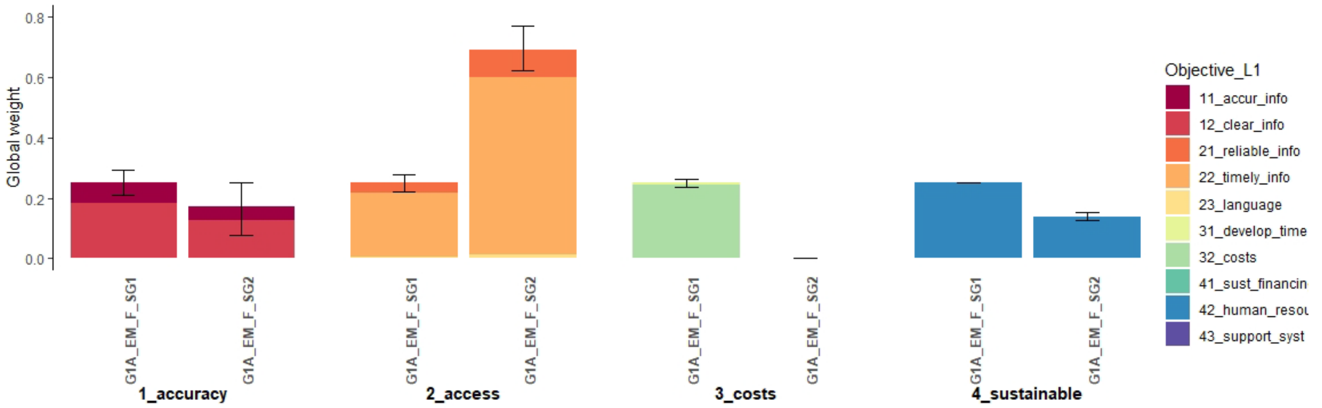


Figure S3. Comparison of global weights (y-axis) from the two sub-groups within the French speaking emergency managers, where the consistency check revealed inconsistent weight preferences. For each higher level objective (e.g., 1_accuracy, 2_access, …), sub-group 1 is depicted to the left (G1A_EM_F_SG1), and sub-group 2 to the right (G1A_EM_F_SG2). Colored stacked bars indicate lower level objectives. Error bars denote the minimum and maxim weights, which result from the ranges in points given by stake-holders in Swing weight elicitation.

Our approach for the further sensitivity analyses depended on the method used to elicit the weights. For Swing weight elicitation, we had allowed the workshop participants to state ranges rather than giving precise numbers if they wished. In this case, a sensitivity analysis towards weight changes of each individual objective was required (sensitivity analyses S22). We defined a threshold of $\Delta = 0.02$ for the ranges; i.e., the difference of the maximum or minimum weight from the average global weight had to be greater than 0.02, else we did not expect a strong effect on the MCDA results. The cases exceeding the defined threshold are highlighted in Table S31. For these cases, we visually investigated whether the respective objective (for the respective stakeholder group) was prone to be sensitive towards weight changes. This is easily possible in the ValueDecisions app, which provides weight sensitivity plots with a mouse click. These plots visualize for which configurations severe changes in values are to be expected if the weight of the objective is altered (for the respective stakeholder group).

An example of these sensitivity plots is provided for the objectives of stakeholder group G2B_HY_F (Figure S4). We identified several cases of weight ranges exceeding the threshold of Δ=0.02. Consulting the weight sensitivity plots, we could exclude three cases (*21_reliable_info*, *22_timely_info*, *41_sust_financin*), which were not sensitive towards weight changes in the corresponding objectives (highlighted green in Table S31). However, we identified two objectives (*11_accur_info* and *12_clear_info*) which seemed to be sensitive towards weight changes.

Thereafter, we individually tested each identified case by recalculating the MCDA. As input into the MCDA model, we utilized the maximum, respectively minimum weight of the sensitive objective, as given by the ranges in the sessions with stakeholders. In the MCDA, the ratios of the weights of all the other objectives are kept constant, and are then normalized to keep the sum of all weights equal to 1. This is done automatically in the ValueDecisions app if new weights are entered for local sensitivity analyses. For a thorough explanation of the method we refer to Eisenführ et al. (2010). Thus, we reran the ValueDecisions app with the new weights and re-calculated the aggregated values of all system configurations (local sensitivity analyses (S22_11_min/max and S22_12_min/max). The results then reveal whether larger changes in the values of configurations and/or rank reversals of the configurations are provoked by these weight changes.

For the three stakeholder groups whose weights were elicited using Simos' card method, the sensitivity analyses on weight ranges were more straightforward. The Z-value was elicited as range (sect. S1.2.2). We recalculated MCDA results for each group, using the two additional weight sets resulting from the minimum, respectively maxim Z-value (sensitivity analysis S231 and S232). For each group, we compared these two additional MCDA results with the one resulting from the weight set of the focus Z-value and identified any major changes in the system configurations' values and any apparent rank reversals. Using the maximum Z value was only necessary for group G3A (S232), because Z min was used as focus value for groups G4A and G5A (i.e., as S0).

It is common to test other interesting objectives by doubling the elicited weight. We did this for "Several languages", because its importance might have been underestimated (S31). For each setting, we compared results (i.e., ranks of configurations) with those of the default MCDA (S0; Table 3, main text), using the nonparametric Kendall's τ correlation coefficient (Kendall, 1938) to measure rank reversals (as in Zheng et al., 2016). As input in the correlation analyses, we used the mean ranks resulting from 1'000 Monte Carlo simulation runs with the changed parameter settings.

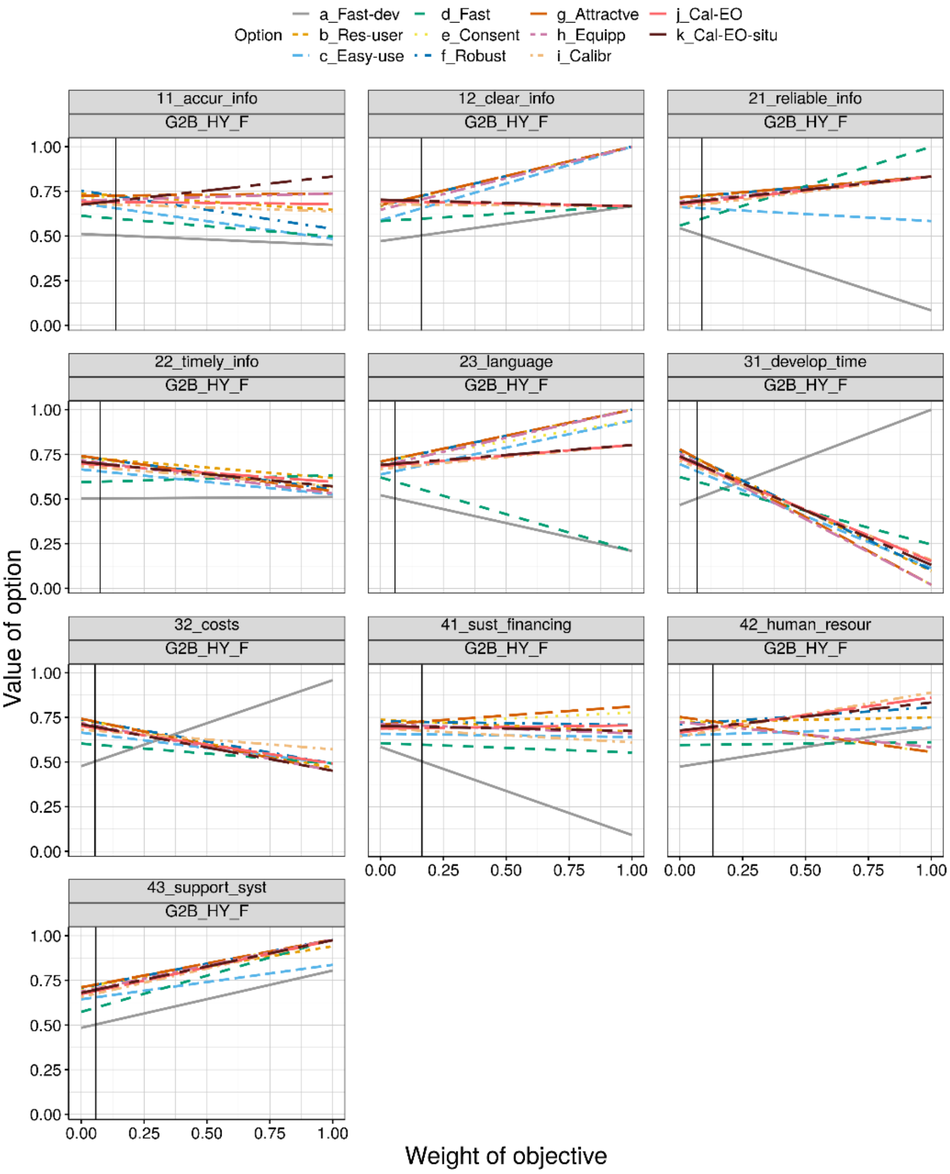


Figure S4. Weight sensitivity plots for stakeholder group G2B_HY_F. For each objective (boxes), the weight (x-axis) can vary between 0 (objective does not count in this decision) to 1 (only this objective is important, all other objectives do not count). Colored lines: FANFAR configurations, and the value (y-axis) that is achieved by an configuration for a given weight. Vertical black line: actual weight assigned by this stakeholder group. If configuration value lines intersect near the vertical line, a sensitivity of the configurations' ranking towards weight change is to be expected. If configuration value lines run parallel or diverge, no sensitivity towards weight changes is to be expected.

1.5 Discuss results with stakeholders, feedback

We organized two half day online rainy season reflection workshops from 21–22 January 2021, inviting the same stakeholders from the 17 West and Central African countries that were invited during the three co-design workshops. The workshops took place on the Zoom platform (https://zoom.us/), with simultaneous translation between English and French. The aim of the reflection workshops was to share experiences with floods and the use of the FANFAR system over the 2020 rainy season.

During the first workshop day, stakeholders were given one hour to fill out an online survey (provided in English and French) on the Limesurvey platform (https://www.limesurvey.org/). Using this survey, we aimed to elicit perceived and expected performance of the FANFAR system. To this end, the survey contained three questions for each of the 10 objectives:

a. How much does FANFAR currently fulfil this objective?
b. Would you use the FANFAR system in the future if it remains as is?
c. What is the minimum acceptable to you? This means: below which level would you NOT use the FANFAR system?

Answers to questions a and c could be provided using a 7-point Likert scale for objective *11_accur_info*, and a 5-point Likert scale for all other objectives. The results generated by Limesurvey were shared and discussed during the second workshop day. In addition to the direct question b, data were analyzed to assess whether the FANFAR system is perceived to meet each objective by assigning descending numbers between 1–5 (or 1–7) to Likert scale answers from best to worst. Then, the answer of c was subtracted from the answer of a for each response (c – a).

2 Supplement: Results

2.1 Stakeholder analysis

Table S4. Overview of the 68 stakeholders that were mentioned to play a role in developing the FANFAR forecast and alert system. The columns relate to the specific questions asked in the survey (see main text and details in Silva Pinto and Lienert (2018)); the task number is given in square brackets, e.g., task number 6 for importance of considering key West African organizations involved in producing and operating flood forecasts and early warning systems in co-design, and 14 for downstream stakeholders, respectively: [6, 14]. SH = stakeholder; Impt. = Importance of considering the SH's needs and interests; Infl. = Influence (power) in the implementation, Affect. = How strongly SH is affected by FANFAR system. The Likert scale for Impt., Infl., and Affect. ranges from 0 ("stakeholder has no influence/is not at all affected") to 10 ("stakeholder decides/ is very strongly affected by"); we present the average. Count = total number of respondents that mentioned this SH in the survey. The survey was completed by 18 respondent groups, usually two (to three) people from the same country, with a total of 31 workshop participants.

Stakeholder [1, 9]	Main interest [3, 11]	Why use system [4, 12]	Information profile	Decisional level	Field	Impt. [6, 14]	Infl. [7, 15]	Affect. [8, 16]	Count
ABN	Resource planning	Forecast refinement	Hydro-innovation	Supranational	Water resources	8,8	7,6	6,2	5
ABV	Resource planning	Forecast refinement	Hydro-innovation	Supranational	Water resources	9,3	7,7	9,3	6
ACF	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	6,0	6,0	5,0	2
ACMAD	Technical	Meteorological data	Hydro-innovation	Supranational	Meteorology	8,3	8,3	8,0	6
AGRHYMET	Technical	Forecast production	Hydro-innovation	Supranational	Water resources	9,8	9,1	8,7	10
ALG	Economic service operations and planning	Alert information	Downstream	Supranational	Agriculture	4,0	4,0	2,0	1
ARC	Economic service operations and planning	Alert information	Downstream	Supranational	Other	10,0	4,0	2,0	1
ASECNA	Economic service operations and planning	Meteorological data	Downstream	Supranational	Transportation	6,0	4,0	5,0	1
CBLT	Resource planning	Forecast refinement	Hydro-innovation	Supranational	Water resources	8,0	8,0	8,0	1
Community	Civil society	Alert information	Downstream	Local	Development	6,6	4,3	10,0	7
CRS	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	8,0	8,0	8,0	1
Dam Management Entity	Economic service operations and planning	Water related information	Downstream	National	Energy	8,2	7,6	8,4	5
ECOWAS	Economic service operations and planning	Other	Downstream	Supranational	Industry and Commerce	10,0	10,0	10,0	1

Stakeholder [1, 9]	Main interest [3, 11]	Why use system [4, 12]	Information profile	Decisional level	Field	Impt. [6, 14]	Infl. [7, 15]	Affect. [8, 16]	Count
Educational Institution	Resource planning	Alert information	Hydro-innovation	National	Research and education	0,0	0,0	10,0	1
Electricity Utility	Economic service operations and planning	Water related information	Downstream	National	Energy	8,4	9,2	8,4	5
Environmental Research Institution	Environment	Other	Hydro-innovation	National	Research and education	5,5	5,0	8,0	3
EU	Resource planning	Other	Downstream	Supranational	Other	10,0	10,0	10,0	1
FEWSNET	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	8,0	8,0	8,0	1
Firestone	Economic service operations and planning	Water related information	Downstream	National	Agriculture	4,0	2,0	2,0	1
GWP/AO	Environment	Water related information	Hydro-innovation	Supranational	Water resources	4,0	4,0	2,0	1
Industry and Commerce Entities	Economic service operations and planning	Alert information	Downstream	Local	Industry and Commerce	6,0	6,0	3,0	3
IUCN	Environment	Water related information	Hydro-innovation	Supranational	Other	4,0	6,0	2,0	1
Local Administrative Entity	Civil society	Alert information	Downstream	Local	Administration	8,7	8,6	8,2	14
Local Association for Agriculture	Economic service operations and planning	Alert information	Downstream	Local	Agriculture	5,7	3,0	4,9	9
Local Entity for Civil Security Enforcement	Rescue and aid	Alert information	Downstream	Local	Civil protection	10,0	10,0	10,0	1
Local Entity for Development	Resource planning	Other	Downstream	Local	Development	6,0	4,0	3,5	2
Local Entity for Water Resources Planning	Resource planning	Forecast refinement	Hydro-innovation	Local	Water resources	4,3	3,8	7,3	4
Media	Civil society	Alert information	Downstream	NA	Other	10,0	6,0	2,0	2
Metal Steel	Economic service operations and planning	Meteorological data	Downstream	National	Industry and Commerce	4,0	2,0	2,0	1
MNG	Economic service operations and planning	Meteorological data	Downstream	National	Industry and Commerce	4,0	2,0	2,0	1
National Administrative Entity	Civil society	Alert information	Downstream	National	Administration	7,2	8,0	5,8	5

Stakeholder [1, 9]	Main interest [3, 11]	Why use system [4, 12]	Information profile	Decisional level	Field	Impt. [6, 14]	Infl. [7, 15]	Affect. [8, 16]	Count
National Agency for Disaster Management Planning	Disaster management	Alert information	Downstream	National	Civil protection	8,7	8,4	7,3	12
National Agency for Meteorology	Technical	Meteorological data	Hydro-innovation	National	Meteorology	10,0	9,2	6,7	7
National Agency for Water Resources	Resource planning	Forecast refinement	Hydro-innovation	National	Water resources	7,0	8,3	6,5	8
National Association for Agriculture	Economic service operations and planning	Alert information	Downstream	National	Agriculture	9,0	9,0	9,0	4
National Entity for Aerial Transportation	Economic service operations and planning	Meteorological data	Downstream	National	Transportation	10,0	10,0	0,0	1
National Entity for Agriculture Management	Economic service operations and planning	Water related information	Downstream	National	Agriculture	5,9	5,3	9,1	7
National Entity for Civil Security Enforcement	Rescue and aid	Alert information	Downstream	National	Civil protection	8,0	8,1	6,9	7
National Entity for Development	Resource planning	Alert information	Downstream	National	Development	6,0	9,0	2,0	2
National Entity for Energy Planning	Resource planning	Alert information	Downstream	National	Energy	10,0	10,0	10,0	1
National Entity for Transportation	Economic service operations and planning	Alert information	Downstream	National	Transportation	10,0	10,0	10,0	1
National Entity for Water Infrastructure	Resource planning	Water related information	Hydro-innovation	National	Water services	7,0	10,0	1,0	2
National Entity for Waterways Transport	Economic service operations and planning	Water related information	Downstream	National	Transportation	6,3	5,5	6,5	4
National Environment Protection Entity	Environment	Water related information	Hydro-innovation	National	Other	6,8	6,0	6,3	4
National Governmental Entity for Agriculture Planning	Economic service operations and planning	Alert information	Downstream	National	Agriculture	7,3	6,0	7,3	7
National Governmental Entity for Disaster	Disaster management	Alert information	Downstream	National	Civil protection	8,9	9,1	7,0	9

Stakeholder [1, 9]	Main interest [3, 11]	Why use system [4, 12]	Information profile	Decisional level	Field	Impt. [6, 14]	Infl. [7, 15]	Affect. [8, 16]	Count
Management Planning									
National Governmental Entity for Meteorology	Technical	Meteorological data	Hydro-innovation	National	Meteorology	10,0	9,3	9,0	3
National Governmental Entity for Water Resources	Resource planning	Forecast refinement	Hydro-innovation	National	Water resources	9,6	9,4	8,5	14
National Governmental Entity for Water Services	Economic service operations and planning	Water related information	Downstream	National	Water services	7,3	9,3	6,3	3
National Health Service Entity	Rescue and aid	Alert information	Downstream	National	Other	0,0	0,0	10,0	1
National Humanitarian Aid Entity	Rescue and aid	Alert information	Downstream	National	Humanitarian aid	6,0	4,0	4,0	1
NGO	Rescue and aid	Alert information	Downstream	NA	Humanitarian aid	8,3	8,4	8,3	7
OCHA	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	8,0	9,0	8,0	2
OMVG	Resource planning	Forecast refinement	Hydro-innovation	Supranational	Water resources	8,0	8,0	9,3	4
OMVS	Resource planning	Forecast refinement	Hydro-innovation	Supranational	Water resources	7,4	7,4	8,8	5
OXFAM	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	6,0	6,0	5,0	2
PAM	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	8,0	8,0	8,0	1
Red Cross	Rescue and aid	Alert information	Downstream	Supranational	Humanitarian aid	7,5	6,5	5,4	8
Regional Dam Management Entity	Economic service operations and planning	Water related information	Downstream	Supranational	Energy	10,0	10,0	10,0	2
Regional Entity for Development	Resource planning	Alert information	Downstream	Supranational	Development	4,0	10,0	2,0	1
Research Institution	Economic service operations and planning	Other	Hydro-innovation	National	Research and education	5,5	4,5	7,0	2
Statistics Institution	Technical	Other	Downstream	National	Research and education	7,0	6,0	7,0	3
TOR	Economic service operations and planning	Meteorological data	Downstream	National	Energy	4,0	0,0	0,0	1

Stakeholder [1, 9]	Main interest [3, 11]	Why use system [4, 12]	Information profile	Decisional level	Field	Impt. [6, 14]	Infl. [7, 15]	Affect. [8, 16]	Count
Tullow	Economic service operations and planning	Meteorological data	Downstream	Supranational	Energy	4,0	0,0	0,0	1
UN	Resource planning	Other	Downstream	Supranational	Other	10,0	10,0	10,0	1
WASCAL	Environment	Alert information	Hydro-innovation	Supranational	Research and education	4,0	4,0	2,0	1
Water Resources Development Programs	Resource planning	Water related information	Downstream	National	Water resources	4,0	8,5	3,0	3
Water Utility	Economic service operations and planning	Water related information	Downstream	National	Water services	8,5	8,3	8,8	4

2.2 Objectives and attributes

Table S5. Fundamental objectives and corresponding attributes, to quantify the system configurations performance, including short names, attribute unit, attribute range in parentheses (worst and best possible case), and short description of attribute.

No.	Higher level / lower level objective full name	Short objective name	Attribute full name	Short attribute name	Unit (range)	Attribute description
1	**High information accuracy and clarity**	**1_accuracy**				
1.1	High accuracy of information	11_accur_info	Value of KGE index for 1, 3, 10 day forecasts	11_kge	Value (0 – 1)	Level of accuracy consisting of KGE for 1, 3, and 10 day forecasts; index transformed into [0:1] value by aggregation with a weighted sum, using different weights for the sub-attributes.
1.2	Clear flood risk information	12_clear_info	Clarity of flood risk thresholds in risk management	12_info	Value (0 – 1)	Levels of flood risk threshold of three categorical sub-attributes: 1) location, 2) possibility to calculate risk, and 3) ease of applying to flood risk management; aggregated into [0:1] value, using an equally weighted sum.
2	**Good information access**	**2_access**				
2.1	Reliable access to information	21_reliable_info	Stability of distribution channel	21_channel	Score (0 – 2.2)	Access to information calculated by the weighted sum of categorical stability of access (i.e., stable, sporadic, none) for each distribution channel (i.e., SMS, Email, fanfar.eu).

No.	Higher level / lower level objective full name	Short objective name	Attribute full name	Short attribute name	Unit (range)	Attribute description
2.2	Timely production, distribution, and access to info	22_timely_info	Time to produce, distribute, and access information	22_time	Value (0 – 1)	Time required to produce forecast, access distribution channels, and access IVP; transformed to [0:1] value with linear interpolation, aggregated with equally weighted sum.
2.3	Several languages	23_language	Number of languages available for system components	23_langue	Score (0 – 17.64)	Number of languages (En, Fr, Pt, Ara) for different system components and channels with different weights for languages and system components; weighted sum resulting in a score.
3	**Low costs**	**3_costs**				
3.1	Short development time	31_develop_time	Time required to develop and implement system components	31_devlptime	Days (0 – 1095)	Total time needed for the development and implementation of all technical system components in days.
3.2	Low operation costs	32_costs	Annual operation and maintenance costs	32_costs	Value (0 – 1)	Annual costs (€ / year) to operate and maintain the system for entire West Africa; estimated costs based on operation and maintenance of individual system components included in each configuration; transformed into [0:1] value.
4	**High sustainability**	**4_sustainable**				
4.1	Long-term financing secured	41_sust_financing	Secured financing beyond 2020	41_finance	Value (0 – 1)	Level of secured financing beyond 2020 consisting of two sub-attributes (costs covered, and duration of financing) transformed to [0:1] value, aggregated with equally weighed sum.
4.2	Skillful human resources available	42_human_resour	Number of people in W Africa for development, O&M, and access & interpretation	42_experts	Value (0 – 1)	Skillful human resources available, based on three sub-attributes (number of people in West Africa available to 1) develop & maintain, 2) operate, and 3) access & interpret the system) transformed to a [0:1] value, aggregated with equally weighted sum.
4.3	Good support system	43_support_syst	Information quantity in support system and time of response to resolve issues	43_suppsys	Value (0 – 1)	Support system quality consisting of two sub-attributes: 1) information quantity, and 2) time for response of support system; transformed into a [0:1] value, aggregated with equally weighted sum.

2.3 FANFAR system configurations

The FANFAR system configurations, were generated using the Strategy Generation Table method, and Brainwriting 635 combined with Cadavre Exquis. The system configurations are combinations of different system characteristics. These were generated separately for the Hydrology-TEP (Table S6), and the Interactive Visualization Portal (IVP; Table S7). The latter is the interface of FANFAR, with which most users will normally interact. In the FANFAR co-design workshops, all emergency managers interacted with the IVP, and most of the hydrologists that need the FANFAR system to create forecasts and alerts. The Hydrology-TEP, i.e., the actual forecast production system, was only accessed by very experienced hydrologists, which in the co-design workshops were members of the FANFAR consortium from AGRHYMET.

Table S6. Characteristics of the system configurations for the Hydrology-TEP (development stage summer 2019). The columns display the key components of the FANFAR system with relevance for Hydrology-TEP including hydrological observation data types, meteorological analysis and forecast data, hydrological model (WWH: World-Wide HYPE: Arheimer et al., 2020) (Niger HYPE: Andersson et al., 2017), forecasted output variables, distribution options of derived information (forecasts and alerts), degree of automatization of information distribution, language, and support system. In situ: local hydrometric gauge observation data; EO: Earth Observations; HydroGFD: Hydrological Global Forcing Data (merged data set of historical precipitation and temperature from meteorological reanalysis and global observations; Berg et al., 2020;Berg et al., 2018); HydroGFD-West Africa from AGRHYMET: HydroGFD2 adjusted by West African meteorological observations; GFS: Global Forecast System (https://www.ncdc.noaa.gov/); ECMWF: European Centre for Medium-Range Weather Forecasts (https://www.ecmwf.int/en/forecasts); En: English; Fr: French; Pt: Portuguese; Ar: Arabic.

Code	System Configuration	Hydrological observation types	Meteorological input/forcing data	Hydrological models	Forecast output variables	Distribution configurations of derived information (forecasts and alerts)	Automatization (of information distribution)	Language	Support system
a_Fast-dev	Least resources for development	None	Current data from SMHI (HydroGFD v2)	Niger HYPE	Streamflow	Web visualization	Automatic push of data to distribution channels	En	• Forum; • Knowledge base; • Help desk
b_Res-user	Least resources for users	In situ country level data: water level, discharge	Improved data from SMHI (HydroGFD v3)	WWH	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP • SMS alert • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt; • Ar	• Forum • Knowledge base
c_Easy-use	Most easy-to use	EO: water level	Current data from SMHI (HydroGFD v2)	Niger HYPE	Water level	• Web visualization; • SMS alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic push of data to distribution channels	• En; • Fr; • Pt; • Ar	• Forum • Knowledge base

Code	System Configuration	Hydrological observation types	Meteorological input/forcing data	Hydrological models	Forecast output variables	Distribution configurations of derived information (forecasts and alerts)	Automatization (of information distribution)	Language	Support system
d_Fast	Fastest system	EO: water level	Current data from SMHI (HydroGFD v2)	Niger HYPE	Water level	• Web visualization; • SMS automatic push of data; • SMS alert • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic push of data to distribution channels	En	• Forum • Knowledge base; • Help desk; • Demos, tutorials; • SMS, Email; • Phone
e_Consent	Highest consensus	System can handle all types of data: in situ, EO	• American meteo (e.g. GFS); • ECMWF ensemble forecasts; • HydroGFD-West Africa from AGRHYMET	• Niger HYPE; • WWH	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt; • Ar	• Forum • Knowledge base; • Help desk; • Demos, tutorials; • SMS, Email; • Phone
f_Robust	Most robust	EO: water level	Current data from SMHI (HydroGFD v2)	• Niger HYPE; • WWH (or WWH alone)	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt; • Ar	• Forum • Knowledge base; • Help desk; • Demos, tutorials; • SMS, Email; • Phone
g_Attractve	Most attractive	System can handle all types of data: in situ, EO	• American meteo (e.g. GFS); • ECMWF ensemble forecasts; • HydroGFD-West Africa from AGRHYMET	• Niger-HYPE; • WWH	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt; • Ar	• Forum • Knowledge base; • Help desk; • Demos, tutorials; • SMS, Email; • Phone
h_Equipp	Fully equipped	System can handle all types of	• American meteo (e.g. GFS);	• Niger HYPE; • WWH	• Streamflow; • Water level;	• Web visualization; • Login to H-TEP;	Automatic processing, with possibility to choose	• En; • Fr;	• Forum • Knowledge base;

Code	System Configuration	Hydrological observation types	Meteorological input/forcing data	Hydrological models	Forecast output variables	Distribution configurations of derived information (forecasts and alerts)	Automatization (of information distribution)	Language	Support system
		data: in situ, EO	• ECMWF ensemble forecasts; • HydroGFD-West Africa from AGRHYMET		• Precipitation; • Evaporation; • Soil moisture	• FANFAR and national FTP; • API; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	automatic and manual control of distribution by operator	• Pt; • Ar	• Help desk; • Demos, tutorials; • SMS, Email; • Phone
i_Calibr	Recalibrated HYPE models	None	Current data from SMHI (HydroGFD v2)	WWH calibrated for West Africa	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt	• Forum; • Knowledge base; • Help desk
j_Cal-EO	Recalibrated HYPE models and EO data	EO: water level	Current data from SMHI (HydroGFD v2)	WWH calibrated for West Africa	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt	• Forum; • Knowledge base; • Help desk
k_Cal-EO-situ	Recalibrated HYPE models and EO data	System can handle all types of data: in situ, EO	Current data from SMHI (HydroGFD v2)	WWH calibrated for West Africa	• Streamflow; • Water level; • Precipitation; • Evaporation; • Soil moisture	• Web visualization; • Login to H-TEP; • SMS alert; • Email alert; • WhatsApp alert; • Radio, TV; • Traditional word of mouth	Automatic processing with manual control of distribution by operator	• En; • Fr; • Pt	• Forum; • Knowledge base; • Help desk

Table S7. Characteristics of the system configurations for the Interactive Visualization Portal (IVP; development stage summer 2019). Columns indicate observed variable, displayed model performance (accuracy of forecast), forecasted variable, flood hazard reference threshold type, data download, interactivity, language, and alert notification system. NSE: Nash-Sutcliffe Efficiency; KGE: Kling-Gupta Efficiency; En: English; Fr: French; Pt: Portuguese; Ar: Arabic.

Code	System Configuration	Observed variable	Model performance, accuracy	Forecasted variable	Flood hazard reference threshold type	Data download	Interactivity	Language	Alert notification to operating agencies and stakeholders
a_Fast-dev	Least resources for development	None	No performance metrics shown	River discharge	Return periods (simulation)	None	Interactive website (zoom, click, etc.)	En	None
b_Res-user	Least resources for users	In situ: water level	No performance metrics shown	River discharge	• Return periods (simulation); • Return periods (observations; only stations); • Selected historic years (e.g., 2012); • User defined thresholds for specific location	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt; • Ar	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders
c_Easy-use	Most easy to use	• In situ: water level; • EO: water level	Performance metric shown (e.g., NSE, KGE; with colored levels)	• River discharge; • Water level; • Precipitation; • Evaporation	• Return periods (simulation); • Selected historic years (e.g., 2012)	• Excel table for selected station; • Charts, Graphs (e.g., histogram, pie, bars)	Website with static images (no zoom, no click)	• En; • Fr; • Pt	• Automatic alerts to agencies; • Automatic alerts sent directly to stakeholders
d_Fast	Fastest system	None	No performance metrics shown	River discharge	Return periods (simulation)	None	Website with static images (no zoom, no click)	En	• Automatic alerts to agencies; • Automatic alerts sent directly to stakeholders
e_Consent	Highest consensus	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation; • Evaporation	• Return periods (simulation); • Selected historic years (e.g., 2012)	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders

Code	System Configuration	Observed variable	Model performance, accuracy	Forecasted variable	Flood hazard reference threshold type	Data download	Interactivity	Language	Alert notification to operating agencies and stakeholders
f_Robust	Most robust	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation	• Return periods (simulation); • Return periods (observations; only stations); • Selected historic years (e.g., 2012); • User defined thresholds for specific location	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt; • Ar	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders
g_Attractve	• Most attractive	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation; • Evaporation • Soil moisture storage; • Water quality	• Return periods (simulation); • Return periods (observations; only stations); • Selected historic years (e.g., 2012); • User defined thresholds for specific location	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt; • Ar	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders
h_Equipp	Fully equipped	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation; • Evaporation • Soil moisture storage; • Water quality	• Return periods (simulation); • Return periods (observations; only stations); • Selected historic years (e.g., 2012); • User defined thresholds for specific location	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	• Interactive website (zoom, click, etc.); • and with static images	• En; • Fr; • Pt; • Ar	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders
i_Calibr	Recalibrated HYPE models	• In situ: water level, river discharge;	• Performance metric shown (e.g., NSE, KGE; with colored levels)	• River discharge; • Water level; • Precipitation; • Evaporation	Return periods (simulation)	• Excel table for selected station; • Map of displayed variables (PNG);	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels;

Code	System Configuration	Observed variable	Model performance, accuracy	Forecasted variable	Flood hazard reference threshold type	Data download	Interactivity	Language	Alert notification to operating agencies and stakeholders
		• EO: water level, precipitation	• Blank areas with too low forecast performance (based on NSE, KGE)			• Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)			• Automatic alerts to agencies that can authorize system to forward alert to stakeholders
j_Cal-EO	Recalibrated HYPE models and EO data	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation; • Evaporation	Return periods (simulation)	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders
k_Cal-EO-situ	Recalibrated HYPE models and EO data	• In situ: water level, river discharge; • EO: water level, precipitation	• Performance metric shown (e.g., NSE, KGE; with colored levels) • Blank areas with too low forecast performance (based on NSE, KGE)	• River discharge; • Water level; • Precipitation; • Evaporation	Return periods (simulation)	• Excel table for selected station; • Map of displayed variables (PNG); • Map of variables (Shapefile); • Charts, Graphs (e.g., histogram, pie, bars)	Interactive website (zoom, click, etc.)	• En; • Fr; • Pt	• Automatic alerts to agencies, that can forward alerts to stakeholders via existing distribution channels; • Automatic alerts to agencies that can authorize system to forward alert to stakeholders

2.4 Predicting performance of each system configuration

2.4.1 Attribute details

1.1 High accuracy of information

Description: *Hydrological forecasts and flood risk alerts should consistently reach high accuracy (matching hydrological observations and local knowledge of flood occurrence and magnitude).*

Attribute: *Value of KGE index for 1, 3, 10 day forecasts (11_kge)*

Unit: *value [0:1]*

Description: *Level of accuracy consisting of KGE for 1, 3, and 10 day forecasts index transformed into [0:1] value by aggregation with a weighted sum using different weights for the sub-attributes.*

Best case: *Operational forecasts are always 100% accurate and have no errors (KGE = 1) across the 1, 3, and 10 lead days.*

Worst case: *Forecasts are 0% accurate and misleading (KGE = -∞ | -1'000) across the 1, 3, and 10 lead days.*

Addit. Info: *The accuracy refers to the streamflow (fluvial floods) information sent out (disseminated forecasts and alerts). The manual operator may decide to send it or not.*

Attribute details:

11_kge is an artificial (constructed) attribute, consisting of the three sub-attributes (KGE index for 1 day, 3 day, and 10 day forecasts).

There are several indices used for accuracy in hydrology (see e.g., Deliverable 3.2, section 5 in https://fanfar.eu/wp-content/uploads/sites/4/2020/05/FANFAR-D3.2-Hydrological-Models.pdf). The FANFAR attribute accuracy however is solely measured through the KGE index (Kling-Gupta Efficiency; Gupta et al., 2009), which is commonly used in hydrology for model evaluation; i.e., to estimate the error of a set predicted values vs observed values. To predict the performance of the FANFAR forecasting system towards the observed streamflow, we used expert estimates. These were elicited from Jafet Andersson (SMHI, hydrologist from FANFAR) in July 2019. He estimated the KGE index over three typically used lead days: the 1 day, 3 day, and 10 day forecasts. These lead days indicate the time between the forecast production and the actual flood event. We used 7 levels for the assessment, which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1. The expert assigned a KGE index number to each level for each of the three lead days (

Table S8). Thereafter, each configuration (FANFAR system configuration) received a KGE index number for each of the three lead days (

Table S9). This KGE index number was transformed to a value using a nonlinear value function between all the above levels from 0 (worst) to 1 (best; with linear interpolation between the levels from worst to best;

Table S8; Figure S5). We then aggregated all three lead day values into a single value (between 0 and 1) with a weighted sum, where the lead times (1, 3, and 10 days) received different weights. The accuracy of the 1 day forecast was weighted the highest [w = 0.5], the 3 day forecast received a slightly lower weight [w = 0.4], and the 10 day forecast was given small importance [w = 0.1]. As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA modelling.

Uncertainty of predictions:

We dealt with the uncertainty of each single lead-day estimate using triangle distributions. These contain a best (most probable) expert estimate for the 1, 3, and 10 day estimate, and a lower (minimum) / upper bound (maximum) of uncertainty in the expert's statement. The uncertainty of the sub-attributes is not symmetrical around the prediction. The uncertainty distributions of the sub-attributes are skewed towards lower values, as for many reasons, the lower values in the uncertainty range are more likely, than higher values. To aggregate over the three sub-attributes, we used 1'000 Monte Carlo simulation runs, each of them drawing from the triangle uncertainty distribution of each sub-attribute. For the subsequent MCDA, we used a normal distribution to model the aggregated uncertainty of the three sub-attributes. To estimate the shape of the aggregated uncertainty distribution, we chose a normal distribution with mean = mean of Monte Carlo simulation, and standard deviation = ¼ of the 95% confidence interval resulting from the Monte Carlo simulation with 1'000 runs. This resulted in an individual uncertainty distribution for each configuration for this attribute. These distributions are symmetrical but have a lower mean than the aggregated prediction to account for the skewedness of uncertainty (Figure S6).

Table S8. Expert estimates for levels of sub-attributes: KGE-index for 1, 3, and 10 day forecasts (lead days).

Error of forecast level	Value	KGE (lead days)		
		KGE-1	KGE-3	KGE-10
Best	1.000	1.00	1.00	1.00
Very good	0.833	0.90	0.80	0.50
Good	0.667	0.70	0.60	0.25
Neutral	0.500	0.40	0.30	0.00
Bad	0.333	0.00	-0.50	-1.00
Very bad	0.167	-1.00	-2.00	-50.00
Worst	0.000	∞ | -1'000.00	∞ | -1'000.00	∞ | -1'000.00

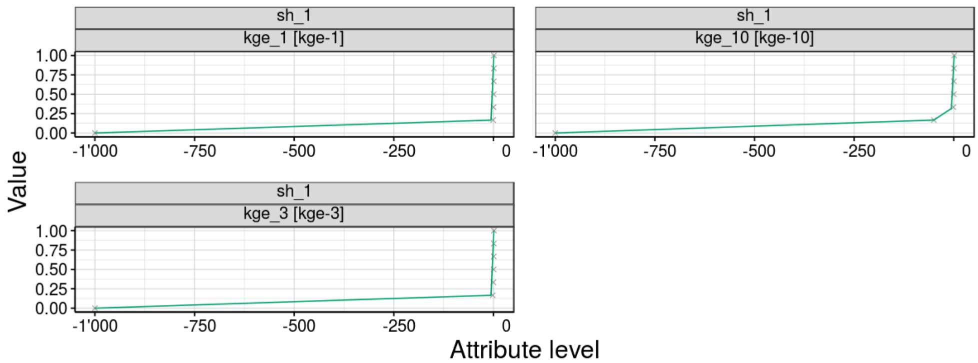


Figure S5: Value functions to aggregate the three sub-attributes kge-1, kge-3, and kge-10 into one [0:1] value. The functions were approximated from the level estimates by the expert using linear interpolation with 140 approximated points.

Table S9: Predictions of the three sub-attributes kge-1, kge-3, and kge-10 for each system configuration with min/ max values of uncertainty in the estimate. These predictions are expert estimates done by Jafet Andersson and were aggregated into one value (last column), which is the value used as prediction for the aggregated attribute "11_kge" in the subsequent MCDA modelling.

System configurations	Predicted KGE (1 day)	min (1 day)	max (1 day)	Predicted KGE (3 days)	min (3 days)	max (3 days)	Predicted KGE (10 days)	min (10 days)	max (10 days)	Aggregated value of prediction
a_Fast-dev	0.30	0.00	0.60	0.15	-0.45	0.35	-2.00	-20.00	0.20	0.450
b_Res-user	0.79	0.00	0.88	0.49	0.00	0.61	-1.94	-19.58	0.22	0.646
c_Easy-use	0.42	0.00	0.67	0.24	-0.31	0.42	-1.99	-19.90	0.20	0.484
d_Fast	0.46	0.00	0.69	0.26	-0.26	0.43	-1.98	-19.87	0.20	0.497
e_Consent	0.86	0.00	0.92	0.58	0.00	0.68	0.40	-3.20	0.84	0.738
f_Robust	0.46	0.00	0.69	0.26	-0.26	0.43	0.40	-3.20	0.84	0.540
g_Attractve	0.86	0.00	0.92	0.58	0.00	0.68	0.40	-3.20	0.84	0.738
h_Equipp	0.86	0.00	0.92	0.58	0.00	0.68	0.40	-3.20	0.84	0.738
i_Calibr	0.65	0.00	0.80	0.49	0.00	0.61	0.40	-3.20	0.84	0.638
j_Cal-EO	0.72	0.00	0.84	0.58	0.00	0.68	0.40	-3.20	0.84	0.679
k_Cal-EO-situ	0.93	0.00	0.96	0.75	0.00	0.81	0.40	-3.20	0.84	0.833

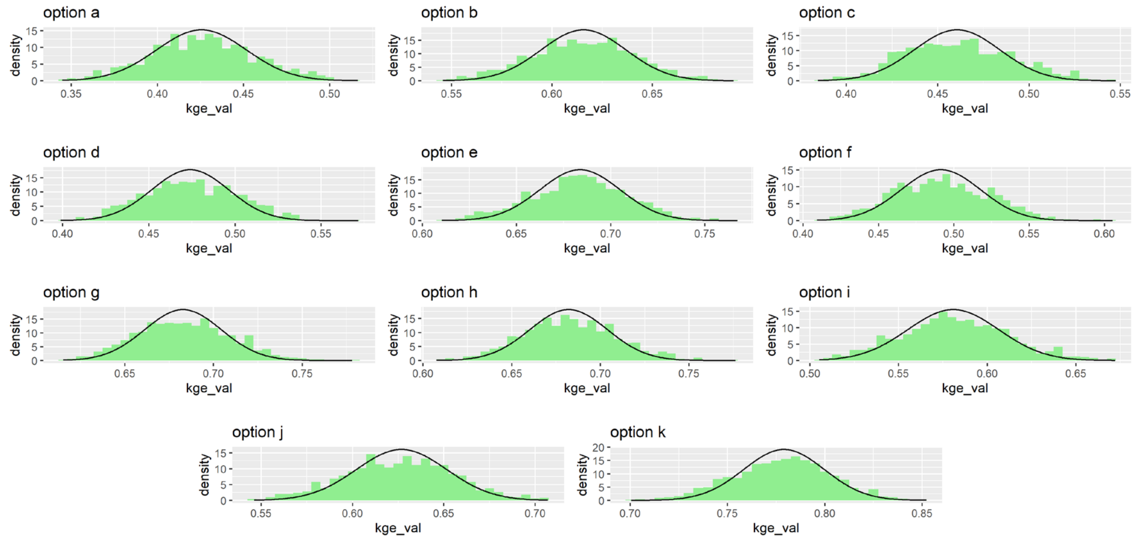


Figure S6: Distribution of resulting uncertainties (green) after aggregating the three sub-attributes and 1'000 Monte Carlo simulation runs. Results for the 11 FANFAR system configurations a – k. A normal distribution with mean = mean of Monte Carlo simulation, and sd = ¼ of the 95% confidence interval (black curve) fits the resulting uncertainty well. A normal distribution is thus used as input of the uncertainty of the aggregated prediction of the attribute 11_kge in the subsequent MCDA. The means of these distributions are lower than the aggregated predictions in

Table S9 to account for skewedness of the uncertainty in the expert estimates.

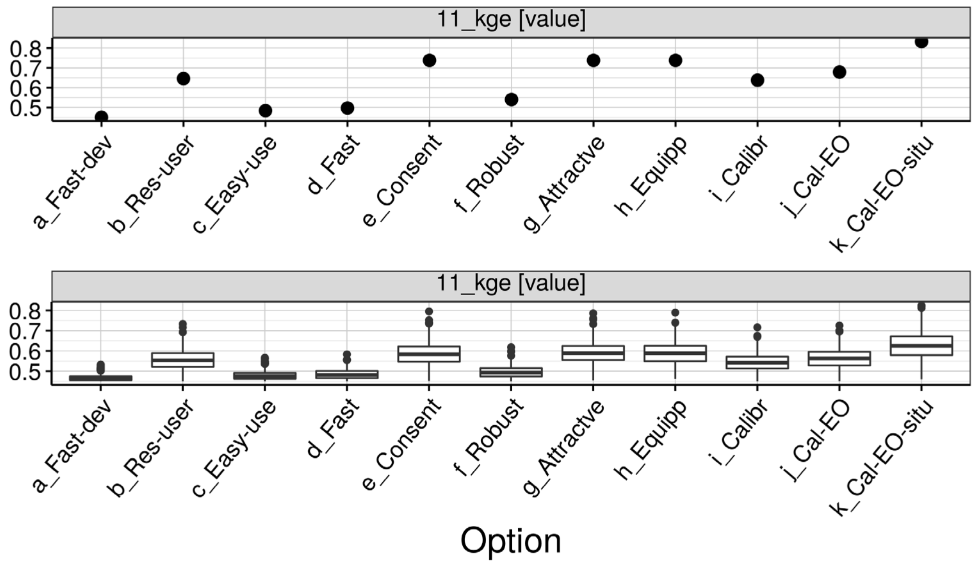


Figure S7: Visualization of the aggregated value of predictions of all system configurations for the attribute 11_kge. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1000 runs.

1.2 Clear flood risk information

Description: *Flood risk thresholds should be clearly defined to be used to calculate flood risks, and are easy to understand and apply in flood management*

Attribute: *Clarity of flood risk thresholds in risk management (12_info)*

Unit: *value [0:1]*

Description: *Levels of flood risk threshold of three categorical sub-attributes: 1) location, 2) possibility to calculate risk, 3) ease of understanding in applied risk management aggregated into [0:1] value, using an equally weighted sum.*

Best case: *Flood risk thresholds are clearly defined everywhere. No ambiguities of calculating flood risk or understanding and application.*

Worst case: *No flood risk definitions. No adequate notifications can be derived, distributed, understood or applied.*

Attribute details:

12_info is an artificial (constructed) attribute, consisting of the three categorical sub-attributes location, possibility of flood risk calculation, and ease of application to flood management.

Our definition of the flood risk thresholds depends on: 1) place where they are defined (e.g., everywhere or gauged locations); 2) possibility to calculate the flood risk; and 3) how easy the flood risk thresholds are to apply in flood management. Estimates on these categorical sub-attributes were elicited from Bernard Minoungou (AGRHYMET, hydrologist from FANFAR) in July 2019. We used 7 levels for the assessment, which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1. To each level, the expert assigned a combination of the three categorical sub-attributes (

Table S10). Note, when giving the predictions for each of our FANFAR system configurations, the expert later decided that our configurations are all defined "everywhere", and that it is possible to calculate the flood risk for all configurations. For these two sub-attributes a value of 1 is achieved by all configurations in our case, and thus they are not needed to evaluate the currently chosen FANFAR system configurations. Therefore, we only used yes (1) / no (0) categorical input data for the sub-attribute on ease of application to flood management. Consequently, the respective values of the configurations are either 1 (best-possible level), or 0.667 (good level;

Table S10 bold text). As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA (Figure S8).

Uncertainty of predictions:

For this attribute there was no uncertainty in the experts' estimate.

Table S10: Expert estimates for levels of the flood risk threshold for the three categorical sub-attributes 1) location, 2) possibility of flood risk calculation, and 3) ease of application to flood management. In our case we only considered sub-attribute 3, as all configurations are identical for sub-attribute 1, and 2.

Flood risk threshold definition level	Value	Place where flood risk is defined	Possible to calculate flood risk?	Possible to apply to flood management?
Best	1.000	Everywhere	Yes	Yes
Very good	0.833	Gauged loc.	Yes	Yes
Good	0.667	Everywhere	Yes	No
Neutral	0.500	Gauged loc.	Yes	No
Bad	0.333	Everywhere	No	Yes
Bad	0.333	Gauged loc.	No	Yes
Very bad	0.167	Everywhere	No	No
Very bad	0.167	Gauged loc.	No	No
Worst	0.000	Not defined	No	No

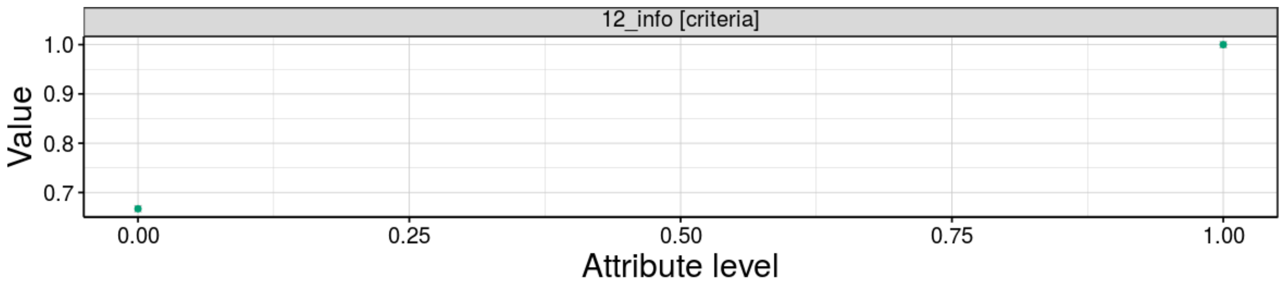


Figure S8: Discrete value function for the transformation of the categorical sub-attribute on ease of application to flood management into a [1:0] value. A sub-attribute category "No" (0) achieved the level "Good" (value = 0.667), and a category "Yes" (1) resulted in the

best case (value = 1). If there would have been differences between configurations concerning the levels of the other two sub-attributes, a more complex aggregation approach would be required.

Table S11: Predictions on the three categorical sub-attributes of 12_info for each configuration. All configurations are defined everywhere and enable to calculate flood risk, thus only the ease of understanding in applied risk management is important for the FANFAR system configurations. Consequently, the aggregated value for the subsequent MCDA is either 0.667 or 1.

System configurations	Prediction (place)	Prediction (calculation risk)	Prediction (understandable for flood management)	Aggregated value of prediction
a_Fast-dev	Everywhere	Yes	No	0.667
b_Res-user	Everywhere	Yes	Yes	1.000
c_Easy-use	Everywhere	Yes	Yes	1.000
d_Fast	Everywhere	Yes	No	0.667
e_Consent	Everywhere	Yes	Yes	1.000
f_Robust	Everywhere	Yes	Yes	1.000
g_Attractve	Everywhere	Yes	Yes	1.000
h_Equipp	Everywhere	Yes	Yes	1.000
i_Calibr	Everywhere	Yes	No	0.667
j_Cal-EO	Everywhere	Yes	No	0.667
k_Cal-EO-situ	Everywhere	Yes	No	0.667

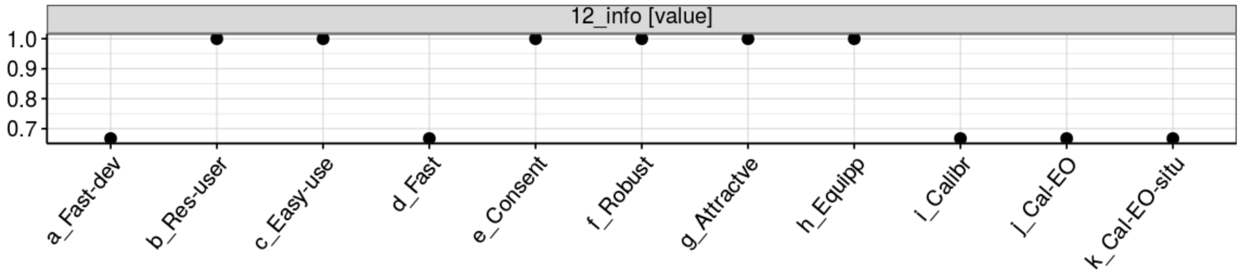


Figure S9: Visualization of the aggregated value of predictions of all system configurations for the attribute 12_info. There was no uncertainty defined for this attribute.

2.1. Reliable access to information

Description: *Access to forecasts and flood risk notifications should be reliable: always reach intended recipients / key stakeholders.*

Attribute: *Stability of distribution channel (21_channel)*

Unit: *score [0:2.2]*

Description: *Access to information calculated by the weighted sum of categorical stability of access (i.e., stable, sporadic, none) for each distribution channel (i.e., SMS, Email, fanfar.eu).*

Best case: *Forecasts and flood risk notifications always reach the intended stakeholders [$\Sigma = 2.2$].*

Worst case: *Forecasts and flood risk notifications do not reach the intended stakeholders [$\Sigma = 0$].*

Attribute details:

21_channel is an artificial (constructed) attribute, consisting of the weighted sum of the reliability of the most important distribution channels: SMS, Email, and FANFAR homepage.

For each configuration (FANFAR system configuration) and each distribution channel, i.e., SMS, Email, FANFAR homepage (fanfar.eu), the reliability of access was assigned categorically as a value, i.e., stable [v=1], sporadic [v=0.7], none [v=0]. This estimate was done by Francisco Silva Pinto (Eawag) in June 2019 (Table S13). To obtain a single overall score for each configuration, we calculated the weighted sum over all three distribution channels, with following weights: SMS [w=1], Email [w=0.7], homepage [w=0.5]. Thereafter, we used 7 assessment levels, which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best; Table S12Table S13). After the transformation into values from 0 to 1, these levels serve as a nonlinear marginal value function based on the expert estimate for the subsequent MCDA (Figure S10).

Uncertainty of predictions:

According to Francisco Silva Pinto, who did the predictions, the estimates have an uncertainty of a normal distribution with 10% of the prediction as standard deviation.

Table S12: Expert estimates for levels of the weighted sum (score) of the reliability of access to the three distribution channels, SMS, Email, and homepage (fanfar.eu). Note: These levels are directly used for the marginal value function in the subsequent MCDA modelling.

Reliability of distribution channel level	Value	Weighted sum (score)
Best	1.000	2.20
Very good	0.833	2.05
Good	0.667	1.84
Neutral	0.500	1.54
Bad	0.333	1.19
Very bad	0.167	0.70
Worst	0.000	0.00

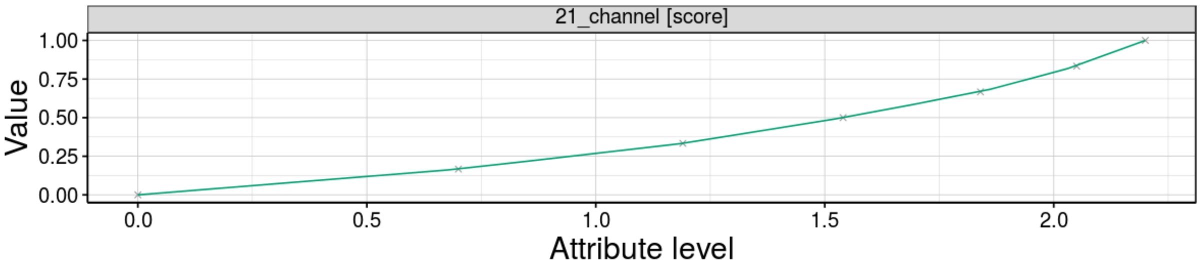


Figure S10: Nonlinear marginal value function for the subsequent MCDA modelling of attribute "reliable access to information". The shape of the value function is based on the experts' estimates on levels (
Table S12).

Note: In contrast to the previous attributes, the levels here were not used for aggregation of sub-attributes, but as a nonlinear single attribute value function for the subsequent MCDA. In other words, here the predictions were directly adopted for the input of the subsequent MCDA.

Table S13: Predictions on the reliability of the three distribution channels SMS, Email, and FANFAR homepage. The predictions were transformed into one score for the subsequent MCDA using a weighted sum with different weights for the distribution channels.

System configurations	Predicted reliability of SMS	Predicted reliability of Email	Predicted reliability of fanfar_eu	Total weighted sum (score)
a_Fast-dev	none	none	sporadic	0.35
b_Res-user	stable	stable	sporadic	2.05
c_Easy-use	sporadic	sporadic	stable	1.69
d_Fast	stable	stable	stable	2.20
e_Consent	stable	stable	sporadic	2.05
f_Robust	stable	stable	sporadic	2.05
g_Attractve	stable	stable	sporadic	2.05
h_Equipp	stable	stable	sporadic	2.05
i_Calibr	stable	stable	sporadic	2.05
j_Cal-EO	stable	stable	sporadic	2.05
k_Cal-EO-situ	stable	stable	sporadic	2.05

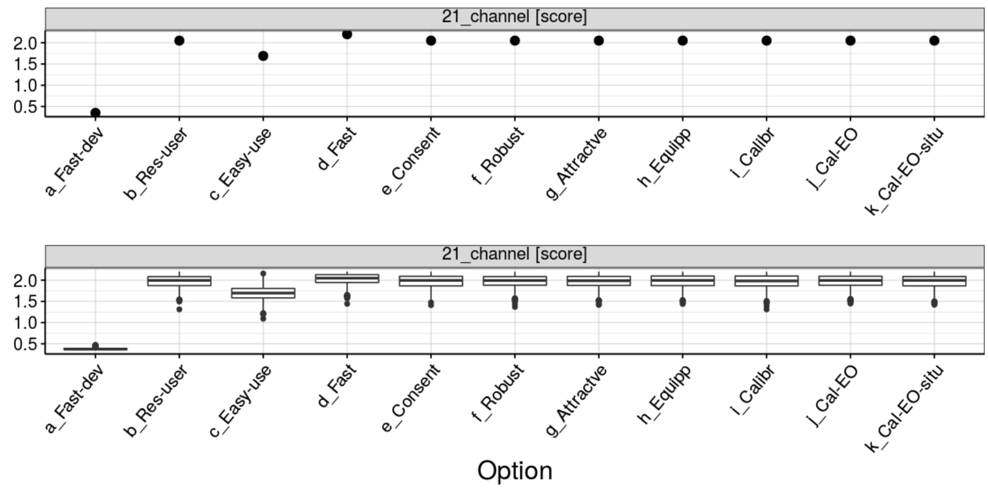


Figure S11: Visualization of the total weighted sum of all system configurations used as predictions for the attribute 21_channel. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

2.2. Timely production, distribution, and access to information

Description: *The time to produce, distribute and access forecasts and flood risk notifications. A faster system allows for a more timely response.*

Attribute: *Time to produce, distribute, and access information (22_time)*

Unit: *value [0:1]*

Description: *Level of time required to produce forecast, access distribution channels, and access IVP; transformed to [0:1] value with linear interpolation, aggregated with equally weighted sum.*

Best case: *Flood forecasts and hazard notifications are produced in 0.5 h, and to access the information in West Africa it takes less than 30 seconds on all platforms.*

Worst case: *Flood forecasts and hazard notifications are produced in more than 24 h. In West Africa, it takes more than 10 min to access the information through the IVP, SMS, and Email, and more than 15 min to access through the production system.*

Additional information:

Production time: *whenever the first component starts until the last component finishes (Hydrology-TEP) in an optimized production chain. From when there is data on the SMHI FTP until the forecast is finished and sent out.*

Access time through distribution channels: *time that it takes for a user from when it was sent until it is possible to be seen in West Africa (as in the FANFAR workshops).*

Access time through production system: *loading time of the Hydrology-TEP and navigation time (imagine someone navigating in West Africa throughout the H-TEP).*

Attribute details:

22_time is an artificial (constructed) attribute consisting of three sub-attributes, the time it takes to: 1) produce a forecast; 2) access the information through the distribution channels; and 3) access the information in the forecast production system.

The predictions for these sub-attributes were estimated by Emmanuel Mathot (Terradue) in July 2019. For the production time, he estimated a basic production time (i.e., the time a system with the least resources (configuration a) requires). To that basic production time, additional time was added depending on whether the system configuration utilizes satellite data (+ 0.5 h), local observations (+ 0.75 h), or manual control (+ 1 h). Criteria for the estimated time needed for access through the distribution channels, were whether the configuration uses SMS, or Email (which both reduce the access time). The access time through the production system depends on the complexity of the components used in each configuration e.g., diversity in observational data inputs, number of models used, type of information derived, and language. The more complex the configuration the more time is needed to load menus and output (Table S14).

Note: During the operational period of the FANFAR pre-operational system in 2020, the production times were generally much higher than what was estimated by the expert in July 2019. However, the assumption, that a more complex system configuration requires more time, still holds. Therefore, the relative differences of the system configurations are still valid.

For each sub-attribute, levels were created, which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best; Table S14). After the transformation into values from 0 to 1 we used these levels as value functions for the aggregation of the three sub-attributes into one single [0:1] value using a weighted sum with different weights for the sub-attributes. As the production time and access via the distribution channels is more important for the users to access every day, these two sub-attributes received a slightly higher weights (each w=0.4), and access via the production system received a lower weight of 0.2. (Figure S12). This resulting [0:1] value of the aggregation forms the prediction for the subsequent MCDA. As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA modelling.

Uncertainty of predictions:

We dealt with the uncertainty of each sub-attribute estimate using a triangle distribution. This contains a best (most probable) expert estimate for the three time estimates, and a lower (minimum) / upper bound (maximum) for each, which are both very unlikely. The expert initially stated uncertainty ranges exceeding the value function range. These uncertainty ranges were trimmed off to lie within the value function, as any predictions outside the value function range will achieve a value of 0 (if undershooting) or 1 (if overshooting). To aggregate over the three sub-attributes, we used 1'000 Monte Carlo simulation runs, each of them drawing from the triangle uncertainty distribution of each sub-attribute. For the subsequent MCDA, we used a normal distribution to model the aggregated uncertainty of the three sub-attributes. To estimate the shape of the aggregated uncertainty distribution, we chose to utilize a normal distribution with mean = mean of Monte Carlo simulation, and standard deviation = ¼ of the 95% confidence interval resulting from the Monte Carlo simulation with 1'000 runs. This resulted in an individual uncertainty distribution for each configuration for this attribute (Figure S13).

Table S14: Expert estimates for levels of the three sub-attributes time to 1) produce, 2) access via distribution channels, and 3) access the flood risk information in the forecast production system.

Timely information level	Value	Production time (h)	Access via distribution channel (min)	Access via production system (min)
Best	1.000	0.5	0.5	0.5
Very good	0.833	1.0	1.0	2.0
Good	0.667	2.0	2.0	5.0
Neutral	0.500	3.0	3.5	7.5
Bad	0.333	6.0	5.0	10.0
Very bad	0.167	12.0	7.5	12.5
Worst	0.000	24.0	10.0	15.0

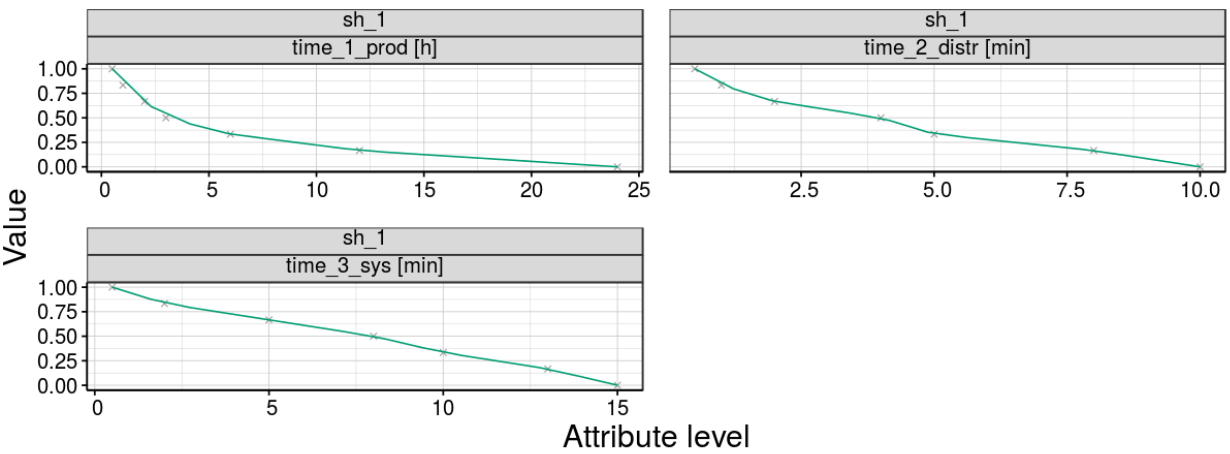


Figure S12: Value functions used for the aggregation of the three sub-attributes into one [0:1] value. The functions were approximated from the level estimates by the expert using linear interpolation with 14 approximated points.

Table S15: Predictions of the three sub-attributes production time, access time through distribution channel, and access time through production system for each system configuration with min/ max values of uncertainty in the estimate. These predictions are expert estimates done by Emmanuel Mathot and were aggregated into one value (last column), which is the value used as prediction for the aggregated attribute 22_time in the subsequent MCDA modelling.

System configu-rations	Predicted production time (h)	Min	Max	Predicted access time distr ch (min)	Min	Max	Predicted access time prod syst (min)	Min	Max	Aggregated value of prediction
a_Fast-dev	4.0	2.0	6.0	2.0	1.0	4.0	10.0	5.0	15.0	0.511
b_Res-user	4.0	2.0	6.0	0.5	0.3	1.0	12.0	6.0	15.0	0.618
c_Easy-use	5.8	3.8	7.8	1.0	0.5	2.0	11.0	5.5	15.0	0.526
d_Fast	4.5	2.5	6.5	0.5	0.3	1.0	10.0	5.0	15.0	0.633
e_Consent	4.5	2.5	6.5	0.5	0.3	1.0	15.0	7.5	15.0	0.567
f_Robust	6.3	4.3	8.3	0.5	0.3	1.0	13.0	6.5	15.0	0.557
g_Attractve	5.5	3.5	7.5	0.5	0.3	1.0	15.0	7.5	15.0	0.544
h_Equipp	6.3	4.3	8.3	0.5	0.3	1.0	15.0	7.5	15.0	0.531
i_Calibr	6.3	4.3	8.3	0.5	0.3	1.0	12.0	6.0	15.0	0.571
j_Cal-EO	5.0	3.0	7.0	0.5	0.3	1.0	12.0	6.0	15.0	0.596
k_Cal-EO-situ	5.5	3.5	7.5	0.5	0.3	1.0	13.0	6.5	15.0	0.571

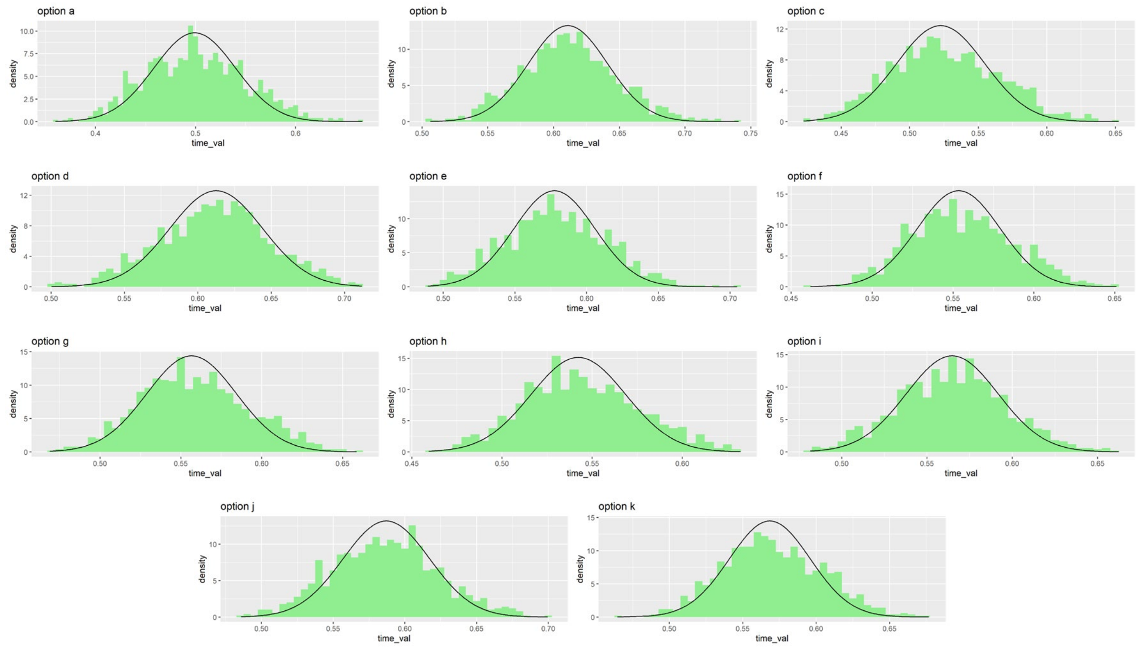


Figure S13: Distribution of resulting uncertainties (green) after aggregating the three sub-attributes and 1’000 Monte Carlo simulation runs. Results for the 11 FANFAR system configurations a–k. A normal distribution with mean = mean of Monte Carlo simulation, and sd = ¼ of the 95% confidence interval (black curve) fits the resulting uncertainty well. A normal distribution is thus used as input of the uncertainty of the aggregated prediction of the attribute 22_time in the subsequent MCDA.

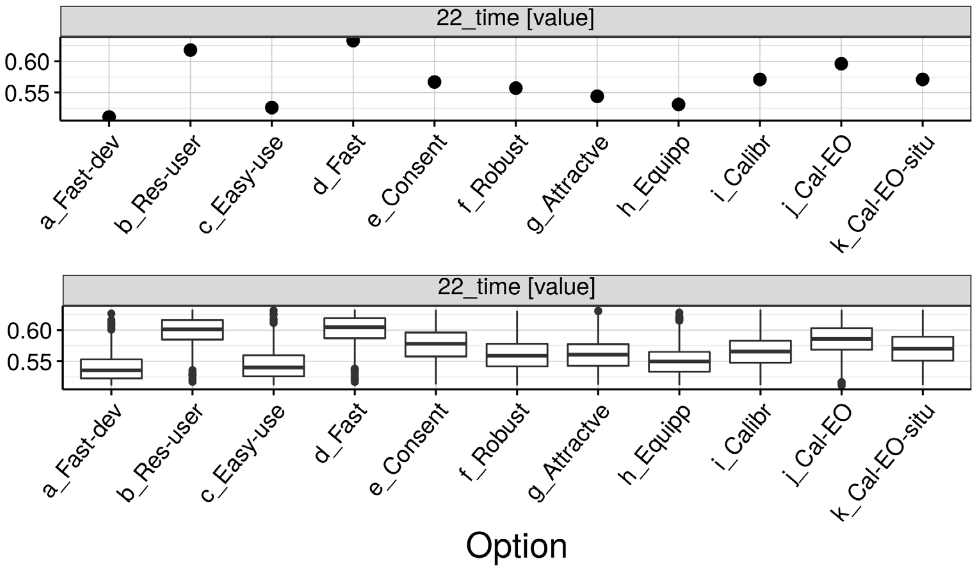


Figure S14: Visualization of the aggregated value of predictions of all system configurations for the attribute 22_time. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

2.3. Several languages

Description: *System components should be available in multiple languages. System = distribution channels (fanfar.eu, IVP, Email, SMS), support (knowledge base, forum, help desk), and Hydrology-TEP (general interface, community page, FANFAR app, processing services).*

Attribute: *Number of languages available for system components (23_langue)*

Unit: *score [0:17.64]*

Description: *Weighted sum of number of languages (En, Fr, Pt, Ar) for different system components and channels with different weights for different languages and system components; weighted sum resulting in a score.*

Best case: *The system and its documentations and manuals are available in English, French, Portuguese, and Arabic.*

Worst case: *The system and its documentations and manuals are only available in English.*

Attribute details:

23_langue is an artificial (constructed) attribute calculated with a weighted sum of languages available in different system components.

Each system component can be available in several languages. Francisco Silva Pinto (Eawag) calculated the weighted sum (with different weights (w) for different languages and system components) of the number of languages available (i.e., En [w=1], Fr [w=0.7], Pt [w=0.3], Ar [w=0.1]) in different system components (i.e., fanfar.eu [w=1], IVP [w=1], Email [w=1], SMS [w=1], Knowledge base [w=1], Forum [w=0.7], Help Desk [w=0.7], H-TEP General Interface [w=0.5], H-TEP Community [w=0.5], H-TEP FANFAR App [w=0.5], H-TEP Proc Services [w=0.5]. This weighted sum resulted in a single score of languages available for each configuration. Different levels for this attribute were created (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best) which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1 (Table S16) in order to achieve a marginal value function for this attribute, which is used for the subsequent MCDA (Figure S15).

Uncertainty of predictions:

For this attribute there was no uncertainty in the experts' estimate.

Table S16: Expert estimates for the levels of weighted sum (score) of languages for the individual system components (fanfar.eu, IVP, Email, SMS, Knowledge base, Forum, Help Desk, H-TEP General Interface, H-TEP Community, H-TEP FANFAR App, H-TEP Proc Services). Note: These levels are directly used as marginal value function in the subsequent MCDA.

Language Level	Value	Score
Best	1.000	17.64
Very Good	0.833	17.10
Good	0.667	15.48
Neutral	0.500	14.58
Bad	0.333	10.99
Very bad	0.167	9.80
Worst	0.000	0.00

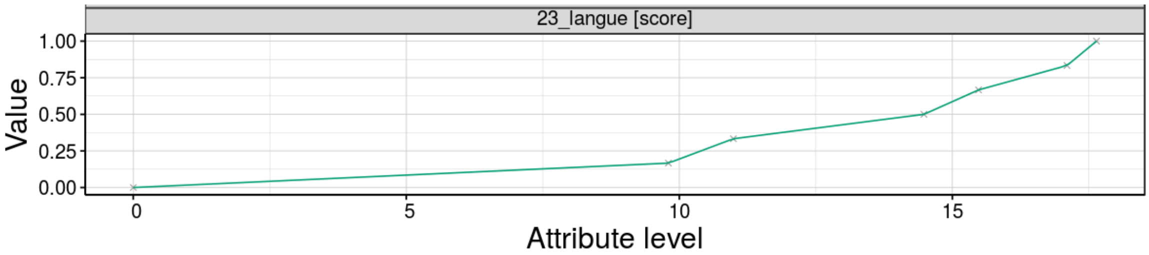


Figure S15: Nonlinear marginal value function for the subsequent MCDA modelling of the attribute “several languages”. The shape of the value function is based on the experts’ estimates on levels.

Table S17: Predictions on the availability of the system components in different languages. The predictions were transformed into one score for the subsequent analysis using a weighted sum with different weights for the different languages and different system components. The resulting total weighted sum (score) was directly adopted for the input of the subsequent MCDA.

System configurations	fanfar_eu	IVP	Email	SMS	knowl. base	forum	help desk	H-TEP intern	H-TEP com	H-TEP app	H-TEP servi	Total weighted sum (score)
a_Fast-dev	efp	e	ef	e	e	e	e	e	e	e	e	10.10
b_Res-user	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	17.64
c_Easy-use	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efp	efp	efp	efp	17.44
d_Fast	efp	e	ef	e	e	e	e	e	e	e	e	10.10
e_Consent	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efp	efp	efp	efp	17.44
f_Robust	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	17.64
g_Attractve	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	17.64
h_Equipp	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	efpa	17.64
i_Calibr	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	16.80
j_Cal-EO	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	16.80
k_Cal-EO-situ	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	efp	16.80

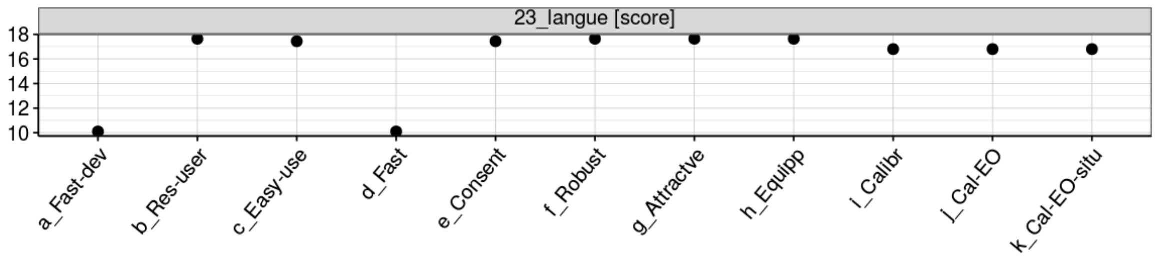


Figure S16: Total weighted sum of all system configurations used as predictions for the attribute 23_langue. The plot represents the predictions without uncertainty. There was no uncertainty defined for this attribute.

3.1. Short development time

Description:	*The time required to develop the system component/ configuration. Each system configuration will demand different amount of time to develop from the FANFAR consortium.*
Attribute:	*Time required to develop and implement system components (31_devlptime)*
Unit:	*days [1:1095]*
Description:	*Total time needed for the development and implementation of all technical system components in days.*
Best case:	*1 day is needed to develop the system component/ configuration.*
Worst case:	*3 years (1'095 days) are needed to develop the system component/ configuration.*

Attribute details:

31_devlptime is a natural attribute directly measured in days needed for the development and implementation of all technical components.

The time required to develop and implement the individual technical system components in days was estimated by Emmanuel Mathot (Terradue) in July 2019. The time required for the development and implementation of the technical components was estimated and different components were assigned to the involved development teams at Terradue, IsardSAT, and SMHI. For each system configuration, the time needed for the technical components included in the configuration was summed up for each development team. The final development time for each configuration resulted from the maximum time requirement of the three development teams; i.e., the development takes as much time as that team needs which requires the longest time (predictions see Table S19). Different levels for this attribute were created (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best), which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1 (Table S18) to achieve a marginal value function for this attribute for the following MCDA modelling (Figure S17).

Uncertainty of predictions:

The expert stated an uncertainty range around the predictions (Table S19). This uncertainty is included in the subsequent MCDA using a triangular distribution with mode = prediction and min/max = min/max estimates.

Table S18: Expert estimate for the levels of the value function for the attribute 3.1 Short development time. Note: These levels are directly used for the marginal value function in the subsequent MCDA.

Development time level	Value	Time (days)
Best	1.000	1
Very Good	0.833	7
Good	0.667	30
Neutral	0.500	90
Bad	0.333	180
Very bad	0.167	365
Worst	0.000	1095

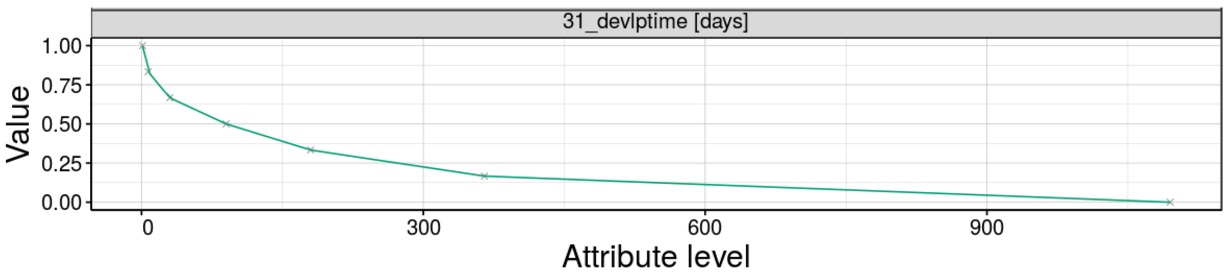


Figure S17: Nonlinear marginal value function for the subsequent MCDA of the attribute "short development time". The shape of the value function is based on the experts' estimates on levels (
Table S18).

Table S19: Predictions on the development time of the system components in days for each FANFAR system configuration with min/ max values of the uncertainty range in the estimate (± 20 % of prediction). The predictions (days) were directly adopted as predictions for the input of the subsequent MCDA.

System configurations	Predicted dev. Time (days)	Min (days)	Max (days)
a_Fast-dev	1	1	1
b_Res-user	672	537	806
c_Easy-use	600	480	720
d_Fast	278	222	333
e_Consent	1008	806	1095
f_Robust	645	516	774
g_Attractve	1008	806	1095
h_Equipp	1008	806	1095
i_Calibr	402	321	482
j_Cal-EO	432	345	518
k_Cal-EO-situ	522	417	626

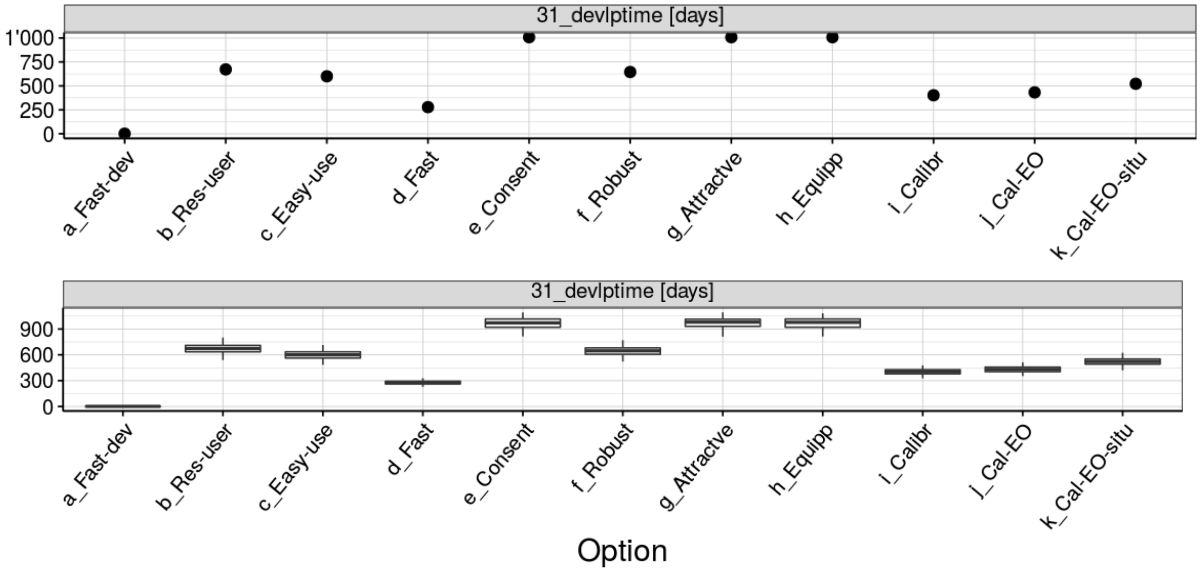


Figure S18 Visualization of the predicted development time of all system configurations. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

2.2. Low operation costs

Description: *The cost to operate and maintain the ICT infrastructure should be as low as possible (e.g., devices, networks, cloud computing resources).*

Attribute: *Annual operation and maintenance costs (32_costs)*

Unit: *Value [0:1]*

Description: *Level of annual costs to operate and maintain the system for entire West Africa.*

Best case: *The operation and maintenance costs are very low (e.g., 10'000 € per year, for entire West Africa).*

Worst case: *The operation and maintenance costs are very high (e.g., 1 million € per year, for entire West Africa).*

Attribute details:

31_costs is a natural attribute measured in € / year required to operate and maintain the system. For confidentiality reasons, the cost estimates of the FANFAR system configurations are not public. Therefore, the costs were transformed into a value from 0:1 representing the relative differences in estimated costs between the system configurations.

The estimate for operation and maintenance costs are calculated based on expert estimates from Aytor Naranjo (IsardSAT), and Fabrizio Pacini (Terradue) in July 2019 for the individual system components included in each configuration. A basic minimum operation cost was estimated and the expected costs to maintain additional system components, i.e., the inclusion of observation data, and utilization of satellite data, were added depending on which components the configuration includes. Additionally, the estimated cost of the work time of the personnel needed to operate and maintain the system was added. This resulted in one final value for operation and maintenance costs for each configuration (Table S21). Different levels for this attribute were created (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best), which initially had a continuous (linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1 to highlight the relative performance of the system configurations for this attribute. Accordingly, the single attribute value function of this attribute in the final MCDA is linear with a range from 0:1 (Figure S19).

Uncertainty of predictions:

The expert stated an uncertainty range around the predictions for each component of the summed operation costs. The costs of operating a system utilizing observation data, the costs to operate satellite data, and the cost estimate on the work time needed to operate the system were each stated with a range of uncertainty. The minimum and maximum total predicted operation costs resulted from adding up the min and max values of these ranges depending on which components the system configuration includes. (see Table S21). This uncertainty is included in the subsequent MCDA using a triangular distribution with mode = prediction and min/max = transformed max/min estimates.

Table S20: Single attribute value function of the transformed cost levels used in the subsequent MCDA.

Cost level	Value	Level of operation costs (value)
Best	1.000	1.000
Very Good	0.833	0.833
Good	0.667	0.667
Neutral	0.500	0.500
Bad	0.333	0.333
Very bad	0.167	0.167
Worst	0.000	0.000

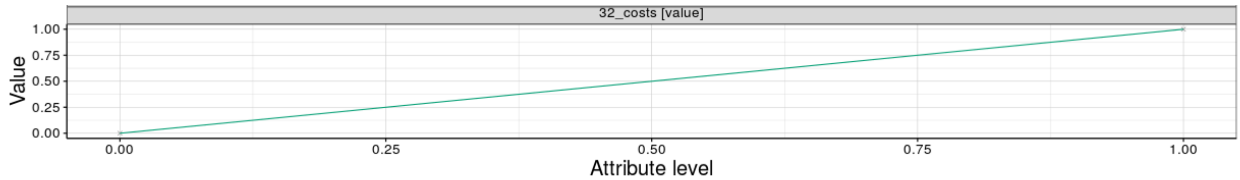


Figure S19: Marginal value function for the subsequent MCDA of the attribute "low operation costs". The shape of the value function is based on the experts' estimates on levels.

Table S21: Predictions on the values of operation costs of the system for each system configuration with min/ max values of the uncertainty range in the transformed estimate.

System configurations	Observation data	Satellite Data	Predicted operation costs (value)	Min (value)	Max (value)
a_Fast-dev	No	No	0.958	0.933	0.983
b_Res-user	Yes	No	0.468	0.441	0.495
c_Easy-use	No	Yes	0.492	0.472	0.587
d_Fast	No	Yes	0.492	0.472	0.587
e_Consent	Yes	Yes	0.451	0.416	0.486
f_Robust	No	Yes	0.492	0.472	0.587
g_Attractve	Yes	Yes	0.451	0.416	0.486
h_Equipp	Yes	Yes	0.451	0.416	0.486
i_Calibr	No	No	0.571	0.497	0.662
j_Cal-EO	No	Yes	0.492	0.472	0.587
k_Cal-EO-situ	Yes	Yes	0.451	0.416	0.486

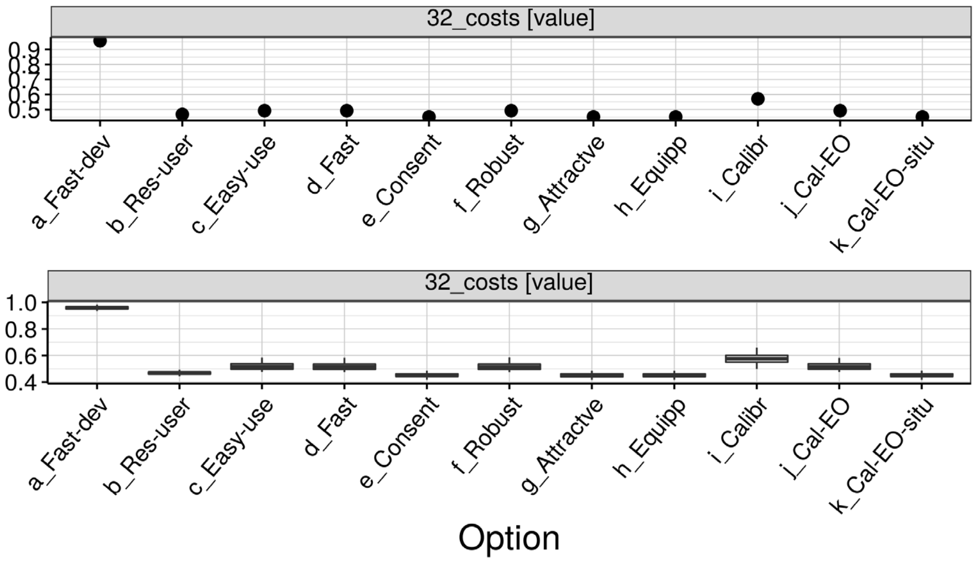


Figure S20: Visualization of the predicted operation costs of all system configurations transformed into a 0:1 value. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

4.1. Long-term financing secured

Description: *The time period for which the financing of the system is secured. This can be only for the duration of the FANFAR project, or for the entire lifespan of the system.*

Attribute: *Secured financing beyond 2020 (41_finance)*

Unit: *value [0:1]*

Description: *Level of secured financing beyond 2020 consisting of two sub-attributes (costs covered, and duration of financing) transformed to [0:1] value integrated with equally weighed sum.*

Best case: *Financing is secured for the entire lifespan of the FANFAR system (100% for 30 years).*

Worst case: *No financing beyond 2020: financing ends when the FANFAR project ends.*

Attribute details:

41_finance is an artificial (constructed) attribute, consisting of two sub-attributes, namely the percentage of costs covered, and the time period for which the financing of the system is secured. This can be only for the duration of the FANFAR project (0 years), or for the entire lifespan of the system (30 years).

The predictions for these two sub-attributes were assessed from an expert, Abdou Ali (AGRHYMET) in July 2019 (Table S23). Seven levels were created for each sub-attribute (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best), which initially had a continuous (linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1 (Table S22). The two sub-attributes were then aggregated into one single value by using the levels as value functions and calculating the weighted sum with equal weight for both sub-attributes. As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA.

Uncertainty of predictions:

The expert stated an uncertainty range of 10 % around the predictions for the first sub-attribute (costs covered). For the duration of financing, the uncertainty ranged from one year less than the prediction to maximally the prediction. This uncertainty is included in the aggregation of the two sub-attributes using a triangular distribution with mode = prediction and min/max = prediction ± 10 %. To aggregate over the two sub-attributes, we used 1'000 Monte Carlo simulation runs, each of them drawing from the triangle uncertainty distribution of each sub-attribute. For the subsequent MCDA, we used a normal distribution to model the resulting uncertainty of the two aggregated sub-attributes. To estimate the shape of the resulting uncertainty distribution after aggregation, we chose to utilize a normal distribution with mean = mean of Monte Carlo simulation, and standard deviation = ¼ of the 95% confidence interval resulting from the Monte Carlo simulation with 1000 runs. This resulted in an individual uncertainty distribution for each configuration for this attribute (Figure S22).

Table S22: Expert estimates for levels of the sub-attributes 1) percentage of costs covered, and 2) duration of financing secured beyond 2020 (years). These levels were used for the aggregation of the two sub-attributes into one [0:1] value.

Long-term financing level	Value	Costs covered	Duration of financing (years)
Best	1.000	100%	30
Very Good	0.833	100%	15
Good	0.667	100%	5
Neutral	0.500	78%	4
Bad	0.333	75%	3
Very bad	0.167	70%	1.5
Worst	0.000	0%	0

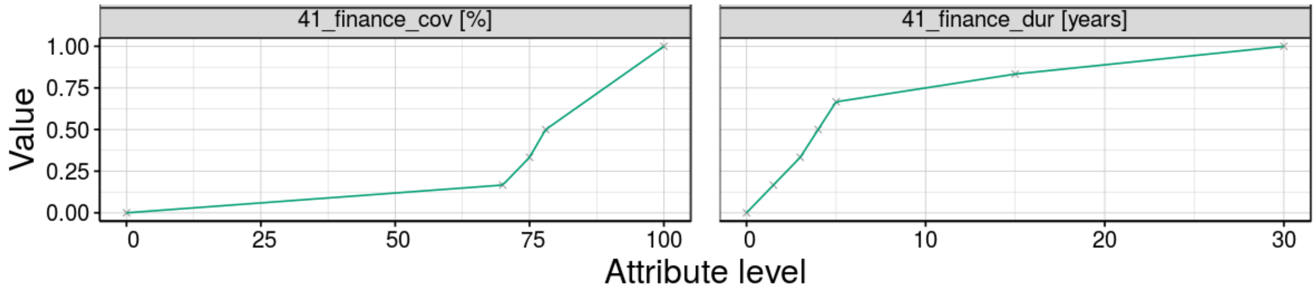
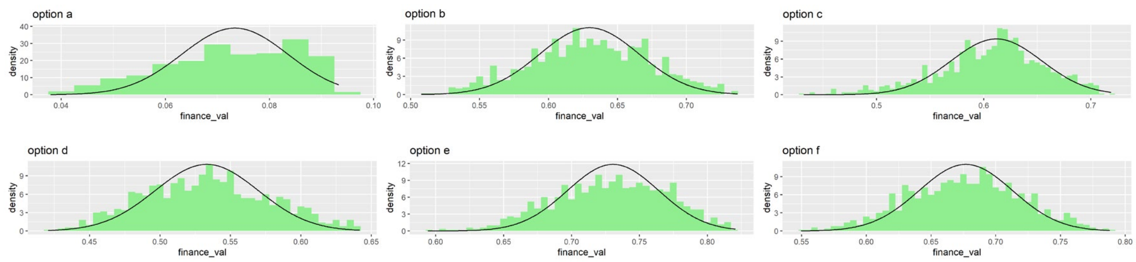


Figure S21: Value functions to aggregate the two sub-attributes into one [0:1] value. The functions were approximated from the level estimates by the expert (
Table S22) using linear interpolation with 14 approximated points.

Table S23: Predictions of the two sub-attributes costs covered, and duration of financing for each system configuration with min/ max values of uncertainty in the estimate. These predictions are expert estimates done by Abdou Ali and were aggregated into one value (last column), which is the value used as prediction for the aggregated attribute 41_finance in the subsequent MCDA modelling.

System configurations	Predicted costs covered	Min	Max	Predicted duration of financing (years)	Min	Max	Aggregated value of prediction
a_Fast-dev	30%	27%	33%	1	0	1	0.091
b_Res-user	93%	83%	100%	4	3	4	0.670
c_Easy-use	83%	74%	91%	5	4	5	0.640
d_Fast	90%	81%	99%	3	2	3	0.553
e_Consent	95%	85%	100%	5	4	5	0.777
f_Robust	89%	80%	97%	5	4	5	0.708
g_Attractve	98%	88%	100%	5	4	5	0.811
h_Equipp	84%	75%	92%	6	5	6	0.660
i_Calibr	95%	85%	100%	3	2	3	0.610
j_Cal-EO	96%	86%	100%	4	3	4	0.705
k_Cal-EO-situ	86%	77%	94%	5	4	5	0.674



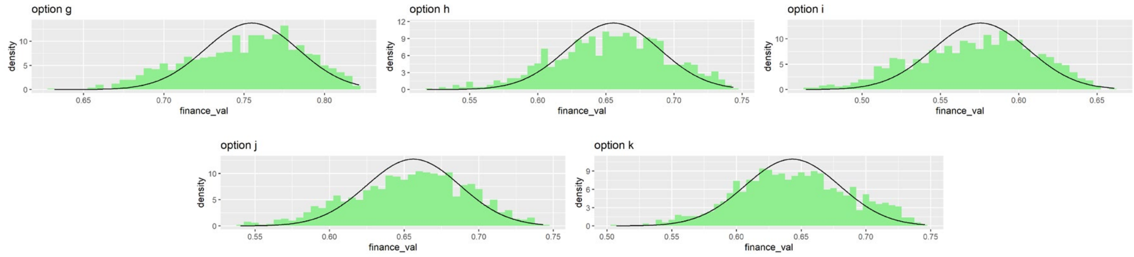


Figure S22: Distribution of resulting uncertainties (green) after aggregating the three sub-attributes and 1'000 Monte Carlo simulation runs. Results for the 11 FANFAR system configurations a – k. A normal distribution with mean = mean of Monte Carlo simulation, and sd = ¼ of the 95% confidence interval (black curve) fits the resulting uncertainty well. A normal distribution is thus used as input of the uncertainty of the aggregated prediction of the attribute 41_finance in the subsequent MCDA.

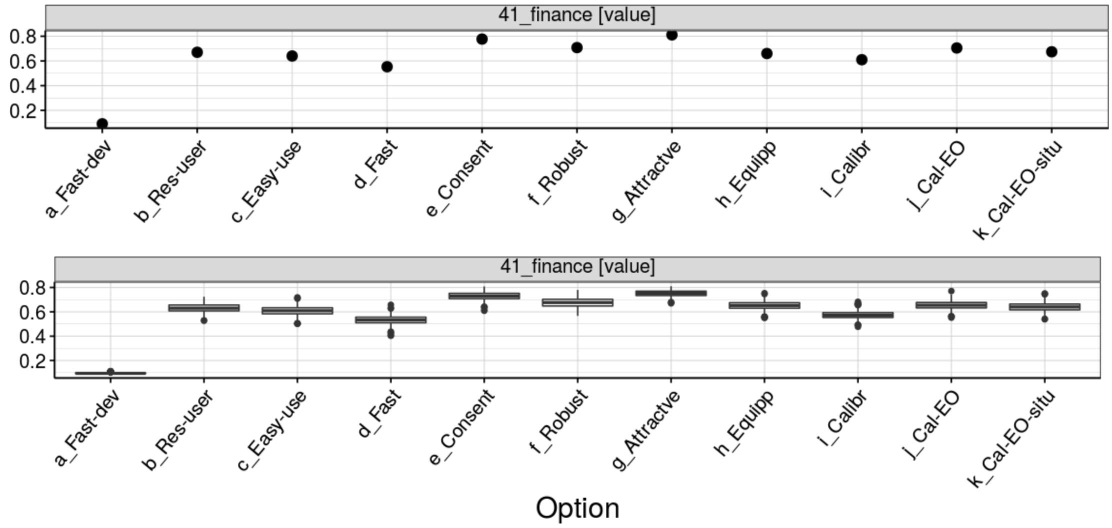


Figure S23: Visualization of the aggregated value of predictions of all system configurations for the attribute 41_finance. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

4.2. Skillful human resources available

Description: *At the end of the FANFAR project (and included training), the availability of skilled staff at West African institutions that are able to access, interpret, operate, maintain, and develop the forecasting and alert system.*

Attribute: *Number of people in W Africa for development, O & M, and access & interpretation (42_experts)*

Unit: *value [0:1]*

Description: *Level of skillful human resources available of three sub-attributes (number of people in West Africa available to 1) develop & maintain, 2) operate, and 3) access & interpret the system) transformed to a [0:1] value aggregated with equally weighted sum.*

Best case: *5 persons are able to develop and maintain the system, 10 to operate it. 100 persons are able to access and interpret the forecasts and flood risk notifications (at least 3 in each country).*

Worst case: *Nobody is able to develop, maintain, and operate the forecasting system, and nobody is able to access and interpret its outputs.*

Attribute details:

42_experts is an artificial (constructed) attribute consisting of three sub-attributes, namely the number of persons: 1) able to develop and maintain the system, 2) able to operate it, and 3) able to access and interpret the forecasts and flood risk notifications.

To predict the outcomes of the individual FANFAR system configurations, we used expert estimates. These were elicited from Jafet Andersson (SMHI, hydrologist from FANFAR) in July 2019. He estimated the number of people able to 1) develop and maintain, 2) able to operate, and 3) able to access and interpret the forecasts and flood risk notifications for each system configuration (predictions see Table S25). Seven levels were created for each sub-attribute (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best), which initially had a (continuous linear) level ranging from 0 (worst), over very bad (16.667), (...), to very good (83.333), and 100 (best). We later transformed the levels to a value from 0 to 1 (Table S24). These three sub-attributes were transformed to one single [0:1] value using nonlinear value functions based on the estimated levels, and aggregated with a weighted sum, where each sub-attribute received the same weight of 0.33. As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA modelling.

Uncertainty of predictions:

For each sub-attribute, the expert gave a different uncertainty range. For the estimate of people able to develop and maintain the system, he stated an uncertainty of 10 % around his prediction. The prediction of people able to operate it lies within an uncertainty range of 20 % around the predicted number. And with 40 % around the prediction, even more uncertainty was expected in the estimate of how many people are able to access and interpret the system. The expert initially stated uncertainty ranges exceeding the value function range. These uncertainty ranges were trimmed off to lie within the value function, as any predictions outside the value function range will achieve a value of 0 (if undershooting) or 1 (if overshooting). The uncertainty ranges were included in the aggregation of the three sub-attributes using a triangular distribution with mode = prediction, and min/max = min/max of stated uncertainty range. To aggregate over the three sub-attributes, we used 1'000 Monte Carlo simulation runs, each of them drawing from the triangle uncertainty distribution of each sub-attribute. For the subsequent MCDA, we used a normal distribution to model the aggregated uncertainty of the three sub-attributes. To estimate the shape of the aggregated uncertainty distribution, we chose to utilize a normal distribution with mean = mean of Monte Carlo simulation, and standard deviation = ¼ of the 95% confidence interval resulting from the Monte Carlo simulation with 1'000 runs. This resulted in an individual uncertainty distribution for each configuration for this attribute (Figure S25).

Table S24: Expert estimates for levels of the three sub-attributes: people able to 1) develop and maintain, 2) operate, and 3) access and interpret the FANFAR system. These levels were used to shape the value functions for aggregating the three sub-attributes.

Human resource level	Value	People able to develop and maintain	People able to operate	People able to access and interpret
Best	1.000	5	10	100
Very Good	0.833	3	7	80
Good	0.667	2	5	60
Neutral	0.500	1	3	40
Bad	0.333	0	1	20
Very bad	0.167	0	0	5
Worst	0.000	0	0	0

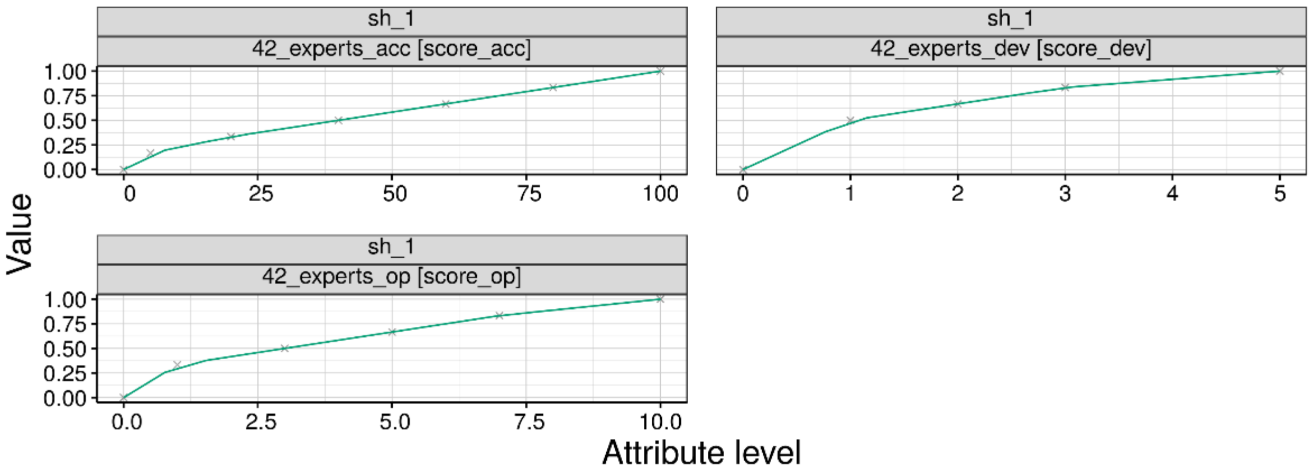


Figure S24: Value functions to aggregate the three sub-attributes into one [0:1] value. The functions were approximated from the level estimates by the expert using linear interpolation with 14 approximated points.

Table S25: Predictions of the three sub-attributes people able to 1) develop and maintain the system, 2) operate the system, and 3) access and interpret the forecasts and notifications for each system configuration with min/ max values of uncertainty in the estimate. These predictions are expert estimates done by Jafet Andersson and were aggregated into one value (last column), which is the value used as prediction for the aggregated attribute 42_experts in the subsequent MCDA modelling.

System configurations	Predicted people able to develop & maintain	Min	Max	Predicted people able to operate	Min	Max	Predicted people able to access & interpret	Min	Max	Aggregated value of prediction
a_Fast-dev	3	2	4	5	4	6	50	30	70	0.694
b_Res-user	3	2	4	3	2	4	90	54	100	0.750
c_Easy-use	2	1	3	4	3	5	80	48	100	0.694
d_Fast	1	0	2	4	3	5	70	42	98	0.611
e_Consent	0	0	0	6	4	8	90	54	100	0.556
f_Robust	2	1	3	7	5	9	90	54	100	0.806
g_Attractve	0	0	0	6	4	8	90	54	100	0.556
h_Equipp	0	0	0	6	4	8	100	60	100	0.583
i_Calibr	4	3	5	7	5	9	90	54	100	0.889
j_Cal-EO	4	3	5	6	4	8	90	54	100	0.861
k_Cal-EO-situ	4	3	5	5	4	6	90	54	126	0.833

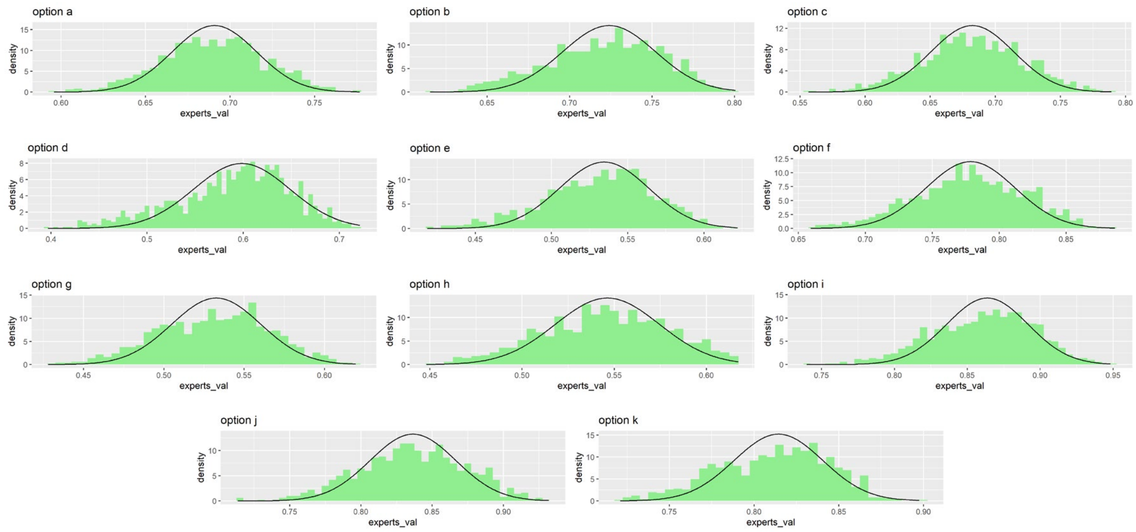


Figure S25: Distribution of resulting uncertainties (green) after aggregating the three sub-attributes and 1'000 Monte Carlo simulation runs. Results for the 11 FANFAR system configurations a – k. A normal distribution with mean = mean of Monte Carlo simulation, and sd = ¼ of the 95% confidence interval (black curve) fits the resulting uncertainty well. A normal distribution is thus used as input of the uncertainty of the aggregated prediction of the attribute 42_experts in the subsequent MCDA.

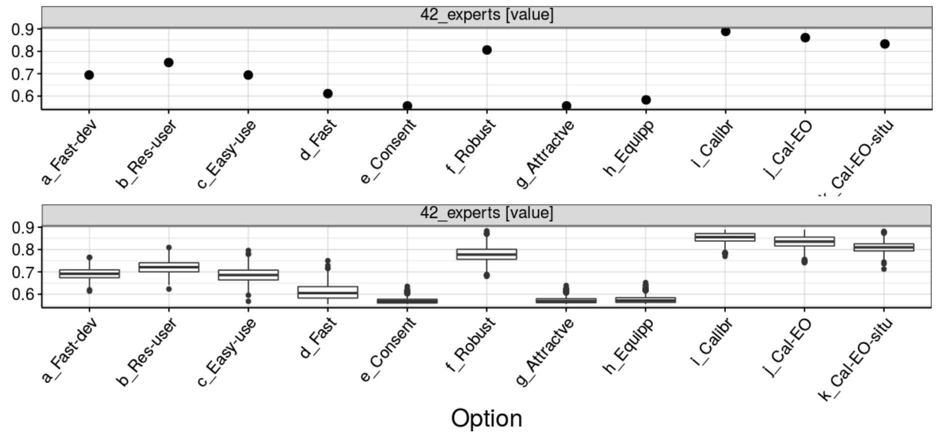


Figure S26: Aggregated value of predictions of all system configurations for the attribute 42_experts. The plot on top represents the predictions without uncertainty, the plot below shows the uncertainty distribution resulting from 1'000 Monte Carlo Simulation runs.

4.3. Good support system

Description:	*The support system should quickly help users with any questions or problems they have regarding fanfar.eu, IVP, Email, SMS, Knowledge base, Forum, Help Desk, H-TEP General Interface, H-TEP Community, H-TEP FANFAR App, and H-TEP Proc Services.*
Attribute:	*Information quantity in support system and time of response to resolve issues (43_suppsys)*
Unit:	*value [0:1]*
Description:	*Level of support system quality consisting of two sub-attributes (1) Information quantity, and 2) time for response of support system; transformed into a [0:1] value aggregated with equally weighted sum.*
Best case:	*The support system is informative (instructive answers to all issues are available in the knowledge base or forum, Score = 6.8). The support team responds quickly to questions posed (1 hour).*
Worst case:	*The support system is not informative. The support team does not react when there are queries and does not help to solve problems.*

Attribute details:

43_suppsys is an artificial (constructed) attribute consisting of two sub-attributes: 1) quantity of information available, and 2) time of response to solve issues.

For sub-attribute 1, we calculated a weighted sum to achieve a single score of the availability of information for each configuration. In this weighted sum, we assigned different weights (w) for different system components of the estimated information available in the knowledge base *(i.e., fanfar.eu [w=0.3], IVP [w=1.75], Email [w=0.8], SMS [w=0.8], Knowledge base [w=1], Forum [w=0.5], Help Desk [w=0.5], H-TEP General Interface [w=0.5], H-TEP Community [w=0.1], H-TEP FANFAR App [w=0.7], and H-TEP Proc Services [w=0.6])*. The estimates were elicited from Aytor Naranjo (IsardSAT) in July 2019. For the second sub-attribute, Aytor Naranjo estimated the minimum and maximum time required for a response to resolve issues. The average from these two estimates formed the prediction (Table S27). Levels were created for each sub-attribute (Worst, Very Bad, Bad, Neutral, Good, Very Good, Best; Table S26). These levels were then used as value functions for the aggregation of the two sub-attributes into one final value [0-1] using the weighted sum with equal weights for both sub-attributes (Figure S27). As consequence of the transformation of the sub-attributes into one [0:1] value, a linear value function [0:1] is used for this attribute for the subsequent MCDA modelling.

Uncertainty of predictions:

For the first sub-attribute "Quantity of information available", there was no uncertainty in the estimate. For the second sub-attribute "Time of response", minimum and maximum values were estimated, which were then averaged to calculate the prediction. The minimum and maximum times of response were included as uncertainty range in the aggregation of the two sub-attributes using a triangular distribution with mode = prediction, and min/max = min/max estimated time. To aggregate over the two sub-attributes, we used the fixed predictions for sub-attribute 1), and 1'000 Monte Carlo simulation runs, each of them drawing from the triangle uncertainty distribution of sub-attribute 2). For the subsequent MCDA, we used a normal distribution to model the resulting uncertainty of aggregated values. To estimate the shape of the aggregated uncertainty distribution, we chose to utilize a normal distribution with mean = mean of Monte Carlo simulation, and standard deviation = ¼ of the 95% confidence interval resulting from the Monte Carlo simulation with 1000 runs. This resulted in an individual uncertainty distribution for each configuration for this attribute (Figure S28).

Table S26: Expert estimates for levels of the two sub-attributes 1) quantity of information available, and 2) time of response to resolve issues. These levels were used to shape the value functions for the aggregation of the two sub-attributes into one [0:1] value.

Support system level	Value	Quantity of information available (score)	Time of response (h)
Best	1.000	6,8	1
Very Good	0.833	5,7	6
Good	0.667	4,5	24
Neutral	0.500	3,4	72
Bad	0.333	2,3	168
Very bad	0.167	1,1	720
Worst	0.000	0,0	26'280

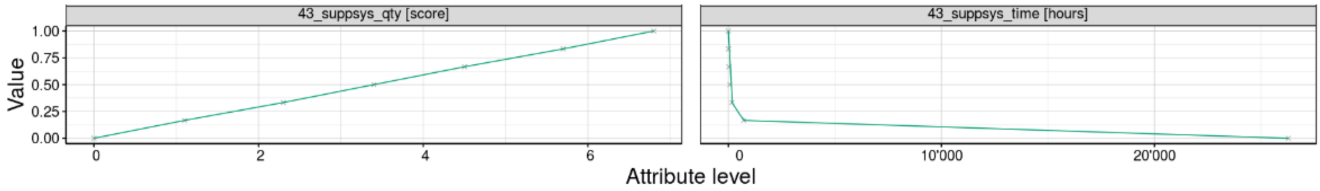


Figure S27: Value functions used to aggregate the two sub-attributes into one [0:1] value. The functions were approximated from the level estimates by the expert using linear interpolation with 140 approximated points.

Table S27: Predictions on: 1) the availability of information for different system components. These predictions were transformed into one score for the subsequent analysis using a weighted sum with different weights for the different system components. Predictions on: 2) the time of required for a response to resolve issues. These predictions result from averaging the min/max estimates by the expert. The total summed score and the predicted time of response were aggregated into one [0:1] value for the aggregated attribute 43_suppsys in the subsequent MCDA modelling (last column).

System configu-rations	fanfar_eu	IVP	Email	SMS	knowl. base	forum	help desk	H-TEP inter	H-TEP com	H-TEP app	H-TEP servi	Total summed score	Predicted time of response	Min	Max	Aggregated value of pre-diction
a_Fast-dev	1	1	0	0	1	0	0	1	1	1	1	4.46	2.50	1.0	4.0	0.805
b_Res-user	1	1	1	1	1	1	0	1	1	1	1	6.35	2.50	1.0	4.0	0.941
c_Easy-use	1	1	0	0	1	1	0	1	1	1	1	4.91	2.50	1.0	4.0	0.837
d_Fast	1	1	1	1	1	1	1	1	1	1	1	6.80	2.25	0.5	4.0	0.979
e_Consent	1	1	1	1	1	1	1	1	1	1	1	6.80	2.25	0.5	4.0	0.979
f_Robust	1	1	1	1	1	1	1	1	1	1	1	6.80	2.25	0.5	4.0	0.979
g_Attractve	1	1	1	1	1	1	1	1	1	1	1	6.80	2.25	0.5	4.0	0.979
h_Equipp	1	1	1	1	1	1	1	1	1	1	1	6.80	2.25	0.5	4.0	0.979
i_Calibr	1	1	1	1	1	1	1	1	1	1	1	6.80	2.50	1.0	4.0	0.975
j_Cal-EO	1	1	1	1	1	1	1	1	1	1	1	6.80	2.50	1.0	4.0	0.975
k_Cal-EO-situ	1	1	1	1	1	1	1	1	1	1	1	6.80	2.50	1.0	4.0	0.975

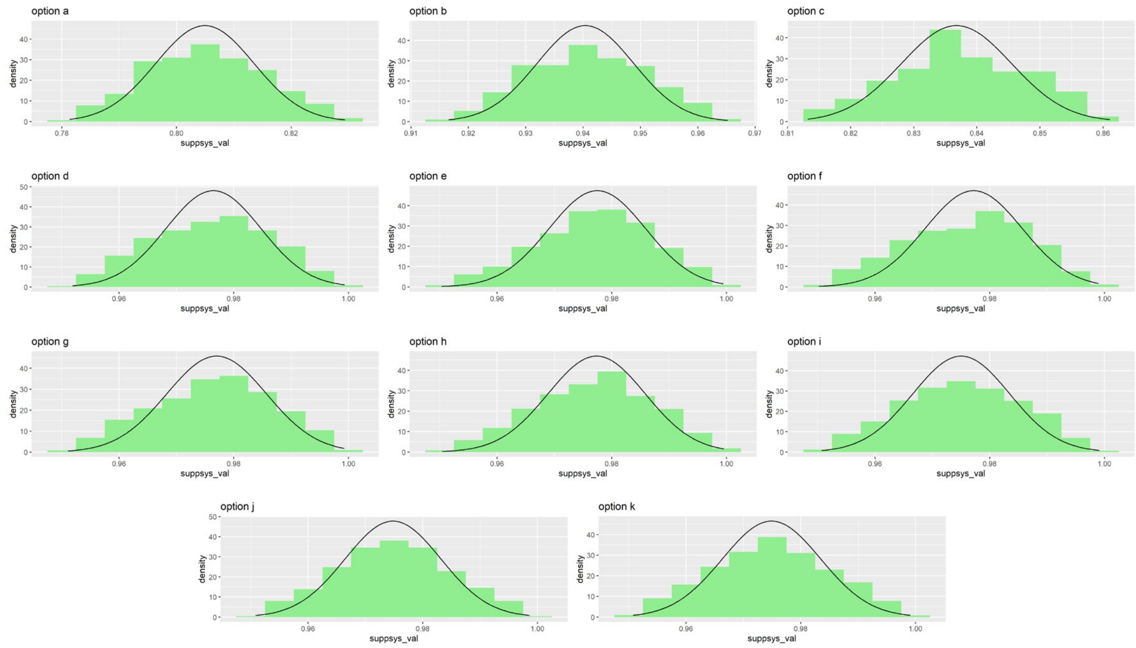


Figure S28: Distribution of resulting uncertainties (green) after aggregating the two sub-attributes and 1'000 Monte Carlo simulation runs. Results for the 11 FANFAR system configurations a – k. A normal distribution with mean = mean of Monte Carlo simulation, and sd = ¼ of the 95% confidence interval (black curve) fits the resulting uncertainty well. A normal distribution is thus used as input of the uncertainty of the aggregated prediction of the attribute 43_suppsys in the subsequent MCDA.

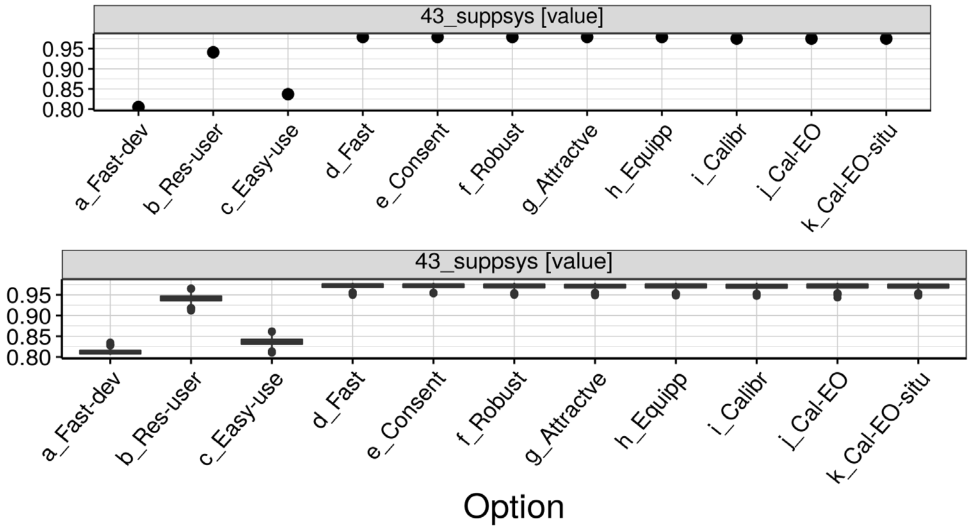


Figure S29: Visualization of the aggregated value of predictions of all system configurations for the attribute 43_suppsys. The plot on top represents the predictions without uncertainty, while the plot below shows the uncertainty distribution resulting from the Monte Carlo Simulation with 1'000 runs.

2.4.2 Summary of predictions

We summarized the information on how the predictions were calculated for each attribute given above (sect. S2.4.1) in overview tables: overview for the attributes belonging to the objectives *High information accuracy and clarity*, and *Good information access* (Table S28), and for the attributes of the remaining objectives of *Low costs*, and *High sustainability* (Table S29). To be able to model the MCDA, e.g., in the ValueDecisions app, this information has to be summarized again, and uncertainty needs to be included. These input data of the predicted performance of each FANFAR system configuration, including uncertainty, are presented in Table S30. Furthermore, we visualized the predicted performance on each attribute with and without uncertainty in the output graphs from the ValueDecisions app (Figure S30).

Table S28: Summary raw data of how predictions were calculated for each FANFAR system configuration (left column) based on sub-attributes for the objectives *High information accuracy and clarity*, and *Good information access*. Indicated bold are the aggregated values (v), ranging from [0,1], or the aggregated score for each attribute, stemming from the predictions on sub-attributes.

	Objective																
	High information accuracy and clarity								Good information access								
	High accuracy of information				Clear flood risk information				Reliable access to information				Timely production, distribution, and access to info				Several languages
Configuration	KGE (1 day)	KGE (3 days)	KGE (10 days)	Aggregated value (v)	Place where flood risk is defined	Possible to calculate flood risk?	Possible to apply to flood management?	Aggregated value (v)	Reliability of SMS	Reliability of Email	Reliability of web fanfar_eu	Weighted sum (score)	Production time (hour)	Access via distribution channel (min)	Access via H-TEP (min.)	Aggregated value (v)	Weighted sum (score)
a_Fast-dev	0.30	0.15	-2.00	0.45	Everywhere	Yes	No	0.667	none	none	sporadic	0.35	4.00	2.00	10.00	0.511	10.10
b_Res-user	0.79	0.49	-1.94	0.65	Everywhere	Yes	Yes	1.000	stable	stable	sporadic	2.05	4.00	0.50	12.00	0.618	17.64
c_Easy-use	0.42	0.24	-1.99	0.48	Everywhere	Yes	Yes	1.000	sporadic	sporadic	stable	1.69	5.80	1.00	11.00	0.526	17.44
d_Fast	0.46	0.26	-1.98	0.50	Everywhere	Yes	No	0.667	stable	stable	stable	2.20	4.50	0.50	10.00	0.633	10.10
e_Consent	0.86	0.58	0.40	0.74	Everywhere	Yes	Yes	1.000	stable	stable	sporadic	2.05	4.50	0.50	15.00	0.567	17.44
f_Robust	0.46	0.26	0.40	0.54	Everywhere	Yes	Yes	1.000	stable	stable	sporadic	2.05	6.30	0.50	13.00	0.557	17.64
g_Attractve	0.86	0.58	0.40	0.74	Everywhere	Yes	Yes	1.000	stable	stable	sporadic	2.05	5.50	0.50	15.00	0.544	17.64
h_Equipp	0.86	0.58	0.40	0.74	Everywhere	Yes	Yes	1.000	stable	stable	sporadic	2.05	6.30	0.50	15.00	0.531	17.64
i_Calibr	0.65	0.49	0.40	0.64	Everywhere	Yes	No	0.667	stable	stable	sporadic	2.05	6.30	0.50	12.00	0.571	16.80
j_Cal-EO	0.72	0.58	0.40	0.68	Everywhere	Yes	No	0.667	stable	stable	sporadic	2.05	5.00	0.50	12.00	0.596	16.80
k_Cal-EO-situ	0.93	0.75	0.40	0.83	Everywhere	Yes	No	0.667	stable	stable	sporadic	2.05	5.50	0.50	13.00	0.571	16.80

Table S29: Summary raw data of how predictions were calculated for each FANFAR system configuration (left column) based on sub-attributes for the objectives *Low costs*, and *High sustainability*. Indicated bold are the aggregated values (v), ranging from [0,1], or the aggregated score for each attribute, stemming from the predictions on sub-attributes.

	Objective													
	Low costs				High sustainability									
	Short d. time	Low operation costs			Long-term financing secured			Skillful human resources available				Good support system		
Configuration	Development time (days)	On site observation data	EO (satellite) data	Operation costs (v)	Costs covered (%)	Duration of financing (years)	Aggregated value (v)	People able to develop & maintain	People able to operate	P. able to access & interpret	Aggregated value (v)	Quantity of information available (score)	Time of response (hours)	Aggregated value (v)
a_Fast-dev	1	No	No	0.958	30%	1	0.091	3	5	50	0.694	4.46	2.50	0.805
b_Res-user	672	Yes	No	0.468	93%	4	0.670	3	3	90	0.750	6.35	2.50	0.941
c_Easy-use	600	No	Yes	0.492	83%	5	0.640	2	4	80	0.694	4.91	2.50	0.837
d_Fast	278	No	Yes	0.492	90%	3	0.553	1	4	70	0.611	6.80	2.25	0.979
e_Consent	1008	Yes	Yes	0.451	95%	5	0.777	0	6	90	0.556	6.80	2.25	0.979
f_Robust	645	No	Yes	0.492	89%	5	0.708	2	7	90	0.806	6.80	2.25	0.979
g_Attractve	1008	Yes	Yes	0.451	98%	5	0.811	0	6	90	0.556	6.80	2.25	0.979
h_Equipp	1008	Yes	Yes	0.451	84%	6	0.660	0	6	100	0.583	6.80	2.25	0.979
i_Calibr	402	No	No	0.571	95%	3	0.610	4	7	90	0.889	6.80	2.50	0.975
j_Cal-EO	432	No	Yes	0.492	96%	4	0.705	4	6	90	0.861	6.80	2.50	0.975
k_Cal-EO-situ	522	Yes	Yes	0.451	86%	5	0.674	4	5	90	0.833	6.80	2.50	0.975

Table S30. Predicted performance of all 11 FANFAR system configurations (a–k, columns) on each objective, respectively attribute (first column), based on expert estimates. Prediction: predicted performance of each configuration; unc_distr: uncertainty distributions used for MCDA modeling with corresponding parameters: mean and standard deviation (SD) for normal distribution; mode, min, and max for triangular distribution (triang).

Attribute	Parameter	a_Fast-dev	b_Res-user	c_Easy-use	d_Fast	e_Con-sent	f_Robust	g_Attrac-tve	h_Equipp	i_Calibr	j_Cal-EO	k_Cal-EO-situ
11_kge value [0:1]	Prediction	0.45	0.646	0.484	0.497	0.738	0.54	0.738	0.738	0.638	0.679	0.833
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.437	0.554	0.461	0.471	0.588	0.485	0.587	0.588	0.542	0.564	0.625
	SD	0.027	0.049	0.030	0.031	0.056	0.034	0.055	0.058	0.043	0.048	0.068
12_info value [0:1]	Prediction	0.667	1	1	0.667	1	1	1	1	0.667	0.667	0.667
	unc_distr	none	none	none	none	none	none	none	none	none	none	none
21_channel score [0:2.2]	Prediction	0.35	2.05	1.69	2.20	2.05	2.05	2.05	2.05	2.05	2.05	2.05
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.35	2.05	1.69	2.2	2.05	2.05	2.05	2.05	2.05	2.05	2.05
	SD	0.035	0.205	0.169	0.22	0.205	0.205	0.205	0.205	0.205	0.205	0.205
22_time value [0:1]	Prediction	0.511	0.618	0.526	0.633	0.567	0.557	0.544	0.531	0.571	0.596	0.571
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.499	0.611	0.528	0.612	0.579	0.558	0.558	0.546	0.564	0.587	0.571
	SD	0.040	0.031	0.031	0.031	0.028	0.027	0.027	0.026	0.027	0.029	0.029
23_langue score [0:17.64]	Prediction	10.1	17.64	17.44	10.1	17.44	17.64	17.64	17.64	16.8	16.8	16.8
	unc_distr	none	none	none	none	none	none	none	none	none	none	none
31_devlptime days [0:1095]	Prediction	1	672	600	278	1008	645	1008	1008	402	432	522
	unc_distr	triang	triang	triang	triang	triang	triang	triang	triang	triang	triang	triang
	Mode	1	537	480	222	806	516	806	806	321	345	417
	Min	1.01	806	720	333	1095	774	1095	1095	482	518	626
	Max	1.001	672	600	278	1008	645	1008	1008	402	432	522
32_costs Value [0:1]	Prediction	0.958	0.468	0.492	0.492	0.451	0.492	0.451	0.451	0.571	0.492	0.451
	unc_distr	triang	triang	triang	triang	triang	triang	triang	triang	triang	triang	triang
	Mode	0.933	0.441	0.472	0.472	0.416	0.472	0.416	0.416	0.497	0.472	0.416
	Min	0.983	0.495	0.587	0.587	0.486	0.587	0.486	0.486	0.662	0.587	0.486
	Max	0.958	0.468	0.492	0.492	0.451	0.492	0.451	0.451	0.571	0.492	0.451
41_finance value [0:1]	Prediction	0.091	0.67	0.64	0.553	0.777	0.708	0.811	0.66	0.61	0.705	0.674
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.073	0.632	0.606	0.534	0.731	0.676	0.754	0.653	0.572	0.654	0.641
	SD	0.01025	0.036	0.038	0.037	0.033	0.038	0.028	0.034	0.032	0.032	0.035
42_experts value [0:1]	Prediction	0.694	0.75	0.694	0.611	0.556	0.806	0.556	0.583	0.889	0.861	0.833
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.69	0.72	0.685	0.591	0.531	0.78	0.532	0.546	0.862	0.836	0.81
	SD	0.025	0.029	0.032	0.048	0.028	0.034	0.030	0.029	0.029	0.031	0.025
43_suppsys value [0:1]	Prediction	0.805	0.941	0.837	0.979	0.979	0.979	0.979	0.979	0.975	0.975	0.975
	unc_distr	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal	normal
	Mean	0.806	0.941	0.837	0.977	0.977	0.976	0.976	0.976	0.975	0.975	0.975
	SD	0.009	0.008	0.008	0.009	0.009	0.009	0.009	0.009	0.008	0.008	0.009

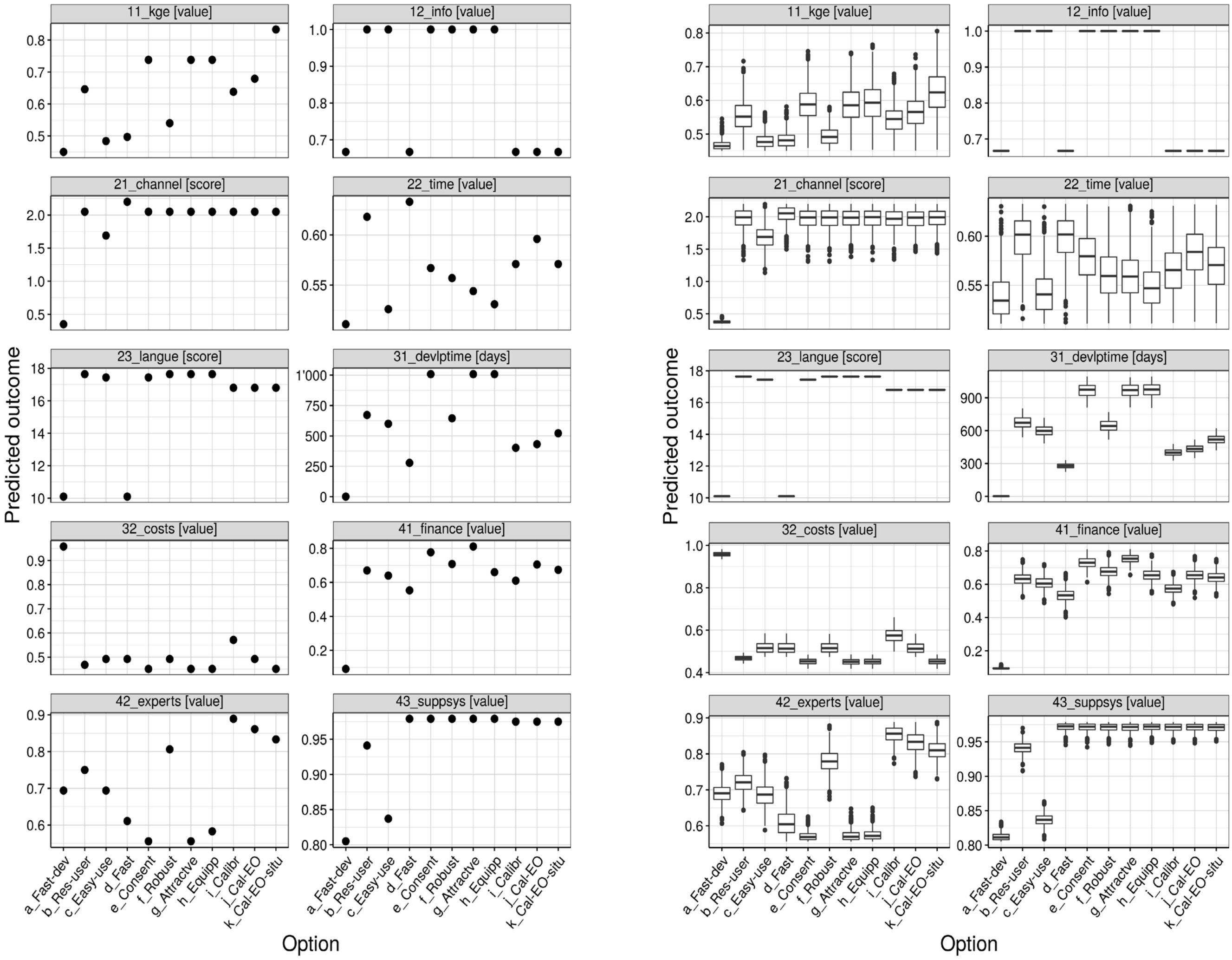


Figure S30. Predicted performance (y-axis) of each FANFAR system configuration (x-axis) on each attribute (boxes), according to the expert elicitation. Left panel: without uncertainty, right panel: with uncertainty).

2.5 Marginal value functions

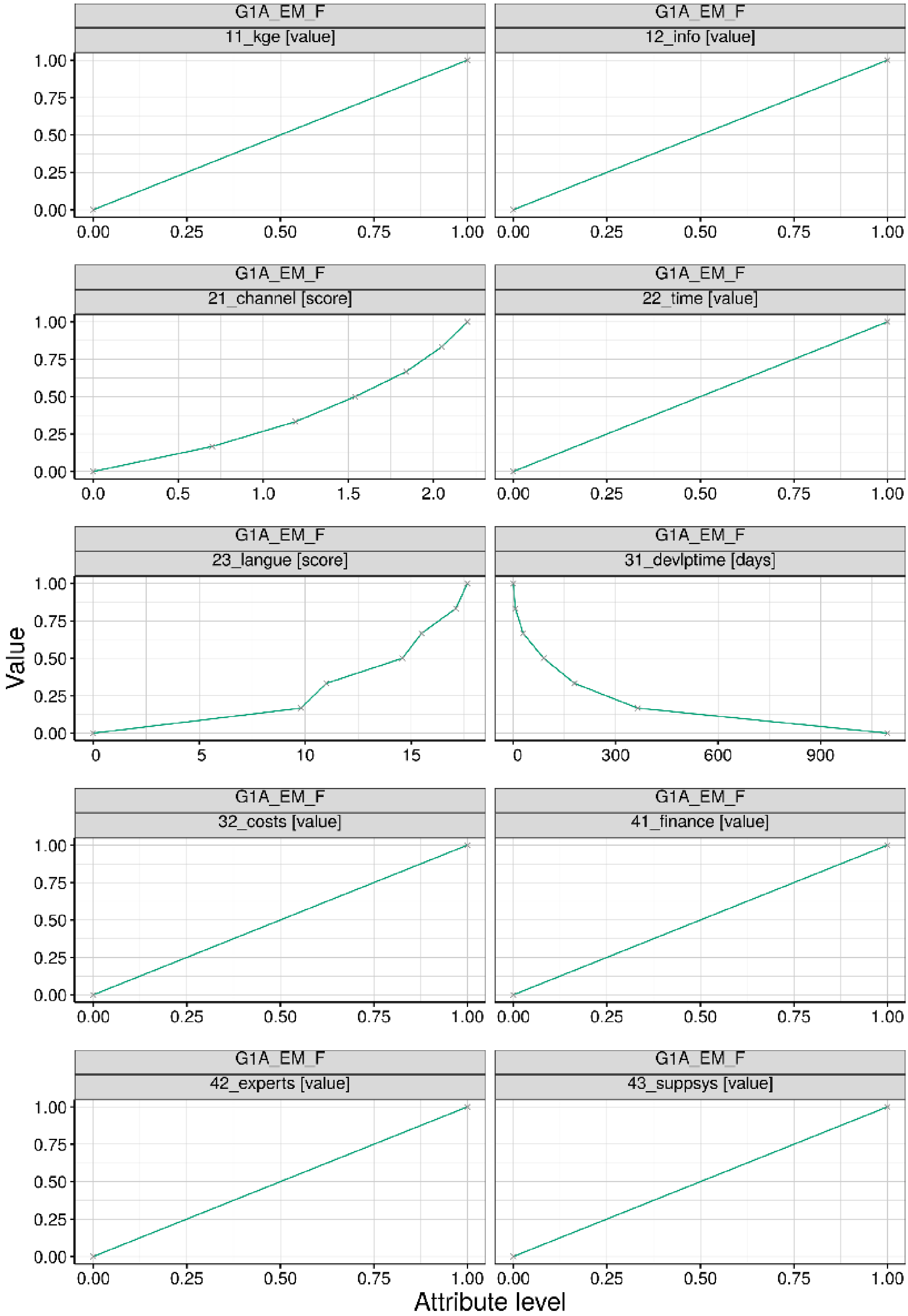


Figure S31. Marginal value functions used to transform predicted outcome of attributes into values. The example shows stakeholder group G1A_EM_F, but the marginal value functions were the same for all stakeholders in our case (see main text).

2.6 Weights

Over all groups, highest weights were assigned to the two objectives within the branch *1. High information accuracy and clarity*: *High accuracy of information (11_accur_info*; median over all groups = 0.150; 0.25 lower quartile = 0.134; 0.75 upper quartile = 0.169), and *Clear flood risk information (12_clear_info*; median=0.152; lower=0.134; upper=0.161; Figure S32*)*. For *11_accur_info*, there were two outliers: G1A_EM_F assigned an exceptionally low weight of 0.065, and G2A_HY_F assigned an exceptionally high weight of 0.251. Generally, slightly lower weights were assigned to *Timely production, distribution and access to info* (*22_timely_info*; median = 0.138; lower = 0.102; upper = 0.150), closely followed by *Reliable access to information* (*21_reliable_info*; median = 0.116; lower = 0.094; upper =0.142), and *Skillful human resources available* (*42_human_resour*; median = 0.106; lower = 0.102; upper = 0.126*).* Group G1A_EM_F, with an exceptionally high weight of 0.25, formed the outlier in *42_human_resour*.

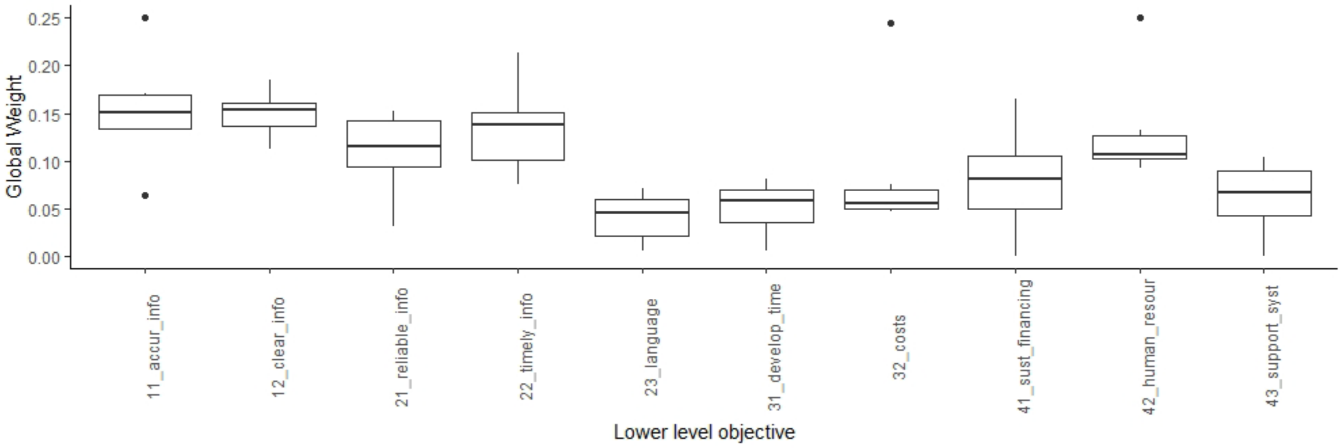


Figure S32. Global weight of each lower level objective over all stakeholder groups. The boxplot shows the 0.25, 0.5, and 0.75 quartiles of weights for each lower level objective. The whiskers extend to maximum and minimum points within 1.5 times the interquartile range. Points outside whiskers are outliers. Uncertainty of preference statements within groups are not represented.

Intermediate weights were assigned to *Long-term financing secured* (*41_sust_financing*; median = 0.081; lower = 0.050; upper = 0.105), and *Good support system* (*43_support_*syst; median = 0,067; lower = 0.042; upper=0.090). Over all groups, low weights were assigned to *Several languages* (*23_language*; median = 0.045; lower = 0.021; upper = 0.060), *Short development time* (*31_develop_time*; median = 0.059; lower = 0.035; upper = 0.070), and *Low acquisition and operation costs* (*32_costs*; (median = 0.056; lower = 0.036; upper = 0.070). G1A_EM_F was again an outlier for *32_costs* with an exceptionally high weight of 0.244.

The difference between groups in the assigned importance of objectives was notably larger for some objectives, e.g., *21_reliable_info*, *22_timely_info*, *41_sust_finance*, and *43_support_syst* (large interquartile ranges in Figure S32). Moreover, the strongly differing preferences of group G1A_EM_F is visualized by outliers in the boxplot of the higher level objectives, namely for *11_accur_info*, *32_costs*, and *42_human_resour* (Figure S33).

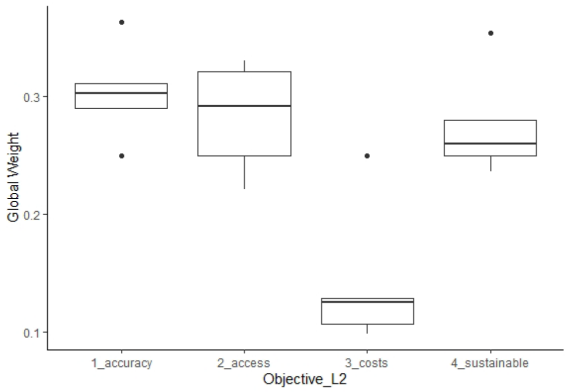


Figure S33. Global weight assigned to each higher level objective over all stakeholder groups. The boxplot shows the 0.25, 0.5, and 0.75 quartiles of weights for each lower level objective. The whiskers extend to the maximum and minimum points within 1.5 times the interquartile range. Points outside the whiskers are outliers. Uncertainty of preference statements within individual groups are not represented.

Table S31. Elicited weights and ranges for each stakeholder group (G1A_EM_F to G5A_AGRHYMET), elicited with the Swing or Simos' revised card procedure (see main text). We indicate the average local and global weights for each sub-objective, and the minimum and maximum range of the local and global weights. Note: the *local weight* is that weight assigned to a sub-objective within one branch of the hierarchy, which sums up to one. For instance, the local weights of sub-objective "*1.1 High accuracy of information*" and "*1.2 Clear flood risk information*" sum up to 1 for the higher level objective "*1 High information accuracy and clarity*". Over all ten sub-objectives of the hierarchy, the *global weights* sum up to 1. We also indicate the minimum and maximum, respectively, of the weight ranges as elicited in the stakeholder workshops, and the deviation from the average in percent. Yellow highlights: deviation >0.02, which was tested in local sensitivity analyses. Green highlights: deviation >0.02, but not tested in sensitivity analyses because objective was not sensitive to weight changes (see Methods).

G1A_EM_F (Swing)

Objective	Average local weight (w)	Local weight range min (wmin)	Local weight range max (wmax)	Average global weight (W)-	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	0.259	0.167	0.333	0.065	0.042 (-0.023)	0.083 (+0.019)
1.2 Clear flood risk information	0.741	0.667	0.833	0.185	0.167 (-0.019)	0.208 (+0.023)
2.1 Reliable access to information	0.128	0.087	0.167	0.032	0.022 (-0.010)	0.042 (+0.010)
2.2 Timely production, distribution, access to info	0.851	0.800	0.909	0.213	0.200 (-0.013)	0.227 (+0.015)
2.3 Several languages	0.021	0.000	0.043	0.005	0.000 (-0.005)	0.011 (+0.006)
3.1 Short development time	0.024	0.000	0.048	0.006	0.000 (-0.006)	0.012 (+0.006)
3.2 Low acquisition and operation costs	0.976	0.952	1.000	0.244	0.238 (-0.006)	0.250 (+0.006)
4.1 Long-term financing secured	0.000	0.000	0.000	0.000	0.000 (-0)	0.000 (+0)
4.2 Skillful human resources available	1.000	1.000	1.000	0.250	0.250 (-0)	0.250 (+0)
4.3 Good support system	0.000	0.000	0.000	0.000	0.000 (-0)	0.000 (+0)

G2A_HY_F (Swing)

Objective	Average local weight (w)	Local weight range min (wmin)	Local weight range max (wmax)	Average global weight (W)	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	0.690	0.690	0.690	0.251	0.238 (-0.013)	0.265 (+0.014)
1.2 Clear flood risk information	0.310	0.310	0.310	0.113	0.107 (-0.006)	0.119 (+0.007)
2.1 Reliable access to information	0.400	0.385	0.417	0.109	0.096 (-0.013)	0.123 (+0.014)
2.2 Timely production, distribution, access to info	0.340	0.313	0.365	0.093	0.078 (-0.015)	0.108 (+0.016)
2.3 Several languages	0.260	0.250	0.271	0.071	0.063 (-0.008)	0.080 (+0.009)
3.1 Short development time	0.556	0.556	0.556	0.071	0.060 (-0.011)	0.082 (+0.012)
3.2 Low acquisition and operation costs	0.444	0.444	0.444	0.057	0.048 (-0.009)	0.066 (+0.009)
4.1 Long-term financing secured	0.449	0.435	0.465	0.106	0.093 (-0.013)	0.121 (+0.014)
4.2 Skillful human resources available	0.393	0.378	0.409	0.093	0.081 (-0.012)	0.106 (+0.013)
4.3 Good support system	0.157	0.136	0.178	0.037	0.029 (-0.008)	0.046 (+0.009)

G2B_HY_F (Swing)

Objective	Average local weight (w)	Local weight range min (wmin)	Local weight range max (wmax)	Average global weight (W)	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	0.459	0.429	0.487	0.138	0.111 (-0.027)	0.168 (+0.030)
1.2 Clear flood risk information	0.541	0.513	0.571	0.163	0.133 (-0.030)	0.197 (+0.035)
2.1 Reliable access to information	0.400	0.385	0.417	0.088	0.067 (-0.021)	0.112 (+0.023)
2.2 Timely production, distribution, access to info	0.340	0.313	0.365	0.075	0.055 (-0.020)	0.098 (+0.023)
2.3 Several languages	0.260	0.250	0.271	0.058	0.044 (-0.014)	0.073 (+0.015)
3.1 Short development time	0.556	0.556	0.556	0.069	0.056 (-0.013)	0.084 (+0.015)
3.2 Low acquisition and operation costs	0.444	0.444	0.444	0.055	0.044 (-0.011)	0.067 (+0.012)
4.1 Long-term financing secured	0.465	0.455	0.476	0.165	0.147 (-0.018)	0.187 (+0.022)
4.2 Skillful human resources available	0.372	0.364	0.381	0.132	0.117 (-0.014)	0.149 (+0.018)
4.3 Good support system	0.163	0.143	0.182	0.058	0.046 (-0.012)	0.071 (+0.014)

G3A_HY_E (Simos' card)

Objective	rank	Focus global weight (W) [Z=10]	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	1	0.171	0.159 (-0.013)	0.178 (+0.007)
1.2 Clear flood risk information	5	0.133	0.127 (-0.006)	0.136 (+0.003)
2.1 Reliable access to information	3	0.152	0.143 (-0.009)	0.157 (+0.005)
2.2 Timely production, distribution, access to info	3	0.152	0.143 (-0.009)	0.157 (+0.005)
2.3 Several languages	17	0.017	0.009 (-0.008)	0.032 (+0.015)
3.1 Short development time	15.5	0.032	0.025 (-0.007)	0.044 (+0.012)
3.2 Low acquisition and operation costs	11	0.075	0.072 (-0.002)	0.079 (+0.004)
4.1 Long-term financing secured	12.5	0.060	0.057 (-0.004)	0.067 (+0.007)
4.2 Skillful human resources available	8	0.104	0.103 (-0.001)	0.104 (+0.0)
4.3 Good support system	8	0.104	0.103 (-0.001)	0.104 (+0.0)
Max/Sum	17	1	-	-

G4A_EM_E (Simos' card)

Objective	rank	Focus global weight (W) [Z=3.33]	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	4.5	0.133	0.126 (-0.007)	0.148 (+0.0)
1.2 Clear flood risk information	1	0.157	0.122 (-0.012)	0.157 (+0.0)
2.1 Reliable access to information	6	0.123	0.118 (-0.005)	0.135 (+0.0)
2.2 Timely production, distribution, access to info	2.5	0.147	0.135 (-0.010)	0.147 (+0.0)
2.3 Several languages	15	0.061	0.044 (-0.0)	0.069 (+0.008)
3.1 Short development time	12	0.081	0.054 (-0.0)	0.085 (+0.004)
3.2 Low acquisition and operation costs	17	0.047	0.047 (-0.0)	0.083 (+0.011)
4.1 Long-term financing secured	17	0.047	0.047 (-0.0)	0.073 (+0.011)
4.2 Skillful human resources available	8	0.109	0.103 (-0.002)	0.109 (+0.0)
4.3 Good support system	10	0.095	0.095 (-0.0)	0.103 (+0.001)
Max/Sum	17	1	-	-

G5A_AGRHYMET (Simos' card)

Objective	rank	Focus global weight (W) [Z=5]	Global weight range min (Wmin)	Global weight range max (Wmax)
1.1 High accuracy of information	1	0.162	0.151 (-0.012)	0.162 (+0.0)
1.2 Clear flood risk information	3	0.149	0.140 (-0.009)	0.149 (+0.0)
2.1 Reliable access to information	3	0.149	0.140 (-0.009)	0.149 (+0.0)
2.2 Timely production, distribution, access to info	6	0.129	0.124 (-0.005)	0.129 (+0.0)
2.3 Several languages	20.5	0.032	0.032 (-0.0)	0.045 (+0.013)
3.1 Short development time	18	0.049	0.049 (-0.0)	0.059 (+0.010)
3.2 Low acquisition and operation costs	18	0.049	0.049 (-0.0)	0.059 (+0.010)
4.1 Long-term financing secured	10	0.102	0.102 (-0.0)	0.102 (+0.0)
4.2 Skillful human resources available	10	0.102	0.102 (-0.0)	0.102 (+0.0)
4.3 Good support system	14	0.076	0.076 (-0.0)	0.080 (+0.005)
Max/Sum	20.5	0.999	-	-

2.7 MCDA results: values and ranks

Table S32. Total aggregated value and rank of all 11 FANFAR system configurations (first column) for each stakeholder group (G1A_EM_F, G2A_HY_F, etc.), without uncertainty of predictions. The higher the value, and the lower the rank, the better the configuration achieves the objectives, given the expert predictions and the stakeholders' preferences. A total value of *v = 1* indicates that for this configuration (and stakeholder group), all objectives were able to achieve the best level (given the available system configurations). In contrast, *v = 0* indicates that only the worst level of all objectives was achieved by this configuration and this stakeholder group. Rank 1 is the best rank, i.e., the best performing configuration, rank 2 the second best, etc., and rank 11 is the worst performing configuration.

System Configuration	G1A_EM_F		G2A_HY_F		G2B_HY_F		G3A_HY_E		G4A_EM_E		G5A_AGRHYMET	
	value	rank	value	rank	value	rank	value	rank	value	rank	value	rank
a_Fast-dev	0.644	5	0.398	11	0.400	11	0.433	11	0.457	11	0.398	11
b_Res-user	0.671	1	0.653	2	0.674	1	0.704	1	0.671	1	0.694	1
c_Easy-use	0.625	7	0.569	9	0.609	9	0.609	10	0.594	9	0.605	10
d_Fast	0.590	11	0.545	10	0.565	10	0.641	9	0.588	10	0.616	9
e_Consent	0.608	8	0.617	7	0.629	7	0.676	6	0.609	7	0.660	7
f_Robust	0.669	2	0.631	5	0.671	2	0.686	4	0.659	2	0.680	3
g_Attractve	0.603	10	0.620	6	0.634	6	0.674	7	0.609	6	0.661	6
h_Equipp	0.607	9	0.608	8	0.615	8	0.667	8	0.604	8	0.648	8
i_Calibr	0.668	3	0.633	4	0.645	5	0.682	5	0.646	5	0.664	5
j_Cal-EO	0.648	4	0.647	3	0.658	4	0.689	3	0.651	4	0.676	4
k_Cal-EO-situ	0.632	6	0.665	1	0.660	3	0.698	2	0.654	3	0.684	2

Table S33. Results of the MCDA without uncertainty. For each FANFAR system configuration (e.g., a_Fast-dev, etc.) and each stakeholder group (e.g., G1A_EM_F, etc.). We show the partial values of each lower level objective (partial), and the total value aggregated over all objectives (aggregated). Partial values were calculated with objectives' weights according to the aggregation model (attribute value$^{0.2}$ × objective weight), and the total aggregated according to the aggregation model (sum of partial values$^{1/0.2}$).

		G1A_EM_F		G2A_HY_F		G2B_HY_F		G3A_HY_E		G4A_EM_E		G5A_AGRHYMET	
configuration	objective	partial	aggregated	partial	aggregated	partial	aggregated	partial	aggregated	partial	aggregated	partial	aggregated
	11_accur_info	0.055		0.214		0.118		0.146		0.113		0.138	
	12_clear_info	0.171		0.104		0.150		0.123		0.145		0.137	
	21_reliable_info	0.019		0.066		0.054		0.093		0.075		0.091	
	22_timely_info	0.186		0.081		0.066		0.133		0.129		0.113	
	23_language	0.004		0.052		0.042		0.012		0.045		0.023	
	31_develop_time	0.006		0.071		0.069		0.032		0.081		0.049	
	32_costs	0.242		0.057		0.055		0.074		0.047		0.049	
	41_sust_financing	0.000		0.066		0.102		0.037		0.029		0.063	
	42_human_resour	0.232		0.086		0.123		0.097		0.101		0.095	
a_Fast-dev	43_support_syst	0.000	0.644	0.035	0.398	0.056	0.400	0.100	0.433	0.091	0.457	0.073	0.398
	11_accur_info	0.060		0.230		0.126		0.157		0.122		0.148	
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.193		0.084		0.068		0.138		0.134		0.117	
	23_language	0.005		0.071		0.058		0.017		0.061		0.032	
	31_develop_time	0.004		0.045		0.043		0.020		0.051		0.031	
	32_costs	0.210		0.049		0.047		0.064		0.040		0.042	
	41_sust_financing	0.000		0.098		0.152		0.055		0.043		0.094	
	42_human_resour	0.236		0.088		0.125		0.098		0.103		0.096	
b_Res-user	43_support_syst	0.000	0.671	0.037	0.653	0.057	0.674	0.103	0.704	0.094	0.671	0.075	0.694

configu-ration	objective	G1A_EM_F partial	G1A_EM_F aggre-gated	G2A_HY_F partial	G2A_HY_F aggre-gated	G2B_HY_F partial	G2B_HY_F aggre-gated	G3A_HY_E partial	G3A_HY_E aggre-gated	G4A_EM_E partial	G4A_EM_E aggre-gated	G5A_AGRHYMET partial	G5A_AGRHYMET aggre-gated
c_Easy-use	11_accur_info	0.056	0.625	0.217	0.569	0.119	0.609	0.148	0.609	0.115	0.594	0.140	0.605
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.029		0.098		0.079		0.136		0.110		0.134	
	22_timely_info	0.187		0.082		0.066		0.134		0.129		0.113	
	23_language	0.005		0.070		0.057		0.017		0.060		0.032	
	31_develop_time	0.004		0.046		0.045		0.021		0.052		0.032	
	32_costs	0.212		0.049		0.048		0.065		0.041		0.043	
	41_sust_financing	0.000		0.097		0.151		0.055		0.043		0.093	
	42_human_resour	0.232		0.086		0.123		0.097		0.101		0.095	
	43_support_syst	0.000		0.036		0.056		0.100		0.092		0.073	
d_Fast	11_accur_info	0.057	0.590	0.218	0.545	0.120	0.565	0.149	0.641	0.116	0.588	0.141	0.616
	12_clear_info	0.171		0.104		0.150		0.123		0.145		0.137	
	21_reliable_info	0.032		0.109		0.088		0.152		0.123		0.149	
	22_timely_info	0.194		0.085		0.068		0.139		0.134		0.118	
	23_language	0.004		0.052		0.042		0.012		0.045		0.023	
	31_develop_time	0.005		0.054		0.052		0.024		0.061		0.037	
	32_costs	0.212		0.049		0.048		0.065		0.041		0.043	
	41_sust_financing	0.000		0.094		0.147		0.053		0.042		0.091	
	42_human_resour	0.227		0.084		0.120		0.094		0.099		0.092	
	43_support_syst	0.000		0.037		0.058		0.104		0.095		0.076	
e_Consent	11_accur_info	0.061	0.608	0.236	0.617	0.130	0.629	0.161	0.676	0.125	0.609	0.152	0.660
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.190		0.083		0.067		0.136		0.131		0.115	
	23_language	0.005		0.070		0.057		0.017		0.060		0.032	
	31_develop_time	0.003		0.032		0.032		0.015		0.037		0.022	
	32_costs	0.208		0.049		0.047		0.064		0.040		0.042	
	41_sust_financing	0.000		0.101		0.157		0.057		0.045		0.097	
	42_human_resour	0.222		0.083		0.117		0.092		0.097		0.091	
	43_support_syst	0.000		0.037		0.058		0.104		0.095		0.076	
f_Robust	11_accur_info	0.057	0.669	0.222	0.631	0.122	0.671	0.151	0.686	0.118	0.659	0.143	0.680
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.189		0.083		0.067		0.135		0.131		0.115	
	23_language	0.005		0.071		0.058		0.017		0.061		0.032	
	31_develop_time	0.004		0.045		0.044		0.020		0.051		0.031	
	32_costs	0.212		0.049		0.048		0.065		0.041		0.043	
	41_sust_financing	0.000		0.099		0.154		0.056		0.044		0.095	
	42_human_resour	0.239		0.089		0.126		0.100		0.104		0.098	
	43_support_syst	0.000		0.037		0.058		0.104		0.095		0.076	

		G1A_EM_F		G2A_HY_F		G2B_HY_F		G3A_HY_E		G4A_EM_E		G5A_AGRHYMET	
configu-ration	objective	partial	aggre-gated	partial	aggre-gated	partial	aggre-gated	partial	aggre-gated	partial	aggre-gated	partial	aggre-gated
	11_accur_info	0.061		0.236		0.130		0.161		0.125		0.152	
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.189		0.082		0.066		0.135		0.130		0.114	
	23_language	0.005		0.071		0.058		0.017		0.061		0.032	
	31_develop_time	0.003		0.032		0.032		0.015		0.037		0.022	
	32_costs	0.208		0.049		0.047		0.064		0.040		0.042	
	41_sust_financing	0.000		0.102		0.158		0.058		0.045		0.098	
	42_human_resour	0.222		0.083		0.117		0.092		0.097		0.091	
g_Attractve	43_support_syst	0.000	0.603	0.037	0.620	0.058	0.634	0.104	0.674	0.095	0.609	0.076	0.661
	11_accur_info	0.061		0.236		0.130		0.161		0.125		0.152	
	12_clear_info	0.185		0.113		0.163		0.133		0.157		0.149	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.188		0.082		0.066		0.134		0.130		0.114	
	23_language	0.005		0.071		0.058		0.017		0.061		0.032	
	31_develop_time	0.003		0.032		0.032		0.015		0.037		0.022	
	32_costs	0.208		0.049		0.047		0.064		0.040		0.042	
	41_sust_financing	0.000		0.098		0.152		0.055		0.043		0.094	
	42_human_resour	0.224		0.083		0.118		0.093		0.098		0.092	
h_Equipp	43_support_syst	0.000	0.607	0.037	0.608	0.058	0.615	0.104	0.667	0.095	0.604	0.076	0.648
	11_accur_info	0.059		0.229		0.126		0.156		0.122		0.148	
	12_clear_info	0.171		0.104		0.150		0.123		0.145		0.137	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.190		0.083		0.067		0.136		0.131		0.115	
	23_language	0.005		0.068		0.055		0.016		0.058		0.031	
	31_develop_time	0.004		0.049		0.048		0.022		0.056		0.034	
	32_costs	0.218		0.051		0.049		0.067		0.042		0.044	
	41_sust_financing	0.000		0.096		0.149		0.054		0.043		0.092	
	42_human_resour	0.244		0.091		0.129		0.102		0.106		0.100	
i_Calibr	43_support_syst	0.000	0.668	0.037	0.633	0.058	0.645	0.103	0.682	0.095	0.646	0.076	0.664
	11_accur_info	0.060		0.232		0.128		0.158		0.123		0.150	
	12_clear_info	0.171		0.104		0.150		0.123		0.145		0.137	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.192		0.084		0.068		0.137		0.133		0.116	
	23_language	0.005		0.068		0.055		0.016		0.058		0.031	
	31_develop_time	0.004		0.049		0.047		0.022		0.056		0.034	
	32_costs	0.212		0.049		0.048		0.065		0.041		0.043	
	41_sust_financing	0.000		0.099		0.154		0.056		0.044		0.095	
	42_human_resour	0.243		0.090		0.128		0.101		0.106		0.099	
j_Cal-EO	43_support_syst	0.000	0.648	0.037	0.647	0.058	0.658	0.103	0.689	0.095	0.651	0.076	0.676

configu-ration	objective	G1A_EM_F partial	G1A_EM_F aggre-gated	G2A_HY_F partial	G2A_HY_F aggre-gated	G2B_HY_F partial	G2B_HY_F aggre-gated	G3A_HY_E partial	G3A_HY_E aggre-gated	G4A_EM_E partial	G4A_EM_E aggre-gated	G5A_AGRHYMET partial	G5A_AGRHYMET aggre-gated
k_Cal-EO-situ	11_accur_info	0.063		0.242		0.133		0.165		0.128		0.156	
	12_clear_info	0.171		0.104		0.150		0.123		0.145		0.137	
	21_reliable_info	0.031		0.105		0.085		0.147		0.119		0.144	
	22_timely_info	0.190		0.083		0.067		0.136		0.131		0.115	
	23_language	0.005		0.068		0.055		0.016		0.058		0.031	
	31_develop_time	0.004		0.047		0.046		0.021		0.054		0.033	
	32_costs	0.208		0.049		0.047		0.064		0.040		0.042	
	41_sust_financing	0.000		0.098		0.152		0.055		0.043		0.094	
	42_human_resour	0.241		0.090		0.127		0.100		0.105		0.098	
	43_support_syst	0.000	0.632	0.037	0.665	0.058	0.660	0.103	0.698	0.095	0.654	0.076	0.684

Table S34. Results of the MCDA with uncertainty. MCDA values resulting from Monte Carlo simulation of uncertainty on predictions with 1’000 runs. For each stakeholder (e.g., G1A_EM_F, etc.) and each configuration (a_Fast-dev, etc.), the calculated mean and median value are presented with parameters of the according uncertainty distribution.

		a_Fast-dev	b_Res-user	c_Easy-use	d_Fast	e_Con-sent	f_Rob-ust	g_At-tractve	h_Equip	i_Calibr	j_Cal-EO	k_Cal-EO-situ
G1A_EM_F	mean value	0.652	0.651	0.633	0.584	0.604	0.666	0.600	0.599	0.652	0.638	0.615
	0.05 quantile	0.639	0.635	0.613	0.566	0.588	0.644	0.585	0.584	0.632	0.620	0.596
	0.95 quantile	0.666	0.666	0.653	0.606	0.619	0.688	0.615	0.615	0.674	0.658	0.632
	median value	0.651	0.652	0.633	0.583	0.604	0.666	0.600	0.598	0.652	0.638	0.615
	0.25 quartile	0.646	0.645	0.624	0.576	0.597	0.657	0.594	0.593	0.644	0.630	0.607
	0.75 quartile	0.657	0.659	0.642	0.592	0.610	0.675	0.606	0.605	0.661	0.647	0.621
G2A_HY_F	mean value	0.408	0.616	0.566	0.525	0.581	0.609	0.584	0.574	0.596	0.606	0.608
	0.05 quantile	0.402	0.586	0.545	0.505	0.544	0.585	0.550	0.540	0.568	0.578	0.573
	0.95 quantile	0.415	0.646	0.588	0.545	0.617	0.633	0.618	0.609	0.622	0.634	0.642
	median value	0.408	0.616	0.566	0.525	0.582	0.609	0.585	0.574	0.597	0.606	0.609
	0.25 quartile	0.405	0.604	0.557	0.516	0.567	0.599	0.570	0.559	0.585	0.592	0.594
	0.75 quartile	0.411	0.628	0.574	0.534	0.596	0.620	0.599	0.590	0.608	0.617	0.622
G2B_HY_F	mean value	0.408	0.645	0.604	0.547	0.607	0.652	0.612	0.597	0.616	0.627	0.622
	0.05 quantile	0.402	0.621	0.583	0.528	0.575	0.629	0.582	0.568	0.593	0.605	0.595
	0.95 quantile	0.414	0.667	0.625	0.566	0.636	0.673	0.639	0.625	0.636	0.649	0.646
	median value	0.407	0.645	0.604	0.547	0.608	0.652	0.612	0.598	0.617	0.627	0.622
	0.25 quartile	0.405	0.636	0.596	0.539	0.596	0.643	0.600	0.584	0.607	0.618	0.612
	0.75 quartile	0.410	0.655	0.612	0.555	0.620	0.661	0.624	0.609	0.626	0.636	0.632
G3A_HY_E	mean value	0.444	0.670	0.609	0.612	0.646	0.666	0.645	0.638	0.650	0.654	0.653
	0.05 quantile	0.437	0.637	0.583	0.586	0.611	0.635	0.612	0.605	0.616	0.622	0.618
	0.95 quantile	0.453	0.701	0.637	0.636	0.680	0.695	0.677	0.670	0.678	0.686	0.686
	median value	0.444	0.671	0.609	0.613	0.646	0.667	0.646	0.638	0.651	0.655	0.653
	0.25 quartile	0.441	0.657	0.597	0.602	0.633	0.653	0.631	0.623	0.637	0.640	0.638
	0.75 quartile	0.448	0.685	0.620	0.624	0.661	0.680	0.661	0.652	0.663	0.669	0.668

		a_Fast-dev	b_Res-user	c_Easy-use	d_Fast	e_Con-sent	f_Rob-ust	g_At-tractve	h_Equipp	i_Calibr	j_Cal-EO	k_Cal-EO-situ
G4A_EM_E	mean value	0.467	0.644	0.594	0.565	0.593	0.642	0.593	0.588	0.621	0.624	0.619
	0.05 quantile	0.460	0.616	0.573	0.544	0.556	0.617	0.559	0.555	0.594	0.599	0.591
	0.95 quantile	0.475	0.669	0.617	0.585	0.625	0.666	0.625	0.620	0.644	0.649	0.646
	median value	0.466	0.645	0.594	0.565	0.594	0.643	0.594	0.588	0.622	0.624	0.619
	0.25 quartile	0.464	0.633	0.585	0.556	0.580	0.632	0.579	0.575	0.611	0.613	0.607
	0.75 quartile	0.470	0.656	0.603	0.575	0.607	0.653	0.608	0.602	0.632	0.636	0.631
G5A_AGRHYMET	mean value	0.407	0.661	0.603	0.588	0.632	0.658	0.633	0.622	0.632	0.641	0.639
	0.05 quantile	0.401	0.628	0.576	0.562	0.596	0.628	0.599	0.587	0.600	0.610	0.605
	0.95 quantile	0.415	0.690	0.631	0.611	0.666	0.686	0.666	0.655	0.659	0.671	0.672
	median value	0.407	0.662	0.603	0.588	0.633	0.660	0.633	0.622	0.634	0.641	0.639
	0.25 quartile	0.404	0.648	0.592	0.577	0.619	0.646	0.619	0.608	0.620	0.627	0.625
	0.75 quartile	0.410	0.675	0.614	0.600	0.647	0.672	0.648	0.637	0.646	0.655	0.654

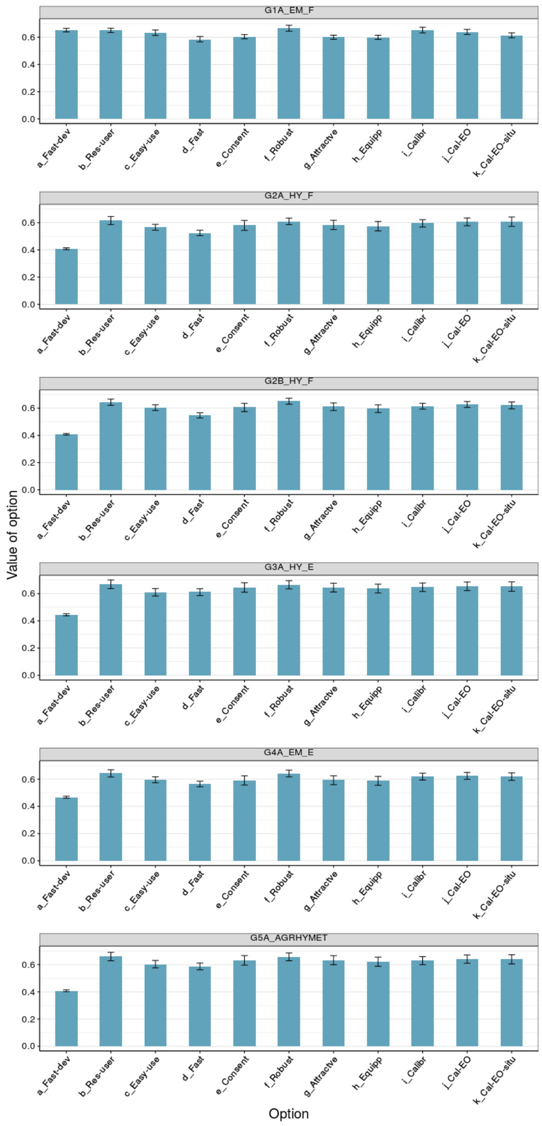


Figure S34. MCDA results with uncertainty. Mean (average) total value (y-axis) calculated by the MCDA for each configuration (x-axis) and each stakeholder (boxes). Uncertainty of the predictions is included with 1'000 Monte Carlo simulation runs of the MCDA for each stakeholder and each configuration. The mean overall value (y-axis) can be anywhere between 0 (none of the objectives are achieved at all) and 1 (all objectives are fully achieved). Error bars show the 0.05 and 0.95 quantiles of the overall value.

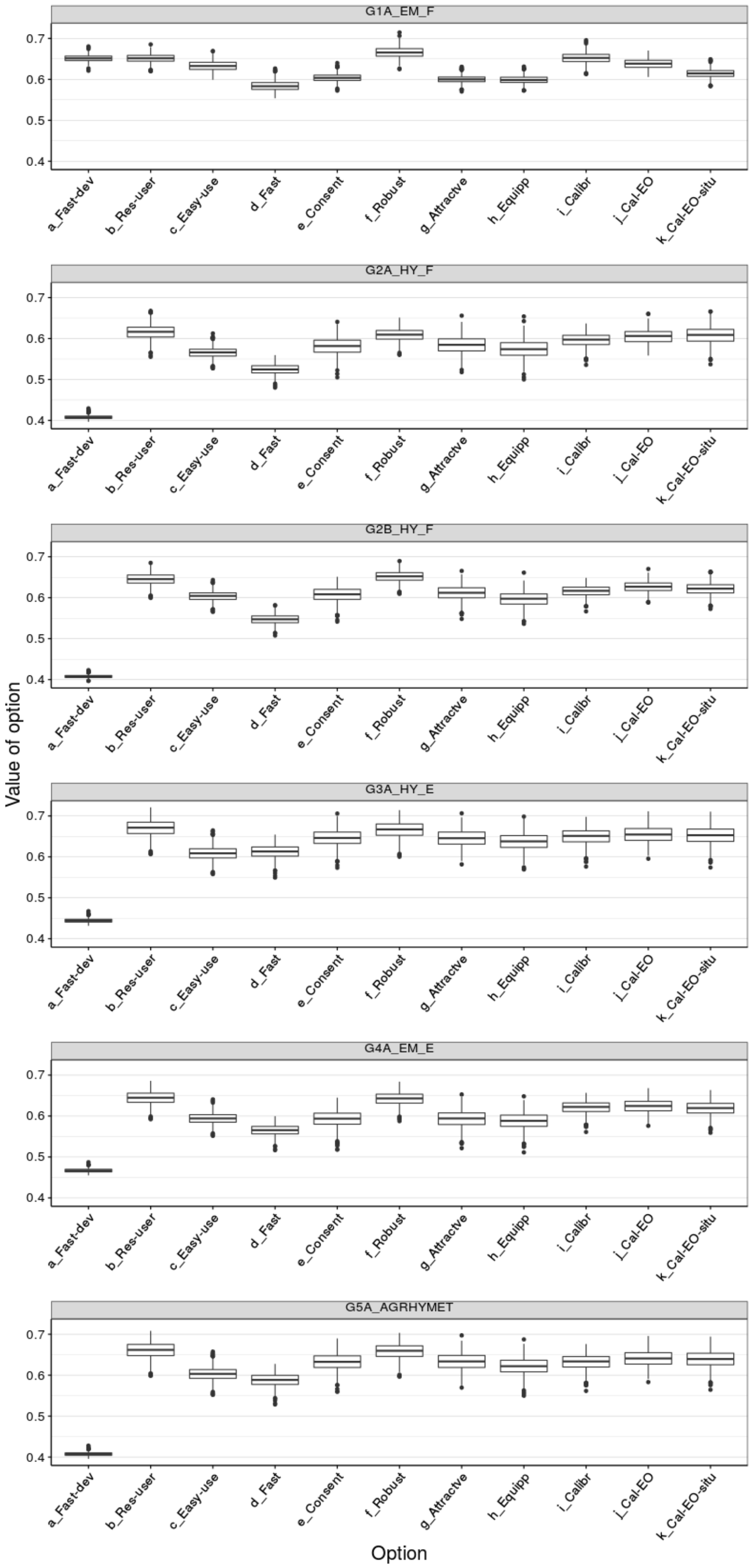


Figure S35. MCDA results with uncertainty: Median overall (total) value (y-axis) calculated by the MCDA for each configuration (x-axis) and each stakeholder (boxes). Uncertainty of the predictions is included with 1'000 Monte Carlo simulation runs of the MCDA calculation for each stakeholder and each configuration. The values (from 0 to 1) are displayed on the y-axis, where 0 means that none of the objectives are achieved and 1 that all objectives are fully achieved. The boxplots show the 0.25 (lower), 0.5 (median), and 0.75 (upper) quartiles of the values as result of the Monte Carlo simulation.

2.8 MCDA results: sensitivity analyses

Table S35: Results of the sensitivity analyses. For each setting (column S0–S31), the mean total value and mean rank resulting from the 1'000 Monte Carlo simulation runs is provided for each stakeholder group (G1A_EM_F, etc.) and configuration (a_Fast-dev, etc.). Yellow highlight: ranking of our standard analysis (S0) compared with additive aggregation model (S11) for best-performing configurations across stakeholder groups *b_Res-user* and *f_Robust*; green highlight: rankings for the also well-performing configuration *i_Calibr*.

			S0	S11	S12	S13	S14	S21	S22				S231	S232	S31
									11_min	11_max	12_min	12_max			
G1A_EM_F	value	a_Fast-dev	0.65	0.69	0.64	0.67	0.68	0.49	0.66	-	-	0.65	-	-	0.65
	rank		3.13	1.88	3.71	2.38	2.03	11.00	2.83	-	-	3.58	-	-	3.51
	value	b_Res-user	0.65	0.67	0.65	0.66	0.67	0.68	0.66	-	-	0.66	-	-	0.65
	rank		3.16	3.05	3.18	3.19	3.13	1.65	3.28	-	-	2.58	-	-	3.04
	value	c_Easy-use	0.63	0.65	0.63	0.64	0.65	0.61	0.64	-	-	0.64	-	-	0.64
	rank		5.41	5.21	5.39	5.40	5.27	8.19	5.20	-	-	4.77	-	-	5.28
	value	d_Fast	0.58	0.59	0.58	0.59	0.59	0.61	0.59	-	-	0.59	-	-	0.58
	rank		10.56	10.94	10.23	10.80	10.90	8.38	10.36	-	-	10.75	-	-	10.72
	value	e_Consent	0.60	0.63	0.60	0.61	0.62	0.64	0.60	-	-	0.61	-	-	0.61
	rank		8.58	8.12	8.77	8.40	8.24	4.83	8.63	-	-	8.27	-	-	8.55
	value	f_Robust	0.67	0.69	0.66	0.68	0.68	0.66	0.67	-	-	0.68	-	-	0.67
	rank		1.73	1.68	1.69	1.68	1.67	3.33	1.53	-	-	1.34	-	-	1.61
	value	g_Attrac-tve	0.60	0.63	0.59	0.61	0.62	0.63	0.60	-	-	0.61	-	-	0.60
	rank		9.05	8.53	9.20	8.81	8.65	5.90	9.03	-	-	8.71	-	-	8.92
	value	h_Equipp	0.60	0.63	0.59	0.61	0.62	0.63	0.60	-	-	0.61	-	-	0.60
	rank		9.19	8.70	9.33	8.95	8.80	6.66	9.29	-	-	8.94	-	-	9.15
	value	i_Calibr	0.65	0.66	0.65	0.66	0.66	0.63	0.66	-	-	0.65	-	-	0.65
	rank		3.08	4.08	2.71	3.52	3.89	5.93	3.27	-	-	3.65	-	-	3.02
	value	j_Cal-EO	0.64	0.65	0.64	0.64	0.65	0.64	0.64	-	-	0.64	-	-	0.64
	rank		4.77	5.58	4.56	5.15	5.40	4.53	5.05	-	-	5.32	-	-	4.83
	value	k_Cal-EO-situ	0.62	0.63	0.61	0.62	0.62	0.63	0.61	-	-	0.62	-	-	0.62
	rank		7.35	8.24	7.23	7.73	8.04	5.60	7.53	-	-	8.09	-	-	7.37
G2A_HY_F	value	a_Fast-dev	0.41	0.50	0.38	0.45	0.48	-	-	-	-	-	-	-	0.39
	rank		11.00	11.00	11.00	11.00	11.00	-	-	-	-	-	-	-	11.00
	value	b_Res-user	0.62	0.67	0.60	0.64	0.66	-	-	-	-	-	-	-	0.64
	rank		2.66	2.94	2.83	2.79	2.90	-	-	-	-	-	-	-	2.21
	value	c_Easy-use	0.57	0.61	0.55	0.58	0.60	-	-	-	-	-	-	-	0.59
	rank		7.93	8.50	7.10	8.56	8.56	-	-	-	-	-	-	-	7.79
	value	d_Fast	0.53	0.55	0.52	0.54	0.55	-	-	-	-	-	-	-	0.49
	rank		9.94	10.00	9.42	10.00	10.00	-	-	-	-	-	-	-	10.00
	value	e_Consent	0.58	0.66	0.54	0.62	0.65	-	-	-	-	-	-	-	0.60
	rank		6.29	3.70	7.48	4.83	4.00	-	-	-	-	-	-	-	6.20
	value	f_Robust	0.61	0.66	0.59	0.63	0.65	-	-	-	-	-	-	-	0.63
	rank		3.26	3.39	3.34	3.30	3.40	-	-	-	-	-	-	-	2.69
	value	g_Attrac-tve	0.58	0.66	0.55	0.62	0.65	-	-	-	-	-	-	-	0.61
	rank		6.03	3.21	7.25	4.36	3.47	-	-	-	-	-	-	-	5.57
	value	h_Equipp	0.57	0.65	0.54	0.61	0.64	-	-	-	-	-	-	-	0.60
	rank		6.94	4.27	7.90	5.51	4.63	-	-	-	-	-	-	-	6.60
	value	i_Calibr	0.60	0.63	0.59	0.61	0.62	-	-	-	-	-	-	-	0.61
	rank		4.72	7.16	3.77	6.16	6.91	-	-	-	-	-	-	-	5.49
	value	j_Cal-EO	0.61	0.64	0.60	0.62	0.63	-	-	-	-	-	-	-	0.62

			S0	S11	S12	S13	S14	S21	S22				S231	S232	S31
									11_min	11_max	12_min	12_max			
	rank		3.73	6.12	2.87	4.89	5.76	-	-	-	-	-	-	-	4.28
	value	k_Cal-EO-situ	0.61	0.64	0.60	0.62	0.63	-	-	-	-	-	-	-	0.62
	rank		3.51	5.72	3.04	4.61	5.37	-	-	-	-	-	-	-	4.19
G2B_HY_F	value	a_Fast-dev	0.41	0.51	0.38	0.45	0.49	-	0.41	0.41	0.40	0.42	-	-	0.39
	rank		11.00	11.00	11.00	11.00	11.00	-	11.00	11.00	11.00	11.00	-	-	11.00
	value	b_Res-user	0.65	0.70	0.63	0.67	0.69	-	0.65	0.64	0.63	0.66	-	-	0.66
	rank		2.15	2.65	2.41	2.27	2.53	-	2.18	2.31	2.61	1.96	-	-	1.95
	value	c_Easy-use	0.60	0.65	0.59	0.62	0.64	-	0.61	0.60	0.59	0.62	-	-	0.62
	rank		7.03	7.61	6.25	7.82	7.77	-	6.58	7.26	7.22	6.25	-	-	6.67
	value	d_Fast	0.55	0.58	0.54	0.56	0.57	-	0.55	0.55	0.54	0.55	-	-	0.52
	rank		9.99	10.00	9.50	10.00	10.00	-	9.98	9.98	9.94	10.00	-	-	10.00
	value	e_Consent	0.61	0.69	0.57	0.65	0.68	-	0.61	0.61	0.59	0.62	-	-	0.62
	rank		6.43	3.57	7.70	4.59	3.77	-	6.71	6.39	6.93	5.98	-	-	6.30
	value	f_Robust	0.65	0.71	0.63	0.68	0.70	-	0.66	0.65	0.64	0.67	-	-	0.67
	rank		1.62	1.91	1.72	1.65	1.82	-	1.37	1.71	1.76	1.40	-	-	1.47
	value	g_Attrac-tve	0.61	0.70	0.57	0.65	0.68	-	0.61	0.61	0.60	0.63	-	-	0.63
	rank		6.00	2.97	7.38	3.93	3.18	-	6.15	5.89	6.49	5.41	-	-	5.41
	value	h_Equipp	0.60	0.68	0.56	0.64	0.67	-	0.60	0.60	0.58	0.61	-	-	0.62
	rank		7.57	4.63	8.29	6.01	4.97	-	7.67	7.37	7.79	7.06	-	-	7.19
	value	i_Calibr	0.62	0.65	0.61	0.63	0.64	-	0.62	0.61	0.61	0.62	-	-	0.63
	rank		5.43	7.90	4.40	7.08	7.72	-	5.35	5.53	4.73	6.38	-	-	6.12
	value	j_Cal-EO	0.63	0.66	0.62	0.64	0.65	-	0.63	0.62	0.63	0.63	-	-	0.64
	rank		4.05	6.70	3.20	5.46	6.40	-	4.03	4.10	3.44	4.93	-	-	4.49
	value	k_Cal-EO-situ	0.62	0.66	0.61	0.64	0.65	-	0.62	0.62	0.62	0.62	-	-	0.63
	rank		4.74	7.07	4.16	6.19	6.86	-	4.98	4.47	4.10	5.64	-	-	5.42
G3A_HY_E	value	a_Fast-dev	0.44	0.53	0.42	0.48	0.52	-	-	-	-	-	0.45	0.44	0.44
	rank		11.00	11.00	11.00	11.00	11.00	-	-	-	-	-	11.00	11.00	11.00
	value	b_Res-user	0.67	0.71	0.66	0.69	0.70	-	-	-	-	-	0.66	0.68	0.67
	rank		2.77	2.98	2.67	2.85	2.94	-	-	-	-	-	2.64	2.84	2.63
	value	c_Easy-use	0.61	0.64	0.60	0.62	0.63	-	-	-	-	-	0.60	0.61	0.62
	rank		9.09	9.11	8.82	9.18	9.15	-	-	-	-	-	8.61	9.17	8.76
	value	d_Fast	0.61	0.64	0.61	0.62	0.63	-	-	-	-	-	0.60	0.62	0.60
	rank		8.88	9.27	8.28	9.14	9.24	-	-	-	-	-	9.20	8.67	9.50
	value	e_Consent	0.65	0.70	0.62	0.67	0.69	-	-	-	-	-	0.63	0.65	0.65
	rank		5.31	3.93	6.38	4.53	4.11	-	-	-	-	-	6.04	5.16	5.40
	value	f_Robust	0.67	0.70	0.66	0.68	0.70	-	-	-	-	-	0.66	0.67	0.67
	rank		3.12	3.13	3.02	3.13	3.16	-	-	-	-	-	2.65	3.14	2.97
	value	g_Attrac-tve	0.65	0.69	0.62	0.67	0.69	-	-	-	-	-	0.63	0.65	0.65
	rank		5.47	4.04	6.58	4.73	4.24	-	-	-	-	-	5.98	5.29	5.34
	value	h_Equipp	0.64	0.69	0.62	0.66	0.68	-	-	-	-	-	0.62	0.65	0.64
	rank		6.30	4.68	7.16	5.40	4.88	-	-	-	-	-	6.79	6.03	6.17
	value	i_Calibr	0.65	0.67	0.64	0.66	0.67	-	-	-	-	-	0.64	0.65	0.65
	rank		4.98	6.25	4.29	5.66	6.06	-	-	-	-	-	4.52	5.26	5.06
	value	j_Cal-EO	0.65	0.68	0.65	0.67	0.67	-	-	-	-	-	0.65	0.66	0.66
	rank		4.46	5.71	3.63	5.05	5.49	-	-	-	-	-	4.01	4.60	4.41
	value	k_Cal-EO-	0.65	0.68	0.64	0.66	0.67	-	-	-	-	-	0.64	0.66	0.65

			S0	S11	S12	S13	S14	S21	S22				S231	S232	S31
									11_min	11_max	12_min	12_max			
	rank	situ	4.63	5.90	4.17	5.36	5.74	-	-	-	-	-	4.56	4.85	4.76
G4A_EM_E	value	a_Fast-	0.47	0.55	0.44	0.50	0.54	-	-	-	-	-	0.46	-	0.45
	rank	dev	11.00	11.00	11.00	11.00	11.00	-	-	-	-	-	11.00	-	11.00
	value	b_Res-	0.64	0.70	0.62	0.67	0.69	-	-	-	-	-	0.64	-	0.66
	rank	user	2.02	2.20	2.33	1.94	2.10	-	-	-	-	-	2.13	-	1.78
	value	c_Easy-	0.59	0.65	0.58	0.62	0.64	-	-	-	-	-	0.59	-	0.61
	rank	use	7.09	8.11	6.37	8.09	8.18	-	-	-	-	-	6.83	-	6.83
	value	d_Fast	0.57	0.60	0.56	0.58	0.59	-	-	-	-	-	0.56	-	0.53
	rank		9.52	9.99	8.13	9.97	9.99	-	-	-	-	-	9.56	-	10.00
	value	e_Consent	0.59	0.69	0.55	0.64	0.67	-	-	-	-	-	0.58	-	0.61
	rank		7.09	3.99	8.30	5.37	4.32	-	-	-	-	-	7.45	-	6.95
	value	f_Robust	0.64	0.70	0.62	0.67	0.69	-	-	-	-	-	0.64	-	0.66
	rank		2.12	2.24	2.38	2.04	2.18	-	-	-	-	-	1.99	-	1.82
	value	g_Attrac-	0.59	0.69	0.55	0.64	0.67	-	-	-	-	-	0.59	-	0.61
	rank	tve	7.10	3.84	8.25	5.27	4.20	-	-	-	-	-	7.19	-	6.54
	value	h_Equipp	0.59	0.68	0.54	0.64	0.67	-	-	-	-	-	0.58	-	0.61
	rank		7.59	4.35	8.55	5.88	4.71	-	-	-	-	-	7.79	-	7.16
	value	i_Calibr	0.62	0.66	0.61	0.64	0.65	-	-	-	-	-	0.62	-	0.63
	rank		4.18	6.99	3.51	5.59	6.67	-	-	-	-	-	3.96	-	4.71
	value	j_Cal-EO	0.62	0.66	0.61	0.64	0.65	-	-	-	-	-	0.62	-	0.64
	rank		3.85	6.50	3.17	5.12	6.12	-	-	-	-	-	3.68	-	4.27
	value	k_Cal-EO-	0.62	0.66	0.60	0.64	0.65	-	-	-	-	-	0.61	-	0.63
	rank	situ	4.43	6.80	4.02	5.73	6.54	-	-	-	-	-	4.44	-	4.94
G5A_AGRHYMET	value	a_Fast-	0.41	0.50	0.38	0.45	0.48	-	-	-	-	-	0.42	-	0.40
	rank	dev	11.00	11.00	11.00	11.00	11.00	-	-	-	-	-	11.00	-	11.00
	value	b_Res-	0.66	0.70	0.65	0.68	0.70	-	-	-	-	-	0.65	-	0.67
	rank	user	2.49	2.88	2.45	2.65	2.83	-	-	-	-	-	2.40	-	2.30
	value	c_Easy-	0.60	0.64	0.59	0.62	0.63	-	-	-	-	-	0.60	-	0.61
	rank	use	8.44	8.49	7.92	8.62	8.54	-	-	-	-	-	7.99	-	8.08
	value	d_Fast	0.59	0.61	0.58	0.60	0.61	-	-	-	-	-	0.58	-	0.57
	rank		9.52	9.81	8.82	9.74	9.80	-	-	-	-	-	9.62	-	9.97
	value	e_Consent	0.63	0.70	0.60	0.66	0.69	-	-	-	-	-	0.62	-	0.64
	rank		5.40	3.56	6.81	4.27	3.73	-	-	-	-	-	6.20	-	5.45
	value	f_Robust	0.66	0.70	0.64	0.68	0.69	-	-	-	-	-	0.65	-	0.67
	rank		2.66	2.89	2.57	2.75	2.86	-	-	-	-	-	2.30	-	2.46
	value	g_Attrac-	0.63	0.70	0.60	0.66	0.69	-	-	-	-	-	0.62	-	0.64
	rank	tve	5.36	3.44	6.74	4.20	3.65	-	-	-	-	-	5.95	-	5.16
	value	h_Equipp	0.62	0.69	0.59	0.65	0.68	-	-	-	-	-	0.61	-	0.63
	rank		6.50	4.36	7.61	5.30	4.63	-	-	-	-	-	7.00	-	6.29
	value	i_Calibr	0.63	0.66	0.62	0.64	0.65	-	-	-	-	-	0.63	-	0.64
	rank		5.49	7.08	4.57	6.45	6.91	-	-	-	-	-	4.92	-	5.71
	value	j_Cal-EO	0.64	0.67	0.63	0.65	0.66	-	-	-	-	-	0.63	-	0.65
	rank		4.51	6.21	3.52	5.38	5.96	-	-	-	-	-	4.04	-	4.60
	value	k_Cal-EO-	0.64	0.67	0.63	0.65	0.66	-	-	-	-	-	0.63	-	0.64
	rank	situ	4.65	6.29	4.00	5.64	6.10	-	-	-	-	-	4.58	-	4.98

Table S36: Overview of Kendall's τ rank correlation coefficient between ranks of configurations in the main MCDA (setting S0) and different settings for sensitivity analyses (S11 to S31), averaged (Aver) over all involved stakeholder groups (SH; e.g., group G1A).

		S11 (all SH)	S12 (all SH)	S13 (all SH)
Kendall's τ	Aver	0.667	0.903	0.770
		S14 (all SH)	S21 (G1A)	S22 (G1A, G2A, G2B)
Kendall's τ	Aver	0.685	0.309	0.955
		S231 (G3A, G4A, G5A)	S232 (G3A, G4A, G5A)	S31
Kendall's τ	Aver	0.891	0.964	0.952

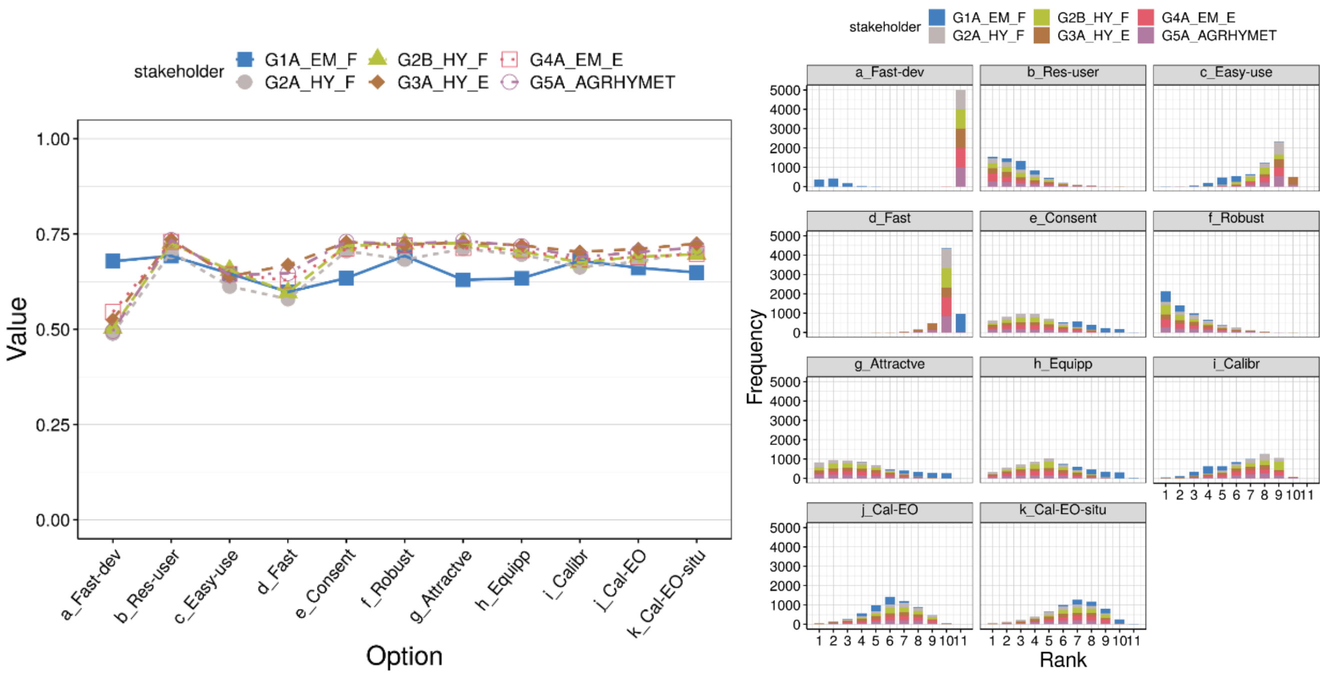


Figure S36: Results of the sensitivity analysis setting S11, additive model for all six stakeholder groups, γ = 1.

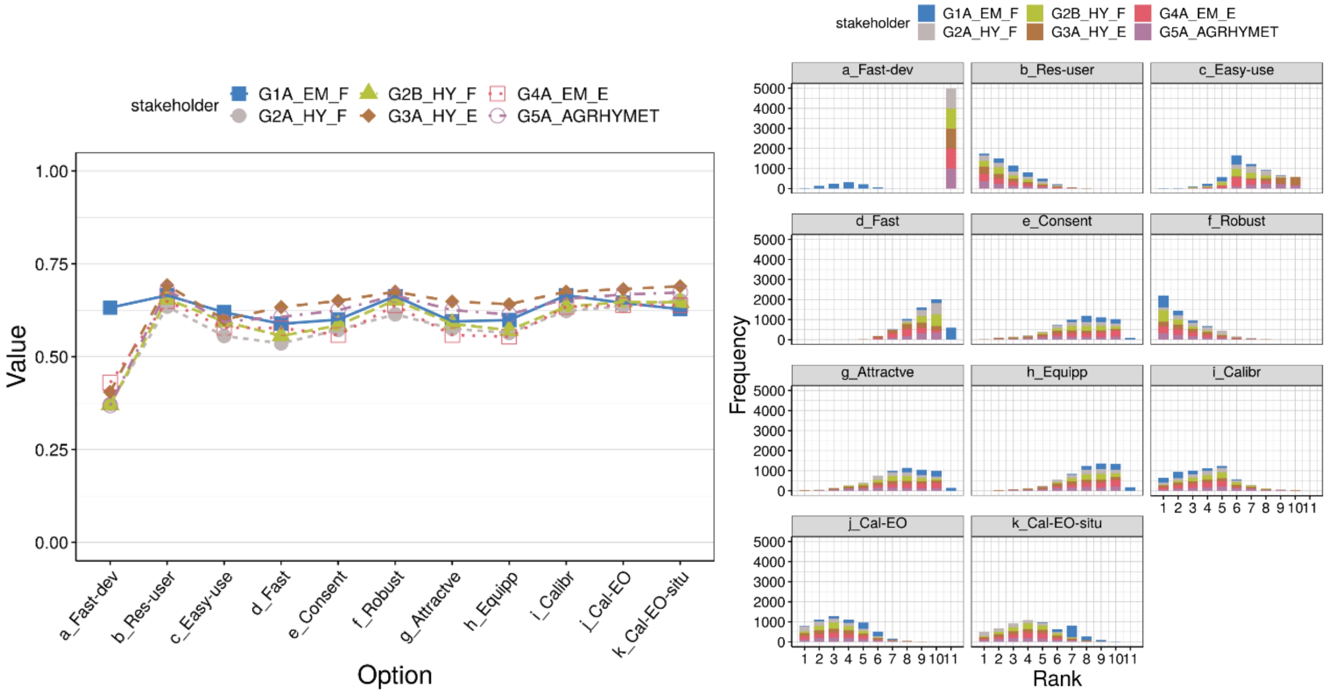


Figure S37: Results of the sensitivity analysis setting S12, weighted geometric mean for all six stakeholder groups, γ → 0.

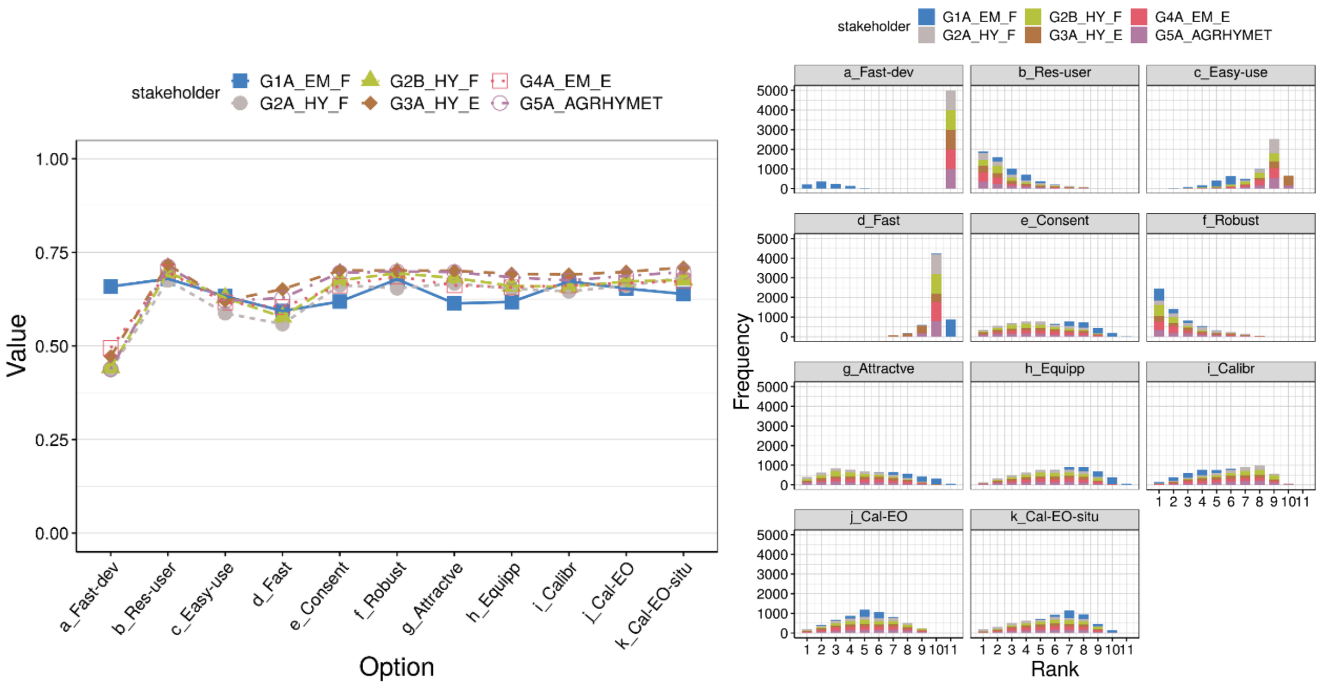


Figure S38: Results of the sensitivity analysis setting S13, mixture model for all six stakeholder groups, γ = 0.5

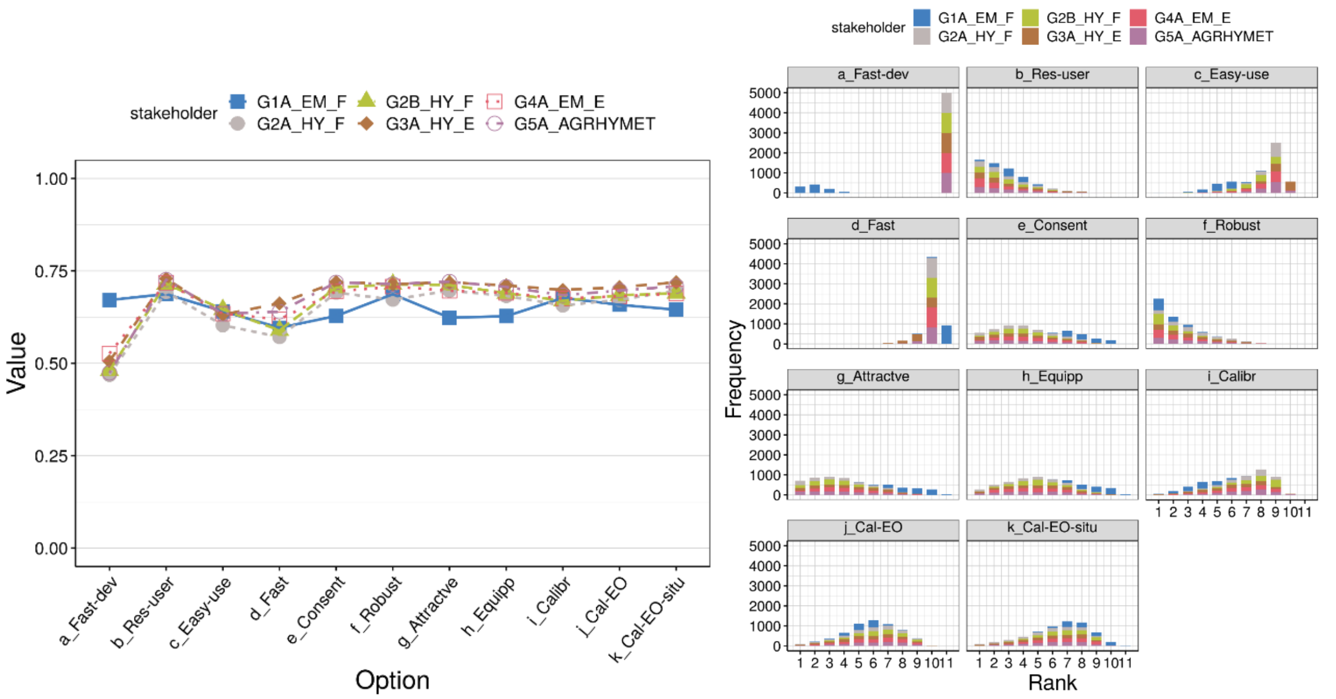


Figure S39: Results of sensitivity analysis setting S14, weighted power mean close to additive for all six stakeholder groups, γ = 0.8.

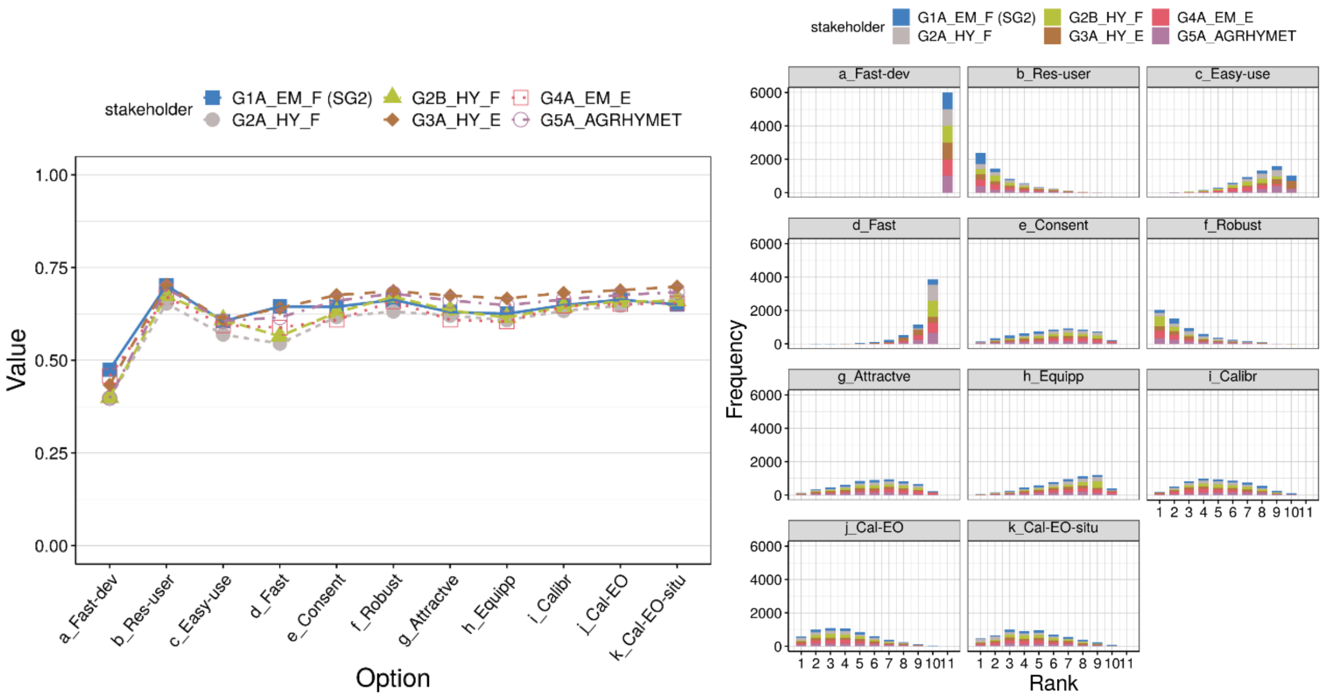


Figure S40: Results of the sensitivity analysis setting S21, alternative weight set for group G1A, resulting from consistency check.

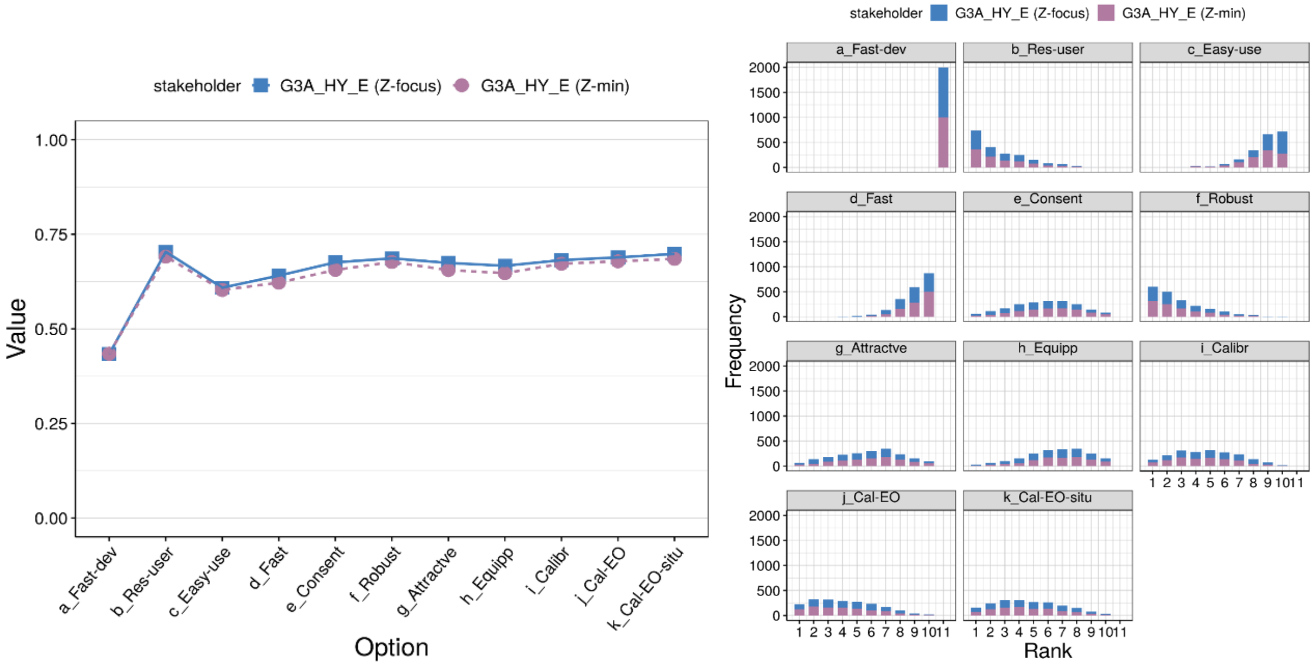


Figure S41: Results of the sensitivity analysis setting S231 for group G3A_HY_E, resulting from ranges assigned to Z min.

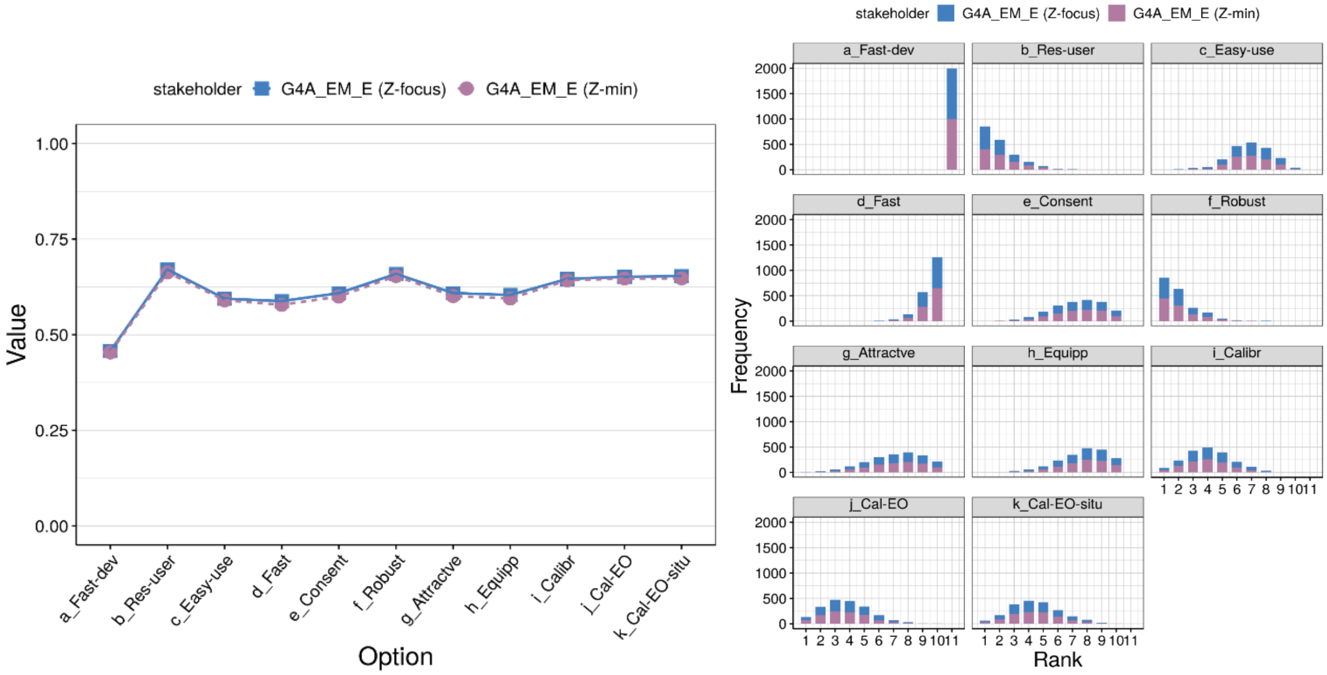


Figure S42: Results of the sensitivity analysis setting S231 for group G4A_EM_E, resulting from ranges assigned to Z min.

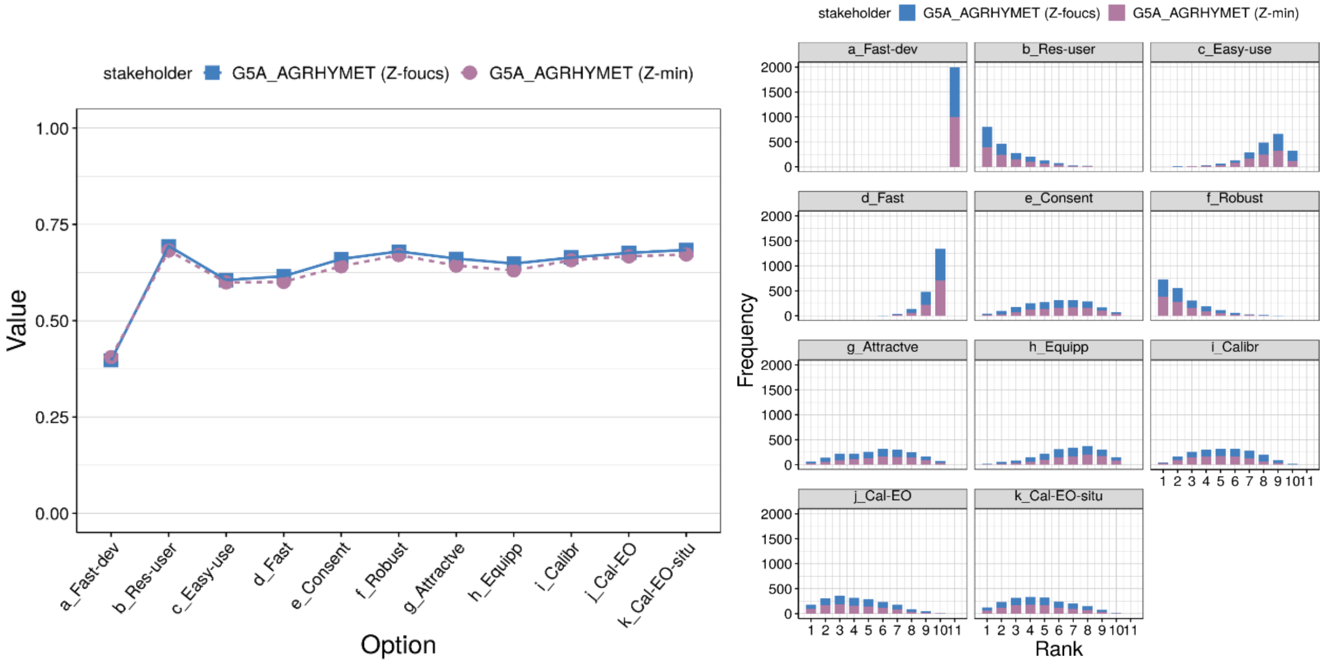


Figure S43: Results of the sensitivity analysis setting S231 for group G5A_AGRHYMET, resulting from ranges assigned to Z min.

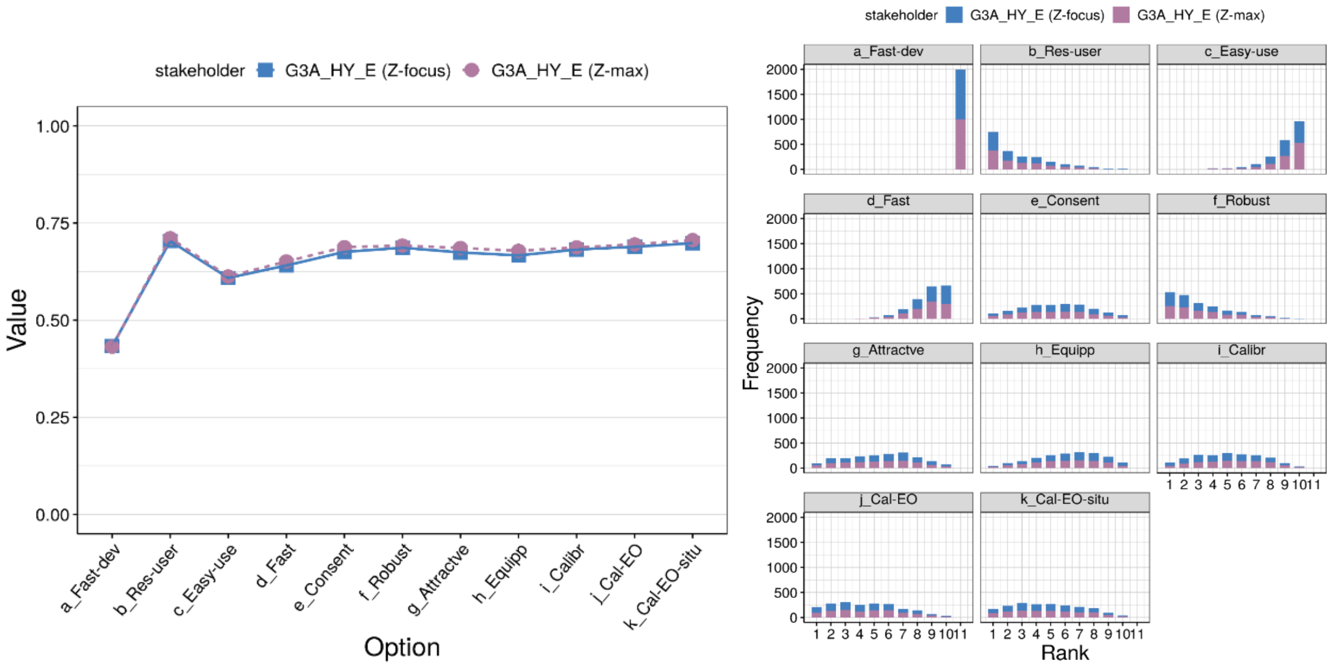


Figure S44: Results of the sensitivity analysis setting S232 for group G3A_HY_E, resulting from ranges assigned to Z max.

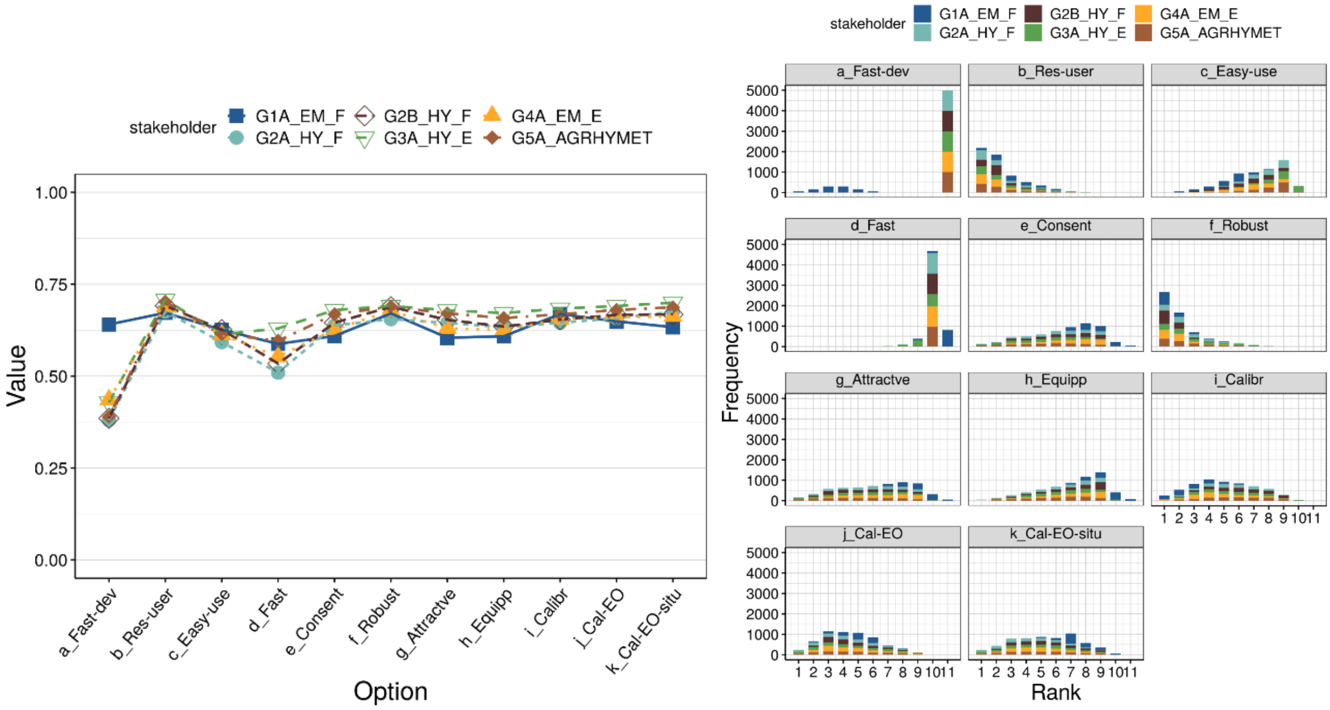


Figure S45: Results of the sensitivity analysis setting S31, doubling weight of "Several languages" for all six groups.

2.9 MCDA results: cost-benefit visualizations

A cost-benefit visualization using MCDA allows to check, whether configurations are outperformed by other configurations, given the underlying conditions of the MCDA model used. The ValueDecisions app easily allows to visualize such cost-benefit visualizations on chosen attributes with one mouse click. We illustrate three such analyses here. First, we carried out a "classic" cost-benefit visualization, where the total operation and maintenance costs (i.e., attribute *32_costs*) of each FANFAR system configuration were plotted against all other "benefits". In our case, the benefits were the total aggregated value that an configuration achieved for the other nine objectives when aggregating over all nine objectives, using our standard MCDA model (and in including the stakeholder preferences; Figure S46). The more to the left the configuration is situated, the more expensive it is (achieving a low value). The configurations with highest performance on both the x- and y-axis were then linearly connected. This interpolation visualizes a possible efficient frontier for the given configurations, and any configuration below this line is outperformed by one or more other configurations in this set. Note that we cannot show the estimated costs in € as normally done for confidentiality reasons. Obviously, the cheapest configuration (to the far right) was *a_Fast-dev*, the status quo configuration at the beginning of the FANFAR project. Not surprisingly, it also achieved the lowest value of all configurations for the other nine objectives.

Potentially interesting configurations are those situated on the efficient frontier (solid black line), and strictly speaking, configurations below the efficient frontier should not be considered. To give some guidance: FANFAR system configuration *d_Fast* (green dot) should definitely not be chosen. It was similarly expensive for operation and maintenance as *c_Easy-use*, *j_Cal-EO*, and *f_Robust*, which all achieved higher values for the other nine objectives, for all stakeholder groups. Comparing these four configurations, the best choice is *f_Robust*, which was located on the efficient frontier for all stakeholder groups. Configuration *i_Calibr* could also be a good choice, it was somewhat cheaper, but also achieved a high value on the other objectives for all stakeholder groups. Of the most expensive configurations in terms of operation and maintenance costs were *b_Res-user*, *e_Consent*, *g_Attractve*, *h_Equipp*, and *k-Cal-EO-situ*. For these, it is difficult to decide which is best, because of differences between stakeholder groups. For instance, *g_Attractve* achieved high values on the nine other objectives for stakeholder groups G2A, G2B, G3A, and G5A, but was outperformed by configuration *b_Res-user* and *f_Robust* for group G4A, and clearly performed poorly for group G1A. For this group of the French speaking emergency managers, G1A, *b_Res-user* would be the best performing of the expensive configurations.

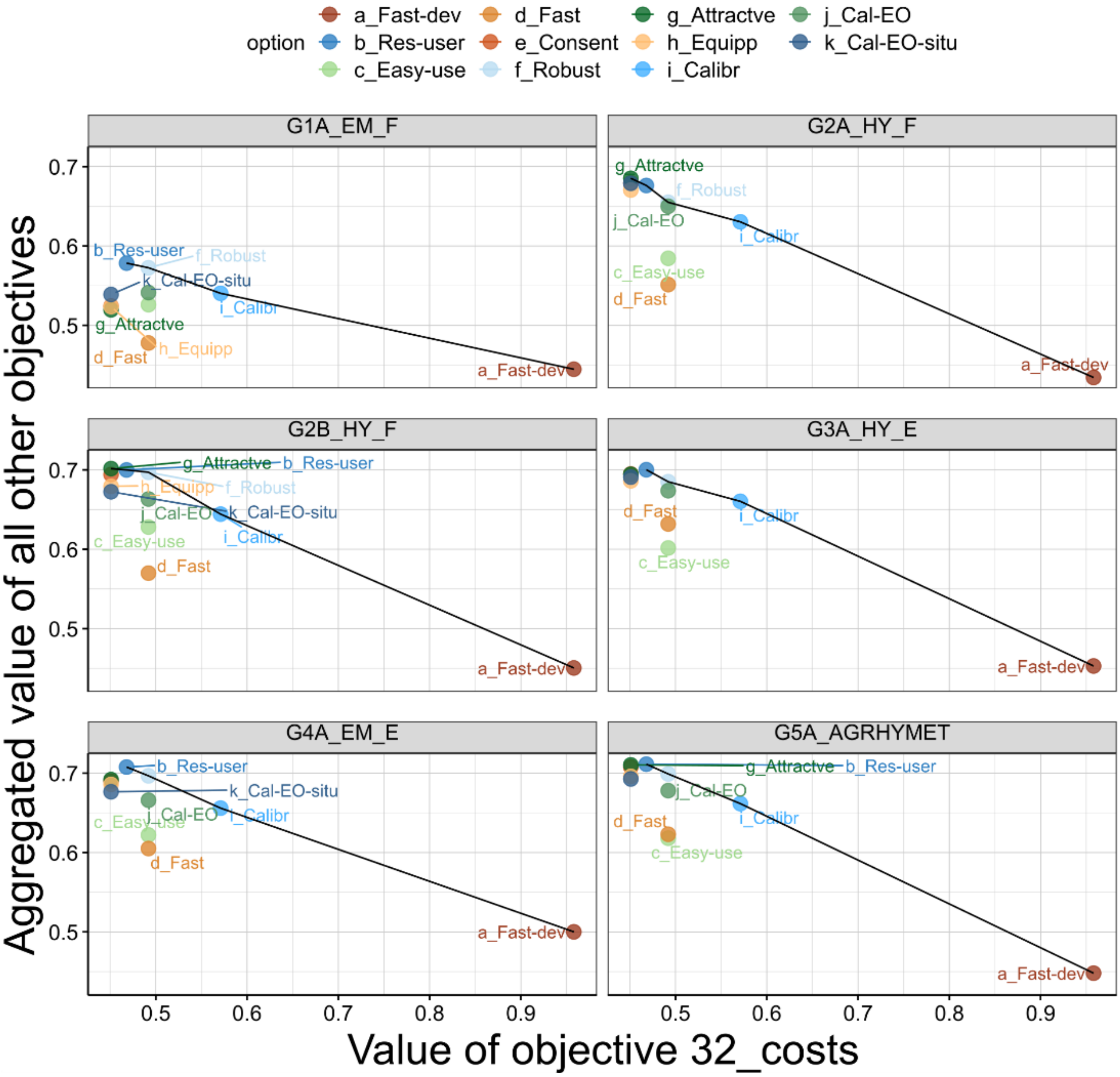


Figure S46. Cost-benefit visualization of 11 FANFAR system configurations (colored dots) for six stakeholder groups (blocks). The value of the total operation and maintenance costs (i.e., attribute *32_costs*; x-axis) is plotted against the total aggregated value of all other nine objectives (y-axis). For both axes, values range from 0 (worst case) to 1 (best). This means that the most expensive configurations are to the left (i.e., achieving low values), and the cheaper configurations to the right (achieving high values). The efficient

frontier (solid black line) visualizes the best performing configurations for a given cost level. Configurations situated below the efficient frontier are outperformed by better configurations, which achieve a higher performance on the other nine objectives for same level of operation and maintenance costs.

We repeated the cost-benefit visualization for both lower level objectives of the objective *3 Low costs*, i.e., *3.1 Short development time*, and *3.2 Low operation and maintenance costs* (Figure S47) as costs, plotted against the total aggregated value of all other eight (respectively seven) objectives. Interestingly, configuration *d_Fast* now appeared on the efficient frontier for four stakeholder groups, while it had been outperformed by other configurations in the previous analysis. Reason is that *d-Fast* is not only the fastest system for producing and distributing forecasts, but because it has no fancy features, it is also the fastest configuration in terms of development time (apart from the status quo configuration *a_Fast-dev*). Thus, *d-Fast* performed quite well for those stakeholder groups who had given a relatively higher weight to the objective *31_develop_time*, namely groups G2A, G2B, G4A, and G5A. Previously recommendable configurations remained on the efficient frontier for all stakeholder groups: *b_Res-user*, *f_Robust*, and *i_Calibr*.

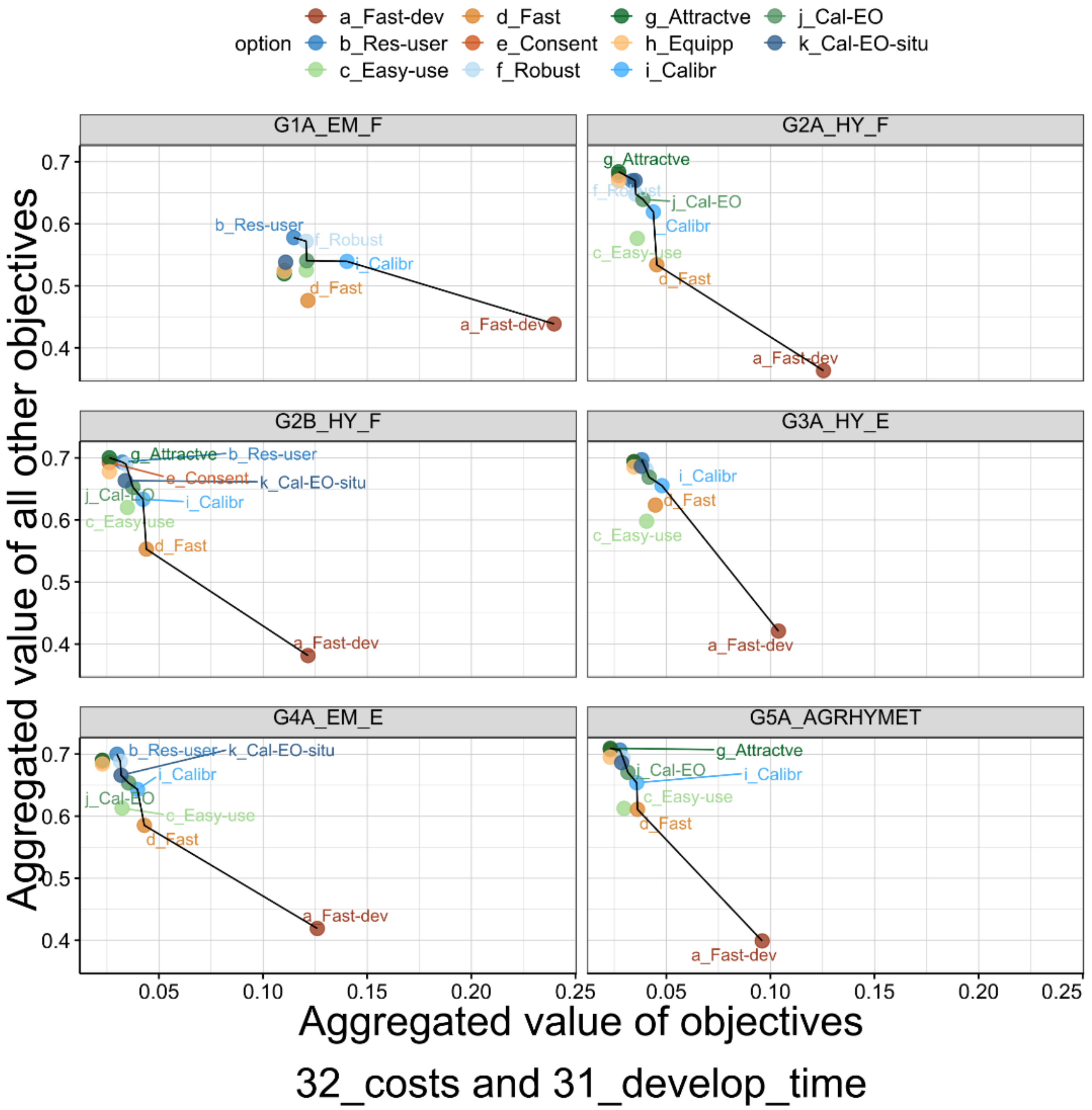


Figure S47. Cost-benefit visualization of 11 FANFAR system configurations (colored dots) for six stakeholder groups (blocks). The costs (x-axis) are the aggregated value of the total operation and maintenance costs (i.e., attribute *32_costs*) and the time needed to develop the FANFAR system (*31_develop_time*). Costs are plotted against the total aggregated value of all other eight objectives (y-axis). Solid black line: efficient frontier.

In the third cost-benefit visualization, we used the two previous cost attributes and included another attribute that concerns costs, namely whether long-term financing is secured (*41_sust_financing*; Figure S48). Thus, the "costs" displayed on the x-axis are the value achieved when aggregating over all three objectives, using our standard MCDA model (and including the stakeholder preferences). Of the previously recommendable configurations *b_Res-user*, and *f_Robust* are still (nearly) on the efficient frontier for all stakeholder groups. However, for groups G2A, G2B, and G5A, *i_Calibr* were now outperformed. Configuration *g_Attractve* performed well for five stakeholder groups, but – again – not for the French speaking emergency managers, group G1A. This can be explained in that they had given zero weight, i.e., zero importance to *41_sust_financing*. Thus, if we are looking for a consensus configuration over all groups, *g_Attractve* should not be chosen.

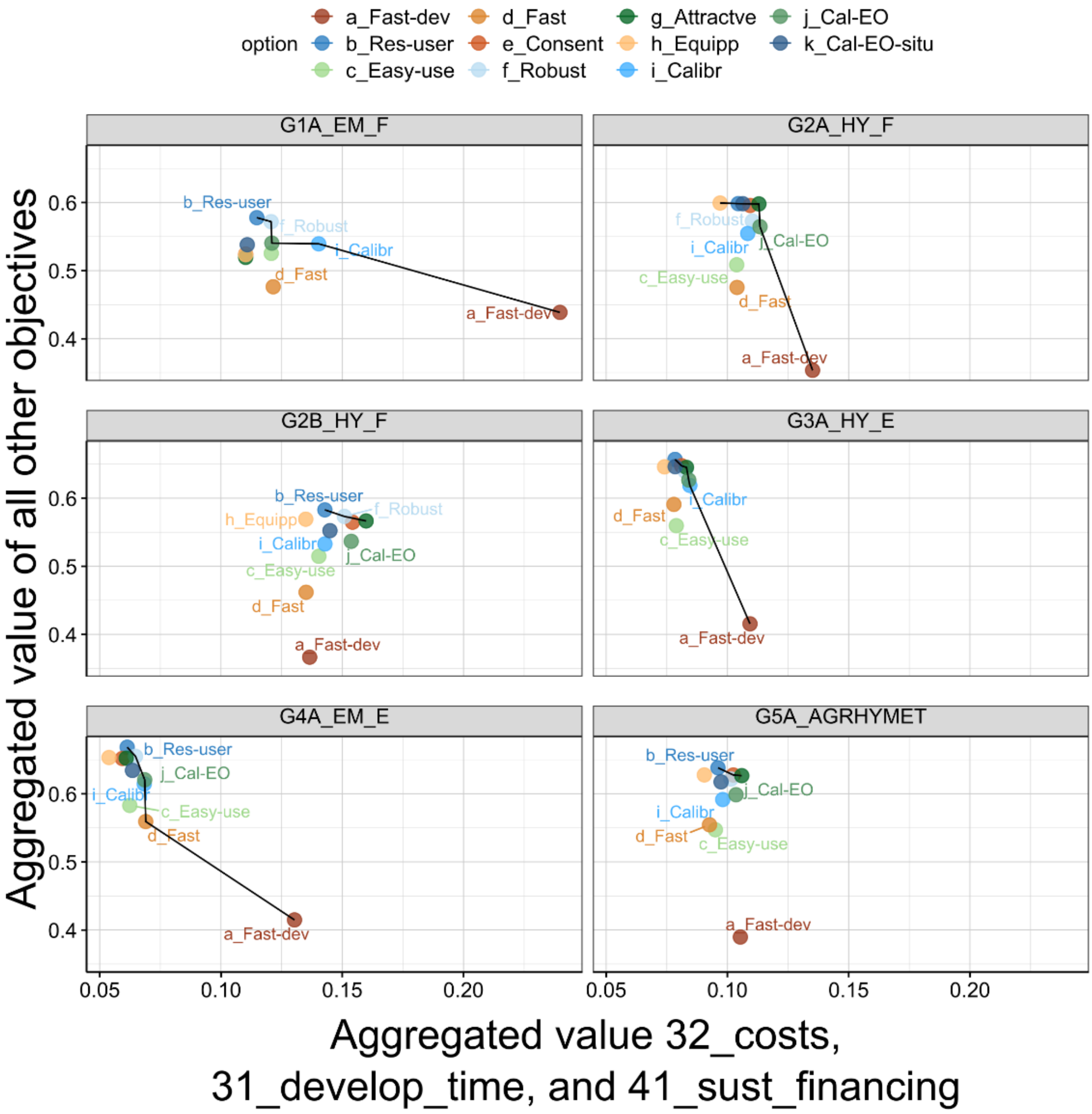


Figure S48. Cost-benefit visualization of 11 FANFAR system configurations (colored dots) for six stakeholder groups (blocks). The costs (x-axis) are the aggregated value of the total operation and maintenance costs (i.e., attribute *32_costs*), the time needed to develop the FANFAR system (*31_develop_time*), and secured long-term financing (*41_sust_financing*). Costs are plotted against the total aggregated value of all other eight objectives (y-axis). Solid black line: efficient frontier.

2.10 Results: stakeholders' perceived satisfaction with FANFAR system

12 workshop participants filled out the survey, resulting for our 10 objectives in 10 x 12 = 120 responses to each question (Figure S49). Both from the directly asked question, whether respondents would use the FANFAR system in future if it remains as is (b), as well as the inferred difference between the acceptable level (c) and the perceived current level (a) of each objective (c – a), the majority of respondents seemed to perceive FANFAR's current performance as sufficient for all objectives. Across all objectives, 79 responses were positive, 16 negative, and 25 did not respond to question b (Figure S49). Regarding the difference (c – a), 97 responses indicated higher perceived current performance (a) than the minimum required performance (c), while 23 responses were opposite (Figure S49). However, although this inferred measure of sufficiency of the FANFAR system performance and question b should measure the same opinion, only in four cases did negative responses to b correspond to inferred negative responses. This indicates that respondents were not always consistent in their answers. Differences between performances of our ten objectives were moderate. Objective *42_human_resour* performed best with only two (out of 21, as three NA) combined negative responses to b and (c – a). Objectives *21_reliable_info* and *22_timely_info* performed worst, with each six combined negative responses (out of 22, as two NA).

11_accur_info			
A	B	Suff.	C
5	Yes	1	4
5	Yes	-1	6
4	Yes	-1	5
5	Yes	1	4
5	Yes	-1	6
6	Yes	1	5
5	Yes	2	3
4	Yes	-1	5
6	Yes	5	1
4	Yes	1	3
5	Yes	0	5
5	Yes	2	3
3	Yes	0	3
	N/A	0	
4	Yes	1	3

12_clear_info			
A	B	Suff.	C
5	Yes	0	5
5	Yes	0	5
3	Yes	-1	4
5	Yes	0	5
3	No	1	2
5	Yes	0	5
3	Yes	0	3
5	Yes	2	3
5	Yes	4	1
3	Yes	0	3
4	Yes	-1	5
	N/A	0	
2	No	0	2
	N/A	0	
3	Yes	1	2

21_reliable_info			
A	B	Suff.	C
3	No	-1	4
5	Yes	0	5
2	No	-2	4
3	No	0	3
3	Yes	0	3
5	Yes	0	5
4	Yes	2	2
4	Yes	1	3
5	Yes	4	1
3	Yes	1	2
5	Yes	0	5
	N/A	0	
2	No	0	2
	N/A	0	
4	Yes	2	2

22_timely_info			
A	B	Suff.	C
4	Yes	0	4
5	Yes	0	5
4	Yes	0	4
1	No	-2	3
4	No	1	3
4	Yes	-1	5
3	Yes	-1	4
4	Yes	1	3
5	Yes	4	1
3	Yes	0	3
4	Yes	3	1
	N/A	0	
2	No	0	2
	N/A	0	
3	Yes	0	3

23_language			
A	B	Suff.	C
4	Yes	3	1
5	Yes	2	3
4	Yes	0	4
3	Yes	0	3
1	Yes	-2	3
4	Yes	-1	5
4	Yes	0	4
3	Yes	0	3
4	Yes	3	1
3	Yes	0	3
5	Yes	0	5
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

31_develop_time			
A	B	Suff.	C
4	Yes	0	4
4	Yes	0	4
4	Yes	0	4
2	Yes	0	2
4	No	2	2
4	Yes	0	4
3	Yes	0	3
4	Yes	1	3
3	Yes	2	1
3	Yes	-1	4
4	Yes	0	4
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

32_costs			
A	B	Suff.	C
4	Yes	-1	5
5	Yes	0	5
5	Yes	0	5
3	Yes	1	2
2	No	1	1
3	Yes	-2	5
4	Yes	0	4
4	Yes	0	4
3	Yes	2	1
5	Yes	0	5
4	Yes	-1	5
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

41_sust_financing			
A	B	Suff.	C
5	Yes	0	5
4	Yes	0	4
3	Yes	-2	5
2	Yes	0	2
5	Yes	2	3
1	Yes	-3	4
2	Yes	0	2
1	No	-3	4
4	Yes	3	1
5	Yes	0	5
4	Yes	0	4
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

42_human_resour			
A	B	Suff.	C
5	Yes	0	5
4	Yes	0	4
5	Yes	1	4
3	Yes	2	1
4	Yes	2	2
5	Yes	1	4
4	Yes	0	4
5	Yes	0	5
5	Yes	4	1
4	Yes	-1	5
5	Yes	0	5
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

43_support_syst			
A	B	Suff.	C
5	Yes	0	5
5	Yes	0	5
4	Yes	0	4
4	Yes	2	2
4	Yes	3	1
4	Yes	-1	5
4	Yes	0	4
2	Yes	-3	5
5	Yes	4	1
5	Yes	0	5
3	Yes	0	3
	N/A	0	
2	No	0	2
	N/A	0	
	N/A	0	

Figure S49. Stakeholder perceived satisfaction with the performance of the FANFAR system during the 2020 rainy season. Survey results for each objective and survey respondent (N = 12). A, B and C refer to the three questions asked for each objective. Answers are represented by descending numbers from "best" to "worst" performance. 'Suff.' represents the calculated difference between the answer of C and A. Green colored cells represent a positive response (respondent indicating their perceived performance is sufficient for use of FANFAR), while red indicates the opposite.

3 References

Andersson, J. C. M., Arheimer, B., Traoré, F., Gustafsson, D., and Ali, A.: Process refinements improve a hydrological model concept applied to the Niger River basin, Hydrological Processes, 31, 4540-4554, https://doi.org/10.1002/hyp.11376, 2017.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, Hydrol. Earth Syst. Sci., 24, 535-559, https://doi.org/10.5194/hess-24-535-2020, 2020.

Berg, P., Donnelly, C., and Gustafsson, D.: Near-real-time adjusted reanalysis forcing data for hydrology, Hydrol. Earth Syst. Sci., 22, 989-1000, https://doi.org/10.5194/hess-22-989-2018, 2018.

Berg, P., Almén, F., and Bozhinova, D.: HydroGFD3.0 (Hydrological Global Forcing Data): a 25 km global precipitation and temperature data set updated in near-real time, Earth Syst. Sci. Data, 13, 1531-1545, https://doi.org/10.5194/essd-13-1531-2021, 2021.

Eisenführ, F., Weber, M., and Langer, T.: Rational Decision Making, Springer, Berlin, Heidelberg, New York, 2010.

Figueira, J., and Roy, B.: Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure, European Journal of Operational Research, 139, 317-326, https://doi.org/10.1016/S0377-2217(01)00370-8, 2002.

Gregory, R., Long, G., Colligan, M., Geiger, J. G., and Laser, M.: When experts disagree (and better science won't help much): Using structured deliberations to support endangered species recovery planning, Journal of Environmental Management, 105, https://doi.org/10.1016/j.jenvman.2012.03.001, 2012.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Haag, F., Lienert, J., Schuwirth, N., and Reichert, P.: Identifying non-additive multi-attribute value functions based on uncertain indifference statements, Omega-International Journal of Management Science, 85, 49-67, https://doi.org/10.1016/j.omega.2018.05.011, 2019.

Haag, F., Aubert, A. H., and Lienert, J.: ValueDecisions, a web app to support decisions with conflicting objectives, multiple stakeholders, and uncertainty, Environmental Modelling & Software, 150, 105361, https://doi.org/10.1016/j.envsoft.2022.105361, 2022.

Howard, R. A.: Decision-analysis - Practice and promise, Manage. Sci., 34, 679-695, https://doi.org/10.1287/mnsc.34.6.679, 1988.

Kendall, M. G.: A new measure of rank correlation, Biometrika, 30, 81-93, https://doi.org/10.2307/2332226, 1938.

Litcanu, M., Prostean, O., Oros, C., and Mnerie, A. V.: Brain-writing vs. brainstorming case study for power engineering education, in: Proceedings of 6th World Conference on Educational Sciences, edited by: Uzunboylu, H., Procedia Social and Behavioral Sciences, 387-390, 2015.

Marttunen, M., Belton, V., and Lienert, J.: Are objectives hierarchy related biases observed in practice? A meta-analysis of environmental and energy applications of Multi-Criteria Decision Analysis, European Journal of Operational Research, 265, 178-194, https://doi.org/10.1016/j.ejor.2017.02.038, 2018.

Paulus, P. B., and Yang, H. C.: Idea generation in groups: A basis for creativity in organizations, Organizational Behavior and Human Decision Processes, 82, 76-87, https://doi.org/10.1006/obhd.2000.2888, 2000.

Pictet, J., and Bollinger, D.: Extended use of the cards procedure as a simple elicitation technique for MAVT. Application to public procurement in Switzerland, European Journal of Operational Research, 185, 1300-1307, https://doi.org/10.1016/j.ejor.2006.05.051, 2008.

Silva Pinto, F., and Lienert, J.: Report activities to establish co-design committee, stakeholder analysis, European Union, Horizon 2020, Innovation Action ICT programme, Project: 780118 FANFAR, Persistent URL: https://www.dora.lib4ri.ch/eawag/islandora/object/eawag%3A23886 Deliverable D2.1, 48, 2018.

Zheng, J., Egger, C., and Lienert, J.: A scenario-based MCDA framework for wastewater infrastructure planning under uncertainty, Journal of Environmental Management, 183, Part 3, 895-908, https://doi.org/10.1016/j.jenvman.2016.09.027, 2016.