Expert Elicitation

Research question:
How can expert experience and judgement reduce the uncertainty from the analysis of climate change impacts on water-related climate services?

Expert Elicitation Protocol

Authors

Jens Christian Refsgaard, Eva Sebök, Ernesto Pasten-Zapata, Hans Jørgen Henriksen

Version/date

Draft2/2019.02.18

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

Table of Contents

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

1. Background and purpose of this note

The idea of applying expert elicitation in model assessments was launched at the AQUACLEW kick-off meeting at SMHI, Norrköping, 11-13 October 2017. The basic idea is to use expert elicitation to assess the probability of model ensemble members to be the most reliable ones for projecting climate change and climate change impacts.

The expert elicitation ideas were elaborated during 2018 and discussed at the AQUACLEW General Assembly at GEUS, Copenhagen, 16-17 October 2018. In the present note the feedbacks received during the GA2018 have been incorporated.

The purpose of this note is to describe the methodology and the action plan for performing the expert elicitation in AQUACLEW.

2. Methodology

2.1 What is Expert Elicitation (EE)?

Expert Elicitation (EE) is a well proven technique widely applied in social science as well as in natural science, see e.g. Meyer and Booker (2001), Ye et al. (2008) and Sebok et al. (2016).

Expert elicitation implies using the ideas, brainstorms and 'gut feeling' of experts (people with special knowledge about the subject of the study) to assess probabilities of events, risks or parameter values. Expert elicitation is using the subjective assessment of experts and translating their opinion into probabilistic form following strict protocols. The elicitation is considered successful if the elicited data truly reflects the experts' opinion, not if the answers are correct.

2.2 Basic considerations

2.2.1 Model types

The aim of the elicitation is to assess probabilities of models included in an ensemble of models. AQUACLEW follows the calculation chain from emission scenarios to hydrological impacts, typically including the following types of models:

- General Circulation Models, also sometimes denoted Global Climate Models (GCM)
- Regional Climate Models (RCM)
- Bias correction and downscaling models (BC)
- Hydrological models (HM)

The first two model types (GCM and RCM) deal with the climate signal, the third type (BC) deal with bias correction of climate models, while the fourth type (HM) assesses the impact of the climate signal on hydrological issues. Therefore, we will apply EE to the following two groups of models

- Climate models, i.e. in principle the downscaled and bias corrected climate projection (GCM, RCM and BC). Depending on the decisions made in the case studies this may be based on GCM projections alone or GCM projections dynamically downscaled using an RCM, and it may or may not include bias correction/downscaling the output from the GCM/RCM.
- Hydrological models (HM), i.e. the impact model. This may include one model, or a sequence of coupled models, and it may include an ensemble of (coupled) models.

2.2.2 Case study specific assessments

It is generally recognised that no model is universally valid and that different models have different strengths and weaknesses with respect to types of application and areas for which they are used.

One climate model may for instance perform better than another one for Mediterranean regions, while the opposite may be the case for regions in Northern Europe. Similarly, one climate model may be considered more reliable for simulating average annual precipitation, while another model may be considered better for handling convective and extreme rainfall events.

Therefore, we will make separate assessments for the different case studies, implying that one model may be considered more reliable for one case study than for another case study. Case study assessments will be done in one process by the same experts case by case the same day (lay-out is explained in later section 2.6.2).

2.2.3 Overall approach

EE has been used for a variety of purposes and hence a broad range of procedures have been applied in practice. Our aim is to elicit model probabilities. Our approach has been inspired by Ye et al. (2008) who elicit probabilities of an ensemble of five hydrological (groundwater recharge) models in connection with a study on deposition of nuclear waste.

The EE work includes the six steps described in the following sections.

2.3 Step 1: Define the main issues of elicitation

We want to end up assigning probabilities to models. We do this by asking experts to respond to two series of questions:

- Block 1 – Warm up. Here we want the experts to consider the models based on general questions on the appropriateness of the models in general as well as specifically to deal with the case study issues (including considerations on geographic region, simulation variable, etc.). Block 1 concludes with a request to rank the members of the model ensemble for the specific case studies.
- Block 2 – Probabilities. Here we ask experts to assign probabilities to the individual models of the ensemble for the specific case studies.

The same logic is used for both climate models and hydrological models, but the specific questions are different.

2.4 Step 2: Select experts

The selection of experts is crucial for the outcome of the elicitation. The selection is based on the following criteria:

- Two groups of experts are selected: climate modellers and hydrological modellers.
- The experts may be from a combination of i) AQUACLEW participants; ii) AQUACLEW partner organisations but not participants in the project; and iii) external experts from other organisations. It is important that some of the experts are not project participants with a stake in model performances.
- Experts shall be selected based on their scientific experience (publication or academic records, recommendation by peers) and in theory should encompass the highest possible variability (geography, modelling experience, academic experience, experience with models etc.)

Elicitations can take place either individually, by interviews or questionnaires, or in group work or a combination of the two. We choose to make individual elicitations using questionnaires, as this in practice allows collection of independent EE inputs in a time-efficient way from a wider group of external experts. We evaluate that at least five experts in each modeller group is a minimum requirement for a solid EE.

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

Experience shows that more reliable results are produced by applying a weighting scheme for expert opinions instead of giving equal weight to each expert opinion. We can experiment by different weighting schemes and test the sensitivity/robustness of the result to the weighting. Different weighting schemes will be applied in an approach where we ask experts to score each other and themselves anonymously.

The motivational bias in expert opinion will be tested in advance by self-scoring (some experts may have been involved in work related to one or the other climate model, hydrological model etc.). We do not intend to disqualify experts with motivational bias but we can assess the sensitivity of the bias to the expert elicitation results, and we will discuss the extent and impact of possible motivational biases on the results of the expert elicitation.

2.5 Step 3: Plan the elicitation

2.5.1 Questions on climate models

The questionnaire is described in more details in Section 6 and consists of block 1 and 2 with the aim of setting the scene, introducing the approach to the participating experts and collecting qualitative information/metadata from block 1 part and quantitative information (probabilities) from block 2 part of questionnaire.

The aim of the questions in block 1 is to bring the experts' attention to a variety of factors they should have in mind when ranking models at the end of block 1 and assigning probabilities to them in block 2.

The questions in block 1 should include some general questions on processes of importance for projecting the climate signal, e.g. parameterisation for cloud formation and convective rainfall, as well as some questions relating to case studies, e.g. which model would be best or worst to project climate change in the Mediterranean region or in Northern Europe.

The aim of block 2 is to assess which models in the ensemble are the most plausible for specific case studies. Ranking the climate models for specific regions and modelling purposes gives a general idea about which climate models perform best in specific regions and for specific purposes. Ranking prepares the experts for the probability questions, as it is much easier to rank than to assign probability. The experts' ranking for the regions and modelling purposes could be used to cross check the consistency of their answers in the final question about model probabilities. In addition to this it can be used to collect information of prior knowledge/assumptions.

The final question in block 2 will be to ask experts to fill in numbers in a probability chart looking as in Table 1.

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

Table 1 Probability chart for eliciting model probabilities. The total probability of all climate models should be equal to 100% and several climate models can have the same probability value.

Climate model	Case study 1	Case study 2	Case study 3	Case study 4	Case study n	Confidence in general in climate models (1: low - 5: high)
1						
2						
3						
4						
N						
Total probability	100%	100%	100%	100%	100%	
Confidence in ranking (1: low - 5: high)						

The two confidence columns refer to confidence in the climate model in general (vertical column to the right) and confidence in the ranking of climate models for a specific case study (horizontal row at the bottom). Confidence intervals could be given verbally using a scale of 5, where 1 means not confident at all and 5 means very confident (not confident all/ less confident/somewhat confident/confident/very confident).

The vertical confidence column gives again an impression in the general plausibility of climate models which could be cross-checked with the questions of block 1 to be sure that the experts are consistent in their answers.

2.5.2 Questions on hydrological models

The questions on hydrological models follows the same logic as for questions on climate models.

For the elicitation of model probabilities experts are asked to fill in numbers in a chart looking as Table 2.

Table 2. Probability chart for eliciting model probabilities. The total probability of all hydrological models should be equal to 100% and several hydrological models can have the same probability value.

Hydrological model	Case study 1	Case study 2	Case study 3	Case study 4	Case study n	Confidence in general in hydrological models (1: low - 5: high)
1	X					
2	X	X		X		
3		X		X		
4		X	X	X		
n				X		
Total probability	100%	100%	100%	100%	100%	
Confidence in ranking (1: low - 5: high)						

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

For some case studies only a limited number of hydrological models will be available and model probabilities will only be elicited for case studies including at least three hydrological models.

2.5.3 Comments on elicitation

Generally scientific literature is quite against eliciting probabilities directly, so it is good to have two blocks of questions with similar content, but a bit different format, so that the answers could be checked for consistency. Experts are generally better at ranking than giving probabilities.

Ranking, e.g. rank members of the climate model ensembles/hydro models in order of plausibility, is very simple and fast, but does not give the relative probability between the models. It is generally not recommended to rank more than 7 items at a time. As two of our case studies plan to each include 10 climate models we may face a challenge here that we need to look into.

Assigning probabilities to the models is a bit more time consuming, but two models can have the same probability and the probability of models relative to each other is more defined

During the verbal elicitation of probability/confidence the experts could see the verbal description of probability and also a probability value assigned to it. It could help not only in the verbal assessment, but also filling in the probability chart. It is easier to think in terms of phrases like virtually certain, than in terms of probability values.

2.6 Step 4: Training of the experts

2.6.1 Training material

The following training material will be prepared and made available for the experts prior to the EE workshop:

- Case studies
    - Information on case studies with focus on key issues for climate and hydrological change projections. This will be based on materials on http://aquaclew.eu/ *to be prepared by case study owners.*
- Climate models
    - Description of climate models included in the climate model ensemble with emphasis on processes. *This will be included in a SMHI deliverable on evaluation of climate models.*
    - Performance metrics for how good the climate models are in simulating present climate including results from the Differential Split-Sample tests for bias correction methods. *This will be prepared by case study owners.*
- Hydro models
    - Descriptions of hydrological models included in the hydrological model ensembles in the case studies.
    - Description of calibration and validation results from the case studies, including results from the Differential Split-Sample tests. *This will be prepared by case study owners.*

2.6.2 EE workshop

The EE workshop is expected to last a full day with a short break between the training of experts and the elicitation exercise.

The first half day of the EE workshop will be devoted to training of the experts. The program will include:

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

- An introduction to expert elicitation and the workshop
- "Pecha kucha" presentations on case studies with emphasis on climate change and hydrological change aspects.
- An overview presentation on climate models with emphasis on the model capabilities and process description of importance for climate change projections in the case studies.
- An overview presentation on hydrological models with emphasis on the model capabilities and process description of importance for hydrological change projections in the case studies.
- Questions from experts – discussion.

2.6.3 Considerations and pitfalls

There are a number of possible biases in expert elicitations that should be avoided:

- Terminology ambiguity. As the workshop includes experts and models from a variety of disciplines, it is important that the various concepts and definitions are clear and understood by all experts.
- Motivational bias. Developers or promoters of a model or who worked mainly with one of them could be biased during elicitation, both over- and underestimating the model probabilities. This could be partly checked by the self-scoring questionnaires addressing the work/modelling experience of experts and the confidence attributed to the models in block 2
- Anchoring. Some experts will most likely use probability values given for case study 1 as a basis to assign probabilities for case study 2. It makes the elicitation faster, but what if they are too influenced by the original numbers and do not evaluate the cases independently? Could be potentially checked by comparing results of block 1 and block 2.
- Availability. Experts' decision can be biased by previous experience from similar case studies.
- Peer-pressure. Depending on the elicitation environment experts could be pressured to hurry up if other experts are faster to finish etc.

It should be emphasized that it is not compulsory to answer the questions and that there are no right and wrong answers, it the most precise description of their subjective opinion that the study is interested in.

We should show an example of the probability table with imaginary data, best if not even related to climate and hydrology.

2.7 Step 5: Elicitation

The elicitation questionnaire should be tested beforehand with an expert to see where questions/misunderstandings can arise, how much time the elicitation takes, and if the questions and the scenarios make sense to the experts.

During the elicitation, experts respond individually to the questions outlined in Section 6 below. The experts provide the answers in an online version of the questionnaire.

An EE session is ideally shorter than 1½ hours. If time allows, climate and hydrological modellers can, as an experiment, do each other's questions in addition to their own questions. Facilitators should be available during the elicitation to help participating experts understanding the tasks and the questions.

When the questionnaires have been completed by all experts the first raw results will be made available and subject to discussion among the experts. This discussion should be recorded, because many good ideas may surface during such discussion and serve as inspiration for the discussion in the journal paper that is planned as an output. Furthermore, the discussions among the experts may be

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

the source for the qualitative/grounded research component (Strauss & Corbin 1998) in order to properly evaluate the findings from the applied EE methodology and eventual biases beyond the above pitfalls. One example could be to which degree previous experiences (AR4/Ensembles models) bias the outputs from EE-sessions, and in which way various expert knowledge frames influence the EE results. Another example could be that some experts judge that indirect impacts which are not included in models result in low probability and/or ranking of a given model.

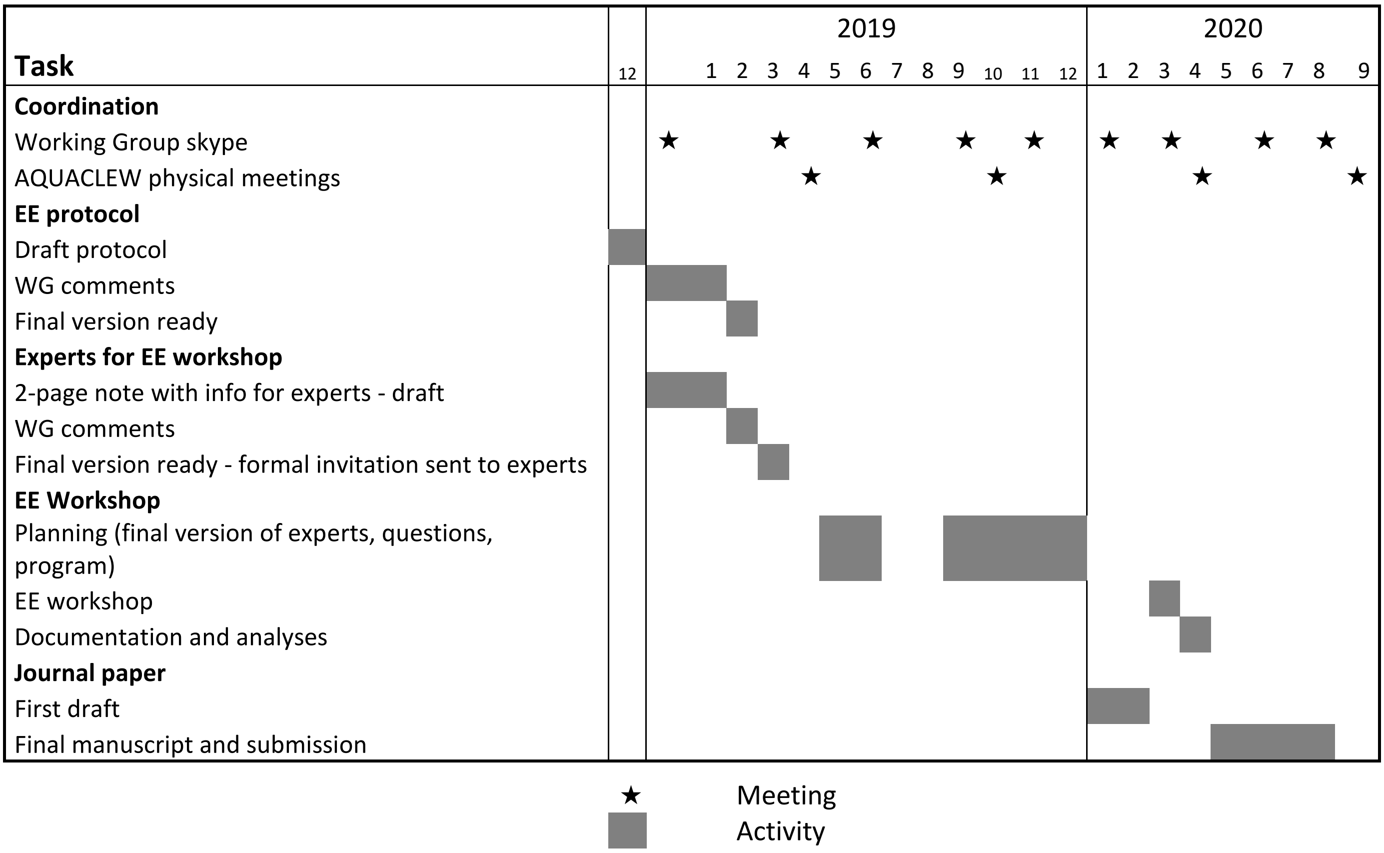
Technically, experts can do the elicitation alone (remote), but participating in a common training session is a must. This allows that some of experts can join the EE workshop via video links, but in order to ensure coherence in the discussions during the training session and during the final session on results, we need the majority of the experts to physically attend the EE workshop – hence in practice we can afford a couple of experts to attend via videolink.

2.7 Step 6: Aggregation and analysis of results

The aggregation of results should ideally include weighting of individual results. We should then experiment with alternative weighting schemes thus testing the sensitivity of the outcome to different assumptions on weighting of experts.

The further analysis should make use of the plenary discussion at the EE workshop after the results have been shown for the first time. Subsequently, the experts should be invited to participate in the evaluation and discussion of the results and in this way co-author a final joint journal paper.

3. Work plan and time schedule



4. Case studies

An overview of the expected use of model ensembles in the seven case studies is provided in Table 3. As can be seen from the table, all case studies expect to make use of EE to evaluate climate models,

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

while EE is only relevant for hydrological models in the two case studies that use ensembles comprising at least three hydrological models.

Table 3 Use of models and expected participation in EE

Case study		Climate models	Hydrological models	Participating in EE?	
Country	Key issues			Climate	Hydrology
SE	Water quality in lake	10	2	Yes	No
DK	Irrigation and drainage requirements	10	3	Yes	Yes
DE	Urban flash floods	0	0	No	No
A	Pluvial flash floods in pre-alpine regions	0	0	No	No
FR	Hydropower	11	2	Yes	No
ES-UCO	Drought and water resources allocation	4	4	Yes	Yes
ES-UGR	Fluvial and coastal interactions	7	2	Yes	No
				5	2

5. Selected experts

The climate and hydrological modelling experts expected to participate in the elicitation are listed in Tables 4 and 5. If only two cases (Spain and Denmark) ultimately decides to join the EE for hydrological models (Table 3) we may decide to replace some of the hydrological modelling experts with some that have a more in-depth knowledge of the catchments used for the two case studies.

Table 4 Status of climate modelling experts expected to participate in the elicitation

Name	Organisation	Link to AQUACLEW	Status
NN1	SMHI/Rossby Centre	Partner organisation	SMHI to identify/invite
NN2	BOKU	Partner organisation	BOKU to identify/invite
NN3	IRSTEA	Partner organisation	IRSTEA to identify/invite
NN4	Spanish climate group	External	UCO to identify/invite
NN5	EDF	External	IRSTEA to identify/invite one external person from France
	Recherche CNRS	External	
Jens Hesselbjerg Christensen	University of Copenhagen	External	JCR talked to Jens – he is interested, waiting for a formal invitation and a date

Table 5 Status of Hydrological modelling experts expected to participate in the elicitation

Name	Organisation	Link to AQUACLEW	Status
Maria	UCO	Participant	Confirmed
NN1	SMHI	Participant or partner organisation	SMHI to identify/invite
	IRSTEA	Participant or partner organisation	IRSTEA to identify/invite

	GEUS	Participant or partner organisation	GEUS to identify/invite
	UIBK	Participant or partner organisation	UIBK to identify/invite

6. Questionnaires

6.1 Scoring of experts

To enable different weights for experts in the aggregation of results we will make two scorings of experts derived from two fundamentally different procedures: i) objective data on scientific experience of experts, and ii) subjective assessments, where each expert is asked to assess the quality of all experts.

1. Provide information on your scientific experience in the table

	Response
Scientific competence	
Scientific experience (years in research positions since PhD)	
h-index on Web of Science	
Climate modelling	
Experience with developing code for climate models (None/0-3 years/4-9 years/>10 years)	
Experience with running/operating climate modelling (None/0-3 years/4-9 years/>10 years)	
Experience from climate change research projects, where you had no role in operating the climate model but contributed to design model studies and evaluate/use results from climate model simulations (None/0-3 years/4-9 years/>10 years)	
Experience with bias correction and/or downscaling of climate models (None/0-3 years/4-9 years/>10years)	
Number of papers directly linked to climate model topics in the last 10 years (on Web of Science)	
Hydrological modelling	
Experience with developing code for hydrological models (None/0-3 years/4-9 years/>10 years)	
Experience with running/operating hydrological modelling (None/0-3 years/4-9 years/>10 years)	
Experience with conducting hydrological modelling research projects, where you had no role in operating the hydrological model but contributed to design model studies and evaluate/use results from hydrological model simulations (None/0-3 years/4-9 years/>10 years)	
Experience with climate change impact and/or adaptation modelling projects (including Disaster Risk Reduction/Nature Based Solutions) (None/0-3 years/4-9 years/>10 years)	
Number of papers directly linked to hydrological model topics in the last 10 years (on Web of Science)	

2. Provide your views on how much weight should be assigned to each of the experts for eliciting probabilities for climate models and hydrological models, respectively. Bio data/CV will be provided in the process.

		Weight in assessing hydrological models	Weight in assessing climate models
Expert 1			
Expert 2			
Expert 3			
Expert 4			
Expert 5			
Expert 6			
Expert 7			
Expert 8			
Expert 9			
Expert 10			
Expert 11			
Total weight		100%	100%

6.2 Climate models

<u>Block 1 – general questions about each climate model</u>

1. To what degree are the main assumptions and included process parameterizations of each model combinations (GCM-RCM) solid and reasonable (high/intermediate/low)? (GCMs and RCMs are evaluated separately from BC)

2. How well does each model (GCM-RCM combination) represent the atmospheric circulation patterns over Europe (high/intermediate/low)?

3. Based on the description of each model (GCM-RCM as a pair), to what degree is the model capable of simulating precipitation at a day to multi-day timescale, and its temporal variations in general (high/intermediate/low)?

4. Based on the description of each model (GCM-RCM as a pair), to what degree is the model capable of simulating precipitation at a sub-daily timescale, and its temporal variations in general (high/intermediate/low)?

5. Based on the description of each model (GCM-RCM as a pair), to what degree is the model capable of simulating persistency of dry periods (meteorological droughts), and its temporal variations in general (high/intermediate/low)?

6. To what degree are the main assumptions of each BC method solid and reasonable (high/intermediate/low)?

<u>Block 2 – specific questions in relation to applications in case studies</u>

7. Danish case: Based on your opinion about the general model structures and the performance metrics of the complete chain of models (GCM-RCM-BC), to what degree is the model suitable to project changes in drought conditions (leading to agricultural drought) as well as wet periods (leading to groundwater flooding) as a result of climate change? (high, intermediate, low)

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

8. Danish case: Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the N models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for projecting changes in drought conditions (leading to agricultural drought) as well as wet periods (leading to groundwater flooding) as a result of climate change? Specify why.
9. Spanish (UCO) case: Based on your opinion about the general model structures and the performance metrics of the complete chain of models (GCM-RCM-BC), to what degree is the model suitable to project changes in snow conditions as a result of climate change? (high, intermediate, low)
10. Spanish (UCO) case: Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the N models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for projecting changes in snow conditions as a result of climate change? Specify why.
11. French case: Based on your opinion about the general model structures and the performance metrics of the complete chain of models (GCM-RCM-BC), to what degree is the model suitable to project changes in seasonal precipitation (average amount as well as extremes in the form of droughts and very wet periods) and temperature (average and extremes) as a result of climate change? (high, intermediate, low)
12. French case: Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the N models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for projecting changes in seasonal precipitation (average amount as well as extremes in the form of droughts and very wet periods) and temperature (average and extremes) as a result of climate change? Specify why.
13. Swedish case: (awaiting more input on the main processes studied)
14. Spanish (UGR) case: Based on your opinion about the general model structures and the performance metrics of the complete chain of models (GCM-RCM-BC), to what degree is the model suitable to project changes in sea-level and storms that affect wave height and surges as a result of climate change? (high, intermediate, low)
15. Spanish (UGR) case: Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the N models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for projecting changes in sea-level and storms that affect wave height and surges as a result of climate change? Specify why.

For the modelling purpose of case study 1 (ex. drought and water resource allocation in Spain) assign a probability to the climate models which expresses to what degree the experts believe that the climate model gives plausible/useful results for the case study. The total probability of all climate models should be equal to 100% and several climate models can have the same probability value.

Climate model	Case study 1	Case study 2	Case study 3	Case study 4	Case study n	Confidence in climate model in general (1: low - 5: high)
1						
2						
3						
4						
N						
Total probability	100%	100%	100%	100%	100%	

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

Confidence in the ranking of the climate models (1: low - 5: high)						

Aim of block 2: Assess which climate models in the ensemble are the most plausible for specific case studies.

The vertical confidence column gives again an impression in the general plausibility of climate models, which could be cross-checked with the questions of block 1 to be sure that the experts are consistent in their answers.

6.3 Hydrological models

The questions below are tentative and need to be updated based on more information on case studies (selection of models and intended use of models for projections).

Block 1 – general questions about each hydrological model

1. To what degree is the hydrological model assumptions solid and reasonable (high/intermediate/low)
2. To what degree is the model likely to represent conditions under future climate change, against which it cannot presently be calibrated or validated (high/intermediate/low)
3. To what degree does the model (concept, assumptions, implementation and results) agree with your knowledge and experience? (high, intermediate or low). If your answer is ‘‘low’’, please specify the reason.
4. Based on the description of physical processes and assumptions in the model, to what degree is the model capable of simulating *streamflow and its temporal variations (average runoff, flow duration curve, dynamics)* in general? (high, intermediate, low)
5. Based on the description of physical processes and assumptions in the model, to what degree is the model capable of simulating *drought conditions (low flows, low groundwater table)* in general? (high, intermediate, low)
6. Based on the description of physical processes and assumptions in the model, to what degree is the model capable of simulating *groundwater flooding (need for drainage in agricultural soils)* in general? (high, intermediate, low)

Block 2 – specific questions in relation to applications in case studies

7. Danish case. Based on your opinion about the general model structure and the performance metrics of the calibrated model, to what degree is the model suitable to predict changes in *drought conditions (low flows, low groundwater table) as well as groundwater flooding (need for drainage in agricultural soils)* as a result of climate change? (high, intermediate, low)
8. Danish case. Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the 5 models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for predicting changes in *drought conditions (low flows, low groundwater table) as well as groundwater flooding (need for drainage in agricultural soils)* as a result of climate change. Specify why.
9. Spanish-UCO case. Based on your opinion about the general model structure and the performance metrics of the calibrated model, to what degree is the model suitable to predict changes in *streamflow and its temporal variations (average runoff, flow duration curve, dynamics)* as a result of climate change? (high, intermediate, low)

10. Spanish-UCO case. Based on your opinion about the general model structure and the performance metrics of the calibrated model, rank the 4 models (with rank 1 as the least plausible and rank N as the most plausible model) which according to your belief are the most suitable for predicting changes in *streamflow and its temporal variations (average runoff, flow duration curve, dynamics)* as a result of climate change. Specify why.
11. What is the probability value that best represents your confidence in each hydrological model given the model performance and purposes of model application in each case study. Different models may be assigned the same probability, but the total of probabilities for all models in each case study must be 100%. Use the table below and also assign confidence in the hydrological model in general (vertical column to the right) and confidence in the ranking of hydrological models for a specific case study (horizontal row at the bottom). Confidence intervals could be given verbally using a scale of 1 to 5 (very low/low/intermediate/high/very high).

Hydrological model	Case study DK	Case study ES-UCO
1	X	
2	X	
3	X	
4	X	
5	X	
6		X
7		X
8		X
9		X
Total probability	100%	100%
Confidence in ranking (1: low - 5: high)		

16. References

Meyer MA, Booker JM (2001) Eliciting and Analyzing Expert Judgment: A Practical Guide. ASA-SIAM Series on Statistics and Applied Probability.

Sebok E, Refsgaard JC, Warmink JJ, Stisen S, Jensen KH (2016) Using expert elicitation to quantify catchment water balances and their uncertainties. Water Resources Research, 52, 5111-5131.

Ye M, Pohlmann KF, Chapman JB (2008) Expert elicitation of recharge model probabilities for the Death Valley regional flow system. Journal of Hydrology, 354, 102–115.

Strauss A and Corbin J (1998) Basics of Qualitative Research. London: Sage.

Project AQUACLEW is part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).