

How Does Seasonal Forecast Performance Influence Decision-Making?

Insights from a Serious Game

Louise Crochemore, Carolina Cantone,
Ilias G. Pechlivanidis, and Christiana S. Photiadou

ABSTRACT: In a context that fosters the evolution of hydroclimate services, it is crucial to support and train users in making the best possible forecast-based decisions. Here, we analyze how decision-making is influenced by the seasonal forecast performance based on the Call For Water serious game in which participants manage a water supply reservoir. The aim is twofold: 1) train participants in the concepts of forecast sharpness and reliability, and 2) collect participants' decisions to investigate the levels of forecast sharpness and reliability needed to make informed decisions. In the first game round, participants are provided with forecasts of varying reliability and sharpness, while in the second round, they have the possibility to pay for systematically reliable and sharp forecasts (improved forecasts). Exploitable answers were collected from 367 participants, predominantly researchers, forecasters, and consultants in the water resources and energy sectors. Results show that improved forecasts led to better decisions, enabling participants to step out of purely conservative strategies and successfully take risks. Reliability levels of 60% are necessary for decision-making while both reliability levels above 70% and sharpness are required for informed risk-prone strategies. Improved forecasts are judged more valuable in extreme years, for instance, when hedging against water shortage risks. Additionally, participants working in the energy, air quality, and agriculture sectors, as well as traders, decision-makers, and forecasters, invested the most in forecasts. Finally, we discuss the potential of serious games to foster capacity development in hydroclimate services and provide recommendations for forecast-based service development.

KEYWORDS: Forecast verification/skill; Forecasting; Seasonal forecasting; Climate services; Communications/decision making; Education

<https://doi.org/10.1175/BAMS-D-20-0169.1>

Corresponding author: Louise Crochemore, louise.crochemore@outlook.fr

In final form 25 April 2021

©2021 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

Hydrometeorological forecasts stand at the core of early warning systems and hydroclimate services that support disaster risk prevention (Street et al. 2019), energy production (Terrado et al. 2017; Troccoli 2018), agriculture (Rembold et al. 2019; Vaughan et al. 2019), environmental protection, or water resources management (Arnal et al. 2018; Crochemore et al. 2020). While short- to medium-range forecasts, which extend up to 7–15 days ahead, can be key tools in day-to-day operations, long-range [i.e., (sub-)seasonal and decadal] forecasts can support strategic planning (Bruno Soares et al. 2018a). One of the current priorities in climate research is to define strategies to tailor hydroclimate information and address local needs through hydroclimate services (Hov et al. 2017). The past decade has thus seen the creation of a number of services based on seasonal hydrometeorological forecasts (e.g., Arnal et al. 2018; Sutanto et al. 2019; Samaniego et al. 2019; Contreras et al. 2020), which should ideally be codeveloped by data providers, service developers, and users (Buontempo et al. 2018; Lavers et al. 2020).

The lack of reliability and the high levels of uncertainty in seasonal forecasts have been identified as two of the main barriers in the uptake of these products to support decision-making in climate-sensitive sectors (Bruno Soares and Dessai 2016; Hewitt et al. 2017). Probabilistic forecasts have long been used to convey these uncertainties through ensembles of possible future outcomes. Efficiently communicating forecast uncertainties and performance is essential to overcome the barriers of credibility and trust that data providers confront (Hartmann et al. 2002; Coelho and Costa 2010; Cloke and Pappenberger 2009; Lavers et al. 2020). However, this communication remains a challenge in the context of climate services, and requires tailored visualization, metrics, and vocabulary to ease user understanding, meet their needs, and thus foster the uptake of climate services (Vaughan and Dessai 2014).

Understanding the role of forecast performance in decision-making can help tailor hydroclimate information into impactful services. Currently, in Europe, only a few decision-making organizations provide clear guidelines on the level of forecast performance required to take action (Bruno Soares et al. 2018a), and that level has not been investigated in an empirical and systematic way. This paper aims at exploring the following research questions: 1) How does seasonal forecast performance influence decision-makers' behavior, and 2) how is the forecast performance information perceived? The analysis is based on participants' game plays and answers to a serious game that sets a constrained decision-making environment. Additionally, this paper highlights the usefulness of serious games to train hydroclimate service users and enhance the marketability of these services.

Forecast uncertainty, communication, and value for decision-making

Despite decades of development in meteorological, hydrological, and climate forecasts (Yuan et al. 2015), and extensive research targeting forecast improvements for local use (e.g., Lucatero et al. 2018; Musuza et al. 2020; Baker et al. 2020; Macian-Sorribes et al. 2020), the skill of these forecasts is bound to remain imperfect (Girons Lopez et al. 2021; Pechlivanidis et al. 2020). Forecasts are inherently uncertain due to the chaotic nature of the atmospheric processes and to the limitations in physical process understanding and representation. Nevertheless, all disciplines subject to forecasting practices experience challenges related to forecast uncertainty identification, quantification, and communication to users (Ramos et al. 2010).

Probabilistic representations of forecasts tend to be favored by the general public, who appears to be aware of and thus expect uncertainties, for instance, in weather forecasts (Joslyn and Savelli 2010; Morss et al. 2018). Visualizing uncertainties in forecasts was even found to foster risk-avoiding behaviors (Ramos et al. 2013). In the field of financial forecasting, Önkál and Bolger (2004) concluded that, out of four visualization methods (i.e., point, directional, 50% interval and 95% interval), the 95% probabilistic range was considered the most useful by forecast users and providers. Similarly, Stephens et al. (2019) found that a representation of temperature forecasts showing both the interquartile range and the 90% probabilistic interval led to fewer misinterpretations than other visualization methods (i.e., table, line graph, 90% interval alone). Explicit probabilities should ideally complement forecast visualizations instead of having users implicitly perceive probabilities from the forecasts and, consequently, potentially misinterpret stand-alone probabilities (Stephens et al. 2019; Joslyn et al. 2009). Furthermore, LeClerc and Joslyn (2012) concluded that communicating odds based on climatology rather than just probabilities supported decision-making in extreme event detection.

In addition to the communication of forecast uncertainties, forecast performance is one of the factors influencing the value of hydroclimate forecasts for decision-making (Bruno Soares et al. 2018b). The relationship between forecast performance and economic value is complex, nonlinear, and context dependent (Cassagnole et al. 2021). Nevertheless, recent research has focused on establishing this relationship in local case studies, showing that hydrologic conditions, as well as end users' attitude toward risk, influence the relationship between forecast skill and value, with potentially stronger relationships in extreme conditions than in normal conditions (Giuliani et al. 2020; Peñuela et al. 2020).

Serious games, hydroclimate forecasts, and decision-making

Serious games, i.e., games explicitly designed for educational purposes and not solely for entertainment (Mendler de Suarez et al. 2012; Aubert et al. 2018), are efficient means of training decision-making (Aubert et al. 2019; Caird-Daley et al. 2007). By communicating advanced concepts that require hands-on experience, such as climate adaptation (Flood et al. 2018) and water systems (Savic et al. 2016), games contribute to the training efforts needed to build the capacity of users in order to properly ingest hydroclimate information and thus toward climate services uptake (Nkiaka et al. 2019; Bruno Soares and Dessai 2015). Serious games may also serve in the early stages of codevelopment as platforms for service developers or data providers to discuss hydroclimatic data and service design. When such platforms rely on sequential experimental decisions, participants can train their forecast-based decision-making while informing researchers on how users engage with forecast information (Neumann et al. 2018; Crochemore et al. 2016). Game and activity sessions have notably been used to research the role of forecast uncertainty in decision-making (e.g., Ramos et al. 2013; Stephens et al. 2019), communicate the added value of seasonal climate forecasts (Terrado et al. 2019), and analyze the perceived value for decision-making of probabilistic forecasts exhibiting varying biases (Arnal et al. 2016).

The Call for Water game

Decision-making context. Call for Water follows a storyline that sets the decision-making context (Fig. 1, left). Participants play the role of reservoir managers of a fictional reservoir which supplies water to a town. The main objectives are primarily to ensure a water supply to the town for the summer season, and secondarily to manage an available budget while securing the water supply to the town. To supply sufficient water to the town for the whole summer season (from June onward), the reservoir has to contain an adequate water volume on

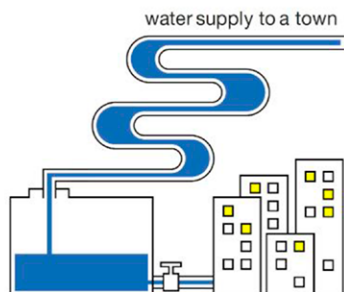
Introduction

You are the newly appointed **water manager** for Lake Bottiglietta, a reservoir that serves water uses for Thursty Town.

Thursty Town needs water for the period **June, July and August**. To ensure all water uses of Thursty Town all summer, Lake Bottiglietta **has to contain at least 60 Mm³ on June 1st**.

You are responsible to **secure sufficient water** (60 Mm³) for Thursty Town on June 1st.

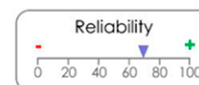
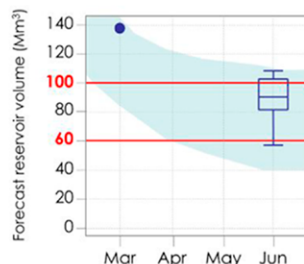
You will manage the reservoir for **two rounds of 5 years each**.



It is March 1st of Year 1

Your current balance: **50000**

What do you decide for June 1st?



WAIT AND SEE

DO NOTHING

SELL SURPLUS WATER

CONTACT NEIGHBOUR

Fig. 1. Game (left) introduction page and (right) decision page in the online version of the game.

1 June. The decision-making takes place from March to May and its general context has been simplified so that the only water use is the supply of drinking water and conflicts between uses or regulation between sectors are thus not accounted for.

From March to May, participants are presented with the latest observed reservoir volume (shown as single dots), seasonal probabilistic forecasts of the reservoir volume for 1 June (shown as boxplots), as well as forecast performance information (Fig. 1, right, and Fig. 2a). Based on this information, participants can decide to

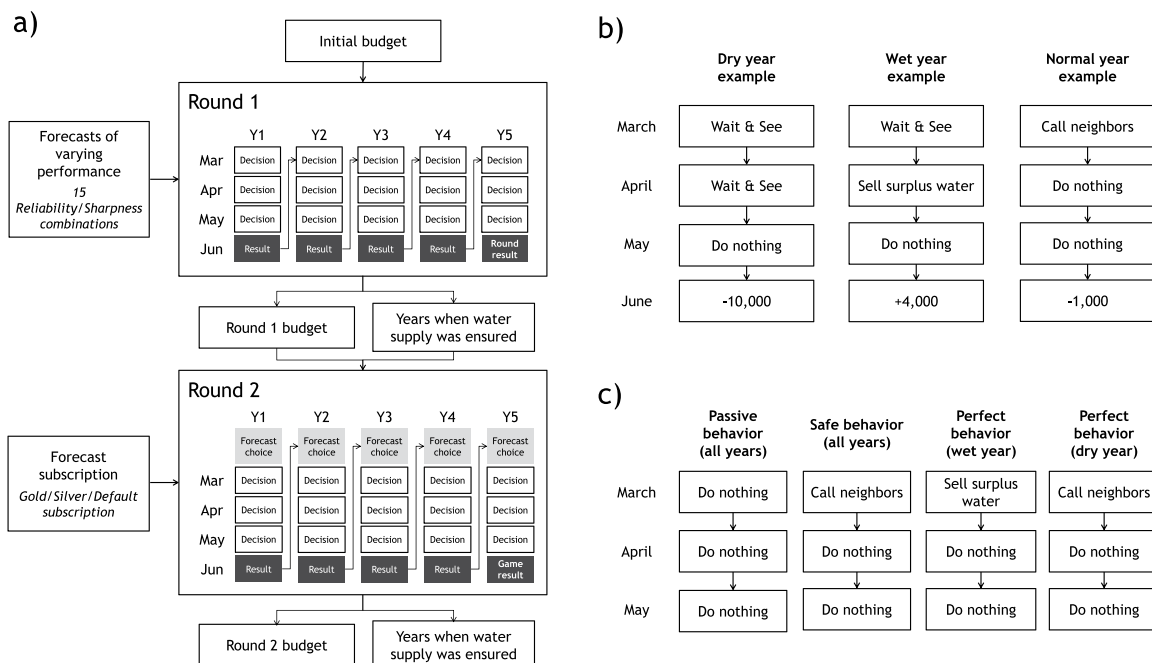


Fig. 2. (a) Game rounds, forecasts, and decisions flowchart, (b) examples of budget calculation, and (c) illustration of the three general behaviors defined.

- (i) *do nothing* if they think reservoir volumes will be sufficient to ensure water throughout summer,
- (ii) *call neighbors* to ask the neighboring reservoir to take over the water supply if participants think there is a risk of not reaching the adequate reservoir volume in June,
- (iii) *sell surplus water* if the reservoir volume is likely to exceed a high-volume threshold on 1 June, or
- (iv) *wait and see* if they judge that the current forecast information is not adequate to base a decision on.

All decisions only take effect in June and thus do not affect the reservoir volume itself, but affect the ability to supply water to the town and the participant's budget. Participants are given an initial budget of 50,000 tokens. This budget is affected by the participants' decisions and their associated costs (Table 1 and Fig. 2b), which are made explicit in the game rules and remain available to participants at all times (Fig. 4). The decision to *call neighbors* induces costs, while not ensuring sufficient water to the town causes a fine. The decision to *sell surplus water* when the reservoir does not reach the high-volume threshold in June causes a reputation loss, which is monetized in the game, in addition to the fine for not ensuring sufficient water; the decision to *sell surplus water* in a wet year yields gains. Last, *wait and see* has no cost per se; nevertheless, late decisions are penalized since the earlier the decision is made in the season, the less costly or more rewarding it is.

Forecasts, communication, and game rounds. Two forecast quality attributes, namely, sharpness and reliability, were chosen to convey forecast performance (Fig. 3, left, appendix A). Sharpness conveys the uncertainty in the forecasts, i.e., the concentration of the predictive distribution, and is an intrinsic attribute of the forecast. In the game, it is represented by a boxplot identifying the forecast spread (minimum value, 25th percentile, median, 75th percentile, and maximum value). In the analysis of results, sharpness was associated with a value ranging from 0 (forecast spread corresponding to the historical range of June reservoir volumes) to 100 (deterministic forecast; appendix A). Sharpness answers the question, How confident is the forecast in predicting future outcomes? Reliability (%) represents how often the observation has fallen within the forecast range in the past, thus ranging between 0 (forecast never covering the subsequent observation) and 100 (forecast systematically containing the observation; appendix A). Note that this is a simplification of the exact definition of reliability in forecasting chosen to make the game accessible to nonexperts. Reliability, as defined here, answers the question, Can one trust that the forecast range will contain the future scenario? These forecast quality attributes were chosen because they complement one another, and when combined, provide a comprehensive picture of forecast performance (Gneiting et al. 2007).

Table 1. Impact of decisions on budget balance.

Action	Conditions	Time of decision	Gain/loss	Type
Call neighbors	All years	March	−1,000	Cost
		April	−2,000	
		May	−5,000	
Sell surplus water	Wet years	March	+6,000	Gain
		April	+4,000	
		May	+2,000	
	Normal/dry years	Anytime	−2,000	Reputation loss
Not call neighbors	Dry years	Anytime	−10,000	Fine

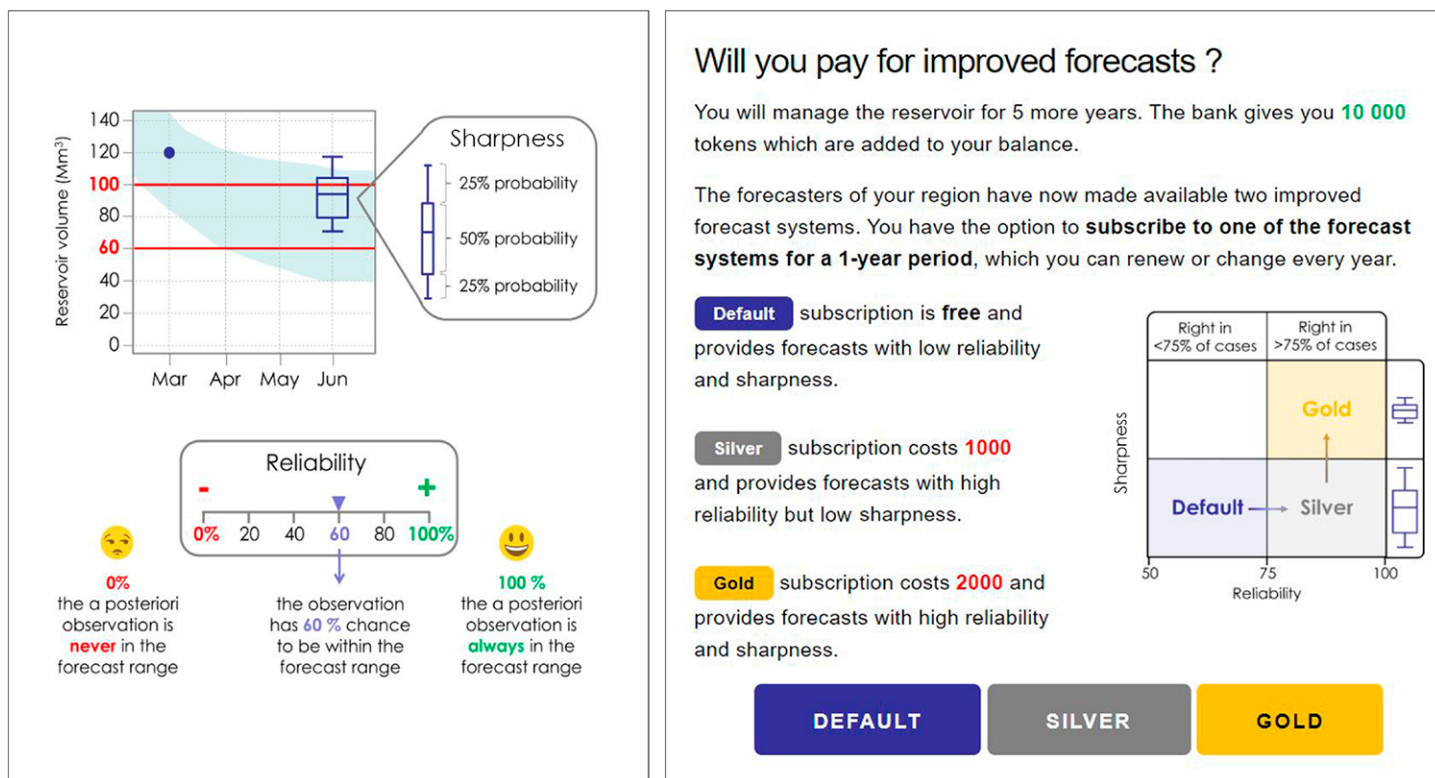


Fig. 3. (left) Forecast performance representation and explanation and (right) forecast subscription page in the online version of the game.

The game is divided into two rounds of 5 years each (Fig. 2a). Each round comprises two dry years when the participants should *call neighbors*, two wet years when the participants have the possibility to *sell surplus water*, and one normal year when the participants do not need to take action (*do nothing*). The participants do not have this information at the beginning of the game and their success depends on whether and how fast they detect the type of year (i.e., wet, dry, or normal) and its corresponding action. In the first round, the participants are presented with fictional forecasts of varying sharpness and reliability. The values of reliability and sharpness were chosen to cover reasonable but diverse combinations, i.e., high sharpness and high reliability, high sharpness and low reliability, low sharpness and high reliability, and low sharpness and low reliability (appendix B). In the second round, participants have the possibility to pay for a gold subscription that provides reliable and sharp forecasts or a silver subscription that provides reliable forecasts with low sharpness (Fig. 3, right). By default, participants have a subscription that provides forecasts of medium reliability with low sharpness. In both rounds, forecasts were created so that, over one game round, the a priori outcome was consistent with the declared forecast reliability (appendix B).

Behavior categorization. Three magnified behaviors yielding different game outcomes were defined to analyze participants' answers and locate game plays on a scale from passive to perfect based on their final budget. "Perfect" behavior corresponds to always securing water for the town while minimizing costs (i.e., *call neighbors* in the first month in dry years) and maximizing gains (i.e., *sell surplus water* in the first month in wet years) (Fig. 2c). This behavior yields a gain of 10,000 tokens in one game round. In practice, participants are not given adequate information to reach this behavior because forecasts are not perfect. "Safe" behavior corresponds to a risk-averse behavior that results in preventively choosing to *call neighbors* in the first month and never *sell surplus water*. This behavior yields a loss of 5,000 tokens in one game round. "Passive" behavior consists of never taking action regardless of the reservoir conditions and

Seasonal forecasting for water management in different sharpness and reliability contexts



A PRELIMINARY QUESTIONS

Professional background :

☐ Hydrology / ☐ Climate / ☐ Meteorology / ☐ Economy / ☐ Other:

Sector:

☐ Energy / ☐ Water resources / ☐ Risk prevention / ☐ Agriculture / ☐ Health /
☐ Other:

Role / Task in the company :

☐ Forecaster / ☐ Decision-maker / ☐ Trader / ☐ Researcher / ☐ Consultant /
☐ Other:

Experience :

☐ < 5 years / ☐ 6-10 years / ☐ 11-15 years / ☐ 16-25 years / ☐ > 25 years

Country :

REMEMBER THE RULES!

Your initial budget is: **50 000** tokens

If you cannot provide water to Thursty town for the season (60 Mm³ on June 1st),
you get a fine of - **10 000**

If you decided to sell surplus water:

- and you have more than 100 Mm³ on June 1st, you gain:

+ **6 000** if you first sold in March

+ **4 000** if you first sold in April

+ **2 000** if you first sold on May 1st

- and you don't reach 100 Mm³ on June 1st, you lose - **12 000** because you did
not provide Thursty Town and lose reputation

If you had to contact neighbouring water providers, you pay:

- **1 000** if you informed them as early as March

- **2 000** if you informed them in April

- **5 000** if you informed them on May 1st

B ROUND 1

Start budget: 50 000

		Wait & see	Do nothing	Sell surplus water	Contact neighbours
Year 1	March	☐	☐	☐	☐
	April	☐	☐	☐	☐
	May	☐	☐	☐	☐
	June 1 st	50 000 + - =			
Year 2	March	☐	☐	☐	☐
	April	☐	☐	☐	☐
	May	☐	☐	☐	☐
	June 1 st	 + - =			
Year 3	March	☐	☐	☐	☐
	April	☐	☐	☐	☐
	May	☐	☐	☐	☐
	June 1 st	 + - =			
Year 4	March	☐	☐	☐	☐
	April	☐	☐	☐	☐
	May	☐	☐	☐	☐
	June 1 st	 + - =			
Year 5	March	☐	☐	☐	☐
	April	☐	☐	☐	☐
	May	☐	☐	☐	☐
	June 1 st	 + - =			

Final budget:

Seasonal forecasting for water management in different sharpness and reliability contexts



C ROUND 2

Default 0
Silver -1000
Gold -2000

Previous budget + 10 000 :

Silver -1000 Gold -2000			Wait & see	Do nothing	Sell surplus	Contact neighbours	
Y 1	Membership:	March	☐	☐	☐	☐	☐ worthy ☐ acceptable ☐ not worthy
	☐ Default	April	☐	☐	☐	☐	
	☐ Silver	May	☐	☐	☐	☐	
	☐ Gold	June 1	+.....-.....=.....				
Y 2	Membership:	March	☐	☐	☐	☐	☐ worthy ☐ acceptable ☐ not worthy
	☐ Default	April	☐	☐	☐	☐	
	☐ Silver	May	☐	☐	☐	☐	
	☐ Gold	June 1	+.....-.....=.....				
Y 3	Membership:	March	☐	☐	☐	☐	☐ worthy ☐ acceptable ☐ not worthy
	☐ Default	April	☐	☐	☐	☐	
	☐ Silver	May	☐	☐	☐	☐	
	☐ Gold	June 1	+.....-.....=.....				
Y 4	Membership:	March	☐	☐	☐	☐	☐ worthy ☐ acceptable ☐ not worthy
	☐ Default	April	☐	☐	☐	☐	
	☐ Silver	May	☐	☐	☐	☐	
	☐ Gold	June 1	+.....-.....=.....				
Y 5	Membership:	March	☐	☐	☐	☐	☐ worthy ☐ acceptable ☐ not worthy
	☐ Default	April	☐	☐	☐	☐	
	☐ Silver	May	☐	☐	☐	☐	
	☐ Gold	June 1	+.....-.....=.....				

Final budget:

D QUESTIONS & DISCUSSION

How useful did you find the following information in your decision-making?

	Not useful	Slightly useful	Useful	Very useful	Did not use it
Reliability information					☐
Forecast range					☐

What minimum level of reliability would allow you to make a forecast-based decision?

☐ 0 (always wrong) / ☐ 20 / ☐ 40 / ☐ 60 / ☐ 80 / ☐ 100 (perfect)

What minimum level of sharpness would allow you to make a forecast-based decision?

☐ 0 (climatology) / ☐ 20 / ☐ 40 / ☐ 60 / ☐ 80 / ☐ 100 (deterministic)

From which month were you willing to make a decision for June 1st?

☐ Mar / ☐ Apr / ☐ May

Have your decision criteria changed during the game? If so, how?

Comments

Fig. 4. Worksheet filled in by participants in the paper version of Call for Water.

forecasts and corresponds to an absence of management, i.e., always choosing to *do nothing*. This behavior yields a loss of 20,000 tokens in one game round. These magnified behaviors, even if they are exaggerated, are meant to frame and categorize the range of behaviors observed in the game context and translate these observed behaviors into real-world attitudes and tendencies.

Game sessions and participants. Answers from 496 Call for Water participants were collected through 1) paper sheets when playing in groups (189 participants) and 2) an online game platform (307 participants) (see the sidebar “Where to access Call for Water”). Answers were collected on 13 occasions, such as user workshops gathering stakeholders, service developers, and researchers, research project meetings, and university courses (Table 2), and continuously through the online platform. Both in the online and paper versions, participants start by filling in their profile (i.e., background, sector, role, and experience; Fig. 4, top). At the end of the game, they are invited to complete a survey and reflect on their perception of forecast performance (i.e., usefulness of the performance information and minimum levels needed) and the evolution of their decision-making throughout the game, while an open field allows them to provide feedback (Fig. 4, bottom). In the second round of the paper version, participants also had the possibility to rate at the end of each year the subscription they had picked (Fig. 4, bottom).

Answer sheets with a clear misunderstanding of the rules were

Where to access Call for Water

The online version of the Call for Water game is currently available as a tutorial alongside a range of hydroclimate services from SMHI's Hypeweb platform (<https://hypeweb.smhi.se/call4water-game/>). On the online platform, participants are autonomous throughout the rules presentation and the game.

The paper version of the game is available online and can be freely used for teaching or training (<https://hepex.inrae.fr/resources/hepex-games/>). This version of the game is based on slides presented by a facilitator that narrates the story and guides participants throughout the game. Answers are collected on paper worksheets. The game lasts about 45 min, including the introduction and presentation of the rules, the decision rounds, and a debrief.

Table 2. Description of game sessions.

Occasion	Date	Location	Participants type	Answers
CLARA project Second Multiuser Forum	13–14 Nov 2018	Córdoba, Spain	Stakeholders, service developers, managers, decision-makers	41
S2S4E Second General Assembly	21–22 Nov 2018	Toulouse, France	Stakeholders in the energy sector, service developers, climate scientists, IT	29
SMHI Hydropower days	28–29 Nov 2018	Norrköping, Sweden	Stakeholders in the energy sector, hydrologists	15
EGU Geosciences Game Night	10 Apr 2019	Vienna, Austria	Researchers, students in geosciences	11
EGU General Assembly	7–12 Apr 2019	Online (Vienna, Austria)	Researchers, students in geosciences	~12
Course at Lund University	11 Apr 2019	Lund, Sweden	Students in M.S. Water, Society and Climate Change	18
Course at Politecnico di Milano	May 2019	Milano, Italy	Students in M.S. Environmental Engineering	14
Game session with SMHI staff	5 Jun 2019	Norrköping, Sweden	Consultants in hydrology, scientists	11
Met Éireann training days at SMHI	14 Jun 2019	Norrköping, Sweden	Forecasters in hydrometeorology	8
IMPRES Third General Assembly	18 Jun 2019	Norrköping, Sweden	Scientists in hydrometeorology and climate	42
Science is Wonderful exhibition at the EC Research and Innovation days	25–26 Sep 2019	Online (Brussels, Belgium)	General public, employees from the European Commission	~15
Delft-FEWS international user meeting 2020	2 Nov 2020	Online	Stakeholders in the water resources and flood protection sectors	91
Course at Universitat Politècnica de Valencia	24 Nov 2020	Online	Students in M.S. and Ph.D. Water, academic staff	11

filtered out resulting in 175 (93%) exploitable answer sheets from the paper version. Online replies that were incomplete or that were replays from participants who had already played the game were also filtered out resulting in 192 (62%) exploitable answers from the online version. The final 367 participants whose answers are analyzed hereafter are predominantly researchers, forecasters, and consultants in the water resources and energy sectors, with backgrounds in hydrology and climate, and with varying years of experience (Fig. 5).

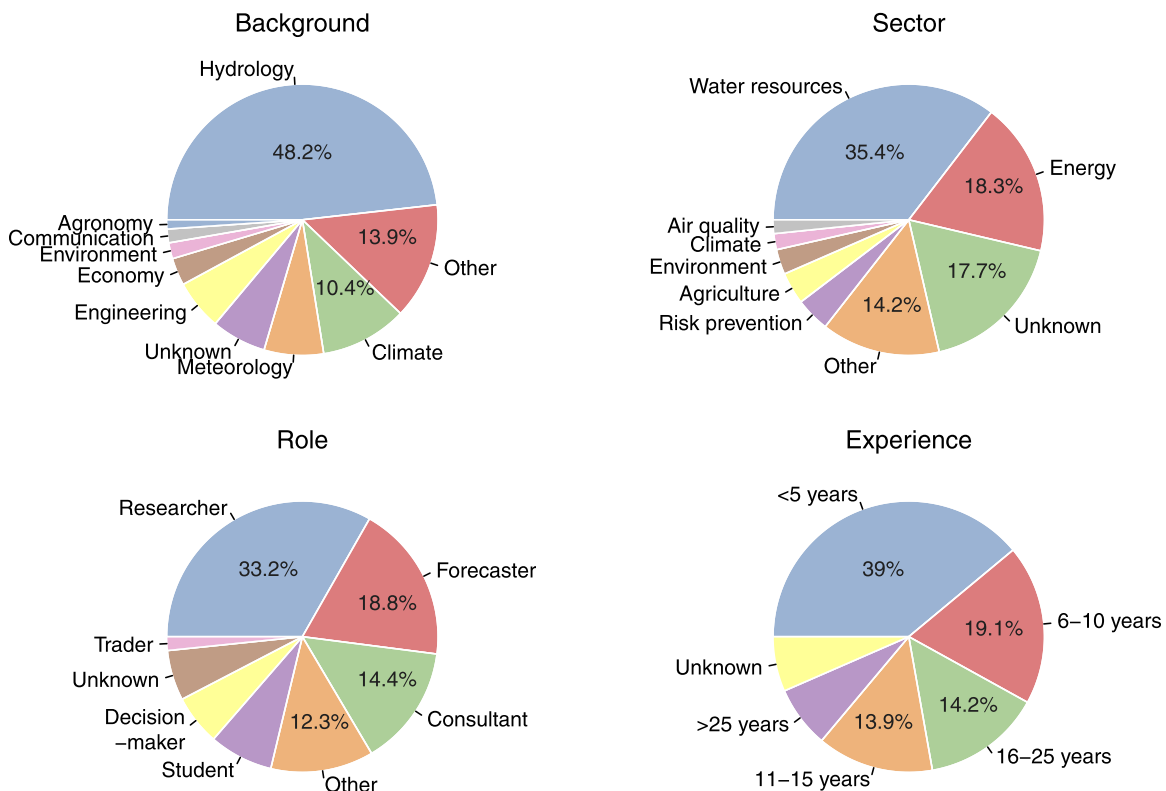


Fig. 5. Proportions in background, sector, role, and experience of the game participants.

How does forecast performance influence decision-makers' behavior?

Improved forecasts foster perfect behaviors. In the first game round, the first dry year was forecasted with a reliability and sharpness greater than 70%, while the second year was forecasted with performances lower than 70%. In the first case, 87% of participants successfully provided the town with water, while in the second case, only 39% managed to foresee the water shortage. In the case of wet years, the first wet year had unsharp but reliable forecasts, while the second wet year had sharp but unreliable forecasts. In the first year, 74% of participants successfully took the risk to *sell surplus water*, and in the second year, 66% of participants took that risk. Therefore, in these cases, the decision outcomes based on sharp and/or reliable forecasts always improve upon the outcome based on the unsharp and unreliable forecast (Table 3).

Table 3. Success rate (i.e., providing water in dry years and selling in wet years) based on forecast performance in the first round of the game.

	Low reliability	High reliability
Low sharpness	39% success (dry year)	74% success (wet year)
High sharpness	66% success (wet year)	87% success (dry year)

In the second game round, participants could subscribe to reliable forecasts on a yearly basis. Out of the 367 participants, 162 (22) systematically chose the gold (silver) subscription, while 32 systematically chose not to subscribe. After removing subscription costs, participants with a gold subscription were able to reach significantly higher gains than those without a subscription and than all participants in the first game round with forecasts of varying performance (Fig. 6). More than 73% of participants with a gold subscription made decisions between “perfect” and “safe,” while 81% of participants without a subscription showed behaviors between “passive” and “safe.”

Forecast performance plays a complementary role to participants’ backgrounds. In the first game round, behaviors were not significantly different¹ across sectors and roles, i.e., within each group 53%–81% of the participants adopted a behavior between “safe” and “passive.” In the second round, participants from the risk prevention sector (15/367), decision-makers (22/367), traders (6/367), and students (28/367) outperformed the other participants with more than 66% adopting behaviors between “perfect” and “safe.” Among the sectors and roles that paid for improved forecasts, the energy participants (67/367) ranked third out of the eight sector groups, and the decision-makers ranked second out of the seven role groups. These results highlight that the sector and role of the participants complemented forecast performance in game outcomes. Traders, decision-makers and forecasters, who were likely to be the most familiar with the decision-making process, invested the most, proportionately, in gold and silver subscriptions (“expertise” in Fig. 6). Nevertheless, the difference in outcomes between experts (“always gold expertise”) and nonexperts (“always gold”) is negligible and not statistically significant compared to the difference between participants with subscriptions (“always gold”) and participants without (“always default”). This shows that, in the context of Call for Water, expertise played a role in the strategic choice to invest or not in high quality forecasts, but had a negligible role in game outcomes when participants were given forecasts of similar quality.

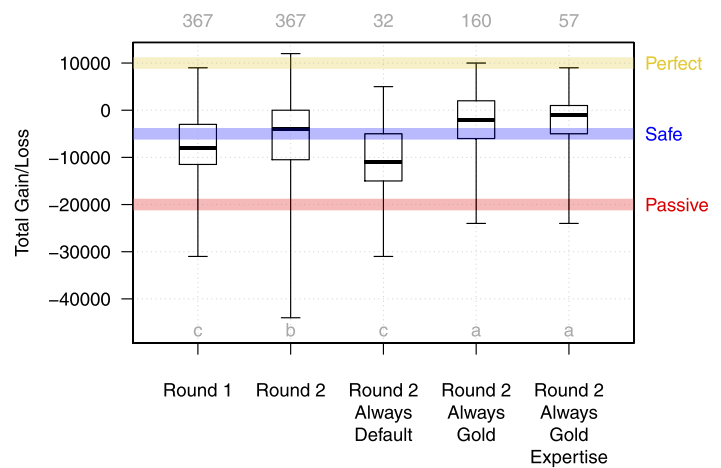


Fig. 6. Balance at the end of round 1, round 2, round 2 restricted to participants who never subscribed to any subscription, round 2 restricted to participants who systematically chose the gold subscription, and then further restricted to traders, decision-makers, and forecasters (expertise). The subscription costs have been removed from these balances. Numbers in gray indicate the sample sizes. Letters in gray indicate the results of the Kolmogorov–Smirnov test at the 5% level. Two boxplots sharing a letter are not significantly different. Colored lines indicate the three stereotypical behaviors.

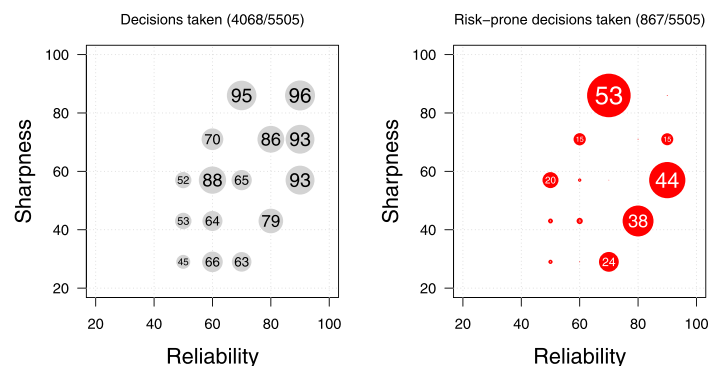


Fig. 7. Proportion (%) of participants making a decision for each combination of reliability and sharpness when considering (left) all decision types (*do nothing, call neighbors, and sell surplus*, i.e., all but *wait and see*) and (right) only risky decisions (*sell surplus water*). The number of opportunities to make a decision for each of the 367 participants was 15 (3 each year), resulting in 5,505 opportunities.

¹ Based on the Kolmogorov–Smirnov test resulting in no p value lower than 0.36 across sectors and lower than 0.12 across roles.

Table 4. Minimum and ideal combined levels of reliability and sharpness required for any type of decisions and risk-prone decisions.

Decision type	Minimum reliability	Minimum sharpness	Ideal reliability	Ideal sharpness
Any decision	60%	30%	70%	60%
Risk-prone	70%	40%	—	—

Minimum reliability levels are necessary to enable decisions. The forecast performance levels needed for participants to make a decision in the Call for Water decision-making context were inferred from participants' decisions in the first game round (Fig. 7 and Table 4). Between 47% and 55% of participants did not make decisions based on forecasts with reliability levels lower than 60% (i.e., forecasts that contained the observation 6 times out of 10 in the past). A reliability threshold of 60% or above would allow at least 64% of participants to make a decision instead of *wait and see*. Similarly, a reliability threshold of 80% allowed at least 79% of participants to make decisions. When considering sharpness and reliability simultaneously, forecasts with reliability and sharpness levels greater than 60% and 30%, respectively, allowed at least 63% of participants to make decisions, while levels greater than 70% and 60% (at the exclusion of this combination) allowed at least 86% of participants to make decisions. Reliability and sharpness levels greater than 70% and 40% were necessary to observe more than a third (38%) of risk-prone decisions.

How is the forecast performance information perceived?

Participants find the sharpness and reliability information useful. When asked in the final survey, 81% of participants considered that reliability and sharpness levels of at least 60% were necessary to inform decisions. This is consistent with the analysis of their game decisions, except that up to 79% of the participants actually made decisions with lower sharpness levels. When asked about the usefulness of the performance information, a large majority of the participants (86% for sharpness and 79% for reliability) found the information useful to very useful (Fig. 8). However, “very useful” was the most frequent answer for sharpness (46%) while “useful” was the most frequent answer for reliability (47%), suggesting that participants valued sharpness more than reliability.

Most participants pay for improved forecasts.

The subscription choice in the second round was fairly stable across the 5 years: 44% of participants systematically paid for the gold subscription, 6% systematically paid for the silver subscription, and 9% never subscribed. The remaining 41% changed their subscription at least once during the 5 years. There was a slight tendency to unsubscribe. At the beginning of the second round, 85% of the participants paid for improved forecasts, while by the end of the round, 79% were still paying. Participants from the energy, air quality, and agriculture sectors invested the most in forecasts (91%–92% subscribed), while participants from the climate sector invested the least (65% subscribed) (Fig. 9). In terms of role, traders

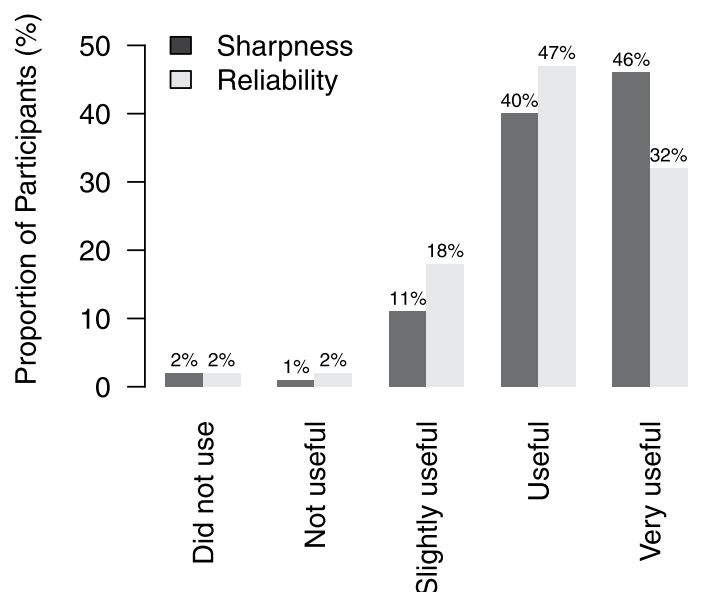


Fig. 8. Appreciation of the forecast performance information by the 265 participants who answered this question.

and decision-makers invested the most in forecasts, while consultants and researchers invested less than others (Fig. 9).

Participants rate forecasts of equivalent performance differently depending on the context. In each year of the second game round, participants who played the paper version could rate the subscription they had chosen as “worthy,” “acceptable,” or “not worthy” (Fig. 10). When looking at normal years, the default subscription received the highest number of “worthy” and “acceptable” ratings. The gold (silver) subscription received the highest number of “not worthy” (“acceptable”) ratings. In dry years, the gold subscription was judged “worthy” in a majority of cases (57%), the silver subscription was most often judged “acceptable” (47%), and the default subscription was equally judged “acceptable” or “not worthy” (37%). Ratings of the gold subscription were the highest during wet years. This only applied to both reliable and sharp forecasts since the silver subscription received the highest proportion of “not worthy” ratings. Participants who chose the default subscription possibly adopted risk-averse behaviors, while those choosing the silver subscription might have taken risks with inadequate sharpness levels.

The role of forecast performance in decision-making

Overall, forecasts of high reliability and sharpness were linked with an increase in decisions taken. Decisions based on sharp and reliable forecasts yielded higher gains and allowed participants to successfully go beyond simple risk-averse decision-making and take risks when adequate. In the absence of reliable and sharp forecasts, purely risk-averse decision-making was hard to beat. When looking at participants’ decisions, high reliability values were necessary but not sufficient to support risk-prone decisions (i.e., taking the risk to *sell surplus water*, Fig. 7), and both sharpness and reliability were required to observe almost half of the participants (44%) taking a risk. Moreover, participants did not take action based on forecasts with reliability levels lower than 60%. A similar outcome was drawn by Bruno Soares et al. (2018a), who noted that “reliability estimates below 50% were generally disregarded by the users interviewed.” However, when directly asked in the end survey, participants judged sharpness more useful than reliability. Participants might have had a more conscious understanding of the forecast range than of the reliability scale, even though, based on the analysis of game plays, reliability appeared to play a key role in the decisions. Additionally, the proximity of the forecasts to the decision thresholds, which is influenced by the forecast range, likely affected users’ attitude toward risk, as demonstrated

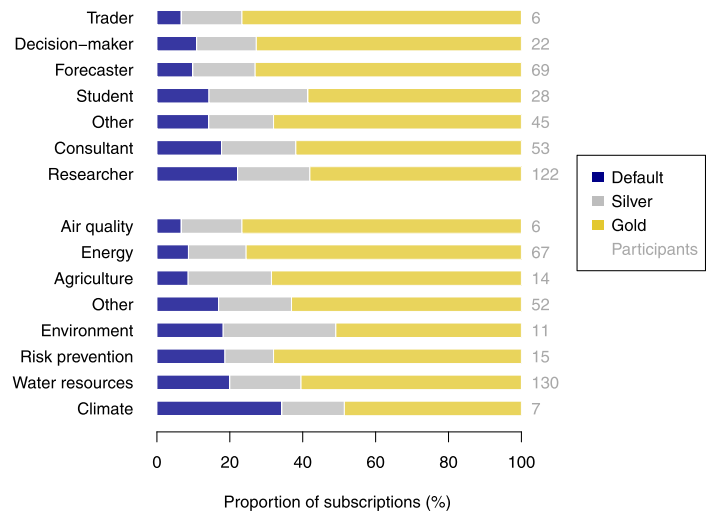


Fig. 9. Proportion of gold, silver, and default subscriptions (%) per (top) sector and (bottom) role in round 2.

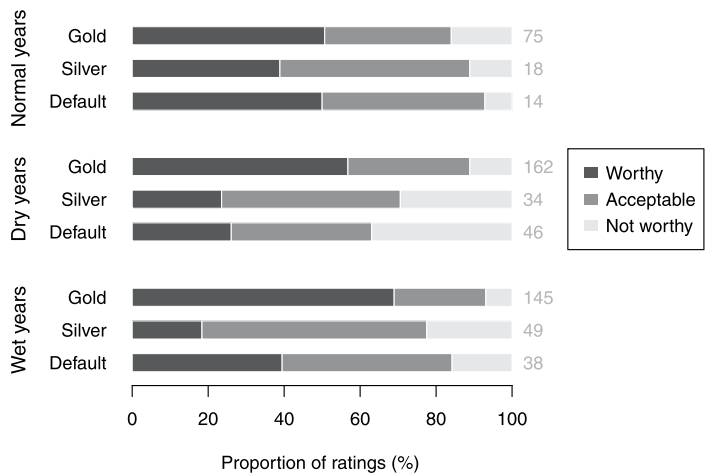


Fig. 10. Evaluation of the subscriptions in the second game round in normal, dry, and wet years. Light gray numbers indicate the number of ratings available for each subscription and year type.

by Ramos et al. (2013), possibly leading to this conclusion. Finally, the background of the participants influenced their choice in paying for forecasts, and complemented forecast performance in their game outcomes.

The feedback collected in the game answers and the end survey showed that many participants grew risk averse as the game progressed, for example, after experiencing failure. The notions of forecast performance and uncertainty were taken more seriously, as commented by two participants in the end survey: “If a forecast was too uncertain, I called my neighbor and did not take risks like I did in the first round,” and “when I knew my subscription had lower reliability I was more conservative and defensive with my choices.” Improved forecasts fostered risk-prone behaviors; a participant stated “I became less risk averse when I started paying for the best forecasts.” Participants may have felt a sense of security when paying for better forecasts, allowing them to move their focus from securing water supply to managing their budget. Some participants mentioned that better forecasts enabled earlier decisions in the season, also reflecting a shift toward risk-prone behaviors. However, others noted that taking risk-prone decisions when they paid for improved forecasts led to poor decisions nonetheless.

The expected value of forecasts

Each participant had a unique background and experience, and consequently, a personal perception of forecast value. Overall, participants from the water resources and climate sectors, as well as consultants and researchers, invested less than other participants. Participants from the energy, air quality, and agriculture sectors invested the most, along with traders, decision-makers, and forecasters. Participants from the energy sector may have come into the game with a positive attitude and a certain familiarity toward incorporating forecasts in the decision-making process, since energy companies already tend to invest in such services, often having in-house forecasting departments (Cassagnole et al. 2021; Foster et al. 2018; Torralba et al. 2017). Traders, forecasters, and decision-makers may also have been more comfortable in interpreting forecasts and incorporating them in their decision process, while also recognizing the added value of forecasts of good quality. These results are consistent with results from Neumann et al. (2018) who observed that water resources managers were more likely to choose a “wait and see” approach, while forecasters were more likely to seek further information and take action. Several participants never subscribed to improved forecasts and adopted a “safe” behavior instead, i.e., *call neighbors* as early as possible without taking the risk to *sell surplus water*. Others, including some climate researchers, started the game with a negative perception of seasonal forecasts. Statements like “I am not confident in seasonal precipitation forecasts so not sure I took consideration of the forecasts enough” indicate that some had in mind the skill of atmospheric seasonal forecasts, which rarely extends beyond 2 weeks to 1 month (Greuell et al. 2019; Buizza and Leutbecher 2015). This observation is also in line with the findings of Önköl and Bolger (2004), who found that forecast providers gave less credibility to their own forecasts than forecast users did. Other participants did not use the forecast information and solely based their decisions on the last observed reservoir level, which they found adequate.

Improved forecasts were judged more favorably in extreme years. As one participant stated: “Subscriptions are perhaps more suitable for offensive strategy.” In dry years, their cost was small compared to the cost of not taking the right action. In wet years, the ratings of improved forecasts were the highest, most likely because they enabled risk-prone decisions leading to gains compensating or exceeding the subscription cost. In years with normal conditions, default forecasts were the most valued, most likely because they were free of charge and because of the absence of potential gain in normal years. However, whether conditions would be extreme or normal was not foreseeable by the participants when they decided whether to use subscriptions.

Therefore, some participants mentioned that they decided never to change the subscription because they wanted “to learn to work with the same forecast system.”

The game as a learning tool for hydroclimate services

One goal of the game is to build knowledge and allow potential service users to become familiar with the displayed forecasts, uncertainty, and skill information through the service platform itself. Such tools could help services toward becoming stand-alone products by giving users access to knowledge and not only data. Moreover, the game can contribute to service codevelopment and foster discussions between service developers and users around suitable forecast performance, but also around data visualization. This was the case during the first Call for Water session organized at a Multiuser Forum of the Climate forecast enable knowledge services (CLARA) project. The forum gathered developers and users of 14 hydro-climate services who were invited to exchange and share their expertise in order to improve the services’ cogeneration process and outcomes.

Designing the game was a balance between comprehensiveness (to collect in-depth results on decision-making), and simplicity (to have a game accessible to a wide variety of audiences and sectors). Following retrospectively the game evaluation framework proposed by Aubert et al. (2019), the paper version offered a more immersive experience and more levity than the online version, because the narration was given by a game facilitator making the experience livelier. Motivational affordance was low since the game rewarded good decisions (i.e., to supply the town with water) with an absence of fine. Action–consequence feedback was fairly high though since participants saw the results of their actions each year. Finally, the level of difficulty was not tailored to the level of knowledge of participants; an exception was made for the EU Research and Innovation days for which the game was simplified to reach a broader audience including children. Adapting game complexity to participants’ knowledge might be a path to follow when integrating games into hydroclimate services in order to facilitate learning and broaden the audience.

Limitations and ideas for future work

The concepts involved in Call for Water are broader than its storyline portraying a water management case informed by seasonal reservoir forecasts. The game was played by participants from a wide variety of sectors, and hence could be used in various fields of environmental sciences involving probabilistic forecasts. Nevertheless, the storyline, game rules and their implied cost-loss relationship provide an artificial decision-making environment, which frames the validity of the presented results. For instance, the participants’ budget may have been an additional factor influencing attitudes toward taking risks (Arnal et al. 2016; Ramos et al. 2013). Moreover, because results were collected through a game with no real consequences for the participants, there could be a bias toward risk-prone behaviors, with potentially inflated risk-taking, spending and volatility in behaviors. Game sessions on the occasion of the EU Research and Innovation days showed, for example, that children, who may project themselves more easily in the role-playing aspect of the game, were significantly more risk averse than adults who sometimes played around with the subscriptions. These limitations linked with the artificial context of the game environment (also referred to as the magic circle of play; Huizinga 1949) impact the capacity for participants to transfer the knowledge gained through the game to real-life contexts (Aubert et al. 2019) and limit the potential generalization of the results.

While Call for Water always displays the same combination of years, further research on forecast perception could benefit from online serious games that allow the generation of randomized forecasts in a defined performance, event likelihood, and uncertainty space. Randomization would enable games to remain short while covering all possible combinations of investigated parameters. In the current version of Call for Water, participants base their decisions on reservoir volume forecasts and on the reliability and sharpness forecast attributes. Future work could

investigate decision-making when participants are directly presented with impact-based forecasts (e.g., economic) instead of forecasts of physical variables. Future work could also explore different forecast quality attributes such as bias, accuracy, and, in particular, discrimination that allows assessing the declared forecast probabilities with respect to critical thresholds (e.g., based on the relative operating characteristic, the Brier score, or the contingency table; Jolliffe and Stephenson 2003). Finally, the interpretation of decisions contradicted the declared usefulness of the forecast performance information. This highlights the need for comprehensive and user-friendly definition of the forecast performance attributes and metrics communicated both effectively and in a scientifically sound manner. To address this, future efforts could focus on the reliability and sharpness concepts, while introducing a controlling process could help monitor the participants' knowledge and understanding of the terms before and after the game.

Conclusions

This paper highlights how serious games can support and complement hydroclimate services by providing training to relevant users and exemplifies how games can be designed to answer research questions around decision-making, user behavior and judgement. Answers from the Call for Water serious game on forecast-based decision-making enabled analyses of the role of seasonal forecast performance in the controlled Call for Water decision-making context.

This work shows evidence that, in this controlled game environment, improved sharpness and reliability led to better decisions, enabling participants to step out of purely conservative strategies. Most participants tended not to base decisions on forecasts with reliability levels lower than 60%, and risk-prone decisions required reliability levels above 70%. Participants from the energy, air quality, and agriculture sectors, as well as traders, decision-makers and forecasters invested the most in forecasts, while participants from the climate and water resources sectors, as well as consultants and researchers, invested the least. Finally, participants tended to pay for reliable and sharp forecasts, which were judged more useful in years requiring extreme decisions.

Following these conclusions, some recommendations for the evolution of hydroclimate services can be made. First, when developing forecast-based services, assessing and communicating forecast performance are crucial steps in the uptake of the service information for decision-making. Forecast reliability levels higher than 60% are needed for informed decision-making (i.e., forecasts historically successful 6 times out 10). For sectors involving risk-prone decisions, reliability levels higher than 70% are needed and both reliability and sharpness are required. Finally, efforts should focus on building decision-making expertise, for instance, via training and serious games such as Call for Water. The perception of the forecast value was better in years requiring actions, which should be considered when developing pilot services and testing forecasts with users.

Acknowledgments. We thank all game participants that took part in this experiment and all volunteers who agreed to play in front of different audiences. We would especially like to thank Matteo Giuliani for organizing the game session at Politecnico di Milano and Micha Werner and Patricia Trambauer for organizing the Delft-FEWS session. This research has been conducted with support from the European Union's Horizon 2020 projects Climate forecast enabled knowledge services (CLARA) under Grant Agreement 730482 and S2S4E (Subseasonal to seasonal forecasting for the energy sector) under Grant Agreement 776787. This game is also part of the Hydrologic Ensemble Prediction Experiment (HEPEX).

Data availability statement. The Call for Water game material is freely available under the license conditions Creative Commons CC-BY-NC-ND 4.0. All the collected answers analyzed in this manuscript can be made available upon request.

Appendix A: Forecast performance metrics

The sharpness and reliability metrics ($V_{\text{sharpness}}$ and $V_{\text{reliability}}$, respectively) used to fabricate the forecasts and build the sharpness and reliability scales are defined as

$$V_{\text{sharpness}} = \frac{1}{N} \sum_{i=1}^N s_i,$$

$$s_t = \begin{cases} 100 \left[1 - \frac{\max(x_{\text{fst},i}) - \min(x_{\text{fst},i})}{\max(x_{\text{clim},i}) - \min(x_{\text{clim},i})} \right], & \text{if } \frac{\max(x_{\text{fst},i}) - \min(x_{\text{fst},i})}{\max(x_{\text{clim},i}) - \min(x_{\text{clim},i})} < 1, \\ 0, & \text{otherwise,} \end{cases}$$

$$V_{\text{reliability}} = \frac{1}{N} \sum_{i=1}^N r_i,$$

$$r_i = \begin{cases} 1, & \text{if } \min(x_{\text{fst},i}) < x_{\text{obs},i} < \max(x_{\text{fst},i}), \\ 0, & \text{otherwise,} \end{cases}$$

where N is the number of events, $x_{\text{fst},i}$ is the forecast ensemble at time step i , $x_{\text{obs},i}$ is the observation, and $x_{\text{clim},i}$ is the climatological ensemble. Both $V_{\text{sharpness}}$ and $V_{\text{reliability}}$ range between 0 and 100, with 100 being their optimum value, i.e., in the case of sharp and perfectly reliable forecasts.

Appendix B: Announced performance in fabricated forecasts

The announced sharpness and reliability drove the construction of the artificial forecasts, with the objective not to give a false sense of the metrics to the participants metrics (e.g., forecasts with a reliability of 90% should not always be wrong). In the case of sharpness, the boxplots extend from narrow intervals close to deterministic forecasts (100% on the scale) to intervals equal or wider than climatology (0% on the scale). In round 1, the width of the boxplots covers a breadth of values (Fig. B1), while in round 2, the width of the boxplots depends on the subscription. In the case of reliability, forecasts were designed so that, over one game round, the announced reliability was coherent with the a priori outcome. More details on a priori outcomes in round 1 (Fig. B1) and round 2 depending on the subscription (Table B1) are provided below.

Finally, the forecast sample was too small (15 in each round and 2 to 4 for each reliability level) to ensure a strict correspondence between the announced reliability and the actual outcomes. Exact estimates of reliability (or corresponding outcomes) in round 1 likely had an influence on whether participants paid for forecasts in round 2 but had little influence otherwise.

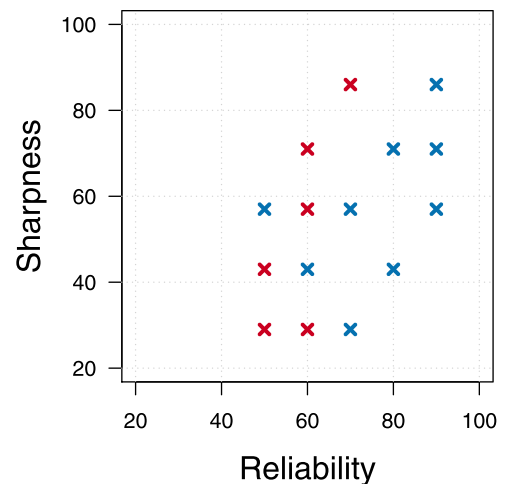


Fig. B1. Combinations of sharpness and reliability considered in the first game round. Blue (red) crosses indicate the forecasts that captured (missed) the June outcome.

Table B1. Number of cases when the forecasts captured the June outcome depending on the lead month, overall reliability, and announced reliability for the three subscriptions in round 2.

	June forecast issued in March	June forecast issued in April	June forecast issued in May	Total	Announced reliability
Gold forecasts	4/5 (80%)	4/5 (80%)	5/5 (100%)	13/15 (86.7%)	75%–100%
Silver forecasts	3/5 (60%)	4/5 (80%)	5/5 (80%)	12/15 (80%)	75%–100%
Default forecasts	2/5 (40%)	3/5 (60%)	3/5 (60%)	8/15 (53.3%)	50%–75%

References

- Arnal, L., M.-H. Ramos, E. Coughlan, H. L. Cloke, E. Stephens, F. Wetterhall, S.-J. van Andel, and F. Pappenberger, 2016: Willingness-to-pay for a probabilistic flood forecast: A risk-based decision-making game. *Hydrol. Earth Syst. Sci.*, **20**, 3109–3128, <https://doi.org/10.5194/hess-20-3109-2016>.
- , H. L. Cloke, E. Stephens, F. Wetterhall, C. Prudhomme, J. Neumann, B. Krzeminski, and F. Pappenberger, 2018: Skilful seasonal forecasts of stream-flow over Europe? *Hydrol. Earth Syst. Sci.*, **22**, 2057–2072, <https://doi.org/10.5194/hess-22-2057-2018>.
- Aubert, A. H., R. Bauer, and J. Lienert, 2018: A review of water-related serious games to specify use in environmental multi-criteria decision analysis. *Environ. Modell. Software*, **105**, 64–78, <https://doi.org/10.1016/j.envsoft.2018.03.023>.
- , W. Medema, and E. J. A. Wals, 2019: Towards a framework for designing and assessing game-based approaches for sustainable water governance. *Water*, **11**, 869, <https://doi.org/10.3390/w11040869>.
- Baker, S. A., A. W. Wood, and B. Rajagopalan, 2020: Application of postprocessing to watershed-scale subseasonal climate forecasts over the contiguous United States. *J. Hydrometeor.*, **21**, 971–987, <https://doi.org/10.1175/JHM-D-19-0155.1>.
- Bruno Soares, M., and S. Dessai, 2015: Exploring the use of seasonal climate forecasts in Europe through expert elicitation. *Climate Risk Manage.*, **10**, 8–16, <https://doi.org/10.1016/j.crm.2015.07.001>.
- , and —, 2016: Barriers and enablers to the use of seasonal climate forecasts amongst organisations in Europe. *Climatic Change*, **137**, 89–103, <https://doi.org/10.1007/s10584-016-1671-8>.
- , M. Alexander, and S. Dessai, 2018a: Sectoral use of climate information in Europe: A synoptic overview. *Climate Serv.*, **9**, 5–20, <https://doi.org/10.1016/j.cliser.2017.06.001>.
- , M. Daly, and S. Dessai, 2018b: Assessing the value of seasonal climate forecasts for decision-making. *Wiley Interdiscip. Rev.: Climate Change*, **9**, e523, <https://doi.org/10.1002/wcc.523>.
- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, <https://doi.org/10.1002/qj.2619>.
- Buontempo, C., and Coauthors, 2018: What have we learnt from EUPORIAS climate service prototypes? *Climate Serv.*, **9**, 21–32, <https://doi.org/10.1016/j.cliser.2017.06.003>.
- Caird-Daley, A. K., D. Harris, K. Bessell, and M. Lowe, 2007: Training decision making using serious games. Human Factors Integration Defence Technology Centre Rep. HFIDTC/2/WP4, 66 pp.
- Cassagnole, M., M.-H. Ramos, I. Zalachori, G. Thirel, R. Garçon, J. Gailhard, and T. Ouillon, 2021: Impact of the quality of hydrological forecasts on the management and revenue of hydroelectric reservoirs—A conceptual approach. *Hydrol. Earth Syst. Sci.*, **25**, 1033–1052, <https://doi.org/10.5194/hess-25-1033-2021>.
- Cloke, H. L., and F. Pappenberger, 2009: Ensemble flood forecasting: A review. *J. Hydrol.*, **375**, 613–626, <https://doi.org/10.1016/j.jhydrol.2009.06.005>.
- Coelho, C. A. S., and S. M. S. Costa, 2010: Challenges for integrating seasonal climate forecasts in user applications. *Curr. Opin. Environ. Sustain.*, **2**, 317–325, <https://doi.org/10.1016/j.cosust.2010.09.002>.
- Contreras, E., J. Herrero, L. Crochemore, I. Pechlivanidis, C. Photiadou, C. Aguilar, and M. J. Polo, 2020: Advances in the definition of needs and specifications for a climate service tool aimed at small hydropower plants' operation and management. *Energies*, **13**, 1827, <https://doi.org/10.3390/en13071827>.
- Crochemore, L., M.-H. Ramos, F. Pappenberger, S.-J. van Andel, and A. W. Wood, 2016: An experiment on risk-based decision-making in water management using monthly probabilistic forecasts. *Bull. Amer. Meteor. Soc.*, **97**, 541–551, <https://doi.org/10.1175/BAMS-D-14-00270.1>.
- , —, and I. G. Pechlivanidis, 2020: Can continental models convey useful seasonal hydrologic information at the catchment scale? *Water Resour. Res.*, **56**, e2019WR025700, <https://doi.org/10.1029/2019WR025700>.
- Flood, S., N. A. Craddock-Henry, P. Blackett, and P. Edwards, 2018: Adaptive and interactive climate futures: Systematic review of 'serious games' for engagement and decision-making. *Environ. Res. Lett.*, **13**, 063005, <https://doi.org/10.1088/1748-9326/aac1c6>.
- Foster, K., C. Bertacchi Uvo, and J. Olsson, 2018: The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers. *Hydrol. Earth Syst. Sci.*, **22**, 2953–2970, <https://doi.org/10.5194/hess-22-2953-2018>.
- Girons Lopez, M., L. Crochemore, and I. G. Pechlivanidis, 2021: Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden. *Hydrol. Earth Syst. Sci.*, **25**, 1189–1209, <https://doi.org/10.5194/hess-25-1189-2021>.
- Giuliani, M., L. Crochemore, I. Pechlivanidis, and A. Castelletti, 2020: From skill to value: Isolating the influence of end-user behaviour on seasonal forecast assessment. *Hydrol. Earth Syst. Sci.*, **24**, 5891–5902, <https://doi.org/10.5194/hess-24-5891-2020>.
- Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc.*, **69**, 243–268, <https://doi.org/10.1111/j.1467-9868.2007.00587.x>.
- Greuell, W., W. H. P. Franssen, and R. W. A. Hutjes, 2019: Seasonal streamflow forecasts for Europe—Part 2: Sources of skill. *Hydrol. Earth Syst. Sci.*, **23**, 371–391, <https://doi.org/10.5194/hess-23-371-2019>.
- Hartmann, H. C., T. C. Pagano, S. Sorooshian, and R. Bales, 2002: Confidence builders: Evaluating seasonal climate forecasts from user perspectives. *Bull. Amer. Meteor. Soc.*, **83**, 683–698, [https://doi.org/10.1175/1520-0477\(2002\)083<0683:CBESCF>2.3.CO;2](https://doi.org/10.1175/1520-0477(2002)083<0683:CBESCF>2.3.CO;2).
- Hewitt, C., C. Buontempo, P. Newton, F. Doblas-Reyes, K. Jochumsen, and D. Quadfasel, 2017: Climate observations, climate modeling, and climate services. *Bull. Amer. Meteor. Soc.*, **98**, 1503–1506, <https://doi.org/10.1175/BAMS-D-17-0012.1>.
- Hov, Ø., D. Terblanche, G. Carmichael, S. Jones, P. M. Ruti, and O. Tarasova, 2017: Five priorities for weather and climate research. *Nature*, **552**, 168–170, <https://doi.org/10.1038/d41586-017-08463-3>.
- Huizinga, J., 1949: *Homo Ludens: A Study of the Play-Element in Culture*. Routledge and Kegan Paul, 219 pp.
- Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley and Sons, 240 pp.
- Joslyn, S., and S. Savelli, 2010: Communicating forecast uncertainty: Public perception of weather forecast uncertainty. *Meteor. Appl.*, **17**, 180–195, <https://doi.org/10.1002/met.190>.
- , L. Nadav-Greenberg, and R. M. Nichols, 2009: Probability of precipitation: Assessment and enhancement of end-user understanding. *Bull. Amer. Meteor. Soc.*, **90**, 185–194, <https://doi.org/10.1175/2008BAMS2509.1>.
- Lavers, D. A., and Coauthors, 2020: A vision for hydrological prediction. *Atmosphere*, **11**, 237, <https://doi.org/10.3390/atmos11030237>.
- LeClerc, J., and S. Joslyn, 2012: Odds ratio forecasts increase precautionary action for extreme weather events. *Wea. Climate Soc.*, **4**, 263–270, <https://doi.org/10.1175/WCAS-D-12-00013.1>.
- Lucatero, D., H. Madsen, J. C. Refsgaard, J. Kidmose, and K. H. Jensen, 2018: On the skill of raw and post-processed ensemble seasonal meteorological forecasts in Denmark. *Hydrol. Earth Syst. Sci.*, **22**, 6591–6609, <https://doi.org/10.5194/hess-22-6591-2018>.
- Macian-Sorribes, H., I. Pechlivanidis, L. Crochemore, and M. Pulido-Velazquez, 2020: Fuzzy postprocessing to advance the quality of continental seasonal hydrological forecasts for river basin management. *J. Hydrometeor.*, **21**, 2375–2389, <https://doi.org/10.1175/JHM-D-19-0266.1>.
- Mendler de Suarez, J., and Coauthors, 2012: Games for a new climate: Experiencing the complexity of future risks. Boston University Frederick S. Pardee Center for the Study of the Longer-Range Future Rep. 978-1-936727-06-3, 119 pp., <https://scienceimpact.mit.edu/games-new-climate-experiencing-complexity-future-risks>.

- Morss, R. E., J. L. Demuth, and J. K. Lazo, 2008: Communicating uncertainty in weather forecasts: A survey of the U.S. public. *Wea. Forecasting*, **23**, 974–991, <https://doi.org/10.1175/2008WAF2007088.1>.
- Musuuzza, J. L., D. Gustafsson, R. Pimentel, L. Crochemore, and I. Pechlivanidis, 2020: Impact of satellite and in situ data assimilation on hydrological predictions. *Remote Sens.*, **12**, 811, <https://doi.org/10.3390/rs12050811>.
- Neumann, J. L., L. Arnal, R. E. Emerton, H. Griffith, S. Hyslop, S. Theofanidi, and H. L. Cloke, 2018: Can seasonal hydrological forecasts inform local decisions and actions? A decision-making activity. *Geosci. Commun.*, **1**, 35–57, <https://doi.org/10.5194/gc-1-35-2018>.
- Nkiaka, E., and Coauthors, 2019: Identifying user needs for weather and climate services to enhance resilience to climate shocks in sub-Saharan Africa. *Environ. Res. Lett.*, **14**, 123003, <https://doi.org/10.1088/1748-9326/ab4dfe>.
- Önköl, D., and F. Bolger, 2004: Provider–user differences in perceived usefulness of forecasting formats. *Omega*, **32**, 31–39, <https://doi.org/10.1016/j.omega.2003.09.007>.
- Pechlivanidis, I. G., L. Crochemore, J. Rosberg, and T. Bosshard, 2020: What are the key drivers controlling the quality of seasonal streamflow forecasts? *Water Resour. Res.*, **56**, e2019WR026987, <https://doi.org/10.1029/2019WR026987>.
- Peñuela, A., C. Hutton, and F. Pianosi, 2020: Assessing the value of seasonal hydrological forecasts for improving water resource management: Insights from a pilot application in the UK. *Hydrol. Earth Syst. Sci.*, **24**, 6059–6073, <https://doi.org/10.5194/hess-24-6059-2020>.
- Ramos, M.-H., T. Mathevet, J. Thielen, and F. Pappenberger, 2010: Communicating uncertainty in hydro-meteorological forecasts: Mission impossible? *Meteor. Appl.*, **17**, 223–235, <https://doi.org/10.1002/met.202>.
- , S. J. van Andel, and F. Pappenberger, 2013: Do probabilistic forecasts lead to better decisions? *Hydrol. Earth Syst. Sci.*, **17**, 2219–2232, <https://doi.org/10.5194/hess-17-2219-2013>.
- Rembold, F., and Coauthors, 2019: ASAP: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. *Agric. Syst.*, **168**, 247–257, <https://doi.org/10.1016/j.agsy.2018.07.002>.
- Samaniego, L., and Coauthors, 2019: Hydrological forecasts and projections for improved decision-making in the water sector in Europe. *Bull. Amer. Meteor. Soc.*, **100**, 2451–2472, <https://doi.org/10.1175/BAMS-D-17-0274.1>.
- Savic, A. D., S. M. Morley, and M. Khoury, 2016: Serious gaming for water systems planning and management. *Water*, **8**, 456, <https://doi.org/10.3390/w8100456>.
- Stephens, E. M., D. J. Spiegelhalter, K. Mylne, and M. Harrison, 2019: The Met Office weather game: Investigating how different methods for presenting probabilistic weather forecasts influence decision-making. *Geosci. Commun.*, **2**, 101–116, <https://doi.org/10.5194/gc-2-101-2019>.
- Street, R. B., C. Buontempo, J. Mysiak, E. Karali, M. Pulquério, V. Murray, and R. Swart, 2019: How could climate services support disaster risk reduction in the 21st century. *Int. J. Disaster Risk Reduct.*, **34**, 28–33, <https://doi.org/10.1016/j.ijdr.2018.12.001>.
- Sutanto, S. J., H. A. J. Van Lanen, F. Wetterhall, and X. Llort, 2019: Potential of pan-European seasonal hydrometeorological drought forecasts obtained from a multihazard early warning system. *Bull. Amer. Meteor. Soc.*, **101**, E368–E393, <https://doi.org/10.1175/BAMS-D-18-0196.1>.
- Terrado, M., N. Gonzalez-Reviriego, L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, 2017: Climate services for affordable wind energy. *WMO Bull.*, **66**, 48–53.
- , and Coauthors, 2019: The Weather Roulette: A game to communicate the usefulness of probabilistic climate predictions. *Bull. Amer. Meteor. Soc.*, **100**, 1909–1921, <https://doi.org/10.1175/BAMS-D-18-0214.1>.
- Torralba, V., F. J. Doblas-Reyes, D. MacLeod, I. Christel, and M. Davis, 2017: Seasonal climate prediction: A new source of information for the management of wind energy resources. *J. Appl. Meteor. Climatol.*, **56**, 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1>.
- Trocchi, A., 2018: *Weather & Climate Services for the Energy Industry*. Springer International Publishing, 197 pp.
- Vaughan, C., and S. Dessai, 2014: Climate services for society: Origins, institutional arrangements, and design elements for an evaluation framework. *Wiley Interdiscip. Rev.: Climate Change*, **5**, 587–603, <https://doi.org/10.1002/wcc.290>.
- , J. Hansen, P. Roudier, P. Watkiss, and E. Carr, 2019: Evaluating agricultural weather and climate services in Africa: Evidence, methods, and a learning agenda. *Wiley Interdiscip. Rev.: Climate Change*, **10**, e586, <https://doi.org/10.1002/wcc.586>.
- Yuan, X., E. F. Wood, and Z. Ma, 2015: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development. *Wiley Interdiscip. Rev.: Water*, **2**, 523–536, <https://doi.org/10.1002/wat2.1088>.