Climate Services 28 (2022) 100335

Contents lists available at ScienceDirect

Climate Services

journal homepage: www.elsevier.com/locate/cliser

Co-evaluation of climate services. A case study for hydropower generation

E. Delpiazzo [a,b,c,*], F. Bosello [a,b,c], P. Mazzoli [d], S. Bagli [d], V. Luzzi [d], F. Dalla Valle [e]

[a] CMCC@CàFoscari, Venice, Italy
[b] Department of Environmental Sciences, Informatics and Statistics University Ca' Foscari of Venice, Venice, Italy
[c] RFF-CMCC European Institute on Economics and the Environment, Venice, Italy
[d] GECOsistema S.r.l., R&D Unit, Bozen, Italy
[e] Enel Green Power Italia S.r.l. Hydrology and Hydraulic Analysis, Italy

ARTICLE INFO

Keywords:
Climate services
Value of information
Decision theory
Economic value of climate services

ABSTRACT

Climate services are attracting growing attention and interest as instruments to promote climate change adaptation. The transparent assessment of the potential value brought by the services can play a major role. It can foster the commitment of the user towards a co-generation process increasingly central to climate services creation, can provide developers important information to better tailor the service to the user needs, and can finally increase recognition of the value of the service boosting confidence and trust in the tool.

This study presents and then demonstrates the applicability of an evaluation methodology based on the Bayesian framework derived from the information value theory. The specific case study is the Smart Climate Hydropower Tool (SCHT), a climate service designed to support management decisions in hydropower generation. The service uses freely available seasonal forecasts and machine learning algorithms to predict incoming discharge to hydropower reservoirs. The user is ENEL Green Power Italy, and the testing environments are two water basins in Colombia.

The study defines the expected value of perfect information, the expected value of the information currently used by the hydropower producer and the expected value of the service information. It then discusses pros and cons of the applicability of the method.

1. Practical implications

The "partial" or "relative" nature of the evaluation process and the related caveats will be highlighted in the methodological part. This section discusses some further challenges and opportunities emerged during the evaluation phase that are worth considering when similar exercises are conducted.

Definition of the payoffs. The procedure is based on the possibility to evaluate the gains for the user enabled by the availability of different information sets. This requires that the whole process, from the acquisition of the information to the decisions, and the identification of the outcomes of the decisions, are clearly identifiable. This is not always possible: the decision process can be not well structured or defined, and/or the user may find it difficult to isolate the role of the information in leading to the result. Furthermore, in the case of private companies, gains expressed in terms of economic performances can be sensitive data hard to share and disclose. The only way to address these difficulties is to engage in an effective co-design-creation process with the user. The user has anyway the best knowledge of the environment in which operations take place to indicate the payoff measure which is more informative and adherent to the reality. This however requires a non-negligible level of commitment in reciprocal listening and learning by the evaluator, and the user of the service.

Definition of the decision process. Often, the decision process enabled by the availability of service-generated information and that currently followed by the user are quite different. So, it may be difficult for the user to clearly identify what could be effectively done when the service is accessible and compare the two situations. This is particularly true when decisions are complex with many possible "states of the world" and uncertainty sources. Nonetheless, the reconstruction of the alternative decision processes is crucial. Once again, the collaboration between the evaluators and users of the service is the only way to address the complexity, reduce it to a level that is manageable, but still sufficient to produce an informative evaluation. Interestingly, during the

* Corresponding author at: Centro Euro-Mediterraneo sui Cambiamenti Climatici, Università Ca' Foscari Venezia, RFF-CMCC European Institute on Economics and the Environment (EIEE), Edificio Porta dell'Innovazione, Via della Libertà 12- 30175, Venezia Marghera, Venice, Italia.

E-mail address: elisa.delpiazzo@cmcc.it (E. Delpiazzo).

https://doi.org/10.1016/j.cliser.2022.100335
Received 11 May 2021; Received in revised form 25 August 2022; Accepted 13 October 2022
Available online 1 November 2022

exchange, the user, forced to think in a structured way to the decision process, can sometimes get a better understanding and awareness about the procedures usually implemented and can find room for improvements.

Definition of the application domain. Climate services are rather new and not yet of widespread use in many contexts. This lack of practice may originate gaps between the functionalities the service develops and their direct usability. Similarly, potential users of the service may lack awareness about the full range of its applications. In the present context, for instance, the initial focus is on the enabled efficiency gains in hydropower production. Benefits from the service can however be extended to a broader context, for instance, considering reduced costs from "wrong" production prediction when buying and selling contracts to dispatch agreed energy volumes are concluded. The difficulty to define the application domain of the service may neglect important sources of value. Once again, a structured thinking of the decision process "from the cradle to the grave" is the solution to minimize this risk.

The role of the service producer. A proactive involvement by the service producer is fundamental for a successful evaluation. In the present case, the availability of technical competence for the hindcasting exercise and the determination of the skill of the service were necessary for the implementation of the whole procedure. Participating to the co-evaluation process, although engaging, can offer important opportunities to the service producer. He can gain important information to better tailor the service to the user needs, understand where to invest in improvement, and, sometimes, get a better awareness of the real relevance of the product it delivers. The interaction with the users can improve trust and facilitate the uptake of the service.

2. Introduction

Major international agendas like, for instance, the 2030 Agenda for Sustainable Development, the Paris Agreement, the Sendai Framework, call explicitly for communities, critical infrastructure, businesses, and ecosystems more resilient to climate impacts (World Meteorological Organization (WMO) (2017); UNFCCC, 2017). Since the launch of the Global Framework for Climate Services (GFCS) in September 2009, there has been a growing interest in promoting the use of climate services in support of climate risk management in many areas: agriculture, disaster risk reduction, health, water management, energy, and a variety of other climate-vulnerable sectors (Vaughan and Dessai, 2014; World Meteorological Organization (WMO) (2016); Street et al., 2019; Hewitt et al., 2020). The demand for products delivering that information will continue to grow in the years to come driven by both concerns over climate change and the occurrence of extreme events such as heat waves, storms, flooding, and drought. It will also reflect the need to respond to new human-induced vulnerabilities such as the growth of megacities and coastal developments (Lugen, 2016; World Meteorological Organization (WMO) (2016)). Nonetheless, the GFCS mission to strengthen the production, availability, delivery and application of science-based climate prediction and services is not yet accomplished. There has been little progress in providing evidence on the value added from tailored climate information once conveyed to users-decision makers (Tall et al., 2018; Vaughan et al. 2018). World Meteorological Organization (WMO) (2015) is a milestone in climate service evaluation with its compendium of methods used to obtain the end value of climate and weather information for the end user. Nonetheless, monitoring and evaluation of the value of specific services remains a significant challenge.

Climate services are decision support tools, based on a process of transforming climate-related information (i.e., science based) into advisory services that assist decision-making by individuals and organizations of a society (i.e., user- specific) (Pope et al., 2017). They assist policymakers and decision-makers operating in climate-sensitive sectors to take practical actions based on the best available climate-related information (from climatic as well as other relevant scientific and socio-economic research). In this way, climate services can also help society to become more resilient and to cope with the growing impacts of climate change. Practically, climate services have been defined in multiple ways (Hewitt, Mason and Walland, 2012; Perrels et al., 2013; Vaughan and Dessai, 2014) but in a broad categorization we could identify operational climate services, which support short-term operations, and "adaptation climate service", that with a longer time horizon, supports adaptation planning and strategies. Here, with the term "climate service" we mainly refer to seasonal climate service, as we tailor our assessment on a service of this type.

Clearly identifying benefits for final users and profits for the providers is fundamental to support the development of climate services and to mainstream their utilization. An inclusive, flexible, and collaborative evaluation procedure is particularly decisive to support the co-generation process (Vincent et al., 2018), increasingly central to climate services creation, where developers and users collaborate to maximize services utility and uptake in investment and decision planning (Clements et al., 2013; von Flotow and Ludolph, 2013). During a collaborative evaluation process the developers can better understand how their services enter in the user decision process and which of the services' features are the most relevant to the generation of value. This provides important information to better tailor the service to the user needs reducing the usability gap (Lemos et al., 2012). The user, on its turn, by directly participating and understanding how the assessment evolves, increases its awareness and recognition of the value of the service boosting confidence and trust in the tool.

In what follows, we apply a Bayesian framework derived from the information value theory (Winkler et al., 1983; Wilks, 2014) to the evaluation of climate services. The methodology appears quite flexible to be applied for services in different sectors. In this vein, for instance, Hamlet and Huppert (2002) evaluated that the use of long-lead stream flow forecasts in the management of hydroelectric dams on the Columbia River could increase energy production by 5.2 million MWh per year, resulting in a US$153 million increase in net revenues. Meza and Wilks (2004) estimated the value of perfect sea-surface temperature anomaly forecasts for fertilizer management for potato farmers in Chile to be between $5 and $22 per hectare, compared to a no forecast context. Berrocal et al. (2010) found that the use of probabilistic weather forecasts for predicting ice conditions reduced costs for the Washington State Department of Transportation by 50 %. However, as surveyed in Bruno et al. (2018), there is a variety of methodologies to assess the value of climate services, from ex-ante to ex-post, and applied to a variety of sectors (agriculture, energy, water management, and transportation). Related to hydropower generation, there are studies applying simulation models. Maurer and Lettenmaier (2004) demonstrated that that use of climate forecast information can improve the hydropower production with a maximum achievable benefit of $25.7 million, and $6.8 million when realistic streamflow predictability is used. Block (2011) found that using forecasts to manage hydropower operations in the upper Blue Nile basin (Ethiopia) produces cumulative decadal benefits between $1 and $6.5 billion, compared to a climatological based approach. Other studies focused on cost models, such as Graham et al. (2022), who assessed the value added in Scotland of 7-day forecasts ranging between £ 2.20/MWh and £ 1.40/MWh, and between £ 2.70/MWh and £ 1.10/MWh for 2-week forecasts with different time-lead (2 and 6 weeks, respectively). Analysis of the value added for the hydropower sector in Colombia could be found in Poveda et al. (2003), although the retrospective analysis is based on a rather old period 1977–1992 and associated to the incorporation of ENSO in the forecasts. In this case, they suggested savings in operating costs ranging between 35 % and 40 % for the hydroelectric plant at Guatapé.

Currently, however, most of the assessment of the economic benefit of integrating climate services into decision making processes focuses on the agricultural sector in developing and least developed countries (Vaughan et al., 2019a; Vaughan et al., 2019b). The need to gather funding to establish operational climate services forces the providers to demonstrate their potential value added.

Here we implement the methodology to quantify the value of a hydropower generation related climate service: the "Smart Hydropower Climate Tool", hereto SCHT (www.https://gecosistema.com/climate-tools/scht-smart-climate-hydropower-tool/ and Essenfelder et al., 2020) examined within the CLARA H2020 project (www.clara-project.eu). Our purpose is to demonstrate the viability of the method, suggest a procedure that can be replicable in different contexts and eventually discuss its pros and cons.

The methodology enables us to extract two values for the service: the "maximum potential value" and the "effective expected value". These two notions play a role at different stages of the co-generation phase. The "maximum potential value" intervenes at the beginning of service development. It represents the gain a hypothetical perfect forecast may convey to the specific user. It is a sort of benchmark that indicates the potential contribution of the service as a production factor in the user production process. The "effective expected value" stands at the end of the service production process and gives the final user's specific value.

The paper is structured as follows. Section 3 illustrates the methodology and introduces the case study. Section 4 shows and discusses the results of the evaluation. Section 5 concludes.

3. Material and methods

3.1. The conceptual framework

The methodology applied for the evaluation of the economic benefit of SCHT in this pilot case study is a Bayesian probabilistic framework that compares the value of alternative information sets (with and without climate service) in the context of decision making. Here, not only the ability of the information to convey the exact forecast matters, but also the "direction" of the possible mistake. Indeed, from the point of view of the end-user, underestimating or overestimating the frequency or magnitude of an event has different consequences in terms of decisions and payoff. The expected value of the service information, which coincides with the value of the service, is computed in relative terms. Once, the gains or payoff from the gain-maximizing (or cost-minimizing) action associated with each information source are determined, the expected value of the information is obtained by comparing the expected payoff of using the climate service against the alternative business-as-usual knowledge. The added value of the service information is given by the difference of the two (for a mathematical description of the methodology see Appendix A).

In the application of the methodology the collaborative approach among the actors involved, namely evaluator, service provider, and end user, is fundamental both to improve the evaluation itself and in gathering the input data needed. Their roles interconnects at different layers, as Fig. 1 shows.

The analysis develops along different phases. The first collects time series on effective realizations, climatological and climate service-based forecasts of inflows that could feed the simulation model and be translated into states of the world. In this phase, statistics on the forecasts enable the computation of the skill of the alternative knowledge sources. This information is primarily produced by the climate service providers and the end user. Secondly, a simulation model translates the forecasts into energy production according to the "technical" knowledge of the reservoir features and its management rules. This is the fundamental contribution provided by the end user and, at the same time, private and sensitive data that could be not disclosed plainly and simply. Thirdly, the end user also provides the information on actions that would be taken on the basis of the prediction and their skill. Fourthly, each combination of action and probability of the forecasts to fail is associated to a payoff. The core work of the evaluator is finally to apply the Bayesian framework using all the information acquired.

3.2. The case study

The case study concentrates on the SCHT experimental service developed to explore added value of seasonal forecasts for ENEL Green Power (EGP) hydropower producer in two reservoirs in Colombia: Betania and Guavio feeding ENEL power plants.

All the information has been collected through interaction between service developers and EGP. This includes collecting monthly river discharge records to the reservoirs and setting up an efficient management and production scheme, based on the hydropower plant characteristics and regulation capabilities. A detailed description of the SCHT service is presented in Appendix B.

Information on water volumes is necessary to plan the energy production. For each reservoir, the manager seeks to maximize the produced energy supposing it has a constant level during the whole month. Furthermore, for each catchment the manager should consider the next month incoming water volume (forecasted), the volume at the beginning of present month, the ecological runoff, and finally decide how much energy generate, according to each plant production scheme, to keep the reservoir storage in a reasonable working range at the end of the month.

A simple simulation model for production has been setup with EGP to translate forecast information of incoming discharge in operational decisions ("actions") on how much energy to produce, according to the technical characteristics of the plant (such as reservoir storage curves, number and efficiency of the turbine groups) and optimal operational management rules. Feeding this production model with different forecasts generates different operational decision, and subsequently different production values and reservoir volumes at the end of the forecast period ("states of the world").

In practice, we first identify the possible states of the world. In this

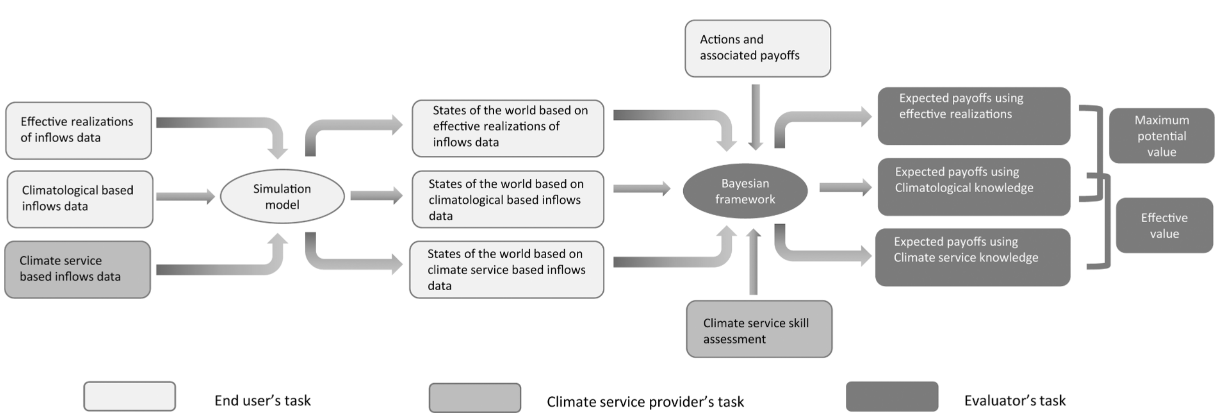

Fig. 1. Workflow of the methodology and role of the key actors.

application, they are possible water volumes in each basin that can fall under three cases: the water volume is inside (*I*), above (*A*) or below (*B*) the “normal range”.

The second step is to identify actions. These are three different production/pumping regimes decided monthly, that depend upon the specific plant regulation capabilities and the expected reservoir water volume at the end of the incoming month. The operational rules respond to “water thresholds” that have been defined by EGP basing on historical values of production.

Operational rules definition entails in-depth knowledge of the plant-reservoir system including commercially sensitive technical and financial information. For this reason, ENEL directly took care of setting up the rules. For the purposes of the present paper the management rules could be briefly described in:

- Considering the forecast incoming discharge in the incoming period.
- Targeting to keep the reservoir in a “normal” range between a maximum and a minimum acceptable value, in this case 10 % and 90 % of reservoir exploitable volume (such optimal thresholds have been identified by EGP basing on simulations of productions in the period 2001/2010).
- Activating one or more different production groups to use all available discharge according to predictions, while keeping the expected reservoir level at the end of the forecast period inside the range (this decision once taken cannot be reversed until the end of the forecast period). Regulation capabilities of a real plant are still limited and minimum flow (corresponding to one single group running) and maximum flow (all groups running) may lead to impossibility of achieving the desired optimal result.
- Checking “a posteriori” with real observed values if the target has been reached or not and moving to the next forecast period.

Finally, the payoff matrix in Table 1 reports the gain the energy producer gets given the combination action (rows) and effective realization of the state of the world (columns). Clearly there are actions that are the best match for each state of the world. These are indicated in the main diagonal of Table 1 that reports the highest payoffs. Reading the cells by rows, out of diagonal, indicates payoff associated to the “wrong” actions. Note that cell entries can be also interpreted in terms of payoff induced by right or wrong predictions (or perfect, higher, lower skill of the information/service).

Since an explicit monetary indication of the payoff is function of the decision-making process and it is not already in place as the service is not applied yet, values are substituted by a more generic “performance indicator index” ranging from 0 to 10, that anyway correctly reflects the payoff ranking across decisions. Furthermore, the performance relates to energy potentially produced and not to “money” directly. The payoff matrix is the same for both reservoirs.

Assigned payoffs corresponds to the following qualitative judgments discussed with EGP on possible consequences and related feasible countermeasures of acting upon forecast knowledge:

- In case of correct forecast (effective realization correspond to the forecast, i.e. values along diagonal of the matrix), higher values are assigned to the better situation, lower ones to the less appealing. Such values are still higher than those assigned in case of incorrect forecast, given the chance of setting up countermeasures (i.e., financial energy buy/sell operations to balance shortage of production) for less than optimal, but correctly predicted, conditions.
- In case of incorrect forecast worst conditions (reservoir out of optimal range in the exact opposite situation than what has been forecasted) 0 points have been assigned given the arguable difficulty in compensating such unexpected event with reasonable countermeasures. Predicting to be in the range (row 1 in the table) and ending with more water in the reservoir is less dramatic than the opposite (leading for example to waste of water for reservoir overtopping instead of water scarcity and impossibility to match energy demand). In the same way (Column “I” in the table), forecasting shortage and ending with more resource is still more favourable than the other way round.

Table 1
Payoff Matrix.

	Effective realization *I*	Effective realization *A*	Effective realization *B*
Action I according to prediction *I*	10	5	3
Action A according to prediction *A*	3	8	0
Action B according to prediction *B*	5	0	6

The exercise then consists in assessing three different values: the “maximum potential expected value” of the climate service, corresponding to a perfect forecast able to correctly predict water volumes in the two basins; the “effective expected[1] value” of (the information provided by) SCHT based on its estimated skill, and the “effective expected value” of an alternative information set. This last is the observed climatic trend over the 30-year period. This is the standard information set that the energy producer uses to form predictions on water volumes in the basins. Evaluations are based on a sample from 2000 to 2019; the period 2000–2016 is the “training set”. That means the machine learning has been fed with historic observations of that period and trained to replicate them. The period 2017–2019 is the test set. Thus, for the economic evaluation the training set is the period where the skill of the service is computed, while we consider the test set as the effective evaluation period. This avoids an “overfitting issue” (i.e. model optimally performing in the training and poorly in the test), that could generate a high total value biased by the training set, although such a service should have a zero value. Eventually, the sample to assess the economic potential and effective value of SCHT is limited to 36 monthly observations.

Data for the assessment thus include the effective probabilities of the states of the world in the reference period, the skills of the service and of the historic based knowledge. This last coincides with that of the climatic trends. The skill of the service is determined by values from the training set period.

Table 2 shows the expected performance (or skill) of the historic

Table 2
Skill of the historic based knowledge in Betania and Guavio reservoirs.

	–	Effective realizations		
		I	*A*	*B*
Betania reservoir				
Historic based knowledge predictions	*I*	0.41	0.00	0.00
	A	0.21	*1.00*	0.00
	B	0.38	0.00	*1.00*
Guavio reservoir				
Historic based knowledge predictions	*I*	0.63	0.00	0.00
	A	0.15	*1.00*	0.00
	B	0.21	0.00	*1.00*

[1] The possibility to observe what effectively happened in the 2017–2019 period enables also to test the payoff performance that would have been enabled in practice by the service. In this case we would shift from an expectation to a fully deterministic context. It can thus occur that in a specific time frame, a service that performs better than an alternative information set “on average”, in fact performs worse. For completeness, we report the results of this additional analysis in the supplementary material.

Table 3
Skills of SCHT hindcasted values in Betania and Guavio reservoirs.

		Effective realizations		
	–	*I*	*A*	*B*
Betania reservoir				
SCHT hindcasted values' predictions	*I*	0.60	0.08	0.00
	A	0.12	0.92	0.00
	B	0.28	0.00	*1.00*
Guavio reservoir				
SCHT hindcasted values' predictions	*I*	0.72	0.00	0.00
	A	0.12	*1.00*	0.00
	B	0.16	0.00	*1.00*

knowledge (past climate average) in predicting water volumes in Betania and Guavio reservoirs in 2017–19. In both reservoirs, when states of the world *A* and *B* occur, they are also predicted. Nonetheless, state *I* is correctly predicted 63 % of times in Guavio reservoir and only 41 % of the times in Betania.

Table 3 reports the expected performance (or skill) of the SCHT climate service. The prediction performance of state of the world *I* improves for both reservoirs (from 41 % to 60 % in Betania and from 63 % to 72 % in Guavio). Predictions of *A* and *B* remain the same in the Guavio basin. *A* predictability decreases slightly in the Betania basin.

Apparently high values along the diagonal for *A* and *B* cases (in italic in the previous tables) shall not be misunderstood, as, by construction of the optimization model, those are rare cases referring to extreme working conditions of the reservoirs (very small samples number). It is much more significative to evaluate results in column "*I*", corresponding to most of real observable situations.

4. Results and discussion

This section assesses the value of the three different information sets. The first, ideally, would enable perfect forecasts of water volumes in the basins in the test period. This would correspond to the value originated by "perfect information" or, in other words, by a service with 100 % skill. The second refers to the historical experience, that in the present exercise corresponds to the information set that the energy producer is using to form his forecasts and plan energy production in the absence of the climate service. The third is what expected by SCHT.

4.1. The added value of perfect information in 2017–2019

Fig. 2 reports the payoffs originated by a perfectly forecasting climate service (the blue lines), that by history-based forecasts (the red lines), and the difference of the two showing the added value of the former against the latter (the grey bars). By construction, the former is also the maximum value an information set can provide, of course against an history-based knowledge with the specific skill reported in Tables 2 and 3. As already evident examining these tables, the history-based knowledge is a quite a good predictor of states *A* and *B* in both basins and rather good of *I* in Guavio. This reservoir is indeed characterized by a low variability in water volumes in the 3-year period considered. Thus, gains from perfect information mostly depend upon the ability to predict *I* and can be expected to be higher in the more volatile Betania reservoir.

On a yearly average basis, the improvement in the performance index enabled by perfect information is 4.42 % in the Guavio basin and 5.96 % in the Betania reservoir. The last is mostly concentrated in 2017, a year where the *I* state occurred more frequently and the production performance could have been increased by 9.04 %.

4.2. SCHT expected value in 2017–2019

Fig. 3 adds to Fig. 2 the expected performance enabled by the SCHT service. In the Guavio reservoir, in the specific period considered, the expected value of the service information almost coincides with that of the history-based forecast. This may seem counterintuitive given that the service better predicts state of the word *I*. However, the gain for the energy producer in adopting the action associated to *I* 72 % rather than

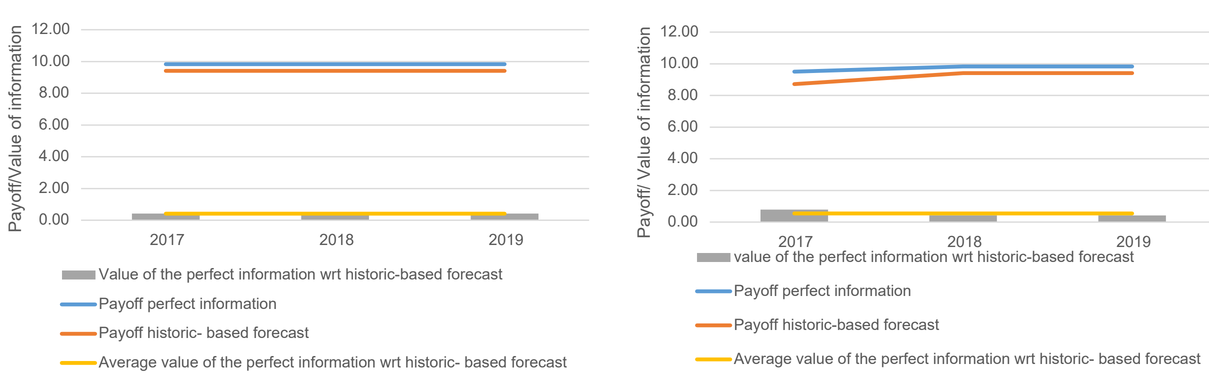

Fig. 2. Yearly expected value of information: payoffs from historical- based forecast and perfect information in Guavio (left) and Betania (right).

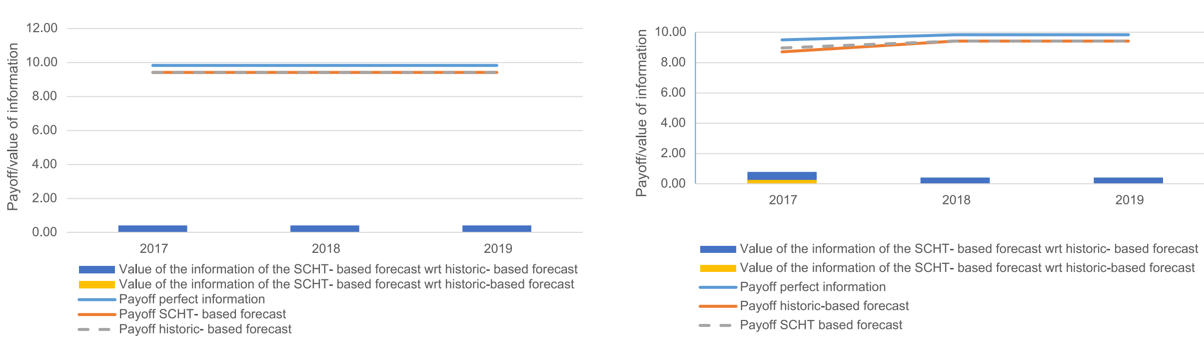

Fig. 3. Yearly expected value of information and maximum potential value: payoffs from historical- based forecast and perfect information in Guavio (left) and Betania (right).

63 % of times of I occurring, is negatively compensated by the service suggestions of A and B 12 % and 16 % of times respectively in the presence of I rather than the 15 % and 21 % suggested by the historic experience. In other words, according to the payoff matrix, when the service is "wrong" mistakes are more costly than when the history-based experience is "wrong".[2] In determining this outcome, a role is played by the stability of the water volumes over time that makes the past a good predictor of the future.

Conversely, the more volatile water volumes of Betania reservoir originate a larger expected value of service information. This is mostly due to the SCHT predictive performance in 2017 where the service enables a production performance index 3 % larger than that of the historic-based information (or just 4.7 % lower than that enabled by a perfect information).

These results confirm that there is a positive expected gain and thus value in using forecast provided by SCHT respect to use the climatological mean-based information. The utility of the service clearly magnifies in those situations of "higher" variability, like in the Betania reservoir in our case study, where "the past" or "experience" cannot be used as a good predictor for the future. Considering that the evaluation is heavily dependent upon the choice of the test period of three years only, and that the climatic variability is very likely to increase, there are good reasons to believe that, over time, the service can offer a valuable contribution to improve the energy producer performance even though we cannot present a real economic value of the service, but just order of magnitudes in payoff improvements. Moving from these results to an economic quantification is quite straightforward when data on gains and losses in monetary terms are available.

5. Conclusions

This paper presents one practical application of the value of information theory, to the estimation of the value of a specific climate service. The cases study is the "Smart Hydropower Climate Tool" (SCHT) designed to support with seasonal forecasts hydropower producers. The value of the service is tested on two Colombian water reservoirs and catchments in the period 2017–2019 in collaboration with ENEL Green Power company.

The case study demonstrates that SCHT service has a positive expected value against what currently used by the hydropower producer. It would allow a 3 % increase in the expected production performance in the Betania basin, while in Guavio, given the intrinsic characteristics of the reservoir, the two forecasting systems are practically equivalent. Results suggest that increasing variability in hydrological conditions due to climate change should generate a higher value of the service. This is implicitly evident comparing the two reservoirs: Betania – higher service value – has a higher variability in its responses to hydrological conditions, Guavio – lower service value – has a more stable behavior because of the characteristics of the basin- reservoir combination. They could be examples of "high variability conditions" and a "low variability conditions", respectively.

The methodology is particularly useful as it also highlights the maximum gains that the service can potentially generate. This information is useful either to the producer or the user of the service. The former can use it to understand where it would be more efficient to concentrate effort to improve the service performance, and better meet the user needs, the latter can get an immediate measure of the added value the service can produce and a transparent description of its functioning.

The methodological steps suggested can be rather easily extended to different contexts and in relation to different reservoirs. Indeed, the methodology could be applied in case of multi-objective reservoirs with more than one constraint and/or generating benefits for many end users.

Nonetheless, the viability of the evaluation method crucially depends upon the proactive engagement of the users, the producer, and the evaluators of the service. On the one hand this is challenging. It requires a non-negligible investment of time, open minded thinking, and availability to share transparently information. On the other hand, it can produce benefits that go well beyond the evaluation itself improving trust and uptake of the service.

CRediT authorship contribution statement

E. Delpiazzo: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **F. Bosello:** Conceptualization, Methodology, Writing – review & editing. **P. Mazzoli:** Conceptualization, Data curation, Software, Writing – review & editing. **S. Bagli:** Conceptualization, Data curation, Software. **V. Luzzi:** Conceptualization, Data curation, Software. **F. Dalla Valle:** Conceptualization, Data curation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge funding from the European Union's Horizon 2020 research and innovation programme for the CLARA project under grant agreement no. 730482. Valuable and constructive comments by two anonymous reviewers were greatly acknowledged.

Appendix A. : Mathematical formulation

The mathematical explanation of the methodology is presented using a simplified example (similar to Murphy, 1993; Katz and Murphy, 1997), where a decision maker (the climate service user) faces a state space X which summarizes mutually exclusive future states of the world. To simplify these are x_1 and x_2, occurring with probabilities $p(x_1)$ and $p(x_2) = 1 - p(x_1)$, respectively.

Differently from standard examples when the value of perfect and imperfect information is assessed, it is assumed that the true values of $p(x_1)$ and $p(x_2)$ are not known to the decision maker[3].

She anyway faces a decision space A, where only two options are available: A_1 and A_2. Each of them is associated to x_1 or x_2 in a payoff matrix

[2] As an example, in Guavio when the state of the world I is predicted by the historic-based forecast, the expected payoff related is 17.40 while the same combination using SCHT gives an expected payoff of 19.73. Nonetheless the historic-based forecast suggests more often than SCHT to take the "wrong" production decision related B when I occurs that, albeit being wrong, gives a higher payoff than the still wrong production decision related to A when I occurs.

[3] In the standard framework, the value of perfect information is associated to the payoff the decision maker can get resolving in advance the uncertainty about the future states of the world, while endowed with perfect knowledge of the probabilities of states.

(Table A1). Payoffs, π_n, are "exits", function of the states of the world and actions $\pi_n(A_n, x_n)$. Typically, on the main diagonal of the payoff matrix the "best" or "correct" combination of action and event is reported.

Table A1
2×2 Payoff (cost/loss) matrix

	Event x_1	Event x_2
Action A_1	$\pi(A_1, x_1)$	$\pi(A_1, x_2)$
Action A_2	$\pi(A_2, x_1)$	$\pi(A_2, x_2)$

A rational decision maker decides the action that maximizes her payoff (minimizes costs or maximizes revenues) given her knowledge, in our example referring to probabilities $p(x_1)$ and $p(x_2)$. This on its turn can derive from some previous information set (y_1) already available to the decision maker, such as for instance experience from past climatology or past observation, or from information (y_2), brought by a climate service. Both information sets are not able to perfectly predict future states of the world and can induce some mistakes.

In our set up, the value of the information sets y_1 and y_2 coincides with the expected gains each can bring to the user. Said differently, y_1 and y_2, the pre-existing knowledge and "climate service knowledge", will suggest different actions to the decision maker, with different expected payoffs.

These are higher the better prediction of the true state of the world the information sets allow. The ability to predict, or the skill of the information sources, can be computed retrospectively comparing predictions with the effective weather/climate realizations in a given time period.

Then, if the skills stay constant, and associating the payoff of the course of action that would have been suggested by y_1 and y_2, it is possible to assess the expected added value of one against the other[4].

Furthermore, it is also possible to use the value associated to (originated by) the knowledge of effective weather/climate realization as the benchmark that defines what we call the Expected Value of Perfect Information *EVPI*. It represents the case when a knowledge source always correctly predicts the occurrence of the uncertain events and enables, accordingly, the best response. In the framework of Table A1, *EVPI* is computed summing the products of the minimum loss times the frequency of the event:

$$EVPI = \sum\nolimits_{n=1}^{2} min_n p(x_n)\pi(A_n, x_n) \tag{A1}$$

Being based upon the best available knowledge, *EVPI* also represents the maximum value an information set can originate. The performance of the alternative information sets depends on their skill, or, in other words, on the ability of the respective forecasts (hereto X_1 and X_2)to correctly predict the events. The skills of each information set can be represented by a contingency matrix (Table A2) where its performance is reported in terms of numbers N of correct forecasts, misses and false alarms.

Table A2
Contingency matrix.

	Event x_1	Event x_2
Forecast X_1	$N_{1,1}$	$N_{1,2}$
Forecast X_2	$N_{2,1}$	$N_{2,2}$

From the contingency matrix it is possible to derive the skill in predicting any state of the world x_1 , x_2 as "conditional probability", here is the "Bayesian" part of the method, from two data: the real frequency of an event (N_n/N) and the frequencies of the forecasts ($N_{n,m}/N_n$). From Table A2, there are four conditional probabilities:

$$P_n(x_n|X_n) = \frac{N_{n,m}/N}{\left(\left(\sum_{n=1}^{2} N_{n,m}\right)/N\right)} \tag{A2}$$

(A2) expresses the probability the information set y_n correctly predicts (predicted) x_n, or tells us how many times x_n effectively occurs when it is forecasted by y_n.

In this illustrative case, the two sources of knowledge y_1 and y_2 (say accumulated experience observing historical climate trend, and the climate service) originate specific conditional probabilities, $P_1(x_n|X_n)$ and $P_2(x_n|X_n)$, respectively. Both convey "imperfect information" as some mistakes in prediction are possible.

Denote as $p_{y_1}(x_1)$ and $p_{y_1}(x_2)$ the predicted frequency of the events based on historical record, hence the Expected Value of Historical Information (*EVHI*) is:

$$EVHI_{y_1} = \sum\nolimits_{n=1}^{2} p_{y1,n}(X_n) min_A \sum\nolimits_{n=1}^{2} \pi(A_n, x_n) P_{y1,n}(x_n|X_n) \tag{A3}$$

With $p_{y_2}(x_1)$ and $p_{y_2}(x_2)$ the predicted frequency of the events based on the climate service.

The Expected Value of the Climate Service information (*EVCS*) is:

$$EVCS_{y_2} = \sum\nolimits_{n=1}^{2} p_{y2,n}(X_n) min_A \sum\nolimits_{n=1}^{2} \pi(A_n, x_n) P_{y2,n}(x_n|X_n) \tag{A4}$$

Equations A3 and A4 state that the value of the information associated with y_1 and y_2 depends upon the fixed payoff, the probabilities p_{y_n}, and a third factor, the conditional probability, that measures the skill of the information, accounting for its "imperfection".

[4] In this setting the value of information can be assessed only in relative terms comparing if, how and with what consequences "new" or "different" information change the behavior of the information recipient. This requires a comparison with what she is doing which is on its turn determined by a pre-existing information set or knowledge. Thus, having just one information set would not enable the evaluation process. At least one alternative needs to be specified.

The value of the information embedded in the historic-based knowledge and in climate services can be also expressed in terms of difference with the Expected Value of Perfect Information (equations A5 and A6 respectively).

$$ValueHistoricInformation = EVHI - EVPI \tag{A5}$$

$$ValueClimateServiceInformation = EVCS - EVPI \tag{A6}$$

The correct interpretation of the results from the application of this method requires some disclaimers. Firstly, it is focused on quantifying the value of information provided by “the climate service”. The cost of producing this information, whether and how much a potential user is willing to pay for it is not pertinent for this analysis. Secondly, the evaluation of the service is “user-focused”. Thus, it can vary according to her/his characteristics. Some of these, although individual, can be “objective” such as what she/he already knows the gains and losses stemming from the different decisions, while other can be subjective, such as different degrees of risk aversions. Even though we do not consider subjectivity, our evaluation remains relative and not absolute. Thirdly, by the same token, the evaluation does not measure the total value of the service but refers to the value that the service can originate for one or a group of identified users in each time period. Fourthly, this assessment does not consider the feedback, or second order effects triggered by the decisions the informed user takes. On the one hand, it neglects potential “imitation processes” by other users, on the other hand it assumes that the fact that potentially a large number of agents act according to the information received does not “rebound” on their payoff.

Appendix B. : Smart Climate Hydropower Tool

To produce SCHT forecasts, the service is based on two sets of information:

(i) *Seasonal meteorological forecasts* (monthly precipitation and temperature) were provided by the Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC) since the service was developed before the Copernicus Climate Data Store (CDS) became operative.
(ii) “*Seasonal forecast monthly statistics on single levels*” were downloaded from the dedicated CDS page[5]. It has been selected due to both availability by the time of this research, and time scale of interest. This “product” provides worldwide seasonal forecasts (from 1 to 6 months lead time) operated at forecast centers in several European countries. For this application, used data include 2m temperature (°K) and Total precipitation (m s-1) at the horizontal resolution of 1° x 1°. The data include forecasts created in real-time (since 2017) and retrospective forecasts (hindcasts) initialized at equivalent intervals during the period 1993-2016. Such forecasts use ensembles (51 members) to reflect a distribution of outcomes and provide statistical variability. Nonetheless, in the current application a deterministic output of discharge is requested. Accordingly, the ensemble forecast, or ensemble mean, has been used. For further details on used meteorological forecast see the dedicated CDS page. Forecast meteorological variables have been extracted as separate time series for each pixel inside every plant upstream catchment (15 to 20 pixels per catchment).

These inputs were used in the Machine Learning algorithm as depicted in Fig. B1. Firstly, the forecast ML algorithms have been trained using as input features historical values of target variables, i.e. time series of monthly river discharge (up to the day of forecast) and seasonal monthly precipitation and temperature (P-T) hindcasts from Copernicus Climate Data store (CDS). This is possible as CDS provides forecasts in real-time (since

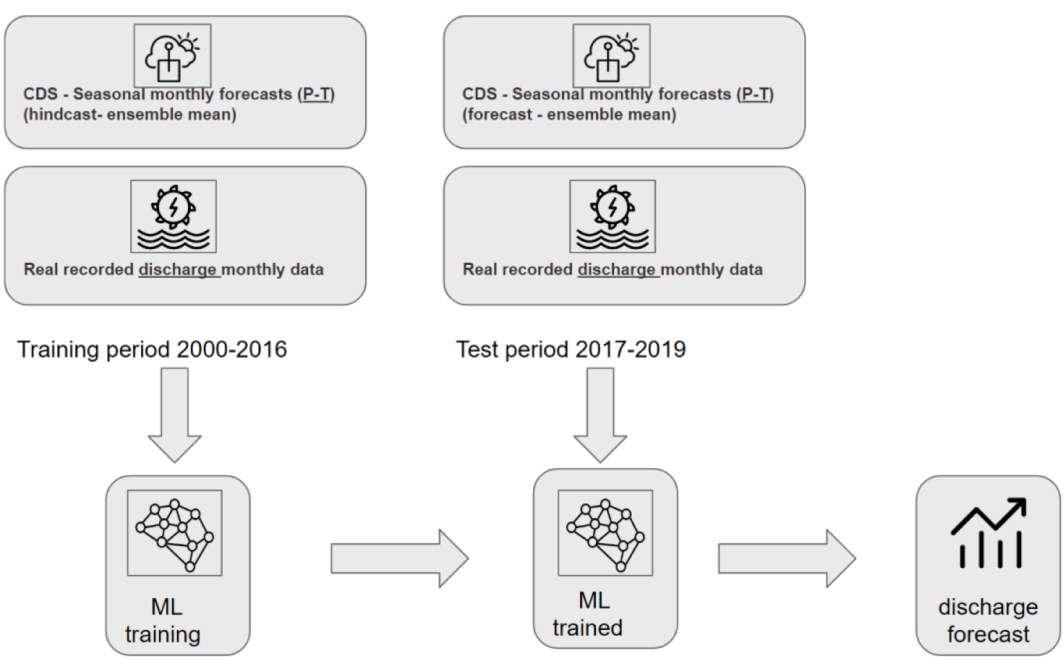

Fig. B1. Training – testing workflow for discharge forecasting.

[5] https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-monthly-single-levels?tab=overview

2017) and retrospective forecasts (hindcasts) initialized at equivalent intervals during the period 1993-2016.

Thus, the hindcast period 2000- 2016 is the "*training set*" used to train the machine learning algorithms. To complete the training phase, the time series of input features (hindcast of P-T for the incoming months for each of the 15-20 pixel inside the catchment extracted from CDS, plus discharge at catchment output up to the current month provided by ENEL) have been normalized and analyzed through AutoML algorithms (in this case using the ones available in H2O platform[6]) in order to rank them and select a subset of informative features (roughly half of the original input set). Then, multiple algorithms have been trained, using the same AutoML platform, on the train dataset and leaderboard on the separate test set to select best performing ones.

An ensemble of 3 to 4 best algorithms for each catchment, ranging from regressors to neural network families, is the final trained ML model used for setting up the forecast.

The forecast period 2017- 2019 is instead the "*test set*" used to get realistic performance of the trained forecast algorithm. This ML model, launched with the same input features, but from the test set period, provides the forecasts used for the proposed analysis.

No observed value of P or T has been used for training nor testing; and no further bias correction has been applied to forecast. The ML model has been trained on hindcasts (ensemble mean) and then launched using forecast (again ensemble mean).

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cliser.2022.100335.

References

Berrocal, V., Raftery, A., Gneiting, T., 2010. Probabilistic weather forecasting for winter road maintenance. Journal of the American Statistical Association 105 (490), 522–537. https://doi.org/10.1198/jasa.2009.ap07184.

Block, P., 2011. Tailoring seasonal climate forecasts for hydropower operations. Hydrology and Earth System Sciences 15 (4), 1355–1368. https://doi.org/10.5194/hess-15-1355-2011.

Bruno, S.M., Daly, M., Dessai, S., 2018. Assessing the value of seasonal climate forecasts for decision-making. WIREs Climate Change 9, e523. https://doi.org/10.1002/wcc.523.

Clements J., Ray A., Anderson G., 2013. *The Value of Climate Services Across Economic and Public Sectors: A Review of Relevant Literature*. United States Agency for International Development (USAID) Report. http://www.climate-services.org/wp-content/uploads/2015/09/CCRD-Climate-Services-Value Report_FINAL.pdf.

Essenfelder, A.H., Larosa, F., Mazzoli, P., Bagli, S., Broccoli, D., Luzzi, V., Mysiak, J., Mercogliano, P., dalla Valle F, 2020. Smart Climate Hydropower Tool: A Machine-Learning Seasonal Forecasting Climate Service to Support Cost-Benefit Analysis of Reservoir Management. Atmosphere 11 (12), 1305. https://doi.org/10.3390/atmos11121305.

Graham, R.M., Browell, J., Bertram, D., White, C.J., 2022. The application of sub-seasonal to seasonal (S2S) predictions for hydropower forecasting. Meteorological Applications 29 (1). https://doi.org/10.1002/met.2047.

Hamlet, A., Huppert, D., 2002. Economic value of long-lead streamflow forecasts for Columbia River hydropower. Journal of Water Resources Planning and Management 128 (2), 91–101. https://doi.org/10.1061/(ASCE)0733-9496(2002)128:2(91).

Hewitt, C.D., Allis, E., Mason, S.J., Muth, M., Pulwarty, R., Shumake-Guillemot, J., Bucher, A., Brunet, M., Fischer, A.M., Hama, A.M., Kolli, R.K., Lucio, F., Ndiaye, O., Tapia, B., 2020. Making Society Climate Resilient. International Progress under the Global Framework for Climate Services. Bulletin of the American Meteorological Society 101 (2), E237–E252. https://doi.org/10.1175/BAMS-D-18-0211.1.

Hewitt, C., Mason, S., Walland, D., 2012. The Global Framework for Climate Services. Nature Climate Change. Nature Research 2 (12), 831–832. https://doi.org/10.1038/nclimate1745.

Katz, R., Murphy, A., 1997. Forecast value: prototype decision making models. In: Katz, R., Murphy, A. (Eds.), Economic Value of Weather and Climate Forecasts. Cambridge.

Lemos, M.C., Kirchhoff, C.J., Ramprasad, V., 2012. Narrowing the climate information usability gap. Nature Climate Change 2, 789–794. https://doi.org/10.1038/nclimate1614.

Lugen M., 2016. *The role of climate services for adaptation to climate change in developing countries, with a case study from South Africa*. KLIMOS Working Paper No. 10. KLIMOS- ACROPOLIS, Brussels, Belgium.

Maurer, E.P., Lettenmaier, D.P., 2004. Potential Effects of Long-Lead Hydrologic Predictability on Missouri River Main-Stem Reservoirs. Journal of Climate 17 (1), 174–186. https://doi.org/10.1175/1520-0442(2004)017<0174:PEOLHP>2.0.CO;2.

Murphy, A.H., 1993. What is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. Weather and Forecasting 8, 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Perrels, A., Frei, T.h., Espejo, F., Jamin, L., Thomalla, A., 2013. Socio-economic benefits of weather and climate services in Europe. Advances in Science and Research 10, 65–70. https://doi.org/10.5194/asr-10-65-2013.

Pope, E.C.D., Buontempo, C., Economou, T., 2017. Quantifying how user-interaction can modify the perception of the value of climate information: A Bayesian approach. Climate Services 6, 41–47. https://doi.org/10.1016/j.cliser.2017.06.006.

Poveda, G., Mesa, O.J., Waylen, P.R., 2003. Nonlinear Forecasting of River Flows in Colombia based upon ENSO and its Associated Economic Value for Hydropower Generation. In: Diaz, H.F., Morehouse, B.J. (Eds.), Climate and Water. Advances in Global Change Research, vol 16. Springer, Dordrecht.

Street, R.B., Buontempo, C., Mysiak, J., Karali, E., Pulquério, M., Murray, V., Swart, R., 2019. How could climate services support disaster risk reduction in the 21st century. International Journal of Disaster Risk Reduction 34, 28–33. https://doi.org/10.1016/j.ijdrr.2018.12.001.

Tall, A., Coulibaly, J.Y., Diop, M., 2018. Do climate services make a difference? A review of evaluation methodologies and practices to assess the value of climate information services for farmers: Implications for Africa. Climate Services 11, 1–12. https://doi.org/10.1016/j.cliser.2018.06.001.

UNFCCC, 2017. *Opportunities and options for integrating climate change adaptation with the Sustainable Development Goals and the Sendai Framework for Disaster Risk Reduction 2015–2030*. https://unfccc.int/sites/default/files/resource/techpaper_adaptation.pdf.

Vaughan, K., Dessai, S., 2014. Climate services for society: origins, institutional arrangements, and design elements for an evaluation framework. WIREs Climate Change 5, 587–603. https://doi.org/10.1002/wcc.290.

Vaughan, K., Dessai, S., Hewitt, C., 2018. Surveying Climate Services: What Can We Learn from a Bird's-Eye View? Weather, Climate, and Society 10 (2), 373–395. https://doi.org/10.1175/WCAS-D-17-0030.1.

Vaughan, K., Hansen, J., Roudier, P., Watkiss, P., Carr, E., 2019a. Evaluating agricultural weather and climate services in Africa: Evidence, methods, and a learning agenda. WIREs Climate Change 10, e586. https://doi.org/10.1002/wcc.586.

Vaughan, K., Muth, M.F., Brown, D.P., 2019b. Evaluation of regional climate services: Learning from seasonal-scale examples across the Americas. Climate Services 15, 1–4. https://doi.org/10.1016/j.cliser.2019.100104.

Vincent, K., Daly, M., Scannell, C., Leathes, B., 2018. What can climate services learn from theory and practice of co-production? Climate Services 12, 48–58. https://doi.org/10.1016/j.cliser.2018.11.001.

von Flotow P. and Ludolph M., 2013. Climate Information as an Object of Economic Research: State and Perspective. Available at SSRN: https://ssrn.com/abstract=2423337 or http://dx.doi.org/10.2139/ssrn.2423337.

Wilks D., 2014. Probability Forecasting, Quantitative Decision Modeling, and the Economic Value of Forecasts. American Meteorological Society (AMS) Proceedings. https://ams.confex.com/ams/94Annual/webprogram/Paper232199.html.

Winkler, R.L., Murphy, A.H., Katz, R.W., 1983. The Value of Climate Information: A Decision-Analytical Approach. Journal of Climatology 3, 187–197. https://doi.org/10.1002/joc.3370030208.

World Meteorological Organization (WMO), 2015. *Valuing weather and climate: Economic assessment of meteorological and hydrological services*. Geneva, Switzerland. https://library.wmo.int/doc_num.php?explnum_id=3314.

World Meteorological Organization (WMO), 2016. *Climate Services for Supporting Climate Change Adaptation. Supplement to the Technical Guidelines for The National Adaptation Plan Process*. https://library.wmo.int/doc_num.php?explnum_id=7936.

World Meteorological Organization (WMO), 2017. White Paper on the Contribution of the Global Framework for Climate Services to Transforming our World: the 2030 Agenda for Sustainable Development (Agenda 2030). https://library.wmo.int/doc_num.php?explnum_id=3586.

[6] https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html