Delft University of Technology

How to evaluate a monitoring system for adaptive policies
criteria for signposts selection and their model-based evaluation

Raso, Luciano; Kwakkel, Jan; Timmermans, Jos; Panthou, Geremy

DOI
10.1007/s10584-018-2355-3

Publication date
2019

Document Version
Final published version

Published in
Climatic Change

Climatic Change (2019) 153: 267–283
https://doi.org/10.1007/s10584-018-2355-3

How to evaluate a monitoring system for adaptive policies: criteria for signposts selection and their model-based evaluation

Luciano Raso[1] · Jan Kwakkel[1] · Jos Timmermans[1] · Geremy Panthou[2]

Received: 21 December 2017 / Accepted: 19 December 2018 / Published online: 21 January 2019

Abstract

Adaptive policies have emerged as a valuable strategy for dealing with uncertainties by recognising the capacity of systems to adapt over time to new circumstances and surprises. The efficacy of adaptive policies hinges on detecting on-going change and ensuring that actions are indeed taken if and when necessary. This is operationalised by including a monitoring system composed of signposts and triggers in the design of the plan. A well-designed monitoring system is indispensable for the effective implementation of adaptive policies. Despite the importance of monitoring for adaptive policies, the present literature has not considered criteria enabling the a-priori evaluation of the efficacy of signposts. In this paper, we introduce criteria for the evaluation of individual signposts and the monitoring system as a whole. These criteria are relevance, observability, completeness, and parsimony. These criteria are intended to enhance the capacity to detect the need for adaptation in the presence of noisy and ambiguous observations of the real system. The criteria are identified from an analysis of the information chain, from system observations to policy success, focusing on how data becomes information. We illustrate how models, in particular, the combined use of stochastic and exploratory modelling can be used to assess individual signposts, and the whole monitoring system according to these criteria. This analysis provides significant insight into critical factors that may hinder learning from data. The proposed criteria are demonstrated using a hypothetical case, in which a monitoring system for a flood protection policy in the Niger River is designed and tested.

Keywords Monitoring · Climate change · Adaptive policies · Dynamic adaptive policy pathways · Signposts · Evidence based · Monitoring · Information · Flood management · Extremes · Deep uncertainty · Niger River

✉ Luciano Raso
l.raso@tudelft.nl

1 Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands

2 Université Grenoble-Alpes, 38 058 Grenoble Cedex 9, Grenoble, France

1 Introduction

Policy failures are often due to the omission of critical uncertainties when preparing the policy (Funtowicz and Ravetz 1990). Climate, environmental, and socioeconomic changes are largely uncertain in the long term, posing a challenge for planners: policies that would be satisfactory for one particular future may fail in many other futures. Adaptive policies can help policy-makers in designing policies which are robust in the face of uncertainty. In adaptive policies, a coherent long-term plan is assembled: this plan indicates how to adapt over time to different possible future system evolutions. The design of adaptive policies takes into account that uncertainty about the validity of critical assumptions affecting policy success will resolve over time, and decisions will be adapted in response to this. Advantages of adaptive policies are their capacity to value correctable (or scalable) decisions, modulate response to evidence of change, coordinate short- and long-term actions, and delay decisions to keep future options open (Hallegatte 2009; Lee 1994; Moser and Ekstrom 2010).

Various approaches for developing adaptive policies have been put forward, such as assumption-based planning (Dewar et al. 1993), dynamic adaptive policies (Kwakkel et al. 2010), real options (Hertzler 2007; Woodward et al. 2014; Jeuland and Whittington 2014; Blyth et al. 2007), adaptive policy-making (Walker et al. 2001; Hamarat et al. 2013), adaptation options (Wilby and Dessai 2010), adaptation tipping points, and adaptation pathways (Wise et al. 2014; Haasnoot et al. 2012), and dynamic adaptive policy pathways (Haasnoot et al. 2013). All these approaches use some type of signpost and trigger to identify when the policy must be adapted.

Detecting change and the consequent adaptation of a policy should not be slower than the change itself. The capacity to adapt on time, i.e. reacting sufficiently before the occurrence of negative consequences, depends on how much information will be available when the decision to adapt has to be taken. For this reason, an effective monitoring system is a pivotal element for the success of an adaptive policy. It has been suggested to base this monitoring system on "signposts" (Dewar et al. 1993) and "triggers" (Walker et al. 2001). Signposts specify "the information that should be tracked in order to determine whether the policy is meeting the conditions for its success", triggers specify "critical values of signpost variables beyond which additional pre-specified actions should be implemented" (Haasnoot et al. 2013; Walker et al. 2001; Kwakkel et al. 2010).

Different applications of signpost selection can be found in the literature. These are generally expert based, sometimes supported by scenario discovery (Bryant and Lempert 2010) or similar vulnerability analysis techniques. The literature, however, lacks a systematic approach to guide the analyst in designing a monitoring system for adaptive policies. Hamarat et al. (2013) select signposts on the basis of an extensive model-based vulnerability analysis using scenario discovery, while trigger values are based on expert opinion. In a follow-up study, Hamarat et al. (2014) and Kwakkel et al. (2016) fine-tune trigger values using many-objective robust optimisation. Zeff et al. (2016) embed the selecting of triggers within a broader many objective optimisation, and Haasnoot et al. (2015) present an in-depth analysis of developing signposts and triggers for climate adaptation for the Netherlands. In Herman and Giuliani (2018), signposts and triggers are identified jointly using optimisation. Lempert and Groves (2010) and Tariq et al. (2017) use signposts based on expert judgement. Ceres et al. (2017) investigate the level of confidence provided by the 100-year peak storm surge, used as signpost, in detecting climatic change.

In this paper, we introduce criteria for signposts evaluation, specifically: relevance, observability, completeness, and parsimony. We present how models, in particular, the use of stochastic models in an exploratory manner, can be employed for evaluating these criteria. The identification of these criteria is based on an analysis of the information chain, from system observations to policy success, and based on principles of information theory (Cover and Thomas 2012), focusing on how data becomes information. This analysis provides significant insight into critical factors that may hinder learning from data.

The paper is structured accordingly. In Section 2, we analyse the problem of monitoring in adaptive policies, we introduce criteria for selecting signposts and for testing the effectiveness of the whole monitoring system. In Section 3, we discuss the role of models in evaluating these criteria. In Section 4, the proposed criteria are used to define the monitoring system for an adaptive policy for flood protection. In Section 5, we present our conclusions.

2 The problem of monitoring in adaptive policies

In complex systems, learning from evidence can be difficult. As a result, learning can be weak and slow (Sterman 2006). This constitutes a barrier to adaptation (Moser and Ekstrom 2010). Information on specific variables or relations can be either ambiguous or noisy. Ambiguity arises from the presence of multiple valid interpretation of data. Noise arises from the presence of nonmeaningful information in the data. Delay in learning is exacerbated in case of monitoring for protection against extreme events: their rarity reduces the capacity to observe and identify them. For example, detecting change in flood risk is hampered by the large interannual variability and the scarcity of valuable data points. As a consequence, an ill-conceived monitoring system can jeopardise the efficacy of the entire adaptive policy.

The design of the monitoring system for an adaptive policy involves identifying the information that is necessary for adapting the policy in order to ensure its continued success over time. The design of a monitoring system happens in the context of a specific decision-making problem, in which the system, the set of possible actions, the critical uncertainties, and the criteria for the policy success have been already identified. In this case, the analyst can build up the monitoring system by defining the information on which the implementation of adaptive actions has to be based (Haasnoot et al. 2013; Walker et al. 2001; Kwakkel et al. 2010). When the information to be used in the monitoring system is not readily available, economic criteria should be used to gather the new information that has the higher marginal net benefit (Raso et al. 2018).

A monitoring system is effective if it is capable of *detecting relevant change on time.* “Relevant” refers to the possible influence on policy success, and “on time” refers to the moment at which a decision on adapting the policy must be taken. Ultimately, a planner is interested in identifying if and when to implement a new action. Taking a new action is required if the changing conditions threaten the necessary conditions for success of the policy. That is, policy success is no longer guaranteed. In the literature on adaptive planning, this situation is sometimes called the “adaptation tipping point” (Haasnoot et al. 2013). Related ideas are scenarios that illuminate vulnerabilities of a policy as identified by scenario discovery (Bryant and Lempert 2010) or through decision scaling (Brown et al. 2012). Designing an effective monitoring system involves the identification of signposts and triggers which enable the timely detection of relevant changes.

A signpost is a statistic, i.e. a rule that aggregates, transforms, or filters data in order to extract from it the relevant information. An example of a signpost, later utilised in the test case, is the 90th percentile of annual maximum discharge at a certain location over a window of 30 years.

We need to separate the object of observation from its meaning: using the terminology borrowed from probability theory (Dekking 2005), a signpost is the "parameter" that we track using our best "estimate" of its true value: the parameter is the quantity of interest that a signpost is intended to track; the estimate is the observed value obtained from data. Continuing on our example, if the parameter is the 10-year flood, i.e. the flood that is expected to return once every 10 years, one possible estimate of this parameter would be the 90th percentile of annual maximum discharge.

Figure 1 zooms in on the relationship between the signpost and the critical uncertainties. Critical uncertainties are conditions that are presently uncertain or subject to change in the future, which strongly affect the success of the policy. The signpost estimate, notwithstanding being a transformation of raw data, is still data. The signpost parameter is the link between the signpost estimate and the critical uncertainty. In this sense, the signpost parameter allows the signpost estimate to be the interpreted from, and assimilated into, the existing knowledge (Liu and Gupta 2007), hence be used to assess the state of the critical uncertainties.

Signpost parameters are only partially measurable: a signpost estimate contains both a signal about the parameter that we are interested in, and the noise that we want to ignore. The larger the noise the more difficult it will be to detect the signal. The noise must thus be properly "filtered" (Papoulis 1977) in order to retain the information about the signpost parameters. When filtering the data in order to extract the signal, one must bear in mind that it is not possible to extract more information than there is in the original raw data (Cover and Thomas 2006; Weijs 2011). Nonetheless, when synthesising information using a signpost estimate, the information extraction process must be kept under control in order to minimise information loss. This has important consequences on criteria for signposts selection, as it will be discussed in the following.

2.1 Signposts selection: relevance and observability

A signpost is to be selected according to its *informativeness*. Informativeness refers to the capacity of a signpost estimate to identify the value of critical uncertainties from data. Signpost informativeness can be decomposed into *relevance* and *observability*, as in shown in Eq. 1, and demonstrated in Appendix A, Eq. A.1.

$$\text{Informativeness} = \text{relevance} \times \text{observability} \tag{1}$$

Relevance is the capacity of a signpost parameter to track the critical uncertainties that it is intended to monitor. Observability is the capacity of a signpost to determine the signpost parameter from the signpost estimate. Observability can be further decomposed into *accuracy* and *precision*. Accuracy refers to the closeness of a signpost estimate to the true value of the parameter. Accuracy is a description of the systematic error. ISO calls this "trueness" (1994). Precision is a measure of the statistical variability, i.e. the noise present in the observations.

Fig. 1 Decomposition of the relationship between signpost and critical uncertainties, and role of stochastic and exploratory models in the evaluation of observability and relevance

If a critical uncertainty can be observed directly, then the signpost parameter is the critical uncertainty itself, directly tracked by the signpost estimate. Otherwise, the signpost is a proxy, i.e. its parameter is indirectly related to the critical uncertainty. In this case, the analyst must relate the first to the latter. Continuing the flood risk example, assuming that the 100-year flood is the critical uncertainty, then the 90th percentile of annual maximum discharge over a 30-year window is a proxy.

A signpost is rarely dominant on all criteria (i.e. relevance, accuracy, and precision). Instead, a typical signpost is a tradeoff among them. Consider, for example the selection of the appropriate window length in a moving window statistics. A larger window length will use more data points, resulting in more precision, but it will also be less reactive to change and, being biased, its accuracy will be lower. Generally speaking, the further removed a signpost is in the causal chain from the critical uncertainty, the lower the relevance of the signpost. Getting closer to the critical uncertainty, however, may result in a loss of observability due to a decrease in accuracy and/or precision. Using again the flood risk example, monitoring the 100-year flood directly would have the maximum relevance, but also an extremely low observability that would require a much lunger time window and which would likely fail to detect nonstationary change timely.

2.2 How many signposts? Completeness and parsimony

The success of an adaptive policy hinges almost always on multiple critical uncertainties; therefore, monitoring often requires more than one signpost. The number of signposts should be as small as possible (i.e. *parsimony*), conditional on the capacity of guaranteeing *completeness* of the overall monitoring system. Completeness is the capacity to track *all* critical uncertainties which affect the success of a policy. All critical uncertainties must be monitored in order to adapt in case any of these uncertainties unfolds in an undesirable way. This does not mean, however, that each critical uncertainty requires a signpost. Ambiguity about which of the critical uncertainties is unfolding in an undesirable way is allowed as long as singling out critical uncertainties is not required for taking action. As an example, consider a river flood protection policy where the adaptive action is raising higher levees: future precipitation and future land use may be uncertain, but rather than monitoring both of those, one could just monitor change in streamflow only.

The schema in Fig. 2 gives an example of the relation between the number of signposts, critical uncertainties, and actions. Ambiguity between critical uncertainty A and B

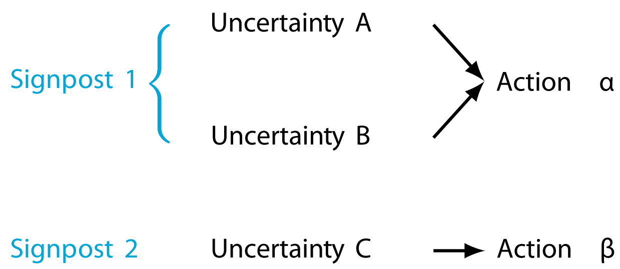

Fig. 2 Correspondence between the number of signposts, critical uncertainties, and actions, illustrative example. Signpost 1 tracks critical uncertainties A and B; signpost 2 tracks critical uncertainty C. The arrows represents the link between critical uncertainties and their conditional actions to be implemented

is allowed because distinguishing between them is not required for taking action α. Therefore, a single signpost is sufficient for tracking both uncertainties. A single signpost for all three uncertainties, however, would not enable the identification of whether action α or β is to be implemented. Ambiguity between C and A or B is not allowed. Therefore, a second signpost is necessary.

Guaranteeing the completeness of the monitoring system requires considering not just the total number of signpost; it also requires considering the mutual dependency of two or more signposts (Cover and Thomas 2006). If two signposts have high mutual dependency, one of them is in principle redundant. Redundant signposts, apart from reducing parsimony without necessity, could have the additional disadvantage of providing an illusory sense of confirmation, which is instead just the result of the signposts mutual dependency.

Apart from redundant, signposts can be more or less synergic (Goodwell and Kumar 2017). In this case, identifying the critical uncertainty can only be attained by knowing the joint value of the synergic signposts. The XOR logic operator is a case of purely synergic interaction; continuing on the river flood protection policy case, an example of partially synergic signposts is when one signpost tracks change in precipitation intensity, another signpost tracks change in land use impermeability, and only the combined increase of both signposts indicates that the policy success is jeopardised. Synergic signposts cannot be reduced, and they must be analysed together.

Notwithstanding the need to reduce the number of signposts, some additional signposts more than the strictly required number may be appropriate. An additional signpost may carry the extra information required to detect a situation that is not explicable by the considered hypotheses. Such a situation can trigger a reflection about the limits of the present adaptive policy, leading to its reassessment (Pahl-Wostl 2009).

3 Role of models

Models can support the analysis of signposts with respect to the criteria of relevance, accuracy, and precision, and the whole monitoring system with respect to the criteria of completeness and parsimony. This requires models capable of both exploring beyond the present behaviour of the system, and representing the uncertainty left after observations, hence requiring features from both exploratory and stochastic modelling.

Exploratory modelling is an approach for developing and using models to map critical uncertainties to their consequences. It relies on the use of computational experimentation to systematically explore the multidimensional uncertainty space. The use of exploratory modelling for policy assessment and design is well-documented in the literature (Bankes 1993; Kwakkel and Pruyt 2013). Exploratory modelling underpins various model-based approaches for designing adaptive policies, such as (many-objective) robust decision-making (Lempert et al. 2006; Kasprzyk et al. 2013; Kwakkel 2017). Stochastic models (Sims et al. 1982) are used to represent the system uncertainty due to inherent variability or imperfect knowledge (Weijs et al. 2010; Nearing et al. 2016).

Exploratory and stochastic approaches have complementary features. In the exploratory modelling approach, models are used to map possible values of inputs, considered uncertain, to their outputs. Stochastic models, in contrast, represent the uncertainty of model output for a certain value of input. In long-term planning problems, both the inputs and in the inputs-outputs relationship are uncertain; therefore, both exploratory and stochastic features are relevant. Exploratory models enable representing a system at points beyond the observed behaviour, stochastic models allow the representation of the observational uncertainty at that point.

3.1 Evaluate relevance and observability

The model can be employed to assess the relevance and observability of candidate signposts. *Ex-ante* evaluation of the relevance of a signpost entails exploring how the signpost parameter changes for different values of the critical uncertainties: the model can be used to explore the value of the signpost parameter for different values of the critical uncertainties. This relationship can be used backwards to track the value of the critical uncertainty given the signpost parameter, and test whether the critical uncertainty has changed.

Ex-ante evaluation of the observability of a signpost requires knowing possible values of the signpost estimate for a given value of the signpost parameter. The variation of a signpost estimate is its sampling variability, which can be represented by a probability density function, a range, or a set of possible values. The sampling variability of the signpost estimate must be quantified for all values of parameters of signpost, for the entire range of the critical uncertainty, in order to map the signpost observability for all possible conditions. Representing the sample variability in the estimate of the parameter requires a stochastic model that reproduce both system variability and observational uncertainty. This model can be used backwards to assess the likelihood of a signpost parameter given the estimate.

When evaluating relevance and observability for parameter values outside the range of observed behaviour, the parameter-estimate relationship is used in extrapolation. This introduces the need to make assumptions about the capacity to detect the parameter from the estimate. These assumptions can be that existing trends will continue beyond the observed range of signpost parameters, or other conjectures.

Figure 1 summarises the role of stochastic and exploratory models in evaluating the relationship between signpost estimate and signpost parameter, and between the signpost parameter and the critical uncertainties. The symbol of inversion applied to the models expresses that they are used “backwards”, i.e. we estimate the value of uncertainties for a given output, rather than the reverse.

3.2 Evaluate completeness and parsimony

Models can be used to analyse the completeness of the monitoring system. Models can be employed to test the effects of variating critical uncertainties on each signpost. This analysis can be used to test if each critical uncertainty is tracked by at least one signpost. If this is the case, then the monitoring system as a whole is capable of tracking all critical uncertainties, hence guaranteeing completeness. Additionally, this analysis can be used to understand which signpost tracks which critical uncertainty. It can be used to identify if a signpost tracks more than one critical uncertainty. Such signpost, all other things being equal, is preferable to multiple signposts, as long as resolving ambiguity among critical uncertainties is not required for deciding among different actions.

Models can be used to test the effect of varying each critical uncertainty on multiple signposts. This analysis permits testing the degree of mutual dependency between signposts. That is, if in response to varying a single critical uncertainty, two or more signposts show high mutual dependency, then both signposts track the same uncertainty and one of them may be redundant. In this case, parsimony can be increased by removing one of the redundant signpost.

4 Application

In this section, we apply the proposed criteria on a flood protection problem in the Niger River for a hypothetical adaptive policy designed in 1985 and applied thereafter. This hypothetical setting allows us to use past data as if the adaptive policy was in its operational phase.

4.1 Policy problem

The city of Niamey, the capital of Niger, is located along the shore of the Niger River. The irregular flow of the river poses a risk of flood for the city. At the inception of the adaptive policy, i.e. in 1985, the levees at Niamey protect the city until a maximum discharge of 2000 m^3/s. The frequency of inundation considered acceptable is below an expectation of one over 25 years, i.e. $F^* < 1/25\ \mathrm{yr}^{-1}$.

In the future, climate change may put at risk the capacity of levees to offer sufficient protection. The adaptive policy envisages the construction of higher levees, if flood hazard exceed the acceptable value, that would raise the maximum discharge to 2700 m^3/s. Raising the levees is an expensive action; therefore, it will be taken only if it is justified by an increase in flood hazard. The underlying critical assumption in this policy is the stationarity of the hydrological process at Niamey.

4.2 System

Flood at Niamey may happen because of two different hydrological processes: the Guinean, or “black” flood, and the Sahelian, or “red” flood (Aich et al. 2016). In this analysis, we will focus on the red flood because of its unpredictable long-term change. The change in distribution of flood magnitude poses a difficulty for the flood defence strategy at Niamey.

4.3 Signposts selection

Discharge at Niamey is directly available, and therefore, it is selected as the signpost variable: We consider the red flood peak discharge taking the maximum discharge during the rainy season of each year (Wilcox et al. 2018).

The selected signposts are as follows:

S_E Average of annual maximum discharge at Niamey, on a 15-year moving window.
$S_{\max}$ Max of annual maximum discharge at Niamey, on a 25-year moving window.

S_E parameter is the “yearly flood intensity at Niamey”, while $S_{\max}$ parameter is the “25-year flood at Niamey”. $S_{\max}$ intends to observe the policy success directly; S_E intends to partially filter out the noisy signal.

4.4 Model

The model used is a generalised extreme value (GEV) stochastic model used in a quasi-stationary fashion. The GEV model estimates the frequency of extreme events given the maximum values over a fixed period (Coles et al. 2001). In flood frequency analysis, the period is generally a year, and the variable the maximum yearly discharge. The GEV model parameters are generally estimated from historical time series, assuming stationary conditions. Here, instead, we use the GEV in an exploratory mode, i.e. we explore the model response for a large set of possible values of critical uncertainties, assuming quasistationary conditions. The critical uncertainties that we consider are the GEV location and scale parameters. The GEV shape parameter is particularly difficult to estimate from a short-time series, requiring a much longer horizon than the period for which we have data, or some regional analysis (Panthou et al. 2012). To reduce ambiguity in data interpretation, we will constrain the shape parameter assuming it to be zero. In this case, the GEV reduces to a Gumbel distribution.

Figure 3 shows the "critical region" spanned by the critical uncertainties where the policy is successful. The blue line in Fig. 3 correspond to the boundary of policy success: if both μ and σ fall below it, the policy is successful; otherwise, there is a need for action. The frontier line corresponds to the critical flood frequency $F^* = 1/25\ \text{yrs}^{-1}$, derived analytically in Appendix B.1. The point in Fig. 3 represents the location and scale parameters estimated from historical data in 1985.

4.5 Signposts evaluation

First, we evaluate their relevance and observability of the two signposts using the GEV model. Second, we evaluate a monitoring system made of these two signposts used together. Here, we assess the completeness of the monitoring system by evaluating its capacity to track all critical uncertainties. We also analyse the mutual dependency between signposts.

4.5.1 Relevance

Relevance is how well a signpost tracks one or more critical uncertainties. In this case, the critical uncertainties are the location and the scale parameters. Figure 4 shows how parameters influence signpost values. In each plot of Fig. 4, the slope is a measure of signpost relevance for that parameter.

The "S_E vs. location μ" plot (top left) and the "$S_{\max}$ vs. location μ" plots (bottom left) show the same slope, meaning that μ influences S_E and $S_{\max}$ alike: the two signposts have the same good relevance for μ. In the "S_E vs. scale σ" plot (top right), instead, the slope is almost flat. This means that σ has very little influence on the value of S_E, i.e. S_E has very

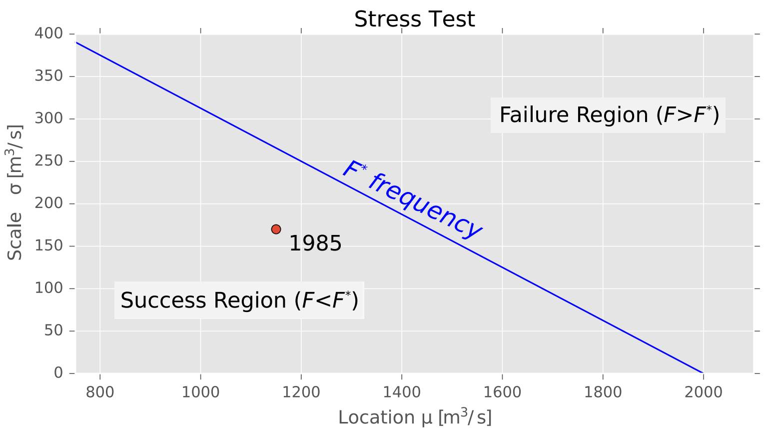

Fig. 3 Critical region, including historical values of parameters (red point). Critical region boundary (blue line) separates success region (below the line) from failure region (above the line)

low relevance for σ. In the "$S_{\max}$ vs. scale σ" plot (bottom right), the higher value of the slope implies that, $S_{\max}$ has good relevance for σ, in this outperforming S_E.

4.5.2 Observability

Both signposts use a moving window over a past period. A moving average is a "quasi-stationary" statistic, used here to detect nonstationarity (Kharin et al. 2007). Use of a moving window leads to some bias, i.e. a loss in accuracy.

Selecting the window size involves making a trade off between precision and accuracy: a longer window results in higher precision and lower accuracy. Accuracy depends on the model's parameter rate of change. The rate of change is not included in the model, hence accuracy cannot be evaluated quantitatively.

In Fig. 4, precision about the signposts estimate is represented by the 90% uncertainty bands, estimated as in Appendix B.2: the larger the band, the lower the precision. The plots

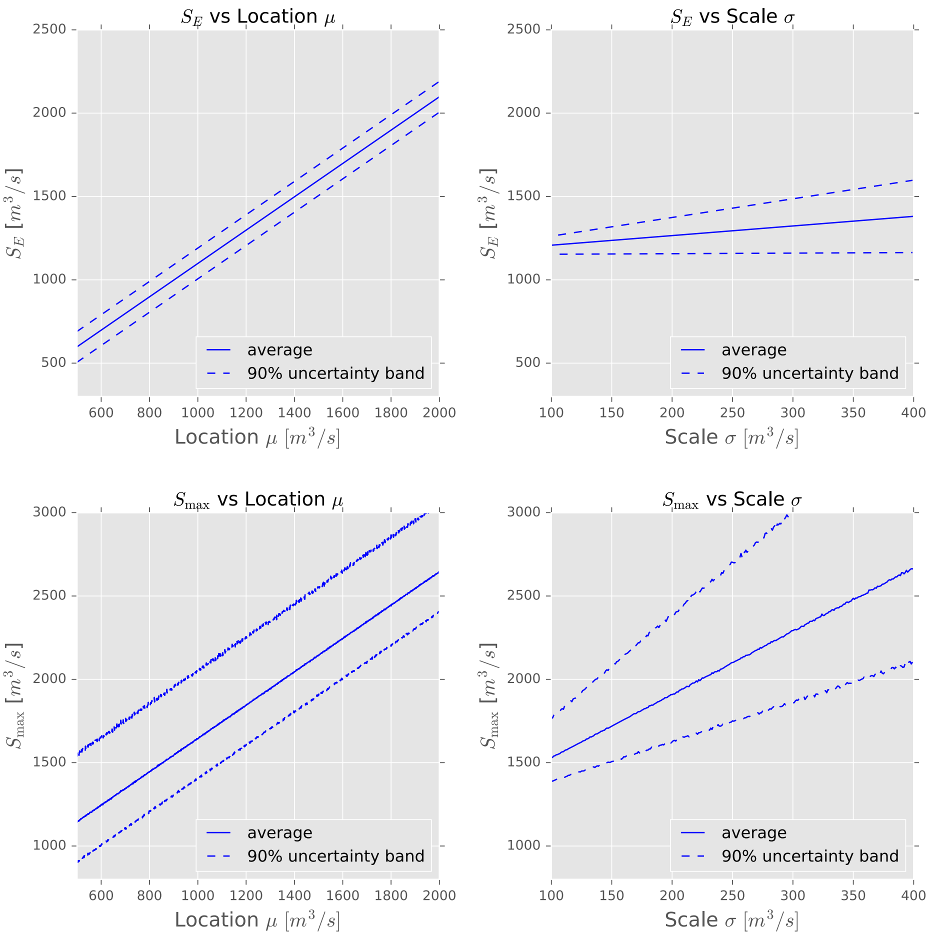

Fig. 4 Top: S_E expected value and uncertainty bands for different values of critical uncertainties, i.e. location μ (left) and scale σ (right). Bottom: $S_{\max}$ expected value and uncertainty bands for different values of critical uncertainties, i.e. location μ (left) and scale σ (right)

for signposts $S_{\max}$ (bottom) present a larger uncertainty band than the plots for signpost S_E (top): $S_{\max}$ has lower precision than S_E. Since $S_{\max}$ has lower accuracy and precision than S_E, therefore it has overall a lower observability.

4.5.3 Completeness

We analyze here the completeness of a monitoring system based on a single signpost, either S_E or $S_{\max}$. From the analysis on signpost relevance, it emerges that S_E can track μ only, whereas $S_{\max}$ is able to track both μ and σ.

A monitoring system made of S_E only will not completely resolve the uncertainty about the consequences on flood frequency and the need for action, which depend on both parameters. A monitoring system made only of $S_{\max}$ will be ambiguous about which parameter has changed. Importantly, however, there will be no ambiguity about the consequences on flood frequency and the need for action. For this reason, a monitoring system made of $S_{\max}$ only, despite this ambiguity, is complete.

4.5.4 Parsimony

Figure 4, plots on the left, shows that change on μ results in the same change on the two signposts. Hence, if μ is the only parameter to change, the correlation between the two signposts would be close to one, meaning that mutual dependency would be very high. The difference between signposts is due to the scale parameter, which affects $S_{\max}$ only. If used together, the information they provide is to some extent overlapping; therefore, one of the two will be partially redundant.

4.5.5 Synthetic evaluation

We give here an qualitative evaluation, where the each criteria is classified as either positive or negative. This highly synthetic appraisal does not reflect the completeness of signpost evaluation as seen in the previous section, nonetheless it provides a clear synoptic overview.

Table 1 summarises the evaluation of signposts with respect to the criteria of relevance and observability. Table 1 shows that no signpost dominates the other: S_E has higher observability, but low relevance for the scale parameter, and, if used alone, its lack of completeness can lead to ambiguity; on the other hand, the advantage of $S_{\max}$ in catching changes for all parameters is counterbalanced by its lower observability, both in accuracy and precision.

Table 2 summarises the analysis of different monitoring systems made of one or two signposts for completeness and parsimony. Monitoring systems made of S_E and $S_E + S_{\max}$ have either low completeness or low parsimony, whereas a monitoring system made of only $S_{\max}$ satisfies both criteria.

Table 1 Synthetic qualitative evaluation of signposts S_E and $S_{\max}$ for different criteria

Signpost	Relevance	Observability	
	μ/σ	Accuracy	Precision
S_E	✓ / ✗	✓	✓
$S_{\max}$	✓ / ✓	✗	✗

Table 2 Synthetic qualitative evaluation of monitoring systems made of S_E, $S_{\max}$, and both, for different criteria

Signposts	Completeness	Parsimony
S_E	✗	✓
$S_{\max}$	✓	✓
$S_E + S_{\max}$	✓	✗

4.6 Signposts in action

The capacity of the signposts S_E and $S_{\max}$ to detect change are tested on a data series from the hypothetical policy implementation, i.e. in 1985, until the year for which we have data, i.e. 2013. Figure 5 presents the signposts value, the 90% uncertainty bands and the a set of possible trigger values. Signposts' 90% uncertainty bands communicates uncertainty about the signpost parameter, meaning that the real signpost parameter has 90% probability of falling within these bands. Band values are calculated using the procedure defined in Appendix B.2, using historical values of μ and σ. Trigger points are calculated using a range of possible parameters value corresponding to the acceptable flood frequency, as in the procedure defined in Appendix B.3.

At the inception of the policy in 1985, S_E suggest with sufficient confidence that the policy is successful, whereas the noisier $S_{\max}$, for which a large part of the band exceeds the trigger value, does not offer the same level of confidence. From 1997, however, S_E starts

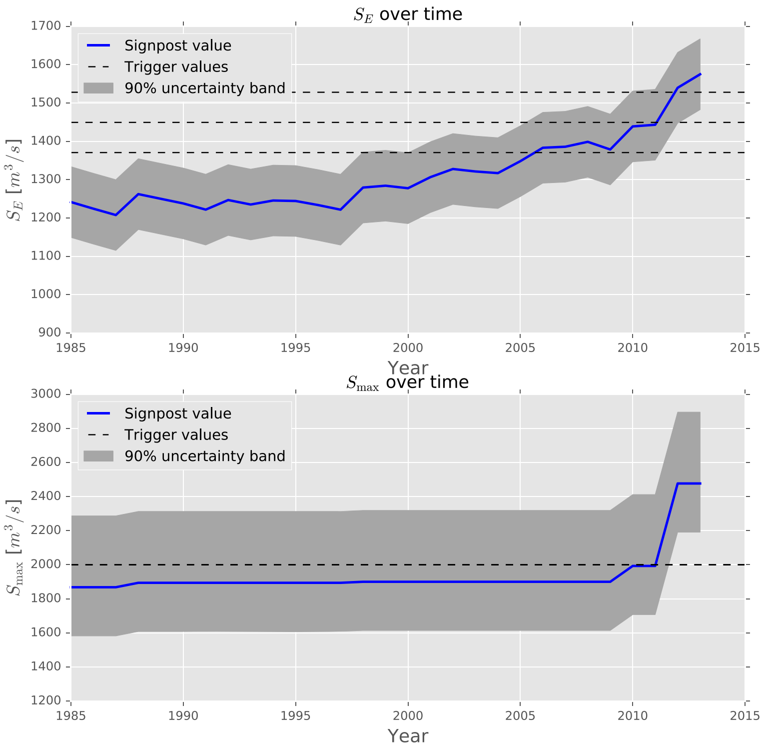

Fig. 5 Signposts value over time

to rise, exceeding the highest trigger value in 2012. $S_{\max}$ is always below the lowest trigger values until 2010, when it raises, and from 2012, it exceeds its trigger value with sufficient confidence. This implies that in 2012, the need for adaptation is sufficiently evident. For both signposts, change is relatively rapid: the quasistationarity assumption, that has been used for designing and assessing signpost, is not a strong one. This rapid change penalises the precision of both signposts, but $S_{\max}$ is penalised more because of its larger window size.

Both signposts are successful in detecting the occurring change. S_E however, offer some anticipation, and it can be considered as early warning signal. $S_{\max}$, instead, provides more confidence about the need for change, and it can be considered as confirmation signal. Uncertainty in S_E is mostly due to its ambiguity in identifying whether the policy is successful, as reflected in the large spread in trigger values. For $S_{\max}$, instead, which tracks the policy success directly, uncertainty is due to the difficulty in estimating the signpost parameter, as evidenced by the larger uncertainty bands.

5 Conclusion and discussions

This paper presented criteria that can be used when designing a monitoring system for adaptive policies based on signposts. These criteria have emerged from the analysis of the information chain from system observations to policy success. The criteria are relevance, observability, completeness, and parsimony. Individual signposts can be evaluated by their relevance and observability. The monitoring system as a whole can be evaluated by its completeness and parsimony. These criteria can be used either for selecting the signposts that will make up the monitoring system, or for pre-selecting a set of candidate signposts to be evaluated and selected within a formal decision problem. In the latter case, the proposed criteria can also be used for an a-posteriori interpretation of the selected signposts, which would enhance their acceptability. We presented how models can be used for the a-priori evaluation of candidate signposts and the overall monitoring system on these criteria. More specifically, we showed how the exploratory use of stochastic models can be employed to understand the response of signposts to change of critical uncertainties.

Single signposts are evaluated according to relevance, defined as the capacity of a signpost parameter to track the critical uncertainties that it is intended to monitor, and observability, defined as capacity to determine the signpost parameter from the signpost estimate. When a model is available, relevance can be evaluated by testing if the signpost parameter is dependent on critical uncertainties, and observability can be evaluated by testing if bias (accuracy) and variability (precision) between signpost estimate and signpost parameter is low. The whole monitoring system is evaluated according to completeness, defined as the capacity to track all critical uncertainties which affect the success of a policy, to be guaranteed, and parsimony, i.e. the number of signposts, to be reduced. When a model is available, completeness can be evaluated by testing that variation of each critical uncertainty is tracked by at least one signpost, and redundant signposts can be identified by testing if they have a high mutual dependency.

The proposed criteria and their model-based evaluation have been demonstrated on a hypothetical case of an adaptive policy for flood protection in the Niger River, West Africa. We identified two signposts and evaluated them using a stochastic model. We found that the two signposts had different degrees of relevance and observability, that a monitoring system made of a single signpost can be more or less complete, depending on the signpost, and that a monitoring system made of both signpost would be less parsimonious but more complete. We applied the signpost to historic data from 1985 to 2013 to test their effectiveness in detecting change. Despite their differences, both signposts provided the information required for adapting the policy. The monitoring system, however, does not offer a clear-cut signal that adaptation is required. In the face of the information provided by the monitoring system, some uncertainty remains, and with them, the risk of taking adaptive actions when these are not really required, or vice versa failing to detect the need for adaptive actions. The model we used is a quasistationary GEV: even if a nonstationary GEV could offer better results, the model used is nonetheless fitted for the analysis; therefore, we leave the testing of better models to further research.

Policy success is an unstable equilibrium, similar to “dancing on the top of a needle” (McInerney et al. 2012), that can be maintained only by continuous feedback and re-adaptation: effective adaptation hinges on an well-designed monitoring system. Evidence of change may emerge at a pace slower than expected; therefore, the capacity to detect change sufficiently on time must be properly investigated. Analysing the effectiveness of a monitoring system to detect change is required to prevent the overestimation of the capacity to adapt over time, i.e. to avoid suggesting an adaptive capacity that cannot be attained in reality.

The proposed criteria provide the foundation for the development of a comprehensive methodology for the integrated evaluation of an adaptive policy and its monitoring system, which will be the aim of future research. Such a comprehensive methodology will be used to test, prior to the policy implementation, whether the monitoring system offers sufficient information for timely adaptation.

Acknowledgements We are grateful to the anonymous reviewers whose precious comments helped improved the manuscript.

Funding information This work is partially supported by the Netherlands Organisation for Scientific Research (www.nwo.nl/en).

Software availability The code and data from this paper are available on Github (https://github.com/luciofaso/Monitoring_DAP/blob/master/1_Niger/Monitoring_Niger.ipynb).

Appendix A: Decomposition of informativeness

In Eq. A.1, we use the law of total probability to decompose the monitoring of critical uncertainties into relevance and observability.

$$P(\lambda_i|d_t) = \int_\delta \mathcal{M}^{-1}(\lambda_i|\delta) \cdot P(\delta|d_t) \cdot d\delta \tag{A.1}$$

In Eq. A.1, λ_i are the ith critical uncertainty, d_t is the signpost estimate, i.e. the data obtained from observations of the real system, and δ_j is the signpost parameter, $\mathcal{M}$ is the system model, and $P(\lambda_i|d_t)$ is the relation between critical uncertainties and signpost estimate. The system model, $\mathcal{M}(\delta|\lambda)$, contains the relationship between the critical uncertainties and the signpost parameter, and it is used backwards to estimate the latter from the earlier. $\mathcal{M}^{-1}(\lambda_i|\delta)$ is a measure of how critical uncertainties change with respect to the signpost parameter, hence the signpost *relevance*. $P(\delta|d_t)$ represents the possible values of the signposts parameter given its estimate, hence a measure of its *observability*.

Appendix B: Test-case values identification

Equation B.1 represents the Gumbel distribution.

$$\mathcal{F}(q;\mu,\sigma) = \exp\left(-\exp\left(-\frac{q-\mu}{\sigma}\right)\right) \tag{B.1}$$

where q is the yearly maximum discharge, μ and σ the location and scale parameter. The Gumbel distribution is equivalent to a generalised extreme value distribution for $\xi = 0$, a.k.a. type-I GEV.

B.1. Critical region

The boundary of the critical region in μ and σ is the set of $[\mu^*, \sigma^*]$, space in R^2, can be found by inverting Eq. B.1, conditioning the flood frequency to be equal to the critical flood frequency, i.e. $F = F^*$, and the discharge equal to the flood threshold level, i.e. $q = q_{\text{flood}}$. Then, one can find the relation at Eq. B.2.

$$\mu^* = q_{\text{flood}} + \log\left(-\log\left(1 - F^*\right)\right) \cdot \sigma^* \tag{B.2}$$

In Eq. B.2, the relation $[\mu^*, \sigma^*]$ is a straight line, in which q_{flood} is the intercept for $\sigma = 0$, and $\log(-\log(1 - F^*))$ its slope.

B.2. Signpost

Signposts distributions are derived analytically from Eq. B.1, assuming quasistationary condition.

S_E is distributed according to Eq. B.3.

$$S_E \dot{\sim} \mathcal{N}\left(\mathbb{E}(q), \frac{\text{VAR}(q)}{n}\right) \tag{B.3}$$

In Eq. B.3, $\mathcal{N}$ is the normal distribution, $\mathbb{E}(q)$ and $\text{VAR}(q)$ are the expected value and variance of Eq. B.1, being $\mathbb{E}(q) = \mu + \sigma \cdot \gamma$, and $\text{VAR}(q) = (\pi^2 \cdot \sigma^2)/6$, where $\pi \simeq 3.14$ is Greek pi, and $\gamma \simeq 0.577$ is the Euler–Mascheroni constant.

$S_{\max}$ is distributed according to Eq. B.4.

$$S_{\max} \sim \mathcal{F}^{-1}\left(\hat{\mathcal{F}}_n(t)\right) \tag{B.4}$$

where

$$\hat{\mathcal{F}}_n(t) \dot{\sim} \mathcal{N}\left(\mathcal{F}(t), \frac{\mathcal{F}(t)\cdot(1-\mathcal{F}(t))}{n}\right) \tag{B.5}$$

Equation B.4 use the property that quantile q convergences to der Vaart (2000) to In Eq. B.4, $F^{-1}(\cdot)$ is the inverse of the original Gumbel distribution, as defined in Eq. B.1, and $\hat{F}_n(t)$ its empirical distribution. The empirical distribution converge to the original Gumbel distribution as in Eq. B.5 (der Vaart 2000). In Eq. B.5, $\mathcal{N}$ is the normal distribution, t the quantile, and n the sample size; for S_{10}, $t = 24/25$, and $n = 25$. Because of the low rate of convergency, however, in this application, we estimated the quantile distribution by a montecarlo approach, using 15000 sampling for each pair of (μ, σ).

B.3. Trigger point selection

A trigger point is the signpost value at which adaptation is required. In the test case, policy requires adaptation if flood frequency F exceeds the threshold level F^*, of one over

25 years. Flood frequency level corresponds to a $[\mu^*, \sigma^*]$, that is a two-dimensional space in model parameters, as in Eq. B.2.

Trigger points are selected by sampling three equidistant combinations of location and scale parameter from the parameters space, in proximity of the parameters historically observed. Then, the three sets of parameters are mapped over the expected signpost values, finding the signpost values that would be measured, on average, for each set of parameters.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Aich V, Koné B., Hattermann FF, Paton EN (2016) Time series analysis of floods across the Niger river basin. Water 8(4):165

Bankes S (1993) Exploratory modeling for policy analysis. Oper Res 41(3):435–449

Blyth W, Bradley R, Bunn D, Clarke C, Wilson T, Yang M (2007) Investment risks under uncertain climate change policy. Energy Policy 35(11):5766–5773

Brown C, Ghile Y, Laverty M, Li K (2012) Decision scaling: linking bottom-up vulnerability analysis with climate projections in the water sector. Water Resour Res 48:W09537. https://doi.org/10.1029/2011WR011212

Bryant BP, Lempert R (2010) Thinking inside the box: a participatory, computer-assisted approach to scenario discovery. Technol Forecast Soc Chang 77(1):34–49

Ceres RL, Forest CE, Keller K (2017) Understanding the detectability of potential changes to the 100-year peak storm surge. Clim Change 145:221. https://doi.org/10.1007/s10584-017-2075-0

Coles S, Bawa J, Trenner L, Dorazio P (2001) An introduction to statistical modeling of extreme values, vol 208. Springer, Berlin

Cover TM, Thomas JA (2006) Elements of information theory, 2nd edn., Wiley-interscience, New York

Cover TM, Thomas JA (2012) Elements of information theory. Wiley, New York

Dekking M (2005) A modern introduction to probability and statistics: understanding why and how. Springer, Berlin

der Vaart AW (2000) Asymptotic statistics, vol 3. Cambridge University Press, Cambridge

Dewar JA, Builder CH, Hix WM, Levin MH (1993) Assumption-based planning; a planning tool for very uncertain times, Tech rep, DTIC Document

Funtowicz SO, Ravetz JR (1990) Uncertainty and quality in science for policy, vol 15. Springer Science & Business Media, Berlin

Goodwell AE, Kumar P (2017) Temporal information partitioning: characterizing synergy, uniqueness, and redundancy in interacting environmental variables. Water Resour Res 53(7):5920–5942

Haasnoot M, Middelkoop H, Offermans A, Van Beek E, Van Deursen WPA (2012) Exploring pathways for sustainable water management in river deltas in a changing environment. Clim Chang 115(3–4):795–819

Haasnoot M, Kwakkel JH, Walker WE, ter Maat J (2013) Dynamic adaptive policy pathways: a method for crafting robust decisions for a deeply uncertain world. Glob Environ Chang 23(2):485–498

Haasnoot M, Schellekens J, Beersma JJ, Middelkoop H, Kwadijk JCJ (2015) Transient scenarios for robust climate change adaptation illustrated for water management in The Netherlands. Environ Res Lett 10(10):105, 008

Hallegatte S (2009) Strategies to adapt to an uncertain climate change. Glob Environ Chang 19(2):240–247

Hamarat C, Kwakkel JH, Pruyt E (2013) Adaptive robust design under deep uncertainty. Technol Forecast Soc Chang 80(3):408–418

Hamarat C, Kwakkel JH, Pruyt E, Loonen ET (2014) An exploratory approach for adaptive policymaking by using multi-objective robust optimization. Simul Model Pract Theory 46:25–39

Herman JD, Giuliani M (2018) Policy tree optimization for threshold-based water resources management over multiple timescales. Environ Model Software 99:39–51

Hertzler G (2007) Adapting to climate change and managing climate risks by using real options. Aust J Agric Res 58(10):985–992

ISO ISO5725-6 (1994) Accuracy (trueness and precision) of measurement methods and results-Part 6: Use in practice of accuracy values. International Organization for Standardization, Geneva, 1994

Jeuland M, Whittington D (2014) Water resources planning under climate change: Assessing the robustness of real options for the blue Nile. Water Resour Res 50(3):2086–2107
Kasprzyk JR, Nataraj S, Reed PM, Lempert R (2013) Many objective robust decision making for complex environmental systems undergoing change. Environ Model Software 42:55–71
Kharin VV, Zwiers FW, Zhang X, Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J Clim 20(8):1419–1444
Kwakkel JH (2017) The exploratory modeling workbench: an open source toolkit for exploratory modeling, scenario discovery, and (multi-objective) robust decision making. Environ Model Software 96:239–250
Kwakkel JH, Pruyt E (2013) Exploratory modeling and analysis, an approach for model-based foresight under deep uncertainty. Technol Forecast Soc Chang 80(3):419–431
Kwakkel JH, Walker WE, Marchau V (2010) Adaptive airport strategic planning. EJTIR 10(3):249–273
Kwakkel JH, Haasnoot M, Walker WE (2016) Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. Environ Model Software 86:168–183
Lee KN (1994) Compass and gyroscope: integrating science and politics for the environment, Island Press, Washington
Lempert R, Groves DG (2010) Identifying and evaluating robust adaptive policy responses to climate change for water management agencies in the American west. Technol Forecast Soc Chang 77(6):960–974
Lempert R, Groves DG, Popper SW, Bankes SC (2006) A general, analytic method for generating robust strategies and narrative scenarios. Manag Sci 52(4):514–528
Liu Y, Gupta HV (2007) Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. Water Resour Res 43:W07401. https://doi.org/10.1029/2006WR005756
McInerney D, Lempert R, Keller K (2012) What are robust strategies in the face of uncertain climate threshold responses? Clim Change 112(3–4):547–568
Moser SC, Ekstrom JA (2010) A framework to diagnose barriers to climate change adaptation. Proc Natl Acad Sci 107(51):22, 026–22, 031
Nearing GS, Tian Y, Gupta HV, Clark MP, Harrison KW, Weijs SV (2016) A philosophical basis for hydrological uncertainty. Hydrol Sci J 61(9):1666–1678
Pahl-Wostl C (2009) A conceptual framework for analysing adaptive capacity and multi-level learning processes in resource governance regimes. Glob Environ Chang 19(3):354–365
Panthou G, Vischel T, Lebel T, Blanchet J, Quantin G, Ali A (2012) Extreme rainfall in West Africa: a regional modeling. Water Resour Res 48:W08501. https://doi.org/10.1029/2012WR012052
Papoulis A (1977) Signal analysis, vol 191. McGraw-Hill, New York
Raso L, Weijs SV, Werner M (2018) Balancing costs and benefits in selecting new information: efficient monitoring using deterministic hydro-economic models. Water Resour Manage 32: 339. https://doi.org/10.1007/s11269-017-1813-4
Sims CA, Goldfeld SM, Sachs JD (1982) Policy analysis with econometric models. Brook Pap Econ Act 1982(1):107–164
Sterman JD (2006) Learning from evidence in a complex world. Am J Public Health 96(3):505–514
Tariq A, Lempert R, Riverson J, Schwartz M, Berg N (2017) A climate stress test of Los Angeles? water quality plans. Clim Chang 144(4):625–639
Walker WE, Rahman SA, Cave J (2001) Adaptive policies, policy analysis, and policy-making. Eur J Oper Res 128(2):282–289
Weijs SV (2011) Information theory for risk-based water system operation, 658 Ph.D. thesis. Delft University of Technology, Delft, The Netherlands
Weijs SV, Schoups G, Van De Giesen N (2010) Why hydrological predictions should be evaluated using information theory. Hydrol Earth Syst Sci 14(EPFL-ARTICLE-167375):2545–2558
Wilby RL, Dessai S (2010) Robust adaptation to climate change. Weather 65(7):180–185
Wilcox C et al (2018) Trends in hydrological extremes in the Senegal and Niger rivers. J Hydrol 566:531–545
Wise RM, Fazey I, Smith MS, Park SE, Eakin HC, Van Garderen ERMA, Campbell B (2014) Reconceptualising adaptation to climate change as part of pathways of change and response. Glob Environ Chang 28:325–336
Woodward M, Kapelan Z, Gouldby B (2014) Adaptive flood risk management under climate change uncertainty using real options and optimization. Risk Anal 34(1):75–92
Zeff HB, Herman JD, Reed PM, Characklis GW (2016) Cooperative drought adaptation: integrating infrastructure development, conservation, and water transfers into adaptive policy pathways. Water Resour Res 52(9):7327–7346