Climate Risk Management 34 (2021) 100375

Contents lists available at ScienceDirect

Climate Risk Management

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

Playing the long game: Anticipatory action based on seasonal forecasts

David MacLeod [a], Dominic R. Kniveton [b,*], Martin C. Todd [c]

[a] School of Geographical Sciences, University of Bristol, United Kingdom
[b] School of Global Studies, University of Sussex, United Kingdom
[c] Department of Geography, University of Sussex, United Kingdom

ARTICLE INFO

Keywords:
Climate adaptation
Seasonal climate forecasting
Early warning
Forecast-based financing
Disaster risk reduction
Anticipatory action

ABSTRACT

Acting in advance of floods, drought and cyclones often requires decision-makers to work with weather forecasts. The inherently probabilistic nature of these forecasts can be problematic when deciding whether to act or not. Cost-loss analysis has previously been employed to support forecast based decision-making such as Forecast-based Financing (FbF), providing insight to when an FbF system has 'potential economic value' relative to a no-forecast alternative. One well-known limitation of cost-loss analysis is the difficulty of estimating losses (which vary with hazard magnitude and extent, and with the dynamics of population vulnerability and exposure). A less-explored limitation is ignorance of the temporal dynamics (sequencing) of costs and losses. That is, even if the potential economic value of a forecast system is high, the stochastic nature of the atmosphere and the probabilistic nature of forecasts could conspire over the first few forecasts to increase the expense of using the system over the no-forecast alternative. Thus, for a forecast-based action system to demonstrate value, it often needs to be used over a prolonged length of time. However, knowing exactly how long it must be used to guarantee value is unquantified. This presents difficulties to institutions mandated to protect those at risk, who must justify the use of limited funds to act in advance of a potential, but not definite disaster, whilst planning multi-year strategies. Here we show how to determine the period over which decision makers must use forecasts in order to be confident of achieving 'value' over a no-forecast alternative. Results show that in the context of seasonal forecasting it is plausible that more than a decade may pass before a FbF system will have some certainty of showing value, and that if a particular user requires an almost-certain guarantee that using a forecast will be better than a no-forecast strategy, they must hold out until a near-perfect forecast system is available. The implication: there is potential value in seasonal forecasts, but to exploit it one must be prepared to play the long game.

1. Introduction

In recent years, the humanitarian community has expressed an increased desire to use weather and climate forecasts to guide their anticipatory actions (see Wilkinson et al., 2018). However, past research has highlighted a number of problems in integrating forecasts into climate risk management. These can be considered to be broadly related to: (i) the perception of forecasts, including a perceived or real lack of saliency, credibility and legitimacy (Hansen et al., 2011, Patt et al., 2007, Cash et al., 2003); and, (ii) difficulties in decision

* Corresponding author.
E-mail addresses: david.macleod@bristol.ac.uk (D. MacLeod), d.r.kniveton@sussex.ac.uk (D.R. Kniveton), m.todd@sussex.ac.uk (M.C. Todd).

https://doi.org/10.1016/j.crm.2021.100375
Received 8 May 2021; Received in revised form 30 October 2021; Accepted 30 October 2021
Available online 9 November 2021
2212-0963/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

making using uncertain information, including overcoming behavioural, institutional, social and technological barriers within relevant organisations (Hammer et al., 2000, Hammer et al., 2001, Hillier and Dempsey, 2012, Ngugi, 2002, Patt and Gwata, 2002, Phillips et al., 2002, Podestá et al., 2002, Patt et al., 2005, Bharwani et al., 2005, Rengalakshmi, 2007, Hansen et al., 2009, Roncoli et al., 2009). Such challenges may be particularly pronounced in the application of seasonal climate forecasts, in which forecast skill is often relatively low and timescales for building adequate experience and confidence are very long (decades). Thus, despite decades of provision and dissemination of seasonal climate forecasts, direct uptake remains limited. While many of the observed problems can be envisaged to be mitigated through greater efforts towards co-production, the issue of uncertainty is inherent to forecasts (Lemos et al., 2012). In practice this means that when acting in advance of a warning of a hazard using a forecast, for each event there is a chance of acting in vain (commonly referred to as a 'false alarm') or failing to act (commonly referred to as a 'miss') which both carry reputational and socio-economic costs. To maximise the benefits of early action and minimise the impact of the false alarms and misses, decision makers must decide at which forecast probability levels to act upon (the forecast trigger) and what actions are most appropriate. These decisions constitute the Early Action Plan (EAP) of automated, objective Forecast based Finance (FbF) systems (Coughlan de Perez et al., 2016, hereafter CdP16).

The FbF approach, developed by CdP16 and further refined in Lopez and Haines (2017) Lopez et al., (2018, hereafter L18) is a systematic approach to help decision-makers utilise uncertain probabilistic forecasts. Based on a 'value of information' approach the FbF formulation utilises knowledge of the forecast skill and the monetary costs and benefits of forecast-based actions, including the disaster losses, costs of actions and acting in vain, losses avoided by the actions and unavoidable losses. Through an optimisation approach (e.g. Eq. (2) of L18) the 'optimal' forecast probability thresholds are derived, that if acted on, in the long run, will lead to net benefits with respect to not acting on forecasts.

FbF initiatives have developed rapidly in recent years. However, despite the promise of 'optimised' forecast, the formal FbF approach as specified by CdP16 and L18 has been rarely implemented in full, at least in risk management systems of humanitarian organisations. There are likely various reasons for this, including access to the required information (forecast skill information, robust estimates of costs/losses) as well as inertia in institutions. There is a fundamental issue of estimating non-monetary costs and losses: where these cannot be estimated, they may be excluded from utility evaluations. In the absence of such information, early action-early warning asks that decision-makers utilise uncertain forecast information without clear evidence of the benefits.

Whilst the FbF approach inherently demands a long term perspective (through the use of forecast skill information derived from hindcast information) the actual time that would be required to realise the benefits of acting on forecasts is not known, given that the actual sequence of weather/climate events is largely stochastic. The case of using seasonal climate forecasts is especially challenging given that the hindcast sample size is orders of magnitude smaller than for weather forecasting. The net long term benefits of acting based on short-lead weather forecasts may be realised rather quickly, for example, a single season may see many trials of forecasts for tropical cyclones, but just a single trial for a seasonal rainfall forecast. Here, we extend the work of CdP16 and L18 to consider in detail a case of seasonal climate prediction and to assess how long it may be expected to take to realise benefits of acting of forecasts, even when forecast skill is known.

2. Methodology

2.1. Potential economic value of a forecast

To estimate the potential benefit for using a forecast to guide decision-making, the potential economic value has been defined, based on the cost-loss framework (Murphy 1977). This assumes that it is possible to take action with a cost (C) to protect against the loss (L) experienced when a hazard occurs with no protection. The expected expense associated with a particular forecast system can be calculated and if that expense is lower than a default decision-making strategy, that forecast can be said to have potential economic value.

To estimate the expense of a forecast system based on historical forecasts for some hazard, it is necessary to determine when that system might have triggered action. For a probabilistic forecast system, this can be done by determining a probability trigger threshold: forecasts with probabilities above that threshold are assumed to trigger action and those below do not. This sequence of actions/no actions can then be compared to occurrences of the specified hazard. If action is triggered a cost C is incurred: a 'hit' occurs if a hazard follows, a 'false alarm' or 'action in vain' otherwise. If no action is triggered but the hazard occurs a 'miss' occurs, incurring cost L, with a 'correct rejection' if no action is taken and no hazard occurs. These outcomes are summarized in Table 1.

A FbF system becomes financially worthwhile when condition of Eq. (1) below is met (CdP16). L18 modified this equation to include a term to account for the unavoidable losses that occur despite the preparedness actions. Here we disregard these unavoidable loses as they exist whether there was a hit or a missed event.

Table 1
Contingency table based on triggering early action if the forecasted probability exceeds a threshold, p (adapted from CdP16).

	Hazard occurs	Hazard does not occur
Forecast triggers action	Hit, *a(p)*	Acting in vain, *b(p)*
No action taken	Miss, c(p)	Correct rejection, *d(p)*

$$L \bullet \frac{a(p) + c(p)}{n} > C \bullet \frac{a(p) + b(p)}{n} + \Delta C \bullet \frac{b(p)}{n} + L \bullet \frac{c(p)}{n} \quad (1)$$

In Eq. (1) a, b and c are the number of event 'hits', 'false alarms' and 'misses' using a particular forecast over some period from Table 1. The left-hand side of represents the financial losses experienced when unprotected hazards occur and the right-hand side represents the average cost accrued when using a forecast system to guide decision-making. The first two terms on the right-hand side of Eq. (1) represent the combined cost of worthy action and action in vain, whilst the third term represents the losses associated with missed events. In the following we assume zero extra cost of action in vain (i.e. $\Delta C = 0$).

Eq. (1) represents the condition for a forecast system to show value over a no-forecast alternative, averaged over a period. However, it may still take many forecast trials for a system to show value, even if this condition is met. To quantify this, we then use a Monte Carlo sampling approach to simulate the impact of sequencing of hazards and non-hazards, hits, misses and false alarms, described below. We provide an answer to the question: given a specification of forecast skill, how long will it take for an FbF system based on that system to clearly show value?

2.2. Using a Monte Carlo approach to evaluate the relationship between the skill of a forecast-based decision system and its long-term value

The skill of a probabilistic forecast system is often defined for some event, such as the seasonal rainfall total falling below the lower tercile of climatological rainfall. A variety of complementary verification scores have been defined but we choose metrics compatible with the formal mathematical formulation of an 'worthy' FbF system (eq (1)): specifically:

(i) The Hit rate (HR), defined as the percentage of events which were successfully forecast ($HR = a/a + c$)
(ii) The False Alarm Ratio, defined as the proportion of actions taken which are in vain. $FAR = b/(a + b)$

Luck plays a role in probabilistic forecasting. Even if the underlying frequency of a hazard is precisely known, the exact sequence of events and non-events is subject to significant uncertainty. For example: one may roll a die three times and get six three times or roll it ten times without a single six. Similarly, a hazard with a one in five-year return period may happen several times in the next decade, or not at all. The exact sequence of events and non-events has a clear impact on hazard-related losses, along with the perceived value of an FbF system. Acting on a probabilistic forecast it is quite possible to experience a string of 'bad luck' (misses or false alarms), even if the system is highly skilful in the long term.

In the context of a decision taken based on a seasonal forecast ahead of an annual rainy season, such a decision would be taken once per year. Hence a string of bad luck may last many years and the long term the value of the system could take several years or longer to reveal itself. To use another gambling analogy: with a card counting system for Blackjack that shifts the odds slightly in our favour it is not unlikely that we lose several hands after sitting down at the table. But if our system works, we will beat the house – as long as we play for long enough. The question here is: how long is long enough?

This impact of 'luck' (or more precisely, sampling uncertainty) on the realized economic value of a forecast system has not been considered before. However, it is essential to evaluate its importance: users setting up a forecast-based action system should be clearly aware of the length of time needed before a system will show value. A clear demonstration of a worthwhile system is far from guaranteed within the first few forecasts. The previous approaches of CdP16, L18 are idealized analyses and implicitly correspond to the expected value of the system obtained over an infinite number of trials. Here we inject some further realism by using a Monte Carlo approach to evaluate the time an FbF system would take to 'beat the house', and how this expected time-to-value is dependent on the skill of the underlying forecast system.

We generate a large ensemble of one thousand 'climate event trajectories' each of 100 years duration, consistent with the underlying probability of the event. These take the form of a binary sequence of events and non-events, each generated with an independent probability, p, equal to the baseline frequency. We carry out separate simulations for events with different return periods: one in three years, five years and ten years ($p = 33\%$, 20% and 10% respectively). Results for a one in five year hazard are shown in the main manuscript, comparison with one in three and one in ten year events is presented in supplementary material.

Two strategies are possible in the absence of any forecast information:

- "Always act", in which early preparedness actions are taken every year. The accumulated expense of this strategy here is simply the cost of acting multiplied by the number of years and is independent of the event trajectory.
- "Never act", in which early preparedness actions are never taken, and losses are faced when an event occurs. The accumulated expense of this strategy depends on the specific event trajectory; no expense is incurred when no event occurs, whilst losses accrue if an event occurs.

These two strategies provide a benchmark to measure the value of a forecast system; if more expense is accrued when guided by a seasonal forecast compared to either of these two strategies then it is difficult to claim the forecast is useful. To calculate the expenses of different strategies we must first define costs and losses, noting that expenses scale linearly with a multiplication of both C and L. Instead it is the ratio C/L is the crucial factor. We therefore carry out simulations for a range of C/L ratios. We focus in the main manuscript on results where C/L is set to equal to the event frequency, as this C/L ratio is an equilibrium point where the expected expenses of the default strategies are equal, noting that 'always act' always costs C and 'never act' costs pL on average. These are equal where $C = pL$, or $C/L = p$. We also show results in supplementary material for $C/L = 0.5p$ and $C/L = 2p$, situations which favour

'always act' and 'never act' respectively. In all cases we explicitly show nominal expenses associated with L=$1000.

We add a third reference strategy to these two 'no-forecast' default options: the expense associated with a theoretical perfect forecast. This is calculated for each event trajectory by accruing a cost every time an event occurs and $0 otherwise. In reality this strategy is impossible (requiring perfect advance knowledge of hazard occurrence) but is a useful indication of the maximum possible saving that a seasonal forecast could provide. Finally, we calculate the expense accrued if an (imperfect) forecast system is relied upon. This is calculated for a theoretical system with particular values of both HR and FAR. For each of the 1000 event trajectories we calculate a 1000-member ensemble in order to sample possible "forecast trajectories". Multiple forecast trajectories are necessary, as the forecast system is probabilistic: e.g. with HR = 50% when an event occurs there is 50% chance a warning would have been given. Each forecast trajectory is then generated as a sequence of warnings, where the probability of issuing a warning is mathematically conditional on the occurrence of an event or non-events and is a function of HR and/or FAR.

The conditional probability that a warning is issued given the occurrence of an event is defined as:

$$P(warning|event) = HR \tag{2}$$

and the conditional probability that no warning is issued before an event is:

$$P(no\ warning|event) = 1 - HR \tag{3}$$

Eqs. (2 and 3) enable the classification of the events in each event trajectory into hits and misses.

Classifying non-events as false alarms or correct rejections using FAR is slightly more complicated. It would be trivial if the false alarm ***rate*** *(*the fraction of non-events followed by a warning), was specified, however here we use the false alarm ***ratio****,* as it is more relevant to FbF practitioners, as it indicates the chance an early action will be in vain on average (CdP16). Since FAR is a function of both false alarms and hits, the probability of classifying a non-event as a false alarm is dependent on both FAR and HR.

In order to derive conditional probabilities for warnings in advance of non-events we use the notation of Table 1 to define FAR explicitly, as the fraction of warnings which are in vain:

$$FAR = \frac{b}{a+b} \tag{4}$$

We then express the number of hits as a function of the hit rate and the total number of trials:

$$a = p \bullet \mathrm{HR} \bullet \mathrm{n} \tag{5}$$

where p is the baseline frequency of the event (specified previously as 20%, or 0.2). Substituting Eq. (5) into Eq. (4) gives:

$$FAR = \frac{b}{p \bullet HR \bullet n + b} \tag{6}$$

This may be rearranged as an equation for the number of false alarms:

$$\frac{b = p \bullet HR \bullet FAR \bullet n}{1 - FAR} \tag{7}$$

The marginal probability of a false alarm could now be calculated by dividing both sides of the equation by n. However, we require the **conditional** probability of a warning, given no event. This is calculated by dividing (7) by the number of non-events, $(1-p)n$, giving:

$$\frac{P(warning|noevent) = p \bullet HR \bullet FAR}{(1 - FAR)(1 - p)} \tag{8}$$

The conditional probability of no warning given no event is then given by:

$$\frac{P(nowarning|noevent) = 1 - p \bullet HR \bullet FAR}{(1 - FAR)(1 - p)} \tag{9}$$

Eqs. (2), (3), (8 and 9) now allow the classification of each event and non-event in each event trajectory to be converted to a sequence of warnings and non-warnings. The 1000 forecast trajectories generated for each event trajectory quantify the possible forecasts which might be issued ahead of a particular sequence of events, each consistent with a particular value of HR and FAR. The associated forecast expense of each forecast trajectory using the established values for cost and loss.

In total one million potential series of forecast expense are calculated for a particular value of HR and FAR. These allow the temporal characteristics of the forecast expense to be calculated. From this sample the central estimate indicates average expected reduction in expense from default action provided by the forecast system, relative to the reduction in expense provided by a perfect forecast (in the long term this will asymptote to the traditional estimation of forecast value as in L18). Finally, for each member the sample we calculate the number of trials pass before forecast value is (and remains) positive relative to default action; the percentiles of this metric across the sample then quantify the number of trials one must wait in order to have confidence of the system displaying positive value.

This Monte Carlo sampling process is then repeated for all possible sequences of HR and FAR, ranging from 0 to 100%, in intervals of 5%. Metrics described above are then displayed as a function of HR and FAR, providing a direct visualization of the relationship

between expected forecast skill and economic value.

2.3. Theoretical restriction on allowed combinations of HR and FAR

FAR and HR are not mutually independent, as both depend on forecast hits. The implication of this is an exclusion of combinations of HR and FAR, found when both HR and FAR and sufficiently large. We exclude these regions in our analysis and describe the delineation of this part of HR, FAR space here.

The space of 'impossible' combinations of FAR and HR can be defined by starting with the definition of the number of false alarms (Eq. (7)) and considering that the ratio b/n is bounded by [0,1]. It is not possible for values of HR and FAR in the range [0,100%] to lead to values of this ratio below zero. However there are values which lead to impossible values b/n >1. Considering this possibility, Eq. (7) gives:

$$\frac{1 > p \bullet HR \bullet FAR}{1 - FAR} \tag{10}$$

which can be rearranged as an identity for disallowed values of FAR:

$$FAR > \frac{1}{p \bullet HR + 1} \tag{11}$$

giving a first exclusion on HR/FAR space.

A second restriction on allowed regions of HR/FAR space comes from the restriction that the values of Table 1 must sum to n.

$$a + b + c + d = n \tag{12}$$

whilst the total of hits and misses must equal the number of events

$$a + c = pn \tag{13}$$

Eq. (12) and (13) can then be combined:

$$b = (1 - p)n - d \tag{14}$$

which when combined with Eq. (7) gives an expression for the fraction of the total number of trails which are correct rejections:

$$\frac{d/n = 1 - p - p \bullet HR \bullet FAR}{1 - FAR} \tag{15}$$

d/n is also bounded by [0,1]. Values greater than one would require an impossible negative value of HR or FAR, however value below zero is possible when:

$$FAR > \frac{1 - p}{p(HR - 1) + 1} \tag{16}$$

Eqs. (11 and 16) define non-identical impossible regions of HR/FAR space. These both depend on p, with a larger region of the space disallowed with a greater frequency of events.

Intuitively this restriction can be thought of as contrary influence of hits on HR and FAR: an increasing number of hits increases HR, forcing ever-larger numbers of false alarms in order to obtain high FAR. At some point there are insufficient trials to obtain these high values. As a simple example: as soon as there is one hit, FAR = 100% is impossible. If all events are hits, the maximum possible FAR depends on the number of remaining non-events (indicated by p). In an extreme case of p = 1 it is clear to see that any value of FAR beyond FAR = 0% is undefined: when an event occurs with every trial, a false alarm is impossible.

It can be shown that of Eqs. (11 and 16), Eq. (16) is a more restrictive condition on HR/FAR space and so this is used to define the limit of HR/FAR sampling.

2.4. Plausible seasonal forecasting skill

In order to provide an indication of real-world HR and FAR values, we calculate scores from the European Centre for Medium-Range Weather Forecast (ECMWF) SEAS5 seasonal forecasting system. Full technical details are provided in Johnson et al 2019: here analysis of the hindcast over 1981–2020 is used to estimate skill. Skill scores are over Kenya, focusing on one in five year wet events for the two seasons. These are chosen as a neat example of low and high skill: the long rains (March-May) has relatively low predictability as relationships to large-scale predictable climate drivers are weak (e.g. Kilavi et al., 2018, Vellinga and Milton 2018, MacLeod 2018, MacLeod 2019), whilst the short rains (October-December) has high predictability, particularly for wet events, with strong forcing from the Indian Ocean Dipole and El Niño Southern Oscillation (MacLeod and Caminade, 2019; MacLeod et al., 2021). In each case forecasts at a lead time of one month are used (corresponding to initialization of 1-February and 1-September respectively), and for each, three HR-FAR pairs are estimated based on triggering on a forecast probability of 50%, 30% and 10% separately.

3. Results

An indicative example of potential expense accumulation is shown in Fig. 1. A single event trajectory is shown in Fig. 1a, with accumulated expense for default strategies, along with a single forecast trajectory consistent with HR of 80% and FAR of 40%. The full range of forecast expense associated with this specific event trajectory is shown in Fig. 1b, according to the distribution of the 1000-member forecast trajectory ensemble.

The uncertainty in event trajectories can then be quantified by considering the distribution across all event trajectories. This is shown in Fig. 1c. Here spread in the baseline strategies of "perfect forecast" and "never insure" are generated from 1000 possible event trajectories, consistent with a one in five year return period event. The spread in possible forecast expenses corresponds here to the uncertainty both in the 1000 possible event trajectories as well as the forecast response to these trajectories (consistent with the specified HR = 80% and FAR = 40%).

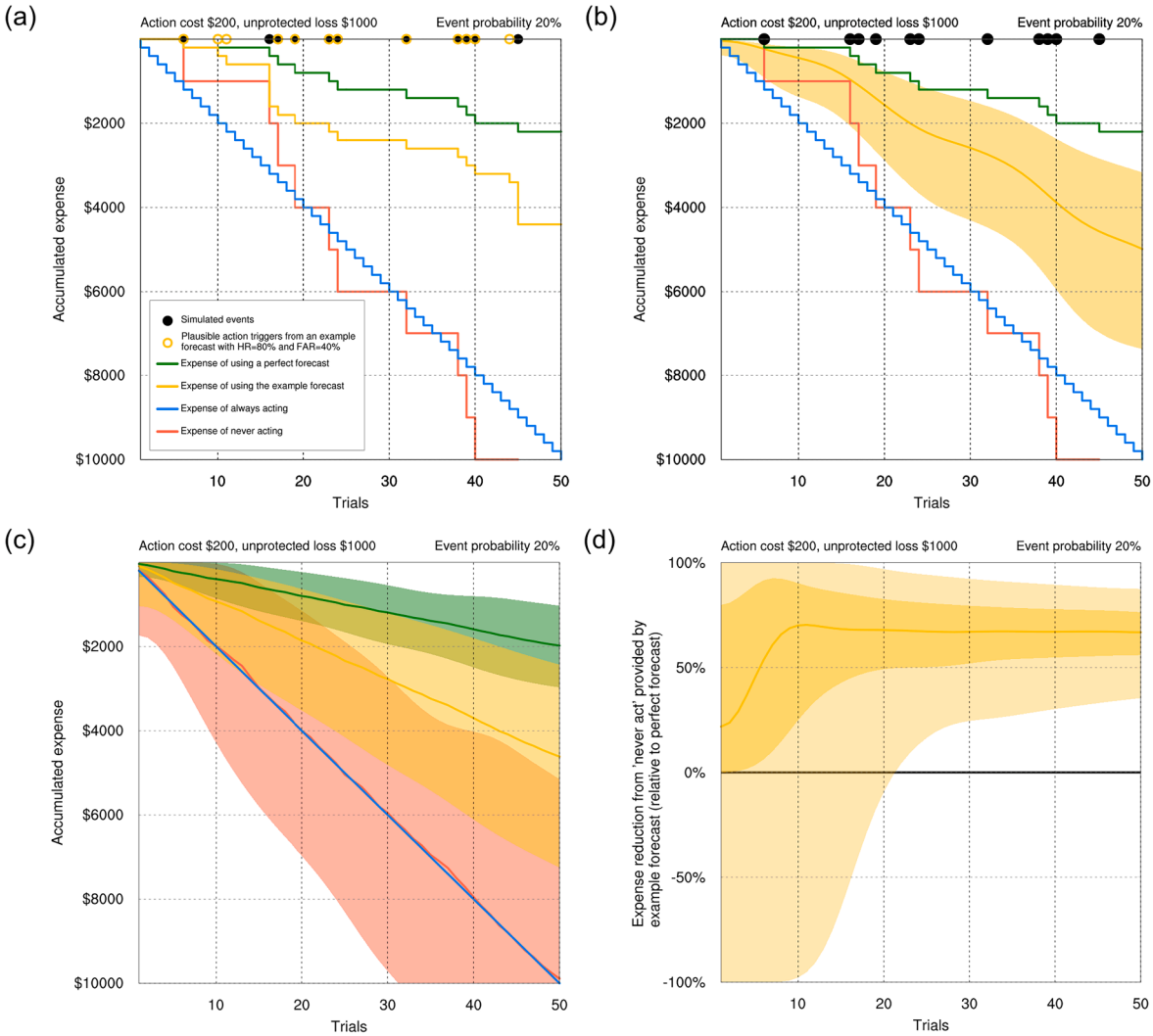

Fig. 1. Estimating the possible future value of a forecast-based action system. Example here shows the expense incurred by different strategies for a 1 in 5-year event (where action costs $200 to avoid a $1000 loss from a hazard). (a) shows a possible trajectory of events (black dots on top axis) and actions based on a forecast consistent with HR = 80% and FAR = 40% (yellow circles). Red and blue lines show the expenses incurred by never and always acting, whilst the yellow and green lines show the expense incurred by acting based on a forecast system with HR = 80% and FAR = 40% or a perfect system. Figure (b) shows the same sequence of events as (a), but with the 5–95% confidenceFbF interval of the expense of the forecast system, estimated by resampling 1000 possible action sequences consistent with the specified HR and FAR. (c) shows the 5–95% confidence interval of the expense of all strategies, calculated by sampling over 1000 event trajectories consistent with the baseline event frequency (NB 1000 action sequences are sampled for each of 1000 event sequences. (d) Shows the reduction in expense from "Never act" when using the forecast system. This reduction is expressed as a percentage of the expense reduction which would be achieved if the forecast were perfect. Lighter shading indicates the 5–95% range of subsamples, darker shading indicates the 25–75% range and the line indicates the median.

In order to show the added value of the forecast system, relative to never acting, Fig. 1d shows the difference in expense between the forecast and "never act", scaled by the expense of the perfect forecast. From this example, most ensemble members indicate positive value relative to never acting, from the very first forecast trial. However around 25% of the samples show negative value in the first year, and around 20 trials are necessary to have a 95% confidence that expense associated with the forecast will be lower than never insuring.

Next the accumulated expense is plotted as a function of HR and FAR in Fig. 2. Fig. 2a–c show the median expense of the forecast system after 10, 20 and 30 trials in US$. As expected, the lowest expenses are seen when forecast skill is high: when HR is close to 100% (no misses) and FAR is close to zero (no false alarms). Naturally, the expense increases over time, with the expense after 30 years ranging around $2000 to over $5000, depending on forecast skill.

The expense saved by a forecast system relative to never acting is shown in Fig. 2d–f as a function of forecast skill, after 10, 20 and 30 years. As expected, more money is saved on average when the system is used for longer; after 30 years a perfect forecast (HR = 100%, FAR = 0%) saves on average well over $2000 in this context. With lower forecast skill less money is saved, although most values of HR and FAR ultimately provide value in the long term, on average. For instance, even a forecast system anticipating with HR = 20% and FAR = 60% could save between $200–1000 after 30 years. However it should be noted that Fig. 2 shows the expected expense reduction; within the bootstrapped ensemble there is significant variability and in practice a particular realization of a forecast system can accrue larger expenses compared to a default "never act" strategy (particularly for systems with lower forecast skill).

Finally we estimate the length of time necessary to be sure a system will show less expense of the alternative of never or always insure. This is shown by calculating the timing that a certain percent of the 1 million ensemble members show less expense that an alternative strategy. This is repeated for each HR/FAR combination and shown in Fig. 3. Fig. 3a and b show the number of years until at least 75% and 95% of the ensemble show less expense than never insuring, whilst 3c and 3d show the number of years for 75% and 95% to show less expense than always insuring. Along with the time-to-value estimate, some indicative skill scores from a real example of a seasonal forecast with high and low skill are included.

Clearly the number of forecasts needed before a system clearly shows value is a strong function of forecast skill. If the system is more skilful with HR >80% and FAR <50%, then there is at least a 75% chance that the forecast will be better than never insuring after only two forecasts (Fig. 3a). This level of skill is possible from the real-world high skill example. However, even with this system there is still

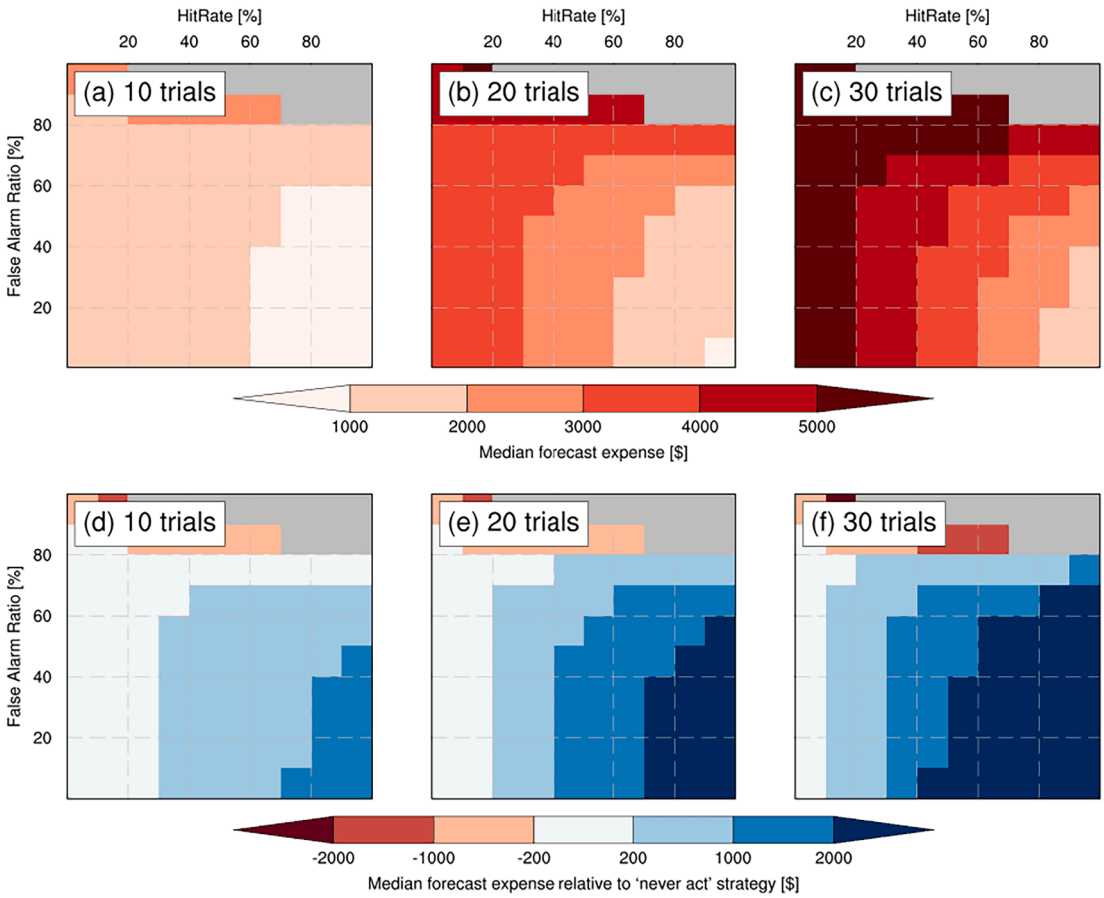

Fig. 2. (a-c) Median expense incurred after 10, 20, 30 trials when using a forecast to trigger action. (d-f) the expense reduction provided by a forecast (relative to a "never act" strategy). The grey region indicates undefined regions of HR/FAR space (see methodology for details).

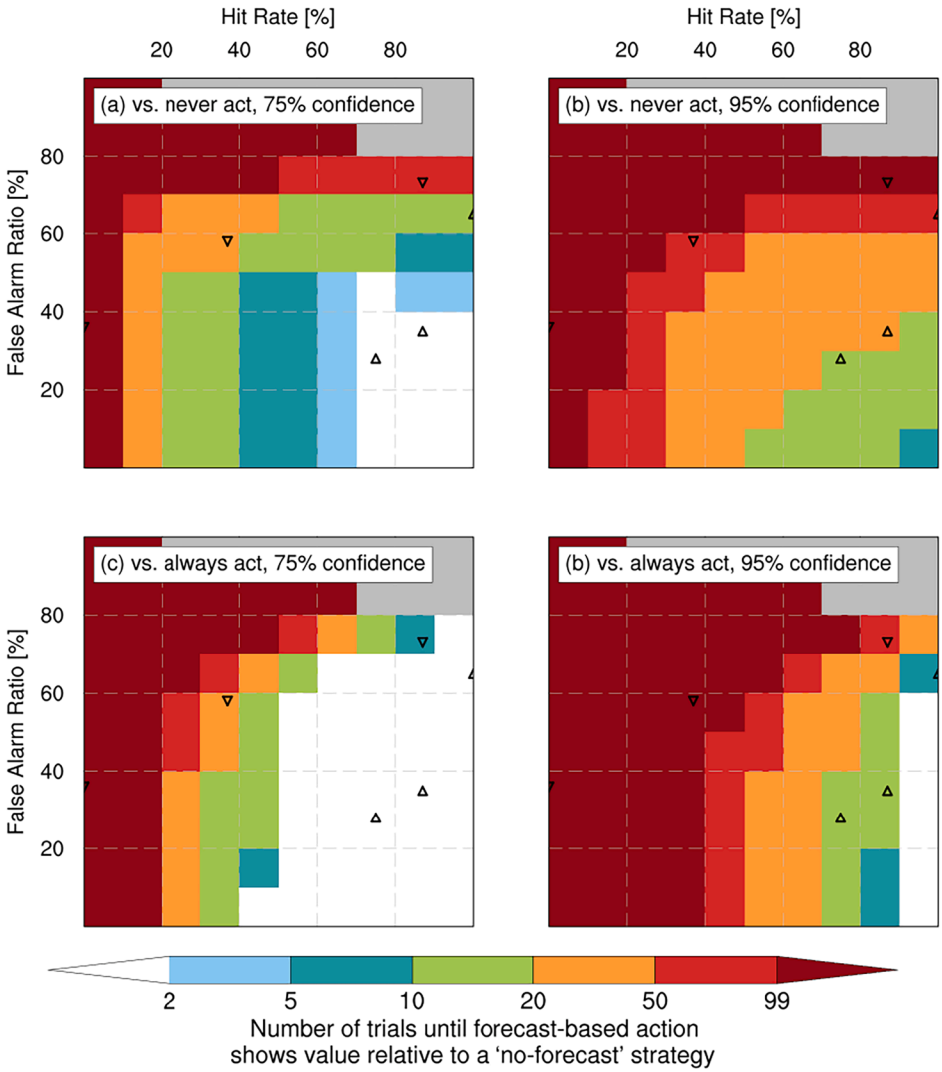

Fig. 3. (a & b) show the number of trials necessary before the expense associated with using a forecast will be lower than the 'never act', in 75% or 95% of cases respectively. Fig. 3 (c & d) show the same metric when measured against the 'always act' scenario. Also included on all figures are six triangles, indicating plausible skill from a seasonal forecast. Three triangles are shown for a high and low seasonal forecast (upward and downward pointing triangles respectively); these results are obtained from ECMWF SEAS5 forecasts of one in five wet seasons over Kenya, for the short and long rains respectively (see methods for details). In each case three instances of HR and FAR are calculated, based on acting with a probability trigger threshold of 50, 30 and 10% (associated with triangles appearing from left to right).

25% chance that the forecast will be more expensive than the alternative after two forecasts; one must use this system for at least 10 forecasts to ensure that money is saved over never insuring (Fig. 3b). For a system with lower skill (e.g. HR = 40% and FAR = 60%) such as is achieved with the real-world low skill example, the long-term expected value of the system is still positive (Fig. 2), however one must use the system for at least 50 years to have 95% certainty that money is saved over the alternative (95% chance Fig. 3b).

Fig. 3c and d show the time a forecast system will clearly show value against to a strategy of "always act". Compared to Fig. 3a and b, HR/FAR space is clearly demarcated into regions which quickly show value and those for which value is not guaranteed even after 100 forecasts. The reason for this is due to the independence of an "always act" strategy to the exact timing of hazards. When using a baseline of "always act", using a forecast with HR below 20% will not show positive value after 100 forecasts in at least 25% of cases, whilst a HR of at least 50% will demonstrate value immediately on using the forecast in 75% of cases (unless FAR is particularly high). However, if one requires 95% certainty that a forecast system will show value compared to always acting, a near-perfect HR is required, along with a FAR of <60%. This stringent condition is because when measuring against "always act", the forecast can only reduce expense by advising not to act. It is then heavily penalized for any misses, which requires many correct worthy actions to recoup this loss.

For reference, additional versions of Fig. 3 are shown in supplementary material which compare results for different hazard frequencies and C/L ratios. Results demonstrate that as hazard frequency increases, the time-to-value decreases. This reflects the decrease

in the time the forecast system must "wait" until it can prove itself; even a forecast with perfect discrimination ability must wait for a hazard to occur before it can beat a forecast of "never acting". Another feature noted is that with increasing hazard frequency, the range of FAR which still provides a reasonable time-to-value also decreases. Where C/L = p, this limit of 'still-useful FAR' is equal to 1-p: this corresponds to the same FAR one would achieve by taking action at random: varying the overall frequency of such a 'random action' provides any desired HR. It is hopefully obvious that any forecast system with FAR is equivalent (or higher) than that achieved by acting at random (or worse) will not show positive economic value!

Finally, supplementary figures show the implications of different C/L ratios which favour either 'always act' or 'never act'. When 'always act' is favoured (C/L = 0.5p), this strategy becomes very difficult to beat: forecasts must have HR >80% to have 75% chance of showing value over this strategy (supplementary Fig. 3, top row); to have 95% confidence they must have HR > 90% and still this will take at least a decade to show value. When 'never act' is favoured (C/L = 2p), FAR determines the possibility of providing certainty of value: with a one in five year event to have 75% certainty then FAR below 50% is required to show value within 50 trials, but to have 95% certainty of this then FAR below 30% is required for HR >50%.

4. Discussion and conclusion

Weather and climate forecasts are inherently probabilistic in nature. This can be problematic to users whose actions (and inactions) are traditionally viewed in the binary terms of being either being right or wrong. This paper attempts to reframe this mismatch by promoting a decision-making approach whereby forecast use is assessed as a longer-term climate risk management strategy. In particular, this reframing places the metrics of skill of the forecasts system at the centre of determining the return. Clearly, a forecast system with a high level of skill leads to reduction in costs and losses, compared to never or always acting: and these returns are realised relatively quickly. When forecast skill is low the use of forecasts can still be theoretically better than a no-forecast strategy, although this may take a very long time to be realized. This is a factor which must be acknowledged as limiting the uptake of such information.

In this paper we place the decision to use forecasts to decide whether to take early preparedness actions in the context of alternative 'no-forecast' decision strategies, i.e., of always acting or never acting, in the absence of forecasts. In particular, we find that there is less uncertainty when estimating the returns on using forecasts when compared against always acting as opposed to never acting. This arises because there is zero uncertainty in the future expense of always acting. When comparing against never acting the uncertainty in realized value of using a forecast system is much higher, as it depends on the sequence of climate events. For example, a decade could pass without a single one in five year event. It is quite possible that a relatively skilful system could unnecessarily trigger in this decade, which would make it look worse. In this context, the risk tolerance of the user becomes key factor in determining whether to use forecasts. For example, a user with a higher risk tolerance can take advantage of a larger range of skilful systems. If the user only requires 75% certainty that using the forecast will show value compared to never acting after 10 years, our research shows that this is possible as long as HR >40% and FAR <50%. However if the user almost guaranteed (95%) certainty of value within 10 years, they must use nothing less than a perfect forecast.

The analysis here provides an estimate of the expected value of a system given certain skill metrics. In practice, estimates of forecast skill are themselves uncertain. For example, with seasonal forecasts, skill metrics are generally calculated by evaluating reforecasts which generally have a sample size below 40. Sampling uncertainty leads to non-trivial uncertainty in both the estimation of the thresholds for a 1 in 5 year event, along with the estimates of HR/FAR themselves. So in reality an envelope of possible HR/FAR should be applied to the analysis presented here. The approach shown in this paper recognises that when using probabilistic information to make a decision, chance helps determine how quickly the user experiences the benefits of using forecasts to guide their decisions making.

Another assumption to the results presented here is that the frequency of the hazard will stay constant in the long term. There is no reason to think this will be the case: significant internal or forced climate variability can significant change the frequency of hazards. The impact on the expected time for a system to show value will depend on how exactly the frequency of the hazard changes – which may be highly nonlinear, and non-monotonic. However analysis presented for hazards with different return periods indicates a general rule of thumb: as hazard frequency increases, a forecast with reasonable skill will show value quicker, as it is given more frequency opportunity to prove itself.

In practice, costs and particularly losses are notoriously difficult to estimate beyond purely financial decisions (there is also an inevitable limitation to the framework as those experiencing losses are not the same as those paying preparedness costs: international aid donors fund FbF programmes, but do not experience the greatest losses when disaster hits). For simplicity we have focused on a scenario where C/L = p, as it balances the expected costs of the two no-forecast strategies; that is, they are equally bad. However results for C/L ratios which diverge from C/L = p reflect the much tougher test of one of the default actions. That is, as C/L moves away from the equilibrium point, one of the strategies becomes much more effective at the expense of the other. This improved default strategy then becomes correspondingly harder for a forecast system to beat: this can be seen when potential economic value curves are plotted as a function of C/L ratio; maximum value is achieved where C/L = p (e.g. Richardson 2000). The implications of this: if C, L and p can be estimated and C/L is significantly different from p, all but near-perfect forecast systems may be excluded as potential triggers for taking early actions.

Nuances and external factors may also result in the real-world 'value' of an FbF system diverging from the idealised toy example presented here. As well as uncertainty in estimating action costs and event losses, these two factors can be non-stationary and dependent on the external, variable factors. Furthermore, event preparation is not an isolated activity: within the forecast horizon additional actions based on shorter-range forecasts may limit the loss of a missed event, whilst false alarms could see some costs

recuperated (e.g. by re-collecting pre-positioned supplies as shorter range forecasts show the hazard risk has decreased). In addition, although our results may indicate that a particular forecast should show value after twenty years, a series of back-to-back false alarms and misses soon after the set-up may lead to drop in trust and program cancellation. Therefore, we do not present these results as a statement about the actual real-world value of FbF, which is highly contextual and related not only to forecast skill and preparedness actions, but determined by specifics of program design, implementation, stakeholder buy-in and communication. Instead, we present the results to delineate reasonable expectations of 'return' from an FbF system, reframing forecast skill as the key determinant of this 'time-to-value'.

The reframing of using forecasts exemplified in this paper puts the onus on the forecasters to quantify the skill of their forecast as much as the user to determine the costs and expected losses to justify the benefits of using forecasts. Clearly, to many of the institutions mandated to reduce disaster risk the cost of inaction when a disaster strikes is politically high, as is the loss of reputation if a disaster is predicted but never occurs. Although the cost-loss framework simplifies real-world decision-making, its value is less in determining where forecasts show value, but in the minimum forecast skill which must be achieved. In this paper we have used the framework to look further: to guide expectations on how long until a forecast-based decision-making system such as FbF must be used to show value. In the case of an annual potential decision from a seasonal forecast, the results are clear: to exploit the potential value of seasonal forecasts, one must be prepared to play the long game.

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.

Acknowledgements

This research was supported by the Science for Humanitarian Emergencies and Resilience (SHEAR) consortium project "Towards Forecast-based Preparedness Action'' (ForPAc, www.forpac.org), Grants NE/P000673/1, NE/P000568/1, and NE/P000428/1. The SHEAR programme is funded by the U.K. Natural Environment Research Council, the Economic and Social Research Council, and the U.K. Department for International Development. DM was also supported by funding from the European Union's Horizon 2020 program through the DOWN2EARTH project (grant 869550).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crm.2021.100375.

References

Bharwani, S., Bithell, M., Downing, T.E., New, M., Washington, R., Ziervogel, G., 2005. Multi-agent modelling of climate outlooks and food security on a community garden scheme in Limpopo, South Africa. Philosophical Transactions of the Royal Society B: Biological Sciences 360 (1463), 2183–2194.

Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D.H., Jäger, J., Mitchell, R.B., 2003. Knowledge systems for sustainable development. Proceedings of the national academy of sciences 100 (14), 8086–8091.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M.K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., Mendler de Suarez, J., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., Zsoter, E., 2016. Action-based flood forecasting for triggering humanitarian action. Hydrology and Earth System Sciences 20 (9), 3549–3560.

Hammer, G., Carberry, P. and Stone, R., 2000. Comparing the value of seasonal climate forecasting systems in managing cropping systems. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems* (pp. 183-195). Springer, Dordrecht.

Hammer, G.L., Hansen, J.W., Phillips, J.G., Mjelde, J.W., Hill, H., Love, A., Potgieter, A., 2001. Advances in application of climate prediction in agriculture. Agricultural systems 70 (2-3), 515–553.

Hansen, J.W., Mishra, A., Rao, K.P.C., Indeje, M., Ngugi, R.K., 2009. Potential value of GCM-based seasonal rainfall forecasts for maize management in semi-arid Kenya. Agricultural Systems 101 (1-2), 80–90.

HANSEN, JAMES.W., MASON, SIMON.J., SUN, LIQIANG, TALL, ARAME, 2011. Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. Experimental Agriculture 47 (2), 205–240.

Hillier, D., Dempsey, B., 2012. A Dangerous Delay: The cost of late response to early warnings in the 2011 drought in the Horn of Africa. Oxfam Policy and Practice: Agriculture, Food and Land 12 (1), 1–34.

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S.P.E., Mogensen, K., Zuo, H., Monge-Sanz, B.M., 2019. SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development 12 (3), 1087–1117.

Kilavi, M., MacLeod, D., Ambani, M., Robbins, J., Dankers, R., Graham, R., Titley, H., Salih, A.A., Todd, M.C., 2018. Extreme rainfall and flooding over central Kenya including Nairobi city during the long-rains season 2018: causes, predictability, and potential for early warning and actions. Atmosphere 9 (12), 472.

Lemos, M.C., Kirchhoff, C.J., Ramprasad, V., 2012. Narrowing the climate information usability gap. Nature climate change 2 (11), 789–794.

Lopez, A., Haines, S., 2017. Exploring the usability of probabilistic weather forecasts for water resources decision-making in the United Kingdom. Weather, Climate, and Society 9 (4), 701–715.

MacLeod, D., 2018. Seasonal predictability of onset and cessation of the east African rains. Weather and climate extremes 21, 27–35.

MacLeod, D., 2019. Seasonal forecasts of the East African long rains: insight from atmospheric relaxation experiments. Climate Dynamics 53 (7-8), 4505–4520.

MacLeod, D., Caminade, C., 2019. The moderate impact of the 2015 El Niño over East Africa and its representation in seasonal reforecasts. Journal of Climate 32 (22), 7989–8001.

MacLeod, D., Graham, R., O'Reilly, C., Otieno, G., Todd, M., 2021. Causal pathways linking different flavours of ENSO with the Greater Horn of Africa short rains. Atmospheric Science Letters 22 (2), e1015.

Murphy, A.H., 1977. The value of climatological, categorical and probabilistic forecasts in the cost-loss situation. Mon. Wea. Rev. 105, 803–816.

Ngugi, R.K., 2002. Climate forecast information: the status, needs and expectations among smallholder agro-pastoralists in Machakos District, Kenya.
Patt, A., Gwata, C., 2002. Effective seasonal climate forecast applications: examining constraints for subsistence farmers in Zimbabwe. Global Environmental Change 12 (3), 185–195.
Patt, A., Suarez, P., Gwata, C., 2005. Effects of seasonal climate forecasts and participatory workshops among subsistence farmers in Zimbabwe. Proceedings of the National Academy of Sciences 102 (35), 12623–12628.
Patt, A.G., Ogallo, L., Hellmuth, M., 2007. Learning from 10 years of climate outlook forums in Africa. Science 318 (5847), 49–50.
Phillips, J.G., Deane, D., Unganai, L., Chimeli, A., 2002. Implications of farm-level response to seasonal climate forecasts for aggregate grain production in Zimbabwe. Agricultural Systems 74 (3), 351–369.
Podestá, G., Letson, D., Messina, C., Royce, F., Ferreyra, R.A., Jones, J., Hansen, J., Llovet, I., Grondona, Martín, O'Brien, J.J., 2002. Use of ENSO-related climate information in agricultural decision-making in Argentina: a pilot experience. Agricultural Systems 74 (3), 371–392.
Rengalakshmi, R., 2007. In: Climate Prediction and Agriculture. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 129–134. https://doi.org/10.1007/978-3-540-44650-7_13.
Richardson, D.S., 2000. Skill and relative economic value of the ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society 126 (563), 649–667.
Roncoli, C., Jost, C., Kirshen, P., Sanon, M., Ingram, K.T., Woodin, M., Somé, L., Ouattara, F., Sanfo, B.J., Sia, C., Yaka, P., Hoogenboom, G., 2009. From accessing to assessing forecasts: an end-to-end study of participatory climate forecast dissemination in Burkina Faso (West Africa). Climatic Change 92 (3-4), 433–460.
Vellinga, M., Milton, S.F., 2018. Drivers of interannual variability of the East African "Long Rains". Quarterly Journal of the Royal Meteorological Society 144 (712), 861–876.
Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D. and Cabot Venton, C., 2018. Forecasting hazards, averting disasters: implementing forecast-based early action at scale. Overseas Development Institute (ODI).