Grantham Research Institute on Climate Change and the Environment

Managing basis risks effectively: balancing the trade-offs from uncertain forecasts for anticipatory action and disaster risk financing

Erica Thompson and Goodwin Gibbins

July 2026

Grantham Research Institute on
Climate Change and the Environment
Working Paper No. 454

ISSN 2515-5717 (Online)

Grantham Foundation for the Protection of the Environment

LSE Global School of Sustainability

The Grantham Research Institute on Climate Change and the Environment was established in 2008 at the London School of Economics and Political Science. The Institute brings together international expertise on economics, as well as finance, geography, the environment, international development and political economy to establish a world-leading centre for policy-relevant research, teaching and training in climate change and the environment. It is hosted by the Global School of Sustainability at LSE and is funded by the Grantham Foundation for the Protection of the Environment.

www.lse.ac.uk/granthaminstitute

Authors' declaration of AI use: The authors declare that no generative AI or AI-assisted tools were used in the research, analysis or preparation of this publication.

This paper was first published in July 2026 by the Grantham Research Institute on Climate Change and the Environment at the London School of Economics and Political Science.

Suggested citation:

Thompson E and Gibbins G (2026) *Managing basis risks effectively: balancing the trade-offs from uncertain forecasts for anticipatory action and disaster risk financing*. Grantham Research Institute on Climate Change and the Environment Working Paper 454. London: London School of Economics and Political Science

DOI: 10.21953/researchonline.lse.ac.uk.00139006

Managing basis risks effectively: balancing the trade-offs from uncertain forecasts for anticipatory action and disaster risk financing

Dr Erica Thompson[1,2,3] and Dr Goodwin Gibbins[3]

1. **UCL Department of Science, Technology, Engineering and Public Policy**
2. **LSE Data Science Institute**
3. **LSE Grantham Institute on Climate Change and the Environment**

Abstract

Anticipatory humanitarian action for weather-related hazards is underpinned by weather and climate information, which is subject to many kinds of uncertainty. This paper is a conceptual exploration of the use of uncertain information in anticipatory action and anticipatory disaster risk financing, grounded in detailed survey and interview data gathered from a cross-section of expert practitioners. We identify three primary trade-offs in system design for anticipatory action: (i) confidence with lead time, (ii) accuracy with simplicity, and (iii) adaptability with consistency. Design of anticipatory action and anticipatory financing systems require difficult choices about balancing these factors to maximise value and gain the trust of all participants. Acting on uncertain forecasts also exposes the stakeholders to basis risk: the mismatch between intended and actual outcomes. We distinguish two types of basis risk which arise respectively from the quantifiable uncertainty in a forecast (Type I) and from the possible model error of a forecasting system (Type II). We discuss strategies for quantifying, reducing, and mitigating the effects of the two types of basis risk, noting that financial stakeholders have different priorities and perspectives from operational stakeholders with regard to the three trade-offs listed above. On the technical side, managing basis risk relies on robust and appropriate forecast evaluation procedures chosen with humanitarian outcomes as objectives. This is necessary but not sufficient for the development of trust in a system, which also requires transparency, communication, codesign, reliability, local input, and accountable processes.

JEL Classification codes

D81: Criteria for Decision-Making under Risk and Uncertainty
G23: Financial Instruments
H84: Disaster Aid

Keywords: Basis risk; Anticipatory action; Disaster risk financing; Decision-making under uncertainty.

Acknowledgements: The authors thank Paolo Avner, Moussa Sidibe, Bramka Arga Jafino, Rashmin Gunasekera and Michael Staudinger for their valuable input to this paper. This study is part of a broader analytical work-program led by the World Bank on the topic of the "Socioeconomic benefits of a better weather information and decision-making ecosystem". It was financially supported by the World Bank's Global Facility for Disaster Reduction and Recovery (GFDRR) and the Austrian Federal Ministry of Finance.

1 Introduction

Advance notice of a hydrometeorological hazard (such as drought, storm, heatwave or flood) is valuable because it permits individuals, communities, governments and aid organisations to act in **anticipation**, to reduce the impact. Anticipation opens up options for a wide range of protective actions (such as distribution of supplies, reinforcement of protective infrastructure, early harvesting of crops, provision of shelter, evacuation) and - if appropriately designed and informed - has the potential to be much more effective than traditional responsive modes of humanitarian action (Suarez, 2010 and Griffiths, 2021, outline the rationale and Pappenberger, 2015, is an example of a detailed benefit analysis).

This paper is a conceptual exploration of the use of uncertain information in anticipatory action and anticipatory disaster risk financing, grounded in detailed survey and interview data gathered from a cross-section of expert practitioners.

The crucial challenge that we focus upon here, is that weather and climate forecast information, especially about extreme or unusual events, is unavoidably limited. The complexity and nonlinearity of the weather system means that on relevant timescales (days, weeks and months) there is **uncertainty** which varies by event, lead time and geographical region. There is important humanitarian value in the uncertain information available, but leveraging it requires care.

Because the forecast is uncertain, users are exposed to **basis risk** - the possibility of unexpected outcomes (either positive or negative) happening due to the forecast not being correct (Harris and Cardenes, 2020; Clement et al, 2018). This can happen in different ways. Here we distinguish two important types: (I) when a probabilistic forecast goes the unexpected way, and (II) when the forecast probabilities themselves are incorrect.

If forecasts were perfect, choosing which actions to take when a hazard is impending and designing anticipatory action systems would be a relatively simple consideration of the cost and benefit to an action. Instead, the **anticipatory action (AA) systems** need to be designed to take into account the probabilities of different outcomes, and minimise or mitigate their exposure to basis risk. To achieve this, probabilistic evaluation methods are therefore also required, as we discuss below.

Over the last decade, a range of experience (Insuresilience 2023, Wagner 2024, Anticipation Hub, 2022 and 2023) has been built up in anticipatory action projects by organisations such as the Red Cross Red Crescent (IFRC 2022, Tozier de la Poterie et al. 2023), Start Network, through their Start Fund Anticipatory Alerts and the Start Ready risk pool (Start Network 2024), WFP (WFP 2023), and UN OCHA (Chaves-Gonzales 2022, Getliffe 2021) and less formal projects in place at the national and local level. Early warnings support informal and ad hoc decisions (Ayers et al. 2024), but “anticipatory action” (also known as forecast-based action or early action) increasingly refers to formal forecast-triggered response plans with pre-arranged financing (Plichta, 2023), which are pre-designed to react to a prediction of an event in a planned way.

Much of the literature in this field focuses on the efficacy of having a pre-established forecast-based action plan to reduce the impact of a predicted disaster (e.g. Coughlan de Perez (2015), Tozier de la Poterie et al. (2023)) and on empirical benefits of early cash transfer (e.g. Pople et al, 2021, Balana et al, 2023, Rahman, 2024). There are many advantages and disadvantages, which were corroborated in our study: (i) the burden of pre-arranging detailed action plans, (ii) the co-benefit of capacity building from developing these plans, (iii) the efficiency and welfare benefits of being able to respond to disasters more quickly, (iv) the logistical challenge of responding within short lead times, communication barriers, insufficient response capacity in-country, (v) the challenge of coordination between institutions using different triggers, and (vi) the potential competition with more general disaster readiness funding.

The primary contribution of this paper is to advance understanding of how anticipatory action systems interact with available data and information, with a particular focus on managing forecast uncertainty—a dimension that has received limited attention in the existing literature. Section 2 describes the methodology. In Section 3, we explore the trade-offs between exposure to basis risk and operational considerations. In Section 4, we distinguish two types of basis risk and in Section 5, we explore how basis risk can be quantified, reduced and mitigated. There are many strategies - some more obvious than others - which improve our ability to leverage limited weather and climate information on hydrometeorological hazards to support effective action. Section 6 offers some key messages, conclusions, and policy implications.

2 Methods

In this paper, we offer an overview from a technical and applied perspective of some of the opportunities and challenges of using weather and climate information for anticipatory decision-making in disaster response contexts, including anticipatory action (AA) and disaster risk financing (DRF). This piece is based on two interlinked research activities:

1. qualitative data gathered from an online survey (Appendix A) and semi-structured interviews (Appendix B) of practitioners working in a range of roles related to design and implementation of AA and DRF, and
2. theoretical advancement of the principles of information use in AA and DRF, particularly related to the trade-off between simultaneous goals and distinguishing different forms of basis risk.

Our theoretical development is grounded in evidence from the practical experience shared with us during six semi-structured interviews and in written survey responses from 28 people who work across the hydrometeorological disaster risk response sector. We asked about the full pathway of how information is created, accessed, understood, distributed, and processed into decisions by institutions with different roles and priorities, focused primarily on anticipatory action, including financing of AA schemes (see survey questions in Appendix A and interview structure in Appendix B). The inclusion criteria for the survey were self-reported experience of working in a relevant role. Table 1 shows the range of roles, from National (Hydro-)Meteorological Services to international aid organisations (non-governmental organisations, NGOs) to local government Disaster Risk Management Agencies. These are not necessarily a representative sample; as such, we do not use survey data to infer anything quantitative or universal about experience within the sector.

The inclusion criteria for the interviews were similar, but we invited a range of participants and then selected six with the aim of hearing from a range of perspectives (geographical regions, experience of different hazards, and roles within the sector). Our interviewees had experience in countries across the globe facing a wide variety of risks (see Table 1). Again, while they were selected for variety, these are not necessarily a representative sample; as such, we do not use interview data to infer anything quantitative or universal about experience within the sector.

Region	S	I
Africa	8	✓
South Asia	8	✓
East Asia and Pacific	2	✓
Europe and Central Asia	2	
Latin America and Caribbean	6	✓
Global	5	✓

Hazard	S	I
Drought	11	✓
Heatwave	5	✓
Cyclone	7	✓
River Flood	14	✓
Flash Flood	9	✓
Glacial Lake Outburst Flood	2	
Coldwave	1	
Other severe weather	1	

Role	S	I
Humanitarian NGO	1	✓
National met service		✓
Scheme design	14	✓
Evaluator	5	
Operations	6	✓
Donor	1	
Beneficiary	1	
Disaster management agency	3	✓

Table 1: An overview of the regions, hazard expertise and roles covered by the 28 respondents to the survey (column S) and the six interviewees (I).

In what follows, interview and survey responses are referenced with brackets, e.g., "(I5, S22)". We transcribed, collated, annotated and anonymised themes from the interviews and surveys and use these to illustrate and give specific examples of our points below.

3 Three key trade-offs in trustworthy information use for AA and DRF

There are three key trade-offs involved in designing systems which use forecast information effectively to support successful anticipatory action and anticipatory disaster risk financing, sketched in Figure 1. Operational considerations, such as for **simplicity** and long enough **lead time**, trade off against considerations which reduce exposure to basis risk, like **accuracy** and **confidence**. Systems also need to trade off **adaptability** with **consistency**. Any plan to make decisions based on uncertain information will need to balance these six attributes in a way that takes account of the needs of practitioners, funders and beneficiaries. An effective and well-understood balance is also critical in gaining the **trust** of all parties, and in achieving the desired **value** in the outcomes.

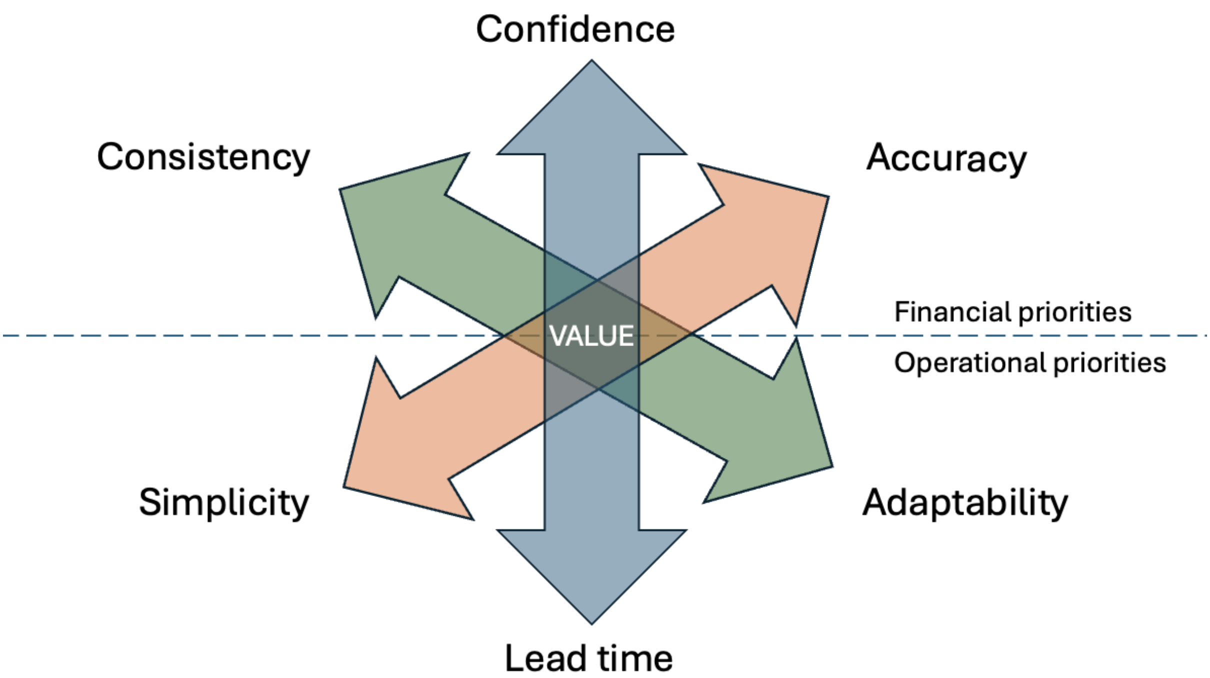

Figure 1: Diagrammatic representation of the three trade-offs discussed in this section: confidence with lead time; accuracy with simplicity; and consistency with adaptability. In the middle, the outcome we are seeking to achieve by balancing these six virtues: value of the system. In section 3.3, we discuss how the trade-offs are balanced by different stakeholders.

3.1 A trade-off of confidence with lead time

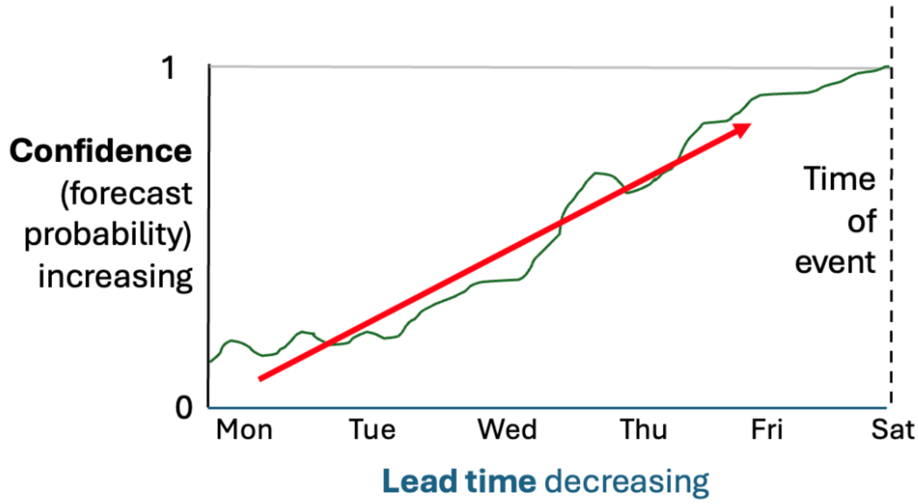

Figure 2: As lead time decreases closer to an event, the confidence in the forecast (usually) becomes greater. The relationship between confidence and lead time is a function of the skill of the forecasting system, and can also vary for different events and different locations. In this sketch, the event happens on Saturday, with confidence increasing in the days leading up to the event on Saturday.

Anticipation of an event, sketched in Figure 2, comes with a trade-off of **confidence** with **lead time** (e.g. Wilkinson, 2018; Anand, 2024). If we wait until *after* an event has happened, we know for certain that it has happened, and we have a good idea (subject to data collection) what and how big it was, who was affected and in what ways. But in order to act *before* an event has impact, we have to rely on **forecasts** which come with some level of **confidence** (and uncertainty). There is some probability of the event not happening in the way that is forecast. Longer lead times permit more interventions, but the forecast will, in general, be of lower confidence. This is summarised in Figure 3, which shows the development of forecast probability over time for a hypothetical event with an anticipatory trigger threshold of 75% probability.

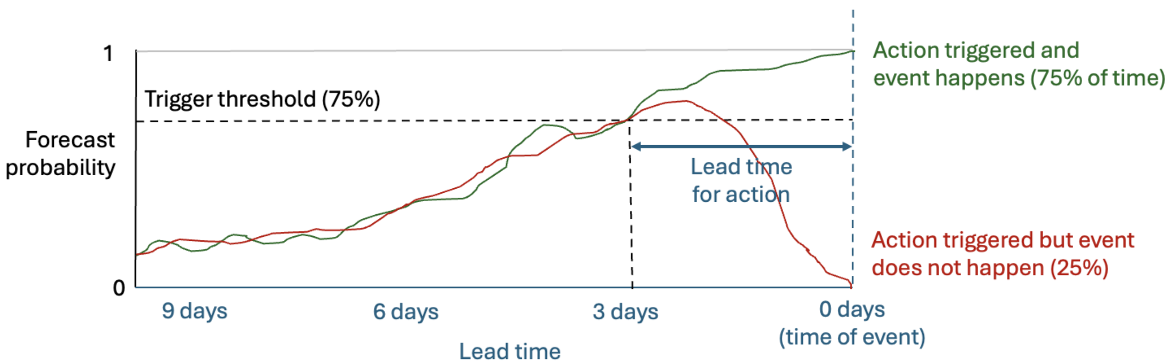

Figure 3: Sketch of a well-calibrated forecast-based anticipatory trigger. The forecast probability is monitored over time and action is triggered if the probability meets a threshold. If that trigger threshold is set at 75% forecast confidence of the event happening, then – if the

forecast is reliable – for 75% of triggers the event will happen as forecast, and for 25% of triggers, it will not happen as forecast.

Although we have used forecast probability here - and most respondents (22 of 28 survey respondents and all of the interviewees) did state that they used probabilistic forecasts - not all forecasts are probabilistic in nature. Nevertheless, the level of confidence will still vary in the same way and could in principle be measured as forecast skill, even for deterministic forecasts.

When a trigger is designed, it defines the accepted trade-off of confidence for lead time. The rationale for anticipatory action is then that in the long run, *the net benefit of acting early in those situations where action is correctly triggered outweighs the loss incurred when action is incorrectly triggered or incorrectly fails to trigger* (a traditional "Cost-Loss" problem). Understanding the frequency and (positive and negative) consequences of each possible outcome in the matrix in Figure 4 is an important step towards defining the situations where an anticipatory approach is indeed worth taking.

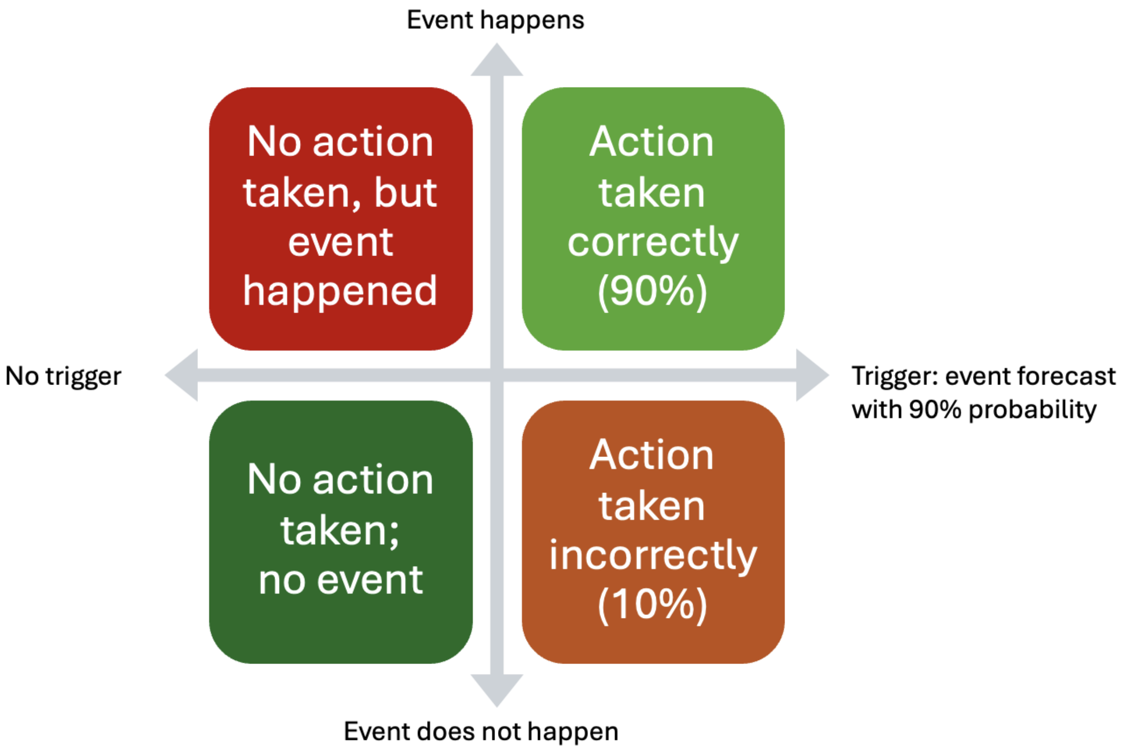

Figure 4: Possible outcomes of a binary (action / no-action) trigger with a 90% probability threshold. Note that we do not know either the base rate (frequency) of the event, or the confidence of a no-trigger forecast, which would be required to turn this into a cost-loss table. The bottom right outcome is sometimes called a "false alarm" and the top left a "missed event". Note also that many events are not binary in this way, so the vertical axis is more of a continuum than a binary outcome.

When action is incorrectly triggered, this is sometimes called a "false alarm". However, it is important to understand that **these "false alarms" are an inherent part of a working anticipatory system**. There cannot be quantitative anticipation without "false alarms", since we do not have access to a perfect forecast. The language of “false alarm” unhelpfully implies something negative, and to be avoided. Institutionally, many organisations are very sensitive both to false alarms and to missed events. Many respondents mentioned examples when actions had either been triggered incorrectly or failed to trigger, for various reasons (I1, I3, I4, S11, S17, S19, S29, S34), and the difficulty of explaining the perceived failure of the scheme to react correctly. In the longer term, false alarms, missed events, and their consequences erode **trust** in a system (I4, S11, S19 and see discussion in Section 4.4), which reduces its effectiveness. Some potential options for dealing with these trade-offs are outlined below in Section 4.2.3.

The opportunity to take action is in the window between the minimum time needed for action and the maximum time available for a confident forecast, as shown in Figure 5a. However, for some hazards and possible actions, it may be that there is no overlap and in fact a gap, as shown in Figure 5b. Until recently, tropical cyclones largely fell into the latter category, but efforts to improve the speed of activation have created a short window of opportunity for effective action (e.g. Start Network 2023, CERF 2022). Rapid-onset hazards such as cyclones often leave only 2-3 days of lead time for action, meaning that decisions have to be made very quickly (S13) and principles or plans need to be outlined in advance (S7, S14, S19, I2, I3) to make this possible.

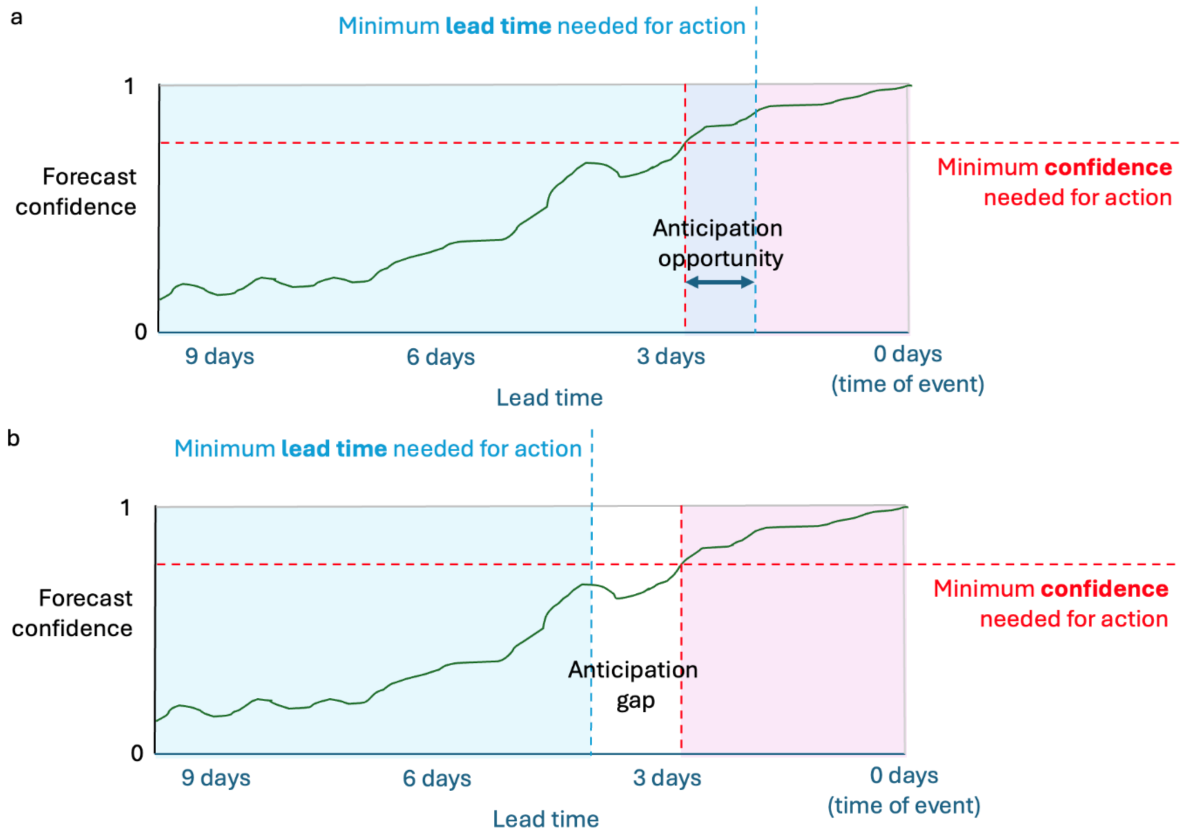

Figure 5: Sketch showing possible outcomes of the trade-off between confidence and lead time. Either there is an anticipation opportunity (panel a), or an anticipation gap (panel b). A gap can be turned into an opportunity either by speeding up the activation process (reducing the lead time needed for anticipatory operations), by reducing the level of confidence for the trigger (accepting a higher number of inaccurate triggers), or by improving the quality (confidence) of the forecast.

An anticipation gap can be turned into an opportunity in three ways:

1. By **reducing the lead time** needed for anticipatory operations. This depends on speeding up both activation processes and action. **Automated triggers** can help with the former, and predefined operational plans, situational awareness and operational readiness (e.g. prepositioned supplies) with the latter. Speeding up the lead time might also mean **restricting the scope of actions** to those which can be done effectively in a short time.
2. By **reducing the level of confidence** for the trigger (accepting a higher number of inaccurate triggers). We will discuss below some problems with accepting less accurate triggers but note here that low-confidence triggers could effectively be used to support **low-cost and low-regret activities** which are less time-sensitive, such as risk awareness programmes.
3. By **improving the quality of the forecast** in the trigger-relevant aspects. Though actual weather forecasting is outside the scope of AA, there is significant improvement of forecast skill over time (reflected by the increasing interest in AA), which is not quantitatively evaluated and tracked except for a small number of hazard types (e.g., Coughlan de Perez et al, 2018, for extreme heat). This could be done more comprehensively for other types of events. Improvements in forecast quality require research and development, which could be targeted by collaboration between scientific and humanitarian communities.

It is *essential* for AA schemes to have an understanding of the quantitative relationship between forecast skill and lead time (even where forecasts are presented in a deterministic format) for the particular forecast, hazard and location targeted. Without this, the expected utility of a scheme and particular action plans cannot be understood or planned effectively.

Overall, AA would benefit from a clearer understanding of expected levels of skill for different lead times, broken down by event types and geographical regions (similar to the work of Coughlan de Perez et al, 2018, for extreme heat).

3.2 A trade-off of accuracy with simplicity

There is a very clear practical understanding among our interviewees that **simplicity** (particularly at the point of interface of decision-makers with the forecast system) is necessary for AA to work effectively, both for technical manageability and for communication/explanation of anticipation decisions (e.g., I1, I3, I6). Achieving simplicity in information systems, however, requires a trade-off with the **accuracy** (comprehensiveness, detail, and exactness) of the information.

Much of the valuable information present in data and forecasts is complicated in nature. The data we have access to as well as the confidence we have in it will depend on location, state of the climate system (El Nino/La Nina, pressure fronts), time of year, model source and much more. There are very valuable insights within that complexity, especially given the corresponding range of choices in the response activities. However, there is limited ability to use complicated information for practical reasons. As one interviewee explained, they were a "*proponent of simplicity. We should rather something simple that more people can* *understand than something fancy*", especially when scaled to other institutions and sectors (I1). Simplicity and consistency also facilitate easier updating, interoperability, and comparisons between projects, as well as institutional coordination and long-term support.

The strongest example of this is the emphasis on simplicity in trigger design. By construction, a trigger is a binary decision-point which takes complicated forecast information and condenses it in some way into a go/no-go decision on a particular action. The triggers themselves are preferred if they are simple[1] (Pichon 2019), with interviewees (I1, I5) flagging issues with buy-in and trust when a trigger criterion is opaque or not sufficiently related to on-the-ground experience: "*An obsession with making forecasts and triggers more accurate means that understanding is concentrated among very few people, so if you have a hiccup, the forecast and the trigger get blamed because people don't understand it*" (I1). Interviewees also cited institutional reasons for not wanting to construct multi-stage triggers, preferring fewer activation criteria to apply to more actions (I1). Where different actions require different lead times and incur different costs, for example in the case of different drought metrics, complexity was preferred (S17). In contrast, the initial stages of another scheme needed to prioritise simplicity in the interests of generating trust: "*We realised that [..] they would value what we were doing more with a process that they understood*." (I6).

This also applies to more complex catastrophe insurance mechanisms: **"***what we have learned in the last five to seven years is that the countries that we work with do not understand a modelled loss approach and that is a bigger barrier than the potential increased basis risk from a simpler approach*" (I6, and see further discussion below). The need for consistency between organisations also favours common-denominator triggers, datasets, and labelling, even if they are not the most fit for purpose (I1, I3, I6). This is not trivial – a trigger which is not shared and well-understood can feel arbitrary and disappointing (e.g., I1, I5).

This preference for simplicity is balanced by recognition of the need to distinguish different events through more complex, more accurate models. In the case of drought, for example, one respondent indicated that "*Using a combination of trigger definitions and data as inputs (delayed season onset, below average total [rainfall], observational data once season starts) enabled us to be more flexible and respond when various conditions warranted*" (S17) and that different kinds of drought would warrant different kinds of responses (S17). In another complex situation, although tropical cyclones are often so unpredictable that little or no action can be taken, there are a subset of storm events which are usefully predictable at longer lead times, if they are identified. While significant improvements to skill have been achieved over recent years (e.g. Heming, 2019), this is not always reflected in practices based on accumulated experience. Coughlan de Perez (2022) also highlight the heterogeneity of forecast skill across different regions (for temperature and precipitation forecasts), meaning that different systems need to be developed to take advantage of information in different contexts.

For some hazards, it may be that there is currently no overlap between the minimum complexity needed to adequately scientifically represent the event, and the maximum

[1] "Early action plans require clear triggers, validated nationally, to reduce uncertainty and remove bureaucratic barriers to action" (Pichon 2019)

complexity tolerable by decision-makers and affected communities. Complex and ill-defined events such as drought can fall into this category. There are two ways to navigate this gap:

1. **Co-production**, by means of a two-way dialogue between users and technical experts, may be able to negotiate an effective compromise: *"It's very important [..] among decision makers to create scenarios so that they anticipate for it and communicate it to the public. It's some kind of combination of the science and the behavioural context."* (I3)
2. **Technicalisation**. A more technical option is to move more of the decision-making into the complicated dataset, using the full richness of the forecasts for cost-benefit analysis and returning targeted advice - such as where to position aid given the tangle of possible future paths of a hurricane. This presents other challenges though, as it disenfranchises human decision-makers in favour of models, which may or may not hold the necessary contextual information or be able to take responsibility for the decision taken.

In practice both options might be taken simultaneously, using co-production to ensure that more technical models and decision tools are integrated carefully, their limitations understood, and contextual factors taken into account where possible.

In summary, evaluation should be routine and regularly updated to take into account new developments in forecast skill as well as additional data and recent events. Codesign of AA frameworks is needed to find the right balance between simplicity and accuracy; there is no "optimal" balance which will be right for all situations. Communication is key. Clearer narratives about the rationale and operation of AA frameworks could simplify explanations and strengthen understanding of key concepts, increasing the ability of the system to realise more potential benefits of complexity. Technicalisation and development of complex decision support systems may be useful in some cases, but cannot bypass the need for codesign.

3.3 A trade-off of adaptability with consistency

Some formal anticipation systems such as the Start Fund[2] are committee-based, relying on the synthesis of a decision-making group of experts to release funds rather than committing to a specific trigger. The primary advantage of qualitative mechanisms is **adaptability**: the ability to integrate multiple sources of information, to take account of contextual factors such as pre-existing needs, compound events, political situations, or even the state of the funding pot, without having to pre-plan a response without fuller information. In the case of the Start Fund, accountability is achieved through the rotating membership and representativeness of the committee, and the transparency of decisions. The adaptability thereby achieved means that events which are not necessarily modellable or forecastable can still be anticipated, such as "anticipation of disease outbreak" (Start Fund alert 790, Madagascar) or "anticipation of spike in cross-border displacement" (Start Fund alert 802, Costa Rica). Disadvantages of this system include potential lack of consistency in the use of information, slower reaction times, and cognitive biases in individual and group information processing (e.g. Pforr 2018).

[2] https://startnetwork.org/funds/global-start-fund/start-fund-anticipation

Conversely, predetermined triggers such as the IFRC Forecast-based Financing protocols (IFRC, 2022) have the advantages of being (potentially) faster to react and more consistent in the use of information and activation of anticipatory action, and the disadvantage of lack of adaptability.

Despite the benefits on both sides of this trade-off, one factor turns out to be particularly important: **predetermined triggers are a necessary precondition for passing on anticipatory financing risks to global markets, which require objective, quantifiable and consistent metrics**. As one interviewee put it, "*you can't design your trigger mechanism without understanding what the requirements are of the funding instrument*." (I1) Unlike committee-based triggers, the total financial risks entailed by forecast-based triggers can (at least in principle) be quantified, with some level of uncertainty reflecting the quality of data available. As such, when schemes become larger there is a tendency to move towards predetermined, quantitative forecast-based triggers - for example, the Start Ready Risk Pools (Start Network, Start Ready[3]) which formalise some anticipatory action that would previously have come under the committee-based Start Fund. Start Ready is therefore able to take out reinsurance to cover the possibility of activations using up all of the initial pot. In general, instances of committee-based triggers, which make decisions according to the judgement of people sitting in the committee (and hence less consistent but more adaptable) are usually publicly-funded or grant-based. By contrast, market-based instruments require clear, quantifiable, objective, and reproducible triggers which can be written into formal contracts (hence less adaptable, but more consistent).

This move towards financialization, and engagement with global markets for risk and capital, requires transactions to be based on contractual obligations rather than on a set of (contextually reinterpretable) moral or ethical principles for disbursement of a fund. As such, it requires quantification. We now proceed to discuss the aspect of basis risk in anticipatory risk financing.

4 Distinguishing Type I and Type II basis risks

Applied to anticipatory humanitarian action, **basis risk** refers to the mismatch between the action that actually happens and the action that was expected or needed (Harris and Cardenes, 2020; Clement et al, 2018). Basis risk is a problem, both because this mismatch prevents effective humanitarian action in single events, and also because humanitarian funds would like to be able to calculate risk profiles in advance, to diversify risk effectively across multiple events. If what happens is not what was intended or expected, then managing risk will be difficult.

Basis risk encompasses the failure modes of an anticipatory system, because it is linked to the technical ability to carry out its purpose. In addition, the basis risk summarises the "failures" that are observed or perceived by participants and beneficiaries of an anticipatory system,

[3] https://startnetwork.org/funds/start-ready

and so minimising basis risk is also critical for ensuring that a system is trusted and trustworthy.

In passing, we note a significant problem of what we might call "Type 0" basis risk resulting from failures of communication, when people don't understand what a system was actually supposed to do in the first place. "*The biggest part is around mismanaged expectations. [Countries] will assume that in a major disaster, they will get money from the policy, and if that doesn't happen, the trust goes*." (I6). However, as it is not directly related to information quality, this is outside the scope of the current paper, but Section 4.4 discusses the importance of trust.

4.1 Uncertainty of single events (Type I basis risk)

Anticipatory action is, by definition, based on the output of some kind of forecasting system. In this paper, we are interested in weather forecasts. As forecasts are not perfect, there is **uncertainty** about the actual outcome, and therefore the possibility of taking an action that is, once further information arrives in the form of the event happening or not, incorrect in hindsight. The consequence of this we label *Type I basis risk.* This forecast uncertainty at the point of action could only be fully eliminated by returning to a responsive mode, intervening only after the disaster is definitively observed. Anticipatory action is chosen when the value of being able to act in advance correctly most of the time is expected to outweigh the risk and cost of getting it wrong some of the time.

As discussed in Section 2.1, at an overview system level the rationale is that in expectation, the benefits outweigh the harm. These "errors" are then just part of a functioning system which operates probabilistically and therefore should not be viewed in isolation as failures.

Nevertheless, the harm of false alarms and missed events are real and significant. Missed events mean humanitarian needs are not met, at least unless a responsive backstop is in place (highlighted by I5). There are strong reputational risks to any scheme in these cases, and also if it pays out or "*cries wolf*" (S13) when no event occurs. Over time, this undermines confidence and depletes funding for real events.

4.2 Model error at the aggregate level (Type II basis risk)

At the level of a forecast-based anticipatory action or financing scheme, the risk that any particular event is wrong can be aggregated and the actions chosen such that even with missed events and false alarms, the scheme is on balance successful. However, the unreliability of forecasting exposes the scheme to another risk - that the model of the system used to generate forecasts and decide actions may be *systematicall*y wrong. Even on balance, the scheme may be taking "incorrect" decisions in hindsight, perhaps under-supplying aid or, in the case of a financing scheme, running the risk of going bust. This we label Type II basis risk.

This can be very challenging to deal with, because often the **model error**, the difference between the model description and a fair probabilistic description, is not quantified. It may not be known until the model or trigger scheme has been compared to further observations, or it may require extra analysis of the model against historical observations to identify. One interviewee dealt with the possibility in an ad hoc way by recalibrating trigger thresholds in order to achieve a certain level of in-practice activation rather than relying on the return period numbers the model claimed (I1).

For individual events, basis risk (Type I) is simply the possibility of taking the wrong action. Where a scheme covers multiple events, there can also be an aggregate basis risk (Type II), evident over time by looking at whether the forecast probabilities match up with observed frequencies.

Trade-offs can make it difficult to mitigate Type II basis risk. Higher **quality** forecasts or more carefully crafted schemes may mathematically minimise basis risk, but can also be complex and difficult to communicate - one interviewee described consciously choosing a scheme with higher basis risk on the grounds of an overriding need for **simplicity** (I6).

All stakeholders would benefit from clearer understanding of the difference between uncertainty and model error and the different kinds of basis risk that they imply, and the prospects for reducing and mitigating those basis risks.

4.3 The experience of basis risk for financial and operational stakeholders

Risk management for anticipatory finance and action is not a single solvable problem, because basis risks are not experienced in the same way by different stakeholders, as summarised in Table 2.

Financial stakeholders (first row, Table 2) experience basis risk as the difference between expected and actual financial commitments. For financial stakeholders such as insurance providers and risk pools, Type I basis risk can be minimised by risk pooling, by ensuring that both sides of the transaction are fully forecast-based, and indeed reduced to zero by ensuring that a reliable forecast (one that forecasts correct probabilities) is used. While there will still be uncertainty about individual events, in the long run they can be averaged out. In addition, all of the costs incurred in taking anticipatory action are clearly contingent on the forecast rather than the event itself. As long as there is no Type II (aggregate) basis risk then the financial outcomes are known and can be accounted for. Type II basis risk is potentially a larger problem for these stakeholders, and is often difficult to estimate quantitatively due to scarce data, but it can be dealt with in practice by writing short-term contracts, leaving the option for models to be updated in light of new information or events.

Practical/operational stakeholders (second row, Table 2) experience basis risk as the difference between expected and actual outcomes on the ground. Since every event and outcome is important, we can't meaningfully "offset" negative outcomes of one event (such as lives lost) with positive outcomes in another in the way that we can for purely financial costs. In addition, the outcomes experienced are necessarily a function of the actual event and not just the forecast. For affected communities (and humanitarian or governmental stakeholders who work on their behalf), the Type I basis risk could only be reduced to zero by a perfect (no-uncertainty) forecast, which is impossible to achieve. Even if there is no aggregate basis risk, and the forecast is reliable, the single-event basis risk is <u>always</u> important for affected communities.

	Single event uncertainty (Type I basis risk)	Aggregate model error (Type II basis risk)	Values prioritised
Financial stakeholders (insurance, risk pools, budget-holders, catastrophe bond-holders)	No problem – minimised (to zero) by portfolio design, risk pooling.	Potentially big problem, but difficult to assess a priori given scarce data. Dealt with through regular contract renewal cycles.	Accuracy Confidence Consistency
Practical/operational stakeholders (vulnerable communities, first responders)	Big problem (for **trust** as well as outcomes). In principle can be assessed, but lack of capacity to do so.	Not visible, because second-order to (or conflated with) the unsolved uncertainty problem	Adaptability Lead time Simplicity
Some humanitarian stakeholders take both roles simultaneously (AA implementation, national governments)	Multiple perspectives on same issue of uncertainty. Analysis often constrained by capacity.	Not visible, because second-order to (or conflated with) the unsolved uncertainty problem	All: must negotiate trade-offs

Table 2: The impact of basis risk on different groups of stakeholders in anticipatory humanitarian action, and the values that are prioritised by those groups. Impacts may be small (green), larger (orange/red), or invisible/difficult to assess (light yellow). While purely-financial and purely-operational stakeholders have clear value priorities (indicated in green), those who take both roles simultaneously have to negotiate conflicting priorities in the form of the trade-offs discussed in this paper.

This suggests that financial and operational stakeholders have a lack of alignment. Financial stakeholders are able to "solve" their problem of uncertainty but struggle to engage with operational perspectives on the non-financial problems that uncertainty causes – primarily key questions about building reputation, trust and confidence that schemes can support delivery of positive humanitarian outcomes as well as financial ones. Basis risk experienced by practical/operational stakeholders is more qualitative, since in addition to financial outcomes and real loss and damage, it includes subjective experiences, perceptions of success or failure and the ongoing trust in a scheme.

Trust is not completely outside the view of financial services, because it forms the basis for ongoing buy-in to a risk pool or renewal of a product such as insurance. Narrative-building about confidence therefore seeks to emphasise the average or expected outcomes. Increasing financialization and risk-pooling in the humanitarian space cannot overcome the fact that individuals and communities experience single events, not an average or expected value.

Returning to the scheme of six interconnected values described in Figure 1, we can identify (Table 2, last column) that while purely financial stakeholders tend to prioritise accuracy, confidence and consistency, and purely operational stakeholders tend to prioritise adaptability, lead time and simplicity, there is a conflict and trade-offs that occur for stakeholders who are simultaneously engaged in financial management and operational management, such as national governments and humanitarian organisations.

All stakeholders would benefit from clearer understanding of the difference between financial basis risk and operational/practical basis risk and the ways to reduce (where possible) or mitigate their effects.

5 Managing basis risks (Type I and Type II)

This section discusses three ways of managing basis risks: quantifying them (5.1); reducing them (5.2); and mitigating their effects (5.3). None of these can be done perfectly, since our risk models are not perfect, but we give recommendations for risk management. We also note that building trust is critically important (5.4) and that human resources in the form of connected technical expertise are often a limiting factor (5.5).

5.1 Quantifying basis risks

To use a forecast, it is essential to have an understanding of how good the forecast is likely to be: in more technical terms, determining how much **predictive skill** the forecast has. To come to that understanding, we either need long **experience** with using the system, or some kind of systematic **evaluation** process (which also requires sufficient historical **data**).

Predictive skill varies with the type of event, lead time, and geographical region among other factors, so the evaluation of a forecast must take into account these factors as well as the intended use of the forecast. Most hydrometeorological forecasts are evaluated scientifically for their average performance on specific hydrometeorological variables compared to past weather (Risbey et al. 2021) but these are not necessarily the outputs of interest to decision-makers in anticipatory action and disaster risk applications in a changing climate (Chaves-Gonzalez et al, 2022), especially for forecasts of extreme events (Stephenson et al. 2008).

Our data suggest that forecasts used in AA systems are not routinely evaluated for predictive skill in the decision variable (I1, I2, I4, S27) and AA system operators often do not have research capacity to evaluate skill in house (I1, I3, S27). Forecasts are therefore by necessity often chosen on the basis of availability and provider reputation rather than performance. The nuances of skill metrics and variation of skill with lead time and return period are noticed in operational impact (I1, I6, S27), but not well understood or communicated (I1). Average skill information is regularly published by global forecasting or forecast visualisation systems such as ECMWF[4], GloFAS[5] and the IRI Maprooms[6], but its importance and the need for tailored skill metrics are underappreciated by users.

Forecast users did emphasise, however, that *local* evaluation of forecasts is needed rather than global or averaged skill metrics, particularly for contexts such as small islands or mountainous areas where weather patterns can be highly localised (I4, S4, and see also Dookie, 2023) and skill is highly variable. Downscaling approaches and local models can provide more granular data, but require even more data for evaluation and testing – subject to the prerequisite investment in observing and data systems, this is potentially an opportunity for AI systems to contribute. The confidence level of seasonal forecasting is generally low (but again very variable), and without specific evaluations for a variable and decision (e.g., Giuliani et al. 2020), users were unclear whether seasonal forecasts could or should be used operationally (I1).

Instead of formal skill metrics, local contextual knowledge and experience are sometimes built into informal systems of expert judgement, for example to deal with known biases in a forecast (I4). This approach may depend on a single individual, leading to concerns about robustness and capacity.

Using forecasts without evaluation of skill will result in undesirable rates of false alarms and missed events, and potential misallocation of resources. Furthermore, skill for AA decision-relevant variables is likely to be worse than for hydrometeorological variables (see discussion about impact-based forecasting in Section 4.2.1).

For a limited number of event types, some decision-relevant evaluation has been performed and helps to inform decisions about where and how to implement anticipatory action. The first step is to define decision-relevant variables which are extractable from forecast data and usefully linked to impact (I4; see also Nissan et al 2017 and WMO Hazard Information Profiles (Murray et al. 2021)). For example, for heatwave this could include a heat index related to physiological impacts of extreme heat on the human body, which incorporates both temperature and humidity, or a measure of climatological extremeness relative to the normal heat experienced at a location, or a measure of cooling needs over a longer period. These different measures will support identification of different kinds of needs, for example the impact on outdoor rural labourers will be different from the impact on less mobile people (e.g., elderly, pregnant) in urban areas. The predictability of those variables can then be evaluated using any available datasets. The methods of Coughlan de Perez et al (2018), which are used to map predictability of extreme temperature events and define where early warning, anticipatory action and seasonal preparedness measures would be most effective, could be transferred to other contexts.

We have noted above that anticipatory action must use probabilistic forecasts (or assign implicit probability/confidence levels to deterministic forecasts), and therefore requires probabilistic evaluation. Skill metrics should be chosen for maximum relevance to the decision

[4] https://www.ecmwf.int/en/forecasts/quality-our-forecasts

[5] https://global-flood.emergency.copernicus.eu/technical-information/glofas-skill/

[6] https://iridl.ldeo.columbia.edu/maproom/index.html

question - for example, the use of root-mean-square error metrics rewards smoothness of forecasts and makes them less likely to capture extreme events well (Lopez-Gomez et al. 2023). Simple measures of association such as correlation or anomaly correlation scores also have undesirable mathematical properties. Instead, **proper scores** (such as information metrics, the Brier score, or Continuous Ranked Probability Skill Score) should be encouraged (Gneiting and Raftery, 2007). Further connection with the extensive statistical literature on scoring rules would be valuable, to ensure that chosen metrics actually reflect the intended outcomes of evaluation (especially where complex decision support systems or any forms of machine learning are employed).

Regular open publication of these metrics of forecast skill would both allow potential AA users to understand the strengths of different forecast systems *and* encourage forecast system developers to prioritise improvements that enable better AA decision making. Metrics and visualisation should be chosen to make some intuitive sense, for example (following Coughlan de Perez 2018, Fig 3), "*lead time (days) at which there is a minimum skill level for forecast of extreme heat, plotted by location*".

The following steps could be taken to achieve this outcome:

- Development of a consensus shortlist of basic decision-relevant forecast variables for anticipatory action, by hazard/event type.
- Development of a consensus shortlist of decision-relevant skill metrics for anticipatory action
- Publication and intercomparison of these decision-relevant forecast skill metrics for the identified variables.

5.2 Reducing basis risks

Reducing basis risk can be done through improvement of forecasts, which requires **data** for calibration and evaluation. Weather forecasts are continually evaluated and improved at global and local levels (see, for example, the suite of forecast evaluation charts provided by ECMWF[7]). But with particular reference to humanitarian contexts and decisions, the key opportunities are in targeted additional data collection for evaluation and improvement (4.2.1), and ensuring effective use of data to support decisions (4.2.2). A further opportunity to reduce basis risks is targeting specific data collection to ensure that specific triggers do not end up being highly influenced by a small number of data points (4.2.3).

5.2.1 Targeted data collection

The raw materials which allow us to do weather forecasting and disaster anticipation are observations. Even the most sophisticated weather models, which leverage state-of-the-art scientific understanding about how the system behaves, will be limited in areas where there is poor observational coverage (S4, and see also Pauley (2022) on the value of land-based observations). For some weather patterns and regions, such as extreme rainfall in small tropical islands, a well-placed observing station can be the difference between not knowing what is coming and having time to evacuate flood-prone canyons (I2). Local observational

[7] https://charts.ecmwf.int/ (Apply filter to show "Verification").

data is often lacking in disaster-prone areas, limiting the ability to down-scale global forecasts well (I4, S4, S31) or do appropriate forecasts of local weather patterns (I3) or even evaluate the quality of a new forecast model before it is trusted (S28).

Our informants, especially those involved in the technical work of coordinating and delivering actionable forecasts to decision-makers, were very clear about the need for more hydrometeorological data measurement (as were interviewees in Tozier de la Poterie (2023)). Several could pinpoint where they wished to place new observing stations, what hazard they would help them forecast, and had already begun the procurement process using the limited funds available (I2, I4). Digitisation, common data formats and shared openly-accessed data magnify the value that this information would deliver, especially as our computational capacity is increasing.

Deciding which datasets are most value-adding to collect is very context dependent. Projects which digitise and make available **historical observational records** (also called "data rescue"[8]) allow decision-makers to understand the local climatology and potential non-stationarity under climate change (S13, S14). Many respondents mentioned this: "*historical data about recurrence of the event and past impacts in function of its characteristics are also needed to make informed decisions*" (S19) and wanted "*more quality ground data of events to correlate with past events to characterise better the triggers*" (S23). Digitised historical observations can be used to run weather models retrospectively (hindcasts) and provide essential context for planning forecast-based actions - what might happen, when, where it can be predicted and what actions might mitigate the impacts. **Real-time on-the-ground data** becomes historical data in the future, but can also be used immediately to improve model forecasts of a region such as by correcting for biases in humidity and temperature (I4). Hydrological data particularly, including river gauge data for flood projections and dam-level data, is of irreplaceable value - one informant told a colourful story of a time when a model predicted a dry landscape but the ground truth was a significant flood (I5). Many respondents also discussed the value of high-coverage and high-resolution satellite and aerial data (S7, S9, S14).

There are technical and institutional challenges to be navigated in the scale-up of data collection. One is that the technology available may not be appropriate; for example, one interviewee described problems with autonomous observing stations developed in another global region which glitched when the larger rainfall levels that were routine in their region occurred (I2). Also, "*when weather stations are funded by different donors they may produce different data outputs difficult to integrate,* [and also] *maintenance of equipment after project completion (sustainability of funding) is an issue*" (S18). Another challenge is the in-country infrastructure; observing stations require power and internet connection if the data is going to be accessed real-time, which can be particularly difficult at times of natural disaster (I2). Overarching all of this is the ownership of data and funding entanglements. One particular recommendation was to support digitisation hosted by a national met service to mirror and integrate data from the region so it is accessible (I3) although some international NGO respondents mentioned "*access to national met service data is often incredibly difficult to get which prevents us from using their data for triggers.*"(S27)" and another mentioned that

[8] For example at https://datarescue.climate.copernicus.eu/

politically-challenging inter-country relationships can prevent coordination between national met services (I5).

Locally-led, ongoing funding for hazard-forecasting focussed observation networks designed to complement global systems could help reduce basis risk, along with digitisation and open data sharing strategies to further leverage existing observations alongside clear quality and uncertainty metainformation.

5.2.2 Using data more effectively

Above, we have made broad recommendations about what kinds of considerations may be relevant, what makes for useful models and appropriate treatment of uncertainty, but when these are put into practice, **the specifics depend on local context**. For this reason, locally-based expertise is often needed to make use of contextual information (S27)**.** This may be at the national level (e.g. national meteorological services), but respondents also discussed the value of further localisation (I1, I3), because weather patterns can vary so much within countries and because at a hyper-local context there is less institutional siloing between departments responsible for different kinds of data and decisions (I3). Local attention can also make use of insight from other sources, offering opportunities to integrate indigenous knowledge (S25) and the experience of populations living in a region. Again, this speaks to the need for collaboration and coordination between global and local initiatives, with an appropriate balance of leadership and funding.

Global datasets and forecasts (such as those from ECMWF) are useful and often widely or freely available, but naive application can result in misleading information that can be actively detrimental (I4). A NMS respondent suggested that funding for model development at the local and global level are comparably important: the global models are used to provide boundary conditions and to extend coverage for regional downscaled models and in-situ observations but were systematically biased (I4). That required local experience and datasets, as well as an understanding of the merits and weaknesses of the global model. Models and data sources produced outside a particular region may not be well-suited to capture local risks (Martinez-Diaz 2019). Global models can particularly struggle with the context of small islands, such that specialised meteorologists may have to make significant adjustments to capture local effects (I2, I4). Solutions either through development of local models or using AI methods for downscaling all rely on sufficient local observational data for calibration and evaluation – where observations are not available, the value of investment in further modelling capability is difficult to evaluate and likely to be minimal. Definitions can also be difficult to translate - what constitutes a drought in one context may not be applicable in another context (I1, I5, see also meteorological discussion in Lloyd-Hughes, 2014 and humanitarian perspective in Sandstrom, 2017).

To more effectively leverage the local and international expertise to create effective forecasting tools, we recommend **defining the aim and measures of success** before development based on stakeholder participation. The metrics used to calibrate and evaluate forecasts should be applicable where the forecast is to be used. This requires ongoing technical conversations and **co-production** between decision-makers and scientists with local and global focuses, opposing the tendency to produce something useful and “*leave it on the* *loading dock [and] say, there it is [and] walk away*" (Cash 2006). One respondent particularly recommended in-country self-managed working groups: it "*allows us to have international knowledge and local knowledge and cross-check*" (I5).

External expert institutions (mostly in the Global North) are often engaged to supply targeted climate information used in AA projects (Coughlan de Perez, 2022), from a diverse range of technical sources including academic institutions, technical consultancies, insurance brokers and catastrophe model vendors. Shifting the responsibility for co-production of decision-support information into an entity which has consulting and technical development responsibilities can create valuable resources and potentially improve the **quality** of a system, but **transparency** and **explainability,** both useful for deciding how to leverage information towards decision-making, may be lost (see FAIR guidelines for example, Wilkinson, 2016), and **relevance** may not be prioritised. Paywalled models are less accessible in some applications and can block collaborative improvements and honest evaluations. Inertia can perpetuate reliance on these models (I6), so their inability to continually improve can cause problems.

To maximise the long-term contribution of models, we suggest the following aspirational goals:

- Make models open access permanently (pro bono offers can be revoked). This begs the question of who should pay for development and maintenance of open access models, but the cost-benefit equation suggests that it would be an effective use of funding for end users.
- Make models with active leadership from experts close to usage scenarios, and ensure that local capacity for model operation is sustainably developed.
- Avoid building black box models into critical systems.

5.2.3 Reducing trigger sensitivity to varying models and scarce data

Just because a metric or measurement is available, does not necessarily make it a good or robust measurement. The use of single well-defined parameters or indices may obscure the existence of significant uncertainty, or divergence between alternate models. Triggers which depend on public indices such as the National Hurricane Center (NHC) estimate of hurricane central pressure may even put an uncomfortable feeling of responsibility on that provider, as described by one NHC forecaster who said in a media interview[9] that he had deliberately not looked at catastrophe bond payout schemes in order to "*evaluate the meteorological data without knowing what any financial consequences might be*."

Where data are scarce and thresholds binary in nature, it is also possible that the overall outcome may be very sensitive to a small amount of observational data. This possibility was highlighted by the non-payout of a catastrophe bond to Mexico following Hurricane Odile in 2014 and a reduced payout following Hurricane Patricia in 2015 (Franco, 2024). In both cases the eventual central pressure estimation was significantly influenced by unofficial observations taken by "hurricane chaser" Josh Morgerman (Muir Wood, 2017). As Muir-Wood notes, this creates obvious potential for conflict of interest anywhere that such

[9] https://www.univision.com/univision-news/latin-america/mexicos-disaster-bonds-were-meant-to-provide-quick-cash-several-catastrophes-later-that-hasnt-happened

mechanisms are used in data-scarce environments. For example, an insurance company or client could send someone specifically to make additional measurements and (in the absence of regulation) could choose to register the reading (or not) according to the consequences for their payout.

Even without a conflict of interest, the inability to constrain the actual central pressure with certainty (either due to model or data limitations) means that there could be technical disagreement about whether a mechanism “should” have triggered, which could lead to legal disputes or loss of trust.

To improve the basis for parametric contracts, we recommend that

- Where parametric triggers are used to determine large payouts, sensitivity to model structure and observational data should be considered. Ideally, only official measurements should be used and the official observing systems should be able to measure with sufficient accuracy and confidence to determine the outcome unambiguously.
- The wording of a financial contract will be clear about the resolution of any form of uncertainty – it is also important for this to be understood by all parties.
- Given the large sums of money involved in schemes such as catastrophe bonds, it is advisable for parties involved to check possible data-sensitivity and invest in upgrading observation systems so that the resolution of the contract is not in any doubt. Any reduction in observing capabilities potentially leaves parametric contracts at greater risk of manipulation.

5.3 Mitigating the effect of basis risks

5.3.1 Refining triggers and thresholds to get more operational benefit from marginal information

False alarms and missed events cannot be avoided. They can be minimised by taking only very high confidence forecasts, but this will reduce the amount of lead time available for action. Instead, we could try to **reduce the impact of false alarms and missed events** by AA system design, for example:

- **Graded triggers** which begin when an event is somewhat likely (Figure 2). Relatively low-cost actions to take at low-confidence, long-lead-time events (A) could include awareness, distribution of information, checking staff availability, reading through protocols, liaising with forecast service for more information. Higher-cost actions to take when confidence is higher (D) might be specifically targeted and could include setting up shelters, evacuating affected areas, distributing protective materials or activating community mechanisms. Graded triggers are already familiar in informal contexts, for example through the Watch/Warning/Advisory (terminologies differ across jurisdictions) systems for hurricanes and other extreme weather. Seasonal forecasting systems (eg for drought) can also make use of graded triggers according to confidence levels - the sketch would correspondingly have longer lead times, but the overall picture would look similar.

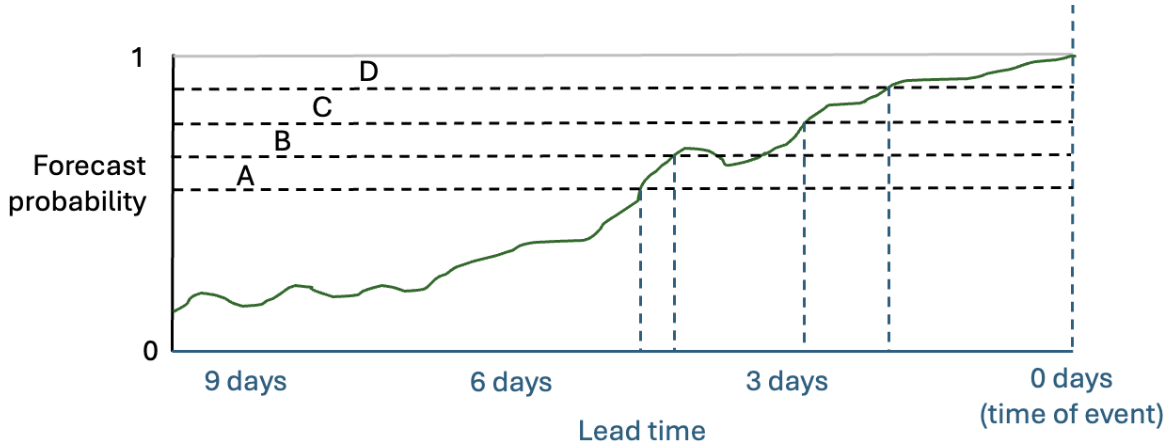

Figure 2: A graded trigger with levels A, B, C, D in increasing order of confidence. Lower levels of confidence (A) trigger "low-regret" actions of low cost and ongoing usefulness. Higher levels of confidence (D) trigger more expensive or difficult actions tailored to the specific event, but also generally come with shorter lead times for action. Lead time in days is only shown for ease of interpretation – this could be longer or shorter for different types of event.

- Allowing **withdrawal/deactivation of a trigger** (Figure 3). One respondent noted that an ability to deactivate a trigger would have been helpful (I1) for a slow-onset event, and this was also implied by some other responses mentioning false alarms (S13, S14). Another respondent also suggested the desirability to "*use stopping mechanisms once a trigger is met if new information indicates a change in direction*." (S17)

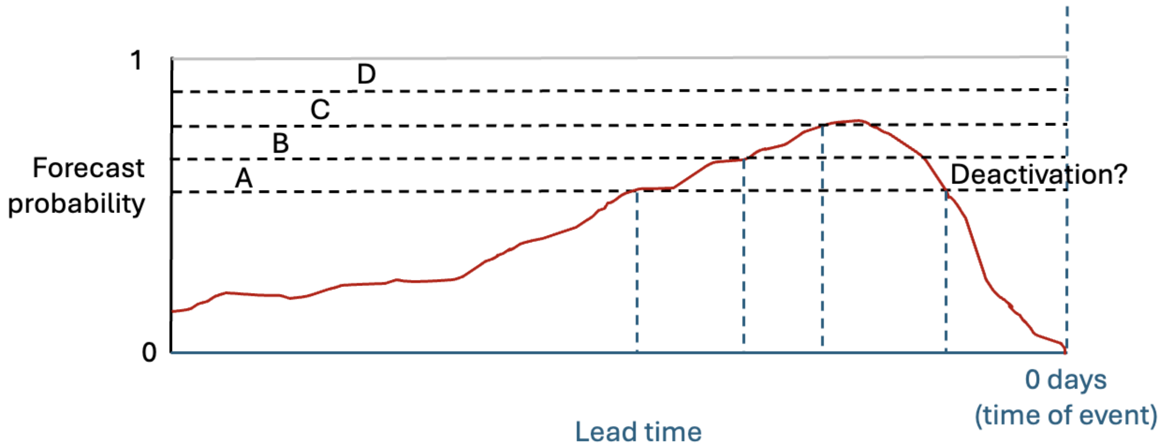

Figure 3: Deactivation of a trigger - conceptual diagram only. The conditions for deactivation would need to be considered carefully to avoid flip-flopping.

These possibilities could be appropriate for some more well-developed examples of AA which have enough experience to judge the costs and benefits of introducing additional complexity. For new schemes, the **practical need for simplicity** is likely to outweigh the efficiency benefits of these options.

5.3.2 Local data, calibration and trust

The "hurricane chaser" example above (4.2.3) highlights the need for reliable and official local data. Even more controversial scenarios could be envisaged, for example if data assimilation systems downweigh or ignore a "low-quality" local measurement and thereby prevent the triggering of some funding mechanism. The failure of the African Risk Capacity parametric mechanism to trigger for Malawi drought in 2016 (Reeves, 2017) was related to inappropriate data about crop planting used as input to the (also criticised) model. While it is easy to identify deficiencies in retrospect, if the aim is to get ahead of foreseeable model risks, then local data and local scrutiny are needed.

In some cases, the trigger is determined by a model provided by a commercial vendor (for example, catastrophe modelling companies). As many models are commercially sensitive, when the structure of the model is proprietary intellectual property and not made publicly available, it cannot be directly checked by others. As one interviewee noted, "*There is no alternative. We could get a different answer from a different model, but wouldn't know which is more valid."* (I6) - though this interviewee also noted that if the risk could be placed with an insurer, it had at least been approved by someone else with technical capability. Other interviewees, however, were stronger about the need to incorporate local information into such systems: "*Local judgments, expert knowledge and understanding and experience of the local context, it feels important for that to go back into these global systems, which otherwise can be very homogenising*." (I4)

In summary,

- If extra complexity is warranted by the benefits of doing so, parametric structures could seek to avoid cliff-edge thresholds (large payout above a threshold and zero payout below it) and instead have a graded or continuously-varying payout. This does not solve the problem of measurement uncertainty but it avoids some of the consequences by reducing the sensitivity of the payout.
- Where large-scale payouts are based on certain measurements, it would be beneficial to all parties to ensure (perhaps by providing funding) that observing systems are robust and sufficiently dense to be able to capture the event, with a focus on observing systems in areas where a larger impact could reasonably be anticipated, such as coastal cities.
- Where a third-party information provider (such as a local meteorological service, international meteorological service, or commercial model vendor) is involved in the determination of triggers and payouts, the basis of their model and the use of data should be transparent to ensure replicable and undisputable outcomes.
- Where a third-party information provider (such as a local meteorological service, international meteorological service, or commercial model vendor) is involved in the determination of triggers and payouts, the third party should be explicitly contracted to take responsibility for the trigger, or at least a resolution process specified in the event that information is withdrawn or manipulated.

5.4 Building trust is critically important

A key message from the framing of AA as a trade-off between confidence and lead time is the importance of **trust**, specifically trust that accepts and can cope with the false alarms and missed events that (without perfect forecasts) are an unavoidable consequence of gaining sufficient lead time to act in anticipation.

Since the design and operation of anticipatory action schemes are complex and involve many stakeholders, trust in the quality and alignment of the process is needed from all sides. With regard to the Type I basis risk identified above, having a trusted narrative about the design and operation of the scheme is the only way that (inevitable) individual failures can be rationalised and worked through.

Firstly, trust in hydrometeorological (hazard) forecasting systems is a precondition for (justified) trust in anticipatory action schemes. Trust does not require perfect information, but it does require relevant uncertainties to be understood accurately. From our interviews, we found that trust is primarily a function of **reputation** and **perceived expertise**, with quantitative evaluation not being regularly carried out by AA decision makers but either devolved to experts (S10, S13, S14, S20, S27) or assumed to be the role of the information provider only (S13, S14, S33, I6).

Next, the trust of intended beneficiaries in the overall AA system is critical. Respondents described the importance of engaging the public (I3), local communities (S7, S24) and traditional or indigenous knowledge (IS7, S25) to create a sense of **ownership**. Putting the system on a par with traditional or indigenous knowledge and forecasting systems, which also often have in-built uncertainty, can help to manage expectations (S17), rather than trying to oversell the ability of the system. Mismanaged expectations mean that high-profile failures can be confusing, embarrassing and reputationally damaging both for the system and for the implementing organisation as well as for anticipatory action more generally (I1, I4, I5, I6, S11, S19).

For some stakeholders, trust may be supported by self-corrections, for example the reassessment of a trigger after a false alarm or missed event, to “stop it from happening again”. But from a technical perspective, justified trust arises from a period of consistent operation long enough to develop experience and generate an archive of data for evaluation, to assess whether the frequency of false alarms and missed events reflects the chosen balance. Changing underlying conditions can make it hard to generate this kind of technical trust in an AA system. Physical hydrometeorological hazards may be changing due to climate change and impact varies with population-level changes; not all models can capture this, but ones that do add a lot of information beyond assuming things are unchanging.

In summary, with regard to the support of AA schemes, we highlight the following aspects of technical trustworthiness:

- There should be **transparency** about processes, including the expected basis risk.
- The chosen balance between confidence and lead time, and the balance between accuracy and simplicity, should be clear to all stakeholders - requiring a process of **communication** and **codesign**.

- The operation of the system should then reflect that balance - requiring the forecast model to have sufficient **reliability**.
- Local input should be part of **accountable** processes which oversee the design and implementation of a scheme.
- Schemes should consider whether stakeholders – particularly in terms of funding arrangements – and their capacity are better suited to an adaptable plan or a pre-arranged, consistent plan.

Trust is much discussed in other literatures, and the concept of trust in a sociotechnical system (like an AA mechanism) is at the forefront of debate about algorithmic accountability and the ethics of artificial intelligence (e.g. Andras et al, 2018). These debates are highly relevant for AA and worth further exploration in this context (e.g., van den Homberg et al.(2020) and Bierens et al. (2020)), but are beyond the scope of the current study. The increasing interest in artificial intelligence (AI) in this space reflects strong potential for improvement of weather forecast models, with evaluations for NVidia's FourCastNet (Pathak et al. 2022), Deepmind's GraphCast (Lam et al. 2023) and Huawei's Pangu-weather (Bi et al. 2023) already approaching the performance of numerical weather predictions from traditional physics-based models (Bouallègue et al. 2024, Lam, 2023). Leveraging new AI-based skill will be limited by the same concerns we raise here, with further challenges related to the transparency and accountability of "black box" mechanisms (Reichstein, 2024; Huynh,2023).

5.5 Human resourcing at local forecast-policy interfaces

A consequence of the needs and recommendations uncovered in this study is that to manage basis risk and support effective anticipatory action schemes, **more well-resourced technical experts are key**. In our conversations, these were often expected to reside with NMSs, which are seen as "*a credible and reputable source*" (S14) of quality and relevant information, with the responsibility for communicating it (I3, I5, S10, S13 S14, S20, S27) – but do not necessarily have expertise in financial products. More data collection, better leverage of forecast-relevant and impact-relevant variables, local downscaling and involvement, communicating across institutional boundaries, dataset discovery, evaluation and validation, uncertainty explanation and quantification – all of these are technical projects which will require the time of already under-resourced meteorological institutions. Some efficiency can be gained by digitisation or international coordination, but there is no shortcut to replace the need for expert work.

Interviewees flagged to us that "*the hydromet services are under-budgeted*" (S21) with uncertain funding streams (Vaughan et al. 2018) and the problem of high turnover in NMSs which limits the ability to use experience to contextualise and improve forecasts (I2). Resources were flagged as a main limitation preventing gathering wider types of data to move from hazard to impact-based forecasting (I3). We heard stories of individuals juggling the trade-off between rolling out a forecast in response to an ongoing need and doing careful evaluation of quality from a model to develop better capability in the long run (I4). Support needs to be a true capacity expansion: "*training activities for forecasters are extremely important, including attachments, etc. but there needs to be follow-up activities to ensure sustainability*" (S18). Challenges and recommendations related to the sustainability of observing systems can be found elsewhere (e.g. Grimes et al, 2022) - we note that for anticipatory action specifically, the local context is particularly important, not least because of the need to build trust through local relevance and solid performance.

We recommend that:

- To be effective and sustainable, funding has to support capacity for in-situ people and training, including innovative business and service delivery models (not just the delivery of models and datasets) which support individuals and teams to make sustainable ongoing use of data beyond the conclusion of short-term funding streams.
- Technical expertise at national and regional levels would benefit from greater peer-to-peer connections, for example between NMSs and other technical experts at different regional levels and within different institutions. Regional forums and collaborations led by local partners can help to ensure that funding is directed towards real local priorities.
- Data and model dependency can cause reliance on external institutions, so the long-term sustainability and reliability of that relationship should be taken into account when commissioning or purchasing data and models.

6 Conclusions

6.1 Key findings

Anticipatory humanitarian action is underpinned by weather and climate information which is subject to many kinds of uncertainty. We have identified three primary trade-offs in system design for anticipatory action: confidence with lead time; accuracy with simplicity; adaptability with consistency. Design of anticipatory action and anticipatory financing systems require difficult choices about balancing these factors to maximise value and gain the trust of all participants, and to reduce and mitigate basis risks. Uncertainty cannot be avoided. To be able to access the value in uncertain information, careful, quantitative attention must be directed towards characterising, evaluating and responding to it in ways that capitalise on opportunities and limit potential downsides.

The recommendations from the foregoing sections are summarised in Table 3 below.

Recommendations
Cross-cutting
Clearer narratives about the rationale and operation of AA frameworks.
Clearer understanding of the difference between uncertainty and model error and the different kinds of basis risk that they imply, and the prospects for reducing and mitigating those basis risks.
Codesign of AA frameworks is needed to find the right balance between simplicity and accuracy; confidence and lead time; consistency and adaptability. In each case, there is no “optimal” balance which will be right for all situations.

Technicalisation and development of complex decision support systems may be useful in some cases, but cannot bypass the need for codesign.

Quantifying basis risk

Evaluation should be routine and regularly updated.

Development of a consensus shortlist of basic decision-relevant forecast variables for anticipatory action, by hazard/event type.

Development of a consensus shortlist of decision-relevant skill metrics for anticipatory action.

Publication and intercomparison of these decision-relevant forecast skill metrics for the identified variables.

Reducing basis risk

Locally-led, ongoing funding for hazard-forecasting focussed observation networks designed to complement global systems

Digitisation and open data sharing strategies alongside clear quality and uncertainty metainformation.

Make models open access permanently.

Make models with active leadership from experts close to usage scenarios, and ensure that local capacity for model operation is sustainably developed.

Avoid building black box models into critical systems.

Where parametric triggers are used to determine large payouts, sensitivity to model structure and observational data should be considered.

Mitigating basis risks

If extra complexity is warranted by the benefits of doing so, parametric structures could seek to avoid cliff-edge thresholds (large payout above a threshold and zero payout below it) and instead have a graded or continuously-varying payout.

Ensure that observing systems are robust and sufficiently dense to be able to capture events subject to large financial contracts, with a focus on observing systems in areas where a larger impact could reasonably be anticipated, such as coastal cities.

Where a third-party information provider is involved in the determination of triggers and payouts, the basis of their model and the use of data should be transparent to ensure replicable and undisputable outcomes.

Where a third-party information provider is involved in the determination of triggers and payouts, the third party should be explicitly contracted to take responsibility for the trigger, or at least a resolution process specified in the event that information is withdrawn or manipulated.

Trust

There should be transparency about processes, including the expected basis risk.

The chosen balance between confidence and lead time, and the balance between accuracy and simplicity, should be clear to all stakeholders - requiring a process of communication and codesign.

The operation of the system should then reflect that balance - requiring the forecast model to have sufficient reliability.

Local input should be part of accountable processes which oversee the design and implementation of a scheme.

Schemes should consider whether stakeholders – particularly in terms of funding arrangements – and their capacity are better suited to an adaptable plan or a pre-arranged, consistent plan .
Human Resources
To be effective and sustainable, funding has to support capacity for in-situ people and training, including innovative business and service delivery models (not just the delivery of models and datasets). Technical expertise at national and regional levels would benefit from greater peer-to-peer connections. Data and model dependency can cause reliance on external institutions, so the long-term sustainability and reliability of that relationship should be taken into account when commissioning or purchasing data and models.

Table 3: Summary of recommendations from the foregoing sections.

In addition to these specific recommendations, we offer some key messages by way of conclusion.

First, the importance of meteorological forecast evaluation supported by local data. Anticipatory action is a sufficiently complex topic that opportunities for beneficial action can best be identified through robust quantitative evaluation (such as the methods of Coughlan de Perez, 2018 and 2022) - importantly, with enough local data to ensure that locally-relevant weather and climate features are well-represented.

Second, high-quality meteorological forecast performance is not enough for effective anticipatory systems. Wider technical evaluation of anticipatory schemes should take relevant humanitarian metrics as targets and be supplemented by qualitative consideration of the appropriate balance of trade-offs between financial and operational priorities. The decision of when and where to take anticipatory approaches should focus squarely on where the uncertainty in the forecast allows the trade-offs of lead time, confidence, accuracy and simplicity to be balanced and clear value to be created. The nature of that uncertainty, and codesign with beneficiaries, can inform whether an adaptable or consistent system is preferred. Development of trust in a system requires transparency, communication, codesign, reliability, local input, and accountable processes.

Third, we identified a tension between financial and operational perspectives in anticipatory action, related to the three trade-offs, which may explain why some aspects of system design can seem so intractable. Where financial stakeholders prioritise confidence, accuracy and consistency in order to minimise basis risk experienced over a lot of events, operational stakeholders prioritise lead time, simplicity and adaptability in order to best meet humanitarian needs during all individual events. Basis risks can be managed by quantifying them, reducing them, and mitigating their effects, but cannot be eliminated. Stakeholders who manage both operational and budgetary responsibilities (such as national governments and international humanitarian organisations) must negotiate conflicting aims and find an appropriate balance which promotes the ongoing financial and social sustainability of the system.

Fourth, there is a clear need for effective communication and support for training and capacity building. Investments in data collection and better models are foundational but are wasted without the sustainable ability to integrate and communicate information. Equally importantly, better and more widespread understanding of design principles for anticipatory action and financing structures would improve trust.

6.2 Policy implications: managing basis risks in practice

Building on the above key findings, we have four main policy recommendations.

First, that **continued locally-led investment in robust basic observing systems in under-resourced regions** benefits all aspects of anticipatory action: the data foundation of models; the ability to evaluate model and system performance; the human resource that underpins all successful schemes; and the development of trust.

Second, that **data-led technical evaluation is vital** for operational effectiveness, and must be supplemented with **practical evaluation of the contribution of systems to humanitarian priorities**, including perceptions of trustworthiness and effectiveness. In both cases, action on feedback from evaluation can identify and minimise emerging risks.

Third, that designing an anticipatory system involves balancing a set of trade-offs between operational priorities (adaptability, lead time, simplicity) and financial priorities (consistency, confidence, accuracy). **There is no "optimal" solution which meets the needs of all stakeholders**; instead, **participatory co-design can help to identify the appropriate choice of scheme parameters** to manage operational and financial risks.

Fourth, that **training and communication about the purpose, goals and design principles of anticipatory risk financing schemes helps to support participatory co-design** and ensure that basis risks are understood and can be managed appropriately.

7 References

Anand, V., Poole, L., Godziff, A., Bekele, B. and Coughlan de Perez, E., 2024. When does Forecast-based Insurance benefit? An Economic Analysis of Drought Risk Anticipatory Insurance. An Economic Analysis of Drought Risk Anticipatory Insurance (October 03, 2024).

Andras, P., Esterle, L., Guckert, M., Han, T.A., Lewis, P.R., Milanovic, K., Payne, T., Perret, C., Pitt, J., Powers, S.T. and Urquhart, N., 2018. Trusting intelligent machines: Deepening trust within socio-technical systems. IEEE Technology and Society Magazine, 37(4), pp.76-83.

Anticipation Hub. *Anticipatory Action in 2022; A Global Overview*. https://www.anticipation-hub.org/advocate/anticipatory-action-overview-report/overview-report-2022 (2022).

Anticipation.Hub. *Anticipatory Action in 2023: A Global Overview*. https://www.anticipation-hub.org/download/file-3995.

Ayers, J., Chadburn, O. & Hoque, K. A. Behaviourally-Informed Early Warning and Anticipatory Action: Unleashing the Potential of Locally-Led Pre-emptive Response and the Forecast-Based Financing it Requires. in *Current and Emerging Trends in the Management of International Disasters* (2024).

Balana, Bedru; Adeyanju, Dolapo; Clingain, Clare; Andam, Kwaw S.; de Brauw, Alan; Yohanna, Ishaku; Olarewaju, Olukunbi; and Schneider, Molly. 2023. Anticipatory cash transfers for climate resilience: Findings from a randomized experiment in northeast Nigeria. NSSP Working Paper 69. Washington, DC: International Food Policy Research Institute (IFPRI). https://doi.org/10.2499/p15738coll2.136812

Bi, K. *et al.* Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538 (2023).

Bierens, S., Boersma, K. & Homberg, M. J. C. van den. The Legitimacy, Accountability, and Ownership of an Impact-Based Forecasting Model in Disaster Governance. *Politics Gov.* **8**, 445–455 (2020).

Bouallègue, Z. B. *et al.* The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context. *Bull. Am. Meteorol. Soc.* **105**, E864–E883 (2024).

Cash, D. W., Borck, J. C. & Patt, A. G. Countering the Loading-Dock Approach to Linking Science and Decision Making. *Sci., Technol., Hum. Values* **31**, 465–494 (2006).

CERF. *Philippines: CERF Mid-Term Review - Super Typhoon Rai (Odette) Caraga and Southern Leyte, March 2022*. https://reliefweb.int/report/philippines/philippines-cerf-mid-term-review-super-typhoon-rai-odette-caraga-and-southern-leyte-march-2022.

Chaves-Gonzalez, J. *et al.* Anticipatory action: Lessons for the future. *Front. Clim.* **4**, 932336 (2022).

Clement, K.Y., Botzen, W.W., Brouwer, R. and Aerts, J.C., 2018. A global review of the impact of basis risk on the functioning of and demand for index insurance. *International Journal of Disaster Risk Reduction*, *28*, pp.845-853.

Coughlan de Perez, E *et al.* Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Nat. Hazards Earth Syst. Sci.* **15**, 895–904 (2015).

Coughlan de Perez, E. *et al.* Global predictability of temperature extremes. *Environ. Res. Lett.* **13**, 054017 (2018).

Coughlan de Perez, E. de *et al.* Adapting to climate change through anticipatory action: The potential use of weather-based early warnings. *Weather Clim. Extremes* **38**, 100508 (2022).

Dookie, Denyse S., Declan Conway, and Suraje Dessai. "Perspectives on climate information use in the Caribbean." Frontiers in Climate 5 (2023): 1022721.

Farkas, H., Linsenmeier, M., Talevi, M., Avner, P. & Jafino, B. A. The Economic Value of Weather Forecasts: A Quantitative Systematic Literature Review. *Forthcoming.*

Gettliffe, E. *UN OCHA Anticipatory Action. Lessons from the 2020 Somalia Pilot*. https://www.disasterprotection.org/publications-centre/un-ocha-anticipatory-action-lessons-from-the-2020-somalia-pilot (2021).

Griffiths, M. Opening Remarks by the Under-Secretary-General for Humanitarian Affairs and Emergency Relief Coordinator at the UN OCHA High-Level Event on Anticipatory Action. (2021).

Giuliani, M., Crochemore, L., Pechlivanidis, I. & Castelletti, A. From skill to value: isolating the influence of end user behavior on seasonal forecast assessment. *Hydrol. Earth Syst. Sci.* **24**, 5891–5902 (2020).

Gneiting, T. & Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *J. Am. Stat. Assoc.* **102**, 359–378 (2007).

Grimes, David R.; Rogers, David P.; Schumann, Andreas; Day, Brian F.. 2022. Charting a Course for Sustainable Hydrological and Meteorological Observation Networks in Developing Countries. World Bank, Washington, DC. http://hdl.handle.net/10986/38071

Harris, C. and Cardenes, I. (2020) 'Basis risk in disaster risk financing for humanitarian action: Potential approaches to measuring, monitoring, and managing it.' Centre for Disaster Protection Insight paper, Centre for Disaster Protection, London.

Heming, J. T. *et al.* Review of Recent Progress in Tropical Cyclone Track Forecasting and Expression of Uncertainties. *Trop. Cyclone Res. Rev.* **8**, 181–218 (2019).

Homberg, M. J. C. van den, Gevaert, C. M. & Georgiadou, Y. The Changing Face of Accountability in Humanitarianism: Using Artificial Intelligence for Anticipatory Action. *Politics Gov.* **8**, 456–467 (2020).

Huynh, B. Q. & Kiang, M. V. AI for Anticipatory Action: Moving beyond Climate Forecasting. *AAAI-SS* **2**, 78–84 (2023).

IFRC. World Disaster Report: Come Heat or High Water. (2020).

International Federation of Red Cross and Red Crescent Societies (IFRC). (2022). Anticipatory Action: Using forecasts to prevent hazards from becoming disasters. https://

www.ifrc.org/sites/default/files/2022-02/220203_IFRC_
Anticipatory%20Action_Brochure_final.pdf

InsuResilience. Linking Anticipatory Action to Risk Financing. (2023).

Lam, R. *et al.* Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421 (2023).

Lloyd-Hughes, Benjamin. "The impracticality of a universal drought definition." Theoretical and applied climatology 117 (2014): 607-611.

Lopez-Gomez, I., McGovern, A., Agrawal, S. & Hickey, J. Global Extreme Heat Forecasting Using Neural Weather Models. *Artif. Intell. Earth Syst.* **2**, (2023).

Murray, V. *et al.* Hazard Information Profiles: Supplement to UNDRR-ISC Hazard Definition & Classification Review:Technical Report. (2021) doi:10.24948/2021.05.

Nissan, H., Burkart, K., Coughlan de Perez, E., Aalst, M. V. & Mason, S. Defining and Predicting Heat Waves in Bangladesh. *J. Appl. Meteorol. Clim.* **56**, 2653–2670 (2017).

Pappenberger, F., Cloke, H.L., Parker, D.J., Wetterhall, F., Richardson, D.S. and Thielen, J., 2015. The monetary benefit of early flood warnings in Europe. *Environmental Science & Policy*, *51*, pp.278-291.

Pathak, J. *et al.* FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operators. *arXiv* (2022) doi:10.48550/arxiv.2202.11214.

Pauley, P.M., Ingleby, B. (2022). Assimilation of In-Situ Observations. In: Park, S.K., Xu, L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV). Springer, Cham. https://doi.org/10.1007/978-3-030-77722-7_12

Pforr, T. *Navigating Uncertainty: An Approach for Start Fund Decision-Makers | Start Network*. https://startnetwork.org/learn-change/resources/library/navigating-uncertainty-approach-start-fund-decision-makers (2018).

Plichta, M. and Poole, L. (2023). The state of prearranged financing for disasters 2023. Centre for Disaster Protection, London.

Plichta, M. and Poole, L. (2024). The State of Pre-Arranged Financing for Disasters 2024. Centre for Disaster Protection, London.

Pople, A., Hill, R. V., Dercon, S., and Brunckhorst, B. (2021) ‘Anticipatory Cash Transfers in Climate Disaster Response’, Working paper 6, Centre for Disaster Protection, London.

Rahman, F., Murshed, S.B., Salehin, M., Sakib, F.M. and de Perez, E.C., 2024. Does forecast-based financing (FbF) lower women's vulnerability to flooding?. *Progress in Disaster Science*, *24*, p.100389.

Reichstein, M. *et al.* Early warning of complex climate risk with integrated artificial intelligence. (2024) doi:10.21203/rs.3.rs-4248340/v1.

Risbey, J. S. *et al.* Standard assessments of climate forecast skill can be misleading. *Nat. Commun.* **12**, 4346 (2021).

Sandstrom, Sofie, and Sirkku Juhola. "Continue to blame it on the rain? Conceptualization of drought and failure of food systems in the Greater Horn of Africa." Environmental Hazards 16, no. 1 (2017): 71-91.

Stephenson, D. B., Casati, B., Ferro, C. A. T. & Wilson, C. A. The extreme dependency score: a non-vanishing measure for forecasts of rare events. *Meteorol. Appl.* **15**, 41–50 (2008).

Suarez, P. & Tall, A. Towards forecast-based humanitarian decisions: Climate science to get from early warning to early action. *Humanitarian Futures Programme. London, Kings College* (2010).

Start Network. START Fund Pilot: Anticipation of Cyclones Forewarn Madagascar. (2023).

Start Network. *Anticipatory Action for Cyclones in Madagascar: Evaluation of the Impact and Suitability of the DRF Model*. https://startnetwork.org/learn-change/resources/library/anticipatory-action-cyclones-madagascar-evaluation-impact-and (2024).

Tozier de la Poterie, A. T. de la *et al.* Anticipatory action to manage climate risks: Lessons from the Red Cross Red Crescent in Southern Africa, Bangladesh, and beyond. *Clim. Risk Manag.* **39**, 100476 (2023).

Vaughan, C., Dessai, S. & Hewitt, C. Surveying climate services: What can we learn from a bird's eye view? *Weather, Clim., Soc.* **10**, 373–395 (2018).

Wagner, M. (2024). 'Early Action: The State of Play 2023', Risk-informed Early Action Partnership, Geneva. https://www.early-action-reap.org/early-action-state-play-2023

Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).

WFP. WFP Somalia: Anticipatory Action & Integrated Climate Risk Management. (2023).

8 Appendix A: Survey Questions

1. Describe your role in anticipatory action (multiple choice):
    - Beneficiary (individual or community organisation)
    - Evaluator (internal MEAL or external consultant or academic)
    - Operations (involved in actually carrying out anticipatory action)
    - National disaster management agency
    - National hydrometeorological service
    - AA scheme design including financing
    - Other: __________

2. What anticipatory action scheme do you have most experience with? Please choose **one**, as you will refer to it in the next questions. If you wish, you can fill in the survey a second time referring to another scheme. (multiple choice)
    - Geographical region: Latin America and Caribbean, North America, Sub-Saharan Africa, Middle East and North Africa, Europe and Central Asia, South Asia, East Asia and Pacific
    - Weather-related hazard type: Cyclone/Typhoon/Hurricane or tropical storm, Flood, Flash flood, Storm surge, Drought, Heatwave, Coldwave, Dzud, GLOF, Wildfire, El Nino, other: _________
    - Action type: trigger-based Early Action Protocol or SOP, committee-based anticipatory funding, parametric insurance, regional risk pool, national disaster management, Early Warning System, other: _________

3. Thinking about this scheme, what information or experience supported your choice to use an anticipatory or early-action approach? (multiple choice)
    - Learning within the organisation
    - Learning from other organisations
    - Academic research
    - Support from national meteorological service
    - Expectation that it would be cheaper
    - Expectation that it would be more effective
    - Wanted to innovate/experiment
    - Other: ___________

4. Thinking about this scheme, are the weather forecasts you use deterministic (single best-guess forecast) or probabilistic (state the chance of different outcomes occurring)?
    - Deterministic
    - Probabilistic

- Don’t know
- Other: ___________

5. Thinking about this scheme, how do you know whether the forecast is reliable? Do you evaluate it yourself or rely on some other source of expertise? (free text)

6. Thinking about this kind of hazard (eg cyclone, flood, drought, etc), does it have any specific or unique challenges for anticipatory action? (free text)

7. If the scheme has been activated in response to an event, what have you learned about your decision processes? What worked well, and what problems were encountered?

8. Thinking about this scheme, what would be your highest priority for improving weather forecasts? (multiple choice)
    - more skill in the forecast (less uncertainty in the outcome)
    - more reliable forecast (better understanding of uncertainty in the outcome)
    - same forecast skill at longer lead time
    - forecast of different variables than the ones normally provided
    - more accessible presentation of forecast
    - timely digital accessibility of forecast data
    - none of these, because the weather forecast isn’t the main barrier to improving the scheme
    - Don’t know / Don’t yet have enough experience to be able to say
    - Other: __________

9. If you could transfer resources away from the improvement of the weather forecast and put them somewhere else to improve the outcomes of AA or DRF programmes, what areas would be a priority for you? (free text)

10. Thinking about this scheme, what additional data (if any) would you like to have access to, in order to enable more informed decisions? Please be specific, and feel free to mention as many kinds of data as you like. (free text)

11. Thinking about this scheme, were there any challenges you have experienced in the communication of forecast information between agencies? If so, what caused these challenges, and how could they be addressed? What worked well? (free text)

12. Thinking about this scheme, were there any challenges you have experienced in the financing or funding arrangements? If so, what caused these challenges, and how could they be addressed? What worked well? (free text)

13. In the same way that forecasts can be wrong and cause false alarms, estimated event probabilities can turn out to be wrong and result in risks to the financial sustainability of a

scheme. Do you have any thoughts on how to manage those financial risks? (free text)

14. What recent or forthcoming developments in anticipatory action are you most excited about? (free text)

Is there anything else you'd like to say?

9 Appendix B: Interview Protocol

These were semi-structured interviews, with the protocol below used as a guide to ensure that relevant topics were covered rather than a list of questions to be asked verbatim. Interviewees with different relevant experience were asked a different subset of questions.

Introduce interviewer, check that consent form has been signed, recap the purpose and format of the interview, ask if willing to proceed, switch on recording.

Questions:

1. Can you describe your involvement with a forecast-based disaster risk reduction scheme? What was it called, and what are/were your responsibilities?
   Interviewer to ask about type of hazard, country or geographical region, institutional backing, financing/funding (if known), etc
2. Tell me about how you came to learn about the forecast-based disaster risk reduction scheme (interviewer to use their words to refer to it, as given in previous answer).
   a. Did you help to design or commission the scheme?
   b. [if yes] What was your role in designing the scheme? What outcomes were you hoping to achieve?
   c. What information or experience supported your choice to use a forecast-based approach?
3. Tell me about the forecasts that you use.
   a. Where do you find weather forecasts for this event? Are they deterministic (single forecast) or probabilistic (state the chance of different outcomes occurring)?
   b. In your experience, are you (or would you be) able to take action based on the available information?
   c. How do you know whether the forecast is reliable?
4. Has the scheme been used for an active event?
   a. [if yes] Can you describe how the forecast-based scheme worked?
   b. How was the decision made to take action?
   c. Did it work the way you expected it to?
5. What did you learn from this event about your decision processes? What worked well, and what problems were encountered? (Decision support system)
6. If a better weather forecast could help the scheme to achieve better outcomes, what would be your highest priority for improving forecasts: for example, more reliable forecast, more skilful forecast, same forecast skill at a longer lead time, forecasts of different variables? Or something else? (Quality)
7. If you could transfer resources away from the improvement of the weather forecast and put them somewhere else to improve the outcomes of AA or DRF programmes, what areas would be a priority for you? (Relative importance)

8. Is the forecast information directly relevant for your decision? Do you have to make further assumptions, or process forecast data through another model? What is your view on the trade-off between the relevance of impact-based forecasts and the reliability of meteorological forecasts? (Relevance)
9. Thinking about this scheme, what additional data (if any) would you like to have access to, in order to enable more informed decisions? (Relevance)
10. From your perspective, were there any barriers to communication of forecast information between agencies? If so, what caused these challenges, and how could they be addressed? (Communication)
11. Were you involved with the financing element?
    a. [if not directly involved with financing] Where does/did the money come from to pay for forecast-based actions? Do you think the financing of the scheme is/was appropriate?
    b. [if directly involved with financing] What advantages and disadvantages have you seen in the way that the financing of this scheme works?

    Interviewer to probe for more detail here depending on responses.
12. In the same way that forecasts can be wrong and cause false alarms, estimated event probabilities can turn out to be wrong and result in risks to the financial sustainability of a scheme. Do you have any thoughts on how to manage those financial risks?
13. We would like to use your experience to help other people design better forecast-based disaster risk reduction schemes. What advice would you give?
14. What recent or forthcoming developments in anticipatory action are you most excited about?

Stop recording, thank interviewee for participation, tell them when to expect further information.