Decide Now or Wait for the Next Forecast? A Decision Framework Based on an Extension of the Cost-Loss Model

Stephen Jewson[1], Sebastian Scher[2], Gabriele Messori[2,3]

[1]Risk Management Solutions Ltd, London, UK
[2]Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
[3]Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Corresponding author: Stephen Jewson (stephen.jewson@gmail.com)

Keywords: cost-loss, forecast change, forecast volatility, decision making**,** expected utility, probabilistic forecasts, ensemble forecasts

Abstract

Users of meteorological forecasts are often faced with the question of whether to make a decision now based on the current forecast or whether to wait for the next and hopefully more accurate forecast before making the decision. One would imagine that the answer to this question should depend on the extent to which there is a benefit in making the decision now rather than later, combined with an understanding of how the skill of the forecast improves, and information about the possible size and nature of forecast changes. We extend the well-known cost-loss model for forecast-based decision making to capture an idealized version of this situation. We find that within this extended cost-loss model, the question of whether to decide now or wait depends on two specific aspects of the forecast, both of which involve probabilities of probabilities. For the special case of weather and climate forecasts in the form of normal distributions we derive a simulation algorithm, and equivalent analytical expressions, for calculating these two probabilities. We apply the algorithm to forecasts of temperature and find that the algorithm leads to better decisions relative to three simpler alternative decision-making schemes. Similar problems have been studied in many other fields, and we explore some of the connections.

1. Introduction

Meteorological forecasts such as weather forecasts, seasonal forecasts and climate forecasts can be used to inform decision-making in various ways (Stewart (1997), Palmer (2002), Fundel et al. (2019)). For all except the simplest situations, decision-making requires information about the probabilities of different future weather or climate outcomes, and this has led to the development of forecast systems that provide estimates of such probabilities. For example for weather forecasts of large-scale conditions, probabilities can be derived from the rates of occurrence of different situations among the members of an ensemble forecast, such as those produced by ECMWF (Molteni, et al., 1996), NCEP (Kalnay & Toth, 1993) and others. For site-specific weather forecasting, probabilities can be created from single numerical model forecasts, or from an ensemble mean forecast, by using linear regression with a specified distribution of errors as part of the calibration to observed values (Glahn & Lowry, 1972). More recently, improved site-specific probabilistic weather forecasts have been created from ensembles of numerical model forecasts using extensions of linear regression that incorporate information from the ensemble spread (Jewson, et al., (2004), Gneiting, et al., (2005)). In climate prediction, probabilities have been created from single model ensembles (Stainforth, et al., 2005) and multi model ensembles (Taylor, et al., 2012), via various different algorithms (Chen, et al., 2019).

In climate science the use of probabilities from forecasts to make decisions has been studied by Murphy (1969, 1985), using the cost-loss model. This model has been analysed, applied and extended by various other climate-science authors (Kernan (1975), Buizza (2001), Richardson (2001), Roulin (2007), Matte (2017)). The model is also used in other fields: for an example in economics see Chambers and Lambert (2015). The climate science cost-loss model is used as an idealized model for making types of decisions that are analogous in terms of logical structure to the binary decision of whether or not to cancel an event based on a forecast. In this simple model there are just two possible weather outcomes (which we will refer to as good or bad weather), with predicted probabilities, and a single binary decision that needs to be taken based on those probabilities (which we will refer to as cancel or go ahead). The four combinations resulting from the two weather outcomes and the two possible choices lead to different levels of benefit or harm, measured in the model using the concept of utility. Choosing the decision

that maximises the expected utility leads to the conclusion that the event organizer should cancel if the predicted probability of bad weather is above a certain threshold, where the threshold depends in a simple way on the parameters that define the utilities of the different outcomes.

However, knowing probabilities of future weather and climate outcomes may not always be enough information to make rational decisions. In particular in this article we investigate situations in which, in addition to probabilities of future outcomes, the decision maker requires information about the skill of subsequent forecasts and/or information about the likely sizes and nature of future forecast changes. This information is needed to inform the choice of whether to decide now or wait for the next forecast.

A specific example in which knowledge of the distribution of possible weather forecast changes is necessary for decision making was described in Jewson and Ziehmann (2004). More generally, consider the following idealized examples of situations where information about forecast skill and/or the size of forecast changes would seem to be relevant, from a purely intuitive perspective:

a) An event is planned for Saturday. If the weather conditions at the start of the event are unsuitable then the event will have to be cancelled, leading to various expenses, known as the 'loss' in the cost-loss framework. Daily weather forecasts are available in the run-up to the event and are used by the event organizer to decide whether to cancel in advance or not. Cancelling on Thursday leads to only small cancellation charges, while cancelling on Friday leads to larger charges. Both sets of cancellation charges are lower than the potential loss due to last-minute cancellation on Saturday, and this leads to a nuanced set of decisions around whether to cancel on Thursday, Friday or not at all. On Thursday, the organizer needs to decide whether to cancel (and take advantage of the lower cancellation charges) or wait for Friday's presumably more skilful forecast. If they wait, then on Friday they need to decide whether to cancel (and suffer higher cancellation charges) or go ahead and take the risk of the loss if the weather is bad.

b) A farmer is deciding which crop to plant in April. Seasonal forecasts for April are available in February and March and can be used to support the decision. There is an advantage to making a decision in February because the price of seed will be lower. But a decision in February instead of March has a greater risk of being the wrong decision, since the forecast is less skilful. Should they decide in February or wait until March?

c) A local government is deciding whether to build a coastal flood defence that will provide a certain standard of protection until 2100. Building now, based on current sea level rise projections, has the benefit of providing protection sooner but increases the risk that the sea level rise projections will change in the future and lead to the defence falling below the standard desired. Waiting for later sea level rise projections takes the risk of waiting for a longer period without defences but has the advantage of being able to use later, and presumably better, estimates of future climate. Should they build the defence now, or wait?

In all these examples rational decision making requires not only an estimate of the probabilities of future outcomes, but also an understanding of how those estimates, and their skill, might change with subsequent forecasts. These examples are idealised, and one could imagine factors that complicate the real-world decision-making situation, such as:

a) In reality, there may be forecasts available at greater frequency than daily, that allow a further option of cancellation late on Friday or early on Saturday, or the organizer may have the option to take out weather insurance to mitigate the loss if it occurs
b) In reality, buying seed early incurs a risk of spoilage, that should be factored in.
c) In reality, it may be possible to design a flood defence that can be built now and improved later if the projections worsen, although this is likely to be more expensive.

In fact, for real-world decisions, it is seldom possible to write down every factor that influences the decision, let alone code them all into a mathematical framework, and practically all actual decisions are ultimately made using a subjective evaluation based on multiple inputs. As a result, these examples should not be taken too literally. They nevertheless illustrate that the *decide now or wait for the next forecast* dilemma is an essential part of many decision-making situations related to weather and climate.

Similar decision-making problems arise in many other fields. For instance, economists have studied whether investors should buy a stock today or wait and see how the stock price evolves as new information becomes available. There is a general mathematical framework for solving this and other multi-stage decision problems in fields as varied as computer science and engineering known as *dynamic programming*, which was introduced by Richard Bellman in the 1950s (see e.g., Bellman (1957)). Dynamic programming is now often considered as a part of the larger field of *operations research* (Ventura, 2019). However, the concepts that arise in multi-stage decision making appear to be largely unknown within the climate science community. For instance, the well-known text book by Wilks (Wilks, 2011) discusses the basic cost-loss model in detail but makes no mention of multi-stage decision making problems or dynamic programming. We believe that a greater awareness of multi-stage decision making would lead to the development of more useful weather and climate forecasts. In an attempt to catalyse the climate science community to think about these issues we will derive our results by extending the climate-science version of the cost-loss model using arguments and terminology that will be familiar to climate scientists. We will explain the intuition from the perspective of someone accustomed to using weather and climate forecasts. In addition, we will then make connections with the broader theory and other applications, focussing on methods used in economics in particular.

In section 2 we describe the basic climate-science cost-loss model in more detail, in the context of the first illustrative example given above (that of an event organised for Saturday), within which the decision to make is whether to cancel the event or not. We then describe how this basic cost-loss model, which depends on a single forecast, can be extended to include the case of two forecasts. The extended cost-loss model is used to explore how to make a rational decision as to whether to cancel based on the first forecast or wait for the second forecast. We will show that the decision that maximises the expected utility is based on two probabilities derived from the forecast and its properties. To estimate these two probabilities for any given forecast one needs to know the skill of the forecast and the distribution of possible forecast changes.

In section 3, we consider the case in which the forecasts are normally distributed and well calibrated, which allows certain simplifications in the modelling of the two probabilities and leads to a straightforward implementation algorithm by which the two probabilities can be calculated and used to make the cancel-or-wait decision. We will see that parts of the mathematics are very similar to methods used in physics and in the economic modelling of stock prices.

In section 4 we test the implementation algorithm from section 3 using a long series of synthetic weather forecast data. The synthetic data is created in such a way as to capture the relevant

statistical structure of real forecasts. We find that, for this synthetic data, the average utility from applying the decision algorithm is always as high as, and often higher, than that from various alternative simpler decision strategies.

In section 5 we perform similar tests on real forecast data. Overall, the decision algorithm gives better decisions than the simpler decision strategies, although for some parameter settings it is beaten by the best of the alternative simpler methods.

Finally, in section 6 we summarize the results and discuss the implications for weather and climate forecasting.

2. Cost-Loss Modelling

2.1 The Basic Cost-Loss Model

The basic cost-loss model as used in climate science (Murphy, 1969) assumes that a probabilistic forecast is available which gives the probability of the two possible weather outcomes: p for bad weather and $1 - p$ for good weather. The forecast probabilities are assumed to be well calibrated (i.e., we assume they have been adjusted based on what can be learnt from past performance of the forecast system) and so can be taken as the best estimate probabilities we have, and do not require further adjustment.

To analyse the model, one can consider the different possible outcomes as a function of the choices that could be made by the event organizer. Each outcome has a probability, based on the forecast, and a utility, based on the definition of the problem. The probabilities and the utilities can be combined to calculate the expected utility for each of the organizer's possible choices, and the assumption in the model is that the organizer will opt for the choice with the higher expected utility. The utilities for each outcome are given in Table 1 and discussed below. The expected utility framework is not, in fact, the only framework that one can use for analysing options, although it is the framework we will use in this article: a brief description of other frameworks is given below.

To apply the expected utility framework, first, we consider the choice in which the organizer goes ahead with the event. In this case there are two possible outcomes, depending on the weather, which are given different utilities in the model: good weather (probability $1 - p$) leads to no cost and no loss, and so is given a utility of zero, while bad weather (probability p) leads to a loss, and so is given a utility of $-L$, where L is positive. The *expected* utility of going ahead with the event ($E_{go\ ahead}$) is the sum of each probability multiplied by the corresponding utility, giving $E_{go\ ahead} = (1 - p)(0) + (p)(-L) = -pL$.

Now we consider the choice in which the organizer cancels the event. In this case there are again two possible outcomes but this time both are given the same utility of $-C$, the cost of cancellation. The expected utility for cancellation (E_{cancel}) is therefore $E_{cancel} = (1 - p)(-C) + (p)(-C) = -C$.

Going ahead, and experiencing good weather, is given a utility of zero rather than a positive value that captures the benefit of running the event in order to reduce the number of parameters in the model. The loss of benefit caused by not running the event is implicitly included in the cost and the loss parameters C and L. Setting up the problem in this way, as is customary in the climate literature, does not affect the decision recommended by the method. This is discussed in more detail below.

If the organizer seeks to maximise their expected utility, then the decision to cancel would be taken if the expected utility of cancelling is greater than the expected utility of going ahead, $E_{cancel} > E_{go\ ahead}$, which gives $-C > -pL$. Rearranging this expression leads to $p > C/L$.

The conclusion is that for the organizer to maximise their expected utility they should cancel if the probability of bad weather is greater than a critical probability given by $p_{crit} = C/L$.
If C is greater than L, then p_{crit} is greater than one, and the event will never be cancelled because cancellation always has a lower utility than bad weather on the day. The interesting cases arise when $C < L$ and there is a tradeoff between cancelling and incurring the cost of cancellation, on the one hand, and not cancelling and incurring the risk of bad weather and associated loss, on the other. The cost-loss model is not particularly designed to be realistic, but rather to be the simplest possible mathematical formulation of a decision situation that captures the essence of this trade-off.

Connections to Economics

Models similar to the climate science cost-loss model have been studied in other fields, in which case the probabilities involved are not from weather or climate forecasts, but from some other source. For instance, economists have studied the question of whether to buy a stock today, given estimates of the probability of the stock price increasing or decreasing to certain levels tomorrow. See, for example, the discussion on page 80 of Chambers and Lambert (2015), or the discussion on page 95 of the textbook by Dixit and Pindyck (1994).
The mathematics in the Chambers and Lambert example is analogous to the climate science cost-loss analysis. For those who wish to make a detailed comparison: the probability of bad weather is replaced by the probability of the stock going down while the probability of good weather is replaced by the probability of the stock going up; cancelling the event is replaced by not buying the stock while going ahead is replaced by buying the stock, and the utilities of different outcomes are replaced by the one day profit or loss from buying the stock. One difference is that the monetary values in the economic example are all shifted upwards by an additive constant relative to the utilities used in the weather case. As a result, not buying the stock leads to no profit or loss, while in the weather case the analogous act of cancelling leads to a negative cancellation fee. In this way the economic example makes it clearer that having the option to choose whether to buy the stock or not itself may have positive value, as one would expect intuitively, in the sense that having any set of options to choose from is always better than having no options (all else being equal). The value of the option to choose whether to buy the stock or not is given by the expected value of making the trade, or zero, whichever is the larger. The value of the option cannot be negative, since if the expected value of making the trade is negative, the trade would not be made: this non-negativity of the option value corresponds to the idea that having more options can never be a bad thing. In the climate case the value of the option is the value conferred by having the choice as to whether to cancel or not, relative to the alternative situation in which there is no option to cancel. Numerically the value of the option to cancel or not is given by the difference between the expected utilities of cancelling and going ahead, or zero, whichever is the greater. This can be written as $max(pL - C, 0)$ and is greater than zero when $p > C/L$.
The Dixit and Pindyck example is not precisely analogous to the weather cost-loss situation, but illustrates what is essentially the same reasoning in a slightly different situation. The differences are that they consider the benefit from owning the stock to accrue indefinitely, from all future dividends, and they also discount future cash-flows to present day values. As with the weather cost-loss model, both these economic models are only stylized representations of the real world.

Option pricing paradigms

We have mentioned above that the expected utility framework is not the only framework available for option pricing. We now give a brief introduction to the broader topic of option pricing: some readers may wish to skip this section and continue with section 2.2 below.

We have stated that the option to buy a stock or not itself can be considered to potentially have value. This is because it may increase the expected utility relative to not having the option. Equivalently, in the weather case, the option to cancel or not can be considered to potentially have value, again because it may increase the expected utility relative to not having the option to cancel. In economics there are two main paradigms that are commonly used for deciding the best course of action when faced with an option, and for calculating values for options. One paradigm, which is the paradigm used in climate science for understanding the cost-loss model, which we have used above in our own description of the cost-loss model, and which we will continue to use in our analysis of the extended cost-loss model below, is based on the concept of expected utility. This paradigm assumes that the best course of action is the one with the highest utility on average, and the increase in expected utility due to having the option, if positive, determines the value of the option. The other paradigm assumes that the option is a tradable financial instrument, that the financial outcomes involved in the option can be replicated, or hedged, using other financial contracts, and that the market price of the option adjusts so that no arbitrage (i.e., no risk-free profit) is possible in the wider market place from trading the option. In this case the value of the option is determined using the theory developed by Black and Scholes (Black & Scholes, 1973), or one of its variants, which do not use expected utility of uncertain outcomes, but assume that uncertainty is hedged away (see standard textbooks such as Wilmott et al., (1995), Baxter and Rennie (1996) and Hull (2017)). The value of an option then depends on both the expected financial outcome of the option and the expected cost of the hedging and is in general different from the value calculated using the expected utility paradigm. In the problems we consider in this article we assume that the uncertainty in the financial outcomes due to the unpredictability in the weather forecast and in the weather cannot be replicated or hedged using any financial instruments, which is almost always the case at present, and which is why we apply the expected utility paradigm. However, we note that there is some (albeit limited) relevance for no-arbitrage pricing in applied meteorology. This arises because there is a small financial market for weather derivatives, including weather options, that can be used to hedge the financial outcomes due to weather in certain very specific cases. This market is currently mostly used for insuring against seasonal fluctuations in weather, although one might imagine that it could expand over time to cover other timescales of weather variability. Within this market, no-arbitrage pricing is relevant, at least to some extent since weather options (one kind of weather derivative contract) can potentially be hedged with weather swaps (another kind of weather derivative contract). The price of weather options can then be calculated as a function of the price of weather swaps using the weather derivatives version of the Black-Scholes equation (see Jewson et al., (2005), chapter 11, for details).

In addition to these two main paradigms for option pricing there is also a third paradigm, known as real options analysis, that uses the Black-Scholes type of no-arbitrage pricing methods for non-financial options that cannot obviously be hedged. These methods are sometimes considered controversial because the assumptions behind the Black-Scholes type models no longer hold true (Gupta, 2011).

2.2 Extending the Cost-Loss Model

We can extend the basic climate-science cost-loss model, staying with the expected utility framework, as follows. To make the explanation as readily understood as possible, we will continue to use our illustrative example based on an event organized for Saturday. We now

assume that *two* weather forecasts are available for Saturday, one on Thursday and one on Friday. The utilities for each outcome are given in Table 2 and discussed below. The decision framework we now derive applies equally well to other types of forecast and other time periods, such as weather forecasts from Monday and Friday for Saturday, or climate forecasts for 2050 produced in 2020 and 2035.

On Friday, the organizer faces the same decision as is described in the basic cost-loss model: whether to cancel Saturday's event or not. We will now write the utility of cancellation on Friday as $-C_1$, where the subscript 1 indicates cancellation 1 day in advance of the event, or, more generally, 1 forecast step in advance. The critical probability then becomes $p_{crit} = C_1/L$, and the organizer should cancel the event on Friday if the probability of bad weather exceeds p_{crit} as before.

We now, additionally, move backwards one step in time and imagine the organizer considering a weather forecast on Thursday, at which point they have the choice to either cancel there and then, or wait for Friday's forecast. This is the decision that we will now analyse in detail. Cancelling on Thursday leads to a cancellation utility of $-C_2$ and the interesting cases arise in this problem when cancellation on Thursday is cheaper than cancellation on Friday, which is in turn cheaper than last minute cancellation on Saturday ($C_2 < C_1 < L$). Cancellation on Thursday being cheaper than cancellation on Friday ($C_2 < C_1$) leads to a dilemma for the organizer, particularly when the weather forecast on Thursday (for Saturday) is looking bad, since there is now a trade-off for them between either cancelling on Thursday and benefitting from Thursday's cheaper cancellation fee or waiting for Friday to make a more informed decision. Waiting until Friday might, however, be followed by having to cancel on Friday and paying a higher cancellation fee than would have been paid on Thursday, or, worse still, might be followed by going ahead with the event and suffering bad weather, and an even higher loss.

From a mathematical point of view the decision on Thursday is complex because it may be followed by, and needs to take account of the possibility of, having to make another decision on Friday, and that second decision would be based on information (Friday's forecast for Saturday) that is not available on Thursday. To help make the decision we must analyse how much we do know already on Thursday about what Friday's forecast might be.

To analyse the trade-off involved in Thursday's decision using expected utility, we first define four probabilities, p_1, p_2, p' and $\hat{p}$. In the basic cost-loss model described above p is used to represent the forecast probability of bad weather on Saturday, as evaluated on Friday. In the extended cost-loss model, we will now write the same probability as p_1 to indicate a 1-day forecast. We will also define the forecast probability of bad weather on Saturday, as evaluated on Thursday, as p_2.

From the point of view of Thursday p_1 is now not a single probability but is a random variable and has a range of possible probability values which are all the possible values that Friday's forecast might take, given what we know on Thursday. For instance, if the forecast for Saturday, created on Thursday, is already saying that bad weather is very likely then p_2 will be known, and high, and we would already be able to predict that p_1 will most likely have high values, even though we would not know exactly the value it would take until Friday. Conversely if the forecast for Saturday, created on Thursday, is saying that bad weather is very unlikely, then p_2 will be low and we would already be able to predict that p_1 will most likely have low values, although again we would not know the exact value until Friday.

In this sense one could imagine creating a probabilistic forecast on Thursday for the range of values that p_1 might have on Friday, and indeed the simulation algorithm described in section 3 below and applied in sections 4 and 5 involves making just such a probabilistic forecast. This

probabilistic forecast captures how we think the probability of bad weather on Saturday will change from what we are predicting on Thursday, to what we might predict on Friday. From this probabilistic forecast for p_1, we will then evaluate the probability that p_1 will exceed the critical value p_{crit}, and we will call this new probability p'. Since exceeding the critical value leads to cancellation of the event on Friday, p' is the probability that we would cancel the event on Friday, as assessed on Thursday.

In the basic cost-loss model, if the event organizer chooses to go ahead, because $p < p_{crit}$, then there is still the chance that the weather will turn out bad during the event. This happens with probability p in that model. In the extended cost-loss model, we will again need to consider the chance that the organizer goes ahead but the weather turns out bad during the event, but we now need to evaluate it on Thursday so that it can form part of the basis for the decision to be made on Thursday. We will call this probability $\hat{p}$. From Thursday's point of view, going ahead, yet having bad weather, can arise from a range of values of p_1. For instance, we can imagine one case (on Friday) in which p_1 turns out only just below the threshold p_{crit}. In this case, the organizer would go ahead, but bad weather on Saturday is not that unlikely, since p_1 is still fairly high. On the other hand, we can imagine another case in which p_1 may be turn out far below the threshold p_{crit}, in which case bad weather on Saturday is more unlikely. $\hat{p}$ is the mean of the probability of bad weather over all such cases, conditional on going ahead, for different levels of p_1 in the range $[0, p_{crit})$. In summary, $\hat{p}$ is the probability, evaluated on Thursday, that if on Friday p_1 does not exceed p_{crit}, the weather on Saturday will nevertheless be bad.

Table 3 summarizes the definitions of p_1, p_2, p' and $\hat{p}$ for reference. The meanings of p' and $\hat{p}$ will become clearer in the context of the normal distribution example, discussed in section 3 below.

Connections to Economics

In the economics analogy, the equivalent question becomes should we buy the stock today or wait and decide whether to buy the stock tomorrow. See, for example, page 81 of Chambers and Lambert (2015) or page 96 of Dixit and Pindyck (1994). The Chambers and Lambert example is now no longer 100% analogous to the weather example, as it was for the basic cost-loss problem. Understanding the difference is a little complex: the issue is that the two choices available on Thursday (buy or not buy the stock) do not precisely correspond to the two choices available in the weather case (cancel or wait). We can see this by analysing cancellation on Thursday, which turns out to be neither analogous to not buying the stock on Thursday, nor analogous to buying the stock on Thursday. This is because cancelling on Thursday is a final decision, while not buying the stock on Thursday is not, since the stock can still be bought on Friday. Conversely, cancelling on Thursday leads to a fixed outcome, while buying the stock on Thursday leads to a random outcome. The logic of the two problems is just slightly different (and it is not surprising that different problems would have slightly different logic). Apart from these technical differences, however, the approach to solving the two problems would be exactly the same.

A more superficial difference between the two problems is that whereas in the weather cost-loss problem we refer to probabilities from externally sourced forecasts, Chambers and Lambert refer to beliefs about future success or failure of an investment, and represent those beliefs using probabilities. Whereas in the weather cost-loss problem we then think about how the external forecast probabilities may change, Chambers and Lambert think about how those beliefs may change in time due to the arrival of good or bad news.

Finally we note that Chambers and Lambert refer to problems in which decisions can only be made at one point in time, like the cost-loss problem, as *static* decision problems, and problems in which decisions can be made at more than one point in time, like the extended cost-loss problem, as *dynamic* decision problems.

2.3 Expected Utility Analysis

Given the definitions of p_1, p_2, p' and $\hat{p}$ we can now derive an expression for the expected utility of the two possible choices in the extended cost-loss model. The decision to be analysed in this case is the decision taken on Thursday as to whether to cancel or wait for Friday's forecast.

First, we consider the choice of cancelling on Thursday. This leads to a 100% chance of a utility of $-C_2$, and hence an expected utility of cancellation on Thursday ($E_{cancel\ Thursday}$) of $E_{cancel\ Thursday} = -C_2$.

Second, we consider the choice of waiting for Friday's forecast. Having waited to Friday, there are two outcomes: cancel on Friday, or decide to go ahead. These occur with different probabilities, which must be evaluated from the point of view of Thursday in order to feed into Thursday's decision. The first of these outcomes, cancelling on Friday, occurs if $p_1 > p_{crit}$, and incurs a utility of $-C_1$. From Thursday's point of view the probability of $p_1 > p_{crit}$ occurring is p' (by the definition of p' given above), and so the contribution of cancelling on Friday to the expected utility for waiting on Thursday is $-p'C_1$.

The second of these outcomes on Friday, deciding to go ahead, is more complicated since the utility is then affected by the weather outcome. Deciding to go ahead on Friday will only occur if $p_1 < p_{crit}$, which occurs with probability $1 - p'$. If the weather is good, the utility outcome is then zero, and the contribution to the expected utility is zero. If the weather is bad, which occurs with probability $\hat{p}$ (by definition of $\hat{p}$ given above) then the utility outcome is $-L$. The contribution to the expected utility of waiting on Thursday from going ahead on Friday is therefore $-(1 - p')\hat{p}L$. The probabilities in this expression can also be understood using the definition of conditional probability, which states that $p(a\ AND\ b) = p(a)p(b|a)$, and which we can apply here to say that the probability of going ahead AND having bad weather is equal to the probability of going ahead $(1 - p')$ multiplied by the probability of having bad weather, given that we have gone ahead $(\hat{p})$.

Based on the above considerations the overall expected utility of waiting on Thursday is made up of three contributions from the three possible outcomes that waiting on Thursday may lead to. These are: cancelling on Friday $(-p'C_1)$, going ahead and having good weather (0) and going ahead and having bad weather $(-(1 - p')\hat{p}L)$. Combining the three contributions to the expected utility of waiting (and noting that one of them is zero) gives a total expected utility of waiting ($E_{waiting}$) of $E_{waiting} = -p'C_1 - (1 - p')\hat{p}L$.

We have now derived expressions for the expected utility for both of the choices that present themselves on Thursday, and hence can proceed to the final step in the analysis, which is to compare the expected utilities of the two choices. If the organizer seeks to maximise their expected utility, then the decision on Thursday to cancel would be taken if the expected utility of cancelling is greater than the expected utility of waiting, in which case $E_{cancel\ Thursday} > E_{waiting}$, implying

$$-C_2 > p'C_1 - (1 - p')\hat{p}L \qquad (1)$$

If we fully understand our forecasting system, the forecast skill and how forecasts can change in time then we can calculate p' and $\hat{p}$, since they are just properties of the forecast. This inequality then determines whether to cancel or not, as a function of L, C_1 and the new parameter C_2. If we decide to wait, then come Friday the complexity of the decision on Thursday can be forgotten, and the decision on Friday can be made using the basic cost-loss model.

In the basic cost-loss model the decision to cancel depends on only one aspect of the forecast, the probability p (the probability of bad weather on Saturday). As a result, the decision can be expressed in the simple expression $p > p_{crit} = C/L$. In the extended cost-loss model the decision to cancel or wait on Thursday depends on two aspects of the forecast, the probabilities p' and $\hat{p}$. As a result, it is not possible to write the decision to cancel on Thursday in such a simple form. We can, however, derive some insight by rearranging equation 1 to give:

$$p' > \frac{C_2 - \hat{p} L}{C_1 - \hat{p} L} = p_{crit2}$$

The right-hand side (RHS) of this expression can be thought of as a new critical probability, p_{crit2}. Unlike the critical probability in the basic cost-loss problem, however, it depends not only on the parameters of the problem (L, C_1, C_2) but also on knowledge of the forecast and the forecast system (via $\hat{p}$).

Curiously, p_2 (Thursday's forecast for Saturday) does not appear explicitly in these two expressions. However, Thursday's forecast is in fact highly relevant, since p' and $\hat{p}$ can only be calculated given knowledge of Thursday's forecast. This will become more apparent in the normal distribution example given in section 3 below.

The above inequality for p' matches intuition in various ways: for instance if $C_2 = C_1$, then the RHS is 1, and we will never cancel on Thursday: we see that cancellation on Thursday is only rational if $C_2 < C_1$, which is when there is some early cancellation benefit. Also, if C_1 is very large, then the RHS is small, and cancellation on Thursday is more likely, to avoid the possibility of high cancellation costs on Friday.

In summary, we have derived an expression that solves the extended cost-loss problem of whether to cancel on Thursday or wait for another forecast on Friday. It depends on the calculation of two forecast quantities that are extensions of what is normally included in a probabilistic forecast. The first is p', the probability (evaluated on Thursday) that the probability (evaluated on Friday) of bad weather (on Saturday) exceeds a critical threshold. The second is $\hat{p}$, the conditional probability (evaluated on Thursday) of bad weather (on Saturday), given that the probability (on Friday) of bad weather (on Saturday) does not exceed the critical threshold.

p' and $\hat{p}$ can be considered as properties of a probabilistic forecast and forecast system. They are both functions of two dimensions: these dimensions are the threshold level (of e.g., rainfall, temperature or wind) that defines bad weather, and the threshold probability from the basic cost-loss problem applied to Friday's decision. In principal one could imagine routinely calculating numerical approximations to these two-dimensional functions every time a forecast is created. Values could then be read off to solve specific extended cost-loss problems as they arise.

In the next section we will consider the special case of normally distributed forecasts in which the calculation of p' and $\hat{p}$ becomes somewhat straightforward.

Connections to Dynamic Programming

In the analysis above we considered the outcomes that follow Friday's forecast, and then moved back one step in time and considered them from the point of view of Thursday. This stepping backwards in time is the essence of the dynamic programming approach to solving multi-stage decision problems (Bellman, 1957) and can be extended to solve problems that involve many time-steps.

3. The Normal Distribution Case

To illustrate the extended cost-loss model derived above in the simplest possible context, we will now consider a forecast system that produces forecasts consisting of normal distributions that are made on Thursday and Friday for Saturday. We will first discuss the statistical properties of these forecasts in some detail before presenting methods that can be used for applying the decision-making framework.

3.1 Forecast Properties

Since they consist of normal distributions, both Thursday's and Friday's forecasts can be described using a mean and a standard deviation: the mean represents the best single forecast, and the standard deviation represents the uncertainty around that forecast. For the forecast made on Thursday we write the mean and standard deviation as m_2 and s_2, and for the forecast made on Friday we write the mean and standard deviation as m_1 and s_1, where $s_1 < s_2$ since we assume Friday's forecast is more accurate on average. We write the observation for Saturday as a (for 'actual') and define the forecast errors as $e_2 = m_2 - a$ and $e_1 = m_1 - a$.

We will assume that the forecasts are well calibrated, by which we mean that they cannot easily be improved by further statistical processing based on past forecasts and past forecast errors. This leads us to make 3 calibration assumptions about the statistical properties of the forecast. The first calibration assumption is that the means of the forecasts are unbiased, and so $E(a|m_1) = m_1$ and $E(a|m_2) = m_2$. Taking expectations of both these expressions and using the law of iterated expectations gives $E(a) = E(m_1)$ and $E(a) = E(m_2)$ from which we can deduce that $E(m_1 - m_2) = 0$ and $E(e_1 - e_2) = 0$.

The second calibration assumption is that the standard deviations match the standard deviations of the actual forecast errors.

The third calibration assumption is slightly more complex. To introduce it, we first define the change in the mean forecast from Thursday to Friday as $\delta = m_1 - m_2$. Using the assumptions given above, $E(\delta) = E(m_1 - m_2) = 0$. We also note that

$$\delta = m_1 - m_2 = (m_1 - a) - (m_2 - a) = e_1 - e_2.$$

At the point in time that the decision is being made on Thursday, Thursday's forecast, and hence m_2 and s_2, are known. The details of how m_2 and s_2 are created are not relevant, as long as they satisfy the assumptions given above. For instance, s_2 could have been estimated simply from analysis of past forecast errors or could have been derived from a statistical calibration scheme that merges information from past forecast errors with information from the ensemble spread (Jewson, et al., (2004), Gneiting, et al., (2005)).

Friday's forecast, however, will not be known on Thursday. We do nevertheless need to be able to estimate s_1 already on Thursday in order to estimate the variance of δ, $V(\delta)$, since $V(\delta)$ is required for the algorithm described below. The simplest method for estimating s_1 on Thursday would be to use past forecast errors. Alternatively, one could investigate whether there might be information in Thursday's ensemble spread to help predict s_1 (i.e., to predict the uncertainty around the *next* forecast, given the *current* ensemble spread), although this has never, to our knowledge, been explored. Another approach would be to estimate $V(\delta)$ directly from the ensemble spread: this approach has been considered in Jewson and Ziehmann (2004).

In order to derive an expression for $V(\delta)$ we will assume, as the third calibration assumption, that the forecast error e_1 must be independent of the change in the forecast δ. The justification for this assumption is that if this were not the case then, on Friday, having observed the change from m_2 to m_1 (and hence the value of δ) one would have information about e_1 that would allow one to improve the forecast m_1. We are assuming that any such improvements have already been made as part of the forecast calibration process, and hence that there is no longer any information about e_1 contained in δ, and that δ and e_1 must be independent. Writing

$$e_2 = e_1 - \delta$$

we can take variances of both sides. Since e_1 and δ are independent, by the argument above, there are no correlation terms on the RHS, giving:

$$V(e_2) = V(e_1) + V(\delta)$$

and hence

$$V(\delta) = V(e_2) - V(e_1) = s_2{}^2 - s_1{}^2 \qquad (2)$$

and in this way, we are now able to calculate the variance of the change δ from the variances of the forecast errors. This variance is used in the algorithm described below.
A final set of assumptions we make, in addition to the assumptions related to good forecast calibration given above, is that the forecast errors e_1 and e_2, and the change in the forecast δ, are all normally distributed.

Connections to Economics

The assumption that e_1 and δ are independent is analogous to the efficient market hypothesis in economics, which says that stock price changes are unpredictable (see for example Wilmott et al., (1995), Baxter and Rennie (1996) and Hull (2017)). The argument for the efficient market hypothesis is that if they were predictable then trading would have already taken place that would have affected the stock price in such a way as to eliminate the predictability. Our argument for the independence of e_1 and δ is essentially the same, in that we assume that if they were not independent then calibration (instead of trading) would have already taken place that would have changed the forecast in such a way as to eliminate the non-independence.

3.2 Solutions for p' and $\hat{p}$

We now describe how we can calculate p' and $\hat{p}$ given a forecast with the properties described above, which can then be used to make the cancel-or-wait decision. We present two methods: the first is based on simulation, and the second based on numerical integration. The two

methods have pros and cons. The simulation method is intuitively simple and forms a framework that could be readily extended to encompass different distributions and more complex decision problems, such as including distributions to capture the uncertainty in the parameters s_1 and s_2. The numerical integration method, on the other hand, is faster and more accurate, but would be more difficult to extend. For the problems we study below in sections 4 and 5 we have found that the simulation method is both fast enough and accurate enough, given standard computing resources, but it could be that in the extension to multiple time steps simulation methods would be prohibitively slow and numerical integration might be preferable.

To define good and bad weather we will assume there is a given threshold value θ of the forecast variable that separates bad weather from good weather, with values higher than θ giving bad weather. An example would be temperature, where values above a given high threshold (i.e., a heatwave) may lead to the cancellation of the event.

3.2.1 Simulation solution

Given the forecasts defined above we now describe a simulation algorithm that can be run on Thursday for calculating p' and $\hat{p}$ for this forecast system. The simulation algorithm estimates p' and $\hat{p}$ in a conceptually straightforward way by simulating many possible versions of Friday's forecast for Saturday, given the information available on Thursday, and calculates p' and $\hat{p}$ from these many simulated forecasts.

We start by considering the forecast mean on Thursday, m_2, and, in the first part of the simulation method we model how the forecast mean might change from Thursday to Friday as m_2 changes to m_1.
Since we know that $E(m_1 - m_2) = 0$, we have $E(m_1) = E(m_2) = m_2$ (this latter step because on Thursday m_2 is no longer random but is fixed by Thursday's forecast) and we see that the distribution of possible values for m_1 will be centred around m_2.
We also know $V(\delta)$, the variance of the change in the forecast means, from equation 2 above, and we have assumed that δ is normally distributed. As a result we can model the distribution of values that m_1 might take on Friday as a normal distribution with mean m_2 and variance $V(\delta)$, which we write as $N(m_2, V(\delta))$. Each possible value of m_1 in this distribution corresponds to a possible probability forecast on Friday consisting of a normal distribution centred around that value of m_1 with standard deviation s_1. We are modelling a distribution of possible distributions for Friday's forecasts.

This leads to an algorithm that can be used on Thursday for the calculation of p' and $\hat{p}$, as follows:
1) Derive s_1 and s_2, either from analysis of past forecast errors, or from the ensemble spread. We will assume single values for both s_1 and s_2, rather than distributions, although the method could be generalized to deal with distributions of parameter uncertainty by simulating over different values for s_1 and s_2. Continue only if $s_1 < s_2$.
2) Calculate $V(\delta)$ using $V(\delta) = s_2{}^2 - s_1{}^2$
3) Given m_2, simulate Q values for m_1, using $N(m_2, V(\delta))$, where Q should be chosen large enough for good convergence of the results of the algorithm
4) For each of the Q simulated values of m_1, use the corresponding forecast $N(m_1, s_1{}^2)$ to calculate a value of p_1 (the probability of exceeding θ), using the Cumulative Distribution Function (CDF) for the normal distribution

5) Count how many of the Q values of p_1 exceed p_{crit}, to give R
6) Estimate p' as R/Q
7) For each of the $Q - R$ forecasts for which p_1 does not exceed p_{crit}, calculate the probability that the forecast variable exceeds θ, using the CDF for the normal distribution
8) Estimate $\hat{p}$ as the mean of these $Q - R$ probabilities

In this normally distributed case we see that p' and $\hat{p}$ are easily derived from the modelling of the distribution of possible future forecast distributions, which in turn is derived from an understanding of possible changes in the mean forecast, which in turn is derived from knowledge of the properties of the forecast errors.

Connection to Economics

The modelling of the change in forecast means in the description above can be described mathematically as a single step of a random walk, or as a type of Markov process known as a martingale. Similar models can be used to model Brownian motion (in physics), or the logarithm of the changes in stock prices (in economics). When modelling stock prices there is usually an additional component to represent systematic drift. In section 4 below we will also model the change from the forecast mean m_1 to the observed value a in the same way. A more detailed discussion of using random walks for both stock prices and expected weather outcomes is given in Jewson et al., (2005), and for stock prices in many standard textbooks such as Wilmott et al., (1995), Baxter and Rennie (1996) and Hull (2017).

3.2.2 Numerical integration solution

We now describe how p' and $\hat{p}$ can, alternatively, be calculated using analytical expressions and numerical integration, which would be faster and more accurate. We start by deriving an expression for p'.

On Friday, the probability of bad weather on Saturday is given by p_1. Given values for m_1 and s_1 (with s_1 assumed to be a single value) and using the assumption that the forecast consists of a normal distribution, p_1 can be written using the cumulative distribution function of the standard normal distribution Φ as one minus the probability of good weather, giving:

$$p_1 = 1 - \Phi\left(\frac{\theta - m_1}{s_1}\right) = \Phi\left(\frac{m_1 - \theta}{s_1}\right)$$

If $p_1 > p_{crit}$ then the event will be cancelled. The function Φ is monotonically increasing, and so this will occur for large values of m_1. Instead of using a threshold for p_1 we can therefore use a threshold for m_1 that we will call m_{crit}, defined by

$$p_{crit} = \Phi\left(\frac{m_{crit} - \theta}{s_1}\right)$$

Or the inverse:

$$m_{crit} = s_1 \Phi^{-1}(p_{crit}) + \theta$$

The decision to cancel can now be based on $m_1 > m_{crit}$, instead of $p_1 > p_{crit}$.

The probability density of m_1 given m_2, as evaluated on Thursday, can be given in terms of the probability density of the standard normal distribution ϕ, as

$$p(m_1|m_2) = \frac{1}{s}\phi\left(\frac{m_1 - m_2}{s}\right)$$

where we have defined $s = \sqrt{V(\delta)}$.

This is the distribution that we simulate from in step 3 of the simulation algorithm above.

The probability p' that m_1 will exceed m_{crit} is equal to one minus the probability m_1 will be less than m_{crit}. The probability m_1 will be less than m_{crit} is given by the CDF of the distribution of m_1, evaluated at $m_1 = m_{crit}$, and so:

$$p' = 1 - \Phi\left(\frac{m_{crit} - m_2}{s_1}\right) = \Phi\left(\frac{m_2 - m_{crit}}{s_1}\right)$$

This expression for p' can be used to replace step 6 of the simulation algorithm.

We now derive an expression for $\hat{p}$. The probability $\hat{p}$ is the probability that the weather is bad, even if the event goes ahead, and we will write this as $P(\text{bad} \mid \text{goes ahead})$. Using the law of total probability this can be decomposed into the integral over all possible values of m_1 as

$$\hat{p} = P(\text{bad} \mid \text{goes ahead}) = \int P(\text{bad} \mid m_1)\, p(m_1 \mid \text{goes ahead})\mathrm{d}m_1$$

Where P is used to indicate a cumulative probability and p to indicate a probability density. The first term inside the integral, $P(\text{bad} \mid m_1)$, is the cumulative probability of bad weather given m_1, and is given by the expression for p_1 given above. The second term inside the integral, $p(m_1 \mid \text{goes ahead})$, is the probability density of m_1, given that we go ahead. The probability density $p(m_1 \mid \text{goes ahead})$ is proportional to the probability density of m_1 given as $p(m_1|m_2)$ above in the range from minus infinity to m_{crit}, and is zero elsewhere. With renormalisation of $(1/(1 - p'))$ to ensure that it is a probability density, the probability density $p(m_1 \mid \text{goes ahead})$ is thus given by

$$p(m_1 \mid \text{goes ahead}) = \frac{1}{s(1 - p')}\phi\left(\frac{m_1 - m_2}{s}\right)$$

Putting these together in the integral gives

$$\hat{p} = \frac{1}{s(1 - p')}\int_{-\infty}^{m_{crit}} \Phi\left(\frac{m_1 - \theta}{s_1}\right)\phi\left(\frac{m_1 - m_2}{s}\right)\mathrm{d}m_1$$

This expression can be evaluated using numerical integration as replacement for steps 7 and 8 in the simulation algorithm.

4. A Synthetic Forecast Example

We now test the extended cost-loss decision algorithm using synthetic randomly generated forecasts and observations with appropriate statistical properties. There are two reasons for first using synthetic, rather than real, forecast data. First, by using synthetic data we can test the logic of the extended cost-loss decision framework and the normal distribution implementation

algorithm without also having to test whether any particular real forecast dataset fits the assumptions made in the derivation of the implementation algorithm. Second, by using synthetic forecasts we can derive results which are presumably as good as the results from the decision framework could ever be, because the data can be constructed so that the assumptions will be satisfied perfectly. These results can then be used as a benchmark against which to compare results from real forecast data. We use the simulation algorithm rather than the numerical integration solutions because we have the ambition to extend the model to allow for distributions for s_1 and s_2, and to forecast distributions other than normal.

4.1 Constructing the Synthetic Forecast Data Set

The synthetic data needs to satisfy various statistical properties in order to represent real forecasts sufficiently realistically. There are four conditions the synthetic data needs to meet: the synthetic forecasts and forecast errors need to be normally distributed, the forecasts need to be unbiased, the mean squared error (MSE) values have to be realistic in the sense that the one day forecast should on average be more accurate than the two day forecast, and finally e_1 and δ need to be uncorrelated, following the discussion in section 3.1 above.

To achieve these properties the synthetic forecasts and observations are created using the following steps, which work by first simulating Thursday's forecast, then Friday's forecast conditional on Thursday's forecast, and then the observations conditional on Friday's forecast. Simulating in this order makes it straightforward to create synthetic forecast data with the properties required.

1) Assign values to s_1 and s_2, the forecast error standard deviations, with $s_1 < s_2$
2) Calculate $V(\delta)$ using $V(\delta) = s_2{}^2 - s_1{}^2$
3) Simulate D values for Thursday's forecast mean m_2 using $N(0, s_2{}^2)$
4) For each value of m_2, simulate a corresponding value of Friday's forecast mean m_1 using $m_1 = m_2 + \delta$ where δ is simulated using $N(0, V(\delta))$
5) For each value of m_1, simulate a corresponding value of the observation a using $a = m_1 - e_1$ where e_1 is simulated using $N(0, s_1{}^2)$

That the synthetic forecasts generated in this way have the required statistical properties can be demonstrated as follows:

1) Friday's forecast is unbiased because $E(m_1 - a) = E(e_1)$, and e_1 is simulated with mean zero
2) Thursday's forecast is unbiased because $E(m_2 - a) = E(m_1 - \delta - a) = E(e_1) - E(\delta)$, and both e_1 and δ are simulated with mean zero
3) e_1 and δ are uncorrelated because they are simulated from independent normal distributions
4) The MSE of Friday's forecast is $s_1{}^2$ because e_1 is simulated with variance $s_1{}^2$
5) The MSE of Thursday's forecast is $s_2{}^2$ because $V(m_2 - a) = V(m_1 - \delta - a) = V(e_1) - V(\delta) = s_2{}^2$

We use the above algorithm to simulate $D = 10,000$ sets of two forecasts and one observation and we define 'bad weather' to be temperatures over a threshold defined by the 70th percentile of the observed temperature distribution. To give a practical interpretation of this definition, one could imagine an event for which high temperatures on the day of the event might lead to immediate cancellation for health and safety reasons and would incur a loss in terms of refunds to paying participants. As a real-world example, the 2019 New York city triathlon, due to be

held on the 28th July 2019, was cancelled at the last minute due to a prediction of a heat wave, and all participants were refunded their entry fees (CNN, 2019).

We use values of $s_1 = 3.79$ and $s_2 = 4.21$, giving $V(\delta) = 1.84^2$. These values are chosen to match the standard deviations of the real forecasts used in the next section. The values of s_1 and s_2 could be very different for different forecast variables, and for different weather, seasonal and climate forecasts.

We compare the average utility from applying the extended cost-loss model with that from three less sophisticated strategies, which are: always ignore Thursday's forecast and wait until Friday before making a decision using the basic cost-loss model (which we label as *always-fc1*); always decide on Thursday using the basic cost-loss decision model and then ignore Friday's forecast (which we label as *always-fc2*); and the more subtle strategy of using the basic cost-loss decision model on Thursday and then again on Friday if the event has not already been cancelled (which we label as *basic-twice*). The *basic-twice* method is the most similar to the extended cost-loss decision model that we have derived, but, from a theoretical point of view, neglects to take into account a proper analysis of the potential value of waiting for the next forecast when making Thursday's decision. This is taken into account in the extended model.

4.2 Synthetic Forecast Results

Results for C_1 in the range from 0.1 to 0.6 are shown in the six panels in Figure 1. In each case C_2 is varied such that the ratio C_1/C_2 takes values from 1 to 1.6, while L is always fixed at 1. *Extended* gives the best results, as would be expected from the theory and derivations given in the previous sections. For most parameter settings tested it gives higher average utilities than the simpler methods. In some cases one or more of the simpler methods give equal utilities, but *extended* is never beaten. The relative ranking of the other methods varies with the parameter values: no method is consistently in second place. This shows that *extended* does a good job of making use of the forecast data to make good decisions across a wide range of situations.

The results in Figure 1 can be interpreted in more detail as follows. We start by considering the results for $C_1 = 0.1$, which corresponds to cancelling on Friday being cheap relative to the loss that might be incurred on Saturday: this makes cancelling on Friday potentially attractive. When C_1/C_2 is 1 (the left-hand end of the horizontal axis) cancellation on Thursday is no cheaper than cancelling on Friday, and so there is no reason not to wait to Friday. As a result, always waiting to Friday to make the decision (*always-fc1)* gives as good results as the extended decision method. The other two decision methods perform less well, because they use Thursday's forecast, and this can only hinder making an optimal decision when cancelling on Thursday is no cheaper than cancelling on Friday.

For values of C_1/C_2 greater than 1 cancellation on Thursday becomes cheaper than cancellation on Friday, and the decision becomes a complicated one. All the factors now come into play: the various costs, the skill of the forecasts, and the logic by which any decision made on Thursday needs to take into account how the forecast might change between Thursday and Friday (and what decision that might lead to on Friday). As a result, *extended,* which is the only method that takes all these factors into account, beats the other three methods.

The ranking of the simpler methods varies with C_1/C_2. As C_1/C_2 increases from 1 to 1.1 *always-fc1* is soon overtaken by *basic-twice* and then by *always-fc2*. This is because *always-fc1* ignores Thursday's forecast, and this becomes increasingly unhelpful as cancelling on Thursday becomes cheaper.

For the limiting case of large C_1/C_2, *extended, always-fc2* and *basic-twice* perform more or less equally well, and only *always-fc1* performs badly. This is because cancelling on Thursday

becomes very cheap, and the whole decision effectively becomes a question of whether to cancel on Thursday or not. Only methods that allow for that can do well.

Considering the other panels in Figure 1: as C_1 increases, cancelling on Friday becomes more expensive. Up to $C_1 = 0.5$ the results are qualitatively the same as they are for $C_1 = 0.1$, although the margin by which *extended* beats the other methods is reduced. This is because as cancelling on Friday becomes more expensive, the value of making a good decision as to whether to wait until Friday or not reduces.

For $C_1 = 0.6$ the results from the *extended* and *basic-twice* are very similar, except for a small region near $C_1/C_2 = 1.0$. This is because cancellation on Friday is so expensive that waiting to Friday to make the decision makes little sense: the decision should be made on Thursday. The subtle logic that *extended* uses to decide whether to wait until Friday or not becomes more or less irrelevant.

Overall, we see from these results that *extended* always does well but has the most impact in the situations in which all the options are potentially reasonable and the decisions on Thursday and Friday are both trade-offs. In the limiting cases in which cancelling on Thursday or Friday is either cheap or expensive then the results from one or other of the simple methods are nearly as good as *extended*. In a real situation, without detailed analysis, one would not know whether the parameters are in a limiting case or not and hence always using *extended* would make the most sense because it is the only method that works well in all cases.

5 A Real Forecast Example

We now consider an example in which we apply the extended cost-loss decision framework derived in section 2, using the implementation algorithm described in section 3, to 30 years of real numerical model generated temperature forecasts from the European Centre for Medium Range Weather Forecasts (ECMWF) and station observations for Stockholm from the Swedish Meteorological Agency (SMHI). We use the same set of decision problems and parameter ranges as was used to test synthetic forecasts in the previous section.

5.1 Constructing the Real Forecast Data-Set

Our forecast and validation datasets consist of 36-hour and 108-hour ERA-interim forecasts and related observed temperature values from 1979 to 2018. We only consider the 92-day period from June 1st to 31st August in each year. We use 1979-1988 for calibration of the forecasts (giving 920 calibration cases) and 1989-2018 for out of sample validation (giving 2760 validation cases). We use forecasts with a wide spacing in lead time to magnify the impact of the method, so that the benefits can be seen more clearly above the noise in the results. The forecasts are calibrated using linear regression by regressing the observed values onto both the forecast and the previous forecast. The regression parameters are considered constant in time. We calibrate the mean of the ensemble while the standard deviation of the forecasts is derived from past forecast errors from the calibration period. This calibration method would be sufficient to ensure good calibration (as defined in section 3 above) if the forecast errors are genuinely normally distributed and homogeneous, and the calibration adjustments required are genuinely constant in time. In reality, the forecast errors are unlikely to be exactly normally distributed, may exhibit nonhomogeneity, and the ideal adjustments would likely not be constant time. Dependent on the level of misfit of these approximations, this might be expected to impact the effectiveness of the decision algorithm for this data set. The calibrated forecasts

consist of a mean and standard deviation for each of the two lead times, m_1, s_1 and m_2, s_2. The standard deviations of the forecasts, given by s_1 and s_2, are $s_2 = 4.21$ and $s_1 = 3.79$. From s_1 and s_2 we calculate the variance of δ as $V(\delta) = 1.84^2$. We again define 'bad weather' to be temperatures over a threshold defined as the 70th percentile of the observed data.

5.2 Real Forecast Results

We compare the average utility from applying the extended cost-loss model over the 2760 validation cases with that from the three less sophisticated strategies as before, and the results are shown in Figure 2. The results are very similar to those shown for the synthetic data in Figure 1. Overall, *extended* dominates the other methods. In this case however, *extended* does not give the highest utility score for every parameter setting. For example, for $C_1 = 0.5$ and $C_1 = 0.6$ we can see that for some values of the ratio C_1/C_2 *basic-twice* does better than *extended.* Conversely, for $C_1 = 0.3$ and $C_1 = 0.4$ we can see that *extended* exceeds the other methods to a greater extent than it does for the synthetic data. Since these variations in the results relative to the synthetic data go in both directions in this way, it seems reasonable to attribute these differences mostly to the noise in the results from having a smaller sample size than used to calculate the synthetic results. Other factors, related to deviations from the assumptions used in the method, may also play a role.

6 Discussion and Conclusions

Probabilistic weather and climate forecasts can be used as input to decisions in various situations. The basic cost-loss model is an idealized representation of a class of decisions which can be represented by the situation in which an event organizer has to make a forecast-based decision by considering the trade-off between the cost of cancellation of an event one day in advance, and the risk of going ahead with the event and the weather turning out bad and causing a loss. Analogous situations, with the same logical structure, appear in many aspects of forecast based decision making, whether using weather forecasts, seasonal forecasts or climate projections. They also occur in many other branches of science, engineering and economics.
We have generalized the basic cost-loss model by adding a previous day of forecast. The first decision is then whether to cancel two days in advance or wait for the 1-day ahead forecast. If a decision is made to wait, the second decision is then the same as that in the basic cost-loss model. Again, similar generalisations arise in other fields, and the general mathematical framework of dynamic programming provides an over-arching framework in which they can be addressed.
We have analysed the cancel-or-wait decision that needs to be made two days in advance using expected utility. The process of using a forecast to make this decision turns out to be different to that involved in the basic cost-loss model. In particular it requires the calculation of two new forecast quantities. One is p', the probability (evaluated two days in advance of the event) that the probability (evaluated one day in advance of the event) of bad weather will exceed a critical probably p_{crit} derived from the utilities of the different outcomes. The second is $\hat{p}$, the probability (evaluated two days in advance of the event) that, if (one day in advance of the event) we decide to go ahead with the event, the weather at the event will turn out bad.
These two quantities are non-trivial to calculate in general and require detailed analysis of the statistical properties of the forecast system. However, we have derived an implementation framework for the case in which forecasts and forecast changes consist of normal distributions. The framework is very similar to one particular application of dynamic programming used in economics, which is for the expected utility-based calculation of the pricing of financial options. In our framework the error statistics of the forecasts can be used to derive the variance of

forecast changes, and the variance of forecast changes can be used, two days in advance of the event, to run simulations of the possible distribution of probabilistic forecasts that will be available one day in advance. Based on this distribution of distributions, p' and $\hat{p}$ can easily be calculated. Solutions for p' and $\hat{p}$ based on numerical integration are also possible in this case. We have tested our extended cost-loss decision algorithm on synthetic forecast data. The algorithm worked as expected and gave decisions that are always as good as and for many of the parameter values tested clearly better than simple alternatives. This validates the logic behind the method and also validates the effectiveness of the implementation algorithm for the case where the forecast errors are genuinely normally distributed and well calibrated.

We have also tested the method on real forecast data. The extended cost-loss method outperformed simpler decision methods for most parameter values tested. However, for some parameter values it was beaten by a method that consists of applying the basic cost-loss model twice on consecutive days. Understanding why the method works slightly less well on the real data than on the synthetic data in some cases (but also slightly better in others) would require detailed investigation. In may be simply that the real data results, which are based on a smaller sample, are more affected by noise in the data. It may also be because some of the assumptions on which the method is based do not hold completely for the real forecast data.

The theory derived in Section 2 above applies to any type of forecast variable with any distribution, although we have only derived an implementation algorithm for forecasts with normally distributed errors. For non-normal forecasts, such as forecasts for daily rainfall, careful statistical analysis and modelling of the behaviour of forecast changes, and the development of an alternative implementation algorithm, would be required.

We have not considered including more than two steps of forecast in the analysis. However, such an extension can be achieved by adding further levels to the analysis while working backwards in time, in similar way to how we added the analysis of Thursday's forecast to the analysis of Friday's forecast when we extended the basic cost-loss model to the extended cost-loss model. This stepping backwards in time approach to solving multi-step decision problems is the essence of the general theoretical framework known as dynamic programming. A financial example of this approach is given in Dixit and Pindyck (1994), page 98.

In summary, this work demonstrates that there is a level of complexity that is needed in the logical interpretation of weather and climate forecasts that has perhaps not been fully appreciated before. In particular, we have shown that in a situation in which waiting for the next forecast is an option, the information in the probabilities in a probabilistic forecast is not enough to make a rational decision as to whether to wait or not and the forecast needs to be supplemented with additional information to help make that decision. In the case of well calibrated forecasts with normally distributed forecast errors, we have shown that the forecast mean squared error (MSE) at the relevant lead times is sufficient to provide this information. The MSE can be used to derive the variance of the size of forecast changes, which in turn can be used to determine whether to wait for the next forecast or not. For non-normal forecasts the picture is more complicated. For some distributions, there may be simplifying assumptions that can be used to derive simple algorithms as we have done for the normal distribution. In general, however, a full understanding of how forecasts change from one lead time to the next would be required.

One implication of this work is that, in order to realize the full potential of probabilistic forecasts for decision making, forecast providers may need to consider providing additional information along with the forecasts that they supply. Forecast users could then use that information to make more logical decisions around the question of whether to wait for the next forecast or not. Those decisions might be made subjectively, with the additional information

as inputs, or they might be made objectively using the extended cost-loss model we have described.

Acknowledgements

We would like to thank the anonymous reviewers for making very helpful suggestions to improve the article from its original version.

References

Baxter, M. & Rennie, A., 1996. *Financial Calculus: An Introduction to Derivative Pricing.* Cambridge: CUP.

Bellman, R., 1957. *Dynamic Programming.* Dover: Dover Publications.

Black, F. & Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy,* Volume 81, pp. 637-654.

Buizza, R., 2001. Accuracy and economic value of categorical and probabilistic forecasts of discrete events. *Monthly Weather Review,* Volume 129, p. 2329–2345.

Chambers, C. & Lambert, N., 2015. Introduction to Dynamic Belief Elicitation. *ACM SIGecom Exchanges,* Volume 14, pp. 80-83.

Chen, J. et al., 2019. Bias correcting climate model multi-member ensembles to assess climate change impacts on hydrology. *Climatic Change,* 153(3), pp. 361-377.

CNN, 2019. *The New York City Triathlon has been canceled because of the heat,* New York: CNN.

Dixit, A. & Pindyck, R., 1994. *Investment Under Uncertainty.* Princeton: Princeton University Press.

Fundel, V. J. et al., 2019. Promoting the use of probabilistic weather forecasts through a dialogue between scientists, developers and end-users. *QJRMS,* Volume 145, pp. 210-231.

Glahn, H. & Lowry, D., 1972. The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology,* Volume 11, pp. 1203--1211.

Gneiting, T., Raftery, A., Westveld, A. & Goldman, T., 2005. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. *Monthly Weather Review,* 133(5), pp. 1098-1118.

Gupta, M., 2011. *The Promises and Perils of Real Options Valuation Technique.: An in-depth analysis of Real Options Valuation technique, in Pharmaceutical R&D industry. By means of a case study research method.* Riga: Lambert.

Hull, J., 2017. *Options, Futures, and Other Derivatives.* London: Pearson.

Jewson, S., Brix, A. & Ziehmann, C., 2004. A new parametric model for the assessment and calibration of medium-range ensemble temperature forecasts. *Atmospheric Science Letters,* Volume 5, pp. 96-102.

Jewson, S., Brix, A. & Ziehmann, C., 2005. *Weather Derivative Valuation.* Cambridge: CUP.

Jewson, S. & Ziehmann, C., 2004. Using ensemble forecasts to predict the size of forecast changes, with application to weather swap value at risk. *Atmospheric Science Letters,* 4(1), pp. 15-27.

Kalnay, E. & Toth, Z., 1993. Ensemble Forecasting at NMC: The generation of perturbations. *BAMS,* Volume 74, pp. 2317--2330.

Kernan, G., 1975. The Cost-Loss Decision Model and Air Pollution Forecasting. *Journal of Applied Meteorology,* 14(1), pp. 8-16.

Matte, S., Boucher, M. A., Boucher, V. & Fortier-Fillion, T. C., 2017. Moving beyond the cost–loss ratio: economic assessment of streamflow forecasts for a risk-averse decision maker. *Hydrol. Earth Syst. Sci,* Volume 21, p. 2967–2986.

Molteni, F., Buizza, R. & Palmer, T., 1996. The ECMWF Ensemble Prediction System: Methodology and validation. *QJRMS,* Volume 122, pp. 73-119.

Murphy, A., 1969. On Expected-Utility Measures in Cost-Loss Ratio Decision Situations. *Journal of Applied Meteorology,* 8(6), pp. 989-991.

Murphy, A. H., 1985. Decision making and the value of forecasts in a generalized model of the cost-loss ratio situation. *Monthly Weather Review,* 113(3), pp. 362-369.

Palmer, T., 2002. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society,* 128(581), pp. 747--774.

Richardson, D., 2001. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Quarterly Journal of the Royal Meteorological Society,* 127(577), p. 2473–2489.

Roulin, E., 2007. Skill and relative economic value of medium-range hydrological ensemble predictions. *Hydrological Earth System Sciences,* 11(2), p. 725–737.

Stainforth, D. et al., 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature,* 433(7024), pp. 403-406.

Stewart, T. R., 1997. Forecast value: Descriptive decision studies. In: R. W. Katz & A. H. Murphy, eds. *Economic Value of Weather and Climate Forecasts.* London: Cambridge University Press, pp. 147-181.

Taylor, K., Stouffer, R. & Meehl, G., 2012. An Overview of CMIP5 and the Experiment Design. *BAMS,* 93(4), pp. 485-498.

Ventura, J. A., 2019. Dyanmic Programming. In: R. Ravindran, ed. *Operations Research Methodologies.* Boca Raton: CRC Press.

Wilks, D., 2011. *Statistical Methods in the Atmospheric Sciences.* 3 ed. London: Elsevier.

Wilmott, P., Howison, S. & Dewynne, J., 1995. *The Mathematics of Financial Derivatives: A Student Introduction.* Cambridge: CUP.

Tables

	Bad weather	Good weather
Cancel	$-C$	$-C$
Go-ahead	$-L$	0

Table 1: utilities for the basic cost-loss model. The rows represent the different choices, and the columns represent the different weather. Each combination of choice and weather leads to a value of the utility.

	Bad weather	Good weather
Cancel Thursday	$-C_2$	$-C_2$
Wait Thursday, cancel Friday	$-C_1$	$-C_1$
Wait Thursday, go ahead Friday	$-L$	0

Table 2: utilities for the extended cost-loss model. The rows represent the different choices, combined across Thursday and Friday. Thursday's choice is cancel or wait, and Friday's choice, if cancellation has not already occurred on Thursday, is to cancel or go-ahead.

p_{crit}	The critical probability in the basic cost-loss model, as used on Friday to decide whether to cancel Saturday's event.
p_1	The probability, evaluated on Friday, of bad weather on Saturday. When considered from the point of view of Friday p_1 takes a single value. When considered from the point of view of Thursday p_1 has a distribution of possible values.
p_2	The probability, evaluated on Thursday, of bad weather on Saturday.
p'	The probability, evaluated on Thursday, that on Friday p_1 will exceed p_{crit}.
$\hat{p}$	The probability, evaluated on Thursday, that if on Friday p_1 does not exceed p_{crit}, the weather on Saturday will nevertheless be bad.

Table 3: definitions of the different probabilities used in the extended cost-loss model.
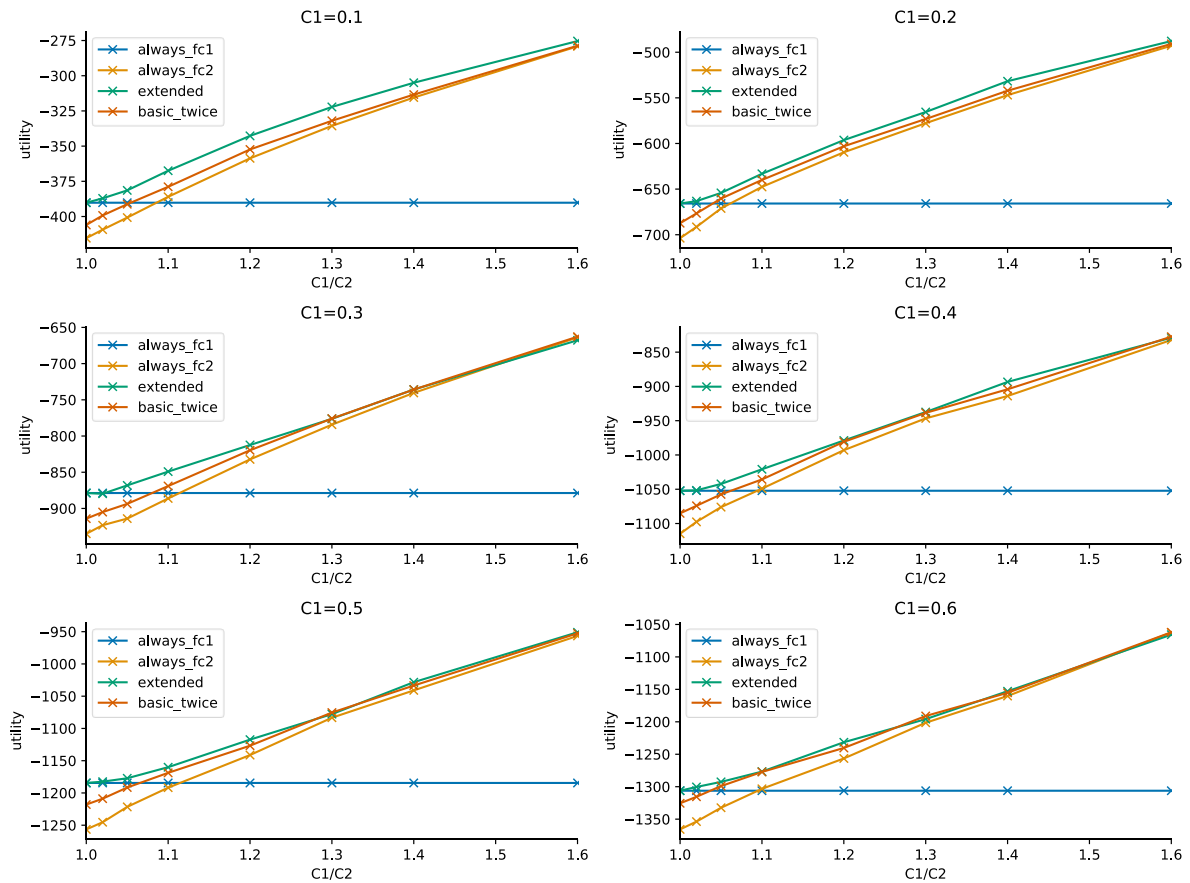
Figures

Figure 1: average utilities for a range of parameter values in the decision problem, calculated from synthetic forecasts, and compared across the four decision making methodologies. The blue lines show average utilities for the always-fc1 model, the orange lines show average utilities for the always-fc2 model, the green lines show average utilities for the extended model and the red lines show average utilities for the basic-twice model.
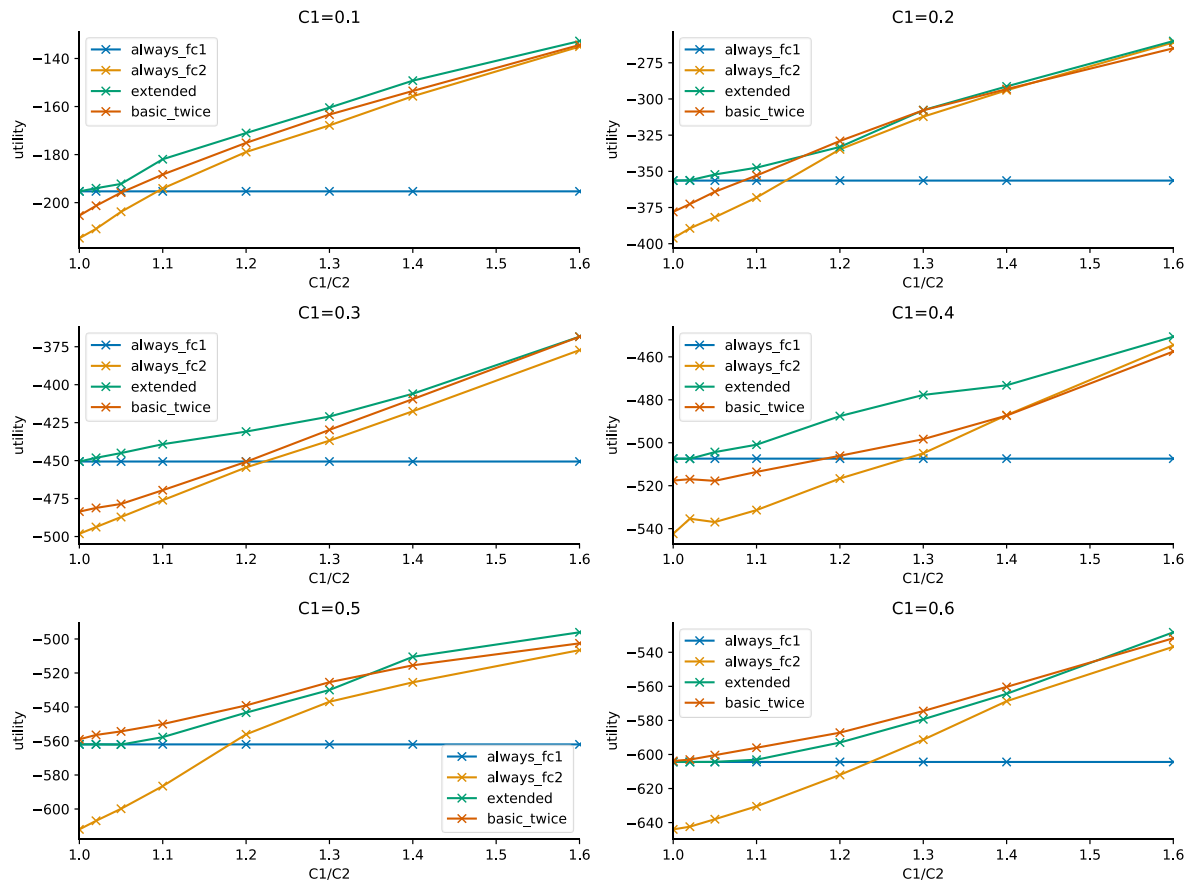
Figure 2: average utilities from the four decision making methodologies, for a real weather forecast data, for a range of parameter values. The colour-coding is the same as in Figure 1.