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)

Abstract

Users of meteorological forecasts are often faced with the question of whether to make a decision now based on the current forecast or 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. We derive an algorithm for calculating these two probabilities for the case of normally distributed weather or climate forecasts and apply it to a series of synthetic weather forecasts of temperature. We find that the algorithm leads to better decisions relative to three simpler alternative decision-making schemes.

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

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).

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 authors (Kernan (1975), Buizza (2001), Richardson (2001), Roulin (2007), Matte (2017)). The cost-loss model is 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 logical 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.

An industry-specific example in which knowledge of the distribution of possible 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 logical 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.

In section 2 we describe the basic 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 logical 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.

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 clearly better than that from various alternative simpler decision strategies, as expected.

Finally, in section 5 we summarize the results and discuss the implications for weather and climate forecasting. We also present some preliminary results from applying the method to real forecast data and discuss some of the additional challenges this creates.

2. Cost-Loss Modelling

2.1 The Basic Cost-Loss Model

The basic cost-loss model (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 has to 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 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.

We first 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. Instead, the loss of benefit caused by not running the event is implicitly included in the cost and the loss parameters C and L. This does not compromise the mathematical analysis.

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 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.

2.2 Extending the Cost-Loss Model

We can extend the basic cost-loss model 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 derive below 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, 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 ($C_2 < C_1 < L$). $C_2 < C_1$ leads to a dilemma for the organizer, particularly when the weather forecast on Thursday 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, lead to cancelling on Friday and paying a higher cancellation fee than would have been paid on Thursday, or, worse still, might lead to 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 lead to another decision having to be made on Friday, and that second decision would be based on information (Friday's forecast for Saturday) that is not available on Thursday.

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 likely have high values, even though we would not know exactly the value it would take until Friday. Similarly 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 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 implementation algorithm described in section 3 below and applied in section 4 involves making just such a probabilistic forecast. This probabilistic forecast would capture 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 could 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.

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

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

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 logical 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.

Given a full understanding of the behaviour of a forecast system, p' and $\hat{p}$ could be calculated: the obvious practical way to do that would be via statistical modelling of past forecasts and their errors. 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.

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 consisting of normally distributed forecasts made on Thursday and Friday for Saturday. We will first discuss the statistical properties of these forecasts in some detail before presenting an algorithm for applying the decision-making framework.

3.1 Forecast Properties

Since they are normally distributed, both Thursday's and Friday's forecast can be described as 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. 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 $E(m_1 - a) = 0$, $E(m_2 - a) = 0$ and hence $E(m_1 - m_2) = 0$ 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 $\delta, V(\delta)$, 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 then 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 hence that δ and e_1 are 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.

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

Given the above forecasts we now describe an algorithm that can be run on Thursday for calculating p' and $\hat{p}$ for this forecast system, which can then be used to make the cancel-or-wait decision. The algorithm estimates p' and $\hat{p}$ in a conceptually straightforward way by simulating many possible versions of Friday's forecast, given the information available on Thursday, and calculating p' and $\hat{p}$ from these many simulated forecasts.

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. We then consider the forecast mean on Thursday, m_2, and, in the first part of the algorithm 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 hence 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
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_1 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.

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 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.

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 $D = 2500$ sets of two forecasts and one observation and we define 'bad weather' to be temperatures over a threshold defined by the 95th percentile. 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.

We use values of $s_1 = 1$ and $s_2 = 2$, giving $V(\delta) = 3$. These values are chosen to be illustrative. For real one- and two-day temperature forecasts, the values of s_1 and s_2 are typically much closer. Using closer values gives qualitatively the same results, but longer simulations are required to demonstrate the benefits of the extended cost-loss algorithm. The actual values of s_1 and s_2 could be very different for different forecast variables, and for different weather, seasonal and climate forecast applications.

4.1 Base Case Results

To test the extended cost-loss decision framework we now specify the parameters of a number of decision problems, and for each set of two forecasts and one observation simulated above we calculate the decision implied by the framework. We also calculate the decisions that would be made by some simpler decision-making strategies, which the extended framework should beat if it is working as intended. We evaluate and compare the quality of the decisions from the different methods by calculating the average utility for each method over the entire set of forecasts and observations.

As a starting point we define a base case in which we define the parameters of the decision problem as $L = 1, C_2 = 0.05$ and $C_1 = 0.1$, giving $p_{crit} = C_1/L = 0.2$. Of the 2500 situations tested, the number of occurrences of each of the different possible outcomes from the decision algorithm are recorded in table 4.

To illustrate in more detail the functioning of the algorithm, we pick examples of each of the 4 utility outcomes from the 2500 situations and show them in table 5. In case 1 m_2 is high. This leads to a high value of p', which leads to cancellation on Thursday when the extended cost-loss decision formula is applied. In case 2 m_2 is high, but not as high as in case 1, and the event is not cancelled on Thursday. However, come Friday m_1 is higher than m_2, and high enough to lead to probabilities that lead to cancellation of the event when the basic cost-loss model is applied. In case 3 m_2 is low, and m_1 is still low. The event is not cancelled, and the final temperatures also turn out low. In case 4 m_2 is low and m_1 is high, but not high enough to lead to cancellation. However, the final temperatures turn out even higher, and a loss is incurred.

We now compare the average utility from applying the extended cost-loss model over the 2500 cases 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.

Results from this comparison are shown in Figure 1 for the base case parameter settings, as both the values of average utility from each of the 4 methods, and as differences in the average

utility between the 3 simpler methods and the extended cost-loss decision algorithm (which we label as *extended*). Statistical significance bars for 5% and 95% levels were generated using a bootstrap. We see that, for these particular values of the parameters L, C_1 and C_2, *extended* performs better than all the other methods (i.e., has less negative average utilities) and the differences versus the three simpler methods are each statistically significant.

4.2 General Results

We now consider a range of parameter values. Results for $C_1 = 0.1$, $C_1 = 0.5$ and $C_1 = 0.8$ are shown in the three panels in figure 2, and in each case C_2 is varied such that the ratio C_1/C_2 takes values from 1 to 4. L is always fixed at 1. For $C_1 = 0.1$, *extended* gives the best results by a clear margin. The relative performance of the other methods varies with the value C_1/C_2, and no method is consistently in second place. For $C_1 = 0.5$ *extended* and *basic-twice* give joint best results. For $C_1 = 0.8$ *extended, basic-twice* and *always-fc2* give joint best results.

The results in Figure 2 can be interpreted in more detail as follows. $C_1 = 0.1$ corresponds to cancelling on Friday being cheap relative to the loss that might be incurred on Saturday, which makes cancelling on Friday potentially attractive. When C_1/C_2 is very close to 1 (left hand end of the horizontal axis) cancellation on Thursday is not much cheaper than cancelling on Friday, and so the decision to wait to Friday is often a good one. As a result, always waiting to Friday to make the decision (*always-fc1*) is a good decision method, and the extended decision method does not add much relative to that. At the other end of the horizontal axis, when C_1/C_2 is large, cancellation on Thursday is now much cheaper than cancellation on Friday and cancelling on Thursday becomes much more attractive as an option. As a result, all methods do well except for never cancelling on Thursday (*always-fc1*). For intermediate values of C_1/C_2, cancellation on Thursday is cheaper than cancellation on Friday, but not so much as to dominate the decision. 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. The decision becomes a complicated trade-off. *Extended* takes all the factors into account, while the other 3 methods all neglect one or more aspects of the trade-off. As a result, *extended* gives significantly better results than any other method in this range.

$C_1 = 0.5$ corresponds to cancelling on Friday being somewhat expensive. When C_1/C_2 is close to one and cancellation on Thursday is also expensive then cancelling on Thursday is generally not a good option and so making the final decision on Thursday (*always-fc2*) does not do well. *Extended* does scarcely better than the other two simple methods, however, because the subtle logic that it adds to Thursday's decision is not particularly relevant in a situation where Thursday's decision is almost guaranteed to be 'wait' because of the high cost of cancellation on Thursday. When C_1/C_2 is large, and cancellation on Thursday is cheap, then always deciding on Thursday does well because cancellation on Friday is unlikely anyway because of the high costs on Friday. For intermediate values of C_1/C_2 *extended* does not perform better than *basic-twice* because the subtle logic that *extended* adds to the decision is not particularly relevant when deciding on Friday is unlikely because it is expensive.

$C_1 = 0.8$ corresponds to cancelling on Friday being almost as expensive as the loss that might be incurred on Saturday. That makes cancelling on Friday an unlikely decision. When C_1/C_2 is close to one and cancellation on Thursday is also expensive then cancelling on Thursday is also not a good option: except for the most extreme forecasts cancellation on either day is unlikely. The four methods work equally well because they all lead to the decision not to cancel

on either day in most cases. When C_1/C_2 takes intermediate or large values, and cancellation on Thursday is relatively cheap then, because cancellation on Friday is expensive, we are close to the limiting case where Friday's forecast is irrelevant, and we are making a final decision on Thursday. As a result, the two simple methods that involve deciding on Thursday (*always-fc2* and *simple-twice)* both perform well, as well as *extended.*

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 various 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 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.

We have generalized the basic cost-loss model by adding a previous day of forecast. The first decision is then whether to cancel 2 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.

We have analysed the cancel-or-wait decision that needs to be made 2 days in advance using expected utility. The logical 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 2 days in advance of the event) that the probability (evaluated 1 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 are normally distributed. In this case 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, 2 days in advance of the event, to run simulations of the possible distribution of probabilistic forecasts that will be available 1 day in advance. Based on this distribution of distributions, p' and $\hat{p}$ can easily be calculated.

We have tested our extended cost-loss decision algorithm on synthetic forecast data. The algorithm worked as expected and gave decisions that are at least as good as and for some parameter ranges 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.

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. In fact the theoretical framework extends straightforwardly to many stages of forecast. The challenge is that each additional day of forecasting adds a new level of probabilities of probabilities. While using two days of forecasts requires probabilities of probabilities, using three days of forecasts requires probabilities of probabilities of probabilities, and so on. This becomes hard to understand, and the simulations required, even for the normally distributed case, become more and more elaborate. However, some simplification may be possible.

An obvious next step is to test the extended cost-loss decision model using real forecast data. We have performed a preliminary investigation using 20 years of daily forecasts of ERA-interim JJA temperatures for Stockholm, Sweden. The forecasts were calibrated using a simple homogeneous linear regression scheme with constant coefficients and predictors from the current and previous forecast. The decision model was implemented using the normal distribution implementation algorithm presented above. Bad weather was defined as temperature exceeding the 95th quantile of the ERA-interim data. The results, however, were inconclusive, and the extended method did not clearly dominate the other decision-making methodologies as it had for the synthetic data. See for example Figure 3, which shows results for one particular range of parameters, in which 3 out of 4 of the decision methods give essentially the same average utility values. For the same range of parameters, the synthetic forecast data gave very different results. There are several possible reasons for the different results from the synthetic and real forecast data. For the real forecast data, the skill of the one and two day forecasts is very close, which may be making it difficult to detect the subtle benefits of the decision algorithm. Also, the real forecast errors may not be well modelled by a normal distribution, in which case using the normal distribution implementation algorithm is not appropriate and a different implementation algorithm would need to be used. Furthermore the forecast calibration scheme may not be working effectively, and may not be creating sufficiently well calibrated forecasts; for instance the forecast errors may not be homogeneous, and should perhaps be calibrated in a way that accounts for that, or the use of constant parameters in the calibration scheme might be inappropriate if the biases in the forecast vary in time. These reasons overlap to some extent: suboptimal calibration could be leading to non-normal forecast errors, for instance. Understanding these real forecast results and improving them would require a more detailed analysis. Doing so may lead to further insights into how to develop implementation algorithms for the extended cost-loss method for non-normal forecasts, and/or further insights into forecast calibration.

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 necessarily enough to make a logical decision as to whether to wait or not and the forecast may need 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 in a simulation algorithm that determines 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

G. Messori was partly supported by the Swedish Research Council Vetenskapsrådet grant no. 2016-03724.

References

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

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.

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.

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. & 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.

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.

Row		Utility	Number of cases	% of all cases
1	Cancel Thursday total	$-C_2$	589	23.56
2	…Cancel Thursday (then bad weather)	$-C_2$	94	3.76
3	…Cancel Thursday (then good weather)	$-C_2$	495	19.80
4	Wait Thursday	n/a	1911	76.44
5	Cancel Friday total	$-C_1$	37	1.48
6	…Cancel Friday (then bad weather)	$-C_1$	22	0.88
7	…Cancel Friday (then good weather)	$-C_1$	15	0.6
8	Go ahead Friday total	n/a	1874	74.96
9	…Go ahead Friday (then bad weather)	$-L$	16	0.64
10	…Go ahead Friday (then good weather)	0	1858	74.32

Table 4: numbers of different outcomes for the extended decision algorithm base case. 2500 cases were tested in all.

	m_2	m_1	p'	$\hat{p}$	a	Decision outcome
Case 1	2.55	2.70	0.35	0.05	3.32	Cancel Thursday
Case 2	1.36	3.58	0.19	0.03	3.02	Cancel Friday
Case 3	-3.87	-2.81	0.00	0.00	-2.85	Go ahead (good weather)
Case 4	0.29	2.80	0.02	0.01	3.94	Go ahead (bad weather)

Table 5: four examples from the extended decision algorithm base case

Figures
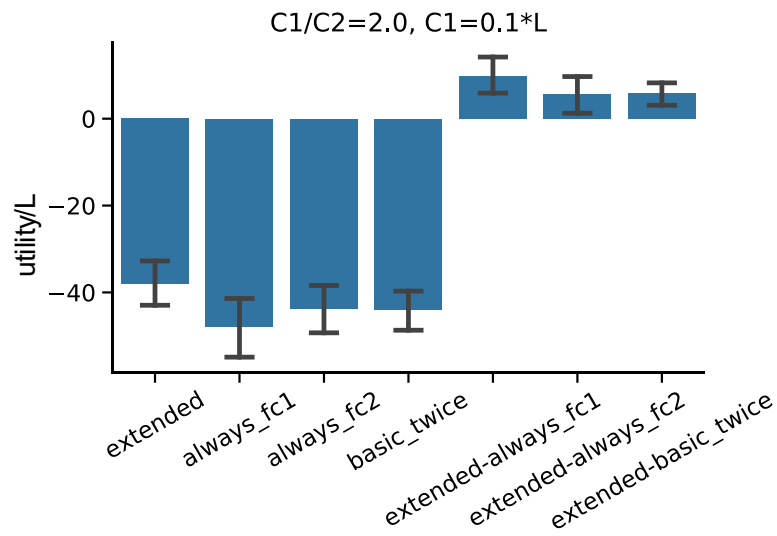
Figure 1: synthetic forecast data results for the average utility achieved across 2500 simulated cases, for four different decision-making methodologies applied to the base case described in the text. The methodologies are 1) the extended cost-loss decision making algorithm derived in this article (labelled extended), 2) the simple decision making methodology based on always ignoring Thursday's forecast and waiting to Friday (labelled always-fc1), 3) the simple decision making algorithm based on always ignoring Friday's forecast and deciding on Thursday (labelled always-fc2) and 4) the decision making algorithm that consists of applying the simple cost-loss framework on Thursday and then if necessary again on Friday (labelled basic-twice). The first four columns show the average utilities of each algorithm on its own. The black bars show the 5%-95% uncertainty intervals estimated with bootstrapping. We see that extended has the highest (least negative) utility. The next three columns show the differences in the utilities between the methods. We see that extended shows a statistically significant positive difference versus the other three methods.
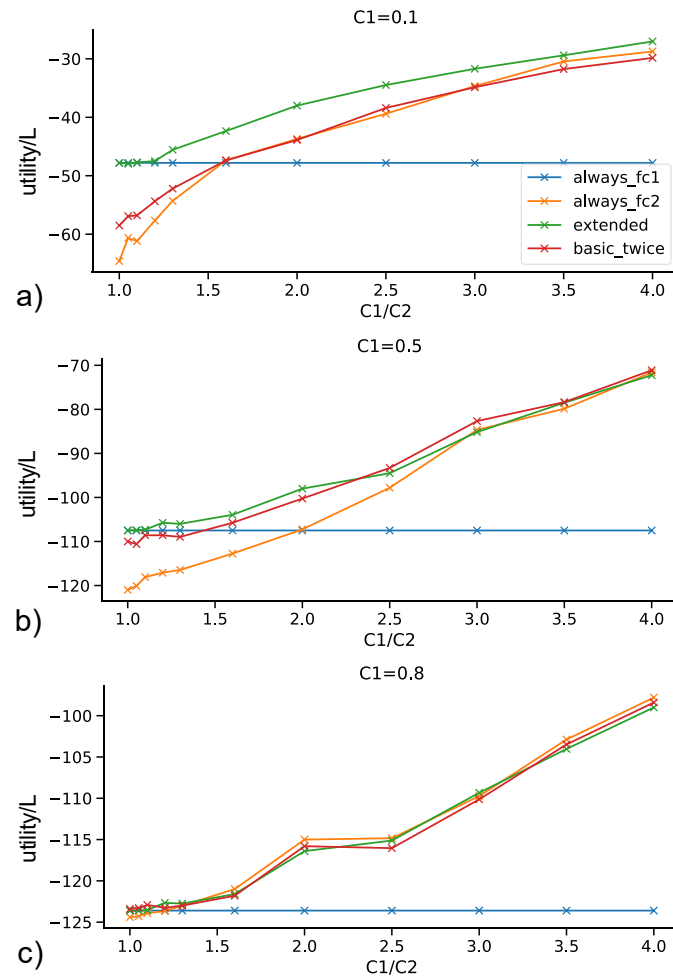
Figure 2: Average utilities for a range of parameter values in the decision problem, calculated from synthetic forecasts, and compared across the four decision making methodologies. Panel (a) is for $L = 1$, $C_1 = 0.1$, and a range of C_2 values such that the ratio C_1/C_2 takes values from 1 to 4. Panel (b) is the same, but for $C_1 = 0.5$ and panel (c) is the same, but for $C_1 = 0.8$. 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.
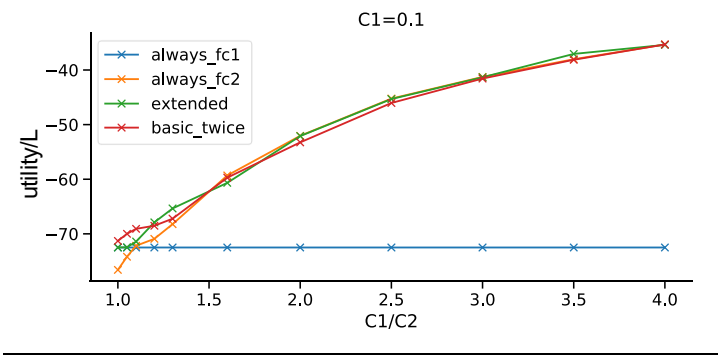
Figure 3: average utilities from the four decision models, for a real weather forecast data, for a range of parameter values.