# Potential Economic Value of Ensemble-Based Surface Weather Forecasts

DANIEL S. WILKS AND THOMAS M. HAMILL

*Department of Soil, Crop and Atmospheric Sciences, Cornell University, Ithaca, New York*

(Manuscript received 2 December 1994, in final form 29 March 1995)

ABSTRACT

The possible economic value of the quantification of uncertainty in future ensemble-based surface weather forecasts is investigated using a formal, idealized decision model. Current, or baseline, weather forecasts are represented by probabilistic forecasts of moderate accuracy, as measured by the ranked probability score. Hypothetical ensemble-based forecasts are constructed by supplementing the baseline set of probabilistic forecasts with lower- and higher-skill forecasts. These are chosen in such a way that mixtures of the forecasts including the lower- and higher-skill subsets with equal frequency exhibit the same accuracy overall as the moderately accurate (conventional, baseline) forecasts. For both simple one-time decisions (static situation) and related sequences of decisions (dynamic situation), these hypothetical ensemble-based forecasts are found to lead to greater economic value in the idealized decision problem when protective actions are relatively inexpensive, corresponding to real-world problems. However, for some decision problems considered, the ensemble-based forecasts are slightly less valuable than the baseline forecasts. This result derives at least in part from the (probably unrealistic) assumption that the ensemble-based forecasts are no more skillful in aggregate than their conventional counterparts, but implies that positive economic value for ensemble forecasts with respect to this baseline will not be automatic. Rather, for ensemble-based forecasts to be at least as valuable for all decision problems, they will need to exhibit sufficiently higher skill in aggregate than the conventional forecasts that could have been produced in their place.

## 1. Introduction

It is sometimes possible to identify, a priori, forecast situations in which future atmospheric behavior is less uncertain, or more uncertain, than usual. One approach to this is through ensemble forecasting, which is implemented by allocating available computer resources to multiple runs (realizations) of a relatively low-resolution dynamical forecast model rather than to the more conventional single high-resolution forecast realization (Brooks and Doswell 1993; Mureau et al. 1993; Tracton and Kalnay 1993). Each of these runs (ensemble members) is started from a slightly different but plausible initial condition. Due to the chaotic nature of atmospheric dynamics, the small perturbations grow and eventually spread through the spectrum of atmospheric motions. The dispersion among the ensemble of forecasts on particular occasions has been found to be related to atmospheric predictability (F. Molteni et al. 1995, unpublished manuscript).

One way to think about the dispersion of forecasts among the ensemble members is as an anticipation of the intrinsic predictability of a given forecast (Kalnay and Dalcher 1987). Forecasts for which the ensemble members are in relatively close agreement are expected to be less uncertain, while cases where ensemble members are very different from each other may be interpretable as indicating higher uncertainty and portending relatively low forecast accuracy. Though work to date has concentrated on the medium range, the methodology may be equally appropriate for shorter-range forecasts (Brooks et al. 1992; Mullen and Baumhefner 1994). In addition to potentially providing information on forecast uncertainty, results from medium-range forecasts indicate that the mean of an ensemble of forecasts is typically more skillful than the majority of individual forecasts, and more skillful than the single higher-resolution forecast that could have been run in their place (Toth and Kalnay 1993).

It is usually assumed implicitly that the enhanced information available from ensemble forecasts will yield greater economic value, but to our knowledge this proposition has not yet been investigated formally. A significant impediment to this kind of investigation is that most real-world weather-sensitive decision problems relate to surface weather variables (e.g., maximum temperature or precipitation amount). While ensemble forecasting of large-scale geopotential height fields is presently practiced at major operational centers (Mureau et al. 1993; Tracton and Kalnay 1993), to our knowledge no ensemble-based forecast guidance for surface weather elements is either available or under

*Corresponding author address:* Daniel S. Wilks, Dept. Soil, Crop and Atmospheric Sciences, 1113 Bradfield Hall, Cornell University, Ithaca, NY 14853.
E-mail: dsw5@cornell.edu

development. Therefore, we resort here to the analysis of hypothetical ensemble-based forecasts of surface weather, which may be broadly representative of the kind of forecast guidance that may eventually become available.

This paper uses decision analysis (e.g., Clemen 1991; Winkler and Murphy 1985; Katz and Murphy 1996) to explore the magnitude and nature of the economic consequences that might derive from quantifying forecast uncertainty for surface weather elements using ensemble methods. Individual decision problems are decomposed into sets of allowable actions (decisions) and unknown future weather events. The economic consequences of each of the possible pairs of the actions and events are computed, and the action is chosen that minimizes expected expense (or, equivalently, maximizes expected income), given forecasts for the future uncertain events. Here "expected" is meant in the statistical sense of a probability-weighted average; and the probabilities are supplied by the weather forecasts, which are assumed to be well calibrated.

Baseline, or conventional, forecasts are regarded as exhibiting a constant, moderate level of accuracy. Ensemble forecasts are assumed to exhibit either low, moderate, or high accuracy, with this being specified a priori, as part of each individual forecast. Hypothetical surface weather forecasts are constructed such that the ensemble-based forecasts (jointly, over all three classes) and the conventional forecasts are equally accurate on average, when the proportion of low- and high-skill forecasts are equal. Therefore, the ensemble forecasts include subsets that are better than the baseline forecasts but also subsets that are of lower quality.

## 2. Forecasts

Because ensemble-based forecasts of surface weather elements are not yet available, this study analyzes hypothetical forecasts of an unspecified surface weather variable. It is assumed that the weather event being forecast is defined over five mutually exclusive and collectively exhaustive categories, $\theta_i$, $i = 1, \cdots, 5$. To fix ideas, it may be helpful to think of these as five bins of daily precipitation amounts, or as five classes of daily maximum temperature outcomes.

Hypothetical sets of climatological probabilities $\pi_i$ for the five events are generated using beta distributions, integrated over five equal partitions of the interval [0, 1],

$$\pi_i = \int_{x^-}^{x^+} B(x; \alpha, \beta)dx = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_{x^-}^{x^+} x^{\alpha-1}(1 - x)^{\beta-1}dx; \quad i = 1, \cdots, 5. \quad (1)$$

Here $B(x; \alpha, \beta)$ indicates the beta probability density function for the random variable $x$, having parameters $\alpha$ and $\beta$, $x^- = (i - 1)/5$, $x^+ = i/5$, and $\Gamma$ is the gamma function. Different sets of climatological probabilities are generated through different choices for the parameters $\alpha$ and $\beta$. For example, $\alpha = 0.75$ and $\beta = 1.25$ results in the climatology vector $\boldsymbol{\pi} = (0.351, 0.224, 0.180, 0.146, 0.099)$, indicating relatively few events in $\theta_5$.

The hypothetical ensemble-based forecasts consist of three sets of probabilities $f_{e,i,j}$, where $e = 1$ indicates the "good," or more accurate forecasts, $e = 2$ indicates forecasts of intermediate accuracy, and $e = 3$ indicates the "poor," or less accurate forecasts. The subscript $i$ corresponds to the event $\theta_i$ being forecast, and it is assumed that each of the three sets of forecasts consists of five conditional probability distributions centered on the corresponding event. For each value of $e$, $f_{e,i,j}$ denotes the conditional probability of $\theta_i$ given forecast $j$; $j = 1, \cdots, 5$. That is, it is assumed that each forecast will be one of only five possible probability vectors. Forecasters in a real-world setting would of course have a much richer suite of possible forecasts from which to choose. The forecast probabilities for the baseline forecasts (i.e., no ensemble forecast information) are assumed to be the same as those for the intermediate ($e = 2$) ensemble forecasts $f_{2,i,j}$.

Allocation of probability for the good ($e = 1$), sharper forecast distributions is made by integrating a Gaussian distribution over the event space

$$f_{1,i,j} = \int_{x^-}^{x^+} \phi(x; \mu_j, \sigma)dx; \quad i, j = 1, \cdots, 5; \quad (2)$$

where $\phi$ indicates the Gaussian probability density function (chosen here for convenience), the mean of which is $\mu_j = (j - 0.5)/5$, and whose standard deviation $\sigma$ can be varied to yield ensemble forecasts of differing quality. The limits of integration are as in (1), except that $x^- = -\infty$ for $i = 1$, and $x^+ = +\infty$ for $i = 5$. It is assumed that the forecasts in (2) are reliable, or calibrated, in the sense that the forecast probabilities correspond exactly to the long-run relative frequencies of the events to which they pertain (e.g., Wilks 1995). For this condition to hold, the frequency of use of each of the five probability vectors $\mathbf{f}_{1,j}$ must be consistent with the vector of climatological probabilities in (1). That is,

$$\mathbf{p} = \mathbf{F}^{-1}\boldsymbol{\pi}, \quad (3)$$

where the matrix $\mathbf{F}$ contains the forecast probability vectors as columns, the column vector $\boldsymbol{\pi}$ is derived from (1); and the column vector $\mathbf{p}$ is the "predictive distribution," containing the relative frequencies of use, $p_j$, of each of the five forecasts, $j = 1, \cdots, 5$. Equation (3) allows determination of whether a given set of forecast and climatological probabilities are mutually consistent, in the sense that $\mathbf{p}$ satisfies the constraints required of a probability distribution. Note that the assumption of forecast reliability is not restrictive,

TABLE 1. Example set of forecast probabilities and corresponding climatological probabilities, constructed using parameters $\alpha = 0.75$, $\beta = 1.25$, and $\sigma = 0.155$. The frequency-of-use vector is $\mathbf{p} = (0.408, 0.169, 0.185, 0.171, 0.067)$, and the resulting RPS is 0.470.

| | $j$ | $\theta_1$ | $\theta_2$ | $\theta_3$ | $\theta_4$ | $\theta_5$ |
|---|---|---|---|---|---|---|
| $f_{1,i,j}$ | 1 | 0.740 | 0.233 | 0.026 | 0.001 | 0.000 |
| | 2 | 0.259 | 0.481 | 0.233 | 0.026 | 0.001 |
| | 3 | 0.027 | 0.233 | 0.481 | 0.233 | 0.026 |
| | 4 | 0.001 | 0.026 | 0.233 | 0.481 | 0.259 |
| | 5 | 0.000 | 0.001 | 0.026 | 0.233 | 0.740 |
| $f_{2,i,j}$ | 1 | 0.667 | 0.231 | 0.055 | 0.028 | 0.019 |
| | 2 | 0.277 | 0.433 | 0.223 | 0.048 | 0.019 |
| | 3 | 0.088 | 0.231 | 0.424 | 0.217 | 0.040 |
| | 4 | 0.067 | 0.063 | 0.223 | 0.418 | 0.229 |
| | 5 | 0.066 | 0.043 | 0.055 | 0.217 | 0.619 |
| $f_{3,i,j}$ | 1 | 0.568 | 0.229 | 0.094 | 0.065 | 0.044 |
| | 2 | 0.300 | 0.367 | 0.210 | 0.079 | 0.044 |
| | 3 | 0.170 | 0.229 | 0.348 | 0.194 | 0.059 |
| | 4 | 0.156 | 0.114 | 0.210 | 0.332 | 0.188 |
| | 5 | 0.155 | 0.100 | 0.094 | 0.194 | 0.457 |
| $\pi_i$ | | 0.351 | 0.224 | 0.180 | 0.146 | 0.099 |

since probability forecasts known to exhibit imperfect reliability can be recalibrated to yield revised probabilities exhibiting this property.

The poor ($e = 3$) forecasts are constructed by relaxing the good forecasts toward the climatological probabilities according to

$$f_{3,i,j} = \left(\frac{\sigma}{0.35}\right)\pi_i + \left(1 - \frac{\sigma}{0.35}\right)f_{1,i,j}; \quad i, j = 1, 2, 3, 4, 5. \quad (4)$$

Thus for high forecast quality overall (small $\sigma$) the good and poor forecasts are quite similar, while for decreasing forecast quality overall (increasing $\sigma$) the poor forecasts progressively resemble the climatological probabilities. Since (4) is a linear combination of (1) and (2), the probabilities $f_{3,i,j}$ and $f_{1,i,j}$ will yield the same frequency-of-use vector $\mathbf{p}$ in (3).

The hypothetical baseline ($e = 2$) forecasts are produced as a weighted average of the good and poor forecasts [Eqs. (2) and (4)]. For differences in forecast value to be attributable only to the difference in information content between ensemble and baseline forecasts, the intermediate forecasts are constructed to have expected [with respect to the probabilities in (3)] accuracy, as measured by the ranked probability score (RPS) (Epstein 1969; Murphy 1969), that is equal overall to the ensemble forecasts. It will be assumed that the forecasts $f_{1,i,j}$ and $f_{3,i,j}$ are issued by the ensemble forecasting system with equal frequency, so the position between them defining $f_{2,i,j}$ is chosen to yield the same RPS as an equal mixture of these two sets of forecasts. Thus, these forecasts are constructed by finding the linear combination of $f_{1,i,j}$ and $f_{3,i,j}$ exhibiting an RPS that is equal to the average of the RPS values for the good and poor forecasts. The resulting forecasts $f_{2,i,j}$ also yield the same probabilities $\mathbf{p}$ in (3).

Table 1 shows an example set of forecasts, generated using $\alpha = 0.75$, $\beta = 1.25$, and $\sigma = 0.155$. The $j = 1$ forecasts, whose probabilities are most concentrated at $\theta_1$, are used most frequently ($p_1 = 0.408$); and the $j = 5$ forecasts, whose probabilities are most concentrated at $\theta_5$, are used least frequently ($p_5 = 0.067$). Again, real-world forecasters would have the flexibility to allocate probability in more ways than are represented by the 15 forecast vectors listed in Table 1. When the sets of five probabilities are forecast with the frequencies specified by the $p_j$'s, RPS for the $e = 2$ (baseline) forecasts is 0.470. Since the RPS possesses a negative orientation (smaller RPS is better) the RPS for the good ($e = 1$) forecasts is lower, and RPS for the poor ($e = 3$) forecasts is higher. However, mixtures of forecasts drawn equally from the good and poor forecasts (and also in proportion to the $p_j$'s) also exhibit average RPS = 0.470.

## 3. Decision-making models

Decision analysis is a well known and logically consistent structure within which to analyze decision problems relating to uncertain future events, to find economically optimal decisions, and to compute the economic value (as distinct from the accuracy) of forecast information (e.g., Katz and Murphy 1996; Winkler and Murphy 1985; Winkler 1972). At its root, decision analysis is based on the simple principle that the optimal decision maker will act to maximize expected (i.e., probability-weighted average) monetary return, or (equivalently) to minimize expected loss. For example, if offered the choice between \$5 for sure versus \$15 on the flip of a coin, an optimal decision maker would choose the coin flip because the expected return for that decision is \$7.50. In this simple example, the probabilities for the relevant events (heads and tails) are clear. In more interesting problems, the probabilities for the relevant events may not be so obvious. Different and possibly competing sources of information (forecasts) may be available. The relative economic worth of these information sources may also be computed, by comparing the expected monetary returns that would be realized by optimal decision makers utilizing each of them. In the present context, these competing information sources are the hypothetical ensemble-based and conventional forecasts described in the previous section.

### a. Static decision problem

The five-action, five-event cost/loss ratio setting, a simple idealized decision problem, is used here to investigate the nature of the potential economic value of ensemble-based surface weather forecasts. For the static decision problem, this is a special case of the problems described by Murphy (1985). The five weather events $\theta_i$ described in section 2 are regarded as causing increasing levels of damage to a hypothetical economic enterprise. That is, they represent increasingly adverse weather, with $\theta_1$ being not adverse and causing no loss, and $\theta_5$ being most adverse and causing a complete loss of magnitude $L$ if no protective measures are employed. Proportional intermediate losses result from events of intermediate adversity if no protective action is taken. For the static problem, it is assumed that each decision is a one-time situation or, if a series of such decisions are made, that each decision in the series is not influenced, nor its outcome affected, by the outcomes of previous or subsequent decisions.

Five levels of protection exist, which range from no protective action at all ($a_1$) to full protection ($a_5$). The cost of protection varies linearly from zero for $a_1$ to the cost of full protection $C$ for $a_5$. For convenience these costs are normalized by the magnitude of the full loss, yielding a relative cost for the $k$th protection level

$$C_k = \left(\frac{C}{L}\right)\frac{k-1}{4}, \quad k = 1, \cdots, 5. \tag{5}$$

Each of the protective actions prevent losses at that and lesser levels of adverse weather, but allow fractional losses to the extent that underprotection has been employed. Specifically, the loss sustained given the event $\theta_i$ and the protective action $a_k$ is

$$L_{i,k} = \begin{cases} (i-k)/4, & k < i \\ 0, & k \geq i \end{cases}; \quad i, k = 1, \cdots, 5. \tag{6}$$

This function is displayed in Table 2.

The problem for the decision maker is to select the level of protection that is optimal, in the sense of yielding the minimum expected expense (cost of protection plus loss), given available probability information regarding the upcoming weather event. If only the (constant) climatological probabilities $\pi_i$ are available, the same action will be optimal on each occasion, and the expected expense is

$$\bar{E}_{\text{clim}} = \min_k \left(C_k + \sum_{i=1}^{5} \pi_i L_{i,k}\right). \tag{7}$$

For each possible action (indexed by $k$), the expected loss is computed (summation within the parentheses) and added to the known cost of protection for that action. The expected expense is then the least of the five possibilities.

Given the conventional baseline forecasts $f_{2,i,j}$, the optimal action may be different for the different probability forecasts (values of $j$). Thus, the overall expected expense must be averaged over all of the possible forecasts, using as weights the frequencies of use $p_j$ of each of the five distinct sets of forecast probabilities. The expected expense associated with the baseline forecasts is then

TABLE 2. The loss function $L_{i,k}$ [Eq. (6)], normalized by the magnitude of the full loss $L$, for the 5 × 5 cost/loss ratio problem; as a function of the events $\theta_i$ and the actions $a_k$.

| | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|---|
| $\theta_1$ | 0 | 0 | 0 | 0 | 0 |
| $\theta_2$ | 1/4 | 0 | 0 | 0 | 0 |
| $\theta_3$ | 1/2 | 1/4 | 0 | 0 | 0 |
| $\theta_4$ | 3/4 | 1/2 | 1/4 | 0 | 0 |
| $\theta_5$ | 1 | 3/4 | 1/2 | 1/4 | 0 |

$$\bar{E}_{\text{base}} = \sum_{j=1}^{5} p_j\left[\min_k\left(C_k + \sum_{i=1}^{5} f_{2,i,j} L_{i,k}\right)\right]. \tag{8}$$

Similarly, if the three sets of ensemble-based forecast probabilities are available, the expected expense associated with their use is the probability-weighted average of the use of all (3 × 5 =) 15 possible forecasts, given by

$$\bar{E}_{\text{ens}} = \sum_{e=1}^{3}\sum_{j=1}^{5} p_{e,j}\left[\min_k\left(C_k + \sum_{i=1}^{5} f_{e,i,j} L_{i,k}\right)\right]. \tag{9}$$

Here it is assumed that the good and poor forecasts are used with equal frequency, and that the frequency of use of each of the probabilities $p_{e,j}$ is proportional to the corresponding probability in the distribution specified by (3). The degree to which different levels of forecast accuracy can be discriminated is then represented by the frequency of use of the three subsets ($e = 1, 2, 3$) of forecast probabilities, indexed here by the parameter $M$ ($0 \leq M \leq 1$) according to

$$\left.\begin{aligned} p_{1,j} &= \frac{1-M}{2} p_j \\ p_{2,j} &= M p_j \\ p_{3,j} &= \frac{1-M}{2} p_j \end{aligned}\right\}, \quad j = 1, \cdots, 5. \tag{10}$$

When $M = 1$, the intermediate forecasts are always used, implying the conventional, baseline forecasts. At the opposite extreme, $M = 0$ implies that the ensemble-based system issues forecast probabilities only from the sets $f_{1,i,j}$ and $f_{3,i,j}$. Because of the method of forecast construction, the value of $M$ does not affect the overall (average) RPS.

The economic value associated with availability of the ensemble forecasts, as compared to availability only of the baseline forecasts is usually (e.g., Katz et al. 1982; Murphy 1985) given by the measure

$$V = \bar{E}_{\text{base}} - \bar{E}_{\text{ens}}. \tag{11}$$

To the extent that use of the ensemble-based forecasts results in a smaller expected expense, their value as measured by (11) will be positive. Since the units of this hypothetical decision problem are arbitrary, we also consider the nondimensional percentage reduction in expected expense, in relation to the expense associated with the climatological information,

$$\%V = \frac{\bar{E}_{\text{base}} - \bar{E}_{\text{ens}}}{\bar{E}_{\text{clim}}} \times 100\%. \quad (12)$$

*b. Dynamic decision problem*

The basic decision problem outlined in the previous section can be generalized to a sequence of related decisions. Fractional losses sustained in this sequence are cumulative, and the full loss can be sustained at most once. Thus available options at a given time, as well as their consequences, depend on the previous time history of the decision sequence. In view of this time dependence, such problems are often called "dynamic." The specific case considered here is the five-action and five-event dynamic cost/loss ratio problem, which has a structure analogous to the static problem described in section 3a, and has been described previously by Wilks (1991).

As a sequence of decisions proceeds through time, the degree of loss sustained is recorded by the state variable $\lambda$. For the simple $5 \times 5$ cost/loss ratio problem considered here, the state variable records the degree of loss sustained to date, the possible values of which are 0, 1/4, 1/2, 3/4, and 1 [cf. (6) and Table 2]. At each stage of the decision process, it is still the case that the action $a_k$ is chosen which minimizes the expected expense, but for the current *and all future* decision periods. In comparison to (8) or (9), this choice is complicated by two additional considerations. First, the optimal action may depend on the value of $\lambda$. That is, the best action can be a function of the degree of loss already sustained as a consequence of previous actions. Effectively, $\lambda = 0$ for all decisions in the static problem described in section 3a, since there are no previous decision periods in which losses could have accumulated. Second, the degree to which the decision maker is willing to expend resources on the protection costs $C_k$ will depend on how many more decisions remain in the sequence. For example, it will certainly not be optimal to protect multiple times if the total of the multiple protection costs is greater than the full loss [cf. Eq. (5)]. Effectively, the static problem described in section 3a corresponds to the final decision in a sequence, in that no further costs or losses can be incurred.

As a consequence of these two complications, the analog of (9) for the dynamic decision problem is a function of both the state variable $\lambda$, and the position in time within the decision sequence, $t$. Somewhat counterintuitively, it is convenient to solve dynamic decision problems of this kind in reverse time, using a process known as "backward recursion" (e.g., Kennedy 1986). Usually, the decision stages are numbered backward as well, so that $t = 1$ indicates the final decision, $t = 2$ the second-to-last decision, etc. Analogous to (9), this recursion is defined by

$$\bar{E}_{\text{ens}}[\lambda, t] = \sum_{e=1}^{3} \sum_{j=1}^{5} p_{e,j}\{ \min_k (C_k + \sum_{i=1}^{5} f_{e,i,j} \bar{E}_{\text{ens}} \times [T(\theta_i, a_k, \lambda), t-1])\}, \quad (13)$$

where $T(\theta_i, a_k, \lambda)$ is the "transfer function." The transfer function is essentially a pointer that specifies the value of the state variable in the next time period $(t-1)$, given that action $a_k$ was taken and event $\theta_i$ occurred. That is,

$$T(\theta_i, a_k, \lambda) = \min(1, \lambda + L_{i,k}). \quad (14)$$

At each stage in the decision process the action $a_k$ is selected as optimal that satisfies the minimization in the curly brackets of (13), for the current and all subsequent decision periods. In addition to depending on the forecast probability $f_{e,i,j}$, as in (9), the optimal action can also depend on the value of the state variable $\lambda$, and the stage of the decision process $t$. The recursion in (13) is "initialized" for $t = 0$ with

$$\bar{E}_{\text{ens}}(\lambda, 0) = \lambda. \quad (15)$$

That is, at the end of the decision sequence $(t = 0)$ the realized loss is given by the final value of $\lambda$.

## 4. Results

*a. Static decision problem*

Consider first the value measure

$$\%V_{\text{clim}} = \frac{\bar{E}_{\text{clim}} - \bar{E}_{\text{ens}}}{\bar{E}_{\text{clim}}} \times 100\%. \quad (16)$$

This equation specifies the relative value of the ensemble-based forecasts with respect to the climatological probabilities, rather than comparing two forecast systems as do (11) and (12). Figure 1 shows $\%V_{\text{clim}}$ as a function of overall forecast quality (RPS) and the ability to forecast forecast skill ($M$), for the cost/loss ratio, $C/L = 0.10$, and the same climatological probabilities obtained in Table 1. The most striking feature of Fig. 1 is that forecast value increases quite strongly with decreasing RPS, as would generally be expected. However, over some portions of the RPS scale, a nonnegligible difference in value is apparent between the top of the figure (baseline forecasts, $M = 1$) and ensemble-based forecasts in the lower portion of the figure, indicating that issuance of the ensemble-based forecasts does affect economic value for forecast users. In Fig. 1, and in other comparable figures (not shown), at each RPS the reduction in expected expenses for $M = 0.5$ is the average of the expected expense reductions for $M = 0$ (no forecasts in the middle category) and $M = 1$ [for which the reduction in expected expense in Eqs.

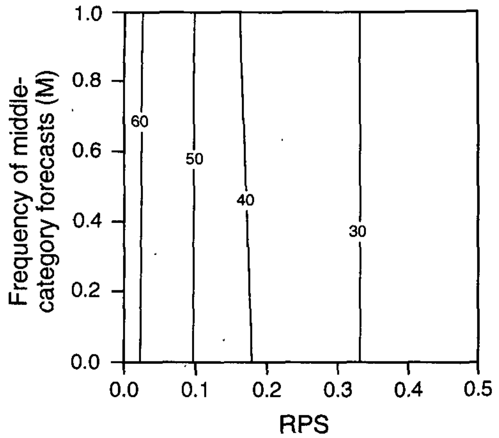



FIG. 1. Percentage reduction in expected expenses for $C/L = 0.10$ in relation to climatological information [Eq. (16)], as a function of overall forecast accuracy as reflected by the RPS, and the frequency of use of the middle-category forecasts $M$ [Eq. (10)]. The climatological probabilities are those corresponding to the parameters $\alpha = 0.75$ and $\beta = 1.25$, as in Table 1.

(11) and (12) is zero]. Accordingly, subsequent results will be presented in the forms of (11) and (12), with $M = 0$ ensemble forecasts only, effectively differencing horizontal cross sections at the top and bottom of Fig. 1.

Figures 2–4 show value of the ensemble forecasts with $M = 0$, according to (a) the expected expense difference (11), and (b) the percentage reduction in expenses relative to the climatological probabilities according to (12). Results in Fig. 2 are for a relatively "optimistic" climate (the harmless event $\theta_1$ is most likely), used also for construction of Table 1. Figure 3 shows results for nearly equal climatological probabilities in each category, and Fig. 4 shows the corresponding results for a "pessimistic" climatology, in which the probabilities for the most damaging events are high. The three figures are broadly similar, indicating little dependence of forecast value on particular climatologies.

As might be expected, all three cases in Figs. 2–4 indicate greatest value for the ensemble forecasts for the smaller $C/L$ ratios: for the large $C/L$ ratios, protection is nearly as costly as allowing the loss. The differences between the (a) and (b) panels are also most noticeable for extremely small $C/L$, for which the value measure in (12) becomes quite large. Inexpensive protection will be chosen here given climatological information, yielding low expected expense in the denominator of (12). For the lower cost/loss ratios, forecast value [Eq. (11)] is greatest for the forecasts with higher overall RPS. Note that for very low RPS (extremely high forecast quality), the good, baseline, and poor forecasts all approach perfect forecasts [$\sigma$ approaches zero in Eqs. (2) and (4)], and the differences between the ensemble and baseline results accordingly approach zero.

The regions in the lower-right portions of these figures are probably the most applicable to real-world decision settings. Moderately low but nonzero $C/L$ ratios are likely to correspond most closely to real-world situations, since high $C/L$ values would suggest marginal economic viability unless the probability of damaging weather is very low indeed. The higher RPS values are indicative of forecasts fairly far from perfection. For example, the forecasts presented in Table 1 correspond to RPS = 0.47 in Fig. 2.

Also apparent in Figs. 2–4 are regions (dashed contours) where the hypothetical ensemble forecasts exhibit *negative* value relative to the baseline forecasts. These are located in portions of the figures that may be of limited relevance to real-world problems, and/or ex-

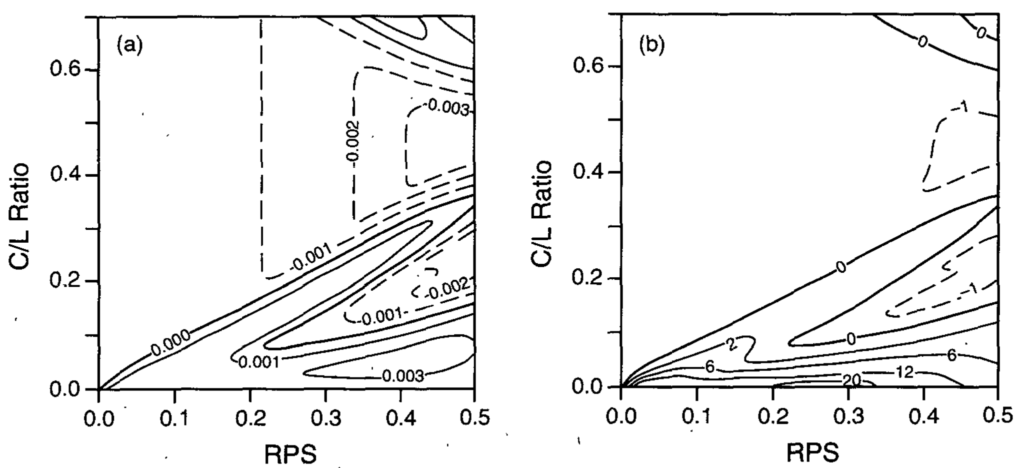



FIG. 2. Value of ensemble forecasts with $M = 0$ using (a) Eq. (11) and (b) Eq. (12), as a function of overall forecast accuracy (RPS) and the $C/L$ ratio. The values of $\alpha$, $\beta$, and the climatological probabilities are the same as for Table 1 and Fig. 1. Dashed contours indicate negative values.

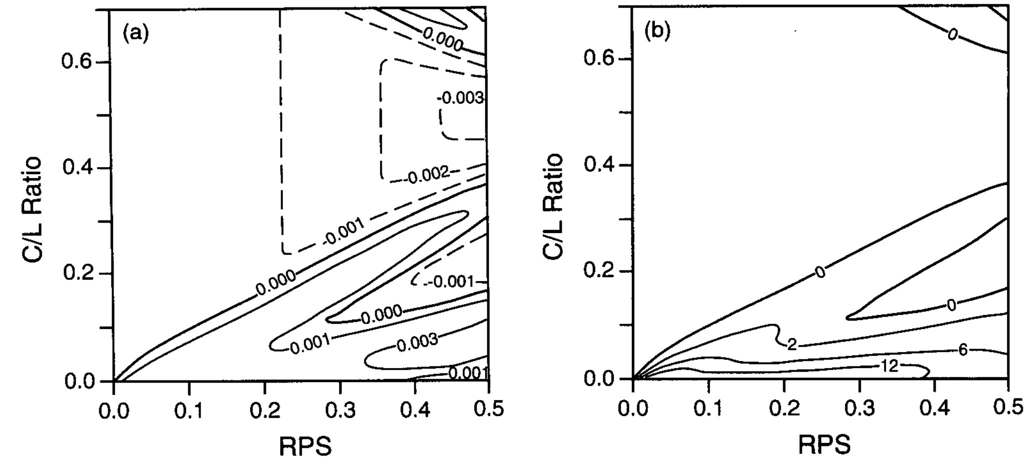



FIG. 3. As in Fig. 2 with $\alpha$ = 1.25, $\beta$ = 1.25, implying $\boldsymbol{\pi}$ = (0.168, 0.218, 0.228, 0.218, 0.168).

hibit negative values that are of small magnitude, but their existence is nevertheless surprising. Recall that the ensemble forecasts are constructed to have overall RPS that is equal to their counterpart baseline forecasts. The regions of negative value evidently correspond to decision situations where the ''good'' forecasts are not sufficiently good to compensate for the low quality of the ''poor'' forecasts in the aggregate.

Figure 5 provides a closer look at the situation for $C/L$ = 0.10 for the climate represented in Fig. 2. The percentage value in relation to climatological information [Eq. (16)] for the baseline forecasts (or, equivalently, ensemble forecasts with $M$ = 1) is shown by the solid line, and the corresponding result for the ensemble forecasts (with $M$ = 0) is shown by the dashed line. Both are monotonically decreasing in RPS. The latter is below the former in the regions (RPS $\leqslant$ 0.11, and 0.24 $\leqslant$ RPS $\leqslant$ 0.33) where the ensemble forecasts, as constructed here, show less value than the baseline forecasts. These regions are also reflected in Fig. 1, in which the slope of the 50% contour is opposite in sign to that of the 40% contour, and the 30% contour is essentially vertical. The multiple crossings of these lines are instances of ''accuracy-value reversals,'' as have been described by Ehrendorfer and Murphy (1988). That is, as a scalar measure the RPS provides an incomplete characterization of forecast performance, and it can occur that forecasts exhibiting lower (better) RPS are of less value for some decision problems.

### *b. Dynamic decision problem*

As was done for the static problem, the results presented in this section pertain to the comparisons between $M$ = 0 and $M$ = 1, so the ensemble-based forecasts consist entirely of probabilities that are equally partitioned between the high-skill ($e$ = 1) and low-skill

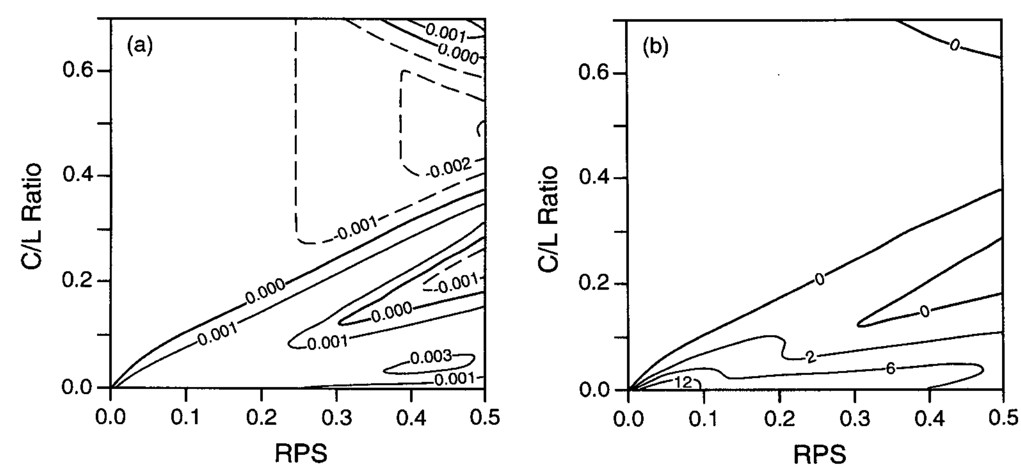



FIG. 4. As in Fig. 2 with $\alpha$ = 1.75, $\beta$ = 1.25, implying $\boldsymbol{\pi}$ = (0.079, 0.176, 0.239, 0.270, 0.236).

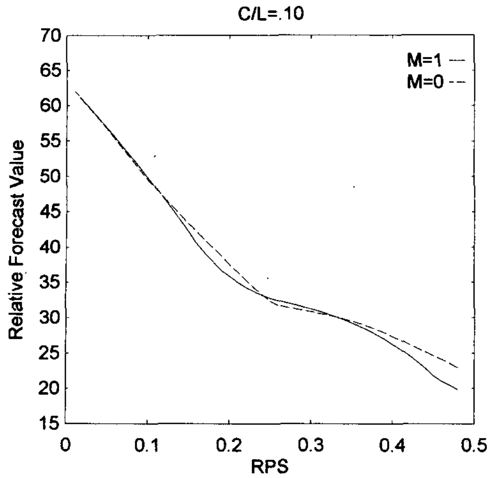



FIG. 5. Relative forecast value in relation to climatological probabilities [Eq. (16)] for ensemble forecasts with $M = 0$ (dashed line) and baseline forecasts (solid line), in the situation corresponding to $C/L = 0.10$ in Fig. 2.

($e = 3$) forecasts. Figure 6 shows relative expense differences [Eq. (12)] at $t = 15$ days from the end of the decision sequence, for (a) the ''optimistic'' climatology (corresponding to Fig. 2), (b) the intermediate climatology (corresponding to Fig. 3), and (c) the ''pessimistic'' climatology (corresponding to Fig. 4). Here it has been assumed that $t = 15$ is the beginning of the decision sequence. Corresponding figures for other decision periods are similar, because of the normalization by the expected expense associated with climatological information, although the absolute values (not shown) diminish as $t$ increases. These estimates of the economic value of the ensemble-based forecasts with respect to the baseline forecasts are appreciable for the small values of $C/L$ that are most realistic for a sequential decision problem of this kind. Regions in the RPS–$C/L$ space for which the modeled ensemble forecasts are computed to have negative value also decrease as $t$ increases, and for $t = 15$ these are very near zero. The overall relative forecast value declines somewhat as the climate becomes progressively less favorable.

The basis for the positive economic value, for the case of $C/L = 0.05$ and the forecasts shown in Table 1 in the ''optimistic'' climatology, is shown in Fig. 7. Here the optimal actions $a_k$ are indicated using different levels of shading, as a function of time $t$ in the decision sequence and the forecast $j$, for the 12 possible combinations of loss $\lambda$ sustained to date and the three levels of forecast quality $e$. Recall that the decision periods are numbered in reverse, so that forward time is from left to right. Larger values of the forecast index $j$ indicate increasing probabilities for the most damaging event $\theta_5$ and decreasing probabilities for the harmless event $\theta_1$, and increasing levels of shading indicate progressively stronger levels of protection.

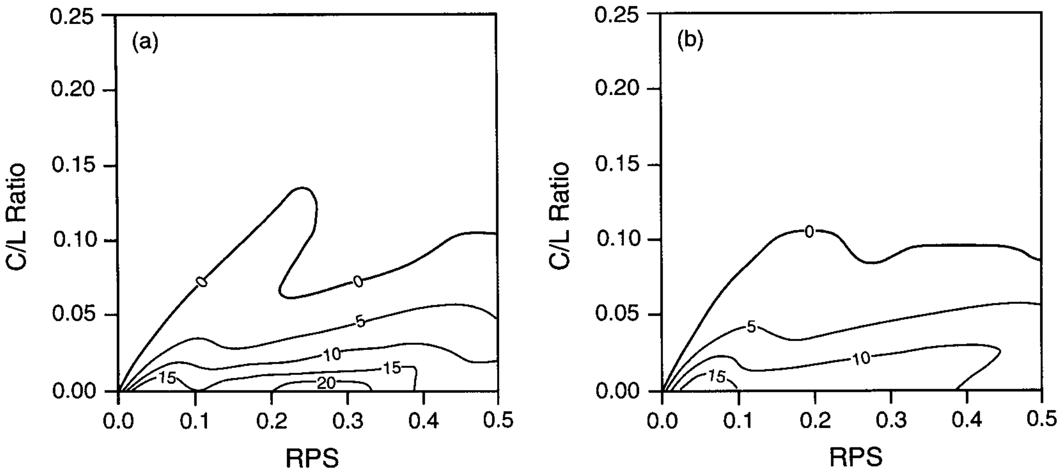
The results in Fig. 2 for the static decision problem, with $C/L = 0.05$ and RPS $= 0.47$, correspond to the actions depicted in Fig. 7 for $\lambda = 0$ and $t = 1$. For the baseline forecasts ($e = 2$), the optimal actions are full protection ($a_5$) given the two forecasts ($j = 4$ and $j = 5$) indicating the highest probabilities for the most damaging events, 75% protection ($a_4$) given the moderately unfavorable forecasts ($j = 2$ and $j = 3$), and 50% protection ($a_3$) given the forecast ($j = 1$) with the lowest probabilities for unfavorable weather. These relatively high levels of protection are optimal because protection is inexpensive relative to the loss. Given only climatological information, the optimal action [Eq. (7)] is always full protection ($a_5$). The unshaded areas in Fig. 7 represent situations where too many additional decision periods remain before the end of the sequence for protection to be worthwhile. Protection in these instances would subject the decision maker to the risk of spending as much or more in protection costs than the avoided losses, by the time of the last decision.

The ensemble forecasts achieve economic value with respect to the baseline forecasts in the static problem on the basis of different actions being optimal for $j = 1$ and $j = 2$ given the more skillful forecasts. In these instances, one level of protection less is justified on the basis of the sharper probabilities in the good ($e = 1$) forecasts. Given the differences in expected losses, these gains outweigh the relatively poorer economic performance of the decisions that must be made when the less accurate forecasts are available. In this latter case, one level of protection greater is optimal for the two forecasts $j = 3$ and $j = 1$, as a consequence of the more diffuse probabilities available when $e = 3$.

In the dynamic decision problem, for which $t > 1$ and $\lambda \geq 0.25$ are also relevant, the same pattern of differences in optimal actions is evident in Fig. 7 for most combinations of $j$ and $t$. That is, forecast value can be realized by avoidance of unnecessary protection when the more skillful forecasts are available, which more than balances the requirement for more caution (higher levels of protection) when the less skillful forecasts occur. This is clear for $\lambda = 0$, for which the optimal action does not change over the full decision sequence. It also occurs for situations where partial losses have been sustained ($\lambda \geq 0.25$), and increasing levels of protection are optimal for decisions nearer to the end of the sequence. In these instances the ability to defer these protection costs when the more skillful forecasts are available more than outweighs the necessity for earlier caution in those decisions when the less skillful forecasts are available.

## 5. Summary and conclusions

This paper has presented an initial, idealized examination of the potential economic value of ensemble-based forecasts of surface weather elements, using hypothetical probability forecasts that may be broadly



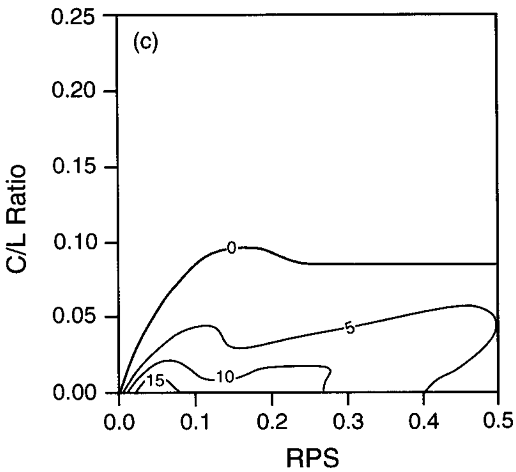




FIG. 6. Relative value of ensemble forecasts [Eq. (12)] in the dynamic decision problem, as a function of the RPS and different values of the *C/L* ratio, at $t = 15$ days from the end of the decision sequence. The climatological parameters $\alpha$, $\beta$, and $\pi$ correspond to (a) those in Fig. 2, (b) those in Fig. 3, and (c) those in Fig. 4.

representative of those that could be available from future ensemble-based forecast guidance. Probability forecasts were constructed exhibiting accuracies both higher or lower than that of baseline forecasts, and the hypothetical ensemble forecasts were drawn from these in such a way that the baseline and ensemble-based forecasts were equally accurate overall, as measured by the RPS. That is, according to this familiar scalar measure of forecast accuracy, the ensemble and baseline forecasts were equivalent. The computed economic value of the ensemble forecasts is accordingly attributed to the additional information that would derive ultimately from discrimination by the forecast ensemble of situations that are less- or more-predictable than usual.

For the most realistic combinations of the cost/loss ratio and overall forecast accuracy, the modeled ensemble-based forecasts show positive economic benefit relative to the baseline forecast probabilities. This is an encouraging result with respect to the eventual operational adoption of ensemble-based forecasts for sensible surface weather elements. The degree of benefit is relatively insensitive to the climatological probabilities for the different events, which correspond to decision problems at different locations or times of year.

It has been assumed here that the conventional and ensemble-based forecast systems are competing, in the sense that information originating from either one or the other, but not both, would be available for use by decision makers. This assumption is consistent with the practical constraint that computing resources are limiting in the operational forecast centers. We see the present results as contributing to the ongoing discussion regarding whether increased computing resources are best allocated to increased model resolution, or to multiple integrations at some lesser resolution (Brooks and Doswell 1993; Mullen and Baumhefner 1994). Of course if guidance from both a single high-resolution dynamical forecast and an ensemble of lower-resolution forecasts were available, a quite different analysis than that presented here would be appropriate.

Both static and dynamic decision models indicated positive economic value for the hypothetical ensemble-

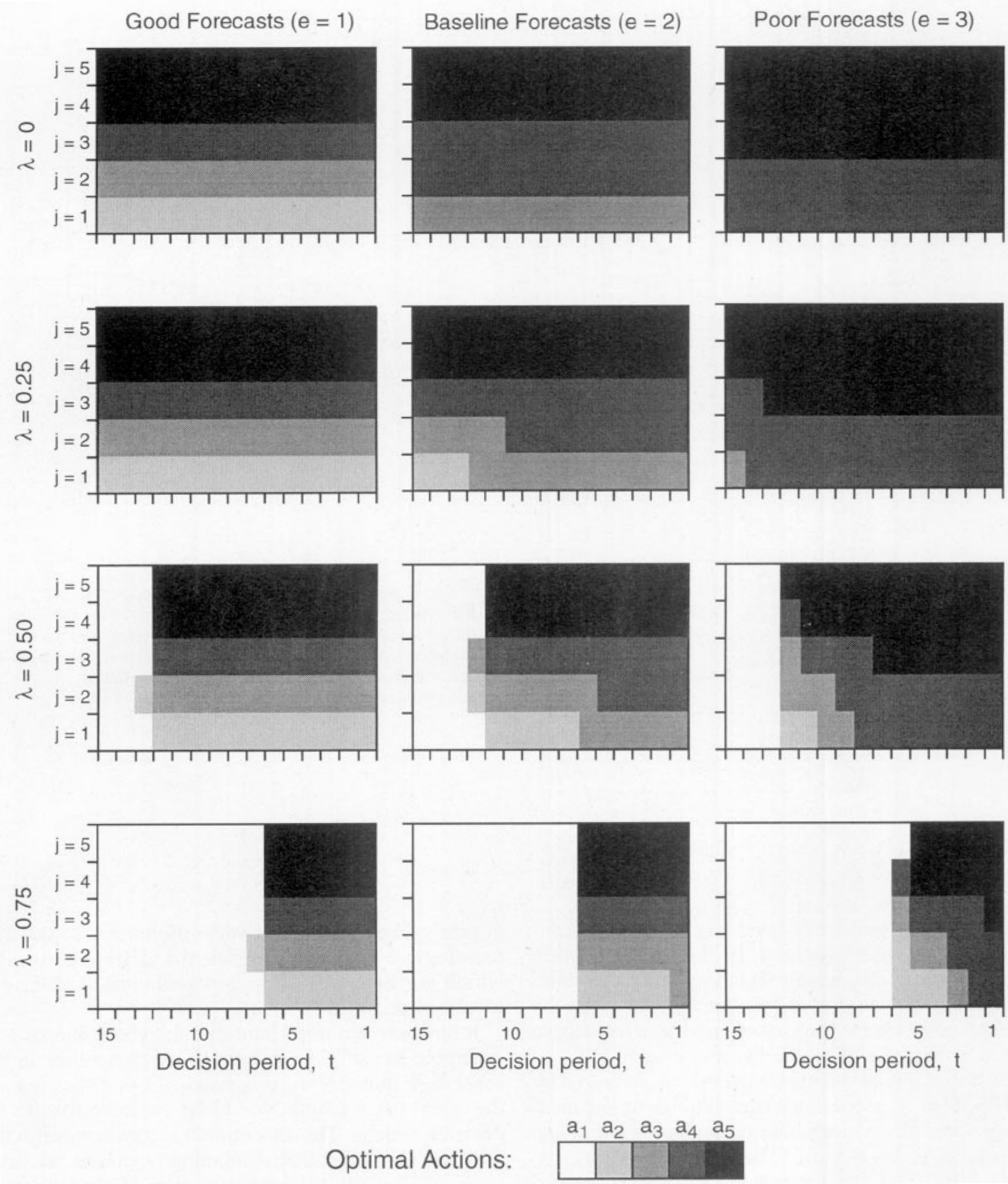



FIG. 7. Optimal actions as a function of time $t$ in the decision sequence, the forecasts in Table 1, and the loss $\lambda$ sustained to date, for the dynamic decision problem with $C/L = 0.05$.

based forecasts. However, because of the idealized nature of the model decision problem, the present results are at best suggestive of eventual economic benefits. Modeling real-world decision problems requires structures that are much more elaborate (e.g., Katz et al. 1982; Wilks et al. 1993), but that are broadly similar to those used here. The possible autocorrelation of forecast skill, which would reflect the existence of persistent regimes with high- and low-skill forecasts, was not treated. However, previous experience with the inclusion of autocorrelation of meteorological events into the dynamic decision problem used here (Wilks 1991) suggests that this aspect of the forecasts should further increase forecast value in relation to sequences of decisions, as it implies the availability of yet more information to decision makers.

Finally, an unexpected result is the existence of combinations of decision problems (i.e., $C/L$ ratios), climates, and overall forecast quality, for which the ensemble forecasts as constructed here yield lower economic value than the baseline forecasts. In part, this result is a consequence of the hypothetical ensemble forecasts being constrained to have overall RPS that is equal to that of the baseline forecasts. It is probably justifiable to assume that ensemble-based forecasts will exhibit greater accuracy in aggregate, because the discrimination of less- and more-predictable forecast situations provides additional information. Furthermore, for forecasts of midtropospheric geopotential height fields, it is observed that the accuracy (as measured by the anomaly correlation) of the average over a forecast ensemble is greater than that of the corresponding single higher-resolution conventional realization. However, the degree to which ensemble-based forecast guidance may be more accurate in aggregate than its conventional counterpart will not be known until such guidance has been constructed and tested.

Even though negative value is not found in the most relevant portions of the parameter space, and the computed negative values are probably overstated, this result has important implications and should be the subject of further investigation. In particular, it implies that positive economic value will not flow automatically from ensemble-based forecasts. Rather, the ensemble-based forecasts must be of sufficiently higher quality in aggregate than their conventional counterparts for them to be universally more valuable. For particular sets of ensemble-based and conventionally produced forecasts, it is possible to determine whether this condition holds (Ehrendorfer and Murphy 1988), but this determination must also await availability of experimental or operational ensemble-based surface weather forecasts.

REFERENCES

Brooks, H. E., and C. A. Doswell III, 1993: Numerical technology and numerical weather prediction—A wasted opportunity? *Weather,* **48,** 173–177.

——, ——, and R. A. Maddox, 1992: On the use of mesoscale and cloud-scale models in operational forecasting. *Wea. Forecasting,* **7,** 120–132.

Clemen, R. T., 1991: *Making Hard Decisions.* Duxbury Press, 557 pp.

Ehrendorfer, M., and A. H. Murphy, 1988: Comparative evaluation of weather forecasting systems: Sufficiency, quality, and accuracy. *Mon. Wea. Rev.,* **116,** 1757–1770.

Epstein, E. S., 1969: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.,* **8,** 985–987.

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.,* **115,** 349–356.

Katz, R. W., and A. H. Murphy, Eds., 1996: *Economic Value of Weather and Climate Forecasts.* Cambridge University Press, in press.

——, ——, and R. L. Winkler, 1982: Assessing the value of frost forecasts to orchardists: A dynamic decision-making approach. *J. Appl. Meteor.,* **21,** 518–531.

Kennedy, J. O. S., 1986: *Dynamic Programming. Applications to Agriculture and Natural Resources.* Elsevier, 341 pp.

Mullen, S. L., and D. P. Baumhefner, 1994: Monte Carlo simulations of explosive cyclogenesis. *Mon. Wea. Rev.,* **122,** 1548–1567.

Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.,* **119,** 299–323.

Murphy, A. H., 1969: On the ranked probability score. *J. Appl. Meteor.,* **8,** 988–989.

——, 1985: Decision making and the value of forecasts in a generalized model of the cost–loss ratio situation. *Mon. Wea. Rev.,* **113,** 362–369.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.,* **74,** 2317–2330.

Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center. *Wea. Forecasting,* **8,** 379–398.

Wilks, D. S., 1991: Representing serial correlation of meteorological events and forecasts in dynamic decision–analytic models. *Mon. Wea. Rev.,* **119,** 1640–1662.

——, 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction.* Academic Press, 464 pp.

——, R. E. Pitt, and G. W. Fick, 1993: Modeling optimal alfalfa harvest scheduling using short-range weather forecasts. *Agric. Systems,* **42,** 277–305.

Winkler, R. L., 1972: *Introduction to Bayesian Inference and Decision.* Holt Rinehart and Winston, 563 pp.

——, and A. H. Murphy, 1985: Decision analysis. *Probability, Statistics, and Decision Making in the Atmospheric Sciences,* A. H. Murphy and R. W. Katz, Eds., Westview, 493–524.