The Cloud Hunter's Problem: An Automated Decision Algorithm to Improve the Productivity of Scientific Data Collection in Stochastic Environments

ARTHUR A. SMALL III

Venti Risk Management, State College, Pennsylvania

JASON B. STEFIK* AND JOHANNES VERLINDE

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania

NATHANIEL C. JOHNSON

International Pacific Research Center, University of Hawaii at Manoa, Honolulu, Hawaii

(Manuscript received 30 July 2010, in final form 13 December 2010)

ABSTRACT

A decision algorithm is presented that improves the productivity of data collection activities in stochastic environments. The algorithm was developed in the context of an aircraft field campaign organized to collect data in situ from boundary layer clouds. Required lead times implied that aircraft deployments had to be scheduled in advance, based on imperfect forecasts regarding the presence of conditions meeting specified requirements. Given an overall cap on the number of flights, daily fly/no-fly decisions were taken traditionally using a discussion-intensive process involving heuristic analysis of weather forecasts by a group of skilled human investigators. An alternative automated decision process uses self-organizing maps to convert weather forecasts into quantified probabilities of suitable conditions, together with a dynamic programming procedure to compute the opportunity costs of using up scarce flights from the limited budget. Applied to conditions prevailing during the 2009 Routine ARM Aerial Facility (AAF) Clouds with Low Optical Water Depths (CLOWD) Optical Radiative Observations (RACORO) campaign of the U.S. Department of Energy's Atmospheric Radiation Measurement Program, the algorithm shows a 21% increase in data yield and a 66% improvement in skill over the heuristic decision process used traditionally. The algorithmic approach promises to free up investigators' cognitive resources, reduce stress on flight crews, and increase productivity in a range of data collection applications.

1. Introduction

This paper presents a decision algorithm developed to improve the efficiency of scientific data collection in stochastic environments. The Atmospheric Radiation Measurement (ARM) Program within the climate science programs of the U.S. Department of Energy has objectives involving the routine collection of data in situ from particular cloud formations by means of specially equipped aircraft (more information available online at http://www.atmos.uiuc.edu/~mcfarq/aavp.whitepaperoverview.pdf). Each day during a field campaign, investigators must decide whether or not to deploy the aircraft on the following day. Investigators traditionally have made these fly/no-fly decisions through a process involving heuristic analysis by experienced human investigators of forecasts of atmospheric conditions. Since budgeted flight hours are limited and expensive, and since available forecasts of suitable conditions are imperfect, investigators view the deployment decisions as having high stakes. In these decisions, two considerations must be balanced: the uncertain data value of the immediate opportunity and the cost associated with using up, from the fixed budget, flight hours that might otherwise be held back for use at a later date. The forecasts provide some information about the estimated value of the immediate prospect. Regarding

* Current affiliation: Risk Management Solutions, Inc., Hackensack, New Jersey.

Corresponding author address: Arthur A. Small III, Venti Risk Management, 200 Innovation Blvd., Suite 253, State College, PA 16803.
E-mail: arthur.small@ventirisk.com

DOI: 10.1175/2010MWR3576.1

© 2011 American Meteorological Society

the more distant opportunities for which flight time might be husbanded, only general statistical information is available. Estimating the second term—the value of *not* flying—appears to represent a particular challenge to even the most experienced investigators. Participants generally view this decision process as time consuming, tedious, stressful, and unpleasant.

This paper presents an alternative process for making these decisions: an automated decision algorithm. In the present study, the investigators' fly/no-fly decision problem is modeled formally, in terms of the investigator's objectives, resource constraints, and available information. This formal model is then analyzed with the aid of dynamic programming, a well-established optimization technique especially suitable for applications involving sequential decisions taken under uncertainty (Bellman 1957; Kamien and Schwartz 1991). Dynamic programming has been used to address problems in fields as diverse as hydrology (how much water to release from a reservoir; Stedinger et al. 1984), financial analysis (when to exercise the option to undertake an irreversible investment; Dixit and Pindyck 1994), meteorology and public safety (when to order a hurricane evacuation; Regnier and Harr 2006), and sports (whether or not, in a game of American football, to punt on a fourth down; Romer 2006). [For a recent survey of the relatively spare body of work at the intersection between meteorology and operations research see Regnier (2008).]

Applied in the present context, dynamic programming is used to derive a decision tool offering bright-line fly/no-fly recommendations, without making extensive use of investigators' time or other expensive resources. The algorithm's performance was tested by means of a retrospective application to a recent ARM field campaign. The analysis indicates that the algorithm, applied mechanically, would have yielded a greater rate of successful data collection per available flight than was achieved by means of the heuristic techniques actually employed by the investigators.

The work introduces two innovations in technique. These include, first, a generalizable method for translating a deterministic forecast produced by a numerical weather prediction model into a calibrated estimate of the probability of encountering specified atmospheric conditions. This approach to probabilistic forecasting is based on the tool of self-organizing maps (Kohonen 2001), a nonparametric numerical clustering technique that has been used in some applications in the atmospheric sciences (Hewitson and Crane 2002; Johnson et al. 2008). The second innovation involves the use of dynamic programming for optimizing the deployment of data collection resources in the context of a field campaign in the atmospheric sciences.

Formal analysis of the sequential fly/no-fly decision problem, and the application of optimization tools in that context, are apparently new in the atmospheric sciences. The analysis does nonetheless have antecedents in other applications involving the optimal deployment of scarce data collection resources. Students of forecasting have, for example, devoted substantial attention to issues of *targeted* or *adaptive* observing strategies (Lorenz and Emanuel 1998; Majumdar et al. 2001; Langland 2005), which concern decisions about where and when to deploy data collection resources in order to improve, through data assimilation, the accuracy of forecasting models. Because it involves a probabilistic analysis of the information value of a data collection program, adaptive observing forms a niche within the field of preposterior analysis (Berger 1993), itself a category within the larger domain of experimental design.

The most fundamental difference between the adaptive observing literature and the present work concerns the goals being pursued. The goal of adaptive observing is to produce better forecasts; observations are collected solely as a means to that end. In the present work, this relationship between means and ends is reversed. Data collection is itself the goal, while forecast quality is important only insofar as it enhances the scientists' ability to collect more data. A second, closely related distinction separating the present work from that on adaptive observing concerns the absence of a feedback loop between forecasts and observations. In adaptive observing, a savvy deployment of data collection resources will lead to a more accurate forecast, which may in turn influence the choice of deployment strategies in future periods. In the present work, there is no such feedback loop. The sampling program being optimized does not influence the forecasting system: deployment decisions have no effect on the forecast information that will be available later. A decision to deploy affects the future progress of the field campaign only as it depletes the stock of available data collection resources.

2. A formal model of the Cloud Hunter's Problem

Developing an automated decision algorithm begins by crafting a formal model of the decision challenge that the investigators confront. We dub this challenge "the Cloud Hunter's Problem."

The Cloud Hunter's basic problem is to decide how to allocate a limited budget of flight time over the course of a field season. Let D denote the length of the field season in days, and let $F \leq D$ denote the number of flights in the Cloud Hunter's budget. Let $d = D, \ldots, 1$ index dates, where by convention d will denote the number of days remaining in the field season. Let x_d be a binary random variable that takes the value 1 if on date d atmospheric conditions are suitable for data collection (d is a "good" day), and 0 otherwise. Let a_d be a binary control variable that takes the value 1 if a flight is made on date d, and 0 otherwise. Let $\mathbf{x} = \langle x_D, x_{D-1}, \ldots, x_1\rangle$ denote the vector of conditions during the field season, and let $\mathbf{a} = \langle a_D, a_{D-1}, \ldots, a_1\rangle$ denote the vector of all decisions taken. Then $\mathbf{a}\cdot\mathbf{x} = \sum a_d x_d$ is the realized number of successful data collection days.

The investigators' challenge can be posed as a problem of constrained optimization under uncertainty: for each day $d = D, \ldots, 1$, make a flight decision $a_d \in \{0, 1\}$ with the goal to maximize in expectation the total number of successful data collection days. This problem can be stated formally:

$$\max_{\mathbf{a}} E[\mathbf{a}\cdot\mathbf{x}] \tag{1}$$

subject to the budget constraint $\sum a_d \le F$.

The Cloud Hunter does not schedule all flights at the beginning of the field season. Rather, each flight decision a_d is made on the basis of information available on the previous day, date $d + 1$. On day $d + 1$, the Cloud Hunter receives a signal s_d that corresponds to a forecast of atmospheric conditions that are expected to prevail on the following day. Each forecast is drawn from a set $\mathbf{S}$ of possible forecast signals. Based on previous experience, the Cloud Hunter has developed a decoding function $p(\cdot)$ that converts a given forecast signal s into an estimated probability $p(s)$ that the next day will offer auspicious conditions for data collection: $p(s) = \text{Prob}\{x_d = 1|s_d = s\}$.

The Cloud Hunter naturally does not know what the forecast signal s_d will be, prior to receiving it. However, the Cloud Hunter does know the climatology of the forecasting process (i.e., the likelihood of receiving any given forecast signal). Let $\pi_d(s)$ denote the unconditional, ex ante probability that the forecast for date d will take the value s. The function $\pi_d(\cdot)$ defines a probability distribution over the set $\mathbf{S}$ of possible forecast signals. We call this function $\pi_d(\cdot)$ the *climatology of forecast signals* for date d. The daily forecasting process can thus be thought of abstractly as a discrete stochastic process on the set $\mathbf{S}$ that generates a sequence of signals $s_D, s_{D-1}, \ldots, s_1$, according to a known sequence of probability distributions $\pi_D(\cdot), \pi_{D-1}(\cdot), \ldots, \pi_1(\cdot)$. In this first-pass treatment we assume that the forecast signals are drawn independently, setting aside the realistic but complicating possibility that the forecasting process may exhibit autocorrelation.[1]

The Cloud Hunter's Problem involves choosing, based on the forecast conditions, whether or not to prepare to fly on the subsequent day. Formally, the goal is to find an optimizing decision rule $a(d, f|s_d)$ that takes a value $a_d = 1$ (fly) or $a_d = 0$ (no fly) as a function of the number d of days left in the field season, the number f of flights remaining in the budget, and the forecast s_d of the next day's atmospheric conditions. A decision rule $a(\cdot)$ is deemed optimal if its consistent application maximizes in expectation the yield of successful flights realized from the given budget.[2]

3. Optimization via dynamic programming

The search for an optimal decision rule employs dynamic programming, a technique suitable in applications involving sequential decisions under uncertainty. The central concept of dynamic programming is captured in the principle of optimality articulated by Bellman (1957): "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." This principle suggests a solution strategy that involves breaking up a problem into two pieces: the immediate next decision and all the decisions that come thereafter.

Suppose, then, that an optimal decision rule, denoted $a(d, f|s)$, has been found. Let $V(d, f)$ denote the expected number of successes that would be realized from repeated application of this rule, starting from initial conditions $\langle d, f\rangle$. By definition, $V(\cdot)$ equals the maximand of the decision problem (1), beginning from these initial conditions, under the substitution $a = a(d, f|s_d)$:

$$V(d, f) = E\left[\sum_{i=d,\ldots,1} a(i, f|s_i)\times x_i\right], \tag{2}$$

where the expectation is taken over the probability distribution of all possible sequences of forecast signals

[1] It is not necessarily assumed that the forecast signals are identically distributed, however. Nor would it be difficult to incorporate the possibility of autocorrelation in the sequence of forecast signals. One would simply replace, at appropriate points, the unconditioned probability distribution function $\pi_d(\cdot)$ with the corresponding function conditioned on realizations at earlier dates.

[2] There could, in practice, be cases in which investigators might rationally pursue objectives other than maximization of expected value. For example, an investigator may wish instead to maximize the probability of collecting at least a specified minimum quantity of useable data—say, an amount determined in advance to be necessary to achieve a desired degree of statistical significance. Such a situation could be modeled by substituting an alternative objective function. Expected value maximization is, however, both analytically tractable and quite realistic for many if not most scientific problems.

$s_d, s_{d-1}, \ldots, s_1$.[3] Bellman's principle of optimality suggests that $V(\cdot)$ be expressed as the sum of two terms: the expected payoff associated with the next day's flight decision, plus the expected number of successful flights realized during the remaining balance of the season, assuming that all decisions are made optimally. Formally, Bellman's principle of optimality implies that the value function $V(\cdot)$ must satisfy a particular recursive relationship:

$$V(d, f) = E_{\{s_d\}}[a_d^* x_d + V(d-1, f-a_d^*)], \tag{3}$$

where $a_d^* = a(d, f|s_d)$ is the optimized choice for the date-d fly/no-fly decision. Here the expectation is taken with respect to the probability distribution $\pi_d(\cdot)$ over s_d, the next forecast signal to be received. The second term in brackets is called the *continuation value* of V, the expected number of successful flights made during the balance of the season (beginning date $d-1$), conditioned on the choice a_d^*, and assuming that all subsequent flight decisions are also made optimally.

Once the forecast s_d is received, the Cloud Hunter must decide whether or not to fly. If the decision is made to fly on date d (if $a_d = 1$), the expected payoff for that day is $E[x_d|s_d] = p(s_d)$. Since one flight has been used from the limited budget, the continuation value for the rest of the season is $V(d-1, f-1)$. If instead the decision is made to not fly ($a_d = 0$), then the payoff for date d is zero, and the flight is saved for possible future use. In this case, the continuation value is $V(d-1, f)$. Since the choice for a_d^* must be one that maximizes value in expectation, we must have

$$V(d, f|s_d) = \max\{p(s_d) + V(d-1, f-1), V(d-1, f)\}. \tag{4}$$

Equation (4) implies a form for an optimal decision rule: $a_d^* = 1$ (''fly'') if and only if

$$p(s_d) \geq V(d-1, f) - V(d-1, f-1); \tag{5}$$

otherwise, $a_d^* = 0$ (''no-fly''). The inequality (5) can be understood as a cost–benefit comparison: $p(s_d)$ is the expected benefit of flying, while $\Delta V = V(d-1, f) - V(d-1, f-1)$ is the expected cost, in units of foregone future successes, associated with beginning the following day (date $d-1$) with one fewer flight remaining in the budget.[4] The term ΔV is called the *hurdle probability* for the state $\langle d, f\rangle$.

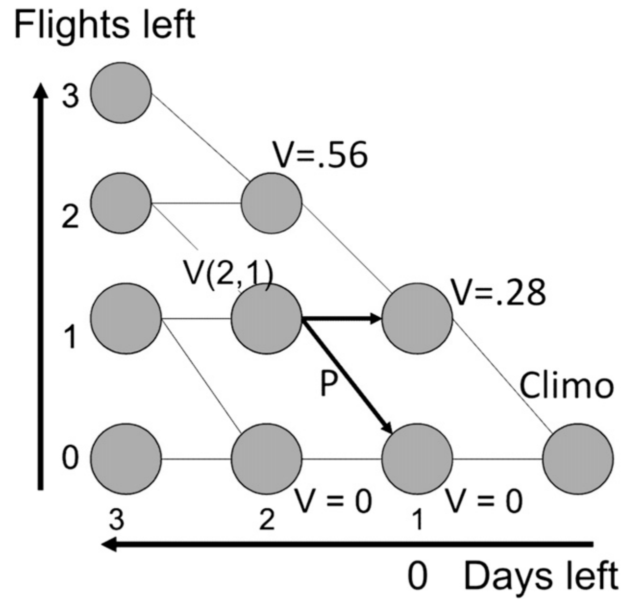

FIG. 1. A graphical representation of the decision algorithm. When there are 2 days remaining in the field season, and only 1 flight remaining in the project budget, the investigator chooses either to fly, or not. If the decision to fly is taken, the investigator has success with probability P, which is estimated based on a forecast of the next day's atmospheric conditions, and then drops to state $d = 1, f = 0$. Alternatively, the investigator can hold the remaining flight in reserve until the last day, when the project will be in state $d = 1, f = 1$. The remaining flight is then used with certainty, yielding success with a probability equal to the long-run climatological average frequency of encountering good conditions (here, 28%). State $\langle 2, 1\rangle$ is more valuable than state $\langle 1, 1\rangle$ exactly because the extra day gives the investigator the option to fly only in case the forecast probability exceeds the average climatological probability. Given knowledge of the long-run distribution of forecasts, the likelihood that the option will be exercised can be computed in advance, prior to the start of the field season.

This decision rule—''fly if and only if the estimated probability of good conditions exceeds the hurdle probability''—defines an optimal policy in terms of the continuation values $V(d-1, f)$ and $V(d-1, f-1)$. If those continuation values were known, then (4) would allow for the computation of $V(d, f) = E_{\{s_d\}}[V(d, f|s_d)]$. Since the choice of initial conditions d and f were arbitrary, (4) offers an inductive procedure for computing the value of the function V for any date and flight budget in terms of the values it could take on subsequent dates. The process is illustrated with the aid of Fig. 1.

[3] The correlation between forecast signals s_i and realized conditions x_i is suppressed in the interest of avoiding notational clutter. Of course, there must be some correlation: a ''forecasting'' system that provided no information about subsequent atmospheric conditions would be useless.

[4] In case the relation (5) holds with equality, the decision maker is indifferent between flying or not. We arbitrarily set the decision rule so that $p = \Delta V$ generates a decision to fly.

Two boundary conditions constrain the values V may take. First, we know that $V(d, 0) = 0$ for all d: if there are no more flights left in the budget, then there will be no more successful flights. A second boundary condition follows from the maintained assumption that unused flights left in the budget at the end of the season have no residual value. This assumption implies that it is always optimal to fly whenever the number of remaining flights equals the number of remaining days (i.e., whenever $f = d$) no matter what forecast signal is received. The expected payoff $V(d, d)$ counting from that date forward is just the long-run average (i.e., climatological) probability of good conditions on the days remaining in the season: for all d, $V(d, d) = E[p(s_d) + p(s_{d-1}) + \cdots + p(s_1)]$.

Together with these boundary conditions, the recursive relationship (4) allows $V(d, f)$ to be computed for arbitrary conditions d and f by taking expectations over the climatology of forecast signals:

$$V(d, f) = E_{\pi(s_d)}[V(d, f|s_d)]. \quad (6)$$

The inequality condition (5) then allows the decision rule $a(d, f|s_d)$ to be implemented for arbitrary conditions $\langle d, f\rangle$.

It bears emphasizing that V is an unconditioned expectation; its values do not depend on forecast realizations. The value the function takes at each node in the $\langle d, f\rangle$ lattice can, therefore, be computed in advance, before the start of a field season. (The details are provided later. Figure 5 displays this idea graphically.) One can likewise compute in advance the hurdle probability ΔV for each node. Finally, given the optimal decision rule $a(d, f|s_d)$ and information about the climatology of forecast signals, one may readily compute the ex ante probability that a campaign in state $\langle d, f\rangle$ will reach a decision to fly:

$$\begin{aligned}\mathrm{Prob}[a(d, f) = 1] &= E_{\{s_d\}}[a(d, f|s_d)]\\ &= \sum_{s_d \in \mathbf{S}} p(s_d) \times a(d, f|s_d).\end{aligned} \quad (7)$$

Repeated application of (7) allows the calculation of derivative terms such as the probability of following any particular path through the $\langle d, f\rangle$ lattice, or the probability that a crew will be required to fly on a specified date. By providing guidance about the likelihood of flying under varying conditions during the season, (7) could evidently be useful to project scientists and crew as they plan their activities during the field campaign.

4. Quantifying success probabilities using self-organizing maps

The Cloud Hunter's Algorithm was applied retrospectively to the flight scheduling challenges associated with the Routine ARM Aerial Facility (AAF) Clouds with Low Optical Water Depths (CLOWD) Optical Radiative Observations (RACORO) campaign, a recent data collection effort within the ARM program (more information is available online at http://acrf-campaign.arm.gov/racoro). For the RACORO investigators, desirable conditions featured the presence of liquid boundary layer clouds above the Southern Great Plains (SGP) ARM Climate Research Facility (ACRF) located near Lamont, Oklahoma.

To operationalize the algorithm requires a technique for estimating, based on a day-ahead forecast, the likelihood that a flight made on the following day will encounter boundary layer clouds over the SGP site. More exactly, it requires that there be a specified set $\mathbf{S}$ of possible forecast signals; a mapping $p(s) = \mathrm{Prob}\{x_d = 1|s_d = s\}$ that converts forecast signals into quantified probabilities of encountering good conditions; and, for each date d during a field season, climatological probabilities $\pi_d(s)$ of receiving each type of forecast signal.

One might consider implementing the decision algorithm for this application by using the output from a standard numerical weather prediction model to create a day-ahead forecast of the likelihood that suitable cloud formations will be observed above the SGP site. However, rather than relying on model physics directly to predict the presence or absence of boundary layer clouds, the implementation strategy adopted relied on the known relationship between vertical profiles of relative humidity (RH) and observed boundary layer cloudiness (Berg and Kassianov 2008). The RH profile forecasts are generated routinely by standard numerical weather prediction models such as the Global Forecast System (GFS) model maintained by the National Centers for Environmental Prediction (Kanamitsu et al. 1991). Taking advantage of this correspondence, the day-ahead RH profile forecast produced by the GFS model was used as the basis for estimating the probability that the next day's sky will feature boundary layer clouds above the SGP site.

This strategy presents at least two difficulties. One difficulty concerns the uncertainty and complexity that characterizes the relationship between RH profiles and the presence or absence of boundary layer clouds. Ideally, one would like to create a predictive model of these relationships, calibrated on the basis of historically observed correlations between the two phenomena. The space of all RH profiles is, however, high dimensional, and the relationship between RH profiles and boundary layer clouds is neither simple nor linear. A given historical RH profile observation may not be directly comparable to a sufficiently large number of others to provide robust guidance about the degree of association between that RH profile type and the observed frequency of boundary layer clouds. This complexity makes it problematic to use standard statistical techniques (e.g., linear regression) to calibrate the mapping from an arbitrary RH profile to an estimated probability of boundary layer clouds.

a. Dimension reduction using self-organizing maps

The dimensionality issue was addressed by deploying the dimension-reduction tool of self-organizing maps (SOMs), a nonparametric clustering technique based on neural networks (Kohonen 2001; Johnson et al. 2008). To create a SOM, the analyst specifies a metric that defines the "distance" between any two elements in the dataset, as well as the number of clusters into which elements will be organized. Applied to the historical RH profile data, the SOM method groups similar RH profiles together into clusters, representing each cluster by a single canonical member. The SOM method presents the canonical members in a structured two-dimensional grid such that clusters with similar members are located close together while clusters with dissimilar members are more widely separated.

Using this approach, a 24-member[5] SOM was created from relative humidity data derived from the North American Regional Reanalysis (NARR; Mesinger et al. 2006) for the period 1979–2008 for the 1000–600-mb layer above the model grid point nearest to the SGP site. The NARR was used because it provides a 30-yr time series, analyzed in a consistent way throughout this period. In the absence of high-frequency, detailed observations of the entire air column above the SGP site, the NARR provides the best available estimate of prevailing atmospheric conditions. The distance metric employed to create the SOM was the standard Euclidean (L^2) norm: the distance between two RH profiles was defined to be the square root of the sum of the squares of the difference in relative humidity at each pressure level (more information is available online at http://www.esrl.noaa.gov/psd/data/gridded/data.narr.html).[6] The model grid point used is located at 36.57°N, 97.56°W, less than 7 km from the SGP facility located at 36.69°N, 97.56°W. The SOM was generated using NARR RH profiles for 1800 UTC on each day. This last choice reflects the fact that the strong majority of RACORO flights were flown during the window 1300–2300 UTC.

To each SOM cluster one can associate a probability of boundary layer clouds being present. A simplifying assumption is imposed that all members of a given SOM cluster are sufficiently similar that they share a common probability of producing boundary layer clouds. If the number of elements in the cluster is not small, then the empirically observed frequency of boundary layer clouds for the members of this cluster provides a good estimate of this common, true underlying probability. In present study, each SOM cluster contains over 100 members for which the presence or absence of boundary layer clouds can be verified. The estimated probability $\tilde{p}(\tilde{s})$ that an arbitrary member of cluster $\tilde{s}$ will be associated with the presence of boundary clouds is thus given by the empirical frequency:

$$\tilde{p}(\tilde{s}) = n_{\tilde{s}}/N_{\tilde{s}}, \tag{8}$$

where $n_{\tilde{s}}$ is the number of relative humidity profiles assigned to SOM cluster $\tilde{s}$ for days on which boundary layer clouds are present, and where $N_{\tilde{s}}$ is the total number of profiles assigned to SOM cluster $\tilde{s}$ over the period of record for which boundary layer cloud data are available.

The presence of boundary layer clouds at the SGP on a given date was verified based on cloud fraction data obtained from ARM's Climate Modeling Best Estimate product (CMBE; Xie et al. 2010). The CMBE algorithm calculates mean hourly cloud fractions for the SGP site at 45-m vertical resolution. These hourly cloud fractions are derived from ARM's Active Remotely-Sensed Clouds Locations (ARSCL) value-added product (Clothiaux et al. 2001, 2000). The ARSCL-derived cloud fractions are assumed to reflect accurately the true state of clouds in the atmosphere. Using the CMBE data, each 1-h interval in the period of record (1998–2008) was coded for

[5] In choosing an SOM grid size an analyst must balance two concerns. Reducing the grid size will tend to increase mean quantization error, the average dissimilarity between elements grouped together in a cluster (Kohonen 2001). Increasing the grid size reduces the number of elements in each cluster and, therefore, the statistical significance of estimates of cluster-specific characteristics. A sensitivity analysis showed that results of the study were essentially unchanged for grid sizes ranging from 9 to 64 (see Stefik 2010).

[6] The pressure levels used in the distance calculation were the same as those used by the NARR model. The NARR uses a vertical resolution 25 mb from 1000–700 mb and 50 mb from 700–600 mb. Hence, the distance calculation was based on differences in RH at the 1000-, 975-, 950- . . . 725-, 700-, 650-, and 600-mb levels.

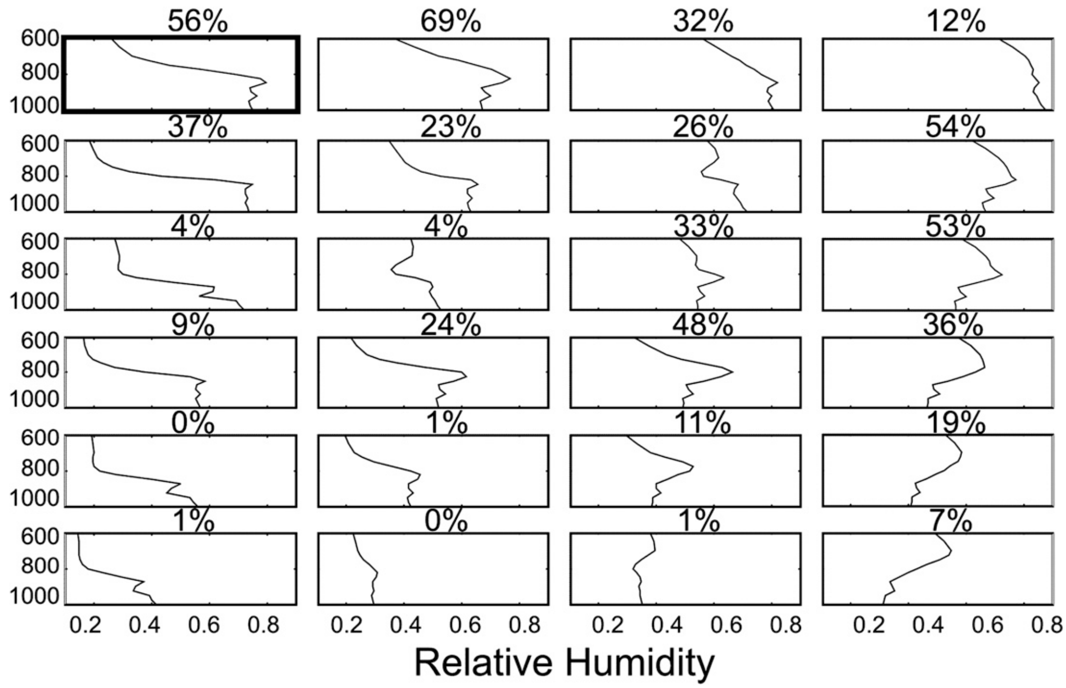

FIG. 2. The SOM grid for relative humidity profiles at the Lamont site, based on daily data derived from NARR (Mesinger et al. 2006) for the period 1979–2008. Each cluster is represented by a canonical member, shown. Percentages reflect the historically observed frequency of boundary layer clouds among members assigned to each respective cluster. Among dates for which relative humidity profiles were mapped to cluster 1 (highlighted), boundary layer clouds were observed 56% of the time.

the presence or absence of boundary layer clouds. Boundary layer clouds were deemed present if and only if all of the following conditions held over the SGP site during the entirety of the hour: (i) the cloud fraction exceeded 10%; (ii) the cloud top was below 3 km; (iii) the cloud base was above 400 m; (iv) no precipitation (less than 1/100th in.) was detected. [The 3-km threshold was used in accordance with the specifications contained in the RACORO proposal (Vogelmann 2008), notwithstanding variability in actual boundary layer depths.] A day was classified as having conditions suitable for data collection if and only if boundary layer clouds were present for at least 4 h during the window 1300–2300 UTC, when most RACORO flights were taken. A sensitivity analysis showed similar results as this threshold was varied between 3 and 6 h (see Stefik 2010).

Using these verification data, (8) was used to estimate, for each SOM state, an associated probability of the presence of boundary layer clouds. Figure 2 shows the resulting SOM for RH profiles at the Lamont SPG site, together with graphs of the representative SOM members and the empirical frequency of boundary layer clouds for the elements of each cluster.

The SOM and (8) together provide an empirically calibrated basis for estimating the likelihood that a given RH profile will be associated with the presence of boundary layer clouds. A specified RH profile can be assigned to 1 of the 24 SOM clusters by matching it with the canonical member to which it is most similar. Equation (8) can then be used to estimate the likelihood that boundary clouds would be present at the SGP site on that same day.

b. Probabilistic forecasting of SOM states

If existing numerical prediction models could offer perfect day-ahead RH profile forecasts, then this two-step technique would provide a sound basis for estimating the probabilities of boundary layer clouds. In practice, existing numerical weather prediction models cannot offer such precision. Figure 3 presents one articulation of the accuracy of 30-h-ahead RH profile forecasts produced by the GFS model. The figure displays the empirical probability distribution for SOM assignments of realized RH profiles, on days for which the GFS model forecast an RH profile that would be assigned to SOM cluster 1. If the GFS model were perfectly accurate, then a day-ahead forecast calling for conditions ''SOM cluster 1'' would be followed the next day by a realized RH profile falling into cluster 1, 100% of the time. In fact, on days for which the GFS model

13.3%	20.6%	4.4%	3.7%
1.5%	8.1%	1.5%	4.4%
0%	3.7%	4.4%	11.8%
2.9%	3.7%	7.4%	6.6%
0%	0%	0.7%	1.5%
0%	0%	0%	0%

FIG. 3. Conditional probability distribution of SOM state realizations following a forecast of SOM state 1. Each cell displays the conditional empirical likelihood of observing a relative humidity profile in the associated SOM cluster, on days for which the GFS model predicted a relative humidity profile falling into SOM Cluster 1. The GFS model accuracy is imperfect: on days for which the GFS model forecasts an RH profile assigned to SOM Cluster 1, the actual RH profile realized the next day falls into Cluster 1 (the upper-left cell) only 13.3% of the time—even less than the observed conditional likelihood of falling into Cluster 2.

forecasts an RH profile assigned to SOM cluster 1, the actual RH profile realized the next day falls into cluster 1 only 13.3% of the time—even less than the observed conditional likelihood of falling into cluster 2.[7]

In light of such forecasting errors, the procedure employed to estimate the probability of boundary clouds does not call for accepting the GFS forecast at face value. Instead, the historical performance of the GFS model is used to calibrate a conditional probability distribution over the set of possible SOM states that might be realized the next day. In notation, let s denote the number of the SOM cluster to which a given GFS RH forecast is mapped. For each $\tilde{s} = 1, \ldots, 24$, let $\pi(\tilde{s}|s)$ denote the conditional probability that the forecast of state s will be followed by a realized state $\tilde{s}$. These conditional probabilities are derived from the empirical joint probability distribution of forecast–observation pairs (Murphy 1997). In terms of this notation, Fig. 3 displays the conditional probabilities $\pi(\tilde{s}|s = 1)$ for each of the 24 SOM clusters.

[7] Fig. 3 shows that the GFS model has a difficult time predicting SOM clusters of RH profiles. This difficulty may arise in part as an artifact of the differences between the GFS and NARR models with respect to numerical techniques, grid size, and physical parameterizations. In the SOM mapping step, the GFS model is in effect used to predict, not the (unobserved) state of the atmosphere itself, but rather the NARRs representation of the atmosphere. Technical differences in model design and output format can be expected to cause at least some difference in the mapping to SOM clusters. (We are grateful to an anonymous reviewer for pointing out this issue.)

To summarize, the technique used to generate a day-ahead estimate of the probability of boundary layer clouds at the SGP site involves four steps. First, the GFS model is used to generate a 30-h-ahead forecast of the vertical relative humidity profile above the model grid point nearest to the SGP site. Second, this forecast RH profile is assigned to 1 of 24 SOM states by matching it with the canonical SOM member it most closely resembles. Based on the historic track record of the GFS model, this point-estimate forecast is then converted into a conditional probability distribution over the set of all SOM states. Finally, the probability of boundary layer clouds is computed as the weighted average of the empirical probabilities observed historically for each SOM state, with weights given by the conditional probabilities determined by the GFS calibration. Formally, if a 30-h-ahead RH forecast produced by the 1200 UTC GFS model run is assigned to SOM Cluster s, then the probability of boundary clouds the next day is estimated by

$$p(s) = \mathrm{Prob}\{x = 1|s\} = \sum_{\tilde{s}=1}^{24} \pi(\tilde{s}|s) \times \tilde{p}(\tilde{s}). \tag{9}$$

In this application, therefore, the set $\mathbf{S}$ of forecast signals is just the set of SOM states, while the mapping $p(\cdot)$ is given by (9). Figure 4 displays the estimated probabilities of observing boundary layer clouds for each of the 24 possible forecast signals.

To compute the values for the function $V(d, f)$ for arbitrary conditions $\langle d, f \rangle$ also requires a specification of the climatological probability distributions governing the generation of forecast signals. For this estimation we used the historical archive of GFS forecasts for the period 2001–08 (see online at http://www.archive.arm.gov, product "sgpncepgfspprofX1.c1"). Each day's 30-h-ahead forecast at 1200 UTC for the RH profile at Lamont was assigned to 1 of the 24 clusters in the previously derived SOM per the mapping procedure described above. The resulting frequency distribution over SOM states was taken as the climatological distribution of forecast signals. Likewise, the empirical joint probability distribution of SOM state forecasts (based on the GFS model output) and SOM state observations (based on the NARR) was used to estimate the conditional probability weights $\pi(\tilde{s}|s)$. For this estimation procedure, historical data from all calendar dates were pooled; possible seasonal variations were not taken into account. The resulting estimated climatological distributions were thus treated as being the same for all dates: the forecast signals were treated as realizations of independent, identically distributed random variables.

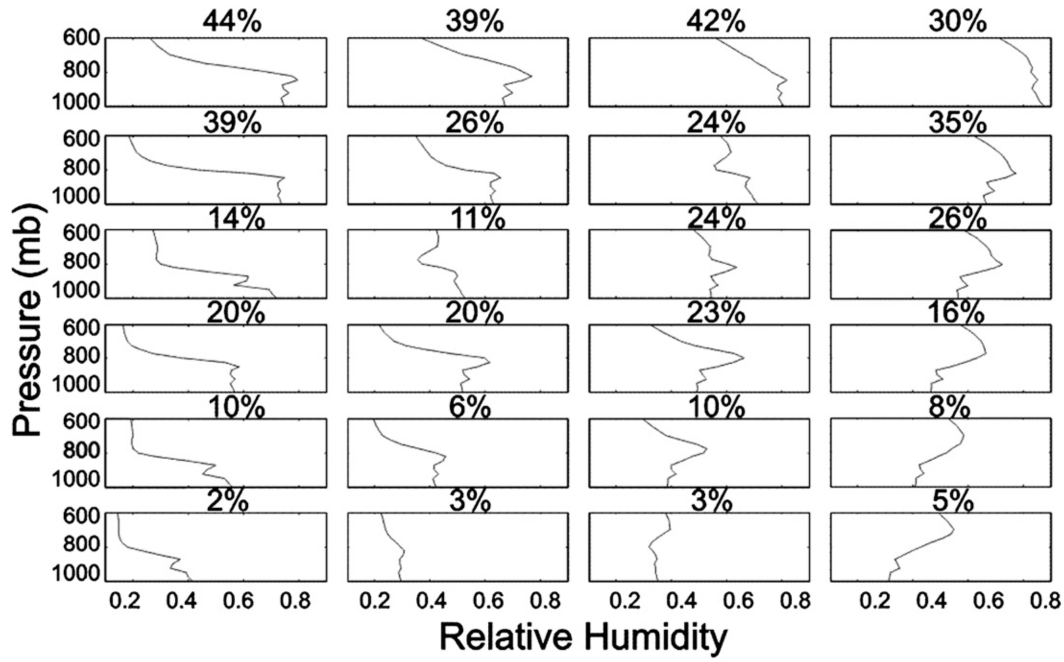

FIG. 4. Estimated probabilities of boundary layer clouds as a function of GFS SOM forecast. Probability estimates based on the GFS forecast are less sharp than those based on RH profile realizations (displayed in Fig. 2). They nonetheless provide a useful basis for discriminating between more vs less promising data collection opportunities, on a day-ahead basis.

5. Comparing decision procedures: Application to the 2009 RACORO field campaign

The RACORO campaign spanned a five-month field season, 22 January–30 June 2009. The objective was to collect measurements from boundary layer clouds over the SGP site by means of a specially equipped aircraft. Each flight required extensive preparations: equipment had to be tested, advance notice given to professional flight crews, and a flight plan filed with the U.S. Federal Aviation Administration. Because of the time required to carry out these preparations, a decision to launch had to be made on the day before the actual launch. Each day, therefore, investigators had to decide whether to prepare the aircraft and crew for a flight on the following day. Given its budget of 300 flight hours and a usage rate of approximately 4.5 h per flight, the RACORO team could expect to fly on approximately 67 days, or 42% of the time. Because of uncertainties concerning future atmospheric conditions, aircraft routinely were prepared, or even sent aloft, only to find no boundary layer clouds over the SGP site. Such type I errors represented an expensive waste of scarce resources, including the time of investigators and flight crew, and of limited supplies of flight hours within the project budget. Conversely, a type II error—a decision to stand down, only to find that boundary layer clouds were present after all—represented an irretrievable loss of a data collection opportunity.

a. RACORO's heuristic procedure for making fly/no-fly decisions

Given the stakes, daily fly/no-fly decisions were a major focus of discussions within the RACORO team. Fly/no-fly decisions were made by means of a labor-intensive process involving heuristic analyses by human investigators of day-ahead weather forecasts for the SGP site as well as other information such as satellite photos. These deliberations consumed on the order of tens of investigator hours per day during the field season (A. M. Vogelmann 2009, personal communication). The decision process was thus itself extremely expensive in terms of its required inputs of human capital and wear and tear. The process also imposed unquantifiable but very real costs in terms of extensive and repeated stress on investigators, flight crews, and other research staff, for whom the uncertainties surrounding flight decisions caused difficulties in planning other activities. Moreover, the heuristic decision process did not take maximum advantage of available resources: the 2009 RACORO season ended with 40 flight hours unused. (Aircraft hours are purchased in advance. There is no cost benefit in conserving flight hours.)

b. Formulation of the counterfactual scenario

The acid test of the algorithm's utility concerns its ability to improve the yield of usable data collected, while holding constant or reducing the cost of the decision-making process. To test its effectiveness, the algorithm was run using conditions that prevailed during the 2009 RACORO field season. A comparison was then made between the actual results achieved by the RACORO team during the 2009 field season, versus the results the team would have obtained through consistent application of the automated decision procedure. In order that the comparison be genuine and fair, the automated decision procedure was calibrated using only pre-2009 data, including the CMBE cloud data, NARR relative humidity profiles, and GFS relative humidity profile forecasts.

1) AUXILIARY ASSUMPTIONS

Applying the automated decision procedure to RACORO further required the imposition of certain auxiliary assumptions and adaptations.

(i) The 5-day rule on flight crews

Regulations promulgated by the U.S. Federal Aviation Administration constrain the frequency with which crews can operate: crew members are not allowed to fly on more than 5 days in any 7-day period. Because only one crew was available, these rules imposed real constraints on the operations of the RACORO campaign. For example, if the crew had flown on each of four previous days, a decision to fly on a fifth consecutive day implied not only an expenditure of scarce flight time, but also the loss of the option to fly on either of the two subsequent days. These regulatory constraints were incorporated into the decision procedure by refining the optimization model. In the dynamic programming procedure, the value function $V(d, f)$ defined by (2) was augmented to condition on the flight history over the previous 7 days: $V = V(d, f|a_{d+1}, a_{d+2}, \ldots, a_{d+7})$. The decision rule codified in the inequality (5) was rephrased to take the form: $a_d^* = 1$ (fly) if and only if $\sum_{k=1}^{6} a_{d+k} < 5$ and

$$p(s_d) \geq V(d-1, f|0, a_{d+1}, \ldots, a_{d+6}) - V(d-1, f-1|1, a_{d+1}, \ldots, a_{d+6}); \qquad (10)$$

otherwise, $a_d^* = 0$ (no-fly). The test (10) was applied daily based on the 30-h-ahead forecast generated by the 1200 UTC GFS model run. Figure 5 displays the values for the function $V(d, f)$ computed for the RACORO campaign.

(ii) Option to cancel a flight

RACORO managers had the option to cancel a scheduled flight at any time up to 3 h prior to a scheduled takeoff.

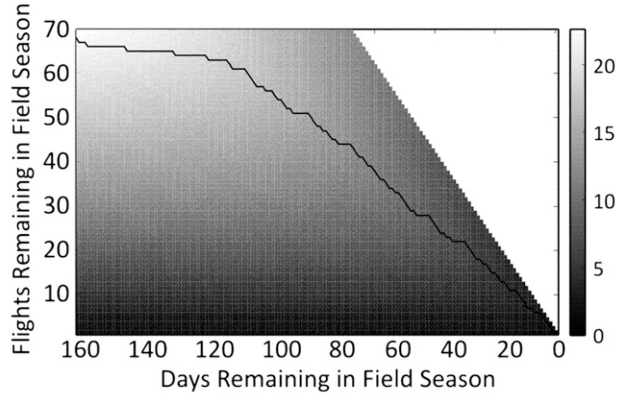

FIG. 5. Computed values for the function $V(d, f)$ for the RACORO campaign. Expected number of successful flights generated by the decision algorithm, as a function of the number of days remaining in the field season (d, horizontal axis) and the number of flights remaining in budget (f, vertical axis), represented by levels of shading. The black line represents the sequence of flight decisions the automated method would have issued had it been implemented under the conditions prevailing during RACORO field season. Diagonal movements correspond to dates on which flights would have been made; horizontal movements correspond to dates on which no flights would have been made.

This option would be exercised if the final weather forecast prior to launch showed a deterioration in cloud conditions that had previously been viewed as promising. A decision to cancel a scheduled flight does not, however, eliminate the burden on the flight and scientific crews associated with preparations for the scheduled launch. A version of the cancellation option was incorporated into the automated decision procedure. On days for which the 30-h-ahead 1200 UTC GFS forecast generated a decision to fly ($a_d = 1$), the test (10) was applied again based on the 6-h-ahead 1200 UTC GFS model run made the following day. This second test used a separate forecast calibration function based on the history of the 6-h-ahead GFS forecast. When running this second test, however, the burden on the flight crew was treated as a sunk cost for purposes of the 5-day rule. Formally, in the first term on the right-hand side of the inequality (10), the first conditioning term $a_d = 0$ is changed to $a_d = 1$. This change has the effect of reducing the hurdle probability required to justify a decision to fly. This secondary check led to flight cancellations on 7 days over the course of the 160-day field season.

(iii) Freezing temperatures

The forecasting system described in section 4 computes a probability that a given flight will encounter boundary layer clouds. The specifications of the RACORO mission stipulate an additional requirement, however, that the cloud be composed entirely of *liquid* water. Clouds with ice present occurred with enough frequency, particularly during the beginning of the field season, that this requirement could not be ignored. The need to avoid freezing temperatures was addressed by checking the GFS model forecast of temperatures in the air column above the SGP site. Because our cloud forecasting system does not provide exact information on the height of clouds (only that, on "good" days, the cloud top will be located below 3 km), the minimum temperature within a cloud was proxied by the GFS forecast of temperature at a height of 2.5 km. Although temperature forecasts were not integrated into the optimization based on dynamic programming, they were incorporated into the decision algorithm using a simple binary approach: if the 30-h GFS forecast predicted subzero temperatures at 2.5 km, then no flight was scheduled. If a flight was scheduled and the 6-h GFS forecast predicted a 2.5-km temperature below freezing, then the flight was cancelled.

2) Rules for scoring flights as successes or failures

The Cloud Hunter's Algorithm, as augmented per the above refinements and applied to data from the 2009 RACORO field season, generates a listing of 66 dates for which a decision would have been made to launch a flight. Records of the RACORO campaign likewise indicate 56 dates on which flights were in fact made and, for those dates, whether the investigators judged the flight a success or a failure. To perform a side-by-side performance comparison between the heuristic and automated decision procedures, a comparable counterfactual log needed to be developed of the results the investigators would have realized had they employed the Cloud Hunter's Algorithm instead. To create this counterfactual track record, some reasonable procedure had to be devised for scoring flights as successes or failures.

For those dates when both the heuristic and automated procedures both generated decisions to fly, a natural rule presents itself: each date was coded a success or failure following the scoring applied to that flight by the RACORO team. There were 31 such dates, of which 22 were scored as successes, 9 as failures. The RACORO records do not, however, include information about the suitability of conditions on dates when the RACORO team did not fly. To compute results for the automated procedure for the remaining 35 such dates required some other means for classifying flights as successes or failures.

To do so, we employed the same definitions and procedures used to calibrate the cloud forecasting system as articulated in section 2, applied now to 2009 data. Again using the CMBE cloud dataset (although this time from January to June 2009), each hour during the RACORO field season was coded for the presence of absence of boundary layer clouds. Each hour was likewise coded for presence or absence of freezing temperatures, as defined by the 2.5-km temperature as recorded in the NARR. An hour was classified as good if boundary layer clouds were present and if the 2.5-km temperature remained above 0°C. A day was classified as good, and a flight on that day scored as successful, if good conditions existed for at least 4 h during the window 1300–2300 UTC.[8]

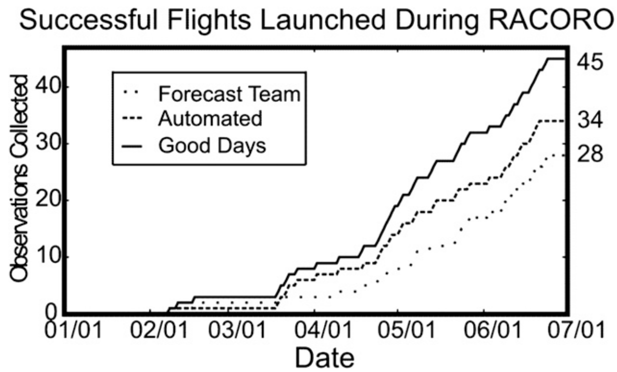

FIG. 6. Results: the algorithm's simulated performance during 2009 field season, compared with realized performance of RACORO campaign. The algorithm achieves a 21% increase in the number of successful flights.

c. *Results*

Results are summarized in Fig. 6. Of the 160 days in the field season, 45 (28%) featured good conditions suitable for data collection, as defined by the above scoring rules.[9] Using the heuristic decision procedure to make fly/no-fly decisions, the RACORO team collected data successfully on 28 of those days, thus harvesting data on 62% of the suitable data collection opportunities. Applying the automated decision rule to the same set of conditions would have yielded useful data on 34 days, constituting 76% of the potential data collection opportunities. In other words, the automated decision procedure showed a 21% improvement in yield over that of the heuristic procedure.

The difference in results is even more pronounced when considered in terms of decision *skill*. If fly/no-fly decisions had been made entirely at random, then one would expect flights to be successful at a rate of approximately 28%, the unconditioned frequency of good days among all days. The expenditure of 67 flights would then be expected to yield approximately 19 days worth

[8] Additional details on the scoring rules used in the RACORO application is presented by Stefik (2010), who also elaborates on the auxiliary assumptions incorporated into the counterfactual analysis.

[9] This classification was confirmed by RACORO project scientists (A.M. Vogelmann, 2009, personal communication).

TABLE 1. Summary of outcomes. Successes are defined as flights launched on days with desired cloud conditions. Type I errors are decisions to fly only to find no clouds, whereas type II errors are decisions to stand down only to find that the desired conditions existed.

	Heuristic procedure	Automated algorithm
Flights launched	56	66
Successes	28	34
Type I errors	28	32
Type II errors	17	11

of useable data, or 42% of the attainable maximum. By collecting data on 28 days, the heuristic decision process showed a (28 − 19)/19, or approximately 47%, improvement in harvest rate as compared with this zero-skill baseline. The automated decision procedure offered a 79% improvement in harvest rate above that of the random decision maker.

Table 1 offers additional perspective on the relative performance of the heuristic and algorithmic approaches. One difference is prominent: the heuristic process left 11 flights unused at the end of the season, while the automated process used up nearly all of the 67 available flights. It is likely that this difference is driven in substantial part by the 5-day restrictions on flight crews—more exactly, by the relative skill of the two approaches to anticipate the importance of these constraints long before they are actively binding. The algorithmic procedure "anticipates" the possibility of bumping up against these constraints, weeks in advance, and automatically adjusts hurdle probabilities downward according to the probability of scenarios in which these constraints would bind. Human analysts likewise attempt to weigh these same factors in their heuristic calculus, but naturally find it challenging to integrate the large number of possible future scenarios into optimized choices. The likelihood increases of miscalculations leading to substantial surpluses or shortfalls at season's end.

Table 1 offers another interesting observation: in this case example, the two approaches perform about equally well in terms of the yield of successful flights realized as a fraction of the number of flights attempted. This fraction was exactly one-half (28/56) for the heuristic approach, and only slightly higher (34/66) for the algorithmic procedure. The algorithmic approach was much better, however, at avoiding type II errors: over the course of the 160-day field season, the algorithm missed only 11 of the days with conditions suitable for data collection.[10] It appears that the heuristic decision process was too conservative at certain times during the season, eschewing launches on days of only marginal promise only to husband a surplus of flights that could not be used before their expiry.

[10] The terms type I error and type II error are adapted from the nomenclature on statistical hypothesis testing.

6. Conclusions

The algorithmic decision procedure outperformed the heuristic decision process used by the RACORO investigators. This superior performance does not appear to depend on improvements in forecasting technique: as shown in Fig. 3, the GFS model upon which the automated forecasts are based does quite an imperfect job of cloud forecasting. The automated system appears to do a better job, not at forecasting per se, but at delivering decision recommendations.

The result may seem paradoxical: how could inferior forecasts yield superior decisions? The key reasons concern the *consistency* and *replicability* of the automated approach. Because the automated system makes forecasts using exactly the same approach each time, its track record provides statistically sound guidance about its future performance. To calibrate a probabilistic forecasting system, one needs a track record of performance that is statistically sound. By giving up the benefits of human input, one gains the opportunity to create a well-calibrated probabilistic forecasting system that works exactly the same way under all conditions. Human intervention may indeed improve the quality of a single forecast viewed in isolation. But a forecasting process that incorporates human judgement produces signals that are difficult to compare statistically with one another. Critically, a decision maker cannot readily see how to place a current forecast signal into the distribution of all possible future signals that may be received over the remainder of a field season. Yet that is precisely the information the decision maker needs, in order to gauge the benefits of alternative courses of action.

It would appear that the combination of self-organizing maps and dynamic programming shows substantial promise for application to other problems of stochastic optimization. The core requirements include that the problem, including its objectives, be sufficiently well structured that it can be framed in terms of a formal mathematical model; that the decision process be sequential; that the realization of stated objectives depend on the resolution of some stochastic phenomena; that the stochastic process be in some way forecastable; and that historical data records of matched forecasts and observations be available in sufficient duration and granularity to allow for calibration of a probabilistic forecasting system.

Most meteorological aircraft field campaigns satisfy these criteria, and are good candidates for application of the approach. Longer-term measurement programs, with their emphasis on consistent sampling, and their proportionately greater demands on human effort, will likely benefit the most. Yet similar strategies seem likely to be beneficial to most meteorological field experiments where decisions on resource allocations are required. At a minimum, the exercise of framing a field sampling campaign formally as a problem of stochastic optimization seems almost certain to deliver useful insights at both the project planning and operational phases.

Creating an automated decision system calls for development of both a forecasting component and an optimization component. Each component needs to be customized carefully to the particular objectives of the field campaign. For many phenomena, self-organizing maps offer a replicable basis for creating a functional system for generating probabilistic forecasts. But the SOM approach is neither necessary for, nor universally applicable to, all forecasting challenges. To be useful in an automated decision system, the forecasting system should be probabilistic, calibrated, and offer some ability to discriminate between more versus less auspicious conditions for data collection.

Care is likewise required to formulate the optimization problem correctly for a given data collection challenge. Dynamic programming offers a general approach well-suited to sequential decision-making subject to uncertainty and resource constraints. But the exact formulation of an optimization approach must be contoured to the particular objectives of an individual field campaign. For example, a campaign that calls for daily decisions about both whether and *where* to fly will require a substantial refinement to the Cloud Hunter model. Nonetheless, we believe that tools of stochastic optimization show great promise in application to data collection in meteorology.

Acknowledgments. Data were obtained from the Atmospheric Radiation Measurement (ARM) Program sponsored by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division. We would also like to thank Fanglin Yang from NOAA who generated the GFS profiles, Shaocheng Xie who expedited the generation of the CMBE data, and the RACORO investigators Andy Vogelmann and Greg McFarquhar for their openness to discuss their project. We also thank two anonymous reviewers for their insightful comments and useful suggestions. Jason Stefik, Johannes Verlinde, and Nathaniel Johnson acknowledge their support through the Office of Science (BER), U.S. Department of Energy Grant DE-FG02-05ER64058. Small gratefully acknowledges support from the Human and Social Dynamics and Decision Making under Uncertainty Programs of the U.S. National Science Foundation under Grant NSF SES-0729413 and Cooperative Agreement NSF SES-0345840. All errors or omissions in this work are the sole responsibility of the authors.

REFERENCES

Bellman, R., 1957: *Dynamic Programming*. Princeton University Press, 342 pp.

Berg, L. K., and E. I. Kassianov, 2008: Temporal variability of fair-weather cumulus statistics at the ACRF SGP site. *J. Climate,* **21,** 3344–3358.

Berger, J. O., 1993: *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. Springer, xvi + 617 pp.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. Miller, and B. E. Martner, 2000: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites. *J. Appl. Meteor.,* **39,** 645–665.

——, and Coauthors, 2001: The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP). DOE Tech. Memo. ARM VAP-002.1, U.S. Department of Energy, Washington, DC, vii + 49 pp.

Dixit, A., and R. Pindyck, 1994: *Investment under Uncertainty*. Princeton University Press, xiv + 468 pp.

Hewitson, B., and R. Crane, 2002: Self-organizing maps: Applications to synoptic climatology. *Climate Res.,* **22,** 13–26.

Johnson, N., S. Feldstein, and B. Tremblay, 2008: The continuum of Northern Hemisphere teleconnection patterns and a description of the NAO shift with the use of self-organizing maps. *J. Climate,* **21,** 6354–6371.

Kamien, M. I., and N. L. Schwartz, 1991: *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*. 2nd ed. North-Holland, xvii + 377 pp.

Kanamitsu, M., and Coauthors, 1991: Recent changes implemented into the Global Forecast System at NMC. *Wea. Forecasting,* **6,** 425–435.

Kohonen, T., 2001: *Self-Organizing Maps*. 3rd ed. Springer, 501 pp.

Langland, R., 2005: Issues in targeted observing. *Quart. J. Roy. Meteor. Soc.,* **131,** 3409–3429.

Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.,* **55,** 399–414.

Majumdar, S. J., C. H. Bishop, B. J. Etherton, I. Szunyogh, and Z. Toth, 2001: Can an Ensemble Transform Kalman Filter predict the reduction in forecast error variance produced by targeted observations? *Quart. J. Roy. Meteor. Soc.,* **127,** 2803–2820.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.,* **87,** 343–360.

Murphy, A. H., 1997: Forecast verification. *Economic Value of Weather and Climate Forecasts,* R. W. Katz and A. H. Murphy, Eds., Cambridge University Press, 19–74.

Regnier, E., 2008: Doing something about the weather. *Omega-Int. J. Manage. Sci.,* **36,** 22–32, doi:10.1016/j.omega.2005.07.011.

——, and P. A. Harr, 2006: A dynamic decision model applied to hurricane landfall. *Wea. Forecasting,* **21,** 764–780.

Romer, D., 2006: Do firms maximize? Evidence from professional football. *J. Political Econ.,* **114,** 340–365, doi:10.1086/501171.

Stedinger, J., B. Sule, and D. Loucks, 1984: Stochastic dynamic programming models for reservoir operation optimization. *Water Resour. Res.,* **20** (11), 1499–1505.

Stefik, J. B., 2010: A decision algorithm to improve data collection in stochastic environments. M. S. thesis, Department of Meteorology, The Pennsylvania State University, University Park, PA, ix + 36 pp.

Vogelmann, A. M., 2008: RACORO Science and Operations Plan. DOE/SC-ARM-0806, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, 39 pp. [Available online at http://www.arm.gov/publications/programdocs/doe-sc-arm-0806.pdf?id=29.]

Xie, S., and Coauthors, 2010: Clouds and more: ARM climate modeling best estimate data— A new data product for climate studies. *Bull. Amer. Meteor. Soc.,* **91,** 13–20.