A DECISION ALGORITHM TO IMPROVE DATA COLLECTION
IN STOCHASTIC ENVIRONMENTS

A Thesis in
Meteorology
by
Jason Stefik

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2010
The thesis of Jason Stefik was reviewed and approved* by the following:

Johannes Verlinde
Associate Professor of Meteorology
Chair of Meteorology Graduate Program
Thesis Co-Advisor

Arthur A. Small III
Associate Professor of Applied Economics and Finance
Thesis Co-Advisor

George Young
Professor of Meteorology

*Signatures are on file in the Graduate School
Field experiments are conducted to collect observations of a particular meteorological condition or phenomena in order to enhance scientific understanding. In some cases, the decision to deploy costly data-collecting resources must be made ahead of time, when the presence of the sought after condition is unknown, resulting in decisions being made from imperfect forecasts.

Traditionally, forecasters are used to predict the likelihood of good data-collecting conditions existing, while a group of scientists use this information to make resource deployment decisions. A new method for resource deployment decisions is presented that shifts the emphasis from the forecasts to the decision-making aspect of the problem. The performance of this new method is evaluated through using the Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) RACORO field campaign, which sought to obtain aircraft measurement of boundary layer clouds (BLCs).

Dynamic programming is used to quantify the expected number of successful deployments yet to be launched for any combination of days and resources remaining in the field season. For a given forecast of boundary layer clouds existing a decision can be made which maximizes the expected number of successful deployments.

To create BLC forecasts Self-Organizing Maps (SOMs) are used to cluster days according to relative humidity profile. Using cloud data available from the experiment site, the likelihood of BLCs existing for each cluster is determined. A numerical weather prediction model predicts the cluster that the relative humidity profile will belong to, from which the probability that good data-collecting conditions will exist is derived.

If the presented, alternative method had been implemented for RACORO, 34 successful flights would have been launched, as opposed to the 28 successful flights that were actually launched. In addition to this 20% increase in the amount of data collected, the new method can
operate efficiently in real time, significantly saving scientists time spent on decision-making and eliminating forecasting costs.
TABLE OF CONTENTS

List of Figures........................................................................................................ vi
List of Tables........................................................................................................ vii
Acknowledgements............................................................................................. viii
Chapter 1 Introduction....................................................................................... 1
Chapter 2 Methodology....................................................................................... 4
  2.1  Problem Overview....................................................................................... 4
  2.2  Dynamic Programming.............................................................................. 5
  2.3  Statistical Forecasts................................................................................... 11
    2.3.1 Self-Organizing Maps........................................................................... 12
    2.3.2 Cloud Forecasts by Self-Organizing Maps........................................... 13
Chapter 3 Application....................................................................................... 18
  3.1  RACORO Details....................................................................................... 18
  3.2  Other Criteria............................................................................................ 21
  3.3  Results........................................................................................................ 22
Chapter 4 Discussion....................................................................................... 28
  4.1  Sensitivities............................................................................................... 28
  4.2  Lucky?........................................................................................................ 32
  4.3  Longer Lead Time Forecasts...................................................................... 32
Chapter 5 Conclusion....................................................................................... 34
References.......................................................................................................... 36
LIST OF FIGURES

Figure 1  A depiction of the decision lattice for algorithm. With two days and one flight remaining one can either choose to fly (getting success with forecast probability Prob), or wait for the next day for which forecast is unavailable. The value at each node represents the expected number of successful flights yet to be flown. Here, the climatological probability of good conditions existing is 22%. .................... 8

Figure 2  A four-by-six Self-Organizing Map of relative humidity profiles from 1000 to 600mb. The percentage above each representative profile corresponds to the probability that good condition requirements were met for the profiles belonging to that cluster. .................... 15

Figure 3  Empirical probability distribution over cluster realizations, conditioned on a prior forecast of cluster 1. Historically, among days when the weather model predicts an RHP in cluster 1, the realized RHP falls into cluster 1 only 13.3% of the time. .................... 16

Figure 4  The same RH-profile SOM as Figure 2, however the corresponding probabilities associated with each pattern are the probabilities that good conditions will exist given the 30-hour GFS RH-profile forecast is most similar to that cluster. These probabilities are obtained using equation (11). .................... 17

Figure 5  The expected number of flights yet to be launched is shown for each combination of number of days and flights remaining in the RACORO field season when using our decision algorithm. The white line represents the sequence of flight-decisions our method would have issued had it been implemented during RACORO. .................... 23

Figure 6  Results from the field campaign RACORO. There were 45 total days where BLCs existed. During the experiment flights were launched on 28 of these days. Our algorithm would have launched flights on 34 of these days, a 25% increase in successful flights. .................... 26

Figure 7  Displays the mean quantization error of the SOMs as a function of total number of clusters (black line) and the average number of profiles within each cluster as a function of the total number of SOM clusters (red line). These two measurements are used to find the optimal SOM size that minimizes the mean quantization error while maximizes the average number of profiles in each cluster. .................... 30

Figure 8  A sensitivity analysis of our results. The total number of successful flights (as classified solely by our definition of good) is plotted against the total number of flights allocated for different SOM sizes, ranging from 3x3 to 8x8. .................... 31
TABLE OF CONTENTS

List of Tables ................................................................................................................... vii
List of Figures ..................................................................................................................... viii
List of Acronyms .................................................................................................................. ix
Acknowledgements ............................................................................................................... x
1. Introduction ...................................................................................................................... 1
2. Methodology .................................................................................................................... 4
3. Results ............................................................................................................................. 9
4. Conclusion ...................................................................................................................... 10
5. References ...................................................................................................................... 11

List of Tables

Table 1 Shows the relevant dates on which we would have launched a successful flight (Fly | Good), an unsuccessful flight (Fly | Bad), and days when good conditions existed but we did not fly (No-Fly | Good). Days in bold represent days in which flights were actually launched during RACORO, and an asterisk denotes the 7 days we did not classify as good, but nonetheless successful RACORO flights were launched. ............... 27
ACKNOWLEDGEMENTS

There are several people to which I am deeply grateful for, and without whom I would not be where I am now. First and foremost I must thank my family. I am grateful for my family and the love, encouragement, and support they have given me over not only the past 4 years, but throughout my entire life. Also, to all of my friends who have helped me get through these fun, but sometimes stressful years. I owe a special thanks to all those in 412 Walker, my friends from 846, and those friends from Vestal that I have managed to stay in touch with.

I would also like to thank my committee of Johannes Verlinde, Arthur Small, and George Young. Arthur Small has helped me acquire much knowledge outside the scope of meteorology that I was able to use in my thesis and will also use in my career. I am grateful to have had him as a professor for so many classes. This research would not have been possible without him. I am extremely grateful that Hans Verlinde was not only my thesis advisor but my academic advisor as well. He has pushed me and provides constructive criticism that has made me a better student in many aspects. I would also like to thank the help and advice provided to me by Jerry Harrington and Eugene Clothiaux from the group research meetings we have held over the years.

I would also like to acknowledge all of the help that I have received from Chad Bahrmann over the past 2 years. Without all of his technical support I would still be writing code instead of my thesis. Also, Nat Johnson proved to be an invaluable resource in the area of Self-Organizing Maps. Of all the literature I read on SOMs, Nat’s paper was by far the most clear and comprehensible. I was very fortunate to have such a resource sitting across the room from me.

Finally, I owe a special thanks to the Atmospheric Radiation Measurement (ARM) program, as well as the scientists involved in the field campaign RACORO. They provided us
with vital data needed to complete this thesis, as well as necessary information we needed regarding RACORO.

The research was supported by the Office of Biological and Environmental Research of the U.S. Department of Energy grant DE-FG02-05ER64058 as part of the Atmospheric Radiation Measurement Program and by the Human and Social Dynamics and Decision Making Under Uncertainty Programs of the U.S. National Science Foundation under grant award number NSF SES-0729413 and cooperative agreement NSF SES-0345840.
Chapter 1

Introduction

Meteorology is often described as an observational science. Atmospheric observations can be used to support or refute hypotheses that scientists currently hold regarding a particular atmospheric phenomenon. When the specific topic of interest is something as complex as clouds, field campaigns are conducted to sample characteristics of the particular type of clouds under investigations. These field campaigns are costly, as they typically utilize relatively expensive data-collecting resources, such as aircraft or radiosondes, along with the associated support staff. As a consequence, campaigns are allocated only a limited number of days and resources to collect sufficient data to draw conclusions about their hypotheses. Making good resource deployment decisions becomes critically important in collecting a sufficient amount of data with the limited amount of data-collecting resources available to the scientists.

Adding to the difficulty of making resource deployment decisions is that decisions must be made in advance; when the decision whether or not to deploy a resource is made, it is unknown if the sought after conditions will be present. These decisions must therefore rely on uncertain forecasts. Due to the variable nature of clouds, it is often difficult to forecast for a specific cloud type existing at a particular time in the future. Campaign managers typically allocate a significant amount of resources, such as scientist time and money, towards making cloud forecasts.

The U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program has conducted several field campaigns involving aircraft to measure cloud properties. The recent ARM field campaign RACORO was a long-term, 160-day experiment (Vogelmann et al., 2008). The main objective was to obtain aircraft observations of boundary layer clouds.
(BLCs). Flight deployment decisions were made similar to other experiments: a forecasting team predicted the likelihood of good data-collecting conditions existing (the presence of BLCs), and the scientists would make fly/no-fly decisions based on these forecasts.

We present an alternative method for making these fly/no-fly decisions. Flight deployment decisions are a resource-allocation problem; decisions depend on the forecast and the number of days and resources (flights) remaining in the field season. Having better forecasts does not necessarily translate into a more successful field season if poor decisions are made from these forecasts.

In this study we take a dynamic programming (Bellman, 1957) approach to the problem, to optimally allocate flight time across the field season in order to maximize the number of successful flights launched. Dynamic programming makes it possible to calculate the expected number of successful flights still to be launched for any given combination of days and flights remaining in the experiment. For a given forecast the fly/no-fly decision will be made that maximizes the expected total of successful flights.

We create our own statistical forecasting method to determine the probability that good data-collecting conditions will exist the next day. Self-organizing maps are used to cluster the atmosphere into different relative humidity (RH) profile states above the experiment site. Using historical cloud fraction data, it is determined how likely sufficient amounts of boundary layer clouds will exist for each atmospheric state. A numerical weather prediction model is then used to predict the atmospheric state that will exist, from which the probability of good data-collecting conditions existing is derived.

The effectiveness of our method is compared to the current method of flight decision making by using the ARM field campaign RACORO as an example. When combining our forecasts with dynamic programming, we determine the days that we would have decided to launch flights. We compute the number of successful flights our algorithm would have launched,
and compare this to the number of successful flights actually launched during the RACORO field season.
Chapter 2

Methodology

Our objective is to design an algorithm that makes optimizing flight deployment decisions for probabilistic forecasts of good conditions existing. In order to make optimizing decisions we conduct a cost-benefit analysis of launching a flight. The benefit of launching a flight is the potential for that flight to observe the sought-after conditions. The cost of a flight is the opportunity cost of that flight (not being able to use that flight on some later date in the field season). We use dynamic programming to quantify the cost and benefit of launching a flight, and a statistical forecasting method using self-organizing maps to predict the likelihood of good data collecting conditions.

2.1 Problem Overview

Meteorological field experiments involving aircraft typically are similar in structure. The field season has a set length a fixed number of flights to collect data. The main objective is for the aircraft to observe the presence of a specific meteorological condition or feature. Experiments involving aircraft pose unique challenges because flights must be scheduled at least one day-ahead of time. Therefore, when the decision to fly \((a=1, \text{ or } a_1)\) or not fly \((a=0, \text{ or } a_0)\) is made, the presence \((x=1, \text{ or } x_1)\) or absence \((x=0, \text{ or } x_0)\) of the sought-after meteorological phenomena is unknown.

The standard practice for making flight decisions in field experiments rely on forecasters to provide the probability that good conditions will exist on the following day \(\text{prob}(x_1)\). The forecasters rarely provide a quantified probability: rather they provide a more qualitative
prediction of the future conditions to the decision-makers. The decision-makers use the forecast information to decide if the likelihood of good conditions existing is high enough to launch a flight. In addition to the forecast, the scientists must also take into account the number of flights $F$ and days $D$ remaining in the field season when making their decision. This method results in decisions being made primarily from the scientists’ intuitive judgments. Despite a decision process that is forced to rely heavily on the scientists own discretion, the majority of the focus is typically placed on improving the forecasts. In some experiments multiple teams of forecasters are utilized to predict the likelihood of $x_t$.

Our methodology shifts the focus from making accurate forecasts to making better decisions, with the ultimate goal of improving the success of a field season through using a statistically robust method to make better decisions from less accurate forecasts. For the purpose of this study, success will be defined as the total amount of successful flights (number of days in which a flight is launched and good conditions exist).

2.2 Dynamic Programming

Dynamic programming is a technique that solves complex problems by breaking them into simpler, solvable sub-problems (Dasgupta et al., 2006). Dynamic programming is a statistically robust method that can solve complex stochastic optimization problems, and is utilized across diverse sets of fields, from the managing of pension funds (Haberman and Sung, 1994) to determining whether or not to punt on fourth down in football (Romer, 2002). The complex problem that faces a decision-maker in a field experiment is how to efficiently allocate a fixed $F$ number of flights across $D$ days. Dynamic programming allows the decisions maker to efficiently allocate resources across time under quantified uncertainty.
The cost and benefit of flying is quantified through dynamic programming by first assigning a value to each combination of \( D \) days and \( F \) flights remaining in the field season. This value is expressed in terms of the expected number of successful flights \( V \) still to be launched, conditioned on following an optimal path. That is, the expected number of successful flights still to be launched with \( D \) days and \( F \) flights remaining is conditioned on making the optimal decision on all future days.

For the purpose of this study, a node will be defined as any combination of \( D \) days and \( F \) flights remaining in the field season. The value of any node \( V(D, F) \) depends on the value of the node for the following day \( V(D-1, F - a) \), where \( a = 1 \) if a flight is launched, and \( a = 0 \) if no flight is launched. Therefore, backward induction must be used to solve for these values. Backward induction is possible because the value of the final day of the experiment can be determined from the specific constraints placed on the problem.

One such constraint is that there must be no flights remaining after the last day of the experiment. This constraint exists because all flight hours are purchased before the experiment begins. Any unused flights at the end of the experiment have no value and are wasted. Also, we assume that the duration of the field season and total number of flights cannot be altered. Therefore, if the optimal decision is made on each day, for \( D = 0 \) there must be no flights remaining. Therefore, on the last day \( (D = 1) \), there must either be zero or one flight remaining. If there are zero flights remaining, then \( V(1,0) = 0 \) because the expected number of successful flights with no more flights remaining must be zero. If there is one flight left with one day remaining, then a flight will be launched regardless of the forecast. Because a flight must be launched, \( V(1,1) \) is equal to the climatological probability of good conditions existing.

The values of the remaining nodes may now be determined by backward induction. If there are two flights remaining on the second-to-last day of the experiment, then a flight will be
launched on both days, and the expected value of $V(2,2)$ is the climatological probability multiplied by two.

A flight deployment decision must be made if there is only one flight and two days remaining. If the last flight is launched with two days remaining, the expected value of successful flights is equal to the forecast probability of good conditions existing when the flight is launched on the penultimate day ($D=2$). It is assumed that forecasts are only available for the next day (in this example, a forecast is only available for $D=2$). This assumption makes the value of not flying $V(1,1)$, which is the climatological probability of $x_1$. The optimizing decision that maximizes the expected number of successful flights at node $V(2,1)$ will be to fly if the forecast probability is greater than climatology. This scenario is depicted in Figure 1.

More generally, the value of flying is equal to the forecast probability of good conditions existing plus the value of being one day into the future with one less flight. The value of not launching a flight is being one day into the future with the same number of flights. Thus, the general decision rule is to choose the maximum of the two quantities $V(D-1, F-1) + prob(x_1|f)$ and $V(D-1,F)$.

Through algebraic manipulation, the decision rule can be expressed more conveniently. A flight will be launched if and only if

$$prob(x_1|f) > V(D-1,F) - V(D-1,F-1),$$

where $prob(x_1|f)$ is the probability of good conditions existing for a forecast signal $f$. The term $V(D-1,F) - V(D-1,F-1)$ is referred to as the hurdle probability ($P_{HD}$), which is the minimum forecast probability necessary to launch a flight optimally.
Figure 1 A depiction of the decision lattice for algorithm. With two days and one flight remaining one can either choose to fly (getting success with forecast probability $Prob$), or wait for the next day for which forecast is unavailable. The value at each node represents the expected number of successful flights yet to be flown. Here, we assume the climatological probability of good conditions existing (Climo) is 22%.
Through backward-propagation, this same method is used to determine the value for all nodes (D,F). However, we must use a slight modification in determining the value of future days because the forecast \( f \) is unknown. Because the forecast is a random variable, the decision that will be made on any future day is also unknown. To account for the uncertainty in the eventual action taken, the value of flying is multiplied by the probability of launching a flight \( (a_1) \), and the value of not flying by the probability of not launching a flight \( (a_0) \), resulting in the following equation

\[
V(D,F) = \text{prob}(a_1) \cdot [V(D-1,F-1) + \text{prob}(x_1|f)] + \text{prob}(a_0) \cdot V(D-1,F)
\]  

(2)

When \( f \) is unknown, the distribution from which the forecast signal is drawn must be known to solve for (2). That is, all possible forecast probabilities \( \text{prob}(x_1|f) \) must be known, along with the likelihood of each of the forecasts being issued, \( \text{prob}(f) \). If this information is available, then all terms in equation (2) can be solved without knowing the forecast. The probability of launching a flight \( \text{prob}(a_1) \) is determined by finding all forecast signals that would result in optimally launching a flight, and summing over the probability of these forecast signals being issued.

Assume there are \( K \) forecast signals that could possibly be issued. For each forecast \( f_i \), there is an associated probability that good conditions will exist \( \text{prob}(x_1|f_i) \). Also, each forecast signal will be issued with \( \text{prob}(f_i) \). To simplify, the forecasts are sorted from the most favorable for good conditions to exist, to the least favorable, such that

\[
\text{prob}(x_1|f_1^*) > \text{prob}(x_1|f_2^*) > \ldots > \text{prob}(x_1|f_K^*)
\]  

(3)

where \( K \) is the total number of forecast signals that could possibly be issued. We define the critical forecast \( f_K^* \) as the last forecast signal for which launching a flight remains the optimal decision, such that
\[ prob(x_1|f'_c) > P_H > prob(x_1|f'_{c+1}) \]  

(4)

where \( P_H \) is the hurdle probability defined by equation (1). If the probability of each forecast being issued can be determined (for example, by having an extensive track record of the forecasting system), then the probability of each individual forecast signal being issued can be determined. The probability of launching a flight on any given node is

\[ prob(a_1) = \sum_{i=1}^{c} prob(f_i) \]  

(5)

Before we can evaluate (2) for a random variable \( f_i \), the term \( prob(x_1|f_i) \) must also be known. The probability of good conditions existing is only relevant if a flight will be launched, and a flight will only be launched if the forecast exceeds the hurdle probability. We therefore must determine the probability of good conditions existing conditioned on the forecast exceeding the hurdle probability. We can now express \( prob(x_1|f_i) \) as the conditional probability that good conditions will exist given that the forecast probability exceeds the hurdle probability (ie a flight is launched)

\[ prob(x_1|f_i > P_H) = \frac{1}{prob(a_1)} \sum_{i=1}^{c} prob(f'_i) \cdot prob(x_1|f'_i) \]  

(6)

Using equations (5) and (6), equation (2) can be solved for all nodes, even when the future forecast is unknown, provided that we quantify the uncertainty in the forecast. Intuitively, the expected value of successful flights \( V \) is a function of the number of days \( D \) and flights \( F \) remaining. \( V \) is also dependent on how often good conditions present themselves (climatological probability of good conditions existing) and of the sharpness of the probabilistic forecasts.
2.3 Statistical Forecasts

Dynamic programming can be utilized for stochastic processes if the uncertainty can be quantified. The need for quantified uncertainty makes having probabilistic forecasts a necessity. It is also necessary to know the likelihood of each forecast being issued. While there are numerous different methods to create cloud forecasts, these values must be quantifiable.

There are many weather forecasting models that produce forecasts of cloudiness: unfortunately these products tend to be inaccurate, and typically concentrate on predicting cloud cover. The heights and depths of the clouds are often not predicted to high resolution, which is a problem because the definition of what constitutes a good day is rather specific. A more important concern is our need that the uncertainty of the forecasts, which generally is lacking for cloud products. For these reasons, we create our own statistical forecasting system to predict the occurrence of good data-collecting conditions.

To build the forecasting system, we obtain cloud data collected above the DOE – ARM Climate Research Facility (ACRF) at the Southern Great Plains site, the location of the RACORO experiment. The Atmospheric Radiation Measurement (ARM) program has long-term high-resolution cloud data extending back to 1998. Hourly cloud fractions are available from ARM’s Climate Modeling Best Estimate (CMBE; Xie et al., 2010) product. Using these cloud fractions we classify each day as either having good($x_1$) or bad($x_0$) data-collecting conditions. An hour is classified as good if it simultaneously meets the following conditions:

1. Cloud fraction exceeding 10%
2. Cloud top below 3km
3. Cloud base above 400m
4. No precipitation (< 0.01 inches/hour)

If there are more than four good hours in a day between 13 UTC and 23 UTC then the entire day is classified as good. Otherwise, the day is classified as “bad”.

A useful forecasting system must use the atmospheric variables that influence the presence or absence of boundary layer clouds. After attempting several different weather variables, we found that the relative humidity (RH) profile serves as the best indicator for boundary layer clouds. To distinguish between different RH profiles self-organizing maps (SOMs) are used.

2.3.1 Self-Organizing Maps

The method of SOMs incorporates an unsupervised neural network algorithm to produce a discrete set of relative humidity profiles that represent the continuous distribution of profiles found in the data set. SOMs are a type of cluster analysis in which similar profiles are grouped together in the same representative pattern.

SOMs are typically created using an iterative training approach. They are trained by determining and updating the representative pattern that closest resembles a particular profile from the data set. This is done through minimizing the Euclidean distance between the best matching pattern \( m^*_c \) and a data vector \( z \) by:

\[
|m^*_c - z| = \min \{|z - m^*_c|\}
\]

for \( i = \{1,...,K\} \), where \( K \) is the total number of SOM patterns. An important distinction between SOMs and most traditional clustering techniques is that SOMs use a neighborhood function (Hewitson and Crane, 2002). The neighborhood function allows patterns close to \( m^*_c \) to also be updated by \( z \), which creates a smoothing effect between patterns on the map. Also, the neighborhood function creates a spatially ordered map within a two-dimensional grid, where similar patterns tend to be located near each other.
The SOM algorithm also tends to minimize the difference between the representative patterns to each sample within the data set (Johnson et al., 2008). That is, the mean quantization error (MQE) will tend to be minimized as described by

\[ MQE = \frac{1}{N} \sum_{t=1}^{N} |z_t - m^*_t| \] (8)

where \( N \) is the total number of samples in the data set. Thus, the SOM algorithm tends to maximize the similarity between the samples within the data set to the representative patterns (Kohonen 2001). Therefore, we can be confident that our SOM patterns are physically representative of the actual profiles within the data set.

**2.3.2 Cloud Forecasts using Self-Organizing Maps**

Self-organizing maps are used to cluster the relative humidity profiles above the experiment site. RH profile data are obtained from the North American Regional Reanalysis (NARR) dataset from 1979-2008. The grid point used is located at (36.57° N, 97.56° W), whereas the actual coordinates of the ACRF facility is (36.69° N, 97.56° W), less than 7 km away from the grid point used to create the SOM. The SOM is created from only the 18 UTC profiles because the majority of flights were flown between 13 UTC and 23 UTC.

The RH profile SOM is used to predict the occurrence of boundary layer clouds. RH profiles extend from the surface to 600mb. Profiles extending above the depth of the boundary layers are used to distinguish ideal BLCs from deep, convective clouds and clouds that are more likely to precipitate. Figure 2 displays the 4-by-6 RH profile SOM, along with the corresponding probability that boundary layer clouds existed for days belonging to each cluster. Additional
information into the choice of SOM dimension and its effect on the overall results can be found in the discussion section.

Each 18 UTC profile is matched to the most similar representative profile through the minimization of Euclidean distance. The probability of a good day $x_i$ existing for a particular pattern $i$ is

$$prob(x_i) = \frac{n_{ii}}{n_{ti}}$$ (9)

where $n_{ii}$ is the number of profiles within SOM pattern $i$ that corresponds to a good day, and $n_{ti}$ is the total number of profiles belonging to SOM pattern $i$.

To create a probabilistic forecast, we use a numerical weather prediction model to forecast the RH profile (Global Forecasting System: Kanamitsu et al., 1991). This RH profile can be classified as a particular SOM pattern by determining the cluster, designated as $f_i$, that the forecasted profile is most similar to. Using historical model output, we find the distribution of realized patterns for each forecast pattern. The historical model performance allows us to determine the pattern that actually existed $r$ when a particular forecast pattern $f$ was issued. Thus we can find the conditional probability distribution of

$$Prob(r_j|f_i)$$ (10)

for $i = \{1, \ldots, K\}$ and $j = \{1, \ldots, K\}$, where $i = j$ represents a correct forecast and $i \neq j$ represents an incorrect forecast. Figure 3 displays a forecast verification for when pattern 1 was the forecast. Note that the most probable pattern was not 1, but pattern 2, and that there are several other patterns with significantly high probabilities. However, whether the right pattern is realized is of lesser importance than whether good conditions exist given the forecasted pattern.
When combining the probability distributions of (9) and (10), the probability that good conditions will exist for a particular GFS forecast $f_i$ is given by

$$prob(x_1|f_i) = \sum_{j=1}^{24} prob(x_1|\tau_j) \cdot prob(\tau_j|f_i)$$

(11)

Figure 4 displays the same SOM map as in Figure 2 but now with the conditional forecast probabilities found by (11). These new probabilities are derived from (11), and represent the probability that good conditions exist given that the GFS forecast belongs to that particular cluster.

**Figure 2** A four-by-six Self-Organizing Map of relative humidity profiles from 1000 to 600mb. The percentage above each representative profile corresponds to the probability that good condition requirements were met for the profiles belonging to that cluster.
Figure 3  Empirical probability distribution over cluster realizations, conditioned on a prior forecast of cluster 1. Historically, among days when the weather model predicts an RHP in cluster 1, the realized RHP falls into cluster 1 only 13.3% of the time.

<table>
<thead>
<tr>
<th></th>
<th>56%</th>
<th>69%</th>
<th>32%</th>
<th>12%</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.3%</td>
<td>20.6%</td>
<td>4.4%</td>
<td>3.7%</td>
<td></td>
</tr>
<tr>
<td>37%</td>
<td>23%</td>
<td>26%</td>
<td>54%</td>
<td></td>
</tr>
<tr>
<td>1.5%</td>
<td>8.1%</td>
<td>4.4%</td>
<td>4.4%</td>
<td></td>
</tr>
<tr>
<td>4%</td>
<td>4%</td>
<td>33%</td>
<td>53%</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>3.7%</td>
<td>4.4%</td>
<td>11.8%</td>
<td></td>
</tr>
<tr>
<td>9%</td>
<td>24%</td>
<td>48%</td>
<td>36%</td>
<td></td>
</tr>
<tr>
<td>2.9%</td>
<td>3.7%</td>
<td>7.4%</td>
<td>6.6%</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>1%</td>
<td>11%</td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>0%</td>
<td>0.7%</td>
<td>1.5%</td>
<td></td>
</tr>
<tr>
<td>1%</td>
<td>0%</td>
<td>1%</td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4  The same RH-profile SOM as Figure 2, however the corresponding probabilities associated with each pattern are the probabilities that good conditions will exist given the 30-hour GFS RH-profile forecast is most similar to that cluster. These probabilities are obtained using equation (11).
Chapter 3

Application

We implement this methodology to make fly or no-fly decisions for the field experiment RACORO. Using our methodology, we determine the days on which we would have launched flights during RACORO. We use ARM’s Climate Modeling Best Estimate cloud data to determine the days when good conditions existed and so would have yielded a successful flight. The total number of days in which we would have launched a flight and for which good conditions existed is the number of successful flights we would have had. To evaluate the performance of our algorithm, we compare the number of successful flights we would have launched to the number of successful flights actually launched during RACORO.

3.1 RACORO Details

RACORO was an experiment that took place over ARM’s Southern Great Plains (SGP) facility from January 22, 2009 to June 30, 2009. Three-hundred total flight hours were allocated for use across the 160-day experiment. The 300 flight hours translates to approximately 67 flights that could have been launched to obtain observations of BLCs. The previously discussed methodology is, for the most part, applied exactly as described to the RACORO field season. However, there were some restrictions that apply specifically to RACORO that forced us to modify our methodology.

One such restriction was that during the RACORO experiment, flights could only be launched on a maximum of 5 out of every 7 days. This limitation was due to flight crew hour
restrictions mandated by the Federal Aviation Administration (FAA). This restriction required a modification to our original dynamic programming setup.

The actions taken on the previous 7 days influence the decision for the present day because of the 5 of 7 day restriction. For example, if flights were launched on the previous 4 days, the cost of launching a flight is not only not having that flight to use on a future date, but additionally not having the option to launch flights for the next 2 days. Therefore, we must consider all the actions taken at $t$, where $t$ in the number of days into the past, from $t = 1$ to $t = 7$.

Each node value is now dependent upon the actions taken on the previous 7 days. Originally $V(1,1)$ was assigned a value of climatology. However, if flights were launched on 5 of the previous 6 days then $V(1,1) = 0$ because a flight cannot be launched. Likewise, the value of $V(3,1)$ if flights were launched on the previous 5 days, because flights cannot be launched for the next 2 days. In this case $V(3,1)$ has the same value as $V(1,1)$.

We solve this problem by assigning each node $V(D,F)$ a separate value based upon the decisions made on the previous 7 days. There are therefore 120 different possible paths that could be taken to any node, represented by

$$
\sum_{k=0}^{5} \binom{7}{k}
$$

(12)

The value at each node is now defined by the number of days left, number of flights left, and the action taken on the previous 7 days, or $V_{D,F}^{a_1,a_2,a_3,a_4,a_5,a_6,a_7}$. As a consequence, each node has 120 different values corresponding to each of the 120 paths that could possibly be taken in arriving at that particular node. The hurdle probability that will determine whether or not the forecast is high enough to optimally launch a flight is found by

$$
P_h = V_{D=1,F=1}^{0,a_2,a_3,a_4,a_5,a_6,a_7} - V_{D=1,F=1}^{1,a_2,a_3,a_4,a_5,a_6,a_7},
$$

(13)
where 0 represents a no-fly decisions and 1 represents a fly decision. Equation (2) can be evaluated knowing all 120 potential hurdle probabilities at node \( V(D,F) \), because the terms \( \text{prob}(a_1), \text{prob}(a_0), \) and \( \text{prob}(x_i|f) \) are solved for each hurdle probability. After the expected number of successful flights still to be launched is calculated for all 120 potential paths corresponding to each node \( V(D,F) \), flight decisions are made for RACORO by comparing the forecast probability of good conditions to the hurdle probability found by (13).

During the RACORO experiment the decision team had the option to cancel a flight up to 3 hours prior to its scheduled takeoff. Such cancellations would be a burden to the pilot and flight crew, causing a real, yet difficult to quantify, cost to cancelling a flight. We give ourselves this same opportunity for day-of cancellations. The 12 UTC 6-hour GFS model forecast is used for the day-of forecast to determine whether to cancel a scheduled flight. The same methodology for the day-ahead decisions is implemented for day-of cancellations; however, a separate forecast verification is implemented where we use the history of the 6-hour GFS forecast instead of the 30-hour GFS forecast.

The flight cancellation option is only exercised on days when a flight is scheduled. If the 6-hour probabilistic forecast remains above the hurdle probability the flight will still be launched. If the 6-hour forecast probability drops below the hurdle probability then the flight is cancelled.

However, we use a slight modification to (13) when determining the hurdle probability for the day-of forecast. Because a flight has already been scheduled, we give credit to the flight crew and pilots for working that day, whether or not a flight is actually launched. This was done to take account of the real, yet difficult to quantify, cost to cancelling a flight on pilot and ground crew effort. Therefore, with our method, not only will flight crews and pilots be required to work a maximum of 5 out of every 7 days, but they will only be required to be on stand-by a maximum of 5 out of every 7 days.
During RACORO we would have cancelled a flight on 7 occasions. While we do not know how many times flights were actually cancelled during RACORO, 7 cancelled flights is a relatively small number for a 160 day experiment. This speaks well for the consistency of the GFS between the 6-hour and 30-hour lead time forecasts.

3.2 Other Criteria

Our methodology thus far has used an oversimplified definition of a good day. Our forecasting system is essentially predicting the likelihood that a sufficient amount of non-precipitating BLCs exist over the experiment site. The extent, depth, and height of clouds and precipitation is tightly linked to the RH profile. However, there are other scenarios besides the absence of BLCs that would constitute a bad day.

One such scenario is severe weather in the area. Severe weather within the vicinity of the experiment site will cause a flight cancellation. It is unlikely that good conditions and severe weather would simultaneously present themselves in the same area. One requirement of good conditions is that cloud top height must be below 3km, and because thunderstorm cloud tops extend well above 3km, it is unlikely that both good conditions and severe thunderstorms would coexist in the same area. Therefore, we ignore severe weather as criteria when classifying the days as having good ($x_1$) or bad ($x_0$) conditions.

Another scenario that would result in a bad day is data-collecting instrumentation failure. Despite good conditions existing, data cannot be collected if the instruments are not functioning properly. These instrumentation failures are random occurrence and are thus unpredictable, therefore impossible to account for in our decision algorithm. Additionally, it is unknown on which days instrument failure occurred during RACORO. Although it is unknown exactly how
many flights were cancelled due to severe weather or instrumentation failure, these are relatively infrequent occurrences, and thus have little effect on our overall results.

The RACORO experiment goal was to observe not just any BLCs, but more specifically liquid BLCs. The temperature of the cloud becomes critically important because the entire cloud must be comprised of liquid water. If there is any ice present in the cloud, then that day is classified as a bad day, even if all other cloud specifications are met. Clouds with ice present occurred with enough frequency (particularly during the beginning of the experiment), that it cannot be ignored.

While the presence of ice in clouds cannot be ignored, it is also independent of our forecasting system that predicts the probability of a sufficient amount of BLCs existing. The GFS model is used in forecasting the cloud top temperature, but because our cloud forecasting system does not provide information on the height of the cloud (just that cloud top will be located below 3 km) we use the GFS to forecast the temperature at 2.5 km. If the 30-hour GFS predicts the 2.5km temperature to be below freezing, then no flight is scheduled. If a flight is scheduled and the 6-hour GFS forecast predicts a 2.5 km temperature below freezing then the flight is cancelled.

3.3 Results

Our algorithm is constructed using only data up until the RACORO experiment began. That is, the Climate Modeling Best Estimate product cloud data, North American Regional Reanalysis relative humidity profiles, and GFS relative humidity profile forecasts used were all from before the RACORO experiment (pre-2009). Thus, the performance of the algorithm is not evaluated with the same data used in building the algorithm.

Dynamic programming is then applied specifically to RACORO (160 total days, 67 total flights) using our statistical forecasts. The expected number of successful flights still to be
launched from any possible combination of D and F during RACORO is displayed in Figure 5, along with the path (white line) we would have taken through the season had we made the flight decisions.

**Figure 5** The expected number of flights yet to be launched is shown for each combination of number of days and flights remaining in the RACORO field season when using our decision algorithm. The white line represents the sequence of flight-decisions our method would have issued had it been implemented during RACORO.
To determine the flight decisions our algorithm would have made, we use archived GFS 30-hour and 6-hour temperature and relative humidity profile forecasts from RACORO (1/22/09 – 6/30/09). Each RH-profile forecast during RACORO has an associated probability that good conditions will exist, as determined by our statistical forecasting system. A flight will be scheduled/launched if the forecast probability exceeds the hurdle probability corresponding to that particular day \((D,F)\), and if the forecasted temperature at 2.5 km is above freezing. Although the temperature forecast is not taken into account when designing the decisions algorithm, it is implemented with a simple binary approach when applied to RACORO. Using the temperature and RH-profile GFS forecasts for the RACORO time period we reconstruct the sequence of actions we would have made during the experiment, thus providing the days on which we would have launched flights.

Once the sequence of actions we would have made is known, the number of successful flights we would have had can be calculated by classifying each day during RACORO as good or bad. Again, using the CMBE cloud dataset (although this time from 1/09 to 6/09), each hour during RACORO is classified as good or bad using the same definition of good from our statistical forecasts (1#-4). In addition to fulfilling requirements 1-4, the cloud top temperature must be above freezing for the hour to be classified as good. Again, if more than 4 good hours existed between 13 UTC and 23 UTC, then the entire day is classified as good.

The success of the experiment is judged solely on the number of observations collected. We assume that an observation is collected on days in which both a flight is launched and good conditions exist. During the 160 days of the experiment, 38 of the days were classified as having good conditions. The 38 days we classified as good were confirmed by personal communication with the scientists involved in the RACORO experiment (Vogalman and McFarquhar, 2009). The RACORO team launched flights on 21 of these 38 good days. All of these 21 flights were
verified to be successful in obtaining useful cloud data from flight logs kept throughout the experiment. Our algorithm would have launched flights on 28 of these 38 good days.

However, the flight logs indicate that there are 7 additional days on which the RACORO flight team flew and collected useful cloud data that we failed to classify as a good day. On 5 of these 7 days the aircraft observed boundary layer clouds away from the ARM experiment site, while the conditions over the experiment site were not classified as good. On the other 2 days RACORO flights sampled good conditions above the site, but there was not at least 4 good hours between 13 UTC and 23 UTC.

These 7 days pose a challenge in classification due to their ambiguity. Using our definition of what constitutes a good day these 7 days should be classified as bad, but during RACORO flights were launched, and cloud data was obtained on these days. Although our classification methodology classify these 7 days as being bad, we change their classification to good. Ultimately, the classification of these days has little effect when comparing our algorithm’s success to the actual success of RACORO. When including these additional days the total number of good days increases from 38 to 45, the number of successful flights actually flown during RACORO increases from 21 to 28, and the number of successful flights our algorithm would have produced increases from 28 to 34. Our method increases the amount of observations collected by 20% over the amount actually collected during RACORO. Our method would have flown on 34 of the 45 total good days compared to the 28 observations actually collected during RACORO. These results are displayed in Figure 4. Table 1 presents an in-depth analysis of Figure 6, showing the specific days that flights were launched or good conditions existed.
Figure 6  Results from the field campaign RACORO. There were 45 total days where BLCs existed. During the experiment flights were launched on 28 of these days. Our algorithm would have launched flights on 34 of these days, a 25% increase in successful flights.
<table>
<thead>
<tr>
<th>Our decision</th>
<th>Actual State</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>6/19*, 6/20, 6/21, 6/23*, 6/26*</td>
</tr>
<tr>
<td>Fly</td>
<td>Bad</td>
<td>1/23, 1/26, 2/24, 3/7, 3/14, 3/15, 3/21, 3/30, 4/1, 4/4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4/11, 4/12, 4/14, 4/16, 4/17, 4/28, 5/1, 5/5, 5/10, 5/11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/12, 5/20, 5/21, 5/25, 6/1, 6/2, 6/4, 6/6, 6/27, 6/28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6/29, 6/30</td>
</tr>
<tr>
<td>No-Fly</td>
<td>Good</td>
<td>2/11, 2/17, 3/19, 4/18, 4/27, 4/30, 5/7, 5/14, 5/24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5/26, 6/24*</td>
</tr>
</tbody>
</table>

Table 1 Shows the relevant dates on which we would have launched a successful flight (Fly | Good), an unsuccessful flight (Fly | Bad), and days when good conditions existed but we did not fly (No-Fly | Good). Days in bold represent days in which flights were actually launched during RACORO, and an asterisk denotes the 7 days we did not classify as good, but nonetheless successful RACORO flights were launched.
Chapter 4

Discussion

The primary result from this study reveals that using our flight deployment method, 34 successful flights would have been launched, compared to the 28 successful flights actually launched during RACORO. Intuitively, the number of successes from our decision algorithm is dependent on our methodology. Thus, changing particular parameters to make our forecasts (particularly in the SOMs), will change our overall results. We therefore demonstrate that our results are not sensitive to these details. We also explain our choice of SOM size and provide insight into why our number of successful flights (28, not including the 6 that were added later) is greater than the 22.5 expected number of successful flights calculated through dynamic programming ($V(160,67) = 22.5$).

4.1 Sensitivities

While not highly sensitive, our results (the number of successful flights) will change as we alter our statistical forecasting method. It is possible to tweak the parameters in the self-organizing maps to produce slightly different forecasts. These slightly different forecasts could potentially translate into different decisions, which will ultimately impact the total number of successful flights launched. We construct several SOMs while altering parameters in the learning rate and neighborhood functions, and find that these alternations have a minimal impact on what the self-organizing map and the corresponding probabilities look like. However, altering the dimension of the SOM, can have a large impact on the forecasts.
The particular SOM dimensions we use are 4-by-6. These dimensions are chosen semi-arbitrarily, and are not used to maximize our number of successful flights. The dimensions are less important than the total number of patterns (in our case, 24). When choosing the total number of patterns, there is a balance between minimizing the mean quantization error (MQE) while also having robust statistics. As the number of patterns increases the MQE decreases, meaning all the profiles belonging to the same cluster will closely resemble each other. The tradeoff is that with a limited amount of data the number of cases in each cluster will decrease, thus reducing the statistical significance of the corresponding probabilities that good conditions will exist. The optimal size would be the point at which the benefit gained from the marginal decrease in MQE due to increasing the number of SOM patterns no longer outweighs the marginal decrease in the number of profiles per cluster. When choosing our SOM size, we deemed the MQE of maps with less than 12 patterns to be too great, and maps with over 40 clusters to have too few profiles in each cluster. The choice of 24 patterns is semi-arbitrary, but is chosen because it falls between 12 and 40.

Ultimately, our results are not highly dependent on the SOM dimensions chosen. To test how sensitive our results are to the SOM size, we repeated the experiment for SOM sizes ranging from 9 to 64 patterns. We calculated how many successful flights would have been obtained as a function of total flights allocated for SOM sizes ranging from 3-by-3 to 8-by-8, which can be seen in Figure 8.

As anticipated, different SOM sizes produce different results. However, these differences are minimal and appear to be random in nature. That is, there is no evidence to support that the algorithm performance systematically improves or worsens as SOM size increases.
Figure 7 Displays the mean quantization error of the SOMs as a function of total number of clusters (black line) and the average number of profiles within each cluster as a function of the total number of SOM clusters (red line). These two measurements are used to find the optimal SOM size that minimizes the mean quantization error while maximizes the average number of profiles in each cluster.
Figure 8  A sensitivity analysis of our results. The total number of successful flights (as classified solely by our definition of good) is plotted against the total number of flights allocated for different SOM sizes, ranging from 3x3 to 8x8.
4.2 Lucky?

Our methodology yields the expected number of successful flights that will be launched during the experiment. The value of the first node of the season ($V(160, 67)$ in Figure 5) is the expected number of successful flights that will be launched during the entire field season. This value is very useful to the scientists in that it reflects the expected number of observations that will be obtained during the experiment.

For RACORO, the expected number of successful flights using our statistical forecasting method is 23. The actual number of successful flights we obtained, excluding the 6 additional days, was 28. One reason the actual number exceeded the expected number is because the actual number of good days exceeded the climatological expectation. Based on climatology there would have been 34 good days, but there were actually 38 days we classify as good. In addition, the expected number of successful flights is calculated assuming the day-ahead decision is executed, and does not take into account for the option of day-of-flight cancellations. The ability to cancel flights along with an above average occurrence of boundary layer clouds yielded more observations collected than expected.

4.3 Longer Lead Time Forecasts

In this thesis we only use the day-of and day-ahead forecasts, and ignore all forecasts with longer lead times. It may reasonably be asked if we can improve our performance using these readily available long lead time forecasts.

One reason we ignore long-range forecasts is because the sharpness of the GFS forecasts decreases with lead time, so the two-day forecast will have more uncertainty than the one-day forecast, and little more information is gained. It is possible that should one use a higher skill
model that some advantage may be gained: however, it purpose of this study was to show the utility of the decision algorithm even when using a less accurate forecast.

Furthermore, the decision made from the one-day forecast is likely not to change, even if the GFS could issue a perfect two-day forecast, because the decision whether or not to fly tomorrow is based solely on the day-ahead forecast and the hurdle probability. Except for the end of the field season, knowing the action taken in two days will have a negligible impact on the hurdle probability. Hypothetically, if it was known with absolute certainty a flight would be flown two days \((D-2)\) from now, then hurdle probability for today’s decision would become \(V_{D-2,F-1} = V_{D-2,F-2}\) instead of \(V_{D-1,F} = V_{D-1,F-1}\). The reason two-day forecasts do not provide much value is that for much of the experiment

\[
V_{D-2,F-1} - V_{D-2,F-2} \approx V_{D-1,F} - V_{D-1,F-1} \tag{14}
\]

This approximation will only fail for small values of \(D\) and \(F\). For instance, at \(D = 100\) and \(F = 37\), the above approximation is \(0.2461 \approx 0.2436\). The approximation still holds well for relatively small values. Even for \(D = 10\) and \(F = 5\), this approximation becomes \(0.2465 \approx 0.2281\).
Chapter 5

Conclusion

Obtaining atmospheric observations is crucial in enhancing the scientific understanding of atmospheric phenomena. Thus, field experiments are often conducted to collect observations of a particular meteorological condition or phenomena. In many cases, the decision to deploy data-collecting resources must be made on a daily basis with significant lead times. These decisions are therefore made using the aid of imperfect forecasts. Relatively large amounts of time and money are spent on the forecasts, while less attention is devoted to flight deployment decisions made from the forecasts. We present an alternative method to making flight decisions, which shifts the focus from having better forecasts to making better decisions.

The ARM field experiment RACORO is used to evaluate the performance of our new decision making method. RACORO was a field campaign with the objective to obtain aircraft observations of boundary layer clouds. Flights deployment decisions were made in a traditional manner; forecasts for BLCs were provided to a team of scientists, who then made fly/no-fly decisions. We compare the number of successful flights we would have launched using our method for decision-making to the number of successful flights actually launched during RACORO.

Our method conducts a cost-benefit analysis to launching a flight for each day of the experiment. The benefit from launching a flight is in the probability that the flight will be successful in collecting cloud data. The cost of launching a flight is being one day into the future with one less flight. A dynamic programming approach is implemented to quantify the expected number of observations still to be collected with \( d \) day and \( f \) flights remaining in the field season.
A flight will be launched if the forecast probability of success is greater than the *hurdle probability*. The hurdle probability is defined as the difference in the expected number of flights still to be launched between being one day into the future having the same number of flights and having one less flight.

We create a statistical forecasting method to predict the probability that good data collecting conditions will exist the following day. Days are grouped into different clusters based on the vertical profile of relative humidity. The probability of good data collecting conditions is found for each cluster. A numerical weather prediction model predicts the RH profile, which is then matched to the most similar cluster. A forecast verification is done on the model to determine the conditional probability that a realized profile belonging to cluster \( r \) will exist for a given model prediction belonging to cluster \( f \). Using cloud fraction data from ARM’s CMBE dataset the probability of good conditions existing for each cluster \( r \) is known.

During the field season good data collecting conditions existed on 45 of the 160 days, as verified by scientists involved in RACORO. During RACORO, flights were launched on 28 of the 45 total good days. The decisions made from our method would have resulted in flights being launched on 34 of these 45 good days, an increase of 20% over the number of successful flights launched during the experiment. These results are not highly sensitive to the details of our forecasting system.

Another advantage to our algorithm is that it can operate effectively in real-time and save on human capital currently spent in the forecasting and decision making process. With future research this method can be implemented for experiment planning purposes to help determine the optimal amount of total flights and experiment length.
References


