Emergency Response Transport Forecasting Using Historical Wind Field Pattern Matching

ROGER G. CARTER AND ROBERT E. KEISLAR
NOAA Air Resources Laboratory, Field Research Division, Idaho Falls, Idaho

(Manuscript received 14 May 1998, in final form 9 February 1999)

ABSTRACT

Historical pattern matching, or analog forecasting, is used to generate short-term mesoscale transport forecasts for emergency response at the Idaho National Engineering and Environmental Laboratory. A simple historical pattern-matching algorithm operating on a database from the spatially and temporally dense Eastern Idaho Mesonet is used to generate a wind field forecast, which then is input to an existing puff diffusion model. The forecasts are rated both by a team of meteorologists and by a computer scoring method. Over 60% of the forecasts are rated as acceptable. The forecasts also are compared with a persistence method, using both a subjective human evaluation and root-mean-square error calculations.

1. Introduction

Weather forecasting by historical pattern matching or analog forecasting employs a simple concept: a historical weather pattern (analog) is found that matches selected characteristics of the current weather pattern as closely as possible and then the weather that followed the analog in time is used for the forecast. The technique has been employed in many different applications: general circulation forecasting (Radinovic 1975; Gutzler and Shukla 1984; van den Dool 1989); long-range weather (Schaumans 1973; Livezey and Barnston 1988; Toth 1989), temperature (Bergen and Harnack 1982), and precipitation (Christensen 1981) forecasting; 1–6-day temperature forecasting (Kruizinga and Murphy 1983); long-range prediction of sea ice anomalies (Chapman and Walsh 1991); short-term visibility forecasts in the United States (Tahark 1975; Chisholm 1976; Miller et al. 1977) and Canada (Esterle 1992); and the El Niño–Southern Oscillation index forecasts (Drosdowsky 1994).

The popularity of analog forecasting has waxed and waned since World War II. Namias (1951) summarizes the early history and identifies pros and cons of the approach that still are cited and argued in more recent work (Bergen and Harnack 1982; van den Dool 1989). Analog forecasting gained sufficient credibility to appear in textbooks (Taylor 1954). It was used for operational long-term forecasting in Great Britain (Bowen 1976) from 1963 through the 1970s, but discontinued thereafter, as Bergen and Harnack (1982) state, “presumably due to lack of skill.” There is general agreement that analog forecasting skill on the synoptic scale over 1–6 days is still subject to the criticism of Lorentz (1969): “there are numerous mediocre analogs but no truly good ones.”

Although analog forecasts sometimes demonstrate modest skill (Esterle 1992; Hansen and Riordan 1998), they often display no significant, long-term advantage over the benchmark standards of persistence (i.e., the forecast event will be identical to the current event) and “climatology” (here defined to mean that the forecast event will be identical to the average of past events for the same time period over some historical subset of time periods) (Brody 1981; Chapman and Walsh 1991; Toth 1991; Drosdowsky 1994) or even over random chance (Gutzler and Shulka 1984). However, operational uses of analog forecasting for long-term forecasts currently exist (Livezey and Barnston 1988; Kerr 1989), and statistically significant success has been reported by many researchers for weather phenomena on a more regional scale (van den Dool 1989). As outlined by Namias (1951), in addition to objectivity, ease of implementation, and fast operation, a distinct advantage of analog matching is that local climatological influences such as mesoscale-γ and microscale terrain features that sometimes are difficult for prognostic numerical models to resolve are “already built into the analog(s)” (van den Dool 1989).

The algorithm described in this paper applies analog forecasting to the generation of mesoscale wind fields for the purpose of forecasting transport and diffusion in...
an emergency response setting. The algorithm operates on data collected from the Eastern Idaho Mesonet (Fig. 1). This network consists of 32 meteorological towers located on and around the Idaho National Engineering and Environmental Laboratory (INEEL). The towers are spread over an area nearly 200 km long and 100 km wide. Wind, temperature, and humidity are collected every 5 min via a single-frequency radio network. Data also include barometric pressure, precipitation, and solar radiation at selected towers. The system has been in operation since 1993 and typically achieves over 99% data recovery. The dataset is dense, both spatially and temporally. The number of candidate analogs and the strong topographic forcing of the region (discussed below) suggested that analog forecasting may warrant consideration for this application.

The INEEL site is located in the northeastern end of the Snake River plain, which drains from the northeast to the southwest. Mountains to the northwest rise above 3000 m above mean sea level (MSL), and individual ranges generally are oriented southeast to northwest, with canyons in between that drain onto the INEEL site. The general surface of the Snake River plain is rolling high desert, is about 1500 m MSL elevation, and is punctuated by volcanic features such as lava fields and buttes. Plants are generally sagebrush and grasses. It is classified as arid to semiarid. The relatively dry air and infrequent low clouds permit intense solar heating and then rapid radiative cooling at night, providing a large diurnal change in temperature and fostering a strong, thermally driven diurnal pattern of upvalley (up the Snake River valley), southwesterly flow during the day and downvalley, northeasterly flow at night. The mountains on both sides of the Snake River valley help to channel any prevailing flow into either a southwesterly or northeasterly direction (Clawson et al. 1989). These predominant flows are complicated by the thermally driven flows in the canyons at the northwest edge of the site. These flows converge or diverge at roughly right angles to the flow in the river valley, and gyres and other complexities can arise.

The motivation for the development of this algorithm is emergency response requirements. The Eastern Idaho Mesonet is operated by the National Oceanic and Atmospheric Administration (NOAA) Air Resources Laboratory, Field Research Division (FRD) to support
INEEL activities that include emergency response. The INEEL uses both hazardous and nuclear materials. In the unlikely event that these materials were to be released into the atmosphere, forecasting their transport would be crucial. To facilitate this forecasting, data from the mesonet is provided to INEEL and State of Idaho emergency operation centers via computer workstations. These workstations display the current mesonet data and three years of historical data. They also allow users to run a simple mesoscale puff diffusion model (MDIFF for mesoscale diffusion; Start and Wendell 1974). MDIFF was developed specifically for INEEL and directly ingests the data from the Eastern Idaho Mesonet. The locations of some mesonet towers were selected to help MDIFF to resolve the wind flows around the complex terrain in eastern Idaho.

The display of near-real-time data with the integrated model has proven to be a very valuable tool. The model does not provide any prognostic capability, however, because it stops calculations at the current time. The model calculates where the material has been but not where it will go next. When considering possible evacuations and coordinating emergency response field teams, forecasts of transport over the next few hours are extremely valuable. Currently, this forecast is provided verbally by an FRD meteorologist who participates in the emergency operations center (EOC). In the time-critical environment of emergency response, communication of a verbal forecast so it is understood correctly can be difficult. There is a need for a printable projection of transport to aid in communication of the forecast. An automated wind forecast fills this need without adding to the workload of the meteorologist. It offers the additional advantage of being available 24 h per day, so it may be used before the meteorologist has arrived at EOC and developed a forecast.

Historical pattern matching was selected for testing for a number of reasons. First, it was very easy to integrate into the system. It could search directly on the existing database. Once a match was selected, data could be copied directly from the database, with no conversion, to create the input for MDIFF. Also, it did not require the acquisition of datasets (e.g., upper-atmospheric gridded data) from outside sources or the acquisition of additional computer resources.

The limited time period required for the forecast made the use of historical matching feasible. Synoptic weather patterns often travel across the Eastern Idaho Mesonet area in a few hours. For a historical match to be successful, it must have access to data from an area that is large enough that synoptic changes likely to affect the weather over the desired forecast period will be reflected in part of the data when the forecast is made. For example, a frontal passage in eastern Idaho often is accompanied by increasing winds. The wind speed increase typically shows up first at the southwest stations in the mesonet and then moves across the network over the next several hours. Because the required forecast time is less than the time typically required for the change to move across the network, synoptic changes that are likely to affect the wind patterns usually should be reflected in part of the stations when the historical match forecast is generated.

One concern with analog forecasting is that the dataset may not contain a match or analog to the current conditions (van den Dool 1989; Gedzelman 1994). The temporally dense dataset from the Eastern Idaho Mesonet provides a large number of cases to examine, which helps to increase the probability of finding a match (5-min data provides 105 120 measurements per instrument per year). Possibly more important are the wind field characteristics at INEEL. Previous work has shown that regional analog forecasting can be more effective when weather patterns are affected strongly by local conditions (van den Dool 1989). As described earlier, local effects often dominate the wind flow in the region around INEEL. Clawson et al. (1989) identify 12 basic flow patterns that account for 74% of the flow patterns at INEEL. This forcing of wind flow into a small number of patterns increases the chances of finding a good analog to current conditions.

Another option would be to use a numerical forecast model instead of historical pattern matching. A good numerical model that incorporates synoptic, regional, and local data may prove to be more accurate than historical pattern matching, especially for longer time periods. Implementing an automated forecast with a mesoscale model at a grid resolution that is fine enough to resolve the local effects can be considerably more costly, however. For example, it may require gaining access to several regional and upper-air datasets, gathering land use data, and acquiring a new workstation on which the model can run (Anthes and Warner 1978). Considerable expertise would be required to set up the model for a fine grid over the region, and run times could be on the order of a few hours to produce a forecast. FRD does not have the resources to do this in the near term. Historical pattern matching was implemented because it could be done quickly with existing resources and also would provide an interesting test of the technique.

2. Algorithm description

The algorithm finds the best historical match by minimizing the difference between wind fields. It uses the average magnitude $A_{tm}$ of the vector difference between the wind at the target time $t$ (i.e., the time to be matched), and the wind measured at the potential match time $m$:

$$A_{tm} = \sum_{j=1}^{n} w_j [(u_{ij} - u_{mj})^2 + (v_{ij} - v_{mj})^2]^{1/2},$$

where $(u_{ij}, v_{ij})$ is the wind vector at the target time at
If I was working in the EOC, I would be: (choose one)

Happy with this forecast: ___
Satisfied with this forecast: ___
Can’t decide if I would be satisfied with this forecast: ___
Not satisfied with this forecast: ___

FIG. 2. An evaluation sheet for rating a forecast generated by the historical pattern matching algorithm. One sheet was generated for each case in the test set and the entire set was rated by five meteorologists.
Table 1. Average meteorologist rating for two test sets of forecasts. Each set contains approximately 40 forecasts and was rated by five meteorologists.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Adequate or good ($\geq 2.5$)</th>
<th>Bad ($&lt; 1.5$)</th>
<th>Undecided</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 h</td>
<td>63%</td>
<td>13%</td>
<td>24%</td>
</tr>
<tr>
<td>2 h</td>
<td>67%</td>
<td>13%</td>
<td>20%</td>
</tr>
</tbody>
</table>

The $j$th tower, $(u_{mj}, v_{mj})$, is the wind vector at the match time at the $j$th tower, $n$ is the number of towers, $w_j$ is the weighting for the $j$th tower, which is 2 for towers at INEEL and 1 for all others. INEEL towers are weighted more because this area is the one of primary concern for the release of hazardous materials.

The quantity minimized $D_m$ is a combination of the average difference between the target time and the potential match time and the average difference between the hour previous to the target time and the hour previous to the potential match time:

$$D_m = (A_{t,m} + 0.5A_{t-1,m-1})P_m$$

where $P_m$ is a penalty for missing stations and is defined by

$$P_m = 1 + (2 - 2f_m)^3,$$

$f_m$ being the fraction of towers reporting wind speed and direction. The functional form of $P_m$ was selected to be close to unity if $f_m$ is greater than 0.8 but increases rapidly as $f_m$ decreases below 0.8, effectively eliminating potential match times for which a large amount of wind data is missing.

The previous hour’s data are included to give the algorithm some information on the changes occurring in the wind field. It is less important to match the previous data than to match the current data, so they are weighted less. Without any experience or mathematical basis to select the weightings, 0.5 was selected arbitrarily. More past datasets could be included, but every included set adds complexity to the algorithm. One set was chosen as a reasonable compromise.

The algorithm searches the last 3 yr of online data to look for the minimum value of $D_m$. Potential match times are considered only if they fall within these windows:

1) target time $\pm 4$ h, and
2) target date $\pm 45$ days.

For example, if the target time was 1700, then the time of the match must fall between 1300 and 2100. If the target date was 20 April, only days between 6 March and 4 June would be considered. These windows help the algorithm to match the seasonal and diurnal patterns and to reduce the search time. Even with the windowing, each run considers 26 481 candidates. The current implementation of the algorithm runs on Pentium-Pro 200-MHz computers and completes in approximately 2 min.

Table 2. Results of paired $t$ tests comparing persistence and the historical pattern matching algorithm. Each test set contains 40 cases.

<table>
<thead>
<tr>
<th>Test set</th>
<th>Human rating</th>
<th>Confidence level</th>
<th>Computer score</th>
<th>Confidence level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 h</td>
<td>2.05</td>
<td>95%</td>
<td>2.11</td>
<td>95%</td>
</tr>
<tr>
<td>4 h</td>
<td>1.67</td>
<td>89%</td>
<td>2.16</td>
<td>96%</td>
</tr>
</tbody>
</table>
3. Evaluating performance

Evaluation of the algorithm’s performance was more challenging than the implementation was. The primary goal of the evaluation was to determine what percent of the forecasts the algorithm produced were usable. The majority of the forecasts must be usable, or the algorithm is useless for emergency response applications.

There are several statistical methods that have been used to evaluate model performance (Fox 1980; Willmott 1982; Willmott et al. 1985; Steyn and McKendry 1988). These include root-mean-square error (rmse), mean average error, and several other measures. These measures are very useful for comparing two or more models that have been used to make the same prediction. However, they do not provide a means to say that the model output is good enough to use operationally. There is no way to say a forecast is acceptable if the rmse (for example) meets a given criterion. Because the primary goal of this project was to produce transport forecasts for use in emergency response, an evaluation of the suitability of the transport forecasts is presented first. Rmse calculations for the wind fields are presented in section 5 as part of the comparison with persistence.

The approach used to evaluate the fraction of forecasts that is acceptable was to ask the users to evaluate the results. This approach was deemed to be appropriate since they are the experts on what is acceptable. They know what decisions are being made, what parameters are important, and what information would be misleading or unacceptable under these circumstances. For this evaluation, a test set of 46 randomly selected cases was compiled. For each case, an evaluation sheet like the one shown in Fig. 2 was created. The sheet shows the wind field, the results of an MDIFF run that uses the actual measured winds, and the results of an MDIFF run that uses the forecast. Four rating choices were provided: good, adequate, undecided, or bad. The four-choice rating system was chosen specifically to meet the goal of this evaluation. Asking the participants to rank the quality of each forecast numerically (e.g., a 1 to 10 ranking) would not identify clearly the division between acceptable and unacceptable forecasts. One
FIG. 8. Time-integrated concentration isopleths from a 4-h MDIFF run for case 35 calculated with MDIFF using (a) real, (b) historical match forecast, and (c) persistence.

Participant may choose to rank acceptable forecasts as greater than 5 and another may adjust the ranking so that acceptable forecasts are ranked as greater than 8. A two-level “acceptable” or “unacceptable” ranking was considered, but this approach leaves the participants with the question of what to do with the marginal cases. There are cases that are acceptable but clearly not in the same class as the really good forecasts. Should these
Fig. 9. The [(a) and (c)] actual and [(b) and (d)] historical match forecast wind fields for case 35; (a) and (b) are at the target time and (c) and (d) are 2 h later.

The four-level ranking provides a way to divide the marginal cases without mixing them in with the clearly good and clearly bad groups.

The rankings were labeled descriptively in the following manner (see Fig. 2 for exact wording) to convey the intent for division into usable and nonusable groups and to keep the context focused on usability for emergency response:
If I were working in the emergency operations center, I:
  • would be happy with this forecast.
  • would be satisfied with this forecast.
  • cannot decide if I would be satisfied with the forecast.
  • would not be satisfied with the forecast.

Those forecasts that were scored in the top two categories we considered acceptable.

To ensure that these rating categories were not misunderstood, an instruction sheet was included with the evaluation sheet. It gave a brief explanation of the forecast method, specified the goal of the evaluation (i.e., to determine what fraction of the forecasts was usable), explained the content of the three graphs, and described the four rating categories in more detail. Copies of the evaluation sheets and an instruction sheet were distributed to the five meteorologists who work in EOC. They were informed verbally about the goals of the project and the need for the evaluation and had the rating sheet verbally explained to them. They each rated the forecasts without discussing the cases and without seeing the ratings given by the other meteorologists.

The first set was 1-h forecasts. The process was repeated with a different set of 2-h forecasts. The ratings assigned by the meteorologists then were given a numerical value: 4 is happy, 3 is satisfied, 2 is undecided, and 1 is not satisfied. Because an undecided forecast may be considered to be better than one that clearly is unsatisfactory yet not as good as one that clearly is satisfactory, undecided cases were included with a rating of 2. The ratings for each case were averaged. The results are shown in Table 1. Generally, the historical match forecast performed better in cases with high wind speeds. This outcome is as expected because the higher wind speeds often are characteristic of strong synoptic forcing. These conditions tend to persist for several hours, making short-term wind forecasting relatively easy. It was apparent that, even with medium wind speeds, the algorithm is more successful than in light wind conditions.

This subjective evaluation method has several disadvantages. It takes a considerable amount of time to distribute the rating sheets, to wait for the ratings to be determined, and then to collect and to compile the ratings. Subjectivity itself is a drawback, because it may be difficult for the meteorologists to remain unbiased in their ratings; after rating several sets, they may develop opinions about the forecasting technique that could influence their ratings. Future development of the historical pattern matching algorithm will require the rating of many test sets to evaluate improvements to the algorithm.

To overcome these problems, a computer scoring algorithm was developed to evaluate the similarity of two MDIFF model runs. It was developed using the data from the previously discussed test sets and was designed to emulate the human expert evaluation. It reads the output files from two model runs, referred to as run 1 and run 2, and calculates a comparison score $S$ that is based on the angle, distance, and spread of the puffs and uses the equation

$$ S = \frac{W_A \left(1 - \frac{A_l - A_s}{\pi}\right)^2 + W_L \frac{|L_1 \cap L_2|}{\max(|L_1|, |L_2|)} + W_W \frac{1}{2} \left[ \frac{|X_1 \cap X_2|}{\max(|X_1|, |X_2|)} + \frac{|X_1| + |X_2|}{\max(|X_1|, |X_2|)} \right]}{W_A + W_L + W_W}, $$

where $W_A$, $W_L$, and $W_W$ are the weights for angle, length, and width, respectively, and which are set at 1.0, 0.5, and 0.25; $A_l$ and $A_s$ are the central angles for the set of puff positions for model runs 1 and 2, in radians (see Fig. 3); $L_1$ and $L_2$ are the lengths for the set of puff positions for model runs 1 and 2; each length is defined by a line segment along the central angle that contains the distances along that line of all puff locations in the set. Then $|L_1 \cap L_2|$ is the length of the overlap (hence the use of set intersection notation) of the two line segments that define $L_1$ and $L_2$. Here $|L_i|$ denotes the length of line segment $L_i$; $X_1$ and $X_2$ are the widths for the set of puff positions for model runs 1 and 2 within the first half of the length (each width is defined by a line segment perpendicular to the line along the central angle that contains all the distances perpendicular to the central line of all the puff locations in the set); and $Y_1$ and $Y_2$ are the widths for the set of puff positions for model runs 1 and 2 within the second half of the length.

This equation is designed to compare model runs using the same factors that we believe humans subjectively consider when comparing model runs, that is, angle, length, and width. It is used to evaluate algorithm improvements as they are implemented and thus to minimize the number of times FRD meteorologists will need to rate the forecasts. The functional form of $S$ was selected from more than a dozen different comparison scoring methods because it agreed best with the meteorologist ratings. Several different values for the weights $W_A$, $W_L$, and $W_W$ were tested. The values used here gave the best agreement with the meteorologist evaluations. It is interesting that these weights reflect the relative importance of each of these parameters to someone who would be involved in emergency response.
operations. The direction of transport is the most important parameter since that determines which population centers are in the path of transport and thus may need to consider evacuation or protective actions. The angle of the plume center line consequently is weighted the highest.

Figure 4 shows the computer-based scores plotted against the meteorologist ratings for the two test sets of Table 1. The scores show reasonable agreement with the average meteorologist ratings, and there is no difference between the scores from the 1- and the 2-h test sets. The cutoff rating of 2.5 for an acceptable forecast corresponds to a computer score of 0.7, and the “not satisfied” cutoff rating of 1.5 corresponds to a computer score of approximately 0.5. In testing improvements to the algorithm, the key statistic is the fraction of cases above the “adequate” cutoff. The scatter evident in the plot may be explained largely by the variation in the meteorologist scores, which have an average standard deviation on the order of 0.6 for any one case.

4. Comparison with the persistence method

Analog forecasting has had difficulty improving on persistence, or the assumption that conditions remain constant into the future (Brody 1981). The forecasts from this algorithm were compared with persistence using a paired \( t \) test (Snedecor and Cochran 1967). Conceptually, this process is simple:

1) assign a numerical score \( f \) to the comparison between the forecast and the actual weather conditions,
2) assign a numerical score \( p \) to the comparison between persistence and the actual weather conditions, and
3) perform a paired \( t \) test using the difference \( f - p \) between these two numerical scores.

In practice, the difficulty is in assigning the scores. Two methods were used for this evaluation: a human rating and the computer score discussed in the previous section.

For the human rating, printouts of MDIFF model results using the actual weather, the forecast, and persistence were made for each case in the test set. These three were compared manually and difference values were assigned as

1) \( f - p = 1 \) if the forecast agreed better with the actual weather,
2) \( f - p = -1 \) if persistence agreed better with the actual weather, and
3) \( f - p = 0 \) if there was not a clear difference.

For the second method, the score calculated with Eq. (4) was used for \( f \) and \( p \). The mathematical difference then was used for the \( t \) test except when both \( f \) and \( p \) were less than 0.5. In this event, neither the forecast nor persistence shows any correlation to the actual weather conditions, and it is not appropriate to define a difference that suggests one is better than the other. The quantity \( f - p \) was defined to be 0 for these cases.

The comparison was completed on two test sets, a 4-h forecast and a 2-h forecast. The results are shown in Table 2. Both scoring methods gave comparable results and indicated that the historical pattern matching is statistically better than persistence. Only the human rating of the 4-h forecast dropped below the 95% confidence level,
to 89%. One-hour forecasts were not used for persistence comparisons because differences between the two methods are not apparent in 1-h forecasts, making the human rating difficult and not very informative.

5. Wind field evaluation

An evaluation of the wind fields provides useful insight into the operation of the historical match algo-
A perfect wind forecast obviously will provide the best possible transport forecast. A good transport forecast, however, may be made with some degree of imperfection in the wind field. A number of statistical measures for agreement between forecast and actual data have been used (Fox 1980; Willmott 1982; Willmott et al. 1985; Steyn and McKendry 1988). Rmse is probably the most common. For this evaluation, the statistical measures for vector quantities described by Willmott et al. (1985) were used. These measures are based on the vector difference between the observed and predicted wind vectors, so lower values are better and 0 is a perfect match. The rmse described by Willmott et al. (1985) is presented here. Figures 5 and 6 show rmse calculated for each 5-min period for two cases from the test set used to compare 4-h forecasts with persistence. These statistics were calculated using 12 stations on or near INEEL, which is the area of primary concern for release of hazardous materials. Rmse values for both persistence and the historical match forecast are included in the figures. Persistence starts with an rmse of 0, which rapidly increases to be similar to that from the historical match forecast. For case 35 (Fig. 5), both values are intermixed, and in case 37 (Fig. 6) persistence is lower for most of the 4-h period.

Space limitations prevent inclusion of plots from all 40 cases used in the 4-h comparison. Figure 7 shows the average rmse values for the 40 cases. Persistence is much lower (i.e., a closer match) for the first 1.5 h. After that, there is little difference, but persistence is slightly lower for most of the period.

This result is in opposition to the persistence comparisons discussed in the previous section. A closer examination of an example, case 35, reveals the reason for this. Figure 5 shows no clear difference between the rmse for persistence and for the historical match forecast for this case. Part of the time persistence is a little lower, and part of the time the historical match forecast is a little lower. Figure 8 shows the MDIFF output using the actual data, the historical match forecast, and persistence. The persistence output shows a very small plume footprint and transport that is in the opposite direction to that of the actual data. The historical match forecast shows transport in the same direction as that of the actual data and a plume footprint that is closer to the plume footprint calculated with the actual data. Although not perfect, the historical match forecast clearly is better than persistence in this case, yet the rmse calculations did not reflect this difference.

This result may be understood best by examining the time evolution of the wind field. Figure 9 shows the wind fields for this case. At the start of the period, the actual wind field in Fig. 9a and the historical match forecast (i.e., from the match time) in Fig. 9b show light and variable winds on the INEEL site but a strong southwesterly flow in the southern part of the mesonet area. Two hours later (Figs. 9c and 9d), the southwesterly flow has spread over most of the site. The historical match forecast correctly predicted the shift to a southwesterly flow for which persistence could not account.

Figure 10 shows the time evolution of the wind speed and direction at the building 690 tower (labeled as 690 in Fig. 1), which is near the release point used in the MDIFF model runs. Both the actual winds and the historical match forecast winds are included. Both show a direction change from about 80° to about 200° and a large increase in speed. In the historical match forecast, however, the direction shift is later and the speed increase is earlier. Also, the wind speed fails to drop off at the end of the period to the same extent that the actual winds do. These differences keep the rmse values relatively high and comparable with the rmse for persistence. Because the historical match forecast does get the trends in the wind field correct, the transport calculations were rated more highly by the meteorologists.

It appears that the rmse calculation does not consider the trends in the wind field, which may account for it showing persistence as performing slightly better than the historical match algorithm, although the meteorologist rating indicates that the historical match forecast is better. Willmott et al. (1985) indicate that the rmse calculation may be applied to measurements made at a single station at different times. Figure 11 shows the rmse values for a number of stations calculated using all data in the last 3-h of the evaluation period. The first hour is omitted because persistence is much better for the first hour, and including it may bias the results in favor of persistence. Five stations have lower rmse values for persistence and seven have lower rmse values for the historical match forecast. The magnitude of the difference is about the same when persistence is lower as it is when the historical match forecast is lower. Although this finding does seem to agree better with the transport results, it still does not give a clear indication
Fig. 13. The wind field at (a) the target time of the good example case, (b) the 1-h model output using actual winds, (c) the wind field at the match time, and (d) the 1-h model output using winds from the match time. Panels (b) and (d) are zoomed in to show the model output clearly (see Table 3.)
of which method would be better. It seems that even the time-based rmse does not reflect the trends adequately.

Another example of this effect is case 37 in the same set. The rmse values are shown in Fig. 6. In this case persistence clearly seems to be better. However, the MDIFF output in Fig. 12 shows the persistence calculations taking a sharp turn to the northeast. While neither the persistence nor the historical match forecast shows great agreement, the historical match forecast is preferred for emergency response operations because it predicts the direction better. In emergency response operations, direction is the most important factor since it determines for which facilities and population centers protective actions will be planned. In summary, although rmse is a useful measure for many applications, the rmse of the wind field may not be an adequate evaluation tool for this application.

6. Detailed evaluation: “Good” and “bad” match examples

To improve the algorithm and to give the meteorologist some guidance on when to place confidence in the resulting historical match, meteorological explanations were sought for the successes and failures of the algorithm. Four cases with meteorologist ratings of 4, 3, 2, and 1 were selected from the 1-h evaluation set for further study. These cases are summarized in Table 3. They were selected because of good agreement on the rating among the meteorologists. Cases with high, steady winds that are associated with strong synoptic forcing were not considered for detailed evaluation.

Figures 13a,b show the wind field and associated model isopleths for the good (rating 4) target time, and Figs. 13c,d show the same for the match time selected by the algorithm. The 500-hPa charts for the target and match dates (Figs. 14a,b) show similar patterns over eastern Idaho. Satellite images (not presented) show similar IR signatures over eastern Idaho for these two days. The surface weather maps show highly similar northeasterly surface gradients of 8 hPa in 250 km (1016 to 1008 hPa in both target and match surface maps) across eastern Idaho and western Wyoming. Combined with the channeling of the topography adjacent to the upper Snake River plain, the northeasterly winds in Figs. 13a,c may be attributed to highly similar forcing.

Figure 15, in the same format as Fig. 13, shows the bad (rating 1) match. The ambiguous wind fields alone are sufficient reason to suspect the match, but Figs. 15b,d clearly show the diverging patterns in the plume evolution between the target and match times. Also, the 500-hPa maps for the target and match times (Figs. 16a,b) show noticeably different patterns over eastern Idaho. Furthermore, the satellite pictures (not presented) show that the target time is cloudy (surface maps show snow), but the match time is clear and therefore more likely to follow drainage flow patterns as the evening sets in.

Figure 17 shows two 1200 UTC soundings from Boise, Idaho (250 km west of the INEEL site), corresponding to the target and match times for the adequate (rating 3) case. At the target time, an upper-level low was located off the British Columbia coast with strong west-southwest flow from the trailing edge of a ridge over eastern Idaho. At the match time, a cutoff low was located over southern Arizona with lighter northerly flow.
over eastern Idaho. Figure 17 shows these considerable upper-level differences, but the near-surface flow reflects the typical nocturnal drainage pattern down the Snake River plain for both soundings. Surface maps also show that, in both situations, the skies are sufficiently clear for radiative cooling to dominate the flow regime. The fact that the case was rated as adequate rather than good, however, may be due to the different upper-air forcing over the upper Snake River plain and the canyons that drain onto the INEEL site from the northwest.
Barr and Orgill (1989) have shown that even subtle differences in flow above the valley affect drainage flow depth within the valley.

7. Conclusions

The historical pattern matching algorithm generated transport forecasts that were rated as usable over 60% of the time. This result is encouraging, given the simple algorithm used here, and is higher than may have been expected from historical applications of analog forecasting. A number of factors may have contributed to its success: the availability of a high-quality, spatially and temporally dense dataset; a tower network designed specifically to resolve the wind flows around INEEL; and wind fields that tend to follow a limited number of characteristic patterns because of mesoscale terrain features (Clawson et al. 1989).

The subjective human evaluation of the transport forecasts showed the results to be better statistically than those results based on persistence; the rmse calculations of the wind field forecasts showed no clear difference from persistence-based forecasts. This difference may be due to rmse not considering the trends in the wind field, which are of key importance in emergency response applications.

The simplicity of the algorithm allows it to execute very quickly and run on existing computer systems. It was inexpensive and very easy to implement. With improvement, it may be a useful tool for emergency response applications. It does provide a set of baseline results with which to compare future improvements and other methods. It is hoped that future work will include the application of cluster analysis, neural networks, or other methods to this same problem.

Acknowledgments. The authors would like to acknowledge the dedicated work of Shane Beard and Rayanne Ridenour during the completion of this project. This work was supported by the U.S. Department of Energy under Interagency Agreement DE-AI07-88ID12717.

REFERENCES


Bergen, R. E., and R. P. Harnack, 1982: Long-range temperature...


