OPTIMIZING MATCHING CRITERIA FOR START TIME ANALYSIS OF REAL-TIME "SPOT" FORECASTS AND CLIMATOLOGY FOR LONG RANGE OIL SPILL TRAJECTORY RESEARCH

Christopher H. Barker, Ph.D., Marc K. Hodges*
Office of Response and Restoration
Hazardous Materials Response Division, Seattle, Washington

1. INTRODUCTION:

In the arena of oil spill trajectory forecasting, standard nowcast/forecast models predict the movement of spilled oil out to 36-48 hours. Generally, most models are unreliable if used beyond 36-48 hours due to the cumulative errors that compound due to the uncertainty in the input parameters. Research conducted at NOAA's Hazardous Materials Response Division (HAZMAT) indicates that the primary error that prohibits accurate long-range forecasting is the wind uncertainty. This paper will investigate a method to analytically compare persistence and forecast winds to historical winds for two particular spill events.

This paper will present a simplified approach for comparing wind events, persistence or forecast, to a region's climatological record and ascertaining whether a predictable pattern in regional winds exists.

2. PROCEDURE:

The procedure involves selecting a set of past wind records corresponding to the forecasted wind records, to the lowest possible threshold value (E_g). This set of past wind records can then be extrapolated into the future, yielding a distribution of likely future wind patterns.

If we turn the time dependent forecast into a complex wind variable A_{kj}

$$A_{kj}(t = t_j) = U_k(t_j) + iV(t_j)$$

Here $U_k(t_j)$ is the kth forecasted x-component of the wind at time t_j. $V(t_j)$ represents the y-component. A similar set of historical wind segments B_{kj} can be generated by choosing different starting times in the past wind history of the location of interest.

The mean difference between the forecasted and historical winds can be calculated by

$$E_k = \frac{\sum_{t_j < T} | A_{kj} - B_{kj} |}{T}$$

E_k is then calculated for ever hour time t in the historical record B_{kj} . The result is a time series of E_k values. This time series is examined to determine all the local minimum values. Minimums are chosen, as when A_{kj} is a close match to a particular segment of the record, as you get closer to the best match time, E_k decreases in value, and then increases as you move past the optimum match time. Therefore, the best match will be a local minimum in E_k time series. Many of the local minimums are clustered together (see figure 1). As an optimum match is sought, each minimum is checked to see if it is the lowest local minimum within the length of the match record, T, and is only retained if it is the lowest value. The result is a set of times for which the match record can be considered a local best match to the measured record, each with an associated error, E_k. The threshold value, E_g is chosen by selecting the user-defined number of the minimums with the lowest associated errors. E_g is the largest error of the selected set of minimums, D_j, which are considered the match set. An example of a plot of E_k versus time is seen in Figure 1., with the local minimums highlighted.

Figure 2, is a schematic of forecast and persistence error over time. At some time T, the dynamic forecast error will become greater than

* Corresponding author address:
Marc K. Hodges, NOAA/HAZMAT, 7600 Sand Pt. Way NE Bld 3, Seattle, WA. 98115;
Email: marc.hodges@noaa.gov

the variability of climatology and persistence error will become asymptotic with that variability. Depending on the region, T is generally 36 to 48 hours. The variability of the mean error E_k, in relation to the data, will give a general scale of the variability of regional wind patterns.

D_j is an array of minimum errors in time. Let J_i be the array of times associated with each minimum error D_j. Hence, this J_i represents the times in history of the best possible matches between either persistence from a spill start time of length T or a given forecast from a spill start time of length T.

The measured records, M_i, that follow each of the matched records, are the set of possible wind patterns that could follow the original match pattern A_{kj}. A statistical oil spill model that uses these patterns, should result in a bound on possible spill behavior following the match time (see section 6).

The question remains as to whether this chosen set of patterns will result in a narrower set of possibilities than simply randomly selecting time periods from the climatological record.

To test this hypothesis, two regions were statistically examined.

3. THE REGIONS:

Two topographically different areas were chosen. Both of these areas had archived oil spill data or on-going oil spill data. The first spill region analyzed was Tampa Bay, Florida. On August 10, 1993, two barges and a freighter collided in the entrance to Tampa Bay. The barge Bouchard 155 spilled approximately 210,000 gallons of #6 fuel oil into the bay.
The second spill region analyzed was 12 miles west of the entrance to San Francisco Bay. The SS Luckenbach sank in 1953 due to a collision but only recently has the vessel started to leak. A salvage operation began in May 2002 to lighter the oil from the sunken remains. Since then, several small oil spills have occurred.

4. ANALYSIS:

To test the predictive capability of this approach, data from the two regions were examined, and set of randomly times were chosen for each. These times represent what would operationally by an oil spill or other event for which a forecast of the wind behavior is desired. For each of these times, the wind record both before and after the event is known. Operationally, only the wind record before the event time would be known. The record before the event time, A_{kj}, is then matched against the entire record, as outlined in section 2, resulting in a set of matches times J_i, and associated patterns, M_i, of length Q. Q is the time period for which a forecast is desired. Each M_i is compared to the measured record following the event time, resulting in a set of associated E_k values. The same procedure is repeated, using a randomly selected set of patterns N_i, from the historical record. The E_ks from the selected set and the random set are compared. If the errors are consistently lower for the selected set, then the method is shown to reduce the bounds on possible oil spill behavior following the match time, A_{kj}.

This approach was followed for the two regions of interest, Tampa and San Francisco Bays. For this analysis, the match period, T, is 36 hours, the forecast period, Q is 72 hours, and the 20 matches with the lowest E_k values were found for each record. The analysis was repeated for 20 randomly selected event times in each record. the mean of the results over each of these 20 were computed.

Tampa Bay:
mean E_g : 3.79
mean E_k for matched set: 8.44
mean E_k for the random set: 10.58

San Francisco Bay:
mean E_g : 4.17
mean E_k for matched set: 10.38
mean E_k for the random set: 13.10

All of these values are in knots. In an oil spill, the oil is moved by the wind at approximately 3% of the wind speed. Thus, the E_k values given above represent a variation of the distance that the oil might be moved by the wind over the 36 hour prediction period. The lower value for the matched set represents a smaller range of possible locations the oil could impact.

For example, for the Tampa results, the possible impact zone is reduced by about 2.3 miles. For San Francisco, the zone is reduced by about 2.9 miles.

This approach provides a tool to help determine the range of possible oil spill movement for about 36 hours. However, 36 hours into the future from any given event time is well covered by dynamic forecasting. The goal is to extend the period in which we can predict something about oil spill behavior beyond the range of accurate dynamic forecasting. This can be accomplished by leveraging a spot forecast for the event. In this case, the spot forecast is used as A_{kj} to find matching patterns in the historical record, rather than the recently measured data. this can extend the prediction period by the duration of the spot forecast.

5. FORECAST INTERPRETATION:

The method for the interpretation of the wind forecasts is described in (Lehr, et al, 2002). Here are the two oil spill incident examples.

The initial forecast during the Bouchard spill in Tampa Bay was E-NE winds at 15 knots shifting to be from the west at 5 knots by the afternoon hours. E-SE winds at 15 knots were predicted through the evening hours. E-NE winds at 10-15 knots were forecast throughout the next day. Using this method, the forecast translates into day, month, year, hour, minute, wind speed, wind direction format, as follows:

```
10,08,1993,06,00,15,070
10,08,1993,13,00,05,270
10,08,1993,17,00,05,270
10,08,1993,18,00,15,110
11,08,1993,06,00,12,070
12,08,1993,06,00,12,070
```

For the Luckenbach spill off the coast of San Francisco Bay, the initial forecast was NW winds at 15-20 knots increasing to 20-25 knots by the afternoon. NW winds at 20-25 knots were predicted through the evening hours with NW winds at 20 knots forecast for the next day. The forecast was interpreted to the following text file:

```
30,05,2002,06,00,18,315
30,05,2002,11,00,18,315
30,05,2002,12,00,22,315
30,05,2002,17,00,22,315
30,05,2002,18,00,22,315
31,05,2002,05,00,22,315
31,05,2002,06,00,20,315
01,06,2002,05,00,20,315
```

For Tampa Bay, an error E_k versus time plot was computed (see Figure 3) and minimum error values D_j below a threshold E_g=5.6, calculated using 20 matches, was produced. This demonstrates that an forecast record can adequately match the historical record.

6. LONG-RANGE OIL TRAJECTORY FORECASTING:

General NOAA Oil Modeling Environment, GNOME (Beegle-Krause, 2001) is the nowcast/ forecast model that is used by HAZMAT. This is a simple model that uses two dimensional physical processes to move lagrangian elements (LE's), representing quantities of oil, throughout the water. GNOME uses tides, hydrology, currents, winds and diffusion to move the LE's. Trajectory Analysis Planner (TAP) (Barker, 2000) and Extended Outlook are, respectively, HAZMAT's area contingency planner and long-range forecasting model. Extended Outlook uses three dimensional arrays (cubes) generated by GNOME and other post processors and displays the output on a waterbased grid.

The data for Extended Outlook is generated by running GNOME using the selected measured records M_i. These results show the bounds of possible oil movement.

7. CONCLUSION

This paper presented a simple approach for comparing a given wind pattern to a region's climatology and, by selecting a few variables (T, Q and E_g), accumulating specific measured wind patterns that can be used to enhance long-range spill trajectories and ascertain the nature of a region's wind variability.

There are obvious limitations to this analysis. This process can only be used in data rich areas, specifically, areas with long robust historical wind records (C-Man stations, buoys, ASOS and AFOS stations).

Future research should include applying this technique to several other topographically different spill areas, seasonalizing the historical winds, building in an analytical method to optimize Q and T to obtain the best match results.

6. REFERENCES

Barker, CH., J. Galt, 2000. Analysis of Methods Used in Spill Response Planning: Trajectory Analysis Planner TAP II, Spill Science & Technology Bulletin, Vol. 6, No. 2, pp. 145-152.

Beegle-Krause, CJ, 2001. General NOAA Oil Modeling Environment (GNOME): A New Spill Trajectory Model. IOSC 2001 proceedings, Tampa Florida March 26-29, 2001.

Lehr, W., D. Simecek-Beatty, M. Hodges. Wind Uncertainty in Long Range Trajectory Forecasts. IOSC 2003.

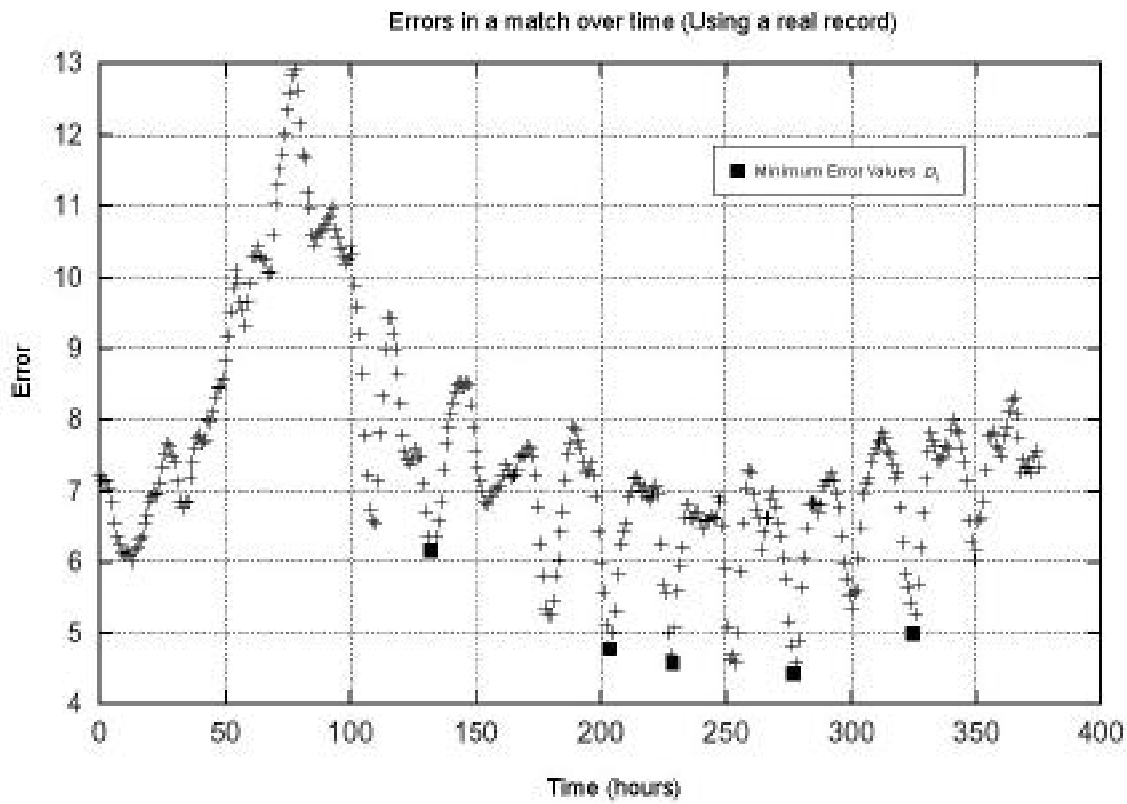

Figure 1

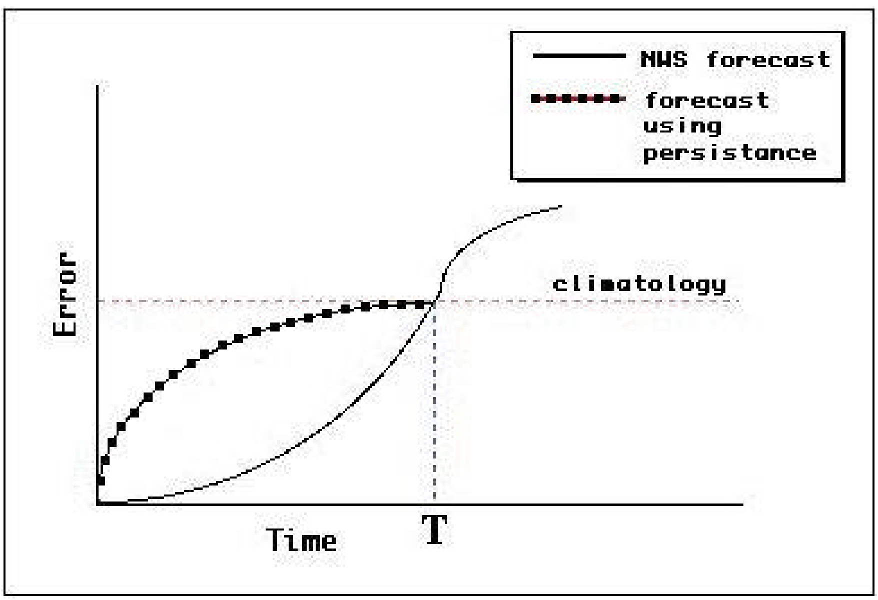

Figure 2

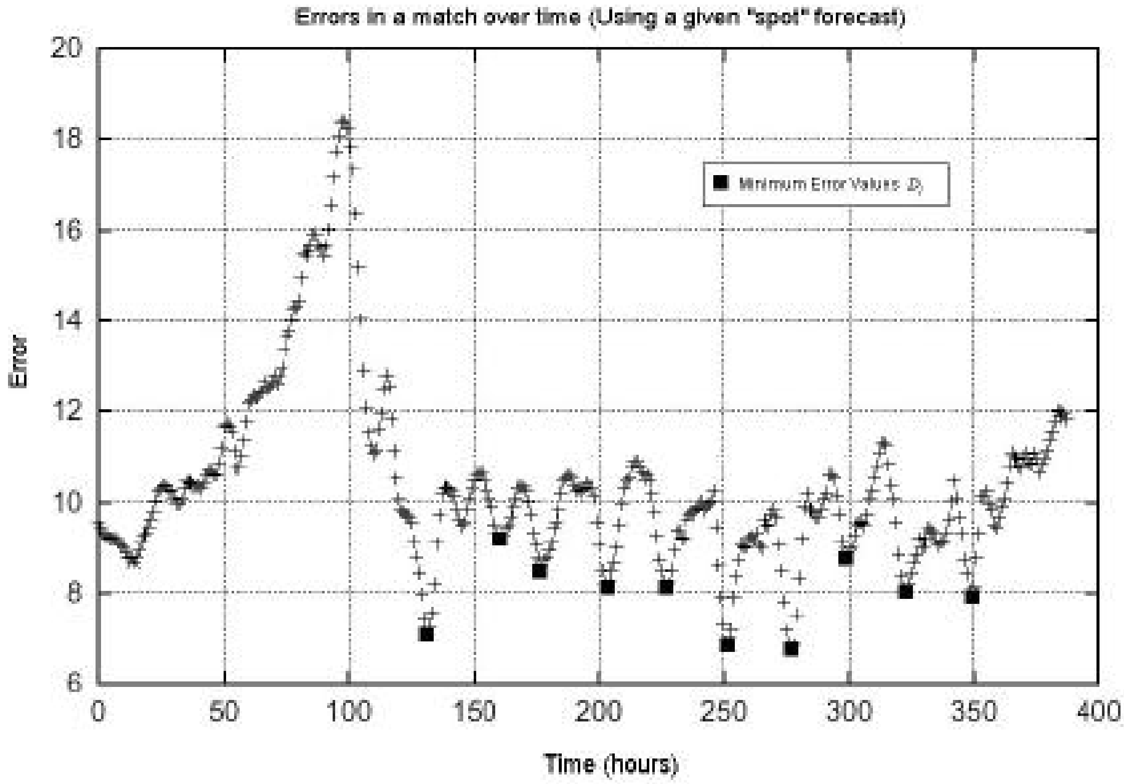
Errors in a match over time (Using a given "spot" forecast)
Minimum Error Values
Error
Time (hours)
20
18
16
14
12
10
8
6
0
50
100
150
200
250
300
350
400

Figure 3