ESTIMATING THE SUCCESS OF TRAJECTORY FORECASTS[1]

William J. Lehr, and Debra Simecek-Beatty
Office of Response and Restoration
National Oceanic and Atmospheric Administration
7600 Sand Point Way NE
Seattle, Washington 98115

A common tool in response to oil spills is trajectory forecasting. While the essential equations of oil spill trajectory modeling have not changed in four decades, the assessment of their effectiveness has only recently undergone quantitative analysis. The authors assert that estimating the success of any forecast can be approached in two different ways: (1) operationally, where the measured location of the spill is used to evaluate previous forecasts and improve the trajectory model in future forecasts, and (2) assessing the utility of the forecast, where spill location dependent cleanup decisions are affected by the model results, either favorably or unfavorably.

The challenge of determining operational success often lies in the method of estimating trajectory error in a defensible way that incorporates the uncertainty inherent in the reported slick observations and highlights the flaws in either the intrinsic model parameters or environmental input. The authors discuss the limitations of simply using the distance between the center of masses of the observed and forecasted spill locations. Some alternative supplementary approaches are discussed.

Adoption of a metric for spill modeling error allows the ability to assign anticipated accuracy to the forecast and highlights the input or model parameters most sensitive to forecast improvement. However, it does not measure the increased value of a better forecast to the actual response. To have an impact, spill forecasts must affect spill response decisions by providing additional information, often in the form of conditional probabilities, regarding oil amount and location. The authors review the likelihood that these probabilities would be, or could be, known, given different spill scenarios. If known, these probabilities can be combined with probabilities of success and anticipated rewards for different cleanup scenarios. By handling the problem in a probabilistic manner, trajectory analysis can assist analysts using risk-reward calculations to determine maximum possible gain while minimizing risk. Assessment in improved trajectory prediction would then be translatable into assessment in improved spill response. Whether existing technology is compatible with such an approach to trajectory forecasting and spill response remains an unanswered question.

INTRODUCTION

Modern spill response, like modern weather forecasting, places a high value on accurate information. Just as accurately determining the present and future location of a hurricane center is useful to an emergency management official, the future distribution of the spilled oil is useful to the on-scene commander in a spill incident. Therefore, a typical response to a significant spill will incorporate some form of trajectory analysis and forecasting. Often this is done through the use of computer models (Brebbia, 2001). Past IOSC proceedings have evaluated these model predictions (Daniel et al. 2001) for specific spill incidents but good general quantitative measures of spill forecasts are still under development. The authors assert that estimating the success of any forecast can be approached in two different ways: (1) operationally, where the measured location of the spill is used to evaluate previous forecasts and improve the trajectory model in future forecasts, and (2) assessing the utility of the forecast, where spill location dependent cleanup decisions are affected by the model results, either favorably or unfavorably.

OPERATIONAL MEASURES

To construct operational measures of success, spill observational data can be compared with spill location forecasts. Both the forecast and the observations are likely to be irregular and discontinuous. Therefore, it is not a trivial exercise to define a metric for forecasts success.

The simplest approach for oil spill modeling is to determine the geographical center of the spill forecast and relate it to the center of the observational data. This provides an easily explainable scalar measure. Larger separations between the forecast and observation centroids represent larger errors in the forecast. Tracking the centroid difference over a time series of forecasts and observations can indicate whether the model dynamics authentically represents the physical dynamics. Most models have adjustable parameters that can be altered to reduce future error. Moreover, systematic error can often be correlated with expected error in the model input data. A typical example is the case of inaccuracy in the predicted surface wind above the slick due to a spatial separation between the spill site and normal wind forecast location (often an airport). Adjustment to the ratio between wind velocity and wind-induced surface transport, an adjustable term in most models, can correct for this. However, this ratio may naturally change over time due to weathering of the oil so that continuing analysis and skilled calibration is required by the modeler.

Simplicity of calculation with the centroid difference method comes at the price of completeness and accuracy. For example, the approach does not provide an indication of the oil distribution about the centroid. It is possible that the oil is widely scattered about the centroid with heavier oil concentrations some distance away at the leading edge. In this instance, the approach does not measure the accuracy of the predicted arrival time of either the heavier concentrations of oil or the leading edge of the slick. While the approach measures general location of the slick, it gives no indication whether the shape of the forecast footprint matches the observational footprint. Also, there is no possible way to ascertain if the observational data itself contains false information.

One way to compare the shape of the predicted and observed slicks is to use an area matching approach to the error analysis. Based on comparison between observational maps and trajectory prediction maps, the modeler determines which coastal water areas are occupied by the forecast, by the observed surface oil, or by both. The overlap is a measure of the accuracy of the forecast. Let be the area forecast to have oil, the observed oil area and the common area (both forecasted and observed). Then one possible measure of accuracy is

$$f_N = \frac{2N_{of}}{N_o + N_f} \tag{1}$$

Figures 1 and 2 give an illustration of both techniques. On December 30, 2003, approximately 4800 gallons of IFO 380 were spilled into the Puget Sound near Edmonds, Washington. Initially, the oil drifted south into the main channel off North Seattle and Shoreline. Eventually, the oil crossed the shipping lanes to the western side of Puget Sound and into the Port Madison area. Figure 1 shows the overflight map for December 31, 2003 at 0910 Pacific Time and the trajectory for December 31, 2003 at 0800

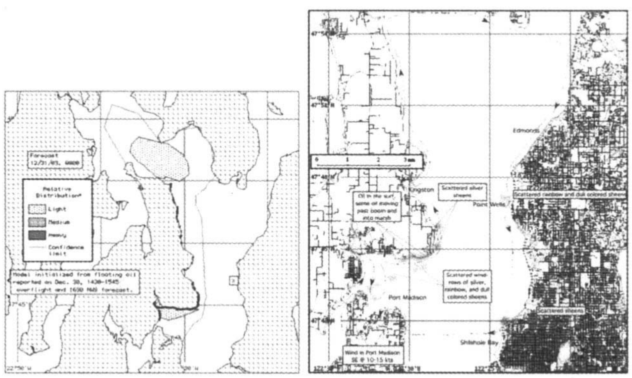

FIGURE 1. TRAJECTORY FORECAST (LEFT) AND OVERFLIGHT (RIGHT) MAPS FOR PT. WELLS SPILL, DEC. 2003.

For calculation of the centroid difference, a polygon was created enclosing the Lagrangian elements (sometimes called spillets) generated in the trajectory model run. The overflight was also digitized with a polygon drawn around the observed surface oil. The two smaller sheens at the spill site and to the south were not included in the analysis as they did not represent the bulk of the oil. The polygons and the centroids of the slick are displayed in Figure 2. By combining both methods, the modeler would have both an estimate for the mean separation and the degree of overlap. For this example, the distance between the two centroids is around 2 nautical miles. The overlap accuracy, f_N, is 42 %. A slightly different way of using the two methods is to transform the two maps so that centroids are forced to coincide. Method 2 in this circumstance would then give a purer measure of the accuracy of the modeled dispersive processes.

A third approach provides a topological rather than numerical result with regard to forecast performance measurement. This approach, used experimentally during the Exxon Valdez spill, also divides the coastal region into observed and forecasted areas. However, instead of simply determining what fraction of observed and forecasted area overlap, the forecasted area is mapped into the observed area. The method of mapping is as follows: Trajec-

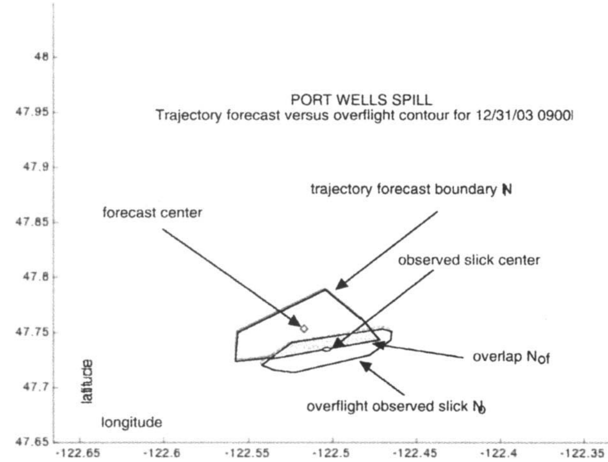

FIGURE 2. PT. WELLS SPILL COMPARISON OF TRAJECTORY VERSUS OBSERVATION.

tory forecast area sections that match oil observed area sections are mapped into those sections. In this respect, the procedure is similar to the second approach. Next the remaining forecast area sections are mapped into the closest observed area and the distance between the two is noted. Once a forecast section is mapped into an observed section, that section is no longer available for further mapping. This process is continued until all the forecast sections are mapped into the observed sections. A graph of number of forecast sections versus distance between the sections is generated. While the mapping strategy is difficult to optimize, only the functional form is important so non-unique mappings can be used, neglecting the actual numerical results. Figure 3 shows this graph for the Pt. Wells spill data.

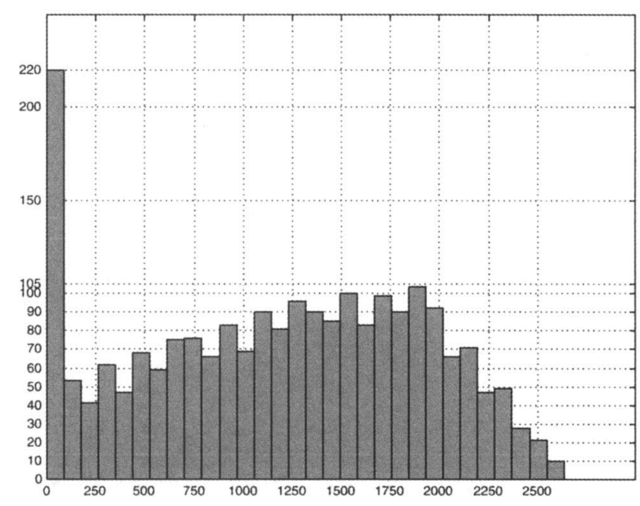

FIGURE 3. NUMBER OF OBSERVED POLYGON SECTIONS VERSUS DISTANCE MOVED IN MAPPING TO FORECAST POLYGON

The further the sections have to be moved, the worse the fit. However, the key information is contained within the shape of the graph. Normally, we would expect a generally declining smooth function with one significant peak as shown in Figure 3. Occasionally a second or third large peak appears. Such extra peaks are an indication of a possible false positive in the observational data or an unknown affect on the transport of the spilled oil. Thus this method provides a flag for looking for possible flaws in the observed oil or transport field input data.

UTILITY MEASURE

Adoption of a metric for spill modeling error allows the ability to assign anticipated accuracy to the forecast and highlights the input or model parameters most sensitive to forecast improvement. However, it does not measure the increased value of a better forecast to the actual response. To have an impact, spill forecasts must affect spill response decisions by providing additional information, often in the form of conditional probabilities, regarding oil amount and location. Figure 4 shows the risk information and decision flow of a typical spill event. Environmental forecasts will affect cleanup choices both directly, by affecting oil recovery efficiencies and indirectly, by altering the trajectory forecast and hence the resources that need protection or cleanup. This paper will only consider the latter.

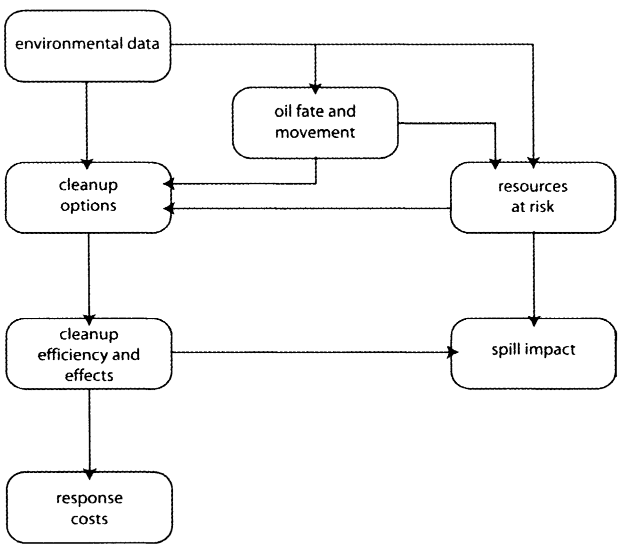

FIGURE 4. RISK INFORMATION AND DECISION FLOW OF A TYPICAL SPILL EVENT

Roulston and Smith (2002) have determined that the benefit of a weather forecast requires the creation of a cost function that takes into account the decisions the responder can make and the utility associated with possible outcomes. The authors suggest that this same approach should be applied to oil spill trajectory forecasts. Let i index the number of cleanup choices and designate a different trajectory of the spilled oil. Let C_{ij} be the cost of implementing a particular cleanup choice and E_{ij} be the expected benefit for cleanup choice i if the trajectory is j. Given a trajectory, j, the logical option of the responder is to choose cleanup option such that the net cost function

$$NC_{i,j} = C_{i,j} - E_{i,j} \tag{2}$$

is minimized. It is likely that both the benefits and costs may remain unchanged for nearby spill trajectories. Let represent all the close trajectories that have the same benefits and costs. Then if $j, j' \in k$, the error difference as defined in the first part of this paperbetween trajectory forecast j and j' is irrelevant as far as the utility to the cleanup is concerned. However, if a different forecast j'' resulted in different net environmental costs, then the error in trajectory prediction could result in higher cleanup costs (Figure 5).

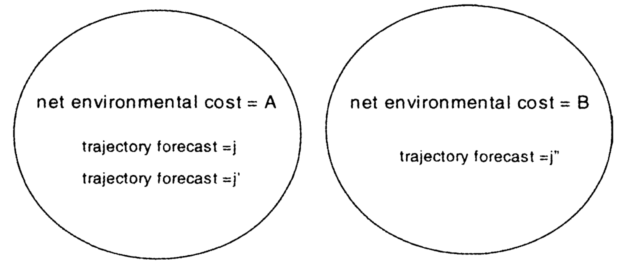

FIGURE 5. ONLY THOSE TRAJECTORY FORECASTS THAT SHOW DIFFERENT NET ENVIRONMENTAL COSTS ARE DIFFERENT WITH REGARD TO RESPONSE DECISIONS.

Furman (1982) has suggested that the absolute utility of weather forecasts be measured against climatology. The equivalent method for oil spill forecasts would be to select a cleanup option based on historical impacts rather than trajectory forecast analysis. These historical impacts can be determined either by recording natural collection areas, past spills, or by computer simulation of multiple spills using historical wind and currents (Barker and Wesley, 2005). Let $NC_{i',c}$ be the net cost of cleanup based upon such a 'climatological' forecast. The prime on the subscript signifies that the choice of cleanup may be different than either the optimum selection based upon the actual trajectory or the cleanup choice based upon trajectory analysis using forecasted, rather than historical, data. If $NC_{i'',o}$ represents the minimum net cost where the trajectory is exact and the response is the best choice, then in general we would expect

$$NC_{i',c} \geq NC_{i,j} \geq NC_{i'',o} \tag{3}$$

One estimate of the maximum benefit,, that could be achieved by trajectory forecasting is

$$MB = NC_{i',c} - NC_{i'',o} \tag{4}$$

However, it is possible that the trajectory forecast is so wrong that the responders would have been more successful using the climatological forecast option. Since the underlying physics of surface slick transport is basically understood and incorporated into existing models, this situation usually occurs when the input data to the model contains gross errors. It is not uncommon for the initial spill location report to be incorrect by 100 kilometers, the initial spill size estimate to be off by an order of magnitude, or the wind direction forecast to be incorrect by a quadrant or more. While little other than better training can be done to reduce the first two error cases, probability theory can help assess the spill location uncertainties involved in wind and current estimates. Increasingly, the National Weather Service provides its wind forecasts in probabilistic format. These probabilities or ones generated from past wind records (Barker and Galt, 2000) can be used to construct conditional probabilities for oil location given a present weather forecast. Let l be a particular weather forecast with probability P_l. Let $P_{j,l}$ be the conditional probability that spill trajectory j will happen, given the forecast l. Then the best estimate for the net environmental cost, NCE_i, of optimum cleanup strategy i given a forecast l is the minimum over all cleanup strategies of

$$NCE_i = \sum_{j,l} NC_{i,j} P_{j,l} P_l \tag{5}$$

Provided this number is lower than the net environmental cost based upon the 'climatology' $NC_{i',c}$, the trajectory analysis provides a positive return to the cleanup effort.

Lehr et al. (2003) discuss the requirements for calculating the conditional probabilities $P_{j,l}$, given the wind forecast probabilities l. In general, the longer the trajectory forecast in time, the more smeared is the probability distribution. For very long-range forecasts, the predictions revert to the climatological case. Barker and Wesley (2005) provide some values for $NC_{i,j}$ for a specific site, San Diego Bay. However, in most cases these numbers would not be known. Whether these numbers could be sufficiently estimated in an actual spill remains questionable.

CONCLUSION

Trajectory performance measures are useful to determine the accuracy of the forecast but are not trivial to compute. Different measures will provide different pieces of information and be useable in different ways. Operational measures can indicate errors in input data to the model or possible false positives in the spill observations as well as providing an overall accuracy metric. Utility measures determine whether increased resolution of the model will provide a net benefit to the response.

BIBLIOGRAPHY

William J. Lehr is with the Hazardous Materials Response Division of the National Oceanic and Atmospheric Administration (NOAA/HAZMAT). Dr. Lehr has a Ph.D. in Physics from Washington State University and has been conducting oil spill research for more than twenty years, particularly in the area of spill modeling. Debra Simecek-Beatty is a Physical Scientist with NOAA/HAZMAT. Ms. Simecek-Beatty has an M.S. degree in Marine Affairs from the University of Washington and has been active in the oil spill technology field for more than fifteen years, specializing in spill response and modeling.

REFERENCES

Barker, C.H. and D. Wesley, 2005, The Trajectory Recovery Analysis Planner (TRAP):

A tool for modeling the potential success of response plans, this proceedings.

Barker, C.H., and J.A. Galt, 2000. Analysis of methods used in spill response planning: trajectory analysis planner TAP II. *Spill Science & Technology Bulletin*, 6:145-152.

Brebbia, C.A. (ed) 2001. *Oil Spill Modelling and Processes*. WIT Press, London, 161 pp.

Daniel, P., P. Josse, and P. Dandin, 2001. Forecasting the Erika oil spills, Proceedings of the 2001 International Oil Spill Conference, American Petroleum Institute, Washington D.C. pp 649-655.

Furman, R.W. 1982. The effectiveness of weather forecasts in decision making: an example, *Journal of Applied Meteorology*, 21:532-536

Lehr, W.J., D. Simecek-Beatty, and M. Hodges, 2003. Wind Uncertainty in Long Range Trajectory Forecasts. Proceedings of the 2003 International Oil Spill Conference, American Petroleum Institute, Washington D.C.

Roulston, M. and Smith, 2002. Evaluating probabilistic forecasts. *Monthly Weather Review*, 20:1653-1660.

1 Although released by NOAA, the information in this paper does not reflect, represent, or form any part of the support of the policies of NOAA or the Department of Commerce. Further, release by NOAA does not imply that NOAA or the Department of Commerce agree with the information contained herein.