ADDRESSING WIND DIRECTION UNCERTAINTY IN SOURCE ESTIMATION THROUGH DYNAMIC TIME WARPING

G. Cervone[*1] J. Lin[2] P. Franzese[3]
[1]Dept. of Geography and Geoinformation Science, George Mason University
[2]Dept. of Computer Science, George Mason University
[3]Center for Earth Observing and Space Research, George Mason University

1. Introduction

Detecting the source of a pollutant release in the atmosphere, and identifying its characteristics, is an important problem due to the necessity to locate the source in order to take action or to correctly assess the potential damages caused by the release. The problem can be summarized as follows. Given a few measurements of pollutant concentrations and some basic meteorological information, the goal is to identify the characteristics of the release such as location, emission mass rate, temporal evolution, in order to be able to predict the fate of the contaminants [12, 5].

Source detection algorithms can be based on backward or forward simulation techniques. Backward techniques use reverse transport and dispersion simulations from the receptor to the source. Forward techniques use transport and dispersion simulations from different candidate sources, and compare the resulting concentrations to the available measurements. The algorithms search the characteristics of the source that minimizes the error between simulated and measured concentrations. An appealing characteristic of the forward techniques is that they do not require modifications to the dispersion model. Therefore, they can be used with any available dispersion model, independent of the complexity of the problem. We will apply a forward simulation technique.

Several forward iterative methods for source estimation have been developed [9, 14, 17, 8, 5]. In particular, evolutionary or genetic algorithms were employed to drive a search process based on forward numerical simulations, and it was shown that source characteristics were correctly identified for synthetic cases and for a controlled field experiment [4, 6].

Different measures of the error between the simulated and observed values were investigated to quantify the performance of the new candidate solutions. The error function is the only feedback that the algorithm receives on the quality of the newly generated solutions. It is usually referred to as error or fitness function, and its value is also called the skill score.

The correct wind direction is paramount to source estimation problems. It was observed that errors in wind direction of only a few degrees drastically worsen the source estimation. Even when the wind direction is carefully measured at the time of the release, as for example in a field experiment, the wind variability over the time of the release can be very large leading to large uncertainty and noise in the data.

*Corresponding author address: Dept. of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030; gcervone@gmu.edu

To address this problem, previous research investigated two different approaches. The first method consisted in choosing an error function that compares the simulated and observed values without taking into account their spatial distribution. In general the method performed poorly because the spatial location of the concentration plays a crucial role in correctly identifying the characteristics of the source. A second approach consisted in making the wind direction an unknown in the source estimation problem. This method generated good results, at the cost of increasing the complexity of the search process.

This paper introduces a third approach, namely the use of Dynamic Time Warping (DTW) [18, 28] to compute the error between simulated and observed concentrations. DTW is a distance measure that is commonly used in time series databases and mining [19, 26, 27, 28, 29] and signal processing communities [30, 32, 33, 31]. DTW uses dynamic programming techniques to determine the best alignment that minimizes the distance/cost/error between sequences. Its ability to produce nonlinear alignment between sequences makes it shift-invariant, and addresses the problem of errors in wind direction.

Since its introduction by Bellman in 1959 [18], DTW has been used extensively in the speech processing community [30, 32, 33, 31]. In 1994, Berndt and Clifford introduced DTW as a time series similarity measure to the database community [19, 28]. Due to its ability to minimize the effects of shifting on the time axis, DTW has been widely used in diverse fields. For example, Kuzmanic and Zanchi used DTW for hand shape (sign language) classification [20]; Corradini used DTW to recognize gestures and human activities [21]; Keogh et al. adapted DTW for various time series data mining tasks such as classification, clustering, and similarity search, on various applications such as motion capture matching and shape matching [34, 28, 29]; Niennattrakul and Ratanamahatana adapted DTW for k-means clustering for multimedia time series data [22]; Muller et al. proposed a multiscale DTW for music synchronization [23]; Aach and Church applied DTW on RNA and protein expression data [24]; and Zhang et al. compared DTW to other similarity measures for surveillance trajectory clustering [25]. While DTW is a robust similarity measure that outperforms many existing approaches, it is also computationally intensive. To mitigate this issue, several techniques for indexing DTW have been proposed [26, 27, 28]. In fact, Ratanamahatana and Keogh show that with indexing, DTW can be achieved in linear time when searching large databases [29].

This paper is structured as following: Section 2 discusses the methodology, including the different error measures used

and the numerical simulation performed; Section 3 describes the experiments performed and their results; Section 4 discusses the findings and suggests applications for the proposed method.

2. Methodology

2.1 *Transport and Dispersion Simulations*

The dispersion simulations are performed using a Gaussian reflected dispersion model, which determines the predicted mean concentration c_s at a location x, y and z of an atmospheric tracer released from a source located at x_s, y_s, and z_s:

$$c_s = \frac{Qg_yg_z}{2\pi U[(\sigma_s^2 + \sigma_y^2)(\sigma_s^2 + \sigma_z^2)]^{1/2}} \quad (1)$$

with

$$g_y = \exp\left[-\frac{(y - y_s)^2}{2(\sigma_s^2 + \sigma_y^2)}\right]; \quad (2)$$

$$g_z = \exp\left[-\frac{(z - z_s)^2}{2(\sigma_s^2 + \sigma_z^2)}\right] + \exp\left[-\frac{(z + z_s)^2}{2(\sigma_s^2 + \sigma_z^2)}\right] \quad (3)$$

where Q is the source mass emission rate, U is the wind speed, $\sigma_y(x, x_s; \psi)$ and $\sigma_z(x, x_s; \psi)$ are the crosswind and vertical dispersion coefficients (i.e. the plume spreads) where ψ describes the atmospheric stability class (i.e., $\psi = A$ to $\psi = F$), and $\sigma_s^2 = \sigma_y^2(x_s, x_s, \psi) = \sigma_z^2(x_s, x_s, \psi)$ is a measure of the area of the source. The result of the simulation is the concentration field generated by the release along an arbitrary wind direction θ. The dispersion coefficients are computed from the tabulated curves of Briggs [2].

2.2 *Prairie Grass Experiment*

The methodology was tested on data from the Prairie Grass field experiment [3]. The experiment consisted of 68 consecutive releases of trace gas SO_2 of 10 minutes each from a single source. The mean concentration was measured at sensors positioned along arcs radially located at distances of 50 m, 100 m, 200 m, 400 m and 800 m from the source. Information on the atmospheric conditions at the time of each release is available, and each experiment could be classified according to Pasquill's atmospheric stability classes [15, 10].

2.3 *Synthetic Dataset*

To test the methodology, in addition of the observed Prairie Grass measurements, we have created a synthetic dataset simulating each of the 68 releases using the model described in Equation (1) along with the meteorological and release characteristics of the original Prairie Grass experiment. The simulated concentrations are recorded at the corresponding sensor locations for the original experiments. For each of the 68 releases, a study on the effect of the wind direction was performed by varying the wind angle from -20 to +20 degrees, in a 1 degree increment. Consequently, for each of the original Prairie Grass release there are 41 synthetic releases. One case simulates the Prairie Grass experiments using exactly the parameters observed at the time of the experiments, while the other 40 vary the wind direction, and keep all other parameters constant.

This synthetic dataset allows to perform sensitivity studies to wind direction, determining for each experiment what is the wind direction that generates smaller error between the simulated and observed concentration. Even in a controlled experiment, like Prairie Grass, there are discrepancies between the observed wind direction and the angle that generates smallest error. This is because there can be errors in measuring wind at the time of the experiment, and because wind direction is usually not constant. The value reported in the official experiment summary is the average of the wind direction during the entire time of the release, and might contain errors.

2.4 *Error Functions*

The quantitative comparison of observed and synthetic concentrations is performed by applying several statistical measures of error which reflect different aspects of the spatial distribution of concentration. We considered two functions: the normalized root mean square error NRMSE [11, 7], and AHY2 [13, 1, 5].

$$NRMSE = \sqrt{\frac{\overline{(c_o - c_s)^2}}{\overline{c_o}\ \overline{c_s}}} \quad (4)$$

$$AHY2 = \sqrt{\frac{\overline{[\log_{10}(c_o + 1) - \log_{10}(c_s + 1)]^2}}{\overline{[\log_{10}(c_o + 1)]^2}}} \quad (5)$$

where c_o and c_s are the observed and simulated concentration at the sensors, respectively.

NRMSE is expressed in terms of variances, reflecting both systematic bias and relative random errors, which are estimated on a linear scale. NRMSE is strongly affected by infrequently occurring large overprediction or large observed outliers. AHY2, defined in [13] and [1] as metrics for the cost function of a genetic algorithm for source detection, computes the error on a logarithmic scale.

### 2.5 *Dynamic Time Warping (DTW)*

The error functions Eq. (4) and Eq. (5) are efficient to compute. However, both are sensitive to slight spatial distortions. To illustrate this, consider the dataset shown in figure 1a. The observed Prairie Grass measurements and the synthetic sequence generated for the same experiment look similar in shape; however, the change in the wind direction has caused the simulated data to shift slightly to the right. This slight shifting on the x-axis will result in large errors being computed by NRMSE and AHY2, since both error functions require one-to-one mapping of data points in space. To mitigate this problem, we propose to use DTW, a well-known distance measure for signal and time series data, as our error function. Given the two sequences $X = x_1, x_2, ..., x_n$ and $Y = y_1, y_2, ..., y_m$, DTW aligns the sequences by constructing a $n \times m$ matrix M, where each entry $M(i, j)$ represents the distance $d(x_i, y_j)$ between points x_i and y_j. The entry $M(i, j)$ also corresponds to an alignment between x_i and y_j [28]. To determine the best alignment between two sequences, DTW finds a path, $W = w_1, w_2, ..., w_k$, through the matrix that minimizes the warping cost, and satisfies the following constraints [18, 28]:
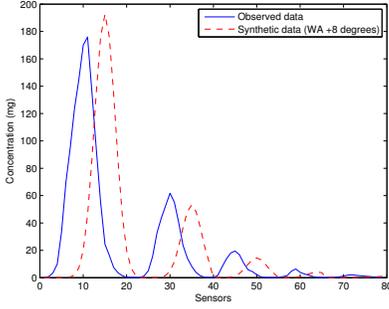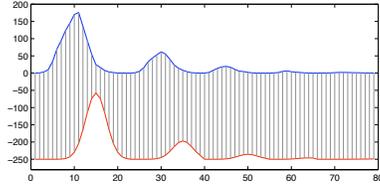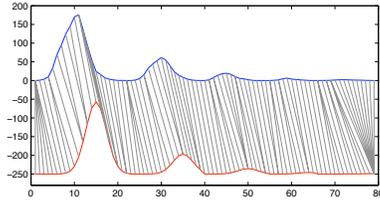
1. boundary conditions: $w_1 = (1, 1), w_k = (m, n)$. This requires that the warping path starts and finishes in the first and the last points, respectively, of the sequences.

2. continuity: Let $w_i = (a, b)$ then $w_{i-1} = (a', b')$ where $a - a' \leq 1$ and $b - b' \leq 1$. This confines the allowable steps in the warping path to neighboring points.

3. monotonicity: Let $w_i = (a, b)$ then $w_{i-1} = (a', b')$ where $a - a' \geq 0$ and $b - b' \geq 0$. This requires that the points in the warping path be monotonically ordered with respect to time.

The warping cost can be computed using dynamic programming with the following recurrence [28]:

$$f(i, j) = d(x_i, y_j) + min \begin{cases} f(i-1, j-1) \\ f(i-1, j) \\ f(i, j-1) \end{cases} \qquad (6)$$

In other words, the cumulative distance $f(i, j)$ is the sum of the distance between current points (x_i, y_j) and the minimum of the cumulative distance in the neighboring points [18].

figures 1b and c show two different alignments for the dataset shown in figure 1a. In figure 1b, the sequences are aligned using an error function such as NRMSE that requires one-to-one mapping, i.e. no warping allowed. As the figures illustrate, the peaks are not aligned properly, thus resulting in a large error. In contrast, in figure 1c, the sequences are aligned using DTW. The non-linear mapping allows the peaks to match, thus minimizing the error and the effect of wind direction.

Note that the concentration series that we analyze are not true time series, since the releases are continuous and the concentration field is stationary. Instead, the series describes the evolution of the concentration in space (not in time). The shift which is identified by DTW is the displacement of the simulated concentration field with respect to the observed values. Essentially, we are replacing the variable 'time' with the variable 'space', and instead of analyzing time-series, we are analyzing 'space-series'. (In this case the technique could be more correctly referred to as 'Dynamic Space Warping').

figure 2 shows the DTW distance matrix for input sequences X and Y. Each cell (i, j) represents the distance between X_i and Y_j. The white curve along the diagonal denotes the best warping path, i.e. one that minimizes the cumulative distance.

Some global constraint on the warping path is typically specified to restrict the warping paths. The advantages of using a global constraint are two-folds: (1) it produces more intuitive alignment, and (2) it speeds up the computation by narrowing the search space. A large warping window causes the search to become prohibitively expensive, as well as possibly allowing meaningless matching between points that are far apart. On the other hand, a small window might prevent us from finding the best solution. It has been shown by Ratanamahatana and Keogh [29] that by learning the best size and shape of the global constraint for different datasets, higher accuracy can be achieved. In our work, we use the Sakoe-Chiba Band [30], with 10% of the series length as the warping window length.

(a) Original and Synthetic data

(b) One-to-one error mapping

(c) DTW error mapping

FIG. 1: Example computing error between the synthetic and observed data using one-to-one and DTW mapping for Prairie Grass experiment 23. The synthetic data is shifted by eight degrees with respect to the observed parameters.
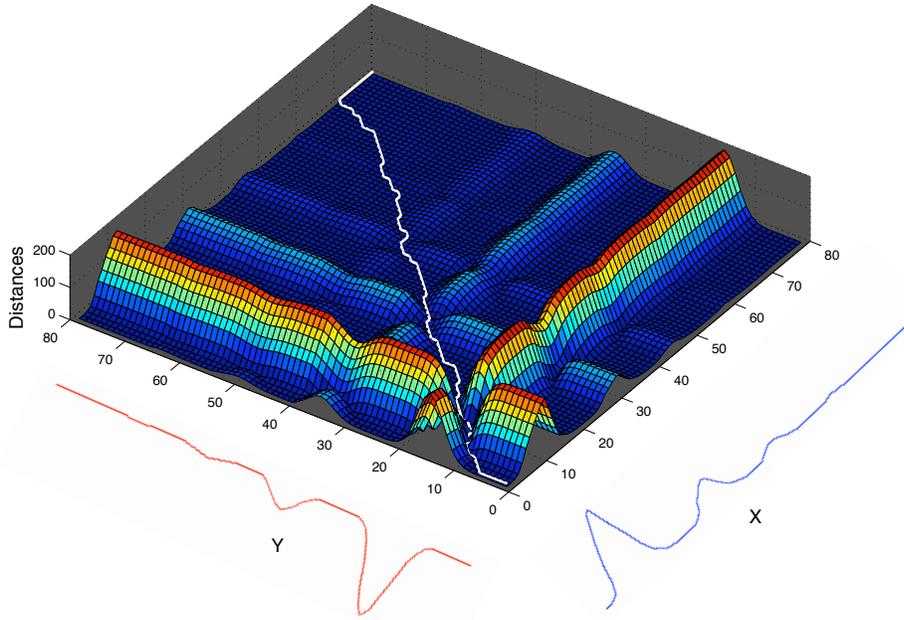
FIG. 2: DTW distance matrix for input sequences X and Y. Each cell (i, j) represents the distance between X_i and Y_j. The white curve along the diagonal denotes the best warping path, i.e. one that minimizes the cumulative distance.

3. Results

The proposed method was tested by performing a sensitivity analysis of the wind angle, computing the error between the observed and simulated concentrations. Experiments were performed for all 68 Prairie Grass experiments, and repeated for each of the three methods, DTW, NRMSE and AHY2. The hypothesis is that the DTW method, because of its ability to detect shifts in time-series, is less prone to errors in wind angle. In the Prairie Grass experiments the sensors are positioned along five concentric arcs, located at 50 m, 100 m, 200 m, 400 m and 800 m from the source. The measurements are transformed into a time-series by sorting each arc counterclockwise, from the inner arc to the outer arc. The footprint of each experiment changes due to the atmospheric class, the wind characteristics and the amount of release. Therefore each release is measured by a varying number of sensors, leading to time-series that vary in length, from tenths to hundreds of observations.

figure 3 shows the results for Prairie Grass experiment 23 (left) and 29 (right). The top figures show an interpolation of the release as measured at the time of the experiment. The sensor location are indicated with black dots, and the source of the release is at 0,0. The center panels show the results of the wind angle sensitivity analysis for each of the three methods performed. The graphs show the error between the simulated and observed concentrations as the simulated wind angle changed from −20 to +20 degrees with respect to the wind direction recorded at the time of the experiment. In both cases DTW obtained smaller error for a wide range of wind angles, whereas both NRMSE and AHY2 obtained best results for a very small range of values. For case 23, the best results are found when the wind angle change is close to 0, meaning that

the observed concentrations are consistent with the observed wind angle. For case 29, there is a variation between 6 and 9 degrees between the observed concentrations are consistent with the observed wind angle, indicating some noise in the observed data. The bottom panels illustrate the Cross-wind profiles of concentration for the releases, plotted as functions of the sensor number. The top graph shows the observed data and the synthetic data generated using the observed wind angle. The bottom graph shows the observed data and the synthetic data generated using an adjusted wind angle. It is evident that for case 29, a wind angle of -10 degrees with respect to the observed wind direction better approximates the observed values. This is an indication of noise in the observed data, most likely to be attributed to fluctuations in wind direction during the time of the experiment.

Table 1 summarizes the results for all the 68 Prairie Grass experiments sorted by atmospheric type. The original experiment identifier (PG ID) ia also reported in the table. NRMSE and AHY2 behave similarly, finding best results with a very small (usually 1 or 2) degrees of wind angle, DTW is able to find best results with a much higher number of wind angles.

4. Conclusions

This preliminary study shows that DTW can be effectively used as the error function driving algorithms for source detection. A current shortcoming of the available error functions is that they have difficulties recognizing simple spatial shifts in the simulated distribution of concentration. This results in the error functions reporting large errors even though the simulated

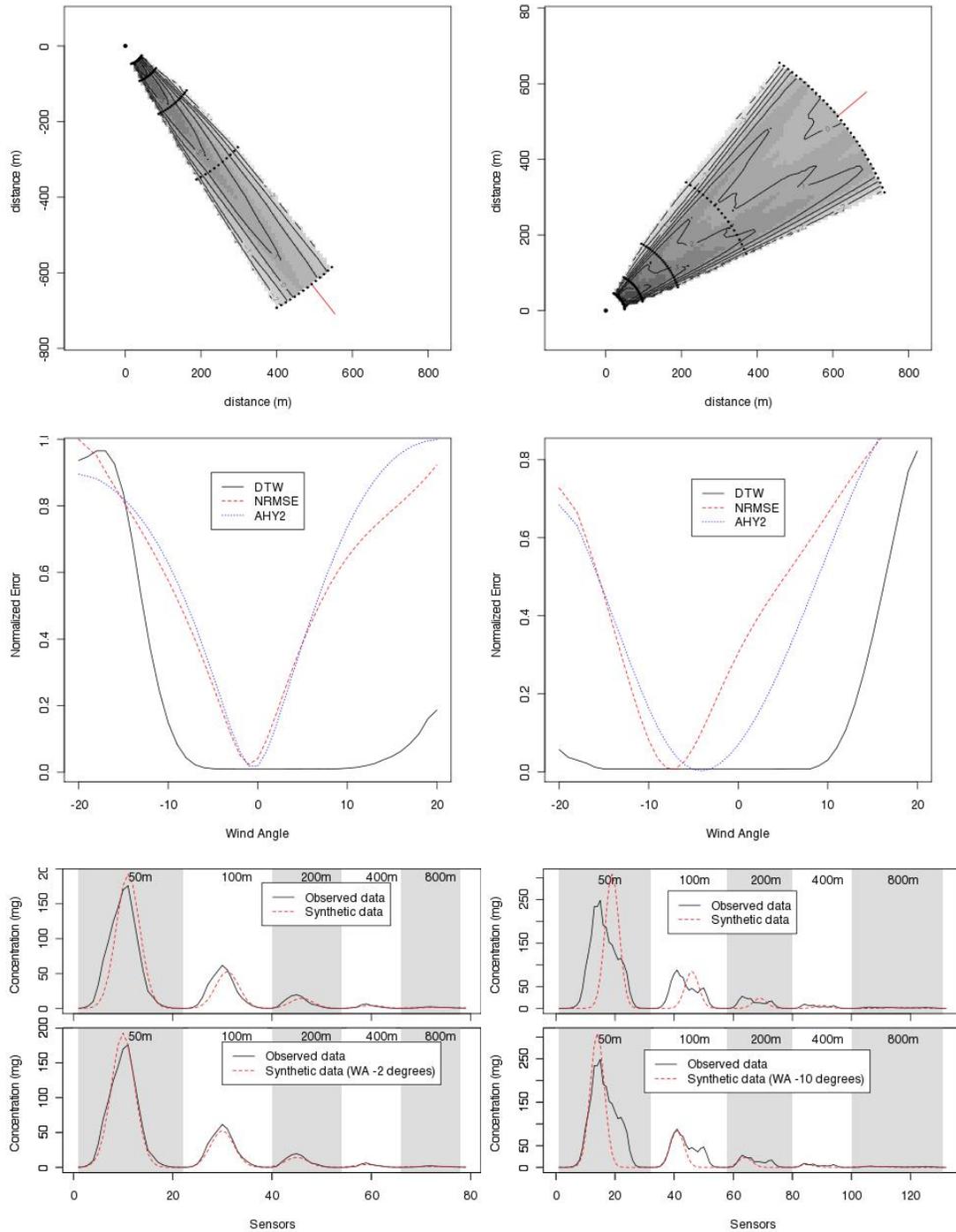Fɪɢ. 3: Results for Prairie Grass release 23 (left) and 29 (right). The graphs show a-TOP) the original releases, interpolated from the measurements made at the receptors (black circles); b-MIDDLE) the error found by AHY2, NRMSE and DTW as a function of wind angle; and c-BOTTOM) the observed and simulated concentrations using the observed wind angle, and the wind angle found by AHY2.

ψ	PG ID	DTW	NRMSE	AHY2
A	15	25	5	4
A	16	14	3	3
A	25	1	4	5
A	47	15	3	3
A	52	4	4	4
B	1	7	1	4
B	2	1	1	1
B	7	12	2	3
B	10	15	2	4
B	48S	3	2	3
C	5	19	2	3
C	8	29	2	2
C	9	23	1	3
C	19	25	2	3
C	27	27	3	3
C	43	21	3	4
C	44	13	3	4
C	49	23	2	3
C	50	23	2	2
C	62	27	2	1
D	6	29	2	2
D	11	21	1	1
D	12	27	2	2
D	17	13	3	2
D	20	29	2	3
D	21	21	2	1
D	22	17	2	2
D	23	19	1	2
D	24	23	2	1
D	26	17	3	4
D	29	25	2	3
D	30	19	2	3
D	31	21	3	3
D	33	28	2	2
D	34	25	1	2
D	35S	17	2	1
D	37	17	2	1
D	38	17	2	1
D	42	17	2	2
D	45	25	3	2
D	46	23	1	2
D	48	23	2	2
D	51	21	2	3
D	54	13	2	1
D	55	21	1	2
D	56	17	1	1
D	57	29	2	3
D	60	19	2	2
D	61	17	2	3
D	65	17	4	1
D	67	17	2	1
E	18	2	2	1
E	28	1	1	2
E	41	1	3	1
E	66	1	2	1
E	68	1	2	1
F	3	28	3	3
F	4	11	3	4
F	13	21	2	4
F	14	7	2	2
F	32	1	2	1
F	35	1	7	1
F	36	1	2	2
F	39	1	1	1
F	40	33	2	2
F	53	2	2	1
F	58	1	2	1
F	59	1	2	2

Table 1: Wind angle range in degrees for which best results were obtained using DTW, NRMSE and AHY2. The results are shown for each of the Prairie Grass experiments, identified by PG ID, and are sorted by atmospheric class

cloud is in fact very close to the measured one in terms of extension, shape, and magnitude, but not in the alignment. The results of the sensitivity study support the hypotheses that using DTW to compute the error between observations with simulations is less sensitive to wind direction changes. Furthermore, because some of the Prairie Grass experiments contained errors in the observed wind direction, the DTW method also works well in the presence of noise. The advantage of DTW over NRMSE and AHY2 is largest for atmospheric class A (unstable) through D (neutral). For stable atmosphere (E and F) DTW also finds best results for a rather small number of wind angles. This is most likely due to the more limited data available (smaller time-series) caused by a smaller footprint of the release, thus measured by fewer sensors.

5. Acknowledgments

REFERENCES

C. T. Allen, G. S. Young, and S. E. Haupt. Improving pollutant source characterization by better estimating wind direction with a genetic algorithm. *Atmospheric Environment*, 41(11):2283–2289, 2007.

P. S. Arya. *Air pollution meteorology and dispersion*. Oxford University Press, 1999.

M. Barad. Project Prairie Grass, a field program in diffusion. Technical Report Geophysical Research Paper, No. 59, Report AFCRC-TR-58-235, Air Force Cambridge Research Center, 218pp, 1958.

G. Cervone and P. Franzese. Machine learning for the source detection of atmospheric emissions. In *Proceedings of the 8th Conference on Artificial Intelligence Applications to Environmental Science*, number J1.7, January 2010.

G. Cervone and P. Franzese. Monte Carlo source detection of atmospheric emissions and error functions analysis. *Computers & Geosciences*, 36(7):902–909, 2010.

G. Cervone, P. Franzese, and A. Gradjeanu. Characterization of atmospheric contaminant sources using adaptive evolutionary algorithms. *Atmospheric Environment*, 44:3787–3796, 2010.

J. C. Chang, P. Franzese, K. Chayantrakom, and S. R. Hanna. Evaluations of CALPUFF, HPAC, and VLSTRACK with two mesoscale field datasets. *Journal of Applied Meteorology*, 42(4):453–466, 2003.

L. Delle Monache, J. Lundquist, B. Kosović, G. Johannesson, K. Dyer, R. Aines, F. Chow, R. Belles, W. Hanley, S. Larsen, G. Loosmore, J. Nitao, G. Sugiyama, and P. Vogt. Bayesian inference and Markov Chain Monte Carlo sampling to reconstruct a contaminant source on a continental scale. *Journal of Applied Meteorology and Climatology*, 47:2600–2613, 2008.

A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, 2003. 668 pp.

S. R. Hanna, J. C. Chang, and G. D. Strimaitis. Uncertainties in source emission rate estimates using dispersion models. *Atmospheric Environment*, 24A(12):2971–2980, 1990.

S. R. Hanna, J. C. Chang, and G. D. Strimaitis. Hazardous gas model evaluation with field observations. *Atmospheric Environment*, 27A:2265–2285, 1993.

S. E. Haupt. A demonstration of coupled receptor/dispersion modeling with a genetic algorithm. *Atmospheric Environment*, 39(37):7181–7189, Dec. 2005.

S. E. Haupt, G. S. Young, and C. T. Allen. Validation of a receptor/ dispersion model coupled with a genetic algorithm using synthetic data. *Journal of Applied Meteorology*, 45:476–490, 2006.

G. Johannesson, B. Hanley, and J. Nitao. Dynamic bayesian models via monte carlo - an introduction with examples -. Technical Report UCRL-TR-207173, Lawrence Livermore National Laboratory, October 2004.

F. Pasquill and F. Smith. *Atmospheric Diffusion*. Ellis Horwood, 1983.

S. K. Rao. Source estimation methods for atmospheric dispersion. *Atmospheric Environment*, 41(33):6964–6973, 2007.

I. Senocak, N. Hengartner, M. Short, and W. Daniel. Stochastic event reconstruction of atmospheric contaminant dispersion using Bayesian inference. *Atmospheric Environment*, 42(33):7718–7727, 2008.

R. Bellman and R. Kalaba On Adaptive Control Processes. *IRE Transactions on Automatic Control*, Vol. 4, No. 2, pp.1 9, 1959.

D. Berndt and J. Clifford Using Dynamic Time Warping to Find Patterns in Time Series. In *Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases*, pp. 229-248.

A. Kuzmanic and V. Zanchi Hand Shape Classification Using DTW and LCSS as Similarity Measures for Vision-Based Gesture Recognition System. In *Proceedings of EUROCON, 2007, the International Conference on Computer as a Tool*, pp. 264-269.

A. Corradini. Dynamic Time Warping for Offline Recognition of a Small Gesture Vocabulary. In *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, July 13. Vancouver, Canada.

V. Niennattrakul and C. A. Ratanamahatana. On Clustering Multimedia Time Series Data Using k-Means and Dynamic Time Warping. In *Proceedings of the 2007 International Conference on Multimedia and Ubiquitous Engineering*, pp. 733-738. Apr 26-28, Seoul, Korea.

M. Muller, H. Mattes and F. Kurth. An Efficient Multiscale Approach to Audio Synchronization. In *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR)*. pp. 192-197. Oct 10-12, 2006. Victoria, Canada.

J. Aach and G. Church. Aligning Gene Expression Time Series with Time Warping Algorithms. *Bioinformatics*, 17:495-508.

Z. Zhang, K. Huang and T. Tan. Comparison of Similarity Measures for Trajectory Clustering in Outdoor Surveillance Scenes. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*. pp. 1135-1138. Washington DC.

B. Yi, K. Jagadish and C. Faloutsos. Efficient Retrieval of Similar Time Sequences Under Time Warping. In *Proceedings of the 14th International Conference on Data Engineering (ICDE)*. Feb 23-27, 1998. Orlando, FL.

S. Kim, S. Park and W. Chu. An Index-Based Approach for Similarity Search Supporting Time Warping in Large Sequence Databases. In *Proceedings of the 17th International Conference on Data Engineering (ICDE)*. Apr 2-6, 2001. Heidelberg, Germany.

E. Keogh. Exact Indexing of Dynamic Time Warping. In *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB)*. Hong Kong, China. August 20-23, 2002.

C. A. Ratanamahatana and E. Keogh. Three Myths About Dynamic Time Warping. In *Proceedings of SIAM International Conference on Data Mining (SDM'05)*. pp. 506-510. Newport Beach, CA. April 21-23.

H. Sakoe and S. Chiba. Dynamic Programming Algorithm Optimization for Spoken Word Recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing*. 26(1):43-49.

L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ, Prentice Hall.

L. Rabiner, A. Rosenberg and S. Levinson. Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition. *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-26, 575-582. 1978

C. Myers, L. Rabiner and A. Rosenberg. Performance Tradeoffs in Dynamic Time Warping Algorithms for Isolated Word Recognition. *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, 623-635.

E. Keogh and M. J. Pazzani. Scaling Up Dynamic Time Warping to Massive Datasets. In *Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery*, pp. 1-11. September 15-18, 1999. Prague, Czech Republic.