27
12-20-78
25 copy to NTIS

MASTER

UCID-17982

Lawrence Livermore Laboratory

APPLICATION OF PRINCIPAL COMPONENTS ANALYSIS (PCA)
FOR LONG-TERM ASSESSMENT OF OPERATING RELEASES
FROM THE NUCLEAR POWER INDUSTRY

Marvin H. Dickerson
John J. Walton
Dieter R. Tuerpe

November 1978

THIS DOCUMENT IS UNCLASSIFIED

APPLICATION OF PRINCIPAL COMPONENTS ANALYSIS (PCA)
FOR LONG-TERM ASSESSMENT OF OPERATING RELEASES
FROM THE NUCLEAR POWER INDUSTRY

Marvin H. Dickerson, John J. Walton and Dieter R. Tuerpe
Lawrence Livermore Laboratory, University of California,
Livermore, CA 94550

November 1978

INTRODUCTION

NOTICE
This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

The objective of this study was to investigate the application of the Principal Components Analysis (PCA) methodology to the long-term assessment of operating releases from the nuclear power industry. Preceding this study we had compared the Gaussian calculations with MATHEW/ADPIC calculations for a simulated 24-hour release in the Hudson River Valley and an area in the southeast (Lange et al., 1976). Considerable differences were noted in the comparisons, which were attributed to the more realistic simulation of the time and space varying wind fields by the MATHEW/ADPIC computer codes.

The MATHEW/ADPIC computer codes require an unrealistic amount of computer time for long-term, e.g., monthly assessments; therefore we became interested in using the PCA methodology for pattern recognition studies. By defining the wind patterns and their temporal relationships

within a given month we could reduce the amount of computer time required to calculate a monthly assessment by a considerable margin.

The PCA technique for diurnal pattern recognition studies was initiated by the Division's Wind Power Assessment Program approximately two years ago. For the present study we built upon and extended the research and development effort of the Wind Power Assessment Project and applied the methodology to a month of data collected in the San Francisco Bay Area. The typical day calculated by the PCA was used to estimate the average monthly concentration for the month of August 1973. These concentration values were compared to the monthly average calculated by using all the hourly data for the month.

The wind-velocity data analyzed to determine a typical day was obtained from the Bay Area Air Pollution Control District (BAAPCD). These data consist of mean hourly wind speed and direction at ten locations in the San Francisco Bay Area. The winds reflect the influence of the complex topography and complicated inversion profiles characteristic of this area.

In order to study the statistical properties of the wind field Hardy (Hardy, 1977; Hardy and Walton, 1977) has generalized the method of principal components analysis (PCA) (also known as empirical eigenvector analysis) to treat two-dimensional vector fields. In this technique a series of simultaneous observations at a set of measurement locations provide the basis for a matrix similar to a correlation matrix. The eigenvectors of this matrix form a complete orthonormal set in terms of which the original observations can be uniquely represented. Other important properties of these eigenvectors are:

1. They are uniquely determined by the original data and are derived from that data through an objective mathematical procedure;
2. The eigenvectors provide a particularly efficient basis set in which to represent the observations;
3. The relative magnitudes of the eigenvalues make it possible to rank order the eigenvectors in terms of decreasing significance in representing the data;
4. The most significant eigenvectors can be identified with physically important patterns in the original data.

Given the complete set of eigenvectors calculated using PCA, the problem remains to represent the characteristics of a month in a compact way. The approach which has been taken is to obtain a typical day which best represents the common characteristics of a day in the month. This is accomplished by representing each hour of the day individually with a best fit of the eigenvectors. The best fit is calculated by minimization of the deviation of the eigenvector approximation from the observation averaged over the whole month. Such a fit will not be exactly like any one day but will exhibit the general or average properties the temporal pattern for the month. This approach also permits the identification of atypical days within the month, since these would be days for which the error measures substantially exceed those for the rest of the days of the month.

Figures 1-6 show the results of the above steps when applied to the Bay Area for the month of August 1973. In each case, figure (a) is the actual wind field observed on August 1, 1973, and figure (b) is the PCA representation of the corresponding hour. The figures show the wind field

at 4-hour intervals through the typical day. The basic westerly characteristic and the topographic channeling of this area are apparent. In addition, the diurnal rise through the morning and into the afternoon can be seen.

We used the MESODIFF model developed at the Air Resources Laboratory at Idaho Falls (Start and Wendell, 1974) to test the PCA method. MESODIFF is a sequential puff model which uses measured winds interpolated to the grid points to track pollutants released into the atmosphere. The center of each puff is tracked by this method and the puff then diffuses relative to its center by the standard prescription.

This model was brought to LLL to test out the PATRIC code (Lange, 1978) during its development. For uniform winds in both the vertical and horizontal directions the models compared favorably, as one would expect. For more complicated vertical structures in the wind fields the two models differ since in MESODIFF diffusion occurs with respect to a trajectory, whereas in PATRIC no trajectory is explicitly used. Since the PATRIC model was still in its development stage, we used the MESODIFF model for this study.

Figure 7 shows the integrated air concentrations for a 1 unit/second continuous release from the Golden Gate Bridge (point x in Fig. 7) obtained through MESODIFF using the complete hourly data set for the month of August. The continuous release was approximated by 6 puffs/hour, and the grid dimensions were 20 x 20 with a spacing of 10 km in each direction. Figure 8 shows the integrated air concentrations for the same conditions as in Fig. 7 using the "typical day" wind data obtained from the PCA method. To establish an equilibrium concentration, we ran the problem

for four days using the same typical day data set for each of the four days. The concentrations were then multiplied by 7.75 to obtain the integrated concentrations for a full month.

It is seen from a comparison of Figs. 7 and 8 that the model results utilizing the typical day data do not precisely reproduce the results obtained with the full data, but it has to be kept in mind that the release point is located near the point of maximum divergence in the wind field.

These methods have more recently been applied to data for March 1975 for an area in the southeastern United States (Walton and Sherman, 1978). This area presented a different problem than the Bay Area in that there was a great deal of day-to-day variability in the wind field. This led to the identification of several different groups of similar days rather than a single one. No group had more than 5 days and 10 days could not be matched with any others. The results of the pattern identification are listed in Table 1.

A typical day was produced for subgroup 1 and was used to drive the PATRIC code to predict, after 24 hours, the concentration of an inert tracer released from a constant source. For comparison, the original data for days 2, 3, 8, 20 and 25 were used to produce patterns. From the concentration patterns could be seen the similarities in the days as well as differences permitted by the comparison criteria.

TABLE 1. Groupings of similar days in March 1975.

Subgroup	Days				
1	2	3	8	20	25
2	7	12	13	24	29
3	14	19	22	30	
4	16	18	27		
5	6	21	23		
6	1				
7	26				
8	4				
9	5				
10	9				
11	10				
12	11				
13	15				
14	17				
15	28				

CONCLUSIONS

Principal Components Analysis does seem to provide a viable basis for the objective identification of similar wind field patterns. When similar patterns have been found, the method permits the computation of a best fit approximation. When this is done, the reduction in the computations can be quite striking. In the southeast was a factor of 2. In the San Francisco Bay Area a single computation did fairly well in representing 30 days worth of computation.

It is important to understand that the degree of data reduction and approximation is at the discretion of the analyst. That is, by broadening the acceptance criteria more days may be grouped together, reducing the total number of model calculations required but at the same time giving a cruder approximation for the driving wind field. On the other hand, the criteria can be made so stringent that unless two days are truly identical they will not be grouped together. The analyst will also have to exercise judgement based on the kind of problem he is considering, that is, the relative importance of accuracy in wind speed and direction can be different for different types of studies.

Some of the questions yet to be explored are: How does one identify and treat observations which are unrepresentative of the regional flow? What time intervals other than 24 hours might prove useful in finding recurring patterns? Given a number of subgroups of similar days, how should the results of the transport calculations be combined to give concentration patterns characteristic of a typical month?

ACKNOWLEDGMENT

This work was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore Laboratory under contract no. W-7405-Eng-48, and supported in part by the U. S. Nuclear Regulatory Commission under contract No. AO-158.

REFERENCES

- Hardy, D. M., "Empirical Eigenvector Analysis of Vector Observations," Geophys. Res. Letters, 4, 319, 1977.
- Hardy, D. M. and J. J. Walton, "Principal Components Analysis of Vector Wind Measurements," J. Appl. Meteor., 17(8), pp 1153-1162, 1978.
- Lange, R., "PATRIC, A Three-Dimensional Particle-In-Cell Sequential Puff Code for Modeling the Transport and Diffusion of Atmospheric Pollutants," Lawrence Livermore Laboratory Report UCID-17701, January 1978.
- Lange, R., M. H. Dickerson, K. R. Peterson, C. A. Sherman, and T. J. Sullivan, "Particle-in-Cell vs. Gaussian Calculations of Concentration and Deposition Out to 70 km for Two Sites," Lawrence Livermore Laboratory Report UCRL-52133, 1976.
- Start, G. E. and L. L. Wendell, "Regional Effluent Dispersion Calculation Considering Spatial and Temporal Meteorological Variations," NOAA Technical Memo ERL-ARL-44, 1974.
- Walton, J. J. and C. A. Sherman, "Objective Wind Field Pattern Characterization," Lawrence Livermore Laboratory Report UCRL-80901, presented at the 9th International Technical Meeting on Air Pollution Modeling and Its Application, Toronto, Canada, August 28-31, 1978.

FIGURE 1. Observed (a) and typical (b) winds at 4:00 a.m. in August 1973 in the San Francisco Bay Area.

FIGURE 2. Observed (a) and typical (b) winds at 8:00 a.m. in August 1973 in the San Francisco Bay Area.

FIGURE 3. Observed (a) and typical (b) winds at noon in August 1973 in the San Francisco Bay Area.

FIGURE 4. Observed (a) and typical (b) winds at 4:00 p.m. in August 1973 in the San Francisco Bay Area.

FIGURE 5. Observed (a) and typical (b) winds at 8:00 p.m. in August 1973 in the San Francisco Bay Area.

FIGURE 6. Observed (a) and typical (b) winds at midnight in August 1973 in the San Francisco Bay Area.

FIGURE 7. Integrated air concentrations for a unit/second continuous release from the Golden Gate Bridge, using the complete hourly data set for the month of August 1973.

FIGURE 8. Integrated air concentrations for a unit/second continuous release from the Golden Gate Bridge, using the typical day data for the month of August 1973.