Weather Uncertainty Total Deposition Novelty Detection

Kayla Schroeder[1], Donald Lucas[2], Giuliana Pallotta Goldhahn[2], Nipun Gunawardena[2]

[1]University of California, Los Angeles [2]Lawrence Livermore National Laboratory, National Atmospheric Release Advisory Committee

Abstract and Motivation

Background: The impact of an accidental or intentional release of hazardous materials is a matter of pressing importance due to the potential resulting catastrophe. As such, it is necessary to analyze weather variability and its effects on releases of hazardous materials.

Focus: Currently, there are many existing models to predict the motion of such a release. These weather models, however, are uncertain and quite time intensive requiring hundreds of cpu-hours per simulation. Thus, it is beneficial to quickly determine which models are the most effective predictors to reduce computational intensity.

Objectives

- Develop an effective model to characterize weather uncertainty and material deposition data.
- Investigate the capability of this model through novelty detection and predictive accuracy
- Determine the most effective parameters or combination of parameters for the model to best characterize the data.

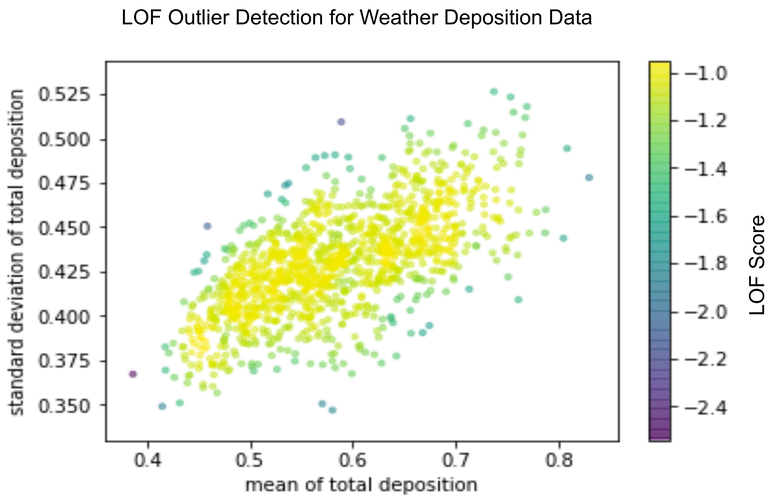
Methods

1. Local Outlier Factor
   Algorithm Theory:
   - Local reachability density is used to determine the necessary traveling distance from a point to reach the next point or cluster of points.
   - The average ratio of the local reachability density of a point to that of its neighboring points. If the density is much smaller than that of its neighbors then the point is classified as an outlier (since it is far from dense areas).

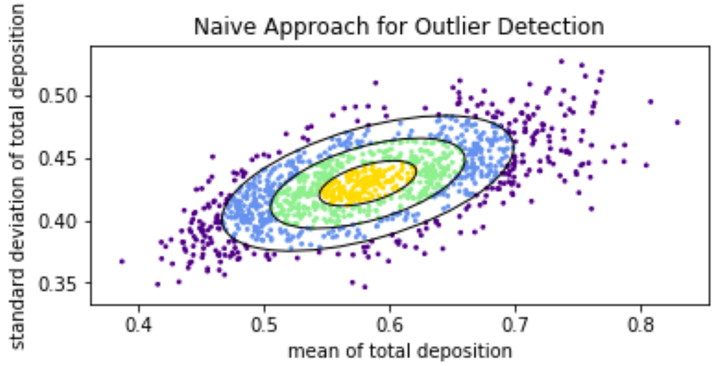
Reachability density is visualized through radii surrounding each point. This method results in incorrect predictions for only two of the points in this arbitrary data set.

2. Naïve Approach
   Algorithm Theory:
   - The simplest method to predict outliers is to determine the extremity of a point based off its distance from the mean in standard deviations.

Analysis



The increase in intensity of color along the color bar denotes a more negative LOF score which corresponds to the designation of a point as an outlier in the data. From this distribution of color, we can see that the predicted outliers lie on the edges of the data's distribution.



Each increase in ellipse size denotes an increase in standard deviation from the mean. Given the probabilities of being such a distance from the mean as stated by the Empirical Rule, any point beyond the third standard deviation ellipse would be considered an outlier.

Conclusion

- The analysis plots show that the naïve approach fails to accommodate the deviation of the data from the standard elliptical shape while the Local Outlier Factor algorithm is effectively able to model this variation.
- As a result, the Local Outlier Factor methodology is superior to the naïve approach with regard to accurate outlier prediction, thereby confirming that the LOF model has strong predictive power.
- The implementation of an LOF model into the process of determining the most effective weather models will help significantly reduce the time required to model a future release of an element without sacrificing the prediction capability.

Future Work

- Explore other models in search of improved explanatory ability for the given data set
- Use the most powerful of these outlier detection models to predict deposition values from the given categorical input

This work was supported by LDRD 17-ERD-045.

LLNL-POST-782204

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.c