

Date: November 18, 2019.
Author: Michael Goodliff.
Institution: Colorado State University - CIRA

Article: "Detection of non-Gaussian behaviour using Machine Learning Techniques"
Authors: Michael Goodliff, Steven Fletcher, Anton Kliwer, John Forsythe, Andrew Jones

Abstract:

An important assumption made in most variational, ensemble and hybrid based data assimilation systems is that all minimised errors are Gaussian random variables. There has been theory developed at the Cooperative Institute for Research in the Atmosphere (CIRA) that enables for the Gaussian assumption for the different types of errors to be relaxed to a lognormally distributed random variable. While this is a first step towards using more consistent distributions to model the errors involved in numerical weather/ocean prediction, we still need to be able to identify when we need to assign a lognormal distribution in a mixed Gaussian-lognormal approach.

In this paper, we present some machine learning techniques and experiments with the Lorenz 63 model. Using these machine learning techniques, we show detection of non-Gaussian distributions can be done using two methods; a support vector machine, and a neural network. This is done by training past data to classify 1) differences with the distribution statistics (means and modes) and 2) the skewness of the probability density function.

Description:

This repository contains eleven data files for the Lorenz63 model:

- testdata: contains the x, y and z variables test trajectory
- traindata: contains the x,y and z variables training trajectory
- skewdata9skew: contains the z-value and p-value on the training trajectory, window length of 9 (radius 4)
- skewdata17skew: contains the z-value and p-value on the training trajectory, window length of 17 (radius 8)
- skewdata25skew: contains the z-value and p-value on the training trajectory, window length of 25 (radius 12)
- MLPrediction9diff: Class values for the support vector machine and neural network, window length 9 points (radius 4), differences data classes
- MLPrediction17diff: Class values for the support vector machine and neural network, window length 17 points (radius 8), differences data classes
- MLPrediction25diff: Class values for the support vector machine and neural network, window length 25 points (radius 12), differences data classes
- MLPrediction9skew: Class values for the support vector machine and neural network, window length 9 points (radius 4), skewness data classes
- MLPrediction17skew: Class values for the support vector machine and neural network, window length 17 points (radius 8), skewness data classes
- MLPrediction25skew: Class values for the support vector machine and neural network, window length 25 points (radius 12), skewness data classes

Format of data files: .csv

Sponsors:

This work is supported by the National Science Foundation grant AGS-1738206 at CIRA/CSU.