Principal Component Analysis (PCA) for anomaly detection in time-series of gamma spectra

ELLINOORA VIKMAN AND NKS-PRICOMOB

Structure

1. Background
2. Goals
3. Theory
4. Methods
5. Summary

Background

- Time series of mobile gamma measurements exhibit significant variation due to the detector moving, such as passing different rock compositions with different U and Th concentrations.
- Time series of monitoring measurements exhibit significant variation due to changing environmental conditions, such as rain events or varying snow cover.
- The variations in the background cause problems for methods that analyze one spectra at a time.

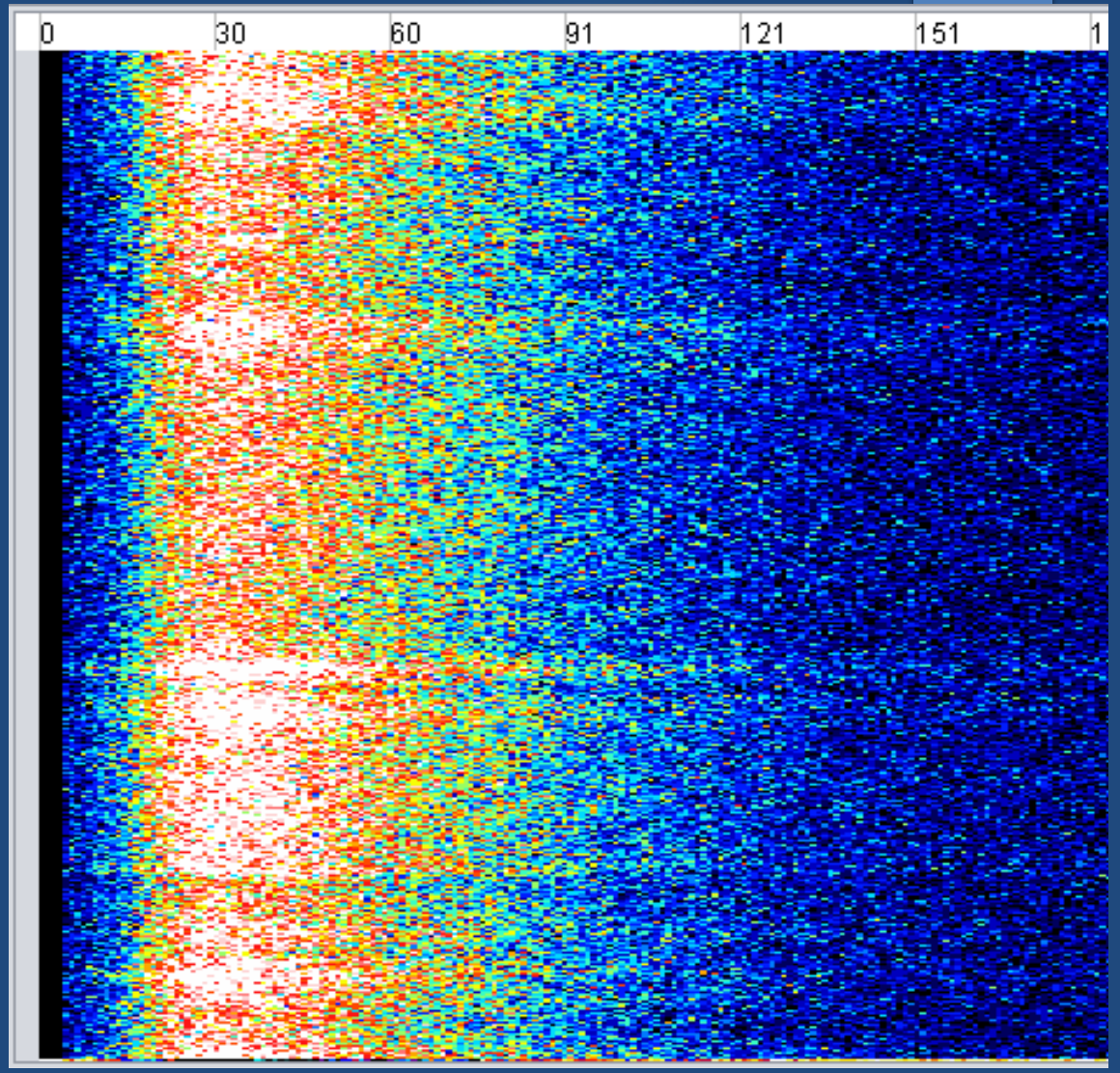

Fig 1: Waterfall display of background data measured by NaI-detector

The goal is…

- to reduce noise from time series.
- to trigger alarms only for relevant anomalies in time series, specifically those caused by artificial nuclides, while adapting to the changing background.
- to reduce the amount of Type I and Type II errors (false-positive and false-negative).

Theory

- PCA reduces dimensions of high-dimensional data by constructing a new set of variables that captures the most variation of the data.
- New variables (called as Principal Components) explain the variation in the order that first PC explains the most variation, 2nd PC second most etc..
- By omitting PCs from the end of the set, we can reduce the unnecessary information of the data (such as noise)
- kth PC is the linear combination of original variables $V_1, V_2, \dots, V_p$

$$PC_k = a_{k1}V_1 + a_{k2}V_2 + \dots + a_{kp}V_p,$$

where a_k is a constant vector

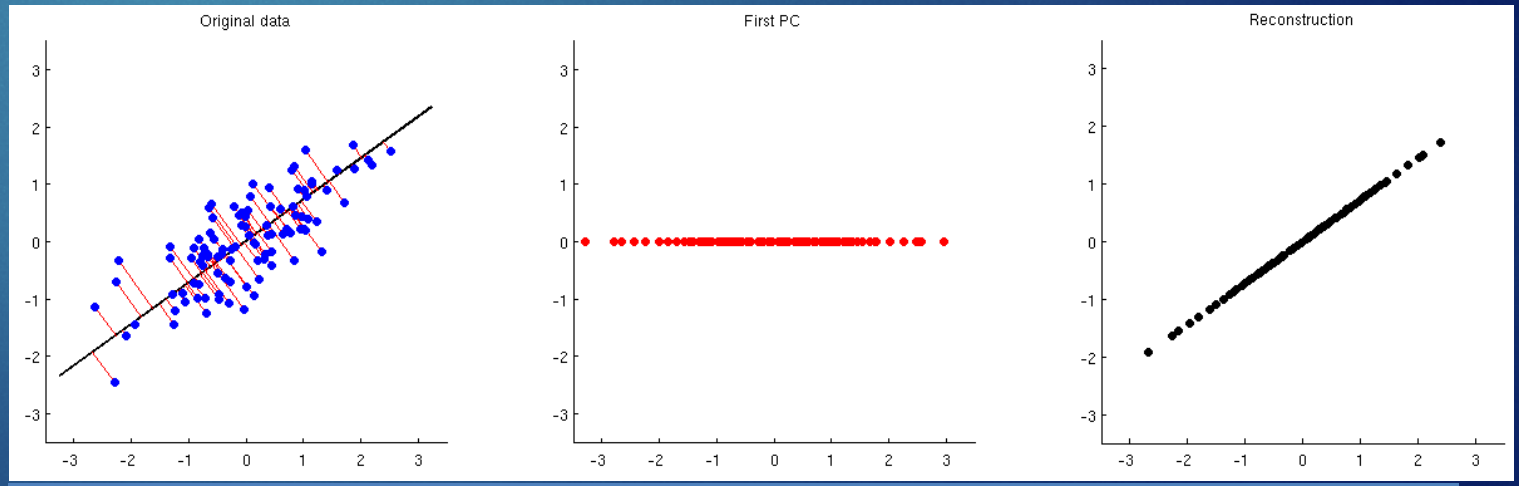

Fig 2: Illustrative image of principal components and PCA reconstruction [1]

Method

1. Implement PCA for background radiation data and take first few PCs → contains the most important part of background radiation

2. Reconstruct unknown spectrum with PCs

   → The reconstruction will explain the background part of the spectrum

4. Then calculate residual between unknown and reconstructed spectra

5. Trigger alarm if residual is above pre-defined threshold

Summary

- There are a lot of features in PCA that should be studied to optimize the method
  - i.e. the number of PCs, distributions of principal components and PCA reconstruction
- Also some preprocessing functions would help
  - for example channel multiplications, variance scaling and normalization
- Tests with simulated and synthetic spectra injected into measured time series suggest that these anomalies can be detected without the background causing excessive false alarms.
  - * The false alarm rate and limits of detection still need to be further studied.

References

- [1] Stack Exchange
- [2] I. T. Jolliffe, Principal Component Analysis, Second Edition
- [3] PRICOMOB DSA data sets