Online spatio-temporal anomaly detection using random forests

Six-month internship in statistics and applied mathematics

Subject

In many real-world applications, data are generated continuously over time, sometimes arriving only gradually or sporadically. More precisely, each observation can be viewed as a time series that progressively grows as new measurements become available.

The objective is therefore to perform continuous anomaly detection as new observations arrive over time, in order to detect abnormal measurements or behaviors as early as possible. Such a problem arises in a wide range of applications, for instance, in maritime surveillance, where the continuous monitoring of vessel trajectories can be used to identify unusual behaviors and support the detection of potentially illegal, unreported, and unregulated fishing activities. Similarly, in epidemiological surveillance, continuously updated information about virus mutations or infection trajectories can be monitored to identify unusual patterns and emerging events.

Traditional statistical methods that consist of building a model on the whole dataset, or building a new model each time new data arrive, are not well suited to this setting. Here, we aim to construct an anomaly detection model that is updated sequentially as new data arrive. Moreover, since the time series may be spatially correlated, the model should also take into account the spatiotemporal dynamics of the data.

Anomaly detection is an important topic in statistics, and a wide range of statistical methods have been developed to address this problem. One of the most widely used methods is the Isolation Forest (IF), introduced by Liu et al. (2008).

More recently, Levêque et al. (2026) introduced a new IF algorithm, referred to as the density-based Isolation Forest (dbiForest). This algorithm aims to overcome some of the main limitations of the original IF by integrating multivariate density estimation directly into the structure of isolation trees. Based on the assumption that outliers are both rare and different from normal observations, the algorithm identifies outliers as points located in low-density regions of the feature space.

In this internship, we aim to develop an online version of dbiForest that can handle time series that progressively grow as new observations become available and that may be spatially correlated.

An online version of dbiForest has already been developed. However, this algorithm assumes that the data acquired at each new observation are independent. It is therefore not well suited to the analysis of temporally dependent and spatially correlated data. The objective of this internship is thus to propose a new online version of dbiForest that is adapted to this setting.

The internship will begin by evaluating the performance of the online dbiForest algorithm on temporally dependent and spatially correlated data. Then, based on a review of ensemble methods suited to sequential data analysis and on the performance of the existing online dbiForest, the second part of the internship will focus on developing a new method adapted to this context. The proposed approaches will be implemented in Python.

All the methods investigated will be evaluated using both synthetic and real-world datasets.

The internship may lead to a PhD thesis.

Keywords : Density based isolation forest, sequential anomaly detection, spatio-temporal data.

Candidate Profile

We are looking for a motivated Master's student in Statistics (or equivalent) who :

— has a solid background in statistics, particularly in statistical learning ;
— has excellent Python programming skills ;
— is interested in methodological development, applied statistical research, and pursuing a PhD in Statistics ;
— has good scientific communication skills in French or English, including speaking, reading, and writing.

Interest in and knowledge of ensemble methods and/or spatio-temporal data analysis will be appreciated.

Practical Information

The internship is expected to last 6 months, with a start date between January and April 2027, depending on the student's availability.

The internship will take place at the laboratoire de Mathématiques de Bretagne Atlantique (LMBA), at Université Bretagne Sud, in Vannes.

The student will be supervised by François Septier (LMBA), Emmanuel Frénod (LMBA) and Audrey Poterie (LMBA).

Interested candidates are invited to send a CV and a cover letter to françois.septier@univ-ubs.fr and audrey.poterie@univ-ubs.fr by **20 December 2026**.

Références

Levêque, N., Frénod, E., Lacaille, J., Poterie, A., Septier, F., and Yahia-Bacha, M. (2026). dbiforest : an isolation forest based on density estimation.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *Eighth IEEE International Conference on Data Mining*.