21st International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
27-30 September 2022, Aveiro, Portugal

DETECTION OF LOW LEVEL CONCENTRATIONS OF HAZARDOUS MATERIALS IN THE AIR USING SEQUENTIAL MULTIVARIATE METHODS

Victor Watson[1,2], François Septier[1], Patrick Armand[2] and Christophe Duchenne[2]

[1]Univ Bretagne Sud, CNRS UMR 6205, LMBA, F-56000 Vannes, France
[2]CEA, DAM, DIF, F-91297 Arpajon, France

Abstract: Detecting weak signals in a noisy environment is a very important matter in many application cases relating to atmospheric dispersion or radiation exposure. Let us mention the detection of abnormal chemical concentrations in the air (due for instance to a leakage in a pipe network) or the detection of abnormal levels of radioactivity or concentrations of radionuclides in the air. When the environment is noisy, it is very difficult to monitor a variation in the measurement of a quantity (concentration, radiation…) corresponding to a threat before the potential hazard materialises. However, by accumulating the data over time and space using several sensors, it is possible to say if what is measured is likely to be noise or not. Being able to detect small variations early may be a way to react before the hazard occurs. Here we propose a new method especially designed to detect weak events characterised by surreptitious releases into the air transported and dispersed by the wind over a field equipped with sensors. We use the CUSUM (Cumulative Sum) as a base for our method and we extend it to multivariate cases and asynchronous monitoring. The extension to multivariate cases is not straightforward. Indeed, if some sensors only measure noise, they will decrease the overall signal to noise ratio and thus lower the probability of detection. We propose a way to prevent this by selecting the sensors that are most likely monitoring the signal. The non-synchronicity of the monitoring between the sensors using online statistical detection methods has never been addressed before our own work. The main assumption in the literature is that when a sensor starts monitoring an event, it will continue until detection. However, for releases spreading in the field, it is very possible that a sensor stops monitoring before some others do. Our method tackles the non-synchronicity issue while keeping the required computational power low enough for online detection. This method has been validated using academic data. Now, it will be tested in a realistic twin experiment consisting in the fictitious release of a low amount of radionuclides in an urban district where sensors are supposedly set up. Both the dispersion in the complex built-up environment and the radionuclides variously affecting the sensors make this configuration of particular interest as will be described in the paper.

Key words: *Detection, CUSUM, multivariate, Low-SNR.*

INTRODUCTION

Detection of weak signals in noisy data streams is a challenge in our societies' current effort to prevent pollution risks of any type. The releases of hazardous compound into the air can have dramatic effects if an appropriate response comes too late. In this work, we present a solution to detect weak signals in noisy time series. Based on the Cumulative Sum (CUSUM) technique (Page, 1954) and its adaptation to multivariate cases (Mei et al. 2010) and (Watson et al. 2022), we propose to use the developed test statistics on a twin experimental release of a hazardous compound in an urban area. In the second section, we describe the principle of sequential detection, the CUSUM and its extension to multivariate cases. More precisely, we discuss on how we can adapt the method to deal with space-sparsity (only some sensors are monitoring the signal) and time-sparsity (sensors do not monitor the signal at the same time). In the third section, we present the experimental release that generated the data we use to test the detection techniques in the fourth section.

SEQUENTIAL DETECTION

In order to detect signal in a low signal to noise ratio case, it is sometimes impossible to tell the signal contribution based on one measurement. Sequential detection gathers the information about the signal contribution along time and make it possible to be detected. The CUSUM is a technique where a likelihood ratio is recursively computed with each measurement so that when a change occurs in the noise distribution, this likelihood ratio rises sample after sample and reaches a threshold that cannot be reached (ideally) if the measurements only contain noise.

Let us consider a sensor which monitors a data stream x only withholding noise (following a distribution $f_0(x_k, \theta_0)$) until the change-point $k = \nu$ and then, after this change point, a signal added to the noise so the distribution becomes $f_1(x_k, \theta_1)$, then the likelihood ratio to determine if a change has likely occurred at time ν is:

$$\Lambda_\nu^n = \prod_{k=\nu+1}^{n} \frac{f_1(x_k,\theta_1)}{f_0(x_k,\theta_0)} \quad (1)$$

Because ν is unknown, it is necessary to generalize the likelihood ratio:

$$V_n = \max_{0 \leq \nu < n} (\Lambda_\nu^n) = \max_{0 \leq \nu < n} \prod_{k=\nu+1}^{n} L_k \quad (2)$$

Testing V_n with an adapted threshold allows determining whether there is a signal in presence. It is possible to compute the generalized likelihood ratio sequentially (Tartakovsky et al., 2014):

$$V_n = max(1, V_{n-1})L_n, n \geqslant 1, V_0 = 1 \quad (3)$$

Univariate detection

The CUSUM is obtained by using the log-likelihood instead of the likelihood in (3). This eases the computation especially for distributions from the exponential family.

$$W_n = Log(V_n) = max(0, W_{n-1}) + log(L_n), n \geqslant 1, W_0 = 0 \quad (4)$$

The CUSUM can be used to detect any kind of change in the distribution of a data-stream but it is often a change in the mean or in the standard deviation that one seeks to find. In our application case, a change in the concentration of a radionuclide in the air is consistent with an increase in the monitored mean value. **Figure 1** shows an example of how the CUSUM works on univariate data.

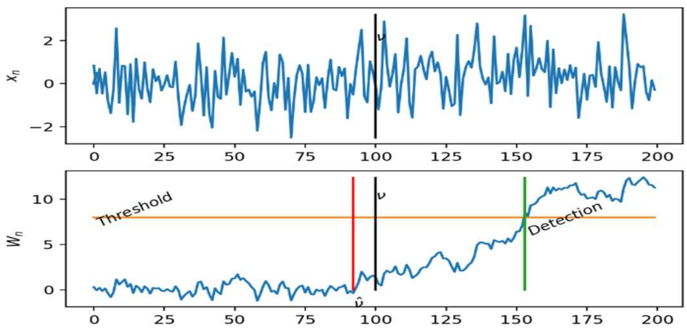

Figure 1. Detection with CUSUM of a change in the signal mean with a low signal to noise ratio (SNR). The change in the mean occurs at the time marked ν. On top, the evolution of the monitored variable x does not allow to detect directly the change in the mean. The CUSUM variable however changes its behaviour as the change-point has been passed and increases until it reaches the detection threshold. It is also possible to estimate the change-point $\hat{\nu}$ from the last time W_n crosses 0.

Multivariate detection

In the application we consider, many sensors are monitoring the area. The easiest way to associate the sensors is to construct the global test variable as the sum of all the local test variables. This technique is called SumCUSUM (Mei et al. 2010). Because the signal can be spatially sparse (only a few of the sensors monitor the signal), Mei et al. (2010) propose a censoring technique based on a rough knowledge of the signal energy establishing local thresholds. This would only sum the local variables where a signal is most likely present and increases the performances of the methods. Watson et al. (2022) propose a relative threshold technique so that the prior knowledge of the signal energy is no longer necessary. The global variable of this censored-SumCUSUM is defined as:

$$T_{cSC}(n) = \frac{1}{\sum_{l=1}^{L} 1_{W_{n,l} \geqslant c_n}} \sum_{l=1}^{L} W_{n,l} \geqslant c_n, \tag{5}$$

With:

$$c_n = \alpha \max_{l \in [1,\ldots,L]} \left(W_{n,l}\right) \tag{6}$$

The case of a time sparse problem has been considered in Watson et al (2022). Indeed, in the case of an event passing over a field of sensors, it is possible that the monitoring between the sensors is not synchronized. Bringing together these time limited exposures is possible while keeping online computation using the test variable presented hereafter:

$$T_{TESC}(n) = \frac{1}{L} \sum_{l=1}^{L} \max_{[\nu_l, N_l]} \sum_{k=\nu_l}^{N_l \leq n} log \frac{f_{1,l}(X_{k,l}, \theta_{1,l})}{f_{0,l}(X_{k,l}, \theta_{0,l})} \tag{7}$$

Watson et al. (2022) show that it is equivalent to rewrite local variables such that G_n acts as the memory of the last highest value of W_n:

$$G_n = max(G_{n-1}, W_n) = \max_{0<k \leqslant n} W_k \tag{8}$$

Thus, it is possible to simplify (7) so that:

$$T_{TESC}(n) = \frac{1}{L} \sum_{l=1}^{L} G_{n,l} \tag{9}$$

These techniques can be combined in order to be used in both time and space-sparse problems:

$$T_{CTESC}(n) = \frac{1}{\sum_{l=1}^{L} 1_{G_{n,l} \geqslant c_n}} \sum_{l=1}^{L} G_{n,l} \geqslant c_n, \tag{10}$$

TWIN EXPERIMENT OF AN INSIDIOUS RELEASE

Having available data is of considerable importance if we want to test our method of finding signals with low signal-to-noise ratios. In our application, these data are the volumetric concentrations on sensors after the transport and dispersion of a release from a point source. One can naturally think of two types of data. The first type would correspond to the real situation of a release whose source goes unnoticed and which leads to sparse detections in space and time. Although this situation certainly exists, we do not have such data. The second type could be data acquired under full-scale experimental conditions or in a wind tunnel. Nevertheless, these data would be complicated and expensive to obtain because of the need to find an available site, devices to release the tracer and measure the signals, the workforce, etc. In the absence of a real dispersion situation available, we had the idea of considering a twin experiment, i.e. to generate synthetic signals from CFD computations.

The Parallel Micro-SWIFT-SPRAY (PMSS) modelling system has been used to carry out the simulations. It is the parallel version of Micro-SWIFT-SPRAY (MSS) (Tinarelli et al. 2013) which was developed to provide a CFD solution of the flow and dispersion in the atmospheric environment in a limited amount of time. MSS is composed of the high-resolution local scale versions of the SWIFT and SPRAY models:

- SWIFT is a 3D diagnostic mass-consistent model using a terrain-following coordinate. Large scale meteorological data, local meteorological measurements, and analytical results in building-modified flow areas are interpolated and adjusted to generate 3D wind fields. Other meteorological data such as temperature or humidity are also interpolated. Eventually, the turbulent flow parameters are computed by SWIFT to be used by SPRAY.
- SPRAY is a Lagrangian particle dispersion model (LPDM) able to take into account the presence of obstacles. The dispersion of the release is simulated by following the trajectories of a large number of fictitious particles. Trajectories are obtained by integrating in time the particle velocity, which is the sum of a transport component defined by the local averaged wind generally provided by SWIFT, and a stochastic component, representing the dispersion due to atmospheric turbulence.

Both SWIFT and SPRAY can handle complex terrains and changing meteorological conditions, as well as specific release features, such as heavy gases. More recently, SWIFT and SPRAY were parallelized across time, space, and numerical particles, resulting in the PMSS system (Oldrini et al. 2017). The parallelism was shown to be very efficient, both on a multi-core laptop and on clusters of several hundreds or thousands of cores in the case of a high-performance computing centre (Oldrini et al. 2019) (Armand et al. 2021). PMSS was systematically validated over numerous experimental wind tunnel and field campaigns for both short and long releases (Trini Castelli et al. 2018). In all configurations, the PMSS results comply with the statistical acceptance criteria defined by Hanna and Chang (2012) used for validating dispersion models in built-up environments.

In the twin experiment, we used meteorological data produced by the AROME model of Météo France at a resolution of 0.025° (around 2.5 km at mid-latitude). Meso-scale weather predictions were downscaled with the PMSS modelling system in order to zoom in the central urban district south of Republic Square, in Paris (France). More precisely, the AROME meteorological profiles were extracted at the four closest points to PMSS simulation domain to nudge the flow simulations, also explicitly accounting for the whole buildings of the urban district. The 3D urban domain has dimensions of 1,100 m x 1,100 m x 1,600 m and 551 x 551 x 34 grid nodes. The horizontal mesh size is 2 m. The vertical mesh size is 2 m up to 22 m and becomes looser going up to the top of the domain. **Figure 2** shows the horizontal cross-section of the urban domain near the ground and the buildings. During the time sequence of the twin experiment, meteorological conditions varied, notably the wind speed and wind direction as indicated in **Table 1**, The fictitious source is denoted by the red dot in **Figure 2**. It was supposed to release an amount of 10^9 Bq of a radioactive tracer (gas or fine particles) at a height of 2 m for 20 minutes. Virtual sensors were also supposed to be set up at the same height as the source in the urban district. They are shown as blue dots in **Figure 2**. Dispersion simulations were carried out for 45 minutes. The results of these simulations are volumetric concentrations averaged over 1 minute and output each minute. They are used to test the detection techniques discussed in this paper.

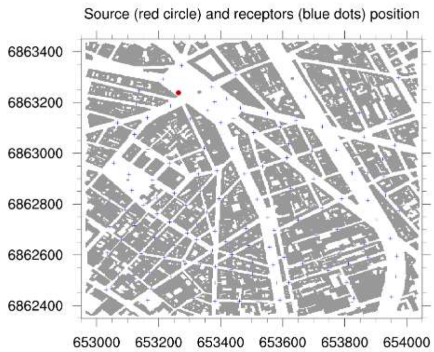

Figure 2. Horizontal cross section of the simulation domain showing the buildings of the urban district south of the Republic Square in Paris (France) and the positons of the fictitious source (red dot) and virtual sensors (blue dots).

Table 1. Evolution of the wind speed and wind direction at 10 m (AROME results not considering the effect of the buildings) during the time sequence considered for the PMSS simulation (t_0 designates the instant of the fictitious release).

Time	Wind speed	Wind direction
t_0	2.6	287
t_0 + 9 min	0.95	327
t_0 + 18 min	1.05	5
t_0 + 27 min	1.8	17
t_0 + 36 min	1.0	28
t_0 + 45 min	1.9	25

SEQUENTIAL DETECTION ON EXPERIMENTAL RELEASE

In order to test the different detection techniques, we will consider a system with 25 sensors. Among these 25 sensors the main proportion only holds noise where 0 to 10 data streams hold a signal. These signals have been randomly picked out of 12 different signals extracted from the twin experiment. In the following results, all the methods are tuned to have a probability of false alarm of 1 to 1,000 during the experiment. The noise on each data-stream is Gaussian and follows the rule $N(0,1)$. Every detection rate value is evaluated with 10,000 runs with random noise generation and random sensors selection for each run.

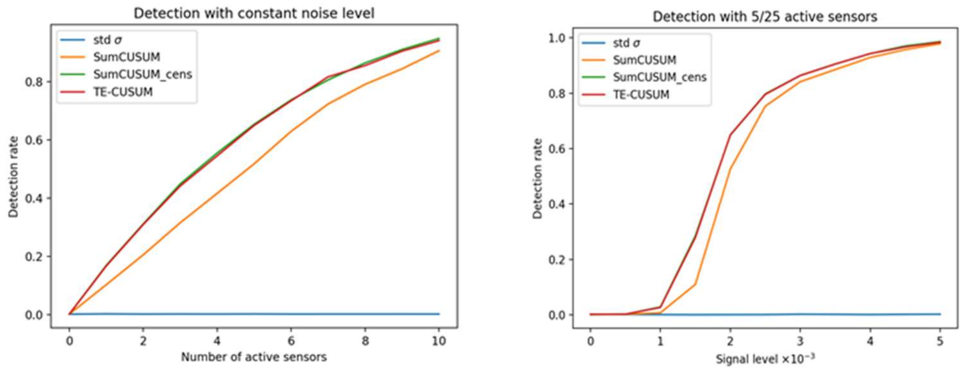

Figure 3. Detection rate depending on the number of affected sensors (left) and depending on the signal to noise ratio (right). The lower line in blue shows the results using a simple threshold solution on the data, the orange one the detection rate of the standard SumCUSUM, the green one the results taking into account the space sparsity, and the red one the results when both time and space sparsity are considered.

Figure 3 (left) shows the detection rate of each method (except for the standard threshold method) with a constant signal to noise ratio and a variable proportion of affected sensors. The probability of detecting a signal computed empirically over the 10,000 runs increases with the number of sensors monitoring a signal. It is also notable that the consideration of space sparsity improves the results over the simple SumCUSUM. Of course, this improvement reduces if the number of sensors affected by the signal is close to the total number of sensors, but such a case is unlikely in the considered application. The consideration of both space and time sparsity gives the same results than considering only space sparsity because in this experiment, there is a lot of overlapping in the exposed sensors.

Figure 3 (right) shows the detection rate of each method (except for the standard threshold method) with a constant number of impacted sensors and a variable signal to noise ratio. When the signal gain decreases, the consideration of the space sparsity shows a lesser degradation of the detection rate than the simple SumCUSUM. Again, the addition of the time sparse consideration changes quite nothing because of the data structure.

When all signals are synchronized, the test variable considering time sparsity does not decrease the detection rate. This indicates that adding the sensitivity to time differential events does not conflict with the censoring technique. An experiment where sensors monitoring is not overlapping will be necessary to show this last technique interest. This could be the case for a more time-localized event traveling quicker in space.

CONCLUSION

In this work, we have shown the interest of using sequential detection technique to quickly identify the variation of a signal characterized by a low signal-to-noise ratio. This signal could be the concentration of a hazardous compound transported and dispersed into the air. Because these techniques can detect weak signals, they can be used as powerful prevention tools. We presented various techniques considering sparsity in space when not all the sensors monitor the signal but also sparsity in time when the sensors do not monitor the signal at the same time. The results on the experimental data presented here show how powerful these techniques are even though time sparsity was not present in the experiment we used. This time sparse technique has shown good results on simple synthetic data in Watson et al. (2022). It will need a different experimental data set to be proven worthy. Besides using these techniques on more complex data set, our future work will show its efficiency on different noise distributions such as Poisson, which is more realistic for some of the detectors used in the targeted applications.

REFERENCES

Armand, P., O. Oldrini, C. Duchenne and S. Perdriel, 2021: Topical 3D modelling and simulation of air dispersion hazards as a new paradigm to support emergency preparedness and response, *Environmental Modelling & Software*, 143, 105129.

Hanna, S. and J. Chang, 2012: Acceptance criteria for urban dispersion model evaluation. *Meteorological Atmospheric Physics*, 116, 133–146.

Mei, Y., 2010: Efficient scalable schemes for monitoring a large number of data streams. *Biometrika*, 97 (2), 419–433.

Oldrini, O., P. Armand, C. Duchenne, C. Olry, J. Moussafir and G. Tinarelli, 2017: Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas, *Environmental Fluid Mechanics*, 17 (5), 997–1014.

Oldrini, O., P. Armand, C. Duchenne and S. Perdriel, 2019: Parallelization performances of PMSS flow and dispersion modelling system over a huge urban area, *Atmosphere*, 10 (7), 404.

Page, E.S., 1954: Continuous inspection schemes. *Biometrika*, 41 (1-2), 100–115.

Tinarelli, G., L. Mortarini, S. Trini Castelli, G. Carlino, J. Moussafir, C. Olry, P. Armand and D. Anfossi, 2013: Description and preliminary validation of the PMSS fast response parallel atmospheric flow and dispersion solver in complex built-up areas. American Geophysical Union (AGU) 200 (A), 311–327.

Skamarock, W.C. et al., 2008: A description of the Advanced Research WRF Version 3. No. NCAR/TN-475+STR, University Corporation for Atmospheric Research, Boulder, CO, USA, p. 113.

Tartakovsky, A., I. Nikiforov and M. Basseville, 2014: Sequential analysis: Hypothesis testing and change point detection, p. 111.

Trini Castelli, S., P. Armand, G. Tinarelli, C. Duchenne and M. Nibart, 2018: Validation of a Lagrangian particle dispersion model with wind tunnel and field experiments in urban environment, *Atmospheric environment*, 193, 273–289.

Watson, V., F. Septier, P. Armand and C. Duchenne, 2022: Sequential detection of a temporary change in multivariate time series. *Digital Signal Processing*, 127, 103545.