On obtaining information about the structure and quality of investigated objects revealed from streams of data through the method for identifying structures

N A Antonyuk\(^1\), S G Kataev\(^2,3\), S S Kataeva\(^4\), H Kelderman\(^1\), M Klein Wolterink\(^1\), A A Miller\(^5\)

\(^1\) Staal Technologies b.v., Eindhoven, the Netherlands
\(^2\) Tomsk State Pedagogical University, Tomsk, Russia
\(^3\) Institute of Monitoring of Climatic and Ecological Systems, Tomsk, Russia
\(^4\) Tomsk State University, Tomsk, Russia
\(^5\) St. Petersburg Mining University, St. Petersburg, Russia

sskataeva@sibmail.com

Abstract. This work is devoted to the study of the possibility of using the Method of Identifying Structures in solving important practical problems. The first task is to recognize the type of raw material on the moving conveyor belt based on the reflected radar signal. The second is to determine the structure of the subsurface layer based on data obtained by gamma-ray logging and borehole measurements. The conducted research shows the effectiveness of the proposed method and possibility to obtain more information about the quality and state of the object by not only looking at a single measurement, but instead looking at consecutive measurements as on a data stream.

1. Introduction

Many measurements of random variables are a temporal, spatial, or other sequence of values and can be written as an ordered series. Let’s call these measurements a data stream. A data stream can be interpreted as known from the theory of Queueing event flow. All events are equal, and each event is characterized only by the moment it occurs. The intensity of flow is the frequency of occurrence of an event or the number of events per unit of time. For the simplest Poisson flow, the intensity is approximately the same for any of its sections. A more complex flow controlled by a Markov chain (MC flow) can be represented as a sequence of Poisson flows with different intensities. The APFLOW method described in this paper – a modification of the structure detection method (MRS) [1] – was developed to solve one of the main tasks of the Queueing system – determining the characteristics of the incoming event flow. APFLOW allows you to approximate the flow of random events by a MC flow, i.e. describe its structure. This article presents some results obtained using an adapted version of the APFLOW method for some tasks with data streams of different nature.
2. APFLOW Algorithm

The problem of approximating an arbitrary flow of $N$ events by an MC flow reduces to finding its structure i.e., such moments of time $\tau_p$, that the flow of events arriving for each interval $\tau_{m+1} - \tau_m$ would be approximately Poisson. It is customary to call such intervals intervals of stationarity (IS).

APFLOW is easy to adapt to the task of finding uniform intervals on a flow data. In fact, the fingerprints of signals are a series of random numbers $n_1, n_2, ..., n_m$ and they can easily be represented as a stream of moments of occurrence of random events, if we assume that the quantities $n_1, n_2, ..., n_m$ are analogous to the differences between the moments of arrival of neighboring events. Thus, the task is to define the boundaries between IS, that is, find the numbers of boundary events $m_j, j = 1, ..., k$, separating $k+1$ IS. To find IS it is necessary and sufficient to build a matrix of relations between objects. The objects in the problem are the numbers of spatial points $x_i$, $i = 1,..., N$, and the intensity values i.e., values of the information fingerprint of the signal $t_i, i = 1,..., N$. In order for the flow to be ordered, the values of $t_i$ are constructed recursively, i.e. $t_i = t_{i-1} + n_e$.

As a characteristic of the relationship between objects (events) $i$ and $j$, it is convenient to choose the “distance” between them i.e., the quantity $d_{ij}:

$$
d_{ij} = \begin{cases} 
(j_{j+1} - t_{i-1})/(j - i + 1), & i \leq j, j, i = 1,..., N, \\
0, & i > j, j, i = 1,..., N.
\end{cases}
$$

The elements $d_{ij}$ defined in this way form an upper triangular matrix $D = \left[d_{ij}\right]_{N \times N}$. We associate with each $i^{th}$ event of the flow, $i = 1,..., N$, a vertex with the same number $i$ of some graph $G = (V, E)$, where $V$ is the set of vertices of the graph of cardinality $N$, and $E$ is the set of pairs of vertices i.e., set of edges of the graph. The cost of each edge of the graph connecting the $i^{th}$ and $j^{th}$ ($i < j$) vertices is the quantity $d_{ij}$ i.e., the corresponding element of the matrix $D$. The cardinality of the set of edges of the graph is $N(N-1)/2$. Further, since MRS (and APFLOW) look for predefined structures on the graph, it is necessary to determine what is the structure in this problem. The definition of the structure can be formulated based on the concept of the IS. Namely, if a group of events with numbers $i_1^p, i_2^p, ..., i_k^p$ makes up a stationarity section of a stream with intensity $\lambda_p$, then this means that

1. the average value of the costs of all the edges connecting the vertices with the numbers $i_1^p, i_2^p, ..., i_k^p$, is approximately equal to the intensity $\lambda_p$:

$$2(k(k-1)) \sum_{(i,j) \in I} d_{ij} = \lambda_p$$

where $I$ is the set of all kinds of pairs ordered by numbers from $\{i_1^p, i_2^p, ..., i_k^p\}$;

2. numbers $i_j^p, j = 1,..., k$, go in a row, i.e., $i_j^p = i_{j-1}^p + 1$;

3. all $d_{ij}, i, j \in I$, belong mainly to a certain interval $\left(\lambda_p - \Delta \lambda, \lambda_p + \Delta \lambda\right)$.

It should be noted that the intensity estimate obtained for the entire group of events should give the most accurate estimate $\lambda_p$.

The above considerations define the idea of an algorithm for allocating IS in a flow. If you select an interval $\left(\lambda_p - \Delta \lambda, \lambda_p + \Delta \lambda\right)$ and remove edges with costs that do not belong to this interval from the graph $G$, the graph is divided into connected components $G_i = (V_i, E_i)$. If the interval is chosen correctly, these components correspond to the IS related to the $p^{th}$ state of the MC flow. $G_i$ is a subgraph of graph $G$. It is reasonable to expect that the IS must corresponds to a subgraph in which, firstly, the maximum number of edges, and secondly, each vertex must be connected to the largest
number of vertices. After selecting several ISs, the matrix $D$ is cleared of “noise” by removing those elements whose refer to different ISs.

The method is iterated. At the first iteration, the matrix $D$ is scanned and all ISs that unconditionally satisfy the structure requirement are found. Then, from the graph, all edges connecting the vertices belonging to different ISs are removed, and the next iteration is performed. The quality of the partition is estimated by calculating the degree of reliability of the correspondence of the found ISs with Poisson using the Pearson criterion in the presence of a sufficient number of events in the IS.

In the following statements (in the examples discussed below), we use the concept of the IS, meaning the boundaries of the selected homogeneous sections

3. Conveyor belt

There is a certain need for obtaining information on the quality of media that are being monitored. One such an application would be monitoring of road conditions; snow, ice, water film, and grip [2], [3]. Furthermore, to further moisture control systems in products where a drying process is involved such as food, chemicals, paper, minerals, textiles, etc. A sensor placed above the conveyor belt such a device can be invaluable, giving operators realtime information on the moisture content of the product [4]. A more far fetching application would be to do remote sensing of a ‘plastic soup collection systems’ [5] which can be done with Thermal sensing [6]. Further experiments are needed to find out if the latter is possible with current system. Here we are presenting the method, algorithms involved and experiments mimicking a conveyor belt. The task of on-line monitoring by non-invasive methods of the state of the material moving along the conveyor belt is currently being solved in different ways [7], [8] and [9]. In this paper, we propose using the signals of the integrated radar system reflected from the surface and the APFLOW algorithm [1] as a data processing method for this purpose. The APFLOW algorithm was previously developed by the authors to determine the structure of a flow of homogeneous random events or to approximate it by a Markov Chain flow (a flow with switching) [10]. Adapted APFLOW variants were used to solve a variety of problems to interpret the data of downhole research using the gamma-ray logging method [11]. When moving the conveyor belt, the signal reflected from the objects located on it changes randomly, both due to fluctuations in the distance between the radar and the surface of the object, and due to a change in the state of the object. If the conveyor belt moves at a constant speed $v$, and a radar placed above it at a certain height captures the reflected signals after a time $\Delta t$, then we have information at points $x_1$, $x_2$, ..., the distances between which are $\Delta x = v\Delta t$.

To use the APFLOW method, it is necessary for each measurement at a spatial point to associate one value, which will be a sort of fingerprint of the signal, we will refer to this as information fingerprint or just fingerprint. Then the values of this attribute can be represented in the form of a time-ordered series of numbers $n_1, n_2, ..., n_M$, which we will call the “measurement flow”. Even for a homogeneous medium, due to the presence of noise, this stream will be a sequence of random numbers oscillating around its average value. In the case where sections containing different materials (or different states of the same material) fall into the stream, fluctuations will occur near different average values. To find an information fingerprint, experimental studies were conducted in which the signals reflected from different media were studied. Studies have shown that their extreme values can act as information fingerprints of reflected signals [12], [13] and [14]. The received reflected signals from different media in a different number of spatial points allowed us to create a variety of data streams, modeling this or that real situation. As an example, figure 1 shows the maximum and minimum (in absolute value for ease of comparison) values of signals received from different media for flow $F_1$.

It can be seen that the extremes behave synchronously; therefore, either its maximum, or minimum, or their difference can be taken as an information fingerprint of the signal. It is clear that the average value of the information feature in each homogeneous section depends on the physical properties of the medium, and its variability depends on the degree of surface roughness.
3.1 Experimental setup

Staal Technologies’ RIC60A has been used as a radar. RIC60A is a 60 GHz BiCMOS single-chip millimeter-wave frequency modulated continuous wave (FMCW) radar with on-chip integrated antennas with analog and digital outputs [15].

In order to mimic a conveyor belt with changing objects a radar module has been scanned along changing material surfaces. At each point the radar is stopped and a stationary measurement is taken. Note that typical speeds would be limited in comparison to the chirp-time in any case.

![Experimental setup diagram]

**Figure 2.** Experimental setup.

Radar is mounted under a cart that can be moved on a rail. Radar is moved along rail and has a beam angle of about 30 degrees. Measurements are taken at constant intervals of along the horizontal at stationary points.

The absorber on the left and the concrete floor on the right provide a baseline to the measurements done on the materials laid in a tray of uniform size. The object material in the tray gave such a uniform response that it is warranted to use sections of changing materials that give the impression of changing materials under the radar. The time scan of a typical reflected (IF-) signal (in this case, from a concrete...
surface) is shown on the left part of figure 3. A range profile can be obtained via Fast Fourier Transformation (FFT) as is shown on the right part of figure 3.

**Figure 3.** Time scan of typical signal reflected from the surface (on the left section) and range profiles of concrete and sand (on the right section). Concrete is at 47 cm from the radar, and sand is at 36 cm.

### 3.2. Measurement

Several data streams were generated from radar measurements for different objects and then processed by APFLOW. And as an example, figure 4 shows the flow of model data $F_2$ of fingerprints composed of various media (concrete floor, empty tray, sand and sawdust of various degrees of moisture, etc.). The solid line shows the result of the APFLOW – found ICs.

**Figure 4.** The behavior of the maxima (dashed line) for the model flow $F_2$. Solid line – average values (intensities) of the selected ISs by the APFLOW algorithm.

### 4. Interpretation the data of downhole research using the gamma-ray logging method

One of the basic applications of the method of gamma-ray logging (GL) is the lithological dissection of the section, that is, the selection of lithological boundaries in the section with reference to their depth. In this case, as a rule, it is assumed that the cut is composed of horizontal (normal well axis) layers with clear boundaries within the layers of rock properties do not change. In addition, it should be kept in mind, that the flow of gamma rays reaching the detector device in the well, is a random process. If we assume, that by a random variable we mean the number of pulses from the output of the recording device per unit time, then these values $n_1, n_2, ..., n_m$ will be distributed according to the Poisson's law. Thus, when advancing the device through the borehole, i.e. when carrying out
measurements of the number of pulses at the output of the photomultiplier tube in a second changed as a result of the random nature of the processes of decay of radioactive isotopes in the rock, and because of the advancement of the instrument through the borehole. In this case, at each subsequent time the device is in a new position in the well against rocks that generally have a different content of radioactive isotopes. Thus, logging measurements are a series of random numbers and they can be easily represented as a flow of random events, if we assume that the values are the differences of the “arrival times of adjacent events.”

Given the properties of the Poisson flow, namely stationarity and homogeneity, it can be argued that these intensities should be approximately the same within one layer. This means that the density distribution of values $n'_i, n'_j, \ldots, n'_w$, where $j$ is the layer number will not be a Poisson distribution.

And it’s clear that the more time the intensity $n'_i, n'_j, \ldots, n'_w$, is determined the closer these values are to each other and the smaller the variance of this series should be. These considerations allow us to formulate a criterion for determining the boundaries of layers, based on the idea of minimum dispersion for $n'_i, n'_j, \ldots, n'_w$, belonging to one layer $m_j, j = 1, \ldots, k$, where $k$ is the number of layers separating the various “stationary intervals”. And here we keep this concept, understanding under the IC sequence of consecutive numbers of the series, the intensities at which are approximately the same, and the total variance of the intervals thus allocated is minimal.

4.1 Model experiment

The method was first tested on data obtained from targeted laboratory experiments [11] and then applied to the interpretation of field data. To conduct this experiment, the logging probe was connected to a digital registration system and the number of pulses coming from the device in 5000 consecutive intervals of 1 second was recorded. At the same time, the radiation background recorded by the detector of the device was changed with the help of an exemplary source of gamma radiation, which was located at various distances from the detector of the device. This distance changed after every 1000 points were recorded. Thus, 5 sequences of points of 1000 pieces were recorded; within each sequence the radiation background did not change. All 5000 points were recorded without interruption. As a result of the algorithm’s work, 5 intervals of stationary, related to 4 states, were highlighted. The final boundaries of the layers were obtained by applying the principle of minimum dispersion to a pair of adjacent layers, starting with the first. As shown in figure 5, the method very well reproduces the layer boundaries for laboratory data.

![Figure 5](image-url)
4.2 Real data
The logging of a well with a depth of 31 m, drilled for the purpose of water extraction in a suburban area of one of the districts of the Leningrad region, was carried out. The logging of the GL was performed using a downhole device and developed ground equipment that allows measurements of the number of pulses coming from the downhole device for a time interval of 0.63 s. These data are recorded in a file. Also in this file depth marks are recorded every 5 m. Then a software linking of each reference depth, with the movement of the device from mark to mark is assumed to be uniform. The lifting speed is approximately 200 m/h.

The data obtained were processed using APFLOW. Baseline data and treatment results (IS boundaries and average values of signals on these IS) are presented in figure 6. The well at the site was not drilled, but the results of measurements made by other methods (apparent specific electrical resistance method with potential and gradient probes) allow to approve with confidence that the spacing limits are sufficiently precise. Figure 6 also shows not only agreement between different experiments, but clarity in the allocation of boundaries.

![Figure 6. Initial measurement data and processing results for real data. The solid line is the original data of gamma-ray logging (GL), and the dashed line is the intensity of the final version of the division into layers.](image)

5. Conclusions
1) At the first iteration, the APFLOW method distinguishes well the boundaries of objects in the presence of sufficient contrast (about 10%) in the amplitude of the reflected signal of the contacting media. The number of layers with sufficient contrast does not affect the quality of the layer selection.

2) Clarification of the boundaries, if necessary, requires a second and, possibly, subsequent iterations. The first iteration can be performed in automatic mode, the subsequent ones in interactive mode.

3) The degree of detail of the allocation of ICs is regulated in the APFLOW method by controlling parameters, which makes it possible to isolate fairly short ICs.

4) A methodology has been shown extract more information from a radar-signal about the state of the object apart from distance and velocity.
5). It should be noted that the program that implements the described algorithm finds the initial partition of the flow into the intervals of stationary in automatic mode. Clarification of the boundaries of the layers is made by the interpreter in an interactive mode.

References
[1] Kataev S G 2012 Algorithmic and software of analytical and noninvasive control of parameters of the environment based on a method of detection of structures Ph. D. dissertation (Tomsk: National research Tomsk Polytechnic University) p 360
[2] Ng J R, Wong J S, Goh V T, Yap W J, V. Yap T T and Ng H 2019 Computational Science and Technology (Singapore) 481 (Singapore: Springer Singapore) pp 259-68
[3] Arbapour Bidgoli M, Golroo A, Sheikhzadeh Nadjar H, Ghelmani Rashidabad A and Ganji M R 2019 Automation in Construction 104 pp 140–52
[4] Khamtree S, Ratanawilai T and Nuntadusit C 2019 Drying Technology vol 37 no 16 pp 2116-25
[5] Serranti S and Bonifazi G 2019 Techniques for separation of plastic wastes Use of Recycled Plastics in Eco-efficient Concrete Elsevier pp 9–37
[6] Goddijn-Murphy L and Williamson B 2019 Remote Sensing 11 no 18 p 2159
[7] Bartelt T L M 2011 Industrial automated systems: instrumentation and motion control (Clifton Park, NY: Delmar/Cengage Learning) p 746
[8] Irudayaraj J and Reh C Eds 2007 Nondestructive Testing of Food Quality (Oxford UK: Blackwell Publishing Ltd)
[9] Murugesan R 2006 Evolution of industrial automation Int. J. of Comp. Appl. in Technol. 25 no 4 p 169
[10] Bekkerman E N, Kataev S G and Kataeva S S 2013 Automation and Remote Control 74 no 9 pp 1449–59
[11] Kataev S, Kataeva S and Miller A 2019 On a New Method of Detecting Layer Boundaries at the Borehole Measurements by Gamma Logging The Hague [Online]. Available:https://www.earthdoc.org/content/papers/10.3997/2214-4609.201902433
[12] Obeid D, Sadek S, Zaharia G and El Zein G 2009 A tunable-frequency system for touch-less heartbeat detection and HRV extraction Romania: IEEE [Online]. Available:http://ieeexplore.ieee.org/document/5206132/
[13] Karrenberg U 2002 An Interactive Multimedia Introduction to Signal Processing. (Berlin, Heidelberg: Springer Berlin Heidelberg) p 412
[14] Schultz W 2015 Physiological Reviews 95 no 3 pp 853–951
[15] Adela B B, van Zeijl P T M, Johanssen U and Smolders A B 2016 IEEE Transactions on Antennas and Propagation 64 no 8 pp 3281–91