Invariant to Observation Conditions, the Algorithm for Processing Spatially Distributed Data from a Monitoring Network Consisting of Three Sensors

E Chernetsova¹, V Burlov¹, A Shishkin¹

¹Faculty of Information Systems and Geotechnology, Russian State Hydrometeorology University, Voronezh st. 79, St-Peterburg, 192007, Russia

E-mail: chernetsova@list.ru

Abstract. A multi-sensory environmental monitoring system is being considered, in which pollution data is transmitted via radio from three sensors. It is assumed that the centralized processing of information at the point of observation and control in order to detect a signal indicating the presence of contamination. The proposed processing algorithm is invariant to any transformations of data coming from sensors. It keeps the property of the data to be equally or differently distributed. Such transformations are any non-monotonic functions of the original observations. It is shown that with respect to these transformations, the maximum invariant statistics is the set of ranks of quantities taken as elements of a single sample. The developed detection algorithm has a sufficiently high resistance to observation conditions, since it retains its working capacity when the observation conditions are non-stationary. The algorithm is also invariant to the law of fluctuations of the pollution signal and its location among the last two observed sites and provides a constant probability of false alarm in any (a priori unknown) noise distribution.

1. Introduction

To monitor the environment, it is necessary to conduct continuous observations in time, based on a thoughtful distribution of measuring means in space, for which it is necessary to use a stationary distributed multi-touch remote monitoring system [1]. It should work quickly, preferably in real time [2]. The stationary network of stations included in the monitoring system requires the availability of communication channels with the Monitoring and Control Point (MCP) [3]. Laying the cable network connection is often unprofitable. Therefore, for communication purposes, it is necessary to use a radio channel or satellite communications [4]. Since the monitoring network sensors receive energy from batteries, in order to save energy in the monitoring network, it is often justified not to pre-process the signal at the sensor, but to send analog signals to the MCP, which is responsible for processing the sensor signals and detecting the monitoring object [5]. Information exchange over the radio channel raises the problem of detecting an analog signal with an unknown law of fluctuations on the background of noise with an unknown distribution [6]. To solve this problem in this paper it is proposed to develop uniformly the most powerful invariant algorithm for incoherent object detection using signals from a monitoring network consisting of three sensors.
2. Theoretical analysis

The basis for constructing decision rules for detecting signals that are stable with respect to the conditions of observation are methods for testing complex hypotheses developed in mathematical statistics with the presence of “interfering” parameters [7, 8, 9], which allow in some cases to find the optimal rules for detecting signals in the background noise of unknown intensity (or even with an unknown distribution) [10]. An important role in overcoming the “a priori difficulty” is played by rank decisive detection rules, which are highly resistant to observation conditions, since they use rank statistics that take into account the most stable features of the distribution of source data [11].

One of the main assumptions adopted in [12] when constructing such rules is the assumption of the equality of the distributions for oscillations in the two analyzed areas when observing only the noise in them; the presence of a signal in one (strictly one) of these sections violates this equality, since the signal component is superimposed on the component of the noise background.

Let us make an attempt to weaken the indicated premise in the sense that the appearance of a signal is allowed not only strictly in one of the two observed areas, but also on the border between them (that is, in both areas). The corresponding ranking rule of detection in this case becomes more adapted to the actual conditions of observation.

Simultaneous processing of statistically independent fluctuations coming from three adjacent monitoring sites is assumed. This allows you to prevent the appearance of a signal at the border of two sections. It is believed that a signal with an unknown law of fluctuations can appear either only in the second or in the third section, or in both. The distribution of the background noise is assumed to be unknown, but the same for all three sites. For concreteness, it is assumed (although this is not necessary) that the MCP receiver performs non-coherent processing of incoming oscillations, so that the initial observations are independent reports \(U_1, U_2, U_3\) of envelopes from the detector output, respectively, to the three observed sites.

It is natural to assume that the presence of a signal in the second and third sections leads to a stochastic increase in magnitudes \(U_2\) and \(U_3\) in comparison with the value \(U_1\), in other words, the family of distributions of the original observations is stochastically increasing [13].

The latter condition corresponds to the practice and at the same time does not cause a noticeable narrowing of the class of the considered distributions of input quantities.

With unknown distribution of observations \(U_i, i = 1, 2, 3\), it is natural to consider (as in [14]) the class of decision rules that are invariant to any transformations of the original quantities, preserving their property of being equally or differently distributed. Such transformations are any non-monotonic functions of the original observations. Regarding these transformations, the maximum invariant statistics (Maximal Invariant - MI) is the set of ranks of values \(U_i, i = 1, 2, 3\) taken as elements of a single sample, so any invariant detection rule is based on this rank statistics [15]. Note that the ranks of the values \(U_i, i = 1, 2, 3\) are unchanged with all the transformations indicated, and their use allows us to get rid of the need to know the exact distribution of the sample. Among the three ranks of values \(U_i, i = 1, 2, 3\) it is enough to consider only two of them, for example, ranks \(r_2\) and \(r_3\) of values \(U_2\) and \(U_3\) since they completely determine and rank the values \(U_1\).

In accordance with the initial prerequisites, the task of detecting an object of observation from the signals coming over the radio channel to the MCP from three contact sensors can be formulated as a test of the following complex hypotheses: \(H_0\) - «distribution of ranks \(r_2, r_3\) is uniform (no signal)»; \(H_1\) - «distribution of ranks \(r_2, r_3\) is irregular, where one of the quantities \(U_2, U_3\) or both are stochastically larger than \(U_1\) (the signal is present in the second or third section, or both)». Each of the ranks \(r_2, r_3\) can take just three values: 1, 2, 3, so under the hypothesis \(H_0\) their uniform distribution is equal \(P_0(r_2, r_3) = 1/6\).
Denote the distribution of ranks \( r_2, r_3 \) with alternatives \( H_1 \) as \( P(r_2, r_3) \). According to the fundamental Neumann-Pearson lemma, the most powerful rank rule for testing hypotheses, based on statistics \( (r_2, r_3) \), has a critical area of view \( P(r_2, r_3) > C \), where the constant \( C \) is determined according to given probability of false alarm.

It is easy to note from here that, in general, there is no Uniformly Most Powerful (UMP) in terms of the signal-to-noise ratio in the sections of the rank rule for testing these hypotheses, based on statistics \( r_2, r_3 \), since it is clear that probabilities \( P(r_2, r_3) \) depend on the signal-to-noise ratio in the second and third sections, this dependence is different for different values \( r_2 \) and \( r_3 \) and may be nonmonotonic in magnitude \( r_2, r_3 \).

You can, however, narrow down the specified class of rank decision rules for testing initial hypotheses and find the RNM decision rule in this narrower class. Indeed, a signal is allowed to appear in either one (second or third) observation section or in both sections at once, and the signal-to-noise ratio in these sections is arbitrary. Therefore, the decision rule for detection should not give preference to any of the values \( U_2 \) and \( U_3 \), i.e. it must be invariant to the location of the signal among the second and third sections. In terms of ranks \( r_2 \) and \( r_3 \) this means that the problem is symmetric with respect to the permutation of the quantities \( r_2 \) and \( r_3 \) where statistics \( r = r_2 + r_3 \) serves as an MI with respect to this permutation, so that any invariant to the specified transformation of quantities \( r_2, r_3 \) decision rule should be based on \( r \).

It is easy to see that in such a narrower class of rank decision rules, the UMP hypothesis rule exists [16]. For this it is enough to note that statistics can take only three values: 5, 4, 3, moreover, its distribution \( P(r) \) (uniform under hypothesis \( H_0 \) and equal \( P_0(r) = 1/3 \)) is monotone on \( r \) with any alternative \( H_1 \). The latter follows from the fact that the probabilities \( P(r_2, r_3) \) monotone with respect to each of the values \( r_2 \) and \( r_3 \) separately, since the family of distributions of initial observations increases stochastically in each of the quantities \( U_2 \) and \( U_3 \) separately.

Thus, the UMP invariant rank hypothesis testing rule has a test function of the form

\[
\begin{align*}
\varphi(r) &= \gamma, r = C, \\
0, r < C,
\end{align*}
\]

where the threshold number \( C \) and probability \( \gamma \) uniquely are found by given probability of false alarm \( \alpha \) from the condition

\[
\alpha = E_0[\varphi(r)]
\]

here averaging « \( E_0 \) » is produced by distribution \( P_0(r) \). This rule does not depend on signal and noise parameters and is characterized by a constant probability of false alarm for any noise distribution (threshold level during operation). In addition, it is insensitive to the "movement" of the signal among the second and third sections of observation.

The effectiveness of the detection rule (1), (2) is determined by its power function, which shows the dependence of the probability of correct detection on the signal-to-noise ratio in areas. This function is equal to

\[
\beta = E[\varphi(r)].
\]

where averaging « \( E \) » is produced by distribution \( P(r) \) of statistics \( r \) for the case of alternatives \( H_1 \).
As an example, consider the calculation of the probability of correct detection of a signal with Rayleigh fluctuations against the background of normal noise. In the absence of a signal envelope distribution $U_1$ are the same and equal to
\[ p(U_i) = 2\theta U_i \exp(-\theta U_i^2), i = 1, 2, 3, \theta = 1/2 \sigma^2, \]
\[ (4) \]
where $\sigma^2$ — noise dispersion at the output of the linear path of the receiver. In the presence of a signal in the second and third sections of the distribution of envelopes $U_2$ and $U_3$ are equal to
\[ g_i(U_i) = [2\theta U_i / (1 + q_i)] \exp[-\theta U_i^2 / (1 + q_i)], \]
\[ (5) \]
where $q_i = \nu_i^2 / \sigma^2$ — signal-to-noise ratio in the second and third sections;
\[ \nu_i^2 — signal strengths in these areas, \ i = 2, 3. \]

Probabilities $P(r_2, r_3)$, and consequently the probability $P(r)$ can now be calculated on the basis of general results from the theory of rank statistics [17, 18]. So, the probability
\[ P(r_2, r_3) = \frac{1}{3!} E \left[ \frac{g_2(x_2)}{p(x_2)} \frac{g_3(x_3)}{p(x_3)} \right], \]
\[ (6) \]
where $x_2 = x^{(r_2)}_2 < x_3 = x^{(r_3)}_3$ - ordered sample taken from the density distribution $p(\cdot)$; the averaging $E$ is produced by the distribution of this ordered sample, which is equal to
\[ w(x_{1,2}) = \frac{3! p(x_2) p(x_3)}{(r_2 - 1)! (r_3 - r_2 - 1)! (3 - r_3)!} \left[ F(x_2) \right]^{r_2 - 1} \times [F(x_3) - F(x_2)]^{3 - r_2 - 1} [1 - F(x_3)]^{3 - r_3}, \]
\[ (7) \]
where $F(\cdot)$ - distribution function of a random variable with a density $p(\cdot)$.

3. Simulation results

Using formulas (6), (7) and densities $p(\cdot)$ and $g_i(\cdot)$, one can get that $i = 2, 3, P(r=5)=P(r_2 = 2, r_3 = 3) + P(r_2 = 3, r_3 = 2) = (1 + q_2 + q_3 + q_2 q_3)/b; \ P(r=4)=P(r_2 = 1, r_3 = 3) + P(r_2 = 3, r_3 = 1) = (1 + q_2)^2 / (2 + q_2) + (1 + q_3)^2 / (2 + q_3) / b; \ P(r=3)=P(r_2 = 1, r_3 = 2) + P(r_2 = 3, r_3 = 1) = (1 + q_2)^2 / (2 + q_2) + (1 + q_3)^2 / (2 + q_3) / b; \ b = 3 + 2q_2 + 2q_3 + q_2 q_3. \]
\[ (8) \]
So, with $\alpha = 1/3$ the critical domain of the detection decision rule (1), (2) contains a single point. $r=5$ (there is no randomization) and the power function (3) in this case is equal to $\beta(q_2, q_3) = P(r = 5) = (1 + q_2 + q_3 + q_2 q_3) / (3 + 2q_2 + 2q_3 + q_2 q_3). \]
\[ (9) \]
Note that the power function is symmetric with respect to the values and, as expected, from the construction of a decision rule that is invariant to the location of the signal among the second and third observation plots. It is maximum at $q_2 = q_3$ and is minimum at $q_2 = 0$ or at $q_3 = 0$. Its corresponding values for $\alpha = 1/2$ and $q_2 = q_3 = q$ where $q$ — the signal-to-noise ratio in the segment for the decision rule in [19] is equal to $(2 + q) / (4 + q)$ and $(12 + 17 q + 6 q^2) / (2 + q) / (3 + 2 q). \]
\[ (10) \]
Comparison of these values with the power function of the rule in [20] using a couple of sections shows a slight decrease in the effectiveness of the rule (1), (2) compared to the rule in [20-21]. This decrease is the fee for the additional properties of the invariance of the rule (1), (2), which consists in the fact that this rule allows for the "appearance" of a signal in any of the last two among the three
observation plots, as well as in both sections simultaneously (including on the border of the plots), without demanding a prior information about this fact.

4. Conclusions
The developed detection algorithm has a rather high resistance to observation conditions, since it:
1) provides a constant probability of false alarm for any (a priori unknown) noise distribution;
2) is invariant to the law of signal fluctuations and its location among the last two observed plots;
3) is characterized by the probability of correct detection, depending only on the signal-to-noise ratio and, moreover, this probability is maximum for each value and;
4) maintains its performance at unsteady noise and signal.

The noted properties of the developed algorithm are important from a practical point of view and allow you to use it when creating automatic detectors of a pollution signal.

5. References
[1] Burlov V, Andreev A, Gomazov F 2018 Development of a model for the management of environmental safety of the region, taking into account of the GIS capacity MATEC Web of Conferences
[2] Burlov V G, Popov N N 2017 Management of the application of the space geoinformation system in the interests of ensuring the environmental safety of the region vol 161 of the Advances in the Astronautical Sciences Series DyCoSS'2017 751-759
[3] Chernetsova E A 2008 Remote monitoring of oil pollution in the aquatic environment Monograph (Spb.: RSHU) 179
[4] "Automation of the environmental monitoring processes" 2017 Systems of Signal Synchronization, Generating and Processing in Telecommunications (SINKHROINFO) (Kazan) 1-3
[5] Beskid P, Shishkin A D, Chernetsova E A 2016 Algorithm of sensor access to the communication channel, maximizing the lifetime of a freely drifting monitoring subsystem of the environment Scientific Herald 1(7) 149-156
[6] Hartmann W M and Jon Pumplin 1988 Noise power fluctuations and the masking of sine signals J. Acoust. Soc. Am. 83(6) 2277-2289
[7] Lehman E 1979 Testing statistical hypotheses Per. from English Yu V Prokhorov (M.: Publishing house "Science") 408
[8] Akimov P S, Bakut P A, Bogdanovich V A and dr. 1984 Theory of signal detection Ed. P A Bakuta (M.: Radio and communication) 440
[9] Prokofiev V P 1971 Asymptotically optimal detector of fluctuating signals in noise of unknown intensity Izv. universities of the USSR Radioelektronika 14 1-6
[10] Shishkin A D Invariant target detection rule for radar with high Resolution Radio engineering and electronics-1974 19 7S 1524-1526
[11] Novikova N M, Lyalikova V G Parametric and non-parametric models Signal Detectors Voronezh State Technical Bulletin University vol 6 8 122-126
[12] Gayek J, Shidak Z 1971 Theory of rank criteria Publishing house "Science" 376
[13] Bogdanovich V A 1973 Application of the principle of invariance in detection problems with a priori uncertainty Izv. universities of the USSR Radio electronics vol 16 1 43-47
[14] Dubrovsky S A 1982 Applied multivariate statistical analysis (M.: Finance and Statistics) 216
[15] Brooks R R, Ramanathan P, Sayeed A 2002 Distributed target tracking and classification in sensor networks IEEE Signal Processing Magazine vol 19 2 17-29
[16] Chernetsova E A, Shishkin A D 2008 Optimum Rank Detection Rule oil pollution of the aquatic environment through a network of contact sensors Scientific works of the XI International Scientific and Practical Conference "Fundamental and applied problems of instrument engineering, informatics and of Economics" (Moscow, MGUPI) 196-200
[17] 2009 Lectures courses at the REC Mathematical Institute V A Steklov RAS (MIAN) (M.:
MIAN) vol 14 Lectures on the asymptotic theory of rank criteria D M Chibisov - 176
[18] Khitsenko V E 2012 Rank criteria for information security tasks Reports of Tusur 1(25) 2 29–33
[19] Orlov A I 2017 Model of the analysis of coincidence in the calculation of nonparametric Rank Statistics Factory Lab. Diagnostics of materials 83 11 66-72
[20] Orlov A I, Tolecheev V O 2011 On the use of non-parametric statistical criteria for assessing the accuracy of classification methods (summarizing article) Factory laboratory Diagnostics of materials 77 3 58-66
[21] Solovev D B, Kuzora S S 2016 Implementation of noise-immune Rogowski coils for busbar differential protection modernization Electric Power Systems Research 138 pp. 223-232. [Online]. Available: http://dx.doi.org/10.1016/j.epsr.2016.03.039.