

DOI: 10.18721/JPM.12314

УДК 519.24

## THE SHANNON INFORMATION QUANTITY IN THE TASKS ASSOCIATED WITH LINEAR REGRESSION: USAGE PATTERN

*Yu.A. Pichugin*

Saint Petersburg State University of Aerospace Instrumentation,  
St. Petersburg, Russian Federation

The article discusses the use of the Shannon information quantity (SIQ) in the tasks associated with linear regression. It has been shown that the SIQ contained in the response components with respect to stochastic parameters is expressed through the Fisher information matrix, is a convex functional on the set of response components, and is equivalent to the use of the  $D$ -criterion in the problems of planning the experiment at a sufficiently large scale of parameters. The SIQ relatively constant regression parameters were determined. An alternative formulation of the optimal experiment planning (OEP) problem was considered and its relation to the traditional formulation was analyzed. The problem of information ordering of data using regression on the basis of principal components was considered. Some algorithms taking into account the value of information in the presence of partial data gaps were proposed.

**Keywords:** Shannon information quantity, linear regression, stochastic regression parameter

**Citation:** Pichugin Yu.A., The Shannon information quantity in the tasks associated with linear regression: usage pattern, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 150–161. DOI: 10.18721/JPM.12314

## ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ ИНФОРМАЦИИ ПО ШЕННОНУ В ЗАДАЧАХ, СВЯЗАННЫХ С ЛИНЕЙНОЙ РЕГРЕССИЕЙ

*Ю.А. Пичугин*

Санкт-Петербургский государственный университет аэрокосмического приборостроения,  
Санкт-Петербург, Российская Федерация

В статье рассматривается использование количества информации по Шеннону (SIQ) в задачах, связанных с линейной регрессией. Показано, что SIQ, содержащееся в компонентах отклика относительно стохастических параметров, выражается через информационную матрицу Фишера, является выпуклым функционалом на множестве компонент отклика и при достаточно большом масштабе параметров эквивалентно использованию  $D$ -критерия в задачах планирования эксперимента. Определено SIQ относительно постоянных параметров регрессии. Рассмотрена альтернативная постановка задачи оптимального планирования эксперимента (ОЕП) и проанализирована ее связь с традиционной постановкой. Рассмотрена задача информационного упорядочивания данных при использовании регрессии на базис главных компонент. Предложены алгоритмы, учитывающие ценность информации при наличии частичных пропусков данных.

**Ключевые слова:** количество информации по Шеннону, линейная регрессия, стохастический параметр регрессии

**Ссылка при цитировании:** Пичугин Ю.А. Особенности использования информации по Шеннону в задачах, связанных с линейной регрессией// Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 12. № 3. С. 164–176. DOI: 10.18721/JPM.12314

### Introduction

The study considers Shannon information as its main subject. Let us briefly describe how this concept evolved in literature.

The quantity of information in the transmitted message, first defined by Claude Elwood Shannon, is closely related to the concept of entropy. Average entropy  $H(\xi)$ , or the information transmitted in a message  $\xi$  by  $n$  characters (values)

$$\{x_1, x_2, x_n\},$$

occurring with probabilities

$$\{P_\xi(1), P_\xi(2), P_\xi(n)\}$$

follows the expression [1]:

$$H(\xi) = -\sum_{i=1}^n P_\xi(i) \cdot \log_2 P_\xi(i) = \sum_{i=1}^n P_\xi(i) H_i,$$

where  $H_i(\xi) = -\log_2 P_\xi(i)$  is the individual entropy.

Theoretically, any number  $a > 1$ , which sets the scale, can be used as the base of the logarithm.

If there is another random variable  $\eta$  taking  $m$  values, i.e.,

$$\{y_1, y_2, y_m\},$$

with probabilities

$$\{P_\eta(1), P_\eta(2), P_\eta(m)\}$$

then, according to Shannon, the quantity of information  $I(\xi, \eta)$  contained in the message  $\xi$  relative to the message  $\eta$  or in  $\eta$  relative to  $\xi$  (there is symmetry in this case) is expressed as

$$I(\xi, \eta) = -\sum_{i=1}^n \sum_{j=1}^m P_{\xi\eta}(i, j) \times \log_a \frac{P_{\xi\eta}(i, j)}{P_\xi(i)P_\eta(j)},$$

where  $P_{\xi\eta}(i, j)$  is the probability that  $\xi$  takes the value  $x_i$  and  $\eta$  takes the value  $y_j$ , respectively.

If  $P_{\xi\eta}(i, j) = 0$ , the corresponding member of the sum is assumed to equal zero.

Gelfand and Yaglom [2] proved that if  $\xi$  and  $\eta$  are two random Gaussian vectors, then the amount of information  $I(\xi, \eta)$  contained in the vector  $\xi$  relative to the vector  $\eta$  and vice versa is

$$I(\xi, \eta) = -\sum_{i=1}^n \sum_{j=1}^m P_{\xi\eta}(i, j) \times \log_a \frac{P_{\xi\eta}(i, j)}{P_\xi(i)P_\eta(j)}, \quad (1)$$

where  $V_\xi, V_\eta$  are the covariance matrices of the components of vectors  $\xi$  and  $\eta$ , respectively;  $\mathbf{I}$  is the unit matrix of the corresponding dimension;  $V_{\xi\eta}$  is the covariance matrix of the components of vector  $\xi$  and vector  $\eta$ ,

$$V_{\eta\xi} = V_{\xi\eta}^T,$$

where  $T$  is the transpose operator.

It was established in [2] that Shannon's quantity of information was independent of the scale of variables and invariant with respect to linear transformations of the vectors  $\xi$  and  $\eta$ . This allows to represent Eq. (1) in the form

$$I(\xi, \eta) = -\log_a |\sin\alpha_1 \sin\alpha_2 \dots \sin\alpha_h|, \quad (2)$$

where  $\cos\alpha_j = r_j (j = 1, 2, \dots, h)$  is no other than a geometric interpretation of the correlation coefficient  $r_j$  between a pair of independent (principal) components of these vectors. In this case,  $h = \min(n, m)$ . Apparently, Eq. (1) allows to use SIQ in multidimensional problems.

A notable applied study by Pokrovsky [3] applied SIQ to solving the problem of optimization of satellite observations based on the linear regression model. This study is very close to the problem of optimal planning of a regression experiment (referred to as optimal experiment planning (OEP) from now on). However, the case of stochastic regression parameters (see the section "Relationship between SIQ and the Fisher information matrix" below) is considered in [3], departing from the traditional statement of the general OEP problem (see the section "Traditional and alternative statements of the OEP problem").

The goal of this study has consisted in understanding the relationship between SIQ, the Fisher information matrix and the OEP criteria in an alternative statement of the problem, developing the ideas of [2, 3], and finding new possibilities for using this mathematical framework in different practical applications.

### Relationship between SIQ and the Fisher information matrix

The classical statement of the regression analysis problem includes the following model:

$$\mathbf{y} = \mathbf{F}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad (3)$$

where  $\mathbf{y}$  is the measurement vector with the dimension  $n$ ;  $\boldsymbol{\theta}$  is the vector of parameters to be estimated, with the dimension  $m$ ;  $\mathbf{F}$  is the  $(n \times m)$  matrix ( $n \geq m$ );  $\boldsymbol{\varepsilon}$  is the random error vector.

It is assumed here that the components of the vector  $\varepsilon$  obey the multidimensional normal distribution, have zero mean values, are mutually independent and have the same variance  $\sigma^2$ , i.e.,

$$\varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$$

where  $\mathbf{0}$  is a zero vector.

Therefore,

$$\mathbf{y} \sim N(\mathbf{F}\boldsymbol{\theta}, \sigma^2 \mathbf{I}),$$

i.e.,  $\mathbf{V}_y = \mathbf{V}_\varepsilon$ .

In addition, we shall assume that  $\mathbf{F}$  is a matrix of full rank.

The Fisher information matrix of the regression parameter vector  $\boldsymbol{\theta}$ , which is a derivative of the information contained in the response components relative to these regression parameters, is determined by the following formula [4]:

$$\mathbf{M} = -E \left\{ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right\},$$

where  $E$  is the expected value operator,  $L$  is the log-likelihood, i.e.,

$$L = \ln \left\{ (2\pi)^{-\frac{n}{2}} \det^{-\frac{1}{2}} (\mathbf{V}_\varepsilon) \times \exp \left[ -\frac{1}{2} (\mathbf{y} - \mathbf{F}\boldsymbol{\theta})^T \mathbf{V}_\varepsilon^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\theta}) \right] \right\}.$$

It follows then that

$$\mathbf{M} = \sigma^{-2} \mathbf{F}^T \mathbf{F} = \mathbf{V}_{\hat{\boldsymbol{\theta}}}^{-1}, \quad (4)$$

where

$$\mathbf{V}_{\hat{\boldsymbol{\theta}}} = \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1} \quad (5)$$

is the covariance matrix of OLS (ordinary least squares) estimates of the components of the regression parameter vector  $\boldsymbol{\theta}$ .

Different convex functionals of the information matrix  $\mathbf{M}$  (see below) can serve as a measure of the information contained in the response components with respect to regression parameters (3). In this case, the vector of OLS estimates of these regression parameters is expressed as

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y} \quad (6)$$

(see monograph [5]).

If the assumptions about the covariance matrix of regression residuals are violated when

this matrix has an arbitrary structure, i.e.,

$$\varepsilon \sim N(\mathbf{0}, \mathbf{V}_\varepsilon) \quad (\mathbf{V}_\varepsilon \neq \sigma^2 \mathbf{I}),$$

we have, instead of Eqs. (4)–(6), the following expressions, respectively:

$$\mathbf{M} = \mathbf{F}^T \mathbf{V}_\varepsilon^{-1} \mathbf{F} = \mathbf{V}_{\hat{\boldsymbol{\theta}}}^{-1}, \quad (7)$$

$$\mathbf{V}_{\hat{\boldsymbol{\theta}}} = (\mathbf{F}^T \mathbf{V}_\varepsilon^{-1} \mathbf{F})^{-1}, \quad (8)$$

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^T \mathbf{V}_\varepsilon^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{V}_\varepsilon^{-1} \mathbf{y}. \quad (9)$$

Eq. (9) is known as the generalized method of OLS estimation.

It follows from Eqs. (5) and (8) that the OLS estimate vector  $\boldsymbol{\theta}$  has a normal distribution

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2 (\mathbf{F}^T \mathbf{F})^{-1})$$

or

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, (\mathbf{F}^T \mathbf{V}_\varepsilon^{-1} \mathbf{F})^{-1})$$

(depending on the structure of the matrix  $\mathbf{V}_\varepsilon$ ).

Each of these distributions defines a corresponding likelihood function. It is easy to verify that the Fisher information matrices of the vector  $\boldsymbol{\theta}$  calculated from these likelihood functions coincide with the matrices given by Eqs. (4) and (7), respectively.

Let us assume that the regression parameters have a stochastic nature and obey a multidimensional normal distribution, i.e.,

$$\boldsymbol{\theta} \sim N(E\boldsymbol{\theta}, \mathbf{V}_\theta).$$

Then, even though the distribution of the vector  $\mathbf{y}$  is different, Eqs. (5), (6) or (8), (9) are still valid for OLS estimates of the components of the vector  $\boldsymbol{\theta}$  constructed of the parameters of regression (3), depending on the structure of the matrix  $\mathbf{V}_\varepsilon$ , and the form of the probability distribution of the vector  $\hat{\boldsymbol{\theta}}$  and the Fisher information matrix do not change (see Eqs. (4), (7)). However, any convex functional of this matrix serves as a measure of information with respect to the parameter vector  $\boldsymbol{\theta}$ , contained in its OLS estimate. It is this information matrix (as more invariant) that is considered in case of stochastic parameters, for which the following theorem holds.

**Theorem 1.** *The quantity of information  $I(\mathbf{y}, \boldsymbol{\theta})$  contained in the response components in model (3) with respect to stochastic normally distributed parameters, provided that  $\mathbf{F}$  is a full-rank matrix,*



is related to the Fisher information matrix  $\mathbf{M}$  by the formula

$$I(\mathbf{y}, \theta) = \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}\mathbf{V}_\theta). \quad (10)$$

**Proof.** Assuming that regression parameters (3) are stochastic, we have the following equalities:

$$\mathbf{V}_y = \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T + \mathbf{V}_\varepsilon,$$

$$\mathbf{V}_{y_0} = \mathbf{F}\mathbf{V}_\theta \text{ and } \mathbf{V}_{\theta y} = \mathbf{V}_\theta\mathbf{F}^T.$$

Let us use these expressions to transform Eq. (1):

$$\begin{aligned} I(\mathbf{y}, \theta) &= -\frac{1}{2} \log_a \det(\mathbf{I} - \mathbf{F}\mathbf{V}_\theta\mathbf{V}_\theta^{-1}\mathbf{V}_\theta\mathbf{F}^T\mathbf{V}_y^{-1}) = \\ &= -\frac{1}{2} \log_a \det(\mathbf{V}_y\mathbf{V}_y^{-1} - \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T\mathbf{V}_y^{-1}) = \\ &= -\frac{1}{2} \log_a \det([\mathbf{V}_y - \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T]\mathbf{V}_y^{-1}) = \\ &= -\frac{1}{2} \log_a \det(\mathbf{V}_\varepsilon\mathbf{V}_y^{-1}) = \frac{1}{2} \log_a \det(\mathbf{V}_\varepsilon^{-1}\mathbf{V}_y) = \\ &= \frac{1}{2} \log_a \det(\mathbf{V}_\varepsilon^{-1}[\mathbf{V}_\varepsilon + \mathbf{F}\mathbf{V}_\theta\mathbf{F}^T]) = \\ &= \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{V}_\varepsilon^{-1}\mathbf{F}\mathbf{V}_\theta\mathbf{F}^T) = (*) \\ &= \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{F}^T\mathbf{V}_\varepsilon^{-1}\mathbf{F}\mathbf{V}_\theta) = \\ &= \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}\mathbf{V}_\theta). \end{aligned}$$

We used the algebraic identity

$$\det(\mathbf{I} + \mathbf{A}\mathbf{B}) = \det(\mathbf{I} + \mathbf{B}\mathbf{A})$$

for the equality (\*), which is obvious in case of square matrices at least one of which is invertible.

Indeed, let the matrix  $\mathbf{A}$  be invertible. Then, multiplying the left-hand side of this equality under the determinant by  $\mathbf{A}^{-1}$  on the left and by  $\mathbf{A}$  on the right, we obtain the identity. It is known that this transformation, called the similarity transformation, does not change the values of the determinant.

Now let  $\mathbf{A}$  and  $\mathbf{B}$  be rectangular matrices;  $\mathbf{A}$  is a matrix with dimensions  $n \times m$ ,  $\mathbf{B}$  is a matrix with the dimensions  $m \times n$  ( $n > m$ ).

Based on the statement of the theorem, we may assume that  $\mathbf{A}$  is a matrix of full rank. Without loss of generality, let us then assume that the first  $m$  rows of the matrix  $\mathbf{A}$  are linearly independent. We extend the rectangular matrices  $\mathbf{A}$  and  $\mathbf{B}$  to square ones with the dimensions  $n \times n$ :

$$\mathbf{A}^* = (\mathbf{A}|\mathbf{T}) \text{ и } (\mathbf{B}^*)^T = (\mathbf{B}^T|\mathbf{T}),$$

where block  $\mathbf{T}$  has the dimensions  $n \times (n - m)$ .

Let us fill block  $\mathbf{T}$  by the following rule:

$$t_{i+m,j} = a_{i+m,j+m}^* = \delta \quad (\delta \neq 0),$$

$$\text{if } i = j, j = 1, 2, \dots, n - m;$$

the remaining elements of  $\mathbf{T}$  are filled with zeros.

The given identity then holds for matrices  $\mathbf{A}^*$  and  $\mathbf{B}^*$  (see above). Next, tending  $\delta$  to zero, we have the required identity for the original matrices  $\mathbf{A}$  and  $\mathbf{B}$ .

Theorem 1 is completely proved.

Eq. (10) and its formulation were previously given in [6, 7].

Assuming that parameters of regression (3) are stochastic, formulating Eq. (10) yields another equation, namely,

$$I(\mathbf{y}, \theta) = \frac{1}{2} \log_a \det \mathbf{V}_y - \frac{1}{2} \log_a \det \mathbf{V}_\varepsilon.$$

This equality provides a good description of SIQ in model (3) with stochastic parameters.

### Traditional and alternative statements of OEP problem

**Traditional statement.** The initial statement of the OEP problem (that has become traditional) under the conditions of model (3) is reduced to calculating a set of quantities [8, 9]:

$$\mathbf{p} = \{p_1, p_2, \dots, p_n\},$$

$$p_i \geq 0; \quad i = 1, 2, \dots, n; \quad \sum_{i=1}^n p_i = 1, \quad (11)$$

called a plan.

These quantities are interpreted as relative frequencies at which the response components are measured (which is why it is termed the plan), similar to the law of discrete probability distribution. This concept was apparently borrowed from the theory of matrix games for mixed strategies [10], where a similar situation occurs. OEP theory uses the values comprising the plan to transform the information matrix by the following scheme (within the framework of this theory):

$$\mathbf{M} \rightarrow \mathbf{M}(\mathbf{p}) = (\mathbf{F}^T \mathbf{P} \mathbf{F}), \quad (12)$$

where  $\mathbf{P} = \text{diag}(p_1, p_2, \dots, p_n)$ .

A requirement for maximization of some convex functional (criterion) from this transformed matrix (see below) is imposed on the plan.

Thus, a transition is made from the initial values  $p_i = 1/n$  ( $i = 1, 2, \dots, n$ ) to optimal values at which some criterion of the transformed (see above) information matrix reaches a maximum, i.e.,

$$\text{crit} \mathbf{M}(\mathbf{p}) = \text{crit } \mathbf{F}^T \mathbf{P} \mathbf{F} \rightarrow \max.$$

The criteria (functionals of  $\mathbf{M}(\mathbf{p})$ ) considered include the determinant ( $D$  criterion), trace ( $A$  criterion), the minimum eigenvalue of the matrix  $\mathbf{M}(\mathbf{p})$  ( $E$  criterion) and other convex functionals of this matrix [8, 9].

**Clarification.** It was easy to notice above that the information matrix is understood in OEP theory not exactly as the Fisher information matrix equal to  $\sigma^{-2} \mathbf{F}^T \mathbf{F}$  (see Eq. (4)) but rather as the proportional matrix  $\mathbf{F}^T \mathbf{F}$  [8, 9] or  $\mathbf{F}^T \mathbf{P} \mathbf{F}$  (see above), which is of no fundamental importance for the problem of maximization of the functionals of this matrix. On the other hand, it follows from the regression analysis described, for example, in monograph [5] that transition to the matrix  $\mathbf{F}^T \mathbf{P} \mathbf{F}$  is only practical if we use parameter estimate

$$\hat{\boldsymbol{\theta}} = \left( \mathbf{F}^T \mathbf{P} \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{P} \mathbf{y},$$

known as the generalized OLS estimate (see Eq. (9)), or, if  $\mathbf{P}$  has a diagonal structure, as the weighted OLS estimate, which is what we actually have here. It follows from the Gauss–Markov theorem whose proof is given in [9] that it is correct to apply such an estimate only when, firstly,  $\mathbf{V}_\varepsilon \neq \sigma^2 \mathbf{I}$ , and, secondly, when the matrix  $\mathbf{P}$  is inversely proportional to the matrix  $\mathbf{V}_\varepsilon$ , i.e.,

$$\mathbf{P} = c \mathbf{V}_\varepsilon^{-1} \quad (c \neq 0).$$

However, no mention of these conditions is found in OEP theory [8, 9]. We can only assume that it is the result of discarding the factors  $\sigma^2$  and  $\sigma^{-2}$  in the equations for  $\mathbf{V}_\theta$  and  $\mathbf{M}$ , and sometimes  $\varepsilon$  in some regression equations (see [9]) that transformation (12) turned out to be unrelated to the structure of covariance matrix composed of errors of regression (3), even though the final goal is precisely in estimating the parameters.

**Alternative statement.** The following statement is equally important from a practical standpoint.

Let us consider different proper subsets of the form

$$\mathbf{q} \subset \{1, 2, \dots, n\}$$

( $\text{card} \mathbf{q}$  indicates the number of elements in  $\mathbf{q}$ ).

Let  $\mathbf{y}_q$  be a vector with the dimension  $\text{card} \mathbf{q}$

containing components  $\mathbf{y}$  only with numbers from the set  $\mathbf{q}$ ; let  $\mathbf{F}_q$  be a matrix that contains rows only with numbers from the set  $\mathbf{q}$ .

Using the vector  $\mathbf{y}_q$ , we can also calculate the OLS estimate of the vector  $\boldsymbol{\theta}$ , i.e.,

$$\hat{\boldsymbol{\theta}} = \left( \mathbf{F}_q^T \mathbf{F}_q \right)^{-1} \mathbf{F}_q^T \mathbf{y}_q. \quad (13)$$

Then the covariance matrix of the vector  $\boldsymbol{\theta}$  follows the expression

$$\mathbf{V}_\theta = \sigma^2 \left( \mathbf{F}_q^T \mathbf{F}_q \right)^{-1},$$

and the Fisher information matrix the expression

$$\mathbf{M}_q = \sigma^{-2} \mathbf{F}_q^T \mathbf{F}_q. \quad (14)$$

In the case  $\mathbf{V}_\varepsilon \neq \sigma^2 \mathbf{I}$ , Eqs. (13) and (14) change similarly to Eqs. (4) and (6) (see Eqs. (7) and (9)). Taking a fixed value  $\text{card} \mathbf{q} = k$  ( $k < n$ ), let us select a subset  $\mathbf{q}$  that yields the maximum of some convex functional on the information matrix  $\mathbf{M}_q$ .

This statement is fairly significant for the problem of saving observation tools with minimum loss in the accuracy with which all  $\mathbf{y}$  components are reproduced. Indeed, as we substitute the OLS estimate of the vector  $\boldsymbol{\theta}$  to model (3) by Eq. (13), we are interested in ensuring that all response values are reproduced with a minimum loss in accuracy, which is what minimizing the variance of the OLS estimate of the vector  $\boldsymbol{\theta}$  should ensure.

**Comparison of two statements.** Let us establish that the alternative statement is reduced to the traditional one by imposing the corresponding restrictions. Indeed,  $p_i = 1/k$  if  $i \in \mathbf{q}$  and  $p_i = 0$  if  $i \notin \mathbf{q}$ , which does not violate the basic requirements for plan (11). Then the information matrix corresponding to the traditional statement of the OEP problem follows the expression

$$\mathbf{M}(\mathbf{p}) = \mathbf{F}^T \mathbf{P} \mathbf{F} = \frac{1}{k} \mathbf{F}_q^T \mathbf{F}_q = \frac{\sigma^2}{k} \mathbf{M}_q,$$

where  $\mathbf{M}_q$  is determined by Eq. (14), i.e., it is the Fisher information matrix.

However, we also obtain a parameter estimate expressed by Eq. (13):

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \left( \mathbf{F}^T \mathbf{P} \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{P} \mathbf{y} = \\ &= k \left( \mathbf{F}_q^T \mathbf{F}_q \right)^{-1} \frac{1}{k} \mathbf{F}_q^T \mathbf{y} = \left( \mathbf{F}_q^T \mathbf{F}_q \right)^{-1} \mathbf{F}_q^T \mathbf{y}. \end{aligned}$$

This estimate does not contradict the Gauss–Markov theorem.

Evidently, using, for example, the  $D$  criterion to solve the problem in this statement is only possible given  $k \geq m$  when

$$\det(\mathbf{F}_q^T \mathbf{F}_q) \neq 0.$$

Therefore, the most convenient criteria in the alternative statement are  $A$  criterion,  $E$  criterion (see above), etc. However, as Seber observed in [5], the criteria that are independent of the scale of quantities are preferable. As noted above, SIQ does not depend on the scale of variables (see Eq. (2)). On the other hand, using SIQ implies that the regression parameters are stochastic in nature. We are going to confirm below that SIQ can be used in the alternative statement of the OEP problem if the parameters of regression (3) are constant values.

**Properties of SIQ as an OEP criterion in the alternative statement**

It follows from Eqs. (10) and (14) that the SIQ contained in the components of the vector  $\mathbf{y}_q$  relative to stochastic components of  $\boldsymbol{\theta}$  is expressed as

$$I(\mathbf{y}_q, \boldsymbol{\theta}) = \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}_q \mathbf{V}_\theta). \quad (15)$$

Eq. (15) allows to determine the SIQ contained in the vector  $\mathbf{y}_q$  relative to the vector of regression parameters  $\boldsymbol{\theta}$  for the case when the parameters of regression (3) are not stochastic in nature, i.e., have constant values.

**Definition.** *If the parameters in model (3) are constant, then  $I(\mathbf{y}_q, \boldsymbol{\theta})$  is expressed by the formula*

$$\begin{aligned} I(\mathbf{y}_q, \boldsymbol{\theta}) &= \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}_q \mathbf{V}_\theta) = \\ &= \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}_q \mathbf{M}^{-1}). \end{aligned} \quad (16)$$

**Comments.**

1) In this definition, we equated  $I(\mathbf{y}_q, \boldsymbol{\theta})$  to  $I(\mathbf{y}_q, \hat{\boldsymbol{\theta}})$ , i.e., reduced the problem to the case when both vectors are stochastic, as required by Eq. (1).

2) If  $a = 2$  (the bit is the unit of information in the binary numeral system), then

$$0 \leq I(\mathbf{y}_q, \boldsymbol{\theta}) \leq \frac{m}{2} \quad (m = \dim \boldsymbol{\theta}).$$

Next, let us define the following functional on subsets of the form  $\mathbf{q}$  (see above):

$$\Phi(\mathbf{q}) = I(\mathbf{y}_q, \boldsymbol{\theta}).$$

Then the following theorem holds true.

**Theorem 2.** *The functional  $\Phi(\mathbf{q})$  is convex, i.e., the inequality*

$$\Phi(\alpha \mathbf{p} + \beta \mathbf{q}) \geq \alpha \Phi(\mathbf{p}) + \beta \Phi(\mathbf{q}),$$

where  $\alpha, \beta \geq 0$ ;  $\alpha + \beta = 1$  holds true;  $\mathbf{p}, \mathbf{q} \subseteq \{1, 2, \dots, n\}$ .

*Proof.* Let us transform Eq. (15) as follows:

$$\begin{aligned} I(\mathbf{y}_q, \boldsymbol{\theta}) &= \frac{1}{2} \log_a \det(\mathbf{V}_\theta^{-1} + \mathbf{M}_q) + \\ &+ \frac{1}{2} \log_a \det \mathbf{V}_\theta. \end{aligned} \quad (17)$$

Let us prove the inequality imposed by the theorem for the functional  $\Psi(\mathbf{q})$ :

$$\Psi(\mathbf{q}) = \frac{1}{2} \log_a \det(\mathbf{V}_\theta^{-1} + \mathbf{M}_q). \quad (18)$$

Let  $\mathbf{p}, \mathbf{q} \subseteq \{1, 2, \dots, n\}$ , as expected. Then, according to the definition of  $\Psi(\mathbf{q})$ , we have:

$$\begin{aligned} \Psi(\alpha \mathbf{p} + \beta \mathbf{q}) &= \\ &= \frac{1}{2} \log_a \det(\mathbf{V}_\theta^{-1} + (\alpha \mathbf{M}_p + \beta \mathbf{M}_q)) = \\ &= \frac{1}{2} \log_a \det((\alpha + \beta) \mathbf{V}_\theta^{-1} + (\alpha \mathbf{M}_p + \beta \mathbf{M}_q)) = \\ &= \frac{1}{2} \log_a \det(\alpha (\mathbf{V}_\theta^{-1} + \mathbf{M}_p) + \\ &\quad + \beta (\mathbf{V}_\theta^{-1} + \mathbf{M}_q)) \geq^{(**)} \\ &\geq \frac{1}{2} \log_a (\det^\alpha (\mathbf{V}_\theta^{-1} + \mathbf{M}_p) \times \\ &\quad \times \det^\beta (\mathbf{V}_\theta^{-1} + \mathbf{M}_q)) = \\ &= \alpha \frac{1}{2} \log_a \det(\mathbf{V}_\theta^{-1} + \mathbf{M}_p) + \\ &\quad + \beta \frac{1}{2} \log_a \det(\mathbf{V}_\theta^{-1} + \mathbf{M}_q) = \\ &= \alpha \Psi(\mathbf{p}) + \beta \Psi(\mathbf{q}). \end{aligned}$$

Here the inequality  $(**)$  follows from the general inequality

$$\det(\alpha \mathbf{A} + \beta \mathbf{B}) \geq \det \mathbf{A}^\alpha \cdot \det \mathbf{B}^\beta;$$

the general inequality holds true under the given restrictions with respect to  $\alpha$  and  $\beta$  in case of symmetric positive-definite matrices  $\mathbf{A}$  and  $\mathbf{B}$  (the proof is given in [9]), which the matrices  $\mathbf{V}_\theta^{-1}$  and  $\mathbf{M}_q$  and, therefore, their sum are.

Thus, we have the inequality

$$\Psi(\alpha\mathbf{p} + \beta\mathbf{q}) \geq \alpha\Psi(\mathbf{p}) + \beta\Psi(\mathbf{q}).$$

Adding the term

$$\frac{1}{2} \log_a \det \mathbf{V}_\theta,$$

to both sides of this inequality, we obtain the following inequality:

$$\begin{aligned} & \Psi(\alpha\mathbf{p} + \beta\mathbf{q}) + \frac{1}{2} \log_a \det \mathbf{V}_\theta \geq \\ & \geq \alpha\Psi(\mathbf{p}) + \beta\Psi(\mathbf{q}) + (\alpha + \beta) \frac{1}{2} \log_a \det \mathbf{V}_\theta. \end{aligned}$$

In view of Eq. (17), we have the required inequality for the functional  $\Phi(\mathbf{q})$ .

Theorem 2 is completely proved.

**Remark.** It is evident from Eq. (16) and the proof of Theorem 2 that the functional

$$\Phi(\mathbf{q}) = I(\mathbf{y}_q, \theta)$$

is convex given constant regression parameters, when the quantity of information is determined by Eq. (16).

Let us return to the case of stochastic regression parameters. Here the following almost obvious theorem relating  $I(\mathbf{y}_q, \theta)$  to the  $D$  criterion (determinant of the information matrix, see above) holds true. Notably, while OEP theory traditionally does not consider the case of stochastic parameters, it does not explicitly prohibit it.

**Theorem 3.** *If stochastic regression parameters have a sufficiently large scale, maximizing the SIQ contained in the response components relative to the regression parameters is equivalent to the  $D$  criterion.*

**Proof.** Let  $\mathbf{p}, \mathbf{q} \subseteq \{1, 2, \dots, n\}$  and  $\text{card}\mathbf{p} = \text{card}\mathbf{q}$ . It follows from Eq. (17) that it is also possible in this case to consider the functional  $\Psi(\mathbf{q})$  defined by Eq. (18), instead of the functional

$$\Phi(\mathbf{q}) = I(\mathbf{y}_q, \theta).$$

It is fairly evident that if the stochastic regression parameters have a sufficiently large scale and, accordingly, the elements of the matrix  $\mathbf{V}_\theta^{-1}$  are small, the following inequalities are certainly equivalent:

$$\begin{aligned} & \det(\mathbf{M}_p + \mathbf{V}_\theta^{-1}) > \det(\mathbf{M}_q + \mathbf{V}_\theta^{-1}) \Leftrightarrow \\ & \Leftrightarrow \det(\mathbf{M}_p) > \det(\mathbf{M}_q), \end{aligned}$$

which, given that the logarithm is monotonic, also means the required.

Theorem 3 is completely proved.

Notably, Eqs. (15), (16) and the statement of Theorem 2 are quite sufficient to justify using SIQ in an alternative statement of the OEP problem, for both stochastic and constant parameters of regression (3).

If we solve the OEP problem in an alternative statement for successive values of  $k$  (see above) from  $k = 1$  to  $k = n$ , we obtain a sequence of response components by descending information contribution, as well as an increasing sequence of information quantities  $I_j$ . It is convenient to represent the latter, excluding the dependence on the choice of the base of logarithm  $a$ , in the following form:

$$I_j := (I_j / I(\mathbf{y}, \theta)) 100\% \quad (j = 1, 2, \dots, n),$$

where  $I(\mathbf{y}, \theta)$  is the total quantity of information when  $\mathbf{q} = \{1, 2, \dots, n\}$ .

Considerations arguing that the statement and solution of such a problem are valid are given at the end of the next section.

### Using principal component regression in information ordering and accounting for the value of information with data gaps

In most practical problems, we at best only have a sample of observations of a certain vector

$$\{\mathbf{y}_j, j = 1, 2, \dots, N\}.$$

Such sample is typically represented as a so-called selective matrix  $\mathbf{Y}$  of dimension  $n \times N$ . Thus, the columns of the matrix  $\mathbf{Y}$  are realizations of a random vector  $\mathbf{y}$ . Let us assume that the vector  $\mathbf{y}$  obeys a multidimensional normal distribution with the parameters  $\theta_y$  and  $\mathbf{V}_y$ , i.e.,

$$\mathbf{y} \sim N(\theta_y, \mathbf{V}_y).$$

Let us calculate the estimates of the distribution parameters

$$\begin{aligned} \hat{\theta}_y &= \frac{1}{N} \sum_{j=1}^N \mathbf{y}_j, \\ \hat{\mathbf{V}}_y &= \frac{1}{N-1} \sum_{j=1}^N (\mathbf{y}_j - \hat{\theta}_y)(\mathbf{y}_j - \hat{\theta}_y)^T. \end{aligned} \quad (19)$$

Next let us calculate an orthogonal matrix  $\mathbf{Q}$ , such that the equality

$$\mathbf{Q}^T \hat{\mathbf{V}}_y \mathbf{Q} = \hat{\mathbf{\Lambda}} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_n)$$

holds and  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$ .

Let us define a vector of dimension  $m$  ( $m < n$ ) in the form

$$\mathbf{z} = \mathbf{Q}_{(m)}^T (\mathbf{y} - \hat{\boldsymbol{\theta}}_y),$$

where the matrix  $\mathbf{Q}_{(m)}$  contains only the first  $m$  columns of the matrix  $\mathbf{Q}$ .

The components of the vector  $\mathbf{z}$  are mutually independent, are called selective principal components, and are interpreted as hidden factors.

The matrix

$$\hat{\mathbf{V}}_z = \hat{\boldsymbol{\Lambda}}_{(m)} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_m)$$

is essentially the estimate of the covariance matrix of components  $\mathbf{z}$ .

The word *selective* is usually omitted but it should be borne in mind that we would have true principal components only with the known distribution parameters  $\boldsymbol{\theta}_y$  and  $\mathbf{V}_y$ .

The most correct choice of dimension for the vector  $\mathbf{z}$  ( $m = \dim \mathbf{z}$ ) is connected to verifying the statistical hypothesis

$$H: \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq \lambda_{m+1} = \lambda_{m+2} = \dots = \lambda_n.$$

However, there are relatively simple methods for finding  $m$  [11].

Let us now use centered values of the vector  $\mathbf{y}$ , i.e.,

$$\mathbf{y} := \mathbf{y} - \hat{\boldsymbol{\theta}}_y,$$

and consider the regression

$$\mathbf{y} = \mathbf{Fz} + \boldsymbol{\varepsilon}, \quad (20)$$

where  $\mathbf{F} = \mathbf{Q}_{(m)}$ .

Let us assume that the residual vector  $\boldsymbol{\varepsilon}$  has the same properties as in model (3), i.e.,

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Model (20) differs from model (3) primarily in that the regression parameters in the first case (in model (20)) are, generally speaking, stochastic. As noted above, this model allows to solve different practical and theoretical problems related to information ordering by sequentially selecting response components with maximized SIQ at each selection step, i.e.,

$$I(\mathbf{y}_q, \mathbf{z}) = \frac{1}{2} \log_a \det(\mathbf{I} + \mathbf{M}_q \hat{\boldsymbol{\Lambda}}_{(m)}),$$

or any of the classical criteria, for example,  $\text{tr} \mathbf{M}_q$  [12].

In addition, model (20) allows to apply different approaches to determining the value of information with data missing due to instability of information sources.

Let the initial sample matrix  $\mathbf{Y}$  contain gaps in individual components. We construct a matrix  $\mathbf{N}$  of the same dimension as  $\mathbf{Y}$  ( $n \times N$ ), where each missing measurement corresponds to 0,

and each existing (not missing) measurement to 1. Then each element  $n_{ij}$  ( $i = 1, 2, \dots, n; j = 1, 2, \dots, N$ ) of the matrix  $\mathbf{N}$  is equal to zero or unity. The gaps in the matrix  $\mathbf{Y}$  are also filled with zeros. Let us calculate the mean values of the components of the vector  $\mathbf{y}$  by the formula

$$\bar{y}_i = \sum_{j=1}^N y_{ij} / \sum_{j=1}^N n_{ij} \quad (i = 1, 2, \dots, n)$$

and proceed to the centered matrix  $\tilde{\mathbf{Y}}$ , where

$$y_{ij} := y_{ij} - \bar{y}_i, \text{ if } n_{ij} \neq 0,$$

leaving the values that correspond to  $n_{ij} = 0$  unchanged, i.e., zeros.

**Algorithm 1.** Now, if stable sources of information are preferred over unstable for some reason, the estimate for the covariance matrix of components of the vector should be calculated by the formula

$$\tilde{\mathbf{V}}_y = \frac{1}{N} \tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T. \quad (21)$$

If there are no gaps and the divisor is equal to  $N - 1$ , estimate (21) coincides with estimate (19). After calculating the matrix  $\mathbf{F} = \mathbf{Q}_{(m)}$  and performing informational ordering of the components of the vector  $\mathbf{y}$ , the numbers of the response components corresponding to unstable information (gaps) are likely to be at the end of this sequence (see above, the end of the previous section). Estimates of individual elements of the mutual covariance matrix by Eq. (21) with missing data are similar to estimates of autocovariance proposed by Jenkins and Watts in their well-known study [13]. The divisor in these estimates equals, regardless of the magnitude of the shift, and, accordingly, of the number of terms, the length of the series  $N$ , which, as established in [13], does not yield overestimated wavelengths. Such a constant divisor can reduce the value of unstable information in our case.

**Algorithm 2.** However, it is precisely the unstable sources of information that may prove to have the greatest value to the researcher. In this case, the elements of the covariance matrix should be estimated by the formula

$$[\tilde{\mathbf{V}}_y]_{ij} = [\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^T]_{ij} / [\mathbf{N} \mathbf{N}^T]_{ij} \quad (22)$$

$$(i, j = 1, 2, \dots, n)$$

with  $[\mathbf{N} \mathbf{N}^T]_{ij} \neq 0$  and  $[\tilde{\mathbf{V}}_y]_{ij} = 0$  otherwise.

Here  $[\cdot]_{ij}$  is the operator taking a matrix element from the  $i$ th row and from the  $j$ th column.

Eq. (22) ensures that the divisor is equal to the number of nonzero terms of the corresponding sum. If the estimate of the covariance matrix is calculated this way, the numbers of the response components with data gaps can appear at the beginning of the ordered sequence [14].

Recall that SIQ is calculated using biased estimates and has a geometric interpretation for a Gaussian distribution (see Eq. (2)). Therefore, a divisor equal to  $N$  could also be used in Eq. (19), while a biased estimate  $\sigma^2$  should be used in model (20) [11].

**Remark.** The only serious criticism regarding SIQ as a measure of information quantity is that it does not take into account the subjective value of information to the consumer.

The algorithms proposed above solve this problem to some extent, since they can be also used in the most general case for calculating SIQ by Eq. (1). Moreover, calculating the estimates of the matrices  $\mathbf{V}_{\xi}$ ,  $\mathbf{V}_{\eta}$  and  $\mathbf{V}_{\xi\eta}$ , we should construct and use the corresponding data availability matrices:  $\mathbf{N}_{\xi}$  and  $\mathbf{N}_{\eta}$  (see above). Evidently, if unstable information has high value (algorithm 2), the estimate  $\mathbf{V}_{\xi\eta}$  should be calculated using the corresponding matrix elements  $\mathbf{N}_{\xi}\mathbf{N}_{\eta}^T$  as divisors (see Eq. (22)). Any subjective preferences for the value of information with respect to individual components of the given vectors can be satisfied by combining algorithms 1 and 2.

Before we discuss the results, we should note that using principal components regression and information ordering based on this regression (see the end of the previous section) allowed to solve a number of applied problems. The most important of these problems are the following:

- optimal integration of satellite and ground-based observations [12];
- detecting the core of the corruption structure [14];
- detecting the zones with dynamic instability of atmospheric circulation [15];
- optimization of environmental monitoring [16];
- optimization of microelectronic production control [17].

This approach can be used for solving a wide range of applied problems, which confirms the practical significance of our study.

### Results and discussion

Reviewing the situation as a whole, we should first of all pay tribute to Pokrovsky's studies (see Introduction and Ref. [3]), estab-

lishing that both the case of stochastic regression parameters and the alternative statement of the OEP problem are relevant for theoretical and practical problems. Pokrovsky was the first to use SIQ in the problem of optimizing of atmospheric sensing via satellites (see also introduction and Ref. [3]). However, that study did not address the relationship between SIQ and the Fisher information matrix (see Eqs. (10), (15), (16)), as well as other issues considered in our paper.

The traditional statement of the OEP problem has generated some interesting mathematical results, as this approach allows to solve the problems based on continuous plans by involving methods from mathematical analysis. Therefore, a theoretically substantiated answer to the clarification made in the section "Traditional and alternative statements of the OEP problem" would make these results more valuable for practical problems.

While the alternative statement of the OEP problem is fairly simple and clear, the following issue is still poorly understood. With a fixed

$$\text{card}\mathbf{q}=k_* (k_* < n),$$

there is obviously a proper subset  $\mathbf{q}$  which gives the maximum of the optimal criterion

$$\max_{\text{crit}\mathbf{M}_q} (\text{crit}\mathbf{M}_q = \text{tr}\mathbf{M}_q, I(\mathbf{y}_{q^*}), \text{etc.}).$$

However, we cannot be certain and neither can we obtain rigorous proof that we will find this exact value  $\max_{\text{crit}\mathbf{M}_q}$  by solving this problem by successively increasing  $k$  from 1 to  $k_*$  (see the end of the section "Properties of SIQ as an OEP criterion in the alternative statement") maximizing the given criterion at each step of sequential selection.

A similar situation is observed in constructing optimal regression [5]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k^*} x_{k^*} + \varepsilon,$$

when we select the regressors providing the maximum determination coefficient  $R^2$  with a fixed number of  $k_*$  from the set of potential regressors  $\mathbf{x}$  ( $\text{card}\mathbf{x} \geq k^*$ ).

The problem can also be solved by sequential selection of regressors, maximizing the determination coefficient  $R^2$  at each step [5]; similarly, the question of whether we can obtain the maximum value of  $R^2$  by this procedure remains open.

It seems plausible enough that the answer is positive with a normal distribution. In practice,



this problem is usually solved in both cases by the inclusion and rejection procedure, stopping the process if repeated steps are detected [5].

### Conclusion

We have obtained the following results.

1. We have proved that the SIQ contained in the response components with respect to stochastic parameters

a) is directly related to the Fisher information matrix;

b) is a convex functional on the set of response components;

c) is equivalent to using the  $D$  criterion in the OEP problem given a sufficiently large scale of parameters.

2. SIQ has been found with respect to constant regression parameters, where SIQ is

also a convex functional on the set of response components;

3. We have analyzed the relationship between traditional and alternative statements of the OEP problem.

4. We have considered principal component regression in the problem of information ordering of data.

5. We have proposed algorithms accounting for the subjective value of information with partially missing data.

Finally, we should stress that the above studies [3, 12, 14–17] conclusively prove that the regression model with stochastic parameters can be used in the alternative statement of the OEP problem; practical significance of the results obtained in this study is fully confirmed by [14–17].

### REFERENCES

1. Shannon C.E., Weaver W., A mathematical theory of communication, Urbana University Press, Chicago, 1949.
2. Gelfand I.M., Yaglom A.M., Computation of the amount of information about a stochastic function contained in another such function, Uspekhi Matematicheskikh Nauk. 12 (1 (73)) (1957) 3–52.
3. Pokrovsky O.M., Ob optimalnykh usloviyakh kosvennogo zondirovaniya atmosfery [On the optimum conditions of the atmosphere indirect sensing], Reports of AS USSR, Physics of Atmosphere and Ocean. 5 (12) (1969) 1324–1326.
4. Kendall M.G., Stuart A., The advanced theory of statistics, Vol. 2: Inference and relationship, Charles Griffin & Company limited, London, 1958–1966.
5. Seber G.A.F., Linear regression analysis, John Wiley & Sons, New York, London, Sydney, Toronto, 1977.
6. Pichugin Yu.A., O svyazi kolichestva informatsii po Shennonu s informatsionnoy matritsey Fishera v zadache planirovaniya regressionnogo experimenta [On the relation of the amount of Shannon's information with Fisher's information matrix in the problem of regression experiment planning], The Issue of Inter-University Scientific Articles "Informatics – Research and Innovation", No. 3, Saint-Petersburg, Herzen University (1999) 32–36.
7. Pichugin Yu.A., Notes on selecting data in the tasks associated with linear multiple regression, Industrial Laboratory. Diagnostics of materials. 5(2002) 61–62.
8. Matematicheskaya teoriya planirovaniya experimenta [Mathematical theory of experiment planning], Ed. S.M. Ermakov, Nauka, Moscow, 1983.
9. Ermakov S.M., Zhiglyavskiy A.A., Matematicheskaya teoriya optimalnogo experimenta [Mathematical theory of optimal experiment], Nauka, Moscow, 1987.
10. Von Neumann J., Morgenstern O., Theory of games and economic behavior, Princeton University Press, Princeton, 1953.
11. Pichugin Yu.A., Notes on using the principal components in the mathematical simulation, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 11 (3) (2018) 74–89.
12. Pichugin Yu.A., Pokrovsky O.M., On the method of complexing the ground and spaceborne meteorologic information, Remote Sensing, (6) (1992) 25–31.
13. Jenkins G.M., Watts D.G., Spectral analysis and its applications, Holden-Day, San Francisco – Cambridge – London – Amsterdam (1969).
14. Pichugin Yu. A., Malafeyev O.A., Rylow D., Zaitseva I. A statistical method for corrupt agents detection, International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017), AIP Conf. Proc. (1978) 100014-1–100014-4.
15. Pichugin Yu.A., Geografiya dinamicheskoy neustoychivosti tsirkulyatsii atmosfery v Severnom polusharii (modelirovanie i analiz) [Geography of dynamic instability of atmospheric circulation in the Northern hemisphere (simulation and

analysis)], Reports of Russian Geographical Society. 137 (3) (2005) 12–16.

16. **Pichugin Yu.A.**, Ekologicheskiy monitoring i metody mnogomernoy matematicheskoy statistiki [Environmental monitoring and methods of multivariate mathematical statistics], Astrakhan Bulletin

of Environmental Education (2) (2012) 101–105.

17. **Gusman Yu.A., Pichugin Yu.A., Smirnov A.O.**, Shannon's informative value for control of microelectronic productions, Voprosy Radioelektroniki [Radio Electronics Issues]. (10) (2018) 6–10.

*Received 22.05.2019, accepted 14.06.2019.*

## THE AUTHOR

**PICHUGIN Yury A.**

*Saint-Petersburg State University of Aerospace Instrumentation*

61 Bolshaya Morskaya St., St. Petersburg, 190000, Russian Federation

yury-pichugin@mail.ru

## СПИСОК ЛИТЕРАТУРЫ

1. **Shannon C.E., Weaver W.** A mathematical theory of communication, Urbana University Press, Chicago, 1949. (Перевод на русский язык см. в Сборнике переводов «Теория передачи электрических сигналов при наличии помех» под ред. С.А. Железнова. М.: Изд-во иностранной литературы, 288 .1953 с.)

2. Гельфанд И.М., Яглом А.М. О вычислении количества информации о случайной функции, содержащейся в другой такой функции // Успехи математических наук. 1957. Т. 1 № .12 (73). С. 3–52.

3. Покровский О.М. Об оптимальных условиях косвенного зондирования атмосферы // Известия АН СССР. Физика атмосферы и океана. 1969. Т. 12 № .5. С. 1324–1326.

4. Кендал М.Дж., Стьюарт А. Статистические выводы и связи. М.: Наука, 899 .1973 с.

5. Себер Дж. Линейный регрессионный анализ. М.: Мир, 456 .1980 с.

6. Пичугин Ю.А. О связи количества информации по Шеннону с информационной матрицей Фишера в задаче планирования регрессионного эксперимента // Межвуз. сб. науч. трудов «Информатика – исследования и инновации». Вып. 3. СПб.: РГПУ им. А.И. Герцена, 1999. С. 32–36.

7. Пичугин Ю.А. Замечания к отбору данных в задачах, связанных с линейной множественной регрессией // Заводская лаборатория. Диагностика материалов. .2002 5 №. С. 61–62.

8. Математическая теория планирования эксперимента. Под ред. С.М. Ермакова. М.: Наука. Гл. ред. физ.-мат. лит., 1983. 392 с.

9. Ермаков С.М., Жиглявский А.А. Математическая теория оптимального эксперимента. М.: Наука. Гл. ред. физ.-мат. лит.-ры., 320 .1987 с.

10. Фон Нейман Дж., Моргенштерн О. Теория игр и экономическое поведение. М.: Наука. Гл. ред. физ.-мат. лит., 1970. 707 с.

11. Пичугин Ю.А. Замечания к использованию главных компонент в математическом моделировании // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2018. Т. 3 № .11. С. 89–74.

12. Пичугин Ю.А., Покровский О.М. О методе комплексирования наземной и спутниковой метеорологической информации // Исследование Земли из космоса. 6 № .1992. С. 25–31.

13. Дженкинс Г., Ваттс Д. Спектральный анализ и его приложения. Выпуск 1. М.: Мир, 316 – .1971 с.; Выпуск 2. М.: Мир, 1972. – 287 с.

14. **Pichugin Yu. A., Malafeyev O.A., Rylow D., Zaitseva I.** A statistical method for corrupt agents detection // International Conference on Numerical Analysis and Applied Mathematics (ICNAAM 2017). AIP Conf. Proc. 1978. Pp. 100014-1–100014-4.

15. Пичугин Ю.А. География динамической неустойчивости циркуляции атмосферы в Северном полушарии (моделирование и анализ) // Известия Русского географического общества. 2005. Т. 3 № .137. С. 12–16.

16. Пичугин Ю.А. Экологический мониторинг и методы многомерной



математической статистики // Астраханский вестник экологического образования. .2012 2 №. С. 105–101.

17. Гусман Ю.А., Пичугин Ю.А., Смирнов

А.О. Информативность по Шеннону в контроле микроэлектронной продукции // Вопросы радиоэлектроники. 10 № .2018. С. 10–6.

*Статья поступила в редакцию 22.05.2019, принята к публикации 14.06.2019.*

### **СВЕДЕНИЯ ОБ АВТОРЕ**

**ПИЧУГИН Юрий Александрович** – доктор физико-математических наук, профессор Института инноватики и базовой магистерской подготовки Санкт-Петербургского государственного университета аэрокосмического приборостроения.

190000, Российская Федерация, г. Санкт-Петербург, Большая Морская ул., 61.  
yury-pichugin@mail.ru