Discrete Dynamics in Nature and Society, Vol. 1, pp. 255–264
Reprints available directly from the publisher
Photocopying permitted by license only

Predictability Problems of Global Change as Seen through Natural Systems Complexity Description

2. Approach

VLADIMIR V. KOZODEROV [a,*], VICTOR A. SADOVNICHII [b], SERGEY A. USHAKOV [b] and OLEG A. TIMOSHIN [b]

[a] *Institute of Computational Mathematics, Russian Academy of Sciences, Gubkine Street 8, 117333 Moscow, Russia;*
[b] *M.V. Lomonosov Moscow State University, 119899 Moscow, Vorobiovy Gory, MSU, Russia*

(Received in final form 2 June 1997)

Developing the general statements of the proposed global change theory, outlined in Part 1 of the publication, Kolmogorov's probability space is used to study properties of information measures (unconditional, joint and conditional entropies, information divergence, mutual information, etc.). Sets of elementary events, the specified algebra of their sub-sets and probability measures for the algebra are composite parts of the space. The information measures are analyzed using the mathematical expectance operator and the adequacy between an additive function of sets and their equivalents in the form of the measures. As a result, explanations are given to multispectral satellite imagery visualization procedures using Markov's chains of random variables represented by pixels of the imagery. The proposed formalism of the information measures application enables to describe the natural targets complexity by syntactically governing probabilities. Asserted as that of signal/noise ratios finding for anomalies of natural processes, the predictability problem is solved by analyses of temporal data sets of related measurements for key regions and their background within contextually coherent structures of natural targets and between particular boundaries of the structures.

Keywords: Probability space, Information measures, Complexity

General statements of research programmes, concerning global change issues, were outlined in Part 1 of the publication. The statements were considered as composite parts of possible discrete dynamics application to study global change by the induced representations about information sub-spaces taking stringent terms of sets, measures and metrics (SMM) into account. Order and chaos categories in dynamic systems were described to develop conceptual models of global analysis, interpretation and modelling using the major framework concerning the SMM categories.

* Corresponding author.

Correct and incorrect problems were mathematically set up to find ways of their solutions existence, uniqueness and stability relative to disturbances of initial data, given by data of remote sensing measurements.

Below we present our approach to employ information measures and entropy metrics for describing the complexity of natural targets and structures construction to find ordering procedures while processing multispectral satellite imagery of the targets/structures. This will be needed to come to analysis of temporal data sets of the imagery that should approach us to understanding the predictability problems of global change.

KOLMOGOROV'S PROBABILITY SPACE

We shall operate in our considerations with the chaos and order characteristics of the conception of Kolmogorov's probability space. This space is defined when the following three categories are accepted known $(\Omega, \mathcal{F}, \mu)$, Ω is a set of elementary events with their composite elements ω, which are only considered in the classical probability theory; $\mathcal{F}$ is a special algebra, called as σ-algebra of the Ω sub-sets, that is associated with the σ-algebra of the A_k events; the latter are defined provided both their conjunction (sum) $\cup_k A_k$ and cross-section (product) $\cap_k A_k$ exist in the infinite sequence of the events; μ is a probability measure on the $\mathcal{F}$ algebra.

Random variables X with their particular meanings x on a finite set $\mathcal{X}$ can be then defined as a result of the following transformation $X: \Omega \to \mathcal{X}$, so that $X^{-1}(x) \in \mathcal{F}$ for all $x \in \mathcal{X}$. The probability of such an event in terms of these random variables is the μ-measure of a corresponding sub-set A of the Ω set, i.e.

$$\mathbf{P}\{X \in A\} = \mu(\{\omega: X(\omega) \in A\}).$$

We shall suppose hereafter that the major probability space $(\Omega, \mathcal{F}, \mu)$ is "sufficiently rich" in the sense that for any pair of infinite sets $\mathcal{X}$ and $\mathcal{Y}$, any random variable X with its meanings on the $\mathcal{X}$ and any distribution P on $\mathcal{X} \times \mathcal{Y}$ (this notation means the Cartesian product of the infinite sets), the striction of which on the $\mathcal{X}$ is coincident with the given P_X distribution, the Y random variable exists such as $P_{XY} = P$. This supposition is sure to be valid if Ω is the unit interval $(0, 1)$, $\mathcal{F}$ is a family of its Borel's sub-sets in the Euclidean space and μ is the Lebesgue measure (Kolmogorov and Fomin, 1976).

Let us identify sets of all probability distributions on the finite set $\mathcal{X}$ as sub-sets of the $n = |\mathcal{X}|$-measure Euclidean space that consists in all vectors with components $P_k \geq 0$ such as $\sum_k P_k = 1$. Linear combinations and convexity are then understood in accordance with the supposition. For instance, the convexity of a real function $f(P)$ from the probability distribution on $\mathcal{X}$ means that

$$f(\alpha P_1 + (1 - \alpha)P_2) \leq \alpha f(P_1) + (1 - \alpha)f(P_2)$$

for any distributions P_1, P_2 and $\alpha \in (0, 1)$.

It is possible to use topological terms for probability distributions on $\mathcal{X}$ assuming that they are related to a metrical topology characterized by Euclidean distance. In particular, the convergence $P_n \to P$ implies that $P_n(x) \to P(x)$ for any $x \in \mathcal{X}$.

We can introduce after these notations and definitions information measures having in mind their relevance to the well-known probability theory, from one side, and remote sensing data applications, from the other side. Of particular interest here for us are the forthcoming measures: unconditional, joint and conditional entropies, information divergence, mutual information, etc.

INFORMATION MEASURES

The scalar quantity of the amount of information in any set of measurements is defined through the mathematical expectance operator **E** in the following way:

$$H(X) = \mathbf{E}\left(\log\frac{1}{P(X)}\right) = \mathbf{E}(-\log P(X)) \equiv H(P) = -\sum_{x\in\mathcal{X}} P(x)\log P(x).$$

This is the alternative representation of the entropy of a random variable X or a probability distribution $P_X = P$. The entropy is the measure of an a priori uncertainty that is contained in the variable X before its measuring or observing; the main property of the measure of the amount of information is

$$0 \le H(P) \le \log|\mathcal{X}|.$$

Understanding of the entropy as the measure of uncertainty about a process under study is meant that a "more homogeneous" distribution would possess of larger entropy values. If two distributions P and Q are given on $\mathcal{X}$, then saying about the above "homogeneity", we imply that $P > Q$ provided that for any two non-decreasing orderings $p_1 \ge p_2 \ge \cdots \ge p_n$, $q_1 \ge q_2 \ge \cdots \ge q_n$ ($n = |\mathcal{X}|$) of probabilities from these distributions, the following inequality is valid for any k, $1 \le k \le n$:

$$\sum_{i=1}^{k} p_i \le \sum_{i=1}^{k} q_i,$$

so that from the condition $P > Q$, the other inequality entails:

$$H(P) \ge H(Q).$$

The information divergence, that is connected with statistical hypotheses testing, is the measure of differing between distributions P and Q and is also given by the $\mathbf{E}$ operator:

$$D(P,Q) = \mathbf{E}\left(\log\frac{P(x)}{Q(x)}\right) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}.$$

The conditional entropy is the measure of additional amount of information that contains in the random variable Y, if X has already been known, and is expressed through the joint entropy of pair of the variables and the unconditional entropy:

$$H(Y|X) = H(X,Y) - H(X).$$

Due to the definition

$$P_{X|Y}(x \mid y) = \frac{P_{XY}(x,y)}{P_X(x)},$$

for any $P_X(x) > 0$, the conditional entropy can be also written as:

$$H(Y \mid X) = \sum_{x\in\mathcal{X}} P_X(x) H(Y \mid X = x),$$

where

$$H(Y \mid X = x) = -\sum_{y\in\mathcal{Y}} P_{Y|X}(y \mid x)\log P_{Y|X}(y \mid x),$$

i.e. properties of the information categories enable to express the conditional entropy $H(Y \mid X)$ as the mathematical expectance of the entropy of the conditional distribution Y under the condition $X = x$.

Mutual information of the X and Y variables

$$I(X \wedge Y) = H(Y) - H(Y \mid X) = H(X) - H(X \mid Y)$$

serves as the measure of a stochastic dependence between these variables. We use the other letter for the measure notation (I instead of H in all other cases) just to follow traditional principles to do that. In particular, the formula

$$I(X \wedge X) = H(X)$$

expresses the amount of information that is contained in X relative to its own.

Other information measures of the type

$$H(X\,|\,Y, Z = z) = \sum_{y\in\mathcal{Y}} \mathbf{P}\{Y = y\,|\,Z = z\} H(X\,|\,Y = y, Z = z),$$

$$I(X \wedge Y|Z) = \sum_{z\in Z} \mathbf{P}\{Z = z\} I(X \wedge Y|Z = z), \ldots,$$

are often studied in various theoretical approaches (Csiszar and Korner, 1981).

It can be found that all the listed and any other information measures have the following general properties:

(1) are non-negative;
(2) are additive, i.e.

$$H(X, Y) = H(X) + H(Y\,|\,X),$$
$$H(X, Y\,|\,X) = H(X\,|\,Z) + H(Y\,|\,X, Z), \ldots;$$

(3) satisfy the "chain rules" for sequences of random variables:

$$H(X_1, \ldots, X_k) = \sum_{i=1}^{k} H(X_i\,|\,X_1, \ldots, X_{i-1}),$$

$$I(X_1, \ldots, X_k \wedge Y) = \sum_{i=1}^{k} I(X_i \wedge Y\,|\,X_1, \ldots, X_{i-1}), \ldots;$$

(4) $H(P)$ is the concave function of P and $D(P, Q)$ is the convex function of the pair (P, Q), i.e. if

$$P(x) = \alpha P_1(x) + (1 - \alpha) P_2(x)$$

and

$$Q(x) = \alpha Q_1(x) + (1 - \alpha) Q_2(x) \quad \text{for any } x \in \mathcal{X} \text{ and } 0 < \alpha < 1,$$

then

$$\alpha H(P_1) + (1 - \alpha) H(P_2) \leq H(P)$$

and

$$\alpha D(P_1, Q_1) + (1 - \alpha) D(P_2, Q_2) \geq D(P, Q).$$

Owing to their additive properties, these information measures can be considered as formal identities for the random variables. The adequacy has been proven (Csiszar and Korner, 1981) to exist between these identities to be valid for an optional additive function f and their equivalents in the form of the information measures. Denoting the sign of such adequacy by $\Leftrightarrow$, we can represent the proven facts as

$$H(X) \Leftrightarrow f(A), \quad H(X, Y) \Leftrightarrow f(A \cup B),$$
$$H(X|Y) \Leftrightarrow f(A\backslash B), \quad I(X \wedge Y) \Leftrightarrow f(A \cap B),$$

where $\cup$ and $\cap$ mean the conjunction and product of the A and B sets, respectively. In general, the theorem was proven in the cited reference that any pair of the information measures would be adequate to an expression of the following type $f((A \cap B)\backslash C)$ with the sign "backslash" denoting the sets disjunction, where A, B, C are infinite conjunctions of sets (supposed that A and B are not empty, C may be empty). And vice versa: any expression of the same type is adequate to the related information measure.

As a result of the facts, the following quantity $(A\backslash B) \cup (B\backslash A) = A \triangle B$, called as the symmetrical difference between sets A and B (Kolmogorov and Fomin, 1976), can be used in the studies. This quantity is a metrics of sub-sets A and B on the initial Ω set of Kolmogorov's space. The function $f((A\backslash B) \cup (B\backslash A))$ does not have any direct analog in the theory of information. However, it is the metrics for the quantity

$$d(X, Y) = H(X|Y) + H(Y\,|\,X)$$

on the random variables space. The said can be convinced by realizing that the metrics properties are correspondent to those given by initial axioms of metrical spaces:

$$d(X, Y) \geq 0, \quad d(X, X) = 0,$$
$$d(X, Y) = d(Y, X),$$
$$d(X, Y) + d(Y, Z) \geq d(X, Z).$$

It is not difficult to find that the information measures are continuous relative to the entropy metrics:

$$|H(X_1) - H(X_2)| \leq d(X_1, X_2),$$
$$|H(X_1|Y_1) - H(X_2 \mid Y_2)| \leq d(X_1, X_2) + d(Y_1, Y_2),$$
$$|I(X_1 \wedge Y_1) - I(X_2 \wedge Y_2)| \leq d(X_1, Y_1) + d(X_2, Y_2).$$

However, the information divergence $D(P, Q)$ cannot serve as the measure that satisfies the Euclidean distance requirements on the probability distribution space since it is not symmetrical. Even the "symmetrized" divergence $J(P, Q) = D(P, Q) + D(Q, P)$ is not the distance once such probability distributions P_1, P_2, P_3 can be found, for which simultaneously the following inequalities are valid:

$$D(P_1, P_2) + D(P_2, P_3) < D(P_1, P_3),$$
$$D(P_3, P_2) + D(P_2, P_1) < D(P_3, P_1).$$

Following the listed results, we can study properties of the information sub-spaces as imbedded into the main probability space that comprises random variables. This would require additional explanations to consider the sub-spaces relative to the distribution space because of the above concavity of the entropy and the convexity of the information divergence in the space. In fact, an opportunity is emerged in the first case to invent a unified description of different data sets representation for selected classes of natural targets using their transformed images, given by remote sensing measurements. The description is based on the SMM categories giving rise to imagery visualization procedures, which are usually implied while saying about the thematical interpretation of the images in a particular subject area. These procedures enable to find an analog to the subjective analysis of single satellite pictures by eyes of an experienced interpretor when an analyzed picture is displayed on the computer screen or as a hard copy. The rigorous definition of such visualization that also includes multispectral analysis, practically not accessible for the subjective interpretation, would originate from the SMM considerations in the metrical information sub-spaces.

IMAGERY VISUALIZATION

Of particular importance for an objective interpretation of natural targets variability on their space imagery are Markov's chains as an effective tool to identify "a recipe" of pixels ordering within a spatial structure. The above mutual information is the quantity that is the most profitable for an analysis of alterations on sets of pixels to be considered as sequences of these random variables. In accordance with its definition, a finite or infinite sequence of variables $X_1, X_2, \ldots$ with final sets of their values is called Markov's chain (Pougachev, 1979) if for any i the variable X_{i+1} is conditionally independent on $(X_1, \ldots, X_{i-1})$ relative to X_i. The latter notation is common-used in the information theory: if information measures are dependent on a set of random variables and these variables can be represented by the only symbol, the set is written as an argument without any parentheses. The parentheses are used to emphasize mutual information between the variables. Random variables $X_1, X_2, \ldots$ would generate a conditional Markov's chain relative to the variable Y if for any i the variable X_{i+1} is conditionally independent on $(X_i, \ldots X_{i-1})$ provided (X_i, Y). Both types of the chains serve to find elements of ordering on the images.

Since according to the definition of mutual information $I(X \wedge Y) = 0$ if variables X and Y are independent, and $I(X \wedge Y \mid Z) = 0$ if X and Y are conditionally independent variables relative to Z, then it can be stated that elements (pixels) of a multispectral image represented by $X_1, X_2, \ldots$ would make up Markov's chain there and only there where

$$I(X_1, \ldots, X_{i-1} \wedge X_{i+1} | X_i) = 0$$

for any i. The similar form of ascertaining the conditional Markov's chain relative to Y looks as

$$I(X_1, \ldots, X_{i-1} \wedge X_{i+1} | X_i, Y) = 0$$

Assuming random variables X_{i-1}, X_i, X_{i+1} as three levels of delineating pixels of one spectral band for an image while identifying rules of decision making as to the separability of the pixels and variable(s) Y in the above sense as related to the second band of the image, the search of pixels of its structure, for which the last "chain rule" is valid, would represent the essence of the visualization procedure for the two-band image. Finding of contextually coherent structures (Kozoderov, 1997) of natural targets and accounting for the related measures of the targets complexity description would be the result of these rules application. Considering random variables $X_1, \ldots, X_i$, X_{i+1}, Y as characteristics of a particular spectral band, the number of which is $i+2$, this search of the ordering measures using the chain rules would serve to elucidate the optimal selection of the number of bands and the efficiency of the relevant instruments called imaging spectrometers (Mission to Planet Earth, 1996). Both these aspects of the information measures applicability are needed to be realized in constructing new versions of special computer languages.

NATURAL TARGETS COMPLEXITY

The most pattern recognition and scene analysis techniques that would present the scientific basis for multispectral imagery processing are divided into two groups: one of them is tackled from the decision making position (Tou and Gonzalez, 1978) and the second is considered within the syntax approach (Fu, 1977). Natural objects (specific targets) are characterized by sets of numbers in the first case. These numbers are digital equivalents of results of remote sensing measurements. Pattern recognition as a procedure of attributing of each pixel on an image to some classes is carried out in this case by sub-dividing the entire space of characteristic features on selected areas to be delineated by sets of such rather standard procedures. Classes are to be defined in accordance with the probability distribution functions for sets of pixels on the scene under processing. It is required in the second case (the structural description of each pattern) that the recognition procedure would enable not only to take an object to a particular class, but to describe those peculiarities of the object, which would exclude its taking to any other class.

Developing the known metrical pattern recognition theory (Grenander, 1976; 1978), we can extend the definition of pixels in the techniques of the first case to elements $x \in \mathcal{X}$ in the second case. The recognition of the images in the second case is based on an analogy between "the structural patterns" (hierarchical or in a tree form) and the syntax of a computer language. The recognition in this case is in a syntax analysis of "a grammatical sentence" that describes a concrete scene under analysis. This scene is reflected by sets of various objects to be quantitatively described by the information measures. Such elements that can be called as generators are natural to be used for constructing configurations. It means that inducing a group of transformations on the set $\mathcal{X}$, a set of objects to be recognized is divided on classes of their equivalency. The configurations are determined by the composition and structure of their generators and by the combinatorial theory of the configurations construction on particular imagery of natural objects to be analyzed by the proposed treatments.

If it is possible to assume a structural combination of the generators into configurations, then these combined objects being characterized by the composition of bonds between the elements and by their own structures are initial to study new classes of the metrical images. The direct problem of studying processes of the images formation through mathematical operations of combination, identification and deformation is usually called the imagery synthesis whilst the inverse problem of selecting particular configurations on the images is called the imagery analysis (Grenander, 1978).

Denoting by $\mathcal{R}$ a system of rules or restrictions that are to define what configurations are regular, we can write the following symbolic expression for a computer language representation on a set of such regular configuration $Q(\mathcal{R})$:

$$L(\mathcal{R}) = (G, S, \Sigma, \rho),$$

where G is the generators set, S is the transformation set of the generators, Σ is the type of the bonds for the taken sets of generators, ρ is the ratio of consistency between the possible bonds in their structural connections. These bonds and connections may serve in the first approximation as a measure of the structure complexity. More comprehensive definition of the complexity in the matrix form will be given below. The regularity of these configurations on particular imagery is supposed in the studies as their consisting in specific structural connections not purely random for the elements; otherwise, no opportunity could be found in traditional supervising procedures of pattern recognition techniques and all attempts to create "an artificial intelligence" by finding regular rules of the element connections would be ambiguous.

There is the theorem (Grenander, 1976) that the tree type bonds Σ and the equality for the ratio of consistency ρ induce the above Markov's properties of the probability measures on sets of their regular configurations. These sets are understood in the sense of the existence of a topology τ on Kolmogorov's initial space (on the Ω set) when any system of sub-sets F should satisfy the following requirements (Kolmogorov and Fomin, 1976):

(i) the set Ω and the empty set $\varnothing$ belong to τ;
(ii) the sum $\cup_\alpha F_\alpha$ of any finite or infinite set and the cross-section $\cap_{k=1}^{n} F_k$ of any finite numbers of such sets from τ belong to the topology.

Three known axioms of separability are valid for such topological spaces $T = (\Omega, \tau)$ (Sadovnichii, 1979):

(1) neighbourhood $O(x)$ of a point x, not containing another point y, and neighbourhood $O(y)$ of the point y, not containing the point x, exist for any two points of the T space;
(2) any two points x and y of the T space have disjoint neighbourhoods $O(x)$ and $O(y)$ (the known Hausdorff's axiom);
(3) any point and any closed set, not containing the point, have disjoint neighbourhoods.

The above regularity of the configurative probability space is accepted by us as satisfying the axioms (1) and (3). Having these rules in mind, the syntactically governing probabilities can be defined as

$$1 > p_r > 0, \qquad \sum_{r \in \mathcal{R}_\xi} p_r = 1 \qquad \text{for any } \xi \in N,$$

where r is altered from 1 to ρ, N is the number of elements of the syntactical variables.

Now it is possible to introduce the complexity matrix for the grammatical rules of the description of the bonds for the regular configurations:

$$M = \{m_{ij}: i, j = 1, 2, \ldots, v\},$$

where $m_{ij} = \sum_{r \in \mathcal{R}_{\xi_i}} p_r n_j(r)$; $\mathcal{R}_{\xi_i}$ is the set of the permutation rules for the variable ξ with the i-index; $n_1(r), n_2(r), \ldots, n_v(r)$ are numbers of the appearance of the first, second, ..., vth syntactical variables, the total number of which is v for a testing grammatics.

Recalling that the entropy is a measure of any ordering, we can use it to write the following expression for the syntactically governing probabilities p_r in the attempts to order the grammatic rules:

$$h_i \equiv h(\xi_i) = -\sum_{r \in \mathcal{R}_{\xi_i}} p_r \log p_r, \quad i = 1, 2, \ldots, v.$$

The entropy of a style of imagery description can then be introduced as:

$$H_j \equiv H(\xi_j) = -\sum_{B \in \xi_j} P_j(B) \log P_j(B),$$

where probabilities P_j are to be known for any possible chain $J \in Q(\mathcal{R})$ for sets of outputs while fitting the style of the imagery description in the tree form. B denotes that using the proposed conception of generators, bonds and sets of regular configurations for the information measures, the rules of forming the description style of the taken computer language are to be selected only from the induced syntactically governing probabilities.

Returning to the above information measures relation

$$H(X, Y) = H(X) + H(Y|X)$$

and considering X as a random variable for a resulting grammatics and Y as that for the style of its description, one can obtain

$$H_i = h_i + \sum_{r \in \mathcal{R}_{\xi_i}} p_r \sum_{j=1}^{v} n_j(r) H_j$$

or in the matrix form

$$\mathbf{H} = [\mathbf{I} - \mathrm{M}]^{-1}\mathbf{h},$$

where the matrix $\mathbf{H}$ of ordering of the style description is expressed through the matrix $\mathbf{h}$ of ordering of the syntax of imagery by the inverse matrix that is equal to the difference between the unit matrix $\mathbf{I}$ and the complexity matrix $\mathbf{M}$. This gives a rule for testing a special language of the structural imagery description (Kozoderov, 1997).

To sum up the results, we can say about the applications of the related languages for testing them while describing the contextually coherent structures mentioned above. It is worthwhile to evolve the techniques for computer work stations to proceed from these improvements to the final stage of the multispectral satellite imagery analysis. This stage is concerned the signal/noise ratios extraction from temporal sets of the consequent images with the structures, predescribed in accordance with the given complexity procedures. Analyzing temporal sets of satellite imagery, given by multispectral radiometers of different spatial resolutions, and describing the imagery structures by the proposed techniques, we are able to proceed from the regional structures description for selected natural targets to their global changes.

Let us add a few words about the complexity category.

Discussing the incorrect problems of data sets interpretation in Part 1 of the paper, we used the complexity functionals in the regularization techniques. These functionals were applied for finding the models, which satisfying to the general formalism of solutions of the inverse problems would be of "the minimal complexity". The last term was utilized there to select those specific models from sets of similar other models, which would be comparable with accuracies of data of the observations that were to be fitted to theoretical results of modelling. More complicated models in this sense could be less consistent with the relevant set of observations than these models of the minimal complexity.

Saying about rules and restrictions in the structural conjunction of the proposed "standard blocks" (generators) here, we have introduced the structural complexity of configurations, which are regular in the sense of how one set of these "details" could be imbedded into the other of the higher level construction. We can use, if necessary, the definition of "the quantitative complexity" of a configuration, just simply counting the number of generators in the configuration. By the general expression of the complexity in the matrix form above, we determined grammatic rules of the complexity description by using the syntax and style of the language in the analysis of the structures on multispectral satellite images for selected classes of natural targets. The term "complexity of terrestrial ecosystems" in the International Global Change and Terrestrial Ecosystems (GCTE) project (Kozoderov, 1995) is in fact identical to the biological diversity. Our intentions are to describe changes in terrestrial ecosystems by the overall natural systems complexity matrix while using regular observations of the systems. Biodiversity changes would inevitably result in the observable changes on remote sensing images. Thus, we do have an opportunity to filter out all seasonal harmonics of vegetation growth on the images and deal with "signals" of their possible change by the proposed below application of discrete dynamics techniques.

PREDICTABILITY OF GLOBAL CHANGE

The scientific basis for solving the predictability problems is given by the cross-correlation techniques to find asynchronous correlations of anomalies of the fields under study (outgoing long-wave radiation, the biomass amount of vegetation, etc.). These correlations are represented in the following form of the signal/noise ratio for two autoregression Markov's processes of the first order (Marchuk *et al.*, 1990):

$$\alpha(\tau) = \frac{R_{XY}(\tau)}{\mathrm{SD}(R_{XY}(\tau))}.$$

Here

$$R_{XY}(\tau) = \frac{1}{n-\tau}\sum_{k=1}^{n-\tau}(X_k - \bar{X}_T)(Y_{k+\tau} - \bar{Y}_T)/S_{XY}$$

are cross-correlations of the anomalies (deviations from the quadric standard) of the studied quantities for n observations and different shifts with time τ, $\mathrm{SD}(C_{XY})$ is the standard deviation of observation covariances for two intervals on time,

$$C_{XY} = R_{XY}(\tau)S_{XY}$$

is the sampling estimate of the cross-covariance for the two discrete processes X and Y,

$$S_{XY} = \left(\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_T)^2 \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \bar{Y}_T)^2\right)^{1/2};$$

$$\bar{X}_T = \frac{1}{n}\sum_{k=1}^{n} X_k, \quad \bar{Y}_T = \frac{1}{n}\sum_{k=1}^{n} Y_k.$$

The index T characterizes the averaging procedure for the two data sets.

The analytical solution for the problem with two Markov's random processes is known as (see the cited reference)

$$\mathrm{SD}(C_{XY}) = \sigma_{\mathrm{F}}\left(\frac{2}{n(1-\exp[-(\lambda_X + \lambda_Y)])}\right)^{1/2},$$

where

$$\mathrm{Cov}(C_{XY}(\tau_1), C_{XY}(\tau_2)) = \frac{1}{N_{XY}}\sum_{k=-\infty}^{\infty}(\gamma_{XX}(k)\gamma_{YY}(k+\tau_1+\tau_2) + \gamma_{XY}(k+\tau_1)\gamma_{YX}(k-\tau_2))$$

is the covariance of the processes under two shifts τ_1 and τ_2, N_{XY} is the effective number of pairs of the processes under the shifts,

$$\gamma_{XX}(\tau) = \sigma_{\mathrm{F}}^2(X)\exp(-\lambda_X\tau),$$
$$\gamma_{YY}(\tau) = \sigma_{\mathrm{F}}^2(Y)\exp(-\lambda_Y\tau);$$

$\sigma_{\mathrm{F}}^2(X)$ and $\sigma_{\mathrm{F}}^2(Y)$ are dispersions of the analyzed fields for different periods of time.

The statistical confidence of the mutual correlation coefficients from the formulae on the 95% of the confidence level is given by

$$R_{XY}(0) > 2\mathrm{SD}(C_{XY})/\sigma_{\mathrm{F}}.$$

The predictability of the process $X(t)$ via $Y(t)$, both are represented in the discrete form with time t, is then determined by the expression

$$P_{XY}(\tau) = \frac{\alpha_{XY}(\tau)}{\alpha_{YY}(0)} = \frac{R_{XY}(\tau)}{R_{YY}(\tau)}\left(\frac{N_{XY}(1-\exp[-(\lambda_X+\lambda_Y)])}{N_{YY}(1-\exp(-2\lambda_Y))}\right)^{1/2}$$

The statistical significance of the anomalies of the fields under study is defined here by the number of independent samples that is equal to

$$N_{\mathrm{f}} = \frac{N_{XY}}{\sum_{k=-\infty}^{\infty} R_{XX}(k)R_{YY}(k)}.$$

In the final run, the predictability problem is reduced to finding the cross-correlations between "particular points" of multispectral remote sensing imagery and their background. Selection of these points and analysis of the structures, that comprise the points, is the subject of the above visualization techniques. The next stage of imagery processing is to retrieve state parameters of natural objects, classified in accordance with routine pattern recognition and scene analysis techniques using relevant imagery transformation procedures (Curran *et al.*, 1990). The final stage is in the temporal data sets analysis that would enable to understand the predictability problem in the way presented here. All these stages of information and mathematical applications for data sets of satellite imagery interpretation are an example of advances in the multidisciplinary description of natural processes.

CONCLUSION

The information and mathematical aspects of global change, presented in the publication, are demonstrated by the unified approach how to compare sets of data in information sub-spaces and to understand predictive capabilities in solving the problem. In spite of the formal information theory does not enable to solve all problems of satellite imagery interpretation, we have elaborated techniques to describe the complexity problem of natural structures using the mathematical formalism of tackling with sets of data, information measures and entropy metrics. We employ the known axiomatics of Kolmogorov's probability space to emphasize the discrepancy of our approach with tendencies of pure numerical applications in current international scientific programmes of global change. Our studies are designed to remove deficiencies given by the common-used information theory, which are due to the assumption that a structure under study is finite. Our improvements of the classical theory gets possible owing to the proven possibilities to unify different data of observations in terms of sets, measures and metrics. The opportunity to account for scientifically the comparability, ordering and calculation measures enables us to extend existing knowledge about the natural structures description. Adherent to updated views on order, chaos and similar other categories in natural dynamic systems, we have shown how these categories are represented in metrical spaces for our purpose to find a regularity on structures of natural targets represented in information sub-spaces by these targets imagery. Our main intention in future is to elucidate the problem of "genesis" of information and situations using the unified description of natural processes by the proposed way. Thus, we are approaching the understanding of global change based on major achievements in information and mathematical sciences.

References

Curran, P.J., Foody, G.M., Kondratyev, K.Ya., Kozoderov, V.V. and Fedchenko, P.P. (1990). *Remote Sensing of Soils and Vegetation in the USSR*. Taylor & Francis, London, New York, Philadelphia, p. 204.

Csiszar, I. and Korner, J. (1981). *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, Budapest, Hungary, p. 398.

Grenander, U. (1976). *Pattern Synthesis: Lectures in Pattern Theory*, Vol. I. Springer-Verlag, New York, Heidelberg, Berlin, p. 384.

Grenander, U. (1978). *Pattern Analysis: Lectures in Pattern Theory*, Vol. II. Springer-Verlag, New York, Heidelberg, Berlin, p. 446.

Fu, K. (1977). *Structural Methods in Pattern Recognition*. Mir Publ., Moscow, p. 320.

Kolmogorov, A.N. and Fomin, S.V. (1976). *Elements of Theory of Functions and the Functional Analysis*. Nauka Publ., Moscow, p. 544.

Kozoderov, V.V. (1995). A scientific approach to employ monitoring and modelling techniques for Global Change and Terrestrial Ecosystems and other related projects. *J. Biogeography* **22**, 927–933.

Kozoderov, V.V. (1997). Information aspects of studies of the Earth from space. *Earth Observations and Remote Sensing* **2**, 27–36.

Marchuk, G.I., Kondratyev, K.Ya. and Kozoderov, V.V. (1990). *Radiation Budget of the Earth: Key Aspects*. Nauka Publ., Moscow, p. 224.

Mission to Planet Earth: *Science Research Plan*. (1996). NASA, Washington DC, USA, p. 158.

Pougachev, V.S. (1979). *Theory of Probabilities and Mathematical Statistics*. Nauka Publ., Moscow, p. 496.

Sadovnichii, V.A. (1979). *Theory of Operators*. Moscow State University Publ., Moscow, p. 296.

Tou, J. and Gonzalez, R. (1978). *Pattern Recognition Principles*. Mir Publ., Moscow, p. 412.