ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2023, Vol. 84, No. 1, pp. 64–81.

INTELLECTUAL CONTROL SYSTEMS, DATA ANALYSIS

Randomized Machine Learning Algorithms to Forecast the Evolution of Thermokarst Lakes Area in Permafrost Zones

Yu. A. Dubnov[*,**,a], **A. Yu. Popkov**[*,b], **V. Yu. Polishchuk**[***,c], **E. S. Sokol**[****,d], **A. V. Melnikov**[****,e], **Yu. M. Polishchuk**[****,f], **and Yu. S. Popkov**[*,g]

[*] *Federal Research Center "Computer Science and Control," Russian Academy of Sciences, Moscow, Russia*
[**] *National Research University Higher School of Economics, Moscow, Russia*
[***] *Institute of Monitoring of Climatic and Ecological Systems, Siberian Branch, Russian Academy of Sciences, Tomsk, Russia*
[****] *Yugra Research Institute of Information Technologies, Khanty-Mansiysk, Russia*
e-mail: [a] *yury.dubnov@phystech.edu,* [b] *apopkov@isa.ru,* [c] *liquid_metal@mail.ru,* [d] *sokoles@uriit.ru,* [e] *melnikovav@uriit.ru,* [f] *yupolishchuk@gmail.com,* [g] *popkov@isa.ru*

Received April 20, 2022
Revised June 21, 2022
Accepted September 29, 2022

Abstract—Randomized machine learning focuses on problems with considerable uncertainty in data and models. Machine learning algorithms are formulated in terms of a functional entropy-linear programming problem. We adapt these algorithms to forecasting problems on an example of the evolution of thermokarst lakes area in permafrost zones. Thermokarst lakes generate methane, a greenhouse gas affecting climate change. We propose randomized machine learning procedures using dynamic regression models with random parameters and retrospective data (climatic parameters and remote sensing of the Earth's surface). The randomized machine learning algorithm developed below estimates the probability density functions of model parameters and measurement noises. Randomized forecasting is implemented as algorithms transforming the optimal distributions into the corresponding random sequences (sampling algorithms). The randomized forecasting procedures and technologies are trained, tested, and then applied to forecast the evolution of thermokarst lakes area in Western Siberia.

Keywords: thermokarst lakes, remote sensing, information entropy, balance equations, dynamic regression, optimization, Lyapunov-type problem, sampling, randomized forecasting, randomized machine learning

DOI: 10.25728/arcRAS.2023.21.29.001

1. INTRODUCTION

The problems, technologies, and algorithms of machine learning were considered in numerous publications. For general ideas, see the monographs [1–3]. Randomized machine learning (RML) focuses on reconstructing the parameters of dependencies *under uncertainty* in initial data (incomplete data, errors with unknown characteristics, different timescales, etc.) and models (insufficient knowledge of the processes, structural ambiguity, uncertain key parameters, etc.) [4]. In contrast to machine learning, the randomized version is based on estimating the probability density functions (PDFs) of the model parameters and measurement noises corresponding to the maximum uncertainty in terms of information entropy.

Thermokarst lakes are an object with high-level uncertainty. Note that the formation and evolution processes of thermokarst lakes have been insufficiently studied [5, 6] and historical data about them, especially satellite-derived ones, are accompanied by considerable errors [7–9].

This paper proposes a randomized technology to forecast the evolution of thermokarst lakes area. Its essence consists in generating optimized ensembles of forecast trajectories by the sampling of the entropy-optimal PDFs of the parameters of randomized dynamic models. The optimal characteristics of the models are determined using real historical data based on conditionally maximal information entropy estimates.

2. THE GENERAL STRUCTURE OF THE ENTROPY-RANDOMIZED MODELING AND FORECASTING PROCEDURE

Randomization as a means of imparting artificial, rationally organized random properties to inherently non-random events, indicators, and methods is a fairly common and effective technique. There are many examples in various fields of science, management, and economics: randomized numerical optimization methods [10, 11], mixed (random) strategies in stock trading [12], the randomized prediction of population dynamics [13], and vibrational control of industrial processes [14]. Randomization implies giving non-random objects artificial stochastic properties with optimal probabilistic characteristics in a definite sense. Choosing quantitative characteristics of optimality always turns out to be debatable and ambiguous. It requires arguments that somehow reflect the important peculiarities of the randomized object. In particular, a fundamental feature of modeling and forecasting procedures is the uncertainty in the data used, predictive models, forecasting methods, etc.

In this paper, we adopt *information entropy* [15] as a characteristic of uncertainty. According to the first law of thermodynamics, entropy is a natural function describing universal evolutionary processes; for example, see [16–18]. By the second law of thermodynamics, entropy maximization determines the best state of an evolutionary process under the worst-case impact on it (the maximum uncertainty). Also, we mention another quality of information entropy, related to measurement and other errors, which are important characteristics of data. With information entropy used to consider the effect of such errors, it is possible to estimate the probabilistic characteristics of noises with the worst-case impact on forecasting procedures [19].

2.1. Randomized Modeling

Randomized modeling rests on mathematical input-output models with random parameters. This stage of the technology yields a basic randomized parametric model (*RPM-B*), associated with the available historical input-output data, and an auxiliary model of the same class (*RPM-A*), intended to reproduce the input of the basic model via a "suitable" input process for *RPM-A*. As such, a random sequence with optimized properties is often used.

Consider a basic randomized parametric model (*RPM-B*) and historical data arrays compatible with it. *RPM-B* transforms a historical input data array $X = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}]$, where $\mathbf{x}^{(j)} \in \mathrm{R}^n$, into a model output with a matrix $\hat{Z} = [\hat{\mathbf{z}}^{(1)}, \ldots, \hat{\mathbf{z}}^{(s)}]$, where $\mathbf{z}^{(j)} \in \mathrm{R}^m$.

Generally, this transformation is assumed dynamic: the model output observed at a time instant j depends on the input observed on some *historical* time interval $j - \varrho, \ldots, j$, i.e., on the matrix $X_\varrho^{(j)} = [\mathbf{x}^{(j-\varrho)}, \ldots, \mathbf{x}^{(j)}]$. This relationship is mathematically expressed through a vector functional $\hat{\Omega}(X_\varrho^{(j)}, \mathbf{a} \,|\, P(\mathbf{a}))$ with random parameters $\mathbf{a} \in R^d$ of the interval type:

$$\mathbf{a} \in \mathcal{A} = [\mathbf{a}^-, \mathbf{a}^+]. \tag{2.1}$$

The probabilistic properties of the parameters are characterized by a continuously differentiable PDF $P(\mathbf{a})$.

The *RPM-B* output at a time instant j (measurements) is an ensemble $\hat{\mathbb{Z}}^{(j)}(P(\mathbf{a}))$ of random vectors

$$\hat{\mathbf{z}}^{(j)}(\mathbf{a}) = \hat{\Omega}(X_\varrho(j), \mathbf{a} \,|\, P(\mathbf{a})), \qquad j = \overline{1,s}. \tag{2.2}$$

The observed *RPM-B* output can be represented as

$$\mathbf{v}^{(j)}(\mathbf{a}, \xi^{(j)}) = \hat{\mathbf{z}}^{(j)}(\mathbf{a}) + \xi^{(j)}, \qquad j = \overline{1,s}, \tag{2.3}$$

where measurement noises $\xi^{(j)} \in R^m$ of the interval type,

$$\xi^{(j)} \in \Xi_j = [\xi_-^{(j)}, \xi_+^{(j)}], \qquad j = \overline{1,s}, \tag{2.4}$$

have continuously differentiable PDFs $Q_j(\xi^j)$, $j = \overline{1,s}$. In accordance with these functions, the model generates an ensemble $\mathbb{K}^{(j)}(Q_j(\xi^j))$ for each measurement instant of the object's output.

The random vectors (2.3) form an ensemble $\mathbb{V}^{(j)}(P(\mathbf{a}), Q_j(\xi^{(j)}))$ with the mathematical expectation

$$\begin{aligned}\mathcal{M}\left\{\mathbf{v}^{(j)}(\mathbf{a}, \xi^{(j)})\right\} &= \int_{\mathcal{A}} \hat{\mathbf{z}}^{(j)}(\mathbf{a}, P(\mathbf{a}))d\mathbf{a} + \int_{\Xi_j} Q_j(\xi^{(j)})\xi^{(j)}d\xi^{(j)} \\ &= \varphi^{(j)}\left[P(\mathbf{a}), Q_j(\xi^{(j)})\right], \quad j = \overline{1,s}.\end{aligned} \tag{2.5}$$

2.2. Training of RPMs

An RPM is trained by *estimating the PDFs* of its parameters and measurement noises based on available data. This stage is implemented using randomized machine learning algorithms (*RML-A*); for details, see [19, 20].

For **RPM-B**, the algorithm has the form

$$[P^*(\mathbf{a}), Q^*(\xi)] = \arg \max_{P(\mathbf{a}), Q(\xi)} \mathcal{H}[P(\mathbf{a}), Q(\xi)] \tag{2.6}$$

on the set of all normalized functions $P^*(\mathbf{a})$ and $Q(\xi)$ satisfying the empirical balance conditions

$$\varphi_j\left[P(\mathbf{a}), Q_j(\xi^{(j)})\right] = \mathbf{y}^{(j)}, \quad j = \overline{1,s}, \tag{2.7}$$

for the expectations (2.5), where $\mathbf{y}^{(j)} \in R^m$ is the real measurement vector of the object's output and $\xi = \{\xi^{(1)}, \ldots, \xi^{(s)}\}$.

The estimation quality of these *PDFs* is characterized by an entropy functional of the form

$$\mathcal{H}[P(\mathbf{a}), Q(\bar{\xi})] = -\int_{\mathcal{A}} P(\mathbf{a}) \ln P(\mathbf{a})\, d\mathbf{a} - \sum_{j=1}^{s} \int_{\Xi_j} Q_j(\xi^{(j)}) \ln Q_j(\xi^{(j)})\, d\xi^{(j)}. \tag{2.8}$$

Problem (2.6), (2.7) belongs to the class of functional entropy-linear problems of the Lyapunov type [21]. It therefore has an analytical solution obtained using the Lagrange multipliers $\Theta = [\theta^j,\ j = \overline{1,s}]$ (vectors $\theta^j \in R^m$):

$$\begin{aligned}P^*(\mathbf{a}) &= \frac{\exp\left(-\sum\limits_{j=1}^{s} \langle \theta^{(j)}, \hat{\mathbf{z}}^{(j)}(\mathbf{a})\rangle\right)}{\mathcal{P}(\Theta)}, \\ Q_j^*(\xi^{(j)}) &= \frac{\exp\left(-\langle \theta^{(j)}, \xi^{(j)}\rangle\right)}{\mathcal{Q}_j(\theta^{(j)})},\ j = \overline{1,s}; \quad Q(\xi) = \prod_{j=1}^{s} Q_j^*(\xi^{(j)}).\end{aligned} \tag{2.9}$$

The denominators of these expressions contain the normalizing constants

$$\mathcal{P}(\Theta) = \int\limits_{\mathcal{A}} \exp\left(-\sum_{j=1}^{s}\langle\theta^{(j)}, \hat{\mathbf{z}}^{(j)}(\mathbf{a})\rangle\right) d\mathbf{a},$$

$$\mathcal{Q}_j(\theta^{(j)}) = \int\limits_{\Xi_j} \exp\left(-\langle\theta^{(j)}, \xi^{(j)}\rangle\right) d\xi^{(j)}, \quad j = \overline{1,s}. \tag{2.10}$$

The optimal PDFs and the normalizing constants are parametrized by the Lagrange multipliers satisfying the balance equations

$$\frac{\mathcal{U}_j(\Theta)}{\mathcal{P}(\Theta)} + \frac{\mathcal{T}_j(\theta^{(j)})}{\mathcal{Q}_j(\theta^{(j)})} = \mathbf{y}^{(j)}, \quad j = \overline{1,s}, \tag{2.11}$$

where

$$\mathcal{U}_j(\Theta) = \int\limits_{A} \hat{\mathbf{z}}^{(j)}(\mathbf{a}) \exp\left(-\sum_{j=1}^{s}\langle\theta^{(j)}, \hat{\mathbf{z}}^{(j)}(\mathbf{a})\rangle\right) d\mathbf{a},$$

$$\mathcal{T}_j(\theta^{(j)}) = \int\limits_{\Xi_j} \xi^{(j)} \exp\left(-\langle\theta^{(j)}, \xi^{(j)}\rangle\right) d\xi^{(j)}, \qquad j = \overline{1,s}. \tag{2.12}$$

2.3. Testing of the Trained RPMs

The optimized RPMs are tested using historical data sequences by the sampling of the entropy-optimized PDFs of the model parameters and measurement noises. Sampling involves transforming the PDFs into suitable sequences of random vectors.

The general method for generating sequences of random vectors with a given PDF was described in [22]. As a result, Monte Carlo simulations yield an ensemble of random trajectories of the observed RPM output under the worst-case measurement noises (in entropy terms).

The empirical trajectories in the testing procedures are averaged over the ensemble (the *mean* trajectories):

$$\bar{\mathbf{v}}[k] = \frac{1}{M}\sum_{i=1}^{M} \hat{\mathbf{v}}^{(i)}[k], \qquad k = \overline{0,N}, \tag{2.13}$$

where M denotes the number of trajectories in the ensemble.

The training quality of these RPMs is characterized by the absolute error

$$\Delta = \sqrt{\sum_{k=1}^{M} \left(\bar{\mathbf{v}}[k] - \mathbf{v}^r[k]\right)^2} \tag{2.14}$$

and (or) the relative error

$$\delta = \frac{\sum\limits_{k=1}^{M} \left(\bar{\mathbf{v}}[k] - \mathbf{v}^r[k]\right)^2}{\sqrt{\sum\limits_{k=1}^{M} \bar{\mathbf{v}}^2[k]} + \sqrt{\sum\limits_{k=1}^{M} (\mathbf{v}^r[k])^2}}, \tag{2.15}$$

where $\mathbf{v}^r[k]$ are historical test data.

2.4. Randomized Forecasting

The application of dynamic input-output models to forecasting requires some modification: in this case, we need a forecast of the input process. It can be obtained using the concept of entropy-randomized machine learning.

Consider an auxiliary randomized parametric model (*RPM-A*). In contrast to *RPM-B*, its input is a random matrix $\Gamma = \left[\gamma^{(1)}, \ldots, \gamma^{(s)}\right]$ composed of random vectors $\gamma^{(j)} \in R^n$ of the interval type:

$$\gamma^{(j)} \in \mathcal{G}_j = [\gamma_-^{(j)}, \gamma_+^{(j)}], \qquad \Gamma \in \mathcal{G} = \bigcup_{j=1}^{s} \mathcal{G}_j. \tag{2.16}$$

RPM-A transforms a random matrix Γ into another random matrix $X = \left[\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(s)}\right]$. By analogy with *RPM-B*, this relationship is mathematically expressed through a vector functional $\hat{\Psi}\left(\Gamma_\varrho^{(j)}, \mathbf{b} \,|\, W(\Gamma_\varrho^{(j)}, \mathbf{b})\right)$ with a matrix $\Gamma_\varrho(j) = \left[\gamma^{(j-\varrho)}, \ldots, \gamma^{(j)}\right]$ and parameters $\mathbf{b} \in R^p$ of the interval type:

$$\mathbf{b} \in \mathcal{B} = [\mathbf{b}^-, \mathbf{b}^+]. \tag{2.17}$$

We characterize the probabilistic properties of the functional $\hat{\Psi}$ by a joint PDF of the parameters $\mathbf{b}$ and matrix $\Gamma_\varrho^{(j)}$, denoted by $W(\Gamma_\varrho^{(j)}, \mathbf{b})$. Assume that this PDF is continuously differentiable as well.

The *RPM-A* output at a time instant j (measurements) is an ensemble $\hat{\mathbb{X}}^{(j)}(W(\Gamma_\varrho(j), \mathbf{b})$ of random vectors

$$\hat{\mathbf{x}}^{(j)}\left(\Gamma_\varrho^{(j)}, \mathbf{b}\right) = \hat{\Psi}\left(\Gamma_\varrho^{(j)}, \mathbf{b} \,|\, W(\Gamma_\varrho^{(j)}, \mathbf{b})\right), \qquad j = \overline{1, s}. \tag{2.18}$$

The observed *RPM-A* output can be represented as

$$\mathbf{f}^{(j)}\left(\Gamma_\varrho^{(j)}, \mathbf{b}, \eta^{(j)}\right) = \hat{\mathbf{x}}^{(j)}\left(\Gamma_\varrho^{(j)}, \mathbf{b}\right) + \eta^{(j)}, \qquad j = \overline{1, s}, \tag{2.19}$$

where measurement noises $\eta^{(j)} \in R^m$ of the interval type,

$$\eta^{(j)} \in \mathcal{E}_j = \left[\eta_-^{(j)}, \eta_+^{(j)}\right], \qquad j = \overline{1, s}, \tag{2.20}$$

have continuously differentiable PDFs $E_j(\eta^{(j)})$, $j = \overline{1, s}$. In accordance with these functions, the model generates an ensemble $\mathbb{E}^{(j)}(E_j(\eta^{(j)}))$ for each measurement instant of the object's output.

The random vectors (2.19) form an ensemble $\mathbb{F}^{(j)}(W(\Gamma_\varrho^{(j)}, \mathbf{b}), E_j(\eta^{(j)}))$ with the mathematical expectation

$$\mathcal{M}\left\{\mathbf{f}^{(j)}(\Gamma_\varrho^{(j)}, \mathbf{b}, \eta^{(j)})\right\} = \int\limits_{\mathcal{B}\bigcap\mathcal{G}} W\left(\Gamma_\varrho^{(j)}, \mathbf{b}\right) \hat{\mathbf{x}}^{(j)}\left(\Gamma_\varrho^{(j)}, \mathbf{b},\right) d\Gamma_\varrho^{(j)}\, d\mathbf{b}$$

$$+ \int\limits_{\mathcal{E}_j} E_j(\eta^{(j)}) \eta^{(j)} d\eta^{(j)} = \psi^{(j)}\left[W(\Gamma_\varrho^{(j)}, \mathbf{b}), E_j(\eta^{(j)})\right], \quad j = \overline{1, s}. \tag{2.21}$$

For **RPM-A,** the randomized machine learning algorithm has the form

$$\left[W^*(\Gamma_\varrho^{(j)}, \mathbf{b}), E^*(\eta)\right] = \arg \max_{W(\Gamma_\varrho^{(j)}, \mathbf{b}), E(\eta)} \mathcal{H}\left[W(\Gamma_\varrho, \mathbf{b}), E(\eta)\right] \tag{2.22}$$

on the set of all normalized functions $W(\Gamma_\varrho^{(j)}, \mathbf{b})$ and $E(\eta)$ satisfying the empirical balance conditions

$$\psi_j \left[W(\Gamma_\varrho^{(j)}, \mathbf{b}), E_j(\eta^{(j)}) \right] = \mathbf{y}^{(j)}, \quad j = \overline{1,s}. \tag{2.23}$$

The entropy functional in (2.22) has the form

$$\mathcal{H}[W(\Gamma_\varrho, \mathbf{b}), E(\eta)] = -\sum_{j=1}^{s} \int\limits_{\mathcal{B}\bigcap\mathcal{G}_j} W(\Gamma_\varrho^{(j)}, \mathbf{b}) \ln W(\Gamma_\varrho^{(j)}, \mathbf{b})\, d\Gamma_\varrho^{(j)}\, d\mathbf{b}$$
$$- \sum_{j=1}^{s} \int\limits_{\mathcal{E}_j} E_j(\eta^{(j)})\, \ln E_j(\eta^{(j)})\, d\eta^{(j)}. \tag{2.24}$$

We denote by $\Lambda \left\{ \lambda^{(1)}, \ldots, \lambda^{(s)} \right\}$, $\lambda^{(j)} \in R^m$, the Lagrange multipliers for problem (2.22, 2.23). Then its solution can be written as

$$W^*(\Gamma_\varrho^{(j)}, \mathbf{b}) = \frac{\exp\left(-\langle \lambda^{(j)}, \hat{\mathbf{x}}^{(j)}(\Gamma_\varrho^{(j)}, \mathbf{b})\rangle\right)}{\mathcal{W}_j(\lambda^{(j)})}, \quad W^*(\Gamma_\varrho, \mathbf{b}) = \prod_{j=1}^{s} W^*(\Gamma_\varrho^{(j)}, \mathbf{b}),$$
$$E_j^*(\eta^{(j)}) = \frac{\exp\left(-\langle \lambda^{(j)}, \eta^{(j)}\rangle\right)}{\mathcal{E}_j(\lambda^{(j)})}, \quad j = \overline{1,s}; \qquad E(\eta) = \prod_{j=1}^{s} E_j^*(\eta^{(j)}). \tag{2.25}$$

The denominators of these expressions contain the normalizing constants

$$\mathcal{W}_j(\lambda^{(j)}) = \int\limits_{\mathcal{B}\bigcap\mathcal{G}} \exp\left(-\langle \lambda^{(j)}, \hat{\mathbf{x}}^{(j)}(\Gamma_\varrho^{(j)}, \mathbf{b})\rangle\right) d\Gamma_\varrho^{(j)}\, d\mathbf{b},$$
$$\mathcal{E}_j(\lambda^{(j)}) = \int\limits_{\mathcal{E}_j} \exp\left(-\langle \lambda^{(j)}, \eta^{(j)}\rangle\right) d\eta^{(j)}, \quad j = \overline{1,s}. \tag{2.26}$$

Hence, the input auxiliary noises and the model parameters have interconnected probabilistic properties for some nonlinear functional $\hat{\Psi}$.

Due to (2.7), the optimal PDFs are parametrized by the Lagrange multipliers Λ satisfying the balance equations

$$\frac{\mathcal{N}_j(\lambda^{(j)})}{\mathcal{W}_j(\lambda^{(j)})} + \frac{\mathcal{S}_j(\lambda^{(j)})}{\mathcal{E}_j(\lambda^{(j)})} = \mathbf{y}^{(j)}, \quad j = \overline{1,s}, \tag{2.27}$$

where

$$\mathcal{N}_j(\lambda^{(j)}) = \int\limits_{\mathcal{B}\bigcap\mathcal{G}_j} \hat{\mathbf{z}}(\Gamma_\varrho^{(j)}, \mathbf{b}) \exp\left(-\langle \theta^{(j)}, \hat{\mathbf{z}}(\Gamma_\varrho^{(j)}, \mathbf{b})\rangle\right) d\Gamma_\varrho^{(j)}, d\mathbf{b},$$
$$\mathcal{S}_j(\lambda^{(j)}) = \int\limits_{\mathcal{E}_j} \eta^{(j)} \exp\left(-\langle \lambda^{(j)}, \eta^{(j)}\rangle\right) d\eta^{(j)}, \qquad j = \overline{1,s}. \tag{2.28}$$

Let the PDF of the model parameters and the input sequence be determined. Then the ensembles of forecast trajectories are generated by their sampling; for details, see [22].

Sampling implies transforming the PDF into a suitable sequence of random vectors. The following empirical probabilistic and numerical characteristics of ensembles are used in randomized forecasting procedures:

- the empirical probability density functions (*ePDFs*) $\mathcal{P}_k(\hat{\mathbf{v}}[k])$;
- the empirical probability distributions (*ePDs*) $\mathfrak{P}_k(\hat{\mathbf{v}}[k])$;

- the maxima trajectories of the *ePDFs* of the model parameters and measurement noises (*max-pn*), given by

$$\begin{aligned}\hat{\mathbf{y}}[k] &= \mathbb{B}(\mathbf{a}^*, X_{(k-\rho)}), \quad \mathbf{a}^* = \arg\max P^*(\mathbf{a}),\\ \xi[k] &= \arg\max Q_k^*(\xi[k]),\\ \hat{\mathbf{v}}[k] &= \hat{\mathbf{y}}[k] + \xi[k], \qquad k = \overline{0,N};\end{aligned} \tag{2.29}$$

- the maxima trajectories of the *ePDFs* of the observed RPM outputs for $k = \overline{0,N}$ (*max-ePDF*), given by

$$\check{\mathbf{v}}[k] = \arg\max \mathcal{P}_k(\hat{\mathbf{v}}[k]), \qquad k = \overline{0,N}; \tag{2.30}$$

- the mean trajectories (*mean*), given by

$$\bar{\mathbf{v}}[k] = \frac{1}{M}\sum_{i=1}^{M} \hat{\mathbf{v}}^{(i)}[k], \qquad k = \overline{0,N}, \tag{2.31}$$

where M denotes the number of trajectories in the ensemble;

- the median trajectories (*med*), given by

$$\hat{\mathbf{v}}^{(i^*)}[k] \Rightarrow \sum_{i=1}^{i^*} \mathfrak{P}_k(\hat{\mathbf{v}}^{(i)}[k]) = \sum_{i=i^*+1}^{M} \mathfrak{P}_k(\hat{\mathbf{v}}^{(i)}[k]), \tag{2.32}$$

where M denotes the number of trajectories in the ensemble.

Other important characteristics of ensembles are the so-called *confidence* sets, which cover all ensemble trajectories with a given deviation or a given probability. Among them, the most informative ones are as follows:

- the variance pipe $\mathcal{D}(\bar{\mathbf{v}}[k]),\ k = \overline{0,N}$, given by

$$\begin{aligned} d(\bar{v}_j[k]) &= \sqrt{\frac{1}{M-1}\sum_{i=1}^{M}\left(\hat{v}_j^{(i)}[k] - \bar{v}_j[k]\right)^2}, \quad j = \overline{1,m},\\ \mathcal{D}(\bar{\mathbf{v}}[k]) &= \left\{\hat{\mathbf{v}}[k] : \bar{v}_j[k] - d(\hat{v}_j[k]) \leqslant \hat{v}_j[k] \leqslant \bar{v}_j[k] + d(\hat{v}_j[k]), j = \overline{1,m};\ k = \overline{0,N}\right\}; \end{aligned} \tag{2.33}$$

- the interquartile set $\mathcal{I}_{(\kappa_1,\kappa_2)}(\hat{\mathbf{v}}[k]),\ k = \overline{0,N}$, given by

$$\mathcal{I}_{(\kappa_1,\kappa_2)}(\hat{\mathbf{v}}[k]) = \mathcal{I}_{\kappa_2}(\hat{\mathbf{v}}[k]) \setminus \mathcal{I}_{\kappa_1}(\hat{\mathbf{v}}[k]), \quad \kappa_1 < \kappa_2 < 1, \tag{2.34}$$

$$\begin{aligned} \mathcal{I}_{\kappa_2}(\hat{\mathbf{v}}[k]) &= \left\{\hat{\mathbf{v}}[k] : \hat{v}_j[k] < \hat{v}_j^{\kappa_2}[k]\right\},\ \hat{v}_j^{\kappa_2}[k] \Rightarrow \mathfrak{P}_k(\hat{v}_j[k]) = \kappa_2,\\ \mathcal{I}_{\kappa_1}(\hat{\mathbf{v}}[k]) &= \left\{\hat{\mathbf{v}}[k] : \hat{v}_j[k] < \hat{v}_j^{\kappa_1}[k]\right\},\ \hat{v}_j^{\kappa_1}[k] \Rightarrow \mathfrak{P}_k(\hat{v}_j[k]) = \kappa_1,\\ k &= \overline{0,N}, \qquad j = \overline{1,m}. \end{aligned} \tag{2.35}$$

We adopt these probabilistic and numerical characteristics of the entropy-optimal ensembles to describe the randomized modeling and forecasting procedures for the evolution of thermokarst lakes area.

3. STRUCTURES OF RANDOMIZED MODELS OF THERMOKARST LAKES STATE

An important stage in the randomized forecasting technology is to construct randomized models of the thermokarst lakes area $S[n]$ and randomized models of the average annual temperature $T[n]$

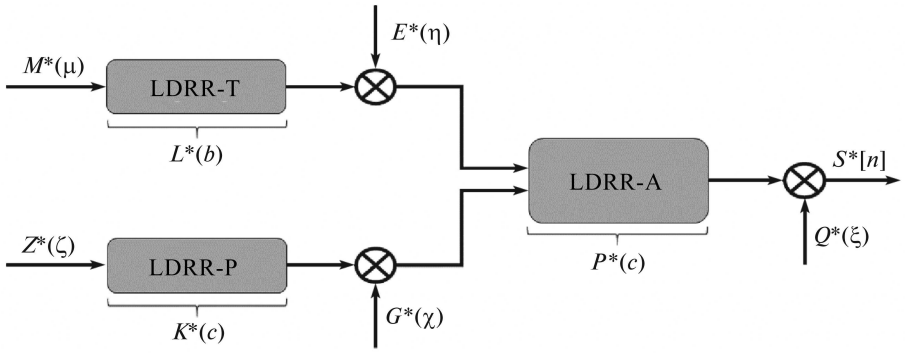

Fig. 1. The structural diagram of LDRR.

and the annual precipitation $R[n]$ (climatic parameters affecting the evolution of such lakes). These variables will be modeled using linear dynamic regressions [23] *with random parameters* (*LDRR*).

The models below are phenomenologically based on the studies described in the paper [8]. As shown therein, the lakes area depends on two climatic factors (the average annual temperature and average annual precipitation), and the past states of the object have "memory."

Figure 1 demonstrates the structural diagram of *LDRR* in accordance with the object's phenomenology. It consists of the randomized models of the lakes area (*LDRR-A*), the temperature (*LDRR-T*), and the precipitation (*LDRR-P*). Measurement errors in the data are simulated by the noises ξ, η, and ζ, respectively.

The evolution of the thermokarst lakes area $S[n]$ (*LDRR-A*) is described by the following dynamic randomized regression equation with two influencing factors, the average annual temperature $T[n]$ and the annual precipitation $R[n]$:

$$S[n] = a_0 + \sum_{k=1}^{p} a_k S[n-k] + a_{(p+1)} T[n] + a_{(p+2)} R[n]. \tag{3.1}$$

Here, p is the memory depth of the model. The parameters

$$a_k \in \mathcal{A}_k = [a_k^-, a_k^+],\ k = \overline{0,(p+2)}, \quad \mathbf{a} = \{a_0, \ldots, a_{p+2}\} \in \mathcal{A} = \bigcup_{k=0}^{p+2} \mathcal{A}_k \tag{3.2}$$

are random with a PDF $P(\mathbf{a})$.

The variable

$$v[n] = S[n] + \xi[n] \tag{3.3}$$

is the observed output of the model, and the values of the random measurement noise $\xi[n]$ at different time instants n may belong to different ranges:

$$\xi[n] \in \Xi_n = [\xi^-[n], \xi^+[n]], \tag{3.4}$$

with a PDF $Q_n(\xi[n])$, $n = \overline{0,N}$.

The randomized model *LDRR-T* of the temperature $T[n]$ with the observed output $t[n]$ has the form

$$T[n] = b_0 + \sum_{k=1}^{p} b_k T[n-k] + \mu[n], \quad t[n] = T[n] + \eta[n]. \tag{3.5}$$

The random parameters of this model are of the interval type:

$$b_k \in \mathcal{B}_k = [b_k^-, b_k^+],\ k = \overline{0,p}, \quad \mathbf{b} = \{b_0, \ldots, b_p\}, \quad \mathbf{b} \in \mathcal{B} = \bigcup_{k=0}^{p} \mathcal{B}_k. \tag{3.6}$$

The random input sequence is

$$\begin{gathered} \mu[n] \in \mathcal{M}_n = [\mu^-[n], \mu^+[n]],\ n = \overline{0,N}; \\ \mu = \{\mu[0], \ldots, \mu[N]\},\ \mu \in \mathcal{M} = \bigcup_{n=0}^{N} \mathcal{M}_n. \end{gathered} \tag{3.7}$$

The random parameters of this model and the random input sequence are characterized by a joint PDF $W(\mathbf{b}, \mu)$.

The measurement noise $\eta[n] \in \mathcal{E}_n = [\eta^-[n], \eta^+[n]]$ in (3.5) is of the interval type with PDFs $E_n(\eta[n])$.

The precipitation model (*LDRR-P*) $R[n]$ with the observed output $r[n]$ has the form

$$R[n] = c_0 + \sum_{k=1}^{p} c_k R[n-k] + \zeta[n], \qquad r[n] = R[n] + \chi[n]. \tag{3.8}$$

The random parameters of this model are independent and of the interval type:

$$c_k \in \mathcal{C}_k = [c_k^-, c_k^+],\ k = \overline{0,p}; \quad \mathbf{c} = \{c_0, \ldots, c_p\},\ \mathbf{c} \in \mathcal{C} = \bigcup_{k=0}^{p} \mathcal{C}_k. \tag{3.9}$$

The random input sequence is

$$\zeta[n] \in \mathcal{Z}_n = [\zeta^-[n], \zeta^+[n]], \quad \zeta = \{\zeta[0], \ldots, \zeta[N]\},\ \zeta \in \mathcal{Z} = \bigcup_{n=0}^{N} \mathcal{Z}_n. \tag{3.10}$$

The random parameters and the input sequence are characterized by a joint PDF $F(\mathbf{c}, \zeta)$. The measurement noise $\chi[n] \in \mathcal{G}_n = [\chi^-[n], \chi^+[n]]$ is of the interval type with PDFs $G_n(\chi[n])$.

Note that data on the lakes area, temperature, and precipitation are available at the training stage of the model *LDRR-A*. But there are no data on the temperature and precipitation at the forecasting stage. *LDRR-T* and *LDRR-P* are intended to forecast the average annual temperature and average annual precipitation using auxiliary random sequences $\mu[n]$ and $\zeta[n]$. Their joint probability characteristics are determined by training *LDRR-T* and *LDRR-P*.

4. RML ALGORITHMS FOR ESTIMATING PDFS

The model was trained using data from the array [24], structured by three geographic zones (the superscript r) and the time intervals 1973–2007. The memory parameters p of the corresponding models were determined from the historical data by correlation analysis. As a result, the following data arrays (matrices) were formed:

$$S_p^{(r)} = \begin{pmatrix} \mathbf{S}_{(0)}^{(r)} \\ \cdots \\ \mathbf{S}_{(24-p)}^{(r)} \end{pmatrix}, \quad T_p^{(r)} = \begin{pmatrix} \mathbf{T}_{(0)}^{(r)} \\ \cdots \\ \mathbf{T}_{(24-p)}^{(r)} \end{pmatrix}, \quad R_p^{(r)} = \begin{pmatrix} \mathbf{R}_{(0)}^{(r)} \\ \cdots \\ \mathbf{R}_{(24-p)}^{(r)} \end{pmatrix}, \tag{4.1}$$

where

$$\begin{gathered}\mathbf{S}_{(n-p)}^{(r)}=\{1,S^{r}[n-1],\dots,S^{(r)}[n-p]\},\quad \mathbf{S}_{(p)}^{(r)}=\{S^{(r)}[p],\dots,S^{(r)}[24]\},\\ \mathbf{T}_{(n-p)}^{(r)}=\{1,T^{(r)}[n-1],\dots,T^{(r)}[n-p]\},\quad \mathbf{T}_{(p)}^{(r)}=\{T^{(r)}[p],\dots,T^{(r)}[24]\},\\ \mathbf{R}_{(n-p)}^{(r)}=\{1,R^{(r)}[n-1],\dots,R^{(r)}[n-p]\},\quad \mathbf{R}_{(p)}^{(r)}=\{R^{(r)}[p],\dots,R^{(r)}[24]\},\\ n=\overline{p,24}.\end{gathered}\tag{4.2}$$

These arrays were used in RML algorithms to estimate the PDFs of the corresponding models.

1. For *LDRR-A*, the RML algorithm has the form

$$\begin{gathered}\mathcal{H}[P(\mathbf{a},Q(\xi)]=-\int\limits_{\mathcal{A}}P(\mathbf{a})\,\ln P(\mathbf{a})\,d\mathbf{a}\\ -\sum_{n=p}^{24}\int\limits_{\Xi_n}Q_n(\xi[n])\,\ln Q_n(\xi[n])\,d\xi[n]\Rightarrow\max\end{gathered}\tag{4.3}$$

subject to the constraints

$$\begin{gathered}\int\limits_{\mathcal{A}}P(\mathbf{a})\,d\mathbf{a}=1,\quad \int\limits_{\Xi_n}Q_n(\xi[n])\,d\xi[n]=1,\\ \int\limits_{\mathcal{A}}P(\mathbf{a})\mathbf{D}_{(n-p)}^{(r)}\,\mathbf{a}\,d\mathbf{a}+\int\limits_{\Xi_n}Q(\xi[n])\xi[n]\,d\xi[n]=S^{(r)}[n],\quad n=\overline{p,24},\end{gathered}\tag{4.4}$$

with the block row vector

$$\mathbf{D}_{(n-p)}^{(r)}=\left[\mathbf{S}_{(n-p)}^{(r)},T^{(r)}[n],R^{(r)}[n]\right].\tag{4.5}$$

The solution of this problem, parametrized by the Lagrange multipliers $\theta=\{\theta_p,\dots,\theta_{24}\}$, is given by

$$P^{*}(\mathbf{a},\theta)=\frac{\exp\left(-\langle\theta,D_p^{(r)}\,\mathbf{a}\rangle\right)}{\mathcal{P}(\theta)},\ \mathcal{P}(\theta)=\int\limits_{\mathcal{A}}\exp\left(-\langle\theta,D_p^{(r)}\,\mathbf{a}\rangle\right)d\mathbf{a},\tag{4.6}$$

with the block matrix

$$D_p^{(r)}=\left(\begin{array}{ccc}S_p^{(r)} & \mathbf{T}_{(p)}^{(r)} & \mathbf{R}_{(p)}^{(r)}\end{array}\right).\tag{4.7}$$

For *LDRR-A*, the measurement noise has the entropy-optimal PDFs

$$Q_n^{*}(\xi[n],\theta_n)=\frac{\exp(-\xi[n]\theta_n)}{\mathcal{Q}_n(\theta_n)},\ \mathcal{Q}_n(\theta_n)=\int\limits_{\Xi_n}\exp(-\xi[n]\,\theta_n)d\,\xi[n].\tag{4.8}$$

The Lagrange multipliers θ are obtained from the system of equations

$$\begin{gathered}\mathcal{P}^{-1}(\theta)\int\limits_{\mathcal{A}}\exp\left(-\langle\theta,D_p^{(r)}\,\mathbf{a}\rangle\right)\,\mathbf{D}_{(n-p)}^{(r)}\mathbf{a}\,d\,\mathbf{a}\\ +\mathcal{Q}_n^{-1}(\theta)\int\limits_{\Xi}\exp(-\xi[n]\,\theta_n)\,d\,\xi[n]=S^{(r)}[n],\\ n=\overline{p,24}.\end{gathered}\tag{4.9}$$

For *LDRR-T* and *LDRR-P*, the RML algorithm differs from (4.3)–(4.4) by the need to estimate the joint PDFs $W(\mathbf{b},\mu)$ and $F(\mathbf{c},\zeta)$ of the model parameters and input random sequences and the PDFs of the measurement noises $E(\eta)$ and $G(\chi)$.

2. Adapting the RML algorithm to *LDRR-T*, we obtain

$$\mathcal{H}[W(\mathbf{b},\mu),E(\eta)] = -\int\limits_{\mathcal{B}\bigcap\mathcal{M}} W(\mathbf{b},\mu)\,\ln W(\mathbf{b},\mu)\,d\mathbf{b}\,d\mu$$
$$-\sum_{n=p}^{24}\int\limits_{\mathcal{E}_n} E_n(\eta[n])\,\ln E_n(\eta[n])\,d\eta[n] \Rightarrow \max \tag{4.10}$$

subject to the constraints

$$\int\limits_{\mathcal{B}\bigcap\mathcal{M}} W(\mathbf{b},\mu)\,d\mathbf{b}d\mu = 1, \quad \int\limits_{\mathcal{E}_n} E_n(\eta[n])\,d\,\eta[n] = 1, \tag{4.11}$$

$$\int\limits_{\mathcal{B}\bigcap\mathcal{M}} W(\mathbf{b},\mu)\left[\mathbf{T}^{(r)}_{(n-p)}\mathbf{b}+\mu[n]\right]\,d\,\mathbf{b}\,d\,\mu[n] + \int\limits_{\mathcal{E}_n} E_n(\eta[n])\eta[n]\,d\,\eta[n] = T^{(r)}[n],$$
$$n=\overline{p,24}.$$

In the expressions (4.10)–(4.11), $\mu=\{\mu[p],\ldots,\mu[24]\}$.

Let $\vartheta=\{\vartheta_p,\ldots,\vartheta_{24}\}$ denote the Lagrange multipliers for this problem. Recall that the model is linear, and its random parameters are independent of the auxiliary random sequence elements. Therefore, the solution of this problem—the optimal PDFs—can be written as

$$W^*(\mathbf{b},\mu,\vartheta) = L^*(\mathbf{b},\vartheta)\,M^*(\mu,\vartheta), \quad M^*(\mu,\vartheta)=\prod_{n=p}^{24} M_n^*(\mu[n],\vartheta_n),$$

$$L^*(\mathbf{b},\vartheta)=\frac{\exp\left(-\langle\vartheta,T_p^{(r)}\,\mathbf{b}\rangle\right)}{\mathcal{L}(\vartheta)},\ \mathcal{L}(\vartheta)=\int\limits_{\mathcal{B}}\exp\left(-\langle\vartheta,T_p^{(r)}\,\mathbf{b}\rangle\right)d\,\mathbf{b}, \tag{4.12}$$

$$M_n^*(\mu[n],\,\vartheta_n)=\mathbb{M}_n^{-1}(\vartheta_n)\exp\left(-\vartheta_n\mu[n]\right), \quad \mathbb{M}_n(\vartheta_n)=\int\limits_{\mathcal{M}_n}\exp\left(-\vartheta_n\mu[n]\right)d\,\mu[n],$$

$$E_n^*(\eta[n],\vartheta_n)=\mathbb{E}_n^{-1}(\vartheta_n)\exp\left(-\eta[n]\,\vartheta_n\right),\ \mathbb{E}_n(\vartheta_n)=\int\limits_{\mathcal{E}_n}\exp\left(-\eta[n]\,\vartheta_n\right)d\,\eta[n].$$

The Lagrange multipliers ϑ are obtained from the system of equations

$$\mathcal{L}^{-1}(\vartheta)\int\limits_{\mathcal{B}}\exp\left(-\langle\vartheta,T_p^{(r)}\,\mathbf{b}\rangle\right)\mathbf{T}^{(r)}_{(n-p)}\,\mathbf{b}\,d\mathbf{b}$$
$$+\,\mathbb{M}_n^{-1}(\vartheta_n)\int\limits_{\mathcal{M}_n} M_n^*(\mu[n],\,\vartheta_n)\,\mu[n]d\mu[n]$$
$$+\,\mathbb{E}_n^{-1}(\vartheta_n)\int\limits_{\mathcal{E}_n} E_n^*(\eta[n],\vartheta_n)\,\eta[n]d\eta[n] = T^{(r)}[n], \quad n=\overline{p,24}. \tag{4.13}$$

3. For *LDRR-P*, similar to (4.12)–(4.13), the RML algorithm has the form

$$F^*(\mathbf{c},\zeta,\lambda)=V^*(\mathbf{c},\lambda)\,Z^*(\zeta,\lambda),\quad Z^*(\zeta,\lambda)=\prod_{n=p}^{24}Z_n^*(\zeta[n],\lambda_n),$$

$$V^*(\mathbf{c},\lambda)=\frac{\exp\left(-\langle\lambda,R_p^{(r)}\,\mathbf{c}\rangle\right)}{\mathcal{V}(\lambda)},\ \mathcal{V}(\lambda)=\int\limits_{\mathcal{C}}\exp\left(-\langle\lambda,R_p^{(r)}\,\mathbf{c}\rangle\right)\,d\,\mathbf{c},\tag{4.14}$$

$$Z_n^*(\zeta[n],\lambda_n)=\mathbb{Z}_n^{-1}(\lambda_n)\exp\left(-\lambda_n\,\zeta[n]\right),\quad \mathbb{Z}_n(\lambda_n)=\int\limits_{\mathcal{Z}_n}\exp\left(-\lambda_n\,\zeta[n]\right)d\,\zeta[n],$$

$$G_n^*(\chi[n],\lambda_n)=\mathbb{G}_n^{-1}(\lambda)\exp\left(-\chi[n]\lambda_n\right),\ \mathbb{G}_n(\lambda_n)=\int\limits_{\mathcal{G}_n}\exp\left(-\chi[n]\,\lambda_n\right)d\,\chi[n].$$

The Lagrange multipliers $\lambda=\{\lambda_p,\ldots,\lambda_{24}\}$ are obtained from the system of equations

$$\begin{aligned}\mathcal{V}^{-1}(\lambda)\int\limits_{\mathcal{C}}\exp\left(-\langle\lambda,R_p^{(r)}\,\mathbf{c}\rangle\right)\mathbf{R}_{(n-p)}^{(r)}\,\mathbf{c}\,d\mathbf{c}\\ +\mathbb{Z}_n^{-1}(\lambda_n)\int\limits_{\mathcal{Z}_n}Z_n^*(\zeta[n],\lambda_n)\,\zeta[n]d\zeta[n]\\ +\mathbb{G}_n^{-1}(\lambda_n)\int\limits_{\mathcal{G}_n}G_n^*(\chi[n],\lambda_n)\,\chi[n]\,d\chi[n]=R^{(r)}[n],\quad n=\overline{p,24}.\end{aligned}\tag{4.15}$$

According to (4.6), (4.8), (4.12), and (4.14), the entropy-optimal PDFs belong to the exponential class. They are parametrized by the corresponding Lagrange multipliers, which satisfy the balance equations (4.9), (4.13), and (4.15).

5. THE EVOLUTION OF THERMOKARST LAKES AREA IN WESTERN SIBERIA: THE RESULTS OF TRAINING, TESTING, AND FORECASTING

1. Randomized training (1973–1997). At the preliminary training stage, we determined the orders p of the corresponding models using historical data from a training collection. For this purpose, the autocorrelation functions $r[k]$ were calculated; $p=k_{\max}$ for which $r[k_{\max}]\leqslant\delta$, where $\delta=0.1$.

As a result, the entropy-optimal PDFs of the model parameters, auxiliary random sequences, and measurement noises were found.

LDRR-A ($p=4$). The analytical expressions for the corresponding parametrized PDFs are given by (4.6), (4.8). Since *LDRR-A* is linear, these PDFs belong to the exponential class:

$$\begin{aligned}P^*(\mathbf{a},\theta)=\prod_{k=0}^{(p+2)}P_k^*\left(a_k\right),\quad P_k^*(a_k)=\frac{\exp(-q_k\,a_k)}{\mathcal{P}_k},\quad \mathcal{P}_k=\int\limits_{\mathcal{A}_k}\exp(-q_k\,a_k)\,da_k,\\ q_0=\sum_{n=p}^{24}\theta_n,\quad q_k=\sum_{n=p}^{24}\theta_n\,S^{(r)}[n-k],\quad k=\overline{1,p},\\ q_{p+1}=\sum_{n=p}^{24}\theta_n\,T^{(r)}[n],\quad q_{p+2}=\sum_{n=p}^{24}\theta_n\,R^{(r)}[n],\\ Q^*(\xi,\bar\theta)=\frac{\exp(-\bar\theta\,\xi)}{\mathcal{Q}},\quad \mathcal{Q}=\int\limits_{\Xi}\exp(-\bar\theta\,\xi)\,d\,\xi,\quad \bar\theta=\frac{q_0}{24-p}.\end{aligned}\tag{5.1}$$

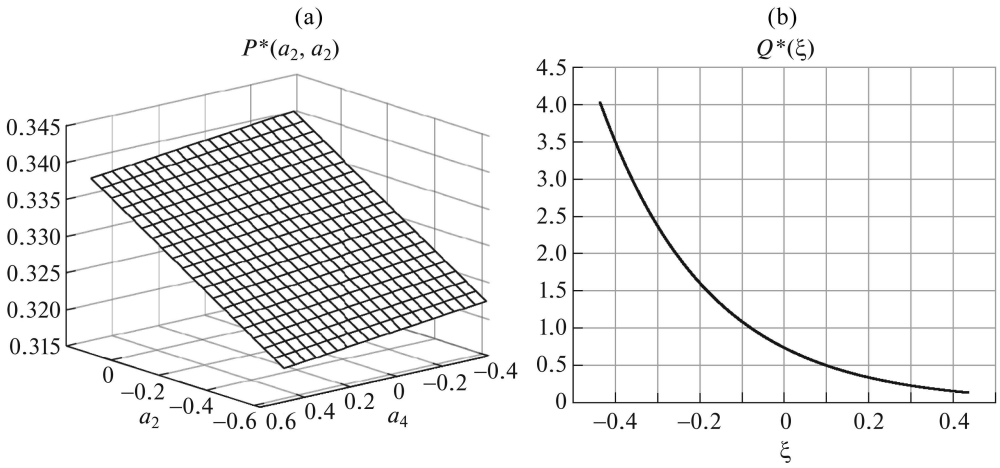

Fig. 2. Examples of reconstructed PDFs.

Figure 2 shows examples of the entropy-optimal PDFs: (a) $P^*(a_2, a_4)$ (model parameters) and (b) $Q^*(\xi)$ (noise); they were reconstructed by training based on the historical data for the northern zone of thermokarst lakes. Table 1 presents the average values of the parameter estimates of *LDRR-A* obtained by the RML procedure and the ordinary least squares (OLS) method. Obviously, the OLS estimates almost coincide with the average RML ones, which is due to the linear RPMs used. However, even in this case, the RML procedure allows generating an ensemble of model output trajectories and determining its numerical characteristics.

Table 1. *LDRR-A*: estimated parameters

The northern zone of thermokarst lakes				
Model	Order	Parameters	OLS estimates	RML estimates
S	4	a_0	−0.2888	−0.2750
		a_1	0.1069	0.1126
		a_2	−0.2224	−0.2212
		a_3	−0.1289	−0.1333
		a_4	0.0535	0.0533
		a_5	0.8330	0.8322
		a_6	0.6245	0.6080

LDRR-T. The analytical expressions for the corresponding PDFs are given by

$$W^*(\mathbf{b}, \mu, \vartheta) = L^*(\mathbf{b}, \vartheta)\, M^*(\mu, \vartheta), \quad M^*(\mu, \vartheta) = \prod_{n=p}^{(24)} M_n^*(\mu[n], \vartheta_n),$$

$$L^*(\mathbf{b}, \vartheta) = \prod_{k=0}^{p} L_k^*(b_k), \quad L_k^*(b_k) = \frac{\exp(-w_k\, b_k)}{\mathcal{L}_k}, \quad \mathcal{L}_k = \int\limits_{\mathcal{B}_k} \exp(-w_k\, b_k)\, db_k,$$

$$w_0 = \sum_{n=p}^{24} \vartheta_n, \quad w_k = \sum_{n=p}^{24} \vartheta_n\, T^{(r)}[n-k], \quad k = \overline{1,p}, \tag{5.2}$$

$$M_n^*(\mu[n], \vartheta_n) = \frac{\exp(-\vartheta_n\, \mu[n])}{\mathfrak{M}_n}, \quad \mathfrak{M}_n = \int\limits_{\mathcal{M}_n} \exp(-\vartheta_n\, \mu[n])\, d\mu[n],$$

$$E^*(\eta, \bar{\vartheta}) = \frac{\exp(-\bar{\vartheta}\, \eta)}{\mathfrak{E}}, \quad \mathfrak{E} = \int\limits_{\mathcal{E}} \exp(-\bar{\vartheta}\, \eta)\, d\eta, \quad \bar{\vartheta} = \frac{w_0}{24-p}.$$

Table 2. Relative testing errors

Estimate	*LDRR-A*	*LDRR-T*	*LDRR-P*
RelErr	0.0089	0.0801	0.1675

LDRR-P. The analytical expressions for the corresponding PDFs are given by

$$
\begin{gathered}
F^*(\mathbf{c},\zeta,\chi)=V^*(\mathbf{c},\lambda)\,Z^*(\zeta,\lambda),\quad Z^*(\zeta,\lambda)=\prod_{n=p}^{(24)}Z_n^*(\zeta[n],\lambda_n),\\
V^*(\mathbf{c},\lambda)=\prod_{k=0}^{p}V_k^*(c_k),\quad V_k^*(b_k)=\frac{\exp(-s_k\,c_k)}{\mathcal{V}_k},\\
\mathcal{V}_k=\int\limits_{\mathcal{C}_k}\exp(-s_k\,c_k)\,dc_k,\\
s_0=\sum_{n=p}^{24}\lambda_n,\quad s_k=\sum_{n=p}^{24}\lambda_n\,R^{(r)}[n-k],\quad k=\overline{1,p},\\
Z_n^*(\zeta[n],\lambda_n)=\frac{\exp(-\lambda_n\,\zeta[n])}{\mathfrak{Z}_n},\quad \mathfrak{Z}_n=\int\limits_{\mathcal{Z}_n}\exp(-\lambda_n\,\zeta[n])\,d\zeta[n],\\
G^*(\chi,\bar\lambda)=\frac{\exp(-\bar\lambda\,\chi)}{\mathfrak{G}},\quad \mathfrak{G}=\int\limits_{\mathcal{G}}\exp(-\bar\lambda\,\chi)\,d\,\chi,\quad \bar\lambda=\frac{w_0}{24-p}.
\end{gathered}
\tag{5.3}
$$

2. Testing (1998–2007). The testing procedure was applied to the combination of the trained models (Section 7) with the test data collections. This procedure is based on the sampling of the optimal PDFs [22] and the further generation of the ensembles of random trajectories on the testing interval. The quality of the models was assessed by the closeness of the mean model trajectories

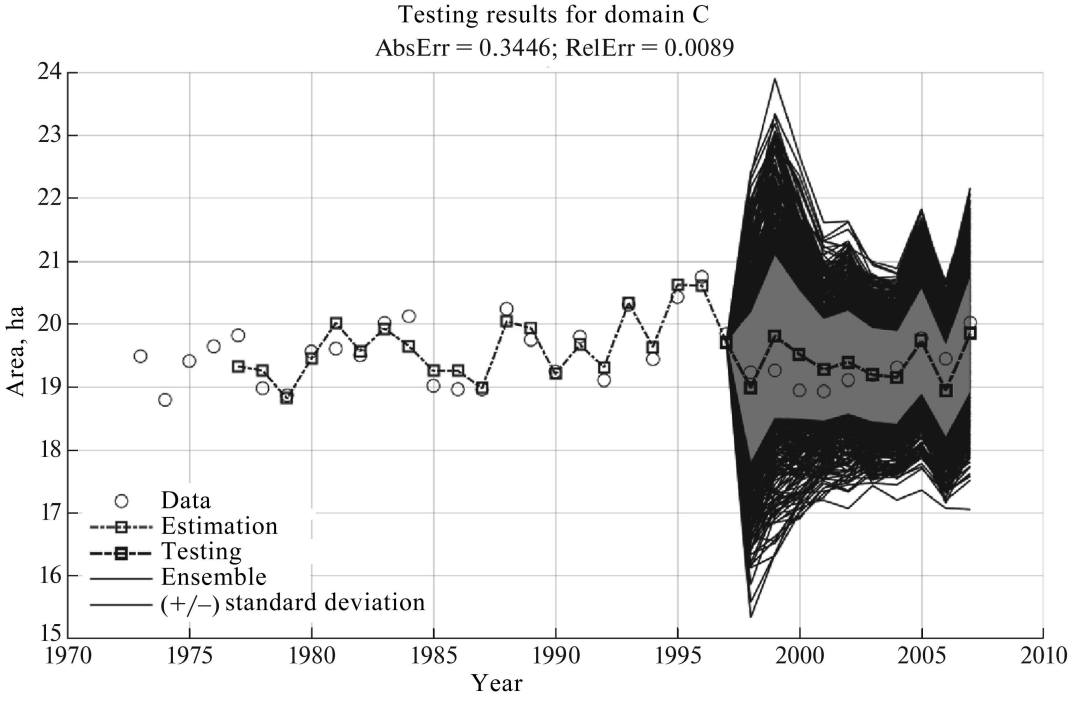

Fig. 3. Dynamics of the thermokarst lakes area.

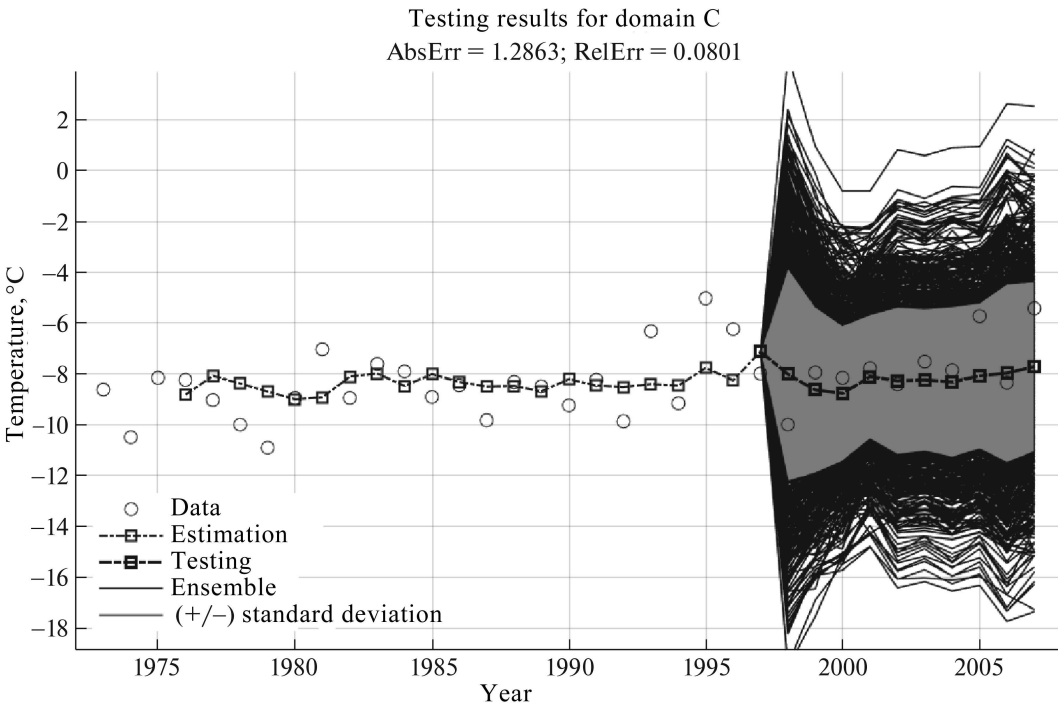

Fig. 4. Dynamics of the average annual temperature.

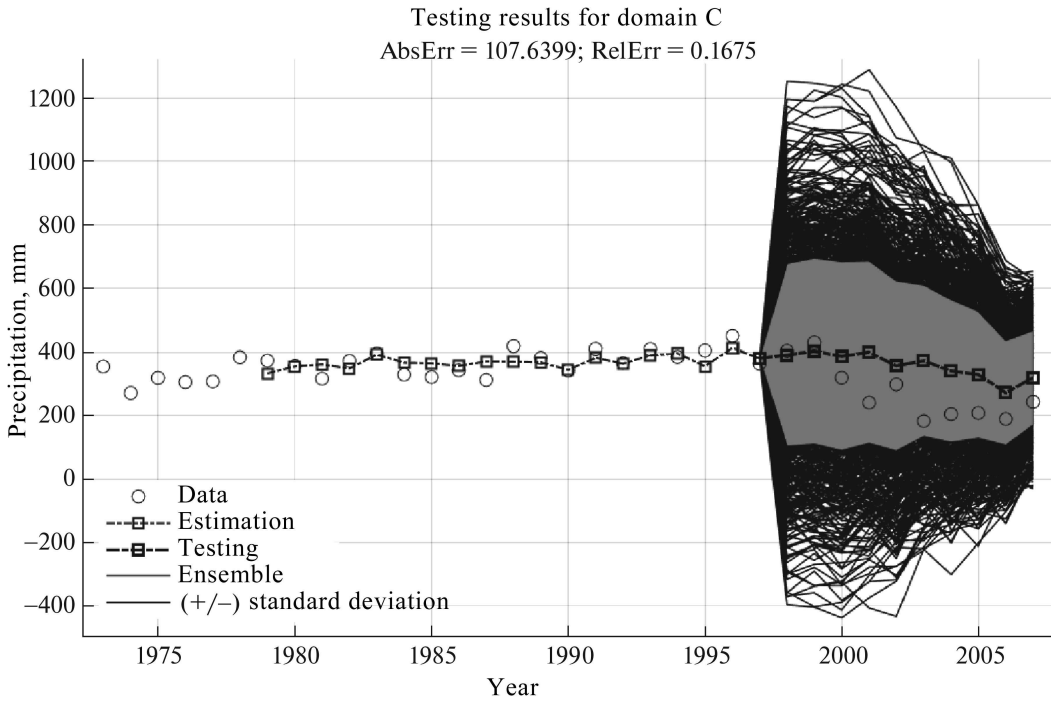

Fig. 5. Dynamics of the annual precipitation.

to the real data on the observation interval in terms of the relative error (2.15). Table 2 shows the relative testing errors for the northern zone. In addition, Figs. 3–5 demonstrate examples of the tested trajectories, including the corresponding errors.

3. Randomized forecasting (2008–2023). All forecasts were constructed using the composite model (Fig. 1) consisting of *LDRR-A*, *LDRR-T*, and *LDRR-P*. The entropy-optimal PDFs gen-

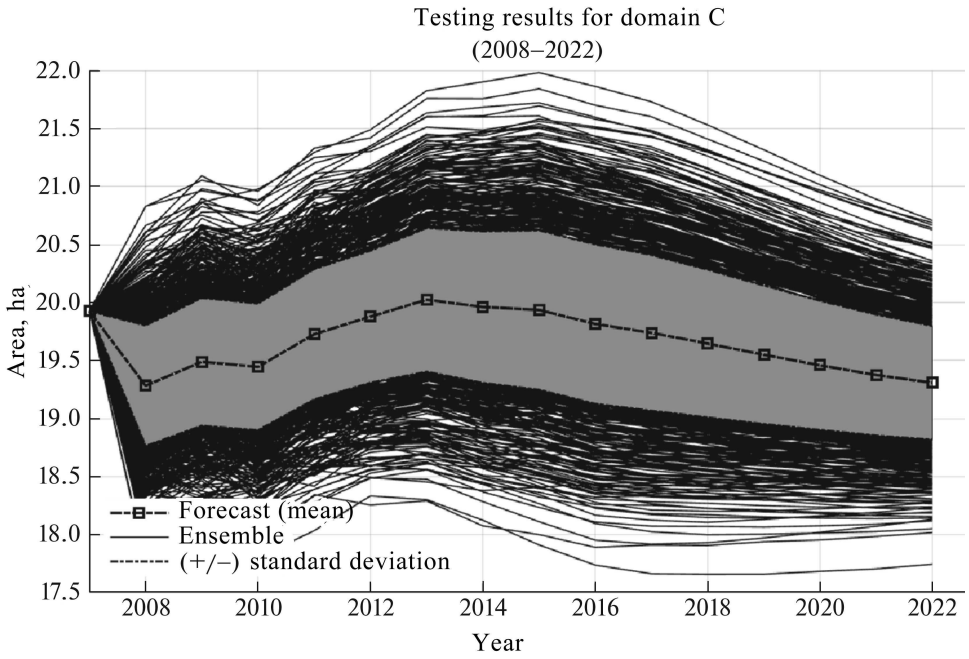

Fig. 6. Forecast trajectories of the thermokarst lakes area: ensemble, variance pipes, and mean trajectories.

erate the ensembles of trajectories describing the evolution of the thermokarst lakes area on the forecasting interval. The difference between the forecasting and testing intervals is that in the former case, all data points are used to train the models.

For each point, the mean value (*mean*) and the standard deviation (*std*) were calculated over the entire ensemble. The resulting forecasts for the northern zone are shown in Fig. 6. They include the ensemble of forecast trajectories for the thermokarst lakes area, the mean trajectory (the dark curve), and the variance pipe (the grey zone).

6. DISCUSSION OF THE RESULTS

The formation and evolution processes of thermokarst lakes in permafrost zones have been insufficiently studied so far, both in terms of their internal geological dynamics and external climatic factors affecting them. Procedures for transforming satellite images into the number and area of thermokarst lakes are accompanied by very significant errors. Therefore, the model information and data on the object's state are uncertain.

RML algorithms serve for estimating the characteristics of such models based on data with unknown-nature errors. These algorithms reconstruct the PDFs of the model parameters and measurement noises corresponding to the maximum uncertainty.

The sampling of the optimal PDFs and Monte Carlo simulations allow generating ensembles of trajectories that describe thermokarst lakes area evolution. Statistical processing of these ensembles yields various numerical characteristics, first of all, the mean trajectories and the set of trajectories bounded by the root-mean-square trajectories (variance pipes). According to the results of this paper, the mean trajectories differ from the real permafrost zone data by 8–17% in terms of the relative root-mean-square (integral) errors.

The conducted study of the object with the application, adaptation, and testing of randomized machine learning algorithms has demonstrated a sufficiently high accuracy in reconstructing the mean trajectories (8–17%). The maximum dimensions of the variance pipes calculated by the generated ensembles were ±9%.

All studies have involved linear dynamic randomized regression models with the memory $p=3$–5. For the thermokarst lakes area, they have provided quite acceptable simulation results. For the climatic parameters, however, the testing errors have turned out to be significantly higher. Apparently, the transition to nonlinear models will increase the accuracy of reproducing real data.

The balance equations with the so-called integral components still remain a general difficulty in the RML procedure. These components are multidimensional definite integrals with parameters calculated on simple sets (parallelepipeds). In this study (linear models), they have been calculated analytically with the software integration of the corresponding analytical expressions. However, such an approach will be eliminated when passing to nonlinear models.

7. CONCLUSIONS

This paper is one of those devoted to the effectiveness and performance of randomized machine learning. It has considered the problems of applying optimal randomized models to forecasting. For such models, the forecasted input sequence is often unknown. It has been proposed to generate this sequence using an auxiliary random sequence with entropy-optimal characteristics reconstructed by the machine learning of the corresponding model.

FUNDING

This work was supported by the Russian Science Foundation, project no. 22-11-20023.

REFERENCES

1. Vapnik, V.N., *Statistical Learning Theory*, John Willey & Sons, 1998.
2. Bishop, C., *Pattern Recognition and Machine Learning*, New York: Springer, 2007.
3. Friedman, J., Hastie T., and Tibshirani, R., *The Elements of Statistical Learning*, Springer Series in Statistics, vol. 1, Berlin: Springer, 2009.
4. Popkov, Yu.S., Dubnov, Yu.A., and Popkov, A.Yu., Randomized Machine Learning: Statement, Solution, Applications, *Proc. IEEE Int. Conf. on Intelligent Systems*, 2016, pp. 27–39.
5. Zuidhoff, F.S. and Kolstrup, E., Changes in Palsa Distribution in Relation to Climate Change in Laivadalen, Northern Sweden, Especially 1960–1997, *Permafrost and Periglacial Processes*, 2000, vol. 11, pp. 55–69.
6. Kirpotin, S., Polishchuk, Y., and Bruksina, N., Abrupt Changes of Thermokarst Lakes in Western Siberia: Impacts of Climatic Warming on Permafrost Melting, *Int. J. Environmental Studies*, 2009, vol. 66, no. 4, pp. 423–431.
7. Karlson, J.M., Lyon, S.W., and Destouni, G., Temporal Behavior of Lake Size-Distribution in a Thawing Permafrost Landscape in Northwestern Siberia, *Remote Sensing*, 2014, no. 6, pp. 621–636.
8. Bryksina, N.A. and Polishchuk, Yu.M., Analysis of Changes in the Number of Thermokarst Lakes in Permafrost of Western Siberia on the Basis of Satellite Images, *Cryosphere of Earth*, 2015, vol. 19, no. 2, pp. 114–120.
9. Liu, Q., Rowe, M.D., Anderson, E.J., Stow, C.A., and Stumpf, R.P., Probabilistic Forecast of Microcystin Toxin Using Satellite Remote Sensing, *in situ* Observation and Numerical Modeling, *Environment Modelling and Software*, 2020, vol. 128, p. 104705.
10. Vidyasagar, M., Statistical Learning Theory and Randomized Algorithms for Control, *IEEE Control System Magazine*, 1998, vol. 1, no. 17, pp. 69–88.
11. Granichin, O.N. and Polyak, B.T., *Randomizirovannye algoritmy otsenivaniya i optimizatsii pri pochti proizvol'nykh pomekhakh* (Randomized Estimation and Optimization Algorithms under Almost Arbitrary Noises), Moscow: Nauka, 2002.
12. Biondo, A.E., Pluchino, A., Rapisarda, A., and Helbing, D., Are Random Traiding Strategies More Successful than Technical Ones?, *PLoS ONE*, 2013, vol. 6, no. 7, p. e68344.

13. Lutz, W., Sandersen, S., and Scherbov, S., The End of World Population Growth, *Nature*, 2001, vol. 412, no. 6846, pp. 543–545.
14. Tsirlin, A.M., *Metody usrednennoi optimizatsii i ikh primenenie* (Average Optimization Methods and Their Application), Moscow: Fizmatlit, 1997.
15. Shannon, C., Communication Theory of Secrecy Systems, *Bell System Technical Journal*, 1949, vol. 28, no. 4, pp. 656–715.
16. Jaynes, E.T., Information Theory and Statistical Mechanics, *Physics Review*, 1957, vol. 106, pp. 620–630.
17. Jaynes, E.T., *Papers on Probability, Statistics and Statistical Physics*, Dordrecht: Kluwer Academic Publisher, 1989.
18. Jaynes, E.T., *Probability Theory. The Logic and Science*, Cambridge University Press, 2003.
19. Popkov, Yu.S., Popkov, A.Yu., and Dubnov, Yu.A., *Entropy Randomization in Machine Learning*, Chapman & Hall/CRC, 2022.
20. Popkov, Y. and Popkov, A., New Method of Entropy-Robust Estimation for Randomized Models under Limited Data, *Entropy*, 2014, vol. 16, pp. 675–698.
21. Ioffe, A.D. and Tihomirov, V.M., *Theory of Extremal Problems*, Amsterdam–New York–Oxford: North-Holland, 1979.
22. Darkhovsky, B.S., Popkov, Y.S., Popkov, A.Y., and Aliev, A.S., A Method of Generating Random Vectors with a Given Probability Density Function, *Autom. Remote Control*, 2018, vol. 79, no. 9, pp. 1569–1581. https://doi.org/10.1134/S0005117918090035
23. Aivazyan, S.A., Enyukov, I.S., and Meshalkin, L.D., *Prikladnaya statistika: Issledovanie zavisimostei* (Applied Statistics: Study of Dependencies), Moscow: Finansy i Statistika, 1985.
24. https://cloud.uriit.ru/index.php/s/0DOrxL9RmGqXsV0.

This paper was recommended for publication by A.N. Sobolevski, a member of the Editorial Board