Article

New Method of Randomized Forecasting Using Entropy-Robust Estimation: Application to the World Population Prediction

Yuri S. Popkov [1,2,3,*], Yuri A. Dubnov [1,2] and Alexey Yu. Popkov [1,2]

[1] Institute for Systems Analysis of Russian Academy of Sciences, Moscow 117312, Russia; yury.dubnov@phystech.edu (Y.A.D.); apopkov@isa.ru (A.Y.P.)

[2] Moscow Institute of Physics and Technology (State University), Moscow 141700, Russia

[3] National Research University Higher School of Economics, Moscow 101000, Russia

[*] Correspondence: popkov@isa.ru; Tel.: +7-499-135-42-22

Abstract: We propose a new method of randomized forecasting (RF-method), which operates with models described by systems of linear ordinary differential equations with random parameters. The RF-method is based on entropy-robust estimation of the probability density functions (PDFs) of model parameters and measurement noises. The entropy-optimal estimator uses a limited amount of data. The method of randomized forecasting is applied to World population prediction. Ensembles of entropy-optimal prognostic trajectories of World population and their probability characteristics are generated. We show potential preferences of the proposed method in comparison with existing methods.

Keywords: entropy; randomized model; randomized forecasting; the exponential World population model

1. Introduction

For a studied process, forecasting as a procedure consists of four consecutive stages: modeling (model design), learning (estimation of model characteristics), testing (of the "learned" model) and prediction of future development.

Forecasting is based on retrospective data analysis with its subsequent extrapolation to future periods. Consider the state of a studied process at moment t_0, and suppose that the problem is to forecast further evolution of the process on a time interval $\mathcal{T}_{frc} = [t_0, T]$. Then, it is necessary to operate existing data on its past dynamics on a time interval $\mathcal{T}_{rts} = [T^-, t_0]$, where $T^- < t_0$ (the so-called retrospective data). Generally, retrospective data and the time interval $\mathcal{T}_{rts}$ are divided into two groups, namely data serving for the estimation of the model's characteristics (on a time interval $\mathcal{T}_{est} = [T^-, t_e]$) and model testing (on a time interval $\mathcal{T}_{tst} = [t_e, t_0]$).

There exist at least three forecasting techniques differing in the objectification degree of constructed forecasts. The *first technique*, referred to as *scenario forecasting* [1], proceeds from the scenario approach whose objectification is replaced by the opinion of an expert group. Actually, it implements only the stages of modeling and prediction: learning and testing are eliminated owing to the opinion of experts, who choose an appropriate mathematical model of a studied process and form value sets (scenarios) of the model parameters. Then, the model with the scenario parameter values generates forecasting trajectories. As a matter of fact, such a forecasting technique is most widespread in demographic prediction [2–4]. Note that real retrospective data about a studied process are not utilized. They are indirectly reflected by the knowledge and experience of invited experts.

The *second technique* of forecasting explicitly involves real data for model learning and testing. The framework of mathematical statistics provides numerous estimation methods for model parameters; see [5–8]. Here, a major assumption is that the model possesses deterministic parameters, the values of which are defined using sets of real retrospective data. The latter are treated as a stochastic object with certain properties (a sample from a universe, normal distribution, *etc*.). In this case, one may assign different probabilistic characteristics (variances, confidence intervals, and so on) to the derived estimates of model parameters.

Sometimes, these characteristics assist in constructing the probabilistic characteristics of forecasting trajectories. The described technique will be termed *probabilistic forecasting (PF)* [9–14]. We emphasize that the above hypotheses regarding the stochastic properties of real datasets are almost impossible to verify, especially under small data arrays. It follows for the low efficiency of forecasts [15,16].

Finally, the *third technique* of forecasting proposed in this paper stems from the *randomized model (RM)* of a studied process, where model parameters are supposed random. Hence, we characterize RMs using the probability density functions (PDFs) of their parameters. At the learning stage, the PDFs of the model parameters are estimated on the basis of real retrospective data.

A randomized model generates an ensemble of forecasting trajectories, where each trajectory corresponds to a set of random realizations of parameters *with the derived estimates of the PDFs*. Computer simulation of such models employs the Monte Carlo method. Below, this technique will be called *randomized forecasting (RF)*.

As opposed to existing methods [17–19], the proposed method of randomized forecasting is based on entropy-optimal estimations of PDFs for real datasets. The structure of the randomized dynamic model, used in this method, is based on ordinary differential equations.

The developed method serves for obtaining randomized predictions of the World population dynamics. Modeling the World population variations in time and space forms a major problem of demographic analysis [20–22].

Throughout the paper, we describe the above mentioned dynamics by the exponential model incorporating several parameters associated with fertility and mortality rates, as well as its change in time. In the randomized setting, they are assumed random, whereas World population is measured with random errors. To find the corresponding PDFs, the method involves the retrospective population data provided by the UN (see UNdata service at `https://data.un.org/`). In addition, we perform the comparative analysis of the PF- and RF-based approaches.

2. Randomized Model: Linear Differential Form

Consider a dynamic object having an input $\mathbf{f}(t) = \{f_1(t), \ldots, f_m(t)\}$ and an output $\mathbf{x}(t) = \{x_1(t), \ldots, x_n(t)\}$. The components of the input and output vectors can be observed (measured) on a time interval $\mathcal{T}_{rts} = [T^-, t_0]$.

The relationship between the input and output of the object is described by the linear nonautonomous system of ordinary differential equations:

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t) + B\mathbf{f}(t), \qquad \mathbf{x}(T^-) = \mathbf{x}^0 \tag{1}$$

where $\mathbf{x} \in R^n$; $\mathbf{f} \in R^m$, $m \leq n$; $A = [a_{ij} \,|\, (i,j) = \overline{1,n}]$ and $B = [b_{ik} \,|\, i = \overline{1,n}, k = \overline{1,m}]$ denote matrices of appropriate dimensions.

The object's output is observed with inevitable disturbances modeled by a vector noise $\bar{\xi}(t) = \{\xi_1(t), \ldots, \xi_n(t)\}$. Therefore, the observed output of the model acquires the form:

$$\mathbf{v}(t) = \mathbf{x}(t) + \bar{\xi}(t) \tag{2}$$

where $\mathbf{v}(t) \in R^n$.

Equations (1) and (2) define a *linear dynamical randomized model (**LDRM**)* if:

- the matrix A is a random matrix (with independent random elements or elements with independent random components) of the interval type:

$$\mathcal{A} = [A : A^- \leq A \leq A^+] \tag{3}$$

  where A^-, A^+ mean given matrices;
- there exists a probability density function (PDF) $P(A)$, $A \in \mathcal{A}$;
- the vector $\bar{\xi}$ is a random vector (*i.e.*, contains independent random components) of the interval type:

$$\Xi = [\bar{\xi}^- \leq \bar{\xi} \leq \bar{\xi}^+] \tag{4}$$

- there exists a probability density function (PDF) $Q(\bar{\xi})$, $\bar{\xi} \in \Xi$;
- the matrix B possesses fixed known elements.

Under the stated conditions, the LDRM generates an ensemble of random trajectories on the time interval $\mathcal{T} = [T^-, T]$.

Let us rewrite the LDRM (2) in the input-output representation using the notion of matrix exponent [23]:

$$W(A|t - \tau) = \exp[A(t - \tau)] \tag{5}$$

The input and output are measured at discrete moments with step h. Hence, on the time interval $\mathcal{T}_{est}$, we have:

$$
\begin{aligned}
\mathbf{x}[T^- + ih] &= W(A|ih)\mathbf{x}^0 + \\
&+ \int_{T^-}^{T^- + ih} W(A|T^- + ih, \tau)\, B\, \mathbf{f}(\tau)d\tau, \\
&i \in \overline{0, N_{est}}
\end{aligned}
\tag{6}
$$

here $N_{est} = [(t_e - T^-)/h]$ and $[\bullet]$ indicates the integer part of $\bullet$.

The LDRM output Equation (3) observed at discrete moments has the following form:

$$\mathbf{v}[T^- + ih] = \mathbf{x}[T^- + ih] + \bar{\xi}[T^- + ih], \qquad i = \overline{0, N_{est}} \tag{7}$$

Let us denote $\bar{\xi}^{(i)} = \bar{\xi}[T^- + ih]$, $i = \overline{0, N_{est}}$. They are random vectors with independent and interval components. Now, we introduce the block-vector $\hat{\xi} = \{\bar{\xi}^{(0)}, \ldots, \bar{\xi}^{(N_{est})}\}$. As we assume that these vectors and their components are independent, then the joint PDF is:

$$Q(\hat{\xi}) = \prod_{i=0}^{N_{est}} Q_i(\bar{\xi}^{(i)}) = \prod_{i=0}^{N_{est}} \prod_{j=1}^{n} q_{ij}(\xi_j^{(i)}) \tag{8}$$

The domain of this function is:

$$\hat{\Xi} = \underbrace{\Xi \times \Xi \cdots \times \Xi}_{(N_{est} + 1) \text{ multipliers}} \tag{9}$$

Here, $\bar{\xi}(T^-), \bar{\xi}(T^- + h), \ldots, \bar{\xi}(T^- + N_{est}h)$ gives a sequence of n-dimensional independent random vectors of the interval type, associated with corresponding PDFs.

As soon as the matrix A and the noise vector $\hat{\bar{\xi}}$ are random and characterized by the PDFs $P(A)$ and $Q(\bar{\xi})$, respectively, so the observed output of the LDRM represents an ensemble $\mathcal{V}$ of random trajectories $\mathbf{v}[T^- + ih]$, $i = \overline{0, N_{est}}$.

3. $\mathcal{S}_{PQ}^1$ Entropy-Robust Estimation

The first stage of the randomized forecasting (RF) is an estimation of the PDFs of the RMs' parameters and of the measurement noises. It is a classical problem of the Bayesian approach, and there exists the classical maximum likelihood method (or maximum relation of likelihood; see [6]) for its solving.

Let us recall the definition of the function of the relation of likelihood (FRL) in the terms of Section 2. The *a priori* PDFs $P^0(A), Q^0(\hat{\xi})$ and the *a posteriori* PDFs $P(A), Q(\hat{\xi})$ are the basic notations of the FRL. The FRL takes the form:

$$L(A) = \ln \frac{P(A)}{P^0(A)}, \qquad L(\hat{\xi}) = \ln \frac{Q(\hat{\xi})}{Q^0(\hat{\xi})} \tag{10}$$

If the PDFs in these expressions are restored as functions of the parameters, then maximization of these functions gives "optimal" estimations. The principle of maximization of the FRL takes the form:

$$\hat{A} = \arg \max_A L(A) \tag{11}$$

As a declaration, this principle is fine. However, how is it possible to restore the PDFs $P(A)$ and $P^0(A)$ (also $Q(\hat{\xi})$ and $Q^0(\hat{\xi})$)? The problem of the restoration of the PDFs remains outside of the FRL. Let us consider the *functional* of the likelihood relation (FuRL) in the following form:

$$\mathcal{L}[P(A)] = \int_{\mathcal{A}} P(A) \ln \frac{P(A)}{P^0(A)} \, dA, \qquad \mathcal{L}[Q(\hat{\xi})] = \int_{\hat{\Xi}} Q(\hat{\xi}) \ln \frac{Q(\hat{\xi})}{Q^0(\hat{\xi})} \, d\hat{\xi} \tag{12}$$

From Equation (12), we can see that the FuRL is the mathematical expectation of the FRL. On the other side, the FuRL is the opposite generalized information Boltzmann entropy (Kullback–Leibler distance) [24,25], that is:

$$\mathcal{H}_A[P(A)] = -\mathcal{L}[P(A)], \qquad \mathcal{H}[Q(\hat{\xi})] = -\mathcal{L}[Q(\hat{\xi})] \tag{13}$$

According to [26], maximization of entropy functions gives the best robust solution under high uncertainty. This idea with the addition of real data balance conditions forms the basis of the $\mathcal{S}_{PQ}^1$ entropy-robust estimation method [27].

The $\mathcal{S}_{PQ}^1$ entropy-robust estimation can be reformulated as a problem of functional nonlinear programming [28,29]:

$$\begin{aligned} -\mathcal{H}[P(A), Q(\bar{\xi})] &= \int_{\mathcal{A}} P(A) \ln P(A) \, dA + \\ &+ \int_{\bar{\Xi}} Q(\hat{\xi}) \ln Q(\hat{\xi}) \, d\hat{\xi} \Rightarrow \min \end{aligned} \tag{14}$$

subject to the constraints imposed on:

- the class of (normalized) PDFs:

$$\int_{\mathcal{A}} P(A) \, dA = 1$$
$$\int_{\hat{\Xi}} Q(\hat{\xi}) \, d\hat{\xi} = 1 \tag{15}$$

- the balance between the first moment vector of the observed output $\mathbf{v}[T^- + ih] = \mathbf{v}^{(i)}$ in the LDRM Equation (7) and the real data vector $\mathbf{y}[T^- + ih] = \mathbf{y}^{(i)}$:

$$
\begin{aligned}
\mathcal{M}\{\mathbf{v}^{(i)}\} &= \bar{\mathbf{v}}^{(i)}[P(A), Q(\hat{\xi})] = \int_{\mathcal{A}} \mathbf{u}^{(i)}(A) P(A) dA + \\
&+ \int_{\hat{\Xi}} \bar{\xi}^{(i)} Q(\hat{\xi}) d\hat{\xi} = \mathbf{y}^{(i)}, \quad i \in \overline{0, N_{est}}
\end{aligned}
\tag{16}
$$

where:

$$
\mathbf{u}^{(i)}(A) = W(A|ih)\mathbf{x}^0 + \int_{T^-}^{T^- + ih} W(A|T^- + ih - \tau) \, B \, \mathbf{f}(\tau) d\tau
\tag{17}
$$

$\mathcal{S}_{PQ}^1$ entropy-robust estimation uses the first moment vector of the output LDRM. It is possible to use moments of higher order. This depends on measurable real data. If they represent the k-moment, then the balance condition can be formulated in the following form:

$$
\left(\mathcal{M}\{\mathbf{v}^k\} \right)^{1/k} = \mathbf{y}
\tag{18}
$$

where $\mathbf{v}^{(k)}$ is a vector of *k*-moment components. In this case, we will have $\mathcal{S}_{PQ}^k$ entropy-robust estimation.

The problem Equations (14)–(17) are related to the Lyapunov problem [28,29] where the goal functional and constraints are of an integral type. Here, we will use the necessary condition of equality to zero of the integral equation:

$$
\int_X h(\mathbf{x}) \, g(\mathbf{x}) d\mathbf{x} = 0
\tag{19}
$$

where the function $h(\mathbf{x})$ is continuous and is equal to zero on the boundary of the set X (the class $\tilde{\mathcal{C}}$); the function $g(\mathbf{x})$ is differentiable (the class $\mathcal{D}$). Then, this equality will be valid *for any function $h(\mathbf{x})$* with the mentioned properties if:

$$
g(\mathbf{x}) = 0, \qquad \forall \mathbf{x} \in X
\tag{20}
$$

Now, we return to the problem Equations (14)–(17) and introduce the Lagrange functional:

$$
\begin{aligned}
L[P(A), Q(\hat{\xi})] &= -\mathcal{H}[P(A), Q(\hat{\xi})] + \lambda \left(\int_{\mathcal{A}} P(A) \, dA - 1 \right) + \mu \left(\int_{\hat{\Xi}} Q(\hat{\xi}) \, d\hat{\xi} - 1 \right) + \\
&+ \sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \mathbf{m}^{(i)}[P(A), Q(\hat{\xi})] \rangle
\end{aligned}
\tag{21}
$$

where:

$$
\mathbf{m}^{(i)}[P(A), Q(\hat{\xi})] = \bar{\mathbf{v}}^{(i)}[P(A), Q(\hat{\xi})] - \mathbf{y}^{(i)}
\tag{22}
$$

Sign $\langle \bullet, \bullet \rangle$ denotes a scalar product.

As the solution of the problem Equations (14)–(17) is searched in the class of differentiable functions, then the Gato derivation can be used for determination of the variation of Lagrange functional Equation (21).

Let us denote the solution of the problem Equations (14)–(17) as $P^*(A)$ and $Q^*(\hat{\xi})$. Furthermore, introduce the functions $\phi(A) \in \tilde{\mathcal{C}}, \psi(\hat{\xi}) \in \mathcal{D}$ and two scalar variables α, β to present the functions $P(A), Q(\hat{\xi})$ in the following form:

$$P(A) = P^*(A) + \alpha\phi(A), \qquad Q(\hat{\xi}) = Q^*(\hat{\xi}) + \beta\psi(\hat{\xi}) \tag{23}$$

Functions $P^*(A), Q^*(\hat{\xi})$, as a solutions of the problem Equations (14)–(17), are fixed. The optimality conditions for the problem Equations (14)–(17) take the form:

$$\left.\frac{dL}{d\alpha}\right|_{\alpha=\beta=0} = 0, \qquad \left.\frac{dL}{d\beta}\right|_{\alpha=\beta=0} = 0 \tag{24}$$

The application of these conditions leads to the following systems of integral equations:

$$\int_{\mathcal{A}} \phi(A) \left(\ln P(A) + 1 + \lambda + \sum_{i=0}^{N_{est}} \langle \hat{\theta}^{(i)}, \mathbf{u}^{(i)}(A) \rangle \right) dA = 0$$

$$\int_{\hat{\Xi}} \psi(\hat{\xi}) \left(\ln Q(\hat{\xi}) + 1 + \mu + \sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \bar{\xi}^{(i)} \rangle \right) d\hat{\xi} = 0 \tag{25}$$

According to Equations (19) and (20), we obtain the following equations, which are necessary optimality conditions (necessary conditions of Lagrangian-stationarity) for the problem Equations (14)–(17):

$$\ln P(A) + 1 + \lambda + \sum_{i=0}^{N_{est}} \langle \hat{\theta}^{(i)}, \mathbf{u}^{(i)}(A) \rangle = 0$$

$$\ln Q(\hat{\xi}) + 1 + \mu + \sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \bar{\xi}^{(i)} \rangle = 0 \tag{26}$$

The solution of the problem Equations (14)–(17) takes the form:

$$P^*(A) = \frac{\exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \mathbf{u}^{(i)}(A) \rangle \right)}{\mathcal{R}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right)}$$

$$Q^*(\hat{\xi}) = \frac{\exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \bar{\xi}^{(i)} \rangle \right)}{\mathcal{Q}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right)} \tag{27}$$

where:

$$\mathcal{R}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right) = \int_{\mathcal{A}} \exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \mathbf{u}^{(i)}(A) \rangle \right) dA \tag{28}$$

$$\mathcal{Q}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right) = \int_{\hat{\Xi}} \exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \bar{\xi}^{(i)} \rangle \right) d\bar{\xi}^{(0)} \cdots d\bar{\xi}^{(N_{est})}$$

The vectors of Lagrange multipliers are determined from the following equations:

$$\mathbf{U}(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})}) = \int_{\mathcal{A}} \mathbf{u}^{(i)}(A) \frac{\exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \mathbf{u}^{(i)}(A) \rangle \right)}{\mathcal{R}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right)} dA + \tag{29}$$

$$+ \int_{\hat{\Xi}} \bar{\xi}^{(i)} \frac{\exp\left(-\sum_{i=0}^{N_{est}} \langle \bar{\theta}^{(i)}, \bar{\xi}^{(i)} \rangle \right)}{\mathcal{Q}\left(\bar{\theta}^{(0)}, \ldots, \bar{\theta}^{(N_{est})} \right)} d\bar{\xi}^{(0)} \cdots d\bar{\xi}^{(N_{est})} - \mathbf{y}^{(i)} = \mathbf{0}, \quad i \in [0, N_{est}]$$

Calculation of the vectors $\hat{\theta}^* = \{\bar{\theta}_*^{(0)}, \ldots, \bar{\theta}_*^{(N_{est})}\}$ is turned into a search for the global minimum of the residual function:

$$J(\hat{\theta}) = \|\mathbf{U}(\hat{\theta})\|_{L_2} \tag{30}$$

A global optimization algorithm is based on the simple Monte Carlo trials proposed in [30]. However, as soon as the L_2 metric is a convex function, one of the traditional gradient-based local optimization methods can be used for its solving.

4. Randomized Forecast Implementation

We comprehend a randomized forecast as an ensemble of trajectories on a forecasting interval $\mathcal{T}_{frc} = [t_0, T]$, which has to be generated using the model Equations (6) and (7) with the random matrix A and noise $\bar{\xi}$ described by the PDFs $P^*(A)$ and $Q^*(\bar{\xi})$, respectively; see formulas Equations (27) and (28). The matrices A and the vector noise $\bar{\xi}$ belong to the parallelepipeds $\mathcal{A}$ from Equation (3) and Ξ from Equation (4), respectively.

Let us study the generation problem of random matrices with the PDF Equation (27). First, we transform a matrix into a vector through concatenation of its rows. This procedure yields a vector $\mathbf{a}$ of length $m = n^2$. Additionally, the domain of random matrices becomes an m-dimensional parallelepiped:

$$\mathbb{A} = [\mathbf{a}^- \leq \mathbf{a} \leq \mathbf{a}^+] \tag{31}$$

where the vectors $\mathbf{a}^-$ and $\mathbf{a}^+$ result from the row concatenation of the matrices A^- and A^+, respectively.

Lets consider a transformation of a vector $\mathbf{a}$ into a vector $\mathbf{q}$ belonging to the m-dimensional unit nonnegative cube $\mathbb{Q}$:

$$\mathbf{a} = (\mathbf{a}^+ - \mathbf{a}^-)\mathbf{q} + \mathbf{a}^-, \quad \mathbb{Q} = \{\mathbf{q} : \mathbf{0} \leq \mathbf{q} \leq \mathbf{1}\} \tag{32}$$

Therefore, the entropy-optimal PDF undergoes the following chain of transformations:

$$P^*(A) \Rightarrow \mathbf{P}(\mathbf{a}) \Rightarrow \mathbb{P}(\mathbf{q}) \tag{33}$$

Therefore, it is necessary to generate random vectors $\mathbf{q} \in \mathbb{Q}$ with PDFs $\mathbb{P}(\mathbf{q})$. The generation was implemented by the acceptance-rejection algorithm [24].

5. Application of the RF Method for World Population Prediction

5.1. The World Population Prediction Problem

The state of an isolated population is characterized by its size $E(t)$ on a calendar time interval $\mathcal{T} = [T^-, t_0]$. Population size varies under the impact of fertility and mortality processes, since World population is an isolated system. Within the framework of the linear population dynamics model, fertility and mortality are described by corresponding rates, whereas the flows of newborns and decedents appear proportional to population size, while fertility (b) and mortality (m) rates are considered as linear time-dependent parameters [22].

5.1.1. Randomized population model

World population evolves according to the following differential equation that has an analytical solution:

$$
\begin{aligned}
\frac{dE(t)}{dt} &= (b(t) - m(t))E(t), \quad b(t) = b_0 + u_b t, \quad m(t) = m_0 + u_m t \\
E(t) &= E_0 \exp\left(\left((b_0 - m_0) + \frac{1}{2}(u_b - u_m)t\right)t\right), \qquad \text{where} \\
E_0 &= E(T^-), \quad b_0 = b(T^-), \quad m_0 = m(T^-)
\end{aligned}
\tag{34}
$$

Real measurements of the population size dynamics modeled by Equation (34) take place at discrete moments. Hence, the population size at discrete moments ih (where h specifies a given increment) is defined by the expression:

$$
\begin{aligned}
\Phi_i(r, u_r \mid E_0) &= E_0 \exp\left[(r + u_r i)ih\right], \qquad i \in I \\
r &= b_0 - m_0, \qquad u_r = \frac{1}{2}(u_b - u_m)h
\end{aligned}
\tag{35}
$$

with parameters r and u_r, which describe the result of the difference between fertility and mortality flows.

World population is measured in *billions* of people. Fertility and mortality processes aggregate many factors, including measurement errors, whose quantitative analysis is impossible or complicated. On the other hand, the mass nature of fertility and mortality processes admits their modeling based on the probabilistic approach.

Thus, the resultant flow rate and its changing in time are supposed to be random variables with a joint probability density function $P(r, u_r)$ defined on the rectangle:

$$
\mathcal{J} = \mathcal{I}_r \bigcup \mathcal{I}_{u_r}, \qquad \mathcal{I}_r = [r^-, r^+], \quad \mathcal{I}_{u_r} = [u_r^-, u_r^+]
\tag{36}
$$

Generally, measurement errors are modeled by an additive noise $\xi[ih]$ of the interval type:

$$
\xi[ih] \in \Xi_i = [\xi_i^-, \xi_i^+], \qquad i \in I
\tag{37}
$$

where I indicates a time interval of such measurements. By assumption, the PDFs $q_i(\xi[ih])$, $i \in I$ are specified on the intervals Ξ_i from Equation (37). Due to the independence of the set of random variables $\xi[0], \ldots, \xi[ih]$, their joint PDF has the form:

$$
Q(\bar{\xi}) = \prod_{i \in I} q_i(\xi[ih])
\tag{38}
$$

Therefore, the *randomized* model of World population dynamics can be described by:

$$
v[ih] = \Phi_i(r, u_r | E_0) + \xi[ih], \qquad i \in I
\tag{39}
$$

where the function $\Phi_i(r, u_r | E_0)$ meets Equality (35).

5.1.2. Real and forecasting data

For PDF estimation, address the World population measurements for the period from 1960–1995 (`http://data.un.org/`; see Table 1).

The entropy-optimal RM is tested via the measurements of the World population dynamics during the period from 1995–2015 and the values E_{1985}^{prn} for this period (according to the UN forecast announced in 1985) (`http://www.irbis.vegu.ru/repos/1002/Html/27.htm`; see Table 2).

The World population prediction till 2050 using the UN data is illustrated by Table 3 (`http://data.un.org/`; [31]). United Nations' projections both for the testing interval and the forecasting interval were made in accordance with the commonly-used cohort-component method, based on age-specific estimates of the components of population change (fertility, mortality and international migration) [32,33]. We will compare this prediction with its randomized analog.

World population is measured in billions of people. In the sequel, the subscript *real* indicates the measured data of the population.

To summarize, the RM has the following forms on corresponding time intervals:

- on the estimation interval $\mathcal{T}_{est}$ (see Table 1):

$$v[ih] = \Phi_i(r, u_r | E_{real}^{est}[0]) + \xi[ih], \qquad i \in [0, 7] \tag{40}$$

- on the testing interval $\mathcal{T}_{tst}$ (see Table 2):

$$v[ih] = \Phi_i(r, u_r | E_{real}^{tst}[0]) + \xi[ih], \qquad i \in [0, 4] \tag{41}$$

- on the forecasting interval $\mathcal{T}_{frc}$ (see Table 3):

$$v[ih] = \Phi_i(r, u_r | E_{real}^{frc}[0]), \qquad i \in [0, 5] \tag{42}$$

where r and u_r are random parameters with the entropy-optimal PDFs $P^*(r, u_r)$ and $\bar{\xi}$ is a vector of random noise with entropy-optimal PDF $Q^*(\bar{\xi})$.

Table 1. Estimation interval $\mathcal{T}_{est}$.

i	0	1	2	3	4	5	6	7
year	1960	1965	1970	1975	1980	1985	1990	1995
E_{real}^{est}	3.026	3.358	3.691	4.070	4.449	4.884	5.320	5.724

Table 2. Testing interval $\mathcal{T}_{tst}$.

i	0	1	2	3	4
year	1995	2000	2005	2010	2015
E_{real}^{tst}	5.724	6.128	6.514	6.916	7.359
E_{1985}^{prn}	5.666	5.962	6.450	6.985	7.469

Table 3. Forecasting interval $\mathcal{T}_{frc}$.

i	0	1	2	3	4	5
year	2015	2020	2025	2030	2040	2050
E_{UN}^{frc}	7.359	7.644	7.964	8.284	8.924	9.564

5.1.3. The entropy-optimal PDFs of the parameters and noise

According to the general entropy-robust estimation procedure of PDFs (see Section 2), we have:

- the PDF of the parameters r and u_r in the form:

$$P^*(r, u_r) = \frac{1}{\mathcal{R}(\bar{\theta} \mid E_{real}^{est}[0])} \prod_{j=0}^{7} p_j^*(r, u_r | \theta_j), \quad p_j^*(r, u_r | \theta_j) = \exp\left(-\theta_j \Phi_j(r, u_r | E_{real}^{est}[0])\right) \tag{43}$$

- the PDF of the noise in the form:

$$Q^*(\bar{\xi}) = \frac{1}{\mathcal{Q}(\bar{\theta})} \prod_{j=0}^{7} q_j^*(\xi[jh]|\theta_j), \quad q_j^*(\xi[jh]|\theta_j) = \exp\left(-\theta_j \xi[jh]\right) \tag{44}$$

where:

$$\mathcal{R}(\bar{\theta} \mid E_{real}^{est}[0]) = \int_{\mathcal{J}} \prod_{j=0}^{7} \exp\left(-\theta_j \Phi_j(r, u_r | E_{real}^{est}[0])\right) dr \, du_r \tag{45}$$

$$\mathcal{Q}(\bar{\theta}) = \prod_{j=0}^{6} \int_{\xi_j^-}^{\xi_j^+} \exp(-\theta_j \xi[jh]) d\xi[jh] =$$

$$= \prod_{j=0}^{6} \frac{1}{\theta_j} \left(\exp(-\theta_j \xi_j^-) - \exp(-\theta_j \xi_j^+) \right) \tag{46}$$

To calculate the Lagrange multipliers, we will solve the system of balance equations (see Equations (16) and (17)):

$$\frac{1}{\mathcal{R}(\bar{\theta} \mid E_{real}^{est}[0])} \int_{\mathcal{J}} \Phi_i(r, u_r | E_{real}^{est}[0]) \prod_{j=0}^{7} \exp\left(-\theta_j \Phi_j(r, u_r | E_{real}^{est}[0])\right) \, dr \, du_r +$$

$$+ \quad \frac{1}{\mathcal{Q}(\bar{\theta})} \int_{\Xi} \xi[ih] \prod_{j=0}^{7} \exp\left(-\theta_j \xi[jh]\right) d\xi[ih] - E_{real}^{est}[ih] = 0, \qquad i \in [0,7] \tag{47}$$

We will denote:

$$\mathcal{N}_i(\bar{\theta} \mid E_{real}^{est}[0]) = \int_{\mathcal{J}} \Phi_i(r, u_r | E_{real}^{est}[0]) \prod_{j=0}^{7} \exp\left(-\theta_j \Phi_j(r, u_r | E_{real}^{est}[0])\right) dr \, du_r \tag{48}$$

Then, Equation (47) can be rewritten as:

$$\mathcal{G}_i(\bar{\theta} \mid E_{real}^{est}[0]) = \frac{\mathcal{N}_i(\bar{\theta} \mid E_{real}^{est}[0])}{\mathcal{R}(\bar{\theta} \mid E_{real}^{est}[0])} + \mathcal{L}_i(\theta_i) - y_i = 0 \tag{49}$$

$$y_i = E_{real}^{est}[ih], \quad i = [0,7]$$

where:

$$\mathcal{L}_i(\theta_i) = \frac{\exp(-\theta_i \xi_i^-)(\xi_i^- + \frac{1}{\theta_i}) - \exp(-\theta_i \xi_i^+)(\xi_i^+ + \frac{1}{\theta_i})}{\exp(-\theta_i \xi_i^-) - \exp(-\theta_i \xi_i^+)} \tag{50}$$

We endeavor to solve these equations through minimizing the residual function:

$$J(\bar{\theta}) = \|\mathcal{G}(\bar{\theta})\| \Rightarrow \min \tag{51}$$

5.2. The Results of Computer Experiment

5.2.1. General conditions

We have performed calculations for the stages of estimation, testing and prediction with the following ranges for the model parameters, which include the possibility for both positive and negative trends of World population growth:

$$\mathcal{I}_r = [-0.025; 0.075], \quad \mathcal{I}_{u_r} = [-0.002; 0.001] \tag{52}$$

and identically the range for the measurement noises:

$$\Xi_j = [-0.5; 0.5], \qquad j \in [0, 7] \tag{53}$$

Under the above ranges of the model parameters and measurement noises, the computer experiment has constructed the entropy-optimal PDFs and generated the ensembles of corresponding random trajectories for RM testing and randomized forecasting.

5.2.2. Estimation

On the estimation interval, we employ the data from Table 1. The residual function $J(\bar{\theta})$ Equation (51) is a function of eight variables, and it contains integral component Equations (45)–(49) to be estimated only numerically. For this, we have selected the so-called tiled method of two-dimensional integral estimation, which represents a combination of several quadrature formulas [34].

The idea of the tiled method consists of: (1) partitioning the whole domain of integration into a set of smaller-area subdomains having the rectangular or trapezoidal shape; and (2) applying appropriate quadrature formulas on each subdomain. The described method is implemented in *MATLAB* by the function quad2d.

Minimization of the residual function $J(\bar{\theta})$ (51) runs by the nonlinear trusted region method implemented by the function lsqnonlin from the package *Optimization Toolbox*. The function lsqnonlin has been optimized [35] for nonlinear least-squares problems.

The function lsqnonlin have several user-defined options for stopping criteria, such as function evaluation tolerance (10^{-6}), step size tolerance (10^{-6}) and maximum number of iterations (500). Table 4 presents the calculated Lagrange multipliers for the above ranges.

Table 4. Calculated Lagrange multipliers.

Range $\mathcal{I}_r$, $\mathcal{I}_{u_r}$								
Measurements	0	1	2	3	4	5	6	7
$\bar{\theta}$	0.0000	-0.3833	-0.3984	-0.5839	-0.3802	-0.4679	-0.1812	0.8881

Figures 1 and 2 demonstrate the entropy-optimal PDFs for the parameters of Model (35) and noise components. The functions $P^*(r, u_r, \bar{\theta})$ and $q^*(\xi[ih])$, $i \in [0, 7]$ represent an entropy-optimal distribution of random variables at the corresponding intervals and will be used for making randomized predictions.
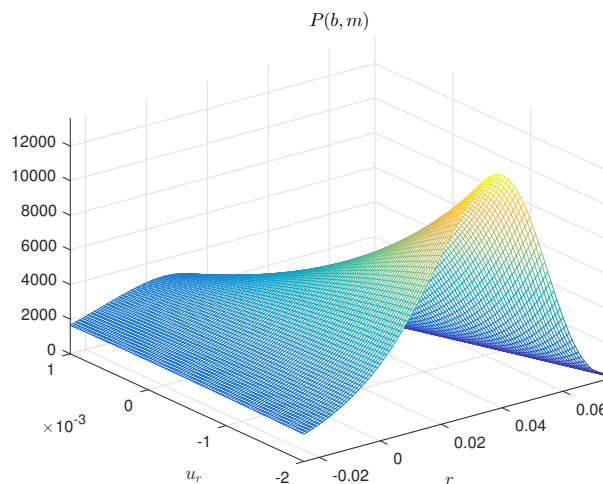

Figure 1. The joint PDF of the random parameters r and u_r for the range $\mathcal{J} = \mathcal{I}_r \bigcup \mathcal{I}_{u_r}$.
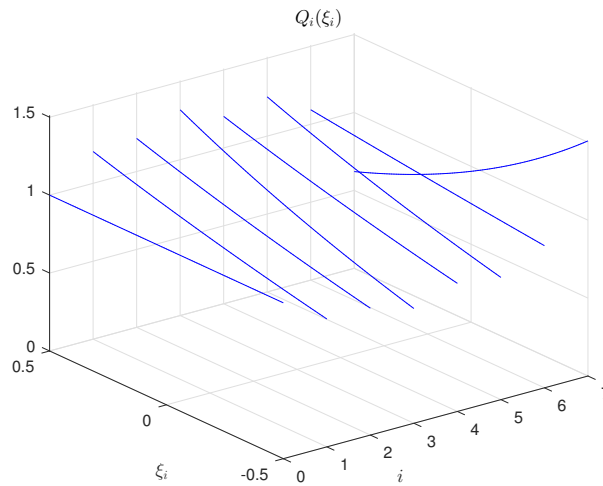
$$Q_i(\xi_i)$$

Figure 2. The family of the PDFs of the noise ξ_i, $i \in [0, 7]$.

5.2.3. Testing

Testing of the model has been made using the data from Table 2. The population size has been evaluated by the Formula (41), where r, u_r are the random variables with the PDF $P^*(r, u_r)$ and $\xi[ih]$ are the random noises with the PDFs $q_i(\xi[ih])$, $i \in [0, 4]$ (see Figures 1 and 2).

To generate an ensemble of random variables, the two-dimensional modification of the Ulam–von Neumann method has been used [24]. The size of the generated sample is $k = 100,000$.

Each pair of the random values r and u_r defines a separate exponential growth curve; moreover, for each point ih, a random value of the noise $\xi[ih]$ is added according to its PDF. As a result, the constructed trajectory of World population dynamics is not an exponential function.

The test procedure yields the probabilistic characteristics of the parameters r, u_r and ξ. It depends on the initial population size for the testing interval. Figure 3 shows the ensemble of the model-based trajectories for the parameter range from Equation (52) and the noise range from Equation (53), with a data-based selection of the initial point:

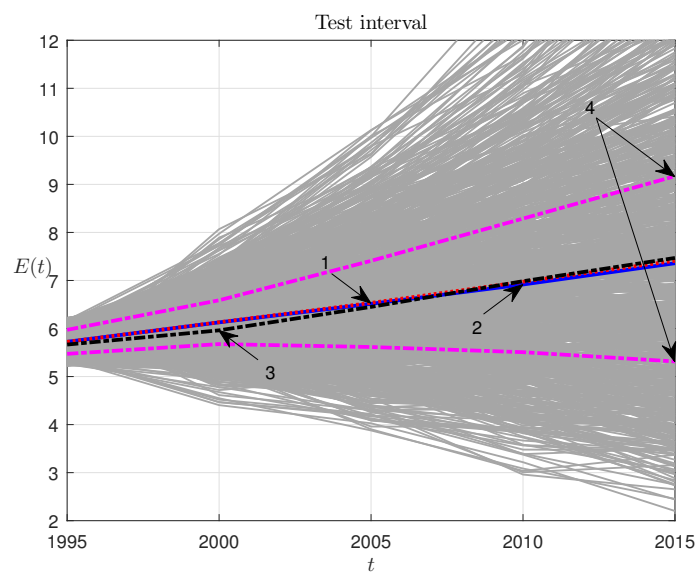$$E^{tst}[0] = E^{tst}_{real}[0] \tag{54}$$

Figure 3. The ensemble of projection trajectories on the testing interval.

Figure 3 has the following notation: 1, the ensemble-average trajectories of population dynamics; 2, real population dynamics on the testing interval; 3, population dynamics on the testing interval according to the UN forecast of the year 1998; 4, the boundaries of the first and the third quartiles. The same model-generated ensemble for testing interval can be presented as a box plot with median mark and interquartile ranges; see Figure 4.

Testing quality is assessed by the root-mean-square error of the average trajectory with respect to its real counterpart:

$$\delta = \|E^{tst} - E^{tst}_{real}\| = \sqrt{\sum_{i=0}^{4}(E^{tst}_{real}[ih] - E^{tst}[ih])^2} \tag{55}$$

as well as by the relative error:

$$\varepsilon = \frac{\delta}{\|E^{tst}_{real}\| + \|E^{tst}\|} \tag{56}$$

For instance, the UN forecast for the testing interval (Table 2) has the following errors:

$$\delta_{1985} = 0.228, \qquad \varepsilon_{1985} = 0.008 \tag{57}$$

In our case, the deviation of the model-average trajectory from the real one (Figure 3) appears to be appreciably smaller and demonstrates the following errors:

$$\delta_{RM} = 0.079, \qquad \varepsilon_{RM} = 0.003 \tag{58}$$

5.2.4. Prediction

This simple RM has been applied to predict World population dynamics for the period from 2015–2050. The trajectory ensemble corresponding to the UN predictions (Table 3) is illustrated by Figure 5: 1, the ensemble-average trajectory; 2, UN projection; 4, the boundaries of the interquartile range (IQR zone). The corresponding box-plot is presented in Figure 6.
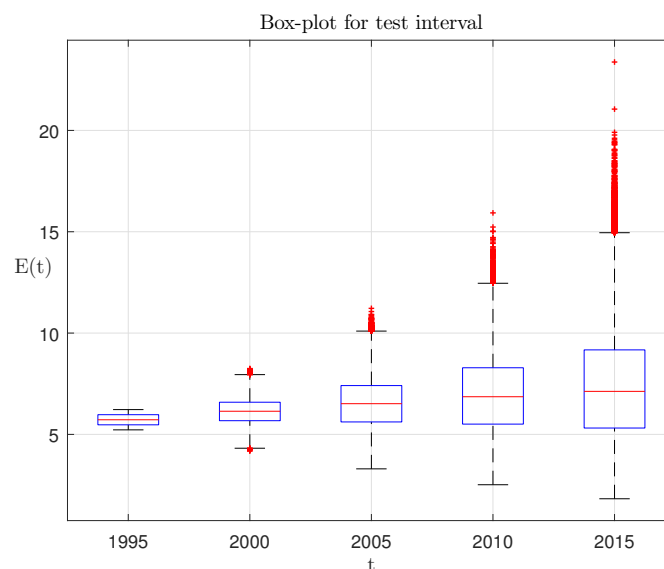

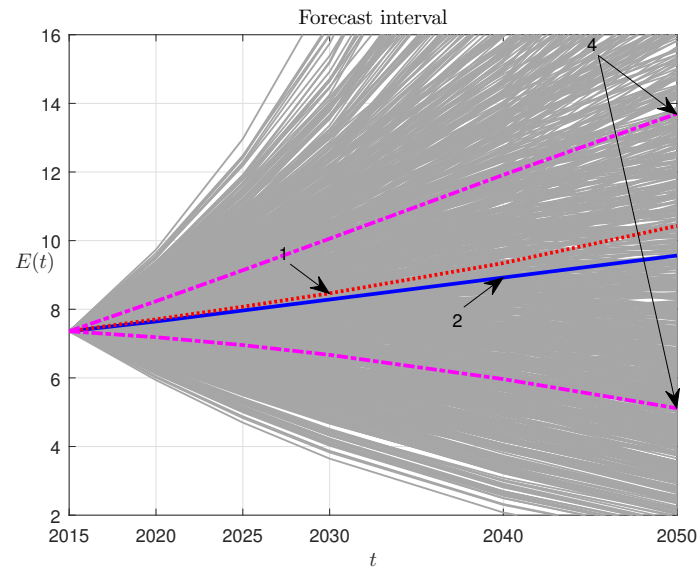**Figure 4.** Box plot for the ensemble of projection trajectories on the testing interval.

Figure 5. The ensemble of projection trajectories on the prediction interval.
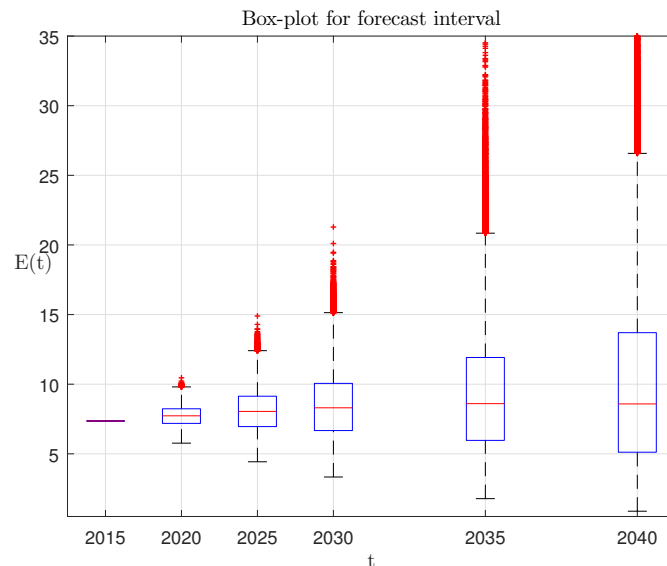

Figure 6. Box plot for the ensemble of projection trajectories on the testing interval.

The presented results testify that the randomized forecasting as opposed to existent methods gives a set of probability characteristics of the World population prediction, which is calculated by using the ensemble of prognostic trajectories. The latter is generated by the randomized dynamic model with entropy-optimal PDFs of parameters. The randomized projection algorithm shows significantly closer to real data numerical results for testing interval, as well as stable projection that is higher, but close to the modern UN forecast for the future. According to our randomized model, 2026 will be the first eight billion year.

6. Conclusions

In this paper, we have suggested a randomized forecasting method that operates dynamic models described by linear ordinary differential equations with random parameters. Entropy-robust estimation has been developed for the probability density functions (PDFs) of model parameters and noisy measurements based on entropy maximization. It has been shown that the above PDFs belong to the exponential class. The randomized forecasting technique has been applied to randomized

prediction of the World population dynamics. It has been demonstrated that randomized forecasting gives a set of probability characteristics of the World population dynamics.

Author Contributions: All authors have contributed significantly to the work reported by this article. Y.S. developed and declared the method; Y.S. and Y.A. conceived and designed the experiments; Y.A. performed the experiments; Y.A. and A.Y. analyzed the data; A.Y. contributed reagents/materials/analysis tools; Y.S., Y.A and A.Y wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Beer, J., van Wissen, L., Eds. *Europe: One Continent, Different Worlds. Population Scenarios for the 21th Century*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999.

2. Hilderink, H. *World Population in Transition. An Integrated Regional Modelling Framework*; Groningen University Press: Groningen, The Netherlands, 2000.

3. Feichtinger, G., Ed. *Vienna Yearbook of Population Reseach*; Vienna Institute of Demography, Austrian Academy of Sciences: Vienna, Austria, 2003.

4. Wissen, L.J.G. PROFILE: Population Research in the Netherlands. *Public Serv. Rev. Eur. Union* **2013**, *25*, 33.

5. Kendall, M.G.; Stuart, A. *The Advanced Theory of Statistics; Vol. II, Inference and Relationship*, 3rd ed.; Griffin: London, UK, 1973.

6. Cramer, H. *Mathematical Methods of Statistics*; Princeton Univ. Press: Princeton, NJ, USA, 1999.

7. Aivazyan, S.A.; Mkhitryan, V.S. Prikladnaya statistika i osnovy ekonometriki. In *Applied Statistics and Foundations of Econometrics*; Interreklama: Moscow, Russia, 2003.

8. Kobzar, A.I. Prikladnaya matematicheskaya statistika. In *Applied Mathematical Statistics*; Fizmatlit: Moscow, Russia, 2006.

9. Cho, S. A Linear Bayesian Stochastic Approximation to Update Project Duration Estimation. *Eur. J. Oper. Res.* **2009**, *196*, 585–593.

10. Zellner, A. Bayesian Shrinkage Estimation and Forecasts of Individual and Total or Aggregate Outcomes. *Econ. Model.* **2010**, *27*, 1392–1397.

11. Horvath, R. Research & Development and Growth: A Bayesian Model Averaging Analysis. *Econ. Model.* **2011**, *28*, 2669–2673.

12. Kim, M.J.; Jiang, R.; Makis, V.; Lee, C.-G. Optimal-Bayesian Fault Prediction Scheme for Partially Observable System Subject to Random Failure. *Eur. J. Oper. Res.* **2011**, *214*, 331–339.

13. Musal, R.M.; Soyer, R.; McCabe, C.; Kharroubi, S.A. Estimating the Population Utility Function: A Parametric Bayesian Approach. *Eur. J. Oper. Res.* **2012**, *218*, 538–547.

14. Borisov, A.V. A posteriori Minimax Estimation with Limit Likelihood. *Autom. Remote Control* **2012**, *9*, 49–56.

15. Lawrence, M.; Goodwin, P.; O'Connor, M.; Öncal, D. Judgemental Forecasting: A Review of Progress over the Last 25 years. *Int. J. Forecast.* **2006**, *22*, 493–518.

16. Kociecki, A.; Kolasa, M.; Rubaszek, M. A Bayesian Method of Combining Judgmental and Model-Based Density Forecasts. *Econ. Model.* **2012**, *29*, 1349–1355.

17. Lahiri, K.; Peng, H.; Zhao, Y. Testing the Value of Probability Forecasts for Calibrated Combining. *Int. J. Forecast.* **2015**, *31*, 113–129.

18. Jacobson, M.Z. *Fundamentals of Atmospheric Modeling*; Cambridge University Press: New York, NY, USA, 2005; p. 828.

19. Allen, M.R.; Stainforth, D.A. Towards Objective Probabilistic Climate Forecasting. *Nature* **2002**, *419*, doi:10.1038/nature01092a.

20. Popkov, A.Y.; Popkov, Y.S.; van Wissen, L. Positive Dynamic Systems with Entropy Operator: Application to Labour Market Modeling. *Eur. J. Oper.l Res.* **2005**, *164*, 811–828.

21. Kapitsa S.P. Obshchaya teoriya rosta naseleniya Zemli. In *The General Theory of World Population Growth*; Nauka: Moscow, Russia, 1999.

22. Popkov, Y.S. *Mathematical Demoeconomy. Integrating Demographic and Economic Approach*; De Gruyter: Berlin, Germany, 2014; p. 534.

23. Kaashoek, M.A.; Seatzu, S.; van der Mee, C. *Recent Advances in Operator Theory and Its Applications*; Springer: Berlin, Germany; Heidelberg, Germany, 2006; p. 478.

24. Rubinstein, R.Y.; Kroese, D.P. *Simulation and the Monte Carlo Method*; John Wiley & Sons: Chichester, UK, 2008.

25. Popkov, Y.S. *Macrosystems Theory and its Applications* ; Springer: New York, NY, USA, 1995; p. 340.

26. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630.

27. Popkov, Y.S.; Popkov, A.Y. New Methods of Entropy-Robust Estimation for Randomized Models under Limited Data. *Entropy* **2014**, *16*, 675–698.

28. Ioffe, A.D.; Tihomirov, V.M. *Theory of Extremal Problems (Studies in Mathematics and Its Applications)*; Nauka: Moscow, Russia, 1979.

29. Alekseev, V.M.; Tihomirov, V.M.; Fomin, S.V. *Optimal Control*; Nauka: Moscow, Russia, 1987.

30. Darkhovskii, B.S.; Popkov, Y.S. Popkov, A.Y. Monte Carlo Method of Batch Iterations: Probabilistic Characteristics. *Autom. Remote Control* **2015**, *76*, 775–784.

31. Gonzalo, J.A.; Munoz, F.-F.; Santos, D.J. Using a Rate Equation Approach to Model World Population Trends. *Simul. Trans. Soc. Model. Simul. Int.* **2013**, *89*, 192–198.

32. Shryock, H.S.; Siegel, J.S. *The Methods and Materials of Demography*; United Nations Department of Commerce: Washington, DC, USA, 1973.

33. *World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections*; Working Paper No. ESA/P/WP.242; Department of Economic and Social Affairs, Population Division: New York, NY, USA, 2015.

34. Shampine, L.F. MatLab Program for Quadrature in 2D. *Appl. Math. Comput.* **2008**, *202*, 266–274.

35. Coleman, T.F.; Li, Y. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM J. Optim.* **1996**, *6*, 418–445.