



Entropy-Randomized Forecasting of Stochastic Dynamic Regression Models

Yuri S. Popkov ^{1,2,3,*}, Alexey Yu. Popkov ¹, Yuri A. Dubnov ^{1,4} and Dimitri Solomatine ⁵

- ¹ Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 119333 Moscow, Russia; popkov@isa.ru (Y.S.P.); apopkov@isa.ru (A.Y.P.); yury.dubnov@phystech.edu (Y.A.D.)
- ² Institute of Control Sciences of Russian Academy of Sciences, 117997 Moscow, Russia
- ³ Department of Software Engineering, ORT Braude College, 2161002 Carmiel, Israel
- ⁴ National Research University "Higher School of Economics", 101000 Moscow, Russia
- ⁵ IHE Delft Institute for Water Education, 2601 Delft, The Netherlands; d.solomatine@un-ihe.org
- * Correspondence: popkov@isa.ru

Received: 20 May 2020; Accepted: 6 July 2020; Published: 8 July 2020



Abstract: We propose a new forecasting procedure that includes randomized hierarchical dynamic regression models with random parameters, measurement noises and random input. We developed the technology of entropy-randomized machine learning, which includes the estimation of characteristics of a dynamic regression model and its testing by generating ensembles of predicted trajectories through the sampling of the entropy-optimal probability density functions of the model parameters and measurement noises. The density functions are determined at the learning stage by solving the constrained maximization problem of an information entropy functional subject to the empirical balances with real data. The proposed procedure is applied to the randomized forecasting of the daily electrical load in a regional power system. We construct a two-layer dynamic model of the daily electrical load. One of the layers describes the dependence of electrical load on ambient temperature while the other simulates the stochastic quasi-fluctuating temperature dynamics.

Keywords: forecasting; randomization; dynamic regression; information entropy; empirical balance; randomized machine learning

1. Introduction

Due to the gradually increasing resources and computational power of computers, huge amounts of data can be accumulated and stored, both in natural and digitized formats. Then, the following question arises immediately: what should be done with these data, except for storage? Extracting new *knowledge* from data seems to be a very interesting idea. The concepts of Data Mining (DM) [1,2], Big Data (BD) [3] and Data Science (DS) [4] were formulated and further developed by researchers accordingly.

A very tempting goal—extracting new knowledge from data—inevitably leads to the verbal or formal (mathematical) modeling of the "expected" knowledge. Therefore, any model has some predictive properties, which can be implemented only under known values of its quantitative characteristics (parameters). Data are a fundamental component of the three concepts above: data are adopted for estimating the characteristics of a model using *machine learning (ML)* procedures, which allows extracting new knowledge.

Unlike DM, BD, and DS, the concept of ML has a rich history of over 70 years as well as vast experience in solving numerous problems. The first publication in this field of research dates back to 1957; see [5]. The notion of empirical risk, a key element of *ML* procedures, was introduced in 1970 in the monograph [6]. The method of potential functions for classification and recognition problems



was also presented in 1970 in another monograph [7]. The modern concept of *ML* is based on the *deterministic* parametrization of models and estimates using data sets with postulated properties. The quality of estimation is characterized by empirical risk functions, and their minimization gives optimal estimates [8,9].

As a rule, real problems that are solved by ML procedures are immersed in some uncertain environment. If the matter concerns data, they are acquired with inevitable errors, omissions or low reliability. The design and parametrization of models is a non-formalizable and subjective process that depends on the individual knowledge of a researcher. Therefore, in the mass application of *ML* procedures, the level of uncertainty is quite high.

All these circumstances indicate that it is necessary to somehow compensate uncertainty. Here a general trend is the application of its stochastic description to parametrized models and data. This means that the model parameters are assumed to be random (appropriately randomized), and the data are assumed to have random errors. Machine learning procedures with these properties belong to the class of *randomized machine learning* (*RML*) procedures. Their difference from the conventional ML procedures is that optimal estimates are constructed not for the parameters but for the probability density functions (PDFs) of the random parameters and the PDFs of the worst-case random errors in the data. In the entropy-based RML procedures, the functional of generalized information entropy [10] is used as an optimality criterion for estimates.

The core of an RML procedure is a parametrized predictive model designed for simulating the temporal (or spatiotemporal) evolution of a process under study. Therefore, such a model belongs to the class of dynamic models. *Parametric dynamic regression models (PDRMs)* are most widespread representatives of this class, in which the current state of a model is determined by its past states on a certain time interval [11,12]. A formal image of PDRMs is difference equations, in the general case of the *p*th order [13]. Most applications are described by linear PDRMs. In particular, they naturally occur in many problems of macroeconomic modeling and forecasting, e.g., time series analysis of economic indices [14], adequacy analysis of PDRMs [15], and prediction of exchange rates [16]. Linear PDRMs are effective enough for short-term forecasting, yet causing significant errors on large forecasting horizons. Therefore, attempts to improve forecasts by introducing various nonlinearities into *PDRMs* are quite natural. The monograph [17] was dedicated to a general approach to the formation and use of nonlinear *PDRMs*. However, applications require a more "personalized" approach to choosing the most useful and effective nonlinearity. On this pathway, it seems fruitful, e.g., to forecast exchange rates using logistic and exponential nonlinearities [18], or to predict the daily electrical load of a power system using periodic autoregressive models [19] or multidimensional time series [20].

Since forecasting is performed under uncertainty, the resulting errors caused by some unaccounted factors are often compensated, if possible, by assigning some probabilities to forecasts [21,22]. The most common approach is the use of Bayes' theorem on posterior probability. Let a parametrized conditional probability density function of data and an a priori probability density function of parameters be specified; then their normalized product will determine the posterior probability density function of the parameters under fixed data. Fundamental problems in this field of investigations are connected with the structural choice of the conditional and prior PDFs. Typically, Gaussian PDFs or their mixture are selected, and the mixture weights are estimated using retrospective data [23–25]. A similar approach was adopted in applications: population genetics [26], where the method of numerical approximation of posterior PDFs was developed; the interaction between the financial sector of the economy and the labor market [27], where the Metropolis–Hastings algorithm was used to estimate the parameters of the above PDFs; population dynamics [28], where a hierarchy of Bayesian models was designed to predict fertility, mortality and migration rates. Probabilistic forecasts are constructed by other methods, taking into account the specifics of applications. In meteorology, retrospective weather forecasts are accumulated for estimating the PDFs; subsequently, these PDFs are used for short-term forecasting [29–32]. A rather interesting procedure is to form a probabilistic forecast as a mixture of forecasts obtained by different methods [33].

In this paper, we propose a fundamentally different forecasting method—the so-called *entropy-randomized forecasting (ERF)*. In accordance with this method, an ensemble of random forecasts is generated by a *predictive dynamic regression model (PDRM)* with random input and parameters. The corresponding probabilistic characteristics, namely the probability density functions, are determined using the entropy randomized machine learning procedure. The ensembles of forecasting trajectories are constructed by the sampling of the entropy-optimal PDFs.

The proposed method is adopted for randomized prediction of the daily electrical load of a regional power system. A hierarchical randomized dynamic regression model that describes the dependence of the load on the ambient temperature is constructed. The temporal evolution of the ambient temperature is represented by an oscillatory second-order dynamic regression model with a random parameter and a random input. The results of randomized learning of this model on the GEFCom2014 dataset [34] are given. A randomized forecasting technology is suggested and its adequacy is investigated depending on the length of forecasting horizon.

2. Procedure of Entropy-Randomized Forecasting

Randomization as a means of imparting artificial and rationally organized random properties to naturally nonrandom events, indicators, or methods is a fairly common technique that yields a positive effect. There exist many examples in various fields of science, management, economics: randomized numerical optimization methods [35,36]; the mixed (random) strategies of trading on a stock exchange [37]; the randomized forecasting of population dynamics [38]; vibration control of industrial processes [39]. As the result of randomization, nonrandom objects gain artificial stochastic properties with optimal probabilistic characteristics in a chosen sense. The question on appropriate quantitative characteristics of optimality has always been controversial and ambiguous. It requires arguments that would somehow reflect the important specifics of a randomized object. In particular, a fundamental feature of forecasting procedures is *uncertainty* in the data, predictive models, methods for generating forecasts, etc.

In what follows, *information entropy* [40] will be used as a characteristic of uncertainty. In the works [41–43], using the first law of thermodynamics it was demonstrated that entropy is a natural functional describing the processes of universal evolution. Moreover, in accordance with the second law of thermodynamics, entropy maximization determines the best state of an evolutionary process under the worst-case external disturbance (maximum uncertainty). Also, note another quality of information entropy associated with measurement errors and other types of errors, which are important attributes of data: with the factor of such errors being considered in terms of informational entropy, the probabilistic characteristics of noises exerting the worst-case impact on forecasting procedures can be estimated in explicit form.

The technology of entropy-randomized forecasting consists of the following stages. At the beginning (the first stage), *a predictive randomized model (PRM)* of a studied object is formed and its parameters are designed. A PRM transforms real data into a model output. In the general case, these transformations are assumed to be dynamic, i.e., the model output observed at a time instant *n* depends on the states observed on some past interval. The PRM parameters are assumed to be of the interval type and random, and their probabilistic properties are characterized by the corresponding PDFs.

The second stage of the technology under consideration—randomized machine learning (more specifically, its entropy version)—is intended to estimate the PDFs. At this stage, the estimates of the PDFs are calculated using learning data sets and also a learning algorithm in the form of a functional entropy-linear programming problem.

At the third stage, the optimized PRM (with the entropy-optimal PDFs) is tested using a test data set and accepted quantitative characteristics of the quality of learning. The optimized PRM actually generates an ensemble of random trajectories, vectors, or events with the entropy-optimal values of their parameters.

The learned and tested PRMs serve for forecasting. In this case, the ensembles of random forecasted trajectories generated by the entropy-optimal PRMs are used to calculate their numerical characteristics such as mean trajectories, variance curves, median trajectories, the PDF evolution of forecasted trajectories, etc.

3. Randomized Dynamic Regression Models with Random Input and Parameters

Randomized dynamic regression models (RDRMs) form a class of dynamic models with random parameters that describe a parametrized dependence of the object's state at a given time instant on external factors and its states at some past time instants.

The structures of models are designed on the basis of existing knowledge and hypotheses about the properties of an object, which often turn out to be very inaccurate. Moreover, the external factors themselves can change over time and therefore should be predicted to model the object's dynamics. Reliable information on realistically measured impacts leading to the temporal evolution of external factors is often unavailable. The aforementioned indicates the presence of *uncertainty*, both in the development and further use of models. In [10], a method for reducing the influence of uncertainty based on the randomization of models (including the class of RDRMs) was proposed. In the latter case, this method extends the idea of randomization to the modeling of external factors and their evolution.

The structure of the RDRM is shown in Figure 1. It consists of a model of the main object (RDRM-O) with random parameters $\mathbf{a} \in R^p$ and a model of external factors (RDRM-F) with random parameters $\mathbf{b} \in R^s$ and a random input $\zeta \in R^q$. The states of the object and its model belong to the vector space R^m , in which $\hat{\mathbf{x}}[n]$ are the state vectors of the object and $\mathbf{x}[n] \in R^m$ are the state vectors of RDRM-O. The external factors are characterized by the vector $\hat{\mathbf{y}}[n] \in R^q$ while the changes in the state of RDRM-F over time by the vector $\mathbf{y}[n] \in R^q$. The variable *n* denotes discrete time taking integer values on the interval $\mathcal{L} = [n^-, n^+]$.



Figure 1. Structure of the RDRM.

Consider *the linear version* of RDRM-O. Its state $\mathbf{x}[n]$ at a time instant n is changing under the influence of p retrospective states $\mathbf{x}[n-1], \ldots, \mathbf{x}[n-p]$ and measurable external factors $\mathbf{z}[n] \in \mathbb{R}^{q}$. The corresponding equation has the form

$$\mathbf{x}[n] = \mathbf{X}^{(n,p)} \mathbf{A}^{(p)} + A_{(p+1)} \mathbf{z}[n], \tag{1}$$

with the following notations:

•

$$\mathbf{A}^{(p)} = [A_1, \dots, A_p]^\mathsf{T} \tag{2}$$

as the block column vector of parameters, where A_i is a random matrix of dimensions $(m \times m)$ with random elements of the interval type, i.e.,

$$A_i \in \mathcal{A}_i = [A_i^-, A_i^+], \qquad i = \overline{1, p}; \tag{3}$$

• $A_{(p+1)}$ as a matrix of dimensions $(m \times q)$ with random elements of the interval type, i.e.,

$$A_{(p+1)} \in \mathcal{A}_{(p+1)} = [A_{(p+1)}^{-}, A_{(p+1)}^{+}];$$
(4)

•

$$\mathbf{X}^{(n,p)} = \lfloor \mathbf{x}[n-1], \dots, \mathbf{x}[n-p] \rfloor$$
(5)

as the block row vector of p retrospective states, where $\lfloor \bullet \rfloor$ denotes a block row vector.

The probabilistic properties of the block vector $\mathbf{A}^{(p)}$ and the matrix $A_{(p+1)}$ are characterized by a joint PDF $P(\mathbf{A}^{(p)})$ and a PDF $F(A_{(p+1)})$, respectively.

The state of RDRM-O is assumed to be measurable at each time instant *n* and also to contain an additive noise $\mu[n]$:

$$\mathbf{v}[n] = \mathbf{x}[n] + \mu[n]. \tag{6}$$

The random vectors $\mu[n]$ are of the interval type, i.e.,

$$\mu[n] \in \mathcal{M}_n = [\mu_n^-, \mu_n^+],\tag{7}$$

with a PDF $M_n(\mu[n])$. The random vectors measured at different time instants are assumed to be statistically independent.

Consider the linear version of RDRM-F, which has a similar structure described by the equation

$$\mathbf{y}[n] = \mathbf{Y}^{(n,s)}\mathbf{B}^{(s)} + \boldsymbol{\zeta}[n]$$
(8)

with the following notations:

•

$$\mathbf{B}^{(s)} = |B_1, \dots, B_s|^\mathsf{T} \tag{9}$$

as a block column vector formed by matrices B_i of dimensions ($q \times q$) with random elements of the interval type, i.e.,

$$B_i \in \mathcal{B}_i = [B_i^-, B_i^+], \qquad i = \overline{1, s};$$
(10)

•

$$\mathbf{Y}^{(n,p)} = \lfloor \mathbf{y}[n-1], \dots, \mathbf{y}[n-s] \rfloor$$
(11)

as a block row vector.

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

The random vector $\zeta[n]$ is of the interval type, i.e.,

$$\zeta[n] \in \mathcal{E}_n = [\zeta_n^-, \zeta_n^+],\tag{12}$$

with a continuously differentiable PDF $Q_n(\zeta[n])$. The random vectors $\zeta[n]$ measured at different time instants are statistically independent.

By analogy with RDRM-O, the state of RDRM-F is assumed to be measurable at each time instant n and also to contain an additive noise $\xi[n]$:

$$\mathbf{z}[n] = \mathbf{y}[n] + \boldsymbol{\xi}[n]. \tag{13}$$

The random vectors $\xi[n]$ are of the interval type, i.e.,

$$\xi[n] \in \Xi_n = [\xi_n^-, \xi_n^+],\tag{14}$$

with a continuously differentiable PDF $G_n(\xi[n])$. The random vectors measured at different time instants are assumed to be statistically independent.

Thus, in the RDRM (RDRM-O and RDRM-F), the unknown characteristics are the PDFs $P(\mathbf{A}^{(p)})$, $F(A_{(p+1)})$ and $W(\mathbf{B}^{(s)})$ of the model parameters and also the PDFs $M_n(\mu[n])$, $Q_n(\zeta[n])$ and $G_n(\xi[n])$ of the measurement noises, $n \in \mathcal{L}$.

4. Models of Learning Data Sets

The desired PDFs (see the previous section) are estimated using the learning data sets that are obtained on a learning interval $n \in \mathcal{L} = [n^-, n^+]$ and are consistent with the RDRM.

Consider RDRM-O. On the learning interval,

$$\mathbf{x}[n^{-}] = \mathbf{X}^{(n^{-},p)}\mathbf{A}^{(p)} + A_{(p+1)}\mathbf{z}[n^{-}],$$

$$\mathbf{x}[n^{-}+1] = \mathbf{X}^{(n^{-}+1,p)}\mathbf{A}^{(p)} + A_{(p+1)}\mathbf{z}[n^{-}+1],$$

$$\cdots = \cdots \cdots \cdots ,$$

$$\mathbf{x}[n^{+}] = \mathbf{X}^{(n^{+},p)}\mathbf{A}^{(p)} + A_{(p+1)}\mathbf{z}[n^{+}].$$
(15)

The observable states of RDRM-O on the learning interval $\mathcal L$ represent the collection of vectors

$$\mathbf{v}[n] = \mathbf{x}[n] + \mu[n], \qquad n = \overline{n^-, n^+}.$$
(16)

Hence, the learning data set consists of the data on retrospective states of the object,

$$\hat{\mathbf{X}}^{(n^{-},p)}, \hat{\mathbf{X}}^{(n^{-}+1,p)}, \dots, \hat{\mathbf{X}}^{(n^{+},p)},$$
(17)

and the data on observable current states,

$$\hat{\mathbf{v}}[n^{-}],\ldots,\hat{\mathbf{v}}[n^{+}],\qquad \hat{\mathbf{z}}[n^{-}],\ldots,\hat{\mathbf{z}}[n^{+}].$$
(18)

Consider RDRM-F. On the learning interval,

$$\mathbf{y}[n^{-}] = \mathbf{Y}^{(n^{-},s)} \mathbf{B}^{(s)} + \zeta[n^{-}],$$

$$\mathbf{y}[n^{-}+1] = \mathbf{Y}^{(n^{-}+1,s)} \mathbf{B}^{(s)} + \zeta[n^{-}+1],$$

$$\cdots = \cdots \cdots \cdots ,$$

$$\mathbf{y}[n^{+}] = \mathbf{Y}^{(n^{+},p)} \mathbf{B}^{(s)} + \zeta[n^{+}].$$
(19)

The observable states of RDRM-F on the learning interval \mathcal{L} represent the collection of vectors

$$\mathbf{z}[n] = \mathbf{y}[n] + \boldsymbol{\xi}[n], \qquad n = \overline{n^-, n^+}.$$
(20)

Hence, the learning data set consists of the data on retrospective states of the factors,

$$\hat{\mathbf{Y}}^{(n^{-},s)}, \hat{\mathbf{Y}}^{(n^{-}+1,s)}, \dots, \hat{\mathbf{Y}}^{(n^{+},s)},$$
(21)

and the data on observable current states,

$$\hat{\mathbf{z}}[n^{-}],\ldots,\hat{\mathbf{z}}[n^{+}]. \tag{22}$$

Thus, the learning procedure of the RDRM involves three data sets, (18), (21) and (22).

5. Algorithm of Randomized Machine Learning

The entropy version [10] of RML algorithms is used for estimating the PDFs of the model parameters and measurements noises of RDRM-O and RDRM-F. For RDRM-O, the corresponding algorithm has the form

$$\mathcal{H}_{O} = -\int_{\mathcal{A}} P(\mathbf{A}^{(p)}) \ln P(\mathbf{A}^{(p)}) d\mathbf{A}^{(p)} - \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) \ln F(A_{(p+1)}) dA_{(p+1)} - \sum_{n=n^{-}}^{n^{+}} \int_{\mathcal{M}_{n}} M_{n}(\mu[n]) \ln M_{n}(\mu[n]) dM_{n}(\mu[n]) \Rightarrow \max$$
(23)

subject to the following constraints:

- the normalization conditions of the PDFs given by

$$\int_{\mathcal{A}} P(\mathbf{A}^{(p)}) d\mathbf{A}^{(p)} = 1, \quad \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) dA_{(p+1)} = 1,$$
$$\int_{\mathcal{M}_n} M_n(\mu[n]) dM_n(\mu[n]) = 1, \ n = \overline{n^-, n^+};$$

- the empirical balances given by

$$\int_{\mathcal{A}} P(\mathbf{A}^{(p)}) \hat{\mathbf{X}}^{(n,p)} \mathbf{A}^{(p)} d\mathbf{A}^{(p)} + \int_{\mathcal{A}_{(p+1)}} F(A_{(p+1)}) A_{(p+1)} \hat{\mathbf{z}}[n] dA_{(p+1)} + \int_{\mathcal{M}_n} M_n(\mu[n]) \mu[n] d\mu[n] = \hat{\mathbf{v}}[n], \quad n = \overline{n^-, n^+}.$$
(24)

Please note that the empirical balances represent a system of $(n^+ - n^-)$ blocks composed of *m* equations. With each block an *m*-dimensional vector of Lagrange multipliers $\theta^{(n)}$ is associated. This optimization problem belongs to the class of entropy-linear programming problems of the Lyapunov type [44]. It has an analytic solution parametrized by the Lagrange multipliers:

$$P^{*}(\mathbf{A}^{(p)}) = \frac{\exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, \hat{\mathbf{X}}^{(n,p)} \mathbf{A}^{(p)} \rangle\right)}{\mathcal{P}(\theta)},$$

$$F^{*}(A_{(p+1)}) = \frac{\exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, A_{(p+1)} \hat{\mathbf{z}}[n] \rangle\right)}{\mathcal{F}(\theta)},$$

$$M_{n}^{*}(\mu[n]) = \frac{\exp\left(-\langle \theta^{(n)}, \mu[n] \rangle\right)}{\mathcal{M}_{n}(\theta^{(n)})}, \quad n = \overline{n^{-}, n^{+}}.$$
(25)

In the above formulas,

$$\mathcal{P}(\theta) = \int_{\mathcal{A}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, \hat{\mathbf{X}}^{(n,p)} \mathbf{A}^{(p)} \rangle\right) d\mathbf{A}^{(p)},$$

$$\mathcal{F}(\theta) = \int_{\mathcal{A}_{(p+1)}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, A_{(p+1)} \hat{\mathbf{z}}[n] \rangle\right) dA_{(p+1)},$$

$$\mathcal{M}_{n}(\theta^{(n)}) = \int_{\mathcal{M}_{n}} \exp\left(-\langle \theta^{(n)}, \mu[n] \rangle\right) d\mu[n], \quad n = \overline{n^{-}, n^{+}}.$$
(26)

The matrix of Lagrange multipliers $\theta = [\theta^{(n^-)}, \dots, \theta^{(n^+)}]$ is determined by solving the balance equations

$$\frac{1}{\mathcal{P}(\theta)} \int_{\mathcal{A}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, \hat{\mathbf{X}}^{(n,p)} \mathbf{A}^{(p)} \rangle\right) \hat{\mathbf{X}}^{(n,p)} \mathbf{A}^{(p)} d\mathbf{A}^{(p)} + \\ + \frac{1}{\mathcal{F}(\theta)} \int_{\mathcal{A}_{(p+1)}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \theta^{(n)}, A_{(p+1)} \hat{\mathbf{z}}[n] \rangle\right) A_{(p+1)} \hat{\mathbf{z}}[n] dA_{(p+1)} + \\ + \frac{1}{\mathcal{M}_{n}(\theta^{(n)})} \int_{\mathcal{M}_{n}} \exp\left(-\langle \theta^{(n)}, \mu[n] \rangle\right) \mu[n] d\mu[n] = \hat{\mathbf{x}}[n], \qquad n = \overline{n^{-}, n^{+}}.$$
(27)

From (25)–(27) it follows that the PDFs $P^*(\mathbf{A}^{(p)})$ and $F^*(A_{(p+1)})$ of the model parameters of RDRM-O and the PDFs $M_n^*(\mu[n])$, $n = \overline{n^-, n^+}$, of the measurement noises are found using the retrospective learning data sets $\hat{\mathbf{X}}(n^-, p)$, $\hat{\mathbf{X}}(n^- + 1, p)$,..., $\hat{\mathbf{X}}(n^+, p)$, the current state data sets $\hat{\mathbf{x}}[n^-], \ldots, \hat{\mathbf{x}}[n^+]$ and the data sets $\hat{\mathbf{z}}[n^-], \ldots, \hat{\mathbf{z}}[n^+]$ generated by RDRM-F.

For obtaining the latter collections, the RML algorithm is applied to estimate the PDFs of the model parameters and measurement noises of RDRM-F. In accordance with [10],

$$\mathcal{H}_{F} = -\int_{\mathcal{B}} W(\mathbf{B}) \ln W(\mathbf{B}) d\mathbf{B} -$$

$$- \sum_{n=n^{-}}^{n^{+}} \int_{\mathcal{E}_{n}} Q_{n}(\zeta[n]) \ln Q_{n}(\zeta[n]) dQ_{n}(\zeta[n]) -$$

$$- \sum_{n=n^{-}}^{n^{+}} \int_{\Xi_{n}} G_{n}(\xi[n]) \ln G_{n}(\xi[n]) dG_{n}(\xi[n]) \Rightarrow \max$$
(28)

subject to the following constraints:

- the normalization conditions of the PDFs given by

$$\int_{\mathcal{B}} W(\mathbf{B}) d\mathbf{B} = 1,$$

$$\int_{\mathcal{E}_n} Q_n(\zeta[n]) dQ_n(\zeta[n]) = 1, \qquad \int_{\Xi_n} G_n(\zeta[n]) dG_n(\zeta[n]) = 1, \quad n = \overline{n^-, n^+};$$

- the empirical balances given by

$$\int_{\mathcal{B}} W(\mathbf{B}^{(s)}) \hat{\mathbf{Y}}^{(n,s)} \mathbf{B}^{(s)} d\mathbf{B}^{(s)} + \int_{\mathcal{E}_n} Q_n(\zeta[n]) \zeta[n] d\zeta[n] + \int_{\Xi_n} G_n(\zeta[n]) \zeta[n] d\zeta[n] = \hat{\mathbf{z}}[n], \quad n = \overline{n^-, n^+}.$$
(29)

This problem is from the same class as (25)–(27). It has the following analytic solution in terms of Lagrange multipliers:

$$W^{*}(\mathbf{B}^{(s)}) = \frac{\exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \eta^{(n)}, \hat{\mathbf{Y}}^{(n,s)} \mathbf{B}^{(s)} \rangle\right)}{\mathcal{W}(\eta)},$$

$$Q_{n}^{*}(\boldsymbol{\zeta}[n]) = \frac{\exp\left(-\langle \boldsymbol{\zeta}[n], \eta^{(n)} \rangle\right)}{\mathcal{Q}_{n}(\eta^{(n)})},$$

$$G_{n}^{*}(\boldsymbol{\xi}[n]) = \frac{\exp\left(-\langle \boldsymbol{\xi}[n], \eta^{(n)} \rangle\right)}{\mathcal{G}_{n}(\eta^{(n)})}, \quad n = \overline{n^{-}, n^{+}}.$$

$$\eta = [\eta^{(n^{-})}, \dots, \eta^{(n^{+})}].$$
(30)

In the above formulas,

$$\mathcal{W}(\eta) = \int_{\mathcal{B}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \eta^{(n)}, \hat{\mathbf{Y}}^{(n,s)} \mathbf{B}^{(s)}\right) d\mathbf{B}^{(s)}, \qquad (31)$$

$$\mathcal{Q}_{n}(\eta^{(n)}) = \int_{\mathcal{E}_{n}} \exp\left(-\langle \zeta[n], \eta^{(n)} \rangle\right) d\zeta[n], \quad \mathcal{G}_{n}(\eta^{(n)}) = \int_{\Xi_{n}} \exp\left(-\langle \zeta[n], \eta^{(n)} \rangle\right) d\zeta[n], \qquad (31)$$

$$n = \overline{n^{-}, n^{+}}.$$

The matrix of Lagrange multipliers η is determined by solving the balance equations

$$\mathcal{W}^{(-1)}(\eta) \int_{\mathcal{B}} \exp\left(-\sum_{n=n^{-}}^{n^{+}} \langle \eta^{(n)}, \hat{\mathbf{Y}}^{(n,s)} \mathbf{B}^{(s)}\right) \hat{\mathbf{Y}}^{(n,s)} \mathbf{B}^{(s)} d\mathbf{B}^{(s)} + \mathcal{Q}_{n}^{(-1)}(\eta^{(n)}) \int_{\mathcal{E}_{n}} \exp\left(-\langle \zeta[n], \eta^{(n)} \rangle\right) \zeta[n] d\zeta[n] + \mathcal{G}_{n}^{(-1)}(\eta^{(n)}) \int_{\Xi_{n}} \exp\left(-\langle \xi[n], \eta^{(n)} \rangle\right) \tilde{\xi}[n] d\xi[n] = \hat{\mathbf{z}}[n], \qquad (32)$$
$$n = \overline{n^{-}, n^{+}}.$$

6. Entropy-Randomized Forecasting of Daily Electrical Load of Power System

The daily electrical load L of a power system depends on many various factors. The analysis below is restricted to one of the most significant external factors—the ambient temperature T. The daily temperature variations are fluctuating [45,46]. These fluctuations affect electrical load, but with some time delay due the inertia of a power network supplying electrical energy from the generator to consumers.

(1). Dynamic Regression Model.

In accordance with the general structure of the RDRM (see Section 2), the electrical load model (the *LT* model) describes the dynamic relationship between electrical load and ambient temperature while the ambient temperature model (the $T\xi$ model) describes the daily dynamics of ambient temperature. There exist quite a few versions of the *LT* model, albeit all being static, i.e., describing the relationship between electrical load and ambient temperature at current time instants [47]. The daily temperature dynamics are fluctuating, and such fluctuations are described, in particular, by the periodic autoregressive model [48].

Please note that the effect of ambient temperature on electrical load is dynamic, i.e., the change in load due to temperature at a given time instant depends on its value at a previous time instant. A similar property applies to ambient temperature fluctuations.

Therefore, following the general randomized approach, the *LT* model is designed as a first-order dynamic regression model with random parameters, while the $T\xi$ model is designed as a second-order dynamic regression with a random parameter and a random input ξ . Then the *LT* ξ model is the composition of the two models above.

In the class of linear models, the randomized dynamic regression load–temperature model (*LT* model) of the first order can be written in the form

$$L[n] = aL[n-1] + bT[n],$$

$$v[n] = L[n] + \mu[n], \qquad n = \overline{n^-, n^+},$$
(33)

where random independent parameters *a* and *b* take values within intervals

$$a \in \mathcal{A} = [a^{-}, a^{+}], \ b \in \mathcal{B} = [b^{-}, b^{+}].$$
 (34)

The probabilistic properties are characterized by PDFs P(a) and F(b) defined on the sets A and B, respectively. The random noise $\mu[n]$ that simulates electrical load measurement errors is of the interval type as well. In the general case, for each time instant the intervals may have different limits, i.e.,

$$\mu[n] \in \mathcal{M}_n = [\mu^-[n]], \mu^+[n]], \tag{35}$$

with PDFs $M_n(\mu[n])$,], $n = \overline{n^-, n^+}$.

Consider the $T\xi$ model. The fluctuating character of the daily temperature variations is described by the randomized dynamic regression model of the second order

$$\tau[n] = c (2.1\tau[n-1] - 1.1\tau[n-2]),$$

$$T[n] = t + \tau[n] + \xi[n],$$
(36)

where *t* is the mean daily temperature. These parameters are random and take values within given intervals $c \in [c^-, c^+]$. The probabilistic properties of the parameters are characterized by PDFs W(c) defined on corresponding intervals.

Equation (36) contains random noises described by independent random variables $\xi[n]$; in each measurement *n*, their values may lie in different intervals, i.e.,

$$\boldsymbol{\xi}[n] \in \boldsymbol{\Xi}_n = [\boldsymbol{\xi}_n^-, \boldsymbol{\xi}_n^+]. \tag{37}$$

The probabilistic properties of the random variable $\xi[n]$ are characterized by a PDF $Q_n(\xi[n])$, $n = n^-, n^+$.

Thus, Equations (33) and (36) describing electrical load dynamics in a power system are characterized by the following PDFs:

- the *LT* model, by the PDFs P(a) and F(b) of the model parameters and the PDF $M_n(\mu[n])$, of the measurement noises, $n = \overline{n^-, n^+}$;
- the $T\xi$ model, by the PDFs W(c) of the model parameters and the PDFs $Q_n(\xi[n])$ of the measurement noises, $n = \overline{n^-, n^+}$.

(2). Learning Data Set.

For estimating the PDFs, the normalized real data from the GEFCom2014 dataset (see [34]) on daily electrical load variations $0 \le L_r^{(i)}[n] \le 1$, mean daily temperature variations $0 \le t_r^{(i)} \le 1$ and temperature deviations $0 \le \tau_r^{(i)}[n] \le 1$ from the mean daily value can be used. (Here normalization means the reduction to the unit interval.)

The normalization procedure is performed in the following way:

$$L_{r}^{(i)}[n] = \frac{\hat{L}_{r}^{(i)}[n] - \hat{L}_{min}^{(i)}}{\hat{L}_{max}^{(i)} - \hat{L}_{min}^{(i)}},$$

$$\tau_{r}^{(i)}[n] = \frac{\hat{\tau}_{r}^{(i)}[n] - \hat{\tau}_{min}^{(i)}}{\hat{\tau}_{max}^{(i)} - \hat{\tau}_{min}^{(i)}},$$

$$t_{r}^{(i)} = \frac{1}{n^{+} - n^{-}} \sum_{n=n^{-}}^{n^{+}} \tau_{r}^{(i)}[n],$$
(38)

where $\hat{L}_{min}^{(i)} = \min_{n} \hat{L}^{(i)}[n]$, $\hat{L}_{max}^{(i)} = \max_{n} \hat{L}^{(i)}[n]$, $\hat{\tau}_{min}^{(i)} = \min_{n} \hat{\tau}^{(i)}[n]$, $\hat{\tau}_{max}^{(i)} = \max_{n} \hat{\tau}^{(i)}[n]$. In accordance with (33) and (36), the model variables and the corresponding real data on the

In accordance with (33) and (36), the model variables and the corresponding real data on the learning interval $n \in T_l$ are described by the vectors

$$\mathbf{L}^{(i)}(\mathcal{T}_{l}) = \{ L^{(i)}[1], \dots, L^{(i)}[24] \}, \qquad \mathbf{L}^{(i)}_{r}(\mathcal{T}_{l}) = \{ L^{(i)}_{r}[1], \dots, L^{(i)}_{r}[24] \}, \\ \mathbf{L}^{(i)}(\mathcal{T}_{l}-1) = \{ L^{(i)}[0], \dots, L^{(i)}_{r}[23] \}, \qquad \mathbf{L}^{(i)}_{r}(\mathcal{T}_{l}-1) = \{ L^{(i)}_{r}[0], \dots, L^{(i)}_{r}[23] \}, \\ \mathbf{V}^{(i)}(\mathcal{T}_{l}) = \{ v^{(i)}[1], \dots, v^{(i)}_{r}[24] \}, \qquad \mathbf{V}^{(i)}_{r}(\mathcal{T}_{l}) = \{ v^{(i)}_{r}[1], \dots, v^{(i)}_{r}[24] \}, \\ \mathbf{T}^{(i)}(\mathcal{T}_{l}) = \{ \tau^{(i)}[1], \dots, \tau^{(i)}_{r}[24] \}, \qquad \mathbf{T}^{(i)}_{r}(\mathcal{T}_{l}) = \{ \tau^{(i)}_{r}[1], \dots, \tau^{(i)}_{r}[24] \},$$

$$\begin{split} \tilde{\mathbf{T}}^{(i)}(\mathcal{T}_{l}-1,\mathcal{T}_{l}-2) &= \{2,1\tau^{(i)}[0]-1,1\tau^{(i)}[-1],\ldots,2,1\tau^{(i)}[23]-1,1\tau^{(i)}[22]\},\\ \tilde{\mathbf{T}}^{(i)}_{r}(\mathcal{T}_{l}-1,\mathcal{T}_{l}-2) &= \{2,1\tau^{(i)}_{r}[0]-1,1\tau^{(i)}_{r}[-1],\ldots,2,1\tau^{(i)}_{r}[23]-1,1\tau^{(i)}_{r}[22]\},\\ \mu^{(i)}(\mathcal{T}_{l}) &= \{\mu^{(i)}[1],\ldots,\mu^{(i)}[24]\}, \quad \xi^{(i)}(\mathcal{T}_{l}) &= \{\xi^{(i)}[1],\ldots,\xi^{(i)}[24]\}. \end{split}$$

In terms of (39), the *LT* and $T\xi$ models on the learning interval T_l have the form

$$\mathbf{L}^{(i)}(\mathcal{T}_{l}) = a\mathbf{L}^{(i)}(\mathcal{T}_{l} - 1) + b\mathbf{T}^{(i)}(\mathcal{T}_{l}),
\mathbf{V}^{(i)}(\mathcal{T}_{l}) = \mathbf{L}^{(i)}(\mathcal{T}_{l}) + \mu^{(i)}(\mathcal{T}_{l}),
\mathbf{T}^{(i)}(\mathcal{T}_{l}) = c\,\tilde{\mathbf{T}}^{(i)}(\mathcal{T}_{l} - 1, \mathcal{T}_{l} - 2),
\mathbf{T}^{(i)}(\mathcal{T}_{l}) = t + \tilde{\mathbf{T}}^{(i)}(\mathcal{T}_{l}) + \xi^{(i)}(\mathcal{T}_{l}).$$
(39)

The random parameters take values within the intervals

$$\mathcal{A} = [0.05, 0.15], \quad \mathcal{B} = [0.5, 1.0], \quad \mathcal{C} = [0.75, 0.85].$$
 (40)

The measurement noises take values within the intervals

$$\mathcal{M}_n = [-0.1, 0.1], \quad \Xi_n = [-0.1, 0.1].$$
 (41)

(3). Entropy-Optimal Probability Density Functions of Parameters and Noises.

In accordance with the approach described in Section 5, for the *LT* model (33)–(35) the PDFs parametrized by the Lagrange multipliers $\theta^{(i)} = \{\theta_1^{(i)}, \dots, \theta_{24}^{(i)}\}$ have the form

$$P_{i}^{*}(a,\theta^{(i)}) = \frac{l_{r}^{(i)}(\theta)\exp\left(-al_{r}^{(i)}(\theta)\right)}{\exp(-a-l_{r}^{(i)}(\theta)) - \exp(-a+l_{r}^{(i)}(\theta))},$$

$$F_{i}^{*}(b,\theta^{(i)}) = \frac{h_{r}^{(i)}(\theta)\exp\left(-bh_{r}^{(i)}(\theta)\right)}{\exp(-b-h_{r}^{(i)}(\theta)) - \exp(-b+h_{r}^{(i)}(\theta))},$$

$$M_{i,n}^{*}(\mu[n]) = \frac{\theta_{n}^{(i)}\exp\left(-\theta_{n}^{(i)}\mu[n]\right)}{\exp(-\mu^{-}[n]\theta_{n}^{(i)}) - \exp(-\mu^{+}[n]\theta_{n}^{(i)})}, \quad n = \overline{1, 24},$$
(42)

where

$$l_r^{(i)}(\theta) = \sum_{n=1}^{24} \theta_n L_r^{(i)}[n-1], \quad h_r^{(i)}(\theta) = \sum_{n=1}^{24} \theta_n T_r^{(i)}[n].$$
(43)

The Lagrange multipliers $\theta^{(i)}$ are calculated by solving the system of balance equations

$$L^{(i)}(\theta^{(i)}) + T^{(i)}(\theta^{(i)}) + M_n^{(i)}(\theta_n^{(i)}) = L_r^{(i)}[n], \qquad n = \overline{1, 24},$$
(44)

where

$$L^{(i)}(\theta^{(i)}) = \frac{\exp(-a^{-}l_{r}^{(i)}(\theta^{(i)}))(a^{-}l^{(i)}(\theta^{(i)}) + 1) - \exp(-a^{+}l_{r}^{(i)}(\theta^{(i)}))(a^{+}l^{(i)}(\theta^{(i)}) + 1)}{\exp(-a^{-}l^{(i)}(\theta^{(i)})) - \exp(-a^{+}l^{(i)}(\theta^{(i)}))},$$

$$T^{(i)}(\theta^{(i)}) = \frac{\exp(-b^{-}h_{r}^{(i)}(\theta^{(i)}))(b^{-}h^{(i)}(\theta^{(i)}) + 1) - \exp(-b^{+}h_{r}^{(i)}(\theta^{(i)}))(b^{+}h^{(i)}(\theta^{(i)}) + 1)}{\exp(-b^{-}h^{(i)}(\theta^{(i)})) - \exp(-b^{+}h^{(i)}(\theta^{(i)}))},$$

$$M_{n}^{(i)}(\theta_{n}^{(i)}) = \frac{\exp(-\mu^{-}[n]\theta_{n}^{(i)})(\mu^{-}[n]\theta_{n}^{(i)}) + 1) - \exp(-\mu^{+}[n]\theta_{n}^{(i)})(\mu^{+}[n]\theta_{n}^{(i)}) + 1)}{\theta_{n}^{(i)}\left(\exp(-\mu^{-}[n]\theta_{n}^{(i)}) - \exp(-\mu^{+}[n]\theta_{n}^{(i)})\right)}.$$
(45)

Consider the $T\xi$ model. The corresponding entropy-optimal PDFs parametrized by the Lagrange multipliers have the form

$$W_{i}^{*}(c,\eta^{(i)}) = \frac{\tilde{h}_{r}^{(i)}(\eta)\exp\left(-c\tilde{h}_{r}^{(i)}(\eta)\right)}{\exp(-a^{-}\tilde{h}_{r}^{(i)}(\eta)) - \exp(-a^{+}\tilde{h}_{r}^{(i)}(\eta))},$$
$$Q_{i,n}^{*}(\xi[n]) = \frac{\eta_{n}^{(i)}\exp\left(-\eta_{n}^{(i)}\xi[n]\right)}{\exp(-\xi^{-}[n]\eta_{n}^{(i)}) - \exp(-\xi^{+}[n]\eta_{n}^{(i)})}, \quad n = \overline{1, 24},$$

where

$$\tilde{h}_{r}^{(i)}(\eta) = \sum_{n=1}^{24} \eta_{n}(2, 1T_{r}^{(i)}[n-1] - 1, 1T_{r}^{(i)}[n-2]), \quad q^{(i)}(\eta^{(i)}) = \sum_{n=1}^{24} \eta_{n}^{(i)}.$$
(46)

The Lagrange multipliers $\eta^{(i)}$ are calculated by solving the system of balance equations

$$D^{(i)}(\eta^{(i)}) + N^{(i)}(\eta^{(i)}) + K_n^{(i)}(\eta_n^{(i)}) = T_r^{(i)}[n], \qquad n = \overline{1, 24},$$
(47)

where

$$D^{(i)}(\eta^{(i)}) = \frac{\exp(-t^{-}q^{(i)}(\eta^{(i)}))(t^{-}q^{(i)}(\eta^{(i)})+1) - \exp(-t^{+}q^{(i)}(\eta^{(i)}))(t^{+}q^{(i)}(\eta^{(i)})+1)}{\exp(-t^{-}q^{(i)}(\eta^{(i)})) - \exp(-t^{+}q^{(i)}(\eta^{(i)}))},$$

$$N^{(i)}(\eta^{(i)}) = \frac{\exp(-c^{-}\tilde{h}_{r}^{(i)}(\eta^{(i)}))(c^{-}\tilde{h}^{(i)}(\eta^{(i)})+1) - \exp(-c^{+}\tilde{h}_{r}^{(i)}(\eta^{(i)}))(c^{+}\tilde{h}^{(i)}(\eta^{(i)})+1)}{\exp(-c^{-}\tilde{h}^{(i)}(\eta^{(i)})) - \exp(-c^{+}\tilde{h}^{(i)}(\eta^{(i)}))},$$

$$K_{n}^{(i)}(\eta^{(i)}) = \frac{\exp(-\xi^{-}[n]\eta_{n}^{(i)})(\xi^{-}[n]\eta_{n}^{(i)})+1) - \exp(-\xi^{+}[n]\eta_{n}^{(i)})(\xi^{+}[n]\eta_{n}^{(i)})+1)}{\eta_{n}^{(i)}\left(\exp(-\xi^{-}[n]\eta_{n}^{(i)}) - \exp(-\xi^{+}[n]\eta_{n}^{(i)})\right)}.$$
(48)

(4). Results of Model Learning.

Using the available data on daily variations of electrical load and ambient temperature (see Figure 1) for the three days indicated above, the balance Equations (44), (45), (47) and (48) were formed. Their solution was determined by minimizing the quadratic residual between the left- and right-hand sides of the equations. Since the equations are significantly nonlinear, the resulting values of the Lagrange multipliers (see Table 1) correspond to a local minimum of the residual. All calculations were implemented in MATLAB; optimization was performed using the fsolve function.

Time Instants	$ heta^{(1)}$	$ heta^{(2)}$	$ heta^{(3)}$	$\eta^{(1)}$	$\eta^{(2)}$	$\eta^{(3)}$
1	-29.72	7009.28	1038.07	14.63	21.22	17.34
2	1.58	230.89	35.35	19.52	26.71	28.32
3	-4.09	369.96	26.23	35.91	31.60	26.33
4	-4.68	29.93	11.96	55.83	127.82	52.08
5	-7.21	24.25	1.03	96.85	642.35	110.94
6	-9.26	13.72	-15.76	592.99	7009.28	4729.52
7	-59.09	-5.96	-7009.28	7009.28	183.92	7009.28
8	-7009.28	-33.99	-767.99	48.21	39.94	23.16
9	-766.00	-1409.28	-22.91	66.58	12.28	-1.26
10	-50.90	-4229.90	-4.27	37.38	2.35	-19.78
11	-18.97	-45.22	3.72	22.51	-8.82	-22.73
12	-11.42	-15.07	9.17	7.16	-27.06	-23.06
13	-13.94	2.59	17.38	5.72	-172.25	-27.06
14	-17.62	5.82	14.94	2.83	-65.29	-23.02
15	-18.18	9.33	17.74	-0.30	-57.45	-23.15
16	-27.28	11.35	21.85	-1.24	-482.69	-47.78
17	-49.55	4.50	22.68	-5.49	-889.02	-130.49
18	-25.41	-7.09	29.39	-0.89	-28.12	-60.71
19	-8.20	-4.66	98.03	-4.23	-14.20	-270.17
20	0.95	-4.89	52.27	3.70	-6.41	-31.47
21	1.01	-16.37	8.15	21.16	2.48	-1.23
22	22.00	-8.24	7.45	17.85	12.15	14.86
23	2881.43	17.00	902.73	26.78	9.98	26.39
24	512.14	36.30	355.47	27.32	24.65	121.44
$l_r^{(i)}(heta^*)$	9.71	6.04	6.09			
$h_r^{(i)}(heta^*)$	0.06	0.58	0.81			
$ ilde{h}^{(i)}_r(\eta^*)$				0.41	0.05	0.19

Table 1. Lagrange multipliers θ , η .

Because the parameters of the *LT* model are independent, the joint PDFs $U_i^*(a, b) = P_i^*(a)F_i^*(b)$ of the parameters and noises have the form

$$U_{1}^{*}(a,b) = 53.09 \exp(-9.72a) \exp(0.06b),$$

$$U_{2}^{*}(a,b) = 55.49 \exp(-6.04a) \exp(0.58b),$$

$$U_{3}^{*}(a,b) = 65.81 \exp(-6.09a) \exp(-0.81b),$$

$$(a,b) \in [0.05, 0.15] \bigcup [0.5, 1.0], \quad \mu \in [-0.1, 0.1], \quad i = \overline{1, 24}.$$
(49)

Clearly, the PDFs are of exponential type. For i = 1, the graphs are shown in Figure 2. For the $T\xi$ model, the PDFs of the parameters and noises have the form

$$\begin{split} W_1^*(c) &= 13.90 \exp(-0.41c), \\ W_2^*(c) &= 10.43 \exp(-0.05c), \\ W_3^*(c) &= 11.65 \exp(-0.19c), \end{split}$$

$$c \in [0.75, 0.85], \quad \xi \in [-0.1, 0.1], \quad i = \overline{1, 24}.$$

For i = 1, the graphs can be seen in Figure 3.

Thus, the randomized $LT\xi$ model generates random trajectories with the entropy-optimal PDFs of the model parameters and measurement noises:

$$L[n] = aL[n-1] + bT[n], \quad (P^*(a), F^*(b)),$$

$$v[n] = L[n] + \mu[n], \quad M^*_n(\mu[n]);$$

$$\tau[n] = c (2.1\tau[n-1] - 1.1\tau[n-2]), \quad W^*(c), \ i = \overline{1,3},$$

$$T[n] = t + \tau[n] + \xi[n], \quad Q^*_n(\xi[n]).$$
(50)

The corresponding ensembles are generated by the sampling procedure of the resulting PDFs of the parameters and noises using the acceptance-rejection (AR) method (also known as rejection sampling (RS); see [49]). During calculations, 100 samples for each parameter and 100 samples for each noise were used; in other words, the ensemble consisted of 10^4 trajectories.



Figure 2. PDFs of the parameters and noises for i = 1.



Figure 3. PDFs of the parameters and noises for i = 1.

(5). Model Testing.

The adequacy of the model was analyzed by the self- and cross-testing of the *LT* and $T\xi$ models on the real load–temperature data for 3–5 July 2016 (i = 1, 2, 3). Self-testing means generating an ensemble of trajectories with the entropy-optimal parameters and noises for day i, calculating the mean (*mean*) and median (*med*) trajectories and also the variance curve ($std\pm$) of the ensemble, and comparing the mean trajectory with the real counterparts by electrical load and ambient temperature *for the same day i*. The quality of approximation is characterized by relative errors,

$$\delta_L^{(i)} = \frac{\sum\limits_{n=1}^{24} \left(L_{mean}^{(i)}[n] - L_r^{(i)}[n] \right)^2}{\sum\limits_{n=1}^{24} (L_{mean}^{(i)}[n])^2 + \sum\limits_{n=1}^{24} (L_r^{(i)}[n])^2}, \quad i = \overline{1, 3},$$
(51)

in electrical load and

$$\delta_T^{(i)} = \frac{\sum\limits_{n=1}^{24} \left(T_{mean}^{(i)}[n] - T_r^{(i)}[n] \right)^2}{\sum\limits_{n=1}^{24} (T_{mean}^{(i)}[n])^2 + \sum\limits_{n=1}^{24} (T_r^{(i)}[n])^2}, \quad i = \overline{1, 3},$$
(52)

in ambient temperature.

Cross-testing represents a similar procedure in which the mean trajectories are compared with the real counterparts in terms of electrical load and ambient temperature for days $j \neq i$. The quality of approximation is characterized by relative errors,

$$\delta_{L}^{(i,j)} = \frac{\sum_{n=1}^{24} \left(L_{mean}^{(i)}[n] - L_{r}^{(j)}[n] \right)^{2}}{\sum_{n=1}^{24} \left(L_{mean}^{(i)}[n] \right)^{2} + \sum_{n=1}^{24} \left(L_{r}^{(j)}[n] \right)^{2}}, \quad i = \overline{1,3}, \ i \neq j,$$
(53)

in electrical load and

$$\delta_T^{(i,j)} = \frac{\sum\limits_{n=1}^{24} \left(T_{mean}^{(i)}[n] - T_r^{(j)}[n] \right)^2}{\sum\limits_{n=1}^{24} (T_{mean}^{(i)}[n])^2 + \sum\limits_{n=1}^{24} (T_r^{(j)}[n])^2}, \quad i = \overline{1,3}, i \neq j,$$
(54)

in ambient temperature.

Self-testing. For the *LT* model, the real ambient temperature data $T_r^{(i)}[n]$ as well as the entropy-optimal PDFs $P_i^*(a)$ and $F_i^*(b)$ of the parameters (a, b) and the PDFs $M_1^*(\mu[1]), \ldots, M_{24}^*(\mu[24])$ of the measurement noises $\mu[n]$ were used. The ensembles $\mathbb{L}^{(i)}$ were generated using the sampling procedure of the above PDFs. The mean trajectory $L_{mean}^{(i)}[n]$, the median trajectory $L_{med}^{(i)}[n]$ and also the trajectories $L_{std\pm}^{(i)}[n]$ corresponding to the limits of the variance graph were found. The errors $\delta_L^{(i)}$ were calculated. The resulting ensembles and relative errors $\delta_L^{(i)}$ for the three indicated days are demonstrated in Figure 4.

The $T\xi$ model was tested by generating the ensemble $\mathbb{T}^{(i)}$ of random trajectories $T^{(i)}[n]$, $n = \overline{1,24}$ with the entropy-optimal PDFs $W^{(i)}(c)$ and $Q_1^*(\xi[1]), \ldots, Q_{24}^*(\xi[24])$ through sampling. The mean trajectory $T_{mean}^{(i)}[n]$, the median trajectory $T_{med}^{(i)}[n]$ and the trajectory $T_{std\pm}^{(i)}[n]$ corresponding to the limits of the variance curve are calculated. The resulting ensembles and relative errors $\delta_T^{(i)}$ for the three days are shown in Figure 5.

Cross-testing. For cross-testing, the *LT* and *LT* ξ models learned on the data for day *i* were used, and their mean trajectories were compared with the data for days $j \neq i$. The resulting errors are combined in Tables 2–4.

Table 2. Values δ_L obtained by cross-testing of *LT* model. Mean value $\delta_L = 0.0530$.

i/j	1	2	3
1		0.0495	0.1052
2	0.0858		0.1428
3	0.0569	0.0364	

Table 3. Values δ_T obtained by cross-testing of $T\xi$ model. Mean value $\delta_L = 0.0757$.

i/j	1	2	3
1		0.1051	0.1079
2	0.1506		0.1185
3	0.1315	0.0676	

Table 4. Values δ_T obtained by cross-testing of $LT\xi$ model. Mean value $\delta_L = 0.1478$.

i/j	1	2	3
1		0.1437	0.2659
2	0.1756		0.2322
3	0.3475	0.1655	











Figure 4. Ensembles of *LT* model.



Figure 5. Ensembles of $T\xi$ model.

(6). Randomized Prediction of N-Daily Load.

In the randomized prediction of the *N*-daily load, the $LT\xi$ model learned on the interval T_l was used. The quality of the forecast was characterized using the $LT\xi$ model with the entropy-optimal PDFs obtained on the real data for the *first* (*i* = 1) day.

The 1-day ($n \in [1, 24]$), 2-day ($n \in [1, 48]$) and 3-day ($n \in [1, 72]$) ensembles were constructed by the sampling procedure of the above PDFs. For these ensembles, the mean trajectories $L_{mean}[n]$, the median trajectories $L_{med}[n]$ and also the limiting trajectories $L_{std\pm}[n]$ of the variance curve were found. The forecast results were compared with the real data for 3–7 July 2006 ($i = \overline{1,4}$). The forecasting quality was characterized by the relative errors calculated similar to (53) and (54).

The resulting 24-h, 48-h and 72-h randomized forecasts of electrical load and their probabilistic characteristics (the mean and median trajectories, the limit trajectories of the variance curves) are presented in Figure 6. The errors, i.e., the deviations between the model forecasts and real data, can be seen in Table 5.

Table 5. Accuracy of 24-h, 48-h and 72-h forecasts using $LT\xi$ model.

$\delta_L^{(2)}$	$\delta_L^{(3)}$	$\delta_L^{(4)}$
0.1509	0.2515	0.2133



Figure 6. 24-h, 48-h and 72-h forecasts using $LT\xi$ model.

7. Conclusions

The article proposes a new forecasting approach based on the idea of generation not the only forecast and not a set of forecasts with the scenario's model parameters, and not forecasts with assigned probabilities, but the ensemble of random forecasts with entropy-optimal model parameters and measurement noises.

For the randomized forecasting we propose a structure of predictive dynamical model which uses as real data as optimized noises. The latter are the source of ensemble of predictive trajectories, which allow computing deterministic trajectories of its different numerical characteristics and probabilistic estimates.

Author Contributions: Conceptualization, Y.S.P.; Data curation, A.Y.P.; Methodology, Y.S.P., A.Y.P., Y.A.D. and D.S.; Software, A.Y.P. and Y.A.D.; Supervision, D.S.; Writing—original draft, Y.S.P., A.Y.P., Y.A.D. and D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially supported by Russian Foundation for Basic Research (project Nos. 20-07-00223, 20-07-00683, 20-07-00470).

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

References

- Frawley, W.J.; Piatetsky-Shapiro, G.; Matheus, C.J. Knowledge discovery in databases: An overview. *AI Mag.* 1992, 13, 57.
- 2. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2005.
- 3. Campbell, P. Editorial on special issue on big data: Community cleverness required. Nature 2008, 455, 1.
- 4. Dhar, V. Data Science and Prediction. Commun. ACM 2013, 56, 64–73. [CrossRef]
- 5. Rosenblatt, F. *The Perceptron, a Perceiving and Recognizing Automaton Project Para;* Cornell Aeronautical Laboratory: Buffalo, NY, USA, 1957.
- 6. Tsypkin, Y.Z. Osnovi Teorii Obuchaiuschichsia Sistem (Foundations of the Theory of Learning Systems); Nauka: Moscow, Russia, 1970.
- 7. Ayzerman, M.A.; Braverman, E.M.; Rozonoer, L.I. *Metod Potencialnikh Funkcii v Teorii Obuchenia Mashin* (*Method of Potential Functions in the Theory of Machine Learning*); Nauka: Moscow, Russia, 1970.
- 8. Vapnik, V.N. Statistical Learning Theory; Wiley: New York, NY, USA, 1998.
- 9. Bishop, C. *Pattern Recognition and Machine Learning (Information Science and Statistics),* 1st ed.; Springer: New York, NY, USA, 2006; Reprint in 2007.
- Popkov, Y.S.; Popkov, A.Y.; Dubnov, Y.A. Randomizirovannoe Mashinnoe Obichenie: Ot Empiricheskoi Veroiatnosti k Entropiinoi Randomizacii (Randomized Machine Learning: From Empirical Probability to Entropy Randomization); LENAND: Moscow, Russia, 2019.

- 11. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: Berlin, Germany, 2001; Volume 1.
- 12. Aivazyan, S.A.; Mhitaryan, V.S. Prikladnaia Statistika i Osnovi Econometriki (Applied Statistics and Basics of Econometrics); Unity, Moscow, 1998.
- 13. Tarassow, A. Forecasting U.S. money growth using economic uncertainty measures and regularisation techniques. *Int. J. Forecast.* **2019**, *35*, 443–457. [CrossRef]
- 14. Marcellino, M.; Stock, J.H.; Watson, M.W. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *J. Econ.* **2006**, *135*, 499–526. [CrossRef]
- 15. Eitrheim, Ø.,; Teräsvirta, T. Testing the adequacy of smooth transition autoregressive models. *J. Econ.* **1996**, 74, 59–75. [CrossRef]
- 16. Molodtsova, T.; Papell, D. Out-of-sample exchange rate predictability with Taylor rule fundamentals. *J. Int. Econ.* **2009**, *77*, 167–180. [CrossRef]
- 17. Granger, C.; Teräsvirta, T. *Modelling Non-Linear Economic Relationships*; Oxford University Press, Oxford, UK, 1993.
- 18. Wang, R.; Morley, B.; Stamatogiannis, M.P. Forecasting the exchange rate using nonlinear Taylor rule based models. *Int. J. Forecast.* **2019**, *35*, 429–442. [CrossRef]
- 19. Bessec, M.; Fouquau, J. Short-run electricity load forecasting with combinations of stationary wavelet transforms. *Eur. J. Oper. Res.* **2018**, *264*, 149–164. [CrossRef]
- 20. Clements, A.E.; Hurn, A.; Li, Z. Forecasting day-ahead electricity load using a multiple equation time series approach. *Eur. J. Oper. Res.* 2016, 251, 522–530. [CrossRef]
- 21. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
- 22. Wheatcroft, E. Interpreting the skill score form of forecast performance metrics. *Int. J. Forecast.* 2019, 35, 573–579. [CrossRef]
- 23. Canale, A.; Ruggiero, M. Bayesian nonparametric forecasting of monotonic functional time series. *Electron. J. Stat.* **2016**, *10*, 3265–3286. [CrossRef]
- 24. Dubnov, Y.; Boulytchev, A.V. Bayesian Identification of a Gaussian Mixture Model. *Inform. Tekhn. Vychisl. Sist.* **2017**, *1*, 101–114.
- 25. Frazier, D.T.; Maneesoonthorn, W.; Martin, G.M.; McCabe, B.P. Approximate bayesian forecasting. *Int. J. Forecast.* **2019**, *35*, 521–539. [CrossRef]
- 26. Beaumont, M.A.; Zhang, W.; Balding, D.J. Approximate Bayesian computation in population genetics. *Genetics* **2002**, *162*, 2025–2035.
- 27. McAdam, P.; Warne, A. Euro area real-time density forecasting with financial or labor market frictions. *Int. J. Forecast.* **2019**, *35*, 580–600. [CrossRef]
- 28. Alkema, L.; Gerland, P.; Raftery, A.; Wilmoth, J. The United Nations probabilistic population projections: An introduction to demographic forecasting with uncertainty. *Foresight (Colch. VT)* **2015**, 2015, 19.
- 29. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3. [CrossRef]
- Bröcker, J.; Smith, L.A. From ensemble forecasts to predictive distribution functions. *Tellus A Dyn. Meteorol. Oceanogr.* 2008, 60, 663–678. [CrossRef]
- 31. Christensen, H.; Moroz, I.; Palmer, T. Evaluation of ensemble forecast uncertainty using a new proper score: Application to medium-range and seasonal forecasts. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 538–549. [CrossRef]
- 32. Gneiting, T.; Katzfuss, M. Probabilistic forecasting. Annu. Rev. Stat. Appl. 2014, 1, 125–151. [CrossRef]
- 33. Lahiri, K.; Wang, J.G. Evaluating probability forecasts for GDP declines using alternative methodologies. *Int. J. Forecast.* **2013**, *29*, 175–190. [CrossRef]
- 34. Hong, T.; Pinson, P.; Fan, S.; Zareipour, H.; Troccoli, A.; Hyndman, R.J. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *Int. J. Forecast.* **2016**, *32*, 896–913. [CrossRef]
- 35. Vidyasagar, M. Randomized Algorithms for Robust Controller Synthesis Using Statistical Learning Theory: A Tutorial Overview. *Eur. J. Control* **2001**, *7*, 287–310. [CrossRef]
- 36. Granichin, O.N.; Polyak, B.T. Randomizirovannie Algoritmi Ocenivania i Optimizacii pri Pochti Proizvolnikh Pomekhakh (Randomized Algorithms of Estimation and Optimization under Almost Arbitrary Disturbances); Nauka: Moscow, Russia, 2003.

- 37. Biondo, A.E.; Pluchino, A.; Rapisarda, A.; Helbing, D. Are random trading strategies more successful than technical ones? *PLoS ONE* **2013**, *8*, e68344. [CrossRef] [PubMed]
- 38. Lutz, W.; Sanderson, W.; Scherbov, S. The end of world population growth. Nature 2001, 412, 543. [CrossRef]
- 39. Tsirlin, A.M. Metody Usrednennoi Optimizatsii i Ikh Primenenie (Average Optimization Methods and Their Application); Fizmatlit: Moscow, Russia, 1997.
- 40. Shannon, C.E. Communication theory of secrecy systems. Bell Labs Tech. J. 1949, 28, 656–715. [CrossRef]
- 41. Jaynes, E.T. Information theory and statistical mechanics. Phys. Rev. 1957, 106, 620-630. [CrossRef]
- 42. Rosenkrantz, R.D.; Jaynes, E.T. *Papers on Probability, Statistics, and Statistical Physics*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
- 43. Jaynes, E.T. Probability Theory: The Logic of Science; Cambridge University Press: Cambridge, UK, 2003.
- 44. Joffe, A.D.; Tihomirov, A.M. Teoriya Ekstremalnykh Zadach (Theory of Extreme Problems); Nauka: Moscow, Russia, 1974.
- 45. Wang, P.; Liu, B.; Hong, T. Electric load forecasting with recency effect: A big data approach. *Int. J. Forecast.* **2016**, *32*, 585–597. [CrossRef]
- 46. Gaillard, P.; Goude, Y.; Nedellec, R. Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *Int. J. Forecast.* **2016**, *32*, 1038–1050. [CrossRef]
- 47. Fiedner, G. Hierarchical Forecasting: Issues and Use Guidelines. *Ind. Manag. Data Syst.* 2001, 101, 5–12. [CrossRef]
- 48. Amaral, L.F.; Souza, R.C.; Stevenson, M. A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting. *Int. J. Forecast.* **2008**, *24*, 603–615. [CrossRef]
- 49. Von Neumann, J. 13. Various Techniques Used in Connection With Random Digits. *Appl. Math. Ser.* **1951**, 12, 36–38.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).