Robust Deep Gaussian Process-based Probabilistic Electrical Load Forecasting against Anomalous Events

Cao, Di; Zhao, Junbo; Hu, Weihao; Zhang, Yingchen; Liao, Qishu; Chen, Zhe; Blaabjerg, Frede

Published in:
I E E E Transactions on Industrial Informatics

DOI (link to publication from Publisher):
10.1109/TII.2021.3081531

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Robust Deep Gaussian Process-based Probabilistic Electrical Load Forecasting against Anomalous Events

Di Cao, Junbo Zhao, *Senior Member, IEEE*, Weihao Hu, *Senior Member, IEEE*, Yingchen Zhang, *Senior Member, IEEE*, Qishu Liao, Zhe Chen, *Fellow, IEEE*, Frede Blaabjerg, *Fellow, IEEE*

***Abstract*—The abnormal events, such as the unprecedented COVID-19 pandemic, can significantly change the load behaviors, leading to huge challenges for traditional short-term forecasting methods. This paper proposes a robust deep Gaussian processes (DGP)-based probabilistic load forecasting method using a limited number of data. Since the proposed method only requires a limited number of training samples for load forecasting, it allows us to deal with extreme scenarios that cause short-term load behavior changes. In particular, the load forecasting at the beginning of abnormal event is cast as a regression problem with limited training samples and solved by double stochastic variational inference DGP. The mobility data are also utilized to deal with the uncertainties and pattern changes and enhance the flexibility of the forecasting model. The proposed method can quantify the uncertainties of load forecasting outcomes, which would be essential under uncertain inputs. Extensive comparison results with other state-of-the-art point and probabilistic forecasting methods show that our proposed approach can achieve high forecasting accuracies with only a limited number of data while maintaining excellent performance of capturing the forecasting uncertainties.**

***Index Terms*—Probabilistic load forecasting, limited data, anomalous events, deep Gaussian process regression, uncertainty quantification.**

I. Motivations and Contributions

The abnormal events, such as the ongoing coronavirus disease 2019 (COVID-19) pandemic, can lead to a profound influence on power system operations [1]. It has been found that there are significant changes in electrical load consumption behaviors across the world during the pandemic [2]. Take the data of Northern Italy after the announcement of second phase containment measures for example, the electrical load demand of this zone during the third week of March 2019 and 2020 are displayed in Fig. 1 [3]. It can be observed that the load demand experienced significant drops during the pandemic compared with that in the same period of 2019. Moreover, the consumption pattern also changes. The changes in both magnitude and consumption patterns bring great challenges for the forecasting of load demand. During this week, the mean absolute percentage error (MAPE) and maximum MAPE of day-ahead load forecasting provided by the system operator increase by 61% and 40% compared with those in the same period in 2019. Although the base load decreases sharply, the mean absolute error (MAE) of day-ahead prediction increases from 329.5 MW to 428.2 MW, increasing by 30% as compared to that in 2019. Moreover, the MAPE obtained by the generalized adaptive models utilized by the main French electricity operator during the first few weeks of

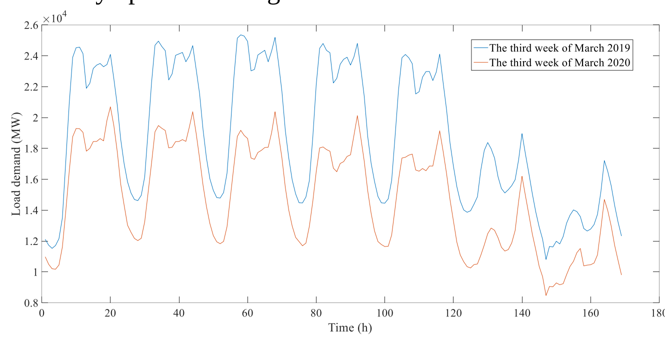


Fig. 1. Load patterns of Northern Italy during the third week of March in 2019 and 2020.

lockdown is five times of that achieved under normal conditions, demonstrating the challenges for load forecasting caused by the pandemic [4]. The sharp decrease of forecasting accuracy is caused by two main reasons: 1) the short-term load prediction models typically depend on the relatively long-term patterns, and therefore they have large inertia against unprecedented change in peak demand and consumption pattern in a short term. The existing forecasting models are not flexible enough to adapt to these scenarios, 2) the dramatic changes happened in a relatively short time and no similar event has been ever observed in history. As a result, the volume of the recorded data is not enough to train an accurate forecasting model.

The sharp decrease load forecasting accuracy during abnormal events may lead to the technical risks for the operation of power systems, the balance of which is maintained through the multi-stage power generation dispatches. Since the aim of the power generation dispatch is to meet the load demand, good forecasting results are required by the operator. For example, the load forecasting of 15-39 hours ahead is a basis for the scheduling of day-ahead generation. The occurrence of anomalous events may significantly decrease the accuracy of current electrical load forecasting tools, which brings great challenges for the balance of power system. This is even aggravated by the increasing proportion of intermittent

This work was supported by the National Key Research and Development Program of China (2018YFE0127600). Corresponding author: Weihao Hu.

Di Cao, Weihao Hu, Qishu Liao are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China (e-mail: caodi@std.uestc.edu.cn; whu@uestc.edu.cn; liaoqishu@std.uestc.edu.cn)

J. Zhao is with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762 (e-mail: junbo@ece.msstate.edu).

Y. Zhang is with National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: yingchen.zhang@nrel.gov).

Zhe Chen, Frede Blaabjerg are with the Department of Energy Technology, Aalborg University, Aalborg, Denmark (e-mail: zch@et.aau.dk; fbl@et.aau.dk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

renewable generations due to their stochasticity [5]. In this context, more flexible margins and accurate probabilistic analysis are required to hedge against the uncertainty. This calls for a probabilistic load forecasting method to make effective inductive reasoning under anomalous conditions with limited data.

Load forecasting has been widely investigated in the literature and various approaches have been proposed. However, previous works focus mainly on the load forecasting during the normal operation period and there are rare studies to deal with extreme events, such as the COVID-19. The statistical analysis techniques face huge challenges of providing accurate forecasting results when dealing with the complex nonlinear system. By contrast, the parametric methods like deep neural networks (DNN) require a large amount of data to find the precise patterns among samples [6]. Note that these methods are typically utilized for point forecasting, which cannot capture the uncertainties caused by the pandemic.

Gaussian process (GP) regression, a non-parametric method based on the Bayesian theory, has advantages over parameterized methods in regression analysis with a limited number of training samples [7]. Furthermore, GP allows us to quantify the uncertainty of the forecasting values, which is very helpful for secure system planning and operation considering uncertainties. However, the selection of the covariance function of GP is cumbersome, requiring both hands-on tuning and a deep understanding of the data. This may not be possible if there are no prior knowledge of the specific problem to be solved. Note that the multilayer hierarchy models, such as DNN have structural advantages in modeling complex dynamic functional features by stacking multiple layers of NNs and sequentially warping the latent variables. A hierarchical generalization of GP leads to the development of deep GP. Specifically, by hierarchically warping and stretching the input space, a covariance function with "self-tuning" capability that can fit data without much human intervention can be formulated [8]. DGP integrates the flexibility of deep structure models and the uncertainty quantification and effective inductive reasoning ability of the GP method [9].

This paper aims to develop DGP for the probabilistic load prediction that can deal with anomalous events with limited data. The main contributions are:

1) The load prediction considering extreme events is formulated as a regression problem with limited training samples. It is further solved by the enhanced DGP method, i.e., double stochastic variational inference DGP considering the impacts of mobility data. The proposed approach can achieve effective inductive reasoning based on only a limited number of recorded data and thus can quickly adapt to the unprecedented changes. The utilization of mobility data further enhances its ability to address uncertainties.
2) The proposed method can effectively quantify load forecasting uncertainties caused by multiple uncertainties during the anomalous events. This allows system planners and operators to make uncertainty-aware decisions. It distinguishes from typical point forecasting methods that may cause large deviations when facing high uncertainties.
3) Comprehensive comparisons with other benchmark machine learning methods have been carried out using a series of datasets at both the city and country level. It shows that our proposed approach has a better capability of dealing with anomalous events with a limited number of data while being able to quantify the uncertainties of forecasts.

The remainder of this article is organized as follows: In Section II, a literature review is provided, followed by the description of the proposed method in Section III. Section IV shows and analyzes the simulation results. Finally, Section V provides some conclusions of this paper.

II. Literature Review

The electric load prediction can be classified into two main categories: statistical analysis method-based techniques and artificial intelligence algorithm-based ones. Statistical methods, such as multiple linear regression models (MLR) [10], exponential smoothing methods [11], autoregressive moving average (ARMA) methods [12] are widely used. MLR methods use multiple independent variables for the linear regression analysis. It does not lead to a good prediction accuracy in the face of a complex nonlinear system. ARMA type methods can make an accurate prediction for time series data with high stationarity and periodicity. When the load is affected by some complex factors and shows strong randomness and non-stationary characteristics, the prediction precision significantly declines.

Among artificial intelligence-based approaches, support vector regression (SVR) [13-14] and NN methods [15-17] are popular. SVR suffers from a high computational burden when dealing with a large amount of data. Various NN based approaches include back-propagation NN (BPNN) [15], recurrent NN (RNN) [16] and convolutional NN (CNN) [17]. BPNN can directly learn an affine rule from historical data without specifying an exact function form. For time-series data, such as electric load, the relationships between inputs contain valuable information. By taking inputs as the state values of the neuron of the previous time-step, RNN can learn the mapping between the inputs. Among them, the long-short term memory (LSTM) NN is the most popular one owing to the alleviation of gradient explosion. LSTM based approaches achieved promising results in load forecasting applications [17-18]. Note that NN based approaches are parametric methods, which typically require a large amount of data to learn the relative patterns from samples as they have many parameters to be optimized. Therefore, the direct application of NN is not appropriate as the data during the anomalous events are scarce.

The aforementioned methods belong to the point forecasting model, which only predicts a deterministic value. In recent years, with the increasing penetration level of flexible demand, the uncertainty quantization of load forecasting becomes more and more important for many applications, such as optimal bidding in the electricity market, probabilistic optimal power flow, and reliability planning [19]. Various probabilistic load forecasting approaches have been proposed that can be divided into three categories [20]: post-processing of the point forecast results [21], probabilistic forecasting methods [22], and scenario-based methods [23]. **However, most probabilistic forecasting methods require a relatively large amount of data for training and therefore are not suitable for load prediction during the anomalous events, where only a**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

handful of data can be utilized.

As a non-parametric method, GP can provide flexible function estimation considering uncertainties and thus is suitable for prediction tasks in safety-critical fields, such as power systems. Several studies have applied GP for the load prediction and it has been demonstrated that GP can produce better estimates than benchmark methods, like NN and SVR [24-27]. It is worth noting that when GP is utilized to solve the regression problem, the forecasting accuracy is sensitive to the selection of the covariance function. In [25], a covariance kernel is developed to incorporate daily/weekly patterns and weather conditions for the prediction of future loads. Three different kernels have been evaluated in [26]. The selection of covariance kernel function requires expert knowledge of the dataset, which is difficult to obtain when dealing with unforeseen scenarios, such as the COVID-19. This paper builds on the DGP and develops a new load forecasting method that can deal with anomalous events with limited recorded data. It is worth noting that DGP has not been investigated in the relevant power system applications.

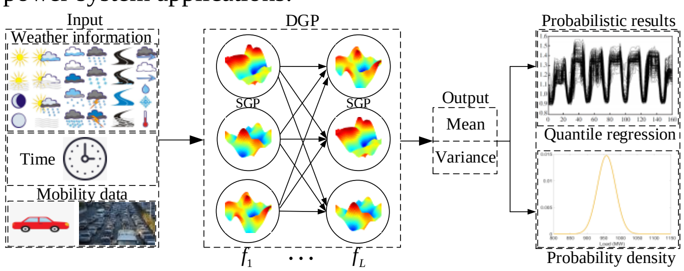


Fig. 2. Schematic of the proposed method.

III. Proposed DGP Model For Load Forecasting

In power systems, the load consumption data can be denoted as a time series $\{y_t\}$. To forecast future load demand y_{t+1}, the historical load consumption data $y_{t-H},...,y_t$ and other related variables, such as weather data, are utilized for constructing the training set. In this paper, they are expressed as x_t. Load forecasting aims to find a model $f(\cdot)$ which maps from x_t to y_{t+1} according to the gradient obtained by minimizing the pre-determined loss functions.

In general, parametric methods require a large amount of data to find relative patterns since those methods need to optimize many parameters. During the anomalous situations, such as the COVID-19 pandemic, the unprecedented changes in consumption patterns and the magnitude of load demand make it difficult to construct a good dataset that is sufficient for the training of parametric methods, especially at the beginning of the pandemic. Also, the parametric methods cannot be directly used for probabilistic forecasting.

Instead of looking for a parameterized model, we model the distribution of function $f(\cdot)$ based on Bayesian nonparametric methods. It allows us to deal with the uncertainties concerning new data. Motivated by that, this paper develops the DGP-based approach. In the sequel, GP modeling is first introduced. Then, the enhanced GP methods, i.e., sparse GP (SGP) and double stochastic variational inference DGP to improve the computational efficiency and generalizability to wide distributions are developed for probabilistic load prediction. The schematic of the proposed approach is shown in Fig. 2 and the detailed descriptions of the algorithms are shown below.

A. Introduction of GP Regression

Given a training data set $D=\{(x_n,y_n)\}_{n=1}^N$, where x_n represents the input of data-point n and y_n is the associated output; $X=[x_1,x_2,...,x_n]^T$ and $Y=[y_1,y_2,...,y_n]^T$ are the input and output sets. Assuming the state set of the stochastic process of input variables $f(X)=(f(x_i))$ obey the n-dimensional joint Gaussian distribution, f is thus a GP. It is specified by the mean function $m(x)$ and covariance function $K(X,X')$:

$$f(X)\sim GP(m(X),K(X,X')) \tag{1}$$

The GP regression model takes the mapping from X to Y as a GP. A standard GP regression model can be obtained by taking into account the homoscedastic Gaussian noise:

$$Y=f(X)+\varepsilon \text{ with } \varepsilon\sim N(0,\sigma_n^2 I) \tag{2}$$

where ε is the homoscedastic Gaussian noise; σ_n^2 is the variance; I is an identity matrix with appropriate dimensions. Since the noise is independent, y is also a GP:

$$Y\sim GP(m(X),K(X,X')+\sigma_n^2 I) \tag{3}$$

Then, the prior distribution of Y, i.e., Y^p, can be obtained on the training set according to Bayesian theory:

$$Y^p\sim N(0,K(X,X')+\sigma_n^2 I) \tag{4}$$

GP aims to forecast f_* given input x_*. The joint distribution of training set output Y and the test output f_* are

$$\begin{bmatrix} Y \\ f_* \end{bmatrix}\sim N(0,\begin{bmatrix} K(X,X)+\sigma_n^2 I_n & K(X,x_*) \\ K(x_*,X) & K(x_*,x_*) \end{bmatrix}) \tag{5}$$

where $K(X,X)=(k(x_i,x_j))$ represents the $N\times N$ covariance matrix on inputs of training set D; $k(x_i,x_j)$ is the kernel function; $K(X,x_*)=(k(x_i,x_*))$ represents the $N\times 1$ vector of covariance between the test point x_* and the training inputs in D; $K(x_*,x_*)$ is the covariance of the test points. The posterior distribution of f_* is given by

$$p(y_*\mid x_*,D)\sim N(\bar{f}_*,\mathrm{cov}(f_*)) \tag{6}$$

$$\bar{f}_*=k(x_*,x)^T(K+\sigma^2 I)^{-1}y \tag{7}$$

$$\mathrm{cov}(f_*)=k(x_*,x_*)-k(x_*,x)^T(K+\sigma^2 I)^{-1}k(x_*,x_*) \tag{8}$$

where $\bar{\mu}_*=\bar{f}_*$ and $\bar{\sigma}_{f_*}^2=\mathrm{cov}(f_*)$. The GP model can select different covariance functions, among which the Radial Basis Function (RBF) is the one that has been widely used:

$$k(x,x')=\sigma_f^2\exp(-\frac{1}{2}(x-x')^T M^{-1}(x-x')) \tag{9}$$

where σ_f^2 represents the variance of time series; $M=diag(l^2)$, where l is the variance scale. Both σ_f^2 and M are parameters to be optimized during the training. Since $\sigma^2 I$ is typically together with the covariance matrix K, it is also treated as the learnable parameters. The parameters to be optimized can be defined as $\theta=\{M,\sigma_f^2,\sigma^2\}$, which are updated by maximizing the following marginal likelihood function:

$$\hat{\theta}=\arg\max_{\theta} p(D\mid\theta) \tag{10}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

When the optimal parameters are obtained, the mean $\bar{\mu}_*$ and variance $\bar{\sigma}_{f_*}^2$ of the test point x_* can be calculated according to equations (7) and (8). The variance allows us to quantify the load forecasting uncertainties.

B. Computationally Efficient Sparse GP Regression

Since the training time complexity of GP is $O(N^3)$, it is not suitable for applications with a large number of training data. In this context, the SGP method is developed. By introducing *M* inducing points to approximate the original GP, the complexity can be reduced to $O(NM^2)$.

Define a set of *M* inducing points with inputs $Z=(z_1,...,z_M)^T$ and their corresponding outputs $U=(u_1,...,u_M)^T$. Since *U* and *f* are generated by the same GP, we get

$$p(f,u\,|\,X,Z)=N([f,u]\,|\,[0,0],K([X,Z],[X,Z])) \tag{11}$$

$$p(u\,|\,Z)=N(u\,|\,0,K(Z,Z)) \tag{12}$$

$$p(f\,|\,u;X,Z)=N(f\,|\,K_{nm}K_{mm}^{-1}u,K_{nn}-K_{nm}K_{mm}^{-1}K_{mn}) \tag{13}$$

Variational inference is introduced to reduce the computational burden. It aims to look for an approximated posterior $q(U,f)$ through the minimization of Kullback Leibler divergence between the true posterior *p* and the variational posterior *q*. The minimization problem is equivalent to the maximization of the following lower bound [28]:

$$\ell=E_{q(f,U)}[\log\frac{p(y,f,U)}{q(f,U)}] \tag{14}$$

The variational posterior is $q(U,f)=p(f\,|U)q(U)$ and $q(U)=N(\mu,\Sigma)$. When $q(U,f)=p(U,f\,|\,y)$ holds, the Gaussian marginal likelihood is formulated as

$$\begin{aligned}q(f)&=\int p(f\,|u)q(u)du=N(f\,|\,K_{nm}K_{mm}^{-1}u,K_{nn}+K_{nm}K_{mm}^{-1}(\Sigma-K_{mm})K_{mm}^{-1}K_{mn})\\&=\prod_i N(m_i,S_i^2)\end{aligned} \tag{15}$$

According to [28], the lower bound can be simplified and formulated as

$$\log p(y)\ge\sum_{i=1}^N \mathrm{E}_{q(f_i)}\log p(y_i\,|\,f_i)-KL(q(u)\,\|\,p(u)) \tag{16}$$

There are two sets of parameters to be optimized: the variational parameters $\theta_i=\left\{\left\{\mu^{(n)}\right\}_{n=1}^M,\left\{\Sigma^{(n)}\right\}_{n=1}^M,Z\right\}$ and the parameters of the kernel $\theta_h=\{M,\sigma_f^2,\sigma^2\}$, both of which are optimized by maximizing the lower bound.

C. Double Stochastic Variational Inference DGP Regression

DGP is a hierarchical generalization of GP that integrates the uncertainty quantification ability of non-parametric methods and the power of the DNN structure. GP relies on a sophisticated definition of the covariance function, which is a laborious process and also requires the empirical knowledge of the dataset. DGP overcomes this limitation by adopting a multilayer hierarchy and successive warping and stretching the input space, leading to a covariance function with a self-tuning ability that can fit arbitrary data. The deep structure of GP is achieved through a feed-forward and fully-connected manner. This paper further extends the SGP with multiple layers and the double stochastic variational inference allows more general distribution types.

For a DGP of depth *L*, each layer represents a GP to model function F^l that takes F^{l-1} and F^l as inputs and outputs for $l=1,...,L$. We consider the DGP model that stacks multiple-layer SGP, and the inducing points with inputs $\left\{Z^l\right\}_{l=0}^{L-1}$ and corresponding outputs $\left\{U^l\right\}_{l=1}^{L}$ are introduced. The joint probability density function of DGP can be denoted as:

$$p(Y,\left\{F^l,U^l\right\}_{l=1}^L)=\prod_{i=1}^N p(y_i\,|\,f_i^L)\prod_{l=1}^L p(F^l\,|U^l;F^{l-1},Z^{l-1})p(U^l;Z^{l-1}) \tag{17}$$

where $F_0=X$. In the DGP model, the marginal likelihood can also be obtained by the variational inference similar to the SGP. The inferred posterior is

$$q(\left\{F^l,U^l\right\}_{l=1}^L)=\prod_{l=1}^L p(F^l\,|U^l;F^{l-1},Z^{l-1})q(U^l) \tag{18}$$

where $q(U^l)=N(\mu^l,\Sigma^l)$. Then, the marginal likelihood is

$$\begin{aligned}q(\left\{F^l\right\}_{l=1}^L)&=\prod_{l=1}^L\int p(F^l\,|U^l)q(U^l)dU^l\\&=\prod_{l=1}^L q(F^l;F^{l-1},Z^{l-1})=\prod_{l=1}^L N(F^l\,|\,\tilde{m}^l,\tilde{S}^l)\end{aligned} \tag{19}$$

where $q(F_l;F_{l-1},Z_{l-1})$ is formulated as:

$$q(F^l;F^{l-1},Z^{l-1})=N(F^l\,|K_{l_{nm}}K_{l_{mm}}^{-1}U^l,K_{l_{nn}}+K_{l_{nm}}K_{l_{mm}}^{-1}(\Sigma^l-K_{l_{mm}})K_{l_{mm}}^{-1}K_{l_{mn}}) \tag{20}$$

and the marginal likelihood of the last layer is derived as:

$$q(f_i^L)=\int\prod_{l=1}^{L-1}q(f_i^l;f_i^{l-1},Z^{l-1})df_i^l \tag{21}$$

The evidence lower bound of the DGP is

$$\ell_{DGP}=\mathrm{E}_{q(\left\{F^l,U^l\right\}_{l=1}^L)}\left[\frac{p(Y,\left\{F^l,U^l\right\}_{l=1}^L)}{q(\left\{F^l,U^l\right\}_{l=1}^L)}\right] \tag{22}$$

which can be further simplified and formulated as [28]:

$$\ell_{DGP}=\sum_{i=1}^N\mathrm{E}_{q(f_i^L)}\left[\log p(y_n\,|\,f_n^L)\right]-\sum_{l=1}^L KL\left[q(U^l)\,\|\,p(U^l)\right] \tag{23}$$

Then the parameters for inference functions $\theta_i=\left\{\left\{\left\{\mu_l^{(n)}\right\}_{n=1}^M,\left\{\Sigma_l^{(n)}\right\}_{n=1}^M\right\}_{l=1}^L,Z_{l=0}^{L-1}\right\}$ and kernels $\theta_h=\left\{\left\{\sigma_{ard}\right\}_{l=1}^L,\left\{\omega_1,...\omega_{Q_l}\right\}_{l=1}^L,\left\{\sigma_l\right\}_{l=1}^L\right\}$ are optimized by maximizing the lower bound. When the learnable parameters are fixed, the distribution of the test points can be obtained according to

$$q(f_*^L)=\int\prod_{l=1}^{L-1}q(f_*^l;f_*^{l-1},Z_*^{l-1})df_*^l \tag{24}$$

The DGP model assumes that the posterior distribution to be Gaussian, which may not be true in practice. To this end, the sampling method is adopted. In particular, we sample $\xi_i^l=N(0,I)$ first, after which we draw the sampled variables $\hat{f}_i^l\sim q(f_i^l;\hat{f}_i^{l-1},Z^{l-1})$ for $l=1,...,L-1$ recursively as

$$\hat{f}_i^l=m_{\mu^l,Z^{l-1}}(\hat{f}_i^{l-1})+\xi_i^l\odot\sqrt{S_{\Sigma^l,Z^{l-1}}(\hat{f}_i^{l-1},\hat{f}_i^{l-1})} \tag{25}$$

where $\odot$ is the elementwise operator and *m* and *S* are expressed as follows:

$$m_{\mu^l,Z^{l-1}}(f_i^l)=[\tilde{m}^l]_i;\;\; S_{\Sigma^l,Z^{l-1}}(f_i^l,f_j^l)=[\tilde{S}^l]_{ij} \tag{26}$$

After obtaining the marginal distribution of each layer $q(\hat{f}_i^l)$, the learnable parameters can be optimized by maximizing the evidence lower bound. During the test stage, the distribution of the test points can be obtained:

$$q(f_*^L)=\frac{1}{S}\sum_{s=1}^S q(f_*^L;f_*^{(s)^{L-1}},Z^{L-1}) \tag{27}$$

The double stochastic variational inference DGP model includes two sources of stochasticity: i) the expectation is approximated using samples from the variational posterior; ii) the model requires sub-sampling the data for scenarios with a large number of training data.

D. Uncertainty Quantification and Algorithm Implementation

After the distribution of the test points is obtained from (27), n samples $\{s_1,...,s_n\}$ are generated from that distribution. Then, the confidence interval of the prediction can be obtained via

$$\underline{c}=\overline{s}-\frac{t(\alpha/2,n-1)*\sigma}{\sqrt{n}},\ \overline{c}=\overline{s}+\frac{t(\alpha/2,n-1)*\sigma}{\sqrt{n}} \tag{28}$$

where $\underline{c}$ and $\overline{c}$ represent the lower and upper bounds of the confidence interval, respectively; $\overline{s}$ and σ represent the mean and variance for n samples; α is the degree of confidence; $t(\cdot)$ is the t distribution, the value of which can be obtained by looking up the table. The implementations of the proposed method are shown in Algorithm I.

Algorithm 1 Implementation of the proposed method

Training procedure

Input: Training set $T_r=\{X,Y\}$

Output: Parameters θ_i and kernel parameters θ_h

Initialize: Randomly initialize all parameters

for epoch=1:stop_epoch **do**
  Samples a mini-batch size data from T_r
  for iteration=1:stop_iteration do
    Calculate variational distribution of each layer $q(U^l)=N(\mu^l,\Sigma^l)$
    Calculate the marginal distribution of each layer $q(f^l)$ according to (20) and (21)
    Samples from marginal according to (25)
    Calculate ℓ_{DGP} according to (23)
    Update θ_i through $\theta_i \leftarrow \theta_i - \nabla_{\theta_i}\ell_{DGP}$
    Update θ_h through $\theta_h \leftarrow \theta_h - \nabla_{\theta_h}\ell_{DGP}$
  end for
end for

Test procedure

Input: Test data points T_*

Output: Probabilistic forecasting result

Load parameters θ_i and kernel parameters θ_h
Calculate the marginal distribution of each layer $q(f^l_{i,*})$
Sample from marginal of each layer according to (25)
Calculate marginal distribution of output layer according to (27)
Obtain interval prediction result via (28)

IV. Results and Discussions

In this section, tests are carried out on different real-world data sets to illustrate the effectiveness of the proposed approach. In particular, comparative tests among various point forecasting methods utilizing small-scale and middle-scale training samples are first illustrated. Then, comparative results for probabilistic forecasting are presented and analyzed.

A. Experimental Setup

Hourly electricity load demand data of regions with different sizes are utilized: metropolitan-level data, including four cities in America (Boston, Seattle, Chicago, and Philadelphia) and country-level data, including two countries in Europe (Germany and France). The data are obtained from [29] and are publicly available. The metropolitan-level city data have 46-dimensional features, including timing information, weather, and mobility data. The weather information includes temperature, humidity, cloud cover, precipitation, and air pressure. The timing features can encode information, such as month index, day index, hour index, and distinction features of weekday and holidays. Mobility data include six location-specific metrics achieved from Google (parks, workplaces, residential, retail & recreation, workplaces, and grocery & pharmacy) and three mobility features from Apple (walking, driving, and transit). Since the weather and mobility data of several cities are concatenated to reflect patterns in the country-level dataset, they contain 60 input features for each instance. All data are at a one-hour interval and range from February 15th to May, 15th in 2020. The data are split into training, validation, and test sets proportionally. The training set is used to train the load forecasting model, and the validation set is applied to select the model with the best forecasting accuracy, while the test set is utilized to evaluate the performance of the method. The simulations are conducted on a workstation with an Intel 2.2 GHz Xeon E5-2630 CPU. The code is written in Python with TensorFlow and gpFlow.

B. Evaluation Metrics

Comparative tests are carried out among various methods on both point and probabilistic forecasting. To evaluate the performances of those methods, two metrics are utilized:

1) ***Metrics for point forecasting****:* The mean absolute percentage error (MAPE) is widely used metric by the power industry owing to its transparency and simplicity [18]. It calculates the absolute difference between the forecasting values and actual ones in %, and it is defined as:

$$MAPE=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_{i,p=50}-y_i}{y_i}\right| \tag{29}$$

where y_i represents the actual load value; n is the number of instances of the test set; $\hat{y}_{i,p=50}$ represents the 50th percentile value of the forecasted load.

2) ***Metrics for probabilistic forecasting****:* Reliability, sharpness, and resolution are three commonly used attributes for the evaluation of probabilistic forecasting. As an error measure for quantile prediction, Pinball loss is a comprehensive metric to measure the results of probabilistic forecasting, which is defined as [19]:

$$Pinball(\hat{y}_{i,q},y_i,q)=\begin{cases}(y_i-\hat{y}_{i,q})q, & \hat{y}_{i,q}<y_i\\ (\hat{y}_{i,q}-y_i)(1-q), & \hat{y}_{i,q}>y_i\end{cases} \tag{30}$$

where $\hat{y}_{i,q}$ represents the load forecasting value at the qth quantile. The pinball loss can be obtained by summing $Pinball(\hat{y}_{i,q},y_i,q)$ across all quantiles (q=0.01, 0.02, …, 0.99) over the prediction horizon. A lower score implies better performance.

C. Point Forecasting Evaluations with Limited Data

The first test is carried out among various methods when only three days' data during the pandemic are used for training. This test is used to illustrate that the proposed method can learn a better forecasting model than other approaches with limited training data. The data are split into the training set (from 07-May-2020 to 09-May-2020), validation set (from 10-May-2020 to 12-May-2020), and test set (from 13-May-2020 to 15-May-2020). The benchmark methods include: 1) support vector regression (SVR) [13]; 2) back-propagation NN (BPNN) [15]; 3) SGP; 4) VAE-DGP [9]. The RBF is selected as the kernel function for SGP, VAE-DGP, and the proposed method. The hyper-parameters of these methods are tuned utilizing the validation set and the model with the best performance of each method is adopted. **We would like to emphasize that SGP and VAE-DGP have not been applied for load forecasting in the literature and it is our effort to adopt them for this application**. The MAPE of each method on the validation/test set is listed in Table I. The best performance (test set) of each case is shown in a bold color. Note that the results are average values of five repeated tests. It can be observed that GP based approaches outperform BPNN methods by a large margin when only 3 days' data are utilized for training. This may be because the BPNN is parametric method, which involves many parameters to be optimized. Therefore, they cannot find all hidden patterns among samples from relatively limited training data. Furthermore, since BPNN has too many parameters (the weights and bias of NN) to be optimized from comparatively few training samples, they are vulnerable to over-fitting issues, leading to an unstable performance on the test data. By contrast, the kernel function based SVR and Bayesian GP based approaches can achieve effective inductive reasoning based on only a limited number of training data. This characteristic makes them a suitable alternative for the load prediction during the anomalous events. Thanks to the "self-tuning" covariance function by hierarchically warping and stretching the input space, the proposed method outperforms the one-layer SGP method in 5 out of 6 cases, demonstrating its better flexibility to various scenarios. Since the proposed DGP method learns a representation non-parametrically with very few parameters to estimate, it avoids the overfitting problem faced by the parametric methods in the presence of increased number of layers. The integration of the flexibility of deep structure models and the effective inductive reasoning ability of GP method makes the proposed method a promising alternative for the electrical load forecasting against the extreme events. Both the proposed and SGP methods achieve better performance than the VAE-DGP based approach in most cases. It is worth noting that the gap between the MAPEs achieved by the BPNN on validation and test sets are larger than the non-parametric methods. The main reason for this phenomenon may be that the time index (e.g., the day index) of the validation set is not seen by the forecasting model during training. Parametric methods, such as the neural networks are prone to over-fit the training data, while the kernel function and Bayesian theory-based methods have better inference ability.

The second test is carried out among various methods when middle-size samples are utilized to train the models. The data are split into the training set (from 15-Feb-2020 to 29-Apr-2020), validation set (from 02-May-2020 to 08-May-2020), and test set (from 09-May-2020 to 15-May-2020). The MAPEs of various methods on the test set are displayed in Table II. Since more data are utilized for training, the prediction accuracies of all methods are improved as compared to those in Test 1. In this test, the proposed method outperforms the SVR, BPNN, VAE-DGP methods in all cases, and the SGP method in 5 cases, demonstrating that the proposed method can achieve high forecasting accuracy when both small-scale and middle-scale data are utilized for training. The gaps between the MAPEs achieved by the BPNN method on validation and test set are also reduced as compared to those in Test 1. This is because a good number of samples is used by the neural network during the training process. The forecasting results of various methods on test data of Seattle when middle-size data are utilized for training are shown in Fig. 3. It can be observed that the BPNN suffers from larger deviations from the actual load demand compared with GP based methods. The proposed approach gets closer to the actual load than the SGP and VAE-DGP methods. The results are consistent with those in Table II.

D. Probabilistic Forecasting Evaluations with Limited Data

The pinball losses achieved by various probabilistic forecasting methods on test data under different training data scales are shown in Table III and Table IV. The pinball loss is a comprehensive measure for the evaluation of probabilistic forecasts. It can be observed from Table III that when only three days' data are utilized for training, the proposed method can achieve the best performance in most cases, demonstrating that the proposed method can better quantify the uncertainties than other methods utilizing a limited number of training samples. When the amount of training data is increased, the pinball loss of all the methods decrease. The pinball loss achieved by the proposed approach is at most 23.0% lower than that obtained by the SGP and VAE-DGP method. The probabilistic results are consistent with those in point forecasting tests.

The probabilistic forecasting results of various methods on test data are shown in Fig. 4, where the results for BPNN and SVR are not shown as they are only point forecasting approaches. The actual load demand is represented by the red dot. The 70%, 80%, 90%, and 95% confidence intervals are represented by the gradually increasing depth values of blue color. It can be observed from Fig. 4 (c) that the 95% confidence interval of the proposed approach can cover most of the actual load demands. By contrast, there are many actual load points not covered by the 95% confidence intervals of the SGP and VAE-DGP, see t=50-70 in Fig. 4 (a) and t=1-25, t=60-65 in Fig. 4 (b) for example. This demonstrates that the proposed approach can better capture the uncertainty of load forecasting. It is worth noting that the peak and valley values of the load demand, which are crucial for the operation and management of power systems, are well predicted by the proposed method, which is not the case for the SGP and VAE-DGP methods.

The probability density curve and the point forecasting value of various methods of Seattle when the load demand is at curve peak and valley are plotted in Fig. 5 to further illustrate the uncertainty capture ability of the proposed approach. The prediction results at the peak (t=16:00 at 15th May) are shown in Fig. 5 (a). It can be observed that the deterministic methods do not provide accurate point forecasting outcomes due to the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

Table I. MAPE achieved by various methods on validation/test data when 3 days' data are utilized for training

	Boston	Chicago	Seattle	Philadelphia	Germany	France
SVR	4.1%/4.1%	5.6%/5.2%	6.97%/5.3%	4.8%/4.6%	13.5%/8.8%	10.8%/14.0%
BPNN	7.9%/5.6%	10%/6.6%	20.1%/14.1%	7.1%/5.7%	**12.4%/7.8%**	14.5%/12.4%
SGP	4.1%/4.5%	6.1%/4.3%	5.4%/4.3%	4.81%/3.8%	16.5%/12.7%	**6.7%/9.0%**
VAE-DGP	4.3%/4.3%	6.5%/6.5%	7.1%/5.5%	7.0%/7.9%	16.5%/12.8%	12.4%/14.1%
Proposed	**4.0%/3.7%**	**5.6%/4.1%**	**5.7%/4.1%**	**4.35%/3.7%**	14.3%/7.9%	7.2%/10.6%

Table II. MAPE achieved by various methods on validation/test data when 75 days' data are utilized for training

	Boston	Chicago	Seattle	Philadelphia	Germany	France
SVR	6.4%/5.6%	2.5%/4.6%	4.6%/5.6%	3.8%/4.8%	6.2%/6.9%	7.6%/8.4%
BPNN	3.3%/3.2%	3.5%/5.3%	4.6%/6.4%	4.6%/5.4%	4.2%/4.3%	4.3%/6.9%
SGP	3.0%/3.2%	2.7%/5.1%	4.6%/5.1%	3.7%/5.0%	**3.2%/3.0%**	3.99%/5.3%
VAE-DGP	3.4%/2.9%	3.4%/4.5%	4.5%/6.9%	4.5%/4.6%	3.9%/4.1%	2.3%/6.5%
Proposed	**3.0%/2.8%**	**2.4%/4.1%**	**3.7%/4.5%**	**3.2%/4.2%**	3.9%/3.4%	**3.6%/5.1%**

Table III. Pinball loss achieved by various methods on validation/test data when 3 days' data is utilized for training

	Boston	Chicago	Seattle	Philadelphia	Germany	France
SGP	32.1/37.8	186.8/137.2	16.4/13.1	399.8/406.3	2796.2/2301.7	**1001.4/1489.5**
VAE-DGP	32.2/33.8	189.4/203.5	19.6/15.6	574.1/742.4	2882.4/2394.1	2010.2/2598.3
Proposed	**33.3/31.2**	**181.7/132.7**	**16.9/12.9**	**390.4/396.5**	**2362.1/1529.5**	1146.4/1891.1

Table IV. Pinball loss achieved by various methods on validation/test data when 75 days' data is utilized for training

	Boston	Chicago	Seattle	Philadelphia	Germany	France
SGP	23.4/26.2	88.4/184.8	16.8/17.6	321.6/513.5	**572.9/525.3**	598.8/868.1
VAE-DGP	25.9/22.1	104.4/162.3	15.2/22.9	381.7/497.2	724.0/770.8	326.9/1169.8
Proposed	**21.8/21.5**	**75.9/142.3**	**13.2/13.8**	**265.3/409.3**	700.9/572.4	**540.99/854.1**

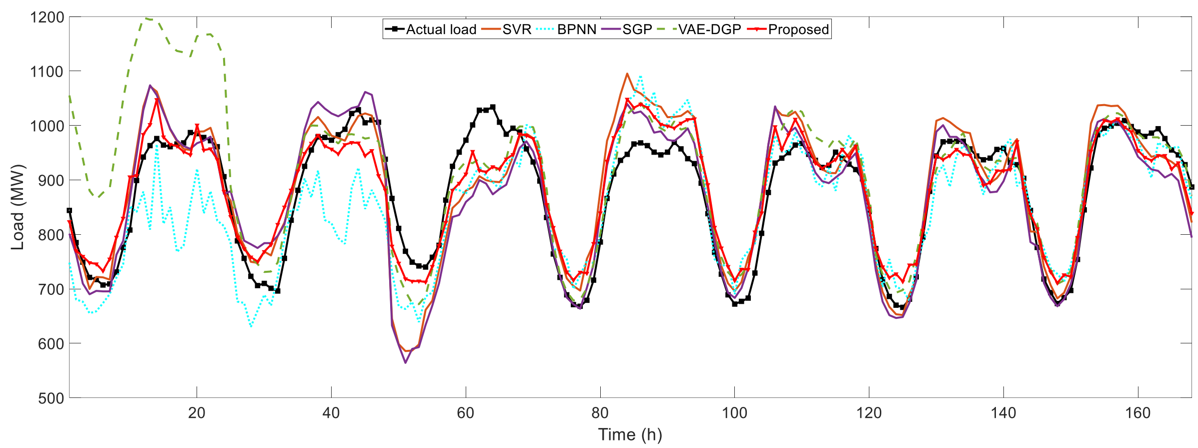
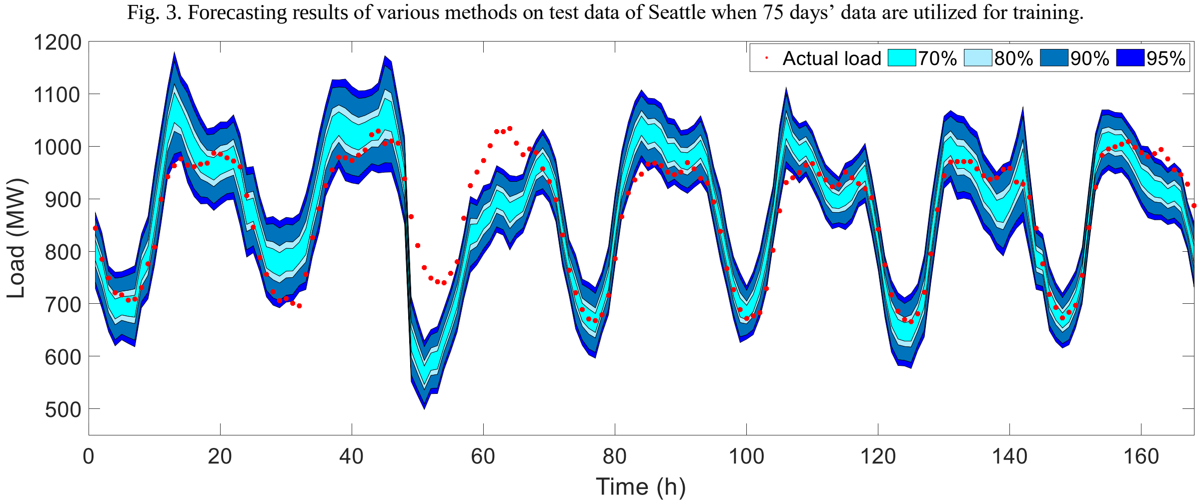
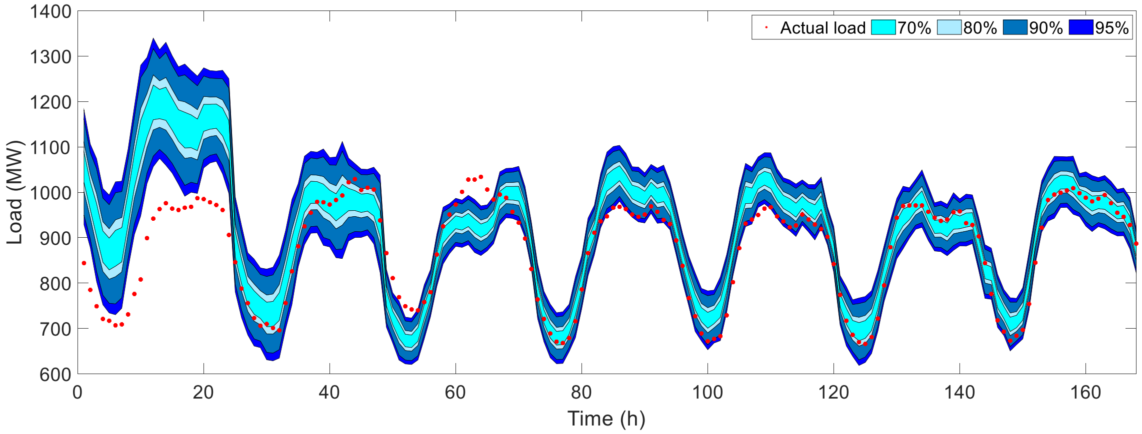


Fig. 3. Forecasting results of various methods on test data of Seattle when 75 days' data are utilized for training.

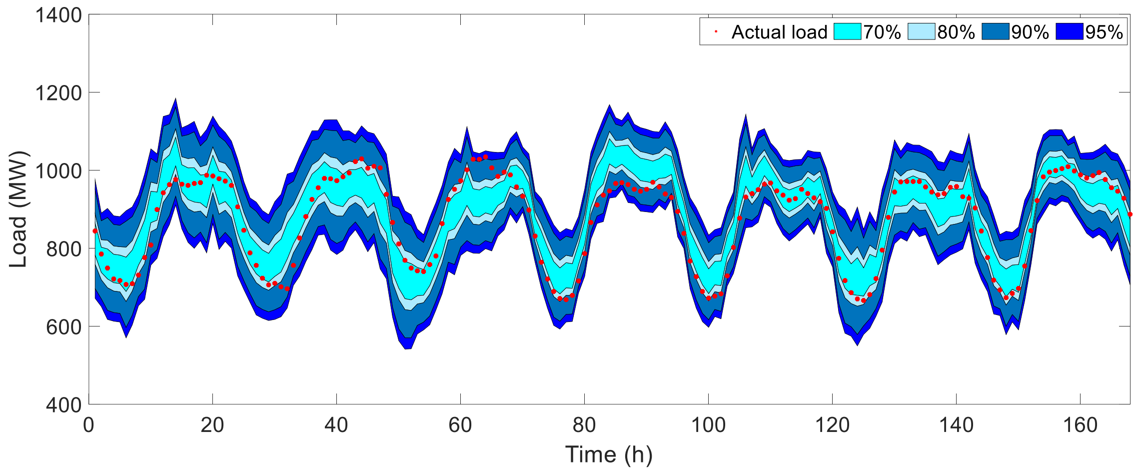


(a) SGP

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics



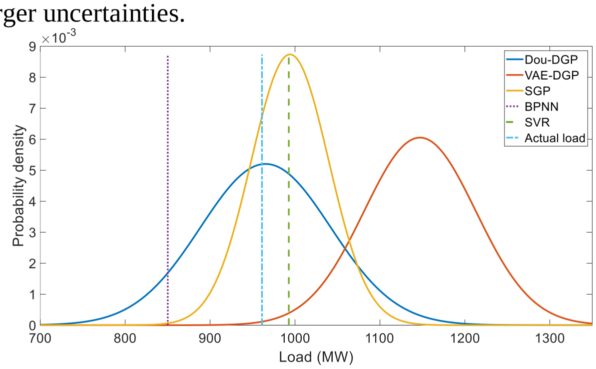
(b) VAE-DGP



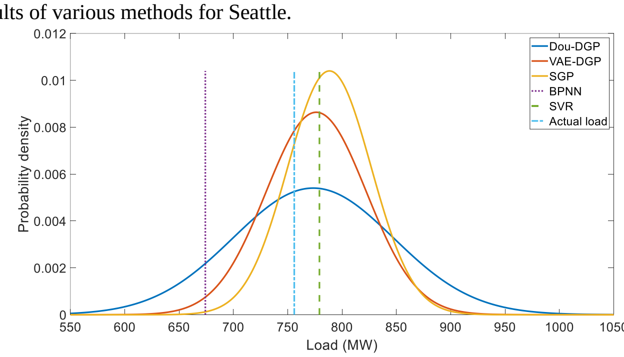
(c) Proposed method

Fig. 4. Probabilistic load forecasting results of various methods for Seattle.

uncertainty caused by the pandemic. By contrast, the GP based probabilistic methods can capture the uncertainties. Compared with the SGP and the VAE-DGP, the 50th percentile prediction results of the proposed method get closer to the actual electrical load demand. A similar phenomenon can be observed in Fig. 5 (b) when the load demand is at the curve valley (t=3:00 at 16th May). It can also be observed that the probability density curve of the proposed method is less steep than SGP and VAE-DGP methods. This may be due to multiple uncertainties that can affect the load consumption behaviors, which make it difficult to accurately predict the peak and valley values. As a result, the proposed method learns a relatively conservative strategy, which would be good to make a conservative decision under larger uncertainties.



(a) Results at curve peak.



(b) Results at curve valley.

Fig. 5. Probabilistic density curve and point forecasting results of various methods of Seattle on 5th May.

E. Tests on Northern Italy Data

To further evaluate the performance of the proposed method, extensive tests are carried out on electrical load demand data of Northern Italy. The containment measures of Italy can be divided into two stages: 1) the first stage started from 23th February, including the closure of restaurants, bars, and schools after 6 p.m. Since the first stage containment measures had limited effects on the control of the pandemic, more restrictive measures take effects after 9th March in the whole country; 2) in the second stage, the government announced to shut down all the nonessential production activities. Strong decreases of electric demand in Northern Italy have been observed since the total lockdown is implemented. The electrical load demand in this period is utilized in this paper to evaluate the performance

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

of the proposed forecasting method. Hourly electrical load data of Northern Italy can be found from [3]. The input features include the historical electrical load demand data of the past 24 hours. Two tests with different size of data are carried out to evaluate the performances of various methods: 1) the first test utilizes small-scale data during the pandemic to train the forecasting models. The data range from 20th April to 28th April, which are split into training set (20-Apr-2020 to 22-Apr-2020), validation set (23-Apr-2020 to 25-Apr-2020), and test set (26-Apr-2020 to 28-Apr-2020); 2) The second test utilizes middle-size data ranging from 11-Jan-2020 to 14-Apr-2020. The data are divided into three non-overlapping intervals representing the training set (11-Jan-2020 to 25-Mar-2020), validation set (26-Mar-2020 to 4-Apr-2020) and test set (5-Apr-2020 to 14-Apr-2020), respectively.

The comparison results by different methods when different sizes of data are utilized are listed in Tables V and VI. It can be observed from Table V that the proposed method can always obtain the lowest prediction error when small-scale and middle-sale data are used for training. It achieves maximum 165% and 495% improvements as compared to other methods under the two tests.

Table V. MAPE achieved by various methods on validation/test data when 3 days' data and 75 days' data are utilized for training

	Test 1	Test 2
SVR	16.7%/17.4%	14.3%/14.5%
BPNN	7.3%/6.2%	3.1%/3.5%
SGP	6.3%/5.8%	2.6%/3.1%
VAE-DGP	8.0%/6.9%	2.4%/3.0%
Proposed	**6.3%/5.2%**	**2.4%/2.8%**

Table VI. Pinball loss achieved by various methods on validation/test data when 3 days' and 75 days' data are utilized for training

	Test 1	Test 2
SGP	356.3/356.6	281.1/257.1
VAE-DGP	404.7/381.0	**271.5/242.2**
Proposed	**355.6/345.6**	276.9/253.8

V. Conclusions And Future Works

This paper proposes a new probabilistic forecasting method for electric load prediction during abnormal event when only a limited number of data is utilized. The key idea is to cast the load forecasting as a regression problem and develop an advanced double stochastic variational inference DGP method. The mobility data that reflect the behavior pattern are also used as additional features to deal with the uncertainties caused by the stay-at-home order. The proposed method can learn the relative patterns from only a handful of training data and capture the uncertainties caused by the pandemic. Both point and probabilistic forecasting measures are used for the performance evaluation of the proposed method. Comparative tests on a series of datasets demonstrate that: 1) when only small-scale training data are used, the performance achieved by the proposed method significantly outperforms the parametric methods in most scenarios. Its performance also outperforms the kernel-based point forecasting method and Bayesian theory-based probabilistic methods in most scenarios. The prediction error can be reduced by at most 70.9% via the proposed method when only three days' data are utilized for training; 2) when more data are available and utilized for training, our proposed method outperforms both the parametric and Bayesian theory-based methods in most scenarios. The prediction error can be reduced by 50.7% at most; 3) the proposed method can better capture the uncertainties than other probabilistic forecasting methods. The pinball loss obtained by the proposed method is at most 46.6% and 39.7% lower than other probabilistic forecasting methods when small-scale and middle-scale training data are utilized, respectively. Future works will be on extending the proposed method for other power system applications, where not too many data are available.

References

[1] A. Paaso, *et. al*, "Sharing knowledge on electrical energy industry's first response to COVID-19," 2020.

[2] IEA, "Covid-19 impact on electricity," 2020. [Online]. Available: https://www.iea.org/reports/covid-19-impact-on-electricity.

[3] "Entsoe transparency platform.", 2020. [Online]. Available: https://transparency.entsoe.eu.

[4] O. David, D. V. Joseph, G. Yannig, [2020, Sep.]. Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France. [Online]. Available: https://arxiv.org/pdf/2009.06527.

[5] H. Zhong, Z. Tan, Y. He, L. Xie and C. Kang, "Implications of COVID-19 for the electricity industry: A comprehensive review," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 489-495, Sept. 2020.

[6] H. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: a review and evaluation." *IEEE Trans. Power Syst*., vol. 16, no. 1, pp. 1-55, 2001.

[7] C. E. Rasmussen, C. K. I. Williams. "Gaussian processes for machine learning." MIT Press, 2005.

[8] M. Havasi, *et. al*., "Inference in deep Gaussian processes using stochastic gradient Hamiltonian Monte Carlo." *32nd Conference on Neural Information Processing Systems*, Montreal, Canada, 2018.

[9] Z. Dai, A. Damianou, J. Gonzalez, "Variational auto-encoded deep Gaussian processes." *International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.

[10] T. Hong, P. Wang, H. Willis, "A naïve multiple linear regression benchmark for short term load forecasting." *IEEE Power & Energy Society General Meeting*, Detroit, Michigan, USA, 2011.

[11] K. B. Song, S. Ha, J. Park, D. Kweon, K. Kim, "Hybrid load forecasting method with analysis of temperature sensitivities," *IEEE Trans. Power Syst.*, vol. 21, no. 2, pp. 869-876, May 2006.

[12] S. J. Huang and K. R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst*., vol. 18, no. 2, pp. 673-679, May 2003.

[13] B. J. Chen, M. W. Chang, and C. J. Lin, "Load forecasting using support vector Machines: a study on EUNITE competition 2001," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1821-1830, Nov. 2004.

[14] C. N. Ko, and C. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413-422, Jan. 2013.

[15] F. Xu, X. Cun, M. Yan, H. Yuan, Y. Wang, and L. Lai, "Power market load forecasting on neural network with beneficial correlated regularization," *IEEE Trans. Ind. Informatics*., vol. 14, no. 11, pp. 5050-5059, Nov. 2018.

[16] M. Tan, S. Yuan, S. Li, Y. Su, H. Li and F. He, "Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning," *IEEE Trans. Power Syst*., vol. 35, no. 4, pp. 2937-2948, July 2020.

[17] X. Dong, L. Qian and L. Huang, "Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach," *IEEE International Conference on Big Data and Smart Computing*, 2017.

[18] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[19] T. Hong, and F. Shu. "Probabilistic electric load forecasting: A tutorial review." *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, 2016.

[20] M. Sun, T. Zhang, Y. Wang, G. Strbac and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188-201, Jan. 2020.

[21] J. Xie, T. Hong, T. Laing, and C. Kang, "On normality assumption in residual simulation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1046–1053, May 2017.

[22] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, C. K, "Probabilistic individual load forecasting using pinball loss guided LSTM", vol. 235, pp. 10-20, Feb. 2019.

[23] J. Xie and T. Hong, "Temperature scenario generation for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1680–1687, May 2018.

[24] G. Xie, X. Chen and Y. Weng, "An integrated Gaussian process modeling framework for residential load prediction," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7238-7248, Nov. 2018.

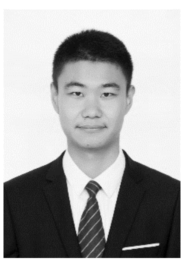
[25] M. Blum and M. Riedmiller, "Electricity demand forecasting using Gaussian processes," in *Proc. 15th AAAI Conf. Trading Agent Des. Anal.*, 2013,

[26] J. R. Lloyd, "GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes" *International Journal of Forecasting*, vol. 30, no. 2, pp. 369-374, 2012.

[27] G. Xie, X. Chen and Y. Weng, "An integrated Gaussian process modeling framework for residential load prediction," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7238-7248, Nov. 2018.

[28] H. Salimbeni, M. Deisenroth. "Doubly stochastic variational inference for deep Gaussian processes," 31st *Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017.

[29] "Load forecasting during COVID-19" 2020. [Online]. Available: https://github.com/chennnnnyize/Load-Forecasting-During-COVID-19.



Di Cao is currently working toward the Ph.D. degree in control science and engineering at the University of Electronic Science and Technology of China. His research interest includes optimization of distribution network and applications of machine learning in power systems.

Junbo Zhao (SM'19) has been an Assistant Professor at Mississippi State University, Starkville, MS, USA since 2019. He received the Ph.D. degree from the Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, in 2018. He was a Research Assistant Professor at Virginia Tech from May 2018 to August 2019. He did the summer internship at Pacific Northwest National Laboratory from May to August 2017. He is currently the chair of the IEEE Task Force on Power System Dynamic State and Parameter Estimation and the IEEE Task Force on Cyber-Physical Interdependency for Power System Operation and Control, co-chair of the IEEE Working Group on Power System Static and Dynamic State Estimation, the Secretary of the IEEE PES Bulk Power System Operation Subcommittee.

He has published three book chapters and more than 100 peer-reviewed journal and conference papers, where more than 60 appear in IEEE Transactions. His research interests are cyber-physical power system modeling, estimation, security, dynamics and stability, uncertainty quantification, renewable energy integration and control, robust statistical signal processing and machine learning. He serves as the editor of *IEEE Transactions on Power Systems*, *IEEE Transactions on Smart Grid* and IEEE Power and Engineering Letters, the Associate Editor of *International Journal of Electrical Power & Energy Systems*, and the subject editor of *IET Generation, Transmission & Distribution*. He is the receipt of best paper awards of IEEE PES General Meeting at 2020 and 2021, and 2019 IEEE PES ISGT Asia. He received the Top 3 Associate Editor Award from IEEE Transactions on Smart Grid and IEEE PES Outstanding Engineering Award in 2020. He has been listed as 2020 World's Top 2% Scientists released by Stanford University.

Weihao Hu (S'06–M'13–SM'15) received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in electrical engineering, and Ph. D. degree from Aalborg University, Denmark, in 2012.

He is currently a Full Professor and the Director of Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China (UESTC). He was an Associate Professor at the Department of Energy Technology, Aalborg University, Denmark and the Vice Program Leader of Wind Power System Research Program at the same department. His research interests include artificial intelligence in modern power systems and renewable power generation. He has led/participated in more than 15 national and international research projects and he has more than 170 publications in his technical field.

He is an Associate Editor for IET Renewable Power Generation, a Guest Editor-in-Chief for Journal of Modern Power Systems and Clean Energy Special Issue on Applications of Artificial Intelligence in Modern Power Systems, a Guest Editor-in-Chief for Transactions of China Electrical Technology Special Issue on Planning and operation of multiple renewable energy complementary power generation systems, and a Guest Editor for the IEEE TRANSACTIONS ON POWER SYSTEM Special Section on Enabling very high penetration renewable energy integration into future power systems. He was serving as the Technical Program Chair (TPC) for IEEE Innovative Smart Grid Technologies (ISGT) Asia 2019 and is serving as the Conference Chair for the Asia Energy and Electrical Engineering Symposium (AEEES 2020). He is currently serving as Chair for IEEE Chengdu Section PELS Chapter. He is a Fellow of the Institution of Engineering and Technology, London, U.K. and an IEEE Senior Member.

Yingchen Zhang (Senior Member, IEEE) received the B.S. degree from Tianjin University, China, in 2003 and the Ph.D. degree from Virginia Polytechnic Institute and State University, in 2010. He is Chief Scientist and Manager of the Sensing and Predictive Analytics group at the National Renewable Energy Laboratory. He is also a Visiting Research Assistant Professor at the University of Denver and an Adjunct Faculty at Colorado State University. He has authored or coauthored more than 150 peer reviewed publications and holds one U.S. patent. His key areas of expertise lie in sensing and measurement for power systems, artificial intelligence, predictive analytics, advanced energy management system for future grids, the impact of large-scale integration of renewable energies on power system operations. Dr. Zhang previously worked for the California ISO developing and implementing State Estimator and Network Applications. Dr. Zhang serves as an Editor for the IEEE Transactions on Sustainable Energy and IEEE Power Engineering Letters.

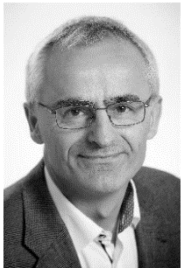
Qishu Liao is currently working toward the M.S. degree in electrical engineering at the University of Electronic Science and Technology of China. His research interest is applications of machine learning in power systems.

Zhe Chen (M'95-SM'98-F'19) received the B.Eng. and M.Sc. degrees from Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph.D. degree from University of Durham, Durham, U.K.

He is a Full Professor with the Department of Energy Technology, Aalborg University, Denmark. He is the Leader of Wind Power System Research Program in the Department of Energy Technology, Aalborg University and the Danish Principle Investigator for Wind Energy of Sino-Danish Centre for Education and Research. His research areas include power systems, power electronics and electric machines; and his main current research interests are wind energy and modern power systems. He has led many research projects and has more than 400 publications in his technical field.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2021.3081531, IEEE Transactions on Industrial Informatics

Dr. Chen is an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS, a Fellow of the Institution of Engineering and Technology, London, U.K., a Chartered Engineer in the U.K., and a Fellow of the IEEE.



Frede Blaabjerg (S'86–M'88–SM'97–F'03) was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. From 1988 to 1992, he got the PhD degree in Electrical Engineering at Aalborg University in 1995. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. From 2017 he became a Villum Investigator. He is honoris causa at University Politehnica Timisoara (UPT), Romania and Tallinn Technical University (TTU) in Estonia.

His current research interests include power electronics and its applications such as in wind turbines, PV systems, reliability, harmonics and adjustable speed drives. He has published more than 600 journal papers in the fields of power electronics and its applications. He is the co-author of four monographs and editor of ten books in power electronics and its applications.

He has received 32 IEEE Prize Paper Awards, the IEEE PELS Distinguished Service Award in 2009, the EPE-PEMC Council Award in 2010, the IEEE William E. Newell Power Electronics Award 2014, the Villum Kann Rasmussen Research Award 2014, the Global Energy Prize in 2019 and the 2020 IEEE Edison Medal. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON POWER ELECTRONICS from 2006 to 2012. He has been Distinguished Lecturer for the IEEE Power Electronics Society from 2005 to 2007 and for the IEEE Industry Applications Society from 2010 to 2011 as well as 2017 to 2018. In 2019-2020 he serves as President of IEEE Power Electronics Society. He is Vice-President of the Danish Academy of Technical Sciences too. He is nominated in 2014-2019 by Thomson Reuters to be between the most 250 cited researchers in Engineering in the world.