Generating scenarios from probabilistic short-term load forecasts via non-linear Bayesian regression

Markus Löschenbrand
Energy Systems
SINTEF Energy
Trondheim, Norway
markus.loschenbrand@sintef.no

Sébastien Gros
Department of Engineering Cybernetics
NTNU
Trondheim, Norway
sebastien.gros@ntnu.no

Venkatachalam Lakshmanan
Energy Systems
SINTEF Energy
Trondheim, Norway
venkatachalam.lakshmanan@sintef.no

***Abstract*—In this paper we present a simple and intuitive method for fitting a non-linear Bayesian regression model on short-term load forecasts. Such models have been implemented via Bayesian neural networks, which are known for their hyper-parameter sensitivity. We instead show a more general method to fit any regression model and demonstrate this by using a tree-model. Further, we evaluate the results against non-linear quantile regression, a common technique in probabilistic load forecasting. The resulting model allows to generate samples for future scenarios and thus can be applied to operations problems such as dynamic control of battery storage, an application that quantile regression is unfit for.**

***Index Terms*—load forecasting, Bayesian regression, non-linear regression, scenario generation**

I. Introduction

Load forecasting, i.e. the accurate prediction of future load patterns, plays a critical role in the planning and operation of electric power systems. Generally, and as shown in Ref. [1], such load forecasts can be classified as long-, medium- and short-term. The related horizons are years for long-term, betweena week up to years as medium-term and horizons below a week as short-term.

A. Motivation and Background

Various forecast horizons also have different applications. Long-term load forecasting, for example, is used in network planning activities. Medium-term load forecasts are applied in seasonal generation planning [2]. Short-term forecasts are used in the operation and control of power systems. An example of such is given by a cost-effective operation of a battery storage system, a dynamic operations problem. Early works on this topic highlight the dependency on uncertainty, specifically in form of scenarios [3]. This is also supported by current works which use similar scenario formulation of load uncertainties [4]. This means, that optimal operation of such systems requires accurate representation of uncertainty.

This work is funded by CINELDI – Centre for intelligent electricity distribution, an 8 year Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257626/E20). The authors gratefully acknowledge the financial support from the Research Council of Norway and the CINELDI partners. Lede (formerly Skagerak Nett) is gratefully acknowledged for sharing of load data.

978-1-7281-7660-4/21/$31.00 ©2021 IEEE

Nomenclature

t	period		s	branching values
x	input matrix		a, b	matrix index
f	function		L	loss function
y	output matrix		q	quantile
ξ	error/noise		P	probability
M	mapping		τ	percentile
R	region		$\mathcal{N}$	Gaussian
j	variable index		σ^2	standard deviation

B. Relevant Literature

A common technique to consider future uncertainty in forecasting is that of probabilistic forecasting. Ref. [5] provides an overview of probabilistic load forecasting techniques. It discusses a wide range of methods. The later presented regression trees are, however, only discussed in the context of clustering. Similarly, in the literature on deterministic load forecasting, regression trees have mostly been used as non-linear regressors [6]. Current state-of-the-art methods which consider uncertainty are however build on neural networks as non-linear regressors.

Most of the existing approaches for probabilistic forecasting are built on the concept of quantile regression, which will be used as a baseline model to assess the performance of our proposed algorithm. Examples of such models are provided in Refs. [7]–[11]. Similarly, Ref. [12] applies the linear version of quantile regression on the problem of probabilistic load forecasting.

As mentioned above, such multi-period samples of potential outcomes of electric loads, labelled scenarios here, are crucial to optimize the system operation over multiple future periods at once [13], and are especially important for problems considering electricity storage (e.g. via batteries or electric vehicles) [14]. However, quantile regression, as mentioned a popular technique for probabilistic load forecasting, does not allow to sample scenarios from the resulting prediction. This is the problem that we attempt to solve in this paper.

Due to the popularity of quantile regression, however, this paper still utilizes it as a baseline model. Specifically we present a Bayesian model and show how it can perform similar to quantile regression whilst allowing for sampling of individual scenarios.

Bayesian methods utilizing neural networks as non-linear

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288

approximators have been proposed in the literature [15]–[17]. However, beyond the issue of interpretability of neural networks, such models are inherently hard to train and are additionally sensitive to hyperparameters. Other probabilistic approaches for load forecasting such as Gaussian process regression suffer from similar issues [18], [19].

C. Contributions and Organization

Instead of such deep learning-based approaches instead we present a simple extension to traditional decision tree regression that allows for approximation of a parameterized distribution. This in turn allows to sample scenarios which can then be used in operational problems, which will be demonstrated by applying it on real-world data from a system utilized to test optimal battery control methods on.

In summary, we propose a simple framework that allows generating samples from a non-linear probabilistic forecast which does not require adjustments characteristic for deep-learning approaches such as using special loss functions to accurately model distributions or introducing additional hyperparameter sensitivity and difficulties in learning weights. Similarly, the framework could also be applied on deterministic neural network models, allowing probabilistic sampling from such models as well, providing a starting point for future research.

The paper is organized the following: In section II we introduce the problem analyzed in the paper. In section III we introduce decision trees and the comparison model we validate our proposed solution against. In section IV we introduce Bayesian regression in order to extend the previously defined comparison model. In section V we compare the models on a case study based on industry data. Finally, section VI concludes the paper.

II. The Problem

Consider a decision vector y where the entries y_t correspond to a single load value for a given time t, and an input matrix x where each column x_t corresponds to a vector of coefficients (e.g. day of the week, hour of the day, etc.). The regression problem can be formulated as finding the parameterized function f that solves the following approximation problem best:

$$y = f(x) + \xi \tag{1}$$

Here, the residuals ξ has to be minimized under some penalty functions by adjusting the parameters of function f, with $\xi = 0$ meaning a perfect fit of the approximation to the data.

For the specific application of load forecasting, there exist numerous techniques ranging from linear to non-linear models, with or without exogeneuous variables or auto-regressive components [5]. For the sake of simplicity, in this paper we will focus on traditional regression with purely exogeneuous variables as described in Eq. 1. However, the proposed approach can similarly applied on auto-regressive or ARX-style models (with latter being a mixture of both models), and are expected to perform well. Further we assume f to be non-linear, in line with the recent literature in load-forecasting, where the focus is mainly on non-linear regression models such as deep neural networks. For the sake of simplicity and to not encounter the above mentioned issues connected to hyperparameter sensitivity in probabilistic methods, we will utilize a decision tree regressor over neural networks. Nonetheless, future research could build on the here presented method and train probabilistic forecasts based on deterministic neural networks.

III. Probabilistic Decision Trees

A decision tree can be formulated as a mapping M that returns a given value for the specific region R (a subset constrained by boundaries) which the values of x are located in.

$$\begin{aligned} f(x) &= M(R) \\ x &\subseteq R \end{aligned} \tag{2}$$

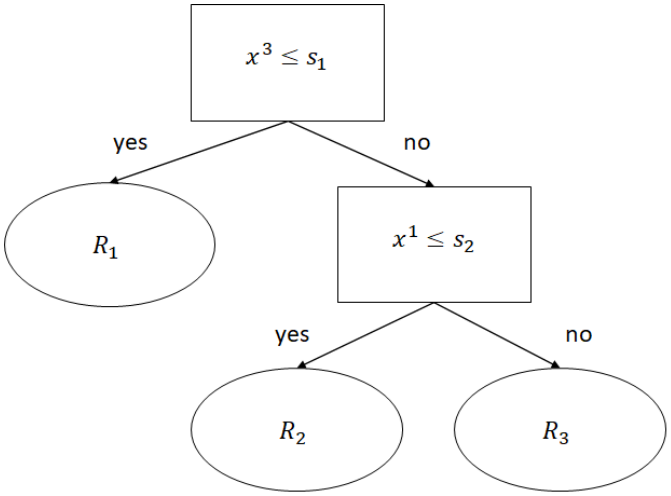

Fig. 1: Regression Tree Example

An example of such is provided by Fig. 1, which shows a regression tree consisting of three regions. x^j corresponds to the variable j of the input. In this example tree, there are (at least) three input variables: x^1, x^2, x^3. s_1 and s_2 are the branching values that decide which region (out of regions R_1, R_2, R_3) a specific data set x belongs to. In the given example tree, only the input variables x^3 and x^1 are used to define the regions, whereas the classification happens independent of value x^2.

This figure also highlights the main advantage of decision trees over neural networks as non-linear approximators, i.e. the interpretability of the results.

In literature several algorithms to yield decision trees that focus on dealing with large data sets can be found. In its essence, the process of deriving such decision trees can be described as in Ref. [20]. This principle is shown in Algorithm 1. Such a branching method starts with empty sets B^* and M^* and the input variables in set B. Then, the algorithm uses an optimization problem minimizing the distance of the real output point y to the mean of the respective region $\bar{y}_{R_1}$ or

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288

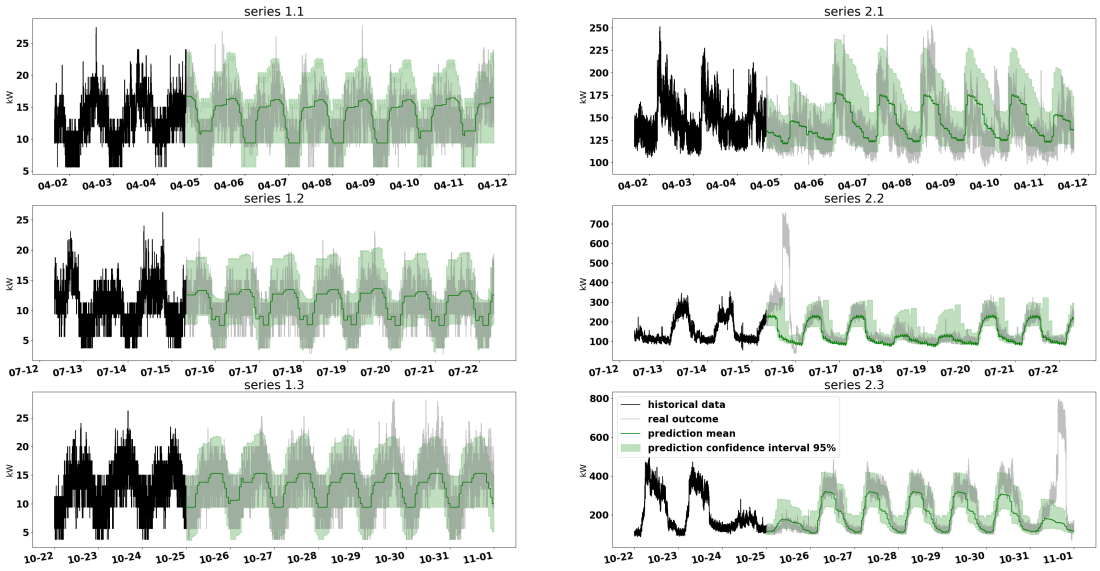

Fig. 2: Quantile Regression Results

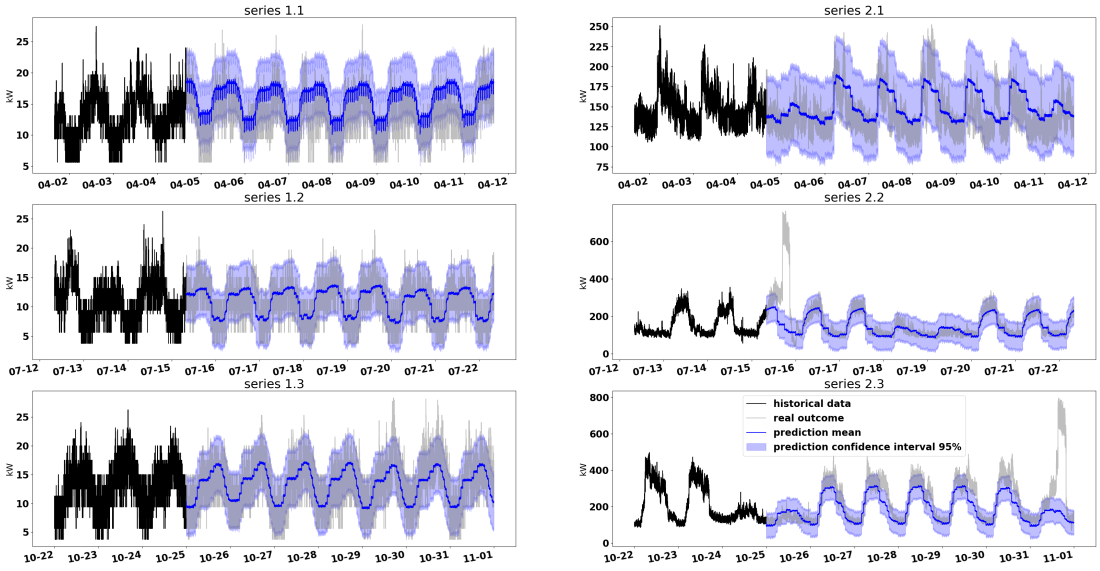

Fig. 3: Bayesian Regression Results

$\bar{y}_{R_2}$ that branches from the given point. This is done until all values are sufficiently branched, as shown in Fig. 1.

A more in-depth discussion on such tree based algorithms is presented in Ref. [21].

Independent of the algorithm, however, the result of this procedure is a tree B^* consisting of a number of regions $R \in B^*$ that are mapped via M to specific outcome predictions.

Minimizing the approximation loss requires the utilization of a loss function L. For traditional, deterministic regression, a popular loss function is the Mean Squared Error (MSE). Using y^a and y^b as the two different series that are to be evaluated against each other (e.g. y^a being real values and y^b being values generated by the tree, or vice versa), this loss

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288

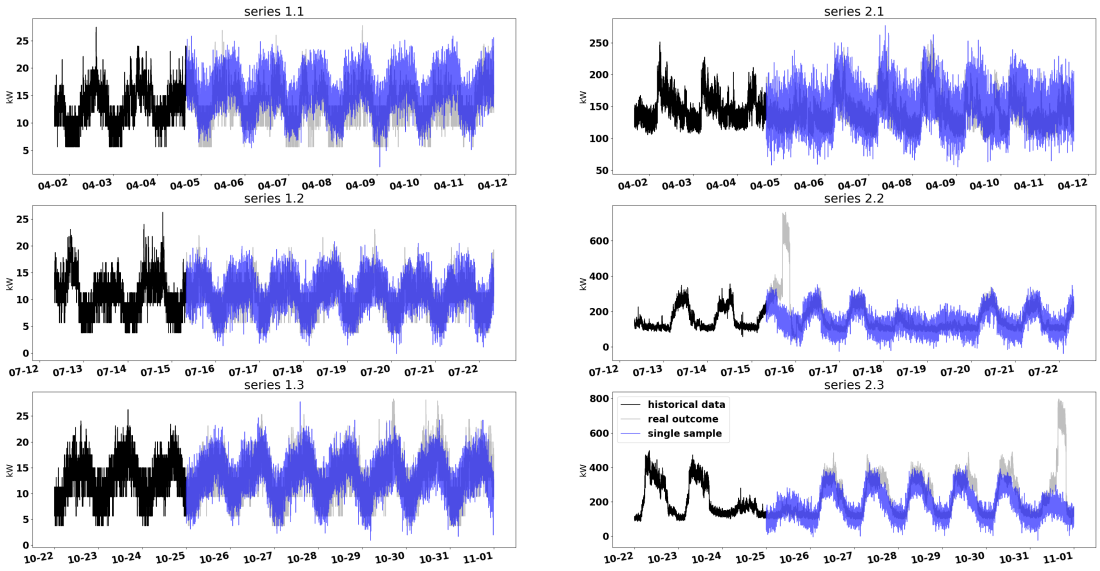

Fig. 4: Bayesian Regression - Single Sample

ALGORITHM 1
Decision Tree Regression for Tree B^*

initialize $B = \{x\}$, $B^* = \{\emptyset\}$, $M = \{\emptyset\}$
while $B \neq \{\emptyset\}$ **do**
  for all $b \in B$ **do**
    solve

$$\begin{aligned} \min_{s,j} \quad & L([y_t \forall t \in R_1], \bar{y}_{R_1}) + L([y_t \forall t \in R_2], \bar{y}_{R_2}) \\ \text{s.t.} \quad & R_1 = \{t | b_t^j \leq s\} \\ & R_2 = \{t | b_t^j > s\} \\ & \bar{y}_{R_1} = \text{mean}([y_t \forall t \in R_1]) \\ & \bar{y}_{R_2} = \text{mean}([y_t \forall t \in R_2]) \end{aligned} \tag{3}$$

    remove b from B
    for all $R = R_1, R_2$ **do**
      if $y_t = \bar{y}_R \forall t \in R$ **then**
        add R to B^*
        add mapping $M : R \rightarrow \bar{y}_R$
      else
        add R to B
      end if
    end for
  end for
end while

function can be described in the following:

$$L(y^a, y^b) = \mathbb{E}((y^a - y^b)^2) \tag{4}$$

The results of a tree trained via MSE will thus be a fit to the mean of the given data. However, in order to accurately model uncertainty, single point estimates are not sufficient. In load forecasting, a popular method to describe the uncertainty in future loads is via quantile regression. This method can be applied via adjusting the loss function.

Following the notation found in the introduction of Ref. [10] a quantile q^τ can be described the following:

$$P(f(x) \leq q^\tau) \leq \tau \tag{5}$$

where P labels a probability and $0 < \tau < 1$. The quantile loss function can then be formulated by selecting a specific τ:

$$\begin{aligned} & L_\tau(y^a, y^b) = \\ & \mathbb{E}\Big(\begin{cases} \tau(y^a - y^b) & \text{if } (y^a - y^b) \geq 0 \\ (1-\tau)(y^a - y^b) & \text{if } (y^a - y^b) < 0 \end{cases} \Big) \end{aligned} \tag{6}$$

A special case is given by $\tau = 0.5$ which then results in this quantile loss function returning the same results as using the mean squared error as a loss function.

However, fitting such a quantile regression model has a crucial downside in using it in operational models: in order to formulate the loss function, a specific numerical value for τ has to be selected. This means that the problem cannot be solved continuously, but instead the tree algorithm has to be conducted again for any desired change in τ.

Thus, albeit such quantile regression models provide useful tools to predict uncertainty in load problems, they do not provide useful tools to provide scenarios to operational problems. Ref. [22] goes further into detail on these stochastic optimization models and the role of scenarios in them.

We thus propose an extension of the deterministic tree model using the loss function from Eq. (4) that is able to

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288

approximate the distributions, specifically expressed via location and scale parameters of a Bayesian regression problem. We introduce this method next.

IV. Bayesian Regression

The Bayesian regression model can be formulated as finding the optimal parameters for the location function f and the scale given by standard deviation σ for a given distribution (here a Gaussian). This is done by sampling the error term ξ from this distribution:

$$\begin{aligned} & y = f(x) + \xi \\ \text{where} \quad & \xi \sim \mathcal{N}(0, \sigma^2) \end{aligned} \tag{7}$$

Assuming the regression problem yields optimal parameters for the function f ('optimal' as in minimizing the loss function from Eq. (4)) allows taking the residuals and yielding the minimum difference ξ as $y - f(x) = \xi$. The standard deviation can then be calculated via approximating it similarly via the mean squared error:

$$\sigma^2 = \mathbb{E}((y - f(x))^2) \tag{8}$$

This approximation is also referred to as the squared error risk.

In this specific case, however, Algorithm 1 results in a perfect fit with $L(y^a, y^b) = 0$. The additional risk to such a perfect fit on the test set is the risk of overfitting, which will here not be addressed and instead interested readers be referred to Ref. [20]. To circumvent the issue of not having a perfect fit of $y = f(x)$ and thus variance $\sigma^2 = 0$, the data y and coefficients x can instead be split into two sets y^1, x^1 and y^2, x^2. Thus, Algorithm 1 can be applied on y^1, x^1 to yield the parameters for f. Then, Eq. (8) can be solved via y^2, x^2 to yield the solution for $\sigma^2 = 0$. In the here provided case study, the data sets were split randomly into both sets.

In addition, a starting point for future research might be also provided in experimenting with using the same data set for both algorithms, but this was considered out of scope of the here presented work.

The non-linear Bayesian approach introduced here will be applied on real load data with its results being shown and compared to non-linear quantile regression in the next section.

V. Case Study

The utilized data set was two sets of (non-residential) load data from a commercial site in Norway managed by the company Lede (formerly Skagerak Nett) over the course of 11 months in 2020. The data series are from an office block (series 1) and a commercial sports facility (series 2), with latter showing a higher variance of the loads due to sports events requiring more electricity.

The original data set consisting of two time series was each split into three series of even length and predictions were made on each separate series individually. The length of the prediction was a single week with a minute resolution, i.e. 7×1440 data points. The requirement for the resolution came from the potential application in real-time systems.

TABLE I: Values outside of 95% Quantile

data	Bayesian regression	Quantile regression
series 1.1	0.218	0.024
series 1.2	0.051	0.059
series 1.3	0.125	0.077
series 2.1	0.056	0.241
series 2.2	0.058	0.109
series 2.3	0.218	0.162

TABLE II: Mean Error

data	Bayesian regression	Quantile regression
series 1.1	3.133	2.064
series 1.2	1.95	1.996
series 1.3	2.478	2.556
series 2.1	20.801	18.275
series 2.2	32.737	35.413
series 2.3	52.014	51.24

The results of the quantile decision tree regression as described above can be observed in Fig. 2. The results are shown within a confidence interval of 95%.

Similarly, the results of the Bayesian decision tree regression is shown in Fig. 3. The figure shows the 95% confidence interval for 1000 taken samples.

The ratio of values outside of the 95% intervals are given in Tab. I.

The errors for both of the algorithms is compared in Tab. II. Both data sets show comparable error values. However, and as mentioned in the method description above, the Bayesian approach has an advantage over the quantile regression method in that it allows for sampling.

This is demonstrated in Fig. 4 which shows a single scenario sample drawn from the Bayesian regression results. As discussed previously, being able to sample from the resulting distribution (here a Gaussian distribution parameterized with a non-linear function for its location) is an important feature in utilizing these load results in operational models. Examples of such operations are scheduling of charging and discharging of batteries or electrical vehicles.

The correlation coefficients of the models are shown in Tab. III. Similar to the error values in Tab. II and the visual results in Fig. 3, these results also indicate a good fit of the prediction.

Nonetheless, this approach has a single disadvantage over the quantile regression method. This can be observed by comparing the confidence intervals of Fig. 3 with that of Fig. 2. Since the Bayesian regression in Eq. (7) is formulated via normal distributions, these tails will be symmetrical. However, series 2 indicates this to not be the case but instead the distribution to be skewed towards lower values. In practice, this could

TABLE III: Correlation Coeffients

data	Bayesian regression	Quantile regression
series 1.1	0.602	0.607
series 1.2	0.63	0.62
series 1.3	0.626	0.608
series 2.1	0.531	0.537
series 2.2	0.54	0.498
series 2.3	0.659	0.628

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288

be solved via utilizing different distributions (such as Poisson for a skewed representation). This can be considered a starting point for future research on the topic. Nonetheless, and as discussed in this section, the current results still indicate good performance using Gaussian distributions as approximations.

VI. Conclusion

In this paper, we expand on the current literature on probabilistic load forecasting by proposing an extension via Bayesian regression to traditional non-linear methods. We show that a tree-based non-linear regression method can be utilized to incorporate uncertainty with no conceivable loss in accuracy compared to models using quantile loss functions.

The resulting model is simple to apply and less hyperparameter sensitive than other Bayesian methods such as Bayesian neural networks. We demonstrate the capabilities visually and quantitatively by using two heterogeneous series from a commercial site in Norway. Further, we discuss the strengths and weaknesses. In this process we find non-Gaussian distributions to be a viable starting point for future research.

References

[1] M. G. Shete and D. P. M. Joshi, "Practical implementation of load forecasting in smart grid: operational domain feature," *3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, p. 6, 2019.

[2] W. Xue, C. Li, X. Mao, X. Li, L. Zhao, and X. Zhao, "Medium and Long Term Load Forecasting of Regional Power Grid in the Context of Economic Transition," in *2018 2nd IEEE Conference on Energy Internet and Energy System Integration (EI2)*. Beijing: IEEE, Oct. 2018, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/8582055/

[3] K. Kwan and D. Maly, "Optimal battery energy storage system (BESS) charge scheduling with dynamic programming," *IEE Proceedings - Science, Measurement and Technology*, vol. 142, no. 6, pp. 453–458, Nov. 1995. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/ip-smt_19951929

[4] Y. Zhu, D. Zhao, X. Li, and D. Wang, "Control-Limited Adaptive Dynamic Programming for Multi-Battery Energy Storage Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4235–4244, Jul. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8409320/

[5] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169207015001508

[6] H. Mori and N. Kosemura, "Optimal regression tree based rule discovery for short-term load forecasting," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, vol. 2. Columbus, OH, USA: IEEE, 2001, pp. 421–426. [Online]. Available: http://ieeexplore.ieee.org/document/916878/

[7] Y. He, "Short-term power load probability density forecasting based on quantile regression neural network and triangle kernel function," *Energy*, vol. 114, pp. 498 – 512, 2016.

[8] D. Gan, Y. Wang, S. Yang, and C. Kang, "Embedding based quantile regression neural network for probabilistic load forecasting," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 244–254, Mar. 2018. [Online]. Available: http://link.springer.com/10.1007/s40565-018-0380-x

[9] T. Zhao, J. Wang, and Y. Zhang, "Day-Ahead Hierarchical Probabilistic Load Forecasting With Linear Quantile Regression and Empirical Copulas," *IEEE Access*, vol. 7, pp. 80 969–80 979, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8736321/

[10] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-Term Load Forecasting With Deep Residual Networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8372953/

[11] H. Aprillia, H.-T. Yang, and C.-M. Huang, "Statistical Load Forecasting Using Optimal Quantile Regression Random Forest and Risk Assessment Index," *IEEE Transactions on Smart Grid*, pp. –, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9240942/

[12] Y. Wang, D. Gan, N. Zhang, L. Xie, and C. Kang, "Feature selection for probabilistic load forecasting via sparse penalized quantile regression," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 1200–1209, Sep. 2019. [Online]. Available: http://link.springer.com/10.1007/s40565-019-0552-3

[13] D. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Massachusetts: Athena Scientific, 2012, vol. 2, no. 2.

[14] W. B. Powell and S. Meisel, "Tutorial on Stochastic Optimization in Energy—Part II: An Energy Storage Illustration," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1468–1475, Mar. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7100937/

[15] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1156–1166, May 2008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0196890407003032

[16] M. Bessani, J. A. Massignan, T. M. Santos, J. B. London, and C. D. Maciel, "Multiple households very short-term load forecasting using bayesian networks," *Electric Power Systems Research*, vol. 189, p. 106733, Dec. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0378779620305368

[17] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei, "Deep-Based Conditional Probability Density Function Forecasting of Residential Loads," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3646–3657, Jul. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8988175/

[18] M. Shepero, "Residential probabilistic load forecasting: A method using Gaussian process designed for electric load data," *Applied Energy*, p. 14, 2018.

[19] L.-L. Li, J. Sun, C.-H. Wang, Y.-T. Zhou, and K.-P. Lin, "Enhanced Gaussian process mixture model for short-term electric load forecasting," *Information Sciences*, vol. 477, pp. 386–398, Mar. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S002002551830865X

[20] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.

[21] W.-Y. Loh, "Fifty Years of Classification and Regression Trees," *International Statistical Review*, vol. 82, no. 3, pp. 329–348, Dec. 2014. [Online]. Available: http://doi.wiley.com/10.1111/insr.12016

[22] N. Löhndorf, "An empirical analysis of scenario generation methods for stochastic optimization," *European Journal of Operational Research*, vol. 255, no. 1, pp. 121–132, Nov. 2016. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0377221716303411

This is the accepted version of an article published in 2021 International Conference on Smart Energy Systems and Technologies - SEST
http://dx.doi.org/10.1109/SEST50973.2021.9543288