SPATIAL AND TEMPORAL PREDICTIONS FOR POSITIVE VECTORS

Omar Graja

A thesis
in
The Department
of
Concordia Institute for Information Systems Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Applied Science in Quality Systems Engineering
Concordia University
Montréal, Québec, Canada

December 2020

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Omar Graja**

Entitled: **Spatial and Temporal Predictions for positive vectors**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science in Quality Systems Engineering

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the final examining commitee:

Dr. Jia Yuan Yu _____ Chair and Internal Examiner

Dr. Jamal Bentahar _____ Examiner

Dr. Ciprian Alecsandru _____ Examiner

Dr. Nizar Bouguila _____ Supervisor

Approved _____
Dr. Mohammad Mannan, Graduate Program Director

_____ 2020 _____ _____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Spatial and Temporal Predictions for positive vectors

Omar Graja

Predicting a given pixel from surrounding neighboring pixels is of great interest for several image processing tasks. To model images, many researchers use Gaussian distributions. However, some data are obviously non-Gaussian, such as the image clutter and texture. In such cases, predictors are hard to derive and to obtain. In this thesis, we analytically derive a new non-linear predictor based on inverted Dirichlet mixture. The non-linear combination of the neighbouring pixels and the combination of the mixture parameters demonstrate a good efficiency in predicting pixels. In order to prove the efficacy of our predictor, we use two challenging tasks, which are; object detection and image restoration.

We also develop a pixel prediction framework based on a finite generalized inverted Dirichlet (GID) mixture model that has proven its efficiency in several machine learning applications. We propose a GID optimal predictor, and we learn its parameters using a likelihood-based approach combined with the Newton-Raphson method. We demonstrate the efficiency of our proposed approach through a challenging application, namely image inpainting, and we compare the experimental results with related-work methods.

Finally, we build a new time series state space model based on inverted Dirichlet distribution. We use the power steady modeling approach and we derive an analytical expression of the model latent variable using the maximum a posteriori technique. We also approximate the predictive density using local variational inference, and we validate our model on the electricity consumption time series dataset of Germany. A comparison with the Generalized Dirichlet state space model is conducted, and the results demonstrate the merits of our approach in modeling continuous positive vectors.

Acknowledgments

I would like to extend my most sincere appreciation and gratitude to my supervisor, Professor Nizar Bouguila, who encouraged and supported me genuinely on my entire journey in this Master's program. As a knowledgeable, respectful and genius supervisor, he always led and motivated me with endless patience. His guidance and bits of advice in the personal, academic and professional scale will remain engraved in my thoughts forever. I have been fortunate to have worked with one of the best and most prestigious and talented supervisors at Concordia. I can't thank you enough!

I also want to express my sincere gratitude to Fatma Najar for her endless support and her high sense of determination in this splendid journey and for her fun and fruitful company.

I want to thank my colleagues, Dr. Mohamad Azam and, Rima Nasfi for their support and motivation!

I owe my most massive thanks to all my lab mates; Nuha Zamzami, Oumayma, Narges , Kamal, Pantea, Basim, Shuai, Hieu, Sunny, Jaspreet, Waleed, Eddy, Hafsa, Xavier, Hussain and Meeta for the great time I spent with them during this adventure.

Last but not least I am deeply grateful to my parents who supported me to achieve and accomplish all my dreams and ambitions and I feel very blessed to be their son. I want to express my heartfelt thanks to my brother in law Hassen karaa and my two sisters Dr. Imen Graja and Ameni Graja who guided and inspired me unconditionally.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Pixel prediction has shown to be one of the most needed tools to perform several applications in image processing such as anomaly detection [1, 2], object detection [3, 4], edge detection [5], video compression [6, 7], semantic segmentation [8, 9], image restoration [10, 11] and keypoint prediction [12]. Meanwhile, pixel prediction is represented by approximating the predicted pixel using its neighbors. It is usually represented by a linear or a non-linear combination of the neighboring pixels plus an error value [13, 14]. In our second chapter, inspired by the work proposed in [15], we take the optimal predictor of x_n as the conditional expectation $E\big(x_{ij} \mid \forall x_{i'j'} \in N_{i,j}\big)$, where $x_{i'j'}$ are the neighbors of x_{ij} within the set of pixels $N_{i,j}$.

Exploiting the ease of analytical derivations, the authors in [16] derived optimal predictors for Gaussian distribution and a mixture of Gaussians. However, the field of non-Gaussian distributions exhibited an exciting expansion in terms of mathematical theorems [17, 18]. Indeed, a large number of researchers proved that Gaussian assumption is generally inappropriate and other alternative distributions are more effective in modeling data than Gaussian distribution by unveiling more appropriate patterns and correlations among data features [19, 20, 21, 22]. However, the image clutter and texture are usually non Gaussian. Thus, to handle the problem cited above, we propose to model an image with a mixture of inverted Dirichlet distributions that has shown outperforming results compared to the Gaussian mixture in modelling positive vectors [23]. This is due to the fact that the inverted Dirichlet

distribution is more scalable and flexible, compared to the Gaussian, and presents different symmetric and asymmetric modes.

In the second work, we have used the same approach as our first publication to create the Generalized inverted Dirichlet optimum predictor [24]. We proved that finite inverted Dirichlet mixture model effectively represents positive vectors [25, 26, 27]. However, it suffers from significant drawbacks such as its minimal, strictly positive covariance structure. The generalized inverted Dirichlet which is part of the Liouville family of distributions will be applied to overcome this limitation [28]. This distribution provides a more decent representation of the variability of the data [29]. Indeed, considering the fact that generalized inverted Dirichlet could be factorized into a set of inverted Beta distributions [30], gives more flexibility for modeling data in real-world applications. Therefore, we derive a novel optimal predictor based on generalized inverted Dirichlet distribution which results in a linear combination of the neighboring pixels. Meanwhile, we evaluate the proposed approach on image inpainting application. We choose a publicly available dataset namely Paris StreetView to validate our approach [31]. For the purpose of proving the efficiency of our proposed optimal predictor, we consider two types of pixel discarding. The first pixel removal is random, whereas, in the second experiment, we discard lines from the image. We use a 3rd order non-symmetrical half-plane casual (NSHP) neighborhood system to compute the missing pixel [10]. Then, we perform two image comparison metrics to evaluate our proposed model and compare it to other similar optimal based predictors.

Finally, our optimal predictors can be applied within a wide range of neighboring systems. Hence, we derive an analytical expression for the optimal predictor from a mixture of inverted Dirichlet distributions and Generalized inverted Dirichlet that is a linear combination of the neighboring pixels. To test the performance of our predictor, we conduct two challenging tasks, which are object detection and image pixel restoration for the Inverted Dirichlet model and image inpainting for the Generalized inverted Dirichlet model.

The third work is about time series forecasting. A novel state space model based on Inverted Dirichlet power state model is proposed. The main purpose of the state space model (SSM) is to infer the relation between the relevant properties of a series of such observations. Other well known targets of SSM are forecasting, smoothing , and parameters estimation. Indeed, state space models are known by their efficiency

in forecasting the missing data of multivariate dynamic systems with deterministic inputs. However, the challenging side when building the model is the optimization and inference of parameters. State space models are tractable and by the introduction of the latent variable we can decouple past from future observations. The generally known state space models include autoregressive moving average (ARMA) time series, autoregressive integrated moving average (ARIMA) models [32], exponential smoothing (ES) [33], and Kalman filters [34]. The state space formulation for ARMA and ARIMA models are formulated as stationary time series where the characteristic properties remains unchanged through time changing. However, the simple form of these models does not provide an explicit information about trend and periodicity as the structural time series approaches [35]. In the same line, the exponential smoothing approach is proposed basically for forecasting purpose. ES models with a simple structure are able to predict the future time series given univariate time series in one step. Though these approaches are practical for forecasting times series in various fields, they are based on linear Gaussian model for the estimation of parameters. When we move to a more complex times series, linear state space models are not adequate for forecasting and smoothing. In this regard, Grunwald et al. [36] proposed a new approach for forecasting time series where the observations are supposed to follow a Dirichlet distribution. Given the limitations of the Dirichlet distribution related to the structure of the covariance and the correlation of the variables, the generalized Dirichlet was proposed also for the state space model in [37]. The generalized Dirichlet power steady model (GDPSM) [37] presents a new forecasting model for compositional time series data where the authors applied the new GDPSM to web service selection application. Indeed, considering applications with positive vectors, it becomes of great interest to consider an adequate distribution. For this purpose, we address in this paper the case of time series of positive vectors where we consider the inverted Dirichlet distribution for modeling the observations.

1.2 Contributions

In this thesis, we derive two optimal predictors based on Inverted Dirichlet and Generalized Inverted Dirichlet. We also develop a methodology for modeling, forecasting,

and estimating times series of continuous positive vectors assumed to follow an inverted Dirichlet distribution. We summarize our contributions as the following:

1. **Finite Inverted Dirichlet Mixture Optimal Pixel Predictor**:

   We analytically derive a new non-linear predictor based on inverted Dirichlet mixture. The non-linear combination of the neighbouring pixels and the combination of the mixture parameters demonstrate a good efficiency in predicting pixels. In order to prove the efficacy of our predictor, we use two challenging tasks, which are; object detection and image restoration.

   This work has been published in the 7th IEEE Global Conference on Signal and Information Processing (GlobalSIP) [38].

2. **Generalized Inverted Dirichlet optimal predictor for image pixel prediction**:

   In this work, we develop a pixel prediction framework based on a finite generalized inverted Dirichlet (GID) mixture model that has proven its efficiency in several machine learning applications. We propose a GID optimal predictor, and we learn its parameters using a likelihood-based approach combined with the Newton-Raphson method. We demonstrate the efficiency of our proposed approach through a challenging application, namely image inpainting, and we compare the experimental results with related-work methods.

   This work is going to be published in the 15th International Symposium on Visual Computing (ISVC 2020).

3. **Inverted Dirichlet power steady model for continuous positive time series forecasting**:

   We propose an inverted Dirichlet power steady state space model. While building the model, we introduce a new maximum a posteriori estimation of Inverted Dirichlet state space model's parameters and we estimate the prediction density by local variational inference. We validate our model on time series application by forecasting the electricity consumption.

1.3 Thesis Overview

- In chapter 1, we introduce the optimal predictors that we derived and the state space time series model. It includes a brief overview of various concepts related to the work. It also includes the motivations behind the conducted research work.

- In chapter 2, we explain in detail the derivation of the optimal predictor finite inverted Dirichlet based distribution. We perform two different applications, object detection and image restoration to prove the efficiency of the proposed model.

- In chapter 3, we explain again the derivation of the generalized inverted Dirichlet optimal predictor and we perform image inpainting application to demonstrate the performance of the proposed method. We explain the outstanding performance of the model and we compare it to other related models including the inverted Dirichlet optimal predictor.

- In chapter 4, we explain the building of the inverted Dirichlet state space power steady model in details. Maximum a posteriori was applied to estimate the latent variable of the model. We also apply local variational inference to approximate the predictive density. We perform different simulations on the electricity consumption time series dataset of Germany and we validate the performance of our proposed approach by averaging the results over the multiple simulations.

- In conclusion, we briefly summarize our contributions.

Chapter 2

Finite Inverted Dirichlet Mixture Optimal Pixel Predictor

In this chapter, we detail the derivation of the inverted Dirichlet optimal predictor. We demonstrate our model and learning algorithm with some real applications for image processing. A comparisons with comparable recent approaches have shown the worth of our proposed model.

2.1 Model specification

In this section, we first present the inverted Dirichlet mixture model. A significant step that we discuss next is the learning method that is based on a maximum likelihood estimation, and we present analytical derivations of the optimal predictor.

2.1.1 Parameter learning method

Let us consider that we have N independent identically distributed vectors generated from Inverted Dirichlet distribution $\mathcal{X} = \left(\vec{X}_1, \ldots, \vec{X}_N\right)$. Each sample is a D-dimensional vector $\vec{X}_n = \left(x_{n1}, \ldots, x_{nD}\right)$ having a probability density function given by:

$$ID(\vec{X}_n \mid \vec{a}_k) = \frac{\Gamma(\mid \vec{a}_k \mid)}{\prod_{d=1}^{D+1} \Gamma(a_{kd})} \prod_{d=1}^{D} x_{nd}^{a_{kd}-1} (1 + \sum_{d=1}^{D} x_{nd})^{-\sum_{d=1}^{D+1} a_{kd}} \quad (1)$$

where $x_{nd} > 0$ for $d = 1, ..., D$. The shape parameters are $\vec{a}_k = (a_{k1}, \ldots, a_{kD+1}), a_{kd} > 0$ for $d = 1, ..., D+1$ and $\mid \vec{a}_k \mid = \sum_{d=1}^{D+1} a_{kd}$.

Considering $\mathcal{X}$ as the composition of K different clusters, we can model it by a

finite Inverted Dirichlet mixture model, where the weight of component k is denoted by p_k, with the constraints $p_k \geq 0$ and $\sum_{k=1}^{K} p_k = 1$. We note that $\Theta_k = \{p_k, \vec{a}_k\}$, for $k = 1, \ldots, K$, represents the set of weight and shape parameters of component k and the complete model parameters are denoted by $\Theta = \{p_1, \ldots, p_k, \vec{a}_1\ , \ldots, \vec{a}_k\}$.

$$p\Big(\vec{X}_n|\Theta\Big) = \sum_{k=1}^{K} p_k ID(\vec{X}_n|\vec{a}_k) \tag{2}$$

where the ID is the inverted Dirichlet distribution.

In mixture models, a latent indicator vector $\vec{Z}$ is commonly introduced to represent the cluster to which the vector $\vec{X}_n$ is assigned. Mathematically, $\vec{Z}_n = \big(Z_{n1}, \ldots, Z_{nK}\big)$ where Z_{nk} is 1 if the expectation that $\vec{X}_n$ belongs to the component k is higher than the other components and 0, elsewhere.

We have the following complete log-likelihood:

$$L(\Theta, Z, \mathcal{X}) = \sum_{k=1}^{K}\sum_{n=1}^{N} Z_{nk}\Big(\log p_k + \log p(\vec{X}_n|\vec{a}_k)\Big) \tag{3}$$

To estimate the mixture parameter, the gradient of the log-likelihood has to be maximized. To solve this optimization problem, we need to find a solution of the following equation:

$$\frac{\partial \log L(\Theta, Z, \mathcal{X})}{\partial \Theta} = 0 \tag{4}$$

After calculating the first and second derivatives of the complete log likelihood, we do not reach a closed-form solution. Therefore, we solve it using an iterative approach namely Newton-Raphson method, that is defined as follows:

$$\hat{a_k}^{new} = \hat{a_k}^{old} - {H_k}^{-1} G_k \tag{5}$$

where G_k is the first derivatives vector with respect to a_{kd}, $k = 1, \ldots, K$, and H_k is a Hessian matrix that is expressed as follow:

$H_k = \sum_{n=1}^{N} p\big(k \mid \vec{X}_n, \vec{a}_k\big) \times$

$$\begin{pmatrix} \Psi'(\mid \vec{a}_k \mid) - \Psi'(\vec{a}_{k1}) & \cdots & \Psi'(\mid \vec{a}_k \mid) \\ \Psi'(\mid \vec{a}_k \mid) & \Psi'(\mid \vec{a}_k \mid) - \Psi'(\vec{a}_{k2}) & \Psi'(\mid \vec{a}_k \mid) \\ \vdots & \vdots & \vdots \\ \Psi'(\mid \vec{a}_k \mid) & \cdots & \Psi'(\mid \vec{a}_k \mid) - \Psi'(\vec{a}_{kd+1}) \end{pmatrix} \tag{6}$$

Algorithm 1 Initialization algorithm

1: Apply the k-means on the N D-dimensional vectors to obtain initial M clusters.
2: Calculate $p_k = \frac{\text{number of elements in class}}{N}$
3: Apply the moments method for each component k to obtain a vector of parameters $\vec{\alpha_k}$ and $\vec{\beta_k}$ using Eq (1) and Eq (2).

$\Psi(.)$ and $\Psi'(.)$ denote the digamma and trigamma functions, respectively. Those functions are defined as follow:

$$\Psi(X) = \frac{\Gamma'(X)}{\Gamma(X)}, \Psi'(X) = \frac{\Gamma''(X)}{\Gamma(X)} - \frac{\Gamma'(X)^2}{\Gamma(X)^2} \tag{7}$$

We initiate our learning method by using the method of moments defined in [39] by:

$$a_{kD+1} = \frac{E(\vec{X_d})^2 + E(\vec{X_d})}{Var(\vec{X_d})} + 2 \tag{8}$$

$$a_{kd} = E(\vec{X_d})(a_{kd+1} - 1) \tag{9}$$

Where $\vec{X_d}$ is a D dimensional sample of observations.

Finally, the estimated values of mixing proportions are expressed, in the following, as it has a closed-form solution:

$$p_k = \frac{\sum_{n=1}^{N} p(k|\vec{X_n}, \vec{a_k})}{N} \tag{10}$$

The initialization and estimation steps are define by algorithm 1 and algorithm 2 respectively.

2.1.2 Optimal predictor

As defined in the introduction, the optimal predictor is the conditional expected value of a pixel given its neighbors:

$$\hat{x}_n = E\big(x_n \mid x_{n-1}, x_{n-2}, \ldots, x_1\big) \tag{11}$$

We suppose that:

$$x_{_n} \sim ID(a_{n+1}, a_n, \ldots, a1)$$

Algorithm 2 Estimation algorithm

1: INPUT: D-dimensional data $\vec{X}_n$, $n = 1, \ldots, N$ and a specified number of clusters K.
2: Initialization algorithm 1
3: E-step: Compute the posterior probability $p(k|\vec{X}_n, \vec{a_k})$.
4: Maximization step:
   - ☐ Update the $\vec{a_k}$ Eq (5), $k = 1, \ldots, K$.
   - ☐ Update the p_k using Eq (10), $k = 1, \ldots, K$.
5: If $p_k < \epsilon$, discard component k and go to 3.
6: If the convergence criterion passes terminate,
   else go to 3.

As given in [40] (see page 179), we note that any marginal distribution of a subvector of the inverted Dirichlet follows an inverted Dirichlet distribution:

$$x'_n = \frac{x_n}{1 + \sum_{j=1}^{n-1} x_j} \tag{12}$$

$$\left(x'_n \mid x_1, x_2, \ldots, x_{n-1}\right) \sim ID(a_n, \sum_{j=1}^{n-1} a_j + a_{n+1}) \tag{13}$$

Knowing that:

$$E(x_i) = \frac{a_i}{a_{n+1} - 1}, a_{n+1} > 1 \tag{14}$$

Thus, we obtain the following equation:

$$E\left(x'_n \mid x_1, \ldots, x_{n-1}\right) = \frac{a_n}{\sum_{k=1}^{n-1} a_k + a_{n+1} - 1} \tag{15}$$

Based on the equations given by Eq (12) and Eq (14), we conclude that:

$$E\left(x_n \mid x_1, \ldots, x_{n-1}\right) = \left(1 + \sum_{j=1}^{n-1} x_j\right) \frac{a_n}{\sum_{j=1}^{n-1} a_j + a_{n+1} - 1} \tag{16}$$

Since:

$$E\left(x_n \mid x_{n-1}, \ldots, x_1\right) = \int x_n p(x_n \mid x_{n-1}, \ldots, x_1)\, \mathrm{d}x_n \tag{17}$$

Then, following the proposal given in [41]:

$$E\big(x_n \mid x_{n-1}, \ldots, x_1\big) = \tag{18}$$
$$\sum_{k=1}^{K} p'_k \int x_n p_k(x_n \mid x_1, \ldots, x_{n-1})\, \mathrm{d}x_n =$$
$$\sum_{k=1}^{K} p'_k E_k\big(x_n \mid x_1, \ldots, x_{n-1}\big)$$

where

$$p'_k = p_k \frac{\int p_k(x_1, \ldots, x_n)\, \mathrm{d}x_n}{\int p(x_1, \ldots, x_n)\, \mathrm{d}x_n} \tag{19}$$

$$p'_k = p_k \frac{p_k(x_1, \ldots, x_{n-1})}{\sum_{k=1}^{K} p_k p_k(x_1, \ldots, x_{n-1})} \tag{20}$$

Finally, our optimal predictor is resumed in the following linear combination of x_n neighbours:

$$\hat{x}_n = \sum_{k=1}^{K} p'_k \left(1 + \sum_{j=1}^{n-1} x_j\right) \frac{a_n}{\sum_{j=1}^{n-1} a_j + a_{n+1} - 1} \tag{21}$$

2.2 Experimental results

In order to evaluate the efficiency of the proposed technique, we consider two challenging applications, object detection and image restoration. Then, we compare the model to the widely used Gaussian mixture [41] and to the generalized Dirichlet [42] predictors.

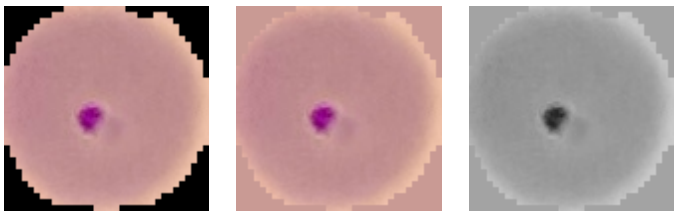
We consider a 2^{nd} order nonsymmetrical half-plane casual neighborhood system [43], which has shown outstanding capabilities in image pixel prediction [42]. In fact, we use two defined algorithms stated in 2.1.1 to generate the mixture parameters. Then, we apply the analytical formula, given by Eq (49), to predict the image pixel. Next, we perform object detection and image restoration.

2.2.1 Object detection

We use the Malaria cell images [1] for object detection. It contains two groups of cells' image infected (see Fig 1), and Uninfected ones (see Fig 2). Our main goal is to

[1] https://ceb.nlm.nih.gov/repositories/malaria-datasets/

detect the small purple dot inside the cell using three different models and compare their performances.



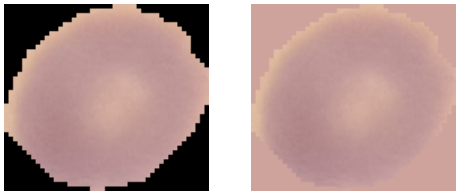
(a) Original (b) Preprocessed (c) Grayscaled

Figure 1: Parasitized image

First, we start by performing the data preprocessing, in which we change zero values of the image by the mean value of non-null pixel values, as given by Fig 1b and Fig 2b. Then, we convert the image to grayscale, as shown in Fig 1c . Next, we apply algorithms mentioned in section 2.1.1 to estimate the model parameters. Furthermore, a prediction error image is generated by computing the difference between the original and the predicted pixels values. After that, we set a threshold to detect the anomaly in the image. We have also used the signal-to-background ratio (LSBR) of a single object as reported in [42] to compare all models' performance, and it is defined as:

$$LSBR = \frac{\sum_{i,j \in object}(e_{i,j} - \mu_n) \div s}{\sigma_n^2} \tag{22}$$

where $e_{i,j}$ is the prediction error, μ_n and σ_n are the mean and standard deviation of the prediction error in the background and s is the object size.



(a) Original (b) Preprocessed

Figure 2: Uninfected image

We compare our technique with predictors generated from Gaussian and generalized Dirichlet mixture models. We take the average value of 30 different infected images and results are show in table 1.

According to table 1, it is clear that Inverted Dirichlet mixture model performs significantly better than the two other models. In fact, this proves that inverted

Table 1: Different models' LSBR for Malaria detection.

Models	LSBR
Inverted Dirichlet mixture	15.03
Generalized Dirichlet mixture	6.02
Gaussian mixture	5.14

(a) Original (b) Noised (c) GMM (d) GDM (e) IDM

(f) Original (g) Noised (h) GMM (i) GDM (j) IDM

Figure 3: Image restoration using three different models.

Dirichlet mixture is more efficient than other distributions in modelling positive, semi bounded vectors.

2.2.2 Image restoration

According to existing studies on image restoration, images can be improved but not completely restored that is why solving image restoration problems has been subject of several research works [44].

We consider 10 different images from UC Berkeley BSDS500 dataset [45]. First, we blur them by adding a white noise to create our test set. Then, we use the structural similarity index (SSIM) stated in [46] to measure the similarities between the original images and the restored ones. In fact, image restoration is done by predicting pixels from their neighbours using Eq (50).

Table 2: SSIM average for the different tested models.

Models	SSIM
Inverted Dirichlet mixture (IDM)	0.6524
Generalized Dirichlet mixture (GDM)	0.5242
Gaussian mixture model (GMM)	0.5312

Based on Fig 3 and Table 2, our proposed method reaches 65.24% which outperforms other comparable models based on Gaussian and generalized Dirichlet mixture models.

Chapter 3

Generalized Inverted Dirichlet Optimal Predictor for Image Inpainting

In this chapter, we detail the derivation of the generalized inverted Dirichlet optimal predictor. We illustrate the model and learning algorithm and we perform image inpainting application to validate the performance of our proposed method. We compare our model to other related models including our inverted Dirichlet optimal predictor and we discuss the merits of using generalized inverted Dirichlet over inverted Dirichlet.

3.1 GID prediction model

The generalized inverted Dirichlet mixture model has shown high flexibility for modeling and clustering positive vectors. In this section, we start by reviewing the finite GID mixture model. Then, we introduce the parameters learning through the EM algorithm and, later, we extend this model to the prediction.

3.1.1 Mixture of generalized inverted Dirichlet distributions

Let $\mathcal{X} = (\vec{X}_1, \ldots, \vec{X}_N)$ be a set of N d-dimensional positive vectors where each vector $\vec{X}_i$ follows a mixture of K generalized inverted Dirichlet (GID) distributions

characterized by parameters $\vec{\theta}_j = (\vec{\alpha}_j, \vec{\beta}_j)$ and mixing weight π_j of the jth component.

$$P(\vec{X}_i|\Theta) = \sum_{j=1}^{K} \pi_j P(\vec{X}_i|\vec{\theta}_j) \tag{23}$$

where $\Theta = (\vec{\theta}_1, \ldots, \vec{\theta}_K, \pi_1, \ldots, \pi_K)$ represents the GID mixture model parameters and $P(\vec{X}_i|\vec{\theta}_j)$ is the generalized inverted Dirichlet distribution, which has the following form [47]:

$$P(\vec{X}_i|\vec{\theta}_j) = \prod_{l=1}^{d} \frac{\Gamma(\alpha_{jl}+\beta_{jl})}{\Gamma(\alpha_{jl})\Gamma(\beta_{jl})} \frac{X_j^{\alpha_{jl}-1}}{(1+\sum_{s=1}^{l} X_{is})^{\gamma_{jl}}} \tag{24}$$

where $\gamma_{jl} = \beta_{jl} + \alpha_{jl} - \beta_{jl+1}$, for $l = 1, \ldots, d$ ($\beta_{jd+1} = 1$). It is to be noted that the GID is reduced to the inverted Dirichlet distribution when the parameter γ_{jl} is set to zero ($\gamma_{j1} = \cdots = \gamma_{jd} = 0$).

The flexibility of the generalized inverted Dirichlet is by dint of the concept of "Force of mortality" of the population where we introduce, here, a doubly non-central Y independent-variables defined as

$$Y_{i1} = 1, \quad Y_{jl} = \frac{X_{il}}{T_{il-1}}, \quad l > 1 \tag{25}$$

where $T_{il} = 1 + X_{i1} + X_{i2} + \cdots + X_{il-1}, \quad l = 1, \ldots, d$

The characteristic function underlying the $\mathcal{Y} = (\vec{Y}_1, \ldots, \vec{Y}_N)$ independent variables follows a product of 2-parameters inverted Beta distribution, where $\theta_l = (\alpha_l, \beta_l)$

$$P(\vec{Y}_i|\vec{\theta}) = \prod_{l=1}^{d} P_{IBeta}(Y_{il}|\theta_l) \tag{26}$$

In which the probability of inverted Beta is given by:

$$P_{IBeta}(Y_{il}|\theta_l) = \frac{\Gamma(\alpha_l+\beta_l)}{\Gamma(\alpha_l)\Gamma(\beta_l)} \frac{Y_{il}^{\alpha_l}}{(1+Y_{il})^{\alpha_l+\beta_l}} \tag{27}$$

Many characteristics of the distribution are defined in [48]. We mention some interesting statistics for this distribution.

First, the mixed moments such as the n^{th} moment is given by:

$$E(Y^n) = \frac{\Gamma(\alpha+\beta)\Gamma(\beta-n)}{\Gamma(\alpha)\Gamma(\beta)} \tag{28}$$

where $\beta - n$ is positive.

Then, the covariance between two variable Y_1 and Y_2 is defined as:

$$COV(Y_1, Y_2) = \frac{(\alpha_1 - 1)(\alpha_2 - 1)}{(\alpha_1\alpha_2 - 1)(\beta_1 - 1)(\beta_2 - 1)} \tag{29}$$

The covariance between two features for inverted Beta is always positive, which means that both they tend to increase or decrease together.

Finally, the variance of a variable Y is conveyed by:

$$VAR(Y) = \frac{(\alpha - 1)(\alpha + \beta - 2)}{(\beta - 1)^2(\beta - 2)} \tag{30}$$

3.1.2 Likelihood-based learning

Theoretically, a plethora of literature agrees on the effectiveness of the likelihood-based approach for estimating the mixture parameters. One of the well-known methodologies is the Expectation-Maximization technique [49], beginning with a tuned initialization for the set of parameters to the expectation step where the posterior is inferred (named often as "responsibilities"), then the iterations are proceeded to update the required variables until convergence. The heart of the matter comes with estimating the parameters based on the second derivative of the log-likelihood function with regards to each parameter. First, we introduce the log-likelihood as follows:

$$\log P(\mathcal{Y}|\Theta) = \sum_{i=1}^{N} \log \Big[\sum_{j=1}^{K} \pi_j \prod_{l=1}^{d} P_{IBeta}(Y_{il}|\theta_{jl}) \Big] \tag{31}$$

Initializing process

As a first step, an unsupervised-method, namely "K-means," is applied to obtain the initial K clusters. Consequently, for each cluster, the method-of-moments is implemented to get the initial $\vec{\alpha}_j$ and $\vec{\beta}_j$ parameters of each component j. The mixing weight is set in the initial phase as the number of elements in each cluster divided by the total number of vectors. As mentioned earlier, with conditionally independent features, the GID is converted by the inverted Beta distribution factorization. Thus, given the moments of inverted Beta distribution [50], the initial α_{jl_0} and β_{jl_0} are deduced by

$$\alpha_{jl_0} = \frac{\mu_{jl}^2(1+\mu_{jl}) + \mu_{jl}\sigma_{jl}^2}{\sigma_{jl}^2} \tag{32}$$

$$\beta_{jl_0} = \frac{\mu_{jl}(1+\mu_{jl}) + 2\sigma_{jl}^2}{\sigma_{jl}^2} \tag{33}$$

where μ_{jl} is the mean and σ_{jl} is the standard-deviation, $j = 1, \ldots, K,\ l = 1, \ldots, D$

Expecting the responsibilities

The responsibilities or posterior probabilities play an essential role in the likelihood-based estimation technique. It affects the update of the parameters in the next following step using the current parameter value.

$$P(j|\vec{Y_i}) = \frac{\pi_j P(\vec{Y_i}|\vec{\theta_j})}{\sum_{m=1}^{K} \pi_m P(\vec{Y_i}|\vec{\theta_m})} \tag{34}$$

Maximizing & upgrading the GID parameters

At the beginning, we set the gradient of log-likelihood function with respect to the mixing weight parameter equals to zero:

$$\frac{\partial \log P(\mathcal{Y}, \Theta)}{\partial \pi_j} = 0 \tag{35}$$

Then, we obtain the update formula for π_j, for $j = 1, \ldots, K$ as

$$\pi_j = \frac{1}{N} \sum_{i=1}^{N} P(j|\vec{Y_i}) \tag{36}$$

where $P(j|\vec{Y_i})$ is the posterior computed in the E-step.

To learn the parameters $\vec{\alpha}_j$ and $\vec{\beta}_j$, the Fisher scoring algorithm [51] is used. Thus, we need to calculate the first and the second derivatives of the log-likelihood function based on the following update [52]:

$$\alpha_{jl}^{t+1} = \alpha_{jl}^{t} - \Big(\frac{\partial^2}{\partial \alpha_{jl}^2} \log P(\mathcal{Y}, \Theta)\Big)^{-1} \times \frac{\partial}{\partial \alpha_{jl}} \log P(\mathcal{Y}, \Theta) \tag{37}$$

$$\beta_{jl}^{t+1} = \beta_{jl}^{t} - \Big(\frac{\partial^2}{\partial \beta_{jl}^2} \log P(\mathcal{Y}, \Theta)\Big)^{-1} \times \frac{\partial}{\partial \beta_{jl}} \log P(\mathcal{Y}, \Theta) \tag{38}$$

The first derivatives of $\log P(\mathcal{Y}, \Theta)$ are given by:

$$\begin{aligned}\frac{\partial}{\partial \alpha_{jl}} \log P(\mathcal{Y}, \Theta) &= \sum_{i=1}^{N} P(j|\vec{Y_i})\Big(P_{IBeta}(Y_{il}|\theta_{jl})[\Psi(\alpha_{jl}+\beta_{jl}) - \Psi(\alpha_{jl}) + \log Y_{il} \\ &- \log(1+Y_{il})]\Big), \end{aligned} \tag{39}$$

$$\frac{\partial}{\partial \beta_{jl}} \log P(\mathcal{Y}, \Theta) = \sum_{i=1}^{N} P(j|\vec{Y_i})\Big(P_{IBeta}(Y_{il}|\theta_{jl})[\Psi(\alpha_{jl}+\beta_{jl}) - \Psi(\alpha_{jl}) - \log(1+Y_{il})]\Big) \tag{40}$$

The second derivative with respect to α_{jl} is given by:

$$\begin{aligned}\frac{\partial^2}{\partial^2 \alpha_{jl}} \log P(\mathcal{Y}, \Theta) &= \sum_{i=1}^{N} P(j|\vec{Y_i})\Big(\frac{\partial P_{IBeta}(Y_{il}|\theta_{jl})}{\partial \alpha_{jl}}[\Psi(\alpha_{jl}+\beta_{jl}) - \Psi(\alpha_{jl}) \\ &- \log(1+Y_{il})] + P_{IBeta}(Y_{il}|\theta_{jl})[\Psi'(\alpha_{jl}+\beta_{jl}) - \Psi'(\alpha_{jl})]\Big), \end{aligned} \tag{41}$$

The second derivative w.r.t β_{jl} is obtained through the same development.

3.1.3 GID optimal predictor

In this section, we present our novel non-linear optimal predictor method based on generalized inverted Dirichlet distribution. We consider the conditional expectation property to predict one random variable from the other neighboring variables.

We consider p data points $(X_i, X_{i+1}, \ldots, X_{i+p-1})$, knowing their values, we predict the neighboring data point $\hat{X}_{i+p}$ on the base of minimizing the *mean squared error* (MSE). Therefore, we model the joint density of X_{i+p} and its neighbors using the generalized inverted Dirichlet. We take $i = 0$ and we derive the equations.

$$\vec{X} \sim GID(\vec{\theta}) \tag{42}$$

Considering generalized inverted Dirichlet properties [28], the conditional random variable Y_p follows an inverted Beta distribution:

$$Y_p = \frac{X_p}{T_{p-1}}|X_{p-1}, \ldots, X_1, X_0 \sim IB(\alpha_p, \beta_p) \tag{43}$$

Consequently, the conditional probability density function of X_p is

$$X_p \mid X_{p-1}, \ldots, X_0 \sim T_{p-1}\, IB(\alpha_p, \beta_p) \tag{44}$$

where $T_{p-1} = 1 + \sum_{k=1}^{p-1} X_k$.
Hence, the conditional expectation expression of X_p is expressed as follows.

$$E(X_p \mid X_{p-1}, X_1, \ldots, X_0) = T_{p-1}\, \frac{\alpha_p}{\beta_p - 1} \tag{45}$$

In the case of mixture models, the optimal predictor expression can be derived directly by following steps defined in [24] (more details are in [16]):

$$\hat{X}_p = E\big(X_p \mid X_{p-1}, \ldots, X_0\big) \tag{46}$$

$$\begin{aligned} \hat{X}_p &= \int X_p\, P(X_p \mid X_{p-1}, \ldots, X_0)\, \mathrm{d}X_p \\ &= \sum_{j=1}^{K} \pi'_j E_j\big(X_j \mid X_{j-1}, \ldots, X_0\big) \end{aligned} \tag{47}$$

where

$$\pi'_j = \pi_j \frac{\int P_j(X_p\ , \ldots, X_0)\, \mathrm{d}X_p}{\int P(X_p\ , \ldots, X_0)\, \mathrm{d}X_p} \tag{48}$$

$$\pi'_j = \pi_j \frac{P_j(X_{p-1}\ , \ldots, X_0)}{\sum_{j=1}^{K} \pi_j P_j(X_{p-1}\ , \ldots, X_0)} \tag{49}$$

Finally, the GID optimal predictor is resumed in the following linear combination of X_p neighbors:

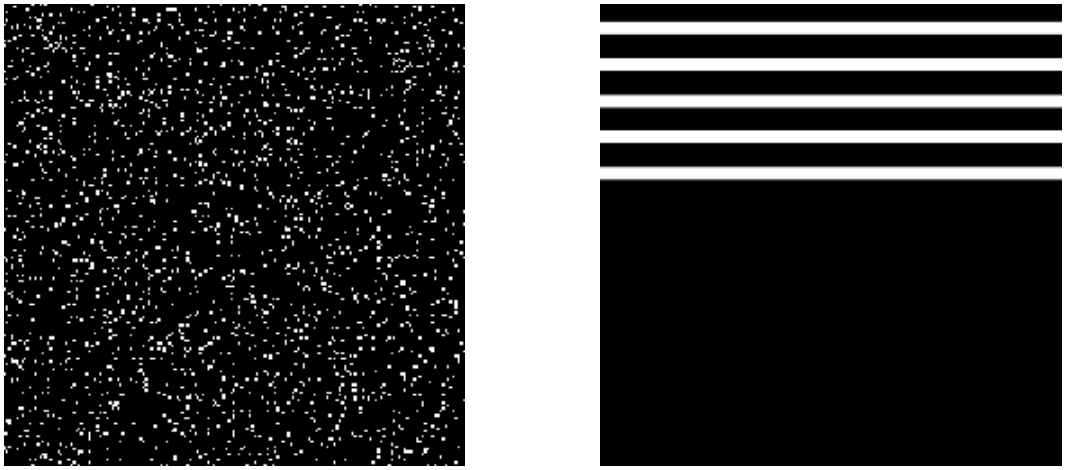
$$\hat{X}_p = \sum_{j=1}^{K} \pi'_j \left(1 + \sum_{k=1}^{p-1} X_k\right) \frac{\alpha_p}{\beta_p - 1} \tag{50}$$

3.2 Experimental results

Image inpainting is the process of restoring deteriorating, damaged or missing parts of an image to produce a complete picture. It is an active area of image processing research [11, 53, 54] where machine learning has exciting results comparable to artists' results. Mainly, in this process, we will be completing a missing pixel by an approximated value that depends on its neighborhood. In our work, we use the 3^{rd}

order non-symmetrical half-plane casual neighborhood system [24, 43]. We apply the model on a publicly available dataset; Paris StreetView [31]. Then, we compare it with the widely used mixture of Gaussian predictor, generalized Dirichlet mixture predictor and inverted Dirichlet mixture predictor. We are not trying to restore the ground-truth image, our goal is to get an output image that is close enough or similar to the ground-truth. Therefore, we use the structural similarity index (SSIM) [10] to gauge the differences between the predicted images and the original ones. We also perform signal to noise ratio (PSNR) [55] to evaluate the performance of the models.

We reduce the size of the original images to 256×256 to minimize the complexity of computing the model's parameters. We train the model on 70% of the database and we test on the rest. We apply two types of masks. The first one is randomly distributed as shown in figure 4a. And, for the second one, we discard lines of the images, as in figure 4b. Finally, we compute the SSIM and PSNR of each test image with its corresponding ground-truth, and we average all over the test set.



(a) Random mask (b) Line mask

Figure 4: Types of image mask

As we mentioned earlier, we discard around 15% of the pixels randomly. Figure 5 reveals that the difference between models' prediction is undetectable visually. Moreover, table 3 shows that the difference between the models is not significant. There is a slight advantage for the use of GIDM model compared to the others. Thus, we

Figure 5: Models' performance on random masked images. 1^{st} column is for the ground truth images, 2^{nd} column is for the masked images , 3^{rd} column is for the GM prediction, 4^{th} column is for the DM prediction , 5^{th} column is for the IDM prediction, 6^{th} column is for the GIDM prediction

conclude that this approach of models' evaluation is not appropriate. For that, we decide to remove slightly thick lines of pixels and re-evaluate the models.

To evaluate the models' performance, we used TensorFlow to calculate the PSNR and Skimage python package for the SSIM metric.

After discarding lines from the images, we are able to generate back again the missing pixels, and figure 6 demonstrates that GIDM is the most efficient model among all the others. This is also clear in figure 6, where we can notice in the chosen images that GIDM is the most accurate re-generator of discarded pixels. Therefore, our work has shown that image data is better represented by generalized inverted Dirichlet. It is noteworthy to mention that these models' performance is hugely dependent on the size of the masks, the hyper-parameters, the type and order of the neighbouring system.

Table 3: Models' evaluation for the randomly masked images.

	PSNR	SSIM
GM	21.832	0.853
DM	25.963	0.878
IDM	25.672	0.875
GIDM	26.126	0.887

Table 4: Models' evaluation for the line masked images.

	PSNR	SSIM
GM	20.366	0.833
DM	25.851	0.856
IDM	27.673	0.868
GIDM	29.398	0.891

Figure 6: Models' performance on line masked images. 1st column is for the ground truth images, 2nd column is for the masked images, 3rd column is for the GM prediction, 4th column is for the DM prediction , 5th column is for the IDM prediction, 6th column is for the GIDM prediction

Chapter 4

Inverted Dirichlet Power Steady model for Time series Forecast with local Variational inference

4.1 Introduction

Time series analysis has been a challenging problem for decades [56, 57]. It is a rich and fast growing field where many techniques have been applied for modeling and analyzing data. It takes into consideration that data points taken over time may have an internal pattern demonstrated in a seasonal, trend, cyclical and noise components. For that purpose, taking advantage of state space representation for modeling and smoothing time series has been extensively proposed in earlier years by Akaike [58], Aoki [59, 60], Hannan [61], Kitagawa and Gersch [62]. Recently, the application of time series using state space models has become of great interest for various applications such as analyzing water quality [63], forecasting electricity and traffic data [64], predicting quality of service performance of web services [37], forecasting business demand [65, 66], and predicting supply chain planning [67].

In this chapter, we develop a methodology for modeling, forecasting, and estimating times series of continuous proportions assumed to follow an inverted Dirichlet distribution. We summarize our contributions as the following:

1. Proposing inverted Dirichlet distribution with state space model.

2. Introducing a new maximum a posteriori estimation of Inverted Dirichlet state

space model's parameters.

3. Predicting Inverted Dirichlet power steady model by local Variational inference.

4. Forecasting electricity consumption using Inverted Dirichlet state space model.

4.2 Exponential form of Inverted Dirichlet distribution

Given a D-dimensional positive vector $\vec{X} = (X_1, \ldots, X_D)$ generated from an inverted Dirichlet distribution characterized by a $D+1$-dimensional parameter vector $\vec{\alpha} = (\alpha_1, \ldots, \alpha_{D+1})$, the probability density function is as follows

$$P(\vec{X}|\vec{\alpha}) = \frac{\Gamma(|\alpha|)}{\prod_{d=1}^{D+1}\Gamma(\alpha_d)} \prod_{d=1}^{D} X_d^{\alpha_d - 1}(1 + |\vec{X}|)^{-|\alpha|} \tag{51}$$

where $|\alpha| = \sum_{d=1}^{D+1}\alpha_d,\ \alpha_d > 0,\ d = 1, \ldots, D+1,\ |\vec{X}| = \sum_{d=1}^{D} X_d, X_d > 0,\ d = 1, \ldots, D$, and $\Gamma(.)$ is the Gamma function defined by $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$. We denote the inverted Dirichlet distribution by InvDir(α).

The Inverted Dirichlet is a member of the exponential family of distributions which have the form:

$$p(\vec{X}|\vec{\alpha}) = H(X)\exp[G(\vec{\alpha})T(X) + \Phi(\vec{\alpha})] \tag{52}$$

where $T(X)$ is a vector of sufficient statistics and $G(\vec{\alpha})$ is the vector for natural parameters.

Then, by letting the likelihood function follows the exponential forms [68], we write down the expression:

$$\begin{aligned} p(\vec{X}|\vec{\alpha}) &= \exp\Bigg[\log(\Gamma(|\alpha|) - \sum_{d=1}^{D+1}\log(\Gamma(\alpha_d)) + \sum_{d=1}^{D}\alpha_d \log(X_d) - \sum_{d=1}^{D} log(X_d) \\ &- |\alpha|\log\left(1 + \sum_{d=1}^{D} X_d\right)\Bigg] \end{aligned} \tag{53}$$

If we consider the exponential family or Koopman–Darmois expression equation (52), we can write the natural parameters as:

$$
\begin{aligned}
H(X) &= 1 && (54)\\
T_d(X) &= \log(X_d) \quad ; \quad d = 1, \ldots, D && (55)\\
T_{D+1}(X) &= \log\left(1 + \sum_{d=1}^{D} X_d\right) && (56)\\
G_d(\vec{\alpha}) &= \alpha_d \quad ; \quad d = 1, \ldots, D && (57)\\
G_{D+1}(\vec{\alpha}) &= -|\alpha| && (58)\\
\Phi(\vec{\alpha}) &= \log(\Gamma(|\alpha|)) - \sum_{d=1}^{D+1} \log\Gamma(\alpha_d) && (59)
\end{aligned}
$$

4.3 Inverted Dirichlet State Space model

We introduce in this section the new proposed state space model where the observations are to follow an Inverted Dirichlet distribution. For that purpose,using exponential family properties, the prior distribution is defined as follows:

$$
p(\vec{\alpha}_t) \propto \exp\left[\sum_{d=1}^{D} \rho_d \alpha_d - \rho_{D+1}|\alpha| + \kappa\Big(\log(\Gamma(|\alpha|)) - \sum_{d=1}^{D+1} \log(\Gamma(\alpha_d))\Big)\right] \qquad (60)
$$

Then, using Bayes's rule the posterior distribution of an unobserved states is defined by:

$$
\begin{aligned}
p(\vec{\alpha}_t|\rho, \kappa, \vec{X}) &\propto p(\vec{\alpha}_t|\rho, \kappa) \times p(\vec{X}|\vec{\alpha}_t) && (61)\\
&\propto \exp\left[\sum_{d=1}^{D}\Big(\rho_d + \log(X_d)\alpha_d\Big) - |\alpha|\Big(\rho_{D+1} + \log(1 + \sum_{d=1}^{D}\log(X_d)\Big)\right.\\
&+ \left.(\kappa+1)\Big(\log(\Gamma(|\alpha|)) - \sum_{d=1}^{D+1}\log(\Gamma(\alpha_d))\Big)\right]
\end{aligned}
$$

According to steady state model, the state $\vec{\alpha}_t$ is assumed to evolve in time where the new state $\vec{\alpha}_{t+1}$ is defined by:

$$
p(\vec{\alpha}_{t+1}|\Delta_t) \propto p(\vec{\alpha}_t|\Delta_t)^{\sigma} \quad ; \quad \sigma \in]0, 1[\qquad (62)
$$

where $\Delta_t = \{I_t, \Delta_{t-1}\}$ and $I_t = \{\vec{X}_t\}$ are the external estimated parameters and all the other relevant information available at t. For instance, we define D_0 the values of the externally estimated parameters and all relevant information available at $t = 0$.

The principle of the state space model is based on two step: the predicting step and the updating step. First, the predicting step consists of determining the $p(\vec{\alpha}_{t+1}|\Delta_t)$ using equation 62 where the mode of the distribution of $(\alpha_t|\Delta_t)$ to $(\alpha_{t+1}|\Delta_t)$ is unchangeable whereas the dispersion gets larger.

for $d = 1, \ldots, D+1$

$$\gamma\rho_{d\{t/t\}} = \rho_{d\{t+1/t\}} \tag{63}$$

$$\kappa_{\{t/t\}} = \kappa_{\{t+1/t\}} \tag{64}$$

Second, the updating step consists of updating the following:

$$p(\vec{\alpha}_{t+1}|\Delta_{t+1}) \propto p(X_{t+1}|\alpha_{t+1})p(\alpha_{t+1}|\Delta_t) \tag{65}$$

Using Bayes's theorem and Markov Random Fields equations [69], we assume that ρ and κ are independent and identically distributed variables. Then, the updating equations are derived from the prior at $t+1$ and the following equation:

$$p(\vec{\alpha}_{t+1}|\Delta_{t+1}) \propto p(_{t+1}|\alpha_{t+1}, \rho_{t+1}, \kappa_{t+1}) \times p(\vec{\alpha}_{t+1}|\vec{X}_t, \rho_{t+1}, \kappa_{t+1}) \tag{66}$$

In regard to the above equation, we derive the updating equations of the hyper-parameters as the following:

$$\rho_{d\{t+1/t+1\}} = \gamma(\rho_{d\{t+1/t\}} + \log X_d) \quad ; \quad d = 1, .., D \tag{67}$$

$$\rho_{D+1\{t+1/t+1\}} = \gamma\big(\rho_{D+1\{t+1/t\}} + \log\,(1 + \sum_{d=1}^{D} X_d)\big) \tag{68}$$

$$\kappa_{\{t+1/t+1\}} = \gamma\kappa_{\{t+1/t\}} + 1 \tag{69}$$

We conduct two procedures to estimate the mode $\hat{\alpha}$ of the inverted Dirichlet conjugate. We use the mixed moments characteristic of the exponential family, where the first moment is defined as $E(T(x)) = -\frac{\partial\phi(x)}{\partial x}$ and the second moment is given by $VAR(T(x)) = -\frac{\partial^2\phi(x)}{\partial^2 x}$. Then, with regards to Theorem.2 in [70], we have:

$$\Psi(\hat{\vec{\alpha}}) - \Psi(|\hat{\alpha}|)\vec{u} = \rho \quad ; \quad \vec{u} = (1, \ldots, 1) \in \Re^{D+1} \tag{70}$$

We conduct the maximum a posteriori to estimate (MAP) $\hat{\vec{\alpha}}$ [71] and we assume that all the components of $\vec{X}$ are independent:

$$\begin{aligned}\vec{\alpha}_{MAP} &= {}_{\alpha}(\log p(\vec{\alpha}|\rho,\kappa,\vec{X})) \\ &= {}_{\alpha}\Bigg[\sum_{d=1}^{D}\Big(\rho_d+\log(X_d)\alpha_d\Big)-|\alpha|\Big(\rho_{D+1}+\log(1+\sum_{d=1}^{D}\log(X_d)\Big) \\ &+ (\kappa+1)\Big(\log(\Gamma(|\alpha|))-\sum_{d=1}^{D+1}\log(\Gamma(\alpha_d))\Big)\Bigg] \end{aligned} \tag{71}$$

Then, we drive the following equations for $\vec{\alpha}_{MAP}$:
for $s = 1, \ldots, D$

$$\begin{aligned}\Psi(|\hat{\alpha}|)-\Psi(\hat{\alpha_s}) &= \frac{\rho_{D+1}+\log\left(1+\sum_{d=1}^{D}X_d\right)-\rho_s-\log X_s}{\kappa+1} \\ \Psi(|\hat{\alpha}|)-\Psi(\hat{\alpha_{D+1}}) &= \frac{\rho_{D+1}+\log\left(1+\sum_{d=1}^{D}X_d\right)}{\kappa+1}\end{aligned} \tag{72}$$

We deploy the Newthon-Raphson approach to get the estimate value of $\vec{\alpha}_{MAP}$.

4.4 Local variational inference

In the Bayesian analysis of statistical modelling, averaging the future predictions over the posterior densities of the unknown parameters is called the predictive distribution. This is leads to more accurate results when comparing to point estimate method in computing the predictive likelihood of the furthcoming data [72].

We used different approach to approximate the prediction distribution. We applied the normal variational inference defined in [73], but, the solution was intractable again. Thus, the new approach of local variational inference (LVI) was used to derive an approximation of the following expression.

$$p(\vec{X}_{t+1}|\Delta_t) = \int p(\vec{X}_{t+1}|\vec{\alpha}_{t+1}) \times p(\vec{\alpha}_{t+1}|\Delta_t) \quad d_{\vec{\alpha}_{t+1}} \tag{73}$$

The local variational inference requires finding limits on a subset of variables or objective function. Hence, to apply it certain conditions must be verified.

If there exists $\{h(\vec{\alpha}_{t+1},\gamma)/ \quad h(\vec{\alpha}_{t+1},\gamma) \geq p(\vec{\alpha}_{t+1}|\Delta_t)\}$. Then, Suppose we have $G(\sigma)$ given by:

$$G(\sigma) = \int p(\vec{X}_{t+1}|\vec{\alpha}_{t+1}) \times h(\vec{\alpha}_{t+1},\gamma) \quad d_{\vec{\alpha}_{t+1}} \tag{74}$$

We have $p(\vec{X}_{t+1}|\Delta_t) \leq G(\sigma)$. Thus, an upper bound of the predictive distribution $G(\sigma)$ is found. Knowing that $G(\sigma)$ is only a function of σ, by calculating the optimal value σ^* which minimizes $G(\sigma)$, we can approximate our predictive distribution.

$$\sigma^* =_\sigma G(\sigma) \tag{75}$$

For that, we have to study the concavity of our posterior distribution. Thus, the expression of the first derivative of the posterior is as follows:

$$\frac{\partial p(\vec{\alpha}_{t+1}|\Delta_t)}{\partial \vec{\alpha}_{t+1}} = \Bigg[(\rho_d + \log(X_d)) - (\rho_{D+1} + \log(1 + \sum_{d=1}^{D+1} X_d)) + (\kappa+1)(\psi(|\alpha|) - \psi(\alpha_d))\Bigg] \times p(\vec{\alpha}_{t+1}|\Delta_t) \tag{76}$$

The sign of the first derivative gives an indication of the monotonicity of the posterior. To study the concavity, we generate the second derivative defined by the following expression.

$$\frac{\partial^2 p(\vec{\alpha}_{t+1}|\Delta_t)}{\partial^2 \vec{\alpha}_{t+1}} = \Bigg[\big[(\rho_d + \log(X_d)) - (\rho_{D+1} + \log(1 + \sum_{d=1}^{D+1} X_d)) + (\kappa+1)(\psi(|\alpha|) - \psi(\alpha_d))\big]^2 + \big[(\kappa+1)(\psi^{'}(|\alpha|) - \psi^{'}(\alpha_d))\big]\Bigg] \times p(\vec{\alpha}_{t+1}|\Delta_t) \tag{77}$$

where $\psi^{'} = \frac{\partial \psi}{\partial \vec{\alpha}}$, the sign of the second derivative depends on the sign of $\psi^{'}(|\alpha|) - \psi^{'}(\alpha_d)$.

Following theorem 1 in [72], it is proven that the logarithm of Multivariate-Inverse-Beta Function is jointly concave. Therefore, our posterior distribution is also concave and, demonstrates the following relation.

$$\frac{\Gamma(\sum_{d=1}^{D+1} \alpha_d)}{\prod_{d=1}^{D+1} \Gamma(\alpha_d)} \leq \frac{\Gamma(\sum_{d=1}^{D+1} \tilde{\alpha}_d)}{\prod_{d=1}^{D+1} \Gamma(\tilde{\alpha}_d)} \times \exp\left(\sum_{d=1}^{D+1} \big[\psi(|\tilde{\alpha}|) - \psi(\tilde{\alpha}_d)\big](\alpha_d - \tilde{\alpha}_d)\right) \tag{78}$$

where $\tilde{\alpha}_d$ for $d \in \{1, .., D+1\}$ is any point from the posterior distribution and $|\tilde{\alpha}| = \sum_{d=1}^{D+1} \tilde{\alpha}_d$. Then, by substituting this last inequality 78 in the prediction equation, we

will have the upper bound which is given as:

$$
\begin{aligned}
p(\vec{X}_{t+1}|\Delta_t) \;\leq\; & \int \exp\Bigg[\sum_{d=1}^{D}\big(\rho_d+\log(X_d)\alpha_d\big)-|\alpha|\big(\rho_{D+1}+\log(1+\sum_{d=1}^{D}\log(X_d)\big) \quad (79)\\
+\; & (\kappa+2)\Big[\big(\log(\Gamma(|\tilde{\alpha}|))-\sum_{d=1}^{D+1}\log(\Gamma(\tilde{\alpha}_d))\big)+\big(\sum_{d=1}^{D+1}\big[\psi(|\tilde{\alpha}|)-\psi(\tilde{\alpha}_d)\big](\alpha_d\\
-\; & \tilde{\alpha}_d)\big)\Big]+\Big(\sum_{d=1}^{D}\alpha_d\log(X_d)+\sum_{d=1}^{D}\log(X_d)-|\alpha|\log\big(1+\sum_{d=1}^{D}X_d\big)\Big)\Bigg]d\vec{\alpha}_{t+1}\\
\leq\; & \exp\Big(\sum_{d=1}^{D}\log(X_d)+(\kappa+2)\Big(\log(\Gamma(|\tilde{\alpha}|))-\sum_{d=1}^{D+1}\log(\Gamma(\tilde{\alpha}_d))\\
-\; & \sum_{d=1}^{D+1}\tilde{\alpha}_d\big[\psi(|\tilde{\alpha}|)-\psi(\tilde{\alpha}_d)\big]\Big)\times\int\exp\Bigg[\sum_{d=1}^{D}\alpha_d\Big(\rho_d+\log_{X_{d,t+1}}\\
+\; & \log_{X_{d,t}}+(\kappa+2)\Big[\psi(|\tilde{\alpha}|)-\psi(\tilde{\alpha}_d)\Big]\Big)-|\alpha|\big(\rho_{D+1}\\
+\; & \log(1+\sum_{d=1}^{D}\log X_{d,t+1})+\log(1+\sum_{d=1}^{D}\log X_{d,t})\big)\Bigg]d\vec{\alpha}_{t+1}
\end{aligned}
$$

For the ease of calculations, we suppose that $\mathbf{B}=\exp\Big(\sum_{d=1}^{D}\log(X_d)+(\kappa+2)\Big(\log(\Gamma(|\tilde{\alpha}|))-\sum_{d=1}^{D+1}\log(\Gamma(\tilde{\alpha}_d))-\sum_{d=1}^{D+1}\tilde{\alpha}_d\big[\psi(|\tilde{\alpha}|)-\psi(\tilde{\alpha}_d)\big]\Big)\Big)$ and, given that $\frac{\exp a}{\exp b}\leq\exp a$ for any real positive b variable. The inequality(79) will be given as follows:

$$
\begin{aligned}
p(\vec{X}_{t+1}|\Delta_t) \;\leq\; & \mathbf{B}\times\int\prod_{d=1}^{D}\exp\Bigg[\alpha_d\Big(\rho_d+\log_{X_{d,t+1}}+\log_{X_{d,t}}+(\kappa+2)\Big[\psi(|\tilde{\alpha}|) \quad (80)\\
-\; & \psi(\tilde{\alpha}_d)\Big]\Big)-|\alpha|\big(\rho_{D+1}+\log(1+\sum_{d=1}^{D}\log X_{d,t+1})\\
+\; & \log(1+\sum_{d=1}^{D}\log X_{d,t})\big)\Bigg]d\vec{\alpha}_{t+1}
\end{aligned}
$$

We set C equal to the value of the integral and we calculate it using induction method. Thus, C is equal to the following:

$$C = \prod_{d=1}^{D} \frac{\exp\left[\alpha_d\Big(\rho_d + \log_{X_{d,t+1}} + \log_{X_{d,t}} + (\kappa+2)\Big[\psi(|\tilde{\alpha}|) - \psi(\tilde{\alpha}_d)\Big]\Big)\right]}{\Big(\rho_d + \log_{X_{d,t+1}} + \log_{X_{d,t}} + (\kappa+2)\Big[\psi(|\tilde{\alpha}|) - \psi(\tilde{\alpha}_d)\Big]\Big)^{D}} \tag{81}$$

Therefore, the resulting upper bound of the predictive distribution is equal to $B \times C$.

4.5 Experiments

In this section, we evaluate the proposed time series model Inverted Dirichlet Power Steady Model (IDPSM) on electricity consumption dataset [74]. In this evaluation, the proposed method is applied to forecast the future electricity production based on previous observations. Therefore, we perform an exploratory data analysis to understand the data features and behaviour. After, we compare our model performance to the Generalized Dirichlet Power Steady Model (GDPSM) proposed in [37].

4.5.1 Exploratory data analysis

Time series analysis has been a challenging task in multiple machine learning problems. Hence, we consider the daily time series of Open Power System Data (OPSD) for Germany to assess the performance of our model [74]. The dataset includes electricity consumption and production of Germany that are stated as daily totals in gigawatt-hours collected from 2006 to 2015 and hourly totals from 2015 to 2020. The columns of the dataset are defined as follows:

- ☐ Date in (yyyy-mm-dd format)
- ☐ Load_actual_entsoe_transparency refers to Consumption in gigawatt-hours
- ☐ Wind_generation_actual refers to wind power in gigawatt-hours
- ☐ Solar_generation_actual refers to solar power in gigawatt-hours

In this work, we focus on the hourly collected part of the electricity consumption and production time series. We start analyzing the data exploratory to study the

variation over time in Germany. In fact, this study demonstrates some time series features of the dataset such as the local and the long-term trends, seasonality, highest and lowest data points.

First, we create line plot to visualize the evolution of the electricity consumption over time. We can see in the figure 7 below that the line plot is congested and unreadable due to the huge number of data points. Yet, we can notice a repeatable pattern for each year known as seasonality of the time series data that oscillates between low and high values on a yearly time scope. We can also notice that the electricity consumption is lowest in summer and highest in winter.

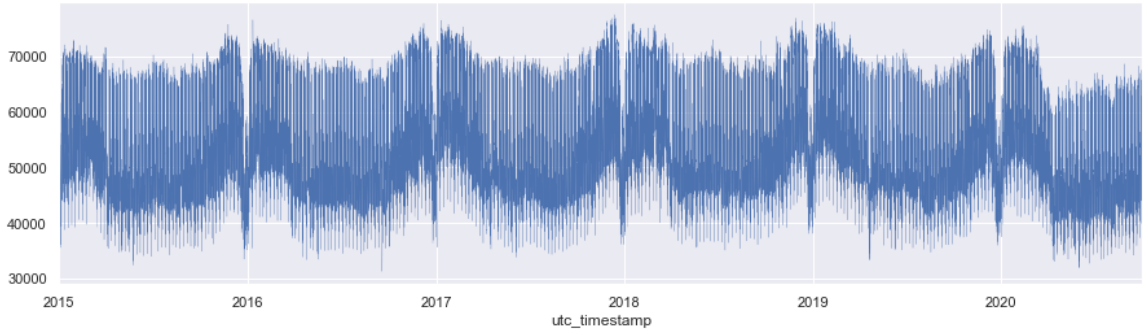


Figure 7: Electricity consumption's hourly variation for Germany 2015-2020

Then, we demonstrate the seasonality for a narrower time scale. Figure 8 shows that the electricity consumption has a periodic pattern that is repeated on a weekly basis.

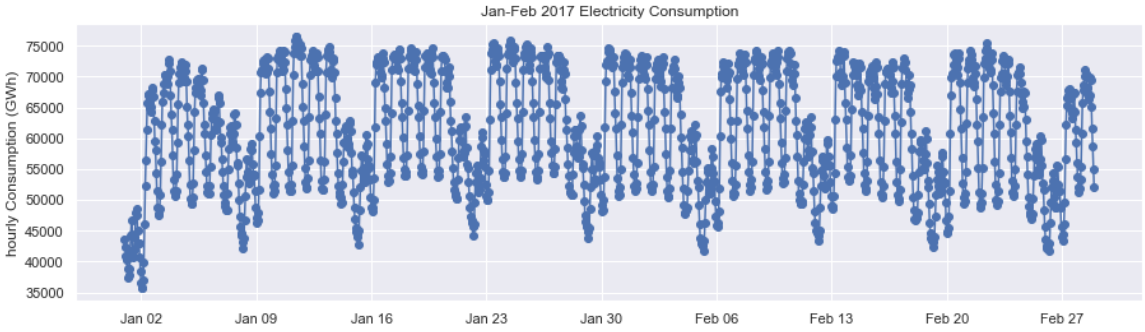


Figure 8: Electricity consumption's hourly variation for Germany Jan-Feb 2017

We can notice the daily oscillations. In addition, it is clear at this point that the electricity consumption is the highest on weekdays and the lowest on weekend and especially on early January which is a holiday. After that, we group the data

points by month and we perform boxplot to study the seasonality of the dataset. The following figure confirms the yearly seasonality of the data. In addition, the electricity consumption does not present data outliers which will make it easier while learning the pattern of the dataset.

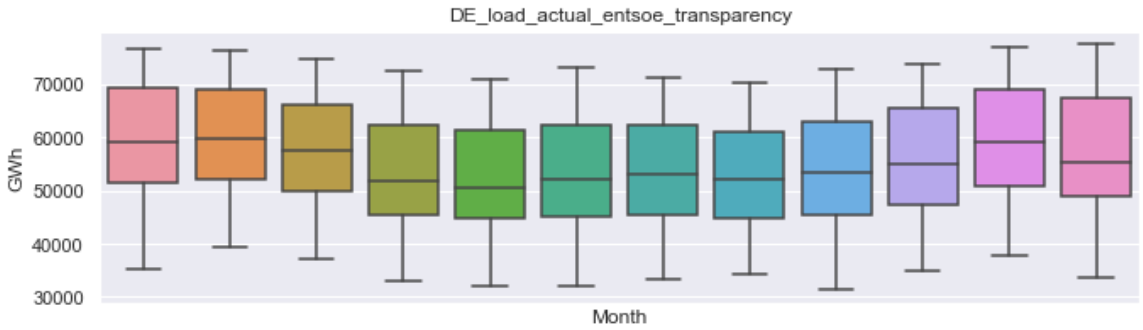


Figure 9: Boxplot of Germany electricity consumption grouped by month

Next, we group the data on daily basis and we generate again the boxplot to explore more the highest and lowest days of electricity consumption. Hence, figure 10 shows clearly how the electricity load is at its highest level during weekdays and at its lowest on weekends.

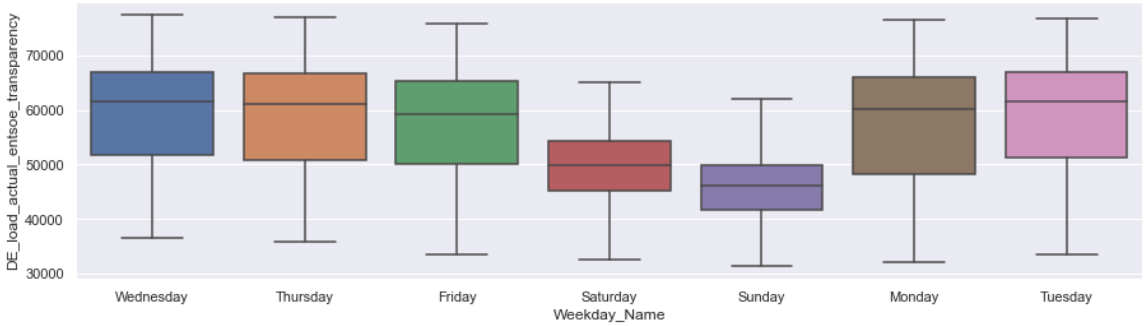


Figure 10: Boxplot of Germany electricity consumption grouped by days

In the figure 11 we downsample the data but taking the average of each day. We can notice that the time series become smoother because we reduce the frequency of the data. Thus, we explore the daily trend.

Finally, we use the rolling window feature to explore the weekly and the long-term trend and seasonality of the electricity load. In fact, the weekly mean resampled data shown in figure 12 is smoother than the daily mean resampled one and it demonstrates

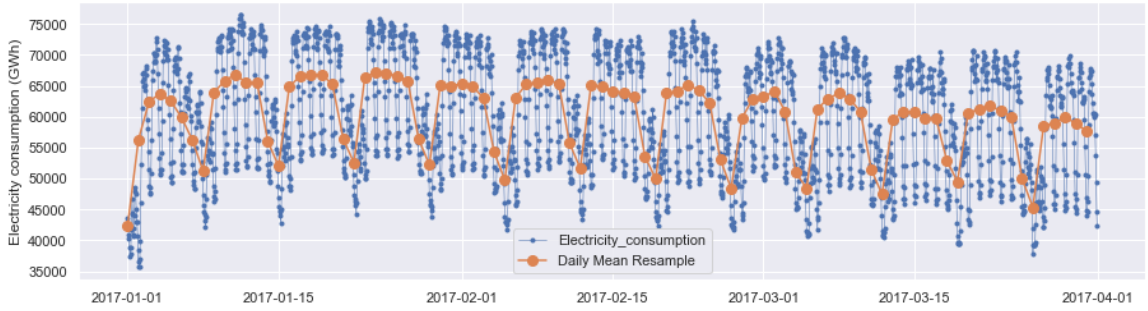


Figure 11: downsampling of the dataset to daily consumption Jan-Mar 2017

again the seasonality and the trend of the time series on a weekly basis. Yet, the long-term trend is horizontal with very small oscillations that decreases abruptly in 2020, because of the COVID-19.

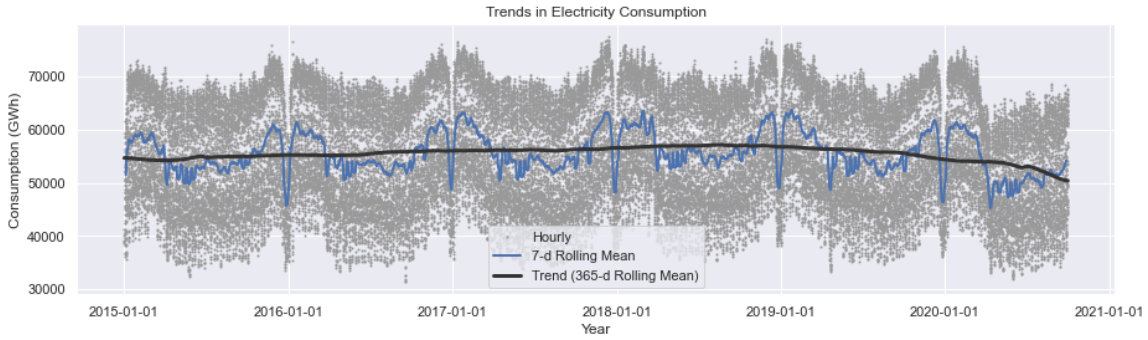


Figure 12: Overall trend of electricity consumption

The exploratory data analysis exhibits the seasonality, trend and the particular lowest and highest curves of the time series for the electricity consumption time series data. We compare the performance of IDPSM and GDPSM models by forecasting the electricity consumption as 1-dimension time series data and predicting wind, solar generation and electricity consumption as 3-dimensions time series.

4.5.2 Model Evaluation

We use the standardized residuals to evaluate the proposed IDPSM, the correlation between the residuals at lag 0, and the mean squared error (MSE) of the predictions. We demonstrate the performances of IDPSM and GDPSM. The standardized

residuals are computed using the following expression [36]:

$$R_t = \frac{X_t - E(X_t|\Delta_{t-1})}{var(X_t|\Delta_{t-1})} \tag{82}$$

Δ_t is explained recursively by $\Delta_t = \{I_t, \Delta_{t-1}\}$, where I_t refers to the observed data at time t as well as all other significant information at time t but not at time $t-1$. In this expression, Δ_{t-1} denotes all the observations provided at time $t-1$. $E(X_t|\Delta_{t-1})$ and $var(X_t|\Delta_{t-1})$ respectively define the posterior mean and variance of the prediction distribution defined in equation (73).

4.5.3 Results and discussions

After exploring the dataset trends, seasonality and particular points, we evaluate our model with the time series data of Open Power System Data (OPSD). We focus on electricity load, wind generation and solar generation of Germany. As described in the previous section, this dataset presents a challenge to test the performance of time series forecasting models as it is collected hourly. Hence, we run multiple simulations for different values of γ, which is the model's hyperparameter. This parameter is not interpretable, and we compare IDPSM and GDPSM with five different simulations by varying γ, which will take the correspondent values $\{0.01\ 0.25\ 0.50\ 0.75\ 0.99\}$. We run 15 different simulations for each value of gamma and we take the average results for 1-dimension data.

First, we start by comparing the models by only forecasting the electricity consumption time series. Table 5 demonstrates the standardized residuals for both models. We average the results over 15 simulation where we change the initial parameters each time. Table 6 also is the average results of the 15 different runs. These two tables below exhibit the merits of our proposed approach over GDPSM and shows that for each γ value we may have different output as shown is figure 13. In fact, the worst performance was for the biggest value of γ. However, this value is still interpretable.

As future work, we can enhance the performance of the state space model by adding dummy variables to represent covariances, trends, seasonality and interventions to the predictive state distribution, we can enhance the performance of our model [36]. In fact, the predictive state distribution of IDPSM will be represented by

Table 5: Standardized residuals for electricity consumption

γ	IDPSM	GDPSM
0.01	0.581	0.638
0.25	0.621	0.696
0.50	0.682	0.661
0.75	0.592	0.613
0.99	0.727	0.751

Table 6: MSE for electricity consumption

γ	IDPSM	GDPSM
0.01	0.089	0.152
0.25	0.082	0.114
0.50	0.112	0.092
0.75	0.072	0.123
0.99	0.159	0.185

the following expression:

$$p(\vec{\tilde{\alpha}}_{t+1}|\Delta_t) = p(\vec{\alpha}_{t+1}|\Delta_t) + M \times \vec{X}_{t+1} \tag{83}$$

where M is the matrix of regression parameters and $\vec{\tilde{\alpha}}_{t+1}$ is the new state model. The incorporation of independent variables to the state equation will change the state model location, and enhance the accuracy of the forecast. However, the model complexity will increase, and further mathematical derivations have to be accomplished.

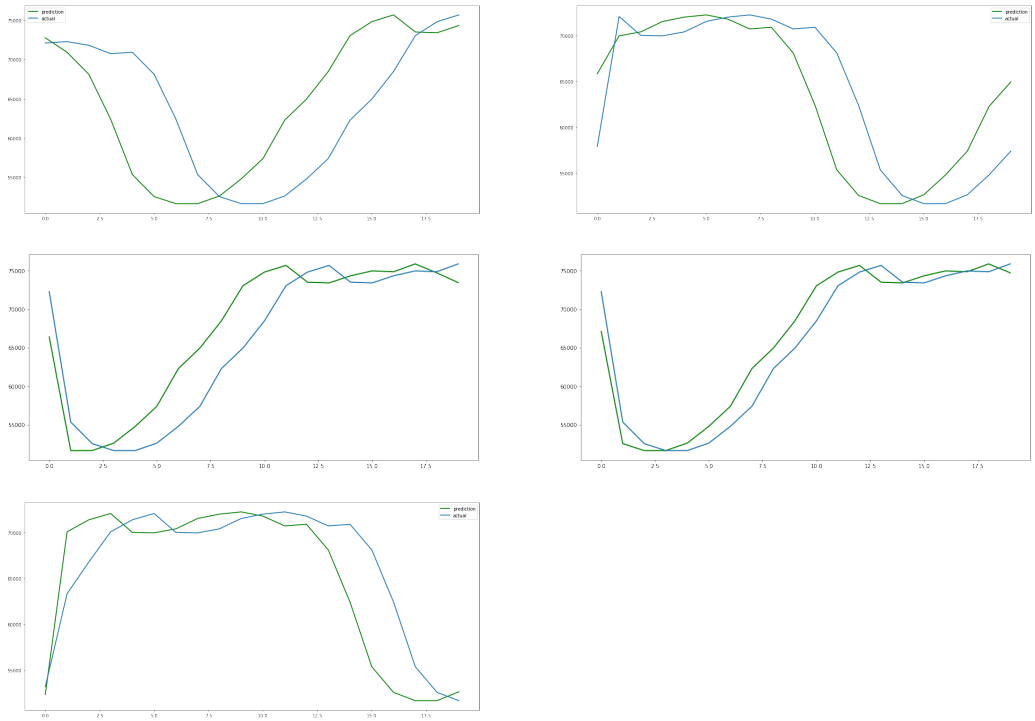


Figure 13: IDPSM performance on electricity consumption dataset. 1$^{\text{st}}$ column is for the smallest value of γ, ranked to the 5$^{\text{th}}$ column which is for the biggest value of γ

Chapter 5

Conclusion

In this thesis, we have developed two different models for image pixel prediction and one model for time series forecasting.

In chapter 2, a new method to model non-Gaussian positive vectors was proposed. First, an analytic expression of the optimal predictor that is based on the inverted Dirichlet mixture is derived. Next, we consider the NSHP neighbouring system and an MLE expectation maximization framework to estimate the model parameters. Then, we conduct two challenging applications to evaluate our predictor's performance that are image pixel restoration and objects detection. Finally, the results illustrate that our model out-performs Gaussian Mixture and Generalized Dirichlet Mixture predictors. In our future work, it would be of interest if we take into account the spatial correlation between pixels to select the set neighbours. This could enhance more the model's performance.

Then, in chapter 3, we have developed a new optimal predictor based on finite inverted Dirichlet mixtures. The GID demonstrates its efficiency in representing positive vectors due to its statistical characteristics through its covariance structure. We learnt the model parameters using MLE approach with Newton Raphson method, and we considered the 3 order NSHP neighbouring system to compute the predicted pixel. We evaluated the GID optimal predictor on image inpainting and we compared the proposed model to other similar related works. The experimental results demonstrate its capability that offers reliable prediction and modeling potential.

In chapter 4, we proposed a new state space model for time series forecasting. The model was based on inverted Dirichlet mixtures. A maximum a posteriori technique

is used to derive the analytical expression of the latent variable. Knowing that, the predictive density was analytically insolvable and intractable, we have applied the local variational inference to approximate it. Finally, we explored the features of the time series electricity consumption dataset of Germany by an exploratory data analysis and we compare our model to the generalized inverted Dirichlet power steady model. The standardized residuals and mean squared error have shown the merits of our proposed model.

The experiments with proposed models are motivating and proved to be a better solution than inverted Dirichlet and Gaussian mixture models for appropriate data. Future works might include trying other data distribution assumptions for image pixel prediction and enhancing the state space model by including some dummy variables in the latent variable equation to represents more the trend, seasonality and interventions of the time series data.

Bibliography

[1] Stefania Matteoli, Marco Diani, and Giovanni Corsini. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):5–28, 2010.

[2] Wentao Fan and Nizar Bouguila. Topic novelty detection using infinite variational inverted dirichlet mixture models. In Tao Li, Lukasz A. Kurgan, Vasile Palade, Randy Goebel, Andreas Holzinger, Karin Verspoor, and M. Arif Wani, editors, *14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, December 9-11, 2015*, pages 70–75. IEEE, 2015.

[3] Tarek Elguebaly and Nizar Bouguila. Finite asymmetric generalized gaussian mixture models learning for infrared object detection. *Computer Vision and Image Understanding*, 117(12):1659–1671, 2013.

[4] Taoufik Bdiri, Nizar Bouguila, and Djemel Ziou. Object clustering and recognition using multi-finite mixtures for semantic classes and hierarchy modeling. *Expert Syst. Appl.*, 41(4):1218–1235, 2014.

[5] Raman Maini and Himanshu Aggarwal. Study and comparison of various image edge detection techniques. *International journal of image processing (IJIP)*, 3(1):1–11, 2009.

[6] Ying Li and Khalid Sayood. Lossless video sequence compression using adaptive prediction. *IEEE Transactions on Image Processing*, 16(4):997–1007, 2007.

[7] Taoufik Bdiri, Nizar Bouguila, and Djemel Ziou. Visual scenes categorization using a flexible hierarchical mixture model supporting users ontology. In *25th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2013,*

Herndon, VA, USA, November 4-6, 2013, pages 262–267. IEEE Computer Society, 2013.

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[9] Ali Sefidpour and Nizar Bouguila. Spatial color image segmentation based on finite non-gaussian mixture models. *Expert Syst. Appl.*, 39(10):8993–9001, 2012.

[10] Nizar Bouguila. Non-gaussian mixture image models prediction. In *2008 15th IEEE International Conference on Image Processing*, pages 2580–2583. IEEE, 2008.

[11] Ali Mosleh, Nizar Bouguila, and A. Ben Hamza. Bandlet-based sparsity regularization in video inpainting. *J. Vis. Commun. Image Represent.*, 25(5):855–863, 2014.

[12] Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1145–1153, 2017.

[13] Charles W Therrien, Thomas F Quatieri, and Dan E Dudgeon. Statistical model-based algorithms for image analysis. *Proceedings of the IEEE*, 74(4):532–551, 1986.

[14] John Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[15] Geoffrey Grimmett, Geoffrey R Grimmett, David Stirzaker, et al. *Probability and random processes*. Oxford university press, 2001.

[16] Jun Zhang and Dehong Ma. Nonlinear prediction for gaussian mixture image models. *IEEE transactions on image processing*, 13(6):836–847, 2004.

[17] Kai Wang Ng, Guo-Liang Tian, and Man-Lai Tang. *Dirichlet and related distributions: Theory, methods and applications*, volume 888. John Wiley & Sons, 2011.

[18] Samuel Kotz, Narayanaswamy Balakrishnan, and Norman L Johnson. *Continuous multivariate distributions, Volume 1: Models and applications*. John Wiley & Sons, 2004.

[19] Taoufik Bdiri and Nizar Bouguila. Positive vectors clustering using inverted dirichlet finite mixture models. *Expert Systems with Applications*, 39(2):1869–1882, 2012.

[20] Taoufik Bdiri and Nizar Bouguila. Bayesian learning of inverted dirichlet mixtures for SVM kernels generation. *Neural Comput. Appl.*, 23(5):1443–1458, 2013.

[21] Taoufik Bdiri, Nizar Bouguila, and Djemel Ziou. Variational bayesian inference for infinite generalized inverted dirichlet mixtures with feature selection and its application to clustering. *Appl. Intell.*, 44(3):507–525, 2016.

[22] Taoufik Bdiri and Nizar Bouguila. Learning inverted dirichlet mixtures for positive data clustering. In Sergei O. Kuznetsov, Dominik Slezak, Daryl H. Hepting, and Boris G. Mirkin, editors, *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing - 13th International Conference, RSFDGrC 2011, Moscow, Russia, June 25-27, 2011. Proceedings*, volume 6743 of *Lecture Notes in Computer Science*, pages 265–272. Springer, 2011.

[23] Taoufik Bdiri and Nizar Bouguila. Bayesian learning of inverted dirichlet mixtures for svm kernels generation. *Neural Computing and Applications*, 23(5):1443–1458, 2013.

[24] Omar Graja and Nizar Bouguila. Finite inverted dirichlet mixture optimal pixel predictor. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5. IEEE, 2019.

[25] Taoufik Bdiri and Nizar Bouguila. An infinite mixture of inverted dirichlet distributions. In Bao-Liang Lu, Liqing Zhang, and James T. Kwok, editors, *Neural Information Processing - 18th International Conference, ICONIP 2011, Shanghai, China, November 13-17, 2011, Proceedings, Part II*, volume 7063 of *Lecture Notes in Computer Science*, pages 71–78. Springer, 2011.

[26] Taoufik Bdiri, Nizar Bouguila, and Djemel Ziou. A statistical framework for online learning using adjustable model selection criteria. *Eng. Appl. Artif. Intell.*, 49:19–42, 2016.

[27] Parisa Tirdad, Nizar Bouguila, and Djemel Ziou. Variational learning of finite inverted dirichlet mixture models and applications. In Yacine Laalaoui and Nizar Bouguila, editors, *Artificial Intelligence Applications in Information and Communication Technologies*, volume 607 of *Studies in Computational Intelligence*, pages 119–145. Springer, 2015.

[28] AK Gupta and D Song. Generalized liouville distribution. *Computers & Mathematics with Applications*, 32(2):103–109, 1996.

[29] Sami Bourouis, Mohamed Al Mashrgy, and Nizar Bouguila. Bayesian learning of finite generalized inverted dirichlet mixtures: Application to object classification and forgery detection. *Expert Systems with Applications*, 41(5):2329–2336, 2014.

[30] Sabri Boutemedjet, Nizar Bouguila, and Djemel Ziou. A hybrid feature extraction selection approach for high-dimensional non-gaussian data clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(8):1429–1443, 2008.

[31] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros. What makes paris look like paris? *ACM Transactions on Graphics (SIGGRAPH)*, 31(4):101:1–101:9, 2012.

[32] George EP Box and Gwilym M Jenkins. Some recent advances in forecasting and control. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 17(2):91–109, 1968.

[33] Rob Hyndman, Anne B Koehler, J Keith Ord, and Ralph D Snyder. *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.

[34] James Douglas Hamilton. *Time series analysis*. Princeton university press, 2020.

[35] James Durbin and Siem Jan Koopman. *Time series analysis by state space methods*. Oxford university press, 2012.

[36] Gary K Grunwald, Adrian E Raftery, and Peter Guttorp. Time series of continuous proportions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):103–116, 1993.

[37] Mohamad Mehdi, Elise Epaillard, Nizar Bouguila, and Jamal Bentahar. Modeling and forecasting time series of compositional data: A generalized dirichlet power steady model. In *International Conference on Scalable Uncertainty Management*, pages 170–185. Springer, 2015.

[38] Omar Graja and Nizar Bouguila. Finite inverted dirichlet mixture optimal pixel predictor. In *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5. IEEE, 2019.

[39] George G Tiao and Irwin Cuttman. The inverted dirichlet distribution with applications. *Journal of the American Statistical Association*, 60(311):793–805, 1965.

[40] Kai Wang Ng, Guo-Liang Tian, and Man-Lai Tang. *Dirichlet and related distributions: Theory, methods and applications*, volume 888. John Wiley & Sons, 2011.

[41] Jun Zhang and Dehong Ma. Nonlinear prediction for gaussian mixture image models. *IEEE transactions on image processing*, 13(6):836–847, 2004.

[42] Nizar Bouguila. Non-gaussian mixture image models prediction. In *2008 15th IEEE International Conference on Image Processing*, pages 2580–2583. IEEE, 2008.

[43] Rima Guidara, Shahram Hosseini, and Yannick Deville. Maximum likelihood blind image separation using nonsymmetrical half-plane markov random fields. *IEEE Transactions on Image Processing*, 18(11):2435–2450, 2009.

[44] Nizar Bouguila and Djemel Ziou. A new approach for high-dimensional unsupervised learning: applications to image restoration. In *International Conference on Pattern Recognition and Machine Intelligence*, pages 200–205. Springer, 2005.

[45] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and

measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[46] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010.

[47] GS Lingappaiah. On the generalised inverted dirichlet distribution. *Demostratio Mathematica*, 9(3):423–433, 1976.

[48] Arak Mathai and Panagis Moschopoulos. A multivariate inverted beta model. *Statistica*, 01 1997.

[49] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[50] Mohamed Al Mashrgy and Nizar Bouguila. A fully bayesian framework for positive data clustering. In *Artificial Intelligence Applications in Information and Communication Technologies*, pages 147–164. Springer, 2015.

[51] David Lindley and Calyampudi R. Rao. Advanced statistical methods in biometric research. 1953.

[52] Mohamed Al Mashrgy, Taoufik Bdiri, and Nizar Bouguila. Robust simultaneous positive data clustering and unsupervised feature selection using generalized inverted dirichlet mixture models. *Knowledge-Based Systems*, 59:182–195, 2014.

[53] Ali Mosleh, Nizar Bouguila, and Abdessamad Ben Hamza. Automatic inpainting scheme for video text detection and removal. *IEEE Trans. Image Process.*, 22(11):4460–4472, 2013.

[54] Ali Mosleh, Nizar Bouguila, and Abdessamad Ben Hamza. A video completion method based on bandlet transform. In *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo, ICME 2011, 11-15 July, 2011, Barcelona, Catalonia, Spain*, pages 1–6. IEEE Computer Society, 2011.

[55] Ines Channoufi, Sami Bourouis, Nizar Bouguila, and Kamel Hamrouni. Image and video denoising by combining unsupervised bounded generalized gaussian

mixture modeling and spatial information. *Multimedia Tools and Applications*, 77(19):25591–25606, 2018.

[56] John Aitchison. The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160, 1982.

[57] Teresa M Brunsdon and TMF Smith. The time series analysis of compositional data. *Journal of Official Statistics*, 14(3):237, 1998.

[58] Hirotugu Akaike. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. In *Selected Papers of Hirotugu Akaike*, pages 223–247. Springer, 1998.

[59] Masanao Aoki. *Optimal control and system theory in dynamic economic analysis*. North Holland publishing company Amsterdam, 1976.

[60] Masanao Aoki. *State space modeling of time series*. Springer Science & Business Media, 2013.

[61] EJ Hannan. The identification and parameterization of armax and state space forms. *Econometrica: Journal of the Econometric Society*, pages 713–723, 1976.

[62] Genshiro Kitagawa and Will Gersch. A smoothness priors–state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association*, 79(386):378–389, 1984.

[63] A Manuela Gonçalves, Olexandr Baturin, and Marco Costa. Time series analysis by state space models applied to a water quality data in portugal. In *AIP Conference Proceedings*, volume 1978, page 470101. AIP Publishing LLC, 2018.

[64] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski. Deep state space models for time series forecasting. In *Advances in neural information processing systems*, pages 7785–7794, 2018.

[65] Matthias Seeger, Syama Rangapuram, Yuyang Wang, David Salinas, Jan Gasthaus, Tim Januschowski, and Valentin Flunkert. Approximate bayesian inference in linear state space models for intermittent demand forecasting at scale. *arXiv preprint arXiv:1709.07638*, 2017.

[66] Matthias W Seeger, David Salinas, and Valentin Flunkert. Bayesian intermittent demand forecasting for large inventories. In *Advances in Neural Information Processing Systems*, pages 4646–4654, 2016.

[67] Nicolas Chapados. Effective bayesian modeling of groups of related count time series. *arXiv preprint arXiv:1405.3738*, 2014.

[68] Taoufik Bdiri and Nizar Bouguila. Bayesian learning of inverted dirichlet mixtures for svm kernels generation. *Neural Computing and Applications*, 23(5):1443–1458, 2013.

[69] Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

[70] Persi Diaconis and Donald Ylvisaker. Conjugate priors for exponential families. *The Annals of statistics*, pages 269–281, 1979.

[71] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[72] Zhanyu Ma, Arne Leijon, Zheng-Hua Tan, and Sheng Gao. Predictive distribution of the dirichlet mixture model by local variational inference. *Journal of Signal Processing Systems*, 74(3):359–374, 2014.

[73] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[74] Francesco Lombardi, Matteo Vincenzo Rocco, and Emanuela Colombo. A multi-layer energy modelling methodology to assess the impact of heat-electricity integration strategies: The case of the residential cooking sector in italy. *Energy*, 170:1249–1260, 2019.