Data-Driven Models for Reliability Prognostics of Gas Turbines

by

Gaurev Kumar

Submitted to the The Center for Computational Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Computation for Design and Optimization

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author .... Signature redacted ....
The Center for Computational Engineering
May 18, 2015

Certified by .... Signature redacted ...
Saurabh Amin
Assistant Professor, Civil & Environmental Engineering
Thesis Supervisor

Accepted by .... Signature redacted
Nicolas Hadjiconstantinou
Professor, Mechanical Engineering
Co-Director, Computation for Design and Optimization
Co-Director, Computational Science and Engineering

Data-Driven Models for Reliability Prognostics of Gas Turbines

by

Gaurev Kumar

Submitted to the The Center for Computational Engineering
on May 18, 2015, in partial fulfillment of the
requirements for the degree of
Master of Science in Computation for Design and Optimization

Abstract

This thesis develops three data-driven models of a commercially operating gas turbine, and applies inference techniques for reliability prognostics. The models focus on capturing feature signals (continuous state) and operating modes (discrete state) that are representative of the remaining useful life of the solid welded rotor. The first model derives its structure from a non-Bayesian parametric hidden Markov model. The second and third models are based on Bayesian nonparametric methods, namely the hierarchical Dirchlet process, and can be viewed as extensions of the first model. For all three approaches, the model structure is first prescribed, parameter estimation procedures are then discussed, and lastly validation and prediction results are presented, using proposed degradation metrics. All three models are trained using five years of data, and prediction algorithms are tested on a sixth year of data. Results indicate that model 3 is superior, since it is able to detect new operating modes, which the other models fail to do.

The turbine is based on a sequential combustion design and operates in the 50Hz wholesale electricity market. The rotor is the most critical asset of the machine and is subject to nonlinear loadings induced from three sources: i) day-to-day variations in total power generated by the turbine; ii) machine trips in high and low loading conditions; iii) downtimes due to scheduled maintenance and inspection events. These sources naturally lead to dynamics, where random (resp. forced) transitions occur due to switching in the operating mode (resp. trip and/or maintenance events). The degradation of the rotor is modeled by measuring the abnormality witnessed by the cooling air temperature within different modes. Generation companies can utilize these indicators for making strategic decisions such as maintenance scheduling and generation planning.

Thesis Supervisor: Saurabh Amin
Title: Assistant Professor, Civil & Environmental Engineering

Acknowledgments

Firstly, I would like to thank my advisor, Dr. Saurabh Amin for all of his support, guidance, and patience throughout my time at MIT. Our weekly discussions–those about mathematics and those not–will always be cherished.

I would like to also thank the Accenture and MIT Alliance in Business Analytics for their financial support for conducting this research. Collaborating with the entire Accenture team was very instructive and helpful. I would like to also thank the electricity production company which provided the data, allowing for analysis and simulations.

Finally, I would like to thank my mother, father, and entire family for always believing in me and encouraging me to strive for perfection, both of craft and mind.

Data analysis and simulations in this thesis were made possible by computational packages written and distributed by Dr. Emily B. Fox and Dr. Matthew J. Johnson.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, three probabilistic reliability modeling approaches are presented. The structures of all three models are discussed and their practical applications are analyzed using data from gas turbine (GT) assets. Although the focus in this thesis is GT-centric, the models are general enough to be applicable to other dynamical systems. The first model is non-Bayesian parametric (NBP) and the remaining two are Bayesian nonparametric (HDP). Results from the three approaches will be discussed and compared.

The models are developed for a real-world GT operating according to the Brayton thermodynamic cycle. The GT forms the part of a commercially operating Combined Cycle Gas Turbine (CCGT) plant, where the hot exhaust from the GT is used to power a steam power plant (operating according to the Rankine cycle). The CCGT operates in the 50 Hz wholesale electricity market, and is a part of the generation fleet of a large electricity producer in Europe. The electricity producer (i.e., the GenCo) continuously collects data from sensors and controllers that are embedded in the GT for the purpose of performance monitoring, maintenance scheduling, and generation planning. In particular, the data provided is a continuous stream of data from 153 sensor-control signals at 5 *min* intervals for a period of six years ($(2009-2014)$).

The objective is to model both intra-day and day-to-day dynamics of the feature signals that are indicative of the condition of GT's rotor, which is the most critical fixed asset of the machine. The day-to-day dynamics are modeled as random switches between a fixed number of *operating modes* (OMs), and are assumed to be generated from a discrete-time stochastic process. The intra-day dynamics of the asset's condition is assumed to evolve according to a linear regression model, dependent on the day's OM for NBP-HMM (model 1) and HDP-HSMM (model 2). In contrast, for the HDP-SLDS-HMM (model 3), the emission structure imbeds the regression model directly. Thus, the system can be viewed as a hierarchy, with each level observing dynamics with a different structure.

1.1 Gas Turbine Overview

Gas turbines are internal combustion engines that are heavily relied upon to generate mechanical energy with chemical energy inputs. A typical gas turbine has five standard components: inlet, compressor, diffuser, combustor, and turbine (with exhaust). Figure 1-1 presents a schematic of the GT used in this study. Gas turbines produce two streams of energy. The first energy stream is used to power an output shaft, while the second stream is used by the gas turbine, itself, to power the compressor. Gas turbines can be viewed as partially self-sustaining in that part of the energy produced is recycled to power internal components.

The process by which gas turbines generate energy can be viewed as a cycle which converts gaseous energy into mechanical energy. Large amounts of clean, unimpeded airflow is provided to the gas turbine by the air inlet. This air, which initially has atmospheric pressure is then converted into high-pressure air. With multiple stages of rotor blades and stator vanes, the compressor incrementally increases the impact pressure (velocity) of the supplied air. The diffuser, through its divergent duct design converts most of the impact pressure of the air output from the compressor into static pressure air. This low velocity, high static pressure air enters the combustor which burns the fuel-air mixture, while the cooling system and liners keep parts safe from the

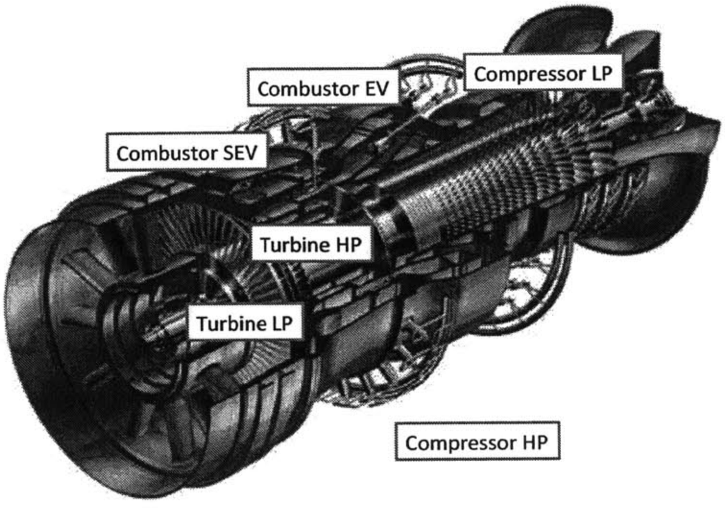

Figure 1-1: Gas turbine components

high flame temperatures produced from the combustion. The gaseous mixture is then run through a convergent duct which accelerates the gas, reducing the static pressure, as well as decreasing the temperature to feed into the turbine. The turbine has the "reverse" deign of the compressor, in that it extracts mechanical energy from the gaseous energy to drive the output shaft, and feeds power back into the compressor. After the gas has passed through the turbine, it exits through the exhaust [1].

The Brayton cycle summarizes the main thermodynamic processes that take place in a gas turbine. It begins with adiabatic compression in the inlet and compressor, followed by constant pressure fuel combustion. This is followed by adiabatic expansion in the turbine, with which some force is taken out of the air to drive the compressor, after which the remaining force is used to accelerate the output shaft. Finally, the cooling of the air at constant pressure returns the cycle to its initial condition [2].

1.1.1 Developing a GT Model

We would like to construct a model for the evolution of the degradation proxy signals by defining the dependency on observed feature signals and mode transitions. Because there are no maintenance activities that target the rotor directly, we cannot directly observe how long the rotor stays "healthy" before it drops below a certain degradation threshold. Instead we must infer, from the features (i.e. observable signals) that are correlated with the degradation of the rotor, how the dynamics of degradation of the rotor are changing. The data is obtained from sensors placed in various locations inside and in the ambient areas outside of the turbine. The sensors capture four types of readings: pressure, power, vibration, and temperature. The data can be organized in three buckets: process, refrigeration, and vibration. Process data refers to data collected by sensors on parts of components which are critical to maintaining thermodynamic parameters so that the turbine is working efficiently. Refrigeration data comes from sensors which are placed in ducts/ cavities in which cooling air passes through for neutralizing high temperatures generated by certain processes. And, vibration data is collected by sensors which measure movements of components in units of mm/s.

1.1.2 Intra-Modal Dynamics

The gas turbine Operation Concept (OC) dictates the governance of temperature and mass air flow of gas turbine, as a function of power output, using limit points. Because the OC is idealized, and not observed in practice, it allows us to define a distance metric characterizing how 'far' the systems operating conditions for a given period are from the idealized. More importantly, it induces a mapping from a particular operating condition to a particular operating mode (OM). At any given time, the system resides in a particular mode, completely characterizing the operating performance of the system at that time.

These operating conditions have a direct impact on the dynamics of the observable features that relate to the health state of the rotor, and by extension the evolution of the health state, itself.

1.1.3 Inter-Modal Transitions

We can view the GT system as one which dwells and transitions between finitely many OMs. The system dwells in exactly one state at any given time and it can transition from one state to another when a *trigger condition* is satisfied, based on the *active* transition law/kernel at any given time. As stated above, there exists a correspondence between the OMs and the operating condition of gas turbine. The transition law between different operating conditions must accommodate non-autonomous, as well as autonomous/stochastic transitions.

The system may transition between different operating conditions autonomously based on some probability distribution, as well as non-autonomously due to forced transitions that may occur because the state of the system (i.e. specific observed signals) reaches a pre-specified abnormal threshold, or due to a planned event that must take place a pre-specified time, which triggers a jump. We can define the transition law which causes the *switches* between modes as a Markov process.

1.2 Gas Turbine Reliability

The GT is governed by an operation concept (OC), which dictates how the machine should ideally perform its startups, normal operation, and shutdowns sequences. Degradation of the machine can be viewed as occurring when the machine operates in conditions that are very different from the OC.

1.2.1 Operation Concept

The machine is fired up when the first combustor, called the Environmental (EV) combustor, starts its operation. Pilot flames are used in the EV burners for initiating the combustion of fuel (gas) mixed with air. At 20 % of base load, the EV burners switch to premix operation, and a second combustor, called the Sequential Environmental (SEV), is ignited. At 40 % load, the stators (Variable Inlet Guide Vanes (VIGV)) are opened, and more fuel is supplied to the two combustors. The exhaust temperature of the turbine is kept constant by controllers, until full load is reached. This is achieved by means of a small increase in the combustor temperature with the VIGV fully open. Optimal operating conditions of the EV burners range from 25 % load to the full load.

The EV combustor is an annular combustion chamber with 30 burners. Combustion air enters the cone through air inlet slots while the fuel is injected through a series of fine holes in the supply pipe. The gaseous pilot fuel and the liquid fuel are injected through nozzles at the cone tip. This ensures that the fuel and air spiral into a vortex form and are mixed. The annular design provides an even temperature profile, resulting in improved cooling, longer blade life and lower emissions. The SEV combustor consists of 24 diffusor-burner assemblies, followed by a single, annular combustion chamber surrounded by convection-cooled walls. The exhaust gas from the high-pressure (HP) turbine enters the SEV combustor through the diffusor area. Due to the elevated temperatures of the HP turbine exhaust, the fuel-air mixture ignites autonomously. Finally, the low-pressure (LP) turbine operates at the tail end of the machine.

As the GT transitions from minimum-load to part-load to base load levels, various temperature limits are utilized in order to actuate appropriate controls to maintain GT efficiency and performance levels. These temperature limits are defined by the OC of the GT; see Fig. 1-2 where these limits are denoted TIT_1, TIT_2, TAT_1, and TAT_2. Essentially, these limits are the ideal levels of the aggregate temperatures of different subcomponents of the machine as a function of the power loading.

1.2.2 Rotor Degradation

As the GT gets older, most components begin to show signs of degradation. Understanding the degradation patterns of parts which are not easily replaceable–such as the rotor shaft–is of crucial importance, because they tend to be the most integral and expensive components. The principal components of a rotor are the shaft, disk, bearings, and seals. The bearings support the rotating components of the system and

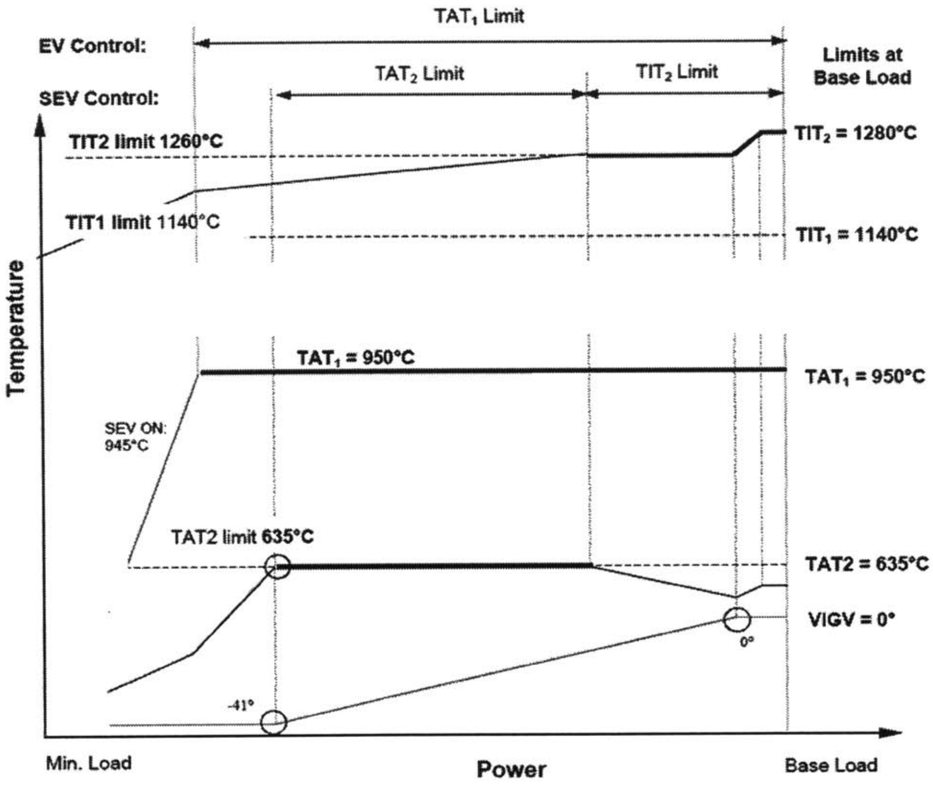

Figure 1-2: GT operation concept

stabilize the rotor vibration, while the seals prevent undesired leakage flows inside the machines of the processing or lubricating fluids [3].

Rotors can begin to degrade in many ways. The condition of the rotor is dependent on two major factors: creep and low-cycle fatigue (LCF). Due to constant loads, there is a risk of creep, which manifests itself in the propagation of cracks within the rotor. When the temperatures pass a certain percentage of the melting point of the material, the probability of creep is high. Furthermore, high temperature gradients, caused by poor casing insulation can lead to elastic rotor bending.

Additionally, LCF accelerates the degradation of the rotor, when it must undergo high amplitude, low frequency strains. This occurs mainly when the system is turned on and shut down multiple times in a short period of time [3]. The rapid switching between different electricity production levels shifts the rotation axis of the shaft by moving the mass center of the rotor. This leads to an increase in rotor vibration [4].

In addition to LCF, long exposure to high levels of mechanical stresses cause rotor shaft deterioration due to creep deformation. Creep also initiates cracks on the surface of the rotor. Under high temperature and pressure loadings that are typical of a GT, risk of creep induced deterioration increases. When the temperature surrounding the rotor surpasses 40 % of the melting point of the rotor, the probability of creep is especially high [2]. Other components of GT such as the disk, bearings, and seals are also subject to damage due to creep.

Rotor degradation dynamics are tied to the different OMs because the behavior of the GT varies drastically under different OMs. Dynamics of the rotor temperature signal differs when the system dwells in different OMs, and change depending on how the turbine transitions between each of these OMs. Specifically, as the turbine transitions from periods of normal operation characterized by loads of approximately 400 MW, to periods of suboptimal operation, to periods when the machine is being shut off, the dynamics of temperature that the rotor faces change. Thus, we can view the rotor temperature, as a time-evolving signal, whose dynamics are governed by the OM in which it dwells.

The mechanistic models of creep and LCF indicate that the variability of loads (stresses) and temperatures in various OMs contribute to progressive deterioration. In our work, random switching rates between the OMs are indicative of the variability in loads on GT. To determine the modes and mode-switching rates, the power output signal from the GT is used. To model the intra-day dynamics, the temperature of the cooling air that circulates in the ducts around the rotor is chosen as the proxy of the rotor temperature.

1.3 Related Work

Many publications have proposed various data-driven methods for estimating degradation trajectories of various subsystems within gas turbines. Yet, past research has dealt mostly with health state trajectory estimation of compressor blades and bearings. Statistical methods such as Monte Carlo, HMMs, neural networks, Gaussian mixture models, and various artificial intelligence methods using fuzzy logic have been used for estimation and prediction purposes [4,5]. Venturini et. al presented a prognostic methodology, which estimates remaining useful life of GTs using statistical techniques, by estimating the propensity of a turbine to transition from an "operable" to "inoperable" state [5]. By sampling the intervals of time that the turbine remained in any one of these states, Weibull distributions are generated, from which simulations are conducted to compute the probability of dwelling in the operable mode. In contrast, the methods we describe automatically learns the number of OMs, without having to a priori posit a number of states. This allows for model flexibility.

The transition matrix/kernel, which governs how OMs switch among one another can be defined deterministically or probabilistically, depending on the application.

Since the OMs in which the GT dwells are tied to random fluctuations induced by the bulk electricity market, we posit a model in which the OMs switch according to a probabilistic relation. This facet has been explored in existing literature which apply Piecewise Deterministic Markov Processes (PDMPs) for reliability prognostics [7,8,9]. Davis introduced as the most general class of continuous-time Markov processes which include both discrete valued and continuous valued processes, except diffusion [7]. A PDMP consists of two components: a discrete component and a continuous component. Here, the OMs take discrete values and the rotor temperature is the continuous valued solution of a OM-dependent difference equation. At discrete moments in time, the discrete valued component may switch or jump to according to specified relation.

Numerous system identification methods are also available for parameter estimation of intra-OM models [10,11]. Morari et al. propose a method of identifying hybrid systems that are assumed to be piecewise affine (PWA). In this case, the OMs map to a finite sequence of polyhedra which partition the space in which the continuous variable take values [10]. Within each of these polyhedra, an affine difference equation governs the evolution path of the continuous variable. And, the transition law is completely determined by the system reaching the boundary of the polyhedra. The algorithm that Morari et al. proposes makes use of support vector machines, and clustering in order to estimate the intra-OM models and the transition law (boundaries of the polyhedra).

1.4 Contributions & Outline

The pragmatic goal of this work is to understand both modeling approaches (NBP and HDP) and comment on how both differ in estimation and prediction power of the condition of critical assets (e.g., rotors, stators, casings) of large machines such as gas turbines, steam turbines, and generators. The operational flexibility of GenCos can be significantly affected if they have limited visibility on the rate at which their critical assets are deteriorating. In recent years, GenCos world-wide are developing a sense of urgency toward prognostic assessment of their generation fleet. The main reason of this urgency is the fact that although the frequency of complete failure of critical assets might be quite low, their replacement costs and delivery times can range from months to years. Ignoring the early indicators of deterioration can lead to huge unexpected economic losses. Moreover, if a large-side generation unit faces disruption due to a failure of a critical asset, the GenCo may even loose its strategic position in the competitive wholesale market.

Especially important is to be able to account for the supply-demand shocks and non-stationarity in production trends [1]. For example, the power plants are expected to have the capability to execute quicker start-up times and switch between one operating point to another when another generator trips and creates a sudden loss of supply. The push for flexible ramp-up and ramp-down schedules is even higher in markets that are gradually accommodating new renewable energy sources. Additional impositions on large machines such as compliance with environmental standards (e.g., cap on emissions) will make their future operating environment even more stringent. Thus, we expect that the machines' critical assets will be subject to new loadings and new temperature and pressure variations. This, in turn, will directly effect their rate of deterioration. Due to the aforementioned reasons, utilizing the available data for building accurate dynamical models of deterioration indicators (i.e., the condition) is an important task, as confirmed by many business leaders in large-scale electricity production [1].

Previous work in modeling deterioration of machine components has primarily focused on replaceable components (e.g., blades within turbine's compressors) [4, 5]. Since these components have shorter lifetimes, the available condition monitoring data can capture their deterioration rates, starting from the time of their installation to the time of their replacement. Major GT manufacturers (e.g., Alstom, GE Energy, Siemens, United Technologies) are fairly specific in recommending the maintenance plans of replaceable components. However, estimating deterioration rates of rotating components such as rotor shaft is not straightforward, as they typically have much longer lifetimes (20-30 years). Hence, data on their actual failure rates is not readily available. GT manufacturers conduct laboratory tests on these assets during design phase, but there is no direct way to conduct tests on these assets once they are installed in an operational machine. Thus, currently there is no practical procedure to estimate the remaining useful life (RUL) of GT rotors.

In chapter 2, we present necessary background on the three models that we will be presenting. The chapter begins with a brief overview of probability distributions and mixture models, followed by structures of parametric and nonparametric hidden Markov models (HMM). In chapter 3, we present the non-Bayesian parametric HMM for rotor degradation. A new detrending method is presented that allows for dimensionality reduction of the HMM emission variable. In chapter 4, we present two Bayesian nonparametric models. The first model is a semi-Markov model, and is able to estimate dwell times in different operating modes, using the exponential distribution. The second model uses an Switching Linear Dynamical System (SLDS) emissions model, which is able provide a richer model structure. In chapter 5, we posit metrics for degradation that are later used to validate all three models, and compare performance. Calculations and comparisons of the cumulative deterioration estimated over the time are presented. The chapter concludes with a discussion of different prediction algorithms, which are tested on the sixth year of data. Finally, concluding remarks end the thesis in chapter 6.

The main contributions of this thesis are the three models. Model 1 is the parametric non-Bayesian hidden Markov model (NBP-HMM). In model 1, each day is classified as an operating mode (OM) by the HMM with a detrended intra-day power signal as the emission. For this model, the number of OMs is assumed to be 3. Model 2 is the nonparametric Bayesian hidden semi-Markov model (HDP-HSMM). In model 2, each day is classified as an operating mode (OM), but the emissions are now the complete intra-day power signals, distributed multivariate gaussian. For this model, the number of OMs learned by the model itself, utilizing an imbedded Dirichlet process. For models 1 and 2, once the OMs are learned, a linear regression model is fit for the cooling air temperature signal, for each OM. Using the residuals from the fitted regression model, deterioration levels are estimated.

Model 3 is the nonparametric Bayesian HMM with a switching linear dynamical system emission model (HDP-SLDS-HMM). In model 3, each day is classified as an operating mode (OM), but the SLDS emission structure allows for simultaneous fitting of intra-day temperature signal. Thus, the regression model is not needed to estimate degradation for model 3, as it was necessary for models 1 and 2.

Chapter 2

Parametric & Nonparametric Models

2.1 Background

This chapter begins by describing common probability distributions that will be used as basic building blocks for the models that we consider in this thesis. This is followed by descriptions of more complex probabilistic graphical models that will ultimately be tailored specifically for the application of GT reliability. For all of the models described, both Bayesian and non-Bayesian forms exist. In a Bayesian framework, appropriate prior distributions are placed on model parameters, while in a non-Bayesian framework, parameters are assumed to be fixed.

2.1.1 Probability Distributions

A. Categorical

The categorical distribution is a generalization of the Bernoulli distribution, and is also called a "discrete distribution" [17]. It is a distribution that describes the result of a random event that can take on one of k possible outcomes, with the probability of each outcome separately specified. If x is a random variable distributed categorical, with parameters $(p_1, \dots, p_k)$, such that $\sum_{i=1}^{k} p_i = 1$, then $x \in \{1, \dots, k\}$
and

$$p(x) = \prod_{i=1}^{k} p_i^{[x=i]} \tag{2.1}$$

Here, $[x = i]$ denotes the Iverson bracket, where

$$[x = i] = \begin{cases} 1 & x = i \\ 0 & otherwise \end{cases} \tag{2.2}$$

B. Binomial/ Multinomial

The binomial distribution is the distribution governing the number of successes for one of just two categories in n independent Bernoulli trials, with the same probability of success on each trial [17]. The multinomial distribution is the generalization of the binomial distribution, where each trial results in exactly one of some fixed finite number k possible outcomes, with probabilities $(p_1, \ldots, p_k)$, such that $\sum_{i=1}^{k} p_i = 1$. If the random variables x_i indicate the number of times outcome number i is observed over the n trials, the vector $x = (x_1, ..., x_k)$ follows a multinomial distribution with parameters n and p, where $p = (p_1, \ldots, p_k)$, where

$$p(x) = p(x_1, \ldots, x_k) = \frac{n!}{x_1! \ldots x_k!} \prod_{i=1}^{k} p_i^{x_i}. \tag{2.3}$$

If $n = 1$, the categorical distribution is obtained.

C. Dirichlet

The Dirichlet distribution is a family of continuous multivariate probability distributions parameterized by a vector of positive reals [13]. It is the multivariate generalization of the beta distribution. Dirichlet distributions are very often used as prior distributions in Bayesian statistics, and is the conjugate prior of the categorical distribution and multinomial distribution. The Dirichlet distribution of order $k \geq 2$

with parameters $\alpha = (\alpha_1, ..., \alpha_K) \geq 0$ has the following probability density function

$$f(x) = f(x_1, \dots, x_k) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} x_i^{\alpha_i - 1} \tag{2.4}$$

Here, $B(\alpha)$ is the multinomial beta function, which is expressed through the gamma function as

$$B(\alpha) = \frac{\prod_{i=1}^{k} \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^{k} \alpha_i)} \tag{2.5}$$

The infinite-dimensional generalization of the Dirichlet distribution is the Dirichlet process, which will be described in more detail later in this chapter.

D. Multivariate Gaussian

The multivariate Gaussian distribution is a generalization of the univariate Gaussian distribution [17]. The multivariate Gaussian distribution is said to be "non-degenerate" when the symmetric covariance matrix Σ is positive definite. In this case, the k-dimensional a Gaussian-distributed random variable $x = (x_1, \dots, x_k)$ with mean vector μ and covariance Σ has probability density

$$f(x) = f(x_1, \dots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} exp\Big(- \frac{1}{2}(x - \mu)^t \Sigma^{-1} (x - \mu)\Big) \tag{2.6}$$

E. Exponential

The exponential distribution is the probability distribution that describes the inter-arrival time between different events in a Poisson process. The Poisson process is a stochastic process in which events occur independently at the same average rate [17]. It is the continuous analog of the geometric distribution, and is notedly memoryless. The probability density function (pdf) of an exponential distribution is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{2.7}$$

2.1.2 Mixture Model

A mixture model is best explained by thinking about a system from which taking measurements is a simple procedure [13]. In some cases, the state of the system, at the time when the measurement is taken is known. Yet, for real-world systems, which are complex, the state of the system is usually unknown. Given a sequence of measurements $\{y_1, y_2, \ldots, y_N\}$, a mixture model posits that the data points are sampled from a finite (or countably infinite) mixture of unobserved states z taking values in $\mathcal{K} = \{1, 2, \ldots, K\}$. In the case of parametric/finite mixture models, the cardinality of $\mathcal{K}$ is assumed known. In the case of nonparametric mixture models, the theoretical cardinality is assumed to be countably infinite, yet empirically a finite number of states will exist. Because the states are unobserved, the data points are used to infer two major qualities of the system. Firstly, each observation is modeled as a random variable, with an underlying distribution coupled with a unique state. Thus, given the observations, parameters governing the underlying distribution can be inferred. And, secondly, the mixture model assumes a prior on the relative importance of each state, which must be inferred, as well. Finally, given knowledge of an observations generating state, the data are assumed to be independent. The exhibit below outlines a Bayesian parametric mixture model.

Bayesian Parametric Mixture Model

K: Number of states (mixture components) $\{z_1, \ldots, z_N\} \in \mathcal{K} = \{1, \ldots, K\}$
N: Number of measurements $\{y_1, \ldots y_N\} \in \mathbb{R}^n$
F: Probability distribution governing observation y_i
θ_k: Parameter governing distribution of observations emitted from kth state
β_k: Mixture weight for kth state. $\mathcal{B} = (\beta_1, \ldots, \beta_K)$
λ: Hyper-parameter governing prior distribution of $\mathcal{B}$

$H(\gamma)$: Prior distribution with hyper-parameter of θ_k for all $k \in \mathcal{K}$

$$\theta_k \sim H(\gamma) \tag{2.8}$$

$$\mathcal{B} \sim Dirichlet(\lambda) \tag{2.9}$$

$$z_i \sim Categorical(\mathcal{B}) \tag{2.10}$$

$$y_i \mid z_i = k \sim F(\theta_k) \tag{2.11}$$

$$\sum_k \beta_k = 1 \ ; \ \beta_k \geq 0 \ \forall \ k \in \mathcal{K} \tag{2.12}$$

$$\tag{2.13}$$

2.2 Parametric Models

2.2.1 Finite Hidden Markov Model

The hidden Markov model (HMM) prescribes a probability distribution over a sequence of observations [13]. The first assumption of the model is that the sequence of observations y_t are sampled at discrete times $t \in \{1, \ldots, T\}$. The second assumption is that the sequence of observations, commonly referred to as "emissions", are produced from a hidden/latent process z_t, at discrete times t. The third assumption is that the hidden process observes the Markov property. The fourth assumption is that the observations in the emission process are independent, conditional on the hidden process, namely

$$p(y_t \mid z_{1:T}, y_{1:T}) = p(y_t \mid z_t) \tag{2.14}$$

With these four assumptions, we can write the joint probability distribution as

$$p(y_{1:T}, z_{1:T}) = p(z_1)p(y_1 \mid z_1) \prod_{t=2}^{T} p(z_t \mid z_{t-1})p(y_t \mid z_t) \tag{2.15}$$

With the model specified, the goal is to use data to infer the parameters of the HMM. The parameters are a triple $\{F, \tilde{\pi}, \tilde{\pi}_0\}$, the emissions distribution parameters (i.e,

$\{\mu_k, \Sigma_k\}$, for Gaussian distributed emissions) , the $K \times K$ time-invariant transition matrix governing the hidden process dynamics, and the initial distribution for the hidden process, respectively. The exhibit below outlines a Bayesian parametric hidden Markov model.

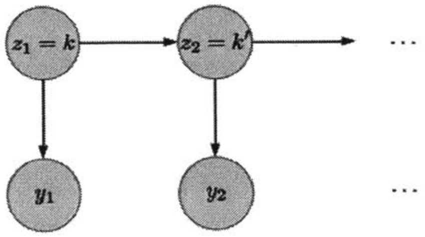

Figure 2-1: Hidden Markov Model

Bayesian Parametric Hidden Markov Model

K: Number of states (mixture components) $\{z_1, \ldots, z_T\} \in \mathcal{K} = \{1, \ldots, K\}$
$(1, \ldots, T)$: Sequence of measurements $\{y_1, \ldots y_T\} \in \mathbb{R}^n$
F: Probability distribution governing observation y_t
θ_k: Parameter governing distribution of observations emitted from kth state
$\tilde{\pi}_{0k}$: Initial probability of kth state. $\tilde{\pi}_0 = (\tilde{\pi}_{01}, \ldots, \tilde{\pi}_{0K})$
$\tilde{\pi}_k$: Transition probability distribution of state k. $\tilde{\pi}_k = (\tilde{\pi}_{k1}, \ldots, \tilde{\pi}_{kK})$
λ: Hyper-parameter governing prior distribution of $\mathcal{B}$
$H(\gamma)$: Prior distribution with hyper-parameter of θ_k for all $k \in \mathcal{K}$

$$\theta_k \sim H(\gamma) \tag{2.16}$$

$$\tilde{\pi}_k, \tilde{\pi}_0 \sim Dirichlet(\lambda) \tag{2.17}$$

$$z_1 \mid \tilde{\pi}_0 \sim \tilde{\pi}_0 \tag{2.18}$$

$$z_t \mid z_{t-1} = k \sim \tilde{\pi}_k \tag{2.19}$$

$$y_t \mid z_t = k \sim F(\theta_k) \tag{2.20}$$

$$\sum_k \tilde{\pi}_{0k} = 1 \; ; \; \tilde{\pi}_{0k} \geq 0 \; \forall \; k \in \mathcal{K} \tag{2.21}$$

$$\sum_j \tilde{\pi}_{kj} = 1 \; ; \; \tilde{\pi}_{kj} \geq 0 \; \forall \; j \in \mathcal{K} \tag{2.22}$$

$$\tag{2.23}$$

2.2.2 Hidden Semi-Markov Model

Although the parametric HMM is capable of modeling a variety of data structures, it has deficiencies. One of the main drawbacks of the HMM is its Markovian assumption. This assumption posits a model in which the latent states observe strictly geometrically distributed dwell times. Depending on the application, this assumption may or may not hold. This limitation leads to improving the HMM to the hidden semi-Markvov model (HSMM) [18]. There are many types of HSMMs, differing on assumptions made about how durations in states are distributed. We will limit our discussion to the "explicit duration" HSMM, in which each state's dwell duration is given an explicit distribution.

The workings of an HSMM are very similar to the HMM, except for an additional random variable which models the amount of time a state is dwelled in, denoted D_t. The distribution of D_t is state-dependent. When the state is entered (using the Markov chain assumptions identical to the HMM), the duration time is drawn from the distribution assigned to the particular state. The system dwells in this state until expiration and the process repeats. Similar to the HMM, at each time step that the system dwells in a particular state an observation is generated, depending on a state-specific emission probability distribution.

The HSMM has three layers/sequences of random variables. The first layer is a sequence of "super-states" z_s. Each super-state takes a discrete value. The index s is at a strictly lower frequency than t, and encapsulates a series of time steps $\{t_s^1, \dots, t_s^2\}$. The next layer is the "label" sequence x_t. x_t takes on the value of the corresponding time's super-state. The final layer is the "emission" sequence y_t, which is generated by a state-specific emission probability distribution. It is assumed that the HSMM has no self-transitions, so that D_t can be interpreted easily. The exhibit below outlines a Bayesian parametric hidden semi-Markov model.

Parametric Hidden Semi-Markov Model

K: Number of states (mixture components) $\{x_1, \dots, x_T\} \in \mathcal{K} = \{1, \dots, K\}$
$(1, \dots, T)$: Sequence of measurements $\{y_1, \dots y_T\} \in \mathbb{R}^n$
$(1, \dots, S)$: Sequence of "super-states" $\{z_1, \dots z_S\} \in \mathcal{K}$
$(1, \dots, S)$: Sequence of durations $\{D_1, \dots D_S\} \in \mathbb{R}$
F: Probability distribution governing observation y_t
θ_k: Parameter governing distribution of observations emitted from kth state
ω_k: Parameter governing distribution of dwell time for kth state
$\tilde{\pi}_k$: Transition probability distribution of state k. $\tilde{\pi}_k = (\tilde{\pi}_{k1}, \dots, \tilde{\pi}_{kK})$
$H(\gamma)$: Prior distribution with hyper-parameter of θ_k for all $k \in \mathcal{K}$
G: Prior distribution of ω_k for all $k \in \mathcal{K}$

$$\theta_k \sim H(\gamma) \tag{2.24}$$

$$\tilde{\pi}_k, \tilde{\pi}_0 \sim Dirichlet(\lambda) \tag{2.25}$$

$$z_1 \mid \tilde{\pi}_0 \sim \tilde{\pi}_0 \tag{2.26}$$

$$z_s \mid z_{s-1} = k \sim \tilde{\pi}_k \tag{2.27}$$

$$D_s \mid z_s \sim G(\omega_{z_s}) \tag{2.28}$$

$$t_s^2 = t_s^1 + D_s - 1 \tag{2.29}$$

$$x_{t_s^1:t_s^2} = z_s \tag{2.30}$$

$$y_t \mid x_t = k \sim F(\theta_k) \tag{2.31}$$

$$\sum_k \tilde{\pi}_{0k} = 1 \ ; \ \tilde{\pi}_{0k} \geq 0 \ \forall \ k \in \mathcal{K} \tag{2.32}$$

$$\sum_j \tilde{\pi}_{kj} = 1 \ ; \ \tilde{\pi}_{kj} \geq 0 \ \forall \ j \in \mathcal{K} \tag{2.33}$$

2.3 Nonparametric Models

2.3.1 Dirichlet Process

The Dirichlet process (DP) is a distribution over a function space, where the functions are probability measures. The defining characteristic of this family of probability measures, is that each of these probability measures have a countably infinite support. More precisely, a DP, denoted $DP(\gamma, H)$ is a distribution on probability measures on

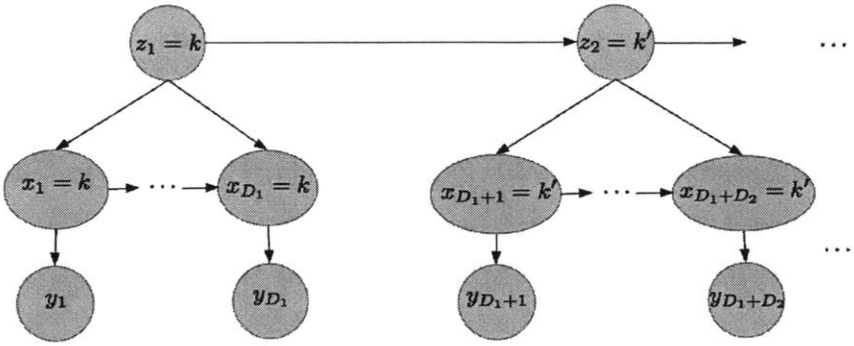

Figure 2-2: Hidden Semi-Markov Model

a measurable space Θ. It is defined by a base measure H on Θ and concentration parameter γ. Consider a finite partition $\{\Theta_1, ..., \Theta_K\}$ of Θ, such that:

$$\cup_{k=1}^{K} \Theta_k = \Theta \tag{2.34}$$

$$\Theta_i \cap \Theta_j = \emptyset \ for \ i \neq j \tag{2.35}$$

Then, probability measure G_0 on Θ is a draw from a DP if its measure on every finite partition follows a Dirichlet distribution, such that:

$$(G_0(\Theta_1), \dots G_0(\Theta_K)) \mid \gamma, H \sim Dirichlet(\gamma H(\Theta_1), \dots, \gamma H(\Theta_k)) \tag{2.36}$$

Here Θ is to be interpreted as a parameter space (i.e., each element $\theta \in \Theta$ is a unique parameter). Each element $\theta \in \Theta$ is assumed to be in a unique partition set $\Theta_k \in \Theta$ which is *H-measurable*. Observe then that $(\gamma H(\Theta_1), \dots, \gamma H(\Theta_k))$ are a set of probabilities, weighted by γ, where $\sum_k H(\Theta_k) = 1$ and $H(\Theta_k) \geq 0$ for all k. For every finite partition, H and γ (through this weighted set of probabilities) fixes a Dirichlet distribution (a probability measure over discrete probability measures with support being the partition set indexes).

G_0 can be thought of weighting regions of Θ "proportional" to H. One can think of fixing different finite partitions of Θ. G_0 maps each partition of length K to a $K-$dimensional vector whose elements sum to one, and are all non-negative. Also, if one fixes set $\Theta_k \in \Theta$, and fix different partitions containing Θ_k, it is expected that the average value that G_0 gives to Θ_k will be $H(\Theta_k)$. In fact, $\mathbb{E}[G_0(\Theta_k) \mid H] = H(\Theta_k)$ [19].

A "stick-breaking" construction, devised by Sethuram is a pedagogical description of the DP which will now be briefly presented. G_0 is a random probability measure distributed $DP(H, \gamma)$. A question that arises is given a sequence of samples $\{\theta'_1, \ldots, \theta'_N\}$, where $N \to \infty$, how does the posterior distribution of G_0 change. It will turn out that as the number of samples limits to ∞, G_0 will only have non-zero mass on finitely many points. The reason is because as more and more samples are produced, the posterior weighs the base measure less and less (i.e. new parameters stop being spawned), while previously spawned parameters are revisited more frequently. Sethuram proved that a realization of $G_0 \sim DP(H, \gamma)$ is actually a discrete probability measure with probability one [22]. Let $\{\beta_k\}_{k=1}^{\infty}$ be a probability mass function on a countably infinite set where the discrete probabilities are defined as follows:

Dirichlet Process: Stick-breaking Construction

N: Number of measurements $\{y_1, \ldots y_N\} \in \mathbb{R}^n$
$H(\lambda)$: Base Measure with hyper parameter λ
γ: Concentration parameter
β_k: Weight for kth component; $\{z_1, \ldots, z_N\} \in \mathcal{K} = \{1, 2, \ldots\}$
θ_k: Parameter governing distribution of observations emitted from kth component
v_k: Auxiliary variable for kth component
F: Probability distribution governing observation y_i

δ_{θ_k}: Unit-mass measure concentrated at θ_k

$$v_k \mid \gamma \sim Beta(1, \gamma) \tag{2.37}$$

$$\beta_k = v_k \prod_{j=1}^{k-1} (1 - v_j) \tag{2.38}$$

$$\theta_k \mid H, \lambda \sim H(\lambda) \tag{2.39}$$

$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k} \tag{2.40}$$

G_0 is denoted as β (i.e., "$\beta \sim GEM(\gamma)$"). Observations are generated:

$$\beta \sim GEM(\gamma) \tag{2.41}$$

$$z_i \mid \beta \sim \beta \tag{2.42}$$

$$y_i \mid z_i, \{\theta_k\}_{k=1}^{\infty} \sim F(\theta_{z_i}) \tag{2.43}$$

$$\tag{2.44}$$

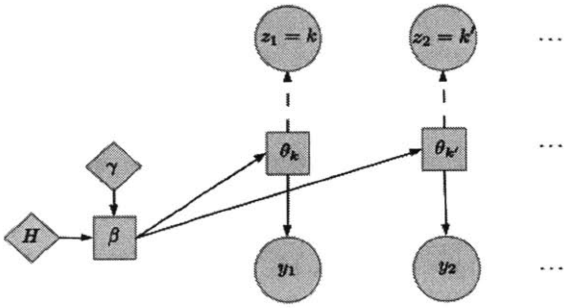

Figure 2-3: Dirichlet Process

2.3.2 Hierarchical Dirichlet Process

A Hierarchical Dirichlet process (HDP) is an extension of a Dirichlet process, used to model groups of data, which are assumed to be generated from a common process, yet have idiosyncrasies unique to each group. This structure is induced by defining different group-specific distributions, which will all be generated by a single Dirichlet process. Each group-specific distribution G_j (for the jth group).

$$G_j \sim DP(\alpha, G_0) \; \forall \; j \in \mathcal{J} = \{1, \dots, J\} \tag{2.45}$$

All of the groups are tied together by the same Dirichlet process G_0.

$$G_0 \sim DP(\gamma, H) \tag{2.46}$$

Given this structure, if one fixes the set $A \in \Theta$, it is expected that the average measure that G_j will give A will the amount that G_0 gives to A will be $G_0(A)$. In fact, for every $A \subset \Theta$, $\mathbb{E}[G_j(A) \mid G_0] = G_0(A)$.

The Stick-breaking construction of the HDP is now presented. Let $\{y_{j1}, \dots, y_{jN_j}\}$ be the set of observations for group j. This extends the stick-breaking representation from the previous section for the hierarchical case. The key fact about an HDP is that the atoms θ_k are shared not only within groups, but also between groups. G_0 establishes an unbounded support of parameters, from which the J groups share their support. In fact, there exists a non-zero probability the different G_j share support points. This is possible because all of the groups are tied to G_0. The exhibit below outlines the Stick-breaking construction of the HDP.

Hierarchical Dirichlet Process: Stick-Breaking Construction

$G_0(\gamma, H(\lambda))$: Dirichlet process with parameters $\alpha, H(\lambda)$
β_k: Weight for kth component, $k \in \mathcal{K} = \{1, 2, \dots\}$
θ_k: Parameter governing distribution of observations emitted from kth component

$$\beta \mid \gamma \sim GEM(\gamma) \left(G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\theta_k} \right) \tag{2.47}$$

$$\theta_k \mid H, \; \lambda \sim H(\lambda) \quad k = 1, 2, \dots \tag{2.48}$$

J: Number of groups, $j \in \mathcal{J} = \{1, \dots, J\}$
G_j : Dirichlet process with parameters α, β

$\tilde{\pi}_{jk}$: Weight for kth component of jth group

$$\left(G_j = \sum_{t=1}^{\infty} \tilde{\pi}_{jk} \delta_{\theta_k}\right) \quad j = 1, \ldots, J \tag{2.49}$$

$$\tilde{\pi}_j \mid \alpha, \beta \sim DP(\alpha, \beta) \tag{2.50}$$

N_j: Number of measurements for jth group $\{y_{j1}, \ldots y_{jN_j}\} \in \mathbb{R}^n$
z_{ji}: Indicator random variable signifying component of y_{ji}, $\{z_{j1}, \ldots, z_{jN_j}\} \in \mathcal{K}$
F: Probability distribution governing observation y_{ji}

$$z_{ji} \mid \tilde{\pi}_j \sim \tilde{\pi}_j \tag{2.51}$$

$$y_{ji} \mid \{\theta_k\}, z_{ji} \sim F(\theta_{z_{ji}}) \tag{2.52}$$

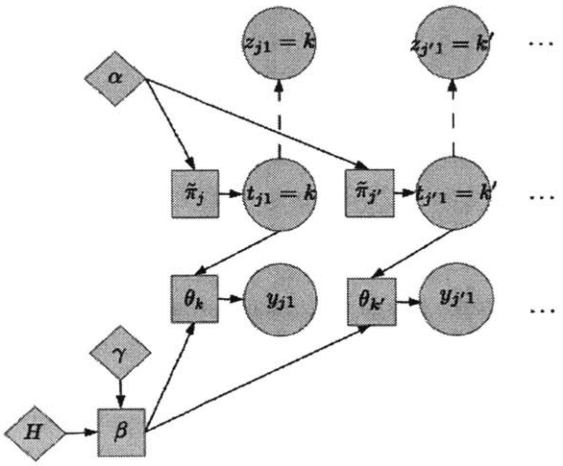

Figure 2-4: Hierarchical Dirichlet Process

Teh, et al. metaphorically compares the HDP structure to a "Chinese restaurant franchise" (CRF) [21]. The CRF is made up of J restaurants, each corresponding to an HDP group. Within each restaurant is an infinite buffet of dishes corresponding to parameters $\{\theta_k\}_{k=1}^{\infty}$. This set of dishes is common among all restaurants.

Each customer, corresponding to observation y_{ji} is pre-assigned to a given restaurant determined by that customer's group j. Upon entering the jth restaurant in the CRF, the customer, corresponding to data point y_{ji} (ith customer in the jth restaurant) sits at one of the currently occupied tables with probability proportional to the number of people sitting at that table (this distribution is denoted $\tilde{\pi}_j$) or creates a new table T_j+1 (within the jth restaurant) with probability α. The table that customer y_{ji} ends up sitting at is denoted t_{ji}.

Whenever a customer is the first customer to sit at a table (implying it is a newly spawned table), in any of the J restaurants, the customer goes to the buffet line and chooses from the current set of K dishes, corresponding to $\{\theta_k\}_{k=1}^{K}$ with probability proportional to to the number of times dish k has been selected across all tables in the franchise or orders a new dish with probability γ (this distribution is denoted β). Once the dish is selected, it is fixed for that table. Once the dish being served at the table t_{ji} (denoted by parameter $\theta_{k_{jt_{ji}}}$) seating y_{ji} is known, y_{ji} is generated by a distribution F with parameter $\theta_{k_{jt_{ji}}}$.

2.3.3 HDP-HMM

Teh, et. al. also presented a formulation of HMMs, extending the HDP (called "HDP-HMM") as a prior distribution on transition matrices over countably infinite state spaces [21]. In the traditional HDP, explained in the previous section, the data points y_{ji} are pre-partitioned into J groups. Once partitioned, depending on their group-specific DP G_j, a parameter θ_{ji}, is chosen, deeding on weights $\tilde{\pi}_k$, induced by G_j. The support of parameters is common across all group-specific DPs.

In contrast, the HDP-HMM posits an equivalence between the groups and parameters. To clarify, assume the previous state $z_{t-1} = k'$. k' designates the distribution from which z_t will be chosen, i.e. $z_t \sim \tilde{\pi}_{k'}$. Given θ_{z_t}, y_t is generated by $F(\theta_{z_t})$. So, now, the groups are equivalent to the parameters, unlike the traditional HDP. To denote this significance, we now subscript the parameters θ with j, instead of k, as done previously.

One serious limitation of the standard HDP-HMM is that it inadequately models the temporal persistence of states. Fox subsequently introduced an extension: *sticky HDP-HMM*, which reduces an HDP-HMM's tendency to rapidly switch between states [19]. This is accomplished by augmenting the HDP-HMM to include a parameter κ that increases the probability of self-transition and a separate prior on this parameter. Below, we present an exhibit that outlines the HDP-HMM.

HDP-HMM

β : Dirichlet process with parameters $\gamma, H(\lambda)$
θ_j: Parameter governing distribution of observations emitted from jth component, (The group and component index sets are now equal: $\mathcal{J} = \mathcal{K} = \{1, 2, \dots\}$)
$\tilde{\pi}_j$: Dirichlet process with parameters α, β (state-specific transition distribution for state j)
$(1, \dots, T)$: Sequence of measurements $\{y_1, \dots y_T\} \in \mathbb{R}^n$
z_t: Indicator random variable signifying component of y_t, $\{z_1, \dots, z_T\} \in \mathcal{K}$
F: Probability distribution governing observation y_t

$$\beta \mid \gamma \sim GEM(\gamma) \tag{2.53}$$

$$\theta_j \mid H, \lambda \sim H(\lambda) \quad j = 1, 2, \dots \tag{2.54}$$

$$\tilde{\pi}_j \mid \beta, \alpha \sim DP(\alpha, \beta) \tag{2.55}$$

$$z_t \mid \{\pi_j\}_{j=1}^{\infty}, z_{t-1} \sim \pi_{z_{t-1}} \quad t = 1, \dots, T \tag{2.56}$$

$$y_t \mid \{\theta_j\}_{j=1}^{\infty}, z_t \sim F(\theta_{z_t}) \quad t = 1, \dots, T \tag{2.57}$$

Sticky HDP-HMM

κ : Self-Transition bias parameter

$$\pi_j \mid \beta, \alpha, \kappa \sim DP\left(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_j}{\alpha + \kappa}\right) \quad j = 1, 2, \dots \tag{2.58}$$

2.4 Inference

Bayesian methods for parameter inference are adopted from two perspectives. The first perspective is based on domain knowledge of the problem. Here, some prior knowledge of the parameters are known. Typically, this would entail a fixed distribution of the parameters, known as a *prior.* In this setting, bayesian inference is straightforward, since it combines information gathered from data, *likelihood* to "update" the prior , producing a *posterior* distribution over the parameters. This posterior distribution is somehow "better" than the prior, since it takes into account the data points.

There is also the pragmatic perspective. Here, the prior distribution over parameters is used to make inference tractable. The issue of tractable inference often motivates the use of *conjugate* priors. in which the prior and posterior distributions are of the same form (i.e. in *exponential family*). Along with tractability, the Bayesian framework allows for models to flexible, namely hierarchical, with multiple layers. In this case, priors do not inherently contain information, but their parametric form will be tailored for tractability. This is why *hyper parameters*, parameters governing the distributions of the parameters governing the prior are typically inferred, as well.

Bayesian inference, gets its name from Bayes' rule. We see how important conditional distributions are by examination of this rule. In fact, the likelihood and posterior distributions are both conditional distributions. Typically, the assumptions a probabilistic model makes in regards to conditional distributions about the random variables involved, can be easily depicted using *graphical models.* Graphical models use nodes to represent random variables and parameters in the probabilistic model and vertices to depicted probabilistic relations between them. We will also see that the structure of the conditional dependencies between the random variables and the parameters of the model is precisely what computational methods such as *Gibbs sampling* take advantage of to iteratively compute parameter estimates.

2.4.1 Weak Limit Approximation

In terms of Gibbs sampling the posterior distributions of the Dirichlet process, it is desirable to maintain a finite approximation to the Dirichlet process mixture model. One approach to producing such a finite approximation is simply to terminate the Stick-breaking construction after some portion of the stick has already been broken and assign the remaining weight to a single component. This approximation is referred to as the truncated Dirichlet process.

Let us assume that there are L components in a finite mixture model and we place a finite-dimensional, symmetric Dirichlet prior on these mixture weights:

$$\beta \mid \gamma \sim Dir(\gamma/L, \ldots, \gamma/L) \tag{2.59}$$

Let $G_0^L = \sum_{k=1}^{L} \beta_k \delta_{\theta_k}$. Then, it can be shown that for every measurable function f integrable with respect to the measure H, this finite distribution G_0^L converges weakly to a countable infinite distribution G_0 distributed according to a Dirichlet process [19].

$$p(z_i = k \mid z_{\setminus i}, \gamma) = \frac{N_k^{-i}}{N - 1 + \gamma} \tag{2.60}$$

for each instantiated cluster k. The probability of generating a new cluster is given the remaining mass $\frac{\gamma}{N-1+\gamma}$.

Another method, motivated by the convergence guarantee of the equation above is to consider the degree L weak limit approximation to the Dirichlet process where L is a number that exceeds the total number of expected mixture components. Both of these approximations, encourage the learning of models with fewer than L components while allowing the generation of new components, upper bounded by L, as new data are observed [24].

As with the Dirichlet process, the HDP mixture model has an interpretation as the limit of a finite mixture model. Placing a finite Dirichlet prior on β induces a finite Dirichlet prior on π_j. As $L \to \infty$, this model converges in distribution to the HDP mixture model [24]. Inference algorithms for the weak limit approximation of the HDP allows for computational ease and efficiency. Given truncation level L, the weak limit approximation is as follows,

$$\beta \mid \gamma \sim Dir(\gamma/L, \ldots, \gamma/L) \tag{2.61}$$

$$\pi_j \mid \alpha, \kappa, \beta \sim Dir(\alpha\beta_1, \ldots, \alpha\beta_j + \kappa, \ldots, \alpha\beta_L) \tag{2.62}$$

Chapter 3

Non-Bayesian Parametric Model

A non-Bayesian parametric (NBP) HMM (model 1) which will be used to estimate the degradation rate of a GT rotor, is now presented. In order to model the evolution of rotor deterioration, a simple model of the the working environment that GenCo's gas turbines face on any given day must first be established. GenCo supplies bulk electricity to the wholesale electricity market, creating a stochastic environment with two main sources of randomness. The first random component is the portion of the demand of the counter-parties that GenCo must supply. The second component is the aggregate internal (known only to GenCo) variables–these include company set thresholds within which the machine must operate, as well as maintenance actions that the machine must undergo. Even though this last component is endogenous, it is nonetheless random, as subsystem failure may occur irregularly.

The impact of both of these forces on the rotor must be mathematically defined. First, notice that these forces affect the evolution of the subsystem at two different frequencies. The uncertainty in the demand manifests at a much lower frequency due to the wholesale market structure. This frequency is assumed to be twenty-four hours (day t) . In contrast, the uncertainty in the endogenous factors operates at a higher frequency. This frequency is set to five minutes (intra-day time τ, with 288 total time units intra-day).

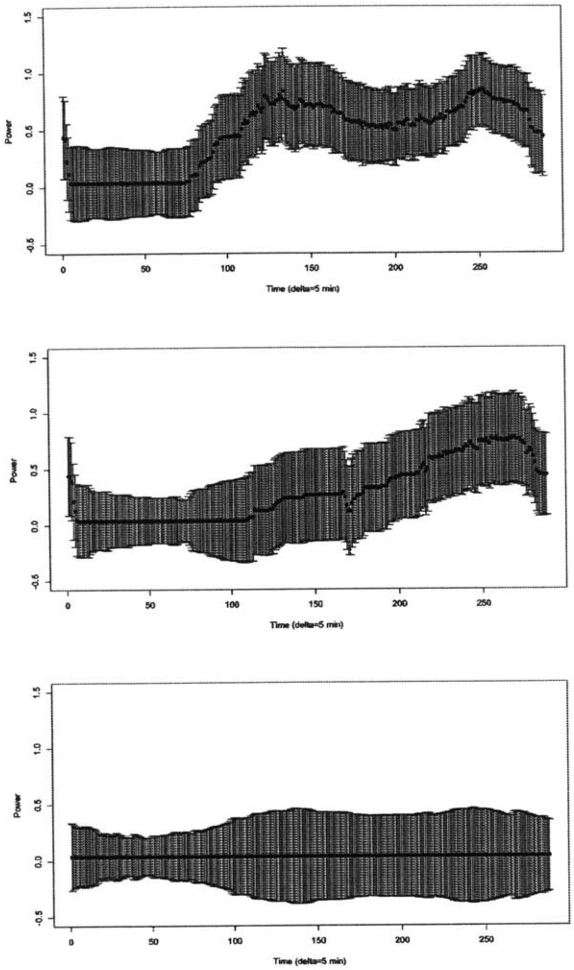

Figure 3-1: Three types of intra-day power-output signals

There are two assumptions, underlying the three models we present in this chapter and the next.

Assumption 1: The aggregation of random exogenous and endogenous factors affecting the machine is summarized by the power output of the machine. Exogenous factors include stochastic demand and endogenous factors include stochastic malfunctions that the machine observes. In other words, the power output of the machine on day t is an approximation for the operating conditions witnessed, and hence there is a correspondence between OMs and the intra-day power output signature.

According to exploratory data analysis, different groups of days observe similar intra-day power-output dynamics. This provides a reason for understanding the dependence of the degradation on day t on the power-output dynamics (corresponding to a specific OM). Figure 3-1 shows three types of power signatures, observed during year 1 of the dataset. The intra-day median values are plotted, with ± 1 standard deviation bands. We see in Panel 1, the power loading is stable for long duration of the day. This in stark contrast to Panel 3, which shows that the machine's power output is low. Thus, the set of days can be naturally bifurcated into different OMs, corresponding to their power output signatures proximity to the ideal operation concept. In this case, the three power signatures in Figure 3-1 can be interpreted as coming from days observing "normal" operation (Panel 1), "semi-normal" operation (Panel 2), and "abnormal" operation (Panel 3). The OMs are abbreviated as $\{N, S, A\}$, respectively.

Assumption 2: The deterioration level of the machine on day t is related the degree of abnormality the cooling air temperature dynamics are undergoing on that day. The level of abnormality is calculated with respect to a base level. The base level of the cooling air temperature dynamics is the kind of dynamics that the machine "should" be witnessing on day t. Since the base level may change with time, we subdivide the data into ordered groups (called "quarter-years"), so that the base level adjusts with time.

Imbedded in the model, are parameters that must be learned: the cardinality of the OM set, the transition matrix governing inter-OM switches, and the model parameters governing the intra-OM degradation evolution. Because of the structure of the parameters that must be estimated, it makes sense that we adopt the framework of an NBP-HMM with parameters $(\tilde{\pi}_0^{ij}, \tilde{\pi}^{ij}, F_{ij})$, for each quarter-year pair ij [13]. F denotes a multivariate Gaussian emission probability density structure. The problem of estimating the cardinality of the OM set is also mitigated in this chapter by positing that three OMs exist $\{N, S, A\}$, based on exploratory data analysis. Using nonparametric Bayesian techniques, explored in the next chapter, this assumption will be removed, and the number of OMs will automatically be estimated, using the data.

3.1 Power Output Detrending

To estimate how the deterioration of the rotor evolves while dwelling in a particular OM, we must first classify day t's residence in a particular OM z_t. As stated in the previous section, the OM of a particular day is signified by its power-output signature u_t. We would like to use the power-output signature of a particular day as the emission of the HMM, in order to learn the OM for that day, using the Expectation-Minimization algorithm for inference. Yet, the large dimensionality of u_t, and the relatively small number data points in the first five years (1420 days) poses a problem. Hence, we reduce the dimensionality of u_t, by projecting it to a four-dimensional vector v_t, by de-trending using the following linear regression [14]:

$$\begin{aligned} u_{t\tau} = v_{t1} + v_{t2}\tau + v_{t3}\tau^2 + v_{t4}|\tau - \frac{288}{2}| + e_{t\tau} \\ e_{t\tau} \sim N(0, \sigma_t^2). \end{aligned} \tag{3.1}$$

A separate regression model of the form Equation 3.1 is run for each day, where $\tau \in \{1, \ldots, 288\}$ is the intra-day time index. For day t, there are 288 observations (assumed to be independently Gaussian distributed with variance σ_t^2). Each observation is made up of tuple: feature vector $(\tau, \tau^2, |\tau - \frac{288}{2}|)$ and independent variable (power signature for time τ) $u_{t\tau}$.

Since day $t's$ power signal is regressed on different degree polynomials of time, along with shifts, and an intercept term, the signature imbedded in the intra-day power signal is captured in the weights v_t. After adding a sufficient number of polynomials, and shifted polynomials to the regression, the intra-day power signal will be detrended leaving only white noise, i.e. $e_{t\tau} \sim N(0, \sigma_t^2)$ [14]. The regressors in Equation 3.1 extract different trends imbedded in the signal. τ extracts the rate at which the signal is changing linearly with time. τ^2: extracts signal accelerations as time increases. And, $|\tau - \frac{288}{2}|$: extracts if the signal exhibits different dynamics during the middle of day verse the day's start and end.

Equation 3.1 is actually a specified version of a much more general detrending regression model. Let a shifted polynomial basis be defined as $(\tau - k)^g$ where τ denotes time, k denotes a shift, and g denotes the degree. And, let a shifted agnostic basis be defined as $|\tau - h|$, where h denotes the shift. The polynomial basis extract trends that emanate from the end points of the time horizon from a given signal. The agnostic basis extracts trends that emanate from the center of the time horizon. The shifts for both cases, allow for trend extraction that begin away from the end points, or away from the center. Using these two bases, a signal can be efficiently detrended using the following procedure.

Power-Output Detrending Procedure

1. Given data u_t for day t
2. Define the polynomial bases as $(\tau - k)^g$ for $k = 1, ..., K;$, $g = 1, ..., G$
3. Define the shift bases as $|\tau - h|$ for $k = 1, ..., H$
4. The mapping is constructed by imposing the following linear regression:

$$u_{t\tau} \sim \sum_{k=0}^{K} \sum_{g=0}^{G} v_{gk}(\tau - k)^g + \sum_{h=0}^{H} v_h |\tau - h| + e_{t\tau} \tag{3.2}$$

$$e_{t\tau} \sim N(0, \sigma_t^2) \tag{3.3}$$

5. Extract compressed signature v_t

3.2 Degradation Model

Having fit the NBP-HMM for each quarter-year, each day is labeled by an OM, thus bifurcating quarter-year days into separate OM "types". In order to isolate amount of degradation that occurs on day t, how the cooling temperature typically behaves on day t's OM type, must be modeled. To accomplish this, for each group (corresponding to OM) of days, a separate linear regression model is fit, with the cooling air temperature being the independent variable. Each model is meant to explain away portions of the variance of the temperature that are correlated with two feature variables: the fluctuations in the power signal and second, those trends that are characterize residence in a particular OM (which are modeled by polynomials of time).

A set of residuals is obtained for each day, from the appropriate regression model. The residual is the amount by which the temperature is deviating or fluctuating away from its expected/typical behavior, given that it belongs to a particular OM. Hence, the residual is a metric the level of abnormal or atypical behavior of the temperature, making it a proxy for the degradation. The terms "idiosyncratic component" and "residual" will be used interchangeably. For the ij quarter-year, we posit the following regression model:

$$y_{t\tau} = \sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t=z\}} m_{t\tau}^{z} + \beta_2 u_{t\tau} + e_{t\tau} \tag{3.4}$$

where

$$m_{t\tau}^{q} = \beta_{11}^{z}\tau + \beta_{12}^{z}\tau^2 + \beta_{13}^{z}|\tau - \frac{288}{2}| \tag{3.5}$$

and

$$\mathbb{1}_{\{z_t=z\}} = \begin{cases} 1 & \text{if } z_t = z \\ 0 & \text{if } z_t \neq z \end{cases} \tag{3.6}$$

The right side of Equation 3.4 has three terms. The first term is the "OM factor". It is made up of two parts: an indicator function defined in Equation 3.6 and the OM factor $m_{t\tau}^{z}$, defined in Equation 3.5. The indicator function groups all of the days of the same OM in quarter-year ij. The OM factor is described by a set of time-dependent polynomials, and is intended to extract the variability of temperature due to dwelling a particular OM. The indicator variable activates the correct $z \in \{1, ..., |\mathcal{Q}_{ij}|\}$ given

Viterbi's classification of day t. The second term is the power loading factor u_{st}. We would like to isolate variance in the temperature that is not explained by the elevated or deflated levels of the power loading. The last term is the idiosyncratic term: $e_{t\tau}$. The purpose of using this regression model is to procure statistics of the distribution of this term, controlling for the first two factors. Sample estimates of the first two statistics of this distribution will provide insights into the deterioration levels of the rotor. Standard Ordinary Least Squares procedure are used for parameter estimation [16]. The exhibit below outlines the full model presented in this chapter.

Parametric Degradation Model

θ_k: Parameter governing distribution of observations emitted from kth component, $\mathcal{K} = \{1, 2, \ldots, K\})$
$\tilde{\pi}_k$: State-specific transition distribution for state j
$(1, \ldots, T)$: Sequence of temperature measurements $\{y_1, \ldots y_T\} \in \mathbb{R}^{288}$
$(1, \ldots, T)$: Sequence of power output measurements $\{u_1, \ldots u_T\} \in \mathbb{R}^{288}$
$(1, \ldots, T)$: Sequence of features extracted from $\{u_t\}_{t=1}^{T}$, $\{b_1, \ldots b_T\} \in \mathbb{R}^4$
z_t: Indicator random variable signifying component of u_t, $\{z_1, \ldots, z_T\} \in \mathcal{K}$

$$z_t \mid \{\pi_j\}_{k=1}^{K}, z_{t-1} \sim \pi_{z_{t-1}} \quad t = 1, \ldots, T \tag{3.7}$$

$$b_t \mid \{\theta_k\}_{k=1}^{K}, z_t \sim \mathcal{N}(\mu_{z_t}, \Sigma^{z_t}) \quad t = 1, \ldots, T \tag{3.8}$$

$$y_{t\tau} \mid u_{t\tau}; t \in \mathcal{Q}_{ij} \sim \mathcal{N}\left(\sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t = z\}} m_{t\tau}^{z} + \beta_2 u_{t\tau}, \sigma_{ij}^2 \right) \tag{3.9}$$

Degradation Level

$d_{t\tau}$: Deterioration level at time τ on day t: $\left(y_{t\tau} - \sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t = z\}} m_{t\tau}^{z} + \beta_2 u_{t\tau} \right)$
D_t: Total deterioration level at time on day t

$$d_{t\tau} \mid t \in \mathcal{Q}_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2) \tag{3.10}$$

$$D_t = \sum_{\tau=1}^{288} d_{t\tau} \tag{3.11}$$

3.2.1 Forced Jumps

In this section, we briefly explain how the NBP-HMM model can be extended to not only include stochastic transitions between different OMs, but also forced transitions that occur due to a temperature set point variable $\psi_{\tau t}$ hitting a guard condition.

Denoting k, k' as two different OMs, we can also model models forced jumps between k and k' by $\rho_{k,k}$. These forced transitions occur when the temperature set point signals $\psi_{\tau t}$ deviate from the operating concept. In general, the levels at which the forced jump activates, known as "guard-lines" can be defined with respect to a function $g(\psi_{\tau t})$. If the guard-line is hit by $g(\psi_{\tau t})$, the OM will be forced to transition, affecting the dynamics of the temperature signal $y_{\tau t}$. From a practical point of view, these forced jumps can be used to model trips.

If we define Γ to be the threshold that governs the forced jump and $\epsilon > 0$ to be the required "distance" required to activate a jump, then we can define a new transition kernel in the following way:

$$\mathbb{P}(k|k') = \begin{cases} \rho_{k,k'}, & \text{if } \Gamma - g(\psi_{\tau t}) \leq \epsilon \\ \pi^{ij}_{k,k'}, & \text{if } \Gamma - g(\psi_{\tau t}) > \epsilon \text{ and } t = T \end{cases} \tag{3.12}$$

$$\rho_{k,k'} = 1 \text{ iff } k = k^* \tag{3.13}$$

As mentioned above, the guard-lines can be defined for a general function of the temperature set point $\psi_{\tau t}$. Using exploratory data analysis, one can derive different statistics such as such as mean, variance, and slope of $\psi_{\tau t}$ before a trip occurs to see which statistic is most correlated with trip occurrence. In our analysis of 10 trip events, we found that the distributions of the mean of $\psi_{\tau t}$ before trips, verses the mean of $\psi_{\tau t}$ under normal conditions were significantly different, allowing use the mean of $\psi_{\tau t}$ as $g(\cdot)$.

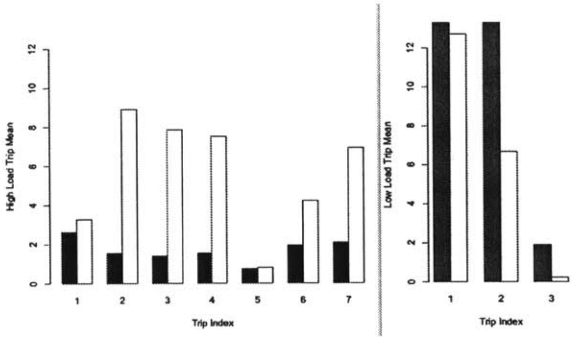

Figure 3-2: Example of $g(\cdot)$. Left set corresponds to High Load Trips. Right set to Low Load Trips. $g(\cdot)$ is the mean of the TAT signal, 2 hours before the trip. Red/Blue=Trip, White=Normal.

In Figure 3-2, we show the different mean values for each trip condition, and corresponding normal condition, for both trips at high loads and low loads. For each trip event, we calculate the mean value using the data of $\psi_{\tau t}$ up to 2 hours before the trip. We group together the 10 most recent normal days prior to the trip day, and calculate the mean value of the averaged $\psi_{\tau t}$ signal across these 10 days. In Figure 3-2, the high load trip events are ordered by time. One can see that over time for high load trips, the means of both normal and trip conditions decrease. Out of these 10 trips, all are trips at high loads, except for trips number 3, 4, 10, which are at low loads, leading to their bar plots being flipped with respect to the others. This information can be used to declare a Γ that minimizes the error of modeling a trip occurrence incorrectly. In this case, Γ may be specified as a decreasing function of time.

In this framework, the system may transition between different OMs both autonomously based on some probability distribution, at the daily frequency, as well as non-autonomously due to forced jumps at intra-day frequency.

Chapter 4

Two Bayesian Nonparametric Models

4.1 HDP-HSMM

Johnson and Willsky introduced a Bayesian nonparametric (HDP) version of the HSMM, which we will refer to as model 2 [20]. It extends the HDP-HMM in cases where strict Markovian assumptions need to be relaxed. The HDP-HSMM is able to learn models in which the dwell times within different modes are not assumed to be geometric. Additionally, the nonparametric nature of the model allows for the number of hidden OMs to be learned. Furthermore, because the Markovian assumption is relaxed, rapidly switching dynamics are also mitigated [24].

For these reasons, a degradation model based on the HDP-HSMM is a natural enhancement of the NBP-HMM described in the previous chapter. Although, the mathematical model is now Bayesian, with an additional semi-Markov dwell-time assumption, the model remains basically the same as the parametric HMM model. Like the NBP-HMM, the HDP-HSMM is used to bifurcate the set of days into different OMs, based on the power output signal.

Let there be a particular interval of days indexed by s: $\mathcal{D}^*{}_s = (t_s^1, \dots, t_s^2)$. Let $z_s \in \mathcal{K}$ denote the particular OM the system dwells in during the interval $\mathcal{D}_s^*$. For the purposes of the degradation model, $\mathcal{D}_s = |\mathcal{D}_s^*|$ is assumed to be distributed $Exponential(\lambda_{z_s})$.

$$\mathcal{D}_s \mid z_s = k \sim Exponential(\lambda_k). \tag{4.1}$$

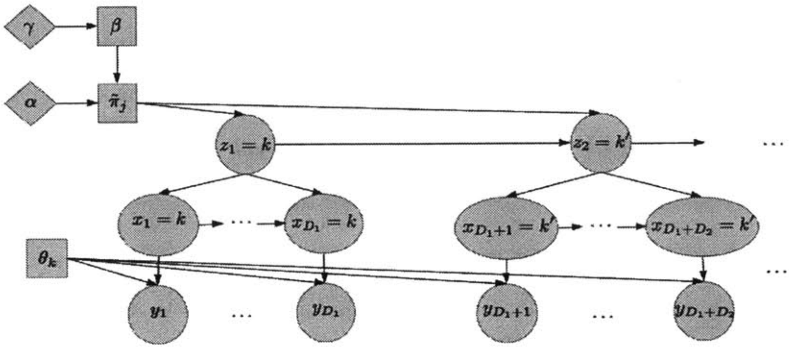

Figure 4-1: HDP-HSMM

Let $t \in \mathcal{D}_s^*$, $u_t \in \mathbb{R}^{\mathcal{T}}$ be the power output signal, $x_t \in \mathcal{K}$ denote the OM, and z_s denote the super-state (*i.e.*, $x_t = z_s$). The powers signal, u_t is assumed to be governed by a multivariate Gaussian distribution with OM-dependent parameters $(\mu_{x_t}, \Sigma_{x_t})$.

$$y_t \mid x_t, z_s = k \sim \mathcal{N}(\mu_k, \Sigma_k) \tag{4.2}$$

In other words, the GT is assumed to be in different OMs during different intervals of days. The length of time it remains in an OM, depends on a OM-specific parameter λ_k. Upon expiration, the system switches to another OM according to a Markov chain. Finally, it is assumed that each day the system resides in a particular OM, the power output is generated by a multivariate Gaussian distribution with OM-dependent parameters. In order to calculate the degradation of the rotor on a particular day, the idiosyncratic component of the temperature for each day is calculated just like in the NBP-HMM case. For the ij quarter-year, we once again posit the following linear regression model:

$$y_{t\tau} = \sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t=z\}} m_{t\tau}^z + \beta_2 u_{t\tau} + e_{t\tau} \tag{4.3}$$

HDP-HSMM Degradation Model

β : Dirichlet process with parameters $\gamma, H(\lambda)$
θ_k: Parameter governing distribution of observations emitted from jth component, (The group and component index sets are now equal: $\mathcal{K} = \{1, 2, \dots\}$)
$\tilde{\pi}_k$: Dirichlet process with parameters α, β (state-specific transition distribution for state k)
$(1, \dots, T)$: Sequence of labels $\{x_1, \dots, x_T\} \in \mathcal{K}$
$(1, \dots, S)$: Sequence of "super-states" $\{z_1, \dots z_S\} \in \mathcal{K}$
$(1, \dots, S)$: Sequence of durations $\{\mathcal{D}_1, \dots \mathcal{D}_S\} \in \mathbb{R}$
$(1, \dots, T)$: Sequence of temperature measurements $\{y_1, \dots y_T\} \in \mathbb{R}^{288}$
$(1, \dots, T)$: Sequence of power output measurements $\{u_1, \dots u_T\} \in \mathbb{R}^{10}$
z_t: Indicator random variable signifying component of u_t, $\{z_1, \dots, z_T\} \in \mathcal{K}$

$$z_t \mid \{\pi_j\}_{j=1}^{\infty}, z_{t-1} \sim \pi_{z_{t-1}} \quad t = 1, \dots, T \tag{4.4}$$

$$u_t \mid \{\theta_k\}_{k=1}^{\infty}, z_t \sim \mathcal{N}(\mu_{z_t}, \Sigma^{z_t}) \quad t = 1, \dots, T \tag{4.5}$$

$$y_{t\tau} \mid u_{t\tau}; t \in \mathcal{Q}_{ij} \sim \mathcal{N}\left(\sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t = z\}} m_{t\tau}^{z} + \beta_2 u_{t\tau}, \sigma_{ij}^2 \right) \tag{4.6}$$

OM Dwell Time

$$\mathcal{D}_s \mid z_s = k \sim Poisson(\lambda_k) \tag{4.7}$$

Degradation Level

$d_{t\tau}$: Deterioration level at time τ on day t: $\left(y_{t\tau} - \sum_{z=1}^{|\mathcal{Q}_{ij}|} \mathbb{1}_{\{z_t = z\}} m_{t\tau}^{z} + \beta_2 u_{t\tau} \right)$
D_t: Total deterioration level at time on day t

$$d_{t\tau} \mid t \in \mathcal{Q}_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2) \tag{4.8}$$

$$D_t = \sum_{\tau=1}^{288} d_{t\tau} \tag{4.9}$$

4.2 HDP-SLDS-HMM

Both the NBP-HMM (model 1) and HDP-HSMM (model 2) can be computationally viewed as a three-step algorithm. The first step classifies each day as a particular OM. The second step groups together days by OM-type and fits a regression model for the cooling air temperature for each group. The final step calculates the daily residuals as an estimate for the degradation level.

The HDP-HMM with switching linear dynamical system (SLDS) emissions (model 3) allows for the first two steps to be combined. Since the OM is never directly measured, it is assumed to be latent, as in the previous cases. But, now, it will be inferred using observations from two correlated variables, instead of one. The OM for day t is denoted z_t, which takes values in discrete set $\mathcal{K}$, whose empirical cardinality will be learned using the estimation procedure.

Assumption 1 (HDP-SLDS-HMM): The aggregation of random exogenous and endogenous factors affecting the machine is summarized by the relationship between the cooling temperature and power output of the machine. How sensitive the cooling air temperature dynamics are to power output dynamics is indicative of the OM. The level of sensitivity will be numerically encapsulated by matrix A.

The HDP-SLDS-HMM posits a two-level SLDS emission model. The first level posits that the power output for day t, denoted u_t, conditional on $z_t = k$ is related to u_{t-1} by a matrix $A^{(k)}$ with OM -dependent additive Gaussian noise $e_t(k)$. The second layer relates the temperature signal y_t to u_t by fixed matrix C and additive Gaussian noise w_t.

$$z_t \sim \pi_{z_{t-1}} \tag{4.10}$$

$$u_t = A^{(z_t)} u_{t-1} + e_t(z_t) \tag{4.11}$$

$$y_t = C u_t + w_t \tag{4.12}$$

HDP-SLDS-HMM Degradation Model

β : Dirichlet process with parameters $\gamma, H(\lambda)$
$\theta_k = (A^{(k)}, \Sigma^{(k)})$: Parameter governing distribution of observations emitted from jth component, (The group and component index sets are now equal: $\mathcal{J} = \mathcal{K} = \{1, 2, \dots\}$)
$\tilde{\pi}_j$: Dirichlet process with parameters α, β (state-specific transition distribution for state j)
$\{z_t\}_{t=1}^{T}$: Sequence of latent discrete variables, $z_t \in \mathcal{K}$

$$\beta \mid \gamma \sim GEM(\gamma) \tag{4.13}$$

$$\theta_j \mid H, \lambda \sim H(\lambda) \quad j = 1, 2, \dots \tag{4.14}$$

$$\tilde{\pi}_j \mid \beta, \alpha \sim DP(\alpha, \beta) \tag{4.15}$$

$$z_t \mid \{\pi_j\}_{j=1}^{\infty}, z_{t-1} \sim \pi_{z_{t-1}} \quad t = 1, \dots, T \tag{4.16}$$

$\Sigma^{(k)}$: Noise covariance matrix for OM k
$A^{(k)}$: Sensitivity matrix for OM k
$\{y_t\}_{t=1}^{T}$: Sequence of temperature time series vectors, $y_t \in \mathbb{R}^{\mathcal{T}}$
$\{u_t\}_{t=1}^{T}$: Sequence of power output time series vectors, $u_t \in \mathbb{R}^{\mathcal{T}}$

$$\Sigma_k \sim IW(n_0, S_0) \tag{4.17}$$

$$A^{(k)} \sim MNIW(\bar{M}, \bar{K}, \Sigma_k) \tag{4.18}$$

$$y_t \mid \{\theta_k\}_{k=1}^{\infty}, x_t, z_t = k \sim \mathcal{N}(A^{(k)} x_t, \Sigma^{(k)}) \quad t = 1, \dots, T \tag{4.19}$$

Degradation Level

$d_{t\tau}$: Deterioration level at time τ on day t: $y_{t\tau} - \hat{y}_{t\tau}$
D_t: Total deterioration level at time on day t

$$d_{t\tau} \mid t \in \mathcal{Q}_{ij} \sim \mathcal{N}(0, \sigma_{ij}^2) \tag{4.20}$$

$$D_t = \sum_{\tau=1}^{288} d_{t\tau} \tag{4.21}$$

Chapter 5

Reliability Prognostics

In this chapter, results from the NBP-HMM (model 1), HDP-HSMM (model 2), and HDP-SLDS-HMM (model 3) are presented. Out of the six years of data, the first five years are used in model development, and the last years data is used for prediction. The first section outlines the metrics that will be used to validate and compare all three models in section 2. The third and fourth sections present the main deterioration results, and analysis of model 3. The final section introduces prediction models that are used to predict degradation rates during the sixth year.

5.1 Deterioration Metrics

In order to understand the level of degradation, estimated by a particular model, the mean and the standard deviation of the absolute value of the model residuals is isolated. For model 1, the power signature is first detrended using the procedure introduced in chapter 3. Using this detrended signal sequence, the OM sequence is estimated for each day in the first five years ($\sim$ 1420 days). For each OM, a separate linear regression model is fit for the cooling air temperature. Then, by subtracting the fitted values from the observed values, we recover a vector of residuals. For model 2, the same procedure is done, except no detrending takes place. Instead, the entire intra-day power signal is taken as an emission for the nonparametric HSMM. For model 3, the emission for each each day is taken to be a tuple made up of the power signature and cooling air temperature signature. The OM sequence is learned for each day, along with the fitted cooling air temperature values. Thus, no separate regression model is needed to produce residuals for model 3.

For day s, we define the level of daily deterioration D_s, volatility of daily deterioration: V_s, and the cumulative deterioration C_s below. As a reminder, T is the number of time steps in a given day: 288.

$$D_s = \frac{1}{T}\sum_{t=1}^{T} |e_{st}| \tag{5.1}$$

$$V_s = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (|e_{st}| - D_s)^2} \tag{5.2}$$

$$C_s = \sum_{k \leq s} D_k \tag{5.3}$$

The definition presented in Equation 5.1 implies that the daily deterioration is the average level of abnormal temperature fluctuation witnessed during a given day. The volatility level V_s is the degree of fluctuation of the temperature, which is indicative of deterioration. These metrics capture the effect of two main causes of rotor degradation: creep and LCF. By obtaining the absolute value of the residuals, the deterioration level accounts for both positive and negative deviations from the expected temperature range.

5.2 Model Validation

For model 1, which is parametric, three latent states are pre-posited corresponding to normal, semi-normal, and abnormal OMs. In contrast, model 2 (with a five-state approximation) learned three different states, and model 3 (with a ten-state approximation) learned six states. The first three states of model 3, and the three states of model 2 correspond to the normal, semi-normal, and abnormal OMs of model 1. In Figure 5-1, columns 1,2,3 correspond to models 1,2,3 respectively. The first row plots the intra-day mean power signature for the normal OM, the second for the semi-normal OM, and the third for the abnormal OM. We notice that for all three models, these three OMs share very similar characteristics.

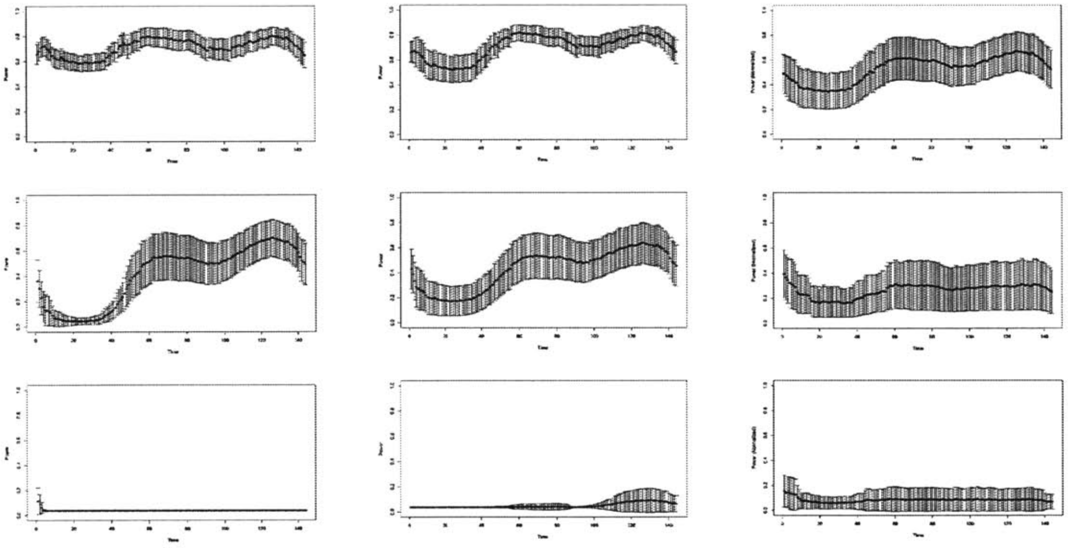

Figure 5-1: Mean power signature, with .5 standard deviation bands

All three models produce accurate predictions in light of non-stationary behavior of GT operating conditions. The non-stationary arises due to fluctuations due to seasonal changes, supply-demand shocks, internal/sub-component states, and maintenance actions – all of which vary across all five years of data. Figure 5-2 shows the fitted values for cooling air temperature from all three models. Notice, the models, along with the actual temperature dynamics change drastically not only across year, but within each year. Specifically, between years 2 and 3 the cooling air temperature observed unstable dynamics, yet was relatively stable dynamics during year 4. All three models produced results which fitted with the true temperature readings very closely. Since the models produce good fits to the data under varying regimes, we can conclude that over fitting is not taking place. In fact, the dependence of the OM in the model, allows for a robust fit.

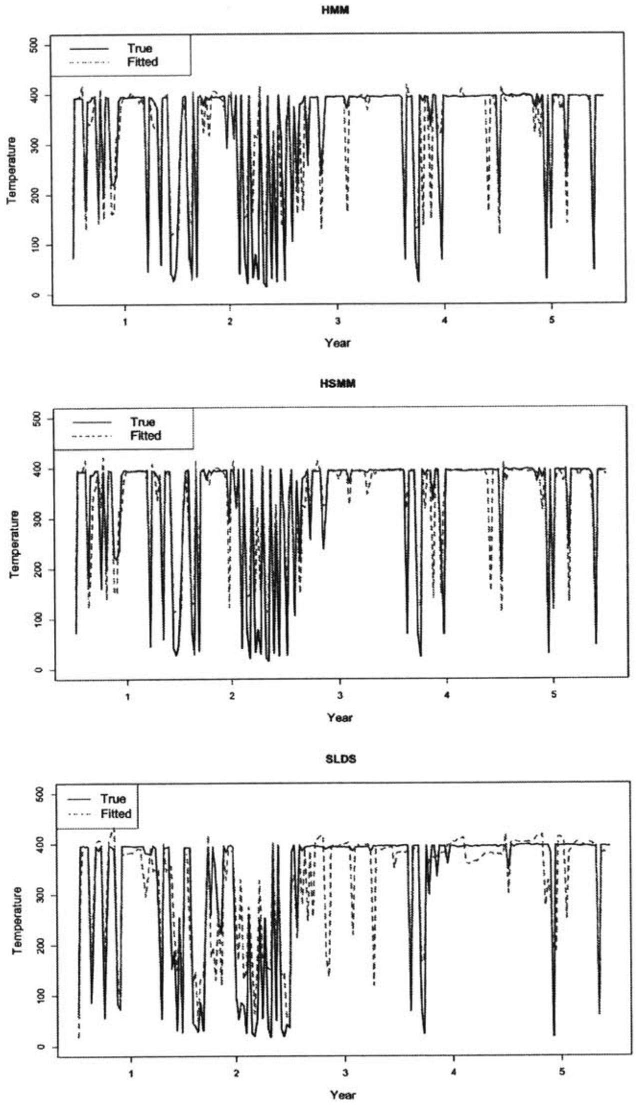

Figure 5-2: Model 1,2,3 fitted models for cooling air temperature across first 5 years

Figure 5-3 depicts the percentage of days the GT dwells in a particular OM for each year for model 1(T), model 2 (M), and model 3 (B). For models 1 and 2, the results are very similar. For model 3, a high percentage of the days were classified as normal. For all three models, we see that across the time horizon, the system transitions out of dwelling in abnormal OMs to normal OMs. Specifically, the GT witnessed many days in abnormal operation during the first two years. The last three years were relatively normal, in comparison. These results agree with Figure 5-2. Particularly, years one and two show an unstable cooling air temperature signal, while year four shows a stable temperature signal.

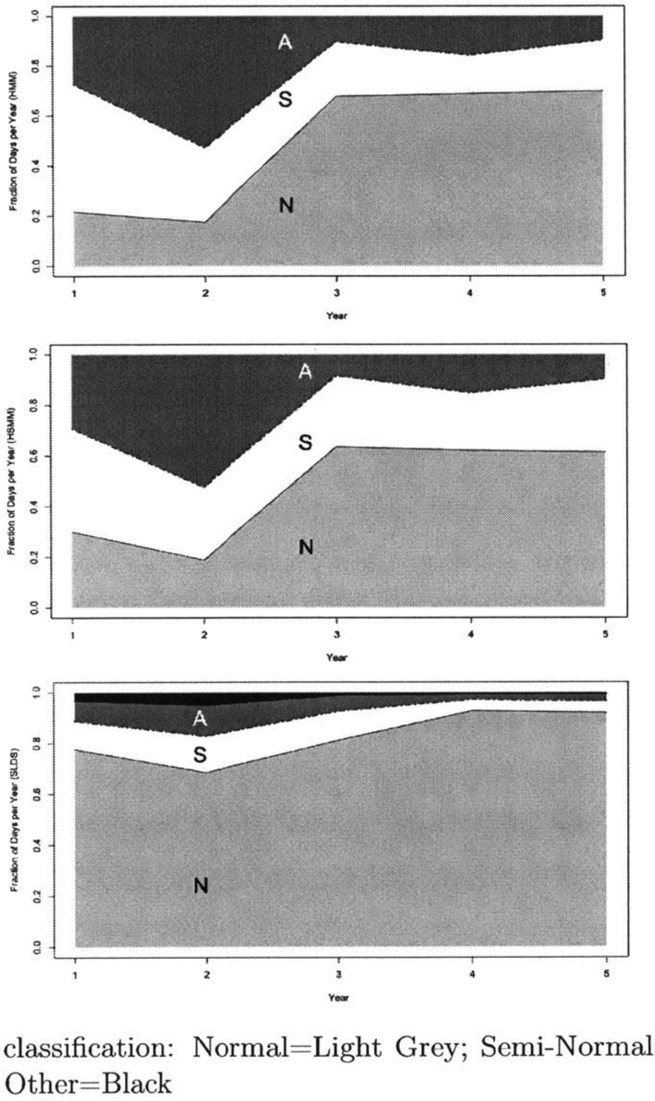

Figure 5-3: OM classification: Normal=Light Grey; Semi-Normal=White; Abnormal=Dark Grey; Other=Black

5.3 Deterioration Results

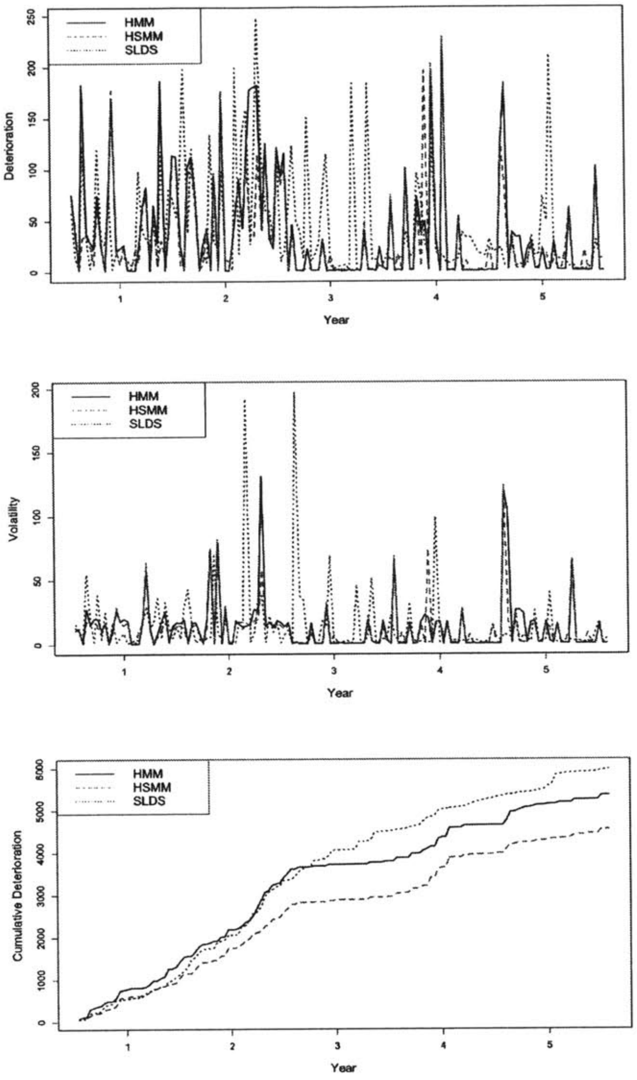

Figure 5-4: Evolution of Rotor Temperature, D_s, V_s, Cumulative Sum of D_s across all 5 years (all graphs are subsampled every 10 days)

In Figure 5-4, we show three subfigures. Each subfigure has 3 lines, corresponding the models 1,2,3. The first subfigure graphs deterioration, D_s as a function of time, the second graphs volatility V_s, and the final subfigure graphs the cumulative deterioration C_s. The first result is that both volatility, V_s and deterioration, D_s are highly correlated, implying that both volatility and magnitude of deterioration evolve together in time. We see that for all three models, D_s values spike during years one, two, and four, whereas years three and five have relatively low levels of deterioration. For V_s, model 1 has overall lower levels than model 2. In particular, for model 1, V_s spikes during year two and four, while for model 2, we observe peaks during years one, two, and four.

The second result is that high dwell times in abnormal and semi-normal modes are correlated with high deterioration levels. From Figure 5-3, it is seen that in year two, there is an increase of the number of days in which the GT dwells in the abnormal OM. Additionally, in the panel of Figure 5-2 corresponding to year four, notice that during quarter two, the rotor temperature exhibits high variance. Hence, the dwell time in the abnormal OM is positively correlated with high deterioration. In addition, there is a correspondence between all three models with respect to the dwell times in semi-normal OM (Figure 5-3) with elevated deterioration levels in years one and five. Note that the observed deterioration levels, when the GT dwells in this OM are not as high as when dwelling in the abnormal OM.

The third result is about the rate at which the deterioration evolves across the time horizon. In the final panel of Figure 5-4 we show how the cumulative deterioration C_s (ref. Equation 5.1) evolves over time for all three models. The important takeaways from these graphs are the different rates at which the deterioration accumulates, during each of the five years. We note that the maximum deterioration accumulation for all models occurs in year two. During years one and two, the slopes are relatively higher than the latter three years indicating higher rates of deterioration. Although models 1 and 2 observe similar deterioration and volatility graphs, across the five years, model 3, observes peaks of deterioration and volatility at differing times. The reason is due to the three new OMs that model 3 detects. In fact, deterioration levels detected by model 3 are due in part to the the existence of the new OMs, as we discuss further in the next section.

A Poisson process is assumed to govern the dwell times for the OMs in model 2 (HDP-HSMM). The exponential distribution, with parameter λ is the probability distribution that describes the time between events in a Poisson process. In Table 5.1, we list the estimated λ for each of the OMs, for each of four data groups. The values estimated correspond to the percentage dwell times depicted in Figure 5-3. In particular, for data group four, corresponding approximately to year five , for the abnormal OM, $\lambda = 70.6715$, meaning that in year five, the amount of time spent upon entering the abnormal OM is extremely short ($\sim 1/70.6715$ days).

Data Group	OM	λ
1	N	7.609
	A	8.4065
2	N	3.8512
	S	0.833
	A	70.6715
3	N	5.2688
	S	2.8638
4	N	17.9668
	A	2.963

Table 5.1: λ for each OM for each data group (model 2

5.4 HDP-SLDS-HMM Results

A weak-limit approximation to model 3 (HDP-SLDS-HMM), with a ten-OM truncation is also fit to the five year data set. Although ten OMs were learned, approximately six OMs had significantly different characteristics. Also, in contrast to the models 1 and 2, the emissions of the HDP-SLDS-HMM is a switching linear dynamical system (SLDS), instead of the power output. Each of the six OMs represent a distinct relationship (by matrix A) between the power output and the cooling air temperature. Figure 5-5 shows six rows of figures, corresponding to each of the OMs learned. The left figure of each row depicts the mean power output signatures with .5 standard deviation bands. The right figure depicts the mean cooling air temperature signatures, also with .5 standard deviation bands. The mean power signatures of the first three OMs, corresponding to Figure 5-5 (panel (L) of rows 1-3) map directly to the "normal", "semi-normal", and "abnormal" days, which were learned by models 1 and 2, as well. The cooling air temperature signatures (panel (R) of rows 1-3) are also distinct in each of these OMs. During normal days, the temperature is stable and high. On semi-normal days, the temperature signature is stable, and decreases through the day. On abnormal days, the temperature signature is stable, yet low, throughout the day.

In addition to these three OMS, the HDP-SLDS-HMM detects three new OMs, corresponding to Figure 5-5 (rows 4-6). Although, all three of these mean power signatures reach the ideal power output level (.75), at some point during the day, each of these three OMs are very volatile compared to the normal, semi-normal, and abnormal OMs. Furthermore, the temperature signatures corresponding to these three OMs follow a very similar "ramp-up" trajectory. The temperature signature of the fourth OM (Figure 5-5: panel 2, row 4) takes almost three-fourths of the day to ramp up after reaching ideal levels. The temperature signature of the fifth OM (Figure 5-5: panel 2, row 5) takes almost one-fourth of the day to ramp up after reaching ideal levels, but then tapers off towards day-end. Finally, the temperature signature of the sixth OM (Figure 5-5: panel 2, row 6) takes almost one-fourth of the day to ramp up after reaching ideal levels.

Although the HDP-SLDS-HMM detects a larger variety of OMs, the percentages of dwell time in the normal, semi-normal, and abnormal OMs are still the highest with 79.6%, 9.2%, 6% respectively. Dwell times in the three new OMs, combined account for less than 5% of the days. In terms of the degradation, the residuals calculated on days residing in OM 4, are very high compared to the average value of residuals with this model. The mean residual value for OM 4 is 293.7, whereas the global mean value of residuals is 44.1. For days in OM 4, we see that the cooling air temperature remains in non-ideal level for most of the day. Additionally, the step-like signature in combination with the variance of the signal lead to high residual levels.

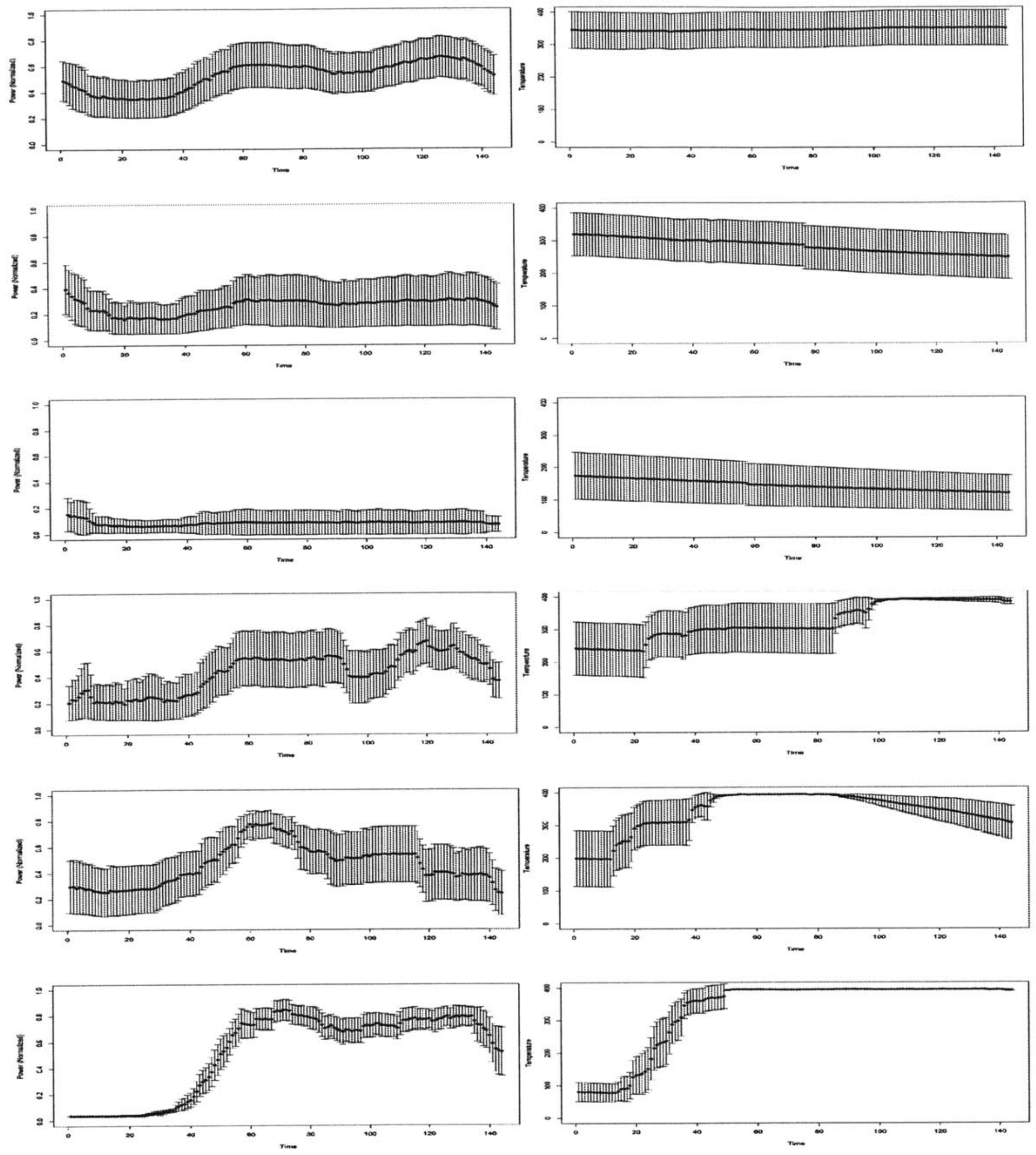

Figure 5-5: HDP-SLDS-HMM OM mean power signatures (left column) with corresponding mean cooling temperature (right column)

5.5 Prognostics

The validity and prediction power of the NBP-HMM (model 1) is now developed using the first five years of data using an out-of-sample data set: the sixth year of data. The results are presented using 70 day "quarters" of the sixth year. Unlike the first five years, the sixth year witnessed a disproportionately high number of normal OM days, as shown in Figure 5-6 (panel 4), with the exception of the first quarter. In line with this observation, the daily deterioration level D_s was generally low as can be seen in Figure 5-6 (panel 2), compared to the levels of deterioration witnessed during years 1-5. Spikes of degradation levels are observed, which manifest at those times when the machine switched from normal to abnormal OMs.

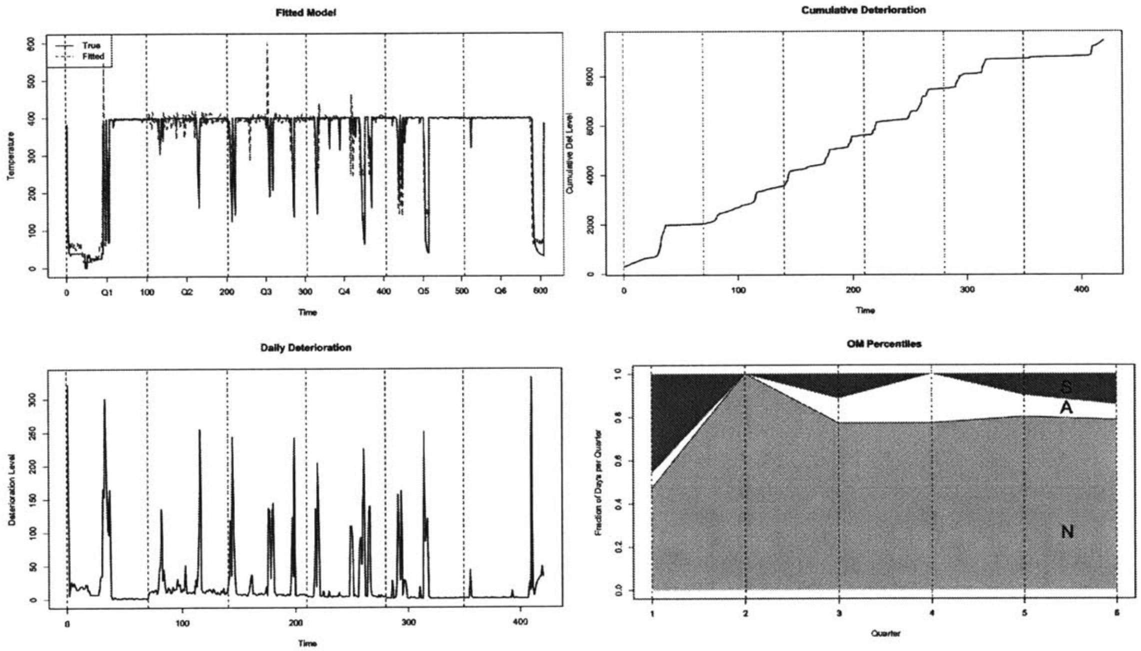

Figure 5-6: Fitted NBP-HMM regression model (panel 1); D_s (panel 2); Cumulative Sum of D_s (panel 3); Viterbi OM classification dwell percentages for the sixth year (panel 4) (subsampled every 10 days)

Two methods to perform online prediction of the rotor temperature were tested. The first method used a model bank $\mathcal{P}$, each learned using in-sample data from quarter-year

ij. To simplify notation, we index each quarter-year's model p by j (i.e., $j = 1, ..., 20$). We subdivide the sixth year of data into overlapping 35 70-day windows, indexed by ω (i.e., $\omega = 1, ..., 35$). The overlap is ten days. For each window, we fit the NBP-HMM presented above. Using the fitted model, we procure the OM labels for last twenty days of the window, and calculate a three dimensional vector p_ω, where the three dimensions correspond to the proportion of normal, semi-normal, abnormal OM days witnessed during window ω's last twenty days. The dimensions of this vector are indexed by i (i.e., $i = 1, ..., 3$). Then, the optimal model for window ω, $p^*(\omega)$ is procured by computing the following min-max optimization.

$$p^*(\omega) = arg \min_{\mathcal{P}} \left(\max_i |p^i_\omega - p^i_j| \right) \tag{5.4}$$

This is followed by forward simulating $p^*(\omega)$ ten days, setting ω's last days OM classification as the initial condition. The predictions, for each of these ten day "chunks" are shown in Figure 5-7 (panel 1). The predictive power of the method is calculated by computing the mean absolute error (MAE) for each 10 day chunk (Figure 5-7 panel 3). As can be seen, the predictive power diminishes towards the end of the sixth year. This is expected since the model bank uses models only from the first five years, and does not update within the sixth year. Yet, for the first fifteen quarters, the errors are relatively low and stable.

The first method presented performs poorly towards the second-half of year six, because none of the models in $\mathcal{P}$ provide a good fit to the dynamics observed in the windows, which arise. To account for this weakness, a second method is formulated which uses the most recent fitted model to forward simulate. The results are presented in Figure 5-7 (panel 2). A drastic improvement in the models stability is seen. Across most of the windows for the year, especially, towards the second half of year six, the MAE is lower than the first methods.

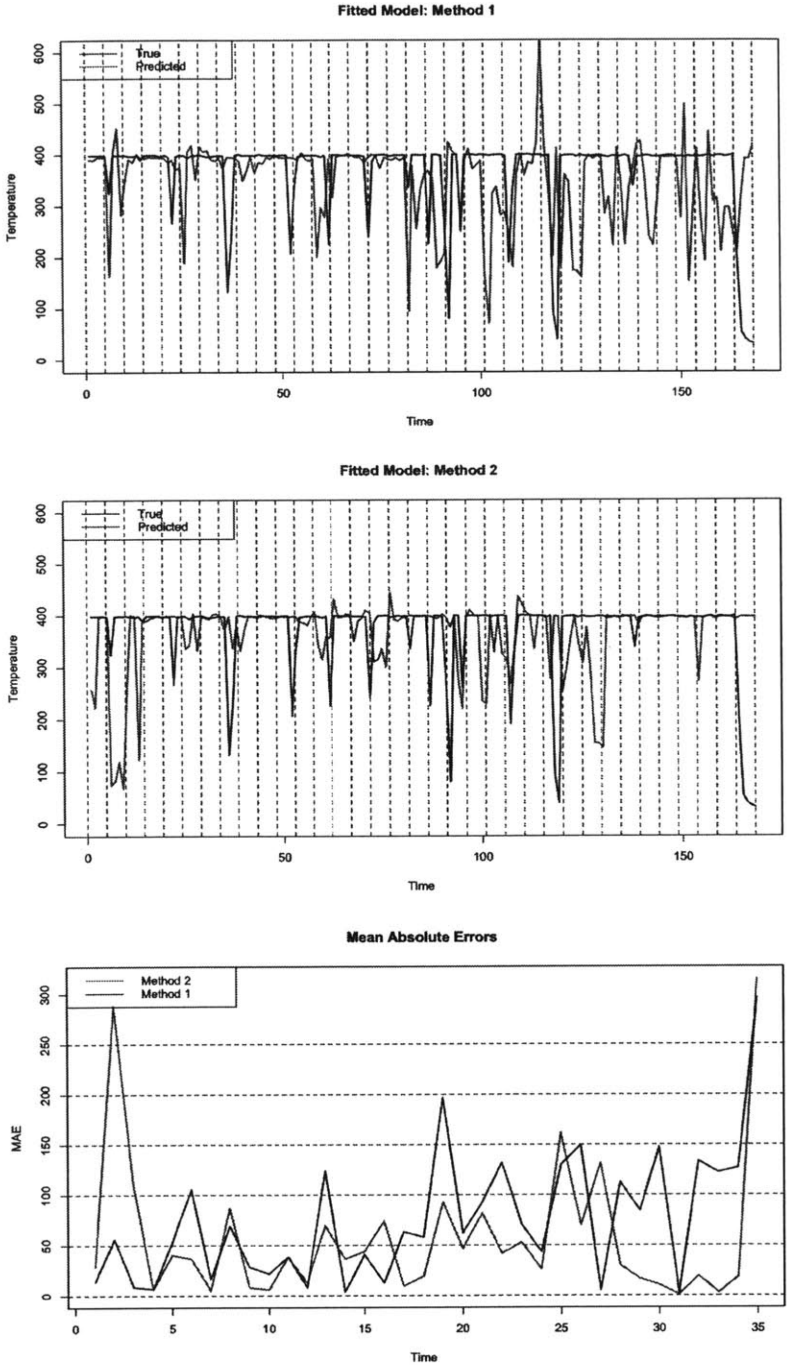

Figure 5-7: Predicted temperature time series using method 1 (panel 1); predictions using method 2 (panel 2); comparison of mean absolute errors (panel 3); (subsampled every 200 data points)

For the HDP-SLDS-HMM (model 3), we can perform rolling prediction of the power signal for the sixth year of data, as well. Similar to the procedure used with model 1, we can split the sixth year of data into blocks of 70-10 days. The first 70 days of every block is used to fit a model, and the adjacent 10 days are used for prediction. We can predict the power signal for 10 day blocks, by simulating the hidden state evolution using the learned transition matrix, and then use the mean power signals as predictions. Preliminary results indicate the trend of the power signal is well predicted by this procedure. Yet, when the power signal reaches extremities, the prediction power decreases.

Chapter 6

Concluding Remarks

The NBP-HMM, HDP-HSMM, and HDP-SLDS-HMM approaches presented in this thesis produce robust estimates of abnormal behavior of temperature signals, whose dynamics are governed by non-stationary dependent variables. The combination of dynamical models and classical linear regression provides a robust procedure that can be used to solve a variety of problems, even outside of reliability analysis. Inherent in all three models was the HMM, which efficiently models the dynamics governing the switch rates between OMs. In the NBP-HMM and HDP-HSMM, regression models were successful in isolating the deterioration within OMs, controlling for factors affecting the subsystem. The BNP-HSMM eliminated the procedure of pre-positing the OM-cardinality, by subsuming it into the Bayesian framework, and allowed for OM dwell time analysis. Results for both models were very similar, shedding light on the overall stability of the approach. The final model explored was the HDP-SLDS-HMM, which detected six OMs. Each of the OMs encapsulated different power-temperature relationships. All three models can be extended to include other critical assets of the GT. Furthermore one can use the modeling framework to do stochastic control, including optimal design of condition-based maintenance. Finally, the information on degradation rates can be used by GenCo to do generation planning, to aid in scheduling its power generation in the face of demand fluctuations in the wholesale market, given that the volatility in generation schedules has a direct impact on the rate at which turbine parts deteriorate.

Bibliography

[1] Alstom. GT24/GT26 Sequential Combustion Gas Turbines. 2012.

[2] Ghafir, A. Performance Based Creep Life Estimation For Gas Turbines. 2011.

[3] T.G. Meyer, T.A. Curse. Low Cycle Fatigue Life Model for Gas Turbine Engine Disks. *J. Eng. Mater. Technol. 102: 45-49* 1980.

[4] A.K.S. Jardine, D. Lin, D. Banjevic. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing 20: 1483-1510* 2006.

[5] N. Puggina, M. Venturini. Development of a Statistical Methodology for Gas Turbine Prognostics. *J. Eng. Gas Turbines Power* 2011.

[6] A. van der Shaft. An Introduction to Hybrid Dynamical Systems. 1999.

[7] M. H. A. Davis. Piecewise-Deterministic Markov Processes: A General Class of Non-Diffusion Stochastic Models. *J. Roy. Statist. Soc. Ser. B* 1984.

[8] M.H.C. Everdij, H.A.P. Blom. Piecewise Deterministic Markov Processes represented by Dynamically Colored Petri Nets. Revised Edition. *National Aerospace Laboratory: Air Transport* 1999.

[9] W. Lair, S. Mercier, M. Roussignol, R. Ziani. Piecewise deterministic Markov processes and maintenance modeling: application to maintenance of a train air-conditioning system. *Pro. Inst. Mech. Eng., Part 0: J. of Risk and Reliability* 2011.

[10] G. Ferrari-Trecatem M. Muselli, D. Liberati, M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica* 39 : 205-217 2003.

[11] J. Roll, A. Bemporad, L. Ljung. Identification of Piecewise Affine Systems via Mixed-Integer Programming. *Automatica*, 2003.

[12] Prokhorov. Hotelling T2-distribution. Encyclopedia of Mathematics. Springer.2001.

[13] Bishop. Pattern Recognition and Machine Learning, 2007.

[14] Shumway, et. al. Time Series Analysis and Its Applications With R Examples. 3rd ed. Springer. 2010.

[15] Johnson, et al. Applied Multivariate Statistical Analysis. 6th ed. Pearson. 2007.

[16] Kutner, et al. Applied Linear Statistical Model. McGraw-Hill. 2005.

[17] Ross, Sheldon. A First Course in Probability, 8th Edition. Pearson Prentice Hall. 2009

[18] Yu, Shun-Zheng. "Hidden Semi-Markov Models". Artificial Intelligence 174 (2): 215-243.

[19] Fox, Emily, et al. "Bayesian nonparametric inference of switching dynamic linear models." Signal Processing, IEEE Transactions on 59.4: 1569-1585. 2011.

[20] Johnson, Matthew J., & Alan S. Willsky. "Bayesian nonparametric hidden semi-markov models." The Journal of Machine Learning Research 14.1: 673-701. 2013.

[21] Teh, Yee Whye, et al. "Hierarchical dirichlet processes." Journal of the american statistical association 101.476. 2006.

[22] Sethuraman, J. A constructive definition of Dirichlet priors. Stat. Sinica, 4, 639-650. 1994.

[23] Beal, M. J., Ghahramani, Z., & Rasmussen, C. E. The infinite hidden Markov model. NIPS (pp. 577-584). 2002.

[24] E.B. Fox, "Bayesian Nonparametric Learning of Complex Dynamical Phenomena," Doctoral Thesis, Massachusetts Institute of Technology, July 2009.