ABSTRACT

Title of Thesis: METHODOLOGY FOR FLEET UNCERTAINTY REDUCTION WITH UNSUPERVISED LEARNING

Ceena Modarres, M.S., 2016

Thesis Directed By: Professor Mark Fuge and Professor Enrique Droguett, Reliability Engineering Program, Department of Mechanical Engineering

Operational and environmental variance can skew reliability metrics and increase uncertainty around lifetime estimates. For this reason, fleet-wide analysis is often too general for accurate predictions on heterogeneous populations. Also, modern sensor based reliability and maintainability field and test data provide a higher level of specialization and disaggregation to relevant integrity metrics (e.g., amount of damage, remaining useful life). Modern advances, like Dynamic Bayesian Networks, reduce uncertainty on a unit-by-unit basis to apply condition-based maintenance. This thesis presents a methodology for leveraging covariate information to identify sub-populations. This population segmentation based methodology reduces fleet uncertainty for more practical resource allocation and scheduled maintenance. First, the author proposes, validates, and demonstrates a clustering based methodology. Afterwards, the author proposes the application of the Student-T Mixture Model (SMM) within the methodology as a versatile tool for modeling fleets with unclear sub-population boundaries. SMM's fully Bayesian formulation, which is approximated with Variational Bayes (VB), is motivated and discussed. The scope of this research includes a new modeling approach, a proposed algorithm, and example applications.

METHODOLOGY FOR FLEET UNCERTAINTY REDUCTION WITH UNSUPERVISED LEARNING

By

Ceena Modarres

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2016

Advisory Committee:
Professor Mark Fuge, Chair and Advisor
Professor Enrique Droguett, Co-Advisor
Professor Ali Mosleh
Professor Balakumar Balachandran

Foreword

The following Master's Thesis is composed of an introduction and conclusion that links two modified journal papers. The first journal paper has been submitted and is under review at Quality and Reliability Engineering International. The second chapter of this thesis corresponds to this first paper. The second journal paper is still in the pre-submission drafting process. The third chapter of this thesis corresponds to the second journal paper. The introduction and conclusion weave together these two papers into a single body of work.

Acknowledgements

The author acknowledges the Petroleum Institute University and Research Center and the Pipeline System Integrity Management project for funding this research. The author also acknowledges the University of Maryland's Center for Advanced Life Cycle Engineering (CALCE) for its help in accessing and wrangling the PHM data.

I would like to take a brief moment to thank a few people that were critical to the completion of this Master's Thesis. First, I thank my advisors Dr. Mark Fuge and Dr. Enrique Droguett. Both were invaluable for their technical and non-technical support. Their mentorship was even more important. Second, I thank Dr. Ali Mohammad-Djafari. He provided a new direction when the project stalled. Third, I thank my peers: Roohollah Heidary, Amin Aria, Anthony Coburger, Guillaume Deis, and Raphael Chinchilla. Most importantly, I thank my parents. Without their support, this research would never have started.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Many modern reliability and maintainability techniques reduce uncertainty around system life and integrity measurements. Particle Filtering (PF) and Bayesian Networks (BN) are two common Prognostic and Health Management (PHM) approaches for reducing unit uncertainty [1-3]. PF can be employed for on line physical system tracking given relevant sensor-based system health data. PF is a generalization of Kalman Filtering that uses particles to approximate the probability density of a relevant system health metric [1]. Static Bayesian Networks and the recursive Dynamic Bayesian Networks can account for numerous operational and environmental conditions that may affect relevant system health variables [4-5]. These techniques provide prognostics that can be leveraged for CBM.

These techniques all rely on the same assumption: reliability metrics extrapolated from an entire fleet (a full population of units) are too imprecise for application to any particular unit. Varying operational loads, environmental stressors, or other unforeseen factors produce inter-fleet variance and heterogeneity. Modern data-driven modeling methods often account for this uncertainty on a unit-by-unit basis. By accounting for uncertainty on a unit-by-unit basis, an analyst can institute cost saving condition-based maintenance (CBM) [6].

CBM is a valuable cost-saving objective, but it should not be considered the primary purpose of uncertainty reduction. Optimizing scheduled maintenance (SM) is still necessary to maximize system life. In some cases CBM is not practical. Existing techniques often identify damage when 80-90% of system life has already been expended [6-7]. The fleet may be too large to conduct analyses on each unit. For example, consider an oil pipeline network and its metallic pipes. Certain parts of the network are likely to see larger cyclic loads and different environmental stressors. These heterogeneous stressors could cause some metallic pipes to fail before others, but the sheer size of the system prevents rapid condition based maintenance. Adjusting schedules could improve maintenance scheduling for pipes that are more likely to fail and prevent wasteful maintenance on more reliable pipes. Also, CBM provides limited information for optimal resource allocation. Precise and variant warranty schedules could systematically save organizations significant costs.

This thesis proposes discrete population segmentation to identify sub-populations within a fleet. By identifying sub-populations within the fleet, a reliability practitioner can reduce uncertainty by factoring larger operational and environmental trends into the analysis. Because identifying underlying patterns, this methodology can improve SM, warranty schedules, and other resource allocation. The methodology can also be used in conjunction with other data-driven techniques to improve SM and CBM planning. Much like BN and PF, this methodology leverages sensor-based covariates to reduce this uncertainty. The proposed methodology employs unsupervised clustering algorithms to segment the population by its covariates. By training a supervised classification model with the sub-population assignments, the methodology can segment still unsold, unused, or healthy units. This thesis explores the advantages of this approach in depth and considers alternative scenarios where sensor data is unavailable.

1.2 Relevant Background and Discussion

Fleets are often subject to different environmental and operational conditions. Moreover, systematic sensor measurement error, reporting inaccuracy, and production process variations often create uncertainty around individual units within a fleet [8]. A probabilistic time-to-failure prediction applied to an entire fleet may over generalize the failure probability of a unit with a different magnitude of stress. By accounting for heterogeneity, it is possible to improve the precision and accuracy of reliability predictions and prognostic health management (PHM).

Past literature has addressed fleet uncertainty in several ways. Kaplan employed a two-step Bayesian procedure to develop a generic prior that specializes predictions with data related to a particular sub-population [9]. Droguett & Mosleh condensed Kaplan's two-step Bayesian procedure by evaluating hierarchical uncertainty around the distribution of a hidden variable's parameters [10]. Liu & Zio trained a unit correction model on a fleet-wide degradation path regression [11].

These approaches only correct for uncertainty in a particular analysis, but the proposed methodology addresses fleet wide heterogeneity. In particular, this thesis recommends a clustering based framework to handle uncertain fleet-wide operational and environmental conditions. Cluster analysis segments data by the similarity or dissimilarity of the observations. An observation assigned to a cluster would be most similar to other observations assigned to the same cluster [9]. This metric of similarity can be any feature or collection of features. In reliability analysis, each cluster would correspond to a sub-population.

Clustering has appeared in the reliability literature previously. Tian (2002) showed that clustering the failure intensities could improve reliability analysis [12]. Similarly, Dindarloo and Siami-Irdemoosa (2016) clustered mining shovel data based on time to failure and repair [13]. Arunajadai et. al clustered the attributes of a product design to identify potential failure modes [14].

One of the more commonly used approaches to counter heterogeneity in reliability is a clustering technique known as mixture modeling (MM). MM assumes the data is produced from some finite number of distributions and each data point is produced by one of the predicted distributions with some probability. Equation 1.1 displays the mathematical representation of mixture modeling, which represents a number of distributions ($F_i(t)$) with a particular mixing proportion (p_i) [9].

$$F(t) = \sum_{i=1}^{n} p_i F_i(t) \tag{1.1}$$

Mixture modeling is a robust tool for time-to-failure analysis as each component distribution can represent a separate sub-population [16]. If there is sufficient heterogeneity within a fleet, a single distribution may not effectively fit the time-to-failure data. A mixture model directly accounts for the heterogeneity of failure events.

Unlike clustering applications in reliability, the proposed methodology capitalizes on the popularization of sensor technology to increase the precision and accuracy of reliability diagnostics and prognostics by segmenting the fleet into sub-populations. The proposed methodology leverages sensor data to systemically inform and reduce fleet uncertainty. Not only can this benefit reliability management, but it also provides a holistic perspective of the fleet's heterogeneity. By clustering covariates directly, a reliability practitioner can assign labels for analysis on unsold, unused, or healthy units.

1.3 Research Objectives

The authors primary objective is to propose a clustering based methodology that reduces fleets uncertainty by identifying sub-populations with sensor-based covariates. The proposed methodology aims to segment the most similar units into sub-populations and separate dissimilar units. Post-methodology time-to-failure analysis should then outperform time-to-failure analysis on an entire fleet. The methodology should introduce a mechanism to classify additional units to a sub-population. Finally, the methodology should achieve objectives despite noise and outliers. In the real world, sub-populations boundaries are not clear. The methodology should obtain the stated objectives despite these possible pitfalls.

1.4 Research Contributions

The research contributions of this thesis are:

- Introduce a methodology for fleet uncertainty reduction that accurately identifies sub-populations, improves time-to-failure distribution fit, and increases precision of reliability metrics.
- Motivate Student-T Mixture Models to identify sub-populations despite real world noise, outliers, and unclear sub-population boundaries.
- Adapt Variational Inference to quickly and efficiently approximate Student-T Mixture Models.
- Discuss Variational Bayes applications in Reliability Engineering.

1.5 Thesis Organization

Chapter 2 introduces, validates, and demonstrates a clustering methodology with the DBSCAN algorithm. Chapter 3 employs SMM as an effective algorithm for applications to cases where the sub-population boundaries are blurred or covariates are unavailable. VB is motivated for reliability contexts. Chapter 4 concludes by summarizing academic contributions and recommending future work. The appendix include additional mathematical derivations and experiments that were not included in the main text.

Chapter 2: A Novel Clustering Based Methodology for Overcoming Heterogeneous Populations for Reliability Prediction

The reliability field is in the process of a transformative change with the advent and popularization of sensor technology to process extra covariate information on a particular system. The ever-increasing amount of information provides an opportunity for the adaption of machine learning to not just improve, but revolutionize, the fit, accuracy, and sheer volume of reliability diagnostics and prognostics [17]. A reliability data set of failure events may now include numerous time series of attributes leading up to and at a conclusive failure event.

The proposed methodology clusters sensor based covariate data for the identification of possible sub-populations for specialized reliability prediction. Applied to a data set with messy, heterogeneous, data, the methodology segments the data set into homogenous sub-fleets. Homogeneity, in this context, can be defined by subgroups with similar reliability characteristics, like failure probability or hazard rate. In these smaller sub-populations, more accurate and precise reliability models (i.e., distributions) can be derived.

This chapter proceeds as follows: section 2.2 reviews the complete methodology and discusses practical concerns for the clustering algorithm. Section 2.3 validates the proposed methodology with simulated data. Section 2.4 demonstrates the methodology on a real power plant data set and discusses results.

2.2 Proposed Methodology

Given a data set that includes failure times and the covariate measures of the respective units, the first step is to cluster the covariate information. After clustering the data into its relevant subgroup, separate reliability metrics are calculated from smaller, more homogenous, sub-populations. Now any unused, unsold, or healthy unit within the same population can be assigned to a subgroup for more accurate and precise prognostics. Thus, in the aforementioned oil pipeline example, uncertainty around reliability predictions for newly installed pipes can be immediately reduced.

In order for the methodology to assess additional units, the cluster assignments should be used as class labels and a supervised learning model (e.g., Logistic Regression) trained. The supervised learning model should be subject to regularization, validation, and testing. The process is graphically displayed in Figure 2-1.

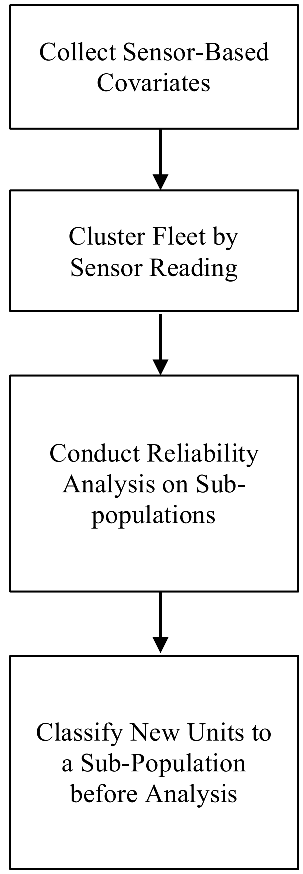

Figure 2-1: Proposed Clustering Methodology

2.2.2 Intricacies of Clustering Covariate Information in Reliability

As displayed in Figure 2-1, the first step of the methodology clusters the data based on covariate information to segment the data into more similar, smaller, subgroups. However, it is important to produce clusters that best represent sub-populations within a larger population. The best representation would group the most similar units into a sub-population and avoid the inclusion of dissimilar units. A number of practical concerns, like class imbalance, are considered in this section.

First, the clustering algorithm should effectively handle class imbalance. If a data set has a large group that contains 95% of the observations and three smaller groups that compose the other 5%, the approach should be able to identify the four heterogeneously sized clusters.

Second, the approach should be scalable. For example, subspace-clustering algorithms, which map higher dimensional data to lower-dimensional representations, may be robust for non-convex data, but require the construction of a similarity matrix [31]. As the number of observations (N) increases, the similarity matrix becomes too large. However, there are a number of approaches that can handle large-scale sub-space clustering through sparsity and sampling, like Scalable Sparse Subspace Clustering (SSSC) and Large-scale Subspace Clustering using Sketching and Validation (SkeVa-SC) [32-33].

Third, similar units should be grouped together and dissimilar units segmented. The similarity measure poses a unique challenge. Similarity or dissimilarity is a critical aspect of many machine-learning algorithms. Observations or data points must be compared to each other by some measure. In the context of this methodology, the approach compares each unit's similarity to another unit by its covariates.

The selected similarity measure is often dependent upon the type of data. If it were a continuous data set, then the simplest choice would be a geometric measure like Euclidean distance. However, for ordinal, binary, categorical, or mixed data sets, the decision becomes more challenging. For ordinal, binary, or categorical data, the dissimilarity could be a function of how many variables the observations have in common. For mixed data set, one possibility is to simply transform continuous data to categorical or ordinal data and employ the Hamming Distance (2.1) [20] or the Gower Similarity Coefficient (2.2) [38].

$$d^{HAD}(i,j) = \sum_{k=0}^{n-1} \left[y_{i,k} \neq y_{j,k} \right] \tag{2.1}$$

$$S_{i,j} = \frac{\sum_{k}^{n} \omega_{ijk} S_{ijk}}{\sum_{k}^{n} \omega_{ijk}} \tag{2.2}$$

Hamming Distance assesses whether each feature of two different data observations is identical ($y_{i,k} = y_{j,k}$) or not ($y_{i,k} \neq y_{j,k}$). The similarity metric increases, or inversely dissimilarity decreases, for each feature the two observations have in common. Hamming Distance has obvious flaws in its application to continuous or mixed data types. The Gower Similarity Coefficient is very effective in its application to mixed data sets. It uses an additional parameter ω_{ijk} to control for invalid comparisons. So, a continuous variable will never be compared to a binary or categorical variable. However, when the two variable types match, the variable S_{ijk} uses an appropriate similarity metric (i.e., Hamming with binary or categorical variables and Euclidian with continuous variables).

Neither Gower Similarity Coefficient nor Hamming Distance is universally applicable. Many clustering algorithms, like k-Means or Mixture Models (MM) [4], cannot cluster mixed data types since they require continuous parameters to iteratively update. Alternatives to traditional algorithms, like k-Prototypes [35], have been proposed to handle this problem. Density Based Spatial Clustering of Applications with Noise (DBSCAN) [36] can, also, operate on various similarity matrices efficiently and accurately. Although the experimental data sets are only composed of continuous variables, DBSCAN is used in upcoming experiments. Euclidean Distance is chosen as the similarity metric.

2.2.3 DBSCAN

The DBSCAN algorithm is useful in reliability contexts because it can handle data of arbitrary shapes and sizes, requires minimal inputs, and is efficient on big databases. DBSCAN also assumes no distribution, which allows it to handle non-convex subsets.

The density-based notion of clusters is rooted in a few definitions. The first is an Eps-neighborhood (N_{Eps}) of a point where is an additional observation [36]:

$$N_{Eps}(p) = \{q \in D | dist(p, q) \leq Eps\} \qquad (2.3)$$

Two observations (p) and (q) lie within the same neighborhood if their similarity, per the selected measure, is within a specified epsilon hyperparameter.

The second is direct density-reachability*:*

$$p \in N_{Eps}(q) \tag{2.4}$$

$$\left|N_{Eps}(q)\right| \geq MinPts \tag{2.5}$$

Direct density reachability implies two points within the same neighborhood. The neighborhood must be greater than a specified minimum points (MinPts) parameter. [20] MinPts determines the least amount of points required to form a cluster [36].

Density Reachability is defined as a group of points $p_1 \dots p_n, p_1 = q, p_n = p$ such that p_{i+1} is directly density-reachable from p_i with respect to (wrt) Eps and MinPts. [36] Density reachability is the transitive direct density-reachability. Now, points can be indirectly connected but belong to the same neighborhood. Observations *p* and *q* are density-connected if they are both density-reachable from a third point *o* wrt Eps and MinPts [36].

A cluster is then defined as a non-empty subset of the data that meets the following two criteria:

1.$\forall p, q$: if $p \in C$ and q is density-reachable from *p* wrt Eps and MinPts, then $q \in C$ [36]

2. $\forall p, q \in C$: *p* is density-connected to q wrt. Eps and MinPts [36]

Finally, an analyst can define all points that are not assigned to a particular cluster, from lack of density reachability or connectability, as noise points.

These definitions are integral to DBSCAN's two-step iterative procedure for clustering. First, an arbitrary point is selected and the core point condition is evaluated. The core point condition requires that the units' neighborhood $N_{Eps}(q)$ contain more than MinPts observations. If the observation is designated a core point, this selected observation is considered a seed. Then, all points density-reachable or density-connected from this seed form a cluster. Each cluster is evaluated against the Eps hyperparameter to determine whether a cluster should be combined or separated. The algorithm is run until convergence [36].

Hyperparameter optimization is a critical task in machine learning. For DBSCAN, the optimal Eps and MinPts parameters produce the most homogenous sub-populations. Manual search, grid search, and randomized search are three conventional approaches to optimizing hyperparameter. However, identifying the "thinnest", or least dense, clusters' Eps parameter is an applicable heuristic for optimizing the selection. Since Eps in DBSCAN is the minimum criteria for the formation of a cluster, the "thinnest" clusters' Eps can be selected as a global parameter because it specifies the lowest possible density that would not be considered noise [36].

In order to assess this value, a function *k-dist* is defined as each point's distance metric to its k-th nearest neighbor. Sorting points and the corresponding distance metric based on descending *k-dist* values produces a *k-dist* graph. This *k-dist* graph can be used to reveal a threshold point that provides the Eps value of the thinnest cluster. In the *k-dist* graph, the large value before the curve's knee will reveal the value of this parameter for a particular *k* [36]. This *k* value is the MinPts parameter. The selection of the optimal MinPts can be optimized or selected based on engineering rationale. At what size is it no longer reasonable to consider something a sub-population? This decision can also be made from the size of the data and engineering or statistical intuition. One concern for an analyst should be over fitting. If the model is over fit to the data, it may not effectively represent additional experiments or field data.

After DBSCAN is applied and cluster assignments are obtained, the population can be segmented into smaller, more homogenous, sub-groups. Now, a reliability engineer can derive separate predictions on each sub-population.

2.3 Validation of Proposed Methodology on Synthetic Data

In this section, the author validates the proposed methodology on a synthetic data set. The data was synthesized from four separate Weibull distributions with five covariates. The covariates were heterogeneous; each sub-population of covariates was sampled from four separate Gaussian distributions. Moreover, the simulated sub-populations were heterogeneously sized with one of the groups accounting for 43.5% of the full data set and another accounting for 13%. The total data set amassed 1150 observations and the covariate matrix used for clustering was five dimensions.

A scatterplot matrix of the synthetic data set is shown in Figure 2-2. The scatterplot matrix displays the data across two dimensions for every possible combination of dimensions. Moreover, across the diagonal, it shows a histogram of the synthetic data within each dimension.

In this simulated case, the covariate information is synthetic. Labeled covariate information is not necessary for segmenting the data of homogenous sub groupings. However, in practice, each collected covariate would represent a variable relevant to failures like humidity, temperature, or air pressure.

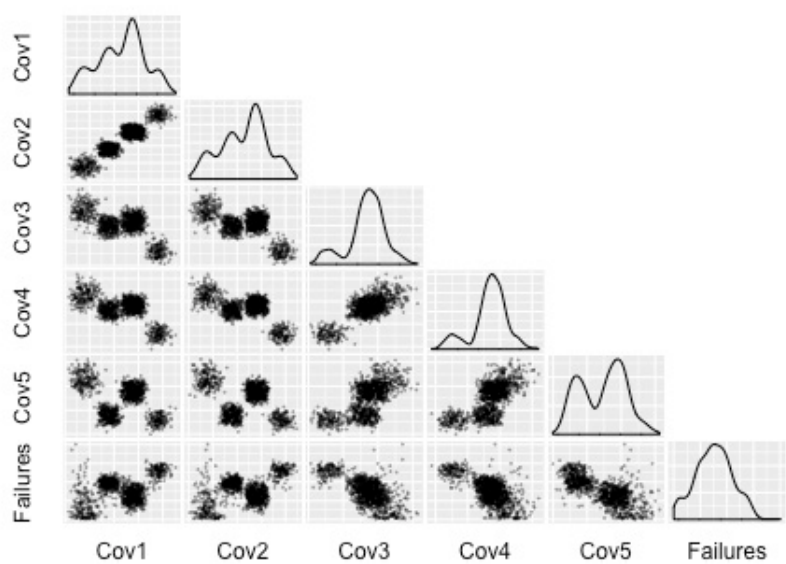

Figure 2-2: Scatter Plot Matrix of Noisy Synthetic Data. Covariates are produced by four different Gaussian distributions, failure times by four different Weibulls

The author tested DBSCAN on covariate data to examine its ability to segment a population. In short, could DBSCAN effectively cluster heterogeneous data? Since the data set was quite small (N=1150), the dimensionality was chosen as the parameter value MinPts (MinPts = 5) and the Eps value (Eps = 0.11) was calculated with the KNN-dist graph.

After implementing DBSCAN on covariates, the failure times were segmented by cluster assignment. As visualized in Figure 2-3, the algorithm effectively segmented the population. All but 14 out of 1150 data points were classified to the correct sub-population, resulting in 98.7% accuracy.

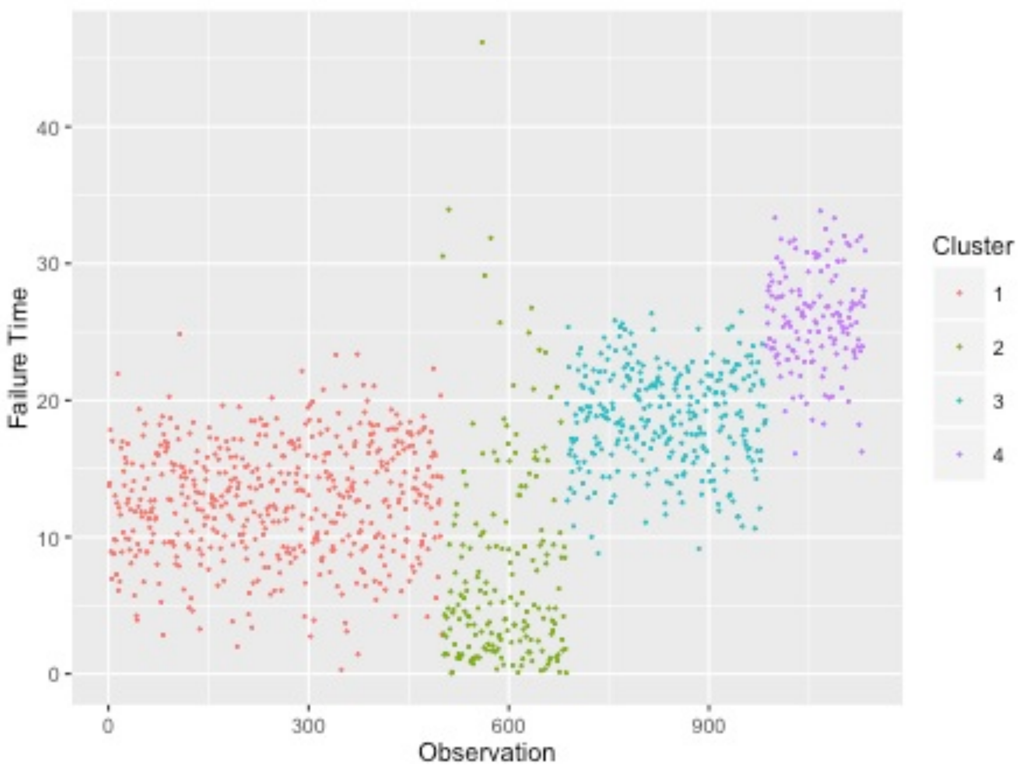

Figure 2-3: DBSCAN Cluster Assignments of Failure Times for Synthetic Data

DBSCAN discovered homogenous subgroups, which indicates the clustering step of the methodology can identify and segment heterogeneous populations. The extent to which clustering improves reliability analysis has yet to be validated. In order to assess the second step of the methodology (Figure 2-1), each identified sub-population's failure times were fit with a Weibull distribution. These sub-population lifetime distributions were compared to a lifetime distribution fit to the full data.

Figure 2-4 shows a fit to the full synthetic data set and Figure 2-5 shows a fit to the largest sub-population. The results provide evidence that clustering the data can improve fit and robustness of prediction. This improvement can be quantified by comparing the log likelihood scores for the aggregate population, random samples from the aggregate population that are the same sample size as each sub-population, and each sub-population. The results are shown in Table 2-1.

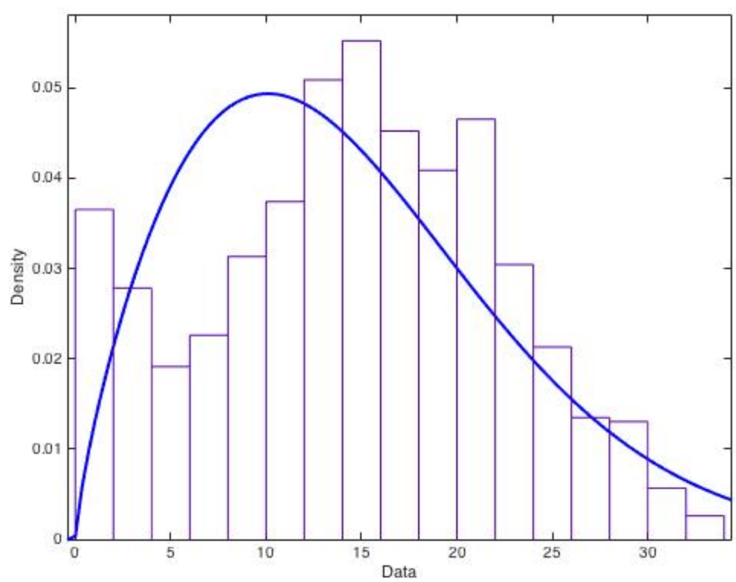

Figure 2-4: Weibull Fit – Full Synthetic Data

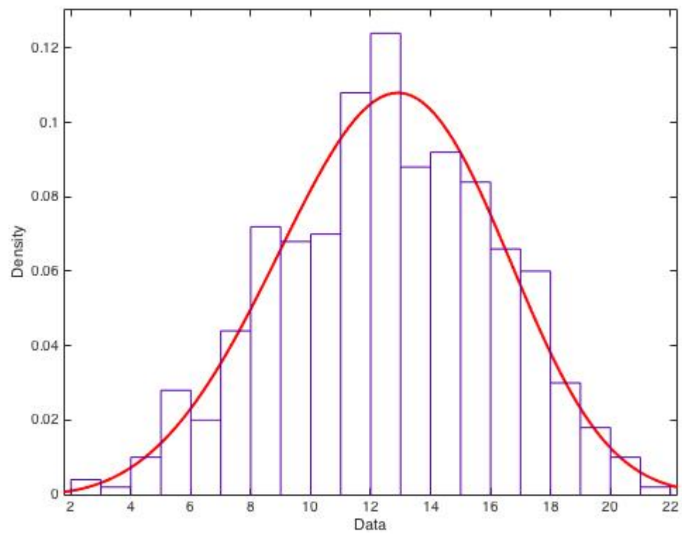

Figure 2-5: Weibull Fit, Largest Sub-Population (n=500)

Table 2-1: Model Comparison

	# Of Data Points (N)	**Log-Likelihood**	**Random Sample Log-Likelihood***	**AIC**	**Random Sample AIC***
Full Population	1150	-4030		8065	
Sub-Population 1	500	-1349	-1756	2702	3517
Sub-Population 2	187	-412	-652	827	1309
Sub-Population 3	300	-796	-1050	1595	2104
Sub-Population 4	149	-401	-517	806	1039
*In this case the same number of random points are sampled from the full population to evaluate the effect of clustering keeping the number of data points constant.					

Likelihood based scores, like the log-likelihood and the AIC, automatically increase with larger samples. Thus, it is important to provide metrics that can hold the sample size constant. The author took random samples from the full data of equal size to each sub-population and compared the results to the clustering methodology. The distributions generated from the methodology were more likely than the random samples. Table 2-1 also includes AIC scores, which is an information criterion for choosing the best possible model that penalizes additional parameters [31]. The methodology avoids additional parameters, so it is not penalized despite the additional step.

Parameters predicted by the methodology, as displayed in Table 2-2, are quite accurate. The simulated distributions resemble the newly estimated distributions.

Table 2-2: Parameter Estimates

Weibull Parameters	Sub-Population	Synthetic Parameter Value	Methodological Parameter Estimates
β (Shape)	1	4	3.93
	2	1	1.07
	3	8	6.49
	4	12	8.32
α (Scale)	1	14	13.90
	2	8	8.38
	3	20	20.24
	4	27	27.61

The fourth cluster's shape parameter is estimated incorrectly. This could be the result of the small sample Monte Carlo simulation. The randomness from a small sample size may have altered the most descriptive shape parameter. Regardless, DBSCAN accurately identified 99.3% of the data points in this sub-population.

Results on synthetic data verified and validated the proposed methodology. In the following section, the clustering methodology is applied to a real data set.

2.4 Application of Methodology to Power Plant Failures

The PHM Data Challenge Competition 2015 (PHM 2015) was a competition to predict failure events in complex power plant systems from unlabeled sensor-based covariates [37]. PHM Society provided a large database of unlabeled, time series sensor measurements for a set of control components within various plant zones. The data set included sensor measurements for non-failure times and a categorical variable defining the variable type, but this information was not used in the presented analysis.

In total, the data set included 16,274 failure events and 30 different covariates. Unlike the synthetic data set, it is too high dimensional to visualize and no longer a mixture of known distributions. The sub-population boundaries are noisier and there are no labels. Figure 2-6 visualizes fleet-wide predictions by fitting a Gaussian distribution to all failure times and demonstrates that a fleet-wide assessment would over or under-estimate failure probability.

The authors applied the proposed methodology to the data collected from this complex power plant system. Per the first step of methodology (Figure 2-1), the population is clustered based on collected covariate data.

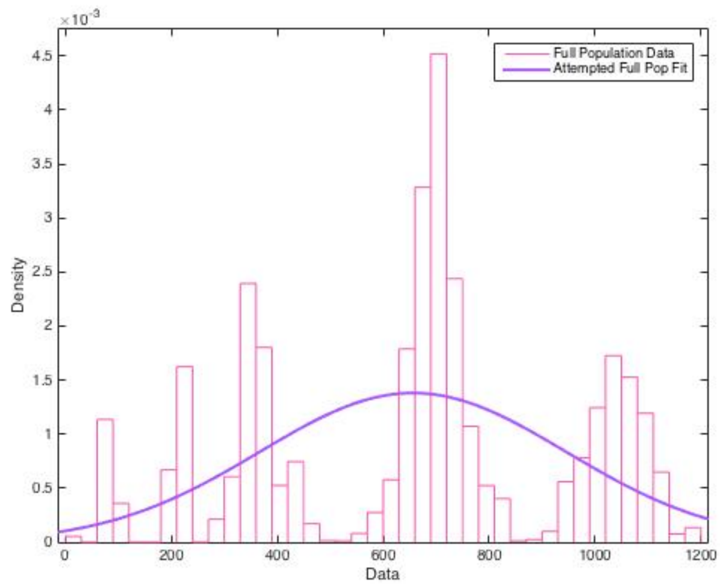

Figure 2-6: Gaussian Fit to PHM Data. Data (X-axis) represents a particular failure time and Density (Y-axis) represents the quantity of failure events at a particular time

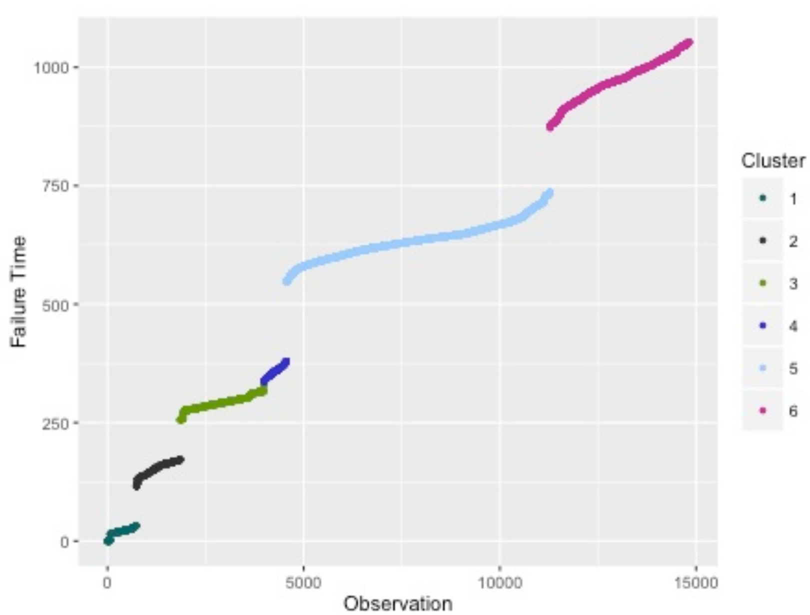

Figure 2-7: Clustering by Covariate

As shown in Figure 2-7, the clustered data has segmented the population into seven sub-populations. Different parametric distributions may best represent each sub-population. Some sub-populations include tails, but acting on these smaller subsets is still far more advantageous and robust than the large, heterogeneous, full data set.

Indeed, Figure 2-8 displays an attempted cumulative distribution function (CDF) fit to a full, heterogeneous data set. The Gaussian distribution was chosen because it provided the best fit of any basic parametric distribution. The Gaussian distribution over and underestimates the probability of failure at various intervals. This would overgeneralize and generate inaccurate predictions.

After clustering the data, an analyst can apply reliability assessments to the sub-population. The CDF fit to the largest sub-population, as displayed in Figure 2-9, has notably improved the fit and, thus, predictive capabilities of the model.

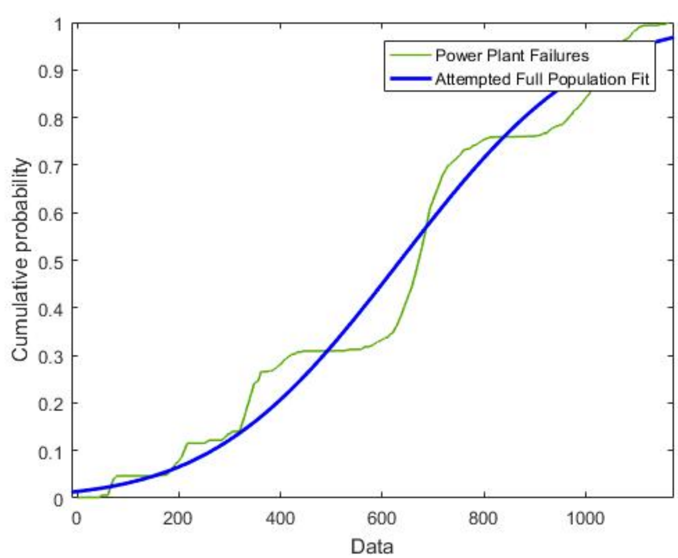

Figure 2-8: Attempted Normal CDF (Best) Fit to Full Data (n=16,274)

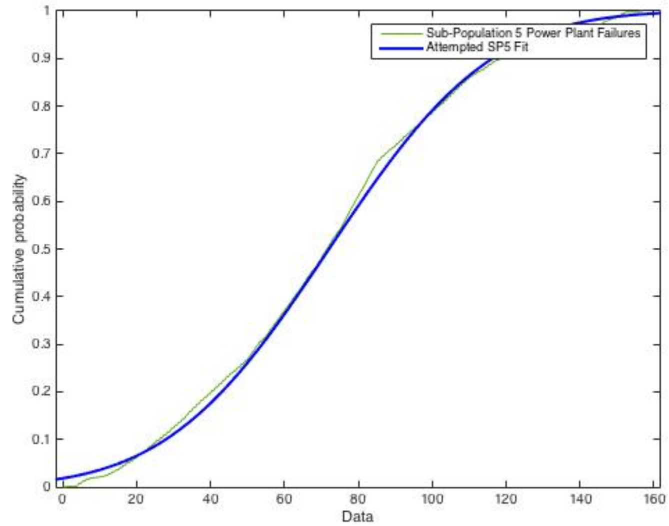

Figure 2-9: Normal CDF (Best) Fit to Largest Sub-Population (N= 6,718)

Table 2-3: Model Selection for PHM 2015 Data

	# Of Data Points (N)	**Distribution (Best Fit)**	**Log-Likelihood**	**Random Sample Log-Likelihood***	**AIC**	**Random Sample AIC***
Full Population	16274	Normal	-115302		230608	
Sub-Population 1	731	Normal	-1520	-5201	3045	10405
Sub-Population 2	1120	Weibull	-4260	-7944	8525	15892
Sub-Population 3**	2140	Normal	-5787	-15213	11578	30431
Sub-Population 4	569	Normal	-1700	-4043	3405	8091
Sub-Population 5	6718	Normal	-31952	-47543	63907	95090
Sub-Population 6	2980	Normal	-14825	-21137	29651	42277
Sub-Population 7	301	Normal	-866	-5201	1735	10405
*Consult Table 1 for details. **Sub-Population clustered a consecutive time.						

Table 2-4: Relevant Reliability Metrics

	MTTF (Days)*	**Conditional Reliability at Full Population MTTF (656 Days)**	**B(10) Life***	**B(90) Life***
Full Population	656	50%	286	1,026
Sub-Population 1	86	0%	82	91
Sub-Population 2	220	0%	203	234
Sub-Population 3	359	0%	344	374
Sub-Population 4	423	0%	412	434
Sub-Population 5	701	91%	656	745
Sub-Population 6	1,038	100%	993	1,083
Sub-Population 7	1,110	100%	1,105	1,116

As Table 2-3 shows, the log-likelihood and AIC scores of the clustered data outperform random samples of the same size from the full data set. The fit of the clustered data outperforms that of a random sample by as much as a factor of six.

Table 4 shows that the B(10) and B(90) life, or the time that 10% and 90% of the population has failed, and the Mean-Time-to-Failure (MTTF) may vary dramatically between sub-populations. For example, the reliability at the MTTF of the full population (656 days) varies by sub-population. This behavior can be considered a sign of high fleet uncertainty. For many units, the application of the methodology changes the MTTF. By leveraging various MTTF estimates for each sub-population, an analyst improves accuracy. Also, the reduced range between the B(10) and B(90) life signals more precision.

An analyst can now assign unsold, unused, or healthy units to a sub-population. The analyst could simply re-cluster the entire data set including the additional point or treat cluster assignments as class labels. A supervised classification algorithm could be trained to predict the cluster assignment of additional observations by covariate. [9] This additional step is demonstrated in Chapter 3.

2.5 Chapter Summary

In this chapter, the author showed the potential of a fleet segmentation methodology. By clustering sensor-based covariates, fleet uncertainty is reduced into discrete sub-populations. The proposed methodology improved reliability life estimates by accounting for uncertainty directly. Now, a reliability analyst can vary resource allocation and warranty schedules by sub-population. DBSCAN was employed as the clustering algorithm.

The author found a serious limitation in the methodology. DBSCAN removed noise points from the segmentation. A number of points were left unassigned. In the real world, sub-population boundaries are unclear and can overlap. For the methodology to be applicable in real world cases, the algorithm should assign these points to a sub-population. For example, components within a pipeline system are not likely to suffer from dramatically different operational and environmental conditions. It is important to account for the uncertain boundaries between two pipeline sub-populations and not ignore these observations when modeling a fleet. In Chapter 3, the author explores the Variational Bayesian Student-T Mixture Model (VBSMM). VBSMM is a heavy tailed mixture of distributions that can account for real world noise and outliers when segmenting a fleet.

Chapter 3: Fleet Uncertainty and the Variational Bayesian Mixture of Student-T Distributions

Thus far, this thesis has argued that clustering relevant covariates with DBSCAN could improve the precision and accuracy of fleet reliability analysis. A reliability practitioner can leverage inter-fleet covariate variance to tackle heterogeneity. However, this data-driven approach assumes sensors will collect information directly relevant to predominant failure mechanisms. There are also circumstances in which sensor installation is impractical. For example, sensor installation can be logistically difficult or cost-ineffective if creep or corrosion is a fleet's leading failure mechanism. Moreover, sub-population boundaries are often blurred. Each unit is not subject to extreme conditions. This extra real-world noise may disrupt the fit of mixture models for many distributions. Unless a distribution has large tails, the mean and variance could be unnecessarily skewed. These blurred lines may force some mixture models to produce an undesirable fit.

In this chapter, the author proposes fleet segmentation with Variational Bayesian Student-T Mixture Models (VBSMM) to counter those real world practical concerns. The Student-T distribution uses a degrees of freedom (υ) parameter to model the size of its tail. As $\upsilon \rightarrow \infty$, the Student-T distribution will converge to a normal distribution. SMM are a robust approach to handling real-world noise and inter-fleet ambiguity that negatively impacts other distributions [40]. VBSMM offers a novel method in reliability and PHM settings.

In this chapter, the author discusses the Expectation-Maximization (EM) and Variational Bayesian Expectation-Maximization (VB-EM) formulation of SMM. The two methods are quite similar, but the Variational Bayesian (VB) formulation introduces prior distributions that can regularize predictions. By regularizing, reliability practitioners can use expert opinion or data from related experiments to improve estimates. VB can be an attractive alternative to Markov Chain Monte-Carlo (MCMC) techniques for approximating intractable Bayesian inference problems. MCMC approaches are computationally expensive in certain circumstances and have difficult to assess convergence properties.

Several previous reliability analysis papers have employed Variational Inference on hierarchical hybrid Bayesian Networks with applications in gas turbine engine prognostics and large-scale integrated circuits [71-73]. VB has been leveraged to approximate Non-Homogenous Poisson Process (NHPP) software reliability modeling [74]. None of these VB reliability applications solve mixture models or address fleet uncertainty.

Since SMM is a parametric distribution that can be integrated to produce a reliability function, it can model failure events directly from time-to-failure data or from sensor-based covariates. A reliability analyst can leverage failure related covariates to segment units to sub-populations with more similar time-to-failure events or degradation paths. An analyst can improve his analysis by assigning unused, unsold, or operational units to a sub-population with relevant covariates.

In section 3.2, the author introduces the proposed methodology, motivates VB, and derives VBSMM. In section 3.3, the author validates the robustness of SMM in handling outliers by comparing it to Gaussian Mixture Models (GMM) on a synthetic signal data set. The author then demonstrates how SMM can segment a population with time-to-failure data or sensor based covariates on real power plant data.

3.2 Proposed Approach

There are two possible approaches to reliability assessments with SMM. SMM could be directly applied to time to failure data or to relevant covariate sensor information. In this case, the analyst can use SMM as the clustering algorithm in the Figure 2-1 methodology or ignore it entirely. As previously mentioned, the clustering methodology is more robust for predictions on additional units. In this chapter, both possibilities are examined.

Equation 3.1 displays the probability density function (PDF) of a multivariate Student-T distribution where Γ is the gamma function, d is the dimensionality of the data, v is degrees of freedom parameter, μ is the mean parameter vector, and $\boldsymbol{\Sigma}$ is the variance.

$$S(x|\mu,\sigma,v) = \frac{\Gamma(\frac{v+d}{2})}{\Gamma(\frac{v}{2})v^{\frac{d}{2}}\pi^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\left(1+\frac{1}{v}(\boldsymbol{x}-\boldsymbol{\mu})^{T}\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}-\boldsymbol{\mu})\right)^{-\frac{v+d}{2}} \qquad (3.1)$$

In order to segment a population with Student-T, the author evokes a finite mixture of Student-T distributions. This clustering procedure, commonly known as SMM, follows the canonical form displayed in Equation 1.1. The hierarchical structure of SMM involves the calculation of a hidden variable (z). In this case, the hidden variable is the probability of a single observation belonging to each particular distribution. The distribution with the highest probability can then be considered the cluster assignment.

Bayesian methods are probabilistic techniques that evoke prior and incoming information to produce distributions that describe some random variable. In this case, Bayesian methods are employed to estimate the parameters of the SMM. Bayes theorem is presented in Equation 3.2.

$$p(\theta|X) = \frac{p(X|\theta)p(\theta)}{p(X)} \propto p(X|\theta)p(\theta) \qquad (3.2)$$

Bayesian methods can be evoked to build distributions around the parameters (θ) that describe the random variable of interest. In Equation 3.2, $p(X|\theta)$ is the likelihood of the evidence X and $p(\theta)$ represents the prior distribution of θ. Thus, the parameter itself is a random variable. To understand the value of this information, consider the case of a Gaussian distribution that describes a set of failure times with a mean parameter (μ). This mean parameter applied to time-to-failure data could describe the reliability metric mean time to failure (MTTF). With this hierarchical procedure, a reliability practitioner can regularize this important metric with prior information before it is derived.

In these cases, the likelihood term $p(D|\theta)$ integrates out parameter θ, since only the prior parameters describing the random variable θ are relevant to the ultimate calculation. Marginal likelihoods are critical to the construction of complicated distributions like SMM. Equation 3.3 displays a marginal likelihood.

$$p(D|\,\alpha) \;=\; \int p(D|\theta)p(\theta|\alpha)\,d\alpha \tag{3.3}$$

Calculating the marginal likelihood is often intractable. The integral can become complicated and multidimensional. Thus, the hierarchical formulation of a SMM cannot be solved analytically. In order to produce parameter estimates, an analyst must approximate the parameters.

3.2.2 Approximation Selection

The previous section has shown how VB can be used to approximate intractable Bayesian problems. However, why use this approximation as opposed to the frequentist Expectation Maximization or MCMC?

The EM algorithm (Algorithm 3-1) is essentially identical to the VB-EM algorithm (Algorithm 3-2). Although the VB-EM algorithm optimizes the KLD, they both essentially maximize the log-likelihood. However, fully Bayesian approaches provide a major benefit that is unavailable in the maximum likelihood formulation. Prior distributions allow reliability practitioners to flexibly fuse other information. By setting informative priors, the distribution is effectively regularized. Thus, a practitioner can account for expert opinion or other data sources. The EM algorithm may provide more accurate results on training data, because it is not regularized by prior information. However, EM is not as well equipped as VBEM to accurately model new units that may vary from the training data.

Algorithm 3-1. Expectation-Maximization:

Expectation Step:

- $q_1^{(t+1)}(z) = p(z|X, \theta^{(t)})$
- $Q(\theta) = \langle ln\, p(X, z, \theta) \rangle_{q_1^{(t+1)}(x)}$

Maximization Step:

- $\theta^{(t+1)} = arg\ \max_{\theta}\{Q(\theta)\}$

Convergence

- $\ln p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}|\mathbf{K})^{(t+1)} - \ln p(\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}|\mathbf{K})^{(t)} \leq Threshold$

Algorithm 3-2. Variational Bayesian Expectation-Maximization:

Expectation Step:

- $q_1^{(t+1)}(z) = p(z|X, \varphi^{(t)})$
- $Q(\theta) = \langle ln\, p(X, z, \theta) \rangle_{q_1^{(t+1)}(x)}$

Maximization Step:

- $q_2^{(t+1)}(\theta) = exp[Q(\theta)]$

Convergence

- $L^{(t+1)} - L^{(t)} \leq Threshold$

Using expert opinion and historical databases as priors can prevent new evidence from dramatically altering uncertainty quantification and important reliability metrics. It is possible that new incoming evidence is an outlier that does not represent true statistical properties. Maybe an experiment was poorly conducted, poorly reported, or corrupted. A prior regularizes these predictions to prevent inaccuracies.

For Bayesian methods, one of the more popular approaches for approximating posterior distributions is MCMC [59, 60]. MCMC algorithms like Metropolis Hastings and Gibbs Sampling select candidate parameters from a posterior and evaluate the likelihood [60]. As the random walk progresses, successive parameter draws are correlated and eventually the value should converge to the posterior distribution.

There are various advantages to using VB instead of MCMC. High dimensional interdependent models, like mixture models, become impractical to employ [63]. VB is deterministic and easy to identify convergence [61]. Convergence of MCMC, in these cases, can be notoriously difficult to monitor, but with VB it is easy to determine the updating values of the lower bound [62]. Moreover, MCMC's computational cost rises dramatically in these cases [62]. For highly dimensional data sets or hierarchical models, VB is valuable for its simplicity and efficiency.

VB is not without its flaws. First, VB tends to underestimate the uncertainty of the true posterior [65]. This clearly clashes with the typical conservatism of reliability predictions. This bias implies that predictions may not represent the true posterior as well as another approximation. Second, VB is arduous to derive and implement.

Finally, VB is sensitive to its priors. As such, it is important to be cautious of this selection.

Nonetheless, VB is an active area of research and numerous approaches have been proposed to improve its simplicity and accuracy [57, 59]. Neither MCMC nor VB are inherently scalable for big data applications. However, Stochastic Variational inference employs stochastic optimization by updating parameters on subsamples or batches of the data. An iteratively dependent step-size hyperparameter reduces the volatility of parameter updates [64]. Finally, Black Box Variational Inference (BBVI) reduces the complexity of deriving Variational Inference algorithms. BBVI optimizes a single MCMC approximated gradient with stochastic gradient descent to infer the Variational posterior [66].

3.2.3 Variational Bayes

Variational Bayesian techniques approximate intractable integrals to solve difficult Bayesian inference problems. Unlike Markov Chain Monte Carlo, VB is a deterministic approximation [52]. Given the same prior distribution and data, VB will always converge to the same posterior distribution.

VB minimizes the Kullback-Leibler Divergence (KLD) between the proposed approximation and the analytic posterior. KLD is a metric for the divergence between two different density functions. Thus, the proposed approximation converges as close as possible to the posterior.

$$\ln p(X) \ = \ L(q) \ + \ KL(q||p) \tag{3.4}$$

$$L(q) \ = \ \int q(z,\theta) \ln\left\{\frac{p(X,z,\theta)}{q(z,\theta)}\right\} dz d\theta \tag{3.5}$$

$$KL(q||p) = -\int q(z,\theta) \ln\left\{\frac{p(z,\theta|X)}{q(z,\theta)}\right\} dz d\theta \tag{3.6}$$

In the equations above, X represents the data, z is a set of hidden variables relevant to the data, and θ are random variables that describe X and z. In Equation 3.4, the log marginal probability of our data is decomposed to produce a lower bound approximation $L(q)$ and the KLD of the proposed and actual distribution [52]. Equation 3.5 involves the complete data posterior $p(X,\theta)$ and a proposed distribution over parameters of interest q(z,θ). Equation 3.6 displays the KLD between the proposed posterior $q(z,\theta)$ and analytic posterior $p(z,\theta|X)$.

$$KL(q||p) \ \geq \ 0 \tag{3.7}$$

$$\ln p(x) \geq L(q) \tag{3.8}$$

As Equation 3.7 shows, the KLD must always be positive. Thus, Equation 3.8 must hold and $L(q)$ acts as a lower bound on $\ln p(x)$. The minimization of the KLD or the maximization of the lower bound function would produce a scenario where $q(z,\theta) \ = \ p(z,\theta|X)$. As such, the proposed distribution (q) would converge to the analytic posterior (p). In practice, a convergence threshold should be set to some very small number.

It is important to note that in VB only q is updated. How can an easily solvable q be defined? For simplicity, a factorized approximation is introduced.

$$q(x,\theta) \;=\; q(x)q(\theta) \tag{3.9}$$

This approximation defines the algorithmic aspect of VB. The factorized distributions switch off updating while the other distribution remains constant. This procedure is commonly known as the Variational Bayesian Expectation-Maximization algorithm (VB-EM) described in Algorithm 3-1.

Algorithm 1 displays the steps in VB-EM. It is important to note that $\varphi^{(t)}$ refers to the expected natural parameters. In many of these cases, the introduction of exponential family conjugate priors makes computations easier. Most popular distributions in reliability tend to be a part of the exponential family. Family members include the Poisson distribution, Gaussian distribution, and exponential distribution. The canonical form of the exponential family is displayed in Equation 3.10 and (η) represents the vector of natural parameters.

$$f(x|\theta) = h(x)exp(\eta^T \cdot T(x) \;-\; A(\theta)) \tag{3.10}$$

3.2.4 Variational Bayesian Student-T Mixture Models

Student-T Mixture Models create a mixture of numerous Student-T distributions. EM or VB-EM assigns each data point to the most representative or likely distribution. The algorithm iteratively updates the distribution assignments and the prior parameters of each distribution. Finally, the prior parameters will stop updating and this will correspond to a maximized lower bound. The general form of a mixture model is given in Equation 3.11.

$$p(x|a,\theta,k) = \sum_{k=1}^{K} a_k p_k(x_k|\theta) \tag{3.11}$$

In equation 3.11, the mixture is the sum of k weighted distributions. Each distribution has some mixing proportion a_k that governs the proportion of assigned observations.

For simplicity, a c_n variable can be introduced to model distribution assignment for each observation n. The value c_n now becomes a discrete distribution around the class assignment. By manipulating a joint distribution over x_n and c_n, transformations can be made to Equation 3.11. The result is the likelihood function in Equation 3.12 [55, 78-79]. Please see the appendix for further derivation.

$$P(X,c|a,\theta,k) = \prod_{n=1}^{N}\prod_{k=1}^{K} p(c_n = k)p(x_n|\theta_k) \tag{3.12}$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} a_k p(x_n | \theta_k) \tag{3.13}$$

The Bayesian formulation to solve for these parameters would allow for prior distributions to regularize predictions. As such, expert opinion, field data, or experimental data could affect the output. The maximum a posteriori (MAP) optimization equation derives point estimates of the parameters from the mode of the posterior distribution. The MAP for SMM is as follows [55]:

$$(\hat{a}, \hat{\theta}) = \arg\max_{(a,\theta)} p(X, c | a, \theta, k) p(a|k) p(\theta|k) \tag{3.14}$$

Variational Bayes can then be leveraged to optimize Equation 3.14.

Priors are set over the relevant parameters governing the hidden and observed random variables. k is a hyper parameter for the number of sub-populations. Setting this k value is a model selection problem. Grid search, random search, and manual search are all possible techniques to solve this problem [56]. In this paper, the author solved for k with a grid search optimization procedure over the Bayesian Information Criterion (BIC) [52]. Grid search simply explores a regularly spaced grid of values over some range to discover the optimal solution.

The directed acyclic graph (Figure 3-1) displays the structure of the hierarchical Bayesian Student-T Mixture Model.

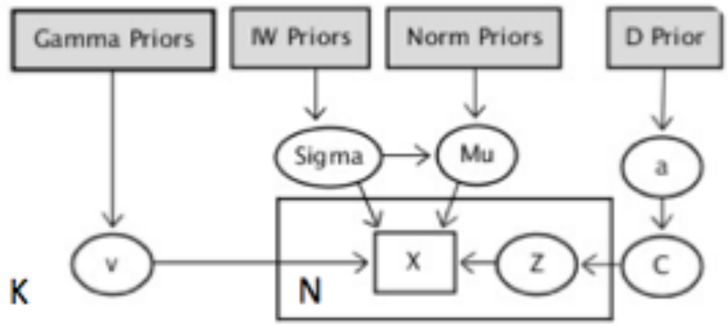

Figure 3-1: Variational Bayesian Student-T Mixture Model

The first step is to re-represent the Student-T PDF into its infinite Gaussian scaled model form [55, 57, 77-79].

$$S(x|\boldsymbol{\mu}, \boldsymbol{\Sigma}, v) = \int_0^\infty N(x|\boldsymbol{\mu}, z^{-1}\boldsymbol{\Sigma})\mathrm{G}(\mathrm{z}|\frac{v}{2}, \frac{v}{2}) \tag{3.15}$$

The Student-T is now represented as a marginal distribution composed of a Multivariate Gaussian and Gamma distributions. In this case, the simplified Gamma distribution only uses one parameter v for both the shape and scale. Using Equations 3.11-3.13, the marginal likelihood can be written as:

$$p(\boldsymbol{X}, \boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}|\mathrm{K}) = \prod\nolimits_{\boldsymbol{n}} \prod\nolimits_{\boldsymbol{k}} a_k \mathrm{N}(\mathrm{x}_n|\boldsymbol{\mu_k}, z_{n,k}^{-1}\boldsymbol{\Sigma_k})\mathrm{G}(\mathrm{z}_{n,k}|\frac{v_k}{2}, \frac{v_k}{2})p(\boldsymbol{\theta}_k) \tag{3.16}$$

where $\boldsymbol{\Theta}$ is a vector of parameters for all distributions and $\boldsymbol{\theta}_k$ is the vector for each component distribution [56]. Finally, the posterior distribution can be written as:

$$p_k(\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}|\mathbf{X}) = \frac{p(\boldsymbol{X}, \boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}|\mathbf{K})}{p(\boldsymbol{X})} \tag{3.17}$$

In order to implement a VB approximation for this otherwise intractable problem, it is necessary to factorize the posterior [56].

$$q(\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}) = q(\boldsymbol{c}, \boldsymbol{Z})q(\boldsymbol{\Theta}) \tag{3.18}$$

$$= \prod_n \prod_k [q(c_n = k|z_{n,k})q(z_{n,k})] \prod_k [q(v_k)q(\boldsymbol{\mu_k}|\boldsymbol{\Sigma_k})q(\boldsymbol{\Sigma_k})] \, \mathrm{q}(\mathbf{a}) \tag{3.19}$$

In this factorized form the VB-EM procedure can be applied. First, the expected values of each distribution's parameters are taken. The distribution assignments are updated. This corresponds to the expectation step. Next, the maximization step updates all other parameters while holding the distribution assignments constant. VB iteratively updates all parameters of interest until the lower bound is maximized.

Since Equation 3.5 is essentially the expectation taken with respect to a function, the following equalities hold:

$$L(q(\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta})) = \langle \ln p(\boldsymbol{X}, \boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}|\mathbf{K}) \rangle_{q(\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta})} \tag{3.20}$$

$$= \prod_k \prod_n \langle \ln p(\boldsymbol{x_n}, c_n, z_{nk}, \boldsymbol{\theta_k}) \rangle_{q(c_n = k|z_k)q(z_{nk})}$$

$$+ \prod_k \langle \ln p(\boldsymbol{x_n}, c_n, z_{nk}, \boldsymbol{\theta_k}) \rangle_{q(\boldsymbol{\theta_k})} \tag{3.21}$$

Thus, maximizing the lower bound is equivalent to assessing whether the factorized distributions have converged. At this point, the KLD is minimized and the lower bound is maximized [52, 56]. This procedure essentially maximizes the log-likelihood $\ln p(\boldsymbol{X}, \boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta}|\mathbf{K})$ with respect to the simplified function $q(\boldsymbol{c}, \boldsymbol{Z}, \boldsymbol{\Theta})$.

In practice, deriving the optimization procedure can be quite arduous. As such, it is advantageous for a reliability analyst to introduce exponential family conjugate priors to easily update posterior parameters. The following priors were introduced in the author's experiments: (The header (˜) signifies a hyperparameter)

$$q(\boldsymbol{a}) = D(\boldsymbol{a}|\widetilde{\boldsymbol{k}}) \tag{3.22}$$

$$q(z_k) = G(z_k|\tilde{v}_k, \tilde{v}_k) \tag{3.23}$$

$$q(\boldsymbol{\mu_k}|\boldsymbol{\Sigma_k}) = N(\boldsymbol{\mu_k}|\widetilde{\boldsymbol{\mu}}, \tilde{\eta}^{-1}\boldsymbol{\Sigma_k}) \tag{3.24}$$

$$q(\boldsymbol{\Sigma_k}) = IW(\boldsymbol{\Sigma_k}|\tilde{\gamma}, \tilde{\gamma}\tilde{\Sigma}) \tag{3.25}$$

With conjugate priors, updating the parameters becomes a far easier task. For example, the Dirichlet prior can evoke a categorical likelihood function to update $\widetilde{\boldsymbol{k}}_{\boldsymbol{k}}$ with the following equation:

$$\boldsymbol{k} = \widetilde{\boldsymbol{k}} + n_k \tag{3.26}$$

Since the amount of observations assigned to a distribution can change, the priors will update accordingly. Finding all the conjugate pairs and expectations is a largely trivial task that can be found in Probability Distributions Used in Reliability Engineering [58].

3.3 Approach Validation

The author previously validated the effectiveness of clustering failure times or related covariates for reliability predictions. Many of these approaches struggle to cluster real world noise. SMM, on the other hand, can cluster data sets with uncertain boundaries and outliers. To validate the proposed SMM based approach, the author examined whether a GMM could effectively serve as a proxy for SMM.

The author synthesized a simple two-dimensional data set of five clusters with 133 observations in each cluster. The data was Student-T distributed and each distribution used the identity matrix as its covariance parameter. In the context of the proposed methodology in Figure 2-1, the data could correspond to two covariates. Figure 3-1 displays the scatter plot of the synthetic data.

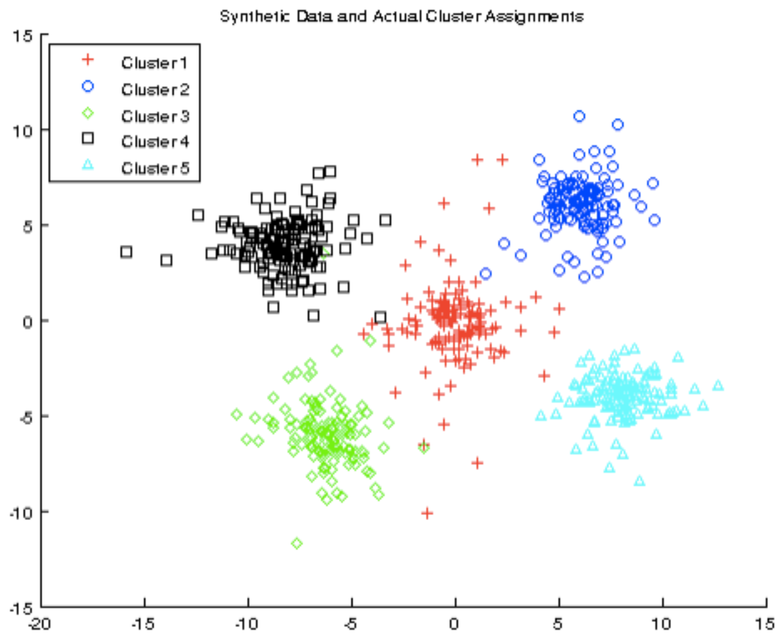

Figure 3-2: Scatter Plot of Synthetic Student-T Distributed Data

In this case, there is no need to conduct a hyper parameter search over the amount of clusters. The author assessed the accuracy and likelihood of the models by comparing the cluster assignment with the distribution from which the data was synthesized. The results displayed in Table 3-1 show that SMM is notably more accurate than GMM on this data set. Table 3-2 compares the predicted parameters to actual parameters.

SMM is more accurate and produces better parameter estimates. Thus, SMM is better able to find the true sub-populations. The visualizations in Figures 3-2 and 3-3 shed light on the reason.

Table 3-1: Mixture Model Accuracy

	VB	EM
GMM	61.5%	72.5%
SMM	97.7%	98.4%

Table 3-2: Mixture Model Parameter Estimates

	π: Mixing Proportion	v: Degrees of Freedom	Σ: Covariance Matrix	μ: Mean
Actual	Cluster 1: 0.20 Cluster 2: 0.20 Cluster 3: 0.20 Cluster 4: 0.20 Cluster 5: 0.20	Cluster 1: 2 Cluster 2: 4 Cluster 3: 5 Cluster 4: 3 Cluster 5: 6	All Clusters: $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	Cluster 1: [0, 0] Cluster 2: [6, 6] Cluster 3: [-6,-6] Cluster 4: [-8,4] Cluster 5: [8,-4]
Estimated - GMM EM	Cluster 1: 0.16 Cluster 2: 0.20 Cluster 3: 0.13 Cluster 4: 0.18 Cluster 5: 0.32	(N/A)	Cluster 1: $\begin{pmatrix} 17.9 & -7.6 \\ -7.6 & 5.4 \end{pmatrix}$ Cluster 2: $\begin{pmatrix} 0.98 & -0.13 \\ -0.13 & 0.78 \end{pmatrix}$ Cluster 3: $\begin{pmatrix} 1.6 & -0.6 \\ -0.6 & 1.8 \end{pmatrix}$ Cluster 4: $\begin{pmatrix} 13 & 12.5 \\ 12.5 & 21 \end{pmatrix}$ Cluster 5: $\begin{pmatrix} 2.11 & 0.04 \\ 0.04 & 1.24 \end{pmatrix}$	Cluster 1: [1.4, .6] Cluster 2: [6.1, 6.2] Cluster 3: [-6.2, -6.1] Cluster 4: [-5.2, 2.4] Cluster 5: [8.1, -4.1]
Estimated - GMM VB	Cluster 1: 0.20 Cluster 2: 0.20 Cluster 3: 0.21 Cluster 4: 0.36 Cluster 5: 0.03	(N/A)	Cluster 1: $\begin{pmatrix} 11701 & -5211 \\ -5211 & 2538 \end{pmatrix}$ Cluster 2: $\begin{pmatrix} 207.7 & 44.3 \\ 44.3 & 264.63 \end{pmatrix}$ Cluster 3: $\begin{pmatrix} 267.2 & -26.2 \\ -26.2 & 310.9 \end{pmatrix}$ Cluster 4: $\begin{pmatrix} 6000 & -2953 \\ -2953 & 2135 \end{pmatrix}$ Cluster 5: $\begin{pmatrix} 21.1 & 75.4 \\ 75.4 & 415.1 \end{pmatrix}$	Cluster 1: [0, -0.5] Cluster 2: [6.1, 6.0] Cluster 3: [-6.2, -6.1] Cluster 4: [-2.1, 0.9] Cluster 5: [1.2, -0.7]
Estimated - SMM EM	Cluster 1: 0.20 Cluster 2: 0.20 Cluster 3: 0.20 Cluster 4: 0.20 Cluster 5: 0.20	Cluster 1: 2.8 Cluster 2: 3.8 Cluster 3: 3.6 Cluster 4: 5.8 Cluster 5: 6.1	Cluster 1: $\begin{pmatrix} 1.37 & -0.06 \\ -0.06 & 1.16 \end{pmatrix}$ Cluster 2: $\begin{pmatrix} 0.93 & -0.06 \\ -0.06 & 0.91 \end{pmatrix}$ Cluster 3: $\begin{pmatrix} 1.2 & -0.39 \\ -0.39 & 1.28 \end{pmatrix}$ Cluster 4: $\begin{pmatrix} 1.82 & 0.03 \\ 0.03 & 1.2 \end{pmatrix}$ Cluster 5: $\begin{pmatrix} 1.66 & -0.03 \\ -0.03 & 0.83 \end{pmatrix}$	Cluster 1: [0,-0.1] Cluster 2: [6.0, 6.1] Cluster 3: [-6.1, -6.1] Cluster 4: [-8.2, 3.9] Cluster 5: [8.1, -4.0]
Estimated - SMM VB	Cluster 1: 0.19 Cluster 2: 0.20 Cluster 3: 0.20 Cluster 4: 0.20 Cluster 5: 0.21	Cluster 1: 127.0 Cluster 2: 140.8 Cluster 3: 140.8 Cluster 4: 140.0 Cluster 5: 141.5	All Clusters: *** $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$	Cluster 1: [-0.1, -0.1] Cluster 2: [6.0, 6.0] Cluster 3: [-6.0, -6.0] Cluster 4: [-8.0, 3.8] Cluster 5: [8.0, -3.9]
*** Covariance round to 0, where it is compensated by a large v.				

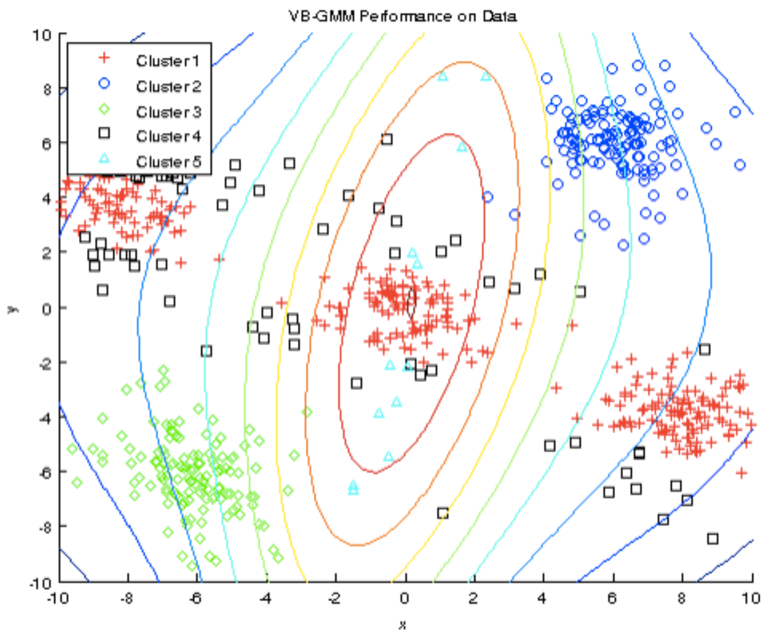

Figure 3-3: VB-EM Gaussian Mixture Model. Contours represent Covariance.

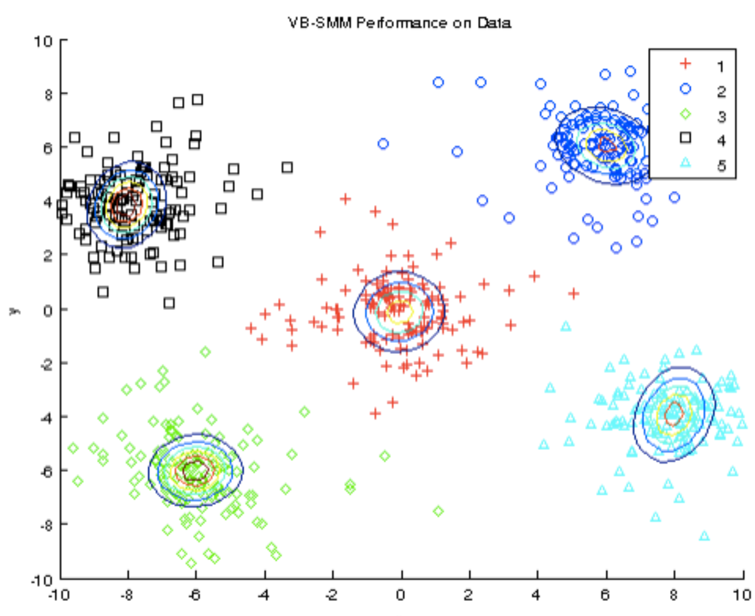

Figure 3-4: VB-EM Student-T Mixture Model. Contours represent Covariance.

The level of outliers begins to cause some observations across clusters to intersect. Without the degrees of freedom parameter to increase the size of the tails, a GMM attempts to overcompensate by increasing the size of the covariance. A single components' covariance then intrudes onto other components, which will produce uncertainties in cluster assignments and reduce accuracy. This phenomenon is visualized by the large contours in Figure 3-4.

For this experiment, the EM algorithm is more accurate than VB-EM. However, the EM algorithm is more sensitive to training data. The prior regularization for VB-EM reduces its sensitivity to training data. VB-EM's reduced bias is an advantage in the long run. Additional units from this hypothetical fleet may not resemble the training data. If data from additional experiments vary from the training set, VB-EM could more accurately identify sub-populations than EM.

Table 3-3: VB-EM Information Criterion and Likelihood Scores

EM Performance	AIC	BIC	Log-Likelihood
GMM	7040	7170	-3491
SMM	6740	6776	-3362

According to Table 3-3, SMM outperforms the GMM in Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores. The information criterion scores control for over fitting from the additional v parameter by introducing a penalty.

This synthetic experiment provides evidence that SMM has higher accuracy than GMM when covariate observations overlap. Many real world applications have this property. An example application on a real data set is discussed in the following section.

3.4 Application to Power Plant Failures

In this section, the methodology is applied to a real data set. The PHM Data Challenge Competition 2015 (PHM 2015) is the same data set used in Chapter 2.3.1.

First, the author evaluated SMM performance against the baseline GMM by applying the model directly on failure events. This corresponds to cases where covariate information is inaccessible or irrelevant. The author used a grid search over the Kullback-Leibler Divergence or AIC to identify the optimal quantity of component distributions before evaluating the performance of each algorithm [56].

It is important to note that the presented problem is an unlabeled and unsupervised pattern recognition problem, thus it is difficult to use accuracy as a validation and testing metric to evaluate the performance. After fitting each distribution, the author used the mean Silhouette Coefficient to compare the quality of the clustering. The Silhouette Coefficient is a validation metric that evaluates the similarity of objects within a cluster [75]. Since the PHM data is unlabeled; the Silhouette Coefficient can serve as a proxy for the self-consistency of the clustering. The Silhouette Coefficient is displayed in equation 3.27.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{3.27}$$

The numerator $b(i) - a(i)$ represents the difference between an observations' mean intra-cluster distance and nearest non-assigned cluster distance. Figure 3-4 displays a fitted single Gaussian distribution to the entire data. The denominator $max\{a(i), b(i)\}$ regularizes this difference, such that a Silhouette Coefficient score close to 1 indicates self-consistent clustering. Conversely, a Silhouette Coefficient score close to -1 indicates incongruous clustering. The large over and under-estimation of various failure intervals provides evidence for the value of mixture modeling. Recall that in some circumstances sensor-based covariates are unavailable. GMM and SMM can be applied directly to time-to-failure data. GMM and SMM clustered failure times are displayed graphically in Figures 3-5 and 3-6. The mean Silhouette Coefficient is tabulated in Table 3-4.

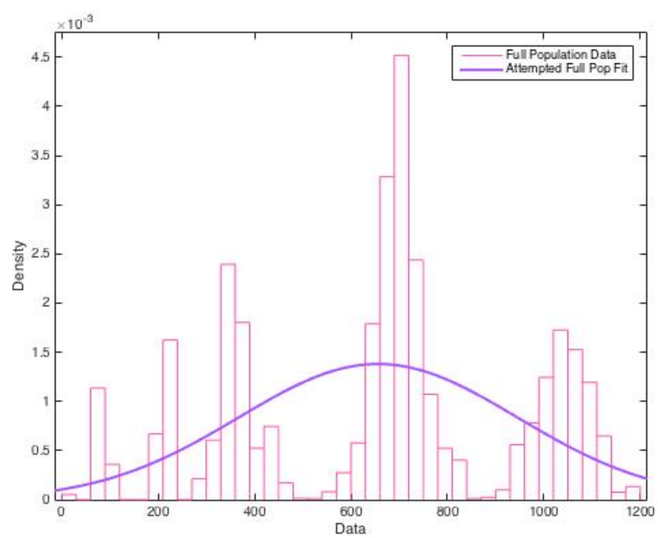

Figure 3-5: Gaussian Fit to all Failure Times

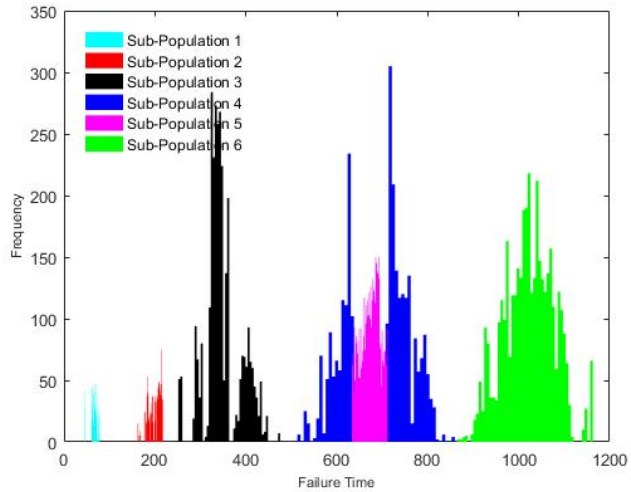

Figure 3-6: Histogram of Failure Times (Days) Segmented with VBGMM. Only Six Distributions Visualized, 10 Distributions (k) Total.

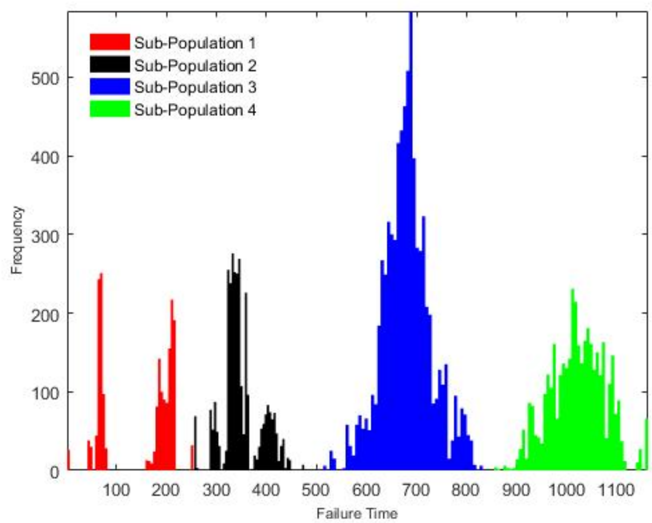

Figure 3-7: Histogram of Failure Times (Days) Segmented with VBSMM. All Four Distributions (k) Visualized.

Table 3-4: Mean Silhouette Coefficient for GMM and SMM Applied to Failure Time

	VB	EM
GMM	.2370	.5531
SMM	.9090	.9159

Table 3-5: Log-Likelihood, AIC, and BIC Scores

Distribution	Log-Likelihood	AIC	BIC
Gaussian	-115,302	230,608	230,633
EMGMM	-78,757	157,574	157,804
EMSMM	-76,033	152,090	152,182

The results provide evidence that SMM finds more self-consistent & distinct clusters than GMM in real world cases. The Silhouette Coefficient score shows that SMM assigned clusters are more similar numerically than GMM assigned clusters. Also, GMM uses more distributions to model the same population, which could result in over fitting.

Table 3-5 shows that clustering can improve the likelihood and relative quality of a distribution fit to heterogeneous data. It also provides evidence that SMM performs better than GMM in this scenario.

In Figure 3-7, the GMM struggles to segment sub-population 4. This heavy tailed data produces uncertainty in the observations' cluster assignment and the results are undesirable.

The author maps clusters of covariates to failure events to explore the application of SMM on the methodology proposed in Figure 2-1. This allows an analyst to assign labels that can be used to train a classification algorithm, so inference can be made on units that have not failed.

The covariates in the PHM data are numerous and this high dimensionality prevents direct analysis on the covariates. In particular, high dimensionality renders the metric distance of each observation to be essentially equal. Thus, similarity related analysis become difficult. Sparse PCA (Equation 3.28) was implemented to decompose the data to only a single dimension [76]. The author conducted a manual model search of dimensionality reduction algorithms, but Sparse PCA produced the best results for both GMM and SMM.

$$(U^*, V^*) = arg \min_{U,V} \frac{1}{2} \|X - UV\|_2^2 + \alpha \|V\|_1 \tag{3.28}$$

The covariates are clustered with GMM or SMM and assignments mapped onto the failure times. GMM performs better than when applied directly to failures times, which may signal smaller tails in the covariate data. Nonetheless, SMM continues to outperform GMM. The author used output parameters of the EM predictions as priors for VB.

Table 3-6: Mean Silhouette Coefficient for GMM and SMM Applied to Covariates

	VB	EM
GMM	.8595	.7422
SMM	.8654	.8812

Table 3-5 provides evidence that SMM on average groups more similar data than GMM. Although VBGMM performs only slightly worse than VBSMM, it uses more than twice as many free parameters. For GMM, the optimal quantity of distributions (k) was 8. Figures 3-7 and 3-8 visualize the two VB models’ histograms.

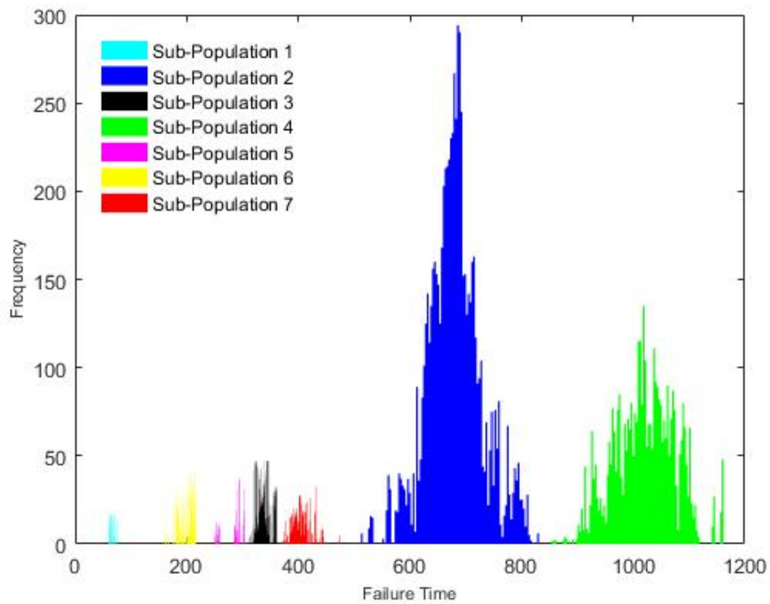

Figure 3-8: Histogram of Clustering Methodology with VBGMM. Only seven largest distributions (n>100) visualized, 8 distributions (k) total.

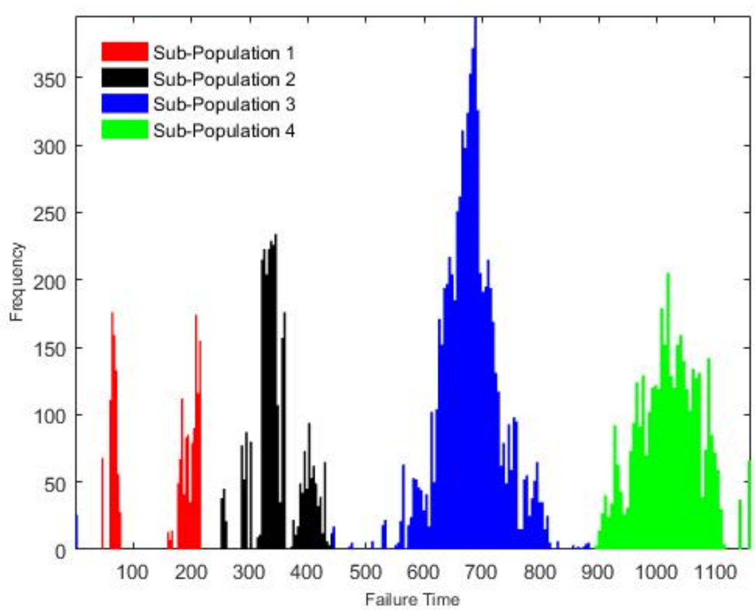

Figure 3-9: Histogram of Clustering Methodology with VBSMM. All four distributions (k) Visualized.

If SMM is applied directly to the failure times, it is possible to conduct a holistic failure estimate for the population. However, the presented methodology clustered covariates and not failures to assign class labels. So, the distributions parameterization is not directly relevant to the failure events. Nonetheless, a separate analysis can be conducted on each sub-population. By analyzing each sub-population separately, an analyst has reduced fleet uncertainty. Variation in manufacturing, operational, and environmental conditions is theoretically less within the sub-population. The author refits each sub-population with a Student-T Distribution and evaluated traditional reliability metrics in Table 3-7.

Table 3-7: Table of Various Relevant Reliability Measures from VBSMM

	MTTF (Days)	**Reliability at the Full Population MTTF (656 Days)**
Full Population	656	50%
Sub-Population 1	146	~0%
Sub-Population 2	345	~0%
Sub-Population 3	679	71%
Sub-Population 4	1022	~100%

Table 3-7 shows dramatic variability in reliability metrics across the entire population. The fleet-wide uncertainty would over or underestimate the MTTF of each sub-population. Thus, the presented methodology improves the predictive capabilities of heterogeneous fleets.

Last, the advantage of using SMM on covariates is to assign units that have not failed to a sub-population. A reliability analyst can leverage the cluster assignments made by the SMM to train a supervised classification algorithm to conduct this task. As an example, the author trained a non-parametric Decision Tree (DT) classifier on the VBSMM cluster assignments [77]. A grid search and 5-fold cross validation optimized the tree length hyperparameter. The DT obtained a validation and test accuracy of **100%** with only 5 branches.

3.5 Chapter Summary

In Chapter 2, the author recommended segmenting a fleet into sub-fleets as a practical approach for efficient resource allocation and maintenance scheduling. However, in real world data, the lines between sub-populations are blurred. In these cases, Gaussian Mixture Models can overestimate the covariance and produce unwanted results. Student-T Mixture Modeling is an attractive heavy-tailed alternative for these real world cases.

In this chapter, the author explored the Variational Bayesian formulation of Student-T Mixture Models, which has several practical advantages over MCMC and EM in solving highly dimensional hierarchical Bayesian inference problems. Recent advances in VB have improved its scalability and accuracy. Finally, the VBEM algorithm, rather than the EM formulation, allows an analyst to use historical databases or expert opinion as a prior to regularize predictions. This regularization is particularly valuable if the new evidence is small in quantity or consists of outliers.

The author validated SMM on both synthetic data and real world power plant failures. In both cases, SMM outperformed the predictive capabilities of a single Gaussian distribution and GMM.

Chapter 4: Conclusion

Fleet uncertainty is a big challenge facing reliability practitioners. Predictions on entire fleets can over or underestimate relevant probabilities for a particular unit. Many data driven PHM approaches reduce uncertainty in a case-by-case basis. Despite these advantages, it is impractical for large fleets and difficult to interpret for resource allocation. This thesis addresses this challenge by identifying similar sub-populations within a fleet and conducting independent analyses on these groups.

In Chapter 2, the author introduced a methodology for clustering fleets to identify and leverage sub-populations for improved reliability and integrity measures. The methodology clusters sensor-based covariate data and derives predictive analytics on each sub-population. Additional units can be assigned to a sub-population by training a supervised classification algorithm with the cluster assignments as labels. The author demonstrated this methodology with DBSCAN clustering. The methodology improved reliability predictions and reduced uncertainty around important integrity metrics (*i.e.* MTTF, B10 life). However, the methodology struggled to segment units that were outliers or could belong to two separate sub-populations. This real world noise limited the predictive capabilities of the methodology.

In Chapter 3, the author introduced Student-T Mixture Models to segment fleets despite real world noise. Using parameterized tails, SMM can more accurately model inter-cluster boundary points without increasing its variance. The author motivated a fully Bayesian formulation of SMM, which allows for the inclusion of prior information like expert opinion or a reliability database. Variational Bayes is an efficient and reliable approximation for solving complicated hierarchical Bayesian problems like the SMM. The author demonstrated the application of SMM to synthetic data and power plant failures. SMM produced a more accurate model with self-consistent sub-populations.

4.2 Research Contributions

The research contributions of this thesis are:

- Introduce a methodology for fleet uncertainty reduction that accurately identifies sub-populations, improves time-to-failure distribution fit, and increases precision of reliability metrics.
- Motivate Student-T Mixture Models to identify sub-populations despite real world noise, outliers, and unclear sub-population boundaries.
- Adapt Variational Inference to quickly and efficiently approximate Student-T Mixture Models.
- Discuss Variational Bayes applications in Reliability Engineering.

4.3 Suggested Future Research

Although the methodology accomplished all its stated goals, there are still notable limitations. First, the author did not explore how other unit-by-unit uncertainty reduction methods, like Particle Filtering, work on segmented fleets. Second, Variational Bayesian approximation infers parameters by reanalyzing all the data iteratively. This limits the applications of the methodology to large quickly streaming sensor data. Third, the collected covariates are arbitrary. Informing covariate selection with physics of failure could improve the effectiveness of this methodology. Fourth, SMM may not be the most representative distribution of a fleet's time-to-failure data. For example Student-t is a symmetric distribution, which may not be representative of population's failure times. If there were no available covariates, Variational Bayesian Mixtures of other standard reliability distributions would be worth exploring.

Fifth, the methodology proposed in this thesis could be generalized and applied to domains other than pipeline and power plant failure events. However, fleet uncertainty modeling may be less effective if a fleet is composed of only a few units. The lack of data may produce unrepresentative or over fit sub-populations. In these cases, the information provided by the prior distribution becomes critical to the clustering performance.

Finally, there is a worthy philosophical discussion surrounding the importance of tails to model sub-populations that should be addressed. SMM encompasses extreme events within its tails to prevent unnecessarily increasing the covariance matrices. For the purposes of this paper, leveraging learnable parameters that model these extreme events was practical. However, outliers are often excluded as the possible product of experimental or rare event error. Could the inclusion of outliers within a sub-population reduce the quality of fleet modeling? This is an important discussion that should be accompanied by experimentation.

To address limitations, the author has suggested future research:

1) Advance the Methodology
   a) The proposed methodology is not intended to replace data driven unit-by-unit techniques like Particle Filtering or Dynamic Bayesian Networks. Rather, it should work in conjunction with these techniques. Employing these techniques in the contexts of the methodology could improve scheduled maintenance and condition based maintenance by reducing fleet-wide uncertainty and unit-by-unit uncertainty.
   b) Currently, the proposed methodology is not scalable for online application. Streaming sensor data can produce large vectors of information. Stochastic Black Box Variational Inference overcomes this challenge by updating variational parameters with only a subsample of the data. [80-81] It also reduces the manual workload of an analyst by eliminating the need to derive update equations.
   c) Collected covariates may not be related to degradation or failure rate uncertainty within a fleet. The methodology should leverage Physics of Failure (PoF) research to identify features that are most likely to contribute to fleet uncertainty. A more relevant set of covariates could improve sub-population identification. Collecting relative humidity and temperature, for

example, could improve the population segmentation and subsequent predictions of corrosion related failures in oil pipelines.

2) Variational Bayesian Mixture Models

    a) The presented variational formulation of the Student-T Mixture Model did not introduce prior parameters over the degrees of freedom (v) parameter. By introducing a prior over v, an analyst can regularize this estimate with reliability field data or expert opinion.

    b) Sensor based covariates are not always available. In this case, other Variational Bayesian mixtures could perform better for time-to-failure analysis. One example is a Variational Bayesian formulation of Weibull Mixture Models.

3) Bayesian Approximation

    a) Both VB and MCMC approximation are limited. VB is biased and MCMC can be slow. Various other approximations to intractable Bayesian inference problems exist. There may be particular circumstances in reliability analysis that these other approximations are more practical, accurate, or efficient. Laplace Approximation and Expectation Propagation are two examples of other approximations. [82-83]

Appendix

Additional Derivation of Equation 3.12:

By introducing $c_n = k$ as a binary variable that identifies assigned distribution, the likelihoods can be segmented into separate likelihoods for each distribution.

$$p(\boldsymbol{c}_n) = \prod_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{K}} a_{\boldsymbol{k}}^{\boldsymbol{c}_{\boldsymbol{nk}}} \qquad (A1)$$

$$\boldsymbol{p}(\boldsymbol{x}_{\boldsymbol{n}}) = \prod_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{K}} \boldsymbol{N}(\boldsymbol{x}_{\boldsymbol{n}} | \boldsymbol{\mu}_{\boldsymbol{k}}, \boldsymbol{z}_{\boldsymbol{nk}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{k}})^{\boldsymbol{c}_{\boldsymbol{nk}}} \qquad (\boldsymbol{A2})$$

From here, it is easy to see Equation 3.12.

Experiment on Synthetic Signals:

5D PCA of Signal Data (N=1026, D=545 -> 5, K=3, $\boldsymbol{a}_{\boldsymbol{1},2,\boldsymbol{3}}$= 0.333).

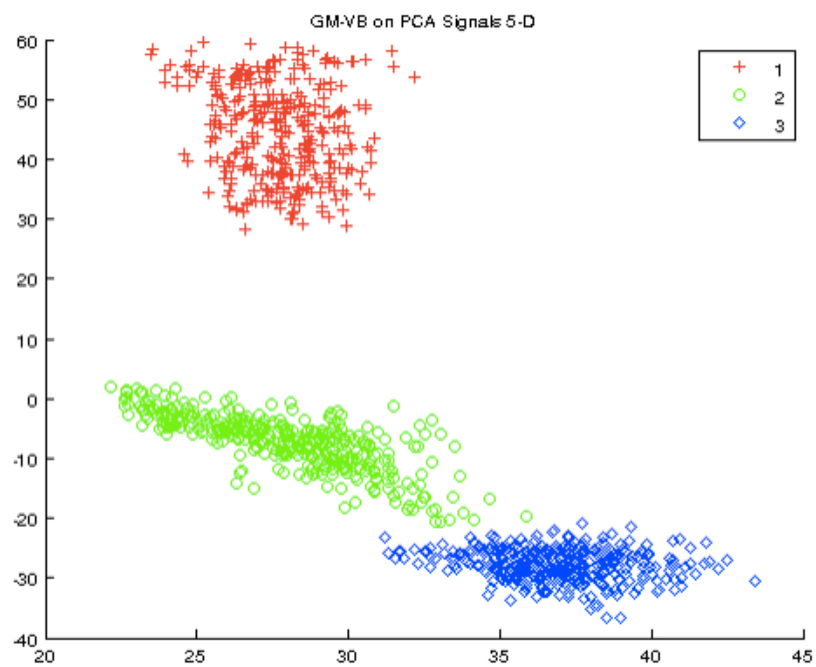

Figure A1: Variational Bayesian Gaussian Mixture Models on Signal Data

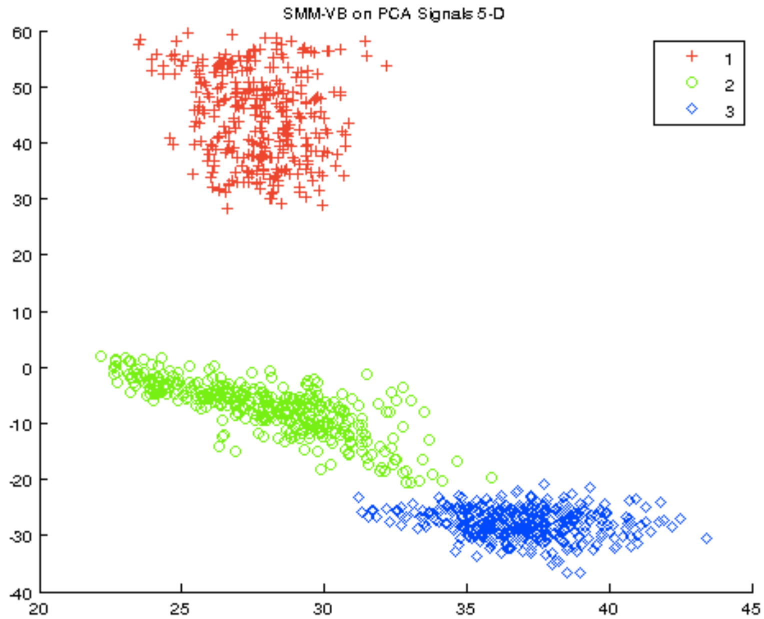

Figure A2: Variational Bayesian Student-T Mixture Models on Signal Data

Table A1: Variational Bayesian Mixture Models Accuracy on Signal Data

	VB	EM
GMM	1.000	0.691
SMM	1.000	1.000

Table A2: Actual and Estimated Parameters for Signal Data

	a: Mixing Proportion	v: Degrees of Freedom	μ: Mean
Actual	Cluster 1: 0.33 Cluster 2: 0.33 Cluster 3: 0.33	N/A	N/A
Estimated - GMM EM	Cluster 1: 0.28 Cluster 2: 0.39 Cluster 3: 0.33	(N/A)	Cluster 1: [27.52, 19.36, 15.50, 15.61, -25.86] Cluster 2: [28.05, 18.37, 19.57, -9.5, 18.21] Cluster 3: [36.88, -27.79, -26.95, .10, -.91]
Estimated - GMM VB	Cluster 1: 0.33 Cluster 2: 0.33 Cluster 3: 0.33	(N/A)	Cluster 1: [36.85, -27.65, -26.82, 0.14, -.99] Cluster 2: [27.69, 44.76, -9.01, -0.26, -0.53] Cluster 3: [27.97, -7.2, 44.8, 1.95, 0.624]
Estimated - SMM EM	Cluster 1: 0.33 Cluster 2: 0.33 Cluster 3: 0.33	Cluster 1: 3.2 Cluster 2: 17.4 Cluster 3: 3.4	Cluster 1: [27.58, -6.76, 45.75, 2.86, .66] Cluster 2: [27.68, 44.60, -8.85, .55, -.38] Cluster 3: [36.75, -27.78, -26.95, 1.17, 1.51]
Estimated - SMM VB	Cluster 1: 0.33 Cluster 2: 0.33 Cluster 3: 0.33	Cluster 1: 359.1 Cluster 2: 359.6 Cluster 3: 359.4	Cluster 1: [22.36, -5.33, 37.04, 5.42, -0.02] Cluster 2: [22.54, 36.13, -7.0, .69, -.95] Cluster 3: [29.94, -22.58, -21.95, 1.05, 0.27]

References

Chapter 1

[1] Tsui, Kwok L., Chen, Nan, Zhou, Qiang, Hai, Yizhen, Wang, Wenbin. "Prognostics and health management: A review on data driven approaches." Mathematical Problems in Engineering 2015 (2015).
[2] Van Trees, Harry L., and Kristine L. Bell. "A Tutorial on Particle Filters for Online Nonlinear/NonGaussian Bayesian Tracking." (2009): 723-737.
[3] Murphy, Kevin Patrick. Dynamic bayesian networks: representation, inference and learning. Diss. University of California, Berkeley, 2002.
[4] Schumann, Johann M., Rozier, Kristin Y., Reinbacher, Thomas, Mengshoe, Ole J, Mbaya, Timmy, Ippolito, Corey. "Towards real-time, on-board, hardware-supported sensor and software health management for unmanned aerial systems." International Journal of Prognostics and Health Management 6 (2015).
[5] Rabiei, Elaheh, Enrique Lopez Droguett, and Mohammad Modarres. "A prognostics approach based on the evolution of damage precursors using dynamic Bayesian networks." Advances in Mechanical Engineering 8.9 (2016): 1687814016666747.
[6] Meeker, William Q., and Yili Hong. "Reliability Meets Big Data: Opportunities and Challenges." Quality Engineering 26.1 (2013): 102-16.
[7] Weiss, Volker, and Anindya Ghoshal. "On the search for optimal damage precursors." Structural Health Monitoring 13.6 (2014): 601-608.
[8] Sun, Bo, Zeng, Shengkui, Kang, Rui, Pecht, Michael G. "Benefits and challenges of system prognostics." IEEE Transactions on reliability 61.2 (2012): 323-335.
[9] Kaplan, Stan. "On a 'Two-Stage' Bayesian Procedure for Determining Failure Rates from Experimental Data." IEEE Power Engineering Review IEEE Power Eng. Rev. PER-3.1 (1983): 43.
[10] Droguett, Enrique Lopez, Frank Groen, and Ali Mosleh. "The combined use of data and expert estimates in population variability analysis." Reliability Engineering & System Safety 83.3 (2004): 311-321.
[11] Liu, Jie, and Zio, Enrico. "Framework for Asset Prognostics from Fleet Data." Proc. Annual. Conference. Prognostic Health Management Society. 2016.
[12] Tian, J. "Better Reliability Assessment and Prediction through Data Clustering." IEEE Transactions on Software Engineering IIEEE Trans. Software Eng. 28.10 (2002): 997-1007.
[13] Dindarloo, Saeid R., and Elnaz Siami-Irdemoosa. "Data Mining in Mining Engineering: Results of Classification and Clustering of Shovels Failures Data." International Journal of Mining, Reclamation and Environment (2016): 1-14.
[14] Arunajadai, Srikesh G., Uder, Scott J., Stone, Robert B., Tumer, Irem Y. "Failure Mode Identification through Clustering Analysis." Qual. Reliab. Engng. Int. Quality and Reliability Engineering International 20.5 (2004): 511-26.
[15] Gower, J. C. "A General Coefficient of Similarity and Some of Its Properties." Biometrics 27.4 (1971): 857.

[16] Zhao, Ke, and Steffey, Duane. "Practical Applications of Mixture Models to Complex Time-to-failure Data." 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS) (2013)

Chapter 2

[17] Meeker, William Q., and Yili Hong. "Reliability Meets Big Data: Opportunities and Challenges." Quality Engineering 26.1 (2013): 102-16.
[18] Meeker, William Q. "Limited Failure Population Life Tests: Application to Integrated Circuit Reliability." Technometrics 29.1 (1987): 51.
[19] Zhao, Ke, and D. Steffey. "Practical Applications of Mixture Models to Complex Time-to-failure Data." 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS) (2013):
[20] Hastie, Trevor J., Robert John Tibshirani, and Jerome H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer, 2016.
[21] "Estimation of the Parameters of a Mixture Weibull Model for Analyze Cure Rate." Ams Applied Mathematical Sciences (2013).
[22] Li, Mingyang, and Jian Liu. "Bayesian Hazard Modeling Based on Lifetime Data with Latent Heterogeneity." Reliability Engineering & System Safety 145 (2016): 183-89.
[23] Kaplan, Stan. "On a ‘Two-Stage' Bayesian Procedure for Determining Failure Rates from Experimental Data." IEEE Power Engineering Review IEEE Power Eng. Rev. PER-3.1 (1983): 43.
[24] Droguett, Enrique Lopez, Frank Groen, and Ali Mosleh. "The combined use of data and expert estimates in population variability analysis." Reliability Engineering & System Safety 83.3 (2004): 311-321.
[25] Hastie, Trevor J., Robert John Tibshirani, and Jerome H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York, NY: Springer, 2016.
[26] Tian, J. "Better Reliability Assessment and Prediction through Data Clustering." IEEE Transactions on Software Engineering IIEEE Trans. Software Eng. 28.10 (2002): 997-1007.
[27] Dindarloo, Saeid R., and Elnaz Siami-Irdemoosa. "Data Mining in Mining Engineering: Results of Classification and Clustering of Shovels Failures Data." International Journal of Mining, Reclamation and Environment (2016): 1-14.
[28] Arunajadai, Srikesh G., Uder, Scott J., Stone, Robert B., Tumer, Irem Y. "Failure Mode Identification through Clustering Analysis." Qual. Reliab. Engng. Int. Quality and Reliability Engineering International 20.5 (2004): 511-26.
[29] Benjelloun, Fatima-Zahra, Lahcen, Ayoub Ait, Belfkih, Samir. "An Overview of Big Data Opportunities, Applications and Tools." 2015 Intelligent Systems and Computer Vision (ISCV) (2015).
[30] Zhao, Ke, and D. Steffey. "Practical Applications of Mixture Models to Complex Time-to-failure Data." 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS) (2013).

[31] Bishop, Christopher M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
[32] Peng, Xi, Zhang, Lei, Yi, Zhang. "Scalable Sparse Subspace Clustering." 2013 IEEE Conference on Computer Vision and Pattern Recognition (2013).
[33] Traganitis, Panagiotis A., Slavakis, Konstantinos, Giannakis, G.B. "Large-scale Subspace Clustering Using Random Sketching and Validation." 2015 49th Asilomar Conference on Signals, Systems and Computers (2015).
[34] Arthur, David, and Sergei Vassilvitskii. "How Slow Is the K -means Method?" Proceedings of the Twenty-second Annual Symposium on Computational Geometry - SCG '06 (2006).
[35] Huang, Zhexue. "Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values." Data Mining and Knowledge Discovery 2.3 (1998): 283-304.
[36] Ester, Martin, Kriegel, Hans-Peter, Sander, Jorg, Xu, Xiaowei. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." (1996).
[37] "PHM Data Challenge 2015." PHM Data Challenge | PHM Society. 7 Oct. 2016.
[38] Gower, J. C. "A General Coefficient of Similarity and Some of Its Properties." Biometrics 27.4 (1971): 857.
[39] Ruhi, S., Sarker, S., Karim, M. R. "Mixture Models for Analyzing Product Reliability Data: A Case Study." SpringerPlus 4.1. (2015).

Chapter 3

[40] Modarres, Ceena, Droguett, Enrique L., Fuge, Mark "A Novel Clustering Based Methodology for Overcoming Heterogeneous Populations for Reliability Prediction." Submission to Quality and Reliability Engineering International. 2016.
[41] Meeker, William Q., and Yili Hong. "Reliability Meets Big Data: Opportunities and Challenges." Quality Engineering 26.1 (2013): 102-16.
[42] Kaplan, Stan. "On a 'Two-Stage' Bayesian Procedure for Determining Failure Rates from Experimental Data." IEEE Power Engineering Review IEEE Power Eng. Rev. PER-3.1 (1983): 43.
[43] Droguett, Enrique Lopez, Frank Groen, and Ali Mosleh. "The combined use of data and expert estimates in population variability analysis." Reliability Engineering & System Safety 83.3 (2004): 311-321.
[44] Sun, Bo, Zeng, Shengkui, Kang, Rui, Pecht, Michael G. "Benefits and challenges of system prognostics." IEEE Transactions on reliability 61.2 (2012): 323-335.
[45] Yavuz, M. Cetin, Sahin, Ferat, Arnavut, Ziya, Uluyol, Onder. "Fault diagnosis for airplane engines using Bayesian networks and distributed particle swarm optimization." Parallel Computing 33.2 (2007): 124-143.
[46] Bartram, Gregory, and Sankaran Mahadevan. "Prognostics and health monitoring in the presence of heterogeneous information." Proc. Annual. Conference. Prognostic Health Management Society. Vol. 3. No. 126. 2012.

[47] Rabiei, Elaheh, Enrique Lopez Droguett, and Mohammad Modarres. "A prognostics approach based on the evolution of damage precursors using dynamic Bayesian networks." Advances in Mechanical Engineering 8.9 (2016): 1687814016666747.
[48] Liu, Jie, and Zio, Enrico. "Framework for Asset Prognostics from Fleet Data." Proc. Annual. Conference. Prognostic Health Management Society. 2016.
[49] Dindarloo, Saeid R., and Elnaz Siami-Irdemoosa. "Data Mining in Mining Engineering: Results of Classification and Clustering of Shovels Failures Data." International Journal of Mining, Reclamation and Environment (2016): 1-14.
[50] Zhao, Ke, and Steffey, Duane. "Practical Applications of Mixture Models to Complex Time-to-failure Data." 2013 Proceedings Annual Reliability and Maintainability Symposium (RAMS) (2013)
[51] Peel, David, and McLachlan, Geoffrey J. "Robust Mixture Modelling using the t Distribution." Statistics and Computing 10 (2000): 339-348.
[52] Bishop, Christopher M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
[53] Mohammad-Djafari, Ali. "Approximate Bayesian Computation for Machine Learning, Inverse problems and Big Data." 36th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt) (2016)
[54] DAGitty - Drawing and Analyzing Causal Diagrams (DAGs). Web.
[55] Mohammad-Djafari, Ali. "Variational Bayesian Approximation Method for Classification and Clustering with a Mixture of Student-t Model." International Conference on Networked Geometric Science of Information. Springer International Publishing, 2015.
[56] Bergstra, James, and Yoshua Bengio. "Random search for hyper-parameter optimization." Journal of Machine Learning Research 13.Feb (2012): 281-305.
[57] Chatzis, Sotirios, and Theodora Varvarigou. "Robust fuzzy clustering using mixtures of Student's-t distributions." Pattern Recognition Letters 29.13 (2008): 1901-1905.
[58] O'Connor, Andrew N. Probability distributions used in reliability engineering. RiAC, 2011.
[59] Lin, Jing. "Bayesian Reliability with MCMC: Opportunities and Challenges." Current Trends in Reliability, Availability, Maintainability and Safety. Springer International Publishing, 2016. 575-585.
[60] Hamada, Michael S., Wilson, Alyson, Reese Shane, Martz, Harry. Bayesian reliability. Springer Science & Business Media, 2008.
[61] Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.
[62] Cowles, Mary Kathryn, and Bradley P. Carlin. "Markov chain Monte Carlo convergence diagnostics: a comparative review." Journal of the American Statistical Association 91.434 (1996): 883-904.
[63] Wan, Muting, James G. Booth, and Martin T. Wells. "Variational Bayes for Hierarchical Mixture Models."

[64] Hoffman, Matthew D., Blei, David M., Wang, Chong, Paisley, John. "Stochastic variational inference." Journal of Machine Learning Research 14.1 (2013): 1303-1347.
[65] Westling, Ted, and Tyler H. McCormick. "Establishing consistency and improving uncertainty estimates of variational inference through M-estimation." arXiv preprint arXiv:1510.08151 (2015).
[66] Ranganath, Rajesh, Sean Gerrish, and David M. Blei. "Black Box Variational Inference." AISTATS. 2014.
[67] Salimans, Tim, Diederik P. Kingma, and Max Welling. "Markov chain Monte Carlo and variational inference: Bridging the gap." International Conference on Machine Learning. 2015.
[68] Turner, Richard E., and Maneesh Sahani. "Two problems with variational expectation maximisation for time-series models." Workshop on Inference and Estimation in Probabilistic Time-Series Models. Vol. 2. No. 3. 2008.
[69] Hoffman, Matthew D., and David M. Blei. "Structured stochastic variational inference." Artificial Intelligence and Statistics. 2015.
[70] "PHM Data Challenge 2015." *PHM Data Challenge | PHM Society*. 7 Oct. 2016.
[71] Langseth, Helge, Nielsen, Thomas D., Rumi, Rafael, Salmeron, Antonio. "Inference in hybrid Bayesian networks." Reliability Engineering & System Safety 94.10 (2009): 1499-1509.
[72] Laurenciu, Nicoleta Cucu, and Sorin D. Cotofana. "Probability density function based reliability evaluation of large-scale ICs." Proceedings of the 2014 IEEE/ACM International Symposium on Nanoscale Architectures. ACM, 2014.
[73] Zaidan, Martha A., Mills, Andrew R., Harrison, Robert F., Fleming, Peter J. "Gas turbine engine prognostics using Bayesian hierarchical models: A variational approach." Mechanical Systems and Signal Processing 70 (2016): 120-140.
[74] Okamura, Hiroyuki, Grottke, Michael, Dohi, Tadashi, Trivedi, Kishor S. "Variational Bayesian approach for interval estimation of NHPP-based software reliability models." 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'07). IEEE, 2007.
[75] Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65.
[76] Mairal, Julien, Bach, Francis, Ponce, Jean, Sapiro, Guillermo. "Online dictionary learning for sparse coding." Proceedings of the 26th annual international conference on machine learning. ACM, 2009.
[77] Quinlan, J. Ross. "Induction of decision trees." Machine learning 1.1 (1986): 81-106.
[78] Gerogiannis, Demetrios, Christophoros Nikou, and Aristidis Likas. "The mixtures of Student's t-distributions as a robust framework for rigid registration." Image and Vision Computing 27.9 (2009): 1285-1294.
[79] Archambeau, Cédric, and Michel Verleysen. "Robust bayesian clustering." Neural Networks 20.1 (2007): 129-138.

Chapter 4

[80] Hoffman, Matthew D., Blei, David M., Wang, Chong, Paisley, John. "Stochastic variational inference." Journal of Machine Learning Research 14.1 (2013): 1303-1347.
[81] Ranganath, Rajesh, Sean Gerrish, and David M. Blei. "Black Box Variational Inference." AISTATS. 2014.
[82] Azevedo-Filho, Adriano, and Ross D. Shachter. "Laplace's method approximations for probabilistic inferencein belief networks with continuous variables." Proceedings of the Tenth international conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 1994.
[83] Minka, Thomas P. "Expectation propagation for approximate Bayesian inference." Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., 2001.