Defence Research and Development Canada
Recherche et développement pour la défense Canada

CAN UNCLASSIFIED

DRDC | RDDC
technologysciencetechnologie

Predictive Analytics for the Royal Canadian Navy Fleet Maintenance Facilities

An application of data science to maintenance task completion times

David Maybury
DRDC – Centre for Operational Research and Analysis

Defence Research and Development Canada

Reference Document
DRDC-RDDC-2018-R150
December 2018

CAN UNCLASSIFIED

Canada

IMPORTANT INFORMATIVE STATEMENTS

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

In November 2016, Cdr Timothy Gibel, Directorate of Naval Strategic Management (DNSM) requested the Centre for Operations Research and Analysis's (CORA) assistance with creating a predictive model for actual work hours per task on Royal Canadian Navy (RCN) vessels at the Cape Scott and Cape Breton Fleet Maintenance Facilities (FMF). He also requested help in finding patterns in maintenance task completions at the vessel level. The FMF provided dataset contains information on 132,292 unique tasks separated into 43,731 order keys. Using regression trees with feature engineering for actual work hours, I find that a simple 14 terminal node tree explains 18% of the variance in the data. Gradient boosted stumps explain as much as 25% of the variance, but at the expense of an inter-pretable structure for schedule validation. Hidden Markov Modelling of monthly order key completion time series data reveals coastal differences between the the RCN's two maintenance facilities, providing an objective motivation to discover the source.

Significance for defence and security

In November 2016, Cdr Timothy Gibel, Directorate of Naval Strategic Management (DNSM) requested DRDC CORA's assistance with creating a predictive model for Actual Work Hours on naval vessels at the Cape Scott and the Cape Breton Fleet Maintenance Facil-ities (FMF). The FMF provided Centre for Operational Research and Analysis (CORA) with two CSV (comma separated variable) files of 62 MB and 44 MB respectively, which contained maintenance data from 5 February 2004 to 6 July 2016. This paper shows the Royal Canadian Navy (RCN) the power of using data science on its operational and trans-actional data. Given the Royal Canadian Navy's (RCN) new SAP data environment, this work provides the timely demonstration of the data science possibilities that the RCN can now exploit.

Résumé

En novembre 2016, le cdr Timothy Gibel, de la Direction de la gestion stratégique de la marine (DNSM), a sollicité l'aide du Centre de recherche et d'analyse opérationnelles (CORA) pour créer un modèle prédictif des heures de travail réelles par tâche sur les navires de la Marine royale canadienne (MRC) des installations de maintenance de la flotte au Cap Scott et Cape Breton (FMF). Il a également demandé de l'aide pour trouver des tendances dans l'achèvement des tâches d'entretien au niveau des navires. L'ensemble de données fourni par FMF contient des informations sur 132 292 tâches uniques, séparées en 43 731 clés d'ordre. En utilisant des arbres de régression avec l'ingénierie des caractéristiques pour les heures de travail réelles, je trouve qu'un simple arbre de 14 nœuds terminaux explique 18% de la variance dans les données. Les *Gradient boosting stumps* expliquent jusqu'à 25% de la variance, mais aux dépens d'une structure inter-simulable pour la validation de la cédule. La modélisation de Markov cachée des données des séries chronologiques d'acheminements mensuels des principales commandes révèle les différences côtières entre les deux installations de maintenance de la MRC, fournissant ainsi une motivation objective à la recherche de la source.

Importance pour la défense et la sécurité

En novembre 2016, le Capf Timothy Gibel, de la Direction de la gestion stratégique de la marine (DNSM), a sollicité l'aide de RDDC CORA pour créer un modèle prédictif des heures de travail réelles des navires de la marine aux installations de maintenance de la flotte de Cape Scott et de Cape Breton. Le FMF a fourni au Centre de recherche opérationnelle et d'analyse (CORA) deux fichiers CSV (variables séparées par des virgules) de 62 et 44 Mo respectivement, contenant les données de maintenance du 5 février 2004 au 6 juillet 2016. Ce document présente la Marine royale canadienne (RCN) la puissance de l'utilisation de la science des données sur ses données opérationnelles et transactionnelles. Compte tenu du nouvel environnement de données SAP de la Marine royale canadienne (MRC), ce travail fournit une démonstration opportune des possibilités en science des données que la MRC peut maintenant exploiter.

1 Statement of results

I build a regression tree (see [1], [2]) with R [3] using packages [4], and [5], based on the log of actual work hours as the target. The key results are:

- The regression tree explains 20% of the variance in data;
- Feature engineering through unsupervised learning gives the regression tree an interpretable structure;
- The R scripts yield prediction quantiles of actual work hours based on task features. The client can include these scripts within the RCN's SAP Business Objects database tool.

Using time series analysis and Hidden Markov Models (HMM) (see [6], [7], and [8]) with R [3] (including package [9]) with monthly order key completion data by vessel data, the key results are:

- FMF Cape Scott show a downward trend on monthly order key completions from 2008 to 2016 while FMF Cape Breton does not show a trend over 2008 to 2013;
- Once accounting for trends, the monthly completion data on each coast shows one-month autocorrelation.
- A five-state HMM describes the data generating process;
- Clustering in the HMM rate space identifies each FMF, suggesting different challenges and business practices between the coasts.

The regression tree is a predictive model which helps FMF planners better understand the risk of schedule creep. Planners can use the model to validate technician built schedules by comparing planned hours to the information in the terminal nodes of the regression trees (e.g., in a 100 job task schedule, roughly 5 tasks should have planning time estimates that exceed the 95% quantile). The time series analysis demonstrates that the maintenance system has memory and that each FMF faces different challenges. Interpretation remains problematic—we do not know the reason for the coastal differences in order key completion rates. Order key completion records occur on the date at which the FMF closes a ticket and each FMF has its own close out polices. Given the downward trend in Figure 15 relative to Figure 16, there are differences between the two facilities. The time series analysis shows that we can identify each FMF by observing historical work patterns at the individual vessel level.

Acknowledgements

I would like to thank LCdr Andrew Sargeant and LCdr Jon Lee for their help in making this work possible. Both officers proved invaluable in helping me understand and interpret the Fleet Maintenance Facility datasets. Without their regular input and patience, this work would not have been possible.

Table of contents

List of figures

List of tables

2 Introduction and Data Overview

Defence Resource Management Information System (DRMIS) unifies Department of National Defence (DND)'s financial information and materials management into a single Enterprise Resource Planning (ERP), replacing both Financial and Managerial Accounting System (FMAS) and Materiel Acquisition and Support Information System (MASIS). The system improves data visibility and transactional efficiency through a harmonized and expandable software system. DND uses a SAP[1] ERP solution which includes SAP's web intelligence and Business Objects.

Given the new visibility of the RCN's data, in November 2016, Cdr Timothy Gibel, Directorate of Naval Strategic Management (DNSM), requested Defence Research and Development Canada (DRDC) Centre for Operational Research and Analysis (CORA)'s assistance with creating a predictive model for actual work hours on naval vessels at the Cape Scott and Cape Breton Fleet Maintenance Facility (FMF). He also requested help in identifying the differences in maintenance activity between the two coasts. LCdr Jon Lee and LCdr Andrew Sargeant of FMF (Cape Scott) provided data on the maintenance tasks by using SAP Business Explorer (BEx) queries. The data consists of two Comma Separated Variable (CSV) files of 62 MB and 44 MB respectively. Together, the data files contain information on 132,292 individual ship specific maintenance tasks in 43,731 unique order

[1] https://www.sap.com/corporate/en/company.html

keys over the date range 5 February 2004 to 6 July 2016. All tasks have both start and end dates—none of the tasks are time censored. Each task contains the following information (parenthetical values indicate the number of covariate levels):

- Order Key initiation time;
- Order Key close out time;
- Actual Hours;
- FMF (2;)
- Vessel name (21);
- PM activity (44);
- Materiel status (2);
- Work center (254) and;
- Functional location (1801).

Tasks belong to order keys and an order key can contain more than one task. Actual Hours record the amount of actual work time on the task while the lapsed order key time indicates the time between order key initiation and close out. In this study, I concentrate only on vessel maintenance activity, and the dataset contains information on 21 vessels (Halifax, Tribal, Protecteur, and Victoria Class). Each maintenance task belongs to a PM activity and a functional location on the vessel. Each task is routed to a work center with a materiel status indicating the presence of missing parts or materiel during the order key's life. The data contain some missing information in the Actual Hours covariate—approximately 30% of the tasks do not have a record for Actual Hours. I split the data analysis into two parts—a regression tree analysis (see [1] and references therein) that predicts of Actual Hours based categorical task covariates, and a time series analysis of monthly order key completions. In the regression tree analysis of Actual Hours, I use only the uncensored dataset, reducing the number of individual tasks to 91,482. In addition to the missing Actual Hours data, some of the covariates area associated with a small numbers of observations. Before beginning the regression tree analysis, I conduct high level feature engineering by coercing all levels within a covariate which has less than 30 observations into a single level called "OTHER". These activities occur rarely and for the purpose of the regression analysis they belong to a single, rare class. The feature engineered dataset contains the following covariate and level information:

- Order Key initiation time;
- Order Key close out time;
- Actual Hours;
- FMF (2);
- Vessel name (21);
- PM activity (32);
- Materiel status (2);
- Work center (173) and;
- Functional location (573).

The final result is a single pruned regression tree with quantile informations for each of the tree's 14 terminal nodes. I grow the tree based on an unsupervised learning step in the covariate levels to reduce level complexity and I compare the result to gradient boosted methods.

In the second case, I analyze time series data on monthly order key completions through the lens of Hidden Markov Model (HMM) (see [7], and [8]). The monthly order keys do not have any missing data although the vessel time series do not all begin on the same date. Based on the analsyis of pseudo-residuals and Akaike Information Criteria (AIC), I build a 5-state Poisson HMM for each vessel's monthly order key completion time series. The final result indicates a difference between the operations at the FMFs in that clustering analysis of the HMM rate vectors separates the vessels based on their coastal assignments.

3 Regression Models for Actual Hours

The covariates associated with the positive continuous Actual Hours target are all categorical. Regression trees are well suited for this type of data [1] which, based on a loss function, cut the target space into regions associated with the covariate information. In this study, I use CART (see [1] and references therein), and gradient boosting implemented in R with Rpart [4] and Gradient Boosted Model (GBM) [10].

Since Actual Hours span the range 0.25 to 11053, I log transform the data, resulting in the new target Log Actual Hours (LAH). The client requires more than a simple prediction of the mean LAH. To be useful for schedule risk assessment, the client needs distributional information on the subpopulations that reside inside the LAH target variable. In this circumstance, the client needs at least a partially interpretable model. I aid interpretation by reducing the covariate space through feature engineering in the mean-standard deviation within the each of the covariate levels.

To gauge the highest amount of variance we can hope to explain, consider Figure 1, which provides distributional comparisons of Actual Hours to the normal distribution. Notice that while we can reject the normal as a fit to the data (the Kolmogorov-Smirnov test confirms the rejection) we see that qualitatively, the data has normal-like properties. The structure in each tail suggests the possible presence of subpopulations but, given the approximate normal distribution description, we should not anticipate that regression techniques will explain most of the variance. The data is noisy. Table 1 summarizes the Actual Hours data.

Table 1: Empirical distributional information on Actual Hours across all tasks.

1st Quartile	Median	Mean	3rd quartile
8	19	39	48

We can detect the presence of subpopulations by using a Gaussian Mixture Model (GMM) (see [1] for details) on the LAH target. Figure 2 shows that the data supports between 8 and 15 Gaussian mixtures [5], which suggests internal structure that a suitable regression tree will discover.

To get a sense of how well regression trees might perform on the data, I fit a gradient boosted model (GBM) of stumps (see [1] for details) with LAH as the target using the covariates (number of levels in parentheses):

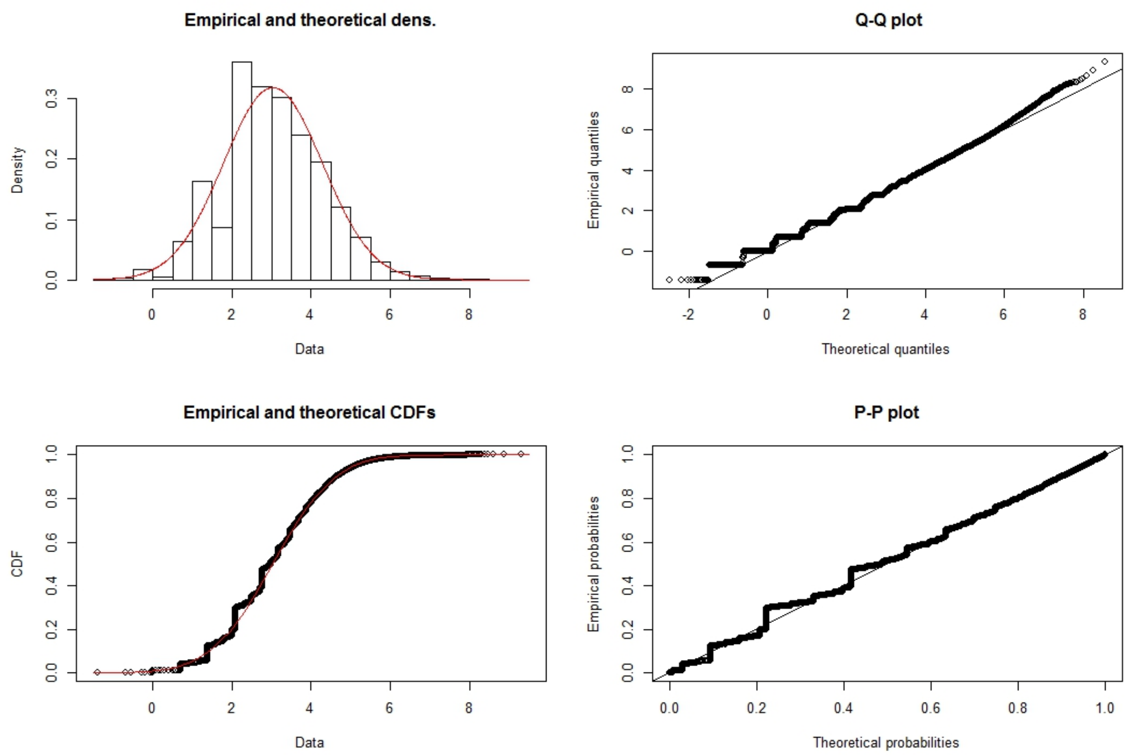

Figure 1: LAH comparisons to the normal distribution. The data are not explained by the normal distribution, but the approximate normal description implies that regression techniques will not explain most of the data's variance.

- Order Key initiation time;
- Order Key close out time;
- FMF (2);
- Vessel name (21);
- PM activity (32);
- Materiel status (2);
- Work center (173);
- Functional location (573).

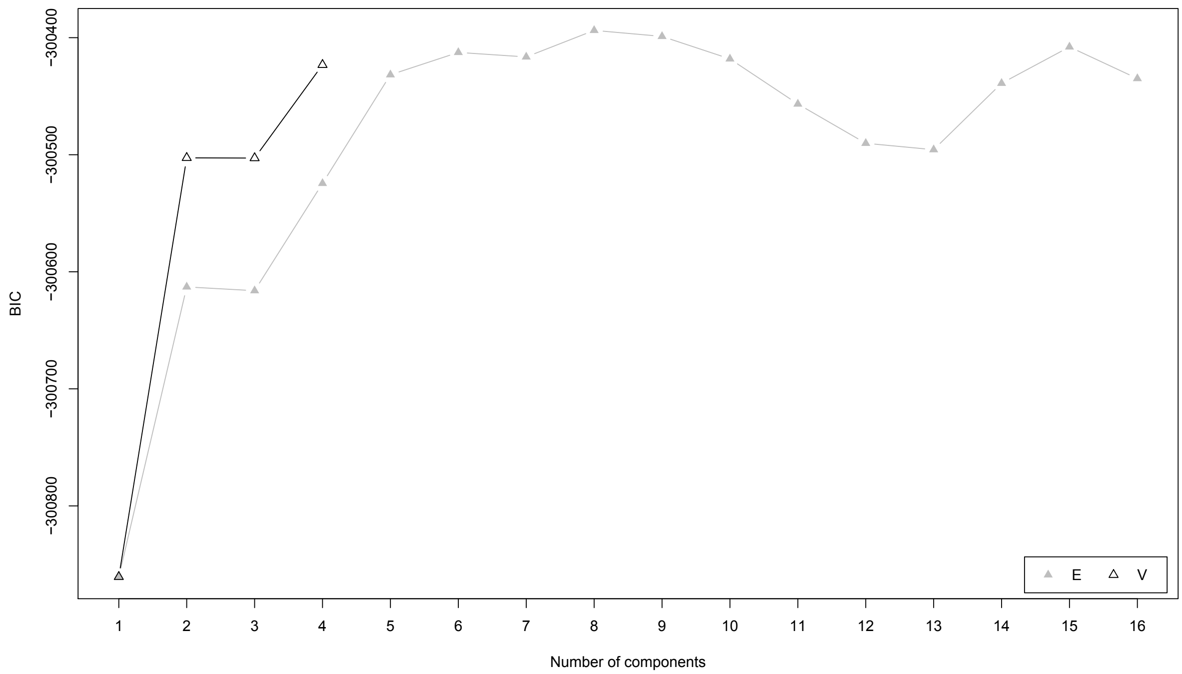

Figure 2: GMM applied to the LAH. Notice the presence of multiple subpopulations. The two curves label equal (E) and unequal (V) variance components.

I train the GBM on 90% of the data, holding back 10% for validation. Using Out-of-bag (OOB) sampling, the GBM contains 7425 stumps and the application of the GBM on the validation set explains 24.8% of the variance in the data. Table 2 shows the variable importance[2] of the GBM.

Table 2: GBM variable importance (24.8% variance explained).

Work center	Functional location	PM Activity	Materiel Status	Customer	FMF
66.8%	30.3%	1.5%	1.4%	0.0%	0.0%

The GBM provides predictions of mean LAH based on covariate information but the client requires a distributional understanding of subpopulations. The client cannot undestand

[2]The model determines variable importance from the variable's strength in improving the loss function at each stump and how often the variable appears across all stumps. See [10] for details.

schedule risk from a point prediction. Given the large amount of unexplained variance, we need a model that captures subpopulation information even if the method sacrifices some of the explained variance. The GBM does not allow us to make a clean interpretation of LAH subpopulations, but it does provide us with an upper bound on how much variance we can expect to explain.

Single regression trees, using Classification and Regression Tree (CART), offer a more interpretable structure in that the terminal nodes of the tree reveal information on the data's subpopulations. Notice that the number of levels vary widely across the covariates even after the level reduction that results from pooling the small number observation levels. To simplify the levels, I feature engineer by using GMM in the mean-standard deviation LAH space of each covariate's levels. In Figures 3, 4, 5, and 6 we see that the unsupervised learning approach finds, by maximizing the Bayesian Information Criteria (BIC), a small number of clusters in the levels associated with each covariate:

- Vessel name (2);
- PM activity (3);
- Work center (2);
- Functional location (2).

Thus, instead of using hundreds of levels in an unbalanced setting, we can collapse the covariate levels down to a handful—assigning the covariate's levels according to its associated cluster. The reduced feature space balances the levels, which helps with variable importance identification. While Bayesian Information Criteria (BIC) favours a low number of levels, I consider three different feature engineered scenarios by expanding out to four Gaussian mixtures in each covariate's mean-standard deviation level space[3]. Figures 3—6 show that BIC disfavours a large number of clusters, but that including up to four does not lead to a large drop in BIC.

In the first scenario, I use the number of levels implied by BIC, and in the second and third scenarios I force the GMM to find exactly three and four clusters in the levels of each covariate respectively. The three different scenarios allow us to build three different regression trees. Table 3 show the results where again I hold back 10% of the data for validation. I have included random forests and GBMs on each featured engineered scenario for comparison.

[3]In the large cluster number limit, we return to the original GBM model—a tree structure without any feature engineering.

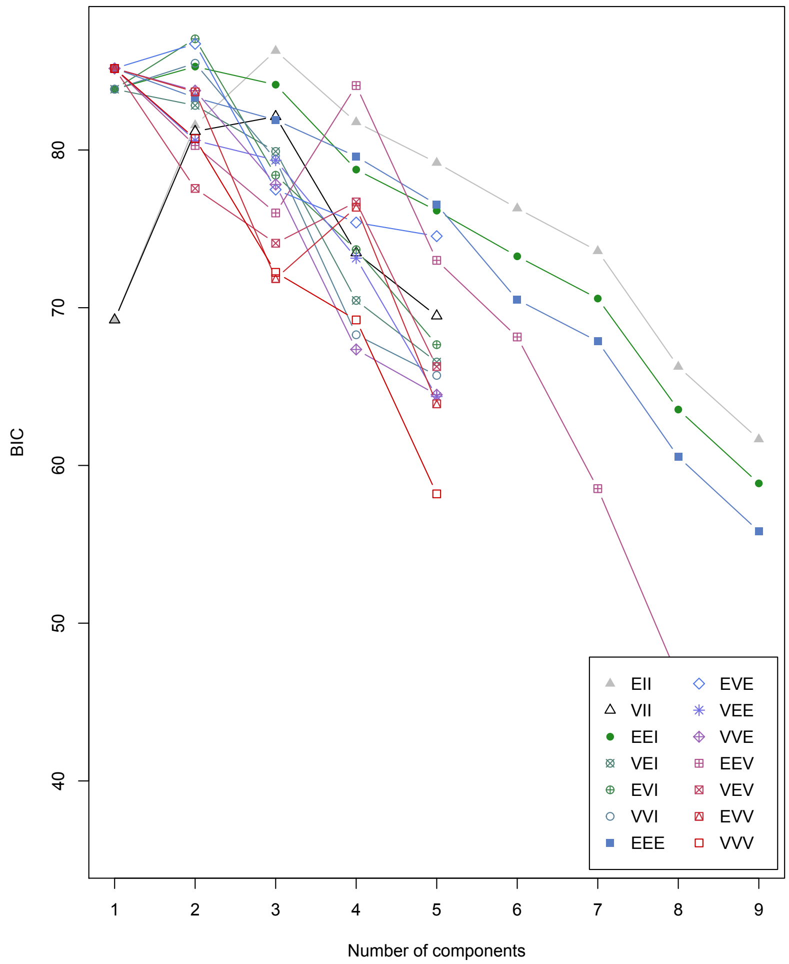

Figure 3: GMM BIC analysis of clusters in mean-standard deviation LAH by vessel name. The different curves label different parameterizations of the GMM covariance matrix. See [5] for details.

Notice that the four cluster model explains 18.6% of the variance while the GBM applied to the unengineered feature space (Table 2) explains 24.8%. The dramatic reduction of in the complexity of the feature space only sacrifices 6.2% of the explained variance, but leaves us with a more interpretable structure. For the rest of this paper, I will focus the analysis on the four cluster model with CART.

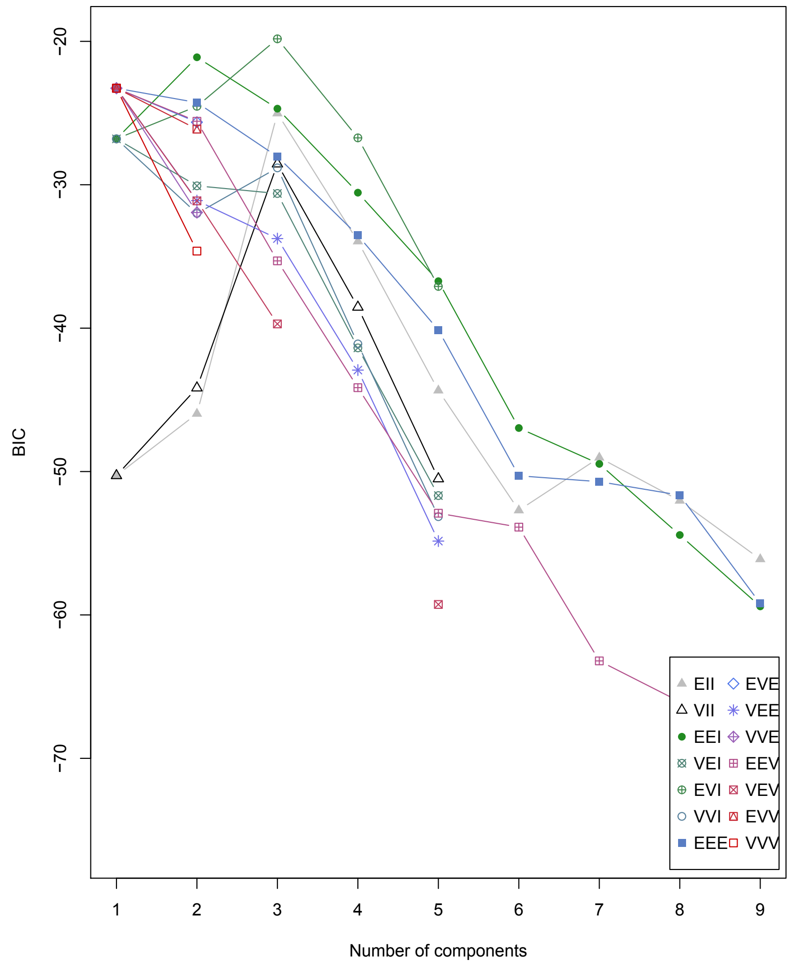

Figure 4: GMM BIC analysis of clusters in mean-standard deviation LAH by PM Activity. The different curves label different parameterizations of the GMM covariance matrix. See [5] for details.

In figures 7, 8, 9, and 10 we see the four clusters in the mean-variance space of each covariate's levels determined by the GMM. Tables 5, 6, 7, and 8 show the cluster membership of the levels. I build the regression tree to a depth of 70 trees (which overfits the data) and then prune back based on Rpart's complexity parameter. Rpart uses 10-fold cross validation to determine the relative error at a fixed complexity parameter values. As a function

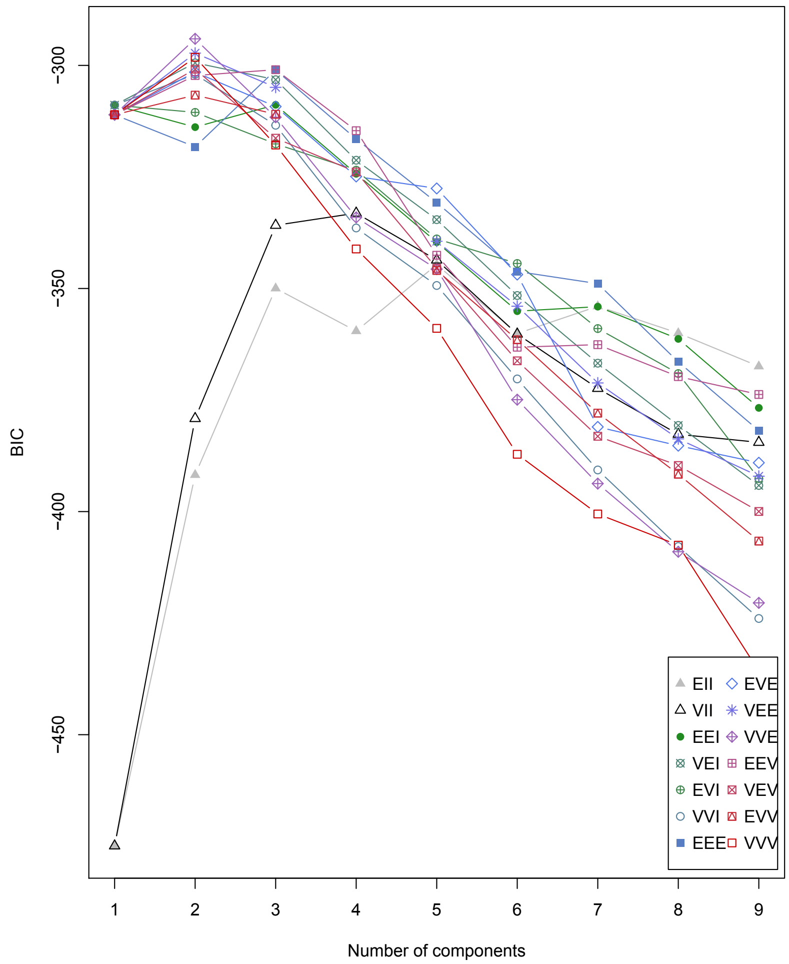

Figure 5: GMM BIC analysis of clusters in mean-standard deviation LAH by work center. The different curves label different parameterizations of the GMM covariance matrix. See [5] for details.

of tree depth, the relative error eventually reaches a plateau which allows us to prune the tree to an optimal depth. I locate the optimal tree by finding the tree depth which has the bottom of its 10 fold cross validation error bar touching the top of the last error bar in the plateau that arises from the overfit. Figure 11 shows the pruning method. Using this technique, I find the optimal tree has 14 terminal nodes, which concords with the approx-

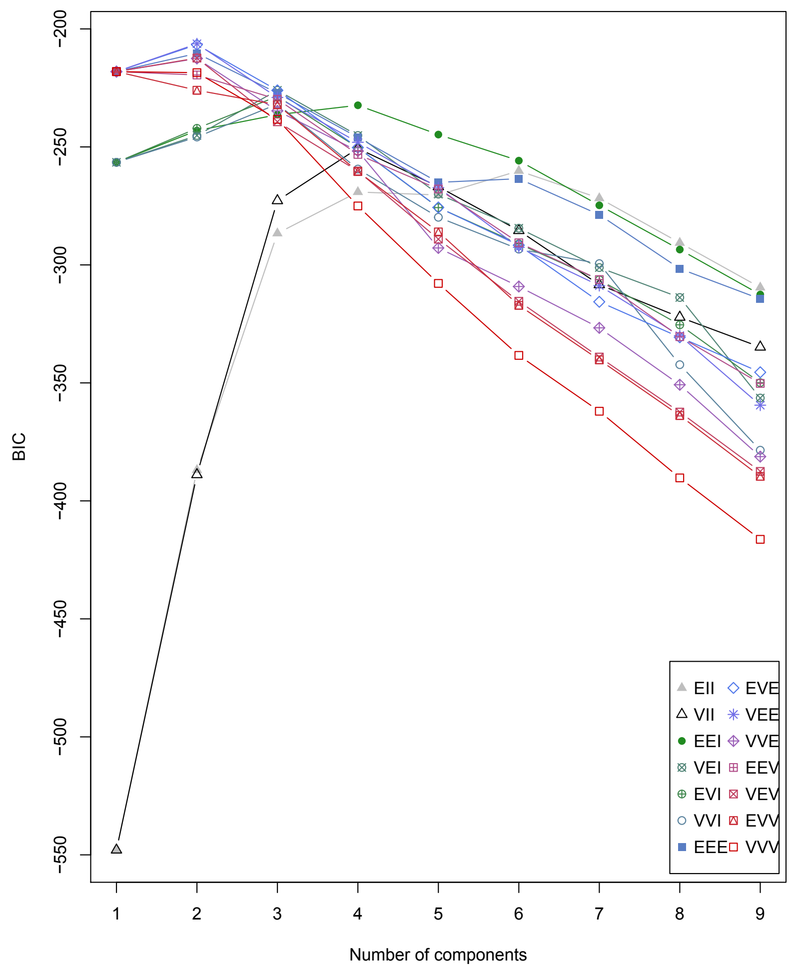

Figure 6: GMM BIC analysis of clusters in mean-standard deviation LAH by functional location. The different curves label different parameterizations of the GMM covariance matrix. See [5] for details.

imate number of subpopulations indicated by the GMM in Figure 2. Table 4 shows the variable importance of the four cluster feature space regression tree which compares well with the unengineered GBM variable importance of table 2. As compared to the GBM, the feature engineering of the four cluster model shares the relative variable importance of the functional location with the PM Activity. Notice that even though we have radically

Table 3: Scenario explained variance (out of sample vaildation)

	Regression method		
Scenario	Random Forest (300 trees)	GBM (OOB)	CART (pruned)
BIC cluster	13.1%	13.4%	13.8%
Three cluster	15.8%	17.0%	16.7%
Four cluster	18.2%	18.9%	18.6%

reduced the covariate levels as compared to the GBM with the unengineered feature space, the order of the variable importance remains the same.

Table 4: Four cluster scenario variable importance (18.6% variance explained).

Work center	Functional location	PM Activity	Materiel Status	Customer	FMF
69.3%	15.1%	10.8%	3.4%	0.9%	0.6%

The 14 terminal node tree allows us to build Actual Hour quantiles for each node, which can help the client understand potential variability and risk within planned schedules. Figure 12 shows the regression tree with the primary nodes[4]. Notice that the Work center forms the first split and continues as a splitting variable throughout the tree. The Work center has the highest relative importance among the covariates. Figure 13 shows the distributional information associated with each of the 14 terminal nodes, including node quantile information.

[4]Rpart builds the regression tree with both primary and surrogate nodes. The tree uses surrogate node information in the absence of primary node information and Rpart incorporates surrogate node strength in determining variable importance (see [4] for details).

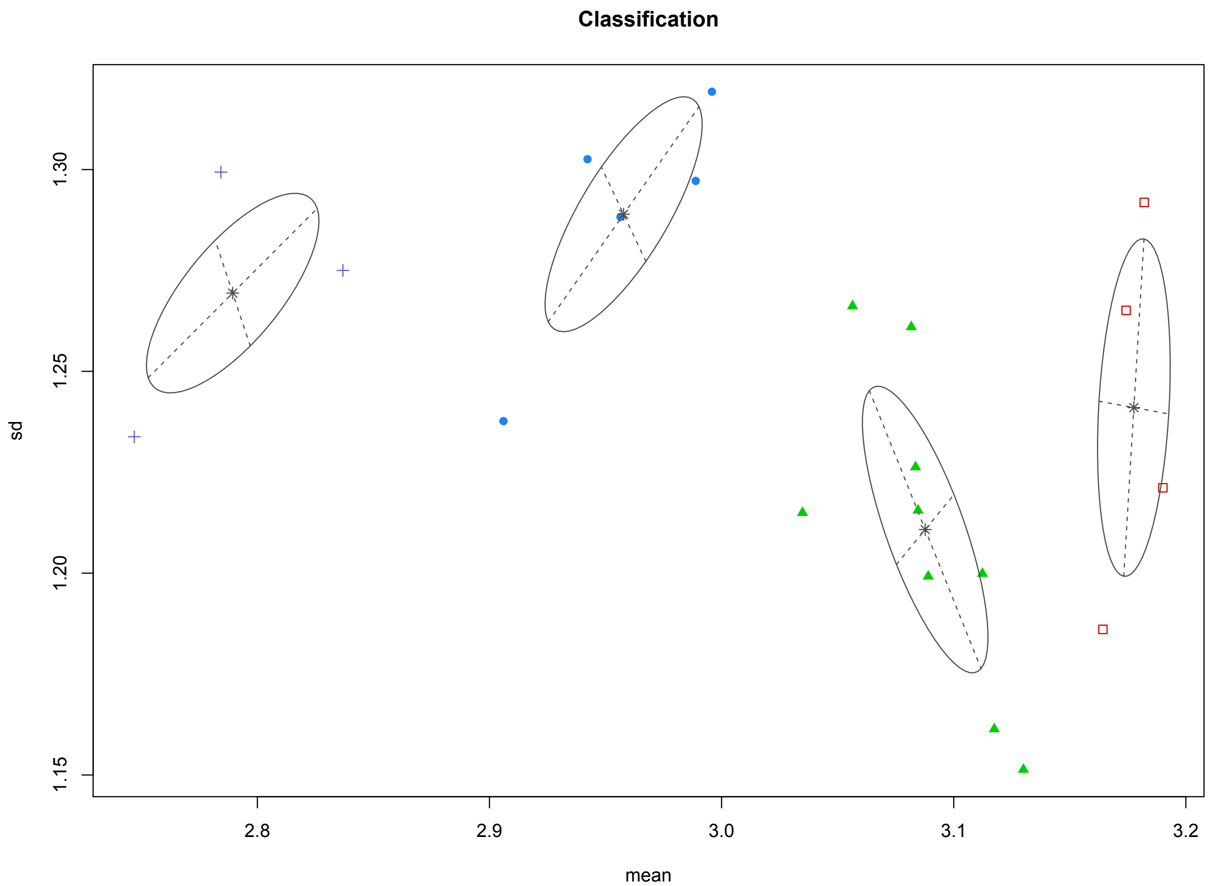

Figure 7: Four cluster GMM applied to vessel name in the LAH mean-standard deviation space. Standard error ellipsoid indicated.

The four cluster model provides the client with distributional results that a straightforward application of boosted models do not. Since the client is interested in understanding the distributional relationships of subpopulations which live in the data, the pruned regression tree offers the client insight that a point prediction from a boost model cannot. The greater interpretability of the regression tree—which has only 14 terminal nodes—relative to the GBM makes up for small sacrifice in explained variance. We see that regardless of the model, the data remains noisy.

To test the model's strength in the face of the noisy data I compute the Naive Bayes Clas-

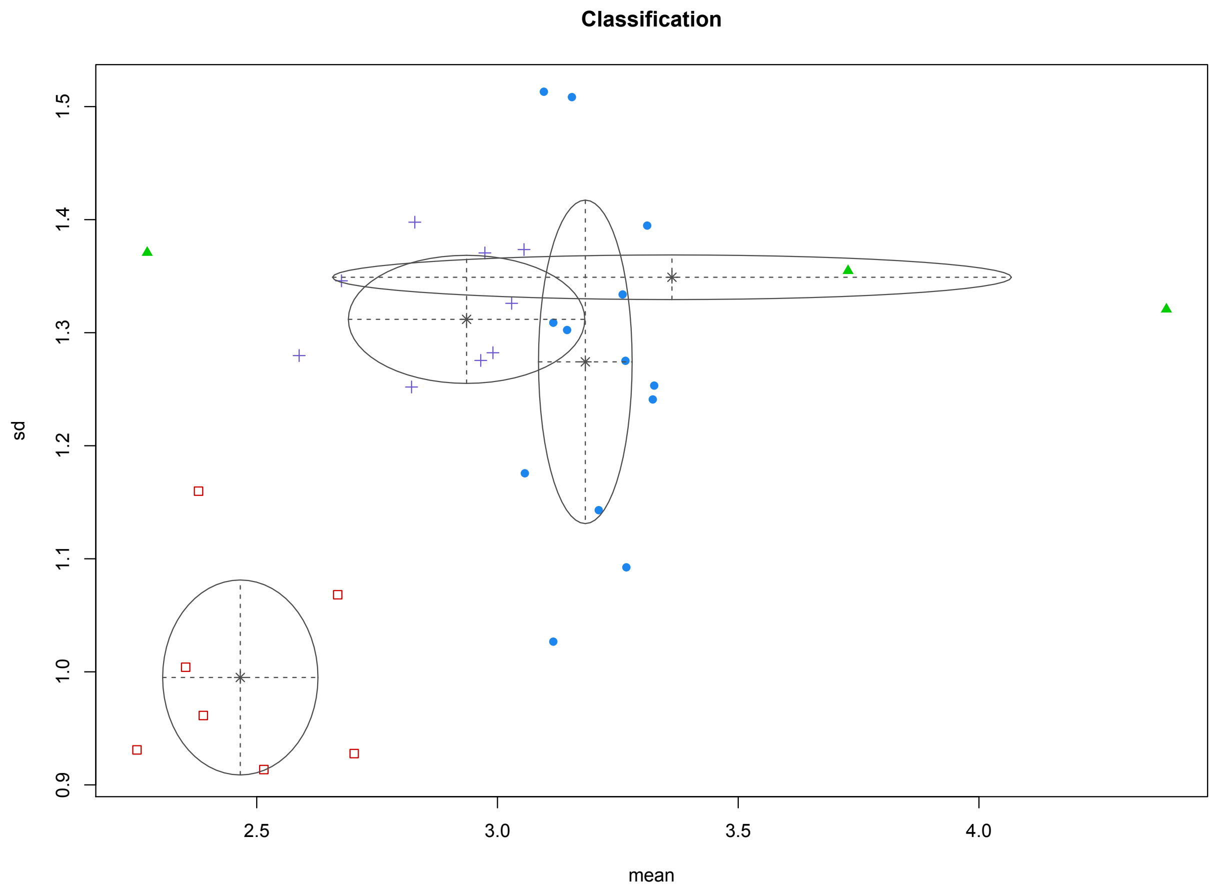

Figure 8: Four cluster GMM applied to PM activity in the LAH mean-standard deviation space. Standard error ellipsoid indicated.

sifier (see [1] for details) of each terminal node as a function of LAH, displayed in Figure 14. The figure shows the probability of reaching a terminal node conditioned on the LAH value. Figure 12 labels the terminal nodes by their respective mean LAH. We see the trend that high values of LAH lead to high probabilities of ending up in terminal nodes with large means. We can also see the effect of the noise in the data—LAH values larger than 5 only imply a probability of less than one chance in three of belonging to the terminal node with the highest mean.

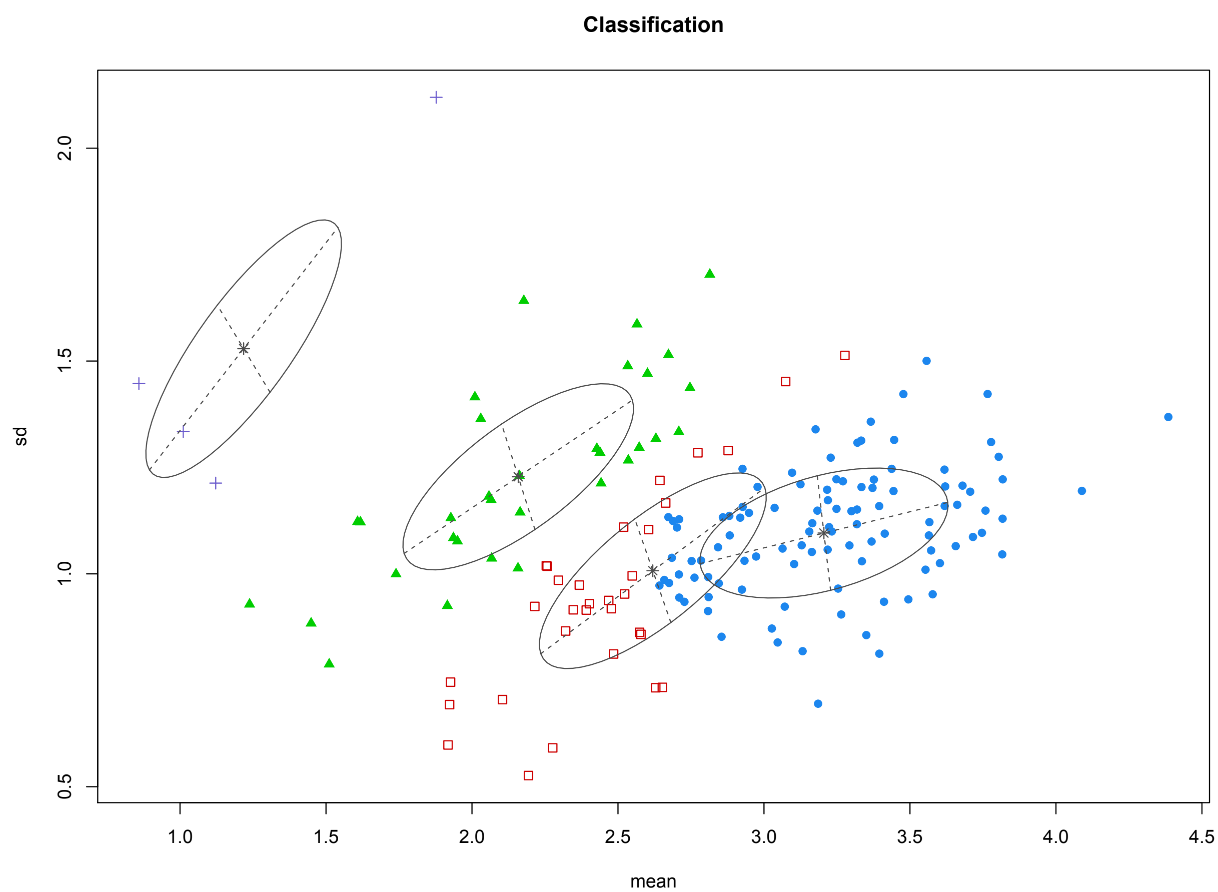

Figure 9: Four cluster GMM applied to work centers in the LAH mean-standard deviation space. Standard error ellipsoid indicated.

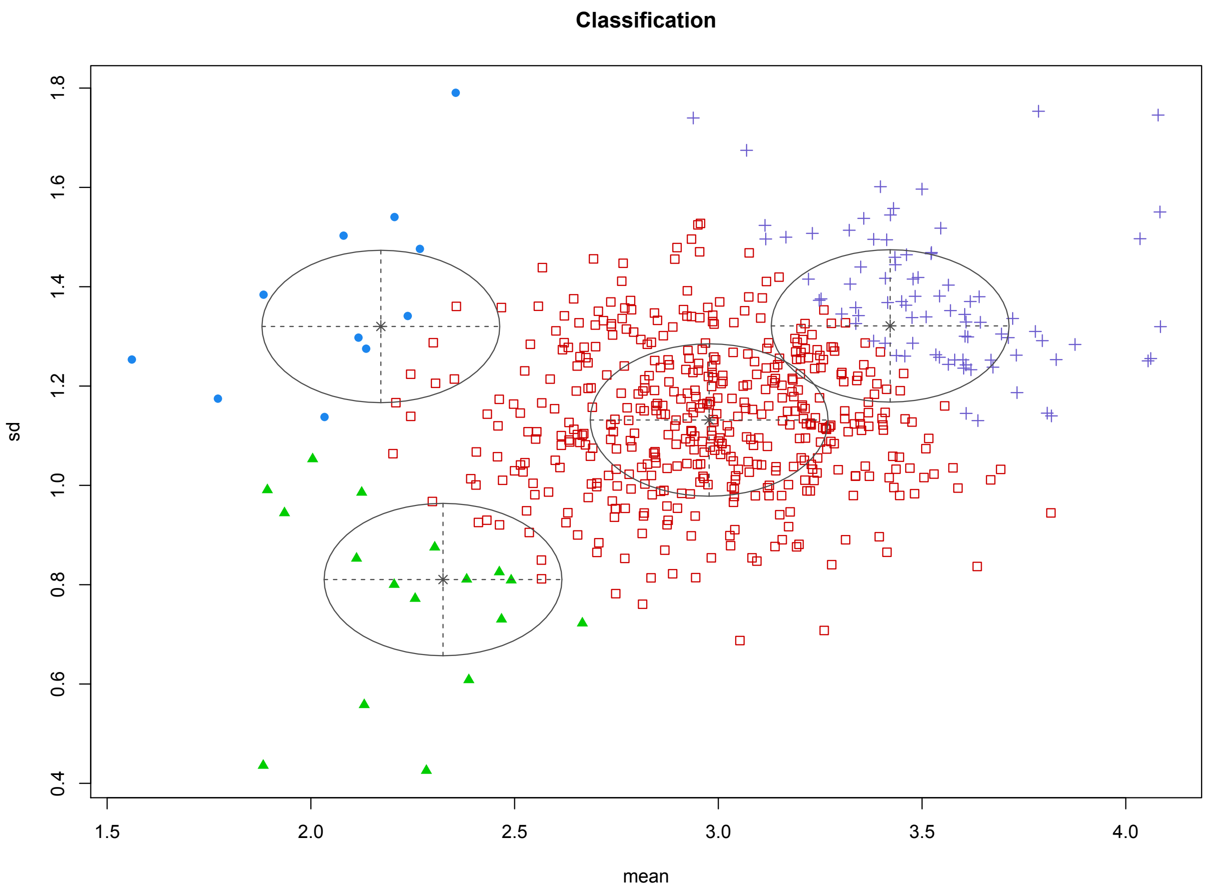

Figure 10: Four cluster GMM applied to functional locations in the LAH mean-standard deviation space. Standard error ellipsoid indicated.

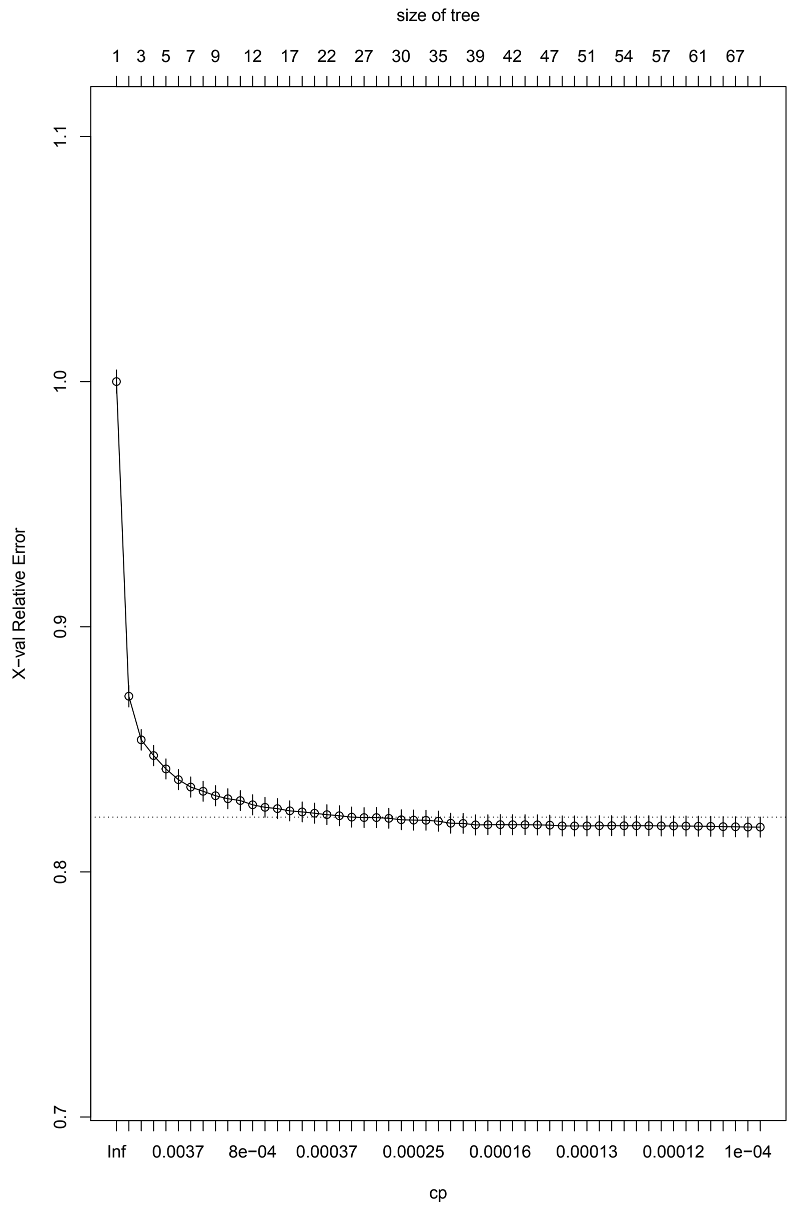

Figure 11: Complexity parameter relative error by 10-fold cross validation. The method selects 14 terminal nodes.

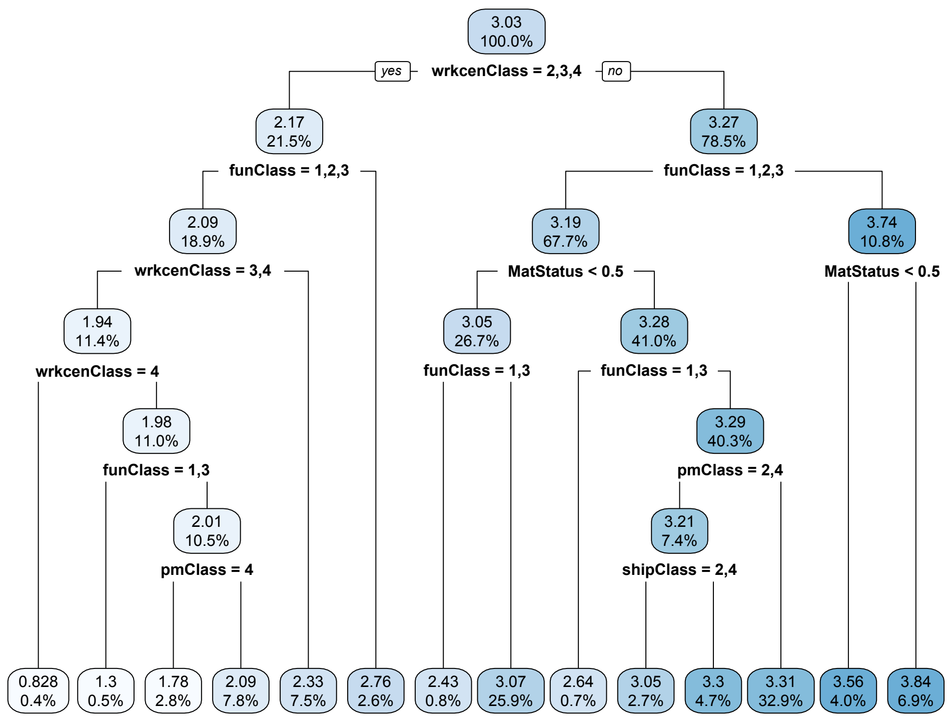

Figure 12: Pruned regression tree built from the four cluster model. The primary nodes are labelled by the LAH mean along with the percentage of the data contained within the nodes.

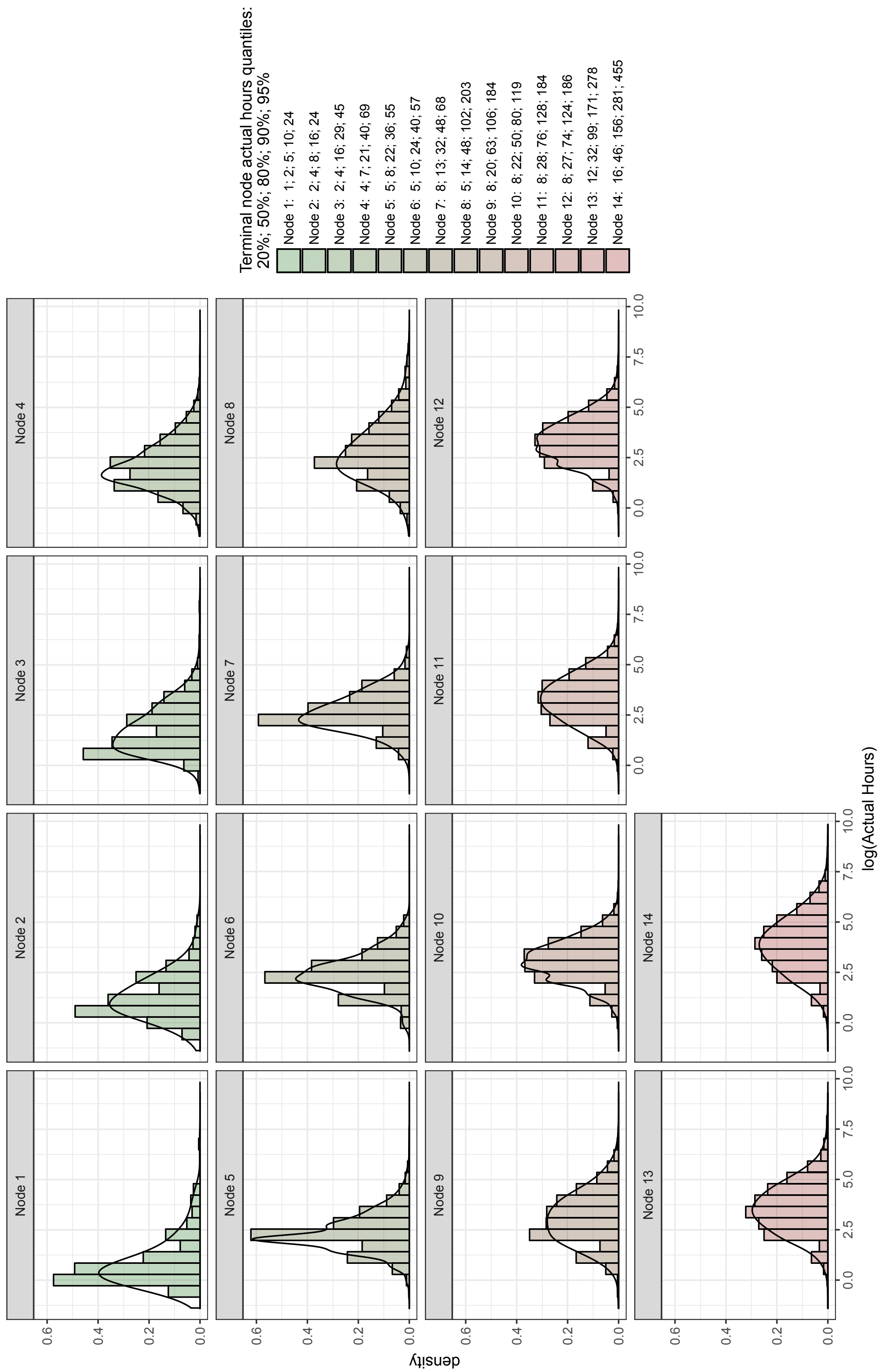

Figure 13: LAH terminal node distributional information.

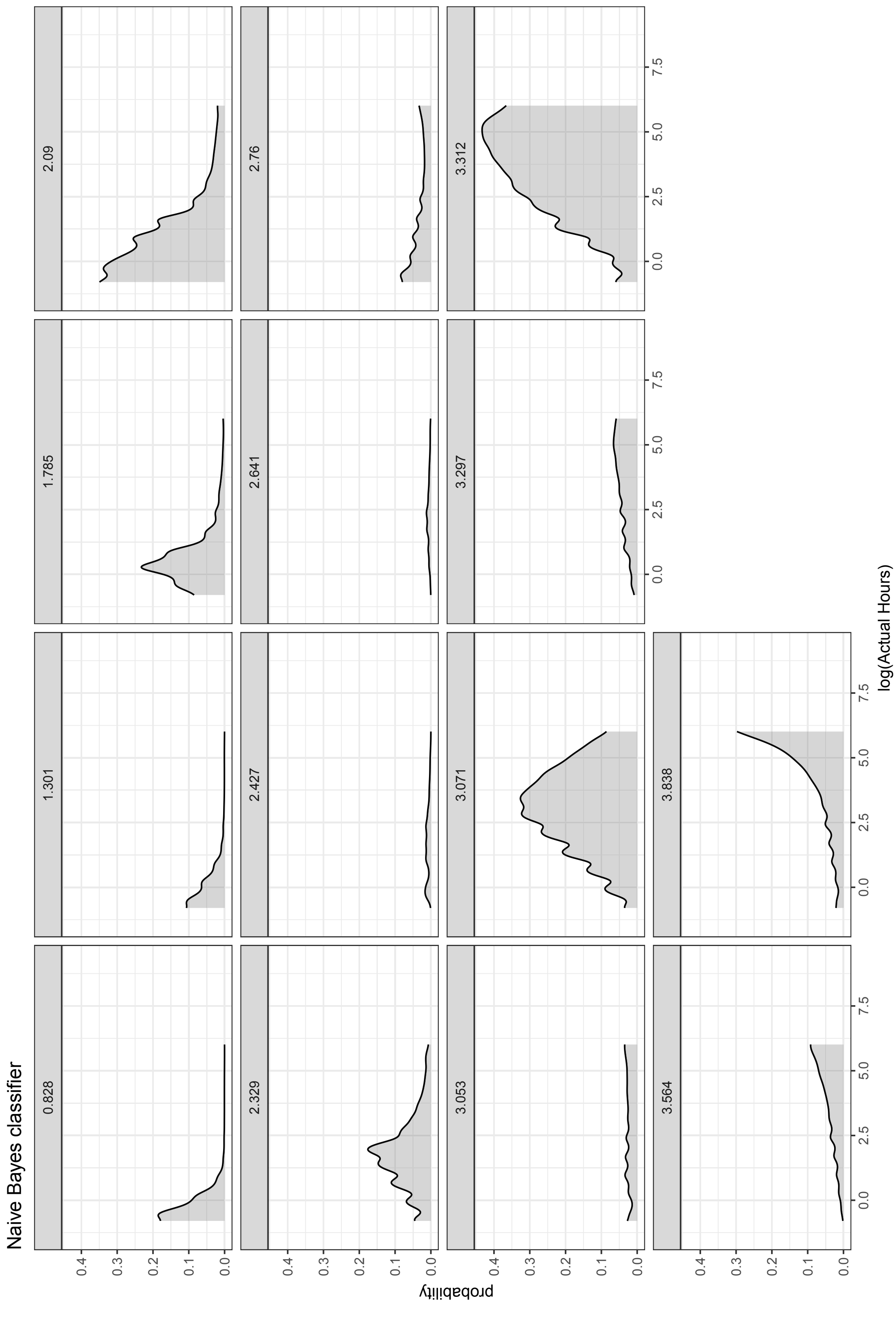

Figure 14: Naive Bayesian Classifier of terminal nodes based on LAH.

Table 5: Vessel features: GMM classification assignments with four clusters.

Blue (1)	Red (2)	Green (3)	Purple (4)
HMCS Algonquin	HMCS Athabaskan	HMCS Charlottetown	HMCS Chicoutimi
HMCS Calgary	HMCS Iroquois	HMCS Cornerbrook	HMCS Vancouver
HMCS Ottawa	HMCS Protecteur	HMCS Fredericton	HMCS Victoria
HMCS Regina	HMCS Windsor	HMCS Halifax	
HMCS Winnipeg		HMCS Montreal	
		HMCS Preserver	
		HMCS St. John's	
		HMCS Toronto	
		HMCS Ville de Quebec	

Table 6: PM Activity features: GMM classification assignments with four clusters.

Blue (1)	Red (2)	Green (3)	Purple (4)
Capital EVAL	Certifications	Costed	NP Canavmod
Capital Major	Condition Based Assessment	NP Major	NP Minor
Capital Minor	Inspections	Trials	Other Capital
Corrective	MRS Particularization		Other Work
Growth Work	Safety		PM Arising
Manufacturing	Surveys/Audits		Preventive RF
NP EVAL	Technical Assistance		Preventive SS
Other Non Capital Work			Temporary (M/F)
PMRS			Warranty
Re-Work			
Repair & Overhaul 3rd Line			
Services			
Supplemental			

Table 7: Work center features: GMM classification assignments with four clusters.

Class	Work Centers			
Blue (1)	*****DELETED***** - USE 12148	ELECTRICAL	MARINE MACHINERY SURFACE	SHIPWRIGHT
	250 BARGE	ELECTRICAL SYSTEMS EC'S & PROJECTS	MECHANICAL FITTERS	SHIPWRIGHTS
	ABOVE WATER WEAPONS	ELECTRICAL TEST	MOTOR	SHIPWRIGHTS & STAGING
	ANTENNA MFWC	EW / CCS / Nav Aids	MULTOTS	SONAR
	ATS	FIRE CONTROL SHOP	NAV AIDS SUPPORT	SONAR CANTASS
	AUXILIARIES 2	FOUNDRY	NAV AIDS SUPPORT METER SECTION	SONAR ELECTRONIC
	AWSE -Inspection Naval Ordnance	GAS TURBINE	NAV AIDS SUPPORT SHOP	SONAR MECH
	BATTERY	GAS TURBINES	NAV ARC AND MATERIAL ENGINEERING (NA/ME)	SUB PIPE FABRICATION
	BELOW WATER WEAPONS	GENSETS & FIRE PUMPS (SURFACE)	NESTRA	SUB ELECTRICAL
	BOILERMAKERS	GOVERNOR SHOP	Optics	SUB PLATE
	CABLE	GUN WEAPONS SYSTEMS	PIPE FABRICATION	SUB SHEET METAL
	CABLING	GUNS AND MISSILES - ELECTRICAL	PIPEFITTERS	SUB SONAR *****DELETED*****
	CCS ENGINEERING	GUNS AND MISSILES - ELECTRONIC	PLANT MAINTENANCE	SUB WELDING
	COATINGS	GUNS AND MISSILES - MECHANICAL	PLATE	TEMPORARY LIGHT
	COMM / NAV ENG	HEAVY ELECTRICAL	RADAR / FIRE CONTROL FITTERS	TILE PREPARATION
	COMM Ship	HULL SURVEYORS	RADAR / IFF / TACAN / IR	TILING
	COMMUNICATION	HYDRAULICS	REFRIGERATION & AIR CONDITIONING	UNDERWATER RANGE
	COMMUNICATIONS	IMCS	RIGGERS	UNDERWATER WEAPONS - ELECTRICAL
	CONTROL SYSTEMS	LAGGERS	RIGGING	UNDERWATER WEAPONS - ELECTRONICS
	CONTROL SYSTEMS Electronics	LAGGING	RIGGING LOFT	UNDERWATER WEAPONS - MECHANICAL SYSTEMS
	CONTROL SYSTEMS ENGINEERING	MACHINE SHOP	SAIL/LIFE RAFT/CANVAS	UNDERWATER WEAPONS -MECHANICAL
	CONTROL SYSTEMS Mechanical	MACHINING SERVICES	SCALERS & CLEANERS	VCS FC / SONAR TECHNOLOGISTS
	CREW - CCS/Radar/Electronic Warfare	MAIN PROP & AUX (SURFACE)	SENSORS - ELECTRONICS	WELDERS
	DI/EHM/TRIALSTION	MARINE ELECTRICAL ENGINEERING	SENSORS - MECHANICAL SYSTEMS	WELDING
	DIESEL TECHS	MARINE MACHINERY	SHEET METAL	
	DIESELS	MARINE MACHINERY (NAT INV)	SHEETMETAL	
	DRAWING OFFICE	MARINE MACHINERY SUB-SURFACE	SHIPS SYSTEM ENGINEERING	
Red (2)	AUXILIARIES 1	GUIDED WEAPONS	PNEUMATICS	SUBMARINE ELECTRICAL
	COMM	INO	POWER SYSTEMS	SUBMARINE HULL SURVEYOR
	CONO - PAINT & TILE	LIFTING APPLIANCE CERTIFICATION	PRESSURE VESSELS	SUBMARINE MECHANICAL
	CONO - FIRE SENTRY	MAIN PROP TECHS	RADAR/EW TECHNOLOGISTS	SURFACE SONAR / NAV AIDS TECHS
	CONTROLS SYSTEMS	MECHANICAL PROCESS ENGINEERING	Ranges Manager	TEMPORARY VENT
	ELECTRICAL POWER SYSTEMS	NON-DESTRUCTIVE TESTING	SHIP INSPECTION	VCS / SURFACE UW WEAPONS TECHNOLOGISTS
	ELECTROPLATORS	OPTICS	SIGN & GRAPHICS	VIBRATION ANALYSIS
	FIRE CONTROL	PAINTERS	SUBMARINE COMMUNICATIONS	WELDING INSPECTION
Green (3)	ABOVE WATER SYSTEMS ENGINEERING	CONO - O & M	LIFTING APPLIANCE	PLANT MAINTENANCE ELECTRICAL
	AWSE - ELEC	CRANE OPERATORS	MAIN MACHINERY ENGINEERING	QAR Windsor EDWP
	AWSE - MECH	DIESEL INSPECTION CELL	NACE INSPECTORS	QUALITY CONTROL INSPECTORS
	C4I	ENGRAVING SHOP	NAV	SAIL/CANVASS & MARINE SURVIVAL SYSTEMS
	CANTASS	HULL INSPECTORS	NDT/NACE INSPECTORS	SURVEIL / ELEC. WARFARE ENG.
	CHEMICAL CLEANING	HULL SYSTEMS	NET-C	SURVEILLANCE SYSTEMS SENIOR ENGINEER
	COMBAT MANAGEMENT SYSTEMS TECHNOLOGISTS	HULL SYSTEMS ENGINEER	Other Center	UW ENG ELEC
	COMBAT SYSTEMS ENGINEERING OFFICER	HULL SYSTEMS TECH	PETROLEUM, OILS & LUBRICANTS	UW ENG MECH
Purple (4)	MECHANICAL AUXILIARIES			
	SAWS			
	TRIALS CPO			
	UWW SENIOR ENGINEER			

Table 8: Functional location features: GMM classification assignments with four clusters.

Class	Functional Locations			
Blue (1)	AUTOMATIC IDENTIFICATION SYSTEM (AIDS).	COMMUNICATION SYSTEM, MISCELLANEOUS	LIFERAFT, INFLATABLE, 20 MAN	SHAFT SEAL COOLING & LUBN
	AUXILIARY VENT & BLOW SYSTEM	FAN SYSTEM, VANEAXIAL, SF-12, 13, 14, 15	MAIN CABLING	SOUNDING SET, SONAR
	BOOMS & LIFTING APPLIANCES	FIRE FIGHTING (HALON GAS)	PUMP SYSTEM, CENTRIFUGAL (CHILLED WATER)	
Red (2)	1 DECK (WEATHER)	DECKS	HYDRAULIC STEERING SYSTEM BROWN BROTHERS	REPLENISHMENT-AT-SEA EQUIPMENT GROUP
	140KW MOTOR GEN & ASSOC. EQUIP (FWD/AFT)	DECKS (GENERIC)	HYDRAULIC SYSTEM TUBE 2 (WDS)	REPLENISHMENT AT SEA EQUIPMENT
	15 TON DECK CRANES	DECKS (SUPER STRUCTURE)	HYDROGEN DETECTION ASSEMBLY	RETRACTABLE POST
	2 DECK	DEGAUSSING SYSTEM (GENERIC)	HYDRONIC SYSTEM	REVERSE OSMOSIS PLANT
	2D SEARCH AND SURVEILLANCE RADAR, AN/SPS	DETECTING RANGING SET SONAR	HYDROPLANES & STEERING HYDRAULIC SYSTEM	REVERSE OSMOSIS UNIT, SKID MOUNTED
	3 DECK	DIDSBURY TORPEDO HANDLING SYSTEM	HYDROPLANES STEERING GEAR HYD PP ANC	RIGGING & CANVASS
	4 DECK	DIESEL EXHAUST SYSTEM	HYDROPLANES STEERING OPERATING GEAR	SACRIFICIAL ANODES
	440/60 HZ DIST & MAIN AC SWITCHBOARD	DIESEL GENERATOR SET, 850 KW	IFF SYSTEM GROUP, MK 12 (TRL, PTR, FSE,	SALT WATER SANITARY SYSTEM
	5004 SATELLITE TELEVISION SYSTEM	DIESEL GENERATORS & CONTROL EQUIPMENT	INDICATOR BUOY RELEASE MECHANISM GENERIC	SANITARY SYSTEM
	5KHZ/25KHZ UHF SATCOM DAMA TERMINAL	DIRECTION FINDER SET	INERTIAL NAVIGATION GROUP	SATELLITE COMMUNICATION GROUP
	ACCOMMODATION SPACES (GENERIC)	DIRECTION FINDER SET, AN/SRD-504A	INSTRUMENTS & SAFETY SYS (PIELSTIC ENG)	SATELLITE TELEVISION SYSTEM
	ADVANCED HARPOON WEAPONS CONTROL SYSTEM	DISTRIBUTION SYSTEM, AC, 450V, 3 PH	INSULATION,PAINT,DECK & HULL COVERINGS	SCUPPERS & DRAINS
	AIR CONDITIONING & VENTILATION SYSTEM	DOMESTIC HOT & COLD FRESH WATER	INTERIOR COMMUNICATIONS & ALARM	SEA WATER SYSTEM
	AIR CONDITIONING HYDRONIC SYSTEM	DOOR SPECIAL PURPOSE (HANGER)	INTERIOR COMMUNICATIONS & ALARM SIGNAL	SEARCHLIGHT, 2.5 KW XENON
	AIR CONDITIONING SYSTEM (PIELSTICK ENG)	DOORS & FRAMES	INTERNAL COMMUNICATION SYSTEM (SHINCOM)	SEARCHLIGHT, SIGNALLING TYPE
	AIR DISTRIBUTION SYSTEM	DOORS, HANGAR, VERTICAL ROLLING	INTERROGATOR-TRANSPONDER SET	SEASEARCH SYSTEM
	AIR PURIFICATION & LIFE SUPPORT	DOORS, HATCHES AND MANHOLES	LADDER, ACCOMMODATION, ARTICULATED	SEAWATER CIRC PUMP SYS,
	AIR SUPPLY SYSTEM	DOORS,HATCHES,MANHOLES & SCUTTLES	LADDERS & GRATINGS	SEAWATER CIRCULATING SYSTEM
	AIRCRAFT FUELLING SYSTEM	DOPPLER SPEED LOG SYSTEM	LADDERS, BOOMS AND STAFFS	SEAWATER SERVICE SYSTEM, FIREMAIN
	AIRCRAFT SUPPORT EQUIPMENT GROUP	DRAINS, SANITARY SYSTEM	LADDERS, GANGWAYS, RAILS, STAFFS & GRATI	SEPARATOR, OILY WATER, HYDROMEM
	AN/SRD-504 RADIO DIRECTION FINDING SET	DRIER, AIR-GAS DESICCANT	LAUNCHING SYSTEM	SERVICE STEAM & DRAIN SYSTEM
	AN/ULR-502 (SUB ESM)	DRYER, AIR-GAS DESICCANT HP	LAUNDRY EQUIPMENT GROUP	SHELL PLATING (GENERIC)
	AN/USC-69(V)3 X AND KU BAND SATELLITE CO	EBS, BIBS, HIS, SBSS	LIFE SAVING EQUIPMENT	SHELL, HULL (GENERIC)
	ANCHOR AND ASSOCIATED EQUIPMENT	ECHO SOUNDER TYPE 778JJ	LIFTING & LOAD HANDLING	SHINCOM 2128 SWITCH SYSTEM (DUAL CENTRAL
	ANCHORING,MOORING & TOWING GEAR	ELECTRONIC SUPPORT MEASURES AN/SLQ-501	LINK 16	SHIP'S SERV 120V 60 HZ LTG & PWR DIST SY
	ANTENNA AS-5196/U	ELECTRONIC SUPPORT MEASURES SET	LOBBYS AND PASSAGEWAYS (GENERIC)	SHIPBOARD ELECTRIC CLOCK SYSTEM
	ANTENNA GROUP OE-5034/SR	ELECTRONIC WARFARE EQUIPMENT GROUP	LOCKER (DAMAGE CONTROL)	SHIPBOARD ELECTRO-OPTIC SURVEILLANCE SYS
	ANTENNA GROUP OE-5036/SR	ELEVATOR AMMUNITION	LONG RANGE SURVEILLANCE RADAR SYSTEM	SHIPBOARD INTEGRATED NAVIGATION/DISPLAY
	ANTENNA GROUP, OE-273A(V)/URN	EMERGENCY BREATHING SYSTEM	LONGITUDINAL STRUCTURAL BULKHEADS	SHIPBOARD LAN (SHIPLAN)
	ANTENNA GROUP, OE-5023/SLQ-501	ENGINE DIESEL	LP BLOW & VENTILATION EXHAUST SYSTEM	SHIPBOARD LAN (SHIPLAN) (GENERIC)
	ANTENNA GROUP, OE-5039/SPS-505	ENGINE DIESEL, PORT	LP COMPRESSED AIR SYSTEM	SHIPS FRESH WATER COOLING SYSTEM
	ANTENNA GROUP, OE-82C/WSC-1(V)	ENGINE DIESEL, STBD	LUBE OIL PURIFIER	SHIPS LOG OUTFIT SSL AGILOG
	ANTENNA GROUP, OE-82D/WSC	ENGINE GROUP, GAS TURBINE	LUBE OIL SYSTEM	SHIPS SEAWATER COOLING SYSTEM
	ANTENNA, AS-3263/SPS-49(V)	ENGINE, DIESEL (850KW GENERATOR SET)	MACHINERY CONTROL CONSOLE	SHORE CONNECTION BOXES
	ANTENNA, AS-5169/SRC	ENGINE, DIESEL (SEAWATER SERVICE SYSTEM)	MACHINERY SPACE VENTILTN/SNORT INDUC SYS	SHORE SUPPLY EQUIPMENT GROUP
	ARMAMENT, COMMANT & CONTROL FOUNDATIONS	ENGINE, GAS TURBINE	MACHINERY SPACES, MARINE SYSTEMS	SICK BAY EQUIPMENT GROUP
	AUTOMATIC CONTROL (PIELSTICK ENGINE)	EXHAUST & IR SUPPRESSION	MAIN GEAR TRANSMISSION ASSEMBLY	SIRIUS - LONG RANGE SEARCH/TRACK SYSTEM
	AUTOMATIC LUBRICATION SYSTEM	EXHAUST DUCTING	MAIN ENGINE, GEARING & SHAFTING SYSTEM	SLIDING TROLLEY BLOCK ASSEMBLY
	AUXILIARY SEAWATER CIRCULATING SYSTEM	EXHAUST SYSTEM	MAIN GUN SYSTEM, 57 MM	SLOW SPEED PROPULSION
	AV& B REDUCING STATIONS	EXHAUST SYSTEM (PIELSTICK ENGINE)	MAIN GUN SYSTEM, 57 MM, MK III	SNORT INDUCTION MAST
	AWNINGS,AWNINGS STANCHIONS & MISC COVERS	EXT HYD POWER PLANT ANC EQUIPMENT	MAIN HYD POWER PLANT ANCILLARY EQUIPMENT	SONAR
	BALLAST PUMPS & COMPENSATING UNITS	EXTERNAL COMMUNICATIONS EQUIPMENT	MAIN HYDRAULIC SYSTEM	SONAR 2041
	BATTERY COOLING SYSTEM	EXTERNAL HYDRAULIC POWER PLANT	MAIN LUBRICATING OIL SYSTEM	SONAR 2071 ADA SUB ACOUSTIC AUGMENTOR
	BATTERY SWITCHBOARD	EXTERNAL HYDRAULIC SYSTEM	MAIN NON-STRUCTURAL TANKS	SONAR DOME ASSEMBLY, 163 INCH DOME
	BATTERY VENTILATION	EXTERNAL WEATHER DECKS & PLATFORMS	MAIN PROPULSION EQUIPMENT	SONAR HULL OUTFIT C5 SYSTEM
	BILGE PUMPS & COMPENSATING UNITS	FAN SYSTEM, VANEAXIAL, MUC-1 & MCU-2	MAIN PROPULSION MOTOR COOLING	SONAR SET 2007AE
	BINOCULAR SYSTEM, MK 3, MOD 5 (BIG EYE).	FAN SYSTEM, VANEAXIAL,EF-9, EF-11, EF-13	MAIN PROPULSION SYSTEM (GENERIC)	SONAR SET 2040AA
	BLACK & GREY WATER COLLECTION SYSTEM	FILTRATION UNIT 1-1A, 4-4A	MAIN REFRIGERATION SYSTEM	SONOBUOY PROCESSING SYSTEM
	BLACK & GREY WATER SEWAGE TREATMENT PLNT	FILTRATION UNIT 2-3	MAIN SEAWATER CIRCULATING SYSTEM	SOUND PROOF CASING (PIELSTICK ENGINE)
	BLACK AND GREY WATER COLLECTION SYSTEM (	FILTRATION UNITS, NBC	MAIN SHAFTING SYSTEM (GENERIC)	SPEED LOG DOPPLER SONAR SYSTEM
	BLANKER-VIDEO MIXER SET AN/SLA-10B	FINE WATER SPRAY SYSTEM (FIRE SUPPRESSIO	MAIN STRUCTURE GENERAL	SSE SYSTEM AFT MK 9
	BOAT DAVIT SYS (RIGID INFLATABLE BOAT)	FIRE AND PREWET PUMP UNIT	MARINE ENGINEERING AUX EQUIP GROUP (GEN)	SSE SYSTEM FWD MK 9
	BOAT, INFLATABLE MAT, SEA RESCUE, 10 MAN	FIRE FIGHTING EQUIPMENT, PORTABLE	MARINE ENGINEERING AUXILIARY EQUIP GROUP	STABILIZED HORIZON REF SYS, AN/SJN-5-501
	BOATS HANDLING & STOWAGE	FIRE MAIN SYSTEM	MARINE SANITATION DEVICE	STEAM GENERATOR
	BOILER, AUXILIARY	FIREMAIN SYSTEM	MARINE SYSTEMS	STEERING GEAR AND CONTROL SYSTEMS
	BOTTLE ASSEMBLIES	FIXED EXTINGUISHING SYSTEMS	MASTS (GENERIC)	STOWAGE-HANDLING TOWED ARRAY OK-410 (V)1

BOUYANT WIRE AERIAL SYSTEM
BUILT-IN BREATHING SYSTEM
BULKHEADS
BUS ACCESS MODULES LOOPS
CAMERA SET, TELEVISION, AN/WXY-501
CANSCOT SHF SATCOM
CAPSTAN AFTER (MOORING TOWING ASSEMBLY)
CAPSTAN SYSTEM, AFT
CAPSTAN WINDLASS ASSY
CAPSTAN, WINDLASS SYSTEM
CASUALTY POWER SYSTEM (GENERIC)
CATHODIC PROTECTION SYSTEM (GENERIC)
CDN REC AND ASST SEC & TRAV SYSTEM
CENTERING SYSTEM
CHILLED WATER SYSTEM
CHILLER SET, 85 TON
CIWS MK 15 MOD 21 BLOCK 1B
CLOSE-IN WEAPON SYSTEM BLOCK 1B
CLOSE IN WEAPON SYSTEM MK15 MOD 21 BLOCK
CLOSED CIRCUIT TELEVISION SURVEILLANCE
CLOSURES, WATERTIGHT/GASTIGHT/OILTIGHT
COLD FRESH WATER SYSTEM
COMBAT SYSTEMS
COMMAND & CONTROL EQUIPMENT GROUP
COMMAND AND COMBAT SYSTEM EQUIP SPACES
COMMAND AND COMBAT SYSTEM EQUIPMENT SPAC
COMMAND AND CONTROL MISC EQUIP GROUP
COMMUNICATION SET, SONAR, AN/WQC-501(V)
COMMUNICATIONS CONTROL & MONITORING SYS
COMMUNICATIONS GROUP (EXTERNAL)
COMMUNICATIONS GROUP, EXTERNAL
COMMUNICATIONS SYSTEM, AN/SSC-510
COMPRESSED AIR SYSTEM, HIGH PRESSURE
COMPRESSED AIR SYSTEM, HP (GENERIC)
COMPRESSOR ASSEMBLY TYPE 5436N
COMPRESSOR UNIT, RECIPROCATING
COMPRESSOR UNIT, RECIPROCATING (HP AIR)
COMPRESSOR UNIT, RECIPROCATING (HP)
COMPRESSOR UNIT, ROTARY (LP AIR)
CONTINUOUS WAVE ILLUMINATOR SYSTEM
CONTROLLABLE PITCH PROPELLERS AND PROPEL
COOLER, FLUID, INDUSTRIAL, OA-5181/SP
COUNTERMEASURES SET, AN/SLQ-502
COUNTERMEASURES SET, AN/SLQ-502A
COUNTERMEASURES SET, AN/SLQ-503
CROSS CONNECT GEARING AND PLATFORM
CRUISE ENG ENCL COOLING AIR INTKE & EXHS
DAMAGE CONTROL SYSTEM
DATA / VIDEO / AUDIO COLLECTION
DATA LINK PROCESSING SYSTEM (DLPS)
DATA TERMINAL SET, P/O AN/USQ-125(V)
DAVIT, BOAT BOOM, RIGID INFLATABLE BOAT
DE-ICING SYSTEM
DECK & HULL EQUIPMENT GROUP
DECK AND HULL FITTINGS
DECK COVERING TREADMATS TREADSTRIPS ETC
DECK CRANE ASSEMBLY
DECK CRANE GENERAL PURPOSE (HIAB 61)
DECK, THREE (GENERIC)
DECK, TWO (GENERIC)
FLIGHT DECK LIGHTING SYSTEM (GENERIC)
FLT DECK CCTV
FOAM SYSTEM, AQUEOUS FILM FORMING
FORWARD MOORING/TOWING WINDLASS
FOUNDATIONS/SEATINGS (GENERIC)
FRESH WATER SYSTEM
FUEL FILLING AND COMPENSATING SYSTEM
FUEL OIL CENTRIFUGE
FUEL OIL FILLING & TRANSFER SYSTEM
FUEL OIL SERVICE SYSTEM
FUEL SYSTEM (PIELSTICK ENGINE)
FWD/AFT CAPSTAN & EQUIPMENT
FWSS SUPPRESSION SYSTEM
GALLEY & SERVERY EQUIPMENT GROUP
GARBAGE EJECTOR ASSEMBLY
GENERATOR AC
GENERATOR GROUP, 750 KW
GENERATOR TEST, TERMINAL BOXES
GENERATOR, AC (850 KW)
GLOBAL COMMAND AND CONTROL SYS MARITME
GOVERNOR (PIELSTICK ENGINE)
GUARD RAILS & STANCHIONS
GUN ASSEMBLY
HANDLING SYSTEM, TORPEDO
HANDLING SYSTEM, TORPEDO, SHIPBOARD
HATCHES, DOORS, MANHOLES
HEAVY MACHINE GUN SYSTEM, .50 CALIBRE.
HIGH SPEED DATA CONNECTIVITY (HSDC)
HMCS ATHABASKAN
HMCS CHARLOTTETOWN
HMCS CORNERBROOK
HMCS FREDERICTON
HMCS HALIFAX
HMCS IROQUOIS
HMCS MONTREAL
HMCS PRESERVER
HMCS REGINA
HMCS ST. JOHN'S
HMCS TORONTO
HMCS VANCOUVER
HMCS VILLE DE QUEBEC
HORIZON BAR SYSTEM, STABILIZED (GENERIC)
HP AIR REDUCING STATIONS
HP AIR SYSTEM
HP AIR VALVES & FITTINGS
HP COMPRESSED AIR SYSTEM
HP DIRECT BLOW SYSTEM
HPAC ARRANGEMENT
HPB VALVES
HULL INSULATION
HULL NON-PENETRATING MAST
HULL NON-PENETRATING MASTS
HULL PENETRATING MASTS, WEAPONS
HULL STRUCTURE
HULL STRUCTURE (GENERIC)
HULL SYSTEMS
HULL SYSTEMS & FITTINGS
HULL VALVES
HVAC SYSTEM
HYDRAULIC OIL FILLING & TRANSFER SYSTEM
MBT VENTING ARRANGEMENTS
MESSAGE HANDLING & DISTRIBUTION SYSTEM
METEOROLOGICAL SYSTEM
MF/HF ANTENNA SYSTEM AJU
MOISTURE SEPARATOR SYSTEM
MOTOR PROPULSION
MOTOR, ELECTRIC, 60/30 H.P.
NAVAL AEHF SYSTEM
NAVAL EXTERNAL COMMUNICATIONS SYSTEM
NAVAL FIREFIGHTING HOSES, NOZZLES
NAVAL INFORMATION SYSTEMS (NAVIS)
NAVAL INFORMATION SYSTEMS (NAVIS).
NAVIGATION DISTRIBUTION SYSTEM
NAVIGATION SET,TACAN (AN/URN-25)
NAVIGATION SYSTEM EQUIPMENT GROUP
NAVIGATION SYSTEM GROUP
NAVIGATION SYSTEM, RING LASER GYRO NAVIG
NAVIGATION SYSTEMS GROUP
NAVIGATIONAL LIGHTING
NERA SATURN BM MARINE MK2 INMARSAT
NON-STRUCTURAL BULKHEADS & CLOSURES
Not assigned
NULKA ACTIVE MISSLE DECOY SYSTEM
OFFICE SPACES (GENERIC)
OIL POLLUTION ABATEMENT AND BILGE STRIP
Other Location
PADEYE INSTALLATION
PLASTIC WASTE PROCESSOR
PORTLIGHTS & WINDOWS
POWER SYSTEM, AFTER CAPSTAN
PRIMARY AIR CONDITIONING PLANTS NO.1 & 2
PRIMARY ELEC POWER GENERATION & DIST SYS
PRIMARY SWITCH GEAR (GENERIC)
PRIMING SYSTEM
PROPULSION SYSTEM, MAIN
PUBLIC ADDRESS SET, AN/UIC-502
PUMP SYSTEM (MAIN LUBE OIL)
PUMP UNIT, CENTRIFUGAL (COLD FRESH WTR)
PUMP UNIT, CENTRIFUGAL (DIES FIRE PUMP)
PUMP UNIT, CENTRIFUGAL (FIRE PUMP).
PUMP UNIT, CENTRIFUGAL (JOCKEY, FIRE)
PUMP UNIT, F-401 (DFO)
RADAR SET AN/SPG-503
RADAR SET GROUP (LTWT RDR-ASSOP DIR SYS)
RADAR SET, AN/SPG-501
RADAR SET, AN/SPS-49A(V)5
RADAR SET, AN/SPS-502
RADAR SET, AN/SPS-505
RADAR SET, AN/SPS-506
RADAR SET, NAVIGATION, KH 1007
RADIO SET, AN/SRC-513
RADIO SET, AN/SRC-515
RADIO SET, HF AN/URC-507
RADIO SET, MARINE, GMDSS DSC CLASS D VHF
RAILS, STANCHIONS & LIFELINES
RAPID SECURING DEVICE
RATIONALISED INTERNAL COMMUNICATIONS
RECEIVER, RADIO, R-5095B/G.
REFRIGERATION SYSTEM FOOD
REPLENISHMENT-AT-SEA DECK MACHINERY
STRUCTURAL BULKHEADS (GENERIC)
STRUCTURAL TANKS & ENCLOSURES
SUB ESCAPE & RESCUE, LIFE SAVING SUPPORT
SUBMARINE CONTROL CONSOLE
SUBMARINE ESCAPE EQUIPMENT
SUBMARINE PRESSURE HULL
SUBMERGED SIGNAL EJECTOR MK9 SYSTEMS SSE
SUPERCHARGING (PIELSTICK ENGINE)
SUPERSTRUCTURE
SUPPLY TO MAIN MOTOR AIR COOLERS
SURFACE & AIR WEAPONS SYSTEM
SWITCHBOARD PROPULSION
SWITCHBOARD, MAIN & ASSOCIATED EQUIPMENT
SWITCHBOARDS, FORWARD & AFTER
SYS, 24V DC DISTRIBUTION
SYSTEM VALVES AND FITTINGS
SYSTEM, ATTACK PERISCOPE
SYSTEM, SEARCH PERISCOPE
TEMP EC - COMMAND & CONTROL EQUIPMENT
TEMP EC - COMMUNICATIONS GROUP
TEMP EC - COMMUNICATIONS GROUP, EXTERNAL
TEMP EC - HULL SYSTEMS
TEMP EC - NAVAL EXTERNAL COMMUNICATIONS
TEMP EC - NAVAL INFORMATION SYSTEM NAVIS
TEMP EC - UNDERWATER COMBAT SYSTEM
TORPEDO HANDLING & STOWAGE SYSTEM
TORPEDO LAUNCHING SYSTEM
TORPEDO SETTING PANEL SYSTEM
TORPEDO TUBE ASSEMBLY, MK 32, MOD 9
TOWED ARRAY SYSTEMS
TOWED ARRAY, SHIPBOARD SESS2-1
TOWER ESCAPE ONE MAN FWD & AFT (STRUCT)
TPS-250 PORTABLE VHF/AM TRANSCEIVER SYS
TRANSMITTER GROUP MK 73 MOD 3
TRANSMITTING SET, COUNTERMEASURES
TRANSMITTING SET, RADIO AN/SRT-504
TRIM SYSTEM
TRIM SYSTEM VALVES
TRIM, BILGE, BALLAST PRIMING SYSTEMS
TURBINE, STEAM(FOR 1000 KW TURBINE GNRT)
UNDERWATER COMBAT SYSTEM
UNDERWATER DETECTION SYSTEM
VALVES & FITTINGS NITROGEN CHARGING SYST
VENT OVERFLOW AND SOUNDING SYSTEM
VENTILATION SYSTEM
VERTICAL CAPSTAN - LINE HANDLING WINCH
VERTICAL STORES CONVEYOR
VHF/UHF AM/FM RADIO SYSTEM AN/PRC-117F
VPDC DISTRIBUTION
WASHPLACES AND SHOWERS
WATER CIRCUIT (PIELSTICK ENGINE)
WATER PURIFICATION UNIT SKID MOUNTED
WEAPONS DISCHARGE SYSTEMS (WDS)
WEAPONS GROUP MK 16 MOD 3
WEATHER DECK ARRANGEMENTS
WEATHER DECK STOWAGES & LOCKERS
WINCH, LINE HANDLING
WIPER ASSEMBLY,WINDOW

Green (3)

AIR START SYSTEM
ANTENNA AS-5171/SRC
ANTENNA GROUP OE-5035/SR
BATHYTHERMOGRAPH SYSTEM, AN/WQQ-502 (V)
BEACON, RADIO, EMERGENCY POSITION INDICA
BOAT, RIGID INFLATABLE, ALUMINUM HULL
BOAT, RIGID, INFLATABLE, 24FT.
COOLER UNITS, AIR (KU 3000) (GENERIC)
COOLER UNITS, AIR (TYPE KU) (GENERIC)
DETECTOR, WIND DIRECTION AND SPEED
DRIER, AIR-GAS, DESICCANT, HD-5029/SRC
FAN COIL UNITS (GENERIC)
FAN SYSTEM, VANEAXIAL, SF-10 & SF-11
FUEL OIL CENTRIFUGE ASSEMBLY
RADAR SET
SUBMARINE INDICATOR BUOY (SARSAT)
WHISTLE SYSTEM, ELECTRIC (PORT & STBD)

Purple (4)

ACCOMODATION SPACES
AIR CONTROL SYSTEM (LOW PRESSURE)
ANTENNA COUPLER GROUP
BOILER, HIGH PRESSURE, AUXILIARY
BOILER, MAIN STEAM, HP
BREATHING APPARATUS, OXYGEN GENERATING,
BRIDGE FIN, CASING & EXTERNAL STRUCTURES
CENTRAL SURVEILLANCE SYSTEM
COMPUTER SYSTEM, INTEGRATED SHIPBOARD
CONTROL, SERVO CONTROLLABLE PITCH PROP
DAMAGE CONTROL MONITORING & ALARM SYS
DECK, ONE (GENERIC)
ELECTRICAL PROPULSION MACHINERY
ELECTRONIC STEERING CONTROL SYSTEM
ENGINE DIESEL, (850 KW GENERATOR SET)
ENGINE, DIESEL
ENGINE, GAS TURBINE, MAIN AND CRUISE
ENGINE, GAS TURBINE, MODEL 570 KF
EVAPORATION & DISTILLATION PLANT
FUTURE EXPANSION
GALLEYS, SERVERYS AND DINING AREAS
GALLEYS, SERVERYS AND DINING AREAS (GEN)
GENERATOR SET, DIESEL ENGINE (1000 KW)
GUN MOUNT, SUPER RAPID, 76/62 MILLIMETRE
GYRO COMPASS SYSTEM
HALON FIRE SUPPRESSION SYSTEM
HELICOPTER HAULDOWN SYSTEM
HMCS ALGONQUIN
HMCS CALGARY
HMCS CHICOUTIMI
HMCS OTTAWA
HMCS PROTECTEUR
HMCS WINDSOR
HMCS WINNIPEG
HULL
INTEGRATED PLATFORM MANAGEMENT SYSTEM
LAUNCHER SYSTEM GUIDED MISSILE VERTICAL
LAUNCHING SYSTEM, GUIDED MISSILE
LIGHTING GENERAL
LM2500 GAS TURBINE ASSEMBLY
LOBBIES AND PASSAGEWAYS
LOCKERS, AMMUNITIONS
LUBE OIL SYSTEM (PIELSTICK ENGINE)
MAIN BATTERY SYSTEM
MASS (MULTI-AMMUNITION SOFTKILL SYSTEM)
MEDIUM RANGE RADAR, AN/SPQ-501
MINOR STRUCTURE (GENERIC)
MODEM COMMUNICATIONS MD-5077(V)3/U
NAVIGATION SYSTEMS EQUIPMENT GROUP
PASSIVE RANGING SONAR
RADIO SET, AN/SRC-514
RADIO SET, AN/SRC-516
RADIO SET, TACAN, AN/URN-20A(V)
RECEIVING SET, RADIO AN/SRR-505
RECOVERY ASSIST SECURE TRAVERSE SYSTEM (
REPLENISHMENT-AT-SEA SYSTEMS
SCBA SYSTEM
SCUTTLES, WINDOWS & ASSOCIATED EQUIPT
SECONDARY ELEC POWER GENERTN & DIST SYS
SECURE LAN (SECLAN)
SHAFT AND PROPELLER
SHINCOM 3100
SHIP ARRANGEMENTS
SHIPS FURNITURE, NON-STANDARD
SOLID WASTE SYSTEM
SONAR TYPE 2008 UWT, VARIANT NG
STOWAGE & HANDLING SPACE, AMMUNITION
STOWAGES AND LOCKERS
SUBMARINE FIRE CONTROL SYSTEM (SFCS)
SUBMARINES,HMCS VICTORIA
SUPERSTRUCTURE (GENERIC)
TEMP EC - AIRCRAFT SUPPORT EQUIPMENT
TEMP EC - DECK & HULL EQUIPMENT GROUP
THERMAL INSULATION & ACOUSTIC DAMPING (V
TORPEDO WARSHOT, MK48, MOD 4
TOWED ARRAY GROUP, SONAR, 0A-9056 (V)2
TUBE CONTROL & INDICATOR
TURBINE, GAS
UNDERWATER DETECTION EQUIPMENT
UNDERWATER WEAPONS EQUIPMENT GROUP
VERTICAL LAUNCHING SYSTEM, MK 41 MOD C
WATER CHILLER UNIT, 125 TON CAPACITY
WEAPONS HANDLING SYSTEM (WHS)

4 Hidden Markov Models for Monthly Order Key Completions

The FMF creates task assignments based on unique order keys. Each order key can have subtasks and the FMF records an order key as complete once all its subtasks are finished. The elapsed time does not reflect the actual work hours committed to the task. Order keys have priority rankings during maintenance activities which can result in a long order key lifetime even though the task may require only a small amount of time to complete. An order key appears complete in the system only once the FMF closes the ticket—the actual work may finish well before the close date, depending on the filing practice of the FMF. Thus, order key completion records depend on the actual time need to compelete the task(s), the priority of the work, and the business close out procedure. The lifetime of an order key does not not simply reflect the required work effort.

The client needs an understanding of patterns in the order key completions by vessel and by FMF. The data contain 43,731 uncensored order key completions on Royal Canadian Navy (RCN)'s vessels (see Table 5) with start dates ranging from 14 November 2003 to 6 July 2016. The dataset contains work from both the Cape Scott and Cape Breton fleet maintenance facilities. From the raw data, I count the number of order key completions per month at each FMF. Figures 15 and 16 shows the monthly completed time series data. Notice that Figure 15 shows a downward, approximately linear, trend while Figure 16 does not. In Figures 17 and 18 we see the evidence of autocorrelation with a one month time lag—shocks to the system have a memory of one month.[5] Approximating the job count data as continuous leads to an autoregressive time series model [6] with one lag (an Autoregressive (AR)(1) model). While in principle the AR(1) models provide a predictive framework for monthly job completions at each FMF, the downward linear trend in the Cape Scott data will eventually end. The nature of the stochastic process may change at the inflection point. If the FMF planners are confident that the trend will continue, the time series model can help predict the monthly job completions. Interpretation remains a difficult problem.

At the individual ship level, we can no longer ignore the discrete nature of the order key completion counts. Simple continuous time series modelling will not work with the count data. To account for the autocorrelation in the data while respecting the discreteness of the

[5] I computed the Cape Scott Autocorrelation function (ACF) by using the residuals after removing the linear trend. The ACF shows some evidence of autocorrelation at 12, 16, and 19 month lags but in building time series models the AIC prefers a one lag model.

count data, I model each ship as a Poisson HMM [7]. Examination of the AIC and pseudo-residuals, shown in Figures 19 and 20, points to a five hidden Poisson rates associated with each ship. The pseudo-residuals concord with their theoretical quantiles while exhibiting no autocorrelation. The five dimensional rate vectors point in the positive orthant above the 45-degree hyperplane (the rate vectors form order tuples). K-means clustering (see [1] for details) of the rate vectors find three clusters and Figure 21 shows the best two dimensional representation (based on rank reduced linear discriminant analysis, see [1] for details) of the five dimensional clusters. We see that the HMM finds the coast to which each vessel belongs. Each FMF completes order keys differently but the HMCS Victoria sits alone in her own cluster. Given the teething problems with the Victoria Class, it is not surprising that one of the RCN's submarines has an entirely different work flow pattern compared to the rest of the RCN's vessels. Figure 22 shows the monthly completion job counts for each ship with cluster colour and rate memebership indicated.

The time series analysis demonstrates that the maintenance system has memory and that each FMF faces different challenges. Interpretation remains problematic—we do not know the reason for the coastal differences in order key completion rates. Order key completion records occur on the date at which the FMF closes a ticket. It might be the case that each FMF closes tickets based on different policies. It might also be the case that each FMF faces different challenges associated with each vessel or with its workforce. Given the downward trend in Figure 15 relative to Figure 16, there are differences between the two facilities. The time series analysis shows that we can identify each FMF by observing historical work patterns at the individual vessel level.

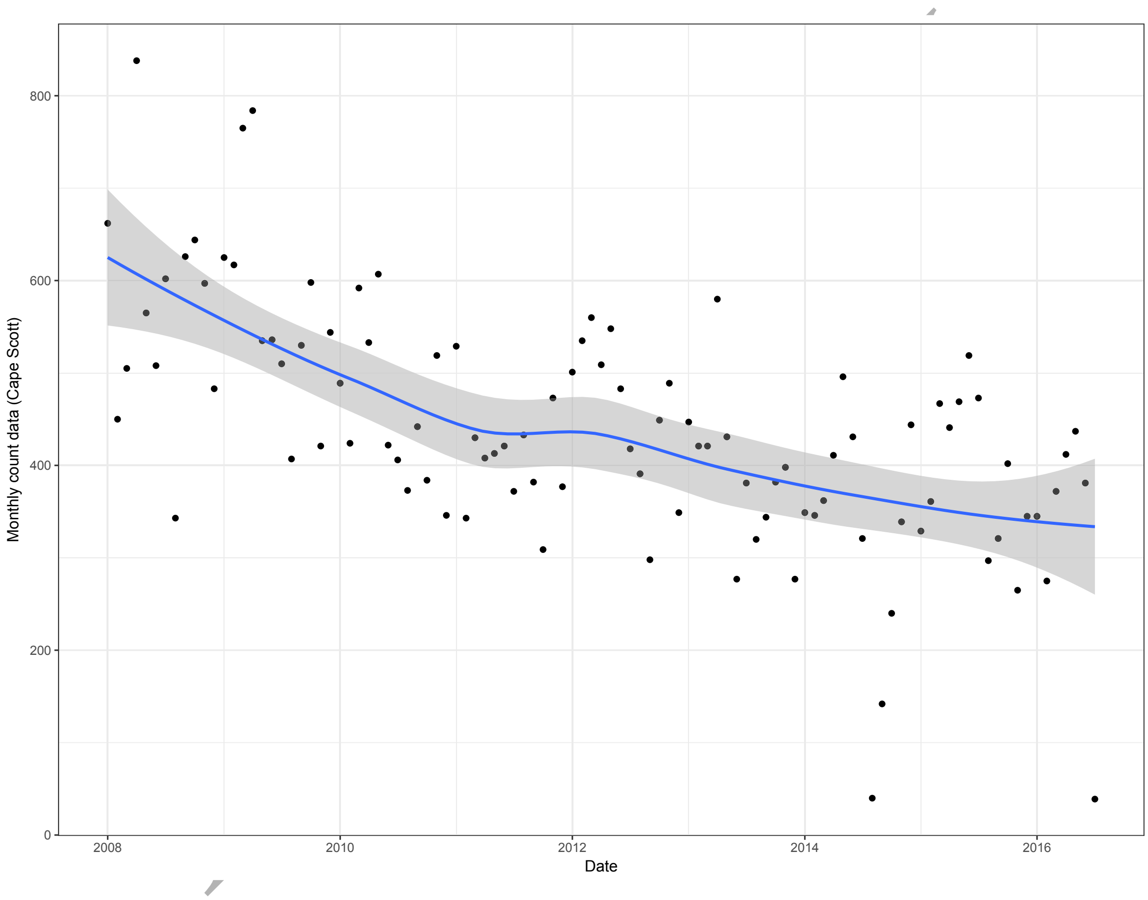

Figure 15: Cape Scott monthly completion time series with smoothing spline and error bar. Notice the presence of an approximately linear trend.

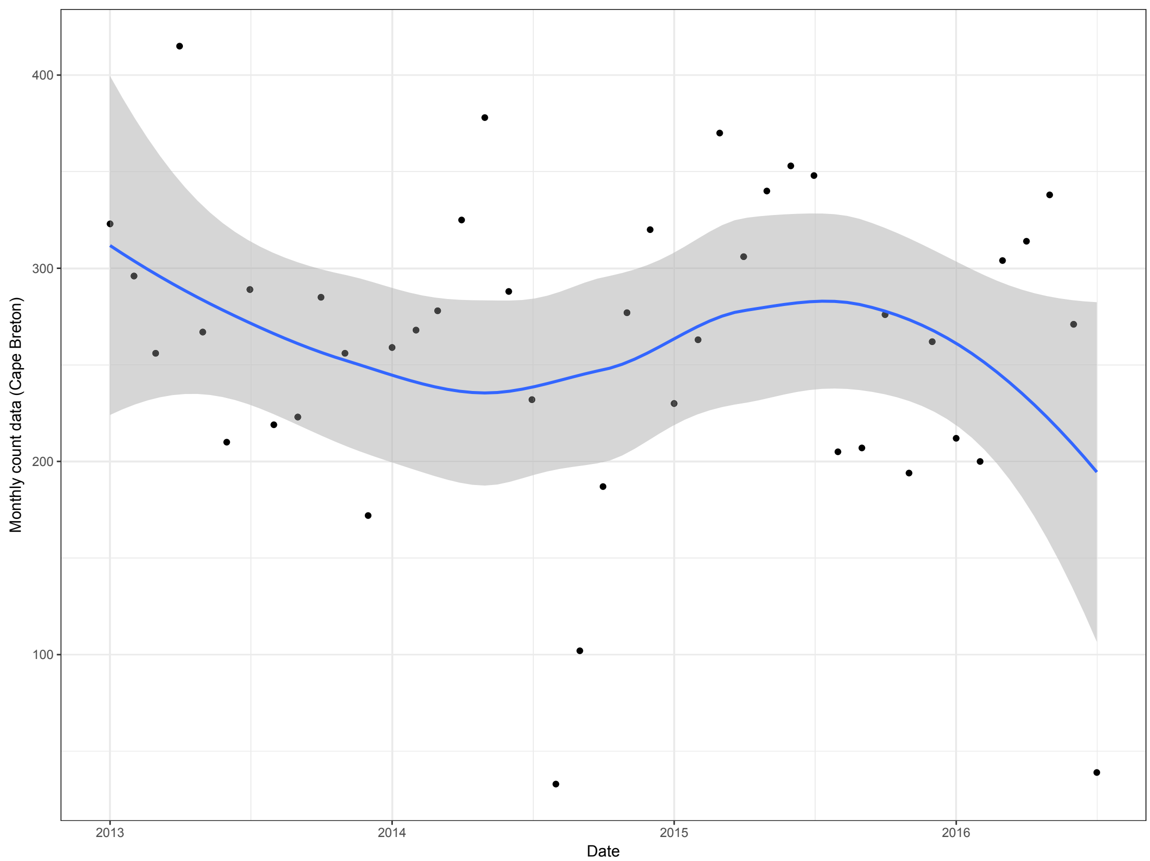

Figure 16: Cape Breton monthly completion time series with smoothing spline and error bar. Notice the absence of any trend.

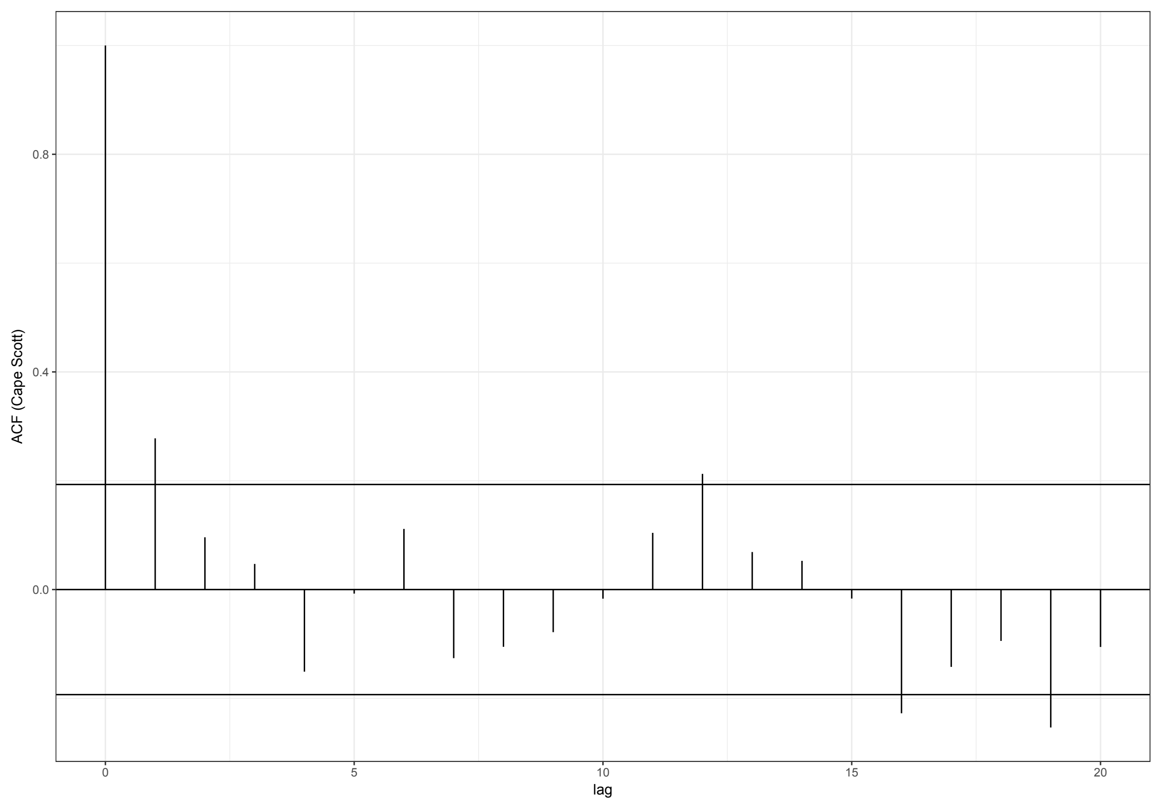

Figure 17: Cape Scott autocorrelation of monthly completions. The horizontal lines indicate the 95% confidence limits.

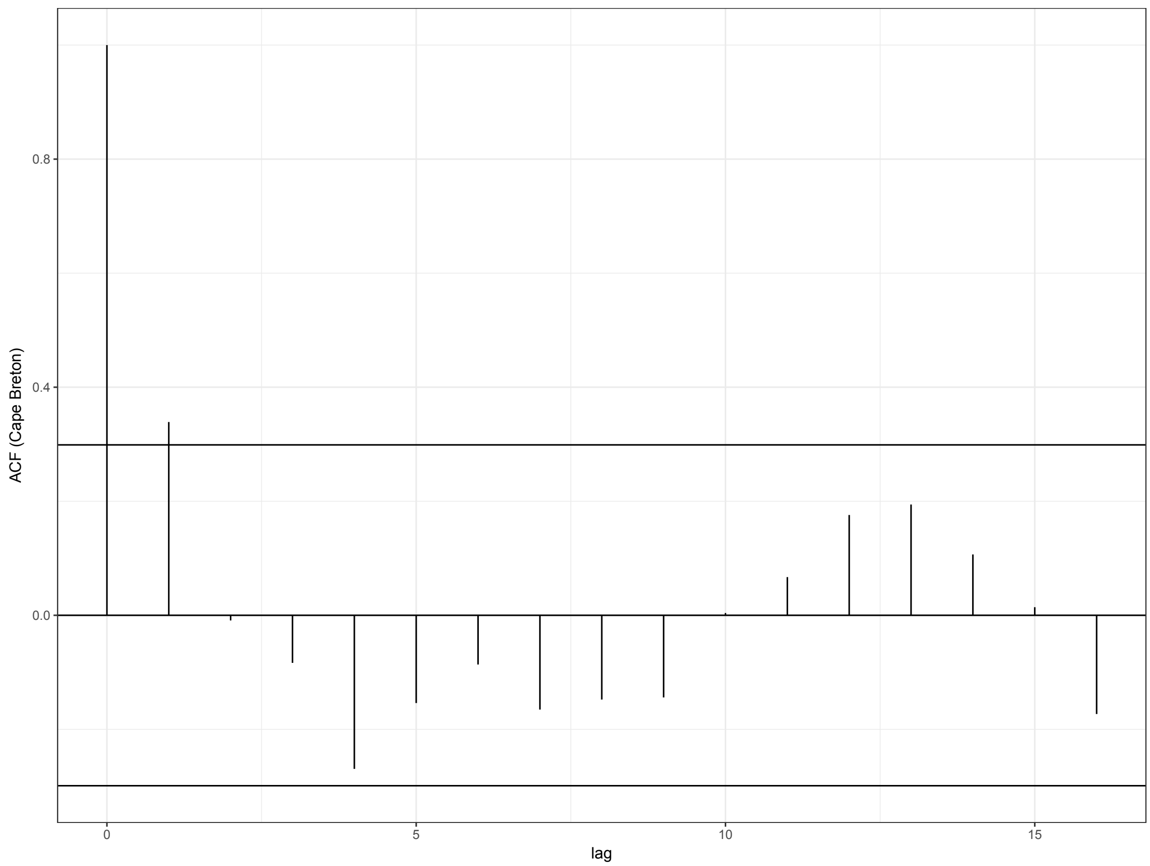

Figure 18: Cape Breton autocorrelation of monthly completions. The horizontal lines indicate the 95% confidence limits.

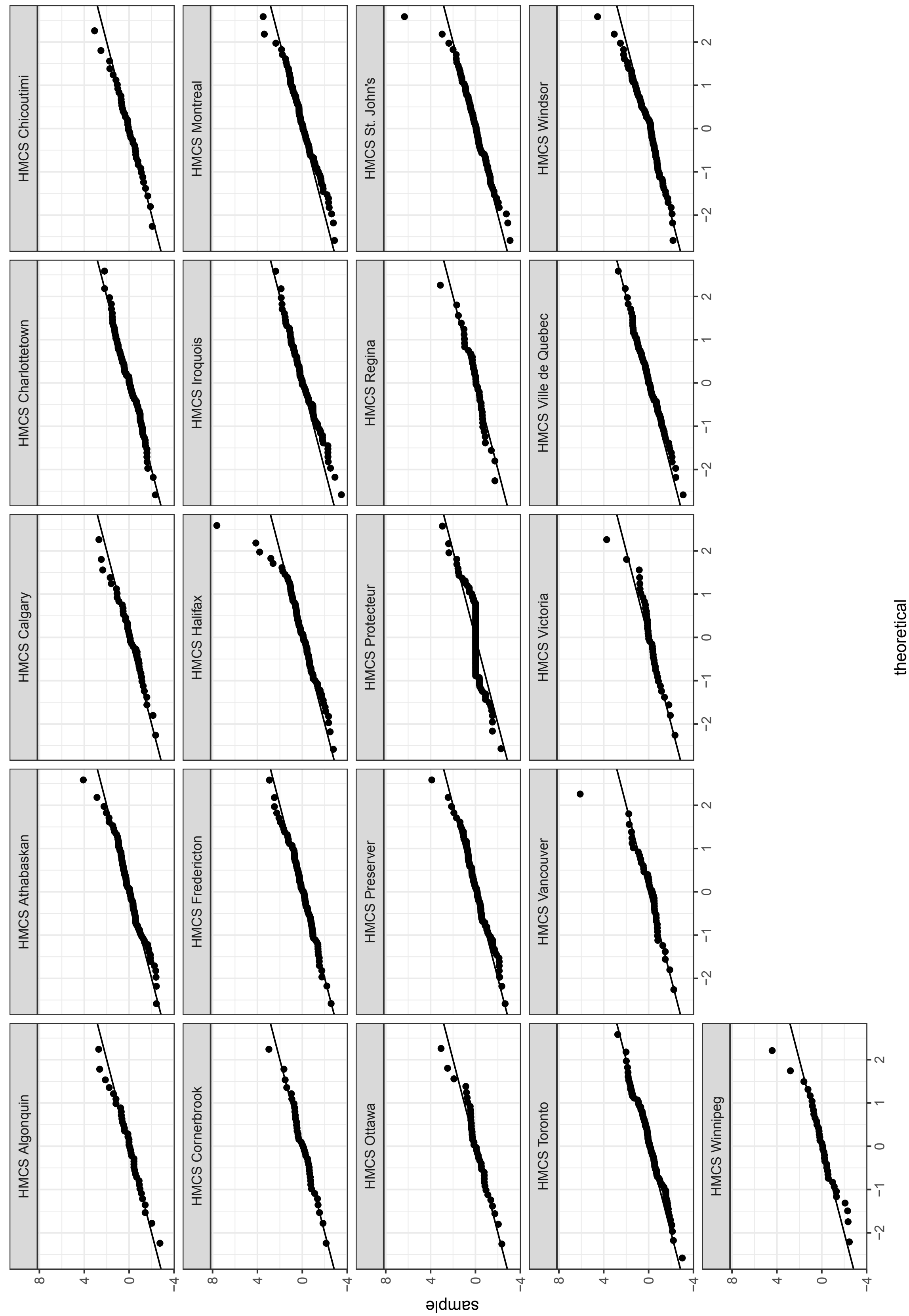

Figure 19: Pseudo-residual quantile analysis of the HMM estimation with 5 hidden states.

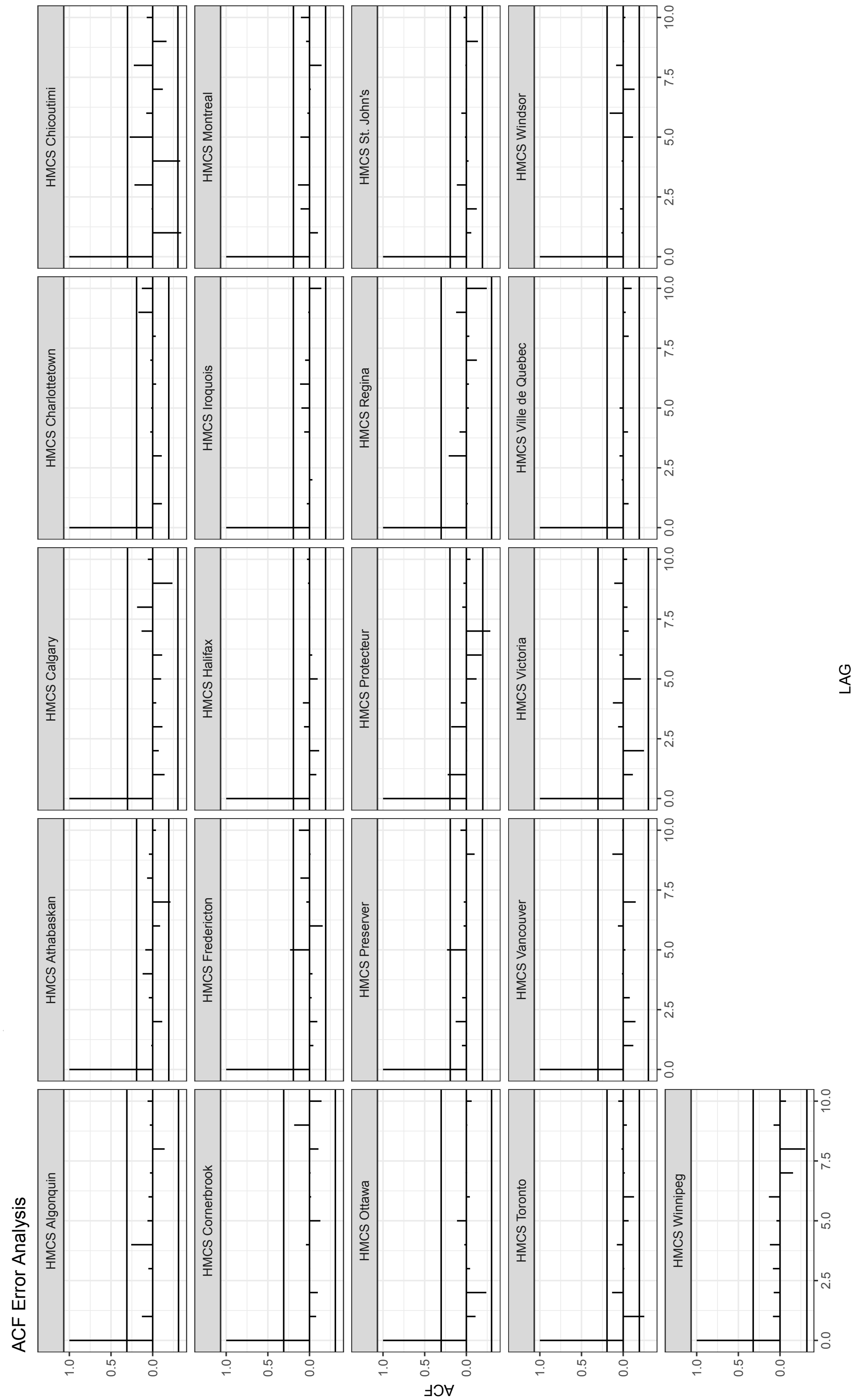

Figure 20: Pseudo-residual autocorrelation analysis of the HMM estimation with 5 hidden states. No evidence of serial correlation in the pseudo-residuals.

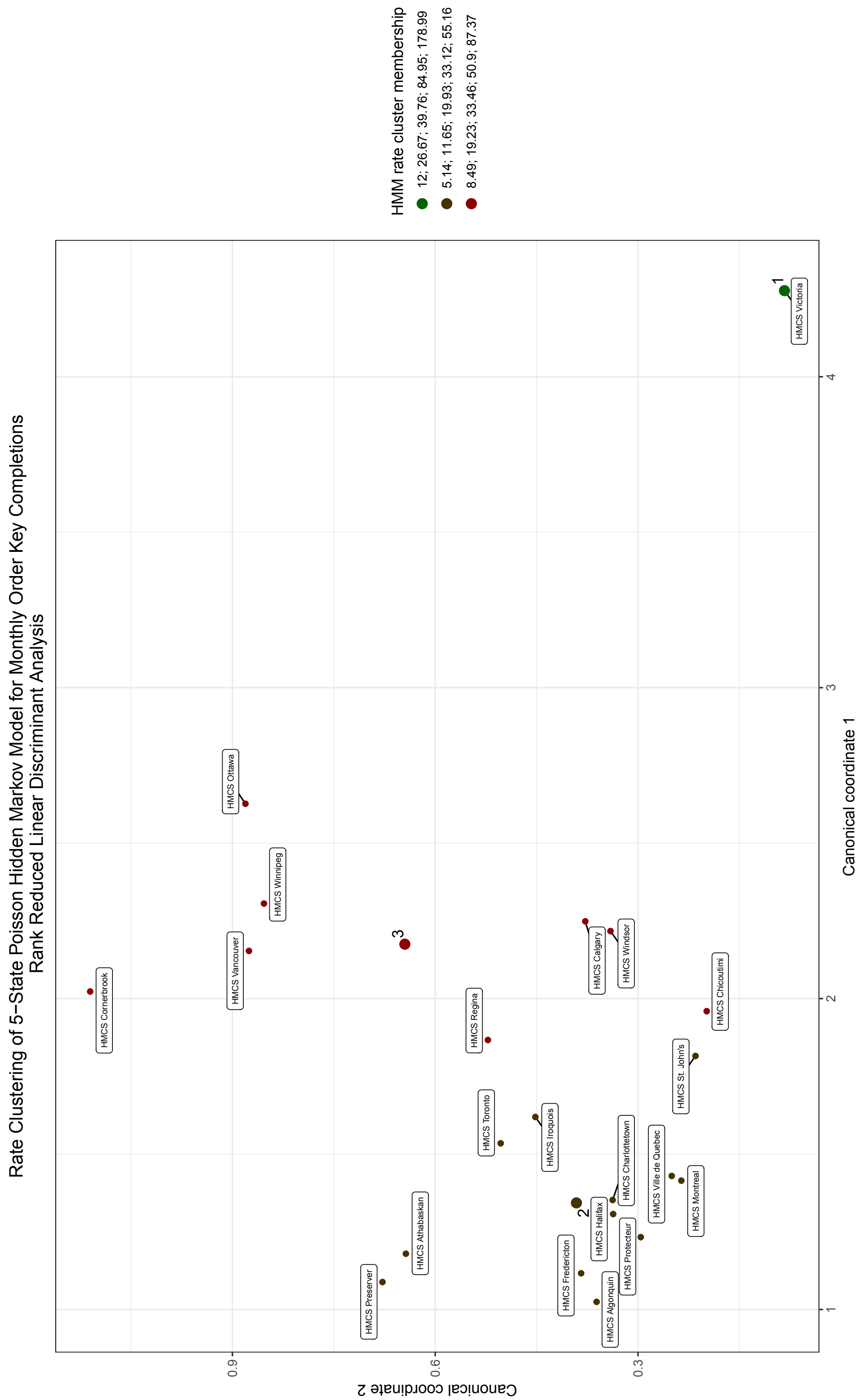

Figure 21: Clustering of the five dimensional rate vectors projected into the best two dimension plane by rank reduction. Cluster center rates are given in the legend.

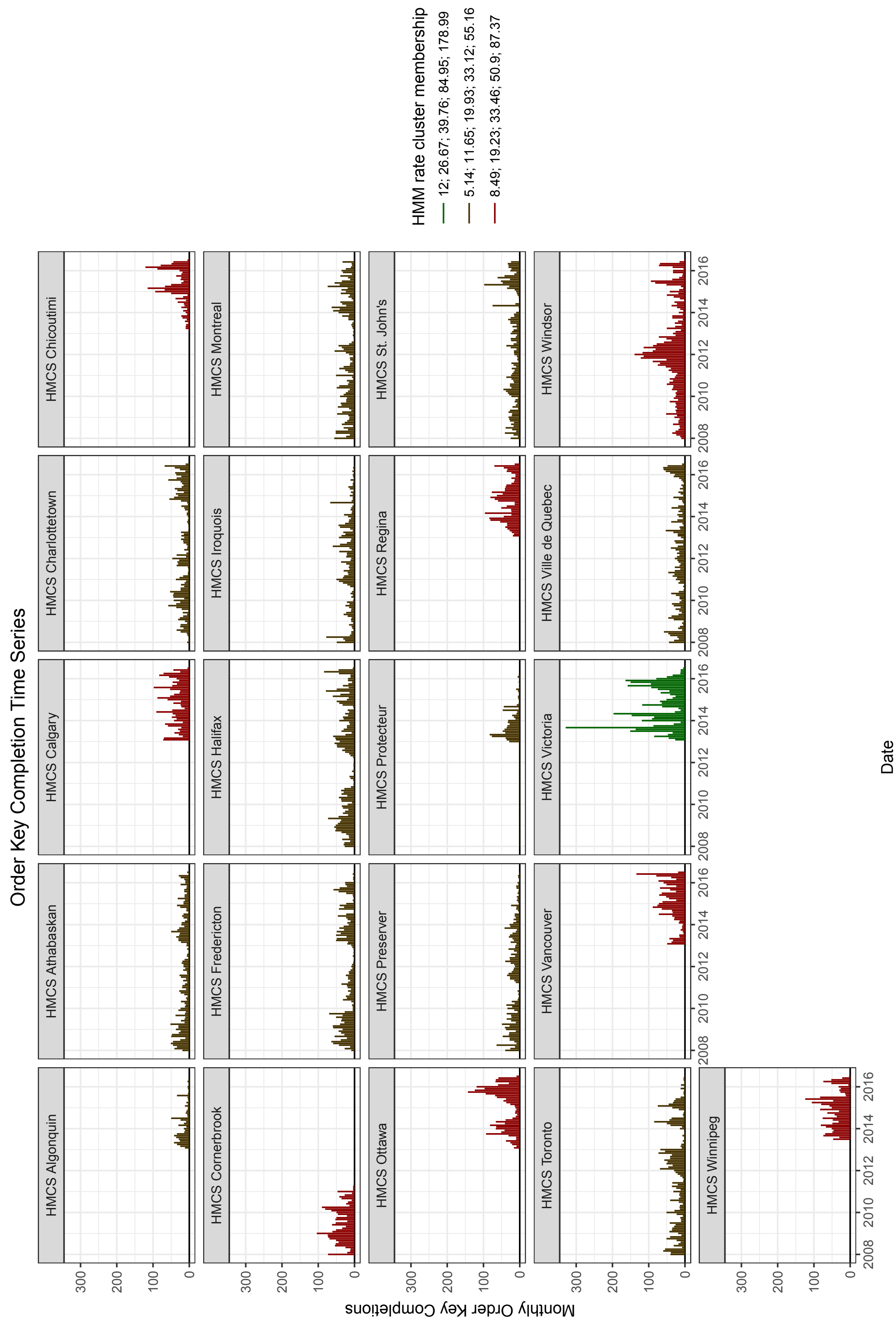

Figure 22: Time series counts by vessel with clustering.

5 Discussion

The regression tree provides the client with a predictive model. Given a task with covariate features, the model assigns the task to a terminal node. Using the quantile information within each terminal node, the planner can better understand the risk of schedule creep. The planner can also use the model to validate technician built schedules by comparing planned hours to the terminal node quantiles (e.g., in a 100 job task schedule, roughly 5 tasks should have planning time estimates that exceed the 95% quantile.)

The client can easily incorporate this tree model as an R script in the RCN's SAP Business Object framework. The full implementation to date includes three trees based on three different feature clustering assignments and all the trees can be used together to predict actual hour quantiles through a convex sum. The regression analysis places work centers at the top of the feature importance list.

The time series analysis demonstrates that the maintenance system has memory and that each FMF faces different challenges. Interpretation remains problematic—we do not know the reason for the coastal differences in order key completion rates. Order key completion records occur on the date at which the FMF closes a ticket. It might be the case that each FMF closes tickets based on different policies. It might also be the case that each FMF faces different challenges associated with each vessel or with its workforce. Given the downward trend in the Cape Scott monthly completion data relative to Cape Breton, there are differences between the two facilities. The time series analysis shows that we can identify each FMF by observing historical work patterns at the individual vessel level through an HMM construction.

This paper shows the RCN the power that data science can bring to operational data. The RCN is uniquely positioned to take advantage of statistical learning techniques in its new SAP environment. The demonstration of these methods in this paper is more important that any single result displayed. The RCN needs far more than one off results generated by CORA on task request basis. To bear fruit, the RCN requires a dedicated data science team that works in conjunction with SAP data maintainers. We have an enormous opportunity to fully exploit the SAP ERP capability, which will lead to operational treasure in the RCN's datasets. We are at the beginning of a journey of marrying data science and operational defence data, not just within the RCN but across all of DND.

Abbreviations and Acronyms

ACF Autocorrelation function

AIC Akaike Information Criteria

AIS Automatic Identification System

AR Autoregressive

BEx Business Explorer

BIC Bayesian Information Criteria

CART Classification and Regression Tree

CSV Comma Separated Variable

CORA Centre for Operational Research and Analysis

DNSM Directorate of Naval Strategic Management

DRMIS Defence Resource Management Information System

DND Department of National Defence

DRDC Defence Research and Development Canada

ERP Enterprise Resource Planning

FMAS Financial and Managerial Accounting System

FMF Fleet Maintenance Facility

GBM Gradient Boosted Model

GMM Gaussian Mixture Model

HMM Hidden Markov Model

LAH Log Actual Hours

MASIS Materiel Acquisition and Support Information System

OOB Out-of-bag

RCN Royal Canadian Navy

References

[1] Hastie, T., Tibshirani, R., and Friedman, J. (2008), The Elements of Statistical Learning, Springer.

[2] Rokach, L. and Maimon, O. Z. (2014), Data Mining with Decision Trees Theory and Applications, World Scientific.

[3] R Core Team (2016), R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.

[4] Therneau, T., Atkinson, B., and Ripley, B. (2015), rpart: Recursive Partitioning and Regression Trees. R package version 4.1-10.

[5] Fraley, C., Raftery, A. E., Murphy, T. B., and Scrucca, L. (2012), mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.

[6] Hamilton, J. (1994), Time Seris Analysis, Princeton.

[7] Zucchini, W. and MacDonald, I. L. (2009), Hidden Markov models for time series : an introduction using R, Monographs on statistics and applied probability, Boca Raton: CRC Press. A Chapman & Hall book.

[8] Cappé, O., Moulines, E., and Ryden, T. (2005), Inference in Hidden Markov Models (Springer Series in Statistics), Secaucus, NJ, USA: Springer-Verlag New York, Inc.

[9] Harte, D. (2015), HiddenMarkov: Hidden Markov Models, Statistics Research Associates, Wellington. R package version 1.8-4.

[10] Ridgeway, G. (2017), gbm: Generalized Boosted Regression Models. R package version 2.1.3.

This page intentionally left blank.

DOCUMENT CONTROL DATA

*Security markings for the title, authors, abstract and keywords must be entered when the document is sensitive

1. ORIGINATOR (Name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or tasking agency, is entered in Section 8.)

DRDC – Centre for Operational Research and Analysis
Defence Research and Development Canada
Carling Campus, 60 Moodie Drive, Building 7S.2
Ottawa, Ontario K1A 0K2
Canada

2a. SECURITY MARKING (Overall security marking of the document including special supplemental markings if applicable.)

CAN UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A

3. TITLE (The document title and sub-title as indicated on the title page.)

Predictive Analytics for the Royal Canadian Navy Fleet Maintenance Facilities: An application of data science to maintenance task completion times

4. AUTHORS (Last name, followed by initials – ranks, titles, etc., not to be used)

Maybury, D.

5. DATE OF PUBLICATION (Month and year of publication of document.)

December 2018

6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)

56

6b. NO. OF REFS (Total references cited.)

10

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter.)

Reference Document

8. SPONSORING CENTRE (The name and address of the department project office or laboratory sponsoring the research and development.)

DRDC – Centre for Operational Research and Analysis
Defence Research and Development Canada
Carling Campus, 60 Moodie Drive, Building 7S.2
Ottawa, Ontario K1A 0K2
Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. DRDC PUBLICATION NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)

DRDC-RDDC-2018-R150

10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

Data Analytics; Classification and Regression Trees; Data Science; Gaussian Mixture Models; Hidden Markov Models

13. ABSTRACT (When available in the document, the French version of the abstract must be included here.)

13. ABSTRACT (When available in the document, the French version of the abstract must be included here.)

In November 2016, Cdr Timothy Gibel, Directorate of Naval Strategic Management (DNSM) requested the Centre for Operations Research and Analysis's (CORA) assistance with creating a predictive model for actual work hours per task on Royal Canadian Navy (RCN) vessels at the Cape Scott and Cape Breton Fleet Maintenance Facilities (FMF). He also requested help in finding patterns in maintenance task completions at the vessel level. The FMF provided dataset contains information on 132,292 unique tasks separated into 43,731 order keys. Using regression trees with feature engineering for actual work hours, I find that a simple 14 terminal node tree explains 18% of the variance in the data. Gradient boosted stumps explain as much as 25% of the variance, but at the expense of an interpretable structure for schedule validation. Hidden Markov Modelling of monthly order key completion time series data reveals coastal differences between the CN's two maintenance facilities, providing an objective motivation to discover the source.

....

En novembre 2016, le cdr Timothy Gibel, de la Direction de la gestion stratégique de la marine (DNSM), a sollicité l'aide du Centre de recherche et d'analyse opérationnelles (CORA) pour créer un modèle prédictif des heures de travail réelles par tâche sur les navires de la Marine royale canadienne (MRC) des installations de maintenance de la flotte au Cap Scott et Cape Breton (FMF). Il a également demandé de l'aide pour trouver des tendances dans l'achèvement des tâches d'entretien au niveau des navires. L'ensemble de données fourni par FMF contient des informations sur 132 292 tâches uniques, séparées en 43 731 clés d'ordre. En utilisant des arbres de régression avec l'ingénierie des caractéristiques pour les heures de travail réelles, je trouve qu'un simple arbre de 14 nœuds terminaux explique 18% de la variance dans les données. Les *Gradient boosting stumps* expliquent jusqu'à 25% de la variance, mais aux dépens d'une structure inter-simulable pour la validation de la cédule. La modélisation de Markov cachée des données des séries chronologiques d'acheminements mensuels des principales commandes révèle les différences côtières entre les deux installations de maintenance de la MRC, fournissant ainsi une motivation objective à la recherche de la source.