Population-Level Modelling for Truck Fleet Survival Analysis

L.A. Bull[1], M. Dhada[2], O. Steinert[3], T. Lindgren[4], A.K. Parlikad[2], and
M. Girolami[1,5]

[1] The Alan Turing Institute, The British Library, London, NW1 2DB, UK
[2] Institute for Manufacturing, Department of Engineering, University of Cambridge, UK
[3] Strategic Product Planning and Advanced Analytics, Scania CV, Scania AB (publ),
SE-151 87 Södertälje, Sweden
[4] Department of Computer and Systems Sciences, Stockholm University, P.O. Box 7003,
SE-164 07 Kista, Sweden
[5] Department of Engineering, University of Cambridge, United Kingdom, UK, CB3 0FA

Abstract. Population-level modelling is used to address issues of data sparsity in the survival analysis of a simulated truck fleet. Specifically, hierarchical Bayes with mixed effects improves the predictive capability of hazard models. A set of correlated functions is learnt over the vehicle population – in a combined model – to approximate fleet predictors. Model uncertainty is reduced when sub-fleets of vehicles are allowed to share correlated information. In turn, vehicle groups with incomplete data (automatically) borrow statistical strength from *data-rich* groups.

Keywords: Hierarchical Bayes · Mixed Effects Model · Multi-Task Learning · Uncertainty Quantification · Asset Management

1 Population-level models for SHM

Data sparsity issues can be significant in practical applications of Structural Health Monitoring (SHM) – especially for systems recently in operation [17, 4]. Typically, there is little information *a priori*, instead, data arrive incrementally throughout use. This fact motivates the idea of *sharing* information between similar systems; specifically, can systems with comprehensive data (or established models) support those with incomplete information.

These concepts have led to the development of population-based SHM [3, 11, 9]. Existing work mostly investigates *similarity* measures between systems [11] and develops tools for the *transfer* of data and/or models from source to target structures

[13, 2, 8]. The work here extends population-level modelling [6], whereby inferences are made from collected population/fleet data in a combined model.

The case study presented here considers survival analysis of simulated data based on a truck fleet dataset provided by Scania. A population-level model is learnt using hierarchical Bayes with mixed effects, providing robust predictions compared to independent models learnt from sub-groups of vehicles. The *multi-task learning* approach [14, 15] automatically shares information between correlated domains (sub-fleets of assets).

2 Survival Analysis of truck fleets: Hazard Curves

Survival analysis is critical for evaluating the time to failure of industrial assets, and therefore fundamental while designing a maintenance plan. It involves studying failure occurrences in a population of assets (here trucks) over some time period. The period must be sufficiently long, such that reliability functions can be evaluated based on the observed failures or drop-outs [1].

This paper focuses on the log-hazard function evaluated for a large fleet of trucks. Hazard is defined as the instantaneous rate of failure of an asset. In other words, the hazard is the probability of a truck failing at time t, given that it has survived until time t [1]. During survival analysis, it is calculated as the fraction of trucks that failed in a given time interval to the number of trucks that survived until that interval. Mathematically, hazard at time t (i.e. $\lambda(t)$) is:

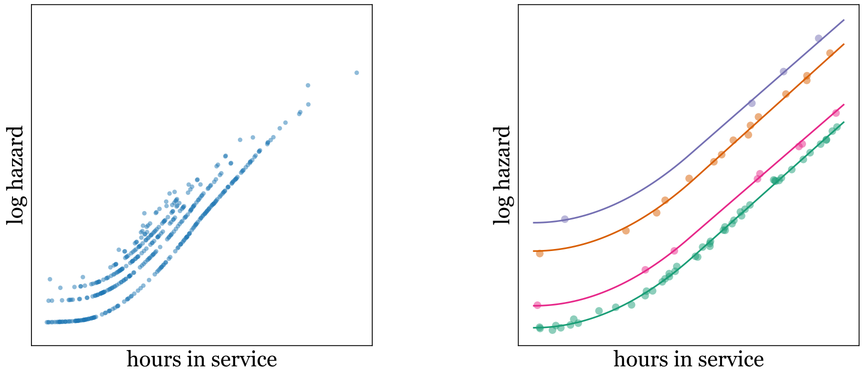
$$\lambda(t) = \frac{P(t \leq T < t + dt | T \geq t)}{dt} \tag{1}$$

where T denotes the *time of failure*.

3 Hazard-curve simulations

Data are simulated to imitate log-hazard curves for a fleet of trucks owned by Scania. The simulated points are compared to typical in-use data in Figure 1 – verified by Scania Engineers. The data are normalised, with axis values removed throughout in view of data sensitivity.

Colours correspond to different sub-fleets of vehicles – K groups are simulated, where $K = 4$. Each sub-fleet might correspond to various categorisations: vehicle use-type (e.g. mining, forestry, retail), component specification (e.g. turbochargers, alternators), or some other labelling. Certain domains are more sparse than others, therefore, a population model looks to utilise data-rich domains with more information (green), to support sparse domains (pink, purple, and orange).



(a) Operational log hazard data (b) Simulated log hazard data

Fig. 1: Log hazard curves from operational data vs. simulated curves

4 Why share information?

To motivate why sharing information over the fleet is useful, regression models are learnt from each sub-fleet independently (no pooling). On can select linear regression with polynomial bases (up to x^3) to approximate the nonlinear response[6],

$$y_{i,k} = \alpha_1^{(k)} + \alpha_2^{(k)} x_{i,k} + \beta_1^{(k)} x_{i,k}^2 + \beta_2^{(k)} x_{i,k}^3 \tag{2}$$

where $i \in \{1, \dots, N_k\}$ (N_k observations in each domain) and $k \in \{1, \dots, K\}$ (K sub-groups in the fleet). This can be written in terms of the linear and non-linear bases,

$$\begin{aligned} y_{i,k} &= \boldsymbol{\alpha}_k^\top \boldsymbol{\phi}_i + \boldsymbol{\beta}_k^\top \boldsymbol{\psi}_i + \epsilon_{i,k} \qquad (3)\\ &= f(x_{i,k}) + \epsilon_{i,k} \qquad (4) \end{aligned}$$

where $\epsilon_{i,k}$ is additive noise $\epsilon_{i,k} \sim \mathrm{N}(0, \sigma_k^2)$ and,

$$\boldsymbol{\phi}_i \triangleq \begin{bmatrix} 1 \\ x_i \end{bmatrix}, \qquad \boldsymbol{\psi}_i \triangleq \begin{bmatrix} x_i^2 \\ x_i^3 \end{bmatrix} \tag{5}$$

The likelihood of the output can then be defined,

$$y_{i,k} \sim \mathrm{N}\left(f(x_{i,k}), \sigma_k^2\right) \tag{6}$$

$$\tag{7}$$

---

[6] The basis functions can be selected by formal validation in practice.

In a Bayesian manner, priors are placed over the parameters of the model,

$$\alpha_1^{(k)} \sim \mathrm{N}(\mu_{\alpha_1}, \sigma_{\alpha_1}^2), \qquad \alpha_2^{(k)} \sim \mathrm{N}(\mu_{\alpha_2}, \sigma_{\alpha_2}^2) \tag{8}$$

$$\beta_1^{(k)} \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2), \qquad \beta_2^{(k)} \sim \mathrm{N}(\mu_\beta, \sigma_\beta^2) \tag{9}$$

$$\sigma_k^2 \sim \mathrm{IG}(3, 1) \tag{10}$$

And hyperpriors[7],

$$\mu_{\alpha_1} \sim \mathrm{N}(-0.5, 1) \qquad \sigma_{\alpha_1}^2 \sim \mathrm{IG}(3, 0.5) \tag{11}$$

$$\mu_{\alpha_2} \sim \mathrm{N}(1, 0.5) \qquad \sigma_{\alpha_2}^2 \sim \mathrm{IG}(3, 0.5) \tag{12}$$

$$\mu_\beta \sim \mathrm{N}(0, 1) \qquad \sigma_\beta^2 \sim \mathrm{IG}(3, 0.5) \tag{13}$$

Here, the prior distributions are defined such that the prior function is centred around a linear response[10]; i.e. without training data, the expected functions have zero-coefficients for the nonlinear $\boldsymbol{\psi}_i$ bases.

Figure 2a shows the graphical model for K independent regressions. The parameters are inferred using MCMC, via the no U-turn implementation of Hamiltonian Monte Carlo [12] in the probabilistic programming language Stan.

Figure 3a shows the resulting domain-wise regression models (no pooling). The independent models poorly approximate the latent hazard functions, as each inference fails to consider valuable information that can be shared between sub-groups of assets in the fleet. Additionally, the posterior predictive $p(y_* | x_*, \mathbf{x}, \mathbf{y})$ presents large uncertainties, especially in sparse domains – this makes sense. (Variance in the parameter estimates is considered later.)

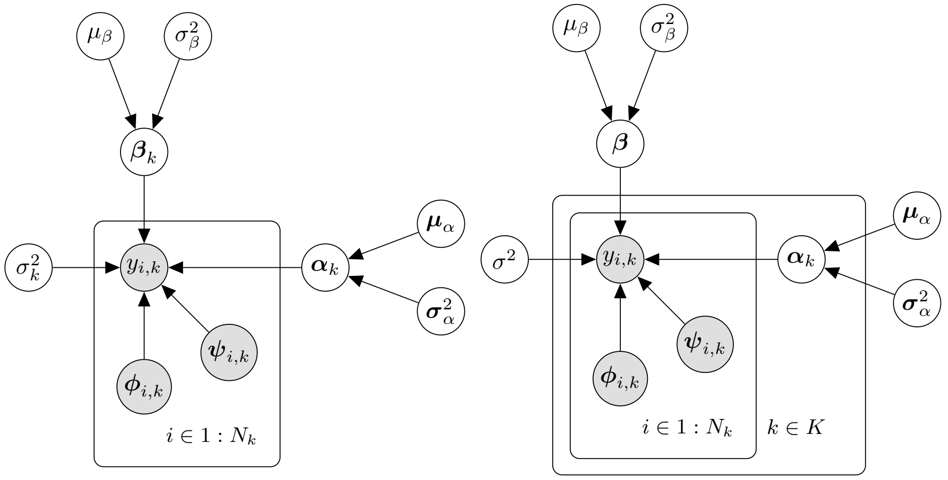
5 Hierarchical Bayes for multi-task learning

When inspecting the data it is clear that regressions mapping are similar between sub-fleets. In consequence, each model might improve by learning the parameters in a joint inference over the whole population. In machine learning, this is referred to as *multi-task learning*; in statistics, such data are usually modelled with hierarchical Bayes [14, 5, 18, 10].

The results so far demonstrate insufficient information for robust independent models, while the data are too dissimilar to be approximated by a single function for the collected observations (complete pooling). Instead, using hierarchical Bayes, a separate (but correlated) regression can be learnt in each domain in combined inference, while the parameters are 'encouraged' to be similar.

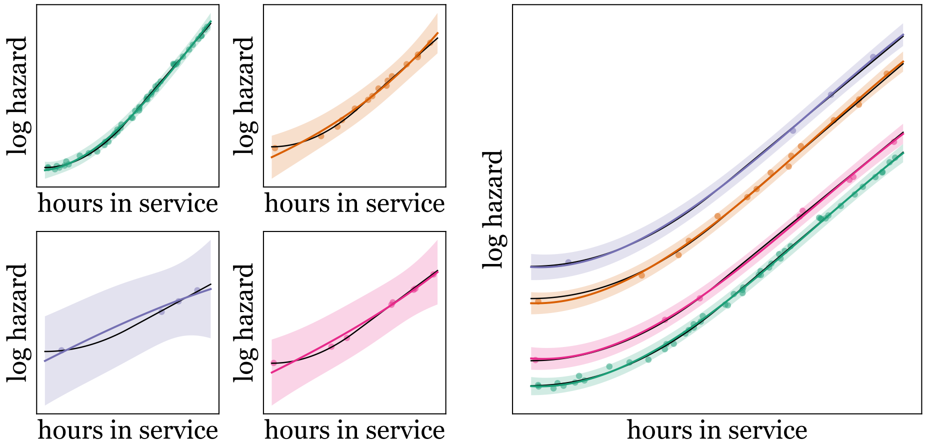
---

[7] Motivations behind the hierarchical structure will become clear at the population level



(a) (b)

Fig. 2: (a) K independent regression models (no pooling). (b) Multi-task learning with hierarchical Bayes and linear mixed effects.



(a) (b)

Fig. 3: Posterior predictive $p(y_{*,k}|x_*, \mathbf{x}, \mathbf{y})$, mean and one-sigma deviation. (a) for K independent regression models (no pooling) and (b) for multitask learning with mixed effects.

Considering the hazard curves, there appears to be a similar *but varied* underlying linear trend for each sub-fleet,

$$\alpha_1^{(k)} + \alpha_2^{(k)} x_{i,k}$$

By setting a shared prior over the parameters of each group $k \in \{1, \ldots, K\}$,

$$\alpha_j^{(k)} \sim \mathrm{N}(\mu_{\alpha_j}, \sigma^2_{\alpha_j})$$

$$\mu_{\alpha_j} \sim \mathrm{N}(m, s) \qquad \sigma^2_{\alpha_j} \sim \mathrm{IG}(a, b)$$

the regression models can be correlated via common latent variables μ_{α_j} (i.e. parent nodes in the graphical model). In turn, sparse domains can borrow statistical strength from data-rich domains. The variable $\sigma^2_{\alpha_j}$ determines how much the parameter $\alpha_j^{(k)}$ depends on the common parents, while the s hyperparameter controls the strength of the overall prior [14]. Crucially, to allow information *sharing*, μ_{α_j} and $\sigma^2_{\alpha_j}$ must be inferred from the population data – if they are constants, each $\alpha_j^{(k)}$ will be conditionally independent and information cannot *flow* between models [14].

Mixed effects

Referring to Figure 1, the nonlinear effect and measurement noise appear consistent over the population – translated by shifts in the linear trend. One can also observe poor performance for independent models in Figure 3a, whereby the nonlinear effects are weakly modelled in sparse domains (i.e. purple and pink). As such, coefficients corresponding to the nonlinear response and measurement noise can be learnt at the population level as common (or *shared*) parameters. The regression model becomes,

$$\begin{aligned} y_{i,k} &= \alpha_1^{(k)} + \alpha_2^{(k)} x_{i,k} + \beta_1 x_{i,k}^2 + \beta_2 x_{i,k}^3 + \epsilon_i \\ &= \boldsymbol{\alpha}_k^\top \boldsymbol{\phi}_i + \boldsymbol{\beta}^\top \boldsymbol{\psi}_i + \epsilon_i \end{aligned}$$

Notice that $\boldsymbol{\beta}$ and ϵ_i are no longer indexed by $k \in \{1, \ldots, K\}$. This implies that the size of $\boldsymbol{\beta}$ and ϵ_i remain the same while the number of the sub-fleets (K) in the population increases (unlike $\boldsymbol{\alpha}_k$). Intuitively, $\boldsymbol{\beta}$ and ϵ_i are termed *tied* parameters [14].

Statistically speaking, the $\boldsymbol{\beta}$ coefficients and measurement noise ϵ_i are referred to as *fixed effects*, since they are learnt at the population-level and shared; on the other hand, $\boldsymbol{\alpha}_k$ are *random effects*, as they vary between groups. Intuitively, a model with both fixed and random effects is considered a *mixed effects model* [16, 7],

$$y_{i,k} = \underbrace{\boldsymbol{\alpha}_k^\top \boldsymbol{\phi}_i}_{\text{random}} + \underbrace{\boldsymbol{\beta}^\top \boldsymbol{\psi}_i + \epsilon_i}_{\text{fixed}}$$

Figure 2b shows the modified graphical model for the multi-task learner with mixed effects. The key difference is the inclusion of a second plate – this collects the data and parameters for each sub-fleet, $k \in \{1, \ldots, K\}$, such that the population model is learnt in a joint inference.

As with the independent models, the posterior distribution is inferred via gradient-based MCMC sampling (no U-turn [12]) in the Stan programming language. The posterior predictive functions are sampled, and the mean and deviation are shown in Figure 3b. The predictive distribution $p(y_{*,k}|x_*, \mathbf{x}, \mathbf{y})$ better approximates the underlying functions by leveraging information between domains. In particular, information from the (data-rich) green domain informs the (fixed) nonlinear effect – parametrised by $\boldsymbol{\beta}$.

6 Analysis of the posterior distribution

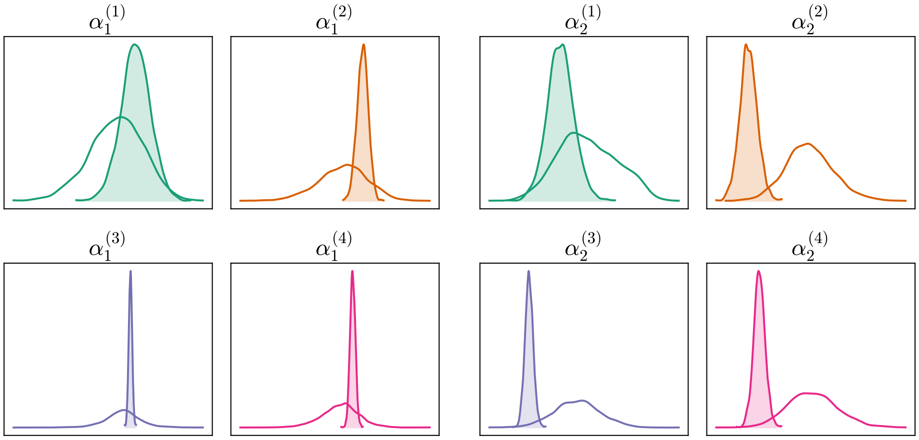
Model improvements are visualised further by reductions in the posterior variance for each parameter. Figures 4a and 4b show the posterior distribution for the slope and intercepts respectively: these parameters correspond to the random (linear) effect $\alpha_1^{(k)} + \alpha_2^{(k)} x_i$. Variance reductions are less significant in the data-rich domain (green) and more significant in sparse domains (orange, pink, and purple). This is intuitive since the population model allows sparse domains to *borrow* information via the shared parent nodes $\boldsymbol{\mu}_\alpha$ and $\boldsymbol{\sigma}_\alpha^2$.

On the other hand, Figure 5 shows the posterior distribution for the fixed effects (i.e. nonlinear bases $\boldsymbol{\beta}$ and noise σ). Building on intuition, when the parameters are tied their variance is reduced and the expected values shift towards the expectation of the data-rich independent model (green).

Finally, Figure 6 is insightful, since it informs which correlations in the hierarchy *transfer* or *share* information between groups of assets. Specifically, while Figure 6a shows little correlation between the intercepts, Figure 6b shows increased correlation between the slope parameters. As such, correlations between the slope parameters appear to contribute more significantly to reductions in model uncertainty.

7 Concluding Remarks

Hierarchical Bayes with mixed effects is demonstrated as an effective method for sharing information within a (simulated) fleet of trucks. Model uncertainty is reduced in a survival analysis case study – based on hazard curves – where selected parameters are inferred at the population-level, rather than vehicle subgroups. The method builds on engineering intuition, as correlations in the hierarchy can be inspected to determine which groups of vehicles are similar (i.e. correlated) for which effects (i.e. model parameters). For example, concerning the rate of change of the (log) hazard,



(a) (b)

Fig. 4: Variance reduction in the posterior distribution of $\boldsymbol{\alpha}_k$. Independent models (hollow) compared to population-level modelling (shaded). (a) shows the intercepts $\alpha_1^{(k)}$ and (b) shows the slope parameters $\alpha_2^{(k)}$.

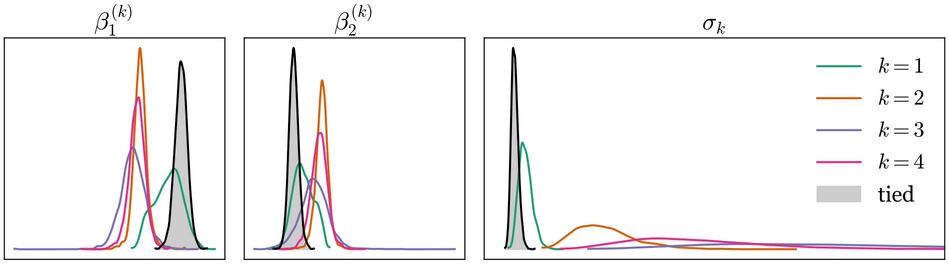
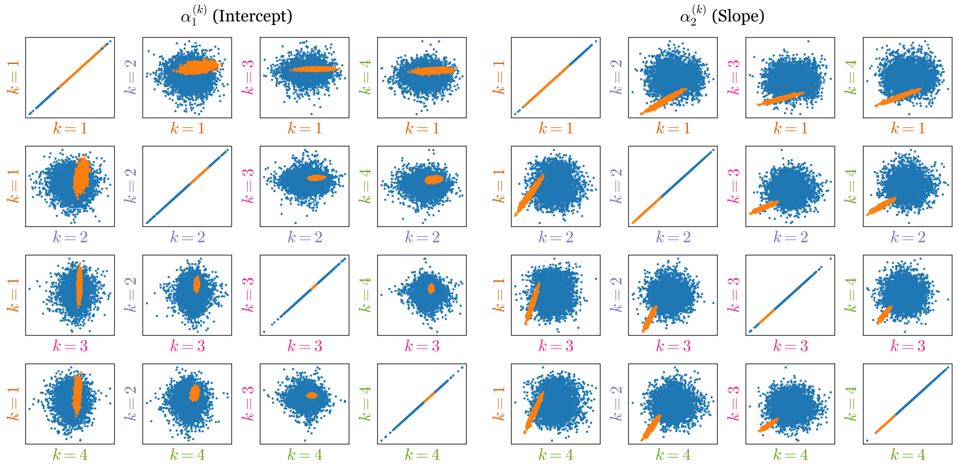


Fig. 5: Posterior distribution of the parameter estimates $\beta_1^{(k)}$ and σ_k. Comparisons between the tied population level parameters (grey shaded) and independent models (hollow) for each domain $k \in \{1, \ldots, 4\}$.


$\alpha_1^{(k)}$ (Intercept)

$\alpha_2^{(k)}$ (Slope)

(a)

(b)

Fig. 6: Conditional posterior distributions for the linear coefficients $\boldsymbol{\alpha}_k$. Pair-wise comparisons across the fleet – blue/orange samples from the independent/population models respectively. (a) shows little correlation in the population model between the intercepts $\alpha_1^{(k)}$, while (b) show higher correlation between slope estimates $\alpha_2^{(k)}$.

mining and forestry vehicles might be more correlated (with each other) than retail vehicles.

Future work must define an objective method to categorise sub-fleets in a practical setting. Specifically, the definition of labels that categorise observations into distinct hazard curves is a non-trivial procedure and requires investigation.

8 Acknowledgements

The first and last authors were supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/W006022/1, particularly the *Ecosystems of Digital Twins* theme within that grant and The Alan Turing Institute. The second and fifth authors were supported by the Next Generation Converged Digital Infrastructure project (EP/R004935/1) funded by the Engineering and Physical Sciences Research Council and BT. This research was supported by Scania CV, Sweden.

References

1. Birolini, A.: Reliability engineering: theory and practice. Springer Science & Business Media (2013)
2. Bull, L., Gardner, P., Dervilis, N., Papatheou, E., Haywood-Alexander, M., Mills, R., Worden, K.: On the transfer of damage detectors between structures: An experimental case study. Journal of Sound and Vibration **501**, 116072 (2021)
3. Bull, L., Gardner, P., Gosliga, J., Rogers, T., Dervilis, N., Cross, E., Papatheou, E., Maguire, A., Campos, C., Worden, K.: Foundations of population-based shm, part i: Homogeneous populations and forms. Mechanical Systems and Signal Processing **148**, 107141 (2021)
4. Bull, L., Rogers, T., Wickramarachchi, C., Cross, E., Worden, K., Dervilis, N.: Probabilistic active learning: An online framework for structural health monitoring. Mechanical Systems and Signal Processing **134**, 106294 (2019)
5. Daumé III, H.: Bayesian multitask learning with latent hierarchies. arXiv preprint arXiv:0907.0783 (2009)
6. Dhada, M., Girolami, M., Parlikad, A.K.: Anomaly detection in a fleet of industrial assets with hierarchical statistical modeling. Data-Centric Engineering **1** (2020)
7. Di Francesco, D., Chryssanthopoulos, M., Faber, M.H., Bharadwaj, U.: Consistent and coherent treatment of uncertainties and dependencies in fatigue crack growth calculations using multi-level bayesian models. Reliability Engineering & System Safety **204**, 107117 (2020)
8. Gardner, P., Bull, L., Dervilis, N., Worden, K.: Overcoming the problem of repair in structural health monitoring: Metric-informed transfer learning. Journal of Sound and Vibration p. 116245 (2021)
9. Gardner, P., Bull, L., Gosliga, J., Dervilis, N., Worden, K.: Foundations of population-based shm, part iii: Heterogeneous populations–mapping and transfer. Mechanical Systems and Signal Processing **149**, 107142 (2021)
10. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian data analysis. Chapman and Hall/CRC (1995)
11. Gosliga, J., Gardner, P., Bull, L., Dervilis, N., Worden, K.: Foundations of population-based shm, part ii: Heterogeneous populations–graphs, networks, and communities. Mechanical Systems and Signal Processing **148**, 107144 (2021)
12. Hoffman, M.D., Gelman, A., et al.: The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res. **15**(1), 1593–1623 (2014)
13. Michau, G., Fink, O.: Domain adaptation for one-class classification: monitoring the health of critical systems under limited information. arXiv preprint arXiv:1907.09204 (2019)
14. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT press (2012)
15. Wand, M.: Semiparametric regression and graphical models. Australian & New Zealand Journal of Statistics **51**(1), 9–41 (2009)
16. West, B.T., Welch, K.B., Galecki, A.T.: Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman and Hall/CRC (2006)
17. Worden, K., Manson, G.: The application of machine learning to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **365**(1851), 515–537 (2007)

18. Xue, Y., Liao, X., Carin, L., Krishnapuram, B.: Multi-task learning for classification with dirichlet process priors. Journal of Machine Learning Research **8**(1) (2007)