

Probabilistic Aging Pipe Strength Estimation Using Multimodality Information Fusion

Jie Chen¹, and Yongming Liu²

^{1,2}Arizona State University, Tempe, AZ, 85281, US

jchen378@asu.edu
Yongming.Liu@asu.edu

ABSTRACT

Accurate pipe material strength estimation is critical for the integrity and risk assessment of aging pipeline infrastructure systems. To predict the strength without interrupting the serviceability of the pipeline, inference methods are used through the relationship between the bulk yield tensile strength and surface material properties from nondestructive testing, such as chemical composition, microstructure images, and hardness testing. In order to make the best of information provided by multimodality surface measurements, Bayesian model averaging (BMA) method is used in this paper to integrate the information from various types of surface measurements for a more accurate bulk strength estimation. The models being considered are constructed by randomly combining the multimodality surface measurements and each case of linear combinations is included. The models considered are assessed by assigning different weights based on the posterior model probability. Markov Chain Monte Carlo sampling provides an effective way for numerically computing the marginal likelihoods, which are essential for obtaining the posterior model probabilities. To avoid the risk of overfitting, BMA is implemented to account for model uncertainty. The predictive performance of a single model and BMA are compared by logarithmic scoring rule. The data collected from industry are used for demonstration and model predictive performance assessment. It is shown that the Bayesian model averaging approach can provide more reliable results in predicting the strength of the aging pipelines.

1. INTRODUCTION

For the pipelines installed decades ago, the mechanical properties decrease with time, such as yield and tensile strength. Failures may occur when these properties are below

the critical levels. It is not an easy job to measure the mechanical properties onsite without disturbing the normal operations. In order to measure the mechanical properties of the pipelines, nondestructive testing is needed without destroying the serviceability of the pipeline part or system. Some indirect methods are proposed through the relationship between the yield strength and surface material properties such as chemical composition, microstructure volume fraction and hardness (Dahire, Tahir, Jiao, & Liu, 2018). Several models are proposed based on different material measurements (Bramfitt, 1998; Hashemi, 2011; Z. Li, Schmauder, & Dong, 1999). Each type of measurement contains certain amount of valuable information for pipe strength estimation. Also, they are all associated with different amount of uncertainties. Relying on a single model often leads to predictions that have good performance in some situations, but not others (Duan, Ajami, Gao, & Sorooshian, 2007). Therefore, it is highly needed to fuse multimodality information in a statistical meaningful way for more accurate estimation.

The general method for ensemble of multiple models is to linearly combine individual model predictions according to different weighting strategies (Duan et al., 2007). The weights can be equal for all models in the simplest case. Also, they can be determined through certain regression-based methods. In this way the weights obtained are actually regression coefficients and hard to interpret since they take on arbitrary positive or negative values and cannot represent the performance of each model (A. E. Raftery, Gneiting, Balabdaoui, & Polakowski, 2005). Another method to calculate the weights of models is based on information criterion, for example, Akaike information criterion (AIC, AICc) and Bayesian information criterion (BIC). However, different information criterion may generate different preference ranks among models since the best model selected by either AIC or BIC can be distinctly different (Burnham & Anderson, 2004). Attention should be carefully paid for the choice of information criterion. Recently, Bayesian Model Averaging (BMA) has gained popularity in diverse fields, such as management science, medicine, and meteorology

Jie Chen et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Duan et al., 2007; Guan, Jha, & Liu, 2011; G. Li & Shi, 2010). BMA is an extension of the common Bayesian inference methods in which both parameter uncertainty and model uncertainty are estimated through obtaining posterior parameters and model posteriors using Bayes theorem. Thus, BMA allows for direct model selection, combined estimation and prediction (Fragoso, Bertoli, & Louzada, 2018). The BMA weights, all positive and summing up to one, reflect relative model performance due to the fact that they are the probabilistic likelihood measure of a model being correct given the observations.

This study explores the use of BMA for strength predictions by fusing surface information of the aging pipelines. The models being considered are constructed by randomly combining the multimodality surface measurements from the nondestructive testing and each case of linear combinations is included. The weight assigned to each single model is achieved by Bayesian model regression and model posterior probability calculation. An ensemble model is obtained by averaging the candidate models. The predicted averaged strength distributions are compared with those of single models and the performance of BMA is evaluated. The paper is organized as follows. First, experimental data collected from industry are introduced. Candidate models are generated based on random combination of linear variables. Next, Bayesian model averaging approach is introduced. Several difficulties of implementing BMA in practice are illustrated in the context of strength estimation. Following that, the overall method is demonstrated using the models and data for aged pipelines. The usefulness of different surface information are analyzed. The advantages of the BMA are discussed based on the results.

2. DATA AND MODELS

2.1. Data

Data are collected from a number of 20 existing aging gas pipes. The data of each pipe fall into 5 categories (Figure 1): yield strength, chemical composition, grain size, hardness and volume fraction. The yield strengths are measured from both surface indentation technique and tensile testing. The strengths from the surface measurements are close the experimental strengths but with some deviations. The chemical composition are measured in weight percentage using a portable spectrometer after 0.003 to 0.005 inches of surface grinding for phosphorus (P), chromium (Cr), copper (Cu), and silicon (Si). The grain sizes are measured at about 0.005 inches from the surface and in mm. The Knoop hardness values are taken at 0.005 inches from the surface. The volume fraction is represented by percentage pearlite estimated by the lever rule using the carbon content at the surface.

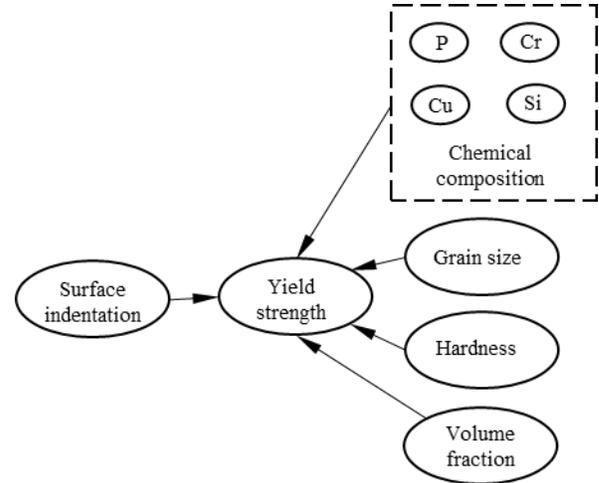


Figure 1. Data categories.

2.2. Models

In Ref. (Bramfitt, 1998), the yield strength of steel is expressed as the linear combination of chemical composition and grain size to the power $-1/2$. Ref. (Z. Li et al., 1999) predicts the yield strengths of two-phase steel by a linear model of volume fraction. Similarly, a linear prediction model of yield strength with hardness is proposed in Ref. (Hashemi, 2011). Based on these existing studies, a general model is proposed as

$$YS_T - YS_S = f(CC, GS^{-1/2}, HD, VF), \quad (1)$$

where YS_T and YS_S are yield strength measured by tensile testing and surface indentation technique, respectively, CC is chemical composition, GS is grain size, HD is hardness, and VF is volume fraction of pearlite. The function $f(\cdot)$ is a linear model. Since YS_S is close to YS_T , YS_S is used as a reference for prediction of YS_T and the deviation between them is modeled as a function of CC , $GS^{-1/2}$, HD , and VF . A total of 7 variables exist in Eq. (1).

Eq. (1) is a full model which contains all the variables but may not be the best model supported by the data. This is because simpler explanations are to be preferred unless there is sufficient evidence in favor of more complicated explanations (Gregory, 2005). Therefore, a class of models to be considered is the collection of all the counterparts of Eq. (1). These models are the reduced models of the full model together with the Eq. (1). In another words, the models considered are linear models of any variable CC (P, Cr, Cu, Si), $GS^{-1/2}$, HD , and VF or any linear combination of these variables.

3. THEORY AND IMPLEMENTATION

3.1. Bayesian model averaging

Suppose there is a set of “reasonable” models M_1, \dots, M_k for estimating a quantity μ from the set of data y . Instead of using one single model for reaching inference for μ , Bayesian model averaging constructs the posterior density of μ given data, $\pi(\mu | y)$, not conditional on any model. This is arrived at via the Bayes formula (Claeskens & Hjort, 2008; G. Li & Shi, 2010; Madigan & Raftery, 1994; A. E. Raftery et al., 2005).

With the prior information on parameter given the model, the integrated likelihood or marginal likelihood $\lambda_{n,j}(y)$ for model M_j is expressed as

$$\lambda_{n,j}(y) = \int f(y | M_j, \theta_j) \pi(\theta_j | M_j) d\theta_j, \quad (2)$$

where $f(y | M_j, \theta_j)$ is the likelihood of data given the model M_j and its parameters θ_j , $\pi(\theta_j | M_j)$ is the prior densities for parameters θ_j of model M_j . The marginal likelihood is larger for a model if more of its parameter space is likely and smaller for a model with large areas in its parameter space having low likelihood values. Thus, a simpler model with compact parameter space will have a larger likelihood than a more complicated model, unless the latter is significantly better explaining the data (van Haasteren, 2014).

Using the Bayes theorem, the posterior density of the model is obtained as

$$P(M_j | y) = \frac{P(M_j) \lambda_{n,j}(y)}{\sum_{j=1}^k P(M_j) \lambda_{n,j}(y)}, \quad (3)$$

where $P(M_j)$ is prior probabilities for models M_1, \dots, M_k under consideration. After computing the posterior density $\pi(\mu | M_j, y)$ of μ for each model assuming that M_j is true, the averaged posterior distributions weighted by posterior model probability is given by

$$\pi(\mu | y) = \sum_{j=1}^k P(M_j | y) \pi(\mu | M_j, y). \quad (4)$$

Eq. (4) shows that the posterior density $\pi(\mu | y)$ is a weighted average and mixture of the conditional posterior densities, where the weights are the posterior probability of each model. By not conditioning on any given model, BMA avoids the mistake of ignoring model uncertainties. The BMA predictions receive higher weights from better performing models since the likelihood of a model is essentially a measure of the agreement between the model predictions and the observations.

Before implementing BMA theory for model averaging, there exist the following difficulties and some points worthy of attention in practice:

The marginal model likelihood $\lambda_{n,j}(y)$ calculated according to Eq. (2) may be analytically intractable. In many cases, no closed form integral is available (Hoeting, Madigan, Raftery, & Volinsky, 1999; G. Li & Shi, 2010). Markov Chain Monte Carlo approach provides a way to overcome this problem. By drawing samples from the required distribution, the integrals in Eq. (2) can be evaluated.

Another issue for BMA is the implementation of averaging according to Eq. (4) in practice. After a class of models is chosen, how to practically implement BMA shown in Eq. (4) remained unsolved. In this paper, the model averaging is achieved through the idea of mixture distributions (Claeskens & Hjort, 2008; Nguefack-Tsague & Zucchini, 2016).

To judge the efficacy of BMA, a metric is needed for assessing and comparing the performance of BMA and the single models. The purpose of this study to predict the strength of aging pipes. Thus, the predictive performance can be used as a meaningful metric to assess a modeling strategy. The logarithmic scoring rule provides a measurement for the predictive ability. This rule compares the probabilities each model assigns to the events that actually occur (Hoeting et al., 1999; Madigan & Raftery, 1994; Piironen, Vehtari, & Computing, 2017; A. Raftery, Madigan, & T. Volinsky, 2000).

The above issues are discussed in the following parts of this section.

3.2. Marginal likelihood calculation

The integrals in Eq. (2) can in general be hard to compute. The MCMC method provides a stochastic method of obtaining samples which simulates a Markov chain converging to the posterior distributions. After some interaction steps, the chain converges and its state can then be used as samples from the desired distribution. Among many available MCMC algorithms, the Metropolis-Hastings algorithm is one of the popular methods (G. Li & Shi, 2010). For a basic introduction to Metropolis-Hastings algorithm, see Ref. (Chib & Greenberg, 1995).

Once the samples of model parameters are drawn from their posterior distributions, the marginal likelihood can then be calculated. The following equations (Newton & Raftery, 1994) are adopted in this paper

$$\lambda_{n,j}(y) = \frac{\frac{\delta}{1-\delta} + \frac{1}{N} \sum_{i=1}^N A \cdot f(y | M_j, \theta_{ji})}{\frac{\delta}{1-\delta} \lambda_{n,j}(y) + \frac{1}{N} \sum_{i=1}^N A}, \quad (5)$$

$$A = 1 / [\delta \lambda_{n,j}(y) + (1-\delta) f(y | M_j, \theta_{ji})],$$

where θ_{ji} ($i = 1, \dots, N$) are randomly sampled from the posterior density $\pi(\theta_j | M_j, y)$ via MCMC sampling, N is the total number of samples θ_{ji} drawn from the posterior

distribution for model M_j , δ is a small number (0.01 for example).

3.3. BMA Implementation

After the MCMC sampling, a set of parameters θ_{ji} ($i = 1, \dots, N$) are sampled from the posterior distributions of parameters of model M_j . Here N is the number of random samples from the posterior distribution of each parameter. Substituting each θ_{ji} into the model, the distribution $\pi(\mu | M_j, y)$ of quantity of interest μ is obtained with the known values of variables. Next, Bayesian model averaging is performed by Eq. (4). The posterior density $\pi(\mu | y)$ for the averaged model is the mixture distribution of the single conditional posterior density $\pi(\mu | M_j, y)$. The mixture distribution based on the weight of each single model is achieved by the following steps.

1. The distributions $\pi(\mu | M_j, y)$ form the columns of a matrix A with dimension of $N \times k$ where k is the number of models considered.
2. Taking a sample of the size N , $b = (b_1, \dots, b_N)$, from the elements of vector $(1, 2, \dots, k)$ randomly according to the probability weight. This series of sample b serves as the indexes for mixture distribution.
3. The mixture distribution is obtained by drawing samples from matrix A according to the indexes from Step 2. The i th sample of $\pi(\mu | y)$ is extracted from the i th sample of $\pi(\mu | M_{b_i}, y)$.

An example of the above algorithm is given by a mixture of two normal distributions, $N_1(0, 8)$ and $N_2(20, 5)$. The weights for distributions N_1 and N_2 are 0.7 and 0.3 respectively. Figure 2 shows the two single distributions and the mixture distribution.

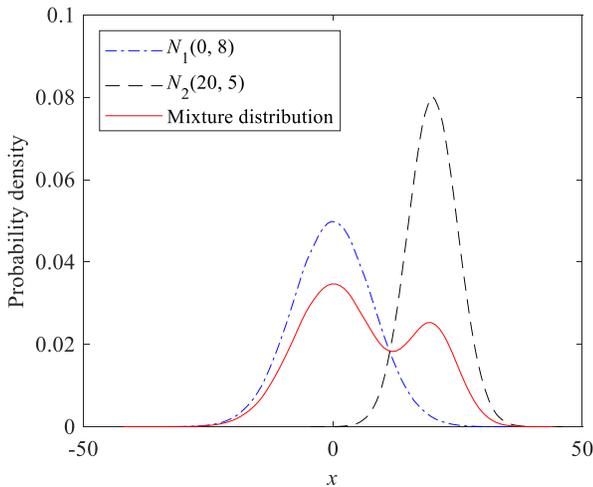


Figure 2. Mixture of two normal distributions.

3.4. Predictive Performance

A primary purpose of statistical analysis is to make forecasts (A. Raftery et al., 2000). Thus, measuring how well a model predicts future observations is one way to judge the efficacy of the BMA strategy. For probabilistic predictions, there exist two types of discrepancies between observed and predicted values: predictive bias and lack of calibration (Claeskens & Hjort, 2008). The predictive bias shows a systematic tendency to predict on the low side or the high side. The lack of calibration shows a systematic tendency to over- or understate predictive accuracy. One measure of predictive ability is the logarithmic scoring rule (Piironen et al., 2017) which is a combined measure of bias and calibration. The smaller the predictive log score for a given model or model average is, the better the predictive performance is.

The data are randomly split into two sets: build data (y^B) which is used for Bayesian model regression and averaging, and test data (y^T) used to measure the model performance. The predictive log score is related with expectation of posterior predictive distribution and expressed as (Madigan & Raftery, 1994)

$$-\sum_{d \in y^T} \log E[\pi(d | M_j, y^B)] \quad (6)$$

for each single model, and

$$-\sum_{d \in y^T} \log E[\pi(d | y^B)] \quad (7)$$

for BMA, where d is one of test data, $\pi(d | M_j, y^B)$ and $\pi(d | y^B)$ are the posterior predictive distribution for model M_j and BMA, respectively. For a single model, $\pi(d | M_j, y^B)$ can be obtained by (Hoff, 2009)

$$\pi(d | M_j, y^B) = \int f(d | M_j, \theta_j) \pi(\theta_j | M_j, y^B) d\theta_j, \quad (8)$$

where $f(d | M_j, \theta_j)$ is the likelihood of the test data, and $\pi(\theta_j | M_j, y^B)$ is the posterior density of θ_j given model M_j . After sampling $\theta_{j1}, \theta_{j2}, \dots, \theta_{jN}$ from $\pi(\theta_j | M_j, y^B)$ through MCMC approximations,

$$E[\pi(d | M_j, y^B)] = \frac{1}{N} \sum_{i=1}^N f(d | M_j, \theta_{ji}) \quad (9)$$

Next, the log score is calculated by Eq. (6). For BMA, $\pi(d | y^B)$ is obtained by applying the method of mixture distribution in Section 3.3. Following that, the log score of BMA is calculated from Eq. (7).

4. RESULTS AND DISCUSSION

4.1. Results for full data

The full data of 20 aging pipes for the chemical composition, grain size, hardness and volume fraction

Table 1. A total of 127 models are linearly constructed from the 7 variables (P, Cr, Cu, Si, $GS^{-1/2}$, HD , and VF).

Table 1 shows the top 10 models ranked by posterior model probability. The rankings are listed in the first column of

Table 1. The best model $YS_T - YS_S = f(Cu, HD)$ with the posterior model probability of 0.169 accounts for 16.9% of the total probability, which is more than 3 times higher than that of the second best model. Another commonly used model selection method is Bayes

Table 1.

are used for the Bayesian model selection, averaging and variable analysis. The results are shown in

information criterion (BIC) (Kuha, 2004; Schwarz, 1978) based on the maximum likelihood estimation. For the best model, BIC selects the same model as the one with the highest posterior model probability. The BIC ranks are also listed in

Table 1. Model ranks and posterior probabilities for the full data.

Model rank	P	Cr	Cu	Si	$GS^{-1/2}$	HD	VF	$P(M_j y)$	BIC rank
1			•			•		0.169	1
2		•	•	•		•	•	0.056	30
3		•	•	•	•	•		0.053	34
4		•	•		•	•	•	0.052	35
5		•		•	•	•	•	0.049	40
6		•	•	•		•		0.044	11
7	•	•		•		•	•	0.042	50
8	•		•					0.038	3
9	•	•	•			•	•	0.036	54
10	•		•			•		0.034	4
$Pr(A_i \neq 0)$	0.34	0.66	0.83	0.42	0.33	0.84	0.39		

In

Table 1, $Pr(A_i \neq 0)$ is the probability that the coefficient A_i of a variable is non-zero. This is calculated by summing all the posterior probabilities of models which contains

Table 1 it can be observed that the variables of best model (Cu and HD) have the highest probabilities that their coefficients are non-zero and show relatively strong predictive usefulness. The results provide positive evidence for Cr and indecisive evidence for P, Si, $GS^{-1/2}$, and VF .

4.2. Predictive performance

The performance of BMA for predictions of pipe strength is assessed by randomly splitting the data from a total of 20 aging pipes into two sets: build data (15 pipes) and test data (5 pipes). The log score serves as the measurement of prediction performance.

shows the results of models selection and averaging for a specific case of data splitting. The models listed in are the top 10 models ranked by posterior model probability. The

this variable. $Pr(A_i \neq 0)$ is used as a metric of evaluating the usefulness of a variable. The variable with a higher probability tends to be more useful. From the bottom row of

best model selected by both posterior model probability and BIC is the $YS_T - YS_S = f(Cr, Cu)$. The probability of the coefficients to be non-zero also show positive evidence for Cr and Cu. It should be noted that the model M_2 which performs the best using the full data is ranked as the second (the third by BIC). This is because both in the subset and full set data case, there is no model showing paramount evidence based on the posterior model probabilities and thus substantial model uncertainty retains.

Predictive log score are shown in . BMA shows better predictive performance than the best model. The improvement can be understood in the following way. The log score of BMA is 0.4 less than that of the best. There are 5 data in the test data set. Therefore the predictive probability of what is actually observed is $[\exp(0.4/5) - 1] \times 100\% = 8.3\%$

larger for BMA than for best model M_1 . However, the model ranked as the second which is the best model for the full data performs better than the best model and BMA from the predictive point of view.

In addition, Figure 3 shows intuitively the comparison of predictive performances of the direct surface indentation technique, lowest BIC model, highest posterior probability model, and BMA. The error bars represent the 95% credible intervals. The dots with different shapes are the expectations of different models. On average, the error of each model is less than that of the sureface indentation technique. BMA shows better performance than the lowest BIC model and highest posterior probability model.

To better compare and analyze the credible interval shown in Figure 3, the predictive distribution of each model is illustrated by taking the data with the largest strength in Figure 3 as an example and shown in Figure 4. It can be seen that the predictive distribution of BMA is wider than that of

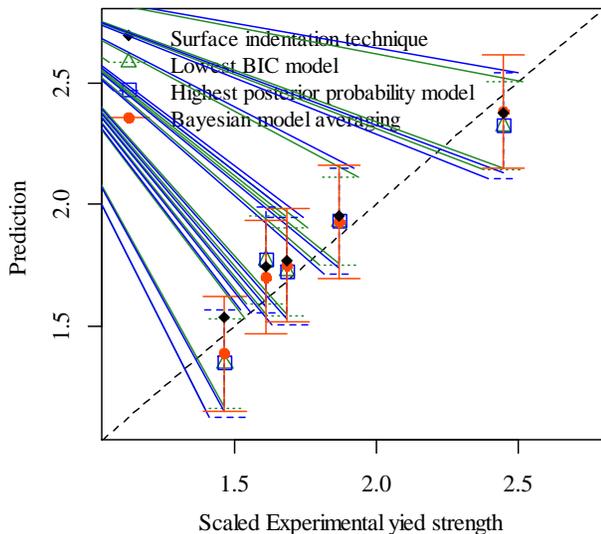
single best model selected by BIC or posterior model probability. This is because that conditioning on a single selected model ignores model uncertainty. The best model in accounting for only 18.9% of the total posterior probability, which shows that model uncertainty is substantial. Ignoring model uncertainty can lead to the underestimation of standard deviation which makes decisions riskier (Hoeting et al., 1999; A. Raftery et al., 2000).

In order to check that the above results are not due to the particular split used, the full data are randomly split for 100 times and the average results are analyzed. Among the repeated experiments of 100 times, BMA have better predictive ability for 71 times. The average predictive log scores are 15.07 and 14.26 for the highest posterior probability model and BMA, respectively. On average BMA can predict the strength with an improvement of 17.6%.

Table 2 . Model ranks and posterior probabilities for the split dada.

Model rank	P	Cr	Cu	Si	$GS^{-1/2}$	HD	VF	$P (M_j/y)$	Log score	BIC rank
1		•	•					0.189	12.9	1
2			•			•		0.065	11.8	3
3		•						0.049	12.6	2
4		•		•	•	•	•	0.046	12.6	46
5		•	•	•		•		0.044	12.3	16
6		•		•				0.043	12.6	11
7		•		•		•	•	0.040	12.2	29
8		•					•	0.038	13.1	10
9		•	•	•	•	•		0.034	12.5	43
10	•		•					0.032	11.8	13
Pr ($A_i \neq 0$)	0.25	0.78	0.69	0.34	0.20	0.27	0.41		12.5	

Figure 3. Predictive performance comparison.



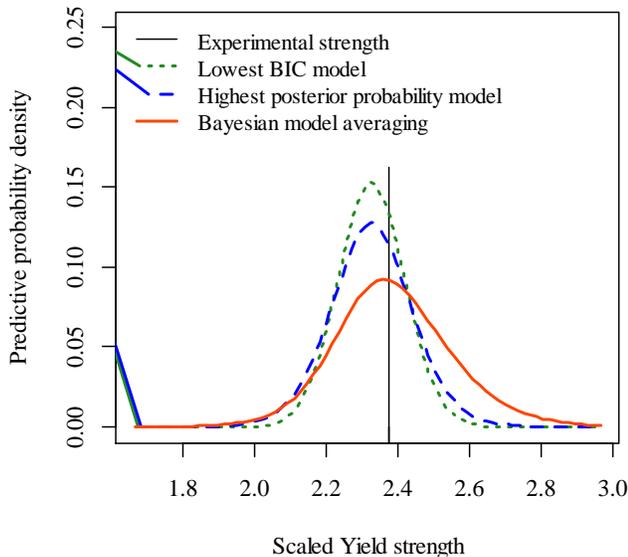


Figure 4. Predictive distributions.

5. CONCLUSIONS

Bayesian model averaging method is used for multimodality information fusion to estimate the probabilistic aging pipe strength. The basic theory and practical implementation of BMA are introduced. Each model considered to be averaged is assigned a weight according to its posterior model probability. BMA is implemented through distribution mixing. The predictive performance of BMA and the single model are evaluated and compared according to logarithmic scoring rule. Based on the current study, the following conclusions can be drawn.

1. From the analysis of full data, the best model $YS_T - YS_S = f(\text{Cu}, HD)$ have the highest probabilities. The probabilities of coefficients being non-zero show relatively strong predictive usefulness for Cu and hardness, positive evidence for Cr and indecisive evidence for P, Si, grain size, and volume fraction.
2. In the case of insufficient data, Bayesian model averaging has the advantages of improving the predictive ability by considering model uncertainty.

Future work may concentrate on the other forms of model uncertainties. This paper considers model uncertainty due to variable selection by constructing linear models. BMA may extent to consider model uncertainties from the functional forms. Also, the likelihood for the data is assumed to be Gaussian distribution. The influence of different likelihood distributions can be further studied.

Acknowledgement

The work in this study was supported by DOT PHMSA through Gas Technology Institute (693JK31810003: Program Manager: Daniel Ersoy). The support is greatly acknowledged.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of Transportation and Gas Technology Institute.

REFERENCES

- Bramfitt, B. (1998). Structure/Property relationships in irons and steels. *Metals Handbook Desk Edition*, 20, 153-173.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *33*(2), 261-304. doi:10.1177/0049124104268644
- Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4), 327-335. doi:10.2307/2684568
- Claeskens, G., & Hjort, N. L. J. C. B. (2008). *Model selection and model averaging*. Cambridge: Cambridge University Press.
- Dahire, S., Tahir, F., Jiao, Y., & Liu, Y. (2018). Bayesian Network inference for probabilistic strength estimation of aging pipeline systems. *International Journal of Pressure Vessels and Piping*, 162, 30-39. doi:doi.org/10.1016/j.ijpvp.2018.01.004
- Duan, Q., Ajami, N., Gao, X., & Sorooshian, S. (2007). Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging. *30*, 1371-1386. doi:10.1016/j.advwatres.2006.11.014
- Fragoso, T. M., Bertoli, W., & Louzada, F. (2018). Bayesian Model Averaging: A Systematic Review and Conceptual Classification. *86*(1), 1-28. doi:10.1111/insr.12243
- Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge: Cambridge University Press.
- Guan, X., Jha, R., & Liu, Y. (2011). Model selection, updating, and averaging for probabilistic fatigue damage prognosis. *Structural Safety*, 33(3), 242-249. doi:doi.org/10.1016/j.strusafe.2011.03.006
- Hashemi, S. H. (2011). Strength-hardness statistical correlation in API X65 steel. *Materials Science and Engineering: A*, 528(3), 1648-1655. doi:doi.org/10.1016/j.msea.2010.10.089
- Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14(4), 382-401.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods* (Vol. 580): Springer.
- Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *33*(2), 188-229. doi:10.1177/0049124103262065
- Li, G., & Shi, J. (2010). Application of Bayesian model averaging in modeling long-term wind speed

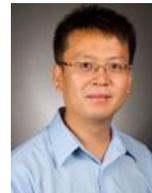
- distributions. *Renewable Energy*, 35(6), 1192-1202. doi:doi.org/10.1016/j.renene.2009.09.003
- Li, Z., Schmauder, S., & Dong, M. (1999). A simple mechanical model to predict fracture and yield strengths of particulate two-phase materials. *Computational Materials Science*, 15(1), 11-21. doi:doi.org/10.1016/S0927-0256(99)00014-2
- Madigan, D., & Raftery, A. E. (1994). Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window. *Journal of the American Statistical Association*, 89(428), 1535-1546. doi:10.2307/2291017
- Newton, M. A., & Raftery, A. E. (1994). Approximate Bayesian Inference with the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(1), 3-48.
- Nguefack-Tsague, G., & Zucchini, W. (2016). *A Mixture-Based Bayesian Model Averaging Method* (Vol. 06).
- Piironen, J., Vehtari, A. J. S., & Computing. (2017). Comparison of Bayesian predictive methods for model selection. 27(3), 711-735. doi:10.1007/s11222-016-9649-y
- Raftery, A., Madigan, D., & T. Volinsky, C. (2000). *Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance* (Vol. 5).
- Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian Model Averaging to Calibrate Forecast Ensembles. 133(5), 1155-1174. doi:10.1175/mwr2906.1
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
- van Haasteren, R. (2014). Marginal Likelihood Calculation with MCMC Methods. In *Gravitational Wave*

Detection and Data Analysis for Pulsar Timing Arrays (pp. 99-120). Berlin, Heidelberg: Springer.

BIOGRAPHIES



Jie Chen is a Ph.D. student majored in Mechanical Engineering in the School of Engineering of Matter, Transport and Energy at Arizona State University. He completed his Master's and Bachelor's Degrees in Civil Engineering from Beihang University, Beijing, China in 2018 and 2015, respectively. His research interest is in reliability and probabilistic methods, Bayesian method, and corrosion and fatigue.



Yongming Liu is a professor in the School for Engineering of Matter, Transport and Energy at Arizona State University. He completed his Ph.D. at Vanderbilt University in 2006, and obtained his Master's and Bachelor's degrees from Tongji University, China in 2002 and 1999, respectively. His research interests include probabilistic methods, diagnostics and prognostics, risk management, materials and structures. He has published over 100 journal articles in the general area of prognostics and health management. He has served on many technical committees in AIAA, ASME, and ASCE. He is an associate fellow of AIAA.