RESEARCH ARTICLE | FEBRUARY 05 2019

Prediction of moisture saturation levels for vinylester composite laminates: A data-driven approach for predicting the behavior of composite materials FREE

Youssef K. Hamidi; Abdelaziz Berrado; M. Cengiz Altan

AIP Conf. Proc. 2065, 030027 (2019)
https://doi.org/10.1063/1.5088285

Prediction of Moisture Saturation Levels for Vinylester Composite Laminates: A Data-Driven Approach for Predicting the Behavior of Composite Materials

Youssef K. Hamidi [a,b,*], Abdelaziz Berrado [c], M. Cengiz Altan [a]

[a] *School of Aerospace and Mechanical Engineering, University of Oklahoma, USA*
[b] *Ecole Nationale Supérieure des Mines de Rabat, Morocco*
[c] *Ecole Mohammadia d'Ingénieurs, AMIPS, Mohamed V University in Rabat, Morocco*

**Email: hamidi@ou.edu*

Abstract: This paper introduces a comprehensive, data-driven method to predict the properties of composite materials, such as thermo-mechanical properties, moisture saturation level, durability, or other such important behavior. The approach is based on applying data mining techniques to the collective knowledge in the materials field. In this article, first, a comprehensive database is compiled from published research articles. Second, the Random Forests algorithm is used to build a predictive model that explains the investigated material response based on a wide variety of material and process variables (of different data types). This advanced statistical learning approach has the potential to drastically enhance the design of composite materials by selecting appropriate constituents and process parameters in order to optimize the response for a specific application. This method is demonstrated by predicting the moisture saturation level for vinylester-based composite laminates. Using 90% of the available published data available as the training dataset, the Random Forests algorithm is used to develop a regression model for the moisture saturation level. Variables considered by the model include the manufacturing process, the fiber type and architecture, the fiber and void contents, the matrix filler type and content, as well as the conditioning environment and temperature. On this training data, the model proved to be a good fit with a prediction accuracy of $R^2_{training}$=94.96%. When used to predict the moisture saturation level for the remaining unseen 10% of the compiled data, the model exhibited a prediction accuracy of R^2_{test}=85.28%. Furthermore, the Random Forests model allows the assessment of the impact of the different variables on the moisture saturation level. The fiber type is found to be the most important determinant on the moisture saturation level in vinylester composite laminates.

Keywords: Data-driven, composite materials, statistical learning, behavior prediction.
PACS: 81.05.Qk, 83.60.-a

INTRODUCTION

Due to their light weight and high performance, composite materials are being increasingly used in structural applications in key industries such as aerospace, automotive, energy and infrastructure [1]. Predicting and characterizing the behavior of composite materials is a key problem in materials research and development. Mico- and macro-mechanical models have been developed to predict some of the critical properties of composite parts, such as the elastic modulus, based on constituent materials properties (i.e. reinforcing fibers and resin). However, these models often consider ideal microstructures and perfect interfacial conditions that limit their accuracy. Other phenomenological constitutive models were developed to predict different properties of composites by utilizing analytical expressions that describe the observed physical phenomena. For instance, the hindered diffusion model is known to accurately capture the moisture absorption behavior of a variety of polymer composites [2]. Nevertheless, these models usually depended on several empirical parameters that are recovered, using experimental data, and as such, are very specific to the studied composite type [2].

For several decades, considerable research has been conducted towards experimentally characterizing composite materials based on constituent materials and process variables. The ultimate goal being to facilitate the design process for these materials to withstand different loads encountered during their life-span for specific applications. Yet, each of these investigations is focused on a precise combination of constituent materials. For instance, many studies focus on a specific type of reinforcement such as glass fibers or carbon fibers, with a specific type of polymeric matrix such as epoxy or vinylester, among others [3]. In addition, these studies often consider one manufacturing process while investigating the effects of process parameters on the composite behavior, or alternatively examine the effect of different manufacturing processes on the fabricated composite parts [4]. Other studies focus on the effects of external conditions, such as hygrothermal or environmental history, on the behavior of the studied composites [5].

While there is an abundance of published body of work on composite materials behavior, each study presents a unique viewpoint that is specific to either (i) constituent materials including fiber type and architecture and matrix

Proceedings of PPS-34
AIP Conf. Proc. 2065, 030027-1–030027-5; https://doi.org/10.1063/1.5088285
Published by AIP Publishing. 978-0-7354-1793-9/$30.00

type and cure process, (ii) a manufacturing process and process parameters, or (iii) external environmental or loading conditions. Although each of these studies could be very insightful for the design using the studied material under similar conditions, the collective knowledge gathered by the scientific community would not be fully used in the design process. Trial-and-error procedures are often used in the design of composites, which has been mostly guided by experiences and predefined rules in materials classification, selection, and property predictions. Avoiding costly trial-and-error experiments can be achieved if the collective knowledge in the materials field is properly utilized to select, for any specific composite application, the best constituent materials and manufacturing process. Data mining and statistical learning techniques have been recently used to exploit material databases and discover trends and mathematical relations for material design. For instance, these techniques have been applied successfully to predict specific properties of metals, such as microstructure [6,7], plastic deformation [8], and non-linear elastic strain-stress relation [9]. However, to the best of the authors' knowledge, no such attempt has been made in the composite materials field.

In this paper, we propose a comprehensive, data-driven method to predict the behavior of composite materials, including thermo-mechanical properties, moisture saturation level, and durability. This advanced statistical learning approach has the potential to significantly improve the design of composite materials by selecting appropriate constituents and process parameters in order to optimize its response for a specific application. The proposed approach is demonstrated on a case study for predicting moisture saturation level for vinylester-based laminates, since moisture saturation levels may dictate the in-service mechanical performance of composite parts.

PROPOSED APPROACH

The proposed data-driven approach for predicting the behavior of composite materials consists of three sequential phases. First, a comprehensive database is compiled from published research articles. Second, the Random Forests algorithm is used on part of the dataset to build a predictive model that explains variations of the investigated material response based on a wide variety of material and process variables. Finally, the model is tested by predicting the behavior of composites using the remaining unseen data for validation.

The first phase of data preparation consists of three steps: (i) data collection, where composite material data is collected from published research articles; (ii) data structuring, where relevant material and process variables (or inputs) are determined; and (iii) data cleaning, where unrealistic values were systematically checked and eliminated. In Phase 2, an appropriate statistical learning technique is applied to most of the collected dataset, labeled training data, in order to build a predictive model that can infer a relationship between a set of potentially explanatory variables, referred to as independent variables and a specific material property, also referred to as dependent variable, or response. The prediction accuracy of the model is assessed using the training data to evaluate the ability of the input variables to explain the variation in the targeted composite property.

In the last phase, the predictive performance of the model is tested using the remaining data, also referred to as unseen or test data. The Random Forests model is used to predict the response of the unseen dataset, and the predicted responses are compared to the real responses. The goals of Phase 3 are to evaluate the predictive performance of the model and distinguish important input variables by ranking them based on their impact on the targeted composite property. Further details on each phase are illustrated on the case study presented hereafter.

CASE STUDY: PREDICTING MOISTURE SATURATION LEVEL FOR VINYLESTER-BASED LAMINATES

One of the critical issues of prime interest during the design of composite materials is durability. Regardless of the manufacturing process, composite materials are prone to humidity, to which they are often exposed. Moisture absorption is known to detrimentally affect the mechanical integrity and durability of composite materials. For example, a moisture saturation level as low as 0.75% can lead to substantial reductions in the tensile strength, elastic modulus, and interfacial shear strength of polymer composites [10]. In addition, higher moisture saturation levels are reported to induce higher reductions in polymer composite performance [11]. Consequently, predicting moisture saturation level in these materials is critical when predicting their service life behavior, especially for structural load-bearing applications designed for long service life. Therefore, this critical property is chosen as a case study to demonstrate how the proposed approach can help predict moisture saturation level in vinylester-based laminates.

Phase 1

Data collection

A comprehensive database is compiled from published research articles between 1977 and 2017. In the process, more than 520 articles reporting moisture absorption in polymer composites were reviewed, and relevant data was collected from those investigating vinylester-based composite laminates. The data was mostly recovered from graphs using the *WebPlotDigitizer* application. Hence, 68 data points were compiled reporting the moisture saturation level, and the composite specifics, the manufacturing process parameters, and hygrothermal conditions.

Data structuring

In this step, explanatory variables affecting the response at hand (i.e. moisture saturation level defined as the maximum percentage weight gain due to absorption of moisture) need to be carefully determined. Based on the literature, all parameters that are documented to affect the moisture saturation level in polymer composites were included in this case study. Hence, reinforcement parameters such as type, content, and architecture were considered. Similarly, matrix filler type and content were considered. The manufacturing processes used to fabricate the composite laminates can also affect moisture absorption, and were considered as well. In addition, the hygrothermal environment was represented by the conditioning environment (water, sea water or relative humidity), and conditioning temperature. Other important variables, i.e. void content and part thickness, were also considered since they are known to affect moisture absorption in these materials [10,11]. It is worth mentioning that each publication reported only some of these variables, and that missing values exist in the compiled database. A recap of the 10 considered variables and their respective ranges, as well as for the observed output -moisture saturation level reported in each study- are presented in TABLE 1.

TABLE 1. Different explanatory variable and moisture saturation level data compiled for the case study.

Input Variable	Number of compiled data points	Range of variable / cases considered (number of recovered data points for each case)			
Reinforcement type	68	Glass (59)	Carbon (10)		
Reinforcement architecture	60	Unidirectional (31)	Multidirectional (15)	Random (4)	Hybrid (10)
Fiber content	55	Up to 74.5% by vol.			
Matrix filler type	68	Nanofillers (3)	Resin (12)	None (53)	
Matrix filler content	68	Up to 7% by weight			
Manufacturing process	48	Pultrusion (27)	Hand Lay-up (11)	LCM (10)	
Conditioning environment	51	Water (44)	Sea water (4)	RH (3)	
Conditioning temperature	65	5 to 100°C			
Void content	20	Up to 6% by vol.			
Part thickness	49	1.41 to 10.16 mm			
Moisture saturation level	68	0.16% to 3.54%			

As presented in TABLE 1, further classifications were considered for different explanatory variables. For instance, reinforcement architecture was further classified to unidirectional, multidirectional, random, and hybrid architectures. Matrix filler type, on the other hand, was categorized into nanofiller (including nanoclay and carbon nanotubes), resin (including epoxy and thermoplastic fillers), and no fillers added. The manufacturing processes were classified into pultrusion, hand lay-up, and liquid composite molding (LCM) processes. While the conditioning environment was classified into water (including distilled and tap water), sea water (including salt water and actual sea water), and relative humidity (RH) conditioning at any concentration. This step is critical as a careful selection of explanatory variables, and a well-defined classification for each of them is essential for a successful use of any learning algorithm.

Data cleaning

In this step, all reported values were checked for consistency to ensure data quality. Erroneous data are known to limit the accuracy of predictive models. The physical phenomena of moisture absorption within vinylester-based composite laminates is well understood, and the reported values are double checked with the literature to verify that no erroneous value was mistakenly included in the database. In case of uncertainty, publications of the same authors and/or research group are checked to verify the reported value. A similar inspection process was performed for void content, fiber content, matrix filler content, conditioning temperature, and part thickness data.

Phase 2

In phase 2, an advanced statistical learning technique is applied to most (90%) of the compiled dataset, labeled training data, in order to build a predictive model that can infer a relationship between a specific material property, also referred to as dependent variable or response and a set of independent (or explanatory) variables. The Random Forests algorithm is chosen for this case study since it copes well with missing values and mixed data types; it is robust to outliers, handles well non-linearity and usually results in predictive models with a highly competitive accuracy on most datasets. Furthermore, it provides an assessment of the contribution of each of the independent variables to explaining the variability of the response in the resulting predictive model. A brief overview of decision trees and Random Forests for regression is given here in order to enhance the understanding of the adopted modeling approach and its resulting conclusions. Decision trees for regression, or simply regression trees, form the core technology of the Random Forests algorithm. A regression tree is a learning algorithm capable of fitting complex datasets and performing both classification and regression tasks. The idea behind a tree is to search for a pair of variable-value within the training set and split it in such a way that will generate the best two child subsets. The goal is to create branches and leafs based on an optimal splitting criteria, a process called tree growing [12]. After each split, this task is performed recursively until the maximum depth of the tree is reached or an optimal tree is found. Regression trees are relatively fast algorithms where variable selection is automatic, and interactions between explanatory variables are naturally considered. Unlike linear models for regression, regression trees map non-linear relationships quite well [13,14]. A Random Forests model [13] for regression consist of a collection of single regression trees each grown built in parallel on different "views" of the training data. The overall prediction is determined by averaging the predictions (aggregating the results) of the individual trees in the forest. Each "view of the data" called a bootstrap sample, is built by performing sampling with replacement from the same training data set. The Random Forests algorithm introduces more randomness and diversity to the feature space, yielding a more robust model [14].

The Random Forests algorithm is used on the training data, which comprises 90% of the collected data set (i.e. 61 data points out of 68 total) in order to develop a model of the relationship between the moisture saturation level and the explanatory variables. The algorithm is run using the open source software R, used primarily for statistical computing. The resulting Random Forests regression model has a predictive accuracy of $R^2_{training}$ = 94.96%. This high accuracy means that for the training data, the explanatory variables were able to explain the variability within the output well. In other words, the 10 variables selected in this case study can explain ~95% of the variability in the values reported for moisture saturation level.

Phase 3

In this last phase, the predictive performance of the model is tested using the remaining 7 data points (10% of 68 total data points), also referred to as unseen or test data. The Random Forests model is used to predict the response of the unseen dataset, and the predicted values are compared to the real moisture saturation level responses. The obtained accuracy can be a good indicator of the predictive performance of the model for moisture saturation level in future vinylester-based composite laminates to be designed. The accuracy of the model using the test data is calculated to be R^2_{test} = 85.28%. The reported moisture saturation levels and the Random Forests model predicted values are presented in FIGURE 1 for the 7 experimental data points. Although lower than the accuracy registered for the training data, this predictive capability is still very useful during the design process as the predicted moisture saturation level would be within an interval of ±15% of the actual moisture level.

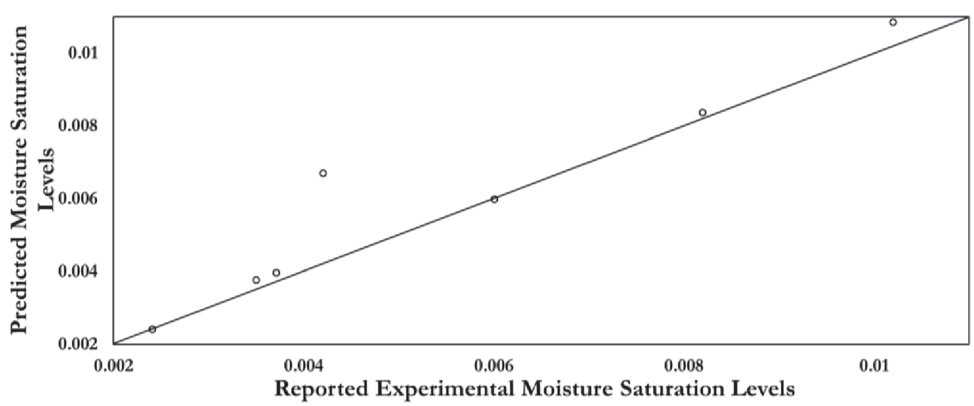

FIGURE 1. Random Forests Model Predicted Moisture Saturation Levels as a Function of the Reported Experimental Values in the Test Dataset Compared the X = Y Line.

It is worth mentioning that five data points are predicted quite accurately by the Random Forests model, while two data points present a relatively higher error. A closer look at the specific data points of interest show that their void content is not reported, and a higher void content would explain a higher moisture saturation level. This emphasizes the importance of expanding the collected database to include the maximum number of cases possible, so that the developed predictive model would be more accurate.

Another interesting feature of the Random Forests technique is its ability to distinguish the impact of input variables on the predicted property. In fact, the model calculates the variability explained by each variable using the percent increase in MSE (noted %IncMSE) between the case where the variable of interest is considered and the case where it is not considered in building the Random Forests model. Using this parameter, the explanatory variables can be classified according to their impact on the moisture saturation level. TABLE 2 shows the variables with high impacts on the moisture saturation level in vinylester-based laminates. The remaining variables have very little effect on the output variability.

TABLE 2. Impact of explanatory variables on the moisture saturation level.

Variable	Reinforcement Type	Void Content	Manufacturing Process	Matrix Filler Type	Conditioning Temperature	Matrix Filler Content	Conditioning Environment
%IncMSE	19.50	7.39	6.13	4.47	3.88	3.84	3.43

For instance, the most important variable by far is observed to be the reinforcement type. Since only glass and carbon fibers were encountered in the database, it follows that moisture saturation levels are fundamentally different between the two types of reinforcement. This can be explained by the different sizings commercially applied to each fiber type, combined with a much higher moisture absorption rate along the fibers [10,11]. Void contents and manufacturing process are also found to influence significantly the moisture saturation levels for vinylester-based composites. This is expected since voids serve as storage locations for water molecules [2], and the manufacturing process would define, more or less, the level of void content in the fabricated composite [15]. What was not expected however, is that fiber content and conditioning environment and temperature did not seem to influence extensively the moisture saturation level of vinylester based composite laminates. These findings would be of great interest while designing vinylester-based composite parts, especially when the intended application involves immersion in water or a humid environment.

CONCLUSION

A data-driven approach is proposed to predict the properties of composite materials. The approach, based on advanced statistical learning techniques applied to the collective knowledge in the composite materials field, involves 3 phases. First, a comprehensive database is compiled from published research articles. Second, the Random Forests algorithm is used to build a predictive model that explains the investigated material response based on a several material and process variables. Third, the predictive performance of the built model is assessed, and the importance of the used explanatory variables is assessed.

This approach is demonstrated on predicting the moisture saturation level of vinylester-based composites. Using 90% of the compiled data as the training dataset, the Random Forests algorithm is used to develop a predictive model for the moisture saturation level. Variables considered by the model include the manufacturing process, the fiber type and architecture, the fiber and void contents, the matrix filler type and content, as well as the conditioning environment and temperature. The model showed a predictive accuracy of $R^2_{training}$=94.96% on the training data, inferring that the variables almost fully explain the variation in the moisture saturation levels. The model displayed a predictive performance of R^2_{test}=85.28% on the unseen 10% of the compiled data. Furthermore, the impact of the different variables on the moisture saturation level can be assessed using the random forest model. The reinforcement type is found to be the most important determinant on the moisture saturation level in vinylester composite laminates.

This proposed approach has the potential to drastically enhance the design of composite materials by leveraging the available published data in the composite material field to select appropriate constituents and process parameters in order to optimize the response of a composite part for a specific application.

REFERENCES

1. H. J. Barraza, K. Olivero, Y. Hamidi, E. A. O'Rear, M. C. Altan, *Compos. Interf.* **9**, 477-507 (2002).
2. G. E. Guloglu, Y. K. Hamidi, M. C. Altan, *Polym. Engr. & Sci.* **57**, 921-931 (2017).
3. H. J. Barraza, Y. K. Hamidi, L. Aktas, E. A. O'Rear, M. C. Altan, *Compos. Interf.* **24**, 125-148 (2017).
4. M. A. Yalcinkaya, E. M. Sozer, M. C. Altan, *Compos. A* **102**, 336-346 (2017).
5. L. Aktas, Y. Hamidi, M. C. Altan, *J. Mat. Proc. & Manuf. Sci.* **10**, 255-267 (2002).
6. H. Xu, R. Liu, A. Choudhary, W. Chen, J. Mech. Design **137**/051403, 1-10 (2015).
7. M.A. Bessa, R. Bostanabad, Z. Liu, A. Hub, D. W. Apley, C. Brinson, W. Chen, W. K. Liu, *Comput. Meth. Appl. Mech. Engr.* **320**, 633-667 (2017).
8. D. Versino, A. Tonda, C. A. Bronkhorst, *Comput. Meth. Appl. Mech. Engr.* **318**, 981-1004 (2017).
9. A. Leygue, M. Coret, J. Rethore, L. Stainier, E. Verron, *Comput. Meth. Appl. Mech. Engr.* **331**, 184-196 (2018).
10. L. Aktas, Y. Hamidi, M. C. Altan, *J. Mat. Proc. & Manuf. Sci.* **10**, 239-254 (2002).
11. H. J. Barraza, L. Aktas, Y. K. Hamidi, J. Long, E. A. O'Rear, M. C. Altan, *J. Adhesion Sci. & Techn.* **17**, 217-242 (2003).
12. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, *Classification and Regression Trees*, NY: Taylor & Francis, 1984, 1-386.
13. L. Breiman, *Machine Learning*, **45**, 5-32 (2001).
14. A. Berrado, G. C. Runger, *Data Mining & Knowledge Discovery* **14**, 409-431 (2007).
15. Y. K. Hamidi and M. C. Altan, *International Polymer Processing*, **32**, 1-18 (2017).