

Prediction of Acrylonitrile–Butadiene–Styrene Mechanical Properties through Compressed-Sensing Techniques

Jonah Poort, Milad Golkaram,* Pieter Janssen, and Jan Harm Urbanus



Cite This: *J. Chem. Inf. Model.* 2024, 64, 7257–7272



Read Online

ACCESS |



Metrics & More

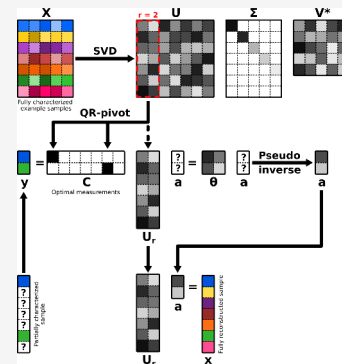


Article Recommendations



Supporting Information

ABSTRACT: One of the challenges in the plastic industry is the cost and time spent on the characterization of different grades of polymers. Compressed sensing is a data reconstruction method that combines linear algebra with optimization schemes to retrieve a signal from a limited set of measurements of that signal. Using a data set of signal examples, a tailored basis can be constructed, allowing for the optimization of the measurements that should be conducted to provide the highest and most robust signal reconstruction accuracy. In this work, compressed sensing was used to predict the values of numerous properties based on measurements for a small subset of those properties. A data set of 21 fully characterized acrylonitrile–butadiene–styrene samples was used to construct a tailored basis to determine the minimal subset of properties to measure to achieve high reconstruction accuracy for the remaining nonmeasured properties. The analysis showed that using only six measured properties, an average reconstruction error of less than 5% can be achieved. In addition, by increasing the number of measured properties to nine, an average error of less than 3% was achieved. Compressed sensing enables experts in academia and industry to substantially reduce the number of properties that must be measured to fully and accurately characterize plastics, ultimately saving both costs and time. In future work, the method should be expanded to optimize not only individual properties but also entire tests used to simultaneously measure multiple properties. Furthermore, this approach can also be applied to recycled materials, of which the properties are more difficult to predict.



1. INTRODUCTION

The experimental characterization of plastic properties can be both time-consuming and costly.¹ Therefore, as an alternative, many modeling approaches have been developed for computationally predicting material properties. One method is through quantitative structure–property relationships (QSPRs),^{2–7} where the chemical structure of a molecule is linked to corresponding physical, mechanical, and chemical properties. Conventionally, QSPRs have been modeled through physical relations and high-fidelity molecular modeling techniques. The recent progress in the field of machine learning (ML) has substantially contributed to the rise of data-driven models for QSPRs. With the help of large amounts of data, hidden correlations can be identified and complemented by physical relationships.^{8,9}

However, QSPR models have limited training data, leading to low predictivity.¹⁰ Moreover, the use of structure–property relationships relies on the knowledge of the chemical structure, in this context, of polymers. However, in practice, commodity and engineering polymers are limited and despite possessing limited chemistry, their properties can vary depending on the crystallization, copolymer ratio, molecular weight, additives, etc. Regardless of the chemical structure, several properties can vary by changes in the polymer's molecular weight.⁷ Because of chain folding, polymers typically can only possess partially crystalline structures. In addition, the degree of crystallinity substantially impacts a polymer's mechanical properties.¹¹

Therefore, researchers have strived to predict individual properties independent of the chemical structure, although with limited success. Qiu et al. predicted the viscosities of a common bisphenol A-based epoxy resin at room temperature and high temperature.¹² Comparing their results with those obtained using the classical viscosity–temperature model, the robustness of the predictions was substantiated.¹² In another study, an ML algorithm was trained using 46 training data points to predict the optimal tensile strengths of polyurethane–wood ash composites with a high degree of precision.¹³ An ML model was recently developed for predicting polypropylene (PP) properties based on the morphology's contribution to the polymer. Four ML algorithms were applied to predict the α - and β -polymorph contents and mechanical properties of PP injection moldings. The study results demonstrated the relationships among the injection pressure, shear rate, and crystal formation.¹⁴ ML models were also applied to predict the tensile strengths of polymeric films. The results suggested the superior classi-

Received: April 11, 2024

Revised: September 3, 2024

Accepted: September 4, 2024

Published: September 19, 2024



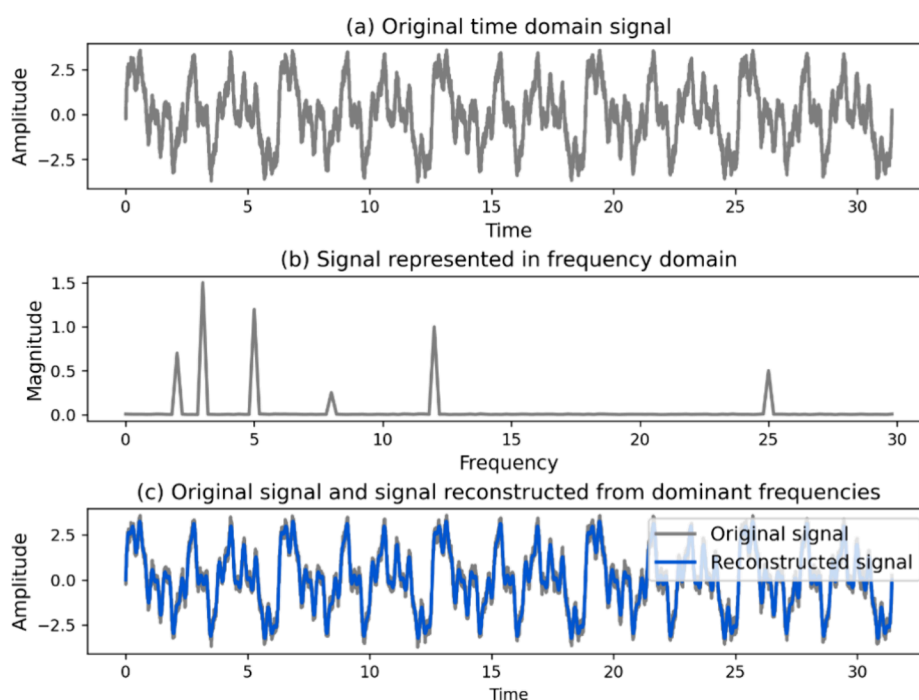


Figure 1. Use of the Fourier transform to transform a complex time signal to a sparse representation. The top plot (a) shows an example signal in the time domain. The middle plot (b) shows this same signal in the frequency domain after application of the Fourier transform. The bottom plot (c) shows the original signal in the time domain and a nearly identical signal reconstructed through the inverse Fourier transform from only the dominant frequencies seen in the middle plot (b).

fication capability of the algorithms for sorting films into conforming and nonconforming parts.¹⁵ The polydispersity index (PDI) was also investigated using a similar approach to relate the chain length distribution with the macroscopic mechanical properties (e.g., tensile modulus, elongation at break, and tensile strength at break).¹⁶

Although these models have powerful predictive capabilities, they typically must be tuned using numerous data to accurately generalize to new previously unseen instances. Unfortunately, large amounts of good data are often lacking. One approach to solve this problem is to combine data from multiple sources, but this often leads to data gaps, where one source contains data for a certain set of properties that are not included in a different data set and *vice versa*.

In this work, we propose a method based on the concept of compressed sensing²³ (sometimes also referred to as *compressive* sensing) that allows for the reconstruction of large sets of properties based on a much smaller set of carefully chosen measured properties.

Compressed sensing is based on the assumption of signal sparsity: the fact that most of (if not all) the natural signals can be transformed to signals in a different domain, where they can be represented in a more compact (sparse) manner. For instance, a temporal waveform (as shown in Figure 1a) can be transformed using a universal transform, such as the Fourier transform, to a waveform in the frequency domain, where it can be represented by only a few dominant frequencies (Figure 1b). By discarding all the other frequencies besides these dominant ones, the time signal can be stored using only a few parameters and reconstructed by simply applying the inverse Fourier transform (Figure 1c). The goal of compressed sensing is to find these sparse frequencies by measuring only a small subset of the original signal. This is typically done using an optimization scheme, and is explained in Section 2.1.

As applied in this work, the set of all the properties of a plastic can be interpreted as the complete signal, while the known measured properties are the sampled parts of the signal. This is schematically depicted using a visual analogy in Figure 2. In addition, instead of a universal transform, such as the Fourier transform, a tailored transform basis is used in this work. The tailored basis is derived from example data (e.g., a complete set of measured properties) and allows for more efficient solving of the compressed-sensing problem. Therefore, the optimization of the measurements that should be performed to maximize the signal reconstruction's accuracy becomes computationally tractable, as further detailed in Section 2.2.

Previously, compressed sensing has been widely applied in various fields, particularly audio and visual signal processing. For instance, Sreenivas and Kleijn used compressed sensing to accurately reconstruct speech patterns from a moderate number of audio signal samples.²⁴ In the medical field, compressed sensing has been used to reduce the number of samples required for measuring electrocardiogram (ECG) signals, reducing the costs of the otherwise sampling-intensive measurements.²⁵ In addition, Manohar et al. used compressed sensing for accurately reconstructing the total two-dimensional fluid flow field comprising 90,000 points from only 42 known point measurements.³⁰ A good overview of various applications of compressed sensing has previously been published.²⁷

Acrylonitrile–Butadiene–Styrene (ABS) is an engineering thermoplastic block copolymer used in automotives, toys, and electric and electronic equipment. Compared with other plastics, ABS is less studied because of its complex structure. ABS is synthesized by coupling the styrene acrylonitrile copolymer (SAN) with polybutadiene (PB) (Figure 3). Several properties, including hardness, tensile and flexural moduli, and tensile and impact strengths, have been reported as key ABS

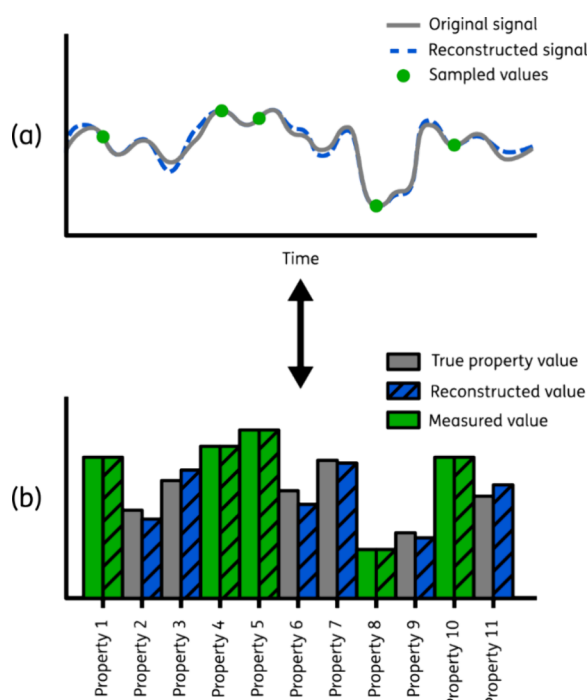


Figure 2. Visual analogy between (a) classical application of compressed sensing for (continuous) signal reconstruction and (b) application of compressed sensing for unknown (discrete) property estimation.

property reconstruction accuracy indicators.^{17–20} ML models, such as linear regressions, decision trees, random forests (RFs), and Adaptive Boosting (AdaBoost), have previously been used to predict ABS hardness values. The accuracy of the RF model was superior to those of the other ML models. For the various models, the coefficients of determination (R^2) ranged from 0.8437 to 0.9136, with the RF model possessing the highest R^2 value (0.9136).^{21,22}

The literature overview reveals a clear need for methods that can help in predicting the values of unknown molecular properties based on known property data. Because of their powerful and flexible modeling capabilities, data-driven methods have attracted increased research interest. While many methods applied in previous studies require large amounts of data and extensive tuning of parameters to provide satisfactory results, the proposed compressed-sensing method is data efficient and requires neither hyperparameter tuning nor extensive training. As such, the proposed method is simple to set up and fast to execute, even with limited available data. To the best of the authors' knowledge, this manuscript describes the first instance of compressed-sensing applied for predicting polymer properties, providing a simple and effective tool for

reducing the costs and time associated with characterizing polymer properties.

The remainder of this paper is structured as follows: In Section 2, the theoretical background is provided for universal and data-driven tailored-basis compressed sensing. Section 3 describes the test case used to demonstrate the methods and provides the results of three analyses, starting with an investigation of the effects of the total number of measurable properties on the reconstruction accuracy. Using only six well-chosen measured properties, the total data set was reconstructed to within an average error of 5%. In the second analysis, the impact of the original size of the data set used to develop the tailored basis on the reconstruction accuracy was studied. Using only 16 fully characterized samples, good accuracy, comparable to those achieved using more data samples, was achieved. Finally, the effect of the data scaling on the reconstruction accuracy was evaluated. The results suggested that data scaling offered no additional benefits over retaining the original data magnitude (and, in fact, decreased the reconstruction accuracy). Lastly, in Section 4, the main findings and their importance are summarized, and recommendations are provided for future work.

2. METHODS

In this section, the theory behind the applied methods is provided, starting with an introduction of universal compressed sensing. Then, the more specific data-driven tailored-basis compressed-sensing problem is discussed. After this, the type of tailored-basis used in this work (the singular value decomposition) is further explained. Finally, the last subsection provides a brief description of the experimental methods used to generate the data reported in this paper.

2.1. Universal Basis Compressed Sensing. Compressed sensing is based on sparsity (where a signal vector/matrix primarily contains zeros).²⁸ More concretely, it is based on the fact that most of (if not all) the natural “signals” (e.g., images, waveforms, or measurements) occupy only a very small segment of their possible signal space. For instance, a 1000×1000 pixel image can be used to display millions of possible objects from dogs to mountains and cars, but for a randomly generated set of pixels, the probability is close to 100% that the pixel set will contain nothing more than random noise.¹ In fact, the probability that even a small subset of pixels lines up to form any recognizable structure is incredibly slim.

Therefore, most signals can be transformed to signals in a different domain, where they can be represented in a much more compact, sparser manner. For instance, a complex waveform can be represented with just a few dominant frequencies by transformation through a Fourier transform,²⁹ as shown in Figure 1a,b. Clearly, the signal is much sparser in the frequency domain, where only a few frequencies are

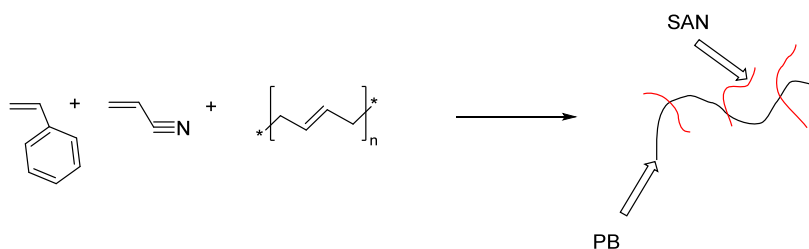


Figure 3. ABS synthesized through emulsion polymerization of styrene and acrylonitrile with polybutadiene (PB).

important, than in the time domain. By discarding all but these nonzero frequencies and performing the inverse Fourier transform, the original time signal can be accurately reconstructed from limited information.

Sparsity in different domains not only forms the basis for many file compression algorithms but also allows for the reconstruction of complex signals by measuring only a few select samples, i.e., compressed sensing.²⁸

Referring to the above example, the goal of compressed sensing is to find the dominant frequencies from the middle plot (Figure 1b) by measuring only a small subset of the original signal in the top plot (Figure 1a). Mathematically, the aim is to reconstruct a (large) signal (x) from a much smaller subset (y) of measurements of x . Using a universal basis transformation matrix (Ψ , such as a Fourier or wavelet transform), the original signal (x) is transformed to a sparse representation (s), which is a vector primarily comprising zero entries and a limited number of nonzero elements. Through optimization methods, (an approximation of) the sparse vector (s) can be recovered from the known y and Ψ arrays and used to reconstruct x with high accuracy.

The measurement array ($y \in x$) is given by

$$y = Cx \quad (1)$$

where C is the measurement matrix and specifies the elements of x that are measured and known in y . Assuming it is sparse in some other domain, signal x can be transformed to sparse vector s using transform Ψ as follows:

$$x = \Psi s \quad (2)$$

The measurement vector (y) can be similarly transformed to a sparse vector:

$$y = (C\Psi)s = \Theta s \quad (3)$$

where Θ is a subset of the transformation matrix (Ψ) given by the measurement matrix (C). Because eqs 2 and 3 both share the common sparse vector (s), x can be reconstructed from y by first solving eq 3 for s before substituting it in eq 2 to produce x . However, although Ψ (and, therefore, Θ) and y are both known, eq 3 provides an underdetermined system of equations, meaning that there are infinitely many possible solutions for s , while only one version of s returns x when used in eq 2. As such, eq 3 cannot simply be solved for s .²⁸

In order to circumvent this issue, the fact that s is expected to be sparse is used. By minimizing the l_0 zero-norm (the number of nonzero elements in an array) of s while simultaneously solving eq 3, the sparsest vector ($\hat{s}$) that solves eq 3 can be obtained, as summarized by the following optimization problem:

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\| s \right\|_{l_0} \text{ subject to } y = (C\Psi)s \quad (4)$$

Unfortunately, finding the minimum l_0 zero-norm is a combinatorial optimization problem that can only be globally solved by evaluating all the possibilities for $\hat{s}$, which is practically intractable for most real-world problems where s is large. Instead, under certain conditions on the C matrix, the l_1 taxicab norm can be used in place of the zero-norm. The taxicab norm defines the distance between two vectors as the distance traversed between them on a right-angled grid (such as how a taxi would drive in a grid-based city layout). This is otherwise expressed as the sum of the absolute values of the elements of s . Using the l_1 norm, eq 4 becomes

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\| s \right\|_{l_1} \text{ subject to } y = (C\Psi)s \quad (5)$$

and the optimization problem becomes convex, rendering the problem much quicker (and practically feasible) to solve using standard optimization methods while still ensuring a sparse solution is found. This does not happen when using the Euclidean l_2 norm, which still produces solutions where all the coordinates of the solution vector are active.²⁸

For the l_1 norm simplification to work, certain conditions must be satisfied. The main condition is on the C matrix, and specifies that the resulting Θ matrix should function as a near-isometry map on the sparse vector s (i.e., Θ preserves the distances of transformed vector s).²⁸ This comes down to the fact that the number of measurements in C should be sufficient and incoherent (produce a small inner product with the columns of Ψ). Intuitively, this means that the information in the measurements should not be too structured or concentrated in a few rows. In practice, most random measurement matrices produce sufficiently incoherent measurements such that the conditions for the l_1 simplification are satisfied.²⁸ For example, referring to Figure 1, if the measurements would be evenly spread over the entire signal in the top plot (or bunched together within a small window), they would likely miss many of the finer signal details. Of course, by increasing the number of measurements, the probability for capturing more of these details increases. However, a more efficient approach would be to simply distribute the samples randomly over the signal instead, as this is more likely to capture different levels of detail with fewer measurements.

Although these conditions can be further quantified and formalized by introducing a coherence measure and the restricted isometry property,²⁸ as this work primarily uses tailored-basis compressed sensing (introduced in the next section), where these conditions are less important, the theory behind these additional conditions will not be provided herein. Interested readers are referred to the publication by Brunton and Kutz.²⁸

Once $\hat{s}$ is found, it can be used in eq 2 to reconstruct x . It should be noted that because $\hat{s}$ approximates s , x will not be perfectly accurately reconstructed. For instance, although solution $\hat{s}$ may be found using the right sparse elements, these elements' values might be incorrect. As such, although the correct frequencies are used in a Fourier transform, their relative magnitudes might slightly deviate, producing a reconstructed signal approximating, but not necessarily perfectly matching, the true value of x . Thus, although it is a powerful tool, compressed sensing is not a magical solution for flawless signal reconstruction.

2.2. Tailored-basis Compressed Sensing. Universal basis compressed sensing, as discussed in the previous section, is already a powerful signal reconstruction technique. However, the use of a universal basis also constrains the number and distribution of the measurements required to feasibly and practically solve the reconstruction problem.

Typically, in compressed sensing, a universal basis is used where the structure of the signal to be reconstructed is not well understood. For example, when the image to be reconstructed is unknown, e.g., a dog or a car, a universal basis allows for the reconstruction of both possibilities. However, to ensure all these possibilities are covered in the universal basis approach,

certain measurement conditions must be satisfied; primarily that there are sufficient randomly spaced measurements.

In contrast, when the type of signal to be reconstructed is known (e.g., only pictures of dogs have to be reconstructed), the number of measurements can be significantly reduced, as measurements no longer have to cover all the possibilities. This can be achieved by slightly changing the compressed-sensing formulation of (eq 3) and using a tailored transform basis instead of a universal transform basis.

In essence, a tailored basis is a transform basis constructed based on example data from the signal type to be reconstructed (e.g., a database of dog images). Using the information in the example data on the structure of the signal to be reconstructed, a more informed decision can be made for the measurements that produce the most accurate reconstruction of signals similar to those in the example database.³⁰

Mathematically, in tailored-basis compressed sensing, the universal transform matrix (Ψ) is replaced by a smaller tailored library matrix (Ψ_r). This tailored library is constructed from X , which is a set of examples of the signal x . In Ψ_r , r denotes the rank of the tailored basis. Because the tailored basis is an approximation of X , the rank denotes how many terms of this approximation comprise Ψ_r (which is similar to the number of terms used in a Fourier/Taylor series approximation). Depending on the problem, different methods for constructing Ψ_r can be used, such as the proper orthogonal decomposition, dynamic mode decomposition, and singular value decomposition (SVD).²⁸ In this work, the SVD was used for constructing Ψ_r (see Section 2.3 for more details).

Using a tailored basis, eqs 2 and 3 respectively become the following expressions³⁰:

$$x = \Psi_r a \quad (6)$$

$$y = (C\Psi_r)a = \theta a \quad (7)$$

where s in the original equation is replaced with the tailored-basis coefficients (a), a vector of length r that is no longer necessarily sparse and indicates the linear combination of Ψ_r (the approximation of X) that is needed to reconstruct signal x from measurements y . When the number of measurements in y is the same as the rank (r), matrix θ becomes square, and the system of equations is no longer underdetermined. Therefore, the pseudoinverse of θ can be used to relatively easily solve for a .²⁸

To further clarify, in the universal basis approach, the reason for requiring sparsity in the universal basis approach is that the main system of equations to be solved (eq 3) is underdetermined and, thus, has infinitely many solutions. To still find a solution that can be used to reconstruct x from y , the universal basis must transform the original signal to a signal in a domain where the signal is sparse (with nonsparse entries being the dominant modes, e.g., signal frequencies for waveforms), as this condition can be used as a constraint to find the best solution of eq 3. That is, if it is known that the solution to eq 3 should be sparse, an optimization scheme (possessing the proper objective function) can be used to find the sparsest solution among all the (infinitely many) possible solutions to eq 3. If the transform basis does not provide a sparse result, s is unconstrained and, therefore, cannot be used to narrow the solution space of eq 3. In this case the right solution would be impossible to find without evaluating all possible solutions (which is practically infeasible for most problems).

However, when a tailored basis is used, the system of equations (eq 3) is no longer underdetermined, and a solution can simply be found using the pseudoinverse, directly providing the solution to a required for reconstructing the original signal (x). Because the solution to eq 3 is now easy to find, the “trick” of requiring a sparse solution is no longer required. The reason for using a data-derived tailored basis (instead of any other random transform matrix) is that the tailored basis provides an (optimal) approximation (Ψ_r) of the original signal matrix (X). Then, the direct solution for a provides a linear combination of this approximation that can be used to reconstruct x . If a random transform was used, matrix Ψ_r would not approximate the original data (X), and thus, the use of this basis for reconstructing x would lead to very poor results.

Therefore, as the solution for a no longer requires solving an optimization problem, the optimization of the measurements (C) that maximize the reconstruction accuracy of x under noisy conditions in y becomes feasible instead. This is typically conducted by minimizing the condition number of (θ), which indicates the degree to which an output vector is affected by the noise in the input vector when θ is used to transform the input vector. Although the condition number of (θ) can be minimized by correctly choosing C , this again leads to a combinatorial optimization problem that is intractable to solve for large problems.²⁸ Instead, the QR pivot matrix decomposition can be used to efficiently find a greedy near-optimal solution.²⁸

Regular QR factorization decomposes a matrix into a unitary Q matrix and an upper triangular R matrix as follows:

$$\Psi_r = QR \quad (8)$$

The introduction of pivot columns (C^T) as follows:

$$\Psi_r^T C^T = QR \quad (9)$$

forces the QR matrix to become better conditioned, which is the original criterion for choosing the measurement matrix (C). In essence, the pivoting procedure sequentially (i.e., greedily in descending order of importance) selects the rows in Ψ_r that contain the most columnar variance and, therefore, samples the signal's dominant modes the most efficiently.^{28,30} As such, these pivot points provide a near-optimal solution for the choice of C . In addition, QR factorization is a ubiquitous and fast algorithm, making it a very practical method for optimizing C .

QR factorization can be implemented using numerous methods, such as the Gram–Schmidt process or Householder reflections.³⁷ In this work, QR factorization was implemented using LAPACK (linear algebra routines in the Python package SciPy).³⁸ For simplicity, the detailed specifications of these methods are omitted herein. Interested readers are referred to the literature²⁸ for more details.

Using a tailored basis, measurements can be optimized to reconstruct entire signals with highest accuracy and robustness possible.

2.3. The Singular Value Decomposition. As mentioned in the previous subsection, many data-driven transforms can be used as a tailored basis for optimizing the choice of measurements for compressed sensing–based signal reconstruction. Because of its general applicability to a variety of different problems, the singular value decomposition (SVD)^{28,31} was used in this work instead of other decompositions, which are typically the most useful for specific

types of systems or data sets. In addition, the SVD is guaranteed to exist for any matrix and is, therefore, a reliable method for generating tailored bases.²⁸

SVD-based matrix decomposition can provide low-rank approximations of high-dimensional matrices. The primary application of the SVD is in dimensionality reduction, allowing for large data sets to be accurately represented by much smaller approximation matrices. However, the SVD has been applied in many other areas, such as forming the foundation for other decomposition methods and as the tailored basis in compressed sensing.²⁸

For an original data set in the following matrix form:

$$\mathbf{X} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_m^T] \quad (10)$$

where each column in $\mathbf{X}$ comprises an individual signal ($\mathbf{x}$) of length n (such as a reshaped image, an audio waveform, or a set of experimentally measured properties of a single plastic sample). The SVD decomposes the $n \times m$ matrix $\mathbf{X}$ into the following three matrices, which, when multiplied, return $\mathbf{X}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \quad (11)$$

where $\mathbf{U}$ is an $n \times n$ left-singular matrix, $\mathbf{V}$ is an $m \times m$ right-singular matrix, $\mathbf{\Sigma}$ is an $n \times m$ matrix, and $*$ indicates the complex conjugate transpose. $\mathbf{U}$ and $\mathbf{V}$ are both unitary matrices (meaning that their conjugate transposes are also their respective inverses, such that, for example, $\mathbf{U}^*\mathbf{U} = \mathbf{U}\mathbf{U}^* = \mathbf{I}$) containing orthonormal columns. $\mathbf{\Sigma}$ is a matrix comprising zeros everywhere apart from on its diagonal, which contains the singular values of the SVD.²⁸

For all matrices larger than 2×2 , the SVD must be calculated using an iterative approach. One such example is the one-sided Jacobian algorithm, which uses a Jacobian matrix to iteratively transform the original matrix to one containing orthogonal columns.³⁹ The approach used in this work, is a numerical approach that makes use of the facts that the left-singular columns in $\mathbf{U}$ are orthonormal eigenvectors of matrices $\mathbf{X}\mathbf{X}^*$, the right-singular vectors $\mathbf{V}$ are orthonormal eigenvectors of the matrix $\mathbf{X}^*\mathbf{X}$, and the singular values of $\mathbf{\Sigma}$ are the square roots of the nonzero eigenvalues of both $\mathbf{U}$ and $\mathbf{V}$.²⁶ More specifically, this work makes use of the LAPACK implementation of these routines (accessed through the Python package NumPy) for solving the SVD.³⁸

The SVD can be interpreted as being a matrix analog of the Fourier analysis of periodic time signals. A Fourier series decompose complex waveforms into sums of differently weighted sine and cosine functions, where, with each additional term in the sum, the original waveform is approximated more accurately. Similarly, the SVD approximates complex matrices, where each additional column and row in $\mathbf{U}$ and $\mathbf{V}$, respectively, scaled by corresponding singular values in $\mathbf{\Sigma}$ increasingly accurately approximate the original matrix ($\mathbf{X}$) on which SVD was based. Thus, like Fourier transform, the SVD allows for the determination of a low-dimensional approximation (i.e., the dominant modes) of a complex high-dimensional signal. However, although the Fourier transform is universal, the SVD directly extracts the low-dimensional approximation from the data.

In addition, matrix $\mathbf{U}$ can be used as an effective tailored transform for optimizing the measurement location in compressed sensing to transform an original high-dimensional data set to a lower-dimensional approximation, similar to a Fourier or wavelet transform.²⁸

2.4. Experimental Methods. To demonstrate the proposed methods, ABS compounds were prepared by mixing six grades of SAN (possessing various molecular weights and acrylonitrile percentages) with seven rubber loadings. A standard additive package, including stabilizers and a mold release agent, was used in all the formulations. The compounds were prepared using a ZSK-25 twin-screw extruder (length: 46 L/D) operating at 250 rpm and 17 kg/h.

Tensile bars and plaques were molded using a Demag Ergotech 80/420–200 injection-molding machine.

Molecular weights were determined using gel permeation chromatography with an Agilent 1260 high-performance liquid chromatography system. The styrene, acrylonitrile, and rubber contents were measured using Fourier-transform infrared (FTIR) spectroscopy with a PerkinElmer Spectrum One spectrometer. Tensile and flexural tests were performed according to ISO 527 and ISO 178 standards, respectively, using a Zwick/Roell Z010 tensile testing machine. Charpy notched impact tests were conducted according to the ISO 179–1eA standard using a Zwick/Roell HIT5.SP pendulum impact tester. Falling-dart impact properties were measured according to the ISO 6603–2 standard using a Zwick/Roell HIT230F drop-weight tester. Melt volume indices (MVIs; 220 °C; 10 kg) were measured according to the ISO 1133 standard using a Zwick/Roell Aflow extrusion plastometer. Vicat softening temperatures were measured according to the ISO 306 standard using a Zwick/Roell HDT/Vicat A apparatus.

3. RESULTS

In this section, the results obtained using the methods described in the previous section are discussed. In the first subsection, the problem is defined, and the associated data sets are described. Then, the main error metrics used to evaluate the methods' property reconstruction accuracies are briefly described. Finally, the last three subsections discuss the results of the different analyses performed, which investigate the impact of various factors on the outcomes of the method.

3.1. Problem Definition and Data Description. The characterization of the chemical and mechanical properties of plastics (that is, additive-modified polymers, such as ABS) is key in determining the effectiveness of the plastic for different practical applications. Ideally, all the possible (or, at least, the most relevant) properties of a plastic are measured to obtain the optimal insights into the plastic's practical utility. However, the measurement of these properties is both time and cost intensive, and, in many cases it is economically not feasible to measure a large number of properties for all samples involved in a process.

Of course, because they depend on the plastic's chemical composition, many properties are related. Therefore, the value of one property may be predicted based on the values of others, allowing for a reduction in the number of properties that must be measured. Using compressed sensing, not only can an entire set of properties be reconstructed from a known smaller subset of these properties but through the use of a tailored basis it is even possible to determine the optimal set of measurements that maximize the reconstruction accuracy. This enables the following question to be answered: If only five of the 18 properties are measurable, which five should be measured to most accurately predict the remaining 13 nonmeasured properties?

In this work, a data set comprising extensive measurements of 18 ABS properties each for 42 samples was used to

demonstrate the compressed-sensing methods. Table 1 provides an overview of all the properties in the data set alongside some basic data statistics.

Table 1. Overview and Basic Statistics of All the Properties in the ABS Dataset^a

property	mean	st. dev.	min.	max.
butadiene %	26.00	6.07	17.00	35.00
Mn	63.96	12.16	48.00	82.90
tensile yield stress	43.68	3.84	37.57	51.56
tensile elongation at yield	2.19	0.07	2.04	2.33
tensile rupture stress	34.17	2.95	29.56	40.06
tensile elongation at rupture	16.35	4.07	9.10	28.00
tensile modulus	2423.83	244.68	2043.00	2843.00
flexural strength	69.81	6.78	59.48	83.76
strain at flexural strength	4.56	0.13	4.37	4.90
flexural modulus	2420.59	238.52	1987.80	2881.20
Charpy impact	19.71	7.87	5.75	33.33
Dart impact (energy to peak)	23.70	10.36	1.52	34.08
Dart impact (energy to break)	31.46	15.24	1.73	50.50
displacement to peak	10.98	2.72	4.90	14.17
displacement to break	13.17	3.91	5.17	17.72
force at peak	3618.64	869.58	903.63	4499.43
MVI	28.26	18.12	8.72	85.66
Vicat	96.07	3.24	91.80	113.30

^aMn and MVI refer to the number-average molecular weight of the polymer and melt volume index, respectively.

To show the potential for predicting certain property values based on the values of others, Figure 4 provides a scatterplot matrix where all the properties are plotted against each other. As can be seen, numerous properties show clear linear or nonlinear relationships to other properties. For instance, the flexural strength is clearly correlated with the tensile yield stress. As such, by learning the appropriate relations between the properties, all the properties can be reconstructed using only a limited set of the key properties. Compressed sensing not only enables such property reconstruction in a rapid and relatively easily implementable manner but also allows for the optimization of the known properties for most accurately reconstructing the other properties.

Mn does not have any trivial relationships with any of the other properties. Because many properties have been proven to have an asymptotic relationship with Mn, the relationship disappears above a certain threshold.³² A relationship was previously reported between viscosity (η) and Mn for ABS, where $\log(\eta)$ plotted as a function of $\log(\text{Mn})$ exhibited a slope of 3.64 (i.e., η scales with $\text{Mn}^{3.64}$).³³ In this study, Mn simultaneously increased with changes in the copolymer ratios, which affected the average monomer friction coefficient (ζ_0), which, in turn, influenced both the viscosity and MVI.³⁴ Therefore, the reported relationship is more complicated to obtain. Nevertheless, in line with other reports, with increasing butadiene content, MVI decreased.³⁵ The tensile modulus and strength both decreased with increasing butadiene content and increased with increasing styrene and acrylonitrile contents. The impact strength had the opposite relationships, as previously reported.^{35,36}

To demonstrate the tailored-basis compressed-sensing workflow, the data set was divided into two subsets: one where all 18 properties were used to create the singular value

decomposition matrices, and the other for evaluating the property reconstruction accuracy. For the test subset, only the properties determined using the optimal measurement selection workflow were known; the remaining properties were reconstructed using the SVD, and the property reconstruction accuracy was evaluated. Numerous analyses were conducted to evaluate the effects of different factors, including varying the number of measured properties used to reconstruct the remaining properties, the effect the size of the data set used to create the SVD matrices has on performance, and the impact of scaling the data in various ways.

3.2. Error Metric. In this work, because the number of included properties was relatively high and the properties covered a wide range of values, the mean absolute percentage error (MAPE) was used to evaluate the property reconstruction accuracy of the compressed-sensing method. The MAPE calculates the error as a percentage of the true value and, thus, provides an easily interpretable number independent of property magnitudes. The MAPE is defined as follows:

$$\text{MAPE} = \frac{\sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}}{n} \times 100\% \quad (12)$$

where y_i and $\hat{y}_i$ are the true and predicted values of sample i , respectively, and n is the number of samples.

As the MAPE provides dimensionless relative errors, it facilitates easier comparisons between the properties. However, the exclusive use of the MAPE can be risky, as it suffers a lot from outliers, is not symmetrical (although underpredictions cannot exceed 100% of the error, overpredictions can become arbitrarily large), and cannot be calculated for true values of 0 (since it leads to division by zero issues). In addition, for true values near 0, the MAPE can become very large for minor absolute errors.

The analysis results revealed that these disadvantages negligibly influenced the study's outcomes and conclusions. Moreover, because the results obtained using the mean absolute error (MAE) and MAPE metrics showed the same main trends, only the results obtained using the MAPE are included herein. For completeness, the Supporting Information provides Figures S1–S3, showing all the analysis results obtained using the MAE instead of the MAPE.

3.3. Influence of Number of Property Measurements on Property Reconstruction Accuracy. The main factor influencing the property estimation accuracy is the number of measured properties available to reconstruct the remaining unknown properties. Likely, the more properties that are measured, the better the reconstruction for the unknown properties. In this study, tailored-basis compressed sensing was used to determine the properties to measure for different numbers of measurements. For each number of measurements, the average reconstruction accuracy per property was provided, and the results were compared to quantify the influence of the number of properties to measure on the property reconstruction accuracy.

To set up the SVD decomposition matrices used to optimize the properties to measure and reconstruct the remaining properties, all 18 properties of 21 out of the total 42 available samples were used. All 21 samples were randomly selected in a stratified manner to ensure they properly represented all the ABS types in the data set. The remaining 21 samples were used for evaluating the property reconstruction accuracy. Only the properties selected using the tailored-basis optimization were

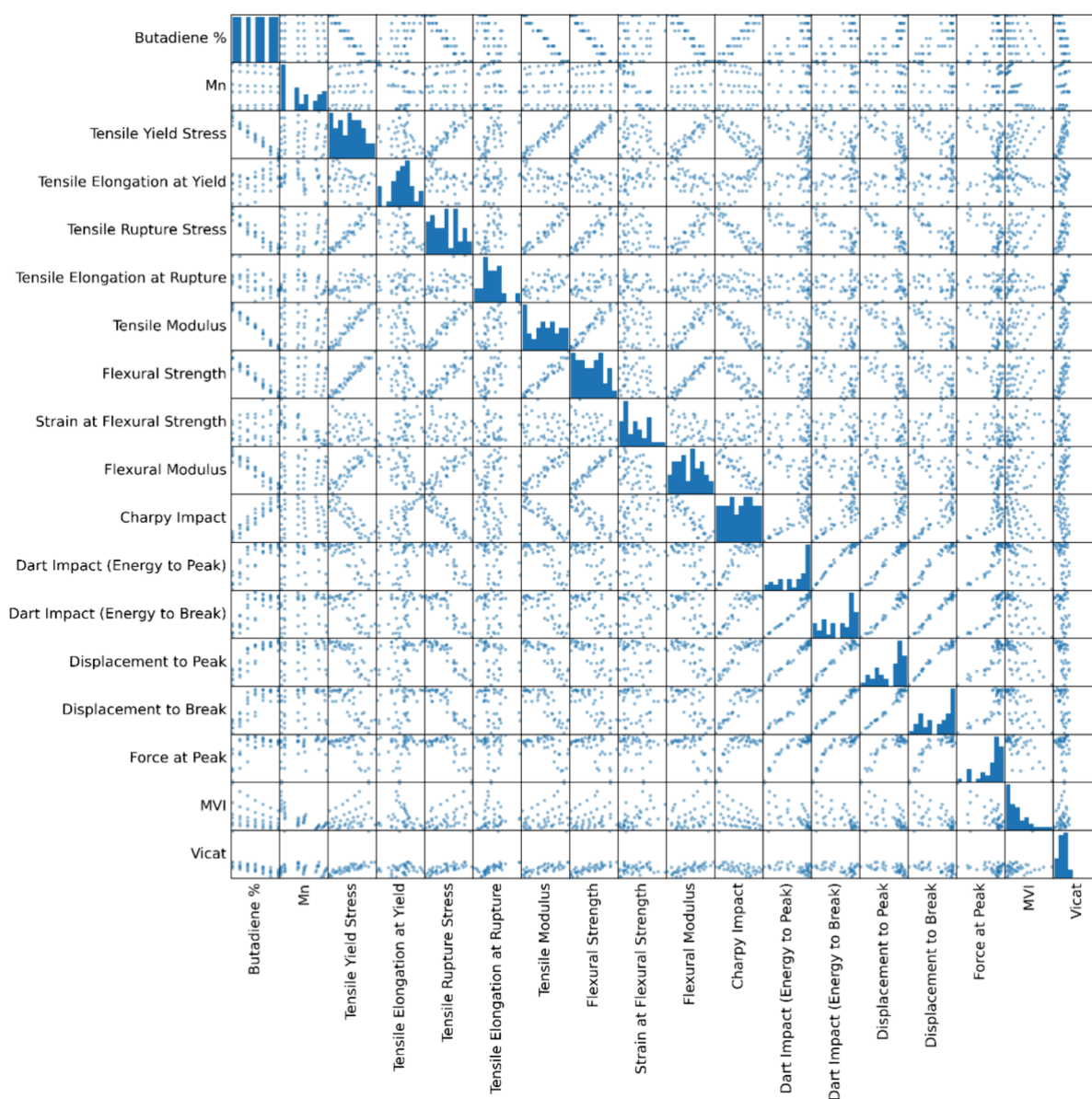


Figure 4. Scatterplot matrix of all the properties in the ABS data set.

known, while the remaining properties were predicted using these known values and the SVD matrix.

Figure 5 provides an overview of the optimal properties that the tailored-basis property reconstruction method selected for different sizes of the measured property set. To interpret the information in Figure 5, the number of measured properties is listed at the bottom. Then, for each number of measurements, all the elements shown in green in the corresponding column indicate the properties the method selects to include in the set of measured properties. Thus, when only one property is measurable, that property should be force at peak; if two properties are measurable, they should be the force at peak and tensile modulus; if five properties are measurable, they should be the force at peak, tensile and flexural moduli, MVI, and Mn.

Clearly, the method usually simply expands the existing set of measured properties. However, in a few instances, the method initially adds a property to the list before choosing another property instead when the total number of measured properties increases, indicating that the optimal solution might not always be simply adding more measurements but may also

depend on the specific combination of measured properties. Thus, when the method determines that another property combines more favorably with the other measured properties (for the current total number of property measurements), a property that was previously included might sometimes be excluded.

Notably, the characterization results are not derived from standalone measurements. For instance, using a simple tensile measurement, modulus, strain at break, stress at break, strain at yield and stress at yield can all be measured at once. Thus, even though the algorithm helps choosing one individual property over the other (to be measured), practically, they might still have to be measured together.

According to the first few properties, including the peak force, tensile and flexural moduli, Mn, and MVI, that the method selects, some choices are more easily interpretable than others. For instance, as shown in Figure 4, both the flexural and tensile moduli are clearly correlated with many of the other properties. As such, the method logically selects these, as they can provide insights into the values of many

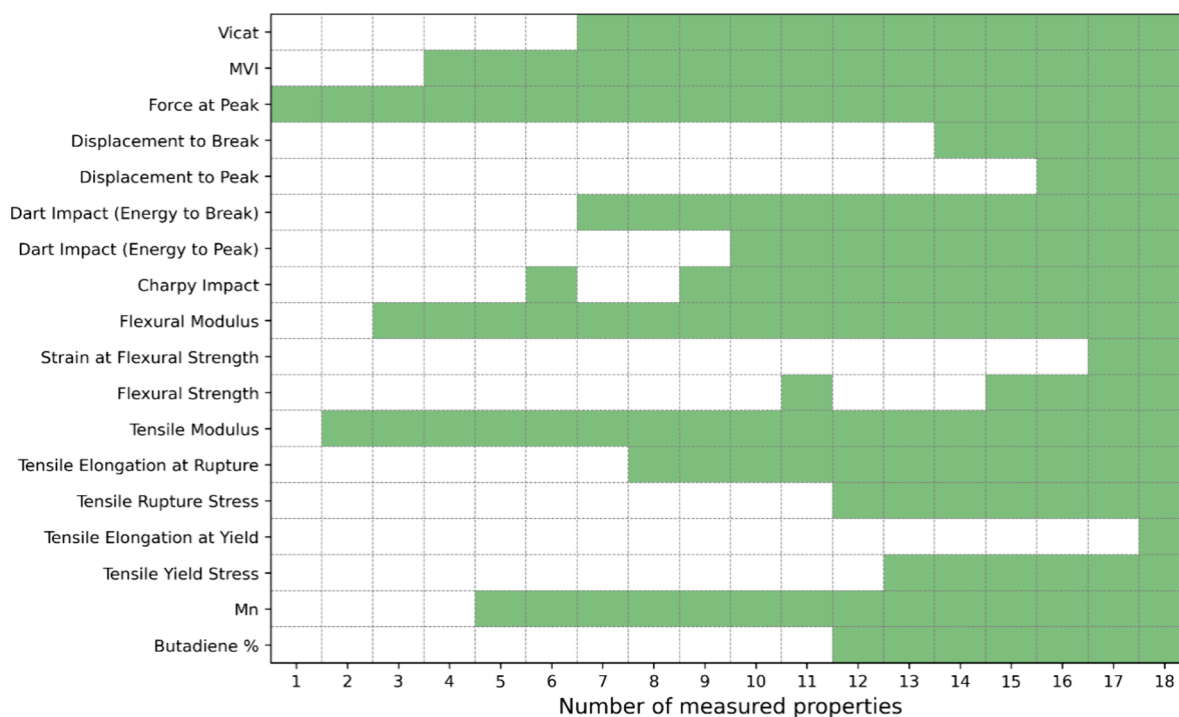


Figure 5. Properties that compressed sensing indicate are optimal to include in sets of different total numbers of measured properties.

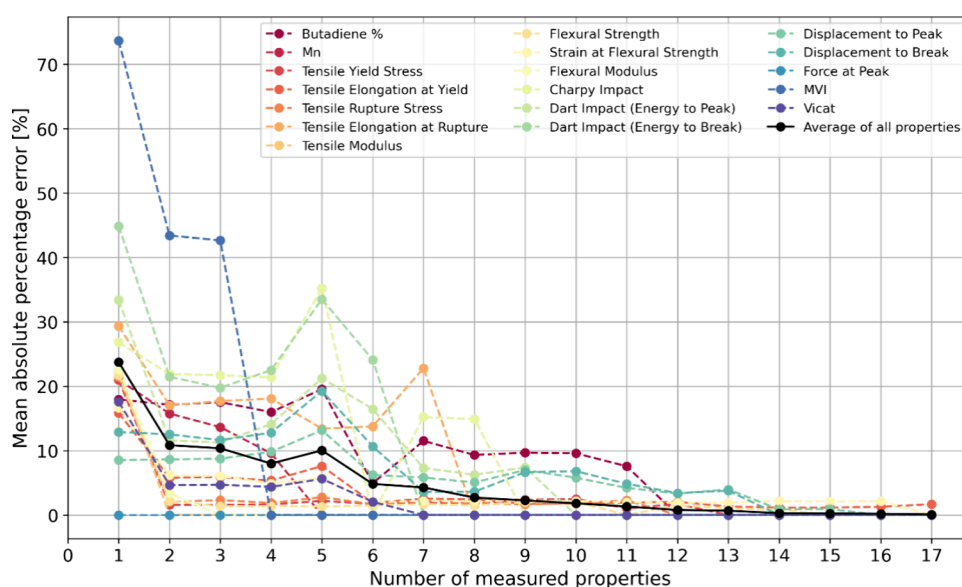


Figure 6. Mean absolute percentage error (MAPE) plotted as functions of the total number of properties measured for reconstructing all the properties.

other properties. Similarly, the force at peak property is also clearly correlated with other properties not covered as well by the tensile and flexural moduli, which somewhat explains why the method chooses it. However, because the MVI and Mn are only slightly correlated with the other properties, understanding why the method chooses these properties is less intuitive. One reason could be that because these and other properties are negligibly correlated, the method can only accurately reconstruct these properties by including them in the list of properties to measure.

In addition, the method clearly emphasizes some properties the least, including the tensile elongation at yield, strain at flexural strength, and displacement to peak. As shown in Figure

4, for a high number of measurable properties, the reason the method chooses to include these properties is difficult to understand because they are negligibly correlated with the other properties; thus, the method should prioritize their selection. However, as shown in Figure 6, these properties are already relatively accurately reconstructed, even after only a few measurements. As such, the method prioritizes the selection of other properties over these ones.

Another interesting aspect of how the method chooses properties is related to the property magnitudes. Table 1 shows that many of the earliest-chosen properties possess relatively high magnitudes, with the first three chosen properties possessing the highest magnitudes. Similarly, most of the

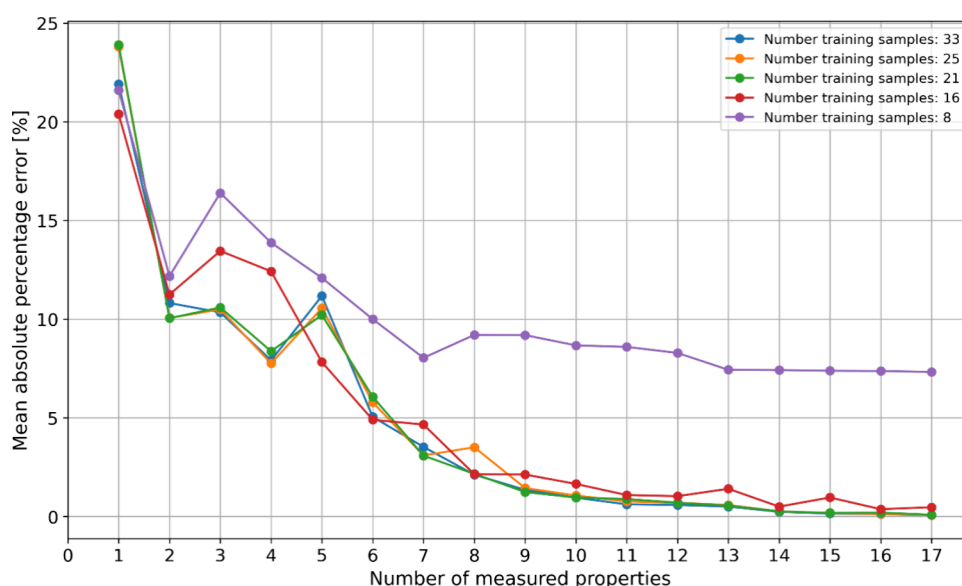


Figure 7. Mean absolute percentage error (MAPE) plotted as functions of the total number of measured properties for different training data set sizes used for reconstructing all the properties.

last-chosen properties possess relatively low magnitudes. Thus, while the method does not just choose only based on magnitude, it does seem to have a big impact on the preferred properties. One simple reason is that because the method aims to reduce the effects of noise on the data reconstruction, this is the most easily accomplished by selecting high-magnitude properties whose absolute levels of noise are comparatively higher than those for low-magnitude properties. Therefore, as described in Section 3.5, the analysis was repeated for a data set where all the properties were normalized in the same range to investigate the impact of the normalization on the method's decisions about the properties to choose.

An examination of the properties selected by the method (in addition to the data presented in Figure 4) provides some insight into why the method might have selected these properties. However, as this method is data driven, most of the results are attributed to the underlying data structure and, therefore, case specific and inherently difficult to interpret. Thus, these findings should not be generalized or extrapolated to other plastics without first evaluating the compressed sensing method's performance on them.

To further investigate the property reconstruction accuracy, Figures 6 and S1 (in the Supporting Information) show the MAPEs and MAEs plotted as functions of the number of properties measured to reconstruct the 18 properties, respectively. As expected, the more property measurements that were included, the more accurate the property reconstruction was, especially for the overall combined average. However, there are a few interesting deviations from this general behavior. Most notably at the 5 measured property mark, there is an increase in error, both for the average error as well as for the errors of individual property. Because tailored-basis compressed sensing is data driven, this behavior can be difficult to fully explain. One reason could be that because the measurements are imperfect, the addition of a measured property introduces additional noise to the available measured property data, which could slightly worsen the overall property reconstruction accuracy. The fact that these deviations are potentially related to noise in the data is further substantiated by the fact that when a different random seed is used to divide

the data between training and test sets, the deviations either appear for different measurement set sizes, or are not present at. That said, when present, the effects of this noise on performance are typically minor.

Ultimately, these results show that by measuring only six of the 18 properties, the average reconstruction error for all the properties was approximately 5%. However, the variation in the property reconstruction error for the individual properties was still rather high, with some properties still exhibiting reconstruction errors above 20–30%. By increasing the number of property measurements to nine, the overall average error percentage was further decreased to below 3%, while all the individual properties exhibited mean error percentages below 10%. Finally, for 12 property measurements, the average error decreased to 1%, and all the individual properties were reconstructed to within a 4% error. Beyond 12 property measurements, although the overall average error would continue improving (and eventually reaches 0% when all the properties are included in the measured set), the number of measured properties would be so high that the trade-off between the number of additional measured properties and error improvement might no longer be worthwhile.

3.4. Influence of Training Data Set Size on Property Reconstruction Accuracy. This section describes the effect of the data set size on the selection of the optimal set of properties to measure and reconstruct the missing properties. The previous methods were used with different “training” data set sizes. To ensure a fair comparison between the various data set sizes, the set of samples used for evaluating the property reconstruction error was identical for each analysis.

Figure 7 shows the MAPE plotted as functions of the size of the measured property data sets. For simplicity, only the average MAPE over all properties are given (corresponding to the black line in Figure 6). Figure S2 in the Supporting Information shows the MAE plotted as functions of the size of the measured property data sets. As expected, with increasing training set size, the property reconstruction accuracy improved. For six or more measured properties, the MAPE was almost identical for training sets comprising 33, 25, 21, and, to a lesser extent, 16 measured properties.

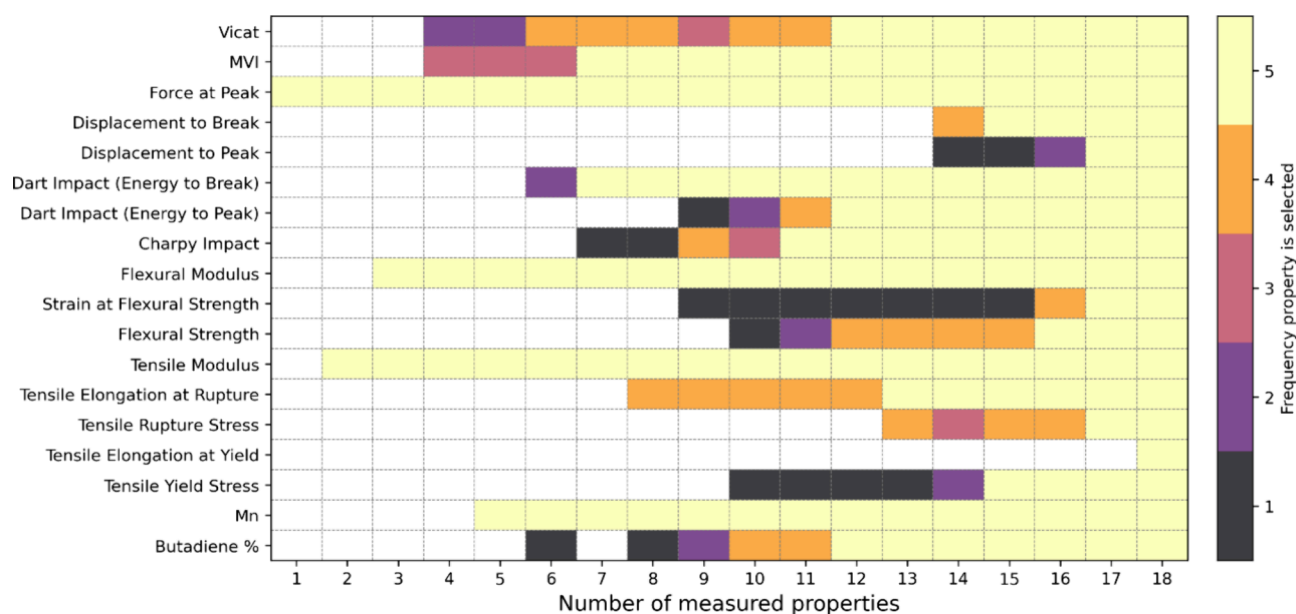


Figure 8. Frequency at which each property is selected depending on the size of the training set for increasing size of the set of measured properties.

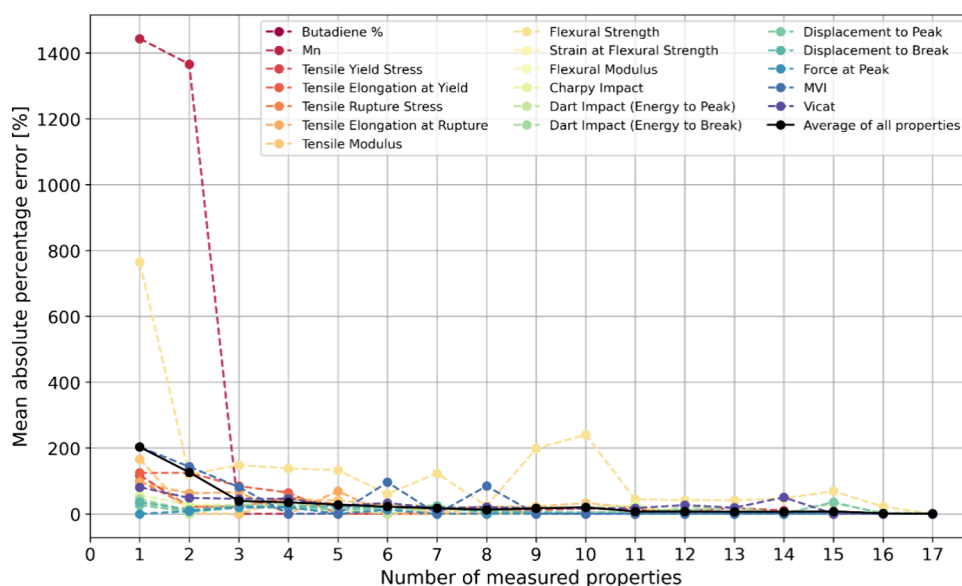


Figure 9. Mean absolute percentage error (MAPE) plotted as functions of the total number of measured properties for reconstructing all the properties using the same data set as shown in Figure 6 but scaled such that all the properties are in the range (0,1).

Notably, both the 8 sample and 16 sample data set sizes stop significantly improving in MAPE at a certain point. This is due to the nature of the tailored basis compressed sensing method itself, in which the number of measured properties the method can optimize for is limited by the size of the training set. For instance, when the training data set is limited to only eight data samples (e.g., eight fully characterized plastic samples), the method can only optimize eight measured properties. Thus, when the method is used to select more than eight properties, only the first eight are optimized, and the others are simply selected in the order in which they are originally defined in the data set. As such, as can be seen in Figure 7, the reconstruction accuracy for the training data set of size 8 stops improving after 8 measurements have been included.

The mathematical reason for the fact that the method stops improving is that the number of measurements that can be optimized is limited to the number of singular values that can be extracted from the training data in X . As the number of singular values corresponds one-to-one to the number of samples in X (X and Σ are both $n \times m$ matrices, where m is both the numbers of data samples in X and singular values on the diagonal in Σ), the number of training samples directly limits the method's effectiveness. Therefore, when less training samples are present than the number of measurements that must be optimized, not all the measurements are optimizable.

The training data set's size not only influences the property reconstruction accuracy of the tailored-basis compressed sensing but also directly impacts the properties the method selects as the optimal combinations for reconstructing

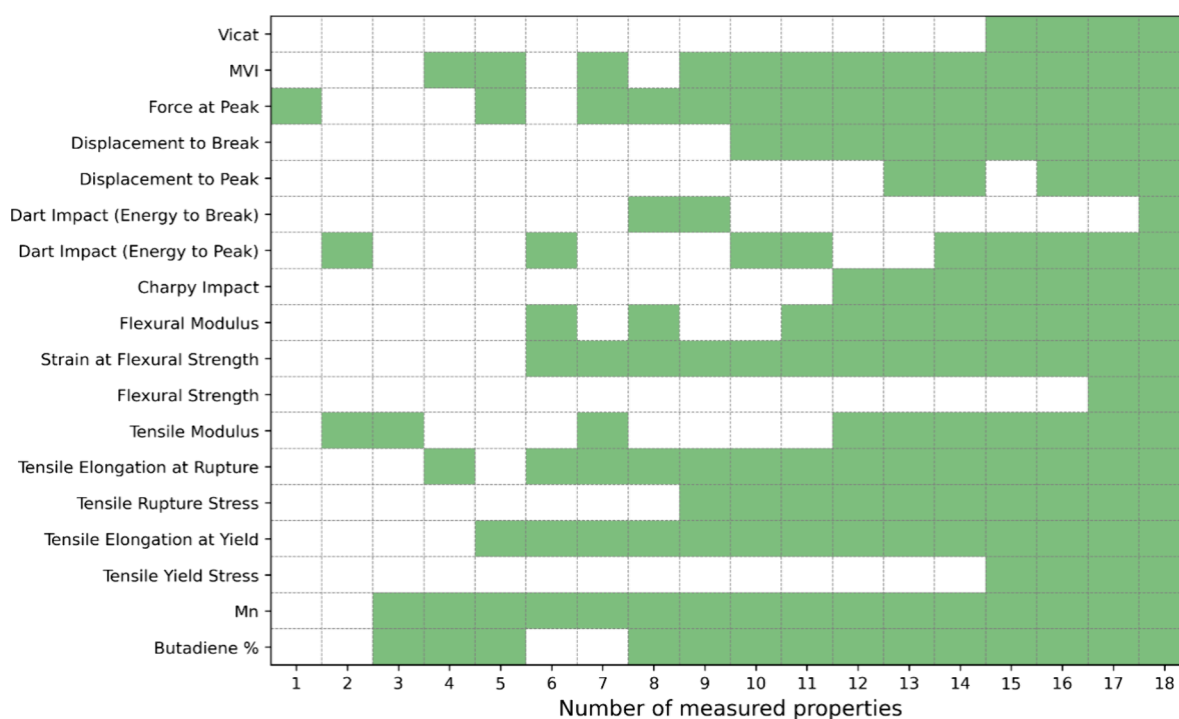


Figure 10. Properties that compressed sensing indicates are optimal to include in sets of measured properties for different total numbers of measured properties using the same data as shown in Figure 5 but scaled such that all the properties are in the range (0,1).

properties. To investigate the importance of this impact, the number of times each property was selected based on the size of the measured property set is plotted in Figure 8. Essentially, the same analysis as presented in Figure 5 was performed for all five training set sizes, and the results were summed and combined into one figure. Thus, if a property was always in the set of measured properties for a given set size, the property was scored as 5. A good example is the force at peak property, which was selected first for all the training set sizes. Similarly, the tensile and flexural moduli were always selected second and third, respectively. Additional variation appears in, for instance, the Vicat and MVI properties. For these properties the plot indicates that the MVI property is selected fourth for three out of five data set sizes, while in the other two out of five of these cases the Vicat property was selected fourth instead. On the other hand, although the strain at flexural strength was usually chosen as one of the final properties, in one case it was already selected when the size of the measured property set was nine. It is also clear that the method considered the tensile elongation at yield as being the least important property for reconstructing properties, as this property was consistently added last to the set of measured properties.

Figure 8 shows that the most (and least) important properties to measure are mostly independent from the training set size. In contrast, the properties that fall somewhere in between showed considerably more variation. However, as shown in Figure 7, in this range, the property reconstruction accuracies were mostly independent from the training set size, indicating that once the most important properties have been included, the specific combination of the other properties is less important to the overall property reconstruction error.

3.5. Influence of Relative Property Magnitude. Section 3.3 discussed how tailored-basis compressed sensing typically selects high-magnitude properties first. This phenomenon was further investigated by applying compressed sensing to the

same data scaled such that the absolute magnitudes of all the properties were equal. The data were scaled using the following standard minimum–maximum (minmax) approach, where for each property, the minimum value was subtracted from all the other values before they were divided by the difference between the maximum and minimum property values:

$$x_{\text{scaled}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

This produced a data set where each property value was in the range (0,1).

Figure 9 shows the MAPEs of the property reconstructions using the minmax-scaled data (the corresponding MAEs are shown in Figure S3 in the Supporting Information). A comparison of the results obtained using the scaled and nonscaled data revealed that the method's overall property reconstruction errors were much worse using the scaled rather than the nonscaled data, with the initial overall error being approximately 200% and Mn reaching an error of approximately 1400%. Although this overall error did decrease with increasing number of property measurements, many more measurements were required using scaled data rather than nonscaled data. For example, using scaled and nonscaled data, 12 and four measurements were required respectively before the overall error decreased below 10%, and 16 and six measurements were required before it decreased below 5%. Furthermore, as shown in Figure 10, the chosen property sets were very inconsistent. Often, the method selected, quickly deselected, and then reselected properties. As such, the results were much less robust for scaled rather than nonscaled data.

Thus, these results show that predictions made using tailored-basis compressed sensing are quite sensitive to data scaling and that nonscaled data reconstruct properties far more accurately than scaled data. Although not reported herein, the same analysis was performed for the same data scaled using a

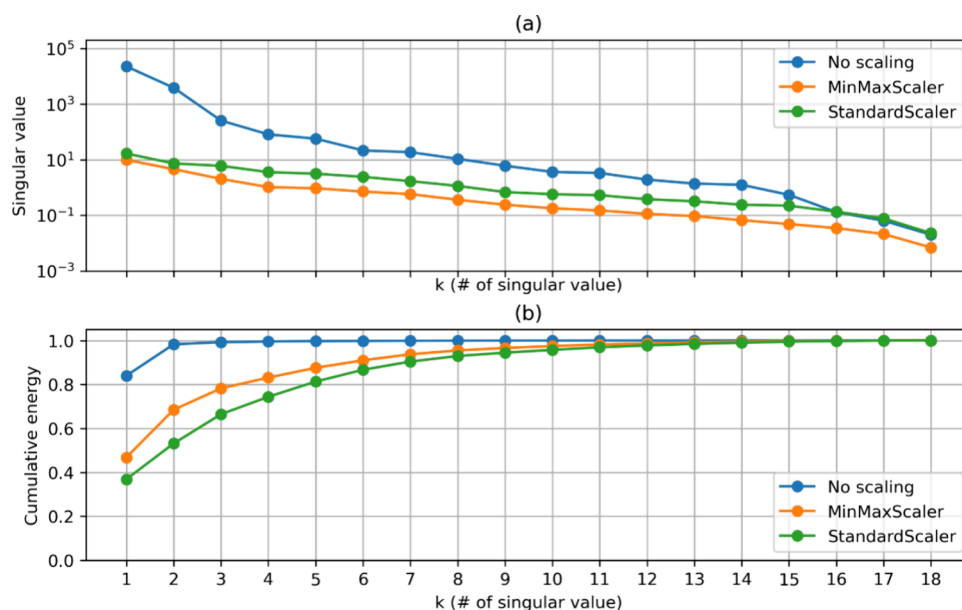


Figure 11. (a) First 18 singular values and (b) corresponding cumulative energies of the SVD for differently scaled data sets.

standard-scaling approach (by subtracting the mean of the property data before dividing by the standard deviation), which also gave poor and unstable results.

To help explain why scaled data produced far worse property reconstruction accuracies than nonscaled data, the first 18 SVD singular values produced using nonscaled, minmax-scaled, and standard-scaled data are plotted in Figure 11a. In addition, the plot also shows the respective cumulative energies (indicating the fraction of the total sum achieved by adding each subsequent singular value to all the previous ones) for the different scaling approach (Figure 11b). For example, for nonscaled data, the first singular value contributed approximately 80% of the total sum of all the singular values.

When the SVD is used for reducing the number of data, the magnitude of each singular value determines the degree to which the original data set is represented by the corresponding columns and rows of matrices U and V , respectively (Section 2.3). Therefore, if the first few singular values are much higher than the remaining ones (i.e., they provide a very high percentage of the cumulative energy), fewer columns and rows of U and V , respectively, are required to accurately approximate the original data set (similar to how in a Fourier transform, certain frequencies might dominate most of the others, such that the original signal can be accurately approximated using only a few terms).

Similarly, when the SVD is used for compressed sensing, if the first few singular values are dominant, far fewer property measurements are required to accurately reconstruct properties. As such, as shown in Figure 11b, for nonscaled data, the cumulative energy was already close to 1 when fewer than five singular values (corresponding to fewer than five measurements) were included. However, for both minmax- and standard-scaled data, either 12 or 13 measurements, respectively, were required to approximate a cumulative energy of 1, which explains why the method performs far worse using these data.

Finally, as for why the singular value distribution changes so much when scaling the data, this is a known (and primary) downside of the SVD: it can be very sensitive to the data format in X and depends heavily on the data's coordinate

system.²⁸ Thus, when the columns in X are translated, rotated, or scaled (which, of course, happens when a scaling transformation is applied), all the resulting U , V , and Σ matrices can change drastically (and often detrimentally),²⁸ as clearly revealed in the singular values shown in Figure 11a.

However, if the importance for data processing is considered and the proposed method is consciously applied to consistently generated and processed data, this SVD downside does not necessarily impact the method's practical applicability.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions. Because the characterization of plastic properties is time-consuming and costly, only as few properties as required should be measured. As such, data-driven predictive methods have attracted considerable research interest, as they allow for the prediction of unknown properties based on known properties by directly learning (nonlinear) correlations between properties from example data. This allows them to be used to fill gaps in nonmeasured properties solely based on the values of properties that have been measured.

In this work, one such data-driven method was demonstrated for reconstruction of a large number of plastic properties based on measurements of a much smaller subset of these properties. The method was based on compressed sensing using a tailored basis constructed from a limited set of fully characterized plastic samples (for which all the properties were measured). Tailored-basis compressed sensing enabled the optimization of a set of properties to measure such that the remaining nonmeasured properties can be estimated at maximum accuracy and robustness. In contrast to many other data-driven methods, the compressed-sensing framework is easy to set up, requires no hyperparameter tuning, and can accurately reconstruct properties from very few "training" data. In addition, the use of a tailored basis allows for the efficient optimization of the measured properties to reconstruct the nonmeasured properties as accurately as possible.

The method was demonstrated using a data set comprising 42 ABS samples possessing a total of 18 measured properties. One-half of this data set was used for developing the tailored-basis SVD, while the other half was used for evaluating the

property reconstruction accuracy. The use of this data set revealed that when only six properties were measured (force at peak, flexural and tensile moduli, MVI, Mn, and Charpy impact), all 18 properties were reconstructed within an average of 5% error (although some properties were more difficult to reconstruct than others). By adding three additional properties (energy to break, Vicat, and tensile elongation at rupture) to the set of measured properties, increasing the number of measured properties to nine, the error was further reduced to below 3% on average.

The force at peak and flexural and tensile moduli were the most important properties to measure using compressed sensing, as they were always the first properties selected for most analyses. On the other hand, the tensile elongation at yield, strain at flexural strength, and displacement to peak were typically selected last, most likely because these properties were substantially correlated with many of the other properties and, thus, provided less additional information for reconstructing properties.

An analysis of the effect of the data set size (used to develop the tailored basis) on the property reconstruction accuracy revealed that properties were accurately reconstructed using as few as 16 samples. However, the use of too few samples substantially reduced the property reconstruction accuracy, as compressed sensing only allows for the optimization of as many measurements as the number of samples used to develop the tailored basis.

Finally, properties were the most accurately reconstructed using unscaled rather than scaled data. Data scaling, such that all the properties were in the range (0,1), detrimentally affected the SVD, and the method's property reconstruction accuracy substantially dropped. Thus, to practically apply the proposed method, the data should be carefully handled and consistently gathered and used.

In theory, the proposed method only requires one to fully characterize a specific plastic for a limited number of samples (the exact number will, of course, vary); however, in as shown in this work, approximately 20 samples were required to be measured to accurately reconstruct the 18 properties. Of course, this will initially incur substantial monetary costs and time investment but must be performed only once. Moving forward, according to the outcomes obtained by applying compressed sensing, a substantially smaller set of properties can be characterized to estimate the values of the other nonmeasured properties, eventually saving considerable costs and time. Additionally, if another plastic must be analyzed, although the full characterization must be repeated for the additional plastic, this is likely to be an upfront cost that will eventually pay for itself in the long term.

4.2. Recommendations. The focus of this work has been on optimizing individual properties to find the best combination of properties to measure and estimate the remaining nonmeasured properties. However, in reality many properties are measured simultaneously with a single test. For instance, in a tensile test, the force at peak, displacement at break, and displacement at peak are simultaneously measured. Thus, in future work, the optimization of not only individual properties but also entire tests would be interesting and could, in turn, further reduce the requirement for certain equipment and its associated costs.

In addition, the results of this work show that measurement noise can substantially impact the reconstruction error. In

future work, this effect should be investigated and substantiated in a more detailed and structural manner.

■ ASSOCIATED CONTENT

Data Availability Statement

Although the source code for implementing the proposed compressed-sensing method is not publicly available, the Electronic Supporting Information contains a detailed description of the Python packages and versions used to develop the method. In addition, a pseudocode snippet is provided for the method's core, covering everything from the loading of the data to the reconstruction of the properties of a single sample from a limited number of property measurements. In addition, all the data produced and used in this study are publicly available at the following repository: https://github.com/poortjp/eABS_data_JCIM.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00622>.

Computational methods used for the compressed sensing approach in the form of pseudo-code; plots of the mean absolute error (MAE) for the reconstruction of each property as a function of the total number of properties measured, MAE for the reconstruction of all properties as a function of the total number of properties measured for different sizes of the training data set, and MAE for the reconstruction of each property as a function of the total number of properties measured for the same data set as used in the main paper but scaled in two different manners (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Milad Golkaram – Netherlands Organization for Applied Scientific Research (TNO), 3584 CB Utrecht, The Netherlands; Circular Plastics, Department of Circular Chemical Engineering, Faculty of Science and Engineering, Maastricht University, 6200 MD Maastricht, The Netherlands; orcid.org/0000-0002-6627-4683; Email: milad.golkaram@tno.nl

Authors

Jonah Poort – Netherlands Organization for Applied Scientific Research (TNO), 3584 CB Utrecht, The Netherlands

Pieter Janssen – Trinseo Netherlands BV, 4542 NH Hoek-Terneuzen, The Netherlands

Jan Harm Urbanus – Netherlands Organization for Applied Scientific Research (TNO), 3584 CB Utrecht, The Netherlands

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00622>

Funding

The project has received funding from the Grant Agreement number: 101058636—ABSolEU—HORIZON-CL4—2021.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the contribution of Trinseo and their support in providing the samples and the characterization thereof. The authors are grateful for the support of Lisanne

Heemskerk, Annemieke Runstraat, and Jorg Roosma. The authors express their gratitude to the European Health and Digital Executive Agency (HADEA) under the power delegated by the European Commission for its support of the project ABSoleU — ‘Paving the way for an ABS recycling revolution in the EU’.

REFERENCES

- (1) Joo, C.; Park, H.; Kwon, H.; Lim, J.; Shin, E.; Cho, H.; Kim, J. Machine Learning Approach to Predict Physical Properties of Polypropylene Composites: Application of MLR, DNN, and Random Forest to Industrial Data. *Polymers (Basel)* **2022**, *14* (17), 3500.
- (2) Yan, C.; Li, G. The Rise of Machine Learning in Polymer Discovery. *Adv. Intell. Syst.* **2023**, *5* (4), 2200243.
- (3) Khan, P. M.; Rasulev, B.; Roy, K. QSPR Modeling of the Refractive Index for Diverse Polymers Using 2D Descriptors. *ACS Omega* **2018**, *3* (10), 13374–13386.
- (4) Xu, P.; Chen, H.; Li, M.; Lu, W. New Opportunity: Machine Learning for Polymer Materials Design and Discovery. *Adv. Theor. Simul.* **2022**, *5* (5), No. 2100565.
- (5) Yu, M.; Shi, Y.; Liu, X.; Jia, Q.; Wang, Q.; Luo, Z. H.; Yan, F.; Zhou, Y. N. Quantitative Structure-Property Relationship (QSPR) Framework Assists in Rapid Mining of Highly Thermostable Polyimides. *Chem. Eng. J.* **2023**, *465*, No. 142768.
- (6) Ishikiriyama, K. Polymer Informatics Based on the Quantitative Structure-Property Relationship Using a Machine-Learning Framework for the Physical Properties of Polymers in the ATHAS Data Bank. *Thermochim. Acta* **2022**, *708*, No. 179135.
- (7) Andraju, N.; Curtzwiler, G. W.; Ji, Y.; Kozliak, E.; Ranganathan, P. Machine-Learning-Based Predictions of Polymer and Postconsumer Recycled Polymer Properties: A Comprehensive Review. *ACS Appl. Mater. Interfaces* **2022**, *14* (38), 42771–42790.
- (8) Krauklis, A. E.; Karl, C. W.; Rocha, I. B. C. M.; Burlakovs, J.; Ozola-Davidane, R.; Gagini, A. I.; Starkova, O. Modelling of Environmental Ageing of Polymers and Polymer Composites-Modular and Multiscale Methods. *Polymers (Basel)* **2022**, *14* (1), 216.
- (9) Patra, T. K. Data-Driven Methods for Accelerating Polymer Design. *ACS Polym. Au* **2022**, *2* (1), 8–26.
- (10) Chung, E.; Russo, D. P.; Ciallella, H. L.; Wang, Y. T.; Wu, M.; Aleksunes, L. M.; Zhu, H. Data-Driven Quantitative Structure-Activity Relationship Modeling for Human Carcinogenicity by Chronic Oral Exposure. *Environ. Sci. Technol.* **2023**, *57* (16), 6573–6588.
- (11) El-Hadi, A.; Schnabel, R.; Straube, E.; Müller, G.; Henning, S. Correlation between Degree of Crystallinity, Morphology, Glass Temperature, Mechanical Properties and Biodegradation of Poly (3-Hydroxyalkanoate) PHAs and Their Blends. *Polym. Test.* **2002**, *21* (6), 665–674.
- (12) Qiu, H.; Zhao, W.; Pei, H.; Li, J.; Sun, Z. Y. Highly Accurate Prediction of Viscosity of Epoxy Resin and Diluent at Various Temperatures Utilizing Machine Learning. *Polymer (Guildf)* **2022**, *256*, No. 125216.
- (13) Okafor, C. E.; Okafor, E. J.; Ikebodu, K. O. Evaluation of Machine Learning Methods in Predicting Optimum Tensile Strength of Microwave Post-cured Composite Tailored for Weight-Sensitive Applications. *Eng. Sci. Technol. An Int. J.* **2022**, *25*, No. 100985.
- (14) Wu, F. Y.; Yin, J.; Chen, S. C.; Gao, X. Q.; Zhou, L.; Lu, Y.; Lei, J.; Zhong, G. J.; Li, Z. M. Application of Machine Learning to Reveal Relationship between Processing-Structure-Property for Polypropylene Injection Molding. *Polymer (Guildf)* **2023**, *269*, No. 125736.
- (15) Altarazi, S.; Allaf, R.; Alhindawi, F. Machine Learning Models for Predicting and Classifying the Tensile Strength of Polymeric Films Fabricated via Different Production Processes. *Materials (Basel)* **2019**, *12* (9), 1475.
- (16) Cravero, F.; Schustik, S. A.; Martínez, M. J.; Díaz, M. F.; Ponzoni, I. How Can Polydispersity Information Be Integrated in the QSPR Modeling of Mechanical Properties? *Sci. Technol. Adv. Mater. Methods* **2022**, *2* (1), 1–13.
- (17) Kumar, N.; Ukey, P. D.; Francis, V.; Singh, R. P.; Sahu, S. Chapter 16 Plastic Pellets. In *Plastics Design Library*; Izdebska-Podsiadly, J. B. T. P. for 3D P, Ed.; William Andrew Publishing, 2022; 307–323.
- (18) Bhaskar, R.; Butt, J.; Shirvani, H. Investigating the Properties of ABS-Based Plastic Composites Manufactured by Composite Plastic Manufacturing. *J. Manuf. Mater. Process.* **2022**, *6* (6), 163.
- (19) P, J.; Pionteck, J.; Hässler, R.; George, S. M.; Cvelbar, U.; Thomas, S. Studies on Stress Relaxation and Thermomechanical Properties of Poly(Acrylonitrile-Butadiene-Styrene) Modified Epoxy-Amine Systems. *Ind. Eng. Chem. Res.* **2011**, *50* (8), 4432–4440.
- (20) Rossato, J. H. H.; Lemos, H. G.; Mantovani, G. L. The Influence of Viscosity and Composition of ABS on the ABS/SBS Blend Morphology and Properties. *J. Appl. Polym. Sci.* **2019**, *136* (8), 47075.
- (21) Veeman, D.; Sudharsan, S.; Surendhar, G. J.; Shanmugam, R.; Guo, L. Machine Learning Model for Predicting the Hardness of Additively Manufactured Acrylonitrile Butadiene Styrene. *Mater. Today Commun.* **2023**, *35*, No. 106147.
- (22) Stoll, A.; Benner, P. Machine Learning for Material Characterization with an Application for Predicting Mechanical Properties. *GAMM-Mitteilungen* **2021**, *44* (1), No. e202100003.
- (23) Manoni, L.; Turchetti, C.; Falaschetti, L.; Crippa, P. A Comparative Study of Computational Methods for Compressed Sensing Reconstruction of EMG Signal. *Sensors (Basel)* **2019**, *19* (16), 3531.
- (24) Sreenivas, T. V.; Kleijn, W. B. *Compressive Sensing for Sparsely Excited Speech Signals*; IEEE: Taipei, Taiwan, 2009.
- (25) Poh, K. K.; Marziliano, P. Compressive Sampling of EEG Signals with Finite Rate of Innovation. *EURASIP J. Adv. Signal Process.* **2010**, *2010* (1), No. 183105.
- (26) Golub, G. H.; Kahan, W. Calculating the Singular Values and Pseudo-Inverse of a Matrix. *SIAM J. Numer. Anal. Ind. Appl. Math.* **1965**, *2*, 205–224.
- (27) Upadhyaya, V.; Salim, M. Compressive Sensing: Methods, Techniques, and Applications. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1099* (1), No. 012012.
- (28) Brunton, S. L.; Kutz, J. N. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, 2019.
- (29) Smith, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing*; Scientific Research Publishing, 1999; 261–344.
- (30) Manohar, K.; Brunton, B.; Kutz, J.; Brunton, S. Data-Driven Sparse Sensor Placement for Reconstruction: Demonstrating the Benefits of Exploiting Known Patterns. *IEEE Control Syst. Mag.* **2018**, *38* (3), 63–86.
- (31) Kühl, N.; Fischer, H.; Hinze, M.; Rung, T. An Incremental Singular Value Decomposition Approach for Large-Scale Spatially Parallel & Distributed but Temporally Serial Data—Applied to Technical Flows. *Comput. Phys. Commun.* **2024**, *296*, No. 109022.
- (32) Fox, T. G., Jr.; Flory, P. J. Second-Order Transition Temperatures and Related Properties of Polystyrene. I. Influence of Molecular Weight. *J. Appl. Phys.* **1950**, *21* (6), 581–591.
- (33) Dreval, V. E.; Vasil'ev, G. B.; Borisenkova, E. K.; Kulichikhin, V. G. Rheological and Mechanical Properties of ABS Plastics Prepared by Bulk Polymerization. *Polym. Sci. Ser. A* **2006**, *48* (3), 338–345.
- (34) Martinetti, L.; Macosko, C. W.; Bates, F. S. Rouse-Bueche Theory and the Calculation of the Monomeric Friction Coefficient in a Filled System. *J. Polym. Sci. B Polym. Phys.* **2016**, *54* (15), 1437–1442.
- (35) Jin, F. L.; Lu, S. L.; Song, Z. B.; Pang, J. X.; Zhang, L.; Sun, J. D.; Cai, X.-P. Effect of Rubber Contents on Brittle-Tough Transition in Acrylonitrile-Butadiene-Styrene Blends. *Mater. Sci. Eng., A* **2010**, *527* (15), 3438–3441.
- (36) Melton, G. H.; Peters, E. N.; Arisman, R. K. Engineering Thermoplastics. *Applied Plastics Engineering Handbook: Processing and Materials*; Elsevier, 2011; 7–21.
- (37) Golub, G.; Golub, G. H.; Van Loan, C. F. *Matrix Computations*; The Johns Hopkins University Press: Baltimore, 2013; 721.

- (38) LAPACK — Linear Algebra PACKage. <https://www.netlib.org/lapack/> (accessed 2024-07-17).
- (39) de Rijk, P. P. M. A one-sided Jacobi algorithm for computing the singular value decomposition on a vector computer. *SIAM J. Sci. Stat. Comput.* **1989**, *10*, 359–371.