

RESEARCH ARTICLE 

# A data-driven method for automated data superposition with applications in soft matter science

Kyle R. Lennon<sup>1</sup> , Gareth H. McKinley<sup>2</sup> and James W. Swan<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup>Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

**Corresponding author:** Gareth H. McKinley; Email: [gareth@mit.edu](mailto:gareth@mit.edu)

**Received:** 30 March 2022; **Revised:** 21 September 2022; **Accepted:** 01 March 2023

**Keywords:** Bayesian statistics; Gaussian process regression; method of reduced variables; self-similarity


## Abstract

The superposition of data sets with internal parametric self-similarity is a longstanding and widespread technique for the analysis of many types of experimental data across the physical sciences. Typically, this superposition is performed manually, or recently through the application of one of a few automated algorithms. However, these methods are often heuristic in nature, are prone to user bias via manual data shifting or parameterization, and lack a native framework for handling uncertainty in both the data and the resulting model of the superposed data. In this work, we develop a data-driven, nonparametric method for superposing experimental data with arbitrary coordinate transformations, which employs Gaussian process regression to learn statistical models that describe the data, and then uses maximum a posteriori estimation to optimally superpose the data sets. This statistical framework is robust to experimental noise and automatically produces uncertainty estimates for the learned coordinate transformations. Moreover, it is distinguished from black-box machine learning in its interpretability—specifically, it produces a model that may itself be interrogated to gain insight into the system under study. We demonstrate these salient features of our method through its application to four representative data sets characterizing the mechanics of soft materials. In every case, our method replicates results obtained using other approaches, but with reduced bias and the addition of uncertainty estimates. This method enables a standardized, statistical treatment of self-similar data across many fields, producing interpretable data-driven models that may inform applications such as materials classification, design, and discovery.

## Impact Statement

The dynamics of certain physical systems exhibit parametric self-similarity, meaning that experiments conducted at different states produce data sets that can be superposed to produce a universally valid model. Performing this analysis by hand may lead to human bias, however, and does not acknowledge experimental uncertainty in the data, which can lead to unstandardized and poorly reproducible results. Here, we develop an automated method for shifting data sets onto a single curve, which combines Gaussian process regression and maximum a posteriori estimation to represent the experimental data and coordinate transforms statistically. Our method standardizes the popular practice of data superposition and the construction of master curves, accounts for uncertainty in the data, and provides meaningful and physically interpretable predictions.

The author, J.W.S., is deceased.

 This research article was awarded an Open Materials badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

### 1. Introduction

For many of the physical processes encountered in scientific and engineering studies, the underlying governing equations are unknown, oftentimes despite tremendous scientific effort. Even in systems for which it is possible to write down the governing equations, the dynamical response often remains too complicated to admit a compact analytical solution, and frequently even numerical solutions to such complicated equation sets are too cumbersome to form the basis of useful tools for the prediction or analysis of experimental observations. In such cases, insights and predictions are instead often made by using domain knowledge and heuristics to identify patterns in experimental data. These patterns may then inform empirical models that describe the behavior of systems quite accurately, over a wide range of conditions. With increasing popularity, these patterns and empirical models are being identified automatically by computational algorithms, a process commonly referred to as *machine learning* (Ferguson, 2017; Butler et al., 2018; Carleo et al., 2019).

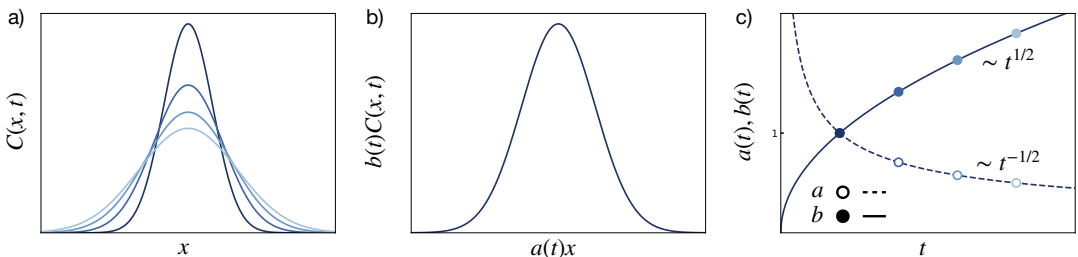
There are many ways in which patterns in data may inform useful mathematical models without any physical considerations. Here we consider one special case, which we exemplify by examining a simple physical system whose governing equations, and the analytical solution to those equations, are known: one-dimensional diffusion from an instantaneous point source. In this system, a fixed mass  $M$  of a passive scalar is released at the origin ( $x=0$ ) at time  $t=0$  in an infinite domain, and allowed to diffuse with a constant diffusion coefficient  $D$ , such that the concentration  $C$  of the diffusing species is governed by

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2}. \tag{1}$$

The analytical solution to this diffusion problem for all times  $t > 0$  is known

$$C(x,t) = \frac{M}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right). \tag{2}$$

However, consider if the underlying diffusion process and analytical solution for this system were *not* known. Instead, we have access to a (possibly noisy) device that measures the concentration profile  $C(x,t)$  at an instant in time. **Figure 1a** presents some of these fictional measurements. A trained researcher may notice that the concentration profiles at each measured time form a set of self-similar curves, each with a different width and maximum value. Noting this pattern, they might rescale the height of each concentration by a factor  $b(t)$ , and rescale the width by a factor  $a(t)$ , so that all the rescaled curves fall on top of one another. This superposition creates the so-called “master curve” shown in **Figure 1b**. The researcher



**Figure 1.** Example of fictitious measurements of the concentration profile of a diffusing species from an instantaneous point source. (a) The concentration profile was measured instantaneously at four different times, with lighter-shaded curves representing later times. (b) A “master curve” constructed by rescaling the width and height of the concentration profiles by time-dependent “shift factors”  $a(t)$  and  $b(t)$ , respectively. (c) The shift factors  $a(t)$  and  $b(t)$  plotted as a function of time. The earliest-time concentration profile is taken as the reference, so its shift factors are unity. The remaining shift factors exhibit the trends  $a(t) \sim t^{-1/2}$  and  $b(t) \sim t^{1/2}$ .

records the horizontal and vertical “shift factors,”  $a(t)$  and  $b(t)$ , respectively, and plots them as a function of time, noting that  $a(t) \sim t^{-1/2}$  and  $b(t) \sim t^{1/2}$ . Therefore, using ideas of self-similarity (Barenblatt, 2003) they propose the following model for the concentration profile:

$$C(x,t) = \frac{A}{\sqrt{t}} g\left(\frac{x}{\sqrt{t}}\right). \quad (3)$$

We recognize that, for this problem, the functional form for the master curve  $g(z)$  is a simple Gaussian in the reduced variable  $z \equiv x/\sqrt{t}$ . An approximation for this form might be determined from the rescaled data through the same sort of methods that determined forms for  $a(t)$  and  $b(t)$ , at which point a self-similar solution to the diffusion equation will have been constructed directly from data.

In nonlinear or higher-dimensional systems, it may not be possible to choose simple functional forms that describe a master curve or the shift factors  $a(t)$  and  $b(t)$ . However, parametric self-similarity in data—that is, the observation that we may construct a master curve by transforming discrete data observed at different “states” as illustrated in the diffusion example above—is a widespread and salient principle in the analysis of thermophysical property data in many fields, including rheology, solid mechanics, and material science (Leaderman, 1944; Tobolsky and Andrews, 1945; Ferry, 1980). In fact, the same principle of dynamical self-similarity finds applications in many different fields as well, such as electromagnetism and even finance (Wagner, 1915; Lillo et al., 2003).

Self-similar data sets may be employed to make accurate predictions of material characteristics at intermediate states, or to learn physical features about the underlying dynamics or structure of the system under study (de Gennes, 1979; Barenblatt, 2003), often times by traditional methods such as direct interpolation between data or by constitutive modeling. However, as demonstrated by the previous example, there is another method for analyzing self-similar data. This method is most accurately referred to as the “method of reduced variables,” but is commonly referred to as “data superposition” or the development of “master curves” (Leaderman, 1944; Ferry, 1980). When applicable, this method is potent, as it is agnostic to selection of an underlying constitutive model, and may encompass a broader and more physically robust set of hypotheses than direct interpolation between specific (possibly noisy) data. Thus, when new cases of data self-similarity are explored, data superposition can be immediately applied to gain physical insight and develop predictive tools, without the need to develop new models.

In many cases, including the diffusion example discussed above, an experiment measuring a property with parametric self-similarity  $C(x;t)$ , over an independent variable  $x$  and with a state parameter  $t$ , is related to a master curve  $g(z)$  in the reduced variable  $z \equiv a(t)x$  by the similarity relation:

$$C(x;t) = \frac{1}{b(t)} g(a(t)x) + ch(x). \quad (4)$$

Here  $a(t)$  and  $b(t)$  are state-parameter-dependent shift factors, while  $h(x)$  and  $c$  are a state-parameter-independent offset function and multiplicative constant, respectively. Together, these transformations collapse the experimental measurements onto the master curve (denoted generically  $g(z)$ ) for all values of  $x$ . Most often, the master curve is constructed from superposition of noisy experimental data. This process uses experimental data for  $C(x;t)$  taken for a finite set of state parameters  $\{t_j\}$  and over a fixed range of  $x$ . Then one determines the set of shift factors,  $\{a_j = a(t_j)\}$ ,  $\{b_j = b(t_j)\}$ , and  $c$  so that these finite data sets across experiments are brought into registry without advance knowledge of the functional form of the master curve  $g(z)$ .

Typically the process of master curve construction is done by an expert in the art who determines the shift factors,  $\{a_j\}$ ,  $\{b_j\}$ , and  $c$ , by eye. This can be done reliably, but it is a time-consuming process and may not be robust to experimental uncertainties. If one aims to construct master curves from large sets of data in an unbiased fashion—where bias could come in the form of presuming an existing model for any of  $g(z)$ ,  $a(t)$  or  $b(t)$ , or through the preconceptions and biases of the expert analyzing the data—then an algorithm for the automated construction of such master curves is necessary. In this work, we present such an algorithm relying on Gaussian process regression (GPR) and maximum a posteriori estimation. We

then apply the algorithm to examples using rheological data for soft material systems drawn from the existing literature.

At its core, the process of constructing a master curve from self-similar data represents a task in pattern recognition. Therefore, the automated method for learning coordinate transformations that result in a master curve presented in this work represents a form of *unsupervised learning* (Carleo et al., 2019)—a class of machine learning algorithms that identify patterns in data without explicit supervision, meaning that the algorithm does not require a set of labeled examples (in this case, a set of data sets labeled with the resulting master curves) for training. However, the patterns learned in this unsupervised context—the coordinate transformations and master curve—are themselves useful for making predictions. For instance, one can use the master curve and shift factors as a map  $x, t \mapsto C$  for values of  $x$  and  $t$  outside of the “training” set used to construct the master curve. This higher-level problem resembles *supervised learning*, a class of machine learning algorithms that use a set of labeled examples (here, measurements of  $C$  at particular  $x$  and  $t$ ) to train a predictive model (the master curve and shift factors). The method presented in this work differs from many so-called “black box” approaches to supervised learning, in that the predictive model is itself the result of a pattern recognition task, and may therefore provide value to the user beyond its ability to make accurate predictions. Thus, this method represents a versatile tool for both modeling and understanding the behavior of systems with parametric self-similarity.

### 1.1. Examples in soft matter science

Many thermomechanical properties of interest in soft materials satisfy principles of self-similarity in the time domain or with respect to the rate of deformation on variation of an appropriate state parameter (Plazek, 1965; Struik, 1977; Larsen and Furst, 2008; Gupta et al., 2012; Dekker et al., 2018; Caggioni et al., 2020; Lalwani et al., 2021), which may be attributed to the wide range of temporal and spatial scales that govern the underlying dynamics (Markovitz, 1975; de Gennes, 1979; Barenblatt, 2003). Some examples include self-similar curves for:

- the creep compliance as a function of time resulting from variation of the temperature in polymer melts (*time–temperature superposition*) (Plazek, 1965),
- the linear viscoelastic modulus as a function of frequency on variation of extent of reaction during gelation (*time–cure superposition*) (Larsen and Furst, 2008),
- the shear stress as a function of shear rate on variation of packing fraction in an emulsion (*shear rate–volume fraction superposition*) (Dekker et al., 2018),
- the relaxation modulus as a function of time on variation of the waiting time  $t_w$  since sample preparation in an aging clay suspension (*time–age–time superposition*) (Gupta et al., 2012).

It is often challenging to design rheometric experiments and equipment capable of measuring these evolving self-similar mechanical properties over the entire dynamic parameter range on which they vary. Instead, the principles of self-similarity (Barenblatt, 2003) and dynamic scaling (de Gennes, 1979), coupled with a judicious choice of physiochemical state parameters such as temperature or pH are used to expand the effective range of time scales accessible in experiments.

Furthermore, the construction of master curves and the interpretation of those curves plays a critical role in extrapolating from direct measurement of material properties across a limited number of experiments to fundamental knowledge of the microscale physical processes giving rise to these material properties. For instance:

- in time–temperature superposition,  $C(x;t)$  would be the linear creep compliance (or the relaxation modulus),  $x$  would be the temporal coordinate, the state variable  $t$  would be the temperature in a given set of experiments,  $a(t)$  and  $b(t)$  would be the horizontal and vertical shift factors that might be used to learn about the temperature dependence of relaxation processes in the soft material. In such a

- representation,  $c$  would be a fluidity contribution that is independent of the temperature, and  $h(x) = x$  would indicate linear creep at long times (Plazek, 1965),
- in the time-cure superposition principle that underpins gelation (Adolf and Martin, 1990),  $C(x;t)$  would be the complex modulus,  $x$  would be the frequency,  $t$  would be perhaps the concentration of gelator,  $a(t)$  and  $b(t)$  would be horizontal and vertical shift factors that might be used to learn about the mechanism of gelation,  $c$  would be an offset accounting for the solvent viscosity, and  $h(x) = ix$  would reflect this high-frequency viscous mode (Larsen and Furst, 2008),
  - in the shear rate-volume fraction superposition of an emulsion,  $C(x;t)$  would be the steady shear stress,  $x$  would be the imposed shear rate,  $t$  would be the packing fraction ( $\phi$ ) of the emulsion,  $a(t)$  and  $b(t)$  would be the horizontal and vertical shift factors that might be used to learn about the packing fraction dependence of the yield stress,  $c$  would be a high shear viscosity and  $h(x) = x$  would reflect Newtonian flow of the emulsion at very high shear rates (Dekker et al., 2018; Caggioni et al., 2020),
  - in time–age–time superposition of an aging clay suspension,  $C(x;t)$  would be the linear relaxation modulus,  $x$  would be the effective or material time coordinate ( $\tilde{t}$ ),  $t$  would be the wait time ( $t_w$ ) after mixing or preshear of the aging suspension,  $a(t)$  and  $b(t)$  would be the horizontal and vertical shift factors that might be used to learn about the rates of microstructural yielding in the suspension,  $c$  would be an elastic plateau modulus and  $h(x) = 1$  would reflect the nonaging elastic energy stored in the microstructure (Struik, 1977; Gupta et al., 2012).

Note that in this last case, the independent variable  $x$  represents an *effective*, or dynamically rescaled, time coordinate,  $x = \zeta(t; \nu)$ , representing a nonlinear transformation of the laboratory time coordinate  $t$  with some scaling exponent  $\nu$  (Joshi and Petekidis, 2018). Despite this added complexity, the method that we present in this work is sufficiently general to accommodate nonlinear coordinate transformations, and can indeed infer optimal values of the scaling parameters defining these transformations (such as the aging exponent  $\nu$ ) at the same time as the shift factors  $a(t)$  and  $b(t)$ .

The present work is organized as follows. In [Section 2](#), we develop the mathematics behind the algorithm for automatic construction of master curves. This includes a brief description of GPR, a development of maximum likelihood and maximum a posteriori estimation, and the discussion of Monte Carlo cross-validation (MCCV) for hyperparameter optimization. Then, [Section 3](#) applies this method to the four specific soft materials science examples listed in this section, which are of increasing complexity. Finally, [Section 4](#) demonstrates how the learned master curves, combined with the inferred shift factors  $a(t)$  and  $b(t)$ , may be used to inform forward predictions of data with automatic uncertainty estimates.

## 2. Method

A number of previous works have made efforts to automate the process of master curve construction. These include methods based on minimizing the mean squared error between data at every state and a single basis expansion (Honerkamp and Weese, 1993; Buttlar et al., 1998; Sihm and Tsai, 1999), methods for minimizing the area between linear interpolants defined by data at different states (Barbero and Ford, 2004; Gergesova et al., 2011; Gergesova et al., 2016), methods for minimizing the mean squared error between the derivatives of spline interpolants fit to the data (Hermida and Povolo, 1994; Naya et al., 2013), methods to minimize the arc length between data sets at different states (Cho, 2009; Maiti, 2016), and a method that leverages mathematical constraints on the Fourier transform of real-valued time-series data (Rouleau et al., 2013). However, only a few of these methods have been demonstrated for simultaneous horizontal and vertical shifting, and many require either parameters, an appropriate interpolant, or a set of basis functions to be specified by the user, thereby introducing elements of subjectivity into the methodology. Moreover, few of these methods explicitly account for the fact that experimental data possesses a finite amount of noise and measurement uncertainty, which may affect both the resulting master curve and confidence in the inferred shift factors. Because noise is not treated directly by these approaches, systematic parametric sensitivity analysis has required computationally intensive

approaches, such as bootstrap resampling (Maiti, 2019), which may require hundreds of thousands of evaluations of the full superposition algorithm.

Here, we propose an approach to automated master curve construction that is both data-driven and statistically motivated, thereby enabling rapid bidirectional superposition of data with automatic uncertainty quantification. We rely on GPR to infer a continuous and probabilistic description of the underlying data, and apply maximum a posteriori estimation to infer the best set of parameters that shift the noisy data onto a common master curve. This methodology is not limited to any specific form for the master curve, nor is it constrained to any specific form for the transformations applied to the data to obtain the master curve. Thus, it may be applied to cases of simple horizontal shifting, simultaneous horizontal and vertical shifting, as well as nonlinear coordinate transformations such as material time dilation in rheologically aging systems. Moreover, this approach is nonparametric, meaning that users need only supply data and the functional form of the transformations that should be applied to obtain the master curve. Finally, because we use a probabilistic description of the data and statistical comparisons of different data sets, the underlying noise in the data and the resulting uncertainties in the inferred shift factors are handled naturally within this framework.

In the remainder of this section, we will develop our approach, beginning with a brief discussion of GPR, then continuing to discussions of inference. We develop our maximum a posteriori approach by first considering the slightly simpler formalism of maximum likelihood estimation. We subsequently incorporate prior expectations about the shifting parameters  $a(t)$  and  $b(t)$  into the framework in the form of prior distributions, which turns the maximum likelihood estimates of the shift parameters to maximum a posteriori estimates.

### 2.1. Gaussian process regression

GPR—known in some fields as *kriging* (Matheron, 1963)—is a machine-learning method for obtaining a probabilistic model of a stochastic process from data (Rasmussen and Williams, 2006). GPR assumes that the data  $C(x;t)$  at fixed  $t$  are described by a Gaussian process (GP):

$$C(x;t) \sim \text{GP}(\mu(x;t), K(\theta, x, x';t)), \tag{5}$$

with a mean function  $\mu(x;t)$  and a covariance function, commonly referred to as the kernel of the GP,  $K(\theta, x, x';t)$ , between any two points  $x$  and  $x'$ , with a set of hyperparameters  $\theta$ . In this section, we use the symbol “ $\sim$ ” to signify that a variable (or series of variables) behaves according to some statistical distribution (or stochastic process). Informally, a GP represents a distribution in *function space*, or an ensemble of functions most likely to describe the data set. The form of the *prior* mean and covariance functions,  $\mu(x;t)$  and  $K(\theta, x, x';t)$  encode prior expectations about the functions in this distribution, such as noise structure, smoothness, and variations over  $x$ .

With prior expectations over fitting functions specified by the form of  $\mu(x;t)$  and  $K(\theta, x, x';t)$ , GPR proceeds by storing data points in the kernel matrix  $\mathbf{K}$  with  $K_{ij} = K(\theta, x_i, x_j; t)$ , where  $x_i$  and  $x_j$  represent values of  $x$  present in the input data set, and determining the hyperparameters  $\theta$  that maximize the log-marginal likelihood of observing the data from the GP model (Rasmussen and Williams, 2006). The GP model now represents a distribution over functions that fit the data set subject to prior expectations, with *posterior* mean  $m(x^*;t)$  and variance  $s(x^*;t)^2$  at an unmeasured point  $x^*$  describing a Gaussian distribution  $\mathcal{N}$  over the predicted value,  $C(x^*;t) \sim \mathcal{N}(m(x^*;t), s(x^*;t)^2)$ . For a GP with zero priormean ( $\mu(x;t) = 0$ ), these posterior mean and variance functions are:

$$m(x^*;t) = \underline{\mathbf{K}}(\theta, x^*, \underline{\mathbf{x}}; t) \mathbf{K}^{-1} \underline{\mathbf{C}}(\underline{\mathbf{x}}; t), \tag{6}$$

$$s(x^*)^2 = K(\theta, x^*, x^*; t) - \underline{\mathbf{K}}(\theta, x^*, \underline{\mathbf{x}}; t) \mathbf{K}^{-1} \mathbf{K}^T,$$

where  $\underline{\mathbf{x}}$  represents the vector of all values of  $x$  in the training data set,  $\underline{\mathbf{C}}(\underline{\mathbf{x}}; t)$  represents the vector with elements  $C_i = C(x_i; t)$  for all  $x_i$  in the training set, and  $\underline{\mathbf{K}}(\theta, x^*, \underline{\mathbf{x}}; t)$  represents the vector with elements  $K_i = K(\theta, x^*, x_i; t)$  for all  $x_i$  in the training set.

In this work, we assume a zero-mean ( $\mu(x;t)=0$ ) GP and a combination of constant, white noise, and rational quadratic kernels (Rasmussen and Williams, 2006):

$$K(\theta, x, x'; t) = \zeta(t) \left( 1 + \frac{(x - x')^2}{2\alpha(t)l(t)^2} \right)^{-\alpha(t)} + \beta(t) + \left( \sigma(t)^2 + \sigma_u^2 \right) \delta_{x,x'}, \tag{7}$$

where  $\theta = [\alpha(t), \beta(t), \zeta(t), l(t), \sigma(t)]$  are the hyperparameters which may depend on the state parameter  $t$ , and  $\delta_{x,x'}$  is the Kronecker delta function between points  $x$  and  $x'$ . Notably, the form of this kernel allows for the GP to automatically adapt to the level of noise in the data set, via the hyperparameter  $\sigma(t)$ . The state-independent hyperparameter  $\sigma_u$  represents an experimental uncertainty level which is not fit by GPR, but rather can be specified in advance to encode additional experimental uncertainty that is not evident in variations of the data (Ewoldt et al., 2015; Singh et al., 2019). Other authors have noted that a reasonable limit for the relative uncertainty level in high-quality rheological data is 4% (Freund and Ewoldt, 2015); therefore, unless otherwise stated, we take  $\sigma_u = 0.04$  when the GP model describes the logarithm of experimental data. This value may be assumed by any user to keep the method nonparametric; however, if the relative uncertainty of specific classes of experimental data is known in advance, then  $\sigma_u$  may be adjusted by the user.

We will not discuss the details of the fitting protocol for GPR or hyperparameter optimization here, but refer interested readers to the following references for more information on GP models and GPR: (Rasmussen and Williams, 2006; Duvenaud, 2014; Görtler et al., 2019).

## 2.2. Maximum likelihood estimation

### 2.2.1. Linearly scaled multiplicative shifting

The first step in our automated superposition algorithm is to fit each data set at distinct state  $t$  to its own GP model. Once this regression is complete, our goal is to optimally superpose these GP models to create a master curve. Because GPs are probabilistic in nature, it is sensible to define a statistical criterion for the optimal superposition.

To begin in developing this criterion, consider the problem of registering a single data point,  $(x_i^{(j)}, C_i^{(j)})$  from a data set at state  $t_j$  with a GP model trained on a data set at a different state,  $t_k$ . For simplicity, consider  $t_k$  as the reference state, such that all shift factors required to collapse the data onto the master curve are applied relative to this state. We apply a horizontal shift  $a_{jk}$  and vertical shift  $b_{jk}$  to the data point in  $t_j$ , bringing it to the coordinates  $(a_{jk}x_i^{(j)}, b_{jk}C_i^{(j)})$ . This shifted data point is now treated as an observation drawn from the GP model at state  $t_k$ :

$$b_{jk}C_i^{(j)} \sim \mathcal{N} \left( m_k \left( a_{jk}x_i^{(j)} \right), s_k \left( a_{jk}x_i^{(j)} \right)^2 \right). \tag{8}$$

The likelihood  $p_k \left( x_i^{(j)}, C_i^{(j)} | a_{jk}, b_{jk} \right)$  of this observation may be evaluated from the Gaussian probability density function taken from the GP at the coordinate  $a_{jk}x_i^{(j)}$ . It will be convenient to work with the negative of the logarithm of this likelihood (abbreviated NLL, for negative log-likelihood):

$$-\ln p_k \left( x_i^{(j)}, C_i^{(j)} | a_{jk}, b_{jk} \right) = \frac{1}{2} \left( \frac{b_{jk}C_i^{(j)} - m_k \left( a_{jk}x_i^{(j)} \right)}{s_k \left( a_{jk}x_i^{(j)} \right)} \right)^2 + \ln s_k \left( a_{jk}x_i^{(j)} \right) + \ln \sqrt{2\pi}. \tag{9}$$

The NLL defines a *loss function* for the superposition of the single data point  $(x_i^{(j)}, C_i^{(j)})$  at state  $t_j$  and the entire data set at state  $t_k$ , which is captured by the GP model with mean and variance functions  $m_k(x)$  and  $s_k(x)^2$ , respectively. The likelihood is conditioned on particular values of the shift factors  $a_{jk}$  and  $b_{jk}$  applied to data at state  $t_j$ ; thus, minimizing the NLL gives the maximum likelihood estimates of  $a_{jk}$  and  $b_{jk}$  for this particular data point.

Before we extend this approach to include an entire data set, it is worthwhile to consider what exactly the maximum likelihood estimate aims to achieve. We see that the first term in equation (9) is minimized when  $b_{jk}C_i^{(j)} = m_k(a_{jk}x_i^{(j)})$ —that is, when the data point at  $t_j$  is shifted directly onto the mean of the GP at  $t_k$ . This matches our intuitive sense of optimal superposition. However, this is not the only term in the NLL that is affected by the shift parameters. The second term in equation (9) penalizes shifting the data point to coordinates at which the GP has large variance. This term will become particularly important when we shift entire data sets, as it prioritizes superposing two data sets in the region with the least uncertainty in their GP models. Because the covariance in our selected GP kernel depends on the Euclidean distance between an unmeasured coordinate  $x^*$  and the measured coordinates  $x'$ , uncertainty in the GP model is typically greatest in regions where data are sparse, or in regions outside of the range of measured data. Penalizing superposition in these regions relative to regions in which data are dense provides a natural means for regularizing excessive extrapolation.

Equation (9) develops the loss function for a single data point, based on maximum likelihood estimation. This loss may be applied independently to each data point  $(x_i^{(j)}, C_i^{(j)})$  in the data set at state  $t_j$ , which we denote  $\mathcal{D}_j$ . We are now interested in the likelihood of observing this entire data set from the GP model at state  $t_k$ , equivalent to the joint likelihood of observing each data point from that GP. If we treat the data points in  $\mathcal{D}_j$  as independent observations, then this joint likelihood is simply the product of the individual likelihoods. The joint NLL is therefore the sum of the individual NLLs:

$$-\ln p_{jk}(\mathcal{D}_j|a_{jk}, b_{jk}) = \sum_{(x_i^{(j)}, C_i^{(j)}) \in \mathcal{D}_j} -\ln p_k(x_i^{(j)}, C_i^{(j)}|a_{jk}, b_{jk}). \tag{10}$$

This equation now defines a loss function over an entire data set, rather than a single data point. It is not yet the final loss function to consider for obtaining a maximum likelihood estimate of the shift parameters, however. Equation (10) describes the likelihood of observing the data set  $\mathcal{D}_j$  to align with the GP model at state  $t_k$ , but as we shift  $\mathcal{D}_j$ , we are similarly bringing the data set at the state  $t_k$ , denoted  $\mathcal{D}_k$ , close to the GP model trained at state  $t_j$ . The final loss function should reflect the symmetric nature of this shifting. Thus, we may define the counterparts to equations (9) and (10):

$$-\ln p'_j(x_i^{(k)}, C_i^{(k)}|a_{jk}, b_{jk}) = \frac{1}{2} \left( \frac{C_i^{(k)}/b_{jk} - m_j(x_i^{(k)}/a_{jk})}{s_j(x_i^{(k)}/a_{jk})} \right)^2 + \ln s_j(x_i^{(k)}/a_{jk}) + \ln \sqrt{2\pi}, \tag{11}$$

$$-\ln p'_{jk}(\mathcal{D}_k|a_{jk}, b_{jk}) = \sum_{(x_i^{(k)}, C_i^{(k)}) \in \mathcal{D}_k} -\ln p'_j(x_i^{(k)}, C_i^{(k)}|a_{jk}, b_{jk}). \tag{12}$$

The shift parameters  $a_{jk}$  and  $b_{jk}$  now divide  $x_i^{(k)}$  and  $C_i^{(k)}$ , to bring them from state  $t_k$  to state  $t_j$ , and the mean and variance functions  $m_j(x^*)$  and  $s_j(x^*)$  are now taken from the GP trained at  $t_j$ .

To complete the loss function, we simply take the joint likelihoods over data sets  $\mathcal{D}_j$  and  $\mathcal{D}_k$ , again assuming independence, such that the likelihood function that we observe both data sets from the same master curve given shifting parameters  $a_j$  and  $b_j$  is:

$$-\ln p(\mathcal{D}_j, \mathcal{D}_k|a_{jk}, b_{jk}) = -\ln p_{jk}(\mathcal{D}_j|a_{jk}, b_{jk}) - \ln p'_{jk}(\mathcal{D}_k|a_{jk}, b_{jk}). \tag{13}$$

Minimizing this joint NLL loss leads to the maximum likelihood estimates  $\hat{a}_{jk}$  and  $\hat{b}_{jk}$ , the shifting parameters for state  $t_j$  relative to state  $t_k$ :

$$\{\hat{a}_{jk}, \hat{b}_{jk}\} = \underset{a_{jk}, b_{jk}}{\operatorname{arg\,min}} [-\ln p(\mathcal{D}_j, \mathcal{D}_k|a_{jk}, b_{jk})]. \tag{14}$$

The preceding equations present a symmetric objective for registering two data sets onto each other. In practice, data sets at more than two states are present in the master set. It is straightforward to extend the NLL



loss in equation (13) to include more than two data sets, by taking the sum of every pairwise NLL possible between two data sets in the master set. However, this treatment requires optimization over all shifts  $\{a_{jk}\}$  and  $\{b_{jk}\}$  simultaneously, a high-dimensional global optimization problem whose complexity may grow exponentially with the number of states. Instead of shifting all data sets simultaneously, it is most computationally efficient to perform maximum likelihood estimation pairwise, developing a set of relative shift factors  $\{a_{jk}\}$  and  $\{b_{jk}\}$  between two curves with consecutive states:  $t_j$  and  $t_k = t_{j+1}$ . Once all relative shift factors have been computed, a global reference state may be selected, and the product or quotient of certain shift factors taken to obtain the global shifts. Although this sacrifices the more general calculation of the mastercurve from all data simultaneously, it reduces the complexity from exponential to linear in the number of states. In the coming sections, we will demonstrate how this approach may be used to solve more complex shifting problems in an efficient manner.

The problem of pairwise shifting requires optimizing a nonlinear function in two dimensions—the horizontal shift  $a_{jk}$  and the vertical shift  $b_{jk}$ . A close examination of equations (9) and (11) reveals that the optimization problems in these two dimensions are quite different. The nonlinearity in  $b_{jk}$  first arises from its inversion in equation (11), and in both cases by taking its square, whereas the nonlinearity in  $a_{jk}$  arises from the nonlinear functions  $s_j(x)$  and  $m_j(x)$  (or  $s_k(x)$  and  $m_k(x)$ ). While optimization in these dimensions may be performed simultaneously, a slight reformulation of the problem reveals a natural bilevel structure to the optimization problem, where the simpler problem of finding the optimal  $b_{jk}$  for a fixed value of  $a_{jk}$  is solved inside a higher-level optimization of  $a_{jk}$ . The following section presents this reformulation.

### 2.2.2. Logarithmically scaled multiplicative shifting

The problem formulation in this section has assumed that the data  $(x_i, C_i)$  represent linear-scaled coordinates, and that the desired shifting protocols involve multiplying these coordinates by factors  $a_{jk}$  and  $b_{jk}$  at each state pair:  $t_j, t_k$ . This multiplicative rescaling of data is the most common in many fields of the physical sciences. However, because of the wide underlying spectrum of time and length scales representing the dynamics of soft multiscale materials, rheological data is often represented on logarithmic scales, with the  $x$ -coordinate representing a temporal coordinate (or its inverse such as frequency or shear rate) that spans orders of magnitude, and with the response coordinate  $C$  (commonly a stress or modulus) spanning many orders of magnitude as well. In this case, we may replace the coordinates  $x_i$  and  $C_i$  in equations (5)–(13) with their logarithms,  $\ln x_i$  and  $\ln C_i$ . That is, the GP models are now regressed to the logarithm of the data, and the joint NLL loss is computed based on the logarithm of the data. In this case, the expressions for the rescaled data now involve addition of the logarithm of the shift factor, rather than direct multiplication by the shift factor:

$$x_i^{(j)} a_{jk} \rightarrow \ln x_i^{(j)} + \ln a_{jk}, \quad x_i^{(k)} / a_{jk} \rightarrow \ln x_i^{(k)} - \ln a_{jk}, \tag{15}$$

$$C_i^{(j)} b_{jk} \rightarrow \ln C_i^{(j)} + \ln b_{jk}, \quad C_i^{(k)} / b_{jk} \rightarrow \ln C_i^{(k)} - \ln b_{jk}. \tag{16}$$

This substitution is especially convenient in the case of the vertical shift factor, because it reduces the joint NLL in equation (13) to a quadratic expression in  $\ln b_{jk}$ . At a fixed value of the lateral shift  $\ln a_{jk}$ , optimization of the vertical shift factor  $\ln b_{jk}$  is now a convex problem whose global minimum can be found analytically. In fact, the minimizer  $\ln \hat{b}_{jk}(a_{jk})$  as a function of  $a_{jk}$  is given by:

$$\frac{1}{\sigma^2} = \sum_{x_i^{(j)} \in \mathcal{D}_j, s_k} \frac{1}{(\ln x_i^{(j)} + \ln a_{jk})^2} + \sum_{x_i^{(k)} \in \mathcal{D}_k, s_j} \frac{1}{(\ln x_i^{(k)} - \ln a_{jk})^2}, \tag{17}$$

$$\ln \hat{b}_{jk}(a_{jk}) = \sigma^2 \left[ \sum_{(x_i^{(j)}, C_i^{(j)}) \in \mathcal{D}_j} \left( \frac{\ln C_i^{(j)} - m_k(\ln x_i^{(j)} + \ln a_{jk})}{s_k(\ln x_i^{(j)} + \ln a_{jk})^2} \right) \right] \tag{18}$$

$$+ \sum_{(x_i^{(k)}, c_i^{(k)}) \in \mathcal{D}_k} \left( \frac{\ln C_i^{(k)} - m_j (\ln x_i^{(k)} - \ln a_{jk})}{s_j (\ln x_i^{(k)} - \ln a_{jk})^2} \right) \Bigg]. \tag{19}$$

Note that, since the GP models are trained on the logarithm of the original data,  $m_j$  and  $s_j$  predict the mean and standard deviation of the logarithm of the material response for state  $t_j$  (and similarly for  $t_k$ ). This minimizer function for  $\ln b_{jk}$  can be interpreted intuitively as choosing the weighted average of the differences between the data and the opposing GP model, where each difference is weighted inversely by its relative contribution to the total uncertainty in the GP predictions. Also, note that the means by which we apply the horizontal shifting do not affect the complexity of analytically computing the minimizer for the vertical shifts; therefore, we may use an analog of equation (19) with any type of horizontal shifting, as long as the vertical shift is multiplicative and performed on the logarithm of the data.

The realization that the maximum likelihood estimate for the vertical shift factor  $b_{jk}$  may be determined analytically for a specified horizontal shift  $a_{jk}$  is significant, as it reduces the pairwise shifting optimization problem from a 2D, nonconvex problem to two separate 1D optimization problems, one of which is convex and solved analytically (vertical shifting), and one that is nonconvex (horizontal shifting). This gives rise to a natural bilevel structure for the optimization problem. The inner problem of optimizing  $b_{jk}$  for a particular  $a_{jk}$  is fast, allowing for an efficient implementation of the outer problem of optimizing  $a_{jk}$ . There are many methods for solving this outer optimization problem. For all examples in this work, we solve it numerically by first generating an initial guess close to the global minimum via a grid search over feasible values for  $a_{jk}$ , and use this guess to seed a gradient-based algorithm (here, the BFGS algorithm).

### 2.2.3. Nonmultiplicative shifting

In some circumstances, developing a master curve requires performing transformations that are not simple multiplication by constant scaling factors  $a_{jk}$  and  $b_{jk}$ . In rheology, this may include, for example, subtracting out state-independent viscous contributions from steady flow curve data,  $\sigma - \eta\dot{\gamma}$  (Plazek, 1996), or a nonlinear “dilation” in the laboratory time coordinate due to power-law rheological aging,  $t^{1-\nu} - t_{\text{ref}}^{1-\nu}$  (Gupta et al., 2012; Joshi and Petekidis, 2018). Nonlinear rescaling of temporal and spatial coordinates is also common in the construction of similarity solutions in fluid dynamics and transport phenomena (Barenblatt, 2003; Eggers and Fontelos, 2015). In many such cases, the nonmultiplicative shifts are accompanied by subsequent multiplicative shifts. It is possible to generalize the maximum likelihood approach discussed previously to accommodate these more complicated data transformations, with the result again being a high-dimensional, nonconvex optimization problem, which is amenable to global optimization techniques such as simulated annealing. However, in many such cases, it is possible to leverage the computationally efficient technique developed in the previous section for multiplicative shifting to solve these more complex problems in less time.

To make use of the previous results for nonmultiplicative shifting, we may separate the data transformations into a set of nonmultiplicative transformations (such as subtracting a state-independent viscous mode, or applying state-independent time dilation) and a set of multiplicative transformations. Typically, there is a single nonmultiplicative transformation, which involves a single parameter (such as  $\eta$  or  $\nu$  in the previously described examples), which we call generically  $\varphi$ . We may therefore separate the optimization over  $\varphi$  and over all multiplicative shift factors  $\{a_{jk}\}$  and  $\{b_{jk}\}$  into an outer-loop optimization over  $\varphi$ , and an inner-loop optimization over  $\{a_{jk}\}$  and  $\{b_{jk}\}$ . This technique is reminiscent of hyperparameter optimization, which is typical in many statistical and machine-learning approaches, including GPR (Gelman et al., 2013). In the outer loop, we may perform a grid search over a specified range of  $\varphi$ , and in the inner loop, we perform maximum-likelihood estimation as described in the previous sections for a specific value of  $\varphi$ , storing the sum of the NLL loss over each pair of data sets for each  $\varphi$ . Thus, we obtain minimizers  $\{\hat{a}_{jk}(\varphi)\}$  and  $\{\hat{b}_{jk}(\varphi)\}$  as a function of  $\varphi$ , and then choose the optimal  $\hat{\varphi}$  to minimize the summed NLL loss over all pairwise shifts in the inner loop. This approach is effective due to the efficiency of the inner-loop optimization. Moreover, with a grid search over  $\varphi$ , the outer-loop

optimization is trivially parallelizable; thus it is possible to incorporate nonmultiplicative shifts with minimal increase in runtime.

### 2.3. Maximum a posteriori estimation

The maximum likelihood estimation approach described so far represents a systematic and robust method for superposing data, which naturally incorporates our intuitive sense of the role of uncertainty in the superposition process. This approach is sufficient to obtain estimates of the optimal shifting parameters in many, if not most, cases. However, the maximum likelihood approach does not provide estimates of the uncertainty in the inferred shift factors, and it is not effective in the limited cases where the NLL loss is degenerate over multiple values of the shift factor. Both of these deficiencies can be addressed by introducing the notion of a *prior distribution* over the shifting parameters, turning the maximum likelihood estimation approach to one of maximum a posteriori estimation.

From Bayes' theorem, the likelihood of observing two data sets from a master curve given certain shift factors,  $p(\mathcal{D}_j, \mathcal{D}_k | a_{jk}, b_{jk})$ , when multiplied by a prior distribution over  $a_{jk}$  and  $b_{jk}$ ,  $p(a_{jk}, b_{jk})$ , is transformed to the *posterior distribution* over the shifting parameters:

$$p(a_{jk}, b_{jk} | \mathcal{D}_j, \mathcal{D}_k) \propto p(\mathcal{D}_j, \mathcal{D}_k | a_{jk}, b_{jk}) p(a_{jk}, b_{jk}). \quad (20)$$

This transformation brings two substantial improvements to our method. Firstly, the posterior distribution  $p(a_{jk}, b_{jk} | \mathcal{D}_j, \mathcal{D}_k)$  now provides a distributional measure of the uncertainty in the shifting parameters. We may either plot this distribution to understand the certainty in the inferred shifts, or we may summarize it by computing the concavity of its logarithm around its maximum. Secondly, this formalism allows one to directly encode prior expectations regarding the shift parameters in order to break degeneracy in certain circumstances. For instance, some authors have noted that in the case of ambiguous shifting in time-cure superposition, it is sensible to minimize the extent of horizontal shifting, which may be encoded by a simple Gaussian prior in  $\ln a_{jk}$  centered about zero (Larsen et al., 2008):

$$-\ln p(a_{jk}, b_{jk}) \propto \lambda^2 (\ln a_{jk})^2. \quad (21)$$

When employing the negative logarithm of the posterior distribution in estimation, this prior is equivalent to introducing  $L^2$ -regularization in the maximum likelihood framework (Hoerl and Kennard, 1970). The parameter  $\lambda$  in this case may be selected using one of many methods for hyperparameter optimization (Zhang, 1993; Gelman et al., 2013). In this work, we employ MCCV for optimization of  $\lambda$  (Xu et al., 2004), which will be discussed briefly in the next section.

The maximum likelihood framework developed previously is extended to maximum a posteriori estimation simply by replacing the likelihood function within the objective in equation (14) with the posterior distribution defined in 20. The minimizers  $\hat{a}_{jk}$  and  $\hat{b}_{jk}$  upon this substitution become maximum a posteriori estimates, which are accompanied by posterior distributions that enable us to quantify their uncertainty. For these posterior distributions to be meaningful, they must be normalized such that their integral over their entire domain is unity.

When our prior expectations about the shift factors are limited, we may employ a uniform prior over  $a_{jk}$  and  $b_{jk}$ . This uniform prior is the maximum entropy (or equivalently, minimum information) prior possible when  $a_{jk}$  and  $b_{jk}$  are bounded. With a uniform prior (i.e., assigning a constant value of  $p(a_{jk}, b_{jk})$ ), the posterior distribution is equal to the likelihood function, and therefore the maximum likelihood and maximum a posteriori estimates are equivalent. In the examples which follow, we assume a uniform prior unless otherwise noted, and bound the region of shift parameters to be that which results in nonzero overlap between data sets.

### 2.4. Monte Carlo cross-validation

In cases where a uniform prior over the shift parameters is not appropriate, we instead employ a prior, such as the Gaussian prior in equation (21), that typically contains one or more hyperparameters ( $\lambda$  in the case of

equation (21)). This prior is meant to encode an expectation for the shift factors; for example, in cases where a wide range of shift factors result in equivalent superposition of two data sets, a Gaussian prior favors the minimum extent of shifting that also superposes the data. In order for the prior to work effectively, however, an appropriate value of the hyperparameter must be selected, without advanced knowledge of the “tightness” of the prior distribution that will properly balance the preference for minimal shifting against the quality of superposition. To find such a value, we employ one of many methods for hyperparameter optimization.

A general class of common methods for hyperparameter optimization is cross-validation, in which a fraction of the original data set is left out (the “validation” set), and the model is fit to the remaining data (the “training” set) (Zhang, 1993; Gelman et al., 2013). The optimal hyperparameter value is typically chosen as that value which, when used to fit the training data, results in the best predictive performance over the validation set. The methods for partitioning the data into validation and training sets vary. Here, we use a technique called MCCV, in which training data are selected randomly and without replacement from the original data set, with a fraction  $\alpha$  of the data retained for training and the remaining fraction  $(1 - \alpha)$  held out for validation (Xu et al., 2004). To apply MCCV to our method, we randomly partition each data set (i.e., the data at a particular state,  $t_j$ ) individually, thus building an ensemble training set consisting of training data from all states, and a corresponding ensemble validation set. The sampling and validation are repeated for  $K$  MCCV “folds” to obtain an estimate of the true cross-validation error associated with a particular hyperparameter value.

A single MCCV fold consists of partitioning the data, applying our automated superposition algorithm using a Gaussian prior with a particular value of  $\lambda$  to the training data in order to obtain a master curve, fitting this master curve to its own GP model, and then evaluating the joint NLL of observing the validation data from this model, denoted  $\mathcal{L}_k(\lambda)$ . We repeat this for  $K$  MCCV folds, and compute the average joint NLL over all folds,  $\mathcal{L}(\lambda) = K^{-1} \sum_k \mathcal{L}_k(\lambda)$ . In the limit that  $K$  is large, the value of  $\lambda$  that minimizes  $\mathcal{L}(\lambda)$ , denoted  $\lambda_m$ , is optimal. However, evaluating a large number of MCCV folds is expensive, thus the estimate of  $\lambda_m$  is subject to some variation due to the sampling procedure. With a limited number of folds, a more conservative estimate of the optimal  $\lambda$  is the maximum value that is within one standard error of the joint NLL loss at  $\lambda_m$  (Hastie et al., 2009; Silmore et al., 2019):

$$\hat{\lambda} = \max \left\{ \lambda \mid \mathcal{L}(\lambda) \leq \mathcal{L}(\lambda_m) + K^{-1} \sqrt{\sum_k (\mathcal{L}_k(\lambda_m) - \mathcal{L}(\lambda_m))^2} \right\}. \quad (22)$$

When hyperparameter optimization is complete, we then apply the superposition approach with a Gaussian prior and optimal hyperparameter  $\hat{\lambda}$  to the entire data set.

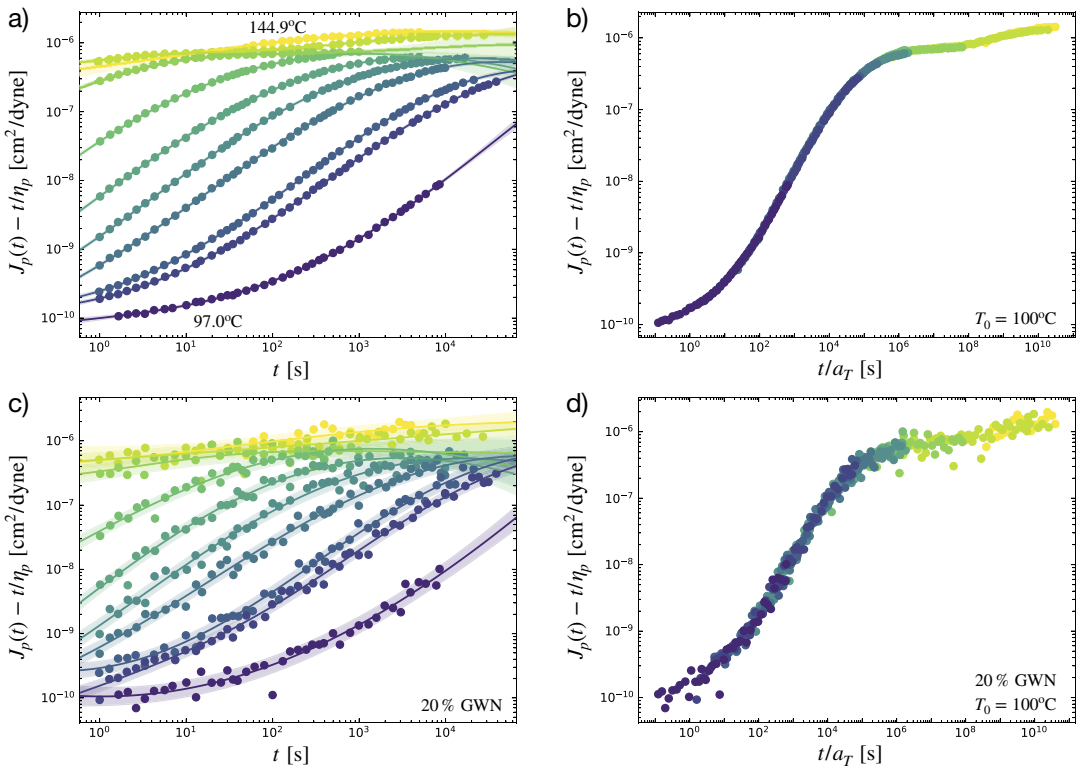
### 3. Detailed Examples Drawn from Rheology

In this section, we present a number of illustrative examples taken from the extensive literature on the application of the method of reduced variables in order to demonstrate the performance of our data superposition methodology. These examples are presented in increasing order of complexity—beginning with horizontal-only shifting for time–temperature superposition, continuing to vertical and horizontal shifting for time–cure superposition, and then concluding with two examples of introducing nonmultiplicative shifting: one in shear rate—packing superposition of a liquid–liquid emulsion, and another in time–age–time superposition of a physically aging soft glassy material. In each case, we highlight the accuracy of our method by comparing the results of our algorithm to results obtained by other means, and emphasize how the trained models may be interpreted to gain insight into the physics underlying the observed data. We also provide timing benchmarks, recorded on a 2019 MacBook Pro (2.4 GHz 8-Core Intel Core i9 processor, 64 GB RAM), to demonstrate the computational efficiency of our approach.

Each of the examples presented in this section employs the default GP kernel described by equation (7), with a zero prior mean function ( $\mu(x; t) = 0$ ). The software implementation of our algorithm, which has been used to obtain the results in this section, is available for download at <https://github.com/krleannon/mastercurves>.

### 3.1. Time–temperature superposition of a polymer melt

Time–temperature superposition is an ubiquitous method for analyzing data pertaining to the thermo-mechanical response of polymeric materials (Markovitz, 1975; Ferry, 1980). Data at different temperatures may be used to construct master curves that span tens of orders of magnitude in time or frequency, a range far beyond that achievable in a single experiment. Here, we analyze canonical creep data taken for an entangled melt of polystyrene by Plazek (1965). The digitized data, shown in Figure 2a, were rescaled to the recoverable creep compliance by the original author, who first subtracted out a response mode



**Figure 2.** Automated superposition of recoverable creep compliance data acquired for a polystyrene melt at different temperatures. (a) The recoverable creep compliance data in laboratory time (circles), for temperatures of  $T = 97.0, 100.6, 101.8, 104.5, 106.7, 109.5, 114.5, 125.0, 133.8,$  and  $144.9^\circ\text{C}$ , shown vertically in increasing order (digitized by Plazek, 1965). Solid lines and shaded regions show the posterior mean  $m(t)$  and uncertainty bounds corresponding to one standard deviation,  $m(t) \pm s(t)$ , determined via Gaussian process regression. (b) Automatically constructed master curve using horizontal shifting, with a reference temperature of  $T_0 = 100^\circ\text{C}$ . (c) The recoverable creep compliance with added 20% relative Gaussian white noise, along with the associated posterior mean  $m(t)$  and one standard deviation bounds  $m(t) \pm s(t)$  determined by Gaussian process regression. (d) Automatically constructed master curve using horizontal shifting with 20% relative Gaussian white noise added to the raw data.

proportional to the measured background viscosity  $\eta(T)$ , then multiplied the result by a temperature and density-dependent factor:

$$J_p(t; T) - t/\eta_p(T) \equiv \left( \frac{T\rho}{T_0\rho_0} \right) [J(t; T) - t/\eta(T)], \tag{23}$$

where  $T$  and  $\rho$  are the temperature and density describing a specific data set, and  $T_0$  and  $\rho_0$  are, respectively, the reference temperature (373.2 K) and the density at the reference temperature. Because the vertical rescaling is accomplished using independently measured quantities such as  $\rho(T)$ , superposition of data sets at different temperatures requires only horizontal, multiplicative shifting.

In the notation developed in the previous section, the individual data sets are described by the state parameter  $t=T$ , and the measured data points are  $(x_i, C_i)_j = (t_i, J_p(t_i; T_j) - t_i/\eta_p(T_j))$ . We take the logarithm of this data for GPR and inference, and perform a grid search over values of the horizontal shift parameter,  $a_{jk}$ , with shifting performed pairwise in order of decreasing temperature. Including the time required to fit the GP models, the superposition algorithm converges in 1.6 s. The master curve produced by the algorithm is shown in Figure 2b, with a reference temperature selected to be  $T_0 = 100^\circ\text{C}$  as in Plazek (1965). Visually, the data have been brought into registry to create a single master curve with minimal deviations. Comparing Figure 2b with Figure 4 of Plazek (1965) confirms that the automatically and manually generated master curves are visually similar, and span nearly the same range in the shifted time coordinate.

A more quantitative comparison of the automated and manual shifting results is seen in Table 1, which lists the manually inferred shifts from Plazek (1965) and the maximum a posteriori estimates from our automated method. The manual and automated shifts are quantitatively close, particularly at lower temperatures, where the slope of the data is greater. At higher temperatures, however, our automated procedure identifies that smaller shifts than those determined manually are required to optimally superpose the data. Moreover, we have obtained uncertainty estimates on these shift factors by computing the inverse Hessian of the joint negative log-posterior over each shift factor  $a_T$ , under the assumption that the posterior is approximately Gaussian. With this assumption, each diagonal element of the inverse Hessian matrix,  $H_T^{-1}$ , is related to the variance  $\sigma_T^2$  in the inferred value of the corresponding  $a_T$  (relative to the value of  $a_T$  inferred at the next lower temperature, due to pairwise shifting):  $H_T^{-1} = \sigma_T^2$ . The shift factor at  $T = 97.0^\circ\text{C}$  is taken from Plazek (1965) and assumed to be exact, and the relative uncertainty in the shifts is accumulated at successively higher temperatures, thus the

**Table 1.** Manually and automatically inferred horizontal shift factors for the recoverable creep compliance of an entangled polystyrene melt (Plazek, 1965).

$T$ ( $^\circ\text{C}$ )	$a_T$ (Manual)	$a_T$ (Automated)	$a_T$ , 20% GWN (Automated)
97.0	$1.35 \cdot 10^1$	$1.35 \cdot 10^1$	$1.35 \cdot 10^1$
100.6	$6.03 \cdot 10^{-1}$	$(5.98 \pm 0.05) \cdot 10^{-1}$	$(6.2 \pm 0.3) \cdot 10^{-1}$
101.8	$2.60 \cdot 10^{-1}$	$(2.84 \pm 0.03) \cdot 10^{-1}$	$(3.1 \pm 0.2) \cdot 10^{-1}$
104.5	$4.12 \cdot 10^{-2}$	$(4.22 \pm 0.05) \cdot 10^{-2}$	$(4.4 \pm 0.3) \cdot 10^{-2}$
106.7	$1.10 \cdot 10^{-2}$	$(1.14 \pm 0.02) \cdot 10^{-2}$	$(1.20 \pm 0.09) \cdot 10^{-2}$
109.5	$2.63 \cdot 10^{-3}$	$(2.71 \pm 0.05) \cdot 10^{-3}$	$(2.7 \pm 0.2) \cdot 10^{-3}$
114.5	$3.24 \cdot 10^{-4}$	$(3.3 \pm 0.1) \cdot 10^{-4}$	$(3.3 \pm 0.6) \cdot 10^{-4}$
125.0	$1.35 \cdot 10^{-5}$	$(1.5 \pm 0.1) \cdot 10^{-5}$	$(1.2 \pm 0.5) \cdot 10^{-5}$
133.8	$2.14 \cdot 10^{-6}$	$(1.4 \pm 0.1) \cdot 10^{-6}$	$(1.1 \pm 0.6) \cdot 10^{-6}$
144.9	$3.89 \cdot 10^{-7}$	$(1.6 \pm 0.2) \cdot 10^{-7}$	$(1.2 \pm 0.7) \cdot 10^{-7}$

*Note.* Manual shift factors are those presented in Plazek (1965). The automated shift factors represent the maximum a posteriori estimates under a uniform prior, with the reported uncertainty representing one standard deviation,  $\sigma$ . All shifts are computed for a reference temperature of  $100^\circ\text{C}$  to match the convention in Plazek (1965). The third column lists the automatically computed shift factors with no added noise, and the fourth column lists the automatically computed shift factors with 20% relative Gaussian white noise (GWN) added to the raw data.

relative uncertainty in the shift factors increases monotonically with temperature. In [Table 1](#), these uncertainty estimates are shown to be quite small relative to the inferred  $a_T$ . Thus the automatically generated results are both accurate, and precise.

We next test the robustness of our approach to the presence of noise in the underlying data by adding synthetic Gaussian white noise (GWN) to the raw data. The data with added 20% GWN are shown in [Figure 2c](#), along with the posterior mean and one standard deviation uncertainty bounds of the regressed GP models. The GP models naturally incorporate the noise in the data via larger standard deviations, leaving the posterior mean function largely unaffected, as compared to the case with no added noise. Thus, the superposition of the noisy data, depicted in [Figure 2d](#), is visually similar to that in the noise-free case, creating a master curve that spans the same range in the temporal coordinate, and maintains the same shape. The inferred shift factors for the added noise case are listed in [Table 1](#) as well, and remain close to both the noise-free and manually inferred values. Although these maximum a posteriori estimates are not substantially affected by a relatively low signal-to-noise ratio, the increased uncertainty in the learned GP models broadens the posterior distributions over the horizontal shift factors, leading to increased uncertainty bounds. In particular, the uncertainties of the inferred shifts in the added noise case are between one and two orders of magnitude greater than those for the noise-free case. These results confirm the robustness of our automated shifting technique to noise in the underlying data, and demonstrate that this noise is appropriately transferred to uncertainty in the inferred shifts by virtue of our statistical approach.

This example represents a success of our method in interpretable machine learning. The learned master curve and shift factors together form a predictive model, which can be employed for subsequent material design or discovery. This model is not all that is learned from our algorithm, however. The shift factors in [Table 1](#) may also be used to infer physical features of this polystyrene melt. For instance, fitting the shift factors to the Williams, Landel, and Ferry equation reveals the free volume and coefficient of thermal expansion of the melt (Plazek, 1965).

Before we continue to other applications of our method, it is worthwhile to consider how this method compares to previous methods for automatic time–temperature superposition. In particular, we compare the results presented in this section to those obtained by a popular method known as the closed-form shifting (CFS) algorithm (Gergesova et al., 2011), and to the minimum arc length method (Maiti, 2016). A detailed comparison is presented in the Appendix, with the results summarized briefly here. While our algorithm converged in 1.6 s, the CFS algorithm had a runtime of 4 ms, and the minimum arc length method converged in 2.7 s. Firstly, we note that our method produces shift factors that are at least as close to those found manually for nearly every temperature, compared to the shift factors determined by the CFS and minimum arc length methods. However, a particular strength of our method is its robustness to noise, thus we compare the relative change in the shift factors inferred by each method upon adding 20% relative GWN to the compliance data. Compared to the CFS algorithm, the relative deviations in the shifts inferred by our algorithm in the presence of noise were more than a factor of 10 lower on average, and more than a factor of 20 lower in the worst case ([Table A2](#)). Therefore, the speedup of the CFS algorithm comes at an apparently high cost in robustness to noise. Compared to the minimum arc length method, the deviations in shifts inferred by our algorithm were more than a factor of two lower on average and in the worst case. Thus, the accuracy is improved two-fold, while still providing runtime savings of nearly a factor of two.

The runtimes noted in the above section all correspond to a single call of each algorithm. However, a single call to our algorithm produces estimates for the shift factors along with their uncertainties, while the CFS and minimum arc length methods provide estimates of the shift factors alone. A fairer comparison of runtimes is one that takes into account the number of calls needed to produce accurate uncertainty estimates for the latter two algorithms. Previous attempts at such uncertainty quantification have focused on second-order bootstrap resampling, noting that 100 samples per order are typically sufficient (Maiti, 2019). This procedure requires many thousands of calls to the underlying shifting algorithm. Thus, to compute uncertainty estimates of the shifts in this way requires more than 6 min of computation for the CFS algorithm, and more than 75 hr for the minimum arc length method. Thus, when performing uncertainty quantification, the runtime savings from our method are more than two orders of magnitude

compared to the CFS algorithm, and more than five orders of magnitude compared to the minimum arc length method.

The previous two paragraphs clearly identify the strengths of our method compared to existing algorithms—ours is both substantially more robust to noise, and substantially faster when computing uncertainty estimates of the shift factors. In addition to its superiority in these aspects, one of the key algorithmic advances of our method is that it is agnostic to the type of coordinate transformations being employed. The CFS algorithm is inherently restricted to multiplicative shifting in a single dimension, and therefore cannot be applied to vertical and horizontal shifting problems unless the vertical and horizontal shifts can be decoupled (Gergesova et al., 2016). Other types of shifts presented later in this work are excluded entirely from this algorithm. Similarly, computational challenges preclude the optimization of the arc length in two dimensions; thus the minimum arc length method has been applied only to multiplicative horizontal shifting. No such restrictions apply to our method—it can be readily applied to a variety of superposition problems, as we will demonstrate in the coming sections.

### 3.2. Time-cure superposition during gelation

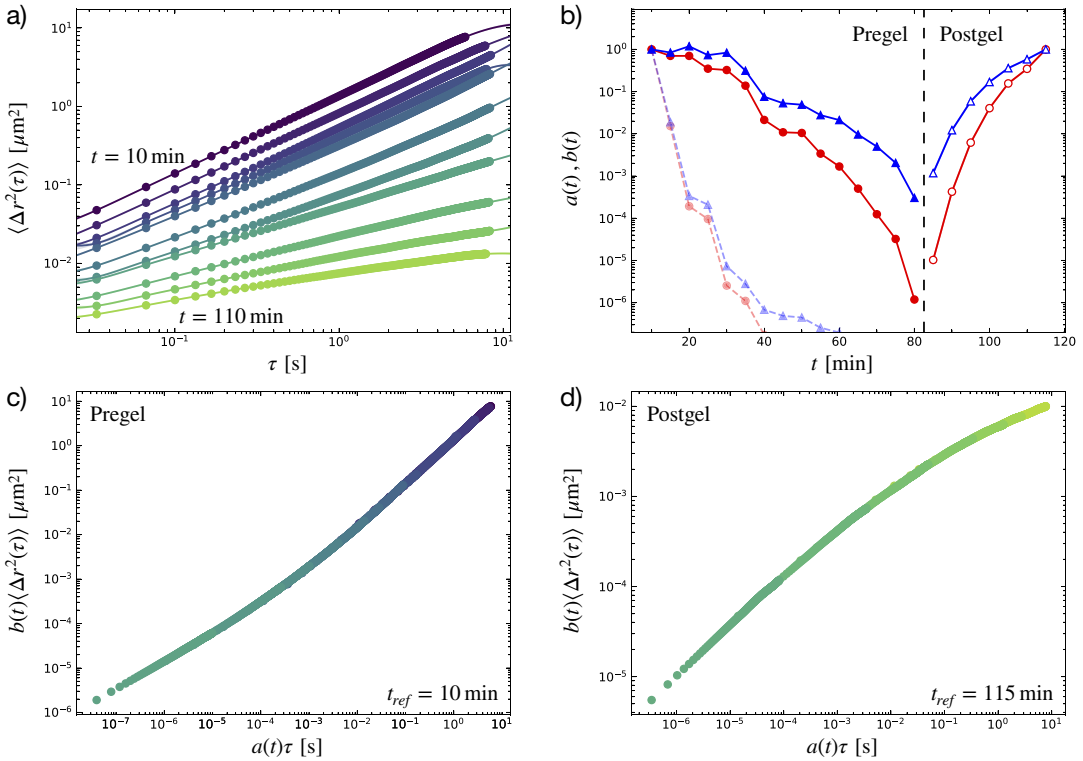
The previous time–temperature superposition example demonstrated the effectiveness of our algorithm for multiplicative shifting of self-similar curve segments along the abscissa only. This is an important and widely applicable reference case; however, many instances of data superposition involve a rescaling of both the abscissa and the ordinate with changes in the state parameter. Simultaneous shifting along both axes complicates the shifting problem, in particular, because it may result in a manifold of candidate shift parameters in two dimensions. Many previous methods for automatic generation of master curves cannot accommodate this more complicated case, either due to explicit limitations in their objective functions, or because they are not able to efficiently explore the two-dimensional domain of shift parameters. Our algorithm, however, extends naturally to simultaneous shifting on both axes, and its efficiency makes this two-dimensional optimization problem computationally tractable. Moreover, in the majority of circumstances involving rheological data, data are logarithmically scaled, so we may leverage the analytical solution for the optimal vertical shift presented in Section 2.2.2 to optimize shifts in two dimensions with minimal additional computational effort.

In this section, we present an example which encompasses simultaneous vertical and horizontal multiplicative shifting. The data set in this example represents a collection of mean-squared displacement versus lag-time trajectories that were obtained by particle tracking passive microrheology in a peptide hydrogel undergoing gelation (Larsen and Furst, 2008). The state parameter in this data set represents the time  $t$  since initial sample preparation (i.e., the “cure time”), ranging from 10–115 min in 5 min increments. A subset of this data (those trajectories collected at 10 min increments) is shown in Figure 3a. Because the uncertainty characteristics of microrheological experiments differ from those of bulk rheology, we fit these data to GP models with  $\sigma_u = 0$ , to let the noise level be inferred solely from the data.

There are multiple particularly interesting features of this data set. One feature of importance, which has been noted in previous time-cure superposition studies (Suman et al., 2021), is that the shift factors tend to increase with distance from the gel point,  $t_c$ . This results in the creation of two master curves from the data set—one for the pregel states  $t < t_c$  and one for the postgel states  $t > t_c$ . The gel point can be inferred by determining the binary partition of the data leading to the optimal superposition into two master curves (Larsen and Furst, 2008).

To construct the pregel and postgel master curves, and determine the optimal partition of the data set, we apply our superposition algorithm both in the forward (increasing  $t$ ) and backward (decreasing  $t$ ) directions, enforcing that the horizontal shifts decrease monotonically, and recording the pairwise posterior loss for each successive pair of curves in each direction. The total loss across a master curve is the sum of all pairwise losses; thus we may determine the optimal partition of the data to two master curves as that which minimizes the joint sum of pairwise losses associated with both curves. Specifically,





**Figure 3.** Automated superposition of mean-squared displacements measured by particle tracking passive microrheology in a peptide hydrogel undergoing physical gelation (Larsen and Furst, 2008). (a) The mean squared displacement  $\langle \Delta r^2(\tau) \rangle$  as a function of lag-time  $\tau$  (circles). Data were obtained for cure times  $t \in [10, 115]$  minutes at 5 min increments, though only data at 10 min increments are shown here, presented in vertically descending order. Solid lines and shaded regions show the posterior mean  $m(\tau)$  and uncertainty bounds corresponding to one standard deviation,  $m(\tau) \pm s(\tau)$ , determined via Gaussian process regression. (b) Maximum a posteriori estimates of the horizontal (red circles) and vertical (blue triangles) shift factors. Lightly shaded points and lines represent values inferred under a uniform prior. Darker shaded points and lines represent values inferred with a Gaussian prior over  $a(t)$  with hyperparameters optimized pairwise using Monte Carlo cross-validation. The shifts are optimally partitioned into a pregel curve and postgel curve. (c) Pregel master curve obtained with an optimized Gaussian prior over the shift factors  $a(t)$ . (d) Postgel master curve obtained with an optimized Gaussian prior over  $a(t)$ .

if  $\mathcal{L}_f(t_k)$  represents the loss incurred by adding the curve at state  $t_k$  to the pregel (forward) master curve, and  $\mathcal{L}_b(t_k)$  represents the loss incurred by adding that curve to the postgel (backward) master curve, then:

$$\hat{t}_c = \arg \min_{t_c} \left\{ \sum_{t_k < t_c} \mathcal{L}_f(t_k) + \sum_{t_k > t_c} \mathcal{L}_b(t_k) \right\}. \tag{24}$$

Upon applying the algorithm to this data set, however, one notices immediately that certain early cure time data sets are nearly linear trajectories in the log–log plots. Thus, there exists for these states a manifold of shifts resulting in nearly degenerate loss when applying a uniform prior to both the horizontal and vertical shift factors. In this case, the algorithm tends to prefer aligning curves at their ends, in order to minimize the overlap between the low-uncertainty (data-dense) regions of the underlying GP models where deviations between curves is most heavily weighted in the joint NLL loss. This behavior results in

extreme values of the shift factors  $a(t)$  and  $b(t)$ . This case of excessive shifting is demonstrated by the horizontal and vertical shifts inferred with a uniform prior, shown as the lightly shaded points in Figure 3b. For early times, these shifts decrease precipitously, even when no shifting at all would result in a similar likelihood of the master curve.

We may remedy these extreme shift factors by introducing a more informative prior over the shift factors, specifically, the Gaussian prior presented in equation (21). This prior is aligned with physical intuition, as previous authors have noted that in the case of ambiguous shifts it is sensible to limit shifts along the abscissa in favor of shifts along the ordinate, corresponding to an increased viscosity of a solution undergoing gelation (Larsen et al., 2008). For each successive pair of data sets, we optimize the hyperparameter  $\lambda$  for this prior using MCCV. For this example, it was found that  $\alpha = 0.1$ ,  $K = 20$ , and a grid search over 31 values of  $\lambda \in [0.01, 10]$  (spaced logarithmically) were sufficient.

The darker shaded markers and lines in Figure 3b represent the maximum a posteriori estimates of the horizontal (red circles) and vertical (blue triangles) shift factors,  $a(t)$  and  $b(t)$  respectively, inferred from a Gaussian prior with optimized  $\hat{\lambda}$ . The introduction and optimization of this prior has clearly resulted in moderate values of shifts when superposing two nearly linear data sets (i.e., data obtained at early cure times), while still allowing for larger shifts when necessary (i.e., later cure times). The inferred shifts now closely resemble those inferred manually (cf. Figure 3a in Larsen and Furst, 2008). Importantly, our algorithm and the condition specified in equation (24) have also reproduced the manual determination of the gel point (Larsen and Furst, 2008), where the optimal partition of the data set to two master curves occurs with  $t_c$  between 80 and 85 min. This sensitive determination of a critical point demonstrates the capability of our automated shifting algorithm to detect meaningful, physically relevant features in rapidly evolving rheological data. The resulting pregel and postgel master curves are shown in Figure 3c,d, both of which show excellent superposition, as well as close agreement with the manually constructed master curves reported in Larsen and Furst (2008). Moreover, the shift factors depicted in Figure 3b may be interpreted in terms of the divergence in the longest relaxation time and creep compliance of the hydrogel near the gel point. The exponents in this power-law divergence may be fit from the learned shift factors and compared to predicted results for certain universality classes of percolated gels, providing insight into the nature of hydrodynamic interactions in these evolving hydrogels (Larsen and Furst, 2008; Joshi and Petekidis, 2018).

### 3.3. Shear rate-volume fraction superposition of an emulsion

In Section 2.2.3, we discussed how our automated, maximum a posteriori shifting approach can be extended beyond simple multiplicative vertical and horizontal shifting. Due to the efficiency of the optimization over multiplicative shifts, for example, it is straightforward and still computationally tractable to perform other non-multiplicative coordinate transformations in an outer optimization problem. Moreover, this optimization hierarchy fits naturally within a Bayesian framework, so that the posterior distribution can easily incorporate a new parameter from the outer optimization problem. In this section, we demonstrate one such problem: the superposition of steady flow curves obtained for castor oil-in-water emulsions at varying oil volume fractions (Dekker et al., 2018).

Recently, a three-component model has been proposed to describe these steady flow curves, composed of elastic, plastic, and viscous contributions to the total shear stress  $\sigma(\dot{\gamma}; \phi)$  in the emulsion (Caggioni et al., 2020):

$$\sigma(\dot{\gamma}; \phi) = \sigma_y(\phi) + \sigma_p(\phi) \left( \frac{\dot{\gamma}}{\dot{\gamma}_c(\phi)} \right)^{1/2} + \eta_{bg} \dot{\gamma}, \quad (25)$$

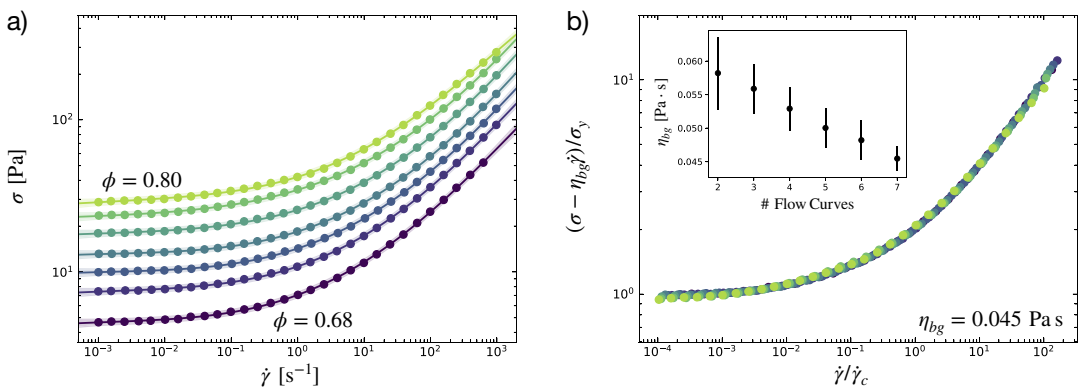
where the yield stress  $\sigma_y(\phi)$  and critical shear rate  $\dot{\gamma}_c(\phi)$  both are assumed to vary with the oil volume fraction  $\phi$ , while the background viscosity  $\eta_{bg}$  is typically assumed to be independent of  $\phi$ . This model assumes a true yield stress  $\sigma_y$  and square root dependence of the plastic contribution on the shear rate. More generally, we may relax these assumptions to write the relationship:

$$\sigma(\dot{\gamma}; \phi) = \sigma_y(\phi) g\left(\frac{\dot{\gamma}}{\dot{\gamma}_c(\phi)}\right) + \eta_{bg} \dot{\gamma}, \tag{26}$$

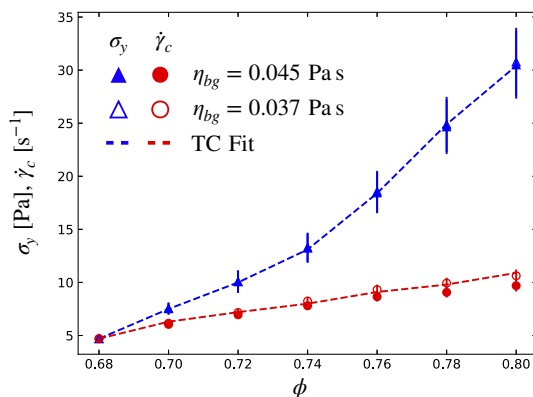
where  $g(\dot{\gamma}_r)$ , with the reduced shear rate  $\dot{\gamma}_r \equiv \dot{\gamma}/\dot{\gamma}_c(\phi)$ , captures the nonlinear viscoplastic response of the emulsion. For example, an ideal Herschel–Bulkley fluid would be described by  $\eta_{bg} = 0$  and  $g(\dot{\gamma}_r) = 1 + \dot{\gamma}_r^n$ . The parametric self-similarity of this model is now clear, as the above equation is of the form of equation (4). To obtain the master curve  $g(\dot{\gamma}_r)$ , we must first subtract any purely viscous mode  $\eta_{bg} \dot{\gamma}$  (e.g., arising from the background solvent) from  $\sigma(\dot{\gamma}; \phi)$ , then apply vertical and horizontal multiplicative shifting.

In Figure 4a, we show steady flow curves for castor oil-in-water emulsions at seven different oil volume fractions spaced evenly between 0.68 and 0.80, digitized by Dekker et al. (2018) and each fit to a GP model. Optimization of the volume fraction-dependent vertical and horizontal shift factors,  $\sigma_y$  and  $\dot{\gamma}_c$ , as well as the volume fraction-independent background viscosity  $\eta_{bg}$  proceeds hierarchically as described previously. In particular, we perform a grid search over 100 linearly spaced values in  $\eta_{bg} \in [0, 0.1]$ . At each value, we compute the optimal  $\sigma_y$  and  $\dot{\gamma}_c$  (selecting a reference state of  $\phi = 0.68$  with  $\sigma_y = 4.7$  Pa and  $\dot{\gamma}_c = 4.7$  s<sup>-1</sup>, as computed in Caggioni et al., 2020) using our algorithm for multiplicative-only shifting, and we record the total negative log posterior loss. Under a uniform prior, the maximum a posteriori estimate of  $\eta_{bg}$  corresponds to the numerical value minimizing this loss, and the uncertainty in  $\eta_{bg}$  can be estimated from a finite-difference approximation for the Hessian of the recorded loss about the minimum.

Upon performing this optimization, we obtain a master curve with optimal vertical and horizontal shifts, as well as the maximum a posteriori estimate of the background viscosity. At each value of  $\eta_{bg}$  in the grid search, the optimization takes approximately 1.3 s to converge. The resulting master curve is presented in Figure 4b, which corresponds to a maximum a posteriori estimate of  $\eta_{bg} = (4.5 \pm 0.2) \times 10^{-2}$  Pa·s. This estimate is only slightly greater than that obtained via direct fits of the three-component model,  $\eta_{bg} = 3.7 \times 10^{-2}$  Pa·s (Caggioni et al., 2020), but now incorporates all data sets in its determination and is accompanied by an uncertainty estimate. To emphasize the benefit of employing the entire data set in robust estimation of  $\eta_{bg}$ , we perform the same optimization using pruned data sets, consisting of a subset of the seven flow curves in Figure 4a. The inset in Figure 4b presents the inferred value of  $\eta_{bg}$ , and its associated



**Figure 4.** Automated superposition of steady flow data for castor oil-in-water suspensions with varying oil volume fraction. (a) The steady shear stress  $\sigma$  measured over a range of steady shear rates  $\dot{\gamma}$  (circles), for oil volume fractions of  $\phi = 0.68, 0.70, 0.72, 0.74, 0.76, 0.78,$  and  $0.80$ , shown vertically in increasing order (digitized by Dekker et al., 2018). Solid lines and shaded regions show the posterior mean  $m(\dot{\gamma})$  and uncertainty bounds corresponding to one standard deviation,  $m(\dot{\gamma}) \pm s(\dot{\gamma})$ , determined via Gaussian process regression. (b) Automatically constructed master curve using subtraction of a state-independent purely viscous contribution to the stress with  $\eta_{bg} = 4.5 \times 10^{-2}$  Pa·s, followed by horizontal and vertical shifting. The reference state is taken as  $\phi = 0.68$  with  $\sigma_y = 4.7$  Pa and  $\dot{\gamma}_c = 4.7$  s<sup>-1</sup>. Inset shows the optimal value and estimated uncertainty of  $\eta_{bg}$  inferred by applying the method to an increasing number of flow curves from (a), averaged over all possible combinations.



**Figure 5.** The critical shear rate and yield stress values for castor oil-in-water emulsions inferred by the automated algorithm with the maximum a posteriori estimate of  $\eta_{bg} = 4.5 \times 10^{-2}$  Pa·s (filled symbols), and with  $\eta_{bg} = 3.7 \times 10^{-2}$  Pa·s (unfilled symbols), as well as values fit via the three-component (TC) model (dashed lines) (Caggioni et al., 2020). Vertical bars depict one standard error in the estimates.

uncertainty, averaged over all possible combinations of these subsets of a given size. The value of  $\eta_{bg}$  inferred with fewer flow curves deviates progressively from the value that is optimal for the entire data set, and has substantially higher uncertainty. That the parameter estimate becomes more precise when including more data highlights a salient feature of our methodology—namely, that its inferences and predictions may be continually refined by including more data.

The inferred horizontal and vertical shifts, corresponding to the critical shear rate  $\dot{\gamma}_c$  and yield stress  $\sigma_y$ , are presented in Figure 5. The values determined via individual fits to the three-component model with  $\eta_{bg} = 3.7 \times 10^{-2}$  Pa·s are shown by dashed lines for clarity (Caggioni et al., 2020). In parallel with the values determined by our automated shifting procedure for the maximum a posteriori estimate of  $\eta_{bg} = 4.5 \times 10^{-2}$  Pa·s, we also present using hollow symbols the corresponding shifts inferred by our algorithm with a fixed value of  $\eta_{bg} = 3.7 \times 10^{-2}$  Pa·s to enable a direct comparison to the previously reported three-component values in Caggioni et al. (2020). The yield stresses (triangles) determined by each method are remarkably similar and unaffected by variation in  $\eta_{bg}$ . The critical shear rates are also quite similar for all cases. The maximum a posteriori estimates for  $\dot{\gamma}_c(\phi)$  with  $\eta_{bg} = 3.7 \times 10^{-2}$  Pa·s are very close to those from the three-component fit—within one standard error for all  $\phi$ —confirming that our method produces quantitatively similar results to parameter estimation within a constitutive model. The optimal values of  $\dot{\gamma}_c(\phi)$  with  $\eta_{bg} = 4.5 \times 10^{-2}$  Pa·s are slightly lower, demonstrating some sensitivity of the method to the choice of background viscosity and again highlighting the importance of including the entire data set for parameter estimation.

Within our interpretable machine learning framework, the learned shift factors in Figure 5 may be explained in terms of a power-law dependence of the yield stress and critical shear rate on the distance to jamming (Dekker et al., 2018). The exponents in such a dependence typically suggest that these emulsions belong to a certain universality class describing a transition from liquid-like to solid-like dynamics (Paredes et al., 2013).

### 3.4. Time-age-time superposition of an aging soft glassy material

The previous examples encompass multiple cases of vertical and horizontal multiplicative shifting, with the additional feature of state-independent nonmultiplicative shifting illustrated in the previous section. These linear coordinate transformations are ubiquitous in the analysis of rheological data; however, there are other more complex material responses that require other, nonlinear coordinate transformations. One example is that of power-law rheological aging in soft glassy materials, in which the modulus and

relaxation dynamics become parametrically self-similar only in an effective time domain (Struik, 1977; Gupta et al., 2012):

$$\frac{\tilde{t}}{t_{\text{ref}}} = \frac{1}{1-\nu} \left[ \left( \frac{t}{t_{\text{ref}}} \right)^{1-\nu} - \left( \frac{t_w}{t_{\text{ref}}} \right)^{1-\nu} \right]. \tag{27}$$

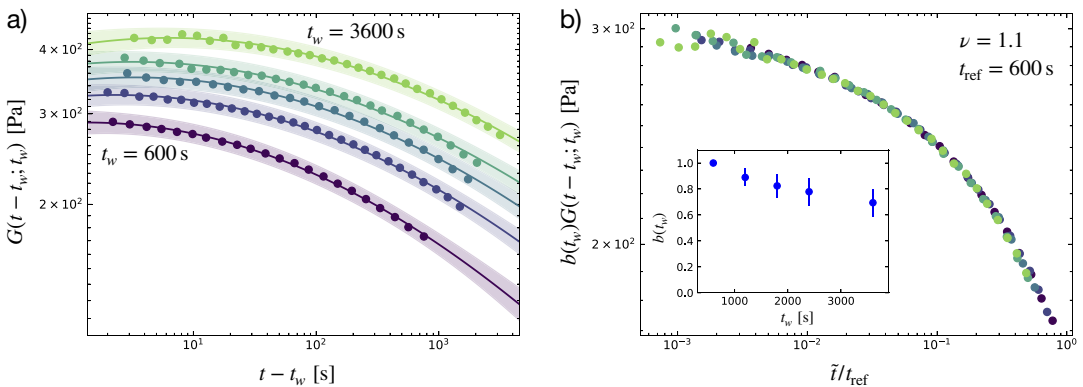
The effective time interval  $\tilde{t}$  is related by a power law in a state-independent parameter  $\nu$  (provided that  $\nu \neq 1$ ) to the laboratory time  $t$  and the “wait time”  $t_w$  (which represents roughly the elapsed time between material preparation and the beginning of a measurement), with  $t_{\text{ref}}$  representing the selected reference state.

For a soft glassy material undergoing power-law aging, one can measure the relaxation modulus  $G(t - t_w; t_w)$  by applying step strains at various times  $t_w$ . In this case, the state parameter is the wait time  $t_w$ . However, due to the power-law dilation of time described by equation (27), the system is not time-translation-invariant (Fielding et al., 2000; Joshi and Petekidis, 2018), and therefore the relaxation modulus is no longer related to a master curve by the relation specified in equation 4 with  $x = t$ . Instead, it is self-similar in the effective time domain:

$$g(\tilde{t}) = b(t_w)G(t - t_w; t_w), \tag{28}$$

where  $\tilde{t}$  depends on  $t$ ,  $t_w$ , and the exponent  $\nu$  as per equation (27). Given this similarity relation, we may apply the same hierarchical optimization approach as in the previous section to the superposition of stress relaxation data in an aging soft material. In an outer optimization loop, we perform a grid search over  $\nu$ , transforming the data into the effective time domain  $\tilde{t}(\nu)$  for each value of  $\nu$  in the search. In the inner loop, the optimal vertical shifts are computed analytically using the approach described in Section 2.2.2.

In Figure 6a, we show stress relaxation data taken for a 2.8 wt% aqueous suspension of Laponite-RD clay, digitized by Gupta et al. (2012). This clay suspension is known to exhibit soft glassy characteristics,



**Figure 6.** Automated superposition of stress relaxation data for a physically aging suspension of Laponite-RD clay with varying wait time  $t_w$  between mixing and preshearing, and imposition of a step strain with  $\gamma_0 = 0.03$  (which is always in the linear viscoelastic range). (a) The relaxation modulus  $G(t - t_w; t_w)$  as a function of the time since the step strain, is  $t - t_w$  parametric in the wait time  $t_w$ . Data (circles) are shown for  $t_w = 600$  s, 1,200 s, 1,800 s, 2,400 s, and 3,600 s, shown vertically in increasing order (digitized by Gupta et al., 2012). Solid lines and shaded regions show the posterior mean  $m(t - t_w)$  and uncertainty bounds corresponding to one standard deviation,  $m(t - t_w) \pm s(t - t_w)$ , determined via Gaussian process regression. (b) Automatically constructed master curve using transformation to the effective or material time interval  $\tilde{t}/t_{\text{ref}} = \left[ \left( t/t_{\text{ref}} \right)^{1-\nu} - \left( t_w/t_{\text{ref}} \right)^{1-\nu} \right] / (1-\nu)$  with  $\nu = 1.1$ , and subsequent vertical multiplicative shifting by a factor  $b(t_w)$ , with a reference wait time of  $t_{\text{ref}} = 600$  s. The vertical shift factors  $b(t_w)$  are shown in the inset, with vertical bars representing one-standard-error uncertainty estimates.

including power-law aging. Data are obtained for five different wait times  $t_w$  between 600 s and 3,600 s. We fit each data set to a GP model, and perform a grid search over 100 linearly spaced values of the state-independent parameter  $\nu \in [0.1, 1.9]$ , representing a range of values typical for aging clay suspensions (Joshi and Petekidis, 2018). At each value of  $\nu$  in the grid search, optimization of the multiplicative vertical shift factors proceeds analytically and in a pairwise fashion. Finally, starting from the optimal value of  $\nu$  determined by the grid search, we perform gradient descent to refine the estimation, and compute the Hessian for uncertainty analysis. In total, the grid search and subsequent gradient descent converges in just under 1 s, producing the maximum a posteriori estimate of  $\nu = 1.1 \pm 0.1$ . The resulting master curve for a chosen reference time of  $t_{\text{ref}} = 600$  s is presented in Figure 6b.

The inferred value of  $\nu$  is virtually identical to that obtained manually— $\nu \approx 1.1$  (Gupta et al., 2012), but its determination comes with the additional benefits of an uncertainty estimate and analytic determination of  $b(t_w)$  (shown in the inset of Figure 6b). This demonstrates that our automated shifting algorithm effectively handles more complicated coordinate transformations, such as time dilation via power law aging, in a manner consistent with expert manual analysis. The precise determination of  $\nu$  is also significant, because  $\nu$  is often used to delineate “hyperaging” ( $\nu > 1$ ) and “subaging” ( $\nu < 1$ ) materials. Therefore, the estimate of  $\nu = 1.1$  also serves to identify that this clay suspension is in the universality class of hyperaging materials, and this automated classification is consistent with the expert manual classification. By collating a series of such *time–age–time superposition (tats)* measurements at different temperatures, the inferred value of  $\nu$  and its dependence on temperature may be interpreted in terms of the microscopic yielding dynamics of soft glassy materials, and subsequently related to characteristic length scales and yielding energy barriers associated with the aging colloidal gel (Gupta et al., 2012).

### 3.5. Additional examples and considerations

The previous examples were selected to demonstrate the diversity of superposition problems to which our method can be applied, and these examples were presented in an order that introduced increasingly complex features of our method sequentially. There are many variations of these problems that appear in science and engineering applications, and these often present their own complexities. It is not practical to consider all of the possible complexities of shifting problems here; however, based on the previous examples and the general formulation of our method, we expect its performance to resemble that demonstrated above for most problems of interest. Moreover, the flexibility of the framework, such as in the choice of prior distributions over the shift parameters or in the choice of the coordinate transformations themselves, leaves space for adaptations to problem-specific challenges.

For instance, one complexity that is not present in the previous examples is nonmonotonicity in either the underlying master curve, or in the dependence of the shift factors on the state parameters. This feature is occasionally seen, for example, in steady flow curves (Mari et al., 2015) or in measurements of the viscoelastic loss modulus (Gergesova et al., 2016) for certain complex fluids. Because our method makes no assumptions about the relationship between the independent and dependent quantities, or between the shift parameters and state variables, nonmonotonic behavior can be easily accommodated. Nonmonotonic vertical shift factors, for instance, will be identified automatically by virtue of the analytical solution for log-scaled multiplicative shifting. Nonmonotonic horizontal shift factors can be identified by conducting a grid search over a range centered about  $\ln a_j = 0$ . If the underlying master curve is nonmonotonic, there still will exist global optima for the horizontal and vertical shift factors that construct the optimal master curve, and these optima will be found by simply conducting a large enough grid search on the horizontal shift factors. Such nonmonotonicity in the master curve may give rise to multiple local minima in the search for the horizontal shift factor, and if these minima are nearly degenerate in value, the “optimal” master curve may no longer be unique. However, as we have demonstrated in Section 3.2, the method can be biased toward either of the solutions as necessary, by introducing nontrivial physical or empirical priors on the shift factors.

Although we do not present examples that highlight many of the other considerations that may arise in superposition problems in this work, we note that more examples are available on the project GitHub

repository, including one that demonstrates nonmonotonic behavior in the master curve and shift factors measured during time–temperature–water superposition (Lalwani et al., 2021).

#### 4. Forward Predictions Using Automated Shifting

In the previous section, we provided four canonical examples of automated construction of rheological master curves using real experimental data taken from the literature, demonstrating the robustness of our automated method to different types of data, types of shifting, and levels of noise. The automated shifting algorithm is thus a useful tool in the analysis of linear and nonlinear thermorheological data, as the inferred shifting parameters provide physical insight to key material properties—such as the static yield stress of an emulsion, the gel point of a reacting or curing polymeric solution, or the rate of aging of a soft glassy material. Another potent application of this algorithm is to make reliable forward predictions of material rheology at a previously unstudied state. In this section, we present an example of one such forward prediction, using the castor oil-in-water flow curve data (Dekker et al., 2018).

Within the mathematical framework employed in this work, the forward prediction problem is as follows. Given data for a material property  $C(x;t)$  at a set of states  $\{t_j\}$ , including a reference state  $t_k$ , we apply the method outlined in Section 2 to infer optimal shifting parameters— $\{a_{jk}\}$ ,  $\{b_{jk}\}$ , and  $c$  in the scenario described by equation (4)—and apply these shifts to the data as in equation (4) to construct a master curve approximating  $g(z)$  in the reduced variable  $z = a(t)x$ . To make a forward prediction at a new state  $t_l$ , we must estimate the shift parameters  $a_{lk}$  and  $b_{lk}$  for that state, and then use equation (4) to transform the inferred master curve  $g(z)$  into a prediction of  $C(x;t_l)$  at state  $t_l$ . Making predictions continuously in the independent variable  $x$  requires a continuous approximation for  $g(z)$ , which may be obtained by fitting the automatically constructed master curve to a new GP model. Similarly, predicting  $a_{lk}$  and  $b_{lk}$  requires an interpolant between the inferred set of  $\{a_{jk}\}$  and  $\{b_{jk}\}$ ; this interpolant can also be treated using a GP model.

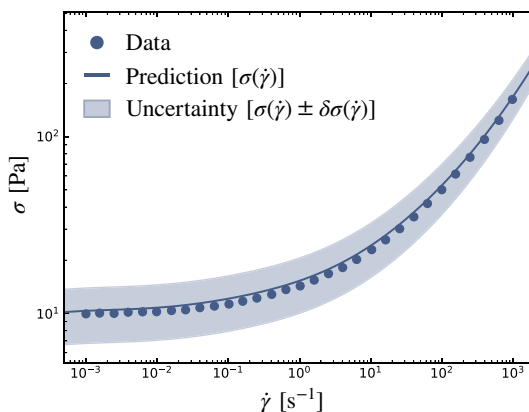
To demonstrate this forward prediction procedure, we automatically construct a master curve for the castor oil-in-water data set shown in Figure 4, but excluding the data set at  $\phi = 0.72$ . The master curve constructed while deliberately omitting this data set out is still very similar to that depicted in Figure 4b, as are the inferred shift factors  $\sigma_y$  and  $\dot{\gamma}_c$  for the remaining volume fractions, and the estimate of  $\eta_{bg} = (4.5 \pm 0.2) \times 10^{-2}$  Pa·s. These shift factors are fit to GP models as a function of  $\phi$ , and these GPs are used to make a forward prediction of the yield stress and critical shear rate at  $\phi = 0.72$ . These GPs again have a zero-mean prior, and to ensure that the GPs fit to these shifts incorporate the uncertainty in the underlying shift factors, we set the GP kernel to that in equation (7) with the fixed noise level  $\sigma_u^2$  equal to the greatest variance of any shift factor (shown by the error bars in Figure 5). The predictions from the resulting GPs are:  $\sigma_y = 9 \pm 1$  Pa and  $\dot{\gamma}_c = 6.7 \pm 0.1$  s<sup>-1</sup>, both in close agreement with the values obtained by automated shifting of the data set at  $\phi = 0.72$  and the values determined from fits to the three-component model (cf. Figure 5).

Finally, the master curve obtained from all data sets besides that for  $\phi = 0.72$  is fit to its own GP model in the reduced coordinate  $z$  with the prior mean  $\mu(z) = 0$  and kernel function  $K(\theta, z, z')$  described by equation (7). The self-similar posterior mean function  $g(\dot{\gamma}_r)$  and variance  $\delta g(\dot{\gamma}_r)$  of this GP model are then computed for a range of reduced shear rates  $\dot{\gamma}_r \equiv \dot{\gamma}/\dot{\gamma}_c$ , and the self-similar results are shifted back to the original  $(\sigma, \dot{\gamma})$  parameter space by first applying vertical and horizontal shifts with the estimated  $\sigma_y$  and  $\dot{\gamma}_c$  at  $\phi = 0.72$ , then adding in the viscous mode with the inferred optimal value of  $\eta_{bg}$ :

$$\sigma(\dot{\gamma}) = \sigma_y g(\dot{\gamma}/\dot{\gamma}_c) + \eta_{bg} \dot{\gamma}. \quad (29)$$

Uncertainty in the value of the vertical shift  $\delta\sigma_y$ , the horizontal shift  $\delta\dot{\gamma}_c$ , and the background viscosity  $\delta\eta_{bg}$  can also be propagated to the total uncertainty expected in the prediction of the resulting flow curve:

$$(\delta\sigma)^2 = g(\dot{\gamma}/\dot{\gamma}_c)^2 (\delta\sigma_y)^2 + \sigma_y^2 (\delta g(\dot{\gamma}/\dot{\gamma}_c))^2 + \left( \frac{\sigma_y g'(\dot{\gamma}/\dot{\gamma}_c) \dot{\gamma}}{\dot{\gamma}_c^2} \right)^2 (\delta\dot{\gamma}_c)^2 + \dot{\gamma}^2 (\delta\eta_{bg})^2, \quad (30)$$



**Figure 7.** Forward predictions of the steady flow curve for a castor oil-in-water emulsion with oil volume fraction  $\phi=0.72$ . The value of  $\eta_{bg} = (4.5 \pm 0.2) \times 10^{-2} \text{ Pa}\cdot\text{s}$  is inferred during construction of a master curve from the remaining data sets ( $\phi = 0.68, 0.70, 0.74, 0.76, 0.78, 0.80$ ). The yield stress  $\sigma_y = 9 \pm 1 \text{ Pa}$  and critical shear rate  $\dot{\gamma}_c = 6.7 \pm 0.1 \text{ s}^{-1}$  are estimated from Gaussian process models fit to the automatically inferred shift factors at the remaining states. The inferred master curve is fit to a Gaussian process model, whose predictions are shifted to the  $\phi=0.72$  state using the predicted values of  $\sigma_y$ ,  $\dot{\gamma}_c$ , and  $\eta_{bg}$ . The mean predicted values are shown with a solid line, and single standard deviation uncertainty bounds are shown with a shaded region. Experimental data for  $\phi=0.72$  from Dekker et al. (2018) are shown by the solid circles.

with  $g'(\dot{\gamma}_r) = dg(\dot{\gamma}_r)/d\dot{\gamma}_r$ . Figure 7 depicts the forward prediction for the steady shear stress of an emulsion with  $\phi=0.72$ . The expectation for the flow curve  $\sigma(\dot{\gamma})$  is shown as a solid line, along with uncertainty bounds  $\sigma \pm d\sigma$  in the shaded region, over six orders of magnitude in the imposed shear rate  $\dot{\gamma}$ . The predictions are compared to the original withheld data, shown with filled circles. The data fall nearly exactly on the mean predictions over the entire range of  $\dot{\gamma}$ , and in all cases are well within the uncertainty bounds, which represent a conservative estimate of the total uncertainty due to the built-in noise level of  $\sigma_u = 0.04$  in the GPs. In this case, the automatically generated master curve, in combination with GP models for the shift factors as a function of  $\phi$ , serves as a highly effective predictive tool. Notably, these accurate predictions have been made without assuming or specifying a specific constitutive model for the master curve,  $g(\dot{\gamma}_r)$ . Rather, the predictions are *data-driven*, using GPR as a machine learning tool to represent the underlying master curve. Moreover, the predictions are statistical in nature; the uncertainty bounds in Figure 7 are similarly derived directly from the available data, and are obtained with minimal extra work. Viewed holistically, this example of forward prediction highlights many of the salient features of our data superposition approach, and demonstrates how it may be straightforwardly adapted as a predictive tool for material design and formulation considerations.

## 5. Conclusions

In this work, we have developed a data-driven, statistical method for the inference of material master curves from parametrically self-similar data sets. Such responses are observed ubiquitously across the field of soft materials science and have led to the development of effective but labor-intensive approaches that are collected under the generic term: *the method of reduced variables*. Our new approach is inherently flexible, as the negative log-posterior objective is independent of the transformations used to bring different data sets into registry and can therefore accommodate a wide variety of data-shifting and superposition transformations. It is also computationally efficient, with optimal values of the vertical multiplicative shifts of (logarithmically scaled) data computed analytically. This allows for quick one-



dimensional horizontal shifting and even outer-loop optimization of state-independent parameters (such as we illustrated using the background viscosity in a model for emulsion rheology). The method is robust to the presence of noise, as noise is handled explicitly by the GP model to which the data is regressed. Finally, the uncertainty estimates of the shift parameters, together with continuous uncertainty bounds for forward predictions, are obtained automatically from the method at minimal added cost, due to the formulation of the superposition problem as one of statistical inference.

Our method for inferring master curves not only facilitates the automation of a popular method for rheological data analysis and extrapolation that has endured for more than 80 years, but also represents the development of a novel probabilistic, data-driven predictive and analytic tool. We may therefore readily extend this method to new material systems, refine its predictions with more extensive data sets, and, potentially, build an open shared library of predictive models that are unbiased by user preferences and preconceptions. Moreover, this automation has been accomplished without sacrificing the physical interpretability of the resulting models. As we have demonstrated, the learned models themselves often provide valuable insight to the underlying physics governing the material systems under study. In all fields of the physical sciences, the adoption of robust, open, and unbiased data-driven tools such as this will be critical in developing both accessible and reproducible scientific insight across large data sets.

**Author contribution.** Conceptualization: K.R.L., G.H.M., J.W.S.; Data curation: K.R.L., G.H.M., J.W.S.; Methodology: K.R.L., J.W.S.; Software: K.R.L.; Visualization: K.R.L., G.H.M., J.W.S.; Writing—original draft: K.R.L., J.W.S.; Writing—review and editing: K.R.L., G.H.M., J.W.S.

**Competing interest.** The authors declare no competing interests exist.

**Data availability statement.** The data used in this work, along with a software implementation of the proposed methodology and the code used in all demonstrations, are available in the following GitHub repository: <https://github.com/krlennon/mastercurves>.

**Funding statement.** K.R.L. was supported by the U.S. Department of Energy (DOE) Computational Science Graduate Fellowship program under Grant No. DE-SC0020347.

## References

- Adolf D and Martin JE** (1990) Time-cure superposition during crosslinking. *Macromolecules* 23(15), 3700–3704.
- Barbero EJ and Ford KJ** (2004) Equivalent time temperature model for physical aging and temperature effects on polymer creep and relaxation. *Journal of Engineering Materials and Technology* 126(4), 413–419.
- Barenblatt GI** (2003) *Scaling*. Cambridge: Cambridge University Press.
- Butler KT, Davies DW, Cartwright H, Isayev O and Walsh A** (2018) Machine learning for molecular and materials science. *Nature* 559(7715), 547–555.
- Buttlar WG, Roque R and Reid B** (1998) Automated procedure for generation of creep compliance master curve for asphalt mixtures. *Transportation Research Record* 1630(1), 28–36.
- Caggioni M, Trappe V and Spicer PT** (2020) Variations of the Herschel–Bulkley exponent reflecting contributions of the viscous continuous phase to the shear rate-dependent stress of soft glassy materials. *Journal of Rheology* 64(2), 413–422.
- Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L** (2019) Machine learning and the physical sciences. *Reviews of Modern Physics* 91, 045002.
- Cho K-S** (2009) Geometric interpretation of time-temperature superposition. *Korea-Australia Rheology Journal* 21(1), 13–16.
- de Gennes P** (1979) *Scaling Concepts in Polymer Physics*. Ithaca, NY: Cornell University Press.
- Dekker RI, Dinkgreve M, de Cagny H, Koeze DJ, Tighe BP and Bonn D** (2018) Scaling of flow curves: Comparison between experiments and simulations. *Journal of Non-Newtonian Fluid Mechanics* 261, 33–37.
- Duvenaud DK** (2014) *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, Cambridge.
- Eggers J and Fontelos MA** (2015) *Singularities: Formation, Structure, and Propagation*. Cambridge: Cambridge University Press.
- Ewoldt RH, Johnston MT and Caretta LM** (2015) Experimental challenges of shear rheology: How to avoid bad data. In Spagnolie SE (ed.), *Complex Fluids in Biological Systems*. New York, NY: Springer, pp. 207–241.
- Ferguson AL** (2017) Machine learning and data science in soft materials engineering. *Journal of Physics: Condensed Matter* 30(4), 43002.
- Ferry JD** (1980) *Viscoelastic Properties of Polymers*. New York: John Wiley & Sons.
- Fielding SM, Sollich P and Cates ME** (2000) Aging and rheology in soft materials. *Journal of Rheology* 44(2), 323–369.

- Freund JB and Ewoldt RH** (2015) Quantitative rheological model selection: Good fits versus credible models using Bayesian inference. *Journal of Rheology* 59(3), 667–701.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB** (2013) *Bayesian Data Analysis*, 3rd edn. Boca Raton, FL: CRC Press.
- Gergesova M, Saprunov I and Emri I** (2016) Closed-form solution for horizontal and vertical shiftings of viscoelastic material functions in frequency domain. *Rheologica Acta* 55(5), 351–364.
- Gergesova M, Zupančič B, Saprunov I and Emri I** (2011) The closed form t-T-P shifting (CFS) algorithm. *Journal of Rheology* 55(1), 1–16.
- Görtler J, Kehlbeck R and Deussen O** (2019) A visual exploration of Gaussian processes. *Distill*. Available at <https://distill.pub/2019/visual-exploration-gaussian-processes> Accessed May 16, 2023.
- Gupta R, Baldewa B and Joshi YM** (2012) Time temperature superposition in soft glassy materials. *Soft Matter* 8, 4171–4176.
- Hastie T, Tibshirani R and Friedman J** (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer Science & Business Media.
- Hermida ÉB and Povolo F** (1994) Analytical-numerical procedure to determine if a set of experimental curves can be superimposed to form a master curve. *Polymer Journal* 26(9), 981–992.
- Hoerl AE and Kennard RW** (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1), 55–67.
- Honerkamp J and Weese J** (1993) A note on estimating mastercurves. *Rheologica Acta* 32(1), 57–64.
- Joshi YM and Petekidis G** (2018) Yield stress fluids and ageing. *Rheologica Acta* 57(6), 521–549.
- Lalwani SM, Batys P, Sammalkorpi M and Lutkenhaus JL** (2021) Relaxation times of solid-like polyelectrolyte complexes of varying pH and water content. *Macromolecules* 54(17), 7765–7776.
- Larsen TH and Furst EM** (2008) Microrheology of the liquid-solid transition during gelation. *Physical Review Letters* 100, 146001.
- Larsen T, Schultz K and Furst EM** (2008) Hydrogel microrheology near the liquid-solid transition. *Korea-Australia Rheology Journal* 20(3), 165–173.
- Leaderman H** (1944) *Elastic and Creep Properties of Filamentous Materials and Other High Polymers*. Washington, DC: The Textile Foundation.
- Lillo F, Farmer JD and Mantegna RN** (2003) Master curve for price-impact function. *Nature* 421(6919), 129–130.
- Maiti A** (2016) A geometry-based approach to determining time-temperature superposition shifts in aging experiments. *Rheologica Acta* 55(1), 83–90.
- Maiti A** (2019) Second-order statistical bootstrap for the uncertainty quantification of time-temperature-superposition analysis. *Rheologica Acta* 58(5), 261–271.
- Mari R, Seto R, Morris JF and Denn MM** (2015) Nonmonotonic flow curves of shear thickening suspensions. *Physical Review E* 91, 052302.
- Markovitz H** (1975) Superposition in rheology. *Journal of Polymer Science, Polymer Symposia* 50(1), 431–456.
- Matheron G** (1963) Principles of geostatistics. *Economic Geology* 58(8), 1246–1266.
- Naya S, Meneses A, Tarrío-Saavedra J, Artiaga R, López-Beceiro J and Gracia-Fernández C** (2013) New method for estimating shift factors in time-temperature superposition models. *Journal of Thermal Analysis and Calorimetry* 113(2), 453–460.
- Paredes J, Michels MAJ and Bonn D** (2013) Rheology across the zero-temperature jamming transition. *Physical Review Letters* 111, 015701.
- Plazek DJ** (1965) Temperature dependence of the viscoelastic behavior of polystyrene. *The Journal of Physical Chemistry* 69(10), 3480–3487.
- Plazek DJ** (1996) 1995 Bingham medal address: Oh, thermorheological simplicity, wherefore art thou? *Journal of Rheology* 40(6), 987–1014.
- Rasmussen CE and Williams CKI** (2006) *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.
- Rouleau L, Deü J-F, Legay A and Le Lay F** (2013) Application of Kramers–Kronig relations to time-temperature superposition for viscoelastic materials. *Mechanics of Materials* 65, 66–75.
- Sihn S and Tsai SW** (1999) Automated shift for time-temperature superposition. *Proceedings of the 12th International Committee on Composite Materials* 51, 47.
- Silmore KS, Gong X, Strano MS and Swan JW** (2019) High-resolution nanoparticle sizing with maximum a posteriori nanoparticle tracking analysis. *ACS Nano* 13(4), 3940–3952.
- Singh PK, Soulaiges JM and Ewoldt RH** (2019) On fitting data for parameter estimates: Residual weighting and data representation. *Rheologica Acta* 58(6), 341–359.
- Struik L** (1977) *Physical Aging in Amorphous Polymers and Other Materials*. PhD thesis, Delft University of Technology, Delft, The Netherlands.
- Suman K, Shanbhag S and Joshi YM** (2021) Phenomenological model of viscoelasticity for systems undergoing sol-gel transition. *Physics of Fluids* 33(3), 033103.
- Tobolsky AV and Andrews RD** (1945) Systems manifesting superposed elastic and viscous behavior. *The Journal of Chemical Physics* 13(1), 3–27.
- Wagner KW** (1915) Dielektrische Eigenschaften von verschiedenen Isolierstoffen. *Elektrotechnische Zeitschrift* 12, 120,123, 135–137, 163,165.

**Xu Q-S, Liang Y-Z and Du Y-P** (2004) Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration. *Journal of Chemometrics* 18(2), 112–120.

**Zhang P** (1993) Model selection via multifold cross validation. *The Annals of Statistics* 21(1), 299–313.

## A. Appendix. Time–Temperature Superposition by Other Methods

We compare the performance of our method in unidirectional multiplicative shifting (time–temperature superposition) against two previously developed methods: the CFS algorithm (Gergesova et al., 2011) and the minimum arc length (min- $\delta$ ) method (Maiti, 2016). The CFS algorithm derives a set of algebraic expressions for the logarithms of the horizontal shift factors  $\ln a_j$  obtained in time–temperature superposition based on minimization of the area between linear interpolants of two data sets at consecutive states, computed within the range of overlap in the  $y$ -coordinate. Therefore, it is a pairwise algorithm by nature, and is necessarily limited to the case of logarithmically scaled multiplicative shifting in a single dimension. The min- $\delta$  method computes all shift factors simultaneously by minimizing the vertical arc length of the master curve (i.e., the sum of vertical distances between points at consecutive  $x$ -coordinates). This algorithm is similarly defined only for multiplicative shifting in a single dimension.

Table A1 presents the shift factors computed by these algorithms for the same data set as in Section 3.1, both with and without the same synthetic 20% GWN added to the compliance. These shift factors were computed in 4 ms by the CFS algorithm, and in 2.7 s by the min- $\delta$  method, compared to the 1.6 s runtime of our automated algorithm. Upon comparison with Table 1, we see that our algorithm more accurately reproduces the manually inferred shift factors for nearly all temperatures compared to the CFS algorithm, and similarly matches or exceeds the accuracy of the min- $\delta$  method for nearly all temperatures.

A comparison to manually inferred values may not represent the best evaluation of the accuracy of each method. Therefore, we also evaluate the robustness of each method to noise in the data by adding 20% relative GWN and repeating the inference. Table A2 presents the relative deviation in the inferred shift factors relative to the values inferred without synthetic noise. The CFS algorithm is apparently very sensitive to noise in the underlying data, with a worst-case relative error of more than 600%, and an average error of more than 100%. The min- $\delta$  algorithm improves on the CFS algorithm with a worst-case error of just over 60%, and an average error of just over 20%. However, our method exceeds both cases, with a worst-case error of 25% and an average error just below 10%.

We have demonstrated that our algorithm outperforms the min- $\delta$  algorithm by more than a factor of two in both the average and worse case for this example. It does this while also outperforming that algorithm in runtime by nearly a factor of two. Moreover, our method provides uncertainty estimates for the shifts automatically, while both the CFS and min- $\delta$  methods provide only a single estimate of these shift factors. Previous attempts to quantify uncertainty in these shift factors have required second-order bootstrap resampling of the data, resulting in thousands of repeated calls to the subroutine implementing the algorithm. For 100 resampling trials at each order, the min- $\delta$  algorithm has a runtime of 75 hr for this data set. Our algorithm need only be run once to obtain these uncertainty estimates—therefore, the true speedup is more than five orders of magnitude compared to the min- $\delta$  algorithm. While a single call of the CFS algorithm takes only 4 ms compared to our 1.6 s runtime, this speedup comes at the cost of accuracy. To obtain uncertainty estimates using the same second-order resampling procedure increases the runtime to more than 6 min. Thus, for inference with uncertainty quantification, our algorithm represents a speedup of more than two orders of magnitude compared to the CFS algorithm.

**Table A1.** The shift factors inferred by the closed-form shifting (CFS) and minimum arc length (min- $\delta$ ) algorithms for time–temperature superposition of recoverable creep compliance data for a polystyrene melt (Plazek, 1965), both without and with 20% relative Gaussian white noise added to the raw compliance data.

$T$ (°C)	$a_T$ (CFS)	$a_T$ (min- $\delta$ )	$a_T$ , 20% GWN (CFS)	$a_T$ , 20% GWN (min- $\delta$ )
97.0	$1.35 \cdot 10^1$	$1.35 \cdot 10^1$	$1.35 \cdot 10^1$	$1.35 \cdot 10^1$
100.6	$5.58 \cdot 10^{-1}$	$5.98 \cdot 10^{-1}$	$5.43 \cdot 10^{-1}$	$7.20 \cdot 10^{-1}$
101.8	$2.47 \cdot 10^{-1}$	$2.71 \cdot 10^{-1}$	$2.58 \cdot 10^{-1}$	$2.71 \cdot 10^{-1}$
104.5	$3.57 \cdot 10^{-2}$	$4.22 \cdot 10^{-2}$	$4.27 \cdot 10^{-2}$	$3.66 \cdot 10^{-2}$
106.7	$9.64 \cdot 10^{-3}$	$1.04 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$	$1.32 \cdot 10^{-2}$
109.5	$2.29 \cdot 10^{-3}$	$2.84 \cdot 10^{-3}$	$2.29 \cdot 10^{-3}$	$2.84 \cdot 10^{-3}$
114.5	$2.68 \cdot 10^{-4}$	$3.19 \cdot 10^{-4}$	$6.21 \cdot 10^{-4}$	$2.65 \cdot 10^{-4}$
125.0	$1.19 \cdot 10^{-5}$	$1.62 \cdot 10^{-5}$	$2.56 \cdot 10^{-5}$	$1.48 \cdot 10^{-5}$
133.8	$8.79 \cdot 10^{-7}$	$1.26 \cdot 10^{-6}$	$6.50 \cdot 10^{-6}$	$1.91 \cdot 10^{-6}$
144.9	$1.01 \cdot 10^{-7}$	$1.02 \cdot 10^{-7}$	$1.36 \cdot 10^{-7}$	$4.02 \cdot 10^{-8}$

Note. All shifts are computed for a reference temperature of 100°C to match the convention in Plazek (1965).

**Table A2.** The relative deviation  $|\Delta a_T|/a_T^{(0)} = |a_T^{(\sigma)} - a_T^{(0)}|/a_T^{(0)}$  in the shift factors computed with 20% relative Gaussian white noise ( $a_T^{(\sigma)}$ ) and without noise ( $a_T^{(0)}$ ), relative to the noiseless values, for the closed-form shifting (CFS) and minimum arc length (min- $\delta$ ) algorithms, compared to the shifts computed by the algorithm presented in this work.

$T$ (°C)	$ \Delta a_T /a_T^{(0)}$ (CFS)	$ \Delta a_T /a_T^{(0)}$ (min- $\delta$ )	$ \Delta a_T /a_T^{(0)}$ (present work)
100.6	0.03	0.20	0.04
101.8	0.04	0.00	0.09
104.5	0.20	0.13	0.04
106.7	0.52	0.26	0.05
109.5	0.00	0.00	0.00
114.5	1.32	0.17	0.00
125.0	1.16	0.09	0.20
133.8	6.39	0.52	0.21
144.9	0.35	0.61	0.25
Average	1.11	0.22	0.10