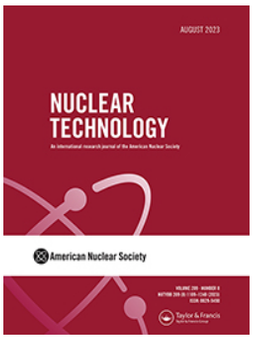

Nuclear Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/unct20

Multiphase Turbulence Modeling Using Sparse Regression and Gene Expression Programming

S. Beetham & J. Capecelatro

To cite this article: S. Beetham & J. Capecelatro (2023): Multiphase Turbulence Modeling Using Sparse Regression and Gene Expression Programming, Nuclear Technology, DOI: 10.1080/00295450.2023.2178251

To link to this article: https://doi.org/10.1080/00295450.2023.2178251

NUCLEAR TECHNOLOGY

DOI: https://doi.org/10.1080/00295450.2023.2178251

Multiphase Turbulence Modeling Using Sparse Regression and Gene Expression Programming

S. Beetham*[a] and J. Capecelatro[b,c]

[a]Oakland University, Department of Mechanical Engineering, 115 Library Drive, Rochester, Michigan 48309
[b]University of Michigan, Department of Mechanical Engineering, 2350 Hayward Street, Ann Arbor, Michigan 48109
[c]University of Michigan, Department of Aerospace Engineering, 1320 Beal Avenue, Ann Arbor, Michigan 48109

Received September 30, 2022
Accepted for Publication February 3, 2023

Abstract — *Turbulence in two-phase flows drives many important natural and engineering processes, from geophysical flows to nuclear power generation. Strong interphase coupling between the carrier fluid and disperse phase precludes the use of classical turbulence models developed for single-phase flows. In recent years, there has been an explosion of machine learning techniques for turbulence closure modeling, though many rely on augmenting existing models. In this work, we propose an approach that blends sparse regression and gene expression programming (GEP) to generate closed-form algebraic models from simulation data. Sparse regression is used to determine a minimum set of functional groups required to capture the physics, and GEP is used to automate the formulation of the coefficients and dependencies on operating conditions. The framework is demonstrated on homogeneous turbulent gas-particle flows in which two-way coupling generates and sustains carrier-phase turbulence.*

Keywords — *Multiphase flow, data-driven modeling, turbulence, gene expression programming.*

Note — *Some figures may be in color only in the electronic version.*

I. INTRODUCTION

In the last decade, data-driven approaches have become the predominant tool for developing turbulence models.[1] Of these approaches, neural networks have gained a considerable amount of traction.[2–11] In contrast, fewer approaches have elected to pursue strategies that enable a compact, algebraic closure. Formulating models in this way has several important properties including increased interpretability, ease of dissemination, and straightforward integration into existing solvers. These techniques generally fall into two categories: (1) symbolic regression and (2) gene expression programming (GEP).

Reference 12 developed a strategy based on sparse regression that identifies the underlying functional form of the nonlinear systems by optimizing a coefficient matrix that acts upon a matrix of trial functions. This construct has the important benefit of including the user in the modeling process through selection of the trial functions. References 13 and 14 extended the sparse identification framework of Ref. 12 to infer algebraic stress models for closure of the Reynolds-averaged Navier-Stokes (RANS) equations. In Ref. 13, the models were written as tensor polynomials and built from a library of candidate functions. In Ref. 14, Galilean invariance of the resulting models was guaranteed through thoughtful tailoring of the feature space.

Gene expression programming, a data-driven technique inspired by the Darwinian concept of survival of the fittest, heuristically evolves symbolic models until the error is reduced beyond a threshold. GEP has demonstrated success in modeling large-eddy simulation

*E-mail: sbeetham@oakland.edu

subgrid-scale closures,[15,16] boundary layer theory,[17] turbulent pipe flow,[18] and informing RANS closures.[19,20]

While these data-driven techniques have been increasingly utilized for modeling single-phase turbulence, their application to multiphase turbulence modeling is still relatively uncharted. Despite this, multiphase flows present a rich and diverse class of problems for which machine learning can prove useful. Because of the large parameter space frequently attributed to such flows, traditional modeling techniques have historically failed, especially beyond dilute regimes, where models extended from single-phase turbulence breakdown (see, e.g., Ref. 21 and discussion therein). At moderate volume fractions, particles are capable of generating turbulent kinetic energy (TKE) at the smallest scales.[22,23] This is the direct antithesis to the classical notion of the turbulent energy cascade and precludes the use of the Boussinesq hypothesis for modeling the Reynolds stress. This motivates the need for methodologies capable of formulating closures "from scratch" rather than augmenting existing models. These challenges, along with the societal importance and pervasiveness of these flows,[24] make them excellent candidates for improvements in data-driven modeling techniques.

In this work, we propose a blended modeling approach that combines the strengths of both sparse regression and GEP to inform turbulence closures in a manner that leverages the physical knowledge of the modeler and automates the determination of model components for which physical insight is not obvious or does not exist. The approach is applied to disperse two-phase flows where the carrier phase is laden with spherical rigid particles at moderate volume fractions. Strong two-way coupling between fluid and particles generates and sustains turbulence. This configuration has been discussed extensively in prior work (see Refs. 14, 25, and 26 for more details) and serves as a case study here.

II. METHODOLOGY

It is well established[27] that any tensor quantity $\mathcal{D}_{ij}$ can be exactly described by an infinite sum given as

$$\mathcal{D}_{ij} = \sum_{n=1}^{\infty} \beta^{(n)} \mathcal{T}_{ij}^{(n)}, \tag{1}$$

where $\beta^{(n)}$ represents the n'th coefficient associated with a corresponding basis tensor $\mathcal{T}_{ij}^{(n)}$. The coefficients may range in complexity from constants to nonlinear functions of the principal invariants of the tensor bases. For many configurations, this infinite sum can be reduced to a finite sum by leveraging the Cayley-Hamilton theorem. This results in a reduced set of tensors termed a "minimal invariant basis" (e.g., Ref. 28). Using knowledge of the system physics, a minimal invariant basis can be derived. Once this basis is established, modeling exercises must address two key questions. (1) Which of the basis tensors are most important for capturing the physics at play? (2) How do the coefficients depend on principal invariants or system parameters?

Sparse regression has been shown to be successful at addressing the first task[14,29] and works well for the second task when constant coefficients are sufficient. However, when the system has a complex and large parameter space, as is the case for multiphase turbulence, constant coefficients are no longer sufficient for capturing physics across scales. In this situation, sparse regression is not efficient at determining the form of the coefficients and requires the modeler to supply all potential test functions to the algorithm manually. While this has important benefits for embedding physics-based reasoning and properties into the resultant model (e.g., form invariance), it implies a tedious, "guess-and-check" exercise if physics-based arguments can no longer be used to supply test functions. In the present method, we propose to offload this work to a gene expression algorithm when naivety in functional form is unavoidable. This effectively automates the process of evaluating trial functions for the coefficients while preserving the benefits of using sparse regression to inform the tensorial building blocks of the model.

The method can be summarized by three distinct modeling steps, as shown in Fig. 1, and outlined here for data spanning s unique conditions in the parameter space (in the context of multiphase, turbulent flows, these parameters might include the solids volume fraction and Reynolds number):

1. Use sparse regression to identify the basis tensors required to describe physics by optimizing

$$\widehat{\beta} = \min_{\beta} ||\mathbb{D} - \mathbb{T}\beta||_2^2 + \lambda||\beta||_1 \tag{2}$$

and assuming constant coefficients. Each base associated with a nonzero coefficient in $\widehat{\beta}$ is deemed to be "essential" and is retained in the final model. The surviving bases are then condensed into a subset of $\mathbb{T}$, denoted $T^{\subset}$.

2. For each of the s conditions, compute the ideal constant coefficients associated with the p essential bases, using ordinary least squares (OLS):

$$\widehat{\beta}_s^{\subset} = \min_{\widehat{\beta}_s^{\subset}} ||D_s - T_s^{\subset}\beta_s^{\subset}||_2^2, \tag{3}$$

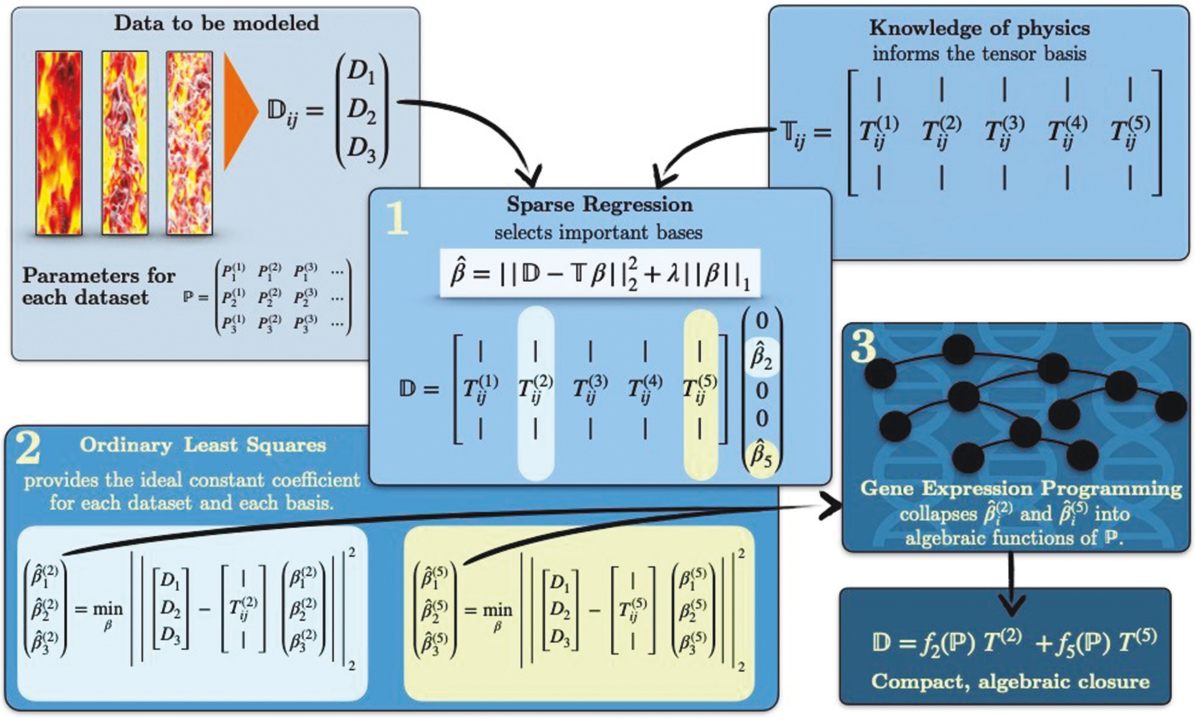

Fig. 1. The modeling approach has three steps. 1: Sparse regression identifies the important basis tensors. 2: OLS provides the ideal coefficients for each of the data sets for each of the identified bases. 3: GEP collapses the ideal coefficients for each case into a compact, algebraic closure.

where $\widehat{\beta}_s^{\subset}$ is of size $p \times 1$; D_s is of size $q \times 1$, where q is the amount of data in the dataset (e.g., the number of time steps); and $T_s^{\subset}$ is of size $q \times p$. Note that both $T^{\subset}$ and D require tensorial data to be reorganized as vertical vectors; see Ref. 14 for details. After this process has been done for all s conditions, concatenate each of the $\widehat{\beta}_s^{\subset}$ vectors into a matrix of size $p \times s$.

3. Finally, provide each p'th row of $\widehat{\beta}^{\subset}$ and matrix of parameters $\mathbb{P}$ associated with the s conditions as input to the GEP algorithm. The resulting functional model for $\widehat{\beta}^{\subset}$ effectively collapses the vector of discrete values for $\widehat{\beta}^{\subset}$ to a continuous, closed form with algebraic dependence on system parameters.

This modeling flow is illustrated in Fig. 1, where $s = 3$ and $p = 2$ for demonstration purposes. Pseudocode for this workflow is provided in the Supplementary Material.

III. SYSTEM UNDER CONSIDERATION

Multiphase flows span large parameter spaces, making modeling challenging. Thus, we use a simple gas-solid flow in which two-way coupling between the phases drives the underlying turbulence as a case study to evaluate the effectiveness of the proposed modeling framework. The same methodology will span bubbly flows in future work. In this configuration, rigid spherical particles are suspended in an unbounded (triply periodic) domain containing an initially quiescent gas. As particles settle under the influence of gravity, they spontaneously form clusters. Because of momentum exchange between phases, particles entrain the fluid, generating turbulence therein.[25,26,30] A frame of reference with the fluid phase is considered such that the mean streamwise fluid velocity is null. Key nondimensional numbers that characterize the system include the Reynolds number, as well as the Archimedes number, defined as

$$\mathrm{Ar} = (\rho_p/\rho_f - 1) d_p^3 g / \nu_f^2 \quad . \tag{4}$$

Alternatively, a Froude number can be introduced to characterize the balance between gravitational and inertial forces, defined as $\mathrm{Fr} = \tau_p^2 g / d_p$, where $\tau_p = \rho_p d_p^2 / (18 \rho_f \nu_f)$ is the particle response time.

The mean particle-phase volume fraction is varied from $0.001 \leq \langle \varepsilon_p \rangle \leq 0.05$, and gravity is varied from $0.8 \leq g \leq 8.0$. Here, angled brackets denote an average in all three spatial dimensions and time. Because of the

large density ratios under consideration, the mean mass loading, $\varphi = \langle \varepsilon_p \rangle \rho_p / \left(\langle \varepsilon_f \rangle \rho_f \right)$, ranges from $\mathcal{O}(10)$ to $\mathcal{O}(10^2)$, and consequently, two-way coupling between the phases is important. A large enough domain with a sufficiently large number of particles is needed to observe clustering. To enable simulations on this scale, we use an Eulerian-Lagrangian approach.[31] The fluid equations are solved on a staggered grid with second-order spatial accuracy and advanced in time with second-order accuracy using the semi-implicit Crank-Nicolson scheme. Particles are tracked individually in a Lagrangian frame of reference and integrated using a second-order Runge-Kutta method. Particle data (volume fraction and drag) are projected to the Eulerian mesh using a Gaussian filter described in Ref. 31.

IV. MULTIPHASE RANS EQUATIONS

Derivation of the single-phase RANS equations is done by directly averaging the Navier-Stokes equations. Derivation of the multiphase RANS equations, however, will retain additional physics if averaging is performed on the volume-filtered Navier-Stokes equations.[21] Volume fraction–weighted averages, or phase averaging (PA), analogous to Favre averaging of variable density flows, was previously derived.[26] For the relatively simple configuration used here, which is homogeneous in all spatial directions, statistically stationary in time, and symmetric in the counter-gravity direction, the transport equations for the fluid-phase Reynolds stresses can be reduced to two unique, nonzero components. In the streamwise (gravity-driven) direction, this equation is given as

$$\begin{aligned}\frac{1}{2}\frac{\partial \langle u_f'''^2 \rangle_f}{\partial t} = &\underbrace{\frac{1}{\rho_f}\left\langle p_f \frac{\partial u_f'''}{\partial x} \right\rangle}_{\text{pressure strain (PS)}} - \underbrace{\frac{1}{\rho_f}\left\langle \sigma_{f,1i} \frac{\partial u_f'''}{\partial x_i} \right\rangle}_{\text{viscous dissipation (VD)}} \\ &+ \underbrace{\frac{\varphi}{\tau_p^\star}\left(\langle u_f''' u_p'' \rangle_p - \langle u_f'''^2 \rangle_p \right)}_{\text{drag exchange (DE)}} + \underbrace{\frac{\varphi}{\tau_p^\star}\langle u_f''' \rangle_p \langle u_p \rangle_p}_{\text{drag production (DP)}} \\ &+ \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f''' \frac{\partial p_f'}{\partial x} \right\rangle_p}_{\text{pressure exchange (PE)}} - \underbrace{\frac{\varphi}{\rho_p}\left\langle u_f''' \frac{\partial \sigma_{f,1i}'}{\partial x_i} \right\rangle_p}_{\text{viscous exchange (VE)}}, \end{aligned} \tag{5}$$

where u_p is the particle-phase velocity in an Eulerian frame of reference and σ is the viscous stress tensor. The modified particle response time $\tau_p^\star$ is given as the particle response time τ_p normalized by the particle-phase–averaged drag correction $\langle F_d \rangle_p$ (see Ref. 29 for additional details). Here, $\langle (\cdot) \rangle_p = \langle \varepsilon_p (\cdot) \rangle / \langle \varepsilon_p \rangle$. Fluctuations about PA terms are denoted with a double prime. In a similar fashion, the PA operator in the fluid phase is defined as $\langle (\cdot) \rangle_f = \langle \varepsilon_f (\cdot) \rangle / \langle \varepsilon_f \rangle$. Fluctuations about the PA fluid velocity are given by $\boldsymbol{u}_f''' = \boldsymbol{u}_f(\boldsymbol{x}, t) - \langle \boldsymbol{u}_f \rangle_f$. With this, the fluid-phase TKE is given by $k_f = \langle \boldsymbol{u}_f''' \cdot \boldsymbol{u}_f''' \rangle_f / 2$. These definitions and notations are illustrated further in Fig. 2.

It is notable that all the terms appearing on the right-hand side of Eq. (5) are unclosed and require modeling. This work has already been carried out using sparse regression exclusively.[29] Here, we select the drag production (DP) term to demonstrate the present methodology. This term is chosen due to its importance in this class of flows. In the absence of mean shear, it is the only source of fluid-phase TKE. Additionally, all components of the DP tensor are identically zero, except for the contribution in the gravity aligned direction.

Several recent efforts have proposed closures for DP across both gas-solid and bubbly flows. These closures are generally postulated in the following form:

$$\frac{\varphi}{\tau_p^\star}\langle u_f''' \rangle_p \left(\langle u_p \rangle_p - \langle u_f \rangle_f \right) = C_I \, C_d |\langle u_p \rangle_p - \langle u_f \rangle_f|^3, \tag{6}$$

where C_d is the drag coefficient, $|\langle u_p \rangle_p - \langle u_f \rangle_f|$ is the mean slip velocity between the phases, and $0 \leq C_I \leq 1$ is a model constant. In the case of bubbly flows, C_I has been postulated as $\min(0.18\mathrm{Re}_p^{0.23}, 1) 0.75 \rho_f \langle \varepsilon_p \rangle / d_p$ (Ref. 32), $0.25(0.75 C_d \rho_f \langle \varepsilon_p \rangle / d_p$ (Ref. 33), and $1.44(0.75 C_D \rho_f \langle \varepsilon_p \rangle / d_p)$ (Ref. 34). For gas-solid flows, a similar closure structure was previously postulated by Ref. 21 in which the drift velocity was modeled as $\langle u_f''' \rangle_p = C_g \left(\langle u_p \rangle_p - \langle u_f \rangle_f \right)$, where $C_g \approx 0.6$ for homogeneous gravity-driven flows[26] and has a nonlinear dependence on $\langle \varepsilon_f \rangle$ in the channel flow.[35]

V. RESULTS AND DISCUSSION

We now demonstrate the modeling methodology presented in Sec. II on the multiphase case study summarized in Sec. III, focusing on the DP term $\mathcal{R}^{\mathrm{DP}}$ in particular. Here, we follow the three modeling steps as outlined previously.

In the first step, we conduct modeling of DP using sparse regression with embedded invariance and the assumption of constant coefficients to inform the bases

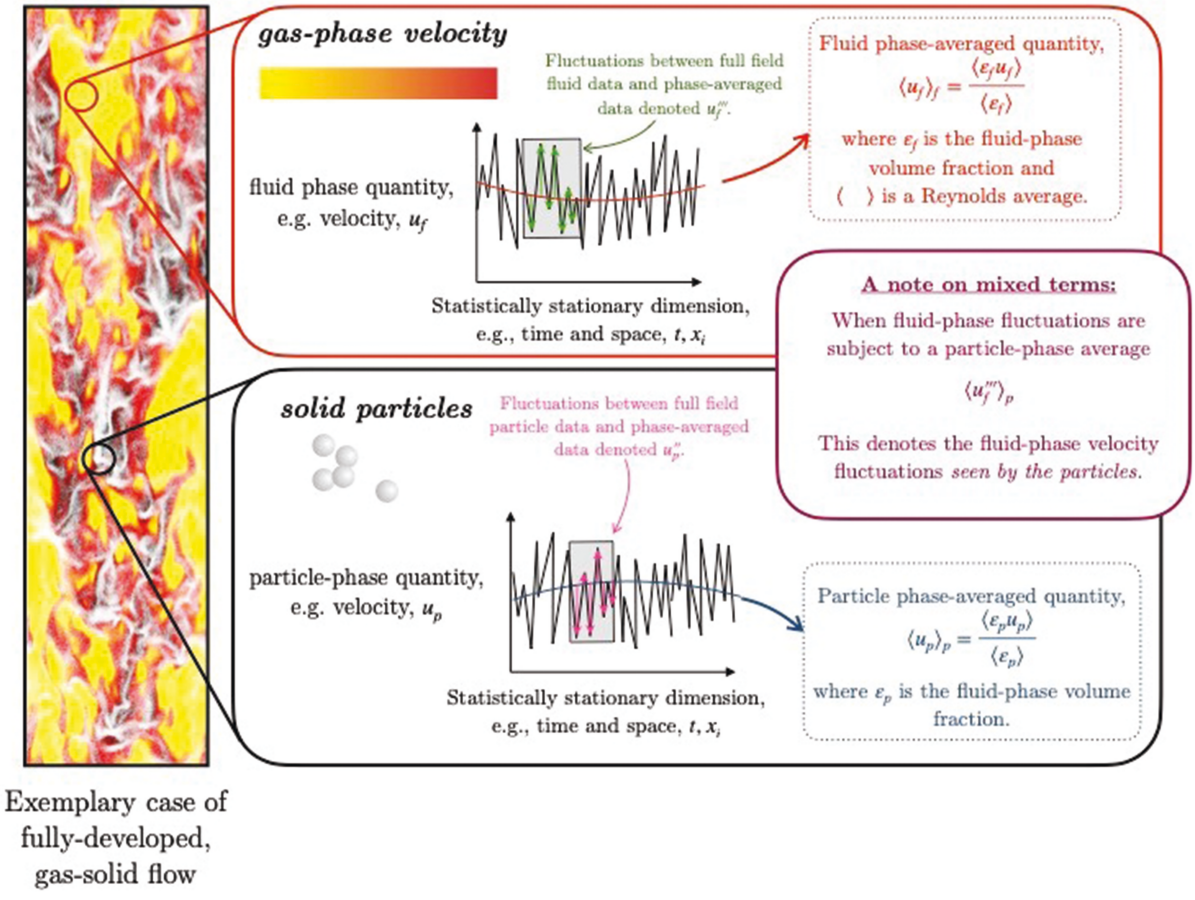

Fig. 2. Illustration of the PA definitions and notations used in the multiphase RANS equations, Eq. (5).

that comprise the reduced set $T^{\subset}$. The model consisting of the reduced basis tensors is given as

$$\mathcal{R}^{\mathrm{DP}} = \beta_1 \mathbb{I} + \beta_2 \widehat{\mathbb{U}}_r \quad , \tag{7}$$

where $\widehat{\mathbb{U}}_r$ is a tensor formulated using the mean slip velocity between the phases, $\mathbb{I}$ is the identity tensor, and the coefficients β_1 and β_2 have functional dependency upon configuration parameters that are unknown and cannot be informed by physics-based reasoning.

Next, we evaluate the ideal constant coefficients for each unique configuration studied, by conducting OLS and allowing the coefficients β_1 and β_2 to take on unique values for each configuration. In other words, the values of β_1 and β_2 associated with the case for $\langle \varepsilon_p \rangle = 0.001$ and $g = 0.8$ need not be the same as the values for $\langle \varepsilon_p \rangle = 0.05$ and $g = 2.4$.

As described in Sec. II, the ideal coefficients are arranged into two vectors: one for each of the basis tensors $\mathbb{I}$ and $\widehat{\mathbb{U}}_r$. Each vector of ideal coefficients is used as input, along with the associated parameters and invariants, to the GEP algorithm.[36] Here, the GEP algorithm selects models that reduce the R^2 between the ideal coefficient values and the candidate models, which are all functions of the parameters and invariants. This effectively collapses the vector of ideal coefficients to a single, compact algebraic expression.

The resultant model learned from this methodology is given as

$$\begin{aligned} \mathcal{R}^{\mathrm{DP}} = &\left(0.258\varphi + (0.03\varphi)^3 + 1.9\frac{\langle \varepsilon_p \rangle}{\mathcal{S}^{(2)}}\right)\mathbb{I} \\ &+ \left(1.9\varphi - 5.8\varphi^{1/2}\right)\widehat{\mathbb{U}}_r \quad , \end{aligned} \tag{8}$$

where $\mathcal{S}^{(2)}$ is a principal invariant, defined as $\mathrm{tr}\left(\widehat{\mathbb{U}}_r \widehat{\mathbb{R}}_f \widehat{\mathbb{R}}_p^2\right)$, and the basis tensor $\widehat{\mathbb{U}}_r$ is defined by the normalized slip tensor. This slip tensor is given as the outer product of the mean slip velocity, $\widehat{\mathbb{U}}_r = \left(\langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f\right) \otimes \left(\langle \boldsymbol{u}_p \rangle_p - \langle \boldsymbol{u}_f \rangle_f\right)$. The two other

basis tensors $\widehat{\mathbb{R}}_f$ and $\widehat{\mathbb{R}}_p$ are the anisotropic stress tensors associated with the fluid and particle phases, respectively. In terms of solution variables, the mean phase velocities are solved by associated momentum equations, and the Reynolds stresses are informed by transport equations in the multiphase RANS equations (see Ref. 29).This model has an error of 0.012, where the error is defined as

$$\epsilon = \frac{||\mathbb{D} - \mathbb{T}\widehat{\beta}||_2^2}{||\mathbb{D}||_2^2} . \tag{9}$$

This is comparable performance to the model learned using sparse regression exclusively ($\epsilon = 0.013$); however, the proposed method does not require a manual selection of trial functions for the coefficients, thus making it far more efficient from a modeling perspective.

Following the same method as is shown for DP, all the other terms appearing in the fluid-phase multiphase RANS equation [shown in Eq. (5)] are also modeled. Since the remaining terms in this equation [namely, pressure strain (PS), viscous dissipation (VD), and drag exchange (DE)] demonstrate more complex models when carrying out the sparse regression–only approach, some additional considerations are needed.

In order to determine a minimal set of basis tensors, sparse regression is carried out with a decreasing sparsity parameter λ until error is moderately reduced and a small subset of the basis tensors is selected. Since GEP will be used to infer the coefficients, it is no longer necessary to select λ such that the model error is minimized to the limit of zero. To choose λ such that the error approaches zero would unnecessarily increase the number of basis tensors included in the model. In other words, the sparse regression portion of the blended modeling approach acts as a "coarse tuning" knob, and the GEP portion conducts "fine tuning."

Once the reduced basis is identified, OLS is carried out to infer the ideal coefficients for each realization of the data. It is worth noting here that for this problem, and many others in engineering systems, the basis tensor matrix is rank deficient, making the OLS computation underdetermined. The reason OLS is underdetermined is that the number of candidate tensor bases exceeds the number of nonzero entries in $\mathbb{D}$.

Thus, a minimum-norm solution to OLS is carried out to circumvent this issue. Namely, the basis tensor subset, $\mathbb{T}^{\epsilon}$ with rank k, is decomposed using singular value decomposition (SVD) as $\mathbb{T}^{\epsilon} = U\Sigma V^T$. Then, the leading submatrix of Σ, which includes the k singular values of $\mathbb{T}^{\epsilon}$, is used to write a compact SVD of $\mathbb{T}'^{\epsilon}$ as $\mathbb{T}^{\epsilon} = U'\Sigma' V'^T$. While this step introduces an unrecoverable error, since $\mathbb{T}'^{\epsilon}$ is properly an approximation of $\mathbb{T}^{\epsilon}$, it also allows all of the idealized coefficients for the basis subset to have nonzero values, without the ad hoc partitioning of data.

Once the idealized set of coefficients corresponding to each basis has been determined, GEP is used to infer the compact, algebraic model. A summary of these steps and resulting models for all the nontrivial terms that appear in Eq. (5) are summarized in Tables I through V. For all the terms considered, the blended modeling approach results in models with similar or better accuracy and similar or increased simplicity compared with a sparse regression–only approach. Specifically, the model errors using a sparse regression–only approach for PS, VD, DE, and DP are 0.15, 0.07, 0.15, and 0.01, respectively. Using the blended GEP and sparse regression approach, resultant models for the previously listed terms are 0.06, 0.09, 0.08, and 0.01, respectively. Additionally, the number of basis tensors required for the blended models is maintained at 3 and 2 for PS and DP and reduced from 4 to 2 basis tensors for each of viscous diffusion and DE.

TABLE I

Summary of Reduced Order Forms for the Fluid-Phase Reynolds Stress Transport Equation*

Term	Reduced Model Form
Pressure strain	$\mathcal{R}^{\mathrm{PS}} = \beta_1^{\mathrm{PS}}\,\widehat{\mathbb{U}}_r + \beta_2^{\mathrm{PS}}\,\widehat{\mathbb{R}}_f + \beta_3^{\mathrm{PS}}\,\widehat{\mathbb{R}}_p$
Viscous dissipation	$\mathcal{R}^{\mathrm{VD}} = \beta_1^{\mathrm{VD}}\left(\widehat{\mathbb{R}}_f\widehat{\mathbb{R}}_p + \widehat{\mathbb{R}}_p\widehat{\mathbb{R}}_f\right) + \beta_2^{\mathrm{VD}}\left(\widehat{\mathbb{R}}_f\widehat{\mathbb{R}}_f\widehat{\mathbb{R}}_p + \widehat{\mathbb{R}}_p\widehat{\mathbb{R}}_f\widehat{\mathbb{R}}_f\right)$
Drag exchange	$\mathcal{R}^{\mathrm{DE}} = \beta_1^{\mathrm{DE}}\,\mathbb{I} + \beta_2^{\mathrm{DE}}\,\widehat{\mathbb{U}}_r$
Drag production	$\mathcal{R}^{\mathrm{DP}} = \beta_1^{\mathrm{DP}}\,\mathbb{I} + \beta_2^{\mathrm{DP}}\,\widehat{\mathbb{U}}_r$

*Here, sparse regression is used to determine a minimal basis set for modeling. Coefficients β are then determined using OLS and GEP.

TABLE II

Summary of the Coefficients Learned by the GEP Algorithm for PS*

Coefficient	Learned Closure
$\beta_1^{PS} =$	$0.07\varphi - 1.46e^{-\mathcal{S}^{(3)}} - 0.85\mathcal{S}^{(3)}\varphi - (0.026\varphi)^2 + 1.26$
$\beta_2^{PS} =$	$0.06\varphi - 0.72\mathcal{S}^{(3)}\varphi - (0.022\varphi)^2 - 0.03$
$\beta_3^{PS} =$	$1.29\frac{\langle\varepsilon_p\rangle}{\mathcal{S}^{(1)}} - 57.2\langle\varepsilon_p\rangle - \left(25.16\langle\varepsilon_p\rangle\right) - 0.04$

*The resultant model using the blended approach yields a model error that is over twofold less than that with the sparse regression approach (0.06 and 0.15, respectively).

TABLE III

Summary for the Coefficients Learned by the GEP Algorithm for Viscous Diffusion*

Coefficient	Learned Closure
$\beta_1^{VD} =$	$0.788\mathcal{S}^{(1)}\varphi - 56.9\langle\varepsilon_p\rangle + 0.0002\langle\varepsilon_p\rangle^2 - 0.158$
$\beta_2^{VD} =$	$0.407\mathcal{S}^{(3)} - 15.4\langle\varepsilon_p\rangle - 20.8e^{-\varphi} - 3.29\ln\left(\langle\varepsilon_p\rangle\right) + 0.776\mathcal{S}^{(3)}\varphi - 15.7$

*Model errors between a sparse regression–only approach and the blended approach are similar (0.07 and 0.09, respectively), However, the blended modeling method resulted in a model with increased simplicity (two basis tensors required compared with five), for a similar accuracy.

TABLE IV

Summary of the Coefficients Learned by the GEP Algorithm for DE*

Coefficient	Learned Closure
$\beta_1^{DE} =$	$0.142 - 1.6\left(\mathcal{S}^{(1)}\varphi\right)^3 - 0.29\varphi$
$\beta_2^{DE} =$	$-1.7\left(\mathcal{S}^{(3)}\varphi\right)^2 - 2.88$

*These closures reduce model error by nearly half compared with sparse regression alone ($\in =$ 0.08 and 0.15, respectively).

In considering the resultant models, it is notable that in the limit of no particles (i.e., $\widehat{\mathbb{U}}_r$, $\widehat{\mathbb{R}}_p = 0$), the PS reduces to a form similar to what is seen in traditional single-phase turbulence modeling.[37] Namely, for single-phase PS, the slow PS correlation is most generally written as

$$\mathcal{R}^{PS} = \beta_1 b_{ij} + \beta_2\left(b_{ik}b_{kj} - \frac{1}{3}II_b\delta_{ij}\right), \tag{10}$$

where b_{ij} is the Reynolds stress anisotropy tensor (equivalent to $\widehat{\mathbb{R}}_f$ in the case of no particles). The first term in this model, which is echoed in the multiphase PS model in this work, is responsible for models following linear paths on the anisotropy invariant map in single-phase flows. The second term captures nonlinear effects.

V.A. Comparison with a Nonblended Approach

As a counter argument to the blended modeling approach, we also allowed the GEP algorithm to learn the full model without sparse regression (i.e., the mean values of DP, all 24 basis tensors, the principal invariants, and configuration parameters from the Euler-Lagrange simulations were provided to the GEP algorithm as input). The learned model is given as

TABLE V

Summary of Model Error and Complexity for Each of the Terms Arising in Eq. (5)*

Term	Sparse Regression		Sparse Regression + GEP	
	ϵ	Number of Terms	ϵ	Number of Terms
Pressure strain	0.15	3	0.06	3
Drag production	0.01	2	0.01	2
Viscous dissipation	0.07	4	0.09	2
Drag exchange	0.15	4	0.08	2

*In comparing sparse regression against a blended sparse regression and GEP approach, the blended approach results in similar or better model errors and fewer tensor bases required for accuracy.

$$\mathcal{R}^{\mathrm{DP}} = 24.4\widehat{\mathbb{U}}_r + 30.4e^{-\widehat{\mathbb{R}}_p} - 1.42\left(\widehat{\mathbb{U}}_r\widehat{\mathbb{R}}_p + \left(\widehat{\mathbb{U}}_r\widehat{\mathbb{R}}_p\right)^{\mathrm{T}}\right)^{1/2} + 1.41\times 10^5\left(\widehat{\mathbb{U}}_r^2\widehat{\mathbb{R}}_f + \left(\widehat{\mathbb{U}}_r^2\widehat{\mathbb{R}}_f\right)^{\mathrm{T}}\right)^2\langle\varepsilon_p\rangle^2 - 30.4\,, \tag{11}$$

with associated error 0.13 (an order of magnitude higher than the blended or sparse regression–only approach). This degradation in performance can be attributed to the fact that the model now depends on $\widehat{\mathbb{R}}_p$ and $\widehat{\mathbb{R}}_f$, the particle and fluid-phase Reynolds stress tensors, in addition to $\widehat{\mathbb{U}}_r$. On a fundamental level, since the DP is a gravity-based phenomenon (i.e., TKE is generated solely due to the presence of gravity in this configuration), we can anticipate that $\widehat{\mathbb{U}}_r$ would be the predominant tensor from the basis for describing the physics. Additionally, since the Reynolds stresses contain nonzero diagonal entries, including these terms makes it difficult to drive the cross stream directions of the DP model to zero. Finally, and perhaps most importantly, GEP does not enforce the relation that the resultant model be linear with respect to the basis tensors. The stipulation of linearity in the basis tensors is critical for ensuring form invariance in the resultant model and for ensuring a physics-based description of the data, as described by Eq. (1). These results suggest that sparse regression and GEP are both needed in order to select a minimal set of tensors to describe physics and automate the complex dependencies of the coefficients when physical intuition cannot guide this process.

VI. CONCLUSIONS

As evidenced by this work, a modeling framework that leverages both sparse regression and GEP has shown preliminary success in modeling complex flows. This methodology combines the benefits of both modeling approaches. Namely, it leverages the fact that sparse regression can incorporate known physical constraints (such as linearity in a specified basis) and results in a compact algebraic model. GEP automates the determination of algebraic models for coefficients for which physical constraints are not known. Combining these two modeling strategies results in models with a high level of accuracy and, in some cases, improved simplicity compared with a sparse regression–only approach. Importantly, this method automates the portions of the modeling process for which the practitioner cannot reasonably provide superior oversight (for example, no physical insights exist to constrain the solution).

Extending this blended framework to a more expansive two-phase data set, spanning gas-solid flows and liquid-gas flows, is the subject of current work. Because of the automation properties of GEP combined with the physics-based structure of sparse regression, we hypothesize that developing closures for data spanning particle-laden to bubbly flows will result in models that are invariant, intepretable, easy to use, and toward universal applicability for two-phase flow.

Supplemental Data

Supplemental data for this article can be accessed online at https://doi.org/10.1080/00295450.2023.2178251.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF CAREER, CBET-1846054). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

S. Beetham http://orcid.org/0000-0002-2823-2394

References

1. K. DURAISAMY, G. IACCARINO, and H. XIAO, "Turbulence Modeling in the Age of Data," *Annu. Rev. Fluid Mech.*, **51**, 357 (2019).
2. B. TRACEY, K. DURAISAMY, and J. J. ALONSO, "A Machine Learning Strategy to Assist Turbulence Model Development," *Proc. 53rd AIAA Aerospace Sciences Mtg.*, AIAA 2015-1287, American Institute of Aeronautics and Astronautics (2015).
3. M. MILANO and P. KOUMOUTSAKOS, "Neural Network Modeling for Near Wall Turbulent Flow," *J. Comput. Phys.*, **182**, 1 (2002).
4. C. LU, "Artificial Neural Network for Behavior Learning from Meso-Scale Simulations, Application to Multi-Scale Multimaterial Flows," PhD Thesis, University of Iowa (2010).
5. E. RAJABI and M. R. KAVIANPOUR, "Intelligent Prediction of Turbulent Flow over Backward-Facing Step Using Direct Numerical Simulation Data," *Eng. Appl. Comput. Fluid Mech.*, **6**, *4*, 490 (2012).
6. K. DURAISAMY and P. A. DURBIN, "Transition Modeling Using Data Driven Approaches," *Proc. Summer Program*, p. 427, Center for Turbulence Research (2014); https://web.stanford.edu/group/ctr/Summer/SP14/09_Large-eddy_simulation/07_duraisamy.pdf.
7. K. DURAISAMY, Z. J. ZHANG, and A. P. SINGH, "New Approaches in Turbulence and Transition Modeling Using Data-Driven Techniques," *Proc. 53rd AIAA Aerospace Sciences Mtg.*, AIAA 2015-1284, American Institute of Aeronautics and Astronautics (2015).
8. M. MA, J. LU, and G. TRYGGVASON, "Using Statistical Learning to Close Two-Fluid Multiphase Flow Equations for Bubbly Flows in Vertical Channels," *Int. J. Multiphase Flow*, **85**, 336 (2016).
9. J. LING, A. KURZAWSKI, and J. TEMPLETON, "Reynolds Averaged Turbulence Modeling Using Deep Neural Networks with Embedded Invariance," *J. Fluid Mech.*, **807**, 155 (2016).
10. M. BODE et al., "Deep Learning at Scale for Subgrid Modeling in Turbulent Flows: Regression and Reconstruction," *arXiv:1910.00928v1* (2019).
11. W. LIU and J. FANG, "Iterative Framework of Machine-Learning Based Turbulence Modeling for Reynolds-Averaged Navier-Stokes Simulations," *arXiv:1910.01232v1* (2019).
12. S. L. BRUNTON, J. L. PROCTOR, and J. N. KUTZ, "Discovering Governing Equations from Data by Sparse Identification of Nonlinear Dynamical Systems," *Proc. Natl. Acad. Sci.*, **113**, *15*, 3932 (2016).
13. M. SCHMELZER, R. P. DWIGHT, and P. CINNELLA, "Discovery of Algebraic Reynolds-Stress Models Using Sparse Symbolic Regression," *Flow Turbul. Combust.*, **104**, *2*, 579 (2020).
14. S. BEETHAM and J. CAPECELATRO, "Formulating Turbulence Closures Using Sparse Regression with Embedded Form Invariance," *Phys. Rev. Fluids*, **5**, 084611 (2020).
15. M. REISSMANN et al., "Application of Gene Expression Programming to A-Posteriori LES Modeling of a Taylor Green Vortex," *J. Comput. Phys.*, **424**, 109859 (2021).
16. C. KASTEN et al., "Modeling Subgrid-Scale Scalar Dissipation Rate in Turbulent Premixed Flames Using Gene Expression Programming and Deep Artificial Neural Networks," *Phys. Fluids*, **34**, *8*, 085113 (2022).
17. J. DOMINIQUE et al., "Inferring Empirical Wall Pressure Spectral Models with Gene Expression Programming," *J. Sound Vib.*, **506**, 116162 (2021).
18. S. SAMADIANFARD, "Gene Expression Programming Analysis of Implicit Colebrook-White Equation in Turbulent Flow Friction Factor Calculation," *J. Pet. Sci. Eng.*, **92**, 48 (2012).
19. J. WEATHERITT and R. SANDBERG, "A Novel Evolutionary Algorithm Applied to Algebraic Modifications of the RANS Stress-Strain Relationship," *J. Comput. Phys.*, **325**, 22 (2016).
20. Y. ZHAO et al., "RANS Turbulence Model Development Using CFD-Driven Machine Learning," *J. Comput. Phys.*, **411**, 109413 (2020).
21. R. O. FOX, "On Multiphase Turbulence Models for Collisional Fluid-Particle Flows," *J. Fluid Mech.*, **742**, 368 (2014).
22. S. BALACHANDAR and J. K. EATON, "Turbulent Dispersed Multiphase Flow," *Annu. Rev. Fluid Mech.*, **42**, 111 (2010).
23. J. CAPECELATRO, O. DESJARDINS, and R. O. FOX, "On the Transition Between Turbulence Regimes in Particle-Laden Channel Flows," *J. Fluid Mech.*, **845**, 499 (2018).
24. D. C. MILLER et al., "Carbon Capture Simulation Initiative: A Case Study in Multiscale Modeling and New Challenges," *Annu. Rev. Chem. Biomol. Eng.*, **5**, 301 (2014).
25. J. CAPECELATRO, O. DESJARDINS, and R. O. FOX, "Numerical Study of Collisional Particle Dynamics in Cluster-Induced Turbulence," *J. Fluid Mech.*, **747**, R2, 1 (2014).

26. J. CAPECELATRO, O. DESJARDINS, and R. O. FOX, "On Fluid-Particle Dynamics in Fully Developed Cluster-Induced Turbulence," *J. Fluid Mech.*, **780**, 578 (2015).

27. S. B. POPE, *Turbulent Flows*, Cambridge University Press (2000).

28. A. J. M. SPENCER and R. S. RIVLIN, "The Theory of Matrix Polynomials and Its Application to the Mechanics of Isotropic Continua," *Arch. Ration. Mech. Anal.*, **2**, 309 (1958).

29. S. BEETHAM, R. O. FOX, and J. CAPECELATRO, "Sparse Identification of Multiphase Turbulence Closures for Coupled Fluid-Particle Flows," *J. Fluid Mech.*, **914**, A11 (2021).

30. F. DABBAGH and S. SCHNEIDERBAUER, "Anisotropy Characterization of Turbulent Fluidization," *Phys. Rev. Fluids*, **7**, *9*, 094301 (2022).

31. J. CAPECELATRO and O. DESJARDINS, "An Euler-Lagrange Strategy for Simulating Particle-Laden Flows," *J. Comput. Phys.*, **238**, 1 (2013).

32. T. MA et al., "Direct Numerical Simulation–Based Reynolds-Averaged Closure for Bubble-Induced Turbulence," *Phys. Rev. Fluids*, **2**, *3*, 034301 (2017).

33. M. COLOMBO and M. FAIRWEATHER, "Multiphase Turbulence in Bubbly Flows: RANS Simulations," *Int. J. Multiphase Flow*, **77**, 222 (2015).

34. R. T. LAHEY and D. A. DREW, "An Analysis of Two-Phase Flow and Heat Transfer Using a Multidimensional, Multi-Field, Two-Fluid Computational Fluid Dynamics (CFD) Model," *Proc. Japan/US Seminar on Two-Phase Flow Dynamics*, Santa Barbara, California, June 5–8, 2000.

35. J. CAPECELATRO, O. DESJARDINS, and R. O. FOX, "Strongly-Coupled Gas-Particle Flows in Vertical Channels. Part II: Turbulence Modeling," *Phys. Fluids*, **28**, 1 (2016).

36. D. SEARSON, "GPTIPS: Genetic Programming & Symbolic Regression for MATLAB," http://gptips.sourceforge.net (2009).

37. S. SARKAR and C. G. SPEZIALE, "A Simple Nonlinear Model for the Return to Isotropy in Turbulence," *Phys. Fluids: Fluid Dyn.*, **2**, *1*, 84 (1990).