RESEARCH

DNS-Based Turbulent Closures for Sediment Transport Using Symbolic Regression

Yvonne Stöcker[1,2] · Christian Golla[1] · Ramandeep Jain[1] · Jochen Fröhlich[1] · Paola Cinnella[3]

Abstract

This work aims to improve the turbulence modeling in RANS simulations for particle-laden flows. Using DNS data as reference, the errors of the model assumptions for the Reynolds stress tensor and turbulence transport equations are extracted and serve as target data for a machine learning process called SpaRTA (Sparse Regression of Turbulent Stress Anisotropy). In the present work, the algorithm is extended so that additional quantities can be taken into account and a new modeling approach is introduced, in which the models can be expressed as a scalar polynomial. The resulting corrective algebraic expressions are implemented in the RANS solver SedFoam-2.0 for cross-validation. This study shows the applicability of the SpaRTA algorithm to multi-phase flows and the relevance of incorporating sediment-related quantities to the set of features from which the models are assembled. An average improvement of ca. thirty percent on various flow quantities is achieved, compared to the standard turbulence models.

Keywords Turbulence modeling · Multi-phase flows · Machine learning · Data-driven

1 Introduction

Sediment transport is a phenomenon that arises in a multitude of situations in nature and technological applications, ranging from the morphological evolution of coastal and riverine landscapes (Ribberink and Al-Salem 1994; Da Silva and Yalin 2017) to the mechanical degradation of industrial machinery due to its contribution to erosion and siltation (Chauchat et al. 2017; Drew 1983). Consequently, the prediction of fluid and sediment behavior is crucial for designing and optimizing many industrial processes (Yang et al. 2019). To accurately describe the transport of sediment, it is necessary to not only account for the mean flow, but also the turbulent fluctuations of the fluid, as these have a substantial impact

✉ Christian Golla
christian.golla@tu-dresden.de

1 Institute of Fluid Mechanics, Technische Universität Dresden, Dresden, Germany

2 Institute of Fluid Mechanics and Heat Transfer, Technische Universität Wien, Vienna, Austria

3 Institute Jean le Rond d'Alembert, Sorbonne Université, Paris, France

on the motion of particles (Sommerfeld 2001). Furthermore, the shape of the particles has a strong impact on the flow regimes (Allen 1985; Sommerfeld and Qadir 2018; Jain et al. 2021). However, due to the complexity of near-bed processes and the chaotic nature of turbulent flows, computation of these phenomena poses significant challenges (Vowinckel 2021).

During the past decades, direct numerical simulation (DNS) has emerged as a promising approach for investigating sediment transport. With modern hardware, DNS can resolve a relatively wide range of active spatial and temporal scales of the flow. It has been employed to study various aspects of sediment transport, including the mobilization of single particles (Vowinckel et al. 2016), the evolution of the sediment bed formed by spherical particles in laminar and turbulent flow (Kempe et al. 2014; Kidanemariam and Uhlmann 2014), the effect of the roughness of channel walls (Jain et al. 2017; Saccone et al. 2022) or the particle shape on the flow (Jain et al. 2021), to name but a few examples. These simulations provide valuable insight into the complex interaction between the particle and the fluid phase in the various circumstances. However, due to their extremely high computational cost, DNS are often infeasible for simulating flows of engineering interest.

Usually, in engineering applications, averaged or macroscopic flow information is sufficient. The fluctuations are only important in the extent of their effect on the mean flow, such as Reynolds stresses in turbulent flows (Drew 1983). Hence, it is preferred to solve the governing equations only for mean flow quantities, governed by the Reynolds-averaged Navier–Stokes (RANS) equations, with the mean velocity as main unknowns and the impact of turbulent fluctuations on the mean flow being modeled entirely. This makes the RANS simulations much more cost-efficient than DNS, but also more uncertain, because of the drastic modeling assumptions introduced to close the averaged system of equations.

For two-phase flows encountered in sediment transport problems, the modeling is even more complex than in single-phase flow models due to the interaction between the phases. The continuous fluid phase is treated in an Eulerian way. For the disperse particle phase two approaches are commonly used. Euler-Lagrange methods track individual particles which is demanding and costly. Euler-Euler methods represent both phases (fluid and sediment) as a continuum filling the domain with a certain volume fraction, as the result of some local averaging procedure. Both phases obey own dynamical equations for mass and momentum. The respective momentum equations are coupled through buoyancy and drag forces, and further unclosed terms represent particle stresses and fluid-particle interaction (Cheng et al. 2017). Accurate modeling of fluid turbulence, fluid-particle and particle-particle interaction is crucial for predicting the behavior of both phases and constitutes the fundamental challenge when devising such methods.

Extensive theoretical and computational effort has been made to increase the accuracy in capturing the complex characteristics of turbulent multi-phase flows (Zhu et al. 2022). As a result, numerous closure models are available, and have been utilized during the past decades for fluid turbulence in two-phase flows, including mixing length (Jenkins and Henes 1998; Dong and Zhang 1999), Reynolds stress (Jha and Bombardelli 2009), k-ε (Hsu et al. 2003; Bakhtyar et al. 2009; Cheng et al. 2017) and k-ω models (Jha and Bombardelli 2009; Yoon and Kang 2005). The particle stress can be modeled, e.g., by the kinetic theory of granular flows (Jenkins and Henes 1998; Cheng et al. 2017) or dense granular flow rheology (Revil-Baudard and Chauchat 2013; Chauchat 2018; Lee et al. 2016).

The parameters of the closure models are typically calibrated with data from experiments or DNS of simple canonical flows. However, the accuracy of predictions usually decreases when the models are applied to problems that differ from the calibration setup. Following Duraisamy et al. 2019, errors arising from RANS closure models can be

classified into four categories, originating from different stages of modeling: (1) loss of information due to averaging the Navier–Stokes equations, (2) uncertainties in the assumption that the unclosed terms can be represented as a function of mean flow quantities, (3) uncertainties in the specific functional form of the model and (4) uncertainties in the model coefficients.

Ways to quantify model uncertainties in RANS simulations are reviewed in Xiao and Cinnella (2019), where a distinction between parametric and non-parametric approaches is made. The former focus on the model coefficients (i.e., the fourth uncertainty category) while neglecting errors in the functional form of the model. It is, therefore, limited to the accuracy of the underlying model form, as the calibration of coefficients can not correct the inherent inadequacy. A non-parametric approach addresses these structural uncertainties of models or model terms (i.e., the third uncertainty category), which allows a more general estimation of the model deficiency. This can be used to investigate, e.g., uncertainties in eddy viscosity models (Dow and Wang 2011; Oliver and Moser 2011; Schmelzer et al. 2020) or single terms in the turbulent transport equations (Singh and Duraisamy 2016).

The knowledge of the uncertainties can then be utilized to improve existing closure models and develop more comprehensive models. For this purpose, machine learning (ML) provides a flexible way to extract complex patterns and information from the large datasets (Zhu et al. 2022). The objective of supervised ML is to translate input variables, such as averaged flow quantities, to appropriate output data. The algorithm learns the relation from training data by comparing the output to the desired result (e.g., an unclosed term or the error of an existing model) and optimizes the free parameters to improve its accuracy constantly.

Numerous research efforts have been published which are concerned with supervised ML to improve RANS predictions of single-phase flows. Examples include neural networks that learn corrections to the Reynolds stress anisotropy (Ling et al. 2016, Duraisamy et al. 2017), or turbulence model transport equations in single-phase flows (Duraisamy et al. 2017, Tracey et al. 2015). Random forest regression has been employed to learn corrective functional terms for the modeled Reynolds stress tensor (Wang et al. 2017; Wu et al. 2018). Gene Expression Programming has been used for symbolic regression of the anisotropy of the Reynolds stress tensor (Weatheritt and Sandberg 2016, 2017). Sparse regression has been applied to learn corrective functional terms for single-phase k-ω turbulence model (Schmelzer et al. 20202020).

For multi-phase flows, the literature is scarcer. It includes attempts of learning the interfacial momentum transfer in bubbly flows by means of neural networks (Bao et al. 2021), as well as symbolic regression of closure models for the transport equations of the fluid-phase Reynolds stresses in gas-solid flows (Beetham et al. 2021). ML has also been applied to coastal sediment transport to predict sediment properties, concentration and flux (Goldstein et al. 2019), for example. Applications of ML to disperse turbulent multi-phase flows, like Beetham and Capecelatro (2023), are still rare.

This paper aims to enhance the accuracy of closure models in turbulent flows involving sediment transport. We focus in particular on the representation of the Reynolds stress tensor and on a two-phase version of the k-ε turbulence model. The study is based on the DNS results of Jain et al. (2021), which investigated the influence of the particle shape on sediment transport in turbulent open-channel flows. Three simulations were performed using different non-spherical particle shapes and one simulation with spherical particles. The results showed the formation of moving rough sediment beds with dune-like structures.

In the present work, RANS simulations were conducted with the aim of reproducing the DNS results. For this purpose, the closure models for particle and fluid stress available

in SedFoam-2.0 were utilized. SedFoam-2.0 is a three-dimensional two-phase solver implemented by Cheng et al. (2017), Chauchat et al. (2017) in the open-source CFD toolbox Open-FOAM. It has already been applied to various scenarios, such as sheet flows (i.e., a thick layer of particles is suspended and transported above the static sediment bed, and no bed ripples form), particle sedimentation, and other configurations, such as scour development downstream of an apron (Cheng et al. 2017, Chauchat et al. 2017, Nagel et al. 2020, Mathieu et al. 2022).

When applied to the cases described in Jain et al. (2021), however, the accuracy of the closure models is not sufficient. With the aim of improving turbulence modeling for the class of flows considered, the error generated by the functional form of the model is extracted, which refers to the third category of errors mentioned above. A non-parametric approach is then used to find relationships between such errors and a suitable set of features. The resulting corrective terms are then added to the baseline turbulence model.

Specifically, in the present work we propose an extension of the deterministic symbolic regression algorithm SpaRTA (Sparse Regression of Turbulent Stress Anisotropy), initially introduced by Schmelzer et al. (2020) for correcting the predictions of RANS models in separated single-phase flows. SpaRTA is a sparse symbolic regression algorithm, a form of "open-box" ML allowing to learn tangible mathematical expressions of corrective terms for the constitutive equation of the Reynolds stress tensor and for the turbulence transport equations. SpaRTA is first modified to address two-phase flows, and then trained against DNS data from Jain et al. (2021). The resulting corrective terms are then integrated into the SedFoam-2.0 solver, and CFD simulations are conducted for validation. The extended SpaRTA approach, termed Sed-SpaRTA, is applied to the case with spherical particles as a first proof-of-concept and may later be extended to other test cases in Jain et al. (2021) to account for the effect of particle shape or examine entirely different flow configurations.

This paper is structured as follows: The second section describes the Euler-Euler framework and the utilized modeling procedure, focusing on the Reynolds stress tensor and turbulence transport equations of the k-ε turbulence model. The setup of the DNS and RANS simulations is presented in Sect. 3. Section 4 compares the DNS and baseline RANS results for the spherical and ellipsoidal test cases, and the model errors are quantified for the spherical case in Sect. 5. A description of the ML approach SpaRTA is given in the sixth section, including a summary of the methodology presented in Schmelzer et al. (2020) and the modifications applied here regarding two-phase flows. The models obtained from Sed-SpaRTA are then implemented in the CFD solver, and new simulations are performed for validation. The results are presented in Sects. 7 and 8. A summary of the findings is given in Sect. 9, followed by an outlook on further studies.

2 Euler-Euler framework and RANS modeling

To make the paper self-contained, this section briefly recalls the Euler-Euler approach of Chauchat et al. (2017) which constitutes the framework of the subsequent modeling. The strategy to describe the sediment as a continuous phase on a fixed, Eulerian grid is based on the phase-averaged Navier-Stokes equations (Sommerfeld 2017; Bakhtyar et al. 2009). The local average of a quantity ψ is computed as

$$\overline{\psi}(\mathbf{x}, t) = \frac{1}{V_0} \int_{V_0} \psi(\mathbf{x}, t) \mathrm{d}V, \tag{1}$$

where V_0 is the averaging volume.

In a multi-phase flow, the phase indicator function χ^{p} is employed, with p = f for the fluid and p = s for the solid, i.e. the sediment. If $\overline{\chi^{\mathrm{p}}} \neq 0$, the phase-weighted local average then reads (Burns et al. 2004)

$$\widetilde{\psi}^{\mathrm{p}} = \frac{\overline{\chi^{\mathrm{p}}\psi}}{\overline{\chi^{\mathrm{p}}}}. \tag{2}$$

Fluctuations relative to the phase-average are defined as

$$\psi^{\mathrm{p}'} = \psi - \widetilde{\psi}^{\mathrm{p}}. \tag{3}$$

The volume fraction of fluid, ϕ, and the volume fraction of sediment, $\alpha = 1 - \phi$, correspond to the locally averaged phase indicator functions for fluid and solid, respectively. These definitions result in the following formulation of the governing Euler-Euler equations for each phase (Chauchat et al. 2017)

$$\frac{\partial \phi}{\partial t} + \frac{\partial \phi \widetilde{u}_i^{\mathrm{f}}}{\partial x_i} = 0, \tag{4}$$

$$\frac{\partial \alpha}{\partial t} + \frac{\partial \alpha \widetilde{u}_i^{\mathrm{s}}}{\partial x_i} = 0, \tag{5}$$

$$\rho^{\mathrm{f}}\frac{\partial \phi \widetilde{u}_i^{\mathrm{f}}}{\partial t} + \rho^{\mathrm{f}}\frac{\partial \phi \widetilde{u}_i^{\mathrm{f}}\widetilde{u}_j^{\mathrm{f}}}{\partial x_j} = -\phi\frac{\partial \widetilde{p}^{\mathrm{f}}}{\partial x_i} + \rho^{\mathrm{f}}\frac{\partial \phi \widetilde{R}_{ij}^{\mathrm{f}}}{\partial x_j} + \rho^{\mathrm{f}}\frac{\partial \phi \widetilde{r}_{ij}^{\mathrm{f}}}{\partial x_j} + \phi\widetilde{f}_i + \phi\rho^{\mathrm{f}}g_i + \widetilde{\mathcal{M}}_i^{\mathrm{fs}}, \tag{6}$$

$$\rho^{\mathrm{s}}\frac{\partial \alpha \widetilde{u}_i^{\mathrm{s}}}{\partial t} + \rho^{\mathrm{s}}\frac{\partial \alpha \widetilde{u}_i^{\mathrm{s}}\widetilde{u}_j^{\mathrm{s}}}{\partial x_j} = -\alpha\frac{\partial \widetilde{p}^{\mathrm{f}}}{\partial x_i} + \rho^{\mathrm{s}}\frac{\partial \alpha \widetilde{R}_{ij}^{\mathrm{s}}}{\partial x_j} + \rho^{\mathrm{s}}\frac{\partial \alpha \widetilde{r}_{ij}^{\mathrm{s}}}{\partial x_j} + \alpha\widetilde{f}_i + \alpha\rho^{\mathrm{s}}g_i - \widetilde{\mathcal{M}}_i^{\mathrm{fs}}, \tag{7}$$

with $i = x, y, z$ the streamwise, vertical and spanwise component, respectively. In these equations, u_i^{p} is the fluid or sediment velocity, and $\widetilde{p}^{\mathrm{f}}$ the fluid pressure, also occurring in the equation of the sediment, $\widetilde{f}_i$ a volume force driving the flow and g_i the gravitational acceleration. The fluid stresses consist of a large-scale component $\widetilde{R}_{ij}^{\mathrm{f}} = -\overline{u_i^{\mathrm{f}'}u_j^{\mathrm{f}'}}$, i.e. the Reynolds stress tensor, and the fluid grain-scale stress tensor $\widetilde{r}_{ij}^{\mathrm{f}}$, which accounts for the viscous stress and the effects of fluid-particle interactions on the grain scale. The momentum exchange between the fluid and the sediment is denoted $\widetilde{\mathcal{M}}_i^{\mathrm{fs}}$ (Hsu et al. 2003).

For closure, the particle stress terms $\widetilde{R}_{ij}^{\mathrm{s}}$ and $\widetilde{r}_{ij}^{\mathrm{s}}$ in Eq. (7) are combined and then decomposed into a deviatoric and an isotropic part, which correspond to the particle normal stress (or pressure) and the particle shear stress, respectively. The particle pressure can be split into a collisional or shear-induced part and a permanent contact component. The model expression for the latter can be found in Chauchat et al. (2017). The contact component becomes dominant in regions of sediment concentration close to the maximum packing limit.

There are two main modeling approaches for the collisional particle normal stress: The kinetic theory of granular flows and the dense granular flow rheology. The kinetic theory is appropriate for low to moderate particle concentration. The dense granular flow rheology is

also called $\mu(I)$-rheology, which relates the dynamic friction coefficient μ to a dimensionless inertial number I (MiDi 2004; Forterre and Pouliquen 2008). Both approaches have corresponding treatments for the particle shear stress. In the dense granular flow rheology, I is computed differently for the grain inertia and viscous regimes, depending on the local Stokes and particulate Reynolds number (Andreotti et al. 2013). Details on the closure models for particle normal and shear stress can be found in Hsu et al. (2003), Chauchat et al. (2017).

Following Schmelzer et al. (2020), the Reynolds stress tensor $\widetilde{R}_{ij}^{\mathrm{f}}$ is decomposed into a deviatoric part $2kb_{ij}$ and an isotropic part $\frac{2}{3}k\delta_{ij}$

$$\widetilde{R}_{ij}^{\mathrm{f}} = 2k\left(b_{ij} - \frac{1}{3}\delta_{ij}\right) \tag{8}$$

with $k = k^{\mathrm{f}} = \frac{1}{2}\overline{\chi^{\mathrm{f}}u_i^{\mathrm{f}\prime}u_i^{\mathrm{f}\prime}}$ the turbulent kinetic energy of the fluid (TKE). The Boussinesq assumption

$$b_{ij} \approx b_{ij}^{\mathrm{mod}} = \frac{\nu_{\mathrm{t}}}{k}\widetilde{S}_{ij}^{\mathrm{f}} \tag{9}$$

is employed to model the Reynolds stresses in terms of average quantities, so that

$$\widetilde{R}_{ij}^{\mathrm{f,mod}} = 2\nu_{\mathrm{t}}\widetilde{S}_{ij}^{\mathrm{f}} - 2/3\,k\delta_{ij}. \tag{10}$$

The deviatoric part of mean fluid strain rate tensor, $\widetilde{S}_{ij}^{\mathrm{f}}$, is

$$\widetilde{S}_{ij}^{\mathrm{f}} = \frac{1}{2}\left(\frac{\partial\tilde{u}_i^{\mathrm{f}}}{\partial x_j} + \frac{\partial\tilde{u}_j^{\mathrm{f}}}{\partial x_i}\right) - \frac{1}{3}\frac{\partial\tilde{u}_k^{\mathrm{f}}}{\partial x_k}\delta_{ij}. \tag{11}$$

The eddy viscosity ν_{t} is determined using a k-ε-turbulence model, employing $\nu_{\mathrm{t}} = C_\mu k^2/\varepsilon$, with $C_\mu = 0.09$.

The transport equation for the TKE reads (Hsu et al. 2003)

$$\frac{\partial k}{\partial t} + \tilde{u}_j^{\mathrm{f}}\frac{\partial k}{\partial x_j} = \underbrace{\widetilde{R}_{ij}^{\mathrm{f}}\frac{\partial\tilde{u}_i^{\mathrm{f}}}{\partial x_j}}_{P} - \underbrace{\overline{2\chi^{\mathrm{f}}\nu^{\mathrm{f}}S_{ij}^{\mathrm{f}\prime}\frac{\partial u_i^{\mathrm{f}\prime}}{\partial x_j}}}_{\varepsilon}$$

$$+ \underbrace{\frac{\partial}{\partial x_j}\left(\overline{2\chi^{\mathrm{f}}\nu^{\mathrm{f}}S_{ij}^{\mathrm{f}\prime}u_i^{\mathrm{f}\prime}} - \overline{\frac{1}{2}\chi^{\mathrm{f}}u_i^{\mathrm{f}\prime}u_i^{\mathrm{f}\prime}u_j^{\mathrm{f}\prime}} - \overline{\frac{1}{\rho^{\mathrm{f}}}\chi^{\mathrm{f}}u_j^{\mathrm{f}\prime}p^{\mathrm{f}\prime}}\right)}_{T} \tag{12}$$

$$+ \underbrace{\overline{\frac{p^{\mathrm{f}\prime}}{\rho^{\mathrm{f}}}\frac{\partial\chi^{\mathrm{f}}u_i^{\mathrm{f}\prime}}{\partial x_i}}}_{N} - \underbrace{\frac{\beta}{\rho^{\mathrm{f}}}\overline{\chi^{\mathrm{s}}u_i^{\mathrm{f}\prime}\left(u_i^{\mathrm{f}\prime} - u_i^{\mathrm{s}\prime}\right)} + \frac{\beta}{\rho^{\mathrm{f}}}\overline{\chi^{\mathrm{s}\prime}u_i^{\mathrm{f}\prime}}(\tilde{u}_i^{\mathrm{f}} - \tilde{u}_i^{\mathrm{s}})}_{I}.$$

In this equation, β is a drag parameter, ν^{f} the fluid viscosity, and the terms are already identified as production P, dissipation ε, and transport T. The last two terms take into account the effect of the sediment, and their combination is called interfacial term I. Modeling is required for the above equation. The term denoted N is neglected due to the lack of

information about the correlations. Details about the modeling procedure of the remaining terms can be found in Cheng et al. (2017), Chauchat et al. (2017). The final modeled equation reads

$$\frac{\partial k}{\partial t} + \tilde{u}_j^{\mathrm{f}} \frac{\partial k}{\partial x_j} = \underbrace{\widetilde{R}_{ij}^{\mathrm{f,mod}} \frac{\partial \tilde{u}_i^{\mathrm{f}}}{\partial x_j}}_{P^{\mathrm{mod}}} - \varepsilon + \underbrace{\frac{\partial}{\partial x_j}\left[\left(\nu^{\mathrm{f}} + \frac{\nu_{\mathrm{t}}}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right]}_{T^{\mathrm{mod}}}$$
$$\underbrace{-\frac{2}{\rho^{\mathrm{f}}}\alpha\beta k\left(1 - t_{\mathrm{mf}}\right)}_{I_1^{\mathrm{mod}}} \underbrace{-\frac{1}{\phi}\frac{\nu_{\mathrm{t}}}{\sigma_c}\frac{\partial \alpha}{\partial x_i}\left(\frac{\rho^{\mathrm{s}}}{\rho^{\mathrm{f}}} - 1\right)f_i}_{I_2^{\mathrm{mod}}}.$$

(13)

The dissipation ε is obtained from its own transport equation. It is analogous to Eq. (13) and reads

$$\frac{\partial \varepsilon}{\partial t} + \tilde{u}_j^{\mathrm{f}} \frac{\partial \varepsilon}{\partial x_j} = C_{1\varepsilon}\frac{\varepsilon}{k}P^{\mathrm{mod}} - C_{2\varepsilon}\frac{\varepsilon^2}{k} + \frac{\partial}{\partial x_j}\left[\left(\nu^{\mathrm{f}} + \frac{\nu_{\mathrm{t}}}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right]$$
$$+ C_{3\varepsilon}\frac{\varepsilon}{k}I_1^{\mathrm{mod}} + C_{4\varepsilon}\frac{\varepsilon}{k}I_2^{\mathrm{mod}}.$$

(14)

The turbulent drag parameter $t_{\mathrm{mf}} = \exp(-B\,St)$ is quantified by the Stokes number St (Benavides and van Wachem 2008) and an empirical coefficient B, which characterizes the degree of correlation between fluctuations in particle and fluid velocity. The values used in the present work for the closure coefficients in the above equations are listed in Table 1.

For the modeling of the interfacial term I, the literature generally assumes that the presence of the sediment dissipates turbulence (Drew 1976; Hsu et al. 2003, 2004; Yu et al. 2010; Cheng et al. 2017; Chauchat et al. 2017).

3 Setup

3.1 DNS

This work is based on a DNS study of particle-laden flows described in Jain et al. (2021) and briefly summarised here. The unsteady three-dimensional Navier–Stokes equations for incompressible fluids were discretized on a staggered cartesian grid with a second-order finite volume scheme (Jain et al. 2021) and the equations of motion for the particles coupled to the fluid by a continuous forcing immersed boundary method proposed in Tschisgale et al. (2018).

Table 1 Parameters for the k-ε model according to Chauchat et al. (2017), with the first value of $C_{4\varepsilon}$ for a stable, the second for an unstable situation

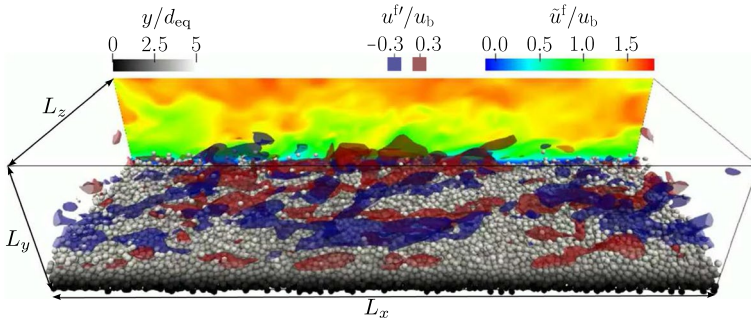| Parameter	C_μ	$C_{1\varepsilon}$	$C_{2\varepsilon}$	$C_{3\varepsilon}$	$C_{4\varepsilon}$	σ_c	σ_k	σ_ε	B
Value	0.09	1.44	1.92	1.2	0 or 1	1.0	1.0	1.3	0.25

Fig. 1 Instantaneous snapshot of the simulation of Jain et al. (2021) with spherical particles, flow from left to right. The back side of the domain shows a two-dimensional contour plot of the streamwise fluid velocity. The mobile particles are colored according to their wall-normal position y. Three-dimensional isosurfaces for the instantaneous streamwise fluid velocity fluctuations are plotted for $u^{f\prime}/u_b = 0.3$ in red and for $u^{f\prime}/u_b = -0.3$ in blue

Turbulent open-channel flow simulations with mobile particles transported in bed-load mode over a fixed rough wall were reported in that reference. Four simulations with different particle shapes were performed, with all other numerical and physical parameters identical, so that the effect of the particle shape on the fluid flow and the sediment bed could be investigated. In the present study, two of these cases are considered, the one with spherical particles and the one with general ellipsoidal particles. The latter have half-axes $a : b : c = 1 : 0.67 : 0.44$ resulting in a Krumbein-sphericity of 0.66 and a nominal roughness of $k_s^+ = 85$. A snapshot for the case with spherical particles is shown in Fig. 1. The computational domain was discretized with an isotropic spatial resolution of $\delta_{DNS} = d_{eq}/18$ and a size of $L_x \times L_y \times L_z = 108d_{eq} \times 23d_{eq} \times 36d_{eq}$, where L_x, L_y and L_z is the channel length in the streamwise, wall-normal and spanwise direction, respectively, and d_{eq} the volumetrically equivalent diameter of the particles. The boundary conditions imposed were periodic in the streamwise and spanwise direction, a rigid-lid condition at the top, which is equal to a free slip condition, and a no-slip condition at the bottom wall. Fixed spherical particles with a uniform diameter d_{eq} were irregularly arranged close to the bottom wall as described in Jain et al. (2017). The relative density of the particles with respect to the fluid $\rho^s/\rho^f - 1 = 1.55$ corresponds to quartz sand in water (Jain et al. 2021). The time step Δt was adjusted so that $CFL = u_{max}\Delta t/\delta_{DNS} = 0.6$. A given flow rate Q^f was maintained by adjusting the volume force $f_x(t)$ that drives the flow. The bulk velocity is $u_b = Q^f/H$, with the submergence height $H = L_y - H^s$, and H^s the mean height of the sediment bed. The latter is defined as the elevation where $\langle \phi \rangle = 0.9$ and assumes values of $5 \ldots 6d_{eq}$ (Jain et al. 2021).

Throughout, the operator $\langle \cdot \rangle$ denotes the averaging procedure to obtain wall-normal profiles defined as

$$\langle \psi \rangle(y) = \frac{1}{T_{av}} \frac{1}{L_x L_z} \int_{T_{in}}^{T_{in}+T_{av}} \int_0^{L_x} \int_0^{L_z} \psi(x, y, z, t)\, dz\, dx\, dt, \tag{15}$$

where T_{in} is the time when initialization is terminated, and T_{av} the duration of averaging. The bulk Reynolds number

$$Re_b = u_b H / \nu^f = Q^f / \nu^f = 3432 \tag{16}$$

was imposed in all cases. The particle density selected yields a Galilei number of

$$Ga = \frac{1}{\nu^f} \left[\left(\frac{\rho^s}{\rho^f} - 1 \right) g_y d_{eq}^3 \right]^{1/2} = 44.7. \tag{17}$$

The total shear stress τ_{tot} was calculated following an approach proposed by Nikora et al. (2007)

$$\tau_{tot}(y) = \rho^f \langle f_x \rangle \int_y^{L_y} \langle \phi \rangle (y) \, dy = \rho^f \nu^f \frac{\partial \langle u^f \rangle}{\partial y} - \rho^f \langle u^{f\prime} v^{f\prime} \rangle + \tau^{fs}, \tag{18}$$

which balances the total shear stress with the driving volume force f_x. The total shear stress consists of a viscous part and a turbulent part, which are evaluated from their explicit functional expressions. The third contribution, τ^{fs}, represents effects of fluid-sediment interaction and was obtained from the above balance equation.

The friction velocity u_τ is computed from the total shear stress τ_{tot} at an elevation $y = H^s$. With this, the friction Reynolds number is $Re_\tau = u_\tau H / \nu^f$. The Shields number $Sh = \tau_{tot}(H^s)/(\rho^s - \rho^f) g_y d_{eq}$ is used to assess the mobility of the particles.

As a starting point, the wall-normal profiles of the RHS terms in the transport equation for the turbulent kinetic energy (12) were computed using the DNS data. The methodology employed is based on the approach proposed in Vreman and Kuerten (2018), for which no fluctuation field is explicitly necessary. However, as discussed in the same reference, evaluating the interfacial term I poses difficulties in this formulation because it requires the evaluation of gradients at the interface between the phases in the Cartesian grid. The interfacial term could be evaluated as in Santarelli et al. (2016), but this is cumbersome in regions of high sediment concentration. As a remedy, the developed flow is considered, such that $Dk/Dt = 0$. As a consequence, I can be determined from the balance equation

$$\frac{Dk}{Dt} = 0 = P - \varepsilon + T + I. \tag{19}$$

3.2 RANS

The RANS simulations were carried out in two-dimensional manner. This was decided because the spanwise variations in the DNS were much less pronounced than the streamwise variations and to focus on the formation of dune-like structures as a hint to realistic transition between flat bed and undulating bedform. The density ratio, bulk Reynolds number Re_b, and Galilei number Ga were chosen to be the same as in the DNS. The flow was controlled in a similar way as in the DNS by adjusting the volume force to obtain the desired bulk Reynolds number. The computational domain used for these simulations is of size $L_x \times L_y \times L_z = 108 d_{eq} \times 23 d_{eq} \times 2 d_{eq}$ and was discretized with a step-size $\delta_{RANS} = d_{eq}$ and a local refinement in wall-normal direction within the bed region

$$\delta_{RANS,y} = \begin{cases} d_{eq}/5, & 3 \leq y/d_{eq} \leq 13, \\ d_{eq}, & \text{otherwise.} \end{cases} \tag{20}$$

The jump in step size is uncritical as all components of the RANS solution are linear or even constant at $y/d_{eq} = 3$ and $y/d_{eq} = 13$. The finally used spatial grid contains $108 \times 63 \times 2$ grid points. The time step was adjusted to maintain $CFL \leq 0.3$, throughout.

The RANS simulations were initialized with zero velocity for fluid and sediment and the wall-normal profile of the sediment volume fraction set to

$$\alpha(y) = \begin{cases} 0.58, & y/d_{eq} < 2, \\ \frac{0.58}{2}\left(1 + \tanh\left(10\,\frac{y/d_{eq}-5}{2-y/d_{eq}}\right)\right), & \text{otherwise.} \end{cases} \tag{21}$$

This was adapted from Bonamy et al. (2021) and devised such that the overall fluid volume fraction is similar to the DNS, with a deviation below 2%.

For the k-ε turbulence model constants, the default parameters were used as defined in Chauchat et al. (2017) and equal to the parameters in the clear fluid model. The additional parameter $C_{3\varepsilon}$ was calibrated with numerical simulations in that reference and $C_{4\varepsilon}$ set to 0 for stably stratified situations (Chauchat et al. 2017). All parameters are assembled in Table 1. The particle phase stress was modeled with the approach described in Chauchat et al. (2017), with the particle normal stress based on the $\mu(I)$-rheology (i.e., dense granular flow rheology) in the grain inertia regime. This is justified based on the Stokes number $St = d\sqrt{\rho^s p^a}/(\rho^f \nu^s)$, with p^a the particle shear stress, which reaches values up to 50 in the region close to H^s. The particle Reynolds number $Re_P = |\mathbf{u}^f - \mathbf{u}^s|d/\nu^f$ assumes values below 25. The parameters employed for the $\mu(I)$-rheology and the reference from which the values are taken are listed in Table 2.

4 Results with unmodified RANS

Figure 2 shows the results obtained from the DNS and the RANS simulations of two cases, one with spherical the other with ellipsoidal particles. The latter are presented, since they exhibit significant differences in the behavior of the sediment bed. The wall-normal profiles of the streamwise velocity of the fluid, $\tilde{u}^f$, the sediment velocity, $\tilde{u}^s$, the fluid volume fraction ϕ, the turbulent kinetic energy of the fluid, k, and the total shear stress, τ_{tot}, are presented. Additionally, the three contributions to the total shear stress are shown, the viscous, the turbulent, and the fluid-solid interaction contribution, according to Eq. (18). The markers indicate the mean height of the sediment bed H^s in the respective cases. It differs as a result of different particle arrangement and void fraction distribution.

For the spherical particles, the velocities of fluid and sediment agree fairly well between the DNS and the RANS. Note that hardly any particle reaches $y > 8d_{eq}$, so that the curves for the sediment velocity are terminated where the particle fraction is too low, here below 0.1 %. The fluid volume fraction ϕ differs somewhat inside the bed. While in the DNS a slightly looser packing and an impact of particle layering is noticed, ϕ is constant and minimal over the bottom of the simulation, then transiting monotonously to the free-stream value of 1. The rough wall at the bottom of the DNS, which was constructed by fixed particles, causes a high sediment concentration in that region, which is not shown in the plot. Combined with the small deviation in the prescribed initial sediment volume profile, this leads to a shift between the sediment bed height in the RANS with respect to the DNS of approximately $0.5d_{eq}$. This does not have a significant influence on the following observations since it is negligible compared to the submergence.

Table 2 Rheology parameters employed for $\mu(I)$-rheology, with μ_s the static friction coefficient, μ_2 an empirical dynamical coefficient, I_0 an empirical constant of the granular rheology, B_ϕ a parameter in the dilatancy law, and Ψ the shape factor of the particles (Chauchat et al. 2017)

Parameter	μ_s	μ_2	I_0	B_ϕ	Ψ
Spheres	0.16	0.70	0.3	0.31	1
Reference	Jain et al. (2021)	Revil-Baudard and Chauchat (2013)	Revil-Baudard and Chauchat (2013)	Maurin et al. (2016)	
Ellipsoids	0.38	0.82	0.6	0.66	0.66
Reference	Chauchat et al. (2017)	Chauchat et al. (2017)	Chauchat et al. (2017)	Chauchat et al. (2017)	Jain et al. (2021)

While the velocities in the RANS are slightly below the ones of the DNS, the fluctuations are larger by a substantial amount, as seen from k and the turbulent shear stress. For the RANS results, $\langle u^{f\prime} v^{f\prime} \rangle$ was determined from the Boussinesq assumption. Compared to the other two contributions to the total shear stress, the viscous part (f) is negligible. The turbulent contribution (g) dominates and exhibits a deviation of about 60 % with respect to the DNS, and the part due to the fluid-sediment interaction is overestimated inside the bed, while underestimated above H^s. Note that if the profiles for the shear stress were scaled with the friction velocity u_τ, they would not differ much from each other since u_τ is defined based on the total shear stress at the elevation $y = H^s$.

With the ellipsoidal particles in the DNS, moving dunes and troughs form which lead to non-zero fluid and sediment velocity in the bed region and a much smoother distribution of the sediment velocity. These dunes are not present in the RANS simulations, so that this case is even harder for the RANS. The sediment is immobile inside the bed, with the fluid fraction minimal and constant. In contrast to the case with spheres, the TKE from the RANS simulation is smaller than the DNS values here (Fig. 2d). The shear stresses for the ellipsoidal case, on the other hand, behave similarly as for the spheres, just on a somewhat higher level, with similar amounts of deviations between DNS and RANS.

All in all, the RANS results cannot be said to capture the case with the spherical particles satisfactorily, nor the case with ellipsoidal particles.

5 Budget terms from the DNS data

To improve the accuracy of the RANS simulations, the approach in this study is to enhance the modeling of fluid turbulence. In this way, we aim to revise the shear stresses, especially in the region of the sediment bed height and by this measure to correct the prediction of the behavior of both phases. In the following, this study focuses on the case with spherical particles, since this is the case with the most DNS data available.

The DNS data can be used to evaluate the RHS terms in both versions of the k-transport equation (exact Eq. (12) and modeled Eq. (13)) to quantify the errors introduced by the modeling procedure. The resulting wall-normal profiles of the TKE budget terms are shown in Fig. 3a. The deficiency of the ε-model is more complex to assess, so that the same will not be done for the modeling of the ε-equation.

The RHS terms of the TKE budget have to add up to zero for a fully developed channel flow, according to Eq. (19). Indeed, this has to hold for the DNS and is used to obtain
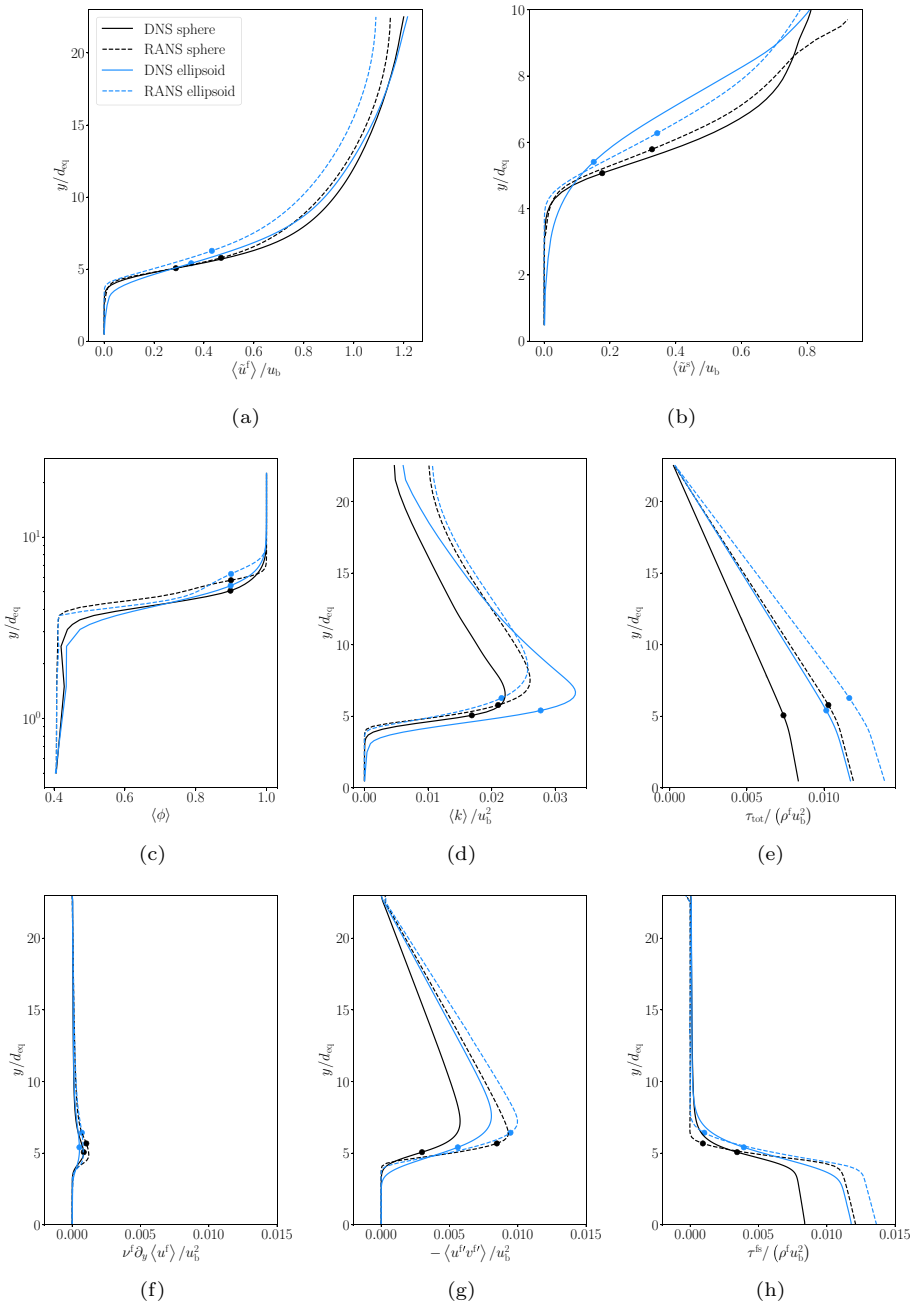
Fig. 2 Wall-normal profiles of various quantities obtained in Jain et al. (2021) (DNS) for the case with spheres and ellipsoids, together with the corresponding results of the RANS simulations using the unmodified SedFoam-2.0 solver. Note the logarithmic y-axis in (c) and the smaller vertical range in (b) to focus on regions of particle presence, $\langle \phi \rangle < 0.999$. The markers indicate the mean height of the sediment bed H^s
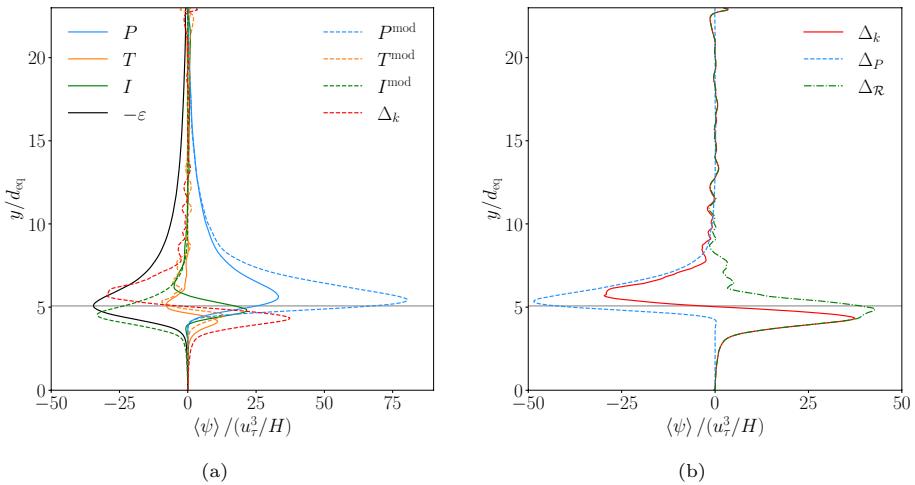
Fig. 3 Wall-normal profiles of RHS of k-transport equation using the DNS data. **a** Exact terms calculated from Eq. (12) and modeled terms according to Eq. (13). **b** Residual term Δ_k, contribution caused by the error in production Δ_P and remaining deviation $\Delta_{\mathcal{R}}$. The horizontal line in both graphs represents the height of the sediment bed H^s in the DNS

I which could not be obtained directly. These terms are shown in Fig. 3a by continuous lines. The graph suggests that the production and interfacial term play a strong role in the TKE-transport.

The DNS data can also be inserted into the model expressions of the respective terms, Eq. (13). The results for the respective contributions are shown as well in Fig. 3a by broken lines. Since these terms are models, hence only approximations, it cannot be expected that the balance Eq. (13) is satisfied. The residual term is named Δ_k and is calculated from

$$0 = P^{\mathrm{mod}} - \varepsilon + T^{\mathrm{mod}} + I^{\mathrm{mod}} + \Delta_k, \qquad (22)$$

where the dissipation ε is taken directly from the DNS data without using the model formulation. The profile of Δ_k is also shown in Fig. 3a. The graph reveals that Δ_k mainly results from an inaccurate modeling of the production P and the interfacial term I, even with some error compensation. The modeled production P^{mod} is largely overestimated in the region $4 < y/d_{\mathrm{eq}} < 10$, i.e., in the region right above and below the height of the sediment bed H^s. Both interfacial terms, I and I^{mod}, act primarily right below H^s, but I^{mod} is of the wrong sign. In the DNS, the presence of the particles increases turbulent fluctuations in the region of the sediment bed height, which is not reproduced by the model expression. The assumption to model the interfacial term as turbulence attenuation is not correct in the particular flow situation this paper is focused on. The effect of the particles to increase turbulence in some situations has been reported before (Shao et al. 2012; Vowinckel et al. 2014). Consequently, a re-calibration of model coefficients (fourth category of uncertainties listed above) can not overcome the modeling error. Instead, it is necessary to revise the functional form of the model, which amounts to addressing the third category of model uncertainties. Considerable effort had been invested to derive the expressions for modelling the interfacial term (Hsu et al. 2003), and the SedFoam-2.0 solver was validated on a wide range of flow scenarios. Additionally, large deviations from the DNS target were observed

in the present study for other modelled terms. Correcting each term individually would require a significant amount of high-fidelity data to learn the corresponding terms, while the output solution will depend on nonlinear interactions of the various terms through the RANS equations. Given the uncertainties in the different terms, and in the aim of providing a first proof-of-concept, we do not attempt individual term corrections. Instead, we follow the approach of Schmelzer et al. (2020) for single-phase flows, and we introduce an additive error term in the TKE equation that accounts for all modeling inaccuracies simultaneously. Other strategies could be imagined, which are left for further research.

A closer inspection of the modeling errors results in the following observations. The residual in the modeled k-transport equation Δ_k can be split into a part that originates in the error of the production model

$$\Delta_P = P - P^{\mathrm{mod}} = 2k\left(b_{ij} - b_{ij}^{\mathrm{mod}}\right)\frac{\partial \tilde{u}_i^{\mathrm{f}}}{\partial x_j} \tag{23}$$

and the remaining errors

$$\Delta_{\mathcal{R}} = \Delta_k - \Delta_P, \tag{24}$$

which are mainly caused by the erroneous modeling of the interfacial term. The wall-normal profiles of these three quantities are shown in Fig. 3b.

The error in the production model, Δ_P, results from erroneous modeling of the Reynolds stress tensor, in particular from the representation of b_{ij}, since this is the only difference between the exact term and the model formulation for the production. The error in b_{ij}^{mod} can be extracted from the DNS data by

$$\Delta_b = b_{ij} - b_{ij}^{\mathrm{mod}} \tag{25}$$

and will be a point of attack in the subsequent strategy for improvement.

6 Employing SpaRTA for model improvement

The SpaRTA approach (Schmelzer et al. 2020) was initially developed for correcting standard RANS models to improve predictions of flow separation and reattachment in separated incompressible single-phase flows. The idea is to add corrective terms to the model expression of the Reynolds stress tensor and the RHS of the modeled turbulence transport equations. Ideally, these terms would be equal to the errors Δ for any situation and, thus, compensate the errors made during the modeling process. As this is impossible, an ansatz $M \approx \Delta$ is made and optimized by ML.

In Schmelzer et al. (2020), corrections were applied separately to the Reynolds stress tensor and the turbulence transport equations. A counterpart of the corrective term for the k-transport equation was also included in the ε-transport equation. Additionally, the combination of both corrections was investigated. However, it was found that the simultaneous application of both corrections does not yield additional improvement compared to using only a single correction. Hence, the Reynolds stress and turbulence transport equations are only corrected separately in this work. Only the effects of Δ_b and Δ_k are considered, and it is refrained from investigating $\Delta_{\mathcal{R}}$ since this term does not consider the error of the model production and is only valuable when combined with a correction for the production.

It is assumed that the corrective terms depend on the mean strain-rate tensor and the mean rotation-rate tensor reading

$$\widetilde{S}_{ij} = \frac{1}{2}\left(\frac{\partial \tilde{u}_i^{\mathrm{f}}}{\partial x_j} + \frac{\partial \tilde{u}_j^{\mathrm{f}}}{\partial x_i} \right) - \frac{1}{3}\frac{\partial \tilde{u}_k^{\mathrm{f}}}{\partial x_k}\delta_{ij} \ , \quad \widetilde{\Omega}_{ij} = \frac{1}{2}\left(\frac{\partial \tilde{u}_i^{\mathrm{f}}}{\partial x_j} - \frac{\partial \tilde{u}_j^{\mathrm{f}}}{\partial x_i} \right), \tag{26}$$

respectively, normalized by multiplication with the time scale $\tau = C_\mu k/\varepsilon$.

In the following, quantities denoted by M_i are models targeting the term i, which will be optimized by ML. To represent the tensor Δ_b, a model M_{Δ_b} is built from the invariants $\mathcal{I}_m$ and base tensors $\mathcal{T}_{ij}^{(n)}$ that form the minimal integrity basis (Spencer 1971) of the normalized tensors $\tau\widetilde{S}_{ij}$ and $\tau\widetilde{\Omega}_{ij}$.

Only two invariants and three base tensors are needed in two-dimensional flows (Gatski and Speziale 1993; Pope 1975)

$$\begin{aligned}
\mathcal{I}_1 &= \tau^2 \widetilde{S}_{mn}\widetilde{S}_{nm}, \quad \mathcal{I}_2 = \tau^2 \widetilde{\Omega}_{mn}\widetilde{\Omega}_{nm}, \\
\mathcal{T}_{ij}^{(1)} &= \tau\widetilde{S}_{ij}, \qquad \mathcal{T}_{ij}^{(2)} = \tau^2\left(\widetilde{S}_{ik}\widetilde{\Omega}_{kj} - \widetilde{\Omega}_{ik}\widetilde{S}_{kj} \right), \\
\mathcal{T}_{ij}^{(3)} &= \tau^2\left(\widetilde{S}_{ik}\widetilde{S}_{kj} - \frac{1}{3}\delta_{ij}\widetilde{S}_{mn}\widetilde{S}_{nm} \right).
\end{aligned} \tag{27}$$

The model approximating Δ_b can then be written as

$$\Delta_b \approx M_{\Delta_b}\left(\tau\widetilde{S}_{ij}, \tau\widetilde{\Omega}_{ij} \right) = \sum_{n=1}^{3} f^{(n)}\left(\mathcal{I}_1, \mathcal{I}_2 \right) \mathcal{T}_{ij}^{(n)} \quad . \tag{28}$$

This approach, combined with the set (27), yields a modeling ansatz that only requires functional expressions for the scalars $f^{(n)}$.

To investigate the relevance of sediment-related quantities when correcting the turbulence transport equations, the ML process for models of Δ_k is employed in two variants. One accounts for the sediment volume fraction α, the other does not.

Two different modeling approaches were employed: The tensorial approach, called *turbulent production analogy*, represents Δ_k in a similar way as the turbulent production (Schmelzer et al. 2020)

$$\Delta_k = 2k\left(b_{ij}^k - \frac{1}{3}\,\delta_{ij} \right)\frac{\partial \tilde{u}_i^{\mathrm{f}}}{\partial x_j}, \tag{29}$$

where a model approximating the tensor b_{ij}^k has to be found. This approach is analogous to Eq. (28), but with the option to include α in the functional expressions $f^{(n)}$, so that it reads

$$b_{ij}^k \approx M_{b_{ij}^k}\left(\tau\widetilde{S}_{ij}, \tau\widetilde{\Omega}_{ij}, \alpha \right) = \sum_{n=1}^{3} f^{(n)}\left(\mathcal{I}_1, \mathcal{I}_2, \alpha \right) \mathcal{T}_{ij}^{(n)}. \tag{30}$$

The second ansatz is a newly introduced *dissipation analogy*, simply reading

$$\Delta_k = c\,\varepsilon. \tag{31}$$

The model for the scalar c is assembled by a similar approach as for b_{ij}^k, but the base tensors are not needed, so that its functional form reads

Table 3 Summary of the model errors which serve as target data for the ML process and the method to evaluate them, the modeling ansatzes, which term has to be modeled and how these models are formulated

Error term obtained by DNS		Description	Ansatz	Term to be modeled	Name of model term	Model formulation
Δ_b	(25)	Model error of b_{ij}^{mod} leading to an error in P^{mod} and $\widetilde{R}_{ij}^{\mathrm{f}}$	(28)	Δ_b	M_{Δ_b}	(28)
Δ_k	(22)	Combined model error of all terms in the k-transport equation	(29)	b_{ij}^k	$M_{b_{ij}^k}$	(30), α is optional
			(31)	c	M_c	(32), α is optional

$$c \approx M_c\left(\tau\widetilde{S}_{ij}, \tau\tilde{\Omega}_{ij}, \alpha\right) = f\left(\mathcal{I}_1, \mathcal{I}_2, \alpha\right). \tag{32}$$

Incorporating α in the model is optional.

With this scalar approach, including new quantities, e.g., scalars like α, is easy. It would also allow the inclusion of additional vectors (e.g., the gradient of α) or tensors. This increases the number of invariants, but since it is a scalar approach it is not necessary to consider the resulting additional base tensors of the minimal integrity basis. However, this also requires investigation of appropriate normalization to ensure extrapolative capabilities of the learned models (Wu et al. 2018) and is beyond the scope of the present paper.

The above Eqs. (29)–(32) combined with the set (27) yield two different modeling ansatzes for Δ_k, that only require functional expressions for the scalars $f^{(n)}$ or f, respectively.

Table 3 gives an overview of the error terms that are the target data for the ML process and from which equation they are obtained. Furthermore, the respective formulation of the modeling ansatzes and which term in the ansatz has to be optimized, are indicated. Finally, the symbol denoting the corresponding model expression and how these expressions are formulated is given.

In each approach the required single or multiple scalar, $f^{(n)}$ or f, is written as a linear combination of monomials of their input variables (i.e., $\mathcal{I}_1$, $\mathcal{I}_2$ and optionally α) up to degree six (Schmelzer et al. 2020). The respective coefficients for the monomials are optimized by ML. The methodology of selecting relevant candidate terms and inferring appropriate model coefficients is employed as described in Schmelzer et al. (2020). First, elastic net regularization (Zou and Hastie 2005) is used to identify relevant terms and to promote sparsity of the models since this method only allows few non-zero model coefficients. In the second step, the model coefficients are optimized using ridge regression. This enforces small coefficients and, thus, yields more stable models and prevents overfitting (Brunton and Kutz 2019).

7 Model variants and assessment

The result of the ML process are multiple model expressions for each model approach, with a wide variety in the number of active model terms. Exemplarily, a set of ten learned models $M_{\Delta_b}^i$ that regress Δ_b using the approach in Eq. (28) are shown in Fig. 4. The upper
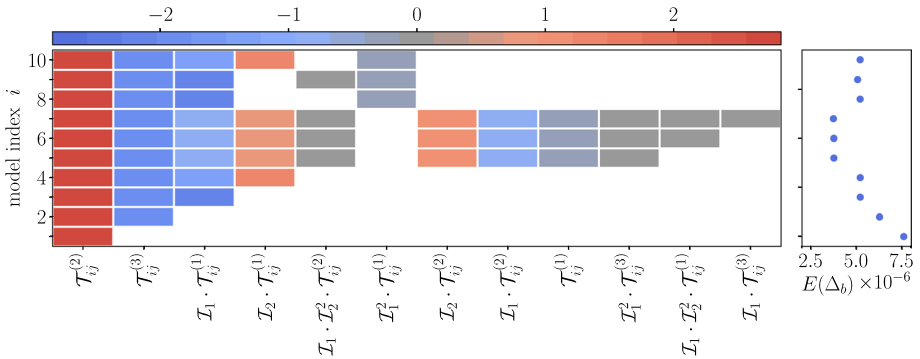
Fig. 4 Model structure of ten exemplary models $M_{\Delta_b}^i$ obtained with the present procedure. **a** Matrix indicating the values of the active (colored) candidate functions (x-axis) for each model $M_{\Delta_b}^i$ with model index i (y-axis). The color scale corresponds to the magnitude of the respective coefficient of the candidate function. A field is left blank if the coefficient is zero. **b** Mean-squared error E on the training data Δ_b for each of the ten models

index i is introduced to distinguish between the different model expressions. Each row represents a model structure and the colors indicate the non-zero value of the coefficient corresponding to the selected candidate function term indicated on the x-axis. The mean-squared error E on the training data Δ_b is shown in the right section. The learned models vary in model structure (complexity) and accuracy. However, a true validation of the models can only be performed by implementing them in a CFD solver. The role of the accuracy on the training data should, therefore, not be overstated.

To assess the quality of the models that resulted from the above procedure, they were implemented in the SedFoam-2.0 solver. The simulation setup is the one described in chapter 3 above.

With this strategy, models of the three classes listed Table 3 were generated. Of the tensorial models $M_{\Delta_b}^i$ in total 58 different variants were obtained. When implemented in the CFD solver, however, most simulations did not yield stable results. Although the model variants were obtained by using training data from the DNS with spherical particles, the resulting modified turbulence models were also used in the RANS setup corresponding to the ellipsoidal particles for validation. If the simulations were not stable, the models were discarded because they were deemed too sensitive to even small changes in rheology parameters. With this procedure, only two simulations were not discarded .

It is possible that the divergence can be avoided by decreasing the time step. On the other hand, it was reported in Chauchat et al. (2017) that a CFL number of 0.3 is generally appropriate for this type of simulation and this method, with only few cases where a smaller limit was needed. Hence, for practical reasons it was decided to stick with the value of 0.3, based on the argument that a new model should not cause a stricter limitation of the time step compared to the baseline models. Further analysis of the stability of the models is beyond the scope of this paper.

Applying the production analogy (Eqs. 29 and 30) for the Δ_k-correction without incorporating α resulted in 65 different model expressions, of which 45 yielded stable CFD simulations. When considering α during the learning process, this resulted in 148 models, but only 2 simulations did not diverge. With the dissipation analogy for Δ_k (Eqs. 31 and 32),

Table 4 Friction Reynolds number and Shields number for the results of the DNS, the RANS simulations using the unmodified turbulence model, and the RANS simulation in which the turbulence model is modified by adding the term in Eq. (34)

Case	Re_τ	Sh
DNS	295	0.13
Unmodified RANS	349	0.21
Modified RANS	325	0.17

112 models were obtained without α incorporated, and 246 models when α was considered. Only 9 and 3 simulations with these models did not diverge, respectively.

The results of the simulations with the modified solver are now compared to the results of the unmodified SedFoam-2.0 solver and evaluated using the DNS results as reference. The reference data with the unmodified RANS solver will be denoted with an upper index 0, those with one of the new models with the upper index M.

The main objective of this work is to improve the modeling of the Reynolds stress tensor and the transport equation for the turbulent kinetic energy, but the other quantities should not be deteriorated. Thus, validation is performed on a set of five quantities: the turbulent kinetic energy k, the streamwise velocity of the fluid $\tilde{u}^f$ and of the sediment $\tilde{u}^s$, the total shear stress τ_{tot} and the fluid volume fraction ϕ. The mean-squared error E on a flow quantity is evaluated with respect to the DNS data. With this, the performance Π of a model M with respect to a flow quantity ψ is assessed by comparing the mean-squared error of the results using the unmodified solver, $E(\psi^0)$, to the error of the results with the model implemented, $E(\psi^M)$ according to

$$\Pi_\psi = \frac{E(\psi^0) - E(\psi^M)}{E(\psi^0)} \quad . \tag{33}$$

A performance $\Pi_\psi < 0$ indicates deterioration of the quantity ψ, while $\Pi_\psi = 1$ signifies that the results of the modified solver perfectly match the DNS results. The average performance Π_{av} is the arithmetic mean of the single performances applied to the five quantities mentioned before.

The average performances of the stable models for Δ_b and Δ_k are shown in Fig. 5a. In Fig. 5b the individual performances of the model with the best performance on k, Π_k, are shown for each modeling approach. For Δ_k, the results are reported for those models using the production analogy (squares) and the dissipation analogy (circles), and it is indicated if the model is learned including α (filled) or without α (unfilled).

The Δ_b-correction does not perform well, but rather deteriorates the performance, as revealed by an average performance below zero. Only the models with index 1 and 2 shown in Fig. 4 yielded stable results when implemented in the CFD solver. In these models, only the base tensors $\mathcal{T}_{ij}^{(2)}$ and $\mathcal{T}_{ij}^{(3)}$ are active. In a channel flow, the shear component $\partial\tilde{u}^f/\partial y$ largely dominates over the other components. Thus, even if the normal stresses are wrong they are small compared to the shear stress. On the other hand, it can be shown analytically that the two selected tensors only act on the normal stresses and do not correct the shear stress, which remains as given by the Boussinesq term. Additionally, the models do not contribute to the turbulent production, which can also be proven analytically. Therefore, the models do not alter the turbulent kinetic energy k. As a consequence, the correction does not affect the most important contributor in the Reynolds stress tensor, i.e., the shear stress.
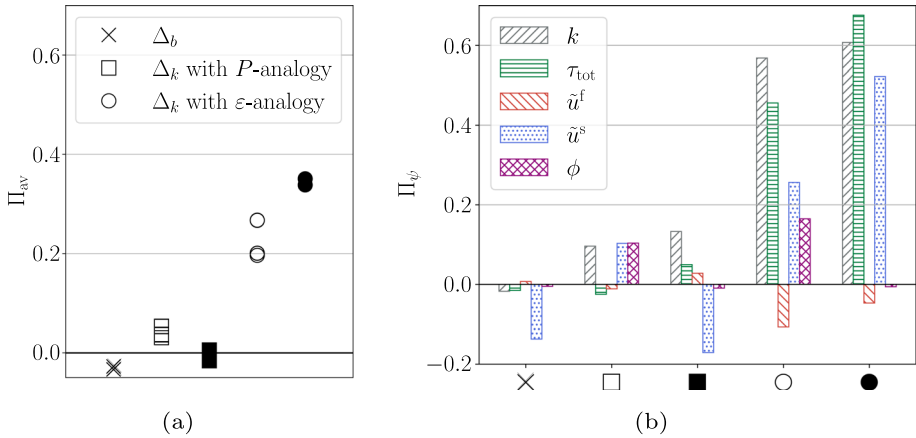
Fig. 5 Performance of the discovered models for Δ_b and Δ_k. **a** Average performance Π_{av} for all discovered models. **b** Performance Π_ψ on the single flow values ψ of the model with the highest Π_k. The modeling approach from which the model is obtained is indicated by the marker at the x-axis. The form of the markers indicates on which modeling analogy the models for Δ_k are based. Unfilled: without considering α, filled: including α

The Δ_k-correction, in contrast, directly acts on k, i.e. on the eddy viscosity ν_t and, therefore, on the shear stress. Figure 5a shows that when using dissipation analogy improvement is obtained when α is disregarded but absent when additionally considering α. Figure 5b reveals that this is mainly due to the negative effect on the sediment velocity the model introduces (only the variant performing best in Fig. 5a is analyzed in part b).

The newly introduced dissipation analogy for Δ_k, on the other hand, performs much better than the production analogy. Additionally including the phase indicator α improves the results further and yields an average performance, i.e. an increase of model quality, of up to 35 %. In this case, it is demonstrated that providing sediment-related quantities is important for the performance of a model. Although not all quantities can be improved, particularly the turbulent fluctuations are better captured as shown by the right-most part of Fig. 5b (filled circle). The error on k is reduced by ≈ 60 %.

8 Final model

It is now interesting to closer look at the best performing model for Δ_k, which is based on the dissipation analogy and includes α. The model expression reads

$$\Delta_k \approx M_c\, \varepsilon = (8.15\, \mathcal{I}_2 + 5.14\, \alpha)\, \varepsilon \quad . \tag{34}$$

The wall-normal profiles of $M_c\, \varepsilon$ and its contributions are shown in Fig. 6 juxtaposed to the wall-normal profile of the targeted term Δ_k. It is instructive to see how both parts of the sum contribute differently to the final model. In fact, the individual terms in the sum constituting $M_c\, \varepsilon$ resemble the profiles shown in Fig. 3a, where Δ_k was split into the part that originates from the error in the production model and the error in the model for

Fig. 6 Wall-normal profile of the target Δ_k, the learned approximation $\Delta_k \approx M_c\,\varepsilon$, and its individual contributions according to Eq. (34). All profiles were evaluated using the DNS data. The horizontal line represents the height of the sediment bed H^s in the DNS
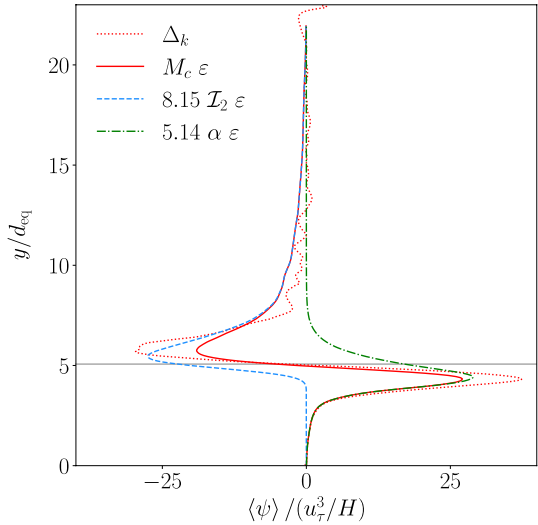
the interfacial term. Without explicitly providing this information, the ML process also identifies two contributions to Δ_k. The first term of M_c depends on $\mathcal{I}_2$ and, thus, is related only to the fluid velocity. It mainly acts close above the sediment bed height H^s and has a turbulence-damping effect. The second term employs α and is connected to fluid-sediment interaction. It increases the turbulence right beneath the height of the sediment bed.

Figure 7 now shows the RANS simulation results obtained with the model (34), labeled M. For comparison, the results of the unmodified RANS method and the DNS results are shown as well. The streamwise fluid velocity shows very small deterioration near the sediment bed height and very small improvement further up compared to the standard RANS model. The sediment velocity, on the other hand, is well-captured, particularly beneath the height of the sediment bed. The volume fraction ϕ is almost unchanged. Note that the model employed here is the correction model that corresponds to the filled circle in Fig. 5b, which provides the best overall improvement of the five quantities investigated. When employing a different model, e.g. the model corresponding to the hollow circle in Fig. 5b, the influence on ϕ would be larger. The turbulent kinetic energy k and the total shear stress τ_{tot} are improved but still leave room for further improvement.

In Table 4, the friction Reynolds number Re_τ and the Shields number Sh are listed for the DNS case, the unmodified RANS solution and the results of the RANS solver with model (34) implemented. The modification leads to an improvement of both quantities, which indicates that the shear stresses and the sediment transport are better captured.

9 Conclusions

The work at hand was concerned with the error introduced by the functional form of the turbulence closure in the two-phase RANS simulation of sediment transport. The aim was to augment standard closure models implemented in the open-source CFD solver SedFoam-2.0 (Chauchat et al. 2017) by inferring additive corrective algebraic terms using the data-driven ML algorithm SpaRTA (Schmelzer et al. 2020). Two additive terms were considered. A tensor Δ_b to correct the modeling of the Reynolds stress tensor and a scalar Δ_k to compensate
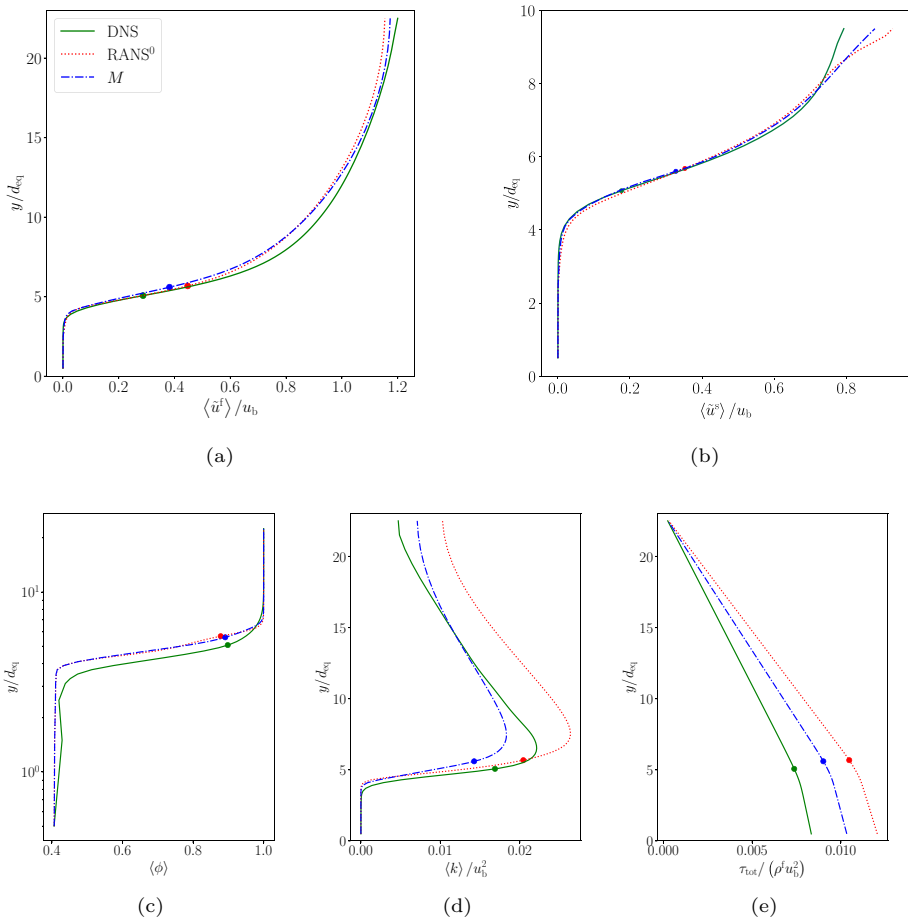
Fig. 7 Wall-normal profiles of various quantities obtained from different simulations of the sediment test case with spheres. The DNS data of Jain et al. (2021) constitute the reference. The results of the RANS simulations using the unmodified SedFoam-2.0 solver are labeled RANS[0]. The results using Eq. 34 are indicated by the label M. Note the logarithmic y-axis in (c) and the smaller vertical range in (b) to focus on regions of particle presence, $\langle\phi\rangle < 0.999$. The markers show the mean height of the sediment bed H^s for each simulation

errors of the modeled k-transport equation. Training data for the ML process were obtained from DNS of sediment transport (Jain et al. 2021). For the scalar term Δ_k, two different modeling ansatzes were used, the turbulent production analogy (Schmelzer et al. 2020), which is a tensorial approach, and a newly introduced scalar approach, called dissipation analogy. The dissipation analogy provides the possibility to easily consider new quantities (scalars, vectors, or tensors) during the model selection process without drastically increasing the number of candidate terms for the models. This is an advantage which might also be explored in other situations different from sediment transport.

To assess the validity of the models, they were introduced in the CFD solver, and new simulations were performed. The results show that the models for Δ_b did not perform well. For Δ_k, the dissipation analogy performed substantially better than the production analogy,

reducing the error on the turbulent kinetic energy k and the total shear stress τ_{tot} by up to 65 % and providing an average performance of up to 35 %.

The present work addressed the closure of the Reynolds stress model and the transport equation of k. The models presented in this paper are a correction trained for a specific class of flows. When providing training data sets for different flow situations, different model terms could be obtained, and a classifier would be needed to decide which correction to use. It would also be possible to target the individual terms in the TKE budget separately. For example, a completely new expression for the interfacial term could be learned.

As discussed before, also the closure for the sediment phase and fluid-sediment interaction is based on multiple modeling assumptions with numerous model coefficients. The values for these coefficients were either set to the default values or, if possible, derived from the DNS or experiments. A re-calibration of parameters for the present test case might further increase the accuracy of the RANS results.

The results presented in this study prove that the SpaRTA approach is applicable to multi-phase flows and demonstrate a proof-of-concept for the newly introduced dissipation analogy. The improvements, however, are not as good as for the separated flows considered in Schmelzer et al. (2020). The performance is expected to increase when additional flow quantities are considered in the models, e.g., the gradient of α or the velocity of the sediment phase. Determining the relevance of such quantities and applying the approach to the other test cases of Jain et al. (2021) is the objective of further studies.

Data Availability N/A

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical Approval N/A

Informed Consent N/A

References

Allen, J.: Principles of physical sedimentology. Springer, ??? (1985). https://doi.org/10.1007/978-94-010-9683-6

Andreotti, B., Forterre, Y., Pouliquen, O.: Granular Media: Between Fluid and Solid. Cambridge University Press, (2013)

Bakhtyar, R., Yeganeh-Bakhtiary, A., Barry, D.A., Ghaheri, A.: Two-phase hydrodynamic and sediment transport modeling of wave-generated sheet flow. Adv. Water Resour. **32**(8), 1267–1283 (2009). https://doi.org/10.1016/j.advwatres.2009.05.002

Bao, H., Feng, J., Dinh, N., Zhang, H.: Deep learning interfacial momentum closures in coarse-mesh CFD two-phase flow simulation using validation data. Int. J. Multiphase Flow **135**, 103489 (2021). https://doi.org/10.1016/j.ijmultiphaseflow.2020.103489

Beetham, S., Capecelatro, J.: Multiphase turbulence modeling using sparse regression and gene expression programming. Nuclear Technol. (2023). https://doi.org/10.1080/00295450.2023.2178251

Beetham, S., Fox, R.O., Capecelatro, J.: Sparse identification of multiphase turbulence closures for coupled fluid-particle flows. J. Fluid Mech. **914**, 11 (2021). https://doi.org/10.1017/jfm.2021.53

Benavides, A., van Wachem, B.: Numerical simulation and validation of dilute turbulent gas-particle flow with inelastic collisions and turbulence modulation. Powder Technol. **182**(2), 294–306 (2008)

Bonamy, C., Chauchat, J., Hsu, T.-J., Cheng, Z., Nagel, T., Mathieu, A., Puig-Montella, E., Chassagne, R., Higuera, P., Tsai, B., Salimi-Tarazouj, A.: Reynolds-averaged flow tutorials. 1DBedloadTurb: Turbulent bedload transport. Accessed May 01, 2023. https://sedfoam.github.io/sedfoam/tutorials_RAS.html#BedLoadTurb (2021)

Brunton, S.L., Kutz, J.N.: Data-Driven science and engineering: Mach. Learn. Dyn. Syst. Control, (2019). https://doi.org/10.1017/9781108380690

Burns, A., Frank, T., Hamill, I., Shi, J.-M.: The favre averaged drag model for turbulent dispersion in Eulerian multi-phase flows. 5th International Conference on Multiphase Flow, ICMF2004 **392** (2004)

Chauchat, J.: A comprehensive two-phase flow model for unidirectional sheet-flows. J. Hydraul. Res. **56**(1), 15–28 (2018). https://doi.org/10.1080/00221686.2017.1289260

Chauchat, J., Cheng, Z., Nagel, T., Bonamy, C., Hsu, T.-J.: Sed-Foam-2.0: a 3-D two-phase flow numerical model for sediment transport. Geosci. Model Develop. **10**(12), 4367–4392 (2017). https://doi.org/10.5194/gmd-10-4367-2017

Cheng, Z., Hsu, T.-J., Calantoni, J.: SedFoam: A multi-dimensional Eulerian two-phase model for sediment transport and its application to momentary bed failure. Coastal Engineering **119**, 32–50 (2017). https://doi.org/10.1016/j.coastaleng.2016.08.007

Da Silva, A.M.F., Yalin, M.S.: Fluvial Processes. CRC Press (2017). https://doi.org/10.4324/9781315206189

Dong, P., Zhang, K.: Two-phase flow modelling of sediment motions in oscillatory sheet flow. Coast. Eng. **36**(2), 87–109 (1999). https://doi.org/10.1016/S0378-3839(98)00052-0

Dow, E., Wang, Q.: Quantification of Structural Uncertainties in the k-ω Turbulence Model. In: 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 19th AIAA/ASME/AHS Adaptive Structures Conference 13t, p. 1762 (2011). https://doi.org/10.2514/6.2011-1762

Drew, D.A.: Production and dissipation of energy in the turbulent flow of a particle-fluid mixture, with some results on drag reduction. J. Appl. Mech. **43**(4), 543–547 (1976). https://doi.org/10.1115/1.3423926

Drew, D.A.: Mathematical modeling of two-phase flow. Annu. Rev. Fluid Mech. **15**(1), 261–291 (1983). https://doi.org/10.1146/annurev.fl.15.010183.001401

Duraisamy, K., Singh, A.P., Zhang, Z.J.: Augmentation of turbulence models using field inversion and machine learning. In: 55th AIAA Aerospace Sciences Meeting, p. 0993 (2017). https://doi.org/10.2514/6.2017-0993

Duraisamy, K., Iaccarino, G., Xiao, H.: Turbulence modeling in the age of data. Ann. Rev. Fluid Mech. **51**(1), 357–377 (2019). https://doi.org/10.1146/annurev-fluid-010518-040547

Forterre, Y., Pouliquen, O.: Flows of dense granular media. Ann. Rev. Fluid Mech. **40**(1), 1–24 (2008). https://doi.org/10.1146/annurev.fluid.40.111406.102142

Gatski, T.B., Speziale, C.G.: On explicit algebraic stress models for complex turbulent flows. J. Fluid Mech. **254**, 59–78 (1993). https://doi.org/10.1017/S0022112093002034

Goldstein, E.B., Coco, G., Plant, N.G.: A review of machine learning applications to coastal sediment transport and morphodynamics. Earth-Science Rev. **194**, 97–108 (2019). https://doi.org/10.1016/j.earscirev.2019.04.022

Hsu, T.-J., Jenkins, J.T., Liu, P.L.-F.: On two-phase sediment transport: dilute flow. J. Geophys. Res. Oceans (2003). https://doi.org/10.1029/2001JC001276

Hsu, T..-J., Jenkins, J..T., Liu, P.L..-F.: On two-phase sediment transport: sheet flow of massive particles. Proc. Royal Soc. London Ser. A Math. Phys. Eng. Sci. **460**(2048), 2223–2250 (2004). https://doi.org/10.1098/rspa.2003.1273

Jain, R., Vowinckel, B., Fröhlich, J.: Spanwise particle clusters in DNS of sediment transport over a regular and an irregular bed. Flow Turbul. Combust. **99**(3), 973–990 (2017). https://doi.org/10.1007/s10494-017-9850-x

Jain, R., Tschisgale, S., Fröhlich, J.: Impact of shape: DNS of sediment transport with non-spherical particles. J. Fluid Mech. **916**, 38 (2021). https://doi.org/10.1017/jfm.2021.214

Jenkins, J.T., Henes, D.M.: Collisional sheet flows of sediment driven by a turbulent fluid. J. Fluid Mech. **370**, 29–52 (1998). https://doi.org/10.1017/S0022112098001840

Jha, S.K., Bombardelli, F.A.: Two-phase modeling of turbulence in dilute sediment-laden, open-channel flows. Environ. Fluid Mech. **9**, 237–266 (2009). https://doi.org/10.1007/s10652-008-9118-z

Kempe, T., Vowinckel, B., Fröhlich, J.: On the relevance of collision modeling for interface-resolving simulations of sediment transport in open channel flow. Int. J. Multiphase Flows **58**, 214–235 (2014)

Kidanemariam, A.G., Uhlmann, M.: Direct numerical simulation of pattern formation in subaqueous sediment. J. Fluid Mech. **750**, 2 (2014). https://doi.org/10.1017/jfm.2014.284

Lee, C.-H., Low, Y.M., Chiew, Y.-M.: Multi-dimensional rheology-based two-phase model for sediment transport and applications to sheet flow and pipeline scour. Phys. Fluids **28**(5), 053305 (2016). https://doi.org/10.1063/1.4948987

Ling, J., Kurzawski, A., Templeton, J.: Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. J. Fluid Mech. **807**, 155–166 (2016). https://doi.org/10.1017/jfm.2016.615

Mathieu, A., Cheng, Z., Chauchat, J., Bonamy, C., Hsu, T.-J.: Numerical investigation of unsteady effects in oscillatory sheet flows. J. Fluid Mech. **943**, 7 (2022). https://doi.org/10.1017/jfm.2022.405

Maurin, R., Chauchat, J., Frey, P.: Dense granular flow rheology in turbulent bedload transport. J. Fluid Mech. (2016). https://doi.org/10.1017/jfm.2016.520

MiDi, G.: On dense granular flows. Eur. Phys. J. E **14**(4), 341–365 (2004). https://doi.org/10.1140/epje/i2003-10153-0

Nagel, T., Chauchat, J., Bonamy, C., Liu, X., Cheng, Z., Hsu, T.-J.: Three-dimensional scour simulations with a two-phase flow model. Adv. Water Resour. **138**, 103544 (2020). https://doi.org/10.1016/j.advwatres.2020.103544

Nikora, V., McEwan, I., McLean, S., Coleman, S., Pokrajac, D., Walters, R.: Double-averaging concept for rough-bed open-channel and overland flows: theoretical background. J. Hydraul. Eng. **133**(8), 873–883 (2007). https://doi.org/10.1061/(ASCE)0733-9429(2007)133:8(873)

Oliver, T.A., Moser, R.D.: Bayesian uncertainty quantification applied to RANS turbulence models. J. Phys. Conf. Ser. **318**(4), 042032 (2011). https://doi.org/10.1088/1742-6596/318/4/042032

Pope, S.B.: A more general effective-viscosity hypothesis. J. Fluid Mech. **72**(2), 331–340 (1975). https://doi.org/10.1017/S0022112075003382

Revil-Baudard, T., Chauchat, J.: A two-phase model for sheet flow regime based on dense granular flow rheology. J. Geophys. Res. Oceans **118**(2), 619–634 (2013). https://doi.org/10.1029/2012JC008306

Ribberink, J.S., Al-Salem, A.A.: Sediment transport in oscillatory boundary layers in cases of rippled beds and sheet flow. J. Geophys. Res. Oceans **99**(C6), 12707–12727 (1994). https://doi.org/10.1029/94JC00380

Saccone, D., Marchioli, C., De Marchis, M.: Effect of roughness on elongated particles in turbulent channel flow. Int. J. Multiphase Flow **152**, 104065 (2022). https://doi.org/10.1016/j.ijmultiphaseflow.2022.104065

Santarelli, C., Roussel, J., Fröhlich, J.: Budget analysis of the turbulent kinetic energy for bubbly flow in a vertical channel. Chem. Eng. Sci. **141**, 46–62 (2016). https://doi.org/10.1016/j.ces.2015.10.013

Schmelzer, M., Dwight, R.P., Cinnella, P.: Discovery of algebraic Reynolds-stress models using sparse symbolic regression. Flow Turbul. Combust. **104**(2), 579–603 (2020). https://doi.org/10.1007/s10494-019-00089-x

Shao, X., Wu, T., Yu, Z.: Fully resolved numerical simulation of particle-laden turbulent flow in a horizontal channel at a low Reynolds number. J. Fluid Mech. **693**, 319–344 (2012). https://doi.org/10.1017/jfm.2011.533

Singh, A.P., Duraisamy, K.: Using field inversion to quantify functional errors in turbulence closures. Phys. Fluids **28**(4), 045110 (2016). https://doi.org/10.1063/1.4947045

Sommerfeld, M.: Validation of a stochastic Lagrangian modelling approach for inter-particle collisions in homogeneous isotropic turbulence. Int. J. Multiphase Flow **27**(10), 1829–1858 (2001). https://doi.org/10.1016/S0301-9322(01)00035-0

Sommerfeld, M.: Numerical methods for dispersed multiphase flows. Particl. Flows (2017). https://doi.org/10.1007/978-3-319-60282-0_6

Sommerfeld, M., Qadir, Z.: Fluid dynamic forces acting on irregular shaped particles: Simulations by the lattice-Boltzmann method. Int. J. Multiphase Flow **101**, 212–222 (2018). https://doi.org/10.1016/j.ijmultiphaseflow.2018.01.016

Spencer, A.J.M.: Part III - Theory of Invariants. Mathematics (1971). https://doi.org/10.1016/B978-0-12-240801-4.50008-X

Tracey, B.D., Duraisamy, K., Alonso, J.J.: A Machine learning strategy to assist turbulence model development, 53rd AIAA Aerospace Sciences Meeting, p. 1287 (2015). https://doi.org/10.2514/6.2015-1287

Tschisgale, S., Kempe, T., Fröhlich, J.: A general implicit direct forcing immersed boundary method for rigid particles. Comput. Fluids **170**, 285–298 (2018)

Vowinckel, B.: Incorporating grain-scale processes in macroscopic sediment transport models: a review and perspectives for environmental and geophysical applications. Acta Mech. **232**(6), 2023–2050 (2021). https://doi.org/10.1007/s00707-021-02951-4

Vowinckel, B., Kempe, T., Fröhlich, J.: Fluid-particle interaction in turbulent open channel flow with fully-resolved mobile beds. Adv. Water Res. **72**, 32–44 (2014)

Vowinckel, B., Jain, R., Kempe, T., Fröhlich, J.: Entrainment of single particles in a turbulent open-channel flow: a numerical study. J. Hydraul. Res. **54**(2), 158–171 (2016)

Vreman, A.W., Kuerten, J.G.M.: Turbulent channel flow past a moving array of spheres. J. Fluid Mech. **856**, 580–632 (2018). https://doi.org/10.1017/jfm.2018.715

Wang, J.-X, Wu, J.-L., Xiao, H.: Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. Phys. Rev. Fluids **2**, 034603 (2017). https://doi.org/10.1103/PhysRevFluids.2.034603

Weatheritt, J., Sandberg, R.: A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship. J. Comput. Phys. **325**, 22–37 (2016). https://doi.org/10.1016/j.jcp.2016.08.015

Weatheritt, J., Sandberg, R.D.: The development of algebraic stress models using a novel evolutionary algorithm. Int. J. Heat Fluid Flow **68**, 298–318 (2017). https://doi.org/10.1016/j.ijheatfluidflow.2017.09.017

Wu, J.-L., Xiao, H., Paterson, E.: Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. Phys. Rev. Fluids **3**, 074602 (2018). https://doi.org/10.1103/PhysRevFluids.3.074602

Xiao, H., Cinnella, P.: Quantification of model uncertainty in RANS simulations: a review. Progr. Aerosp. Sci. **108**, 1–31 (2019). https://doi.org/10.1016/j.paerosci.2018.10.001

Yang, Y., Peng, H., Wen, C.: Sand transport and deposition behaviour in subsea pipelines for flow assurance. Energies **12**(21), 4070 (2019). https://doi.org/10.3390/en12214070

Yoon, J.-Y., Kang, S.-K.: A numerical model of sediment-laden turbulent flow in an open channel. Can. J. Civ. Eng. **32**(1), 233–240 (2005). https://doi.org/10.1139/l04-089

Yu, X., Hsu, T.-J., Hanes, D.M.: Sediment transport under wave groups: Relative importance between nonlinear waveshape and nonlinear boundary layer streaming. J. Geophys. Res. Oceans (2010). https://doi.org/10.1029/2009JC005348

Zhu, L.-T., Chen, X.-Z., Ouyang, B., Yan, W.-C., Lei, H., Chen, Z., Luo, Z.-H.: Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors. Ind. Eng. Chem. Res. **61**(28), 9901–9949 (2022). https://doi.org/10.1021/acs.iecr.2c01036

Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Royal Stat. Soc. Ser. B (Statistical Methodology) **67**(2), 301–320 (2005). https://doi.org/10.1111/j.1467-9868.2005.00503.x