Spring 4-12-2023

Development of models for the velocity-pressure gradient correlations in incompressible wall-bounded planar turbulent flows using multiple linear regression

Juampablo E. Heras Rivera
University of New Mexico - Main Campus

Juampablo Enrique Heras Rivera

Candidate

Mechanical Engineering

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Svetlana V. Poroseva , Chairperson

Peter Vorobieff

Daniel Banuti

Development of models for the velocity-pressure gradient correlations in incompressible wall-bounded planar turbulent flows using multiple linear regression

by

Juampablo Enrique Heras Rivera

B.S. Mechanical Engineering and B.S. Mathematics, 2022

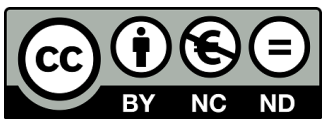
THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science in
Mechanical Engineering

The University of New Mexico
Albuquerque, New Mexico

May, 2023



DEDICATION

Firstly, I would like to thank my wonderful parents, Dr. Eva Linette Rivera Lebrón and Dr. Alfonso Enrique Heras Llanos, for their support and guidance throughout my life and educational career. I am also immensely grateful to my professor, Dr. Svetlana V. Poroseva, for believing in me and mentoring me as a young student as I transition to a professional. Finally, I am thankful to my family and friends for keeping my best interests in mind and filling my life with joy.

ACKNOWLEDGEMENTS

The work of J. E. Heras Rivera has been supported in part by the National Science Foundation EPSCoR Cooperative Agreement OIA-1757207 and by the New Mexico Space Grant Consortium (NMSGC), National Aeronautics and Space Administration (NASA) Cooperative Agreement Number 80NSSC20M0034.

Development of models for the velocity-pressure gradient correlations in incompressible wall-bounded planar turbulent flows using multiple linear regression

by

Juampablo Enrique Heras Rivera

B.S. in Mechanical Engineering and B.S. in Mathematics,

M.S. in Mechanical Engineering

University of New Mexico, 2023

ABSTRACT

In this manuscript, the multiple linear regression method is applied to clarify terms and their coefficients in data-driven models for velocity/pressure-gradient (VPG) correlations in the Reynolds stress transport equations. Additionally, a method for developing universal linear models for the VPG correlations with unchanging coefficients when complex conditions arise in a flow is introduced. The generated models were assessed using residual analysis to ensure an appropriate level of accuracy. Data from direct numerical simulation in an incompressible fully-developed turbulent channel flow at $Re_\tau = 392$ with an adverse pressure gradient is used in the study as the data source.

CONTENTS

LIST OF FIGURES

–

LIST OF TABLES

INTRODUCTION

Current modeling of turbulent flows relevant to engineering applications involves the use of the Reynolds Averaged Navier-Stokes (RANS) equations. In RANS equations, the high resolution of the Direct Numerical Simulation (DNS) method, where all turbulence scales are calculated, is given up in exchange for fast calculations. However, even the simpler RANS equations invoke a significant challenge for modelers since they include the Reynolds stresses that have to be modeled. This makes a solution of the RANS equations sensitive to the modeling strategy chosen.

When applied to engineering problems, Reynolds stresses are modeled using the linear-eddy viscosity models based on the Boussinesq hypothesis. In this hypothesis, the Reynolds stresses are modeled as proportional to the mean strain rate of a flow by a factor of the "turbulence viscosity". This approach leads to a family of one- and two-equation models such as the k-ω model (first suggested by Kolmogorov [1], translated to English by Spalding and later improved in Wilcox [2]), k-ε model [3], [4], Spalart-Allmaras model [5], and their variations [6]. However, these models are known to underperform in flows with high degrees of anisotropy, significant streamline curvature, flow separation, recirculation, and flows influenced by mean rotational effects [7]. Also, the linear-eddy viscosity models cannot reproduce the behavior of turbulent flows approaching the rapid distortion theory (RDT) limit or the return to isotropy behavior in decaying turbulent flows [8].

Reynolds-stress transport (RST) models where transport equations for the Reynolds stresses are included have the potential to provide better predictions than simpler one- and two-equation models at a computational expense much lower than methods which resolve turbulence scales. In particular, such models may be able to reproduce the directional effects of the Reynolds stresses and additional complex interactions in turbulent flows [9], and accurately describe the return to isotropy of decaying turbulence and the behavior of turbulence in the RDT limit [10]. For these reasons, RST models are considered in this study.

In RST models, three terms have to be modeled, namely the velocity-pressure gradient (VPG) correlations, turbulent diffusion, and dissipation terms. In homogeneous turbulence, modeling is reduced to the pressure-strain correlations and the dissipation terms. However, this study is concerned with inhomogeneous turbulent flows and, particularly with model-

ing VPG correlations in such flows.

The VPG correlations, Π, are the terms that describe the interaction of turbulent velocity and pressure fields. The more familiar pressure-strain correlations, Φ, are a convenient substitution for the VPG correlations for modeling purposes only in *homogeneous* flows (an impractical assumption for engineering flows). To highlight the subtle yet significant difference, expressions for VPG and pressure-strain correlations in the RST equations for incompressible flow are compared below, written in Cartesian coordinates with $\vec{x} = (x, y, z)$ and $\vec{u} = (u, v, w)$:

$$\Pi_{ij} = -\frac{1}{\rho}\left[\overline{u_j \frac{\partial p}{\partial x_i}} + \overline{u_i \frac{\partial p}{\partial x_j}}\right] \quad \textbf{vs.} \quad \Phi_{ij} = \frac{1}{\rho}\left[\overline{p\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right)}\right]. \tag{1}$$

In these equations, p is the flow pressure fluctuation, ρ is the fluid density, and $\overline{\omega}$ represents an ensemble average of parameter ω.

Current efforts for modeling the pressure-strain correlations include a wide variety of traditional linear and nonlinear modelling approaches [11]–[13], and machine learning models [14], [15]. These models are broadly categorized as empirical or data-driven, and require the use of many coefficients or training of neural networks. For decades, numerous attempts to improve the performance of Φ models in inhomogeneous flows through the use of empirical, semi-empirical, and artificial correlations resulted in rather modest success in predicting the behavior of simpler flows. With such evidence in place, it is reasonable to return to modeling the VPG correlations, since they are apt for inhomogeneous flows by nature.

The effort of modeling the VPG correlations was never completely abandoned and can be traced back to [16], where it was suggested to derive models for the VPG correlations from the analysis of their exact integro-differential expressions striving to preserve the tensor properties of relevant two-point turbulence statistics in a model for one-point VPG correlations. Various linear and non-linear models were developed in such a manner over the years [10] including for the correlations up to the fourth order, applicable to complex flows with separation and under rotation (see e.g. [17]–[22]). However, these models (like many others for turbulent flows) rely on adjustment of their model coefficients in different flows to achieve more accurate results when compared to reference data. Additionally, it was concluded in [20] that matching the tensor-invariant properties of two-point correlations may not be sufficient for developing one-point VPG correlation models accurate in

the near-wall flow regions.

For these reasons, a data-driven approach to modeling VPG correlations in wall-bounded flows was proposed by Poroseva and Murman in [20], [23]. The model expressions for the correlations up to the fourth order were obtained from the analysis of data from DNS data of incompressible planar wall-bounded turbulent flows. In [24]–[27], the models were further clarified and successfully validated using available datasets for wall-bounded flows of different geometries [28]–[31], at different Reynolds numbers, and in the presence of flow separation. Of particular importance was the demonstration that the linear models for VPG correlations that include only the existing terms in RANS equations for velocity moments of corresponding orders can accurately reproduce the correlation behavior in the entire flow area including the wall vicinity *without* variation of the model coefficients.

Whereas the models were shown to work excellently in planar wall-bounded turbulent flows, the models have not yet been validated in free-shear flows, the scalar coefficients have no physical interpretation yet, and the models have not been generalized to three dimensions. This requires fixing model coefficient values in a manner that accounts for DNS data inaccuracies. Indeed, it was shown in [26], [27] that the choice for the initial DNS dataset used to determine model coefficients slightly affects the coefficient values for VPG correlation models. Additionally, the models include diffusion terms with a non-negligible contribution to the models, but are in some regions in the same order of magnitude as balance errors in the transport equation budgets collected from DNS [32]. This brings to question the physical meaning of the diffusion term coefficients.

The work in this manuscript discusses the modeling of the velocity-pressure gradient using the data-driven linear model approach proposed initially by Poroseva and Murman [20], [23] supplemented with Multiple Linear Regression to determine model terms and their respective coefficients. In the first chapter, multiple linear regression is used to generate models using DNS data for an unstrained channel flow [31], with residual distributions utilized to quantify a model's quality. For each VPG correlation, models are simplified following a stepwise procedure until a final model is obtained that is simple while preserving accuracy. In the second chapter, a new method for applying regression is presented, where a single model is generated for each VPG correlation by optimizing the model performance over multiple channel strain rates. After that, the initial optimized model undergoes simplification using a stepwise reduction procedure. The model quality is quantified using residual analysis similar to the one described in the first chapter.

DATA

0.1. Mathematical Framework

For the majority of flows relevant to industry, CFD simulations using Reynolds-Averaged Navier-Stokes (RANS) equations are the only practical option. The RANS approach is appealing because it allows for the use of coarse computational grids and in many practical situations the flow is steady in the mean so the problem reduces to solving spatial variations [10].

Obtaining the RANS equations involves the use of the Reynolds decomposition, in which an instantaneous variable is expressed as the sum of the mean and the fluctuation around that mean. The decomposed parameters replace the instantaneous parameters in the Navier-Stokes equations, and then the equations are averaged in time, which results in the following equations of motion:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho U_i)}{\partial x_i} = 0, \tag{2}$$

$$\frac{\partial (\rho U_i)}{\partial t} + \frac{\partial (\rho U_j U_i)}{\partial x_j} = \sum_n F_i^n - \frac{\partial P}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i} - \rho\overline{u_i u_j}\right)\right]. \tag{3}$$

In these equations, t is the time variable, ρ is the fluid density, P is the mean flow pressure, μ is the fluid dynamic viscosity, and F_i^n are body forces acting on the flow. Additionally, U_i is the mean flow velocity, x_i is the Cartesian coordinate, with $i = 1, 2, 3$ corresponding to streamwise, wall-normal, and spanwise directions, respectively. The final term in the square brackets, $\rho\overline{u_i u_j}$ is referred to as the Reynolds stress tensor. The components of this tensor are unknown terms in the RANS equations of motion and require modeling to close the equations. Hereafter, the body force term is not considered because there are no external forces acting on the flow being studied in this work.

An exact equation for the transport of the Reynolds stress, $\overline{u_i u_j}$ is derived from the equation for velocity fluctuations u_i obtained by subtracting the RANS momentum equation (3) from that for the instantaneous velocity. This equation, named the Reynolds stress transport equation, is as follows:

$$\frac{D\overline{u_i u_j}}{Dt} \equiv \frac{\partial \overline{u_i u_j}}{\partial t} + U_k \frac{\partial \overline{u_i u_j}}{\partial x_k}$$
$$= -\left(\overline{u_i u_k}\frac{\partial U_j}{\partial x_k} + \overline{u_j u_k}\frac{\partial U_i}{\partial x_k}\right) - \frac{1}{\rho}\left(\overline{u_i \frac{\partial p}{\partial x_k}} + \overline{u_j \frac{\partial p}{\partial x_i}}\right) - 2\nu\overline{\frac{\partial u_i}{\partial x_k}\frac{\partial u_j}{\partial x_k}}$$
$$+ \frac{\partial}{\partial x_k}\left[\nu\frac{\partial \overline{u_i u_j}}{\partial x_k}\right] - \frac{\partial}{\partial x_k}\left[\overline{u_i u_j u_k}\right] \quad (4)$$

Terms in this equation can be grouped to allow for physical interpretation of the processes with the following labels:

$$P_{ij} = \left(\overline{u_i u_k}\frac{\partial U_j}{\partial x_k} + \overline{u_j u_k}\frac{\partial U_i}{\partial x_k}\right),$$
$$\Pi_{ij} = \frac{1}{\rho}\left(\overline{u_i \frac{\partial p}{\partial x_k}} + \overline{u_j \frac{\partial p}{\partial x_i}}\right),$$
$$\varepsilon_{ij} = 2\nu\left(\overline{\frac{\partial u_i}{\partial x_k}\frac{\partial u_j}{\partial x_k}}\right),$$
$$D^M_{ij} = \frac{\partial}{\partial x_k}\left[\nu\frac{\partial \overline{u_i u_j}}{\partial x_k}\right],$$
$$D^T_{ij} = \frac{\partial}{\partial x_k}\left[\overline{u_i u_j u_k}\right].$$

In this grouping, P_{ij} is the stress production due to the mean velocity gradient, Π_{ij} is the velocity-pressure gradient correlation, ε_{ij} is the stress dissipation rate, D^M_{ij} is the molecular diffusion of the Reynolds stress tensor, and D^T_{ij} is the turbulent diffusion of the Reynolds stress tensor [10]. Three terms in this equation require modeling, namely Π_{ij}, ε_{ij}, and D^T_{ij}. The work in this manuscript is concerned with modeling the velocity-pressure gradient correlation, Π_{ij}. The grouped terms are also referred to as "budget" terms in the RANS equations.

0.2. DNS Dataset Description

The data used for this study is from DNS of a fully-developed channel flow at Reynolds number $Re_\tau = \delta u_\tau/\nu = 392$ (based on the channel friction velocity and the channel width) conducted in [33]. The data includes budget terms in the RANS equations in the three Cartesian directions, x, y, and z, where x is the streamwise direction, y is the wall-normal direction, and z is the spanwise direction, as shown in Figure 1. Alternatively, the three Cartesian directions x, y, and z can be denoted by 1, 2, and 3, respectively.

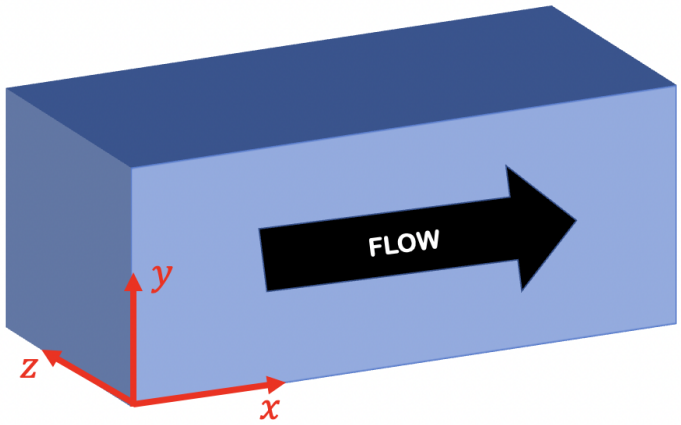


Figure 1: Diagram showing channel flow coordinates

DNS was conducted for a conventional turbulent channel flow and for a strained channel flow in [33]. The channel was strained to reproduce properties of an adverse pressure gradient (APG) boundary layer such as those observed under flow separation. In the DNS, the channel was progressively strained in time by simultaneously applying a sliding motion and a straining motion to the flow domain as shown in Figure 2.

The strain rate, $A_{ij} = \partial U_i/\partial x_j$, is zero at $t = 0$ for the unstrained channel case and at $t > 0$, it is spatially uniform and constant with time, such that the resulting strain increases linearly with time. This tensor has two non-zero components: $A_{11} = \frac{\partial U}{\partial x} < 0$ (streamwise compression) and $A_{22} = \frac{\partial V}{\partial y} > 0$ (wall-normal expansion), which satisfy the condition $A_{11} + A_{22} = 0$.

The wall sliding motion is synchronized with the strain to reproduce the bulk deceleration of the APG, ultimately leading to flow reversal at the walls. The difference between the mean centerline velocity, $U_c(t)$, and the channel wall velocity, $U_w(t)$, diminishes expo-

nentially following the relation $U_c(t) - U_w(t) = U_c \exp(A_{11}t)$.

The strain values used in this study are at $A_{22}t = 0$ (unstrained), 0.020, 0.281, 0.365, 0.675, and 0.772, where t is time. After the strain of $A_{22}t = 0.675$, there is separation in the flow.

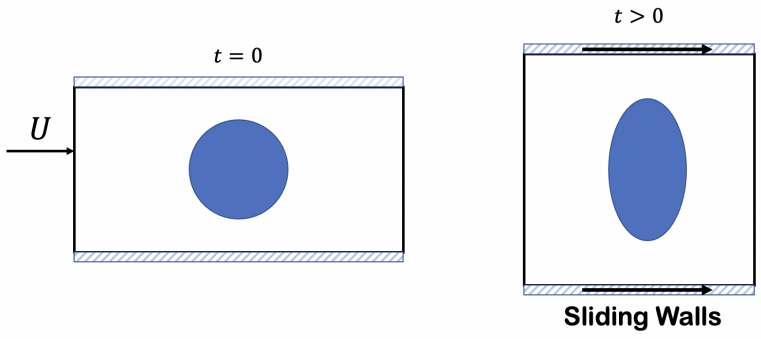


Figure 2: Strained channel progression with time

Part I

Model Development for an Incompressible Channel Flow

1. METHODS

In this section, the method of multiple linear regression is applied to developing models for the VPG correlations in the unstrained channel flow.

1.1. Multiple Linear Regression using Least Squares

In this study, multiple linear regression (MLR) using the least squares method was used for developing data-driven models for VPG correlations in the Reynolds Stress Transport Equations. In the MLR approach, the best fit, $\hat{\boldsymbol{\Theta}}$, is sought for the observed variable, $\boldsymbol{\Theta}$, as a weighted sum of multiple regressors:

$$\hat{\boldsymbol{\Theta}} = a_1 \boldsymbol{r_1} + a_2 \boldsymbol{r_2} + \cdots + a_t \boldsymbol{r_t} \tag{1.1}$$

where $a_i, i = 1, 2, \cdots, t$ are coefficients found to optimize the fit using the method of least squares, outlined in this section. Variables $\boldsymbol{\Theta}$, $\hat{\boldsymbol{\Theta}}$, and r_i are $d \times 1$ vectors. The scalar coefficients a_i are concatenated into a $t \times 1$ vector, $\boldsymbol{A} = [a_1\, a_2\, \cdots\, a_t]^T$. Also, regressor vectors are concatenated to form a $d \times t$ matrix, $\boldsymbol{R}$, with each of regressor vector as a column:

$$\begin{bmatrix} r_{11} & \cdots & r_{1t} \\ \vdots & \ddots & \vdots \\ r_{d1} & \cdots & r_{dt} \end{bmatrix}.$$

As a result, (1.1) can be expressed more compactly as $\hat{\boldsymbol{\Theta}} = \boldsymbol{RA}$.

Now to quantify the difference between $\hat{\boldsymbol{\Theta}}$ and $\hat{\boldsymbol{\Theta}}$, the residual, ε, is defined as:

$$\boldsymbol{\varepsilon} = \boldsymbol{\Theta} - \hat{\boldsymbol{\Theta}} = \boldsymbol{\Theta} - \boldsymbol{RA}. \tag{1.2}$$

Then, the scalar coefficients in (1.1) are found by minimizing the sum of the squares of the residuals, $\sum_{j=1}^{d} \varepsilon_j^2$, with the objective function,

$$\boldsymbol{S(A)} = \sum_{j=1}^{d} \varepsilon_j^2 = \boldsymbol{\varepsilon^T \varepsilon} = (\boldsymbol{\Theta - RA})^T(\boldsymbol{\Theta - RA}) = \boldsymbol{\Theta^T\Theta - \Theta^T RA - A^T R^T \Theta + A^T R^T RA}. \tag{1.3}$$

Note that each term in (1.3) is of dimension 1×1. Now since $(\boldsymbol{\Theta}^T \boldsymbol{R} \boldsymbol{A})^T = \boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{\Theta}$, and the transpose of a 1×1 matrix is the matrix itself, the second and third term in (1.3) can be combined as $-\boldsymbol{\Theta}^T \boldsymbol{R} \boldsymbol{A} - \boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{\Theta} = -2\boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{\Theta}$. Thus, $\boldsymbol{S}(\boldsymbol{A})$ can be written as:

$$\boldsymbol{S}(\boldsymbol{A}) = \boldsymbol{\Theta}^T \boldsymbol{\Theta} - 2\boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{\Theta} + \boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{R} \boldsymbol{A}. \tag{1.4}$$

To minimize the objective function, $\boldsymbol{S}(\boldsymbol{A})$, take the derivative of $\boldsymbol{S}(\boldsymbol{A})$ with respect to $\boldsymbol{A}$ and set it equal to zero:

$$\frac{d\boldsymbol{S}(\boldsymbol{A})}{d\boldsymbol{A}} = \frac{d}{d\boldsymbol{A}} \left[\boldsymbol{\Theta}^T \boldsymbol{\Theta} - 2\boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{\Theta} + \boldsymbol{A}^T \boldsymbol{R}^T \boldsymbol{R} \boldsymbol{A} \right] = 0. \tag{1.5}$$

Further simplification leads to:

$$-2\boldsymbol{R}^T \boldsymbol{\Theta} + 2\boldsymbol{R}^T \boldsymbol{R} \boldsymbol{A} = 0 \tag{1.6}$$

This equation is rearranged to form the least-squares estimator of $\boldsymbol{A}$ as:

$$\boldsymbol{A} = (\boldsymbol{R}^T \boldsymbol{R})^{-1} \boldsymbol{R}^T \boldsymbol{\Theta}. \tag{1.7}$$

After the coefficients vector, $\boldsymbol{A}$, are found, the fitting process is complete. Further, the quality of the fit is determined by inspection of the residual ($\boldsymbol{\varepsilon}$) plots. Specifically, a "homoscedastic" fit is preferred where model residuals have the same finite variance through all predictions made. A homoscedastic fit is preferred because it implies that the model errors are normally distributed. An example of such a fit is shown in 1.1(a). In contrast, when residuals have differing variance through all predictions made, the residual distribution is said to be "heteroscedastic". A heteroscedastic residual distribution implies that there is an unaccounted bias, possibly explained by an incomplete set of regressors or model-form deficiency (e.g. a non-linear model would be more appropriate). Examples of heteroscedastic fits are shown in 1.1(b) and 1.1(c).

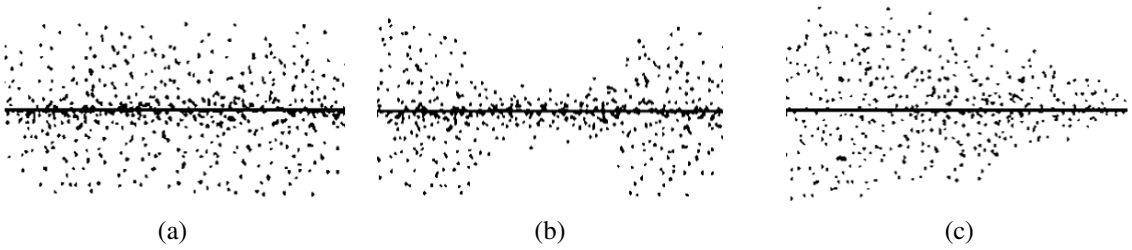


Figure 1.1: Distributions of residuals: a) homoscedastic, b) and c) heteroscedastic.

The calculations used for the model development as described in this section were completed using MATLAB [34].

2. RESULTS

The results of an initial analysis of the application of the standard MLR procedure described in section 1.1 to the modeling of the velocity/pressure-gradient correlations are shown in this section. The data used in this section is DNS data from [31] of the velocity/ pressure-gradient correlations at a channel strain rate of $A_{22}t = 0$. At this strain rate value, the channel flow has no adverse pressure gradient, which results in a reduced flow complexity in comparison to flows at higher strain rates. This simple flow will serve as a benchmark to decide the applicability of the MLR method to the modeling of the VPG correlations.

For each VPG correlation, the initial models found in previous works are compared to the models found using the MLR procedure. Then, if the residual distributions indicate that the models are unsatisfactory, adjustments are made to the list of regressors until a satisfactory fit is found.

2.1. Π_{xy} Correlations

Figure 2.1(a) compares the DNS data for the correlation Π_{xy} from [31] (black dots) with the model profile for this correlation (solid line) obtained using the Π_{xy} model from [26]:

$$\Pi_{xy} = -0.3D^M_{xy} - 0.92D^T_{xy} - 0.92P_{xy}. \tag{2.1}$$

In Figure 2.1(b), the same DNS data are compared with the model from [27]:

$$\Pi_{xy} = -0.3D^M_{xy} - 0.8D^T_{xy} - 0.92P_{xy} - 0.2D^M_{yy}. \tag{2.2}$$

In these figures and those that follow, all budget terms are normalized with respect to friction velocity, u_τ, and kinematic viscosity, ν.

Adding the term with molecular diffusion in (2.2) appears to improve the agreement between the model profile and the DNS data near the channel wall (Fig. 2.1(b)). The residual distribution near the wall also becomes more random (Fig. 2.1(e)) to compare with that on Fig. 2.1(d). Nevertheless, residual distributions for both models are heteroscedastic

(see Fig. 1.1 for reference), which indicates that these models are likely to be incomplete. In figures 2.1(d)-2.1(f) and those below with the residual distributions, the parameter i runs from 1 to 97, with $i = 1$ and 97 corresponding to locations at the channel wall and the channel axis, respectively. MLR applied to the regressors of model (2.1) to clarify the coefficient values gives the following expression:

$$\Pi_{xy} = -0.276D^M_{xy} - 0.757D^T_{xy} - 0.908P_{xy}. \tag{2.3}$$

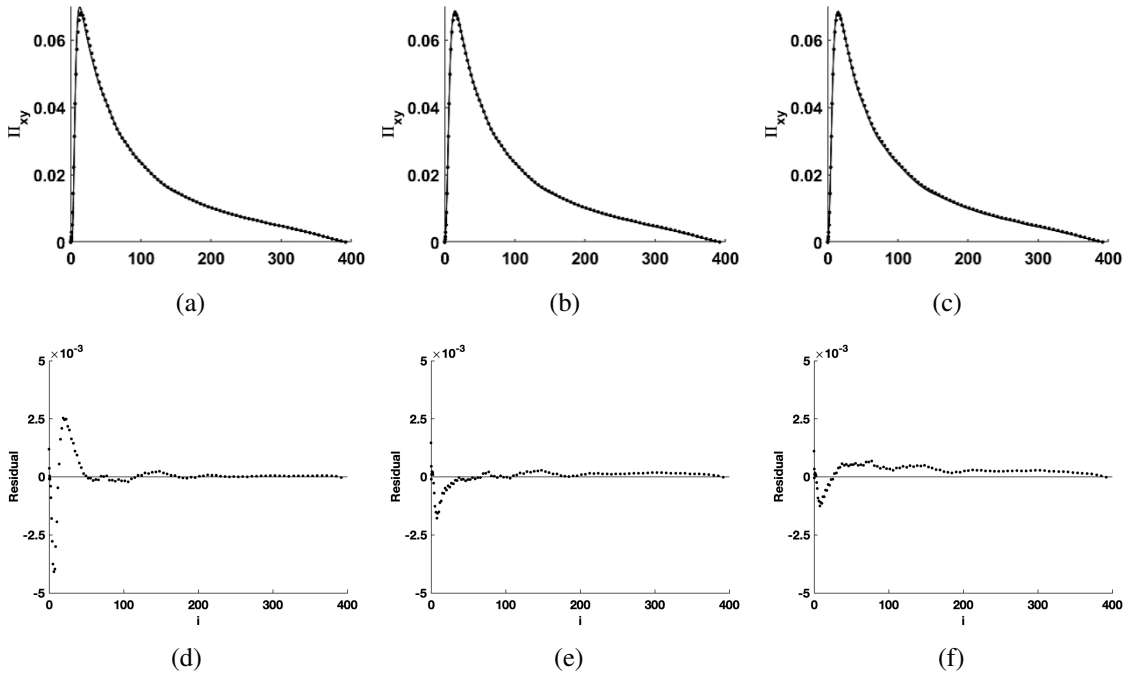
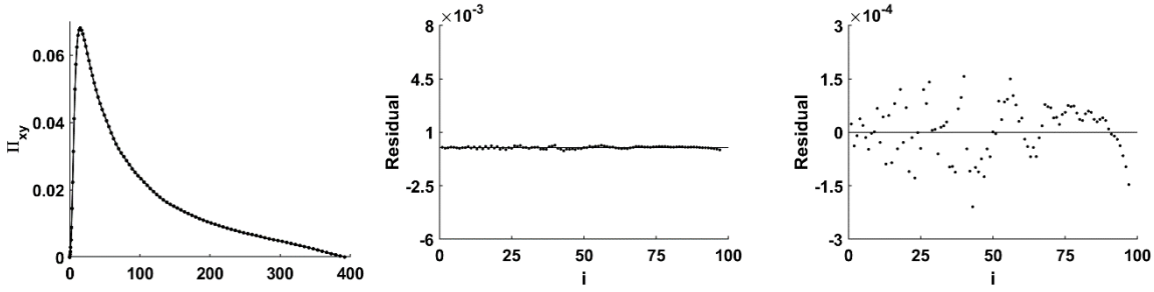


Figure 2.1: Comparison of the DNS (black dots) and model (solid lines) profiles for Π_{xy}: (a), (b) and (c). The residual distributions for models (2.1) - (2.3): (d), (e) and (f). Figures (a), (d) - model (2.1); (b), (e) - model (2.2); (c), (f) – MLR fit (2.3).

Based on these observations, we increased the set of regressors to include all terms in the Reynolds Stress Transport Equation, such as: turbulence production, molecular diffusion, and turbulent diffusion. We then fit a MLR model with this new set of regressors to obtain the following model:

$$\begin{aligned}\Pi_{xy} = -0.212D^M_{xy} - 0.790D^T_{xy} - 0.934P_{xy} - 0.005P_{xx} - 0.003D^T_{xx} - 0.064D^T_{yy} + \\ 0.552D^M_{yy} + 0.242D^T_{zz} + 0D^M_{xx} + 0D^M_{zz}.\end{aligned} \tag{2.4}$$

Visual comparison of the fit from model (2.4) with the DNS data does not reveal much change in comparison to the results shown in Figures 2.1(b) and 2.1(c). However, there is a clear improvement in the residual distribution as is illustrated in Figures 2.2(a) and 2.2(b). The magnitude of the residuals decreases significantly, and the residual distribution become more homoscedastic for model (2.4).



(a) (b) (c)

Figure 2.2: Results for MLR fit model (2.4): a) comparison of the DNS profiles (circles) and MLR fit model (2.4) (solid lines), b) and c) the residual distribution for model (2.4) with different scales.

In MLR fit (13), values of coefficients for D^M_{xx} and D^M_{zz} are shown as zero after rounding to three figures after the decimal point. Due to their small contributions, these terms were removed from the set of regressors as the next step. The resulting MLR fit is the following:

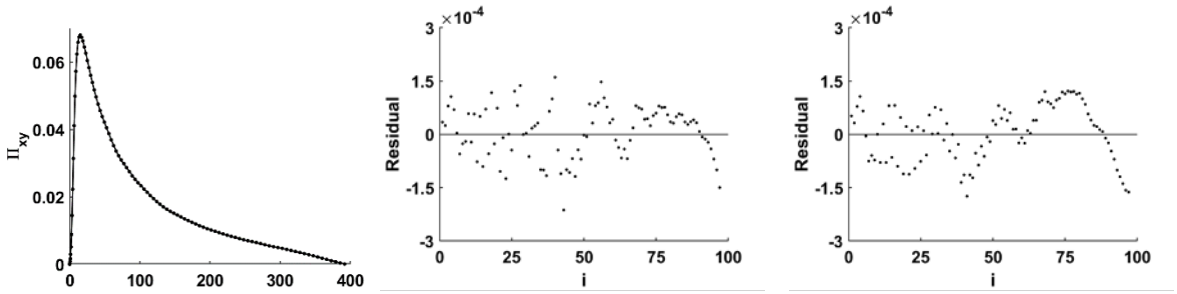
$$\Pi_{xy} = -0.211D^M_{xy} - 0.790D^T_{xy} - 0.934P_{xy} - 0.005P_{xx} - 0.004D^T_{xx} - 0.073D^T_{yy} + 0.607D^M_{yy} + 0.261D^T_{zz}. \tag{2.5}$$

This procedure did not change values of the coefficients for the first four terms in (2.4) and slightly altered the coefficients for the remaining four terms. The plots for this new model are shown in Figure 2.3 and indicate no significant penalty in residual magnitude or homoscedasticity as a result of removing the extraneous terms. To analyze the stability of the coefficients in model (2.5), an MLR fit with the same regressors list was done using data with 1000 additional flow realizations from [27], to generate the following model:

$$\Pi_{xy} = -0.207D^M_{xy} - 0.782D^T_{xy} - 0.930P_{xy} - 0.004P_{xx} - 0.001D^T_{xx} - 0.061D^T_{yy} + 0.625D^M_{yy} + 0.300D^T_{zz}, \tag{2.6}$$

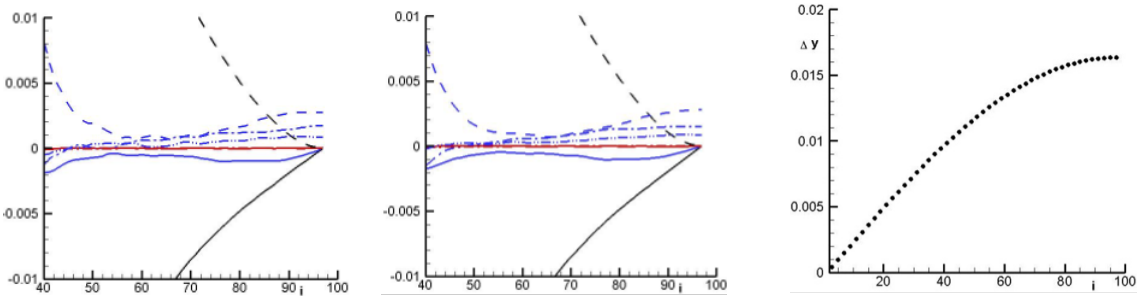
with differing coefficient values from (2.5) shown in red. As shown in Figure 2.3, the residual values for model (2.6) are reduced towards the channel wall and the residual distribution becomes more homoscedastic near the channel wall as well in comparison to (2.5). Therefore, a longer simulation may lead to a more homoscedastic fit, which is expected as computational systematic errors should decrease as variables converge further.

The heteroscedastic residual distribution near the channel axis can potentially be attributed to the grid resolution used in DNS. The grid step in DNS [31], [33] increases significantly from the channel wall to its axis (Fig. 5c). This rate may be too fast for accurate calculation of the derivatives, whose values are close to zero. Additional studies are required to quantify the effects of the grid resolution used in the DNS.



(a) (b) (c)

Figure 2.3: Results for MLR fits (2.5) and (2.6): (a) comparison of the DNS profiles (circles) and the MLR fits (solid lines), (b) the residual distribution for (2.5), and (c) the residual distribution for (2.6).



(a) (b) (c)

Figure 2.4: Budget terms in the Reynolds stress transport equations included as the regressors in (a) MLR fit (2.5) using DNS data from [31] and (b) MLR fit (2.6) using DNS data from [33]. The grid step growth used in DNS [31], [33] is shown in (c). Lines in (a) and (b): solid black - P_{xy}, dashed black - P_{xx}, red - D^M terms, blue solid - D^T_{xy}, blue dashed - D^T_{xx}, blue dashed-dotted - D^T_{yy}, and blue dashed-dot-dot - D^T_{zz}.

2.2. Effects of the Balance Errors on the Π_{xy} Models

In the previous sections, the effect of inaccuracies in the DNS data has not been considered when selecting the set of regressors used for creating a MLR model. However, the effect of this inaccuracy is not negligible for modeling purposes [32]. In DNS datasets, the balance errors account for the inaccuracy of terms in the transport equations (hereafter referred to as budget terms). For example, in the budget of the shear stress $\langle uv \rangle$, the balance errors, Err_{xy}, are defined as $\mathrm{Err}_{xy} = P_{xy} + \Pi_{xy} + D^T_{xy} + D^M_{xy} - \epsilon_{xy}$.

Figure 2.5(a) shows shows that in the dataset [31], two molecular diffusion terms, D^M_{xy} and D^M_{yy} used as regressors in the MLR fits for Π_{xy} above are smaller than the balance errors Err_{xy} in the transport equation for this Reynolds stress. Their values exceed that of the balance errors only close to the wall. Some of the turbulent diffusion terms, D^T, are also comparable in value with the balance errors in half of the flow area (not shown here). This issue persists for all mentioned budget terms when using data from [33] where more flow realizations were used to collect statistics (Figure 2.5(b)). Therefore, one should expect for the coefficient values in any MLR fit to be affected by the DNS data inaccuracy.

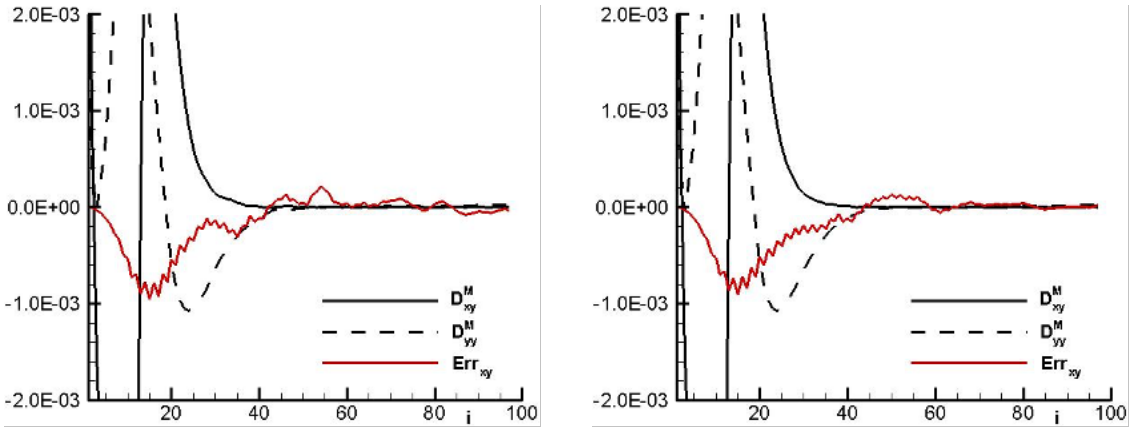
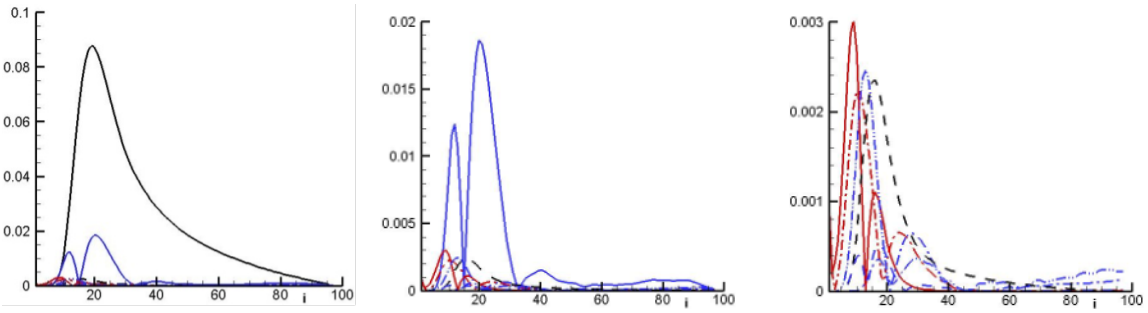


Figure 2.5: Molecular diffusion terms in the Reynolds stress transport equations in comparison with Err_{xy}, (a) DNS data [31], (b) DNS data [33]



(a) (b) (c)

Figure 2.6: Contributions $a_i r_i$ (absolute values) in the MLR fit (2.5) (a) including contributions from P_{xy} (black solid line) and D^T_{xy} (blue solid line), (b) without contribution from P_{xy}, (c) without contributions from P_{xy} and D^T_{xy}. Other lines: black dashed - P_{xx}, red solid – D^M_{xy}, red dashed-dotted – D^M_{yy}, blue dashed - D^T_{xx}, blue dashed-dotted - D^T_{yy}, and blue dashed-dot-dot - D^T_{zz}

It was shown in [32] that adding the balance errors as an additional term when solving RANS equations with all terms but the molecular diffusion substituted with their DNS profiles (RANS-DNS simulations), helps dramatically in obtaining the right solutions for the equations. Following that approach, **we propose to add the balance errors as regressors in models for the VPG correlations with the purpose of determining more accurate coefficients for the remaining terms**. We hope these accurate coefficients will allow for a physical interpretation of the model coefficients and provide information on the contribution of each regressor to the VPG correlations.

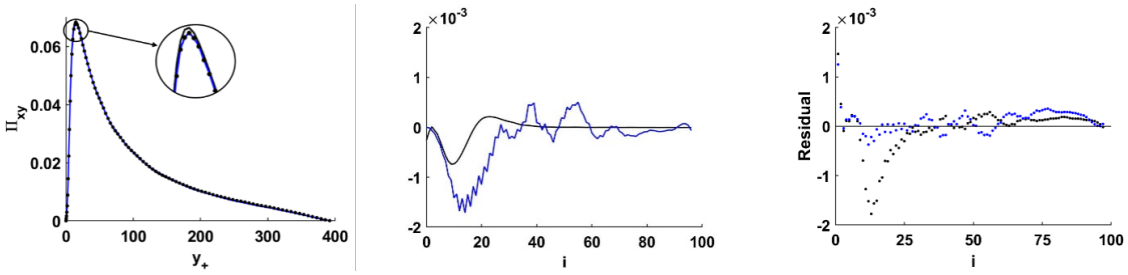
Adding the balance error term, Err_{xy} to the set of regressors used in model (2.3) and applying MLR generates the following model:

$$\Pi_{xy} = -0.314 D^M_{xy} - 0.766 D^T_{xy} - 0.919 P_{xy} + 2.010\,\mathrm{Err}_{xy}\,. \tag{2.7}$$

In (2.7), the values of coefficients for D^M_{xy} and P_{xy} are -0.143 and -0.919 to compare with those in (2.3): -0.276 and -0.908, respectively. When rounded, these are the same values that were used in models (2.1) and (2.2) obtained without using any algorithm. The coefficient value in front of D^T_{xy} is now 0.766, which falls between values used in (2.2) and (2.3).

Figure 2.7 compares profiles of Π^{MLR5}_{xy} and $\Pi^{(M2)}_{xy}$ (Figure 2.7(a)) and the terms $2.01\,\mathrm{Err}_{xy}$ in (2.7) and $-0.2 D^M_{yy}$ in (2.2) (Figure 2.7(b)). The regression distributions for (2.2) and (2.7) are compared in Figure 2.7(c). Figure 2.7 demonstrates that not only the residual values are

reduced substantially for the MLR fit (2.7), but also the residual distribution for (2.7) tends to be more homoscedastic to compare with that of (2.2). Therefore, the presence of D^M_{yy} in the set of regressors can be attributed to data inaccuracies rather than to physical processes, and the data inaccuracies are better absorbed by the balance errors included in the model.



(a) (b) (c)

Figure 2.7: Comparisons of (a) MLR fits (2.2) and (2.7) for Π_{xy}, (b) terms $-0.2D^M_{yy}$ in (2.2) and $2.01\,\mathrm{Err}_{xy}$ in (2.7), and (c) residual distributions for (2.2) and (2.7). Lines: black for (2.2) and blue for (2.7)

Following the procedure described above, we can understand the contribution of each regressor used in (2.7), particularly the significance of small terms such as molecular diffusions. Terms with small contribution are removed for the sake of model simplicity. First, when removing D^M_{xy} from the set of regressors in (2.7), the MLR fit is as follows:

$$\Pi_{xy} = -0.822D^T_{xy} - 0.914P_{xy} + 0.752\,\mathrm{Err}_{xy}\,. \tag{2.8}$$

Although model (2.8) fits the data well (Figure 2.8(a)), the fit has a poor residual distribution near the wall region in comparison to model (2.7), as shown in Figure 2.8(d). This indicates that although D^M_{xy} has a small magnitude (comparable to that of balance errors), its presence in the set of regressors is beneficial to the model.

Next, we will consider the effect of adding balance errors as regressors to the full set of regressors used in model (2.4). The balance errors added correspond to the regressors in the original list (i.e. Err_{xx} is only added as a regressor if a term in this direction, like D^M_{xx}, is in the original regressors list). In this case, model (2.4) includes terms from the transport equations for all Reynolds stresses as regressors, so each corresponding balance error is added to the set of regressors. The fit with this set of regressors is as follows:

$$\begin{aligned}\Pi_{xy} = -0.221D^M_{xy} - 0.791D^T_{xy} - 0.935P_{xy} - 0.005P_{xx} - 0.003D^T_{xx} + 0.563D^M_{yy} - \\ 0.047D^T_{yy} + 0.237D^T_{zz} + 0D^M_{xx} + 0.001D^M_{zz} + 0.925\,\mathrm{Err}_{xy} + \\ 0.086\,\mathrm{Err}_{xx} + 0.294\,\mathrm{Err}_{yy} - 0.369\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.9}$$

Results for this fit are shown in Figures 2.8(b) and 2.8(e).

Three terms in model (2.9), namely D^M_{xx}, D^T_{xx}, and D^M_{zz}, have negligible contributions to the fit as a result of the small coefficient values and are removed to produce the following fit:

$$\begin{aligned}\Pi_{xy} = -0.213D^M_{xy} - 0.775D^T_{xy} - 0.926P_{xy} - 0.002P_{xx} + 0.614D^M_{yy} - 0.056D^T_{yy} + \\ 0.281D^T_{zz} + 1.009\,\mathrm{Err}_{xy} + 0.096\,\mathrm{Err}_{xx} + 0.621\,\mathrm{Err}_{yy} - 0.223\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.10}$$

The results for this model are very similar to those of model (2.9) and thus are not shown here.

Note that balance errors should be included in the fit when the magnitude of the balance errors is similar to that of a regressor in the same direction. In (2.10), the only term from the Reynolds stress equation in the streamwise direction (xx) is P_{xx} which is significantly larger than Err_{xx}. Therefore, we expect that removing Err_{xx} from the list of regressors should have a minimal impact on the fit. This reduced MLR fit model is as follows:

$$\begin{aligned}\Pi_{xy} = -0.212D^M_{xy} - 0.779D^T_{xy} - 0.925P_{xy} - 0.002P_{xx} + 0.613D^M_{yy} - 0.059D^T_{yy} + \\ 0.296D^T_{zz} + 0.918\,\mathrm{Err}_{xy} + 0.748\,\mathrm{Err}_{yy} - 0.224\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.11}$$

As expected, this reduction step produced results similar to model (2.9) (Figures 2.8(c) and 2.8(f)). The removal of other terms was explored while seeking an improved fit, but every option increased the heteroscedasticity of the residual distribution and the magnitude of residual values. Therefore, model 2.11 is the current final selection for VPG correlation Π_{xy}.

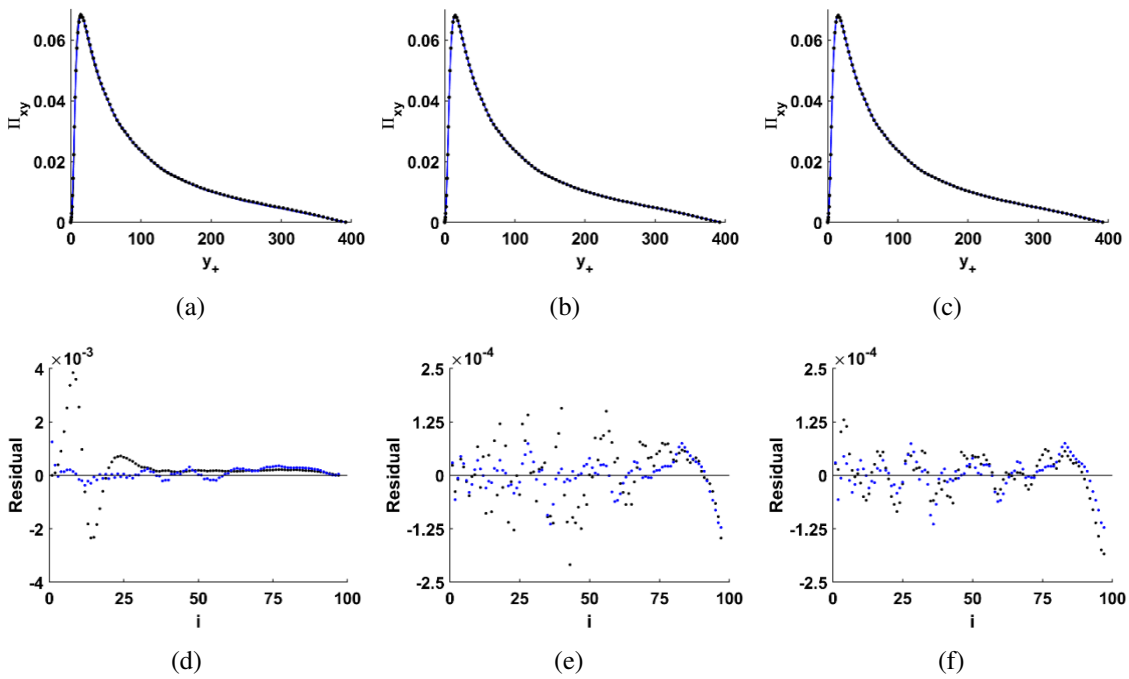


Figure 2.8: Comparisons of MLR fits and DNS data for Π_{xy}: (a) model (2.7) (blue line) and model (2.8) (black line), (b) model (2.4) and model (2.9), (c) model (2.11) and model (2.9), and the residual distributions for (d) model (2.7) (blue dots) and model (2.8) (black dots), (e) model (2.4) (black dots) and model (2.9) (blue dots), (f) model (2.11) (black dots) and model (2.9) (blue dots).

2.3. Π_{xx} Correlations

In this section, we discuss the modeling of Π_{xx} correlations using MLR, and the results are compared with those of models (2.12) - (2.14) suggested in [26], [27].

$$\Pi_{xx} = -0.78\Pi_{xy} - 0.7\Pi_{yy} - 0.25D^T_{xy} + 0.01D^M_{xx}, \tag{2.12}$$

$$\Pi_{xx} = -0.8\Pi_{xy} - 0.6\Pi_{yy} - 0.25D^T_{xy} + 0.01D^M_{xx} + 0.22D^T_{yy}, \tag{2.13}$$

$$\Pi_{xx} = 0.25D^M_{xy} + 0.65D^T_{xy} + P_{xy} + 0.01D^M_{xx} + 0.02P_{xx} + D^T_{yy} - 0.6D^T_{zz}. \tag{2.14}$$

Results for model (2.12) are shown in Figure 2.9: the black line in Figure 2.9(a) are for

the model profile and the black dots in Figure 2.9(b) are for the model residual distribution.

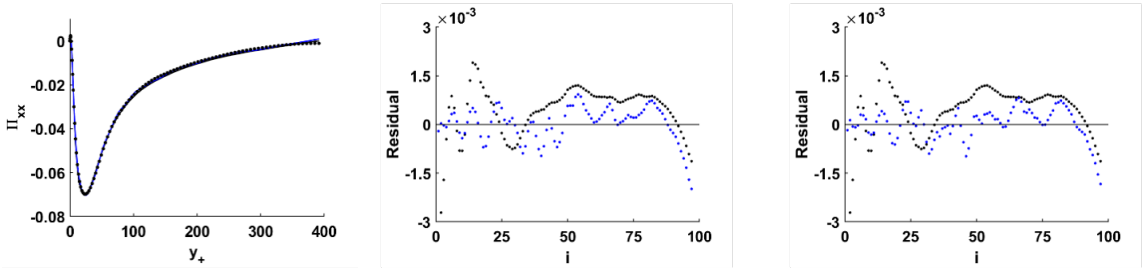
Now, the MLR fit for VPG correlation Π_{xx} with the same regressors list as model (2.12) is as follows:

$$\Pi_{xx} = -0.753\Pi_{xy} - 0.739\Pi_{yy} - 0.265D^T_{xy} + 0.007D^M_{xx}. \tag{2.15}$$

Applying MLR to this set of regressors made little difference to the fit, and the residual distributions for models (2.12) and (2.15) (not shown here) are heteroscedastic and both indicate that the set of regressors is incomplete. Similarly, changing the list of regressors to match that of models (2.13) and (2.14) and fitting an MLR model resulted in heteroscedastic residual distributions. The most improvement was observed with the full set of regressors:

$$\begin{aligned}\Pi_{xx} = 0.860D^M_{xy} + 0.608D^T_{xy} + 1.002P_{xy} + 0.059D^M_{xx} + 0.048D^T_{xx} + 0.022P_{xx} - 2.387D^M_{yy} \\ +1.673D^T_{yy} - 0.153D^M_{zz} - 1.900D^T_{zz}.\end{aligned} \tag{2.16}$$

Results for model (2.16) are shown in Figure 2.9 by the blue line in Figure 2.9(a) and by the blue dots in Figure 2.9(b) for the model residual distribution. Notice that with all its complexity, model (2.16) provides rather modest improvements. The values of model coefficients exceeding one are also difficult to interpret in terms of their relevance to physical processes.



(a) (b) (c)

Figure 2.9: Comparisons of (a) model (2.12) with the MLR fits (2.16) and (2.17) for Π_{xx}, (b) residual distributions for (2.12) and (2.16), (c) residual distributions for (2.12) and (2.17). Lines/dots: black for (2.12) and blue for (2.16) and (2.17).

Following the strategy suggested in the previous section, relevant balance errors were

added to the fit and generated the following model:

$$\Pi_{xx} = 0.911D^M_{xy} + 0.597D^T_{xy} + 0.950P_{xy} + 0.062D^M_{xx} + 0.050D^T_{xx} + 0.013P_{xx} - 2.500D^M_{yy} + 1.718D^T_{yy} - 0.161D^M_{zz} - 2.079D^T_{zz} + 0.348\,\mathrm{Err}_{xy} + 1.090\,\mathrm{Err}_{xx} - 4.875\,\mathrm{Err}_{yy} - 0.369\,\mathrm{Err}_{zz}\,. \tag{2.17}$$

When compared with the DNS profile for Π_{xx}, model (2.16) produced no significant change from the MLR fit (2.16), as shown in Figure 2.9(a). However, the residual distribution improves slightly as seen in Figure 2.9(c), where the residual distribution for model (2.12) (black dots) is compared with that of (2.17) (blue dots). The model coefficients of some small regressors such as D^M_{yy}, D^T_{yy}, and D^T_{zz}, are still larger than one, which is likely to be unphysical. This motivated us to explore the fit performance without these terms starting from those in the spanwise direction (zz).

Removing all regressors in the spanwise direction from the preceding model results in the following MLR model:

$$\Pi_{xx} = -0.097D^M_{xy} + 0.460D^T_{xy} + 1.024P_{xy} + 0.005D^M_{xx} - 0.012D^T_{xx} + 0.044P_{xx} - 0.918D^M_{yy} + 0.924D^T_{yy} - 0.783\,\mathrm{Err}_{xy} + 0.746\,\mathrm{Err}_{xx} - 3.754\,\mathrm{Err}_{yy}\,. \tag{2.18}$$

In the rest of the discussion for Π_{xx}, the direct visual comparison of fits provides little information because of small differences in the models. Therefore, these figures are omitted and the residual distributions are analyzed, as they are more informative. For model (2.18), the residual distribution is shown in Figure 2.10(a) by blue dots, and compared to the residual distribution for model (2.17) shown in black dots. This step does not affect the residual distribution towards the channel wall, but the distribution becomes more heteroscedastic towards the channel axis. On the other hand, the values of all model coefficients but that of P_{xy} are now less than one and the model is simpler. The new fit also includes a very small term (marked by in red in (2.18)), which is the next candidate for removal. This fit and two more steps are shown below, with the terms being removed in each following fit indicated in red:

$$\Pi_{xx} = -0.079D^M_{xy} + 0.489D^T_{xy} + 0.999P_{xy} - 0.026D^T_{xx} + 0.032P_{xx} + 0.007D^M_{yy} + 0.853D^T_{yy} - 1.233\,\mathrm{Err}_{xy} + 0.745\,\mathrm{Err}_{xx} - 3.774\,\mathrm{Err}_{yy}, \tag{2.19}$$

$$\Pi_{xx} = -0.080D^M_{xy} + 0.489D^T_{xy} + 0.999P_{xy} - 0.026D^T_{xx} + 0.032P_{xx} + 0.853D^T_{yy} - 1.235\,\mathrm{Err}_{xy} \\ +0.745\,\mathrm{Err}_{xx} - 3.777\,\mathrm{Err}_{yy}, \tag{2.20}$$

$$\Pi_{xx} = 0.487D^T_{xy} + 0.991P_{xy} - 0.019D^T_{xx} + 0.029P_{xx} + 0.872D^T_{yy} - 2.157\,\mathrm{Err}_{xy} + 0.599\,\mathrm{Err}_{xx} \\ -4.288\,\mathrm{Err}_{yy}\,. \tag{2.21}$$

The residual distributions for models (2.19)-(2.21) are shown in Figures 2.10(a)-2.10(d) by blue dots. Notice that all terms removed in (2.18)-(2.20) are those associated with molecular diffusion. It would be tempting at this point to conclude that the model for the Π_{xx} correlation should not include a molecular diffusion term. However, adding the molecular diffusion term in the spanwise direction along with the corresponding balance error to the list of regressors:

$$\Pi_{xx} = 0.357D^T_{xy} + 0.926P_{xy} - 0.031D^T_{xx} + 0.016P_{xx} + 0.911D^T_{yy} + 0.009D^M_{zz} \\ -0.662\,\mathrm{Err}_{xy} + 0.698\,\mathrm{Err}_{xx} - 6.704\,\mathrm{Err}_{yy} - 2.227\,\mathrm{Err}_{zz}, \tag{2.22}$$

improves the residual distribution scedasticity near that channel axis (2.12(e)). Whether these improvements correspond to a physical phenomenon has yet to be confirmed by testing the model in other flow geometries.

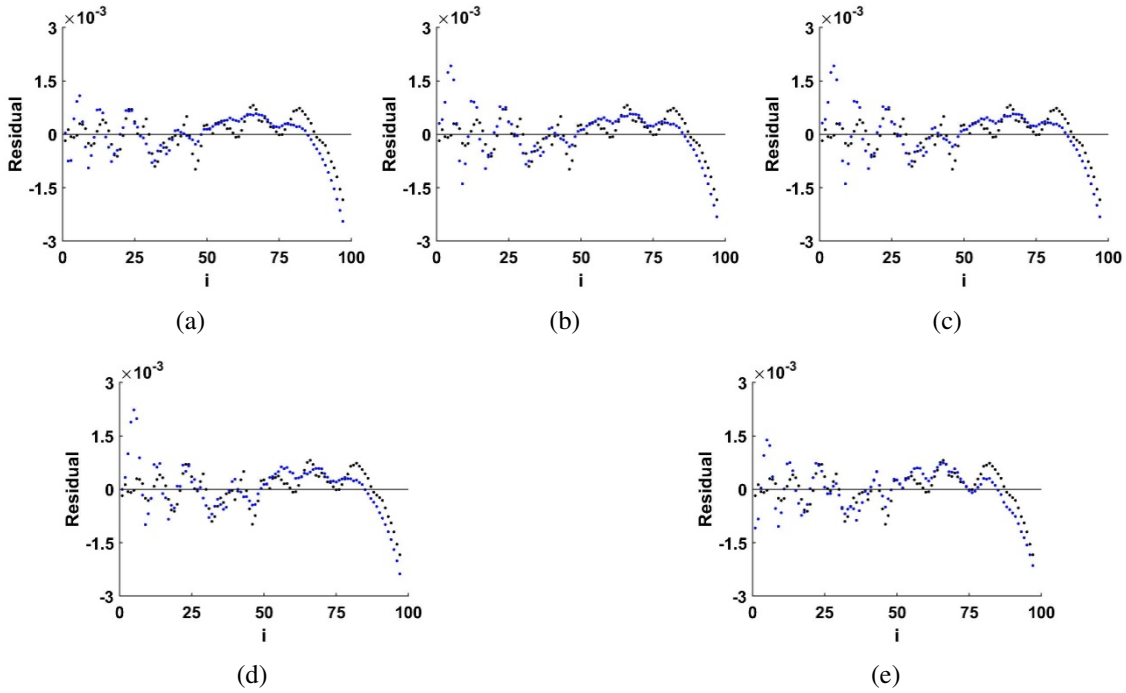


Figure 2.10: Residual distributions of MLR fits (2.18)-(2.22) for Π_{xx} (blue dots) in comparison with that of (2.17) (black dots): (a) model (2.18), (b) model (2.19), (c) model (2.20), (d) model (2.21), (e) model (2.22).

Notice that variations in the models and in the MLR fits have little effect on the residual distribution near the channel axis, similar to what was observed when modeling Π_{xy}. This is another confirmation that this error source is not relevant to modeling, but to the DNS data used.

2.4. Π_{yy} Correlations

In this section the modeling VPG correlation in the wall-normal direction, Π_{yy}, is considered. The original model used for reference is that proposed in [27]:

$$\Pi_{yy} = -0.47D^T_{xy} - 0.45P_{xy} - 0.031P_{xx} + 0.2D^M_{yy} - 1.35D^T_{yy} + 1.15D^T_{zz}, \tag{2.23}$$

with results shown in Figures 2.11(a) (black line) and 2.11(b) (black dots). Applying MLR to the same set of regressors as model (2.23) results in the following model:

$$\Pi_{yy}^{MLR1} = -0.469D_{xy}^{T} - 0.444P_{xy} - 0.030P_{xx} + 0.140D_{yy}^{M} - 1.388D_{yy}^{T} + 1.109D_{zz}^{T}. \quad (2.24)$$

The resulting coefficients for model (2.24) are very similar to those of (2.23), and do not lead to any significant difference in results, so the plots are not shown here.

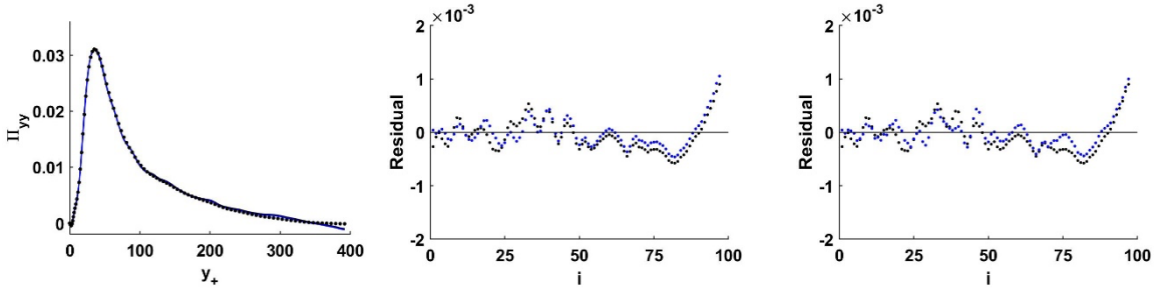
The MLR fit for Π_{yy} with the full set of regressors is as follows:

$$\Pi_{yy} = -0.150D_{xy}^{M} - 0.486D_{xy}^{T} - 0.447P_{xy} - 0.008D_{xx}^{M} - 0.014D_{xx}^{T} - 0.032P_{xx} + 0.358D_{yy}^{M} - 1.468D_{yy}^{T} + 0.021D_{zz}^{M} + 1.228D_{zz}^{T}. \quad (2.25)$$

The results for model (2.25) are shown in Figure (2.25) are shown by a blue line and blue dots. Comparing the residual distribution of (2.25) with that of (2.23) shows that the increase in model complexity is not beneficial to the model as the residual magnitude and scedasticity remain relatively unchanged. Now, adding balance errors to the list of regressors results in the following fit:

$$\Pi_{yy} = -0.133D_{xy}^{M} - 0.475D_{xy}^{T} - 0.426P_{xy} - 0.007D_{xx}^{M} - 0.010D_{xx}^{T} - 0.027P_{xx} + 0.329D_{yy}^{M} - 1.464D_{yy}^{T} + 0.017D_{zz}^{M} + 1.255D_{zz}^{T} + 0.193\,\mathrm{Err}_{xy} - 0.400\,\mathrm{Err}_{xx} + 1.521\,\mathrm{Err}_{yy} + 0.217\,\mathrm{Err}_{zz}\,. \quad (2.26)$$

This increase in complexity also does little to improve the results (Figure 2.11(c)).



(a) (b) (c)

Figure 2.11: Comparisons of (a) model (2.23) with the MLR fits (2.25) and (2.26) for Π_{yy} with DNS data as black dots, (b) residual distributions for (2.23) and (2.25), (c) residual distributions for (2.23) and (2.26). Lines/dots: black for (2.23) and blue for (2.25) and (2.26).

The set of regressors in the MLR fit for Π_{yy} can be clarified following the same proce-

dure suggested in the previous section, that is, by recognizing that in (2.26), there are small terms and some of them have coefficients larger than one. Such terms are removed starting from those in the spanwise (zz) direction. In the fits below, the terms to be removed next are marked in red:

$$\Pi_{yy} = -0.002D^M_{xy} - 0.391D^T_{xy} - 0.494P_{xy} - 0.001D^M_{xx} - 0.010D^T_{xx} - 0.058P_{xx} + 0.333D^M_{yy} - 1.076D^T_{yy} + 0.694\,\mathrm{Err}_{xy} - 0.121\,\mathrm{Err}_{xx} + 1.209\,\mathrm{Err}_{yy}, \tag{2.27}$$

$$\Pi_{yy} = -0.390D^T_{xy} - 0.494P_{xy} \color{red}{-0.001D^M_{xx}} - 0.010D^T_{xx} - 0.058P_{xx} + 0.326D^M_{yy} - 1.076D^T_{yy} + 0.688\,\mathrm{Err}_{xy} - 0.121\,\mathrm{Err}_{xx} + 1.208\,\mathrm{Err}_{yy}, \tag{2.28}$$

$$\Pi_{yy} = -0.387D^T_{xy} - 0.498P_{xy} \color{red}{-0.012D^T_{xx}} - 0.059P_{xx} - 0.193D^M_{yy} - 1.087D^T_{yy} + 0.629\,\mathrm{Err}_{xy} - 0.120\,\mathrm{Err}_{xx} + 1.2087\,\mathrm{Err}_{yy}, \tag{2.29}$$

$$\Pi_{yy} = -0.343D^T_{xy} - 0.482P_{xy} - 0.054P_{xx} - 0.340D^M_{yy} - 1.036D^T_{yy} + 1.499\,\mathrm{Err}_{xy} \color{red}{-0.008\,\mathrm{Err}_{xx}} + 1.946\,\mathrm{Err}_{yy}, \tag{2.30}$$

$$\Pi_{yy} = -0.343D^T_{xy} - 0.482P_{xy} - 0.054P_{xx} \color{red}{-0.339D^M_{yy}} - 1.036D^T_{yy} + 1.505\,\mathrm{Err}_{xy} + 1.934\,\mathrm{Err}_{yy}, \tag{2.31}$$

$$\Pi_{yy} = -0.299D^T_{xy} - 0.471P_{xy} - 0.052P_{xx} - 1.040D^T_{yy} + 2.655\,\mathrm{Err}_{xy} + 3.240\,\mathrm{Err}_{yy}\,. \tag{2.32}$$

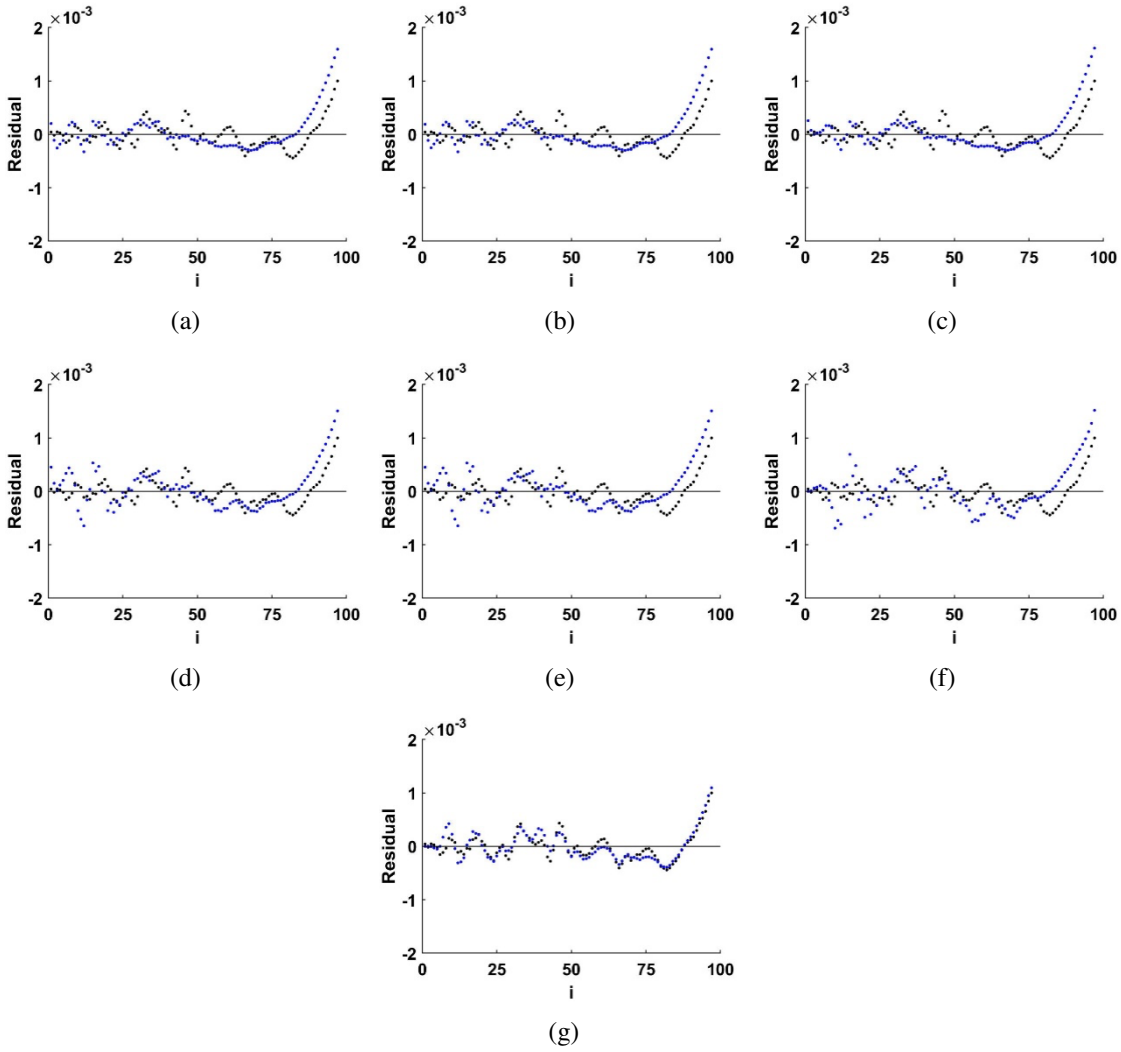


Figure 2.12: Residual distributions of MLR fits (2.27)-(2.32) for Π_{yy} (blue dots) in comparison with that of (2.26) (black dots): (a) model (2.27), (b) model (2.28), (c) model (2.29), (d) model (2.30), (e) model (2.31), (f) model (2.32), (g) model (2.33).

Similar to the previous section, balance error Err_{xx} is removed from the set of regressors since balance errors should only be included in a model if small terms, such as molecular and turbulent diffusions, are among the regressors. In (2.30), the only term from the Reynolds stress transport equation in the streamwise direction (xx) is the production, P_{xx}, which is large in comparison to Err_{xx}.

Figure 2.12(g) shows the results of adding the turbulent diffusion in the spanwise direc-

tion (along with the relevant balance error) to the fit:

$$\begin{aligned}\Pi_{yy} = -0.448D^T_{xy} - 0.448P_{xy} - 0.033P_{xx} - 1.326D^T_{yy} + 0.903D^T_{zz} \\ 0.377\,\mathrm{Err}_{xy} + 0.461\,\mathrm{Err}_{yy} + 0.061\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.33}$$

This step helps to reduce the magnitude of the residuals and produces a residual distribution almost identical to that of (2.26) without the additional complexity. The coefficients in model (2.33) all have magnitude less than one except for the turbulent diffusion in the wall-normal direction, D^T_{yy}. This is something that should be explored in future studies.

Model (2.33) has no contribution from molecular diffusion in the plane of flow development, but it does have a contribution from the turbulent diffusion in the spanwise direction, similar to Π_{xx}. This is in contrast to the fit for Π_{xy} where both molecular and turbulent diffusions were found to be of importance.

2.5. Π_{zz} Correlations

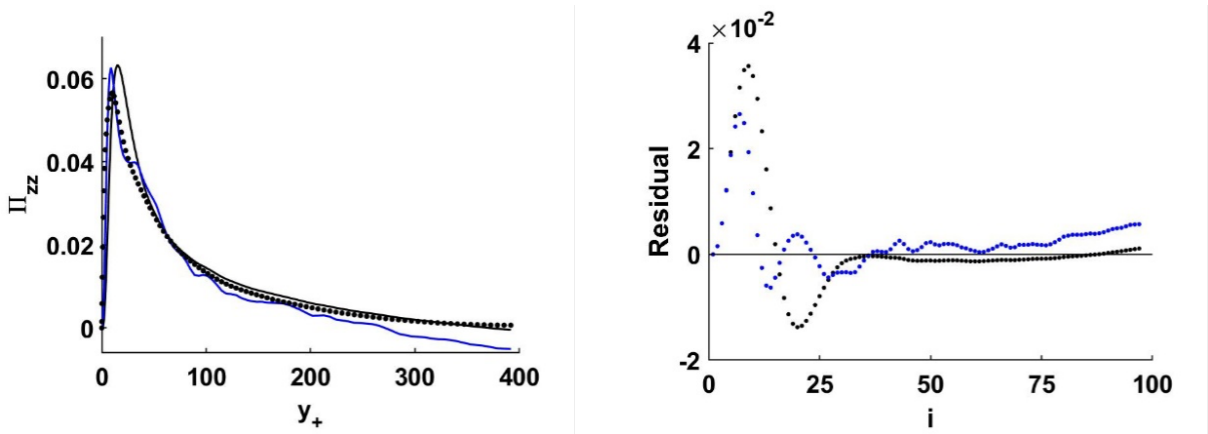
In this section, modeling the VPG correlation in the spanwise direction, Π_{zz}, is considered. The initial model is the three-regressor model found in [23]:

$$\Pi_{zz} = -0.500D^T_{zz} + 0.025P_{xx} - 0.550P_{xy}. \tag{2.34}$$

The plot and residual distribution for this model are shown in Figure 2.13 as black lines and black dots respectively. Now, applying MLR to the same set of regressors as model (2.34) results in the following model:

$$\Pi_{zz} = -5.402D^T_{zz} - 0.043P_{xx} - 0.564P_{xy}. \tag{2.35}$$

The plot and residual distribution for this model are shown in Figure 2.13 as blue lines and blue dots respectively.



(a) (b)

Figure 2.13: Comparison of models (2.34) and (2.34): (a) black lines: plot of model (2.34), blue lines: plot of model (2.35), DNS data: black dots; (b) black dots: residual distribution for model (2.34), blue dots: residual distribution for model (2.35).

In model (2.35), the coefficient for turbulent diffusion D^T_{zz}, -5.402, has a magnitude greater than one and is thus unphysical. In an attempt to mitigate this, the balance error term Err_{zz} was added to the regressors list but had an adverse effect, since the fit was worse (not shown here).

$$\Pi_{zz} = -0.628P_{xy} - 0.071P_{xx} - 5.384D^T_{zz} - 6.449\,\mathrm{Err}_{zz}\,. \tag{2.36}$$

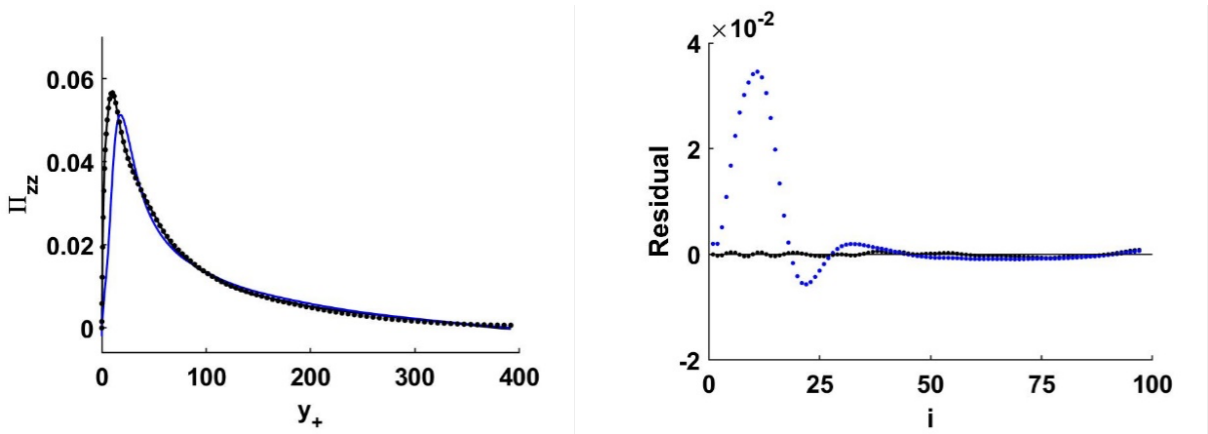
Adding the Err_{xy} term to the regressors was not considered because the only term in that direction is the production term, P_{xy} (which is large in magnitude).

In [27] another model for this correlation with more complexity was considered:

$$\Pi_{zz} = -1.55D^M_{xy} - 0.5D^T_{xy} - 0.54P_{xy} + 0.045D^M_{xx} - 0.05D^T_{xx} + 0.015P_{xx} - 0.07D^M_{zz} - 2.6D^T_{zz}. \tag{2.37}$$

Applying MLR to the full set of regressors results in the following model:

$$\begin{aligned}\Pi_{zz} = {} & -0.030D^M_{xy} - 0.517D^T_{xy} - 0.313P_{xy} + 0.132D^M_{xx} + 0.112D^T_{xx} + 0.118P_{xx} - \\ & 0.511D^M_{yy} - 0.051D^T_{yy} - 0.333D^M_{zz} + 0.529D^T_{zz}.\end{aligned} \tag{2.38}$$



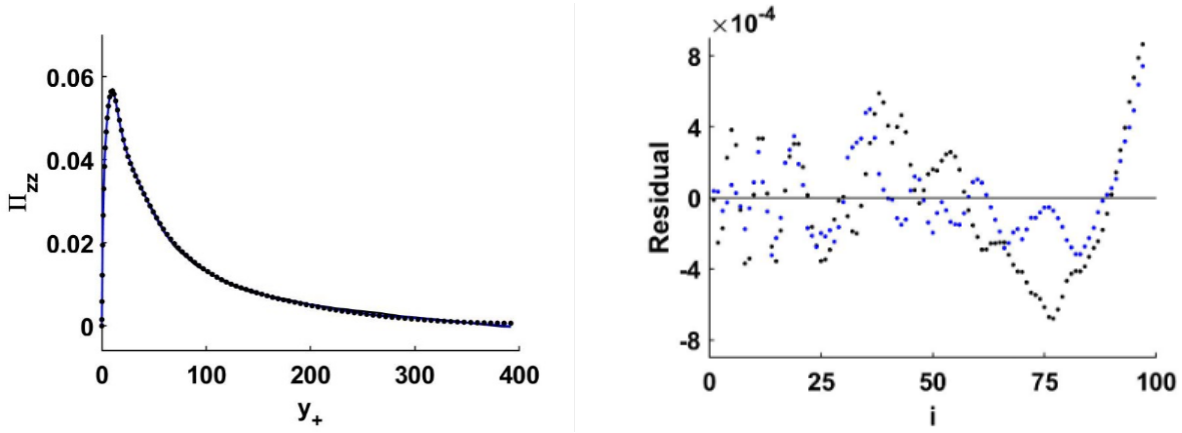
(a) (b)

Figure 2.14: Comparison of models (2.36) and (2.37): (a) black lines: plot of model (2.37), blue lines: plot of model (2.36), DNS data: black dots; (b) black dots: residual distribution for model (2.37), blue dots: residual distribution for model (2.36).

The results shown in Figure 2.14 indicate there was a significant improvement in the fit from model 2.37 when MLR is applied to the full set of regressors, as the residual magnitude decreases significantly. This improvement shows not only the benefit of applying MLR to the model, but also the importance of the terms in the wall-normal direction (yy) that are missing in model 2.37.

Adding relevant errors to the regressors list results in the following model:

$$\Pi_{zz} = 0.229D^M_{xy} - 0.441D^T_{xy} - 0.267P_{xy} + 0.151D^M_{xx} + 0.144D^T_{xx} + 0.135P_{xx} - 0.997D^M_{yy} + 0.047D^T_{yy} - 0.384D^M_{zz} - 0.741D^T_{zz} + 0.078\,\mathrm{Err}_{xy} - 0.582\,\mathrm{Err}_{xx} - 0.151\,\mathrm{Err}_{yy} + 1.741\,\mathrm{Err}_{zz}\,. \tag{2.39}$$



(a) (b)

Figure 2.15: Comparison of models (2.38) and (2.39): (a) black lines: plot of model (2.38), blue lines: plot of model (2.39), DNS data: black dots; (b) black dots: residual distribution for model (2.38), blue dots: residual distribution for model (2.39).

Note all terms representing physical processes in this model only have coefficients with magnitudes less than one. From Figure 2.15, it is apparent that adding the balance errors slightly improved the residual distribution towards the channel axis (large y_+ values), but not as much towards the channel wall.

Now, we are interested in removing extraneous regressors that have little impact on the fit residual distribution for the sake of simplifying the model. The present approach consists of removing one regressor term at a time and comparing the newly generated residual distribution with that generated using the original regressors list. At each step, every term in the regressor list is removed individually, then replaced when another term is being considered for removal. A term is removed when its removal has the least impact on the quality of the residual distribution in comparison to other terms. The first iteration in this process results in the following MLR model, where D^T_{yy} is removed from (2.39):

$$\begin{aligned}\Pi_{zz} = {} & 0.170D^M_{xy} - 0.454D^T_{xy} - 0.267P_{xy} + 0.147D^M_{xx} + 0.139D^T_{xx} + 0.135P_{xx} - 0.891D^M_{yy} \\ & -0.374D^M_{zz} - 0.629D^T_{zz} + 0.003\,\mathrm{Err}_{xy} - 0.599\,\mathrm{Err}_{xx} - 0.234\,\mathrm{Err}_{yy} + 1.683\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.40}$$

As evidenced by Figure 2.16, the removal of this term has very little impact on the residual distribution, and most of the discrepancy comes near the channel wall.

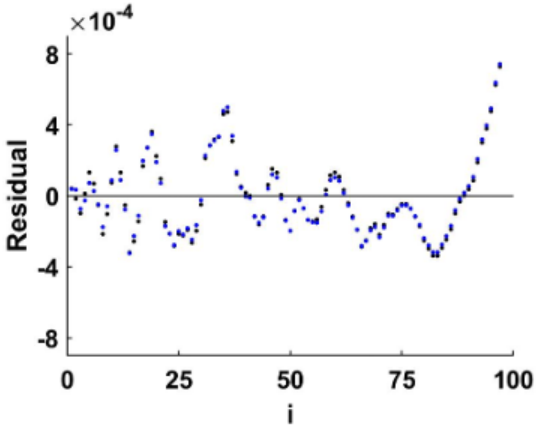


Figure 2.16: Comparison of residual distributions: black - model (2.40), blue - model (2.39)

Following the same procedure with (2.40) as the original fit, it was determined that removing D^M_{xy} had the least effect on the residual distribution. The resulting MLR model is as follows:

$$\begin{aligned}\Pi_{zz} = &-0.464D^T_{xy} - 0.277P_{xy} + 0.139D^M_{xx} + 0.128D^T_{xx} + 0.131P_{xx} - 0.958D^M_{yy} \\ &-0.349D^M_{zz} - 0.658D^T_{zz} + 0.164\,\mathrm{Err}_{xy} - 0.569\,\mathrm{Err}_{xx} - 0.156\,\mathrm{Err}_{yy} + 1.530\,\mathrm{Err}_{zz}\,.\end{aligned} \tag{2.41}$$

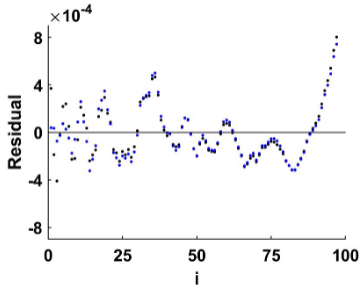


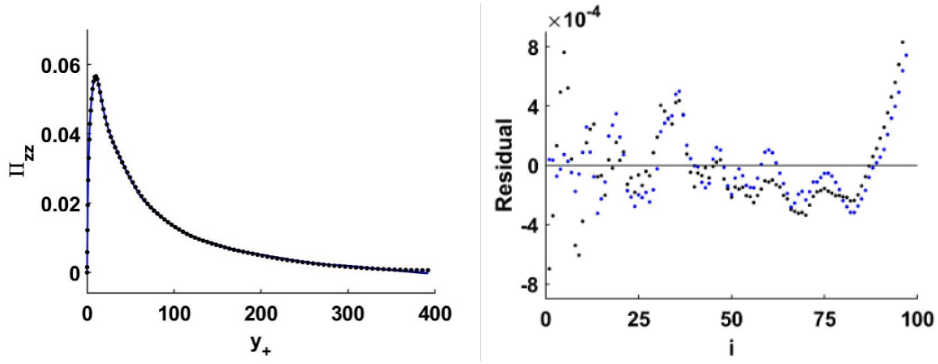
Figure 2.17: Comparison of residual distributions: black - model (2.41), blue - model (2.39)

Figure 2.17 shows that the removal of D^M_{xy} from (2.40) results in a small discrepancy in the model quality with respect to model (2.39).

Further simplification indicated that two more terms, namely D^M_{yy} and Err_{yy}, could be removed from the regressors list without sacrificing fit quality. With this step, the fit no longer depends on terms in the wall-normal direction (yy). Removal of any additional terms resulted in a reduction of fit quality, so the final model for Π_{zz} in this section is as

follows:

$$\Pi_{zz} = -0.426D^T_{xy} - 0.317P_{xy} + 0.131D^M_{xx} + 0.115D^T_{xx} + 0.114P_{xx} - 0.330D^M_{zz} \\ -0.752D^T_{zz} + 0.558\,\mathrm{Err}_{xy} - 0.504\,\mathrm{Err}_{xx} + 1.040\,\mathrm{Err}_{zz}\,. \tag{2.42}$$



(a) (b)

Figure 2.18: Comparison of models (2.42) and (2.39): (a) black lines: plot of model (2.42), blue lines: plot of model (2.39), DNS data: black dots; (b) black dots: residual distribution for model (2.42), blue dots: residual distribution for model (2.39).

The final model has deficiencies near the wall in comparison to model (2.39), evidenced by the increase in residual magnitude, but overall has similar scedasticity and remains in the same order of magnitude.

Part II

Generalization of Model Coefficients

3. METHODS

The results found in Part I indicate that there is great potential in the application of the MLR procedure to developing linear models that describe the VPG correlations using terms in the Reynolds stress transport equations. However, the models developed in the previous section were developed for a turbulent flow without complex conditions such as flow separation and adverse pressure gradients. Hence, the goal in the present chapter is to develop a "universal" model for each VPG correlation that is linear in nature and whose coefficients do not vary when complex conditions arise in a flow.

The DNS data from [31] provides flow parameters at progressively more complex flow states in "time" increments $A_{22}t$. The variable $A_{22}t$ acts as an indicator of the strain value of the channel, and an increasing strain value corresponds to an increasing adverse pressure gradient, with flow separation occurring at $A_{22}t = 0.675$. Here, we analyze 6 specific instances of the flow at strain values of $A_{22}t = 0\,(\text{unstrained}), 0.020, 0.281, 0.365$, 0.675, and 0.772.

We now seek a single model that performs well at every strain value without changing model coefficients that provides an appropriately small residual distribution magnitude and acceptable scedasticity, for a given VPG correlation. Intuitively this avenue of modeling the VPG correlations seems far-fetched because of the rigid and simple nature of the linear models, and there is no guarantee that a single model will produce adequate results at all strain values.

In this chapter, the Least Sum of Least Squares Method is introduced and applied to develop universal linear models for each VPG correlation. The accuracy and viability of applying this method is assessed by analyzing the residuals of the models with regards to the data.

3.1. Least Sum of Least Squares (LSLS)

In the previous section, the least squares method was applied to minimize the sum of the squared errors between a VPG model and a *single* data set. In this section, we are seeking to minimize the sum of the squared errors over *multiple* data sets simultaneously to develop a universal model for a given VPG correlation.

This problem for an arbitrary observed variable, Θ, can be written symbolically as follows:

$$\underset{a,b,c,\cdots}{\text{minimize}}\ S(a,b,c,\cdots) = \sum_{j=1}^{D}\sum_{i=1}^{N}(\Theta_i^{(j)} - ax_{1i}^{(j)} - bx_{2i}^{(j)} - cx_{3i}^{(j)} - \cdots)^2, \tag{3.1}$$

where: S = sum of the squared errors over D data sets of the model with coefficients $a, b, c, \cdots$ and regressors x

D = total number of data sets

N = number of entries in each vector, i.e. length of the vectors

$a, b, c, \cdots$ = constant coefficients corresponding to each regressor in the model

$x_{ri}^{(j)}$ = ith entry of rth regressor from jth data set

$\Theta_i^{(j)}$ = ith entry of observed variable vector from jth data set.

From elementary calculus, assuming S is differentiable, the solution to this problem occurs when:

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = \frac{\partial S}{\partial c} = \cdots = 0. \tag{3.2}$$

Differentiating (3.1) according to (3.2) results in a system of linear equations:

$$\begin{aligned}
\frac{\partial S}{\partial a} &= \sum_{j=1}^{D}\sum_{i=1}^{N} 2(\Theta_i^{(j)} - ax_{1i}^{(j)} - bx_{2i}^{(j)} - cx_{3i}^{(j)} - \cdots)x_{1i}^{(j)} = 0 \\
\frac{\partial S}{\partial b} &= \sum_{j=1}^{D}\sum_{i=1}^{N} 2(\Theta_i^{(j)} - ax_{1i}^{(j)} - bx_{2i}^{(j)} - cx_{3i}^{(j)} - \cdots)x_{2i}^{(j)} = 0 \\
\frac{\partial S}{\partial c} &= \sum_{j=1}^{D}\sum_{i=1}^{N} 2(\Theta_i^{(j)} - ax_{1i}^{(j)} - bx_{2i}^{(j)} - cx_{3i}^{(j)} - \cdots)x_{3i}^{(j)} = 0 \\
&\vdots
\end{aligned} \tag{3.3}$$

Now, (3.3) can be rewritten as:

$$\begin{aligned}
\frac{\partial S}{\partial a} &= \sum_{j=1}^{D}\sum_{i=1}^{N}\Theta_i^{(j)}x_{1i}^{(j)} - a\sum_{j=1}^{D}\sum_{i=1}^{N}x_{1i}^{(j)}x_{1i}^{(j)} - b\sum_{j=1}^{D}\sum_{i=1}^{N}x_{2i}^{(j)}x_{1i}^{(j)} - c\sum_{j=1}^{D}\sum_{i=1}^{N}x_{3i}^{(j)}x_{1i}^{(j)} - \cdots = 0 \\
\frac{\partial S}{\partial b} &= \sum_{j=1}^{D}\sum_{i=1}^{N}\Theta_i^{(j)}x_{2i}^{(j)} - a\sum_{j=1}^{D}\sum_{i=1}^{N}x_{1i}^{(j)}x_{2i}^{(j)} - b\sum_{j=1}^{D}\sum_{i=1}^{N}x_{2i}^{(j)}x_{2i}^{(j)} - c\sum_{j=1}^{D}\sum_{i=1}^{N}x_{3i}^{(j)}x_{2i}^{(j)} - \cdots = 0 \\
\frac{\partial S}{\partial c} &= \sum_{j=1}^{D}\sum_{i=1}^{N}\Theta_i^{(j)}x_{3i}^{(j)} - a\sum_{j=1}^{D}\sum_{i=1}^{N}x_{1i}^{(j)}x_{3i}^{(j)} - b\sum_{j=1}^{D}\sum_{i=1}^{N}x_{2i}^{(j)}x_{3i}^{(j)} - c\sum_{j=1}^{D}\sum_{i=1}^{N}x_{3i}^{(j)}x_{3i}^{(j)} - \cdots = 0 \\
&\vdots
\end{aligned} \tag{3.4}$$

To simplify the visual appearance of (3.4), let us define two terms as follows:

$$Z_{rs} = \sum_{j=1}^{D}\sum_{i=1}^{N}x_{ri}^{(j)}x_{si}^{(j)} = \sum_{j=1}^{D}x_r^{(j)}\cdot x_s^{(j)}$$

and

$$Y_w = \sum_{j=1}^{D}\sum_{i=1}^{N}\Theta_i^{(j)}x_{wi}^{(j)} = \sum_{j=1}^{D}\Theta^{(j)}\cdot x_w^{(j)}.$$

Note that Z_{rs} and Y_w are both scalars and $Z_{rs} = Z_{sr}$.

Then, (3.4) can be rewritten as:

$$\begin{aligned}
\frac{\partial S}{\partial a} &= Y_1 - aZ_{11} - bZ_{21} - cZ_{31} - \cdots = 0 \\
\frac{\partial S}{\partial b} &= Y_2 - aZ_{12} - bZ_{22} - cZ_{32} - \cdots = 0 \\
\frac{\partial S}{\partial c} &= Y_3 - aZ_{13} - bZ_{23} - cZ_{33} - \cdots = 0 \\
&\vdots
\end{aligned} \tag{3.5}$$

Or more compactly as:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{Z}, \tag{3.6}$$

where: $\boldsymbol{Y} = \begin{bmatrix} Y_1 & Y_2 & Y_3 & \cdots \end{bmatrix}$

$$\boldsymbol{Z} = \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots \\ Z_{21} & Z_{22} & Z_{23} & \cdots \\ Z_{31} & Z_{32} & Z_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$\boldsymbol{A} = \begin{bmatrix} a & b & c & \cdots \end{bmatrix}$.

Thus, $\boldsymbol{A}$ can be obtained as follows:

$$\boldsymbol{A} = \boldsymbol{Y}\boldsymbol{Z}^{-1}. \tag{3.7}$$

This "least sum of least squares" procedure minimizes the sum of the sum of least squares for a model over multiple data sets, thus generating one model which provides the best linear fit for multiple data sets simultaneously.

The calculations used for the model development as described in this section were completed using MATLAB [34].

3.2. Regressor List Reduction Procedure

Increasing the list of regressors used in a model of a VPG correlation often results in diminishing returns, such that adding more regressors does not improve the fit quality so the increase in model complexity is not justified. In this study, a procedure informed by the Mean Squared Error (MSE) for systematically removing regressors from a list of regressors with little impact on the fit quality is used. In this procedure, only budget terms from the Reynolds stress transport equations, referred to as "primary regressors" in this work, will be considered for removal. Balance error terms, also referred to as "secondary regressors" in this study, are not considered for removal in the reduction procedure. Relevant balance error terms are added after the list of primary regressors is established.

MSE between two vectors, $\boldsymbol{a}$ and $\boldsymbol{b}$, each of length n is defined as:

$$\mathrm{MSE}(\boldsymbol{a}, \boldsymbol{b}) \triangleq \frac{1}{N} \sum_{i=1}^{n} (a_i - b_i)^2 . \tag{3.8}$$

Note that MSE takes the square of each error, meaning it heavily weighs large errors, so

poor agreement between the two vectors will be more visible.

In the present application, the regressor list reduction procedure should be informed by errors in models from multiple data sets simultaneously to reduce the errors in the universal model. Therefore, we introduce a parameter, $\Omega(s)$, defined as the sum of the MSEs resulting from removing regressor s from the list of regressors over all data sets. Symbolically, this can be expressed as:

$$\Omega(s) \triangleq \sum_{j=1}^{D} \mathrm{MSE}\left(\Theta^{(j)}, \hat{y}^{(j)}_{R\sim s}\right), \tag{3.9}$$

where: s = parameter being considered for removal from regressor list
D = total number of data sets
$\mathrm{MSE}(\mathbf{a}, \mathbf{b})$ = mean squared error between vectors a and b
$\Theta^{(j)}$ = observed variable vector from jth data set
R = list of regressors
$\hat{y}^{(j)}_{R\sim s}$ = MLR model obtained with the regressor list R where regressor s is removed, using data set j.

Note that all MLR models, $\hat{y}^{(j)}_{R\sim s}$, have different coefficients as they are fit individually to $\Theta^{(j)}$ vectors from different data sets. The parameter $\Omega(s)$ is beneficial in this application because it takes into account model deficiencies from all data sets simultaneously, giving increased attention to outliers. In the future, it will also be interesting to use the absolute value, or the "Manhattan distance", between the model and prediction as a metric instead of the MSE, but this is not explored in this study.

Then, the regressor, r, in the regressor list, R, that minimizes $\Omega(s)$:

$$r = \underset{s\in R}{\operatorname{argmin}}\, \Omega(s). \tag{3.10}$$

is removed from the list.

Now, a rigorous procedure must be established to determine how many regressors should be removed from a list. For this reason, the LSLS loss metric, $\mathcal{L}$, is defined as follows:

$$\mathcal{L} = \frac{\sum_{j=1}^{D}\sum_{i=1}^{N}(\Theta_i^{(j)} - ax_{1i}^{(j)} - bx_{2i}^{(j)} - cx_{3i}^{(j)} - \cdots)^2}{\sum_{j=1}^{D}\sum_{i=1}^{N}\left(\Theta_i^{(j)}\right)^2}, \tag{3.11}$$

where: D = total number of data sets

N = number of entries in each vector, i.e. length of the vectors

$a, b, c, \cdots$ = constant coefficients corresponding to each regressor in the LSLS model

$x_{ri}^{(j)}$ = ith entry of rth regressor from jth data set

$\Theta_i^{(j)}$ = ith entry of observed variable vector from jth data set.

This loss metric quantifies the agreement between the LSLS model and the data for all data sets, so as the model complexity is reduced through the reduction of the regressors list it is expected that $\mathcal{L}$ increases. When all regressors are removed from the model, the value of $\mathcal{L}$ is 1. Note that the parameter $\mathcal{L}$ is based on the agreement between the LSLS model and the data, whereas the parameter $\Omega(s)$ is based on the agreement between individual MLR fits and the data. The stepwise reduction procedure is stopped once $\mathcal{L}$ is large (as defined in the Results section), which implies that the model no longer describes the data appropriately. In a later study, it will be interesting to use $\mathcal{L}$ as the regressor reduction parameter instead of $\Omega(s)$.

The calculations used for the model development as described in this section were completed using MATLAB [34].

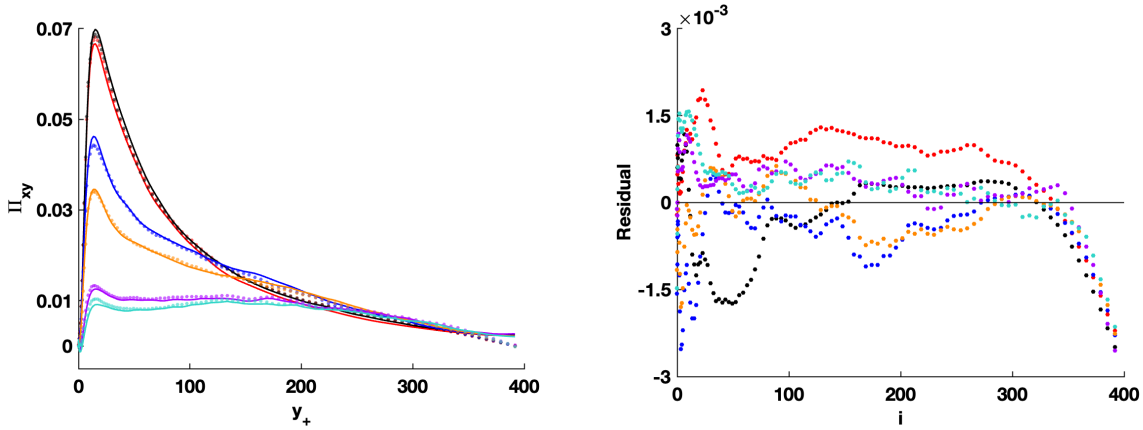
4. RESULTS

4.1. LSLS Applied to Modeling Π_{xy}

This section describes the application of the LSLS method to modelling the Π_{xy} VPG correlation. The full set of regressors was determined in a preliminary study and includes other VPG correlations as well as turbulent and molecular diffusion terms in the RANS equations for all Reynolds stresses, for a total of 11 regressors.

The initial model for Π_{xy} with the full set of primary regressors obtained using the LSLS procedure described in the previous section takes the following form:

$$\begin{aligned}\Pi_{xy} = &-0.780\Pi_{xx} - 0.162\Pi_{yy} + 0.357\Pi_{zz} - 0.239D^T_{xy} - 0.015D^T_{xx} + \\ &0.465D^T_{yy} + 0.491D^T_{zz} + 0.281D^M_{xy} - 0.014D^M_{xx} - 0.069D^M_{yy} + 0.029D^M_{zz}.\end{aligned} \tag{4.1}$$



(a) (b)

Figure 4.1: Results from model (4.1): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Figure 4.1 demonstrates that this LSLS model is indeed a good fit for the VPG correlation Π_{xy} at various strain values. However, the residual distributions are heteroscedastic at all considered strain values.

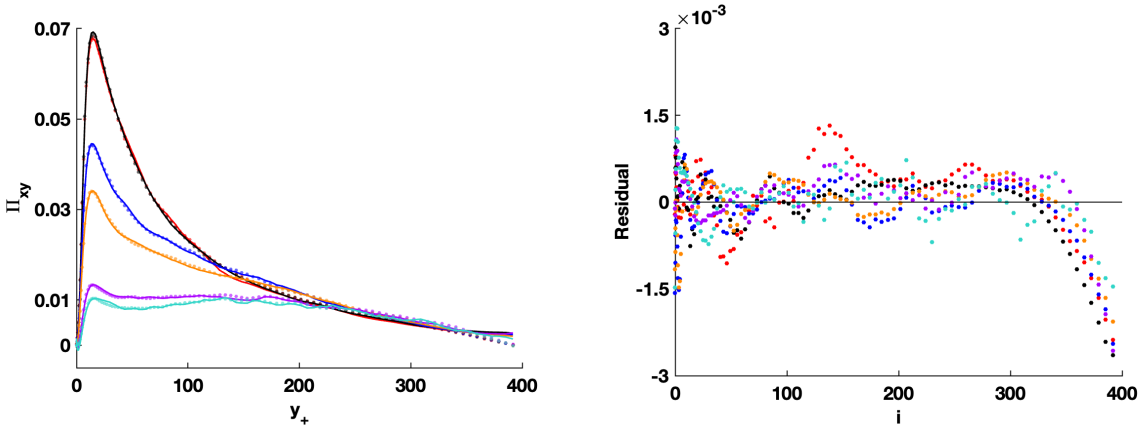
As the next step, the list of regressors was increased by adding the error terms (sec-

ondary regressors). At the same time, the coefficients from model (4.1) for the primary regressors remain unchanged. Note the coefficients for the secondary regressors are found using MLR at each strain value individually. The coefficients for these fits at each strain value are presented in Table 4.1.

Table 4.1: Values of coefficients on secondary regressors corresponding to model (4.1)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-1.628	-0.149	3.134	0.765
0.002	-1.395	-0.077	-2.081	-0.106
0.281	1.294	0.266	-0.973	0.196
0.365	0.679	0.190	-0.449	-0.310
0.675	-0.810	-0.135	0.555	-0.756
0.772	-0.912	-0.158	1.824	-1.097

The resulting fits for these models are shown in Figure 4.2. The results show that the models with the error terms added to the regressors list have improved the residual distribution, especially in the near-wall region, but also outside. The residual distributions become more homoscedastic for all strain values.



(a) (b)

Figure 4.2: Results from model (4.1) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Now, it is of interest to simplify the model as much as possible by reducing the regressors from the list. This step will be done using the stepwise reduction procedure outlined in Section 3.2. Note again that the reduction procedure is only implemented to the primary

regressors. The secondary regressors included in the model are adjusted later based on the remaining primary regressors.

The values of $\Omega(s)$ for the full set of regressors are shown in Table 4.2:

Table 4.2: $\Omega(s)$ values corresponding to regressors in model (4.1)

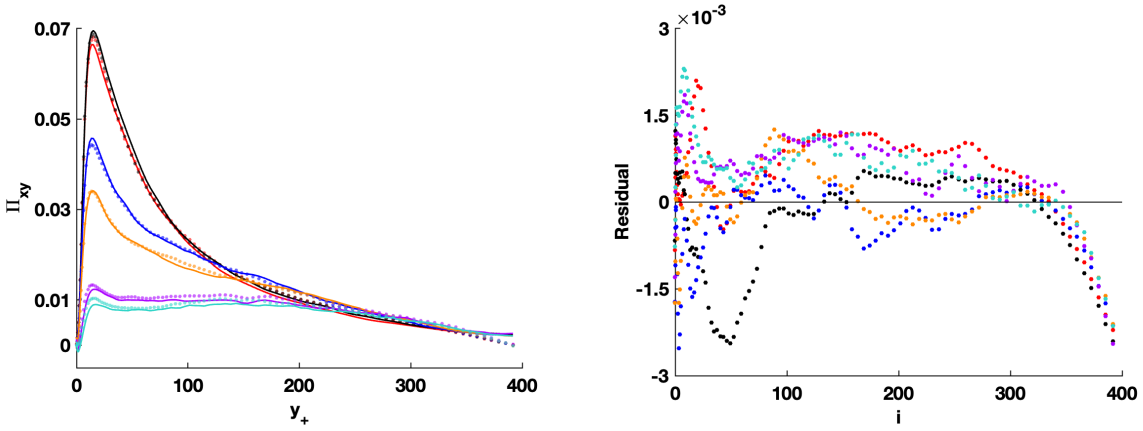
s	Π_{xx}	Π_{yy}	Π_{zz}	D^T_{xy}	D^T_{xx}	D^T_{yy}
$\Omega(s)$	1.965e-6	0.876e-6	0.942e-6	1.223e-6	0.942e-6	0.836e-6

s	D^T_{zz}	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.848e-6	0.895e-6	0.863e-6	0.959e-6	0.864e-6

From the table, the term with the smallest $\Omega(s)$ value is D^T_{yy}, meaning that the removal of this term will have the least immediate impact on the fit based on the MSE of the MLR models at each strain value. Therefore, in the next step, the LSLS procedure is used on the original set of regressors removing D^T_{yy}, to obtain the following model:

$$\begin{aligned}\Pi_{xy} = -0.898\Pi_{xx} - 0.465\Pi_{yy} + 0.372\Pi_{zz} - 0.349D^T_{xy} - 0.016D^T_{xx} + \\ 1.088D^T_{zz} + 0.342D^M_{xy} - 0.015D^M_{xx} + 0.317D^M_{yy} + 0.031D^M_{zz}.\end{aligned} \tag{4.2}$$

This procedure will be iteratively repeated until a satisfactory model is found. Results from model (4.2) are shown in Figure 4.3. The residual distribution for model (4.2) is noticeably worse than that of model (4.1), but the model still shows acceptable agreement with the DNS data (Figure 4.3(a)).



(a) (b)

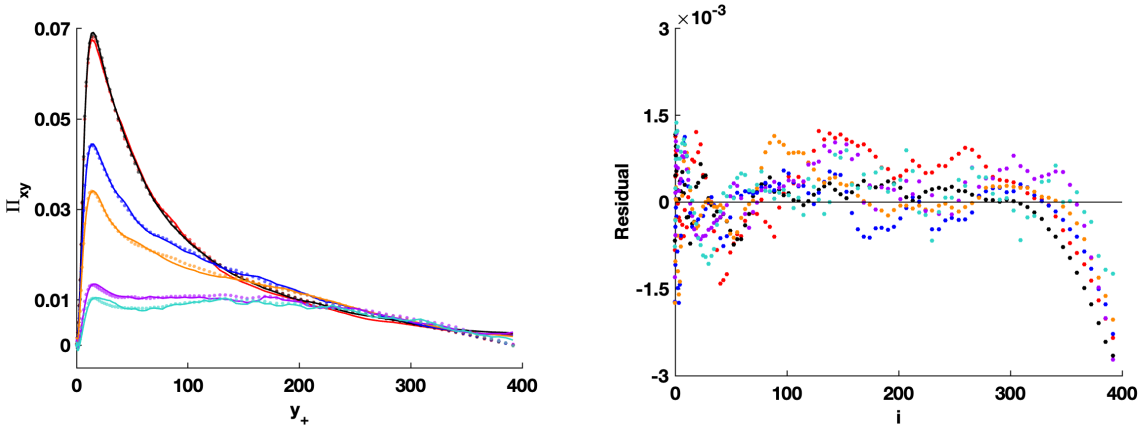
Figure 4.3: Results from model (4.2): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Similar to the previous step, the relevant balance errors were added to the list of regressors to generate the models in Table 4.3.

Table 4.3: Values of coefficients on secondary regressors corresponding to model (4.2)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-1.756	-0.400	1.832	0.841
0.002	-0.997	0.149	-2.601	-0.900
0.281	1.024	0.189	-0.290	0.221
0.365	0.513	0.098	0.125	-0.250
0.675	-0.872	-0.070	2.290	-1.407
0.772	-1.097	0.114	2.620	-1.858

The results for these models are shown in Figure 4.4. Again, the addition of the secondary regressors to the regressors list is beneficial to the fit, as the residual magnitudes and scedasticities improved for all strain values.



(a) (b)

Figure 4.4: Results from model (4.2) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Now, to reduce the model again, a similar procedure is done as before. The $\Omega(s)$ are shown in Table 4.4:

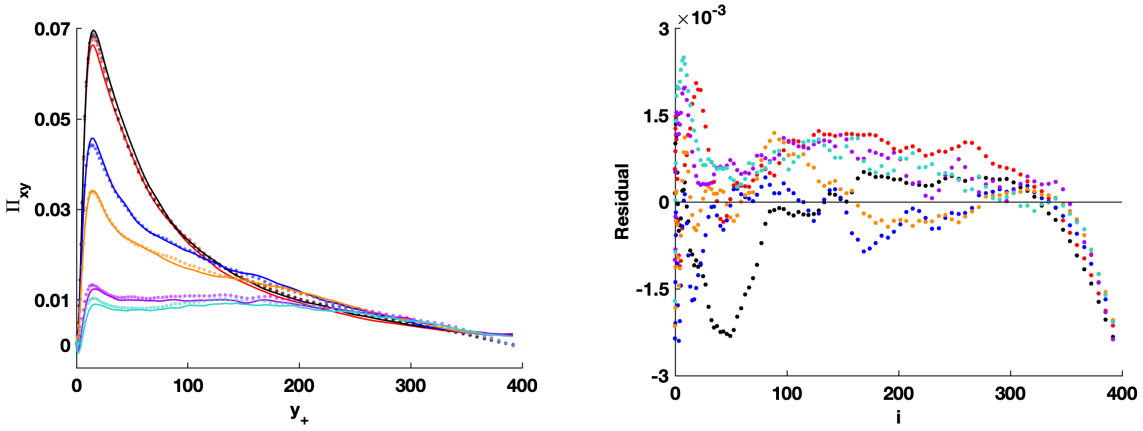
Table 4.4: $\Omega(s)$ values corresponding to regressors in model (4.2)

s	Π_{xx}	Π_{yy}	Π_{zz}	D^T_{xy}	D^T_{xx}
$\Omega(s)$	0.281e-3	0.274e-3	0.284e-3	0.290e-3	0.272e-3
s	D^T_{zz}	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.271e-3	0.270e-3	0.295e-3	0.292e-3	0.299e-3

The results in the table indicate D^M_{xy} is the next term to be removed from the regressors list. The new model generated with the updated regressors list is:

$$\begin{aligned}\Pi_{xy} = -0.934\Pi_{xx} - 0.490\Pi_{yy} + 0.327\Pi_{zz} - 0.349D^T_{xy} - \\ 0.0250D^T_{xx} + 1.027D^T_{zz} - 0.026D^M_{xx} + 0.231D^M_{yy} + 0.059D^M_{zz}.\end{aligned} \tag{4.3}$$

Results from model (4.3) are shown in Figure 4.5. Figure 4.5 indicates that this removal step has very little impact on the fit, as the result is very similar to that from model (4.2). Further, the effect is only slighty noticeable in the residual distribution (4.5(b)).



(a) (b)

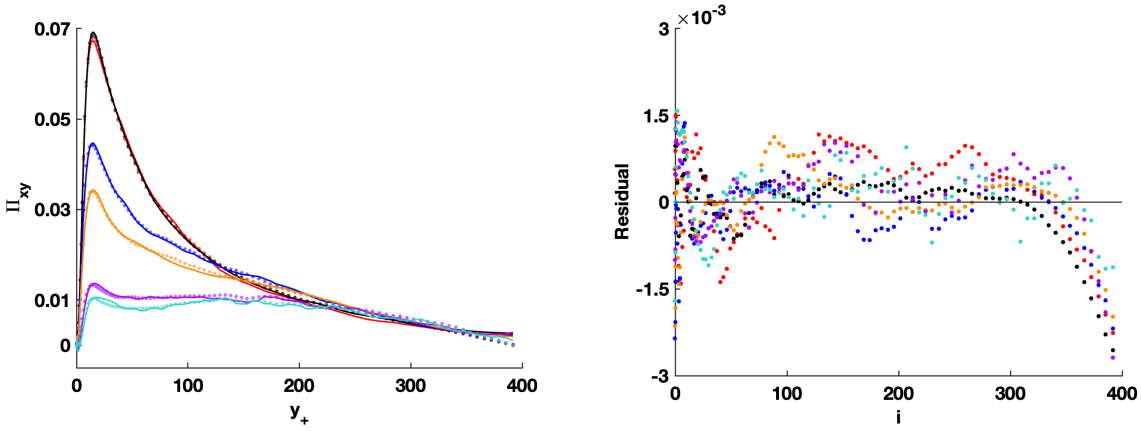
Figure 4.5: Results from model (4.3): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Now, adding the relevant secondary regressors and fitting these terms with MLR generates the models in Table 4.5:

Table 4.5: Values of coefficients on secondary regressors corresponding to model (4.3)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-1.503	-0.485	2.101	0.855
0.002	-0.585	0.191	-2.326	-1.071
0.281	0.809	0.149	-0.418	0.248
0.365	0.210	0.035	-0.053	-0.166
0.675	-1.246	-0.161	2.244	-1.339
0.772	-1.440	0.032	2.460	-1.866

The results for these models are shown in Figure 4.6.



(a) (b)

Figure 4.6: Results from model (4.3) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Model (4.3) as well as the models with errors added still agree with the DNS data, so the reduction procedure can be implemented again. The $\Omega(s)$ values at this step are shown in Table 4.6, and indicate that the regressor D^T_{zz} can be removed.

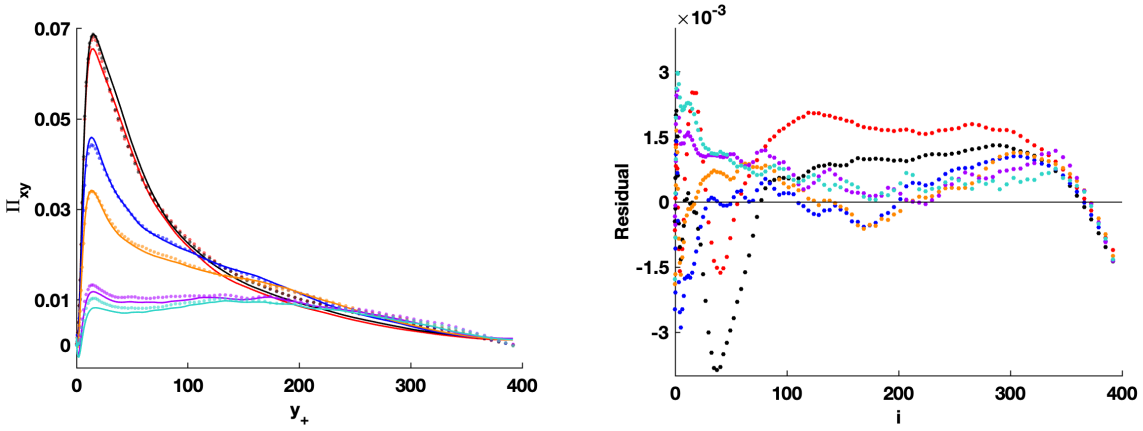
Table 4.6: $\Omega(s)$ values corresponding to regressors in model (4.3)

s	Π_{xx}	Π_{yy}	Π_{zz}	D^T_{xy}	D^T_{xx}
$\Omega(s)$	0.281e-3	0.278e-3	0.302e-3	0.311e-3	0.273e-3

s	D^T_{zz}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.271e-3	0.295e-3	0.299e-3	0.299e-3

Following the reduction procedure then implementing LSLS generates the following model, with results shown in Figure 4.7:

$$\Pi_{xy} = -0.929\Pi_{xx} - 0.414\Pi_{yy} + 0.224\Pi_{zz} - 0.214D^T_{xy} - 0.017D^T_{xx} - 0.016D^M_{xx} \\ -0.030D^M_{yy} + 0.046D^M_{zz}. \tag{4.4}$$



(a) (b)

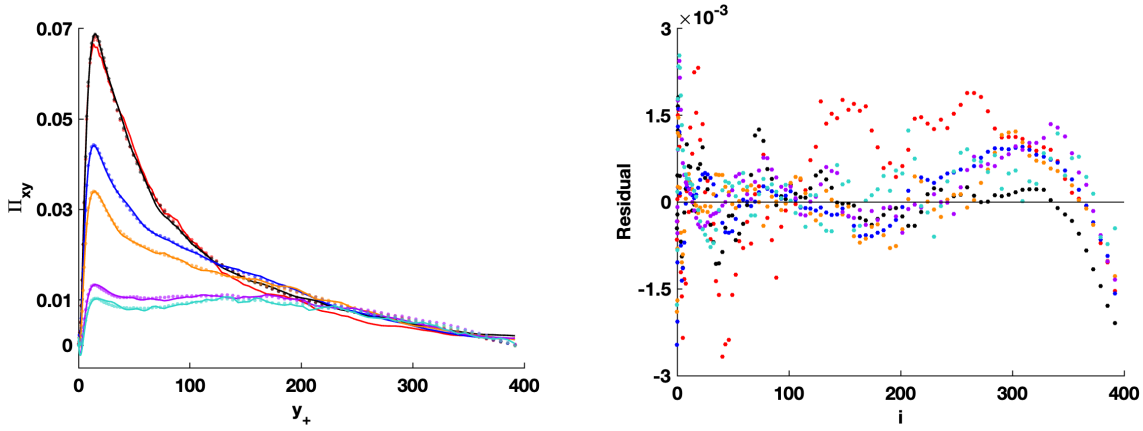
Figure 4.7: Results from model (4.4): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

From this figure we see that this removal step resulted in a noticeable increase in heteroscedasticity, but the residual magnitude stays similar (note the change in the vertical axis of the residual plots). Adding errors to the list of regressors as before results in the models shown in Table 4.7:

Table 4.7: Values of coefficients on secondary regressors corresponding to model (4.4)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-5.612	-0.628	0.451	3.702
0.002	-0.951	0.737	-5.147	-1.931
0.281	1.694	0.308	0.346	0.155
0.365	1.748	0.558	2.491	-1.939
0.675	0.944	-0.267	1.833	-0.623
0.772	-2.514	-0.494	1.052	-1.810

The results of these models (Figure 4.8) show an improvement in the residual scedasticity when the error terms are added. Interestingly, the residual magnitude reduces noticeably for terms at later strain values. This implies that the models actually perform better at conditions where the flow complexities are further developed.



(a) (b)

Figure 4.8: Results from model (4.4) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

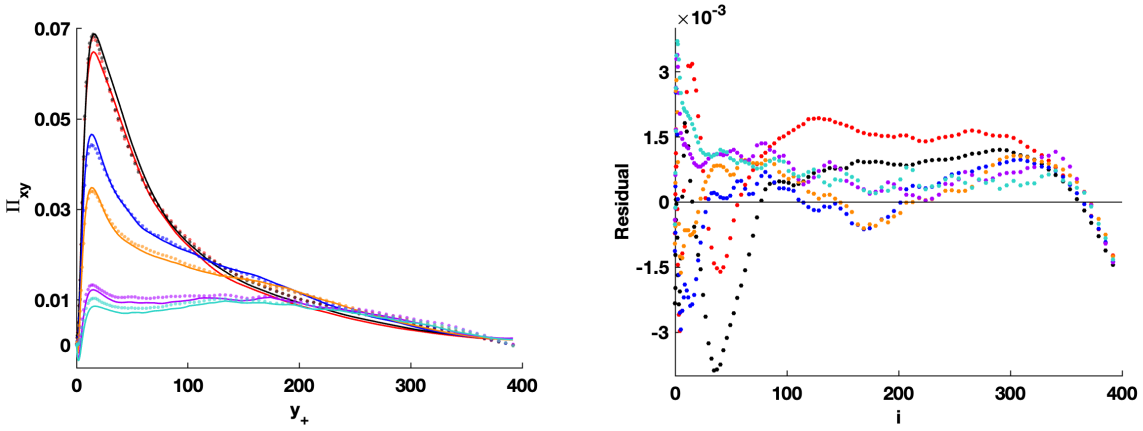
At this point in the modeling, the fits still agree well with the DNS data, as indicated by Figure 4.7, so the reduction process is continued for another iteration. The $\Omega(s)$ values for this step are shown in Table 4.8, and indicate that the regressor D^T_{xx} can be removed.

Table 4.8: $\Omega(s)$ values corresponding to regressors in model (4.4)

s	Π_{xx}	Π_{yy}	Π_{zz}	D^T_{xy}
$\Omega(s)$	0.285e-3	0.279e-3	0.313e-3	0.312e-3
s	D^T_{xx}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.273e-3	0.341e-3	0.344e-3	0.344e-3

The resulting model, with four terms removed and results shown in Figure 4.9, is as follows:

$$\begin{aligned}\Pi_{xy} = &-1.025\Pi_{xx} - 0.552\Pi_{yy} + 0.165\Pi_{zz} - 0.211D^T_{xy} - 0.005D^M_{xx} \\ &+0.070D^M_{yy} + 0.018D^M_{zz}.\end{aligned} \tag{4.5}$$



(a) (b)

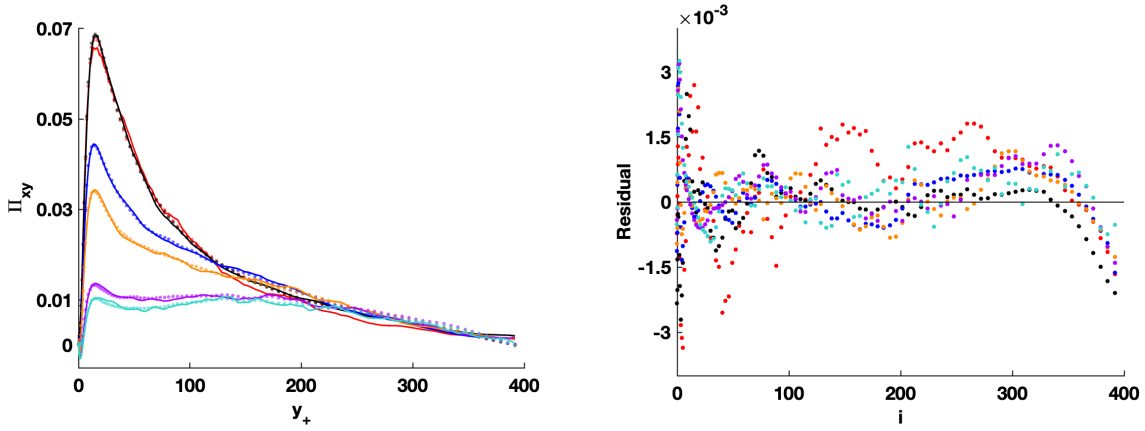
Figure 4.9: Results from model (4.5): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

At this point, the fits are noticeably different from the DNS data for all strain values. Now, adding error terms to the list of regressors results in the models shown in Table 4.9:

Table 4.9: Values of coefficients on secondary regressors corresponding to model (4.5)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-6.040	-0.793	0.101	3.760
0.002	-1.281	0.563	-4.916	-1.496
0.281	2.030	0.474	1.009	-0.367
0.365	1.974	0.652	2.603	-2.107
0.675	0.594	-0.108	1.912	-1.059
0.772	-2.314	-0.279	0.998	-1.974

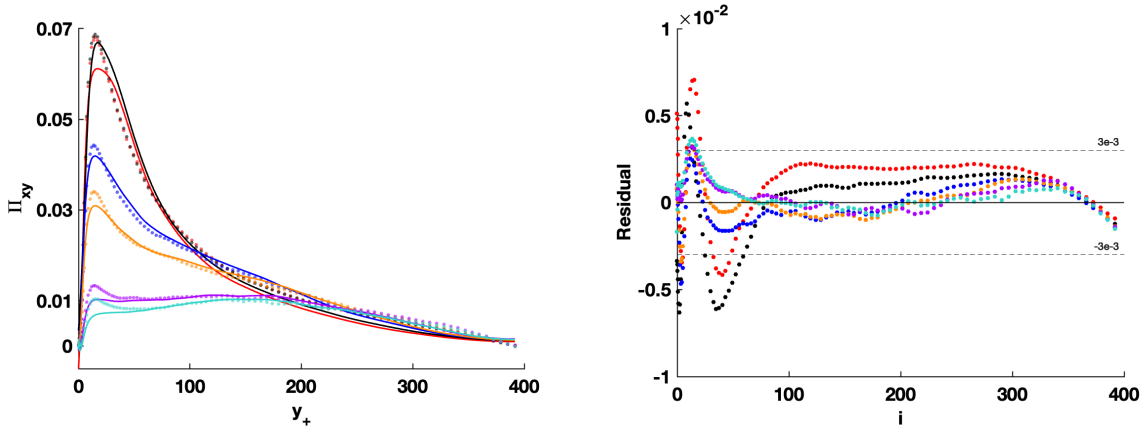
The results for these models are shown in Figure 4.10. Adding error terms to the list of regressors again significantly improved the fits over all strain values. Additionally, the residual distributions have similar magnitude to those in the previous model with error terms, implying that the negative effects of this removal step can be mitigated by the addition of error terms to the regressors.



(a) (b)

Figure 4.10: Results from model (4.5) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Continuing the reduction procedure one more step produces a model with results shown in Figure 4.11. The results indicate that the performance of this model is noticeably worse for all strain values. Therefore, model (4.4) is the final universal model for this VPG correlation.



(a) (b)

Figure 4.11: Results from model with 5 terms removed: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Now the LSLS loss, $\mathcal{L}$, is used to establish a rigorous stopping point for the reduction procedure for other correlations. The stepwise procedure of removing regressors is contin-

ued (not shown) until all regressors are removed from the list. The resulting values of $\mathcal{L}$ are shown in Table 4.10 and Figure 4.12 below. Note that the row with 11 terms removed describes the metric $\mathcal{L}$ for the model with no regressors, which predicts 0 at all points. This serves as a lower bound for model performance and therefore corresponds to 100% loss.

The loss is calculated for the models with secondary regressors as well:

$$\mathcal{L}_{\text{Err}} = \frac{\sum_{j=1}^{D} \sum_{i=1}^{N} (\Theta_i^{(j)} - a x_{1i}^{(j)} - b x_{2i}^{(j)} - c x_{3i}^{(j)} - \cdots - \alpha^{(j)} \mathrm{E}_{1i}^{(j)} - \beta^{(j)} \mathrm{E}_{2i}^{(j)} - \gamma^{(j)} \mathrm{E}_{3i}^{(j)} - \cdots)^2}{\sum_{j=1}^{D} \sum_{i=1}^{N} (\Theta_i^{(j)})^2} \tag{4.6}$$

where: D = total number of data sets,

N = number of entries in each vector, i.e. length of the vectors,

$a, b, c, \cdots$ = constant coefficients corresponding to each regressor in the LSLS model,

$\alpha^{(j)}, \beta^{(j)}, \gamma^{(j)}, \cdots$ = constant coefficients corresponding to each secondary regressor in the model for jth dataset,

$\mathrm{E}_{ri}^{(j)}$ = ith entry of rth secondary regressor from jth data set,

$\mathrm{x}_{ri}^{(j)}$ = ith entry of rth regressor from jth data set,

$\Theta_i^{(j)}$ = ith entry of observed variable vector from jth data set.

and is also shown in Table 4.10 and Figure 4.12 as $\mathcal{L}_{\text{Err}}$.

Table 4.10: Table showing values of $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ for Π_{xy} corresponding to the number of terms removed.

Model	# of terms removed	$\mathcal{L}(\%)$	$\mathcal{L}_{\text{Err}}(\%)$
(4.1)	0	0.14	0.07
(4.2)	1	0.17	0.09
(4.3)	2	0.18	0.10
(4.4)	3	0.29	0.13
(4.5)	4	0.33	0.17
	5	0.74	0.47
	6	3.49	2.58
	7	3.79	2.96
	8	3.91	3.09
	9	11.23	6.32
	10	15.31	13.24
	11	100	100

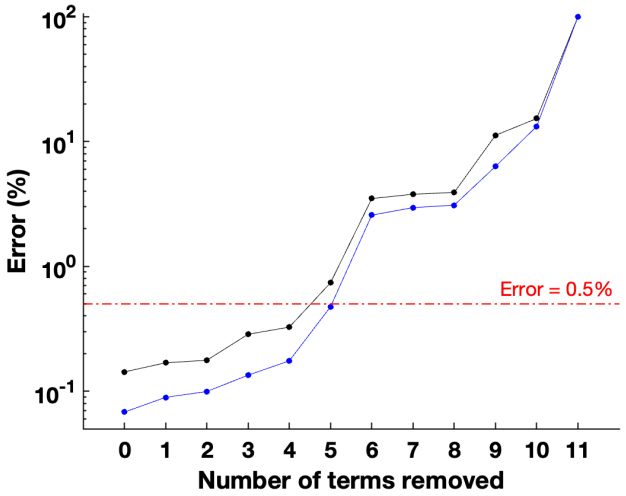


Figure 4.12: Values of $\mathcal{L}$ (shown as –●–) and $\mathcal{L}_{\text{Err}}$ (shown as –●–) for Π_{xy} vs. number of terms removed.

As evidenced by Table 4.10 and Figure 4.12, $\mathcal{L}$ provides similar information as $\mathcal{L}_{\text{Err}}$ while being simpler to calculate. Therefore in future sections, $\mathcal{L}$ is used to determine the threshold for determining when to stop the reduction procedure. The results for Π_{xy} shown in Figures 4.1-4.11 indicate that the model performance reduces significantly after five terms are removed. Hence, the proposed threshold for $\mathcal{L}$ is 0.5%.

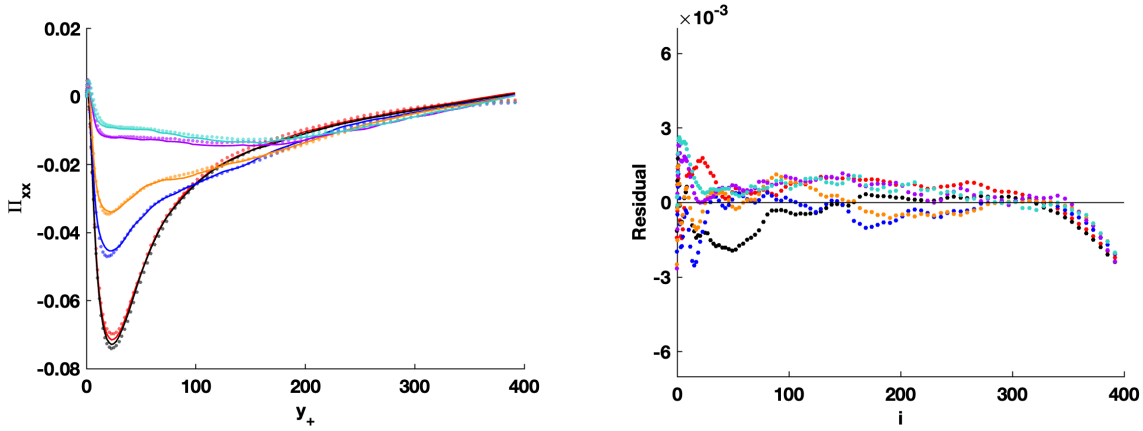
4.2. LSLS Applied to Modeling Π_{xx}

This section consists of an analysis similar to that in the previous section, but for the VPG correlation in the streamwise direction, Π_{xx}. First a model is generated using the LSLS procedure, then a model is generated keeping the regressors and coefficients from the previous model and adding the relevant balance error terms to the list of regressors. If the generated models describe the data well, the model is reduced using the reduction procedure described in Section 3.2. These steps are repeated until an obtained model is no longer satisfactory, and the last satisfactory model is kept as the final fit.

The original set of regressors includes the VPG correlations in the other flow directions and the turbulent and molecular diffusion terms found in the Reynolds Stress Transport Equation. The original model found using this set of regressors and LSLS is:

$$\begin{aligned}\Pi_{xx} = &-0.968\Pi_{xy} - 0.560\Pi_{yy} + 0.224\Pi_{zz} - 0.324D^T_{xy} - 0.015D^T_{xx} + \\ &0.078D^T_{yy} + 0.914D^T_{zz} + 0.055D^M_{xy} - 0.015D^M_{xx} - 0.784D^M_{yy} + 0.039D^M_{zz}.\end{aligned} \tag{4.7}$$

The results for this model are shown in Figure 4.13. The residual distributions indicate that the fits can model the data with an acceptable degree of accuracy, but that all models are heteroscedastic. Additionally, there is a consistent trend for the model to overpredict in regions near the channel axis ($y_+ > 300$).



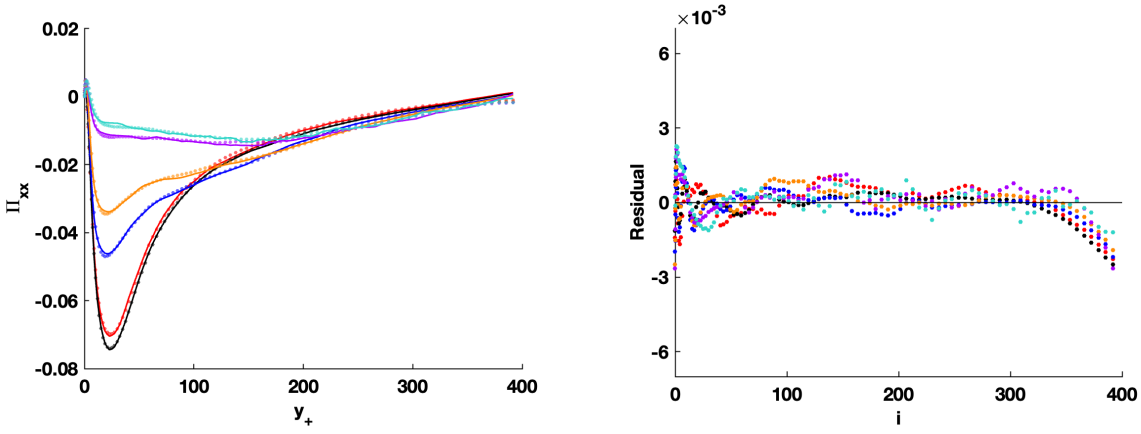
(a) (b)

Figure 4.13: Results from model (4.7): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

One possible explanation for this overprediction trend in the models is errors in the data itself. This serves as additional motivation for adding the set of relevant balance errors to the list of regressors, which generates the fits in Table 4.11, with results shown in Figure 4.14.

Table 4.11: Values of coefficients on secondary regressors corresponding to model (4.7)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-0.613	-0.366	2.566	0.311
0.002	-0.408	-0.058	-1.541	-0.836
0.281	1.170	0.383	0.225	-0.528
0.365	0.347	0.152	-0.155	-0.407
0.675	-1.725	-0.069	1.769	-1.508
0.772	-1.299	0.113	2.154	-1.756



(a) (b)

Figure 4.14: Results from model (4.7) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The residual distribution indicates that the models perform significantly better in the regions between the wall and the channel axis. However, the underprediction trend is still present even with the addition of the balance errors as regressors, indicating that it would be beneficial to add a regressor with different characteristics.

Since the model had acceptable accuracy, the reduction procedure is implemented to obtain the next fit. The values for $\Omega(s)$ of each regressor are shown in Table 4.12.

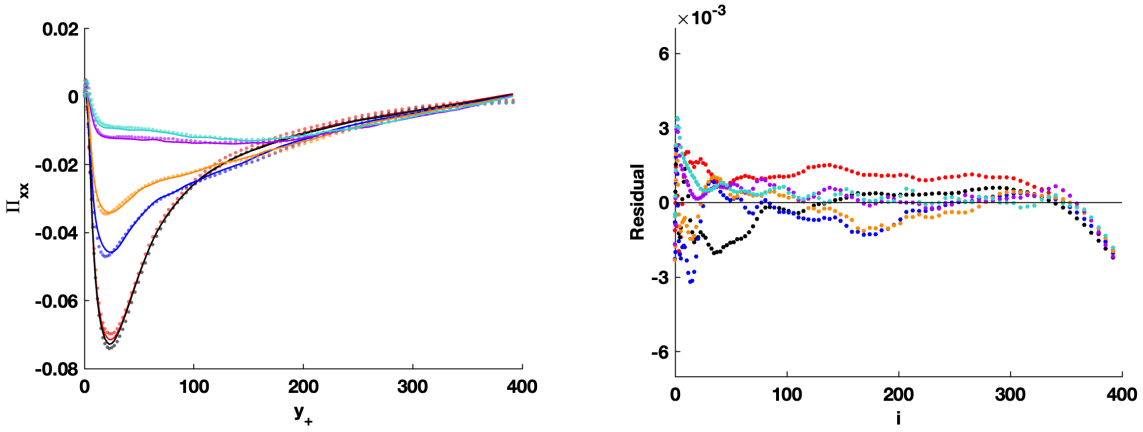
Table 4.12: $\Omega(s)$ values corresponding to regressors in model (4.7)

s	Π_{xy}	Π_{yy}	Π_{zz}	D^T_{xy}	D^T_{xx}	D^T_{yy}
$\mathbf{\Omega(s)}$	0.162e-5	0.139e-5	0.056e-5	0.181e-5	0.065e-5	0.054e-5

s	D^T_{zz}	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\mathbf{\Omega(s)}$	0.047e-5	0.054e-5	0.063e-5	0.062e-5	0.066e-5

Following Table 4.12, the next model should have the D^T_{zz} term removed. The model generated with this set of regressors is as follows:

$$\Pi_{xx} = -1.037\Pi_{xy} - 0.334\Pi_{yy} + 0.148\Pi_{zz} - 0.195D^T_{xy} - 0.015D^T_{xx} + 0.494D^T_{yy} - 0.149D^M_{xy} - 0.014D^M_{xx} + 0.447D^M_{yy} + 0.045D^M_{zz}. \tag{4.8}$$



(a) (b)

Figure 4.15: Results from model (4.8): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

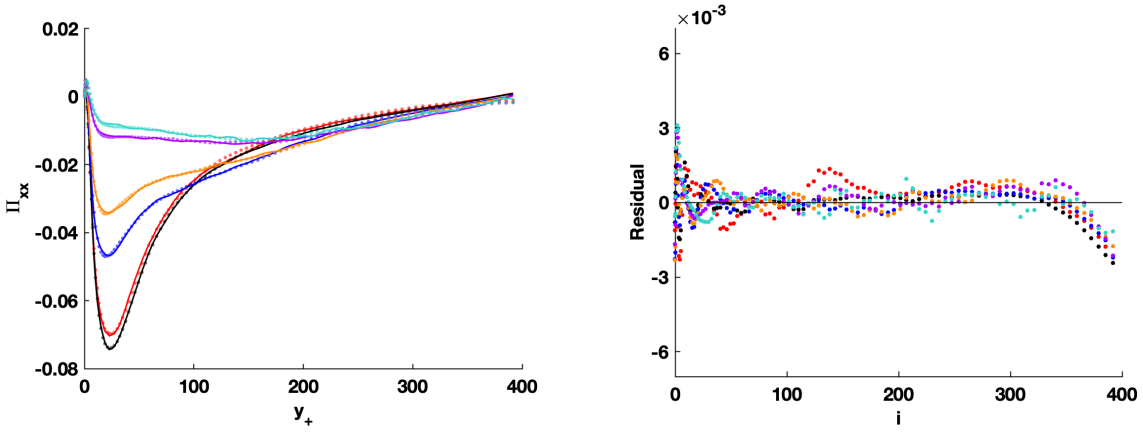
Model (4.8) has very similar performance to model (4.7), but with a slight decrease in accuracy in the unstrained flow (shown as red dots).

Adding errors to the list of regressors results in the fits shown in Table 4.13.

Table 4.13: Values of coefficients on secondary regressors corresponding to model (4.8)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-2.001	-0.261	3.390	1.377
0.002	-0.709	0.096	-2.132	-0.622
0.281	1.810	0.560	0.011	-0.732
0.365	1.030	0.447	0.298	-1.222
0.675	-1.144	-0.203	0.151	-0.760
0.772	-1.827	-0.356	0.928	-1.200

The results from these fits are shown in Figure 4.16. The results show that again, the addition of the error terms as regressors improves the fits in the middle region between the channel wall and axis, but makes little difference outside this region.



(a) (b)

Figure 4.16: Results from model (4.8) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Now, visual inspection of the residuals indicates the model accuracy is acceptable, so the reduction procedure can be repeated for another step. The $\Omega(s)$ values for the set of regressors are shown in Table 4.14.

Table 4.14: $\Omega(s)$ values corresponding to regressors in model (4.8)

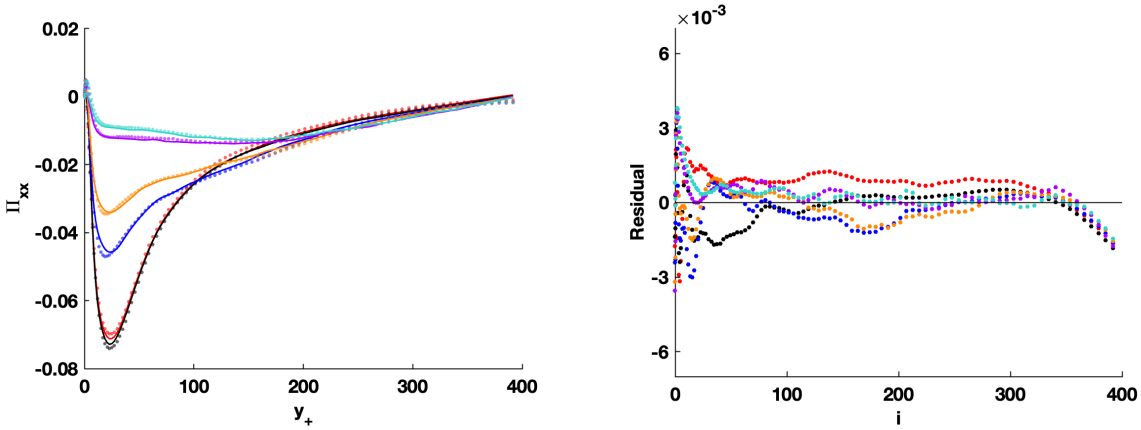
s	Π_{xy}	Π_{yy}	Π_{zz}	D^T_{xy}	D^T_{xx}	D^T_{yy}
$\boldsymbol{\Omega(s)}$	0.440e-3	0.428e-3	0.418e-3	0.424e-3	0.448e-3	0.475e-3

s	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\boldsymbol{\Omega(s)}$	0.455e-3	0.481e-3	0.527e-3	0.482e-3

From the $\Omega(s)$ table, the next regressor to be removed is Π_{zz}. The model generated with the LSLS procedure using these regressors is as follows:

$$\begin{aligned}\Pi_{xx} = &-0.891\Pi_{xy} - 0.446\Pi_{yy} - 0.167D^T_{xy} - 0.015D^T_{xx} + \\ &0.330D^T_{yy} - 0.375D^M_{xy} - 0.014D^M_{xx} + 1.093D^M_{yy} + 0.052D^M_{zz}.\end{aligned} \tag{4.9}$$

The resulting model fit and residual values are shown in Figure 4.17.



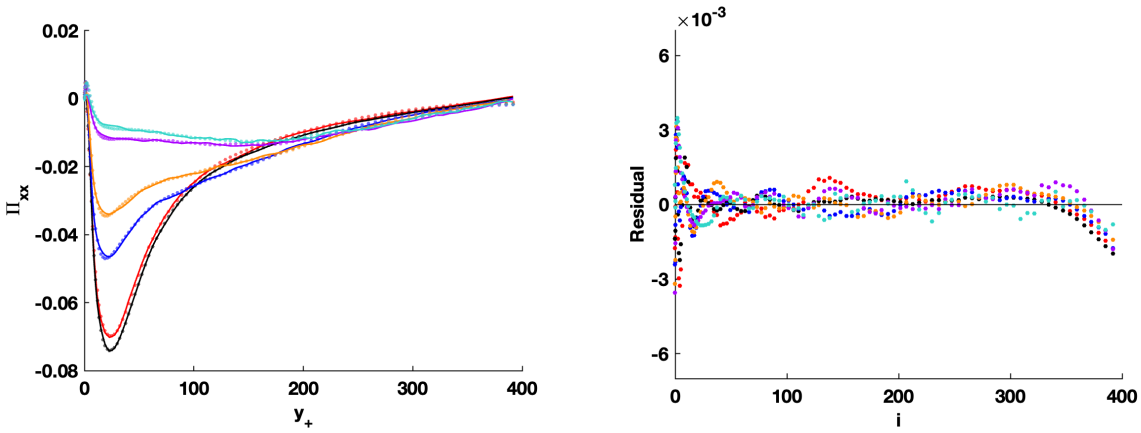
(a) (b)

Figure 4.17: Results from model (4.9): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The residual distribution for this model is very similar to that of preceding models implying that the removal of this term did not have a large effect on the fit quality. Now, adding relevant balance errors to the set of regressors results in the fits in Table 4.15, with plot results shown in Figure 4.18.

Table 4.15: Values of coefficients on secondary regressors corresponding to model (4.9)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-1.318	-0.327	3.115	1.061
0.002	-0.155	0.139	-1.522	-0.828
0.281	1.422	0.554	0.423	-1.083
0.365	0.558	0.380	0.312	-1.268
0.675	-1.650	-0.277	0.240	-0.739
0.772	-2.165	-0.392	0.632	-1.184



(a) (b)

Figure 4.18: Results from model (4.9) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The results again show that the model benefits from the addition of errors as regressors, and there is no significant penalty in fit quality as a result of removing a term. The model still has acceptable results, so the reduction procedure is continued for another step, with $\Omega(s)$ values shown in Table 4.16.

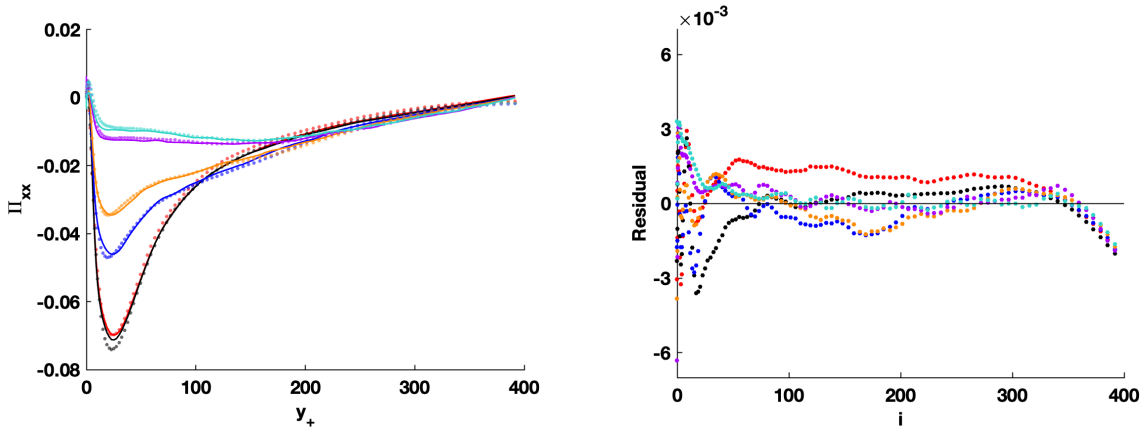
Table 4.16: $\Omega(s)$ values corresponding to regressors in model (4.9)

s	Π_{xy}	Π_{yy}	D^T_{xy}	D^T_{xx}	D^T_{yy}
$\Omega(s)$	0.447e-3	0.428e-3	0.426e-3	0.448e-3	0.478e-3

s	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.505e-3	0.566e-3	0.586e-3	0.583e-3

The $\Omega(s)$ values in in Table 4.16 indicate that the next regressor for removal is D^T_{xy}. The LSLS model (results shown in Figure 4.19) with this updated regressors list is as follows:

$$\begin{aligned}\Pi_{xx} = -0.933\Pi_{xy} - 0.373\Pi_{yy} - 0.011D^T_{xx} + \\ 0.415D^T_{yy} - 0.233D^M_{xy} - 0.011D^M_{xx} + 2.049D^M_{yy} + 0.037D^M_{zz}.\end{aligned} \tag{4.10}$$



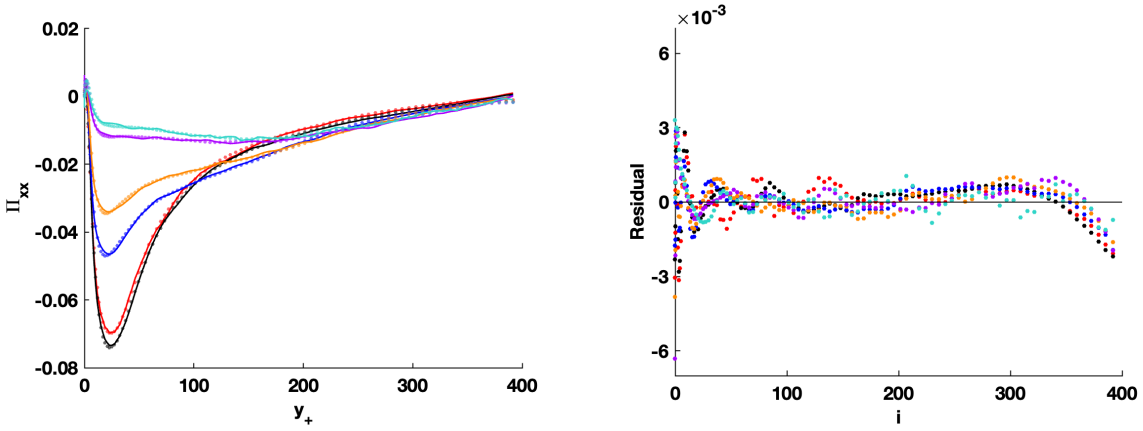
(a) (b)

Figure 4.19: Results from model (4.10): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

At this step, the residual plot indicates a slight decrease in the fit quality near the wall since the residual values there are larger in magnitude. However, the model performs similarly for the rest of the flow as previous models. Now, adding relevant balance errors to the list of regressors results in the models shown in Table 4.17, with results shown in Figure 4.20.

Table 4.17: Values of coefficients on secondary regressors corresponding to model (4.10)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-1.662	0.790	3.445	1.936
0.002	-1.054	-0.329	-1.858	1.195
0.281	1.178	0.392	-0.203	-0.616
0.365	0.146	0.242	0.198	-1.199
0.675	-1.666	-0.618	0.188	-0.231
0.772	-2.929	-0.836	0.818	-1.122



(a) (b)

Figure 4.20: Results from model (4.10) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

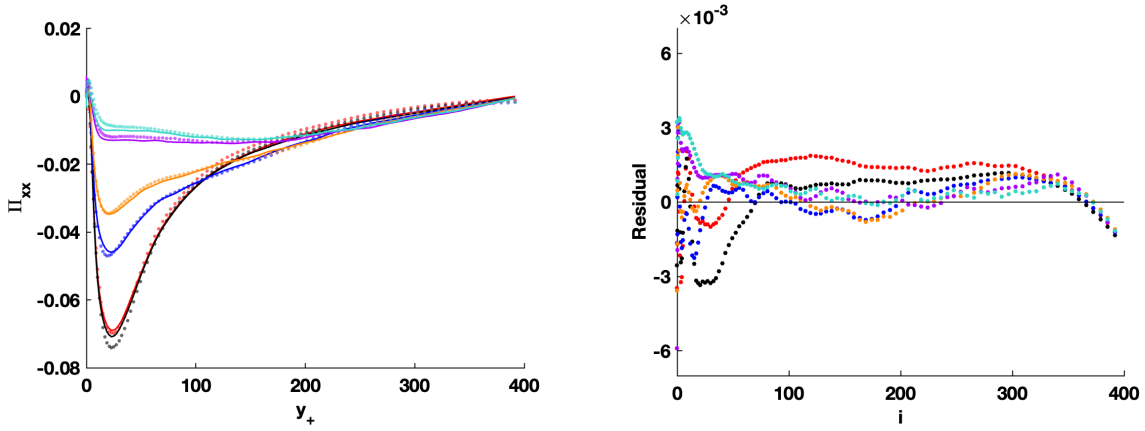
The addition of balance errors to the set of regressors resulted in reduced residual magnitudes for all strain values in the region of the channel between the wall region and the axis. This is expected because the error terms are small in magnitude near the axis and the wall, which implies that their contribution to a model will be small in this region. The models are still performing at an acceptable level, so the reduction procedure is repeated for another step, generating the $\Omega(s)$ values in Table 4.18.

Table 4.18: $\Omega(s)$ values corresponding to regressors in model (4.10)

s	Π_{xy}	Π_{yy}	D^T_{xx}	D^T_{yy}
$\Omega(s)$	0.682e-3	0.557e-3	0.534e-3	0.482e-3
s	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.514e-3	0.567e-3	0.588e-3	0.583e-3

These $\Omega(s)$ values indicate that D^T_{yy} should be removed from the regressors list in the next step. The fit with this new set of regressors, with results shown in Figure 4.21, is as follows:

$$\begin{aligned}\Pi_{xx} = &-0.876\Pi_{xy} - 0.544\Pi_{yy} - 0.008D^T_{xx} \\ &-0.146D^M_{xy} - 0.006D^M_{xx} + 1.900D^M_{yy} + 0.029D^M_{zz}.\end{aligned} \tag{4.11}$$



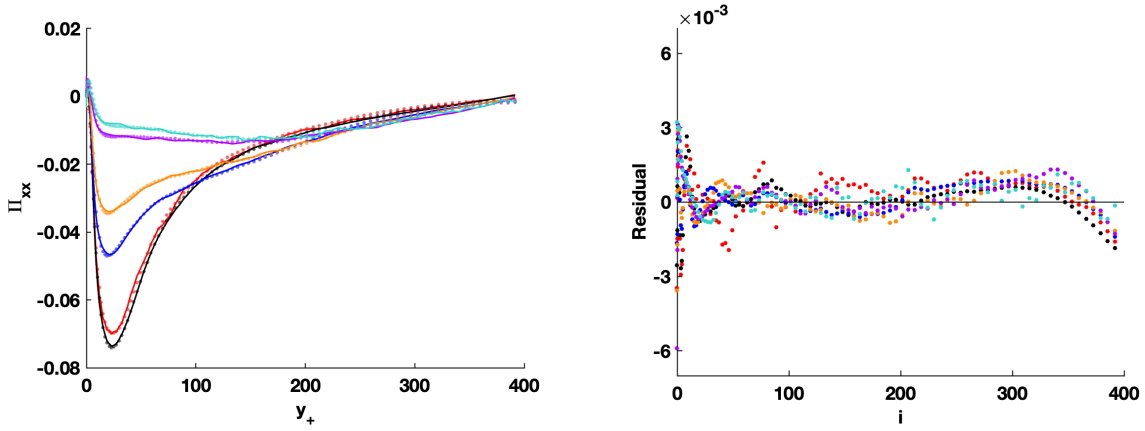
(a) (b)

Figure 4.21: Results from model (4.11): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Model 4.11 performs similarly to the previous models, implying that the removal of D^T_{yy} is not detrimental to performance. Adding balance error terms to the list of regressors results in the fits shown in Table 4.19, with results shown in Figure 4.22.

Table 4.19: Values of coefficients on secondary regressors corresponding to model (4.11)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	-3.911	0.735	1.093	3.655
0.002	-0.946	0.070	-3.754	0.457
0.281	1.388	0.340	0.699	-0.386
0.365	0.883	0.397	2.046	-1.885
0.675	-0.260	-0.572	1.599	-0.399
0.772	-3.399	-0.814	0.819	-1.712



(a) (b)

Figure 4.22: Results from model (4.11) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The resulting fit still has an acceptable level of accuracy; therefore, the reduction procedure is implemented again, producing the $\Omega(s)$ values in Table 4.20.

Table 4.20: $\Omega(s)$ values corresponding to regressors in model (4.11)

s	Π_{xy}	Π_{yy}	D^T_{xx}
$\Omega(s)$	0.149e-3	0.073e-3	0.054e-3

s	D^M_{xy}	D^M_{xx}	D^M_{yy}	D^M_{zz}
$\Omega(s)$	0.053e-3	0.058e-3	0.060e-3	0.060e-3

At this point the reduction procedure is continued until all terms are removed, producing the values for $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ shown in Table 4.21and Figure 4.23.

Table 4.21: Table showing value of $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ for Π_{xx} corresponding to the number of terms removed.

Model	# of terms removed	$\mathcal{L}(\%)$	$\mathcal{L}_{\text{Err}}(\%)$
(4.7)	0	0.15	0.09
(4.8)	1	0.18	0.10
(4.9)	2	0.19	0.13
(4.10)	3	0.24	0.15
(4.11)	4	0.28	0.15
	5	0.28	0.15
	6	0.28	0.16
	7	0.28	0.17
(4.12)	8	0.29	0.22
	9	0.56	0.46
	10	2.76	2.28
	11	100	100

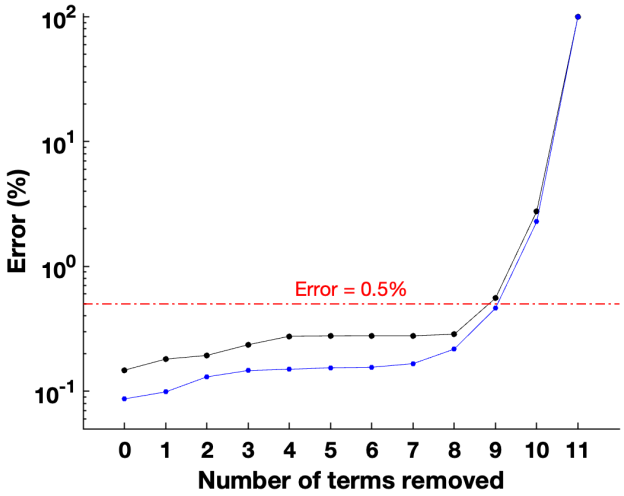


Figure 4.23: Values of $\mathcal{L}$ (shown as –●–) and $\mathcal{L}_{\text{Err}}$ (shown as –●–) for Π_{xx} vs. number of terms removed.

The plot and table indicate that stopping the reduction procedure when 8 regressors are removed from the original list satisfies the previously established 0.5% threshold for $\mathcal{L}$.

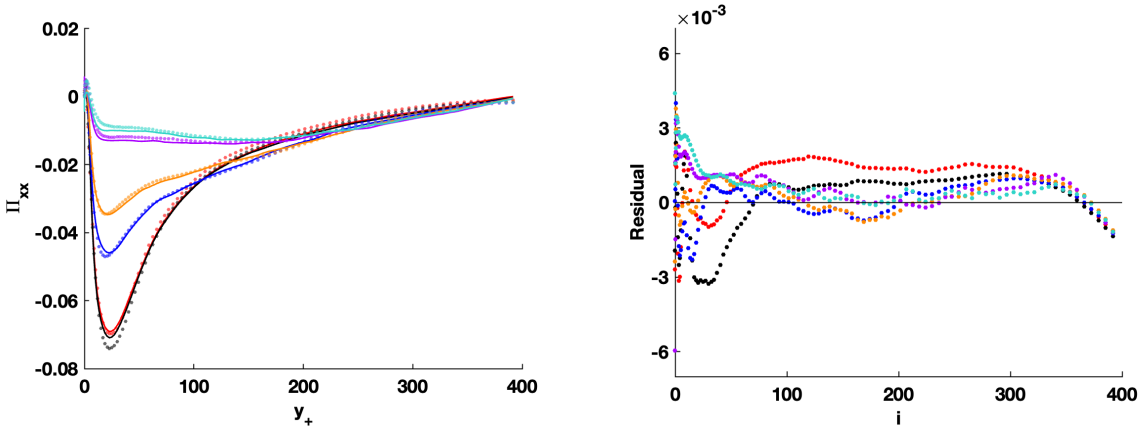
The resulting model after completing 8 steps of the reduction procedure is as follows:

$$\Pi_{xx} = -0.872\Pi_{xy} - 0.551\Pi_{yy} + 2.015D^{M}_{yy}. \tag{4.12}$$

The results for this model are shown in Figure 4.24. Now, adding the relevant secondary regressors to the regressors list as before produces the model coefficient values in Table 4.22. The results for these models are shown in Figure 4.25.

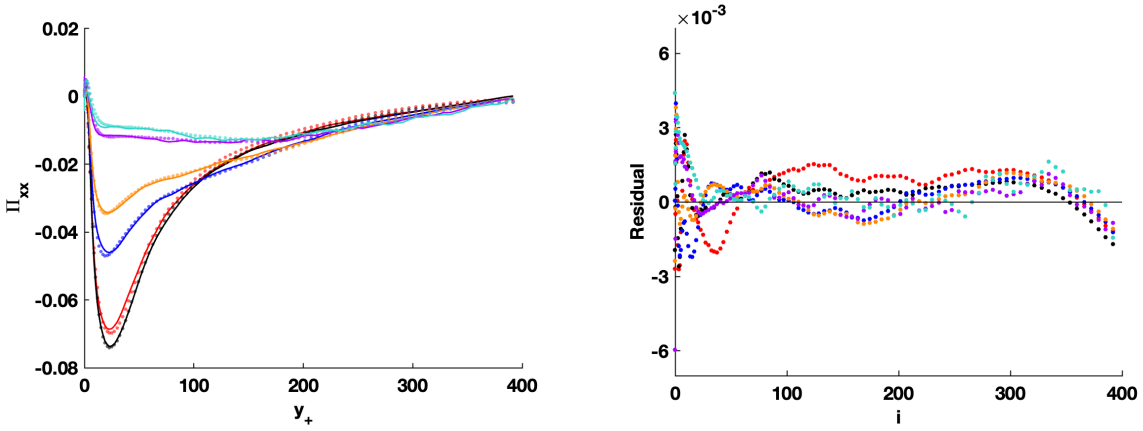
Table 4.22: Values of coefficients on secondary regressors corresponding to model (4.12)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{yy}}$
Unstrained	1.327	3.222
0.002	-0.671	-2.975
0.281	-0.123	-0.052
0.365	0.688	0.349
0.675	3.439	1.083
0.772	3.105	1.522



(a) (b)

Figure 4.24: Results from model (4.12): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .



(a) (b)

Figure 4.25: Results from model (4.12) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

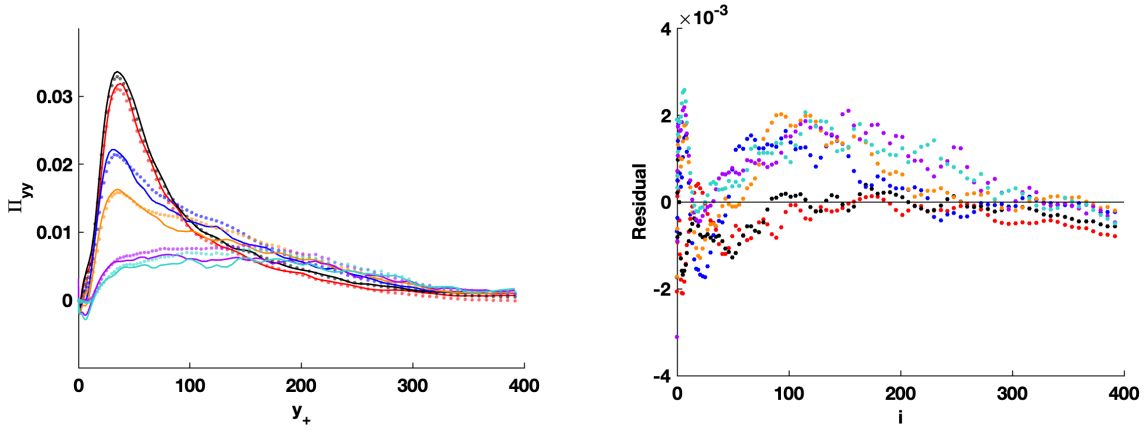
The final model for this VPG correlation is model (4.12). In all models, the addition of the secondary regressors to the list of regressors resulted in improved performance. The model has heteroscedastic residual distributions for all strain values and struggles to predict Π_{xx} in the unstrained channel in the near-wall region. Despite the model limitations, the results of this section indicate that this VPG correlation can be modeled with 3 terms.

4.3. LSLS Applied to Modeling Π_{yy}

This section describes the application of the LSLS method to modeling the Π_{yy} VPG correlation using VPG correlations in the other directions and diffusion terms from the Reynolds Stress Transport Equation as regressors. The original model using the full list of regressors is as follows:

$$\begin{aligned}\Pi_{yy} = &-0.225\Pi_{xy} - 0.626\Pi_{xx} + 0.012\Pi_{zz} - 0.354D^T_{xy} - 0.001D^T_{xx} \\ &-0.969D^T_{yy} + 1.593D^T_{zz} + 0D^M_{xy} - 0.001D^M_{xx} + 1.003D^M_{yy} + 0.008D^M_{zz}.\end{aligned} \tag{4.13}$$

The results for this model are shown in Figure 4.26.



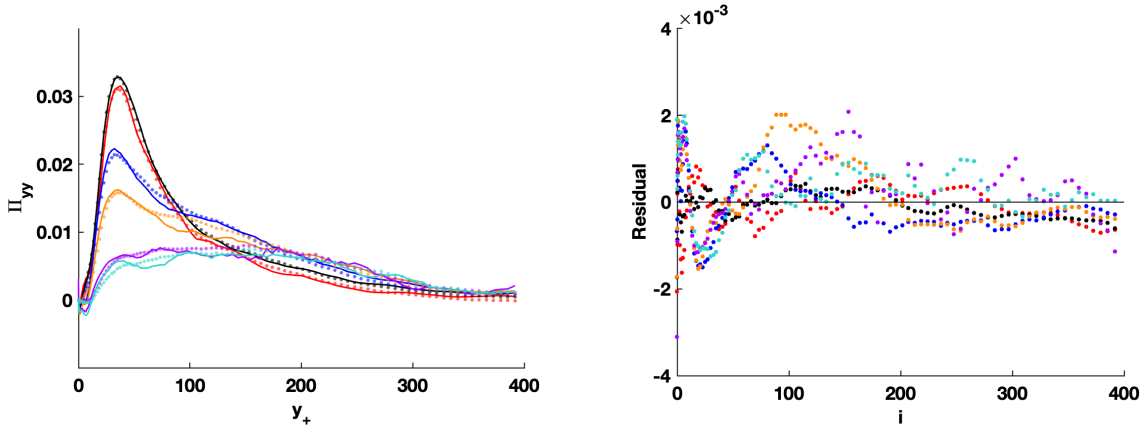
(a) (b)

Figure 4.26: Results from model (4.13): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

Adding the relevant secondary regressors to the list of regressors results in the fits in Table 4.23, with results plots shown in Figure 4.27:

Table 4.23: Values of coefficients on secondary regressors corresponding to model (4.13) without D^M_{xy}.

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	1.327	-0.496	-1.172	-0.547
0.002	1.535	0.316	0.019	-1.523
0.281	0.520	0.255	2.953	-1.105
0.365	0.253	-0.032	1.270	-0.056
0.675	-1.189	0.450	4.434	-2.628
0.772	-0.129	1.100	2.235	-2.369



(a) (b)

Figure 4.27: Results from model (4.13) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The models with secondary regressors included perform noticeably better than model (4.13), as indicated by the reduced residual magnitude. The zero-value of the coefficient for D^M_{xy} indicates that this regressor will be the first term to be removed in the reduction process. Visual comparison of figures 4.26 and 4.27 to those of previous correlations suggest that model (4.13), which is an upper bound for performance in this model framework for this VPG correlation, has poor performance. This suggestion is validated by the large loss values, $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$, of 0.97% and 0.58%, respectively, for this model. These loss values are above the 0.5% quality threshold, implying that a model of similar performance to that of previous sections cannot be obtained with this set of regressors for the Π_{yy} correlation. The remaining loss values for this VPG correlation are shown in Table 4.24 and Figure 4.28.

Table 4.24: Table showing value of $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ for Π_{yy} corresponding to the number of terms removed.

Model	# of terms removed	$\mathcal{L}(\%)$	$\mathcal{L}_{\text{Err}}(\%)$
(4.13)	0	0.97	0.58
(4.14)	1	0.97	0.58
	2	2.08	1.37
	3	2.41	1.26
	4	3.11	1.18
	5	4.64	1.55
	6	10.37	5.69
	7	14.93	7.60
	8	15.37	8.91
	9	17.83	12.41
	10	25.34	19.66
	11	100	100

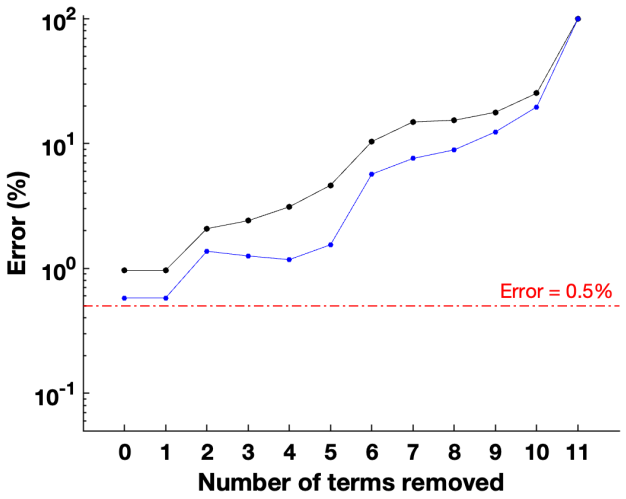


Figure 4.28: Values of $\mathcal{L}$ (shown as –●–) and $\mathcal{L}_{\text{Err}}$ (shown as –●–) for Π_{yy} vs. number of terms removed.

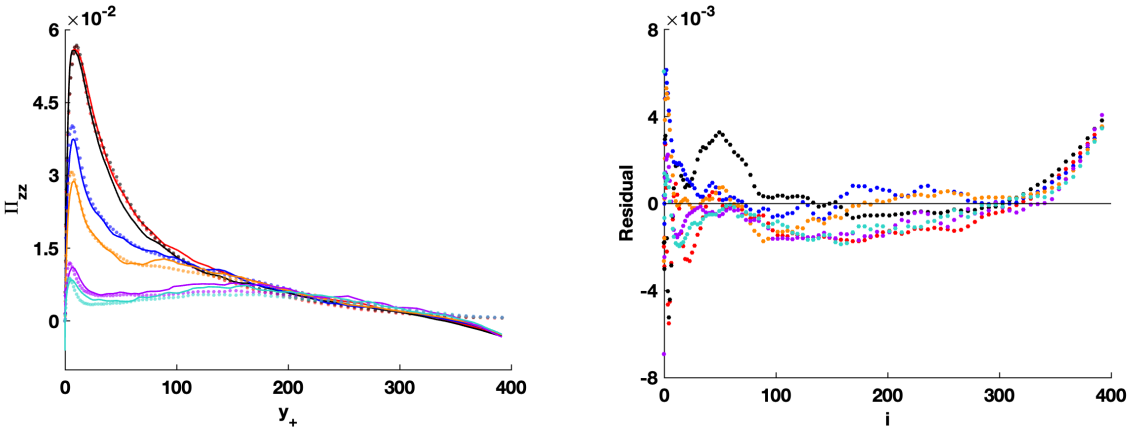
In light of the poor performance of the original model, the modeling process for this correlation will not include a reduction procedure. Therefore, the final model for Π_{yy} is Model 4.13 without the D_{xy}^{M} term:

$$\Pi_{yy} = -0.225\Pi_{xy} - 0.626\Pi_{xx} + 0.012\Pi_{zz} - 0.354D^T_{xy} - 0.001D^T_{xx} \\ -0.969D^T_{yy} + 1.593D^T_{zz} - 0.001D^M_{xx} + 1.003D^M_{yy} + 0.008D^M_{zz}. \tag{4.14}$$

4.4. LSLS Applied to Modeling Π_{zz}

This section describes the application of the LSLS method to modeling the Π_{zz} VPG correlation using VPG correlations in the other directions and diffusion terms from the Reynolds Stress Transport Equation as regressors. The original model using the full list of regressors (results plots shown in Figure 4.29) is as follows:

$$\Pi_{zz} = -1.352\Pi_{xy} + 0.685\Pi_{xx} + 0.033\Pi_{yy} + 0.499D^T_{xy} + 0.009D^T_{xx} \\ -0.479D^T_{yy} - 1.565D^T_{zz} - 1.421D^M_{xy} + 0.008D^M_{xx} + 2.472D^M_{yy} + 0.015D^M_{zz}. \tag{4.15}$$



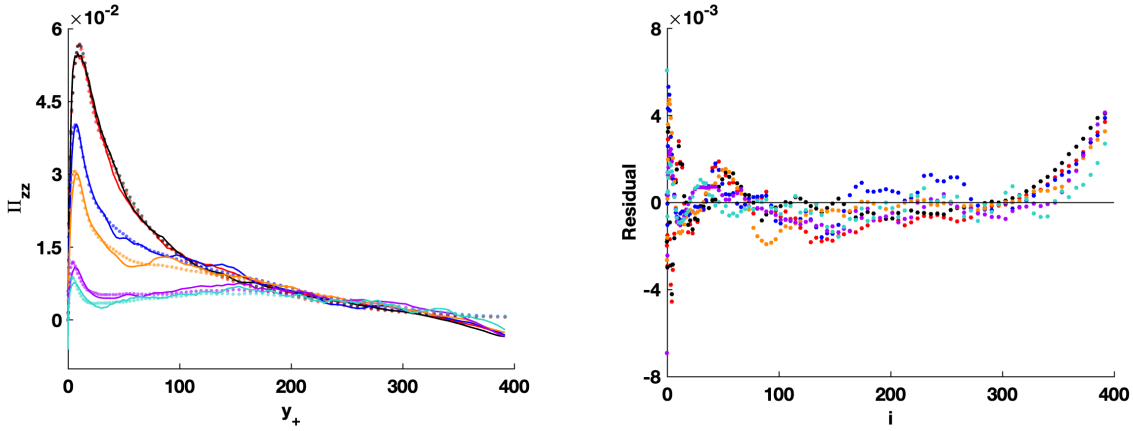
(a) (b)

Figure 4.29: Results from model (4.15): a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

The model performs well in the regions of the flow between the wall region and the axis; however, its performance struggles otherwise. There is a noticeable underprediction trend in the residuals near the channel axis for all strain values. As before, balance errors are now added to the list of regressors, generating the models in Table 4.25, with results shown in Figure 4.30:

Table 4.25: Values of coefficients on secondary regressors corresponding to model (4.15)

	$\mathbf{Err_{xy}}$	$\mathbf{Err_{xx}}$	$\mathbf{Err_{yy}}$	$\mathbf{Err_{zz}}$
Unstrained	2.538	0.015	-2.341	-0.7622
0.002	2.850	0.220	3.618	-0.278
0.281	-2.163	-0.115	1.766	-1.672
0.365	-1.990	-0.151	0.643	-0.796
0.675	-0.854	-0.475	-2.406	1.907
0.772	-1.407	-0.747	-4.269	1.855



(a) (b)

Figure 4.30: Results from model (4.15) with error terms in regressors list: a) comparison of model (lines) and DNS data (dots), b) residual distribution plots. $A_{22}t = 0$ ● – , $A_{22}t = 0.020$ ● – , $A_{22}t = 0.281$ ● – , $A_{22}t = 0.365$ ● – , $A_{22}t = 0.675$ ● – , $A_{22}t = 0.772$ ● – .

All models perform noticeably better with the addition of errors as residuals, but the underprediction trend near the axis persists. It is expected that trends near the axis and wall regions remain unchanged with the addition of the error terms as regressors because of the small magnitude of these terms in these regions. The reduction procedure was continued to generate the values for $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ shown in Table 4.26 and Figure 4.31.

Table 4.26: Table showing value of $\mathcal{L}$ and $\mathcal{L}_{\text{Err}}$ for Π_{zz} corresponding to the number of terms removed.

Model	# of terms removed	$\mathcal{L}(\%)$	$\mathcal{L}_{\text{Err}}(\%)$
(4.15)	0	0.85	0.62
	1	0.89	0.66
	2	1.63	1.07
	3	1.70	0.95
	4	2.62	1.69
	5	2.71	1.70
	6	2.79	2.13
	7	6.77	5.67
	8	8.33	7.23
	9	8.54	7.73
	10	15.31	12.94
	11	100	100

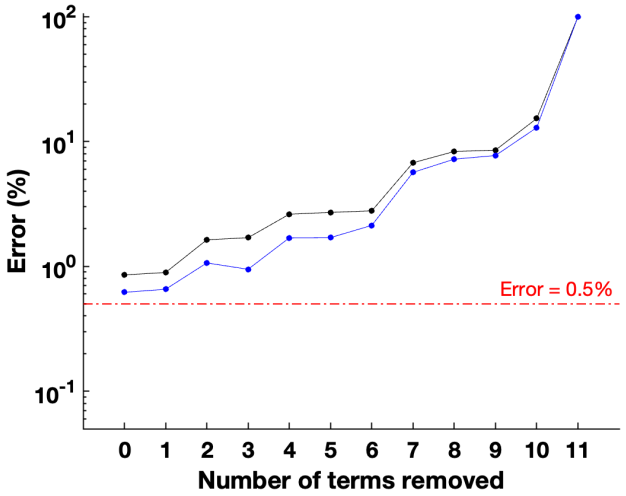


Figure 4.31: Values of $\mathcal{L}$ (shown as –●–) and $\mathcal{L}_{\text{Err}}$ (shown as –●–) for Π_{zz} vs. number of terms removed.

Table 4.26 and Figure 4.31 indicate that all LSLS models for Π_{zz} produce loss values exceeding the previously established threshold of 0.5%. Therefore, the original set of regressors cannot be reduced, and the final model for the Π_{zz} correlation is model (4.15).

CONCLUSIONS

The results of the study demonstrate that the velocity pressure-gradient (VPG) correlations in a channel flow can be accurately represented by linear models generated using the Multiple Linear Regression (MLR) method. It was shown that the residual magnitudes and scedasticities can be used as metrics to quantify the performance of generated models. The residual magnitudes and scedasticities informed the model development; therefore, the proposed method provides a mathematically rigorous and efficient framework for generating such models.

It was demonstrated that it is also possible to develop accurate linear universal models for VPG correlations with unchanging coefficients in an adverse pressure-gradient channel flow (flow with separation) using the Least Sum of Least Squares (LSLS) method. In contrast to the MLR method, the LSLS method optimizes model coefficients for multiple data sets simultaneously. To evaluate the accuracy of such models and to reduce their complexity, the loss metric was proposed and used.

The regressors lists used for modeling the Π_{xy} and Π_{xx} correlations were reduced according to the loss metric, and the resulting models are more accurate than the models for the Π_{yy} and Π_{zz} correlations, which employed the entire regressors list.

Finally, for all considered models and methods, it was shown that the addition of balance error terms as regressors to the linear models produces more accurate models for all channel strain values. This suggests that future data-driven modeling approaches using this DNS data set should incorporate the balance errors in some form.

FUTURE WORK

Future research should include modeling the VPG correlations using data sets from other DNS and exploring generalizing the models beyond only channel flow. Additionally, it will be pertinent to analyze the coefficient values in the generated models to extract physical interpretation from the models. It will be interesting to assess the performance of LSLS models for the Π_{yy} and Π_{zz} correlations with a regressors list containing production and diffusion terms in the Reynolds transport equation budgets, similar to that used for the traditional MLR models in the unstrained case.

BIBLIOGRAPHY

[1] A. N. Kolmogorov, "Equations of turbulent motion of an incompressible fluid," *Izvestiya Academii Nauk USSR: Physics*, vol. 6, no. 1-2, pp. 56–58, 1942.

[2] D. C. Wilcox, "Formulation of the k-ω turbulence model revisited," *AIAA Journal*, vol. 46, no. 11, pp. 2823–2838, 2008. DOI: 10.2514/1.36541.

[3] K. Hanjalić, "Two-dimensional asymmetrical turbulent flow in ducts," Ph.D. dissertation, University of London, 1970.

[4] W. P. Jones and B. Launder, "The prediction of laminarization with a two-equation model of turbulence," *International Journal of Heat and Mass Transfer*, vol. 15, no. 2, pp. 301–314, 1972. DOI: 10.1016/0017-9310(72)90076-2.

[5] P. R. Spalart and S. R. Allmaras, "A one-equation turbulence model for aerodynamic flows," *La Recherche Aérospatiale*, no. 1, pp. 5–21, 1994, (DOI is of later copy). DOI: 10.2514/6.1992-439.

[6] NASA Langley Research Center, *Turbulence Modeling Resource*. [Online]. Available: https://turbmodels.larc.nasa.gov/index.html.

[7] T. J. Craft, B. E. Launder, and K. Suga, "Development and application of a cubic eddy-viscosity model of turbulence," *International Journal of Heat and Fluid Flow*, vol. 17, no. 2, pp. 108–115, 1996. DOI: 10.1016/0142-727X(95)00079-6.

[8] S. B. Pope, *Turbulent Flows*. Cambridge University Press, 2000, ISBN: 9780521598866.

[9] A. V. Johansson and M. Hallback, "Modelling of rapid pressure-strain in Reynolds-stress closures," *Cambridge University Press*, vol. 269, pp. 143–168, 1994. DOI: 10.1017/S0022112094001515.

[10] K. Hanjalić and B. Launder, *Modelling Turbulence in Engineering and the Environment*. Cambridge University Press, Cambridge, UK, 2011, ISBN: 0521845750.

[11] C. Speziale, S. Sarkar, and T. Gatski, "Modelling the pressure– strain correlation of turbulence: An invariant dynamical systems approach," *Journal of Fluid Mechanics*, vol. 227, pp. 245–272, 1991. DOI: 10.1017/S0022112091000101.

[12] B. Launder, G. Reece, and W. Rodi, "Progress in the development of a Reynolds-stress turbulence closure," *Journal of Fluid Mechanics*, vol. 68, no. 3, pp. 537–566, 1975. DOI: 10.1017/S0022112075001814.

[13] A. Mishra and S. Girimaji, "Toward approximating non-local dynamics in single-point pressure–strain correlation closures," *Journal of Fluid Mechanics*, vol. 811, pp. 168–188, 2017. DOI: 10.1017/jfm.2016.730.

[14] S. Taghizadeh, F. D. Witherden, and Y. A. Hassan, "Turbulence closure modeling with data-driven techniques: Investigation of generalizable deep neural networks," *Physics of Fluids*, 2021.

[15] J. P. Panda and H. V. Warrior, "Progress in the development of a Reynolds-stress turbulence closure," 2021.

[16] P. Y. Chou, "On the velocity correlations and the solutions of the equations of turbulent fluctuation," *Quarterly of Applied Mathematics*, vol. 3, pp. 38–54, 1945.

[17] S. V. Poroseva, "Modeling the "rapid" part of the velocity/pressure-gradient correlation in inhomogeneous turbulence," *Annual Research Brief 2001, Center for Turbulence Research, NASA-Ames/Stanford University*, pp. 367–374, 2001.

[18] A. F. Kurbatskii, S. V. Poroseva, and S. N. Yakovenko, "Calculation of statistical characteristics of a turbulent flow in a rotating cylindrical pipe," *High Temperature*, vol. 33, no. 5, pp. 732–742, 1995, Translated from Teplofizika Vysokikh Temperatur. Vol 33(5), pp. 738-748.

[19] A. F. Kurbatskii and S. V. Poroseva, "A model for calculating the three components of the excess for the turbulent field of flow velocity in a round pipe rotating about its longitudinal axis," *High Temperature*, vol. 35, no. 3, pp. 432–440, 1997.

[20] S. V. Poroseva and S. M. Murman, "Velocity/pressure-gradient correlations in a FORANS approach to turbulence modeling," *AIAA2014-2207, Proc. the AIAA Aviation and Aeronautics Forum and Exposition, Atlanta, GA*, 2014.

[21] S. V. Poroseva, "The effect of a pressure-containing correlation model on near-wall flow simulations with RST models," *ASME Journal of Fluids Engineering*, vol. 136, no. 6, pp. 38–54, 2013. DOI: 10.1115/1.4025936..

[22] S. V. Poroseva, S. C. Kassinos, C. A. Langer, and W. C. Reynolds, "Structure-based turbulence model: Application to a rotating pipe flow," *Physics of Fluids*, vol. 14, no. 4, pp. 1523–1532, 2002. DOI: 10.1063/1.1458008.

[23] S. V. Poroseva and S. M. Murman, "On modelling velocity/pressure-gradient correlations in higher-order RANS statistical closures," *Proc. 19th AFMC, Melbourne, Australia*, 2014.

[24] S. V. Poroseva, J. D. Colmenares, and S. M. Murman, "RANS simulations of a channel flow with a new Velocity/Pressure-Gradient model," *AIAA2015-3067, Proc. AIAA Aviation 2015, Dallas, TX*, 2015.

[25] S. V. Poroseva and S. M. Murman, "Reynolds-stress simulations of wall-bounded flows using a new velocity/pressure-gradient model," *Proc. TSHP-9, Melbourne, Australia*, 2015.

[26] S. V. Poroseva and S. M. Murman, "Sensitivity of a new velocity/pressure-gradient model to the Reynolds number," *Proc. TSFP10*, 2017.

[27] S. V. Poroseva and S. M. Murman, "Validation of algebraic models for VPG and dissipation terms in RANS and FANS equations," , *Presentation at the ECCM - ECFD 2018 6th European Conference on Computational Mechanics (Solids, Structures and Coupled Problems), 7th European Conference on Computational Fluid Dynamics, Glasgow, UK*, 2018.

[28] P. R. Spalart, "Direct simulation of a turbulent boundary layer up to $Re_\theta = 1410$," *Journal of Fluid Mechanics*, vol. 187, pp. 61–98, 1988. DOI: 10.1017/S0022112088000345.

[29] J. A. Sillero, J. Jiménez, and R. D. Moser, "One-point statistics for turbulent wall-bounded flows at Reynolds numbers up to $\delta^+ = 2000$," *Physics of Fluids*, vol. 25, no. 10, pp. 1–15, 2013. DOI: 10.1063/1.4823831.

[30] M. Lee and R. D. Moser, "Direct numerical simulation of turbulent channel flow up to $Re_\tau = 5200$," *Journal of Fluid Mechanics*, vol. 774, pp. 395–415, 2015. DOI: 10.1017/jfm.2015.268.

[31] E. Jeyapaul, G. N. Coleman, and C. L. Rumsey, "Higher-order and length-scale statistics from DNS of a decelerated planar wall-bounded turbulent flow," *International Journal of Heat and Fluid Flow*, vol. 54, pp. 14–27, 2015. DOI: 10.1016/j.ijheatfluidflow.2015.04.002.

[32] S. V. Poroseva, J. D. Colmenares, and S. M. Murman, "On the accuracy of RANS simulations with DNS data," *Physics of Fluids*, vol. 28, no. 11, 2016. DOI: 10.1063/1.4966639.

[33] S. V. Poroseva, E. Jeyapaul, S. M. Murman, and J. D. Colmenares, "The effect of the DNS data averaging time on the accuracy of RANS-DNS simulations," *AIAA2016-3940, Proc. AIAA Aviation, Washington DC*, June 13-17, 2016.

[34] MathWorks, *MATLAB*, version 2022b, 2022. [Online]. Available: https://www.mathworks.com/products/matlab.html.

[35] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley Sons, Inc., Wiley, Hoboken, New Jersey, 2012, pp. 67–120, ISBN: 0470542810.

[36] A. N. Kolmogorov, "Equations of turbulent motion in an incompressible fluid," *Proceedings of the USSR Academy of Sciences*, vol. 30, pp. 299–303, 1941.