UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INVESTIGATION OF COINTEGRATION RELATIONSHIPS IN TURBULENT FLOWS

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

JOSHUA HAINES
Norman, Oklahoma
2016

INVESTIGATION OF COINTEGRATION RELATIONSHIPS
IN TURBULENT FLOWS

A THESIS APPROVED FOR THE
SCHOOL OF AEROSPACE AND MECHANICAL ENGINEERING

BY

Dr. Prakash Vedula, Chair

Dr. Peter Attar

Dr. Jivtesh Garg

Dr. Li Song

Dedication

This thesis is dedicated to my loving wife Julie who sacrificed a great deal of time and frustration during the countless nights that were spent in the office working on this endeavor. She is appreciated more than words can ever say and loved ever so deeply.

Table of Contents

List of Tables

List of Figures

Abstract

Turbulence has often been called one of the last unsolved problems of classical physics and impacts a wide variety of fields in nature and engineering. An important area of focus in turbulence research is the discovery and scaling laws in turbulence which facilitate reduced order representations that are useful for efficient computational predictions of turbulent flows.

In this work we seek to utilize the method of cointegration, a tool typically utilized in econometrics for the study of non-stationary time series, to search for and quantify new relationships amongst a variety of physical quantities in (nonstationary or inhomogeneous) turbulent flows. The primary research questions of this work are: (i) Do cointegration relationships exist in non-stationary and/or non-homogeneous turbulent flows?, (ii) If so, what are the cointegration relationships?, (iii) What collections of signals show strong evidence of cointegration relationships? and (iv) What is the utility of cointegration analysis in investigation of turbulent flows? To address these questions, we analyze flow data (available from publicly accessible databases) obtained from direct numerical simulations of two test cases: (a) turbulent channel flow and (b) buoyancy driven flow in a periodic box.

Based on analysis of data obtained from direct numerical simulation of turbulent channel flow and buoyancy driven flow, it is found that significant cointegration relationships exist among several physical quantities. Significant cointegration relationships in the channel flow are found between the instantaneous streamwise velocity (U^+) and streamwise turbulence intensity $u'u'$, $u'u'$ and $v'v'$, and $v'v'$ and $w'w'$. A large number of quantities exhibit significant cointegration relationships in the buoyancy driven flow as well. However,

some disagreement was found in the percentage of cointegration relationships between the two test methods employed (Engle-Granger and Johansen Cointegration Tests). Additional work is needed to fully understand the discrepancy between the various methods and this is one of the suggested future research topics presented.

This work shows that cointegration is a viable method for evaluating potential relationships within turbulent flows and provides identification and quantification of a few relationships between several physical quantities in multiple flow types. The identification of these relationships may lead to a better understanding of the nature of turbulence. More work is needed to refine the application of the various cointegration test methods for turbulence data and to understand the true meaning of the identified relationships.

Chapter 1

Introduction

1.1 Background on Turbulence

Turbulence has often been called one of the last unsolved problems of classical physics and is a phenomenon that impacts a wide variety of fields in nature and engineering [41]. Turbulence is characterized by swirling motion of fluids that occurs irregularly in space and time. Turbulent flows consist of a wide range of length (and time) scales, where the large scales of motion that contain most of the energy (associated with turbulent fluctuations) transfer it through an energy cascade mechanism to the small scales of motion, where the energy is dissipated in the dissipative range [29]. While large scale motions are dependent on the geometry of the flow, small scale motions exhibit scaling behavior and statistics that are universal across a broad range of turbulent flows [30, 39]. Many of the universal statistical features or structures were not only hypothesized in theoretical descriptions of turbulence (especially based on similarity or scaling arguments), but were also subsequently verified in experiments and direct numerical simulations of turbulence.

Several works in the field of turbulence have focused on the discovery and verification of scaling laws in turbulence [3, 47]. This emphasis on exploration of scaling behavior (especially for the small scales of turbulence) is important

not only from the perspective of fundamental understanding of turbulence but also from a modeling perspective. Small scales of motion in turbulence, along with their statistically universal scaling features, appear to be more amenable to modeling. Models of the effects of small scales of motion on the large scales of motion are important for numerical simulation of turbulent flows based on reduced order representations. Such reduced order representations have the potential to enable reliable computational predictions of complex turbulent flows at a relatively low computational cost, in contrast to a fully resolved, direct numerical simulation (DNS) of turbulence which often has a prohibitively high computational cost (especially at high Reynolds numbers where a broad range of length and time scales need to be accurately represented). In the latter case, the computational cost increases with Reynolds number (Re) roughly as $O(Re^3)$.

The fundamental equations governing turbulence are known and are given by the Navier-Stokes equations (assuming an incompressible flow) as

$$\nabla \cdot \mathbf{u} = 0 \tag{1.1}$$

$$\partial_t \mathbf{u} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{f} \tag{1.2}$$

where $\mathbf{u} \equiv \mathbf{u}(\mathbf{x}, t)$ denotes the velocity field and $\mathbf{f}(\mathbf{x}, t)$ denotes the forcing function/mechanism. The symbols p, ρ and ν denote the pressure, density and viscosity respectively. The difficulty of obtaining solutions to these equations is what drives the various analysis methodologies just described.

1.2 Turbulence Closure Problem

Any attempt to develop a set of equations for the statistical quantities of interest for turbulent flows which are described by the Navier-Stokes equations results in a system of equations that is not closed. This problem has become known as the closure problem of turbulence and is a critical issue in the study of turbulence [5]. To resolve this issue, many closure models have been developed that attempt to close the system of equations by relating various physical quantities to one another. These models all reduce the system of equations so that they may be solved but they are also limited in their applicability as they are based on certain assumptions about the flows. For example, one very common model is the κ-ϵ model. This is a two-equation transport model that determines the turbulent viscosity based on the solution of the turbulence-energy and turbulence-energy-dissipation rate transport equations. The basis for this model is that the turbulent shear stress is related to the mean rate of strain as shown in (1.3) below [17].

$$-\rho\overline{u'v'} = \mu_T \frac{\partial u}{\partial y} \tag{1.3}$$

Here, μ_T is defined as:

$$\mu_T = c'_\mu \rho k^{\frac{1}{2}} l \tag{1.4}$$

where c'_μ is a constant. From here an equation for k is solved and the turbulence-energy-dissipation rate equation is solved to find the length scale. This model has proven to be useful in a wide variety of situations but has several disadvantages that limit its use [5]. A further discussion of these limitations and the finer points of this model are beyond the scope of this study.

In addition to the κ-ϵ model, several other models have been developed ranging from algebraic solutions such as the Prandtl mixing length model [33], one equation eddy viscosity model [34], and Rotta second order model using the Boussinesq approximation [38]. Each of these models have their own strengths and weakness and vary in complexity but share a commonality of being based on relationships between various physical quantities. Given this strong dependence of turbulence closure models on relationships amongst various quantities, one potential method for identifying new closure methods may be to analyze turbulent flows for as of yet unidentified relationships.

Kamataki *et al.* examine some of these potential relationships in their work where they evaluate the causal relationship between the sequence of phenomena in turbulent plasma [18]. This is done by evaluating the correlation coefficient and a parameter called directional area which they define. From their analysis they show that an increase in the amplitude of the drift wave results in a drop in the density gradient of the plasma. This work solidified some relationships that were previously only identified visually and provided analytical methods to further understand energy transfer direction in turbulent plasma.

To support the development of closure models, more relationships between various physical quantities may be investigated. In addition to use in closure models, the discovery of relationships between various quantities may be useful in reducing the amount of information that must be measured in experiments/simulations, improve the ability to verify experimental or simulation results, and provide further insight into properties of turbulent flows.

Several recent works have explored the possibility of utilizing methods not typically associated with turbulence analysis to search for new insights. For

example, Faranda *et al.* have investigated the use of Autoregressive Moving Average models in the analysis of various turbulent flow datasets with the intent of relating recorded time series data to stochastic differential equations [10]. They found using this method that it was possible to relate the observed data from an von Kármán swirling flow experiment with Reynolds numbers up to $Re = 2 \times 10^5$ to a stochastic differential equation that described correlation structures within the data. This analysis allowed the researchers to discriminate between flows with comparable mean velocities and analyze flow configurations with multi-stability features.

Other works investigating relationships in turbulent data in include Nikora *et al.* who utilize relationships in stationary turbulent series to enhance measurements with Acoustic Doppler Velocimeter (ADV) systems [31]. These results provide improved performance of ADV measurements and give another example of how identification of relationships in flows may prove useful.

Tao *et al.* also investigate relationships in turbulent flows [43]. Utilizing Holographic Particle Image Velocimetry measurements in a square duct with fully developed flow it is found that there are surprising alignment trends within the filtered vorticity, strain-rate, and subgrid-scale stress tensors. The authors argue that the relationships found in their results should be utilized as a way to check new and more realistic turbulence models.

While all of these examples are encouraging for utilizing new models for evaluating flows and provide examples of how identified relationships may be useful in future turbulence work, they all deal with stationary series or at least make no mention of a requirement regarding the stationarity of the data. However, non-stationary data appears to occur in certain types of turbulent flows and can cause issues when standard analysis techniques are utilized.

Priestly provides an argument discussing the need for analysis methods that can deal with non-stationary series that are often encountered in atmospheric turbulence and other areas [35]. In this work, Priestly develops a method for spectral analysis of non-stationary processes. Along with this development, Thomson has investigated the use of dispersion models in non-stationary turbulent flows [44]. The author makes the argument that several previous studies have attempted to produce models that fail to properly perform in inhomogeneous or non-stationary conditions. These investigations begin to provide some insight into the issues that may be encountered when dealing with non-stationary series and the importance of their occurrence in turbulent systems.

Within these identified works there was no specific work found to investigate relationships amongst physical quantities in non-stationary flows. The remainder of this investigation will provide more analysis of the issues with non-stationary analysis and investigate new methods for evaluating potential relationships in these types of flows.

For some time the theory of turbulence has outpaced the experimental data available in certain areas for testing predictions of properties of turbulent flows and for development and validation of closure models especially in higher Reynolds number flows. Recently there has been an increase in work to create new experimental facilities that can achieve higher Reynolds number flows [22]. In addition to these new facilities, new measurement methods are being developed and utilized that can provide higher resolution sampling of the instantaneous velocity field in these experiments. The availability of high resolution experimental data at higher Reynolds numbers has allowed for the investigation of alternative methods for analysis of turbulent flows and further investigation of theory [43, 42, 45]. Along with the improvements in

experimental work, Direct Numerical Simulation (DNS) of turbulent flows is continually under development and can provide more information for analysis than even the most detailed experiments.

With this additional experimental and simulation data, increased work in investigating potential relationships amongst various physical quantities in turbulent flows is possible. Identification and quantification of these relationships may lead to a reduction in the required data that must be obtained from simulations and experiments and may also shed new light on the nature of turbulence. In addition to the availability of more data for study, analysis methods not traditionally used in the study of turbulence are being applied.

1.3 Objectives and Research Questions

In this thesis, we explore potential relationships that could exist in non-stationary (in time) and/or non-homogeneous (in space) turbulent flows. This work is intended to complement many of the existing works on analysis of turbulent flows that were based largely on the conditions/assumptions of statistical stationarity (in time) or statistical homogeneity (in space). Discovery of the presence (or absence) of relationships between physical quantities in non-stationary (or non-homogeneous) turbulent flows could not only lead to fundamental insights into complex turbulent flows but could also aid in construction of new turbulence models (in non-stationary/non-homogeneous flow configurations), along with potential reduction to complexity of existing models. To explore relationships between physical quantities that vary in space and time in turbulent flows, especially when there are underlying deterministic and stochastic (linear) trends in the signals, we consider the concept of coin-

tegration that has been widely used in the economics literature but has never been used till date in the analysis of turbulence data. The intuition is that there might be a long-run (statistical) equilibrium relationship among a certain class of signals (or physical quantities) if the signals do not drift too far apart from each other. Cointegration analysis can help us discover if there exists a linear combination of non-stationary (and first-order integrated) time-series which results in a stationary (or zeroth-order integrated) time series. Some important research questions explored in this thesis include: (i) Do cointegration relationships exist in non-stationary and/or non-homogeneous turbulent flows?, (ii) If so, what are the cointegration relationships?, (iii) What collections of signals show strong evidence of cointegration relationships? and (iv) What is the utility of cointegration analysis in investigation of turbulent flows?

In order to address these research questions, we consider data obtained from direct numerical simulation (DNS) of turbulence in canonical flows like (a) turbulent channel flow and (b) homogeneous buoyancy driven turbulence in a periodic box. The DNS data for these flow configurations is available via at the Johns Hopkins Turbulence Database (JHTDB) website, which is a portal to an open source numerical turbulence laboratory. In this work, data obtained from these databases are used for investigation of cointegration relationships. In the case of turbulent channel flow, we focus on investigation of cointegration relationships among physical quantities like fluctuations in velocity components, streamwise velocity, pressure and/or their gradients. While cointegration analysis in other fields is often applied to temporally evolving signals, in this thesis we will also use cointegration analysis for spatially evolving signals. In particular, we will analyze cointegration relationships among physical quantities by considering their variation in the wall normal direction,

in the case of turbulent channel flow investigations. In the case of buoyancy driven turbulent flows, we will analyze time series data of signals. In the rest of the thesis, we will consider the usage of the phrase "time-series" in a more general sense to represent variation of signals in time or space depending on the context.

1.4 Organization

The organization of the rest of the thesis is as follows. In chapter 2 we present some background information describing time series analysis and the need for cointegration. In chapter 3 we present an introduction to cointegration and some example synthetic series further illustrating the concept. The application of cointegration testing to turbulent flows is given is chapter 4 and chapter 5 provides a final discussion of the results and overall conclusions. As mentioned earlier, one of the primary goals of this thesis is to use a non-traditional approach of cointegration analysis to explore potential relationships among multiple physical quantities in turbulent flows and our findings on potential cointegration relationships (or the lack thereof) along with the validity and utility of this approach for turbulent flows will be summarized in chapter 5.

Chapter 2

Time Series

In this chapter, we will present a brief introduction to the notions of time series, regression analysis, non-stationarity, unit root testing, spurious regression and integrated series. These topics will help build an appropriate context for cointegration analysis (especially for researchers in fluid turbulence).

A time series can be defined as a series of observations taken over time with a given interval between each observation. Time series are prevalent in analysis where observed series change over time and a great deal of tools have been developed to analyze these series. This type of analysis spans many disciplines including economics, physics, and engineering. As a result, the understanding and study of time series is extremely important.

2.1 Regression Analysis

Regression analysis is a common technique to determine if series have a relationship with one another by determining how a set of independent variables may be related to given dependent variables. If y is taken as the dependent variable and x the independent variable, then a regression analysis will relate

y to a function of x and another term β

$$y \approx f(x, \beta). \tag{2.1}$$

A very common regression method is linear regression which relates y to x using a set of linear parameters β. The simplest example of a linear regression for a time series relates the two quantities using a single parameter and a residual term. For example, for a simple data series consisting of n observations, the relationship can be expressed as

$$y_i = \beta_1 x_i + u_i \quad \text{for } i = 1, \ldots, n \tag{2.2}$$

where u_i is an error term known as a residual. This can be expanded to a more complicated relationship such as

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + u_i. \tag{2.3}$$

Even though this relationship is non-linear in the independent variables, the β parameters are still linear making this a linear regression [40]. As a result of this definition of linearity, there are a wide variety of models that can be explored under the umbrella of linear regressions. However, specifying a model is not of much use unless the parameters can be estimated.

One approach, among a wide variety of options, for estimating β in a linear regression is ordinary least squares (OLS). OLS seeks to minimize the sum of the square of the residuals over a given number of samples T. For the simple

linear regression described in (2.2), the sum of squared residuals is defined as

$$\mathrm{SSR} = \sum_{i=1}^{n} \left(y_i - \beta x_i\right)^2 . \tag{2.4}$$

To complete the estimation, the value of β that minimizes the sum of squared residuals must be found. This minimized parameter, labeled b, can be found by solving (2.5) as [13]

$$b = \left[\sum_{i=1}^{n} x_i x_i\right]^{-1} \left[\sum_{i=1}^{n} x_i y_i\right] . \tag{2.5}$$

Once the minimized sum of residuals is found, a measure is needed to determine the "goodness of fit" for the regression to determine the quality of the estimation. Typically this is done using the square of the multiple correlation coefficient, R^2, parameter which is defined as

$$R^2 = 1 - \frac{\mathrm{SSR}}{\sum \left(y_i - \bar{y}\right)^2}. \tag{2.6}$$

As described by Draper and Smith [7], the R^2 term measures the "proportion of total variation about the mean $\bar{y}$ explained by the regression." R^2 values can range from 0 to 1 with higher values indicated a better fit to the data from the regression. For example, an $R^2 = 0.8124$ would indicate that 81.24% of the variation in the data about the average is explained by the regression. Suitable values of R^2 vary depending on the field of study and the data being evaluated. Simply drawing conclusions based solely on reported R^2 values can drive incorrect conclusions or assumptions and care must be used when evaluating the results of any regression relationship as will be discussed later. Ordinary least squares regression is a relatively straight forward and popular

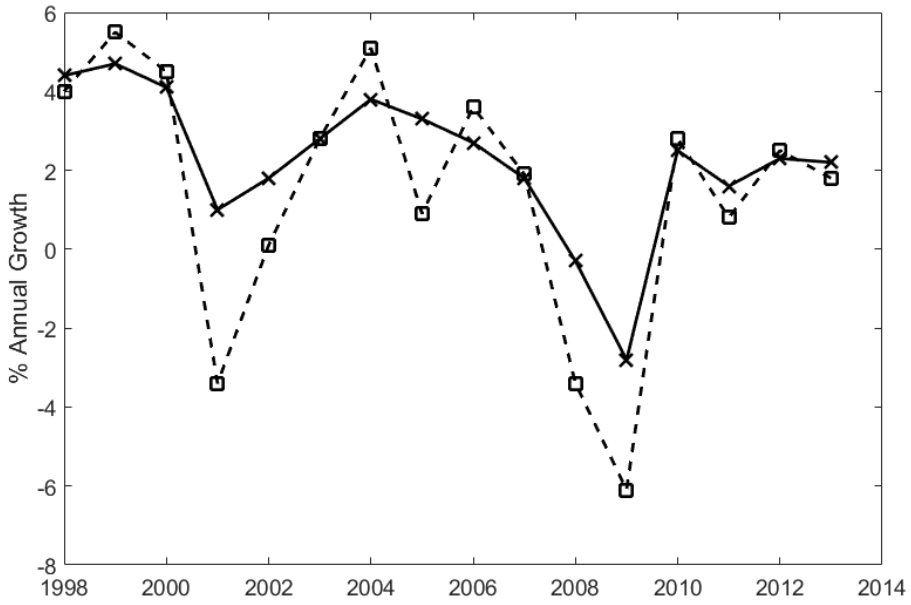


Figure 2.1: United States GDP and industry value added annual percentage growth comparison as reported by the World Bank. GDP annual percentage growth is represented as solid line with X markers. Industry value added annual percentage growth is the dashed line with open box markers. The two series appear to move in a similar pattern suggesting there may be a link between the two processes.

method for determining regression relations and will be utilized as part of methods discussed later.

A simple example of a regression analysis can be illustrated by evaluating the US gross domestic product (GDP) annual percentage growth and US Industry value added annual percentage growth from the World Bank [2]. The two series are shown from 1998 to 2014 in Figure 2.1. Through visual inspection of these series it can be seen that the two series appear to move together and may be related. To determine if this is the case or not, an OLS regression can be computed.

Utilizing the methodology outlined above, the best fit b value is determined

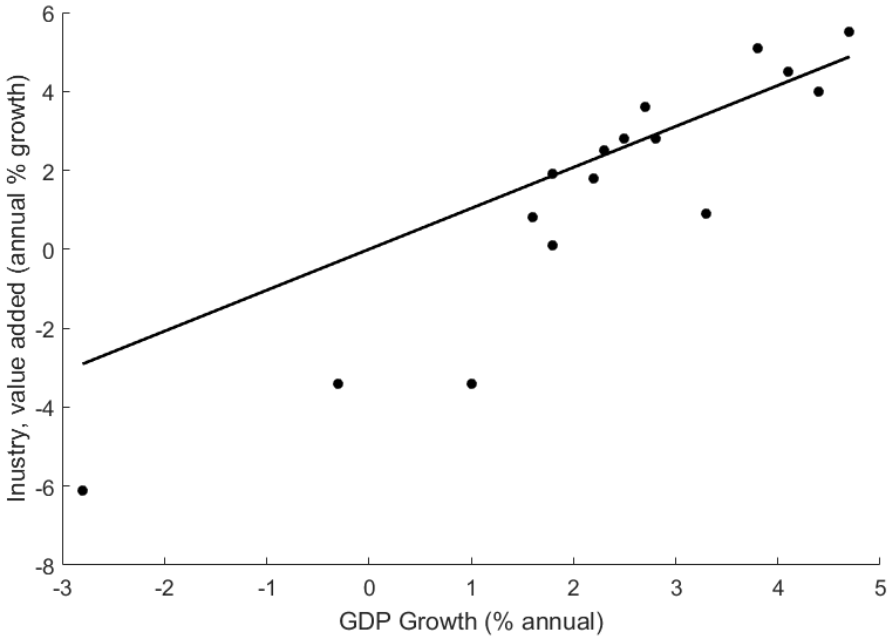


Figure 2.2: Linear regression results of GDP vs industry value added annual percentage growth using linear relationship with no intercept term.

to be 1.0378 with an R^2 value of 0.6742. Figure 2.2 shows scatter plot of the Annual GDP growth vs the Industry Value Added growth along with the OLS regression estimate. The regression relation described in 2.2 can be improved by included a β_0 intercept term in the regression relation to achieve a new regression relation of the form

$$y_i = \beta_0 + \beta_1 x_i + u_i. \tag{2.7}$$

Implementing the regression with the intercept included would result in an R^2 of 0.8621 and suggest a much better fit of the data. Additional examples of regression analyses will be presented later that deal with additional considerations.

There are many more methods for estimating relationships among series

each with their own strengths and weaknesses. Linear regressions can be estimated using methods such as Generalized Least Squares, Total Least Squares, Maximum Likelihood Estimation, Bayesian Linear Regression, and others. In addition to linear regressions, non-linear regressions are also available for data that cannot be explained by a linear combination of parameters. Some of these methods include Exponential Regression, Nonlinear Least Squares, and Binary Logistic Relation just to name a few. The analysis methods discussed in the remainder of this thesis all rely solely on linear regressions using Ordinary Least Squares estimation techniques. As a result, no detailed discussion of the non-linear regression methods is included.

2.2 Stationary and Non-Stationary Time Series

Within the realm of time series, there are two primary types of series that must be distinguished, stationary and non-stationary series. Simply put, a non-stationary series is one whose properties change over time. One type of stationarity is weakly stationary. To be considered weakly stationary, the mean μ and the autocovariances γ_{jt} must not depend on the time of the observation t [13]:

$$E(Y_t) = \mu \qquad \text{for all } t \tag{2.8a}$$

$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_{tj} \qquad \text{for all } t \text{ and any } j. \tag{2.8b}$$

In other words, (2.8) states that the mean is the same regardless of when the observation is taken and that the autocovariances depend only on the distance between the observations and not the date of the observation. If these conditions are not met then the series is considered non-stationary.

In addition to the weakly stationary series previously defined, there are strictly stationary series. A series is said to be strictly stationary if the joint distribution of the series $(Y_t, Y_{t+j_1}, Y_{t+j_2}, \ldots, Y_{t+j_n})$ depends only on the separation of dates $(j_1, j_2, \ldots, j_n)$ and not on the date t. It is much more difficult to prove a series is strictly stationary than it is to prove weak stationarity. For the purposes of this discussion, the term stationary will refer to a weakly stationary processes.

A series can be shown to be non-stationary by demonstrating that it has a unit root. Time series analysis in econometrics is primarily based on difference equations [8]. When solving a difference equation such as $y_t - \sum_{i=1}^{n} a_i y_{t-i} = 0$, a characteristic equation is found for the process. A stationary process is one where the roots of the characteristic equation lie outside of the unit circle. In contrast, a non-stationary process has at least one root that lies on the unit circle. When this occurs the process is referred to as a unit-root process. As a simple example to further explain this concept (adapted from Hamilton [13]), consider a standard autoregressive process modeled by

$$y_t = \rho_1 y_{t-1} + \epsilon_t \tag{2.9}$$

where ρ is an integer value and ϵ_t is a series of independent normal random variables with zero mean and variance of σ^2. The characteristic equation of this process is $\alpha - \rho_1 = 0$ and therefore the system has a unit root when $\rho_1 = 1$. Now, evaluating the process $y_t = y_{t-1} + \epsilon_t$ through repeated substitution

assuming $y_0 = \text{constant}$

$$y_1 = y_0 + \epsilon_1 \tag{2.10a}$$

$$y_2 = y_1 + \epsilon_2 = y_0 + \epsilon_1 + \epsilon_2 \tag{2.10b}$$

$$y_3 = y_2 + \epsilon_3 = y_0 + \epsilon_1 + \epsilon_2 + \epsilon_3 \ldots \tag{2.10c}$$

For all $t > 0$ it follows that

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_t. \tag{2.11}$$

To show that this process is indeed non-stationary, the variance for the process is calculated. Utilizing the property that $\text{Var}(X + a) = \text{Var}(X)$ where X is a random variable and a is a constant, the property that the variance of a sum of random variables is

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j), \tag{2.12}$$

and that by definition the random variables ϵ_t are independent so that their $\text{Cov}(\epsilon_i, \epsilon_j) = 0$, it can be found that

$$\text{Var}(y_t) = \text{Var}\left(y_0 + \sum_{i=1}^{t} \epsilon_i\right) = \sum_{i=1}^{t} \text{Var}\left(\epsilon_i\right) = t\sigma^2. \tag{2.13}$$

As shown in (2.13), the variance of this series is a function of time and therefore this process is non-stationary. This provides a simple example of how a unit root series is non-stationary. More detailed discussion regarding general unit root processes can be found in Hamilton's work [13].

2.2.1 Unit Root Testing

The fact that non-stationary processes are also unit root processes results in several convenient methods for testing series for stationarity. One method commonly utilized to test a series for a unit root is the Dickey-Fuller Test [6]. The basic concept of the test can be illustrated by again starting with a standard auto-regressive process $y_t = \rho_1 y_{t-1} + \epsilon_t$ and then subtracting y_{t-1} from each side. Using the standard delta operator $\Delta y_t = y_t - y_{t-1}$ and defining $\gamma = \rho_1 - 1$, the process can be rewritten as $\Delta y_t = \gamma y_{t-1} + \epsilon_t$. This equation is then estimated using an OLS estimator to obtain an estimate for γ and the standard error of the estimate. These two values are then utilized to obtain a t-statistic value which is used in conjuction with the Dickey-Fuller tables to test the null hypothesis of $\gamma = 0$ (testing $\gamma = 0$ is equivalent to testing $\rho_1 = 1$). The values of the Dickey-Fuller table were calculated using a Monte-Carlo study by Dickey and Fuller. It is important to note that the critical values for the Dickey-Fuller test vary from the standard values. Utilization of the incorrect values can result in erroneous conclusions from the Dickey-Fuller test. Based on the t-statistic values, a probability value (p-Value or pVal) can be determined that gives the probability of a result as or more extreme than the derived estimate. A significance value is then determined and used as a cutoff point for stating whether or not a series is a unit root process.

Expanding on this basic concept, Dickey and Fuller propose three regressive

models used to test a series for a unit root. These models are:

$$y_t = \rho y_{t-1} + \epsilon_t, t = 1, 2, \ldots \tag{2.14a}$$

$$y_t = \mu + \rho y_{t-1} + \epsilon_t, t = 1, 2, \ldots \tag{2.14b}$$

$$y_t = \mu + \beta t + \rho y_{t-1} + \epsilon_t, t = 1, 2, \ldots \tag{2.14c}$$

where $Y_0 = 0$ in all instances, ρ is a real number, and e_t is a sequence of independent normal random variables with zero mean and a variance of σ^2 [8]. Equation (2.14a) represents a random walk process, (2.14b) is a random walk with drift, and (2.14c) is a random walk with drift and deterministic trend. The Dickey-Fuller test is relatively limited in that it was originally developed for a limited set of time series models. Also, it is very important to choose the correct model in the specification of the test as the Dickey-Fuller tables vary depending on the model chosen.

To extend the applicability of the method, the Augmented Dickey-Fuller test was proposed. The Augmented Dickey-Fuller test utilizes the same methodology as the original but employs the regression equation:

$$\Delta Y_t = a_0 + \gamma Y_{t-1} + \sum_{i=2}^{p} \beta_i \Delta Y_{t-i} + e_t \tag{2.15}$$

where p is the order of the autoregressive process, $\Delta\ y_t = y_t - y_{t-1}$, $\gamma = -\left(1 - \sum_{i=1}^{p} a_i\right)$, and $\beta_i = -\sum_{j=1}^{p} a_j$. The Augmented Dickey-Fuller test will be revisited later in this study as it is one of the options for testing the residuals of the cointegration test to determine stationarity and therefore cointegration.

One of the key parameters that must be properly selected when utilizing the Augmented Dickey-Fuller method is the selection of the appropriate lag

length. There are several methods available for selecting the appropriate lag length for a given set of data as described by Enders [8].

The first method is referred to as the general-to-specific method. In this method, a relatively large lag length, p, is selected and tested. If this lag length is insignificant per the usual t-test or F-test then the test is repeated with lag length $p-1$. This process is repeated until a test value that is significantly different than zero is obtained. For purely autoregressive cases, this procedure will find the true lag length with an asymptotic probability of one assuming the true lag length is contained within the initial estimate of p.

An alternative method to the general-to-specific procedure is to use information criteria such as the Akaike Information Criterion (AIC) or the Schwartz Bayesian Criterion (SBC). Both of these schemes seek to estimate parameters of the ARMA model to maximize the log likelihood function while including penalties for each lag parameter estimated. The lag values that minimize the AIC and SBC values are potentially good values to use for Dickey-Fuller testing. It should be noted that each lag added reduces the degrees of freedom as the sample size is reduced by the number of lags used. This is why the information criteria models impose a penalty for the number of lags used [1].

Another popular unit root test is the Phillips-Perron approach [32]. For their test method, Phillips and Perron utilize two least squares regression equations

$$Y_t = \hat{\mu} + \hat{\alpha} Y_{t-1} + \hat{\mu}_t \tag{2.16a}$$

$$Y_t = \tilde{\mu} + \tilde{\beta}\left(t - \tfrac{1}{2}T\right) + \tilde{\alpha} Y_{t-1} + \tilde{\mu}_t \tag{2.16b}$$

where $(\hat{\mu}, \hat{\alpha})$ and $\left(\tilde{\mu}, \tilde{\beta}, \tilde{\alpha}\right)$ are conventional least squares regression coeffi-

cients. Also, there are fewer restrictions on the innovation sequence of the time series Y_t which theoretically increases the applicability of this test method. As was the case with the Dickey-Fuller tests, the t-statistics of the test are critically important and are discussed at length in [32].

Many other approaches have been developed for testing series for unit roots. These include test methods from Sargan and Bhargava, Solo, Dickey and Pantula, Park and Choi, and Schmidt and Phillips just to name a few [13]. These tests will not be discussed here as they are not utilized in the cointegration test methods that will be examined later.

2.3 Spurious Regressions

Regression analysis is a vital tool for economists and other disciplines [12, 14]. Often times researchers apply a regression tool of choice and present results based solely on the R^2 values found during the study or present results despite other indications that the regression may not be applicable. Granger and Newbold discuss this issue where regression analysis with high R^2 values are reported as meaningful even though the reported Durbin-Watson statistic is very low [12]. Granger and Newbold go on to define these relationships as "spurious regressions."

Hendry *et al.* propose a simple theoretical example of how issues in regression may be encountered with non-stationary data [15]. A brief synopsis of their argument is presented here. Considering the regression equation given in (2.2), if x_t is a non-stationary stochastic variable then y_t must share the same stochastic trend if they are to be related in a causal manner. When both x_t and y_t share a stochastic trend, then the residual errors u_t are expected to be

stationary as is typical for regression analysis. If x_t and y_t are not related x_t should have no bearing and it is expected that the hypothesis of $\beta_1 = 0$ should be found to be true. However, if $\beta_1 = 0$ then $y_t = u_t$ meaning that u_t must be non-stationary which violates the initial assumption. This violation of the underlying assumption can create issues when applying this method.

To further see how this may manifest itself an example of a spurious regression is provided in the following paragraphs. Consider two independent random walk plus drift processes, which are non-stationary by definition, generated by the following formulas

$$y_t = 0.2 + y_{t-1} + \varepsilon_{yt} \tag{2.17a}$$

$$x_t = -0.01 + x_{t-1} + \varepsilon_{xt} \tag{2.17b}$$

The relationship between these two processes is then evaluated using a linear regression model of the form

$$\hat{y}_t = \beta_0 + \beta_1 x_t + u_t. \tag{2.18}$$

A single realization of each of these two processes, a scatter plot of the x_t and y_t series with the regression line, and the regression residuals are shown in Figure 2.3. The linear regression for these two series was found to be $y_t = -3.274 + -1.2498x_t$ with an R^2 coefficient of 0.875. By these indications, this is a decent regression of these two series. However, the Durbin Watson statistic found for this regression was only 0.2159. Also, based on their construction, it is known that these are two independent series and it can be seen in the bottom right section of Figure 2.3 that the residuals are non-stationary.

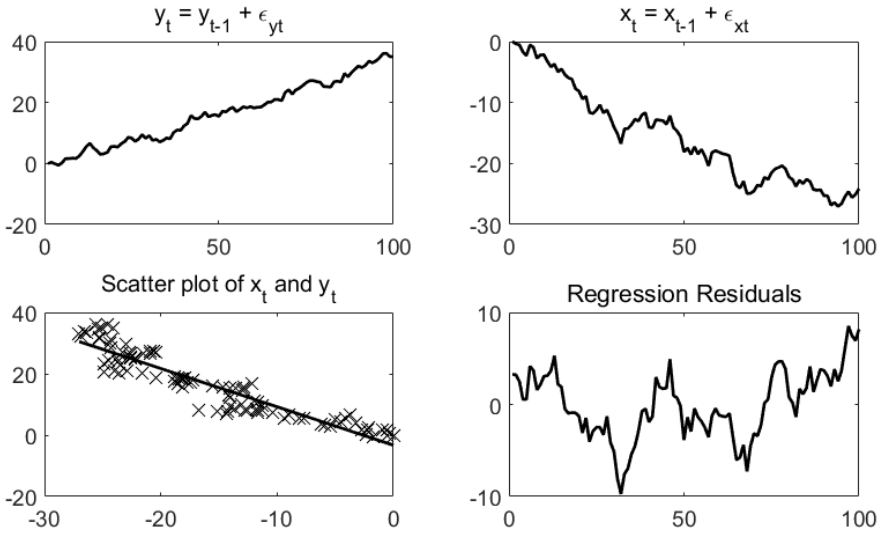


Figure 2.3: Spurious regression example using two random walks with drift and simple linear regression. While these two random walks are independent and not constrained to have any long term relationship, the scatter plot in the bottom left appears to show a strong regression relationship between the series. However, evaluating the residuals from the regression in the bottom right panel reveal non-stationary behavior indicating that the two series exhibit a spurious regression.

This non-stationarity of the residuals was verified using the default Matlab augmented Dickey-Fuller test function which failed to reject the unit root null in the residual series. The fact that the series are independent by construction and the residuals are non-stationary shows that this is a spurious regression and that there is not a significant causal relationship between the two series. This example was specifically chosen to illustrate the possibility of spurious relationships and on its own does not fully demonstrate the magnitude of this potential issue.

Granger and Newbold performed multiple simulations to evaluate the possibility of spurious regressions [12]. Using a similar regression as explained in (2.18), they found that approximately three quarters of the null hypothesis of no relationship would be incorrectly rejected when using a traditional t test at a 5% significance level. They go on to provide further evidence through regressions involving different models that spurious regressions are a major issue in economic and other time series analysis. It is strongly recommended that any regression analysis be accompanied by an evaluation of the residuals using the Durbin-Watson value or some other measure to ensure the regression is meaningful. There have been several methods proposed to address the spurious regression issue especially with regards to non-stationary series. The method of most interest to this study is cointegration as proposed by Granger [11]. More detail regarding cointegration is given in Chapter 3.

2.4 Integrated Series

The concept of an integrated series will also be important in the later discussions of cointegration. A non-stationary series is said to be integrated of order

d, denoted $I(d)$, if after differencing d times the series becomes stationary. Differencing a non-stationary process has historically been used as a method for removing trends from non-stationary time series. Once differenced, the series may be able to be treated as a stationary series and the typical analysis techniques can be used. However, this method may not always be the preferred way to handle non-stationarity in series and additional methods for investigating these cases may be needed. Integrated series will be utilized when defining cointegration methods in Chapter 3.

2.5 Autoregression and Error Correction

As mentioned previoulsy in this chapter, there are alternative representations of time series that allow for additional analysis of systems. An autoregressive (AR) process is one whose values are defined by a sum of the previous values of the series, a constant intercept term, and a stochastic process. A first order autoregression satisfies a first order difference equation of the form:

$$y_t = c + \Phi y_{t-1} + \epsilon_t \tag{2.19}$$

and is denoted as AR(1) [13]. This can be expanded to a pth-order process, AR(p), satisfying the equation

$$y_t = c + \sum_{i=1}^{p} \Phi_i y_{t-i} + \epsilon_t. \tag{2.20}$$

The parameter of the equation, Φ, determines the impact of the ϵ's over time. For $\Phi \geq 1$ the noise accumulates and the process is non-stationary. However, with $\Phi < 1$ the error terms die out over time resulting in a stationary process.

The parameters of the AR process can be evaluated in a multitude of ways such as using OLS as previously described or using the Yule-Walker equations.

While the auto-regressive process is useful for describing many time series models, it falls short when evaluating relationships containing a unit root and for situations where more than one variable are of interest [13]. To examine the utility of autoregressive processes for series with a unit root, the concept of the vector auto-regression (VAR) must be introduced. A vector auto-regression is a simple extension of AR processes to include relationships in a set of variables. The vector auto-regression expresses the series as a sum of the lag values of the series and the lagged values of all other variables. Let y_t be a an $(n \times 1)$ vector of variables.The pth-order vector auto-regression relation, VAR(p), is

$$y_t = c + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \epsilon_t \tag{2.21}$$

Finally, the VAR representation can be expanded to include both short and long term trends in an vector error-correction model (VECM). By introducing lagged differences into (2.21) and re-arranging, the standard error correction representation can be found. The standard form of the error-correction representation VEC(q) is

$$\Delta y_t = C y_{t-1} + \sum_{i=1}^{q} B_i \Delta y_{t-i} + \epsilon_t. \tag{2.22}$$

Utilizing the error correction representation allows modeling of the short term corrections as well as the long term trends in the data. This can provide more insight into the processes than other methods. Estimating VECM parameters is closely related to cointegration testing and will be discussed during the

description of cointegration testing methods in Chapter 3.

It is also possible to represent a VECM in a VAR representation with the appropriate coefficients and order. While this transition from one model form to the other can be very useful in certain situations, it is not necessary for the discussions in this work.

Chapter 3

Cointegration

In this chapter, we will discuss the notion of cointegration and methods to test for cointegration among time-series, along with some examples of cointegrated series.

3.1 Intuition: A Drunk and Her Dog Example

Perhaps the most intuitive explanation of cointegration is the a drunk and her dog example as proposed by Murray [28]. In this paper, Murray considers a drunk woman and her dog that share a common starting point from the same bar. If they both begin to randomly wander they can be assumed to take random walks away from the bar. The drunk and the dog will both meander whatever way they choose, changing directions at a whim. As time goes on, the likelihood that both the drunk and the dog have wandered far from the bar, and from each other, increases. This is equivalent to a variance that grows over time which results in these walks being non-stationary by definition.

Now, instead of just a miscellaneous drunk and dog, consider the case where the dog belongs to the drunk. The drunk and the dog will both again begin wandering in random directions but every so often the drunk calls out to the dog. When the dog hears its owner it attempts to move towards her

until it is distracted by another object and moves off on its own path again. This mechanism of the dog moving back towards its owner when called is a form of error correction that prevents the two walks from wandering too far from one another. Over time, the walks will still randomly move with regards to the bar but they will always attempt to stay close with relation to one another. This concept of two non-stationary processes being linked such that they do not drift away from one another is, in its most basic form, the idea of cointegration.

3.2 Definition of Cointegration

A more formal definition of cointegration is provided by Engle [9]. Let x_t and y_t be vectors of data that are both integrated of order $d > 0$ ($x_t \sim I(d), y_t \sim I(d)$). If there exists a vector a such that the linear combination of x_t and y_t

$$z_t = x_t - ay_t \sim I(d - b) \tag{3.1}$$

is $I(d-b)$, where $b > 0$, then the two series are said to be cointegrated of order d, b ($CI(d, b)$). Or more generally, if x_t is an $N \times t$ vector of time series data, the components of the vector are $CI(d, b)$ if the following two criteria are met:

1. All components of x_t are integrated to the same order $\sim I(d)$.

2. There is a cointegrating vector $\alpha \neq 0$ such that $z_t = \alpha' x_t$ is integrated to a lower order $I(d - b)$ where $b > 0$.

If both of these conditions are true, then the vector x_t is cointegrated, or $x_t \sim CI(d, b)$. Generally there does not exist a cointegrating vector that satisfies all of the requirements. While this method is extensible to series

integrated of any order, the most common, and most applicable to this analysis, is the investigation of series integrated to order one.

Another way to look at this is to consider two non-stationary series, x_t and y_t that are constrained from drifting apart from each other over time. If an estimated long run relationship of the form $y_t = \delta_0 + \delta_1 x_t + u_t$ is assumed, where δ_0 and δ_1 are constants, then u_t can be thought of as a measure of error between the two series. If the error is stationary then the series will stay close to each other over time. This is the basic principal behind the Engle-Granger Cointegration Test that is discussed further in later sections.

It is also possible for there to exists multiple cointegrating vectors if x_t contains N components. In this scenario, there exists a unique cointegrating vector for each one of the relationships and the linearly independent cointegrating vectors, r with $r \leq N - 1$, makeup the $N \times r$ vector α. The rank of α is r which is called the "cointegrating rank" of x_t.

Finally, for cointegrated systems, there exists an autoregressive or error correction representation for the series [9]. This statement is explicitly described by Engle and Granger as the *Granger Representation Theorem:*
Let x_t be a $N \times 1$ vector time series that is $I(1)$. In this case, there will always exist a multivariate Wold representation using the backshift operator B as

$$\Delta x_t = C\left(B\right)\epsilon_t \tag{3.2}$$

where $C(B)$ is a moving average polynomial. Then, if x_t is co-integrated $CI(d, b)$ with $d = 1$, $b + 1$ and with cointegrating rank r:

1. $C(1)$ is of rank $N - r$

2. There exists a vector ARMA representation

$$A(B)x_t = d(B)\epsilon_t \tag{3.3}$$

with the properties that $A(1)$ has rank r and $d(B)$ is a scalar lag polynomial with $d(1)$ finite, and $A(0) = I_N$ (note: I_N is the $N \times N$ identity matrix). When $d(B) = 1$, this is a VAR.

3. There exist $N \times r$ matrices, α, γ, of rank r such that
$\alpha' C(1) = 0, \quad C(1)\gamma = 0, \quad A(1) = \gamma\alpha'$.

4. There exists an error correction representation with $z_t + \alpha' x_t$, an $r \times 1$ vector of stationary random variables:

$$A^*(B)\Delta x_t = -\gamma z_{t-1} + d(B)\epsilon_t \tag{3.4}$$

with $A^*(0) = I_N$.

5. The vector z_t is given by

$$z_t = K(B)\epsilon_t \tag{3.5}$$

$$\Delta z_t = -\alpha'\gamma z_{t-1} + J(B)\epsilon_t \tag{3.6}$$

where $K(B)$ is an $r \times N$ matrix of lag polynomials given by $\alpha' C^*(B)$ with all elements of $K(1)$ finite with rank r, and $det(\alpha'\gamma) > 0$.

6. If a finite vector autoregressive representation is possible, it will have the form given by (3.3) and (3.4) with $d(B) = 1$ and both $A(B)$ and $A^*(B)$ as matrices of finite polynomials.

Engle and Granger provide a thorough proof of this Theorem which will not be recreated here [9].

3.3 Test Methods

3.3.1 Engle-Granger Test

In addition to defining what cointegration is, Engle and Granger proposed a method to test for cointegration between series. The Engle-Granger Cointegration (EGCI) test method is a two step estimator utilizing a null hypothesis test that evaluates the null condition of no cointegration. To achieve this, the series of interest are regressed using an OLS regression process and the residuals of the regression are then evaluated for stationarity. If the regression residuals are found to not contain a unit root, and are therefore stationary, the null hypothesis of no cointegration is rejected. The regression coefficients can then be utilized to estimate the error correction relationship of the data.

The Engle Granger test methodology is straight forward and simple to estimate. The primary implementation of this method used for cointegration testing throughout the remainder of this discussion is the EGCI test method implemented in Matlab [24]. This implementation of the Engle-Granger test method follows the basic flow as outlined above except it uses a linear least squares method in the regression and allows the user the ability to choose what unit root test is used when evaluating the regression residuals for stationarity. While either of the options including Augmented Dickey-Fuller (ADF) test or the Phillips-Perron test could be used in this implementation, the default ADF test was utilized for all investigations discussed in this work. Also, in this implementation the EGCI test contains no initial testing of the series of

interest for a unit root and it is left to the user to ensure the data being investigated is appropriate for the methodology being used.

While the EGCI test is useful, it suffers from some drawbacks. The test requires the use of unique test statistics as opposed to the standard t test values and suffers from low power issues due to the use of unit root testing. Another limitation of the EGCI test is that it can only determine a single cointegration relationship between two variables. Since the introduction of cointegration, several alternative test methods have been developed to address these limitations and more.

3.3.2 Johansen Test

One of the common alternative methods to expand the EGCI test is the Johansen test method [36, 37]. The Johansen method expands upon the Engle Granger method previously discussed by evaluating the time series for multiple cointegration vectors, determining how many cointegrating vectors exist, and estimating each of the relationships. The Johansen test method is based on using maximum likelihood estimates of the parameters in a VEC model of the cointegrated series. The null of cointegration of rank less than or equal to r is evaluated against the alternative. The estimated VEC(p) model is

$$\Delta y_t = \pi y_{t-1} + \sum_{i=1}^{p-1} \pi_i \Delta y_{t-i} + \epsilon_t \tag{3.7}$$

where $\pi = -\left(I - \sum_{i=1}^{p} A_i\right)$ and $\pi_i = -\sum_{j=i+1}^{p} A_j$.
Using this VEC model, the rank of π can be estimated which provides the number of independent cointegrating vectors as previously discussed. There

are a variety of methods for estimating π and a detailed description of them can be found in a variety of texts.

As was the case with the EGCI test, Matlab has a built in Johansen test procedure that is utilized when evaluating series for cointegration [25]. This test package (denoted as JCITest) evaluates a VEC(q) model of the form

$$\Delta y_t = Cy_{t-1} + B_1\Delta y_{t-1} + \cdots + B_q\Delta y_{t-q} + DX + \epsilon_t \tag{3.8}$$

with $C = AB'$, where A is a matrix of error correction speeds and B is a matrix of basis vectors for the cointegrating relations. The user is able to specify the "mode" used which determines the form of the $Cy_{t-1} + DX$ component. The default model is $A(B'y_{t-1} + c_0) + c_1$ which is utilized in all testing unless otherwise specified. The user is also able to select the number of lags (q) to use.

3.4 Examples of Cointegrated Series

To further illustrate the concept of cointegration, some examples of cointegrated series are shown below. Consider the bivariate system described by (3.9)

$$y_{1,t} = \gamma y_{2,t} + \epsilon_{1,t} \tag{3.9a}$$

$$y_{2,t} = y_{2,t-1} + \epsilon_{2,t} \tag{3.9b}$$

where $\epsilon_{t,1}$ and $\epsilon_{t,2}$ are independent standard normal distribution noise series. The equation for y_2 is a simple random walk expression which can be differ-

enced one time to yield

$$\Delta y_{2,t} = \epsilon_{2,t} \tag{3.10}$$

which is a stationary series since $\epsilon_{2,t}$ is defined to have a mean of zero and a variance of 1, $N(0,1)$. Therefore, (3.9b) is found to be an $I(1)$ series. Differencing (3.9a) and then substituting (3.10) for $\Delta y_{2,t}$ results in

$$\Delta y_{1,t} = \gamma \Delta y_{2,t} + \Delta \epsilon_{1,t} = \gamma \epsilon_{2,t} + \epsilon_{1,t} - \epsilon_{1,t-1}. \tag{3.11}$$

The right hand side of (3.11) is the sum of a moving average process, MA(1), and white noise. Again, this series is a stationary series and therefore (3.9a) is also an $I(1)$ process. Since both of these series are integrated to the same order, they satisfy the first requirement for cointegrated series. The linear combination of these two series $(y_{1,t} - \gamma y_{2,t})$ is

$$(y_{1,t} - \gamma y_{2,t}) = \gamma y_{2,t} + \epsilon_{1,t} - \gamma y_{2,t} = \epsilon_{1,t}. \tag{3.12}$$

As was the case in (3.10), the noise process $\epsilon_{1,t}$ is stationary, and therefore the two series in (3.9) are cointegrated. A sample realization of these two processes with 100 timesteps and $\gamma = 2$ is shown in Figure 3.1.

These two series were evaluated using the EGCI test method as implemented by Matlab. The Engle-Granger cointegration test with the lags parameter set to 0, rejected the null hypothesis in favor of the alternative of cointegration with a p-value of 0.001 which is well below the chosen significance level of 0.05. Repeating the test using 1 and 2 lags resulted in rejection of the null in favor of cointegration with a p-value of 0.001 and 0.0012 respectively. The results of the EGCI test agree with the theoretical analysis that

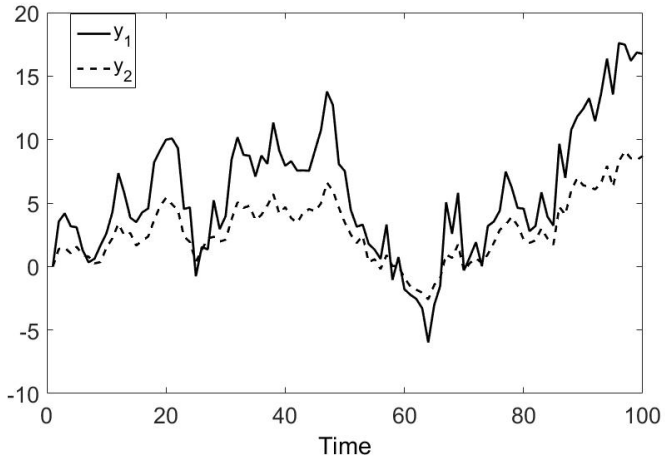


Figure 3.1: Cointegrated series example using bivariate system as described in (3.9). The error-correction between the two series is evident in the similar pattern of movements between the two series.

these two series are cointegrated. Figure 3.2 shows the regression relation between the two series along with the residuals which demonstrate the stationary nature of the residual series.

These series can also be evaluated using the Johansen test method. The test rejects the null of cointegration of rank 0 (no cointegration) and fails to reject the null of cointegration of rank 1 corresponding to a single cointegration relationship as would be expected for two series of interest. The estimated parameters of the model described in (3.8) are:

$$A = \begin{bmatrix} -0.0107 \\ -0.8826 \end{bmatrix}$$

$$B = \begin{bmatrix} -1.9806 \\ 1.0308 \end{bmatrix}$$

$$c_0 = -0.3967$$

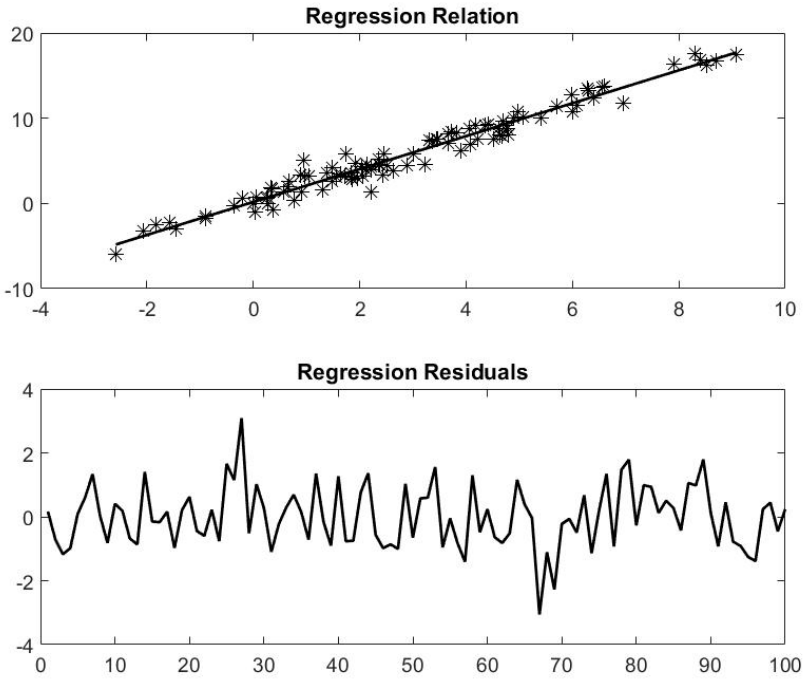


Figure 3.2: Linear regression results of bivariate system described in (3.9). The cointegration between the two series is evident in the stationarity of the regression residuals in the bottom panel.

$$c_1 = \Big[0.0859 - 0.0010\Big]$$

An alternative to this result can be evaluated by returning to the example given in Section 2.3 of two random walk processes with drift. As was previously established, these two series present a spurious regression when evaluated with standard regression techniques. Therefore it is expected that the cointegration tests should reveal no relationship between these two series as this is exactly the scenario this test was designed for. Utilizing the EGCI test with lag values of 0, 1, and 2 resulted in a failure to reject the null of no-cointegration in all cases as expected with p-values well above the 0.05 significance threshold at 0.3461, 0.4453, and 0.4380 respectively.Additionally, the JCI test fails to reject the null of a rank less than or equal to zero meaning no cointegration was found between the series.

While the EGCI test helps to alleviate the issue with spurious regressions it is not perfect. 500 pairs of standard random walks were generated and tested for cointegration utilizing the EGCI test with lag parameters from 0-10. An example of one of the realizations of these random walk pairs is given in Figure 3.3. It was found that up to 6% of the pairs of random walks resulted in a positive cointegration result. As these series were constructed to be independent of one another, the ideal result would have been to find 0% in testing. This information will be utilized in later testing to establish a threshold for determining if there is enough evidence to claim that there is a relationship between various series.

As can be seen by the previous discussion and examples, cointegration is a useful method for finding meaningful relationships between various series when those quantities are non-stationary. Cointegration testing as implemented by

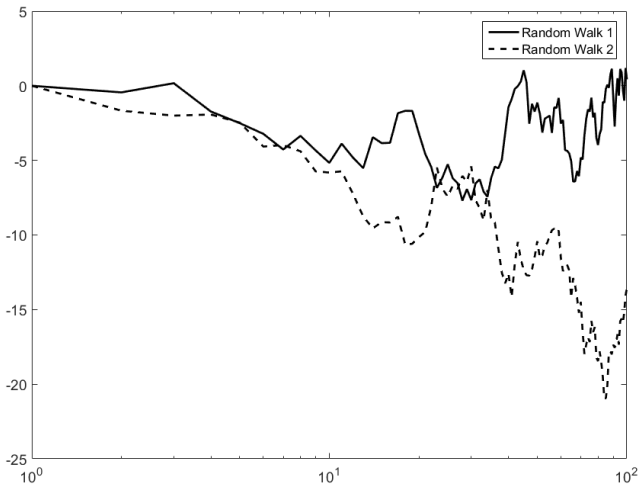


Figure 3.3: Example of random walk pair sample realization showing lack of cointegration

the Engle-Granger test resolves some of the spurious regression issues that were previously identified and provides means for forecasting future relationships. The Johansen method expands upon the EGCI test methodology and alleviates some of the issues that have been identified with the more simple Engle-Granger method. Also, the Johansen test provides an easy and convenient way to estimate a VEC model for the series of interest. The rest of this paper will utilize cointegration testing and relationships to attempt to find relationships between various physical quantities in several types of turbulent flows.

Chapter 4

Cointegration in Turbulence

As previously mentioned, the investigation of relationships between various physical quantities in turbulent flows is of interest to (i) facilitate finding new closure models, (ii) reduce the burden of required data collection and storage for both simulation and experiment, and (iii) gain a better understanding of the basic properties of turbulence. For this investigation a variety of sources of data will be utilized from both experimental measurements and direct numerical simulations (DNS).

While the basic principles of cointegration and the associated test methods have been covered in the previous sections, there are some unique aspects of the data that will be evaluated in this section that require additional investigation.

4.1 Turbulence Data Specific Cointegration Investigation

So far all of the analysis and discussion has been conducted using standard time series with uniform linear spacing between observations. As will be seen during the later analysis of wall bounded flows, turbulence data is often presented with non-uniform, and sometimes logarithmically spaced, observation

locations taken over a spatial dimension instead of time. It must be determined if these differences will have any impact on the results of the cointegration testing before continuing with the use of this method in the analysis of turbulence data.

4.1.1 Non-linear Spacing in Data

The first issue is how the difference in observation spacing may impact the regression portion of the two step EGCI test. The standard linear regression models are independent of the spacing between the observations and simply attempt to minimize the residual error at each observation. As a result, the logarithmic or other non-uniform spacing in turbulence data should be of no impact. To further test this theory two simple series were created. A series, x, of integers from 1 to 100 was initially created. Based on x, series y was then generated using the equation

$$y = 10 + 2x + \epsilon \tag{4.1}$$

where ϵ is a normally distributed noise term with mean equal to 0 and standard deviation of 30. A realization of this series was created and a scatter plot of the data is given in Figure 4.1.

A regression was then performed on the x and y series to obtain estimated values for the coefficients. While this is a needless regression since the relationship between these two series is already known, it serves to provide a control case for regressions and the ability to compare results with alternative series to be discussed later. The regression resulted in a R^2 value of 0.7979 and estimated coefficients of $\beta_0 = 9.9424$ and $\beta_1 = 2.0073$ which provide a

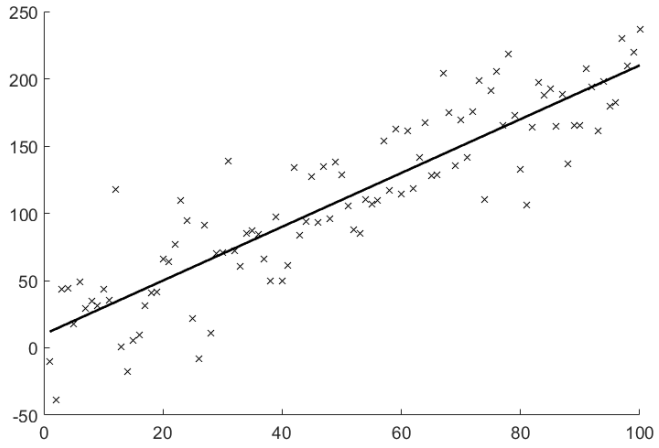


Figure 4.1: Simple linear regression showing results of linear regression of x and y as defined in (4.1) as a control

good estimation of the original parameters from (4.1). The original $x - y$ scatter plot and the deterministic trend without the noise ($y = 10 + 2x$) are given in Figure 4.1. This control example established the performance of the regression analysis with uniformly spaced data. Next, instead of using the linearly spaced x values, logarithmically spaced x values from 1 to 100 were utilized. The regression analysis was repeated and resulted in parameters of $\beta_0 = 10.7696$ and $\beta_1 = 1.9790$ and an R^2 value of 0.7499. Figure 4.2 shows the results of this series and regression analysis. Both the R^2 and estimated parameters are very similar for both the linear and logarithmically spaced series. These results provide additional information supporting the claim that the spacing between the points is irrelevant. However, this is a just a single empirical example demonstrating the concept that the spacing between points is not of concern and a more thorough proof of this should be developed to ensure the methods discussed are still applicable with arbitrarily spaced data.

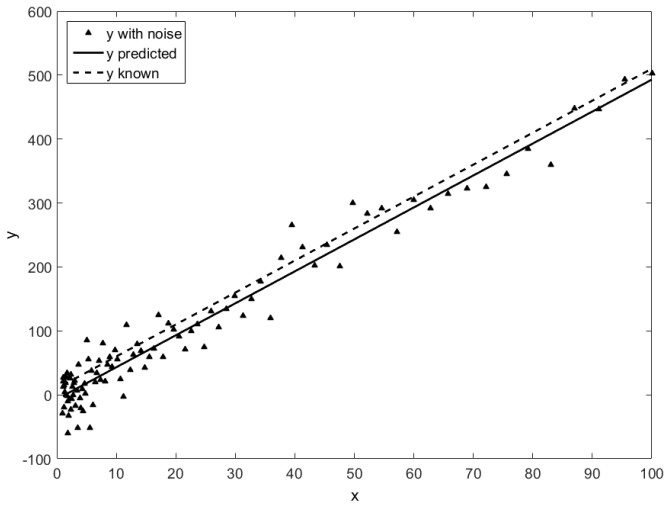


Figure 4.2: Simple linear regression showing results of linear regression of logarithmically spaced x points vs y as defined in (4.1). The estimated regression relation is similar to the known true relation and presents a good fit of the data regardless of the logarithmic spacing between the points in the x and y series.

4.1.2 Interpolation Impacts on Cointegration Testing

Along with this analysis, it was hypothesized that the non-uniformly spaced data could be interpolated to achieve equal spacing and make the data more consistent with a traditional time series. Ultimately it was discovered that interpolating the data impacted the final test results and increased the positive percentage of cointegration relationships identified when using the EGCI test method. To illustrate this, additional synthetic series were developed and various methods of interpolation were applied.

The two series used for the interpolation investigation are based on the analytical expressions for the average streamwise velocity and turbulence intensity which will be discussed in detail later in this chapter. The equations

are given as

$$\frac{1}{\kappa}\log\left(y^{+}\right) + A + \epsilon_1 = \text{Average Streamwise Velocity} \tag{4.2a}$$

$$B_1 - A_1\log\left(\frac{y}{\delta}\right) + \epsilon_2 = \text{Average Turbulence Intensity} \tag{4.2b}$$

where κ, A, A_1, and B_1 are constants to be defined later and ϵ_1 and ϵ_2 are noise terms that will take various forms for this interpolation investigation.

A variety of forms of noise series were used in this analysis to ensure that the noise distribution chosen was not the cause of the change in cointegration results and to see if any of these noise distributions were a good simulation of the channel flow data. The noise distributions added to the above equations are a uniformly distributed noise, normally distributed noise with a mean of zero and an expected value of 0.5, a random walk, and a reversed random walk. The reverse random walk was created by generating a standard random walk series and then the series was flipped so that the last point became the first and the first became the last. A series of standard random walks excluding the theoretical profiles were also tested as a control.

A total of 60 series for each type of noise were generated, added to the theoretical profiles, and then tested for cointegration using the standard EGCI and JCI tests as previously discussed. It should be noted that there is significant disagreement between the EGCI and Johansen test methods regarding the cointegration between these two series. As will be discussed later in this work, more analysis is needed to fully understand the driving force behind this disagreement and what the impact to the results of the cointegration testing truly are. Appendix A provides sample realizations of each of the series along with tables featuring the cointegration testing results for all testing discussed

in this section.

The series were then interpolated using a variety of interpolation techniques and the cointegration testing was repeated. The interpolation methods investigated are a cubic spline interpolation, linear interpolation, and piecewise cubic hermite interpolating polynomial (PCHIP). It was found that with the EGCI test, the interpolated data resulted in higher percentages of cointegration especially in the odd lag cases for all types of noise and interpolation technique investigated. The impact on the Johansen testing was not as straight forward. The percentage of identified cointegration relationships (rejection of the null of rank 0 and failure to reject the null of rank 1) vary for each noise type added. For the uniform noise, the percentages decline with the use of the interpolation schemes. However, with normally distributed noise, the percentages sometimes increase or decrease depending on the lag value evaluated. The two random walks added to the deterministic series appear to not be significantly impacted by the interpolation methods investigated. Finally, the pure random walks exhibit an increase in percentage of cointegration when interpolated. The exact reason for these varying behaviors is unknown at this time.

Given the previous discussion that the regression analysis (the first step of the two step EGCI estimator) is independent of the spacing of the original series, it would be expected that similar results would be obtained with or without interpolation if the various methods of interpolation were valid for use in this scenario. However, as shown in the various tables in Appendix A, the interpolated results present much higher percentages of cointegration than the non-interpolated series with the EGCI test method and varying results with the Johansen method. For the EGCI test, it is hypothesized that this increase in cointegration may be due to an introduced covariance in the data from

the regression that is then impacting the unit root test after the regression. As a result, the final analyses will all be conducted on non-interpolated data directly from the measurement results/simulation data bases.

A large amount of cointegration testing of channel flow data was performed prior to observing the impact of interpolating the data. This testing was conducted in the same manner as the analysis described in the following sections. However, due to the impact of interpolation that was just identified, the results are not presented in the main body of this work. The results of this testing are provided in Appendix B for reference.

4.2 Cointegration Testing of Turbulent Flow Data

Recently Hultmark, Marusic, and McKeon have performed analysis on the logarithmic region of flows in wall bounded turbulence and other turbulent flows [16, 23, 26]. Their investigations centered on the universal nature of the logarithmic region but also show a strong relationship between the mean channel flow velocity ($\overline{U}^+$) and the average turbulence intensity $\left(\overline{u^2}^+\right)$. (Note: Throughout this investigation the term turbulence intensity and Reynolds Stress (R_{uu}) are used interchangeably and defined to be the same physical quantity.)

Marusic and McKeon investigate the logarithmic region profiles over a large Reynolds number range and find that the flows closely follow an analytically predicted logarithmic profile [23, 26]. Along with the agreement with the predicted profile, the relationship between $\overline{U}^+$ and $\overline{u'^2}^+$ can be clearly seen in their work. This relationship that was one of the motivations for additional study into the use of cointegration techniques in turbulence analysis.

While the relationship between the two physical quantities is visually apparent in the Hultmark data, the authors did not address this relationship directly and there appears to be no quantitative measure of the link between streamwise velocity and turbulence intensity. Cointegration testing will be utilized to attempt to provide this quantitative analysis and as a test case to evaluate the utility of this method for investigating additional relationships.

4.2.1 Cointegration Testing of Wall Bounded Turbulence Data

To this point cointegration has been discussed solely in the context of time series. However, the physical quantities of interest just introduced are commonly studied over the dimensionless distance from the wall, y^+, for a given time. From this point forward, y^+ is defined as

$$y^+ = \frac{u^* y}{\nu} \tag{4.3}$$

where u^* is the friction velocity, y is the distance to the wall, and ν is the kinematic viscosity. To relate the time series definition to the data specified in the various data sets, the observations taken over time can instead be looked at as observations taken over y^+. This substitution of the y^+ dimension for the time dimension allows cointegration to be utilized to study the potential relationships between these physical quantities in a channel flow over a spatial dimension instead of time.

Marusic shows that the logarithmic profiles for $\overline{U}^+$ and $\overline{u^2}^+$ are given by

(4.4) and (4.5) respectively. In the logarithmic region, $\overline{U}^+$ is

$$\overline{U}^+ = \frac{1}{\kappa}\log(y^+) + A \tag{4.4}$$

where $\overline{U}^+ = \langle U/U_\tau \rangle$, U_τ is the friction velocity, ν is the kinematic viscosity, κ is the von Kármán constant, and A is a parameter that depends on the surface roughness and is assumed to be a constant for smooth walled flows. Similarly, $\overline{u^2}^+$ is

$$\overline{u^2}^+ = B_1 - A_1 \log\left(\frac{y}{\delta}\right) \tag{4.5}$$

where $\overline{u^2}^+ = \overline{u^2}/U_\tau^2$, δ is the channel half-height (or pipe radius, or boundary-layer thickness), and A_1 and B_1 are constants.

These predicted profiles are not a stochastic process and the use of cointegration to evaluate them does not immediately seem relevant. However, if independent normally distributed noise processes are added to these profiles to simulate experiment and measurement error, then the use of cointegration is more applicable. This addition makes the two series into simple trends with a deterministic component and a noise process.

An initial analytical evaluation of the potential cointegration relationship can be performed. Utilizing a cointegrating vector of $\alpha = (\kappa, 1/A_1)$, the linear sum of these two processes is

$$\kappa\left(\frac{1}{\kappa}\log\left(y^+\right) + A + \epsilon_1\right) + \frac{1}{A_1}\left(B_1 - A_1\log\left(\frac{y}{\delta}\right) + \epsilon_2\right). \tag{4.6}$$

Substituting $y^+ = yU_\tau/\nu$ and simplifying results in

$$\log\left(\frac{yU_\tau\delta}{y\nu}\right) + \kappa A + \frac{B_1}{A_1} + \kappa\epsilon_1 + \frac{\epsilon_2}{A_1} \tag{4.7a}$$

$$\log\left(\frac{U_\tau\delta}{\nu}\right) + \kappa A + \frac{B_1}{A_1} + \kappa\epsilon_1 + \frac{\epsilon_2}{A_1} \tag{4.7b}$$

$$C_1 + \kappa\epsilon_1 + \frac{\epsilon_2}{A_1} \tag{4.7c}$$

where C_1 is a constant equal to $\log(U_\tau\delta/\nu) + \kappa A + B_1/A_1$. The constant C_1 has no bearing on the stationarity of this series and it can be shown that the sum of two independent, normally distributed process is a normally distributed process with mean and variance equal to the sum of the means and variances of the original series. This results in a final series that is stationary and therefore the two series presented in 4.6 can be said to be cointegrated. There is an issue with this analysis that the series are deterministic trends with noise and not fully stochastic processes which are not necessarily unit root processes. That being said, this analysis does provide some initial insight that there is a relationship between these two series. Also, it is not sufficient to rule out cointegration between the actual experimental or simulation data as it is not guaranteed that this is the data generating process for the instantaneous values of these physical quantities.

Example realizations of these two series can be created to visually inspect the potential relationship. The logarithmic profiles from (4.4) and (4.5) are used with the constants and parameters for the LCC experiment data as described in Marusic [23]. For this case, $\kappa = 0.39, A = 4.3, A_1 = 1.26, B_1 = 2, Re_\tau = 68780, \delta = 0.1135, U_\tau = 0.5884,$ and $\nu = 9.710x10^{-7}$. Two independent normally distributed variables with a mean of 0.1 were added to each of

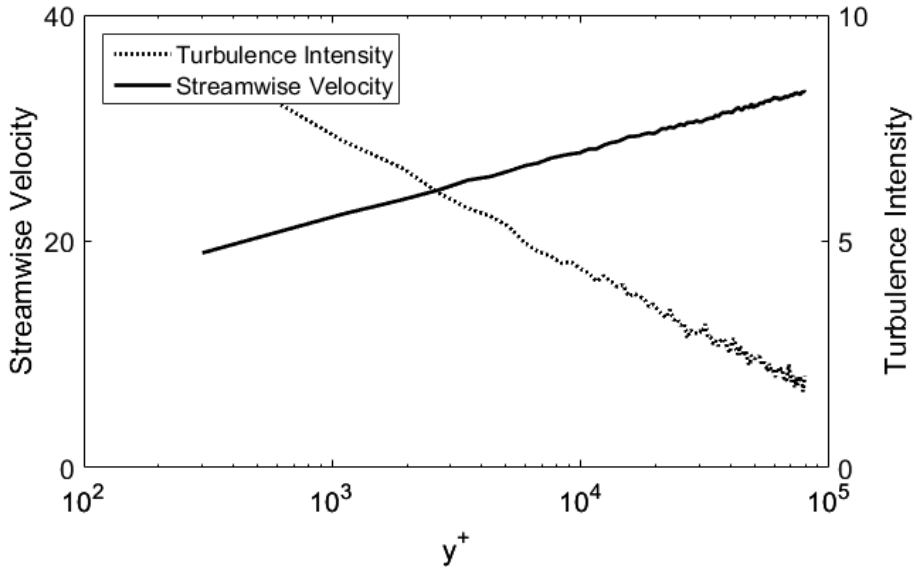


Figure 4.3: Theoretical average channel flow streamwise velocity and turbulence intensity in logarithmic region with normally distributed noise added

the profiles and the resulting series are shown in Figure 4.3.

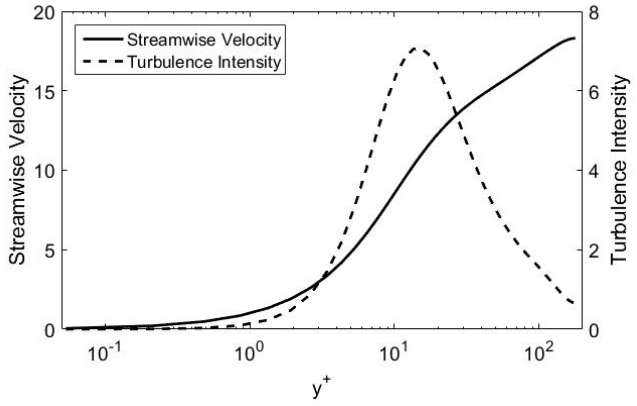
The mean velocity and turbulence intensity only obey these trends in the logarithmic region of the wall bounded flows. There are several definitions of these regions and, as will be shown later, changing the start and stop points of this region can have large impacts on the results of cointegration testing. One example definition of the log region in channel flow is given by Davidson [5] who defines the logarithmic region as the area where $y^+ \gg 1$ and $\eta \ll 1$. However, Marusic points out that several authors have proposed estimates for the start of the logarithmic that range from $y^+ > 30$ to $y^+ > 600$. After evaluating data from several laboratory flows at high Reynolds numbers, Marusic settles on a range of $3Re_\tau^{1/2} < y^+ < 0.15Re_\tau$ for turbulent wall bounded flows [23]. This definition still has its drawbacks as it is only valid for Reynolds numbers greater than 400. These varying perspectives on the applicable logarithmic regions illustrate the difficulty in defining a universal region and a clear region

for use in the varying flows studied herein. For the purposes of this study, the logarithmic region of interest will be defined for each dataset evaluated and the impacts of varying this region will also be studied. The raw data from the Hultmark and Marusic studies was not readily available and therefore no additional investigation was performed. There were several other datasets with raw data available that were used to study this relationship in further detail which are discussed below.

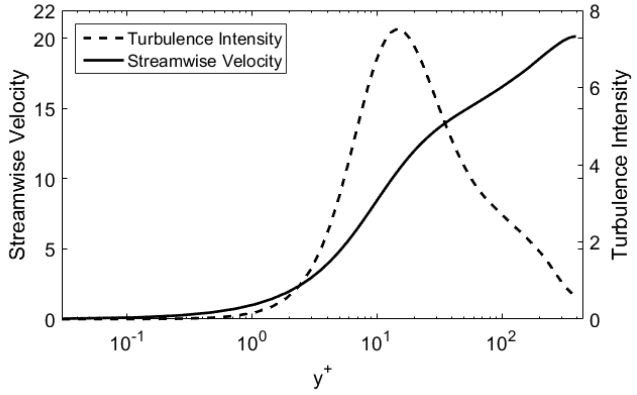
4.2.2 Turbulent Channel Flow Data Analysis: Moser Database

Moser *et al.* performed direct numerical simulations of channel flow at multiple Reynolds numbers utilizing a Chebychev-tau formulation in the wall normal direction, a Fourier representation in the horizontal directions, and periodic boundary conditions in the streamwise and spanwise directions [27]. The simulations were conducted with a planar channel flow at friction Reynolds numbers of 180, 395, and 590. As pointed out by Moser, significant low Reynolds number effects are present in the Re=180 case and will be considered when evaluating the cointegration results. Figures 4.4a to 4.4c show the mean velocity and turbulence intensity profiles over the full y^+ half channel height region and demonstrate the log-law behavior of the data.

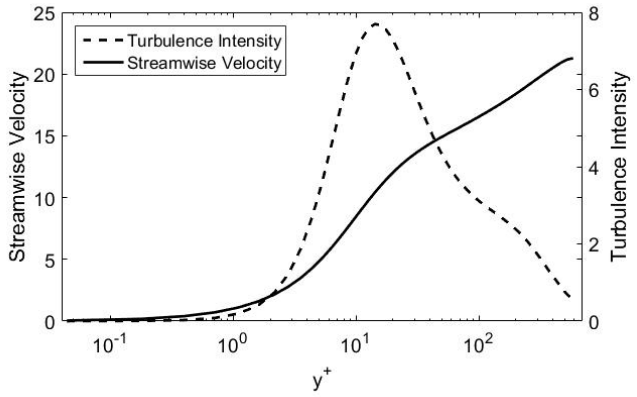
Initial analysis was performed using a logarithmic region of $14 < y^+ < 178$ for $\mathrm{Re}_\tau = 180$, $20 < y^+ < 300$ for $\mathrm{Re}_\tau = 395$, and $25 < y^+ < 600$ for $\mathrm{Re}_\tau = 590$. The average streamwise velocity and average turbulence intensity are shown for these areas in Figures 4.5a to 4.5c. The series in the logarithmic regions appear to show a similar relationship as was previously discussed in the Hultmark and Murasic data.



(a) $Re_\tau = 180$

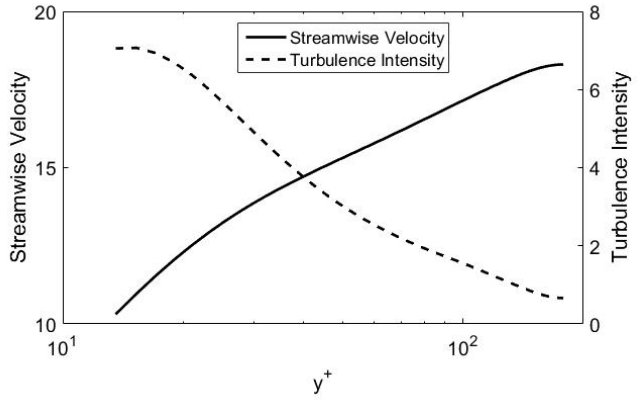


(b) $Re_\tau = 395$

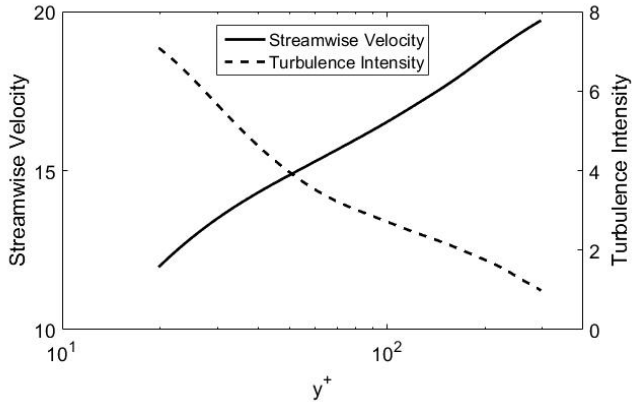


(c) $Re_\tau = 590$

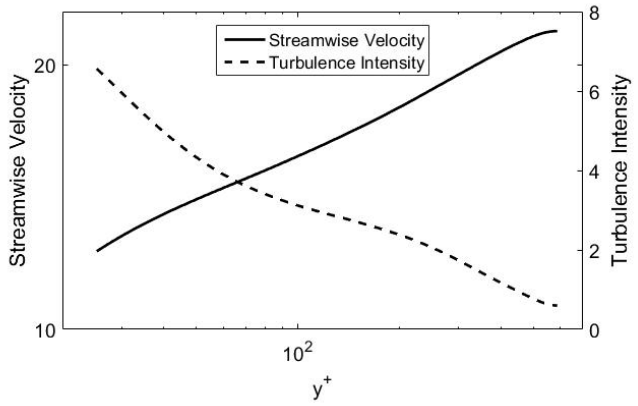
Figure 4.4: Moser channel flow average streamwise velocity and turbulence intensity over half channel height for available Reynolds numbers



(a) $Re_\tau = 180$



(b) $Re_\tau = 395$



(c) $Re_\tau = 590$

Figure 4.5: Channel flow average streamwise velocity and turbulence intensity in logarithmic region for available Reynolds numbers

As Moser provided raw average data in electronic form, a more detailed analysis of the relationships in these flows was possible. To evaluate the potential relationship between the streamwise velocity and turbulence intensity, the two series were tested for cointegration. Before the series can be evaluated they must be tested for a unit root to ensure they satisfy the initial condition that both series are integrated of the same order. For each Reynolds number flow, the streamwise velocity and turbulence intensity were tested for a unit root using the Augmented Dickey-Fuller test. Somewhat surprisingly, the ADF test rejected the null of a unit root for the turbulence intensity for all three Reynolds number simulations. This indicates that per the ADF test, the turbulence intensity series do not contain a unit root. This lack of a unit root could be due to the fact that the averaged data has little stochastic content remaining and behaves more as a deterministic series. Even though this result suggests that these two series are not suited for cointegration testing, the testing was still conducted to evaluate the relationship between the series on the assumption that the data is truly non-stationary and the lack of a unit root from the ADF test is a peculiarity of the test and not a true result. The ADF test can suffer from a lack of statistical power and may not correctly identify a unit root in all cases.

As discussed in Section 2.2.1, it is critical to select the appropriate lag length for use in ADF test which is used in the second step of the EGCI test. To evaluate this, the information criteria method was employed. The Akaike (AIC), Bayesian (BIC), and Hannan-Quinn (HQC) information criteria were evaluated for lags from 0 to 20 when testing average streamwise velocity and turbulence intensity for all three Reynolds number simulations. All three of the information criteria obtained their minimum values at either 2 or 3 lags

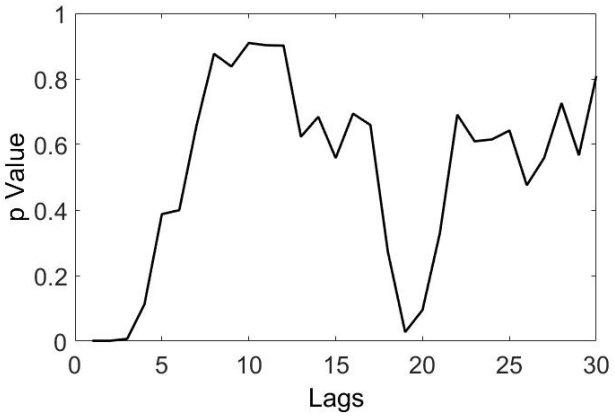


Figure 4.6: pValue result vs lags for general-to-specific lag value determination using Moser channel flow data. This methodology could result in choosing a lag of 19 or less than 4 depending on when the general-to-specific methodology is considered complete.

indicating that these are the most appropriate lag values to use for this data.

Applying the general-to-specific methodology as described by Enders starting with a lag length of 30 and working down to 1 showed that the pValues of the EGCI test showed consistent decrease toward the critical value when the lag value was below five. A dip was observed at 18 lags but quickly returned to the higher pValues. A plot of the p Values vs number of lags used is shown in Figure 4.6. This is in agreement with the lag parameters identified using the information criteria methodology as well.

4.2.2.1 Engle-Granger Cointegration Testing of Turbulent Channel Flow: Moser Database

The series at each of the provided Reynolds numbers was tested for cointegration using the standard Matlab Engle-Granger cointegration test method in the Econometrics Toolbox. The default parameters were used in the test

including the standard significance level of 0.05. A test decision of 1 indicates rejection of the null in favor of the alternative of cointegration while a decision of 0 indicates a failure to reject the null and a result that the series are not cointegrated. Testing was repeated for lag values of 0-10 to evaluate the impact of lag choice beyond the suggested range of the information criteria analysis previously discussed. Lag values greater than 5 failed to reject the null of no cointegration in all cases. For lag of 5 and below, the results varied between the Reynolds number simulations. A full overview of the cointegration results for the three simulations and all lag parameters is found in Table 4.1.

Table 4.1: Cointegration test results of streamwise velocity vs. turbulence intensity using Engle-Granger test method on Moser channel flow data

	$Re = 180$		$Re = 395$		$Re = 590$	
Lag	h	pValue	h	pValue	h	pValue
0	1	0.016	0	0.579	1	0.001
1	1	0.001	1	0.001	1	0.001
2	0	0.366	0	0.313	1	0.001
3	0	0.387	1	0.036	1	0.001
4	0	0.068	0	0.302	1	0.003
5	0	0.063	0	0.402	1	0.023
6	0	0.199	0	0.506	0	0.203
7	0	0.350	0	0.512	0	0.461
8	0	0.904	0	0.740	0	0.632
9	0	0.915	0	0.785	0	0.688
10	0	0.877	0	0.738	0	0.778

The initial testing previously discussed was based on start and stop points for the logarithmic regions chosen by visual inspection and input from other general logarithmic region guidance. The impact of varying the start and stop locations of the logarithmic region is also of interest and can be evaluated by varying the start and stop locations of the log region in the $Re = 590$ simu-

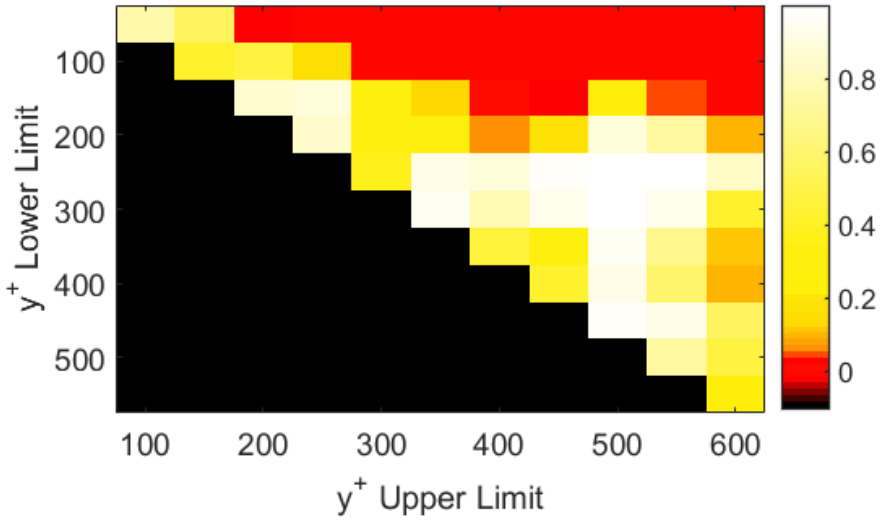


Figure 4.7: Cointegration results of streamwise velocity vs turbulence intensity with varying logarithmic region for $Re = 590$ data using the Engle-Granger cointegration test method. Shading represents pValues with white equal to one and red equal to zero. The more red the shading, the stronger the indication of cointegration for that log region selection. The lower left black region indicates invalid log regions where $y^+_{Upper} \leq y^+_{Lower}$. The greatest cointegration relationships are primarily found with smaller y^+ lower limits.

lation data. The start values are varied in increments of 50 with a range of $50 \leq y^+ \leq 550$. The logarithmic region end value started at the start value plus 50 and went through 600 in steps of 50. Results of this investigation are shown in Figure 4.7 where the shading corresponds to the pValue reported by the EGCI test at that combination of start and stop points. White locations represent a pValue of one and strong rejection of cointegration with red values approaching 0. The black region in the bottom left of the graph show the invalid log regions where $y^+_{upperlimit} \leq y^+_{lowerlimit}$.

These analysis results strongly indicate that there is a relationship between average streamwise velocity and turbulence intensity for these simulations at the lower lag values with a potential favoritism of the odd lag cases. The in-

vestigation of the variable logarithmic region indicates that the lower $y^+_{lowerlimit}$ values are the driving factor behind the cointegration relationship. This would suggest that the flow closer to the wall shows more of a relationship between these two physical quantities than the flow closer to the centerline of the channel.

The limited range of the available Reynolds numbers still leaves questions regarding the universal nature of these relationships and uncertainty about what the low Reynolds numbers impacts may be. Also, the lack of stochastic content in the averaged values and failure of the ADF test to detect a unit root in any of the series also creates a desire to perform further evaluation to corroborate these results. To help alleviate these concerns, data from the Johns Hopkins Turbulence Database was also evaluated.

4.2.3 Turbulent Channel Flow Data Analysis: Johns Hopkins Database

The Johns Hopkins University Turbulence Database (JHTDB) is another repository that provides detailed information from direct numerical simulations of turbulent flows. In addition to the average data provided in the Moser database, JHTDB provides instantaneous values for a variety of physical quantities in a multitude of turbulent flow types. One of these simulations is a channel flow simulation with a friction Reynolds number of 9.99713×10^2. The simulation is performed using an $8\pi \times 2 \times 3\pi$ channel with a $2048 \times 512 \times 1536$ grid [19]. A seventh-order Basis-Spline method was used for representation of solutions in the wall normal direction. The average streamwise velocity and turbulence intensity is provided over the full y^+ range from the wall to mid channel point in Figure 4.8 and shows the log-law behavior of the average

streamwise velocity.

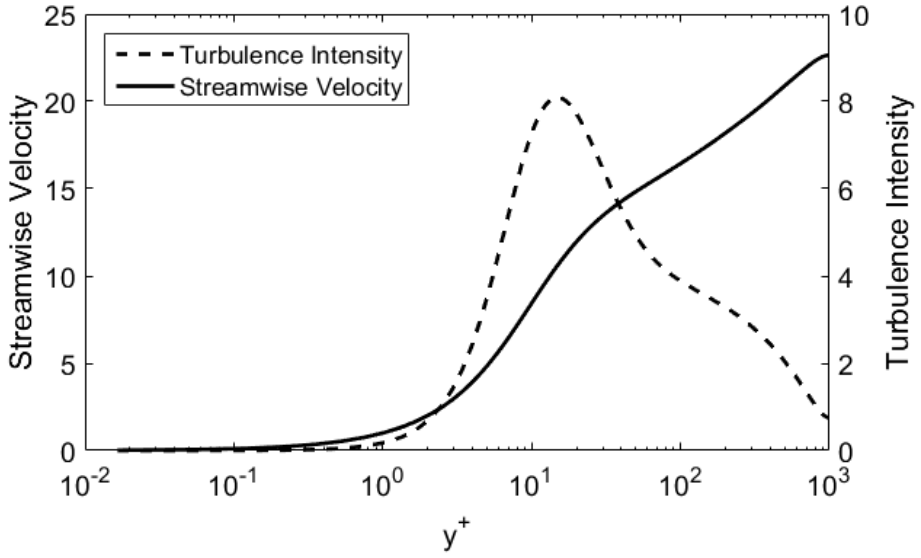


Figure 4.8: Average streamwise velocity and turbulence intensity over half channel height from Johns Hopkins Turbulence Database Channel Flow

The streamwise velocity and turbulence intensity in this region are provided in Figure 4.9. Marusic's defined log region for this flow is $30 < y^+ < 150$. Analysis of this region shows that it would reside primarily within the viscous sublayer area of the flow and is not a good approximation for the log region. Although the region suggested by Marusic's Reynolds number approach is not useful for this data, the concept of establishing the log region based on the deviation of the turbulence intensity from the theoretical values presented in (4.5) is relevant for establishing the region. The basics of utilizing the deviation of the turbulence intensity from the theoretical averages are that the turbulence intensity varies from the theoretical value quicker than the streamwise velocity and a better approximation of the start and stop portions of the log region can be obtained. A visual examination of the data suggests a definition of the logarithmic region as $60 < y^+ < 400$. Figure 4.9 shows

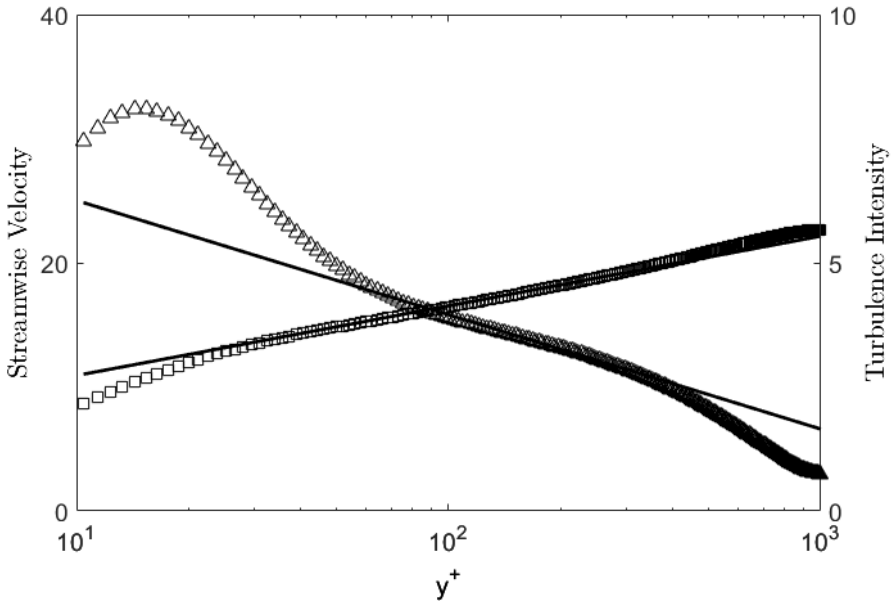


Figure 4.9: Average streamwise velocity (open square markers) and turbulence intensity (open triangle markers) profiles in channel flow compared with theoretical log region profiles as defined in (4.4) and (4.5) (solid lines). The parameters used are $\kappa = 0.41$, $A = 5.3$, $A_1 = 1.0$, and $B_1 = 1.65$. The turbulence intensity data varies from the theoretical profile more rapidly than the average velocity allowing for a better definition of the logarithmic region.

the velocity and turbulence intensity within this log region along with the theoretical profiles from equations (4.4) and (4.5).

Starting with this initial logarithmic region, the instantaneous streamwise and turbulence intensity values can be evaluated for cointegration. To achieve this, a subset of data was obtained from the JHTDB for analysis. A sample of the obtained data from the center of the X-Z plane at the middle time of the simulation is shown in Figure 4.10.

4.2.3.1 Lag Length Determination

One of the critical parameters that must be specified prior to performing an EGCI test is the proper selection of the lag length for use in the ADF test of

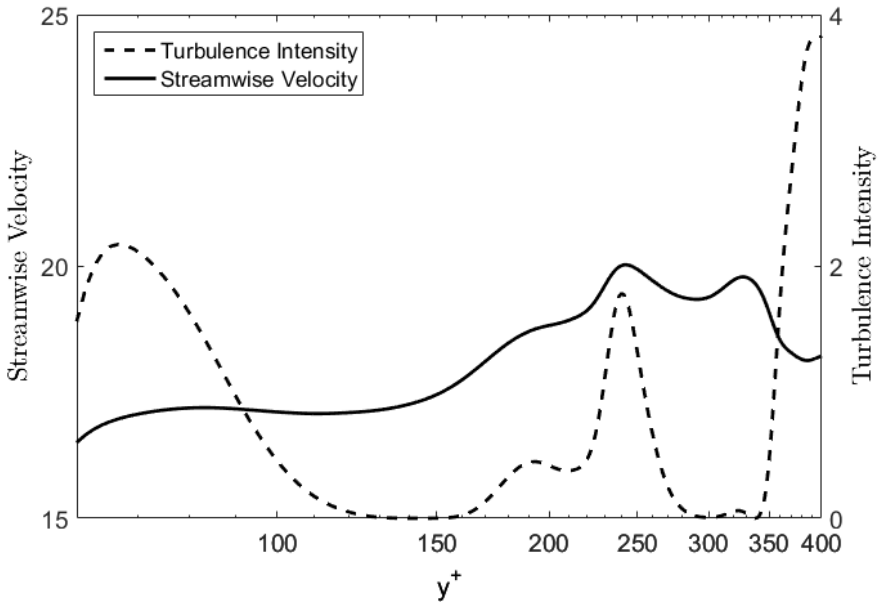


Figure 4.10: Instantaneous streamwise velocity and turbulence intensity in the channel flow logarithmic region at point $X = 4\pi$, $Z = 1.5\pi$, and $t = 13$ from JHTDB. The noise in the instantaneous data is apparent and the general trend of the theoretical profiles previously discussed is obscured in the data.

the regression residuals. If too small of a lag parameter is chosen then serial correlation in the residuals may incorrectly influence the test results. However, for every lag added, the degrees of freedom of the test are reduced and using too many lags can result in the power of the test being significantly impacted. The ADF test checks the null hypothesis of a unit root with the model

$$y_t = c + \delta t + \phi y_{t-1} + \beta_1 \Delta y_{t-1} + \cdots + \beta_p \Delta y_{t-p} + \epsilon_t, \tag{4.8}$$

where Δ is the differencing operator, p is the number of lagged difference terms, and ϵ_t is a mean zero innovation process. The number of lagged terms p is user defined as the lag parameter in the EGCI test. The proper definition of this term is critical as previously discussed.

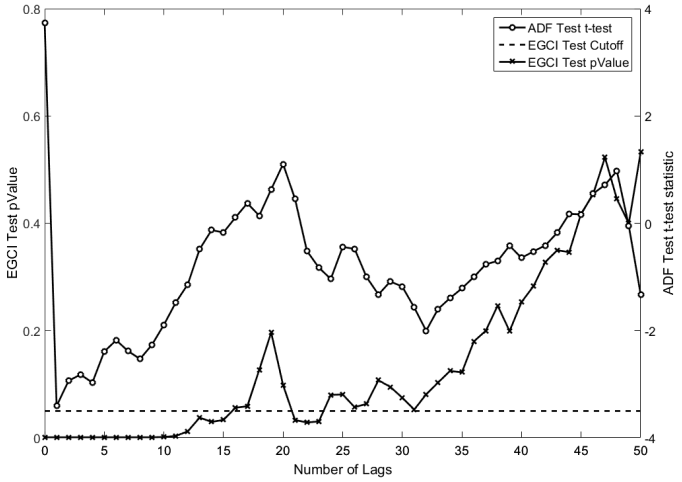


Figure 4.11: Lag length determination results using the general-to-specific methodology for channel flow data. This method suggests that the most appropriate lag lengths for testing are those below 10.

To investigate the utility of both the information criteria and "general-to-specific" methods previously defined, two synthetic series were created based on the analytical values for the streamwise velocity, (4.4), and turbulence intensity, (4.5). Normally distributed noise was added to both of these series and the "general-to-specific" and information criteria tests were conducted. Figure 4.11 shows the t-test statistic as reported by the EGCI test function, the reported pValue of the EGCI test, and the significance pValue cutoff for determining cointegration of 0.05. These two series can analytically be shown to be cointegrated and so the lag length chosen by these methods should also show positive cointegration. Applying the "general-to-specific" methodology, the first instance of a t-test statistic greater than zero comes with a lag length of 48. This lag does not result in a positive cointegration indication and does not appear to be the appropriate choice. Ultimately the largest t-test statistic occurs at the zero lag value which is a good choice. However, starting at a

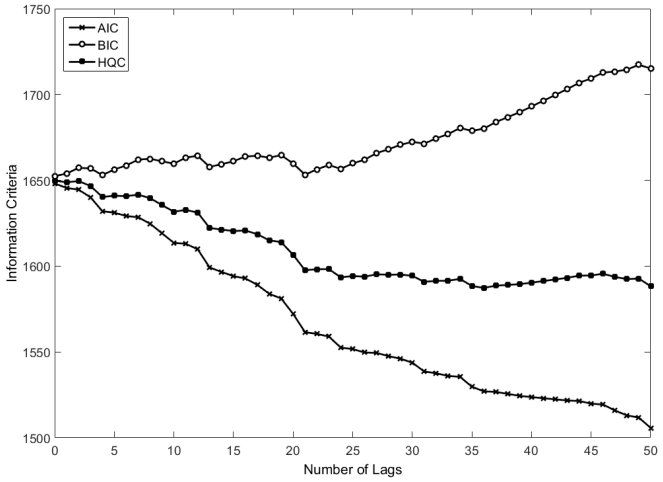


Figure 4.12: Lag length determination results using the information criteria methodology for channel flow data. The various information criteria results lead to conflicting recommended lag lengths with the AIC indicating a lag length of 50, HIC a lag length of 36, and BIC a lag length of 0.

larger lag value and working backwards, it would be possible to choose a lag length of 48 or even 20 which may be inappropriate choices.

The information criteria lag determination process was also evaluated with these series. The various information-criteria results as reported by the EGCI test function are shown in figure 4.12. The lag length that minimizes the information criteria should be chosen as the appropriate lag for testing. This varies greatly depending on the information criteria chosen. The BIC suggests a lag length of zero, while the HIC would result in 36, and the AIC would utilize a lag length of 50. Again, a value of zero lags is presented and is consistent with the general-to-specific results but extreme care must be taken to find this value as alternative options are also given. As a result of these conflicting lag length determination results, the remainder of the investigations will test all lag values from 0-10 to determine how the results are impacted by lag length selection. Also, it is important to note that these results were

obtained by evaluating analytic series with normally distributed noise added. This may not be a precise reconstruction of the actual channel flow data and therefore may not fully reflect the best lag choice for every data point. This fact reinforces the choice of testing a variety of lag parameters.

4.2.3.2 Cointegration Testing Results

Once the appropriate lag values have been determined, the series can be evaluated for any quantifiable relationships. An investigation of the cointegration relationships between a variety of physical quantities was conducted to determine if there were any significant relationships in the initial logarithmic region as previously defined. The primary quantities of interest are the streamwise velocity and turbulence intensity as these components visually appear to be related in the logarithmic region in some of the works previously discussed. A data set of 3,200 unique X-Z points were downloaded from the JHDTB website (40x20 grid over 4 time steps) and the quantities of interest were evaluated at each point. The components were initially tested for a unit root using the ADF test with zero lags and any points where both series of interest exhibited a unit root were then tested for cointegration. The total percentage of points that satisfied these requirements and exhibited cointegration was recorded. Results of these investigations are given in Table 4.2 where the Percent Cointegrated column gives the percentage of the 3,200 points that tested positive for cointegration in accordance with the EGCI test and the Average pVal column lists the average pValue from all points. Lag lengths of 0-10 were tested. (NOTE: For the duration of the cointegration testing discussion, the instantaneous velocity U^+ represents the instantaneous velocity divided by the friction velocity u_τ.) While the percentage of points in which both series contained a

Table 4.2: Cointegration testing results using the Engle-Granger test method for channel flow data. Significant relationships are found for all physical quantity pairs at one lag and also at 3, 5, and 7 lags for three of the four pairs.

	U^+ vs $u'u'$		$u'u'$ vs $v'v'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	9.7%	0.492	0.2%	0.573
1	23.4%	0.279	57.6%	0.026
2	2.4%	0.558	0.7%	0.468
3	5.1%	0.444	31.2%	0.106
4	0.9%	0.584	1.2%	0.413
5	1.9%	0.526	13.9%	0.196
6	0.4%	0.600	1.3%	0.388
7	0.9%	0.572	6.3%	0.261
8	0.3%	0.613	1.3%	0.382
9	0.6%	0.599	3.4%	0.305
10	0.4%	0.626	1.4%	0.383

	$u'u'$ vs $w'w'$		$v'v'$ vs $w'w'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	0.1%	0.542	0.1%	0.635
1	52.0%	0.031	65.2%	0.023
2	0.4%	0.448	0.4%	0.529
3	27.5%	0.114	35.5%	0.103
4	0.8%	0.399	0.5%	0.464
5	11.2%	0.200	15.3%	0.199
6	1.0%	0.377	1.0%	0.430
7	5.4%	0.263	7.1%	0.272
8	1.1%	0.370	1.3%	0.416
9	3.2%	0.301	4.3%	0.323
10	1.1%	0.371	1.6%	0.426

unit was relatively high, the percentage that showed cointegration was lower than anticipated.

At this point several potential reasons may be speculated for the low results. First the log region may be misspecified. This option will be investigated in the next section. Another potential issue is using the incorrect lag lengths. This issue could be further investigated by calculating the serial correlation of the residuals to determine if they adequately approximate a white noise process [8]. If significant correlation exists, longer lag lengths may be required. More detailed analysis of the appropriate lag length for each point tested is also suggested. The lag lengths chosen herein were based on theoretical profiles and assumed to be the same for all points. Establishing unique lag lengths for each point may provide more accurate results. Also, errors due to the low power of the ADF test may be driving the low percentages or this may be an indication that this methodology is not appropriate for evaluating potential relationships in the logarithmic region. This is the least likely option as the percentages for several of the lag parameters are greater than the 6% threshold that was established through testing of random walks previously. Larger percentages of cointegration are seen in the odd lag cases with one lag resulting in the greatest number of positive cointegration relationships. This may provide potential information regarding the structure of the residuals of the two series of interest, but it is difficult to draw any conclusions about the structure of the relationship from this information.

To further explore the relationships identified, the regression coefficients of the initial cointegration regression between U^+ and $u'u'$ were evaluated for all points that tested positive for cointegration. A simple model of $y = B_0 + B_1 x$ is evaluated in the cointegration regression. Average values of the parameters

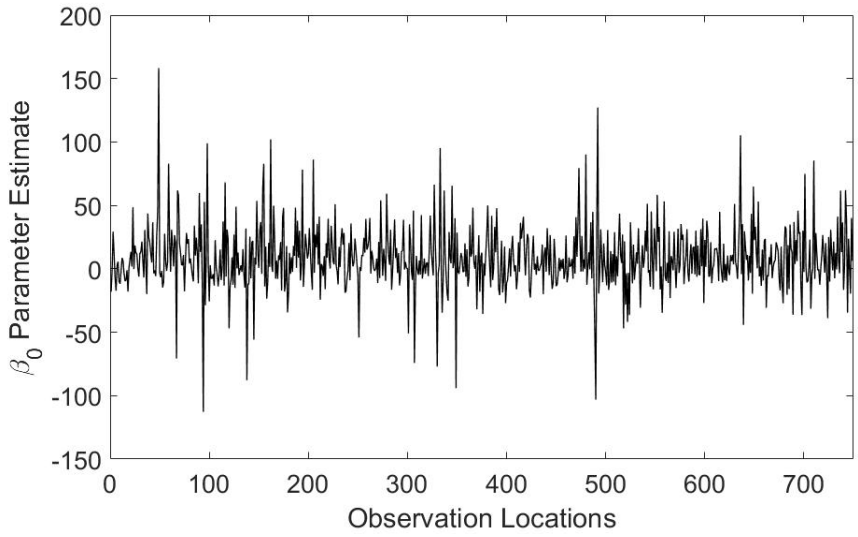


Figure 4.13: Estimated β_0 regression parameter values for each U^+ vs $u'u'$ point tested using the Engle-Granger test method. The β_0 parameter exhibits a relatively large amount of variation but there are no obvious outliers that skew the average or standard deviation calculations.

are $B_0 = 8.4534$ and $B_1 = -0.3264$ with standard deviations of 25.1984 and 1.3918 for B_0 and B_1 respectively. Figure 4.13 and 4.14 show the estimated parameters for each of the points where cointegration was found. There are a few outliers present in the data but nothing that would indicate an issue with the results. The large standard deviation implies that there is very little consistency among the cointegrated points. The conclusion drawn from this is that there is a large amount of variation in the instantaneous data and a universal relationship between the streamwise velocity and turbulence intensity is not obvious from this investigation method.

Testing of the U^+ vs $u'u'$ series was repeated with the Johansen test method using the default model as previously discussed. Overall the percentage of cointegration relations was higher than that found with the EGCI test. The results are shown in Table 4.3. Care must be taken when comparing the pValue

Table 4.3: Cointegration testing results using the Johansen test method on channel flow data. Significant relationships are identified for all physical quantities and all lag cases. Notably the one lag cases exhibit a lower percentage of cointegration than the other lag cases. This is in contrast to the EGCI results where the one lag case was had by far the largest percentage of cointegration.

	U^+ vs $u'u'$		$u'u'$ vs $v'v'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	34.3%	0.161	21.6%	0.084
1	52.7%	0.189	8.9%	0.025
2	25.0%	0.091	18.8%	0.051
3	44.9%	0.155	17.6%	0.042
4	21.5%	0.071	18.3%	0.042
5	31.6%	0.110	21.9%	0.051
6	21.3%	0.074	19.6%	0.046
7	28.2%	0.100	23.2%	0.055
8	23.6%	0.082	21.6%	0.053
9	27.5%	0.099	25.6%	0.067
10	26.7%	0.098	24.3%	0.063

	$u'u'$ vs $w'w'$		$v'v'$ vs $w'w'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	23.4%	0.084	17.8%	0.066
1	8.2%	0.021	6.3%	0.017
2	19.0%	0.052	15.2%	0.039
3	16.6%	0.037	14.6%	0.033
4	18.9%	0.046	15.8%	0.037
5	22.2%	0.052	19.3%	0.042
6	19.5%	0.047	18.3%	0.039
7	22.7%	0.054	21.9%	0.049
8	22.0%	0.054	21.4%	0.053
9	24.2%	0.060	23.5%	0.060
10	24.4%	0.062	22.9%	0.061

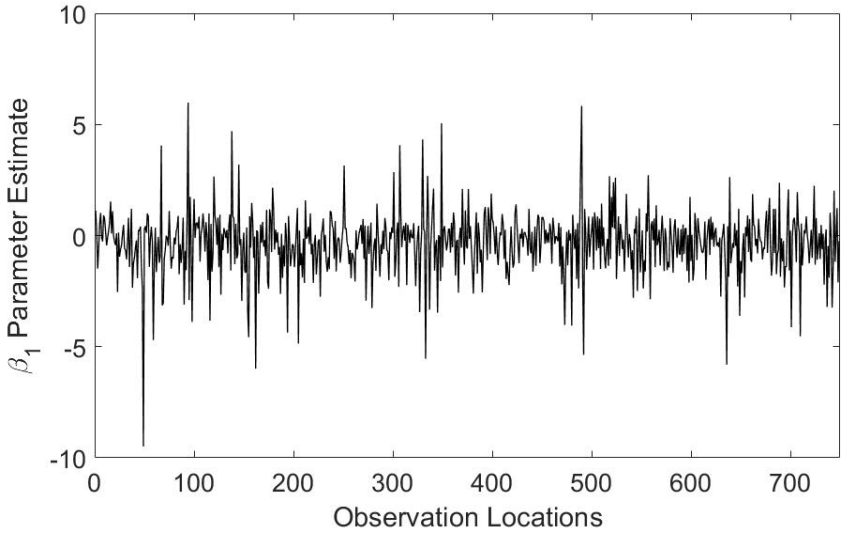


Figure 4.14: Estimated β_1 regression parameter values for each U^+ vs $u'u'$ point tested using the Engle-Granger test method. The β_1 parameter exhibits a relatively large amount of variation. There are a few outliers but nothing of great significance to the results.

results from the Johansen test with those from the Engle-Granger method. The two tests are evaluating opposite hypothesis so the pValue results will vary. Positive cointegration was counted for all points where the test rejected the null of rank less than zero but failed to reject the null of rank less than or equal to one.

It is interesting to note that not only are the percentages different, the pattern where the highest percentages are found is also inconsistent between the two test methods. In the EGCI test results, the one lag case consistently showed the highest percentage of cointegration while the JCI test results show the one lag case as the lowest percentage for all test pairs except for the U^+ vs. $u'u'$ case. Evaluating the various relationship pairs with both tests helps to check the results of the cointegration testing and can add confidence to any conclusions drawn due to the fact that the two tests use opposite null

hypothesis in their testing. The fact that there is such disagreement between results of the two tests indicates that more work may be needed to better specify the models used and understand the underlying structure of the data.

As discussed in 3.3.2, the vector error correction parameters can be evaluated using the Johansen test procedure. The parameters of the estimated VEC model were evaluated for the U^+ vs. $u'u'$ relationship at the one lag case since this showed the highest percentage of cointegration. The average and standard deviations of the model parameters are shown below with the averages listed first followed by the standard deviation in parenthesis.

$$A = \begin{bmatrix} -0.0056 & -0.0885 \\ (0.0230) & (0.0925) \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0190 & 0.6147 \\ (0.7239) & (0.6558) \end{bmatrix}$$

$$c_0 = \begin{matrix} -1.2319 \\ (13.7171) \end{matrix}$$

$$c_1 = \begin{bmatrix} 0.0048 & -0.0003 \\ (0.0042) & (0.0021) \end{bmatrix}$$

With the exception of the c_1 parameter, the standard deviations of the parameters in the VEC model are far lower than those found by the EGCI regression parameters. The greater consistency of the estimated parameters and higher percentages of cointegration may be a result of the higher power of the test and lack of reliance on unit root testing in the JCI test as opposed to the Engle Granger method. Additional investigation is needed to fully

understand the deltas and the driving forces behind the relationships.

Ultimately, significant relationships are found in all of the physical quantity pairs tested with both the EGCI and JCI test methods. Differences are encountered in the most significant lags and percentages of cointegration between the two methods which needs additional work to fully comprehend.

4.2.3.3 Multiple Log-Region Investigations

To evaluate the impact of the start and stop point of the logarithmic region on the relationships of the quantities, the previous analysis were repeated for multiple regions. The start of the logarithmic region was varied from 50 to 300 in increments of 50 and the end of the region was varied from the start value plus 100 to 1000 in increments of 100. For the U^+ vs $u'u'$ and $u'u'$ vs $v'v'$, it was found that as the size of the logarithmic region increased the positive percentage of cointegration increased. The remaining quantities showed an increase in cointegration percentage as the log region grew to a maximum cointegration percentage in the middle of the log region test range and then slightly fell at the largest log regions. As an example of this, the cointegration results of the $u'u'$ vs $v'v'$ for the initial log region of $60 \leq y^+ \leq 400$ are compared with the results from the region of $50 \leq y^+ \leq 950$ in table 4.4. These results fail to support any further definition of the log region as was shown in Figure 4.10. If the cointegration results were solely used to define the log region, they would indicate that the log region should extend for nearly the entire half channel flow which does not align with the definitions of the log region previously discussed.

Table 4.4: Comparison of cointegration results for varying logarithmic region for $u'u'$ vs $v'v'$ using Engle-Granger test method in channel flow. The larger log region results in greater cointegration percentages in all lag cases.

Log Region	Start: 60, Stop: 400,		Start: 50, Stop: 950	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	0.2%	0.573	0.0%	0.497
1	57.6%	0.026	59.4%	0.006
2	0.7%	0.468	1.4%	0.336
3	31.2%	0.106	44.4%	0.039
4	1.2%	0.413	3.3%	0.266
5	13.9%	0.196	28.7%	0.087
6	1.3%	0.388	4.7%	0.240
7	6.3%	0.261	17.9%	0.129
8	1.3%	0.382	5.0%	0.231
9	3.4%	0.305	11.9%	0.162
10	1.4%	0.383	5.4%	0.231

4.2.3.4 Johansen Method: Multivariate Cointegration Testing Results

In addition to providing an independent test methodology from the EGCI test, the Johansen method allows for analysis of cointegration relations among multiple physical quantities. Staying with the quantities previously evaluated, the Johansen method was utilized to evaluate U^+ with $u'u'$, and $w'w'$. The results of the testing are presented as either a rejection of or failure to reject the null of cointegration of rank r for $0 \leq r \leq n-1$ where n is the number of quantities being evaluated ($r = 3$ in this case). Table 4.5 presents the results of the Johansen cointegration testing for all of the possible combinations of null hypothesis ranks of r. The highest percentage of positive cointegration results occur while rejecting rank = 0, and failing to reject the null of rank = 1 or rank = 2. These results indicate that there is a relationship between

these three physical quantities. The estimated VEC model parameters along with their standard deviations in parenthesis are

$$A = \begin{bmatrix} 0.0012 & -0.0012 & -0.0126 \\ (0.0230) & (0.0925) & (0.0971) \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0959 & -0.0106 & 0.0827 \\ (0.6108) & (0.7553) & (1.5975) \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0.9008 & 0.3243 & -0.0046 \\ (0.1974) & (0.9020) & (0.9256) \\ -0.0294 & 0.8615 & 0.0078 \\ (0.0619) & (0.2106) & (0.2719) \\ 0.0100 & -0.0177 & 0.8896 \\ (0.0755) & (0.3733) & (0.1106) \end{bmatrix}$$

$$c_0 = \begin{matrix} -1.7937 \\ (11.3895) \end{matrix}$$

$$c_1 = \begin{bmatrix} c0.0053 & -0.0087 & 0.0025 \\ (0.0076) & (0.0278) & (0.0293) \end{bmatrix}$$

Investigations were also completed for P^+, $u'u'$ and $w'w'$. Table 4.6 presents the results of the Johansen cointegration testing. The estimated VEC model parameters are also presented below.

Table 4.5: Multivariate cointegration results for $U^+, u'u', w'w'$ using the Johansen test method for channel flow. Ranks of 1,0,0 and 1,1,0 indicate that cointegration of at least rank 1 exists between the physical quantities.

Percent Cointegrated								
Lags	h result for Rank 0, Rank 1, Rank 2							
	0,0,0	1,0,0	0,1,0	1,1,0	0,0,1	1,0,1	0,1,1	1,1,1
0	27.4	38.4	0.0	18.8	0.3	1.9	0.0	13.2
1	0.9	14.9	0.0	47.4	0.1	1.2	0.0	35.6
2	36.6	36.3	0.0	12.1	1.2	3.3	0.0	10.5
3	8.5	37.0	0.0	29.5	0.3	4.0	0.0	20.7
4	31.8	35.5	0.0	12.5	1.5	5.2	0.0	13.6
5	15.7	37.9	0.0	20.2	1.0	5.1	0.1	20.0
6	22.8	35.4	0.0	16.9	1.0	5.0	0.1	18.8
7	13.1	35.8	0.0	21.8	1.0	4.9	0.1	23.3
8	13.4	34.3	0.0	22.6	0.9	3.8	0.0	24.9
9	8.2	30.6	0.0	27.9	0.7	3.7	0.0	29.0
10	6.1	27.6	0.0	29.8	0.4	3.7	0.0	32.4

Average pValue								
Lags	h result for Rank 0, Rank 1, Rank 2							
	0,0,0	1,0,0	0,1,0	1,1,0	0,0,1	1,0,1	0,1,1	1,1,1
0	0.217	0.198	0.000	0.002	0.000	0.002	0.000	0.001
1	0.005	0.038	0.000	0.003	0.000	0.002	0.000	0.001
2	0.236	0.144	0.000	0.002	0.002	0.005	0.000	0.001
3	0.042	0.107	0.000	0.005	0.000	0.005	0.000	0.002
4	0.180	0.121	0.000	0.002	0.002	0.007	0.000	0.002
5	0.078	0.109	0.000	0.004	0.002	0.006	0.000	0.002
6	0.121	0.108	0.000	0.003	0.001	0.006	0.000	0.002
7	0.065	0.100	0.000	0.004	0.002	0.006	0.000	0.002
8	0.069	0.094	0.000	0.004	0.002	0.005	0.000	0.002
9	0.039	0.080	0.000	0.004	0.001	0.004	0.000	0.002
10	0.030	0.073	0.000	0.004	0.000	0.004	0.000	0.002

Table 4.6: Multivariate cointegration results for $P^+, u'u', w'w'$ using the Johansen test method for channel flow. Ranks of 1,0,0 and 1,1,0 indicate that cointegration of at least rank 1 exists between the physical quantities.

Percent Cointegrated								
Lags	h result for Rank 0, Rank 1, Rank 2							
	0,0,0	1,0,0	0,1,0	1,1,0	0,0,1	1,0,1	0,1,1	1,1,1
0	9.8	28.7	0.0	44.3	0.0	1.1	0.0	16.0
1	0.4	10.2	0.0	20.1	0.0	2.0	0.0	67.3
2	22.3	34.9	0.0	20.7	1.3	4.7	0.1	16.0
3	5.6	25.9	0.0	20.9	0.7	5.1	0.0	41.9
4	21.5	32.4	0.0	16.0	2.0	6.2	0.1	21.8
5	9.8	29.9	0.0	21.0	1.0	5.0	0.1	33.3
6	14.5	30.9	0.0	18.9	1.5	6.1	0.1	28.0
7	8.1	28.9	0.0	24.3	0.6	5.3	0.1	32.8
8	8.3	28.5	0.0	24.1	0.6	4.5	0.0	33.9
9	4.3	25.5	0.0	27.0	0.2	3.6	0.1	39.3
10	3.0	21.1	0.0	30.8	0.3	2.7	0.0	42.1

Average pValue								
Lags	h result for Rank 0, Rank 1, Rank 2							
	0,0,0	1,0,0	0,1,0	1,1,0	0,0,1	1,0,1	0,1,1	1,1,1
0	0.072	0.117	0.000	0.003	0.000	0.002	0.000	0.001
1	0.003	0.037	0.000	0.001	0.000	0.002	0.000	0.002
2	0.136	0.123	0.000	0.003	0.003	0.006	0.000	0.001
3	0.031	0.079	0.000	0.003	0.001	0.006	0.000	0.003
4	0.110	0.099	0.000	0.003	0.004	0.008	0.000	0.002
5	0.046	0.082	0.000	0.003	0.002	0.006	0.000	0.003
6	0.070	0.081	0.000	0.003	0.002	0.007	0.000	0.003
7	0.039	0.076	0.000	0.004	0.001	0.005	0.000	0.003
8	0.039	0.076	0.000	0.004	0.001	0.005	0.000	0.003
9	0.019	0.063	0.000	0.004	0.000	0.004	0.000	0.003
10	0.014	0.055	0.000	0.004	0.000	0.003	0.000	0.002

$$A = \begin{bmatrix} -0.0031 & 0.0005 & -0.0184 \\ (0.0177) & (0.0858) & (0.0724) \end{bmatrix}$$

$$B = \begin{bmatrix} 0.1469 & 0.0042 & 0.3245 \\ (1.0912) & (0.6476) & (1.3317) \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 0.9503 & -0.0791 & -0.0906 \\ (0.0697) & (0.8343) & (0.5706) \\ -0.0042 & 0.8780 & -0.0068 \\ (0.0434) & (0.1044) & (0.1574) \\ 0.0047 & -0.0029 & 0.8978 \\ (0.0683) & (0.3004) & (0.1097) \end{bmatrix}$$

$$c_0 = \begin{matrix} 0.0539 \\ (2.3255) \end{matrix}$$

$$c_1 = \begin{bmatrix} c0.0004 & -0.0008 & 0.0015 \\ (0.0050) & (0.0159) & (0.0136) \end{bmatrix}$$

The two relationship investigations resulted in very similar results with some slight differences in percentage of cointegration throughout. Both sets of physical quantities exhibit significant signs of cointegration of rank 1 and 2 for all lag values. The estimated VEC model coefficients are also similar suggesting the relationships among these sets of quantities may be comparable. Further analysis of individual points and the flow as a whole may provide more insight into the nature of these relationships and if there is a link between the two sets of series.

4.2.3.5 Engle-Granger Testing of Small Sample Averages

In addition to the instantaneous values at single points previously discussed, the average values are also of interest. The overall ensemble average of the channel is not useful as any stochastic properties are essentially removed due to the large number of points being averaged. An average of a smaller subset of points helps to smooth some of the "noise" in the values while still retaining the stochastic components. To investigate the impact of averaging small numbers of points over short times, 8 distinct groups of 10x10 X-Z points over 4 time steps were queried from the JHTDB system and then averaged to provide 32 unique series. It was found that none of the averaged series were found to contain a unit root utilizing the ADF test prerequisite as previously described. The cause of this is unknown as the series visually look non-stationary when the plots are analyzed. Therefore, testing was conducted without the ADF test prerequisite. This may cause the initial conditions of the EGCI test to be invalidated and the results of these investigations are presented with this caveat in mind. The results of these cointegration tests are shown in table 4.7. The same pattern is seen in the averaged series as was found in the instantaneous data. All of the tested pairs exhibit significant relationships for at least one lag parameter. This further reinforces the potential relationships in these quantities previously found.

Table 4.7: Cointegration results of small sample averages using the Engle-Granger cointegration test method for channel flow. Cointegration percentages of the small sample average are greater than those found for instantaneous points.

	U^+ vs $u'u'$		$u'u'$ vs $v'v'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	0.0%	0.759	43.8%	0.433
1	75.0%	0.070	46.9%	0.217
2	3.1%	0.669	12.5%	0.505
3	31.3%	0.261	21.9%	0.426
4	0.0%	0.574	12.5%	0.559
5	18.8%	0.374	9.4%	0.509
6	3.1%	0.546	0.0%	0.586
7	12.5%	0.459	3.1%	0.587
8	9.4%	0.548	3.1%	0.658
9	6.3%	0.500	3.1%	0.665
10	3.1%	0.562	0.0%	0.688

	$u'u'$ vs $w'w'$		$v'v'$ vs $w'w'$	
Lags	Percent Coint	Average pVal	Percent Coint	Average pVal
0	6.3%	0.689	0.0%	0.868
1	65.6%	0.090	87.5%	0.025
2	0.0%	0.649	0.0%	0.695
3	31.2%	0.309	34.4%	0.220
4	3.1%	0.594	0.0%	0.625
5	3.1%	0.428	9.4%	0.384
6	0.0%	0.590	0.0%	0.590
7	0.0%	0.540	12.5%	0.452
8	0.0%	0.623	0.0%	0.558
9	0.0%	0.620	12.5%	0.491
10	3.1%	0.637	3.1%	0.579

4.2.3.6 Patterns in Cointegration Results of Channel Flow Data

Another area of interest is to determine if there are any patterns in the cointegration relationships over the full channel. To investigate this, a sample of 200×150 points evenly distributed in the x and z dimensions were evaluated over a single time step of the simulation. Figure 4.15 shows the pValues of a one lag EGCI test over the subset of points previously described. pValues of one are shaded white and pValues of 0 are shaded red. Therefore, the darker the region the greater the likelihood of cointegration relationships. The results show that the cointegration relationships are evenly spaced throughout the channel volume and there are no obvious patterns that emerge.

Due to the discrepancy in percentage of cointegration relationships, this analysis was repeated with the Johansen method cointegration test. The results are shown in 4.16. The figure shows the pValue results of the cointegration test for a rank of less than or equal to one. Again, the red regions show higher probabilities of cointegration. The results are consistent with the Engle-Granger results and do not reveal any obvious patterns in the channel. An analysis with more points considered may be able to reveal finer structures in the channel that are not present here.

4.2.3.7 Summary

Cointegration testing is a useful method for investigating potential relationships in turbulent channel flows. It is found that there are significant relationships between the (instantaneous) streamwise velocity (U^+) vs. the turbulence intensity ($u'u'$), turbulence intensity vs $v'v'$, turbulence intensity vs $w'w'$, and

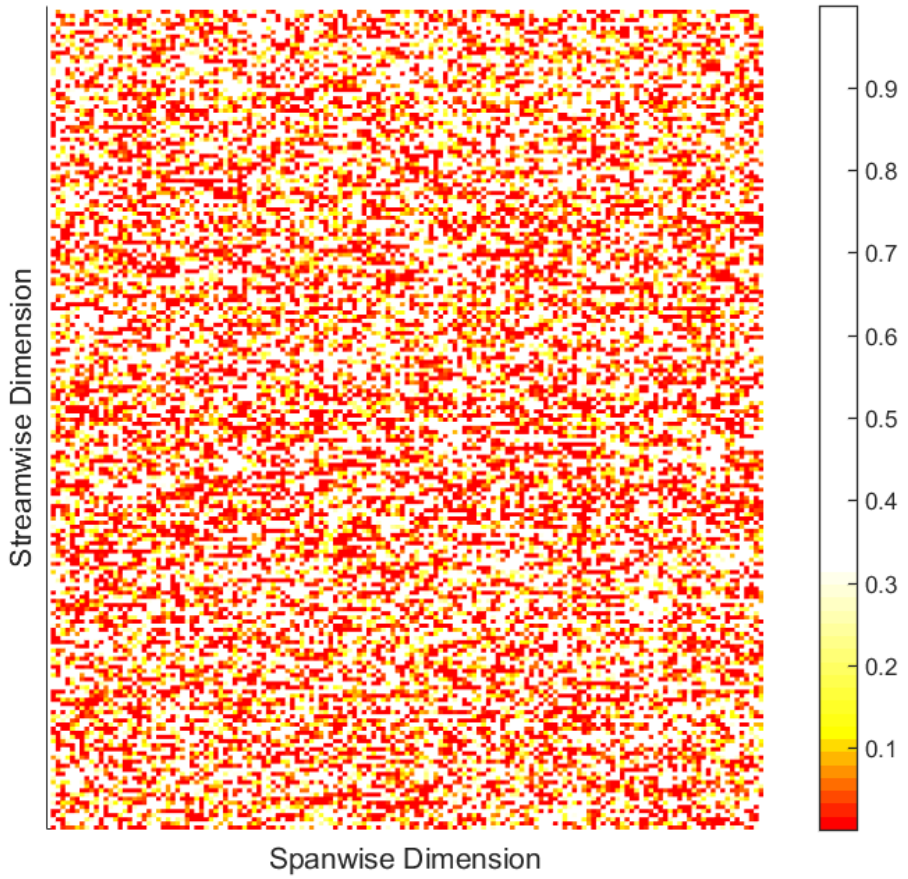


Figure 4.15: Pattern of cointegration test pValue results over X-Z plane in channel flow using Engle-Granger test method. The pValue results from each point evaluated in the X-Z plane are plotted to evaluate any potential patterns in the cointegration relationships throughout the channel. Red values indicate cointegration and white is an absence of a relationship. No obvious patterns are evident in the cointegration relationships throughout the channel flow.

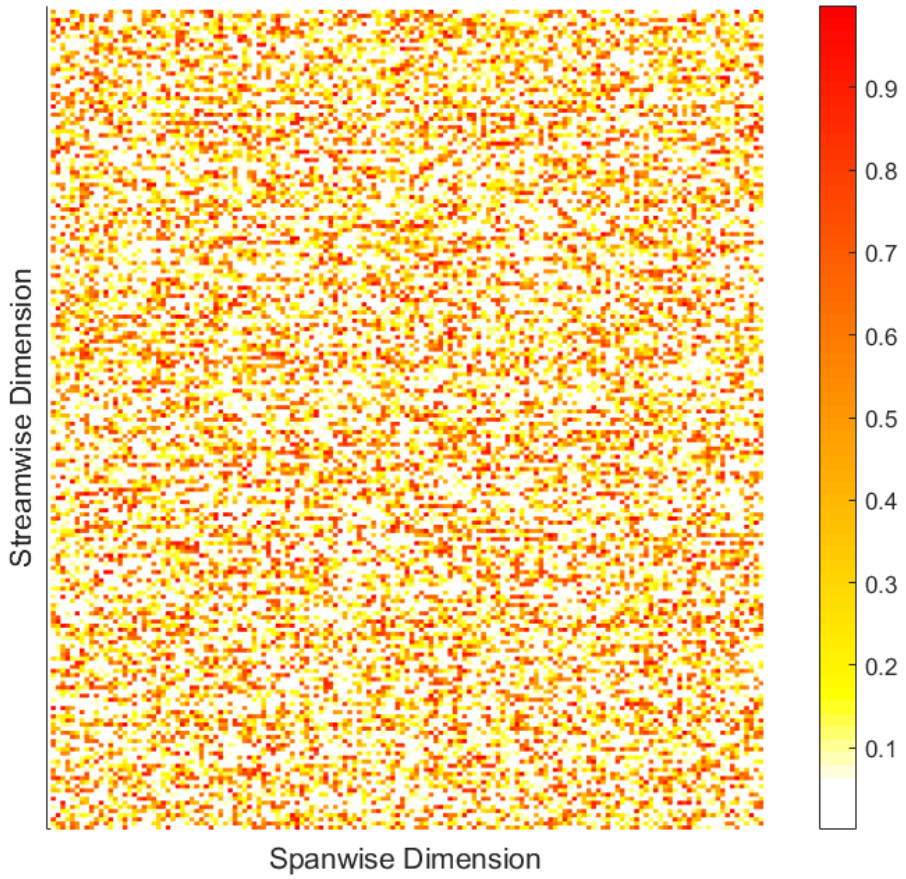


Figure 4.16: Pattern of cointegration test pValue results over X-Z plane in channel flow using Johansen test method. The pValue results from each point evaluated in the X-Z plane are plotted to evaluate any potential patterns in the cointegration relationships throughout the channel. Red values indicate cointegration and white is an absence of a relationship. No obvious patterns are evident in the cointegration relationships throughout the channel flow.

$v'v'$ vs $w'w'$. Both the Engle-Granger and Johansen test methods suggest these relationships. However, there is some inconsistency in the magnitude and most significant lag cases between the two tests. Additionally, the channel flow was evaluated for any patterns in the cointegration relationship between instantaneous U^+ and $u'u'$ but nothing of significance was found. Finally, the estimated regression and VEC parameters of the identified relationships are presented. There is a significant amount of variation in the parameters throughout the flow indicating a universal linear relationship model for the quantities may not be applicable.

4.2.4 Buoyancy Driven Flow Data Analysis: Johns Hopkins Database

Flows involving mixing of fluids with different densities where the primary driver of the flow is the difference in buoyancy of the two fluids are known as buoyancy driven flows. The potential energy in this flow as a result of the buoyancy force can be sufficient to drive the flow to turbulence. Buoyancy driven turbulence can be found in a variety of environmental and manmade situations such as combustion and chemical engineering applications [4, 21]. The ability to investigate relationships between various quantities may reveal additional insights about the structure of these flows in the same manner as was attempted with channel flow.

A common method to simulate buoyancy driven turbulent flows, is to start with a random distribution of two fluids with different densities which are statistically homogeneous and initially defined at rest. The simulation is allowed to progress and motion of the fluid is a result of buoyancy forces driven by gravity. If the resulting Reynolds number of the motion is sufficiently large,

the flow may become turbulent. As the flow progresses, the two fluids will mix together and eventually the system will return to rest as a uniform fluid. Two distinct periods can be defined for this flow. The initial period will be referred to as the increasing period where the velocity and turbulent Reynolds number are increasing. Once the turbulent Reynolds number reaches its maximum, the second period begins where the velocity and Reynolds number begins to decrease. This period is referred to as the decaying period. These two time periods will be essential to the analysis that follows.

For the JHTDB simulation, the miscible two-fluid incompressible Navier-Stokes equations are solved. These equations can be derived from the fully compressible Navier-Stokes equations. For the case of two species of fluid with different molar masses where the densities of the two fluids remain constant, the equations can be shown to be [20]:

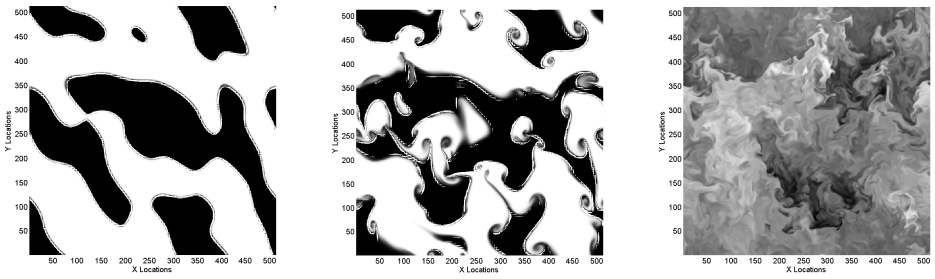
$$\frac{\partial}{\partial t}\rho + (\rho u_j)_{,j} = 0 \tag{4.9}$$

$$\frac{\partial}{\partial t}(\rho u_i) + (\rho u_i u_j)_{,j} = -p_{,i} + \tau_{ij,j} + \frac{1}{Fr^2}\rho g_i \tag{4.10}$$

$$u_{j,j} = -\frac{1}{Re_0 Sc}(ln\rho)_{,jj}, \tag{4.11}$$

where ρ is the density of the mixture, u_i is the mixture's velocity vector field, and p the pressure. The density of the fluid is defined by $\rho = (Y_1/\rho_1 + Y_2/\rho_2)^{-}1$ where ρ_1 and ρ_2 are the individual fluid densities with a mixing fraction of $Y_1 + Y_2 = 1$.

A direct numerical simulation (DNS) of homogeneous buoyancy driven turbulence in the manner previously described was conducted by Johns Hopkins University who made the results of the simulations available via various means

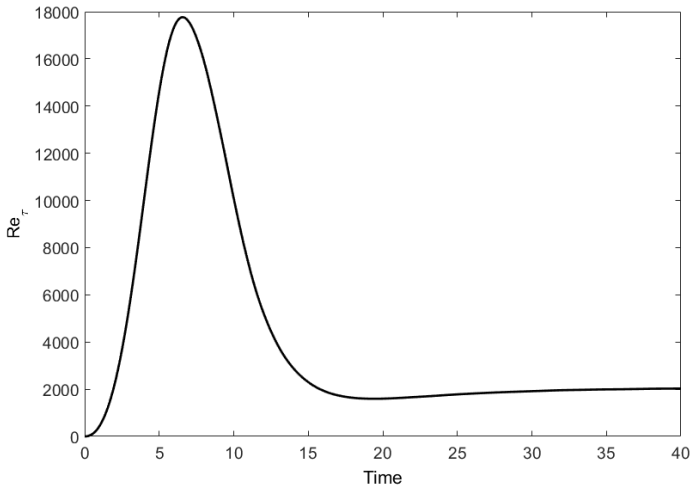


(a) $t = 0$ (b) $t =$ Maximum Re_τ (c) $t = 40.56$

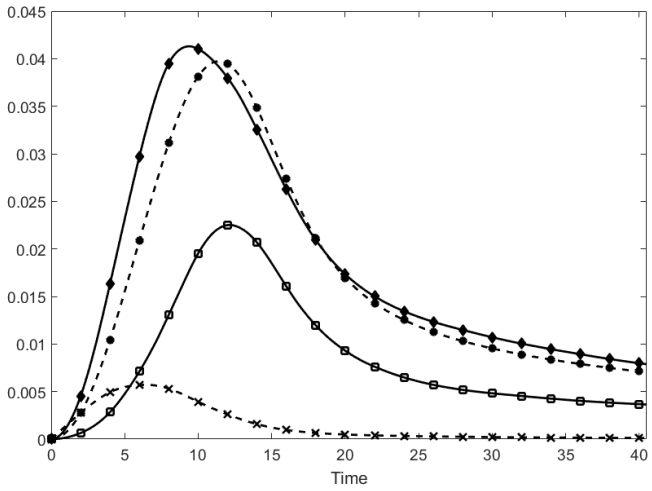
Figure 4.17: Density distribution during turbulent evolution of buoyancy driven flow

including web access. The simulation was conducted on a 1024^3 periodic grid and solved the miscible two-fluid incompressible Navier-Stokes equations. The simulation covers both the increasing and decaying portions of the flow over a total time of 40.56 seconds with a constant time interval of 4×10^{-2}. The buoyancy driven motion results in a maximum turbulent Reynolds number of $Re_t \approx 17,765$ at time 6.56 [20]. For this analysis, the time step at 6.56s is utilized as the cutoff between the increasing and decaying time ranges.

The evolution of the flow can be seen in Figure 4.17. The average turbulent Favre Reynolds number, Reynolds stresses (in the direction of gravity, $R_{vv} = \langle \rho u_1'' u_2'' \rangle$, and perpendicular to gravity, $R_{hh} = \left(\langle \rho u_2'' \rho u_2'' \rangle + \langle \rho u_3'' u_3'' \rangle \right)/2$), and vertical mass flux ($a_v = \langle \rho u_1' \rangle / \langle \rho \rangle$) are given in Figure 4.18. The two cointegration test methods previously discussed were utilized to look for relationships among physical quantities in the simulation data from the Johns Hopkins University dataset. The physical quantities of interest were determined based on the average evolutions previously shown in Figure 4.18. From these figures there appears to be a relationship between Reynolds stresses and Kinetic energy and possibly with vertical mass flux in the average data. The instantaneous values for these components are all tested against each other to



(a) Reynolds number



(b) Vertical Reynolds stress (solid line with filled diamonds), horizontal Reynolds stress (solid line with open squares), Favre kinetic energy (dashed line with filled circles), and vertical mass flux (dashed line with X)

Figure 4.18: Time evolution of average physical quantities in buoyancy driven flow. The Reynolds stress, Favre kinetic energy, and vertical mass flux appear to move together over time and may be candidates for cointegration testing.

investigate potential relationships. Along with these quantities, several other combinations of velocity, density, and their gradients are also evaluated to see if any relationships exist in the instantaneous data that may not be evident in average values.

Two distinct sets of data were analyzed in this investigation. First, results from a single point in the simulation volume were evaluated. Secondly, particle tracks were derived using the JHTDB provided functions and physical quantities over the particle track were evaluated for cointegration. The results of these two different approaches are presented in the following sections. As described in Section 3.4, the threshold for establishing a significant relationship between two physical quantities is roughly 15% of the tested points showing a positive cointegration relation.

4.2.4.1 Single Point Over Time Cointegration Results

A subset of 15,625 points in an equally spaced $25 \times 25 \times 25$ grid of points was queried from the database for the full time duration. At each location, a variety of physical quantities are tested for cointegration. Prior to implementing the EGCI test, the series of interest are initially tested for a unit root using the Augmented Dickey Fuller test. This helps to ensure that the prerequisites of each series being integrated of the same order is satisfied but is not a perfect screening. Those locations where both series of interest contain a unit root are then tested for cointegration using the EGCI test method. This process was repeated for the full time interval, the increasing time range, and the decaying time range. Additionally, a variety of lag parameters were used in the ADF test. Lag values of 0-10 were used with particular emphasis placed on the results of the 2 and 3 lag cases as these were the lag parameters suggested by

the various lag determination methods. The results are given as a percentage of points where both series contain a unit root and are cointegrated per the EGCI test based on the total number of points tested. Tables 4.8 and 4.9 contain the results.

It was found that there was significant variation between the lag parameters used to test for cointegration. As seen in the results presented, the three lag case shows higher percentages of positive cointegration relationships for all components tested. This variability continued for all of the lags 0-10 with the highest percentages of cointegration found in the odd lag cases. The one lag case presented the highest percentages which continued to decrease as the lag values increased. The occurrence of higher cointegration percentages when odd lag values are used may provide some insight into the structure of the regression residuals but exactly what is unknown. The fact that the lag parameter is used during the testing of the regression residuals and not the original series adds difficulty to determining what the differences in results based on the lag used imply about the original series of interest.

Also of interest is the very low percentages of non-stationary series in the full and decaying time ranges for the density vs density gradient and density gradient vs pressure gradient quantities. These results imply that cointegration testing may not be applicable to these quantities in these time regions of this type of flow. More traditional regression analysis or other methods may be better suited to investigating relationships for these quantities.

The odd lag results all indicate that there is a significant relationship amongst all of the physical quantities evaluated for the increasing time range. There is also a significant relationship in the full and decaying time ranges for the velocity vs velocity, velocity vs Reynolds stress, kinetic energy vs Reynolds

Table 4.8: Cointegration results with two lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time			Increasing Time			Decaying Time		
Physical quantities	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	44%	0.335	100%	3%	0.795	61%	34%	0.457
Kinetic Energy vs R_{hh}	66%	31%	0.432	100%	4%	0.782	67%	31%	0.421
Kinetic Energy vs. Mass Flux	93%	0%	1.000	100%	0%	1.000	82%	0%	1.000
R_{hh} vs R_{vv}	55%	22%	0.545	100%	4%	0.780	48%	16%	0.609
u_x vs. u_y	80%	10%	0.498	100%	25%	0.621	69%	10%	0.537
u_x vs. u_z	79%	9%	0.503	100%	25%	0.624	68%	9%	0.549
u_y vs. u_z	69%	15%	0.470	100%	20%	0.691	68%	11%	0.500
u_x vs. R_{vv}	74%	24%	0.476	100%	12%	0.638	61%	22%	0.537
u_y vs. R_{hh}	62%	16%	0.520	100%	11%	0.705	66%	16%	0.485
u_z vs. R_{hh}	61%	18%	0.513	100%	8%	0.737	66%	18%	0.479
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	11%	0.854	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	12%	0.844	0%	0%	0.997
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	11%	0.851	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	50%	6%	0.870	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	7%	0.858	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	5%	0.868	0%	0%	0.999

Table 4.9: Cointegration results with three lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time			Increasing Time			Decaying Time		
Physical quantities	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	67%	0.280	100%	36%	0.437	61%	54%	0.405
Kinetic Energy vs R_{hh}	66%	55%	0.361	100%	40%	0.373	67%	56%	0.351
Kinetic Energy vs. Mass Flux	93%	0%	1.000	100%	0%	1.000	82%	0%	1.000
R_{hh} vs R_{vv}	55%	44%	0.476	100%	40%	0.373	48%	37%	0.545
u_x vs. u_y	80%	38%	0.303	100%	35%	0.463	69%	35%	0.390
u_x vs. u_z	79%	38%	0.311	100%	37%	0.441	68%	34%	0.400
u_y vs. u_z	69%	51%	0.340	100%	37%	0.436	68%	46%	0.366
u_x vs. R_{vv}	74%	47%	0.341	100%	23%	0.490	61%	41%	0.443
u_y vs. R_{hh}	62%	47%	0.414	100%	25%	0.496	66%	49%	0.375
u_z vs. R_{hh}	61%	50%	0.412	100%	24%	0.526	66%	50%	0.373
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	19%	0.728	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	20%	0.723	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	19%	0.728	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	23%	0.688	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	24%	0.679	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	22%	0.694	0%	0%	0.999

stress, and Reynolds stress vs Reynolds stress.

Testing was repeated with the Johansen test method. Initial screening of the series for a unit root with the ADF test is not required and the results presented show the percentage of total test points that rejected the null of rank 0 and failed to reject the null of rank 1. Tables 4.10 and 4.11 show the results of the Johansen testing.

Again there is a difference between the two and three lag cases in the Johansen testing. However, in contrast with the EGCI test results, the three lag cointegration percentages are generally lower than the two lag case. Overall the percentages of cointegration returned by the Johansen test are greater than the EGCI results. Also, the 100% results are suspect but any errors leading to their appearance have not been found to date. These results are questioned as they are significantly higher than any other cointegration percentages found in the rest of this testing. Given the magnitude of the noise in turbulent flows, finding 100% of the points exhibiting cointegration raises curiosity that is in need of further investigation. The disagreement between the EGCI and Johansen results is also interesting. The two test methods evaluate opposite null hypothesis and should serve as a check against each other. The cause behind the fact that there is a disagreement with regards to the percentage of cointegration between the two test methods should be investigated further. This is consistent with the results obtained in the channel flow investigations previously discussed.

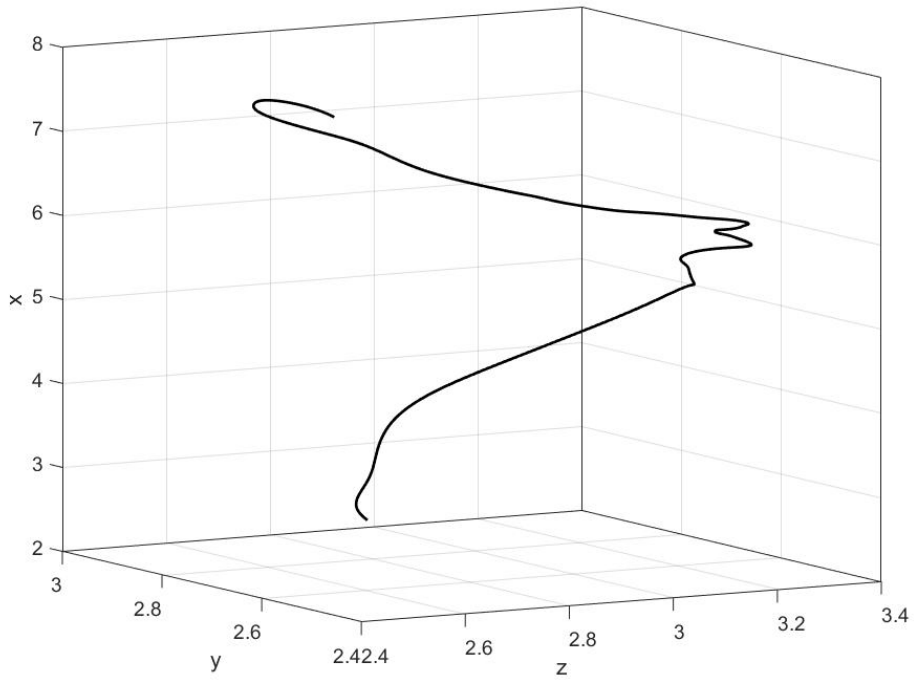


Figure 4.19: Johns Hopkins Turbulence Database buoyancy data - example particle path realization

Table 4.10: Cointegration results with two lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	35%	0.020	100%	0.237	60%	0.017
Kinetic vs. R_{hh}	33%	0.020	100%	0.239	60%	0.017
R_{hh} vs. R_{vv}	70%	0.018	100%	0.237	94%	0.017
u_x vs. u_y	52%	0.066	100%	0.155	80%	0.057
u_x vs. u_z	75%	0.060	88%	0.160	100%	0.056
u_y vs. u_z	64%	0.036	100%	0.144	54%	0.041
u_x vs. R_{vv}	80%	0.106	95%	0.163	77%	0.080
u_y vs. R_{hh}	66%	0.030	100%	0.192	53%	0.032
u_z vs. R_{hh}	43%	0.027	100%	0.193	62%	0.032
ρ vs. $\partial\rho/\partial x$	57%	0.018	76%	0.117	9%	0.004
ρ vs. $\partial\rho/\partial y$	72%	0.021	100%	0.120	6%	0.004
ρ vs. $\partial\rho/\partial z$	70%	0.020	66%	0.124	5%	0.003
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	3%	0.001	100%	0.213	0%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	100%	0.216	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	94%	0.216	0%	0.001

Table 4.11: Cointegration results with three lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	9%	0.006	100%	0.207	2%	0.005
Kinetic vs. R_{hh}	9%	0.006	100%	0.203	2%	0.005
R_{hh} vs. R_{vv}	23%	0.005	100%	0.198	10%	0.006
u_x vs. u_y	50%	0.023	90%	0.190	62%	0.019
u_x vs. u_z	51%	0.020	94%	0.176	28%	0.019
u_y vs. u_z	23%	0.011	100%	0.169	36%	0.013
u_x vs. R_{vv}	78%	0.032	87%	0.190	63%	0.027
u_y vs. R_{hh}	18%	0.009	100%	0.213	23%	0.010
u_z vs. R_{hh}	9%	0.006	100%	0.197	21%	0.008
ρ vs. $\partial\rho/\partial x$	62%	0.022	100%	0.153	6%	0.004
ρ vs. $\partial\rho/\partial y$	72%	0.022	100%	0.148	6%	0.004
ρ vs. $\partial\rho/\partial z$	75%	0.022	92%	0.156	5%	0.004
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.227	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	93%	0.239	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	96%	0.233	0%	0.001

4.2.4.2 Particle Track Cointegration Results

In addition to the single point over time values, the same physical quantities were obtained from the database along particle tracks. A set of 1,000 points were equally distributed in a $10 \times 10 \times 10$ grid and used as starting points. For each starting point the particle track was developed using the provided functions in the JHTDB Matlab code. Figure 4.19 gives an example of one of the particle path realizations.

Cointegration testing was performed on the 1,000 particle track series in the same way as was done for the single point over time data. The averages for these series were also evaluated and found to be very similar to the averages obtained with the previous series. As a result, the same emphasis will be placed on the 2 and 3 lag cointegration results. Tables 4.12 and 4.13 give the results of the cointegration testing.

The same patterns in the lag parameter results were noted in particle track testing as were found in the previous testing. The three lag tests resulted in much higher percentages of cointegration than the two lag case and again the one lag case produced the highest percentages. The percentage of non-stationary series in the density vs density gradient and density gradient vs pressure gradient full time range and decaying time range were still very low. Overall, the odd lag cases have higher percentages of positive cointegration relationships and provide evidence of relationships between most of the physical quantities during the increasing time range.

In the three lag case, using the 10% significance cutoff, the Kinetic energy vs. Reynolds stresses, Reynolds Stress vs. Reynolds Stress, velocity vs. velocity, velocity vs. Reynolds stress, and density vs density gradient all ex-

Table 4.12: Cointegration results with two lags for particle track using Engle-Granger test method in buoyancy driven flow

	Full Time			Increasing Time			Decaying Time		
Physical quantities	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	90%	12%	0.415	100%	2%	0.869	74%	8%	0.549
Kinetic vs. R_{hh}	90%	11%	0.458	100%	2%	0.843	83%	10%	0.516
R_{hh} vs. R_{vv}	85%	7%	0.508	100%	2%	0.842	70%	5%	0.622
u_x vs. u_y	94%	4%	0.624	100%	31%	0.564	78%	4%	0.660
u_x vs. u_z	93%	4%	0.627	100%	28%	0.573	78%	3%	0.660
u_y vs. u_z	92%	5%	0.533	100%	29%	0.583	87%	5%	0.570
u_x vs. R_{vv}	90%	8%	0.569	100%	22%	0.553	75%	8%	0.607
u_y vs. R_{hh}	89%	6%	0.547	100%	16%	0.645	86%	5%	0.553
u_z vs. R_{hh}	89%	6%	0.545	100%	18%	0.626	86%	5%	0.553
ρ vs. $\partial\rho/\partial x$	15%	1%	0.937	27%	11%	0.844	18%	5%	0.881
ρ vs. $\partial\rho/\partial y$	17%	1%	0.935	26%	12%	0.837	20%	5%	0.870
ρ vs. $\partial\rho/\partial z$	15%	1%	0.937	24%	11%	0.851	19%	5%	0.874
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	6%	0.941	41%	12%	0.813	11%	10%	0.895
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	5%	0.952	45%	12%	0.808	8%	7%	0.925
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.964	42%	11%	0.815	7%	6%	0.935

Table 4.13: Cointegration results with three lags for particle track using Engle-Granger test method in buoyancy driven flow

	Full Time			Increasing Time			Decaying Time		
Physical quantities	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal	Perc ADF non-Stationary	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	90%	65%	0.175	100%	21%	0.579	74%	50%	0.330
Kinetic vs. R_{hh}	90%	58%	0.208	100%	26%	0.523	83%	52%	0.283
R_{hh} vs. R_{vv}	85%	54%	0.249	100%	27%	0.520	70%	43%	0.396
u_x vs. u_y	94%	34%	0.313	100%	31%	0.475	78%	27%	0.418
u_x vs. u_z	93%	34%	0.316	100%	32%	0.473	78%	29%	0.419
u_y vs. u_z	92%	48%	0.241	100%	35%	0.459	87%	42%	0.296
u_x vs. R_{vv}	90%	46%	0.282	100%	16%	0.550	75%	39%	0.387
u_y vs. R_{hh}	89%	50%	0.252	100%	18%	0.524	86%	46%	0.279
u_z vs. R_{hh}	89%	51%	0.252	100%	17%	0.550	86%	48%	0.278
ρ vs. $\partial\rho/\partial x$	15%	1%	0.943	27%	10%	0.856	18%	4%	0.886
ρ vs. $\partial\rho/\partial y$	17%	0%	0.944	26%	10%	0.853	20%	4%	0.878
ρ vs. $\partial\rho/\partial z$	15%	1%	0.944	24%	10%	0.864	19%	4%	0.881
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.941	41%	8%	0.848	11%	9%	0.896
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.951	45%	8%	0.844	8%	7%	0.925
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.964	42%	9%	0.848	7%	6%	0.935

hibit significant relationships in the increasing time range. While the trends are similar between the single point and particle track cases, the single point over time results generally show a higher percentage of cointegration in the increasing time range.

Yet again, the testing was repeated with the Johansen method. The results of the two and three lag cases are presented in tables 4.14 and 4.15. The variation between the two and three lag cases for the Johansen test method are less pronounced than some of the other results previously discussed. The largest discrepancies occur in the decaying time range where some of the physical quantities exhibit an increase in cointegration while some experience a decrease. As previously stated, the difference in behavior of the EGCI and Johansen tests is concerning. Ideally similar patterns would emerge from the two tests to bolster the argument of cointegration. These results may indicated that the model being used in one or both of the tests is incorrect for the turbulent data.

A full listing of all lag parameters tested for both the point over time and particle track investigations is given in Appendix C.

4.2.4.3 Decaying Range Start Time Variation: Cointegration Impacts

A brief visual examination of the average turbulence physical quantities presented earlier reveals that not all of the quantities reach their maximum values at the same time as the Reynolds number. Utilizing the time of maximum Reynolds number as the cutoff for the increasing and decaying time ranges will result in some of the non-linear regions of the series potentially being included in the decaying time range analysis. To address this, the start of the decaying

Table 4.14: Cointegration results with two lags for particle track using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	37%	0.041	88%	0.217	59%	0.072
Kinetic vs. R_{hh}	32%	0.041	88%	0.219	62%	0.072
R_{hh} vs. R_{vv}	29%	0.042	89%	0.217	62%	0.072
u_x vs. u_y	50%	0.078	81%	0.142	53%	0.108
u_x vs. u_z	31%	0.082	92%	0.140	50%	0.094
u_y vs. u_z	66%	0.055	100%	0.133	66%	0.086
u_x vs. R_{vv}	52%	0.186	89%	0.160	100%	0.174
u_y vs. R_{hh}	19%	0.065	93%	0.191	74%	0.105
u_z vs. R_{hh}	37%	0.068	87%	0.197	66%	0.098
ρ vs. $\partial\rho/\partial x$	100%	0.260	94%	0.156	71%	0.050
ρ vs. $\partial\rho/\partial y$	100%	0.255	86%	0.146	73%	0.045
ρ vs. $\partial\rho/\partial z$	100%	0.255	80%	0.151	67%	0.045
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	11%	0.003	94%	0.220	0%	0.003
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	4%	0.003	100%	0.229	0%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	98%	0.216	0%	0.002

Table 4.15: Cointegration results with three lags for particle track using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	29%	0.020	100%	0.200	35%	0.046
Kinetic vs. R_{hh}	29%	0.020	100%	0.204	38%	0.046
R_{hh} vs. R_{vv}	29%	0.020	100%	0.202	35%	0.046
u_x vs. u_y	58%	0.064	100%	0.184	57%	0.094
u_x vs. u_z	50%	0.065	91%	0.163	66%	0.086
u_y vs. u_z	31%	0.046	100%	0.169	100%	0.076
u_x vs. R_{vv}	52%	0.081	90%	0.178	100%	0.104
u_y vs. R_{hh}	39%	0.034	100%	0.205	55%	0.065
u_z vs. R_{hh}	36%	0.034	88%	0.189	100%	0.054
ρ vs. $\partial\rho/\partial x$	100%	0.265	93%	0.169	86%	0.048
ρ vs. $\partial\rho/\partial y$	100%	0.258	88%	0.160	71%	0.044
ρ vs. $\partial\rho/\partial z$	100%	0.262	73%	0.161	85%	0.044
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	16%	0.003	96%	0.254	2%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	4%	0.003	100%	0.250	0%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	98%	0.248	1%	0.002

time range was delayed to 13 seconds while the end of the increasing range remained at the maximum Reynolds number time.

Cointegration testing was repeated with this modified time range for both the single point over time and particle track data sets. In all cases it was found that the percentage of positive cointegration decreased. This may imply that the cointegration is primarily being driven during the time periods near maximum Reynolds number. Regardless of the underlying reason, this analysis shows that the relationships in the flow change over time and any statements regarding global relationships should be made carefully.

4.2.4.4 Summary

A multitude of significant cointegration relationships are found in the physical quantities evaluated in turbulent buoyancy driven flows. Some of the notable relationships occurred between the kinetic energy of the flow vs. Reynolds stresses (horizontal and vertical) and between velocity components and Reynolds stresses. The density gradient and pressure gradient terms tested all failed to indicate any significant relationships. The greatest percentage of cointegration was found between the kinetic energy and Reynolds stress at the one lag case. Overall, the Johansen test found the greatest percentage of cointegration in the increasing period of the flow while the Engle-Granger test typically provided the greatest percentages over the full time range. Again, the Engle-Granger and Johansen methods resulted in varying results which needs additional investigation to fully understand the cause and impact of this.

Chapter 5

Conclusion

5.1 Overview of Results

Cointegration has been found to be a useful analysis method for evaluating non-stationary data in a variety of fields. This analysis has shown how cointegration can be applied to a variety of turbulent flow data sets in the search for relationships between a variety of physical quantities.

An analysis of data from the Johns Hopkins Turbulence Database DNS simulation of a turbulent channel flow provided confirmation that cointegration testing is a legitimate method for use in investigations into turbulent flows. Utilizing the Engle-Granger test method, it was shown that there are significant relationships between the (instantaneous) streamwise velocity (U^+) vs. the turbulence intensity ($u'u'$), turbulence intensity vs $v'v'$, turbulence intensity vs $w'w'$, and $v'v'$ vs $w'w'$. These results also corroborate the hypothesized relationship between streamwise velocity and turbulence intensity originally observed in Hultmark's and Marusic's work [23, 16]. However, follow on analysis utilizing the Johansen test method revealed inconsistent results and casts some doubt on the relationships identified. Both methods found significant relationships but did so with different lag parameters and exhibited inconsistent behavior when lag values and other properties were varied. This may indicate

that an alternative model may need to be employed and that more work is needed to further refine the use of cointegration with turbulence data.

An attempt was also made to utilize cointegration test results to further define the logarithmic region in the channel flow. This resulted in somewhat inconclusive results that do not appear to be in agreement with the logarithmic regions proposed by other works. It is concluded that definition of the logarithmic region based on the cointegration relationship between the streamwise velocity and turbulence intensity is not a valid method for defining this region.

To further investigate relationships in turbulent flows, data from the JHTDB buoyancy driven turbulence simulation was evaluated. Again, the Engle-Granger and Johansen cointegration tests were utilized for this analysis. Several pairs of physical quantities were tested for relationships over three distinct time segments of the flow development. Two types of series were evaluated. The first series consists of data from a single location in the simulation volume over the full time of the simulation. The second series is composed of data obtained over a particle path traced for the duration of the simulation. A multitude of physical quantities for both types of series exhibited significant relationships with the largest number of relationships found in the increasing turbulence intensity time range. Some of the notable relationships occurred between the kinetic energy of the flow vs. Reynolds stresses (horizontal and vertical) and between velocity components and Reynolds stresses. The density gradient and pressure gradient terms tested all failed to indicate any significant relationships. Again, the Engle-Granger and Johansen methods resulted in varying results.

5.2 Objectives and Research Questions

Looking back upon the objectives laid out in Section 1.3 we can now answer the initial research questions that were posed. (i) Significant cointegration relationships do exist amongst a variety of physical quantities in multiple types of turbulent flow. (ii) The relationships can be quantified through either linear regression coefficient estimates or estimates of VEC model parameters depending on the cointegration method utilized. (iii) In a turbulent channel flow, physical quantity pairs of U^+ vs. $u'u'$, $u'u'$ vs. $v'v'$, '$u'u'$ vs $w'w'$, and $v'v'$ vs $w'w'$ all exhibit signs of cointegration. Additionally, multivariate relationships were also identified for U^+ vs. $u'u'$ vs. $w'w'$ and for P^+ vs. $u'u'$ vs. $w'w'$. Buoyancy driven flows also exhibited strong signs of cointegration in a multitude of physical quantity pairs. The strongest relationships were identified between Kinetic Energy and Reynolds stress (both horizontal and vertical), u_y vs. u_z, u_x vs R_{vv}, u_y vs R_{hh}, and u_z vs R_{hh}. (iv) Cointegration is a valid method for identifying and quantifying relationships amongst physical quantities in turbulent data.

5.3 Future Work

This work has shown that cointegration is a legitimate method for investigating turbulent flows and for analyzing potential relationships within these flows. This type of analysis may be utilized to learn more about the nature of turbulence and reduce the required workload of post simulation analysis. More work is needed to refine the appropriate models to use when evaluating turbulent flows and to determine the utility of additional alternative cointegration test methods.

To further refine the results presented in this work, several immediate steps are needed. First, a more definitive investigation into the proper lag length for ADF and cointegration testing is needed. Unique lag length determination at every test point as opposed to a universal lag length may help to provide more information into the structure of the turbulence and provide better cointegration results. Also, more work on determining what the lag length results indicate about the structure of the quantities being investigated may also be fruitful.

The Matlab implementations of both the EGCI and JCI test methods allow the user the ability to choose different unit root and model specifications. The single default selection was utilized in this investigation but alternative models may provide better results with more understanding of the structure of the series of interest.

The investigations in this analysis were all conducted over the half channel height or in the full buoyancy flow. However, much of the work completed in investigating the universal properties of turbulence are centered around the dissipation and inertial subranges [46]. Cointegration testing may be useful in evaluating relationships within these ranges both within a singular flow and between flow types to determine if universal relationships do exist.

Finally, the full significance and driving force behind the relationships identified in this paper have not been investigated. More work is needed to fully comprehend the potential impact of what it really means when quantities are identified as being cointegrated.

Bibliography

[1] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 1974.

[2] The World Bank. *World DataBank Jobs*, 2016. Available From http://databank.worldbank.org/data.

[3] GI Barenblatt and Alexandre J Chorin. New perspectives in turbulence: Scaling laws, asymptotics, and intermittency. *SIAM review*, 40(2):265–291, 1998.

[4] G. K. Batchelor, V. M. Canuto, and J. R. Chasnov. Homogeneous buoyancy-generated turbulence. *Jornal of Fluid Mechanics*, 235, 1992.

[5] P.A. Davidson. *Turbulence and introduction for scientists and engineers*. Oxford University Press, New York, 2004.

[6] David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

[7] Norman R. Draper and Harry Smith. *Applied Regression Analysis*. John Wiley & Sons, Inc, New York, New York, 1998.

[8] Walter Enders. *Applied Econometric Time Series*. Wiley, 2010.

[9] Robert F. Engle and C. W. J. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251–276, 1987.

[10] Davide Faranda, Flavio Maria Emanuele Pons, Bérengère Dubrulle, François Daviaud, Brice Saint-Michel, Éric Herbert, and Pierre-Phillippe Cortet. Modelling and analysis of turbulent datasets using auto regressive moving average processes. *Physics of Fluids*, 26, 2014.

[11] C.W.J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.

[12] C.W.J. Granger and P. Newbold. Some comments on the evaluation of economic forecasts. *Applied Economics*, 5:35–47, 1973.

[13] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1994.

[14] Derek D. Headey and Andrew Hodge. The effect of population growth on economic growth: A meta-regression analysis of the macroeconomic literature. *Population and development review*, 35:221–248, 2009.

[15] David F. Hendry and Katarina Juselius. Explaining cointegration analysis: Part 1. *The Energy Journal*, 21, 2000.

[16] Marcus Hultmark. A theory for the streamwise turbulent fluctuations in high reynolds number pipe flow. *Journal of Fluid Mechanics*, 707:575–584, 2012.

[17] W.P. Jones and B. E. Launder. The prediction of laminarization with a two-equation model of turbulence. *International Journal of Heat Mass Transfer*, 15:301–314, 1972.

[18] Kunihiro Kamataki, Kimitaka Itoh, Sanae-I. Itoh, Akihide Fujisawa, Shigeru Inagaki, Yoshihiko Nagashima, Masatoshi Yagi, and Takuma Yamada. Statistical analysis technique for estimation of causal relationship in plasma turbulence. *Journal of the Physical Society of Japan*, 79, 2010.

[19] Y. Li, E. Perlman, M. Wan, Y. Yang, R. Burns, C. Meneveau, R. Burns, S. Chen, A. Szalay, and G. Eyink. A public turbulence database cluster and applications to study lagrangian evolution of velocity increments in turbulence. *Journal of Turbulence*, 9, 2008.

[20] D. Livescu, C. Canada, K. Kalin, R. Burns, and J. Pulido. *Homogeneous buoyancy driven turbulence data set*. http://turbulence.pha.jhu.edu/docs/README-HBDT.pdf.

[21] D. Livescu and J. R. Ristorcelli. Buoyancy-driven variable-density turbulence. *Journal of Fluid Mechanics*, 591, 2004.

[22] I. Marusic, B. J. McKeon, P. A. Monkewitz, H. M. Nagib, A. J. Smits, and K. R. Sreenivasan. Wall-bounded turbulent flows at high reynolds numbers: Recent advances and key issues. *Physics of Fluids*, 22, 2010.

[23] Ivan Marusic, Jason P. Monty, Marcus Hultmark, and Alexander J. Smits. On the logarithmic region in wall turbulence. *Journal of Fluid Mechanics*, 716:R3–1–R3–11, 2012.

[24] Mathworks. Engle-granger cointegration test documentation, 2015. http://www.mathworks.com/help/econ/egcitest.html.

[25] Mathworks. Matlab johansen cointegration test documentation, 2016. http://www.mathworks.com/help/econ/jcitest.html.

[26] B.J. McKeon. Natural logarithms. *Journal of Fluid Mechanics*, 718:1–4, 2013.

[27] Robert D. Moser, John Kim, and Nagi N. Mansour. Direct numerical simulation of turbulent channel flow up to $re_\tau = 590$. *Physics of Fluids*, 11:943–945, 1999.

[28] Michael P. Murray. A drunk and her dog: An illustration of cointegration and error correction. *The American Statistician*, 48:37–39, 1994.

[29] Mark Nelkin. In what sense is turbulence an unsolved problem? *Science*, 255(5044):566, 1992.

[30] Mark Nelkin. Universality and scaling in fully developed turbulence. *Advances in physics*, 43(2):143–181, 1994.

[31] Vladimir I. Nikora and Derek G. Goring. Adv measurements of turbulence: Can we improve their interpretation? *Journal of Hydraulic Engineering*, 124:630 – 634, 1998.

[32] Peter C. B. Phillips and Pierre Perron. Testing for a unit root in time series regression. *Biometrika*, 75:335–346, 1988.

[33] L. Prandtl. Uber die ausgebildete turbulenz. *ZAMM*.

[34] L. Prandtl. Uber ein neues formelsystem fur die ausgebildete turbulenz. *Nacr. Akad. Wiss. Gottingen, Math-Phys.*, KI., 1945.

[35] M. B. Priestley. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B*, 27:204–237, 1965.

[36] Søren Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.

[37] Søren Johansen. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59:1551–1580, 1991.

[38] J. C. Rotta. Statistische theorie nichthomogener turbulenz. *Zeitschrift fur Physik*, 129:547–572, 1951.

[39] Jörg Schumacher, Janet D Scheel, Dmitry Krasnov, Diego A Donzis, Victor Yakhot, and Katepalli R Sreenivasan. Small-scale universality in fluid turbulence. *Proceedings of the National Academy of Sciences*, 111(30):10961–10965, 2014.

[40] George A. F. Seber. *Linear Regression Analysis*. John Wiley & Sons, Inc, Hoboken, New Jersey, 2003.

[41] Katepalli R Sreenivasan. Fluid turbulence. *Reviews of Modern Physics*, 71(2):S383, 1999.

[42] Jielun Sun, Sean P. Burns, Donald H Lenschow, Robert Banta, Rob Newsom, Richard Coulter, Stephen Frasier, Turker Ince, Carmen Nappo, Joan Cuxart, William Blumen, Xuhui Lee, and Xin-Zhang Hu. Intermittent turbulence associated with a density current passage in the stable boundary layer. *Boundary-Layer Meteorology*, 105, 2002.

[43] Bo Tao, Joseph Katz, and Charles Meneveau. Geometry and scale relationships in high reynolds number turbulence determined from three-dimensional holographic velocimetry. *Physics of Fluids*, 12, 2000.

[44] D. J. Thomson. Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*, 180:529 – 556, 1987.

[45] Francis C.K. Ting and James T. Kirby. Observation of undertow and turbulence in a laboratory surf zone. *Coastal Engineering*, 24, 1994.

[46] Z. Warhaft. Turbulence in nature and in the laboratory. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 2002.

[47] Z.Warhaft. Passive scalars in turbulent flows. *Annual Review of Fluid Mechanics*, 32(1):203–240, 2000.

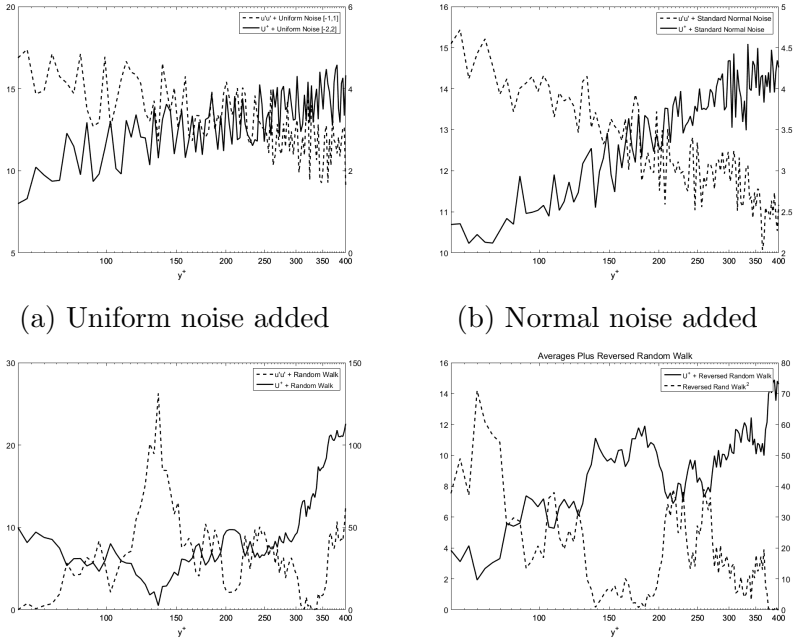
Appendices

Appendix A

Interpolation Impact on Cointegration Testing

The full results of the cointegration testing as discussed in Section 4.1.2 are presented in this Appendix. A sample realization of the theoretical series with each of the various noise distributions are shown in Figure A.1.

Lag values of 0-10, 20, and the lag from 0-50 that minimized the AIC were evaluated for the EGCI test and the percentage of positive cointegration results and average pValue were recorded. The JCI testing only utilized lag values from 0-10. The full results of the EGCI test of the theoretical series with noise added are presented in Table A.1. Table A.2 gives the results of Johansen cointegration testing with uniform noise added to the series (average pValues are omitted).



(a) Uniform noise added

(b) Normal noise added

(c) Random walk added

(d) Reverse random walk added

Figure A.1: Sample realizations of theoretical average channel flow streamwise velocity and average turbulence intensity with various noise types added

Table A.1: Cointegration results of theoretical average streamwise velocity and turbulence intensity with various noise series added using the Engle-Granger test method. The uniform and normally distributed noise exhibit much higher percentages of cointegration than the random walks as expected. However, the trend in the series where the random walks are added to the theoretical profiles also results in higher cointegration percentages than the pure random walks.

	Uniform Noise		Normally distributed		Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue	Positive Coint	Average pValue
0	100%	0.001	100%	0.001	9%	0.402
1	100%	0.001	100%	0.001	11%	0.409
2	96%	0.010	100%	0.002	9%	0.443
3	65%	0.053	92%	0.015	8%	0.468
4	28%	0.135	71%	0.046	7%	0.483
5	10%	0.218	43%	0.110	5%	0.506
6	2%	0.327	19%	0.185	4%	0.533
7	2%	0.394	12%	0.247	4%	0.541
8	1%	0.471	7%	0.326	4%	0.568
9	0%	0.532	2%	0.398	1%	0.553
10	0%	0.596	0%	0.469	4%	0.578
20	0%	0.834	0%	0.762	0%	0.739
Min AIC	2%	0.901	0%	0.879	2%	0.765

	Reversed Random Walk		Pure Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue
0	12%	0.434	5%	0.487
1	9%	0.470	5%	0.482
2	12%	0.489	4%	0.503
3	6%	0.502	4%	0.503
4	4%	0.538	6%	0.506
5	3%	0.55	2%	0.520
6	4%	0.579	1%	0.542
7	4%	0.572	0%	0.558
8	2%	0.599	3%	0.563
9	2%	0.631	2%	0.576
10	1%	0.657	1%	0.579
20	2%	0.738	1%	0.680
Min AIC	2%	0.713	1%	0.742

Table A.2: Cointegration results of theoretical average streamwise velocity and turbulence intensity with uniform noise added. The $r = 1, 0$ case represents cointegration. The Johansen method results in significant relationships being identified but does so with differing lag values and percentages of cointegration than the EGCI method.

Uniform Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	0%	0%	100%
1	0%	0%	0%	100%
2	0%	13%	0%	87%
3	0%	33%	0%	67%
4	0%	45%	2%	53%
5	7%	58%	2%	33%
6	17%	48%	0%	35%
7	20%	37%	5%	38%
8	28%	30%	3%	38%
9	32%	40%	7%	22%
10	30%	32%	7%	32%

After interpolation utilizing the various methods discussed in Section 4.1.2, the cointegration testing was repeated. The results of the EGCI cointegration testing for all interpolation methods are given in Table tables A.3 to A.5. Table tables A.6 to A.8 shows the impacts of interpolation on the JCI test method for the uniform noise case.

Table A.3: Cointegration test results of cubic spline interpolated theoretical averages with noise added using Engle-Granger test method. The interpolation greatly increases the cointegration percentage of the odd lag cases for all noise types.

	Uniform Noise		Normally distributed		Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue	Positive Coint	Average pValue
0	75%	0.038	96%	0.020	4%	0.721
1	100%	0.001	100%	0.001	89%	0.040
2	15%	0.198	32%	0.130	4%	0.729
3	100%	0.001	100%	0.001	50%	0.220
4	99%	0.002	100%	0.001	18%	0.347
5	100%	0.001	100%	0.001	24%	0.303
6	31%	0.164	58%	0.100	5%	0.643
7	30%	0.145	57%	0.091	3%	0.542
8	99%	0.003	100%	0.001	46%	0.231
9	83%	0.032	95%	0.011	11%	0.425
10	69%	0.060	90%	0.027	9%	0.467
20	8%	0.369	22%	0.210	10%	0.466
Min AIC	0%	0.584	3%	0.448	0%	0.505

	Reversed Random Walk		Pure Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue
0	4%	0.766	0%	0.999
1	86%	0.073	76%	0.177
2	4%	0.745	82%	0.094
3	32%	0.271	90%	0.064
4	15%	0.419	85%	0.067
5	21%	0.374	68%	0.101
6	5%	0.680	49%	0.209
7	3%	0.610	45%	0.313
8	32%	0.286	39%	0.409
9	5%	0.500	33%	0.497
10	4%	0.546	28%	0.557
20	2%	0.587	2%	0.866
Min AIC	3%	0.576	1%	0.941

Table A.4: Cointegration test results of linearly interpolated theoretical averages with noise added using Engle-Granger test method. The interpolation increases all cointegration percentages but has less impact on the random walks than the other noise types.

	Uniform Noise		Normally distributed		Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue	Positive Coint	Average pValue
0	72%	0.039	91%	0.023	13%	0.650
1	100%	0.001	100%	0.001	53%	0.136
2	100%	0.001	100%	0.001	19%	0.321
3	100%	0.001	100%	0.001	14%	0.368
4	99%	0.006	100%	0.002	9%	0.469
5	96%	0.009	99%	0.003	8%	0.443
6	100%	0.001	100%	0.001	27%	0.275
7	100%	0.002	100%	0.001	11%	0.406
8	99%	0.006	100%	0.002	11%	0.443
9	75%	0.039	95%	0.013	8%	0.500
10	90%	0.019	98%	0.005	11%	0.445
20	43%	0.094	87%	0.025	7%	0.443
Min AIC	15%	0.239	62%	0.076	10%	0.438

	Reversed Random Walk		Pure Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue
0	2%	0.792	7%	0.476
1	43%	0.171	2%	0.489
2	6%	0.428	3%	0.514
3	3%	0.489	2%	0.517
4	1%	0.616	5%	0.538
5	1%	0.568	4%	0.536
6	17%	0.323	2%	0.551
7	4%	0.485	2%	0.556
8	3%	0.517	2%	0.560
9	2%	0.577	1%	0.561
10	2%	0.510	2%	0.556
20	5%	0.519	0%	0.638
Min AIC	7%	0.469	1%	0.754

Table A.5: Cointegration test results of PCHIP interpolated theoretical averages with noise added using Engle-Granger test method. PCHIP interpolation has more impact on the pure random walk than previously encountered.

	Uniform Noise		Normally distributed		Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue	Positive Coint	Average pValue
0	88%	0.027	99%	0.011	8%	0.704
1	100%	0.001	100%	0.001	84%	0.042
2	89%	0.022	99%	0.007	5%	0.614
3	100%	0.001	100%	0.001	17%	0.319
4	100%	0.001	100%	0.001	13%	0.349
5	100%	0.001	100%	0.001	13%	0.371
6	99%	0.009	100%	0.002	6%	0.474
7	99%	0.006	100%	0.002	12%	0.328
8	100%	0.005	100%	0.001	7%	0.416
9	94%	0.014	99%	0.003	7%	0.433
10	99%	0.008	100%	0.002	11%	0.402
20	25%	0.145	70%	0.043	5%	0.509
Min AIC	52%	0.109	98%	0.010	5%	0.481

	Reversed Random Walk		Pure Random Walk	
Lags	Positive Coint	Average pValue	Positive Coint	Average pValue
0	6%	0.731	3%	0.968
1	77%	0.086	65%	0.267
2	5%	0.651	71%	0.189
3	17%	0.347	73%	0.161
4	13%	0.379	62%	0.196
5	9%	0.407	52%	0.277
6	2%	0.528	47%	0.342
7	11%	0.383	45%	0.396
8	4%	0.468	40%	0.436
9	3%	0.484	35%	0.468
10	5%	0.448	30%	0.510
20	3%	0.489	6%	0.728
Min AIC	5%	0.472	2%	0.873

Table A.6: Cointegration test results of cubic spline interpolated theoretical averages with uniform noise added using Johansen test method.

Uniform Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	2%	0%	98%
1	0%	0%	0%	100%
2	20%	18%	7%	55%
3	0%	0%	0%	100%
4	0%	0%	0%	100%
5	7%	0%	0%	100%
6	8%	35%	3%	55%
7	0%	37%	3%	52%
8	0%	5%	0%	95%
9	0%	25%	0%	75%
10	0%	33%	0%	67%

Table A.7: Cointegration test results of linearly interpolated theoretical averages with uniform noise added using Johansen test method.

Uniform Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	12%	0%	88%
1	0%	0%	0%	100%
2	0%	0%	0%	100%
3	0%	0%	0%	100%
4	0%	3%	0%	97%
5	0%	7%	0%	93%
6	0%	0%	0%	100%
7	0%	0%	0%	100%
8	0%	7%	0%	93%
9	2%	20%	0%	78%
10	2%	12%	0%	87%

Table A.8: Cointegration test results of PCHIP interpolated theoretical averages with uniform noise added using Johansen test method.

Uniform Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	2%	0%	98%
1	0%	0%	0%	100%
2	0%	5%	0%	95%
3	0%	0%	0%	100%
4	0%	0%	0%	100%
5	0%	0%	0%	100%
6	0%	7%	0%	93%
7	0%	7%	0%	93%
8	0%	3%	0%	97%
9	0%	8%	0%	92%
10	0%	7%	0%	93%

Table A.9: Cointegration test results of theoretical averages with normally distributed noise added using Johansen test method.

Normal Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	0%	0%	100%
1	0%	43%	0%	57%
2	0%	40%	0%	60%
3	0%	45%	0%	55%
4	0%	35%	0%	65%
5	2%	27%	2%	70%
6	5%	15%	2%	78%
7	7%	10%	5%	78%
8	8%	7%	7%	78%
9	10%	5%	10%	75%
10	10%	5%	10%	75%

Table A.10: Johansen cointegration results of cubic spline interpolated theoretical average streamwise velocity and turbulence intensity with normally distributed noise added

Normal Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	40%	0%	60%
1	0%	0%	0%	100%
2	18%	23%	8%	50%
3	0%	33%	0%	67%
4	0%	42%	0%	58%
5	0%	40%	0%	60%
6	8%	30%	3%	58%
7	7%	35%	0%	58%
8	0%	50%	0%	50%
9	2%	50%	0%	48%
10	3%	42%	0%	55%

Table A.11: Cointegration test results of linearly interpolated theoretical averages with normally distributed noise added using Johansen test method.

Normal Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	2%	58%	0%	40%
1	0%	0%	0%	100%
2	0%	8%	0%	92%
3	0%	28%	0%	72%
4	0%	45%	0%	55%
5	0%	53%	0%	47%
6	0%	35%	0%	65%
7	0%	57%	0%	43%
8	0%	52%	0%	48%
9	2%	38%	0%	60%
10	0%	48%	0%	52%

Table A.12: Cointegration test results of PCHIP interpolated theoretical averages with normally distributed noise added using Johansen test method

Normal Noise Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	47%	0%	53%
1	0%	0%	0%	100%
2	0%	50%	0%	50%
3	0%	37%	0%	63%
4	0%	38%	0%	62%
5	0%	38%	0%	62%
6	0%	53%	0%	47%
7	0%	52%	0%	48%
8	0%	52%	0%	48%
9	0%	48%	0%	52%
10	0%	48%	0%	52%

Table A.13: Cointegration test results of theoretical averages with random walk added using Johansen test method.

random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	67%	13%	5%	15%
1	62%	18%	10%	10%
2	60%	17%	3%	20%
3	57%	15%	8%	20%
4	60%	18%	5%	17%
5	58%	20%	10%	12%
6	55%	23%	5%	17%
7	60%	17%	12%	12%
8	57%	25%	7%	12%
9	67%	12%	10%	12%
10	72%	8%	8%	12%

Table A.14: Cointegration test results of cubic spline interpolated theoretical averages with random walk added using Johansen test method.

random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	77%	13%	7%	3%
1	8%	12%	0%	80%
2	77%	17%	2%	5%
3	23%	22%	5%	50%
4	52%	18%	7%	23%
5	42%	20%	8%	30%
6	78%	12%	7%	3%
7	77%	10%	7%	7%
8	23%	20%	3%	53%
9	53%	20%	8%	18%
10	58%	20%	8%	13%

Table A.15: Cointegration test results of linearly interpolated theoretical averages with random walk added using Johansen test method.

random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	72%	20%	2%	7%
1	20%	23%	2%	55%
2	52%	18%	8%	22%
3	60%	23%	7%	10%
4	72%	18%	3%	7%
5	67%	18%	5%	10%
6	40%	18%	5%	37%
7	62%	22%	5%	12%
8	63%	22%	7%	8%
9	72%	17%	7%	5%
10	62%	23%	5%	10%

Table A.16: Cointegration test results of PCHIP interpolated theoretical averages with random walk added using Johansen test method.

random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	73%	22%	2%	3%
1	2%	27%	0%	72%
2	87%	10%	2%	2%
3	50%	20%	13%	17%
4	60%	13%	10%	17%
5	63%	12%	8%	17%
6	80%	8%	2%	10%
7	60%	18%	5%	17%
8	72%	10%	3%	15%
9	73%	12%	2%	13%
10	67%	12%	3%	18%

Table A.17: Cointegration test results of theoretical averages with reverse random walk added using Johansen test method.

Reversed random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	65%	18%	8%	8%
1	58%	17%	12%	13%
2	63%	7%	10%	20%
3	67%	7%	5%	22%
4	55%	22%	7%	17%
5	58%	17%	12%	13%
6	63%	12%	8%	17%
7	58%	12%	12%	18%
8	55%	20%	8%	17%
9	58%	17%	3%	22%
10	47%	15%	10%	28%

Table A.18: Cointegration test results of cubic spline interpolated theoretical averages with reverse random walk added using Johansen test method.

Reversed random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	67%	23%	3%	7%
1	0%	17%	0%	83%
2	67%	23%	7%	3%
3	27%	25%	5%	43%
4	58%	15%	5%	22%
5	45%	23%	5%	27%
6	75%	17%	3%	5%
7	65%	15%	8%	12%
8	25%	23%	3%	48%
9	65%	8%	0%	27%
10	62%	10%	7%	22%

Table A.19: Cointegration test results of linearly interpolated theoretical averages with reverse random walk added using Johansen test method.

Reversed random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	62%	27%	2%	10%
1	15%	23%	3%	58%
2	50%	15%	13%	22%
3	57%	12%	15%	17%
4	72%	7%	12%	10%
5	63%	10%	17%	10%
6	35%	15%	12%	38%
7	57%	12%	13%	18%
8	63%	12%	7%	18%
9	67%	12%	13%	8%
10	58%	13%	13%	15%

Table A.20: Cointegration test results of PCHIP interpolated theoretical averages with reverse random walk added using Johansen test method.

Reversed random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	63%	35%	0%	2%
1	3%	15%	0%	82%
2	67%	22%	7%	5%
3	50%	15%	7%	28%
4	53%	15%	7%	25%
5	60%	10%	5%	25%
6	65%	17%	8%	10%
7	53%	15%	5%	27%
8	60%	12%	7%	22%
9	58%	17%	5%	20%
10	60%	13%	5%	22%

Table A.21: Cointegration test results of pure random walk using Johansen test method.

Pure random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	83%	6%	6%	5%
1	74%	7%	9%	10%
2	76%	6%	12%	6%
3	75%	8%	8%	9%
4	75%	9%	6%	10%
5	69%	12%	8%	11%
6	71%	12%	5%	12%
7	71%	12%	4%	13%
8	69%	13%	7%	11%
9	67%	15%	8%	10%
10	66%	12%	7%	15%

Table A.22: Cointegration test results of cubic spline interpolated pure random walk using Johansen test method.

Pure random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	0%	0%	100%
1	2%	41%	0%	57%
2	1%	36%	0%	63%
3	0%	38%	0%	62%
4	1%	48%	0%	51%
5	6%	73%	0%	21%
6	10%	81%	0%	9%
7	15%	81%	0%	4%
8	18%	81%	0%	1%
9	21%	79%	0%	0%
10	21%	79%	0%	0%

Table A.23: Cointegration test results of linearly interpolated pure random walk using Johansen test method.

Pure random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	85%	5%	6%	4%
1	83%	6%	7%	4%
2	80%	4%	10%	6%
3	80%	7%	7%	6%
4	81%	7%	5%	7%
5	76%	8%	8%	8%
6	82%	5%	7%	6%
7	76%	11%	3%	10%
8	76%	13%	6%	5%
9	75%	12%	5%	8%
10	72%	16%	4%	8%

Table A.24: Cointegration test results of PCHIP interpolated pure random walk using Johansen test method.

Pure random walk Percent Cointegrated				
	Cointegration Null Rejection r = 0, r = 1			
Lags	0, 0	1, 0	0, 1	1, 1
0	0%	0%	0%	100%
1	1%	33%	0%	66%
2	1%	31%	0%	68%
3	1%	29%	0%	70%
4	1%	34%	0%	65%
5	1%	48%	0%	51%
6	6%	62%	1%	31%
7	8%	74%	0%	18%
8	15%	75%	0%	10%
9	18%	79%	0%	3%
10	18%	80%	0%	2%

Appendix B

Interpolated JHTDB Channel Flow Cointegration Results

This appendix contains the results from the cointegration testing of the JHTDB channel flow data set using interpolated data. The data herein is provided for reference purposes only and should be viewed with the knowledge that interpolating the data has been found to significantly skew the cointegration test results. Section 4.1 provides more details of the issues encountered when utilizing interpolated data for cointegration testing.

B.1 Engle-Granger Cointegration Test Results

Using the departure points of the simulation turbulence intensity profile from the theoretical profile as discussed in previous sections, the logarithmic region will initially be defined as $60 \leq y^+ \leq 400$. This is consistent with the general requirements for the log region given by Davidson[5], who states that the log-law occurs in areas where $y^+ \gg 1$.

Starting with this initial logarithmic region, the instantaneous streamwise and turbulence intensity values can be evaluated for cointegration. To achieve this, a subset of data was obtained from the JHTDB for analysis. A set of 30,000 points were queried (200 X points by 150 Z points evenly spaced) for a total

Table B.1: Cointegration test results of interpolated channel flow data using Engle-Granger test method

	Percent Cointegrated			
Lags	U^+ vs $u'u'$	$u'u'$ vs $v'v'$	$u'u'$ vs $w'w'$	$v'v'$ vs $w'w'$
0	68%	0%	0%	1%
1	75%	100%	100%	100%
2	41%	2%	2%	3%

of eight time steps throughout the simulation duration from the wall to the centerline of the channel.

The data from the JHTDB is initially spaced in the y^+ dimension according to the locations of the B-splines used in the simulation. To satisfy the requirements previously discussed for applying time series analysis methods to data taken over the y^+ dimension, the data was made to have equally spaced steps using a simple cubic spline interpolation.

An initial evaluation was conducted using lag parameters of zero, one, and two.

An investigation of the cointegration relationships between a variety of physical quantities was conducted to determine if there were any significant relationships in the initial logarithmic region. The primary physical quantities of interest are the streamwise velocity and turbulence intensity. Results of these investigations are given in Table 4.2 where the Physical Quantities column provides the parameters being evaluated, the Percent Unit Root column gives the percentage of the 240,000 points (30,000 X-Z points over eight time steps) that both quantities contain a unit root per the ADF test, and the Percent Cointegrated column gives the percentage of those points with unit roots that tested positive for cointegration in accordance with the EGCI test.

The results obtained using the interpolated data differ significantly from

Table B.2: Cointegration test results of interpolated channel flow data using Engle-Granger test method - additional quantities

	Percent Cointegrated		
Physical Quantities	0 Lags	1 Lag	2 Lags
u^+ vs. $u'u'$	68%	75%	41%
u^+ vs. $v'v'$	55%	84%	37%
u^+ vs. $w'w'$	52%	86%	33%
u^+ vs. $u'v'$	59%	78%	36%
u^+ vs. $u'w'$	55%	79%	34%
$u'u'$ vs. $v'v'$	0%	100%	2%
$u'u'$ vs. $w'w'$	0%	100%	2%
$u'u'$ vs. $u'v'$	1%	100%	3%
$u'u'$ vs. $u'w'$	1%	100%	3%
$v'v'$ vs. $w'w'$	1%	100%	3%
$v'v'$ vs. $u'v'$	1%	100%	3%
$v'v'$ vs. $u'w'$	0%	100%	2%
$w'w'$ vs. $u'v'$	0%	100%	3%
$w'w'$ vs. $u'w'$	1%	100%	4%
$u'v'$ vs. $u'w'$	1%	100%	3%

those presented in Table 4.2. The percentage of positive cointegration for the one lag case is significantly higher with the interpolated data and the percentages for zero and two lags are also slightly exaggerated. In addition to these quantities, a large variety of additional quantities were tested that were not able to be recreated with the non-interpolated data due to a change in data retrieval methods from the JHTDB server and a lack of time. This data is presented in the following tables.

Again, this data shows a similar pattern with unrealistically high percentages of cointegration reported for the one lag case. They are distinctly different than the results presented for the first four quantities shown in Table B.1 which gives doubt as to if any meaningful relationships would be found if the testing was conducted without interpolated data.

Appendix C

Full JHTDB Buoyancy Flow Data Cointegration Results

This appendix contains the full results tables from the cointegration of the Buoyancy Turbulence flow data set. A full compilation of all lag values evaluated for all physical quantities is provided. Section 4.2.4 contains a detailed description of the calculations used to obtain these results along with a discussion of the implications of this data.

C.1 Cointegration Test Results

C.1.1 Single Point Over Time Engle-Granger Cointegration Results

The following tables contain the full results set for the cointegration testing of the JHTDB Buoyancy Driven Turbulence data taken at a single point over the simulation duration using the EGCI test method.

Table C.1: Cointegration results with zero lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	9%	0.540	100%	0%	0.946	61%	7%	0.629
Kinetic Energy vs R_{hh}	66%	4%	0.653	100%	1%	0.944	67%	5%	0.631
R_{hh} vs R_{vv}	55%	0%	0.759	100%	1%	0.946	48%	0%	0.797
u_x vs. u_y	80%	0%	0.801	100%	20%	0.711	69%	0%	0.778
u_x vs. u_z	79%	0%	0.803	100%	17%	0.737	68%	0%	0.789
u_y vs. u_z	69%	0%	0.778	100%	11%	0.810	68%	0%	0.780
u_x vs. R_{vv}	74%	4%	0.726	100%	47%	0.399	61%	5%	0.725
u_y vs. R_{hh}	62%	1%	0.781	100%	15%	0.753	66%	1%	0.743
u_z vs. R_{hh}	61%	1%	0.774	100%	17%	0.734	66%	1%	0.734
ρ vs. $\partial\rho/\partial x$	0%	0%	1.000	39%	9%	0.880	0%	0%	0.998
ρ vs. $\partial\rho/\partial y$	0%	0%	1.000	40%	10%	0.871	0%	0%	0.997
ρ vs. $\partial\rho/\partial z$	0%	0%	1.000	40%	9%	0.874	0%	0%	0.998
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	50%	2%	0.935	0%	0%	1.000
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	1.000	54%	1%	0.940	0%	0%	1.000
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	1.000	53%	1%	0.945	0%	0%	0.999

Table C.2: Cointegration results with one lag for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	72%	0.272	100%	89%	0.072	61%	59%	0.394
Kinetic Energy vs R_{hh}	66%	62%	0.345	100%	92%	0.050	67%	64%	0.334
R_{hh} vs R_{vv}	55%	51%	0.461	100%	92%	0.048	48%	45%	0.529
u_x vs. u_y	80%	66%	0.230	100%	75%	0.168	69%	58%	0.333
u_x vs. u_z	79%	65%	0.235	100%	74%	0.180	68%	55%	0.341
u_y vs. u_z	69%	67%	0.311	100%	77%	0.165	68%	64%	0.329
u_x vs. R_{vv}	74%	63%	0.287	100%	66%	0.266	61%	54%	0.402
u_y vs. R_{hh}	62%	59%	0.389	100%	72%	0.215	66%	63%	0.346
u_z vs. R_{hh}	61%	59%	0.392	100%	68%	0.254	66%	62%	0.349
ρ vs. $\partial\rho/\partial x$	0%	0%	0.998	39%	31%	0.656	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.998	40%	33%	0.646	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	32%	0.649	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	41%	0.573	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	43%	0.536	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	43%	0.550	0%	0%	0.999

Table C.3: Cointegration results with two lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	44%	0.335	100%	3%	0.795	61%	34%	0.457
Kinetic Energy vs R_{hh}	66%	31%	0.432	100%	4%	0.782	67%	31%	0.421
R_{hh} vs R_{vv}	55%	22%	0.545	100%	4%	0.780	48%	16%	0.609
u_x vs. u_y	80%	10%	0.498	100%	25%	0.621	69%	10%	0.537
u_x vs. u_z	79%	9%	0.503	100%	25%	0.624	68%	9%	0.549
u_y vs. u_z	69%	15%	0.470	100%	20%	0.691	68%	11%	0.500
u_x vs. R_{vv}	74%	24%	0.476	100%	12%	0.638	61%	22%	0.537
u_y vs. R_{hh}	62%	16%	0.520	100%	11%	0.705	66%	16%	0.485
u_z vs. R_{hh}	61%	18%	0.513	100%	8%	0.737	66%	18%	0.479
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	11%	0.854	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	12%	0.844	0%	0%	0.997
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	11%	0.851	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	50%	6%	0.870	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	7%	0.858	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	5%	0.868	0%	0%	0.999

Table C.4: Cointegration results with three lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	67%	0.280	100%	36%	0.437	61%	54%	0.405
Kinetic Energy vs R_{hh}	66%	55%	0.361	100%	40%	0.373	67%	56%	0.351
R_{hh} vs R_{vv}	55%	44%	0.476	100%	40%	0.373	48%	37%	0.545
u_x vs. u_y	80%	38%	0.303	100%	35%	0.463	69%	35%	0.390
u_x vs. u_z	79%	38%	0.311	100%	37%	0.441	68%	34%	0.400
u_y vs. u_z	69%	51%	0.340	100%	37%	0.436	68%	46%	0.366
u_x vs. R_{vv}	74%	47%	0.341	100%	23%	0.490	61%	41%	0.443
u_y vs. R_{hh}	62%	47%	0.414	100%	25%	0.496	66%	49%	0.375
u_z vs. R_{hh}	61%	50%	0.412	100%	24%	0.526	66%	50%	0.373
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	19%	0.728	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	20%	0.723	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	19%	0.728	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	23%	0.688	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	24%	0.679	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	22%	0.694	0%	0%	0.999

Table C.5: Cointegration results with four lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	52%	0.311	100%	5%	0.749	61%	40%	0.436
Kinetic Energy vs R_{hh}	66%	40%	0.400	100%	4%	0.728	67%	39%	0.392
R_{hh} vs R_{vv}	55%	29%	0.513	100%	4%	0.725	48%	23%	0.582
u_x vs. u_y	80%	16%	0.428	100%	15%	0.626	69%	15%	0.483
u_x vs. u_z	79%	13%	0.435	100%	15%	0.631	68%	13%	0.494
u_y vs. u_z	69%	26%	0.415	100%	15%	0.657	68%	21%	0.446
u_x vs. R_{vv}	74%	30%	0.424	100%	4%	0.657	61%	27%	0.501
u_y vs. R_{hh}	62%	26%	0.472	100%	6%	0.670	66%	25%	0.439
u_z vs. R_{hh}	61%	28%	0.470	100%	5%	0.707	66%	28%	0.436
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	11%	0.825	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	12%	0.812	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	11%	0.819	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	50%	10%	0.777	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	11%	0.758	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	11%	0.770	0%	0%	0.999

Table C.6: Cointegration results with five lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	60%	0.294	100%	12%	0.627	61%	47%	0.420
Kinetic Energy vs R_{hh}	66%	49%	0.374	100%	15%	0.563	67%	50%	0.367
R_{hh} vs R_{vv}	55%	39%	0.488	100%	15%	0.561	48%	33%	0.558
u_x vs. u_y	80%	27%	0.356	100%	20%	0.551	69%	26%	0.431
u_x vs. u_z	79%	26%	0.363	100%	18%	0.559	68%	23%	0.441
u_y vs. u_z	69%	40%	0.368	100%	22%	0.554	68%	34%	0.397
u_x vs. R_{vv}	74%	39%	0.377	100%	7%	0.596	61%	35%	0.468
u_y vs. R_{hh}	62%	38%	0.436	100%	11%	0.588	66%	37%	0.400
u_z vs. R_{hh}	61%	40%	0.432	100%	8%	0.628	66%	40%	0.397
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	12%	0.797	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	12%	0.785	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	12%	0.793	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	15%	0.728	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	16%	0.710	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	16%	0.709	0%	0%	0.999

Table C.7: Cointegration results with six lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	50%	0.311	100%	4%	0.734	61%	39%	0.437
Kinetic Energy vs R_{hh}	66%	41%	0.396	100%	5%	0.695	67%	38%	0.389
R_{hh} vs R_{vv}	55%	31%	0.508	100%	5%	0.693	48%	23%	0.578
u_x vs. u_y	80%	18%	0.413	100%	13%	0.637	69%	18%	0.472
u_x vs. u_z	79%	16%	0.422	100%	12%	0.641	68%	15%	0.486
u_y vs. u_z	69%	28%	0.405	100%	12%	0.658	68%	23%	0.436
u_x vs. R_{vv}	74%	31%	0.416	100%	3%	0.651	61%	28%	0.494
u_y vs. R_{hh}	62%	28%	0.466	100%	5%	0.661	66%	26%	0.432
u_z vs. R_{hh}	61%	30%	0.461	100%	3%	0.698	66%	28%	0.429
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	8%	0.837	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	9%	0.826	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	8%	0.832	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	9%	0.763	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	11%	0.746	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	11%	0.740	0%	0%	0.999

Table C.8: Cointegration results with seven lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	55%	0.305	100%	7%	0.687	61%	41%	0.431
Kinetic Energy vs R_{hh}	66%	45%	0.386	100%	7%	0.652	67%	43%	0.381
R_{hh} vs R_{vv}	55%	35%	0.498	100%	8%	0.645	48%	27%	0.569
u_x vs. u_y	80%	22%	0.387	100%	11%	0.623	69%	21%	0.452
u_x vs. u_z	79%	21%	0.395	100%	11%	0.626	68%	19%	0.465
u_y vs. u_z	69%	34%	0.387	100%	13%	0.620	68%	27%	0.418
u_x vs. R_{vv}	74%	35%	0.398	100%	4%	0.637	61%	31%	0.483
u_y vs. R_{hh}	62%	32%	0.453	100%	6%	0.626	66%	31%	0.417
u_z vs. R_{hh}	61%	35%	0.447	100%	5%	0.659	66%	33%	0.414
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	7%	0.828	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	9%	0.819	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	8%	0.819	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	10%	0.751	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	11%	0.733	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	12%	0.733	0%	0%	0.999

Table C.9: Cointegration results with eight lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	48%	0.317	100%	4%	0.753	61%	36%	0.442
Kinetic Energy vs R_{hh}	66%	40%	0.398	100%	5%	0.705	67%	37%	0.393
R_{hh} vs R_{vv}	55%	30%	0.509	100%	6%	0.701	48%	21%	0.580
u_x vs. u_y	80%	17%	0.417	100%	9%	0.648	69%	16%	0.474
u_x vs. u_z	79%	14%	0.425	100%	10%	0.661	68%	14%	0.487
u_y vs. u_z	69%	27%	0.411	100%	9%	0.671	68%	22%	0.442
u_x vs. R_{vv}	74%	29%	0.420	100%	4%	0.639	61%	28%	0.496
u_y vs. R_{hh}	62%	25%	0.472	100%	5%	0.654	66%	25%	0.438
u_z vs. R_{hh}	61%	30%	0.464	100%	3%	0.690	66%	27%	0.431
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	6%	0.842	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	7%	0.835	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	7%	0.839	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	7%	0.769	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	9%	0.751	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	9%	0.749	0%	0%	0.999

Table C.10: Cointegration results with nine lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	50%	0.313	100%	6%	0.705	61%	37%	0.440
Kinetic Energy vs R_{hh}	66%	41%	0.394	100%	9%	0.637	67%	38%	0.388
R_{hh} vs R_{vv}	55%	32%	0.506	100%	9%	0.635	48%	24%	0.575
u_x vs. u_y	80%	19%	0.407	100%	10%	0.629	69%	18%	0.466
u_x vs. u_z	79%	17%	0.415	100%	10%	0.650	68%	16%	0.480
u_y vs. u_z	69%	30%	0.402	100%	10%	0.644	68%	24%	0.433
u_x vs. R_{vv}	74%	30%	0.412	100%	5%	0.639	61%	28%	0.492
u_y vs. R_{hh}	62%	28%	0.465	100%	4%	0.650	66%	28%	0.430
u_z vs. R_{hh}	61%	29%	0.458	100%	3%	0.679	66%	29%	0.425
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	6%	0.839	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	6%	0.841	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	6%	0.842	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	8%	0.765	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	9%	0.747	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	9%	0.747	0%	0%	0.999

Table C.11: Cointegration results with ten lags for single point over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	73%	45%	0.324	100%	4%	0.731	61%	35%	0.450
Kinetic Energy vs R_{hh}	66%	38%	0.402	100%	6%	0.680	67%	35%	0.397
R_{hh} vs R_{vv}	55%	29%	0.513	100%	6%	0.678	48%	21%	0.582
u_x vs. u_y	80%	15%	0.426	100%	8%	0.663	69%	14%	0.480
u_x vs. u_z	79%	14%	0.434	100%	8%	0.684	68%	13%	0.494
u_y vs. u_z	69%	25%	0.416	100%	9%	0.677	68%	20%	0.450
u_x vs. R_{vv}	74%	28%	0.428	100%	3%	0.661	61%	25%	0.502
u_y vs. R_{hh}	62%	24%	0.477	100%	4%	0.655	66%	23%	0.443
u_z vs. R_{hh}	61%	27%	0.468	100%	3%	0.691	66%	26%	0.437
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	39%	5%	0.853	0%	0%	0.997
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	40%	5%	0.849	0%	0%	0.996
ρ vs. $\partial\rho/\partial z$	0%	0%	0.998	40%	6%	0.848	0%	0%	0.996
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	0.999	50%	6%	0.772	0%	0%	0.999
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	54%	6%	0.761	0%	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	0.999	53%	8%	0.758	0%	0%	0.999

C.1.2 Single Point Over Time Johansen Cointegration Results

The following tables contain the full results set for the cointegration testing of the JHTDB Buoyancy Driven Turbulence data taken at a single point over the simulation duration using the Johansen test method.

Table C.12: Cointegration results with zero lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	42%	0.061	100%	0.153	65%	0.056
Kinetic vs. R_{hh}	53%	0.060	100%	0.154	64%	0.057
R_{hh} vs. R_{vv}	81%	0.055	100%	0.158	96%	0.054
u_x vs. u_y	35%	0.157	60%	0.095	78%	0.144
u_x vs. u_z	42%	0.162	68%	0.086	51%	0.145
u_y vs. u_z	27%	0.104	69%	0.093	35%	0.118
u_x vs. R_{vv}	43%	0.369	81%	0.150	60%	0.234
u_y vs. R_{hh}	28%	0.121	100%	0.157	73%	0.111
u_z vs. R_{hh}	37%	0.107	100%	0.143	73%	0.104
ρ vs. $\partial\rho/\partial x$	76%	0.027	61%	0.060	16%	0.006
ρ vs. $\partial\rho/\partial y$	88%	0.036	62%	0.070	16%	0.006
ρ vs. $\partial\rho/\partial z$	91%	0.036	49%	0.057	11%	0.005
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	7%	0.003	75%	0.150	20%	0.004
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	7%	0.003	77%	0.158	11%	0.005
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	9%	0.003	83%	0.155	8%	0.004

Table C.13: Cointegration results with one lag for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	5%	0.003	76%	0.159	1%	0.002
Kinetic vs. R_{hh}	4%	0.002	76%	0.159	1%	0.002
R_{hh} vs. R_{vv}	14%	0.002	98%	0.161	2%	0.001
u_x vs. u_y	18%	0.007	100%	0.104	4%	0.005
u_x vs. u_z	5%	0.006	65%	0.099	2%	0.005
u_y vs. u_z	4%	0.003	100%	0.121	7%	0.003
u_x vs. R_{vv}	22%	0.011	92%	0.137	9%	0.009
u_y vs. R_{hh}	0%	0.002	100%	0.126	1%	0.002
u_z vs. R_{hh}	3%	0.002	85%	0.140	1%	0.002
ρ vs. $\partial\rho/\partial x$	49%	0.016	81%	0.107	7%	0.003
ρ vs. $\partial\rho/\partial y$	43%	0.013	61%	0.103	8%	0.003
ρ vs. $\partial\rho/\partial z$	47%	0.013	63%	0.111	3%	0.003
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.157	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	85%	0.142	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	100%	0.145	0%	0.001

Table C.14: Cointegration results with two lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	35%	0.020	100%	0.237	60%	0.017
Kinetic vs. R_{hh}	33%	0.020	100%	0.239	60%	0.017
R_{hh} vs. R_{vv}	70%	0.018	100%	0.237	94%	0.017
u_x vs. u_y	52%	0.066	100%	0.155	80%	0.057
u_x vs. u_z	75%	0.060	88%	0.160	100%	0.056
u_y vs. u_z	64%	0.036	100%	0.144	54%	0.041
u_x vs. R_{vv}	80%	0.106	95%	0.163	77%	0.080
u_y vs. R_{hh}	66%	0.030	100%	0.192	53%	0.032
u_z vs. R_{hh}	43%	0.027	100%	0.193	62%	0.032
ρ vs. $\partial\rho/\partial x$	57%	0.018	76%	0.117	9%	0.004
ρ vs. $\partial\rho/\partial y$	72%	0.021	100%	0.120	6%	0.004
ρ vs. $\partial\rho/\partial z$	70%	0.020	66%	0.124	5%	0.003
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	3%	0.001	100%	0.213	0%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	100%	0.216	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	94%	0.216	0%	0.001

Table C.15: Cointegration results with three lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	9%	0.006	100%	0.207	2%	0.005
Kinetic vs. R_{hh}	9%	0.006	100%	0.203	2%	0.005
R_{hh} vs. R_{vv}	23%	0.005	100%	0.198	10%	0.006
u_x vs. u_y	50%	0.023	90%	0.190	62%	0.019
u_x vs. u_z	51%	0.020	94%	0.176	28%	0.019
u_y vs. u_z	23%	0.011	100%	0.169	36%	0.013
u_x vs. R_{vv}	78%	0.032	87%	0.190	63%	0.027
u_y vs. R_{hh}	18%	0.009	100%	0.213	23%	0.010
u_z vs. R_{hh}	9%	0.006	100%	0.197	21%	0.008
ρ vs. $\partial\rho/\partial x$	62%	0.022	100%	0.153	6%	0.004
ρ vs. $\partial\rho/\partial y$	72%	0.022	100%	0.148	6%	0.004
ρ vs. $\partial\rho/\partial z$	75%	0.022	92%	0.156	5%	0.004
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.227	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	93%	0.239	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	96%	0.233	0%	0.001

Table C.16: Cointegration results with four lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	19%	0.014	100%	0.236	17%	0.013
Kinetic vs. R_{hh}	21%	0.013	100%	0.239	16%	0.013
R_{hh} vs. R_{vv}	49%	0.012	100%	0.241	42%	0.013
u_x vs. u_y	73%	0.050	100%	0.188	79%	0.044
u_x vs. u_z	57%	0.043	98%	0.186	57%	0.043
u_y vs. u_z	64%	0.025	100%	0.184	38%	0.031
u_x vs. R_{vv}	81%	0.071	92%	0.189	72%	0.060
u_y vs. R_{hh}	14%	0.020	97%	0.215	40%	0.022
u_z vs. R_{hh}	33%	0.016	100%	0.206	56%	0.022
ρ vs. $\partial\rho/\partial x$	60%	0.022	100%	0.189	14%	0.004
ρ vs. $\partial\rho/\partial y$	61%	0.022	100%	0.173	5%	0.004
ρ vs. $\partial\rho/\partial z$	64%	0.022	100%	0.185	12%	0.004
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.233	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	90%	0.232	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	100%	0.245	0%	0.001

Table C.17: Cointegration results with five lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	12%	0.010	100%	0.212	10%	0.010
Kinetic vs. R_{hh}	11%	0.010	100%	0.217	10%	0.010
R_{hh} vs. R_{vv}	31%	0.009	100%	0.211	27%	0.009
u_x vs. u_y	62%	0.036	93%	0.180	80%	0.033
u_x vs. u_z	51%	0.032	88%	0.179	48%	0.032
u_y vs. u_z	40%	0.016	100%	0.182	56%	0.021
u_x vs. R_{vv}	84%	0.049	100%	0.196	61%	0.043
u_y vs. R_{hh}	23%	0.013	100%	0.206	38%	0.016
u_z vs. R_{hh}	33%	0.010	100%	0.205	35%	0.014
ρ vs. $\partial\rho/\partial x$	64%	0.025	100%	0.191	12%	0.005
ρ vs. $\partial\rho/\partial y$	73%	0.025	100%	0.193	9%	0.005
ρ vs. $\partial\rho/\partial z$	67%	0.025	100%	NaN	15%	0.006
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.224	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	95%	0.233	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	100%	0.233	0%	0.001

Table C.18: Cointegration results with six lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	15%	0.014	100%	0.234	14%	0.014
Kinetic vs. R_{hh}	15%	0.014	100%	0.233	15%	0.014
R_{hh} vs. R_{vv}	39%	0.013	100%	0.234	38%	0.013
u_x vs. u_y	68%	0.050	93%	0.185	88%	0.046
u_x vs. u_z	56%	0.045	90%	0.174	66%	0.045
u_y vs. u_z	43%	0.024	100%	0.189	46%	0.030
u_x vs. R_{vv}	79%	0.065	95%	0.192	83%	0.058
u_y vs. R_{hh}	17%	0.019	100%	0.196	51%	0.022
u_z vs. R_{hh}	33%	0.014	100%	0.199	57%	0.021
ρ vs. $\partial\rho/\partial x$	75%	0.026	100%	0.241	14%	0.005
ρ vs. $\partial\rho/\partial y$	75%	0.026	100%	0.232	9%	0.006
ρ vs. $\partial\rho/\partial z$	71%	0.026	100%	0.225	9%	0.006
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.236	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	95%	0.241	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	96%	0.235	0%	0.001

Table C.19: Cointegration results with seven lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	12%	0.012	100%	0.213	16%	0.014
Kinetic vs. R_{hh}	13%	0.012	100%	0.218	16%	0.014
R_{hh} vs. R_{vv}	30%	0.011	100%	0.222	42%	0.013
u_x vs. u_y	66%	0.046	95%	0.190	73%	0.043
u_x vs. u_z	68%	0.041	100%	0.171	58%	0.040
u_y vs. u_z	74%	0.022	100%	0.181	52%	0.027
u_x vs. R_{vv}	76%	0.058	93%	0.192	71%	0.052
u_y vs. R_{hh}	23%	0.017	100%	0.196	52%	0.020
u_z vs. R_{hh}	23%	0.013	100%	0.202	54%	0.019
ρ vs. $\partial\rho/\partial x$	81%	0.029	100%	0.255	18%	0.007
ρ vs. $\partial\rho/\partial y$	79%	0.028	100%	0.235	20%	0.006
ρ vs. $\partial\rho/\partial z$	75%	0.028	100%	0.243	11%	0.006
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0.001	100%	0.239	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	95%	0.245	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.001	100%	0.233	0%	0.001

Table C.20: Cointegration results with eight lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	15%	0.015	100%	0.220	20%	0.017
Kinetic vs. R_{hh}	17%	0.015	100%	0.215	21%	0.017
R_{hh} vs. R_{vv}	37%	0.013	100%	0.223	50%	0.016
u_x vs. u_y	56%	0.052	92%	0.184	75%	0.051
u_x vs. u_z	55%	0.048	97%	0.174	65%	0.048
u_y vs. u_z	67%	0.025	93%	0.177	49%	0.031
u_x vs. R_{vv}	71%	0.067	91%	0.179	57%	0.060
u_y vs. R_{hh}	20%	0.020	100%	0.194	50%	0.024
u_z vs. R_{hh}	42%	0.016	100%	0.205	48%	0.024
ρ vs. $\partial\rho/\partial x$	83%	0.030	100%	0.259	12%	0.007
ρ vs. $\partial\rho/\partial y$	77%	0.030	100%	0.237	12%	0.007
ρ vs. $\partial\rho/\partial z$	75%	0.030	100%	0.254	13%	0.007
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	3%	0.001	100%	0.223	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	100%	0.240	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	1%	0.001	96%	0.241	0%	0.001

Table C.21: Cointegration results with nine lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	15%	0.015	100%	0.217	17%	0.018
Kinetic vs. R_{hh}	15%	0.015	100%	0.216	17%	0.018
R_{hh} vs. R_{vv}	44%	0.013	100%	0.214	42%	0.018
u_x vs. u_y	63%	0.053	100%	0.175	78%	0.050
u_x vs. u_z	56%	0.047	97%	0.174	55%	0.048
u_y vs. u_z	71%	0.025	89%	0.186	60%	0.032
u_x vs. R_{vv}	56%	0.064	92%	0.186	59%	0.058
u_y vs. R_{hh}	27%	0.020	94%	0.199	45%	0.024
u_z vs. R_{hh}	21%	0.015	100%	0.199	30%	0.024
ρ vs. $\partial\rho/\partial x$	77%	0.032	100%	0.287	16%	0.008
ρ vs. $\partial\rho/\partial y$	78%	0.033	100%	0.269	23%	0.009
ρ vs. $\partial\rho/\partial z$	81%	0.032	100%	0.282	19%	0.008
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	5%	0.001	100%	0.213	0%	0.001
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	95%	0.243	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	97%	0.248	0%	0.001

Table C.22: Cointegration results with ten lags for single point over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	20%	0.016	100%	0.224	48%	0.021
Kinetic vs. R_{hh}	19%	0.016	100%	0.215	47%	0.021
R_{hh} vs. R_{vv}	48%	0.014	100%	0.211	86%	0.021
u_x vs. u_y	61%	0.056	100%	0.172	72%	0.054
u_x vs. u_z	73%	0.051	100%	0.181	60%	0.054
u_y vs. u_z	72%	0.028	94%	0.181	58%	0.035
u_x vs. R_{vv}	73%	0.070	91%	0.178	54%	0.065
u_y vs. R_{hh}	21%	0.022	100%	0.190	48%	0.026
u_z vs. R_{hh}	26%	0.018	100%	0.189	52%	0.027
ρ vs. $\partial\rho/\partial x$	77%	0.033	100%	0.299	14%	0.008
ρ vs. $\partial\rho/\partial y$	84%	0.034	100%	0.292	24%	0.009
ρ vs. $\partial\rho/\partial z$	79%	0.034	100%	0.295	18%	0.009
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	1%	0.001	100%	0.218	1%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0.001	93%	0.230	0%	0.001
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.001	100%	0.252	0%	0.001

C.1.3 Single Point Over Time Decaying Range Start Time Variation Cointegration Impacts

In an attempt to alleviate any potential discrepancies caused by the transition period from increasing to decaying turbulence intensity, the start of the decaying time range was modified to 13s instead of the maximum Reynolds Stress time at 6.5s as in the previous results. Lag values of four, seven, and ten were not tested as the impact of this change became obvious from the results obtained with the other lag values. Only the results from the decaying time range are presented in the following tables.

C.1.4 Particle Track Engle-Grange Cointegration Results

The following tables contain the full results set for the cointegration testing of the JHTDB Buoyancy Driven Turbulence data taken at along a particle track over the simulation duration.

Table C.23: Cointegration results with zero, one, and two lags for modified decaying time range for single point over time using Engle-Granger test method in buoyancy flow

		Zero Lag		One Lag		Two Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	46%	7%	0.709	46%	0.539	26%	0.586
Kinetic Energy vs R_{hh}	42%	4%	0.764	41%	0.582	19%	0.636
R_{hh} vs R_{vv}	29%	0%	0.871	29%	0.708	9%	0.756
u_x vs. u_y	50%	1%	0.814	44%	0.512	9%	0.652
u_x vs. u_z	52%	1%	0.819	45%	0.494	7%	0.646
u_y vs. u_z	46%	0%	0.835	43%	0.547	7%	0.671
u_x vs. R_{vv}	56%	8%	0.715	50%	0.455	22%	0.572
u_y vs. R_{hh}	46%	2%	0.818	43%	0.550	10%	0.663
u_z vs. R_{hh}	46%	2%	0.806	44%	0.548	11%	0.657
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial z$	0%	0%	0.999	0%	0.999	0%	0.999
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	0%	1.000	0%	1.000
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	1.000	0%	0.999	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	1.000	0%	1.000	0%	1.000

Table C.24: Cointegration results with three, five, and six lags for modified decaying time range for single point over time using Engle-Granger test method in buoyancy flow

		Three Lag		Five Lag		Six Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	46%	42%	0.546	37%	0.556	29%	0.572
Kinetic Energy vs R_{hh}	42%	37%	0.592	32%	0.603	25%	0.619
R_{hh} vs R_{vv}	29%	25%	0.717	20%	0.728	15%	0.742
u_x vs. u_y	50%	31%	0.548	21%	0.575	14%	0.613
u_x vs. u_z	52%	30%	0.533	22%	0.566	13%	0.604
u_y vs. u_z	46%	31%	0.573	22%	0.594	15%	0.627
u_x vs. R_{vv}	56%	40%	0.487	34%	0.514	26%	0.544
u_y vs. R_{hh}	46%	31%	0.573	24%	0.595	15%	0.627
u_z vs. R_{hh}	46%	33%	0.569	24%	0.590	16%	0.619
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial z$	0%	0%	0.999	0%	0.999	0%	0.999
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	0%	1.000	0%	1.000
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	0.999	0%	0.999	0%	0.999
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	1.000	0%	1.000	0%	1.000

Table C.25: Cointegration results with eight and nine lags for modified decaying time range for single point over time using Engle-Granger tet method in buoyancy flow

		Zero Lag		One Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	46%	27%	0.576	28%	0.574
Kinetic Energy vs R_{hh}	42%	23%	0.622	24%	0.621
R_{hh} vs R_{vv}	29%	13%	0.744	14%	0.742
u_x vs. u_y	50%	13%	0.618	13%	0.613
u_x vs. u_z	52%	13%	0.610	15%	0.603
u_y vs. u_z	46%	14%	0.630	15%	0.624
u_x vs. R_{vv}	56%	24%	0.547	26%	0.543
u_y vs. R_{hh}	46%	15%	0.632	16%	0.628
u_z vs. R_{hh}	46%	15%	0.626	16%	0.619
ρ vs. $\partial\rho/\partial x$	0%	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial y$	0%	0%	0.999	0%	0.999
ρ vs. $\partial\rho/\partial z$	0%	0%	0.999	0%	0.999
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	0%	0%	1.000	0%	1.000
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	0%	0%	1.000	0%	1.000
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0%	1.000	0%	1.000

Table C.26: Cointegration results with zero lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	1%	0.742	100%	1%	0.943	74%	1%	0.804
Kinetic Energy vs R_{hh}	90%	1%	0.782	100%	0%	0.924	83%	2%	0.771
R_{hh} vs R_{vv}	85%	0%	0.820	100%	0%	0.925	70%	0%	0.856
u_x vs. u_y	94%	0%	0.891	100%	20%	0.711	78%	1%	0.855
u_x vs. u_z	93%	0%	0.888	100%	19%	0.730	78%	1%	0.852
u_y vs. u_z	92%	0%	0.885	100%	11%	0.823	87%	0%	0.859
u_x vs. R_{vv}	90%	2%	0.834	100%	39%	0.498	75%	4%	0.805
u_y vs. R_{hh}	89%	0%	0.875	100%	15%	0.771	86%	2%	0.826
u_z vs. R_{hh}	89%	0%	0.877	100%	17%	0.736	86%	0%	0.832
ρ vs. $\partial\rho/\partial x$	15%	2%	0.941	27%	8%	0.905	18%	7%	0.872
ρ vs. $\partial\rho/\partial y$	17%	3%	0.932	26%	9%	0.887	20%	8%	0.854
ρ vs. $\partial\rho/\partial z$	15%	2%	0.930	24%	8%	0.902	19%	8%	0.861
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	0%	0.983	41%	3%	0.941	11%	0%	0.968
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	0%	0.986	45%	3%	0.944	8%	0%	0.979
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	0%	0.991	42%	3%	0.943	7%	0%	0.984

Table C.27: Cointegration results with one lag for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	90%	0.104	100%	92%	0.059	74%	74%	0.258
Kinetic Energy vs R_{hh}	90%	89%	0.106	100%	95%	0.039	83%	82%	0.179
R_{hh} vs R_{vv}	85%	84%	0.152	100%	95%	0.038	70%	69%	0.299
u_x vs. u_y	94%	92%	0.065	100%	81%	0.149	78%	76%	0.222
u_x vs. u_z	93%	92%	0.073	100%	81%	0.151	78%	76%	0.222
u_y vs. u_z	92%	91%	0.081	100%	85%	0.124	87%	86%	0.138
u_x vs. R_{vv}	90%	88%	0.107	100%	68%	0.271	75%	73%	0.259
u_y vs. R_{hh}	89%	89%	0.110	100%	82%	0.146	86%	85%	0.140
u_z vs. R_{hh}	89%	88%	0.116	100%	77%	0.176	86%	85%	0.144
ρ vs. $\partial\rho/\partial x$	15%	2%	0.941	27%	16%	0.813	18%	4%	0.886
ρ vs. $\partial\rho/\partial y$	17%	1%	0.939	26%	16%	0.812	20%	5%	0.873
ρ vs. $\partial\rho/\partial z$	15%	1%	0.940	24%	16%	0.825	19%	5%	0.875
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.945	41%	21%	0.743	11%	8%	0.900
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.954	45%	22%	0.732	8%	6%	0.928
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.967	42%	22%	0.738	7%	5%	0.938

Table C.28: Cointegration results with two lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	12%	0.415	100%	2%	0.869	74%	8%	0.549
Kinetic Energy vs R_{hh}	90%	11%	0.458	100%	2%	0.843	83%	10%	0.516
R_{hh} vs R_{vv}	85%	7%	0.508	100%	2%	0.842	70%	5%	0.622
u_x vs. u_y	94%	4%	0.624	100%	31%	0.564	78%	4%	0.660
u_x vs. u_z	93%	4%	0.627	100%	28%	0.573	78%	3%	0.660
u_y vs. u_z	92%	5%	0.533	100%	29%	0.583	87%	5%	0.570
u_x vs. R_{vv}	90%	8%	0.569	100%	22%	0.553	75%	8%	0.607
u_y vs. R_{hh}	89%	6%	0.547	100%	16%	0.645	86%	5%	0.553
u_z vs. R_{hh}	89%	6%	0.545	100%	18%	0.626	86%	5%	0.553
ρ vs. $\partial\rho/\partial x$	15%	1%	0.937	27%	11%	0.844	18%	5%	0.881
ρ vs. $\partial\rho/\partial y$	17%	1%	0.935	26%	12%	0.837	20%	5%	0.870
ρ vs. $\partial\rho/\partial z$	15%	1%	0.937	24%	11%	0.851	19%	5%	0.874
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	6%	0.941	41%	12%	0.813	11%	10%	0.895
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	5%	0.952	45%	12%	0.808	8%	7%	0.925
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.964	42%	11%	0.815	7%	6%	0.935

Table C.29: Cointegration results with three lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	65%	0.175	100%	21%	0.579	74%	50%	0.330
Kinetic Energy vs R_{hh}	90%	58%	0.208	100%	26%	0.523	83%	52%	0.283
R_{hh} vs R_{vv}	85%	54%	0.249	100%	27%	0.520	70%	43%	0.396
u_x vs. u_y	94%	34%	0.313	100%	31%	0.475	78%	27%	0.418
u_x vs. u_z	93%	34%	0.316	100%	32%	0.473	78%	29%	0.419
u_y vs. u_z	92%	48%	0.241	100%	35%	0.459	87%	42%	0.296
u_x vs. R_{vv}	90%	46%	0.282	100%	16%	0.550	75%	39%	0.387
u_y vs. R_{hh}	89%	50%	0.252	100%	18%	0.524	86%	46%	0.279
u_z vs. R_{hh}	89%	51%	0.252	100%	17%	0.550	86%	48%	0.278
ρ vs. $\partial\rho/\partial x$	15%	1%	0.943	27%	10%	0.856	18%	4%	0.886
ρ vs. $\partial\rho/\partial y$	17%	0%	0.944	26%	10%	0.853	20%	4%	0.878
ρ vs. $\partial\rho/\partial z$	15%	1%	0.944	24%	10%	0.864	19%	4%	0.881
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.941	41%	8%	0.848	11%	9%	0.896
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.951	45%	8%	0.844	8%	7%	0.925
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.964	42%	9%	0.848	7%	6%	0.935

Table C.30: Cointegration results with four lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	64%	0.164	100%	7%	0.742	74%	49%	0.320
Kinetic Energy vs R_{hh}	90%	54%	0.196	100%	7%	0.723	83%	46%	0.281
R_{hh} vs R_{vv}	85%	50%	0.243	100%	7%	0.722	70%	37%	0.397
u_x vs. u_y	94%	32%	0.275	100%	26%	0.508	78%	28%	0.395
u_x vs. u_z	93%	34%	0.275	100%	25%	0.511	78%	28%	0.392
u_y vs. u_z	92%	51%	0.196	100%	25%	0.497	87%	42%	0.261
u_x vs. R_{vv}	90%	39%	0.273	100%	7%	0.605	75%	33%	0.386
u_y vs. R_{hh}	89%	47%	0.219	100%	11%	0.598	86%	41%	0.258
u_z vs. R_{hh}	89%	50%	0.222	100%	10%	0.607	86%	43%	0.257
ρ vs. $\partial\rho/\partial x$	15%	1%	0.950	27%	8%	0.877	18%	3%	0.892
ρ vs. $\partial\rho/\partial y$	17%	0%	0.949	26%	8%	0.871	20%	3%	0.884
ρ vs. $\partial\rho/\partial z$	15%	1%	0.949	24%	8%	0.879	19%	3%	0.886
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.942	41%	6%	0.877	11%	8%	0.899
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.951	45%	5%	0.875	8%	6%	0.927
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.965	42%	5%	0.879	7%	5%	0.938

Table C.31: Cointegration results with five lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	61%	0.165	100%	9%	0.702	74%	44%	0.325
Kinetic Energy vs R_{hh}	90%	54%	0.190	100%	11%	0.662	83%	46%	0.275
R_{hh} vs R_{vv}	85%	49%	0.237	100%	11%	0.661	70%	36%	0.392
u_x vs. u_y	94%	26%	0.295	100%	23%	0.523	78%	22%	0.412
u_x vs. u_z	93%	27%	0.299	100%	22%	0.535	78%	22%	0.413
u_y vs. u_z	92%	43%	0.212	100%	23%	0.530	87%	35%	0.280
u_x vs. R_{vv}	90%	37%	0.279	100%	5%	0.627	75%	32%	0.389
u_y vs. R_{hh}	89%	43%	0.231	100%	9%	0.605	86%	37%	0.269
u_z vs. R_{hh}	89%	43%	0.234	100%	8%	0.624	86%	37%	0.268
ρ vs. $\partial\rho/\partial x$	15%	1%	0.952	27%	7%	0.886	18%	3%	0.894
ρ vs. $\partial\rho/\partial y$	17%	0%	0.951	26%	7%	0.877	20%	3%	0.885
ρ vs. $\partial\rho/\partial z$	15%	1%	0.950	24%	7%	0.889	19%	3%	0.887
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.943	41%	5%	0.869	11%	7%	0.901
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.952	45%	5%	0.869	8%	6%	0.928
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	3%	0.965	42%	5%	0.875	7%	5%	0.938

Table C.32: Cointegration results with six lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	54%	0.178	100%	6%	0.727	74%	39%	0.337
Kinetic Energy vs R_{hh}	90%	49%	0.207	100%	7%	0.692	83%	40%	0.291
R_{hh} vs R_{vv}	85%	44%	0.253	100%	7%	0.691	70%	30%	0.407
u_x vs. u_y	94%	20%	0.333	100%	16%	0.562	78%	16%	0.442
u_x vs. u_z	93%	21%	0.336	100%	16%	0.585	78%	17%	0.444
u_y vs. u_z	92%	37%	0.241	100%	16%	0.584	87%	29%	0.307
u_x vs. R_{vv}	90%	34%	0.308	100%	4%	0.634	75%	28%	0.412
u_y vs. R_{hh}	89%	36%	0.256	100%	7%	0.631	86%	31%	0.294
u_z vs. R_{hh}	89%	38%	0.257	100%	6%	0.644	86%	33%	0.291
ρ vs. $\partial\rho/\partial x$	15%	1%	0.952	27%	6%	0.892	18%	3%	0.895
ρ vs. $\partial\rho/\partial y$	17%	1%	0.950	26%	7%	0.881	20%	3%	0.885
ρ vs. $\partial\rho/\partial z$	15%	0%	0.951	24%	6%	0.896	19%	3%	0.887
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.943	41%	4%	0.870	11%	8%	0.900
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.952	45%	5%	0.869	8%	6%	0.929
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	2%	0.965	42%	5%	0.868	7%	4%	0.938

Table C.33: Cointegration results with seven lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	57%	0.172	100%	7%	0.700	74%	41%	0.329
Kinetic Energy vs R_{hh}	90%	51%	0.202	100%	7%	0.691	83%	42%	0.286
R_{hh} vs R_{vv}	85%	46%	0.248	100%	7%	0.690	70%	33%	0.401
u_x vs. u_y	94%	19%	0.339	100%	14%	0.573	78%	15%	0.445
u_x vs. u_z	93%	19%	0.338	100%	14%	0.588	78%	17%	0.442
u_y vs. u_z	92%	35%	0.242	100%	12%	0.605	87%	28%	0.308
u_x vs. R_{vv}	90%	32%	0.304	100%	7%	0.620	75%	28%	0.409
u_y vs. R_{hh}	89%	38%	0.251	100%	7%	0.620	86%	33%	0.291
u_z vs. R_{hh}	89%	38%	0.254	100%	5%	0.639	86%	33%	0.290
ρ vs. $\partial\rho/\partial x$	15%	1%	0.952	27%	5%	0.891	18%	3%	0.894
ρ vs. $\partial\rho/\partial y$	17%	1%	0.949	26%	6%	0.881	20%	4%	0.884
ρ vs. $\partial\rho/\partial z$	15%	0%	0.950	24%	5%	0.897	19%	3%	0.886
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.943	41%	4%	0.864	11%	8%	0.899
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.952	45%	4%	0.858	8%	6%	0.929
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	2%	0.964	42%	5%	0.862	7%	5%	0.937

Table C.34: Cointegration results with eight lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	53%	0.183	100%	5%	0.730	74%	39%	0.341
Kinetic Energy vs R_{hh}	90%	48%	0.210	100%	7%	0.696	83%	38%	0.296
R_{hh} vs R_{vv}	85%	43%	0.256	100%	7%	0.694	70%	29%	0.411
u_x vs. u_y	94%	19%	0.349	100%	13%	0.582	78%	15%	0.452
u_x vs. u_z	93%	18%	0.346	100%	12%	0.600	78%	16%	0.449
u_y vs. u_z	92%	35%	0.247	100%	11%	0.612	87%	27%	0.312
u_x vs. R_{vv}	90%	30%	0.316	100%	6%	0.627	75%	27%	0.419
u_y vs. R_{hh}	89%	36%	0.262	100%	6%	0.630	86%	31%	0.302
u_z vs. R_{hh}	89%	36%	0.260	100%	4%	0.648	86%	31%	0.297
ρ vs. $\partial\rho/\partial x$	15%	1%	0.952	27%	5%	0.892	18%	3%	0.895
ρ vs. $\partial\rho/\partial y$	17%	0%	0.950	26%	6%	0.883	20%	3%	0.884
ρ vs. $\partial\rho/\partial z$	15%	0%	0.950	24%	5%	0.904	19%	3%	0.887
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.942	41%	5%	0.862	11%	8%	0.899
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.952	45%	6%	0.850	8%	6%	0.929
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	2%	0.964	42%	5%	0.864	7%	4%	0.937

Table C.35: Cointegration results with nine lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	52%	0.188	100%	7%	0.686	74%	36%	0.347
Kinetic Energy vs R_{hh}	90%	48%	0.213	100%	12%	0.618	83%	38%	0.298
R_{hh} vs R_{vv}	85%	42%	0.259	100%	13%	0.617	70%	29%	0.412
u_x vs. u_y	94%	19%	0.347	100%	11%	0.582	78%	14%	0.451
u_x vs. u_z	93%	19%	0.342	100%	9%	0.605	78%	16%	0.446
u_y vs. u_z	92%	35%	0.244	100%	11%	0.602	87%	28%	0.310
u_x vs. R_{vv}	90%	30%	0.316	100%	5%	0.640	75%	25%	0.421
u_y vs. R_{hh}	89%	35%	0.263	100%	5%	0.641	86%	30%	0.303
u_z vs. R_{hh}	89%	36%	0.262	100%	5%	0.646	86%	29%	0.298
ρ vs. $\partial\rho/\partial x$	15%	1%	0.952	27%	5%	0.889	18%	3%	0.895
ρ vs. $\partial\rho/\partial y$	17%	0%	0.949	26%	5%	0.881	20%	3%	0.884
ρ vs. $\partial\rho/\partial z$	15%	0%	0.951	24%	4%	0.900	19%	3%	0.886
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.942	41%	5%	0.856	11%	8%	0.898
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.951	45%	5%	0.845	8%	6%	0.929
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	2%	0.965	42%	5%	0.865	7%	4%	0.937

Table C.36: Cointegration results with ten lags for particle track over time using Engle-Granger test method in buoyancy driven flow

	Full Time Range			Increasing Time Range			Decaying Time Range		
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal	% non-Stationary	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	90%	50%	0.190	100%	6%	0.682	74%	33%	0.350
Kinetic Energy vs R_{hh}	90%	45%	0.218	100%	10%	0.633	83%	36%	0.304
R_{hh} vs R_{vv}	85%	41%	0.263	100%	10%	0.632	70%	27%	0.417
u_x vs. u_y	94%	18%	0.353	100%	11%	0.596	78%	14%	0.457
u_x vs. u_z	93%	17%	0.351	100%	9%	0.617	78%	15%	0.455
u_y vs. u_z	92%	34%	0.251	100%	11%	0.611	87%	26%	0.318
u_x vs. R_{vv}	90%	27%	0.326	100%	7%	0.645	75%	22%	0.427
u_y vs. R_{hh}	89%	34%	0.265	100%	5%	0.643	86%	29%	0.305
u_z vs. R_{hh}	89%	34%	0.268	100%	4%	0.647	86%	28%	0.304
ρ vs. $\partial\rho/\partial x$	15%	0%	0.952	27%	5%	0.886	18%	3%	0.895
ρ vs. $\partial\rho/\partial y$	17%	0%	0.949	26%	5%	0.883	20%	3%	0.884
ρ vs. $\partial\rho/\partial z$	15%	0%	0.951	24%	4%	0.906	19%	3%	0.886
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	6%	5%	0.942	41%	4%	0.850	11%	8%	0.899
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	5%	4%	0.951	45%	7%	0.841	8%	6%	0.929
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	4%	2%	0.965	42%	5%	0.864	7%	4%	0.938

C.1.5 Particle Track Johansen Cointegration Results

The following tables contain the full results set for the cointegration testing of the JHTDB Buoyancy Driven Turbulence data taken along a particle track over the simulation duration using the Johansen test method.

Table C.37: Cointegration results with zero lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	21%	0.089	100%	0.121	80%	0.141
Kinetic vs. R_{hh}	21%	0.088	100%	0.121	82%	0.140
R_{hh} vs. R_{vv}	23%	0.089	100%	0.121	80%	0.140
u_x vs. u_y	11%	0.097	67%	0.057	66%	0.149
u_x vs. u_z	14%	0.084	70%	0.054	41%	0.138
u_y vs. u_z	100%	0.099	100%	0.067	43%	0.132
u_x vs. R_{vv}	22%	0.555	91%	0.184	81%	0.396
u_y vs. R_{hh}	22%	0.168	100%	0.154	77%	0.194
u_z vs. R_{hh}	30%	0.150	100%	0.133	55%	0.178
ρ vs. $\partial\rho/\partial x$	100%	0.244	100%	0.082	94%	0.083
ρ vs. $\partial\rho/\partial y$	100%	0.243	49%	0.064	86%	0.092
ρ vs. $\partial\rho/\partial z$	100%	0.244	77%	0.066	82%	0.093
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	57%	0.028	100%	0.204	18%	0.023
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	56%	0.029	89%	0.199	27%	0.024
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	75%	0.026	81%	0.189	21%	0.021

Table C.38: Cointegration results with one lag for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	1%	0.005	80%	0.104	13%	0.013
Kinetic vs. R_{hh}	0%	0.003	90%	0.153	7%	0.006
R_{hh} vs. R_{vv}	0%	0.003	90%	0.154	7%	0.006
u_x vs. u_y	4%	0.022	73%	0.085	20%	0.033
u_x vs. u_z	20%	0.026	68%	0.094	18%	0.037
u_y vs. u_z	9%	0.012	88%	0.097	19%	0.023
u_x vs. R_{vv}	4%	0.004	72%	0.113	26%	0.012
u_y vs. R_{hh}	0%	0.001	100%	0.097	5%	0.006
u_z vs. R_{hh}	0%	0.003	100%	0.094	3%	0.004
ρ vs. $\partial\rho/\partial x$	100%	0.274	100%	0.120	67%	0.061
ρ vs. $\partial\rho/\partial y$	100%	0.272	69%	0.106	100%	0.059
ρ vs. $\partial\rho/\partial z$	100%	0.269	71%	0.120	79%	0.058
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	18%	0.005	74%	0.104	6%	0.003
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	6%	0.004	92%	0.121	2%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	5%	0.004	77%	0.112	1%	0.002

Table C.39: Cointegration results with two lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	37%	0.041	88%	0.217	59%	0.072
Kinetic vs. R_{hh}	32%	0.041	88%	0.219	62%	0.072
R_{hh} vs. R_{vv}	29%	0.042	89%	0.217	62%	0.072
u_x vs. u_y	50%	0.078	81%	0.142	53%	0.108
u_x vs. u_z	31%	0.082	92%	0.140	50%	0.094
u_y vs. u_z	66%	0.055	100%	0.133	66%	0.086
u_x vs. R_{vv}	52%	0.186	89%	0.160	100%	0.174
u_y vs. R_{hh}	19%	0.065	93%	0.191	74%	0.105
u_z vs. R_{hh}	37%	0.068	87%	0.197	66%	0.098
ρ vs. $\partial\rho/\partial x$	100%	0.260	94%	0.156	71%	0.050
ρ vs. $\partial\rho/\partial y$	100%	0.255	86%	0.146	73%	0.045
ρ vs. $\partial\rho/\partial z$	100%	0.255	80%	0.151	67%	0.045
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	11%	0.003	94%	0.220	0%	0.003
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	4%	0.003	100%	0.229	0%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	98%	0.216	0%	0.002

Table C.40: Cointegration results with three lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	29%	0.020	100%	0.200	35%	0.046
Kinetic vs. R_{hh}	29%	0.020	100%	0.204	38%	0.046
R_{hh} vs. R_{vv}	29%	0.020	100%	0.202	35%	0.046
u_x vs. u_y	58%	0.064	100%	0.184	57%	0.094
u_x vs. u_z	50%	0.065	91%	0.163	66%	0.086
u_y vs. u_z	31%	0.046	100%	0.169	100%	0.076
u_x vs. R_{vv}	52%	0.081	90%	0.178	100%	0.104
u_y vs. R_{hh}	39%	0.034	100%	0.205	55%	0.065
u_z vs. R_{hh}	36%	0.034	88%	0.189	100%	0.054
ρ vs. $\partial\rho/\partial x$	100%	0.265	93%	0.169	86%	0.048
ρ vs. $\partial\rho/\partial y$	100%	0.258	88%	0.160	71%	0.044
ρ vs. $\partial\rho/\partial z$	100%	0.262	73%	0.161	85%	0.044
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	16%	0.003	96%	0.254	2%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	4%	0.003	100%	0.250	0%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	98%	0.248	1%	0.002

Table C.41: Cointegration results with four lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	26%	0.017	90%	0.199	32%	0.038
Kinetic vs. R_{hh}	24%	0.017	93%	0.201	28%	0.038
R_{hh} vs. R_{vv}	23%	0.017	93%	0.200	29%	0.038
u_x vs. u_y	54%	0.051	89%	0.181	72%	0.080
u_x vs. u_z	46%	0.053	88%	0.168	57%	0.078
u_y vs. u_z	44%	0.034	100%	0.168	72%	0.061
u_x vs. R_{vv}	64%	0.066	89%	0.180	100%	0.095
u_y vs. R_{hh}	34%	0.022	91%	0.231	55%	0.046
u_z vs. R_{hh}	23%	0.020	88%	0.206	55%	0.040
ρ vs. $\partial\rho/\partial x$	100%	0.272	97%	0.182	85%	0.052
ρ vs. $\partial\rho/\partial y$	100%	0.270	100%	0.185	85%	0.051
ρ vs. $\partial\rho/\partial z$	100%	0.272	100%	0.175	87%	0.049
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	34%	0.003	91%	0.267	2%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	3%	0.003	100%	0.255	1%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	98%	0.253	0%	0.002

Table C.42: Cointegration results with five lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	29%	0.017	95%	0.206	41%	0.037
Kinetic vs. R_{hh}	28%	0.016	95%	0.204	41%	0.037
R_{hh} vs. R_{vv}	28%	0.016	96%	0.204	43%	0.037
u_x vs. u_y	53%	0.051	87%	0.185	73%	0.075
u_x vs. u_z	52%	0.051	92%	0.183	100%	0.075
u_y vs. u_z	44%	0.032	96%	0.190	58%	0.058
u_x vs. R_{vv}	63%	0.062	79%	0.176	100%	0.087
u_y vs. R_{hh}	19%	0.023	97%	0.218	71%	0.046
u_z vs. R_{hh}	30%	0.021	92%	0.194	52%	0.041
ρ vs. $\partial\rho/\partial x$	100%	0.277	93%	0.182	83%	0.053
ρ vs. $\partial\rho/\partial y$	100%	0.275	100%	0.164	79%	0.053
ρ vs. $\partial\rho/\partial z$	100%	0.274	100%	0.171	80%	0.051
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	26%	0.002	96%	0.250	1%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	2%	0.002	100%	0.248	1%	0.003
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	2%	0.002	97%	0.246	1%	0.002

Table C.43: Cointegration results with six lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	26%	0.019	100%	0.213	69%	0.045
Kinetic vs. R_{hh}	25%	0.019	100%	0.214	69%	0.045
R_{hh} vs. R_{vv}	28%	0.019	100%	0.214	68%	0.045
u_x vs. u_y	50%	0.055	84%	0.171	77%	0.081
u_x vs. u_z	54%	0.056	79%	0.165	68%	0.078
u_y vs. u_z	47%	0.036	95%	0.186	85%	0.065
u_x vs. R_{vv}	60%	0.074	100%	0.179	52%	0.094
u_y vs. R_{hh}	36%	0.026	94%	0.217	63%	0.057
u_z vs. R_{hh}	34%	0.026	91%	0.189	55%	0.047
ρ vs. $\partial\rho/\partial x$	100%	0.277	85%	0.207	89%	0.057
ρ vs. $\partial\rho/\partial y$	100%	0.276	87%	0.194	87%	0.056
ρ vs. $\partial\rho/\partial z$	100%	0.274	87%	0.197	90%	0.055
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	28%	0.002	93%	0.228	1%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	2%	0.002	100%	0.258	2%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	1%	0.002	96%	0.241	2%	0.002

Table C.44: Cointegration results with seven lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	31%	0.018	100%	0.207	71%	0.045
Kinetic vs. R_{hh}	30%	0.019	100%	0.212	71%	0.045
R_{hh} vs. R_{vv}	30%	0.019	100%	0.213	71%	0.045
u_x vs. u_y	62%	0.056	100%	0.169	59%	0.084
u_x vs. u_z	58%	0.059	83%	0.165	68%	0.080
u_y vs. u_z	39%	0.036	93%	0.180	100%	0.072
u_x vs. R_{vv}	59%	0.074	100%	0.171	56%	0.096
u_y vs. R_{hh}	34%	0.026	96%	0.206	59%	0.059
u_z vs. R_{hh}	31%	0.026	86%	0.188	49%	0.048
ρ vs. $\partial\rho/\partial x$	100%	0.273	87%	0.224	82%	0.057
ρ vs. $\partial\rho/\partial y$	100%	0.272	85%	0.215	81%	0.056
ρ vs. $\partial\rho/\partial z$	100%	0.269	88%	0.223	88%	0.056
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	28%	0.003	100%	0.202	2%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	1%	0.002	100%	0.245	1%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	1%	0.002	98%	0.229	0%	0.002

Table C.45: Cointegration results with eight lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	29%	0.019	100%	0.216	43%	0.047
Kinetic vs. R_{hh}	28%	0.019	100%	0.218	49%	0.046
R_{hh} vs. R_{vv}	29%	0.019	100%	0.217	49%	0.047
u_x vs. u_y	58%	0.059	100%	0.171	66%	0.089
u_x vs. u_z	62%	0.061	86%	0.157	70%	0.084
u_y vs. u_z	45%	0.036	100%	0.186	100%	0.071
u_x vs. R_{vv}	52%	0.077	100%	0.176	100%	0.104
u_y vs. R_{hh}	33%	0.027	90%	0.218	55%	0.063
u_z vs. R_{hh}	34%	0.025	91%	0.179	58%	0.049
ρ vs. $\partial\rho/\partial x$	100%	0.269	90%	0.223	82%	0.057
ρ vs. $\partial\rho/\partial y$	100%	0.270	84%	0.204	84%	0.059
ρ vs. $\partial\rho/\partial z$	100%	0.265	86%	0.224	81%	0.056
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	28%	0.003	100%	0.217	2%	0.003
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	1%	0.002	100%	0.230	2%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.002	96%	0.229	0%	0.002

Table C.46: Cointegration results with nine lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	27%	0.020	100%	0.207	62%	0.049
Kinetic vs. R_{hh}	27%	0.020	100%	0.210	61%	0.049
R_{hh} vs. R_{vv}	28%	0.020	100%	0.212	70%	0.049
u_x vs. u_y	61%	0.062	79%	0.170	79%	0.093
u_x vs. u_z	58%	0.062	100%	0.166	65%	0.089
u_y vs. u_z	34%	0.036	100%	0.176	100%	0.071
u_x vs. R_{vv}	55%	0.076	83%	0.185	100%	0.107
u_y vs. R_{hh}	26%	0.027	91%	0.200	63%	0.064
u_z vs. R_{hh}	41%	0.025	84%	0.189	62%	0.052
ρ vs. $\partial\rho/\partial x$	100%	0.266	89%	0.218	89%	0.058
ρ vs. $\partial\rho/\partial y$	100%	0.266	85%	0.216	83%	0.058
ρ vs. $\partial\rho/\partial z$	100%	0.260	87%	0.211	88%	0.055
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	27%	0.003	100%	0.219	2%	0.002
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	1%	0.002	89%	0.217	2%	0.002
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	0%	0.002	96%	0.227	1%	0.002

Table C.47: Cointegration results with ten lags for particle track over time using Johansen test method in buoyancy driven flow

	Full Time		Increasing Time		Decaying Time	
Physical Quantities	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal	Perc Cointegrated	Average pVal
Kinetic vs. R_{vv}	30%	0.021	100%	0.201	64%	0.053
Kinetic vs. R_{hh}	30%	0.022	100%	0.206	67%	0.053
R_{hh} vs. R_{vv}	30%	0.021	100%	0.205	65%	0.053
u_x vs. u_y	60%	0.065	78%	0.159	66%	0.099
u_x vs. u_z	58%	0.064	95%	0.173	70%	0.095
u_y vs. u_z	47%	0.038	100%	0.177	100%	0.078
u_x vs. R_{vv}	57%	0.076	86%	0.184	100%	0.105
u_y vs. R_{hh}	23%	0.029	93%	0.192	73%	0.066
u_z vs. R_{hh}	34%	0.028	93%	0.189	68%	0.058
ρ vs. $\partial\rho/\partial x$	100%	0.264	92%	0.238	82%	0.058
ρ vs. $\partial\rho/\partial y$	100%	0.264	94%	0.222	85%	0.061
ρ vs. $\partial\rho/\partial z$	100%	0.258	89%	0.237	88%	0.057
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	27%	0.003	100%	0.209	3%	0.003
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	2%	0.002	86%	0.211	1%	0.003
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	1%	0.002	97%	0.207	1%	0.003

C.1.6 Particle Track Decaying Range Start Time Variation Cointegration Impacts

As was done with the single point over time data, the start of the decaying time range was modified to 13s instead of the maximum Reynolds Stress time at 6.5s as in the previous results. All eleven lag values were tested for this data as this testing was completed prior to the single point over time data previously discussed. The trends identified in this data also provided validation that it was not necessary to test all eleven lag values for the single point over time data.

Table C.48: Cointegration results with zero, one, and two lags for modified decaying time range for particle track over time using Engle-Granger test method in buoyancy flow

		Zero Lag		One Lag		Two Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	55%	1%	0.838	54%	0.456	5%	0.679
Kinetic Energy vs R_{hh}	52%	2%	0.857	52%	0.480	5%	0.708
R_{hh} vs R_{vv}	46%	0%	0.909	45%	0.541	2%	0.766
u_x vs. u_y	59%	3%	0.850	57%	0.415	5%	0.725
u_x vs. u_z	55%	2%	0.871	52%	0.457	4%	0.758
u_y vs. u_z	57%	1%	0.878	56%	0.433	3%	0.726
u_x vs. R_{vv}	67%	7%	0.791	65%	0.339	9%	0.640
u_y vs. R_{hh}	58%	2%	0.863	57%	0.420	3%	0.711
u_z vs. R_{hh}	58%	2%	0.869	57%	0.422	3%	0.716
ρ vs. $\partial\rho/\partial x$	13%	6%	0.914	6%	0.902	6%	0.898
ρ vs. $\partial\rho/\partial y$	14%	7%	0.906	6%	0.894	6%	0.892
ρ vs. $\partial\rho/\partial z$	13%	6%	0.920	6%	0.902	6%	0.896
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	8%	0%	0.979	7%	0.927	7%	0.926
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	7%	0%	0.982	6%	0.932	7%	0.930
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	6%	0%	0.987	4%	0.948	5%	0.945

Table C.49: Cointegration results with three, four, and five lags for modified decaying time range for particle track over time using Engle-Granger test method in buoyancy flow

		Zero Lag		One Lag		Two Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	55%	31%	0.533	29%	0.522	24%	0.525
Kinetic Energy vs R_{hh}	52%	29%	0.576	27%	0.566	26%	0.564
R_{hh} vs R_{vv}	46%	24%	0.633	22%	0.624	21%	0.621
u_x vs. u_y	59%	21%	0.574	21%	0.552	18%	0.562
u_x vs. u_z	55%	19%	0.614	18%	0.588	14%	0.601
u_y vs. u_z	57%	22%	0.576	22%	0.541	21%	0.549
u_x vs. R_{vv}	67%	34%	0.471	30%	0.471	28%	0.474
u_y vs. R_{hh}	58%	25%	0.546	24%	0.526	20%	0.537
u_z vs. R_{hh}	58%	25%	0.549	24%	0.532	20%	0.541
ρ vs. $\partial\rho/\partial x$	13%	5%	0.903	4%	0.909	4%	0.911
ρ vs. $\partial\rho/\partial y$	14%	5%	0.895	4%	0.901	4%	0.902
ρ vs. $\partial\rho/\partial z$	13%	4%	0.903	4%	0.911	3%	0.913
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	8%	6%	0.927	5%	0.930	5%	0.931
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	7%	6%	0.931	5%	0.934	4%	0.935
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	6%	5%	0.946	4%	0.948	3%	0.949

Table C.50: Cointegration results with six, seven, and eight lags for modified decaying time range for particle track over time using Engle-Granger test method in buoyancy flow

		Zero Lag		One Lag		Two Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	55%	21%	0.543	23%	0.538	21%	0.544
Kinetic Energy vs R_{hh}	52%	21%	0.579	22%	0.578	21%	0.585
R_{hh} vs R_{vv}	46%	16%	0.636	17%	0.635	17%	0.641
u_x vs. u_y	59%	14%	0.585	14%	0.592	14%	0.599
u_x vs. u_z	55%	12%	0.622	11%	0.624	11%	0.630
u_y vs. u_z	57%	17%	0.578	14%	0.584	14%	0.587
u_x vs. R_{vv}	67%	24%	0.494	25%	0.494	22%	0.503
u_y vs. R_{hh}	58%	16%	0.563	17%	0.563	14%	0.569
u_z vs. R_{hh}	58%	17%	0.564	18%	0.560	16%	0.562
ρ vs. $\partial\rho/\partial x$	13%	4%	0.912	4%	0.912	4%	0.914
ρ vs. $\partial\rho/\partial y$	14%	4%	0.905	4%	0.905	4%	0.907
ρ vs. $\partial\rho/\partial z$	13%	3%	0.914	3%	0.912	3%	0.913
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	8%	5%	0.931	5%	0.930	5%	0.930
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	7%	4%	0.936	5%	0.935	5%	0.935
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	6%	3%	0.949	4%	0.948	4%	0.948

Table C.51: Cointegration results with nine and ten lags for modified decaying time range for particle track over time using Engle-Granger test method in buoyancy flow

		Zero Lag		One Lag	
Physical Quantities	% non-Stationary	% Cointegrated	Average pVal	% Cointegrated	Average pVal
Kinetic Energy vs R_{vv}	55%	21%	0.549	19%	0.551
Kinetic Energy vs R_{hh}	52%	20%	0.588	19%	0.594
R_{hh} vs R_{vv}	46%	16%	0.643	15%	0.648
u_x vs. u_y	59%	13%	0.600	12%	0.603
u_x vs. u_z	55%	11%	0.631	10%	0.635
u_y vs. u_z	57%	13%	0.587	12%	0.595
u_x vs. R_{vv}	67%	21%	0.507	19%	0.512
u_y vs. R_{hh}	58%	14%	0.571	13%	0.576
u_z vs. R_{hh}	58%	17%	0.567	16%	0.569
ρ vs. $\partial\rho/\partial x$	13%	3%	0.915	3%	0.919
ρ vs. $\partial\rho/\partial y$	14%	4%	0.908	3%	0.908
ρ vs. $\partial\rho/\partial z$	13%	3%	0.913	3%	0.915
$\partial\rho/\partial x$ vs. $\partial p/\partial x$	8%	5%	0.930	5%	0.931
$\partial\rho/\partial y$ vs. $\partial p/\partial y$	7%	5%	0.935	5%	0.934
$\partial\rho/\partial z$ vs. $\partial p/\partial z$	6%	4%	0.948	4%	0.948