Statistical Analysis of Flow Parameters For The Graphical Simulation Outputs

Abstract

System dynamics simulation software, in general, depicts graphical interpretations. The values of the parameters, on the other hand, are required for prediction. The goal of this research is to develop a novel multivariate model that can predict flow parameters while simulating flow under various scenarios. The project involves looking for variations in the streamline and constructing a new multivariate model for each elliptic cylinder system's velocity magnitude. Furthermore, the flow zones were split into three groups based on streamline behavior. As a result, utilizing simulation outputs, new models for flow zones are developed using linear and semiparametric regression. The best fitted model for each flow region was determined using mean square error (MSE), root of mean square error (RMSE), and mean absolute percentage error (MAPE). Based on the fitted smoothing curve of the velocity magnitude, a summary statistic and variability may be assessed. The presented models can be used to predict magnitude in any point of fluid flow using these models.
Keywords: Multivariate model, Semiparametric regression, Streamlines, Linear Regression

1. Introduction

The field of computational fluid dynamics (CFD) makes a significant contribution to understanding the physical events that take place in the flow of fluids around and within designated objects. In addition, these events are connected to the action and interaction of phenomena such as convection, diffusion, boundary layers, and turbulence. In the field of aerodynamics, all of these phenomena are governed by the compressible Naiver-Stokes equation. Applying the fundamental laws of mechanics to a fluid gives the governing equations for a fluid. The conservation of mass equation and the conservation of momentum equation, along with the conservation energy equation, come from a set of coupled, non-linear partial differential equations. It is not possible to solve these equations analytically for most engineering problems. However, it is possible to obtain approximate computer-based solutions to the governing equations for a variety of engineering problems. This is the subject matter of computational fluid dynamics.

Previous studies provide sound knowledge of the flow behaviors of flow past elliptic cylinders and semi-parametric regression could be gained. The flow past cylinders has engineering applications in cooling towers, heat exchangers, chimney stacks, nuclear reactors, and offshore structures. Researchers are now forced to study flow past an elliptic cylinder. Elliptic cylinders are more general geometrical configurations than the canonical circular cylinder and present a richer flow behavior than the typical engineering flow configuration. In 1994, D'Alessio and Dennis proposed a mathematical model for the steady, 2-dimensional flow of a viscous incompressible fluid past a cylinder [4].

Moreover, Mittal and Balachandra have presented the results of two-and three-dimensional simulations for a range of flow and geometric parameters [2]. Furthermore, the results were compared to available experimental data, and it was discovered that important quantities such as Strouhal numbers and drag coefficients correspond well to established values.

De Silva et al. compared the numerical results with theory when one cylinder is present and the parameter relation with Reynolds number; moreover, flow behavior and the parameter relation are given when two cylinders are present [5]. Most scholars mainly use numerical and computational methods to investigate and study the flow behaviors and parameter relations when elliptic cylinders are present.

Fiosina and Maksims have presented distributed parallel versions of some nonparametric and semiparametric regression models [7]. They used the Map Reduce paradigm and described the algorithms in terms of SPARK data structures to parallelize the calculations. The forecasting accuracy of the proposed algorithms was compared with that of the linear regression model, which was the only forecasting model currently having parallel distributed realization within the SPARK framework to address big data problems. The advantages of the parallelization of the algorithm were also provided.

Ludwig et al. aimed to investigate such effects in spline-based semiparametric regression for spatial data [9]. They have discussed estimators' behavior under the traditional spatial linear regression, how the estimates change in spatial confounding-like situations, and how selecting a proper tuning parameter for the spline could help reduce bias.

Li et al. have proposed a semiparametric model based screening algorithm [6]. The model quantifies organism specific infection risks in individual subjects and accounts for the within subject interdependence of the infection outcomes of different organisms and the serial correlations among the repeated assessments of the same organism. Model parameters were estimated by using a penalized likelihood method. For inference, they have developed a likelihood based resampling procedure to compare the bivariate effect surfaces across outcomes. Simulation studies were conducted to evaluate the model fitting performance. A screening algorithm was developed using data collected from an epidemiological study of young women at increased risk of STIs.

Tadesse et al. performed a mathematical investigation of heat and mass transport in a Casson fluid boundary layer flow across an inclined stretching cylinder containing magnetic nanoparticles [12]. It was discovered that by raising the non-Newtonian Casson parameter, the flow velocity can be slowed while the temperature and concentration profiles are improved. Sensitivity analysis with the Monte-Carlo algorithm requires several function evaluations, which are infeasible with high-fidelity computational simulations [13]. Shahane et al. have trained multilayer perceptron neural networks (MLPNN) using fluid dynamics data to estimate lift and drag coefficients efficiently.

In this study, software was used to simulate the elliptic system. The flow simulation performs calculations based on the Navier-Stokes equations to simulate the interaction of fluids with surfaces. This software gives graphical interpretations; however, the values of parameters are needed to predict. The main purpose of this study is to simulate the flow under different conditions and propose a new multivariate model to predict the flow parameters. Consequently, linear regression and semiparametric regression are used to develop new models for flow regions using simulation outputs. Using these models, the approximate value of the velocity magnitude at any point of fluid flow can be calculated. Only the graphical interpretations are presented in much simulation research, but here the new models are developed for each region using statistical methods.

2. Materials and Methods

2.1 Numerical Analysis

The development of the 2-D models is described for the fluid velocity distribution for two cases. The first system developed when an elliptic cylinder was presented, while the two elliptic cylinder system was the second one. Elliptic cylinders are kept at a fixed distance from the inlet.

Fig. 1: Geometry of an elliptic cylinder system

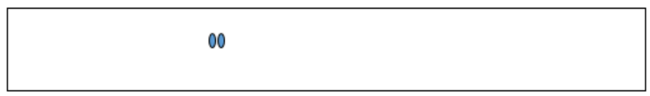


Fig. 2: Geometry of two elliptic cylinders' system

The fluid was Newtonian with constant density and viscosity resulting in the Navier-Stokes equations for laminar flow.

$$\rho(u.v)u = \nabla.\left[-\rho I + \mu\left(\nabla u + (\nabla u)^T\right)\right] + F \qquad (1)$$

Where,

ρ is the density $\left(\mathrm{kgm}^{-3}\right)$

u is the fluid velocity $\left(\mathrm{ms}^{-1}\right)$

v is a vector

μ is the fluid dynamic viscosity $\left(\mathrm{kgm}^{-1}s^{-1}\right)$

F is the external forces $\left(\mathrm{kgm}^{-2}s^{-2}\right)$

I is the identity matrix

For laminar flow, the boundary conditions were zero velocity (no-slip) on the outer sphere and constant angular velocity on the inner sphere. The wall function condition was applied. If we want to give a function, we must define units. The laminar flow interface is found under the single-phase flow branch. We can use this interface to compute the velocity and pressure field for the single-phase fluid in the laminar flow. The laminar interface can be used for stationary and time-dependent analysis.

For turbulent flow, the usual extra terms involving turbulent viscosity and turbulent kinetic energy appear.

.

Navier-Stokes equation for turbulent flow

$$\rho\frac{\partial u}{\partial t} + \rho(u.\nabla)u = \nabla.\left[-\rho I + (\mu + \mu_T)\left(\nabla u + (\nabla u)^T\right) - \frac{2}{3}\rho k I\right] + F \qquad (2)$$

Where,

ρ is the density$\left(\mathrm{kgm}^{-3}\right)$
u is the fluid velocity$\left(\mathrm{ms}^{-1}\right)$
t is the time(s)
μ is the fluid dynamic viscosity$\left(\mathrm{kgm}^{-1}s^{-1}\right)$
μ_{T} is the turbulent viscosity$\left(\mathrm{kgm}^{-1}s^{-1}\right)$
k is the turbulent kinetic energy $\left(\mathrm{kgm}^{2}s^{-2}\right)$
F is the external forces$\left(\mathrm{kgm}^{-2}s^{-2}\right)$
I is the identity matrix

The turbulent flow, *k-ε* interface found under the single-phase flow> turbulent flow branch. Turbulence effects are modeled using the standard two equation *k-ε* model with reliability constants. Flow close to walls is modeled using wall function. This interface can also be used for stationary and time-dependent analyses. We cannot give two boundary condition on a wall. If we give two boundary conditions for a wall, one of them will be overlapped. Other equations that are solved under the turbulent flow interface.

$$\rho\frac{\partial k}{\partial t}+\rho(u.\nabla)k=\nabla.\left[\left(\mu+\frac{\mu_T}{\sigma_k}\right)\nabla k\right]+p_k-\rho\varepsilon \tag{3}$$

$$\rho\frac{\partial \varepsilon}{\partial t}+\rho(u.\nabla)\varepsilon=\nabla.\left[\left(\mu+\frac{\mu_T}{\sigma_\varepsilon}\right)\nabla\varepsilon\right]+C_{\varepsilon 1}\frac{\varepsilon}{k}p_k-C_{\varepsilon 2}\rho\frac{\varepsilon^2}{k}=ep$$

(4)

In this equation p_k is

$$p_k=\mu_T\left[\nabla u\left(\nabla u+(\nabla u)^T\right)\right] \tag{5}$$

The turbulent viscosity was given by

$$\mu_T=\rho C_\mu\frac{k^2}{\varepsilon} \tag{6}$$

Where $C_{\varepsilon 1}, C_{\varepsilon 2}$ and C_μ are turbulent model constants, ε is turbulent dissipation and $\sigma_\varepsilon, \sigma_k$ are turbulent Prandtl number for *ε* and *k* respectively.

The systems described by the Navier Stokes equation, which is a partial difference equation (PDE) with analytic approximations by HAM, which is not based on discrete points, do not have exact form solutions .Instead, an approximation of the equations is based on different types of discretization. These discretization methods approximate the PDEs with numerical model equations, and the model equations can also be solved using numerical methods. Furthermore, the finite element method (FEM) is used to compute this approximation.

2.2 Statistical Analysis

It is difficult to find a general model for the whole region because elliptic cylinders have a serious effect on the velocity of the flow. Hence, the flow region should be divided into three parts, and models fitted for each region. Semiparametric regression [10] is a technique that provides the user with some flexibility in modeling complex data without maintaining stringent assumptions. With this regression, the aim is to construct a properly specified model that integrates the simplicity of parametric estimation with the flexibility provided by nonparametric splines. They are often used in situations where the fully nonparametric model may not perform well or when the researcher needs to use a parametric model but the functional form according to a subset of the regressors or the density of the errors is not known. In semiparametric regression, a smooth function,

$$f(x) = \sum_{j=0}^{m-1} \beta_j x^j + \sum_{k=1}^{K} b_k \left| x - K_k \right|^{2m-1}, m = 1,2,3,... \quad (7)$$

Where,

β_j is the j^{th} regression coefficent

K_k is the location of the k^{th} knot

With

$$b = [b_1, ..., b_k]^T \sim N\left(0, \sigma_b^2 \Omega^{-1/2} \left(\Omega^{-1/2}\right)^T\right), \quad \Omega \equiv \left[\left|K_k - K_{k'}\right|^{2m-1}\right], 1 \le k.k' \le K \quad (8)$$

is estimated using penalized spline smoothing. Penalized spline smoothers come in a number of forms. In a linear spline, weights are put on each of the splines to penalize over fitting of the data while still allowing the splines to fit the data well. The weights are determined based upon some penalization criteria. There are several options for the penalization criteria, but the easiest to implement is to choose a C such that:

$$\sum_{i=1}^{k} b_i^2 < C \quad (9)$$

Since these criteria reduces the overall effect of individual piecewise functions and avoids over fitting the data, it is an excellent minimization criterion. This criteria is more formally stated as minimizing the equation $\left| y - X\beta \right|^2$ subject to $\beta^T D\beta \le C$, where

$$D = \begin{bmatrix} 0_{2\times2} & 0_{2\times K} \\ 0_{K\times2} & I_{K\times K} \end{bmatrix} \quad (10)$$

Using Lagrange multipliers, this is equivalent to minimizing $\left| y - X\beta \right|^2 + \lambda^2 \beta^T D\beta$ for some $\lambda \ge 0$ with respect to β.

3. Results and Discussion

The flow pattern in the meridian plane is shown in following figure. In this case, an elliptic cylinder is fixed and the pressure difference is kept between inlet and outlet.

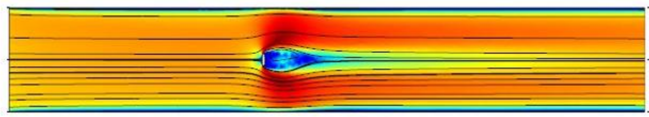


Fig. 3: The flow behavior of the system that an elliptic cylinder is presented

As we can see above figure, there is a serious effect on the velocity of the flow around elliptic cylinder. Thus we cannot find a general model for the velocity magnitude. Then the flow region should be divided into three parts as follows

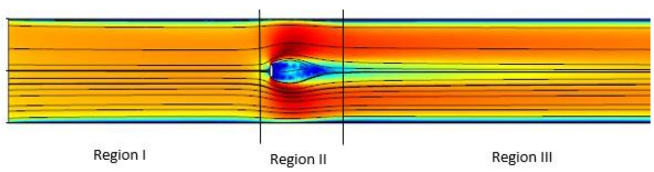
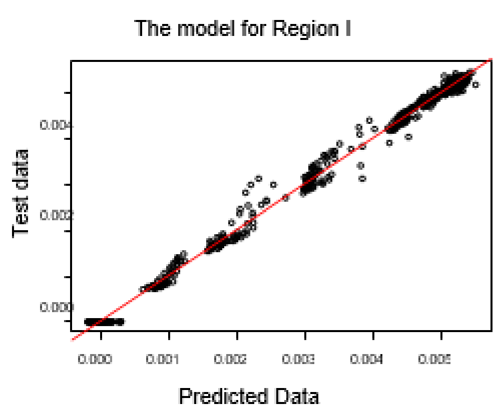


Fig.4: Three flow regions of the system that an elliptic cylinder is presented

According to the change in velocity, the flow regions are categorized. In the second region, there are two stagnation points in front of the elliptic cylinder. These eddies are created by centrifugal effects generated due to the pressure difference. The velocity magnitude of the flow slightly decreases as the distance from the inlet increases in the third region.

According to the statistical analysis, the best model has been chosen and the cross validation has been done for each region. After that, the strong relationship between the test data and predicted data with pressure and velocity cannot be seen in the whole system, but the position of the flow, shear rate, and vorticity make a significant contribution to the velocity magnitude.



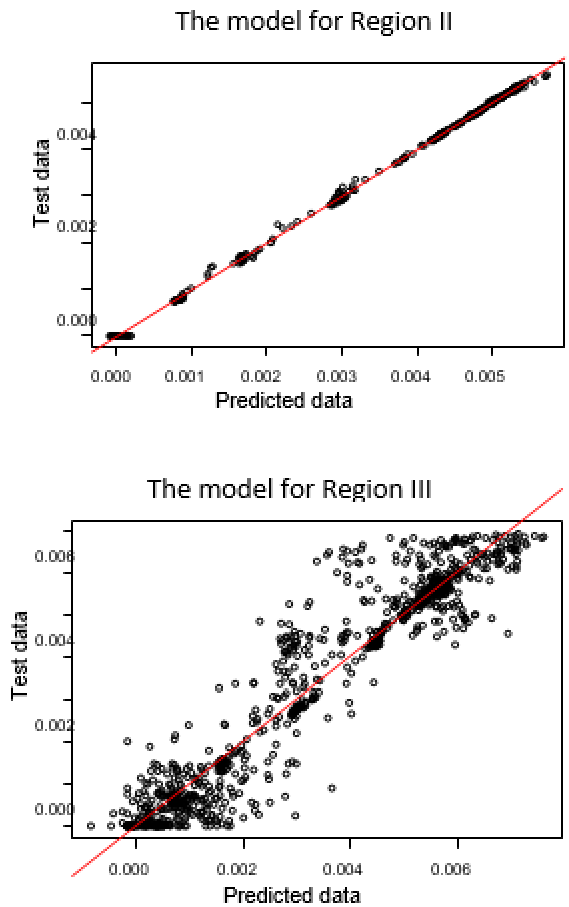


Fig. 5: The cross validations for the semiparametric regression models of first system

Further our problem can be extended to the system that two elliptic cylinders are presented. In here also there is serious changes of the velocity magnitude. It can be seen behaviors of the following streamline. So three regions had to be considered.

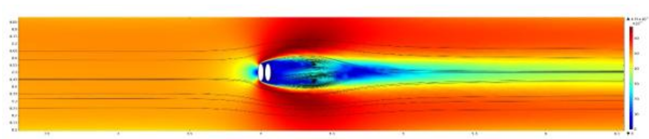


Fig. 6: The flow behavior of the system that two elliptic cylinders are presented

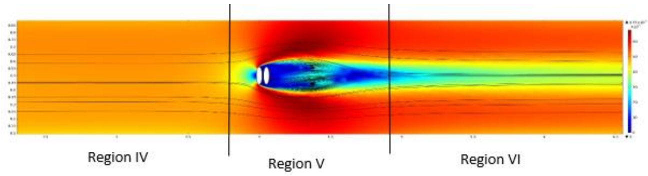


Fig. 7: Three flow regions of the system that two elliptic cylinders are presented

Using statistical analysis, both linear regression and semiparametric regression are applied to the data of all regions. Then the following table could be used to find the best fitted models for each region.

Table I: Summary of the model selection

Region	Linear Regression			Semiparametric Regression		
	MSE	RMSE	MAPE	MSE	RMSE	MAPE
I	3.091299e-07	0.0005559945	0.309768	1.209647e-08	0.000109984	0.1179557
II	4.541549e-06	0.002131091	0.5910871	5.472022e-07	0.0007397312	0.3137074
III	4.414634e-07	0.0006644271	0.2602194	2.348009e-08	0.0001532321	0.1325621
IV	4.147061e-07	0.0006439767	0.3161197	1.208502e-08	0.0001099319	0.1184157
V	1.794775e-07	0.0004236478	0.2184671	1.149482e-08	0.0001072139	0.0979895
VI	4.563871e-06	0.002136322	0.7086085	5.028746e-07	0.0007091365	0.5126755

After selecting best fitted model, the cross validation has done for each region in the second system. Furthermore, while the strong association between the test data and anticipated data with pressure and velocity cannot be seen throughout the system, the position of the flow, shear rate, and vorticity all play a role in the magnitude of velocity.

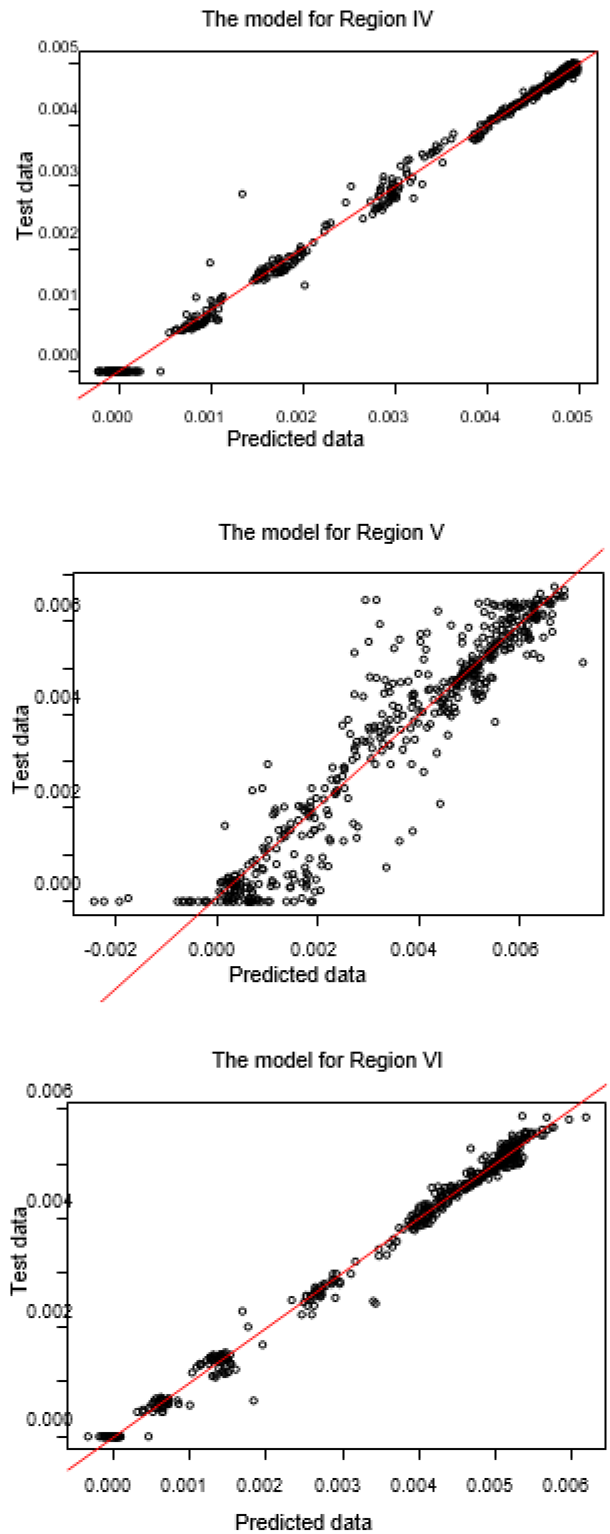


Fig. 8: The cross validations for the semiparametric regression models of the second system

4. Conclusion

Software can be used to solve the flow through elliptic cylinders system numerically. Although this software generates a graphical interpretation, statistical analysis is used to create the mathematical model. In most cases, linear regression is employed to model data; nevertheless, semiparametric regression outperformed linear regression in this investigation. According to the behavior of streamlines in the flow, the initial system in which an elliptic cylinder was presented was separated into three zones. The velocity magnitude is the most essential aspect of the flow. As a result, R was used to visualize the model for each region.

The value of mean square error, as well as RMSE and MAPE, aided in the selection of a good model for each region. Furthermore, model validation demonstrated that the chosen model is the best. The semiparametric regression delivers the optimal model for each region for the first system, according to the results. A similar procedure was used for the second system, which consists of two elliptic cylinders. For each region, a semiparametric model was created. Furthermore, the relationship between the velocity magnitude and the flow pressure is weaker than the relationship between the other parameters and the velocity. In addition, two equations for describing the velocity magnitude of both systems will be identified.

References

[1] Rahni M. T., Esfashanian V., Salary M. Investigation of Flow Around a Confined Elliptical Cylinder Using the Lattice Boltzmann Method. MiddleEast Journal of Scientific Research.

[2] Mittal R., Balachandar S. Direct Numerical Simulation of Flow Past Elliptic Cylinders., Journal of Computational Physics,1996.

[3] Perumal D. A., Kumar G. V. S., Dass A. K. Lattice Boltzmann Simulation of Viscous Flow past Elliptical Cylinder., www.cfdl.issres.net,2012.

[4] D' Alessio S. J. D., Kocabiyik S. Numerical Simulations of the Flow Induced by a Transversely Oscillating Inclined Elliptic Cylinder., Journal of fluid and structures ,2001.

[5] De Silva J. W. E. W., Dhananjaya K. D. E., Weliwita J. A. A Numerical Study on Viscous Flow Past Two Adjacent Elliptic cylinders., Proceeding of the Peradeniya University International Research sessions,2017.

[6] Zhuokai Li, Hai Liu, Wanzhu Tu. Model selection in multivariate semiparametric regression., SAGE Journals .2017.

[7] Jelena Fiosina, Maksims Fiosins. Distributed Nonparametric and Semiparametric Regression on SPARK for Big Data Forecasting. Applied Computational Intelligence and Soft Computing. Volume 2017, Article ID 5134962, 13 pages.

[8] Arnab Maity ,Tatiyana V. Apanasovich. Estimation via corrected scores in general semiparametric regression models with error-prone covariates.Electron jornal of statistics.volume 5(2011).

[9] Guilherme Ludwig, Jun Zhu, Chun-Shu Chen. On fixed effects estimation in spline-based semiparametric regression for spatial data.2015-27th annual conference on applied statistics in agriculture.Department of Kansas State University.

[10] Zhuokai Li, Hai Liu, Wanzhu Tu. A sexually transmitted infection screening algorithm based on semiparametric regression models. Statistics in Medicine. Volume 34(2015).

[11] Thomas S. Shively and Thomas W. Semiparametric Regression Approach to Adjusting for Meteorological Variables in Air Pollution Trends.

[12] Walelign T., Haile E., Kebede T. and Walelgn T., Analytical study of heat and mass transfer in MHD flow of chemically reactive and thermally radiative Casson nanofluid over an inclined stretching cylinder, Journal of physics communications.2020.

[13] Shahanea S., Kumarb P., Vankac S. P., Sensitivity Analysis of Lift and Drag Coefficients for Flow over Elliptical Cylinders of Arbitrary Aspect Ratio and Angle of Attack using Neural Network, Journal of fluid dynamics.2021.