Proceedings of the 11th International Conference on Fluid Flow, Heat and Mass Transfer (FFHMT 2024)
Chestnut Conference Centre - University of Toronto, Toronto, Canada – June 16-18, 2024
Paper No. 012
DOI: 10.11159/ffhmt24.012

Classification Tree Analysis and Manifold Alignment of Manifold Learning-Based Turbulent Flow Abundances for Flow Characterization

Nicholas V. Scott[1], Antoine Mathieu[2], and Tian-Jian Hsu[2]
[1]Apogee Engineering, LLC, Science and Technology Group
4031 Colonel Glenn Highway, Beavercreek Township, Ohio, 45431, USA
nicholas.scott@apogeeusa.com; amathieu@udel.edu; thsu@udel.edu
[2]University of Delaware, Center for Applied Coastal Research
2509 Academy Street, Newark, DE, 19716, USA

Abstract

Turbulent drag on flow structures, whether in air or water, represents a serious impediment to realizing flow efficiency for atmospheric and oceanic structures where there is a serious need to optimize energy expenditures. New research and technology have sought to go beyond mere understanding and characterization of flow structure boundary layers towards actual manipulation of them. Such state-of-the-art flow technology relies on high-resolution models which allow prediction and understanding of boundary layer spatio-temporal eddy structure. Machine learning modeling based on classification tree modeling and manifold alignment is performed as a statistical way of providing insight into flow similarity over both small and large-time scales for the time varying boundary layer eddy structure.

Flow abundance values for velocity fluctuations in the mean flow direction and particle concentration are estimated for a sinusoidally forced flow field containing medium size particles of size 280 microns. This is done use large eddy simulation data cubes which capture the boundary layer and upper free stream turbulent structure over a single sinusoidal phase at 15◦ increments. The t-distributed stochastic neighbor embedding, locality preservation projection mapping, and multidimensional scaling are used to estimate low rank embeddings for velocity and concentration depth profiles over the boundary layer in the simulation data cubes. Consecutive two-dimensional latent space embeddings or manifolds of consecutively occurring velocity and concentration data sub-cubes demonstrate topologies which can be compared via Procrustes analysis, a form of manifold alignment. Rotation, translation, and size scaling of one data cube manifold is performed with respect to the data cube manifold occurring right after it in time with a mean square-based dissimilarity value calculated for the pair.

Initial results show that the velocity abundances from all decompositions have high dissimilarity values throughout the wave cycle. The multidimensional scaling and locality preserving projection velocity abundances demonstrate small dips in dissimilarity at 0-15◦ and 180–195◦ phase transition intervals which are time periods of low sinusoidal turbulent shear stress. The dissimilarity curves for the t-distributed stochastic neighbor embedding velocity abundances are noisy and do not demonstrate strong evidence of local minimum values at this point, suggesting a lack of sensitivity to wave cycle turbulent dynamical changes. The locality preservation projection and multidimensional scaling based-concentration abundances carry high dissimilarity values throughout the sinusoidal phase cycle except at two temporal phases of 0-15◦ and 180–195◦. The low dissimilarity is thought to be due to extremely low stress occurring during the beginning of the wave cycle and flow reversal which fosters topological similarity over the small 15◦ phase time scale. The two low dissimilarity curve values occurring over the first 180◦ of the complete wave cycle suggests a phase asymmetrical turbulent response, with the second part of the complete 360◦ degree cycle being less dissimilar than the first part.

Classification tree analysis of manifold learning abundance values for concentration and velocity in the mean flow direction provide comprehension of the nonlinear relationship of latent space abundance values to 12 distinct time phase intervals equally dividing the 360◦ phase time scale. Mode classification trees show how segmented areas of manifold learning based-latent space are related to one another via the tree graph, ultimately leading to associations with specific flow forcing phase time intervals. Preliminary results suggest that different manifold learning decompositions have different tree graph structures with a tendency for the t-distributed stochastic neighbor embedding and locality preservation projection to possess a concentration-based root node, while the multidimensional scaling always produces a velocity abundance-based

root node. The second order bifurcation for the tree graph for the locality preserving projection and t-distributed stochastic neighbor embedding tends to have only velocity abundance-based nodes while the same bifurcation for multidimensional scaling has both velocity and concentration abundance nodes.

Preliminary results also suggest that the t-distributed stochastic neighbor embedding maps continual maximum values of concentration and velocity abundances toward the first 180◦ part of the 360◦ phase cycle. On the other hand, the locality preserving projection tends to map continual maximum and minimum values of concentration and velocity abundances toward the second 180◦ part of the 360◦ phase cycle. This is irrespective of the type of root node. Multidimensional scaling-based decision trees, on the other hand, tend to map continual maximum values of concentration and velocity abundances toward the second 180◦ part of the 360◦ phase cycle and continual minimum values of both abundances to the first 180◦ part of the 360◦ phase cycle. These results suggest that the locality preservation projection is not sensitive to the asymmetrical turbulent sediment-flow physics while multidimensional scaling is sensitive to such dynamics.

Keywords: *concentration abundance, classification tree analysis, manifold alignment, Procrustes analysis, t-distributed stochastic neighbor embedding, locality preservation projection, multidimensional scaling, velocity abundance*

1. Introduction

Sediment transport under sheet flow conditions due to sinusoidal temporal ambient flow forcing is an important fluid dynamical process pertinent to a wide range of coastal environmental problems including beach erosion/recovery, swash processes, and scour around structures [1,2]. The engineering implications of these processes with respect to safeguarding structures are potentially expensive with millions of dollars involved if the pertinent fluid flow physics is not handled appropriately. Addressing these complex issues hinges on understanding flow-particle dynamics whose intricacy and complexity is not captured completely by the linear theory of sediment transport. Modern theory and research have indicated that nonlinear flow effects are important contributors to sediment transport dynamics over both short and long-time scales [3]. In particular, flow asymmetrical turbulent flow response, driven by nonlinear boundary layer dynamics, is an important dynamical issue responsible for macroscopic sediment transport changes and can be captured by numerical models [4]. The resolution of this process by modern computational models such as large eddy simulations (LESs) allow for practical engineering solutions dealing with sediment transport.

Modern engineering science and technology suggests that the future of operational sediment transport lies in the manipulation of boundary layers to control sediment transport, preventing damage and extravagant costs when possible. Such control relies on both engineering structures capable of physically interacting with fluid-particle flow as well as practical predictive models parameterizing the gross structural behavior of statistical sediment motion over time and space. Machine learning is believed to be a strong practical facilitator towards the latter by its ability to find patterns in complex flow field output emanating from numerical simulations. The goal of this research is first order illustration of how machine learning can help in understanding time varying boundary layer eddy structure within a flow field. Manifold learning and classification tree analysis are applied to velocity and concentration data from a LES of a medium particle saturated water flow field under sinusoidal ambient flow forcing. The approach is to treat turbulence as information and use machine learning to derive structural relationships embedded in the evolution of the turbulent dynamical process which are potentially exploitable by boundary layer manipulation technology. Signal processing-based machine learning modeling is applied here for the purpose of characterizing subtle turbulent information changes in the boundary layer dynamics in the form of turbulent flow asymmetrical response. Considering wider research implications, it is also believed that the investigation and characterization process could lead to a formalism which can be used as a predictive tool in other time varying fluid dynamical problems.

The structure of this paper is as follows. First the LES data structure is briefly explained. This is followed by delineation and explication of the manifold learning and classification tree analysis methods used in pattern analysis of this data. The results of these two machine learning analytical methods are then illustrated with a focus on how both methods characterize turbulent flow asymmetrical response. Summary conclusions are finally given based on the feature parameterization of boundary layer flow over distinct phases of the forcing cycle, demonstrating the nature of the intersection of turbulent flow physics and machine learning.

2. Large Eddy Simulation Data Structure

A turbulence-resolving Eulerian two-fluid model was applied to oscillatory sheet flow involving medium sized sand under sinusoidal ambient free stream flow forcing [4]. The model and its results were used to study fluid particle interactions to resolve issues of long-standing interest in fluid flow-based sediment transport including particle settling, flow instability change, and enhanced boundary layer thickness [4]. In particular, the main objective of the model study was an investigation into the processes responsible for the observed differences in medium and fine sand dynamics in oscillatory sheet flow. The LES model consisted of modeling the Navier Stokes equations for a two fluid flow where the numerical domain with 200 × 260 × 92 elements in x, y and z directions respectively. Here x and z are the coordinates in the mean flow and cross mean flow direction respectively and y is the vertical coordinate in the direction of gravity. The boundary conditions consist of a symmetrical boundary condition applied at the top boundary, a smooth-wall boundary condition applied at the bottom boundary, and cyclic boundary conditions applied for the lateral boundaries. The mesh has a non-uniform grid size distribution along the y axis. Medium size particles were modeled with a diameter d_{50} = 280 μm with density $\rho = 2650$ kg/m^3. The sinusoidal wave flow forcing wave period was T = 5 s where the maximum free-stream velocity Ufm = 1.5 m/s. For this wave condition, the Stokes-layer thickness is $\delta = 1.26 \times 10^{-3}$ m and the maximum excursion length is L= 1.19 m, giving Reynolds number based on these quantities of $Re_\delta = 1890$ and $Re = 1.8 \times 10^6$ respectively. Further details about the numerics can be found in Mathieu et al. [4]

The LES model reproduces some well-known and distinctive aspects of particle saturated turbulence dynamics for oscillatory boundary layers. The change in the concentration profile across the wave period follows the well-documented description proposed by O'Donoghue and Wright [5] with a clockwise (anticlockwise) rotation of the concentration profile during flow acceleration (deceleration) around a 'pivot' of constant concentration. From the analysis of the two-fluid model results, this can be explained by a competition between downward settling flux and upward turbulent Reynolds flux over the wave period. The gross sediment and turbulent dynamics possess a known asymmetry around the point of change from acceleration to deceleration of the flow.

Data cubes were extracted from the LES data with dimensions consisting of 50 x 50 units in the horizontal plane perpendicular to the direction of gravity and 200 units in the vertical direction. The 200-unit vertical scale captures the boundary layer flow extending from just above the particle bed to the free stream flow region. Data cubes were acquired at 15◦ increments along the full 360◦ cyclic phase. The work here demonstrates how machine learning can be sensitive to turbulent asymmetrical dynamical characteristics by demonstrating latent space structure which is sensitive to and consistent with evolving phase changes in the flow.

3. Manifold Learning Processing Methods and Classification Tree Regression Analysis

Three manifold learning decomposition algorithms were used to decompose turbulent velocity fluctuations in the mean flow direction and concentration fluctuations into latent-space topological values which are dubbed abundances. These are the t-distributed stochastic neighbor embedding (t-sne), the multi-dimensional scaling (mds), and the locality preserving projection (lpp). Classical mds is a dimensional reduction technique where the objective is to find a data pattern in a lower dimensional space such that data points close together in the higher dimensional space are also close after dimensionality is reduced. Proximity of data points is measured using a dissimilarity matrix D which is a n x n asymmetric matrix which measures dissimilarity between data point observations [6]. In this work data points, each representing different dimensions, correspond to horizontal slices of the LES field at different depths. Classical mds is a metric-based decomposition where the dissimilarity is the Euclidean distance between points. A value k is needed for the calculation of the k largest eigenvalues and eigenvectors. In this work, the value of k = 2 is used to visualize the projections of the velocity and concentration data in 2 dimensions.

The lpp method is a dimensionality reduction method that is a data driven, local preservation mapping that exhumes weak covariance structure in high dimensional data. It is a form of nonlinear manifold learning which unfolds complex higher dimensional manifold structure in a lower number of dimensions for ease of visualization, preserving nearness (distance) of similar data points. This is done through the process of embedding or projection of the original manifold into a lower dimension space while adhering to the constraint of keeping dissimilar features segregated [7]. As with the mds, the lower dimension of 2 is used in the dimensionality reduction calculation.

The application of the lpp algorithm to velocity and concentration data consists of three steps where each data dimension or depth in the LES field is considered a data element. In the first step, the adjacency graph for all dimensions of the multi-

dimensional data set is calculated. The algorithm connects two data elements with an edge if they are locally close to each other, where closeness is measured via a function-based threshold. Data elements below the threshold are connected via edges where edge weights are applied via the use of a Gaussian heat kernel which has the form:

$$w_{bc} = e^{-\frac{\left\|x_b - x_c\right\|^2}{t}} \tag{1}$$

The input parameter T controls the data element width scale and x_b and x_c designates two selected data element dimensions [7, 8]. The similarity graph threshold T was set to the variance of a data element functioning as a local window. It is used in conjunction with an eigenmodal method to accentuate anomalous spatial structures in the velocity and concentration data in 2 dimensions.

In the third step, the eigenvectors and eigenvalues for the generalized eigenvector problem are computed which has the form:

$$XLX^TA = \lambda XMX^TA\,. \tag{2}$$

The graph Laplacian matrix, M = L-W is a diagonal matrix where $m_{bb} = \sum_c w_{bc}$. The quantity X^T is the data element matrix where the bth column is the bth data element x_b. Column vectors of the matrix A are the solutions of the equation ordered according to the eigenvectors $\lambda_0 < \lambda_1 < \ldots < \lambda_N$ [8]. A mapped data element y_i (i=1 ,2 , …N) is calculated using data element x_i as:

$$y_i = A^T x_i\,. \tag{3}$$

Here A^T is a matrix of row eigenvectors (a_0, a_1, a_2,…, a_N) for each of the N dimensions. The eigenvectors are called eigenfaces allowing for projection of column data vectors x_i and which uncovers low variance data structure. The sensitivity to low variance makes the lpp useful for finding and segregating anomalies in velocity and concentration data in 2 dimensions.

The t-sne is an unsupervised, local, but nonlinear dimensionality reduction technique for embedding high-dimensional data for visualization in a two-dimensional space. The two-dimensional t-sne starts by calculating a pairwise similarity between all data elements in the high-dimensional space using a Gaussian kernel. The points that are far apart have a lower probability of being picked than the elements close together [9]. The algorithm tries to map higher dimensional data elements or depth dependent LES information onto a lower dimensional space while preserving the pairwise similarities. This is achieved by minimizing the divergence between the probability distributions of the original high-dimensional and lower-dimensional spaces. The optimization used allows for the creation of clusters and sub-clusters of similar data elements in the lower-dimensional space where visualization allows understanding of structure and relationships in the original higher-dimensional data.

Procrustes analysis is a statistical shape analysis algorithm where two manifold learning projections are compared allowing for assessment of the mean square difference between them. Comparison is performed by optimally translating, rotating, and scaling one manifold learning projection with respect to another using singular value decomposition [10,11,12]. Singular value decomposition finds the series of matrix transformations that optimally match one manifold projection with another. Each manifold learning projection is compared to the next one in the temporal sequence to gain insight into the changes in the concentration and the along mean flow direction velocity turbulent dynamics in the boundary layer.

Classification tree analysis (CTA) was used to estimate the nonlinear rule relating the 24 phase intervals with their calculated manifold learning projections or abundances for velocity and concentration. The classification tree model uses predictor variables, in this case velocity and concentration abundance values, to build a decision tree providing an output response variable in the form of time phase values [6,13]. The ‘leaves’ of the decision tree represent class labels of time phase interval and the ‘branches’ velocity and concentration abundance features which lead to the time phase class labels. The CTA model algorithm is based on the premise of partitioning the abundance space into increasing smaller subgroups where the relationships between the predictor and response is more lucid. Splitting of abundance feature information into subsets and the creation of nodes in the decision tree is performed using the Gini impurity index which is a measure of data mixing. Gini impurity measures the probability of misclassification of a random instance from a subset labeled according to the majority class. Lower Gini impurity means more purity of the subset [14]. All potential bifurcations at every node are evaluated where minimization of Gini impurity for velocity/concentration abundance subgroups leads to optimal tree

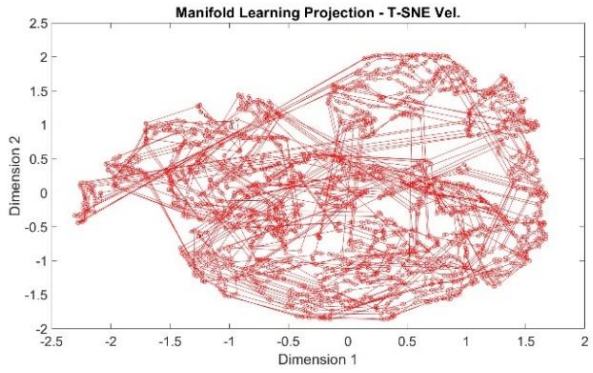
nodal splits. The stopping criteria of maximal tree depth is used to halt the tree growing process producing time-based phase class labels.

The Matlab fitctree algorithm [6] was used to estimate the classification tree using the time phase, and velocity and concentration abundance training data. The parameters used in the calculation include the use of principal component analysis to find the best split in the predictor variables. The maximum number of decision splits was set to 20 and the predictor variable is always split if the abundance variable has at most 5 levels. Pruning was allowed and cross validation was used with 10 folds. It is important to note that the nonlinear classification rule provided by the CTA algorithm is not unique. There are many factors that account for non-uniqueness. Of significance is the algorithmic features where multiple splits in the tree growing process can give the same minimization of impurity, allowing the algorithm to choose arbitrarily. However, though alternative trees can differ in terms of the splitting rules, they tend to achieve comparable levels of accuracy. It is because of this that the statistical functional rules, representing the statistical modes of the predictor-response variable training data set provided by the fitctree algorithm, are the focus of the time phase-abundance relationship analysis in this work. It is believed that physical insight into each decomposition's characteristic tendency is afforded via examination of the mode nonlinear functional mapping of abundance features to time phase label.

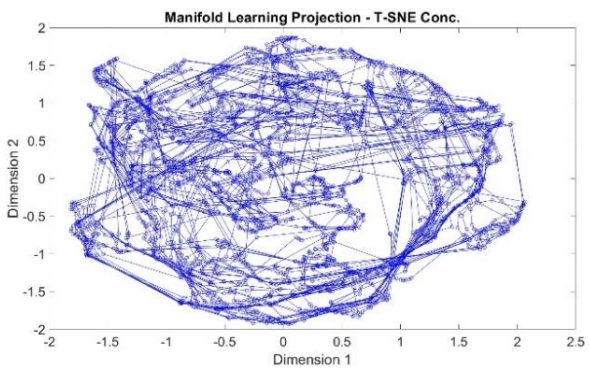
4. Manifold Learning Analysis Results

Manifold learning projections for t-sne, mds, and lpp in two-dimensional space are shown in Figures 1a-f. The projections shown for the mds and lpp are performed using the first 2 eigenvectors extracted from the decomposition. The manifold learning projections are shown for sixth phase transition which is the point of maximum forward forcing of velocity or maximum positive shear. The cohesive structure of the mds concentration manifold is noted and is most likely due to the inertia and interstitial forces along with the ensuing fluid flow viscosity caused by particles which cause a cohesive pattern in the lower dimensional projection of the higher dimensional flow field. Each temporal phase point in the captured forcing cycle admits a manifold learning projection whose similitude can be measured using the Procrustes analysis. It is the measurement of the relative differences in the topological structure of the manifold learning projections of the concentration and velocity abundance fields which provides insight into the turbulent flow response asymmetry.

a)



b)



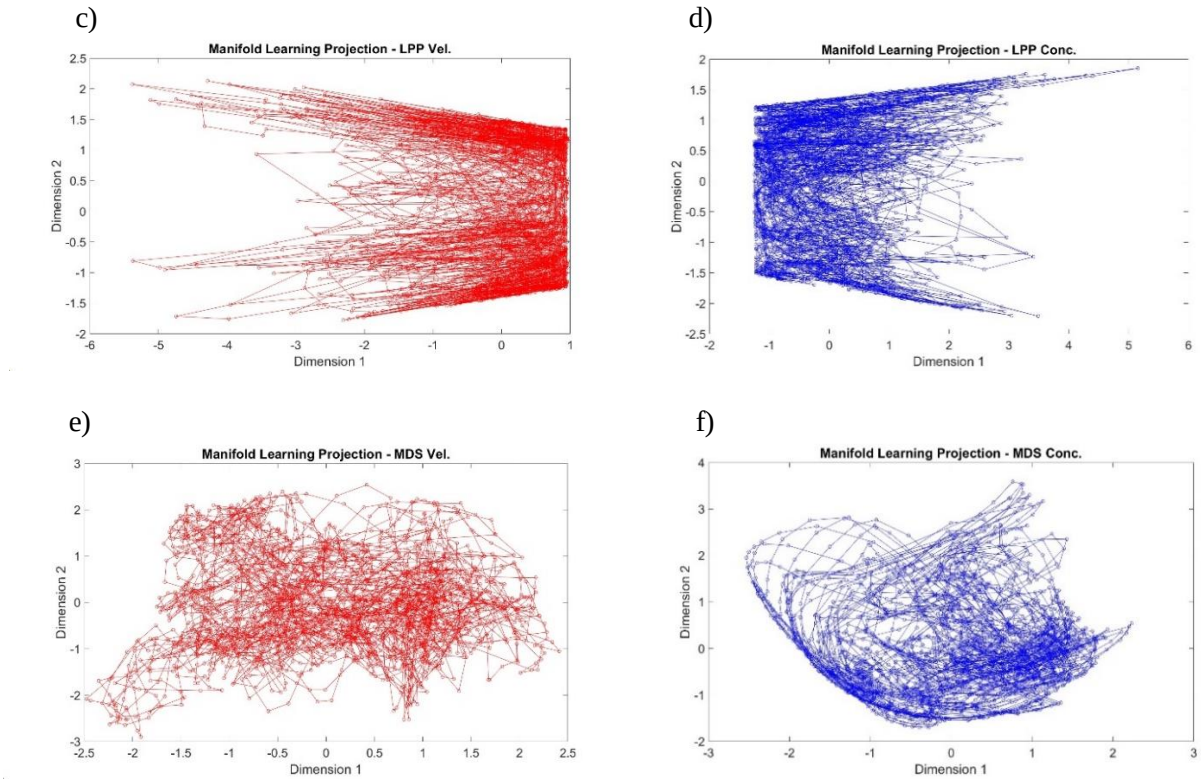


Figure 1: Manifold learning decomposition projections or abundances using the first two eigenvectors/dimensions. a) t-sne velocity and b) concentration abundances, c) lpp velocity and d) concentration abundances, and e) mds velocity and f) concentration abundances.

The LES model reproduces other distinctive aspects of particle saturated turbulence dynamics for oscillatory boundary layers [4] besides the structure stated above. For medium sand, the boundary layer remains turbulent throughout the wave period where strong bed shear generated turbulence occurs at the early stage of the wave period (0◦ to 20◦). Sudden intensification of turbulent kinetic energy (tke) is observed at the 20◦ phase due to the creation of two-dimensional instabilities. The tke reaches a maximum before the flow peak and decays eventually during flow deceleration resulting in the particle deposition without flow re-laminarization, with the tke remaining between 0.04 and 0.08 m^2/s^2. Figures 2a-f shows Procrustes analysis-based dissimilarity curves for velocity and concentration abundances. The pervasiveness of turbulence throughout the wave cycle accounts for the high level of dissimilarity in the Procrustes dissimilarity curves for the t-sne, lpp, and mds velocity abundances. This is consistent with the high tke values demonstrated in the LES model.

For the medium sand condition, the Richardson number is below the threshold value of 0.25 in the sheet flow layer during the latest stage of flow acceleration and the early stage of flow reversal (60◦–120◦). For medium sand particles, the LES model shows that the effect of density stratification is weak where the flow remains turbulent in the boundary layer. This again supports the trend of consistently high values observed in the dissimilarity curves for the velocity abundances. It is noteworthy however that slight dips in dissimilarity occur in the velocity abundance dissimilarity curves at phase interval 1 and 12 for all manifold learning decompositions. The dips correspond to the points of maximum acceleration and deceleration associated with the flow start-up process and reversal. Strong local minima at the temporal phase transition point associated with flow reversal (phase interval 12) are shown in the Procrustes dissimilarity plots for lpp and mds-based concentration abundances. It is strongly believed that the sharp drop in dissimilarity or heightened similarity in the manifold learning projections during this transition is due to the relaxation of ambient shear forcing. Low values of bed shear decrease the amount of mixing in the water column, producing a similitude of manifold structure over this small time interval.

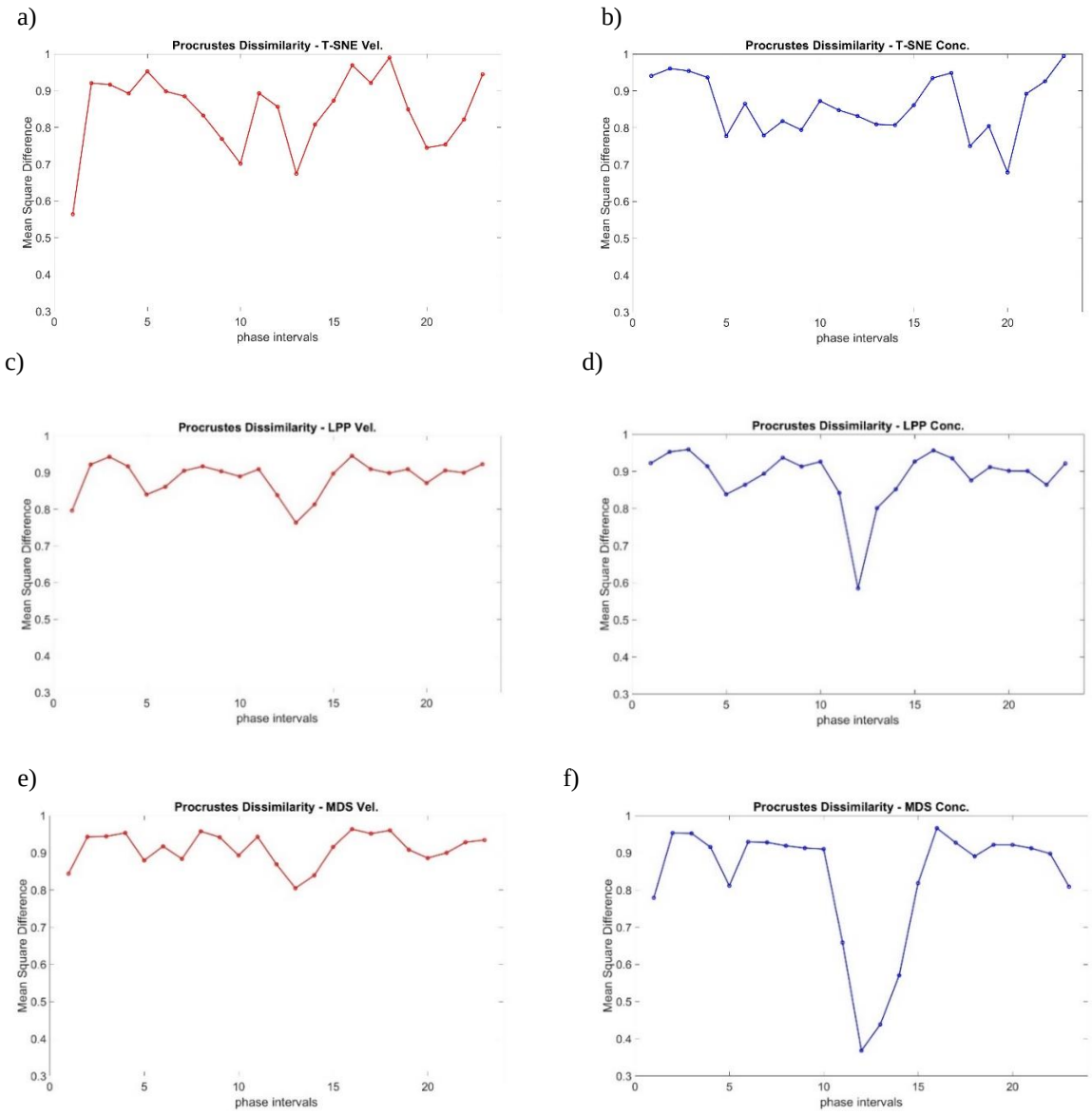


Figure 2: Procrustes analysis dissimilarity curves calculated from manifold learning projections for each of the 23 temporal phase transition intervals. Dissimilarity curves for t-sne-based a) velocity abundance and b) concentration abundance. Dissimilarity curves for lpp-based c) velocity abundance and d) concentration abundance. Dissimilarity curves for mds-based e) velocity abundance and f) concentration abundance.

5. Classification Tree Analysis Results

The statistical mode t-sne analysis-based classification decision tree is shown in Figure 3. It is noted again that the nonlinear rule relating velocity and concentration abundances with temporal phase intervals is not unique. Every iteration of the Matlab classification tree algorithm can provide a different classification rule. However, the variance in the rule generation process is not large with the statistical mode being robust. The mode t-sne-based decision tree possesses a root node that is always a concentration abundance while the second level bifurcation nodes contain only velocity abundance. Characteristic pathways exist in the decision tree that are robust and offer insight into the asymmetrical boundary layer response associated with sediment transport under sinusoidal shear forcing. If the root node is taken as the beginning point and the 'greater than' operation is taken at each node (which translates as movement along an edge to the right), then the leaf

temporal phase class label of the classification tree lies in the first 180◦ part of the 360◦ forcing cycle. The first 180◦ part of the 360◦ forcing cycle is also the destination class region if the ‘less than’ operation (which translates as movement along an edge to the left) is taken at each node starting from the root node. This result suggests that the t-sne decomposition of velocity and concentration abundances does not capture turbulent flow asymmetrical response if the continual extreme decision of ‘greater than’ is used in the decision tree.

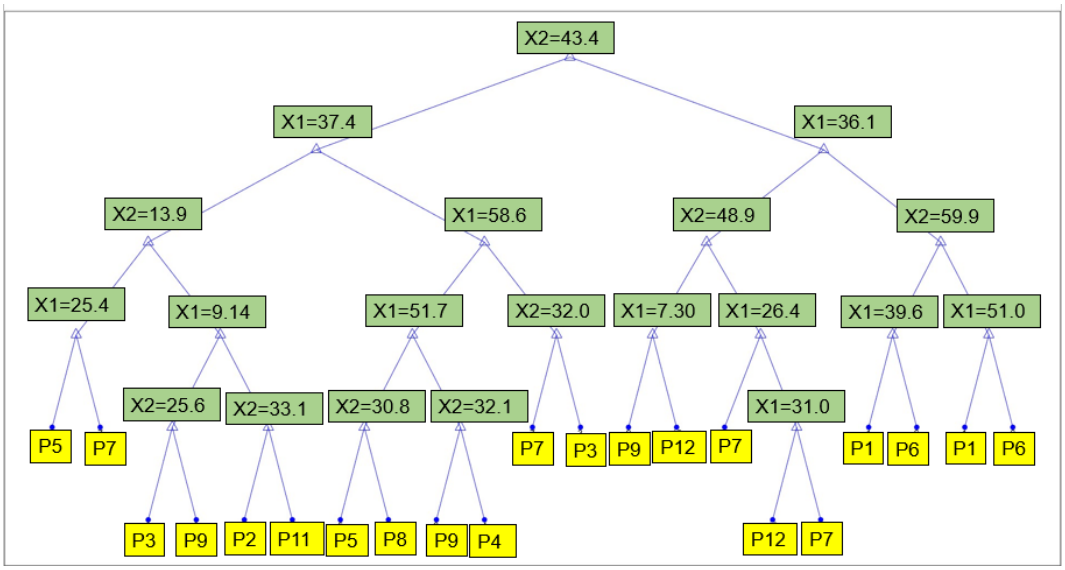


Figure 3: CTA-based decision tree relating velocity (X1) and concentration (X2) abundances to temporal phase intervals of the sinusoidal wave cycle. Mode nonlinear function decision tree rule for the t-sne decomposition is shown. Sinusoidal wave cycle broken up into 12 phase intervals of 30◦ where the final decision tree temporal phase interval mapping label is shown in yellow at the bottom. Nodal tree bifurcations denoted by bold numerals where edge movement to right is associated with velocity/concentration abundance values greater than the nodal value. Edge movement to left is associated with velocity/concentration abundance values which are less than the nodal value.

The mode CTA-based decision tree for the mds analysis is shown in Figure 4. The mode mds-based decision tree possesses a root node that is always a velocity abundance while the second level bifurcation nodes can contain both velocity and concentration abundances. The characteristic pathway based on continually taking the ‘greater than’ operation starting from the root node in the decision tree provides a destination which lies in the second 180◦ part of the 360◦ forcing cycle. Continually taking the ‘less than’ operation leads to a destination which lies in in the first 180◦ part of the 360◦ forcing cycle. This result suggests that the mds decomposition of velocity and concentration abundances along with the CTA does capture the turbulent flow asymmetrical response via the continual extreme decision operations of ‘greater than’ and ‘less than’ in the decision tree. It is stressed that the turbulent boundary layer response is not symmetric with respect to the midpoint of the sinusoidal cycle with particle distributions differing at different phase parts of the forcing cycle. This is a physical characteristic of the sediment transport dynamics.

The mode CTA-based decision tree for the lpp analysis is shown in Figure 5. The mode lpp-based decision tree can possess a root node that is either velocity or concentration abundance based. When the root node is concentration based, as shown in Figure 5, the second level bifurcation nodes contain only velocity abundances. The characteristic pathway based on continually taking the ‘greater than’ operation starting from the root node provides a destination class region which lies in the second 180◦ part of the 360◦ forcing cycle. Continually taken the ‘less than’ operation leads to a destination class region which can lie in either part of the 360◦ forcing cycle. (Figure 5 shows the result of a temporal phase class label in the second 180◦ part of the 360◦ forcing cycle). This result suggests that the lpp decomposition of velocity and concentration abundances is not sensitive to turbulent flow asymmetrical response via the continual use of the extreme decision operation of ‘less than’ in the decision tree.

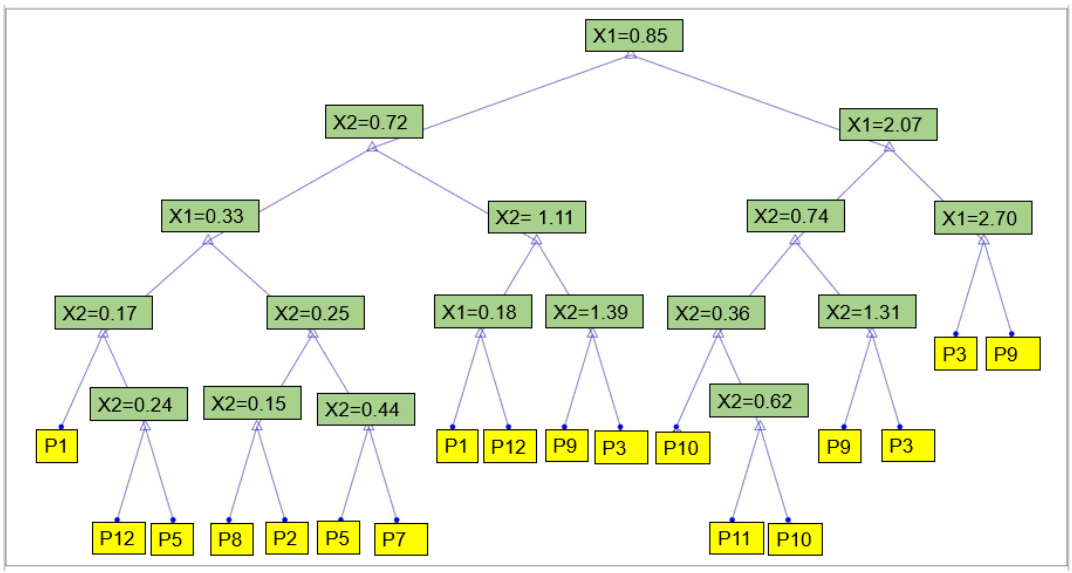


Figure 4: CTA-based decision tree relating velocity (X1) and concentration (X2) abundances to temporal phase intervals of the sinusoidal wave cycle. Mode nonlinear function decision tree rule for the mds decomposition is shown. Sinusoidal wave cycle broken up into 12 phase intervals of 30◦ where the final decision tree temporal phase interval mapping label is shown in yellow at the bottom. Nodal tree bifurcations denoted by bold numerals where edge movement to right is associated with velocity/concentration abundance values greater than the nodal value. Edge movement to left is associated with velocity/concentration abundance values which are less than the nodal value.

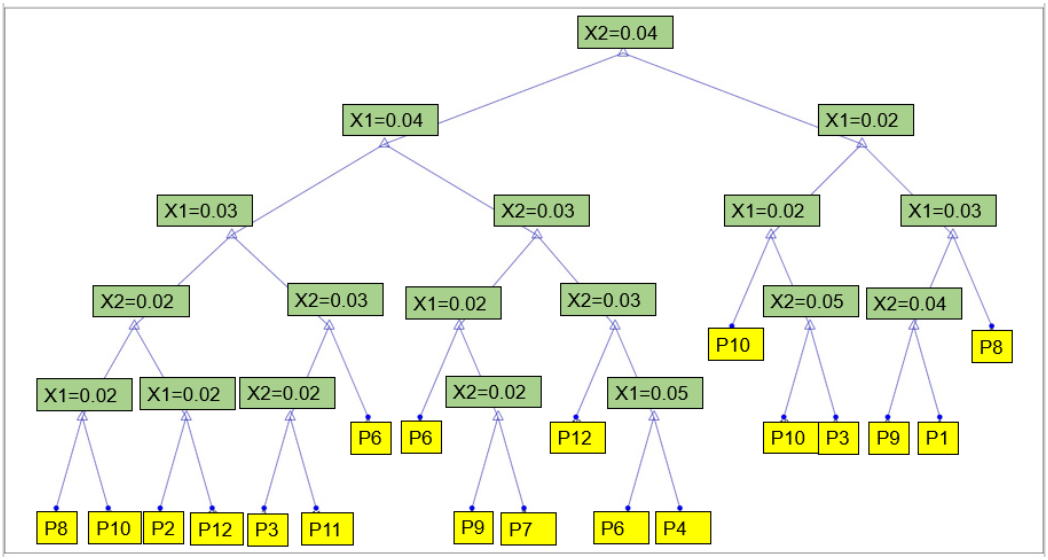


Figure 5: CTA-based decision tree relating velocity (X1) and concentration (X2) abundances to temporal phase intervals of the sinusoidal wave cycle. Mode nonlinear function decision tree rule for the lpp decomposition is shown. Sinusoidal wave cycle broken up into 12 phase intervals of 30◦ where the final decision tree temporal phase interval mapping label is shown in yellow at the bottom. Nodal tree bifurcations denoted by bold numerals where edge movement to right is associated with velocity/concentration abundance values greater than the nodal value. Edge movement to left is associated with velocity/concentration abundance values which are less than the nodal value.

6. Conclusions

Manifold alignment and classification tree analysis are used as small-time scale and large-time scale analytical methods elucidating how changes in small-scale boundary layer physics are connected to large-scale ambient temporal changes

outside the boundary layer. Manifold alignment of concentration abundances is consistent with known turbulent particle physics suggesting the utility of manifold learning decompositions in understanding particle system dynamics under time varying forcing. More importantly, changes in topology captured by manifold alignment associated with multidimensional data can provide insight into system dynamical changes suggesting temporal points where flow modulation can occur. Classification tree analysis of manifold learning abundance values for concentration and velocity in the mean flow direction captured over a single sinusoidal forcing cycle is a way of imbuing a sense of conscious mind into a turbulent flow via calculation of a decision tree graph. (The tree graph allows the potential for decision making manifested in terms of characteristic nodal-edge pathways). Preliminary results demonstrate non unique decision tree graphs for all decompositions used but with mode tree structures providing general rule-based tendencies. Moreover, results suggest that the mds decomposition along with the classification tree analysis algorithm are sensitive to the asymmetrical particle saturated flow physics. The resulting nonlinear rules obtained from estimation of tree graphs coupled to fluid mechanical mechanisms for boundary layer modulation may provide the first step towards machine learning-based optimal manipulation of turbulent flow.

References

[1] J. S. Ribberink, and A. A. Al-Salem, "Sheet flow and suspension of sand in oscillatory boundary layers," *Coastal Engineering*, vol. 25, pp. 205-225, 1995.

[2] T. Lanckriet, J. Puleo, G. Masselink, I. Turner, D. Conley, C. Blenkinsopp, and P. Russell, "Comprehensive field study of swash-zone processes. II: Sheet flow sediment concentrations during quasi-steady backwash," Journal of Waterway, Port, Coastal, and Ocean Engineering, 140(1), doi:10.1061/(ASCE)WW.1943-5460.0000209, pp. 29–42, 2014.

[3] C. R Sherwood, A. Van Dongeren, J. Doyle, C. A. Hegermiller, T.-J. Hsu, T. S Kalra, M. Olabarrieta, A. M. Penko, Y. Rafati, D. Roelvink, M. Van der Lugt, J. Veeramony, and J. C. Warner, "Modeling morphodynamics of coastal response to extreme events–What shape are we in?," Annual Review of Marine Science, vol. 14., https://doi.org/10.1146/annurev-marine-032221-090215, 2022.

[4] A. Mathieu, Z. Cheng, J. Chauchat, C. Bonamy, and T.-J. Hsu, "Numerical investigation of unsteady effects in oscillatory sheet flows," *Journal of Fluid Mechanics*, vol. 943, A7, doi:10.1017/jfm.2022.405, pp. 1-33, 2022.

[5] T. O'Donoghue, and S. Wright, "Concentrations in oscillatory flow for well sorted and graded sands," *Coastal Engineering*, vol. 50, 3, pp. 117-138, 2004.

[6] W. L. Martinez, and A. R. Martinez, *Computational Statistics Handbook with MATLAB, Third Edition*. Boca Raton, FL: Chapman and Hall/CRC, 2015.

[7] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York, NY: Springer, 2008.

[8] A. Vathy-Fogarassy, and J. Abonyi, *Graph-Based Clustering and Data Visualization Algorithms*. Heidelberg, Germany: Springer, doi:10.1007/978-1-4471-5158-6, 2013.

[9] L. van de Maaten, and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 1, pp. 1-48, 2008.

[10] D. G. Kendall, "A Survey of the Statistical Theory of Shape," *Statistical Science*, vol. 4, no. 2, pp. 87–99, 1989.

[11] C. Wang, and S. Mahadevan, "Manifold alignment using Procrustes analysis," *Computer Science Department Faculty Publication Series*, vol. 64, pp. 1–9, 2008.

[12] J. G. Gower, "Procrustes Analysis," eds. N. J. Smelser, P. B. Baltes, *International Encyclopedia of the Social and Behavioral Sciences*, Oxford: United Kingdom, Science Direct, Pergamon Press, pp. 12141-12143, 2001.

[12] E. Alpaydin, *Introduction to Machine Learning, 2nd ed*. Boston, MA: MIT Press, 2010.

[13] L. Breiman, J. H. Friedman., R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York, NY: Chapman and Hall (Wadsworth, Inc.), 1984.

[14] H. H. Patel, and P. Prajapati, "Study and Analysis of Decision Tree Based Classification Algorithms," *JCSE International Journal of Computing Science and Engineering*, vol. 6, no.10, pp. 74-78, 2018.