Proceedings of the 9th International Conference on Fluid Flow, Heat and Mass Transfer (FFHMT'22)
Niagara Falls, Canada – June 08-10, 2022
Paper No. 132
DOI: 10.11159/ffhmt22.132

Machine Learning Based Statistical Characterization of a Turbulence Dissipation Rate Array: A Revisitation

Nicholas V. Scott[1] and Jack McCarthy[2]
[1]Riverside Research, Open Innovation Center, Dayton Research Center
2640 Hibiscus Way, Beavercreek, Ohio, 45431, USA
nscott@riversideresearch.org; jack.mccarthy@duke.edu
[2]Duke University, Department of Statistical Science
214 Old Chemistry, Durham, NC, 27708, USA

Extended Abstract

Comprehension of energy dissipation rates in coastal waters is crucial to understanding such environmental fluid dynamical processes as coastal sediment transport, pollution dispersal, and heat and mass exchange across the air-sea interface. Traditional methods for understanding sub-surface turbulent velocity energetics have focused on the sampling of the turbulent velocity field using acoustic Doppler velocimeters (ADVs) and analyses directed towards the quantification of velocity array variability modeled as statistically independent frequency modes experiencing weak spectral energy transfer. These statistical methods while enlightening have not provided comprehensive insight into spatial ADV array structural dynamics especially in the area of probe system characterization. Understanding the structure of an ADV array as an information system is addressed and accomplished here via the analytical revisitation of three-dimensional velocity data obtained from a four-probe array deployed during the 2001-2003 Coupled Boundary Layers and Air-Sea Transfer (CBLAST) Low Program. The research objective was to show how machine learning algorithms can provide an alternative perspective for the characterization of coastal turbulent velocity information with respect to three important areas – multivariate turbulent kinetic energy level segmentation, nonlinear turbulent velocity modal analysis, and statistical modelling of probe relationships.

The Coupled Boundary Layer and Air-Sea Transfer Low program (CBLAST-LOW) was a field experiment whose goal was to improve understanding of the parameterization of the marine boundary layer and air-sea interaction processes during low winds. During the experiment four ADVs were mounted on a submerged steel beam in a linear array approximately 3.5 meters below the water surface in the Martha's Vineyard Sound. Turbulent kinetic energy (tke) dissipation rates were estimated from power spectra of the vertical velocity component and through the use of the frozen turbulence field hypothesis from data acquired from September 22-23, 2003 [1]. Each dissipation rate value was estimated from 20 minutes of data at 20 minute intervals over a time period characterized by two high and two low tides per day. Dissipation rates were elevated in general due to energy contamination by surface wave fluctuations and possessed local maxima due to the semidiurnal tidal component which inundated the sampling region.

Gaussian mixture modelling (GMM) [2] using two spatially distant probes, probes 2 and 4, showed a linear proportionality covariance structure at high dissipation rates. A second cluster mode exists at low dissipation rates where low values for probe 2 were associated with a large spread of dissipation rate values at probe 4. This is thought to be due to a turbulence wake effect where a large uniform turbulence system sat on top of all the probes at high tide, causing linearly proportional dissipation rates for the two probes. At low tide, dissipation rates at probe 2 were extremely low but a residual medium local turbulence level still existed at probe 4.

Generative topographic mapping (GTM) [3] is a non-linear latent variable model which furnishes a two-dimensional organized representation of noisy, nonlinear dissipation rate data exhibiting data clusters using latent variables constructed under the assumption of an underlying manifold data structure. Latent space is filled with a regular square array of feature nodes where the four-dimensional data space points lying on a manifold are images of the latent space under a local but nonlinear kernel function mapping. Latent space exhibits a segmentation of data points above and below a root mean

square dissipation rate threshold range of 0.0003-0.0004 suggesting a natural threshold range for distinguishing between a high tide enhanced dissipate rate regime and a low tide regime.

Competitive leaky learning (CLL) [4] was used to cluster four-dimensional ADV probe data into five clusters through the use of mean site distributions and correlation matrices which provided cluster labels. These clusters represent nonlinear modes existing within the four-dimensional data. Clusters 4, 2, and 5 are nonlinear modes characterized by mean dissipation rate values which were approximately at the low, medium, and high levels uniformly across each of the four probes. Correlation matrices for clusters 4 and 2 show independence of probes 3 and 4 and strong correlation of probes 1 and 2 respectively. Cluster 5 exhibits strong correlation between probes 1 and 2, and probes 3 and 4. Cluster 3 is a fourth nonlinear mode where probe 3 has a high mean dissipation rate but where strong correlation structure exists between probes 1 and 2. This correlation of probes 1 and 2 is very similar to correlation matrix structure exhibited in clusters 2 and 4. Cluster 1 represents the fifth nonlinear mode dominated by an exceptionally high mean dissipation rate at probe 4 along with strong independence of probe 4 and strong correlation of probes 1-3 exhibited in its correlation matrix. Cluster modes demonstrate intra probe turbulence dynamics where there were three nonlinear modes with semi-uniformly low, medium, and high mean dissipation rates, in addition to two 'nonlinear' modes having exceptionally high dissipation rates for probes 3 and 4.

With sufficient evidence of the correlation of at least two probes in the array, statistical classification was applied to the ADV array to investigate the degree to which probe 4 was statistically related to the amalgamation of probes1-3. This was performed using Bayesian and naïve Bayesian classification [4] for a dual or bi-modal class distribution where the separation threshold level was set at 0.0003. True and estimated Bayesian and naïve Bayesian classification time series for the fourth probe's dissipation rate level (modeled as the causal node) using probes 1, 2, and 3 dissipation rates (modeled as effect nodes) were estimated. This structure does not follow the true cause-effect structure of the probe system and was applied to examine whether knowledge of information at probe 4 implies covariance-based knowledge at probes 1-3. The time series consisted of bimodal class labels for regimes above and below the threshold whose error was then calculated. Bayesian classification possessed a slightly lower error than naïve Bayesian classification, suggesting that prediction of probe dissipation rate levels from probe 1, 2, and 3 is best performed using covariance, substantiating their clustering nature.

Tidally and surface wave forced, array sampled, shear flow turbulence can have dissipation rates possessing complex nonlinear modal structure which can be detected and quantified using advanced machine learning methods of GMM, CLL, and GTM analysis. In addition, covariance-based Bayesian classification of the highly variable, spatially separate dissipation rate levels provides better results than naïve Bayesian classification, substantiating the sub-cluster nature of the ADV probe data. All analyses afford a way to detect, model, and separate nonlinear modes using both probe number and energy dissipation rate structure providing an alternative view of probe relationships.

References

[1] H. Tennekes, and J. L. Lumley, *A First Course in Turbulence*. Massachusetts, USA: MIT Press, 1994.

[2] W. L. Martinez, and A. R. Martinez, *Computational Statistics Handbook with MATLAB, Second Edition*. London, UK: Chapman Hall/CRC, 2007.

[3] J. A. Lee, and M. Verleysen, *Nonlinear Dimensionality Reduction*. New York, USA: Springer, 2007.

[4] S. Theodoridis, A. Pikrakis, K. Koutroumbas, and D. Cavouras, *Introduction to Pattern Recognition, A MATLAB Approach*. Massachusetts, USA: Academic Press, 2010.