Proceedings of the 9th International Conference on Fluid Flow, Heat and Mass Transfer (FFHMT'22)
Niagara Falls, Canada – June 08-10, 2022
Paper No. 175
DOI: 10.11159/ffhmt22.175

Bayesian Belief Network Analysis of a Large Eddy Simulated Ocean Turbulence Field

Nicholas V. Scott[1] and Tobias Kulkulka[2]
[1]Riverside Research, Open Innovation Center, Dayton Research Center
2640 Hibiscus Way, Beavercreek, Ohio, 45431, USA
nscott@riversideresearch.org; tkukulka@udel.edu
[2]University of Delaware, Center for Applied Coastal Research
2509 Academy Street, Newark, DE, 19716, USA

Abstract—An observational space Bayesian belief network analytical formulism is applied to a sub-grid modeled turbulent kinetic energy (tke) field emanating from ocean turbulence large eddy simulation (LES) data containing Langmuir cells but no breaking waves. The purpose of the analysis is to illustrate how statistical machine learning modeling can be used to understand the probabilistic structure of observational space which is needed in demonstrating how the allied latent space can be related statistically to it for the purpose of data generation. The Peter and Clark (PC)-algorithm-based Bayesian belief network (BBN) edge-nodal structure for the observational-space tke subdomains demonstrates a distinctive nonlocal connectivity pattern when the multidimensional scaling graph layout is invoked. When the Chow-Liu algorithm is used, a different tree-based connectivity in the network is revealed. In particular, a dominant parental root node occupies the far upper left region in the observational-space domain with many edge connections flowing toward the right. Quantification of linkage strength through the BBNs provides a way to understand the spatial covariance of observational space subdomains providing hints as to the physically relevant statistical inferential model.

Keywords: *ocean turbulence, large eddy simulation, turbulent kinetic energy, Bayesian belief network, PC algorithm, Chow-Liu algorithm*

1. Introduction

Using oceanic observational array systems is an important method for accruing in situ information about complex upper ocean dynamics, which in turn has great implications for understanding ocean-atmospheric processes, from microscales, all the way up to synoptic scales. With the strong and incessant push towards understanding oceanic processes over vast spatio-temporal scales, there is a need for improving data processing schemes which attempt to model dynamical information assuming complex covariate structure. Bayesian belief network (BBN) analysis is one area that allows for the exposure of latent information structure allowing insight into how to model spatio-temporal dynamical complexity. This statistical signal processing technology seeks to uncover correlative trends among a set of random variables based on a joint probabilistic model linking the variables. It is applied here as an alternative way for understanding the spatial structure of the mixed layer of an oceanic turbulence field setting the stage for future spatio-temporal analyses.

The structure of this paper is as follows: first, the tke mean feature matrix is briefly summarized and a brief description of BBN analysis is given. Next, an explanation of the two BBN algorithms used in producing probabilistic graphical layouts is provided. The focus here is on understanding the rudimentary structure of the LES data obtained through the use of a BBN software package, Bayes Server Ltd. A small discussion of results is provided to illuminate statistical trends present in the LES data features.

2. Observational-Space Data Struture

A. Data Structure

The LES data analyzed in this work originates from a numerical simulation study performed by Kukulka and Brunner [1] who modeled the wave averaged Navier-Stokes equations for a rectangular ocean domain subject to a constant wind stress. The data set used from the complete LES study had a mean wind speed of 5 m/s at 10 m above the water surface. The dominant wave frequency was 0.258 Hz, and the wave age, defined as the ratio of the surface wave phase speed to the air side friction velocity, was 35. The oceanic turbulence field possessed Langmuir turbulence, but no breaking waves. Two hundred seventy LES data cubes comprising 128 points in both the x and y horizontal directions, and 300 points in the vertical direction aligned with gravity, were sequentially selected and used to obtain the surface sub-grid modeled tke field. The physical dimensions of a complete LES data cube for x, y, and z were Lx = 150 m, Ly = 150 m, and Lz = 90 m respectively where the vertical mixed layer depth was 33 m. Further details about the LES numerics can be found in Kukulka and Brunner [1].

Self-organized mapping (SOM), a dimensional reduction method which affords visualization of three-dimensional turbulence data in terms of two-dimensional maps, was used to reduce the 270 turbulence spatial samples to a series of square horizontal images. This was done using the same technique used by Scott and Handler [2]. SOM was used to compress or reduce the vertical depth dimension of the LES data cubes, from the ocean surface to approximately half the mixed layer depth, forming LES horizontal images. From these images, a mean tke feature matrix with dimensions of 16 X 270 was created where the first dimension denotes the numerical label associated with 16 equally spaced horizontal subdomain locations. The subdomains were the results of taking the average tke value at 9 square pixel values over each subdomain location. The locations were labeled in a column-wise fashion in the tke feature matrix. The second dimension of the tke feature matrix designates the time sample location from 1 to 270. The tke feature matrix information was used in the construction of BBNs. Further details about the LES numerics can be found in Scott and Kulkulka [3].

B. Bayesian Belief Network Structure

BBNs are probabilistic graphical models which use edge and nodes to model the joint probability distribution existing between a set of random variables describing a system [4]. They allow for statistical inferences to be made at random variable nodes when evidence is provided to one or more network nodes. Prior to statistical inference, network nodes along with nodal states need to be defined followed by structural learning which derives the directed acyclic graph (DAG) associated with the BBN. This step is concerned with exhuming the BBN topology from the data. Once the network is induced from tke feature information, parameter learning can be performed which provides numerical values to conditional probabilities existing between nodes [5]. The defined nodes and conditional probabilities in turn allow for statistical inference where the effects of evidence at one or more random variable nodes are propagated throughout the BBN to estimate its impact on other nodes.

BBN analysis was performed on the mean tke feature matrix using the software package Bayes Server manufactured by Bayes Server Ltd. which automates much of the statistical analysis including the Bayesian network structural learning and parameter learning. Two structural learning algorithms were implemented in this analysis. The first algorithm is the Peter and Clark (PC) algorithm which uses conditional independence to estimate a DAG associated with the LES feature data [6]. This is a local analysis which works by first assuming conditional dependence between all random variable nodes which represent the 16 locations in observational space. It then breaks linkages between nodes when conditional independence is satisfied with respect to two nodes. By moving all over the observational-space domain and examining nodal linkages, the DAG is estimated.

The second network structural algorithm used is the Chow-Liu algorithm [7]. This is a global structural learning method which searches network structures using a single root initial network structure in the beginning of the network learning process. Variations from this network structure are performed at each step causing the tree to increase in complexity. The tree structure model used is a multivariate probability distribution expressed as a product of conditional probability distributions based on a parent node in the tree. The tree structure that best approximates the real distribution is found by minimizing the difference between the real data-based distribution and the tree approximation. This in turn is done by minimizing the mutual information between any two pairs of nodal variables [5]. The best network structure in other words is selected based on a score measuring how well the model represents the data. The Chow-Liu algorithm assumes a tree structural model is appropriate to the data and seeks a skeletal structure consisting of a low number of dominant parental nodes which provide sub-dominant children nodes. The direction of links is found using higher order dependency tests [5]. Both the PC and Chow-Liu algorithms in principle should obtain the same network structure, provided there is enough information or data. In cases of small amounts such as used here, they can and do differ.

Parameter learning provides numerical values to the nodal-edge structure allowing for statistical inference between nodes. Both the PC and Chow-Liu algorithms use the same relevance tree algorithm which allows for exact statistical inference rather than approximate inference [6].

3. Results

A. Observational-Space Analysis of Subdomains: Bayesian Belief Network Structure

Fig. 1 shows the BBN emanating from the use of the PC algorithm where nodes are depicted as rectangles with bar charts in them showing the state levels of the nodal variable. It employs a multidimensional graph layout which orders the nodes and edges of the BBN such that nodes which possess the highest covariance are in close proximity. For example, note the proximity of nodal pairs L and M near the bottom of the plot, G and J on the left-hand side, and B, H, and O on the right-hand side. The BBN also shows arrows depicting the conditional probabilistic ties that the nodes share. For example, nodes G and J share a strong linkage depicted by a heavy arrow where arrow thickness is proportional to mutual information or shared covariance. This is a clear example of two nodes sharing significant local covariance. Fig. 1 also depicts extreme nonlocal correlation between nodes which are not in close proximity. Nodal pairs B and H, L and M, and H and O are examples of this. These BBN observations are corroborated via the association table for this network (not shown) which provides the mutual information (MI) or conditional probabilistic strength of individual nodal pairs. The MI values for nodal pairs B and H, L and M, and H and O are 0.305, 0.275, and 0.233 respectively. Many other local and nonlocal connections in the BBN can be identified by inspecting Fig. 1. Worthy of note in the PC algorithm-based BBN is node P which is not conditionally related to any other node.

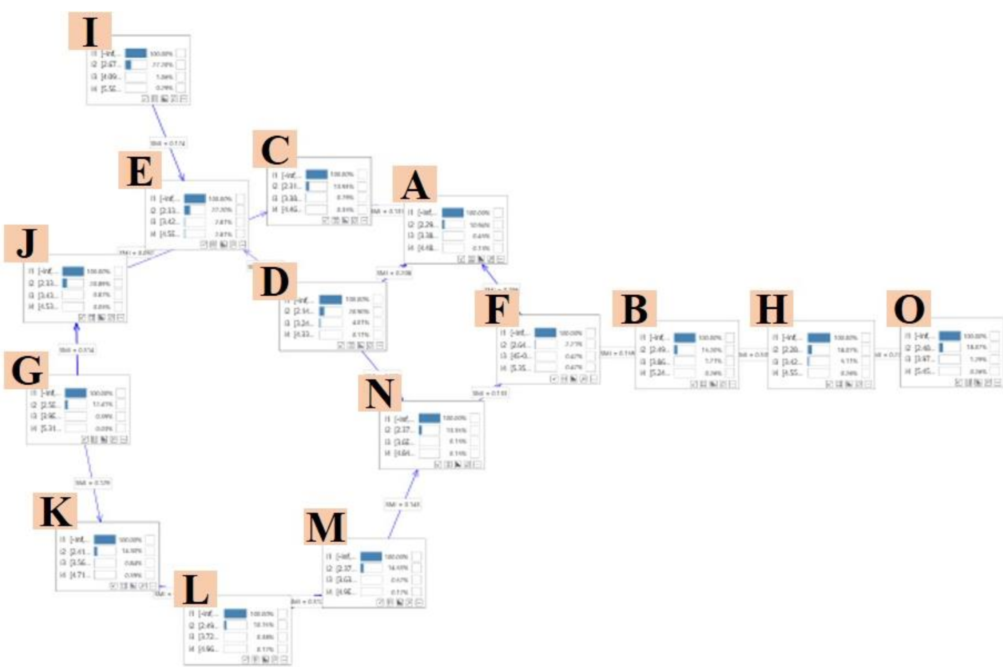

Fig. 1. BBN structure for the mean tke subdomain field based on the application of the PC algorithm. Multidimensional scaling format is shown. Horizontal plane subdomains are labeled with letters. Arrow thickness parameterize the strength of edge connections.

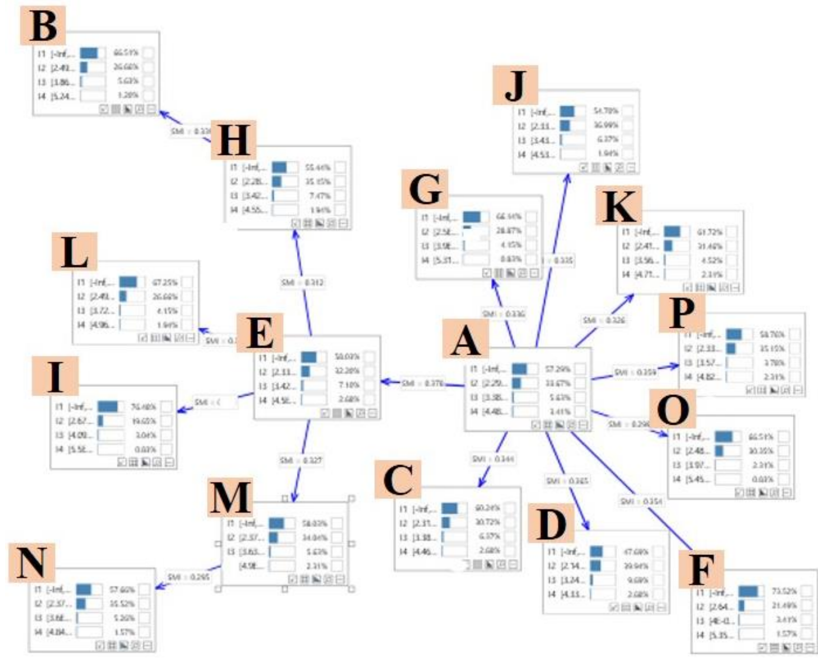

Fig. 2. BBN structure for the mean tke subdomain field based on the application of the Chow-Liu algorithm. Multidimensional scaling format is shown. Horizontal plane subdomains are labeled with letters. Arrow thickness parameterize the strength of edge connections.

Fig. 2 shows the BBN emanating from the use of the Chow-Liu algorithm which also depicts the statistical dependence of nodes. The BBN structure here is very different from the PC algorithm-based structure. Fig. 2 provides a visually clear map of how different nodal spatial subdomains in the observational space are complexly related in the large eddy simulated sea. In particular, in the midst of data paucity where only 270 data realizations are present, statistical structure based on an assumed model can still be estimated. The BBN plot shows that node A, which lies in the upper left corner of the subdomain array, is a strong parent node where many edges emanate from. Edges actually radiate from this node outward towards the right, lower right-hand corner. and lower left-hand corner. The Chow-Liu BBN is observed to have more nodal pairs which possess stronger statistical conditional probability than the PC BBN by noting the presence of more thick arrows. This is also corroborated by the association strength table for this BBN (not shown).

Nodal pairs A and D, A and E, A and P, A and C, and A and J all have high MI scores demonstrating their strong statistical connections. The Chow-Liu BBN suggests that given the 16 x 270 feature matrix, the strong local and nonlocal statistical connections that exist throughout the observational-space turbulence field is predominantly due to the upper left-hand corner of the observational domain. The use of the multidimensional graphical layout in Fig. 2 shows that node E, lying to the left of A, is also a parental node with many edge emanations to the south and the east in the original physical space. Interestingly, node A according to this model is the parent of node E which in turn (as shown in Fig. 2) influences nodes H, I, L, and M with the strongest links being between E and H, and E and M. The Chow-Liu network has the feature that two nodal subdomains can lie adjacent to each other, but possess a convoluted statistical connection pathway. For example, node B, which lies adjacent to node A in the original physical space, requires a pathway through node H and E to reach node A. Nodes which lie along the same line in the array in physical space can also be complexly related. For instance, nodes B and N, both of which occupy the same row in the subdomain array in the original physical space, are second-tier nodes both being separated from their parental node E by a single node (Fig. 4). This complexity in the BBN is due to the model attempting to statistically approximate the nonlinear and nonlocal turbulence dynamics.

No clear evidence is present suggesting a relationship between the two BBN models. Nodal relationships which are dominant in the PC algorithm-based BBN do not appear in Chow-Liu algorithm-based BBN. This is possibly and most likely due to data paucity. Given the extreme difference in the structural learning results, the question arises as to which model possibly and more closely approximates the truth. This topic is addressed in another paper where a HMM is used to connect latent and observational spaces of the turbulent feature field. The flow of high tke energy from latent-space nodes to observational-space nodes and the subsequent strength of the observational subdomain nodal linkages to other nodes is used as a metric for conceptually judging the physical pertinence of each BBN.

4. Conclusions

BBN analysis is applied to LES data features emanating from the dynamics of an upper ocean turbulence field exhibiting nonbreaking ocean surface waves and Langmuir circulation. The PC algorithm and the Chow-Liu algorithm for BBN structural learning of the mean tke feature values demonstrate different types of local and nonlocal connectivity between subdomain locations in the tke field. Nodal linkage strength is estimated in each network via calculation of the MI index with the Chow-Liu algorithm possessing more potent linkages. HMM analysis of maximum tke feature values will provide stronger, but not definitive, evidence as to which network is the more physically appropriate network for high energy flow tke physics. These latent space-observational space results will be contingent on latent space having an observational space characterized by strong covariance when inundated with high tke information, similar to observational space dynamical characteristics.

5. References

[1] T. Kukulka, and K. Brunner, “Passive buoyant tracers in the ocean surface boundary layer: 1. Influence of equilibrium wind-waves on vertical distribution,” *Journal of Geophysical Research: Oceans*, vol. 120, pp. 3837-3858: doi:10.1002/2014JC010487, 2015.

[2] N. V. Scott, and R. A. Handler, “Structural characterization of wind-sheared turbulent flow using self-organized mapping,” in Proceedings of the Society of Photo-Optical Instrumentation Engineers 9827, Ocean Sensing and Monitoring VIII, 98270X, 2016, [Online]. Available: https://doi.org/10.1117/12.2220596

[3] N. V. Scott, and T. Kukulka, “Baysian Latent-Observational Space Analysis of a Large Eddy Simulated Ocean Turbulence Field, S. Sukoriansky, and B. Galerpin”, Oceans 2021 San Diego-Porto, 2021, pp. 1-7, doi: 10.1109/IEEECONF38699.2020.9388980

[4] L. E. Sucar, *Probabilistic Graphical Models*. London, UK: Springer, 2015.

[5] A. Darwiche, Nicholson, *Modeling and Reasoning with Bayesian Networks*. New York, USA: Cambridge University Press, 2009.

[6] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*. Florida, USA: CRC Press, 2010.

[7] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufman, 1988.