Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. MVML 142
DOI: 10.11159/mvml25.142

Bayesian Statistical Priors Estimation Supporting Multidimensional Feature Anomaly Characterization

Nicholas V. Scott
Apogee Engineering, LLC, Science and Technology Group
2611 Commons Blvd., Beavercreek, OH, 45324, USA
nscott.gso@gmail.com

Abstract **-** Knowledge of remotely sensed imagery states and anomalous change plays a crucial role in enhancing remote sensing-based geo-intelligence analysis. Central to this process is the ability to quantify robust modulations in imagery state and its associated spatial-temporal geometric metadata for remote sensing sensors. Bayesian statistical priors, established during periods devoid of extensive number of anomalous structural events, provide a baseline for future Bayesian inference and anomaly assessment. A robust two-pathway methodology for estimating these statistical priors is explicated contingent on the identification of significant and nonsignificant off-diagonal correlative multidimensional conditions for remotely sensed imagery features and associated geometrical metadata variables over a set time scale. For instances possessing strong non-correlative states, Markov chain and Morlet wavelet-based statistics are employed to analyze the structure of imagery features only over time. This analysis yields insight into imagery feature transitions over different intervals as well as high energy event statistics within imagery features as a function of scale. For strong correlative states, Morlet wavelet-based statistics of imagery features as well as Laplacian eigenmap and independent component analysis of the imagery features and metadata are employed for structural exhumation. Furthermore, Hidden Markov modeling allows for estimation of emission matrix models characterizing the relationship of temporal statistical evolution of imagery feature values with the dominant metadata variable. Markovian analysis, Morlet wavelet analysis, Laplacian eigenmap, and independent component analysis combined provide a set of baseline statistical models and parameters for future anomaly assessment in newly acquired multidimensional data.

Keywords: Bayesian priors, correlation matrix, hidden Markov model, independent component analysis, Laplacian eigenmap, Markov chains, Morlet wavelet analysis

1. Introduction

Automatic target recognition (ATR) systems fundamental to mission directives supporting geointelligence initiatives are based on the need for rapid exploitation of large amounts of imagery acquired from state-of-the-art imagery acquisition technology utilized over land and sea. For instance, satellite infrared imaging systems as well as hyperspectral imagery drone systems both possess ATR technology characterized by the acquisition of imagery feature time series along with metadata which is used to determine anomalous structures worthy of further investigation by geointelligence analysts [1]. Some remote sensing imagery missions are based on predetermined scanning pathways where imagery acquisition pathways and the associated sensor Eulerian angles are used by the remote sensing imagery system to search and track interesting phenomenological structures for future analysis. Other sensor data acquisition pathways are more random in nature with payload sensors indiscriminately searching over the geometrical data acquisition space in hopes of finding anomalous structures in imagery. Given the large number of anomalous structures which can exist in both conditions and given that the same ATR system can often function in both modalities, algorithmic methodologies which can facilitate the detection and characterization process of anomalous structures in both scenarios are strongly sought after. An image processing methodology for characterization of multidimensional image feature state (imagery and metadata) evolving over time based on the appearance of anomalous statistical structures in these two distinct conditions is explicated. The methodology is only meant to establish crude Bayesian priors for anomaly structures typically sensed over a time domain, allowing for more robust detection of anomalies by more powerful machine learning algorithms in the future.

Distilled imagery features, in the form of image variance along with the geometrical configuration of the sensor system used to acquire these features, form a multidimensional system state which changes at different points in time as a sensor system captures information. This variability can be informative with respect to anomaly quantification. For instance, significant local and global structure in the multidimensional system state often suggests important underlying dynamical forces warranting more intensive examination. The approach used in the quantification of anomaly statistics from multidimensional imagery feature state information is an *ensemble methodology*. A series of algorithms are used to examine an array of imagery passes where an imagery pass is defined as vector at a single point in time consisting of imagery variance and the roll, pitch, and yaw metadata of the sensor associated with this variance. An array of image passes is first evaluated as being 'correlation strong' or 'correlation weak' with respect to the relationship of imagery variance and its geometric metadata variables. Based on the result of this data assessment, an algorithmic dependent bifurcation in the signal processing pathway takes place where two pathways can be taken. Each pathway contains a series of algorithms directed towards characterization of the global and local structures in an array of imagery passes. Global analytical algorithms consist of Markov chain and hidden Markov analysis of imagery pass data which serve to assess the probabilistic transition structure of anomalies as a Markovian statistical process. Local analysis of imagery pass data, on the other hand, consisting of Morlet wavelet, Laplacian eigenmap, and independent component analysis, serves to examine the 'granular' aspect of anomalous structure by quantifying both the temporal scale structure of anomalies in time as well as their cohesive manifold structure within an array of imagery passes.

The creation of algorithms for the examination of multidimensional system state for Bayesian prior anomaly quantification to be exploited in future assessment is the subject of this work and is explained as follows. First, the numerically simulated data sets used in the algorithmic development are briefly explained. The theory behind the advanced signal processing algorithms used in the signal processing pipeline methodology is then provided. The structure of the signal processing pipeline is explained next where the focus is on ensemble methodological analysis of multidimensional system state for local anomaly detection and assessment. Signal processing examples are then given illustrating how the ensemble methodology works in practice. Finally, preliminary conclusions are provided where suggestions on practical implementation are provided.

2. Data Structure

The signal processing pipeline was applied to two different simulated data sets. The two datasets were generated through a simulation algorithm in which multidimensional data points over a specified time frame were generated simulating a visible light remote sensing payload moving over an aerial pathway. Both datasets are m x n matrices, where row m denotes the observed multidimensional temporal sampling point of an imagery pass characterized by n=4 imagery state features: image variance (σ) as well as sensor payload roll (φ), yaw (ψ), and pitch (θ). The simulation algorithm utilized to produce the data was a simple Kalman state filter to prescribe temporal dependent values characterizing the movement of a sensor along a set trajectory used to provide semi-random changes in Eulerian angles and image variance [2]. For the first data set, a noise saturated manifold for the 4-dimensional data was created allowing for noisy geometric metadata described by φ, ψ, and θ which was correlated with σ. By imbuing even higher levels of random image variance noise not correlated with the Eulerian angle dynamics, the second data set was created. The second dataset comprised 3900 data points that reflect erratic sensor payload behavior characterized by randomly varying geometric orientations during the capture of image variance σ. The first dataset comprised 1700 data points that depict more coherent sensor behavior, characterized by semi-smooth transitions in σ with one of the metadata variables which is taken as the dominant metavariable. The simulations offer a way for analyzing both the erratic and coherent remote sensor behaviors using advanced analytical methodologies aimed at detecting anomalies.

The simulated multidimensional system data were analyzed using Markov chains, hidden Markov modeling, Laplacian eigenmaps, independent component analysis, and Morlet wavelet analysis with the objective of establishing anomaly statistics for Bayesian priors. The following sections delve into the various analytical techniques which form a complete processing pipeline for this training data characterized by low levels of anomalous structures. It is noted again that the true value of this proposed processing pipeline lies in affording prior statistical insight into anomaly structure for future analysis.

3. Data Analysis Theory

3.1 Markov chain analysis of image variance

Markov chains are probabilistic graphical models characterizing how a random process transitions through discrete variable intervals through quantification of the likelihood of moving from one state to another [3]. Markov chains possess the Markov property where the future variable interval depends solely on the current interval thereby describing a memoryless process. The imagery variance or σ time series can be modeled as a Markov chain through the estimation of the transition matrix representing the probabilities of transitioning between σ value intervals. To construct this matrix, σ data is segmented into discrete intervals from which conditional probabilities are calculated based on the frequency of transitions between them. The transition matrix can then be used to address two important statistical questions surrounding the σ Markov chain process. First, what percentage of time does the σ Markov chain spend in each interval. Second, what proportion of σ values exist within each interval? The percentage of time, T_M spent in each interval and the proportion of σ values in each interval, P_M can be estimated from the transition matrix [4]. These two statistics are strong global parameterizations of anomalous behavior.

3.2 Hidden Markov models for geometric data and image variance

A hidden Markov model (HMM) is a double stochastic process consisting of a Markov chain hidden state stochastic process, which is not directly observable, and an observation stochastic process which is observable [5]. The observation stochastic process produces a sequence of observations given the hidden state process. In addition to the Markov property, the HMM possesses the stationarity property where the state and observation statistics do not change over time as well observation independence where the observations only depend on the current Markov chain hidden state. Discrete HMMs are considered here where the parameters of the HMM are estimated using two time series X and Y which are the hidden state variable and the observation variable respectively [6]. The values in each of these variables are divided up into intervals where the transition frequency between intervals in X is counted along with the distribution of the observation variable intervals for Y when the hidden state transition is made. In this problem Y= σ and X is the dominant geometric variable most correlated with σ. With the calculation of HMM model parameters, optimal prediction of future state intervals of σ can be estimated from new values of the dominant geometric variable.

3.3 Laplacian eigenmap and independent component analysis of geometric data and image variance

Laplacian eigenmap (LE) analysis is a dimensionality reduction method consisting of a data driven, local preservation mapping that exhumes weak covariance structure in high dimensional data [7,8]. 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 multidimensional data points. This is done through the process of embedding or projection of the original manifold into a lower dimensional space while adhering to the constraint of keeping dissimilar points segregated. The application of the LE algorithm to multidimensional data here consists of a series of steps applied to data vectors with elements σ, φ, ψ, and θ. In the first step, the adjacency graph of the multi-dimensional data is calculated. The algorithm then connects two data elements with an edge when locally close to each other where closeness is measured via a function-based threshold. Data elements below the threshold are connected via edges where the edge weights are applied via the use of a Gaussian heat kernel. In the last step, the algorithm computes the eigenvectors and eigenvalues for the generalized eigenvector problem using the adjacency graph matrix [9]. The first two graph Laplacian eigenmap eigenvectors are used to visualize the multidimensional data providing a Bayesian reference.

Independent component analysis (ICA) is a blind source separation eigenvector method which seeks to unmix a high dimensional data set into a lower rank set of source signals which are statistically independent [10]. Algorithmically this is done through the maximization of kurtosis or nonGaussianity within the data [11]. As in the case of the LE, ICA requires prescription of the rank of the manifold that the high dimensional data lies near which is taken as 2 for ease of visualization.

The topological structure embodied in the eigenvectors of the simulated multidimensional data provides the reference from which a distance metric can be applied for assessment of future anomalous data points in new multidimensional data.

3.4 Morlet wavelet analysis of image variance

The continuous Morlet wavelet transform of the one-dimensional image variance signal σ(n) is a time-scale decomposition $\overline{W}_n(s)$ which gives the local power or energy of the signal at different scale magnifications, s and time locations, n [12,13]. Computation of the Morlet wavelet transform is performed by taking the convolution of σ(n) with the stretched and compressed versions of the Morlet wavelet function which has localized sinusoidal fluctuations modulated by a Gaussian envelope. Normalization rescales the Morlet wavelet transform results allowing for comparison across scale magnifications. By varying the scale s and the time index n, a Morlet wavelet image of information can be created where a slice of the Morlet wavelet transform, $\overline{W}_n(s)$ at a specific time location gives the energy contained within σ at various spatial scales [12]. At any time point n, the wavelet scalogram, $P(s)$ exists quantifying the spectral energy distribution. This scalogram can be normalized providing the 95% significance levels for all points in the wavelet time-scale space, providing a subset of points which can be used to estimate a constant optimal wavelet energy threshold [14].

The estimation of the optimal wavelet energy threshold is performed by first taking the maximum energy values over the significant Morlet wavelet energy points for each individual time point n and forming a histogram of the frequency of significant maximum energy values. This distribution has a negative exponential or Poisson distribution shape with large frequency values for low wavelet energy bins and low frequency values for high wavelet energy bins. A frequency histogram constant for the maximal number of high wavelet energy events is assumed and set to the arbitrary low value of 35. The first wavelet energy bin possessing this frequency of occurrence or greater is then found. This characteristic wavelet energy bin represents a threshold which captures low levels of anomalous local events in σ and is another Bayesian prior parameter useful in the discovery of future local anomalous events.

4. Signal Processing Methodology and Pipeline

The aforementioned algorithms represent a multicomponent filter forming a geoprocessing pipeline designed to assess the anomalous structure of multidimensional training data. This pipeline is shown in Fig. 1. The data processing pipeline methodology first consists of evaluating the data's correlative nature for σ and φ, ψ, and θ where the results determine which of two possible processing pathways are taken. The evaluation is performed via computation of the correlation matrix and examining if significantly high values exist for the off-diagonal elements. If the multidimensional data tests positive for strong correlation, the bottom processing pathway is taken in Fig. 1. Along this pathway the data is examined to understand both its global and local fluctuation structure. Hidden Markov modeling is first performed to estimate the emission matrix that couples the dominant geometric variable and σ. (The dominant geometric variable is established by finding the geometric metadata variable most correlated with σ). The emission matrix values are stored. ICA and LE analysis are applied next to the multidimensional imagery pass data. The ICA and LE eigenvectors and the embedded manifold weights for the data are estimated and stored. From the topology of these two decompositions, the optimal eigenspace distance thresholds for anomaly detection are estimated from inspection and stored. Finally, the Morlet wavelet transform is applied to the σ signal to estimate the high energy event statistics and the distribution stored. Because the training stage data is assumed to not possess large numbers of anomalies, it is assumed that the lowest wavelet energy level found using the process delineated in section 3.4 is a suitable wavelet threshold to be used in the future processing of new σ data. If the training imagery pass data tests negative for strong off-diagonal correlation, the top processing pathway is undertaken in Fig. 1. Along this pathway the data structure is examined using Markov chain analysis. From the transition matrix of σ, T_M and P_M are calculated and stored. Finally, the Morlet wavelet transform is applied to σ for estimation of the high energy event statistics and the results stored.

Each processing pathway provides algorithmic outputs in the form of model matrices, statistical distributions, parameters, and thresholds which can be amalgamated to achieve the goal of evaluating whether significant data anomalies are most likely present when new data is acquired. It is in this sense that the signal processing pipeline is Bayesian since new image feature data can be evaluated based on what has been learned from past prior training data analysis.

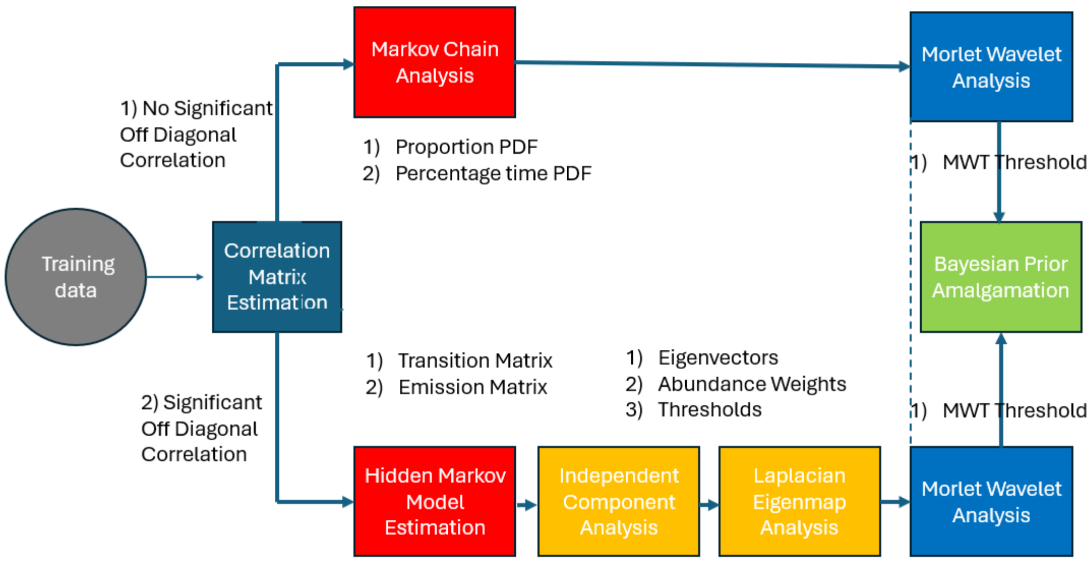

Figure 1. Training stage processing pipeline for multidimensional imagery state data.

3. Results and Discussion

The signal processing pipeline algorithms are illustrated for the two different correlation cases using the aforementioned simulated multidimensional data. The first case is the situation where the correlation matrix shows no significant off-diagonal correlation values and utilizes data set 2 comprising 3900 data points. The second case is for the situation where significant off-diagonal correlation exists in the correlation matrix and utilizes data set 1 comprising 1700 data points. The correlation matrices for the two different cases are shown in Fig. 2a)-b) where no significant off-diagonal correlation characterizes the matrix on the left and significant off-diagonal correlation characterizes the matrix on the right. A measure for the level of off-diagonal correlation for processing pathway delineation is the summation of the off-diagonal mean square values which is 0.31 and 0.82 for the correlation matrices on the left and right respectively as shown in Fig. 2a) and b). If the off-diagonal mean square value exceeds a set threshold, taken as 0.5, then significant off-diagonal correlation exists. The processing pathway results are explicated below.

a) b)

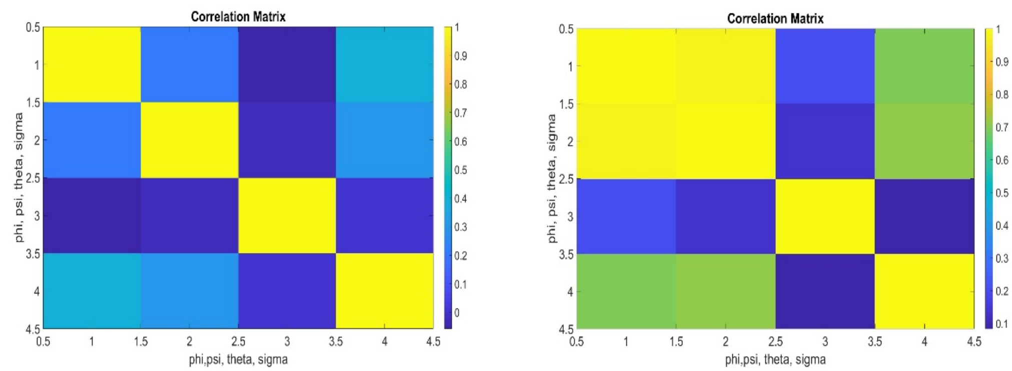

Figure 2. Correlation matrices for the cases of a) no significant off-diagonal correlation and b) significant off-diagonal correlation. Off-diagonal mean square values for a) and b) are 0.31 and 0.82 respectively where the threshold for pathway delineation is 0.5.

5.1 Case 1: No significant off-diagonal correlation

For the case of no significant off-diagonal correlation, the proportion and percentage of time pdfs, P_M and T_M respectively, are calculated for the σ signal and are shown in Fig. 3a). Both curves show that σ has its highest probability in interval 2 which corresponds to the σ transition interval [16.7 22.8] where the units are arbitrary. These statistical curves are Markov chain parameters to be used in future processing. The Morlet wavelet transformation of the complete σ training data is shown in Fig. 3b). The σ data has a very low number of high energy events denoted by white dots. A histogram of the high energy events as a function of scale is shown in Fig. 3c) and is sparse again due to the paucity of high energy events. The wavelet transform threshold of 2.46 is the estimated threshold used to produce this frequency distribution and is used in future processing for anomaly statistical characterization.

a) b) c)

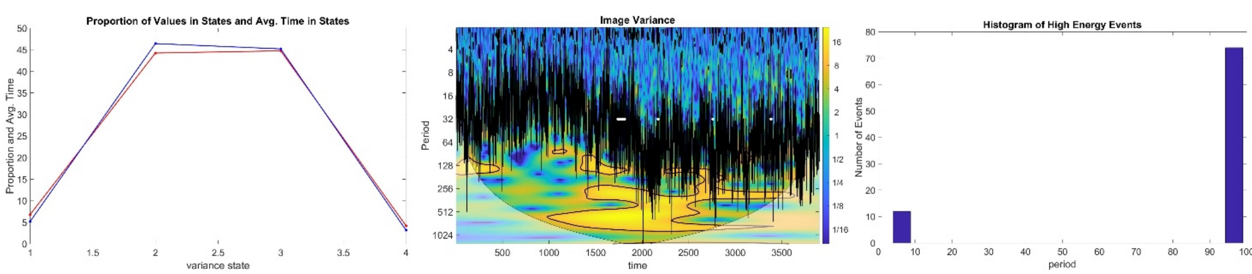

Figure 3. a) P_M and T_M pdfs (red and blue respectively) for the training stage data. X-axis delineates 4 equidistant intervals for σ which span the range of [10.6 35]. b) Morlet wavelet transform for σ with the σ time series overlaid. Contours are 95% confidence intervals for time-scale space. White dots signify data points exceeding set constant wavelet transform threshold of 2.46. c) High energy event histogram of white dots. Period time scale on x-axis and number of events on y-axis. Physical units for time and scale are not applied.

5.2 Case 2: Significant off-diagonal correlation

For the case of significant off-diagonal correlation, the emission matrix emanating from HMM estimation is calculated for σ and ψ, the geometric variable found to be most strongly correlated with σ. The emission matrix is shown in Fig. 4a) where ψ is the state variable whose intervals appear on the y-axis and σ is the observation variable whose intervals appear on the x-axis. Each row of the emission matrix is a pdf quantifying the probability of attaining σ values appearing on the horizontal axis given specific values of ψ appearing on the vertical axis. The emission matrix shows that as the ψ interval increases, there is an increasingly high probability of getting larger values at high σ intervals. This is useful information to be considered as part of the Bayesian prior pool of information. The application of ICA and LE analysis for this case produces the two-dimensional projections of the multidimensional data points or imagery pass data using the first two eigenvectors as shown in Fig. 4b) and c). It is from inspection of the topology shown that suitable ICA and LE eigenvector thresholds for anomaly classification are obtained for future anomaly characterization.

a) b) c)

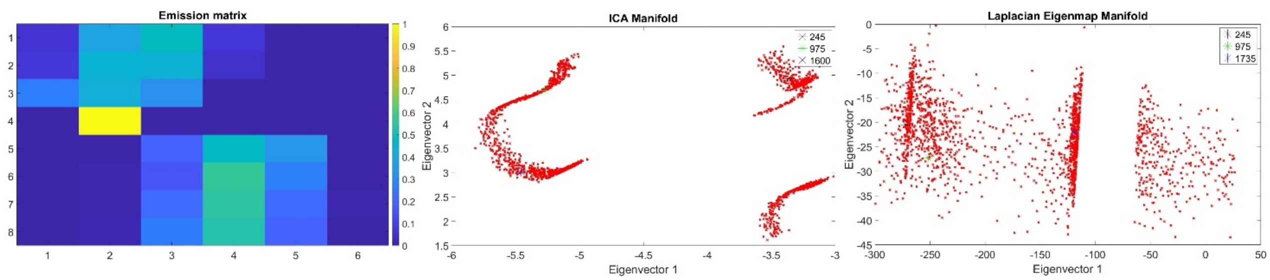

Figure 4. a) Emission matrix for σ - ψ variables (observation and state variables respectively). Observation variable intervals divided up into 6 linearly equidistant intervals from 10.6 to 35. Units are arbitrary. State variable intervals divided up into 8 linearly equidistant intervals from 16 º degrees to 68 º. Each row for the emission matrix sums to 1 making it a pdf. b) ICA of multidimensional data using first two eigenvectors. c) LE analysis of multidimensional data using first two eigenvectors. Three asterisks denote three data points.

Morlet wavelet transformation of the σ data for case 2 shown in Fig. 5a) depicts a series of high energy events. The wavelet threshold of 2.9 was used and produced a histogram of high energy events as shown in Fig. 5b). Note that this case has a higher number of high energy events than the previous case providing a different reference level and provides expected values for future processing of anomalous events for this scenario.

a) b)

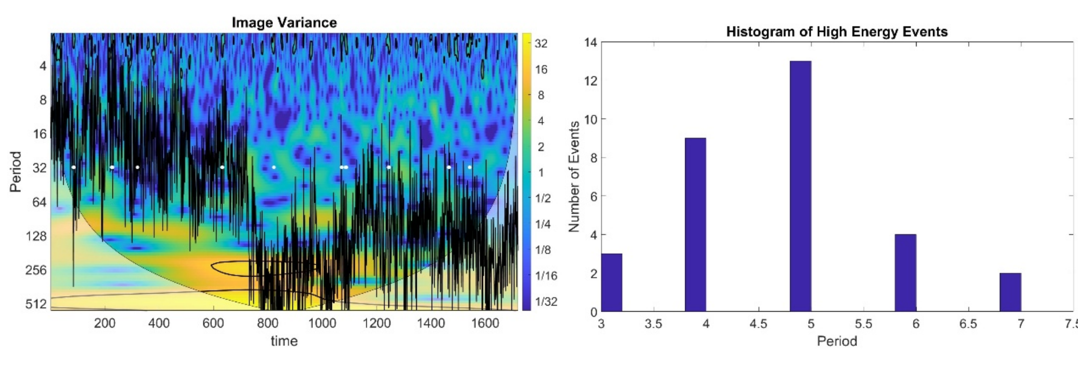

Figure 5. a) Morlet wavelet transform for σ with σ time series overlaid. Contours are 95% confidence intervals for time-scale space. White dots signify data points exceeding set constant wavelet threshold=2.9. b) High energy event histogram. Period scale on x-axis and number of events on y-axis. Physical units for time and scale are not applied.

6. Conclusions

The advent and continual development of robust geo-intelligence processors for distillation of spatio-temporal information shows great promise in providing timely information facilitating geointelligence decision making. While such processors, which rely heavily on using abundant amounts of data to provide state estimation, are powerful, there still is a need for statistical processors which provide initial reference points for data hungry algorithms to exploit. Even the most crude, statistical priors can assist state-of-the-art processors address the curse of dimensionality confronted by modern ATR systems, allowing improved chances of inverting observations and achieving low uncertainty solutions. This works falls within this category. The use of different processing pathways and the algorithms within these pathways afford information easily summarized allowing for improved future detection and characterization of anomalous data structure. Adequate assessment of the algorithmic parameters is needed to ensure future robust anomaly detection and characterization that is consistent across data sets. The pdfs, eigenvectors, emission matrices, and high energy event statistics in addition to the optimal thresholds, however, make the ensemble methodology useful and easily exploitable when future multidimensional data is processed.

References

[1] B. J. Schacter, *Automatic Target Recognition*, Washington, USA: SPIE Press Books, 2020.

[2] N. Kovvali, M. Banavar, and A. Spanias, *An Introduction to Kalman Filtering with Matlab Examples*, London, UK: Springer, 2013.

[3] G. A. Fink, *Markov Models for Pattern Recognition*, London, UK: Springer, 2014.

[4] P. A. Gagniuc, *Markov Chains: From Theory to Implementation and Experimentation*], Hoboken, NJ: John Wiley and Sons, 2017.

[5] L. E. Sucar, *Probabilistic Graphical Models: Principles and Applications*, London, UK: Springer, 2015.

[6] S. Marsland, *Machine Learning: An Algorithmic Perspective*, Boca Raton, FL: CRC Press, 2015.

[7] M. Belkin, and P. Niyogi, "Laplacian eigenmap for dimensionality reduction and data representation," *Neural Computation*, vol. 15, 6, pp. 1373-1396, 2003.

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

[9] A. J. Izenman, *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, New York, NY: Springer, 2013.

[10] W. L Martinez, and, A. R. Martinez, *Computational Statistics Handbook with Matlab*, Cambridge, MA: Cambridge University Press, 2009.

[11] S. S. Sharath, K. N. Balasubramanya Murthy, and S. Natarajan, "Dimensionality Reduction Techniques for Face Recognition," *Intech Open*, 26 October 2010, <https://www.intechopen.com/books/reviews-refinements-and-new-ideas-in-face-recognition/dimensionality-reduction-techniques-for-face-recognition>, 27 July 2011.

[12] P. S. Addison, *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science*, Engineering, Medicine and Finance], Bristol: UK, Institute of Physics Publishing, 2002.

[13] E. Foufoula-Georgiou, and P. Kumar, "Wavelet Analysis in Geophysics: An Introduction," in [*Wavelets in Geophysics*], ed. Foufoula-Georgiou, E. and Kumar, P., San Diego, CA: Academic Press,1994, pp. 1-43.

[14]. A. Grinsted, J. C. Moore, and S. Jevrejeva, *"Application of the cross wavelet transform and wavelet coherence to geophysical time series," Nonlinear Processes in Geophysics*, 11, pp. 561-566, 2004.