Gaussian Process Latent Variable Models Applied to Study Maritime Traffic Patterns from VIIRS Data

Raffaele Grasso
NATO STO CMRE, La Spezia, Italy
raffaele.grasso@cmre.nato.int

***Abstract*—Gaussian process latent variable models are used as a data dimensionality reduction technique and applied to analyze long spatio-temporal series of ship traffic patterns measured from data acquired by the Visible Infrared Imaging Radiometer Suite nighttime sensor on board the NOAA-Suomi National Polar-Orbiting Partnership spacecraft. The results show that these techniques are able to model traffic pattern with a number of variables much lower than the number of cells of the time-spatial grid supporting the input data. The use of a Bayesian formulation allows the introduction of spatio-temporal prior constraints that clearly improve the visualization of the time series in the reduced dimensionality space with respect to the classical principal component analysis.**

***Index Terms*—Machine Learning, Satellite imaging, Gaussian Process Latent Variable Models, time series analysis.**

I. Introduction

Gaussian Process latent variable models [1] (GPLVMs) and their variants are applied to analyze spatio-temporal series of ship contacts radiance measurements acquired by the Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime sensor on board the NOAA-Suomi National Polar-Orbiting Partnership (Suomi NPP) spacecraft.

The VIIRS sensor is able to measure ship self-emitting light during night time and, despite its low spatial resolution, it is possible to automatically detect a ship and assess the quality of the detection from the sensor radiance measurements [2]. The patterns provided by VIIRS data are validated against vessel monitoring systems (VMS) tracks demonstrating that the sensor can be used in enforcing fishery policies [3].

The primary aim of this paper is to analyze long temporal series of ship light mean intensity grids retrieved by binning ship contacts detected from the VIIRS sensor and to visualize the dynamic and the spatial distribution of the intensity using a two/three dimensional vector space representation that contains the majority of the information carried by the data set.

Maritime traffic analysis using historical data of ship tracks is applied in several fields. In maritime situational awareness and traffic surveillance, for instance, it helps to define models of the traffic normal behavior which can be used to develop on-line traffic anomaly detection systems and knowledge based ship tracking algorithms [4]. Moreover, the study of large data sets of traffic data helps in searching for seasonality patterns and trends, and in predicting the traffic over long periods, with important impacts on economy, policy enforcement like in fisheries and geopolitics. Such studies require a large amount of data. The automatic identification system (AIS) [5], for instance, with its network of transponders on board ships and land/satellite based receivers, is able to provide ship track data in a cooperative way, at high rates, allowing detailed traffic monitoring at global scale. The problem with the AIS is that not all the traffic is covered by the system (transponders on board ships are mandatory beyond a given gross tonnage and the transponders are usually switched off during illegal activities) and the coverage is not 100% reliable, especially in open sea areas not covered by land and satellite based AIS receivers. For these reasons, non cooperative sensors, providing ship contact measurements continuously and at global scale, like the VIIRS, are of great interest for complementing the AIS network and supplying historical traffic data to be analyzed.

Data dimensionality reduction techniques are largely applied to analyze spatio-temporal data series and visualize the dynamic and spatial variability of the data set. In [6], for instance, the principal component analysis (PCA) [7] is applied to study the temporal and spatial variability of oceanographic fields by processing long series of ocean variables estimated from remotely sensed data. In techniques like the PCA, the data belong to an observable vector space that is related to a latent space with reduced dimensionality. Given the vector data, the dimensionality reduction is achieved by estimating the corresponding vectors in the latent space, by optimizing a given information criterion.

The relation between the observation and the latent spaces in the PCA and in its probabilistic version [8] is linear. They are simple to apply, but usually they provide noisy estimates of the latent variables and they are less efficient than other techniques in compressing the information in few vector components, thus making the visualization of the variability of the data set in a 2D or 3D space less effective. In this paper, the GPLVM and its extensions are used to improve the visualization of the most informative latent components of a data set. In these techniques, the observation and latent spaces are related by a Gaussian process (GP) [9]. The method is more flexible than the PCA as it is able to allow non-linear relationships between the two vector spaces. The GPLVM is computationally more demanding than the PCA (or probabilistic PCA), but the Bayesian formulation of the technique allows the introduction of latent variable priors to smooth the solution and improve the visualization of the latent component dynamic and spatial variability.

The paper, after this introduction, is organized in a second

section summarizing the theory at the base of GPLVM and in a third section with examples of results on real VIIRS data. Section four draws conclusions and traces the way ahead.

II. Theory

A. GPLVM

The GPLVM can be defined by introducing a latent variable model which relates an observation space to a reduced dimensionality latent space by a generic input-output mapping function. In particular, given a matrix of N vector observations, $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]^\top$, with $\boldsymbol{y}_n = [y_{n,1}, \ldots, y_{n,D}]^\top$ and $n = 1, \ldots, N$, the observation component $y_{n,i}$, with $i = 1, \ldots, D$, can be related to the $n-th$ unknown column vector, $\boldsymbol{x}_n \in \mathbb{R}^M$, with $M < D$, by the following latent model:

$$y_{n,i} = f_i(\boldsymbol{x}_n; \boldsymbol{\Theta}) + \epsilon_n, \tag{1}$$

where ϵ_n is a zero mean Gaussian noise with variance σ^2, $f_i(\cdot)$ is a generic function and $\boldsymbol{\Theta}$ a vector of hyper-parameters. In case f_i is linear and the latent vectors $\boldsymbol{x}_n$ are zero mean multi-variate Gaussian distributed, (1) defines the so called probabilistic principal component analysis (PPCA) [8]. In case the relationship is non-linear, there is not a general way to propagate the uncertainty of the prior distribution of the latent variables and obtain a solution for $\boldsymbol{x}_n$ given the observations, thus achieving the desired reduction of dimensionality.

In the GPLVM [1] [10] a prior distribution is directly placed on the non-linearity, or latent function, which is assumed to be a zero mean Gaussian process [9]:

$$\boldsymbol{f}_i \sim GP(\boldsymbol{0}, \boldsymbol{K}_{N\times N}), \tag{2}$$

where $\boldsymbol{f}_i = [f_{1,i}, \ldots, f_{N,i}]^\top$, $f_{n,i} = f_i(\boldsymbol{x}_n; \boldsymbol{\Theta})$, with $i = 1, \ldots, D$. $\boldsymbol{K}_{N\times N}$ is the process covariance matrix defined from a covariance function kernel , $k(\boldsymbol{x}_l, \boldsymbol{x}_n)$, with $l, n = 1, \ldots, N$, so that [9]:

$$\boldsymbol{K}_{N\times N} = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ \vdots & & \vdots \\ k(\boldsymbol{x}_N, \boldsymbol{x}_1) & \ldots & k(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{bmatrix}. \tag{3}$$

The resulting non-parametric probabilistic model can be marginalized out with respect to the latent function to calculate the data likelihood which can be maximized to estimate the latent variable space and the model hyper-parameters. In particular, the marginal likelihood of the data given the latent variables can be written as:

$$p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta}) = \int p(\boldsymbol{Y}|\boldsymbol{F}) p(\boldsymbol{F}|\boldsymbol{X}; \boldsymbol{\Theta}) \, \mathrm{d}\, \boldsymbol{F}, \tag{4}$$

where $\boldsymbol{F} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_D]$ and $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top$. In this case, $\boldsymbol{\Theta}$ includes the hyper-parameters of the covariance function and the variance, σ^2, of the Gaussian noise components in (1), supposed to be independent identically distributed (iid). In case the GPs $\boldsymbol{f}_i$ are independent and share the same covariance function kernel, (4) can be written as [10]:

$$p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta}) = \prod_{i=1}^{D} p(\boldsymbol{y}_i|\boldsymbol{X}; \boldsymbol{\Theta}), \tag{5}$$

where $p(\boldsymbol{y}_i|\boldsymbol{X}; \boldsymbol{\Theta}) = \mathcal{N}[\boldsymbol{y}_i; \boldsymbol{0}_N, \boldsymbol{K}_{N\times N} + \sigma^2 \boldsymbol{I}_{N\times N}]$ and $\boldsymbol{y}_i$ is the $i-th$ column of the data matrix $\boldsymbol{Y}$, with $\boldsymbol{0}_N$ being the zero vector of size N.

The choice of the covariance function kernel is important in order to fit the model properly. The automatic relevant determination (ARD) technique [9] is used to estimate the dimensionality of the latent space. An example of such an ARD kernel is the widely used weighted squared exponential kernel [9] given by:

$$k(\boldsymbol{x}_l, \boldsymbol{x}_n) = \sigma_f^2 \exp\left[-\frac{1}{2} \sum_{i=1}^{M} \alpha_i (\boldsymbol{x}_{l,i} - \boldsymbol{x}_{n,i})^2 \right], \tag{6}$$

where α_i, with $i = 1, \ldots, M$, is the weight associated to the $i-th$ dimension of the latent vector $\boldsymbol{x}$ and σ_f^2 is a kernel variance parameter. The vector of hyper-parameters in (5) is thus given by $\boldsymbol{\Theta} = [\alpha_1, \ldots, \alpha_M, \sigma_f^2, \sigma^2]$. The latent variables and the hyper-parameters are estimated by maximizing the likelihood in (5). After that, the complexity of the model is reduced by discharging the latent components having very high α_i, ideally approaching infinity, achieving, in this way, a tradeoff with the data fit.

B. Bayesian GPLVM

The basic GPLVM in section II-A can be extended by introducing a prior, $p(\boldsymbol{X})$, for the latent variables. The hyper-parameters of the model are estimated by maximizing the likelihood $p(\boldsymbol{Y}; \boldsymbol{\Theta})$, obtained by marginalizing out $\boldsymbol{X}$ from $p(\boldsymbol{Y}, \boldsymbol{X}; \boldsymbol{\Theta})$ as follows [10]:

$$p(\boldsymbol{Y}; \boldsymbol{\Theta}) = \int p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta}) p(\boldsymbol{X}) \, \mathrm{d}\, \boldsymbol{X}, \tag{7}$$

where $p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta})$ is given by (5). $\boldsymbol{X}$ is estimated by maximizing its posterior $p(\boldsymbol{X}|\boldsymbol{Y}; \boldsymbol{\Theta}) \approx p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta}) p(\boldsymbol{X})$. This approach has several advantages with respect to estimating the model parameters by the maximum a posteriori probability (MAP) method. For example, the marginalization (7) of the latent variables makes the estimation method more robust to data over-fitting. On the contrary, the marginalization in (7) is intractable as the GP covariance contains non-linear terms of $\boldsymbol{X}$. By consequence, approximate inference methods are needed in order to estimate the parameters of the model. A widely used approach in this case is the variational inference (VI) one [8] [10]. It consists in approximating the posterior, $p(\boldsymbol{X}|\boldsymbol{Y})$, with a simple variational distribution, $q(\boldsymbol{X})$, having unknown parameters, and in finding a tractable lower bound for the logarithm of the marginal likelihood in (7). The bound is maximized to estimate the hyper-parameters $\boldsymbol{\Theta}$ and the parameters of $q(\boldsymbol{X})$ (here not included to avoid cluttering the notation). The variational distribution $q(\boldsymbol{X})$ is finally maximized to estimate the latent variables, $\boldsymbol{X}$.

Using the Jensen's inequality, the log-likelihood can be lower bounded as follows:

$$\begin{aligned} \log[p(\boldsymbol{Y}; \boldsymbol{\Theta})] &\geq \int q(\boldsymbol{X}) \log \frac{p(\boldsymbol{Y}|\boldsymbol{X}; \boldsymbol{\Theta}) p(\boldsymbol{X})}{q(\boldsymbol{X})} \, \mathrm{d}\, \boldsymbol{X} \\ &= \tilde{F}\left[q(\boldsymbol{X}); \boldsymbol{\Theta}\right] - \mathrm{KL}\left[q(\boldsymbol{X}) || p(\boldsymbol{X})\right] \end{aligned} \tag{8}$$

where $E_q[\cdot]$ is the statistical mean operator with respect to the distribution $q(\boldsymbol{X})$, $\tilde{F}\left[q(\boldsymbol{X});\boldsymbol{\Theta}\right] = E_q\left[p(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\Theta})\right]$ is a likelihood term and $\mathrm{KL}\left[q(\boldsymbol{X})||p(\boldsymbol{X})\right]$ is the Kullback-Leibler divergence between $q(\boldsymbol{X})$ and $p(\boldsymbol{X})$, defined as:

$$\mathrm{KL}\left[q(\boldsymbol{X})||p(\boldsymbol{X})\right] = E_q\left[\log\frac{q(\boldsymbol{X})}{p(\boldsymbol{X})}\right]. \tag{9}$$

The bound in (8) is a tradeoff between the need to fit the data, represented by the first likelihood term $E_q\left[p(\boldsymbol{Y}|\boldsymbol{X};\boldsymbol{\Theta})\right]$, and the constraint, given by the KL divergence term (9), to take $q(\boldsymbol{X})$ as close as possible to the prior $p(\boldsymbol{X})$. It can be demonstrated that the bound is tight, i.e. the equal sign in the inequality (8) is true, when $q(\boldsymbol{X})$ is equal to the posterior $p(\boldsymbol{X}|\boldsymbol{Y})$. The maximization of the bound with respect to the variational distribution thus provides an approximation to the true posterior. The interested reader can refer to [11] [10] for the detailed evaluation of that bound.

C. GPLVM extensions

Depending on the prior distribution assigned to the latent variables, a number of other forms of GPLVM have been defined in the scientific literature during the years [12] [13] [14] [15]. In particular, the dynamic GPLVM (DGPLVM) one is suitable to analyze temporal sequences of high dimensional vector data. The solution of the model is smoothed by defining a temporal prior over the latent variables which affects the choice of the variational distribution, $q(\boldsymbol{X})$, and the evaluation of the KL divergence (9). The back-constrained GPLVM (BCGPLVM) introduces further constraints to smooth the estimation of the latent space by jointly optimizing a backward model that allows the evaluation of a vector point in the latent space given an observation vector.

In the DGPLVM the data are organized as a multivariate time series, i.e. the index n is associated to a time instant t_n, with $n = 1,\ldots,N$. The latent variables are defined through a temporal latent function $\boldsymbol{x}(t) = [x_1(t),\ldots,x_M(t)]^\top$ where each component is an independent Gaussian process [13]:

$$x_i \sim GP\left[\mathbf{0}, k_t(t_l,t_n)\right], \tag{10}$$

where $k_t(t_n,t_l)$ is a temporal covariance function kernel, with $l,n = 1,\ldots,N$. The prior over the latent vector matrix $\boldsymbol{X}$ is thus given by:

$$p(\boldsymbol{X}|\boldsymbol{t}) = \prod_{j=1}^{M}\mathcal{N}\left[\boldsymbol{x}_{:,j};\mathbf{0},\boldsymbol{K}_x\right], \tag{11}$$

where $\boldsymbol{x}_{:,j}$ is the $j-th$ column of $\boldsymbol{X}$, $\boldsymbol{t} = [t_1,\ldots,t_N]$ and $\boldsymbol{K}_t$ is the $N\times N$ temporal covariance matrix calculated using the covariance function kernel $k_t(t_l,t_n)$ as follows:

$$\boldsymbol{K}_t = \begin{bmatrix} k_t(t_1,t_1) & k_t(t_1,t_2) & \ldots & k_t(t_1,t_N) \\ \vdots & \vdots & & \vdots \\ k_t(t_N,t_1) & k_t(t_N,t_2) & \ldots & k_t(t_N,t_N) \end{bmatrix}. \tag{12}$$

The resulting GPLVM model can be optimized using the variational approach detailed in section II-B.

The second technique to smooth the GPLVM consists in defining a back constraint, i.e. a function that maps a data point into a latent point as follows [15]:

$$x_{n,i} = g(\boldsymbol{y}_n;\boldsymbol{\Omega}), \tag{13}$$

where $\boldsymbol{\Omega}$ is a vector of unknown parameters, $n = 1,\ldots,N$ and $i = 1,\ldots,M$. In practice, the introduction of the back constraint into the optimization of the GPLVM forces points that are close in the data space to be also close in the latent space [15]. Moreover, the back constraint, once jointly trained with the GPLVM, provides a fast way to calculate latent points from new input vectors, that is useful, for instance, to generate features for classification or clustering tasks. The form of the back constraint can be any smooth parametric or non-parametric model including a Gaussian process like the forward mapping of the GPLVM. In this case, $\boldsymbol{\Omega}$ includes the hyper-parameters of the covariance function kernel of the backward GP.

D. Complexity

The complexity of algorithms involving GPs depends on the complexity to invert the kernel matrix, which is $\mathcal{O}(N^3)$. By consequence, the GPLVM typically scales badly for $N > 1000$. In order to improve scalability, some approximations are available. For example, in GPLVM with Nystrom approximation [9] the covariance matrix is approximated by a reduced rank matrix, $\boldsymbol{K} = \boldsymbol{K}_{N\times L}\boldsymbol{K}_{L\times L}^{-1}\boldsymbol{K}_{L\times N}$, with $\boldsymbol{K}_{L\times N} = \boldsymbol{K}_{N\times L}^\top$, whose inversion requires a complexity of $\mathcal{O}(L^2N)$, resulting in a considerable saving of computational time when $L << N$.

III. Results

The purpose of this work is to analyze temporal series of binned ship contact radiance measurements, detected from the day/night band of the VIIRS sensor [2], to study the temporal and spatial variability of the night vessel traffic over an area of interest in a given time range. Dimensionality reduction techniques play an important role in this type of analysis [6] in order to decompose the data spatial and temporal variability in such a way that can be visualized and described by a limited number of informative components. Differently from the PPCA, the GPLVM and its extensions allow the analysis to take into account data non-linearity and to smooth the estimation of the reduced dimensionality vector space through the introduction of suitable constraints.

The two sections below show examples of data analysis on areas delimited by two different exclusive economic zone (EEZ) borders. Section III-A shows an example of temporal data decomposition. The rows of the data matrix, $\boldsymbol{Y}$, are time snapshots of the ship radiance over space. The columns of the matrix are spatial features which are compressed by applying the GPLVM with ARD and retaining the three latent components having the higher sensitivity score (i.e the inverse of the ARD coefficients, α_i, with $i = 1,\ldots,M$, associated to each latent variable component). The retained components are time functions that can be used to visualize and analyze the

temporal variability of the radiance over the area. Section III-B shows an example of spatial decomposition where the analysis is on the transposed data matrix $\boldsymbol{Y}_s = \boldsymbol{Y}^\top$. In this case the matrix columns are temporal features and each row represent a spatial point. The retained latent components are now spatial functions suitable to understand the spatial variability of the radiance and find the traffic most active spots.

The temporal decomposition analysis is performed by using the DGPLVM with back constraints, while the spatial decomposition using the BCGPLVM. In both cases, the GPLVM models are solved using the Bayesian approach of section II-B, with a sparse approximation of the GP covariance matrix. The PPCA is used to initialize the iterative optimization of the latent space and it is compared with the final results. The work makes use of models implemented in the GPflow software library [16] and the GPmat software toolbox [17].

Ship contacts are retrieved from the VIIRS boat detection service [2] and temporally and spatially binned using a time bin of 14 days and a grid cell size of 40 $\times$ 40 km. Each grid spatial cell contains the average radiance of the contacts in a spatial-time bin. The whole data series span a period of about 3 years from 2016-07-01 and is composed of $N = 60$ temporal grids each with a number of spatial cells, D, depending on the extent of the region of interest. Figure 1 shows an example of a temporal series subset over the Vietnamese EEZ. The

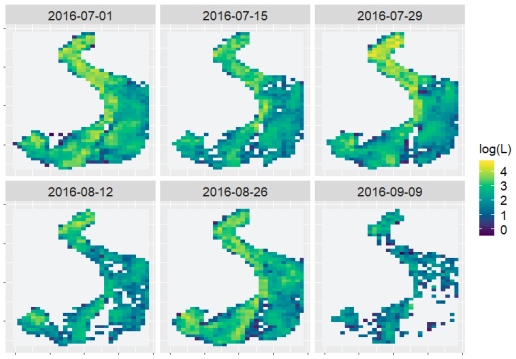

Fig. 1. Example of a short time series of binned ship radiance VIIRS data on the Vietnamese EEZ.

grids are lexicographically ordered to build a time series of $N = 60$ D-dimensional columns vectors which are organized in a $60 \times D$ data matrix. The matrix is pre-processed to evaluate the radiance logarithm and to center and scale the vectors by the data mean and standard deviation, respectively.

A. Temporal decomposition

This section shows an example of temporal data decomposition by using the DGPLVM with back constraints on data collected in the Taiwanese EEZ. In general, the DGPLVM provides smoother estimation of the latent space than PPCA and GPLVM, making the visualization of the temporal variability easier to interpret. In particular, the graph of the first 2 or 3 latent components clearly show the dynamic of the radiance and allow the detection of temporal structures such as attraction points and cycles.

Figure 2 shows the PPCA results on the data series collected on the Taiwanese EEZ. The local bi-weekly variability of the first 3 latent variables is quite high. Some longer temporal structures are visible in the latent variables temporal graphs on the left hand side of fig. 2. In particular, the first two variables, x_1 and x_2, are characterized by an oscillation of roughly the same period, of about one year, but with a relative phase shift. The visualization of the analysis in the latent space (see fig. 2 on the right side) shows how the local variability of the latent components contributes to hide the longer dynamic evolution of the data set. The PPCA in fig. 2 was used to

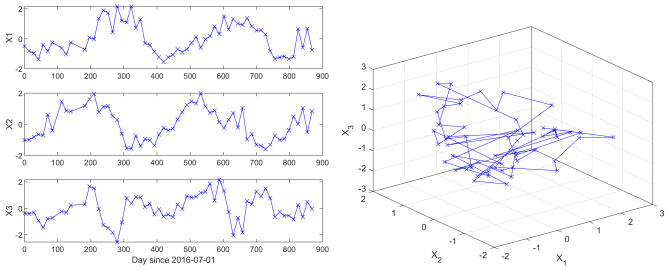

Fig. 2. PPCA applied to the data of the Taiwanese EEZ. Left: graph plot of the first three latent variables vs time. Right: data trajectory in the latent space.

initialize the optimization of the DGPLVM over the same area. Figure 3 shows the results of the optimized model. On the left side, the figure displays the temporal graph of the first three latent components. The stylized structures characterizing the latent variables in fig. 2 are almost still present in the latent variable graph, but this time, thanks to the dynamical constrains introduced by the DGPLVM, the local variability is much smoother than in the PPCA case. This provides the visualization in the latent space (see the right side of fig. 3) of a clear dynamical evolution of the radiance, with a cycling structure, resembling a "potato chip" shape.

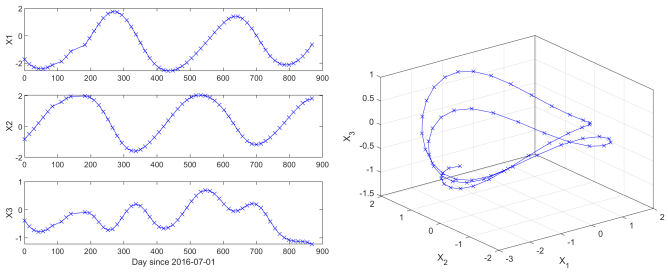

Fig. 3. Back-constrained DGPLVM applied to the data of the Taiwanese EEZ. Left: plots of the first three latent variables having the highest sensitivity. Right: "potato chip" data trajectory in the latent space.

B. Spatial decomposition GPLVM

This section shows an example of applying the BCGPLVM to the transposed data matrix, $\boldsymbol{Y}_s$. In this case, the time samples are considered as features (the columns of $\boldsymbol{Y}_s$), while each row represents a position in the spatial grid of the binned data set. Consequently, each latent variable is a spatial field over the given regular grid.

Figure 4 shows the results of the proposed analysis over the Vietnamese EEZ. The picture displays the first three latent variables over the spatial grid. Additionally, figure 5 shows 2D and 3D scatter plots of the latent variables in which clusters of spatial points are clearly visible.

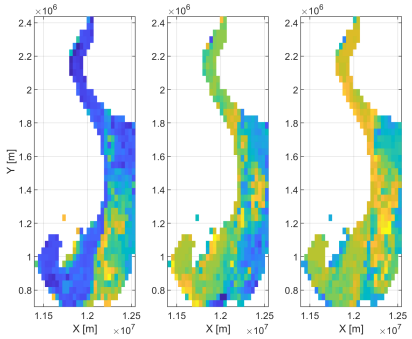

Fig. 4. Spatial distribution of the first 3 latent variables of the optimized BCGPLVM over the Vietnamse EEZ having the highest sensitivity.

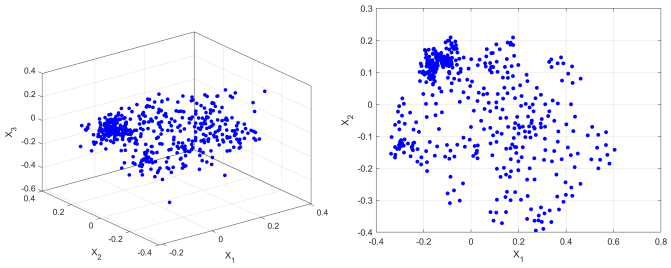

Fig. 5. Scatter plots of the first three latent variables of the BCGPLVM over the Vietnamese EEZ. Left: 3D scatterplot. Right: projection onto the space spanned by the first two latent variables.

The informative features provided by such an analysis are useful to detect spots over the spatial grid characterized by similar radiance field statistical properties.

IV. Conclusions

This report shows preliminary results on the application of GPLVM techniques to the analysis of vessel traffic patterns estimated using ship contacts detected from VIIRS satellite sensor data. The results allow the following observations:

1) the information contained in the input data can be efficiently compressed in a number of latent variables that is much less than the initial dimensionality of the data,
2) the first 3 latent variables contain meaningful temporal-spatial features whose physical interpretation is under investigation,
3) the paper shows a possible approach to study traffic patterns by using non-cooperative satellite sensors that complement the AIS.

Future work includes additional validation and comparison with AIS data patterns. Further work will be focused on the study, design and test of new algorithms for data clustering/segmentation, prediction and anomaly detection making use of the retained latent variables and their probabilistic characterization.

V. Acknowledgment

This work has been funded by the NATO ACT project *Data Knowledge Operational Effectiveness*.

References

[1] P. Li and S. Chen, "A review on Gaussian process latent variable models," *CAAI Transactions on Intelligence Technology*, 11 2016.

[2] C. Elvidge, M. Zhizhin, K. Baugh, and F.-C. Hsu, "Automatic boat identification system for viirs low light imaging data," *Remote Sensing*, 03 2015.

[3] F. Hsu, C. D. Elvidge, K. E. Baugh, M. N. Zhizhin, T. Ghosh, D. Kroodsma, A. Susanto, W. Budy, M. Riyanto, R. Nurzeha, and Y. Sudarja, "Cross-matching VIIRS boat detections with vessel monitoring system tracks in indonesia," *Remote Sensing*, vol. 11, no. 9, p. 995, 2019. [Online]. Available: https://doi.org/10.3390/rs11090995

[4] R. Grasso, L. M. Millefiori, and P. Braca, "Bayesian track-to-graph association for maritime traffic monitoring," *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1042–1046, 2018.

[5] ITU, "Technical characteristics for an automatic identification system using time division multiple access in the vhf maritime mobile band," International Telecommunications Union, Tech. Rep. ITU-R M.1371-5, 02 2014.

[6] A. Baldacci, G. Corsini, R. Grasso, G. Manzella, J. Allen, P. Cipollini, T. Guymer, and H. Snaith, "A study of the alboran sea mesoscale system by means of empirical orthogonal function decomposition of satellite data," *Journal of Marine Systems*, vol. 29, pp. 293–311, 05 2001.

[7] I. Jolliffe, *Principal component analysis*. New York: Springer Verlag, 2002.

[8] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006. [Online]. Available: http://research.microsoft.com/en-us/um/people/cmbishop/prml/

[9] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[10] M. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, Y. W. Teh and M. Titterington, Eds., vol. 9. Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 844–851. [Online]. Available: http://proceedings.mlr.press/v9/titsias10a.html

[11] A. C. Damianou, M. K. Titsias, and N. D. Lawrence, "Variational inference for latent variables and uncertain inputs in gaussian processes," *Journal of Machine Learning Research*, vol. 17, no. 42, pp. 1–62, 2016. [Online]. Available: http://jmlr.org/papers/v17/damianou16a.html

[12] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, Feb 2008.

[13] N. D. Lawrence and A. J. Moore, "Hierarchical Gaussian process latent variable models," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07. New York, NY, USA: ACM, 2007, pp. 481–488. [Online]. Available: http://doi.acm.org/10.1145/1273496.1273557

[14] A. C. Damianou, M. K. Titsias, and N. D. Lawrence, "Variational Gaussian process dynamical systems," in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, ser. NIPS'11. USA: Curran Associates Inc., 2011, pp. 2510–2518. [Online]. Available: http://dl.acm.org/citation.cfm?id=2986459.2986739

[15] N. D. Lawrence and J. Quiñonero Candela, "Local distance preservation in the gp-lvm through back constraints," in *Proceedings of the 23rd International Conference on Machine Learning*, ser. ICML '06. New York, NY, USA: ACM, 2006, pp. 513–520. [Online]. Available: http://doi.acm.org/10.1145/1143844.1143909

[16] A. G. d. G. Matthews, M. van der Wilk, T. Nickson, K. Fujii, A. Boukouvalas, P. Le'on-Villagr'a, Z. Ghahramani, and J. Hensman, "GPflow: A Gaussian process library using TensorFlow," *Journal of Machine Learning Research*, vol. 18, no. 40, pp. 1–6, apr 2017. [Online]. Available: http://jmlr.org/papers/v18/16-537.html

[17] S. M. L. Software, "Sheffieldml/gpmat," Jun 2017. [Online]. Available: https://github.com/SheffieldML/GPmat