LATENT FORCE MODELS FOR EARTH OBSERVATION TIME SERIES PREDICTION

David Luengo, Manuel Campos-Taberner and Gustau Camps-Valls

Dep. Signal Theory and Communications, Universidad Politécnica de Madrid, Spain
Dep. Earth Physics and Thermodynamics, Universitat de València, Spain
Image Processing Laboratory (IPL), Universitat de València, Spain

ABSTRACT
We introduce latent force models for Earth observation time series analysis. The model uses Gaussian processes and differential equations to combine data driven modelling with a physical model of the system. The LFM presented here performs multi-output structured regression, adapts to the signal characteristics, it can cope with missing data in the time series, and provides explicit latent functions that allow system analysis and evaluation. We successfully illustrate the performance in challenging scenarios of crop monitoring from space, providing time-resolved time series predictions.

1. INTRODUCTION
The Earth is a highly complex and evolving networked dynamical system. In the last few hundred years, human activities have precipitated an environmental crisis on Earth, commonly known as global climate change [1]. In this context, Earth-observation (EO) provide a unique source of information to address some of the challenges of the Earth system science [2]. To accomplish this ambitious goal, EO deploys both data acquired by remote sensing airborne and satellite sensors, as well as quantitative in situ measurements of biophysical parameters [3], such as the surface temperature, chlorophyll content in plants, or vegetation coverage just to name a few.

In this scenario, relevant vegetation variables, such as the green leaf area index (LAI) and the related Fraction of Absorbed Photosynthetically Active Radiation (fAPAR), need to be closely monitored from spaceborne satellites. LAI is a key bio-physical parameter which represents half of the total leaf area per unit of ground area [4], while fAPAR accounts for the light absorption across an integrated plant canopy. Both variables are used as indicators of the state and evolution of the vegetation cover. Predictive models use the reflected energy acquired by satellite sensors to predict such biophysical variables, and hence to deploy spatially explicit and time-resolved vegetation maps. LAI and fAPAR actually play a key role in vegetation processes, such as photosynthesis and transpiration, and are connected to meteorological/climate and ecological land processes. LAI/fAPAR have been widely used in many agricultural and remote sensing studies [5-8], and have been used in crop models simulation [9]. Actually, LAI prediction from satellite data is among the main goals of the remote sensing community [10], as evidenced by the variety and usefulness of operational medium resolution products for vegetation monitoring from satellite sensors, such as Moderate Resolution Imaging Spectroradiometer (MODIS) [11] or Système Pour l’Observation de la Terre VEGETATION (SPOT-VGT) [12].

From a methodological point of view, LAI retrieval can be faced following several regression approaches based on either statistical, physical, or hybrid methods [13,14]. In the last decade, non-parametric regression algorithms excelled in biophysical parameter retrieval in general, and vegetation parameter estimation in particular. Current operational vegetation products are typically produced with Neural Networks (NN) [12]. Support vector regression showed high efficiency in modelling LAI, fCOVER and evapotranspiration [15]. Nevertheless, in the recent years, Gaussian processes regression (GPR) [16], provided encouraging results in the framework of bio-physical parameter estimation outperforming the rest [17].

Despite the success of GPR in this domain, some particular problems are foreseen. First, GPR models have been so far applied in static fashion, without encoding the relevant temporal information conveyed by phenological cycles of vegetation. Second, the few time-based covariances available nowadays are mere parametrizations encoding ad hoc simple features such as trends and periodicities, hence no richer relations are learned from the data in a nonparametric way. Third, few attempts of multi-output and structured GPR models are available, which hampers adoption of these models in real EO problems. Finally, we should note that very often the EO time series show uneven sampling because of the acquisition artifacts and the presence of clouds. The latter is typically solved by parametric gap filling interpolation, which alters the time series structure by oversmoothing the series, thus producing unrealistic results.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement 606983, the European Research Council (ERC) under the ERC-COG-2014 SEDAL project (grant agreement 647423), and the Spanish Ministry of Economy and Competitiveness (MINECO) through the projects TIN2015-64210-R and TEC2015-64835-C3-3-R. The Land-SAF (EUMET-SAT) project is also acknowledged.
In this paper, we introduce in the field of EO data processing a latent force model (LFM) for GPR modeling that addresses all these problems simultaneously. The model, based on the GP-LFM described in [18,19], uses Gaussian processes and smoothing kernels obtained from basic characteristics of the problem at hand to combine data driven modeling with a physical model of the system. The LFM presented here performs multi-output structured regression, adapts to the signal characteristics, is able to cope with missing data in the time series, and provides explicit latent functions that allow system analysis and evaluation.

2. LATENT FORCE MODELS GP

2.1. Problem Statement

Let us consider a set of \( Q \) correlated time series, \( y_q(t) \) for \( 1 \leq q \leq Q \), and let us assume that we have \( N \) samples available for each of these signals, taken at sampling points \( t_n \), s.t. \( y_q[n] = y_q(t_n) \) for \( 1 \leq n \leq N \). This is the training set, which is composed of an input vector, \( \mathbf{t} = [t_1, \ldots, t_N] \), and an output matrix, \( \mathbf{Y} = [y_1, \ldots, y_Q] \), with \( y_q = [y_q[1], \ldots, y_q[N]]^T \). Our goal is learning a statistical model for that set of \( Q \) outputs or tasks that can be used thereafter to perform inference on the test set: \( \mathbf{t} = [\tilde{t}_1, \ldots, \tilde{t}_M]^T \) and \( \mathbf{\tilde{Y}} = [\tilde{y}_1, \ldots, \tilde{y}_Q] \) with \( \tilde{y}_q = [\tilde{y}_q[1], \ldots, \tilde{y}_q[M]]^T \) and \( \tilde{y}_q[n] = y_q(t_{\text{test}}) \). The challenge in multi-output time series modelling is constructing, based on the empirical evidence (i.e. the test set) and the prior information available about the problem at hand, a model with a reasonable complexity (i.e. storage requirements and computational cost) that effectively and efficiently captures the interrelations between the outputs. In the following section, we describe a model based on Gaussian process latent functions that is able to achieve these goals.

2.2. Gaussian Process latent force models

Let us assume that a set of \( R \) independent latent functions (LFs), \( f_r(t) \) with \( 1 \leq r \leq R \), are responsible for the observed correlation between the outputs. Then, the cross-correlation between the outputs arises naturally as a result of the coupling between the set of independent LFs, instead of being imposed directly on the set of outputs. Two critical aspects of the model are (1) providing a statistical description for the latent functions and (2) detailing the coupling mechanism between them. With respect to the LFs, we model them as zero-mean Gaussian processes (GPs) with a zero mean function, \( \mu_f(r, t) = 0 \), and an arbitrary auto-covariance function (ACF) \( k_{f_r, f_r}(t, t') \). The key point now is the choice of the LFs’ ACF, since it encodes our a priori knowledge about basic characteristics of the LFs such as continuity, smoothness, stationarity, etc. Although several choices of smooth and non-smooth functions are possible (e.g., see [18–20]), in the sequel we consider only the squared exponential, RBF or Gaussian auto-covariance function:

\[
k_{f_r, f_r}(t' - t) \propto \exp \left( -\frac{(t' - t)^2}{2\ell_r^2} \right),
\]

where the hyperparameter \( \ell_r \) controls the length-scale of the process. On the other hand, the smoothing kernel encodes our knowledge about the linear system that relates the unobserved LFs and the outputs, and can be based on basic physical principles of the system at hand (as in [18,19]) or selected arbitrarily (as in [21,22]). In this paper, we also consider an RBF smoothing kernel:

\[
h_q(t) \propto \exp \left( -\frac{t^2}{2\gamma_q^2} \right).
\]
Finally, note that, since the LFs are zero-mean GPs, the noise is also zero-mean and Gaussian, and all the operators involved in (1) and (2) are linear, hence the joint LFs-outputs process is also a GP. Hence, we only need to obtain the corresponding mean and cross-covariance functions in order to characterize it completely. Regarding the outputs, the mean function of the q-th output is \( \mu_{y_q}(t) = 0 \), whereas the cross-covariance function between two outputs is

\[
\begin{align*}
\kappa_{y_q y_r}(t, t') = & \sum_{p=1}^{P} S_{r, p} S_{q, p} L_{p}[t] \left\{ L_{q}[t'] \left\{ k_{f, f, \{t, t'\}} \right\} \right\} \\
& + \eta^2 \delta[p - q] \delta[t' - t],
\end{align*}
\]

where the term \( L_{p}[t] \left\{ L_{q}[t'] \left\{ k_{f, f, \{t, t'\}} \right\} \right\} \) denotes the application of the convolutional operator twice to the ACF of the LFs, which results in the following double integral:

\[
L_{p}[t] \left\{ L_{q}[t'] \left\{ k_{f, f, \{t, t'\}} \right\} \right\} = \int_{0}^{t} \int_{0}^{t'} h_{p}(t - \tau) h_{q}(t' - \tau') \times k_{f, f, \{\tau, \tau'\}} \mathrm{d} \tau \mathrm{d} \tau'.
\] (6)

Finally, we need the cross-correlation between the LFs and the outputs to complete the characterization of the joint process:

\[
k_{f, y_q}(t, t') = S_{f, q} L_{q}[t'] \left\{ k_{f, f, \{t, t'\}} \right\},
\] (7)

which involves a single one-dimensional integral already required for evaluating (5). Fortunately, all the integrals can be solved analytically when both the LFs and the smoothing kernel have a Gaussian shape, and the cross-covariance between the outputs stored in the matrix \( \mathbf{K}_{yy} \), which is required for learning and inference, as described in the following section.

2.3. Learning and inference

Learning the model amounts to estimating its hyperparameters, \( \theta \). Since the joint PDF of the outputs is Gaussian, the maximum likelihood (ML) estimator of the hyperparameters can be obtained by minimizing the minus log-likelihood function. Denoting as \( \mathbf{y} = [y_1^T, \ldots, y_q^T] \) the vectorized version of matrix \( \mathbf{Y} \) that contains all the samples from all the time series in the training set, the cost function becomes

\[
J(\theta) = -\log p(\mathbf{y}; \theta) \propto \mathbf{y}^T \mathbf{K}_{yy}^{-1} \mathbf{y} + \log |\mathbf{K}_{yy}|.
\] (8)

where the dependence on the hyperparameters is contained within \( \mathbf{K}_{yy} \), \( \log \) denotes the natural logarithm, and we have dropped all the elements that do not depend on the hyperparameters. Unfortunately, the dependence of (8) on \( \theta \) can be quite complex, so it is not possible to obtain an analytical expression for the ML estimate of the hyperparameters. Hence, we use a stochastic gradient descent technique, the scaled conjugate gradient as implemented in NETLAB [23], to minimize (8).

Once the hyperparameters of the model have been learned, inference proceeds by applying standard GP regression formulas. The predictive density is given by [16]:

\[
p(\tilde{\mathbf{y}} | \mathbf{y}) = \mathcal{N}(\mathbf{\bar{y}} | \mu_{\tilde{y}|y}, \mathbf{K}_{\tilde{y}|y}),
\] (9)

where \( \tilde{\mathbf{y}} = [\tilde{y}_1^T, \ldots, \tilde{y}_Q^T] \), and

\[
\begin{align*}
\mu_{\tilde{y}|y} &= \mathbf{K}_{\tilde{y}y} \mathbf{K}_{yy}^{-1} \mathbf{y}, \\
\mathbf{K}_{\tilde{y}|y} &= \mathbf{K}_{\tilde{y}y} - \mathbf{K}_{\tilde{y}y} \mathbf{K}_{yy}^{-1} \mathbf{K}_{\tilde{y}y}.
\end{align*}
\] (10)

Now, since the conditional PDF is Gaussian, the minimum mean squared error (MMSE) prediction is simply given by the conditional mean:

\[
\hat{\mathbf{y}} = \mu_{\tilde{y}|y} = \mathbf{K}_{\tilde{y}y} \mathbf{K}_{yy}^{-1} \mathbf{y},
\] (11)

where \( \hat{\mathbf{y}} = [\hat{y}_1^T, \ldots, \hat{y}_Q^T] \) is the vectorized version of the inferred outputs, which can be expressed in matrix form as \( \hat{\mathbf{y}} = \hat{\mathbf{y}}_q \) with \( \hat{\mathbf{y}}_q = [\hat{y}_q[1], \ldots, \hat{y}_q[M]]^T \) and \( \hat{\mathbf{y}}_q[m] = \hat{y}_q(t_m) \).

3. DATA COLLECTION

In this paper, we are bringing to light the usefulness of the proposed method for filling gaps in time series of two of the main remote sensing bio-physical variables such as LAI and FAPAR. In the experiments, we are using multitemporal LAI and FAPAR products corresponding to the main Mediterranean rice areas. The data were derived in the context of the ERMES European project. ERMES aims to develop a prototype of COPERNICUS down-stream services based on the assimilation of Earth observation (EO) and in situ data within crop modeling solutions dedicated to the rice sector. In this framework, the ERMES study areas have been selected in Spain, Italy and Greece, which are the three countries responsible of 85% of total European rice production. The Spanish study area extends from Tarragona to the region of Valencia, and is one the main rice cultivation areas of the country. The Italian study area is located in the Lomellina rice district, which is located in the south-western Lombardy region, between the Ticino, Sesia and Po rivers. The Greek study area covers the two main rice cultivation areas of the country, Thessaloniki (180,000 ha) and Serres (4,000 ha), where cultivation represents almost the 75% of the Greek rice area. Within each study area, rice is a common crop with a long tradition and cultural and economical value.

For rice, LAI ranges between values close to zero to a maximum of 10 at flowering, although maximum values closer to 6 are the norm, while FAPAR ranges from 0 to 1. In this paper we focus on a set of representative rice pixels of each ERMES area thus allowing us to observe the interannual variability of rice from 2003 to 2014 at coarse spatial resolution (2Km) which is useful for regional vegetation modelling. Concretely, the ERMES LAI and FAPAR products were derived from two optical sensors onboard satellite platforms (MODIS and SPOT-VGT) thus obtaining similar mutitemporal trends computed over different dates from two different sources. The ERMES LAI/FAPAR based on SPOT-VGT data provides estimates every 10 days while the one based on MODIS data provides estimates every 8 days.
4. EXPERIMENTAL RESULTS

This section gives experimental evidence of the proposed GP-LFM for challenging scenarios in EO time series analysis. We study the learned latent forces, the prediction accuracy in several time series of LAI and fAPAR derived from the MODIS sensor, and assess the robustness to high missing data rate regimes.

4.1. Multi-output time series: learning the latent forces

In this section, we focus on learning the latent forces for the multi-output time series composed of the LAI and fAPAR data for Spain and Italy (i.e., the number of outputs is $Q = 4$) from the beginning of 2003 until the end of 2013. We have experimented with a variable number of latent forces, $R \in \{1, 2, 3\}$, yet using more than $R = 1$ LFs the model tends to overfit. Therefore, all the results displayed in the sequel correspond to a single LF (i.e., $R = 1$). We use all the data available from the MODIS sensor ($N = 506$ samples per time series in total) without removing any data, except for truly missing data (marked with negative values in the original time series).

The recovered LF and two examples of the modelled time series are displayed in Fig. 2. Note that the model has succeeded in capturing the dynamics of the data by using a single LF. A quantitative measure of performance is provided by the mean squared error (MSE),

$$\text{MSE}_q = \frac{1}{N_q} \sum_{n=0}^{N_q-1} (y_q[n] - \hat{y}_q[n])^2, \quad (12)$$

and the normalised MSE,

$$\text{NMSE}_q(\%) = \frac{\text{MSE}_q}{\frac{1}{N_q} \sum_{n=0}^{N_q-1} \hat{y}_q^2[n]} \times 100, \quad (13)$$

where $y_q[n]$ denotes the true value of the $n$-th sample from the $q$-th time series, $\hat{y}_q[n]$ is the value predicted by the model, $N_q \leq N$ is the number of samples available for the $q$-th time series ($N_q < N$ if there are true missing data that have been removed) and $q = 1, \ldots, Q$. In this case, for Spain we have $\text{MSE} = 0.1139$ (NMSE = 2.08 %) and $\text{MSE} = 0.0080$ (NMSE = 4.02 %) for LAI and fAPAR respectively, whereas for Italy we have $\text{MSE} = 0.2422$ (5.97 %) and $\text{MSE} = 0.0046$ (NMSE = 2.49 %), again for LAI and fAPAR respectively, showing the excellent modelling results in all cases.

Then, we explore the LAI vs. fAPAR relationship. Figure 3 displays the LAI vs. fAPAR scatter plot, obtained from the modelled time series for Italy. The shaded area corresponds to the uncertainty (that appears now in both axis, as a consequence of the modelling uncertainty in both LAI and fAPAR), whereas the continuous red line shows the well-known exponential model usually assumed, $\text{fAPAR} = 1 - \exp(\alpha \times \text{LAI})$, that has been fitted to the data by performing a simple linear regression in the log-domain, resulting in $\alpha = -0.4653$. Once more, we can see that the model has been able to capture the expected LAI vs. fAPAR relationship.
4.2. Multisite gap filling and domain adaptation

In this second experiment, we investigate the capability of the method of recovering missing data (i.e., filling gaps in the measurements). In order to do so, we use all the available data from LAI (ES) and remove complete years from the other three time series (fAPAR (ES), LAI (IT) and fAPAR (IT)). We start by removing the data from year 2009 and end up removing the data from the years 2009–2013 (i.e., almost half of the available data from the last three time series). The modelled time series are displayed in Fig. 4 for R = 1, whereas Table 1 shows the absolute and normalised MSE. On the one hand, from Fig. 4 we can see that the model has been able to capture the correlations among all the time series, since it is able to reconstruct the second half of the last three outputs when only the first one is available for that time interval. On the other hand, note that the error always remains within the same order of magnitude, even when we completely removed the second half of the data from the last three time series. This confirms the robustness of the proposed approach w.r.t. large amounts of missing data, and its suitability for the gap filling problem in Earth observation applications.

<table>
<thead>
<tr>
<th>Missing years</th>
<th>LAI (IT)</th>
<th>fAPAR (ES)</th>
<th>fAPAR (IT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>0.2469 (6.13 %)</td>
<td>0.0123 (5.79 %)</td>
<td>0.0055 (3.82 %)</td>
</tr>
<tr>
<td>2009–2010</td>
<td>0.2865 (7.02 %)</td>
<td>0.0118 (6.61 %)</td>
<td>0.0073 (3.96 %)</td>
</tr>
<tr>
<td>2009–2011</td>
<td>0.2439 (5.79 %)</td>
<td>0.0127 (6.09 %)</td>
<td>0.0061 (3.27 %)</td>
</tr>
<tr>
<td>2009–2012</td>
<td>0.2734 (6.75 %)</td>
<td>0.0133 (6.31 %)</td>
<td>0.0067 (3.66 %)</td>
</tr>
<tr>
<td>2009–2013</td>
<td>0.2997 (7.26 %)</td>
<td>0.0139 (6.54 %)</td>
<td>0.0069 (3.72 %)</td>
</tr>
</tbody>
</table>

Table 1. Absolute and normalised MSE for the missing data (test set) using R = 1 LFs.

4.3. Dealing with high missing data rates

In the final experiment, we explore the limits of the proposed method by analyzing its performance w.r.t. randomly distributed missing data. In order to fulfill this goal, we set a missing probability (P = 0.1, 0.2, . . . , 0.9) for each sample of each of the two countries (ES and IT), i.e., for each sample of ES and IT we decide randomly and independently (according to the probability P) whether it should be removed and, if we decide to remove it, then we erase both LAI and fAPAR from our training data set for that sample. This simulates the common scenario in EO where products cannot be generated from optical sensors due to the presence of clouds. The evolution of the NMSE as P increases is shown in Fig. 5 for R = 1 latent forces. Let us remark that the MSE for P ≤ 0.5 does not increase substantially w.r.t. the MSE without missing data (except for LAI (ES), but even here the relative MSE is around 4 % for P = 0.5), i.e., the absence of up to 50 % of the data is not a problem at all for the method. For P > 0.5 the performance starts getting worse gradually, but the proposed approach is able to provide reasonably good results even when 90 % of the data are missing (the relative MSE is always below 20 % and even below 10 % in two cases). Fig. 6 shows an example of the fAPAR obtained for P = 0.9, showing that even in this extreme case (where 456 and 453 out of the...
Fig. 6. Example of a high missing data rate scenario ($P = 0.9$, so 90% of the data are missing) using a single LF (i.e., $R = 1$). Training data (red circles), test data (red dashed line), predicted time series (black line) and uncertainty measured by ±2 standard deviations about the mean predicted value (grey shaded area).

506 available samples are missing for Spain and Italy, respectively) the method is able to capture the underlying dynamics of the multi-output time series. Notice also that, even in the large uncertainty regions, where no samples at all are available, the predicted value is not so far from the true value.

5. CONCLUSIONS

In this paper we introduced latent force models for crop monitoring from satellite sensors. The presented model uses Gaussian processes and differential equations to combine data driven modelling with a physical model of the system. The model performs multi-output structured regression, adapts to the signal characteristics, it can cope with missing data in the time series, and provides explicit latent functions that allow system analysis and evaluation. We successfully illustrated the performance, providing time-resolved time series predictions.

6. REFERENCES