Modelling Non-Stationary and Non-Separable Spatio-Temporal Changes in Neurodegeneration via Gaussian Process Convolution

Lorenzi Marco[1], Gabriel Ziegler[2], Daniel C. Alexander[3], and Sebastien Ourselin[1] ⋆

[1] Translational Imaging Group, CMIC, UCL, London, UK
[2] Centre for Medical Image Computing (CMIC), UCL, London. UK
[3] Wellcome Trust Centre for Neuroimaging, UCL, London, UK

Abstract. Modelling longitudinal changes in organs is fundamental for the understanding of biological and pathological processes. Most of the previous works on spatio-temporal modelling of image time series relies on the assumption of stationarity of the local spatial correlation, and on the separability between spatial and temporal processes. These assumptions are often made in order to lead to computationally tractable approaches to longitudinal modelling, but inevitably lead to an oversimplification of the complex spatial and temporal dynamics underlying the biological processes. In this work we propose a novel spatio-temporal generative model of time series of images based on kernel convolutions of a white noise Gaussian process. The proposed model is parameterised by a sparse set of control points independently identified by specific spatial and temporal parameters. This formulation is highly flexible and can naturally account for spatially and temporally varying dynamics of changes. We demonstrate a preliminary application of our non-parametric method on the modelling of within-subject structural changes in the context of longitudinal analysis in Alzheimer's disease. In particular we show that our method provides an accurate description of the pathological evolution of the brain, while showing high flexibility in modelling and predicting region-specific non-linearity due to accelerated structural decline in dementia.

1 Introduction

Modelling longitudinal changes in organs is fundamental for the understanding of biological and pathological processes. For instance the development of a spatio-temporal model of disease progression in Alzheimer's disease (AD) from time

series of magnetic resonance images (MRIs) would be highly valuable for the fundamental understanding of the disease process, for diagnostic purposes and individual predictions, and for testing the efficacy of disease modifying drugs in clinical trials.

The consistent modelling and prediction of spatio-temporal changes in longitudinal MRI is still an important challenge from both methodological and computational perspectives. In fact, flexible modelling instruments are required in order to robustly capture meaningful pathological accelerations specific to sensitive brain regions. Moreover, since a biological model of local brain changes is often unknown, it is important to develop optimal models in terms of statistical complexity. Notably, the spatial dimensionality of MRI time series prevents the straightforward implementation of classical multivariate statistical modelling techniques and often leads to computationally intractable solutions.

We can identify two main approaches to spatio-temporal modelling of image time-series in computational anatomy. The first one is based on non-linear image registration, describing signal differences between images as local spatial transformations [1,2,3,4]. In non-linear registration the spatial changes are usually modelled at a fixed spatial scale defined by the regularization energy at which the transformation is optimized. The temporal modeling usually relies on the definition of a specific model of temporal evolution, which is identified either by fitting parametric progression models on geometric features of the transformation, or by choosing an opportune metric in the space of transformations to characterize specific evolution models in the image space. The second one, usually identified as voxel-based-morphometry (VBM), is based on voxel-by-voxel modelling based on parametric [5], or non-parametric regression frameworks [6]. Models are usually independently fitted for each voxel, and local correlation is usually imposed by applying Gaussian convolution of the images with some apriori kernel size.

The majority of the above mentioned approaches rely on important assumptions concerning the spatial and temporal processes. In fact, by either choosing a global regularization energy in image registration, or a global smoothing parameter in VBM, we usually impose local *stationary* correlation models for the spatial changes. Even though this assumption is often necessary to lead to computationally tractable approaches, it inevitably leads to an oversimplification of the complex spatial properties of the images, for example concerning regionally varying smoothness, and image boundaries.

At the same time, by fitting global longitudinal models, either defined by the registration metric, or by a fixed statistical model complexity, we assume that spatial and temporal processes are *separable*, i.e. that the properties of the temporal variation (for instance following a quadratic or linear behaviour) is independent from the spatial locations. As before, this assumption often leads to simplistic modeling solution, as the progression of the temporal changes in organs is generally highly variable across spatial regions.

Non-parametric Gaussian process (GP) models have emerged as a flexible and elegant Bayesian approach for prediction of continuous and binary vari-

ables in manifold applications [7]. Recently, GPs were successfully introduced to the field of neuroimaging, e.g. in the context of single-case inference in aging [6]. Moreover, it was recently introduced in [8] a generative framework for the modelling of image time series based on Gaussian process regression. This approach was however completely based on the assumption about stationarity and separability of spatial and temporal processes.

In this work we propose a generative model of image time series based on Gaussian processes (GPs), which is characterized by a covariance structure parameterized by a sparse set of control points defined in space and time. Since each control point is governed by specific spatial and temporal parameters, the proposed model is highly flexible and can naturally account for spatially and temporally varying signal changes. The proposed model thus overcomes many limitations of previous spatio-temporal modelling approaches.

The paper is organized as follows. In Section 2 we propose our generative model of longitudinal changes parameterized by a sparse set of control points, with details about parameter optimization and prediction. In Section 3 we then provide a preliminary application of our non-parametric method on the modelling of within-subject structural changes in the context of longitudinal analysis in Alzheimer's disease. In particular we show that our method provides an accurate description of the pathological evolution of the brain, while showing high flexibility in modelling and predicting region-specific non-linearity due to accelerated structural decline in dementia.

2 Generative Model of Spatial Data Through Gaussian Process Convolution

Let u and t be respectively the spatial and temporal coordinates. Given an image time series $y(s)$, $s = (u, t)$, we assume a generative model for the spatio-temporal variations:

$$y(s) = z(s) + \epsilon, \tag{1}$$

where ϵ Gaussian distributed spatial noise $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and where z is a (zero-mean) Gaussian process (GP), identified by the associated covariance form Σ. Following the idea introduced in [9], we model $z(s)$ as the convolution of a white noise process $x(s) \sim \mathcal{GP}(0, \sigma_x^2 Id)$, with a given kernel function k. More specifically, the spatial process $z(s)$ is identified by a sparse set of control points defined in space and time, $\{w_j = (u_j, t_j))\}_{j=1}^{N_w}$, and associated parameters θ_j:

$$z(s) = \sum_{j=1}^{N_w} x(w_j) k(s - w_j | \theta_j). \tag{2}$$

Under these modelling assumptions, the generative model (1) assumes the form:

$$y = Kx + \epsilon, \tag{3}$$

where K is the matrix of the spatial coefficients associated to the control points $K_{s,w_j} = k(s - w_j | \theta_j)$. The image time series y is therefore a realization of the

following process:

$$y \sim \mathcal{GP}(0, \Sigma), \quad \text{with } \Sigma = \sigma_x^2 K K^T + \sigma_\epsilon^2 Id. \tag{4}$$

The model (4) is completely identified by the measurement noise σ_ϵ, by the white process parameter σ_x, and by the control points w_j with associated parameteres θ_j.

We note that the size of the covariance matrix Σ is $N_u N_t \times N_u N_t$, where N_u is the number of voxels, and N_t is the number of temporal observations. For this reason the naive approach to the modelling of (4) can easily lead to prohibitive problems in term of storage and computations. In the following section we show that the proposed kernel parameterization leads to computationally tractable inference schemes whose complexity depends on the number of basis functions solely.

2.1 Efficient Inference in Gaussian Process Convolution Models

The GP-based generative model with kernel structure outlined in this work provides a powerful and extremely flexible framework for prediction an inference in image time series. Let $\boldsymbol{\theta} = \{\sigma_x, \sigma_\epsilon, (\theta_j)_{j=1}^{N_w}\}$ be the set of parameters of the model (4). In the following sections we provide the main results concerning the marginal likelihood computation, the hyper-parameter optimization and the posterior prediction.

2.2 Log-marginal likelihood

The log-marginal likelihood of model (4) is:

$$\log \mathcal{L}(\boldsymbol{\theta}) = -\tfrac{n}{2}\log(2\pi) - \tfrac{1}{2}\log|\Sigma(\boldsymbol{\theta})| - \tfrac{1}{2}y^T \Sigma(\boldsymbol{\theta})^{-1}y. \tag{5}$$

In particular, the determinant and matrix inverse terms can be efficiently computed by using well known matrix algebra properties:

$$|\Sigma(\boldsymbol{\theta})| = |\sigma_x^2 K^T K + \sigma_\epsilon^2 Id_{N_w}| \tag{6}$$

$$\Sigma(\boldsymbol{\theta})^{-1} = \frac{1}{\sigma_\epsilon^2}Id_{N_u} - \frac{1}{\sigma_\epsilon^4}K\left(\frac{1}{\sigma_x^2}Id_{N_w} + \frac{1}{\sigma_x^2}K^T K\right)^{-1}K^T. \tag{7}$$

We note that in this form both inverse and determinant operations are performed on matrices of size N_w, which is magnitude smaller than $N_u N_t$.

2.3 Hyperparameter optimization

The derivative of the log-marginal likelihood (5) with respect to the model parameters $\boldsymbol{\theta}$ is:

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\log\mathcal{L} = -\tfrac{1}{2}Tr\left(\Sigma(\boldsymbol{\theta})^{-1}\frac{\mathrm{d}\Sigma(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}\right) - y^T \Sigma(\boldsymbol{\theta})^{-1}\frac{\mathrm{d}\Sigma(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}\Sigma(\boldsymbol{\theta})^{-1}y \tag{8}$$

It can be shown that formula (8) can be efficiently computed with respect to each model parameters. For instance, the gradient with respect to the noise parameter can be expressed in the form:

$$\frac{\mathrm{d}}{\mathrm{d}\sigma_\epsilon^2} \log \mathcal{L} = -\frac{1}{2} Tr\left(\Sigma(\boldsymbol{\theta})^{-1}\right) - y^T \Sigma(\boldsymbol{\theta})^{-2} y \tag{9}$$

$$= -\frac{\sigma_x^2}{2\sigma_\epsilon^2} \left(Tr(K^T K) - \frac{1}{\sigma_\epsilon^2} Tr(K^T K \left(\frac{1}{\sigma_x^2} Id_{N_w} + \frac{1}{\sigma_x^2} K^T K \right)^{-1} K^T K) \right) \tag{10}$$

$$+ \frac{\sigma_x^2}{2\sigma_\epsilon^2} A^T A, \tag{11}$$

where

$$A = \left(K - \frac{1}{\sigma_\epsilon^2} K \left(\frac{1}{\sigma_x^2} Id_{N_w} + \frac{1}{\sigma_x^2} K^T K \right)^{-1} K^T K \right)^{-T} y.$$

We note that, as for the computation of the marginal likelihood, the above term can be efficiently decomposed in the more convenient product of matrices of lower dimension, thus leading to computationally tractable solutions.

2.4 Prediction

The proposed generative model allows us to consider the predictive distributions of the latent spatio-temporal process at any testing locations u^* and timepoints t^*.

Given image time series $I(u,t)$, we now aim at predicting the image I^* at $N^* \times N_T^*$ testing coordinates $\{u^*, t^*\}$. Let us define $\Sigma_{I,I^*} = \Sigma(u, t, u^*, t^*)$ the cross-covariance matrix of training and testing data, and $\Sigma_{I^*,I^*} = \Sigma(u^*, t^*, u^*, t^*)$ the covariance evaluated on the new coordinates. The joint GP model of training and testing data is:

$$\begin{pmatrix} I(u,t) \\ I^*(u^*,t^*) \end{pmatrix} \sim \mathcal{N} \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma + \sigma^2 Id & \Sigma_{I,I^*} \\ \Sigma_{I^*,I} & \Sigma_{I^*,I^*} + \sigma^2 Id \end{pmatrix} \right], \tag{12}$$

and it can be easily shown that the posterior distribution of I^* conditioned on the observed time series I and parameters $\boldsymbol{\theta}$ is [7]:

$$I^*|I, \{u^*, t^*\}, \boldsymbol{\theta} \sim \mathcal{N}\left(\mu^*, \Sigma^* \right), \text{ where } \mu^* = \Sigma_{I,I^*} \Sigma^{-1} I$$
$$\text{and } \Sigma^* = \Sigma_{I^*,I^*} - \Sigma_{I,I^*} \Sigma^{-1} \Sigma_{I^*,I} + \sigma^2 Id . \tag{13}$$

3 Application: Longitudinal Brain Changes in Alzheimer's Disease

In this section we show an application of the proposed generative model to the analysis of the individual longitudinal brain changes observable in image time
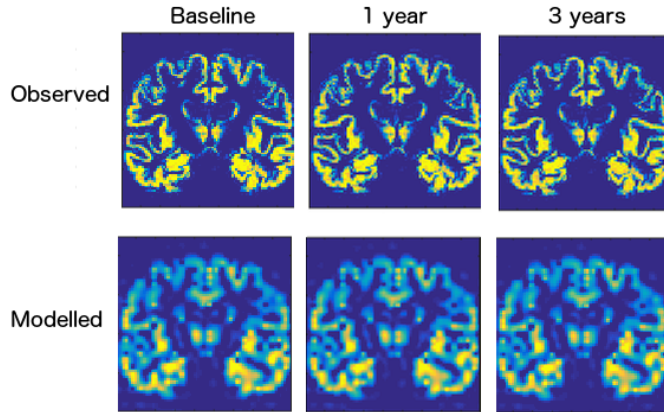
Fig. 1: Observed and modelled image time series by using a grid of 25x25 control points. The predicted progression provides a probabilistic description of the observed data at the grid resolution.

series. We consider the model outlined in Equation (3), with kernel function k associated to the control points $\{w_j\}_{j=1}^{N_w}$ identified by independent spatial and temporal length-scale parameters $\theta_j = \{\theta_j^u, \theta_j^t\}$:

$$k(s - w_j | \theta_j) = \exp(-|u - u_j|^2 / \theta_j^u) \exp(-|t - t_j|^2 / \theta_j^t). \qquad (14)$$

With the proposed parameterization, the spatio-temporal process (3) is completely characterized by the sparse set of spatial and temporal parameters associated to the set of control points. As we shall see in the following experiment, these parameters describe the spatial and temporal complexity of the underlying spatio-temporal signal, and thus they identify the non-stationary and non-separable model of the observed image time series.

3.1   Data analysis and Results.

We selected a patient affected by mild cognitive impairment for which 6 images were available, corresponding to observational time of respectively baseline, 6 months, 1, 1.5, 2 an 3 years.

The images were processed according to established procedures consisting of joint bias correction, tissue segmentation, alignment to the within-subject average anatomy, and non-linear normalization to a group-wise anatomical reference [10]. The final image size was of 100^3 cubic voxels with isotropic resolution of $1.5\,mm$.

Figure 1 shows an application of the proposed approach on the modeling of the coronal slice including temporal regions, by using a grid of 25x25 basis functions. We note that the predicted progression provides a description of the observed data at the grid resolution.

The fitted model parameters are shown in Figure 2, left. It is interesting to note that they provide a description of the spatial and temporal complexity of the observed time series. Indeed, the spatial complexity is higher (decreased spatial length-scale parameter) in the cortical areas, while the temporal complexity is higher in the temporal regions (decreased temporal length-scale). We also note that the model variability is zero outside the brain areas (Figure 2, right).
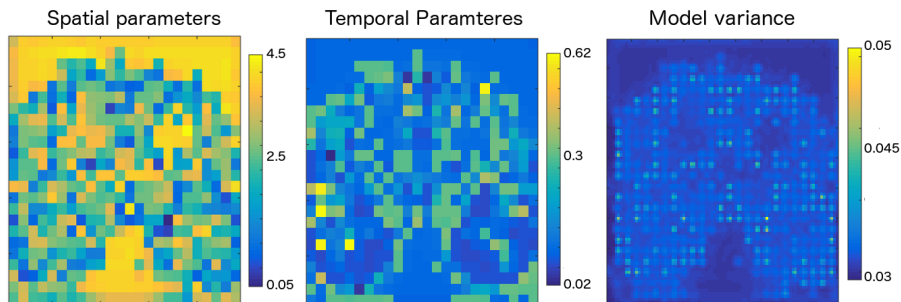

Fig. 2: Fitted model parameters and model variance. We note that the model is capable to adjust the parameters to the spatial and temporal complexity of the data. In particular, the spatial complexity is higher (decreased spatial length-scale parameter) in the cortical areas, while the temporal complexity is higher in the temporal regions (decreased temporal length-scale). We also note that the model variability is zero outside the brain areas.

The accuracy of the proposed approach in modelling the longitudinal changes is shown in Figure 3, where we show the average longitudinal changes measured in respectively temporal areas and thalami, two regions which are characterized by different temporal complexity (Figure 2): the temporal length-scale parameters of the temporal region are low (thus denoting high temporal complexity of this area), while the ones of the thalami are associated to higher length-scale (low temporal complexity). Indeed, the average progressions shown in Figure 3 show an almost constant progression for the thalami, while the temporal area has an accelerated atrophy process.

4 Conclusions

In this work we proposed a novel probabilistic approach to the modelling of non-stationary, non-separable spatio-temporal processes, by means of kernel convolutions of a white noise Gaussian process. The experimental results show that the proposed modelling method leads to an accurate fit to the observed image time series, while at the same time providing a rich description of the spatio-temporal dynamics of the data, encoded by the learned spatial and temporal parameters. Further extensions of the proposed work will aim at improving the computa-
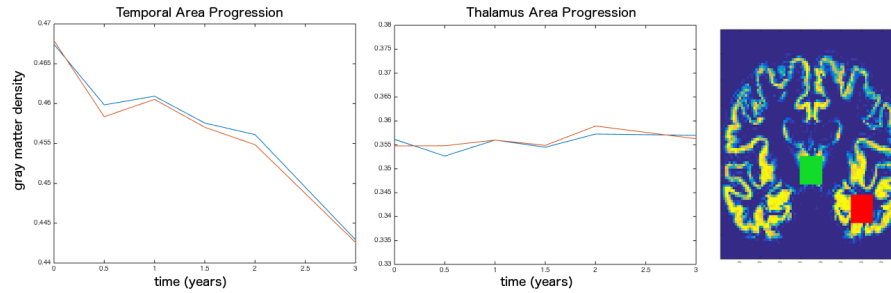
Fig. 3: Left. Modelled (red) and observed (blue) atrophy progressions. The model provides accurate fit, and shows that the temporal areas have higher temporal complexity than the thalami, caused by the process of atrophy acceleration. Right. Reference areas enclosing temporal region (red), and thalami (green).

tionally efficiency of the inferential process, in order to scale to the modelling of high-dimensional time series of 3D images with several time points.

References

1. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.C.: Population shape regression from random design data. International Journal of Computer Vision **90**(2) (2010) 255–266
2. Ashburner, J., Ridgway, G.: Symmetric diffeomorphic modeling of longitudinal structural MRI. Frontiers in Neuroscience **6**(197) (02 2013)
3. Niethammer, M., Huang, Y., Vialard, F.X.: Geodesic regression for image time-series. In: Proc. of Medical Image Processing and Computer Assisted Intervention (MICCAI). Volume 6892. (2011) 655–662
4. Lorenzi, M., Ayache, N., Frisoni, G.B., Pennec, X.: Mapping the effects of $A\beta$ 1-42 levels on the longitudinal changes in healthy aging: hierarchical modeling based on stationary velocity fields. In: Proc. of Medical Image Processing and Computer Assisted Intervention (MICCAI). Volume 6892. (2011) 663–670
5. Friston, K.J., Holmes, A., Worsley, K.J.: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping **2** (1995) 189–210
6. Ziegler, G., Ridgway, G.R., Dahnke, R., Gaser, C.: Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. NeuroImage **97** (April 2014) 333–348
7. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press (2005)
8. Lorenzi, M., Ziegler, G., Alexander, D.C., Ourselin, S.: Efficient Gaussian process-based modelling and prediction of image time series. In: Proc. of Information Processing in Medical Imaging (IPMI). (2015) to appear.
9. Higdon, D.: Space and space-time modeling using process convolutions. In: Quantitative methods for current environmental issues. Springer (2002) 37–56
10. Ashburner, J., Friston, K.: Unified segmentation. NeuroImage **26** (2005) 839–851