Evolutionary Inference for Functional Data: Using Gaussian Processes on Phylogenies to Study Shape Evolution

Jones, Nick S. and Moriarty, John

2011

ISSN 1749-9097

Evolutionary Inference for Functional Data: Using Gaussian Processes on Phylogenies to Study Shape Evolution

Nick S. Jones[1] and John Moriarty[2]

[1]Department of Physics, Oxford Centre for Integrative Systems Biology and CABDyN Complexity Centre,
Clarendon Laboratory, University of Oxford, Parks Road, Oxford. OX1 3PU
[2]School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK

This paper uses the interface between two disciplines—phylogenetics and functional data analysis—to aid the analysis of rich ancestral data like continuous curves. We place Gaussian processes on phylogenies in order to perform evolutionary inference on such functional data objects. Unlike morphological summaries, which reduce data dimension, this approach allows one to make inferential statements about curves themselves. We provide a modified covariance function that corrects for the relationships between states at different points on a phylogeny and discuss its use in inference. In general, this covariance is expressed as the solution of an integral equation and we note that, for a given Gaussian process, a set of solutions sufficient for all phylogenies may be precomputed as a library which, for stationary processes, is one dimensional. This work has relevance for those wanting to perform inference on functional data objects related by an evolutionary process; it also specifies a class of hierarchical clustering algorithms for functional data objects and can be used for multivariate time series forecasting.

Conventional methods for phylogenetic inference [21] take a set of symbolic sequences and attempt to infer a phylogeny which relates them. In this paper we consider the corresponding problem for a set of functions rather than a set of sequences. The sets of functions one might consider may be of diverse form and might represent bird songs, landscapes, zebra stripes or skull shapes. One might attempt to summarize each function by a symbolic sequence (e.g. the presence or absence of certain characters) [9, 28, 38] or by one or more continuous characters (e.g. continuous summary statistics such as the distances between landmarks) [6, 25], and existing methods for phylogenetic inference can then be employed. These methods obviously only give indirect access to the values of ancestral functions, since the map from data to summary is many-to-one. Alternatively, one could impose the restriction that all curves come from a certain parameterized class; while this assumption again allows the use of existing approaches, it will, in general, restrict the palette of curves available for modelling. In the following, we shall adopt a straightforward nonparametric approach. We represent the time evolution of any function on the spatial domain $\mathbb{R}^d$ by a function on the space-time domain $\mathbb{R}^{d+1}$. We assume that the data points observed constitute a sample from a Gaussian process (GP), which may then be specified uniquely by its mean and covariance functions on $\mathbb{R}^{d+1}$. The observations that we have about the function at one point in time and space will thus be encoded in its posterior distribution at another through the covariance function of the GP. We explain below how this GP on $\mathbb{R}^{d+1}$ can be extended to a GP on a phylogeny. This allows us to exploit the considerable existing machinery for inference with GP's [33]. The work presented here may be understood as a generalization of the work of Felsenstein and Martins and Hansen [6, 25] which uses GP's for the study of continuous characters. The methods described here are most directly appropriate for functional data representing signals, curves and patterns; more general shapes lie outside this paper's immediate purview. In the remainder of this introduction we will recall some relevant areas of morphometrics for phylogenetic inference, functional data analysis and GP's, and pattern matching.

Phylogenetic inference for continuous characters: There is an established relationship between the study of morphology and the practice of phylogenetic inference [17, 23] and this has direct bearing on questions in Anthropology, Paleontology and Evolution. In this setting, phylogenetic inference is typically mediated via real-valued morphometric phenotypes, or 'characters'—that is, reduced versions of the data acquired [4]. While discrete versions of characters are widely used e.g. [9, 23, 28, 38] there has also been investigation of continuous characters e.g. [6, 7, 11, 25]. Debate over whether continuous characters are the best way of reasoning about unobserved shapes continues to be lively [13, 22]. Felsenstein [6] presents a method for comparative studies of continuous real-valued phenotypes or characters. The key innovation in this statistical model is a correction for branching phylogenies: in previous studies, the assumption of independence had caused the overstatement of the significance in hypothesis tests. Felsenstein's method requires that the phylogeny is known and that the characters may be modelled by Brownian motion. Despite, in the author's words, the "considerable barriers to making practical use of this technique", his method has become widely used to infer correlations in character evolution. The method, based as it is on Brownian motion, is a GP model (for an implementation see the `corBrownian` function of the `ape` package of the R statistical language [6, 25]). We will see that it satisfies our definition of a phylogenetic GP, supplied in the next section; indeed, since the state is a single real value, it is perhaps the simplest nontrivial phylogenetic GP. Felsenstein discusses key practical issues for inference with his phylogenetic GP model, such as uncertainty over phylogenies, confounding, selection and drift, Gaussian modelling assumptions, non-stationarity, and punctuated evolution. Further, the author hints directly at the possibility of richer GP models but notes "the difficulty is that quantitative characters will evolve at different rates, and in a correlated fashion".

Martins and Hansen [25] extended Felsenstein's model, recasting it in the framework of generalized least squares (GLS). Their work makes clear that GLS models, which are also special cases of GPs, provide a unified approach capable of addressing a wide range of questions for single character evolution. In particular, the authors give simple linear point estimates both for ancestral states and for certain model parameters. The authors show that the GLS framework (and hence the phylogenetic GP framework) is rich enough to include covariance functions motivated by different evolutionary assumptions. These include random genetic drift, directional and stabilizing selection, and environmental fluctuations—in the words of the authors: "In essence, by applying different [covariance] matrices, we create a new phylogenetic comparative method for each situation".

Functional Data Analysis and GP's: Though the study of shape evolution is well established, the disciplinary emphasis has been on the use of a few morphometric characters [4, 23]. In statistics, both the parametric and non-parametric study of curves and contours have been lively, see Refs. [8, 32] for reviews and examples. As a classic example, Functional Data Analysis can be used to discriminate between people who have been asked to draw the same shape [32]. After work in Refs. [5, 35], nonparametric estimation of mean and covariance functions has developed into the area of Functional Principal Component Analysis (FPCA). This approach can be aided by assuming the underlying process to be Gaussian with covariance function drawn from a known class, since the distribution of a GP is uniquely specified by its mean surface and covariance function. The study of GP's is a corner-stone of the theory of stochastic processes [34] (example GP's include the Wiener and Ornstein-Uhlenbeck Processes) and recently GP's have been adopted by those working in statistical machine learning [33]. Before this recent activity they have seen practical use in Geostatistics under the name of kriging [36]. The literature on GP's defines a clear mathematical and computational framework for inference and it will be the integration of this approach within phylogenies that will enable evolutionary inference on functional data in this paper.

Pattern Matching and Evolving Shapes: The approach outlined here may also be viewed in the context of methods for clustering sets of functions: as an evolutionary hierarchical clustering algorithm. If a phylogeny can be inferred then distances between the different functions observed are defined by the the time to a common ancestor. In common with some other methods, the approach we consider is suitable for clustering sparsely or irregularly sampled data [16, 32]. How to register, or align, one functional data object with another is also a lively area of applied mathematics, geometry and computer vision [14, 37, 40, 43]; note that good alignment of sequence data is also an important problem in conventional sequence phylogenetics [21]. For the purposes of this paper, however, we assume that functions are available pre-aligned. As well as aligning and clustering functional data, some authors have considered evolving new shapes from given grammars: these often combine a shape alphabet/grammar with ideas from evolutionary/genetic computing and are used for design tasks [2, 3, 10, 19, 29]. It is possible that the methods outlined below could be run forwards in time to yield new forms.

The paper is organised as follows. Having introduced the reader to a few definitions we obtain the Phylogenetic covariance and discuss challenges and simplifications (Sections I and II). We then consider how these tools could be used to infer the nature of an evolutionary process, given observations and a known phylogeny (Section III A). Our next task is to explain how one can use knowledge of a process and a phylogeny to make predictions about past, future or missing data (Section III B). Having outlined how we might take a set of observed functions and infer an evolutionary tree, we present our discussion (Sections III C and IV).

I. DEFINITIONS

GP models indexed by *time* have found wide applicability, for example in time series analysis. A Gaussian *graphical model* is indexed by the vertices of a graph, and its covariance function is encoded by the presence or absence of edges, which represent conditional independences. In this paper we combine these two approaches, regarding a phylogeny T as having both linear structure and graph structure, and using T as the index set for a GP Y (in the following, all GPs will be understood to have mean 0). The covariance function of Y is then defined via both the linear and graph structure, as follows.

Clearly, any phylogenetic tree may be represented as a planar straight-line graph: the edge lengths may be chosen to represent evolutionary time, while the angles are not important and are chosen arbitrarily. Modulo this choice of angles, we therefore have a branched linear space, which we will call a *phylogeny.*

Definition I.1 (Phylogenetic GP)**.** *Given a rooted phylogenetic tree S, represented as a planar straight-line graph whose edge lengths represent evolutionary time, the corresponding* phylogeny *is the branched linear space T obtained from S by neglecting the choice of angles. A* Phylogenetic GP *on a phylogeny T is a GP indexed by T and a* phylogenetic covariance function *is the covariance function of a phylogenetic GP.*

Note that the branched linear space T described above is topologically equivalent to a loop-free *train track* [30]. We use the graph structure of the phylogeny to represent conditional independences in a manner which is, in general, different to that commonly used in graphical models. For $u, v \in T$, we denote

- the path between u and v by $q(u, v)$,
- the most recent common ancestor (MRCA) of u and v by $c(u, v)$, and
- $\{Y(w) : w \in q(v_0, c(u, v))\}$ by $A(u, v)$ (the *Ancestry of u and v*).

The most recent common ancestor $c(u, v)$ of example points u and v on a phylogeny can be seen in Fig. 1a), and example realisations of the states of u, v and $c(u, v)$ for a simple phylogenetic GP are given in Fig. 1b).

Assumption I.1 (Graph structure)**.** *Conditional on their ancestry $A(u, v)$, the states $Y(u)$ and $Y(v)$ are independent.*

The linear structure of the phylogenetic GP is specified by the choice of a single *marginal covariance* Σ on $\mathbb{R}$. Again for $u, v \in T$, denote

- the evolutionary time coordinate at u by t_u (taking $t_{v_0} = 0$),
- $u \prec v$ if u is a direct ancestor of v.

Assumption I.2 (Linear structure). *If $u \prec v$ then $\sigma(u, v) = \Sigma(t_u, t_v)$.*

If $u \prec v$ then the path $q(u, v)$ is naturally parameterised by the evolutionary time interval $[t_u, t_v] \subset [0, \infty)$, and in this way we may use Σ to specify a covariance function on 'rays' through T. We now use this parameterisation to show that there exists a unique phylogenetic covariance function satisfying assumptions I.1-I.2.

Construction of the phylogenetic GP. First note that if assumptions I.1-I.2 hold then, in equations (6)-(7) below, the terms $E[Y(u)|A(u, v)], E[Y(v)|A(u, v)]$ are specified by Assumption I.2 and so the phylogenetic covariance function is specified fully. Therefore, if a phylogenetic GP can be constructed to satisfy assumptions I.1-I.2 then it is unique. We may construct a realisation from such a GP Y by exploring the edges of T in a breadth-first search, and progressively generating the values of Y on each edge, as follows. First generate $y(v_0)$ according to its unconditional distribution. By induction, when the edge (v, w) (setting $v \prec w$) is first visited, the values $y(u) : u \in q(v_0, v)$ have already been generated; for each $u \in q(v, w)$ we may therefore set $y(u) = z(t_u)$, where z is a realisation of a GP Z on $[0, t_w]$, with covariance Σ, conditioned on $Z(t_u) = y(u)$ for each $u \in q(v_0, v)$ and independent of all other randomness. Assumption I.1 holds for the GP Y by construction, and assumption I.2 holds by the Law of Total Expectation.

It can be seen that under the above definitions, Felsenstein's model in [6] may be interpreted as a phylogenetic Brownian motion. In general we may consider phylogenetic GPs on the multidimensional branched space $T \times \mathbb{R}^d$ by assuming space-time separability—that is, that the covariance factorizes as a product of a covariance on the phylogeny T and a covariance on $\mathbb{R}^d$. Examples can be found in Fig. 1a,b). It is not difficult to show that space-time separability at the phylogenetic level is equivalent to space-time separability at the marginal level: the assumption of separability will therefore be justified when the process which governs the evolution of any single functional data object is space-time separable. Alternatively, we note that the formalism in this section can be extended to include multidimensional phylogenetic GPs that are not space-time separable.

II. PHYLOGENETIC COVARIANCE FUNCTION

Given the preceding assumptions and definitions we are now in a position to obtain the covariance function, $\sigma(u, v)$, between different points u, v on the phylogeny T. We begin by supposing that for each $\tau > 0$ there exists a kernel θ_τ on $(-\infty, \tau] \times [\tau, \infty)$ which makes the covariance Σ an eigenfunction with eigenvalue 1, in the sense that for each $t > \tau > s$ we have

$$\int_{-\infty}^{\tau} \theta_\tau(w, t)\Sigma(w, s)dw \;=\; \Sigma(t, s). \tag{1}$$

We now use the kernel θ_τ to explicitly construct the GP Y as described in Section I, and so obtain an expression for the covariance σ. Consider a GP Z on $(-\infty, \tau]$ with covariance Σ, and for $t > \tau$ define the Gaussian random variable Z_t^τ by

$$Z_t^\tau = \int_{-\infty}^{\tau} \theta_\tau(w,t) Z(w) dw. \tag{2}$$

Then Z_t^τ is the conditional expectation of $Z(t)$ given $\{Z(s) : s < \tau\}$, since for each $s < \tau$ we have by (1) and Fubini's Theorem

$$E[Z_t^\tau Z(s)] = \Sigma(t,s) = E[Z(t)Z(s)]. \tag{3}$$

Choose now $u, v \in T$, set $\tau = t_{c(u,v)}$, and extend the construction of Y given in section I as follows. When constructing Y on the path $(c(u,v), u)$, define

$$\begin{aligned} y(u) &= z(t_u) && (4)\\ y_u^\tau &= \int_{-\infty}^{\tau} \theta_\tau(w, t_u) z(w) dw. && (5) \end{aligned}$$

With this extended construction, we have $Y_u^\tau = E[Y(u)|A(u,v)]$ and $Y_v^\tau = E[Y(v)|A(u,v)]$. Using the fact that $\sigma(u,v) = E[Y(u)Y(v)]$ and the Law of Total Expectation we have

$$\sigma(u,v) = E[E[Y(u)Y(v)|A(u,v)]]. \tag{6}$$

By assumption I.1 we find

$$\begin{aligned} \sigma(u,v) &= E[E[Y(u)|A(u,v)]E[Y(v)|A(u,v)]] && (7)\\ &= E[Y_u^\tau Y_v^\tau]. && (8) \end{aligned}$$

It follows by Fubini's Theorem and (1) that

$$\begin{aligned} \sigma(u,v) &= \int_{w=-\infty}^{\tau} \int_{z=-\infty}^{\tau} \theta_\tau(w,t_u)\theta_\tau(z,t_v)\Sigma(w,z) dw dz && (9)\\ &= \int_{-\infty}^{\tau} \theta_\tau(w,t_u)\Sigma(t_v,w) dw && (10) \end{aligned}$$

(note that, since $t_v > \tau$, (1) cannot be used again in (10)). The phylogenetic covariance sought in this section is thus Equation (10), and in the remainder of this section we show how this equation simplifies in certain special cases.

Time-domain Markov processes. If Σ is the covariance of a Markov process then the edges of the phylogeny represent conditional independences exactly as the edges of a Gaussian graphical model, and $\sigma(u,v)$ also takes a particularly simple form, so it is interesting to recover this case. By equation 2.19 of [33], the kernel $\theta_\tau(w,t)$ defined in (1) takes the form

$$\theta_\tau(w,t) = \frac{\Sigma(t_u,\tau)}{\Sigma(\tau,\tau)} \delta(w - \tau)$$

where δ is the Dirac delta function. Equation (10) then simplifies to

$$\sigma(u,v) = \Sigma(t_u,\tau)\Sigma(\tau,\tau)^{-1}\Sigma(t_v,\tau), \tag{11}$$

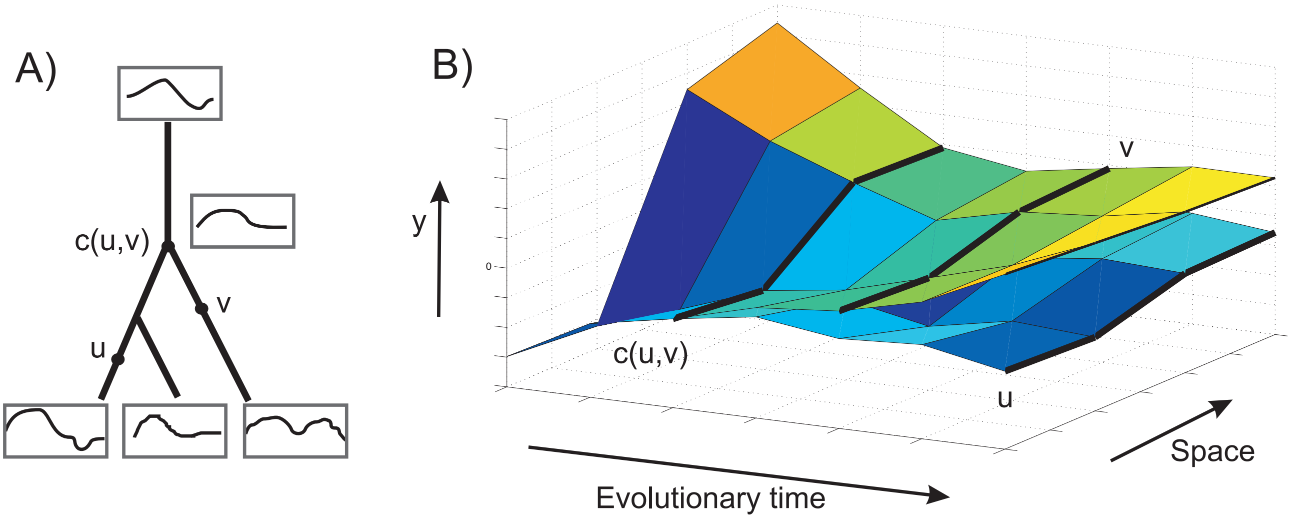

FIG. 1: A) A schematic of a phylogenetic GP. The most recent common ancestor of points u and v, $c(u, v)$, is given. B) A small sample from a Phylogenetic GP which is space-time separable: an Ornstein-Uhlenbeck process in time (7 samples) and with a squared exponential covariance in a 1-d space (4 samples). A notional forking event occurs at time indicated by the left-most solid black line labeled $c(u, v)$. For visualization purposes, after this time a linear trend is respectively added and subtracted from the two sets of points which lie on different rays.

confirming that in this case the phylogenetic covariance σ may be evaluated directly from the time-domain covariance Σ.

Stationary time-domain processes. If Σ is stationary in the sense that $\Sigma(t_u, t_v)$ is a function only of the distance $t_u - t_v$, then the solution of (1) simplifies. If there exists a kernel θ on $(-\infty, 0] \times [0, \infty)$ satisfying, for each $s < 0 < t$,

$$\int_{-\infty}^{0} \theta(w, t)\Sigma(w, s)dw \;=\; \Sigma(t, s) \tag{12}$$

then by stationarity this implies that for each $\tau > 0$, the kernel $\theta_\tau(w, t) := \theta(w - \tau, t - \tau)$ on $(-\infty, \tau] \times [\tau, \infty)$ satisfies (1) for each $s < \tau < t$. Equation (10) then becomes

$$\sigma(u, v) \;=\; \int_{-\infty}^{0} \theta(w, t_u - \tau)\Sigma(t_v - \tau, w)dw. \tag{13}$$

We note that, while the evaluation of the phylogenetic covariance σ for non-Markovian time-domain covariances may require the solution of the integral equations (1) by numerical methods, this is an established problem in numerical analysis [26, 31]. Further, for a stationary time-domain covariance Σ, the single kernel θ solving (12) is sufficient to calculate the phylogenetic covariance for *any* phylogeny.

III. INFERENCE TASKS USING THE PHYLOGENETIC COVARIANCE

Suppose now that we are given a set of observations $\{y(t, x) : (t, x) \in L\}$ (L being the set of co-ordinates of observation) from a GP with space-time separable covariance function K indexed by $T \times \mathbb{R}^d$, so that

$$K((t_1, x_1), (t_2, x_2)) = \sigma(t_1, t_2)k(x_1, x_2).$$

We consider the following three inference tasks, summarized in the table below: A) given the phylogeny T, to infer the evolutionary process as represented by the covariance K; B) given a known evolutionary process and phylogeny, to predict unobserved states $\{Y(t,x) : (t,x) \in M\}$ where $M \subset (T \times \mathbb{R}^d) \setminus L$, whether those states are ancestral, future, missing, or outside the sampled range; C) given knowledge only of the evolutionary process, to infer the phylogeny.

	Problem		
Data type	A) Parameter	B) Prediction	C) Phylogeny
Observed states, $y(L)$	Given	Given	Given
Phylogeny, T	Given	Given	Unknown
Covariance function, K	Unknown	Given	Given

A. Parameter Estimation

Given observations from functional data objects (which may or may not all be sampled at the same point in evolutionary time) and a known evolutionary history relating them, encoded by a phylogeny T, one may wish to make inferences about the evolutionary process that yielded the data. In the GP setting, this evolutionary process is encoded in the covariance matrix and there is a large literature on the estimation of covariance matrices. In this section we discuss maximum likelihood and Bayesian estimation of phylogenetic covariances drawn from a parametrized class, making a number of assumptions which are common in practice. Under these assumptions the analysis simplifies, and the effect of including branched phylogenies in GP regression may be seen more explicitly.

Given a phylogeny T and a time-domain covariance Σ, the phylogenetic covariance σ on T is specified by the construction in section II. Given also a space-domain covariance k, a separable GP with covariance $K = \sigma \cdot k$ may be defined on $T \times \mathbb{R}^d$. Many widely-used covariance functions for GP regression are smooth functions of a parameter vector θ, and we make this assumption in this section. We suppose further that data has been sampled discretely and systematically, in the sense that the set L of co-ordinates of observation has the product form $\mathbf{t} \times \mathbf{x}$, where $\mathbf{t} \subset T$ and $\mathbf{x} \subset \mathbb{R}^d$ are finite sets.

The maximum likelihood estimate of θ may be obtained as follows. Our sample $y = y(L)$ is a Gaussian vector with covariance matrix which we will denote $K_L(\theta)$; its log-likelihood given θ is

$$\log p(y|\theta) = -\frac{1}{2} y^T K_L^{-1} y - \frac{1}{2}\log(\det K_L). \tag{14}$$

This likelihood may be maximised by finding the zeroes of the function

$$\frac{\partial}{\partial \theta_j} \log p(y|\theta) = \frac{1}{2} tr\left((\alpha\alpha^T - K_L^{-1}) \frac{\partial K_L}{\partial \theta_j} \right), \quad \text{where } \alpha = K_L^{-1} y \tag{15}$$

([33], equation (5.9)). Writing $\otimes$ for the Kronecker matrix product, it follows from our assumptions of separability and systematic sampling that

$$K_L = \sigma_{\mathbf{t}} \otimes k_{\mathbf{x}} \tag{16}$$

and therefore

$$\begin{aligned} K_L^{-1} &= \sigma_{\mathbf{t}}^{-1} \otimes k_{\mathbf{x}}^{-1} & (17)\\ \frac{\partial K_L}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j}\sigma_{\mathbf{t}} \otimes k_{\mathbf{x}} + \sigma_{\mathbf{t}} \otimes \frac{\partial}{\partial \theta_j} k_{\mathbf{x}}. & (18) \end{aligned}$$

Given observations from functional data on $\mathbb{R}^d$ which arise from an evolutionary process, the phylogenetic covariance function may be regarded as correcting for branching in the shared evolutionary history. Equations (15)-(18) emphasise that, under the above assumptions, the computational cost of introducing this correction is relatively low. When estimating the phylogenetic covariance matrix, if the time-domain covariance has m unknown parameters then the added computational overhead that comes from taking a branched phylogeny is the inversion of $|\mathbf{t}| \times |\mathbf{t}|$ matrices in (17), and an additional m dimensions in the search space for the zeroes of (18). We note that separable covariance matrices such as (16) can in general impose certain identifiability constraints on the parameter vector θ; general discussion on computational issues for the estimation of covariance functions may be found in Ref. [33].

In the context of Bayesian parameter estimation, given a prior distribution for θ, one might instead attempt to maximize the posterior probability of θ given the observations y. This leads to a modified version of the right-hand side of (14), although the computational issues raised by the introduction of branched phylogenies are essentially unchanged.

In contrast to the theory of sequence evolution, the choice of particular parametrized classes for Σ and k may be, a priori, unclear. The problem of model selection for GP regression is discussed in [33], including exploratory data analysis and the formation of complex covariances from several different kinds of simple covariance function. In certain applications one may, however, wish to specify a priori that the marginal time-domain covariance Σ has the Markov property, which leads to the special class of Gauss-Markov processes (including, for example, Brownian motion and the Ornstein-Uhlenbeck process) and the simplified representation (11) for σ.

B. Prediction of unobserved states

Given the same set of observations $\{y(t,x) : (t,x) \in L\}$ from functional data objects related by the known phylogeny T, and assuming now that the covariance function K is known, one may alternatively wish to give a predictive distribution for a set of unobserved states $\{Y(t,x) : (t,x) \in M\}$, where for each $(t,x) \in M$, the element $t \in T$ may be at a past, present or future evolutionary time, and the element $x \in \mathbb{R}^d$ may be inside or outside the sampled range $\{x : (t,x) \in L\}$. We write K_M for the covariance matrix of the Gaussian vector $Y(M)$, and $K(M,L)$ for the covariance matrix between $Y(M)$ and $Y(L)$. The predictive formula for GP's from [33] then gives

$$Y(M)|y(L) \sim N(A,B)$$

where

$$\begin{aligned} A &= K(M,L)K_L^{-1}y(L), & (19)\\ B &= K_M - K(M,L)K_L^{-1}K(L,M). & (20)\end{aligned}$$

As noted above, these formulae simplify when L and/or M are product sets. Example applications of these predictive distributions for phylogenetic GPs are discussed below in section IV.

C. Phylogeny

For a given phylogeny T, a finite subset $L \subset T \times \mathbb{R}^d$ and a given marginal covariance Σ on $[0,\infty)$ and spatial covariance k on $\mathbb{R}^d$, one can calculate the (separable) covariance $K_L = K_L(T)$ of the Gaussian vector $Y(L)$ using Equation (10). A given set of observations $y = y(L)$ will then occur with probability density

$$p(y|T) = (2\pi)^{-\frac{|L|}{2}}(\det K_L)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y^T K_L^{-1} y\right). \tag{21}$$

One can maximize this likelihood by varying over the set of phylogenies; this is a hard but very standard problem within phylogenetics and is typically addressed with MCMC methods [21]. The phylogeny that maximizes this likelihood should inform about the evolutionary relationships between the observations $y(L)$. Note that the right-hand side of (21) is a function of T, since the phylogenetic covariance σ and hence the separable covariance K_L encode the evolutionary relationships between the co-ordinates of observation L. Since recalculating K_L for each new phylogeny T for a non-Markov marginal covariance Σ involves solving integral equations, it might seem that this will be a difficult task; however, as noted above, if the marginal covariance Σ is stationary then one can use a one-dimensional precomputed library of covariances to fully resolve K_L across the phylogeny T. We further note that in our setting of space-time separable covariances, the spatial part k of K_L is independent of the phylogeny, and further if L has product form then from (16) one need only recalculate the factor $\sigma_{\mathbf{t}}$ of K_L for each new T.

Beyond the above, there is a spectrum of more sophisticated approaches that could be considered. It might be the case that the parameters θ of k and Σ are unknown. The simplest approach would be to maximize $p(y|T, \theta)$ by varying over θ as well as T; as noted, this could be computationally challenging for non-Markov covariances. One might, alternatively, seek to identify the most probable phylogeny given only the data. To achieve this requires a choice of priors over possible T (given, for example, by the coalescent prior) and θ and an integration over θ [21].

IV. DISCUSSION

In this paper we have explained how the powerful inference architecture provided by GP's can in principle be used for evolutionary inference with functional data. Where the data objects are related by an evolutionary process with a given phylogeny, phylogenetic GP's perform regression using a covariance function which is modified to correct for correlations caused by shared evolutionary history. In the evolutionary setting we should therefore expect phylogenetic GP's to improve upon conventional GP inference for problems of parameter estimation and prediction. From a different viewpoint, phylogenetic GP's offer a highly flexible approach to evolutionary inference in that they model a rich palette of functional data objects, and offer straightforward approaches to model selection and prediction of unobserved data.

A new approach to the study of the evolution of curves offers to contribute to a number of areas in biology. We noted in the introduction how morphological information can aid phylogenetic inference in animal evolution; in other areas of biology, it may be the evolution of functional data that is important. Two examples are in disease progression and speech sound evolution. Physicians frequently collect functional data from their patients as diseases progress e.g. gait time series, speech records or heart rhythms. We might suppose that healthy patients start at the root of a tree and disease progression might act as a modulation of their functional data. Different disease variants would correspond to different branches away from a healthy initial state. Language evolution has already been studied from an evolutionary perspective using ideas from sequence phylogenetics [1, 12, 27]. It seems reasonable that functional data, in the form of speech sounds, could also evolve. In this case linguistics can equip us with known phylogenies [44] allowing us to address the parameter or prediction problems.

We now mention some further open and relevant areas. We explain, in Section III C, how one might attempt to infer a phylogeny given knowledge of functional data at a set of leaves. This is practically useful as it provides a method for a hierarchical clustering of functional data in which the position of the internal nodes contains useful information. When the functional data are signals (or time series, where the shape of the time series is the functional data object to be modelled) then the phylogenetic inference described above could be interpreted as an exercise in (evolutionary) network inference. The inference of networks of relatedness from time series data is a lively area of econometrics, neuroscience, and systems biology [39, 41, 42]. In systems biology, one often wants to pass from time series measurements of gene expression via mRNA concentrations to networks summarizing their relationships [24]. This approach would relate the mRNAs through a tree with unseen nodes. As well as its relevance for unseen ancestral functions, the prediction problem has other roles: rather than making statements about the past, a given phylogeny and parameterized covariance give predictive distributions for unobserved present states. As an example, if the functional data objects are time series and one assumes that the fitted covariance function applies in a neighbourhood around the observed signal data then our methods could also provide predictive distributions for data points lying outside the observed range. In this way, phylogenetic GP's allow multivariate time series forecasting from multiple series related by an evolutionary process, again corrected for evolutionary relatedness. We mentioned in our introduction that there is interest in methods of evolving new shapes for applications in design. In this case one evolves an observed present state forward in time; given knowledge of the parameters of a Markovian covariance this evolution is independent of other present states, and the phylogenetic perspective is thus, perhaps, uninteresting. However, for non-Markov processes, or Markov processes with unknown covariance functions, the forward evolution of any observed state is statistically dependent on all other observed states and on the phylogeny itself; this makes for an interesting class of forward shape evolutions which takes families of functions (and possibly phylogenies also) and generates new functions which are dependent on the whole of that family, in a way which is statistically consistent with the fitted phylogeny.

An area for further work is to extend the formalism we have presented from functional data evolution to the evolution of more general shapes.

Acknowledgements: We thank John Aston, Dorothy Buck and Vincent Macaulay for their valuable input

[1] Q.D. Atkinson, A. Meade, C. Venditti, S.J. Greenhill, M. Pagel (2008) Science, 319:588

[2] P. J. Bentley (1996) Generic Evolutionary Design of Solid Objects using a Genetic Algorithm. PhD thesis, Division of Computing and Control Systems, School of Engineering, The University of Huddersfield

[3] P. Bentley (ed) (1999) Evolutionary Design by Computers. Morgan Kaufmann, CA.

[4] J. Claude (2008) Morphometrics with R. Springer, Berlin

[5] P. E. Castro, W. H. Lawton, E. A. Sylvestre (1986) Principal modes of variation for processes with continuous sample curves. Technometrics. 28:329-337

[6] J. Felsenstein (1985) Phylogenies and the Comparative Method. The American Naturalist 125:1-15

[7] J. Felsenstein (1988) Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445-471

[8] F. Ferraty, P. Vieu (2006) Nonparametric Functional Data Analysis: Theory and Practice. Springer, NY

[9] W. L. Fink, M. L. Zelditch (1995) Phylogenetic analysis of ontogenetic shape transformations: a reassessment of the piranha genus Pygocentrus (Teleostei). Syst. Biol. 44:343-360

[10] P. Funes, J. Pollack (1998) Evolutionary body building: Adaptive physical designs for robots. Artificial Life 4:337-357

[11] A. Grafen (1989) The Phylogenetic Regression. Philosophical Transactions of the Royal Society Series B 326:119-157

[12] R. Gray, Q.D. Atkinson (2003) Language-tree divergence times support the Anatolian theory of Indo-European origin. Nature, 426:435

[13] B. E. Hendrixson, J. E. Bond (2009) Evaluating the efficacy of continuous quantitative characters for reconstructing the phylogeny of a morphologically homogeneous spider taxon (Araneae, Mygalomorphae,

Antrodiaetidae, Antrodiaetus). Molecular Phylogenetics and Evolution 53:300-313

[14] D.D. Holm, A. Trouvé, L. Younes (2009) The Euler-Poincaré theory of metamorphosis. Quart. Appl. Math. 67:661-685.

[15] G. S. Hornby, J. B. Pollack (2001) Evolving L-systems to generate virtual creatures. Computers and Graphics 25:1041-1048

[16] G.M. James, C.A. Sugar (2003) Clustering for Sparsely Sampled Functional Data. Journal of the American Statistical Association 98:397-408

[17] R. A. Jenner (2006) Challenging received wisdoms: Some contributions of the new microscopy to the new animal phylogeny. Integrative and Comparative Biology 46:93-103

[18] D.G. Kendall (1984) Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. Bulletin of the London Mathematical Society 16:81-121

[19] R. Kicinger, T. Arciszewski, K. De Jong (2005) Evolutionary computation and structural design: A survey of the state-of-the-art. Computers and Structures 83:1943-1978

[20] S. L. Lauritzen (1996) Graphical Models. Clarendon Press, Oxford

[21] P. Lemey, M. Salemi, A.-M. Vandamme (eds) (2009) The Phylogenetic Handbook (2nd ed.). Cambridge University Press, Cambridge

[22] M. Macholan (2008) The mouse skull as a source of morphometric data for phylogeny inference. Zoologischer Anzeiger 247:315-327

[23] N. MacLeod, P.L. Forey (eds) (2002) Morphology, shape and phylogeny. Taylor and Francis, London.

[24] D. Marbach, R.J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, G. Stolovitzky (2010) Revealing strengths and weaknesses of methods for gene network inference. PNAS early edition

[25] E.P. Martins, T.F. Hansen (1997) Phylogenies and the Comparative Method: A General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific Data. The American Naturalist 149:646-667

[26] M. Masujima (2009) Applied mathematical methods in theoretical physics. Wiley-VCH, Weinheim

[27] L. Nakhleh, T. Warnow, D. Ringe, S.N. Evans (2005) A Comparison of Phylogenetic Reconstruction Methods on an IE Dataset. Transactions of the Philological Society, 3:171-192

[28] M.J. Novacek (1992) Fossils, topologies, missing data, and the higher-level phylogeny of eutherian mammals. Syst. Biol. 41:58-73

[29] M. O'Neill et al (2009) Shape grammars and grammatical evolution for evolutionary design. Proceedings of the 11th Annual conference on Genetic and evolutionary computation table of contents Montreal, Québec

[30] Penner, R. C., with Harer, J. L. (1992). Combinatorics of Train Tracks. Princeton University Press, Annals of Mathematics Studies.

[31] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery (2007) Numerical recipes: the art of scientific computing. Cambridge University Press, Cambridge

[32] J. O. Ramsay, B. W. Silverman (2005) Functional Data Analysis. (2nd ed.). Springer, Berlin.

[33] C. Rasmussen, C. Williams, (2006) Gaussian Processes for Machine Learning. MIT, MA

[34] D. Revuz, M. Yor (1998) Continuous Martingales and Brownian Motion. Springer, Berlin.

[35] J. A. Rice and B. W. Silverman (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society: Series B 53:233-243

[36] M.L. Stein (1999) Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York

[37] M. Vaillant, J. Glaunes (2005) Surface matching via currents. Proc. Conf. IPMI 381-392

[38] J.J. Wiens (2001) Character analysis in morphological phylogenetics: problems and solutions. Syst. Biol. 50:689-699

[39] M. Winterhalder et al. (2005) Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems. Signal Processing 85:21372160

[40] L. Younes, F. Arrate, M.I. Miller (2009) Evolutions equations in computational anatomy. NeuroImage 45:S40-S50

[41] Causality Workbench `http://www.causality.inf.ethz.ch/`

[42] Dialogue for Reverse Engineering Assessments and Methods, `http://wiki.c2b2.columbia.edu/dream/index.php/The_DREAM_Project`

[43] http://iris.usc.edu/Vision-Notes/bibliography/contents.html 12.3.3 Surface Matching, Deformable Surface Matching

[44] http://www.cs.rice.edu/ nakhleh/CPHL/