# Nonparametric Functional Concurrent Regression Models

Arnab Maity*

**Article Type:**

Advanced Review

**Abstract**

Function-on-function regression refers to the situation where both independent and dependent variables in a regression model are of functional nature. Functional concurrent regression is a specific type of function-on-function regression that relates the response function at a specific point to the covariate value at that point and the point itself. Standard functional concurrent models are linear (a linear combination of the covariates is used), and often criticized due to their linearity assumption and lack of flexibility. This gives rise to nonparametric functional concurrent regression that models the response function at a specific point using a multivariate nonparametric function of both the point and the covariate value at that point. Such models allow for much more flexibility and predictive accuracy, especially when the underlying relationship is nonlinear. In the past decade, several methods have been proposed to perform estimation, prediction and inference in the nonparametric concurrent models using various methods such as spline smoothing, Gaussian process regression and local polynomial kernel regression. Such models have been shown to be useful tools in functional regression as well as stepping stone for further development.

*Department of Statistics, North Carolina State University

# INTRODUCTION

Functional data have become more and more common in recent years and as a result various statistical models and techniques are being developed to address problems involving such data type. In regression problems where both response and covariate(s) are of functional nature, an interesting and popular modeling strategy is to use concurrent models. Specifically, the response at a specific point is modeled as a function of the value of the covariate only at that specific point. The standard linear concurrent model assumes a linear relationship between the response and covariate functions at any given point of observation. Such models, also known as varying coefficient models, are useful in many settings such as longitudinal data analysis, and as such have a rich literature spanning over two decades since its introduction for the case with scalar response variable[4] and later for longitudinal data[2,3,15]; readers are referred to the article by Fan and Zhang[9] for a detailed review of such models. There are several methods to estimate the model components in longitudinal setting including local polynomial kernel smoothing[3,8,15], regression and smoothing spline methods[7,16,17], and penalized spline and quadratic inference function based methods[25]. In the functional data analysis setting, Ramsay and Silverman[27] provide a detailed exposition of the linear functional concurrent model and its fitting procedures. There have been many subsequent developments in this area including extension to spatial imaging[48], ridge regression[14] and extension to accommodate generalized response[1,11].

A drawback of the linear concurrent models is the assumption of linear relationship between the response and covariate functions. Nonparametric concurrent models seek to address this issue by specifying a nonparametric relationship that offer more flexibility in modeling and prediction. Various methods for estimation and inference in nonparametric functional concurrent regression have been developed using spline smoothing, Gaussian process regression and local kernel smoothing techniques. This article is a survey of these methods. In what follows, we discuss the three types of methods and their extensions, mention currently available software, provide overall comparison of the procedures followed by a numerical illustration using the Gait data, and finally present some concluding remarks.

## NONPARAMETRIC CONCURRENT MODELS

In what follows, we use $Y(\cdot)$ to denote the response variable and $X_1(\cdot), \ldots, X_Q(\cdot)$ to denote the functional covariates. We denote any other (vector-valued) covariates present in the model by $Z_i$. We assume that $Y(\cdot)$ and $X_q(\cdot), q = 1, \ldots, Q$ are defined on the same domain $[0, T]$ for some $T > 0$. We observe data tuples $[Y_i(t), X_{i1}(t), \ldots, X_{iQ}(t)]$, for $i = 1, \ldots, n$, for a finite number of points, $t \in \{t_{i1}, \ldots, t_{im_i}\}$, where $m_i$ is the number of observations for the $i$-th subject. The observational points may be different for each $i$, and may be irregularly spaced in practice. The covariates can also be contaminated with noise, that is, instead of $X_q(\cdot)$ we only observe $W_{ijq} = X_{iq}(t_{ij}) + e_{ijq}$ for $i = 1, \ldots, n$; $j = 1, \ldots, m_i$, and $q = 1, \ldots, Q$, where $e_{ijq}$ white noise with mean zero and variance $\sigma_e^2$. For simplicity of presentation, we shall first assume that $\sigma_e^2 = 0$, that is, we observe error free covariates and that $t_{ij} = t_j$ and $m_i = m$. We shall discuss the more general case as we present each method in subsequent sections. We assume that $Y(\cdot)$ and $X_q(\cdot), q = 1, \ldots, Q$ are smooth processes. We shall discuss the extension to the case with general response functions in a later section.

### Spline smoothing based methods

The general form of a nonlinear concurrent model with only one covariate $X(\cdot)$ is $Y_i(t) = F\{X_i(t), t\} + \epsilon_i(t)$, where $F(\cdot, \cdot)$ is a smooth unknown bivariate function and $\epsilon_i(\cdot)$ is an error process. The linear concurrent model is a special case of this general concurrent model if one takes $F\{X(t), t\} = \mu(t) + X(t)\beta(t)$. A more general additive model with multiple functional covariates is

$$Y_i(t) = \mu(t) + \sum_{q=1}^{Q} F_q\{X_{iq}(t), t\} + \epsilon_i(t),$$

where $\mu(\cdot)$ is an unknown intercept function, and $F_q(\cdot, \cdot), q = 1, \ldots, Q$ are unknown smooth functions. Such models are considered by Kim et al[19] and by Scheipl, Gertheiss and Greven[31] as a part of a more general framework.

For simplicity, we review the case of one covariate. The same formulation can be applied for multiple covariate case. The bivariate function $F(\cdot, \cdot)$ is modeled using tensor product of

B-splines as basis functions[23]: for any fixed $x$ and $t$,

$$F(x,t) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_t} \theta_{k,l} B_{X,k}(x) B_{T,l}(t), \tag{1}$$

where $\{B_{X,k}(x) : k = 1, 2, \ldots, K_x\}$ and $\{B_{T,l}(t) : l, 2, \ldots, K_t\}$ are B-spline basis functions, and $\theta_{k,l}$ are unknown parameters. The estimation then proceeds by minimizing the penalized sum of squares $\mathcal{L}(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \{Y_i(t_j) - \mathbf{Z}_i(t_j)\Theta\}^2/m + \Theta^T P \Theta$, where $\mathbf{Z}_i(t) = [B_{X,k}\{X_i(t)\} B_{T,l}(t)]_{k=1,\ldots,K_x}^{l=1,\ldots,K_t}$ is the full design matrix derived using the expression of $F(x,t)$, $\Theta = [\theta_{1,1}, \ldots, \theta_{1,K_t}, \ldots, \theta_{K_x,1}, \ldots, \theta_{K_x,K_t}]^T$ is the parameter vector, and $P$ is an appropriate penalty matrix. Kim et al used a penalty of the form $\lambda_x P_x^T P_x + \lambda_t P_t^T P_t$, where $P_x = D_x \bigotimes I_{K_t}$ and $P_t = I_{K_x} \bigotimes D_t$ are row and column penalties, and the parameters $\lambda_x$ and $\lambda_t$ control the smoothness in the directions of $x$ and $t$, respectively[6,23,24]. Here $D_x$ and $D_t$ are matrices representing the row and column second order difference penalties, respectively.

Minimization of $\mathcal{L}(\Theta)$ is straightforward and a closed form expression of the estimator is[19] $\widehat{\Theta} = (\sum_{i=1}^{n} \mathcal{Z}_i^T \mathcal{Z}_i + P)^{-1} \sum_{i=1}^{n} \mathcal{Z}_i^T \mathbf{Y}_i$, where $\mathbf{Y}_i = [Y_i(t_1), \ldots, Y_i(t_m)]^T$ and $\mathcal{Z}_i$ is a matrix such that $j$-th row of $\mathcal{Z}_i$ is $\mathbf{Z}_i(t_j)$. The penalty parameters $\lambda_x$ and $\lambda_t$ are typically chosen by well-known criteria such as generalized cross validation (GCV), restricted maximum likelihood (REML), and maximum likelihood (ML)[30,41–44]. The function $F(\cdot,\cdot)$ can now be constructed by plugging-in $\widehat{\Theta}$ in the basis expansion in (1).

An important issue in nonparametric concurrent models with continuous response is to account for covariance that may be present in the unobserved error process. Accounting for such covariance allows one to correctly construct standard errors for estimated model components and prediction of a new response. Kim et al assume that the error $\epsilon(\cdot)$ has the form $r(\cdot) + \eta(\cdot)$, where $r(\cdot)$ is a smooth process and $\eta(\cdot)$ is white noise. They develop a functional principal component analysis (FPCA)[5,46,47] based approach to estimate the covariance among the errors and subsequently use the estimated covariance to form corrected estimation and prediction errors.

When the sampling design is irregular or sparse, that is, $X(\cdot)$ is only observed at $s_{ij}, j = 1, \ldots, m_i$ and the response functions are observed at $t_{ik}, k = 1, \ldots, m_{Y,i}$, Kim et al propose to apply FPCA on the functional covariate to obtain smooth versions $\widehat{X}(\cdot)$ evaluated at $t_{ik}, k = 1, \ldots, m_{Y,i}$, and use the modified penalized sum of squares $\mathcal{L}(\Theta) = \sum_{i=1}^{n} \sum_{k=1}^{m_{Y,i}} \{Y_i(t_{ik}) -$

$\mathbf{Z}_i(t_{ik})\Theta\}^2/m + \Theta^T P\Theta$. The expressions for estimator of $\Theta$, prediction and their variances are also adjusted accordingly. Such techniques (especially the usage of FPCA) can account for the situation where the covariates are observed with measurement error as well. We refer the readers to Kim et al[19] for a much detailed discussion of these issues and the corresponding methodology.

## Gaussian process regression based methods

Gaussian process regression[39,40] provides a nice tool to accommodate large number of covariates into a regression model that would otherwise be over-saturated and unfeasible to fit. We start by providing a brief overview of Gaussian process regression framework. A Gaussian process can be seen as a collection of random variables where any finite number of them has a multivariate normal distribution[29,40]. We denote such a process by $\mathcal{GP}\{m(\cdot), C(\cdot,\cdot)\}$, where $m(\cdot)$ and $C(\cdot,\cdot)$ are the mean and covariance functions, respectively. A realization $h(\cdot)$ from such a process is a random function with $E\{h(x)\} = m(x)$ and $cov\{h(x), h(x')\} = C(x, x')$, where $x$ and $x'$ are points in the domain of the process. Also, for any finite set of points $x_1, \ldots, x_n$, the vector $[h(x_1), \ldots, h(x_n)]^T$ follows a multivariate normal distribution.

Now consider a simple nonparametric model

$$Y_i = h(x_i) + \epsilon_i,$$

where $Y_i$'s are scalar responses, $x_i$'s are vector-valued covariates, $\epsilon_i$'s are independent random errors from a normal distribution with mean zero and finite variance $\sigma^2$, and $h(\cdot)$ is a nonlinear unknown function. Gaussian process regression starts with the assumption that the unknown function $h(\cdot)$ is a realization from a zero mean Gaussian process[34], that is, $h(\cdot) \sim \mathcal{GP}\{0, C(\cdot,\cdot)\}$, where $C(\cdot,\cdot)$ is a covariance 'kernel', typically known up to a set of parameters. Thus combined with the assumption that the errors $\epsilon_i \sim N(0, \sigma^2)$ and are independent, one can write $(Y_1, \ldots, Y_n)^T \sim N(0, \mathbf{C} + \sigma^2\mathbf{I}_n)$, where $\mathbf{C}$ is an $n \times n$ matrix with $\mathbf{C}_{ij} = C(x_i, x_j)$.

In practice, there are many choices for the kernel function such as, polynomial kernel[40] $C(x, x') = (a + bx^T x')^d$, where $a, b > 0$ are constants and $d$ is an positive integer; Gaussian kernel[40] $C(x, x') = exp(-\gamma||x - x'||^2), \gamma > 0$; spline[13] and B-spline kernels[10], among many others[22,40]. Shi, Murray-Smith and Titterington[34] suggest to use

$$C(x, x') = \nu_0\, exp\left(-\frac{1}{2}\sum_{q=1}^{Q} w_q(x_q - x'_q)^2\right) + a_0 + a_1 x^T x',$$

where $x = (x_1, \ldots, x_Q)^T$ and $x'$ is defined similarly.

One could adopt several approaches for estimation. A Bayesian approach is to specify priors for the parameters and then make inference based on the posterior distribution. A frequentist approach is to consider the parameters ($\sigma^2$ and the parameters in $C(\cdot, \cdot)$) as variance components and use maximum likelihood[34] or restricted maximum likelihood[20] for estimation. Prediction of $Y_{\text{new}}$ for a new covariate value $X_{\text{new}}$ can be conducted by first observing that $(Y_1, \ldots, Y_n, Y_{\text{new}})^T$ follows a multivariate normal distribution, and then computing conditional distribution of $Y_{\text{new}}$ given $(Y_1, \ldots, Y_n)^T$, where one can plug-in the estimated parameters.

It is interesting to note that the Gaussian process regression framework has a strong connection to Reproducing Kernel Hilbert Space (RKHS). This mainly follows from the fact that every mean zero Gaussian process is defined by some RKHS, and the covariance kernel of the process can be identified with the reproducing kernel of the corresponding RKHS[33].

Shi, Murray-Smith and Titterington[34] develop a Gaussian process regression framework for nonparametric concurrent models of the form

$$Y_i(t_j) = F_i\{X_{i1}(t_j), \ldots, X_{iQ}(t_j)\} + \epsilon_i(t_j),$$

where they assume

$$F_i(\cdot) \sim \sum_{\ell=1}^{L} \pi_\ell GP\{0, C(\cdot, \cdot; \theta_\ell)\}.$$

Here $C(\cdot, \cdot)$ is a pre-specified kernel/covariance function depending on parameter $\theta_\ell$, possibly taking different values for each component of the mixture, and $\pi_\ell$ is the weight corresponding to the $\ell$-th component. A Bayesian methodology is developed by putting prior distributions on $\theta_\ell$ and $\pi_\ell, \ell = 1, \ldots, L$, and then using a combination of Gibbs sampling and MCMC procedure to draw samples from the resulting posterior distribution. A method for prediction of new response curves is also provided based on the posterior density of $F_i(\cdot)$.

Shi et al[37] extend the nonparametric concurrent model to include a linear mean component by using the model

$$Y_i(t) = Z_i^T \beta(t) + F_i\{X_{i1}(t), \ldots, X_{iQ}(t)\} + \epsilon_i(t),$$

where $\beta(\cdot)$ is a vector of unknown coefficient functions, and $F_i(\cdot)$ is defined as before. The model components are estimated using a two stage approach. In the first stage, each observed response curve is smoothed using B-spline basis functions: $Y_i(t) = \Phi^T(t)A_i$, where $\Phi(t) = [\Phi_1(t), \ldots, \Phi_{K_Y}(t)]^T$ are B-spline basis functions, and $A_i$ is a $K_Y \times 1$ vector of coefficients. These coefficients are estimated by minimizing $\int (Y_i(t) - \Phi^T(t)A_i)^2\, dt$ with respect to $A_i$ for each $i = 1, \ldots, n$. The unknown coefficient function $\beta(\cdot)$ is also modeled using B-spline basis functions: $\beta(t) = \Phi^T(t)B$, where $B$ is a matrix of unknown coefficients. $B$ is estimated as $(Z^TZ)^{-1}Z^TA$, where $Z = [Z_1, \ldots, Z_n]^T$ and similarly for $A$. In the second stage, the residuals from the first stage are computed as $\widetilde{F}_i(X_i(t)) = Y_i(t) - Z_i^T\widehat{\beta}(t)$. Then, based on the model $\widetilde{F}_i(X_i(t)) = F_i(X_i(t)) + \epsilon_i(t)$, the variance components involved in $F_i(\cdot)$ are estimated either by MLE or a Bayesian approach[34].

The number of basis functions, $K_Y$, controls smoothness of $\beta(\cdot)$. Later, Shi and Choi[33] discuss a general procedure to include a second derivative based roughness penalty terms in the first stage to estimate $A_i$ as well as $\beta(\cdot)$. They propose an iterative procedure that accounts for the covariance structure imposed by the Gaussian process regression framework into the estimation of $A_i$ as well. The approach of Shi et al[37] essentially removes these roughness penalty terms as well as the covariance structure in estimation of $A_i$ and $\beta(\cdot)$ for faster computation.

Other notable works in this framework include extension to the case where data are spatially indexed[36]; incorporating mixed effects of functional variable in the mean model[33,35], and extension to the case with generalized response from exponential family[38].

It should be noted that the two stage approach[37] depends on individual smoothing of each response curves, and as such may face difficulties if the curves are observed using a sparse design.

## Kernel smoothing based methods

Jiang and Wang[18] develop a functional single index model that uses concurrent relationship between a functional response and multiple functional covariates. Specifically, the functional single index model is given by

$$Y(t) = F\{\beta^T X(t), t\} + \epsilon\{X(t), t\},$$

where $X(t) = [X_1(t), \ldots, X_Q(t)]^T$, $F(\cdot,\cdot)$ is an unknown bivariate function, $\beta$ is an unknown coefficient vector, and $\epsilon$ is an error process possibly depending on the covariates. Jiang and Wang extend the standard conditional minimum average variance estimation (MAVE) methodology[45] based on local kernel smoothing to the case of functional single index model.

Denote $Y_{ij} = Y(t_{ij})$ and similarly for $X_{ij}$ for any $i, j$. Given any $t_{j\ell}$ and covariate value $X_{j\ell} = X(t_{j\ell})$, one can expand $F(\beta^T X_{ik}, t_{ik}) = E(Y_{ik}|t_{ij}, X_{ik})$ around $(\beta^T X_{j\ell}, t_{j\ell})$ using a locally linear approximation

$$F(\beta^T X_{ik}, t_{ik}) = a_{j\ell} + b_{j\ell}(t_{ik} - t_{j\ell}) + d_{j\ell}\beta^T(X_{ik} - X_{j\ell}),$$

where $a_{j\ell}, b_{j\ell}$ and $d_{j\ell}$ are quantities that depend on $F(\cdot,\cdot)$ and its first derivatives, respectively, evaluated at $(\beta^T X_{j\ell}, t_{j\ell})$. The conditional variance $E\{Y_{ik} - F(\beta^T Z_{ik}, t_{ik})\}^2$ (at $t_{j\ell}$ and $\beta^T X_{j\ell}$) can be approximated as

$$\sigma^2(\beta^T X_{j\ell}, t_{j\ell}) = \sum_{i=1}^{n}\sum_{k=1}^{m_i}[Y_{ik} - \{a_{j\ell} + b_{j\ell}(t_{ik} - t_{j\ell}) + d_{j\ell}\beta^T(X_{ik} - X_{j\ell})\}]^2 w_{ikj\ell},$$

where

$$w_{ikj\ell} = \frac{K[(t_{ik} - t_{j\ell})/h_t, \{\beta^T(X_{ik} - X_{j\ell})\}/h_x]}{\sum_{i=1}^{n}\sum_{k=1}^{m_i} K[(t_{ik} - t_{j\ell})/h_t, \{\beta^T(X_{ik} - X_{j\ell})\}/h_x]}$$

are kernel based weights, and $h_t$ and $h_x$ are corresponding bandwidths. The MAVE procedure then proceeds to estimate $\beta$ by solving

$$argmin_{a,b,d,\beta} \sum_{j=1}^{n}\sum_{\ell=1}^{m_j}\sum_{i=1}^{n}\sum_{k=1}^{m_i}[Y_{ik} - \{a_{j\ell} + b_{j\ell}(t_{ik} - t_{j\ell}) + d_{j\ell}\beta^T(X_{ik} - X_{j\ell})\}]^2 w_{ikj\ell}.$$

The minimization can be done in an iterative manner until convergence is reached[18]. The bandwidths for the kernel estimator is chosen via $m$-fold cross-validation.

Jiang and Wang show that, under appropriate regularity conditions, the parametric component $\beta$ has an asymptotic normal distribution with root-$n$ convergence rate. For the estimation of the bivariate function $F(\cdot,\cdot)$, two results are presented. It is assumed that the design points $t$ are sampled randomly from some distribution, that is, the design is random. Jiang and Wang show that when only one functional covariate is present, that is, $E[Y(t)|X(t)] = F\{X(t), t\}$, then

$$\sqrt{n\bar{N}h_t h_x}\{\widehat{F}(x,t) - F(x,t)\} \to^d N\{\eta(x,t), \Sigma_F(x,t)\}$$

for any scalars $t$ and $x$. Here $\bar{N} = \sum_i m_i/n$, $\eta(x,t)$ is the asymptotic bias term (shown to be dependent on $h_x/h_t$ and on second derivatives of $F(\cdot,\cdot)$) and $\Sigma_F$ is the asymptotic variance-covariance matrix (shown to be dependent on the kernel function as well as the joint density of $X$ and $t$). This result is of particular interest since it applies exactly to the nonlinear nonparametric concurrent model with one functional covariate. Thus this result provides a nice asymptotic property of the kernel smoothing estimator for a single covariate nonparametric concurrent model. A similar second result concerning the estimation of $F(\cdot,\cdot)$ for the general case with multiple covariates with estimated $\beta$ is also presented.

# GENERALIZED CONCURRENT MODELS

In the Gaussian process regression framework, Wang and Shi[38] develop a nonparametric concurrent modeling approach for non-Gaussian response. They assume that for each $t$, the response $Y_i(t)$ has a distribution from an exponential family with density function

$$f\{y_i(t)|\alpha_i(t), \phi_i(t)\} = \exp\left[\frac{y_i(t)\alpha_i(t) - b\{\alpha_i(t)\}}{a\{\phi_i(t)\}} + c\{y_i(t), \phi_i(t)\}\right],$$

where $\alpha(\cdot)$ and $\phi(\cdot)$ are the canonical and dispersion parameters, respectively. Wang and Shi then define the nonparametric concurrent regression model, conditional on the covariates,

$$E\{Y_i(t)\} = h[\mu_i(t) + F_i\{X_i(t)\}] \quad \text{with} \quad F_i\{X_i(t)\} \sim GP\{0, C(\cdot,\cdot;\theta_i)\},$$

where $\mu_i(\cdot)$ is an unknown function that can depend of other vector valued covariates $Z_i$ (e.g., $Z_i^T\beta(\cdot)$), and $h(\cdot)$ is a known link function. The model proposed in Shi et al[37] is indeed a special case of this model. For the case $\mu_i(t) = Z_i^T\beta(t)$, Wang and Shi propose to model $\beta(\cdot)$ using spline basis functions, and use an empirical Bayes learning approach to estimate $\theta_i$ and the unknown coefficients in the expression of $\beta(\cdot)$. Theoretical results about consistency of the predicted process $\widehat{Y}(\cdot)$ to the true process $Y(\cdot)$ are also presented.

Recently, Scheipl, Gertheiss and Greven[31] develop a generalized additive modeling framework for non-Gaussian functional response that also includes nonparametric concurrent models. They consider the model

$$Y_i(t) \sim \mathcal{F}\{\mu_i(t), \nu\} \quad \text{with} \quad g\{\mu_i(t)\} = \eta_i(t) = \sum_{r=1}^{R} f_r(\mathcal{X}_{ri}, t),$$

where $\mathcal{X}_r$ is a set of predictor that can include part of the functional covariates $X(\cdot)$, $\mathcal{F}$ is some distribution with conditional mean $E[Y_i(t)|\mathcal{X}_i, t, \nu] = \mu_i(t)$. This model is indeed very general, see their Table 1 for different possible choices of $f_r(\cdot)$ that give rise to various functional regression models. While Scheipl, Gertheiss and Greven did not specifically investigate nonparametric concurrent models in their article, their modeling framework does include linear and nonparametric concurrent models as well as concurrent interaction models of the form $v(t)w(t)\beta(t)$ or $f(v(t), w(t), t)$, among others. The unknown functions $f_r(\mathcal{X}_r, t)$ are approximated by sum of tensor product of marginal basis functions for $\mathcal{X}_r$ and $t$. Estimation of unknown coefficients are done by a penalized likelihood approach.

Scheipl, Gertheiss and Greven also presents a numerical comparison (see their Section 4.3) of their method to that of Wang and Shi, where the response process is binary with a functional intercept and observation-specific functional random effects. The two approaches yield similar errors for estimates of $\eta(\cdot)$ (Wang and Shi's method performs better in predicting the random effects but not so for fixed functional intercept, while Scheipl, Gertheiss and Greven's method shows the opposite trend). An investigation of coverage for estimates of $\eta(\cdot)$ shows that both the methods generally do not achieve nominal coverage, but Scheipl, Gertheiss and Greven's method converges to the nominal level faster as sample size increases. In addition, the Gaussian process based method requires much more computation time.

# SOFTWARE DEVELOPMENT

Some Gaussian process regression based procedures for continuous responses have been implemented in the R[26] package `GPFDA`[32]. Kim et al[19] implement their method using the `mgcv` package in R. Implementation details are discussed in their supplementary materials; code is available at `http://www4.stat.ncsu.edu/~maity/` under the software section. R functions for the procedure in Scheipl, Gertheiss and Greven[31] is included in the `pffr` function in the R package `refund`[12]. A MATLAB implementation of the functional single index model approach[18] is available at `http://www.stat.ucdavis.edu/PACE/download.html` under the name `fsim.zip`.

# COMPARISON OF METHODS

## General discussion

We review three types of modeling and fitting procedures for nonparametric functional concurrent regression models, namely, spline smoothing, Gaussian process and local kernel smoothing based approaches. Gaussian process based approaches attempt to directly model the covariance structure within each response function via a pre-specified covariance/kernel function that depends on the functional covariates. These approaches have the flexibility to incorporate multiple (and possibly a large number of) functional covariates simultaneously quite easily and without assuming any specific parametric structure such as additivity, giving these methods much more flexibility over their parametric counterparts. However, the computation complexity and time of such methods are generally greater. Also, the methods proposed by Shi et al[37] and subsequent articles require the functional response curves to be sampled on a sufficiently large number of points, so that smoothing of individual curves is possible. They also use the assumption that the functional covariates are sampled without any additional measurement errors. As a result, such methods may face difficulties when the covariate data is sampled over a sparse design and/or observed with measurement errors. Also, choice of smoothing parameters for the response curves as well as in the estimation step for any linear concurrent components can have heavy computational burden in practice. While the Gaussian process based methods provide good prediction based on all the covariates, interpretation of individual covariates are difficult to quantify and interpret.

Spline based methods for nonparametric concurrent models essentially model the unknown bivariate function $F(\cdot,\cdot)$ using linear combinations of tensor products of marginal basis functions for the two directions, and then propose to estimate the unknown coefficients by minimizing an appropriate penalized (likelihood or sum of squares) criteria. An advantage of such methods is that the response and/or the covariate functions do not have to be observed on a dense grid, and covariates can have measurement errors[19]. The smoothing parameters can be thought of variance components in a mixed model and can be determined easily using REML. A spline based approach also enables faster computational methods while keeping the prediction accuracy similar, as is demonstrated in Scheipl et al[31] in the comparison with Gaussian process based methods, especially if the number of observations per curve increases. A potential disadvantage of such spline based methods arises in presence of multiple functional covariates. If one uses an additive structure for the functional concurrent effects, the computational burden can be much bigger, especially when the number of covariates is moderate of large. This issue is even more pronounced if one wishes to incorporate interactions in the model as well.

The kernel smoothing based method of Jiang and Wang uses a single index modeling framework, where one computes an index by taking a linear combination of the functional covariates (using unknown coefficients), then the response function at time $t$ is modeled using a bivariate function of the index and $t$. The bivariate surface is estimated using kernel smoothing, and the unknown parameter in the single index is estimated using nonlinear optimization. When only one functional covariate in available, the model coincides with the standard nonparametric functional concurrent model. Unlike the Gaussian regression based method, the single-index model can accommodate sparsely observed data, and as such does not require smoothing of each response curve before fitting the model. It does however require that the covariates are observed without any measurement error. This procedure has the advantage that the same bandwidth is used to estimate the nonparametric mean function and the single index parameter, and the estimated parameter still achieves root-$n$ convergence rate (unlike semiparametric regression models with independent response, where one often needs to under-smooth the nonparametric component). The main computational burden of this method comes from two sources: the iterative procedure to estimate the single-index parameter, and the selection of bandwidths for the bivariate kernel smoothing using cross-validation. In our experience in the real data analysis, the kernel smoothing method (implemented in MATLAB) appeared to be more time consuming than the spline based method (implemented in R) with automatic smoothing parameter selection. One important point to be noted that the single-index parameter is identifiable only if specific constraints are imposed, such as $\beta_1 = 1$. Thus one should exercise caution when interpreting the absolute values of the estimates.

## Numerical illustration

For illustration purposes, we apply the spline based method of Kim et al (GFCM), Gaussian process regression based method of Shi et al (GPFR), and kernel smoothing based method of Jiang and Wang (FSIM) to the Gait dataset[27]. The gait dataset is available in the `fda` package[28] in the R software. The R and Matlab programs for the data analysis presented below are available in the software section of the author's web page (http://www4.stat.ncsu.edu/∼maity/).

The gait dataset consists of measurements of hip and knee angles (in degrees) through a movement cycle for $n = 39$ boys over $m = 20$ equally spaced time points. The left and middle panels in the first row of Figure 1 display the observed hip and knee angles, respectively. Data for two specific boys are highlighted in red and blue. In this analysis, we use hip angle as covariate and knee angle as response variable. To investigate the performance of the three methods, we first randomly split the dataset into a training set with $n_{\text{train}} = 30$ subjects and a test set with the remaining $n_{\text{test}} = 9$ subjects. We de-noise the covariate curves in the training set by performing FPCA and obtain smooth covariate curves for each subject in the training set. The FPCA procedure also produces a smooth estimate of the pointwise mean function and the pointwise variance can also be estimated from the output. We then standardize the training set covariates by subtracting the pointwise mean function and dividing by the pointwise standard deviation. Such a pre-processing method is suggested in Kim et al[19] to fit the spline based model. For the sake of comparison, we apply the Gaussian process and kernel smoothing based methods to these de-noised and transformed data.

To fit GFCM, we use tensor products of 7 spline basis functions for both $x$ and $t$ directions, and choose the penalty parameters using REML. For FSIM, the Epanechnikov kernel function is used and bandwidths are chosen using 10-fold crossvalidation. For the GPFR method, we include an intercept function in the model (as the software required such an inclusion to run properly), and use 10 basis functions for the intercept function as well as to smooth each response curve. One practical point to note here is that both the GFCM and FSIM approaches have automatic smoothing parameter selection implemented in the corresponding software. However we could not locate such a procedure in the software for the GPFR method (implemented in R package `GPFDA`[32], version 2.2). To save computing time, we therefore choose the smoothing parameters as $\lambda_1 = 10^{-4}$ (to smooth each response curve) and $\lambda_2 = 10^{-4}$ (to smooth the intercept function). These values are chosen so that the fit obtained from GPFR is reasonable. We note that changing the values of these parameters may impact the results from the GPFR method.

We first investigate the out-of-sample prediction performance of the three methods. For each method, we take the fitted surface obtained from the training set, and predict the responses in the test set using the covariates in the test set. To this end, we de-noise the test set covariates using the same eigen-components obtained from the training set and standardize them using the original transformation used in the training set. We then compute root mean squared prediction error as

$$\text{RMPE} = \left[ \frac{1}{m\, n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \sum_{j=1}^{m} \{\widehat{Y}_i^{\text{pred}}(t_j) - Y_i^{\text{test}}(t_j)\}^2 \right]^{1/2},$$

where $Y_i^{\text{test}}(t_j)$ and $\widehat{Y}_i^{\text{pred}}(t_j)$ denote the $i$-th actual and predicted responses in the test set, respectively, evaluated at time $t_j$. We repeat this process of splitting the data set into training and test sets 50 times, and compute RMPE for each split. Boxplots of the RMPE values for the three methods are displayed in Figure 1 (first row, right panel). The average RMPEs for the GFCM, GPFR and FSIM methods are 5.86, 6.16, and 5.72, respectively. We observe that the out-of-sample predictions errors for the three methods are quite similar for the gait data. One practical point to note here is that during the process of randomly splitting the data set into training and testing sets, we sometimes encounter cases where the test set covariates are not fully inside the range of the training set covariates, and thereby making prediction inappropriate for those test sets. Therefore we perform our analysis using only those 50 splits where this issue is not present.

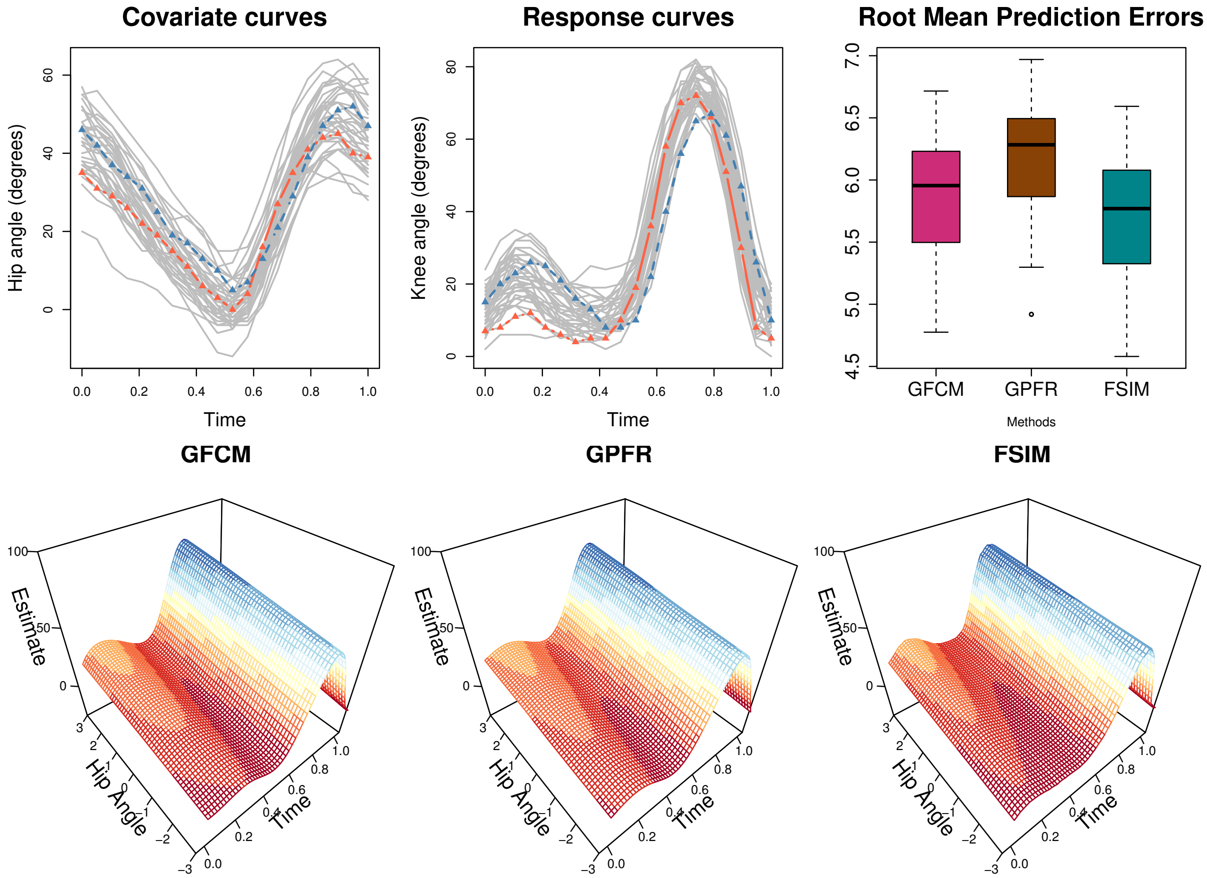


Figure 1: Results from gait data analysis. The first row displays the observed hip (left panel) and knee angles (middle panel), and the out-of-sample root mean squared prediction errors (right panel) of the three methods. The bottom row displays the full estimated bivariate surfaces using spline (left panel), Gaussian process (middle panel) and kernel smoothing (right panel) based methods, respectively.

Next we estimate the full bivariate surface $F(\cdot,\cdot)$ using each of the three methods based on the full data set. To this end, we use a equally spaced grid of 59 points in $[0,1]$ for the $t$-direction and 59 equally spaced points between the minimum and maximum values of the transformed covariates (roughly between -3 and 3). We then estimate $F(\cdot,\cdot)$ for each point on this grid. The estimated surfaces are shown in Figure 1 (second row) for GFCM (left panel), GPFR (middle panel) and FSIM (right panel). Here also it is evident that the estimated surfaces from the three methods are quite similar to each other, and as such indicative of similar performance of the three methods.

It is worth mentioning that the functional variables in the gait dataset are observed on an equally spaced grid of points, and as such all three methods perform similarly. However, this may not be the case in every situation. For example, while the GFCM and FSIM methods are applicable for sparsely observed data, the GPFR procedure is not applicable in such a situation. Furthermore, we have de-noised the covariate curves by applying FPCA for fair comparison since both FSIM and GPFR assume that the covariates are measured without any error. It is not clear how FSIM would perform in the situation where covariates are measured with error and are observed sparsely.

We note that the analysis presented here is meant for illustration purposes, and is not intended to be a fully fledged data analysis. Looking closely at the estimated surfaces, one can see that while $F(\cdot,\cdot)$ is nonlinear in time, it is close to linear in the $x$-direction (hip angle). Thus a simpler model, such as a linear functional concurrent model, might be sufficient for this data set. However, as Kim et al[19] observes in their numerical studies, fitting a nonparametric concurrent model when the actual relationship is linear does not result in significant loss in performance (e.g., prediction error), but fitting linear concurrent model when the true relationship is in fact nonlinear may result is drastic loss of performance. Thus a safer option might be to fit a nonparametric model to begin with and then assessing whether a linear model would suffice by examining the fitted surface.

# CONCLUSIONS

Concurrent models are useful statistical tools to perform function-on-function regression. Due to the limitations imposed by the assumed linearity, it is of practical importance to extend the classical linear concurrent regression to the nonparametric setting. The methods we review in this article provide statistical tools for further development. There are many open research questions in this area. One such question of interest is variable selection in presence of multiple functional covariates. While there are a few procedures for functional variable selection in other functional regression models, to the best of our knowledge, no such procedures are available for nonparametric functional concurrent regression.

Hypothesis testing in nonparametric concurrent models remains a relatively new area of research as well. To the best of our knowledge, no procedure for hypothesis testing has been developed in the contest of nonparametric concurrent regression using Gaussian process framework or using local kernel smoothing approach. In the spline smoothing framework, Kim et al[19] proposed two bootstrap based testing procedures: 1) testing for a global effect, that is, whether the function $F$ depends on both $X(\cdot)$ and $t$ or just $t$ in a single covariate nonparametric concurrent model, and 2) testing for inclusion, that is, in an additive nonparametric concurrent model, whether one needs to include a particular functional covariate or not. For both the tests, Kim et al[19] propose to use $F$-ratio type test statistics, and approximate their null distributions using bootstrap. Their numerical study shows that the proposed tests have close to nominal type one errors and good power consistently over different sample sizes and covariance structures of the error process. A particular problem of interest is to see whether one can use the connection of spline models and linear mixed effects models to develop more efficient tests for these problems.

## Acknowledgments

The author is extremely grateful to the Review Editor and two anonymous referees whose suggestions greatly improved the results and presentation of the article.

# References

1. Brockhaus, S., F. Scheipl, T. Hothorn, and S. Greven (2015). The functional linear array model. Statistical Modelling, 15, 279 - 300.

2. Brumback, B. and Rice, J. A. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion). J. Amer. Statist. Assoc, 93, 961 - 994.

3. Chiang, C. T., Rice, J. A., and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. Journal of the American Statistical Association 96, 605619.

4. Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. In Statistical Models in S (Chambers, J. M. and Hastie, T. J., eds), 309 - 376. Wadsworth & Brooks, Pacific Grove.

5. Di, C., Crainiceanu, C., Caffo, B., and Punjabi, N. (2009). Multilevel functional principal component analysis. Annals of Applied Statistics, 3, 458 - 488.

6. Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. Statistical Science, 11, 89 - 121.

7. Eubank, R. L., Huang, C., Muñoz Maldonado, Y., Wang, N. and Wang, S. and Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models, J. R. Statist. Soc. B, 66, 653 - 667.

8. Fan, J. and Zhang, J. (2000). Two-Step estimation of functional linear models with applications to longitudinal data. Journal of the Royal Statistical Society, Ser. B, 62, 303 - 322.

9. Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models, Statistics and Its Interface, 1, 179 - 195.

10. Fomel, S. Inverse B-spline interpolation, Stanford Exploration Project, 2000. http://sepwww.stanford.edu/public/docs/sep105/sergey2/paper_html/node5.html

11. Gertheiss, J., V. Maier, E. F. Hessel, and A.-M. Staicu (2015). Marginal functional regression models for analyzing the feeding behavior of pigs. Journal of Agricultural, Biological, and Environmental Statistics 20, 353 - 370.

12. Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C. and Reiss, P. T. (2016). refund: Regression with Functional Data. R package version 0.1-16. https://CRAN.R-project.org/package=refund

13. Gunn, S. R. (1998, May). Support vector machines for classification and regression, Technical report, Faculty of Engineering, Science and Mathematics School of Electronics and Computer Science.

14. Hilgert, N., Manrique Chuquillanqui, T., Crambes, C., Mas, A. (2015). Ridge regression for the functional concurrent model. Presented at 60. World Statistics Congress ISI2015, Rio de Janeiro, BRA (2015-07-26 - 2015-07-31). http://prodinra.inra.fr/record/310110

15. Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. Biometrika, 85, 809 - 822.

16. Huang, J. Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. Biometrika, 89, 111 - 128.

17. Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. Statistica Sinica, 14, 763 - 788.

18. Jiang, C.R. and Wang, J.L. (2011): Functional Single Index Model for Longitudinal Data, The Annals of Statistics, 39, 362 - 388.

19. Kim, J., Maity, A. and Staicu-A-M. (2016). General Functional Concurrent Model. Technical report #2660, North Carolina State University. Available at http://www.stat.ncsu.edu/information/library/mimeo.php.

20. Liu, D., Lin, X. and Ghosh, D. (2007) Semiparametric Regression of Multi-Dimensional Genetic Pathway Data: Least Squares Kernel Machines and Linear Mixed Models. Biometrics, 63, 1079-1088.

21. Liu, D., Ghosh, D. and Lin, X. (2008) Estimation and Testing for the Effect of a Genetic Pathway on a Disease Outcome Using Logistic Kernel Machine Regression via Logistic Mixed Models. BMC Bioinformatics, 9, 292. SAS code

22. MacKay D.J.C. 1999. Introduction to Gaussian processes. Technical report. Cambridge University. (Available from http://wol.ra. phy.cam.ac.uk /mackay/GP/)

23. Marx, B. D. and Eilers, P. H. C. (2005). Multivariate penalized signal regression. Technometrics, 47, 13 - 22.

24. McLean, M. W., Hooker, G., Staicu, A. M., Scheipl, F., and Ruppert, D. (2014). Functional generalized additive models. Journal of Computational and Graphical Statistics 23, 249 - 269.

25. Qu, A. and Li, R. (2006). Quadratic inference functions for varying coefficient models with longitudinal data. Biometrics, 62, 379 - 391.

26. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

27. Ramsay, J.O. and Silverman, B.W. (2006). Functional Data Analysis. New York: Springer.

28. Ramsay, J. O., Wickham, H., Graves, S. and Hooker. G. (2014). fda: Functional Data Analysis. R package version 2.4.4. https://CRAN.R-project.org/package=fda

29. Rasmussen, C. E. (2004). Gaussian Processes in Machine Learning, in *Advanced Lectures on Machine Learning*, Editors: Olivier Bousquet, Ulrike von Luxburg, Gunnar Rtsch, Lecture Notes in Computer Science Volume 3176 2004.

30. Reiss, P. T. and Ogden, R. T. (2009). Smoothing parameter selection for a class of semiparametric linear models. J. R. Statist. Soc. B, 71, 505 - 523.

31. Scheipl, F., Gertheiss, J. and Greven, S. (2016). Generalized Functional Additive Mixed Models, Electronic Journal of Statistics, 10, 1455 - 1492.

32. Shi, J. Q. and Cheng, Y. (2014). GPFDA: Apply Gaussian Process in Functional data analysis. R package version 2.2. https://CRAN.R-project.org/package=GPFDA

33. Shi, J. Q. and Choi, T. (2011). Gaussian Process Regression Analysis for Functional Data, Chapman and Hall/CRC.

34. Shi, J. Q., Murray-Smith, R. and Titterington, D. M. (2005). Hierarchical Gaussian process mixtures for regression, Statistics and Computing, 15, 31 - 41.

35. Shi, J. Q., Wang, B., Will, E. J. and West. R. M. (2012). Mixed-effects GPFR models with application to dose-response curve prediction. Statistics in Medicine. 31, 3165 - 3177.

36. Shi, J.Q. and Wang, B. (2008) Curve Prediction and Clustering with Mixtures of Gaussian Process Functional Regression Models.Statistics and Computing, 18, 267 - 283.

37. Shi, J. Q., Wang, B., Murray-Smith, R. and Titterington, D. M. (2007). Gaussian Process Functional Regression Modeling for Batch Data, Biometrics 63, 714 - 723.

38. Wang, B. and Shi, J. Q. (2014) Generalized Gaussian process regression for non-Gaussian functional data. Journal of American Statistical Association 99, 285 - 298.

39. Williams C.K.I. 1998. Prediction with Gaussian processes: From linear regression to linear prediction and beyond. In: Jordan M.I. (Ed.), Learning and Inference in Graphical Models, Kluwer, pp. 599 - 621.

40. Williams C.K.I. and Rasmussen C.E. 1996. Gaussian process for regression. In: Touretzky D.S. et al. (Eds.), Advances in Neural Information Processing Systems 8, MIT Press.

41. Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Am. Statist. Ass., 99, 673 - 686.

42. Wood, S. N. (2006) Generalized Additive Models: An Introduction with R, Chapman & Hall, London

43. Wood, S. N. (2008). Fast stable direct fitting and smoothness selection for generalized additive models. J. R. Statist. Soc. B, 70, 495 - 518.

44. Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, Journal of the Royal Statistical Society Series B, 73, pp. 3 - 36.

45. Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. J. R. Stat. Soc. Ser. B Stat. Methodol. 64, 363 - 410.

46. Xiao, L., Li, Y., and Ruppert, D. (2013). Fast bivariate P-splines: the sandwich smoother. Journal of the Royal Statistical Society: Series B 75(3), 577 - 599.

47. Yao, F., Muller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. Journal of the American Statistical Association 100, 577 - 590.

48. Zhang, J., Clayton, M. K. and Townsend, P. A. (2011). Functional Concurrent Linear Regression Model for Spatial Images, Journal of Agricultural, Biological, and Environmental Statistics, 16, 105 - 130.

## FURTHER READING

Anjyo, K. and Lewis, J. P. (2011). RBF interpolation and Gaussian process regression through an RKHS formulation, Journal of Math-for-Industry, 3, pp. 63 - 71.

Cederbaum J., Pouplier M., Hoole P. and Greven S. (2016). Functional Linear Mixed Models for Irregularly or Sparsely Sampled Data. Statistical Modelling, 16(1), 67-88.

Kim, J., Staicu-A-M., Ruppert, D. and Maity, A. (2016). General Additive Function-on-Function Regression. Available at https://arxiv.org/abs/1606.03775.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model, The Canadian Journal of Statistics, 31, 115 - 128.

Ramsay, J. O., Hooker, G. and Graves, S. (2009). Functional Data Analysis with R and MATLAB, Springer-Verlag New York.

Scheipl, F., Staicu, A.-M. and Greven, S. (2015). Functional Additive Mixed Models. Journal of Computational and Graphical Statistics, 24, 477 - 501.