

Nonlinear dynamic systems modeling using Gaussian processes: Predicting ionospheric total electron content over South Africa

E. R. Ackermann,^{1,2} J. P. de Villiers,^{1,2} and P. J. Cilliers³

Received 14 December 2010; revised 8 June 2011; accepted 11 July 2011; published 4 October 2011.

[1] Two different implementations of Gaussian process (GP) models are proposed to estimate the vertical total electron content (TEC) from dual frequency Global Positioning System (GPS) measurements. The model falseness of GP and neural network models are compared using daily GPS TEC data from Sutherland, South Africa, and it is shown that the proposed GP models exhibit superior model falseness. The GP approach has several advantages over previously developed neural network approaches, which include seamless incorporation of prior knowledge, a theoretically principled method for determining the much smaller number of free model parameters, the provision of estimates of the model uncertainty, and a more intuitive interpretability of the model.

Citation: Ackermann, E. R., J. P. de Villiers, and P. J. Cilliers (2011), Nonlinear dynamic systems modeling using Gaussian processes: Predicting ionospheric total electron content over South Africa, *J. Geophys. Res.*, 116, A10303, doi:10.1029/2010JA016375.

1. Introduction

[2] The ionosphere is the uppermost part of the Earth's atmosphere, stretching from a height of about 80 km to more than 1000 km [Arikan *et al.*, 2003]. At such heights the atmosphere is so thin that free electrons can exist for short periods of time before they are recombined with nearby positive ions. The level of ionization depends primarily on the received solar radiation, and it exhibits strong diurnal and seasonal variations. The number of these free electrons is sufficient to affect radio wave propagation, which influences the reliability and operation of systems which make use of the earth-ionosphere waveguide such as radio communications [Goodman and Aarons, 1990].

[3] Other severely affected systems range from navigational systems (including high-frequency direction finding and satellite navigation) to Global Positioning System (GPS) surveying and space weather forecasts [Kouris *et al.*, 2004; Hoffman-Wellenhof *et al.*, 1992; Tulunay *et al.*, 2004, 2006; Opperman *et al.*, 2007]. The vertical total electron content (TEC), defined as the number of free electrons in the ionosphere contained in a vertical column of unit cross sectional area (1 m^2), provides an indication of the ionospheric electron density and is often used to (partially) characterize the ionospheric variability.

[4] In recent years, dual frequency GPS receivers have been used extensively to monitor the TEC at a large number of locations worldwide [see, e.g., Howe *et al.*, 1998], and much research effort has been expended at modeling the

observed GPS TEC in order to better understand the nature of (and counteract) the effects experienced by the various affected systems [Spalla and Ciralo, 1994]. Two main categories of mathematical models can typically be identified for the modeling of geophysical phenomenon, namely theoretical models, which are based on our physical understanding of the process, and empirical models, which are based primarily on observed data.

[5] The theoretical models are usually derived from the classical physical equations of continuity, conservation of energy, and momentum for electrons, ions, and neutral molecules and include models such as the Time Dependent Ionospheric Model (TDIM) [Schunk *et al.*, 1986]. Such models are often very insightful when studying the underlying physics of interactions between various processes within the system under investigation, but they generally offer only modest practical utility at large computational costs [Spalla and Ciralo, 1994]. In addition, owing to the complexity of many geophysical processes, first-principle models and their analytical approximations can often not be derived or applied efficiently and, as a result, more flexible types of models are needed.

[6] In contrast, empirical models (both parametric and non-parametric) offer very little in the way of understanding the physics of a process but often provide superior practical performance as compared to theoretical models in terms of both accuracy and computational cost. As a consequence, these types of models have received considerable attention in the past few years (the interested reader is referred to the book by Nelles [2001] and the references contained therein for a comprehensive discussion of various nonlinear system identification methods and some of their applications).

[7] Some of the most commonly used approaches to nonlinear dynamic system identification from observed data include neural networks (NNs) [Narendra and Parthasarathy, 1990], fuzzy models [Takagi and Sugeno, 1985], and local

¹Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria, South Africa.

²Defence, Peace, Safety and Security, Council for Scientific and Industrial Research, Pretoria, South Africa.

³Hermanus Magnetic Observatory, Hermanus, South Africa.

model networks (LMNs) [Murray-Smith and Johansen, 1997; Gregorčič and Lightbody, 2008]. However, neural network and fuzzy models are difficult to interpret, and like other black-box approaches they lack sufficient transparency [Gregorčič and Lightbody, 2008]. This means that the resulting models do not provide any physical insight regarding the underlying processes, and as a consequence this lack of transparency hinders the incorporation of prior domain knowledge into the neural network models.

[8] In addition, neither neural network nor fuzzy models are well suited to model high-dimensional data. This is known as the curse of dimensionality [Ažman and Kocijan, 2007] which requires that a large number of neurons and a lot of data must be used to train acceptable models.

[9] Multiple model approaches (and fuzzy LMNs in particular) were proposed to provide increased model transparency and to reduce the severity of the curse of dimensionality [Murray-Smith and Johansen, 1997]. LMNs form a global process model by blending a number of local models, each of which are valid only for a small interval on the operational domain of the process [Johansen and Foss, 1997; Gregorčič and Lightbody, 2008]. Such models can only be expected to accurately model the system close to the points where the local models have been identified, and away from these points the accuracy of the network can decrease rapidly due to unmodeled system dynamics [Murray-Smith et al., 1999; Shorten et al., 1999].

[10] In an attempt to improve the modeling performance in the regions between local models while maintaining a level of transparency an alternative approach to nonlinear dynamic system identification was proposed by Murray-Smith et al. [1999], namely the Gaussian process modeling approach. Gaussian processes are being increasingly used to tackle many of the standard applications usually addressed by artificial neural networks [Kocijan et al., 2003].

[11] Nevertheless, neural networks have found considerable use in the modeling of TEC for over a decade, and in fact, besides the commonly encountered global International Reference Ionosphere (IRI) empirical model [Bilitza, 2001] which suffers from a historic scarcity of data in the Southern Hemisphere [McKinnel, 2002], the most interesting contributions to TEC modeling from a practical point of view have arguably been the development of several regional neural network models (see for example the work by Hernandez-Pajares et al. [1997] which made use of GPS observations, Xenos et al. [2003] which employed Faraday-rotation derived TEC, Tulunay et al. [2006] where NNs were used to predict TEC maps, as well as the work by Leandro and Santos [2007], Habarulema et al. [2009], and Yilmaz et al. [2009]).

[12] In this paper we propose the Gaussian process (GP) modeling approach [Rasmussen and Williams, 2006] as an alternative to the neural network approach for the modeling of GPS derived TEC in an effort to alleviate some of the many difficulties inherent in the current neural network-based modeling approaches. The Gaussian process approach is very closely related to Kriging, which is commonly used as a geostatistical technique to interpolate between values in spatially distributed datasets [see, e.g., Olivera and Webster, 1990].

[13] The rest of this paper is organized as follows. Section 2 describes several approaches to modeling geophysical pro-

cesses, with special emphasis on model validation. Neural networks is then introduced as a popular modeling approach, along with their most important limitations, after which it is shown how these limitations can be handled efficiently and elegantly through the Gaussian process modeling approach. Section 3 presents a number of Gaussian process models for modeling the daily GPS-derived TEC, and the results are compared to existing neural network models developed by Habarulema et al. [2007]. Finally, it is shown in section 4 how the Gaussian process approach can be used to develop adaptive predictive time series models which are better suited to the task of predicting GPS TEC than existing models.

2. Modeling Geophysical Processes

[14] Many geophysical processes are characterized by complex, nonlinear behavior with numerous factors affecting and influencing the dynamics of the system in an often unknown manner. The development of accurate, representative models can help to further our understanding of such geophysical processes through the systematic investigation of model behavior. However, it is not an easy task to construct such models for a number of different reasons.

[15] First, it is often very difficult or impossible to derive accurate analytical models from first principles, especially if many different inputs are involved. Second, the high dimensionality and nonlinear behavior of many geophysical processes make it very difficult to define appropriate parametric models since the contributions of all the inputs have to be considered jointly. As a consequence more flexible types of models and an approach to construct such models are required. In fact, it is precisely this problem of visualizing or characterizing nonlinear, high-dimensional data that led to the development and widespread application of black box system identification approaches such as neural networks.

2.1. Framework for Nonlinear System Identification

[16] A general framework for dynamic system identification is presented in the work of Ažman and Kocijan [2007] and consists of the following six stages: (1) defining the model purpose, (2) model selection, (3) experimental design, (4) data collection and preparation, (5) model training, and (6) model validation. The model identification procedure is iterative in that it may be necessary to return to some previous step at any time in the identification process [Ažman and Kocijan, 2007]. Each of the steps in the system identification framework is discussed very briefly below.

2.1.1. Defining the Model Purpose

[17] The purpose of the desired geophysical process model has to be defined, since it will largely determine which type of model will be the most appropriate for the particular purpose. For example, if our aim is to understand the geophysical process we should consider developing a model from first principles, whereas if our aim is to predict system behavior (such as, e.g., weather forecast models), we might rather wish to use a different modeling approach.

2.1.2. Model Selection

[18] As mentioned previously, the decision to use a specific type of model depends mainly on the purpose of the model. Model selection also involves the structure determination of the particular type of model, such as specifying

the architecture of a neural network or the covariance function of a Gaussian process model.

2.1.3. Experimental Design

[19] The experimental design is of critical importance in any system identification procedure. The quality of the model depends on the system information contained in the measurement data [Ažman and Kocijan, 2007], which in turn depends on the quality of the experimental design.

[20] However, in the context of geophysical processes the experimental design translates almost exclusively to the specification of the measurement setup, since we cannot typically influence the geophysical process itself. In other words, we must determine which variables to measure and how to measure them. Some of the most important design considerations (apart from deciding which variables to measure) include the determination of the sampling interval, the amount of data required, and the necessary range of observations.

2.1.4. Data Collection and Preparation

[21] After the measurement setup and requirements have been specified, it is usually relatively straightforward to collect the necessary data. However, it is often more difficult to obtain a comprehensive set of measurements for geophysical processes than for other types of processes commonly encountered in practice. This is primarily due to the fact that we cannot sample geophysical processes under arbitrary input conditions. That is, we cannot typically provide a set of desired inputs to the geophysical system, but instead we have to wait for those input conditions to occur naturally before we can observe the resulting system behavior.

[22] The amount and type of data preprocessing that is required typically depend on the measurement setup, the nature of the observations as well as the type of model that was selected in step 2 of the system identification procedure.

2.1.5. Model Training

[23] The task of “training” a model simply refers to the process of finding a set of parameters (or hyperparameters in the case of non-parametric Gaussian processes) which minimizes, at least to some extent, the difference between the model and the observed data. Various training strategies exist, and the appropriate choice once again depends primarily on the type and purpose of the selected model.

2.1.6. Model Validation

[24] Model validation concerns the level of agreement between the model and the physical process under consideration; and although it is probably one of the most important steps in system identification, it is often underemphasized [Ažman and Kocijan, 2007]. Several features should ideally be used to assess the validity of any type of model [Murray-Smith, 1998], the most important of which are arguably the model plausibility, model falseness, and model purposiveness [Ažman and Kocijan, 2007].

[25] In order to evaluate the model plausibility, we essentially have to answer two related questions: whether the model “looks plausible” and whether it “behaves logical.” The first question mainly concerns the model structure, that is, it seeks to answer whether or not the model selection and model composition can be adequately justified. The second question should be answered by visually inspecting the model response over a representative range of typical inputs and noting whether the model behavior at least seems logical or plausible.

[26] Model falseness represents the agreement between the process and the model output [Ažman and Kocijan, 2007] and is typically expressed quantitatively in terms of performance measures such as the sum of square errors (SSE), the mean square error (MSE), or more commonly, the root mean square error (RMSE) defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}, \quad (1)$$

where $e_i = \hat{y}_i - y_i$ is the model prediction error for the i th step of the simulation.

[27] It should be emphasized that the model falseness alone is usually not sufficient to establish whether or not a model is in fact a good model, and it should only be considered meaningful if the model has already been found to be plausible. In spite of this, the model falseness is often employed as the only measure of model performance, and decisions to choose one model over another is often made based only on the model falseness.

[28] Model purposiveness simply concerns the usefulness and effectiveness of the model in addressing the problem it was originally developed for. When the model is useful for its intended application it is deemed purposive, whereas if for some reason it is not considered useful, it is termed non-purposive. For example, if a model is required which must be able to predict daily rainfall a week in advance, and instead a model is developed which can only predict rainfall accurately for the next day, then the model is considered not purposive even though it may still be a good model in all other respects.

2.2. Neural Networks

[29] It is not always easy or convenient to specify a simple parametric model, and as a consequence, more flexible approaches such as neural networks are frequently employed in modeling complex nonlinear systems. Neural networks (and multilayer perceptron neural networks in particular) are commonly used as universal function approximators and consist of a number of interconnected artificial neurons, each described by an appropriate transfer function. A neural network is further defined by its architecture, which specifies how the artificial neurons are connected to each other. Each connection has an associated weight which must be “learned” from input-output data pairs.

[30] Multilayer perceptron neural networks typically consist of an input layer, an output layer and at least one hidden layer which connects the input and output layers. An illustration of a simple neural network with one input layer, one hidden layer and one output layer is shown in Figure 1, where a total of $(3 \times 5) + (5 \times 2) = 25$ weights have to be learned.

[31] To determine an appropriate set of weights, the network is presented with multiple input-output data pairs, and the weights are adjusted so as to minimize some training error. This process can be interpreted as “learning” an underlying function which maps the input variables to the output variables. It is then hoped that the learned function will “generalize” well to new, previously unseen examples [Bishop, 1995].

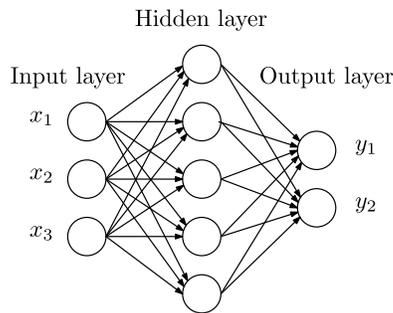


Figure 1. Simple three layered neural network.

[32] More specifically, the process of learning in multi-layer perceptron neural networks can be described as follows. Let $\mathcal{D} = \{\mathbf{x}_i, t_i\}_{i=1}^N$ denote the set of N training data pairs, where t_i denotes the target corresponding to the input \mathbf{x}_i . Each t_i is assumed to be a noise-corrupted version of an underlying y_i . A neural network architecture \mathcal{A} is invented, consisting of a specification of the number of layers, the number of units in each layer, the type of activation function performed by each unit, and the available connections between the units [MacKay, 1992]. If we assign a set of weights \mathbf{w} to the connections of the network, the task of the neural network is to define a (possibly nonlinear) mapping $\mathbf{y} = f(\mathbf{x}; \mathbf{w}, \mathcal{A})$ from the input variables \mathbf{x} to the output variables \mathbf{y} [MacKay, 1992]. The training error is commonly defined as

$$E_{\mathcal{D}}(\mathcal{D}|\mathbf{w}, \mathcal{A}) = \sum_{i=1}^N \frac{1}{2} (f(\mathbf{x}_i; \mathbf{w}, \mathcal{A}) - t_i)^2. \quad (2)$$

The task of “learning” the neural network then simply refers to the process of finding a set of connections \mathbf{w} which minimizes, at least to some extent, the training error $E_{\mathcal{D}}$. The commonly used backpropagation algorithm learns by performing gradient descent on the training error $E_{\mathcal{D}}$ directly in \mathbf{w} space [MacKay, 1992].

[33] Despite being well established, neural networks still lack some properties that would further increase their acceptance [Kocijan et al., 2003]. More specifically, neural networks are not always easy or straightforward to implement in practice, and the approach lacks a principled framework with which to determine the architecture of the network [Rasmussen and Williams, 2006].

[34] Neural networks are further characterized by the requirement of large training datasets [Bishop, 1995], and the architecture and other learning parameters are usually optimized by repeated experimentation [Rasmussen and Williams, 2006]. Neural networks, like other black box modeling approaches, also lack sufficient transparency, and it is very difficult to incorporate any prior knowledge into the model. The number of neurons required to capture highly nonlinear behavior can also become prohibitively large, especially when high-dimensional models must be developed. In an attempt to address some of the many challenges presented by current neural network approaches an alternative supervised learning method is proposed, namely the Gaussian process approach, briefly introduced next.

2.3. The Gaussian Process Approach

[35] The Gaussian process (GP) model can be viewed as a probabilistic, nonparametric model which differs from most other black box identification approaches (including neural networks) in that it does not attempt to approximate the modeled system by fitting the parameters of a set of selected basis functions [Ažman and Kocijan, 2007], but instead the GP model searches for relationships in the measured data. Gaussian processes are stochastic processes which can be thought of as the generalization of the Gaussian probability distribution over a finite vector space to a function space of infinite dimension [MacKay, 1997]. That is, a Gaussian process is a mathematical set comprised of an infinite amount of random variables of which any subset is jointly Gaussian [Rasmussen and Williams, 2006].

[36] Gaussian processes are being increasingly used to tackle many of the standard applications usually addressed by artificial neural networks [Kocijan et al., 2003]. This is not surprising, since the two models are closely related: in the Bayesian treatment under suitable conditions, Gaussian processes correspond to neural networks with an infinite number of hidden units [Neal, 1997].

[37] The observation that, from the Bayesian perspective at least, the choice of a neural network model can be viewed as defining a prior probability distribution over nonlinear functions has motivated the idea of discarding parameterized networks and working directly with Gaussian processes [MacKay, 1997]. Computations in which the parameters of the network are optimized are then replaced by simple matrix operations using the covariance matrix [MacKay, 1997].

[38] Consider again the data set $\mathcal{D} = \{\mathbf{x}_i, t_i\}_{i=1}^N$, and let $\mathbf{X}_N = \{\mathbf{x}_i\}_{i=1}^N$ denote the set of input vectors and let $\mathbf{t}_N = \{t_i\}_{i=1}^N$ denote the set of corresponding target values. In the Bayesian interpretation, a nonlinear function $\mathbf{y} = f(\mathbf{x})$ is assumed to underlie the data \mathcal{D} , and the adaptation of the model to the data corresponds to an inference of the function given the data [MacKay, 1997]. The inference of $f(\mathbf{x})$ is then described by the posterior probability distribution

$$P(f(\mathbf{x})|\mathbf{t}_N, \mathbf{X}_N) = \frac{P(\mathbf{t}_N|f(\mathbf{x}), \mathbf{X}_N)P(f(\mathbf{x}))}{P(\mathbf{t}_N|\mathbf{X}_N)}. \quad (3)$$

[39] If we place a Gaussian prior $P(f(\mathbf{x}))$ directly on the space of functions, it is called a Gaussian process [MacKay, 1997]. A Gaussian process is specified by a mean $\mu(\mathbf{x})$ and a covariance function $C(\mathbf{x}, \mathbf{x}')$ which expresses the expected covariance between the value of the function f at the points \mathbf{x} and \mathbf{x}' [MacKay, 1997]. We denote such a Gaussian process as $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), C(\mathbf{x}, \mathbf{x}'))$. The mean and covariance functions are typically specified by a set of hyperparameters which define the characteristics of the Gaussian process. The hyperparameters can be found by optimizing (minimizing) the negative log marginal likelihood, which in the noise-free case is given by

$$\log p(\mathbf{t}_N|\mathbf{X}_N) = -\frac{1}{2} \mathbf{t}_N^T \mathbf{C}^{-1} \mathbf{t}_N - \frac{1}{2} \log |\mathbf{C}| - \frac{N}{2} \log 2\pi, \quad (4)$$

where $\mathbf{C} = C(\mathbf{X}_N, \mathbf{X}_N)$ is the covariance matrix and $|\mathbf{C}|$ the determinant of \mathbf{C} . After the covariance function has been specified and the hyperparameters have been determined,

the Gaussian process can be used to predict any number N_* of test outputs, denoted here as \mathbf{t}_* , from the posterior distribution [Rasmussen and Williams, 2006]:

$$p(\mathbf{t}_* | \mathbf{X}_*, \mathbf{X}_N, \mathbf{t}_N) \sim \mathcal{N}(\mathbf{m}, \Sigma), \quad (5)$$

where the mean vector \mathbf{m} and the covariance matrix Σ are given by

$$\mathbf{m} = C(\mathbf{X}_*, \mathbf{X}_N)C(\mathbf{X}_N, \mathbf{X}_N)^{-1}\mathbf{t}_N \quad (6)$$

$$\Sigma = C(\mathbf{X}_*, \mathbf{X}_*) - C(\mathbf{X}_*, \mathbf{X}_N)C(\mathbf{X}_N, \mathbf{X}_N)^{-1}C(\mathbf{X}_N, \mathbf{X}_*) \quad (7)$$

[40] Here the predictive mean \mathbf{m} contains the most likely values of the test outputs, while the diagonal of the covariance matrix Σ gives the corresponding predictive variances.

[41] The Gaussian process approach for nonlinear system identification as a structured, supervised learning alternative to the neural network approach presents a number of interesting and important advantages at almost every stage in the system identification framework. In contrast to neural networks, the structure determination of Gaussian processes follows a principled framework in which prior knowledge about the geophysical process can be incorporated through the careful specification of an appropriate covariance function [Rasmussen and Williams, 2006]. This makes the Gaussian process approach particularly attractive in terms of model selection, since it enables other researchers to build on existing knowledge to quickly and effortlessly extend and refine Gaussian process models. This also results in increased model transparency in which the hyperparameters of well chosen covariance functions may be interpreted to some extent. For example, the hyperparameters of a periodic covariance function given by

$$k(x, x') = \theta_1^2 \exp(\sin^2(\theta_2(x - x'))) \quad (8)$$

are physically interpretable in that the amplitude and period of the covariance function can be readily obtained directly from the hyperparameters, θ_1 and θ_2 .

[42] Specifying the structure of the covariance function corresponds to placing a prior directly on the space of functions, which dramatically reduces the number of parameters that need to be determined for Gaussian process models compared to conventional neural networks [Gregorčič and Lightbody, 2008]. As a consequence, Gaussian processes are generally easier to implement and interpret and typically require much less training data than comparable neural networks. Not only do Gaussian processes require less training data, but the process of data collection is also facilitated in that the predictive variance of the Gaussian process can be used to identify and evaluate regions where additional training data should be collected. Regions with a high predictive variance should be augmented with additional training data (if available) or should not be considered as part of the operational range of the model, since model predictions can only be viewed as informative in regions where sufficient training data is available.

[43] Model training in the Gaussian process context usually refers to the minimization of the negative log likelihood (or, equivalently, to the maximization of the log likelihood),

which requires that a number of hyperparameters describing the covariance function must be determined. Since Gaussian processes are typically characterized by much fewer hyperparameters than neural networks have parameters (or weights), it is generally much easier to optimize the hyperparameters of a Gaussian process than it is to optimize the parameters of a neural network. The maximum likelihood (ML) optimization approach does not specify a particular optimization technique or method, and both gradient-based and zero-order methods are commonly employed.

[44] When working with Gaussian processes, model plausibility can partially be assessed by direct inspection of the hyperparameters. Furthermore, model falseness can be expressed in the usual manner, that is, by some standard performance measure such as the RMSE, or in the Gaussian process framework, other more informative performance measures can be defined. For example, the log predictive density error, given by

$$\text{LD} = \frac{1}{2} \log(2\pi) + \frac{1}{2N} \sum_{i=1}^N \left(\log(\sigma_i^2) + \frac{e_i^2}{\sigma_i^2} \right), \quad (9)$$

takes the entire predictive distribution into account, where σ_i^2 is the predictive variance for the i th step of the simulation [Azman and Kocijan, 2007]. The LD approach effectively penalizes errors in overconfident predictions more heavily than acknowledged bad predictions.

[45] Finally, the GP framework also provides an elegant way of performing sensitivity analyses and can easily be used to infer the relative importance of various input parameters (referred to as automatic relevance determination). Inputs with large characteristic length scales are generally considered less important than inputs with small length-scale parameters.

3. Modeling Daily Total Electron Content

[46] A number of neural network models have previously been developed by Habarulema *et al.* [2007] to estimate the daily and hourly GPS-derived total electron content (TEC) at several geographic locations within South Africa. In an attempt to alleviate some of the previously mentioned difficulties associated with neural networks and other common modeling approaches, the Gaussian process framework was used to model the GPS-derived TEC at 12h00 SAST (South African Standard Time), which corresponds to 10h00 UT (Universal Time). More specifically, the Gaussian process modeling approach was applied to the task of estimating the daily GPS TEC over Sutherland (see Figure 2), South Africa (32.38°S, 20.81°E).

[47] Parameters which have previously been shown to influence GPS TEC include the solar activity, geographical receiver location, satellite position, and magnetic activity as well as the time of day [Habarulema *et al.*, 2007; Reddy, 2002]. However, to keep our presentation simple we will only consider a daily GPS TEC model for a single, fixed geographic receiver location.

3.1. Model Purpose

[48] The purpose of the GPS TEC model is simply to provide an accurate estimate of the GPS-derived TEC when only certain input parameters are known. That is, when we have access to some related observations (which are perhaps

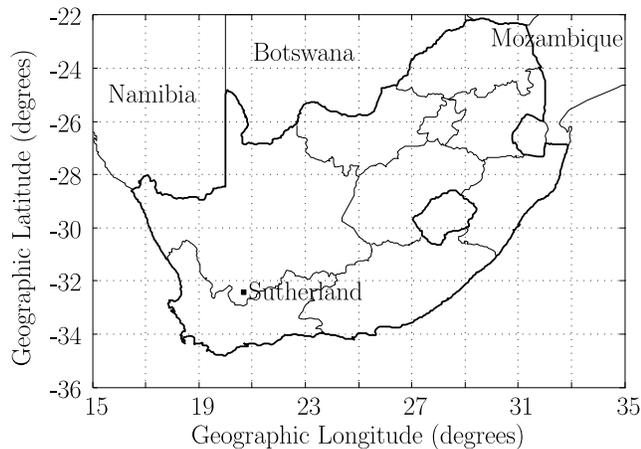


Figure 2. GPS receiver location: Sutherland, South Africa.

easier to obtain than the actual GPS TEC itself), the model should provide an accurate estimate of the GPS TEC.

3.2. Model Inputs

[49] The Gaussian process model is shown in Figure 3 with the day number (DNS and DNC), long-term solar activity ($R4$), and magnetic activity ($A8$) as inputs and the estimated GPS-derived TEC as output. The development of the model input space was taken from *Habarulema et al.* [2007] in order to allow for a direct and objective comparison between the existing neural network model [*Habarulema et al.*, 2007] and the Gaussian process model developed in this paper.

[50] The solar activity is characterized by diurnal, seasonal as well as long-term periodic variations, although the diurnal variations are ignored since we only concern ourselves with a daily GPS TEC model. The seasonal variation is represented by the day number (DN), which is made continuous by applying the following nonlinear transformation:

$$\text{DNS} = \sin\left(\frac{2\pi \times \text{DN}}{365.25}\right), \quad \text{DNC} = \cos\left(\frac{2\pi \times \text{DN}}{365.25}\right). \quad (10)$$

By representing the day number using both sine and cosine components as in (10) above, we have effectively established a constant Euclidean distance between consecutive days in the transformed input space ($\sqrt{\text{DNS}^2 + \text{DNC}^2} = 1$).

[51] The long-term solar activity is commonly represented by the sunspot number (SSN), and in this instance the 4-month running mean ($R4$) of the daily SSN is used. The magnetic index ($A8$) is derived from the K index, which is determined from magnetic flux variations measured at the Hermanus Magnetic Observatory (HMO) in South Africa. In this sense it is a regional measure of the geomagnetic field variations resulting from induced currents in the ionosphere and the magnetosphere.

[52] More specifically the A index is obtained from 3-hourly values of the K -index which has an associated (linearized) a_k value. In this way $A8$ is defined as the value corresponding to the mean (average) of the previous 8 a_k values, which is representative of the previous 24 h.

[53] The derivations of the solar and magnetic activity indices are perhaps rather strange in that the running mean

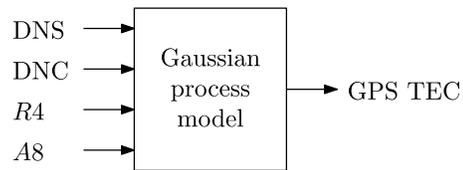


Figure 3. Daily GPS TEC Gaussian process model.

assigns an equal weight to each of the data points. For the solar activity with a k -day running mean this would suggest that the GPS TEC is equally dependent on each of the k previous days but entirely independent of all the days before that.

[54] It would probably make more physical sense to consider an exponentially weighted moving average (EWMA) for example, but we have used the windowed running mean employed by *Habarulema et al.* [2007] to keep a direct comparison between the existing neural network model and the Gaussian process model presented here meaningful.

3.3. Target Output

[55] The same procedure was also followed to obtain the target GPS-derived vertical TEC (VTEC) measurements (under specific input conditions) as employed by *Habarulema et al.* [2007]. The VTEC over Sutherland at 12h00 SAST was derived by interpolation of bias-corrected slant TEC observations (defined as the TEC along the raypath from the satellite to the receiver) sampled at 60 s intervals over a 24 h period. Data for all satellites visible from only the GPS receiver at Sutherland were used in the interpolation. The slant to vertical TEC mapping function which was applied at each ionospheric pierce point (IPP) prior to interpolation and the adjusted spherical harmonic interpolation algorithm are described in [*Opperman et al.*, 2007]. The VTEC at 12h00 SAST for each day was extracted from the resulting 24 h diurnal TEC interpolation [*Habarulema et al.*, 2007], and is shown in Figure 4 as the target VTEC, henceforth referred to as the daily GPS TEC.

3.4. Data Collection and Preparation

[56] The same data was used as in the work of *Habarulema et al.* [2007] and there was no opportunity to influence the data acquisition process or the preprocessing of the data. Several parameters were recorded over Sutherland, South

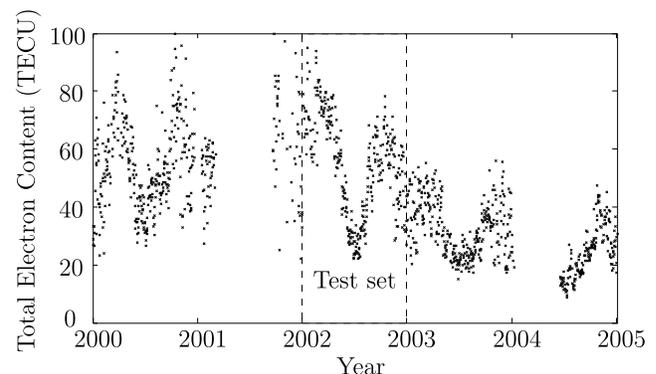


Figure 4. Daily GPS TEC for Sutherland, South Africa.

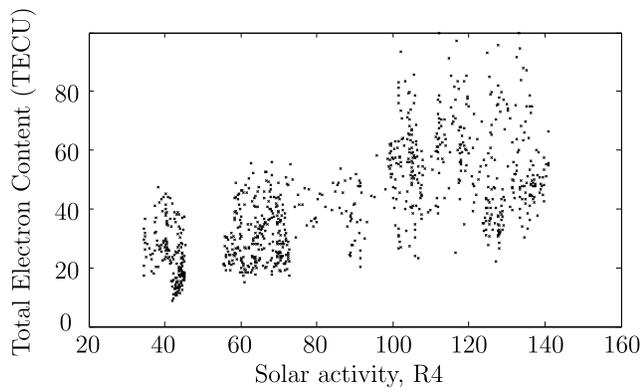


Figure 5. Daily GPS TEC variation as a function of solar activity, $R4$.

Africa (32.38°S , 20.81°E) on an almost daily basis, including the sunspot number, the day number, the magnetic K index, and the GPS-derived TEC at 12h00 SAST. The observed GPS TEC for the entire data set (both testing and training) is shown in Figure 4.

[57] A total of 1268 data points are available for the 5 year period ranging from 2000 to 2004, with some missing observations (primarily during 2001 and 2004). From the 1268 data points, 329 (corresponding to 2002) are kept out for testing and validation purposes, which leaves 939 data points spread across the remaining years (2000, 2001, 2003, and 2004) for training. This same strategy was followed as outlined by *Habarulema et al.* [2007].

3.5. Preliminary Data Analysis

[58] Before constructing a Gaussian process model (or any model, for that matter), it is usually a good idea to get a good overview of the available data and to try to determine how the output varies with changes in different input parameters; this will typically enable us to specify meaningful prior information later on in the model selection process.

[59] With reference to Figure 4, it is noticeable that there seems to be an annual seasonal variation, with peak electron densities at around mid-April and mid-October each year. The GPS TEC also appears to have a general downward trend throughout the duration of the data set. This downward trend could have been expected, since 2001 was the last solar maximum year in cycle 23. In contrast, 2004 was a solar minimum year.

[60] If the GPS TEC is considered purely as a function of the average sunspot number ($R4$), we can see that there seems to be a slight upward trend (see Figure 5). That is, as the sunspot number (or solar activity) is increased, the GPS TEC also increases. This is perhaps also an expected result, since it is well known that solar radiation is the primary cause of ionization within the ionosphere.

[61] If we consider the magnetic index ($A8$) in isolation, we obtain the results shown in Figure 6 from which we might reasonably expect $A8$ to be a much less important input than either the solar activity or the day number. That is, there is no immediately apparent trend or relationship between the magnetic index, $A8$, and GPS TEC. Nevertheless, the magnetic index may still contain valuable information about the GPS-derived TEC when considered jointly

with the other input parameters and should therefore not be discarded immediately.

3.6. Model Selection

[62] Two relatively simple Gaussian process models will be presented in this section mainly for comparative purposes, and it is expected that even better models can be obtained by incorporating more sophisticated prior knowledge. Recall that the task of model selection in the Gaussian process framework involves specifying the structure of the underlying covariance function. Arguably, the most important and widely used covariance function (in the Gaussian process context at least) is the simple squared exponential (SE) function, which has the form

$$k_{\text{SE}}(r) = \exp\left(-\frac{r^2}{2\ell^2}\right), \quad (11)$$

with parameter ℓ defining the characteristic length scale of the input dimension and where r is the distance in input space.

[63] The SE covariance function is infinitely differentiable, which means that a Gaussian process with this covariance function has mean square derivatives of all orders, making it very smooth. It has been argued that such strong smoothness assumptions are unrealistic for modeling physical processes and that the Matérn class of covariance functions should perhaps rather be used [*Rasmussen and Williams, 2006*]. Nevertheless, to keep the presentation simple, the well-known SE covariance function was used to model the daily GPS TEC.

3.6.1. Single Kernel Gaussian Process Model

[64] As an initial attempt to model the daily GPS TEC as a function of the day number (DN), the sunspot number ($R4$), and the magnetic index ($A8$), a Gaussian process with a single squared exponential covariance function and an independent noise term was chosen. More specifically, the covariance function was specified as

$$k(\mathbf{x}, \mathbf{x}') = \theta_5 \exp\left(-\sum_{i=1}^4 \left(\frac{(x_i - x'_i)^2}{2\theta_i^2}\right)\right) + \theta_6 \delta_{\mathbf{x}\mathbf{x}'}, \quad (12)$$

with $\theta_1 \dots \theta_4$ corresponding to the characteristic length scales of the four inputs, $\mathbf{x} = [\text{DNS}, \text{DNC}, R4, A8]^T$, θ_5 the

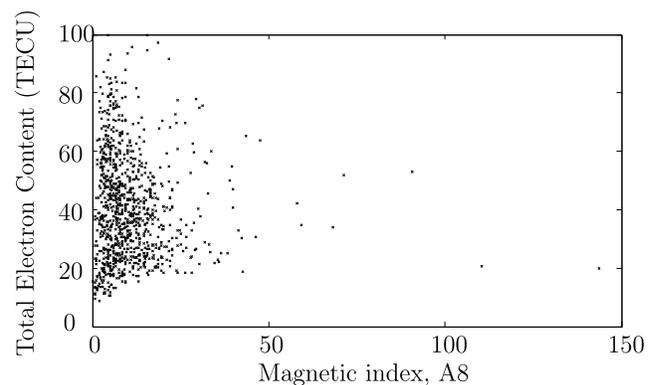


Figure 6. Daily GPS TEC variation as a function of magnetic activity, $A8$.

Table 1. RMSE Performance of Complex Single Squared Exponential Gaussian Process Model^a

Year	IRI RMSE	NN RMSE	GP RMSE	Improvement (%)
2000	12.30	10.28	7.16	30.37
2001	15.59	12.30	7.91	35.73
2002*	13.29	9.09	12.01	-32.16
2003	7.27	6.46	4.41	31.72
2004	5.73	4.40	2.96	32.70

^aYears marked by an asterisk are not included in the training set.

signal variance, and θ_6 the noise variance. Notice that (12) can equivalently be expressed in matrix form as

$$k(\mathbf{x}, \mathbf{x}') = \theta_5 \exp\left(-\frac{(\mathbf{x} - \mathbf{x}')^T \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}')}{2}\right) + \theta_6 \delta_{\mathbf{x}\mathbf{x}'}, \quad (13)$$

where \mathbf{P} is a diagonal matrix with the automatic relevance determination (ARD) parameters $\theta_1^2, \dots, \theta_4^2$ along the main diagonal. The hyperparameters are then simply given as

$$\theta = \begin{bmatrix} \sqrt{\mathbf{P}_{11}} \\ \sqrt{\mathbf{P}_{22}} \\ \sqrt{\mathbf{P}_{33}} \\ \sqrt{\mathbf{P}_{44}} \\ \theta_5 \\ \theta_6 \end{bmatrix} = \begin{bmatrix} \theta_{\text{DNS}} \\ \theta_{\text{DNC}} \\ \theta_{\text{R4}} \\ \theta_{\text{A8}} \\ \sigma_f \\ \sigma_n \end{bmatrix}, \quad (14)$$

and they can be obtained by performing gradient-based descent (or any other suitable optimization strategy) on the maximum-likelihood surface using the 939 training data points from the period 2000–2004. We might reasonably expect that such a simple model might not fit our particular problem very well, and indeed this observation has motivated the development of a more advanced and more flexible model, presented next.

3.6.2. Composite Gaussian Process Model

[65] A more flexible model can be obtained by increasing the number of hyperparameters which characterizes the model. This can be achieved by changing or extending the covariance function. However, more hyperparameters do not always lead to better models; ideally, the covariance function (and therefore also the hyperparameters) should be specified in such a way as to reflect our prior knowledge of the process. For example, since we expect the GPS TEC to have a strong seasonal component, we could probably include a periodic term in the covariance function. Nevertheless, we will only consider a simple extension to the single squared exponential covariance function here.

[66] Consider the composite covariance function obtained by summing three squared exponential covariance functions together with an independent noise term:

$$k(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^3 \theta_i \exp\left(-\sum_{j=1}^4 \left(\frac{(x_j - x'_j)^2}{2\theta_{4i-1+j}^2}\right)\right) + \theta_{16} \delta_{\mathbf{x}\mathbf{x}'}, \quad (15)$$

so that the newly obtained Gaussian process with (15) as its covariance function is specified by 16 hyperparameters, $\theta = [\theta_1, \theta_2, \dots, \theta_{16}]^T$, which once again can be determined by optimization on the likelihood surface. We may now want to answer the key question about which model is

better: the single kernel or the composite Gaussian process model?

3.7. Model Validation

[67] During model validation we should take the model falseness, plausibility, and purposiveness into account, which unfortunately means that model validation often becomes an iterative and subjective process. Indeed, it is not always immediately apparent which model is superior or even desirable for any given context.

[68] When discussing model falseness, we will consider only the RMSE, primarily to facilitate a direct comparison with the existing neural network models, but also to highlight the fact that a low RMSE does not necessarily guarantee a good model; all the criteria for model validation should be considered jointly.

[69] The results of three models will be presented in this section, namely an implausible, single kernel Gaussian process model with a very good training error but a bad validation error, followed by a fairly good (and plausible) composite Gaussian process model, and finally a plausible single kernel Gaussian process model with a respectable training and validation error. In each case, the RMSE is given as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{TEC}_{\text{model}} - \text{TEC}_{\text{target}})^2}, \quad (16)$$

where N is the number of data points, $\text{TEC}_{\text{model}}$ is the output of the model, and $\text{TEC}_{\text{target}}$ is the TEC derived from GPS observations (as described in section 3.3) and the percentage improvement (of the GP model over the NN model) is defined as

$$\text{Improvement} = \frac{\text{NN RMSE} - \text{GP RMSE}}{\text{NN RMSE}} \times 100 (\%). \quad (17)$$

3.7.1. Complex Single SE Gaussian Process Model

[70] The hyperparameters for the single squared exponential Gaussian process given in (12) was determined using all 939 training data points, and the RMSE performance of the model is given in Table 1. Note that from the training error (corresponding to 2000, 2001, 2003, and 2004) the Gaussian process model appears to be superior to both the IRI and the neural network models. However, for the year of 2002 (which is not contained in the training set) the performance of the Gaussian process model seems relatively poor. Nevertheless, the Gaussian process model exhibits good model falseness, with an average improvement over the neural network model of 19.67%.

[71] To establish whether the Gaussian process model is plausible, we need to inspect the model behavior for a number of typical input scenarios. This model behavior is shown in Figure 7 where the magnetic index has been set to zero. (Note that even though we cannot ever expect the magnetic index to be zero, the model behavior is very similar for almost any fixed magnetic index.)

[72] We can clearly see that the model behavior presented in Figure 7 is overly complex: we cannot reasonably expect the GPS TEC to fluctuate in such a seemingly random fashion with only a slight increase in solar activity or from one day

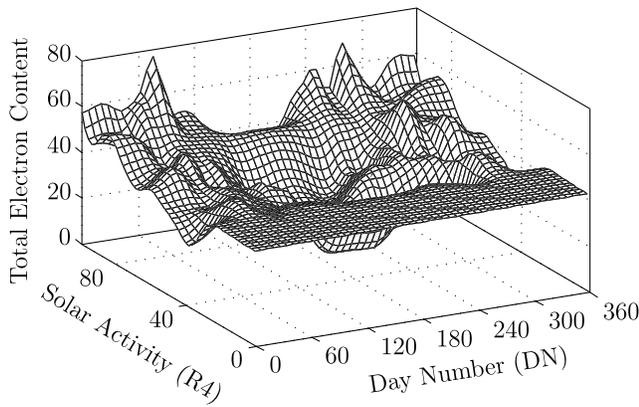


Figure 7. Complex single squared exponential Gaussian process model behavior.

to the next. In addition, for low solar activity regions the model simply reverts back to its prior expectation (the mean GPS TEC), which is expected since there were no training examples for these regions, but it is not plausible behavior.

[73] The actual GPS TEC as well as the GPS TEC estimates returned by the complex (implausible) Gaussian process model are shown in Figure 8, from where it is clear that the model does not provide an accurate estimate of the GPS TEC during the year of 2002. That is, the model is not purposive for estimating the GPS TEC. The observation that the above Gaussian process model is both implausible and nonpurposive motivated the development of the more flexible composite Gaussian process model, presented next.

3.7.2. Composite SE Gaussian Process Model

[74] The RMSE performance of the slightly more flexible composite Gaussian process model given in (15) is presented in Table 2, with an average improvement of 29.81% over the neural network model. We also notice that the RMSE performance during each year (including the previously unseen year of 2002) is better than for the implausible model presented previously, but the RMSE of 2002 is not as good as for the neural network model. Nevertheless, the model falseness is considered very good, and the composite Gaussian process model is clearly superior (in terms of model falseness) to both the IRI and the neural network models.

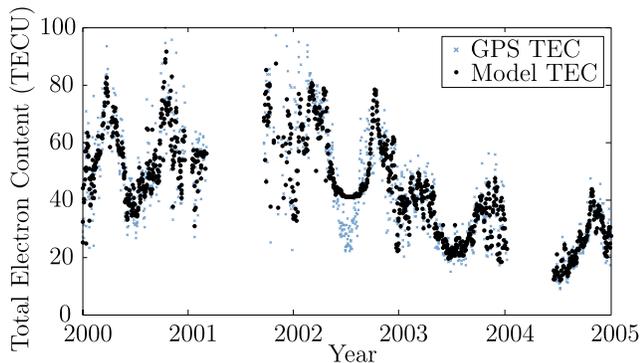


Figure 8. GPS TEC estimates using complex single squared exponential Gaussian process model.

Table 2. RMSE Performance of Composite Squared Exponential Gaussian Process Model^a

Year	IRI RMSE	NN RMSE	GP RMSE	Improvement (%)
2000	12.30	10.28	6.43	37.44
2001	15.59	12.30	7.48	39.18
2002*	13.29	9.09	10.23	-12.55
2003	7.27	6.46	3.81	40.98
2004	5.73	4.40	2.46	44.01

^aYears marked by an asterisk are not included in the training set.

[75] The behavior of the composite Gaussian process model is shown in Figure 9, where the magnetic index has once again been set to zero. We immediately notice that the model behavior is much simpler than shown in Figure 7 (corresponding to the implausible single kernel Gaussian process model). The seasonal variation is relatively smooth, as expected, and the model TEC increases monotonically as the solar activity is increased; in other words, the model seems plausible (even though the magnetic index is seldom truly zero).

[76] The model purposiveness is established by comparing the actual GPS TEC with the modeled TEC. This is shown in Figure 10, where we can see that the composite Gaussian process model performs much better during the previously unseen year of 2002 than the implausible single kernel Gaussian process model of Figure 8.

3.7.3. Plausible Single SE Gaussian Process Model

[77] We have already seen that the composite Gaussian process model is superior to the single kernel Gaussian process in every aspect of model validation process. However, we can in fact adjust the implausible single kernel Gaussian process model so as to compete with the composite Gaussian process model.

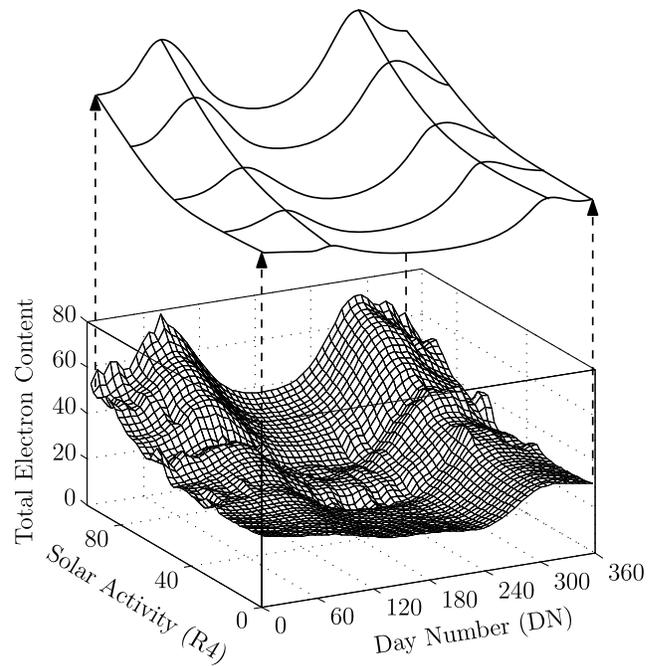


Figure 9. Composite squared exponential Gaussian process model behavior.

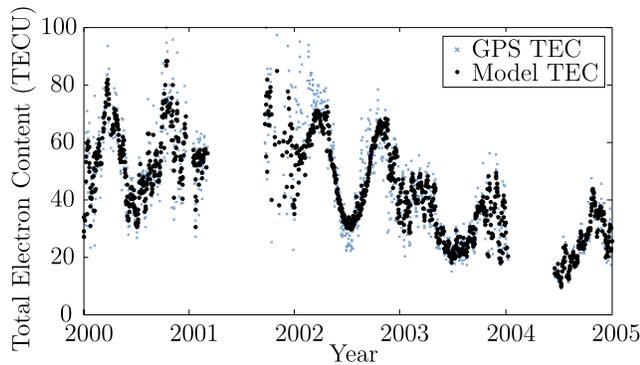


Figure 10. GPS TEC estimates using composite squared exponential Gaussian process model.

[78] From Figure 7 we might suspect that the single kernel Gaussian process model varies too rapidly with changes in the solar activity. By carefully adjusting the hyperparameter corresponding to the solar activity, we can obtain more plausible model behavior. The newly obtained model behavior seems plausible enough as shown in Figure 11, even though it is perhaps not as intuitive as the behavior of the composite Gaussian process model of Figure 9.

[79] The model falseness (in terms of RMSE) for the adjusted single kernel Gaussian process is given in Table 3, with an average improvement of 11.7% over the neural network. We also notice that the model falseness for the unseen year of 2002 is the best of all the models presented so far.

[80] The model TEC is shown in Figure 12, from which we can easily see that the model TEC corresponds fairly well with the actual GPS TEC for the training set (2000, 2001, 2003, 2004) as well as the test set (2002). The model is therefore considered purposive.

[81] Finally, we may want to decide which model is better: the plausible composite Gaussian process model which has a lower overall RMSE or the plausible single kernel Gaussian process model which exhibits the smallest RMSE during the previously unseen year of 2002. The final choice, however, is highly subjective and may very well depend on the exact requirements of the model. Suffice it to say that

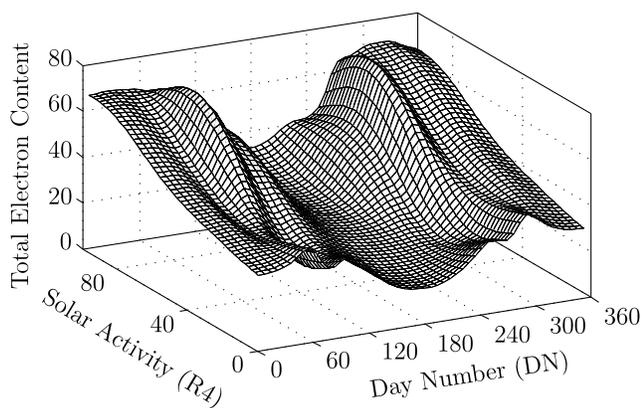


Figure 11. Plausible single squared exponential Gaussian process model behavior.

Table 3. RMSE Performance of Plausible Single Squared Exponential Gaussian Process Model^a

Year	IRI RMSE	NN RMSE	GP RMSE	Improvement (%)
2000	12.30	10.28	8.78	14.56
2001	15.59	12.30	10.48	14.78
2002*	13.29	9.09	9.78	-7.55
2003	7.27	6.46	5.27	18.35
2004	5.73	4.40	3.57	18.81

^aYears marked by an asterisk are not included in the training set.

both models exhibit good model falseness, are considered plausible, and have been shown to be purposive.

4. Predicting Daily Total Electron Content

[82] The GPS TEC models that were presented in section 3, as well as the existing models developed by *Habarulema et al.* [2007], all depend on input parameters which are generally not known in advance; this makes it both difficult and inconvenient to use these models for predicting GPS TEC. Indeed, to use the existing models in a predictive mode would first require the development of several additional models, each capable of predicting a sequence of input parameters, after which the predicted input parameters can be used to estimate the corresponding GPS TEC.

[83] In this section a simple alternative is presented: a Gaussian process model which is designed specifically for the task of predicting GPS TEC. This is done primarily to demonstrate how the GP framework can be used to perform prediction (even though we will only consider the easy task of one-step ahead prediction here) and to demonstrate the importance of considering nonstationary covariance functions when predicting dynamic processes.

4.1. Model Purpose

[84] The purpose of the predictive time series model is of course to predict GPS TEC as accurately and as easily as possible, without having to predict the input parameters first. In addition, the model should be able to adapt to previously unseen trends without having to retrain the model.

4.2. Model Inputs

[85] Several suitable input parameters probably exist, but a common (and quite natural) choice for predictive models

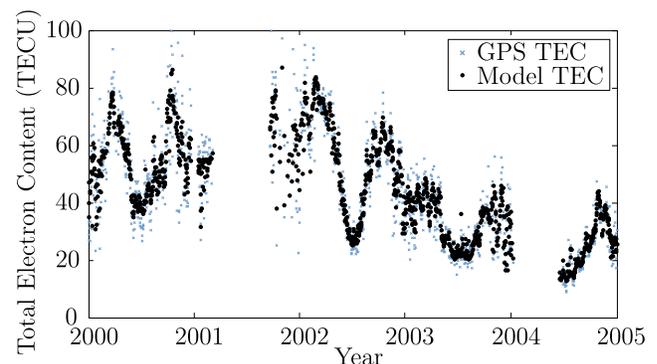


Figure 12. GPS TEC estimates using plausible single squared exponential Gaussian process model.

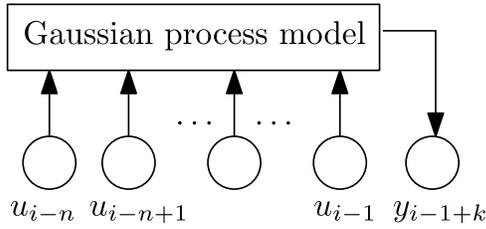


Figure 13. Illustration of an n th order k -step ahead predictive GPS TEC Gaussian process model.

is to simply use past and present target observations as the inputs to the model. Of course, additional sources of information can be incorporated very easily in the Gaussian process framework, but we will only consider target observations as inputs here.

[86] The predictive GPS TEC model is shown in Figure 13, where n past observations, $u_{i-n}, u_{i-n+1}, \dots, u_{i-1}$, are used to predict the target output, y_{i-1+k} . Such a model is generally referred to as an n th order k -step ahead predictor or sometimes simply as a (possibly nonlinear) time series model. We will limit our attention to one-step ahead ($k = 1$) predictors only.

4.3. Data Collection and Preparation

[87] The same data was again used as described by *Habarulema et al.* [2007], and we did not have the opportunity to influence the data acquisition process or the pre-processing of the data. However, since we require $n + 1$ consecutive observations (n inputs, 1 output) for each training sample in an n th order one-step ahead predictive model, we have to discard regions with too many missing data points.

4.4. Model Selection

[88] There are many possible choices of prior covariance functions, but we will again restrict our attention to using a simple squared exponential covariance function with the express aim of demonstrating a very important consideration in model selection, namely choosing between stationary and nonstationary covariance functions.

[89] The issue of stationarity (or lack thereof) becomes especially important in the construction of predictive models, where it is often expected that a predictive model should be able to adapt to unforeseen changes in the process without having to retrain the model. This is also true in our current example, since we might reasonably want to train the predictive model only on an initial data set corresponding to 2000, and we would expect the model to work throughout the remaining years (2001–2005 and beyond).

[90] A simple Gaussian process approach to predict nonstationary time series is presented in the work of *Brahim-Belhouari and Bermak* [2004], and a similar strategy will be followed here. We begin by defining the (stationary) squared exponential covariance function as

$$k_s(\mathbf{x}, \mathbf{x}') = \sigma_f \exp\left(-\sum_{i=1}^n \left(\frac{(x_i - x'_i)^2}{2\theta_i^2}\right)\right) + \sigma_n \delta_{\mathbf{x}\mathbf{x}'}, \quad (18)$$

where σ_f and σ_n are the signal and noise variances, respectively, and $\theta_1, \theta_2, \dots, \theta_n$ are the hyperparameters cor-

responding to the inputs $u_{i-n}, u_{i-n+1}, \dots, u_{i-1}$. This will be our stationary model.

[91] Next we consider what is arguably the simplest nonstationary covariance function, namely the linear trend:

$$k_{ns}(\mathbf{x}, \mathbf{x}') = \theta_0 + \theta_1 \sum_{i=1}^n x_i x'_i \quad (19)$$

with hyperparameters θ_0 and θ_1 . However, we can hardly expect the GPS TEC to be modeled accurately simply with a linear trend, so we have to consider a slightly more sophisticated covariance function.

[92] By adding the stationary squared exponential covariance function (18) to the nonstationary linear trend (19) we obtain the composite, nonstationary function

$$k(\mathbf{x}, \mathbf{x}') = k_s(\mathbf{x}, \mathbf{x}') + k_{ns}(\mathbf{x}, \mathbf{x}'), \quad (20)$$

which is itself guaranteed to be a valid covariance function (a valid covariance function is any function which will generate a nonnegative definite covariance matrix for any set of input points [*Brahim-Belhouari and Bermak*, 2004]).

4.5. Model Validation

[93] We will once again evaluate the Gaussian process models (both stationary and nonstationary) based on model falseness, plausibility, and purposiveness.

4.5.1. Stationary Predictive Model

[94] We first present the results of the stationary Gaussian process model, with covariance function given by (18). A sixth-order model was trained on the first $N = 75$ samples (each consisting of seven consecutive observations), which corresponds roughly to the first 5 months of 2000.

[95] As we might have expected, the predictive model performs quite poorly (see Figure 14). Also note that the model GPS TEC never goes below about 40 TECU, even though the actual GPS TEC often lies between about 15 and 40 TECU during the years of 2002–2005.

[96] In fact, it would appear as though, when faced with observations outside of the training set range, the model simply reverts back to the average of the training set, which is roughly 50 TECU. Clearly then, the stationary model fails to adapt to unforeseen changes in the process and can therefore not be deemed plausible or purposive.

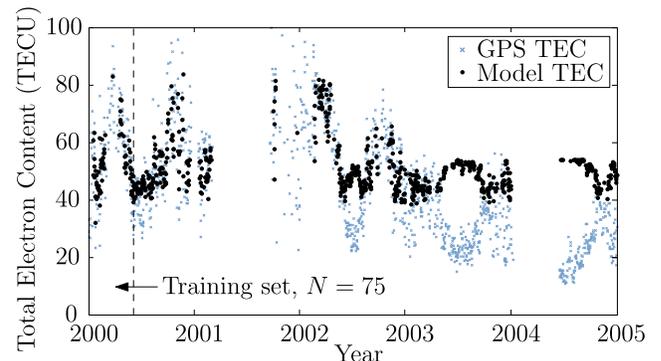


Figure 14. Stationary sixth-order Gaussian process predictive model; $N = 75$ leading training samples.

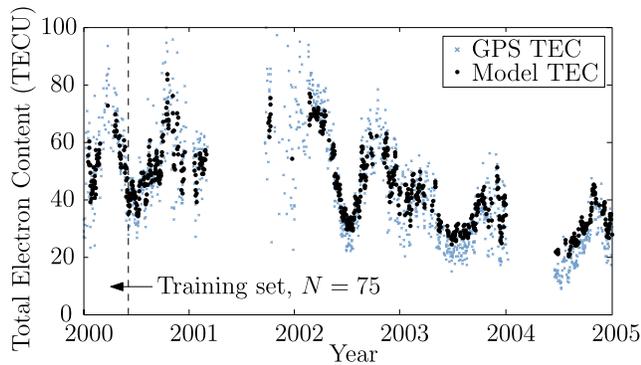


Figure 15. Nonstationary sixth-order Gaussian process predictive model; $N = 75$ leading training samples.

4.5.2. Nonstationary Predictive Model

[97] In contrast to the stationary time series model, the sixth-order nonstationary Gaussian process model with covariance function given by (20) performs quite well, as shown in Figure 15. Furthermore, the nonstationary model seems to adapt to unforeseen changes in the process, which makes it both plausible and purposive.

[98] Even though the nonstationary Gaussian process model can already be seen to be much better than the stationary model, we can further improve the model's fidelity by considering a slightly larger training set (corresponding to $N = 150$ or roughly 7 months during 2000). The results of this new nonstationary model is shown in Figure 16.

[99] The RMSE performance for the nonstationary model (with $N = 150$) is summarized in Table 4, from where it can be seen that in terms of model falseness at least, this model performs better than any other model presented so far, with an average improvement over the neural network model of 28.96% during the 4 years not included in the training set (2001–2004). Of course, this nonstationary model can easily be seen to be both plausible and purposive (see Figure 16), so that we have finally succeeded in our task of developing a simple yet powerful predictive GPS TEC model.

5. Summary and Conclusion

[100] In this paper we have proposed the Gaussian process framework as an attractive alternative to the neural network

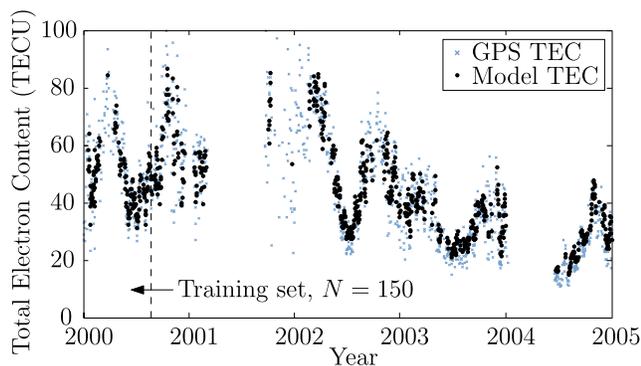


Figure 16. Nonstationary sixth-order Gaussian process predictive model; $N = 150$ leading training samples.

Table 4. RMSE Performance of Sixth-Order Predictor Model, $N = 150$

Year	IRI RMSE	NN RMSE	GP RMSE	Improvement (%)
2000	12.30	10.28	7.20	29.96
2001	15.59	12.30	6.28	48.93
2002	13.29	9.09	5.98	34.26
2003	7.27	6.46	5.27	18.42
2004	5.73	4.40	3.77	14.24

modeling approach for GPS-derived TEC. The Gaussian process framework presents many advantages over competing modeling strategies, such as providing powerful and convenient ways of incorporating prior knowledge and requiring less training data than neural networks.

[101] It was shown that within the principled framework of Gaussian processes it is relatively easy to develop simple Gaussian process models with similar or better model falseness than existing neural network models. In addition, it is expected that by specifying more complex covariance functions the model falseness of the Gaussian process models can be improved even further. It was also shown that model falseness should not be used as the only criterion on which to evaluate empirical models, but that model plausibility and purposiveness should play an equally important role in model validation.

[102] Finally, it was shown that the Gaussian process framework can also be used to develop predictive models which are better suited to the task of predicting GPS TEC than the existing neural network models of *Habarulema et al.* [2007]. Nevertheless, only the one-step ahead predictor was considered here, since it is unlikely that a k -step ahead predictor will perform very well on the current data set due to the very large variability of GPS TEC from one day to the next. However, the GP model can easily be extended to incorporate additional information (such as hourly GPS TEC data or a host of ancillary information), in which case it is expected that the GP framework will provide an attractive means of performing k -step ahead prediction.

[103] **Acknowledgments.** The authors would like to thank John Bosco Habarulema at the Hermanus Magnetic Observatory for making the GPS TEC data available that were used in this study, as well as the anonymous reviewers for their valuable feedback and suggestions for improvements to this paper.

[104] Philippa Browning thanks the reviewers for their assistance in evaluating this paper.

References

- Arikan, F., C. B. Erol, and O. Arikan (2003), Regularized estimation of vertical total electron content from Global Positioning System data, *J. Geophys. Res.*, *108*(A12), 1469, doi:10.1029/2002JA009605.
- Azman, K., and J. Kocijan (2007), Application of Gaussian processes for black-box modelling of biosystems, *ISA Trans.*, *46*, 443–457.
- Bilitza, D. (2001), International Reference Ionosphere 2000, *Radio Sci.*, *36*(2), 261–275.
- Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Oxford Univ. Press, New York.
- Brahim-Belhouari, S., and A. Bermak (2004), Gaussian process for nonstationary time series prediction, *Comput. Stat. Data Anal.*, *47*, 705–712.
- Goodman, J. M., and J. Aarons (1990), Ionospheric effects on modern electronic systems, *Proc. IEEE*, *78*(3), 512–527.
- Gregorčič, G., and G. Lightbody (2008), Nonlinear systems identification: From multiple-model networks to Gaussian processes, *Eng. Appl. Artificial Intell.*, *21*(7), 1035–1055.

- Habarulema, J. B., L. A. McKinnel, and P. J. Cilliers (2007), Prediction of global positioning system total electron content using neural networks over South Africa, *J. Atmos. Sol. Terr. Phys.*, *69*, 1842–1850.
- Habarulema, J. B., L. A. McKinnel, P. J. Cilliers, and B. D. L. Opperman (2009), Application of neural networks to South African GPS TEC modelling, *Adv. Space Res.*, *43*, 1711–1720.
- Hernandez-Pajares, M., J. Juan, and J. Sanz (1997), Neural network modelling of the ionospheric electron content at global scale using GPS, *Radio Sci.*, *32*, 1081–1090.
- Hoffman-Wellenhof, B., H. Lichtenegger, and J. Collins (1992), *Global Positioning System Theory and Practice*, Springer, New York.
- Howe, B. M., K. Runciman, and J. A. Secan (1998), Tomography of the ionosphere: Four dimensional simulations, *Radio Sci.*, *33*(1), 109–128.
- Johansen, T. A., and B. A. Foss (1997), Operating regime based process modeling and identification, *Comput. Chem. Eng.*, *21*, 159–176.
- Kocijan, J., B. Banko, B. Likar, A. Girard, R. Murray-Smith, and C. E. Rasmussen (2003), A case based comparison of identification with neural network and Gaussian process models, paper presented at International Conference on Intelligent Control Systems and Signal Processing, Int. Fed. of Auto. Control, Laxenburg, Austria.
- Kouris, S. S., T. D. Xenos, K. V. Polimeris, D. Stergiou (2004), D. TEC and foF2 variations: Preliminary investigations, *Ann. Geophys.*, *47*(4), 1325–1332.
- Leandro, R. F., and M. C. Santos (2007), A neural network approach for regional vertical total electron content modelling, *Stud. Geophys. Geod.*, *51*(2), 279–292.
- MacKay, D. J. C. (1992), A practical Bayesian framework for backprop networks, *Neural Comput.*, *4*(3), 448–472.
- MacKay, D. J. C. (1997), Introduction to Gaussian processes, technical report, Univ. of Cambridge, Cambridge, U. K.
- McKinnel, L. A. (2002), A neural network based ionospheric model for the bottomside electron density profile over Grahamstown, South Africa, Ph.D. thesis, Rhodes Univ., Grahamstown, South Africa.
- Murray-Smith, D. J. (1998), Methods for the external validation of continuous system simulation models: A review, *Math. Comput. Modell. Dyn. Syst.*, *4*, 5–31.
- Murray-Smith, R., and T. A. Johansen (Eds.) (1997), *Multiple Model Approaches to Modelling and Control*, Taylor and Francis, London.
- Murray-Smith, R., T. A. Johansen, and R. Shorten (1999), On transient dynamics, off-equilibrium behaviour and identification in blended multiple model structures, paper presented at European Control Conference, EUCA, Karlsruhe, Germany.
- Narendra, K. S., and K. Parthasarathy (1990), Identification and control of dynamical systems using neural networks, *IEEE Trans. Neural Networks*, *1*(1), 4–27.
- Neal, R.M. (1997), Monte Carlo implementation of Gaussian process models for Bayesian regression and classification, *Tech. Rep. TR-9702*, Dept. of Statistics, Univ. of Toronto, Toronto, Ontario, Canada.
- Nelles, O. (2001), *Nonlinear System Identification*, Springer, Heidelberg, Germany.
- Olivera, M. A., and R. Webster (1990), Kriging: A method of interpolation for geographical information systems, *Int. J. Geogr. Inf. Sci.*, *4*(3), 313–332.
- Opperman, B. D. L., P. J. Cilliers, L. A. McKinnel, and R. Haggard (2007), Development of a regional GPS-based ionospheric TEC model for South Africa, *Adv. Space Res.*, *39*, 808–815.
- Rasmussen, C. E., and C. K. I. Williams (2006), *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass.
- Reddy, C. D. (2002), Study of space weather effects using GPS, paper presented at 2002 GPS in Atmospheric Sciences, India Int. Cent., New Delhi, India.
- Schunk, R. W., J. J. Sojka, and M. D. Bowline (1986), Theoretical study of the electron temperature in the high latitude ionosphere for solar maximum and winter conditions, *J. Geophys. Res.*, *91*, 11,281–11,290.
- Spalla, P., and L. Ciraolo (1994), TEC and foF2 comparison, *Ann. Geophys.*, *37*(5), 929–938.
- Shorten, R., R. Murray-Smith, and R. Bjorgan (1999), On the interpretation of local models in blended multiple model structures, *Int. J. Control*, *72*(7–8), 620–628.
- Takagi, T., and M. Sugeno (1985), Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Syst. Man Cybernetics*, *15*(1), 116–132.
- Tulunay, E., E. T. Senalp, L. R. Cander, Y. K. Tulunay, A. H. Bilge, E. Mizrahi, S. S. Kouris, and N. Jakowski (2004), Development of algorithms and software for forecasting, nowcasting and variability of TEC, *Ann. Geophys.*, *47*(2–3), 1201–1214.
- Tulunay, E., E. T. Senalp, S. M. Radicella, and Y. K. Tulunay (2006), Forecasting total electron content maps by neural network technique, *Radio Sci.*, *41*, RS4016, doi:10.1029/2005RS003285.
- Xenos, T. D., S. S. Kouris, and A. Casimiro (2003), Time-dependent prediction degradation assessment of neural-networks-based TEC forecasting models, *Proc. Geophys.*, *10*, 585–586.
- Yilmaz, A., K. E. Akdogan, and M. Gurun (2009), Regional TEC mapping using neural networks, *Radio Sci.*, *44*, RS3007, doi:10.1029/2008RS004049.

E. R. Ackermann and J. P. de Villiers, Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Pretoria 0002, South Africa. (etienne.ackermann@ieec.org)

P. J. Cilliers, Hermanus Magnetic Observatory, PO Box 32, Hermanus 7200, South Africa.