(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 July 2010 (15.07.2010)

PCT

(10) International Publication Number
WO 2010/078615 A1

(51) International Patent Classification:
*G06F 17/18* (2006.01) *G01V 1/50* (2006.01)
*G06F 17/00* (2006.01) *G06N 5/02* (2006.01)
*G06N 5/04* (2006.01)

(21) International Application Number:
PCT/AU2009/001668

(22) International Filing Date:
21 December 2009 (21.12.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2009900054 7 January 2009 (07.01.2009) AU

(71) Applicant *(for all designated States except US)*: **THE UNIVERSITY OF SYDNEY** [AU/AU]; Parramatta Road, Sydney, NSW 2006 (AU).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MELKUMYAN, Arman** [AM/AU]; 5/16 Epping Road, Lane Cove, New South Wales 2066 (AU). **RAMOS, Fabio, Tozeto** [BR/AU]; 63/2 Coulson Street, Erskineville, New South Wales 2043 (AU).

(74) Agent: **FREEHILLS PATENT & TRADE MARK ATTORNEYS**; Level 38, MLC Centre, 19-29 Martin Place, Sydney, New South Wales 2000 (AU).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17**:

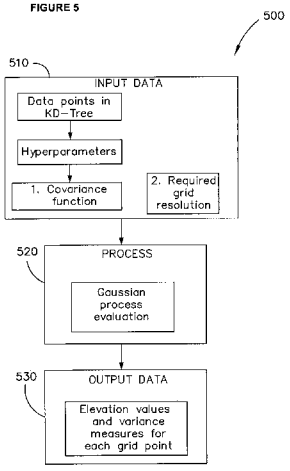
— *of inventorship (Rule 4.17(iv))*

**Published**:

— *with international search report (Art. 21(3))*

(54) **Title**: A METHOD AND SYSTEM OF DATA MODELLING

FIGURE 5




(57) **Abstract**: A system for large scale data modelling is described. The system includes at least one data measurement sensor (230) for generating measured data, a training processor (240) to determine optimized hyperparameter values in relation to a Gaussian process covariance function including a sparse covariance function that is smooth and diminishes to zero outside of a characteristic hyperparameter length. An evaluation processor (260) determines model data from the Gaussian process covariance function with optimised hyperparameter values and measured data. Also described is methods for modelling date, including a method using a Gaussian process including a sparse covariance function that diminishes to zero outside of a characteristic length, wherein the characteristic length is determined from the data to be modelled.

WO 2010/078615 A1

# A Method and System of Data Modelling

## Field of the invention

This invention relates to the field of large scale data modelling. Examples of its application include, but are not limited to spatial data modelling such as, in-ground ore grade estimation for mining and terrain modelling for robotic vehicle navigation.

## Background of the invention

Gaussian Process regression models are used to determine functions that model known data. The known data can be, for example, discrete elevation measurements over a certain area. The model would then provide a means of inferring values for intermediate points that were not measured, called test points.

In a large scale environment like an open-cut mine spatial modelling of geography and geology can have many uses in planning, analysis and operations within the environment. In the case of automated mining, a geographical model or terrain map can be used to guide robotic vehicles, whilst an in-ground geological model of the ore body may be used to determine drilling and blasting operations.

A digital representation of the operating environment in the form of a spatial model is typically generated from sensor measurements which provide a sample of the actual environmental variable being modelled (e.g. elevation in the case of a terrain map, or ore grade in the case of in-ground ore body modelling) at various spatially distinct locations within the operating environment. The measured sample data is then treated in some manner such as by interpolation to determine information about the environment in locations other than those actually measured. Some of the challenges posed by this task include dealing with the issues of uncertainty, incompleteness and handling potentially large measurement data sets.

Gaussian Processes (GPs) are stochastic processes based on the normal (Gaussian) distribution and can be used to good effect as a powerful nonparametric learning technique for modelling. Inference of continuous values with a Gaussian process prior is known as Gaussian process regression, and is sometimes referred to as "Kriging".

Whilst Gaussian processes are a useful and powerful tool for regression in supervised machine learning they are regarded as a computationally expensive technique, which is particularly disadvantageous in the treatment of large measurement data sets. The computational expense is primarily brought on by the need to invert a large covariance matrix during the inference procedure. For problems with thousands of observations, exact inference in normal Gaussian processes is intractable and approximation algorithms are required.

## Summary of the invention

In one embodiment the invention provides a method for modelling data using a Gaussian process including a sparse covariance function that diminishes to zero outside of a characteristic length, wherein the characteristic length is determined from the data to be modelled.

In one embodiment the sparse covariance function is smooth. In another embodiment the sparse covariance function has a continuous fourth derivative. In another embodiment the characteristic length is determined through a machine learning process. In another embodiment the sparse covariance function is multi-dimensional with a separate characteristic length for each dimension. In another embodiment the Gaussian process further includes a non-stationary covariance function. In another embodiment the non-stationary covariance function is a neural network covariance function. In another embodiment the sparse covariance function and non-stationary covariance function are combined in Gaussian process regression computations. In another embodiment the data is measured using one or more measurement sensors. In another embodiment the data represents geographical surface terrain. In another embodiment the data represents in-ground geological mineral distribution.

In one embodiment the invention provides a method for modelling large scale data representing a distribution of physical properties, including obtaining a measured dataset, applying an exactly sparse Gaussian process covariance function to the measured dataset to determine an optimized set of hyperparameter values associated with the Gaussian process covariance function, and generating selected model data by

Gaussian process regression using the optimized hyperparameter set and the measured dataset.

In one embodiment the exactly sparse Gaussian process covariance function includes a sparse covariance function that is smooth and diminishes to zero outside of a characteristic hyperparameter length, and a non-stationary neural network covariance function.

In another embodiment the sparse covariance function is multi-dimensional with a separate characteristic hyperparameter length for each dimension. In another embodiment the measured dataset represents geographical surface terrain measurements. In another embodiment the measured dataset represents in-ground geological mineral distribution measurements.

In one embodiment the invention provides a system for large scale data modelling, including at least one data measurement sensor for generating measured data representing a distribution of physical properties in a selected geographical region; a training processor adapted to determine for the measured data optimized hyperparameter values in relation to a Gaussian process covariance function including a sparse covariance function that is smooth and diminishes to zero outside of a characteristic hyperparameter length; and an evaluation processor adapted to determine model data from the Gaussian process covariance function with optimised hyperparameter values and measured data.

In one embodiment the Gaussian process covariance function also includes a non-stationary neural network covariance function. In another embodiment the measured data comprises geographical surface terrain measurements. In another embodiment the measured data comprises in-ground ore grade distribution measurements.

In another embodiment the invention provides a computational method for modelling a data space, the method including using a computational system, applying to the data space a kernel machine, wherein the kernel machine uses a covariance function which is constructed from a smooth basis function equal to zero outside of a defined interval; one of outputting and storing in memory a model from the kernel machine.

In another embodiment the kernel machine uses a Gaussian learning process. In another embodiment the basis function is a squared cosine wave. In another embodiment the basis function is $cos^2(\pi x)H(0.5 - |x|)$ where H(.) is a step function.

In another embodiment the multi-dimensional data space has N dimensions and the covariance function is

$$K_1(x,x';l,\sigma_0)=\begin{cases}\sigma_0\left[\frac{2+\cos\left(2\pi\frac{d}{l}\right)}{3}\left(1-\frac{d}{l}\right)+\frac{1}{2\pi}\sin\left(2\pi\frac{d}{l}\right)\right] & if\ d<l \\ 0 & if\ d\geq l\end{cases}, d=|x-x'|,$$

extended to multiple dimensions.

in another embodiment the defined interval is a hyperparameter.

In another embodiment the invention provides a method for modelling a data space, the method including using a computational system, applying to the data space a kernel machine, wherein the kernel of the kernel machine is a positive semi-definite function and a symmetric function, equal to zero outside of a defined interval, and at least one of one of outputting and storing in memory a model from the Gaussian regression process.

## Brief description of the drawings

The invention will be more fully understood through the following description of an embodiment thereof together with the illustrations in the accompanying drawings in which:

Figure 1 is an example computing system utilisable to implement a terrain modelling system in accordance withy an embodiment of the invention;

Figure 2 is a diagrammatic illustration of a terrain region and a system adapted for generating and maintaining a corresponding digital terrain model for use in controlling autonomous vehicles;

Figure 3 is a diagrammatic illustration of sensor sample data points representing elevation information across the terrain region of Figure 2;

Figure 4 is a flow-chart diagram showing a training phase for a terrain data modelling process;

Figure 5 is a diagrammatic representation of the evaluation phase for the terrain data modelling process;

Figure 6 is a plot of the covariance function for an exactly sparse Gaussian process for various values of hyperparameter $l$; and

Figures 7 and 8 show a comparison between the outputs of a Gaussian process model utilising a neural network covariance function $K_{NN}$ and the new covariance function $K_M$ that results in an exactly sparse Gaussian process.

Figure 9 shows a normalized squared error for different covariance functions and the standard deviation as a function of the number of inference points.

Figure 10 shows the percentage of zeros in the covariance matrix as a function of the number of inference points.

## Detailed description of the embodiments

It will be understood that the invention disclosed and defined in this specification extends to all alternative combinations of two or more of the individual features mentioned or evident from the text or drawings. All of these different combinations constitute various alternative aspects of the invention.

The present invention may be applied to digital terrain modelling as described hereinbelow in a non-limiting example of its implementation.

### 1. System overview

Referring to Figure 1, an embodiment of a terrain modelling system is implemented with the aid of appropriate computer hardware and software in the form of a computing system 100. The computing system 100 comprises suitable components necessary to receive, store and execute appropriate computer instructions. The components may include a processing unit 102, read only memory (ROM) 104, random access memory (RAM) 106, an input/output device such as disk drives 108, communication links 110 such an Ethernet port, a USB port, etc. and a display 113 such as a video display or any other suitable display. The computing system 100 includes instructions that may be included in ROM 104, RAM 106 or disk drives 108 and may be executed by the processing unit 102. There may be provided a plurality of communication links 110 which may variously connect to one or more computing devices such as a server, personal computers, terminals, wireless handheld computing devices or other devices capable of receiving and/or sending electronic information. At least one of a plurality of communications links may be connected to an external computing network through a telephone line, an Ethernet connection, or any type of communications link. Additional information may be entered into the computing system by way of other suitable input devices such as, but not limited to, a keyboard and/or mouse (not shown).

The computing system may include storage devices such as a disk drive 108 which may encompass solid state drives, hard disk drives, optical drives or magnetic tape drives. The computing system 100 may use a single disk drive or multiple disk drives. A suitable operating system 112 resides on the disk drive or in the ROM of the computing system 100 and cooperates with the hardware to provide an environment in which software applications can be executed.

In particular, the data storage system is arranged to store measurement data received from the sensors, in a suitable database structure 114. The data storage system also includes a terrain model 116, which is utilised with the measurement data to provide a 2.5D "map" of the terrain. The data storage system may be integral with the computing system, or it may be a physically separate system.

In more detail, the data storage system is loaded with a modelling module including various sub-modules (not shown). The sub-modules are arranged to interact with the hardware of the computing system 100, via the operating system 116, to either receive

the data collected by the measurement sensors (generally sent via the communications links 114) and/or process the received data to provide the measured data. In some embodiments, there may be provided a visual display unit, which may be in either local or remote communication with the computing system 100, and is arranged to display information relating to the programs being run thereon. In other embodiments, the data may be used directly by an autonomous robot (e.g. an automated vehicle) to perform particular tasks, such as navigation of the terrain.

Figure 2 is a diagrammatic illustration of a terrain region and a system adapted for generating and maintaining a corresponding digital terrain model for use in controlling autonomous vehicles. The terrain modelling system 210 operates on a terrain region 220 in the form of an open pit mine, and utilises one or more measurement sensors 230 to provide measured terrain data about the region 220. The sensor 230 provides spatial data measured from the terrain region 220 which can be generated by a number of different methods, including laser scanning, radar scanning, GPS or manual survey. One example of an appropriate measurement sensor is the LMS Z420 time-of-flight laser scanner available from Riegl. This form of sensor can be used to scan the environment region and generate a 3D point cloud comprising (x, y, z) data in three frames of reference. In the system 210 illustrated, two separate scanners are shown disposed to opposite sides of the terrain to be modelled so as to minimise occlusions and the like.

The sampled spatial data compiled from the sensors is illustrated diagrammatically in Figure 3, which shows a distribution 300 of individual sample data points 310 collected from across the terrain region representing the surface elevation at various locations. In order to build a terrain model the collection of sample data points are used as described below in a computational system involving Gaussian processes to enable the surface elevation to be inferred at any location in the terrain region.

The measurement sensor data generated by the sensors 230 is provided to a training processor 240 coupled to data storage 250. The training processor 240 is adapted to organise the sensor data and determine a non-parametric, probabilistic, multi-scale representation of the data for use in terrain modelling, which is stored in the data

storage 250. Details of the specific operational procedures carried out by the training processor are described below with reference to Figure 4.

An evaluation processor 260 is also coupled to access the data storage 250 and is adapted to utilise the data from the data storage 250 and terrain model data according to a desired modelling grid resolution. Specific operational details of the evaluation processor are provided below with reference to Figure 5. Once the terrain model data has been generated it can be communicated to an autonomous robotic vehicle 280, operating within the modelled environment, by a suitable communications medium 270, such as a radio communication link.

Figure 4 is a flow-chart diagram showing a training phase procedure 400 for the terrain data modelling process. The procedure 400 that begins with obtaining sensor measurement data at step 410 from an appropriate source, in this case a Riegl LMS Z420 3D laser scanner. The laser scanner mentioned generates in the order of two-million data points for a given scan. Previously known Gaussian processes would typically require such a dataset to be reduced in magnitude in order to reduce the computational complexity and data storage requirements. In other words, a prior art Gaussian process may discard most of the measured data to reduce the number of data points to be used for actually generating the terrain model. By way of example, known Gaussian processes may necessitate the sensor measurement dataset being reduced from millions of data points down to thousands of data points. However, a Gaussian process employing a modified covariance function is described below which allows all of the measured data to be incorporated into the model without arbitrary sampling reduction. This process is thus well-suited for modelling large scale data, and particularly advantageous where the number of data points exceeds approximately 5000.

For ease of storage and retrieval the data storage 250 can make use of an efficient, hierarchical data structure known as a KD-Tree. The use of such a data structure provides the training and evaluation processors with rapid access to the sampled measurement data on demand. After the data has been input to the training processor 420 it is converted to KD-Tree data 430 and stored 440.

The data storage step is followed by a Gaussian process learning procedure at step 450, with the objective of learning a representation of the spatial data. Details of the Gaussian process learning procedure are provided below. This method provides for a non-parametric, probabilistic, multi-scale representation of the data. No explicit assumption of the underlying functional model is made.

Hyperparameters are determined from the data to be modelled. The hyperparameters can be learnt from the measured data using a manual process, i.e. using a trial and error process. The hyperparameters can also be learnt using a machine learning process. Typical methods include using leave-one-out cross-validation (LOOCV), also called rotation estimation, and Bayesian learning such as Maximum Likelihood Estimation.

In this example, a Maximum Likelihood Estimation method is used.

There are also various methods that can be used to optimise the hyperparameters. These include Broyden-Fletcher-Goldfarb-Shanno (BFGS) gradient descent and simulated annealing. In the example described in this specification, a combination of these two methods is used.

The covariance function hyperparameters provide a coarse description of the terrain model, and can be used together with the sensor measurement data to generate detailed terrain model data at any desired resolution, including a statistically sound uncertainty estimate. The optimized covariance function hyperparameters are stored 470, together with the KD-Tree sample data structure, for use by the evaluation procedure.

The hyperparameters determined for this example application are: $\sigma_0$, a scaling factor, $\sigma_n$, a noise component, and $l$, the characteristic length, which describes the distance over which points are correlated in a certain neighbourhood.

It is also possible to obtain an estimate of the hyperparameters without using machine learning. It can be done through an iterative manual estimation process, for example.

Figure 5 is a diagrammatic representation of the evaluation phase procedure 500 for the terrain data modelling process. The sensor measurement spatial data stored in the KD-Tree and the Gaussian process model hyperparameters obtained in the learning phase are used here as the input data to the GP evaluation process 520. Since the Gaussian process representation obtained is a continuous domain one, applying the model for any desired resolution amounts to sampling the model at that resolution, as the data required for the grid resolution (in 510) can be obtained as outlined below.

A grid in the area of interest, at the desired resolution is formed. The objective is to use the learnt spatial model to estimate the elevation at individual points in this grid. Each point in the grid is interpolated with respect to the model learnt in the previous step and the nearest training data around that point. For this step, the KD-Tree naturally and efficiently provides access to the nearest known data. This together with the learnt model provides an interpolation estimate for the desired location in the grid. The estimate is also accompanied with an uncertainty measure that is simultaneously computed in a statistically sound manner. Details of the Gaussian process evaluation and computational procedures are provided below. It is a benefit of the covariance function used with the Gaussian Process as described below that the data neighbourhood can be determined from the data relatively quickly. It is also possible to know with an increased certainty that the impact of the data points outside the neighbourhood is zero. This neighbourhood of relevant points is shown in the KD-Tree structure.

The output 530 of the Gaussian process evaluation 520 is a digital elevation map/grid at the chosen resolution and region of interest together with an appropriate measure of uncertainty for every point in the map. The digital elevation map may be used as is or may be rapidly processed into a digital surface/terrain model and used thereafter for robotic vehicle navigation and the like in known fashion.

Although the foregoing description relates to spatial data modelling, in particular to terrain modelling of a mine, it will be readily appreciated that the data modelling methodologies described herein are not limited to this application. Any modelling problem to which a Gaussian Process is applied may use the modified covariance function described herein, with fields including, and not limited to, weather forecasting,

financial forecasting, bio-medical diagnosis, geometric modelling, stochastic maintenance modelling, astronomy, internet and search analysis, speech recognition, face recognition, audio/video data study, geophysics, mining, hydrology, reservoir engineering and robotics. As an example, the invention is also applicable to in-ground ore grade estimation, also in a mining context, where modelling of in-ground ore distributions are desired on the basis of drill hole analyses or other measurements. In the case of in-ground mineral distribution the data can have more dimensions than when considering surface elevation data.

In other applications, the input data to the process 510 can include multi-dimensional data points 410 of any dimensions and can represent a wide variety of parameters, including and not limited to time, distance, grade of material, temperature, humidity and density. These data points can be obtained in various ways, and need not be obtained by a sensor as such, as shown in 410.

Further details of the data modelling method and system are set forth below, with a particular focus on the details of the Gaussian process and underlying mathematical principles and procedures.

## 2. The Gaussian Process (GP)

Gaussian processes (GPs) provide a powerful learning framework for learning models of correlated and uncertain data. Gaussian process regression provides a robust means of estimation and interpolation of information that can handle incomplete sensor data effectively. GPs are non-parametric approaches in that they do not specify an explicit functional model between the input and output. They may be thought of as a Gaussian Probability Distribution in function space and are characterized by a mean function $m(x)$ and the covariance function $k(x, x')$ where

$$m(x) = E[f(x)] \tag{1}$$

$$k(x,x') = E[(f(x) - m(x))(f(x') - m(x'))]$$

(2)

such that the GP is written as

$$f(x) \sim GP(m(x), k(x,x'))$$

(3)

The mean and covariance functions together provide a distribution over functions. In the context of modelling surface terrain, for example, each $x \equiv (x, y)$ and $f(x) \equiv z$ of the given data.

## 2.1 The covariance function in the Gaussian Process

The covariance function models the covariance between the random variables which, here, correspond to sensor measured data. Although not necessary, for the sake of convenience the mean function $m(x)$ may be assumed to be zero by scaling the data appropriately such that it has a mean of zero. There are numerous covariance functions that can be used to model the spatial variation between the data points. It will be understood that a covariance function will be positive semi-definite and symmetric, that is $k(x, y) = k(y, x)$. The most popular covariance function is the squared-exponential covariance function given as

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq}$$

(4)

where $k_y$ is the covariance function; $l$ is the characteristic length, a measure of how quickly the $f(x)$ value changes in relation to the $x$ value; $\sigma_f^2$ is the signal variance and $\sigma_n^2$ is the noise variance in the data being modelled. The symbol $\delta_{pq}$ represents a Kroeneker Delta defined on indices $p$ and $q$. The set of parameters $l$, $\sigma_f$, $\sigma_n$ are referred to as the hyperparameters and specify what sort of values the parameters might take.

Another covariance function, one that has been found to be successful in the present application, is the *neural network* covariance function. It is specified by

$$k_y(x, x') = \sigma_f^2 \arcsin\left(\frac{\beta + 2x^T \Sigma x'}{\sqrt{(1 + \beta + 2x^T \Sigma x)(1 + \beta + 2x'^T \Sigma x')}}\right) + \sigma_n^2 \delta_{pq} \quad (5)$$

where $\Sigma$ , $\sigma_f$ , $\beta$ and $\sigma_n$ are the hyperparameters associated with the Gaussian process covariance function.

The difference between these two covariance functions is that the squared-exponential covariance function, being a function of $|x - x'|$, is stationary (invariant to translation) whereas the neural network covariance function is not so. In practice, the squared exponential covariance function has a smoothing or averaging effect on the data. The neural network covariance function proves to be much more effective than the squared exponential covariance function in handling discontinuous (rapidly changing) data.

Covariance functions in general and the covariance function described below can also be used in other kernel machines, for example support vector machines and kernel PCA (principal component analysis) machines. Kernel machines using the covariance function of the present invention may be used for machine learning or artificial intelligence techniques for data modelling applications including dimensionality reduction, classification problems, prediction/forecasting problems and regression analysis as described herein for spatial data.

Dimensionality reduction is used to reduce the number of data points that a data analysis uses and can be used for, for example, denoising, image compression or bio-informatics applications. Classification can be used, for example, for image recognition. Prediction methods can be used, for example, for forecasting in environmental sciences, e.g. for weather or pollution, for predicting failure times or rates of structures or devices, or for financial forecasting.

## 2.2 Training the Gaussian Process

Training the GP for a given dataset is tantamount to optimising the hyperparameters of the underlying covariance function. This training can be done using machine learning. It can also be done manually, for example by estimating the values and performing an iterative fitting process.

For a machine learning method, the hyperparameters are optimised with respect to the relevant covariance function. For the squared-exponential covariance function, this amounts to finding the optimal set of values for $\theta = \{ l_x , l_y , l_z , \sigma_f , \sigma_n \}$. For the neural network covariance function, the optimal values must be determined for $\theta = \{ l_x , l_y , l_z , \sigma_f , \beta , \sigma_n \}$. This is done by formulating the problem in a Maximum Likelihood Estimation framework and subsequently solving a non-convex optimization problem.

A Bayesian procedure is used to maximise the log marginal likelihood of the training output ($y$) given the training input ($X$) for a set of hyperparameters $\theta$ which is given by

$$\log(y|X,\theta) = -\frac{1}{2} y^T K_y^{-1} y - \frac{1}{2}\log|K_y| - \frac{n}{2}\log(2\pi) \qquad (6)$$

where $K_y = K_f + \sigma_n^2 I$ is the covariance matrix for the noisy targets $y$. The log marginal likelihood has three terms – the first describes the data fit, the second term penalises model complexity and the last term is simply a normalisation coefficient. Thus, training the model will involve searching for the set of hyperparameters that enables the best data fit while avoiding overly complex models. Occam's razor is thus in-built in the system and prevention of over-fitting is guaranteed by the very formulation of the learning mechanism.

Using this maximum likelihood approach, training the GP model on a given set of data amounts to finding an optimal set of hyperparameters that maximize the log marginal likelihood (Eq. 6). This can be done using standard off-the-shelf optimization approaches. For example, a combination of stochastic search (simulated annealing) and gradient descent (Quasi-Newton optimization with BFGS Hessian update) has been found to be successful. Using a gradient based optimization approach leads to

advantages in that convergence is achieved much faster. A description and further information about these optimization techniques and others can be found in the text *Numerical Optimization*, by J. Nocedal and S. Wright (Springer, 2006), which is hereby incorporated herein by reference.

Applying the GP model amounts to using the learned GP model to estimate the elevation information across a region of interest, characterised by a grid of points at a desired resolution. The 2.5D elevation map can then be used as is or as a surface map for various applications. This is achieved by performing Gaussian Process Regression at the set of query points, given the training dataset and the GP covariance function with the learnt hyperparameters.

For additive independent identically distributed Gaussian noise with variance $\sigma_n^2$ , the prior on the noisy observations becomes

$$\operatorname{cov}(y_p, y_q) = k(x_p, x_q) + \sigma_n^2 \delta_{pq} \tag{7}$$

where $\delta_{pq}$ is a Kroeneker Delta defined on $p$ , $q$ and is equal to $1$ *iff* $p = q$ and $0$ otherwise.

The joint distribution of any finite number of random variables of a GP is Gaussian. Thus, the joint distribution of the training outputs $f$ and test outputs $f^*$ given this prior can be specified by

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{8}$$

The function values $(f_*)$ corresponding to the test inputs $(X_*)$ given the training data $X$, training output $y$ and covariance function $K$, is given by

$$\bar{f}_* = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y$$

(9)

and their uncertainty is given by

$$\operatorname{cov}(f_*) = K(X_*, X_*) - K(X_*, X)\left[K\left(X, X + \sigma_n^2 I\right)\right]^{-1} K(X, X_*)$$

(10)

Denoting $K(X, X)$ by $K$ and $K(X, X_*)$ by $K_*$; for a single test point $x_*$, $k(x_*) = k_*$ is used to denote the vector of covariances between the test point and the set of all training points. The above equations can then be rewritten for a single query point as:

$$\bar{f}_* = k_*^T (K + \sigma_n^2 I)^{-1} y$$

(11)

and

$$V[f_*] = k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_*$$

(12)

Equations (11) and (12) provide the basis for the elevation estimation process. The GP estimates obtained are a best linear unbiased estimate for the respective query points. Uncertainty is handled by incorporating the sensor noise model in the training data. The representation produced is a multi-resolution one in that a terrain model can be generated at any desired resolution using the GP regression equations presented above. Thus, the terrain modelling approach proposed is a probabilistic, multi-resolution one that aptly handles spatially correlated information.

Although the Gaussian process modelling as described above is a powerful method, modelling large-scale stochastic processes still presents challenges. The difficulty

comes from the fact that inference in GPs is usually computationally expensive due to the need to invert a potentially large covariance matrix during inference time which has $O(N^3)$ cost. For problems with thousands of observations, exact inference in normal GPs is intractable and approximation algorithms are required.

Most of the approximation algorithms employ a subset of points to approximate the posterior distribution of a new point given the training data and hyperparameters. These approximations rely on heuristics to select the subset of points, or use pseudo targets obtained during the optimization of the log-marginal likelihood of the model.

## 3. The new covariance function

It is also possible to address the complexity problem differently. Instead of relying on sparse GP approximations, a new covariance function may be employed which provides intrinsically sparse covariance matrices. This allows exact inference in sparse GPs using conventional methods. The new sparse covariance function can be multiplied by any other covariance function and the result is a sparse covariance matrix, enabling a high degree of flexibility to accurately model problems while still preserving the sparseness properties. Gaussian processes constructed using this new sparse covariance function are referred to herein as Exact Sparse Gaussian Processes (ESGPs).

The ESGP utilizes a smooth covariance function whose output equals zero whenever the distance between input observations is larger than a hyperparameter. It will be understood that smooth means continuously differentiable at least once. As with other hyperparameters, this can be estimated by maximising the marginal likelihood to better model the properties of the data such as smoothness, characteristic length, and noise. The new covariance function resembles the popular squared exponential in terms of smoothness. The ESGP compares well with local approximation techniques and can be used together with other (non-stationary) covariance functions in the same modelling framework. The method results in very sparse covariance matrices (over 90% equal to

zero in an in-ground grade estimation example presented below) which requires significantly less memory while providing better performance.

The covariance function sought must vanish out of some finite interval to enable exact sparse GP regression. It must preferably produce smooth sample curves but the sample curves should not be infinitely differentiable so that they will be applicable to inference of processes that have jumps in some derivatives.

The function $f(x) = \cos^2(\pi x)H(0.5 - |x|)$ has been chosen as a basis function.

In view of $\cos^2(\pi x) = (\cos(2\pi x) + 1)/2$, the basis function is actually the cosine function shifted up, normalized and set to zero out of the interval $x \in (-0.5, 0.5)$. The cosine function was chosen as a basis function because it is analytically well tractable, integrals with finite limits containing its combinations, can be calculated in closed form and also the cosine function has proved to be able to provide good approximations for different functions and is a core element of the Fourier analysis. Here and afterwards $H(\cdot)$ represents the Heaviside unit step function. The chosen basis function $f(x)$ is smooth on the whole real axis, vanishes out of the interval $x \in (-0.5, 0.5)$ and has discontinuities in the second derivative.

Using the transfer function $h(x; u) = f(x - u)$ the following 1D covariance function can be obtained:

$$K_1(x,x') = \sigma \int_{-\infty}^{\infty} h\left(\frac{x}{l}; u\right) h\left(\frac{x'}{l}; u\right) du \qquad (13)$$

Due to the chosen form of the basis function the integral in Eq. (13) can be analytically evaluated to result in:

$$K_1(x,x';l,\sigma_0) = \begin{cases} \sigma_0 \left[ \frac{2+\cos\left(2\pi\frac{d}{l}\right)}{3}\left(1-\frac{d}{l}\right) + \frac{1}{2\pi}\sin\left(2\pi\frac{d}{l}\right) \right] & \text{if } d < l \\ 0 & \text{if } d \geq l \end{cases}$$

(14)

where $\sigma_0 > 0$ is a constant coefficient, $l > 0$ is a characteristic length and $d$ is the distance between the points:

$$d = |x - x'|$$

(15)

From Eq. (14) it follows that for any points $x_i$ and any numbers $a_i$ where $i = 1, 2, \ldots, n$ the inequality $\sum_{i,j=1}^{n} a_i a_j K_1(x_i, x_j) = \sigma \int_{-\infty}^{\infty} (\sum_{i=1}^{n} a_i h(\frac{x_i}{l}; u))^2 du \geq 0$ holds, so that the constructed covariance function is positive semi-definite.

Based on Eqs. (14)-(15) it can be calculated that

$$K|_{d=l} = \frac{\partial K}{\partial d}\bigg|_{d=l} = \frac{\partial^2 K}{\partial d^2}\bigg|_{d=l} = \frac{\partial^3 K}{\partial d^3}\bigg|_{d=l} = \frac{\partial^4 K}{\partial d^4}\bigg|_{d=l} = 0$$

(16)

$$\frac{\partial^5 K}{\partial d^5}\bigg|_{d=l} = -4\pi^4 \neq 0$$

(17)

which shows that the covariance function $K_l$ is continuous and has continuous 4th derivative. The function $K_l(d,0;1,1)$ is illustrated in Figure 6.

This covariance function can be extended to multiple dimensions in the following ways:

1. Using direct products for all axes:

$$K_M^{(1)}(x, x'; L, \sigma_o) = \sigma_o \prod_{i=1}^{D} K_1(x_i, x_i'; l_i, 1)$$

(18)

where D is the dimensionality of the points;

$$x=(x_1,x_2,\ldots,x_D)^T,\ x'=(x'_1,x'_2,\ldots,x'_D)^T$$

(19)

are the points and their coordinates; and $L$ is the vector of the characteristic lengths:

$$L = (l_1, l_2, \ldots, l_D)^T$$

(20)

2. Using Mahalanobis distance:

$$K_M^{(2)}(r;\sigma_o,\Omega) = \sigma_o\left[\frac{2+\cos(2\pi r)}{3}(1-r)+\frac{1}{2\pi}\sin(2\pi r)\right] \quad if\ r<1$$

(21)

$$K_M^{(2)}(r;\sigma_o,\Omega) = 0 \qquad if\ r\geq 1$$

(22)

where $\sigma_0 > 0$, $\Omega$ is positive definite and

$$r=\sqrt{(x-x')^T\Omega(x-x')},\ \ \Omega>0$$

(23)

$K_M$ is used in this specification to refer to the function $K_M^{(2)}(r;\sigma_0\Omega)$.

The developed new multi-dimensional covariance function in both its forms $K_M^{(1)}$ and $K_M^{(2)}$ has the following remarkable properties:

1. It vanishes out of some finite region $S$:

$$S_1 = \left\{ r \in \Re^D : K_M^{(1)}(r) \neq 0 \right\} \tag{24}$$

$$S_2 = \left\{ r \in \Re^D : K_M^{(2)}(r) \neq 0 \right\} \tag{25}$$

   Sizes of the regions $S_1$ and $S_2$ can be controlled via the characteristic lengths $l_i$. Moreover, these sizes can be learned from the data using GPs.

2. All the derivatives up to (and including) the fourth order derivative vanish at the perimeter of $S$, which guarantees mean square differentiability up to the corresponding order of the sample curves in GPs. There is a discontinuity at the perimeter for the fifth order gradient.

3. The region $S_1$ is a $D$ dimensional rectangle and the region $S_2$ is a $D$ dimensional ellipsoid.

4. In the case of one dimension $K_M^{(1)}$ and $K_M^{(2)}$ become identical.

5. This covariance function makes it possible to learn the anisotropic properties of the data (i.e. different inner properties in different directions).

6. This covariance function leads to very sparse covariance matrices and makes it possible to study the whole picture of huge datasets with low memory and computational cost requirements and without making local approximations.

Based on Eqs. (14)-(20) the following expressions for the partial derivatives of $K_M^{(1)}(x,x';L,\sigma_o)$ can be calculated:

$$\frac{\partial K_M^{(1)}}{\partial \sigma_0} = \frac{1}{\sigma_0} K_M^{(1)} \tag{26}$$

$$\frac{\partial K_M^{(1)}}{\partial l_i} = \frac{4\sigma_0}{3} \frac{K_M^{(1)}}{K_1} \frac{d_i}{l_i^2} \left[ \pi(1-\frac{d_i}{l_i})\cos(\pi \frac{d_i}{l_i}) + \sin(\pi \frac{d_i}{l_i}) \right] \sin(\pi \frac{d_i}{l_i}) \tag{27}$$

where $i = 1, 2, ..., D$.

If Ω is diagonal and positive definite it can be expressed via the characteristic lengths as follows:

$$\Omega = diag\left(\frac{1}{l_1^2}, \frac{1}{l_2^2}, ..., \frac{1}{l_D^2}\right) \tag{28}$$

From Eqs. (23)-(28) it follows that

$$r = \sqrt{\sum_{k=1}^{D} \left( \frac{x_k - x'_k}{l_k} \right)^2} \tag{29}$$

Based on Eqs. (21)-(23), (28)-(29) the following gradient components of this multi-dimensional covariance function $K_M$ can be obtained:

$$\frac{\partial K_M}{\partial \sigma_0} = \frac{2+\cos(2\pi r)}{3}(1-r) + \frac{1}{2\pi}\sin(2\pi r), \quad if\, 0 \le r < 1$$

(30)

$$\frac{\partial K_M}{\partial l_j} = \frac{4\sigma_0}{3}\left[\pi(1-r)\cos(\pi r) + \sin(\pi r)\right]\frac{\sin(\pi r)}{r}\frac{1}{l_j}\left(\frac{x_j - x'_j}{l_j}\right)^2, \quad if\, 0 < r < 1$$

(31)

$$grad\, K_M = 0, \; if \; r \ge 1$$

(32)

In Eq. (31) there is $r$ in the denominator, so that direct calculations cannot be carried out using Eq. (31) when $r = 0$. However, using the equality

$$\lim_{r\to 0}\frac{\sin(\mu r)}{r} = \mu$$

(33)

one can directly show that

$$\lim_{r\to 0}\frac{\partial K_M}{\partial l_j}(r;\sigma_0,\Omega) = \lim_{r\to 0}\frac{4\sigma_0}{3}\pi^2\frac{1}{l_j}\left(\frac{x_j - x'_j}{l_j}\right)^2 = 0$$

(34)

Based on Eq. (34) it must be taken directly

$$\left.\frac{\partial K_M}{\partial l_j}\right|_{r=0} = 0, j = 1,2,\ldots,D$$

(35)

Equations (30)-(32), (35) fully define the gradient of the new covariance function $K_M$ at every point and can be directly used in calculations.

## 4. Advantages over the neural network covariance function

As described above, the neural network (NN) covariance function is known for its ability to model discontinuities without undesired oscillations. However, numerical computations and comparisons shown in Figures 7 and 8 demonstrate that $K_M$ provides much better results and is independent of axes translations. In these figures the neural network covariance function $K_{NN}$ and the covariance function $K_M$ were each used to model a set of data points and the results are shown side-by-side. As can be seen in the plots of the actual data points and resulting fitted curves, in each instance the sparse covariance function results in a curve that more closely follows the actual data then the neural network covariance function. For example, in Figure 7 the curves fitted to the data as a result of the $K_{NN}$ covariance function can be seen to vary with axis translations, whilst $K_M$ does not. In Figure 7, $K_M$ provides reliable results in each case, whereas $K_{NN}$ is unable to fit a curve to the data points in the last instance. Furthermore, $K_M$ is more reliable, has better performance, requires much less memory usage and has much less computational cost than the neural network covariance function.

Application of this new covariance function to an in-ground ore grade estimation problem has shown that a covariance matrix which has about 2 million cells has over 90% of its cells exactly equal to 0. This provides possibilities for significantly decreasing the memory and computational costs of the computations. Using $K_M$ the number of zero cells is $O(n^2)$ while the number of nonzero cells is $O(n)$. Consequently, the percentage of zeros in the covariance matrix will significantly grow for bigger datasets. It can become over 95%, over 99% and even greater as the dataset becomes bigger and bigger.

To demonstrate the benefits of the proposed approach in reducing storage and number of operations for the same accuracy, the exact sparse GP can be compared with the exact GP with the squared exponential covariance function and the covariance function obtained by the multiplication of both of them. The dataset used for this example is a

popular dataset in statistics for comparing inference procedures and is known as the Spatial Interpolation Comparison dataset (SIC). The dataset consists of 467 points measuring daily rainfall measurements in 2D geographical space. These points are divided into two sets, inference and testing. The inference set contains the points used to perform inference on the testing points. For each case the experiment is repeated 1500 times with randomly selected inference and testing sets. Figure 9 shows the normalized squared error for the different covariance functions and the standard deviation (one sigma for each part of the bar) as a function of the number of inference points. As the number of inference points increases, so does the size of the covariance matrix. The results demonstrate that very similar errors are obtained for the difference covariances. However, the sparse covariance function produces sparse matrices thus requiring much less floating point operations. Figure 10 shows the percentage of zeros in the covariance matrix as a function of the number of inference points. As can be observed, the percentage of zeros grows quickly as more inference points are added and it reaches its limit around 100 inference points. Although the percentage of zeros reaches its limit, the number of zeros in the covariance matrix still increases because the size of the covariance matrix increases with the number of inference points. Also worth noticing is the performance of the multiplication between the two covariance functions. The error is essentially the same as for the sparse covariance function alone but the percentage of zeros for the sparse covariance function is significantly greater than for the product.

The foregoing description shows that large scale data modelling using Gaussian process techniques can be substantially improved in terms of both modelling accuracy and required computational effort by employing a sparse covariance function in isolation or in combination with another stationary or non-stationary covariance function (e.g. the non-stationary neural-network covariance function). The Exact Sparse Gaussian Process is successfully able to model large complex datasets whilst requiring significantly less memory and computational effort as compared to known GP applications.

Throughout this specification, unless the context requires otherwise, the word "comprise", or variations such as "comprises" and "comprising", will be understood to

imply the including of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The foregoing detailed description of the present invention has been presented by way of example only and is not intended to limit the scope of the invention which includes each and every novel feature and combination of novel features herein disclosed.

## CLAIMS

1. A method for modelling data using a Gaussian process including a sparse covariance function that diminishes to zero outside of a characteristic length, wherein the characteristic length is determined from the data to be modelled.

2. The method of claim 1 wherein the sparse covariance function is smooth.

3. The method of claim 1 or claim 2 wherein the sparse covariance function has a continuous fourth derivative.

4. The method of any one of the preceding claims wherein the characteristic length is determined through a machine learning process

5. The method of any one of the preceding claims, wherein the sparse covariance function is multi-dimensional with a separate characteristic length for each dimension.

6. The method of any one of the preceding claims, wherein the Gaussian process further includes a non-stationary covariance function.

7. The method as claimed in claim 1, wherein the non-stationary covariance function is a neural network covariance function.

8. The method as claimed in claim 6 or claim 7, wherein the sparse covariance function and non-stationary covariance function are combined in Gaussian process regression computations.

9. The method of any one of the preceding claims, wherein the data is measured using one or more measurement sensors.

10. The method as claimed in claim 9, wherein the data represents geographical surface terrain.

11. A method as claimed in claim 9, wherein the data represents in-ground geological mineral distribution.

12. A method for modelling large scale data representing a distribution of physical properties, including

obtaining a measured dataset,

applying an exactly sparse Gaussian process covariance function to the measured dataset to determine an optimized set of hyperparameter values associated with the Gaussian process covariance function, and

generating selected model data by Gaussian process regression using the optimized hyperparameter set and the measured dataset.

13. A method as claimed in claim 12, wherein the exactly sparse Gaussian process covariance function includes a sparse covariance function that is smooth and diminishes to zero outside of a characteristic hyperparameter length, and a non-stationary neural network covariance function.

14. A method as claimed in claim 12 or claim 13, wherein the sparse covariance function is multi-dimensional with a separate characteristic hyperparameter length for each dimension.

15. A method as claimed in claim 12, wherein the measured dataset represents geographical surface terrain measurements.

16. A method as claimed in claim 12, wherein the measured dataset represents in-ground geological mineral distribution measurements.

17. A system for large scale data modelling, including:

at least one data measurement sensor for generating measured data representing a distribution of physical properties in a selected geographical region;

a training processor adapted to determine for the measured data optimized hyperparameter values in relation to a Gaussian process covariance function including

a sparse covariance function that is smooth and diminishes to zero outside of a characteristic hyperparameter length; and

an evaluation processor adapted to determine model data from the Gaussian process covariance function with optimised hyperparameter values and measured data.

18. A system as claimed in claim 17, wherein the Gaussian process covariance function also includes a non-stationary neural network covariance function.

19. A system as claimed in claim 17, wherein the measured data comprises geographical surface terrain measurements.

20. A system as claimed in claim 16, wherein the measured data comprises in-ground ore grade distribution measurements.

21. A computational method for modelling a data space, the method including:

using a computational system, applying to the data space a kernel machine, wherein the kernel machine uses a covariance function which is constructed from a smooth basis function equal to zero outside of a defined interval;

one of outputting and storing in memory a model from the kernel machine.

22. The method of claim 21 wherein the kernel machine uses a Gaussian learning process.

23. The method of claim 21, wherein the basis function is a squared cosine wave.

24. The method of claim 21, wherein the basis function is $cos^2(\pi x)H(0.5 - |x|)$ where H(.) is a step function.

25. The method of claim 21, wherein the multi-dimensional data space has N dimensions and the covariance function is

$$K_1(x,x';l,\sigma_0)=\begin{cases}\sigma_0\left[\frac{2+\cos\left(2\pi\frac{d}{l}\right)}{3}\left(1-\frac{d}{l}\right)+\frac{1}{2\pi}\sin\left(2\pi\frac{d}{l}\right)\right] & if\ d<l \\ 0 & if\ d\geq l\end{cases}, \quad d=|x-x'|$$

extended to multiple dimensions.

26. The method of any one of claims 21 to 25, wherein the defined interval is a hyperparameter.

27. A computational method for modelling a data space, the method including:

using a computational system, applying to the data space a kernel machine, wherein the kernel of the kernel machine is a positive semi-definite function and a symmetric function, equal to zero outside of a defined interval;

at least one of one of outputting and storing in memory a model from the Gaussian regression process.

28. A method of data modelling substantially according to any one of the embodiments herein described.

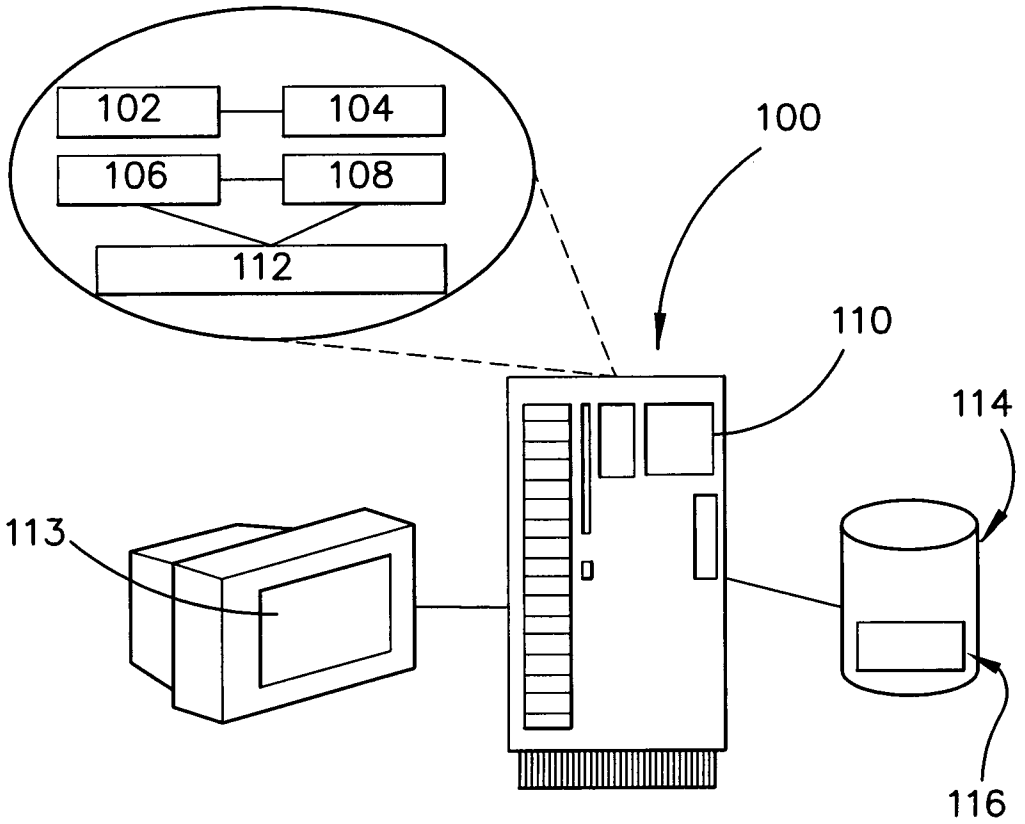



FIGURE 1

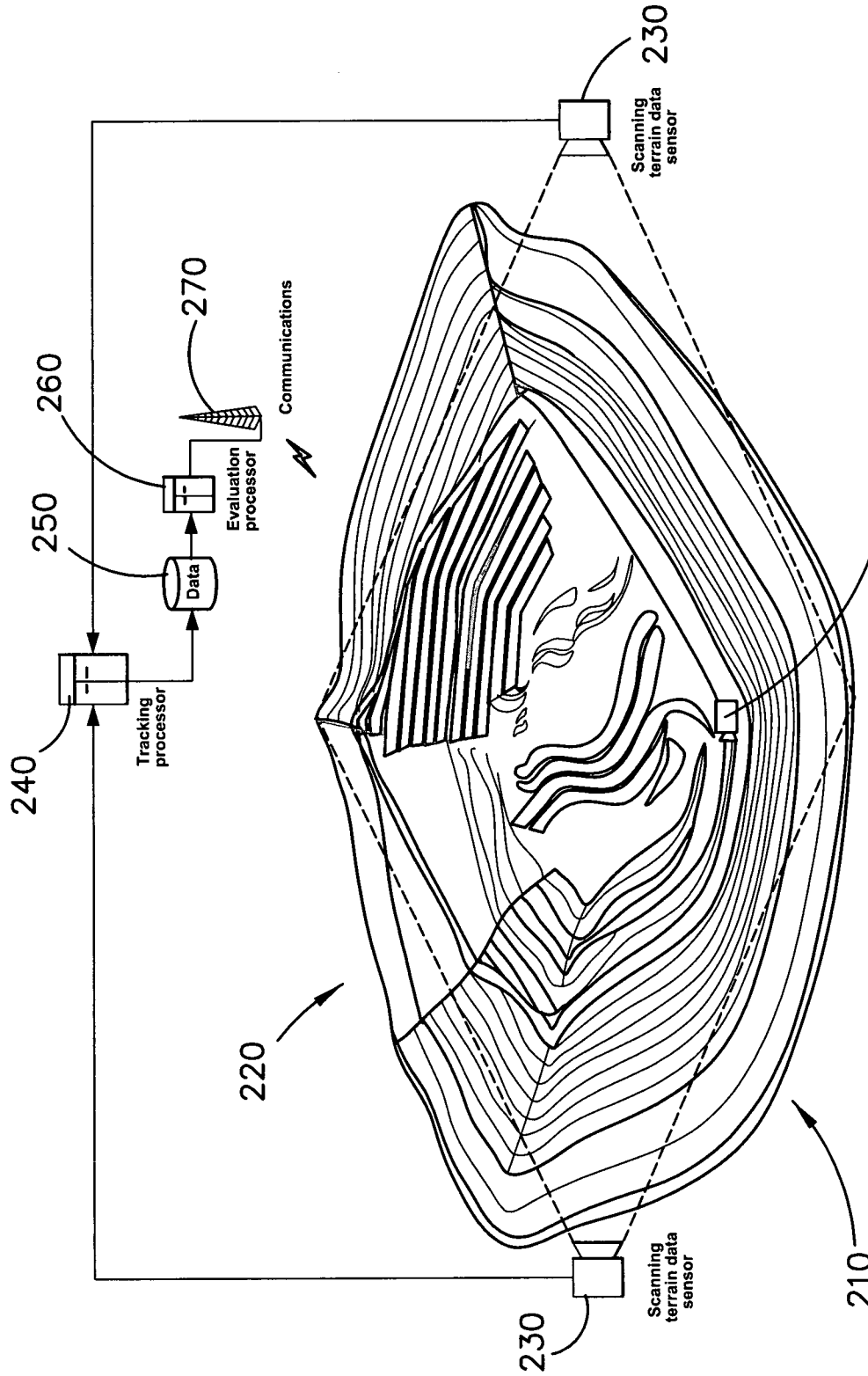

230
Scanning terrain data sensor
270
Communications
260
Evaluation processor
250
Data
240
Tracking processor
220
280
230
Scanning terrain data sensor
210

FIGURE 2

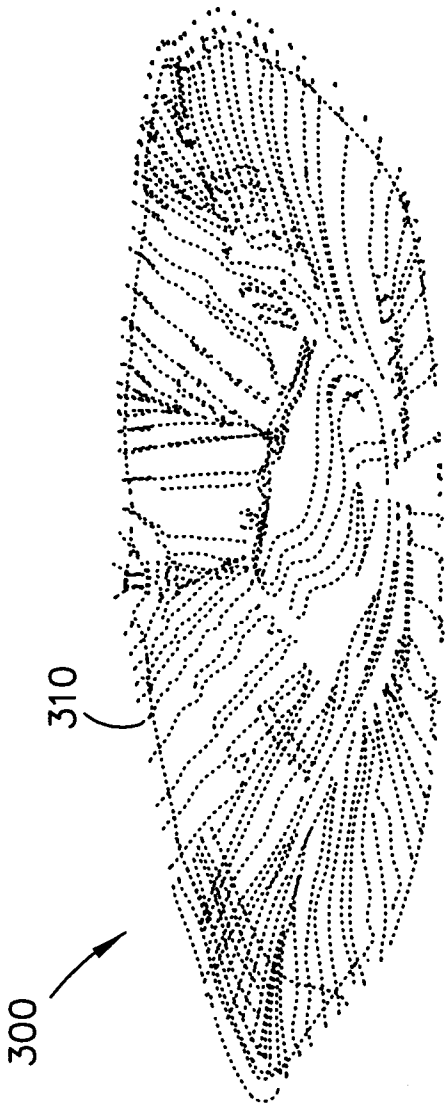



FIGURE 3

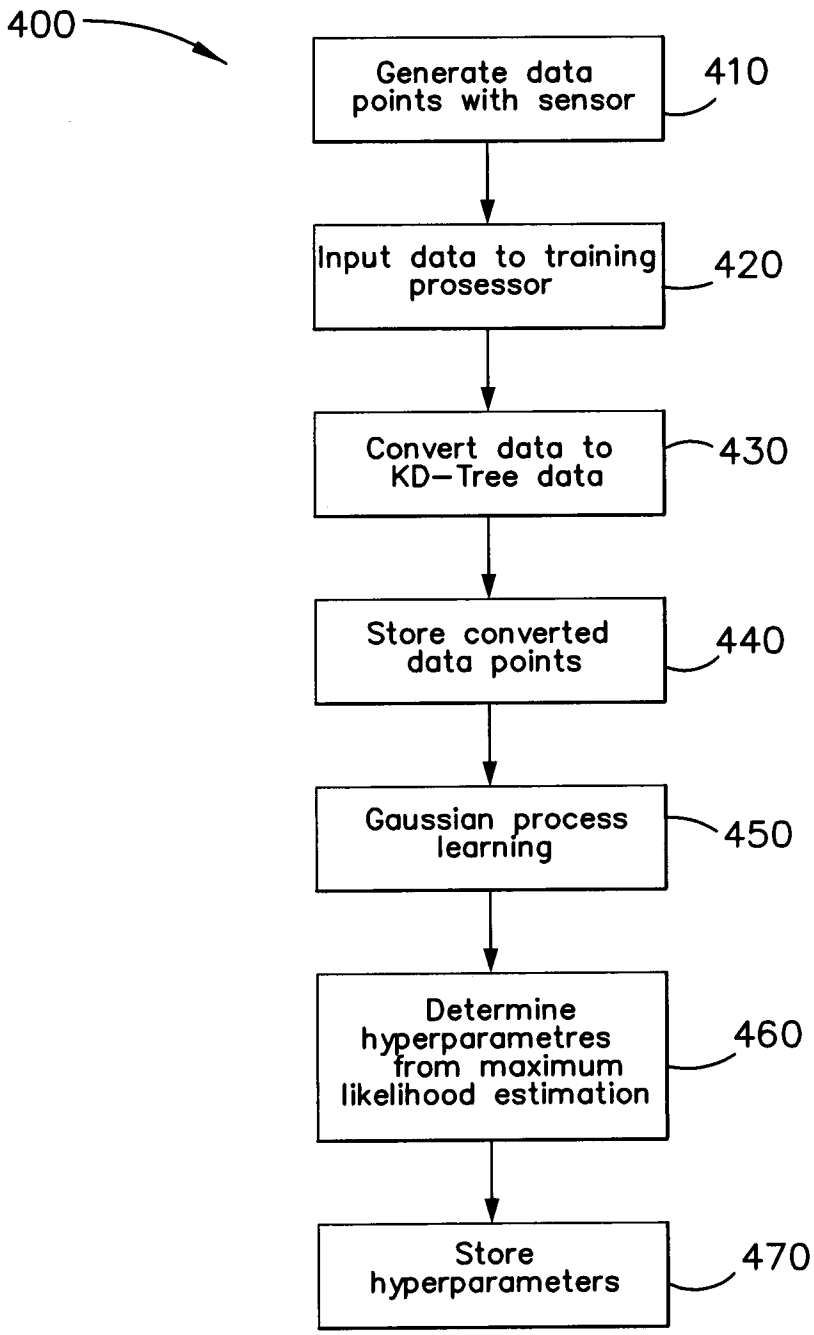

400
Generate data points with sensor
410
Input data to training prosessor
420
Convert data to KD–Tree data
430
Store converted data points
440
Gaussian process learning
450
Determine hyperparametres from maximum likelihood estimation
460
Store hyperparameters
470

FIGURE 4

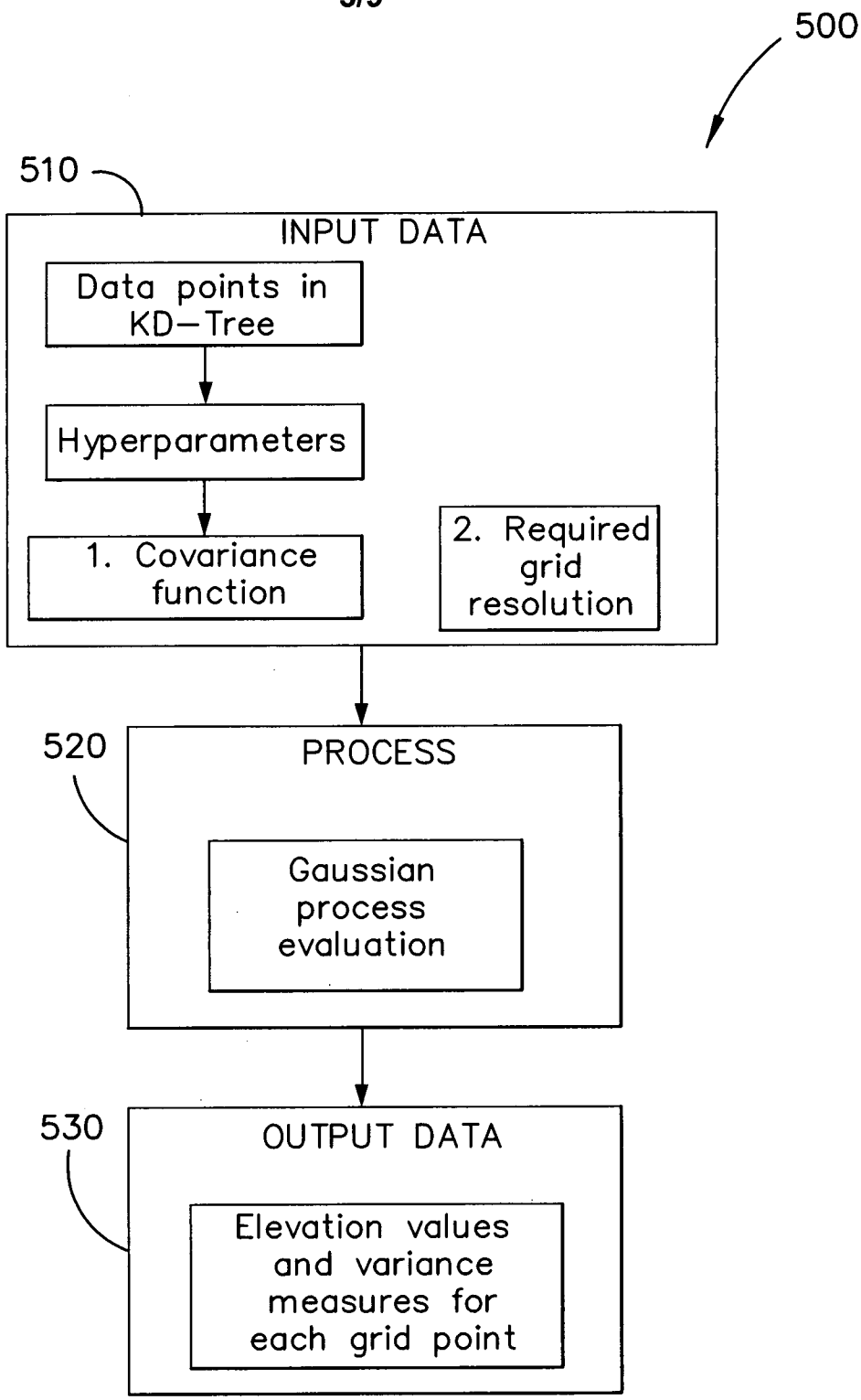

500
510
INPUT DATA
Data points in KD–Tree
Hyperparameters
1. Covariance function
2. Required grid resolution
520
PROCESS
Gaussian process evaluation
530
OUTPUT DATA
Elevation values and variance measures for each grid point

**FIGURE 5**

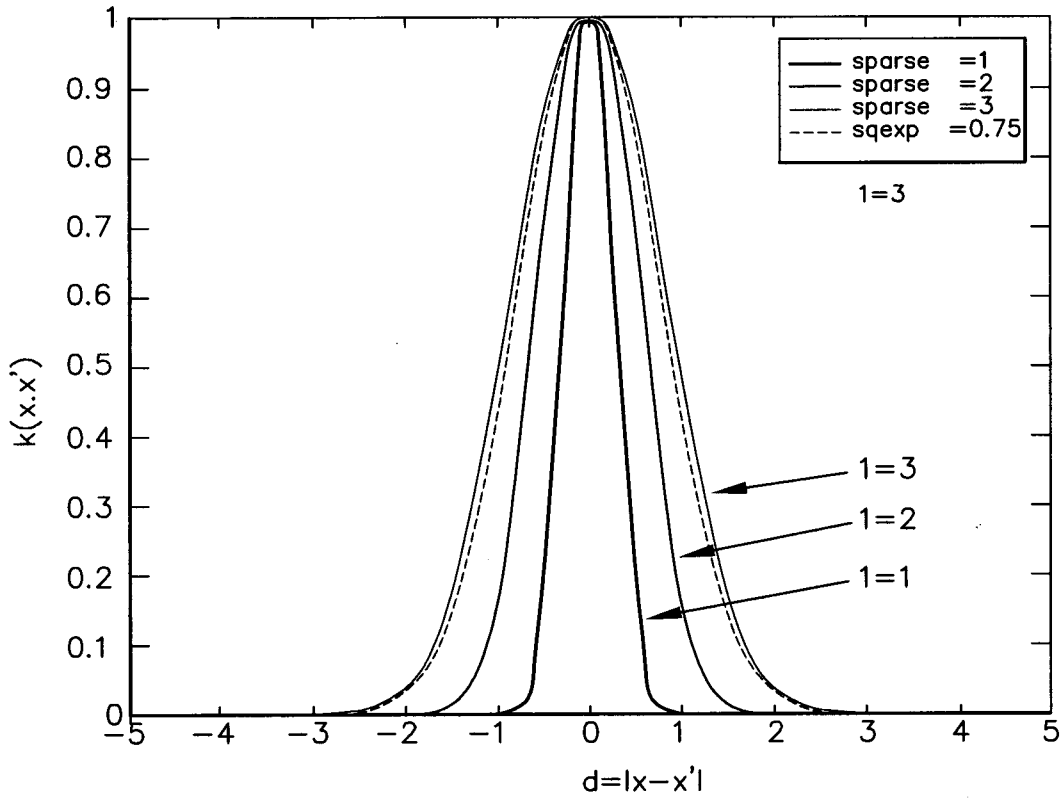
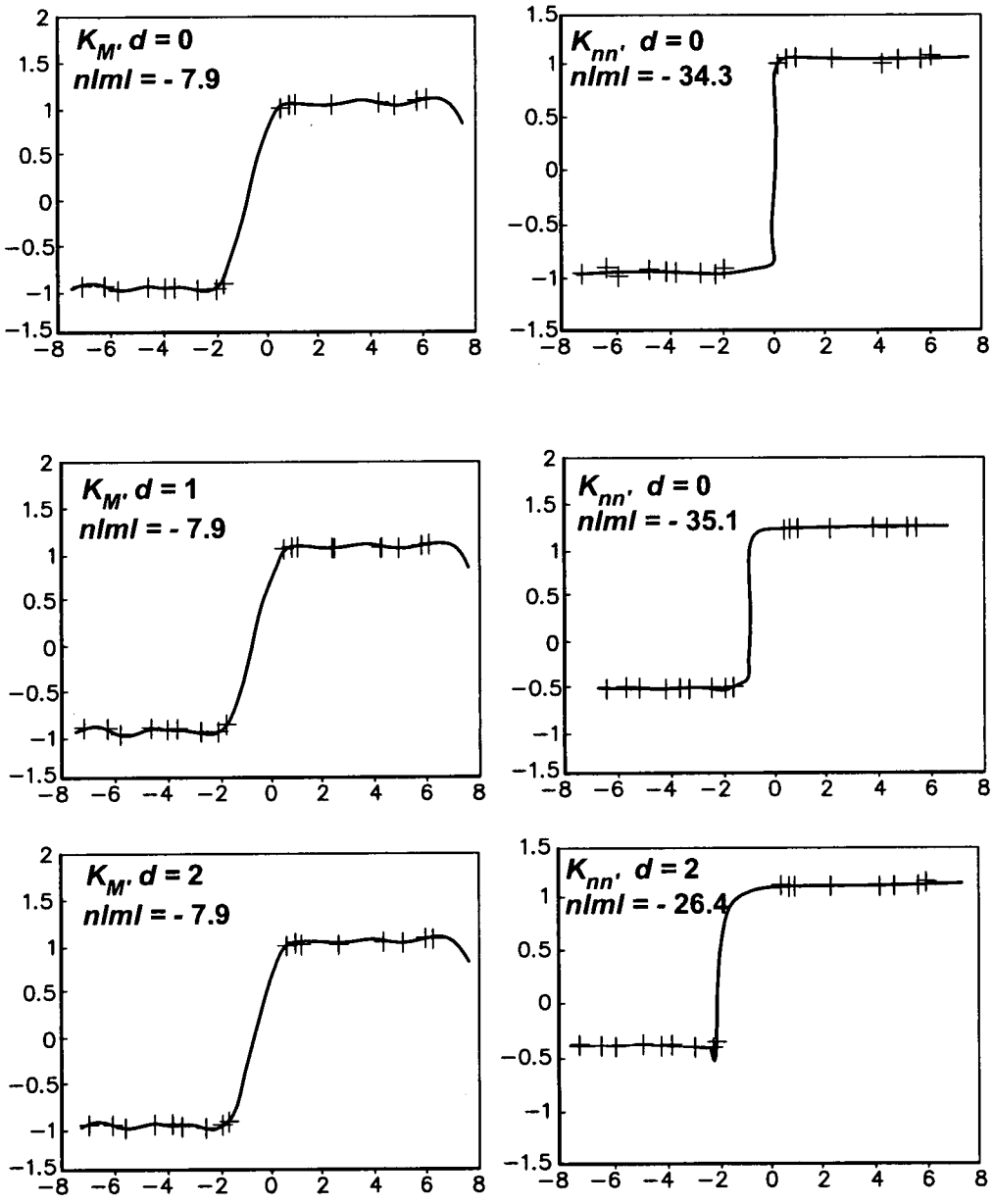
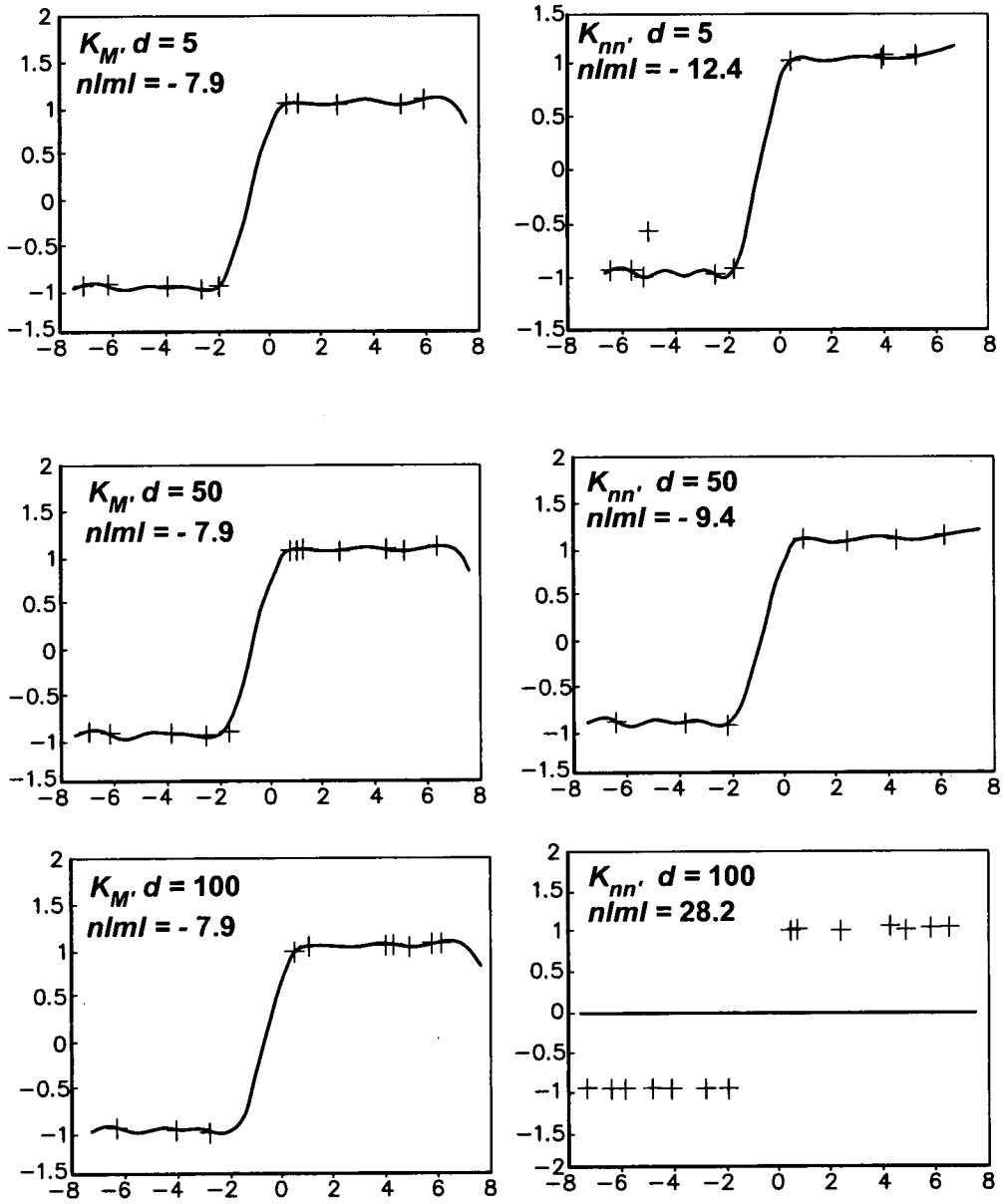

sparse =1
sparse =2
sparse =3
sqexp =0.75
1=3
1=3
1=2
1=1
k(x,x')
d=|x−x'|
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−5
−4
−3
−2
−1
0
1
2
3
4
5

FIGURE 6




**Comparison of GP regression with covariance functions $k_M$ and $k_{NN'}$.**

**FIGURE 7**




**Comparison of GP regression with covariance functions $k_M$ and $k_{NN'}$ (continuation)**

**FIGURE 8**

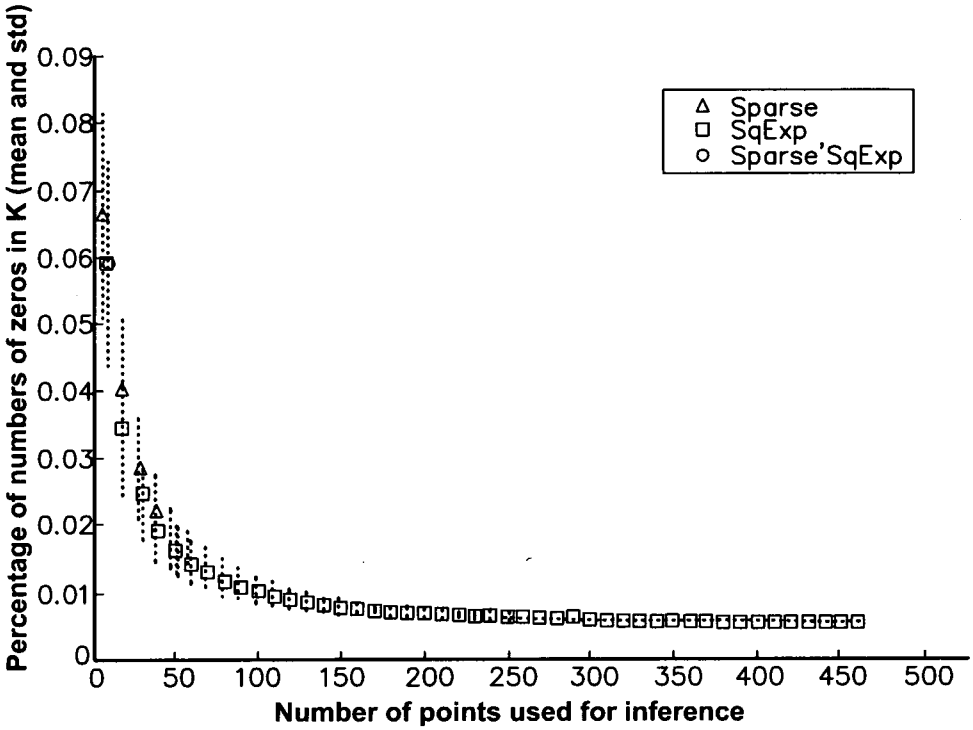



FIGURE 9

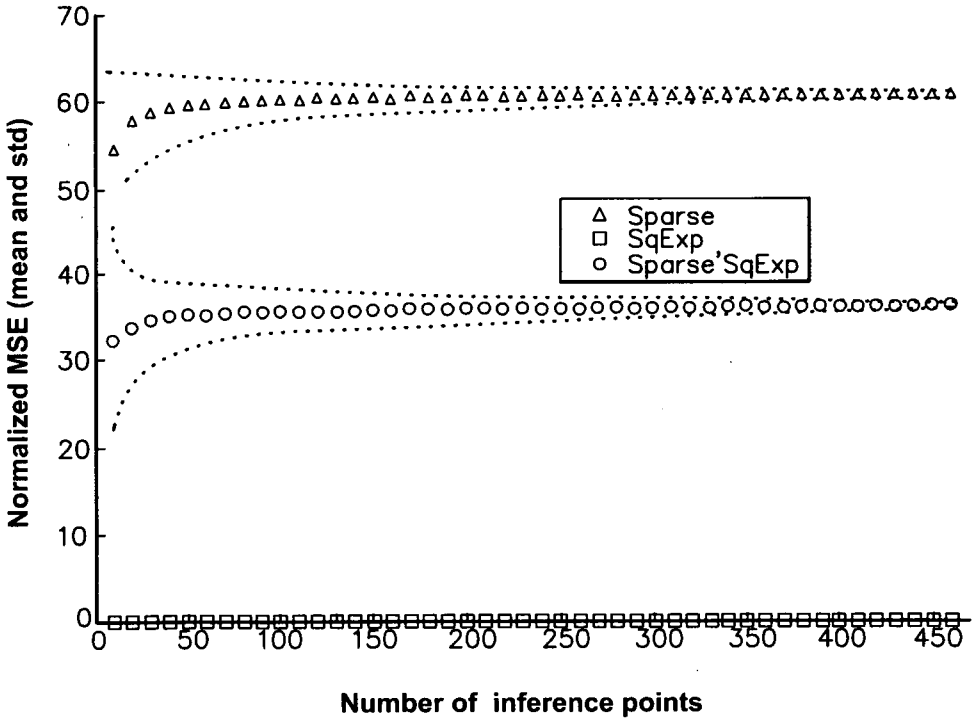



FIGURE 10

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2009/001668

A. CLASSIFICATION OF SUBJECT MATTER

Int. Cl.

***G06F 17/18*** (2006.01) ***G06F 17/00*** (2006.01) ***G06N 5/04*** (2006.01) ***G01V 1/50*** (2006.01) ***G06N 5/02*** (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Epodoc, WPI, Google Patents, Es@cenet, Patent Lens: keywords (Gaussian, covariance, sparse, model) and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | RASMUSSEN, C.E. and WILLIAMS, K.I. Gaussian Processes for Machine Learning (online), 2006 (retrieved on 25 February 2010). Retrieved from the internet: <URL: http://www.GaussianProcess.org/gpml><br>page 4 para. [0002], page 6 para. [0005], page 12 para. [0002], page 70 footnote, page 80 para. [0003], page 87 para. [0003]-page 88 para. [0001], Appendix C | 21-23, 26-27 |
| Y | page 4 para. [0002], page 6 para. [0005], page 20 para. [0002], page 87 para. [0003]-page 88 para. [0001], page 90 para. [0003], page 105-120 | 17-20 |
| Y | LANG et al. 'Adaptive Non-Stationary Kernel Regression for Terrain Modeling'. In: Proceedings of Robotics: Science and Systems (RSS), Atlanta, USA, June 2007<br>column 3 lines 32-35 | 17-20 |
| A | GLACKEN, I. M. and SNOWDEN, D. V.. Mineral Resource Estimation, in Mineral Resource and Ore Reserve Estimation – The AusIMM Guide to Good Practice (Ed: A C Edwards), 2001, pp 189-198<br>Whole document | 17-27 |

[X] Further documents are listed in the continuation of Box C [ ] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| 25 February 2010 | 8 MAR 2010 |
| Name and mailing address of the ISA/AU<br>AUSTRALIAN PATENT OFFICE<br>PO BOX 200, WODEN ACT 2606, AUSTRALIA<br>E-mail address: pct@ipaustralia.gov.au<br>Facsimile No. +61 2 6283 7999 | Authorized officer<br>**SURYA PRAKASH**<br>AUSTRALIAN PATENT OFFICE<br>(ISO 9001 Quality Certified Service)<br>Telephone No : +61 2 6283 2101 |

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2009/001668

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: **1-16**

because they relate to subject matter not required to be searched by this Authority, namely:

**Claims 1-16** define a mathematical algorithm to model data and in a number of dependent claims, the end result is a mere presentation of data. Rule 39.1 (i) and Rule 39.1 (v) of the regulations under the PCT define scientific and mathematical theories and mere presentation of information as being excluded subject matter respectively.

2. [X] Claim No.: **28**

because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

**Claim 28** does not comply with Rule 6.2(a) because it relies on references to the description and/or drawings.

3. [ ] Claims Nos.:

because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. [ ] As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

[ ] The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

[ ] The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

[ ] No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU2009/001668

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | SEEGER, M. 'Gaussian Processes for Machine Learning'. In: International Journal of Neural Systems, 2004, Volume 14<br>Whole Document | 17-27 |
| A | CSATÓ L. and OPPER M. 'Sparse representation for Gaussian process models'. In: Advances in Neural Information Processing Systems, MIT Press, 2001<br>Whole Document | 17-27 |
| A | Tipping, M.E. 'Sparse Kernel Principal Component Analysis'. In : Proceedings of Neural Information Processing System, 2000<br>Whole Document | 17-27 |