Probabilistic Detection of Vital Sign Abnormality with Gaussian Process Regression

David Wong, David A. Clifton, and Lionel Tarassenko
Institute of Biomedical Engineering
University of Oxford
Oxford, United Kingdom
{wong, davidc, lionel}@robots.ox.ac.uk

Abstract— **Vital-sign monitoring of patients within a hospital setting is a big component in the recognition and treatment of early signs of deterioration. Current vital-sign monitoring systems, including both manual early warning systems, and more sophisticated data fusion systems, typically make use of the most recently recorded data, and are unable to deal with missing data in a principled manner. The latter is particularly pertinent in the field of ambulatory monitoring, in which patient movement can result in sensor disconnections and other artefact. This paper presents a Gaussian process regression technique for estimating missing data and how it can be incorporated within an automated data fusion monitoring system. The technique is then demonstrated using vital-sign data from a recent clinical study conducted at the John Radcliffe Hospital, Oxford.**

Keywords - Data Fusion, Gaussian Process, Novelty Detection, Patient Monitoring

I. INTRODUCTION

Abnormalities in vital-signs such as heart rate, respiratory rate, blood pressure, and oxygen saturation have been shown to precede adverse events such as cardiac arrest [1]. For this reason, regular vital-sign observations by nurses are recommended for in-hospital patients, and are analysed using a manual early warning score (EWS) system [2]. This involves the clinician observing the vital-signs, applying univariate scoring criteria to each vital-sign, and then escalating the patient to a higher level of care if pre-determined criteria are met. The disadvantage of most EWS systems is that the escalation criteria are often determined heuristically, according to clinical opinion. Evidence-based systems, such as the centile-based EWS and VitalPac EWS attempt to address this issue [3, 4].

Another disadvantage of EWS systems is that they are typically used after routine observations of the patient vital-signs have been made, which may occur only twice per day in some ward settings. Continuous monitoring using automated devices may allow for a greater quality of care by prompting earlier clinical intervention between nurse observations. Traditionally, these automated devices have been used in conjunction with a bedside monitor. However, ambulatory monitoring, in which vital-sign data are transmitted wirelessly, is becoming increasingly popular due to perceived benefits in patient comfort and mobility [5], which may in turn affect the total recovery time.

To quantify the level of vital-sign abnormality, we have previously developed a data fusion technique which uses multiple channels of continuous vital-sign data, and which calculates a single score, the patient status index (PSI). The latter attempts to summarise the overall patient condition as a single value, such that a high PSI value indicates abnormality [6].

The systems listed above are sub-optimal in the event of missing vital-sign data; in each case, the algorithm assumes that the missing data channel is uninformative. In the case of continuous monitoring, this assumption is enforced by setting the missing variable to the population mean (or using a lower dimensional model) whereas [3] and [4] simply omit the missing variable from the calculation of the resultant score. While these methods may seem reasonable, they can lead to artefactual step-changes in score (shown in Fig. 1) that are unrelated to the underlying physiology.

Results from our clinical study (given in Section IV) have shown that, in the case of bedside monitors, data loss due to sensor disconnection is of the order of 20% (as shown in Table 1). Although we are unaware of any similar studies of ambulatory equipment, it is reasonable to assume that data loss would be even higher in such a setting due to the greater activity of the patient.

We present a method which estimates a distribution of values over the missing data channel by using previous values of the vital-sign. The distribution can then be used within a data fusion model to provide a probabilistic PSI score that is robust to periods of sensor artefact.

II. EXISTING WORK

Intelligent data fusion algorithms for continuous patient monitoring are necessary because of the high false-alert rate generated by simple single-channel threshold alerts. Tsien and Fackler [7] showed that approximately 86% of alerts from a bedside monitor in an intensive care unit (ICU) setting were false alerts.

Oberli et al. proposed an expert systems approach to the data fusion problem [8]. In their system, the vital-signs are first converted into a set of quantitative classes which describe a physiological condition, such as "bradycardia" or "normal heart rate", based on training information given by a set of clinicians. The classes overlap, and were described using fuzzy logic. A patient diagnosis, and resulting alerts, can then

be arrived at via a set of fuzzy logical rules derived from expert knowledge.

Schoenberg et al. [9] took an alternative approach, creating a customizable "logic engine" in which a set of vital-sign features was determined by expert knowledge. Thresholds were set for each feature, again on expert advice, and the sum of the scores was compared to a threshold, which triggers an alert if exceeded.

Zhang [10] proposed a personalised model that attempts to increase alert specificity by automatically tuning alert thresholds on a per-patient basis. Both neural networks and classification trees were tested, with the former found to perform consistently better.

The inclusion of temporal information, which is useful for predicting data drop-out, has been attempted in [11]. Their system uses semi-quantitative temporal vital-sign features, such as "*heart rate increasing*", within a rule-based system that triggers alerts when the trend is deemed to be sufficiently persistent and severe.

Temporal information can also be used to detect artefacts and thereby reduce the number of false alerts. Hoare and Beatty analysed the time series of a set of physiological features, and attempt to predict the next values using Kalman filtering and ARIMA models [12]. A new observation was then classed as artefactual if its value was outside a predetermined range.

Williams et al. extend this theme by using a factorial switching Kalman filter (FSKF) to model vital-sign data in neonatal intensive care [13]. The FSKF extends a standard Kalman filter by using different linear dynamic models that are selected by a switching variable, allowing "normal" and "artefactual" conditions to be modeled. The value of the switching variable changes depending on a number of factors such as the presence of bradycardia or recognition of probe disconnection. Given a set of observations, the FSKF is then used to calculate the most likely switching state.

Tarassenko et al. developed a patient monitoring system based on a novelty-detection principle, called "Visensia", in which the vital-signs from a normal patient population are first modeled, and alerts generated when a set of vital-signs differ significantly from that of the training population [14]. We now consider the Visensia model in more detail, which will be used to derive the PSI scores featured within the results presented here.

The model was trained using 3,500 hours of continuous vital-sign data collected from 150 high-risk patients at the John Radcliffe Hospital, Oxford, between 2001 and 2003 as part of an observational study. The method defines a PSI by first calculating the joint distribution $p(V)$ of vital-signs $V \in \mathbb{R}^4$. The vital-signs measured were: heart rate (HR), respiratory rate (RR), peripheral oxygen saturation (SpO_2), and systolic and diastolic blood pressures. The input to the model was a vector containing the instantaneous observation of each of the vital-signs, apart from the systolic and diastolic blood pressures, which were summarized by their arithmetic mean so that blood pressure did not have undue influence on

the model. Each vital-sign was normalized with respect to its own mean and variance, $V_i' = (V_i - \mu_i)/\sigma_i$.

The joint distribution was estimated using a Parzen windows kernel density estimate, using 400 kernels. The kernels were selected by summarizing the training data set using K-means clustering; this step was necessary due to the large size of the data set. The likelihood of a test data-point is then evaluated with respect to the p.d.f. $p(x|\theta)$, and the corresponding PSI is defined to be

$$-\log [p(x|\theta)] \qquad (1)$$

An alert is generated by the system if the PSI exceeds a given threshold, ζ, for n-minutes out of any m-minute window of data. Following previous work, we use $n = 4$ minutes and $m = 5$ minutes.
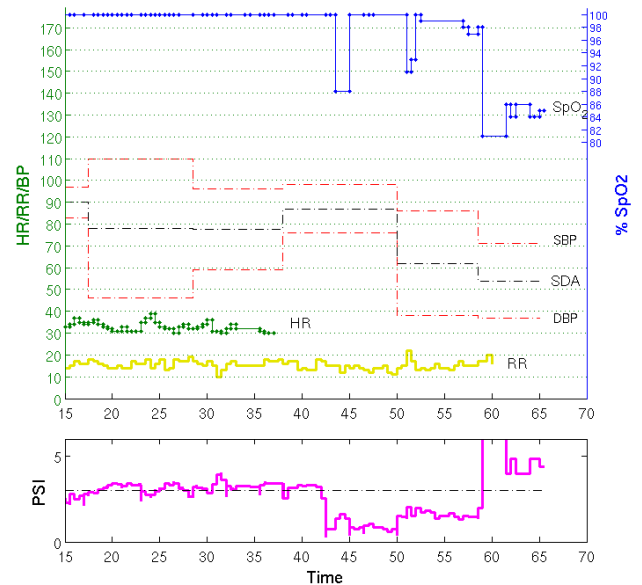

Fig. 1. Vital-sign data for an Emergency Department patient. The heart rate channel is lost $t = 37$ minutes into the record. The last seen value of HR is then internally held for five minutes (from $t = 37$ to $t = 42$ minutes), at which point, the missing variable is replaced by the training population mean. This leads to an artefactual drop in PSI score at $t = 42$ minutes.

III. METHOD

We now propose an adaption of previous work, into the probabilistic, non-parametric framework provided by Gaussian process regression.

The process of generating a probabilistic PSI during a period of missing data involves two stages. Firstly, the posterior distribution of the missing vital-sign data must be estimated at a given point in time. Samples from the posterior distribution can then be used as inputs to the deterministic data fusion model. This will provide a range of PSI scores, from which the empirical distribution of the PSI during the period of missing data may be derived.

A. Gaussian Process Fundamentals

A Gaussian process is defined as a stochastic process for which any finite combination of variables has a joint multivariate Gaussian distribution. To estimate the missing vital-sign data, we assume that the time-series may be adequately modelled using a Gaussian process.

A simple visual example of a time series modelled using a Gaussian process is shown in Fig. 2, which depicts a two-dimensional Gaussian distribution with a mean $[0\ 0]^T$ and a covariance of $\begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}$.

Any point on the distribution, (y_1, y_2), represents the joint probability $p(y_1, y_2)$ of two samples. For instance, the red dot in (a) represents the probability of the two points $y_1=1.5$, $y_2 = -0.2$.

Fig. 2 also demonstrates how the points y_1 and y_2 can be depicted as a two-point time series between x-axis values of x_1 and x_2. The figure also shows example time series for two other points selected from the Gaussian in (a). Let us suppose that these are observed at arbitrary times x_1 and x_2. As the time series grows with more points $y_3, y_4, \ldots, y_n$ it can continue to be modeled as a Gaussian process, in which the dimensionality of the joint Gaussian is increased to *3, 4, …, n,* respectively.
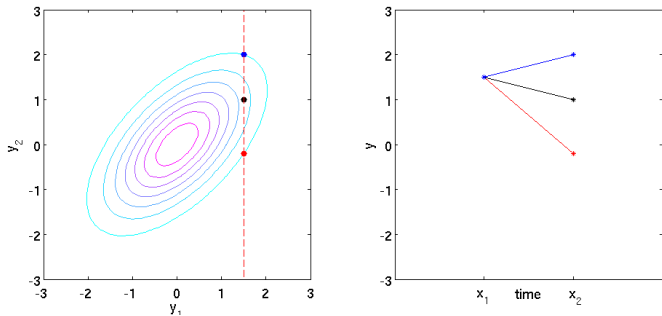

Fig. 2. A simple example of a Gaussian process for two points. The left figure shows the joint probabilities of all possible points as a bivariate Gaussian distribution. The right figure shows the equivalent time series plots for the three points on the Gaussian distribution highlighted in blue, black and red.

B. Covariance Functions

We now consider how the values on the time axis may be determined.

A covariance function is used to define the covariance matrix of the joint Gaussian distribution over the values y_i. A covariance function is a function of the times (x_i, x_j) corresponding to any two values (y_i, y_j). The form of the function is free, but must lead to the production of a valid covariance matrix (i.e. a matrix that is positive semi-definite). In our application, we expect local temporal behavior to be highly correlated, and the correlation to decrease as data samples become increasingly separated in time. Furthermore, in many cases we may also expect the function to be stationary, so that only the difference between x_i and x_j is used, and not their absolute values (i.e., absolute times).

The squared exponential covariance function is often used, which takes the form:

$$Cov(y_i, y_j) = K(x_i, x_j) = \sigma_0^2 \left(-\frac{1}{2} \frac{|x_i - x_j|^2}{\lambda} \right) \quad (1)$$

The covariance function contains two hyperparameters. The amplitude hyperparameter, σ_0, defines the maximum allowable variance, and is large for variables with a high dynamic range. The length-scale parameter, λ, controls how long in time an observation will be correlated to future observations.

While the hyperparameters may be set using prior knowledge, their "optimal" value may be learned from some training data by maximizing the log likelihood over the hyperparameters. The likelihood, which may be derived from Bayes' theorem, is $P(y|x, \lambda, \sigma)$, which is a normal distribution with covariance matrix, K. The log likelihood, for N time series observations, is therefore:

$$L = \frac{1}{2} \log|K| - \frac{1}{2} y^T K^{-1} y - \frac{N}{2} \log 2\pi \quad (2)$$

In the case where further information about the structure of the data is known, other covariance functions may lead to greater accuracy (e.g., see [8]). The results in the remainder of this article use the squared exponential covariance function to derive covariance matrices.

C. Gaussian Process Regression

The example in Fig. 2 is now used to demonstrate how regression onto previously unseen data can be computed. Consider the case when the value of $y_1=1.5$; the selection of the value of y_1 constrains y_2 to a one-dimensional slice, denoted by the dotted red line in Fig. 2. This line represents the conditional distribution $p(y_2 / y_1=1.5)$. From the contours of $p(y_1, y_2)$, we can see that y_2 is likely to have a value between *-0.3* and *2*. Therefore, by conditioning on y_1, the uncertainty in the value of y_2 can be reduced.

In general, for an n-dimensional Gaussian, and for a vector of y known points, and y^* unknown points, the relevant conditional distribution is given by:

$$P(y^*|y) \sim N(C^T B^{-1} y, A - C^T B^{-1} C) \quad (3)$$

where the covariance matrix, K, for $[y, y^*]$ has been divided into components that describe the correlations between the known points y, the correlations between the unknown points y^*, and the cross-terms. These are labeled A, B and C, respectively, so that K is:

$$K = \begin{bmatrix} A & C^T \\ C & B \end{bmatrix} \quad (4)$$

D. Data Fusion Framework

The posterior distribution for the missing vital-sign data is generated using Gaussian process regression. By extending the techniques outlined in Section II, an N-dimensional Gaussian process (for N vital-signs) will allow both inter- and intra-channel dependencies to be modelled. If more than one

channel of data is missing, the posterior distribution output by Gaussian process regression will be a *z*-dimensional Gaussian for *z* missing data channels. Fig 3(a) shows an instance in which one vital-sign is missing, which will lead to a univariate Gaussian posterior over the missing variable.

The output from the Gaussian process regression can now be interpreted within the context of the data fusion model. For this example, we assume that the data fusion model is the Visensia model described in Section II. However, the process should be easily extendible to other early warning score systems.

Fig. 3(b) shows a 2-dimensional representation of the pdf $p(V)$ model described previously. If the vital-signs V_1 and V_2 are known exactly, then a single point on the model can be defined, which can be converted into a PSI using eq. (1). If only one vital-sign is known exactly (for example, $V_1 = 20$), and V_2 has been estimated by Gaussian process regression, as in 3(a), then the resulting output from the Parzen windows model is constrained to:

$$P(V) = P(V_1, V_2 | v_1, \{\mu_{v_2}, \sigma_{v_2}\}) \quad (5)$$

Where $P(V_1, V_2)$ is merely the pdf from Section II, v_1 is the value of V_1, and μ_{v_2} and σ_{v_2} are the mean and variance, respectively, of the Gaussian process estimate of V_2. For the general case, the Parzen windows posterior cannot be calculated analytically, and a discrete estimate of the distribution can be generated by sampling from $V_2 = N(\mu_{v_2}, \sigma_{v_2})$ as the input to the Parzen windows model. As before, the PSI can again be calculated using eq. (1), but will now result in a distribution rather than a single value.

It now remains to show how alerts can be generated using the probabilistic PSI distribution. In the original data fusion model, we decide to alert if *n* out of the previous *m* minutes of data were above an alerting threshold (as shown in Fig 3(c)). Under this scheme, the alerting condition for a PSI score is binary: it is either above or below the threshold.

The framework introduced here outputs a *distribution* over PSI scores, rather than a single value (as shown in Fig 3(d)). An alert may now be generated if $\frac{n}{m}$% of the probability mass of a PSI distribution is above the alerting threshold. An alert should thus be generated if:

$$\sum_{T=0}^{T \leq I} P(PSI > \zeta)_T \geq \frac{n}{m} \quad (6)$$

Where ζ is the alerting threshold, and there are a total of I sets of vital-sign inputs within an *m*-minute period. In the case that the PSI is a single value (i.e. all variables are known), then $P(PSI > \zeta)$ is simply a Dirac delta function centred at the PSI score. More generally, the univariate cumulative probability distribution over PSI scores, $P(PSI)$, is given by:

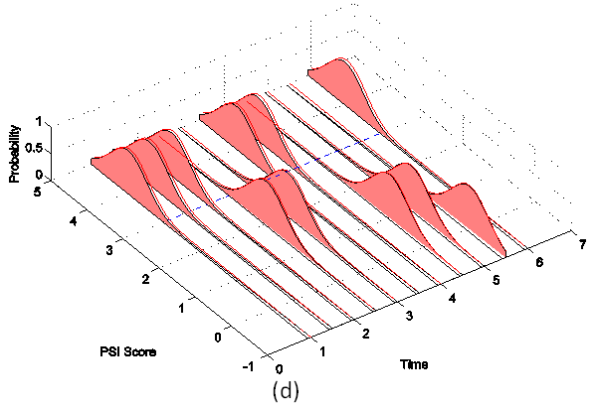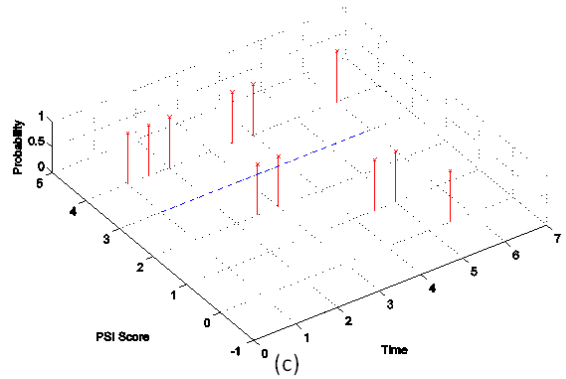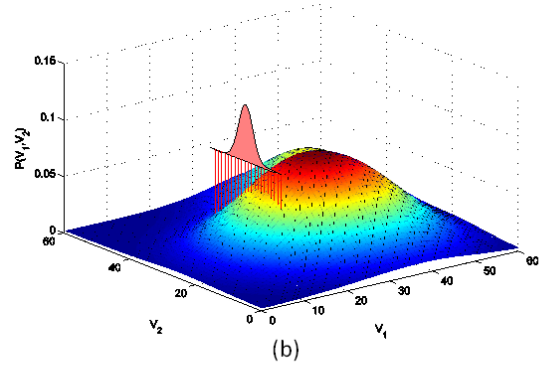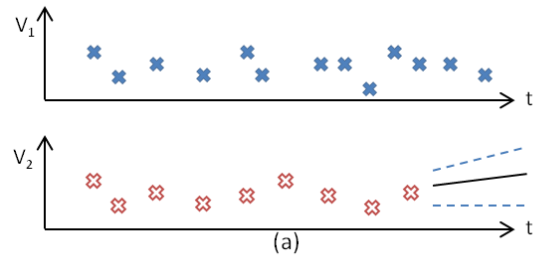$$P(PSI > \zeta) = \int_{e^{-\zeta}}^{\infty} P(V) dV_N \quad (7)$$

Fig. 3. A 2-dimensional pictorial description of the main stages of the Gaussian process framework for producing probabilistic PSI scores. (a) Firstly, V_2 is estimated (shown by blue dashed lines) using a multivariate Gaussian process, from information in the V_1 and V_2 time series. (b) shows how a PSI distribution is then generated. The surface plot represents the training population joint probability over V_1, V_2, on which the data fusion model is based. The red Gaussian slice indicates the allowable values of $P(V_1, V_2)$, given that V_1 is fixed, and the estimate for V_2 is Gaussian distributed according to the result from (a). (c) shows the PSI against time, demonstrating that, under the original system, a the score is fully above or below the alerting threshold of PSI = 3, even when there is uncertainty in the vital-sign values. (d) shows how this is addressed in the Gaussian process framework, which allows for PSI probability distributions

I. THE CLINICAL STUDY

The data fusion framework was tested using data collected from a clinical study conducted at the John Radcliffe hospital, Oxford. During the study, both intermittent observations and continuously monitored vital-sign data were acquired from Emergency Department patients. The study was conducted with the approval of the UK National Research Ethics Service, reference number 08/H1307/56.

Continuous data was acquired using a Phillips Intellivue bedside monitor, and measurements of RR, HR, and SpO$_2$ were recorded at a sampling rate of 30 seconds. These measurements required the physical connection of ECG electrodes, and a finger pulse-oximeter.

Intermittent measurements of blood pressure (BP) were recorded at intervals related to the condition of the patient. The most acute patients had BP recordings taken every 5 minutes, whereas recording could be as infrequent as once per hour for patients with less serious conditions.

In total, 476 patients were recruited to the study. The mean age of the patients was 61 years, and 52% of the patient population was male. Further patient demographics and information regarding data collection are reported in [16].

	HR	RR	SpO$_2$	SBP
Time (hours)	1,645	1,629	1,664	1,776
Data loss (%)	24.2	24.9	23.3	18.2

Table 1. Data quantity and data loss during clinical study data acquisition

II. RESULTS

We present two examples of Gaussian process regression for heart rate data, first showing an example of regression, and then showing a regression estimate is interpreted by probabilistic PSI.
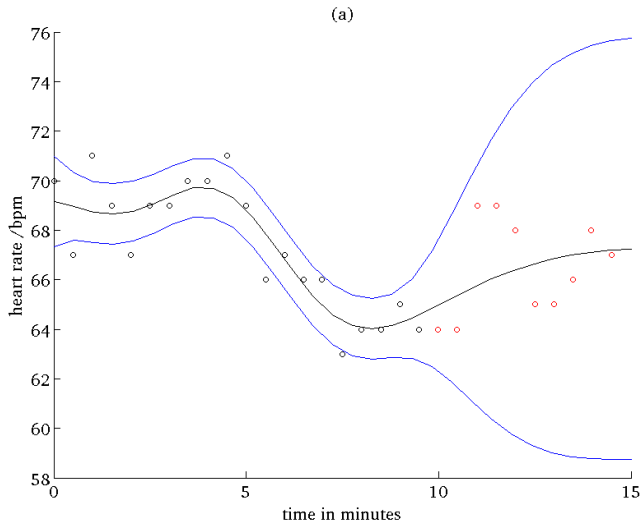
Fig. 4. An example of Gaussian process regression for heart rate data. The data was selected at random from the patients consented to a clinical study at the John Radcliffe hospital, Oxford. The black circles were used to train the Gaussian process model. The mean of the resulting regression estimate is shown in black, and ±2 standard deviations from the mean are shown in blue. The red circles indicate unseen test data.
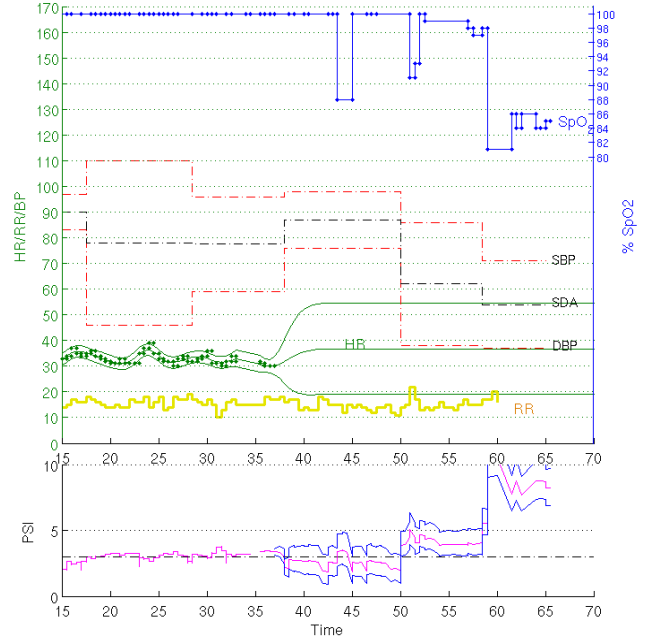
Fig. 5. Vital-sign data for an Emergency Department patient. The heart rate data channel is lost at t = 37 minutes into the record. This time, the value of the missing heart rate data is inferred using Gaussian process regression, leading to a consistent PSI score. As a consequence, the PSI score remains relatively high, and would indicate an alert (for an alert threshold of 3) at t = 50 minutes, due to the drop in blood pressure.

In Fig. 4, 15 minutes of data were selected at random from an arbitrary patient consented to the clinical study. These data were then partitioned into a set of 10 minutes of training data, which are depicted as black circles, and a set of 5 minutes of test data, which are shown in red. Gaussian process regression was implemented using the training data, and the mean of the resulting estimate is shown as a black line, with ±2 standard deviations from the mean in blue.

The method correctly predicts the slight overall upward trend in the heart rate values. In addition, the estimate becomes less certain as time progresses and the points become further away from the last observed data

The second example, shown in Fig. 5, uses the same patient data as for Fig. 1, and demonstrates how Gaussian process regression may be used within the data fusion framework described in Section IV. In comparison to Fig. 1, there are now no artificial step changes in the PSI at $t = 37$ minutes. Instead, the PSI remains at approximately 3, with a rapidly growing uncertainty.

In addition to producing a consistent estimate, the Gaussian process-assisted model is able to generate appropriate alerts when data are missing. At $t = 50$ minutes into the record, the blood pressure drops to an unusually low figure of 87/39 mmHg. This leads to an increase in the PSI, such that the 5th and 95th percentiles of the PSI distribution are greater than 3. In the case where the alerting threshold is set at PSI = 3 (its normal value from [18]), this would lead to an alert. In contrast, Fig. 1 shows a corresponding PSI of approximately

1.5, and the patient would be incorrectly deemed to be "normal".

III. DISCUSSION

In this paper, Gaussian process regression has been introduced as a method for inferring vital-sign data in the case of data drop-out. The Gaussian process model provides a posterior distribution for each unknown data channel, the variance of which may be used to quantify the certainty in the vital-sign estimate.

Following this, a description of how Gaussian process regression may be incorporated into an existing data fusion model was provided. The probabilistic framework has the advantage of providing a PSI distribution which is consistent with all previous data. This is most clearly seen by comparing Fig. 5 to Fig. 1, in which the original data fusion model reverts to a normal state once the pertinent channel of data is lost. This in turn means that the probabilistic PSI has a greater chance of detecting vital-sign deterioration, and providing relevant alerts.

The methods presented in this chapter are limited by the data quality. For instance, it is unclear whether the sampling rate, 30 seconds, of the continuous data presented here is optimal for estimating short term trends. A principled lower-bound on the sampling rate may be derived using raw waveform data. Indeed, Gaussian process regression may be used for "smart sampling" in which a measurement of a vital-sign is made when the uncertainty in the model has reached some threshold value.

Furthermore, improvements may also be made by careful selection of the covariance function. While [15] chose covariance functions heuristically, it may be possible to obtain an objective estimate of the covariance from a set of covariance functions by maximizing the marginal likelihood over the covariance function set and their respective hyperparameters for each vital-sign record. The most selected covariance function can then be considered as the optimal function.

The approach can be further improved using the dependent Gaussian process methodology derived by Boyle and Frean [17], which more accurately models cross-correlation between dependent channels of data.

ACKNOWLEDGMENT

clinical insight into the difficulties of using continuous monitoring in a ward setting.

REFERENCES

[1] R. M. Schein, N. Hazday, M. Pena, B.H. Ruben and C.L. Sprung. Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest*, vol. 98, pp.1388-1392, 1990.

[2] National Patient Safety Association, "Safer care for acutely ill patients: Learning from serious accidents," Technical Report, 2007.

[3] L. Tarassenko, D. Clifton, M. Pinsky, M.Hravnak, J. Woods, and P. Watkinson, "Centile-based early warning scores derived from statistical distributions of vital signs," *Resuscitation*, vol. 82, no. 8, pp. 1013-1018, 2011.

[4] D. Prytherch, G. Smith, P Schmidt, and P Featherstone. ViEWS - Towards a national early warning score for detecting adult inpatient deterioration. *Resuscitation*, vol. 81, no. 8, pp 932-937, 2010

[5] Y. Lin, I. Jan, P. Ko, Y. Chen, J. Wong and G. Jan. A wireless PDA-based physiological monitoring system for patient transport. *IEEE Trans. Information Technology in Biomedicine*. Vol. 8, pp. 439-447, 2004.

[6] L. Tarassenko, A. Hann and D. Young. Integrated monitoring and analysis for early warning of patient deterioration. *British Journal of Anaesthesia*. Vol. 97, p. 64, 2006.

[7] C. Tsien and J Fackler. Poor prognosis for existing monitors in the intensive care unit. *Critical Care Medicine*, Vol. 25, No. 4, p. 466, 1997

[8] C. Oberli, J Urzua, C. Saez, M. Guarini, A.Cipriano, B. Garayar, G. Lema, R. Canessa, C.Sacco, and M. Irarrazeval. An expert system for monitoring alarm integration. *Journal of clinical monitoring and computing*, Vol. 15, pp. 29-35, 1999

[9] R. Schoenberg, D. Sands, and C. Safran. Making ICU alarms meaningful: a comparison of traditional vs. trend-based algorithms. In 1999 *Proc AMIA symposium*. P. 379

[10] Y. Zhang. Real-time development of patient-specific alarm algorithms for critical care. In 2007 *Proc. Engineering in Medicine and Biology Society*, pp. 4351-4354

[11] S. Charbonnier and S. Gentil. A trend-based alarm system to improve patient monitoring in intensive care units. *Control Engineering Practice*, Vol. 15, No. 9, pp. 1039-1050, 2007

[12] S. Hoare and P. Beatty. Automatic artifact identification in anaesthesia patient record keeping: a comparison of techniques. *Medical engineering & physics*. Vol. 22, No. 8, pp. 547-553, 2000

[13] C. Williams, J. Quinn, and N. McIntosh. Factorial switching Kalman filters for condition monitoring in neonatal intensive care. *Advances in Neural Information Processing Systems*, Vol. 18, p. 1513, 2006

[14] L. Tarassenko, A. Hann, A. Patterson, E. Braithwaite, K. Davidson, V. Barber, and D. Young. Biosign: Multi-parameter monitoring for early warning of patient deterioration. In 2005 *Proc. 3rd IEE International Seminar on Medical Application of Signal Processing*, pp. 71-76

[15] O. Stegle, S Fallert, D MacKay, and S. Brage. Gaussian process robust regression for noisy heart rate data. *IEEE Trans. Biomedical Engineering*. Vol. 55, pp 2143-2151, 2008.

[16] S. Wilson, D. Wong, D. Clifton, S. Fleming, R. Way, R. Pullinger and L. Tarrasenko. Track and trigger in an Emergency department: an observational evaluation study. *Emergency Medicine Journal*, submitted for publication.

[17] P. Boyle and M. Frean. Dependent Gaussian processes. In 2005 *Proc. Advances in neural information processing systems 17*. Vol. 17, p. 217