

# Recovering Latent Signals From a Mixture of Measurements Using a Gaussian Process Prior

Felipe Tobar, Gonzalo Rios, Tomás Valdivia, and Pablo Guerrero

**Abstract**—In sensing applications, sensors cannot always measure the latent quantity of interest at the required resolution, sometimes they can only acquire a blurred version of it due the sensor’s transfer function. To recover latent signals when only noisy mixed measurements of the signal are available, we propose the Gaussian process mixture of measurements (GPMM), which models the latent signal as a Gaussian process (GP) and allows us to perform Bayesian inference on such signal conditional to a set of noisy mixture of measurements. We describe how to train GPMM, that is, to find the hyperparameters of the GP and the mixing weights, and how to perform inference on the latent signal under GPMM; additionally, we identify the solution to the underdetermined linear system resulting from a sensing application as a particular case of GPMM. The proposed model is validated in the recovery of three signals: A smooth synthetic signal, a real-world heart-rate time series and a step function, where GPMM outperformed the standard GP in terms of estimation error, uncertainty representation, and recovery of the spectral content of the latent signal.

**Index Terms**—Bayesian inference, convolution process, Gaussian process (GP), sensing applications, underdetermined linear systems.

## I. INTRODUCTION

**O**BSERVATIONS within sensing applications result from the convolution between the latent signal and the sensors’ transfer function; therefore, a desired property of the sensor is to have a transfer function that is close to a Dirac delta function so that the latent signal can be recovered from the observations. We will model this convolution in a discrete manner to give rise to the representation of a sensing application described in Fig. 1, where we model the observations as a (noisy) mixture of (again noisy) measurements and aim to recover the latent signal from the observations. Mixing of the latent signal’s values stems from the inability of the sensor to measure the latent signal at the required resolution, this is due to low quality of the sensors that *color* the observations which have to then be *whiten* in order to

Manuscript received October 11, 2016; accepted November 26, 2016. Date of publication December 7, 2016; date of current version January 26, 2017. The work of F. Tobar was supported in part by CONICYT PAI-82140061 and Basal-CMM. The work of P. Guerrero was supported by FONDECYT Postdoctoral Project 3130729. The work of G. Rios was supported by CONICYT-PCHA Doctorado Nacional 2016-21161789. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mónica F. Bugallo.

F. Tobar, G. Rios, and T. Valdivia are with the Center for Mathematical Modeling, Universidad de Chile, Santiago 8370456, Chile (e-mail: ftobar@dim.uchile.cl; grios@dim.uchile.cl; tomvalh@gmail.com); P. Guerrero is with the Computer Science Department and the Advanced Mining Technology Center, Universidad de Chile, Santiago 8370456, Chile (e-mail: pguerrer@ing.uchile.cl).

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2016.2637312

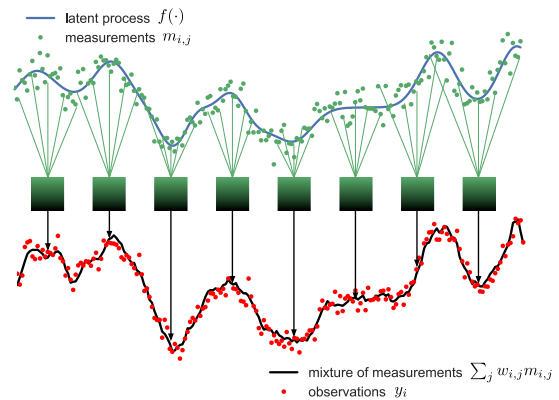


Fig. 1. Sensing application: A latent signal (blue) is measured up to sensor noise (green), these measurements are linearly mixed to yield the mixture-of-observations process (black) which is in turn corrupted by noise (red).

recover the latent process. Observations composed by mixtures of measurements are commonplace in sensing applications in different areas: when using a radar or sonar in robot localization [1], [2], in astronomical applications [3], and in super-resolution image recovery [4], to name but a few.

A workaround to the problem of recovering a latent process from observations composed by mixtures of measurements is to define a set of sensing locations (i.e., a grid) and model the observations as a system of linear equations—recall that the measurements are combined in a linear fashion. However, this approach is rather restrictive since it constrains the measurements to be collected at fixed locations, which is rarely the case in real-world applications, and it leads to heavily underdetermined linear systems. In our view, the key to this problem is to assume spatial correlations in the latent signals so that observations of overlapping regions reveal structure in the signal; we do so in a probabilistic manner by placing a Gaussian process (GP) prior [5] on the signal, which is then updated into the posterior density of the latent signal conditional to such observations. However, GPs in their standard form are not suited to deal with observations comprising a mixture of measurements.

In this paper, we propose a GP-based generative model for the problem of mixture of measurements, illustrate how to fit it to observed data and perform inference on the latent signal in a probabilistic manner. We also interpret the linear system approach to sensing applications as a particular case of the proposed framework and validate our model on real-world and synthetic data.

## A. Background: Gaussian Processes and Related Work

A Gaussian process (GP) [5] is a nonparametric prior distribution on functions  $f : X \mapsto \mathbb{R}$ , for any set  $X$  where a

covariance function can be defined (e.g., a metric space), such that for any finite collection of inputs  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in X^n$ , the corresponding outputs  $f(x_1), f(x_2), \dots, f(x_n) \in \mathbb{R}$  are jointly Gaussian, that is,

$$[f(x_1), f(x_2), \dots, f(x_n)]^T \sim \mathcal{N}(\mu(\mathbf{x}), K(\mathbf{x}, \mathbf{x})) \quad (1)$$

where the mean and covariance functions are denoted by  $\mu(\cdot)$  and  $K(\cdot, \cdot)$ , respectively.

The GP framework is well suited to the sensing setting in Fig. 1, since modeling the observed signal as a GP leads to the posterior distribution of the latent process (given the mixture of measurements) being Gaussian. GP-based convolution models [6]–[9] consider white-noise latent signals, which is unsuitable in sensing applications where spatial correlations are fundamental. Conversely, [10] allows GPs as latent functions but to perform inference on the outputs rather than the latent processes. These methods consider continuous-time convolutions, which are computationally demanding and require, e.g., variational approximations [11], [12], consequently, closed-form and computationally-efficient Bayesian reconstruction of the latent process is still an open problem in sensing applications.

## II. GAUSSIAN PROCESS MIXTURE OF MEASUREMENTS

Consider a sensing application where each observation  $y_i$  is a noisy mixture of, again noisy and hidden, measurements  $m_{i,j}$  of a latent process  $f(\cdot)$  measured at locations  $x_{i,j}$ , that is,

$$m_{i,j} = f(x_{i,j}) + \epsilon_{i,j} \quad (2)$$

$$y_i = \sum_{j=1}^M w_{i,j} m_{i,j} + \eta_i \quad (3)$$

where for the  $i^{\text{th}}$  observation  $y_i$ , we use the following notation:

- 1)  $f(x_{i,j})$  is the value of the latent process at the location  $x_{i,j}$ ;
- 2)  $m_{i,j}$  is the measurement of  $f(x_{i,j})$  acquired by the sensor;
- 3)  $w_{i,j}$  is the weight of the measurement  $m_{i,j}$ ;
- 4)  $M$  is the number of measurements in the observation  $y_i$ ;
- 5)  $\epsilon_{i,j}$  is the measurement Gaussian noise; and
- 6)  $\eta_i$  is the observation Gaussian noise.

We use the terms *measurement* and *observation* defined here consistently in the rest of the letter—see Fig. 1. Note that, in general, different observations, i.e.,  $y_i, y_{i'}$   $i \neq i'$ , correspond to different regions and both the locations  $x_{i,j}, x_{i',j'}$ ,  $\forall i, j, j'$  and the weights  $w_{i,j}, w_{i',j'}$ ,  $\forall i, j, j'$  might be different due to the sensing procedure.

For  $N$  observations, we express (3) as a system of linear equations in block matrix notation:

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \mathbf{w}_1 & & \\ & \ddots & \\ & & \mathbf{w}_N \end{bmatrix}^T}_{\mathbf{W}} \underbrace{\begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_M \end{bmatrix}}_{\mathbf{m}} + \underbrace{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_N \end{bmatrix}}_{\boldsymbol{\eta}} \quad (4)$$

where  $\mathbf{w}_i = [w_{i,1}, \dots, w_{i,M}]^T$ ,  $\mathbf{m}_i = [m_{i,1}, \dots, m_{i,M}]^T$  and the matrix  $\mathbf{W}$  is an  $M$ -diagonal matrix. Column vectors are denoted in bold lowercase font and matrices in bold uppercase font.

### A. Solving the Linear System

Our approach will consider the weights  $w_{i,j}$  to be unknown and learn them from the observations; however, let

us assume they are known in this section (that is, there is a model for the sensor's transfer function) in order to address the problem as a linear system; for ease of notation, we assume there are no common locations across the measurements (i.e.,  $i \neq i', j \neq j' \Rightarrow x_{i,j} \neq x_{i',j'}$ ). With these assumptions, (4) is an underdetermined linear system: There are  $L = MN$  unknowns and  $N$  equations, where  $N \ll L$ , in fact, the general solution to such a system has  $L - N$  free parameters or degrees of freedom—we neglect underdetermined system with no solutions (i.e., *inconsistent*), since we are dealing with a physical signal and a solution *does* exist.

This underdetermined system has infinite solutions, with the minimum-norm solution given by  $\hat{\mathbf{m}} = \mathbf{W}^+ \mathbf{y}$ , where  $\mathbf{W}^+$  is the Moore–Penrose pseudoinverse of  $\mathbf{W}$  [13]. Using this solution to recover the latent signal has a number of drawbacks: (i) It requires the weights  $w_{i,j}$  to be known, (ii) it only recovers the process at the measured locations without providing any insight about regions not measured, (iii) it does not provide a measure of uncertainty for the estimates, e.g., in the form of error bars.

### B. Novel Generative Model for the Mixture of Measurements

As the system in (4) has infinite solutions, a constraint (or regularization criterion) has to be imposed to reduce the number of solutions, or critically, to find a single solution. The Moore–Penrose pseudoinverse imposes the minimum-norm criterion, we instead assume a probabilistic condition (a prior) on the spatial structure of the solution. Specifically, we (i) place a GP prior on the latent signal to then (ii) find the posterior distribution of the latent signal conditional to the mixture of measurements, even at locations that were not measured. A key property of this approach is that these two steps are performed analytically, since the latent signal and the mixture of measurements are jointly Gaussian. We next present a formal description of the proposed generative model.

We model the latent process  $f(\cdot)$  in (2) as a GP over the set of locations  $X$  given by

$$f \sim \mathcal{GP}(\mu_f, K_f) \quad (5)$$

where  $\mu_f : X \mapsto \mathbb{R}$  and  $K_f : X \times X \mapsto \mathbb{R}$  are the GP mean and covariance functions, respectively. As the linear combination of jointly Gaussian random variables (RVs) is Gaussian, the observations in (3) are Gaussian RVs indexed by  $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,M}]^T \in X^M$  with mean and covariance, respectively, given by

$$\mu_y(\mathbf{x}_i) = \mathbb{E}(y_i) = \sum_{j=1}^M w_{i,j} \mu_f(x_{i,j}) \quad (6)$$

$$K_y(\mathbf{x}_i, \mathbf{x}_{i'}) = \mathbb{E}((y_i - \mu_y(\mathbf{x}_i))(y_{i'} - \mu_y(\mathbf{x}_{i'}))) \quad (7)$$

$$= \sum_{j,j'=1}^M w_{i,j} w_{i',j'} (K_f(x_{i,j}, x_{i',j'}) + \sigma_\epsilon^2 \delta(x_{i,j} - x_{i',j'})) + \delta_{i-i'} \sigma_\eta^2$$

where  $\sigma_\eta^2$  and  $\sigma_\epsilon^2$  are the variances of the measurement and observation noises respectively. In the continuous-location case, measurements are taken always at different locations, thus we have  $\delta(x_{i,j} - x_{i',j'}) = 0 \forall i, i', j, j'$  and

$$K_y(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^M \sum_{j'=1}^M w_{i,j} w_{i',j'} K_f(x_{i,j}, x_{i',j'}) + \delta_{i-i'} \sigma_\eta^2 \quad (8)$$

$$= \mathbf{w}_i^T K_f(\mathbf{x}_i, \mathbf{x}_{i'}) \mathbf{w}_{i'} + \delta_{i-i'} \sigma_\eta^2 \quad (9)$$

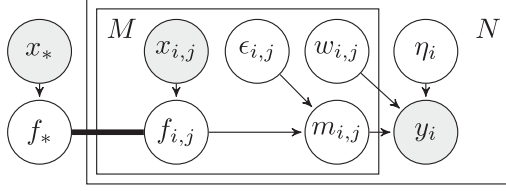


Fig. 2. Graphical model of GPMM. Hidden variables are left blank whereas observed ones are shaded. The inner plate represents the  $(i, j)$ th measurement, the outer one the  $i$ th observation, and  $x_*$  and  $f_*$  are a test location and its value, respectively. The thick bar indicates that the latent GP  $f$  is completely interconnected.

where  $K_f(\mathbf{x}_i, \mathbf{x}_{i'}) \in \mathbb{R}^{M \times M}$  is the Gram matrix evaluated on  $\mathbf{x}_i$  and  $\mathbf{x}_{i'}$ ,  $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,M}]^\top$  and  $\mathbf{w}_i = [w_{i,1}, \dots, w_{i,M}]^\top$ .

The covariance of the observations is a mixture of evaluations of the covariance function of the process  $f(\cdot)$  at the locations measured. This implies that a single entry in common between measurement locations  $\mathbf{x}_i$  and  $\mathbf{x}_j$  is enough for the observations  $y_i(\mathbf{x}_i)$  and  $y_j(\mathbf{x}_j)$  to be correlated. This mixture-of-kernel structure resembles additive GPs [14] and multikernel learning [15]–[17]; however, these methods combine different kernels (evaluated on a common input) for expressive kernel design, whereas the proposed model combines evaluations of a single kernel on different locations to find the spatial structure of the latent signal. We refer to the presented model as Gaussian processes mixture of measurements (GPMM) and give a graphical model illustration in Fig. 2.

### III. INFERENCE FOR GPMM AND RELATIONSHIP TO THE MOORE–PENROSE PSEUDOINVERSE

Fitting the model to the observations  $\mathcal{D} = \{(\mathbf{x}_i, y_i), i = 1, \dots, N\}$  involves finding appropriate hyperparameters for GPMM,  $\theta_{\text{GPMM}}$ , that is, the hyperparameters of the latent signal in (5) and the weights  $w_i$ , which are now hyperparameters. This can be achieved by minimizing the negative log-likelihood  $\log p(\mathbf{y})$ :

$$\theta_{\text{GPMM}} = \arg \min \frac{1}{2} \mathbf{y}^\top K_y^{-1} \mathbf{y} + \frac{1}{2} \log |K_y| + \frac{N}{2} \log 2\pi$$

where  $K_y$  is given in (7),  $\mathbf{y} = [y_1, \dots, y_N]^\top$ , and the optimization can be performed by, e.g., gradient descent. The cost of computing  $K_y$  in (9) and inverting it are, respectively,  $\mathcal{O}(N^2 M^2)$  and  $\mathcal{O}(N^3)$ , meaning that if  $M$  is dominated by  $\mathcal{O}(N^{1/2})$  the cost of training GPMM is  $\mathcal{O}(N^3)$  as standard GPs. With the optimal hyperparameters, we are now ready to calculate the posterior  $f(\cdot)$ .

#### A. The Posterior of the Latent Process

The posterior density of  $f$  given the observations  $\mathcal{D}$ ,  $p(f|\mathcal{D})$ , is Gaussian and determined by the following: (i) the prior mean—assumed to be zero in this case; (ii) the autocovariance of the mixture-of-measurement process  $y$ —given in (7); and (iii) the covariance between the latent process  $f(x)$  and the observation at  $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,M}]^\top$  given by  $y(\mathbf{x}_i) = y_i = \sum_{j=1}^M w_{i,j} m_{i,j}$ , this covariance is

$$K_{fy}(x, \mathbf{x}_i) = \mathbb{E}((f(x) - \mu_f(x))(y_i - \mu_y(\mathbf{x}_i))) \\ = \sum_{j=1}^M w_{i,j} K_f(x, x_{i,j}) = \mathbf{w}^\top K_f(x, \mathbf{x}_i). \quad (10)$$

Denoting  $\mathbf{y} = [y(\mathbf{x}_1), \dots, y(\mathbf{x}_N)]^\top$ , the predictive posterior is

$$\mathbb{E}(f(x)|\mathbf{y}) = K_{fy} K_y^{-1} \mathbf{y} \quad (11)$$

$$\text{Var}(f(x)|\mathbf{y}) = K_f(x, x) - K_{fy} K_y^{-1} K_{yf} \quad (12)$$

thus, the reconstruction of the latent process  $f$  can be computed in closed form (we used the notation  $K_{fy}^\top = K_{yf}$ ).

#### B. Moore–Penrose Solution is a Particular Case of GPMM

Without loss of generality, let us consider that observations  $y_i$  were taken at a grid  $\mathbf{x}_{\text{grid}} = [x_1, \dots, x_M]$  were only a few weights are nonzero per observation. Furthermore, without evidence for spatial correlation of the latent process  $f(\cdot)$ , its covariance matrix is the identity multiplied by  $\sigma_f^2$  (signal power)—i.e.,  $\mathbf{K}(\mathbf{x}_{\text{grid}}, \mathbf{x}_{\text{grid}}) = \sigma_f^2 \mathbf{I}_M$ . Consequently, denoting the mixing weights by  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]$  and  $\mathbf{x}$  the input of the observed process, from (9) and (10) the covariances are given by

$$K_y(\mathbf{x}, \mathbf{x}) = (\sigma_\epsilon^2 + \sigma_f^2) \mathbf{W}^\top \mathbf{W} + \sigma_\eta^2 \mathbf{I}_M \quad (13)$$

$$K_{fy}(x, \mathbf{a}) = \sigma_f^2 \mathbf{W}^\top. \quad (14)$$

Finally, introducing the above two expressions in the posterior mean in (11) gives the solution to the linear system

$$\hat{f}(\mathbf{x}) = \frac{\sigma_f^2}{\sigma_\epsilon^2 + \sigma_f^2} \mathbf{W}^\top \left( \mathbf{W}^\top \mathbf{W} + \frac{\sigma_\eta^2}{\sigma_\epsilon^2 + \sigma_f^2} \mathbf{I}_M \right)^{-1} \mathbf{y}. \quad (15)$$

The connection between solutions to linear systems and GP models is therefore established: When the noise variances are negligible w.r.t. the signal power ( $\sigma_f^2 \gg \sigma_\epsilon^2, \sigma_\eta^2$ ), the ratios in (15)  $\sigma_f^2/(\sigma_\epsilon^2 + \sigma_f^2) \rightarrow 1$  and  $\sigma_\eta^2/(\sigma_\epsilon^2 + \sigma_f^2) \rightarrow 0$ , and the Moore–Penrose inverse is obtained. On the contrary, when the noise variance  $\sigma_\epsilon^2$  is large, the estimate decays to zero, or *reverts to the prior*, since the measurements are not reliable. Furthermore, when the observation noise  $\sigma_\eta^2 = 0$ , eq. (15) is equivalent to the ordinary least squares and when  $\sigma_\eta^2 > 0$  to the regularized least squares (ridge regression).

Finally, we emphasize that, unlike the Moore–Penrose method, the proposed GPMM jointly infers the mixing weights and the complete latent process, while also providing a measure of uncertainty given by the variance in eq. (12).

## IV. SIMULATIONS

The proposed GPMM model was validated using three datasets: A heart-rate time series, a smooth function generated by a GP with square exponential (SE) covariance kernel, and the Heaviside step function. All three experiments consisted in recovering the latent process from noisy mixtures of measurements by fitting GPMM to the observations and then computing the predictive posterior as described in Section III-A, where the learnt weights were constrained to have unit  $L_1$ -norm, to have positive entries and to be symmetric to avoid redundant solutions. The proposed GPMM was compared to the standard GP that considers the observations as a single measurement in a single location, which is the common practice in sensing applications.

#### A. Smooth Synthetic Signal

This first toy example considered a draw from a GP with SE covariance as the latent function and 120 observations with 7 measurements per observation. Fig. 3 shows the GP estimates

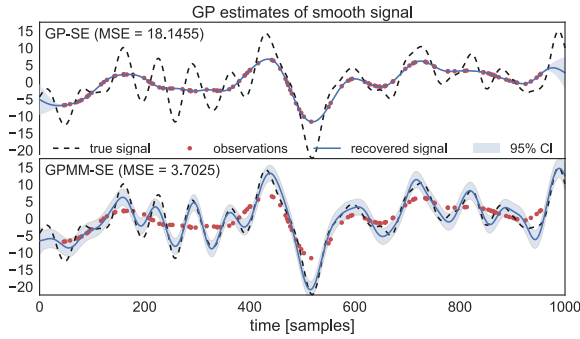


Fig. 3. Recovering a synthetic signal (dashed) from mixture of measurements (red): standard GP-SE (top, blue) and the proposed GPMM-SE (bottom, blue).

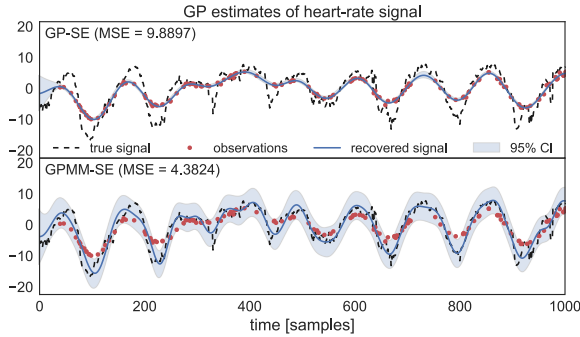


Fig. 4. Recovering a heart-rate signal (dashed) from mixture of measurements (red): standard GP-SE (top, blue) and the proposed GPMM-SE (bottom, blue).

and their mean square error (MSE) for both the standard GP with SE kernel (termed GP-SE) and the proposed GPMM with SE kernel (termed GPMM-SE). Notice how GP-SE fails to recover all the extrema of the latent process and adjust very tightly to the observations, this is because the convolution performed by the sensor smooths out the extrema in the observations, which are then trusted as true values by GP-SE. Conversely, the GPMM-SE was able to recover all the extrema, place appropriate error bars on the latent function and report a lower estimation error.

### B. Heart-Rate Signal

The rationale behind validating the proposed GPMM on a heart-rate signal is that instantaneous frequency is calculated averaging the signal over a small time window, we used the data from [18] and constructed a lower resolution version of it composed by a mixture of measurements. Fig. 4 shows the recovery of a heart-rate signal from such observations for both the GP-SE and the proposed GPMM-SE, where the GPMM-SE again outperforms the GP-SE in terms of estimation MSE and uncertainty representation. The error bars are larger in this case, since, unlike the synthetic example, the heart-rate signal is non-smooth and has irregular plateaus that are attributed to noise when a SE kernel is used. We used 240 observations with 7 measurements each.

### C. Heaviside Step Function

Motivated by edge-detection applications, we then considered the Heaviside step function. To cater for discontinuous signals, we used the neural network (NN) kernel [19] and implemented both the standard GP-NN and the proposed GPMM-NN to recover the latent step function. Observe from Fig. 5 how GPMM-NN successfully recovered the discontinuity with low variance. We used 120 observations with 7 measurements each.

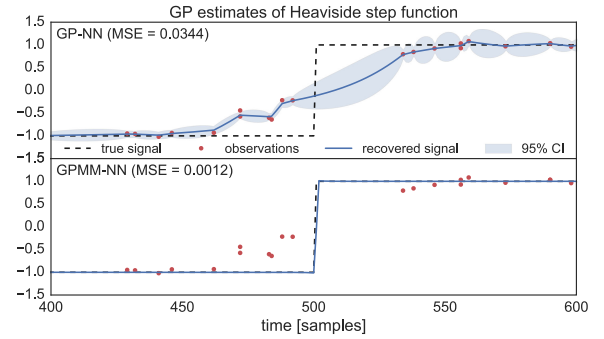


Fig. 5. Recovering a step function (dashed) from mixture of measurements (red): standard GP-SE (top, blue) and the proposed GPMM-SE (bottom, blue).

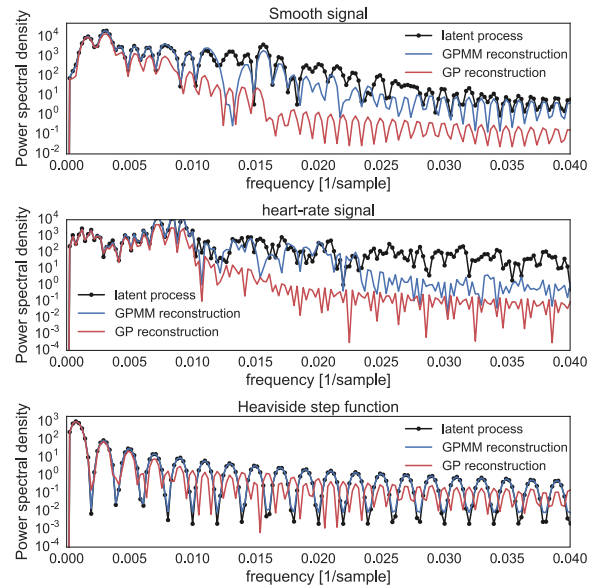


Fig. 6. Power spectral density of true latent signal (black), proposed GPMM (blue), and standard GP (red) for all datasets considered: smooth signal (top), heart-rate signal (middle), and step function (bottom).

### D. Comparing Recovered Signals in Spectral Terms

Fig. 6 shows the PSDs for the smooth, heart-rate, and step function from top to bottom, observe how the GPMM (blue) was able to better recover the spectrum in all the experiments considered unlike the standard GP (red), where the latter failed to identify the spectral content of the latent process that was removed by the convolution sensor.

## V. CONCLUSIONS

We have proposed the GPMM to address the problem of recovering a latent signal from a noisy mixture of measurements of such signal, a common setting in sensing applications. Our contributions are as follows: (i) to model the latent process and the mixture-of-measurements process as jointly Gaussian; (ii) fitting GPMM, including the mixing weights, and deriving the posterior distribution of the latent function in closed form; (iii) interpreting the solution of the underdetermined linear system generated by the sensing application as a particular case of GPMM; and (iv) validating GPMM against the standard GP for synthetic and real-world signals, where the reconstruction accuracy of GPMM was evidenced both in the time and frequency domains.

## REFERENCES

- [1] M. Adams, J. Mullane, and B.-N. Vo, "Laser and radar based robotic perception," *Found. Trends Robot.*, vol. 1, no. 3, pp. 135–252, 2011.
- [2] H. P. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in *Proc. IEEE Int. Conf. Robot. Autom.*, 1985, vol. 2, pp. 116–121.
- [3] J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," *Publ. Astron. Soc. Pac.*, vol. 114, no. 800, 2002, Art. no. 1051.
- [4] L. Baboulaz and P. L. Dragotti, "Exact feature extraction using finite rate of innovation principles with an application to image super-resolution," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 281–298, Feb. 2009.
- [5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Cambridge, MA, USA: MIT Press, 2006.
- [6] F. Tobar, T. D. Bui, and R. E. Turner, "Learning stationary time series using Gaussian processes with nonparametric kernels," in *Advances in Neural Information Processing Systems 28*. Cambridge, MA, USA: MIT Press, 2015, pp. 3483–3491.
- [7] F. Tobar, T. D. Bui, and R. E. Turner, "Design of covariance functions using inter-domain inducing variables," in *Proc. Neural Inf. Process. Syst. Time Series Workshop*, 2015, <https://sites.google.com/site/nipsts2015/accepted-papers>
- [8] P. Boyle and M. Frean, "Dependent Gaussian processes," in *Advances in Neural Information Processing Systems 17*. Cambridge, MA, USA: MIT Press, 2005, pp. 217–224.
- [9] D. Higdon, "Space and space-time modeling using process convolutions," in *Quantitative Methods for Current Environmental Issues*, pp. 37–56. Springer, U.K.: Springer, 2002.
- [10] M. Álvarez and N. D. Lawrence, "Sparse convolved Gaussian processes for multi-output regression," in *Advances in Neural Information Processing Systems 21*. Cambridge, MA, USA: MIT Press, 2009, pp. 57–64.
- [11] M. K. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Proc. Int. Conf. Artif. Intell. Stat.*, 2009, vol. 12, pp. 567–574.
- [12] S. P. Chatzis and D. Kosmopoulos, "A latent manifold Markovian dynamics Gaussian process," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 1, pp. 70–83, Jan. 2015.
- [13] E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bull. Amer. Math. Soc.*, vol. 26, pp. 394–395, 1920.
- [14] D. K. Duvenaud, H. Nickisch, and C. E. Rasmussen, "Additive Gaussian processes," in *Advances in Neural Information Processing Systems 24*. Cambridge, MA, USA: MIT Press, 2011, pp. 226–234.
- [15] M. Gönen and E. Alpaydm, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, 2011.
- [16] F. Tobar, S.-Y. Kung, and D. P. Mandic, "Multikernel least mean square algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 265–277, Feb. 2014.
- [17] M. Yukawa, "Multikernel adaptive filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4672–4682, Sep. 2012.
- [18] A. L. Goldberger and D. R. Rigney, *Theory of Heart: Biomechanics, Biophysics, and Nonlinear Dynamics of Cardiac Function*, chapter Nonlinear dynamics at the bedside. New York, NY, USA: Springer-Verlag, 1991, pp. 583–605.
- [19] C. K. I. Williams, "Computation with infinite neural networks," *Neural Comput.*, vol. 10, no. 5, pp. 1203–1216, 1998.