

Denoising Smooth Signals Using a Bayesian Approach: Application to Altimetry

Abderrahim Halimi, *Member, IEEE*, Gerald S. Buller, Stephen McLaughlin, *Fellow, IEEE*, and Paul Honeine, *Member, IEEE*

Abstract—This paper presents a novel Bayesian strategy for the estimation of smooth signals corrupted by Gaussian noise. The method assumes a smooth evolution of a succession of continuous signals that can have a numerical or an analytical expression with respect to some parameters. The proposed Bayesian model takes into account the Gaussian properties of the noise and the smooth evolution of the successive signals. In addition, a gamma Markov random field prior is assigned to the signal energies and to the noise variances to account for their known properties. The resulting posterior distribution is maximized using a fast coordinate descent algorithm whose parameters are updated by analytical expressions. The proposed algorithm is tested on satellite altimetric data demonstrating good denoising results on both synthetic and real signals. In comparison with state-of-the-art algorithms, the proposed strategy provides a good compromise between denoising quality and necessary reduced computational cost. The proposed algorithm is also shown to improve the quality of the altimetric parameters when combined with a parameter estimation or a classification strategy.

Index Terms—Altimetry, Bayesian algorithm, coordinate descent algorithm (CDA), gamma Markov random fields (gamma-MRFs).

I. INTRODUCTION

IN MANY applications, the development of new sensor technologies allows for high-speed acquisition of a succession of signals, leading to a slight variation from one signal to the next. This is the case for satellite altimetric signals that can be described as a succession of continuous functions corrupted by noise [1], [2]. Indeed, when observing the ocean, the altimetric successive signals show a reduced variation due to the nature of ocean (see Fig. 1 that shows a succession of 800 signals acquired by the Jason-2 mission). This paper aims to exploit this correlation to denoise the observed altimetric signals.

Manuscript received February 10, 2016; revised July 23, 2016 and October 27, 2016; accepted November 10, 2016. Date of publication January 16, 2017; date of current version March 22, 2017. This work was supported by the Engineering and Physical Sciences Research Council under Grant EP/J015180/1, Grant EP/N003446/1, and Grant EP/K015338/1. (*Corresponding author: Abderrahim Halimi.*)

A. Halimi, G. S. Buller, and S. McLaughlin are with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K. (e-mail: a.halimi@hw.ac.uk; g.s.buller@hw.ac.uk; S.McLaughlin@hw.ac.uk).

P. Honeine is with the Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes, Université de Rouen-Normandie, 76000 Rouen, France (e-mail: paul.honeine@univ-rouen.fr).

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

Digital Object Identifier 10.1109/JSTARS.2016.2629516

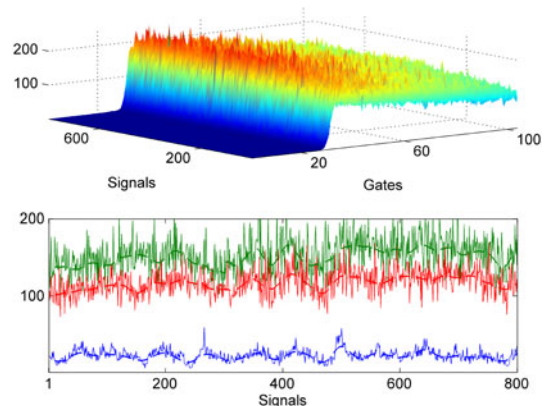


Fig. 1. (Top) Example of 800 noisy Jason-2 signals. (Bottom) Signal evolutions in gates (30, 50, 90) (continuous lines) and their smooth approximation with the proposed SSE algorithm (dashed lines). The blue, green, and red lines are associated with the gates 30, 50, and 90, respectively.

A satellite altimeter is a nadir-viewing radar that emits regular pulses and records the travel time, the magnitude, and the shape of each return signal after reflection on the oceanic surface. This reflected echo provides information about some physical parameters such as the range between the satellite and the observed scene (denoted by τ), the significant wave height (denoted by μ), and the wind speed (related to the signal's amplitude P_u). The shape of the signals carries additional information related to the observed regions (such as coastal zone, iceberg, etc.), which promotes the development of different altimetric classification methods [3], [4].

Altimetric signals are corrupted by speckle noise, and many recent studies and missions have been focusing on improving the quality of these signals by reducing the effects of the noise. This goal can be achieved by improving the instrument technology [5], [6], the physical altimetric models [2], [7]–[9], or the signal processing algorithms [10], [11]. In this paper, we focus on signal processing approaches that can be divided into two categories, as shown in Fig. 2. The first class focuses on the altimetric parameters by providing realistic smooth estimates [10]–[12]. The second class is more general and consists in filtering the noisy signals [13], [14]. The filtered signals can then be processed by several classical techniques, which have different goals such as parameter estimation [7], [10], [11], signal classification [4], and object detection [15]. This study follows

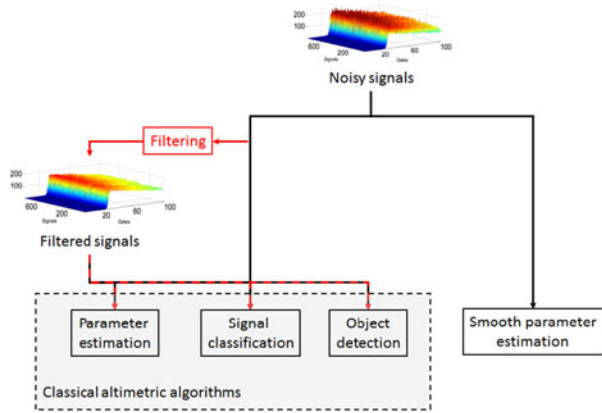


Fig. 2. Scheme summarizing some of the existing altimetric approaches and the proposed filtering algorithm (in red).

the second class and proposes a new filtering algorithm that operates on the observed signals. The resulting algorithm differs from the CD-BM algorithm proposed in [12], since it can be combined with many classical altimetric algorithms for different purposes, while CD-BM is specialized on the extraction of realistic altimetric parameters.

Filtering successive signals can be seen as a classical image denoising problem, which has been tackled using different approaches such as empirical, sparsity-based, and statistical methods, etc. Empirical methods include the empirical mode decomposition [16], [17] and other heuristic algorithms [18]. Sparsity-based methods account for the sparsity of the data under a given signal representation. This class includes optimization-based algorithms such as FISTA [19] and SALSA [20], and collaborative methods as the well-known BM3D [21], and its variant for speckle noise SBM3D [22]. Statistical methods are generally based on the noise statistics of the image. This class includes, as an example, the well-known PPB algorithm [23], Kalman filtering [24], particle filtering methods [25], [26], and Gaussian-process-based methods [27]. This last class is considered in this paper by proposing a new Bayesian algorithm especially designed to denoise altimetric signals under the requirements of a good denoising quality and a reduced computational cost. It should be noted that the provided list is far from being exhaustive, and we refer the interested reader to [18], [28], and [29] for more details regarding other methods.

The first contribution of this paper is a hierarchical Bayesian model to denoise a set of smooth altimetric signals corrupted by a speckle noise. As a result of the averaging on-board the altimetric satellite, the speckle noise is generally approximated by an independent and identically distributed Gaussian noise [7], [11]. The proposed Bayesian model generalizes this assumption by considering that each signal is corrupted by additive independent and nonidentically distributed Gaussian noise. A gamma Markov random field (gamma-MRF) prior [30] is also considered to account for the correlation between the noise variances to better approximate the speckle noise. The signal energies are also assigned a gamma-MRF prior to account for their

continuity. Using the Bayes rule, the likelihood and the prior distributions lead to a posterior distribution that is used to estimate the noiseless signals and the noise parameters (as described in the next paragraph). Note that the proposed Bayesian hierarchy is generic in the sense that it does not assume a specific signal model. Indeed, the signal can be expressed by a numerical formula or by linear/nonlinear analytical function with respect to (w.r.t.) some parameters. Therefore, the proposed algorithm can be easily adapted to coastal altimetric signals [31], [32] or other altimetric technologies such as delay/Doppler altimetry [6], [33], [34]. It can also be generalized to hyperspectral images, which often present a high spectral correlation, as already exploited by many algorithms in the literature to denoise these images [35], [36].

The second contribution of this paper is the derivation of a denoising algorithm associated with the proposed hierarchical Bayesian model. The minimum mean square error (MMSE) and maximum *a posteriori* (MAP) estimators of the unknown signals/parameters cannot be easily computed from the obtained joint posterior. In this paper, the MAP estimator is evaluated by considering a coordinate descent algorithm (CDA) [37], [38] whose convergence to a stationary point is ensured. The proposed algorithm sequentially updates the estimated noiseless signals, noise variances, and other hyperparameters by analytical formulas leading to a reduced computational cost. The proposed Bayesian model and the estimation algorithm are validated using synthetic and real altimetric data acquired during the Jason-2 mission. The obtained results are very promising and show the potential of the proposed denoising strategy.

This paper is organized as follows. Section II introduces the observation model and the considered altimetric signal. The proposed hierarchical Bayesian model and its estimation algorithm are introduced in Sections III and IV. Section V evaluates and compares the proposed technique with the state-of-the-art algorithms using simulated data with controlled ground truth and real Jason-2 altimetric data. Finally, conclusions and future work are reported in Section VI.

II. PROBLEM FORMULATION

Consider M successive signals $\mathbf{S} \in \mathbb{R}^{K \times M}$ and let $\mathbf{Y} \in \mathbb{R}^{K \times M}$ denote their noisy version. Let $\mathbf{y}_{:m} \in \mathbb{R}^{K \times 1}$ be the m th column of \mathbf{Y} and $\mathbf{y}_{k\cdot} \in \mathbb{R}^{1 \times M}$ its k th row, representing the k th temporal gate for all signals. For notation simplicity, we denote $\mathbf{y}_{:m} = \mathbf{y}_m$, for $m = 1, \dots, M$, and $\mathbf{y}_{k\cdot} = \mathbf{y}_k$, for $k = 1, \dots, K$ (the same notation is used for \mathbf{s}). The observation model is given by

$$\mathbf{y}_m = \mathbf{s}_m + \mathbf{e}_m, \text{ with } \mathbf{e}_m \sim \mathcal{N}(\mathbf{0}_K, \boldsymbol{\Sigma}) \quad (1)$$

where \sim means “is distributed according to,” \mathbf{y}_m and \mathbf{s}_m are $(K \times 1)$ vectors representing the m th observed and noiseless signals, and \mathbf{e}_m is a centered Gaussian noise vector with a diagonal covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma}^2)$, with $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_K^2)^T$ a $(K \times 1)$ vector. The signals \mathbf{S} might depend on some parameters (by a linear or nonlinear expression), which are denoted by the $(1 \times H)$ vector $\boldsymbol{\Theta}_m = [\theta_1(m), \dots, \theta_H(m)]$

containing the H parameters of the m th signal. Note, however, that the proposed method does not necessarily require a parametric expression for \mathbf{S} and is valid provided that the signals satisfy some properties (as described in the following). In different applications such as oceanic altimetry [10], [11], the successive signals show a reduced variation mainly because of the correlation between the successive physical parameters $\Theta = (\Theta_1^T, \dots, \Theta_M^T)^T$ (see Fig. 1). This smooth variation can be highlighted by expressing the observed signals (1) as follows:

$$\mathbf{y}_k = \mathbf{s}_k + \mathbf{e}_k, \text{ with } \mathbf{e}_k \sim \mathcal{N}(\mathbf{0}_M, \sigma_k^2 \mathcal{I}_M) \quad (2)$$

where $k \in \{1, \dots, K\}$ indexes the signal samples that are known as ‘‘temporal gates,’’ \mathcal{I}_M denotes the $(M \times M)$ identity matrix, and \mathbf{s}_k is a smooth $(M \times 1)$ vector representing the signal evolution at the k th gate (see Fig. 1(bottom) for examples). The proposed Bayesian method aims to filter the observed signals $\mathbf{y}_k, k \in \{1, \dots, K\}$, to retrieve the noiseless signals $\mathbf{s}_k, k \in \{1, \dots, K\}$. The next section introduces the hierarchical Bayesian model proposed to achieve this goal.

III. HIERARCHICAL BAYESIAN MODEL

This section introduces a hierarchical Bayesian model to denote M successive signals. The Bayesian approach first requires the determination of the likelihood that is based on the statistical model associated with the observed data. Second, the known properties of the parameters of interest are modeled via suitable prior distributions. The Bayes theorem allows the likelihood and the priors to be combined to build the posterior distribution of the statistical model. More precisely, if $f(\mathbf{X})$ denotes the prior distribution assigned to the parameter \mathbf{X} , the Bayesian approach computes the posterior distribution of \mathbf{X} using the Bayes rule

$$f(\mathbf{X}|\mathbf{Y}) = \frac{f(\mathbf{Y}|\mathbf{X})f(\mathbf{X})}{f(\mathbf{Y})} \propto f(\mathbf{Y}|\mathbf{X})f(\mathbf{X}) \quad (3)$$

where \propto means ‘‘proportional to’’ and $f(\mathbf{Y}|\mathbf{X})$ is the likelihood of the observation vector \mathbf{Y} . The parameter \mathbf{X} is then estimated from the posterior distribution by computing its MMSE estimator (the mean of the distribution) or its MAP estimator (the maximum of the distribution). The following sections introduce the likelihood and the prior distributions considered in this paper. The unknown parameters of the proposed model include the $(K \times M)$ matrix representing the noiseless signals \mathbf{S} and the $(K \times 1)$ vector σ containing the noise variances associated with the M considered signals.

A. Likelihood

The observation model defined in (2) and the Gaussian properties of the noise sequence $\mathbf{e}_k, k \in \{1, \dots, K\}$, yield

$$f(\mathbf{y}_k|\mathbf{s}_k, \sigma_k^2) \propto \left(\frac{1}{\sigma_k^2}\right)^{\frac{M}{2}} \exp\left(-\frac{\|\mathbf{y}_k - \mathbf{s}_k\|^2}{2\sigma_k^2}\right) \quad (4)$$

where $\|\cdot\|$ denotes the standard ℓ_2 norm such that $\|\mathbf{x}\|^2 = \mathbf{x}^T \mathbf{x}$. Assuming independence between the temporal samples

of the observed signals leads to

$$f(\mathbf{Y}|\mathbf{S}, \Theta) \propto \prod_{k=1}^K f(\mathbf{y}_k|\mathbf{s}_k, \sigma_k^2). \quad (5)$$

B. Priors for the Observed Signal

As previously assumed, the successive observed signals evolve slowly leading to smooth vectors \mathbf{s}_k , for $k \in 1, \dots, K$ [see Fig. 1(bottom)]. This property is satisfied by considering a conjugate Gaussian prior for \mathbf{s}_k ensuring smoothness as follows:

$$\mathbf{s}_k|\epsilon_k^2 \sim \mathcal{N}(\mathbf{0}_M, \epsilon_k^2 \mathbf{H}) \quad (6)$$

where \mathbf{H} is an $(M \times M)$ matrix representing the squared-exponential covariance function given by $H(m, m') = \exp[-\frac{(m-m')^2}{(\vartheta)^2}]$, which introduces the correlation between the successive signals, ϑ controls the degree of correlation between the signals, and ϵ_k^2 is a variance parameter that is gate dependent. From (6), it is clear that this variance is related to the energy of the signals at the k th gate (via the norm $\mathbf{s}_k^T \mathbf{H}^{-1} \mathbf{s}_k$). Moreover, because of the continuity of the signal \mathbf{s}_m w.r.t. the temporal gates, the signal energies vary smoothly from one gate to another. Therefore, we expect ϵ_k^2 to vary smoothly from one gate to another, which will be introduced by considering a specific prior for ϵ_k^2 , as explained in Section III-D. As a result, the proposed model accounts for the individual signal continuity w.r.t. temporal gates and for the smooth evolution existing between successive signals. An interesting property of this prior is that the signals \mathbf{s}_k and $\mathbf{s}_{k'}, k \neq k'$, are conditionally independent; thus, they can be updated in parallel. Note that the coefficient ϑ depends on some technical parameters such as the speed of the satellite, and the number of emitted signals per second. Its value affects the algorithm behavior where a high value of ϑ leads to oversmoothing and a low value provides noisy signals. Therefore, it should be empirically adapted to the considered data. As a rule of thumb, ϑ should be increased for systems with higher frequency of signals or lower satellite speed.¹

C. Prior for the Noise Variance

The choice of a prior distribution is generally driven by two points: the available knowledge about the parameter of interest and the tractability of the resulting posterior distribution. Regarding the first point, and due to the speckle origins of the corrupting noise, we expect the noise variances $\sigma_k^2, k \in \{1, \dots, K\}$, to vary smoothly. The second point is often fulfilled by considering a conjugate distribution for the parameter of interest. In our case, σ has a conjugate inverse gamma distribution, which is the commonly used variance prior in many Bayesian algorithms [39], [40]. In this paper, we satisfy these two goals by introducing an auxiliary vector \mathbf{w} (of size $K \times 1$) and assigning a gamma-MRF prior for the couple (σ, \mathbf{w}) given

¹The coefficient ϑ can be included in the Bayesian inference, but this should be done by considering a different optimization strategy such as a Markov chain Monte Carlo approach (which is likely to be time consuming).

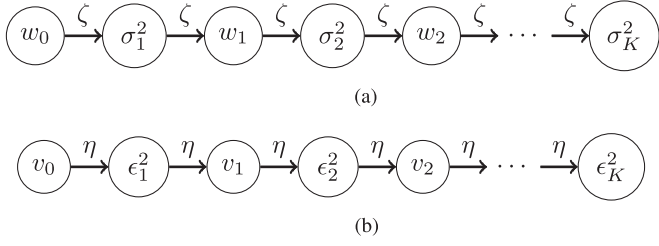


Fig. 3. Proposed first-order gamma-MRF neighborhood structures for (a) the noise variances σ and (b) the signal energies ϵ .

by (see [30] for more details regarding this prior)

$$f(\boldsymbol{\sigma}, \mathbf{w}|\zeta) = \frac{1}{Z(\zeta)} \left(\prod_{k=1}^K \sigma_k^{-2(2\zeta+1)} \right) \times \left(\prod_{k'=1}^K w_{k'}^{(2\zeta-1)} \right) \exp\left(-\frac{\zeta w_0}{\sigma_1^2}\right) \times \prod_{k=1}^{K-1} \exp\left[-\zeta w_k \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_{k+1}^2}\right)\right] \quad (7)$$

where $Z(\zeta)$ is a normalizing constant and $\zeta > 1$ is a fixed coupling parameter that controls the amount of correlation enforced by the gamma-MRF. This prior ensures that each σ_k^2 is connected to two neighboring elements of \mathbf{w} and vice versa [see Fig. 3(a)]. An interesting property of this joint prior is that the conditional prior distributions of $\boldsymbol{\sigma}$ and \mathbf{w} reduce to conjugate inverse gamma (\mathcal{IG}) and gamma (\mathcal{G}) distributions, respectively [30]:

$$\begin{aligned} \sigma_k^2 | w_{k-1}, w_k, \zeta &\sim \mathcal{IG}[2\zeta, \zeta(w_{k-1} + w_k)] \\ \sigma_K^2 | w_{K-1}, \zeta &\sim \mathcal{IG}(\zeta, \zeta w_{K-1}) \\ w_k^2 | \sigma_k^2, \sigma_{k+1}^2, \zeta &\sim \mathcal{G}\left[2\zeta, \frac{1}{\zeta\left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_{k+1}^2}\right)}\right] \end{aligned} \quad (8)$$

where $k \in \{1, \dots, K-1\}$. Equation (8) clearly shows that the variances σ_k^2 and $\sigma_{k'}^2$, for $k \neq k'$, are conditionally independent and that the correlation is introduced via the auxiliary variables \mathbf{w} . It also shows that the maximum of the conditional distributions of $\boldsymbol{\sigma}$ and \mathbf{w} can be analytically evaluated (as detailed in Section IV). These interesting properties will no longer hold if one marginalizes over the latent variables \mathbf{w} , which is not recommended. The coefficients associated with this prior are ζ and w_0 . The condition $\zeta > 1$ ensures the existence of a mode for the gamma distribution (necessary for the considered algorithm). A high value of ζ enforces more correlation between the variances σ_k^2 . In this paper, we used $\zeta = 1000$, and the value of w_0 has been empirically adjusted using the formula $w_0 = \max\{10^{(-2)}, \sqrt{\sum_{m=1}^M [y_{1m} - (\sum_{m'=1}^M \frac{y_{1m'}}{M})]^2}\}$.

D. Hyperparameter Priors

As previously explained, the hyperparameters ϵ_k^2 are closely related to the signal energies via the norm $(\mathbf{s}_k^T \mathbf{H}^{-1} \mathbf{s}_k)$.

Considering this property and the continuity of the signal suggest the presence of a correlation between the parameters ϵ_k^2 . This correlation can be introduced by considering a gamma-MRF prior for $(\boldsymbol{\epsilon}, \mathbf{v})$ as follows [30]:

$$f(\boldsymbol{\epsilon}, \mathbf{v}|\eta) = \frac{1}{Z(\eta)} \prod_{k=1}^K \epsilon_k^{-2(2\eta+1)} \times \left(\prod_{k'=1}^K v_{k'}^{(2\eta-1)} \right) \exp\left(-\frac{\eta v_0}{\epsilon_1^2}\right) \times \prod_{k=1}^{K-1} \exp\left[-\eta v_k \left(\frac{1}{\epsilon_k^2} + \frac{1}{\epsilon_{k+1}^2}\right)\right] \quad (9)$$

where \mathbf{v} are auxiliary variables and $\eta > 1$ is the coupling parameter. A schematic description of the variable correlations is shown in Fig. 3(b), which is similar to that presented in Section III-C. The conjugate conditional prior distributions for $\boldsymbol{\epsilon}$ and \mathbf{v} are given by

$$\begin{aligned} \epsilon_k^2 | v_{k-1}, v_k, \eta &\sim \mathcal{IG}[2\eta, \eta(v_{k-1} + v_k)] \\ \epsilon_K^2 | v_{K-1}, \eta &\sim \mathcal{IG}(\eta, \eta v_{K-1}) \\ v_k^2 | \epsilon_k^2, \epsilon_{k+1}^2, \eta &\sim \mathcal{G}\left[2\eta, \frac{1}{\eta\left(\frac{1}{\epsilon_k^2} + \frac{1}{\epsilon_{k+1}^2}\right)}\right] \end{aligned} \quad (10)$$

where $k \in \{1, \dots, K-1\}$. The parameters (η, v_0) play the same role as (ζ, w_0) and have been fixed to $\eta = 1000$ and $v_0 = w_0$ in the rest of the paper.

E. Posterior Distributions

The parameters of interest associated with the proposed Bayesian model are $\mathbf{X} = (\mathbf{S}, \boldsymbol{\sigma}, \mathbf{w}, \boldsymbol{\epsilon}, \mathbf{v})$. The joint posterior distribution of this Bayesian model can be computed as follows:

$$f(\mathbf{X}|\mathbf{Y}) \propto f(\mathbf{Y}|\mathbf{S}, \boldsymbol{\sigma}) f(\mathbf{S}|\boldsymbol{\epsilon}) f(\boldsymbol{\epsilon}, \mathbf{v}) f(\boldsymbol{\sigma}, \mathbf{w}) \quad (11)$$

where we have assumed *a priori* independence between the parameters. For simplicity, $f(x|\theta)$ has been denoted by $f(x)$ when the parameter θ is a user-fixed parameter. The MMSE and MAP estimators associated with the posterior (11) are not easy to determine. In this paper, and akin to [41], we propose to evaluate the MAP estimator by using an optimization technique maximizing the posterior (11) w.r.t. the parameters of interest. At this point, it should be noted that the proposed algorithm [denoted as smooth signal estimation (SSE)] has many differences from the CD-BM algorithm proposed in [12]. The latter assigns a smooth evolution constraint on the altimetric parameters, while SSE operates on the observed signals by imposing smoothness on the signals and their energies. In addition, the SSE outputs are filtered signals that can be combined with classical altimetric algorithms to achieve classification or other tasks, while CD-BM only provides the estimated altimetric parameters.

IV. COORDINATE DESCENT ALGORITHM

This section describes the optimization algorithm maximizing the posterior (11) w.r.t. the noiseless signals, the noise

Algorithm 1: SSE Algorithm.

-
- 1: **Input**
 - 2: The noisy data \mathbf{Y} , $\zeta > 1$, $\eta > 1$, ϑ
 - 3: **Initialization**
 - 4: Initialize parameters $\mathbf{S}^{(0)}$, $\boldsymbol{\sigma}^{(0)}$, $\mathbf{w}^{(0)}$, $\mathbf{v}^{(0)}$, $\boldsymbol{\epsilon}^{(0)}$ and $t = 1$
 - 5: conv = 0,
 - 6: **Parameter update**
 - 7: **while** conv = 0 **do**
 - 8: Update $\mathbf{s}_k^{(t)}$, for $k \in \{1, \dots, K\}$ according to (33)
 - 9: Update $\boldsymbol{\sigma}^{(t)}$ according to (20)
 - 10: Update $\mathbf{w}^{(t)}$ according to (22)
 - 11: Update $\boldsymbol{\epsilon}^{(t)}$ according to (21)
 - 12: Update $\mathbf{v}^{(t)}$ according to (23)
 - 13: Set conv = 1 if the convergence criteria are satisfied
 - 14: $t = t + 1$
 - 15: **end while**
-

variances, and the hyperparameters. This provides the MAP estimator of the parameters of interest \mathbf{X} . An equivalent problem is to minimize w.r.t. \mathbf{X} , the negative log-posterior $\mathcal{C}(\mathbf{X}) = -\log[f(\mathbf{X}|\mathbf{Y})]$ denoted as ‘‘cost function’’ and given by (after removing unnecessary constants)

$$\mathcal{C}(\mathbf{X}) = \sum_{k=1}^K \left[(\alpha_1 + 1) \log \sigma_k^2 + \frac{\beta_1}{2\sigma_k^2} - (2\zeta - 1) \log w_k + (\alpha_2 + 1) \log \epsilon_k^2 + \frac{\beta_2}{2\epsilon_k^2} - (2\eta - 1) \log v_k \right] \quad (12)$$

where $\alpha_1 = 2\zeta + \frac{M}{2}$, $\alpha_2 = 2\eta + \frac{M}{2}$, $\beta_1 = \|\mathbf{y}_k - \mathbf{s}_k\|^2 + 2\zeta(w_{k-1} + w_k)$, and $\beta_2 = \mathbf{s}_k^T \mathbf{H}^{-1} \mathbf{s}_k + 2\eta(v_{k-1} + v_k)$. Because of the large number of parameters in $\mathbf{X} = (\mathbf{S}, \boldsymbol{\sigma}, \mathbf{w}, \boldsymbol{\epsilon}, \mathbf{v})$, we propose a CDA [37], [38] that sequentially updates the different parameters. More precisely, in each step, the posterior distribution is maximized w.r.t. one parameter, the others being fixed. This process is repeated until the algorithm has converged to a local minimum of the cost function $\mathcal{C}(\mathbf{S}, \boldsymbol{\sigma}, \mathbf{w}, \boldsymbol{\epsilon}, \mathbf{v})$. Thus, the algorithm iteratively updates each parameter by maximizing its conditional distribution, as described in Algorithm 1. The next section describes the suboptimization procedures maximizing the cost function $\mathcal{C}(\mathbf{X})$ w.r.t. the noiseless signal \mathbf{S} , the noise variance $\boldsymbol{\sigma}$, and the hyperparameters $(\mathbf{w}, \boldsymbol{\epsilon}, \mathbf{v})$.

1) *Updating the Parameters:* The noiseless signal \mathbf{S} can be updated by maximizing the conditional distribution associated with each independent \mathbf{s}_k , which is a Gaussian distribution given by

$$\mathbf{s}_k | \mathbf{y}_k, \sigma_k^2, \epsilon_k^2 \sim \mathcal{N}(\bar{\mathbf{s}}_k, \boldsymbol{\Gamma}_k) \quad (13)$$

where

$$\bar{\mathbf{s}}_k = \frac{1}{\sigma_k^2} \boldsymbol{\Gamma}_k \mathbf{y}_k \quad (14)$$

$$\boldsymbol{\Gamma}_k = \left(\frac{\mathbf{H}^{-1}}{\epsilon_k^2} + \frac{\mathbb{I}_M}{\sigma_k^2} \right)^{-1}. \quad (15)$$

Therefore, the noiseless signal \mathbf{S} can be updated using (14), which is the maximum of the Gaussian distribution (an overline is used to denote the mode associated with the updated parameters). Note that this solution corresponds to a least squares solution of the quadratic problem w.r.t. \mathbf{s}_k shown in (12). Note also that the matrix inversion in (14) should be computed at each descent step leading to a high computational cost. Thus, the proposed algorithm considers a useful modification to achieve this computation with less operations, as discussed in the Appendix. The conditional distributions of $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ (respectively, \mathbf{w} and \mathbf{v}) are inverse gamma distributions (respectively, gamma distributions) as follows:

$$\sigma_k^2 | \mathbf{y}_k, \mathbf{s}_k, \mathbf{w}_k \sim \mathcal{IG} \left(\alpha_1, \frac{\beta_1}{2} \right) \quad (16)$$

$$\epsilon_k^2 | \mathbf{y}_k, \mathbf{s}_k, \mathbf{v}_k \sim \mathcal{IG} \left(\alpha_2, \frac{\beta_2}{2} \right) \quad (17)$$

$$w_k^2 | \sigma_k^2, \sigma_{k+1}^2, \zeta \sim \mathcal{G} \left(2\zeta, \frac{1}{\zeta \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_{k+1}^2} \right)} \right) \quad (18)$$

$$v_k^2 | \epsilon_k^2, \epsilon_{k+1}^2, \eta \sim \mathcal{G} \left(2\eta, \frac{1}{\eta \left(\frac{1}{\epsilon_k^2} + \frac{1}{\epsilon_{k+1}^2} \right)} \right). \quad (19)$$

The mode of each distribution is uniquely attained and given by

$$\bar{\sigma}_k^2 = \frac{\beta_1}{2\alpha_1 + 2} \quad (20)$$

$$\bar{\epsilon}_k^2 = \frac{\beta_2}{2\alpha_2 + 2} \quad (21)$$

$$\bar{w}_k^2 = \frac{2\zeta - 1}{\zeta \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_{k+1}^2} \right)} \quad (22)$$

$$\bar{v}_k^2 = \frac{2\eta - 1}{\eta \left(\frac{1}{\epsilon_k^2} + \frac{1}{\epsilon_{k+1}^2} \right)}. \quad (23)$$

These modes are used to update the parameters $\boldsymbol{\sigma}$, $\boldsymbol{\epsilon}$, \mathbf{w} , and \mathbf{v} , as shown in Algorithm 1.

2) *Convergence and Stopping Criteria:* The limit points of the sequence generated by the CDA are stationary point of (12) provided that the minimum of that function w.r.t. \mathbf{X} along each coordinate is unique and that the function \mathcal{C} is monotonically nonincreasing along each coordinate in the interval from \mathbf{x}_i^t to \mathbf{x}_i^{t+1} (see [37, Proposition 2.7.1]). This is easily checked for all the parameters, since they have unimode conditional distributions (Gaussian, gamma, and inverse-gamma distributions), the Gaussian and gamma distributions are log-concave, and the inverse-gamma distribution is decreasing on each side of its maximum. The cost function is not convex; thus, the solution obtained might depend on the initial values that should be chosen carefully. In this paper, the parameters have been initialized as follows: $\boldsymbol{\sigma}^{(0)} = \mathbf{s}_m^{(0)} = \frac{1}{M} \sum_{n=1}^M \mathbf{y}_n$, $\forall m$, $\epsilon_k^2 = 10$, $\forall k$, and $w_k^{(0)} = v_k^{(0)} = 10^{-12}$, $\forall k$. Note that more elaborate initialization procedures can be investigated, but these proposed values

have provided relevant results in the considered simulations (see Section V).

Algorithm 1 is an iterative algorithm that requires the definition of stopping criteria. In this paper, we have considered two criteria, and the algorithm is stopped if either of them is satisfied. The first criterion compares the new value of the cost function to the previous one and stops the algorithm if the relative error between these two values is smaller than a given threshold, i.e.,

$$|\mathcal{C}(\mathbf{X}^{t+1}) - \mathcal{C}(\mathbf{X}^t)| \leq \xi \mathcal{C}(\mathbf{X}^t) \quad (24)$$

where $|\cdot|$ denotes the absolute value and ξ is the threshold that has been fixed to $\xi = 0.001$. The second criterion is based on a maximum number of iterations $T_{\max} = 100$. The next sections study the behavior of the proposed algorithm when considering synthetic and real signals.

V. EXPERIMENTS

This section evaluates the performance of the proposed algorithm with synthetic and real data. It is divided into six parts whose objectives are: a) description of the parametric model used to generate the synthetic altimetric signals, b) introduction of the criteria used for the evaluation of the algorithm quality, c) choice of the number of signals to be processed jointly, d) and e) evaluation on synthetic data of the proposed algorithm when combined with parameter estimation or classification strategies, and f) validation on real Jason-2 altimetric data.

A. Conventional Altimetric Model

The altimetric model, in its simplified version, accounts for three parameters that are the amplitude P_u , the epoch τ , and the significant wave height μ . The resulting mathematical nonlinear model for the altimetric signal is known as the ‘‘Brown model’’ and is given by [2], [7]

$$s(t) = \frac{P_u}{2} \left[1 + \operatorname{erf} \left(\frac{t - \tau_s - \gamma \sigma_c^2}{\sqrt{2} \sigma_c} \right) \right] \exp \left[-\gamma \left(t - \tau_s - \frac{\gamma \sigma_c^2}{2} \right) \right] \quad (25)$$

where

$$\sigma_c^2 = \left(\frac{\mu}{2c} \right)^2 + \sigma_p^2 \quad (26)$$

and where $\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-z^2} dz$ stands for the Gaussian error function, t is the time, $\tau_s = \frac{2\tau}{c}$ (respectively, τ) is the epoch expressed in seconds (respectively, meters), c is the speed of light, and γ and σ_p^2 are two known parameters (depending on the satellite and on the measurement instrument). The nonlinear model described in (28) is commonly used in the altimetric community mainly because of its simplicity [2], [7]. Note that the discrete altimetric signal is gathered in the vector $\mathbf{s} = (s_1, \dots, s_K)^T$, where $s_k = s(kT)$, T is the time resolution, and $\Theta_m = [\theta_1(m), \theta_2(m), \theta_3(m)] = [\mu(m), \tau(m), P_u(m)]$ is a (1×3) vector containing the three altimetric parameters μ, τ, P_u for the m th signal.

The altimetric signals are corrupted by speckle noise that, thanks to the averaging that takes place on-board of the satellite, can be approximated by additive Gaussian noise, as shown

in [42]–[44]. Thus, the altimetric model satisfies (2). The noise variances $\sigma_k^2, k \in \{1, \dots, K\}$, obtained after averaging on-board the satellite, are correlated due to the nature of the speckle noise (this correlation will be considered in the proposed Bayesian scheme). Note that this paper only considers oceanic observations, which generally show a smooth variation between successive signals.

B. Evaluation Criteria

For synthetic signals, the quality of the proposed algorithm can be evaluated by comparing the noiseless signals \mathbf{s}_m to the denoised signals $\hat{\mathbf{s}}_m$ using the reconstruction signal-to-noise ratio (RSNR) given by [45]

$$\text{RSNR} = 10 \log_{10} \left(\frac{\sum_{m=1}^M \|\mathbf{s}_m\|^2}{\sum_{m=1}^M \|\mathbf{s}_m - \hat{\mathbf{s}}_m\|^2} \right). \quad (27)$$

Note that a high RSNR corresponds to a good denoising result. The benefit of the proposed algorithm can also be measured by using the filtered signals as an input to classical altimetric algorithms, as highlighted in Fig. 2. In this paper, we evaluate the enhancement obtained on the estimated parameters and the signal classification results. The latter is evaluated using the accuracy criterion given by

$$\text{accuracy} = \frac{100}{M} \sum_{m=1}^M \delta(z_m - \hat{z}_m) \quad (28)$$

where z_m (respectively, \hat{z}_m) denotes the actual (respectively, estimated) label of the m th signal and δ denotes the Dirac delta function. The effect on the altimetric parameters is evaluated when considering the well-known gradient descent (GD)-based strategy that is commonly used by the altimetric community [7], [11]. For synthetic data, the true altimetric parameters are known, and the estimation quality is evaluated using the root-mean-square error (RMSE) and the standard deviations (STDs) of the estimator $\hat{\theta}_i$ as follows:

$$\text{RMSE}(\hat{\theta}_i) = \sqrt{\frac{1}{N} \sum_{n=1}^N [\hat{\theta}_i(n) - \theta_i(n)]^2} \quad (29)$$

$$\text{STD}(\hat{\theta}_i) = \sqrt{\frac{1}{N} \sum_{n=1}^N \left[\hat{\theta}_i(n) - \left(\frac{1}{N'} \sum_{n'=1}^{N'} \hat{\theta}_i(n') \right) \right]^2} \quad (30)$$

for $i \in \{1, \dots, 3\}$, where $\theta_i(n)$ (respectively, $\hat{\theta}_i(n)$) is the true (respectively, estimated) parameter for the n th signal and N is the number of simulated signals ($N' = N$ for synthetic signals).

When considering real signals, the performance of the proposed algorithm is qualitatively evaluated by a visual comparison between the noisy signals/parameters and the denoised ones [7], [10], [11]. Quantitatively, a modified parameter STD is computed using (30), in which the averaged parameter value is approached by the mean of the estimated parameters along each $N' = 20$ successive signals. This modified STD is called ‘‘STD at 20 Hz’’ [33], [34], [46], [47].

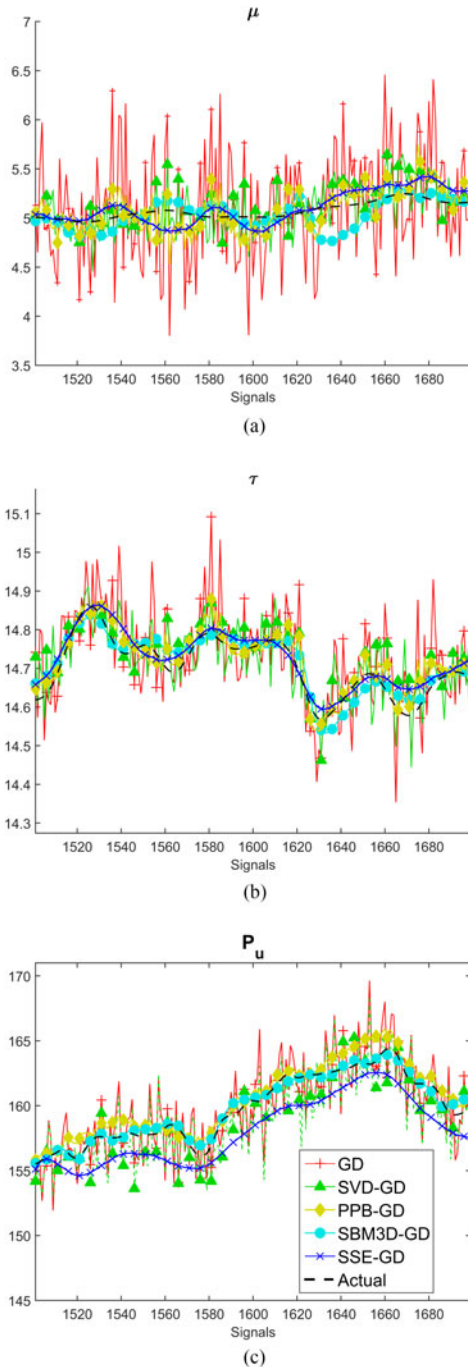


Fig. 4. Example of 1200 ground-truth synthetic parameters (black lines) and their estimations using GD (in red), SVD + GD with $M = 500$ (in green), PPB + GD with $M = 500$ (in yellow), SBM3D + GD with $M = 500$ (in cyan), and the proposed SSE + GD algorithm with $M = 500$ (in blue). (a) μ , (b) τ , and (c) P_u .

C. Choice of the Number of Signals to be Processed Jointly

This section studies the behavior of SSE when varying the number of the jointly denoised signals M . The SSE parameters have been empirically fixed to $\vartheta = 30$, $\zeta = \eta = 1000$ in the rest of the paper. Note, however, that they can be changed according to the recommendations provided in Section III. For this study, $N = 5000$ signals are generated according to the

TABLE I
PERFORMANCE OF THE PROPOSED SSE ALGORITHM W.R.T. THE LENGTH M
(5000 SIGNALS)

	Number of signals: M						
	50	100	250	500	1000	2500	5000
RSNR (dB)	31.1	31.4	31.5	31.6	31.7	31.7	31.7
Time per signal (ms)	0.35	0.24	0.25	0.38	1.03	5.00	17.47

The corrupted data present an RSNR = 19.55 dB.

altimetric model (28), with $K = 104$ gates, using a realistic variation of parameters $\Theta_m = [\mu(m), \tau(m), P_u(m)]$. The sequence of parameters is obtained by applying the CD-BM algorithm [12], on N real Jason-2 signals, where we obtain $\mu \in [3.4, 5.4]$ m, $\tau \in [14.3, 15]$ m, and $P_u \in [150, 190]$ unit (see black lines in Fig. 4). The obtained signals are corrupted by a multiplicative speckle noise generated according to a gamma distribution with shape and scale parameters equal to $(L, 1/L)$, $L = 90$ being the number of looks. This leads to an SNR = 19.55 dB. The noisy signals are processed by the proposed algorithm using different set lengths, as shown in Table I. For example, for a length set $M = 250$, the algorithm is run 20 times to process the $N = 5000$ signals. Overall, these results show an ≈ 11 dB improvement in the processed data with an increasing RSNR w.r.t. M . By increasing the length M , higher computational cost [mainly due to the computation of the singular value decomposition (SVD) in (32)] is required, whereas small values of M lead to more iterations. The value $M = 500$ represents a good compromise, and we consider this value for the rest of the paper.

D. Parameter Estimation Results on Synthetic Data

This section compares the proposed SSE algorithm with some state-of-the-art algorithms when considering a parameter estimation task. We consider three filtering algorithms that are the SBM3D² algorithm [22], the speckle noise variant of PPB³ [23], and the SVD filtering strategy that is commonly used by the altimetric community [13], [14] (with a threshold equal to 84%). We also consider the CD-BM algorithm specialized on parameter estimation [12]. Following [8], [46], and [47], the study is carried out using μ (with fixed $\tau = 14.5$ m and $P_u = 130$). For each value of $\mu \in [0.5, 8]$ m, 500 synthetic signals are generated using the Brown model with different noise realizations (with $L = 90$ looks) and processed using the considered algorithms.

Table II reports the obtained RSNR of the considered algorithms for different values of μ . The best performance is obtained with the CD-BM, since it accounts for more physical information through the considered altimetric model. The PPB and SBM3D provide better results than SSE; however, this improvement is achieved at a price of a higher computational times, as will be seen in the next section (see Table IV). Note

²SBM3D has been run with the number of looks $L = \{1, 30, 90\}$ and we only report the one which provided the best results.

³PPB has been run with the alpha-quantile parameter equals to 0.9999, $L = 90$ looks and standard values for the other parameters.

TABLE II
RSNR (IN DB) W.R.T. μ

	μ (m)								
	0.5	1	2	3	4	5	6	7	8
SVD	26.35	26.43	26.30	26.02	26.03	26.07	26.08	25.92	25.86
PPB	34.89	35.00	34.95	35.01	35.04	35.21	35.04	35.06	35.17
SBM3D	37.96	38.13	37.64	37.72	37.31	36.25	36.31	36.68	36.87
CD-BM	47.34	47.26	47.60	47.47	47.35	46.79	47.61	47.39	47.58
SSE	32.24	32.21	32.22	32.13	32.15	32.10	32.22	32.13	32.07

The corrupted data presents an RSNR = 19.55 dB.

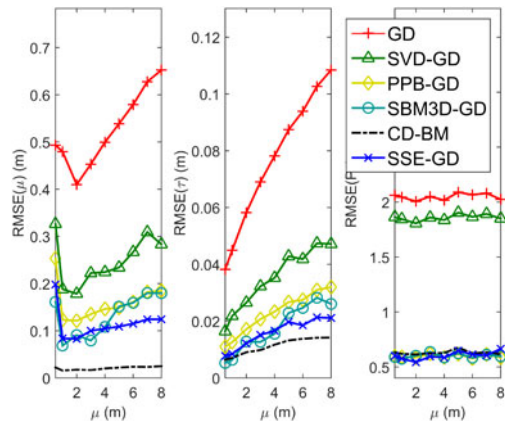


Fig. 5. RMSEs of the altimetric parameters w.r.t. μ for GD, SVD + GD, PPB + GD, SBM3D + GD, CD-BM, and SSE + GD algorithms.

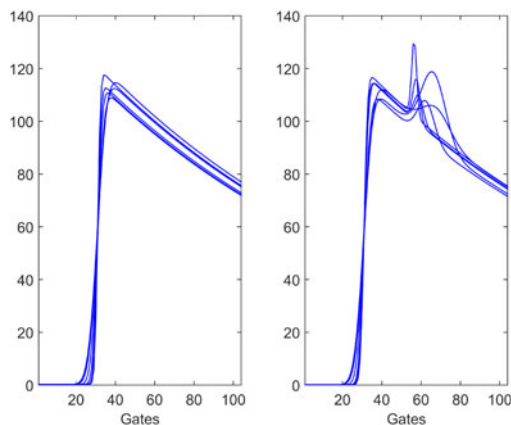


Fig. 6. Example of noiseless (left) Brown echoes denoted by class 1, and (right) Brown plus Gaussian peak denoted by class 13.

also that the algorithms performance are relatively stable w.r.t. the variation of μ .

As previously explained, the results of the filtering algorithms can be used to improve classical tasks such as parameter estimation and signal classification. In this paragraph, we highlight the filtering benefit on the estimated physical altimetric parameters (μ , τ , P_u). The estimated parameters are compared when applying the CD-BM algorithm [12] and the classical GD estimation algorithm [7], [11] on filtered and noisy signals. Fig. 5 shows the obtained parameter RMSEs w.r.t. μ when considering the CD-BM algorithm, the GD algorithm (without filtering), and

TABLE III
CLASSIFICATION RESULTS (IN PERCENTAGE) WITH AND WITHOUT FILTERING

	Class 1	Class 13	Overall accuracy
SVM	90.5	80.0	85.2
SVD + SVM	99.9	83.0	91.4
PPB + SVM	100.0	78.8	89.4
SBM3D + SVM	100.0	82.9	91.5
SSE + SVM	100.0	86.5	93.2

the filtered signals denoted by SVD + GD, PPB + GD, SBM3D + GD, and SSE + GD. As expected, the CD-BM provides the best performance, since it is designed to achieve the parameter estimation task. Apart from CD-BM, the proposed strategy presents the best performance with the lowest RMSEs. Indeed, at a typical μ of 2 m, the proposed SSE-GD improves the RMSEs of the GD algorithm by a factor of 4, 6, and 3 w.r.t. the parameters μ , τ , and P_u , respectively. This parameter improvement is also highlighted in Fig. 4, which shows the estimated parameters when considering the settings of the previous section (for clarity purpose, we only show 200 signal parameters). It is clear from this figure that SSE-GD (blue lines) provides smoother results that better approximate the actual parameters (black lines) than the other algorithms. These results highlight the interest of the proposed strategy in denoising the altimetric signals and improving the estimated altimetric parameters.

E. Classification Results on Synthetic Data

This section highlights the benefit of the SSE in improving the classification of altimetric data. As previously, SSE will be compared to SVD, PPB, and SBM3D algorithms. The altimetric classification algorithms are generally not available. Therefore, we have considered a simple support vector machine (SVM) strategy⁴ to classify two kind of waveforms after the denoising step (see Fig. 2). The first class includes Brown waveforms (i.e., class 1 from the Pistach project [3]), while the second includes the sum of the Brown signal and a Gaussian peak in the trailing edge of the waveform (i.e., class 13 from the Pistach project), where examples are shown in Fig. 6. The SVM has been trained using 2000 noiseless echoes from each class with random parameters in the intervals $\mu \in [2, 4]$ m, $\tau = 31$ gates, $P_u \in [50, 200]$, $a_g \in [10, 200]$, $p_g = 60$ gates, $\sigma_g \in [4, 6]$ gates, where (a_g, p_g, σ_g) denote the amplitude, position, and square root of the Gaussian peak. The classification was achieved on 4000 signals that have been generated using the parameters $\mu \in [2, 4]$ m, $\tau = 31$, $P_u \in [50, 200]$, $a_g \in [10, 200]$, $p_g = 60$, $\sigma_g \in [4, 6]$ and corrupted by a multiplicative noise with $L = 90$ looks, as in the previous sections. Table III reports the obtained classification results with and without the application of the studied filtering algorithms. It can be seen that the filtering algorithms improve the classification results, and that SSE performs better than the other strategies.

⁴The SVM-based MATLAB function “svmclassify” was used to achieve the classification.

TABLE IV
PERFORMANCE ON REAL JASON-2 DATA (45 000 SIGNALS)

		μ (cm)	τ (cm)	P_u
Mean	GD	242	14.68	167.73
	PPB + GD	242	14.65	168.27
	SBM3D + GD	238	14.64	168.87
	SVD + GD	241	14.67	166.62
	CD - BM	270	14.68	167.30
	SSE + GD	248	14.68	164.83
STD	GD	59.9	12.01	6.18
	PPB + GD	29.05	8.21	3.63
	SBM3D + GD	19.7	5.98	3.26
	SVD + GD	18.14	6.02	6.09
	CD - BM	2.01	4.46	1.92
	SSE + GD	9.03	2.94	1.21
Average time per signal (ms)	GD		8.56	
	PPB + GD		32	
	SBM3D + GD		33	
	SVD + GD		9.05	
	CD-BM		6.24	
	SSE + GD		9.63	

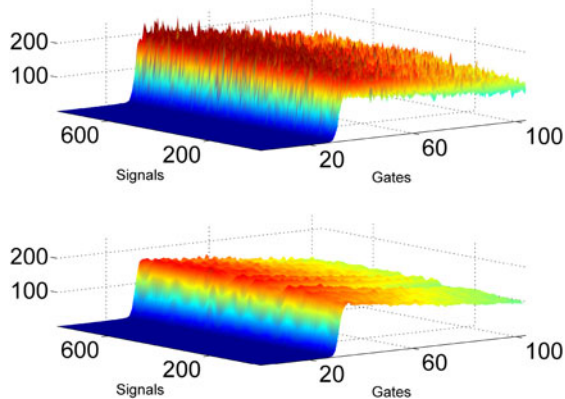


Fig. 7. Example of Jason-2 echoes. (Top) without filtering and (bottom) with SSE filtering.

F. Results on Jason-2 Real Data

This section is devoted to the validation of the proposed SSE denoising algorithm when applied to a parameter estimation task on the oceanic Jason-2 dataset. The data represent a 36-min time series and consist of 45 000 real signals that were extracted from pass 30 of cycle 35. The signals contain $K = 104$ temporal gates, where each gate corresponds to a distance of 46.8 cm. Fig. 7 shows a sequence of 800 Jason-2 signals before and after SSE filtering. Note first that this sequence shows a reduced variation in the altimetric signals, which justifies the use of the proposed strategy. Moreover, it clearly shows a reduction in the noise affecting the signals after the application of the SSE algorithm especially in the tail of the signal (i.e., the decreasing part), which was most affected by the speckle noise. Fig. 8 shows the parameters estimated w.r.t. time when considering the GD (in red), SVD + GD (in green), PPB + GD (in yellow), SBM3D + GD (in cyan), CD-BM (in black), and SSE + GD (in blue) algorithms. As observed for synthetic data in Section V-D, the CD-BM and SSE + GD provide a smooth parameter evolution, which is physically more consistent, while the other

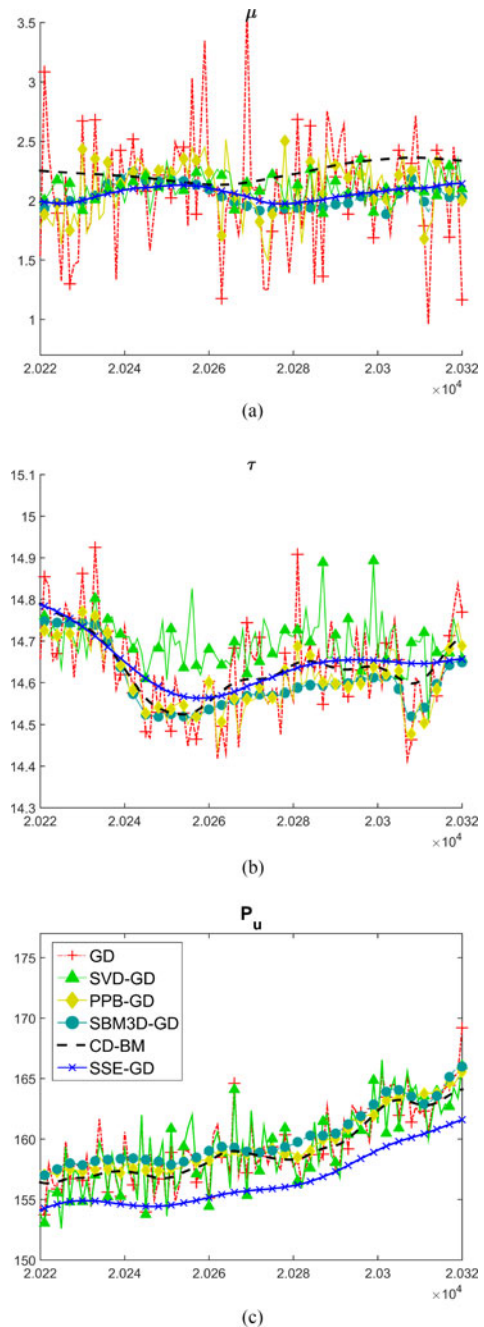


Fig. 8. Estimated parameters using GD (in red), SVD + GD (in green), PPB + GD (in yellow), SBM3D + GD (in cyan), CD-BM (in black), and the proposed SSE + GD algorithm (in blue) for Jason-2 signals. (a) μ , (b) τ , and (c) P_u .

algorithms present high estimation noise. This result is quantitatively confirmed in Table IV, where the smallest STDs are obtained for CD-BM and SSE + GD. This STD reduction has initiated many altimetric missions [5], [6] and is of great importance for many practical applications related to oceanography such as bathymetry. Comparing SSE + GD to GD, Table IV highlights an STD improvement by a factor of 6 for μ , 4 for τ , and 5 for P_u . To better appreciate these STD improvements, it should be noted that new delay/Doppler technology is expected to reduce STDs by a factor lower than 2, as reported in [48]–[50]. This table also shows a good agreement between the means

of the estimated parameters for the different algorithms (except P_u that is slightly reduced by SSE + GD, as shown in Fig. 8). Finally, Table IV also compares the computational costs of the considered algorithms when processing the 45 000 signals (the result is reported for the processing of one signal). As expected, the filtering algorithms require more computational times than GD. However, the proposed SSE algorithm shows a good compromise between the computational cost and the quality of the filtering. These results confirm the good performance of the proposed strategy for the fast denoising of smooth signals such as oceanic altimetric signals.

VI. CONCLUSION

This paper presented a new Bayesian strategy for the estimation of smooth signals corrupted by Gaussian noise. The successive continuous signals can have a numerical expression or be given by a linear/nonlinear function w.r.t. some parameters. A Bayesian model was proposed to take into account the Gaussian properties of the noise and the smooth properties of the signal evolutions. The resulting posterior distribution was maximized using a fast CDA. Experiments on synthetic and real data showed that the proposed algorithm provides good filtering results at a reduced computational cost. An advantage of the proposed algorithm is that it can be combined with the available altimetric algorithms to improve their results. To highlight this, SSE was evaluated by combining it with a commonly used parameter estimation strategy and a classification algorithm. The obtained results showed a clear improvement highlighting the benefit of the proposed algorithm. Future work will include the combination of the proposed algorithm with other classification and target detection altimetric algorithms. It is worth noting that the proposed strategy is fast and generic and thus could be applied when considering other altimetric technologies such as delay/Doppler altimetry [6], [33], [34]. It can also be applied to coastal altimetric signals that have recently been the subject of growing interest [31], [32]. Finally, generalizing the proposed approach for hyperspectral images, that show smooth spectral variations, is also an interesting issue that is currently under investigation.

APPENDIX MATHEMATICAL DERIVATIONS

A. Updating the Noiseless Signal \mathbf{s}_k

The conditional distribution of \mathbf{s}_k in (13) has been computed using the following property:

$$\begin{aligned} x &\sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \times \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \\ &\sim \mathcal{N}[\boldsymbol{\Sigma}(\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2), \boldsymbol{\Sigma}] \end{aligned} \quad (31)$$

with $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}$. The $(M \times M)$ matrix inversion in (15) should be computed at each update of the noiseless signals, which requires a high computational cost. To avoid this cost, we divide this matrix inversion into two parts. Significant computations can be moved outside the “while” loop in Algorithm 1, whereas simple vector multiplications are kept inside the loop. To achieve this, the SVD decomposition is first applied to \mathbf{H}^{-1}

as follows:

$$\mathbf{H}^{-1} = \mathbf{V}\mathbf{D}(r_i)\mathbf{V}^T \quad (32)$$

where $\mathbf{D}(x_i)$ denotes a diagonal matrix with its i th diagonal element equal to x_i , r_i is the i th singular value of \mathbf{H}^{-1} , and \mathbf{V} is a unitary orthogonal matrix, i.e., $\mathbf{V}\mathbf{V}^T = \mathcal{I}_M$. Straightforward computations lead to the following expression for the noiseless signal update:

$$\bar{\mathbf{s}}_k = \mathbf{V}\mathbf{D}\left(\frac{\epsilon_k^2}{r_i\sigma_k^2 + \epsilon_k^2}\right)\mathbf{V}^T\mathbf{y}_k. \quad (33)$$

Note that the operation $\mathbf{V}^T\mathbf{y}_k$ in (33) and the SVD decomposition (32) are only computed once outside the loop, while the remaining vector operations in (33) are achieved inside the loop.

REFERENCES

- [1] B. W. Hapke, “Bidirectional reflectance spectroscopy. I. Theory,” *J. Geophys. Res.*, vol. 86, pp. 3039–3054, 1981.
- [2] G. Brown, “The average impulse response of a rough surface and its applications,” *IEEE Trans. Antennas Propag.*, vol. AP-25, no. 1, pp. 67–74, Jan. 1977.
- [3] F. Mercier *et al.*, “The PISTACH project for coastal altimetry: Status, products and early results,” in *Proc. Ocean Surface Topography Sci. Team Meeting*, Seattle, WA, USA, 1998.
- [4] J.-Y. Tourneret, C. Mailhes, L. Amarouche, and N. Steunou, “Classification of altimetric signals using linear discriminant analysis,” in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, Boston, MA USA, 2008, vol. 3, pp. 75–78.
- [5] P. Vincent *et al.*, “AltiKa: A Ka-band altimetry payload and system for operational altimetry during the GMES period,” *Sensors*, vol. 6, no. 3, pp. 208–234, Mar. 2006.
- [6] R. K. Raney, “The delay/Doppler radar altimeter,” *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1578–1588, Sep. 1998.
- [7] L. Amarouche, P. Thibaut, O. Z. Zanife, J.-P. Dumont, P. Vincent, and N. Steunou, “Improving the Jason-1 ground retracking to better account for attitude effects,” *Marine Geodesy*, vol. 27, nos. 1/2, pp. 171–197, Aug. 2004.
- [8] A. Halimi, C. Mailhes, J.-Y. Tourneret, P. Thibaut, and F. Boy, “Parameter estimation for peaky altimetric waveforms,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1568–1577, Mar. 2013.
- [9] A. Halimi, “From conventional to delay/Doppler altimetry,” Ph.D. dissertation, Inst. Nat. Polytechnique Toulouse, Toulouse, France, 2013. [Online]. Available: <http://thesis.inp-toulouse.fr/archive/00002423/01/halimi.pdf>
- [10] S. Maus, C. M. Green, and J. D. Fairhead, “Improved ocean-geoid resolution from retracked ERS-1 satellite altimeter waveforms,” *Geophys. J. Int.*, vol. 134, no. 1, pp. 243–253, Feb. 1998.
- [11] D. T. Sandwell and W. H. F. Smith, “Retracking ERS-1 altimeter waveforms for optimal gravity field recovery,” *Geophys. J. Int.*, vol. 163, no. 1, pp. 79–89, Oct. 2005.
- [12] A. Halimi, C. Mailhes, J.-Y. Tourneret, and H. Snoussi, “Bayesian estimation of smooth altimetric parameters: Application to conventional and delay/Doppler altimetry,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2207–2219, Apr. 2016.
- [13] A. Ollivier, “Nouvelle approche pour l’extraction de paramètres géophysiques partir des mesures en altimétrie radar,” Ph.D. dissertation, Inst. Nat. Polytechnique Grenoble, Grenoble, France, 2006.
- [14] P. Thibaut, J. Poisson, A. Ollivier, E. Bronner, and N. Picot, “Singular value decomposition applied on altimeter waveforms,” in *Rep. Ocean Surface Topography Sci. Team Meeting*, Seattle, WA, USA, Jun. 2009. [Online]. Available: http://www.aviso.altimetry.fr/fileadmin/documents/OSTST/2009/oral/Thiba_ut_SVD2.pdf
- [15] J. Tournadre, K. Whitmer, and F. Girard-Arduin, “Iceberg detection in open water by altimeter waveform analysis,” *J. Geophys. Res.: Oceans*, vol. 113, no. C8, 2008, Art. no. c08040.
- [16] Y. Kopsinis and S. McLaughlin, “Development of EMD-based denoising methods inspired by wavelet thresholding,” *IEEE Trans. Signal Process.*, vol. 57, no. 4, pp. 1351–1362, Apr. 2009.

- [17] P. Flandrin, P. Goncalvs, and G. Rilling, "Detrending and denoising with empirical mode decompositions," in *Proc. 12th Eur. Signal Process. Conf.*, Sep. 2004, pp. 1581–1584.
- [18] L. Shao, R. Yan, X. Li, and Y. Liu, "From heuristic optimization to dictionary learning: A review and comprehensive comparison of image denoising algorithms," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1001–1013, Jul. 2014.
- [19] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.
- [20] M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
- [21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.
- [22] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.
- [23] C. A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.
- [24] S. Haykin, "Kalman Filtering and Neural Networks," New York, NY, USA: Wiley, 2001.
- [25] M. S. Arulampalam, S. Maskell, and N. Gordon, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [26] A. Doucet and A. M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," Oxford Univ. Press, 2009.
- [27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. London, U.K.: MIT Press, 2006.
- [28] C. Knaus and M. Zwicker, "Progressive image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3114–3125, Jul. 2014.
- [29] H. Talebi and P. Milanfar, "Global image denoising," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 755–768, Feb. 2014.
- [30] O. Dikmen and A. Cemgil, "Gamma Markov random fields for audio source modeling," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 3, pp. 589–601, Mar. 2010.
- [31] C. Gommenginger *et al.*, "Retracking altimeter waveforms near the coasts," in *Coastal Altimetry*, S. Vignudelli, A. G. Kostianoy, P. Cipollini, and J. Benveniste, Eds. Berlin, Germany: Springer, 2011, pp. 61–101.
- [32] M. Passaro *et al.*, "Cross-calibrating ALES Envisat and CryoSat-2 delay-Doppler: A coastal altimetry study in the Indonesian seas," *Adv. Space Res.*, vol. 58, no. 3, pp. 289–303, 2016.
- [33] A. Halimi, C. Mailhes, J.-Y. Tourneret, P. Thibaut, and F. Boy, "A semi-analytical model for delay/Doppler altimetry and its estimation algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4248–4258, Jul. 2014.
- [34] A. Halimi, C. Mailhes, J.-Y. Tourneret, T. Moreau, and F. Boy, "Including antenna mispointing in a semi-analytical model for delay/Doppler altimetry," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 598–608, Feb. 2015.
- [35] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.
- [36] L. Gao, Q. Du, B. Zhang, W. Yang, and Y. Wu, "A comparative study on linear regression-based noise estimation for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 488–498, Apr. 2013.
- [37] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1995.
- [38] J. Sigurdsson, M. Ulfarsson, and J. Sveinsson, "Hyperspectral unmixing with l_q regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6793–6806, Nov. 2014.
- [39] E. Punsakaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 747–758, Mar. 2002.
- [40] N. Dobigeon, J.-Y. Tourneret, and C.-I. Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2684–2695, Jul. 2008.
- [41] A. Halimi, P. Honeine, and J. M. Bioucas-Dias, "Hyperspectral unmixing in presence of endmember variability, nonlinearity or mismodelling effects," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.
- [42] C. Martin-Puig and G. Ruffini, "SAR altimeter retracker performance bound over water surfaces," in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, Cape Town, South Africa, Jul. 12–17, 2009, pp. 449–452.
- [43] O. Germain and G. Ruffini, "A revisit to the GNSS-R code range precision," in *Proc. GNSS-R Conf.*, Noordwijk, The Netherlands, Jun. 14–15, 2006.
- [44] A. Halimi, C. Mailhes, and J.-Y. Tourneret, "Cramér-Rao bounds and estimation algorithms for delay/Doppler and conventional altimetry," in *Proc. 21st Eur. Signal Process. Conf.*, Marrakech, Morocco, Sep. 9–13, 2013, pp. 1–5.
- [45] J. Bioucas-Dias and M. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, Jun. 2010, pp. 1–4.
- [46] K. Giles, D. Wingham, N. Galin, R. Cullen, and W. Smith, "Precise estimates of ocean surface parameters from CryoSat," in *Proc. Ocean Surface Topography Sci. Team Meeting*, Venice Lido, Italy, Sep. 22–29, 2012. [Online]. Available: http://www.aviso.oceanobs.com/fileadmin/documents/OSTST/2012/oral/02_fr_iday_28/05_instr_processing_Ilb/02_IP2B_Giles.pdf
- [47] C. Gommenginger, C. Martin-Puig, S. Dinardo, D. Cotton, M. Srokosz, and J. Benveniste, "Improved altimetric accuracy of SAR altimeters over the ocean: Observational evidence from CryoSat-2 SAR and Jason-2," in *Proc. Ocean Surface Topography Sci. Team Meeting*, San Diego, CA, USA, Oct. 19–21, 2011. [Online]. Available: http://www.aviso.oceanobs.com/fileadmin/documents/OSTST/2011/oral/01_We_dnesday/Splinter%201%20IP/01%20Gommenginger%20SAMOSA OSTST Oct2011.pdf
- [48] K. Xu, P. Liu, L. Shi, L. Wang, and X. Yu, "The altimeter precision comparison between SAR mode and conventional mode through airborne experiment," in *Proc. IEEE Int. Conf. Geosci. Remote Sens.*, Jul. 2015, pp. 3638–3641.
- [49] L. Shi, K. Xu, P. Liu, S. Yang, L. Wang, and X. Yu, "Height precision of SAR altimeter and conventional radar altimeter based on flight experimental data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2676–2686, Jun. 2016.
- [50] L. Fenoglio-Marc *et al.*, "The German Bight: A validation of CryoSat-2 altimeter data in SAR mode," *Adv. Space Res.*, vol. 55, no. 11, pp. 2641–2656, 2015.



Abderrahim Halimi (S'11–M'14) received the Eng. degree in electronics from the National Polytechnic School of Algiers, El Harrach, Algeria, in 2009, and the M.Sc. and Ph.D. degrees in signal processing from the Institut National Polytechnique de Toulouse, Toulouse, France, in 2010 and 2013, respectively.

From October 2013 to September 2015, he was a Postdoctoral Research Associate with the University of Toulouse and the University of Technology of Troyes, France, where his work was supported by the HYPANEMA ANR Project. Since November 2015, he has been a Postdoctoral Research Associate with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, U.K. His research interests focus on statistical signal and image processing, with a particular interest in Bayesian inverse problems with applications to hyperspectral imaging, satellite altimetry, and single-photon depth imaging.



Gerald S. Buller received the B.Sc. (Hons.) degree in natural philosophy from the University of Glasgow, Glasgow, U.K., in 1986, and the Ph.D. degree in physics from Heriot-Watt University, Edinburgh, U.K., in 1989.

He is a Professor of physics at Heriot-Watt University. In 2002, he cofounded Helia Photonics Ltd., Livingston, U.K. His current research interests include single-photon detection methods, particularly at infrared wavelengths, and applications involving the use of single photons. These applications include single-photon depth imaging and novel quantum communications protocols.

Prof. Buller is a Fellow of the Royal Society of Edinburgh, the U.K. Institute of Physics, and the Optical Society of America. He received an EPSRC Established Career Fellowship in Quantum Technology to research sparse photon imaging in 2015.



Stephen McLaughlin (F'11) was born in Clydebank, U.K., in 1960. He received the B.Sc. degree in electronics and electrical engineering from the University of Glasgow, Glasgow, U.K., in 1981, and the Ph.D. degree in electronics and electrical engineering from the University of Edinburgh, Edinburgh, U.K., in 1990.

From 1981 to 1984, he was a Development Engineer in industry. In 1986, he joined the University of Edinburgh, where he studied the performance of linear adaptive algorithms in high noise and nonstationary environments. In 1988, he joined the academic staff at the University of Edinburgh, and from 1991 until 2001, he held a Royal Society University Research Fellowship to study nonlinear signal processing techniques. In 2002, he became a personal Chair in Electronic Communication Systems at the University of Edinburgh. In October 2011, he joined Heriot-Watt University, Edinburgh, as a Professor of signal processing and Head of the School of Engineering and Physical Sciences. His research interests include the fields of adaptive signal processing and nonlinear dynamical systems theory and their applications to biomedical, energy, and communication systems.

Prof McLaughlin is a Fellow of the Royal Academy of Engineering, the Royal Society of Edinburgh, and the Institute of Engineering and Technology.



Paul Honeine (M'07) was born in Beirut, Lebanon, on October 2, 1977. He received the Dipl.-Ing. degree in mechanical engineering and the M.Sc. degree in industrial control from the Faculty of Engineering, Lebanese University, Beirut, in 2002 and 2003, respectively, and the Ph.D. degree in system optimization and security from the University of Technology of Troyes, Troyes, France, in 2007.

From 2007 to 2008, he was a Postdoctoral Research Associate with the Systems Modeling and Dependability Laboratory, University of Technology of Troyes, where he was an Assistant Professor from September 2008 to August 2015, France. Since September 2015, he has been a Full Professor at the Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes, Université de Rouen-Normandie, Rouen, France. Over the past five years, he has authored or coauthored more than 100 peer-reviewed papers. His research interests include nonstationary signal analysis and classification, nonlinear and statistical signal processing, sparse representations, and machine learning. His particular research interests are applications to (wireless) sensor networks, biomedical signal and image processing, hyperspectral imagery, and nonlinear adaptive system identification.

Dr. Honeine is the corecipient (with C. Richard) of the 2009 Best Paper Award at the IEEE Workshop on Machine Learning for Signal Processing.