A Bayesian Mixture-of-Gaussians Model for Astronomical Observations in Interferometry

Lerko Araya-Hernández
Dpt. of Electrical Engineering
Universidad de Chile
Santiago, Chile
email: lerko.araya@ug.uchile.cl

Jorge F. Silva
Dpt. of Electrical Engineering
Universidad de Chile
Santiago, Chile
email: josilva@ing.uchile.cl

Axel Osses
Dpt. of Mathematical Engineering
Universidad de Chile
Santiago, Chile
email: josilva@ing.uchile.cl

Felipe Tobar
Center for Mathematical Modeling
Universidad de Chile
Santiago, Chile
email: ftobar@dim.uchile.cl

***Abstract*—The interferometry problem addresses the estimation of an unknown quantity exploiting the interference among measurements from different sources. These measurements are obtained from the Fourier domain but are sparse and contaminated with noise. We propose a parametric, sum-of-basis, model for these observations together with a Bayesian approach for reconstructing interferometry images. Our main contributions are the construction of a model with a complex-valued noise source, an implementation of an approximate inference method to train the model using Markov chain Monte Carlo and a quantitative comparison against the so-called *dirty* algorithm, where the proposed approach outperformed the considered baseline.**

***Keywords*—Interferometry, Bayesian inference, mixture of Gaussians, signal processing, spectral estimation**

I. Introduction

Interferometry is a class of methods where different signals are combined in order to extract useful information from their sources. Within radio astronomy, interferometry has proven to provide a high-resolution representation of celestial objects by sampling the sky in the frequency (Fourier) domain. In order to do so, interferometers are comprised of an array of antennas, usually identical or coplanar, that simulates an optical lens, where the diameter of this virtual lens is directly related to the largest distance among the antennas in the array. In this way, each pair of antennas yields a *visibility*, i.e., an observation in the frequency space [1], and the set of all visibilities compose the observation in an interferometry setting; the main purpose in this setting is then to convert the observations from the frequency domain to the space domain.

Due to the invertibility of the Fourier transform, if we had complete information of the spectrum (i.e., if the interferometry setting had no missing values or noise) we could determine the spatial representation directly. However, astronomical observations are always subjected to noise and missing data, this is because the region where the spectrum is acquired is given by the topology of the array of antennas and covering dense areas requires an infinite number of them. Therefore, the main challenge arising in real-world applications of interferometry for radio astronomy is to recover the spatial representation of celestial objects given noisy and sparse observations of its spectrum.

A basic approach to recover astronomical images from missing data in interferometry is the *dirty* method [2], which just fills the missing observations with zeros and then applies the discrete Fourier transform. Then, there are two main methods for recovering astronomical images from missing data in interferometry. The first one is CLEAN [2], proposed by Högbom in 1974, which takes a deconvolution standpoint and assumes that the image is a mixture of Gaussian radial basis functions (RBFs). CLEAN finds the locations, magnitude and widths of these RBFs sequentially by fitting a single component to the image and then continuing sequentially with the residual of the previous step. The second method is the maximum entropy method (MEM), proposed by Maisinger et al. [3], which consists on a nonparametric model that is learnt by optimising the entropy of the image.

Although the aforementioned methods are the standard in interferometry for radio astronomy, neither of these methods are able to account for their own uncertainty, a property that arises naturally in the Bayesian approach. In this sense, Lochner et al. [4] proposed the method Bayesian Inference for Radio Observations (BIRO), which models the visibility as a function of atmospheric interference and the interferometer noise. Our proposed model differs from BIRO in that we focus on the image signal directly.

We address the Bayesian recovery of astronomical images in interferometry by proposing a generative model composed of a mixture of Gaussians equiped with a complex-valued observation noise model. We then define the prior distributions for all the elements in the model, discuss the implementation

LAH was supported by Conicyt project PFCHA/MagísterNacional/2017-22171830 and the Department of Electrical Engineering, Universidad de Chile; JFS acknowledges support from Fondecyt 1170854; AO acknowledges support from Fondecyt 1151512; and FT by Conicyt projects PAI-82140061 and Basal-CMM.

requirements of the proposed model and find their posterior distributions in the light of observed data. The model is proved to be superior to the dirty method in terms of the reconstruction mean square error.

The remaining of the paper is organised as follows: Section II presents the basic notions of interferometry, Section III presents the proposed models for the astronomical images and the observation noise, Section IV gives insight into the implementation of the proposed algorithm and Section V presents the simulation results and numerical comparison of the proposed model. Finally, Section VI gives a discussion of our finding and future research directions.

II. Interferometry basics

The supporting concept behind interferometry is the Van Cittert-Zernike Theorem [5], [6]. Denoting by $\mathcal{U}\times\mathcal{V}$ a 2-dimensional frequency space and by $\mathcal{X}\times\mathcal{Y}$ a 2-dimensional location space, the Van Cittert-Zernike Theorem states that if an interferometer has a reception area $A(x,y),\ (x,y)\in X\times Y$, then the visibility $V(u,v),\ (u,v)\in\mathcal{U}\times\mathcal{V}$, of a set of signals with intensity $I(x,y),\ (x,y)\in X\times Y$, is given by:

$$V(u,v)=\int\int_{\mathcal{X}\times\mathcal{Y}}A(x,y)I(x,y)\exp\{-2\pi j(ux+vy)\}\,\mathrm{d}x\,\mathrm{d}y \tag{1}$$

This means that (i) the visibility V and (ii) the intensity I of the astronomical image—in the space domain—masked with the acquisition function A are Fourier pairs. Therefore, a successful reconstruction of the image intensity I requires either a dense acquisition function A (hardware), or an efficient reconstruction algorithm to cater for the points masked by A (estimation theory).

Standard methods to deal with the aforementioned challenge rely on the properties of the Fourier transform and address this problem from a deconvolution point of view. Despite the effectiveness of those approaches, they fail to provide a measure of uncertainty of its estimates. The inherent uncertainty in the observations in radio astronomy is our motivation to focus on the recovery of astronomical observations in interferometry from a Bayesian standpoint.

III. A Bayesian parametric model for observations in interferometry

Our approach relies on a parametric model for the image I that admits closed-form Fourier spectrum, this way, the visibility V is also a parametric model that is learnt from the observations. Then, as the parameters of I and V are linked via the Fourier transform and the Van Cittert-Zernike Theorem, fitting the model to V in the frequency domain provides the parameters of the image I in the space domain and vice versa. We consider a mixture-of-Gaussians for two reasons: First, the square exponential functions are invariant under linear (e.g., Fourier) transformations [7]. Second, they are universal, meaning that they can approximate any continuous functions to a desired degree of precision [8].

A. Generative model

Let us consider an image $I:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ and model it a as a mixture of Gaussian functions of the form

$$I(x,y)=\sum_{i=1}^{N_B}\beta_i\psi_i(x,y)$$
$$\forall i=1,...,N_B\ \ \beta_i\in\mathbb{R};\psi_i:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$$

with $\psi(x,y)$ the basis functions given by

$$\psi_i(x,y)=\frac{1}{\sqrt{2\pi}l}\exp\left(-\frac{1}{2l^2}\left\|\begin{bmatrix}x\\y\end{bmatrix}-\begin{bmatrix}C_x^i\\C_y^i\end{bmatrix}\right\|^2\right), \tag{2}$$

equiped with a lengthscale parameter l and centres of the i^{th} RBF given by C_x^i, C_y^i in the location space.

Owing to Van Cittert-Zernike Theorem, assuming a perfect (constant) acquisition function A, and taking advantage of the linear properties of the Fourier transform, the visibility function V is also a mixture of Gaussians given by eq. (3).

$$\begin{aligned}V(u,v)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_{i=1}^{N_B}\beta_i\psi_i(x,y)\\&\quad\cdot\exp\{-2\pi jux\}\exp\{-2\pi jvy\}\,\mathrm{d}x\,\mathrm{d}y\\&=\sum_{i=1}^{N_B}\beta_i\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_i(x,y)\\&\quad\cdot\exp\{-2\pi jux\}\exp\{-2\pi jvy\}\,\mathrm{d}x\,\mathrm{d}y\\&=\sum_{i=1}^{N_B}\beta_i l\\&\quad\cdot\exp\left\{\frac{-2\pi^2\left(u^2+v^2\right)}{1/l^2}+j2\pi(C_x^iu+C_y^iv)\right\}\end{aligned} \tag{3}$$

Thus, renaming the model parameters according to

$$\alpha_i=\beta_i l \tag{4}$$

$$\phi_i(u,v)=\exp\left\{\frac{-2\pi^2\left(u^2+v^2\right)}{1/l^2}+j2\pi(C_x^iu+C_y^iv)\right\} \tag{5}$$

we have the model for the visibility V by:

$$V(u,v)=\sum_{i=1}^{N_B}\alpha_i\phi_i(u,v).$$

Notice that the proposed model results in a complex-valued RBF model for the visibility. For general hypercomplex-valued kernel models, the reader is referred to [9], [10], [11], [12], [13], [14].

B. Noise model

Recall that the observations are complex-valued, therefore, we represent the noise in a general complex-valued manner, that is, introducing correlations between the real and imaginary channels based on the latent variable representation proposed in [15]. The noise-corrupted observation model is therefore

$$V_{obs}(u,v) = \sum_{i=1}^{N_B} \alpha_i \phi_i(u,v) + \eta(0,C,P) \tag{6}$$

where $\eta(0,C,P)$ is a (possibly noncircular [16]) complex-valued random variable with probability density function

$$f(z) = \frac{1}{\pi|\Sigma|} \exp\left\{ -\frac{1}{2} \begin{bmatrix} z-\mu \\ \bar{z}-\bar{\mu} \end{bmatrix}^T \underbrace{\begin{bmatrix} C & P \\ \overline{P} & \overline{C} \end{bmatrix}^{-1}}_{\Sigma^{-1}} \begin{bmatrix} z-\mu \\ \bar{z}-\bar{\mu} \end{bmatrix} \right\} \tag{7}$$

where $\overline{z}$ denotes the complex conjuate z, and the covariance C and the pseudocovariance P are defined by eq. (8)

$$C = \mathbb{E}[(z-\mu)\overline{(z-\mu)}^T]; \quad P = \mathbb{E}[(z-\mu)(z-\mu)^T] \tag{8}$$

Finally, recall that general complex-valued random variables have non-vanishing pseudocovariances [15].

C. Finding the posterior distribution of the model

We aim to represent the uncertainty in the latent spectrum resulting from the missing observations, therefore, we take a full Bayesian approach and compute the posterior densities of all the parameters of the proposed model. In this sense, let us denote the observation locations by $U = \{u_k, v_k\}_{k=1}^{N_s}$ and the observed visibilities by $V = \{V_k\}_{k=1}^{N_s}$, then the posterior distribution of the proposed model parameters (denoted by $\theta = \{C, P, l, \alpha_i, C_x^i, C_y^i\}$) are given by

$$\begin{aligned} p(\theta|U,V) &= p(\theta) \prod_{k=1}^{N_s} p(V_k, u_k, v_k|\theta) \\ &= p(\theta) \prod_{k=1}^{N_s} \frac{1}{\pi|\Sigma|} \exp\left\{ -\frac{1}{2} e_k^T \Sigma^{-1} e_k \right\} \end{aligned} \tag{9}$$

where Σ is defined by C and P as in eq. (7), and e_k is the error or discrepancy between the proposed model and the k^{th} observation given by

$$e_k = \begin{bmatrix} V_k - V(u_k, v_k) \\ \overline{V_k} - \overline{V(u_k, v_k)} \end{bmatrix}.$$

We have denoted the prior over the parameters by $p(\theta)$ and factorised the likelihood due to the independence of the observations given the model parameters.

Finding the full posterior allows us to compute error bars, modes and higher-order statistics. In particular, we can compute the *maximum-a-posteriori* parameters for point estimates. This is given by

$$\theta^* = \underset{C,P,l,\alpha_i,C_x^i,C_y^i}{\text{argmax}} \quad p(C, P, l, \alpha_i, C_x^i, C_y^i|U,V).$$

These parameters, via the notation introduced in eq. (4), can then be used to compute the parameters of the image model I.

IV. Implementation details

There are a number of practical considerations that need to be addressed to implement our proposed model. These are related to the representation of complex-valued data in the scientific computation packages used, choice of the prior distributions of the model parameters, and the numerical differences between the theoretical Fourier transform and its numerical approximation. We next refer to these aspects.

A. From complex to real valued parameters

In order to avoid numerical stability issues and compatibility with general available functions in Python, we represent the proposed models (both for the image I and the visibility V) in $\mathbb{R}^2$ according to

$$V^{real}(u,v) = \begin{bmatrix} \text{Real}\left(\sum_{i=1}^{N_B} \alpha_i \phi_i(u,v)\right) \\ \text{Imag}\left(\sum_{i=1}^{N_B} \alpha_i \phi_i(u,v)\right) \end{bmatrix}$$

where the noise is now given by

$$\omega(0, \Sigma_\omega) \in \mathbb{R}^2 \tag{10}$$

and the covariance Σ_ω is a function of both the covariance C and pseudocovariance C of the complex-valued noncircular noise process η.

B. Choice of the prior distributions

The computation of the posterior distribution requires a model (likelihood), an approximate inference procedure (in our case Markov chain Monte Carlo) and prior distributions. The prior distributions encode all previous knowledge about the problem at hand and the properties that the practitioner would like to introduce (such as regularisation penalties).

Due to numerical representation of the data to be analysed, we need to ensure positivity of all centres and the weight parameters α. Additionally, the lengthscales have to be positive, they are the *width* of the RBFs. Consequently, we choose the following priors:

$$\alpha \sim \text{Gamma}(\alpha_\alpha, \beta_0) \qquad l \sim \text{Gamma}(\alpha_l, \beta_0)$$
$$C_x^i \sim \text{Gamma}(\alpha_{Cx}, \beta_0) \qquad C_y^i \sim \text{Gamma}(\alpha_{Cy}, \beta_0)$$

In order to ensure positive definiteness of the noise covariance matrix, we use a prior distribution according to the Lewandowski-Kurowicka-Joe method (LKJ) [17]

$$\Sigma_\omega \sim \text{LKJCholesky}(\eta_0, d_0).$$

In this distribution, the parameter η_0 controls the independence of each covariance. Under $\Sigma_\omega \sim \text{LKJCholesky}(\eta_0, d_0)$ samples of the covariance matrix Σ_ω are obtained with a probability proportional to $\det[\Sigma_\omega]^{\eta_0 - 1}$.

For all prior distributions chosen, the parameter β_0 is fixed manually and the remaining hyperparameters[1] have been optimised according to maximum likelihood.

C. Relationship between discrete- and continuous-time Fourier transforms

The observations in interferometry are obtained through the discrete Fourier transform (DFT), whereas the proposed method links the parameters of the image and visibility models according to the continuous-time Fourier transform. The discrepancy between these transformations yields a scaling in the frequencies with respect to one another.

Recall that the 2-dimensional Fourier transform of a function $f(x, y)$ at frequencies (u, v) is given by

$$F_c(u, v) = \int\int f(x, y) \exp\left(-2\pi j(ux + vy)\right) \mathrm{d}x\, \mathrm{d}y \quad (11)$$

and that the discrete Fourier transform by

$$F_d(u, v) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x, y) \exp\left(-2\pi j\left(\frac{ux}{N} + \frac{vy}{M}\right)\right)$$

In this sense, the frequencies u and v have to be scaled according to $\frac{1}{N}$ and $\frac{1}{M}$ respectively, and, as expected, they range between 0 and N and M respectively.

V. SIMULATIONS

We validated the ability of the proposed methodology to recover images from their spectrum with missing data and added (observation) noise. We generated a 50×50 image, shown in Fig. 1, through a mixture of three Gaussians with the following parameters:

- weights: $\alpha = 7.52$
- lengthscale: $l = 3$
- centres: $[(25, 25), (40, 30), (10, 10)]$
- covariances: $\sigma_x^2 = 0.02$, $\sigma_y^2 = 0.02$ and $\sigma_{xy} = 0.0002$

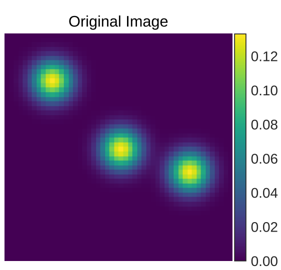
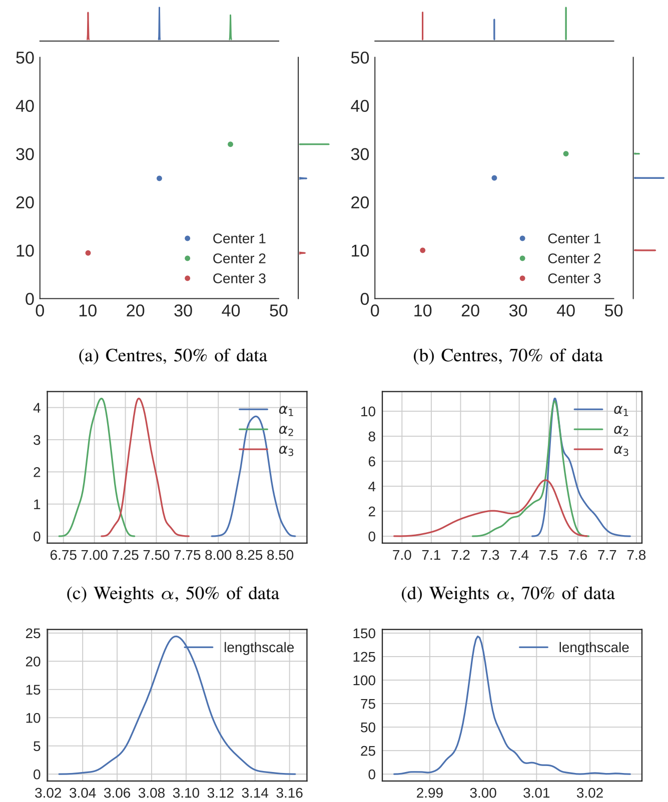


Fig. 1: Synthetic image generated by a mixture of Gaussians.

Then we computed the discrete Fourier transform of this image, added noise and randomly removed some of the data—the remaining datapoints were then regarded as observations from an interferometer. We report results for two variants of this approach using 50% and 70% of the samples to train the models, the aim of this experiment was to show how, in a Bayesian rationale, more data results in narrower posterior distributions.

[1]By hyperparameters we mean parameters of the prior distributions of the parameters of the model.



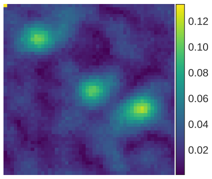
(a) Centres, 50% of data
(b) Centres, 70% of data
(c) Weights α, 50% of data
(d) Weights α, 70% of data
(e) Lengthscale l, 50% of data
(f) Lengthscale l, 70% of data

Fig. 2: Posterior distributions of the model parameters for 50% (left column) and 70% (right column) of data. Using more data results in narrower posterior distributions.

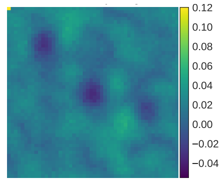
TABLE I: MEAN SQUARE ERROR (MSE) FOR IMAGE RECONSTRUCTION USING THE PROPOSED AND DIRTY METHODS FOR 50% AND 70% OF OBSERVATIONS.

Amount of data	Proposed method MSE	Dirty method MSE
50%	$1.08 \cdot 10^{-1}$	1.16
70%	$1.51 \cdot 10^{-2}$	$9.47 \cdot 10^{-1}$

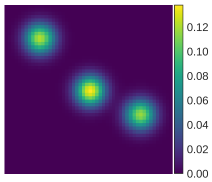
Fig. 2 shows the posterior distributions for both scenarios, where we can see that the model was able to give unbiased estimates, as well as to reduce uncertainty in the estimates as more data are seen. Figs. 3 and 4 show the reconstructions for the 50% and 70% scenarios respectively, where we have compared our proposal to the *dirty* method [2], which consists in setting the missing visibilities to zero and applying the inverse Fourier transform. We have used the same colour scale in all figures to emphasise how the proposed method outperformed the *dirty* one in both scenarios. Furthermore, Table I shows the reconstruction error for both methods, where the proposed approach reported a performance improvement of more than one order of magnitude than the *dirty* approach.



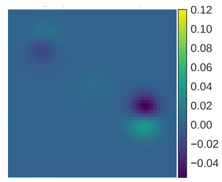
(a) Dirty reconstruction



(b) Dirty method residual

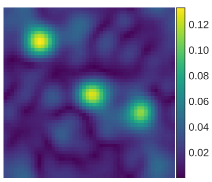


(c) Proposed reconstruction

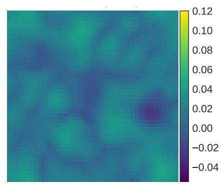


(d) Proposed method residual

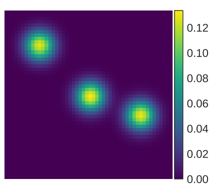
Fig. 3: Image reconstruction using 50% of data. Notice how the proposed method successfully learns the constant regions of the image unlike the dirty method.



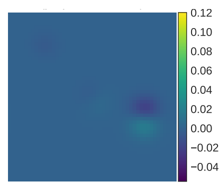
(a) Dirty reconstruction



(b) Dirty method residual



(c) Proposed reconstruction



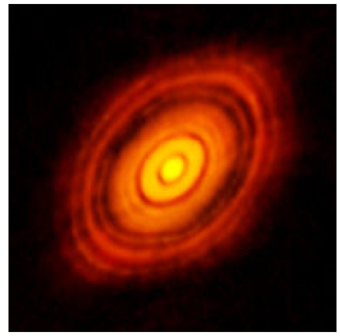
(d) Proposed method residual

Fig. 4: Image reconstruction using 70% of data. The proposed method recovered the rightmost centre much more accurately in the presence of more data (see Fig. 3).

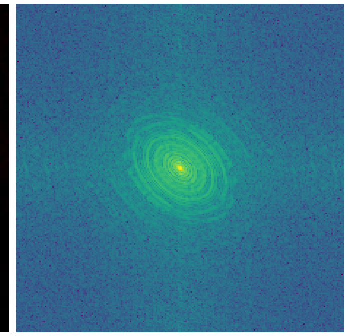
VI. Discussion

We have proposed a Bayesian parametric model for observations in interferometry applications to radio astronomy. The model has desired analytic properties given by its closed form under the Fourier domain, this resulted in having dual models for the space and frequency models with parameters that are linked via the Van Cittert-Zernike Theorem. The model also features a complex-valued observation noise and can be trained using Markov chain Monte Carlo to find the posterior distribution of all the parameters. The proposed methodology was validated against the *dirty* method [2] in terms of the reconstruction MSE for two different levels of missing data. Additionally, we would like to emphasise that another advantage of the proposed approach is its flexibility: the basis functions and prior distributions can be chosen based on expert knowledge.

The next steps to validate the proposed approach is to test it on real-world images of celestial objects such as those shown in Fig. 5 (obtained from ALMA). The main challenge in this setting is the absence of a ground truth and therefore the comparison has to be made based on the recommendation of a specialist (astronomer) and other methods such as CLEAN [2] and MEM [3].



(a) Reconstructed Image



(b) Spectrum of reconstructed image

Fig. 5: A reconstructed image and its spectrum. These real-world data have been obtained from ALMA and therefore there is no ground truth for them.

References

[1] G. Taylor, C. Carilli, and R. Perley, *Synthesis Imagining in Radio Astronomy II*. Astronomical Society of the Pacific, 1999.

[2] J. A. Högbom, “Aperture synthesis with a non-regular distribution of interferometer baselines,” *Astronomy and Astrophysics Supplement*, vol. 15, p. 417, 1974.

[3] K. Maisinger, M. P. Hobson, and A. N. Lasenby, “A maximum entropy method for reconstructing interferometer maps of fluctuations in the cosmic microwave background radiation,” *Monthly Notices of the Royal Astronomical Society*, vol. 290, no. 2, pp. 313–326, 1997.

[4] M. Lochner, I. Natarajan, J. T. L. Zwart, O. Smirnov, B. A. Bassett, N. Oozeer, and M. Kunz, “Bayesian inference for radio observations,” *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1308–1319, 2015.

[5] P. van Cittert, “Die wahrscheinliche schwingungsverteilung in einer von einer lichtquelle direkt oder mittels einer linse beleuchteten ebene,” *Physica*, vol. 1, no. 1-6, pp. 201–210, 1934.

[6] A. R. Thompson, J. M. Moran, and G. W. Swenson, Eds., *Interferometry and Synthesis in Radio Astronomy*. Wiley-Blackwell, 2001.

[7] F. Tobar, T. Bui, and R. Turner, “Learning stationary time series using Gaussian processes with nonparametric kernels,” in *Proc. of Neural Information Processing Systems 28*, 2015, pp. 3483–3491.

[8] I. Steinwart, “On the influence of the kernel on the consistency of support vector machines,” *Journal of Machine Learning Research*, vol. 2, pp. 67–93, 2001.

[9] F. Tobar, A. Kuh, and D. Mandic, “A novel augmented complex valued kernel LMS,” in *Proc. of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2012, pp. 473–476.

[10] F. Tobar and D. Mandic, “Quaternion reproducing kernel Hilbert spaces: Existence and uniqueness conditions,” *IEEE Trans. on Information Theory*, vol. 60, no. 9, pp. 5736–5749, 2014.

[11] R. Boloix-Tortosa, F. J. Payán-Somet, and J. J. Murillo-Fuentes, “Gaussian processes regressors for complex proper signals in digital communications,” in *Proc. of the IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2014, pp. 137–140.

[12] R. Boloix-Tortosa, F. J. Payán-Somet, E. A. de Reyna, and J. J. Murillo-Fuentes, “Complex kernels for proper complex-valued signals: A review,” in *Proc. of the European Signal Processing Conference*, 2015, pp. 2371–2375.

[13] T. K. Paul and T. Ogunfunmi, “Study of the convergence behavior of the complex kernel least mean square algorithm,” *IEEE Trans. on Neural Networks and Learning Systems*, vol. 24, no. 9, pp. 1349–1363, 2013.

[14] ——, “A kernel adaptive algorithm for quaternion-valued inputs,” *IEEE Trans. on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2422–2439, 2015.

[15] F. Tobar and R. E. Turner, “Modelling of complex signals using Gaussian processes,” in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015, pp. 2209–2213.

[16] D. P. Mandic and V. S. L. Goh, *Complex valued nonlinear adaptive filters: Noncircularity, widely linear and neural models*. Wiley, 2009.

[17] D. Lewandowski, D. Kurowicka, and H. Joe, “Generating random correlation matrices based on vines and extended onion method,” *Journal of Multivariate Analysis*, vol. 100, no. 9, pp. 1989 – 2001, 2009.