# MODELLING OF COMPLEX SIGNALS USING GAUSSIAN PROCESSES

*Felipe Tobar and Richard E. Turner*

Department of Engineering
University of Cambridge
{fat25,ret26}@cam.ac.uk

**ABSTRACT**

In complex-valued signal processing, estimation algorithms require complete knowledge (or accurate estimation) of the second order statistics, this makes Gaussian processes (GP) well suited for modelling complex signals, as they are designed in terms of covariance functions. Dealing with bivariate signals using GPs require four covariance matrices, or equivalently, two complex matrices. We propose a GP-based approach for modelling complex signals, whereby the second-order statistics are learnt through maximum likelihood; in particular, the complex GP approach allows for circularity coefficient estimation in a robust manner when the observed signal is corrupted by (circular) white noise. The proposed model is validated using climate signals, for both circular and noncircular cases. The results obtained open new possibilities for collaboration between the complex signal processing and Gaussian processes communities towards an appealing representation and statistical description of bivariate signals.

***Index Terms***— Gaussian process, complex Gaussian process, multi-output GPs, circularity, widely-linear estimation.

## 1. INTRODUCTION

Estimation algorithms for complex-valued signals [1] have become an active research topic within signal processing and have seen application in a number of practical scenarios due to the natural interpretation that the complex-valued representation provides in e.g. wind speed prediction, source separation, and telecommunications applications [2]. The extension of (real-valued) probabilistic estimation approaches to complex signals requires definition of a statistical model for the complex signals, which, under linearity and Gaussianity assumptions, is equivalent to the design of first- and second-order statistics. Unlike real Gaussian random variables (RV), where the only second-order statistic is the covariance, complex-valued ones [3] require both the covariance and the so-called *pseudocovariance*: the covariance between a complex RV and its conjugate. Complex RVs that are uncorrelated with their conjugate therefore have a vanishing pseudocovariance and are referred to as *proper*[1]. The estimation of the pseudocovariance is thus fundamental in complex-valued estimation, as it allows us to calculate the optimal estimator for jointly Gaussian noncircular RVs, that is, the *widely-linear* [5], or *augmented* [6], estimator, which is a linear function of both the regressor (input) and its conjugate.

Existing nonlinear estimators also can exploit the notion of circularity. For instance, neural networks [1, 7] and kernel adaptive filters [8, 9, 10], combine the output of complex-valued neurons and kernels in a widely-linear fashion to yield accurate estimates, yet they fail to provide a probabilistic description of the processes, this is a consistent requirement when assuming a specific statistical setting (second order) for the signals at hand. This issue has been partially solved using Gaussian processes (GP) [11], although only for circular signals. In this sense, a probabilistic model for complex signals suitable for both circular and noncircular cases is still lacking in the open literature.

Gaussian processes [12] approximate a latent function $\mathbf{f}$ by assuming a Gaussian prior density over the space of functions, whereby for any finite collection of inputs $t_1, t_2, \ldots, t_n \in \mathcal{T}$, the corresponding outputs $f_1 = \mathbf{f}(t_1), f_2 = \mathbf{f}(t_2), \ldots, f_n = \mathbf{f}(t_n) \in \mathcal{F}$ are jointly Gaussian, that is,

$$[f_1, f_2, \ldots, f_n]^T \sim \mathcal{N}(\mu, \Sigma) \tag{1}$$

where $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n\times n}$. The mean $\mu$ is usually considered to be zero or set based on *a priori* knowledge of the signal of interest, whereas the covariance matrix $\Sigma$ is parametrised by a positive definite kernel operating on the input space $\mathcal{T}$.

When the codomain (range) of the function $f$ is $\mathcal{F} = \mathbb{R}^m, m > 1$, the vector-valued samples $f_1, \ldots, f_n \in \mathbb{R}^m$ can still be modelled by a Gaussian pdf $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the mean $\boldsymbol{\mu} \in \mathbb{R}^{n\times m}$ is a matrix, and the covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n\times m\times m}$ is a four-dimensional tensor: two dimensions for the input pairs and two dimensions for the output pairs. These methods are referred to as *multi-task* or *multi-output* GPs [13, 14].

For the bivariate-output case ($m = 2$), four covariance functions need to be designed. Alternatively, to avoid using higher dimensional arrays while still accounting for bivariate outputs, we can model the latent function as a **complex** Gaussian process, where finite collections of function values follow a multivariate complex Gaussian density [15]. Since the algebraic topology of the complex field is identical to that of the real filed, except for the conjugate/Hermitian operator, all the properties that make GP methods a standard in Bayesian inference are also inherited in the complex-valued approach; in particular, the *maginalisation* and *conditioning* properties of the GP approach are preserved.

We propose a complex GP predictor the computational complexity of which is equal to that of the two-output GP, since the former operates through a linear transformation of the data. Moreover, our aim is to perform inference on the covariance and the pseudocovariance, as these not only allow for predicting the signal but also for assessing circularity, a desired statistic in complex-valued Signal Processing. We validate the proposed algorithm using real-world noncircular climate signals and a circular process generated by decoupling the climate signal using PCA, where the complex GP captured the second-order statistical information of the signals.

The authors are supported by EPSRC grant number EP/L000776/1.

[1]The more general concept of *circularity* [4] is used for RV with a pdf that is invariant to rotations. As for Gaussian RVs circularity and propriety are equivalent, we use them interchangebly in this paper.

## 2. THE COMPLEX GAUSSIAN DENSITY

Consider the collection of $2n$ jointly Gaussian real-valued RVs

$$F = [f_1, \ldots, f_n, h_1, \ldots, h_n]^T \sim \mathcal{N}(\mu, \Sigma) \tag{2}$$

where $\mu \in \mathbb{R}^{2n}$ and $\Sigma \in \mathbb{R}^{2n \times 2n}$. The RVs $f_i$ and $h_j$ can be seen as the (random) values of two functions at time steps $t_i$ and $t_j$ respectively. Our aim is to find the probability density function of the complex RV defined by the mapping $(f_i, h_i) \mapsto g_i = f_i + jh_i$.

Recall that the density of $F$ is given by (we assume $\mu = \mathbf{0}$)

$$p(F) = \frac{1}{(2\pi)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} F^T \Sigma^{-1} F\right) \tag{3}$$

and consider the linear mapping $\Gamma : \mathbb{R}^{2n} \to \mathbb{C}^{2n}$ defined by

$$F \longmapsto \underbrace{\begin{bmatrix} \mathbf{I}_n & j\mathbf{I}_n \\ \mathbf{I}_n & -j\mathbf{I}_n \end{bmatrix}}_{\Gamma} F = \begin{bmatrix} G \\ G^* \end{bmatrix} = \underline{G} \tag{4}$$

where $g_i = f_i + jh_i$, $G = [g_1, \ldots, g_n]$, $\underline{G}$ is referred to as the *augmented complex vector*, and $\mathbf{I}_n \in \mathbb{R}^{n \times n}$ is the identity matrix.

Observe that the block-wise matrix $\Gamma \in \mathbb{R}^{2n \times 2n}$ is invertible and allows us to express the pdf in (3) as

$$p(F) = \frac{1}{(2\pi)^n \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} F^T \left(\Gamma^H \Gamma^{-H}\right) \Sigma^{-1} \left(\Gamma^{-1} \Gamma\right) F\right) \tag{5}$$

rearranging as a function of $\underline{G} = \Gamma F$, denoting $\underline{\Sigma} = \Gamma \Sigma \Gamma^H$ and using $|\underline{\Sigma}| = 2^{2n}|\Sigma|$, we arrive at the complex Gaussian distribution [15]

$$p(G) = \frac{1}{\pi^n \sqrt{|\underline{\Sigma}|}} \exp\left(-\frac{1}{2} \underline{G}^H \underline{\Sigma}^{-1} \underline{G}\right) \tag{6}$$

where $\underline{\Sigma}$ is the covariance matrix of the augmented complex random vector $\underline{G}$ and is given by

$$\underline{\Sigma} = \Gamma \Sigma \Gamma^H = \begin{bmatrix} C & P \\ P^* & C^* \end{bmatrix} = \mathbb{E}\left[\underline{G}\, \underline{G}^H\right]. \tag{7}$$

The matrix $\underline{\Sigma}$ comprises the covariance of $G$, $C = \mathbb{E}[GG^H]$, and the pseudocovariance of $G$, $P = \mathbb{E}[GG^T]$; when $P = 0$, the RVs $g_i$ are referred to as being *circular*.

## 3. PARAMETRISATION OF THE COVARIANCE AND PSEUDOCOVARIANCE USING KERNELS

Analogously to the real GP definition, we then say that a function $\mathbf{g} : \mathcal{T} \to \mathbb{C}$ is a complex Gaussian process if every finite collection of random variables $g_i = \mathbf{g}(t_i)$ is jointly distributed according to a complex Gaussian distribution and write

$$G = g_1, g_2, \ldots, g_n \sim \mathcal{CN}(\mu, C, P) \tag{8}$$

where the process is defined by the mean $\mu$, covariance $C$ and pseudocovariance $P$. A complex GP is then uniquely defined by its mean, covariance and pseudo-covariance; we now discuss how to parametrise the covariance and pseudo-covariance based on the covariances of the corresponding two-output real GP.

Complex GPs are isomorphic to two-output GPs, with the isomorphism given by (4) or, between the space of covariances, by (7). We next parametrise the covariances of a two-output GP using kernels and then express the covariance and pseudocovariance of a complex GP in terms of the kernels considered.

Consider the complex-valued function $g_t = f_t + jh_t$, where the bivariate process $[f_t, h_t]^T$ is a two-output real GP defined by the covariances (we consider $\mu = \mathbf{0}$)

$$\mathbb{E}[f_s f_t] = K_{rr}(s,t) \qquad \mathbb{E}[h_s h_t] = K_{ii}(s,t) \tag{9}$$

$$\mathbb{E}[h_s f_t] = K_{ir}(s,t) \qquad \mathbb{E}[f_s h_t] = K_{ri}(s,t) \tag{10}$$

where the last two equations and the symmetry of real-valued covariance functions imply[2] $K_{ir}(s,t) = K_{ri}(s,t)$.

Notice that the kernel $K_{ri}$ can be either positive or negative definite[3], as it is not a covariance function. Furthermore, the choice of the kernels in eqs. (9)-(10) has to fulfil the positive definiteness (PD) condition of the covariance—see (2). This can be achieved by modelling the processes $f_t$ and $h_t$ as the output of a latent variable model (LVM) of the form

$$f_t = aX_t + bY_t,\ h_t = cX_t + dZ_t$$

where $X_t, Y_t$ and $Z_t$ are independent unit-variance GPs and $a, b, c, d \in \mathbb{R}$. The processes $Y_t, Z_t$ are referred to as *private* and $X_t$ as *shared*. The LVM formulation provides a parametrisation of the magnitude of $K_{rr}, K_{ii}, K_{ri}$ that fulfils the PD condition of $\Sigma$. Denoting by $|K|$ the magnitude of the kernel $K$, this relationship is given by

$$|K_{rr}| = a^2 + b^2,\ |K_{ii}| = c^2 + d^2,\ |K_{ri}| = ac \tag{11}$$

and allows for choosing the kernels' magnitude in an unconstrained manner, since $a, b, c, d \in \mathbb{R}$.

The covariance and pseudo-covariances of the complex GP $g_t = f_t + jh_t$ can then be expressed in terms of the kernels in eqs. (9)-(10)

$$\begin{aligned} \operatorname{cov}(g_s, g_t) &= \mathbb{E}[g_s \bar{g}_t] = \mathbb{E}\left[(f_s + jh_s)(f_t - jh_t)\right] \\ &= K_{rr}(s,t) + K_{ii}(s,t) + j\left(K_{ir}(s,t) - K_{ri}(s,t)\right) \\ &= K_{rr}(s,t) + K_{ii}(s,t) \end{aligned} \tag{12}$$

$$\begin{aligned} \operatorname{pcov}(g_s, g_t) &= \mathbb{E}[g_s g_t] = \mathbb{E}\left[(f_s + jh_s)(f_t + jh_t)\right] \\ &= K_{rr}(s,t) - K_{ii}(s,t) + j\left(K_{ir}(s,t) + K_{ri}(s,t)\right) \\ &= K_{rr}(s,t) - K_{ii}(s,t) + j2K_{ir}(s,t). \end{aligned} \tag{13}$$

Eqs. (12) and (13) hold the key for parametrising the covariance and pseudocovariance matrices using kernels. The covariance matrix $C$ is given by a positive definite kernel (sum of kernels), whereas the pseudocovariance $P$ is composed of a real part, $\Re P$, which can be positive definite, negative definite or even zero, and an imaginary part, $\Im P$, which corresponds to two times the covariance between the channels multiplied by the imaginary unit $j$. Furthermore, the notion of circularity can be understood from the relationship between $P$ and the real kernels: $P$ is zero only when $K_{rr} = K_{ii}$ and $K_{ri} = 0$, that is, when the (real-valued) processes $\Re g = f$ and $\Im g = h$ have the same covariance and are uncorrelated, this allows for an interpretation, and design, of *circular Gaussian process* in terms real kernels.

## 4. TRAINING AND INFERENCE: EQUIVALENCE BETWEEN $\mathbb{R}^2$ AND $\mathbb{C}$

We now show how to train a complex GP and perform inference via the analogy between the proposed complex GP and its two-output real GP counterpart.

[2] Since $K_{ir}(s,t) = K_{ri}(t,s) = K_{ri}^T(s,t) = K_{ri}(s,t)$.

[3] For instance, $f_t = -h_t$ implies $K_{ri} = -K_{rr}$.

### 4.1. Training

Recall that the hyperparameters of a two-output real GP can be found by maximising the log-likelihood, see (3),

$$L = \log p(F|x,\theta) = -\frac{1}{2}\log|\Sigma| - \frac{1}{2}F^T\Sigma^{-1}F - \frac{n}{2}\log(2\pi).$$

Using the mappings $\Sigma \mapsto \underline{\Sigma} = \Gamma\Sigma\Gamma^H$ and $F \mapsto \underline{G} = \Gamma F$, the minimisation of $L$ can be achieved as a function of the hyperparameters of the complex matrices $P$ and $C$. This allows us to train the complex GP directly in $\mathbb{C}$.

### 4.2. Inference

Recall that if the real RVs $f_{1:n+1}, h_{1:n+1}$ are jointly Gaussian[4], the conditional density $p(f_{n+1}, h_{n+1}|f_{1:n}, h_{1:n})$ is also Gaussian with mean and covariance given by

$$\mu = B_n^T\Sigma_n^{-1}F_{1:n} \tag{14}$$

$$\Sigma = K_{n+1} - B_n^T\Sigma_n^{-1}B_n \tag{15}$$

where $F_{1:n} = [f_1, \ldots, f_n, h_1, \ldots, h_n]^T$, $\Sigma_n$ is the (marginal) covariance of $F_{1:n}$, $B_n = \mathbb{E}[F_{1:n}[f_{n+1}\ h_{n+1}]]$ and $K_{n+1} = \text{cov}([f_{n+1}\ h_{n+1}])$, or in "kernel notation":

$$B_n = \begin{bmatrix} K_{rr}(t_{1:n}, t_{n+1}) & K_{ri}(t_{1:n}, t_{n+1}) \\ K_{ri}(t_{1:n}, t_{n+1}) & K_{ii}(t_{1:n}, t_{n+1}) \end{bmatrix} \in \mathbb{R}^{2n\times 2} \tag{16}$$

$$K_{n+1} = \begin{bmatrix} K_{rr}(t_{n+1}, t_{n+1}) & K_{ri}(t_{n+1}, t_{n+1}) \\ K_{ri}(t_{n+1}, t_{n+1}) & K_{ii}(t_{n+1}, t_{n+1}) \end{bmatrix} \in \mathbb{R}^{2\times 2}. \tag{17}$$

Inference in $\mathbb{C}$ is also possible through the transformations

$$\mu \mapsto \Gamma_1 B_n^T(\Gamma_n^H\Gamma_n^{-H})\Sigma_n^{-1}(\Gamma_n^{-1}\Gamma_n)F_{1:n}$$

$$\Sigma \mapsto \Gamma_1\left(K_{n+1} - B_n^T(\Gamma_n^H\Gamma_n^{-H})\Sigma_n^{-1}(\Gamma_n^{-1}\Gamma_n)B_n\right)\Gamma_1^H$$

where the operator $\Gamma_1 : \mathbb{R}^2 \mapsto \mathbb{C}^2$ (blue) transforms from real to augmented complex representations, and the operator $\Gamma_n : \mathbb{R}^{2n} \mapsto \mathbb{C}^{2n}$ (red), is introduced to express the desired quantities in terms of the covariance and pseudocovariance.

Rearranging gives the augmented statistics

$$\mu \mapsto \underbrace{(\Gamma_n B_n \Gamma_1^H)^H}_{\underline{B}_n}\underbrace{(\Gamma_n\Sigma_n\Gamma_n^H)}_{\underline{\Sigma}_n}{}^{-1}\underline{G}_{1:n} \tag{18}$$

$$\Sigma \mapsto \underbrace{\Gamma_1 K_{n+1}\Gamma_1^H}_{\underline{K}_{n+1}} - \underbrace{(\Gamma_n B_n\Gamma_1^H)^H}_{\underline{B}_n}\underbrace{(\Gamma_n\Sigma_n\Gamma_n^H)}_{\underline{\Sigma}_n}{}^{-1}\underbrace{(\Gamma_n B_n\Gamma_1^H)}_{\underline{B}_n} \tag{19}$$

where:

- $\underline{B}_n$ is the augmented kernel evaluation vector—see (16),
- $\underline{\Sigma}_n$ is the augmented covariance matrix—see (7),
- $\underline{G}_{1:n}$ is the augmented output representation—see (4), and
- $\underline{K}_{n+1}$ is the augmented covariance matrix for the new input $t_{n+1}$—see (17).

Notice that (14)-(15) and (18)-(19) confirm that inference can be performed either in $\mathbb{R}^2$, knowing the covariance matrices, or in $\mathbb{C}$, knowing the covariance and pseudocovariance. This is a consequence of the linear transformation between the real and complex GPs. Moreover, observe that the complex GP estimate is inherently widely-linear: the posterior mean of the complex GP in (18) is a linear transformation of the augmented observations $\underline{G}_{1:n}$. As a consequence, if the off-diagonal blocks of the terms $\underline{B}_n$ and $\underline{\Sigma}_n$ are nonzero (as it is the case for noncircular process) the estimate is linear in both $G_{1:n}$ and $G^*_{1:n}$.

[4]Recall that $f_j = f(t_j)$ and $h_j = h(t_j)$.

## 5. A COMPLEX GAUSSIAN PROCESS MODEL FOR CLIMATE DATA

The proposed Bayesian nonparametric approach for complex signals was validated using historical climate data—usually modelled using complex signals [1, 2, 8]. We considered the maximum monthly temperature and the number of daylight hours per month in Cambridge, UK, between 1959 and 2009[5]. These signals were then preprocessed and combined to produce two complex signals:

- A **noncircular** complex signal, the real and imaginary parts of which were zero-mean, normalised, versions of the aforementioned data.
- A **circular** complex signal composed by the principal components of the climate data, which were also normalised.

Notice that although the real and imaginary parts of the so-constructed signals have zero mean and unit standard deviation, the correlation between channels has been preserved in the noncircular signal. This way, testing for circularity becomes challenging, as it requires identification of correlation between channels rather than just the difference in their variances.

The data were then modelled as samples of a complex Gaussian process. Due to the periodic nature of climate data, we parametrised the covariances in eqs. (9)-(10) using the periodic kernel [16]

$$K_{\text{id}}(s,t) = A_{\text{id}}\exp\left(-\frac{\sin^2(\pi|s-t|/p)}{\sigma^2}\right) + A_n^{\text{id}}\delta_{s,t} \tag{20}$$

for $\text{id} \in \{rr, ii, ri\}$, where $A_n^{\text{id}}$ is the noise variance and the magnitudes $A_{rr}, A_{ii}, A_{ri}$ are parametrised according to (11). The periodic kernel has been readily validated in long-term forecasting of periodic signals [17].

### 5.1. Training and Inference

Two GPs were next trained with the first 60 samples (60 months between Jan. 1959 to Dec. 1963) by setting the scaling of the input to $p = 1/24$ (i.e. one-year periodicity) and the remaining hyperparameters through maximum likelihood (see Table 1). Observe that for the circular signal, $A_{ri} \approx 0$ and $A_{rr} \approx A_{ii}$, whereas for the noncircular case $A_{ri}$ is significantly larger. From (13), these values are in line with the second order statistics of the data, as the pseudocovariance of the GP vanishes for circular training data.

The trained GPs were next implemented for long-term prediction of the normalised climate data. Fig. 1 show the GP predictions and their two-standard-deviation confidence interval (95.4% probability) for the last 24 training samples (crosses) and first 48 validation samples (circles). The GP predictions were more accurate in the noncircular case, where both signals share deterministic periodic components and low noise, whereas for the circular case the second component deviates more form the predicted mean. This was expected, since the second component carries most of the process noise—despite also capturing some of the data periodicity—and is confirmed by the larger noise variances in Table 1.

[5]Data available from `http://www.metoffice.gov.uk/`.

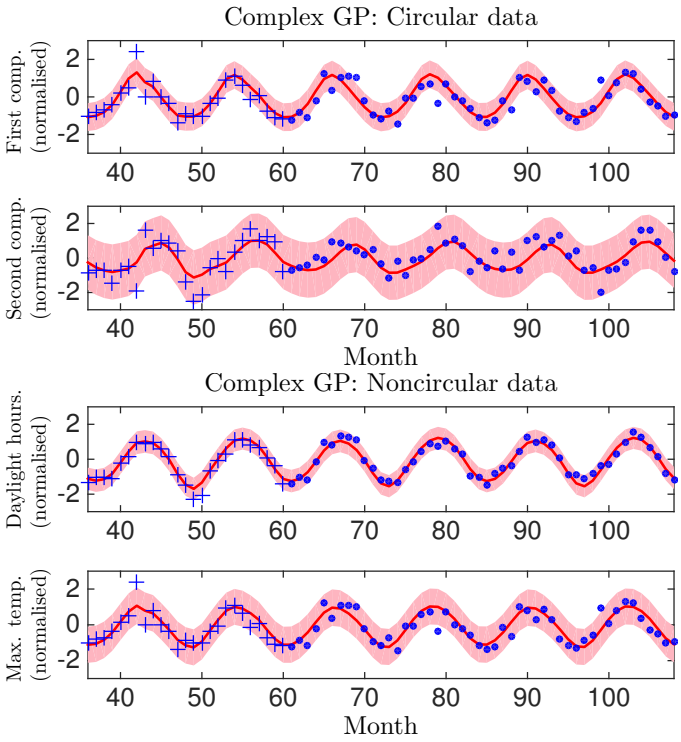


**Fig. 1**: Long term prediction of circular (top plots) and noncircular (bottom plots) data. Training samples (crosses), validation samples (circles), and GP predicted mean (red) and two-standard-deviation interval (light red).

**Table 1**: GP hyperparameters for circular and noncircular data: $A_{rr}, A_{ii}, A_{ri}$ are covariances' magnitudes, $\sigma$ is the kernel width for all kernels, and $A_n^{rr}$, $A_n^{ii}$ the noise variances, see eqs. (12),(13) and (20). The error bars correspond to one standard deviation using a Laplace approximation with uniform priors.

| | Circular | Noncircular |
|---|---|---|
| $A_{rr}$ | $0.76 \pm 0.33$ | $0.86 \pm 0.1145$ |
| $A_{ii}$ | $0.64 \pm 0.35$ | $0.74 \pm 0.1085$ |
| $A_{ri}$ | $0.02 \pm 0.24$ | $0.71 \pm 0.0614$ |
| $\sigma^2$ | $0.14 \pm 0.0045$ | $0.17 \pm 0.0073$ |
| $A_n^{rr}$ | $0.18 \pm 0.0146$ | $0.09 \pm 0.0178$ |
| $A_n^{ii}$ | $0.55 \pm 0.1392$ | $0.23 \pm 0.0458$ |

**Table 2**: Norm of the prediction error (100 validations samples) for the circular and noncircular signals using both the *circular* and *unconstrained* algorithms.

| | Circular signal | Noncircular signal |
|---|---|---|
| Circular algorithm | 4.98 | 3.64 |
| Unconstrained algorithm | 4.95 | 3.46 |

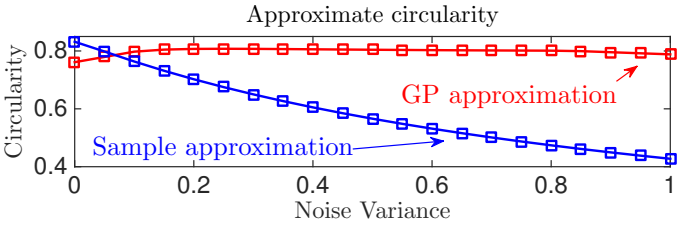


**Fig. 2**: Approximated circularity coefficient using GPs and sample averages for climate data between Jan. 1959 and Dec. 2009. The smaller the circularity coefficient, the more circular the process.

### 5.2. A Circular Predictor

We next trained circular GPs, that is, via maximum likelihood constrained to $A_{rr} - A_{ii} - 2jA_{ri} = 0$ and compared it to the *unconstrained* GP of the previous section. The prediction error is shown in Table 2, where constraining the algorithm to be circular resulted in larger prediction errors, particularly for the noncircular signal where the performance drop was in the order of 5%. This validates the ability of the complex GP approach to exploit second order statistical information encoded in the pseudocovariance.

### 5.3. Learning Circularity Using GPs

Circularity is defined by the coefficient $\rho = \text{cov}(g, g^*)/\text{var}(g)$ [18], where $\rho \approx 0$ (*cf.* $\approx 1$) implies that the process is circular (*cf.* noncircular). For a collection of samples $g_{1:n}$, we considered the following approximations of $\rho$

- Sample approx.: $\rho_S = \sum_{t=1}^{n} g_t g_t / \sum_{t=1}^{n} g_t g_t^*$
- GP approx.: $\rho_{GP} = \left((A_{rr} - A_{ii})^2 + 4A_{ri}^2\right) / (A_{rr} + A_{ii})^2$

where $\rho_S$ is based on averaging the observed samples, and $\rho_{GP}$ through parametrising the covariance and pseudocovariance using kernels—see (12)-(13) and the hyperparameters in Table 1.

To assess the proposed GP approximation of the circularity coefficient, the noncircular complex signal corresponding to the period between Jan. 1959 and Dec. 2009 (600 samples) was contaminated with white circular noise of increasing variance, to then compute the approximations $\rho_S$ and $\rho_{GP}$. Fig. 2 shows the robustness of the GP approximation to increasing levels of observation noise: $\rho_{GP}$ did not decrease with the introduction of circular noise. The empirical estimate $\rho_S$ approximates the circularity of the observed samples without distinguishing between data and noise, whereas the GP models the noise as an additional source of uncertainty—see (20)—and therefore is able to capture the circularity of the data even under the presence of circular noise. We emphasise that the hyperparmeters $A_{rr}, A_{ii}$ and $A_{ir}$, used to compute $\rho_{GP}$, where found by maximum likelihood and were not constrained to the circular/noncircular cases.

## 6. CONCLUSIONS

The Gaussian process approach to complex-valued signals has been validated through learning and forecasting of circular and noncircular complex signals produced using historical climate data. Additionally, the hyperparameters of the proposed model allow for an approximation of the circularity coefficient that is robust to circular noise. The results obtained confirm that GPs are well-suited for capturing statistical information of complex signals, and also position the proposed approach as a basis for further developments of probabilistic inference for complex signals, which arise from the connection between the signal processing and machine learning literature.

The proposed GP model for complex signals allows for designing second order statistics directly in $\mathbb{C}$, by parametrising the covariance and pseudocovariance instead of the real-valued covariances. In the authors' view, however, it remains an open question as to whether the complex representation has analytical or computational advantages over the bivariate one.

## 7. REFERENCES


[1] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. John Wiley & Sons, 2009.

[2] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, 2011.

[3] K. S. Miller, "Complex Gaussian processes," *Siam Review*, vol. 11, no. 4, pp. 544–567, 1969.

[4] B. Picinbono, "On circularity," *IEEE Trans. on Signal Processing*, vol. 42, no. 12, pp. 3473–3482, 1994.

[5] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[6] S. Javidi, M. Pedzisz, S. L. Goh, and D. P. Mandic, "The augmented complex least mean square algorithm with application to adaptive prediction problems," in *Proc. of the Cognitive Information Processing Conference*, 2008, pp. 54–57.

[7] I. Aizenberg, *Complex-valued neural networks with multi-valued neurons*. Springer, 2011, vol. 353.

[8] F. Tobar, A. Kuh, and D. P. Mandic, "A novel augmented complex valued kernel LMS," in *Proc. of the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2012, pp. 481–484.

[9] P. Bouboulis, S. Theodoridis, and M. Mavroforakis, "The augmented complex kernel LMS," *IEEE Trans. on Signal Processing*, vol. 60, no. 9, pp. 4962–4967, 2012.

[10] F. Tobar, "Kernel-based adaptive estimation: Multidimensional and state-space approaches," Ph.D. dissertation, Imperial College London, 2014.

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

[12] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. the MIT Press, 2006.

[13] M. Alvarez and N. D. Lawrence, "Sparse convolved Gaussian processes for multi-output regression," in *Advances in Neural Information Processing Systems 21*, pp. 57–64.

[14] E. V. Bonilla, K. M. Chai, and C. K. I. Williams, "Multi-task Gaussian process prediction," in *Advances in Neural Information Processing Systems 20*, 2008, pp. 153–160.

[15] N. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Annals of mathematical statistics*, pp. 152–177, 1963.

[16] D. J. C. MacKay, "Introduction to Gaussian processes," in *Neural Networks and Machine Learning*, C. M. Bishop, Ed. Springer, 1998, vol. 168 of *NATO ASI Series*, pp. 133–165.

[17] N. Haji Ghassemi and M. Deisenroth, "Analytic long-term forecasting with periodic Gaussian processes," in *Proc. of AISTATS*, 2014, pp. 303–311.

[18] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Processing Letters*, vol. 15, pp. 841–844, 2008.