

A Novel Rayleigh Dynamical Model for Remote Sensing Data Interpretation

Fábio M. Bayer, Débora M. Bayer, Andrea Marinoni, *Senior Member, IEEE*, Paolo Gamba, *Fellow, IEEE*

Abstract—This work introduces the Rayleigh autoregressive moving average (RARMA) model, which is useful to interpret multiple different sets of remotely sensed data, from wind measurements to multitemporal synthetic aperture radar (SAR) sequences. The RARMA model is indeed suitable for continuous, asymmetric, and nonnegative signals observed over time. It describes the mean of Rayleigh-distributed discrete-time signals by a dynamic structure including autoregressive and moving average terms, a set of regressors, and a link function. After presenting the conditional likelihood inference for the model parameters and the detection theory, in this work a Monte Carlo simulation is performed to evaluate the finite signal length performance of the conditional likelihood inferences. Finally, the new model is applied first to sequences of wind speed measurements, and then to a multitemporal SAR image stack for land-use classification purposes. The results in these two test cases illustrate the usefulness of this novel dynamic model for remote sensing data interpretation.

Index Terms—land-use classification, machine learning, Rayleigh distribution, SAR, time series, wind speed.

I. INTRODUCTION

Gaussian distributions are widely employed in remote sensing data analysis [1]–[3] and classical statistical signal processing [4]. Accordingly, in case of image stacks or multitemporal data sets, autoregressive integrated moving average (ARIMA) models [5] are currently used [6]–[8]. However, although the Gaussian assumption provides a reasonable model for signals of interest in many situations [4], [9], it has been recognized that Gaussianity is too restrictive for many applications [10], [11].

A possible solution is to transform the data prior to the modeling step. However, this solution presents some drawbacks, such as the fact that the output values will be interpreted in terms of the mean of the transformed signal and not of the mean of the signal of interest [12]. For this reason, the interest in non-Gaussian time series models has increased [12]–[16].

Manuscript received ... This work was partly supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil, Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA), Norway, and the Research Council of Norway (RCN Grant no. 237906).

F. M. Bayer is with the Department of Statistics and LACESM, Federal University of Santa Maria, Brazil (E-mail: bayer@ufsm.br).

D. M. Bayer is with the Department of Sanitary and Environmental Engineering, Federal University of Santa Maria, Brazil (E-mail: debora.bayer@ufsm.br).

A. Marinoni is with the Department of Physics and Technology, UiT the Arctic University of Norway, NO-9037, Tromsø, Norway (E-mail: andrea.marinoni@uit.no).

P. Gamba is with the Telecommunications and Remote Sensing Laboratory, Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy (E-mail: paolo.gamba@unipv.it).

In many cases of remote sensing data sets the signal of interest is continuous, asymmetric, and nonnegative, and the Rayleigh distribution is a better fit for signal and image modeling [17] than the Gaussian distribution. Rayleigh-distributed signals describe magnetic resonance imaging [18], wind speed [19], radar scattering [20] and speckle distribution [21], and they have been specifically used for synthetic aperture radar (SAR) data [17], [22], [23]. In fact, the Rayleigh distribution is capable to describe all the modalities that use coherent imaging in the amplitude mode, such as SAR, laser, ultrasound, sonar, and others [24].

Despite its importance for these and other several applications, in the current state of the art technical literature there is no mention of a dynamic model capable of modeling a sequence of Rayleigh-distributed discrete-time signals. For this reason, this research proposes a novel Rayleigh-based dynamical model, called Rayleigh autoregressive moving average (RARMA) model, which fills this gap, and allows modeling Rayleigh-distributed time series. In this context, we explore two applications of RARMA model in remotely sensed data sets, namely: wind speed estimation and multitemporal SAR data interpretation. They have been selected due to their relevance.

As for SAR multitemporal data, data sequence interpretation has increased in recent technical literature [38]–[41]. SAR sequences and image stacks have been used in different Earth observation (EO) applications, such as change detection and land-use classification [8], [42]–[44]. In a SAR image observed over time, each pixel can be seen as an univariate time series [8]. As the Rayleigh distribution is suitable for modeling the amplitude values of SAR images, the univariate time series of each pixel can be modeled as a Rayleigh-distributed discrete-time signal. This work provides evidences that, by using the proposed RARMA model and its associated inference and diagnostic tools, it is possible to extract relevant features from SAR multitemporal data. These features can be then used as input to machine learning classifiers such as support vector machines [45], k -nearest neighbor [46], and random forest [47], [48]. Developing methods for extracting information from multitemporal remote sensing data to be used together with machine learning classifiers is a hot topic in EO science [42], [49], [50].

Accordingly, the contribution of this work is two-fold. First, the Rayleigh regression model in [17] is extended, introducing the novel RARMA model. Like other non-Gaussian dynamic models [12], [14], [16], the RARMA model describes the mean of Rayleigh-distributed discrete-time signals by a dynamic structure including autoregressive and moving average

terms, a set of regressors, and a link function. The second contribution is the use of this new model to improve the interpretation of two sets of remotely sensed data, enabling (i) the possibility to provide forecasts, i.e. to generate future values for remotely sensed signals, and (ii) the capability to understand the stochastic behavior of these signals and extract interesting features. To prove these statements, the proposed RARMA model is used in this work to model and forecast wind speed time series. Moreover, it is used to extract multitemporal features from a sequence of SAR images recorded by the European Space Agency (ESA) Sentinel-1 satellite over Beijing, P.R. China. These features are exploited to classify land-use, and specifically urban area extent, in the imaged scene by means of different machine learning methods [49], [50].

The paper is structured as follows. First, in Sections II–V the theoretical concepts of the new statistical model is discussed. Specifically, in Section II, the RARMA model is introduced. Then, Section III presents closed-form expressions for the score vector, large data record results, and other details about the conditional likelihood inference. Section IV discusses some diagnostic measures as well as the forecasting method. In Section V, we present a Monte Carlo simulation study, in order to evaluate the finite signal length inference performance. Finally, Section VI explores and discusses the application of the proposed methods to remotely sensed data. The conclusions of this work can be eventually found in Section VII.

II. THE PROPOSED MODEL

Let's consider a discrete-time signal $\{y[t]\}_{t \in \mathbb{Z}}$, where each $y[t]$ assumes values in \mathbb{R}^+ and $\mathcal{F}[t-1]$ is the previous information set of observations up to time t . Assuming that, conditionally to $\mathcal{F}[t-1]$, each $y[t]$ is distributed according to a Rayleigh distribution with conditional mean $\mu[t]$, the conditional probability density function of $y[t]$, with the mean-based parametrization used in [17], is given by

$$f(y[t]|\mathcal{F}[t-1]) = \frac{\pi y[t]}{2\mu[t]^2} \exp\left(-\frac{\pi y[t]^2}{4\mu[t]^2}\right), \quad (1)$$

where $y[t] > 0$ and $\mu[t] > 0$. The cumulative distribution function and quantile function are, respectively,

$$F(y[t]|\mathcal{F}[t-1]) = 1 - \exp\left(-\frac{\pi y[t]^2}{4\mu[t]^2}\right)$$

and

$$F^{-1}(u|\mathcal{F}[t-1]) = 2\mu\sqrt{\frac{-\log(1-u)}{\pi}}.$$

The conditional mean and variance of $y[t]$ are given by

$$\mathbb{E}(y[t]|\mathcal{F}[t-1]) = \mu[t] \quad (2)$$

and

$$\text{Var}(y[t]|\mathcal{F}[t-1]) = \mu[t]^2 \left(\frac{4}{\pi} - 1\right).$$

The Rayleigh regression model proposed in [17] is given by a linear predictor $\eta[t]$ related to the mean $\mu[t]$ through a

strictly monotonic and twice differentiable link function $g(\cdot)$, where $g: \mathbb{R}^+ \rightarrow \mathbb{R}$, by the following regression structure:

$$\eta[t] = g(\mu[t]) = \mathbf{x}^\top[t]\boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_l)^\top$ is a set of unknown parameters, and $\mathbf{x}[t] = (x_1[t], \dots, x_l[t])^\top$ is a vector of l deterministic independent variables (covariates).

As done in [12], [14] for other classes of models, we extend the Rayleigh regression considering an additional dynamical component, $\boldsymbol{\tau}[t]$, that allows the inclusion of autoregressive (AR) and moving average (MA) terms in the linear predictor. Thus, the structure of the Rayleigh autoregressive moving average model is given by

$$g(\mu[t]) = \eta[t] = \mathbf{x}^\top[t]\boldsymbol{\beta} + \boldsymbol{\tau}[t],$$

where $\boldsymbol{\tau}[t] = \sum_{i=1}^p \phi_i \mathcal{A}(y[t-i], \mathbf{x}[t-i], \boldsymbol{\beta}) + \sum_{j=1}^q \theta_j \mathcal{M}(y[t-j], \mu[t-j])$,

with AR and MA terms represented by functions \mathcal{A} and \mathcal{M} , respectively. The AR coefficients are defined as $\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$, and the MA coefficients as $\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$, where p and q are the orders of the model. There are different options for \mathcal{A} and \mathcal{M} . For the AR term, \mathcal{A} , we consider the structure as in [15] and for the MA term, \mathcal{M} , as in [16]. This way, a more parsimonious structure for RARMA(p, q) model is given by

$$g(\mu[t]) = \eta[t] = \zeta + \mathbf{x}^\top[t]\boldsymbol{\beta} + \sum_{i=1}^p \phi_i g(y[t-i]) + \sum_{j=1}^q \theta_j r[t-j], \quad (3)$$

where $\zeta \in \mathfrak{R}$ is the intercept, and $r[t] = g(y[t]) - g(\mu[t])$ is the MA term on the predictor scale. Additionally, the conditional mean of $y[t]$ is a function of the past observed signals $y[t-i]$, with $i = 1, \dots, p$, covariates $\mathbf{x}[t]$, and moving average error terms. Note that the link function $g(\cdot)$ connects the linear predictor $\eta[t]$ to the mean $\mu[t]$ of data $y[t]$. As $\mu[t] > 0$, a common choice for the link function is the log function, $\eta[t] = \log([\mu[t]])$ [14], [17]. As usual in the context of ARMA modeling [51], we assume that the AR and MA characteristic polynomials do not have common roots and the AR characteristic polynomial does not have unit roots. Invertibility and causality conditions for the ARMA component of the proposed model are not required.

The use of the RARMA model for modeling asymmetric and nonnegative signals avoids the need to transform the signal. Moreover, when the signal is Rayleigh distributed, e.g., like the amplitude values of SAR image pixels, the usual Gaussian-based inference is not appropriate. The RARMA model naturally accommodates asymmetries and the autocorrelation of the signal.

III. CONDITIONAL LIKELIHOOD INFERENCE

The proposed RARMA(p, q) model is given by specifications in (1) and (3). Given an observed discrete-time signal of length T , y_1, \dots, y_T , the problem is to estimate the

$(l + p + q + 1)$ -dimensional vector of unknown parameters, $\boldsymbol{\gamma} = (\zeta, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top)^\top$. To this aim, as in [12], [16], [52] for other dynamical models, we shall consider the conditional maximum likelihood estimation.

The logarithm of the conditional likelihood function for the vector parameter $\boldsymbol{\gamma}$, conditional on $\mathcal{F}[t - 1]$ and null for the first $m = \max(p, q)$ values of t , is

$$\ell(\boldsymbol{\gamma}) = \sum_{t=m+1}^T \log f(y[t] | \mathcal{F}[t - 1]) = \sum_{t=m+1}^T \ell[t](\boldsymbol{\mu}[t]), \quad (4)$$

where

$$\ell[t](\boldsymbol{\mu}[t]) = \log\left(\frac{\pi}{2}\right) + \log(y[t]) - \log(2\mu[t]^2) - \left(\frac{\pi y[t]^2}{4\mu[t]^2}\right).$$

The conditional maximum likelihood estimators (CMLE) are obtained upon maximizing the conditional log-likelihood function. Hence, the CMLE of $\boldsymbol{\gamma}$ can be written as:

$$\hat{\boldsymbol{\gamma}} = \arg \max_{\Theta} \ell(\boldsymbol{\gamma}), \quad (5)$$

where Θ is the parametric space. To this aim, we need to define the score vector, as presented in the next section.

A. Score Vector and Numerical Optimization

By differentiating the conditional log-likelihood function in (4), with respect to each j -th element of the parameter vector $\boldsymbol{\gamma}$, for $j = 1, \dots, (l + p + q + 1)$, the following equations are obtained:

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} = \sum_{t=m+1}^T \frac{d\ell[t](\boldsymbol{\mu}[t])}{d\mu[t]} \frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_j}.$$

Note that

$$\frac{d\ell[t](\boldsymbol{\mu}[t])}{d\mu[t]} = \frac{\pi y[t]^2}{2\mu[t]^3} - \frac{2}{\mu[t]},$$

and

$$\frac{d\mu[t]}{d\eta[t]} = \frac{1}{g'(\mu[t])},$$

where $g'(\cdot)$ is the first derivative of the link function $g(\cdot)$. In particular, for log link we have $g'(\mu[t]) = \mu[t]$. Thus, we can write

$$\frac{\partial \ell(\boldsymbol{\gamma})}{\partial \gamma_j} = \sum_{t=m+1}^T \left(\frac{\pi y[t]^2}{2\mu[t]^3} - \frac{2}{\mu[t]} \right) \frac{1}{g'(\mu[t])} \frac{\partial \eta[t]}{\partial \gamma_j}.$$

The derivatives of η with respect to ζ , and each element of $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}$ are given, respectively, by

$$\begin{aligned} \frac{\partial \eta[t]}{\partial \zeta} &= 1 - \sum_{s=1}^q \theta_s \frac{\partial \eta[t-s]}{\partial \zeta}, \\ \frac{\partial \eta[t]}{\partial \beta_k} &= \mathbf{x}^\top[t] - \sum_{s=1}^q \theta_s \frac{\partial \eta[t-s]}{\partial \beta_k}, \\ \frac{\partial \eta[t]}{\partial \phi_i} &= y[t-i] - \sum_{s=1}^q \theta_s \frac{\partial \eta[t-s]}{\partial \phi_i}, \\ \frac{\partial \eta[t]}{\partial \theta_j} &= r[t-j] - \sum_{s=1}^q \theta_s \frac{\partial \eta[t-s]}{\partial \theta_j}, \end{aligned}$$

with $k = 1, \dots, l$, $i = 1, \dots, p$, and $j = 1, \dots, q$.

Therefore, we can write the score vector in matrix form as

$$\mathbf{U}[\boldsymbol{\gamma}] = (U[\zeta], \mathbf{U}[\boldsymbol{\beta}]^\top, \mathbf{U}[\boldsymbol{\phi}]^\top, \mathbf{U}[\boldsymbol{\theta}]^\top)^\top,$$

where

$$\begin{aligned} U[\zeta] &= \mathbf{a}^\top \mathbf{T} \mathbf{v}, & \mathbf{U}[\boldsymbol{\beta}] &= \mathbf{M}^\top \mathbf{T} \mathbf{v}, \\ \mathbf{U}[\boldsymbol{\phi}] &= \mathbf{P}^\top \mathbf{T} \mathbf{v}, & \text{and } \mathbf{U}[\boldsymbol{\theta}] &= \mathbf{R}^\top \mathbf{T} \mathbf{v}, \end{aligned}$$

with $\mathbf{v} = \left(\left\{ \frac{\pi y[m+1]^2}{2\mu[m+1]^3} - \frac{2}{\mu[m+1]} \right\}, \dots, \left\{ \frac{\pi y[T]^2}{2\mu[T]^3} - \frac{2}{\mu[T]} \right\} \right)^\top$, $\mathbf{T} = \text{diag} \left\{ 1/g'(\mu[m+1]), \dots, 1/g'(\mu[T]) \right\}$, and $\mathbf{a} = \left(\frac{\partial \eta[m+1]}{\partial \zeta}, \dots, \frac{\partial \eta[T]}{\partial \zeta} \right)^\top$. The (i, j) -th elements of matrices \mathbf{M} , \mathbf{P} , and \mathbf{R} with dimensions $(T - m) \times l$, $(T - m) \times p$, and $(T - m) \times q$, respectively, can be defined as

$$\begin{aligned} \mathbf{M}[i, j] &= \frac{\partial \eta[i + m]}{\partial \beta_j}, & \mathbf{P}[i, j] &= \frac{\partial \eta[i + m]}{\partial \phi_j}, \\ \mathbf{R}[i, j] &= \frac{\partial \eta[i + m]}{\partial \theta_j}. \end{aligned}$$

The CMLE of the RARMA(p, q) model parameters in (5) are thus obtained by solving the nonlinear system $\mathbf{U}[\boldsymbol{\gamma}] = \mathbf{0}$, where $\mathbf{0}$ is a null vector with dimension $l + p + q + 1$. Since there is no closed-form solution, numerical methods are necessary. In the present work, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [53] with analytic first derivatives is considered. However, since the iterative BFGS algorithm requires initial values for the constant (ζ), the autoregressive (ϕ) parameters, and the regressors coefficients (β), they are selected by an ordinary least squares estimate via a linear regression, where $\mathbf{Y} = (g(y[m+1]), \dots, g(y[T]))^\top$ are the responses and the covariate matrix is given by:

$$\mathbf{X} = \begin{bmatrix} 1 x_1[m+1] \cdots x_l[m+1] & g(y[m]) & \cdots & g(y[m+1-p]) \\ 1 x_1[m+2] \cdots x_l[m+2] & g(y[m+1]) & \cdots & g(y[m+2-p]) \\ \vdots & \ddots & \ddots & \vdots \\ 1 x_1[T] \cdots x_l[T] & g(y[T-1]) & \cdots & g(y[T-p]) \end{bmatrix}.$$

Instead, as in [16], [52], the initial values for the moving average (θ) parameters are set to zero.

B. Large Data Record Inference

Under usual mathematical regularity conditions [9], [54], the maximum likelihood-based estimators are consistent and satisfy the following conditions:

$$\hat{\boldsymbol{\gamma}} \xrightarrow{d} \mathcal{N}(\boldsymbol{\gamma}, \mathbf{K}^{-1}(\boldsymbol{\gamma})), \quad (6)$$

for $T \rightarrow \infty$, where \xrightarrow{d} means convergence in distribution, $\mathcal{N}(\boldsymbol{\gamma}, \mathbf{K}^{-1}(\boldsymbol{\gamma}))$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{\gamma}$ and covariance matrix $\mathbf{K}^{-1}(\boldsymbol{\gamma})$, and $\mathbf{K}(\boldsymbol{\gamma})$ is the Fisher information matrix. In order to obtain the Fisher information matrix we need to compute the expected values of all second order derivatives of the conditional log-likelihood function. These analytical results are presented in the Appendix.

Additionally, the CMLE asymptotic distribution can be used to construct confidence intervals for the j -th component of $\boldsymbol{\gamma}$,

γ_j , $j = 1, \dots, l + p + q + 1$, with confidence approximately $100(1 - \alpha)\%$ as in [54]:

$$[\hat{\gamma}_j - z_{1-\alpha/2}\text{ep}(\hat{\gamma}_j); \hat{\gamma}_j + z_{1-\alpha/2}\text{ep}(\hat{\gamma}_j)],$$

where $\text{ep}(\hat{\gamma}_j) = \sqrt{\mathbf{K}_{jj}^{-1}(\hat{\gamma})}$ is the estimated standard error of $\hat{\gamma}_j$, $\mathbf{K}_{jj}^{-1}(\hat{\gamma})$ is the j -th element of the diagonal of $\mathbf{K}^{-1}(\hat{\gamma})$, α is the significance level, and z_ρ is the ρ quantile of the standard normal distribution.

Finally, to test hypotheses over the RARMA model parameters, the likelihood-based detection theory [54], [55] can be considered. Let γ_j^0 be a given hypothesized value for the true parameter γ_j . To test

$$\begin{cases} \mathcal{H}_0 : \gamma_j = \gamma_j^0, \\ \mathcal{H}_1 : \gamma_j \neq \gamma_j^0, \end{cases}$$

we can apply the Z statistic, given by [54]

$$Z = \frac{\hat{\gamma}_j - \gamma_j^0}{\text{ep}(\hat{\gamma}_j)}.$$

Based on (6), and under \mathcal{H}_0 , the limiting distribution of Z statistic is standard normal. Therefore, the test is performed by comparing the calculated Z statistic with the usual quantiles of the standard normal distribution for the desired probability of false alarm α . Other commonly applied statistics to test more general restrictions over the model parameters are the Wald (W), likelihood ratio (LR), and Rao's score statistics (S) [55], [56]. The statistics W , LR , and S are all asymptotically chi-squared distributed under the null hypothesis.

IV. DIAGNOSTIC AND FORECASTING

Typically, time series analysis aims at studying and describing the inherent dynamic structure of an interested signal. Diagnostic measures are useful to evaluate the accuracy of the fitted model to describe the interested signal and obtain accurate out-of-signal forecasting [57]. To this aim, in this work residuals and information criteria are used.

The behavior of residuals is useful to evaluate the accuracy of the fitted model [15]. Different types of residuals are considered in literature for several classes of models. The quantile residual [58] is given by

$$e[t] = \Phi^{-1}(F(y[t]; \hat{\mu}[t])), \quad (7)$$

where Φ^{-1} denotes the standard normal quantile function and $\hat{\mu}[t]$ is the estimated value of $\mu[t]$ using the CMLE of γ , $\hat{\gamma}$ in the dynamic structure (3).

In a good model adjustment, the residuals should display white noise behavior, i.e., they should follow a zero mean and constant variance uncorrelated process [15]. To verify such conditions, the standard Ljung-Box [59] test and the residual correlograms can be used.

Information criteria are important tools for automatic model comparison/selection [60]. In order to help in the selection of the most suitable RARMA model, some standard information criteria, such as Akaike's (AIC) [61] and Schwartz's (SIC) [62], can be obtained from the maximized

conditional log-likelihood function. Among a set of competitive models with different orders, the best model is the one minimizing the considered information criteria.

The other important aspect of dynamic models is their ability to predict variables. This is specifically significant for geophysical variables, since it is mandatory to develop accurate approaches that can represent and propagate prediction uncertainties [50]. Long-term forecasting, for instance of climate variables, are crucial to model future scenarios and subsequent adaptation policies. Forecasts in the proposed RARMA(p, q) model can be produced using the theory of traditional ARIMA models [5]. Let H denote the forecast horizon: accordingly, the mean response estimate at $T + h$, with $h = 1, \dots, H$, can be obtained as

$$\begin{aligned} \hat{\mu}[T+h] = g^{-1} & \left(\hat{\zeta} + \mathbf{x}^\top[T+h]\hat{\beta} + \sum_{i=1}^p \hat{\phi}_i \{g(y[T+h-i])\} \right. \\ & \left. + \sum_{j=1}^q \hat{\theta}_j \{r[T+h-j]\} \right), \end{aligned}$$

where

$$\begin{aligned} \{g(y[T+h-i])\} &= \begin{cases} g(\hat{\mu}[T+h-i]), & \text{if } i < h, \\ g(y[T+h-i]), & \text{if } i \geq h, \end{cases} \\ \{r[T+h-j]\} &= \begin{cases} 0, & \text{if } j < h, \\ \hat{r}[T+h-j], & \text{if } j \geq h, \end{cases} \end{aligned}$$

and $\hat{r}[t] = g(y[t]) - g(\hat{\mu}[t])$.

V. MONTE CARLO EXPERIMENTS

This section presents a Monte Carlo simulation performed to assess the finite signal length performance of the conditional likelihood inference applied to the RARMA model parameters. All the implementations were carried out using the R language [63]. The theory validation based on synthetic signals is important for applications in actual remote sensed signals.

The synthetic signal $y[t]$ was generated from a Rayleigh distribution by (1) with the dynamical structure of a RARMA(p, q) model in (3). For simulations, we considered two scenarios in a similar way as in [16], [64]. The parameter values in each scenario guarantee that the AR and MA characteristic polynomials do not have common roots and the AR characteristic polynomial does not have unit roots. First, the RARMA(1, 1) with one covariate was simulated with the following parameter values: $\zeta = 0.50$, $\phi_1 = 0.45$, $\theta_1 = -0.30$, and $\beta_1 = 0.60$. The covariate was generated from the uniform distribution (0, 1) and considered constant for all Monte Carlo replicas. Second, the RARMA(2, 2) with the following parameter values: $\zeta = 0.40$, $\phi_1 = 0.40$, $\phi_2 = -0.20$, $\theta_1 = 0.35$, $\theta_2 = 0.25$, and without covariates was considered. The number of Monte Carlo replicas was set to 5,000 and the signal length was set to $T \in \{100, 300, 500, 1000, 3000, 5000\}$.

In order to evaluate the point estimators, the mean, percentage relative bias (RB%), and mean square error (MSE) of the CMLE were computed. The RB is defined as the ratio between the bias and the true parameter value times 100. To evaluate the interval estimator, the coverage rates (CR) of the confidence interval with a significance level of 5% was computed, too.

TABLE I
MONTE CARLO SIMULATION RESULTS FOR POINT AND INTERVAL ESTIMATION FOR THE RARMA(1, 1) MODEL WITH ONE COVARIATE

Parameter Value	ζ	ϕ_1	θ_1	β_1
	0.500	0.450	-0.300	0.600
$T = 100$				
Mean	0.567	0.391	-0.256	0.605
RB (%)	13.425	-13.169	-14.690	0.847
MSE	0.120	0.065	0.081	0.035
CR	0.863	0.842	0.831	0.946
$T = 300$				
Mean	0.526	0.429	-0.282	0.602
RB (%)	5.197	-4.695	-6.008	0.299
MSE	0.031	0.015	0.019	0.010
CR	0.928	0.919	0.915	0.952
$T = 500$				
Mean	0.515	0.438	-0.290	0.602
RB (%)	2.970	-2.747	-3.463	0.296
MSE	0.017	0.009	0.011	0.006
CR	0.932	0.925	0.923	0.941
$T = 1,000$				
Mean	0.508	0.443	-0.294	0.602
RB (%)	1.681	-1.613	-2.123	0.407
MSE	0.008	0.004	0.005	0.003
CR	0.941	0.926	0.925	0.956
$T = 3,000$				
Mean	0.503	0.448	-0.298	0.601
RB (%)	0.523	-0.535	-0.643	0.172
MSE	0.002	0.001	0.002	0.001
CR	0.951	0.933	0.932	0.952
$T = 5,000$				
Mean	0.502	0.448	-0.299	0.601
RB (%)	0.343	-0.361	-0.451	0.156
MSE	0.002	0.001	0.001	0.001
CR	0.948	0.938	0.935	0.951

TABLE II
MONTE CARLO SIMULATION RESULTS FOR POINT AND INTERVAL ESTIMATION FOR THE RARMA(2, 2) MODEL

Parameter Value	ζ	ϕ_1	ϕ_2	θ_1	θ_2
	0.400	0.400	-0.200	0.350	0.250
$T = 100$					
Mean	0.328	0.574	-0.288	0.165	0.208
RB (%)	-17.941	43.388	44.115	-52.763	-16.953
MSE	0.031	0.219	0.049	0.238	0.048
CR	0.756	0.700	0.746	0.689	0.790
$T = 300$					
Mean	0.371	0.470	-0.231	0.278	0.228
RB (%)	-7.256	17.380	15.627	-20.533	-8.749
MSE	0.013	0.102	0.021	0.105	0.018
CR	0.845	0.801	0.822	0.800	0.847
$T = 500$					
Mean	0.383	0.441	-0.219	0.308	0.238
RB (%)	-4.199	10.200	9.606	-11.950	-4.737
MSE	0.007	0.052	0.012	0.053	0.009
CR	0.888	0.865	0.874	0.865	0.901
$T = 1,000$					
Mean	0.393	0.419	-0.209	0.331	0.245
RB (%)	-1.782	4.684	4.692	-5.456	-1.945
MSE	0.003	0.022	0.006	0.023	0.004
CR	0.913	0.903	0.913	0.904	0.923
$T = 3,000$					
Mean	0.399	0.403	-0.202	0.347	0.249
RB (%)	-0.373	0.850	0.828	-0.964	-0.350
MSE	0.001	0.007	0.002	0.007	0.001
CR	0.943	0.936	0.939	0.937	0.944
$T = 5,000$					
Mean	0.399	0.402	-0.201	0.348	0.250
RB (%)	-0.209	0.376	0.512	-0.455	-0.097
MSE	0.001	0.004	0.001	0.004	0.001
CR	0.943	0.938	0.941	0.935	0.945

The CR is given by the percentage of Monte Carlo replicas in which the parameter is within the confidence interval.

The simulation results are presented in Tables I and II. It is possible to observe that the performance of the CMLE is good, except for the smallest sample size, as expected, in the scenario without covariate. For $T \geq 300$ the mean of the estimate are close to the true parameter values, presenting small relative bias values. The RB and MSE are greatly reduced as the signal length increases. For $T > 1000$ all RB are smaller than 1%, and MSE are close to zero. These figures are numerical evidences of the CMLE consistency.

Regarding the confidence interval evaluation, the CR values are close to the nominal coverage level (0.95) in the first scenario (Table I). In the more complex scenario, considering the RARMA(2,2) in Table II, the CR values are close to the nominal value for $T \geq 1000$. As expected by the asymptotic normality of the CMLE, the CR converges to the nominal level as the signal length increases. In general, the numerical results show that the overall performance of the CMLE for the RARMA model is very good.

VI. REMOTE SENSING DATA INTERPRETATION USING THE RARMA MODEL

As the proposed RARMA model can be used to model continuous and nonnegative discrete-time signal, it can be applied, as mentioned in the Introduction, to many different remote sensing data sets. In this work, we focus our attention on wind speed measurements, presented in Section VI-A, and on multitemporal SAR data sequences, discussed in Section VI-B.

A. Wind Speed Modeling

Wind speed is an important parameter for studies about atmospheric systems [29], ocean-atmospheric mechanisms [31], and especially for wind energy applications. Indeed, wind power is a sustainable alternative to conventional energy sources such as fossil fuels. To understand how to fully exploit this potential, several works have been studying wind speed distributions [32]–[37], and specifically the Rayleigh model have been considered for these models. We expect that the novel approach presented in this paper, in Sections II–IV, capable to capture dynamic behaviors, will be better suited than already proposed ones to model and forecast wind speed.

To this aim, we considered two different sets of wind speed data. The primary time series used in this analysis is the mean of 13 altimeter records already used in [31] for evaluating trends in oceanic wind speeds. The occurrences considered here are related to the Southern Ocean site on 40°S, 250°E and is graphically presented in Fig. 1. The ocean wind speeds range from 4.759 to 12.611 m/s with the unconditional sample mean equals to 8.636 m/s in the period from August 1990 to July 2017. The second one is the monthly mean wind speed data for Yenagoa, Nigeria, and covering the period from January 2013 to December 2017. The dataset was obtained from [19] and is presented in Fig. 2. The signal unconditional mean is equal to 1.719 m/s ranging from 0.579 to 3.392 m/s. The unconditional mean could not be representative, because, as discussed in [19], the wind speed can be different in each

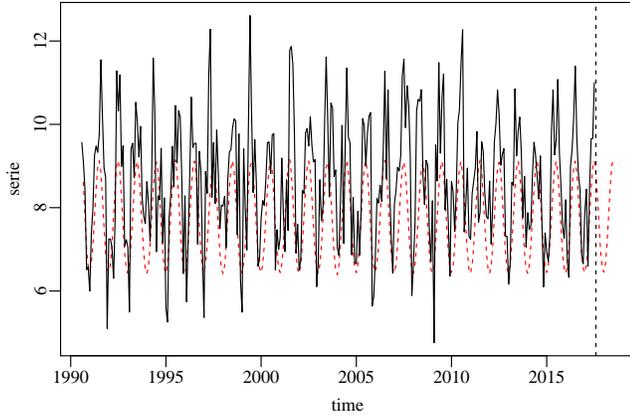


Fig. 1. Observed (solid line) and fitted (dashed line) Southern Ocean (40°S , 250°E) wind speed values.

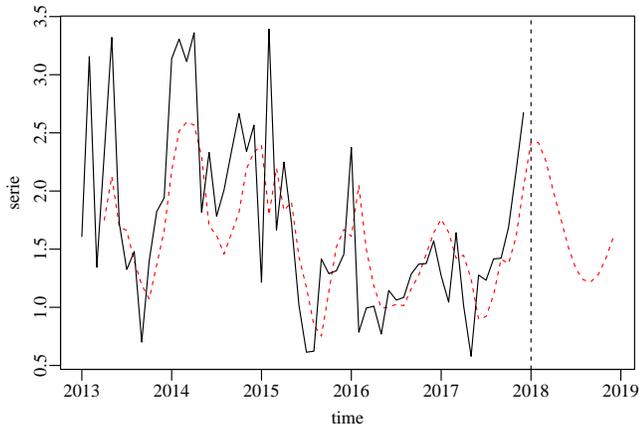
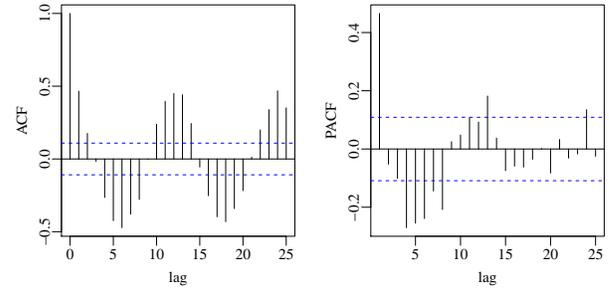


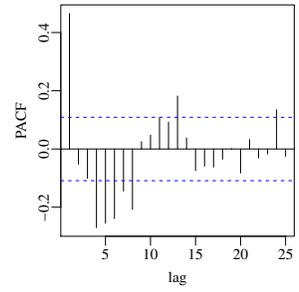
Fig. 2. Observed (solid line) and fitted (dashed line) Yenagoa wind speed values.

month (seasonality). In addition, the autocorrelation function (ACF) and partial ACF (PACF) of both time series (y) (Fig. 3 and Fig. 4) show significant autocorrelations. Accordingly, the proposed RARMA has a big potential for modeling the conditional mean (2) of the wind speed over time in these two situations.

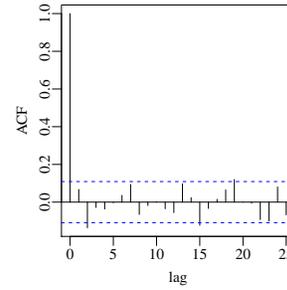
Fig. 3 and Fig. 4 make the seasonal patterns of the measurements quite evident because, as discussed in [19], the wind speed depends on the month of the year. To take into account the seasonality in both signals, the covariate $x_t = \cos(2\pi t/12)$, with $t \in 1, \dots, T+H$, has to be considered, like in a harmonic regression approach [65]. If the p -value of the β_1 in the fitted model in Table III, which is the parameter related to the seasonal covariate x_t , is smaller than a significance level, then the presence of seasonality is statistically significant. To select the best RARMA model to each wind speed signal, different p and q orders have been systematically checked, eventually selecting the p and q values whose AIC and SIC present the smallest values, all parameters in dynamic structure were significant (at 10% significance level), and the residual autocorrelations are null. To check the residual autocorrelations, the Ljung-Box test and residual



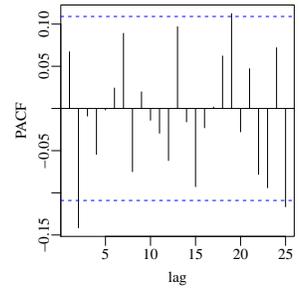
(a) Autocorrelation function (ACF) of original signal (y)



(b) Partial ACF (PACF) of y

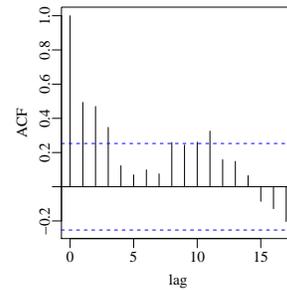


(c) ACF of residuals

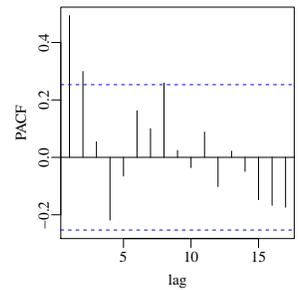


(d) PACF of residuals

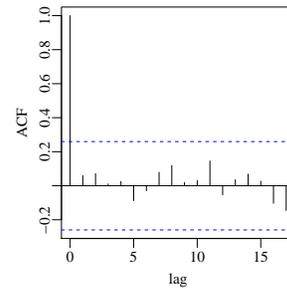
Fig. 3. Correlograms of the Southern Ocean (40°S , 250°E) wind speed time series.



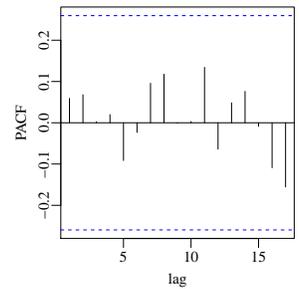
(a) ACF of y



(b) PACF of y



(c) ACF of residuals



(d) PACF of residuals

Fig. 4. Correlograms of the Yenagoa wind speed time series.

TABLE III
FITTED RARMA MODELS FOR THE WIND SPEED DATA IN SOUTHERN OCEAN AND YENAGOA

Ocean data				
	Estimate	Standard error	Z stat	p-value
ζ	0.9793	0.2786	3.5145	0.0004
ϕ_1	0.5197	0.1365	3.8080	0.0001
θ_1	-0.5363	0.2311	2.3203	0.0203
β_1	0.1066	0.0391	2.7243	0.0064
AIC= 1615.8713 SIC= 1634.7750				
Ljung-Box (p -value) = 0.2628 (with 12 lags)				
Yenagoa data				
	Estimate	Standard error	Z stat	p-value
ζ	0.2843	0.1416	2.0084	0.0446
ϕ_3	0.3164	0.1871	1.6907	0.0909
θ_1	0.3772	0.1945	1.9393	0.0525
θ_2	0.3741	0.2011	1.8600	0.0629
β_1	0.2511	0.1434	1.7512	0.0799
AIC= 124.6298 SIC= 137.1959				
Ljung-Box (p -value) = 0.9756 (with 12 lags)				

TABLE IV
MSE AND MAPE FOR IN-SAMPLE PREDICTED VALUES FROM DIFFERENT MODELS IN DIFFERENT SIGNALS

Model	Ocean data		Yenagoa data	
	MSE	MAPE	MSE	MAPE
RARMA	1.9937	12.9332	0.3116	29.8738
Holt-Winters	38.7448	56.5354	2.0110	71.2778
ETS	1.0911	9.8263	0.4359	36.5884

correlograms have been considered. The estimates, standard errors, Z statistics and their associated p -values for the fitted RARMA model in the two considered test cases are presented in Table III. The AIC, SIC, and the results of Ljung-Box test with 12 lags are also presented in the same table. For the ocean times series the RARMA(1, 1) model was eventually selected, while the RARMA(3, 2), with ϕ_1 and ϕ_2 not significant, was found the most suitable for the Yenagoa wind speed signal.

The residual ACF (Fig. 3(c) and Fig. 4(c)) and the residual PACF (Fig. 3(d) and Fig. 4(d)) of the adjusted models for the Southern ocean and Yenagoa signals, respectively, evidence the goodness of fit of the models, enabling the possibility to perform predictions based on the adjusted RARMA models. The predicted (in-sample) and twelve predicted values steps forward (out-of-sample forecast) are presented in Fig. 1 and Fig. 2. We can see that the predicted values are close to the actual values and the forecast values capture the seasonal behavior of the signals. As the signals of interest are asymmetric and nonnegative, the Gaussian-based ARIMA models are not recommended to fit them. For comparative purposes, we adjusted the nonparametric exponential smoothing Holt-Winters (additive) model for each signal using the `forecast` R package [66] and the state space framework (ETS) by `ets` package [67], [68]. Table IV presents the mean square error (MSE) and the mean absolute percentage error (MAPE) computed for each methodology in each signal. It is noteworthy that the RARMA predictions outperforms the Holt-Winters model according to both criteria in both signals and presents competitive results with ETS approach.

B. SAR Land-use Classification

A second set of measurements for which the novel model enable the possibility to capture complex dynamic behaviors is composed by time series of SAR images. These sequences are currently used for land-use classification and change detection, an increasing important topic in Earth observation [42], [69], [70]. Indeed, although several methods have been proposed to extract features for accurate classification and change detection, there is no paper, to the best of our knowledge, that exploits parametric non-Gaussian time series for SAR time series analysis.

The test sequence used in this work is composed by 53 ESA Sentinel-1 acquisitions between 2015 and 2016 over the city of Beijing, P.R. China. The SAR images include both VV and VH polarizations, and have been used in past works to achieve a binary classification between urban and non-urban areas, e.g., in [42]. The selected Beijing subset has 900×626 pixels and presents 28.79% urban points. For classification purposes, known urban points were randomly split into training and validation sets with a 10%-90% proportion. Fig. 5(a) shows the actual map of the urban extents for the considered scene, to be compared with the classification maps extracted by the proposed approach, presented in the same figure.

The RARMA model is used to extract attributes from the SAR sequence, following the procedure described by these steps:

- 1) In the training set, 500 urban pixels and 500 non-urban pixels are randomly selected.
- 2) For each of these 1,000 time series (each pixel), a RARMA(2,2) model is fitted.
- 3) The average of the estimated parameters are computed for urban and non-urban areas. Let's define $\hat{\gamma}_u$ the average of the parameter vector estimates for urban pixels, and $\hat{\gamma}_n$ the average of the estimates for non-urban pixels.
- 4) For each pixel in the validation dataset, the log-likelihood function of $\hat{\gamma}_u$ and $\hat{\gamma}_n$ is evaluated. Thus, a binary variable is created:

$$I = \begin{cases} 1, & \text{if } \ell(\hat{\gamma}_u) > \ell(\hat{\gamma}_n) \text{ (urban is more likely),} \\ 2, & \text{if } \ell(\hat{\gamma}_u) < \ell(\hat{\gamma}_n) \text{ (non-urban is more likely).} \end{cases}$$

- 5) For each pixel, the quantile residuals (7) are extracted, evaluating the model at $\hat{\gamma}_u$ and $\hat{\gamma}_n$. Based on the residuals of the model for the urban training points, $\hat{\gamma}_u$, the mean (m_u), the standard deviation (s_u), the coefficient of skewness (a_u), and the coefficient of kurtosis (k_u) are computed. The same is performed for the non-urban area training points, obtaining the same values, labeled m_n , s_n , a_n , and k_n , respectively.

Finally, each pixel is represented by a vector of 11 features: I , $\ell(\hat{\gamma}_u)$, $\ell(\hat{\gamma}_n)$, m_u , s_u , a_u , k_u , m_n , s_n , a_n , and k_n . Until this step, all the theoretical concepts developed in Sections II–IV have been used. This vector is eventually input to support vector machine (SVM) [45], [71], k -nearest neighbor (KNN) [46], [72], and random forest (RF) [47], [48], [73] classifiers, with the aim to extract urban area extents.

TABLE V
ACCURACY EVALUATION FOR THE MULTITEMPORAL SAR URBAN EXTENT
EXTRACTION RESULTS

Method	Acc	Sens	Spec	Prec	F1
RARMA+SVM	0.9285	0.9566	0.8588	0.9437	0.9501
RARMA+KNN	0.9163	0.9280	0.8872	0.9531	0.9404
RARMA+RF	0.9291	0.9548	0.8655	0.9461	0.9504
IT+RF [42]	0.9249	0.9496	0.8771	0.9371	0.9433
IT+PO [42]	0.9387	0.9389	0.9385	0.9671	0.9528

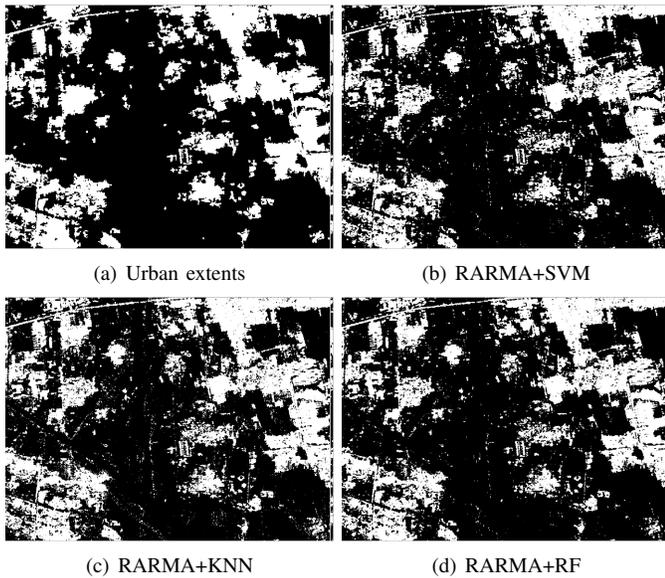


Fig. 5. (a) Ground truth (GT) map for urban extents in the Beijing subset used in this paper. Urban and non-urban areas are shown in white and black, respectively. The GT should be compared with the classification maps achieved by (b) RARMA+SVM, (c) RARMA+KNN, and (d) RARMA+RF.

Specifically, to implement the classifier through SVM, the C classification method with default parameters of R function `ksvm` [71] was used. For classification through KNN, 4-nearest neighbors according to Euclidean distance were considered. Finally, RF was also considered, because it is known to have the computationally most effective procedures to improve unstable estimates, especially in complex problems [74]. In this work a set of 500 regression trees was used.

For the evaluation of the resulting classification maps, standard quantitative measures (overall accuracy (Acc), sensitivity (Sens), specificity (Spec), precision (Prec), and F1 score) were considered. Eventually, to perform a benchmarking comparison, the two state-of-the-art information theory-based (IT) classification methods proposed in [42] were used.

The results show that all the methods present very accurate performances, with F1 score greater than 0.94. Specifically, the performance of the method based on RARMA is similar to those introduced in [42]. We emphasize that the method presented here is not specific for feature extraction in land-use classification problems. It is just one possible application of the RARMA model. The fact that results are accurate even if the procedure was not designed just for urban extents extraction provides an important conformation that the approach is promising and competitive.

In order to evaluate the computational efficiency of the

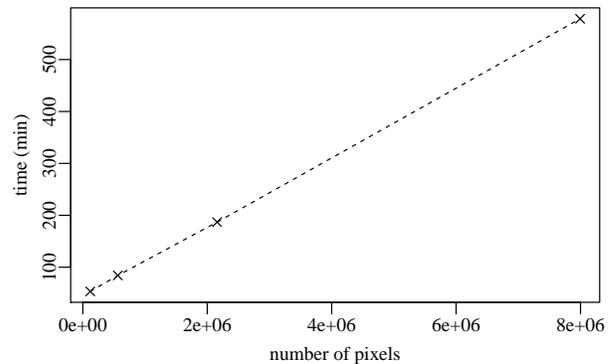


Fig. 6. Execution time for the proposed feature extraction/classification method versus the number of pixels of the input data set.

proposed feature extraction method, we considered subsets of the original sequence with different image sizes, namely: 400×300 , 900×626 , 1700×1200 , and 3265×2448 . The five steps of the proposed method were implemented in R and run on a laptop with 64 bits Intel i7-3537U 2.00 GHz CPU core and 8 GB RAM. The execution times for each image size (i.e., number of pixels) are presented in Fig. 6. From the graph it is clear that the proposed method is efficient, and its computational cost increases linearly with the image size. Accordingly, the use of a RARMA model is compatible with the use of platform with limited performances.

VII. CONCLUSION

This paper introduces a novel dynamic regression model, the Rayleigh autoregressive moving average model (RARMA). RARMA is a non-Gaussian dynamic model for Rayleigh-distributed discrete-time signals. This paper discusses how to estimate the model parameters by means of conditional maximum likelihood approach, and introduces closed form expressions for the score vector and the conditional Fisher information matrix. Monte Carlo simulations were used to assess the finite signal length performance of inference methods applied to the RARMA model parameters. These simulations show that maximizing the conditional log-likelihood function is an adequate choice for parameter estimation, even in case of moderate signal length.

Since the model was developed for remote sensing applications, two tests on remotely sensed data have been introduced. The first one is related to wind speed modeling, which could be useful for wind power production projects and polices. The results show that the RARMA model is able to detect the wind speed behavior providing accurate predicted values. The second test involves multi-temporal SAR data for land-use classification. In this context, the RARMA model proved to be a competitive tool for the extraction of SAR sequence attributes for urban extent extraction

Both example and the previous theoretical analysis show that the RARMA model is a flexible and useful alternative for modeling asymmetric and nonnegative discrete-time signals, and that it has application to accurate forecasts of, and feature extraction from, different classes of remotely sensed time series data.

APPENDIX
CONDITIONAL INFORMATION MATRIX

In this Appendix we introduce the conditional Fisher information matrix for the RARMA(p, q) model. To this aim, let's compute the expected values of all second order derivatives:

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} &= \sum_{t=m+1}^T \frac{\partial}{\partial \mu[t]} \left(\frac{d\ell[t](\mu[t])}{d\mu[t]} \frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_j} \right) \frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_i} \\ &= \sum_{t=m+1}^T \left[\frac{d^2 \ell[t](\mu[t])}{d\mu[t]^2} \frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_j} \right. \\ &\quad \left. + \frac{d\ell[t](\mu[t])}{d\mu[t]} \frac{\partial}{\partial \mu[t]} \left(\frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_j} \right) \right] \frac{d\mu[t]}{d\eta[t]} \frac{\partial \eta[t]}{\partial \gamma_i}. \end{aligned}$$

As shown in [17], $E(d\ell[t](\mu[t])/d\mu[t]|\mathcal{F}[t-1]) = 0$, hence

$$\begin{aligned} E\left(\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}[t-1]\right) &= \sum_{t=m+1}^T E\left(\frac{d^2 \ell[t](\mu[t])}{d\mu[t]^2} \middle| \mathcal{F}[t-1]\right) \\ &\quad \times \left(\frac{d\mu[t]}{d\eta[t]}\right)^2 \frac{\partial \eta[t]}{\partial \gamma_i} \frac{\partial \eta[t]}{\partial \gamma_j}. \end{aligned}$$

The derivatives $\partial \ell[t](\mu[t])/d\mu[t]$, $d\mu[t]/d\eta[t]$ and $\partial \eta[t]/\partial \gamma_j$ have been defined in Section III. The second derivative of $\ell[t](\mu[t])$ with respect to $\mu[t]$ is given by

$$\frac{d^2 \ell[t](\mu[t])}{d\mu[t]^2} = \frac{2}{\mu[t]^2} - \frac{3\pi y[t]^2}{2\mu[t]^4}.$$

Computing the conditional expectation, the result is

$$E\left(\frac{\partial^2 \ell}{\partial \gamma_i \partial \gamma_j} \middle| \mathcal{F}[t-1]\right) = \sum_{t=m+1}^T \frac{-4}{\mu[t]^2} \left(\frac{1}{g'(\mu[t])}\right)^2 \frac{\partial \eta[t]}{\partial \gamma_i} \frac{\partial \eta[t]}{\partial \gamma_j}.$$

Let $\mathbf{V}[i, j] = \text{diag}\left(\frac{-4}{\mu[1]^2} \left(\frac{1}{g'(\mu[1])}\right)^2, \dots, \frac{-4}{\mu[T]^2} \left(\frac{1}{g'(\mu[T])}\right)^2\right)$, the conditional Fisher information matrix for $\boldsymbol{\gamma}$ is

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\gamma}) = - \begin{bmatrix} K_{(\zeta, \zeta)} & \mathbf{K}_{(\zeta, \beta)} & \mathbf{K}_{(\zeta, \phi)} & \mathbf{K}_{(\zeta, \theta)} \\ \mathbf{K}_{(\beta, \zeta)} & \mathbf{K}_{(\beta, \beta)} & \mathbf{K}_{(\beta, \phi)} & \mathbf{K}_{(\beta, \theta)} \\ \mathbf{K}_{(\phi, \zeta)} & \mathbf{K}_{(\phi, \beta)} & \mathbf{K}_{(\phi, \phi)} & \mathbf{K}_{(\phi, \theta)} \\ \mathbf{K}_{(\theta, \zeta)} & \mathbf{K}_{(\theta, \beta)} & \mathbf{K}_{(\theta, \phi)} & \mathbf{K}_{(\theta, \theta)} \end{bmatrix},$$

where $K_{(\zeta, \zeta)} = \mathbf{a}^\top \mathbf{V} \mathbf{a}$, $\mathbf{K}_{(\beta, \zeta)} = \mathbf{M}^\top \mathbf{V} \mathbf{a}$, $\mathbf{K}_{(\phi, \zeta)} = \mathbf{P}^\top \mathbf{V} \mathbf{a}$, $\mathbf{K}_{(\theta, \zeta)} = \mathbf{R}^\top \mathbf{V} \mathbf{a}$, $\mathbf{K}_{(\zeta, \beta)} = \mathbf{a}^\top \mathbf{V} \mathbf{M}$, $\mathbf{K}_{(\beta, \beta)} = \mathbf{M}^\top \mathbf{V} \mathbf{M}$, $\mathbf{K}_{(\phi, \beta)} = \mathbf{P}^\top \mathbf{V} \mathbf{M}$, $\mathbf{K}_{(\theta, \beta)} = \mathbf{R}^\top \mathbf{V} \mathbf{M}$, $\mathbf{K}_{(\zeta, \phi)} = \mathbf{a}^\top \mathbf{V} \mathbf{P}$, $\mathbf{K}_{(\beta, \phi)} = \mathbf{M}^\top \mathbf{V} \mathbf{P}$, $\mathbf{K}_{(\phi, \phi)} = \mathbf{P}^\top \mathbf{V} \mathbf{P}$, $\mathbf{K}_{(\theta, \phi)} = \mathbf{R}^\top \mathbf{V} \mathbf{P}$, $\mathbf{K}_{(\zeta, \theta)} = \mathbf{a}^\top \mathbf{V} \mathbf{R}$, $\mathbf{K}_{(\beta, \theta)} = \mathbf{M}^\top \mathbf{V} \mathbf{R}$, $\mathbf{K}_{(\phi, \theta)} = \mathbf{P}^\top \mathbf{V} \mathbf{R}$, $\mathbf{K}_{(\theta, \theta)} = \mathbf{R}^\top \mathbf{V} \mathbf{R}$.

REFERENCES

[1] K. Zhao, S. Popescu, and X. Zhang, "Bayesian learning with Gaussian processes for supervised classification of hyperspectral data," *Photogrammetric Engineering & Remote Sensing*, vol. 74, no. 10, pp. 1223–1234, 2008.

[2] B. Zhao, Y. Zhong, A. Ma, and L. Zhang, "A spatial Gaussian mixture model for optical remote sensing image clustering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 12, pp. 5748–5759, 2016.

[3] P. Morales-Álvarez, A. Pérez-Suay, R. Molina, and G. Camps-Valls, "Remote sensing image classification with large-scale Gaussian processes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1103–1114, 2018.

[4] A. M. Zoubir, V. Koivunen, E. Ollila, and M. Muma, *Robust Statistics for Signal Processing*. Cambridge University Press, 2018.

[5] G. Box, G. M. Jenkins, and G. Reinsel, *Time Series Analysis: Forecasting and Control*. Hardcover, John Wiley & Sons, 2008.

[6] P. Han, P. X. Wang, S. Y. Zhang, and D. H. Zhu, "Drought forecasting based on the remote sensing data using ARIMA models," *Mathematical and Computer Modelling*, vol. 51, no. 11, pp. 1398–1403, 2010.

[7] A. Fernández-Manso, C. Quintano, and O. Fernández-Manso, "Forecast of NDVI in coniferous areas using temporal ARIMA analysis and climatic data at a regional scale," *International Journal of Remote Sensing*, vol. 32, no. 6, pp. 1595–1617, 2011.

[8] B. G. Palm, D. I. Alves, V. T. Vu, M. I. Pettersson, F. M. Bayer, R. J. Cintra, R. Machado, P. Dammert, and H. Hellsten, "Autoregressive model for multi-pass SAR change detection based on image stacks," in *SPIE Image and Signal Processing for Remote Sensing XXIV*, vol. 10789, 2018.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1993.

[10] E. J. Wegman, S. C. Schwartz, and J. B. Thomas, *Topics in Non-Gaussian Signal Processing*. Springer-Verlag, 1989.

[11] M. L. Tiku, W.-K. Wong, D. C. Vaughan, and G. Bian, "Time series models in non-normal situations: Symmetric innovations," *Journal of Time Series Analysis*, vol. 21, no. 5, pp. 571–596, 2000.

[12] A. V. Rocha and F. Cribari-Neto, "Beta autoregressive moving average models," *Test*, vol. 18, no. 3, pp. 529–545, 2009.

[13] T. Zheng, H. Xiao, and R. Chen, "Generalized ARMA models with martingale difference errors," *Journal of Econometrics*, vol. 189, no. 2, pp. 492–506, 2015.

[14] M. A. Benjamin, R. A. Rigby, and D. M. Stasinopoulos, "Generalized autoregressive moving average models," *Journal of the American Statistical Association*, vol. 98, no. 461, pp. 214–223, 2003.

[15] B. Kedem and K. Fokianos, *Regression Models for Time Series Analysis*. John Wiley & Sons, 2005, vol. 488.

[16] F. M. Bayer, D. M. Bayer, and G. Pumi, "Kumaraswamy autoregressive moving average models for double bounded environmental data," *Journal of Hydrology*, vol. 555, pp. 385–396, 2017.

[17] B. G. Palm, F. M. Bayer, R. J. Cintra, M. Pettersson, and R. Machado, "Rayleigh regression model for ground type detection in SAR imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 10, pp. 1660–1664, 2019.

[18] A. den Dekker and J. Sijbers, "Data distributions in magnetic resonance images: A review," *Physica Medica*, vol. 30, no. 7, pp. 725–741, 2014.

[19] H. N. Amadi, "Wind energy potential assessment of coastal states in South-South Nigeria based on the Weibull distribution model," *European Journal of Electrical and Computer Engineering*, vol. 2, no. 7, 2018.

[20] F. T. Ulaby, T. F. Haddock, and R. T. Austin, "Fluctuation statistics of millimeter-wave scattering from distributed targets," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 3, pp. 268–281, 1988.

[21] D. H. Hoekman, "Speckle ensemble statistics of logarithmically scaled data (radar)," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, no. 1, pp. 180–182, 1991.

[22] M. N. Sumaiya and R. S. S. Kumari, "Unsupervised change detection of flood affected areas in SAR images using Rayleigh-based Bayesian thresholding," *IET Radar, Sonar & Navigation*, vol. 12, no. 5, pp. 515–522, 2018.

[23] G. Lampropoulos, A. Drosopoulos, N. Rey *et al.*, "High resolution radar clutter statistics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 1, pp. 43–60, 1999.

[24] J. W. Goodman, "Some fundamental properties of speckle*," *Journal of the Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, Nov 1976.

[25] J. Jeon and J. W. Taylor, "Using conditional kernel density estimation for wind power density forecasting," *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.

[26] M. P. Clarizia, C. S. Ruf, P. Jales, and C. Gommenginger, "Spaceborne GNSS-R minimum variance wind speed estimator," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 6829–6843, 2014.

[27] A. G. Fore, S. H. Yueh, W. Tang, A. K. Hayashi, and G. S. E. Lagerloef, "Aquarius wind speed products: Algorithms and validation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2920–2927, 2014.

[28] Y. R. Gel, V. Lyubchich, and S. E. Ahmed, *Catching Uncertainty of Wind: A Blend of Sieve Bootstrap and Regime Switching Models for Probabilistic Short-Term Forecasting of Wind Speed*. New York, NY: Springer, 2016, pp. 279–293.

- [29] T. V. La, A. Khenchaf, F. Comblet, and C. Nahum, "Assessment of wind speed estimation from C-Band Sentinel-1 images using empirical and electromagnetic models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 4075–4087, 2018.
- [30] D. Cazau, J. Bonnel, and M. Baumgartner, "Wind speed estimation using acoustic underwater glider in a near-shore marine environment," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2097–2106, 2019.
- [31] I. R. Young and A. Ribal, "Multiplatform evaluation of global trends in wind speed and wave height," *Science*, vol. 364, no. 6440, pp. 548–552, 2019.
- [32] J. Haslett and E. Kelledy, "The assessment of actual wind power availability in Ireland," *International Journal of Energy Research*, vol. 3, no. 4, pp. 333–348, 1979.
- [33] A. Mihelić-Bogdanić and R. Budin, "Specific wind energy as a function of mean speed," *Renewable Energy*, vol. 2, no. 6, pp. 573–576, 1992.
- [34] A. N. Celik, "A statistical analysis of wind power density based on the Weibull and Rayleigh models at the southern region of Turkey," *Renewable Energy*, vol. 29, no. 4, pp. 593–604, 2004.
- [35] F. A. L. Jowder, "Weibull and Rayleigh distribution functions of wind speeds in Kingdom of Bahrain," *Wind Engineering*, vol. 30, no. 5, pp. 439–445, 2006.
- [36] J. Carta, P. Ramírez, and S. Velázquez, "A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands," *Renewable and Sustainable Energy Reviews*, vol. 13, no. 5, pp. 933–955, 2009.
- [37] R. Gupta and A. Biswas, "Wind data analysis of Silchar (Assam, India) by Rayleigh's and Weibull methods," *Journal of Mechanical Engineering Research*, vol. 2, no. 1, pp. 10–24, 2010.
- [38] E. F. Lambin and M. Linderman, "Time series of remote sensing data for land change science," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 7, pp. 1926–1928, 2006.
- [39] E. B. Brooks, V. A. Thomas, R. H. Wynne, and J. W. Coulston, "Fitting the multitemporal curve: A Fourier series approach to the missing data problem in remote sensing analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, pp. 3340–3353, 2012.
- [40] T. Hoberg, F. Rottensteiner, R. Q. Feitosa, and C. Heipke, "Conditional random fields for multitemporal and multiscale classification of optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 2, pp. 659–673, 2015.
- [41] F. Petitjean, J. Inglada, and P. Gancarski, "Satellite image time series analysis under time warping," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 8, pp. 3081–3095, 2012.
- [42] A. Marinoni, G. C. Iannelli, and P. Gamba, "An information theory-based scheme for efficient classification of remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5864–5876, 2017.
- [43] K.-F. Lin and D. Perissin, "Single-polarized SAR classification based on a multi-temporal image stack," *Remote Sensing*, vol. 10, no. 7, 2018.
- [44] C. Marin, F. Bovolo, and L. Bruzzone, "Building change detection in multitemporal very high resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2664–2682, 2015.
- [45] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [46] N. S. Altman, "An introduction to kernel and nearest-neighbor non-parametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.
- [47] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [48] M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.
- [49] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geoscience Frontiers*, vol. 7, no. 1, pp. 3–10, 2016.
- [50] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, 1992.
- [51] A. Karpatne, I. Ebert-Uphoff, S. Ravela, H. A. Babaie, and V. Kumar, "Machine learning for the geosciences: Challenges and opportunities," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2018.
- [52] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed. Springer-Verlag, 1991.
- [53] F. M. Bayer, R. J. Cintra, and F. Cribari-Neto, "Beta seasonal autoregressive moving average models," *Journal of Statistical Computation and Simulation*, vol. 88, no. 15, pp. 2961–2981, 2018.
- [54] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford Science publications, 2001.
- [55] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall, 1998, vol. II.
- [56] A. Buse, "The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note," *The American Statistician*, vol. 36, no. 3, pp. 153–157, 1982.
- [57] W. K. Li, *Diagnostic Checks in Time Series*, 3rd ed. Chapman and Hall/CRC, 2003.
- [58] P. K. Dunn and G. K. Smyth, "Randomized quantile residuals," *Journal of Computational and Graphical Statistics*, vol. 5, no. 3, pp. 236–244, 1996.
- [59] G. M. Ljung and G. E. P. Box, "On a measure of a lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.
- [60] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.
- [61] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [62] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [63] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2018. [Online]. Available: <https://www.R-project.org/>
- [64] C. F. Ansley and P. Newbold, "Finite sample properties of estimators for autoregressive moving average models," *Journal of Econometrics*, vol. 13, no. 2, pp. 159–183, 1980.
- [65] P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*, 2nd ed. Wiley-Interscience, 2013, p. 288.
- [66] R. J. Hyndman, *forecast: Forecasting functions for time series and linear models*, R package version 8.0 ed., 2017. [Online]. Available: <http://github.com/robjhyndman/forecast>
- [67] R. J. Hyndman, A. B. Koehler, R. D. Snyder, and S. Grose, "A state space framework for automatic forecasting using exponential smoothing methods," *International Journal of Forecasting*, vol. 18, no. 3, pp. 439–454, 2002.
- [68] R. J. Hyndman, A. B. Koehler, J. Ord, R. D. Snyder, *Forecasting with Exponential Smoothing: The State Space Approach*. Springer, 2008.
- [69] H. Zhang, H. Lin, and Y. Li, "Impacts of feature normalization on optical and SAR data fusion for land use/land cover classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 5, pp. 1061–1065, 2015.
- [70] D. Lu and Q. Weng, "Use of impervious surface in urban land-use classification," *Remote Sensing of Environment*, vol. 102, no. 1-2, pp. 146–160, 2006.
- [71] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – an S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004.
- [72] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt., *caret: Classification and Regression Training*, 2018, r package version 6.0-81. [Online]. Available: <https://CRAN.R-project.org/package=caret>
- [73] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
- [74] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, no. 2, pp. 197–227, 2016.