

A non-parametric test for detecting the complex-valued nature of time series

Temujin Gautama^{a,*}, Danilo P. Mandic^b and Marc M. Van Hulle^a

^a*Laboratorium voor Neuro- en Psychofysiologie, K.U.Leuven, Campus Gasthuisberg, Herestraat 49 B-3000 Leuven, Belgium*

^b*Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, SW7 2BT, London, UK*

Abstract. Although the emergence of multivariate signals in natural sciences and engineering has emphasised the problem of signal representation, that is, whether signals are by their nature a set of independent observations or multidimensional vectorial quantities, little research has been conducted on detecting the true nature of such signals. It remains unclear, therefore, when the complex-valued approach is to be preferred over the bivariate one, thus, clearly indicating the need for a criterion that addresses this issue. To this cause, we propose a nonparametric statistical test, based on the local predictability in the complex-valued phase space, which discriminates between the bivariate and complex-valued nature of time series. This is achieved in the well-established surrogate data framework. Results on both benchmark and real-world IPIX radar data support the approach.

1. Introduction

Recently, the potential of the use of complex-valued signals, as compared to their real-valued bivariate counterparts, has been highlighted in several branches of physics and digital signal processing (DSP). Consequently, considerable research effort has been directed toward extensions of nonlinear modelling and filtering approaches to cater for complex-valued signals [1–3], the applicability of which has been demonstrated, among others in radar, sonar and phase-only DSP. In practice, the two-dimensional case is frequently encountered and the problem boils down to whether the signals are bivariate scalars or they have full complex-valued (vectorial) representations. In the latter case, a complex-valued approach is likely to exhibit performance advantages over the real-valued bivariate one.

In the field of nonparametric data analysis, the surrogate data method, as originally proposed by Theiler et al. [4], has evolved into a standard technique to test for the presence of nonlinearity in a real-valued time series. In disciplines dealing with nonlinear phenom-

ena, such tests are indispensable, since, in principle, signal nonlinearity should be assessed prior to the utilisation of nonlinear models, the parameters of which are more mathematically involved to determine than those of linear models. Similar reasoning holds for the use of complex-valued signals and models, which require more intricate mathematical description, against the real-valued bivariate ones, as the latter effectively require processing on two separate channels. However, although complex-valued signals are widely used in communications, physics, and biomedical and environmental signal processing, little is known on how to assess the bivariate real-valued versus fully complex-valued nature of such signals, and the common line of thought is based on rather empirical observations. A reliable, statistical test for assessing the complex-valued nature of a signal is still lacking. We therefore aim at providing a deeper insight into how to detect whether an optimal representation of a two-dimensional signal is via its complex-valued representation or as a bivariate scalar.

To that cause, we first extend the surrogate data methodology [4–6] from the nonlinearity analysis of scalar time series toward complex-valued signals (Sec-

*Corresponding author. E-mail: temu@neuro.kuleuven.ac.be.

tion 2). Within this context, a null hypothesis of a complex-valued linear system underlying the time series under study is utilised. Next, based upon the concept of the recently introduced Delay Vector Variance (DVV) technique for quantifying nonlinearity in a signal [7,8], a novel methodology is proposed for characterising (Section 3), and statistically testing (Section 4) the complex-valued nature of a time series. Simulations, which support the analysis, are performed on both benchmark and real-world complex-valued data.

2. Surrogate data

The surrogate data method computes test statistics on the ‘original’ time series (i.e., the time series under study) and a number of so-called ‘surrogates’, which are realisations of a certain null hypothesis, H_0 . These are further used for estimating the distribution of the test statistic under the assumption of H_0 . In this section, a surrogate data generation procedure known as the (real-valued) iterative Amplitude Adjusted Fourier Transform (iAAFT) method [5,6] is shortly introduced, after which an extension of this method toward complex-valued signals is proposed.

2.1. Real-Valued iAAFT

The iAAFT method generates a surrogate for a real-valued (univariate) time series under the null hypothesis that the original time series is generated by a Gaussian linear process, followed by a possibly nonlinear, static (memoryless) observation function, $h(\cdot)$. The surrogates have empirical signal distributions identical to that of the original signal, and amplitude spectra that are approximately identical, or vice versa. Let $\{|S_k|\}$ be the Fourier amplitude spectrum of the original time series, s , and $\{c_k\}$ the (signal value) sorted version of the original time series. Note that k denotes the frequency index for the amplitude spectrum, whereas for a time series, it denotes the time index. At every iteration j of the algorithm, two time series are calculated, namely $r^{(j)}$, which has a signal distribution identical to that of the original, and $s^{(j)}$, which has an amplitude spectrum identical to that of the original. The iAAFT iteration starts with $r^{(0)}$ a random permutation of the time samples of the original time series, and can be described by:

Repeat:

1. compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$

2. compute $s^{(j)}$ as the inverse transform of $\{|S_k| \exp(i\phi_k)\}$
3. compute $r^{(j)}$ by rank-ordering $s^{(j)}$ to match $\{c_k\}$, i.e., sort $\{s_k^{(j)}\}$ in ascending order and set $r_k^{(j)} = c_{\text{rank}(s_k^{(j)})}$

Until error convergence.

The modelling error can be quantified as the mean-square-error (MSE) between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$. The algorithm was extended toward the multivariate case in [6], yielding surrogates that retain not only the amplitude spectra of the variates separately, but also the cross-correlation spectrum. This was achieved by modifying the phase adjustment step (step 1): the cross-correlation between the variates can be retained if the relative phases between the frequency components remains intact. For further details, we refer to [6]. Figure 1(B) shows a real-valued bivariate iAAFT realisation of the Ikeda Map. The chaotic, complex-valued Ikeda Maps in this study are generated from:

$$z_{k+1} = a + b z_k \exp \left(i \left(\phi - \frac{c}{1 + |z_k|^2} \right) \right), \quad (1)$$

with $a = 1$, $b = 0.9$, $\phi = 0.4$ and $c = 6$ (Fig. 1(A) shows an example realisation).

2.2. Complex-Valued iAAFT

A straightforward extension of the univariate iAAFT-method toward complex-valued signals would be to replace the desired amplitude spectrum by the amplitude spectrum of the original complex-valued signal (step 2 in the iAAFT-iteration). In the next step, the desired signal distribution needs to be imposed on the surrogate in the time domain (step 3 in the iAAFT-iteration). This can be achieved by applying the rank-ordering procedure to the real and imaginary parts of the signal separately. However, in practise, for complex-valued signals, it is more important to impose equal empirical distributions on the moduli of the complex-valued samples, rather than on the real and imaginary parts separately. Therefore, we subsequently perform a rank-ordering procedure on the moduli, so as to match the moduli of the original time series. The underlying null hypothesis is that the time series is generated by a linear complex-valued process, driven by Gaussian white noise, followed by a (possibly nonlinear) static observation function, $h(\cdot)$, which operates on the moduli of the complex-valued time samples. We propose the following complex-valued iAAFT (CiAAFT) procedure, using the same conventions as in the iAAFT case:

Repeat:

1. compute the phase spectrum of $r^{(j-1)} \rightarrow \{\phi_k\}$
2. compute $s^{(j)}$ as the inverse transform of $\{|S_k| \exp(i\phi_k)\}$
3. rank-order the real and imaginary parts of $r^{(j)}$ to match the real and imaginary parts of $\{c_k\}$
4. rank-order the moduli of $r^{(j)}$ to match the modulus distribution of $\{c_k\}$

Until error convergence.

The iteration is started with $r^{(0)}$ a random permutation of the complex-valued time samples. Convergence can be monitored as the MSE computed between $\{|S_k|\}$ and the amplitude spectrum of $r^{(j)}$. Simulations suggest that the iteration can be terminated when the MSE decrement is smaller than 10^{-5} , which typically occurs after fewer than 100 iterations. Figure 1(C) shows a CiAAFT realisation of the Ikeda Map, for which the error curve is shown in Fig. 1(D) (the iterative procedure in this case would terminate after 95 iterations).

3. Delay Vector Variance

We have used a complex-valued variant of the Delay Vector Variance (DVV) method [8] for characterising the time series based on its local predictability in phase space over different scales. The advantage of the DVV method is that it yields a representation which captures several aspects of a time series simultaneously (deterministic/stochastic nature, nonlinearity, geometric structure in phase space), and that it can be easily extended toward the complex-valued case. For a given embedding dimension m and resulting time delay embedding representation (i.e., a set of delay vectors (DV), $\mathbf{x}(k) = [x_{k-m}, \dots, x_{k-1}]^T$), a measure of unpredictability, $\sigma^{*2}(r_d)$, is computed for a standardised range of degrees of locality, r_d :

- The mean, μ_d , and standard deviation, σ_d , are computed over all pairwise Euclidean distances between DVs, $\frac{\|\mathbf{x}(i) - \mathbf{x}(j)\|}{\sqrt{\sum_{n=1}^m |x_{i-n} - x_{j-n}|^2}} (i \neq j)$.
- The sets $\Omega_k(r_d)$ are generated such that $\Omega_k(r_d) = \{\mathbf{x}(i) \mid \|\mathbf{x}(k) - \mathbf{x}(i)\| \leq r_d\}$. The range r_d is taken from the interval $[\max\{0, \mu_d - n_d \sigma_d\}; \mu_d + n_d \sigma_d]$, e.g., uniformly spaced, where n_d is a parameter controlling the span over which to perform the DVV analysis.
- For every set $\Omega_k(r_d)$, the variance of the corresponding targets, $\sigma_k^2(r_d)$, is computed as the sum of the variances of the real and imaginary parts. The average over all sets $\Omega_k(r_d)$, normalised by the variance of the time series, σ_x^2 , yields the ‘target variance’, $\sigma^{*2}(r_d)$:

$$\sigma^{*2}(r_d) = \frac{\frac{1}{N} \sum_{k=1}^N \sigma_k^2(r_d)}{\sigma_x^2}. \quad (2)$$

For a more detailed description of the method, we refer to [8]. Note that the computation of the Euclidean distance between complex-valued DVs is equivalent to considering real and imaginary parts as separate dimensions. Since for bivariate time series, a delay vector is generated by concatenating time delay embedded versions of the two dimensions, the complex-valued and bivariate versions of the DVV method are equivalent, and can be conveniently compared, provided that the variance of a bivariate variable is computed as the sum of the variances of each variate.

A DVV plot, \mathcal{D} , is obtained by plotting the target variance, $\sigma^{*2}(r_d)$, as a function of the ‘standardised’ distance, $\frac{r_d - \mu_d}{\sigma_d}$, i.e., standardised to the distribution of pairwise distances between DVs. It is due to this standardisation that DVV plots of different time series can be conveniently compared. The DVV plots for a 1000 sample realisation of the Ikeda Map, \mathcal{D} (thick solid curve), and for the two types of surrogates, \mathcal{D}^b (thin dashed) and \mathcal{D}^c (thin solid), generated using the iAAFT and CiAAFT method, respectively, are shown in Fig. 2(A), using $m = 3$ and $n_d = 3$.

4. Statistical testing

In the framework of surrogate data testing, as introduced by Theiler et al. [4], a time series is characterised by a certain test statistic, which is compared to an empirical distribution of test statistics generated under the assumption of a null hypothesis. This empirical distribution is generated by computing the test statistics for a number of ‘surrogates’, which are different realisations of the original time series, under the assumption of the null. The actual statistical test then evaluates the probability that the test statistic for the original time series is drawn from the empirical null distribution. If this probability is lower than a significance threshold, the null hypothesis is rejected. In the CiAAFT case, a rejection of the null hypothesis that the signal is complex-valued and linear, could be due to a deviation from either of the two properties. Therefore, we propose a different approach: rather than comparing the original time series to the surrogates, we compare surrogates generated under different null hypotheses, namely that of a linear and bivariate time series, H_0^b , and that of a linear and complex-valued time series, H_0^c . The respective surrogates for this test are generated using the bivariate

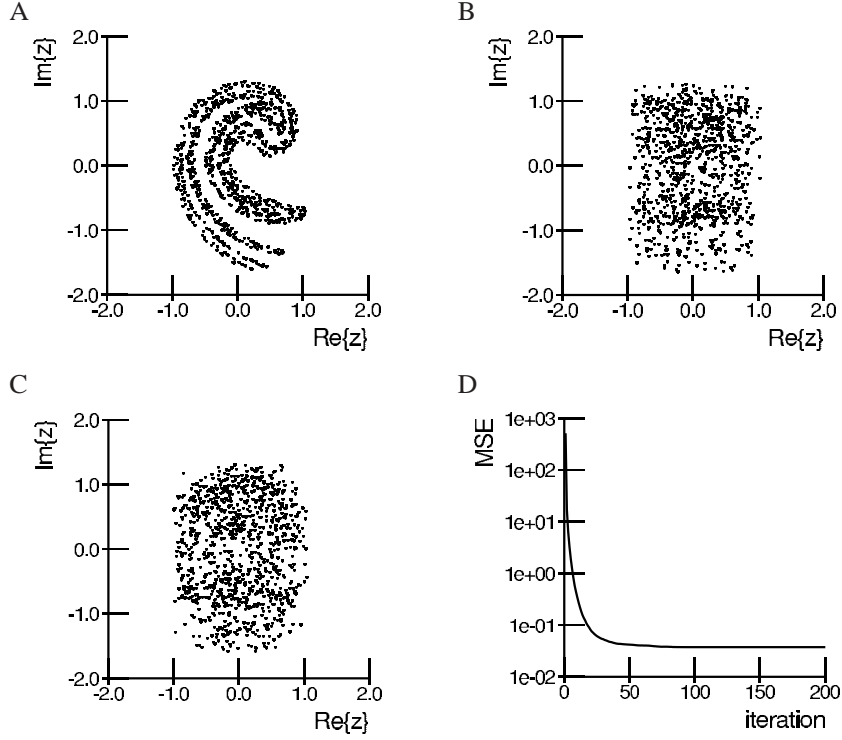


Fig. 1. Realisation of the Ikeda Map (A), iAAFT (B) and CiAAFT (C) surrogates; D) Mean-square-error convergence of the CiAAFT method.

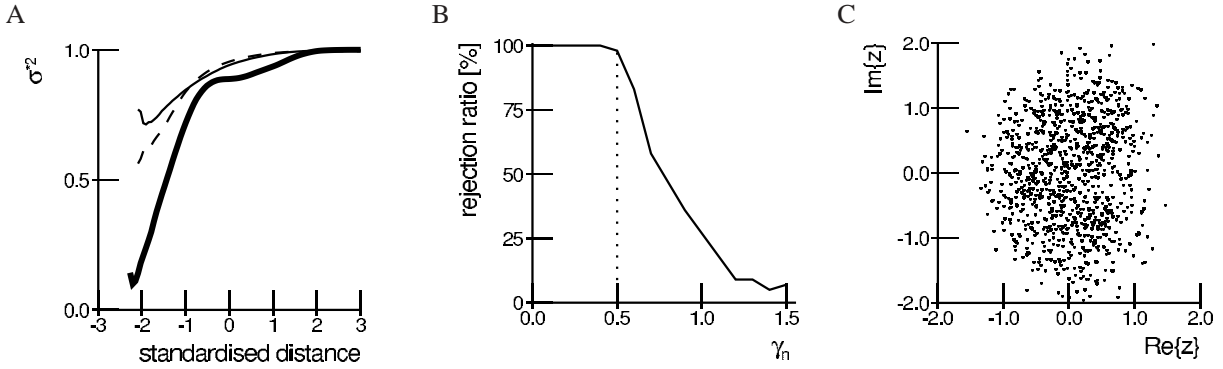


Fig. 2. A) DVV plots for the Ikeda Map (thick solid), for an iAAFT (thin dashed) and a CiAAFT surrogate (thin solid); B) Ratio of (correct) rejections of the null hypothesis for the Ikeda Map, as a function of the noise level γ_n ; C) Example of a noisy Ikeda Map with $\gamma_n = 0.5$ (dotted line in Fig. B).

iAAFT [6] and the proposed CiAAFT method. All time series are characterised using the DVV method, and a significant difference between the two empirical distributions of test statistics is an indication that the original signal is complex-valued. The proposed methodology is the following.

1. Generate $N_{s,\text{ref}}$ CiAAFT surrogates and the average DVV plot $\rightarrow \mathcal{D}^0$;

2. Generate N_s iAAFT surrogates and corresponding DVV plots $\rightarrow \{\mathcal{D}^b\}$;
3. Generate N_s CiAAFT surrogates and corresponding DVV plots $\rightarrow \{\mathcal{D}^c\}$;
4. Compare $(\mathcal{D}^0 - \{\mathcal{D}^b\})$ and $(\mathcal{D}^0 - \{\mathcal{D}^c\})$.

$N_{s,\text{ref}}$ and N_s are the number of surrogates generated for generating, respectively, the ‘reference’ DVV plot and the empirical distributions of test statistics. To perform the final step (4) in a statistical manner, the (cu-

mulative) empirical distributions of root-mean-square distances between $\{\mathcal{D}^b\}$ and \mathcal{D}^0 , and between $\{\mathcal{D}^c\}$ and \mathcal{D}^0 , are compared using a Kolmogorov-Smirnoff (K-S) test. This way, the different types of *linearisations* (bivariate, $\{\mathcal{D}^b\}$, and complex-valued, $\{\mathcal{D}^c\}$) are compared to the ‘reference’ linearisation given a complex-valued nature of the time series, \mathcal{D}^0 . If the two distributions of test statistics are significantly different at a certain level α , the original time series is complex-valued. Note that assumptions regarding the possible nonlinearity of the signal are avoided.

5. Simulations

5.1. Synthetic time series

The proposed algorithm was tested on five sets of synthetically generated benchmark signals containing $N = 1000$ time samples. The two linear sets contained time series 1) consisting of time samples that are drawn from a bivariate Gaussian distribution, $\mathcal{N}([0, 0], [1, 2])$, rotated over an angle of $\frac{\pi}{3}$ (**linear bivariate**, “LB”), and 2) generated by considering a Gaussian ‘amplitude spectrum’, adding random phase and computing the inverse FFT (**linear complex**, “LC”). The two nonlinear sets were generated by a nonlinear system described in [9]:

$$y_k = \gamma \frac{x_{k-1} x_{k-2} (x_{k-1} + 2.5)}{1 + x_{k-1}^2 + x_{k-2}^2} + x_k, \quad (3)$$

where γ is a parameter controlling the prevalence of the nonlinear over the linear part of the signal, which was set to $\gamma = 0.6$, unless stated otherwise. In the first nonlinear set (**nonlinear bivariate**, “NLB”), both dimensions of “LB” were separately passed through the nonlinear system, and in the second set (**nonlinear complex**, “NLC”), x represents the complex-valued time series “LC”. The final set contained realisations of the Ikeda Map (Eq. (1); an example is shown in Fig. 1(A)).

For each of the five sets, 100 realisations of the time series are generated, to each of which the proposed test is applied. For the bivariate sets (LB and NLB), the number of (erroneous) rejections is of the order expected from the $\alpha = 0.05$ level (5/100 and 1/100). The proposed test does not perform well on the LC set: only 16/100 the time series are correctly judged to be complex-valued. However, this is not surprising, since any linear complex-valued system has a bivariate equivalent, though not vice versa. Consequently, the

iAAFT method can represent these time series equally well as the CiAAFT method. For the NLC set, the proposed test correctly judges the time series to be complex-valued in 62/100 cases (the performance increased to 79/100 with $\gamma = 1.0$), and in each of the Ikeda Map realisations.

Furthermore, the robustness of the statistical test with respect to additive noise is analysed. We have considered the Ikeda Map Eq. (1), and have added Gaussian white noise to the real and imaginary parts separately. The standard deviation is set proportional to the standard deviations of the separate components, with a scaling factor of γ_n . For each scaling factor, 100 realisations of $N = 1000$ samples have been generated. Figure 2(B) shows that the ratio of correct rejections remains 100% for $\gamma_n < 0.5$ (dotted line), and decreases for higher levels of noise. The noisy Ikeda signal corresponding to $\gamma_n = 0.5$ is shown in Fig. 2(C), illustrating that the structure of the Ikeda Map can no longer be visually observed at this noise level.

5.2. Radar data

We have further considered real-world data taken from in-phase and quadrature components from the Dartmouth 1993 IPIX radar data, which is publicly available. We have selected data sets where the radar operates in stare mode, and there is a target present, namely a spherical block of styrofoam with a diameter of one meter, wrapped with wire mesh. For more details, we refer to <http://soma.crl.mcmaster.ca/ipix>. Each data set consisted of 14 rangebins containing signal of 131,072 complex-valued time samples. The rangebins in which there is no target present, contained only so-called ‘sea clutter’, i.e., radar backscatter from the ocean surface. The data sets are the following:

- 25,26** recorded during a higher sea state, with the waves moving toward the radar. Target was present in rangebin 7 (to a lesser degree in 6 and 8);
- 30,40** recorded during a lower sea state, with the waves moving toward the radar. Target was present in rangebin 7 (6,8) and 7 (5,6,8), resp.;
- 17,18** recorded during a higher sea state, with the waves moving away from the radar. Target was present in rangebin 9 (8,10,11);
- 54,310** recorded during a lower sea state, with the waves moving away from the radar. Target was present in rangebin 8 (7,9,10) and 7 (6,8,9), resp.

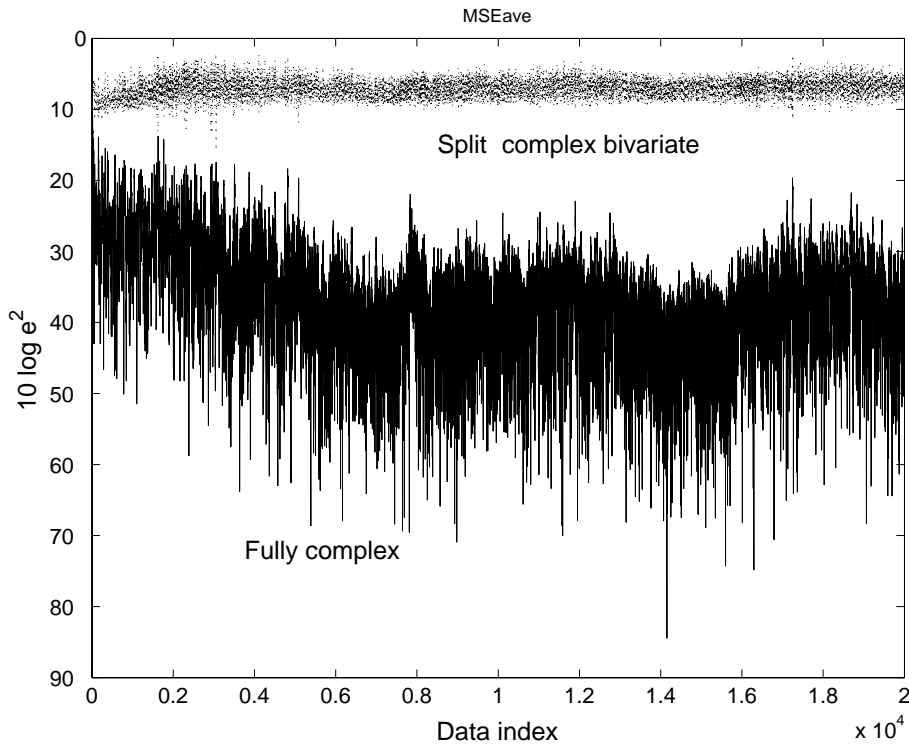


Fig. 3. Fully-complex *versus* split-complex bivariate prediction for radar data.

From every bin, we have considered time segments of $N = 1000$ samples (one second), and have generated 100 non-overlapping time segments (uniformly spaced). The proposed test has been applied on each of these segments.

To gain a first insight into the nature of these data, we performed nonlinear prediction, using a split-complex bivariate and a fully-complex approach [3]. The filter was a simple dynamical perceptron (nonlinear finite impulse response filter of the order of ten), with a sigmoid activation function. The performance criterion was the averaged squared instantaneous prediction error. The results of simulations are shown in Fig. 3 for one range-bin of data set 54. Clearly, using the fully-complex approach provided significantly better performance, than using the bivariate one. The average prediction error for the fully-complex approach was -40 dB whereas for the bivariate approach it was -8 dB.

However, although useful, the approach based upon adaptive prediction can give only a general idea about the complex-valued nature of the signal. To obtain further insight into the nature of the data and provide statistical criteria, we employ the approach based upon surrogate data and DVV analysis, as proposed above.

The number of time series that were judged to be complex-valued, are shown in Fig. 4 for each data set

and rangebin. Overall, there were stronger indications of a complex-valued nature in those bins in which a target was present. There seemed to be only weak evidence of a complex-valued nature for data from the lower sea state with waves moving away from the radar (bottom right panel). The difference between the other sea states and wave directions was not consistent over the data sets under consideration.

6. Conclusions

We have introduced a novel methodology for statistically testing whether or not the processing of a bivariate time series could benefit from a complex-valued representation. A novel algorithm, the Complex iterative Amplitude Adjusted Fourier Transform (CiAAFT) method, has been proposed for generating surrogate time series under the null hypothesis of a linear and complex-valued system underlying the time series. Consequently, surrogates generated using the traditional iAAFT method for bivariate time series can be compared to those generated using the CiAAFT method. Both types of surrogates have been characterised using a complex-valued extension of the Delay

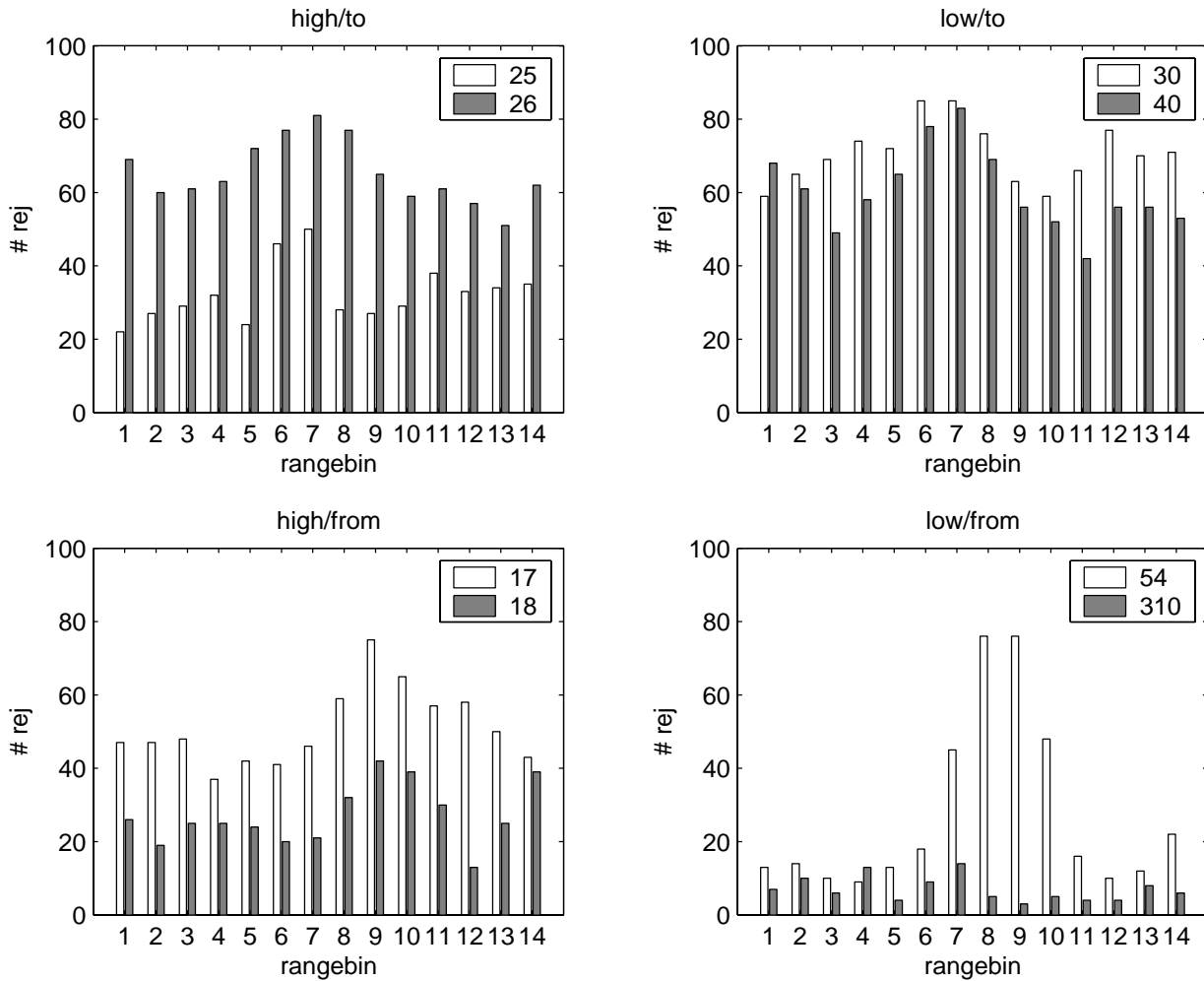


Fig. 4. Number of time series that were judged complex-valued by the proposed method, for every rangebin in the IPIX radar data sets. The results are grouped per sea state and wave direction.

Vector Variance (DVV) method, allowing for a statistical comparison between the two types of surrogates. If the difference is significant, the time series is judged complex-valued, and it is judged bivariate otherwise.

The methodology was validated on synthetically generated time series, and applied to real-world data obtained from the IPIX radar. This type of data has been frequently addressed in the open literature (for an overview, see [10]), and it has been shown that short time segments can be modelled adequately by a complex-valued autoregressive (AR) model. It has been demonstrated using the proposed methodology that there were indications of a complex-valued nature in most radar data sets, and that this proportion increased in the presence of a target. The data sets corresponding to a lower sea state with the waves moving

away from the radar showed only weak indications of a complex-valued nature in the absence of a target.

Acknowledgements

We would like to thank Mo Chen and Su Lee Goh, from Imperial College for their help toward the final version of this manuscript.

T. Gautama and M.M. Van Hulle are supported by research grants received from the Belgian Fund for Scientific Research – Flanders (G.0248.03 and G.0234.04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), and the European Commission, 5th framework programme (QLG3-CT-2000-30161 and IST-2001-32114) and 6th framework programme (IST-001917).

D.P. Mandic is supported by the Royal Society, U.K. (Grant No: 574006.G503/24543/SM).

M.M. Van Hull is further supported by a Royal Society, U.K., International Exchange Grant (Visit no 16513).

References

- [1] H. Leung and S. Haykin, The complex backpropagation algorithm, *IEEE Trans. Signal Processing* **39** (1991), 2101–2104.
- [2] T. Nitta, An analysis of the fundamental structure of complex-valued neurons, *Neural Processing Letters* **12** (2000), 239–246.
- [3] T. Kim and T. Adali, *Universal approximation of fully complex feed-forward neural networks*, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP), 2002.
- [4] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J. Farmer, Testing for nonlinearity in time series: The method of surrogate data, *Physica D* **58** (1992), 77–94.
- [5] T. Schreiber and A. Schmitz, Improved surrogate data for nonlinearity tests, *Phys. Rev. Lett.* **77** (1996), 635–638.
- [6] T. Schreiber and A. Schmitz, Surrogate time series, *Physica D* **142** (2000), 346–382.
- [7] T. Gautama, D. Mandic and M. Van Hulle, Indications of nonlinear structures in brain electrical activity, *Phys. Rev. E* **67** (2003), 046204.
- [8] T. Gautama, D. Mandic and M. Van Hulle, The delay vector variance method for detecting determinism and nonlinearity in time series, *Physica D* **190** (2004), 167–176.
- [9] K. Narendra and K. Parthasarathy, Identification and control of dynamical systems using neural networks, *IEEE Trans. Neural Networks* **1** (1990), 4–27.
- [10] S. Haykin, R. Bakker and B. Currie, *Uncovering nonlinear dynamics: The case study of sea clutter*, in: Proc. of IEEE, Special Issue, 2002, submitted.