The surrogate data test for nonlinearity has been widely used in real applications in order to establish statistically the existence of nonlinear dynamics and justify the use of nonlinear tools in the analysis of time series. The most general null hypothesis $H_0$ of this test so far is that the examined time series $x = [x_1, \ldots, x_n]$ is a realisation of a normal (linear) process $s = [s_1, \ldots, s_n]$ undergoing a possibly nonlinear static transform $h$, i.e., $x_i = h(s_i), \ i = 1, \ldots, n$. To test the null hypothesis, an estimate $q$ from a nonlinear method applied to the original data, $q_0$, is compared to the estimates, $q_1, \ldots, q_M$ on $M$ surrogate time series representing the null hypothesis. A properly designed surrogate time series should possess the same autocorrelation as the original data, even after the use of nonlinear tools in the analysis of time series. The key feature for the successful implementation of the surrogate data test for nonlinearity on a scalar time series is the generation of surrogate data that represent exactly the null hypothesis (statically transformed normal stochastic process), i.e., they possess the sample autocorrelation and amplitude distribution of the given data. A new conceptual approach and algorithm for the generation of surrogate data is proposed, called the statistically transformed autoregressive process (STAP). It identifies a normal autoregressive process and a monotonic static transform, so that the transformed realisations of this process fulfill exactly both conditions and do not suffer from bias in autocorrelation as the surrogate data generated by other algorithms. The appropriateness of STAP is demonstrated with simulated and real world data.

PACS numbers: 05.45.-a, 05.45.Tp, 05.10.Ln
have an analytic form because \( \Phi_X \) is not known analytically, but it can be approximated by an analytic function, e.g. a polynomial \( p_m \) of degree \( m \),

\[
Z_i = g(U_i) \simeq p_m(U_i) = a_0 + \sum_{j=1}^{m} a_j U_i^j, \quad (2)
\]

Low degree polynomials have been used to approximate such transforms \([4],[13]\). Then using the definition for the autocorrelation, the approximate expression for \( \phi \) reads

\[
\rho_X = \phi(\rho_U) = \frac{\sum_{s=1}^{m} \sum_{j=1}^{m} a_s a_t (\mu_{s,t} - \mu_s \mu_t)}{\sum_{s=1}^{m} \sum_{j=1}^{m} a_s a_t (\mu_{s,t} + \mu_s \mu_t)}, \quad (3)
\]

where an arbitrary lag \( \tau \) is implied as argument for the autocorrelations, \( \mu_s \) is the \( s \)th central moment of \( U_i \) being \( \mu_{2k+1} = 0, \mu_{2k} = 1 \cdot 3 \cdots (2k-1), k \geq 0 \), and \( \mu_{s,t} \) is the \( s \)-th \( t \)-th central joint moment of the bivariate standard normal distribution of \( (U_i, U_i + \gamma) \), defined as follows \([16]\)

\[
\mu_{2k,2l} = \frac{(2k!)(2l)!}{2^{k+l}} \sum_{j=0}^{\nu} \left( \frac{\rho_U}{3j+1}\right)^{2j} = \frac{(2k+1)! (2l+1)!}{2^{k+l+1}} \sum_{j=0}^{\nu} \left( \frac{\rho_U}{(2j+1)}\right)^{2j+1}, \quad (4)
\]

where \( \nu = \min(k,l) \). By substituting the expression for the moments in eq.(3), the expression for \( \phi \) can be brought to a polynomial form of the same order \( m \)

\[
\rho_X = \phi(\rho_U) = \sum_{j=1}^{m} c_j \rho_U^j, \quad (4)
\]

where the vector of coefficients \( c = [c_1, \ldots, c_m]^T \) is expressed only in terms of \( a = [a_1, \ldots, a_m]^T \) (the expressions are rather involved and therefore not presented here). Simpler expressions can be derived using the Tchebycheff-Hermite polynomials \([7]\). Thus an analytic expression for \( \rho_U \) is possible if eq.(3) can be solved with respect to \( \rho_U \). From our simulations, we conjecture that if \( g \) is monotonic then \( \phi \) is also monotonic in \([-1,1]\). Then \( \phi^{-1} \) exists and a unique solution for \( \rho_U \) can be found from eq.(4). The proper standard normal process \( U \) is completely defined by \( \rho_U \) and applying the transform \( g \) of eq.(1) to the components of \( U \), the “linear copy” \( Z \) of the given process \( X \) is constructed.

Note that the solution for \( \rho_U \) is given analytically from the polynomial approximation of \( g \) and it requires only the knowledge of the coefficients \( a \) of the polynomial and the autocorrelation \( \rho_X \).

In practice, we operate with a single time series \( x \) rather than a process \( X \) and with the sample estimates \( F_z \) and \( r_z \) for \( \Phi_X \) and \( \rho_X \), respectively. The steps of the algorithm are as follows:

1. Estimate the vector of coefficients \( a = [a_1, \ldots, a_m]^T \) of the polynomial \( p_m \) from the graph of \( x_t = F_z^{-1}(F_0(w_t)) \), i.e. the graph of \( x \) vs \( w \) after their ranks are matched, where \( w = [w_1, \ldots, w_n]^T \) is standard white normal noise.

2. Compute \( c = [c_1, \ldots, c_m]^T \) for the given \( a \) from eqs.(3,4).

3. Find \( r_{+1} \) from eq.(4) for the given \( c \) and \( r_z \) using the sample estimates \( r \) instead of \( \rho \). The common practice is that the solution exists and it is unique. If this is not the case, repeat the steps 1–3 for a new \( w \) until a unique solution is obtained.

4. Generate a realisation \( u \) of a standard normal process with autocorrelation \( r_u \). We choose to do this simply by means of an autoregressive model of some order \( p \), AR\( (p) \). The parameters \( b = [b_0, b_1, \ldots, b_p]^T \) of AR\( (p) \) are found from \( r_u \) using the normal equations solved effectively by the Levinson algorithm \([18]\). The AR\( (p) \) model is run to generate \( u \)

\[
u_{i+1} = b_0 + \sum_{j=1}^{p} b_j u_{i-(j-1)} + c_i, \quad c_i \sim N(0,1).
\]

5. Transform \( u \) to \( z \) by reordering \( x \) to match the rank order of \( u \), i.e. \( z_i = F_{-1}^{-1}(F_0(u_i)) \).

Note that if \( u \) possesses the sample normal marginal cdf \( F_0 \) and the proper \( r_u \), so that \( z \) possesses \( F_z = F_x \), \( r_z = r_x \), and is otherwise random, as desired. In practice however, the equality \( r_z = r_x \) is not exact and \( r_x \) may vary substantially around \( r_z \). Two possible reasons for this are the insufficient approximation of \( g \) in step 1 and the inevitable variation of the sample autocorrelation of the generated \( u \) in step 4, which decreases with the increase of data size. The former is due to the limited power of polynomials in approximating monotonic functions and this shortcoming causes also occasional repetitions of the first steps of the algorithm as stated in step 3 \([21]\). The latter constitutes an inherent property of the so-called “typical realisation” approach (i.e. a model is used to generate the surrogate data) and cannot be controlled. However, less variation in the autocorrelation is achieved when the AR\( (p) \) model is optimised making the following steps, in the same way as for the CAAFT algorithm \([10]\):

1. Apply the algorithm presented above \( K \) times and get \( z^1, \ldots, z^K \) surrogate time series.

2. Compute \( r_{z1}, \ldots, r_{zK} \) and find the one, \( r_{z1} \) closest to \( r_x \) \([21]\).

3. Use the parameters \( b \) of the \( l \)-repetition to generate the \( M \) surrogate data (steps 4–5 of the algorithm above).

The \( K \) repetitions above as well as the occasional repetitions of steps 1–3 of the first part of the algorithm may slow down the algorithm if the time series is long, but they have no impact on the principal function of the algorithm. Simply, some realisations of white noise \( w \) are discarded in the search of the parameters \( b \) of the most suitable AR model that generates the surrogate data (through the \( g \) transform).

The free parameters of the STAP algorithm are the degree \( m \) of the polynomial approximation of \( g \), the order \( p \) of the AR model, the number \( K \) of repetitions for the optimisation of AR\( (p) \) and the maximum lag \( r_{\text{max}} \), used to compare \( r_{z1}, \ldots, r_{zK} \) to \( r_x \). Usually, a small \( m \)
(m ≤ 10) is sufficient. For p, there is no optimal range of values but it may vary with the shape of rᵢ, e.g. a slowly decaying rᵢ may be better modelled by a larger p. In all our simulations, we set K = M = 40 and τₘₐₓ = p.

The proper performance and superiority of STAP over AAFT and IAAFT was confirmed from simulations on different toy models. CAAFT was found to perform very similarly to STAP. We show in Fig. 1 comparative results for AAFT, IAAFT and STAP for three representative synthetic systems: the cube of an AR(1) process (sᵢ₊₁ = 0.3 + 0.8sᵢ + eᵢ, eᵢ ∼ N(0, 1), xᵢ = sᵢ), the square of the same AR(1) (xᵢ = sᵢ²), both being consistent with H₀, and the x variable of the Rössler system [19], not consistent with H₀. For each system, we generate 100 time series of 2048 samples each and for each realisation we generate M = 40 surrogate data of each type.

As discriminating statistics qᵢ, we choose the correlation coefficient (CC) of the fit with the series of Volterra polynomials of degree 2 and order v = 5. The polynomials for the first i terms, where i = 1, . . . , 5, are linear and for terms i = 6, . . . , 20 are nonlinear (see also [13]). To quantify the discrimination we use the significance Sᵢ = |qᵢ − ⟨qᵢ⟩|/σᵢ for each polynomial of i terms, where qᵢ is the statistic qᵢ on the original data, ⟨qᵢ⟩ and σᵢ are the mean and standard deviation of the statistic qᵢ on the M surrogate data. The null hypothesis H₀ is formally rejected at the 0.05 significance level when S > 1.96, under the assumption of normality for the statistic qᵢ, which turns out to hold in general. The percentages of rejections for each of the three systems are shown in Fig. 1. Very similar results were found when deciding the rejection from the rank ordering of q₀, q₁, . . . , q₅M.

For the linear statistics, STAP gives consistently and correctly no rejections, i.e. unbiased match of the linear correlations, whereas AAFT and IAAFT give very large percentages of rejections for all but the first case where AAFT gives about 5% rejections, as expected. For IAAFT, the rejections occur because σᵢ is very small (10 to 20 times smaller than for AAFT and STAP), though the bias q₀ − ⟨qᵢ⟩ is smaller than for STAP.

![FIG. 1: Percentages of rejections of H₀ using as discriminating statistics the fit of Volterra polynomials from 100 realisations for each of the three cases in the three panels, as indicated. Three types of surrogates are used in each test as shown in the legend (for STAP m = 5, p = 5). The vertical gray line distinguishes the linear from the nonlinear statistics.](image)

For the nonlinear statistics, the feature of the linear statistics persists for the two first systems (consistent with H₀) and for all three algorithms, i.e. correctly no rejections with STAP and erroneous rejections with AAFT and IAAFT. We cannot explain why the polynomial fit for the IAAFT surrogates improves with the addition of nonlinear terms (see Fig. 1). For the nonlinear system, STAP properly converges with the addition of few non-

![FIG. 2: (a) The correlation coefficient of the fit with Volterra polynomials on the original normal EEG data (black line) and 40 AAFT, IAAFT and STAP surrogates (gray lines) in the three panels, as indicated (for STAP m = 10, p = 10). (b) The same as in (a) but for the epileptic EEG. (c) The significance for the fits in (a) (upper panel) and in (b) (lower panel). The vertical lines in the plots distinguish the linear from the nonlinear polynomials.](image)
linear terms to the correct 100% rejection level, which AAFT and IAAFT possessed already with linear statistics.

Next, we verify the three algorithms on two human EEG data sets, one recorded many hours before an epileptic seizure accounting for normal brain activity and another recorded during an epileptic seizure. The epileptic EEG seems to exhibit a pattern of oscillations indicating a more deterministic character than the non-epileptic EEG, and this constitutes a well established result in physiology. This is demonstrated also with the fit of Volterra polynomials in Fig. 2b. The fit improves with the inclusion of the first nonlinear terms for the epileptic EEG but not for the non-epileptic EEG. These findings are confirmed statistically by the test with STAP surrogate data while the results of the test with AAFT and IAAFT are more or less confusing.

In particular, the \( H_0 \) on the normal EEG is erroneously rejected with AAFT because the difference in the fit between original and surrogate data is about the same for the linear and nonlinear polynomials (see Fig. 2a and c). The same test result is obtained with IAAFT for large nonlinear polynomials, whereas again there is significant difference in the linear fits between original and IAAFT surrogates (not easily discernible as both bias and variance are very small). Using STAP surrogates, the \( H_0 \) is not rejected for both linear and nonlinear fits.

For the epileptic EEG, there is again a clear difference in the linear fit between original data and AAFT surrogates and a smaller but equally significant difference between original data and IAAFT surrogates (see Fig. 2b and c). The significance \( S \) for both AAFT and IAAFT increases with the addition of the first couple of nonlinear terms, much more for IAAFT due to the small variance of CC. However, the deviation in the linear fit does not support reliable rejection of \( H_0 \). On the other hand, using STAP surrogates, the \( H_0 \) is properly rejected only for the nonlinear statistics and with high confidence (\( S < 2 \) for the linear fit and \( S \approx 5 \) for the nonlinear fit).

In general, the test with STAP surrogates tends to be more conservative, i.e. small discriminations are found less significant, as the data size decreases. For example, the test on 296 sunspot samples (for which a small leap of the polynomial fit with the addition of nonlinear terms was observed) gave rejection of \( H_0 \) for AAFT and IAAFT but not for STAP (not shown here, see also [11]). However, this should not be considered as a drawback of the STAP algorithm, as one expects that the power of the test reduces with the decrease of data size.

A new algorithm that generates surrogate data for the test for nonlinearity has been presented, called statically transformed autoregressive process (STAP). The key feature of STAP is that it finds analytically the autocorrelation of an appropriate underlying normal process for the test. This is the main difference of the STAP algorithm from the corrected AAFT (CAAFT) algorithm, where the autocorrelation is estimated numerically. Both CAAFT and STAP algorithms do not suffer from the severe drawback of the AAFT algorithm, i.e. bias in the match of the original autocorrelation. The AFT algorithm is essentially impractical for real applications because it favours the rejection of \( H_0 \) as a result of the bias in the autocorrelation. From the numerical simulations, it turns out that the IAAFT algorithm may also give small bias in the linear correlations, favouring also the rejection of \( H_0 \). On the other hand, the STAP algorithm performs properly and gives reliable rejections of \( H_0 \), only whenever this appears to be the case.

---

[1] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997).

[2] C. Diks, *Nonlinear Time Series Analysis: Methods and Applications* (World Scientific, 2000).

[3] T. Schreiber and A. Schmitz, Physica D **142**, 346 (2000).

[4] D. Kugiumtzis, in *Modelling and Forecasting Financial Data, Techniques of Nonlinear Dynamics*, edited by A. Soofi and C. Cao (Kwyer Academic Publishers, 2002), chap. 12.

[5] D. Kugiumtzis, Internat. J. Bifur. Chaos Appl. Sci. Engrg. **11**, 1881 (2001).

[6] J. Theiler, S. Eubank, A. Longtin, and B. Galdrikian, Physica D **58**, 77 (1992).

[7] D. Kugiumtzis, Phys. Rev. E **60**, 2808 (1999).

[8] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).

[9] T. Schreiber, Phys. Rev. Lett. **80**, 2105 (1998).

[10] D. Kugiumtzis, Phys. Rev. E **62**, 25 (2000).

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes* (McGraw-Hill, Inc., 1991), p. 113, 3rd ed.

[12] N. Johnson and S. Kotz, *Distributions in Statistics, Continuous Multivariate Distributions* (John Wiley and Sons, 1972), chap. 34, p. 15, Wiley Series in Probability and Mathematical Statistics.

[13] A. Der Kiureghian and P.-L. Liu, J. Eng. Mech. **112**, 85 (1986).

[14] C. D. Elphinstone, Comm. Statist. Theory Methods **12**, 161 (1983).

[15] D. M. Hawkins, Comput. Statist. **9**, 233 (1994).

[16] T. P. Hutchinson and C. D. Lai, *Continuous Bivariate Distributions Emphasising Applications* (Rumby Scientific Publishing, 1990).

[17] N. M. Bhatt and P. H. Dave, J. Indian Statist. Assoc. **2**, 177 (1964).

[18] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer Series in Statistics (Springer-Verlag, New York, 1991).

[19] O. E. Rössler, Physics Letters A **57**, 397 (1976).

[20] Other model classes may give better approximations of \( g \) and should be investigated, but polynomials were used here to derive a simple analytic form for the transform of the autocorrelations.

[21] We found that a robust way to achieve this is to compute the error \( \sum_{i=1}^{K} (r_g(i) - r_g(i))^2 \) for \( j = 1, \ldots, K \) and \( \tau = 1, \ldots, \tau_{\text{max}} \) and then select the trial \( l \) that gives the minimum error most times, where \( \tau_{\text{max}} \) minimums are totally computed, each time over the \( K \) trials.