

# Testing for Phase Synchronization

Carsten Allefeld      Jürgen Kurths

Nonlinear Dynamics Group, Institute of Physics, and Research Group  
“Conflicting Rules”, Universität Potsdam, Am Neuen Palais 19, PF 601553,  
D-14415 Potsdam, Germany

allefeld@ling.uni-potsdam.de

**Reference:** *International Journal of Bifurcation and Chaos*, **14**(2), 405–416, 2004.

## Abstract

We present different tests for phase synchronization which improve the procedures currently used in the literature. This is accomplished by using a two-samples test setup and by utilizing insights and methods from directional statistics and bootstrap theory. The tests differ in the generality of the situation in which they can be applied as well as in their complexity, including computational cost. A modification of the resampling technique of the bootstrap is introduced, making it possible to fully utilize data from time series.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	The measure of phase synchronization . . . . .	3
1.2	The two-samples test setup . . . . .	4
1.3	Accuracy and power of the tests . . . . .	4
<b>2</b>	<b>Parametric Tests</b>	<b>5</b>
2.1	Distributions of directional statistics . . . . .	5
2.2	The variance-stabilizing transformation test . . . . .	7
2.3	Comparison with simulated phase synchronization distributions . . . . .	8
<b>3</b>	<b>A Simple Nonparametric Test</b>	<b>8</b>
<b>4</b>	<b>Bootstrap Techniques</b>	<b>10</b>
4.1	Bootstrap $t$ -test . . . . .	10
4.2	Bootstrap $H_0$ simulation . . . . .	11
4.3	The permutation test . . . . .	11
<b>5</b>	<b>Data from Time Series</b>	<b>12</b>
<b>6</b>	<b>Conclusion</b>	<b>13</b>

# 1 Introduction

Synchronization of coupled or forced chaotic systems has received much attention recently (cf. Pikovsky et al. [2001] and Boccaletti et al. [2002] for a review and references therein). Four basic types of such a complex behavior have been found: complete [Pecora & Carroll, 1990; Fujisaka & Yamada, 1983; Afraimovich et al., 1986; Pikovsky, 1984], generalized [Rulkov et al., 1995], phase [Rosenblum et al., 1996], and lag synchronization [Rosenblum et al., 1997]. Phase synchronization has found various applications in laboratory experiments [Parlitz et al., 1996; Taherion & Lai, 2000; Allaria et al., 2001; DeShazer et al., 2001; Ticos et al., 2000; Kiss & Hudson, 2001] as well as in natural systems, such as cardiorespiratory interaction [Schäfer et al., 1998; Stefanovska et al., 2000; Anishchenko et al., 2000], brain activity of Parkinsonian patients [Tass et al., 1998], EEG measurements from patients with temporal lobe epilepsy [Mormann et al., 2000] and early seizure detection [Jerger et al., 2001], ecology [Blasius et al., 1999], and climate systems [Lunkeit, 2001]. A technique to reconstruct the synchronization diagram from bivariate data has been proposed by Tokuda et al. [2002].

In noisy systems or where the acquisition of empirical data includes a large amount of noise [Tass et al., 1998], as well as in chaotic systems on the borderline of synchronization or with imperfect phase synchronization [Zaks et al., 1999], the basic concept of phase synchronization as boundedness of the phase difference of two oscillators has to be replaced by a quantitative statistical definition. In this perspective, phase synchronization is a gradual phenomenon whose strength corresponds to the “peakedness” of the distribution of the phase difference. Since such a statistical measure is in itself a random variable, a deviation of its empirical value from zero (corresponding to no synchronization) doesn’t suffice to state an effect of phase synchronization and this means that a statistical test has to be applied.

There are some approaches in the literature to test for the significance of a certain synchronization level, but they are mainly based on surrogate data and the corresponding null hypothesis is that the oscillators are unsynchronized [Bhattacharya et al., 2001; Lachaux et al., 1999; Rodriguez et al., 1999; Mormann et al., 2000; Tass et al., 1998; Paluš, 1997; Paluš & Hoyer, 1998].<sup>1</sup> In practice, this null hypothesis often proves to be too weak, because there is a base level of synchronization which is always present in the system and therefore the test is not able to distinguish between different system states.

In this paper, we present a number of tests improving this situation in three respects. 1) We use a measure known from the context of directional statistics [Mardia & Jupp, 2000; Mardia, 1972] and utilize its simple and defined statistical properties. 2) We use a two-sample approach and test against the null hypothesis that the synchronization strength is the same in the two samples. 3) We use testing procedures based on proven statistical theory; in the nonparametric case, they rely mainly on bootstrap techniques [Efron & Tibshirani, 1993]. The different tests are presented in a sequence of increasing generality and precision, but also increasing computational complexity; they are graded with respect to their applicability and usefulness in a given situation and their correctness is checked in numerical simulations.

---

<sup>1</sup>Furthermore, for some of the standard procedures of surrogate data generation like phase randomization, their statistic validity in the context of phase synchronization is doubtful.

## 1.1 The measure of phase synchronization

The system we are concerned with consists of two autonomous oscillators, A and B. For each of them a phase  $\varphi_A, \varphi_B$  is defined; the details of this definition are not important, but it should be done in a way that there is no preferred value of the phase. The oscillators are bi- or unidirectionally coupled such that depending on the strength of coupling, there is a more or less strong dynamical dependence between the phases which leads to a preferred value of the phase difference  $\theta = \varphi_B - \varphi_A$ .<sup>2</sup> If the process of phase synchronization is regarded as a stochastic one (in the deterministic case the randomness belongs to the initial conditions), the phase difference is a random variable characterized by a probability distribution, which is empirically accessible in the form of a number of realizations  $\theta_j$ . The strength of phase synchronization which corresponds to the peakedness of the phase difference distribution may be quantified in different ways (cf. Tass et al. [1998]); in this paper, we follow the approach of directional statistics.

Directional statistics [Mardia, 1972; Mardia & Jupp, 2000] deals with a special case of random variable which is defined on a circular scale, such that values whose difference is an integral multiple of a certain period (in general  $2\pi$ ) are regarded the same, and for convenience all values are wrapped into a single period. The phase difference is an example of such a circular random variable  $\theta$ ; another example are directions in space, which are a common object of investigations in science, for instance the directions of the flight of birds in biology or that of geologic formations. To a circular random variable, standard (linear) statistical measures and moments like mean and variance are not applicable, because they yield different values if the period is added to or subtracted from some values, though the physical meaning of these changed values is the same.

Instead of these classical moments, directional statistics studies the statistical properties of the values of trigonometric functions applied to the circular variable  $\theta$ , the trigonometric moments of order  $p$  ( $p = 1, 2, \dots$ ) of its distribution

$$\alpha_p = \langle \cos p \theta \rangle \quad \text{and} \quad \beta_p = \langle \sin p \theta \rangle. \quad (1)$$

The combined complex form of cosine and sine moments  $\chi_p = \alpha_p + i\beta_p$  is equivalent to the characteristic function of the distribution (its Fourier transform). Its first element is of special importance, because the polar components  $\mu = \arg \chi_1$  and  $\rho = |\chi_1|$  can be regarded as indices of the mean direction of the distribution and of its concentration onto this mean direction.

The corresponding quantities on a sample  $\theta_j$  ( $j = 1 \dots n$ ) are  $a_p = \frac{1}{n} \sum \cos p \theta_j$  and  $b_p = \frac{1}{n} \sum \sin p \theta_j$ , which are estimators of the moments  $\alpha_p$  and  $\beta_p$ . The components of the first empirical moment are

$$\bar{C} = \frac{1}{n} \sum \cos \theta_j \quad \text{and} \quad \bar{S} = \frac{1}{n} \sum \sin \theta_j \quad (2)$$

or, in polar representation,

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2} \quad \text{and} \quad \bar{\theta} = \arctan \frac{\bar{S}}{\bar{C}}. \quad (3)$$

Since  $\bar{R}$  is the length of the mean of the unit vectors corresponding to the sample values, it is called mean resultant length. It is an estimator of the population moment  $\rho$  and as such an empirical measure of the concentration of the underlying distribution. Applied to a distribution of phase differences  $\theta = \varphi_B - \varphi_A$ , we use it as a measure of synchronization strength.  $\bar{R}$  takes on values in the range from 0 to 1, describing a continuum between no and perfect phase synchronization.

<sup>2</sup>Or in the general case of  $m : n$  synchronization, of the generalized phase difference  $m\varphi_B - n\varphi_A$ .

An advantage of this measure is that it is possible to derive some useful facts about its sampling distribution. Because  $\bar{C}$  and  $\bar{S}$  are sums of identically distributed contributions, due to the central limit theorem their asymptotic joint distribution is two-dimensional normal. The moments of this distribution depend on the trigonometric moments of  $\theta$ :

$$\begin{aligned} \langle \bar{C} \rangle &= \alpha_1 & \langle \bar{S} \rangle &= \beta_1 \\ n \operatorname{var}(\bar{C}) &= \frac{1}{2}(1 + \alpha_2 - 2\alpha_1^2) & n \operatorname{var}(\bar{S}) &= \frac{1}{2}(1 - \alpha_2 - 2\beta_1^2) \\ n \operatorname{cov}(\bar{C}, \bar{S}) &= \frac{1}{2}(\beta_2 - 2\alpha_1\beta_1) \end{aligned} \quad (4)$$

For  $\rho > 0$ , the transformation from  $(\bar{C}, \bar{S})$  to  $(\bar{R}, \bar{\theta})$  is locally linear, and so  $\bar{R}$  is asymptotically normally distributed, too.

## 1.2 The two-samples test setup

$\bar{R}$  is a measure of phase synchronization based on a sample of size  $n$  of the phase difference  $\theta$  of the coupled oscillators. It reflects the synchronization strength in a certain state of the system of two oscillators. Since only in special cases it makes sense to compare against a state of unsynchronized behavior, it is necessary to obtain a reference level of synchronization. This is given by a second sample of the phase difference from another state of the system of two oscillators. For each of the samples  $\theta_{1,j}, \theta_{2,j}$  the value of the measure  $\bar{R}_1, \bar{R}_2$  is calculated, and the question to be answered by the test is whether the difference of these two values is significant.

The corresponding null hypothesis is that the population values of the synchronization measure are equal:  $H_0 : \rho_1 = \rho_2$ . The test decides if this hypothesis is to be rejected or accepted based on the given information (the samples), and the important part of the test design is to ensure that the probability for erroneous rejection of the null hypothesis, the so-called error of the first kind, is equal to (or below) a chosen value, the significance level of the test.

In the following, several different tests are presented. In the given form, all of them assume that the size of the two samples is the same, but they are easily generalized to different sample sizes  $n_1$  and  $n_2$ . Initially, all of them assume that the samples consist of independent values, that is they are obtained from independent realizations of the process of phase synchronization. In this way the number of samples is also the number of degrees of freedom inherent in the data (in total  $2n$ ). Since phase synchronization is a dynamical process in time this is a strong constraint, and so in the last part of the paper some of the tests are generalized to the case of partly dependent samples, so that the data from the underlying time series can be fully utilized.

## 1.3 Accuracy and power of the tests

In several respects, the quality of the different tests going to be presented in the following sections will be judged on theoretical grounds. But in most cases, the performance of a test cannot be theoretically deduced. Therefore, the tests have been checked in a numerical simulation using random numbers generated according to the wrapped normal distribution with the specified  $\rho$ .<sup>3</sup> The essential parameter to be calculated is the probability for the rejection of the null hypothesis. If the null

<sup>3</sup>For a definition of the wrapped normal distribution, see below Sec. 2.1. It is important to note that except for the parametric test there is nothing in the derivation of the test procedures that relies on the specific properties of this distribution. For the numerical simulation it has been necessary to choose a specific distribution family, and the wrapped normal has been chosen because random numbers are easily obtained and because it can be considered as a typical circular random distribution. Calculations based on the von Mises distribution produced equivalent results.

hypothesis  $\rho_1 = \rho_2$  is true, this probability is called error of the first kind, which has to be equal to or at least smaller than the significance level. For the general case in which both  $\rho$ s are arbitrarily chosen it is called the power function, because the power of the test to recognize a situation not conforming to the null hypothesis corresponds to this probability for  $\rho_1 \neq \rho_2$ .

Figure 1 presents the results of the simulations. Two samples of size  $n = 100$  from the wrapped normal distribution for the given values of  $\rho_1$  and  $\rho_2$  have been generated, the respective test for a significance level of 5% has been performed and the relative frequency of rejection as an estimator of the probability has been calculated in 4,000 repetitions. Panel a) shows the rejection probability depending on  $\rho$  with  $\rho_1 = \rho_2 = \rho$  (error of the first kind), panel b) depending on  $\rho_1$  with  $\rho_2 = 0$ , and panel c) the same for  $\rho_2 = 0.4$  and  $\rho_2 = 0.8$  (three sections through the two-dimensional power function). Since the scale is much smaller in panel a), the randomness of the probability estimation appears stronger here; the horizontal black lines mark the mean  $\pm$  s.d. to be expected from a binomial distribution.

In Fig. 1 a) it can be seen that all of the tests are valid for all values of  $\rho$  in the sense that the chosen significance level is not exceeded. On the other hand, four of the tests show errors that in the vicinity of  $\rho = 0$  lie strongly below the upper limit. This leads to a decreased power of these tests for testing against  $\rho_2 = 0$ , as can be seen in panel b). For higher values of  $\rho_2$ , this weakness is no longer relevant; the different lines are almost indistinguishable. What also can be seen in panel c) is that the power of the tests increases (the valley of the power function gets narrower) for those higher values. We will refer to these results in more detail in the following sections.

## 2 Parametric Tests

A direct way to obtain a test is to assume that the distribution of  $\theta$  belongs to a certain family which can be described by a small number of parameters, of which the concentration moment  $\rho$  is the most important. This is the approach taken by Mardia in his monographs on directional statistics [Mardia & Jupp, 2000; Mardia, 1972]; it is mathematically justified, but we will see that its applicability is rather limited.

### 2.1 Distributions of directional statistics

In directional statistics, instead of the normal distribution which is central to linear statistics, there are two similar but different distributions, each of which shares some of the special properties of the normal distribution. The first is the von Mises distribution  $M(\mu, \kappa)$ , defined by the probability density function

$$p_M(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}. \quad (5)$$

$I_p$  denotes the modified Bessel function of the first kind of order  $p$ .  $\mu$  specifies the mean direction of the distribution and  $\kappa > 0$  its concentration; the moments are  $\chi_p = I_p(\kappa)/I_0(\kappa) \exp(ip\mu)$  and especially  $\rho = A(\kappa) \equiv I_1(\kappa)/I_0(\kappa)$ . Main properties of this distribution in analogy to the normal are that it is the one with maximum entropy for fixed moments  $\rho$  and  $\mu$  and the one under which the sample moment  $\bar{\theta}$  is the maximum likelihood estimator of a shifting parameter like  $\mu$ . For our application it is of special importance that the stationary solution of the Fokker Planck equation of the simplest stochastic model of phase synchronization  $\dot{\varphi}_{A,B} = \omega_0 + k \sin(\varphi_{2,1} - \varphi_{A,B}) + \xi_{A,B}$  (where  $\xi_{A,B}$  is standard normal white noise) is a von Mises distribution:  $\theta \sim M(0, 4k)$ .

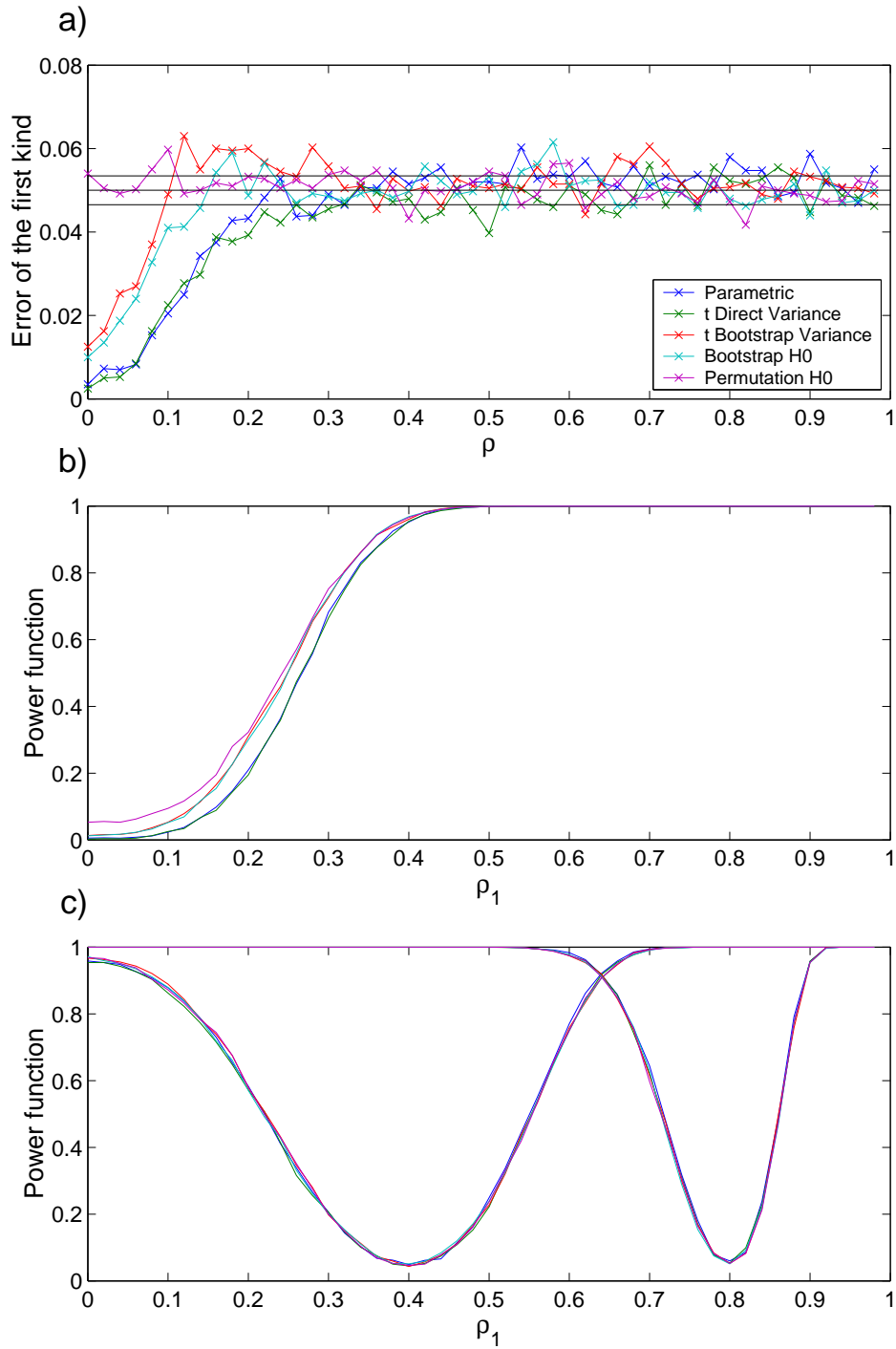


Figure 1: Investigation of the properties of the presented tests in a numerical simulation based on wrapped normally distributed samples of size  $n = 100$ . a) Empirical error of the first kind in samples conforming to the null hypothesis  $\rho_1 = \rho_2 = \rho$ . b) Power function for testing against a sample with  $\rho_2 = 0$ . c) Power functions for testing against a sample with  $\rho_2 = 0.4$  and  $\rho_2 = 0.8$ , respectively. Each probability has been estimated based on 4,000 simulations of the test for a significance level of 5%.

The other distribution corresponding to the linear normal one is the wrapped normal distribution  $W(\mu, \sigma)$  with the probability density

$$p_W(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{k=\infty} e^{-\frac{1}{2} \frac{(\theta - \mu + 2\pi k)^2}{\sigma^2}}. \quad (6)$$

$\mu$  specifies again the mean direction of the distribution and  $\sigma$  its dispersion. The moments are  $\chi_p = \rho^{p^2} \exp(ip\mu)$  with  $\rho = \exp(-\frac{1}{2}\sigma^2)$ . Main properties corresponding to those of the normal distribution are that it is additive and that there exists a central limit theorem which describes convergence to the wrapped normal distribution. The von Mises and the wrapped normal distribution can be related by equating their concentration moments  $A(\kappa) = \rho = \exp(-\frac{1}{2}\sigma^2)$ . For extreme values  $\rho \rightarrow 0$  or  $\rho \rightarrow 1$  they converge to each other and to the uniform or delta distribution, respectively; however, for intermediate values the peak of the von Mises distribution is sharper.

Since the von Mises and the wrapped normal distribution are closely related to basic statistical properties of circular random variables, there is some possibility that empirical distributions like that of the phase difference of coupled oscillators may be described sufficiently well by one of them.

## 2.2 The variance-stabilizing transformation test

We have seen that for  $\rho > 0$ ,  $\bar{R}$  is asymptotically normally distributed. Based on the assumption that  $\theta$  follows one of the two standard circular distributions, the moments of the corresponding distributions of  $\bar{R}$  can be calculated. For the von Mises distribution, they are

$$\langle \bar{R} \rangle \cong A(\kappa), \quad n \text{ var}(\bar{R}) \cong 1 - A(\kappa)^2 - \frac{A(\kappa)}{\kappa} \quad (7)$$

with  $A(\kappa)$  as above and for the wrapped normal distribution we get

$$\langle \bar{R} \rangle \cong \rho, \quad n \text{ var}(\bar{R}) \cong \frac{1}{2}(1 - \rho^2)^2 \quad (8)$$

with  $\rho$  as above, plus terms of order  $1/n$ .

Following Mardia & Jupp [2000], we apply a variance-stabilizing transformation to obtain a suitable test statistic. Such a transformation can be defined for a statistic  $Z$  on a sample of a random variable which follows a distribution with a parameter  $a$ ; then the moments of the statistic are functions of this parameter:  $\langle Z \rangle = f(a)$ ,  $n \text{ var}(Z) = g(a)$ . If the transformation  $h(x) = \int_0^x 1/\sqrt{g(f^{-1}(x))} dx$  is applied to the statistic, we get an approximately constant variance  $n \text{ var}(h(Z)) \cong 1$ . In the case of the statistic  $\bar{R}$  on a von Mises or a wrapped normal distribution, the corresponding transformations are

$$h_M(x) = \int_0^{A^{-1}(x)} \sqrt{A'(\kappa)} d\kappa \quad \text{and} \quad h_W(x) = \sqrt{2} \text{artanh } x, \quad (9)$$

respectively.<sup>4</sup> Interestingly, the latter is up to a factor identical to the variance-stabilizing transformation of the linear correlation coefficient, the well-known Fisher  $Z$  transform [Sheskin, 1997].

The result of these transformations is now used to perform a test for a significant difference of the concentration of two distributions. If we calculate the value

<sup>4</sup>The first integral can not be written in a closed form, but can be calculated numerically.

of the concentration measure on each of the two samples, then under the null hypothesis  $\rho_1 = \rho_2$  the statistic  $\sqrt{\frac{\pi}{2}}(h(\bar{R}_2) - h(\bar{R}_1))$  follows asymptotically a standard normal distribution. The hypothesis of equal concentration in the two samples has to be rejected if the modulus of this quantity exceeds a certain value, which is given by the percentiles of the normal distribution for a chosen significance level (i.e., a Gauss test is to be performed).

The test is applicable if both samples follow the supposed distribution and if the sample size is large enough so that the asymptotic approximation can be applied. An additional restriction follows from the presupposition  $\rho > 0$  in the derivation of the sampling distribution of  $\bar{R}$ ; for a finite sample size it is necessary that circa  $\rho_{1,2} > 3/\sqrt{2n}$ . For smaller values, the asymptotic distribution assumption as well as the variance estimation implicit in the test are no longer correct. In Fig. 1 (blue lines) it can be seen that this leads to a decrease of the error of the first kind below the significance level and a corresponding suboptimal power for testing against  $\rho_2 = 0$ ; still, the test seems to be valid in all cases.

### 2.3 Comparison with simulated phase synchronization distributions

As a classic example of a system exhibiting phase synchronization, we consider two coupled chaotic Rössler oscillators [Rosenblum et al., 1996],

$$\begin{aligned}\dot{x}_{A,B} &= -\omega_{A,B} y_{A,B} - z_{A,B} + \epsilon(x_{B,A} - x_{A,B}), \\ \dot{y}_{A,B} &= \omega_{A,B} x_{A,B} + a y_{A,B}, \\ \dot{z}_{A,B} &= f + z_{A,B}(x_{A,B} - c),\end{aligned}\tag{10}$$

where  $a = 0.15$ ,  $f = 0.2$ ,  $c = 10$ , with a small frequency mismatch  $\omega_{A,B} = 1 \mp 0.01$ . Phases are defined to be  $\varphi_{A,B} = \arctan y_{A,B}/x_{A,B}$  and the phase difference  $\theta = \varphi_B - \varphi_A$ .

Figure 2 shows the distribution of the phase difference in a numerical simulation, which depends on the strength of the coupling. For a small value of  $\epsilon = 0.005$ , the resulting distribution may be well described by a fitted von Mises as well as a wrapped normal distribution (a). If the coupling is increased to  $\epsilon = 0.015$ , the peak of the distribution is sharper than both of the fitted distributions (b). To make the simulation more realistic for empirical data, artificial measurement noise in phase space has been added (two-dimensional normal in the  $(x, y)$ -plane with s.d.  $\sigma = 5$ ). As a result, the distribution for the stronger coupling may again be described by a fitted von Mises distribution, but not as well by a wrapped normal distribution (c). It seems that measurement noise may increase the applicability of the standard distributions in phase synchronization systems.

As can be seen from this, there surely are synchronization phase differences which can be described by one of the standard normal distributions, but this is not always the case. Not only that the parametric test may be not applicable; to be precise, one would have to perform a separate test for goodness of fit in each case. Though the parametric approach of Mardia is elegant and simple, it is therefore necessary to look for a nonparametric means to test for phase synchronization.

## 3 A Simple Nonparametric Test

There is a classic nonparametric approach in statistics which is designed to test for a significant difference of the means of two samples. Mean values and variances are calculated from the sample and the difference of the means is divided by the standard deviation of its estimation, resulting in the  $t$  statistic. If the sample values



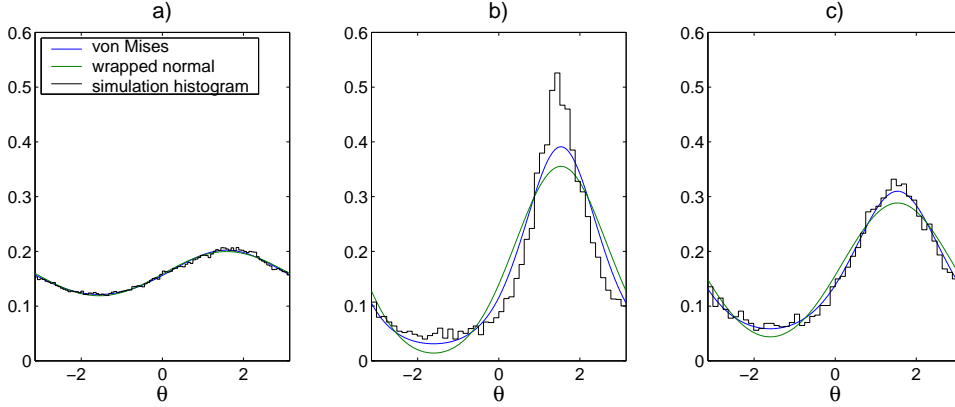


Figure 2: Histograms of the phase difference distribution of two coupled Rössler oscillators in a numerical simulation, with fitted von Mises and wrapped normal distributions. a) Weak coupling  $\epsilon = 0.005$ . b) Stronger coupling  $\epsilon = 0.015$ . c) Stronger coupling with additional measurement noise  $\sigma = 5$ .

are normally distributed,  $t$  is distributed according to a Student distribution with  $f = 2(n - 1)$  degrees of freedom [Sheskin, 1997].

Formally, it is possible to write the equation for  $\bar{R}$  in the form of a mean value,

$$\bar{R} = \frac{1}{n} \sum \cos(\theta_j - \bar{\theta}), \quad (11)$$

and correspondingly, the variance of the estimation can be directly calculated,

$$s_{\bar{R}}^2 \cong \frac{1}{n(n-1)} \sum (\cos(\theta_j - \bar{\theta}) - \bar{R})^2. \quad (12)$$

With this, a  $t$ -like statistic is defined as

$$t = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{s_{\bar{R},1}^2 + s_{\bar{R},2}^2}}. \quad (13)$$

This approach has two faults: The randomness of the mean phase difference  $\bar{\theta}$  is neglected, which leads to deviations especially for small values of  $\rho$  and therefore inaccuracies in the calculation of the variance  $s_{\bar{R}}^2$ ; and the distribution of  $\cos(\theta_j - \bar{\theta})$  is certainly not normal.

However, the  $t$ -test proves to be very robust against deviations from the distribution assumption for large samples, and so one can approximately assume that the given statistic is distributed like a standard  $t$  random variable. Based on this, the hypothesis of equal concentration has to be rejected if the modulus of this quantity exceeds a certain value, which is given by the percentiles of the  $t_f$  distribution for a chosen significance level. In this approximation, the test is applicable for a sufficiently large sample size.

The advantage of this approach is that it is generally applicable, but at the expense of theoretical accuracy. Despite of this, in the simulation (Fig. 1, green lines) the test proves to have properties very similar to those of the parametric test for the wrapped normal distribution. The tests presented in the following are non-parametric with increasing accuracy, but also increasing computational demands, and so the simple  $t$  test may be an option where precision is not that important.

## 4 Bootstrap Techniques

The basis of the two testing approaches presented so far is the theoretical derivation of statistical properties of the used measure. This is also the cause for their limitations, because either the specific distribution of  $\theta$  has to be known or the applied approximations are not generally valid. There is a group of computational techniques introduced by Efron & Tibshirani [1993] under the name bootstrap that make it possible to investigate those statistical properties empirically, replacing theory by the use of computer power. In the following, we will use some of these techniques, and we will do this in a way that additionally includes theoretical knowledge about  $\bar{R}$  to reduce the computational expense.

The basic idea is to generate “bootstrap replications” of the statistic of interest, and to calculate its variance and other properties on the set of these replications. To compute the sampling distribution of the statistic  $\bar{R}$ , it is necessary to somehow estimate the underlying distribution of the sample values, the population distribution of the random variable  $\theta$ . The best knowledge we have about this distribution is the sample  $\theta_j$  itself, and so the best (nonparametric) estimate of the population is a discrete distribution with probability  $1/n$  for each of the sample values. To generate new samples according to this estimate, one simply has to draw values from the original sample with replacement. Formally, if  $k_j$  ( $j = 1 \dots n$ ) are uniformly distributed independent integer random numbers in the range  $1 \dots n$ , then  $\theta_{k_j}$  is a bootstrap replication of the sample, and the statistic calculated on this sample is a bootstrap replication of  $\bar{R}$ . This technique is called resampling.

As a variant of the bootstrap, it is also possible to use a parametric estimate of the distribution of  $\theta$ . The parameters of the distribution are estimated from the sample, and the replications of the sample are taken from random numbers following the distribution with those parameter values. For our application, this would not be a relevant improvement over the parametric approach explained above; it would increase the complexity of the computations but only marginally improve the accuracy of the test. But the special form of our statistic of interest,  $\bar{R}$ , makes it possible to perform a “parametric” bootstrap at an intermediate level. As we have seen above, the asymptotic joint distribution of  $(\bar{C}, \bar{S})$  is two-dimensional normal, and the parameters of this distribution depend on the first and second trigonometric moments of the distribution of  $\theta$  (see Sec. 1.1). This description is an approximation based on the central limit theorem, but it is very good also for relatively small sample sizes (about  $n \geq 30$ ); it is valid for every distribution of  $\theta$  and for all values of  $\rho$ . With this, it is possible to generate “parametric” bootstrap replications of  $\bar{R}$  in a very direct way: The moments  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ , and  $\beta_2$  are estimated from the sample by the empirical moments, replications of  $(\bar{C}, \bar{S})$  are generated according to the corresponding two-dimensional normal distribution (see Eq. 4), and the replication of  $\bar{R}$  is calculated as  $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$ . In this way, it is not necessary to generate replications of the sample itself, which drastically reduces the computation time (in our calculations, up to a factor 40). This is especially important if the percentiles of the sampling distribution are to be calculated.<sup>5</sup>

### 4.1 Bootstrap $t$ -test

A very straightforward application of these techniques is to modify the  $t$ -test (Sec. 3) by replacing the direct variance estimation by the variance estimated from bootstrap replications. About 200 “parametric” replications are generated for  $\bar{R}_1$  and

---

<sup>5</sup>In the following we present tests based on bootstrap variance estimation and  $H_0$  simulation, as well as the related permutation test. It is also possible to perform a test based on bootstrap confidence intervals; these methods proved to be unreliable in the simulation and therefore have been left out.

$\bar{R}_2$  each, and the variance of these replications is used as  $s_{\bar{R},1}^2$  and  $s_{\bar{R},2}^2$ , respectively, in Eq. (13). In this way, the inaccuracy of the direct variance estimation is removed.

The consequence of this becomes visible in the simulation (Fig. 1, red line). The threshold for  $\rho$  below which the error of the first kind is smaller than necessary gets nearer to 0, and therefore the power of the test is increased. But the theoretical objections to the application of the  $t$ -test remain valid, and this improved version is still just a low-precision approximation.

## 4.2 Bootstrap $H_0$ simulation

Another approach is to use the bootstrap techniques to simulate the distribution of the test statistic  $\bar{R}_1 - \bar{R}_2$  under the null hypothesis. The underlying assumption is that if  $\rho_1 = \rho_2$ , then the distributions of  $\theta_1$  and  $\theta_2$  are also the same. This same distribution gets estimated by the distribution of the combined sample  $\theta_{0,k}$ ,  $k = 1 \dots 2n$ , where

$$\theta_{0,k} = \begin{cases} \theta_{1,k} & \text{for } k \leq n \\ \theta_{2,(k-n)} & \text{for } k > n \end{cases} . \quad (14)$$

Bootstrap  $H_0$  “replications” for  $\bar{R}_1$  and  $\bar{R}_2$  are generated by resampling or via the “parametric” approach from this combined sample (but with sample size  $n$ ), and the replication of  $|\bar{R}_1 - \bar{R}_2|$  is calculated. About 200/significance level values (4,000 for 5%) are generated, and the 200th largest of them is used as the rejection threshold of the test. That is, the null hypothesis is to be rejected if the actual value of  $|\bar{R}_1 - \bar{R}_2|$  exceeds this threshold.

It is important to see that the generated  $\bar{R}$  values are not replications of the value on the original sample, but that they correspond to a distribution according to the null hypothesis, which is simulated by the union of the two samples. The theoretical basis of this test is much better than that of the bootstrap  $t$ -test, because there is no assumption for the distribution of the test statistic, but this distribution is simulated explicitly. However, in the simulation (Fig. 1, cyan line) their performances seem to be essentially the same, and this at a much higher computational expense. Like before, there is a range of small  $\rho$  values for which the error of the first kind is smaller than necessary, decreasing the power of the test. The cause for this is that the bootstrap  $H_0$  simulation is not exact. This is improved in the next section.

## 4.3 The permutation test

According to Efron & Tibshirani [1993], the bootstrap simulation of the null hypothesis distribution of  $\bar{R}_1 - \bar{R}_2$  is not exact, because the properties of the original sample are not preserved. In the general bootstrap case that cannot be achieved, because to preserve them exactly would make the random variation impossible that is necessary to estimate the sampling distribution. But in the two-samples situation it is possible to introduce random variation and still preserve the statistical properties of the combined sample representing the null hypothesis distribution. The trick is to generate new samples  $\theta_{1,j}$  and  $\theta_{2,j}$  by randomly exchanging sample values between them. That is, the new samples are the first and second half of a random permutation of the combined sample  $\theta_{0,k}$ . Formally, if  $k_\ell$  is a random permutation of the integers  $1 \dots 2n$ , then the “replication” of  $\bar{R}_1$  is calculated on  $\theta_{0,k_\ell}$  with  $\ell = 1 \dots n$  and of  $\bar{R}_2$  on  $\theta_{0,k_\ell}$  with  $\ell = (n+1) \dots 2n$ . The following calculation is the same as for the bootstrap  $H_0$ . Such a testing procedure for the comparison of two samples is called a permutation test; it is very similar to the standard bootstrap approach, but is much older than bootstrap theory.

The theoretical accuracy of this test reflects itself in the simulation results (Fig. 1, magenta line). The permutation test is the only one that perfectly adheres to the chosen significance level for all values of  $\rho$ . Accordingly, its power for testing against  $\rho_2 = 0$  is the highest of all tests presented, which is especially important for small differences in  $\rho$ . The price of this superior performance is the high computational cost. Since every two simulated samples as well as the sample values in them are not independent of each other because of the permutation underlying their selection, the computation cannot be abbreviated by the “parametric” approach introduced above for the bootstrap computations. The permutations have really to be carried out and the “replications” of the test statistic have to be calculated directly from the generated samples.

## 5 Data from Time Series

As has been explained in Sec. 1.2, in the form presented above all the tests assume that the sample values in every sample are independent of each other, and that means they have to be obtained from independent realizations of the process of phase synchronization. The phase difference in a given realization is a function of time,  $\theta(t)$ . There are a number of different realizations of the process,  $\theta_j(t)$ ,  $j = 1 \dots n$ , and each of the sample values has to be obtained from another realization, typically corresponding to the same time point  $t_0$ :  $\theta_j = \theta_j(t_0)$ . This is necessary because in this way the sample size  $n$  is also the number of degrees of freedom inherent in the data. If one would use time series data from consecutive time points, the amount of statistical dependency reducing the number of degrees of freedom below the sample size would not be known. In the case of the parametric tests, the simple  $t$ -Test, and the “parametric” bootstrap tests this number directly enters into the equations, determining the variance of the test statistic.

In the case of the resampling-based version of the bootstrap tests, the situation is a little bit different. Here, the aim is to produce replications of the sample that have the same statistical properties as the original. Since resampling works by drawing with replacement, the new sample values are statistically independent, and by making the size of the new samples equal to  $n$ , the equivalence is granted. This would no longer work if one would simply put partly dependent data into the original samples and use that for resampling, because the structure of those dependencies would be destroyed in the resampling process.

But there is a version of resampling which takes this into account. If the realization of the process  $\theta_j(t)$  is sampled at certain time points  $t_m$ , then the original sample gets a two-dimensional structure:  $\theta_{jm} = \theta_j(t_m)$ . The resampling now has to be performed in a way that does not destroy the dependency structure within the rows (constant  $j$ ). This is accomplished by simply treating each row as a single vector-valued sample value  $\vec{\theta}_j = (\theta_{j\bullet})$ , and to perform the resampling by drawing from the set of these row vectors in the original sample. The same method can be used to adapt the permutation test to partly dependent data.

In this way, the full information from time series data can be utilized in the test, increasing its power. But it is important to see that one still needs multiple independent realizations of the process; if the sample just contains one vector sample value, there is nothing to resample or to permute. It will just be the case that the number of realizations needed to distinguish two states in a test is decreased because of the increased statistical power inherent in each sample value.

Since in this case the “parametric” bootstrap can no longer be used, but the resampling has actually to be performed, the advantage in terms of computational expense of the bootstrap  $H_0$  simulation versus the permutation test gets lost, and so practically the latter is the method of choice because of its superior accuracy.

Eventually, the bootstrap  $t$ -test may still be an alternative, if it is known in advance that all  $\rho$  values are sufficiently different from zero.

## 6 Conclusion

We have presented several different tests for phase synchronization, improving the procedures currently used in the literature. The improvements have been accomplished by using a test setup in which two samples of phase differences are compared and by applying insights and methods from directional statistics and bootstrap theory.

The presentation of the tests has been started with parametric procedures which derive from the assumption of a specific family of phase difference distributions. These are mathematically elegant and simple but have been shown to be applicable only in special cases of phase synchronization. A first attempt at a nonparametric procedure that relies on the robustness of the  $t$ -test has given the impression to perform reasonably well but has a weak theoretical basis. Following this, mathematically valid nonparametric test procedures have been obtained using bootstrap techniques; here we were able to strongly reduce their high computational cost by introducing a “parametrization” at an intermediate level. As a step beyond the bootstrap, the permutation test has been described as a perfectly exact but computationally intensive nonparametric test method. We have closed the paper with instructions how bootstrap-based and permutation tests are correctly applied to partly dependent samples like time series data.

In summary, the paper has presented a variety of testing procedures and highlighted their specific advantages and drawbacks.

## Acknowledgments

This work has been supported by Deutsche Forschungsgemeinschaft, research group “Conflicting Rules”.

MATLAB implementations of the test procedures which were used in generating Fig. 1 can be obtained from the corresponding author.

## References

- Afraimovich, V. S., Verichev, N. N., & Rabinovich, M. I. [1986] “Stochastically synchronized oscillations in dissipative systems” *Izv. VUZ Radiofiz.* **29**(9), 1050–1060.
- Allaria, E., Arecchi, F. T., Di Garbo, A., & Meucci, R. [2001] “Synchronization of homoclinic chaos” *Phys. Rev. Lett.* **86**, 791–794.
- Anishchenko, V. S., Balanov, A. G., Janson, N. B., Igosheva, N. B., & Bordyugov, G. V. [2000] “Entrainment between heart rate and weak noninvasive forcing” *Int. J. Bifurcat. Chaos* **10**, 2339–2348.
- Bhattacharya, J., Pereda, E., Kariyappa, R., & Kanjilal, P. P. [2001] “Application of non-linear analysis to intensity oscillations of the chromospheric bright points” *Sol. Phys.* **199**, 267–290.
- Blasius, B., Huppert, A., & Stone, L. [1999] “Complex dynamics and phase synchronization in spatially extended ecological systems” *Nature* **399**, 354–359.

- Boccaletti, S., Kurths, J., Osipov, G., Valladares, D. L., & Zhou, C. S. [2002] "The synchronization of chaotic systems" *Phys. Rep.* **366**(1-2), 1–101.
- DeShazer, D. J., Breban, R., Ott, E., & Roy, R. [2001] "Detecting phase synchronization in a chaotic laser array" *Phys. Rev. Lett.* **87**, 4101.
- Efron, B. & Tibshirani, R. J. [1993] *An Introduction to the Bootstrap*. (Chapman & Hall).
- Fujisaka, H. & Yamada, T. [1983] "Stability theory of synchronized motion in coupled oscillator systems" *Progr. Theor. Phys.* **69**(1), 32–47.
- Jerger, K. K., Netoff, T. I., Francis, J. T., Sauer, T., Pecora, L., Weinstein, S. L., & Schiff, S. J. [2001] "Early seizure detection" *J. Clin. Neurophysiol.* **18**, 259–268.
- Kiss, I. Z. & Hudson, J. L. [2001] "Phase synchronization and suppression of chaos through intermittency in forcing of an electrochemical oscillator" *Phys. Rev. E* **64**, 6215.
- Lachaux, J.-P., Rodriguez, E., Martinerie, J., & Varela, F. J. [1999] "Measuring phase synchrony in brain signals" *Hum. Brain Mapp.* **8**, 194–208.
- Lunkeit, F. [2001] "Synchronization experiments with an atmospheric global circulation model" *Chaos* **11**, 47–51.
- Mardia, K. [1972] *Statistics of Directional Data*. (Wiley).
- Mardia, K. & Jupp, P. [2000] *Directional Statistics*. (Wiley).
- Mormann, F., Lehnertz, K., David, P., & Elger, C. E. [2000] "Mean phase coherence as a measure for phase synchronization and its application to EEG of epilepsy patients" *Physica D* **144**, 358–369.
- Paluš, M. [1997] "Detecting phase synchronization in noisy systems" *Phys. Lett. A* **235**, 341–351.
- Paluš, M. & Hoyer, D. [1998] "Detecting nonlinearity and phase synchronization with surrogate data" *IEEE Eng. Med. Biol.* **17**, 40–45.
- Parlitz, U., Junge, L., Lauterborn, W., & Kocarev, L. [1996] "Experimental observation of phase synchronization" *Phys. Rev. E* **54**(2), 2115–2117.
- Pecora, L. M. & Carroll, T. L. [1990] "Synchronization in chaotic systems" *Phys. Rev. Lett.* **64**(8), 821.
- Pikovsky, A., Rosenblum, M., & Kurths, J. [2001] *Synchronization. A Universal Concept in Nonlinear Sciences*. (Cambridge University Press).
- Pikovsky, A. S. [1984] "On the interaction of strange attractors" *Z. Phys. B Con. Mat.* **55**(2), 149–154.
- Rodriguez, E., George, N., Lachaux, J.-P., Martinerie, J., Renault, B., & Varela, F. J. [1999] "Perception's shadow: Long-distance synchronization of human brain activity" *Nature* **397**, 430–433.
- Rosenblum, M. G., Pikovsky, A. S., & Kurths, J. [1996] "Phase synchronization of chaotic oscillators" *Phys. Rev. Lett.* **76**, 1804–1807.
- Rosenblum, M. G., Pikovsky, A. S., & Kurths, J. [1997] "From phase to lag synchronization in coupled chaotic oscillators" *Phys. Rev. Lett.* **78**, 4193–4196.

- Rulkov, N. F., Sushchik, M. M., Tsimring, L. S., & Abarbanel, H. D. I. [1995] "Generalized synchronization of chaos in directionally coupled chaotic systems" *Phys. Rev. E* **51**, 980–994.
- Schäfer, C., Rosenblum, M. G., Kurths, J., & Abel, H. H. [1998] "Heartbeat synchronized with ventilation" *Nature* **392**, 239–240.
- Sheskin, D. J. [1997] *Handbook of Parametric and Nonparametric Statistical Procedures*. (CRC Press).
- Stefanovska, A., Haken, H., McClintock, P. V. E., Hozic, M., Bajrovic, F., & Ribaric, S. [2000] "Reversible transitions between synchronization states of the cardiorespiratory system" *Phys. Rev. Lett.* **85**, 4831–4834.
- Taherion, S. & Lai, Y.-C. [2000] "Experimental observation of lag synchronization in coupled chaotic systems" *Int. J. Bifurcat. Chaos* **10**, 2587–2594.
- Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., & Freund, H.-J. [1998] "Detection of  $n : m$  phase locking from noisy data: Application to magnetoencephalography" *Phys. Rev. Lett.* **81**(15), 3291–3294.
- Ticos, C. M., Rosa, E., Pardo, W. B., Walkenstein, J. A., & Monti, M. [2000] "Experimental real-time phase synchronization of a paced chaotic plasma discharge" *Phys. Rev. Lett.* **85**, 2929–2932.
- Tokuda, I., Kurths, J., & Rosa, E. [2002] "Learning phase synchronization from nonsynchronized chaotic regimes" *Phys. Rev. Lett.* **88**(1), 014101.
- Zaks, M. A., Park, E.-H., Rosenblum, M. G., & Kurths, J. [1999] "Alternating locking ratios in imperfect phase synchronization" *Phys. Rev. Lett.* **82**(21), 4228.