Signal Processing 238 (2026) 110186

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

On cycloergodicity

William A. Gardner

Department of Electrical and Computer Engineering, University of California, Davis, CA, USA

ARTICLE INFO

Keywords:
Cyclostationarity
Cycloergodicity
Non-population probability
Fraction-of-time probability
Stochastic processes

ABSTRACT

Cycloergodicity is the equivalence of sinusoidally weighted time averages of measurement functions on sample paths of a stochastic process exhibiting some form of cyclostationarity to their expected values which, in turn, equal sinusoidally weighted time averages of time varying expected values of those measurement functions. Colloquially, cycloergodicity is a generalization, from unweighted *averages to sinusoidally weighted averages and thereby to periodic and almost periodic averages, of* the property "time averages equal (time averages of) ensemble or population averages". Despite the historical practice of treating ergodicity as a strictly mathematical subject in a theorem/proof format, this article provides a narrative presentation of previously missing cycloergodicity theorems, which are expressed in plain English, with minimization of distracting technical detail to enable readers to use the concepts in their work on probabilistic analysis of time-average statistics derived from single records of time series data without populations. The results obtained do not support the use of stochastic process models for the empirical types of applications addressed. This motivates a brief but hard-hitting perspective on an alternative probability model referred to as Fraction-of-Time Probability, a non-population probability. For technical details required for mathematical proofs of the theorems, readers are referred to a classic book.

1. Introduction to cycloergodicity

In mathematics, the single word *ergodicity* is used to represent the idea that a point in a moving system will eventually visit all parts of the space that the system moves in, in a uniformly random sense. For example, the system can be the mathematical model of an ensemble of trajectories (coordinate-vector-valued functions of time) of physical particles (points) in a gas within a chamber (a time/position space) or an ensemble of sample paths (called sample points—not the analogs of the gas particles) in the abstract sample space of all possible trajectories of a hypothetical stochastic process, such as a scalar-valued thermal noise voltage waveform, for which the term trajectory is used to convey the concept of a point moving along a locus in time/amplitude space. This mental construct is developed around the underlying concept of a hypothetical experiment which, when performed, produces at random one such trajectory. Ergodicity is intended to imply that the average or statistical behavior, over all trajectories at one point in time, of the system can be deduced (with probability equal to one) from the statistical behavior over time of a single randomly selected trajectory.

The concept of ergodicity has led to a substantive literature in the mathematics of probability and more generally measure theory and dynamical systems and also plays a key role in empirical time series analysis in the field of Mathematical Statistics. The objective of this article is to provide solid scientific support for a paradigm shift in Mathematical Statistics involving time series analysis, with no intended impact on the broader mathematical theory of Ergodicity, a field in which I claim no expertise.

The field of study motivating this article is the probabilistic analysis of statistics derived from time series analysis: the measurement of time averages of specified measurement functions of time series data for the purposes of statistical inference and decision making. This field is quite mature and there is no intention of contributing directly in any major way to the body of knowledge summed up in theory and methodology and algorithms for implementing inferences and decisions. Rather, the target here is rooted more deeply in the conceptualization and mathematical definition of probability as used for the basis for the probabilistic analysis of time-average statistics, though this does have an impact on both thinking and some aspects of the practice of probabilistic analysis.

This article is a prelude to a companion article [17], which provides a broad perspective on probabilistic modeling in science and engineering; the present article is motivated by the admonition: "know thine enemy". The meaning here is that the new research results on cycloergodicity reported do not support the use of stochastic process models for the class of time series analysis problems of interest, which permeate the empirical components of science and engineering. This is explained in the final section of the paper, where a brief perspective on an alternative probability model is provided. This perspective is a component of the

E-mail address: bill@cyclicsignals.com.

https://doi.org/10.1016/j.sigpro.2025.110186
Received 14 April 2025; Received in revised form 2 June 2025; Accepted 22 June 2025
Available online 10 July 2025

Table of Acronyms

ACS	Almost Cyclostationary
AMACS	Asymptotically Mean Almost Cyclostationary
AMCS	Asymptotically Mean Cyclostationary
AMS	Asymptotically Mean Stationary
CDF	Cumulative Distribution Function
CS	Cyclostationary
CT	Continuous Time
DT	Discrete Time
LLR	Log Likelihood Ratio
LR	Likelihood Ratio
PDF	Probability Density Function
PQ	Period Quality
QP	Quasi Periodic
S	Stationary

full perspective on the alternative, called Non-Population Probability, that is presented in the companion paper.

For an experimentalist having measured a single time series of data and doing empirical data analysis by calculating time-average statistics, with no interest in hypothetical or real populations of time series, should a stochastic process model of the time series based on orthodox population probability (the Kolmogorov model) be introduced for the purpose of probabilistically quantifying the utility of the statistics or should unorthodox non-population probability based on fraction-of-time (FOT) measurements from the same time series be used, or should a mathematical model of an infinitely long version of the time series be used?

- Answer 1: *Were the empiricist to consult just about any mathematical statistician today (an expert in a formally recognized field of academic study and industrial laboratory work), there would, with high likelihood, be no quandary: the typical answer would be "population probability is the only recognized mathematical model of probability in Mathematical Statistics".*
- Answer 2: *Were the empiricist to consult the Author of this article, who has a long history of innovation in time-series analysis and its applications in engineering, there would, with certainty, be no quandary: the answer would be "use the real data you have to calculate FOT-probabilities of whatever events involving the measured statistics are of interest, or use FOT-expected-values of whatever functions of the statistics are of interest, or, especially if the amount of data is lacking, use a formulaic model of the time series and calculate infinite-time FOT probabilities or FOT-expected values from that model.*

Before discussing the pros and cons of these two conflicting answers, let us consider an example.

1.1. Motivating example

Let us start with a simple motivating example: An example of a time-average-based statistic from a time series is the empirical variance of a data record $x(t)$ for some range of time, t. For specificity, let $x(t)$ represent a thermal noise voltage in some electrical circuit for which the empirical variance is

$$\widehat{\sigma}^2(x) = \frac{1}{W}\int_0^W [x(t) - \widehat{m}(x)]^2\,dt$$

where $\widehat{m}(x)$ is the empirical mean,

$$\widehat{m}(x) = \frac{1}{W}\int_0^W x(t)\,dt$$

(An alternative discrete-time model is completely analogous).

Assuming the data used above on the time interval [*0, W*] is only 10 % of the entire available data record on the time interval [*0, V*], we can calculate 10 independent measurements of empirical variance, indexed by n:

$$\widehat{\sigma}_n^2(x) = \frac{1}{W}\int_{nW}^{(n+1)W} [x(t) - \widehat{m}_n(x)]^2\,dt, \quad n = 0, 1, 2, 3, \cdots, 9$$

As a metric for quantifying how reliable any one of these variance measurements is, we can calculate the percentage P of the 10 blocks of data of length W for which the deviation from the average of the 10 variances exceeds, say, 5 % of that average:

$$P = \frac{1}{10}\sum_{n=0}^{9} U\left[\left|\widehat{\sigma_n^2}(x) - \widehat{m}\left(\widehat{\sigma^2}\right)\right| - 0.05\widehat{m}\left(\widehat{\sigma^2}\right)\right]$$

where U is the unit step function that is 0 when its argument is less than zero and one otherwise.

This quantity P is the fraction of discrete time n for which the empirical variance exceeds the specified error bound. This Fraction of Time (FOT) of the occurrence of an event has all the properties of the standard axiomatic definition of probability and is referred to as *FOT-Probability*. It is an example of an element of a complete theory of finite-time probability (either discrete or continuous time) that can be used for probabilistic analysis of time series data. In addition, using a mathematical model of $x(t)$ that extends over all time along the real line or the integers and that is, in whatever manner possible, based on the actual record of data available and any knowledge about the origin of the data (such as a specified dynamical system driven by what is called white noise (in the FOT probability sense, including continuous-time Gaussian noise or non-Gaussian discrete-time noise), one can in principle use the limit as averaging time approaches infinity to calculate an idealized FOT-probabilistic model of the data—idealized in the sense that all random effects in the data are averaged away in the limit.

Numerous examples of FOT-probability models of communication signals are provided in [5].These models can be comprised of whatever probabilistic parameters are desired, such as individual moments, joint moments at multiple time separations, associated cumulants, first order and higher order cumulative probability distribution functions, spectral densities and higher-order spectral moments and associated spectral cumulants, etc., as shown in [5] and other references provided below that followed [5]. The time averaging performed empirically from the data or mathematically from a model of $x(t)$ can include averaging types that produce stationary, cyclostationary, and almost cyclostationary probabilistic models (the latter two of which capture statistical cyclicity), and generalizations thereof [5,16].

1.2. A scientific approach

The above example illustrates a scientific approach to probabilistic modeling of data which is a relatively mature theory that is today considered unorthodox and is mostly not used in Mathematical Statistics. The history of this anomaly is touched on in this article and treated in more depth in the companion article [17]. What IS orthodox is to *imagine* there is a population of times series (which there actually may be in practice sometimes, but frequently not) and define probability in terms of relative frequency of occurrence of events upon random selection of hypothetical data records from a mathematical model of the population. This is done by axiomatically positing an abstract sample space of typically an uncountable infinity of possible data records and an abstract probability measure defined on all typically uncountably many subsets of interest in this sample space and axiomatically positing special properties of the sample space and measure that facilitate theorem development and theorem proving. In other words, the orthodox approach to probabilistic modeling of time series data introduced by Kolmogorov in 1933 is designed for mathematicians with little concern for empiricism or, for many applications, realism in any sense. This example amply demonstrates that we have a serious problem in Mathematical Statistics that is long overdue for corrective action. This article and its companion [17] comprise my proposed solution, which has already waited 38 years for action to be taken, beyond that of a few colleagues of mine.

- *In this article, the weak link we call ergodicity between empiricism and the Kolmogorov stochastic process model is investigated and shown to be of little use for empiricists working with real data, which is the majority of real scientists and engineers.*
- *It is hard for me to understand how so many people in so many critically important fields of study who have been in need of a better theory for probabilistic analysis can have apparently been complacent with the orthodox for so long. This enigma is further studied in the companion paper [17].*
- *Empiricists do encounter time series analysis problems for which the data record length is too short for calculating useful statistics or, barely long enough but still too short for calculating FOT-probability models of the statistics. In such cases, relying completely on an idealized FOT-probability model, possibly using a data model (as described earlier), is more parsimonious than creating an imaginary population and positing a corresponding stochastic process model.*

1.3. Ergodicity vs. cycloergodicity

Cycloergodicity is the word we use, by analogy with the use and meaning of the word *Ergodicity*, for the property of a stochastic process that guarantees the convergence of sinusoidally-weighted time averaged measurements (functions of the process) on a realization (sample path) of a stochastic process, as averaging time increases without bound, to sinusoidally-weighted time averages of time-varying expected values of those same measurements, which is the same as saying sinusoidally-weighted time averaged measurements on a stochastic process converge to their own expected values. Cycloergodicity also guarantees the convergence of periodically time varying time average measurements, obtained by averaging time samples taken once every period and doing this separately for every starting time throughout one period. Ergodicity applies to time averages of stationary stochastic processes and, more generally, to *Asymptotically Mean Stationary* (AMS) stochastic processes, which are nonstationary processes for which the finite-time averages of time-varying expected values converge to a limit value, the temporal-mean value of the probabilistic-mean (expected value). *Cyclostationary* (CS) processes and more generally *Almost Cyclostationary* (ACS) processes are examples of AMS processes, and they can be generalized to Asymptotically Mean (Almost) Cyclostationary (AM(A) CS) processes. *Almost Cyclostationary Processes* are generalizations of *Cyclostationary Processes* that exhibit at least two incommensurate statistical periodicities: the time functions composed of the time translates of their probability measures are *almost periodic functions* in the classical sense [1] for every event set (with probability 1).

Although ergodicity applies to ACS and AMACS processes, it still means that the time average converges to its expected value, even if the process is not stationary provided that it is AMS, including the various versions of cyclostationary processes. But this is a distinct type of *ergodicity that is generally unrelated to cycloergodicity*. Cycloergodicity applies to stationary, cyclostationary, and, more generally, almost cyclostationary processes, and even more generally to their asymptotic mean versions for which the expected values are, respectively, time-invariant, periodically time varying, or almost periodically time varying, or more generally none of the preceding, but containing one of the preceding as an additive component.

Stationary (S) processes, even those that are ergodic, can fail to be cycloergodic, in which case their sample paths can exhibit statistical cyclicity even though the probabilities and expected values of the process do not. This is not well known but is easily demonstrated without resort to contrived examples. Similarly, cyclostationary processes, even those that are cycloergodic for the period of cyclostationarity, can fail to be cycloergodic for other periods. Consequently, the classic *Theorem of Ergodicity* (and its variations [2]) cannot be manipulated to include cycloergodicity, as has been done for the special case of cyclostationary processes by representing a CS scalar-valued process by a S vector-valued process [2]. This means that a stationary or cyclostationary or almost cyclostationary process may be cycloergodic for one period (which, for discrete-time processes, must be commensurate with the time sampling increment), and still not be cycloergodic with some other incommensurate period. This leads to the following important conclusion:

- *When sample paths of stochastic processes exhibit statistical cyclicity, the existing theory of ergodicity is inadequate to explain the relationship between sinusoidally weighted time averages and their expected values.*

In other words, Cycloergodicity is a genuine generalization of Ergodicity. Except for the case of cyclostationarity (and AMCS) with a single period (which for discrete time, is commensurate with the time sampling increment [2,3]) and except for what might be said to be a contrived variation thereof for what are called *quasi-periodic probability measures* (those AMACS measures with only rational cycle frequencies [3,4]), there are to my knowledge (based on the relevant literature) no cycloergodic theorems for the various other cases mentioned above.

The fact that there is no mention of this incompleteness of ergodicity theory in the classic mathematically oriented book for engineers on ergodicity [2], despite the existence of substantial literature demonstrating the practical utility of ACS process models dating back to the classic engineering book [5] published almost forty years ago, and including the sequel [6], is strange. It seems to speak to the ongoing separation between what might be called pure-mathematics and the applied-mathematics *oriented* practice of engineers, physicists, and other scientists, as well as some applied mathematicians—one of many reflections on the impact of humanity on the conduct of science [7, p.7;8, chap.7].

In an attempt to move this important yet neglected aspect of stochastic process theory forward, this article presents tutorial explanations of necessary and sufficient tests of discrete-time and continuous-time stochastic process models for the property of cycloergodicity. Each of these decomposes both the question and the answer regarding cycloergodicity with multiple arbitrary periods into separate unrelated parts, one for each period of interest. As discussed below, it also is shown that the known ergodicity theorems for discrete-time cyclostationary AMS processes [2,9] and their continuous-time counterparts (characterized by vector-valued discrete-time process [2,10]) can be applied directly to each and every component part in the decomposition to achieve the desired necessary and sufficient condition for cycloergodicity, one period at a time. Although straightforward word descriptions of the individual steps required to test a process model for cycloergodicity are provided, mathematicians may prefer to refer to these theorems as *propositions* because it may be argued that the word descriptions are not sufficiently rigorous; for example, they may not cite all restrictive assumptions precisely determining applicability. Nevertheless, they do provide key necessary and sufficient conditions for cycloergodicity—conditions that can be seen to be natural counterparts to existing ergodicity theorems. But to the astute reader, it should be clear that these theorems are simply the composites of a known ergodicity theorem and one or the other of two lemmas enabling a novel application and insightful explication of theory and methodology from a single source [2]. From this perspective, these results might be labeled *corollaries*. Nevertheless, their explicit recognition is, one might say, long overdue. And the form of presentation by their originator, the Author, is specifically designed to serve empiricists wanting to perform probabilistic analysis of their measured cyclic statistics of real time series. To categorize them as simply appendages to known ergodicity theorems deemphasizes their importance in conceptualization.

Before proceeding with this discussion, it is noted for the readers' benefit that the now-traditional usage of the term *Ergodic Theorems* for the theorems addressing the convergence of time averages to their expected values (such as Birkhoff's Ergodic Theorem) is not used in this article. Being a stickler for precise language, as one component of the pursuit of unambiguous communication, I point out that these theorems are not themselves ergodic: they cannot exhibit the property of ergodicity, as some stochastic processes or their probability measures can, although they can and do address the property of ergodicity. It follows that these should be called *Ergodicity Theorems.* Consequently, the subject of the discussion in this article is referred to as *Cycloergodicity Theorems*, not cycloergodic theorems. (Unfortunately, the equivalent language misuse in applications of cyclostationarity is pervasive: authors often use the adjective "cyclic" in place of the correct nouns/adjectives "cyclicity" or "cycle". The entities too often referred to as *cyclic*, do not cycle. This can be confusing to students of the subject, cf [8, Chap. 1, Definitions].)

The purpose of the next section on background is, as it is sometimes said, to leave no stone unturned. Consequently, the narrative is admittedly extensive and possibly a little redundant here and there, and this is the exact opposite of what is found in existing literature on ergodicity theory which, for the most part, has been contributed by mathematicians and strongly mathematically oriented physicists and engineers. It is no doubt said by some mathematicians that this is its strength. And I cannot argue with this if the objective is to develop more theorems and more proofs, such as in information theory, which is in large part what mathematicians do. But the objective of this article is entirely different. The intended readership is experimentalists and empiricists whose primary objective is to derive information from data and who are wanting to probabilistically model their statistics derived from time series measurements. This is the realm of mathematical statistics. But, dare I say, the emphasis in this field on mathematics may not serve experimentalists and empiricists as well as a more data-centric treatment, such as that briefly introduced in this article and treated expansively in cited books and research papers. Unlike the problems of primary concern to mathematicians, the problems here are data centric, which means that mathematical models must be tied very closely to the real data they represent. This is the intended essence of the concept of ergodicity for empiricists. But this paper shows that ergodicity theory falls short of reaching this goal and that recognition of this is long overdue.

2. Background on cycloergodicity

In 1958, William Bennett [11, p.1510] from Bell Telephone Laboratories wrote "We suggest the name "cyclostationarity" and analogously "cycloergodicity" be used . . ." in connection with a continuous-time stochastic process model for a digital-pulse-amplitude-modulated signal in his statistical analysis of regenerative digital repeaters in communication systems. In fact, as pointed out later, these terms were not needed, mathematically, because this signal model is equivalent to an infinite-dimensional (or, in discrete time, finite dimensional) vector-valued stationary process and the desired property of the model is a conceptually straightforward extension of traditional ergodicity from scalar-valued processes to vector-valued processes. However, this dismissal does not expose the fact that the cycloergodicity test must be applied specifically for the correct period and applied separately for each and every one of the uncountably infinite (or finite) number of time points within a period. Ignoring this detail, this may be the reason the key engineering source on ergodicity theory [2] makes no mention of the term *cycloergodicity* despite addressing this concept albeit only in terms of finite-dimensional vector-valued discrete-time S-process representations of CS processes. Nevertheless, this equivalence is used here in a novel manner as *a necessity*, not an option, for leveraging classical theory [2] to establish tests for cycloergodicity for previously unconsidered process types that are *not* equivalent to vector-valued stationary processes, namely *almost cyclostationary processes*. No previous work has successfully tackled the ergodicity question for almost cyclostationary processes. This term was introduced in 1978 by the Author [12] for processes whose translation indexed probability measure is an *almost periodic function* of the translation time parameter [1]. Such models arise frequently in radio frequency multiuser communication systems (e.g., cellular telephone), which was a driving motivation for the Author's work, who had recently been employed by Bell Telephone Laboratories before returning to academia. This term is also used in this article because it is now standard terminology in fields where these models are used.

The only known attempt to establish cycloergodicity (with probability $= 1$) theorems for almost cyclostationary processes is that of the Author and his coauthor, Russel Boyles, in the 1983 article [3], but success was not achieved for discrete-time processes, and no attempt was made for continuous time. In this present article, the term *cycloergodicity* is adopted but also used in the modified form *T-cycloergodicity* because it is shown that a single process can exhibit ergodic properties associated with some periods and not others, making explicit mention of the period essential. At this point in our understanding, avoidance of the term cycloergodicity as in [2] is no longer advisable. It is shown in this article that establishment of a theory of cycloergodicity for discrete-time almost cyclostationary processes requires stepping outside of the admittedly extensive and mature field of ergodicity theory (cf [13]) by introducing the concepts of interpolating/re-sampling a stochastic process model and the associated probability measure to achieve a *model transformation* that is both necessary and sufficient in order to use classic ergodicity theory to derive a necessary and sufficient condition for T-cycloergodicity for a period T that is incommensurate with the original process's time-sampling increment. This transformation of the process, but not the measure, is also necessary for testing for T-cycloergodicity, and this appears to have significant practical consequences.

The introductory section in the 1983 paper [3] fully defines the class of cycloergodicity problems that had not then and has not until now been addressed by what can be called the classic theory of ergodicity for AMS discrete-time processes and their special cases, which is the domain addressed in [2]. The Author has found no published work on this ignored (vis a vis cycloergodicity) class of processes, here termed *non--contrived AMACS including non-contrived ACS,* in the ensuing 42 years (*non-contrived* means continuous-time or discrete time with any cycle frequencies not exhibiting a finite limit point and, in particular, not limited to rational cycle frequencies only, and is used when necessary to avoid ambiguity with Quasi-Periodic probability measures as explained below). This seems to fly in the face of the great deal of published work on ACS processes in the statistical signal processing engineering literature stemming from the seminal work reported in the paper [12], and subsequent work in three 1985–1994 books [5,14,15], followed 24 and 30 years later by the encyclopedic treatments [16] and [8], all of which study in immense detail this otherwise largely ignored class up until the 1990s, producing volumes of results on development of a comprehensive theory and associated statistical methodology, which has led to resultant technology demonstrating significant abilities to mitigate the undesired effects of signal interference. Much of this work was motivated by the explosion of multiuser communication systems technology, such as cellular telephone, beginning around 1980, because these systems naturally create interfering signals residing in the same time periods, frequency bands, and spatial locations. The ACS model is quintessential for representing interfering signals and enabling the technological development that followed. Most of the technology developed for exploiting cyclostationarity uses algorithms that compute time-average statistics with sinusoidal weighting or its equivalent in terms of periodic time averages. The majority of this published work, excluding the Author's, bases its probabilistic analysis on expected values from the theory of population probability (stochastic processes) without the benefit of cycloergodicity theory establishing the relationship (or lack thereof) between these expected values and the time averages: the topic of this article.

This history brings into question the importance of cycloergodicity theory. The achievements made in interference mitigation technology were accomplished without the cycloergodicity theorems presented in this article. This supports the perspective, presented in the companion paper [17], that the use of stochastic processes in time-series analysis based on time averaging is typically illogical and out of step with current practice in which time-average statistics are the focus and expected values are a distant abstraction. Engineers have apparently succeeded (and continue to) in their work by using typical empirical methods, without any logical support from theory, but rather with intuition that analysis of expected values can suggest algorithms that use time averages instead. We can congratulate these engineers for their progress despite the conceptual and mathematical pitfalls that awaited them (summarized in this article), which arise from unknowing use of potentially non-cycloergodic models. We can only surmise that progress may have been accelerated, or better solutions obtained, had the engineers made logical use of the non-population probability theory based on fraction-of-time probability for their probabilistic analyses of the time-series statistics needed in this development work. That use of this alternative to stochastic processes can be fruitful is amply illustrated with the plethora of signal analysis studies and cyclostationarity-exploiting algorithms invented using fraction-of-time probability theory as documented throughout the preceding references [4–8] in this article, as well as [18–48] and the companion perspective paper [17] (quotes from which are copied in last section of this article).

Nevertheless, the stochastic process theory deserves to be completed with the addition of needed cycloergodicity theorems; so, let us proceed. But it is a statement worthy of note here that the entirety of the Author's development of the now-comprehensive theory of cyclostationarity from the mid-1980s forward, and his many applications of this theory, were built exclusively using fraction-of-time probability and yet the follow-on work by almost everyone so engaged has been conducted using unnecessarily abstract stochastic process models that may or may not exhibit the cycloergodicity that is required for establishing relevance to the motivating empirical time-series analysis problems.

The paper [3] shows that the class of AMACS processes is, perhaps surprisingly, identical to the class of AMS processes, and it extends/generalizes the classic ergodic theorems to cycloergodic theorems. There are no surprises in methods or results, except for one: The class of discrete-time AMACS processes is shown to include ACS processes *only if for any cycle frequencies that are incommensurate with the sampling increment, the cyclic components of expected measurements are zero.* So, the only processes covered by the extended/generalized theory that are not CS (Cyclostationary) are what might be called a *contrived class* referred to as those with *Quasi-Periodic (QP) probability measures* as defined by Blum and Hanson in 1966 [3,4] (discrete-time AMACS or ACS with only rational cycle frequencies). What can be called *non-contrived ACS processes* that are not CS are excluded from the cycloergodicity theorems presented in [3]. A revealing example given in [3] is the independently distributed discrete-time Bernoulli process with an *Almost Periodic* sequence of probably parameters that equals a constant plus a sinusoid with frequency that is incommensurate with the process time-sampling increment. This example is not contrived, and it reveals that we have no ergodicity theory for substantive discrete-time ACS processes that are not CS (or QP). For those unfamiliar with almost periodic functions, they comprise the class of functions that consist of a finite or infinite sum of periodic functions with incommensurate periods.

- *Finally, after 42 years since the only known previous attempt, a set of theorems for these previously omitted processes is introduced in the next section. These theorems reveal necessary and sufficient conditions for discrete-time or continuous-time stochastic process models to exhibit the cycloergodicity properties that are necessary and sufficient for estimating almost periodically time varying expected values using time averages. These theorems accommodate the ubiquitous situation of simultaneous presence in data of statistical periodicities having incommensurate periods which, in the case of discrete time, can also be incommensurate with the time-sampling increment. These results inform the mathematical analyst of the necessary and sufficient analytical tests that must be performed on a stochastic process model to determine its cycloergodicity.*

The approach proposed in these theorems is to decompose any given ACS (i.e., almost periodic) or AMACS (i.e., contains an additive almost periodic component) probability measure into a set of component CS (i.e., periodic) measures (and a residual for the AMACS case); and to then independently apply an existing ergodicity theorem to each component measure by using its vector AMS representation.

In the simplest of ACS cases, a discrete-time ACS process has only one periodicity in its probability measure (or, more accurately, in its family of time-translated measures), but the period is incommensurate with the time-sampling increment; otherwise, the process is cyclostationary. In the next simplest case, a continuous-time ACS process has two periodicities with periods that are incommensurate with each other. Such processes can be constructed, for example, by adding two CS processes with incommensurate periods. *This happens every day all over the world when two propagating cellular telephone radio-frequency signal waves, with distinct carrier frequencies and/or digital symbol rates, both impinge on the same receiving antenna of a cell phone or base station. Such signals typically exhibit only or up to 4 or possibly 6 distinct periods in its autocorrelation function, while exhibiting a countable infinity of periods in its CDFs.* The theory of cyclostationarity can be used to design receivers that separate these interfering signals (see, for example, the topics Cyclic Wiener Filtering, Cyclic Signal Interception, and Nonlinear System Identification in the books [4–8] and other related topics in [18–49]. The importance of this is reflected in one narrow-perspective represented by three examples: (1) there is a representative patent [49], purchased from the inventors by Apple Computers for 6 figures, on an invention that exploits cyclostationarity to separate spectrally overlapping cell phone signals, with the objective of improving system capacity and message quality. (2) Lockheed Martin Corporation purchased from the lead inventor for 7 figures cyclostationarity-exploiting intellectual property (signal processing algorithms, software, and R&D reports) targeting solutions for signal reception and information extraction in highly corruptive radio-frequency environments. (3) 10 agencies and laboratories of the US Government and 10 of its large contractors invested 8 figures over 25 years in support of the R&D conducted by one faculty member's team of graduate students into cyclostationarity exploitation in the field of signals intelligence [7, page 12].

However, the practical value of the cycloergodicity theorem for ACS and AMACS processes is limited by the same constraints encountered in applying traditional ergodicity theory, namely the difficulty that can be encountered in testing for the necessary and sufficient mathematical conditions for cycloergodicity of a stochastic process. These basic theorems, generally speaking, are not of much use in empirical analysis of time series data based on time-averaged or time-averaged/sub-sampled or sinusoidally weighed time-averaged measurements. They simply provide a modicum of comfort to those using stochastic process models in connection with doing empirical time-average based analysis of time-series data. Nevertheless, cycloergodic theorems establishing that certain specific classes of stochastic process models are cycloergodic are of definite help here—see Appendix II. (Also, it should not be forgotten that ergodicity theorems can be useful in theoretical work on stochastic processes not expected to be directly relevant to empirical analysis.)

To go a step further with the cell phone example, consider that a stochastic process model for the additive mixture of two cell phone signals can be cycloergodic or not. This depends, in part, on the modeling of the phases of the carrier frequencies and/or the phases of the digital symbol repetition rates (or any other periodic structure in the signal format). Any randomness of such phase parameters of the model destroys cycloergodicity at the period of the associated frequency or rate. In addition, cycloergodicity is destroyed by any randomness of a time-invariant amplitude parameter multiplying one or both signals or individual components of a signal (see Appendix II). As can be seen, the absence of cycloergodicity is not a contrived situation. It is true that an engineer trained in the theory of cyclostationarity of man-made communications signals (in today's language in the world of machine learning, a person with *domain expertise*—some say a dying breed) can likely determine which parameters in an explicit stochastic process model of such signals will destroy cycloergodicity if modeled as random, but this situation of being intimately familiar with the design and modeling of the signals of interest might well not carry over to data analysis in the sciences for which the signals of interest are not man-made. Nevertheless, biology has enjoyed many applications of the theory of cyclostationarity [16, p.362], as have other fields throughout the sciences [16, chap.10]. Before leaving the topic of utilizing cyclostationarity for separating interfering signals, it should be mentioned that signals with identical cycle frequencies also can sometimes be separated. It is not always necessary for cycle frequencies to be different or incommensurate for exploitation of cyclostationarity.

Readers desiring more motivation for buckling down and addressing cycloergodicity directly are referred to a set of six revealing examples in the next subsection illustrating the surprisingly troublesome anomalous behavior that can arise with non-cycloergodic stochastic process models. These remarks are not proven herein to be correct, but the Author assures readers that he has demonstrated this and suggests that the desired impact of the claimed anomalous behavior will be greater for readers who choose to ponder these apparent mysteries instead of having readily available explanations handed to them.

2.1. Examples of non-cycloergodic processes

This section on background concludes with several remarks illustrating the apparently strange nature and seriousness of issues that can arise with non-cycloergodic stochastic process models. Each remark centers on a fact-example that illustrates how severely the breakdown between empirical data behavior and stochastic process model properties can be in the absence of *cycloergodicity*.

- **Remark 1**: Hidden statistical dependence and causality

For two jointly strictly stationary and ergodic stochastic processes, X and Y, for which Y's future is to be predicted from the past of X, there are examples for which there is no predictive power using any linear or nonlinear time invariant transformation of X if Y is non-cycloergodic, despite the fact that prediction with any desired level of accuracy, given enough data, using a linear least squares periodically time varying predictor is achievable. Example: Let Y be a moving average over a piece of the past of the stationary ergodic process X with periodically time varying coefficients all having the same period and all containing random phases (either all the same or statistically independent) that are uniformly distributed over one period. The probabilistic correlations between Y and any/all linear or nonlinear functions of X are zero because of the random phases, so X has no time-invariant predictive power for Y. Yet there is a periodically time-varying transformation of any sample path of X that produces predictions of the corresponding sample path of Y with accuracy that increases without bound with increasing data segment length. Learning this prediction transformation requires performing a linear least-squares periodically time-varying prediction of sufficient memory length and sufficient harmonic content of predictor coefficients using only the past of both X and Y for training, and then applying the trained predictor to X up to its present value to predict a future value of Y. The harmonic content of the periodic weights can be estimated using an estimator of the cyclic spectral density or cyclic autocorrelation function for Y and simple combinatorics that relate periodicities in Y to those in its lag product. This predictor sees only one realization of the random phase and can estimate it over time using a least-squares procedure. The role of the random phase in this model is to render the process non-cycloergodic in a manner that renders the correlations between Y and all linear and nonlinear functions of X equal to zero.

- **Remark 2**: Hidden strength of periodicities

The strength of cyclostationarity of a CS or ACS stochastic process $Y(t) = X(t+\Theta)$ at any one or more periods can be weakened, relative to that in a single sample path, to any desired extent by design of the PDF of the phase randomization variable Θ. This fact is intimately associated with the fact that there is, in general, no fixed relationship between the strength of cyclicity in a non-cycloergodic stochastic process and the strength of cyclicity in its individual sample paths. This is proven in [12], where the exact impact of the PDF of the phase variable on the expected strength of cyclic features is derived. Any non-degenerate PDF (PDF that is not equal to a single Dirac delta) destroys any cycloergodicity that may have been present prior to phase randomization.).

- **Remark 3:** Hidden strength of spectral redundancy

All CS and ACS stochastic processes exhibit some non-zero degree of spectral redundancy [18–20]. This degree is reduced by phase randomization and also can be modified by amplitude randomization of process components that are the origin of spectral redundancy. For example, since the n-th moment of a randomized amplitude factor on a CS component of a process will generally differ from the n-th power of any realization of that amplitude factor, the degree of n-th order spectral redundancy will differ. More specifically, for $X = Y + AZ$, where processes Y and Z are independent of each other, each have zero-mean value, Y is S and Z is CS, $n = 3$, and the PDF of the factor A is even, the degree of first-order cyclostationarity of X^n and therefore the degree of spectral redundancy among triples of frequencies in X with amplitude randomization A can be zero despite the strengths of redundancy of its various realizations. This phenomenon is relevant to Remark 2 as well.

- **Remark 4**: Mysterious origin of spectral lines

Let the input process X to a stable nonlinear time-invariant transformation be strictly stationary and ergodic and have no spectral lines in its spectral density of expected power or of time-averaged power and be non-cycloergodic. The corresponding output process Y also is strictly stationary and ergodic, but it can exhibit spectral lines in both its spectral density of expected power and of time-averaged power. Example: If the input is the product of a sine wave with random phase uniformly distributed over $[0, 2\pi]$ and a stationary ergodic zero-mean process, and the nonlinear transformation is a squaring operation, then even though the input and output are stationary and ergodic and the input exhibits no spectral lines, the output time-average-power spectrum and expected-power spectrum will both contain spectral lines at frequencies 0 and twice the frequency of the sine wave factor at the input. Without knowing the exact structure of the input process, this raises the question of where the spectral lines came from. In other words, the seemingly reasonable assumption that stationarity and ergodicity of the input and output of a time-invariant stable system guarantees that, with no spectral lines at the input, there can be none at the output is false. The reason is that non-cycloergodicity produces hidden periodicity.

- **Remark 5**: Performance loss of likelihood-ratio detectors

Consider three alternative approaches to detection of the presence of a random signal in random noise, based on three alternative models for the observed data which is actually statistically independent time samples of stationary Gaussian noise, possibly plus cyclostationary Gaussian signal with unknown time origin (phase). Detector-1 uses the Log Likelihood Ratio (LLR) test statistic for exactly this model, with maximization over the unknow phase parameter. Detector-2 uses as a test statistic the Likelihood Ratio (LR) after removal of conditioning on the phase for the resultant non-cycloergodic model by taking the expected value of the LR w.r.t the phase modeled as a random variable uniformly distributed over the period of cyclostationarity. Detector-3 uses the LLR for a stationary Gaussian model for the signal with the same power spectral density as that of the stationarized (phase-randomized) version of the actual signal (which version is not Gaussian because of the phase randomization). It can be shown that the performance improvement achieved, at very little added computational cost, by using the cycloergodic model that Detector −1 uses can be substantial.

- **Remark 6**: Performance loss of minimum-mean-squared-error (MMSE) linear transformations

Consider the sum of two spectrally overlapping stationary ergodic signals, which are inseparable using a time invariant transformation, whether linear or non-linear. If these signals are modeled as non-cycloergodic their optimum linear time invariant processer can fail to separate them even though they may be completely separable using an almost periodically time varying linear transformation. Examples include double-sideband amplitude-modulated sine waves, pulse-amplitude-modulated periodic pulse trains, and amplitude-shift keyed sinewaves. The spectral redundancy in these signals enables either constructive or destructive linear combining of frequency-shifted subbands.

3. Cycloergodicity theory

In the tutorial paper [50], a thorough treatment of non-population probability theory of stationary, cyclostationary, and almost cyclostationary processes and a comparison with the Kolmogorov population-probability theory of stochastic processes is presented. Included in [50, Sec. 3.7] is a brief discussion of cycloergodicity that describes how to formulate cycloergodic theorems for all but one class of almost cyclostationary processes which are discrete-time processes with at least one period of cyclostationarity that is incommensurate with the time-sampling increment of the process. That discussion gives additional insight into the only preceding published treatments of cycloergodicity [3,9] and also expands the classes or processes treated from discrete time to continuous time and, in addition, lends powerful insight into the challenge of the one remaining class of processes for which no known method for developing a cycloergodicity theorem had yet been established. In this third (following [3] and [50]) broad treatment of the subject, all classes of processes exhibiting cyclostationarity, including those left out of previous treatments, are considered: these are discrete-time and continuous-time processes that are asymptotically-mean cyclostationary or asymptotically-mean almost cyclostationary, which includes those that are cyclostationary or almost cyclostationary and discrete-time processes that exhibit cyclostationarity or asymptotically-mean cyclostationarity with at least one period that is incommensurate with the time-sampling increment. These results, like those in [50, Sec. 3.7], are presented in discussion form, not in the form of formal mathematical/symbolic statements of theorems and proofs like the results in [3]; however, like [3], the discussion is detailed in terms of properties of the probability measures of the stochastic processes of interest, and in terms of the mathematical necessary and sufficient conditions (under typical assumptions) on the measures for existence of cycloergodicity. Yet, the presentation remains accessible to those with no substantive knowledge of measure theory and little if any knowledge of stochastic process theory beyond standard engineering introductions to stochastic processes. This goes farther than the classic treatment of ergodicity for engineers [2], which shares the goal here of accessibility for engineers but remains considerably more technical than the treatment given here. For references to related earlier work on ergodicity, readers are referred to, [3, p.106] and also, for mean-square cycloergodicity, [16;3, Sec. II]. Together with the other references in this present article, these additions complete the Author's bibliography of all previous work he has found that is directly or closely related to the topic of cycloergodicity (with probability = 1).

- *The pragmatic approach taken in this article that enables a general treatment of cycloergodicity including discrete and continuous-time processes and all the relevant types of cyclostationarity cited above is to use a notation P that is given alternative interpretations as either finite-dimensional CDFs or "infinitely" more abstract probability measures, as defined in Kolmogorov's definition of a stochastic process. The proofs of the theorems are considerably less abstract for finite-dimensional events and their probabilities than they are in [2] for general abstract events sets in sample space and their probability measures. But the concepts that are relevant to empiricists are the same. It is not an objective of this article to provide mathematically rigorous proofs of theorems. Rather the objective is to communicate in a language that is intelligible to empiricists wanting to solve real practical problems by facilitating an intuitive understanding of cycloergodicity and the necessary and sufficient conditions for a stochastic process model to have this property. Nevertheless, meaningful descriptions of proofs for all theorems included are provided or known proofs are cited in the literature.*

For example, in order to avoid material that requires substantive knowledge of mathematics well above the level of typical PhD graduates in science and engineering, this article takes the long-established result of *Birkhoff's Pointwise Ergodic Theorem* (the term *pointwise* means the desired equality between a time average and its expected value holds for every *point* in the sample space within a set of probability measure equal to 1—abbreviated herein by w.p.1), which most of the theorems included herein rely on, as a given. There are a number of claimed proofs of this theorem in the mathematics literature, two of which I have selected: [2,51]. The necessity part seems easy to prove by counterexample, but the sufficiency part is quite a challenge.

One more preliminary remark is needed before proceeding to the theorems. The question of how many incommensurate periods can be exhibited by the family of time-translated versions of a process's measure should be addressed. The answer is 0 or a countable infinity or anything in between. An example of 1 is the independently distributed discrete-time Bernoulli example already discussed. This example can be generalized to a probability parameter that is any almost periodic function with range in the closed interval [0,1]. So, this one example can exhibit a number of periods equal to any natural number, or countable infinity. Another example of a countable infinity is the mixture of periods that occur when two processes are added together as happens with interfering signals in radio communications. Here, it is easiest to work with the CDF of a discrete-time process, which is the expected value of the indicator function of the event set $S(t)$ for a process with sample paths denoted by $x(t,\omega)$

$$S(t) = \{\omega \in \Omega : x(t,\omega) \le y\} \tag{1}$$

The expected value of the (0,1)-valued indicator of this set equals the first-order CDF:

$$F_x(t,y) = E\{I_{S(t)}(\omega)\} = \int_{\Omega} I_{S(t)}(\omega)\, \mathrm{d}P(\omega) = \int_{S(t)} \mathrm{d}P(\omega) = \mathrm{Prob}\{\mathrm{x}(t,\omega) \le y\} \tag{2}$$

where

$$I_{S(t)}(\omega) = U(y - x(t,\omega)) = \begin{cases} 1, & x(t,\omega) \leq y \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

for which $U(z)$ is the unit step function. This step function $U(z)$ can be approximated by the sum of 1 and $Tanh(z)$, and $Tanh(z)$ admits a power series approximation for $\text{mag}(z) < \pi$, and this power series contains all odd powers. All the odd powers of $z = y - x(t,\omega)$ produce all even and odd powers of $x(t,\omega)$. Consequently, $U(y - x(t,\omega))$ appears to contain all even and odd order powers of $x(t,\omega)$. Therefore, the expected value of $U(y - x(t,\omega))$ contains all even and odd moments of $x(t,\omega)$. This suggests that, if $x(t,\omega)$ exhibits statistical cyclicity of some order with two or more incommensurate periods, then some of its CDFs will likely contain all integer-linear combinations of the reciprocal of the two periods, which comprises a countable infinity. By the classical Daniell-Kolmogorov Extension Theorem [52], the measure for a discrete-time stochastic process is uniquely determined by the set of finite-dimensional (finite order) CDFs. Therefore, we can expect some measures for some practical stochastic process models, for which the process itself contains only a few statistically cyclic terms, to exhibit a countable infinity of cycle frequencies. The Gaussian process can be used as a relatively tractable example to investigate because its CDFs are all functions of only first- and second-order joint moments of the process. Nevertheless, the dependence of the CDF on these moments is highly nonlinear, including the inverse of the covariance matrix. In contrast, the log joint characteristic function of any order for a Gaussian process, which contains exactly the same information as the joint CDF, is simply a linear plus quadratic function with both the mean and covariance matrix entering linearly. But this should be contrasted with the general infinite series expansion of the characteristic function in terms of joint moments for non-Gaussian processes. And even for a zero-mean Gaussian processes, the joint moments depend on potentially numerous products of the autocorrelation, depending on the order of the moment.

For reference further on in this section, the time-average counterpart of the stochastic-process CDF (1) – (3) is defined for n-th order as follows:

The event set of interest is, for each translation t of interest,

$$\widehat{S}(t) = \{t' \in R : x(t+t'+t_1) \leq y_1, x(t+t'+t_2) \leq y_2, \ldots, x(t+t'+t_n) \leq y_n\} \tag{1′}$$

where R is the set of real numbers (the analog of Ω) and the set $\widehat{S}(t)$ is defined for the interval of values of $t + t'$ inside [-V, V] for which the time series $x(t+t'+t_i)$ is defined for $i = 1, 2, \ldots, n$. The Fraction-of-Time CDF is the time average over the time interval [-V, V] of the (0,1)-valued indicator of this set:

$$\widehat{F_{\{x\}}}(t,\{y\}) = \langle I_{\hat{S}(t)}(t')\rangle_V = \int_R I_{\hat{S}(t)}(t')p_V(t')\,dt' = \int_{-V}^{V} I_{\hat{S}(t)}(t')\frac{1}{2V}dt'$$
$$= \text{FOT-Prob}\{x(t+t_1) \leq y_1, x(t+t_2) \leq y_2, \cdots, x(t+t_n) \leq y_n\} \tag{2′}$$

where $p_V(t')dt'$, the analog of $dP(\omega)$, is the time differential times the uniform PDF over the interval [-V, V] on which $x(t+t'+t_i)$ is defined for $i = 1, 2, \ldots, n$, and

$$I_{\hat{S}(t)}(t') = \prod_{k=1}^{n} U(y_k - x(t+t'+t_k)) = \begin{cases} 1, & x(t+t'+t_1) \leq y_1, \ldots, x(t+t'+t_n) \leq y_n \\ 0, & \text{otherwise} \end{cases} \tag{3′}$$

In the limit as $V \to \infty$, the dependence of the FOT-Probability on t vanishes, but this requires an infinite record of time series data. Nevertheless, in practice, t can be selected to maximize the amount of available data used in (2′), and then the left-hand side of (2′) can be treated as independent of t.

This definition of FOT-CDF produces a stationary model. For a T-cyclostationary model $\widehat{F}^{T}_{\{x\}}(t,\{y\})$, we make the replacement in (2′)

$$\int_{-V}^{V} (\cdot)\frac{1}{2V}\,dt' \leftarrow \sum_{t'=-\frac{(M-1)T}{2}}^{\frac{(M-1)T}{2}} (\cdot)\frac{1}{M} \tag{2″}$$

In this case, in the limit as $M \to \infty$, the dependence on t becomes periodic. For finite M, only the single period of the CDF in t in the center of the data record, which is the one that results from the largest number of non-zero terms in the sum of M terms, should be retained, and the CDF over one period in t should be periodically replicated every T units of time.

For an α-sinusoidal cyclostationarity model $\widehat{F^{\alpha}_{\{x\}}}(t,\{y\})$, we make the replacement in (2′)

$$\int_{-V}^{V} (\cdot)\frac{1}{2V}\,dt' \leftarrow \int_{-V}^{V} (\cdot)\frac{1}{2V}\exp\{-i2\pi\alpha t'\}\,dt' \tag{2‴}$$

and select the best values of t covering one period of the sinusoid and then replicate. Finally, for the almost cyclostationary model, we simply use the standard model proposed in [5] in terms of cyclostationary CDF components:

$$F^{\{T_q\}}_{\{x\}}(t,\{y\}) = F_{\{x\}}(t,\{y\}) + \sum_{q=1}^{Q}\left[F^{T_q}_{\{x\}}(t,\{y\}) - F_{\{x\}}(t,\{y\})\right] \tag{2⁗}$$

which is valid for stochastic processes (population probability) and non-population probability for finite and infinite time averaging, and Q may be infinite. Using Fourier series representations, the CS and ACS models can be expressed in terms of only the α-sinusoidal CDFs. More complete discussions of such CDFs for all orders and for traditional time averages as well as sinusoidally-weighted time averages, periodic subsampling averages, and almost periodic averages are given in [5,6,26] and thereafter in many references cited throughout this paper and, for finite V and M, in [7, page 3.5.3,8, Sec. 3.5.3,44].

For reference below, the classic *Fundamental Theorem of Expectation* is expressed here:

$$E\{g(\{x(t+t_1,\omega), x(t+t_2,\omega), \ldots, x(t+t_n,\omega)\})\} \equiv E\{\tilde{g}(t,\omega)\} = \int_{\Omega} \tilde{g}(t,\omega)\,dP(\omega) \tag{4}$$

for which g is some measurement function of the process at n time points with specified separations, abbreviated to $\tilde{g}(t,\omega)$. This theorem avoids finding the measure for g from the measure P for x, which can be a very challenging analytical task. As an alternative, the above can be expressed in terms of the n-th order CDF,

$$E\{g(\{x(t+t_1,\omega), x(t+t_2,\omega), \ldots, x(t+t_n,\omega)\})\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_n)\,dF_{\{x\}}(y_1, y_2, \ldots, y_n) \tag{5}$$

or its n-th order derivative, the PDF (Probability Density Function),

$$E\{g(\{x(t+t_1,\omega), x(t+t_2,\omega), \ldots, x(t+t_n,\omega)\})\} = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_n) f_{\{x\}}(y_1, y_2, \ldots, y_n)\,dy_1\,dy_2 \ldots, dy_n \tag{6}$$

which avoids finding the CDF or PDF for g *from the joint CDF for n time samples of x.* As we shall see, ergodicity theorems can be expressed in terms of the measure of the process or its CDFs in the case for which only a finite number of time points are of interest.

Also, for reference below, the not-so-well-known *Fundamental Theorem of Averages* [5] is expressed here

$$\begin{aligned}&\langle g(\{x(t+t_1,\omega),x(t+t_2,\omega),\cdots,x(t+t_n,\omega)\})\rangle\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(y_1,y_2,\ldots,y_n)\,\mathrm{d}\widehat{F_{\{x\}}}(y_1,y_2,\ldots,y_n,\omega)\\&=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(y_1,y_2,\cdots,y_n)\,\widehat{f}_{\{x\}}(y_1,y_2,\cdots,y_n,\omega)\,dy_1\,dy_2\cdots dy_n\end{aligned}\tag{7}$$

where the circumflex atop the CDF and PDF denotes Fraction-of-Time probability for a single sample path, not stochastic probability for the process, and the angle brackets denote time average instead of expectation. This theorem is valid for any finite time-interval, as well as in the limit as averaging time approaches infinity [7, page 3.5,8, Sec. 3.5]. This theorem is also valid for the FOT Probability models defined above for CS and ACS cases.

In Gray's book [2], the property of a discrete-time process having both ergodicity properties and *N*-ergodicity properties for $N > 1$ (cycloergodicity properties with period *N*) is referred to as *Total Ergodicity.* This may be a misnomer, because it does not account for cycloergodicity (defined below) of a discrete-time process for one or more periods *T* that are not commensurate with the sampling increment and each other. Also, the issue of translating a measure or an event set by some amount other than an integer multiple of the defining time index for a discrete-time process may be a new concept. I do not recognize this concept arising in Gray's treatment [2]. Because the techniques presented in the classical ergodic theory [2] are sufficient to derive ergodicity theorems for discrete-time vector-valued AMS measures, I adopt here two primary objectives:

- Objective 1: Explain how to reduce the objective of seeking cycloergodicity theorems for continuous-time AMACS (or ACS) processes to that of applying ergodicity theorems defined in the traditional manner for each of multiple derived AMCS (or CS) discrete-time processes via their vector-valued AMS (or S) representations.
- Objective 2: Explain how to reduce the objective of seeking cycloergodicity theorems for discrete-time AMACS (or ACS) processes with periods that are incommensurate with the process's time sampling increment to that of applying ergodicity theorems defined in the traditional manner for each of multiple derived AMCS (or CS) discrete-time processes via their vector-valued AMS (or S) representations, for which measure decomposition and interpolation and process interpolation of the original process are used.

Both of these objectives give rise to another apparently new concept in ergodicity theory, namely that of sinusoidally-weighted times averages and their ergodic properties. Although I and a PhD thesis student back in the early 1980s, the late Russell A. Boyles, did tackle the cycloergodicity problem for discrete time [3], we proceeded along the lines of the techniques in Gray's early work, which is now reported in his book and we found that any ACS (or AMACS) discrete-time process in the class of AMS processes is degenerate in the sense that all cyclic components (sinusoidally-weighted time averages) of expected measurements are zero unless the cycle frequency is commensurate with the reciprocal of the sampling increment. Yet, both new Cycloergodicity Theorems 4 and 5 presented below appear to be valid applications of methods presented in [2], even if the applications are novel. This result in [3] is a direct consequence of the fact that an infinitely long discrete-time average of a sinusoid using a time-sampling increment (candidate period) that is an integer multiple of the true period of the sinusoid yields the unique value of the sinusoid at all the sampling times, whereas if this ratio is not an integer, the average value will converge to some other value which, for a non-rational ratio, will be zero.

For completeness, new ergodicity and cycloergodicity theorems that avoid use of the difficult-to-prove Birkhoff ergodicity theorem also are presented. These theorems apply to the property of a stochastic process model that is here called *local ergodicity* in contrast to the classical *global ergodicity*. The term *global* as used here probably does not appear in the literature because it is needed only to distinguish from *local,* and the term *local ergodicity* (with the meaning here) does not, to my knowledge, appear in the literature. The necessary and sufficient conditions provided in these local (cyclo) ergodicity theorems are expected to generally be less of a mathematical challenge to verify.

3.1. Key to integration of new theory into classical ergodic theory

The quantities in classical treatments of the theory of ergodic properties for discrete-time processes with either stationary or AMS measures are all defined for only translations that are equal to the discrete-time process' time-sampling increment and integer (*N*) multiples thereof, which are the only possibilities for time-translation transformations to be measure preserving—a necessary condition for ergodicity. The crucial (as to be seen below) concept of *T-translation invariance* of event sets and their probabilities for *T*-cycloergodic properties, when *T* is incommensurate with the time sampling increment, is therefore outside of classical theory (which includes Nedoma's theory of N-ergodicity [9], which is cycloergodicity with period *N*.). Nevertheless, it is shown in this article that, as claimed in [50], the classical Birkhoff theorem can be applied by using one or the other of two crucial new lemmas that address what is called *Probability Measure Decomposition and Interpolation* and, indirectly, associated *Process Interpolation.* The first of these lemmas described below addresses the more straightforward case of continuous-time (CT) processes, and the second lemma addresses the slightly less straightforward discrete-time (DT) processes. It is noted that this Measure Decomposition, in which a measure is represented by an algebraic sum (with some negative terms) is unrelated to the classical and more abstract Ergodic Decomposition (cf [2]), in which a measure is represented in terms of a mixture of measures by the expected value of a random measure, which is a sum of probability-weighted measures with positive weights adding to 1.

- *The role of this new measure decomposition is that of being the unique tool for representing the almost periodic family of translated probability measures for a (possibly asymptotically-mean) almost cyclostationary process, which measure is NOT preserved by the time-translation transformation of the process (even with the technique of using vector-valued representations), by a set of periodic families of translated probability measures for cyclostationary processes, for which the time-translation transformation can be made measure preserving by using the previously known tool of vector-valued representation. This decomposition renders the almost cyclostationary process indirectly amenable to classic ergodic theory.*

There is an alternative to the periodic decomposition for obtaining component measures that are preserved under the time-translation transformation, and this is the *harmonic* or *sinusoidal decomposition,* also described below. However, the component measures in this decomposition are complex-valued measures instead of probability measures, but the complex-valued measures occur in the representation of the probability measure in conjugates pairs, the sum of any number of which harmonically related pairs—when added to the zero-frequency sinusoidal measure—is a periodic probability measure. The theory of complex-valued measures allows for representation of each complex component as a vector containing real and imaginary parts or a polar representation consisting of a magnitude-measure multiplied by a phase factor [53].

The approach proposed here is to use this sinusoidal decomposition and use time averages to estimate each component complex-valued measure of a specified event and then add the results to obtain an estimate of the almost periodic probability of the specified event (or similarly for the almost periodic expected value of a function), analogous to the proposed method for periodic component probability measures for periods that are commensurate with the sampling increment.

- *Because the real-valued magnitude of these complex-valued measures are preserved by the time-translation transformation (although the phase gets shifted in the phase factor), classical ergodic theory techniques can apparently be applied to derive the necessary and sufficient conditions for each component estimate to converge to its expected value. But, instead of a cycloergodicity test on the probability of translation-invariant event sets being needed, it is the probability of sinusoidally translation-variant event sets that must be tested.*

Because of the parallels between the periodic decomposition approach to cycloergodicity of almost periodic measures and the sinusoidal decomposition approach, only the former is spelled out in some detail below. Nevertheless, this sinusoidal-component approach to estimating moments of almost periodic CDFs representing almost periodic measures is commonplace in the literature on the theory of cyclostationarity, well represented by the references in this article, and dating back to the Author's seminal book on the subject [5]. Although the proof of the sinusoidal cycloergodicity theorem not given herein is less straightforward than the proof of the periodic cycloergodicity theorem given, the practice of estimating the sinusoidal components is just as practical as estimating the periodic components for every time-phase throughout a period as again illustrated in much past work on both periodic and sinusoidal component estimation.

Recall, as discussed above, that the quantity P occurring below, with its various notational embellishments, is a measure, which can be interpreted as an abstract measure defined for all events in the field of abstract subsets of the sample space of a Kolmogorov stochastic process model or as a finite-dimensional CDF defined for all values of its argument.

Lemma 1. *Periodic and Sinusoidal Decompositions for CT*

Consider any nonstationary family of probability measures $P(t)=T_t[P]$, indexed by t, obtained from translation of the event-set argument of a given measure P from a Kolmogorov process for all real numbers t, and recognize that the composition, P(t), of the measure with the translation operation, when evaluated at some event set in its domain, can be treated simply as a real valued function of the real variable t, and assume that for each event set, P(t) can be decomposed into the sum of an almost periodic component, when it exists, and a generally nonstationary residual that contains no sinusoidal components cf [1,46]. Because the residual vanishes in both of the limit averages of interest considered below, it is not of interest and not addressed below, and the notation P(t) is used below for the almost periodic component instead of the original generally nonstationary family of measures. (Because the average of the residual is zero, it has no impact on the ergodic properties of the process, which are fully determined by the average of P(t) which equals the average value of the almost periodic component of P(t) which is denoted by P^0 below.) In the following, all instances of the symbol P, regardless of superscript $\{\alpha\}$, $\{T\}$, 0, or none, represent a measure evaluated at some particular event set and is therefore a real or complex number or, where indicated, a real-valued or complex-valued function of the real time variable. Therefore, (8) below is a standard Fourier series representation for a continuous-time almost periodic real-valued or complex-valued function on the real line for which the union of sets of α over all event sets is assumed to be countable [1]:

$$\begin{aligned} P(t) &= P^{\{\alpha\}}(t) = def\sum_{\alpha}[P^{\alpha}\exp\{i2\pi\alpha t\}] \\ P^{\alpha} &= def\lim_{U\to\infty}\frac{1}{2U}\int_{-U}^{U} P(u)\exp\{-i2\pi\alpha u\}\,\mathrm{d}u \end{aligned} \tag{8}$$

In (8), the sum in the first line is over a possibly infinite but countable set of frequencies. By partitioning the set of frequencies α, called cycle frequencies here, into mutually exclusive subsets in each of which all cycle frequencies are harmonically related, we can re-express the above as follows:

$$\begin{aligned} P(t) &= P^{\{T\}}(t) = def\sum_{T}\left[P^{T}(t) - P^{0}\right] + P^{0} \\ P^{T}(t) &= def\lim_{K\to\infty}\frac{1}{2K+1}\sum_{k=-K}^{K} P(t-kT) \\ &= P^{\{m/T\}}(t) = \sum_{m=-\infty}^{\infty} P^{m/T}\exp\{i2\pi mt/T\} \end{aligned} \tag{9}$$

where the reciprocal of each period T is the generator for each of the subsets of harmonically related cycle frequencies, all of which are integer multiples of the fundamental frequency 1/T. The sum in the first line of (9) is over a possibly infinite but countable set of periods, T. (The reader should observe that the T-periodic averaging operation on RHS of (2′′) and the second equation in (9) are the same.)

Lemma 2. *Periodic and Sinusoidal Decompositions for DT*

Lemma 1 *for continuous time can be converted to the lemma desired here, for discrete time, by simply replacing the integral in the 2nd line of (8) with a discrete sum. The discrete-time P used in the so-modified 2nd line of (8), to calculate the Fourier coefficients $\{P^{\alpha}\}$, generates the continuous-time P(t) appearing everywhere else in (8) and (9) which is defined for all real t. Thus, this representation can be interpreted as a means of interpolating a discrete-time almost periodic function to obtain a continuous-time counterpart whose discrete time samples at the original times agree with the original almost periodic function and whose spectral content is identical. However, unlike continuous-time almost period functions in general, this spectrum is band-limited (or periodically replicated). By sampling any of the continuous-time periodic component measures, as in the 2nd line of (9), at discrete times that are commensurate with any selected period, the awkward situation of the first two lines of (9) being undefined for the original discrete-time P(t), when the sampling increment is not commensurate with the period, is circumvented. This process is referred to here as Probability-Measure Resampling. Its utility in practice is unknown, except for when the probability measure is represented by a set of CDFs of all orders. In this case, the above representations (8) and 9) of CDFs is a central part of the well-developed theory of ACS processes. But, for theoretical use for the measure itself, the lemma enables ergodicity theory proofs for discrete-time processes to be generalized to cycloergodicity, regardless of the relationship between the period of cyclostationarity and the time-sampling increment of the process.*

3.2. Ergodicity theorems

Before getting into the details of the new theorems, it is emphasized here that a process may be non-cycloergodic with a period that is not exhibited by the probability measure—a new concept in ergodicity theory: In other words, sample paths can exhibit statistical cyclicity even though the family of translated probability measures exhibits no periodicity with that period. This can happen as a result of random phase variables. For example, a random process that is a function (e.g., a sum or a product) of a cyclostationary process and a stationary process can be made stationary by adding a time-invariant random variable to the time parameter in the cyclostationary process or the sum of processes [12]. But the sample paths of the process will still exhibit statistical cyclicity due to the originally cyclostationary process. That is, sample paths of such a process or time-invariant functions thereof (such as a lag product) can, with probability equal to 1, contain finite-strength periodic components. The only requirement for such a process to be stationary is that the characteristic function of the random phase variable must be zero at the cycle frequencies in every expected measurement function (such as joint moments and CDFs). The periodic zeros of a sinc function (except at zero frequency) are quite useful here. For every uniformly distributed random phase variable in a sum of statistically independent such phases, we obtain a sinc factor in the characteristic function for the sum. So, all cycle frequencies in any expected measurement function of a poly-cyclostationary process (almost cyclostationary with a finite number of periods) can be annihilated with a single random phase variable equal to a finite sum of uniform random variables with properly chosen PDF widths [12]. (The case of an infinite number of periods may require investigation.)

As a prelude to the new cycloergodicity theorems presented below, the discussion here begins with known theorems of global ergodicity (where the unusual modifier *global* is used, as explained earlier in this section, to distinguish from Local ergodicity, a lesser-known concept, as defined herein).

Theorem 1. *Probability-1 Global Ergodicity Theorem for DT Stationary Processes (attributed to G. D. Birkhoff for pointwise (w.p.1) ergodicity, and paraphrased from https://en.wikipedia.org/wiki/Ergodicity) – A stationary probability measure on a sample space is ergodic for the measure-preserving integer-time-translation transformation if and only if the average over all translations of time-translated versions of each measurable function on the sample space converges to the expected value of that function, with probability measure equal to 1, for all measurable functions; and a necessary and sufficient condition for this ergodicity property is that there exist no translation-invariant subsets of the sample space not having probability equal to 0 (or, as a degenerate case, 1).*

Effectively, this means that the translated measurement function "never" (or w.p.0) equals a constant or it "always" (or w.p.1) equals a constant. The "never" case is typically desirable and the "always" case is uninteresting because the function of interest is then "non-random": it is a time-invariant constant with probability equal to 1. As examples of w. p.1 and w.p.0, if a time series is bounded below with Relative Lebesgue Measure equal to 1, (which is expressed as "for almost every time t") w. p.1 at value -100, then the event $x(t,\omega) < -100$ "never" (throughout the sample space) happens for almost every time t and the event $x(t,\omega) > -100$ "always" (throughout the sample space) happens for almost every time t.

Because the functions in Theorem 1 include event-set indicator functions, whose expected values equal the event-set probabilities, this theorem applies to estimation of event-set probabilities as well as to expected values of measurement functions.

In practice, if the experimenter can prove that the process model for $\{x(t, \omega)$: for all $t\}$ being used "never" produces a constant sample path of the t-translated event-indicator function for any non-zero-probability event set in the sample space, then the finite-time average of the indicator of this event will "always" (with probability equal to 1) converge to the probability of this event as averaging time approaches infinity.

The requirement that the time-translation transformation preserve the measure restricts this theorem to stationary processes. An important generalization treated in depth by Gray [2] extends this to Asymptotically-Mean Stationary (AMS) processes, for which the time-average of a possibly nonstationary measure converges, as averaging time grows without bound, to a stationary measure, which can be interpreted as the additive stationary component of the nonstationary measure, for which the residual—the difference between the nonstationary measure and its asymptotic mean (time average)—has a time average value that is identically zero. This stationary component is P^{α} for $\alpha = 0$ in (8) - (9) and it is said to *dominate* the nonstationary measure in the sense that the latter can equal 0 or 1 only if the former equals 0 or 1, respectively (because $P(t)$ can take on only values between 0 and 1). Less interestingly, P^0 also is dominated by the nonstationary measure $P(t)$ for a.e. t. Also, because the former is assumed to be nonstationary, the time-translation transformation does not preserve this measure, but it does preserve the stationary component. Because of this, the above classic theorem generalizes to the following:

Theorem 2. *Probability-1 Global Ergodicity Theorem for DT AMS Processes – An AMS probability measure on a sample space is ergodic for the integer-time-translation transformation if and only if the average over all translations of time-translated versions of each measurable function on the sample space converges to its own expected value, with probability measure equal to one, for all measurable functions, and a necessary and sufficient condition for this ergodicity property is that there exist no translation-invariant subsets of the sample space not having stationary-component probability equal to zero. Furthermore, the expected value of the time averaged measurement equals, with probability equal to 1, the time average of the nonstationary expected value of the measurement.*

The restriction to discrete time in Theorems 1 and 2 can be removed by representing a scalar-valued CT process by a vector-valued DT process. The vector dimension here is derived from the size of the set of all real times in whatever subinterval one chooses for partitioning the real line representing time into discrete-indexed contiguous subintervals, and this dimension is uncountably infinite (the number of time points in a continuous interval of the real line) in every case. Although the choice of interval length is arbitrary, in the case of interest in cycloergodicity properties with some specified period, choosing the subinterval length equal to that one period avoids the need for further decomposition of the derived vector-valued discrete time process, as discussed further below.

Theorem 3. *Probability-1 Global Ergodicity Theorem for CT AMS Processes – A CT AMS probability measure on a sample space with a vector-valued DT AMS probability-measure representation (with DT increment normalized to unity) is ergodic for the integer-time-translation transformation if and only if the average over all translations of time-translated versions of each measurable function on the sample space for the vector-valued representation converges to its own expected value, with DT probability (or stationary-component probability) measure equal to 1, for all measurable functions, and a necessary and sufficient condition for this ergodicity property is that there exist no translation-invariant subsets of the sample space for the vector-valued representation not having stationary-component probability equal to 0 (or, as a degenerate case, 1).*

For Theorems 1, 2, and 3, it is a simple matter to prove that the (0,1)-probability condition for global ergodicity is necessary, by simply assuming it is not true and demonstrating that the time average of the corresponding time-invariant event-indicator function cannot equal its expected value, the event probability assumed to be not equal to either 0 or 1. The proof of sufficiency is beyond the scope of this article, and readers are referred to the established literature (and wished good luck).

The above theorems address global ergodicity, which means ergodicity for an individual stochastic process for *all individual measurement functions and event sets*. This article does not address ergodicity "in mean square" which is a much weaker less restrictive alternative to "pointwise ergodicity with probability 1". However, it does discuss further below *Local Ergodicity* as a less restrictive alternative to global ergodicity.

Because CT and DT processes that are CS, AMCS, ACS, and AMACS are all AMS processes [3] (cf. Lemmas 1, 2), Theorems 2 and 3 are an essential part of the new probability-1 global cycloergodicity theorem for both DT and CT AMACS processes presented below. By observing, as already mentioned above, that the probability of any event set in the sample space of a process can be represented by the expected value of the (0,1)-valued indicator function for membership in the event set (cf. (1)-(3)), Theorems 1 – 3 apply to time-average estimation of probabilities of arbitrary event sets in the field of subsets of the sample space as well as to estimation of expected values of arbitrary measurable functions defined on the sample space such as, for example, any power of the process, the expected value of which is a moment of the measure (or of the process). Consequently, for efficiency, the following cycloergodic theorems also address only the estimation of expected values of functions using time averages (more specifically periodic time averages and sinusoidally weighted times averages) of those functions. Because all $2 \times 4 = 8$ classes of processes exhibiting cyclostationarity that are listed above are subclasses of DT or CT AMACS processes, we can capture all the cycloergodicity cases of interest in two theorems, one for CT AMACS and one for DT AMACS. Furthermore, since for all eight classes, Lemmas 1 or 2 are used below to replace the original processes of interest with vector-valued DT AMACS processes, one for each period of interest, and since every CT and DT AMACS process represented by a vector-valued DT process is AMS, we need only one generic theorem for DT AMACS processes, which leverages Theorem 2 generalized from scalar-valued to vector-valued DT AMS processes, with finite-dimensional vectors for original processes that are DT and infinite-dimensional vectors for original processes that are CT. However, there is one exception, which requires a corollary, and this is the case of an original DT process with one or more periods of cyclostationarity that is/are incommensurate with the sampling increment. The reason this case is an exception is because only in this case must we implement an interpolation processing step on the original stochastic process in order to define (and test for satisfaction of) the necessary and sufficient condition for cycloergodicity given by the upcoming cycloergodicity theorem. In practice, there may be only one or a few periods of cyclostationarity of interest, each of which would require a cycloergodicity test to be performed. However, for complete cycloergodicity of a process, as many as a countable infinity of tests for individual periods may (theoretically) need to be performed. This is part of the reason that the upcoming *Local Cycloergodicity* Theorems are of some interest. (The reader is reminded of Appendix II, which avoids "re-inventing the wheel" for (cyclo)ergodicity testing in some applications.)

The number of periods that may be of interest will generally depend on the particular measurement function whose expected value is to be estimated. There can be more periods for higher-order moments to be estimated than for lower order moments and different numbers for cumulants and moments; and there can be different numbers of periods for estimation of joint CDFs and joint characteristic functions (less for characteristic functions for Gaussian processes). This remark exposes the important fact mentioned briefly above that no process has a unique *cycle spectrum* (exhaustive set of cycle frequencies), because no process has a unique probabilistic model. Equally valid options include 1) the set of all moments or cumulants (countable for DT and uncountable for CT), 2) the CDFs for all finite sets of discrete-time samples (countable for DT and uncountable for CT), and 3) the set of all joint characteristic functions (uncountable for all DT and CT processes). The size of the cycle spectrum can vary widely from one to another of these optional models. There are also significant differences possible for local cycloergodicity. e.g., for a single function of a process, whose expected value is to be estimated, the relevant probabilistic model is only the mean of the function, which can contain anywhere from 0 to a countable infinity of cycle frequencies. The cycle spectrum of the characteristic function of any order for a Gaussian process can contain no more cycle frequencies than those in its mean and autocorrelation functions, while higher-order moments depend on sums of products of the mean and autocorrelations which create mixing of harmonics of the cycles in the mean and autocorrelation functions. Needless to say, testing for global cycloergodicity of ACS processes can be a huge challenge.

To be explicitly clear, cycloergodicity in general for one period does not imply and is not implied by cycloergodicity for any other (incommensurate) period. And, for a given period, cycloergodicity for one time point within a period does not imply and is not implied by cycloergodicity for some other time point in that period. This can be expressed alternatively as cycloergodicity for one vector-element in the vector representation does not imply and is not implied by cycloergodicity for any other element. Similarly, in general, cycloergodicity for one cycle frequency does not imply and is not implied by cycloergodicity for some other cycle frequency whether or not they are harmonically related.

For empirical work in time series analysis based on cyclic time average statistics, for which only one or a few functions are targeted for determining cycloergodicity, this limited number of targets addressed by what I shall call *local ergodicity* (in the function space or the sample space, not on the time line) may reduce the challenge of performing the necessary and sufficient cycloergodicity test(s). It is not clear in general whether each such local test for cycloergodicity is less of a mathematical challenge than is the global test for all functions, and the added complication of working with multiple periods and vector-valued processes instead of scalar-valued processes can only make matters worse. This pragmatic consideration is returned to in the following section providing a critical perspective on pragmatism.

In general, some prior information or testing of the stochastic process model of interest is needed in order to test for cyclostationarity at candidate periods. Once a candidate period of interest has been identified (see Appendix I), and events or functions of interest have been identified, the local cycloergodicity tests specified by the appropriate theorem below can be performed.

Despite the apparent advantage of local cycloergodicity for applications of interest in this article, for generality, both new global and local cycloergodicity theorems are presented below.

The next subsection begins with a new theory of local ergodicity, before advancing to the new theorems for local and global cycloergodicity.

Local Ergodicity

Theorem 4. *Probability-1 Local Ergodicity Theorem for DT and CT AMS Processes – Let {x(t, ω); for all t in R or all t in Z, for all ω in Ω}, with probability measure P defined on a field of subsets of the sample space Ω, be a Kolmogorov stochastic process [54] for which the class of measurement functions of interest, denoted generically by g(t, ω), is comprised of those functions of any set of time samples of the process, translated by all t, that satisfy the assumption that the limits of time averages over all t, denoted by <g(t, ω)>, exist w.p.1. This assumption for well behaved functions is satisfied by any of three axioms: 1) P is stationary, which is sufficient, or 2) P is AMS, which is necessary and sufficient, or 3) the sample paths of the functions of interest are Relatively Lebesgue Measurable (which means the univariate FOT-CDF of g(t, ω) exists w.p.1), which is sufficient but not necessary. For any such function g(t, ω), assume <g(t, ω)> = c for all ω in S which is a not-necessarily-proper subset of Ω, where c is independent of almost every t in R or Z (and independent of ω inside S), and assume that P(S) = 1. Then and only then we have, w.p.1,* $\langle g(t,\omega)\rangle = E\{\langle g(t,\omega)\rangle\} = \langle E\{ g(t,\omega)\}\rangle$, *which is the meaning of local ergodicity.*

Proof. To show that the assumption $P(S) = 1$ in this theorem is necessary is easily achieved by assuming $P(S)$ not $= 1$ and observing that a contradiction of $\langle g(t,\omega)\rangle = E\{\langle g(t,\omega)\rangle\}$ is obtained. To show that $P(S) = 1$ is sufficient, simply use the fundamental theorem of expectation to evaluate the expected value of the time average to discover that, when $P(S) = 1$, the expression reduces to the time average for all ω in S which means *w.p.1.*

This admittedly obvious local ergodicity Theorem 4, in terms of a condition on the constancy over the sample space of the average of the function, merits comparison with the global ergodicity Theorems 2, 3, in terms of the constancy of the function itself over the sample space. These two theorems are definitely not equivalent. The assumptions required by Theorem 3 for global ergodicity are not necessary for the assumptions of Theorem 4 for local ergodicity. However, they are sufficient. This sufficiency of Theorem 3 conditions for global ergodicity for the validity of conditions of Theorem 4 for local ergodicity is also logically equivalent to the sufficiency of violation of the conditions of Theorem 4 to ensure the violation of the conditions for Theorem 3. These statements are easy to prove by example.

If the function of interest (not yet subjected to time translations) in Theorem 4, is an event-indicator function and is finite-dimensional (involves the process at only a finite number of times) and its probability can be expressed in terms of CDFs, then the abstract probability measure used in the (0,1)-probability test can be replaced with the an expression in terms of these CDFs. Similarly, if other types of functions of interest are finite dimensional, their expected values can be expressed in terms of CDFs instead of probability measures. Theorem 4 and this observation appear to bring the task of testing for ergodicity more down to earth, but "the devil remains in the details". Unfortunately, the following generalizations to cycloergodicity take us in the wrong direction with regard to pragmatism, by further complicating the tests the user needs to perform.

3.3. The AMACS global cycloergodicity theorem

Theorem 5. *Probability-1 Global Cycloergodicity Theorem for DT and CT AMACS Processes – By virtue of Theorems 2 or 3, and Lemmas 2 or 1, respectively, we have the fact that the T-periodic time averages, as per (2′)-(2″) of a measurement function on an AMACS CT or DT process, converge as averaging time increases without bound to their own expected value, with DT probability (or stationary-component probability) measure equal to 1, and a necessary and sufficient condition for this cycloergodicity property is that there exist no translation-invariant subsets of the sample space for the vector-valued representation not having stationary-component probability equal to zero.*

The uninteresting case of having a stationary-component probability of a translation invariant event equal to 1 is simply ignored because this is obviously not a useful stochastic process model.

Regarding the special case for which a period of (AM) cyclostationarity for a DT process is incommensurate with the process sampling increment, Theorem 5 applies as well as it does when this situation is not present. However, the testing of the necessary and sufficient condition is more onerous. When the situation is not present, one must seek translation-invariant event sets of the vector representation OR T-translation-invariant event sets of the original process, which are periodically time-varying translated events sets, and the probability $= 0$ test is applied to any such event set.

When the family of discrete-time translated probability measures is almost periodic and has non-zero periodic components with periods that are incommensurate with the discrete sampling-time increment, Lemma 2, establishing the probability-measure resampling technique, is immediately applicable and replaces the given discrete-time family of translated probability measures with a resampled version. Nevertheless, it is only the stationary component of the probability measure that must be tested for probability $= 1$, and the stationary component can be seen from Lemma 2 to be independent of any resampling. That is, the stationary component of the interpolated process equals the stationary component of the original process (stationarity is preserved by this method of resampling). So, it is concluded that resampling of a probability measure is unnecessary in the cycloergodicity theorem even for a period that is incommensurate with the time-sampling increment of the process. Resampling is simply a conceptual tool for proving the theorem by application of Theorem 2. However*, the translation amounts to be used in the test are not commensurate with the original time sampling increment. This wrinkle appears to require that the process itself be interpolated.* Because an ACS or AMACS process is generally not almost periodic, the interpolation method of Lemma 2 for measures cannot be used here. Instead, it is proposed that the Nyquist sampling theorem be used to interpolate the process with sinc pulses and then the process be resampled in synchronism with the period of interest. It is unclear at this time how much more challenging this special case can be in practice.

A potentially useful tool, at least conceptually, for checking a stochastic process model for ergodicity or cycloergodicity properties is a set of *mixing conditions*. However, in the practice of probabilistic analysis of time-average statistics by empiricists, these sufficient conditions may still, like the tests specified by the above cycloergodic theorem, not be practical. Nevertheless, for at least conceptual benefit, readers are referred to the sufficient conditions in Theorem 3.2 and equation (3.20) in [3] for discrete-time ACS processes not including periods of cyclostationarity that are incommensurate with the sampling increment. It can be seen that the examples in the present article of stochastic processes that are non-cycloergodic due to meeting the sufficient conditions (for non-cycloergodicity) of time-invariant random amplitude factors and time-invariant random time shifts in components of processes, which are equivalent to random phase shifts in the argument of a sinusoid, violate these mixing conditions, even though they are not necessary conditions for (for non-cycloergodicity). It may be true that specific behavior causing any non-cycloergodicity other than these random amplitude factors and random time shifts (and a few related special cases identified in Appendix II) can be relatively complicated(?). (These mixing conditions involve a different mixing concept than that of mixture processes, which are always non-(cyclo)ergodic, though they may be mixtures of ergodic processes (Appendix II)).

3.4. The AMACS local cycloergodicity theorem

To complement Theorem 5, as discussed earlier in this section, we have Theorem 6.

Theorem 6. *Probability-1 Local Cycloergodicity Theorem for DT and CT AMACS Processes – Let{$x(t, \omega)$; for all t in R or all t in Z, for all ω in Ω}, with probability measure P defined on a field of subsets of the sample space Ω, be a Kolmogorov stochastic process [54] for which the class of measurement functions of interest, denoted generically by $g(t, \omega)$, is comprised of those functions of any set of time samples of the process, translated by all t, that satisfy the following assumption: For any period T of the family of translated measures, P(t), let $\mathbf{g}_T(t, \omega)$ with discrete t be the vector-valued representation of $\mathbf{g}(t, \omega)$ with continuous t and assume the limits of the discrete-time averages over all discrete t, denoted by $<\mathbf{g}_T(t, \omega)>$, exist w.p.1. This assumption is satisfied if and only if P(t) is AMCS with period T, which is necessary and sufficient since it is equivalent to the periodic component $P^T(t)$ from Lemmas 1 or 2 being AMS.*

For any such function $\mathbf{g}_T(t, \omega)$, assume $<\mathbf{g}_T(t, \omega)> = \boldsymbol{c}$ for all ω in S which is not-necessarily a proper subset of Ω, where the vector $\boldsymbol{c}$ is independent of almost every t in Z and independent of ω inside S, and assume that $P(S) = 1$.

Then and only then we have, w.p.1, $<\mathbf{g}_T(t, \omega)> = E\{<\mathbf{g}_T(t, \omega)>\} = <E\{\mathbf{g}_T(t, \omega)\}>$, which is the meaning of local cycloergodicity.

Proof. To show that the assumption $P(S) = 1$ in this theorem is necessary is easily achieved by assuming $P(S)$ not $= 1$ and observing that a contradiction of $<\mathbf{g}_T(t, \omega)> = E\{<\mathbf{g}_T(t, \omega)>\}$ is obtained. To show that $P(S) = 1$ is sufficient, simply use the fundamental theorem of expectation to evaluate the expected value of the time average to discover that, when $P(S) = 1$, the expression reduces to the time average for all ω in S which means *w.p.1*.

4. A perspective on cycloergodicity

The research results and associated material from the previous sections of this paper lead us to conclude that the objective of probabilistic analysis of time-average statistics that is generally pursued by experimentalists and empiricists working in time series analysis is not met by the stochastic process model, in part because of the practical failures of the ergodic hypothesis concept. This key link between abstract probability and empirical reality is seriously flawed from a practical standpoint. The purpose of the following discussion is to reflect on this in a constructive manner, by clearly exposing the serious issues that exist with application of ergodicity theory and then drawing attention to what can be seen to be a natural solution to this problem: A solution for which an extensive technical book of the past [5] was devoted to teaching. Four decades hence, we see from the literature that this natural solution has been further developed and is now quite mature but, except for a few disciples and contributors, has apparently been ignored by the entire field of mathematical statistics and apparently all university curricula involving mathematical statistics. This is so despite the prevalence of time-series analysis throughout science and engineering!

The Kolmogorov specification of a stochastic process and the Birkhoff Ergodicity Theorem are generally so abstract that, to my knowledge, no meaningful methodology for implementing the ergodicity test that could be useful for an empiricist except in the most trivial examples has been developed. Gray [2] provides a uniquely accessible treatment of ergodic theorems for the non-ergodic-theory specialist but, for the typical empiricist, I fear it is not accessible enough. Fortunately, past work by mathematicians, extended from ergodicity to cycloergodicity by the Author, has resulted in the identification of certain specific classes of stochastic process models that are cycloergodic and this is of definite help here—see Appendix II. But this list is not comprehensive, so the challenge remains for all process models not in this list or any expanded version of it that surfaces.

When complemented by Theorems 3–6 above, the book [2] represents, better than any other source I know of, the state of knowledge about ergodicity and cycloergodicity in applied fields, including especially engineering but also the sciences. It provides the perfect backdrop for the following alternative perspective on the role of stochastic processes and ergodicity theorems in empirical time series analysis. For this reason, I have included below just enough excerpts from the book's preface to enable the reader to benefit from this key source in understanding the present article.

The author's intent in writing the book [2] is stated to be to provide: "a reasonably self-contained advanced (at least for engineers) treatment of measure theory, probability theory, and random processes, with an emphasis on general alphabets and on ergodic and stationary properties of random processes that might be neither ergodic nor stationary."

The author further explains that, since in the first in 1987of multiple versions of this book [2], "The intended audience was mathematically inclined engineering graduate students and visiting scholars who had not had formal courses in measure theoretic probability or ergodic theory. Much of the material is familiar stuff for mathematicians, but many of the topics and results had not then previously appeared in books."

He reveals that his "Personal experience indicates that the intended audience rarely has the time to take a complete course in measure and probability theory in a mathematics or statistics department, at least not before they need some of the material in their research."

And he adds the following remark, especially relevant here: "Many of the existing mathematical texts on the subject are hard for the intended audience to follow, and the emphasis is not well matched to engineering applications. A notable exception is Ash's excellent text [55], which was likely influenced by his original training as an electrical engineer. Still, even that text devotes little effort to ergodic theorems, **perhaps the most fundamentally important family of results for applying probability theory to real problems**." (Author's emphasis with boldface added.)

I congratulate the author of [2] for giving us this unique book, which I suspect has enabled many, like me, to gain an understanding of ergodicity theory that we would be hard pressed to find anywhere else in a treatment that respected our lack of education or at least training in measure theory. I believe it is fair to paraphrase his final remark quoted immediately above as follows: *Real problems, such as those involving the application of population probability to analysis of time-average statistics derived from empirical time series data require a quantitative link between time averages and expected values and for this reason ergodicity theorems are of fundamental importance to such real problems.* My response to this author's perspective in [2] is twofold: 1) If, indeed, ergodicity theorems are of fundamental importance in real time-series analysis, I am perplexed by the observation that the content of this book falls short of meeting its objective by virtue of having excluded the Cycloergodicity Theorems presented in the present paper, given the spectacular growth of research & development and commercial use of signal processing algorithms based on time-average statistics for (almost) cyclostationary signals throughout the many fields of engineering and science since the appearance of the original and seminal book [5] on this topic nearly 40 years ago, and especially in the field of communication systems.

The second and more important response, given that we now have the missing cycloergodicity theorems, is 2) *If the non-population probability theory based on time averages is used to perform the needed probabilistic analysis as demonstrated in* [5] *40 years ago and in many follow-on publications cited below and described very recently in [48,50], instead of the expected values from the orthodox population probability theory associated with the stochastic process model, then for any such application there would be no need for ergodicity or cycloergodicity theorems and therefore no need to face the challenge of performing mathematical tests of the (cyclo)ergodic hypothesis.*

The simplest case referred to here is that for which the cyclostationary process model is for discrete time data and has period of cyclostationarity equal to an integer multiple of the time-sampling increment in the model, which is exactly equivalent to a finite-dimensional vector-valued stationary process, for which the standard ergodicity theorems apply, as explained in [2]. But this idealized model excludes all the more generally applicable models for which a) the period of cyclostationarity is incommensurate with the time-sampling increment, and/or b) there are two or more mutually incommensurate periods, and/or c) the stochastic process is of continuous-time type and either of one period or two or more incommensurate periods are present. Key applications in communication systems design and analysis requiring these more realistic cyclostationary signal models are briefly discussed and cited in the section *Background on Cycloergodicity*. The cycloergodicity theory needed for these more practical models is given in the preceding section.

Before continuing, a few more words about what may be an exaggerated statement about the importance of ergodicity theory is merited here. If the meaning in [2] of "real problems" is restricted to applications that specifically call for a stochastic process model of time series data while, at the same time, these applications involve time-average statistics obtained from an actual (real) time series that specifically calls for modeling as one member of some population of time series of interest, consisting of realizations of the stochastic process, then the perspective from [2] on the fundamental importance of ergodicity theory for applying probability to real problems is logical.

However, if the real problem of interest consists of empirical time series analysis based on time-average statistics motivated by real science or real engineering and there is no real population of time series, positing a stochastic process model consisting of a population of hypothetical time series for the sole purpose of performing probabilistic analysis of the empirical time-average statistics is *fundamentally non-scientific* because it egregiously violates the principle of parsimony. This abstract model of a real problem would be scientifically acceptable if and only if there were no more realistic alternative probability theory available for the purpose. This assumption, which may have been valid around the middle of the last century, was rendered invalid in the mid-1980s by the introduction in [5] of a comprehensive non-population probability theory of time series based on time averages, assuming the time series arise from phenomena or systems in equilibrium, for which the key characteristics do not change with time or, more generally, cycle with time with single or multiple possibly-incommensurate periods. These lead to probabilistic models of single time series that are stationary, cyclostationary, or almost cyclostationary, and time-dilated versions thereof.

Abstract Population Probability does exhibit some convenient mathematical properties by virtue of being axiomatically willed, but the absence of these properties in non-population probability is a reflection of closer ties with empirical reality, which again renders the former non-scientific as argued in [47]. Its virtues are valuable for the non-empirical field of ergodic dynamical systems but are of questionable appropriateness for empirical time-series analysis.

The treatment of *Non-Population Probability Theory* of time series that began being promoted in earnest with a book [5] published in the same year, 1987, as the first version of [2], explicitly expresses the considered opinion that *the stochastic process is simply the wrong tool for probabilistic modeling of time averages* when real populations of time series are not available or cannot even exist or when they are simply not of interest. By setting aside the stochastic process tool and adopting in its place a comparable and highly analogous probability theory of time averages, as done in [4–8,18–50], the difficult and mathematically challenging concept of (cyclo)ergodicity of stochastic processes becomes irrelevant. Even at this late date, almost 40 years since publication of the proposed probability paradigm shift in the book [5], this statement will likely be shocking to all those who have 1) been brought up believing that the stochastic process is the only tool available for doing probabilistic analysis of time series data and 2) have not read [5] or any of the considerably more the 40 above-cited subsequent publications that were spurred by [5] and that used and usually promoted adoption of the non-population probability model.

As unorthodox as this alternative theory is at present, there is overwhelming evidence that the non-population probability theory of time averages is a better fit, from a scientific perspective, for probabilistic characterization of time-average behavior of time series data. In fact, several of our most accomplished and highly recognized scientists and engineers of the recent past, in directly relevant fields including even Information theory, "saw the writing on the wall" in 1987, when the seminal book [5] appeared. Four striking examples of their perspective in favor of the non-population probability alternative are presented in [50] and are worth repeating here:

The late information theorist Professor James L. Massey (recipient of the Shannon Award in 1988, the IEEE Alexander Graham Bell Medal in 1992, the Marconi Prize in 1999, and the Information Theory Society Distinguished Service Award in 2004, and elected to the U.S. National Academy of Engineering, the Swiss Academy of Engineering Sciences, the European Academy of Sciences and Arts, and the Royal Swedish Academy of Sciences) wrote in a prepublication review of the book [5]

- "*I admire the scholarship of this book and its radical departure from the stochastic process bandwagon of the past 40 years.*"

Complementing this bold statement, the late Professor Enders A. Robinson, who revolutionized exploration geophysics as the originator of the Geophysics Analysis Group at MIT, author of over 20 advanced technical books focused on time-series analysis and probability theory, and a member of the National Academy of Engineering and the European Academy of Sciences, states in a 1987 letter of reference:

- "*Professor Gardner has the ability to impart a fresh approach to many difficult problems. William is one of those few people who can effectively do both the analytic and the practical work required for the introduction and acceptance of a new engineering method. His general approach is to go back to the basic foundations and lay a new framework. This gives him a way to circumvent many of the stumbling blocks confronted by other workers . . .*
- *I am particularly impressed by the fundamental work in spectral analysis done by Professor Gardner. Whereas most theoretical developments make use of ensemble averages, he has gone back and reformulated the whole problem in terms of time-averages. In so doing he has discovered many avenues of approach which were either not known or neglected in the past. In this way his work more resembles some of the outstanding mathematicians and engineers of the past. This approach took some courage, because generally people tend to assume that the basic work has been done, and that no new results can come from re-examining avenues that had been tried in the past and then dropped. William's success in the approach shows the strength of his engineering insight. He has been able to solve problems that others have left as being too difficult.*"

In a 1990 published review of the book [5], Professor Robinson wrote:

- *"This book can be highly recommended to the engineering profession. Instead of struggling with many unnecessary concepts from abstract probability theory, most engineers would prefer to use methods that are based upon the available data. This highly readable book gives a consistent approach for carrying out this task. In this work Professor Gardner has made a significant contribution to statistical spectral analysis, one that would please the early pioneers of spectral theory and especially Norbert Wiener''*

With similar sentiment, the late Professor Ronald N. Bracewell, recipient of the IEEE Heinrich Hertz medal for pioneering work in antenna aperture synthesis and image reconstruction as applied to radio astronomy and to computer-assisted tomography (CAT-scans), in his Foreword to the book [5], makes essentially the same point that Professor Robinson makes:

- *"If we are to go beyond pure mathematical deduction and make advances in the realm of phenomena, theory should start from the data. To do otherwise risks failure to discover that which is not built into the model . . . Professor Gardner's book demonstrates a consistent approach from data, those things which in fact are given, and shows that analysis need not proceed from assumed probability distributions or random processes. This is a healthy approach and one that can be recommended to any reader''*.

The late Akiva M. Yaglom, physicist, mathematician, and professor on the Faculty of Probability Theory at Moscow State University, Russia, member of the USSR Academy of Sciences, recipient of the Lewis Fry Richardson Medal for "eminent and pioneering contributions to the development of statistical theories of turbulence, atmospheric dynamics and diffusion, including spectral techniques, stochastic and cascade models", who specialized in theories of stochastic processes, stated in a published review of the book [5]:

- *"It is important . . . that until Gardner's . . . book was published there was no attempt to present the modern spectral analysis of random processes consistently in language that uses only time-averaging rather than averaging over the statistical ensemble of realizations [of a stochastic process] . . . Professor Gardner's book is a valuable addition to the literature"*

The late Phillip. E. Doak, Founding Editor of the *Journal of Sound and Vibration*, with a tenure as Editor in Chief of 40 years, on 8 March 1990, sent the Author his perspective on non-population probability:

- "In my latter years, I have become more and more convinced of the validity of his [Percy W. Bridgman, Nobel Prize Laureate] outlook. Not only can ergodic mathematical concepts put students off, indeed

I now believe that for physical scientists and engineers, they are "operationally erroneous'', and dangerous to mental health. Interpreting observations through ergodic spectacles is to misinterpret what the observations really mean. Not only does it confuse the issue, but also it inhibits the development of one's intellectual capacity to ask the right questions about what the data means. Thus, in design, development, and research it is a model of reality which is counterproductive in respect to generating concepts which can lead to real progress in the real world''

The substantial evidence cited above and otherwise presented throughout 40 years of the Author's previous work cited herein and references therein to the effect that ergodicity theory is, to quote the editor Doak in [50] who is reflecting on words from Nobel Prize Laureate Percy W. Bridgman, "operationally erroneous to physical scientists and engineers" and "irrelevant to empirical work in time-series analysis*" directly opposes Gray's statement* ". . . ergodic theorems [are] perhaps the most fundamentally important family of results for applying probability theory to real problems" or, at the very least, calls into question what Gray means by "real problems". The meaning certainly does not appear to be particularly inclusive.

Returning now to the absence of cycloergodicity theory in Gray's otherwise powerful book, it is conceivable that this oversight might have been facilitated by Gray's apparent lack of interest in unorthodox Non-Population Probability theory. For example, there appears to be no sign of research or publication by Gray on this topic (excluding the trivial case of N-ergodicity). This explanation is proffered because it was along the path of *Non-Population Probability Theory* development that the long-missing Cycloergodicity Theorems and the unorthodox approach to deriving them that are surfaced in this article were conceived of.

- *I claim that the reason for this direct opposition between two perspectives is Gray's choice, in concert with the great majority of contributors to probabilistic analysis of time series data, of Kolmogorov's now orthodox Population Probability model for ALL applications involving probabilistic analysis of time series data.*

  This opposition to the current paradigm is focused on applications for which 1) the real problems of interest arise from empirical time-series analysis based on time-average statistics and 2) populations of time series are not of interest in and of themselves and do not even exist as part of the real problem.

- *I also claim that the reason these population probability models are almost invariably introduced in time series analysis problem formulations is the failure of our education system to teach our students that the ubiquitous population probability model is NOT THE ONLY OPTION FOR PROBABILISTIC ANALYSIS:*

  This is a major oversight in education and merits the serious attention of university faculties involved with time series analysis toward corrective action.

- *It is conceivable that a major historical influence that arose in the mid-1970s may have precipitated the wholesale rejection of the then-nascent time-average approach to probability modeling, and this source of influence is the popular book on time series analysis [56] (reprinted in the SIAM book series Classics in Applied Mathematics) by David Brillinger. As explained in [17], Brillinger considers the ergodicity condition to "not be overly restrictive for our purposes [time series analysis]", dismisses the FOT-Probability model (which he calls the "functional" model) for being mathematically equivalent to the stochastic process model via ergodicity, and then proceeds to present an excellent treatise on stochastic-process modeling for time series analysis applications. The quality of this book in every other respect, might well be responsible for the apparent broad acceptance of his dictum. The companion article [17] provides a mathematical proof, based on existing published work, that the two models are not mathematically equivalent, and the difference has substantive practical significance. A more comprehensive and up-to-date perspective on this needed paradigm shift is addressed in the forthcoming article [17], as a follow-up to the recently published first phase of a renewed (since [5] in1987) call for a paradigm shift [48,50].*
- *And, finally, I claim that, if the partial (meaning specifically for only empirical time-series analysis involving time average statistics) paradigm shift initially proposed in [5], renewed in [50], and further supported in [48] and this present article and its companion [17] eventually takes place, we must conclude that ergodicity is irrelevant in this field of study.*
- *Under the proposed partial paradigm shift, the new cycloergodicity theorems presented herein, as well as their long-standing classical counterparts, will become of little interest in the field of empirical statistical time series analysis based on time-average statistics, because the Non-Population Probability theory of time-average based statistics renders stochastic processes and ergodicity/cycloergodicity theorems irrelevant.*

When populations of time series data do exist, sample averages from an ensemble of time series should be and typically are used; in this case time averages as well as ergodicity again become at least possibly irrelevant.

Still, in fairness to the field of ergodicity theory more generally, non-empirical theoretical work on stochastic process models exhibiting the various forms of cyclostationarity, for example, information theory involving interfering signals, have benefited and may continue to benefit from cycloergodicity theory. Yet there exists an open question as to what parts of an expanded information theory might be amenable to non-population probability? The channel coding theorem, for example, may not be a candidate for non-population probability because the use of random channel codes in its proof produces non-ergodic signals at the channel output. Nevertheless, the use of random codes might not be required for proving this theorem [57]. Jacob Ziv's celebrated work on his "Individual-sequence approach to information theory" and "universal data compression" is said to have had a profound impact on shaping the landscape of information theory and its applications" [58]. The cycloergodic theorem for ACS processes as well as the FOT-probability theory of ACS time series could be useful in needed future work in this field.

As a summary remark regarding the discussion in this section and the associated comments throughout this article about the irrelevance of ergodicity theorems to empirical work in time-series analysis based on time-average statistics, one, but by no means all, of the key points made herein concerns the challenges posed by the need to implement the tests specified for the types or ergodicity addressed and discussed in detail in the previous section of this article: If the (cyclo)ergodicity tests cannot be implemented for models motivated by empirical work, they are for most practical purposes irrelevant to such work. It is conceivable that many empiricists might find they do not possess the mathematical knowledge/skills to undertake this mathematical work and or they might find that their posited stochastic process models are mathematically too vague to allow for this work to be carried out. Moreover, even if (cyclo)ergodicity testing is facilitated by this article and can be mechanized to some extent (see Appendix II), the approach can be argued to be unscientific because of its unnecessary abstraction and associated requirement for conceptual contortions to accommodate the reliance on population probability when there is no population.

With regard to circumventing tests for cycloergodicity of stochastic processes by seeking refuge in the *ergodic hypothesis*, that is, by simply wishfully presuming all requirements of the theorems are met, this should be judged scientifically poor practice because, although ergodic theorems render this hypothesis falsifiable in principle, falsification may not be within practical reach as explained above. Fortunately, empiricists can use a *cheat sheet* that can considerably facilitate avoidance of the ergodic hypothesis. Such a tool consists of a list of (ideally) broad classes of stochastic process models that are known to be (cyclo)ergodic or non(cyclo)ergodic. This can to some extent mechanize the needed testing. I have not seen any extensive examples of such a tool but, considering potential utility, I think there might be some in the literature. To encourage the community to establish/standardize such a tool, I started a list and put it in Appendix II.

Conclusion: The perspective that ergodicity theory for probabilistic analysis in empirical work on time series analysis based on time-average statistics is, in most cases, irrelevant appears to be unavoidable, while utility of ergodicity in the theory of stochastic processes unrelated to empirical time-series analysis is a different issue or is a non-issue, but one that is of no relevance to the field of study focused on here. With historical reflection, it seems to me a fair assessment that Kolmogorov gave us a model that has proven itself invaluable as the seed of a magnificent mathematical theory of stochastic processes with unbelievably diverse applications, with one especially notable exception: The objective of performing probabilistic analysis of the statistics of engineering and science time series data based on times averages of various measurement functions has not been at all well served, regardless of the help offered by Birkhoff's ergodicity theorem. Its strengths for mathematicians come at the cost of its weaknesses for empirical time series analysts. This theme is pursued further in the companion article [17], where more analysis and logic leads the Author to propose a paradigm shift in the relevant subfield of mathematical statistics.

A final topic meriting brief discussion in this perspective, regarding ergodicity and cycloergodicity, is the role played by all-pervasive numerical evaluations of statistical inference algorithms for time series data processing using computer-based simulations relying on pseudo noise as the source of randomness in the time series of interest. Typically, today, probabilistic analysis leading to proposed signal processing algorithms for statistical inference is based on population probability: stochastic process models of signals and noise. Numerical evaluation of performance metrics for such algorithms is conducted by replacing the mathematically defined stochastic processes from the analytical model with possibly processed pseudo noise sequences. Users may assume the repeated (Monte Carlo) trials of their simulation experiment use a genuine ensemble representing random samples drawn from a probability space, but this is not how a pseudo noise generating algorithm works because we know very little about practical generation of ensembles of actual random samples. Not only are these conceptualized random samples often not part of the real world, but they're apparently not even amenable to creation for testing other than by recording physical noise of an appropriate type (e.g., thermal noise) and digitizing it.

Pseudo noise algorithms typically produce phony ensembles of random sample paths by selecting subsegments of a single long periodic time series produced by a pseudo noise generator using a deterministic algorithm. That is, the so-called ensemble averages performed from Monte Carlo trials of an experiment are actually time averages! But this is good, because it is usually time averages that will be performed when a statistical inference algorithm is implemented in practice in the field of statistical signal processing. But then, what is the point of using stochastic process models? They generally do not relate to either the real-world implementation of a signal processing algorithm or the Monte Carlo testing.

If the models used can be proven to be ergodic or cycloergodic, their expected values might analytically approximate the time averages used in practical implementations—depending on the fidelity of the posited stochastic process model—or they might approximate the Monte Carlo simulations. But do we want the expected values to match the real-world's time averages or the Monte Carlo averages? Perhaps it depends on whether we want to do engineering and science or publish papers. One thing is sure: If the work to validate the (cyclo)ergodic hypothesis is not done, the analytical work based on stochastic processes is scientifically irrelevant in the class of problems addressed in this article.

4.1. Alternatives to narrowly defined concepts

In the Real World, narrowly defined concepts, properties, and associated theories have their limitations. Here are a few ways to broaden the concepts discussed in this article.

1) Ergodicity is inherently an event-by-event property or a function-by-function property. Any event in sample space can be translation invariant and have probability equal (or unequal) to 0 or 1 independently of all non-intersecting events. Similarly, any function can be translation invariant and have a time average that equals (or doesn't equal) its expected value independently of many other functions. This is *Local Ergodicity*, which is defined in the previous section (local in the sample space or in the space of measurement functions on the process, not local in time).
2) A translation invariant event can have probability that is small enough (we cannot use the term "close to zero") to say the process is *nearly ergodic* for that event, and if all translation-invariant events for a specific model have probabilities that are small enough, the process can be said to have *globally nearly ergodic probabilities*. Similarly, a translation invariant function can have a time average that sufficiently closely approximates its expected value to say *the expected value of that function is nearly ergodic* and if all translation invariant functions have time averages that sufficiently closely approximate their expected values, the process can be said to have globally nearly ergodic expected values of functions. While the latter concept for functions appears to be viable for both local and global ergodicity, the former concept seems less generally viable for local ergodicity because it corresponds to the situation where all event sets of interest are highly improbable.
3) Time averages of sample path functions for which the averaging times are long enough to closely approximate their expected values can be said to reveal functions having *Temporally Local Ergodicity*. Because limits are not involved here, this can be considered an ergodicity-like property. Similarly, yet distinctly, time averages of functions of data for which the averaging time is much shorter than the length of a data record but long enough to closely approximate the time average over the entire data record can be said to reveal *Empirically Stationary Functions.* For a function whose expected value is approximately constant over subintervals of time, the term *Locally Mean Stationary* can be used only if the lengths of those subintervals are substantially longer than the coherence times of the function (the effective width of the function's autocovariance or the statistical-dependence time).
4) The various above local and global properties of ergodicity and stationarity, which are outside classical ergodicity theory, generalize to local and global properties of cycloergodicity and cyclostationarity.
5) These observations 1) – 4) reveal that, unlike the standard definitions of stationarity and cyclostationarity, and the implication of the Birkhoff ergodicity theorem and the corresponding cycloergodicity theorem, stationarity and cyclostationarity as well as ergodicity and cycloergodicity need not be "either/or" propositions. This may or may not moderate the level of the challenge of testing models for (cyclo)ergodicity. For example, relatively straightforward numerical methods may become feasible when there are no limits of time averages involved and/or when tests over complete sample spaces or function spaces are not needed. Despite this possible reduction of the magnitude of the challenge of testing for (cyclo)ergodicity, the use of

alternative FOT-probability models retains its attractiveness to me because they are direct, to the point, and more parsimonious.

CRediT authorship contribution statement

William A. Gardner: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Acknowledgment

The Author's thinking, as relayed in this article, has been strongly influenced over the preceding 20 of his 40 years of work on this topic by the work of his colleague, Professor Antonio Napolitano, and his colleague's coauthors, notably Professors Jacek Leskow and Dominique Dehay, beginning with the publication of their seminal 2006 paper [42] and including [43] and [44]. For this he expresses his deep gratitude.

Appendix I. Period Detection/Estimation

In the somewhat strange situation for which an analyst has a stochastic process model and does not know whether or not it exhibits some form of cyclostationarity or at least does not know the period(s), then before the theorems in this article can be used to test for cycloergodicity of the stochastic process model, a test for Period Detection is needed. This is a relatively far-fetched situation (enquired about by a reviewer of this paper) that does not bode well for eventual testing of the model for cycloergodicity. Nevertheless, the issue deserves consideration.

It is assumed that some nonlinear transformation on the process, such as a lag product, can be performed to transform higher-order CS into 1st-order CS (additive finite-strength sine-wave components), so the task here is to detect such components, if present, and estimate their period(s). To address this, we consider the companion problem of having a time series (e.g., a sample path of a stochastic process or real empirical data) and wanting to know if it contains additive periodic component(s). Thus, we need a period-test statistic. This appendix presents a simple straightforward example of such a statistic, a period-quality metric, denoted by *PQ*, without the formality of researching this interesting problem, which has a long history dating back to before the invention of the periodogram specifically for this period detection problem at the turn of the 19th century [5]. Thus, the statistic presented here may be novel (unlikely) or simply one of various viable alternatives that have been devised over time, such as the periodogram. This appendix may find more application by empiricists in their effort to choose time-average statistics to calculate from their data, than by analysts trying to test for cycloergodicity. The ad hoc metric introduced here complements the orthodox statistics from the now-classic theory of cyclostationarity, such as the periodogram, cyclic periodogram, spectral coherence function, and others [5].

The concept here (for the reviewer's question) is that the analyst would use the stochastic process model to produce a sample path for testing. The test described below can be repeated for as many sample paths as desired.

A function of time with minimum period T also has period NT for all integers N that do not render NT larger than half the total record length of the function. When averaging adjacent segments of a function, each of length equal to the hypothesized period NT, the data (function) is used more efficiently the smaller N is. If a period $\widehat{T}$ (with unknown N and true period T) is detected in the data, the candidate periods $\widehat{T}/n = NT/n$ for $n = 2, 3, 4, \ldots$ can be tested, and the *PQ* metric monitored for a maximum. As soon as $n = N$ is exceeded, *PQ* should decrease indicating that the function does not repeat as rapidly as every $\widehat{T}/n$ unit of time. At that point, a reasonable estimate would be that with the largest *PQ* value. When initially searching for a period by calculating the average of adjacent data segments of length, say, $\widehat{T}$, care should be taken to not start with an excessively large value for $\widehat{T}$ in order to not excessively limit the number of candidate periods in the data record. For $n = 2$, if adjacent $\widehat{T}/n$-length segments do not exhibit as much similarity as for $n = 1$, increase n by 1 to $n = 3$ and try again, and continue until the highest similarity is obtained, and select that value of n (which could be $n = 1$). The following *PQ* metric for a candidate period $\widehat{T} = t$ can be used for measuring similarity:

$$PQ(t) = 1 - \frac{\frac{1}{R}\sum_{k=1}^{R}\int_0^t |x(\nu + (k+1)t) - x(\nu + kt)|^2 \, d\nu}{\frac{2}{R}\sum_{k=1}^{R}\int_0^t \left\{ |x(\nu + (k+1)t)|^2 + |x(\nu + kt)|^2 \right\} d\nu} \tag{10}$$

PQ has been designed so that maximum similarity is $PQ = 1$, and minimum similarity is $PQ = 0$ and occurs when one data-segment equals the negative of its adjacent segment, in which case the true period is twice that candidate. The ratio, rounded down to the nearest integer, of the total record length to the candidate period *t is R.* There is a false maximum that occurs as t approaches 0 because, for very small trial periods, there is little variation in $x(\nu)$ from one candidate period to the next. Typically, as t increases from 0, *PQ* will start at 1 and tend to decrease and then begin to increase as the true period is approached and then, as the true period is exceeded, *PQ* will tend to decrease until t approaches 2 times the true period, and so on for each integer multiple of the true period. The strongest peak in *PQ* should occur at the true period. It is likely that multiple local peaks in *PQ* may occur but can be accommodated by doing a sufficiently exhaustive search.

A variation on *PQ* that should perform better replaces the numerator, which contains only the average of adjacent-segment mean-squared differences, which is $R\text{-}1$ in number, with the average of mean-squared differences for *all pairs of segments* obtained by partitioning the data record into adjacent t-length segments, which is approximately $R(R\text{-}1)/2$ in number. This would require some modification to the overall formula for *PQ* to preserve the metric range of [0,1], though this is not necessary.

The *PQ* method proposed here finds periods that maximize the correlation between time segments from the partitioned data whose segment lengths equal a candidate period, which is equivalent to minimizing the mean-squared differences between these segments. The high-performance variation of *PQ* described above uses a total averaging time in its correlation measurements of approximately $R^2t/2$ for the candidate period t. By comparison, the periodogram correlates the data with sine waves of all periods of interest using an averaging time of approximately Rt for each sine-wave frequency. If the periodogram values are summed over all harmonics of $1/t$, which is $2Bt$ in number (where B is the positive-frequency bandwidth of the periodic component) to produce the *harmogram* [5, p. 499], the total averaging time is $2RBt^2$. It follows that high-performance *PQ* has a total averaging time approximately equal to the record length Rt times *half the number of periods*, and the harmogram's total is approximately equal to the record length times the *number of harmonics* in the periodic component. So, the relative performance appears to depend on the number of periods and the number of harmonics. Many periods and few harmonics favor *PQ.* In addition, lack of knowledge of B will degrade the harmogram's performance.

Formula (10) shows PQ for first-order periodicity. For second-order, $x(t)$ must be replaced by $x(t)x^{(*)}(t-\tau)$, where (*) denotes optional conjugation and so on for higher order. Also, $x(t)$ can be replaced with an event-set indicator.

Appendix II. Catalog of (Cyclo)Ergodicity by Class

Classes of (cyclo)ergodic models

CT: stable linear time-invariant or periodically time-varying or almost periodically time-varying non-random transformation of white Gaussian noise or white Poisson impulse noise, with sufficiently transformation-memory decay rate.
DT: stable linear time-invariant or periodically time-varying or almost periodically time-varying non-random transformation of any i.i.d. sequence, with sufficient transformation-memory decay rate.
CT and DT: finitely repeated composition of any of the members of one or the other of the above two classes of transformations alternating with zero-memory-nonlinearities
CT and DT: WGN- or i.i.d.-driven time-invariant Volterra Systems with sufficient memory decay rate of the kernels
CT and DT: periodically time-variant, and almost periodically time-variant generalizations of the above Volterra class
CT or DT: uniform periodic pulse train with n^{th} absolutely integrable pulse modulated in the same manner from pulse to pulse by the n^{th} term in a vector-valued ergodic sequence; or discrete time sampled versions of such pulse trains.

Model properties that destroy (cyclo)ergodicity

- Additive and multiplicative and exponentiation random constants (or periodic functions in time) appearing in otherwise cycloergodic stochastic processes
- Random constants in time added to (only for cycloergodicity) or exponentiating time in an otherwise cycloergodic stochastic process
- Mixtures of processes: process 1 with probability P1 and process 2 with probability P2 (e.g., set 1 of CDFs with P1 and set 2 of CDFs with P2)

Data availability

No data was used for the research described in the article.

References

[1] C. Corduneanu, Almost Periodic Functions, Wiley, New York, 1961.
[2] R.M. Gray, Probability, Random Processes, and Ergodic Properties, Springer Verlag, 1987. Revised 2001, 2006, 2007, 2008.
[3] R.A. Boyles, W.A. Gardner, Cycloergodic properties of discrete-parameter nonstationary stochastic processes, IEEE Trans. Inf. Theory IT-29 (1) (1983) 105–114.
[4] J.R. Blum, D.L. Hanson, On quasi-periodic probability measures, J. Math. Mech 15 (1966) 291–303.
[5] W.A. Gardner, Statistical Spectral Analysis: A Non-Probabilistic Theory, Prentice-Hall, Englewood Cliffs, NJ, 1987, p. 565 (Author Comments: *SIAM Review*, Vol. 36, No. 3, pp. 528-530, 1994, https://epubs.siam.org/doi/epdf/10.1137/1036137.).
[6] W.A. Gardner (Ed.), Cyclostationarity in Communications and Signal Processing, IEEE Press, Piscataway, NJ, 1994 sponsored by *IEEE Communications Society*.
[7] www.cyclostationarity.com, March 2025.
[8] W.A. Gardner, The statistical and probabilistic theory of functions. (eBook due to be published in 2025).
[9] J. Nedoma, Uber die ergodizitat and r-ergodizitat stationarer wahrscheinlichkeitsmasse, Wahrscheinlichkeitstheorie 2 (1963) 90–97.
[10] W.A. Gardner, L.E. Franks, Characterization of cyclostationary random signals processes, IEEE Trans. Inf. Theory IT-21 (1) (1975) 4–14.
[11] W.R. Bennett, Statistics of regenerative digital transmission, Bell Syst. Tech. J. (Nov. 1958) 1501–1542.
[12] W.A. Gardner, Stationarizable random processes, IEEE Trans. Info. Th IT-24 (1) (1978) 8–22..
[13] https://en.wikipedia.org/wiki/Ergodicity.
[14] W.A. Gardner, Introduction to Random Processes With Applications to Signals and Systems, Macmillan, New York, 1985, p. 434.
[15] W.A. Gardner, Introduction to Random Processes With Applications to Signals and Systems, McGraw-Hill Publishing Company, New York, 1990, 546 pages.
[16] A. Napolitano, Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations, Academic Press, 2019.
[17] W.A. Gardner, A critique of the ergodicity-enigma-burdened stochastic process for time series analysis, to be submitted in July 2025 to *Signal Processing* (*EURASIP*).
[18] W.A. Gardner, The spectral correlation theory of cyclostationary time-series, Signal Process. (EURASIP) 11 (1986) 13–36.
[19] W.A. Gardner, Spectral correlation of modulated signals: part I — Analog modulation, IEEE Trans. Commun. COM-35 (6) (1987) 584–594.
[20] W.A. Gardner, W.A. Brown, C.-K. Chen, Spectral correlation of modulated signals: part ii — Digital modulation, IEEE Trans. Commun. COM-35 (9) (1987) 595–601.
[21] W.A. Gardner, Common pitfalls in the application of stationary process theory to time-sampled and modulated signals, IEEE Trans. Commun. COM-35 (5) (1987) 529–534.
[22] W.A. Gardner, Signal interception: a unifying theoretical framework for feature detection, IEEE Trans. Commun. COM-36 (8) (1988) 897–906.
[23] W.A. Gardner, Simplification of music and esprit by exploitation of cyclostationarity, Proc. IEEE 76 (7) (1988) 845–847.
[24] W.A. Gardner, Two alternative philosophies for estimation of the parameters of time-series, IEEE Trans. Inf. Theory 37 (1) (1991) 216–218.
[25] W.A. Gardner, Exploitation of spectral redundancy in cyclostationary signals, IEEE Signal Process Mag 8 (2) (1991) 14–36.
[26] W.A. Gardner, W.A. Brown, Fraction-of-time probability for time-series that exhibit cyclostationarity, Signal Process. (EURASIP) 23 (3) (1991) 273–292.
[27] W.A. Gardner, C.-K. Chen, Signal-selective time-difference-of-arrival estimation for passive location of man-made signal sources in highly corruptive environments. Part I: Theory and method, IEEE Trans. Signal Process. 40 (1992) 1168–1184.
[28] C.-K. Chen, W.A. Gardner, Signal-selective time-difference-of-arrival estimation for passive location of man-made signal sources in highly corruptive environments. Part II: Algorithms and performance, IEEE Trans. Signal Process. 40 (1992) 1185–1197.
[29] W.A. Gardner, Cyclic Wiener filtering: Theory and method, IEEE Trans. Commun. 41 (1993) 151–163.
[30] W.A. Gardner, C.M. Spooner, Detection and source location of weak cyclostationary signals: simplifications of the maximum-likelihood receiver, IEEE Trans. Commun. 41 (1993) 905–916.
[31] S.V. Schell, W.A. Gardner, High-resolution direction-finding, in: N.K. Bose, C. R. Rao (Eds.), Handbook of Statistics, Handbook of Statistics, 10, Elsevier, 1993, pp. 755–817.
[32] W.A. Gardner, T.L. Archer, Exploitation of cyclo-stationarity for identifying the volterra kernels of nonlinear systems, IEEE Trans. Inf. Theory 39 (1993) 535–542.
[33] W.A. Gardner, C.M. Spooner, The cumulant theory of cyclostationary time-series, Part I: Foundation, IEEE Trans. Signal Process. 42 (12) (1994) 3387–3408.
[34] C.M. Spooner, W.A. Gardner, The cumulant theory of cyclostationary time-series, Part II: Development and applications, IEEE Trans. Signal Process. 42 (12) (1994) 3409–3429.
[35] Chapter S.V. Schell, W.A. Gardner, Spatio-temporal filtering and equalization for cyclostationary signals, Chapter, in: C.T. Leondes (Ed.), Control and Dynamic Systems, Control and Dynamic Systems, 64, Academic Press, 1994, pp. 1–85.
[36] S.V. Schell, W.A. Gardner, et al., Blind adaptive antenna arrays for increased capacity in cellular communications, in: T.S. Rappaport, et al. (Eds.), Chapter 2 in Wireless Personal Communications: Trends and Challenges, Kluwer Academic Publishers, 1994.
[37] W.A. Gardner, C.M. Spooner, Cyclostationary signal processing, in: C.T. Leondes (Ed.), Chapter in Control and Dynamic Systems, Chapter in Control and Dynamic Systems, 65, Academic Press, 1994, pp. 1–92.
[38] W.A. Gardner, L. Paura, Identification of polyperiodic nonlinear systems, Signal Process. (EURASIP) 46 (1) (1995) 75–83.
[39] S.V. Schell, W.A. Gardner, Programmable canonical correlation analysis: a flexible framework for blind adaptive spatial filtering, IEEE Trans. Signal Process. 43 (12) (1995) 2898–2908.
[40] A. Napolitano, Cyclic higher-order statistics: input/output relations for discrete- and continuous-time MIMO linear almost-periodically time-variant systems, Signal Process. 42 (2) (March 1995) 147–166, https://doi.org/10.1016/0165 1684(94) 00124-I. *European Association for Signal Processing (EURASIP) Best Paper Award.* ISSN: 0165-1684.

[41] L. Izzo, A. Napolitano, Multirate processing of time-series exhibiting higher-order cyclostationarity, IEEE Trans. Signal Process. 46 (1998) 429–439, https://doi.org/10.1109/78.655427. ISSN: 1053-587X.
[42] J. Leskow, A. Napolitano, Foundations of the functional approach for signal analysis, Signal Process. (EURASIP) 86 (12) (December 2006) 3796–3825.
[43] D. Dehay, J. Leskow, A. Napolitano, Central limit theorem in the functional approach, IEEE Trans. Signal Process. 61 (16) (2013) 4025–4037, https://doi.org/10.1109/TSP.2013.2266324.
[44] D. Dehay, J. Leskow, A. Napolitano, Time average estimation in the fraction-of-time probability framework, Signal Process. (EURASIP) 153 (2018) 275–290, https://doi.org/10.1016/j.sigpro.2018.07.005.
[45] W.A. Gardner, A. Napolitano, L. Paura, Cyclostationarity: half a century of research, Signal Process. (EURASIP) 86 (2006) 639–697.
[46] W.A. Gardner, Statistically inferred time warping: extending the cyclostationarity paradigm from regular to irregular statistical cyclicity in scientific data, EURASIP J. Adv. Signal Process. 2018 (2018), https://doi.org/10.1186/s13634-018-0564-6. Article number: 59.
[47] A. Napolitano, W.A. Gardner, Fraction of time probability: advancing beyond the need for stationarity and ergodicity assumptions, IEEE Access 10 (2022) 34591–34612, https://doi.org/10.1109/ACCESS.2022.3162620.
[48] A. Napolitano, W.A. Gardner, Discovering and modeling hidden periodicities in science data, J. Adv. Signals Process. (2025), https://doi.org/10.1186/s13634-025-01207-w.
[49] W.A. Gardner and S.V. Schell. GMSK signal processors for improved communications capacity and quality. U.S. Patent No 5848,105, 1998.
[50] W.A. Gardner, Transitioning away from stochastic process models, J. Sound Vib. 565 (2023), https://doi.org/10.1016/j.jsv.2023.117871.
[51] D. Alexander, Pointwise ergodic theorem, drp_blogpost_spring2020+-+NotDylan+Alexander+PDF+%28web.ma.utexas.edu%29, accessed 6 July 2025.
[52] https://en.wikipedia.org/wiki/Kolmogorov_extension_theorem.
[53] T. Tao, An Introduction to Measure Theory, American Mathematical Soc, 2011. ISBN 978-0-8218-6919-2.
[54] A.N. Kolmogorov, *Foundations of the Theory of Probability*, 2nd ed., English. Dover Publications, Inc. Mineola New York.
[55] R.B. Ash, Real Analysis and Probability, Academic Press, New York, 1972.
[56] D.R. Brillinger, Time Series: Data Analysis and Theory, SIAM, Philadelphia, 2001.
[57] N. Merhav, On Jacob Ziv's individual-sequence approach to information theory, IEEE BITS Infor. Th. Mag (2025), https://doi.org/10.1109/MBITS.2025.3543321.
[58] W.L. Root, Nonprobabilistic and partly probabilistic channel coding and time-varying channels, in *Communication Systems and Random Process Theory,* J.K. Skwirzynski, ed., NATO Science Series E, Vol 25, Springer Dordrecht.