Transitioning Away from Stochastic Process Models

William A. Gardner

[a] *University of California, Davis, 18 Hahnemann Lane, Napa, CA 94558, USA bill@cyclicsignals.com,*

Abstract

It has been over 30 years since a paradigm shift from abstract stochastic process models to more concrete Fraction-of-Time Probability models for time-series data was called for and was supported by this journal's editor in chief. Yet, little, if any, detectable progress in making this transition has occurred. This paper reviews this needed transition and attempts to facilitate it with a new type of stochastic process model. The primary purpose of this model is to serve as a pedagogical tool for facilitating the conceptual transition from the standard relatively abstract way of thinking to a more concrete alternative. The utility of this parsimonious alternative was thoroughly proven when it was introduced in an advanced 1987 textbook, and the evidence in support has continued to accumulate in subsequent theoretical and applied research publications. But resistance to change is ever present.

Keywords: Vibration data modeling; sound data modeling; stochastic processes; statistical inference; time-series analysis; probabilistic modeling; ergodicity; cyclostationarity

1. Introduction

Because of the length of this introductory section, it has been partitioned into four subsections: Foreword, Level of Presentation, Origins, and Outline.

1.1. Foreword

The standard theoretical foundation for statistical processing of persistent signals, whether they are signals representing sound and vibration, or radio-frequency transmission, or time series of measurements on just about any persistent phenomenon, is presently the discrete-time and continuous-time Kolmogorov stochastic process models and especially, but not exclusively, strongly ergodic and cycloergodic Kolmogorov stochastic process models satisfying the axiom of relative measurability, which guarantees that limits of time averages on functions of sample paths exist. After a brief discussion exposing drawbacks of these generic models for many applications in statistical signal processing, particularly those involving empirical data, an alternative stochastic process model is proposed for statistically stationary signals, and a complementary model for statistically cyclostationary signals also is proposed. For these alternative models, defined first in terms of a parsimonious construction of their samples spaces, their cumulative probability distribution functions (CDFs) are derived from Fraction-of-Time (FOT) Probability calculations on a single member of the sample space, defined in terms of the Kac-Steinhaus relative measure on the Real line, and they are then shown to be valid CDFs over the entire sample space of the process. If all such finite-dimensional CDFs are specified, then this corresponds to a complete probabilistic model for the alternative stochastic process—equivalent to the specification of a probability measure defined directly on the sample space.

The motivating difference between Kolmogorov's model and this alternative parsimonious model is that the alternative is derived from empirical data, at least in principle. It is not posited in an abstract axiomatic manner that typically leads to a number of conceptually confusing and often unanswerable questions about the behavior of the sample paths in the model. These preferred alternative models are also complemented with another empirically derived model, this one for poly-cyclostationary signals that exhibit multiple incommensurate periods of cyclostationarity, but this model does not have an associated sample space for reasons explained herein.

The first applications proposed on page 358 in the first chapter (Chap. 10) of the six-chapter Part II of the 1987 book [1] (available at [2, p. 8.1]) that originated the comprehensive FOT-Probability theory of cyclostationarity, were "mechanical vibrations monitoring and diagnosis for machinery [from which] periodicity arises from rotation, revolution, and reciprocation of gears, belts, chains, shafts, propellers, bearings, pistons, and so on". By using the non-stochastic theory developed in Part II to study another field of applications—that of communications systems design and analysis—it was demonstrated that exploitation of cyclostationarity through signal processing was key to achieving substantial improvements in statistical inference. Since that time, there has been an explosion of applied work in both communications system design and analysis and monitoring and diagnosis of rotating machinery and many other fields of science and engineering (see bibliography in the encyclopedic 2019 book [3, pp. 360-362]). The seminal work on development of the foundation and theoretical framework for signals exhibiting cyclostationarity reported in [1] proved that a signal is cyclostationary if and only if there exist nonlinear transformations of the signal that generate finite-strength additive sine-wave components. This key non-stochastic characteristic led naturally to a comprehensive theory based on FOT Probabilities, without any mention of the more abstract axiomatically defined stochastic process. Yet, the applications of that theory have used, almost exclusively, the unnecessarily abstract—for many applications—stochastic process theoretical framework. This was made possible by my translation of the FOT-Probability theory into a dual Stochastic-Process theory in a 1989 companion book for the sake of completeness. But all the practical and pragmatic reasons given in [1] for practitioners to prefer the FOT-Probability theory do not appear to have resulted in the paradigm shift predicted at that time.

In this article, the conceptual and practical advantages of these three types of alternative stochastic-process models are discussed in some detail, *and then they are done away with*! That is, it is shown that the entire framework of stochastic processes, particularly the standard Kolmogorov processes with their often nonempirical abstraction, can be altogether circumvented by using FOT-Probability models for single signals, without any reference to stochastic processes. These single-signal models are *identical* to the novel alternative stochastic process models introduced here, but they do away with the unnecessary sample space because it is redundant. These most elegant of models provide all the same tools for statistical analysis—including CDFs, their derivatives—probability density functions (PDFs), temporal moments and cumulants, spectral moments and cumulants, and so on—but without any reference to stochastic processes and associated abstractions and confusing technicalities.

In the final analysis, it is recommended that the alternative stochastic process models introduced here be used exclusively as a pedagogical tool that helps in understanding the circumstances under which stochastic process models are unnecessary for statistical signal processing and probabilistic analysis involving stationary, cyclostationary, and poly-cyclostationary signals. These circumstances are, simply stated, any situation in which stochastic processes are appropriate if and only if they are ergodic or cycloergodic or multi-cycle generalizations thereof, possibly conditioned on knowledge of the values of random model parameters.

In contrast, the general situation for which stochastic processes are actually required, rather than avoidable, as a mathematical basis for statistical processing and analysis is that for which the lack of ergodicity or cycloergodicity is an essential characteristic. This is typically those situations for which populations or ensembles of signals are an essential ingredient. Nevertheless, when a stochastic process model is non-ergodic or non-cycloergodic but is *conditionally ergodic or cycloergodic*—meaning conditioned on knowledge of some finite set of parameters of the signal model, the conditional process is ergodic or cycloergodic—and when this conditioning can be either experimentally implemented or mathematically enforced in a data model, then the conditional FOT-CDFs can be measured or calculated and used in the same manner as CDFs for traditional stochastic processes. This enables the incorporation of FOT-Likelihood functions in the FOT-Probability theory.

In summary, the purpose of this paper is to help those, who have been indoctrinated in stochastic processes as the only viable analytical tool for statistical analysis of persistent signals, to make a transition in conceptualization that will enable them to replace this often unnecessarily abstract and conceptually problematic tool with a more elegant alternative that is formulated specifically for empirical data analysis. That this offered help is needed is evidenced by the passage of 35 years since a comprehensive introduction to this alternative was published in tutorial form and a paradigm shift was proposed; despite the passage of all this time, essentially all analysts who publish their work continue to cling tightly to the concept of the more abstract Kolmogorov model. It is the Author's belief that this is a result of shortcomings in education.

As a matter of fact, much applied work in engineering and other applied fields uses probability concepts and tools, such as expected values, autocorrelation functions, and other moments of signals, but does not actually formulate or even explicitly assume the existence of Kolmogorov stochastic process models for the signals of interest. It is often simply stated at the outset of a published research work "let $x(t)$ be a stochastic process, . . ." with the implication being that all the underlying mathematical machinery that may be required for the probability calculations subsequently performed to be meaningful exists . . . when in fact it may not exist in a manner that is consistent with the empirical data being studied or with various assumptions and restrictions on signals despite their being treated as if they arose from stochastic processes. In other words, one might say that the analytical portions of much practical work that treats signals as stochastic processes is a sham. The theoretical quantities, such as autocorrelation function, are not really explicitly defined, although symbolic formulas for them are derived through symbolic manipulation. And when it comes to implementation, these undefined quantities are calculated from empirical data typically using time averages in place of the undefined expected values.

Yet, work gets done, research papers get published, problems presumably get solved and so one might ask "who cares?" The answer is that, as an educator, I care because I know how confused students often are about the concept and effective use of the stochastic process, when times averages are the quantities of interest in practice, and because I see confusion in the minds of authors of applied research papers who have attempted to use stochastic processes in their work. This confusion is often simply buried by a complete disconnect between calculations or simply symbolic manipulations performed using expected values and experiments performed using time averages.

History reveals that the implementation of this pending paradigm shift has been found to be quite a challenge despite the strong support of the likes of Phillip. E. Doak, Founding Editor of this journal with a tenure as Editor in Chief of 40 years. On 8 March 1990, Phillip sent me his perspective on the need for this transition, and I quote:

"*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.*"

As author of the 1987 book [1] that proposes this paradigm shift, I cannot say it any better than this! Phillip's informed perspective is also aligned with that of other leaders in fields based on statistical signal processing, who—like Phillip—have made their informed positions clear. The first mentioned here is Professor Enders A. Robinson, originator of the digital revolution in geophysics, and highest honored scientist in the field of geophysics. In a published review of the book [1] [*Signal Processing*, EURASIP, and *Journal of Dynamical Systems, Measurement, and Control*, ASME, 1990], Enders 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."

Similarly, the following quotation from Professor Ronald N. Bracewell–recipient of the IEEE's Heinrich Hertz medal for pioneering work in antenna aperture synthesis and image reconstruction as applied to radio astronomy and to computer-assisted tomography–taken from his Foreword to the book [1], introducing FOT-Probability theory, makes essentially the same point that Enders 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."

Not to belabor the point, but even the information theorist, Professor James Massey—Professor of Digital Technology at ETH Zurich, IEEE Alexander Graham Bell medalist and member of the National Academy of Engineering—wrote, in a 1986 prepublication review of the book [1],

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

Summing up, despite the accolades given to the proposal for a paradigm shift, it has not yet happened. The intent of this paper is to further motivate the community with additional assistance for understanding the merit of the alternative to the stochastic process standard, and to introduce a new pedagogical tool for making the transition.

1.2. Level of Presentation

The statements of theoretical results and discussion of practical ramifications provided in this article are written for statistical signal processing engineers and like-minded time-series analysts, which may include physicists and other specialists in the physical sciences, and other fields where statistical analysis of empirical time-series is of interest. It is felt that mathematical proofs at any higher level of rigor than that which is presented herein would be distracting and are not included for this reason and others. Because the specific reasoning given in this article is not at odds with the day-to-day reasoning generally used by the intended audience, little of value would be added for this audience if a more mathematically rigorous presentation were provided. The preference acted on here is especially appropriate since the whole point of the effort leading to these new models is to show practitioners that the substantial abstractions and unmet challenges of trying to verify strong ergodicity or cycloergodicity of traditional stochastic process models are in the great majority of applications nothing more than distractions from the reality of empirical data and its processing and analysis and the more elegant theory that is identified here and is based on *Fraction-of-Time (FOT) Probability* for single signals.

Perhaps the most important reason for not getting distracted by rigor is that these new models are intended for only the pedagogical purpose of providing a conceptual transition from stochastic process models to FOT-Probability models of single signals and demonstrating that stochastic process models are often an unnecessary abstraction: they forfeit parsimony and mathematical elegance relative to the alternative single-signal models with fraction-of-time probability calculated directly from the single signal.

To counter the appearance of avoiding technical detail that may be important in comparing the two approaches to stochastic process modeling discussed in this paper, a glimpse into such details is provided in this paragraph and here and there in following sections. The Relative Measure used in [4] for the mathematical foundation of FOT-Probability models is not sigma additive (probabilities of infinite unions of nested event sets do not all converge), but in Kolmogorov's stochastic process probability model, sigma additivity of the proposed probability measure *is only assumed* by virtue of Axiom VI [5]. So, this axiom does not guarantee that, for any particular stochastic process model one adopts, the probability measure will in fact be sigma additive. Kolmogorov simply removes the mathematically undesirable general lack of sigma additivity of measures by axiomatically removing from consideration all probability measures that are not sigma additive. But how often do we encounter practitioners seeking to determine if the probability measure for some stochastic process model they have adopted is sigma additive or even just seeking to explicitly describe the probability measure for their adopted model? This is a very rare event. For the Fraction-of-Time Probability Theory discussed herein, an alternative restrictive assumption is required: the undesirable general lack of relative measurability of functions of time series is avoided by removing from consideration all time series and functions of those time series that are not relatively measurable. Such prohibited time series can be constructed, but they also can be considered to be contrived though an application of such contrived functions to secure communications has been proposed [4, Sec. 7]). This restriction to relative measurability is also required of the sample paths of ergodic stochastic processes, because sample-path time averages cannot converge to expected values if they do not converge at all. These restrictions are discussed further in [4].

In many applications, one starts with a finite-length record of empirical data. All that is actually required in many studies is that it be considered conceivable that there exists a mathematical model of an infinitely long data record that is consistent in some appropriate sense with the empirical data record. It is only in a minority of applied fields where more analytically oriented work is being done that an explicit specification of a mathematical model of an infinitely long signal is required.

1.3. Origins

The three-decade history from the 1930s through the 1950s of time-average statistical theory of time series is traced in [6] but the first approach to more comprehensive *Fraction-of-Time Probabilistic Modeling* of signals seems not to have been introduced until the concise publications of Brennan [7] and Hofstetter [8] in the 1960s. This approach was later developed independently and more comprehensively, including extension/generalization from stationarity to cyclostationarity, with in-depth application to the theory of statistical spectral analysis by myself in 1987 [1] (see also [9]). In the early 1980s, as I was writing the textbooks [1] and [10], I discovered the earlier work [7] and [8] as a result of discussions with Professor Thomas Kailath of Stanford University. I added to the Introduction in my book draft citations of this relatively unknown work from two decades earlier. As discussed in the present article and in more depth at the University of California, Davis website [2], earlier work on time-average theory, including [7] and [8], appears to have been largely forgotten as the *stochastic process bandwagon* trend developed.

The time-average approach was the starting point for the use of statistical time-series analysis in physics but has been largely ignored for well over half a century by many college instructors and criticized by some mathematicians for supposedly being non-rigorous. However, it has recently been shown by Leśkow and Napolitano to have a rigorous basis in measure theory, using mathematical tools dating back to the work of Kac and Steinhaus in 1938 [11]. This basis for measure-theoretic rigor underlying Fraction-of-Time Probability Theory was apparently lost track of in the shadow of Kolmogorov's contributions earlier the same decade. But, well over half a century later, it was uncovered by Leśkow and Napolitano in 2006 [4], where a more complete list of early (1920s to 1940s) contributors to time-average statistical theory is given (see also [3] by Napolitano).

1.4. Outline

1. Introduction
2. Historical Perspective
3. Results
    - 3.1 Kolmogorov's model of a stochastic Process
        - Birkhoff's Ergodic Theorem for Discrete Time
        - Birkhoff's Ergodic Theorem for Continuous Time
    - 3.2 The Measure Theory of FOT-Probability
    - 3.3 Definition of *Stationary FOT-Stochastic Process*
        - Stationary FOT Ergodic Theorem
        - Relation to Wold's Isomorphism
    - 3.4 Comparison of Kolmogorov and FOT stochastic Process Models
    - 3.5 Definition of *Cyclostationary FOT- Stochastic Process*
        - Cyclostationary FOT Cycloergodic Theorem
    - 3.6 The FOT-Probability Model for Almost Cyclostationary Processes
    - 3.7 Cycloergodicity for Multiple Incommensurate Periods
        - How to Generalize Birkhoff's Ergodic Theorem
    - 3.8 Purely Empirical FOT-Probability Models for Regular Cyclicity
    - 3.9 Purely Empirical FOT-Probability Models for Irregular Cyclicity
    - 3.10 The Weakness of Mean-Square Ergodicity
    - 3.11 Optimum and Adaptive Statistical Inference
4. Discussion of Results
5. References

2. Historical Perspective

To put this proposed evolutionary step in larger perspective, some stages of signal modeling that this community has passed through over the last century are briefly summarized. Time-series analysis goes back more than a century, but the time of R. A. Fisher one century ago seems to be a turning point when broader theoretical frameworks began to be formulated. This includes most notably Fisher's Principle of Maximum Likelihood, which is among the most commonly used optimization criteria for designing statistical inference and decision rules—algorithms—in use today within the statistical signal processing community. This includes both signal-parameter estimation and signal detection and classification. Predating Fisher by two centuries was Thomas Bayes, who gave birth to the theory of Minimum-Risk Statistical Inference and Decision (which addresses the same or similar signal parameter estimation and signal detection and classification problems that Maximum-Likelihood addresses, but with the added axiom that prior probabilities [prior to experimentation including observation or data collection] are assumed to exist). More recently, just preceding the middle of last century, Norbert Wiener used his developing statistical theory of single time functions (signals) to derive what we now call the Wiener Filter and related linear time-invariant signal processors, using a time-average counterpart of the Bayes Minimum-Risk design criterion, where risk was specified to be expected squared error, reformulated as time-averaged squared error. This was the continuous-time counterpart of Carl Friedrich Gauss's discrete least-squares optimization criterion used two centuries ago. Wiener's time-average theory and its applications to the nascent field of statistical communication theory was given a boost in visibility and further developed in 1960 with the publication of a book by one of Wiener's previous students at M.I.T, Yuk Wing Lee [12]. That same year, David Middleton's landmark book *An Introduction to Statistical Communication Theory* was published. In contrast to Lee's book, Middleton's was solidly based on the theory of stochastic processes. It had been said to cover a panoramic view unmatched by any other publication in the field [13]. This book was likely instrumental in cementing the place of the stochastic process in statistical signal processing. Middleton states in his preface "The mathematical exposition is for the most part heuristic". Although he does favor obtaining autocorrelation functions from signal models using time-averaging, he then takes an expected value to obtain an ensemble autocorrelation. Because of this approach, he misses the fact that some of his signal models are cyclostationary, not stationary. Nevertheless, he does note that, in general, his approach produces stationary autocorrelations for nonstationary processes. This precedes more theoretical work decades later on what are called *asymptotically mean-stationary processes*, which includes as special cases cyclostationary and almost cyclostationary processes. Middleton, however, does not adopt the Kolmogorov model for stochastic processes. He uses heuristics instead.

Contemporaneously with Wiener in the 1930s and 1940s, Kolmogorov introduced the now-standard theory of the stochastic process as a probabilistic model for time-series. Also contemporaneous was the establishment of *Information Theory* by its originators, Harry Nyquist, Ralph Hartley, and Claude Shannon during the 1920s – 1940s. The landmark event *establishing* the discipline of information theory and bringing it to immediate worldwide attention was the publication of Claude E. Shannon's classic paper "A Mathematical Theory of Communication" in the Bell System Technical Journal in 1948. This theory is strongly probabilistic. From 1960 forward, Wiener's time-average approach quickly faded into the background, and Kolmogorov's expected-value approach grew into the standard we use today. It is conceivable that this was in large part a result of the boom that information theory initiated and possibly also a result of the mathematical rigor of Kolmogorov's book on the theory of stochastic processes. Interestingly, though, information theory involving signals is valid for time-average probabilities, not just ensemble-average probabilities, as discussed further on in this paper.

What has for almost a century been referred to as *statistical time-series analysis* has increasingly come to be relabeled *statistical signal processing*, perhaps because of the lead electrical engineers have taken in developing the technology used for implementation. This field of study, born within the field of electrical engineering, was originally based in large part on what is called *statistical communication theory*, which arose out of the work of Wiener and his contemporaries but was reformulated in terms of expected values and stochastic process models. This theory is more probabilistic than it is statistical, yet it is called a statistical theory by the authors of classic books on the subject, written starting in the 1950s-1960s, particularly Middleton's book. Middleton is, however, precise in his distinction between statistical and probabilistic quantities. But, over time, the language has become less precise. Today, the terms *signal* and *time series* are often used interchangeably by more broadly educated practitioners, with some preference given to *time series* by statisticians and preference given to *signals* by engineers, especially electrical/electronics engineers. The primary difference between time-series analysis and signal processing is that, prior to the communications technology revolution, the term *signal* was not yet being used for essentially any time-record of data. Some authors reserve the term *time series* for discrete-time data.

In communication theory, the stochastic process model of signals was adopted because a key concept was to design inference-making algorithms that optimized expected performance (minimized *expected cost*, which is the definition of Bayesian Risk). That is, performance was to be optimized over the population or ensemble of all sample paths of a stochastic process model of a type of signal of interest. For example, in telecommunications, the Wiener filter—according to modern theory—was the solution to minimum-mean-squared-error estimation of a transmitted signal, given a corrupted version of that signal obtained from a remote receiver. Thus, the statistical averaging of interest, performed by the expectation operation, was performed for example over all speech to be telecommunicated (referring back to the early days of Bell Telephone Laboratories), as well as all noise corrupting the transmitted signal. This eventually included all speaker physiologies, all languages, and all accents. Standardized fixed population-statistics computed empirically and expected values were used for designing channel filters and equalizers, which themselves were fixed or manually adjustable. But, as technology progressed, fixed optimum solutions began to be replaced with adaptive solutions that automatically optimized performance for each and every single signal. This required working with statistics obtained from time averaging single signals, not ensemble averaging multiple signals. This gave impetus to preferring ergodic stochastic process models for signals because then solutions implemented with algorithms that computed and used time-average statistics gave good approximations to the ensemble-averages dealt with in the mathematical models used for deriving the algorithms, and this rendered the stochastic process theory, in which electrical engineers were beginning to be indoctrinated, adequate for these. But despite ergodic theory, most users did not know how to test their mathematical signal models for strong ergodicity. Birkhoff's ergodic theorem provided the ergodicity condition only in terms of the abstract mathematical probability measure defined (possibly only generically specified) in terms of a function of arbitrary subsets in a sigma field—the mathematical sample space—which also was defined (often only generically specified) in terms of sample paths often having no explicit description, e.g., interfering signals known only by their power spectral densities. So, the ergodicity condition was rarely able to be tested. Empirical data was of no use for this purpose because the condition involves only the abstract probability measure; it's a property of the mathematical model, not the empirical data. Practitioners often just invoked the Ergodic Hypothesis and typically left it untested. This is discussed early on by Middleton and remained the status quo up to and including today. *But, once ergodicity was invoked, the stochastic process model was, in principle, no longer the most appropriate model, as explained in this paper and its references.* With time-averages of primary concern, population averages became, in principle but often unknowingly, irrelevant, and the abstraction of stochastic processes became unnecessary and nothing more than a distraction—something not recognized by most users. Although Middleton uses time averages, especially for calculating autocorrelation functions and associated quantities, before he takes the expected value, he does not appear to comment on the broader concept of FOT-Probability.

Although 35 years have passed since a comprehensive development of an alternative probability theory for random signals that is based entirely on time averages was published in textbook form [1], this alternative theory has been largely ignored by all but a small minority of users of stochastic processes. For instructors of courses on statistical signal processing, teaching this alternative requires an introductory textbook, since the only textbook available [1] is written for advanced students. Similarly, a 2^{nd} book (not a textbook with exercises) treating this alternative theory that appeared just two years ago is written for experts or at least mathematically mature readers.

This stagnation in statistical signal processing pedagogy in universities occurred even though this simpler more transparent theory was proven in [1] to be analogous and actually *operationally equivalent* to the probability theory based on abstract and, one might even say, mysterious ergodic stochastic process models and, with regard to calculations, *yields the same results* in all cases for which relative measurability is assumed, which is necessary for the ergodic theorem to prove that expected values can be approximated by time averages. It is hoped that the pedagogical approach taken in this paper, whereby alternative stochastic process models are introduced as a conceptual transition from Kolmogorov's abstract stochastic process to concrete FOT-Probability models for single signals will spark interest in universities in developing new introductory courses based on the time-average theory of signals. Some of the many practical advantages of doing so are discussed in this article.

To be especially clear at the outset about limitations of FOT-Probability Theory, the particularly important area of statistical inference and decision-making based on time-series observations is briefly discussed. Generally speaking, FOT-Probability models are well matched to what might be loosely called *non-parametric inference and decision*, for which no use is made of assumed functional forms of Cumulative Distribution Functions (CDFs) of the data with or without known, unknown, or random parameters of the functional form; the only CDF used is that measured from the observed time-series data. The complementary area of statistical inference and decision-making denoted with the adjective *parametric* partitions into two general types, one of which is accommodated by FOT-Probability models and the other of which is not.

The type of parametric statistical inference and decision making that is not accommodated by FOT-Probability theory is that which is based on non-ergodic stochastic process models and some ergodic models for which probability functions, including CDFs or possibly just some moments, for the data conditioned on knowledge of some model-parameter values and/or hypotheses are needed but cannot be measured or calculated from a model for the observed data. Such cases can arise in Maximum-Likelihood Methods and Bayesian Minimum-Risk Methods of inference and decision making. If such parameters are modeled as random variables, the data must be considered to have arisen from a non-ergodic process since observation of one record of data cannot be used to learn about the influence of other values of the parameters that did not occur in the record of data. For example, if received data consists of signal plus noise under one hypothesis and noise only under an alternative hypothesis, the stochastic process model for the data that is not conditioned on a specific hypothesis cannot be ergodic.

In contrast to these parametric methods based on non-ergodic models, there is a type of parametric inference and decision making that is based on formulaic data models (sample-path models) in which the values of some parameters are unknown but are not treated as random variables. These are stochastic process models that are known only partially. For such models, one can in principle use the expectation operation to mathematically calculate the dependence of theoretical probability functions, such as moments, on the unknown parameters and then equate these theoretical moments to measured sample moments, and finally solve these equations, when possible, for the unknown parameters. This is called the *Method of Moments* for inferring parameter values.

Popular sample-path models used in the Method of Moments are autoregressive (AR), moving average (MA), and ARMA models and their periodic and poly-periodic generalizations. All such parametric methods are accommodated by the theory of FOT-moments associated with FOT-probabilities, for which the expected values in the Method of Moments are replaced with limits of time averages, and the empirical counterparts that were equated with expected values are finite-time averages that are equated with the limits of time averages. A survey of FOT parametric statistical spectral analysis is available in [1]; see also [3], [14], [15]. In addition, a radically different method of moments that has not yet been thoroughly evaluated is described in [2. p. 11.4].

3. Results

3.1. Kolmogorov's model of a stochastic Process

We are interested here in discussing alternatives to both the discrete- and continuous-time versions of Kolmogorov's 1933 definition [5] of a stochastic process consisting of a sample space (the set of all sample paths, or signal realizations), a sigma field of subsets (events) in the sample space with a sigma algebra, and a probability measure on the event sets. These "sigma" requirements, meaning "convergence requirements for countably infinitely many operations", derive from Kolmogorov's Axiom VI in his definition of a stochastic process. In practice, the specification of a particular probability measure is rarely carried out because this is a difficult mathematical challenge for which there is no recipe. Sometimes practitioners will specify some lower order CDFs or Probability Density Functions (PDFs) as a half-hearted attempt. In the special case of a Gaussian process, the specification of the 2^{nd}-order CDF or PDF is all that is needed to derive from it all orders of CDFs and PDFs. Once all orders are specified, one can invoke the Kolmogorov Extension Theorem to conclude that the measure for the sample space has been effectively, if not explicitly, specified.

Although there exist a modest number of well-known specifications of probability measures for stochastic processes, it is fair to say that in much practical work the probability measure for a stochastic process is rarely specified; as a consequence, Axiom VI can only rarely be tested. Consequently, it is common practice to simply assume Axiom VI is satisfied by the selected model and proceed to use the consequences of that axiom in performing calculations involving infinite sums—not a particularly justifiable approach.

In other cases, practitioners will construct a formulaic model of a stochastic process as some combination of specified deterministic functions and some random variables. For example, essentially all digital communications signal models are specified in this manner. Similarly, vibrations from, say, bearing faults in rotating machinery are sometimes modeled as the response of a specified linear time-invariant dynamical system to a nearly periodic train of impulses, with one or two associated random parameters. For time-varying RPM, the impulse rate varies in proportion to the RPM. This typically provides no insight into the probability measure for the process but does often enable the practitioner to calculate some moments and/or cumulants and, much less frequently, some CDFs or PDFs. In a number of cases for which statistical inference using the stochastic process model is of interest, it suffices to calculate only the PDF for the observed data, conditioned on knowledge of the random variables in the model that are to be estimated, or conditioned on hypotheses to be tested. This can be adequate for deriving maximum-likelihood inference rules and in some cases minimum-Bayes-Risk inference rules.

In summary, it is a relatively rare occasion when Kolmogorov's model of a stochastic process is able to be specified and used for time-series analysis, aka statistical signal processing. A particularly egregious consequence of this common practice is having to assume that an adopted and possibly only partially specified model is strongly ergodic. This assumption, when valid, enables one to accurately approximate expected values, calculated from the model, using time averages on sufficiently long finite segments of a single realization of the signal being modeled. Without actually knowing that the model used for calculating expected values is ergodic, such time averages may or may not be accurate approximations. In fact, without the added assumption, which is typically ignored, that limits of time averages of sample paths exist, the ergodic hypothesis—whether true or false—does not guarantee that expected values can be approximated by time averages.

The above less-than-desirable situation concerning the use of Kolmogorov's stochastic process model has been tolerated for nearly a century now. Evidently, we've "gotten by" despite the unsavory facts summarized above. Nevertheless, there do exist alternative approaches to modeling signals for purposes of statistical inference and analysis. The purpose of this paper is to present such a model—the FOT-Probability model of a single signal—and explain how it relates to Kolmogorov's model and how much easier it is to use in practice in a more justifiable manner for applications in statistical signal processing, where complete mathematical specifications of stochastic processes a la Kolmogorov is not possible. It should however be mentioned here that the FOT-Probability model can be used for statistical inference and decision-making involving likelihood functions only when such likelihood functions can be measured or calculated as conditional FOT-PDFs. This is further discussed in Section 4.

An *event set* A for some specified event, such as the event that a stochastic process takes on a value exceeding unity at time 1 sec, is the set of all sample paths for which this event occurs. For the purpose at hand, let $T_t(A)$ denote the time-translation set-operator that shifts, by any real number $t \in R$, typically representing time, all sample paths in an event set A, and let $T_n(B)$ denote the discrete-time counterpart for any integer $n \in Z$. Following are the two ergodic theorems that are assumed to apply in many applications:

Birkhoff's Ergodic Theorem for Discrete Time (BET–DT)

Consider a discrete-time Kolmogorov stochastic process with integer-valued time, satisfying Kolmogorov's six defining axioms [5], for which all event sets E that are translation-invariant, $T_n\{E\} = E$ for all integers n, have probabilities of either $P(E) = 0$ or $P(E) = 1$. By Birkhoff's 1931 Ergodic Theorem [16], this stochastic process is *ergodic w.p.1*, and is also referred to as *strongly ergodic.* Birkhoff's *ergodicity condition* here is necessary for discrete-time-averages of functions of the stochastic process to converge, with probability equal to one (*w.p.1*), to the corresponding expected values, as the averaging time approaches infinity.

Birkhoff's Ergodic Theorem for Continuous Time (BET–CT)

Consider a continuous-time Kolmogorov stochastic process, satisfying Kolmogorov's six defining axioms [5], for which all event sets E that are translation-invariant, $T_t\{E\} = E$ for all real t, have probabilities of either $P(E) = 0$ or $P(E) = 1$. By Birkhoff's 1931 Ergodic Theorem [16], extended from discrete- to continuous-time (e.g., page 1 of [17]), this stochastic process is *ergodic w.p.1*, and is also referred to as *strongly ergodic.* Birkhoff's *ergodicity condition* here is necessary for continuous-time-averages of functions of the stochastic process to converge, w.p.1, to the corresponding expected values as the averaging time approaches infinity.

These theorems require an additional axiom, here labeled *Axiom VII*, or they require a proof of a *proposition* in order to provide the desired necessary and sufficient condition for strong ergodicity. Without this Axiom VII or a proof of the proposition, these theorems are not applicable in the way they have been applied for many years. This needed axiom or proof guarantees that the limits of the time averages of interest in practice exist. If they do exist, then the ergodic theorem establishes that they equal w.p.1 the corresponding expected values. For discrete time, this proposition has been proved at least in some cases such as for finite-alphabet processes. As per my knowledge, it may or may not have been proved for continuous time. The proposition can be stated as follows: For an ergodic Kolmogorov discrete-time (continuous-time) process, the samples paths of well-behaved functions of the process are *relatively measurable*, as defined below.

One example of a sufficient condition for existence of the continuous-time average, which has been assumed in the early work on ergodic theorems, like Birkhoff's work (cf. [16] and references therein) is that the function of time is any well-behaved function of the positions of the particles of a dynamical system described by differential equations for which the sum of kinetic energies of all the particles in the system is time invariant. Unfortunately, this is typically not an appropriate model for the manmade signals used in communication systems and also not appropriate for many other applications like rotating machinery fault diagnosis and monitoring, and biological signals.

3.2. The Measure Theory of FOT-Probability

The material in this subsection is taken from [4], also cf. [3, Chap. 2]. Let us consider the set $A \in \mathcal{B}_{\mathbb{R}}$, where $\mathcal{B}_{\mathbb{R}}$ is the σ-field of the Borel subsets on the real line and let μ be the Lebesgue measure on the real line $\mathbb{R}$. The *relative measure* of A is defined by Kac and Steinhaus [11] as follows

$$\mu_R(A) \triangleq \lim_{T\to\infty} \frac{1}{T} \mu \left(A \cap [t_0 - T/2, t_0 + T/2] \right) \tag{1}$$

provided that the limit exists. In such a case, the limit does not depend on t_0 and the set A is said to be *relatively measurable* (RM). For example, given a function $x(t)$, the event set A consisting of all the time points on the real line for which some event involving $x(t)$ occurs has *Fraction-Of-Time (FOT) Probability* given by $\mu_R(A)$, provided that A is relatively measurable.

Let $x(t)$ be a Lebesgue measurable function on the real line. The function $x(t)$ is said to be relatively measurable [11] if the set $\{t \in \mathbb{R} : x(t) \leq \xi\}$ is

RM for every $\xi \in \mathbb{R} - N_0$, where N_0 is at most a countable set of points. Each RM function $x(t)$ generates the function

$$
\begin{aligned}
F_x(\xi) &\triangleq \mu_R(\{t \in \mathbb{R} : x(t) \leq \xi\}) \\
&= \lim_{T\to\infty} \frac{1}{T} \mu\left(\{t \in [t_0 - T/2, t_0 + T/2] : x(t) \leq \xi\}\right) \\
&= \lim_{T\to\infty} \frac{1}{T} \int_{t_0-T/2}^{t_0+T/2} \mathrm{u}(\xi - x(t))\, \mathrm{d}t
\end{aligned}
\tag{2}
$$

at all points $\xi$where the limit exists. In this equation, $\mathrm{u}(\xi)$ denotes the unit step function:$\mathrm{u}(\xi) = 1$ for $\xi \geq 0$ and $\mathrm{u}(\xi) = 0$ for $\xi < 0$.

The function $F_x(\xi)$ has all the properties of a valid cumulative distribution function (CDF), except for the right-continuity property (at points of discontinuity). It represents the *fraction-of-time* (FOT) that the function $x(t)$ is below the threshold ξ, as illustrated in Fig. 1. For this reason, $F_x(\xi)$ is referred to as *the FOT- distribution* of the function $x(t)$.

Since the relative measure of every finite set is zero, the relative measure of every finite-energy or transient function $x(t)$ has the trivial distribution function $F_x(\xi) = \mathrm{u}(\xi)$. Only finite-average-power or persistent functions, such as almost periodic functions, can have a non-trivial FOT-distribution.

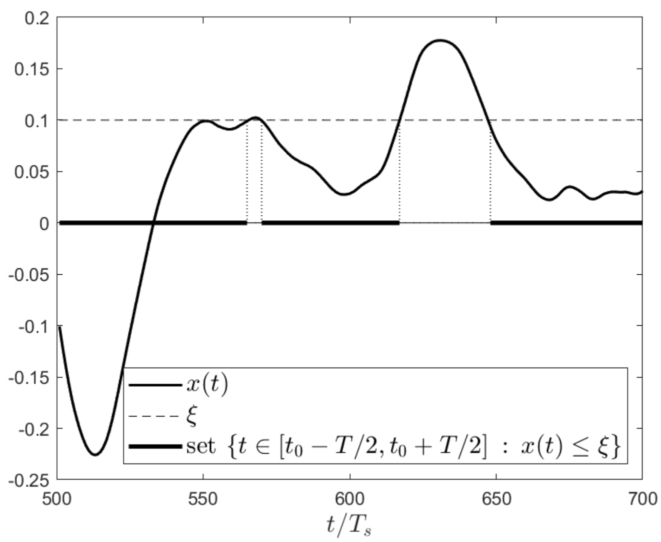

Fig. 1 The measure of the set $\{t \in [t_0 - T/2, t_0 + T/2] : x(t) \leq \xi\}$ *(the length of the thick line) divided by the total time* T *is the fraction of time that the function* $x(t)$ *is below the threshold* ξ *as* t *ranges over* $[t_0 - T/2, t_0 + T/2]$.

If $x(t)$ is a relatively measurable persistent function and not necessarily bounded and $g(\cdot)$ is a well-behaved function, then the following Fundamental Theorem of Time Averages [1] can be verified [4, Theorem 3.2]

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} g(x(t))\,\mathrm{d}t = \int_R g(\xi)\,\mathrm{d}F_x(\xi) \tag{3}$$

where the first integral in the left member is in the Lebesgue sense and does not depend on t_0, and the integral in the right member is in the Riemann-Stieltjes sense. When $F_x(\xi)$ is differentiable, its derivative, denoted by $f_x(\xi)$, is the probability density function, and $dF_x(\xi)$ can be replaced in the right member with $f_x(\xi)\mathrm{d}x$, in which case the integral is in the Riemann sense.

From this theorem, it follows that *the infinite-time average is the expectation operator for the FOT-distribution* $F_x(\xi)$ and for every bounded $x(t)$ we have

$$\langle x(t)\rangle_t \equiv \lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x(t)\,\mathrm{d}t = \int_{\mathbb{R}} \xi\,\mathrm{d}F_x(\xi) \tag{4}$$

The analogy between FOT-Probability and Kolmogorov probability [1], [9] is evident.

For a 1st-order strict-sense stationary process $X(t)$ with distribution $F_X(\xi) \triangleq P[X(t) \leq \xi]$, the stochastic counterpart of the above time-average definition of the distribution is

$$F_X(\xi) = \mathrm{E}\{\mathrm{u}(\xi - X(t))\} \tag{5}$$

where $\mathrm{E}\{\cdot\}$ is the expected value operation, which equals the limit ensemble average operation, and which replaces the time average operation used in the FOT-Probability approach. Similarly, the Kolmogorov counterpart of the *Fundamental Theorem of Time Average* is the following *Fundamental Theorem of Expectation*

$$\mathrm{E}\{g(X(t))\} = \int_{\mathbb{R}} g(\xi)\,\mathrm{d}F_X(\xi). \tag{6}$$

A necessary and sufficient condition for the relative measurability of a function is not known. However, if $x(t)$ is a bounded function, the existence of the time average

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x^p(t)\,\mathrm{d}t. \tag{7}$$

for every positive integer p is a necessary condition for the relative measurability of $x(t)$. In addition, it follows from the Fundamental Theorem of Time Average that, if $x(t)$ is continuous and bounded and the left-hand side of the equation

$$\lim_{T\to\infty}\frac{1}{T}\int_{t_0-T/2}^{t_0+T/2} x^p(t)\,\mathrm{d}t = \int_{\mathbb{R}} \xi^p\ \mathrm{d}F_x(\xi) \tag{8}$$

exists for every positive integer$p$, then $x(t)$ is relatively measurable, and the above equation is valid.

As a final remark, it is noted that the absence of right-continuity of the FOT-distribution is not important in applications where integrals in $\mathrm{d}F_x(\xi)$ are of interest. For stochastic probability, the right-continuity of the distribution is a consequence of the assumed σ-additivity of the probability measure P.

The preceding theory has a completely analogous discrete-time counterpart, which can be obtained by simply replacing integrals over continuous time with sums over discrete time [3, Chap. 2]. The same terminology is used. For example, the relative measure of a finite set A is defined by

$$\mu_R(A) \triangleq \lim_{N\to\infty}\frac{1}{2N+1}\#\left(A\cap[n_0-N, n_0+N]\right) \tag{9}$$

where $\#(A)$ is the counting measure of the finite set A, which equals the number of elements in A.

We can now proceed with the definition of the stationary FOT-stochastic process. As above, $x(t)$ represents a persistent relatively measurable real-valued function of time defined over the entire real line and x_n represents a persistent relatively measurable real-valued sequence indexed by discrete time over the entire set of integers.

Multiple functions are said to be *Jointly Relatively Measurable* if they each are relatively measurable, meaning there FOT-CDFs exist, and their joint FOT-CDFs exist.

3.3. Definition of Stationary FOT- Stochastic Process

Def. S1: The *Sample Space* of the *Stationary FOT-Stochastic Process* is comprised of all the time translates of a single relatively measurable discrete- or continuous-time sample path (persistent real-valued function of a real variable), x, subject to the constraint that replications are disallowed (no two sample paths can be identical):

$$\begin{aligned} \Omega_d &= \{\{x_{n-\omega}; n \in \mathbb{Z}\}; \omega \in \mathbb{Z}\}, \\ \Omega_c &= \{\{x(t-\omega); t \in \mathbb{R}\}; \omega \in \mathbb{R}\} \end{aligned} \tag{10}$$

Def. S2: The probability of any relatively measurable subset of elements from the sample space index set R or Z, called an *event*, is the value of the relative measure of that set.

Def. S3: The FOT-CDF of any relatively measurable discrete- or continuous-time function, $f[x](t)$ or $f[x]_n$, which is jointly relatively measurable, for m real-valued time points $\{t_1, t_2, t_3, ..., t_m\}$ or m integer-valued time points $\{n_1, n_2, n_3, ..., n_m\}$, respectively, of the *Stationary FOT- Stochastic Process* $x(t)$ or x_n is the relative measure of the event set

$$\begin{aligned} E_m^c \triangleq \{&\omega \in \mathbb{R}; f[x](t_1-\omega) \leq \xi_1, \\ &f[x](t_2-\omega) \leq \xi_2, ..., f[x](t_m-\omega) \leq \xi_m\} \end{aligned} \tag{11a}$$

or

$$\begin{aligned} E_m^d \triangleq \{&\omega \in \mathbb{R}; f[x]_{n_1-\omega} \leq \xi_1, \\ &f[x]_{n_2-\omega} \leq \xi_2, ..., f[x]_{n_m-\omega} \leq \xi_m\} \end{aligned} \tag{11b}$$

for all real-valued m-tuples $\{\xi_1, \xi_2, \xi_3, ..., \xi_m\}$.

It follows from Def. S3 that the 1^{st} order FOT-CDF for a continuous-time stationary FOT process is given explicitly by the formula

$$\begin{aligned} &F_x(\xi) \triangleq \mu_R(\{t \in \mathbb{R} : x(t) \leq \xi\}) \\ &= \lim_{U\to\infty} \frac{1}{U} \mu\left(\{t \in [t_0 - U/2, t_0 + U/2] : x(t) \leq \xi\}\right) \\ &= \lim_{U\to\infty} \frac{1}{U} \int_{t_0-U/2}^{t_0+U/2} \mathrm{u}(\xi - x(t))\mathrm{d}t \end{aligned} \tag{12}$$

for all real ξ, and similarly for higher-order FOT-CDFs; and, for discrete-time, the FOT-CDF is given by

$$\begin{aligned} &F_x(\xi) \triangleq \mu_R\left(\{n \in \mathbb{Z} : x_n \leq \xi\}\right) \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \#\left(\{n \in [n_0 - N, n_0 + N] : x_n \leq \xi\}\right) \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{k=n_0-N}^{n_0+N} \mathrm{u}\left(\xi - x_n\right) \end{aligned} \tag{13}$$

As another example, for $m = 2$,we have the 2^{nd} order FOT-CDF

$$\begin{aligned}
F_x(\xi_1, \xi_2) \;&\triangleq\; \mu_R(\{t \in \mathbb{R} : x(t + t_1) \leq \xi_1, \\
&\qquad\qquad x(t + t_2) \leq \xi_2\}\} \\
&= \lim_{U\to\infty} \frac{1}{U} \mu(\{t \in [t_0 - U/2, t_0 + U/2] : x(t + t_1) \leq \xi_1,\; x(t + t_2) \leq \xi_2\}) \\
&= \lim_{U\to\infty} \frac{1}{U} \int_{t_o - U/2}^{t_o + U/2} \mathrm{u}\left(\xi_1 - x(t + t_1)\right) \mathrm{u}\left(\xi_2 - x(t + t_2)\right) \mathrm{d}t
\end{aligned} \tag{14}$$

for all real ξ. Note: The constraint in Def. S1 that disallows replications in the sample space also disallows constant signals, which are a degenerate case of stationary signals. A viable alternative is to remove this constraint.

The probability of the entire sample space of the *Stationary FOT-Stochastic Process* is equal to 1, meaning every *experimental* outcome for this model is one of the members of the sample space. That is, for a discrete sample space Ω_d^N with a finite number N of translates, the probability of each translate is $1/N$ and since these translates are mutually exclusive events, the probability of the entire set of N translates is the sum over N probabilities, each equal to $1/N$, which sum equals 1. In the limit, as the number of translates N included in the sample space approaches infinity, we get the result that the probability of each sample path is 0 and the probability of the total sample space Ω_d is 1. Similarly, for a continuous sample space, the probability of each sample path is 0, because the relative measure of a single point on the real line is 0, and the probability of the total sample space Ω_c is 1, because the relative measure of the entire real line is 1.

For this FOT-stochastic process, any one of the translates, $\{x(t - \omega) : t \in \mathbb{R}\}$ for any particular $\omega \in \mathbb{R}$ or $\{x_{n-\omega} : n \in \mathbb{Z}\}$ for any particular $\omega \in \mathbb{Z}$, can be taken as the *Sample Space Generator*. In practice, the sample space generator would be taken to be the single observed signal, conceptually extended from the finite observation interval to the real line, or to the integers; and when a formulaic specification of the process is made, the sample space generator would be obtained from the formula for any specified set of random samples of the random functions in the formula. So, given a specification of one sample path, we have a specification of the entire sample space. Here are some examples that are commonly encountered in communications systems and various other applications.

Example 1: Binary Amplitude-Modulated Pulse-Train Signal

$$x_1(t) = \sum_{k=-\infty}^{+\infty} a_k p_1(t - kT_1)$$

where $\{a_k\}$ is a sequence of i.i.d. (in the FOT-Probability sense) binary-valued (± 1) numbers and $p_1(t)$ is an absolutely integrable pulse of essentially arbitrary shape, and T_1 is a real number.

Example 2: Amplitude-Modulated Sine-Wave Carrier Signal

$$x_2(t) = a_2(t)\cos(2\pi f_2 t + \theta_2)$$

where $a_2(t)$ is an FOT-stationary Gaussian signal with some specified continuous FOT power spectral density function, and f_2 and θ_2 are real numbers.

Example 3: Amplitude-Shift Keyed Sine-Wave Carrier Signal

$$x_3(t) = \sum_{k=-\infty}^{+\infty} a_k p_3(t - kT_3)\cos(2\pi f_3 t + \theta_3)$$

where $\{a_k\}$ is a sequence of i.i.d. (in the FOT-Probability sense) binary-valued (± 1) numbers and $p_3(t)$ is an absolutely integrable pulse of essentially arbitrary shape, and f_3 and θ_3 are real numbers.

Example 4: Phase-Modulated Sine-Wave Carrier Signal

$$x_4(t) = a_4 \cos(2\pi f_4 t + \theta_4(t))$$

where a_4 is a real number, $\theta_4(t)$ is an FOT-stationary Gaussian signal, with some specified FOT power spectral density function that has been passed through a zero-memory nonlinear device that is linear with slope of 1 over the domain $[-\pi,\ \pi]$ and has output of $-\pi$over the domain $[-\infty, -\pi]$ and $+\pi$ over the domain $[\pi, +\infty]$.

Example 5: Multiplexed Signal with two independent (in the FOT Probability sense) components

$$x_5(t) = x_2(t) + x_4(t)$$

There are numerous examples of calculations of FOT probabilistic parameters for formulaic specifications like those in the above examples; the first extensive catalog appeared in the book [1] and this was recently supplemented with additional examples in the book [3]. The great majority of these are calculations of cyclic autocorrelations and cyclic spectra (spectral correlation functions), but there are also some examples of calculations of higher-order moments and cumulants, both temporal and spectral types, cf. [19]. Calculations of cumulative FOT-Probability distribution functions are less common. The reason is undoubtedly a result of the effort required. It is more practically feasible to use computer simulations to numerically evaluate FOT-CDFs.

Stationary FOT Ergodic Theorem:

1. Every Stationary FOT-Stochastic Process is *Strongly Ergodic*, by construction, meaning the infinite time averages of relatively measurable functions of the process exist and are independent of the particular sample paths selected and are equal to the expected values of those functions obtained using the FOT-CDF or FOT-PDF.
2. Every Finite-Ensemble Average of every function of a Stationary FOT-Stochastic Process is identical to a Finite-Time Average of that function.

The validity of this theorem follows directly from the Definitions. It is noted here that ensemble averages are typically conceived of as being performed on randomly selected ensemble members, which do not occur in any ordered fashion. In contrast, time averages are typically performed on time-ordered time samples or time translates. Item b) in this theorem does not assume any ordering. However, when one approaches the question of convergence of time averages as the length of averaging time approaches infinity, time ordering is desirable and typically assumed (e.g., as in a Riemann integral), but no such ordering can be assumed for random selection of ensemble members. To avoid the technical details involved here (which are of no pragmatic interest), Item b) addresses only finite averages and, like Item a), states a fact that is obvious from the construction of the sample space.

Relation to Wold's Isomorphism

Wold introduced an isomorphism in 1948 [20], which is referred to here in its extended form that accommodates continuous-time processes, between

(1) the sample space of a stochastic process, defined to consist of the collection of all time translates of a single time function, including that time function itself, and (2) this single time function. This isomorphism establishes a distance-preserving relationship between the stochastic process, with its definition of squared distance as the ensemble-averaged squared difference between two processes, and a single sample-path of that stochastic process, with its definition of squared distance as the time-average of the squared difference between two sample-paths. This mapping between the metric space of a stochastic process and the metric space of a single sample path therefore preserves distance and is consequently an isomorphism. The above sample space is identical to that in Def. S1 for a Stationary FOT-Stochastic Process. By complementing this sample space with an FOT-Probability measure satisfying Defs. S2 and S3, we obtain a Stationary FOT-Stochastic Process. Wold did not take this step, and—according to my literature search—apparently did not pursue the conceptual path taken in the present article.

3.4. Comparison of Kolmogorov and FOT-Stochastic Process Models (The Magic Hand)

To illustrate how simple the sample space of a stationary FOT-stochastic process is, compared with one of the simplest examples of the sample space of a Kolmogorov process, consider an infinite sequence of statistically independent finite-alphabet real-valued equally probable symbols, with alphabet size K. The Kolmogorov sample space for a finite sequence of length N contains K^N distinct sequences and the probability of each is $(1/K)^N$. The probability of the entire sample space is the sum of the probabilities of the K^N mutually exclusive and exhaustive sample paths, each having probability $(1/K)^N$, which sum equals 1. In the limit, as the sequence length approaches infinity, we get the result that the probability of each sample path is 0 and the probability of the total sample space is 1. This sample space includes as a strict subset the entire FOT sample space generated from any one of the Kolmogorov sample paths. The Kolmogorov probability of this FOT sample space is the limit, as N approaches infinity, of $N(1/K)^N$. Therefore, the Kolmogorov probability of the entire FOT sample space is 0. This is a result of the fact that the sample space represents a single signal—a single infinite sequence of $K-$ary symbols, not all possible infinite sequences of $K-$ary symbols. The Kolmogorov sample space apparently contains not only the FOT sample space of all translates of one infinite sequence, but also contains the FOT sample spaces of all translates of every possible infinite sequence of $K-$ary symbols. Despite the huge difference in the sizes of these two sample spaces, as N approaches infinity, it is interesting to note that the FOT probability of a subsegment comprised of a specific sequence of length N occurring over the lifetime of the function is $(1/K)^N$, and this is the same as the probability of selecting a sample path from the corresponding Kolmogorov stochastic process that possesses a particular subsegment of length N comprised of this specific sequence. Because the time position in a stationary time series or a stationary stochastic process is of no probabilistic consequence, the difference in sizes of these sample spaces appears to be of no consequence unless one is interested in studying populations of time series.

As a reminder, the Birkhoff ergodic theorem guarantees that the time average of every sample path in this immense sample space equals w.p.1 the expected value and this equals w.p.1 every ensemble average. This mysterious result is not necessary in practice; it is not a prerequisite for having a probability theory for time-series analysis. The much simpler FOT- stochastic process will do for types of applications described earlier in this paper, for which populations of signals are not of primary interest, and further in this Results section, and this means that the entire stochastic process concept can be discarded for these types of applications and replaced with a single signal and its FOT-probability model. Sample spaces are then irrelevant. The cost of abandoning the Kolmogorov stochastic process model is that the FOT-Probability measure is in general not sigma-additive, and the corresponding FOT-expectation operation is not in general sigma-linear. However, the utility of these sigma properties exists only when performing calculations involving infinitely many subsets of the sample space or sums of infinitely many functions of the process. Moreover, to benefit from these properties, *one must verify that a specified probability measure does indeed exhibit these assumed properties*. This is rarely done in practice, except when well-known probability measures, like the Gaussian, which have already been verified, are adopted. But there are no models for manmade communications signals in use that are Gaussian and the same is apparently true for models of naturally occurring biomedical signals, and signals of many other origins. If there is not a large number of independent samples of random variables added together to form a random variable to be modeled, there is generally no reason to expect that random variable to be Gaussian.

Another way to compare these two models of stochastic processes is as follows. Consider, as an example, a Bernoulli sequence with parameter p = 0.3. This is a sequence of statistically independent binary random variables with values of 0 and 1 having probabilities of 0.3 and 0.7, respectively. A sample path for the Kolmogorov model is denoted by $x(n, \omega)$, where n is integer-valued and ω also need only take on a countable infinity of values, and can therefore be taken to be integer valued. The values this function of two integer variables can take on are 0 and 1. The specification of the actual infinitely large 2-dim array of 0's and 1's is such that every possible sequence of 0's and 1's is included once and only once. So, the specification of the sample space is simply exhaustive. But there is a specification of a probability measure for this function of ω for subsets of values of n. The measure tells us the limit, as the number of randomly selected values of ω approaches infinity, of the relative frequency of sets of 0's and 1's at these subsets of discrete time points that will occur as outcomes. This probability measure is like *a magic hand* that guides the selection of experimental outcomes so that at each time point 1's are selected in 70% of the experimental outcomes and 0's are selected in 30% of the outcomes. And, for example, the pair of adjacent outcomes of 0 followed by 1 are selected in $(0.3)(0.7) = 21\%$ of the outcomes. There is an inherent abstractness here, which I call a *magic hand.* It cannot in general be made concrete or given a concrete interpretation. And it is not a property of the sample space. It is simply a specified rule regarding the randomly selected outcomes of an experiment.

It should be clarified here that the strong law of large numbers [5] establishes that averages over ensembles of random samples converge to expected values w.p.1 *not because of replication* in the sample space (which is not allowed), but rather because of the *magic hand*. Replications of entire sample paths occurring with non-zero probability are disallowed in the Kolmogorov model, as they are in the FOT model; however, for any finite set of time samples, the same finite set of sample path values can occur in infinitely many distinct sample paths all of which differ in at least some of the values at other time points. But the numbers of these partial replicas are determined by nothing more than combinatorics. In contrast, the relative frequency of occurrence in random samples of sets of process values at subsets of time points is determined by only the *magic hand*. This fact is often not recognized in the literature. For example, even the classic book by Middleton [13, Sec. 1.3, pp. 26-27] includes invalid attempts at explaining the convergence of ensemble averages to expected values in terms of replications of sample paths in the sample space. Similarly, for the sample space defining the FOT-stochastic process (e.g., continuous time), replications like $\{x(t-\omega_1);\ t \in \mathbb{R}\} = \{x(t-\omega_2);\ t \in \mathbb{R}\}$, $\omega_1 \neq \omega_2$, are disallowed (Def. S1) because they do not produce what we think of as random functions since they imply $x(t)$ is simply periodic with period $= |\omega_1 - \omega_2|$.

In contrast to the Kolmogorov sample space for the Bernoulli process, a sample path for the corresponding FOT-stochastic process is given by (with some abuse of notation) $\{x(n, \omega); n, \omega \in \mathbb{Z}\} = \{x(n - \omega); n, \omega \in \mathbb{Z})\}$ and this function $x(n)$ takes on values of 0 and 1. Given a single sample path$x(n)$on the integers, we have a full but non-exhaustive specification of $x(n, \omega)$ throughout the entire sample space (2 dim array). Because of this, there is no need for a magic hand. We can derive the probability measure by simply calculating (in principle, at least) the limit of the relative frequencies of 1's in $x(n)$. Any statistical dependence of these binary variables in the sequence also can (in principle, at least) be calculated from joint FOT-probabilities. Work on designing sequences that exhibit specified relative frequencies can be found in the early literature (cf. references at [2]).

The above discussion illustrates that the details and level of abstraction of the Kolmogorov stochastic process model are often not observed in applied work in statistical signal processing. Consequently, there is little pragmatic justification for continuing to hang on to the baggage (abstraction) that comes with this standard model when populations of signals are not of primary concern, when we have the much simpler and more concrete alternative, the FOT-Probability model for single signals.

3.5. Definition of Cyclostationary FOT- Stochastic Process

Def. CS1: The *Sample Space* of the *Cyclostationary FOT-Stochastic Process with Period T* is comprised of all the time translates, by integer multiples of the period, of a single relatively measurable discrete- or continuous-time sample path (persistent real-valued function of a real variable), x, subject to the constraint that replications are disallowed (no two sample paths can be identical):

$$\begin{aligned} \Omega_d &= \{\{x_{n-\omega T}; n \in \mathbb{Z}\}; \omega \in \mathbb{Z}\}, \\ \Omega_c &= \{\{x(t - \omega T); t \in \mathbb{R}\}; \omega \in \mathbb{Z}\} \end{aligned} \tag{15}$$

The period T can be any real number for continuous-time processes but must be an integer for discrete-time processes with time index set equal to the set of all integers $\mathbb{Z}$.

Def. CS2: The probability of any relatively measurable subset of elements defined by some common property the sample paths share, from the sample space index set $\mathbb{Z}$, called an *event*, is the value of the relative measure of that set. (If the function x exhibits statistical cyclicity with period T, then probabilities of time-translated events will, in general, vary periodically in the translation parameter. Otherwise, the probabilities will be translation invariant—a degenerate case of periodicity.)

Def. CS3: The FOT-CDF of any relatively measurable discrete- or continuous-time function, $f[x](t)$ or $f[x]_n$, which is jointly relatively measurable, for m real-valued time points $\{t_1, t_2, t_3, ..., t_m\}$ or m integer-valued time points $\{n_1, n_2, n_3, ..., n_m\}$, of the *Cyclostationary FOT-Stochastic Process* $x(t)$ or x_n, with *Period* T, is the relative measure of the event set

$$\begin{aligned} E_m^c \triangleq \{\omega \in \mathbb{Z}; f[x](t_1 - \omega T) \leq \xi_1, \\ f[x](t_2 - \omega T) \leq \xi_2, ..., f[x](t_m - \omega T) \leq \xi_m\} \end{aligned} \tag{16a}$$

or

$$\begin{aligned} E_m^d \triangleq \{\omega \in \mathbb{Z}; f[x]_{n_1 - \omega} \leq \xi_1, \\ f[x]_{n_2 - \omega} \leq \xi_2, ..., f[x]_{n_m - \omega} \leq \xi_m\} \end{aligned} \tag{16b}$$

for all real-valued m-tuples $\{\xi_1, \xi_2, \xi_3, ..., \xi_m\}$, and all these FOT-CDFs are periodic functions of time: If $\{t_1, t_2, t_3, ..., t_m\}$ is replaced with $\{t_1 + T, t_2 + T, t_3 + T, ..., t_m + T\}$ or, if $\{n_1, n_2, n_3, ..., n_m\}$ is replaced with $\{n_1 + T, n_2 + T, n_3 + T, ..., n_m + T\}$, the FOT- CDF remains unchanged.

It follows from Def. CS3 that the first-order FOT-CDF for a continuous-time cyclostationary FOT process is given explicitly by the formula

$$\begin{aligned} F_{x,T}(\xi, t) &\triangleq \mu_R(\{n \in \mathbb{Z} : x(t - nT) \leq \xi\}) \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \#(\{n \in [n_0 - N, n_0 + N] : \\ &\qquad x(t - nT) \leq \xi\}) \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n = n_0 - N}^{n_0 + N} \mathrm{u}\left(\xi - x(t - nT)\right) \end{aligned} \tag{17}$$

for all real t and ξ, and similarly for higher-order FOT-CDFs (cf. Eq. (14)); and the first order FOT-CDF for a discrete-time FOT process is given explicitly by the formula

$$\begin{aligned} F_{x,T}(\xi,n) &\triangleq \mu_R(\{n\in\mathbb{Z}: x_{k-nT}\le\xi\}) \\ &= \lim_{N\to\infty}\frac{1}{2N+1}\#(\{n\in[n_0-N,n_0+N]: x_{k-nT}\le\xi\}) \\ &= \lim_{N\to\infty}\frac{1}{2N+1}\sum_{n\,=\,n_0-N}^{n_0+N}\mathrm{u}\left(\xi-x_{k-nT}\right) \end{aligned} \tag{18}$$

for all real ξ and all integers n. In contrast to the periodicity, with a single period, of these FOT-CDFs, the FOT-CDFs for a stationary FOT-stochastic process remain unchanged for *all* real-valued or integer-valued T. They are periodic with every period and are therefore time-invariant.

Note: The constraint in Def. CS1 that disallows replications in the sample space also disallows periodic signals, which are a degenerate case of cyclostationary signals. A viable alternative is to remove this constraint.

For this FOT-stochastic process, any one of the translates, $\{x(t-\omega T): t\in\mathbb{R}\}$ for any particular $\omega\in\mathbb{Z}$ or $\{x_{n-\omega T}: n\in\mathbb{Z}\}$ for any particular $\omega\in Z$, can be taken as the *Sample Space Generator*. Observe that, whereas the sample space for the stationary FOT process is uncountably infinite for continuous time, it is only countably infinite for the continuous-time cyclostationary FOT process.

Although not immediately obvious, a single sample-space generator (a single signal) can, in general, generate a stationary FOT process or a cyclostationary FOT process with any one of multiple incommensurate periods. If the single signal exhibits no cyclostationarity, all the FOT- CDFs will be time-invariant and identical. If the single signal exhibits only one period, then its cyclostationary FOT-CDF with this period will be periodic, not time-invariant and it will therefore be distinct from the stationary FOT-CDF. And if the single signal exhibits two incommensurate periods, the sample space generator can generate a time invariant FOT-CDF and two distinct periodic FOT-CDFs, by using different sets of translation amounts. And so on. For the five example signal models specified above, we have the following results for the distinct FOT-CDFs that can be produced from each signal.

Example 1: $x_1(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T=T_1$

Example 2: $x_2(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T=1/2f_2$

Example 3: $x_3(t)$ has stationary FOT-CDF and multiple cyclostationary FOT-CDFs with periods $T^{(j)} = 1/(2f_3 + j/T_3)$ for possibly all integers$j$, assuming that f_3 and $1/T_3$are incommensurate

Example 4: $x_4(t)$ has stationary FOT-CDF and one cyclostationary FOT-CDF with period $T = 1/2f_4$

Example 5: $x_5(t)$ has stationary FOT-CDF and multiple cyclostationary FOT-CDFs with periods $T^{(j)} = 1/(nf_2+mf_3)$ for possibly all pairs of integers $(n,\ m)$ (except those for which $(n_2, m_2) = (kn_1, km_1)$ for any integer k) if f_2 and f_3 are incommensurate; otherwise just one cyclostationary FOT-CDF with period $T = 1/nf_2 = 1/mf_3$ for the smallest pair of integers n, m for which this equality holds.

Cyclostationary FOT Cycloergodic Theorem:

1. Every Cyclostationary FOT-Stochastic Process is *Strongly Cycloergodic*, by construction, meaning the infinite time averages, with cyclostationarity period T, of relatively measurable functions of the process exist and are independent of the particular sample paths selected and are equal to the time-periodic expected values of those functions obtained using the periodic FOT-CDF or FOT-PDF.
2. Every Finite-Ensemble Average of every function of a Cyclostationary FOT- Stochastic Process is identical to a Finite-Time Periodic Average of that function.

The validity of this theorem follows directly from the Definitions. It is noted here that ensemble averages are typically conceived of as being performed on randomly selected ensemble members, which do not occur in any ordered fashion. In contrast, time averages are typically performed on time-ordered time samples or time translates. Item b) in this theorem does not assume any ordering. However, when one approaches the question of convergence of time averages as the length of averaging time approaches infinity, time ordering is desirable and typically assumed, but no such ordering can be assumed for random selection of ensemble members. To avoid the technical details involved here (which are of no pragmatic interest), Item b) addresses only finite averages and, like Item a), states a fact that is obvious from the construction of the sample space.

For each value of t, the indicator function $\mathrm{u}(\xi - x(t))$ takes on values of only 0 and 1 for all real ξ, and its range is therefore contained in the closed interval $[0, 1]$. It is easy to demonstrate graphically that, for each t and for all real ξ_2 and ξ_1, if $\xi_2 \geq \xi_1$, then $\mathrm{u}(\xi_2 - x(t)) \geq \mathrm{u}(\xi_1 - x(t))$. Therefore, for each value of t, $\mathrm{u}(\xi - x(t))$ is a nondecreasing function of ξ. Also, since $\xi - x(t) < 0$ (or > 0) for all finite $x(t)$ when $\xi = -\infty$ (or $\xi = \infty$), then $\mathrm{u}(-\infty - x(t)) = 0$ (and $\mathrm{u}(\infty - x(t)) = 1$). Consequently, for each value of t, $\mathrm{u}(\xi - x(t))$ is a valid cumulative probability distribution function (CDF) of the variable ξ; for all integer-valued time t, this is a discrete-time-indexed set of CDFs and, for all real-valued time t, this is a continuous time-indexed set of CDFs.

It can be shown that any discrete or continuous average of CDFs $\{CDF_n : n = 1, 2, ..., N\}$ or $\{CDF(t) : 0 \leq t \leq T\}$, such as

$$\frac{1}{N}\sum_{n=1}^{N} CDF_n \text{ or } \frac{1}{T}\int_0^T CDF(t)dt$$

is also a valid CDF. The above facts hold true for any finite-order CDF for $\{x(t + t_i) : i = 1, 2, ..., I\}$, not just the first-order CDFs referred to above.

In the limit as the averaging interval covers all time, as in the third line of Eq. (12) or Eq. (13), these averages over the *Indicator-Function CDFs* are referred to as *stationary CDFs* of the function $x(t)$. This is in agreement with the definition of a stationary FOT-Stochastic process given in Section 3.3. Similarly, for averages of the form shown in the third line of Eq. (17) or (Eq. 18), the limits are periodic in time with period T and, if not equal to a constant independent of t, are referred to as t *cyclostationary CDFs* of the function $x(t)$. This is in agreement with the definition of a cyclostationary FOT-Stochastic process given in Section 3.5.

The stationary CDF defined by Eq. (12) or Eq. (13) is what is called the *constant component* of the erratically fluctuating Indicator-function CDF, and the cyclostationary CDF defined by Eq. (17) or Eq. (18) when these are not $t-$invariant is what is referred to as the *periodic component* of the erratically time fluctuating indicator-function CDF. When the stationary CDF is subtracted from the erratically fluctuating Indicator-function CDF, the difference, referred to as the residual, contains no constant component. Similarly, the residual for the periodic component of the erratically time fluctuating indicator-function CDF contains no periodic component with the same period.

The preceding reinterpretation of the CDFs for stationary and cyclostationary FOT-Stochastic processes reveals how to define *almost cyclostationary CDFs* for functions of time that exhibit statistical cyclicity with multiple incommensurate periods, even though we do not know how to construct a corresponding FOT-type of stochastic process model because we do not know how to specify the appropriate sample space to accommodate multiple incommensurate periods of cyclicity.

We begin with a little background information on almost periodic functions. An almost *periodic function* $Q(t)$is one that admits a Fourier series representation of the form

$$Q(t) = \sum_{\alpha} q_\alpha \exp\{i2\pi\alpha t\} \tag{19a}$$

where the index α ranges over a countable (possibly countably infinite) set. In the mathematics literature, various distinct types of almost periodic functions have been defined. In the simplest of terms, they differ from each other in the sense in which the above Fourier series representation converges, and the sense in which the formula

$$q_\alpha = \lim_{U \to \infty} \frac{1}{U} \int_{-U/2}^{U/2} Q(t) \exp\{-i2\pi\alpha t\} dt \tag{19b}$$

for the Fourier series coefficients converges (cf. [3, Appendix B]). Almost periodic functions are literally nearly periodic, which can be expressed mathematically (cf. [3, Appendix B]).

If all the values of $\alpha$are integer multiples of a single value $\alpha_o = 1/T_o$, for which T_o is called the *period,* then $Q(t)$ is a periodic function with period T_o. This is a degenerate form of almost periodicity. More generally, because the set of values of α is countable, there exists an at-most-countable set of incommensurate periods $\{T_k\}$ such that the above Fourier series representation can be re-expressed as

$$\begin{aligned} Q(t) &= \sum_k \sum_j q_{kj} \exp\{-i2\pi(j/T_k)t\} \\ &= q_0 + \sum_k [Q_k(t) - q_0] \end{aligned} \tag{20a}$$

where $Q_k(t)$is periodic with period T_k,

$$\begin{aligned} Q_k(t) &= \sum_j q_{kj} \exp\{\mathrm{i}2\pi j/T_k\} = Q_k(t+T_k) \\ &= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-\mathrm{N}}^{N} Q(t+nT_k) \end{aligned} \tag{20b}$$

and $q_{k0} = q_0$ for all k. Periods are *incommensurate* if no two periods have a ratio that is a rational number.

In the event that the almost periodic function $Q(t)$exhibits only a finite number of incommensurate periods $\{T_k :\ k = 1, 2, 3, ..., K\}$, then we have a degenerate case of almost periodicity that is called *Poly-periodicity.*

Returning to almost cyclostationary CDFs, the set of almost periodic functions of time of interest here are

$$Q(\xi, t) = \sum_\alpha q_\alpha(\xi) \exp\{\mathrm{i}2\pi\alpha t\} \tag{21a}$$

for each and every real value of ξ. The Fourier coefficients in this expression are given by

$$q_\alpha(\xi) = \lim_{U\to\infty} \frac{1}{U} \int_{-U/2}^{U/2} u(\xi - x(t)) \exp\{-\mathrm{i}2\pi\alpha t\}\mathrm{d}t. \tag{21b}$$

We shall use the notation

$$Q(\xi, t) \equiv F_x(\xi, t) \triangleq \sum_{\alpha\in\mathrm{A}} F_x^\alpha(\xi) \exp[\mathrm{i}2\pi\alpha t] \tag{21c}$$

and

$$q_\alpha(\xi) \equiv F_x^\alpha(\xi) \triangleq \lim_{U\to\infty} \frac{1}{U} \int_{-U/2}^{U/2} u(\xi - x(t)) \exp\{-\mathrm{i}2\pi\alpha t\}\mathrm{d}t \tag{21d}$$

to be more consistent with the discourse in earlier sections. Then it follows from Eq. (20) that

$$\begin{aligned} F_x(\xi, t) &\triangleq \sum_{\alpha\in\mathrm{A}} F_x^\alpha(\xi) \exp[\mathrm{i}2\pi\alpha t] \\ &= F_x^0(\xi) + \sum_{\mathrm{k}\in\mathbb{Z}} \left\{ \sum_{\mathrm{j}\in\mathbb{Z}} \left(F_x^{j/T_k}(\xi) \exp[\mathrm{i}2\pi(j/T_k)t] - F_x^0(\xi) \right) \right\} \\ &= F_x^0(\xi) + \sum_{\mathrm{k}\in\mathbb{Z}} \left\{ F_{x,T_k}(\xi, t) - F_x^0(\xi) \right\} \end{aligned} \tag{22}$$

The set of frequencies $A = \{\alpha\}$ are called the *Cycle Frequencies*; they are in general harmonics of the fundamental frequencies $\{1/T_k\}$associated with each periodic component. Unless otherwise stated, the set A contains all cycle frequencies for which the Fourier component $F_x^\alpha(\xi)$ is not identically zero.
The Fourier coefficient functions $\{F_x^\alpha(\xi)\}$comprising the almost cyclostationary CDF are complex-valued and are therefore not themselves CDFs. They are, however, a generalization referred to as complex cumulative distributions with range confined to the unit disc in the complex plane instead of the unit interval of the real line as for real CDFs.
For the special case in which an almost cyclostationary CDF $F_x(\xi, t)$is degenerate in the sense of being poly-periodic, this CDF and the underlying function $x(t)$are both referred to as being *poly-cyclostationary.* An example of this type of function is one for which all finite-order CDFs are Gaussian and have poly-periodically time-varying mean(t)and $cov(t, t+\tau)$ for all time-separations τ, with the collection of periods $\{T_{k(\tau)}\}$ over all real τbeing finite. For poly-cyclostationary CDFs, the sum over the index $k$in the second and third lines of Eq. (22) ranges over only a finite subset of the integers Z; however, the harmonic index$j$for each of the periods $T_k$in their finite set must range over all integers. No integer value of$j$for which the associated Fourier coefficient $F_x^{j/T_k}(\xi)$is not identically zero can be omitted from the sum, without possibly violating the defining properties of a CDF (cf. [3, Chap 2]). An exception is the case for which the entire term in Eq. (22) with any specific period index $k = k_o$is omitted, provided that $1/T_{k_o}$is incommensurate with not only all$\{1/T_k; k = 1, 2, 3, \,.\,.\}$but also with all the integral linear combinations $I_1/T_1 + I_2/T_2 + \; I_3/T_3 + \ldots$ for all integers $\{I_q : q = 1, 2, 3, \; \ldots\}$. [3, Chap. 2]. This means that all but a finite set of the countably infinite set of periods $\{1/T_k\}$can be omitted provided that this requirement is met for all omitted periods.
In other words, for a non-degenerate almost cyclostationary function $x(t)$, exhibiting a countably infinite number of periods $\{T_k\}$of cyclicity, it is possible to extract a poly-periodic component with any finite (size K) subset of these periods from its indicator-function CDF of first order, $u(\xi - x(t))$, or of any finite order, and the result will be a valid CDF, assuming the entire poly-periodic component with the specified periods is extracted and any frequencies contained in the residual are incommensurate with all integral linear combinations $I_1/T_1 + I_2/T_2 + \; I_3/T_3 + \ldots + I_K/T_K\}$.
Although the cyclostationary, polycyclostationary, and almost-cyclostationary

CDFs defined up to this point cannot have arbitrarily selected non-zero terms in their complete Fourier series omitted without possibly violating the required properties of the CDF, any terms can be omitted while still retaining an important property that CDFs possess, which is referred to as the *Fundamental Theorem of Almost-Periodic Component Extraction.* To prove this theorem, we first consider a generalization of the *Fundamental Theorem of Time Averages* Eq. (3).

Let $g(\{x(t)\})$be a well-behaved real-valued function of $\{x(t)\}$, of the form

$$g(\{x(t)\}) = g(x(t+t_1), x(t+t_2), ..., x(t+t_m)), \tag{23}$$

for any finite positive integer m and any set of m time samples $\{t_i : i = 1, 2, ..., m\}$and all real-valued time $t \in S$ for some interval$S$(finite or infinite) of the real line. Let $\mathbf{P}$ be an orthogonal projection operator, to be applied to $g(\{x(t)\})$, for projection onto some linear subspace of functions of t on S. Also, consider the set of projections of the indicator-functions

$$F_{\mathbf{P}}(\boldsymbol{\xi}, t) \triangleq \mathbf{P}\left[\prod_{i=1}^{m} u(\xi_i - x(t+t_i))\right] \tag{24}$$

for all real m-tuples $\boldsymbol{\xi}$.

Fundamental Theorem of Orthogonal Projection of Functions of a Function:

The projection $P[g(\{x(t)\})]$ of *any function* $g(\{x(t)\})$ of the form in Eq. (23) can be calculated from the set of projected indicators functions $F_{\mathbf{P}}(\boldsymbol{\xi}, t)$ for all real $\boldsymbol{\xi}$ as follows:

$$\begin{aligned} P[g(\{x(t)\})] &= \int g(\xi_1, \xi_2, ..., \xi_m) \mathrm{d}^m F_{\mathbf{P}}(\boldsymbol{\xi}, t) \\ &= \int \int ... \int g(\xi_1, \xi_2, ..., \xi_m) f_{\mathbf{P}}(\xi_1, \xi_2, ..., \xi_m) \mathrm{d}\xi_1 \mathrm{d}\xi_2 ... \mathrm{d}\xi_m \end{aligned} \tag{25}$$

where $f_{\mathbf{P}}(\xi_1, \xi_2, ..., \xi_m)$ is the density function corresponding to the distribution function:

$$f_{\mathbf{P}}(\xi_1, \xi_2, ..., \xi_m) \triangleq \frac{\partial^m}{\partial \xi_1 \partial \xi_2 ... \partial \xi_m} F_{\mathbf{P}}(\boldsymbol{\xi}, t). \tag{26}$$

Examples of this theorem for $m = 1$ include the following special cases of *Fundamental Theorems of Almost-Periodic Component Extraction:*

1. Stationary Component Extraction (cf. Eq. (12)):

$$P[g(\{x(t)\})] \triangleq \lim_{U \to \infty} \frac{1}{U} \int_{t_0-U/2}^{t_0+U/2} g(\{x(t)\})\mathrm{d}t$$

2. Cyclostationary Component Extraction (cf. Eq. (17)):

$$P[g(\{x(t)\})] \triangleq \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n\,=\,n_0-N}^{n_0+N} g(\{x(t-nT)\})$$

3. Almost Cyclostationary Component Extraction (cf. E2. (21)):

$$P[g(\{x(t)\})] \triangleq \sum_{\alpha\,\in\,\mathrm{A}} g_x^{\alpha} \exp[\mathrm{i}2\pi\alpha t]$$

   where

$$g_x^{\alpha} \triangleq \lim_{U \to \infty} \frac{1}{U} \int_{t_0-U/2}^{t_0+U/2} g(\{x(t)\}) \exp\{-\mathrm{i}2\pi\alpha t\}\mathrm{d}t$$

   and A is a countably infinite set of any real numbers including any incommensurate numbers.

4. Poly-Cyclostationary Component Extraction: Same as Example (3) but for (a) only a finite set A of real numbers or (b) a countably infinite set A, each member of which is an integer multiple of one of only a finite set of incommensurate fundamental frequencies $\{1/T_k; k = 1, 2, ..., K\}$.

Other examples can include subspaces that are finite dimensional or that contain only functions having time domains that are finite intervals of the real line (cf. [2, page 3.5]). Interestingly, this theorem is valid for projections that do not produce CDFs (cumulative probability distribution functions); this includes some cases within Example (4) and an unlimited number of other examples. A frequently used subspace projection in statistical signal processing is that spanned by a subset of eigenvectors of the signal's covariance matrix. Thus, the fundamental theorem of time averaging is a special case of this more general theorem for more general projections. Nevertheless, the Projections must be orthogonal projections in order to apply the above theorem. (Non-orthogonal projections do not extract components of a function, because the residual still includes some of this same component.) For example, sinusoids with incommensurate frequencies are not orthogonal over any finite-length interval of time. Therefore, finite-time CDFs can be orthogonal projections only if all sinusoidal components are harmonics of a single fundamental frequency; i.e., they must be cyclostationary, not poly-cyclostationary, and the time-interval over which the CDF is defined must be an integral number of periods. The required modification of the basis functions $\{\exp[i2\pi\alpha t]\}$ to render the above theorem applicable for any finite-length time interval is described in Section 3.8.

Outline of Proof of Fundamental Theorem of Orthogonal Projection of Functions of a Function:

The density function on the RHS of Eq. (24) is given by

$$\begin{aligned} f_{\mathbf{P}}(\xi_1,\xi_2,...,\xi_m) &\triangleq \frac{\partial^m}{\partial\xi_1\partial\xi_2...\partial\xi_m}F_{\mathbf{P}}(\boldsymbol{\xi},t) \\ &= \frac{\partial^m}{\partial\xi_1\partial\xi_2...\partial\xi_m}\mathbf{P}\left[\prod_{i=1}^{m} u(\xi_i - x(t+t_i))\right] \\ &= \mathbf{P}\left[\frac{\partial^m}{\partial\xi_1\partial\xi_2...\partial\xi_m}\prod_{i=1}^{m} u(\xi_i - x(t+t_i))\right] \\ &= \mathbf{P}\left[\prod_{i=1}^{m} \delta(\xi_i - x(t+t_i))\right] \end{aligned} \tag{27}$$

where the order of the projection operation on a function of t and $\boldsymbol{\xi}$ for each value of $\boldsymbol{\xi}$ and the differentiation operation on this function of t and $\boldsymbol{\xi}$ for each value of t have been interchanged. Substituting this into the RHS of Eq. (24) yields

$$\begin{aligned} P[g(\{x(t)\})] &= \int g(\xi_1,\xi_2,...,\xi_m)\mathrm{d}^m F_{\mathbf{P}}(\boldsymbol{\xi},t) \\ &= \int g(\xi_1,\xi_2,...,\xi_m)\mathbf{P}\left[\prod_{i=1}^{m}\delta(\xi_i - x(t+t_i))\right]\mathrm{d}\xi_1\mathrm{d}\xi_2...\mathrm{d}\xi_m \\ &= \mathbf{P}\left[\int g(\xi_1,\xi_2,...,\xi_m)\left[\prod_{i=1}^{m}\delta(\xi_i - x(t+t_i))\right]\mathrm{d}\xi_1\mathrm{d}\xi_2...\mathrm{d}\xi_m\right] \\ &= \mathbf{P}\left[g(x(t+t_1),x(t+t_2),...,x(t+t_m))\right] \end{aligned} \tag{28}$$

where the order of the projection operation on the slice of the function of t and $\boldsymbol{\xi}$ for each value of $\boldsymbol{\xi}$ and the integration operation on a slice of this function of t and $\boldsymbol{\xi}$ for each value of $t$have been interchanged. Here, the final line in the RHS is the definition of the LHS; so Eq. (24) is verified.

3.7. Cycloergodicity for Multiple Incommensurate Periods

Many communications signals with sample paths specified formulaically exhibit cyclostationarity with multiple incommensurate periods (they are *poly-cyclostationary* or *almost cyclostationary*, but not purely cyclostationary or purely stationary) and, as shown by Boyles and Gardner in 1983 [21], they can be tested for what is here called *Sinusoidal Ergodicity* (SE). This means some such processes can exhibit the strong sinusoidal ergodic properties required to support the commonly assumed convergence of estimates of sinusoidal components (which are typically called *cyclic components*) of their almost-periodically time-varying probabilistic parameters, such as cyclic autocorrelations and cyclic spectral densities (also called spectral correlation functions). However, *these processes cannot be included in the traditional ergodic theory stemming from Birkhoff's work or its extension to the cycloergodic theory of cyclostationary processes of Boyes and Gardner.* This is mathematically proved in [21] and illustrated with the example of a Bernoulli process with a periodically time-varying probability of success having its period incommensurate with the sampling-time increment. What has essentially invariably been done since the introduction of almost cyclostationary processes in 1978 [22] is to specify such processes in a formulaic manner (e.g., Examples 3 and 5 above) and to then invoke a strong cycloergodic hypothesis, sometimes based on the demonstration of a much weaker form of cycloergodicity, such as cycloergodicity in the mean square sense. But we are now going to go beyond this by building on the concepts introduced in earlier sections.

The sample spaces for the cyclostationary FOT- stochastic processes reveal why *there cannot exist a single FOT-stochastic process with more-than-one incommensurate period*: A single sample space cannot consist of only translates of one period if it also consists of only translates of another incommensurate period. What one must therefore do with the FOT model introduced in Section 3.6 is to introduce a unique sample space for each and every incommensurate period of cyclostationarity of interest for a single record of data or a single formulaic model. However, this is just a conceptual aid. For operational purposes, all one needs is the formula for almost cyclostationary CDFs given in Section 3.6 (third line of Eq. (22) and the method presented in Section 3.5 for calculating cyclostationary CDFs for each period. This calculation can be empirical, using a record of observed data, or it can be performed mathematically using a formulaic specification of the time series. This, in turn, provides insight into how to generalize Birkhoff's ergodic theorem to accommodate almost cyclostationary processes of the Kolmogorov

type, as explained next.
But first, let us sum up the situation for formulaic FOT-Probability models for almost cyclostationary time series. Deterministic periodicity with multiple periods, combined in a sample-path formula (such as those in Examples 1 – 5), with stationary FOT time-series components, provides the basis for constructing the CDFs or PDFs from FOT calculations using the time-series model. Nonlinear functions of a time series whose sample-path formula contains multiple periodicities contain in general not only harmonics not originally present, of the fundamental frequencies originally present, but also linear combinations with integer-valued coefficients, of all these harmonics. Consequently, in constructing the CDFs for such a time series, it must be assumed at the outset that the CDFs will contain sinusoidally time-varying components with all these various mixed frequencies.

How to Generalize Birkhoff's Ergodic Theorem for Continuous-Time Almost Cyclostationary Kolmogorov Stochastic Processes

The content of this section does not contribute to the primary objective of this article, but it does follow easily from the concepts introduced in the previous section and it does provide a genuine generalization of ergodic theory of stationary and cyclostationary processes to poly-cyclostationary and almost cyclostationary Kolmogorov stochastic processes. *Strong Cycloergodic theory of Kolmogorov stochastic Processes*, which extends and generalizes existing ergodic theory, is developed in [21], where it is shown that sinusoidal and periodic components of time-varying probabilistic parameters can be consistently estimated w.p.1 from time averages on one sample path. It is also established that a strong theory of cycloergodicity inclusive enough to cover all applications of practical interest had, at that time, not yet be shown to exist. Moreover, it is shown that such a theory cannot presuppose the existence of a dominating stationary measure, as does the theory presented therein. Nevertheless, it would appear that it can be argued that because a continuous-time cyclostationary process can be characterized as a discrete-time vector-valued (or function-valued) stationary process, Birkhoff's Ergodic Theorem [16] for scalar-valued discrete-time stationary processes, if generalized to vector-valued processes, leads to a completely analogous cycloergodic theorem for continuous-time cyclostationary processes. The vector (or function), at any discrete time equal to an integer multiple of the period of cyclostationarity, consists of the infinite set of process values over the period between that discrete time and the previous discrete time.

Furthermore, it is shown in [23, Chap. 7] and refs. therein that Birkhoff's ergodic theorem has been extended from stationary to asymptotically mean-stationary (AMS) discrete-time processes. This extension guarantees the existence of consistent estimators for the discrete-time averages of time-varying probabilistic parameters, such as probability density functions. Because almost-cyclo-stationary (ACS) discrete-time processes are AMS, this extended theorem applies to discrete-time ACS processes (and the same might well be true for continuous-time ACS processes after discrete-time sampling) but it does not apply directly to estimation of the sinusoidal and periodic components of almost-periodically time-varying probabilistic parameters.

Nevertheless, [23, Chap. 7] does discuss ergodicity of $N-$stationary discrete-time processes, which are N-dimensional vector-valued representations for discrete-time cyclostationary processes with period N. Furthermore, the discrete-time infinite-dimensional vector-valued process described above that represents a continuous-time scalar-valued process is AMS if that continuous-time process is ACS (which includes, as special cases, poly-cyclostationary, cyclostationary, and stationary processes).

Consequently, for any selected period of a continuous-time ACS process, one can form a discrete time vector-valued AMS process as explained above. Then the time average of a probabilistic parameter of this vector-valued process will equal the periodic component of the corresponding probabilistic parameter of the original ACS process. In this way any periodic component for any real-valued period $T$of the almost periodically time-varying probabilistic parameters of the original scalar-valued continuous-time ACS process can be guaranteed to be consistently estimable by applying the proposed ergodic theorem to the infinite-dimensional vector-valued discrete-time AMS process.

It follows that the discrete-time AMS version of the Birkhoff ergodic theorem can be extended / generalized to accommodate cycloergodicity for continuous-time ACS processes by requiring that the ergodicity condition for discrete-time AMS processes be satisfied by the vector-valued representation for each and every period $T$of the continuous-time process. In addition, there appears to be a partially cycloergodic version of this proposed theorem that requires the ergodicity condition for some but not all periods be satisfied.

This leaves one class of ACS processes for which a cycloergodic theorem remains to be proposed, and this is the class of discrete-time processes having measures that possess non-zero sinusoidal components with sine-wave frequencies that are incommensurate with the time-sampling rate. Some such processes do indeed allow for consistent estimation of such sinusoidal components, but others do not. A necessary and sufficient condition for consistent estimation has apparently not yet been proposed but the Author suspects one will be discovered by following ideas in the present paper.

3.8. Purely Empirical FOT-Probability Models for Regular Cyclicity

As explained below, we can obtain finite-data probability models by using the FOT-CDF formula (21d) in Section 3.6, but without taking the limit as the averaging time approaches infinity, and still get CDFs that are exactly constant (using only $\alpha = 0$) or periodic (using only $\alpha = j/T$ for all integers j) or poly-periodic (using only $\alpha = j/T_k$ for all integers j and any finite set of incommensurate real-valued periods $\{T_k : \ k = 1, 2, ..., K\}$) for continuous time. However, if we use more than a finite number of integers j we cannot properly call the CDF *empirical*. So, we consider here only finite numbers of cycle frequencies. However, omission of some cycle frequency harmonics of a periodic component for which the Fourier coefficients are not identically zero renders the formula for the CDF only approximate. Such approximations do not necessarily retain all the characteristic properties of valid CDFs, such as having range confined to the closed interval [0,1].

Nevertheless, it is expected that the approach with finite numbers of harmonics for continuous time can produce accurate approximations if the number is sufficiently large. In addition, the *Fundamental Theorem of Orthogonal Projection of Functions of a Function* does apply to such approximate Empirical FOT-CDFs using only $\alpha = 0$ or $\alpha = j/T$ and $\alpha = -j/T$ for finite numbers of integers j because such CDFs are still valid orthogonal projections on finite intervals (of length equal to an integral number of periods T).

More generally, the program of calculation for any probabilistic parameters, such as joint moments, using a finite segment of data $x(t)$, is that everywhere the data occurs, in the infinite-interval formula for the probabilistic parameter of interest [1], for some function of the data that is of interest, such as a lag product, the time support of that data is windowed to the finite observation interval, just like what is done in the conventional correlogram & cyclic correlogram, and periodogram & cyclic periodogram [1]. Then the time-invariant Fourier coefficient of the sinusoidal component, with frequency α of interest, of the function of the time series over the finite observation window is extracted and multiplied by $\exp[i2\pi\alpha t]$ (with t extending over the reals) in the usual manner, but without the limit as integration time approaches infinity. These components when added together for all detected or selected cycle frequencies comprise an almost periodic function over all time and, when restricted to the finite time support of the function of the data, comprise an approximation to that function. The approximation is not a least-squares fit because the sinewave components are not mutually orthogonal except over the entire real line unless their frequencies are commensurate. It also does not equal the limit almost periodic component, but it would hypothetically converge to it as the observation time approaches infinity, provided that the function is relatively measurable. But the theory does not use the limit together with conditions for or assumptions of convergence in the limit. It simply uses the finite time statistics (approximate Fourier components) that are artificially extended over all time. These extracted almost periodic representations can be used just as they are used in the limit theory (while recognizing that there are some approximations involved) and can be calculated from either a finite-time record of$x(t)$or an explicit mathematical model of $x(t)$.

Nevertheless, the finite-harmonic component extracted from the data can be made a least-squares fit by simply recognizing that because the harmonic frequencies $\{\alpha_j\}$ in the extracted component

$$\sum_{j=-J}^{J} F_x^{\alpha_j}(\xi)\exp[\mathrm{i}2\pi\alpha_j t]$$

are not integer multiples of the fundamental frequency that is the reciprocal of the length of the time interval of the data, the basis functions $\{\exp(i2\pi\alpha_j t); -J \leq j \leq J\}$ are not orthonormal and therefore are not self-reciprocal, but there reciprocal basis $\{\theta_l(t) : -J \leq l \leq J\}$ can be calculated using the inverse of the Gram matrix as follows:

$$\boldsymbol{\theta}(t) = \mathbf{G}^{-1}\mathbf{e}(t)$$

where $\mathbf{e}(t)$ is the column vector with elements $\{\exp(\mathrm{i}2\pi\alpha_j t); -J \leq j \leq J\}$, $\boldsymbol{\theta}(t)$ is the column vector with elements $\{\theta_l(t) : -J \leq l \leq J\}$, and $\mathbf{G}^{-1}$ is the inverse of the Gram matrix $\mathbf{G}$ with jl^{th} element

$$G_{jl} \triangleq \int_{-U/2}^{U/2} e_j(t)e_l^*(t)\mathrm{d}t$$

Then by replacing $(1/U)\{\exp(-i2\pi\alpha_j t); -J \leq j \leq J\}$ with $\{\theta_l(t) : -J \leq l \leq J\}$ in Eq. (21d) and omitting the limit operation, Eq. (21c) becomes the least-squares-fitting finite-harmonic component of the time varying indicator function $\mathrm{u}(\xi - x(t))$ for each value of ξ. In this case, the entire component of interest is extracted from the indicator function: the residual contains none of this component. In order for the extracted component described above to be real-valued, it is required that $\alpha_{-j} = -\alpha_j$.

The data windowing used does not affect the theoretical equality of the two calculations of an extracted component—one from a finite-time data record and the other from a mathematical formula for the data—provided that the data record is producible from the mathematical model, except for the difference between the values of the random elements in the mathematical model and the actual values of those elements in the record of data, such as the amplitude sequence in an amplitude modulated periodic pulse-train signal. The link here, which replaces the ergodic theorem, is the assumption that the single data record is indeed a segment of one translate of a single time series and that the functions of this time series that are of interest are relatively measurable. This then enables a standard type of argument that agreement between the two methods of calculation can be made as close as desired to each other and to their infinite-time limit by using a long-enough finite-segment of data [4].

All the usual tools still apply. For example, the proof of the central limit theorem for FOT- probability [24] is applicable to the theory for finite records by simply arguing that for any arbitrarily small error, epsilon, in equality between the limit quantity (Gaussian distribution) and the measured quantity, one can in principle choose a finite record length that is long enough to achieve an error size not exceeding epsilon.

There's nothing here of any technical sophistication. The novelty is in recognizing that finite-time FOT models that are precisely stationary or poly-cyclostationary can be constructed from a finite record of data, and these models can be used for all the usual probability calculations to within some finite accuracy determined by the length of the data segment and particular cycle frequencies used. The sensitivity of the accuracy to the numbers of harmonics of each fundamental frequency that are used increases as the degree of nonlinearity of the function of the data increases. A second-order lag product, for example, has a low degree of nonlinearity, but the step discontinuities of the indicator function used to calculate CDFs results in a high degree of nonlinearity.

In the Fourier-coefficient formulas for the function (of the data) of interest, consisting of a lag product of any finite order, the time-shifted finite segments of data will force the integrand to be zero outside of a subinterval defined by the intersection of the time-translated finite-segment support intervals and the integration interval. Assuming all time-shifts of interest are much smaller than the segment length, this approach is acceptable. But it will window the $n-$dim space of n time shifts. Assuming desired spectral resolution width in any spectral parameters (PSD, SCF, Poly-Spectra, etc.) is larger than the reciprocal of the smallest value,$U - \max\{|t_i - t_j|\}$, for data-segment length U, where $\{t_i\}$ denote the lag values, the achieved spectral resolution can be acceptable. Ideally, we'd like this smallest value to be much larger than the coherence length of $x(t)$ (here meant to be the time separation between time samples that is just large enough to result in negligible statistical dependence) to ensure statistical reliability.
A refinement that should moderately improve reliability and reduce bias is to truncate the integration interval involving time-shifts $\{t_i\}$ to the closest integer multiple of $1/\alpha$ that does not exceed $U - \max\{|t_i - t_j|\}$. For more detail on the definitions of finite-time FOT CDF's, see [2, p.3.5].

3.9. Purely Empirical FOT-Probability Models for Irregular Cyclicity

Cyclicity is ubiquitous in scientific data, but for many if not most natural sources of data, the cyclicity is irregular: the period of cyclic time-variation itself changes with time, slowly in some applications and rapidly in others. One approach to accommodating this is to restrict cyclostationarity modeling to data segments that are short enough for the period to be treated as if it were constant. A more general and less restrictive approach is to hypothesize that the irregularity results from a time-warping of an otherwise regular cyclicity. This is true for some irregularly cyclic data sources and not true for others, such as rotating machine vibrations with time-varying rotational speed as explained in [14]. Fortunately, there is a middle ground of natural sources of data for which the irregular cyclicity—though irregularly fluctuating too rapidly to treat as locally regular—is due to time warping of otherwise regular cyclicity and the rate of variation of the warping function is slow enough to be tracked. A broadly applicable approach to doing this is introduced in [14] and is based on the concept of *property-restoral adaptation.*

Methodology and algorithms for such adaptation are presented therein for restoral of regular cyclicity. The adaptation process produces both a timedewarped version of the original data, which is more nearly cyclostationary, and explicitly identifies the dewarping function. In some applications, identification of the warping function inherent in the data, by inverting the identified dewarping function, is the end goal for this time-series analysis; in other cases, further time-series analysis that exploits the restored cyclostationarity is the end goal. In this latter case, by preprocessing data that exhibits irregular cyclicity to restore cyclostationarity enables the user to go on to construct cyclostationary FOT-Probability models. These models can be used directly for some applications and can be time-warped to obtain irregularly cyclic probability models. A generally applicable rule of thumb for predicting how well this methodology can perform is described in [14] in terms of a comparison between (1) what can be called the *coherence time* (or *statistical dependence time*) of the data or the *data memory length* and (2) the *constancy time* (reciprocal of some measure of the rate of time variation) of the warping function. Best performance is expected when (2) exceeds (1) by a factor much larger than unity. This is akin to the well-known concept of *local stationarity* but generalized to *local cyclostationarity* and similarly for the more esoteric and less precisely defined concept of *local ergodicity* generalized to *local cycloergodicity.* But fortunately, such abstractions are avoided when using FOT-Probability models. Complementary work on property-restoral de-warping has been conducted in [25] and references therein, and [2, pp. 4.2, 4.3].

3.10. The Weakness of Mean-Square Ergodicity

For readers who have been indoctrinated in stochastic process theory, a question that might be arising at this point is: "where does the concept of *mean-square (m.s.) ergodicity* and *ergodicity in probability (weak ergodicity)*, as distinct from the *ergodicity w.p.1* or *strong ergodicity* discussed above, arise in the considerations discussed in the earlier sections of this paper?". Typical engineering textbooks, such as the popular book by A. Papoulis [26], do not treat strong ergodicity. The fact of the matter is that m.s. and weak ergodicity and their extension / generalization to m.s. and *weak cyclo-ergodicity* introduced by Boyles and Gardner [21] (see also [3]) is of some use in analytical work. But it must be realized that these forms of ergodicity are *much* weaker than strong ergodicity. For example, m.s. ergodicity guarantees that the squared difference between a time average and an ensemble average (both possibly modified for cyclostationarity) goes to zero in the limit as averaging time approaches infinity, but *only* on average over the typically infinite ensemble. Therefore, this difference need not go to zero for many members of the ensemble. And these members need not be exotic as may those that may be present but are ignored by using the w.p.1 (with probability one) modifier of equality. One might think that because squared error cannot be negative, the average squared error can be zero only if every individual error is zero. But this is not true because we are considering infinitely many errors and every individual sample path occurs with probability zero: It gets zero weight in the weighted average that is the expected value. This is easier to see for temporal mean squared error. For continuous-time averages, a countably infinite number of errors can be non-zero while the average is still equal to zero. Although less commonly known, the average over all time can be zero even if the error at an uncountably infinite number of times is non-zero. The error can be non-zero throughout any finite interval, while the average error over all time is zero. Such are the vagaries of infinity. Consequently, signal processing engineers designing algorithms based on a theory of expected performance using a model that is only m.s. or weakly ergodic can be surprised by the occurrence of sample paths for which time averages differ greatly from the expected values used in the design.

3.11. Optimum and Adaptive Statistical Inference

If a signal processing algorithm for statistical inference adapts to the data as time progresses, it will adapt using its own time-averages, not expected values. This suggests that FOT-Probability analysis of the solutions that adaptive algorithms converge to might be more relevant than stochastic probability analysis. Yet, the opposite is apparently true for investigating the convergence process itself, since this process is transient, not persistent, and can be modeled as a non-ergodic stochastic process but cannot be usefully modeled in terms of FOT-Probability cf. [27] and [28].

Examples of fixed optimum vs. adaptive algorithms are fixed Wiener filters vs. adaptive filters using least-mean-squares (LMS) or recursive least squares (RLS) adaptation algorithms or some type of property-restoral (PR) adaptation algorithm. Also, for parameter estimators, detectors, and classifiers, as well as filters, there are fixed optimized implementations and there are adaptive implementations using, for example, PR algorithms such as modulus-restoral and cyclostationarity-restoral algorithms [29].

Besides the issue of deciding what type of probability model to use for design and analysis of adaptive signal processing algorithms, similar questions arise for optimum algorithms, such as optimum filters. That is, one can minimize time-averaged squared error using an FOT-Probability model or minimize expected squared error using a stochastic process model. If the stochastic process model is strongly ergodic, the solution and performance of the optimum filter will be the same (w.p.1) as it is for an FOT-Probability model for a sample-path of that process. However, if the model is only mean-square ergodic, the solutions and performances will be equal only in the sense of zero mean-squared differences. However, if the stochastic process model is non-ergodic, there is no known time-series model for which the solution and performance obtained using FOT-Probability would be the same. It comes down to the question "what does the practitioner want to model: averages over time or averages over ensembles?" It depends on the application and real-world objectives. The teachings in our colleges today presuppose that only stochastic process models and associated theory need be learned. This is a mistake that needs to be rectified.

4. Discussion of Results

We have known for nearly a century that Birkhoff's Ergodic Theorem, extended from discrete-time to include continuous-time, provides a condition on the sample space and probability measure of Kolmogorov's generic stochastic process model that makes convergent time-averages of measurements on (functions of) the process converge, with probability equal to 1 (w.p.1), to expected values of those measurements. And, we also have known all this time that Kolmogorov's Law of Large Numbers proves that ensemble averages converge to expected values w.p.1. However, practitioners using these results are generally unable to understand, with any appreciable level of intuition, why these equalities between fundamentally different entities are valid.

In contrast, the alternative and greatly simplified stochastic process models introduced in this paper are transparent. It is obvious why time averages equal ensemble averages, because the sample space consists of time-translated versions of a single signal, and it is obvious why these both equal expected values defined in terms of Fraction-of-Time Probability.

In applications where we are interested in only ergodic processes, there does not appear to be any pragmatic reason for adopting the complicated abstract Kolmogorov model of a stochastic process instead of the simpler more concrete alternative stochastic process model. In fact, once we've accepted the alternative model as sufficient for our purposes, we can take the next step of recognizing that this alternative model is identical to the entity comprised of a single signal and its Fraction-of-Time (FOT) Probabilities which are derived directly from this single signal. The conclusion is that sample spaces and stochastic processes are unnecessary unless non-ergodic models of data are the entities of interest, in which case Kolmogorov's model may be a good choice.

This is a situation where a pragmatic person would ask "what's the point of teaching students of statistical signal processing about the strongly ergodic Kolmogorov stochastic process model as a tool for problem solving, with its unnecessary abstraction and its ergodic hypothesis which can almost never be tested in practice, when the model of a single time series (a persistent function of time), together with the concrete time-average operation is operationally equivalent? If we hold to the principle of scientific parsimony and we value mathematical elegance and we act logically and rationally, shouldn't we terminate this nearly-one-century-long practice immediately? It is relevant here that it has been said:

If elegance in science is just an attractive attribute, then elegance is not a necessary goal but simply something to be admired when it happens. However, if elegance is a requisite feature of good science, then the characteristics defining elegance deserve the same attention given to scientific rigor.

To be sure the ramifications of what is stated above are understood by the reader, it is also stated explicitly here, and shown in [1] (see also [9] and [15, Chap. 1]) that the temporal-expectation (time-average) operation behaves just like the stochastic-expectation operation and produces all probabilistic quantities we are familiar with: cumulative probability distribution functions, probability density functions, moments, characteristic functions, cumulants, etc. For example, both operations obey a Fundamental Theorem of Expectation. It's just that:

For temporal expectation, the term probability means (1) *Fraction of Time (FOT) of occurrence of an event at a set of times with specified time-separations, over all translations of that set covering the temporal lifetime of the time series, instead of* (2) *fraction of repeated experiments (each producing a time-series over a full lifetime) for which an event occurs at a particular set of times.*

There are two exceptions to this equivalence, and they are the sigma linearity property of expectation and the relative measurability property of single time functions; these properties are simply dictated by the creators of these two models: the first by the Kolmogorov Axiom VI and the second by the Kac-Steinhaus Axiom of Relative Measurability. Axiom VI may or may not be satisfied by a stochastic process model that some practitioner specifies. And relative measurability is not necessarily satisfied by all the time-series models practitioners may specify. For example, the samples paths of a strongly ergodic continuous-time stochastic process are not necessarily relatively measurable; so, this property must be assumed (call it Axiom VII) for the strongly ergodic stochastic process for continuous time if the limits of time averages in the Birkhoff ergodic theorem are to exist. Although there's no question that sigma additivity of probability measures and sigma linearity of expectation can be useful mathematically, users can rarely verify that the models they use actually exhibit these properties. Nice mathematical properties for both stochastic processes and single time functions come at a cost of restricted applicability. This is the nature of models, especially those involving infinity. It is not necessarily a basis for arguing the superiority or inferiority of the ensemble-average theory over the time-average theory. More in-depth analysis of this topic is provided in [6]. But it is important to mention here that just because the use of the relative measure (time-averaging operation) does not generally enable the user to interchange the limit in the time-averaging integral with the summation over a countable infinity of additive terms does not mean that one cannot proceed with such a calculation. It's just that the interchange of operations must be executed before an attempt to take the limit is made. In some cases, this is required only for the limit that defines the time average; in other cases, it may be required also for a limit that defines an infinite summation.

For example, some continuous-time functions for which averages over discrete times exist may not be relatively measurable on the real line and therefore may not be averageable over all real time. This requires the addition of a 7^{th} axiom to Kolmogorov's stochastic process model to accommodate Birkhoff's ergodic theorem for continuous time averages. As another example, the Channel Coding Theorem of Information Theory cannot be based on FOT-Probability because it is formulated in terms of a non-ergodic stochastic process: The stochastic-process output from any and every random channel except for a random time-delay, is non-ergodic, regardless of whether or not the channel input is ergodic. (The random-delay exception is not allowed for cycloergodicity.) For example, Middleton's classic models of non-Gaussian noise are non-ergodic, because these noise models depend on random time-invariant parameters such as the random number of noise sources seen by the receiver and their random locations relative to the receiver (see, for example, [18], and references therein).

As another example, the theories of maximum-likelihood parameter estimation and hypothesis testing are based on the likelihood function, which is the PDF of the observed data, conditioned on each specific hypothesis and/or hypothetical parameter value of interest. Also, Bayesian minimum-risk parameter estimation and hypothesis testing inference rules can be expressed in terms of likelihood functions. Consequently, these theories and methodologies can only be based on FOT-Probability if conditional FOT-Probabilities and/or PDFs can be experimentally measured or mathematically calculated from mathematical sample-path (time function) models of the data. Frequently this can indeed be done as demonstrated with many examples in [1], [3], [15]. However, it cannot always be done.

Motivated by a full recognition of the issues surfaced in the above discussion in this Section 4 and underscored by a deep appreciation for the ramifications to the practice of statistical signal processing design and analysis, I developed the comprehensive theory and methodology of FOT-Probability and statistical spectral analysis that is presented in the 35-year-old book [1]. This book extends and generalizes the theory from stationary time series to cyclostationary, poly-cyclostationary, and almost cyclostationary time series, which provide higher fidelity models of many time series encountered in engineering and the sciences, where there is some form of underlying statistical cyclicity. This extension / generalization of theory and method has engendered many new and higher-performing signal processing algorithms over the last 35 years—the application to random vibrations from rotating machinery being one of many applications. The similar-vintage book [10] provides the theory of the stochastic-process counterpart of cyclostationarity. A much more recent and more comprehensive book on both the stochastic-process and time-series models is also available [3] and is recommended. This latter book is encyclopedic and is the most scholarly treatment of cyclostationarity available today.

So, the failure of the community to adopt the more pragmatic and less abstract data models delineated in this literature is not due to any lack of theoretical foundation or lack of detailed theoretical and methodological framework built upon that foundation for conducting statistical signal processing design and analysis. It is solely due to indoctrinated people's propensity to avoid changing their ways of thinking. It has been more than a century since the celebrated physicist Max Planck wrote [2, p. 7.1]:

"A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and

a new generation grows up that is familiar with it."

To my knowledge, stochastic process models of speech have not proven to be of much use in practice, but this makes speech a useful example here for illustrating the problems that can arise when using stochastic process models that are not ergodic.

Non-ergodic models of signals do have their uses. Specifically, when important conditions of an experiment change from one trial of the experiment to another, the impact revealed in an ensemble average of these changes cannot be determined from a time average on the time series from a single experimental trial. In the case of speech, the character of speech differs from one speaker to another due to physiological, language, accent, and even emotional-state differences. So, an ergodic stochastic process model is inappropriate. If one wants to design a speech processing algorithm that provides optimum performance averaged over all speakers in a diverse group, a non-ergodic stochastic process model for the speech can, in principle, be used. However, if one wants to design a *data-adaptive* algorithm that provides optimum performance for each and every speaker, then the expected values analytically derived from a non-ergodic model are irrelevant to the design, analysis, and performance of the algorithm. The speech statistics required by the algorithm are learned and adapted to for each individual signal. If probability models are to be useful for studying the output time series converged to by an adaptive speech processor, they would have to be FOT-Probability models.

The same remarks apply for applications involving communications channels that introduce noise or interfering signals that is collectively modeled in terms of multiple noise and/or signal sources, random in number, and with multiple locations, random in their coordinates, relative to the receiver [18].

To illustrate how far the proposed paradigm shift can take us, its extension from time-series models of infinite length to those of finite length, which is introduced in Section 3.8, is briefly resurfaced here.

Finite-time time-average statistics are ubiquitous in statistical signal processing algorithms, and such algorithms are typically implemented with DSP software and/or hardware, which greatly facilitates adaptivity. The potential for considerably higher fidelity of the FOT-Probability models and the fact that these models, using idealized infinite-time averages follow essentially all the same rules for finite mathematical manipulation as do stochastic process models, should encourage DSP algorithm designers to use FOT-Probability models in place of the traditional stochastic process models. And it is important to note that, as discussed in this paper, the Fundamental Theorem of Time Averaging applies to not only limits of time-average statistics but also finite-time averages: it applies to completely empirical quantities! Yet, there is a caveat: For the models derived from finite-time averages, some properties of the expectation and infinite-time-average models are only approximated. This appears to be more of an issue with poly-cyclostationary models, less so with cyclostationary models, and even less so with stationary models. This is due, at least in part, to the loss of the exact orthogonality of the harmonics of 1) a periodic function on a finite interval that is not an integer multiple of the period, and 2) a poly-periodic function on all finite intervals, and also due to the loss of exact statistical independence of random time series on all finite intervals. Consequently, the accuracy of these approximations becomes an important issue. Another-finite-window effect, which applies to all three classes of time series mentioned here is the "edge effects" on a convolution operation. The finite-time statistics like autocorrelation do not exactly obey the elegant input-output relation for convolution. But, again, these effects become negligible for sufficiently long time-segments of data; that is, long relative to the memory length of the convolution. The detailed definitions of the cumulative CDFs and their moments and other probabilistic functions for finite-time segments of data are provided at [2, p. 3.5].

The difference between the terms *statistical* and *probabilistic* are pointed out here for further clarity. Probabilities and probabilistic parameters, such as means, variances, correlations, probability densities, etc., defined in terms of mathematical expectation calculated from mathematic models of stochastic processes, are theoretical or mathematical constructs. They come from within our heads through our imagination or as solutions to mathematical equations. In contrast, averages of empirical measurements, such as estimates of these theoretical quantities, are statistics. They can be obtained from finite ensemble averages derived from repeated experimentation or from finite-time averages performed on a single time series of measurements. This difference is often ignored in the terminology chosen by users of these tools. This can cause the same type of confusion as that resulting from use of theoretical stochastic process models for implementations based on time-averages from single time series. Because stochastic processes are mathematical entities, no actual single signal can ever be considered to be ergodic or non-ergodic. It is a real statistic, not an imaginary probability model. For example, the Statistical Theory of Communication and Information Theory are both primarily probabilistic theories, but they do deal with statistics to some extent. When the focus is on statistics in communications, the traditional name for these theories is appropriate, but many if not most books on this subject focus on probabilities. In contrast, turbulence studies are especially interested in ensembles, for example, all aircraft of a specified design in all operational environments, or even a single aircraft in all operational environments. Here the ensemble in the definition of a stochastic process can be real, not just imagined. Yet, the stochastic process models used in turbulence studies are not real, only the finite ensembles of actual measured turbulence—the statistics—are real. The example set in Middleton's classic book [13], of being consistently clear about this distinction, has not been as diligently followed as would behoove the statistical signal processing community. It is my belief that the all-too-common lack of distinction between probabilities and statistics is a clear reflection of the confusion caused, at least in part, by the abstraction of the stochastic process model that engineers are indoctrinated in.

Despite this little mini-lecture, the strict rule distinguishing between probabilities and statistics is violated in the case of FOT-Probabilities, and this is what makes these probabilities so relevant to practice. Except for the assumption of infinitely long time series, FOT-probabilistic quantities are indeed empirical and are therefore statistics. And, for the FOT-Probabilities defined for finite-segments of data at [2, p. 3.5], they are statistics without any exceptions.

Before closing this discussion, the topic of fixed optimum vs adaptive algorithms for signal processing is briefly revisited. The technology of signal processing has evolved rapidly and exhibited many advances in capability over the last several decades, and education in this technology has stayed at the forefront. However, this cannot be said with as much conviction of education in the theoretical tools used for advancing this field. Our engineering programs may be keeping up to date on adaptive signal processing algorithms, but they are stuck teaching stochastic processes now much as it was done five decades ago—except for a shift from mostly continuous-time signal models to mostly discrete-time signal models—even though the theory of FOT-Probability models that is often more relevant to adaptive signal processing was made available 35 years ago [1], [9].

The entire subject of this article is but one example of a philosophical challenge of great practical import which we face every day in every endeavor: distinguishing between models of reality that our brains create and the real thing—reality itself, which can be quite elusive in some cases. People generally act on the basis of their models of reality for better or for worse. The effectiveness of interpersonal communication, for example, is dictated by the models in terms of which the communicators think. If their models differ too much, they will likely not communicate well. Further discussion of the impact, of the challenge to better match models with reality, on the conduct of science is available at this University of California, Davis website [2, p. 7].

4.1. Conclusions

The traditional generic Kolmogorov model for stochastic processes consists of a generally abstract ensemble of sample paths (realizations) of the process together with a probability measure on the event sets in the sample space. The process is defined by six axioms which, for many applications, cannot all be verified for specific models adopted for use in practice.

The Birkhoff ergodic theorem establishes a condition on the measure in the Kolmogorov model under which probabilistic parameters of the model, such as mean, covariance, probability density functions, etc., can be approximated by time averages on a single sample path from the ensemble. However, the measures for models specified in practice often cannot be explicitly determined and therefore cannot be tested for Birkhoff (strong) ergodicity. Practitioners generally consider the probability measure and the measure property of ergodicity to be mysterious. And they often resort to simply hypothesizing, without verification, that the model they adopt satisfies Kolmogorov's six axioms and Birkhoff's condition on the measure that guarantees ergodicity.

Not only do these practical limitations exist for stationary models, but similar limitations also exist for cyclostationary and poly-cyclostationary models and models that are potentially ergodic or cyclo-ergodic.

To address this disconnect between today's practice in statistical signal processing and traditional theory, new stochastic process models are proposed in this paper. These models are less abstract than the Kolmogorov model and they can, in fact, be derived directly from empirical data consisting of a single time series or from formulaic models for time series. Consequently, ergodicity and cycloergodicity are automatic and conceptually transparent in these new parsimonious models.

The parsimonious models avoid substantive conceptual challenges that often cannot be met in practice and that cause confusion when practitioners attempt to invoke theoretical properties of the standard models in their work on empirical data (see [6] for an in-depth discussion and mathematical treatment).

Although these new models entail, for each stochastic process of interest, a recipe for specifying a sample space and the equivalent of a probability measure which is automatically ergodic or cyclo-ergodic, the recommendation herein is to use these new models for only pedagogical purposes of understanding the relationship between the old (Kolmogorov) and the new models, and otherwise do away with the entire concept of sample spaces and stochastic processes. That is,

the recommendation for operational use is to adopt in place of the new stochastic process models the completely equivalent concept of a single empirical time series of data or a formulaic model of such and the set of cumulative probability distributions of all orders of interest (or, moments, or cumulants, or characteristic functions of all orders of interest), each of which is derived directly from the data or formulaic model using nothing more than time averages. In this formulation, the probability of an event involving the time series is defined to be the fraction of time, over the lifetime of the time series, that the event of interest occurs.

Previous publications have demonstrated in great detail that this concrete alternative approach is operationally equivalent to the abstract stochastic process approach for processes that are stationary or cyclostationary and ergodic or cyclo-ergodic. So, there is no penalty for the conceptual advantages offered by this alternative approach for this class of processes.

Only when non-ergodic models are specifically of interest is there a possible need to use the more abstract traditional stochastic process approach. This includes all nonstationary processes that are not cyclostationary, polycyclostationary, almost cyclostationary, asymptotically mean-stationary, or asymptotically mean-cyclostationary because no such process can be ergodic or cyclo-ergodic.

It is the intent of this paper to assist readers in recognizing the pragmatic benefits of moving toward a paradigm shift in the teaching and practice of statistical signal processing for all applications in which the class of models delineated here are of interest. It follows as a consequence that this paradigm shift also entails separate treatment of the complementary class of models for which stochastic processes are or may be essential: the non-ergodic (and non-cyclo-ergodic) process models.

Acknowledgment

The Author acknowledges the important role that his collaborative work with Professor Antonio Napolitano on writing the paper [6] played in the development of the Author's thoughts that are conveyed in the present paper.

References

1. W.A. Gardner, Statistical Spectral Analysis: A Nonprobabilistic Theory, Prentice-Hall, Englewood Cliffs, 1987.
2. W. A. Gardner, Cyclostationarity Educational Resource, https://www.cyclostationarity.com, 2018 (accessed 1 April 2023).
3. A. Napolitano, Cyclostationary Processes and Time Series: Theory, Applications, and Generalizations, Academic Press, London, 2019. https://doi.org/10.1016/C2017-0-04240-4.
4. J. Leśkow, A. Napolitano, Foundations of the Functional Approach for Signal Analysis, Signal Processing 86 (2006) 3796–3825.
5. A.N. Kolmogorov, Foundations of the Theory of Probability, 1956. Translated to English, 2^{nd} ed. Chelsea Publishing Co., 1933.
6. A. Napolitano, W.A. Gardner, Fraction-of-Time Probability: Advancing Beyond the Need for Stationarity and Ergodicity Assumptions, IEEE Access, April (2022). doi 10.1109/ACCESS.2022.3162620
7. D.G. Brennan, Probability Theory in Communication System Engineering, in: E. J. Baghdady (Ed.), Communication System Theory, McGraw Hill, New York, 1961, Chap. 2.
8. E.M. Hofstetter, Random Processes, in H. Margenau, G.M. Murphy (Eds.), The Mathematics of Physics and Chemistry, Vol. II, D. Van Nostrand Co, Princeton, 1964.
9. W.A. Gardner, W.A. Brown, Fraction-of-Time Probability for Time-Series That Exhibit Cyclostationarity, Signal Processing 92 (1991) 273–292.
10. W. A. Gardner, Introduction to Random Processes with Applications to Signals and Systems, Macmillan, New York, 1985.
11. M. Kac M, H. Steinhaus, Sur Les Foncions Indépendantes IV, Studia Mathematica 7 (1938) 1–15.
12. Y. W. Lee, Statistical Theory of Communication, Wiley, New York, 1960.
13. D. Middleton, An Introduction to Statistical Communication Theory, 1960. Reissued by Wiley-IEEE Press, 1996. Online ISBN: 9780470544112, Print ISBN: 9780780311787.

14. W. A. Gardner, Statistically Inferred Time Warping: Extending the Cyclostationarity Paradigm from Regular to Irregular Statistical Cyclicity in Scientific Data, EURASIP Journal on Advances in Signal Processing 1(59) (2018). https://doi.org/10.1186/s13634-018-0564-6.
15. W. A. Gardner (Ed.) *Cyclostationarity in Communications and Signal Processing*, IEEE Press, Piscataway, 1994.
16. G. D. Birkhoff, Proof of the Ergodic Theorem, Proceedings of the National Academy of Sciences 17(12) (1931) 656–660. https://doi.org/10.1073/pnas.17.2.656.
17. V. Bergelson, A. Leibman, C.G. Moreira, From Discrete- to Continuous-Time Ergodic Theorems, in Ergodic Theory and Dynamical Systems, Cambridge University Press, 32, 2012 383-426. https://www.cambridge.org/core/journals/ergodic-theory-and-dynamical-systems/article/abs/from-sdiscrete-to-continuoustime-ergodic-theorems/E5CCE8CBDB2FA4445BE2FCFC0A08C7C2
18. D. Middleton, Canonical Non-Gaussian Noise Models: Their Implications for Measurement and for Prediction of Receiver Performance." IEEE Transactions on Electromagnetic Compatibility EMC-21(3) (1979) 209-220.
19. C. M. Spooner, W. A. Gardner, The Cumulant Theory of Cyclostationary Time Series. Part II: Development and Application, IEEE Transactions on Signal Processing 42(12) (1994) 3409-3429
20. H. O. A. Wold, On Prediction in Stationary Time Series, Ann. Math. Statist. 19 (1948) 558–567.
21. R. A. Boyles, W. A. Gardner, Cycloergodic Properties of Discrete Parameter Non-Stationary Stochastic Processes, IEEE Transactions on Information Theory IT-29 (1983) 105–114.
22. W. A. Gardner, Stationarizable Random Processes, IEEE Transactions on Information Theory 24(1) (1978) 8-22.
23. R. M. Gray, Probability, Random Processes, and Ergodic Properties, 2nd ed. Springer, Dordrecht, The Netherlands, 2009.
24. D. Dehay, J. Leśkow, A. Napolitano, Central Limit Theorem in the Functional Approach, IEEE Transactions on Signal Processing 61 (16) (2013) 4025–4037. https://doi.org/10.1109/TSP.2013.2266324.
25. A. Napolitano, Time-Warped Almost-Cyclostationary Signals: Characterization and Statistical Function Measurements. 65 (20) (2017) 5526–5541. https://doi.org/10.1109/TSP.2017.2728499.

26. A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, Kogakusha, Tokyo; 1965. 9th edition, ISBN 0-07-119981-0
27. W. A. Gardner, Nonstationary Learning Characteristics of the LMS algorithm, IEEE Transactions on Circuits and Systems CAS-34(10) (1987) 1199-1207.
28. W. A. Gardner, Learning Characteristics of Stochastic-Gradient-Descent Algorithms: A General Study, Analysis, and critique. Signal Processing 6 (1984) 113-133.
29. S. V. Schell, W. A. Gardner, Programmable Canonical Correlation Analysis: A Flexible Framework for Blind Adaptive Spatial Filtering, IEEE Transactions on Signal Processing 43(12) (1995) 2898-2908.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The research also was not supported by an employer.