JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 26, No. 4, July–August 2003

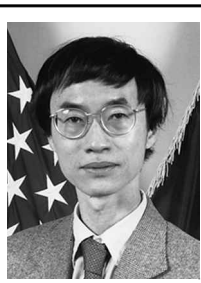
Using Fractional Gaussian Noise Models in Orbit Determination

Winston C. Chow and Paul W. Schumacher Jr.
Naval Network and Space Operations Command, Dahlgren, Virginia 22448-5300

Motivated by a recent requirement to provide accurate covariance matrices as well as orbit estimates for cataloging space objects, a Bayesian estimator to handle autocorrelated process and measurement noises is presented. Although the application is orbit determination, the estimation method itself is general. Techniques are presented to model autocorrelated noises in terms of fractional Gaussian noise and increments of fractional Brownian motion. The Bayesian estimator is derived in both batch and sequential forms, along with explicit formulas for calculating the estimation error covariance matrix. The sequential form generalizes the Kalman filter to the case of autocorrelated noise processes. Also discussed is a batch least-squares estimator formulated with a whitening transformation that accounts for both measurement and process noise. Basic properties of the fractional Gaussian noise model are presented, showing how it contains the special case of white noise in a one-parameter family of self-similar random processes. The parameter characterizes the extent of the autocorrelation and is known as the Hurst parameter. When the measurement noise can be isolated from the process noise by appropriate sensor calibrations, statistical test-of-hypothesis techniques can be used to estimate the Hurst parameter by adjusting it to optimize the p value of the test. For a measurement noise sample that is sufficiently dense in time, the Hurst parameter can be calculated directly from the sample's fractal dimension. In contrast, the process noise usually cannot be isolated from the total noise because the measurements are the only information available. However, assuming the Hurst parameter of the measurement noise is either estimated first or known a priori, test-of-hypothesis techniques can be used to estimate the Hurst parameters of the process noise.

Nomenclature

A_D, A_R	=	true cross-sectional areas for drag and radiation pressure
A_{Da}, A_{Ra}	=	adopted value of cross-sectional areas for drag and radiation pressure
$A(t)$	=	Jacobian matrix of measurements with respect to state
a	=	timescale parameter, used to define self-similarity
$\boldsymbol{a}_D$, $\boldsymbol{a}_G$, $\boldsymbol{a}_R$	=	drag, gravitational, and solar radiation pressure acceleration vectors
$\boldsymbol{a}(k)$	=	intermediate vector variable for Cholesky decomposition
B	=	true ballistic coefficient
B_a	=	adopted value of ballistic coefficient
B_i	=	the ith Borel set in the real line
$B(t)$, $B_H(t)$	=	standard and fractional Brownian motions
$\boldsymbol{b}(k)$	=	intermediate vector variable for Cholesky decomposition
C	=	proportionality constant in the power-law relation for fractal dimension
C_D	=	true drag coefficient (dimensionless)



Winston C. Chow was born in Washington, D.C. He received a B.S. in mathematics from West Virginia Wesleyan College, an M.S. in mathematics from the University of Cincinnati, and a Ph.D. in information technology, specializing in statistical sciences (with emphasis on stochastic processes), from George Mason University. He is a U.S. government employee and has worked for the Navy and Air Force Departments and the U.S. Census Bureau. He is currently a mathematician at the Naval Networks and Space Operations Command, in Dahlgren, Virginia. His interests are in applications of random process theory. He may be contacted at winston.chow@nnsoc.navy.mil.

Paul Schumacher has 15 years experience in space surveillance and space catalog applications, beginning at the Naval Space Surveillance Center and continuing with its successor organizations in Dahlgren, Virginia. Before that, he earned a Ph.D. in aerospace engineering at Virginia Polytechnic Institute and State University with speciality in orbital mechanics. He is currently the Technical Advisor for Operations at Naval Network and Space Operations Command, where most of his work is related to military space control. His research interests are mainly in orbit estimation, multitarget multisensor tracking, and statistical characterization of spacecraft trajectories. He is a Senior Member of AIAA and may be contacted at paul.schumacher@nnsoc.navy.mil.

Received 19 April 2002; revision received 28 February 2003; accepted for publication 6 March 2003. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States. Copies of this paper may be made for personal or internal use, on condition that the copier pay the $10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 0731-5090/03 $10.00 in correspondence with the CCC.

C_{Da} = adopted value of drag coefficient (dimensionless)
c = speed of light; ratio of fractional Gaussian noise scale factor to measurement variance
D = fractal dimension of fractional Gaussian noise; diagonal matrix of reciprocal variances of fractional Gaussian noise
$D(k)$ = Cholesky square-root matrix of the total noise covariance
$d(k)$ = intermediate scalar variable for Cholesky decomposition
$E[\cdot]$ = expected-value operator
F = transformation of random variables
F_a = notional function of satellite body surface parameters (adopted value)
$F(\alpha)$ = notional function of satellite body surface parameters (true value)
$f(t)$ = scalar coefficient of fractional Gaussian noise in the measurements
G = matrix coefficient of fractional Gaussian noise in the state
g_{kjj} = scalar coefficient of jth process noise component at kth time
H = Hurst parameter for fractional Gaussian noise
I = identity matrix
J = Jacobian of transformation among random variables
K = true solar radiation pressure coefficient
K_a = adopted value of solar radiation pressure coefficient
k = discrete-time index for measurement times
$L(x, y)$ = loss function of vectors x and y (for Bayesian estimation)
M = total number of measurements; matrix of orthogonal eigenvectors of measurement noise covariance
$\boldsymbol{M}(k)$ = Jacobian matrix of measurements with respect to initial state
m = total number of measurements or of autocorrelated measurements; true satellite mass
m_a = adopted value of satellite mass
N = number of covering neighborhoods, used to define fractal dimension
P_{XY} = covariance of random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$
$p(\cdot)$ = probability density function
Q_k = covariance matrix of process noise at the kth measurement time
q_{kjj} = jth variance of process noise at the kth measurement time
$q(k)$ = intermediate variable in sequential Bayesian estimation
$\boldsymbol{r}$ = satellite inertial position vector
$r^2(k)$ = measurement noise variance at the kth measurement time
S^2 = diagonal matrix of measurement noise sample variances
$S(k)$ = Cholesky square-root matrix of the measurement covariance matrix
s = time
s^2 = sample variance of measurement noise
T = transformation of random variables
t = time
$\boldsymbol{u}$ = unit vector normal to body surface or toward the sun
V_H = amplitude coefficient in covariance of fractional Brownian motion
V_k = measurement noise at the kth measurement time, as random variable
$\boldsymbol{V}_{\text{rel}}$ = relative velocity vector between satellite and local atmosphere
$\boldsymbol{v}$ = satellite inertial velocity vector
$v(k)$ = intermediate scalar variable for Cholesky decomposition
v_k = measurement noise at the kth measurement time, as realized value
W_k = standard normal random variable for measurement noise at the kth time
w_k = standard normal realization of measurement noise at the kth time
$\boldsymbol{X}$ = linearized state as a random vector
$\boldsymbol{x}$ = linearized state as a realization of a random vector
$\boldsymbol{x}_0$ = initial value of linearized state
$\hat{\boldsymbol{x}}$ = estimate of linearized state
Y_k = standard normal random variable for a measurement at the kth time
y_k = standard normal realization of a measurement at the kth time
$\boldsymbol{Z}$ = linearized measurement vector as a random vector
Z_k = linearized scalar measurement at the kth time, as a random variable
$\boldsymbol{z}$ = linearized measurement vector as a realization of a random vector
z_k = linearized scalar measurement at the kth time, as a realization
α = variable of integration with respect to time
$\boldsymbol{\alpha}$ = notional vector of satellite body surface characteristics
β = state variable for drag effects (dimensionless)
Γ = complete gamma function
γ = state variable for radiation pressure effects (dimensionless)
$\Delta B_H(t)$ = finite increment of fractional Brownian motion
Δt = finite time increment
$\boldsymbol{\delta}(k)$ = total noise vector for measurements in least-squares estimation
ε = length scale, used to define fractal dimension
ζ = dummy time variable
$\boldsymbol{\eta}(t)$ = process noise in the linearized model
λ = variable of integration with respect to time
μ = mean of a random variable
$\nu(t)$ = measurement noise in the linearized model
$\boldsymbol{\xi}(k)$ = vector of covariances of the kth measurement with the state
ρ = true local atmospheric mass density
ρ_a = adopted value of local atmospheric mass density
$\sigma^2(k)$ = variance of the kth measurement
τ = variable of integration with respect to time
τ = dummy time variable
Φ = transition matrix for linearized state
ϕ = true local solar radiation flux
ϕ_a = adopted value of local solar radiation flux
$\boldsymbol{\psi}(k)$ = vector of covariances of the kth measurement with all previous ones
ω = independent variable in the sample space of measurements

Introduction

ESTIMATION methods based on standard least-squares criteria account for measurement noise but not process noise, and neither least squares nor Kalman filtering provides any means to handle time-autocorrelated noise processes. Yet, autocorrelations are commonplace. For example, radar sensors often smooth a large number of raw detection data to report a few refined measurements. The smoothing implies correlations in the set of reported values. Also, models of satellite motion usually omit a variety of persistent physical effects associated with higher-order gravity harmonics, atmospheric drag, or solar radiation pressure, necessarily resulting in time-correlated errors in the satellite ephemeris. Traditionally, noise autocorrelations are neglected so that standard estimation

techniques can be applied. This is the case, for example, with the present-day special-perturbations (SP) catalog of space objects, which uses a weighted batch least-squares technique. However, the SP catalog must supply accurate estimates not only of the state vector itself but also of the estimation error covariance matrix for each cataloged object. This new requirement to report an accurate covariance has led to some reappraisals of the effects of autocorrelated measurement noise and to an approximate procedure for deweighting highly correlated measurements within a single track.[1] A rigorous method of representing measurement correlation has not yet been adopted for this application. Moreover, no representation of process noise has yet been proposed for the SP catalog. Rather, an empirically determined consider-covariance parameter is added to the variance of the estimated ballistic coefficient to compensate for unmodeled atmospheric effects in the prediction of the estimation error covariance matrix. To date, both the track-by-track deweighting and the empirical consider parameter have been implemented without reformulating the estimator to account for correlated noise processes. In this paper, we examine the problem of rigorously accounting for time-correlated measurement and process noises.

We discuss four main aspects of this problem. First, we summarize some essential features of the state model for satellite motion used in updating the SP catalog. We offer simple alternatives to the ballistic coefficient and solar radiation pressure coefficient that may allow more reliable estimation of drag and radiation pressure effects, without increasing the complexity of the state model. The two new variables may also behave more nearly like Gaussian random variables than do the usual coefficients, even in the nonlinear case. In this paper, we do not include noise in the gravitational acceleration because that problem has already been treated elsewhere.[2,3] Moreover, accounting for gravitational noise does not call for the type of empirically based, time-correlated model that we present in this paper. Our formulation does allow additive noise in the gravitational acceleration, either time correlated or not, provided that 1) its autocovariance is known and 2) it is not correlated with the drag or radiation pressure noise.

Second, we discuss a type of time-correlated noise model that can be used with the estimator described later. The estimator calls for Gaussian models, and within this class we know of two specific approaches to representing the noise. The better-known method is the Gauss–Markov approach, in which the correlated noise process is modeled as some postulated deterministic (usually first-order linear) dynamics driven by a white-noise forcing function. Recently, for example, a first-order Gauss–Markov method has been investigated as a way to represent time-correlated random variations of atmospheric density in orbit determination problems (see Ref. 4). The other method, less well known in the astrodynamics community, is to assume that the time integral of the noise is a fractional Brownian motion (FBM) process, so that the noise itself is so-called fractional Gaussian noise (FGN). This noise model has been studied extensively in fields other than orbit determination,[5–9] and is particularly well suited for representing noise correlations that persist over long periods of time. With this noise model, one does not need to postulate extra deterministic dynamics to represent the time correlations. FGN represents the noise as a purely zero-mean stochastic process. Naturally, we would prefer to derive the noise model from the physics of the problem, rather than offer an assumed model a priori. However, that goal is still out of reach for the space-catalog application. In the meantime, we believe that FGN offers at least a conceptual improvement in the handling of some important errors in orbit determination.

Third, we review the derivation of a Bayesian estimator, assuming Gaussian probability density functions for all noise processes. This classical result is perhaps the simplest estimator that can handle time-correlated measurements and model errors. The high computational requirements of Bayesian estimators, such as inversion of large matrices, are a concern for the space-catalog application, and so we mention some possibilities for realizing the estimator. We present batch and sequential versions of the estimator, as well as explicit formulas for calculating the estimation error covariance matrix, the Bayesian counterpart of a Kalman filter covariance. Note that the estimation problem can be formulated as a least-squares problem using a whitening transformation of the data. We show how to account for both measurement noise and process noise in the transformation.

Finally, estimating the FGN model parameters from real observations of satellite motion presents a variety of analytical and statistical problems. For a scalar FGN process, values of two parameters must be assigned. One parameter fixes the amplitude of the noise process, whereas the other, the so-called Hurst parameter, fixes the extent of the autocorrelation. The model includes both white and perfectly autocorrelated noises as special cases, corresponding to the extreme allowable values of the Hurst parameter. We describe theoretical ways to estimate values of these parameters from real data sets and discuss some of the difficulties that will be encountered on any approach. Note also that estimates of the Hurst parameter based on real data automatically provide some indication of the adequacy of the FGN model to describe the real noise, given that valid values of the Hurst parameter must lie between known fixed limits.

Choice of State Variables for Drag and Radiation Pressure Effects

In a realistic system model for orbit determination, it is not usually possible to arrange that either the states or the measurements will have Gaussian distributions. However, it may be possible to choose transformed variables that we could expect to be more nearly Gaussian than the original variables. Here we offer dimensionless state variables that might replace the ballistic coefficient and solar radiation pressure coefficient. These two states are of special concern in orbit determination problems because the physical modeling is not as complete for atmospheric drag and radiation pressure as it is for gravitational forces. This fact occasionally leads to poorly observable or aliased values of these coefficients, including unphysical negative values. The variables we offer do not affect observability; however, they do preclude negative values of either coefficient in a natural way that also happens to be extremely easy to implement if the orbit determination algorithm already estimates the coefficients. Moreover, we can present a simple reason why these new state variables may be more nearly Gaussian than the traditional parameters. This idea was, in fact, our original motivation for devising new state variables, because we hope to model the real noise processes in terms of FGN. However, the new state variables may prove useful for orbit determination even if one neglects process noise entirely.

The acceleration vector of a body due to atmospheric drag is modeled as

$$\boldsymbol{a}_D = -\tfrac{1}{2}\rho(C_D A_D/m)\boldsymbol{V}_{\text{rel}}\|\boldsymbol{V}_{\text{rel}}\| \equiv -\tfrac{1}{2}\rho B\boldsymbol{V}_{\text{rel}}\|\boldsymbol{V}_{\text{rel}}\| \qquad (1)$$

where ρ is the local mass density of the atmosphere, given by a model as a function of position and time, C_D is the dimensionless drag coefficient, A_D is the associated cross-sectional area of the body, m is the mass of the body, and $\boldsymbol{V}_{\text{rel}}$ is the velocity vector of the body relative to the local atmosphere. The drag coefficient, cross-sectional area, and mass of most real space objects are unknown, and the usual expedient is to estimate their combination B, the ballistic coefficient, as part of the dynamic state. However, the best atmospheric models rarely yield density values that are less than 15% uncertain. With disturbed solar conditions, the error in density can exceed 100%. Estimating the atmospheric model parameters themselves as part of the orbit determination problem is usually not feasible and cannot be done even in principle if the satellite body characteristics are unknown. The result is that estimated values for B are always corrupted by errors in the atmospheric density model.

We propose to account for the combined uncertainty in ρ and B by a suitable choice of state variable. We adopt a priori modeled values of these two quantities and denote them as ρ_a and B_a. Here the subscript a signifies that each of the values can vary arbitrarily along the orbital path, although we do not estimate either of them. Rather, we formulate the product of density and ballistic coefficient as

$$\rho B = \rho_a B_a(\rho B/\rho_a B_a) = \rho_a B_a(\rho/\rho_a)(C_D/C_{Da})(A_D/A_{Da})(m_a/m) \qquad (2)$$

Assuming that the true values of both density and all three factors in the ballistic coefficient are random, we recommend, as the new state variable to be estimated, the dimensionless quantity

$$\beta = \ell n(\rho/\rho_a) + \ell n(C_D/C_{Da}) + \ell n(A_D/A_{Da}) + \ell n(m_a/m) \quad (3)$$

This choice is suggested by a procedure often used in regression analysis of power-law relationships, in which taking the logarithm of the relation converts the nonlinear regression problem into a linear one. In any case, the product of density and ballistic coefficient is now formulated in the drag acceleration equation as

$$\rho B = \rho_a B_a \exp\beta \quad (4)$$

When the adopted a priori modeling of density and ballistic coefficient is correct in some mean sense, we expect that the estimated value of β will be small in magnitude. However, β can take on any real value without reversing the sign of the product ρB. The scale factor ($\exp\beta$) measures the adequacy of the adopted values of atmospheric density and ballistic coefficient and could be inserted easily into any orbit determination algorithm that already estimates ballistic coefficient. Moreover, β itself is the sum of several random variables. We appeal to the central limit theorem to conjecture that β will be more nearly Gaussian than any of the random variables ρ, C_D, A_D, or m, or B itself. Although this property is not essential for our method, it may provide extra confidence in the practical utility of our noise model and could be verified during numerical work with real or simulated data. Bryson and Ho[10] outline a simple numerical example that illustrates this conjecture. Notice also that we do not actually have to supply a priori adopted values of C_D, A_D, or m, only of B, and so the estimation problem is no more complicated than before.

The estimation of solar radiation pressure effects can be approached in the same manner. A typical term in a solar radiation pressure acceleration vector model has the form

$$\boldsymbol{a}_R = -(\phi/c)(A_R/m)F(\boldsymbol{\alpha})\boldsymbol{u} \equiv -K\boldsymbol{u} \quad (5)$$

where ϕ is the local solar flux, in units of energy per unit time per unit area, c is the speed of light, A_R is the cross-sectional area of the body, m is the mass of the body, and F is a dimensionless function of a set of body surface parameters symbolized by the vector $\boldsymbol{\alpha}$. The vector $\boldsymbol{u}$ is a unit vector in the direction of either the sun or the body surface normal. Consequently, the solar radiation pressure coefficient K has units of acceleration. Although the fundamental physical modeling difficulties are not as severe with solar radiation pressure as they are with atmospheric drag, estimated values of K are usually small and, therefore, easily aliased by deficiencies elsewhere in the dynamic model.

We propose to formulate the solar radiation pressure coefficient as

$$K = K_a(K/K_a) = K_a \exp\gamma \quad (6)$$

where

$$\gamma = \ell n(\phi/\phi_a) + \ell n(F/F_a) + \ell n(A_R/A_{Ra}) + \ell n(m_a/m) \quad (7)$$

The scale factor ($\exp\gamma$) measures the adequacy of the adopted, deterministic value K_a and is readily inserted into any algorithm that estimates K. The estimated value of γ will be small in magnitude if the adopted value K_a, which can vary arbitrarily along the orbital path, is correct in some mean sense. However, the sign of K does not change, regardless of the value of γ. Moreover, γ is the sum of several random variables and is, therefore, we conjecture, more nearly Gaussian than any of the individual variables or K itself.

If one has done a conscientious job of supplying the a priori values ρ_a, B_a, and K_a, then workable starting values for β and γ in the estimation calculations should be simply $\beta = \gamma = 0$. Although it is not necessary to do so, one could devise a procedure in which the estimates of β and γ are used to revise the a priori modeling of ρ_a, B_a, and K_a such that new estimates of β and γ would be closer to zero, given the same data. This procedure could be repeated indefinitely, if desired, so that β and γ would become, in effect, zero-mean random variables. The fact that ρ_a, B_a, and K_a can vary arbitrarily means that, in most practical cases, it would be reasonable to estimate β and γ as constants (plus noise). However, nothing in this formulation prevents one from estimating arbitrary deterministic variations in β and γ as well. In every case, the only question would be whether the data support this more complicated approach.

System Model

Given the preceding considerations, we take the differential equations of motion to be

$$\dot{\boldsymbol{r}} = \boldsymbol{v}$$
$$\dot{\boldsymbol{v}} = \boldsymbol{a}_G(\boldsymbol{r}, t) + \boldsymbol{a}_D(\boldsymbol{r}, \boldsymbol{v}, t; \rho_a, B_a, \beta) + \boldsymbol{a}_R(\boldsymbol{r}, t; K_a, \gamma) \quad (8)$$

where $\boldsymbol{r}$ is the position vector, $\boldsymbol{v}$ is the velocity vector, and $\boldsymbol{a}_G$ is the gravitational acceleration. Note that $\dot{\beta}$ is fractional Gaussian noise (drag) and that $\dot{\gamma}$ is fractional Gaussian noise (radiation pressure). If process noise were not considered, we would have $\dot{\beta} = \dot{\gamma} = 0$. In the most general case, one might want to allow for noise in the gravitational acceleration. This problem has been analyzed thoroughly by Wright,[2,3] and so we pursue it no further in this discussion. From the point of view of the SP catalog application, the gravitational field of the Earth has been modeled very completely, so that any noise comes essentially from the truncation of known terms in the acceleration. Hence, the calculation of the covariance of additive gravitational process noise for space-catalog orbit determination problems involves no empirical statistics. In contrast, noise in the drag and radiation pressure accelerations arises from unmodeled physical effects, so that we have to resort to empirical statistics.

We linearize the system model to derive a practical estimation procedure, which means only that the working formulas will have to be iterated in a differential correction of the nonlinear estimate. Let the 8×1 vector $\boldsymbol{x}$ represent the first-order variation of the state $(\boldsymbol{r}, \boldsymbol{v}, \beta, \gamma)$ from its nominal value, so that a first-order Taylor's series expansion of the governing equations (8) produces the variational equations of motion in the form

$$\dot{\boldsymbol{x}}(t) = \frac{\partial(\dot{\boldsymbol{r}}, \dot{\boldsymbol{v}}, \dot{\beta}, \dot{\gamma})}{\partial(\boldsymbol{r}, \boldsymbol{v}, \beta, \gamma)}\boldsymbol{x}(t) + G(t)\frac{\mathrm{d}\boldsymbol{B}_{H'}(t)}{\mathrm{d}t} \quad (9)$$

where $\boldsymbol{B}_{H'}(t)$ is an 8×1 vector of fractional Brownian motion processes. The time-derivative notation applied to this function, although not rigorously correct, is a convenient shorthand for fractional Gaussian noise. For our purposes, we set the first six components of $\boldsymbol{B}_{H'}(t)$ equal to zero because we need explicit noise terms only in the variations of β and γ. The 8×8 matrix G contains the scale factors needed to match the FGN amplitude to that of the real noise process. Because of the special form of the governing equations (8), we can set all of the elements of G equal to zero except the main-diagonal elements in rows 7 and 8. Also, the system matrix of the variational equations has the form

$$\frac{\partial(\dot{\boldsymbol{r}}, \dot{\boldsymbol{v}}, \dot{\beta}, \dot{\gamma})}{\partial(\boldsymbol{r}, \boldsymbol{v}, \beta, \gamma)} =$$

$$\begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times1} & 0_{3\times1} \\ \left(\frac{\partial \boldsymbol{a}_G}{\partial \boldsymbol{r}} + \frac{\partial \boldsymbol{a}_D}{\partial \boldsymbol{r}} + \frac{\partial \boldsymbol{a}_R}{\partial \boldsymbol{r}}\right)_{3\times3} & \left(\frac{\partial \boldsymbol{a}_D}{\partial \boldsymbol{v}}\right)_{3\times3} & (\boldsymbol{a}_D)_{3\times1} & (\boldsymbol{a}_R)_{3\times1} \\ 0_{1\times3} & 0_{1\times3} & 0 & 0 \\ 0_{1\times3} & 0_{1\times3} & 0 & 0 \end{bmatrix}$$

Here we have used $\partial \boldsymbol{a}_D/\partial\beta = \boldsymbol{a}_D$ and $\partial \boldsymbol{a}_R/\partial\gamma = \boldsymbol{a}_R$ to emphasize that no extra computations are required for the partial derivatives with respect to the new states. This system matrix is to be evaluated as a function of time along the nominal trajectory. Then, having solved the initial-value problem,

$$\dot{\Phi}(t, t_0) = \frac{\partial(\dot{\boldsymbol{r}}, \dot{\boldsymbol{v}}, \dot{\beta}, \dot{\gamma})}{\partial(\boldsymbol{r}, \boldsymbol{v}, \beta, \gamma)}\Phi(t, t_0), \qquad \Phi(t_0, t_0) = I_{8\times8}$$

we eventually obtain the solution of the variational equations in the general form

$$\boldsymbol{x}(t_{k+1}) = \Phi(t_{k+1}, t_k)\boldsymbol{x}(t_k) + \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau)G(\tau)\,\mathrm{d}\boldsymbol{B}_H(\tau)$$

$$\equiv \Phi(t_{k+1}, t_k)\boldsymbol{x}(t_k) + \boldsymbol{\eta}(t_k) \tag{10}$$

which maps the state variations between successive measurement times via the transition matrix Φ. To represent measurements in the linearized model, we assume that measured values of any data types can be represented as nominal values plus first-order variations. The nominal values of measurements will be predicted as functions of the nominal state at the measurement times, using the nonlinear model. Let z represent the first-order variation of a measurement. Then we write

$$z(t_{k+1}) = A(t_{k+1})\boldsymbol{x}(t_{k+1}) + v(t_{k+1})$$

$$\equiv A(t_{k+1})\boldsymbol{x}(t_{k+1}) + f(t_{k+1})\frac{\mathrm{d}B_H(t)}{\mathrm{d}t}\bigg|_{t_{k+1}} \tag{11}$$

where $B_H(t)$ is a fractional Brownian motion process used to represent the measurement noise v. The scale factor f matches the FGN amplitude to that of the real noise. The matrix A is the appropriate partial-derivative matrix evaluated on the nominal trajectory. Then, condensing the notation in an obvious way, we summarize the linear system model as

$$\boldsymbol{x}(k+1) = \Phi(k+1, k)\boldsymbol{x}(k) + \boldsymbol{\eta}(k)$$

$$z(k+1) = A(k+1)\boldsymbol{x}(k+1) + v(k+1) \tag{12}$$

The problem is to estimate the state $\boldsymbol{x}(m)$ at the most recent measurement time, based on measurements $z(i)$ taken at m discrete times t_i, accounting for autocorrelations in each of the random sequences $\boldsymbol{\eta}(i)$ and $v(i)$. We assume that there are no correlations between $\boldsymbol{\eta}$ and v. Moreover, we assume that both of the random sequences are samples from continuous random processes, $\boldsymbol{\eta}(t)$ and $v(t)$, which can be described in terms of FBM and FGN. For this reason, we digress to summarize the properties of FBM and FGN.

FBM and FGN

FBM, denoted as $B_H(t)$, is a generalization of classical Brownian motion and has properties that are analogous to those of that better-known random process. Like Brownian motion, $B_H(t)$ is almost surely not differentiable and yet approximates many real physical processes.[5] FBM can be simulated by a simple numerical procedure called random midpoint displacement.[5,11] It is sometimes convenient to consider $B_H(t)$ as the time integral of an idealized random process FGN, just as ordinary Brownian motion can be considered the time integral of the idealized white noise. Like white noise, FGN is a stationary process. Unlike white noise, however, FGN is, in general, autocorrelated. If it is necessary to exhibit FGN numerically, we can resort to a finite difference approximation of the (almost surely nonexistent) derivative:

$$\frac{\mathrm{d}B_H}{\mathrm{d}t} \approx \frac{\Delta B_H}{\Delta t}$$

FBM is defined as a transformation of ordinary Brownian motion[7,12]:

$$B_H(t) = \frac{1}{\Gamma\left(H + \frac{1}{2}\right)}\left[\int_{-\infty}^{0} \left(|t-s|^{H-\frac{1}{2}} - |s|^{H-\frac{1}{2}}\right)\mathrm{d}B(s) + \int_0^t |t-s|^{H-\frac{1}{2}}\,\mathrm{d}B(s)\right] \tag{13}$$

$B(\cdot)$ is Brownian motion and $\Gamma(\cdot)$ is the gamma function. For scalar FBM, the dimensionless parameter H, known as the Hurst parameter, is restricted to the interval $0 < H < 1$. The reverse transformation is also available as a theorem[7]:

$$B(t) = \frac{1}{\Gamma\left(\frac{3}{2} - H\right)}\left[\int_{-\infty}^{0} \left(|t-s|^{\frac{1}{2}-H} - |s|^{\frac{1}{2}-H}\right)\mathrm{d}\,B_H(s) + \int_0^t |t-s|^{\frac{1}{2}-H}\,\mathrm{d}\,B_H(s)\right] \tag{14}$$

Notice that the special case $H = \frac{1}{2}$ implies that $B_H(\cdot) \equiv B(\cdot)$, meaning also that FGN reduces to the special case of white noise.

The covariance function of $B_H(t)$ is given by

$$E[B_H(t)B_H(s)] = (V_H/2)(|t|^{2H} + |s|^{2H} - |t-s|^{2H}) \tag{15}$$

where the factor V_H is defined conventionally as

$$V_H = \mathrm{var}[B_H(1)] = E\left[B_H^2(1)\right] = \frac{-\Gamma(2-2H)\cos(\pi H)}{\pi H(2H-1)} \tag{16}$$

The covariance function shows that $B_H(t)$ is not stationary; rather, it depends on the times t and s, as well as on the time difference $t - s$. However, increments of $B_H(t)$ are stationary, besides having zero mean. In particular, the process

$$\frac{\Delta B_H(t)}{\Delta t} \equiv \frac{B_H(t+\Delta t) - B_H(t)}{\Delta t} \tag{17}$$

for any fixed $\Delta t > 0$, is a zero-mean Gaussian process whose covariance function depends only on the time difference $t - s$,

$$E\left[\frac{\Delta B_H(t)}{\Delta t}\frac{\Delta B_H(s)}{\Delta t}\right] = \frac{V_H \Delta t^{2H-2}}{2}\left[\left(\frac{|t-s|}{\Delta t} + 1\right)^{2H} - 2\left(\frac{|t-s|}{\Delta t}\right)^{2H} + \left|\frac{|t-s|}{\Delta t} - 1\right|^{2H}\right] \tag{18}$$

The limit $\Delta t \to 0$ exists for this function:

$$\lim_{\Delta t \to 0} E\left[\frac{\Delta B_H(t)}{\Delta t}\frac{\Delta B_H(s)}{\Delta t}\right] = V_H H(2H-1)|t-s|^{2H-2} \tag{19}$$

We take this result as the covariance function of fractional Gaussian noise. However, the variance of this random process, obtained by setting $t = s$, reduces to

$$E\left\{[\Delta B_H(t)/\Delta t]^2\right\} = V_H \Delta t^{2H-2} \tag{20}$$

As $\Delta t \to 0$, the variance diverges for every value of H in the interval $0 < H < 1$, as might be expected by analogy with white noise.

Note that nonoverlapping increments of $B_H(t)$ are, in general, correlated, so that FGN is, in general, autocorrelated. In the special case $H = \frac{1}{2}$, we recover classical Brownian motion, which happens to have zero autocorrelation. In case $H > \frac{1}{2}$, nonoverlapping increments of $B_H(t)$ have positive autocorrelation. Qualitatively, this means that if an increment of $B_H(t)$ is positive, then subsequent increments tend to be positive. Similarly, if an increment of $B_H(t)$ is negative, then subsequent increments tend to be negative. A graph of positively correlated $B_H(t)$ would appear smoother than a graph of Brownian motion on the same scales of time and amplitude. In case $H < \frac{1}{2}$, nonoverlapping increments of $B_H(t)$ have negative autocorrelation. Qualitatively, this means that if an increment of $B_H(t)$ is positive, then subsequent increments tend to be negative, and if an increment of $B_H(t)$ is negative, then subsequent increments tend to be positive. A graph of negatively correlated $B_H(t)$ would appear rougher than a graph of Brownian motion on the same scales of time and amplitude.[11] Positive autocorrelation of $B_H(t)$, with its associated smoothing effect, allows FBM to mimic the behavior of many real physical processes, especially those whose correlations can persist over long periods of time.[5,12] On the other hand, the violent oscillatory behavior of negatively autocorrelated FBM seems not to mimic the behavior of real physical processes, at least the

processes we are concerned with in orbit determination. This is the main reason why we will assume that the Hurst parameter must be restricted as $H \geq \frac{1}{2}$ in practice. Eventually, we will want to assign a value of H based on empirical statistics for some real physical process. If the empirically derived value of H does not lie in the interval $\frac{1}{2} \leq H < 1$, we will assume that scalar FBM and FGN are not adequate to represent the real-noise process and that some other model must be used.

The form of the covariance function implies that FBM is statistically self-similar. Informally, self-similarity means that a sample path of FBM has the same general appearance on any timescale. The statistical properties of the sample path do not change as the path is examined at different magnifications, provided only that a certain relation is maintained between the timescale and the amplitude scale. To prove this, rescale time as $t \to at$ and $s \to as$ and substitute in the covariance function. It is easy to verify in this manner that the functions $B_H(t)$ and $|a|^{-H} B_H(at)$ do have the same covariance. Because they are also both zero mean and Gaussian, both functions have the same probability density function (PDF). In particular, they are both FBMs on the timescale t. Although the random process $B_H(at)$ is indeed FBM on the timescale at, one would not automatically expect it to be FBM on the timescale t as well, merely through an adjustment of the amplitude. This special property is what is meant by self-similarity. By definition, the random process $B_H(t)$ is self-similar under the transformation $B_H(t) \to |a|^{-H} B_H(at)$ if and only if the two functions are equal in distribution: $B_H(t) \overset{d}{=} |a|^{-H} B_H(at)$. Consequently, the Hurst parameter H characterizes the self-similarity.

Because FBM is statistically self-similar, the sample paths of this process are fractal curves. The fractal dimension, which measures the roughness of a fractal curve, is almost surely the same for all sample paths of a given FBM process and characterizes the random process just as the Hurst parameter does. In fact, there is a simple relation between the two parameters. For a scalar FBM, it can be shown that, almost surely,

$$D + H = 2 \tag{21}$$

where D is the fractal dimension.[13] For a vector FBM, it can be shown that the sum of the fractal dimension and the Hurst parameter must equal the Euclidean embedding dimension of the vector plus one (for time).[14] This means that we can derive an empirical value of the Hurst parameter if we can estimate the fractal dimension of merely one sample path of the FBM process. Note that a value of $H = 1$ implies $D = 1$, which is the fractal dimension of a smooth curve in the plane. We take this as the "perfect-autocorrelation" limit, complementing the "zero-autocorrelation" limit given by $H = \frac{1}{2}$ and $D = \frac{3}{2}$.

The term fractal dimension has a number of different definitions, not all of them equivalent. However, several of them are based on the idea that an inverse power-law relation exists, at least approximately or in a certain limit, between the number N of equal-sized closed spherical balls needed to cover the fractal and the size ε of the balls themselves, namely,

$$N\varepsilon^D \approx C \tag{22}$$

where C and D are positive constants.[15] The motivation for calling D the fractal dimension is the analogy with the special cases of smooth curves ($D = 1$), smooth surfaces ($D = 2$), solid figures ($D = 3$), and so on. Fractals are distinguished by having nonintegral values of the exponent. Notice that the power-law relation (22) is self-similar in the sense that the same values of N and exponent D characterize the relation if ε is rescaled; merely the constant C has to be adjusted. Solving this relation for the exponent, we get

$$D \approx \frac{\ell n C - \ell n N}{\ell n \varepsilon} = \frac{\ell n N - \ell n C}{\ell n (1/\varepsilon)} \tag{23}$$

If ε is chosen to be small, in an effort to derive an accurate value for D, the value of N becomes large compared to any constant C. Consequently, one definition of dimension often used in theoretical analysis of fractals is

$$D = \lim_{\varepsilon \to 0} \frac{\ell n N(\varepsilon)}{\ell n (1/\varepsilon)} \tag{24}$$

where $N(\varepsilon)$ is the minimum number of balls of diameter ε needed to cover the fractal. For fractal curves in the plane, the balls can be replaced with square grids that are more convenient computationally.[15] Then $N(\varepsilon)$ is taken to be the minimum number of square grids, each with side length ε, that are needed to cover the fractal, and D is called the box-count dimension. It can be shown that, for scalar FBM in particular, the box-count dimension is equivalent to the ε-ball dimension, that it is related to the Hurst parameter by $D + H = 2$, and that all smooth curves in the plane have box-count dimension equal to unity.[11,13] In practice, the limit of vanishing ε may not have to be approached closely. If N and ε are, in fact, related by a power law, then the locus of $\ell n N$ vs $\ell n (1/\varepsilon)$ will be a straight line with slope equal to D. Hence, the value of D can be estimated by first computing $N(\varepsilon)$ for several small but finite values of ε and then computing the slope of a straight line fitted to the data in the $\ell n N - \ell n (1/\varepsilon)$ plane. If the true relation between N and ε were not a power law, the data would not be well approximated by a straight line. However, this would imply that a single well-defined fractal dimension does not exist, so that the sample path is not a fractal curve and, consequently, that the random process that generated the data is not self-similar. Therefore, the goodness of fit in the linear regression of the data is a measure of how adequately the idealized FBM model can represent the real noise process.

Estimating the System State and Covariance

A Bayesian estimator involves no assumptions that the noise processes for either the state or the measurements are uncorrelated. We review briefly the derivation of a Bayesian estimator for the case of Gaussian autocorrelated noise processes to show how the FGN noise model would be applied. In this section we assume that values of the FGN Hurst parameters and scale factors are already known. Later we consider how to derive suitable values of these parameters from real measurements. We first present the batch form of the estimator, which is also described by Meditch.[16] Then we show how to put the estimator into sequential form. In the case of zero autocorrelation in the noises, the sequential Bayesian estimator is the Kalman filter and provides a substantial savings of computation compared to batch processing. In contrast, in the case of autocorrelated noises, the sequential Bayesian estimator involves no net savings of computation compared to batch processing, when we are estimating the state at a single time. However, when we need state estimates at a sequence of times, the sequential form of the estimator becomes relatively efficient because it distributes the processing load over the sequence of updates.

Let $\hat{\boldsymbol{x}}(\boldsymbol{z})$ be an estimator of a random n-vector $\boldsymbol{X}$ and defined as a function of a random M-vector $\boldsymbol{Z}$. In our application, the components of $\boldsymbol{Z}$ correspond to scalar measurements at each of m discrete times, and we allow for the possibility of simultaneous measurements at various times by requiring only that $m \leq M$. Let $\boldsymbol{z}$ denote the vector of observations of the components of $\boldsymbol{Z}$ and let $\boldsymbol{x}$ denote realizations of $\boldsymbol{X}$. The notation $L(\boldsymbol{x}, \hat{\boldsymbol{x}}(\boldsymbol{z}))$ denotes an admissible scalar loss function of $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$. Then $\hat{\boldsymbol{x}}$ is called a Bayesian estimator on L if it minimizes the expected value of L conditioned on $\boldsymbol{z}$,

$$E\{L[\boldsymbol{X}, \hat{\boldsymbol{x}}(\boldsymbol{z})] | \boldsymbol{z}\} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L[\boldsymbol{x}, \hat{\boldsymbol{x}}(\boldsymbol{z})] p(\boldsymbol{x}|\boldsymbol{z}) \, \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{25}$$

where the conditional PDF in the integrand is given by the Bayes formula

$$p(\boldsymbol{x}|\boldsymbol{z}) = \frac{p(\boldsymbol{z}|\boldsymbol{x}) p(\boldsymbol{x})}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\boldsymbol{z}|\boldsymbol{x}) p(\boldsymbol{x}) \, \mathrm{d}x_1 \cdots dx_n}$$

$$\boldsymbol{x}^T = [x_1, \ldots, x_n] \tag{26}$$

For a wide class of loss functions, including the quadratic $L[\boldsymbol{x}, \hat{\boldsymbol{x}}(z)] = [\boldsymbol{x} - \hat{\boldsymbol{x}}(z)]^T [\boldsymbol{x} - \hat{\boldsymbol{x}}(z)]$, it can be shown that $\hat{\boldsymbol{x}}$ is just the conditional mean of $\boldsymbol{X}$:

$$\hat{\boldsymbol{x}}(z) = E[\boldsymbol{X}|z] \tag{27}$$

The explicit form of the estimator can be found if $p(\boldsymbol{x}|z)$ is Gaussian. A well-known theorem[16] asserts that if $\boldsymbol{X}$ and $\boldsymbol{Z}$ are Gaussian random vectors, then the conditional PDF $p(\boldsymbol{x}|z)$ is Gaussian with mean given by

$$E(\boldsymbol{X}|z) = E(\boldsymbol{X}) + P_{XZ} P_{ZZ}^{-1} [z - E(\boldsymbol{Z})] \tag{28}$$

and covariance given by

$$\operatorname{cov}(\boldsymbol{X}|z) = P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \tag{29}$$

where the notation P_{UV} means the covariance matrix of random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$. This conditional mean is the Bayesian estimator of $\boldsymbol{X}$ if all noises are Gaussian.

To prove the theorem, observe that the numerator of the Bayes formula is the joint PDF of $\boldsymbol{Z}$ and $\boldsymbol{X}$, $p(z|\boldsymbol{x}) p(\boldsymbol{x}) = p(z, \boldsymbol{x})$, which is Gaussian if $\boldsymbol{Z}$ and $\boldsymbol{X}$ are Gaussian. The denominator reduces to $p(z)$, which is also Gaussian. Moreover, $p(x|z)$ is Gaussian for fixed z. To within a normalizing factor, we have

$$p(\boldsymbol{x}|z) \propto \exp\left\{ -\frac{1}{2} \begin{bmatrix} z - E(\boldsymbol{Z}) \\ \boldsymbol{x} - E(\boldsymbol{X}) \end{bmatrix}^T \begin{bmatrix} P_{ZZ} & P_{ZX} \\ P_{XZ} & P_{XX} \end{bmatrix}^{-1} \begin{bmatrix} z - E(\boldsymbol{Z}) \\ \boldsymbol{x} - E(\boldsymbol{X}) \end{bmatrix} + \frac{1}{2} [z - E(\boldsymbol{Z})]^T P_{ZZ}^{-1} [z - E(\boldsymbol{Z})] \right\} \tag{30}$$

A well-known matrix-inversion identity can be derived by partitioning a system of linear equations once and solving the system formally. The result can be expressed in several different algebraic forms, all of them equivalent to

$$M_{n \times n}^{-1} \equiv \begin{bmatrix} A_{p \times p} & B_{p \times q} \\ C_{q \times p} & D_{q \times q} \end{bmatrix}^{-1} \equiv \begin{bmatrix} (A^{-1} + A^{-1} B F C A^{-1})_{p \times p} & (-A^{-1} B F)_{p \times q} \\ (-F C A^{-1})_{q \times p} & F_{q \times q} \end{bmatrix} \tag{31}$$

where $F = (D - C A^{-1} B)^{-1}$. Using this matrix inversion identity, we can write

$$\begin{bmatrix} P_{ZZ} & P_{ZX} \\ P_{XZ} & P_{XX} \end{bmatrix}^{-1} \equiv \begin{bmatrix} P_{ZZ}^{-1} + P_{ZZ}^{-1} P_{ZX} \left(P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \right)^{-1} P_{XZ} P_{ZZ}^{-1} & -P_{ZZ}^{-1} P_{ZX} \left(P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \right)^{-1} \\ -\left(P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \right)^{-1} P_{XZ} P_{ZZ}^{-1} & \left(P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \right)^{-1} \end{bmatrix} \tag{32}$$

Substituting this expression into $p(\boldsymbol{x}|z)$ and collecting terms, we find that

$$p(\boldsymbol{x}|z) \propto \exp\left[-\tfrac{1}{2} \left(\boldsymbol{x} - \left\{ E(\boldsymbol{X}) + P_{XZ} P_{ZZ}^{-1} [z - E(\boldsymbol{Z})] \right\} \right)^T \left(P_{XX} - P_{XZ} P_{ZZ}^{-1} P_{ZX} \right)^{-1} \left(\boldsymbol{x} - \left\{ E(\boldsymbol{X}) + P_{XZ} P_{ZZ}^{-1} [z - E(\boldsymbol{Z})] \right\} \right) \right] \tag{33}$$

By inspection, we obtain the conditional mean and covariance required by the theorem. Note also that the covariance of the state conditioned on the data is really the estimator's error covariance:

$$\operatorname{cov}(\boldsymbol{X}|z) = \operatorname{cov}\{[\boldsymbol{X} - E(\boldsymbol{X}|z)]|z\} = \operatorname{cov}[\boldsymbol{X} - E(\boldsymbol{X}|z)] = \operatorname{cov}(\boldsymbol{X} - \hat{\boldsymbol{x}}) \tag{34}$$

This result follows because, for any Gaussian random vectors $\boldsymbol{X}$ and $\boldsymbol{Z}$, the vector $\boldsymbol{X} - E(\boldsymbol{X}|z)$ is not correlated with $\boldsymbol{Z}$ (Ref. 16).

To complete the formulation of the batch estimator, it remains to show how to compute $E(\boldsymbol{X})$, $E(\boldsymbol{Z})$, and the partitions of the P matrix from the linearized state and measurement models. We have immediately that $E[\boldsymbol{X}(k)] = \Phi(k, 0) \boldsymbol{x}_0$, where $\boldsymbol{x}_0 = E[\boldsymbol{X}(0)]$ is the initial state, which we assume to be given. Also we have $E[\boldsymbol{Z}(k)] = A(k) \Phi(k, 0) \boldsymbol{x}_0$, where the notation $\boldsymbol{Z}(k)$ means the scalar random variable associated with the measurement at time t_k.

The covariance matrices P_{ZZ}, P_{XX}, and $P_{XZ} = P_{ZX}^T$ are derived from autocovariances of $\boldsymbol{\eta}(k)$ and $v(k)$ for $k = 1, \ldots, m$. For the measurement noise covariance, we have

$$\operatorname{cov}[v(k), v(\ell)] = f(t_k) f(t_\ell) E\left(\left. \frac{\mathrm{d}B_H}{\mathrm{d}t} \right|_{t_k} \left. \frac{\mathrm{d}B_H}{\mathrm{d}t} \right|_{t_\ell} \right) \tag{35}$$

We substitute the covariance of FGN [Eq. (19)] based on the approximation $\mathrm{d}B_H / \mathrm{d}t \approx \Delta B_H / \Delta t$:

$$\operatorname{cov}[v(k), v(\ell)] = f(t_k) f(t_\ell) V_H H(2H - 1) |t_k - t_\ell|^{2H-2} \tag{36}$$

For the process noise covariance, we have

$$\begin{aligned} \operatorname{cov}[\boldsymbol{\eta}(k), \boldsymbol{\eta}(\ell)] &= E\left\{ \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau) G_k \, \mathrm{d}\boldsymbol{B}_{H'}(\tau) \right. \\ &\quad \left. \times \left[\int_{t_\ell}^{t_{\ell+1}} \Phi(t_{\ell+1}, \alpha) G_\ell \, \mathrm{d}\boldsymbol{B}_{H'}(\alpha) \right]^T \right\} \\ &= \int_{t_k}^{t_{k+1}} \int_{t_\ell}^{t_{\ell+1}} \Phi(t_{k+1}, \tau) G_k E\left[\mathrm{d}\boldsymbol{B}_{H'}(\tau) \, \mathrm{d}\boldsymbol{B}_H'^T(\alpha) \right] \\ &\quad \times G_\ell^T \Phi^T(t_{\ell+1}, \alpha) \end{aligned} \tag{37}$$

where the matrix $\mathrm{d}\boldsymbol{B}_{H'}(\tau) \, \mathrm{d}\boldsymbol{B}_{H'}^T(\alpha)$ has all zeros except at the (7, 7) and (8, 8) positions. Each component of the expectation matrix in the integral is of the form

$$E\left[\mathrm{d}\boldsymbol{B}_{H_j'}(\tau) \, \mathrm{d}\boldsymbol{B}_{H_j'}(\alpha) \right] = V_{H_j'} H_j' (2H_j' - 1) |\tau - \alpha|^{2H_j' - 2} \, \mathrm{d}\tau \, \mathrm{d}\alpha \quad \text{for} \quad j = 7, 8 \tag{38}$$

with $E[\mathrm{d}B_{H_7'}(\tau) \, \mathrm{d}B_{H_8'}(\alpha)] = 0 \quad \forall \tau, \alpha$. Also, we will assume that the matrices G_k and G_ℓ are constants over the single time step and that they may take on different values on each time step. The matrix $\Phi(\tau, \varsigma)$ is a function of ς, given τ. Hence, the components of the matrix-product integral must be a linear combination of the scalar terms

$$c \int_{t_k}^{t_{k+1}} \int_{t_\ell}^{t_{\ell+1}} h(\alpha, \tau) |\tau - \alpha|^{2H_j' - 2} \, \mathrm{d}\tau \, \mathrm{d}\alpha$$

for $j = 7, 8$, for some constant c taken from G_k and G_ℓ, and some function h taken from the transition matrix Φ. The intervals $[t_k, t_{k+1})$ and $[t_\ell, t_{\ell+1})$ are either disjointed or coinciding. Without loss of generality, let $[t_\ell, t_{\ell+1})$ be either to the right of $[t_k, t_{k+1})$ or coinciding with it. Because $E(\mathrm{d}B_{H_7'}(\tau) \, \mathrm{d}B_{H_8'}(\alpha)] = 0$ and only the components of the lower right-hand 2×2 submatrix of $E[\mathrm{d}\boldsymbol{B}_{H'}(\tau) \, \mathrm{d}\boldsymbol{B}_{H'}^T(\alpha)]$ are nonzero, and moreover, the lower right-hand 2×2 submatrix of Φ is the identity matrix, then the (7, 7) and (8, 8) components of the integral become, for $j = 7, 8$, respectively,

$$V_{H'_j} g_{kjj} g_{\ell jj} H'_j (2H'_j - 1) \int_{t_k}^{t_{k+1}} \int_{t_\ell}^{t_{\ell+1}} |\tau - \alpha|^{2H'_j - 2} \, d\tau \, d\alpha$$

$$= \begin{cases} \frac{1}{2} V_{H'_j} g_{kjj} g_{\ell jj} \left[(t_{k+1} - t_\ell)^{2H'_j} + (t_k - t_{\ell+1})^{2H'_j} - (t_{k+1} - t_{\ell+1})^{2H'_j} - (t_k - t_\ell)^{2H'_j} \right] & \text{if intervals are disjoint} \\ V_{H'_j} g_{kjj}^2 (t_{k+1} - t_k)^{2H'_j} & \text{if intervals coincide} \end{cases} \quad (39)$$

Successive substitution of the equation for $\boldsymbol{x}(k)$ starting at $\boldsymbol{x}(1)$ gives

$$\boldsymbol{x}(k) = \Phi(k, 0)\boldsymbol{x}(0) + \sum_{j=0}^{k-1} \Phi(k, j+1)\boldsymbol{\eta}(j)$$

showing that $\boldsymbol{x}(k)$ is a linear combination of all process noises up to t_k. Also, $z(k)$, the scalar measurement at time t_k, given by $z(k) = A(k)\boldsymbol{x}(k) + v(k)$, includes a linear combination of the measurement noise at t_k and the process noises up to t_k. If we define

$$\boldsymbol{\eta}(-1) = \int_{t_{-1}}^{t_0} \Phi(t_0, \tau) G_{-1} \, d\boldsymbol{B}'_H(\tau)$$

analogously with the process noise for all other points in time, then we can write $x(0) = x_0 + \eta(-1)$, so that

$$\boldsymbol{x}(k) = \Phi(k, 0)\boldsymbol{x}_0 + \sum_{j=-1}^{k-1} \Phi(k, j+1)\boldsymbol{\eta}(j)$$

and, consequently

$$\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{X}(\ell)]$$

$$= \sum_{j=-1}^{k-1} \sum_{i=-1}^{\ell-1} \Phi(k, j+1) E[\boldsymbol{\eta}(j)\boldsymbol{\eta}^T(i)] \Phi^T(k, i+1) \quad (40)$$

In this expression, the covariance at initial time, $\operatorname{cov}[X(0)] = E[\boldsymbol{\eta}(-1)\boldsymbol{\eta}^T(-1)] = P_0$, is assumed to be given. Then we have also

$$\operatorname{cov}[Z(k), Z(\ell)] = \operatorname{cov}\{[A(k)\boldsymbol{X}(k) + v(k)], [A(\ell)\boldsymbol{X}(\ell) + v(\ell)]\}$$

$$= A(k)\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{X}(\ell)]A^T(\ell) + \operatorname{cov}[v(k), v(\ell)] \quad (41)$$

$$\operatorname{cov}[Z(k), \boldsymbol{X}(\ell)] = \operatorname{cov}\{[A(k)\boldsymbol{X}(k) + v(k)], \boldsymbol{X}(\ell)\}$$

$$= A(k)\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{X}(\ell)] + \operatorname{cov}[v(k), \boldsymbol{X}(\ell)]$$

$$= A(k)\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{X}(\ell)] \quad (42)$$

Components of P_{XX}, P_{XZ}, and P_{ZZ} can also be computed recursively. Besides the preceding formulas, we have

$$\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{X}(\ell)] = \operatorname{cov}\{[\Phi(k, k-1)\boldsymbol{x}(k-1)$$

$$+ \boldsymbol{\eta}(k-1)][\Phi(\ell, \ell-1)\boldsymbol{X}(\ell-1) + \boldsymbol{\eta}(\ell-1)]\}$$

$$= \Phi(k, k-1)\operatorname{cov}[\boldsymbol{X}(k-1), \boldsymbol{X}(\ell-1)]\Phi^T(\ell, \ell-1)$$

$$+ \operatorname{cov}[\boldsymbol{\eta}(k-1), \boldsymbol{X}(\ell-1)]\Phi(\ell, \ell-1)$$

$$+ \Phi(k, k-1)\operatorname{cov}[\boldsymbol{X}(k-1), \boldsymbol{\eta}(\ell-1)]$$

$$+ \operatorname{cov}[\boldsymbol{\eta}(k-1), \boldsymbol{\eta}(\ell-1)] \quad (43)$$

$$\operatorname{cov}[\boldsymbol{X}(k), \boldsymbol{\eta}(\ell)] = \operatorname{cov}[\Phi(k, k-1)\boldsymbol{X}(k-1) + \boldsymbol{\eta}(k-1), \boldsymbol{\eta}(\ell)]$$

$$= \Phi(k, k-1)\operatorname{cov}[\boldsymbol{X}(k-1), \boldsymbol{\eta}(\ell)] + \operatorname{cov}[\boldsymbol{\eta}(k-1), \boldsymbol{\eta}(\ell)] \quad (44)$$

Starting values for these recursions are

$$\operatorname{cov}[\boldsymbol{X}(0), \boldsymbol{\eta}(j)] = \operatorname{cov}[\boldsymbol{\eta}(-1), \boldsymbol{\eta}(j)]$$

$$\operatorname{cov}[\boldsymbol{X}(0), \boldsymbol{X}(j)] = E\left[\boldsymbol{\eta}(-1) \sum_{i=-1}^{j-1} \boldsymbol{\eta}^T(i) \Phi^T(k, i+1) \right] \quad (45)$$

Sequential Bayesian Estimator

The matrix inverse P_{ZZ}^{-1} typically requires an extensive computation. However, if the estimate based on k measurements is available, the estimate based on $k+1$ measurements can be obtained without recomputing the large inverse. The prior estimate has the form

$$\hat{\boldsymbol{x}}(k) = E[\boldsymbol{X}(k)] + P_{XZ}(k) P_{ZZ}^{-1}(k)\{\boldsymbol{z}(k) - E[\boldsymbol{Z}(k)]\} \quad (46)$$

where $\boldsymbol{z}(k)$ is the vector of all measurement values up to time t_k. The next estimate is

$$\hat{\boldsymbol{x}}(k+1) = E[\boldsymbol{X}(k+1)] + P_{XZ}(k+1) P_{ZZ}^{-1}(k+1)\{\boldsymbol{z}(k+1)$$

$$- E[\boldsymbol{Z}(k+1)]\} \quad (47)$$

which we can write in partitioned form as

$$\hat{\boldsymbol{x}}(k+1) = E[\boldsymbol{X}(k+1)] + [P_{XZ}(k) \quad \boldsymbol{\xi}(k+1)]$$

$$\times \begin{bmatrix} P_{ZZ}(k) & \boldsymbol{\psi}(k+1) \\ \boldsymbol{\psi}^T(k+1) & \sigma^2(k+1) \end{bmatrix}^{-1} \begin{Bmatrix} \boldsymbol{z}(k) - E[\boldsymbol{Z}(k)] \\ z(k+1) - E[Z(k+1)] \end{Bmatrix} \quad (48)$$

Here z is the new scalar measurement and σ^2 is the variance of z. The matrix inversion identity (31) allows us to write

$$P_{ZZ}^{-1}(k+1) = \begin{bmatrix} P_{ZZ}^{-1}(k) + q(k+1) P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1)\boldsymbol{\psi}^T(k+1) P_{ZZ}^{-1}(k) & -q(k+1) P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1) \\ -q(k+1)\boldsymbol{\psi}^T(k+1) P_{ZZ}^{-1}(k) & q(k+1) \end{bmatrix} \quad (49)$$

where

$$q(k+1) = 1 \Big/ \left[\sigma^2(k+1) - \boldsymbol{\psi}^T(k+1) P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1) \right]$$

When this formula is substituted for the matrix inverse and terms are collected, it is a straightforward job to obtain the sequential state update formula,

$$\hat{\boldsymbol{x}}(k+1) - E[\boldsymbol{X}(k+1)] = \hat{\boldsymbol{x}}(k) - E[\boldsymbol{X}(k)]$$

$$+ q(k+1)\left\{ \boldsymbol{\xi}(k+1) - P_{XZ}(k) P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1) \right\} \big(z(k+1)$$

$$- E[Z(k+1)] - \boldsymbol{\psi}^T(k+1) P_{ZZ}^{-1}(k)\{\boldsymbol{z}(k) - E[\boldsymbol{Z}(k)]\} \big) \quad (50)$$

The sequential update of the estimation error covariance is derived similarly. The prior covariance has the form

$$\operatorname{cov}[\boldsymbol{X}(k)|\boldsymbol{z}(k)] = P_{XX}(k) - P_{XZ}(k) P_{ZZ}^{-1}(k) P_{XZ}^T(k) \quad (51)$$

We write the new covariance in partitioned form as

$$\operatorname{cov}[\boldsymbol{X}(k+1)|\boldsymbol{z}(k+1)] = P_{XX}(k+1) - [P_{XZ}(k) \quad \boldsymbol{\xi}(k+1)]$$

$$\times \begin{bmatrix} P_{ZZ}(k) & \boldsymbol{\psi}(k+1) \\ \boldsymbol{\psi}^T(k+1) & \sigma^2(k+1) \end{bmatrix}^{-1} [P_{XZ}(k) \quad \boldsymbol{\xi}(k+1)]^T \quad (52)$$

Using the matrix inversion identity (49) and collecting terms, we obtain

$$\operatorname{cov}[\boldsymbol{X}(k+1)|\boldsymbol{z}(k+1)] = \operatorname{cov}[\boldsymbol{X}(k)|\boldsymbol{z}(k)] + P_{XX}(k+1) - P_{XX}(k)$$
$$- q(k+1)\big[\boldsymbol{\xi}(k+1) - P_{XZ}(k)P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1)\big]\big[\boldsymbol{\xi}(k+1)$$
$$- P_{XZ}(k)P_{ZZ}^{-1}(k)\boldsymbol{\psi}(k+1)\big]^T \tag{53}$$

Hence, neither the state update nor the covariance update requires any new matrix inverse but merely an extra scalar division in the calculation of $q(k+1)$. Of course, in the next update from $k+1$ to $k+2$, we need $P_{ZZ}^{-1}(k+1)$ explicitly and similarly throughout the continued sequential update of state and covariance. Updating $P_{ZZ}^{-1}(k)$ to $P_{ZZ}^{-1}(k+1)$ using the matrix inversion identity (49) requires $\mathcal{O}(k^2)$ arithmetic operations. After m such updates have been performed for m autocorrelated measurements, the total number of arithmetic operations is $\mathcal{O}(m^3)$. This is a computational effort equivalent to inverting ab initio the entire $m \times m$ matrix P_{ZZ}. Nevertheless, the sequential approach distributes the computational load so that only $\mathcal{O}(k^2)$, rather than $\mathcal{O}(k^3)$, operations are required at each measurement update. This feature is desirable for applications such as the space catalog, in which it is important always to maintain estimates of the current state conditioned on all available data.

Note that the inverse-measurement-covariance update based solely on the matrix inversion identity, as just shown, is inefficient because P_{ZZ}^{-1} is symmetric. Moreover, the calculation of $q(k+1)$ may not be trouble-free numerically even though P_{ZZ}^{-1} is theoretically positive definite or at least semidefinite. Both problems can be alleviated to some extent by using the Cholesky decomposition to develop a sequential update of the square root of P_{ZZ}^{-1}, although ultimately $\mathcal{O}(m^3)$ calculations are still required. Suppose $P_{ZZ}(k+1)$ is factored as $P_{ZZ}(k+1) = S(k+1)S^T(k+1)$, so that $P_{ZZ}^{-1}(k+1) = S^{-T}(k+1)S^{-1}(k+1)$. Anticipating that we will develop a recursive form of the inverse, we partition the matrix factors as

$$P_{ZZ}(k+1) = \begin{bmatrix} P_{ZZ}(k) & \boldsymbol{\psi}(k+1) \\ \boldsymbol{\psi}^T(k+1) & \sigma^2(k+1) \end{bmatrix}$$
$$= \begin{bmatrix} S(k) & \boldsymbol{a}(k+1) \\ \boldsymbol{b}^T(k+1) & d(k+1) \end{bmatrix} \begin{bmatrix} S^T(k) & \boldsymbol{b}(k+1) \\ \boldsymbol{a}^T(k+1) & d(k+1) \end{bmatrix} \tag{54}$$

where the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ and the scalar d are to be determined. Equating partitions gives the identities

$$P_{ZZ}(k) = S(k)S^T(k) + \boldsymbol{a}(k+1)\boldsymbol{a}^T(k+1)$$
$$\boldsymbol{\psi}(k+1) = S(k)\boldsymbol{b}(k+1) + d(k+1)\boldsymbol{a}(k+1)$$
$$\sigma^2(k+1) = \boldsymbol{b}^T(k+1)\boldsymbol{b}(k+1) + d^2(k+1) \tag{55}$$

These equations verify that the general square-root factorization is not unique. However, we can select a unique factorization by requiring that $P_{ZZ}(k) = S(k)S^T(k)$ for all k. This implies that $\boldsymbol{a}(k+1) = \boldsymbol{0}$ for all k and, consequently, that S will be lower triangular. The remaining two identities can be rewritten as

$$\boldsymbol{b}(k+1) = S^{-1}(k)\boldsymbol{\psi}(k+1)$$
$$d(k+1) = \sqrt{\sigma^2(k+1) - \boldsymbol{b}^T(k+1)\boldsymbol{b}(k+1)} \tag{56}$$

Standard theorems relating to the Cholesky decomposition guarantee that $S^{-1}(k)$ always exists and that $d(k+1)$ is always real valued if the $m \times m$ matrix P_{ZZ} is positive-definite symmetric.[17,18] Then, according to the inversion identity, the matrix

$$S^{-1}(k+1) = \begin{bmatrix} S(k) & 0 \\ \boldsymbol{b}^T(k+1) & d(k+1) \end{bmatrix}^{-1}$$

can be written in explicit form as

$$S^{-1}(k+1) = \begin{bmatrix} S^{-1}(k) & 0 \\ -v(k+1)\boldsymbol{b}^T(k+1)S^{-1}(k) & v(k+1) \end{bmatrix}$$
$$v(k+1) = \frac{1}{d(k+1)} \tag{57}$$

The vector $\boldsymbol{b}(k+1)$ and scalar $d(k+1)$ are computed from Eq. (56). The recursion begins with $S^{-1}(1) = 1/\sqrt{[\sigma^2(1)]}$, where $\sigma^2(1)$ is the leading diagonal entry in the $m \times m$ matrix P_{ZZ}. Then we compute, whenever needed, $P_{ZZ}^{-1}(k+1) = S^{-T}(k+1)S^{-1}(k+1)$. Naturally, this Cholesky square-root update can be adapted to invert the measurement covariance matrix in batch processing as well.

Whitening Transformation

Note that we could solve the estimation problem by formal least squares, accounting for autocorrelations in both the measurement noise and process noise, if we first apply a whitening transformation to the data.[19] Usually only measurement noise is included in the transformation, but it is not difficult to include process noise as well. Recall that

$$\boldsymbol{x}(k) = \Phi(k,0)\boldsymbol{x}_0 + \sum_{j=-1}^{k-1} \Phi(k, j+1)\boldsymbol{\eta}(j)$$

so that an observation can be expressed as

$$z(k) = A(k)\Phi(k,0)\boldsymbol{x}_0 + A(k)\sum_{j=-1}^{k-1} \Phi(k, j-1)\boldsymbol{\eta}(j) + v(k)$$
$$\equiv A(k)\Phi(k,0)\boldsymbol{x}_0 + \delta(k) \tag{58}$$

Hence, $z(k)$ is a mapping of the initial state plus a total noise contribution containing both process and measurement noise. Then we can write the vector of all observations up to time t_k in the form

$$\boldsymbol{z}(k) = \boldsymbol{M}(k)x_0 + \boldsymbol{\delta}(k) \tag{59}$$

where $\boldsymbol{\delta}(k) = [\delta(1)\ \ \delta(2)\ \ \cdots\ \ \delta(k)]^T$ and a typical row of $\boldsymbol{M}$ has the form $A(i)\Phi(i,0)$. Compute the Cholesky decomposition of the total noise covariance, $\operatorname{cov}(\boldsymbol{\delta}) = DD^T$, and transform the data as

$$\boldsymbol{y}(k) = D^{-1}(k)\boldsymbol{z}(k) = D^{-1}(k)\boldsymbol{M}(k)\boldsymbol{x}_0 + D^{-1}(k)\boldsymbol{\delta}(k) \tag{60}$$

The components of the transformed noise are uncorrelated and in the Gaussian case are independent identically distributed (standard normal) random variables because their covariance is

$$\operatorname{cov}[D^{-1}](k)\boldsymbol{\delta}(k)] = D^{-1}(k)\operatorname{cov}[\boldsymbol{\delta}(k)]D^{-T}(k)$$
$$= D^{-1}(k)D(k)D^T(k)D^{-T}(k) \equiv I \tag{61}$$

Therefore, we can estimate $\boldsymbol{x}_0$ in terms of $\boldsymbol{y}$ by unweighted least squares. The batch solution is

$$\hat{\boldsymbol{x}}_0 = \operatorname{cov}(\hat{\boldsymbol{x}}_0)\boldsymbol{M}^T(k)D^{-T}(k)\boldsymbol{y}(k) \tag{62a}$$

where

$$\operatorname{cov}(\hat{\boldsymbol{x}}_0) = [\boldsymbol{M}^T(k)D^{-T}(k)D^{-1}(k)\boldsymbol{M}(k)]^{-1} \tag{62b}$$

Of course, in principle, nothing prevents us from reckoning the initial conditions at any convenient time t_i, so that we could estimate $\boldsymbol{x}(i)$ and its covariance directly rather than having to estimate $\boldsymbol{x}_0$ and propagate to the time of interest. Whereas this transformation method involves no more computation than Bayesian estimation, we prefer the latter for its generality. The whitening transformation might find use as a processing option in applications where least squares has previously been the method of choice and the cost of replacing the estimation software is high. In any case, we can expect that least-squares estimates will differ from Bayesian estimates because the two underlying optimality criteria differ.

Is Bayesian Estimation Practical for the Space Catalog?

The Bayesian estimator ultimately requires computation equivalent to a matrix inverse of size at least equal to the size of the largest set of correlated measurements. Whether the calculations are arranged in batch form or sequential form, eventually $\mathcal{O}(m^3)$ arithmetic operations are required for m correlated measurements. In the past, large matrix inversions would have prevented the realization of Bayesian estimates for most satellite tracking applications, including especially the space-catalog application. However, even if batch processing is to be used, we have available several stable numerical

methods, such as Cholesky decomposition, QR factorization, and Pan–Reif-type iterations,[17,18,20] for inverting the positive-definite symmetric matrix P_{ZZ}. The implementation problem is mainly one of computational throughput, rather than the preservation of numerical accuracy in the inversion algorithm. For this and other reasons, we believe that the space-catalog application may be a good candidate for Bayesian estimation. First, an accurate estimation error covariance is required with every estimate of a cataloged orbit, so that we must account for autocorrelations. Second, although the estimates have to be available in a timely manner, they do not have to be provided in an extremely fast, real-time mode. In most situations, it is adequate to have estimates and error covariances within several minutes after receipt of the last measurement, allowing time for extensive computation. Third, much progress has been made in updating the space catalog using distributed parallel computing facilities.[21] For this application, the rate at which orbit updates are produced can be made almost proportional to the number of processing nodes available, even without adopting parallel algorithms for the fundamental numerical processes themselves. As the processing time per satellite increases because of large-matrix inversion, one can simply increase the number of computing nodes used for the catalog updates to maintain the overall system performance.

Estimation of Noise Model Parameters

We now describe how to estimate the Hurst parameter and scale factor for an FGN model, using the observations. For both measurement and process noise, we assume that the noise variances at each measurement time are known a priori. We assume that the Hurst parameter is constant over the span of measurement times, but we allow the scale factor to be different at each measurement time to match any changes in the noise variance. In practice, one may be faced with temporary changes in sensor performance or solar activity, and it is convenient to have a means to accommodate such effects.

Characterizing Measurement Noise

There are a few cataloged satellites whose orbits are estimated very accurately and precisely by methods and data that are independent of the cataloging system. By comparing sensor measurements with the independent predictions for these reference orbits, we can neglect process noise and, therefore, isolate the measurement noise for the sensor. This procedure is already a routine part of SP catalog maintenance and is used to calculate fixed biases and standard deviations for each sensor's data.[22–24] We will assume that the Hurst parameters and scale factors derived for the measurement noise using these few reference orbits are also characteristic of the sensor and apply any time the sensor tracks any object. Then, because we already know the measurement noise parameters, we can use the catalog tracking data taken on a particular object to determine the process noise parameters for that orbit, as described later. Of course, it is possible that the measurement noise parameter values would change on some timescale, analogously to the weights and biases, and would have to be monitored and periodically updated. We also assume for the sake of discussion that there is no correlation between different scalar measurement types, so that the following analysis is repeated for each type.

Using the measurement model $z(k) = A(k)x(k) + f_k(\mathrm{d}B_H/\mathrm{d}t)_k$, we assume that the value of the measurement noise variance $r^2(k)$ at time t_k is known from comparisons with reference orbits. Note that $r^2(k) = f_k^2 E[(\mathrm{d}B_H/\mathrm{d}t)_k^2]$. Then, using $\mathrm{d}B_H/\mathrm{d}t \approx \Delta B_H/\Delta t$ and Eq. (20), we approximate $E[(\mathrm{d}B_H/\mathrm{d}t)_k^2] \approx V_H \Delta t_k^{2H-2}$ so that

$$r^2(k) = f_k^2 V_H \Delta t_k^{2H-2} \tag{63}$$

We take the time increment Δt_k to be a fixed parameter that characterizes the sensor measurement process. We are free to choose it conveniently, provided only that its value is small enough to allow a good approximation of FGN for the application. For example, we might choose the value of Δt_k to be on the order of the integration time for a radar measurement or even on the order of a single pulse length. In most practical cases, Δt_k would be much smaller than the time between measurements, $t_k - t_{k-1}$. If Δt_k were a fixed constant over all k, then $V_H \Delta t_k^{2H-2}$ would be constant, for a given H yet to be estimated. Hence, the ratio $c^2 = f_k^2/r^2(k)$ must be constant over all k, allowing us to compute fractal dimension.

Using Fractal Dimension to Estimate *H* for Measurement Noise

If measurements were very dense in time, we could use the time between measurements, $t_k - t_{k-1}$, to approximate the derivative of $B_H(t)$ and thereby obtain the approximation

$$c[B_H(t) - B_H(0)] = c\int_0^t \mathrm{d}B_H(\tau) \approx c\sum_i \Delta B_H(\tau_i)$$

In that case, the following fractal property could be used to estimate the Hurst parameter H.

Theorem 1: [Falconer (Ref. 13, p. 246, Theorem 16.7)]: Every FBM sample path $B_H : [0, 1] \to \Re$ (the real numbers) has a graph with box-count dimension $D\{B_H(t) : t \in [0, 1]\} = 2 - H$, almost surely.

Of course, instead of $B_H : [0, 1] \to \Re$, we have $c[B_H(t) - B_H(0)] : [0, T] \to \Re$, for finite T. Proving that sample paths of $c[B_H(t) - B_H(0)]$, $cB_H(t)$, and $B_H(t)$ all have the same box-count dimension on the interval $0 \le t \le T$ as they do on the interval $0 \le t \le 1$ turns out not to be trivial, and so we record the following two theorems.

Theorem 2: The random process $\{B_H(t) : t \in [0, T],\ T > 0\}$ has a graph with box-count fractal dimension $D = 2 - H$, almost surely.

Proof: From Theorem 2, $D\{B_H(t) : t \in [0, 1]\} = 2 - H$, almost surely. Suppose $T < 1$, and let an integer n be chosen such that $nT < 1$ and $(n + 1)T \ge 1$. Then

$$\{B_H(t) : t \in [0, 1]\}$$

$$= \bigcup_{k=0}^{n-1} \{B_H(t) : t \in [kT, (k + 1)T)\} \cup \{B_H(t) : t \in [nT, 1]\}$$

$$D\{B_H(t) : t \in [kT, (k + 1)T)\}$$

$$= D\{B_H(t) - B_H(kT) : t \in [kT, (k + 1)T)\} \qquad \forall k = 0, \ldots, n$$

because the graphs of the two random processes are the same up to an additive constant. Here

$$\{B_H(t) - B_H(kT) : t \in [kT, (k + 1)T)\}$$

$$\stackrel{d}{=} \{B_H(t) - B_H(0) : t \in [0, T)\}$$

because increments are stationary, making

$$D\{B_H(t) : t \in [kT, (k + 1)T)\} = D\{B_H(t) : t \in [0, T)\}$$

almost surely. Thus,

$$D\bigcup_{k=0}^{n-1} \{B_H(t) : t \in [kT, (k + 1)T)\} = D\{B_H(t) : t \in [0, T)\}$$

almost surely, following Barnsley (Ref. 15, p. 183, Theorem 2). Similarly,

$$D\{B_H(t) : t \in [nT, 1)\} = D\{B_H(t) : t \in [0, 1 - nT)\}$$

almost surely. Nevertheless,

$$\{B_H(t) : t \in [0, 1 - nT)\} \subseteq \{B_H(t) : t \in [0, T)\}$$

and following Barnsley (Ref. 15, p. 183, Theorem 1), subsets have equal or lower D. Hence,

$$2 - H \stackrel{\text{as}}{=} D\{B_H(t) : t \in [0, 1]\}$$

$$= D\left\{\bigcup_{k=0}^{n-1} \{B_H(t) : t \in [kT, (k + 1)T)\} \cup \{B_H(t) : t \in [nT, 1]\}\right\}$$

$$\stackrel{\text{as}}{=} D\{B_H(t) : t \in [0, T]\}$$

where superscript as is almost surely. Thus, graphically, $D\{B_H(t) : t \in [0, T]\} = 2 - H \; \forall \, T < 1$, almost surely.

Now suppose $T \geq 1$. Let an integer n be chosen such that $n < T$ and $(n+1) \geq T$. By arguments similar to the $T < 1$ case, $D\{B_H(t) : t \in [k-1, k)\} = D\{B_H(t) : t \in [0, 1)\} \; \forall \, k = 0, 1, \ldots,$ almost surely, and $D\{B_H(t) : t \in [n, T]\} = D\{B_H(t) : t \in [0, T-n]\}$, almost surely. By construction, $\{B_H(t) : t \in [0, T-n)\} \subseteq \{B_H(t) : t \in [0, 1]\}$ and subsets have equal or lower D. Hence,

$$D\{B_H(t) : t \in [0, T]\}$$

$$= D\left\{ \bigcup_{k=0}^{n-1} \{B_H(t) : t \in [k, k+1)\} \cup \{B_H(t) : t \in [n, T]\} \right\}$$

$$\stackrel{\text{as}}{=} D\{B_H(t) : t \in [0, 1]\}$$

When both cases are combined, the graph of B_H has $D\{B_H(t) : t \in [0, T]\} = 2 - H$, almost surely.

Theorem 3: Given B_H and any constant c, the graphs of B_H and cB_H have the same box-count fractal dimension $D = 2 - H \; \forall \, t \in [0, T]$, $T > 0$, almost surely. Also, they have the same box-count fractal dimension $D = 2 - H \; \forall \, t \in [t_1, t_2]$, $t_1, t_2 \in (-\infty, \infty)$, almost surely.

Proof: Multiplying an FBM by c is equivalent to stretching or contracting (with possible negative) the ordinate axis of the Euclidean metric space. Spaces before and after such scaling are metrically equivalent, and therefore, any compact subset that is a fractal has the same D under this mapping,[15] proving the first result. Then also

$$c\{B_H(t) - B_H(t_1) : t \in [t_1, t_2]\} \stackrel{d}{=} c\{B_H(t) - B_H(0) : t \in [0, |t_2 - t_1|]\}$$

because increments are stationary. As before,

$$D\{c\{B_H(t) : t \in [t_1, t_2]\}\} = D\{c\{B_H(t) : t \in [0, |t_2 - t_1|]\}\}$$

$$= D\{B_H(t) : t \in [0, |t_2 - t_1|]\} = 2 - H \quad \forall c > 0$$

almost surely, including $c = 1$, proving

$$D\{c\{B_H(t) : t \in [t_1, t_2]\}\} = D\{B_H(t) : t \in [t_1, t_2]\} = 2 - H$$

almost surely.

With these results, and because $f_k / r(k) \equiv c$ is constant, we can proceed to estimate the box-count dimension of any sample path of the real measurement noise divided by $r(k)$ because we have modeled the noise discretely as $v(k) = f_k (\mathrm{d}B_H / \mathrm{d}t)_k$. We use that value of D to calculate the value of the Hurst parameter for the FGN via $H = 2 - D$. If the computed value of H turns out not to lie in the interval $[\frac{1}{2}, 1)$ then we conclude that scalar FGN may not be an adequate noise model for this data set.

Note that Künsch[6] has devised formulas that allow H to be calculated directly from discrete data without calculating the fractal dimension. The preceding theorems provide another justification of Künsch's methods. Although our observations are discrete, our model is continuous, whereas Künsch's model is discrete with implicitly defined time steps of unity. Nevertheless, Künsch's formulas should still work because of self-similarity. Self-similarity means that changes in timescale are statistically equivalent to an amplitude scaling of FBM. By the preceding theorems, Künsch's formulas must be, almost surely, invariant with respect to any amplitude scaling. However, these formulas require that the data be equally spaced, which is not usually possible for catalog tracking data because of constraints on sensor tasking and coverage. With special tasking, sensors might be able to provide dense, equally spaced data during selected apparitions of satellites.

Using Statistical Tests of Hypothesis to Estimate H for Measurement Noise

Suppose now that the measurement data are sparse. Then we cannot use the fractal-dimension method to estimate the Hurst parameters. Instead, we resort to an alternate method, involving testing of statistical hypotheses. This is done by first choosing a trial value of $H \in [\frac{1}{2}, 1)$, solving for the scale factor from Eq. (63), $r^2(k) = f_k^2 V_H \Delta t_k^{2H-2}$, and computing the p value of the test. If rejection occurs in any test of the hypothesis, then we must make a new choice of H. After searching the entire interval, we choose the most nearly correct value of H as the one with the highest p value from the test of the hypothesis. If the highest p value turns out to be unacceptably low, then we conclude that scalar FGN may not be an adequate model of the real noise. Naturally, the threshold p value has to be set for each individual application, but we do not treat that problem here.

Now consider the measurement noise sample $\{v_1, \ldots, v_m\}$, where the lower case letters represent observations and uppercase letters represent the corresponding random variables. We would like to be able to compute sample statistics for this sequence and then perform tests of hypothesis based on guessed values of H. However, the measurement noise values are autocorrelated rather than independent and identically distributed. To account for this, we use the theorem on conditional PDFs of Gaussian random variables, Eqs. (28) and (29). With this theorem and given a guessed value of H, we know that $p(V_k | v_1, \ldots, v_{k-1})$ is Gaussian with mean and covariance given by

$$E(V_k | v_1, \ldots, v_{k-1}) = E(V_k) + \mathrm{cov}[V_k, (V_1, \ldots, V_{k-1})]$$

$$\times [\mathrm{cov}(V_1, \ldots, V_{k-1})]^{-1} \begin{bmatrix} v_1 - E(V_1) \\ \vdots \\ v_{k-1} - E(V_{k-1}) \end{bmatrix}$$

$$\mathrm{cov}(V_k | v_1, \ldots, v_{k-1}) = \sigma_{v_k}^2 - \mathrm{cov}[V_k, (V_1, \ldots, V_{k-1})]$$

$$\times [\mathrm{cov}(V_1, \ldots, V_{k-1})]^{-1} \mathrm{cov} \left[\begin{pmatrix} V_1 \\ \vdots \\ V_{k-1} \end{pmatrix}, V_k \right]$$

By the definition of conditional probabilities in general, each observation behaves as if it is a single random sample from a PDF conditional on all previous observations. This is a true PDF because all previous observations are fixed constants, no longer random variables, and so can be considered parameters. Furthermore, the joint PDF of the m observations is the product of these conditional PDFs. In particular, given random variables $X_1, \ldots, X_m$, their joint PDF is $p(x_1, \ldots, x_m) = p(x_1) p(x_2|x_1) p(x_3|x_2, x_1) \ldots p(x_m|x_1, \ldots, x_{m-1})$, where we have used the symbol p generically to denote the appropriate PDF in each instance. To prove this, observe that the Bayes rule allows us to write

$$p(x_1) p(x_2|x_1) \cdots p(x_m|x_1, \ldots, x_{m-1})$$

$$= p(x_1) \frac{p(x_1, x_2)}{p(x_1)} p(x_3|x_1, x_2) \cdots p(x_m|x_1, \ldots, x_{m-1})$$

$$= p(x_1, x_2) p(x_3|x_2, x_1) \cdots p(x_m|x_1, \ldots, x_{m-1})$$

$$= p(x_1, x_2, x_3) \cdots p(x_m|x_1, \ldots, x_{m-1})$$

The result follows by induction and shows that the joint statistical origin of the m observations can be analyzed by analyzing the conditional PDF of each one.

Now transform the sample as follows. For all ω in the sample space, we write

$$W_1(\omega) = \frac{V_1(\omega) - \mu(V_1)}{\sigma(V_1)}$$

$$W_k(\omega; v_1, \ldots, v_{k-1}) = \frac{V_k(\omega; v_1, \ldots, v_{k-1}) - \mu(V_k | v_1, \ldots, v_{k-1})}{\sigma(V_k | v_1, \ldots, v_{k-1})}$$

$$\forall k \in \{2, \ldots, m\}$$

Then $W_k(\omega; v_1, \ldots, v_{k-1}) \sim N(0, 1) \quad \forall k \in \{1, \ldots, m\}$. Because $v_j = \sigma(V_j | v_1, \ldots, v_{j-1}) w_j + \mu(V_j | v_1, \ldots, v_{j-1})$ for the observation w_j of the random variable W_j, we can consider W_j as function

of ω with parameters $w_1, w_2, \ldots, w_{j-1}$ instead of $v_1, \ldots, v_{j-1}$, namely, $W_j(\omega; w_1, \ldots, w_{j-1})$. With this construction, we can analyze the statistics of the original observations by analyzing this transformed data. Each transformed sample w_k behaves as if it were a single sample from the random variable $W_k(\omega; w_1, \ldots, w_{k-1})$ and so, by design, each w_k is characterized as if it were a single sample from $N(0,1)$.

We can indeed show that the sample $\{w_1, \ldots, w_m\}$ is characterized as if it is a random sample from $N(0,1)$. However, to do so, we need to establish two properties. The first is a lemma on joint probabilities of transformed random variables. The second is a theorem on joint PDFs of transformed random variables. We present the theorem in a vector version that is more general than actually needed in this section because of its intrinsic interest.

Lemma: Let $X_1, \ldots, X_n$ be random variables and T_k be transformations where $Y_1 = T_1(X_1), \ldots, Y_n = T_n(X_n)$ are random variables. If there exist Borel sets, $B_1, \ldots, B_n$, in the real line such that $T_k^{-1}(B)$ is Borel for any Borel set B and $k = 1, \ldots, n$, and if

$$P\left[X_1 \in T_1^{-1}(B_1), \ldots, X_n \in T_n^{-1}(B_n)\right] = \prod_{x=1}^{n} P\left[X_1 \in T^{-1}(B_k)\right]$$

then

$$P(Y_1 \in B_1, \ldots, Y_n \in B_n) = \prod_{k=1}^{n} P(Y_k \in B_k)$$

Proof:

$$P(Y_1 \in B_1, \ldots, Y_n \in B_n) = P\left[X_1 \in T_1^{-1}(B_1), \ldots, X_n \in T^{-1}(B_n)\right]$$

$$= P\left[X_1 \in T_1^{-1}(B_1)\right] \cdots P\left[X_n \in T^{-1}(B_n)\right]$$

$$= P(Y_1 \in B_1) \cdots P(Y_n \in B_n)$$

Note that Borel sets are required in this formulation because, in probability theory, sets in the real line that are assigned probabilities are defined as Borel sets in the Borel σ algebra.

Theorem 4: Given random vectors $\{\boldsymbol{X}_i : i = 1, \ldots, n\} \ni p_{X_i}(\boldsymbol{x}_i) > 0 \,\forall \boldsymbol{x}_i \in S_i$, for some set S_i, and transformation T with vectors $\boldsymbol{Y}_i = T_i(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = F_i(\boldsymbol{X}_i)$ such that T_i is 1-1 onto $\forall S_{ij}$, disjoint subsets where

$$\bigcup_{j=1}^{k_i} S_{ij} = S_i$$

if transformation images $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ exist where $p_{X_1,\ldots,X_n}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = p_{X_1}(\boldsymbol{x}_1) \cdots p_{X_n}(\boldsymbol{x}_n) \,\forall \boldsymbol{x}_i \in F_i^{-1}(\boldsymbol{y}_i)$, then $p_{Y_1,\ldots,Y_n}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = p_{Y_1}(\boldsymbol{y}_1) \cdots p_{Y_n}(\boldsymbol{y}_n)$.

Proof: Because $\times_{i \neq j} S_i \times S_j$ partitions $\times_i S_i$, from transformation theory of random vectors we have

$$p_{Y_1,\ldots,Y_n}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$$

$$= \sum_{r_n=1}^{k_n} \cdots \sum_{r_1=1}^{k_1} \left| \frac{\partial(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}{\partial(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)} \right| p_{X_1,\ldots,X_n}\left[F_{1r_1}^{-1}(\boldsymbol{y}_1), \cdots, F_{nr_n}^{-1}(\boldsymbol{y}_n)\right]$$

where $F_{ir_i}^{-1}$ is the inverse of F_i that is 1–1 on to $S_{ir_i} \in (S_{i1}, \ldots, S_{ik_i})$. Notice that because $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are vectors, the transformation T gives a Jacobian that is the determinant of a matrix that is block diagonal in the submatrices of the Jacobians $\partial \boldsymbol{x}_i / \partial \boldsymbol{y}_i \equiv J(\boldsymbol{x}_i)$. By the property of determinants of block diagonal matrices,

$$\frac{\partial(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}{\partial(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)} = \prod_{i=1}^{n} J(\boldsymbol{y}_i)$$

Hence,

$$(*) \equiv p_{Y_1,\ldots,Y_n}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$$

$$= \sum_{r_n=1}^{k_n} \cdots \sum_{r_1=1}^{k_1} \prod_{i=1}^{n} \left| J_{ir_i}(\boldsymbol{y}_i) \right| \prod_{i=1}^{n} p_{X_i}\left[F_{ir_i}^{-1}(\boldsymbol{y}_i)\right]$$

since $F_{ir_i}^{-1}(\boldsymbol{y}_i) = \boldsymbol{x}_{ir_i} \in F_i^{-1}(\boldsymbol{y}_i)$, where J_{ir_i} is the Jacobian $\partial x_{ir_i} / \partial y_i$. By the transformation of random variables,

$$\prod_{i=1}^{n} p_{Y_i}(\boldsymbol{y}_i) = \prod_{i=1}^{n} \left\{ \sum_{r_i=1}^{k_i} \left| J_{ir_i}(\boldsymbol{y}_i) \right| p_{X_i}\left[F_{ir_i}^{-1}(\boldsymbol{y}_i)\right] \right\}$$

$$= \sum_{r_n=1}^{k_n} \cdots \sum_{r_1=1}^{k_1} \prod_{i=1}^{n} \left| J_{ir_i}(\boldsymbol{y}_i) \right| p_{X_i}\left[F_{ir_i}^{-1}(\boldsymbol{y}_i)\right]$$

$$= \sum_{r_n=1}^{k_n} \cdots \sum_{r_1=1}^{k_1} \prod_{i=1}^{n} \left| J_{ir_i}(\boldsymbol{y}_i) \right| \prod_{i=1}^{n} p_{X_i}\left[F_{ir_i}^{-1}(\boldsymbol{y}_i)\right] \equiv (*)$$

$$\therefore \quad p_{Y_1,\ldots,Y_n}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = p_1(\boldsymbol{y}_1) \cdots p_n(\boldsymbol{y}_n)$$

Furthermore, suppose there are points where 1–1 onto subsets do not exist, but some components have constant images over positive-probability preimages. We can define a generalized PDF with marginal distributions of positive probabilities at these constants. With the theorem hypothesis using such a generalized PDF, these components on the domain sets with constant images are statistically independent of other components. Using the lemma before this theorem, we could show that the theorem still holds for this type of generalized PDF.

Now statistical tests can be performed to assess whether $\{w_1, \ldots, w_m\}$ really behaves as if it were a random sample from $N(0,1)$, inferring on the correctness of the guessed values of the measurement-noise Hurst parameter and scale factor. Given sufficiently numerous data, a χ^2 goodness-of-fit test can designed for any distribution, where the $N(0,1)$ distribution in our application is merely one specific choice of distribution. In particular, given any PDF, partition the range of its associated random variable into m bins, for example, $(-\infty, a_1]$, $(a_1, a_2]$, $(a_2, a_3]$, $\ldots$, (a_m, ∞). Compute probabilities that observations fall into certain bins via

$$P_k = \int_{a_k}^{a_{k+1}} p(x)\, \mathrm{d}x$$

for PDF p where $a_0 = -\infty$ and $a_{m+1} = \infty$. Given $P_1, \ldots, P_m$, perform a χ^2 goodness-of-fit test to infer the appropriateness of the assumed PDF p.

We note that transforming to independent, identically distributed variables for non-Gaussian cases may be impractical. However, for any random sequence, $(X_j : j = 1, \ldots, n)$, Gaussian or not, we can do statistical inference after deriving distributions conditioned upon past observations without transforming to identical distributions, by partitioning the range into m bins, $(a_0 = -\infty, a_1]$, $(a_1, a_2]$, $\ldots$, $(a_m, a_{m+1} = \infty)$. We model the probability of picking observation j in bin k as

$$P(I = j, x_j \in (a_k, a_{k+1}]) = \left(\frac{1}{n}\right) \int_{a_k}^{a_{k+1}} p_{X_j}(a|x_1, \ldots, x_{j-1})\, \mathrm{d}a$$

for all $j = 1, \ldots, n$, noting that $j = 1$ involves no prior conditioning. We then model the probability that any $x \in \{x_1, \ldots, x_n\}$ falls within a certain bin as if it is the marginal distribution

$$P(x \in (a_k, a_{k+1}]) = \sum_{j=1}^{n} \left(\frac{1}{n}\right) \int_{a_k}^{a_{k+1}} p_{X_j}(a|x_1, \ldots, x_{j-1})\, \mathrm{d}a$$

This is a true probability measure because

$$P(x \in (-\infty, \infty)) = \sum_{k=0}^{m} \sum_{j=1}^{n} \frac{1}{n} \int_{a_k}^{a_{k+1}} p_{X_j}(a|x_1, \ldots, x_{j-1})\, \mathrm{d}a$$

$$= \frac{1}{n} \sum_{j=1}^{n} \sum_{k=0}^{m} \int_{a_k}^{a_{k+1}} p_{X_j}(a|x_1, \ldots, x_{j-1})\, \mathrm{d}a = \frac{1}{n} \sum_{j=1}^{n} 1 = 1$$

With these derived probability measures and counting the numbers of observations per bin of $(X_j : j = 1, \ldots, n)$, a χ^2 goodness-of-fit or a Kolmogorov–Smirnov test can be performed to decide the suitability of a hypothesized distribution or parameter of $(X_j : j = 1, \ldots, n)$. In our specific application, the observations are $v_1, \ldots, v_n$, and there is now no need to standardize into $w_1, \ldots, w_n$. Note also that in other applications one may not be able to transform the random variables into identically distributed ones. However, it is always helpful to subtract out their means if these are known. This reduces the variation of the randomized data, thereby increasing the sensitivity of a test of hypothesis.

Alternatively, under Gaussian assumptions, we could use eigenvector analysis to transform to independent random variables, an approach best suited for sparse data because it is computationally intensive. Let observations $v_1, \ldots, v_m$ be from the noise $V(k), k = 1, \ldots, m$, and let $\boldsymbol{V}$ be the m-dimensional vector of such noises. With eigenvector analysis, a matrix M^T of orthogonal eigenvector rows can be found that diagonalizes $\mathrm{cov}(\boldsymbol{V})$, where $\boldsymbol{V}^T = [V(1), \ldots, V(m)]$ and $\mathrm{cov}(\boldsymbol{V})$ is obtained from the given input variance of each $V(k), k = 1, \ldots, m$, and the autocovariances from FGN properties. In this way, $M^T\boldsymbol{V}$ is a vector of independent random variables when the Hurst parameter is correctly guessed. We can standardize each component of $M^T\boldsymbol{V}$ and test the hypothesis that the components behave as random samples from a Gaussian $N(0, 1)$ distribution. Actually, this idea is usable for any Gaussian process, not just FGN, as long as the means and autocovariances are known. However, in general, the method will not work unless zero correlation is equivalent to independence.

Naturally, the test-of-hypothesis methods can be used even for dense data. Given $v_1, \ldots, v_m$ dense enough to estimate the Hurst parameter by fractal theory, we can still take a (possibly sparse) subsequence of it. The test-of-hypothesis methods can be used as a check on the fractal theory's estimated Hurst parameter, or vice versa.

Special Results for Measurement Noise

Suppose the measurement noise variance $r^2(t)$ is constant $\forall t$, but unknown. In this case the measurement noise is

$$V(k) \equiv V(t_k) = f\frac{\mathrm{d}B_H(t_k)}{\mathrm{d}t_k}$$

where we will explain that f is constant. Barton and Poor[7] described two transformations, one for infinite time intervals and one for finite intervals, which transform FGN into Brownian motion. Because Brownian motion has independent increments, we can perform sample statistics to estimate values for f and, if desired, also r^2. We have already exhibited the transformation for the infinite-time case as Eq. (14). The theorem for the finite time interval case is as follows.

Theorem 5: (Barton and Poor[7]): There exists a standard Brownian motion $B_T = \{B_T(t) : t \in [0, T]\}$ such that

$$B_T(t) = \frac{1}{\Gamma\left(\frac{3}{2} - H\right)} \int_0^t \tau^{H-\frac{1}{2}}\,\mathrm{d}_\tau \int_0^\tau (\tau - \lambda)^{\frac{1}{2}-H}\lambda^{\frac{1}{2}-H}\,\mathrm{d}B_{H|T}(\lambda)$$

for FBM $B_{H|T}$ on $[0, T]$.

With the formula for the noise being $V(t_k) = f[\mathrm{d}B_H(t_k)/\mathrm{d}t_k]$, the variance has the form $r^2 = f^2E[(\Delta B_H(t)/\Delta t)^2]$. With increments of FBM being stationary, $E[(\Delta B_H(t)/\Delta t)^2]$ is constant for fixed Δt. Therefore, f is constant if r is constant.

Thus, with our guessed Hurst parameter value H, and assuming that the data are dense enough for sums to approximate integrals, we can use Theorem 5 to numerically approximate

$$fB_T(t) = \frac{f}{\Gamma\left(\frac{3}{2} - H\right)} \int_0^t \tau^{H-\frac{1}{2}}\,\mathrm{d}_\tau \int_0^\tau (\tau - \lambda)^{\frac{1}{2}-H}\lambda^{\frac{1}{2}-H}\,\mathrm{d}B_{H|T}(\lambda)$$

With a correctly guessed H value, B_T is Brownian motion and, thus, has independent increments. When equal time-interval data are assumed, the increments are also identically distributed with known Gaussian statistics. Thus, we can compute the sample variance of the equal-time-interval increments of the data $fB_T(t)$, namely, $\{f[B_T(t_k) - B_T(t_{k-1})] : k = 1, \ldots, n\}$, and equate it to f^2 times the theoretical variance of $B_T(t)$ increments:

$$s^2_{f\delta B_T} = f^2\sigma^2_{\delta B_T} = f^2(t_k - t_{k-1})$$

Hence, we can simply solve for the positive value of f and no longer depend on an accurate given input variance of the measurement noise.

We must repeatedly do tests of hypothesis on each transformed sample set, $\{B_T(t_k) - B_T(t_{k-1}) : k = 1, \ldots, n\}$, for each guessed Hurst parameter value. We test the hypothesis that the set has values as if it were a random sample of a normal distribution with zero mean and variance of $(t_k - t_{k-1})$. We choose the guessed Hurst parameter that yields a high test-of-hypothesis p value. With the estimated f, we can also solve for the estimated measurement noise variance r^2 if desired.

For constant scale factors, eigenvectors are useful in the absence of input measurement noise variances, when data are too sparse to perform the integration to Brownian motion. Given $V(k) \equiv V(t_k) = f[\mathrm{d}B_H(t_k)/\mathrm{d}t_k]$ with constant scale factor f and the vector $\boldsymbol{V}^T = [V(1), \ldots, V(n)]$, of random variables corresponding to n observations,

$$\mathrm{cov}(V) = f^2\mathrm{cov}\left[\left(\frac{\mathrm{d}B_H}{\mathrm{d}t}\right)\left(\frac{\mathrm{d}B_H}{\mathrm{d}t}\right)^T\right] \ni \left(\frac{\mathrm{d}B_H}{\mathrm{d}t}\right)^T$$

$$\equiv \left[\left(\frac{\mathrm{d}B_H(t_1)}{\mathrm{d}t_1}\right), \ldots, \left(\frac{\mathrm{d}B_H(t_n)}{\mathrm{d}t_n}\right)\right]$$

We know the theoretical covariance of a vector of FGN samples, and thus, it can be diagonalized with a matrix of orthogonal eigenvector rows M^T. Hence, with a correctly guessed H value for the assumed FGN process, $M^T\boldsymbol{V} = fM^T(\mathrm{d}\boldsymbol{B}_H/\mathrm{d}t)$ must be independent and Gaussian, with the diagonal covariance matrix,

$$\mathrm{cov}(M^T\boldsymbol{V}) = f^2M^T\mathrm{cov}\left[\left(\frac{\mathrm{d}\boldsymbol{B}_H}{\mathrm{d}t}\right)\left(\frac{\mathrm{d}\boldsymbol{B}_H}{\mathrm{d}t}\right)^T\right]M$$

Diagonal components of this covariance without the unknown f^2 factor can be standardized to unity by multiplying by a diagonal matrix D, whose diagonal elements are the reciprocals of the diagonals of $M^T\mathrm{cov}[(\mathrm{d}\boldsymbol{B}_H/\mathrm{d}t)(\mathrm{d}\boldsymbol{B}_H/\mathrm{d}t)^T]M$. Thus, with a correctly guessed H, each component of $D^{1/2}M^T\boldsymbol{V}$ must be identically distributed (IID) with the distribution $N(0, 1)$. With the guessed H, f is the only unknown quantity in this covariance. Assuming IID values, the sample variance $s^2_{M^TV}$ of these components can be computed. Then we have

$$S^2_{M^TV} = f^2D^{\frac{1}{2}}M^T\mathrm{cov}\left[\left(\frac{\mathrm{d}\boldsymbol{B}_H}{\mathrm{d}t}\right)\left(\frac{\mathrm{d}\boldsymbol{B}_H}{\mathrm{d}t}\right)^T\right]MD^{\frac{1}{2}} = f^2I$$

where $S^2_{M^TV}$ is the diagonal matrix of identical $s^2_{M^TV}$ values. When only $f > 0$ is used, this implies $f = s_{M^TV}$. Then all quantities of the covariance are computed. Now we can test the hypothesis that components of $M^T\boldsymbol{V}$ behave as $N(0, 1)$ and choose the H value that yields the highest test-of-hypothesis p value. As before, with the estimated f we can also solve for the estimated measurement noise variance r^2 if desired. Note that this idea is applicable not only to FGNs but also any Gaussian processes with constant scale factors provided the autocovariances and means of the Gaussian process without the constant scale factor are known.

Characterizing Process Noise

In the space-catalog application, we can reasonably assume that a given sensor will have the same measurement noise characteristics no matter what satellite is being tracked. Hence, we can use a limited number of high-precision reference orbits to isolate the actual measurement noise for each sensor. However, we do not have an analogous means of isolating the process noise, which is liable

to have different characteristics for each satellite in the catalog because we have only the measurements to work with. Although this prevents us from estimating the fractal dimension of the process noise, we can still estimate the Hurst parameter indirectly from the observations after all parameters for the measurement noise have been estimated. A statistical test-of-hypothesis approach using conditional probability theory will be described.

The process noise covariance is

$$Q_k \equiv \text{cov}[\boldsymbol{\eta}(k)] = \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau) G_k E\left[\text{d}\boldsymbol{B}_{H'}(\tau) \text{d}\boldsymbol{B}_{H'}^T(\alpha)\right] \times \Phi^T(t_{k+1}, \alpha) G_k \tag{64}$$

Recall that only the seventh (atmospheric drag) and eighth (solar radiation) state components are driven by process noise. Hence, we have only the H'_7 and H'_8 Hurst parameters to estimate. We then require that only the seventh and eighth diagonal elements of Q_k, q_{k77} and q_{k88}, be given. Recalling the earlier derivation of the (7, 7) and (8, 8) components of cov$[\boldsymbol{\eta}(k), \boldsymbol{\eta}(\ell)]$, we can write, from Eq. (39),

$$q_{kjj} = V_{H'_j} g_{kjj}^2 H'_j (2H'_j - 1) \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} |\tau - \alpha|^{2H'_j - 2} \, \text{d}\tau \, \text{d}\alpha = V_{H'_j} g_{kjj}^2 (t_{k+1} - t_k)^{2H'_j}, \qquad j = 7, 8 \tag{65}$$

Assuming that the Hurst parameter is known, we solve for the (nonnegative) scale factor g_{kjj}, $j = 7, 8$. Furthermore, with values for H'_7, H'_8, g_{k77}, and g_{k88} available, all of the components of cov$[\boldsymbol{\eta}(k), \boldsymbol{\eta}(\ell)]$ can be computed $\forall k, \ell \in \{0, 1, \ldots, m\}$, as needed in the covariance matrix for the Bayesian estimator. As we did in characterizing measurement noise, we assume that the Hurst parameters are constant over the data span. However, we allow the scale factors to be different at each time to match the possibly changing process noise variances.

With the measurement noise parameters accurately estimated, the complete statistical description of the measurements Z as a function of state $\boldsymbol{X}$ is available. Let $\{z_1, \ldots, z_m\}$ be the set of observations of $Z(t)$. We will infer values of the parameters of the process noise,

$$\boldsymbol{\eta}(k) = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1}, \tau) G_k \, \text{d}\boldsymbol{B}_{H'}(\tau)$$

by doing statistics on the observations, the values of the Z_k. Given the initial state, the state model equation, the measurement model equation, and the estimated measurement noise parameters, make an initial guess for the Hurst parameters H'_7 and H'_8. Evaluate Eq. (65), given the guessed H'_7 and H'_8, $\forall k \in \{1, \ldots, m\}$. Set q_{k77} and q_{k88} on the left side, equal to the corresponding elements on the right side, and solve the two quadratic equations for the two scale factors, g_{k77} and g_{k88}. With the guessed Hurst parameters and their corresponding scale factors, we now have a complete statistical description of both the state and measurement noises.

Note that because Z_k is a linear function of various Gaussian processes, it must be Gaussian, and recall that conditional PDFs of joint Gaussian random variables are Gaussian. Hence, as with the measurement noise, we can find all of the PDFs $p(Z_1)$, $p(Z_2|z_1)$, $p(Z_3|z_1, z_2)$, $\ldots$, $p(Z_n|z_1, \ldots, z_{n-1})$ using the theorem on conditional PDFs given before [Eqs. (28) and (29)]. The conditional PDFs are Gaussian single-variate PDFs with means and variances parameterized by all previous observations, namely, $\mu(Z_k|z_1, \ldots, z_{k-1}) = E[Z_k|z_1, \ldots, z_{k-1}]$ and $\sigma^2(Z_k|z_1, \ldots, z_{k-1}) = E\{[Z_k - \mu(Z_k|z_1, \ldots, z_{k-1})]^2|z_1, \ldots, z_{k-1}\}$. For the same reasons as in analyzing the measurement noise, each Z_k behaves as if it is sampled from a PDF conditioned on all previous observations. Also, the joint statistical origin of all m observations, the joint PDF of all m Z_k, is the product of these PDFs. The random variable corresponding to each conditioned distribution can be written as $Z_k(\omega; z_1, \ldots, z_{k-1}) \sim N[\mu(Z_k|z_1, \ldots, z_{k-1}), \sigma^2(Z_k|z_1, \ldots, z_{k-1})]$, with $Z_1(\omega) \sim N[\mu(Z_1), \sigma^2(Z_1)]$, for all ω in the sample space. We can, as before, transform to standardized Gaussian random variables:

$$Y_1(\omega) = \frac{Z_1(\omega) - \mu(Z_1)}{\sigma(Z_1)}$$

$$Y_k(\omega; z_1, \ldots, z_{k-1}) = \frac{Z_k(\omega; z_1, \ldots, z_{k-1}) - \mu(Z_k|z_1, \ldots, z_{k-1})}{\sigma(Z_k|z_1, \ldots, z_{k-1})} \qquad \forall k \in \{2, \ldots, m\}$$

Hence, $Y_k(\omega; z_1, \ldots, z_{k-1}) \sim N(0, 1)\ \forall k \in \{1, \ldots, m\}$. Just as in the case of the w_k in the measurement noise transformed data, each transformed observation $y_k = [z_k - \mu(Z_k|z_1, \ldots, z_{k-1})]/\sigma(Z_k|z_1, \ldots, z_{k-1})$ behaves as if it were a single random sample from the random variable $Y_k(\omega; z_1, \ldots, z_{k-1}) \sim N(0, 1)$, and the transformed data $\{y_1, \ldots, y_m\}$ can be handled as if they were independent as well as identically distributed. Hence, we can successively guess values of H'_7 and H'_8, compute g_{k77} and g_{k88}, and check their validity as described before in the measurement noise case, using the same test-of-hypothesis techniques.

Another useful fact is that, because Y_k is standard normal,

$$B_k^2 = \sum_{k=1}^{j} Y_k^2$$

has a χ^2 distribution with j degrees of freedom. Hence, the validity of the parameters of the original distribution can be checked by a test of hypothesis on either $Y_k \sim N(0, 1)$ or $B_k^2 \sim \chi^2(j)$, by partitioning the range of the Y_k or B_k^2 as described before in the measurement noise section. Then a χ^2 goodness-of-fit or a Kolmogorov–Smirnov test can be performed. Alternatively, we can partition the ranges of the Z_k into bins, without any transformation of the data, and model each bin as a summation of conditional PDFs as derived in the measurement noise section, testing the hypothesis on this partition.

Conclusions

In this paper, we formulated a Bayesian estimator that can estimate the state of a system having autocorrelated process and measurement noises. We used FGN to model persistently autocorrelated noise. We derived batch and sequential versions of the estimator and showed how the sequential version distributes the computing load over measurement updates. We also showed how to modify least-squares methods to account for process noise by including this noise in a whitening transformation. We discussed methods to invert large matrices and concluded that parallel computing can make Bayesian estimation feasible for the space-catalog application.

We proposed that scale factors of FGN might be suitable for modeling persistently autocorrelated process and measurement noises in orbit determination applications. We described how to tailor the model to data by guessing the Hurst parameter and then solving for the scale factor. Tests of hypothesis on guessed parameter values were derived, in which the p values of the tests are to be optimized with respect to the Hurst parameter. If measurement data were sufficiently dense, the Hurst parameter for a sensor could be estimated by calculating the fractal dimension of a sample of the measurement noise, assuming that the measurement noise is the same for all satellites and can be isolated through sensor calibration. However, process noise is different for each satellite and cannot be isolated because we have only the measurements to work with. Hence, we cannot calculate its fractal dimension.

Preliminary tests of our method can be performed by simulating FGN with known parameters, for example, transforming standard Brownian motion as in Eq. (13) and taking increments. Then calculations of fractal dimension or tests of hypothesis can be used to recover the known parameter values. Afterward, we can test real data to see whether the application is actually characterized by FGN. If FGN adequately represents the real noise, the calculated value of the Hurst parameter will lie between known fixed limits.

The extent to which FGN models real noise processes in orbit determination applications remains unknown. In particular, we need to verify whether it is reasonable to assume that satellite tracking

sensors will have a nearly fixed value of the Hurst parameter for their measurement noises. We also do not know whether measurement autocorrelations may persist for long times or are confined to a single apparition of the satellite.

Acknowledgments

The authors thank John Seago of Honeywell Technology Solutions, Inc., at Naval Research Laboratory for critiquing early versions of this work and for offering many helpful suggestions. James Gentle of George Mason University also provided valuable insight into the statistical methods outlined in this paper.

References

[1]Barker, W. N., Casali, S. J., and Walker, C. A. H., "Improved Space Surveillance Network Observation Error Modeling and Techniques for Force Model Error Mitigation," *Advances in the Astronautical Sciences*, Vol. 103, Pt. 3, Univelt, San Diego, CA, 2000, pp. 1881–1901.

[2]Wright, J. R., "Sequential Orbit Determination with Autocorrelated Gravity Modeling Errors," *Journal of Guidance and Control*, Vol. 4, No. 3, 1981, pp. 304–309.

[3]Wright, J. R., "Orbit Determination Solution to Non-Markov Gravity Error Problem," *Advances in the Astronautical Sciences*, Vol. 87, Pt. 2, Univelt, San Diego, CA, 1994, pp. 1153–1172.

[4]Wilkins, M. P., and Alfriend, K. T., "Characterizing Orbit Uncertainty Due to Atmospheric Uncertainty," *Collection of the AIAA/AAS Astrodynamics Specialist Conference Technical Papers*, AIAA, Reston, VA, 2000, pp. 19–29.

[5]Mandelbrot, B. B., *The Fractal Geometry of Nature (Updated and Augmented)*, W. H. Freeman, New York, 1983.

[6]Künsch, H., "Statistical Aspects of Self-Similar Processes," *Bernoulli*, Vol. 1, 1987, pp. 67–74.

[7]Barton, R. J., and Poor, H. V., "Signal Detection in Fractional Gaussian Noise," *IEEE Transactions on Information Theory*, Vol. 43, No. 5, 1988, pp. 943–959.

[8]Chow, W. C., "Modeling and Inference with Fractional Brownian Motion and Fractional Gaussian Noise," Ph.D. Dissertation, School of Information Technology and Engineering, George Mason Univ., Fairfax, VA, Spring 1994.

[9]Chow, W. C., "Estimation Theory with Fractional Gaussian Noise," NSWCDD TR-95-145, Naval Surface Warfare Center, Dahlgren, VA, Sept. 1995.

[10]Bryson, A. E., Jr., and Ho, Y.-C., *Applied Optimal Control: Optimization, Estimation, and Control*, rev. printing, Hemisphere, Washington, DC, 1975, distributed by Wiley, New York, Sec. 10.6.

[11]Pietgen, H.-O., Jürgens, H., and Saupe, D., *Chaos and Fractals: New Frontiers of Science*, Springer-Verlag New York, 1992, Sec. 9.5.

[12]Mandelbrot, B. B., and Van Ness, J. W., "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, No. 4, 1968, pp. 422–437.

[13]Falconer, K., *Fractal Geometry: Mathematical Foundations and Applications*, Wiley, Chichester, England, U.K., 1990, Chap. 16.

[14]Schroeder, M., *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, W. H. Freeman, New York, 1991, Chap. 5.

[15]Barnsley, M., *Fractals Everywhere*, Academic Press, San Diego, CA, 1988, Chap. 5.

[16]Meditch, J. S., *Stochastic Optimal Linear Estimation and Control*, McGraw–Hill, New York, 1969, Sec. 3.6.

[17]Ralston, A., and Rabinowitz, P., *A First Course in Numerical Analysis*, 2nd ed., McGraw–Hill, New York, 1978.

[18]Stoer, J., and Bulirsch, R., *Introduction to Numerical Analysis*, 2nd ed., Springer-Verlag, New York, 1993.

[19]Bierman, G. J., *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977, Sec. 3.5.

[20]Pan, V., and Reif, J., "Efficient Parallel Solution of Linear Systems," *Proceedings of the 17th Annual ACM Symposium on Theory of Computing*, ACM, Providence, RI, 1985, pp. 143–152.

[21]Coffey, S. L., Neal, H. L., Visel, C. L., and Conolly, P., "Demonstration of a Special-Perturbations-Based Catalog in the Naval Space Command System," *Advances in the Astronautical Sciences*, Vol. 99, Univelt, San Diego, CA, 1998, pp. 227–247.

[22]Gilbreath, G. C., Schumacher, P. W., Jr., Davis, M. A., Lydick, E. D., and Anderson, J. M., "Evaluation of the Naval Space Surveillance Fence Performance Using Satellite Laser Ranging," *Journal of Guidance, Control, and Dynamics*, Vol. 22, No. 1, 1999, pp. 149–155.

[23]Schumacher, P. W., Jr., Gilbreath, G. C., Davis, M. A., and Lydick, E. D., "Precision of Satellite Laser Ranging Calibration of the Naval Space Surveillance System," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 5, 2001, pp. 925–932.

[24]Seago, J. H., Davis, M. A., Reed, A. E., Lydick, E. D., and Schumacher, P. W., Jr., "Results of Naval Space Surveillance System Calibration Using Satellite Laser Ranging," *Advances in the Astronautical Sciences*, Vol. 109, Pt. 1, Univelt, San Diego, CA, 2001, pp. 837–856.