ELSEVIER

Journal of Quantitative Spectroscopy & Radiative Transfer 84 (2004) 3–34

Journal of Quantitative Spectroscopy & Radiative Transfer

www.elsevier.com/locate/jqsrt

# Photon propagation in heterogeneous optical media with spatial correlations: enhanced mean-free-paths and wider-than-exponential free-path distributions

Anthony B. Davis[a,*], Alexander Marshak[b]

[a] *Los Alamos National Laboratory, Space & Remote Sensing Sciences Group (NIS-2), Los Alamos, NM 87545, USA*

[b] *NASA's Goddard Space Flight Center, Climate & Radiation Branch (Code 913), Greenbelt, MD 20771, USA*

Received 26 November 2001; received in revised form 29 December 2002; accepted 27 February 2003

**Abstract**

Beer's law of exponential decay in direct transmission is well-known but its break-down in spatially variable optical media has been discussed only sporadically in the literature. We document here this break-down in three-dimensional (3D) media with complete generality and explore its ramifications for photon propagation. We show that effective transmission laws and their associated free-path distributions (FPDs) are in fact never exactly exponential in variable media of any kind. Moreover, if spatial correlations in the extinction field extend at least to the scale of the mean-free-path (MFP), FPDs are necessarily wider-than-exponential in the sense that all higher-order moments of the relevant mean-field FPDs exceed those of the exponential FPD, even if it is tuned to yield the proper MFP. The MFP itself is always larger than the inverse of average extinction in a variable medium. In a vast and important class of spatially-correlated random media, the MFP is indeed the average of the inverse of extinction. We translate these theoretical findings into a practical method for deciding a priori when 3D effects become important. Finally, we discuss an obvious but limited analogy between our analysis of spatial variability and the well-known effects of strong spectral variability in gaseous media when observed or modeled at moderate resolution.


*Keywords:* Beer's law; Photon propagation; Mean-free-path; Three-dimensional radiative transfer; Spatially correlated heterogeneous media; Cloudy atmospheres

* Corresponding author. Tel.: +1-505-665-6577 (voice); fax: +1-505-667-9208.
*E-mail address:* adavis@lanl.gov (A.B. Davis).

0022-4073/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0022-4073(03)00114-6

## 1. Introduction

### 1.1. Definitions

In the absence of all sources, including multiple scattering, the three-dimensional (3D) radiative transfer (RT) equation reduces to

$$\boldsymbol{\Omega} \cdot \nabla I + \sigma(\boldsymbol{x}) I(\boldsymbol{x}, \boldsymbol{\Omega}) = 0, \tag{1}$$

where $I(\boldsymbol{x}, \boldsymbol{\Omega})$ is the radiance field dependent on position $\boldsymbol{x}$ and direction of propagation $\boldsymbol{\Omega}$. The extinction field $\sigma(\boldsymbol{x})$ describes the local depletion rate of the photon population in a beam passing through $\boldsymbol{x}$; for simplicity, we will not explicitly denote a possible dependence on $\boldsymbol{\Omega}$. We are interested in radiative processes unfolding along a given beam $(\boldsymbol{x}_0, \boldsymbol{\Omega}_0)$, so we can set $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{\Omega}_0 s$, hence $\boldsymbol{\Omega}_0 \cdot \nabla = \mathrm{d}/\mathrm{d}s$ for the directional derivative. Then Eq. (1) reads simply as

$$\frac{\mathrm{d}I}{\mathrm{d}s} = -\sigma(s) I(s). \tag{2}$$

We now introduce the standard change of variables in 1D (plane-parallel) RT theory,

$$\mathrm{d}\tau(s) = \sigma(s)\,\mathrm{d}s \tag{3}$$

that defines optical distance as a function of $s \geqslant 0$. This leads to $\mathrm{d}I/\mathrm{d}\tau = -I$ with $I(0) = I_0$, thus

$$I(s) = I_0 \exp[-\tau(s)]. \tag{4}$$

Restoring the dependence on $(\boldsymbol{x}_0, \boldsymbol{\Omega}_0)$ as fixed parameters (offset throughout this paper by a semicolon), the optical distance appearing in (4) is

$$\tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s) = \int_0^s \sigma(\boldsymbol{x}_0 + \boldsymbol{\Omega}_0 s')\,\mathrm{d}s'. \tag{5}$$

The fundamental law of exponential cumulative extinction in (4) has a simple probabilistic interpretation in kinetic theory. Using "|" to offset given quantities held fixed while drawing random variables, we have the following probability of direct transmission:

$$\mathscr{T}(\boldsymbol{x}_0 \to \boldsymbol{x}_0 | \boldsymbol{\Omega}_0 s) = \mathrm{Prob}\{\mathit{physical\ \ step} \geqslant s | \boldsymbol{x}_0, \boldsymbol{\Omega}_0\} = \mathrm{Prob}\{\mathit{optical\ \ step} \geqslant \tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s)\}. \tag{6}$$

Using (4), we obtain

$$\mathscr{T}(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s) = I(s)/I_0 = \exp[-\tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s)]. \tag{7}$$

In homogeneous media, where $\tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s) \equiv \sigma s$, Eqs. (4) and (7) are known in the literature as Beer's law, or Bouger–Beer law, and sometimes Lambert's name is also associated.

In addition, we can relate the direct transmission $\mathscr{T}(\boldsymbol{x}_0 \to \boldsymbol{x}_0 + \boldsymbol{\Omega}_0 s)$ to the free-path distribution law—probability of a step along the beam $(\boldsymbol{x}_0, \boldsymbol{\Omega}_0)$ to be between $s$ and $s + \mathrm{d}s$ $(\mathrm{d}s \geqslant 0)$—as

$$\mathrm{d}P(s | \boldsymbol{x}_0, \boldsymbol{\Omega}_0) = -\left(\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}s}\right)\mathrm{d}s = \left|\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}s}\right|\mathrm{d}s. \tag{8}$$

From this step distribution, we can compute the mean-free-path (MFP), or other moments, which will of course depend on $(\boldsymbol{x}_0, \boldsymbol{\Omega}_0)$.

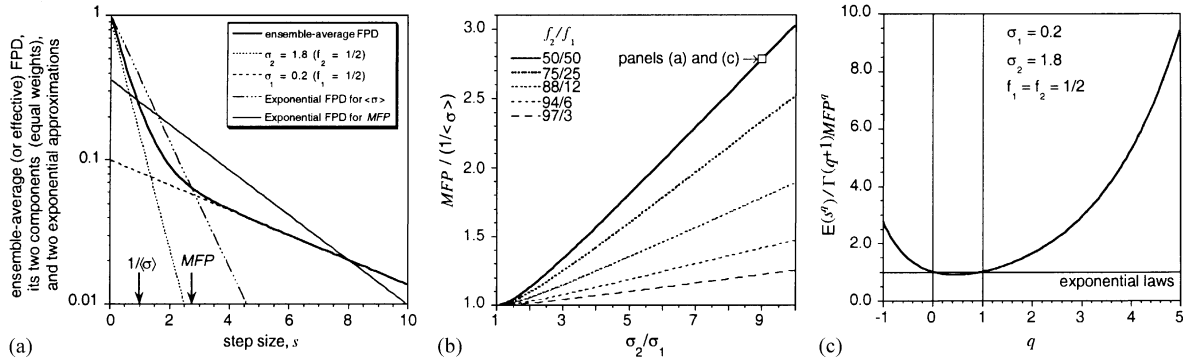


Fig. 1. Illustration of non-exponential transmission/FPD in a binary medium. In this simple model, extinction is $\sigma_1$ with probability $f_1$ and $\sigma_2 \geqslant \sigma_1$ with probability $f_2 = 1 - f_1$. (a) The actual FPD is compared with its two exponential components for $\sigma_1 = 0.2$ and $\sigma_2 = 1.8$ with $f_1 = f_2 = 1/2$ which is a scaled version of the "square-wave" cloud or "Case 1" used in the Intercomparison of 3D Radiation Codes ⟨http://climate.gsfc.nasa.gov/I3RC/⟩ under normal illumination; the short steps are dominated by the dense fraction and the long ones be the tenuous fraction. Two exponential approximations based on $1/\langle\sigma\rangle = 1/(\sigma_1 f_1 + \sigma_2 f_2) = 1$ and on the actual MFP $= \langle 1/\sigma\rangle = f_1/\sigma_1 + f_2/\sigma_2 = 25/9 = 2.77\ldots$ are also plotted. (b) The actual MFP is compared to the prediction $1/\langle\sigma\rangle$ based on mean extinction as $\sigma_2/\sigma_1$ increases from 1 to 10 and as the mixing ratio $f_2/f_1$ varies; the special case used in panels (a) and (c) is highlighted. (c) Statistical moments $\mathsf{E}(s^q)$ of the actual FPD are compared with the exponential prediction $\Gamma(q+1)\langle 1/\sigma\rangle^q$ based on the actual MFP; analytic continuation of the formula in (34) for real, not just integer, $q$ is used here. The under-estimation of moments at both higher-order ($q > 1$) and negative order ($q < 0$) is a result of the wider-than-exponential nature of the actual FPD. Note that, although their plotted ratio is finite, both moments are in fact divergent for $q \leqslant -1$.

### 1.2. Main results and outline

The paper is organized as follows. We review in Section 2 the related concepts of direct transmission and free-path distribution (FPD); in particular, we revisit Beer's law of exponential transmission for uniform and variable media.

In Section 3, we introduce spatial- and ensemble-averages in order to define the "mean-field" FPD which is relevant to bulk transport, at least under some conditions spelled out in Section 4. By recasting the photon propagation problem in terms borrowed from probability theory, we derive rigorously three fundamental properties of mean or "effective" FPDs under quite general conditions on the 3D disorder. They are:

I. Mean-field FPDs are never exponential in variable media and, conversely, exponential FPDs are obtained only in uniform media.
II. The actual MFP is always larger than that predicted by the average extinction.
III The effective FPD is always wider than an exponential law, even if it is based on the actual MFP.

The above properties I–III are illustrated in Figs. 1a–c respectively for a simple binary medium. On a more quantitative note, we show that the average of the inverse extinction is generally a better estimate of the actual MFP than the inverse of the average extinction. However, the one-parameter exponential FPD associated with Beer's law fails in a fundamental way to represent photon propa-

gation in variable media no matter what MFP it is based on; consequently, the MFP is insufficient to characterize photon propagation in general.

In Section 4, criteria based on 2-point statistical considerations are proposed for deciding when 3D effects leading to non-exponential FPDs are likely to be important, and spatial correlations prove to be the critical ingredient. Specifically, strong-enough variability is required on scales commensurate with the actual MFP, which may be quite large due to property "II." This is a form of *resonance* in the response of radiation flow to the spatial variability, but not in a narrow range of scales. In *much slower* variability, photons will sample only a rather narrow range of the possible extinction values and the non-exponential mean transmission law is less relevant than the local exponential one. This is the basis of a class of approximations in 3D radiative transfer. In *much faster* variability, photons will sample almost all possible extinction values between each emission/scattering/absorption event. The mean extinction (and the associated exponential transmission) will therefore dominate the bulk transport.

In Section 5, we draw an important but limited analogy between our results and classic work in long-path spectroscopy of gases at coarse resolution and we offer an in-depth discussion of closely related (and mostly quite recent) work . Finally, we summarize in Section 6 and outline a follow-on paper. The appendix establishes in sufficient statistical technicality the natural prevalence of the requirement of "1-point scale-independence," defined in Section 3 and invoked several times later on, in relation to spatial correlations.

## 2. Beer's law and photon free-path distributions (FPDs)

### 2.1. Optical free paths: exponential probability density function with unit mean

In the Introduction, we presented direct transmission as a *cumulative* probability distribution. From the cumulative probability in Eqs. (6)–(7) for the tail of the distribution, we obtain the probability density function (pdf) for steps of any optical distance $\tau \geqslant 0$:

$$p(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left[1 - \mathrm{Prob}\{\textit{optical step} \geqslant \tau\}\right] = \left|\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}\tau}\right| = \mathrm{e}^{-\tau}. \tag{9}$$

From there, we obtain the mathematical expectation for photon steps in units of optical distance,

$$\mathsf{E}(\tau) = \int_0^\infty \tau\,\mathrm{d}P(\tau) = \int_0^\infty \tau\,p(\tau)\,\mathrm{d}\tau = \int_0^\infty \tau\exp(-\tau)\,\mathrm{d}\tau = \Gamma(2) = 1, \tag{10}$$

i.e., the mean free *optical* path is always unity. We can also estimate optical step variance,

$$\mathsf{D}(\tau) = \mathsf{E}([\tau - \mathsf{E}(\tau)]^2) = \mathsf{E}(\tau^2) - \mathsf{E}(\tau)^2 = \Gamma(3) - \Gamma(2)^2 = 1, \tag{11}$$

it is also unity. In the above, we have used Euler's Gamma function,

$$\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)\,\mathrm{d}t \tag{12}$$

which reduces to $\Gamma(n) = (n-1)!$ for positive integer arguments; so we have $\mathsf{E}(\tau^q) = \Gamma(q+1)$.

### 2.2. Physical free paths, 1: the uniform case

Assume $\sigma(\boldsymbol{x}) \equiv constant$, hence $\tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s) \equiv \sigma s$ in Eq. (5), independent of $\boldsymbol{x}_0$ and of $\boldsymbol{\Omega}_0$. Following the same rule as in (9), we get

$$\mathrm{Prob}\{step \geqslant s\} = \mathscr{T}(s) = \mathrm{e}^{-\sigma s} \tag{13}$$

from (6)–(7), and then

$$p(s) = \left|\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}s}\right| = \left|\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}\tau}\right| \frac{\mathrm{d}\tau}{\mathrm{d}s} = \sigma \mathrm{e}^{-\sigma s} \tag{14}$$

for the free-path pdf, or FPD. From there, we derive the MFP

$$\ell = \mathsf{E}(s) = 1/\sigma \tag{15}$$

and step variance

$$\mathsf{D}(s) = \mathsf{E}(s)^2, \tag{16}$$

a relation we will refer back to as necessary to have an exponential law.

The theory presented so far is entirely captured in the procedure used routinely in Monte Carlo simulations [1] to approximate exponential deviates starting with the output of a pseudo-random number generator that approximates a uniform distribution on (0,1):

$$\begin{aligned} &\text{(i)}\ \ \tau = -\ln \xi \quad \text{with}\ \ \xi \in (0,1)\ \ \text{where}\ \ p(\xi) \equiv 1; \\ &\text{(ii)}\ \ s = \tau/\sigma \quad \text{if the medium is uniform.} \end{aligned} \tag{17}$$

### 2.3. Physical free paths, 2: the general case

Now consider the more interesting case where $\sigma(\boldsymbol{x}) \neq constant$; in the spirit of statistical physics, we will refer to this situation as "disorder." In 3D Monte Carlo coding, the 2nd step in (17) is replaced by a (generally iterative) solution of Eq. (5) for the unknown $s$, unless the "maximum cross-section" algorithm [1] is implemented.

We can still define a FPD,

$$\mathrm{d}P(s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0) = P(s + \mathrm{d}s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0) - P(s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0) = \mathrm{Prob}\{s \leqslant step < s + \mathrm{d}s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0\}, \tag{18}$$

and relate it to optical distance:

$$\begin{aligned} \mathrm{d}P(s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0) &= \left|\frac{\mathrm{d}\mathscr{T}}{\mathrm{d}s}\right| \mathrm{d}s = \mathscr{T}(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s)\sigma(\boldsymbol{x}_0 + \boldsymbol{\Omega}_0 s)\,\mathrm{d}s \\ &= \exp[-\tau(\boldsymbol{x}_0, \boldsymbol{\Omega}_0; s)]\sigma(\boldsymbol{x}_0 + \boldsymbol{\Omega}_0 s)\,\mathrm{d}s. \end{aligned} \tag{19}$$

From there, we can again define a MFP,

$$\ell(\boldsymbol{x}_0, \boldsymbol{\Omega}_0) = \mathsf{E}(s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0) = \int_0^\infty s\,\mathrm{d}P(s|\boldsymbol{x}_0, \boldsymbol{\Omega}_0). \tag{20}$$

and step variance,

$$\mathsf{D}(s|\boldsymbol{x}_0,\boldsymbol{\Omega}_0)=\int_0^\infty s^2\,\mathrm{d}P(s|\boldsymbol{x}_0,\boldsymbol{\Omega}_0)-\mathsf{E}(s|\boldsymbol{x}_0,\boldsymbol{\Omega}_0)^2. \tag{21}$$

However, these quantities are now inherently local and, furthermore, they depend on position $\boldsymbol{x}_0$ as well as on direction $\boldsymbol{\Omega}_0$ (even if the extinction $\sigma$ does not).

The local MFP in (20) is generally different from $1/\sigma(\boldsymbol{x})$ which at least has the dimension of a length. We will call this point-wise estimate of the MFP the local *pseudo*-MFP field, an estimated MFP that can become extremely large in regions of small extinction. The pseudo-MFP field $1/\sigma(\boldsymbol{x})$ plays an essential role in our statistical considerations further on.

## 3. Average properties of FPDs in variable optical media

### 3.1. Statistical considerations

#### 3.1.1. Spatial, spatial–angular, and/or ensemble averaging

Consider a field $f(\boldsymbol{x})$ in a domain $M$ of three-dimensional space. Its spatial average is defined simply by

$$\langle f\rangle=\int_{\boldsymbol{x}\in M} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big/\int_{\boldsymbol{x}\in M}\mathrm{d}\boldsymbol{x} \tag{22}$$

where the denominator is simply the volume of $M$. In atmospheric work, $M$ can be either a single cloud or cloud layer, broken or not, or it can be the whole column, including possible gaps between clouds and cloud layers.

Alternatively, one can first define its pdf and then compute the average; specifically,

$$\langle f\rangle=\int f\,\mathrm{d}P(f), \tag{23a}$$

where

$$\mathrm{d}P(f)=\mathrm{Prob}\{f\leqslant f(\boldsymbol{x})<f+\mathrm{d}f,\ \text{where } f \text{ (as a map from } M \text{ to } \mathbb{R}) \text{ is given, and } \boldsymbol{x} \text{ spans } M\}. \tag{23b}$$

The difference between the averaging methods described in Eq. (22) and (23a) and (23b) is the same as between Riemann's and Lebesgue's theories of integration in mathematical analysis.

This is all that can be done for a situation with "deterministic" variability, meaning in a single given optical medium. The formulation in (23a) and (23b) has the advantage of generalizing immediately to "ensemble" averaging. This means that $\mathrm{d}P(f)$ can be given a priori, without having to be tallied in $\boldsymbol{x}$-space using (23b). A defining aspect of ensemble averaging is that there are many realizations of the fields $f(\boldsymbol{x})$ to choose from and to average over in some (usually quite vast) functional space $\mathscr{B}$:

$$\mathrm{d}P(f)=\mathrm{Prob}\{f\leqslant f(\boldsymbol{x})<f+\mathrm{d}f,\quad \text{where } \boldsymbol{x}\in M \text{ is fixed and the map } \boldsymbol{x}\to f(\boldsymbol{x},\boldsymbol{\Omega}) \text{ spans } \mathscr{B}\}. \tag{23c}$$

So, using (23a) and (23c), one can now evaluate $\langle f(\boldsymbol{x})\rangle$ at any given $\boldsymbol{x}$ over these realizations and ask questions about their dependence on $\boldsymbol{x}$. For instance, if $\langle f(\boldsymbol{x})\rangle$ does not depend on $\boldsymbol{x}$, then $f$ is a candidate for *statistical* homogeneity, the spatial counterpart of "stationarity" in time-series analysis. It is only a candidate because in fact moments of all orders for all $n$-point statistics must be independent of $\boldsymbol{x}$ to ensure stationarity strictly speaking (i.e., the 2- and more-point statistics can only depend on differences $\boldsymbol{x}_2-\boldsymbol{x}_1$, etc.). Section 4 covers statistical homogeneity in more detail.

If the ensemble-average $\langle f(\boldsymbol{x})\rangle$ is independent of $\boldsymbol{x}$ and if its spatial counterpart $\langle f\rangle$ in (22) gives the same answer for every realization (and large enough $M$), then the statistical ensemble has an "ergodic" quality. Technically speaking, ergodicity means that increasingly large spatial averages converge in some sense to the ensemble value. Even without making explicit ergodic hypotheses, we will not need to distinguish between spatial- and ensemble-averages in most of the following. In fact, they can be combined at will.

Averaging may, if required, also be carried over directions $\boldsymbol{\Omega}$ in $\Xi$ (the unitary 3D sphere). Consider now another generic function $f(\boldsymbol{x},\boldsymbol{\Omega})$ in RT theory dependent on position and, this time, on direction too. It can be averaged directly over $M\otimes\Xi$, extending the recipe in Eq. (22), or using (23a) by first computing its pdf over $M\otimes\Xi$, extending the algorithm in Eq. (23c). Viewing $f(\boldsymbol{x},\boldsymbol{\Omega})$ as just one realization of a random function going from $M\otimes\Xi$ to $\mathbb{R}$, and thus generalizing (23c), $\langle f(\boldsymbol{x},\boldsymbol{\Omega})\rangle$ can be estimated as a function of the pair $(\boldsymbol{x},\boldsymbol{\Omega})$. If the result does not depend on $\boldsymbol{\Omega}$, then $f$ is a candidate for *statistical* isotropy.

### 3.1.2. Segment-averaging along beams at a given scale defined by step size s

Consider now the segment-averaged extinction that appears implicitly in Eq. (5). Dropping the "0" subscripts for simplicity in the following, we will use the notation

$$\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)=\tau(\boldsymbol{x},\boldsymbol{\Omega};s)/s=\frac{1}{s}\int_0^s\sigma(\boldsymbol{x}+\boldsymbol{\Omega}s')\,\mathrm{d}s'. \tag{24}$$

As long as the extinction field is finite, we have

$$\lim_{s\to 0}\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)=\frac{\partial}{\partial s}\tau(\boldsymbol{x},\boldsymbol{\Omega};s)|_{s=0}=\sigma(\boldsymbol{x}) \tag{25}$$

for all $(\boldsymbol{x},\boldsymbol{\Omega})$.

The (undistinguished) spatial- and ensemble-averages over the 3D disorder in (22) and (23) will be combined further on with the line-average in (24) as well as with the free-path averaging over an FPD from Section 2. Hence the need for three distinctive notations.

### 3.1.3. One-point scale-independence

We will also require further on that segment-averaged extinction in Eq. (24) be statistically similar to point-valued extinction over a significant range of scales, from 0 to some value $s_{\max}$. More precisely, we require (i) that the spatial/ensemble-average of the line-averaged field be the same, i.e.,

$$\langle\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)\rangle\equiv\langle\sigma(\boldsymbol{x})\rangle-\langle\sigma\rangle \tag{26a}$$

for $0\leqslant s\lesssim s_{\max}$, and (ii) that their centered moments differ at most by a small amount on the order of $s/s_{\max}$ when this ratio is small, i.e.,

$$\langle[\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)-\langle\sigma\rangle]^q\rangle/\langle[\sigma-\langle\sigma\rangle]^q\rangle-1=O(s/s_{\max}), \tag{26b}$$

over a significant range of orders, say, up to a value $q_{\max}$ not too small with respect to unity. This amounts to saying that the 1-point probability distributions of the random fields $\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)$ are similar for all but the most extreme deviations from the mean that naturally dominate the higher-order moments. We will refer to this property of the multi-dimensional extinction field as "1-point scale-independence." Homogeneous media are of course trivially 1-point scale-independent.

For instance, consider the pdf of $\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)$ in Eq. (24). If 1-point scale-independence in Eqs. (26a) and (26b) prevails, then $\mathrm{d}P(\bar{\sigma}|s)$ is independent of the averaging scale $s$ (over a given range), so it can be written $\mathrm{d}P(\bar{\sigma}|s) \equiv \mathrm{d}P(\bar{\sigma}|0) = \mathrm{d}P(\sigma)$. We will use this identification further on.

In Section 4 and in the technical appendix, 1-point scale-independence is related to the *statistical* isotropy and homogeneity of the random field. We show, in particular, that a certain degree of point-to-point correlation is needed to maintain the desired property. Incidentally, 1-point scale-independence is assumed in almost all instrument design since, for obvious signal-to-noise purposes, geophysical fields are not sampled at a point but over a finite duration, area, or volume. Fortunately, there are often long-range correlations in optical media, the cloudy atmosphere being just one example.

The property defined above as "1-point scale-independence" is not to be confused with "scale-invariance" which translates essentially to the systematic occurrence of power-laws in the spatial statistics, at least over a significant range of scales. An example of scale-invariance, in this case along with 1-point scale-independence, is if the right-hand side of Eq. (26b) goes as $(s/s_{\max})^{\alpha}$ with $\alpha > 0$. See appendix for more details.

### 3.1.4. Variance and Jensen's inequality

Consider a concave function $f(\xi)$, i.e., for which $f''(\xi) \leqslant 0$ on the support of an arbitrary pdf $\mathrm{d}P(\xi)/\mathrm{d}\xi$. Now average $f(\xi)$ over the randomness of $\xi$; we find

$$\int f(\xi)\,\mathrm{d}P(\xi) \leqslant f\left(\int \xi\,\mathrm{d}P(\xi)\right). \tag{27a}$$

This well-known inequality is generally traced to Jensen [2]. For a convex function $f(\xi)$ with $f''(\xi) \geqslant 0$, we naturally obtain

$$\int f(\xi)\,\mathrm{d}P(\xi) \geqslant f\left(\int \xi\,\mathrm{d}P(\xi)\right). \tag{27b}$$

In either case, "=" is reached in just two situations. Either $f(\xi)$ is linear (i.e., $f''(\xi) \equiv 0$) or $\xi$ is degenerate (i.e., $p(\xi) = \mathrm{d}P(\xi)/\mathrm{d}\xi = \delta(\xi - \xi^*)$, equivalently, the variance of $\xi$ is 0).

In the same way as we have defined step variance in (11), (16) or (21), we can define here the 1-point spatial/ensemble variance of the extinction field with a different notation to avoid any confusion:

$$\mathrm{var}(\sigma) = \langle[\sigma(\boldsymbol{x}) - \langle\sigma(\boldsymbol{x})\rangle]^2\rangle = \langle\sigma^2\rangle - \langle\sigma\rangle^2 \geqslant 0. \tag{28}$$

Here "=" is obtained only if the medium is uniform: $\sigma(\boldsymbol{x}) \equiv \textit{constant}$ hence $p(\sigma) = \delta(\sigma\text{-}\textit{constant})$. Although it is generally derived from Schwartz's inequality, the one in (28) is in fact a direct consequence of setting $f(\xi) = \xi^2$ in (27b).

### 3.2. Mean-field FPD, 1: importance in statistical (domain-average) transport theory

We presented the FPD in (19) as a function of $s$ that is parametrically dependent on $(\boldsymbol{x},\boldsymbol{\Omega})$. We can also look at it the other way around. Holding the step-size $s$ constant, we average over the independent space/angle variables of RT theory:

$$\langle P(s|\boldsymbol{x},\boldsymbol{\Omega})\rangle = \int_{\boldsymbol{x}\in M}\int_{\boldsymbol{\Omega}\in\Xi} P(s|\boldsymbol{x},\boldsymbol{\Omega})\,\mathrm{d}x\,\mathrm{d}\boldsymbol{\Omega}\Big/\left[4\pi\int_{x\in M}\mathrm{d}\boldsymbol{x}\right]. \tag{29}$$

So this is the averaging *of* a pdf rather than averaging *with* this pdf. Somewhat paradoxically, we are now looking at point/direction-wise FPDs as *random* functions of the special form expressed in (19). Using this expression, the above "mean-field" FPD becomes

$$\langle \mathrm{d}P(s|\boldsymbol{x},\boldsymbol{\Omega})\rangle = \left|\frac{\mathrm{d}}{\mathrm{d}s}\langle\mathscr{T}(\boldsymbol{x},\boldsymbol{\Omega};s)\rangle\right|\mathrm{d}s \tag{30}$$

that we will often denote just as $\langle \mathrm{d}P(s)\rangle$ since either the space/angle variables were used in the averaging, or we will be dealing with probability spaces of statistically homogeneous/isotropic fields.

We can now define *mixed* averages and these can be done in either order: first over the random FPDs then over the disorder, or vice versa. For instance, the MFP for the mean-field FPD (or "mean MFP"), written both ways, is

$$\langle \mathsf{E}(s)\rangle = \left\langle\int_0^\infty s\,\mathrm{d}P(s)\right\rangle = \int_0^\infty s\langle \mathrm{d}P(s)\rangle. \tag{31}$$

and the associated step-variance is

$$\langle \mathsf{D}(s)\rangle = \int_0^\infty s^2\langle \mathrm{d}P(s)\rangle - \langle \mathsf{E}(s)\rangle^2. \tag{32}$$

In the next subsection we invoke some lesser-known probability theory to show that:

I. In variable media, the mean FPD in Eq. (30) is never exponential. Since we already know that constant $\sigma$ leads to an exponential FPD (cf. Section 2.2), we can restate this simply as

> The mean FPD is exponential if and only if the medium is homogeneous.

II. The mean-field MFP in (30) is always larger than the MFP in the homogeneous medium of mean extinction $\langle\sigma\rangle$ or, equivalently, in a uniform medium that has the same total mass, since $\sigma = cross\_section \times particle\_density$, when the former quantity is constant. In short, we have

$$\langle\ell\rangle = \langle \mathsf{E}(s)\rangle \geqslant 1/\langle\sigma\rangle, \tag{33}$$

where "=" implies homogeneity. This means that, if mean particle density is used to predict the MFP, it will be under-estimated.

III. If the line-averaged extinction field in (24) is 1-point scale-independent, then

$$\langle \mathsf{E}(s^n)\rangle = n!\langle 1/\sigma^n\rangle. \tag{34}$$

The mean-field FPD for any such variable medium is therefore "wider" than the exponential FPD associated with the mean extinction $\langle\sigma\rangle$ in the following sense: exponential FPDs predict

$\mathsf{E}(s^n) = n!/\langle\sigma\rangle^n$ and $\langle\sigma^{-n}\rangle$ in (34) exceeds $\langle\sigma\rangle^{-n}$ for all $n > 0$ by Jensen's inequality (27b). In particular, we have

$$\langle \mathsf{D}(s)\rangle \geqslant \langle \mathsf{E}(s)\rangle^2 \tag{35}$$

in contrast with (16). If "=" applies here, then the FPD is then an exponential and, again, the extinction field is uniform (because of property "I"). This shows that, even if an "effective" (lower-than-mean) particle density is used in an exponential FPD with the correct MFP, all the higher-order moments of $s$ will be under-estimated.

Property III is illustrated in Fig. 1c where $\mathsf{E}(s^n)$ in (34) is compared to $n!\langle 1/\sigma\rangle^n$, the exponential prediction using the actual MFP $\langle 1/\sigma\rangle$ rather than the estimate $1/\langle\sigma\rangle$ based on the mean value. Notice from Eq. (34) that, as far as photon propagation in 1-point scale-independent media is concerned, the statistics of pseudo-MFP $1/\sigma$ are more important than those of extinction $\sigma$ itself. Finally, we note that 1-point scale-independence is not a strict requirement for obtaining wider-than-exponential transmission laws and FPDs, only a non-degenerate distribution of *optical* distance $\tau$ for a given *physical* step $s$ in Eq. (5), using (23b) to tally $\mathrm{d}P(\tau|s)$.

### 3.3. Mean-field FPD, 2: interpretation in probability theory

For the moment, we will hold the photon step $s$ constant, like a parameter, but $\tau(s)$ is still a non-negative random variable due to the disorder of the medium. Following the method proposed in (23a,b), we first define $P(\tau|s) = \mathrm{Prob}\{optical\ \ path \geqslant \tau(s)\}$ hence

$$\mathrm{d}P(\tau|s) = \mathrm{Prob}\left\{\tau(s) \leqslant \int_0^s \sigma(\boldsymbol{x} + \boldsymbol{\Omega}s')\,\mathrm{d}s' < \tau(s) + \mathrm{d}\tau(s);\, \boldsymbol{x} \in M, \boldsymbol{\Omega} \in \Xi\right\} \tag{36}$$

to describe the distribution of $\tau(s)$ for a given $s$. This leads to the pdf

$$p(\tau|s) = \mathrm{d}P(\tau|s)/\mathrm{d}\tau, \tag{37}$$

which is not to be confused with $p(\tau)$ in (9) for the exponential distribution of $\tau$ in an direct transmission experiment relating to the FPD in optical distance. The pdf in (37) depends only on the disorder and the definition of $\tau(\boldsymbol{x}, \boldsymbol{\Omega}; s)$ in (5). Also recall that the averaging operations using (9) lead to $\mathsf{E}(\cdot)$'s and $\mathsf{D}(\cdot)$'s while those using (37) lead to $\langle\cdot\rangle$'s and var[ · ]'s. For instance, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\tau(s)\rangle = \left\langle \frac{\partial}{\partial s}\,\tau(\boldsymbol{x}, \boldsymbol{\Omega}; s)\right\rangle = \langle\sigma(\boldsymbol{x} + \boldsymbol{\Omega}s)\rangle = \langle\sigma(\boldsymbol{x})\rangle = \langle\sigma\rangle \tag{38}$$

if the extinction field is statistically homogeneous, and therefore

$$\langle\tau(s)\rangle = \langle\sigma\rangle s. \tag{39}$$

Now consider the Laplace transform of the pdf in Eq. (37), also known as the "characteristic function" of $\tau(s)$:

$$\Phi(q|s) = \langle\exp[-q\tau(s)]\rangle = \int_0^\infty \exp[-q\tau(s)]\,\mathrm{d}P(\tau|s). \tag{40}$$

Some general properties and practical applications of Laplace characteristic functions $\Phi(q)$ of arbitrary non-negative random variables are [3]:

i. $\Phi(q)$ is positive, $\Phi(0)=1$, decreases monotonically, and $\lim_{q\to\infty}\Phi(q)=0$;
ii. to add independent random variables, one has to perform convolution products of their probability density functions, so we just need to multiply their respective $\Phi(q)$'s;
iii. compute successive moments (i.e., integrals) by Taylor expansion (i.e., derivatives)

$$\langle\tau(s)^n\rangle=\left[\left(-\frac{\partial}{\partial q}\right)^n\Phi(q|s)\right]_{q=0}, \tag{41}$$

where we have restored the step size parameter $s$ that is germane to FPD characterization. There is a natural generalization of Eq. (40) to real (hence possibly negative) random variables using the Fourier transform.

The "second" characteristic function or "cumulant-generating function," $\ln\Phi(q)$, is an equally powerful tool in probability theory. Its important properties and utilities are [3]:

a. $\ln\Phi(q)$ is large-sense convex, i.e., $[\ln\Phi(q)]''\geqslant 0$ (this is equivalent to *variance* $\geqslant 0$);
b. $\ln\Phi(q)$ is linear in $q$ if and only if the pdf is degenerate (i.e., *variance* $=0$ in property "a");
c. to add independent random variables, just add their respective $\ln\Phi(q)$ (this corresponds to $\Phi(q)$'s property "ii");
d. compute "cumulants" by successive derivatives

$$\left[\left(-\frac{\partial}{\partial q}\right)^n\ln\Phi(q|s)\right]_{q=0}=n\text{th-order cumulant of }\tau(s)=\begin{cases}n=0: & \int \mathrm{d}P=1,\\ n=1: & \text{mean }\langle\tau(s)\rangle,\\ n=2: & \mathrm{var}[\tau(s)],\end{cases} \tag{42}$$

and higher-order cumulants are related to skewness ($n=3$) and kurtosis ($n=4$). Note that, as in (41), we have restored here the dependence on the parameter $s$ (step size) for future reference.

Because of properties "c" and "d," the cumulants of a sum of independent random variables are sums of the individual cumulants, hence the name. Cumulant-generating functions for random variables on the whole real axis (such as normal deviates) are Fourier- rather than Laplace-transform based; they have the same properties "b"–"d" and for property "a" they are concave rather than convex. They are used extensively in random walk theory [3].

### 3.4. Mean-field FPD, 3: implications for bulk transport

From (7) and (40), we see that

$$\langle\mathscr{T}(s)^q\rangle=\Phi(q|s). \tag{43}$$

Fig. 2 summarizes graphically the general properties of $\langle\mathscr{T}(s)^q\rangle$, as a Laplace characteristic function of the random variable $\tau(s)$ defined in Eq. (5). If, moreover, $\sigma(\boldsymbol{x})$ is 1-point scale-independent, then we get $\langle\mathscr{T}(s)^q\rangle\approx\langle\exp(-\bar{\sigma}sq)\rangle$ from (23)–(26), hence $\Phi(q|s)\equiv\Phi(qs)$ in (43), at least for the

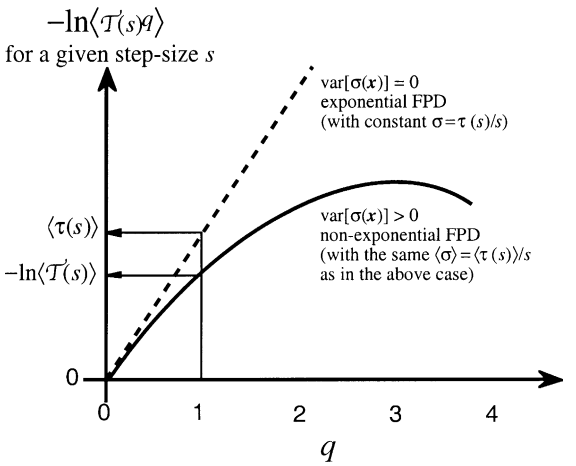


Fig. 2. Cumulant generating function for optical paths associated with a given step size. This schematic shows the (negative of the) cumulant generating function $-\ln \Phi(q|s)$ from (40), equivalently $-\ln\langle \mathscr{T}(s)^q\rangle$ from (43). The independent variable $q$ is the Laplace conjugate of $\tau(s)$, equivalently, the (not necessarily integer) order of the statistical moment of $\mathscr{T}(s)$. The case of an exponential FPD (homogeneous extinction) is in dashes and the more realistic variable extinction case is the continuous line (non-linear in $q$). The inequality highlighted at $q=1$ is proven and interpreted in the main text. Still using $q=1$, in the range of scales where 1-point scale-independence prevails, the abscissa can be re-assigned to the step value $s$ since $\Phi(q|s) \equiv \Phi(qs)$. In this case, the illustrated decrease with respect to $s$ is not truly physical because it can be traced to negative deviations of the extinction field.

designated range of $q$ and $s$. So the horizontal axis Fig. 2 can be taken as the physical distance $s$ where the mean transmission is measured.

In Fig. 2, we also highlight a remarkable inequality at $q=1$, namely, $-\ln\langle \mathscr{T}(s)\rangle \leqslant \langle \tau(s)\rangle$, equivalently,

$$\langle \mathscr{T}(s)\rangle \geqslant \exp[-\langle \tau(s)\rangle] = \exp[-\langle \sigma\rangle s] \tag{44}$$

using (39). This is a direct consequence of Jensen's inequality (27b) for a convex function. In the case of direct transmission, $f$ has the convexity of a decreasing exponential on the positive real axis and the random variable is $\tau(s)$. Conversely, the distance $s$ at which a given mean transmission $\langle \mathscr{T}(s)\rangle$ is reached is minimum for the sure case where $\langle \tau(s)\rangle = \langle \sigma\rangle s$.

From (30) and (43), the corresponding FPD is obtained at $q=1$:

$$p(s) = \frac{\mathrm{d}}{\mathrm{d}s}\langle P(s)\rangle = \left|\frac{\partial}{\partial s}\Phi(1|s)\right|. \tag{45}$$

So mean-field FPDs inherit many properties from this connection with characteristic functions and we are now in a position to prove statements I–III in Section 3.2 and derive the associated inequalities in (33)–(35):

I. The premise of property "b" of $\ln \Phi(q|s)$—that it is linear in $q$—can be restated as "$\langle \mathscr{T}(s)\rangle$ is exponential in $-\langle \tau(s)\rangle$," hence the associated FPD as well. We already know this to be true if $\sigma$ is constant (cf. Section 2.2). We now have the reciprocal statement: an exponential $\langle \mathscr{T}(s)\rangle$

and/or FPD implies $\mathrm{var}[\tau(s)]=0$, for any $s$. Now, if $\tau(\boldsymbol{x},\boldsymbol{\Omega};s)$ is a sure function of $s$ for any $\boldsymbol{x}$ and $\boldsymbol{\Omega}$, then so is $\partial\tau/\partial s|_{s=0}=\sigma(\boldsymbol{x})$, in other words, $\mathrm{var}[\sigma(\boldsymbol{x})]=0$ (homogeneity).

II. From Eq. (42), we can write the Taylor expansion of the cumulant generating function as

$$\ln \Phi(q|s) = -\langle\tau(s)\rangle q + \frac{1}{2}\,\mathrm{var}[\tau(s)]q^2 + \textit{higher-order terms}. \tag{46}$$

Because of property "a," we know that the second-order term and all the higher-order terms cannot make $\ln\Phi(q|s)$ less than its first-order term, similarly for $\Phi(q|s)$. Setting $q=1$ in (46) and using (31), and (45), we see that the mean MFP

$$\langle\ell\rangle = \langle \mathsf{E}(s)\rangle = \int_0^\infty s\left|\frac{\partial\Phi}{\partial s}\right|_{q=1} \mathrm{d}s = \int_0^\infty s\left|\frac{\mathrm{d}}{\mathrm{d}s}\,\mathrm{e}^{\ln\Phi(1|s)}\right|\mathrm{d}s \tag{47}$$

is always greater or equal to

$$\int_0^\infty s\left|\frac{\mathrm{d}}{\mathrm{d}s}\,\langle\tau(s)\rangle\right|\mathrm{e}^{-\langle\tau(s)\rangle}\,\mathrm{d}s = \frac{1}{\langle\sigma\rangle}. \tag{48}$$

If it is equal, then the second-order term and all higher-order terms in (46) must vanish, so we retrieve the case of constant extinction. Another proof uses (30)–(31) on both sides of the inequality in (44).

III. We now consider free-path moments of all orders for the mean-field FPD:

$$\langle \mathsf{E}(s^n)\rangle = \int_0^\infty s^n\langle \mathrm{d}P(s)\rangle. \tag{49}$$

From (45), or by averaging (19) with (24), we have

$$\langle \mathrm{d}P(s)\rangle/\mathrm{d}s = \langle \exp[-\tau(s)]\,\mathrm{d}\tau(s)/\mathrm{d}s\rangle = \langle \exp[-\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)s]\sigma(\boldsymbol{x}+\boldsymbol{\Omega}s)\rangle. \tag{50}$$

So the mean-field FPD is a cross-correlation of $\bar{\sigma}$ at two different scales ($s>0$ and $=0$) and two different positions ($\boldsymbol{x}$ and $\boldsymbol{x}'=\boldsymbol{x}+\boldsymbol{\Omega}s$). We therefore invoke statistical homogeneity (a.k.a. stationarity) and 1-point scale-independence to write $\langle \mathrm{d}P(s)\rangle/\mathrm{d}s=\langle\exp[-\sigma s]\sigma\rangle$. The mean-field FDP then becomes

$$\langle \mathrm{d}P(s)/\mathrm{d}s\rangle = \langle\sigma\exp[-\sigma s]\rangle = \int \sigma\exp[-\sigma s]\,\mathrm{d}P(\sigma). \tag{51}$$

We have deliberately left out the bounds on the averaging integral because the support of $\mathrm{d}P(\sigma)$ need not be specified—it can be part or all of the positive real axis. The easiest way of evaluating the moments in (49) is to first compute the characteristic function of $s$:

$$\Phi(u) = \langle \mathsf{E}(\mathrm{e}^{-us})\rangle = \int_0^\infty \mathrm{e}^{-us}\langle \mathrm{d}P(s)\rangle = \int_0^\infty \mathrm{e}^{-us}\,\mathrm{d}s\int \sigma\exp[-\sigma s]\,\mathrm{d}P(\sigma). \tag{52}$$

Reversing the order of the integrals, we find

$$\begin{aligned}\Phi(u) &= \int \sigma\,\mathrm{d}P(\sigma)\int_0^\infty \exp[-\sigma s - us]\,\mathrm{d}s \\ &= \int [1+u/\sigma]^{-1}\,\mathrm{d}P(\sigma) = \sum_{n=0}^\infty (-u)^n\int\frac{1}{\sigma^n}\,\mathrm{d}P(\sigma).\end{aligned} \tag{53}$$

The result in (34) then follows directly from property "iii" of $\Phi(u)$, i.e., Eq. (41) but with different random and sure variables (specifically, $\tau$ and $q$ are mapped to $s$ and $u$ respectively).

Specifically, we identify term-by-term the expression in (53) with the Taylor expansion of $\Phi(u) = \Sigma\Phi^{(n)}(0)u^n/n! = \Sigma\langle s^n\rangle(-u)^n/n!$ from (41).

We can now show that the mean-field FDP in (51) is wider-than-exponential because, for all integer $n \geqslant 0$, we have

$$\langle \mathsf{E}(s^n)\rangle = n!\langle 1/\sigma^n\rangle \geqslant n!\langle 1/\sigma\rangle^n = \int_0^\infty s^n \exp[-s/\langle 1/\sigma\rangle]\,\mathrm{d}s/\langle 1/\sigma\rangle \tag{54}$$

where we start by restating the result in (34). The last expression is just $\mathsf{E}(s^n)$ for the exponential FPD associated with the MFP $\langle 1/\sigma\rangle$ rather than mean extinction $\langle\sigma\rangle$; to see this, use the definition in (12) for integer arguments. The inequality in (54) follows directly from Jensen's in (27b) as it applies to the convex function $f(\xi) = \xi^n$, $n > 1$ (or $n < 0$), and if "=" is obtained for any $n \neq 0, 1$ this implies $\sigma \equiv constant$. Finally, the conjectured inequality in (35) is equivalent to

$$\langle \mathsf{E}(s^2)\rangle \geqslant 2\langle \mathsf{E}(s)\rangle^2 \tag{55}$$

which in turn follows from setting $n = 1, 2$ in (54): $\langle \mathsf{E}(s^2)\rangle = 2!\langle 1/\sigma^2\rangle \geqslant 2!\langle 1/\sigma^2\rangle = 2\langle \mathsf{E}(s)\rangle^2$, and again "=" occurs if and only if $\sigma$ is constant.

Preliminary versions of these derivations are given in one of our doctoral theses [4].

We have opted to introduce wider-than-exponential transmission laws and associated FPDs using 1-point scale-independent media because they are a natural choice in astrophysical and geophysical applications where highly turbulent flows shape the optical variability, and thus we emphasize the key role of spatial correlations. However, this assumption is not necessary from the purely mathematical perspective. As first noticed by Borovoi [5], all that is needed is a non-trivial distribution of $\tau$ for given $s$, $p(\tau|s)$, to perform the ensemble average of $\exp[-\tau]$. For instance, Lovejoy et al. [6] show that multiplicative cascades, an interesting class of spatially-correlated scale-invariant media that are patently not 1-point scale-independent, lead to transmission laws in the mean that are wider-than-exponential; see also Knyazikhin et al. [7] in an application stemming from the optics of vegetation canopies.

## 4. What kind of variability is conducive to very non-exponential FPDs?

### 4.1. The critical issue of spatial correlations

In 3D RT, numerical techniques included, we are primarily concerned with the variability of $\bar{\sigma}(\boldsymbol{x}, \boldsymbol{\Omega}; s)$, the line-average extinction estimated along the segment from $\boldsymbol{x}$ to $\boldsymbol{x} + \boldsymbol{\Omega}s$. It fluctuates around its spatial mean $\langle\bar{\sigma}\rangle(s)$ carried over the independent space-angle variables $(\boldsymbol{x}, \boldsymbol{\Omega})$ of the radiative transfer equation. Equivalently, we are looking at the variability of the optical distance $\tau(\boldsymbol{x}, \boldsymbol{\Omega}; s) = \bar{\sigma}(\boldsymbol{x}, \boldsymbol{\Omega}; s)s$ between any two points at distance $s$ around $\langle\tau\rangle(s)$ which may or may not be linear in $s$. It is not hard to anticipate that when things get interesting from the standpoint of 3D RT the 1-point variability of $\sigma(\boldsymbol{x})$, as captured in its pdf, is not the only statistical ingredient. How $\sigma(\boldsymbol{x})$ changes as a function of $\boldsymbol{x}$ is also important. Simply put in statistical parlance, spatial correlations matter.

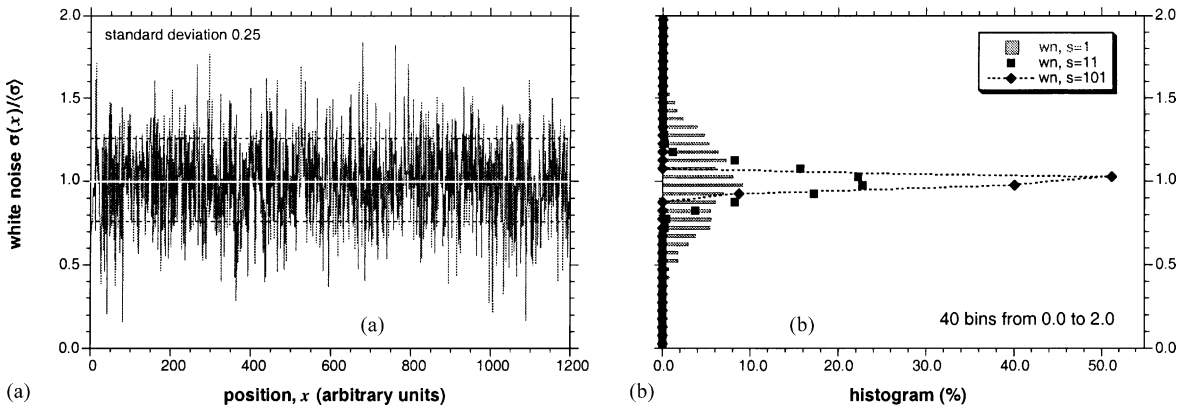


Fig. 3. Gaussian white noise: an extinction field that fails the 1-point scale-independence test. (a) Trace of 1200 samples of independent normally distributed random variables with unit mean and variance 1/16 (mean and $\pm 1$ standard deviation are marked). (b) Probability density functions (normalized histograms) of the extinction field in panel (a) averaged over 1, 11 and 101 grid-points; notice the significant narrowing of the histograms as averaging length increases in contradiction with (26b).

A key assumption in some of the arguments of Section 3 is the 1-point scale-independence of the extinction field implying, in particular, that $\langle\bar{\sigma}\rangle(s)$ is independent of $s$ and therefore that $\langle\tau\rangle$ is proportional to $s$. This is a reasonable assumption but needs closer examination from the standpoint of spatial correlations and of *statistical* homogeneity.

### 4.2. White noise as a counter-example

It is easy to conjure up a situation where $\sigma(x)$ can be almost arbitrarily variable—we will only assume it has finite variance—but the associated variability of $\sigma(\boldsymbol{x},\boldsymbol{\Omega};s)$ is quite trivial: a field of "$\delta$-correlated" random variables or spatially "white" noise (cf. Fig. 3 for an example using Gaussian deviates for specificity). In this case, we know from elementary probability theory (essentially 1D random walk theory) that the variance of $\tau(\boldsymbol{x},\boldsymbol{\Omega};s)$, i.e., the running integral of $\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)$ in Eq. (5), will increase as $s/\ell$, as does its mean $\langle\tau\rangle$. Here $\ell$ denotes the smallest ("pixel") scale of interest, that is, where the variability starts. Thus $\mathrm{var}[\bar{\sigma}](s)$, measuring deviations from the mean $\langle\bar{\sigma}\rangle$ in samples of length $s$, decreases as $\ell/s$ (see technical appendix for a justification of this trend from first principles). So the 1-point scale-independence requirement in (26a) for the mean holds, but not (26b) for variance and other centered moments.

What happens when, at large-enough $s$, the fluctuations of $\bar{\sigma}(\boldsymbol{x},\boldsymbol{\Omega};s)$ around its mean $\langle\bar{\sigma}\rangle$ become small with respect to this mean? Now, if this "large-enough" $s$ is in fact still quite small compared to the MFP, we are converging towards an effectively homogeneous situation. So an exponential Beer's law applies at all but the smallest scales. Indeed, for Gaussian white noise with mean $\langle\sigma\rangle$ and variance $\mathrm{var}[\sigma]$ at the pixel scale $\ell$, a little algebra leads to:

- $\langle\mathscr{T}(s)\rangle \approx \exp[-\langle\sigma\rangle s + \mathrm{var}[\sigma]s^2/2]$ for $s \ll \ell$ (assuming $\mathrm{var}[\sigma]/\langle\sigma\rangle^2 \ll 1$);
- $\langle\mathscr{T}(s)\rangle \approx \exp[-(\langle\sigma\rangle - \mathrm{var}[\sigma]\ell/2)s]$ for $s \gg \ell$ (assuming only that $\mathrm{var}[\sigma]/\langle\sigma\rangle^2 \ll s/\ell$).

The corresponding estimate of the actual MFP is $1/(\langle\sigma\rangle - \mathrm{var}[\sigma]\ell/2)$, from the second and more spatially dominant exponential trend. As predicted, it is larger than $1/\langle\sigma\rangle$, the estimate based on the mean extinction; however, the applicable correction, $\mathrm{var}[\sigma]\ell/2\langle\sigma\rangle = \mathrm{var}[\sigma]/\langle\sigma\rangle^2 \times (\langle\sigma\rangle\ell)/2$, is necessarily quite small since the pixels are assumed optically thin from the start (i.e., $\langle\sigma\rangle\ell \ll 1$). Consequently, 1D RT theory becomes accurate if mean extinction is used in the asymptotic limit ($\ell \to 0$ for any given $\langle\sigma\rangle$ and $\mathrm{var}[\sigma]$) and a small correction to the mean extinction is sufficient for finite (but optically thin) pixels. In short, the variability of $\sigma(\boldsymbol{x})$ is too "fast" to generate interesting (i.e., large) 3D effects at most scales.

Rapid decorrelation is related to the notions of statistical stationarity (in the time domain) and statistical homogeneity (in the spatial domain), which is not to be confused with the meanings of constancy or uniformity often assigned to the term homogeneity—this paper included. Statistical stationarity/homogeneity means that statistical quantities (moments, probability distributions, etc.) do not depend on when/where they are collected. Implicitly, this assumes we are talking about a probabilistic ensemble of different realizations of the random process/field. White noise is stationary/homogeneous by definition. Beyond that, this classification is purely academic because, in operational statistics, all the spatial information is consumed in obtaining the best possible estimate of the average, cf. Eq. (22). However, one can compare averages obtained over $r$-sized portions of the data and attempt a statement about stationarity/homogeneity at scale $r$, varying from the sampling scale $\ell$ to a much larger value. If there is evidence for the existence of a minimum scale $R$ at which data appear to be stationary/homogeneous, it is like saying the data are decorrelated for $r \gtrsim R$ and that $R$ is the "correlation" scale which, depending on the point of view, is sometimes called the "*de*correlation" scale. This characteristic scale $R$ will likely depend on the choice of statistic used in the procedure. A standard definition, the "integral" (de)correlation scale, is based on the auto-correlation function, namely,

$$\rho(r) = \langle [\, f(\boldsymbol{x}+\boldsymbol{r}) - \langle f\rangle][\, f(\boldsymbol{x}) - \langle f\rangle]\rangle = \langle f(\boldsymbol{x}+\boldsymbol{r}) f(\boldsymbol{x})\rangle - \langle f\rangle^2, \tag{56}$$

where $\boldsymbol{r}$ is an arbitrary displacement vector of length $r = \|\boldsymbol{r}\|$. The random field $f(\boldsymbol{x})$ in (56) is assumed statistically homogeneous, hence the lack of dependence on $\boldsymbol{x}$ in $\rho(r)$, and isotropic, hence the lack of dependence on displacement direction $\boldsymbol{r}/r$ in $\rho(r)$. The definition of $R$ then reads as [8]

$$R = \frac{1}{\rho(0)} \int_0^\infty \rho(r)\, \mathrm{d}r \tag{57}$$

assuming $\rho(r)$ decreases fast enough for this integral to converge. One can thus distinguish between non-stationary $r \lesssim R$ and stationary $r \gtrsim R$ regimes. In technical jargon, this is the criterion for "broad-sense" stationarity/homogeneity. For white noise, $\rho_{\mathrm{wn}}(r) \propto \delta(r)$, so $R_{\mathrm{wn}} = 0$: it is (broad-sense) stationary at *all* scales.

In "stochastic" RT theory for binary mixtures [9], the media of interest have a Bernoulli pdf in extinction (hence two variability parameters beyond the mean, as used in Fig. 1) and one coupling constant akin to the integral scale in (57) that describes the spatial frequency of extinction jumps. Coupled RT equations for the domain/ensemble-average radiances are then established and solved numerically by any of a number of techniques (Monte Carlo, discrete ordinates, etc.) When the spatial frequency increases without bound ($R_\sigma \to 0$)—this is known as the "atomic" mixture—the model collapses onto the homogeneous situation. This is exactly what we have found above for

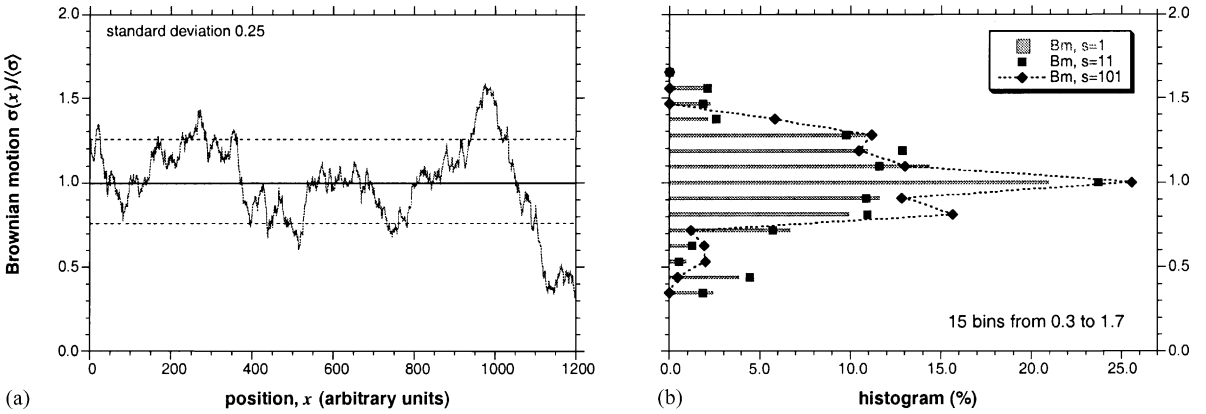


Fig. 4. Brownian motion (a.k.a. Wiener–Levy process): an extinction field that is 1-point scale-independent. (a) Trace of 1200 samples of Brownian motion which is Gaussian by definition with unit mean and variance 1/16 (mean and $\pm 1$ standard deviation are marked). The plot was generated by affine transformation of the running sum of the white noise in Fig. 3 (reduced to a zero mean); note the fractal scale-invariance of the graph (each portion looks like the whole). (b) Probability density functions of the field in panel (a) averaged over 1, 11 and 101 grid-points; note the constant width of the histograms at all averaging lengths in agreement with Eq. (26b).

white noise with thin pixels, reckoning only on the properties of $\langle \mathscr{T}(s) \rangle$ hence of the mean-field FPD (which are quasi-exponential).

### 4.3. Wiener–Levy process and multiplicative cascades as examples

Now consider a field $\sigma(\boldsymbol{x})$ of non-negative 3D "Brownian landscape" with an outer scale $L$ where the graph of any 1D transect looks like a random walk with some (positive) mean value $\langle \sigma \rangle$ (cf. Fig. 4). This is a spatial counterpart of the well-known Wiener–Levy process [3,8]. There is of course a constraint requiring $\sigma(\boldsymbol{x})$ to remain non-negative which puts an upper bound on the overall variance in relation to $\langle \sigma \rangle$. By construction, Wiener–Levy processes—*cumulative* sums of Gaussian white noise—are non-stationary (in the above sense) when they start from a sure value (usually taken to be 0). However, they do have stationary increments—*finite* sums of the white noise.

Here, running averages over a length $s$ will fluctuate wildly, just like the field itself, until $s$ becomes close to $L$; consequently, their mean and other moments will not depend on $s$ under the same condition, especially if many realizations are used. In the definition using (56), this model is nonstationary at all scales up to $R_{\mathrm{WL}} \approx L$. In absence of complications at boundaries, the impact of this relatively "slow" variability on photon propagation will depend (i) on the overall variance with respect to the mean and (ii) on the mean optical distance across the entire field, namely $\langle \sigma \rangle L$.

A direct application of 1-point scale independence in Eqs. (26a) and (26b) is the substitution of point-wise values by local averages. This is now done routinely when going from so-called "structure functions" to "wavelet spectra" [10]. In structure function analysis, only the fields of $r$-scale increments $f(\boldsymbol{x}+\boldsymbol{r}) - f(\boldsymbol{x})$ are required to be statistically stationary/homogeneous and the absolute non-negative moments of this 2-point difference in field values at distance $r$ are often

estimated, namely, the family of $q$th-order structure functions

$$S_f(q,r) = \langle |\, f(\boldsymbol{x}+\boldsymbol{r}) - f(\boldsymbol{x})|^q \rangle \tag{58a}$$

that surely yields $S_f(q,0) = 0$ and increases with $r = \|\boldsymbol{r}\| \geqslant 0$ as soon as $f(\boldsymbol{x})$ has at least a small degree of continuity (a.k.a. Hölder regularity). If the field $f(\boldsymbol{x})$ is scale-invariant (or, at least, has a "scaling" regime), the spatial statistics are power-law in $r$:

$$S_f(q,r) \propto r^{\zeta(q)}. \tag{58b}$$

Then the exponents $\zeta(q) \geqslant 0$ for $q \geqslant 0$ are the same for structure functions and for wavelets, even if higher-order moments are considered, as long as the signal is nonstationary with stationary increments [10]. Being Gaussian, the Wiener–Levy process in Fig. 4 leads to $\zeta(q) = q/2$ since the $q$th-order moment goes as that power of the standard deviation with a proportionality constant that involves a ratio of Euler Gamma functions. However, the increments in (58a) need not be independent, nor Gaussian. By comparison of (40) with (58a) and (58b) and assuming that the proportionality constants ("prefactors") in the latter do not depend much on $q$, we see that $\zeta(q)\ln r$ is related to the cumulant-generating function $\ln \Phi(q)$ of the random variables $\ln(|f(\boldsymbol{x}+\boldsymbol{r}) - f(\boldsymbol{x})|)$. Thus $\zeta(q)$ is necessarily large-sense concave in $q$. For case studies of $\zeta(q)$ inside terrestrial clouds that demonstrate this universal property, we refer to Marshak et al. [11].

In the special case of the second-order structure function, there is another way of writing increment variance:

$$S_f(2,r) = 2[\rho_f(0) - \rho_f(r)] \propto r^{\zeta(2)} \quad \text{as } r \to 0, \tag{59a}$$

but only if $f$ is itself broad-sense stationary/homogeneous, by using (56). In the appendix, we show that any field with $\zeta(2) = 2H > 0$ in (59a) is non-trivially 1-point scale independent, with the right-hand side of (26b) being $\propto (s/s_{\max})^{\alpha}$ with $\alpha = 2H$, and therefore spawn wider-than-exponential transmission laws. In terrestrial clouds, the value of $H = \zeta(2)/2$ for liquid water content, a good proxy for extinction, is found to hover around $1/3$ in both remotely measured time-series of the vertical integral [12] and in aircraft probings [13]. So the internal cloud variability is, as can be expected, 1-point scale independent and their optical properties are conditioned by the general findings in the present study.

However, we must emphasize at this point that, even though it is arguably the most prevailing property in natural media, 1-point scale independence is not a strict requirement to obtain non-exponential transmission laws. The evidence points towards the need for a strictly positive auto-correlation function at scales $r > 0$. Indeed, among all possible random scale-invariant field models many will have $\zeta(q) \equiv 0$ because they are stationary at all scales, but they also have

$$\rho(r) \propto 1/r^{\mu} \quad (0 < \mu < 1). \tag{59b}$$

These can also lead to sub-exponential transmission laws as was demonstrated independently by Lovejoy et al. [6] and Knyazikhin et al. [7] for specific multiplicative cascades models.

### 4.4. Given enough amplitude of variation, how much correlation is required?

The above examples point to the relation of $R_{\bar{\sigma}}$ (integral scale for random line-averages $\bar{\sigma}$ of extinction) to the MFP (the relevant scale for the photon transport in absence of boundaries) as the

key question:

- "Too fast for interesting 3D effects" is the outcome of our analysis for Gaussian-type white noise fields and this translates statistically to "decorrelation over a small scale." Here "small" is clearly with respect to the MFP which we can take as $1/\langle\bar{\sigma}\rangle$ for simplicity (in spite of the bias in this estimate). So we have $\langle\bar{\sigma}\rangle R_{\bar{\sigma}} \ll 1$ in this situation: the optical distance across a decorrelation scale is typically quite small. In this case, we argued that transport is determined by $\langle\bar{\sigma}\rangle$ and the level of variability is not very important (at least if its variance is finite).
- "More interesting 3D effects" were predicted in our analysis of Wiener–Levy fields because of its "long-range correlations," hence relatively slow variability. In this case, we at least need to have $\langle\bar{\sigma}\rangle R_{\bar{\sigma}} \gtrsim 1$ and possibly even $R_{\bar{\sigma}}/\langle 1/\bar{\sigma}\rangle \gtrsim 1$: a typical decorrelation distance is optically thick. Furthermore, we want the overall variance of $\bar{\sigma}$ to be comparable to its mean $\langle\bar{\sigma}\rangle$; this will make the positive difference between $\langle 1/\bar{\sigma}\rangle$ and $1/\langle\bar{\sigma}\rangle$ more significant.

From the above theoretical examples of randomly variable optical media, we clearly need (1) significant deviations with respect to mean properties ("strong enough" variability) and (2) correlations on the scale of the MFP ("slow enough" variability) to obtain significant deviations from Beer's law. From there, classic 1D transport theory for slab media will fail.

### *4.5. Break-down of Beer's law and the on-set of strong 3D effects*

#### *4.5.1. Sure and smooth case*

If the medium of interest is given (deterministically) in every detail and is smooth (has finite gradients everywhere), then is not hard to define a ratio of scales that captures both criteria we isolated above. It tells us how "fast" or "slow" the optical variability is in comparison with the relevant scale in RT, namely, the local pseudo-MFP $\sigma(\boldsymbol{x})^{-1}$. Indeed, the variability itself gives us a local length-scale, namely, $\sigma(\boldsymbol{x})/|\boldsymbol{\Omega}\cdot\nabla\sigma| = |\boldsymbol{\Omega}\cdot\nabla\ln\sigma|^{-1}$, which is the nominal distance needed for $\sigma(\boldsymbol{x})$ to change by once its own value in a displacement starting at $\boldsymbol{x}$ along the direction $\boldsymbol{\Omega}$. The relevant ratio is therefore

$$\kappa_{\mathrm{det}}(\boldsymbol{x}) = \frac{\sigma(\boldsymbol{x})^{-1}}{\sigma(\boldsymbol{x})/|\boldsymbol{\Omega}\cdot\nabla\sigma|} = \left|\boldsymbol{\Omega}\cdot\nabla\left(\frac{1}{\sigma}\right)\right|, \tag{60}$$

where the subscript "det" refers to the deterministic nature of the variability.

So we are looking at the (dimensionless) norm of the gradient of local pseudo-MFP $1/\sigma(\boldsymbol{x})$ along the beam of interest. Equivalently, this is a rough estimate of the optical distance to the point where $\sigma(\boldsymbol{x})$ has changed significantly. Considering a spatial average or representative sample of some kind, we distinguish three regimes for $\kappa_{\mathrm{det}}(\boldsymbol{x})$ where we can anticipate quite different radiative behaviors.

- $\kappa_{\mathrm{det}} \gg 1$, "fast" variability: At almost every free path, photons sample almost all the variability, so $\bar{\sigma} \approx \langle\sigma\rangle$. Consequently, Beer's exponential law applies everywhere to a good approximation, using the mean extinction $\langle\sigma\rangle$. Therefore, and somewhat paradoxically, plane–parallel theory for uniform media will apply for the multiple scattering to a reasonable approximation.
- $\kappa_{\mathrm{det}} \approx 1$, "resonant" variability: In this case, $\sigma(\boldsymbol{x})$ changes significantly over typical photon free paths. So strong deviations from exponential direct transmission will occur and 1D plane–parallel theory fails at the fundamental level.

- $\kappa_{\rm det} \ll 1$, "slow" variability: Most photons sample essentially only one value of the extinction, so a certain Beer's law applies but just locally. Here, one can consider applying standard 1D theory locally and averaging over the outcome.

In atmospheric RT, this last procedure is called the independent pixel/column approximation, "IPA" [14] or "ICA" [15]. A medium produced by stacking a sequence of plane–parallel slabs one over another is the optical analog of a montage of electrical resistors wired in series. Here, a response (such as direct or total transmission) to a fixed potential, or the (e.g., solar) source of irradiance, must cross all the elements; so impedances, or optical depths, have to be summed. Similarly, the ICA/IPA assembly is analogous to a circuit mounted in parallel. Here, all the elements are subjected to the same potential, or (solar) irradiance; so the responses (e.g., transmission as a function of variable optical depth) are added. In short, slow variability can lead to a highly non-exponential FPD by spatial averaging over the 1-point pdf but it may not be as relevant to the 3D problem as a spatial averaging of the radiative responses (such as total transmission, reflection or absorption) computed individually with exponential transport kernels.

In "stochastic" RT for binary media, briefly described above and in Ref. [9] for the details, there is an essentially deterministic limit associated with the spatial frequency of the extinction jumps going to zero (i.e., $R_\sigma \to \infty$). In this case, the two RT equations decouple and the mean radiance field is the weighted sum of the individual solutions. This is simply the ICA/IPA for a Bernoulli pdf in optical depth, exactly in the manner of Fig. 1 but for diffuse as well as direct transmission.

In their survey of 3D RT effects based on photon diffusion theory in conservatively scattering media, Davis and Marshak [16] propose a criterion for strong 3D effects that is very similar to the above interpretation of the ratio of scales in (60). Their ratio is in fact identical to $\kappa_{\rm det}(\boldsymbol{x})$ in (60) apart from being based on the "transport" extinction $(1-g)\sigma(\boldsymbol{x})$, where $g$ is the asymmetry factor of the scattering phase function, and on a directional gradient which is more relevant to their PDE-based formulation of 3D RT. In essence, Davis and Marshak were concerned in Ref. [16] with transport by a mean radiation flow (via multiple scattering) channeled through the 3D medium while here we are concerned with ballistic photon transport (between scatterings) inside the 3D medium.

#### 4.5.2. *Stochastic and/or non-differentiable cases*

Returning to random optical media, derivatives often do not exist. We can nonetheless generalize the scale ratio in (60) to this situation by replacing $\sigma(\boldsymbol{x})^{-1}$ by its spatial/ensemble average, and the directional gradient magnitude $|\boldsymbol{\Omega}\cdot\nabla\ln\sigma|$ by $S_{\ln\sigma}(1,r)/r$ using the definition in (58a). We arrive at the scale-dependent ratio of scales

$$\kappa_{\rm ran}(r) = \frac{\langle 1/\sigma\rangle}{[S_{\ln\sigma}(1,r)/r]^{-1}} = \frac{\langle 1/\sigma\rangle}{r}\, S_{\ln\sigma}(1,r), \tag{61}$$

where $S_{\ln\sigma}(1,r)$ contains the key information about variability magnitude and speed. If the medium is scaling with long-range correlations then, using (58b) with $q=1$, we get

$$\kappa_{\rm ran}(r) \approx \sqrt{\mathrm{var}(\ln\sigma)} \times \frac{\langle 1/\sigma\rangle}{L} \times \left(\frac{L}{r}\right)^{1-H}, \quad l \leqslant r \leqslant L, \quad 0 < H < 1, \tag{62}$$

where $H=\zeta(1)$ is the so-called (global) Holder or Hurst exponent for $\ln\sigma(\boldsymbol{x})$. We have replaced $r$ with $r/L$ in (58b) to restore dimensional consistency, and $S_{\ln\sigma}(1,L)$ in (61) by the standard deviation for $\ln\sigma$, i.e., $\sqrt{\mathrm{var}(\ln\sigma)}$. Any other simple and finite 1-point measure of relative variability for

$\sigma(\boldsymbol{x}) \geqslant 0$ such as $\mathrm{var}(\sigma)/\langle\sigma\rangle^2$ would do just as well since increments at the largest scale of interest here, $L$, will capture the full width of its 1-point pdf. The condition of finiteness is obviously important. The same remark applies to the MFP estimate used in (62) $\langle 1/\sigma\rangle$: if it is infinite, then $1/\langle\sigma\rangle$ is a better choice. For instance, $\langle(\ln\sigma)^q\rangle$ and $\langle 1/\sigma\rangle$ are ill-defined in media with finite volumes of optical vacuum ($\sigma = 0$). Bearing this caveat in mind, we proceed with $\sqrt{\mathrm{var}(\ln\sigma)}$ and $\langle 1/\sigma\rangle$.

Now we consider the extreme values of $H$. Smooth (i.e., random but almost everywhere differentiable) media have $H = 1$ and the scale dependence in the last term disappears, as expected from $\kappa_{\mathrm{det}}$ in (60); hence,

$$\kappa_{\mathrm{ran}} \approx \sqrt{\mathrm{var}(\ln\sigma)} \times \langle 1/\sigma\rangle / L, \quad r \leqslant L. \tag{63}$$

Extremely rough (white-noise-like or sparse fractal) media have $H = 0$ so the decorrelation scale $L$ appears to cancel from $\kappa(r)$ in (62), however, this is not the case because, in actuality, we now have decorrelation at the smallest scale: $R_{\ln\sigma} \approx l$ for the integral scale. The nonstationary scaling range is thus reduced to the pixel scale, so we can simply set $L = l$ in (63); hence,

$$\kappa_{\mathrm{ran}} \approx \sqrt{\mathrm{var}(\ln\sigma)} \times \langle 1/\sigma\rangle / l, \quad r \geqslant l. \tag{64}$$

This supports our radiative conclusions about white noise in Section 4.2 where we made explicit "thin pixel" ($\langle 1/\sigma\rangle \gg 1$) and finite variance ($\mathrm{var}(\ln\sigma) = O(1)$) assumptions, leading to $\kappa_{\mathrm{ran}} \gg 1$.

In summary, we have in the expressions (62)–(64) for $\kappa_{\mathrm{ran}}$:

- a first term representing the (outer-scale) magnitude of the variation, i.e., an overall non-dimensional measure of relative (1-point) variability strength;
- a second term giving the ratio of the actual MFP to the 2-point decorrelation scale for $\ln\sigma(x)$, i.e., an optically relevant overall non-dimensional measure of variability speed; and
- a third term that captures any scale-dependence that is often, but not always, present.

The categories defined in Section 4.5.1 based on the numerical value of $\kappa_{\mathrm{det}}$ carry over to $\kappa_{\mathrm{ran}}$ unchanged.

What if the medium is deterministic but non-differentiable, or otherwise not amenable to the computations prescribed in Eq. (60)? Consider, for instance, media where extinction varies periodically in space. A simple sine-wave medium can be analyzed [16] with $\kappa_{\mathrm{det}}(\boldsymbol{x})$ in (60); its square-wave counterpart however cannot, since gradients are either zero or infinity. This difference in wave-form is of course certainly not fundamental so, in order to draw the same conclusions, we simply use (63) with spatial averages and interpret $L$ as the period of the medium. Note that a judicious sampling rather than straightforward averaging of $\kappa_{\mathrm{det}}(\boldsymbol{x})$ may, at times, be in order due to divergences (this is indeed the case for a sine-wave extinction when the amplitude equals the mean). Alternatively, another moment than $\langle 1/\sigma\rangle$ can be used, e.g., simply revert to $1/\langle\sigma\rangle$. Finally, if the boundaries of the optical medium are at finite range (compared to $L$), then possibly another reference scale that defines medium size should be used in (62)–(64).

In short, some flexibility is required in choosing the various terms that make up the $\kappa$'s in (60)–(64). The essential rule here is to combine non-dimensional measures of variability strength (a 1-point statistical concept) and of variability speed (a 2-point statistical concept), and to capture any systematic scale-dependence anticipated in "strength" or "speed" or both.

## 5. Discussion

### 5.1. A limited analogy with long-path molecular spectroscopy at coarse resolution

In the absence of scattering, the RT problem is reduced to

$$I(\boldsymbol{x},\boldsymbol{\Omega}) = \int_0^{s_{\partial M}} S(\boldsymbol{x}-\boldsymbol{\Omega}s,\boldsymbol{\Omega})\mathscr{T}(\boldsymbol{x}-\boldsymbol{\Omega}s\to\boldsymbol{x})\,\mathrm{d}s \tag{65}$$

where $S(\boldsymbol{x},\boldsymbol{\Omega})$ is the source term and $s_{\partial M}$ is the distance along the inverse beam $\{\boldsymbol{x},-\boldsymbol{\Omega}\}$ from $\boldsymbol{x}$ to the boundary $\partial M$ of the medium, and with direct transmission $\mathscr{T}(\cdot)$ given by Eqs. (5)–(7). This is not the end of the story in many applications, atmospheric in particular, only a recipe for dealing with the spatial aspect of the problem. Indeed, both the source term in (65) and the extinction coefficient in (5), hence $\mathscr{T}(\boldsymbol{x}-\boldsymbol{\Omega}s\to\boldsymbol{x})$ in (65), can be strongly dependent on wavenumber $v$. In order to interpret observations, always made at some finite spectral resolution, we need to compute radiances which are deterministic mixtures of extinction values. Often, a spectral "band" will contain variability over a very wide range of extinction values. So, in practice, there is no essential difference with how we have treated the *spatially* variable extinction and its line-integral in (5). The *spectrally* variable part of (5) is the absorption cross-section per molecule rather than the molecular density. So, as a function of cumulative absorber mass, transmission is non-exponential as soon as there are mixtures of extinctions in the spectral band.

For technical details, we refer to Goody's [17] seminal paper on a statistical treatment of spectral variability which is complicated but fundamentally deterministic (since it results from molecular dynamics and quantum-mechanics). Non-exponential transmission is a direct outcome of Goody's model. It is noteworthy that spectral correlations are part of the *solution* in this occurrence of non-exponential FPDs in spectroscopy (i.e., efficiency and accuracy gains using "correlated-$k$" techniques) while spatial correlations are at the heart of the *problem* in the occurrence of non-exponential FPDs in 3D RT as shown in Section 4.

There are now standard software packages such as MODTRAN based on parameterizations, including non-exponential transmission, of the spectral variability of the atmosphere under a wide variety of physico-chemical conditions. These parameterizations are constantly being upgraded on the basis of improved "line-by-line" calculations which go back to the fundamentals of monochromatic radiative transfer in absence of scattering, i.e., Eqs. (5)–(7). The interesting open question is of course about how the spatial and spectral nonlinear effects are compounded. In the follow-on paper, we will survey the theoretical and empirical evidence that 3D RT in the cloudy terrestrial atmosphere has a systematic effect on total photon pathlength statistics (over all orders of scattering) and that this directly impacts absorption by all molecular species.

### 5.2. A debate surrounding the present study

To the best of our knowledge, only a few studies—most quite recent—address the systematic statistical effects of spatial variability on direct transmission per se. There are many more publications that address these effects in the context of multiple scattering, of thermal emission, or of neutron transport; we will survey this literature in our follow-on paper on the multiple scattering ramifications of the present study.

Romanova [18] states that direct transmission through a variable medium is not as predicted by the mean extinction but larger. Borovoi [5] concurs and relates this fact to Jensen's [2] inequality in probability theory; he also underscores the role of inhomogeneity scale (i.e., spatial correlation) but argues nonetheless for a mean-field radiative transfer equation where an effective exponential transmission law appears because he recognizes only our "too slow" and "too fast" variabilities. Without as much consideration of scale, Stephens et al. [19] make simple and specific assumptions (e.g., uniform probability density) about the variability of extinction in the 1-point statistical sense and compute moments of all orders in order to highlight differences with the exponential law. Lovejoy et al. [6] computed transmission statistics for "multifractal" extinction fields based on a large class of multiplicative cascade models which are, by construction, highly correlated in space; they found laws that are significantly wider than exponential. Knyazikhin et al. [7] independently investigated the special case of extinction fields modeled with "monofractal" measures (i.e., supported only on sparse Cantor-like sets) and found the same trend.

Kostinski [20] investigates ballistic photon propagation through a population of spatially-correlated scatterers/absorbers. This author's approach is a rather radical departure from previous studies because he makes none of the elementary assumptions in radiative transfer such as radiance/flux balance and/or the existence of an extinction field. Instead, he draws on a formula used by Landau and Lifshitz [21] in the theory of critical phenomena to assess the effects of correlations on collision rates for photons in a medium of discrete obstacles. Non-exponential transmission is a natural outcome of the correlations; he even finds a slow power law decay for a specific but illustrative case of *positive* ("clustering") correlations. This contrasts with the exponential law that results directly from a purely random spatial distribution governed by Poisson's point statistics.

Borovoi [22] comments on Kostinski's paper coming from his (and our) more standard perspective based on the RT equation; however, he continues to dismiss the importance of the variability scales that cause the large departures from exponential behavior in direct transmission (whereas we offer a more systematic analysis by showing that his Laplace transform representation applies in fact to all, not just the largest, variability scales). In his reply [23] to Borovoi's commentary, Kostinski insists that discrete-point statistics afford us a deeper look at reality and he provides counter-examples to Borovoi's, ours and his own earlier investigation in Ref. [20] of wider-than-exponential transmission laws. There is no contradiction here since sub-exponential transmission laws necessarily call for *negative* correlations ("anti-clustering") in the optical medium. Shaw et al. [24] investigate the effects of negative correlations in full detail. In the continuum (density-based) formulation used here, negatively correlated point distributions correspond to different ways of implementing the assumption of constant density: in the discrete world, some uniform media are actually more uniform than others. This situation is not directly amenable to the formalism used here, but there is a noteworthy analogy with the way some random quadrature rules used in (so-called "quasi-") Monte Carlo integration achieve faster than "$N^{-1/2}$" convergence by better sampling of space [25].

We hope the present study, started over a decade ago [4], will help resolve this controversy based largely on over- and under-statements, or at least contribute to the debate. Of course, everyone knows that a theory based on discrete particles is, by construction, more general than a theory based on continuum field concepts such as density, extinction, flux, radiance, and so on. Although a continuum theory such as hydrodynamics has little to say about scales smaller that the inter-particle distance, RT is in fact a kinetic theory based on particle interactions and it applies perfectly well in an optical vacuum. The true limitation of RT theory is that it is grounded in geometric optics: wavelengths

need to be small enough compared to inter-particle distances that (1) a photon "beam" can be defined and (2) coherence effects can be neglected inside "elementary volumes" (i.e., they occur at the scattering event but for a mixture of particles it is the energies and not the amplitudes and phases that add). We note that this limitation applies equally to the discrete-particle/discrete-photon approach advocated by Kostinski and co-workers and to the particle-density/photon-flux approach used by Borovoi, ourselves, and many others. Furthermore, both approaches rely on the fundamental notion of a distribution of paths between emission, scattering or absorption events, the FPD, which may or may not be exponential. We assume in most of this study that its mean, the MFP, exists although it can in principle be larger that any other dimension of the system.

The introduction of negatively correlated discrete media therefore does not contradict but complements the general conclusions of RT theory. However, the requirement that correlations—positive or negative—need to exist at least on the scale of the actual MFP seems to have been overlooked at times. In the realm of astrophysical, planetary and geophysical optics, negative correlations in the scattering medium (e.g., clouds, aerosol and molecules in the atmosphere) are not likely to occur on scales commensurate with the MFP simply because of the turbulent fluid- and/or thermo-dynamical processes that shape the 3D variability tend to promote clustering tendencies over a very broad range of scales. This tendency is currently the object of intense study in the case of terrestrial clouds [26–30]. Anti-correlations may occur at much smaller-than-MFP scales (e.g., due to sedimentation in stagnant water or air and electrostatic charge in convective clouds), and conceivably also at much larger scales (clear/sinking air defining the space between clouds caused by updrafts). But, as we demonstrated, these are not the variability scales that dominate the production of non-exponential decay in the FPD. If the optically active particles (i.e., with size $\approx$ wavelength) are densely packed, enough to produce negatively correlated positions by excluding each other from a volume of their own size, then coherence effects will also be at work, thus disabling both of the incoherent/geometrical-optics approaches mentioned here. (We note in passing that the prediction of sub-exponential transmission nonetheless is nonetheless in the right direction, with Anderson's localization of light—inhibited propagation—being the ultimate reduction of transmission; this occurs in disordered media with fluctuations—and transport MFP—on the scale of the wavelength [31].) In the realm of macroscopic RT, individual leaves in vegetation canopies will exclude each other physically but we are, there again, outside of the realm of kinetic transport theories that describe obstacles as "particles" having an interaction cross-section but no volume per se. Interestingly, in the properly reformulated (non-Markovian) RT theory for vegetation canopies and for realistic (fractal) canopy architecture, wider-than-exponential transmission laws are again a natural outcome [7].

In summary, the possibility of negative 2-point correlations is interesting but they are hard to find in nature without defeating the most basic assumptions of either formulation in competition here.

## 6. Summary and outlook

Beer's law predicts an exponential decay in direct transmission with distance from the light source. We have used methods inspired by some lesser-known probability theory to establish that effective transmission laws in 3D radiative transfer are never exponential. We have furthermore established that the actual mean-free-path (MFP) based on the effective free-path distribution (FPD) is always larger than that expected from the mean extinction coefficient when a natural mass-conservation

assumption is imposed on the spatial variability. Specifically, in the more interesting situations where the variability is characterized by relatively long-range correlations, it is better to estimate the MFP using the mean value of inverse extinction rather than the inverse of the mean extinction. We also show that, under the same general conditions, the effective FPD is always wider-than-exponential in the sense that the higher-order moments are systematically under-estimated by the exponential model, even if the MFP is properly evaluated. A simple binary mixture of extinctions is used to illustrate the basic results and we draw an important but limited analogy between spatial and spectral variabilities.

We show how to design criteria for forecasting strong 3D radiative transfer effects (very non-exponential FPDs). These diagnostics highlight the need for a certain "overall strength" and a "typical speed" of the variability that does not overwhelmingly reduce the variance in optical path at a fixed step-size. At any given amplitude of variability, it can indeed be so fast that almost every photon samples almost all of it at almost every step, in which case only the mean extinction matters and the FPD is quasi-exponential. It can also be so slow that photons sample essentially just one value of extinction between emission and detection or escape, in which case a local uniformity assumption can be made. Finally, extinction fluctuations over the MFP scale—give or take an order of magnitude—can be "resonant," meaning that extinction usually changes significantly in the course of a free path, but not so much as to sample all possible values. In this case, the effective FPD will be highly non-exponential.

The focus of this paper has been the spatial (propagation) part of the overall radiative transfer process as opposed to the scattering, emission or absorption components. Indeed, if the scattering kernel in the multiple-scattering source function is isotropic, then the mean-field FPD studied here is all that is needed, beyond boundary conditions, for an integral formulation of a natural mean-field theory for multiple scattering. In a follow-on paper, we will illustrate the general properties of mean-field 3D transport kernels, with an emphasis on radiation processes in the Earth's cloudy atmosphere.

## Acknowledgements

This work was supported in part by the Environmental Sciences Division of U.S. Department of Energy as part of the Atmospheric Radiation Measurement (ARM) program. ABD also acknowledges financial support of the LDRD/DR program at Los Alamos National Laboratory and of the CNRS for funding a visiting position at the Laboratoire de Météorologie Physique (LaMP) in Clermont-Ferrand. We thank Drs. A. Arnéodo, H. Barker, A. Benassi, A. Borovoi, R. Cahalan, H. Frisch, P. Gabriel, C. Jeffery, Y. Knyazikhin, A. Kostinski, M. Larsen, S. Lovejoy, K. Pfeilsticker, F. Sczcap, D. Schertzer, R. Shaw, G. Stephens, B. van Tiggelen, B. Watson, and W. Wiscombe for many fruitful discussions.

## Appendix A. With Albert Benassi (Laboratoire de Météorologie Physique and Departement de Mathématiques, Université Blaise Pascal, Clermont-Ferrand, France): The prevalence of 1-point scale-independence illustrated with scale-invariant variability models

Used quite extensively in Section 4, the property of "1-point scale-independence" is introduced in Section 3 by putting two statistical requirements on a random field $f(x)$. First, its smoothed versions

are defined as

$$f_r(x) = \frac{1}{r}\int_x^{x+r} f(x')\,\mathrm{d}x', \tag{A.1}$$

in analogy with Eq. (24) of the main text (only with different notations). Only one spatial variable is assumed here without any real loss of generality, only for more ease in the ensuing calculations. Echoing Eq. (26a), the first requirement is that

$$\langle f_r(x)\rangle \equiv \langle f(x)\rangle = \langle f\rangle \tag{A.2a}$$

for $0 \leqslant r \lesssim r_{\max}$; this is a necessary but far from sufficient condition for statistical stationarity. Echoing Eq. (26b), the second requirement is that the centered moments of $f_r(x)$ differ at most by a small amount, on the order of $r/r_{\max}$ when that ratio is small, i.e.,

$$\langle[\, f_r(x) - \langle f\rangle]^q\rangle / \langle[\, f - \langle f\rangle]^q\rangle - 1 = O(r/r_{\max}). \tag{A.2b}$$

In this appendix, we make straightforward connections between generally expected behavior in the auto-correlation function of (positively) correlated fields and 1-point scale-independence. We then illustrate that property with scale-invariant variability models that are not necessarily stationary nor Gaussian but do represent the most frequently occurring kind of variability found in nature (i.e., geo- or astro-physical systems). Counter-examples are also provided, selected for their relevance to the topic of the main text, positively correlated extinction fields and sub-exponential mean transmission laws.

### *A.1. A general remark on 1-point scale-independence in stationary random functions*

We will consider broad-sense stationary random functions, $f(x)$. Without loss of generality but gaining simplicity in the following calculations, we can therefore assume that the ensemble averages $\langle f_r(x)\rangle \equiv 0$ for all $x$ (by stationarity) and for all $r$ (by definition). We now ask how the variance of $f_r(x)$ varies with $r$, expecting a monotonic decrease since we are dealing with smoother versions of $f(x)$.

From (A.1), we have

$$\mathrm{var}[\, f_r] = \frac{1}{r^2}\left\langle \int_x^{x+r} f(x')\,\mathrm{d}x' \int_x^{x+r} f(x'')\,\mathrm{d}x'' \right\rangle \tag{A.3}$$

using the fact that the ensemble mean is zero. Reversing the order of the smoothing and the averaging, we have

$$\mathrm{var}[\, f_r] = \frac{1}{r^2}\int_x^{x+r}\int_x^{x+r} \langle(\, f(x')f(x'')\rangle\,\mathrm{d}x'\,\mathrm{d}x'' \tag{A.4a}$$

where

$$\langle(\, f(x')f(x'')\rangle = \rho(x''-x') \tag{A.4b}$$

the auto-correlation function defined in Eq. (56) of the main text, exploiting the vanishing mean assumption. So we are reduced to evaluating

$$\mathrm{var}[\, f_r] = \frac{1}{r^2}\int_0^r\int_0^r \rho(x''-x')\,\mathrm{d}x'\,\mathrm{d}x''. \tag{A.5}$$

We can illustrate first with white-noise which has $\rho(x''-x') \propto \delta(x''-x')$: we find $\mathrm{var}[f_r] \propto 1/r$. This result was used in Section 4.2 of the main text when dealing with the transmission statistics of such decorrelated fields when the spatial noise is Gaussian with a mean $\langle f \rangle > 0$. Then we already know from first principles that variances add, so the standard-deviation of estimates of the mean of length $r$ decrease as $1/\sqrt{r}$. White noise is clearly not 1-point scale-independent (cf. Fig. 3 in the main text).

If the variance of the original field $f(x)$ is finite, then we can write

$$\rho(\delta x) = \langle f(x) f(x+\delta x) \rangle \approx \mathrm{var}[f] \times [1 - a\delta x^{2H}] \tag{A.6}$$

for $\delta x \to 0^+$, with $a > 0$ and $H > 0$. We will take this approximation to be valid for $\delta x \lesssim \delta x^*$, some value on the order of the integral scale in Eq. (57) of the main text, if it exists (i.e., $\rho(r)$ is indeed integrable). We can easily evaluate (A.4) with the functional form of $\rho(\cdot)$ taken from (A.6), leading to

$$\mathrm{var}[f_r] \approx \mathrm{var}[f] \times [1 - br^{\alpha}] \tag{A.7}$$

with $\alpha = 2H$ and $b = a/(H+1)/(2H+1)$ and the approximation in (A.7) will of course prevail for $r \lesssim r_{\max} \approx \delta x^*$. We notice that (A.7) can be rewritten as (A.2b) in the main text for $q = 2$ simply by dropping the zero-mean assumption and writing the right-hand side as $\propto r^{\alpha}$.

Generalization of (A.7) to higher-order moments is immediate for Gaussian processes since any moment can be derived from variance and ratios of Euler's Gamma function in Eq. (12). This completes the proof that the large class of Gaussian stationary random functions that verify (A.6) are endowed with 1-point scale-independence. We illustrate 1-point scale-independence in Gaussian and non-Gaussian fields respectively in Sections A.3 and A.4, but start by showing that there is a characteristic non-stationarity present in the small scale regime of interest here.

### A.2. Scale-invariance and the non-stationary regime of otherwise stationary functions

Based on a remark in Section 4.3 of the main text connecting $\rho(\cdot)$ and the second-order structure function, we know from (A.6) that $2H = \zeta(2)$ in the multi-scaling representation of structure functions in Eqs. (58a) and (58b). So the vast class of finite-variance stationary functions we are interested in here are in fact operationally stationary (effectively decorrelated) only at large scales. There is a large range of scales where their increments yield non-trivial power-law structure functions in Eq. (58b). In this scaling range the functions are, for all practical purposes, in the even vaster class of non-stationary functions with stationary increments.

We must note however that the parameter $H$ used here can be identified exactly with the Hurst exponent $\zeta(1)$ used in the latter part of Section 4 only in the so-called "mono-scaling" or "monofractal" limit where $\zeta(q)/q \equiv \textit{constant}$. This abuse in notation is unfortunately quite frequent in the fractal literature. Robust determination of $H$ is a major concern in fractal data analysis and, interestingly, plotting the left-hand side of (A.2b) versus $r$ in log–log axes has been proposed as a possible method using combinations of $q = 1$ and 2 and an (implicit) monoscaling assumption.

### A.3. One-point scale-independence of additive scale-invariant models

As a first empirical example of 1-point scale-independence, we used in Section 4.3 (Fig. 4) the Wiener–Levy process which has $H = 1/2$, and

$$\zeta(q) = Hq \tag{A.8}$$

generally speaking. This process, the well-known temporal trace of Brownian motion (Bm), is the prototypical nonstationary random function with stationary increments. By enforcing scale-invariance in this example, we have made the model nonstationary throughout practically the whole range of available scales.

Mandelbrot [32] popularized his generalizations of Brownian motion known as "fractional" Brownian motion (fBm) which are in the same broad category as Bm: they obey Eq. (A.7) with $0 < H < 1$ as unique scaling parameter. So they too are nonstationary with stationary increments and therefore 1-point scale-independent. Fig. 5 shows an example for $H = 1/3$ which, considered as a second-order statistic, is the prevailing value for terrestrial clouds [13]. Panel (a) shows a sample trace $f(x)$ of length 8192 pixels with a unit mean and a variance of $1/9$ (standard deviation $1/3$); in (b) histograms using 10 realizations of $f_r(x)$ are plotted for a large range of averaging lengths $r = 1, \ldots, 256$ pixels (by powers of 2). The 1-point scale independence via pdf overlap is obvious even though we do observe the expected retreat as $r$ increases of the extreme values under deliberately limited sampling.

There is a striking difference in construction method between Brownian and fractional Brownian processes.

- The former can be generated simply by summing white noise and more can be generated as needed because the increments are not only stationary but independent. Brownian motion has a Markovian quality (only the present is needed to predict the future).
- The latter have stationary but correlated increments. Indeed, it is easy to show, starting with (58a) and (58b) for $q = 2$ and $2r$ as well as $r$, that the correlation of successive increments in the same direction in $x$ is given by

$$\langle [f(x+2r) - f(x+r)][f(x+r) - f(x)] \rangle \propto [2^{2H-1} - 1] r^{2H}. \tag{A.9}$$

  So successive increments anti-correlate if $H < 1/2$ (cf. Fig. 5) and correlate if $H > 1/2$. As far as we know, this correlation structure can only be achieved by a scale-by-scale hierarchical or "cascade" construction.

All of these above models for scale-invariant/fractal variability are additive in nature: sums (across scales) of independent random variables that are generally assumed to be Gaussian. Benassi et al. [33] recently proposed a general framework for building such models based on wavelet-like entities and a general (not necessarily dyadic) tree structure across scales and positions. They even introduce a degree of intermittency by randomly "clipping" branches in the tree, thus spoiling the Gaussian distribution of the field and of its increments yet leaving (A.6)—hence the 1-point scale-independence—intact.

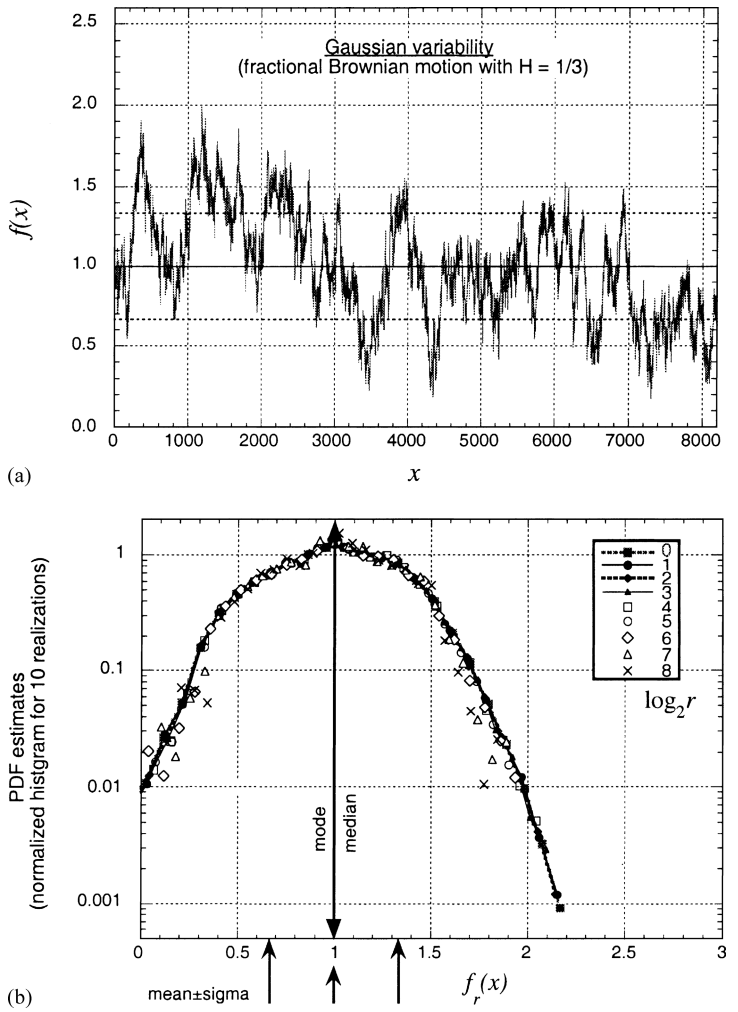


Fig. 5. One-point scale-independence in a scale-invariant Gaussian variability model. (a) Sample trace with 8192 pixels, unit mean and standard deviation 1/3. (b) Normalized histograms for 10 random realizations (truncated infrequently at 0 to enforce the positivity of the extinction field) and a significant range of averaging lengths $r=2^m$ pixels ($m=0,\ldots,8$); we note that the extreme values are naturally contracting as $r$ increases. The clear overlap out to several times the standard deviation is equivalent to the 1-point scale-independence defined in Section 3.

### *A.4. One-point scale-independence in multiplicative scale-invariant models*

Another way of obtaining intermittency in the same sense of broad non-Gaussian increments is to use multiplicative processes. These models are best constructed simply as exponentials of additive processes, without or with incremental correlations of either sign. Marshak et al. [34] use a sequence of Bernoulli random variables to construct the "bounded cascades" of Cahalan et al. [12] and derive their multi-scaling properties:

$$\zeta(q) = \min\{qH, 1\} \tag{A.10}$$

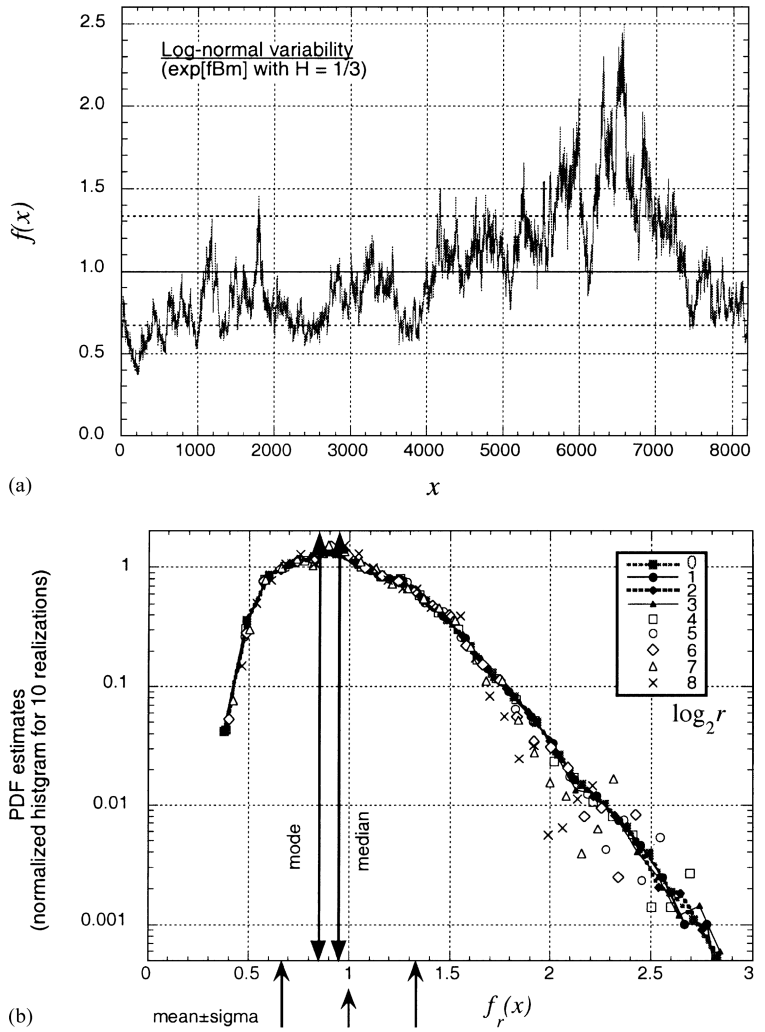


Fig. 6. One-point scale-independence in a scale-invariant log-normal variability model. (a) Sample trace with 8192 pixels, unit mean and standard deviation 1/3 (as in Fig. 5); these assumptions lead to $\sigma = \mathrm{StDev\_of\_ln}\, f = \sqrt{\ln(10/9)} \approx -0.3246$, and $\mu = \mathrm{Mean\_of\_ln}\, f = -\sigma^2(\ln f)/2 \approx -0.0527$. (b) Overlapping histograms for 10 different realizations and a large range of averaging lengths $r = 2^m$ pixels ($m = 0, \ldots, 8$); in contrast with the Gaussian case, the mean ($\exp[\mu + \sigma^2/2]$), the median ($\exp[\mu]$) and the mode ($\exp[\mu - \sigma^2]$) are all different. Apart from the expected contraction of the extremes, we see that the sampling noise is stronger in the bins for high $f(x)$ values than for the Gaussian model in Fig. 5. Nonetheless, roughly the same histogram overlap is observed as in Fig. 5.

in Eq. (58b) with parameter $H > 0$ (Heaviside steps are obtained in the limit $H \to \infty$). As noted in the main text, the multi-scaling function is definitely concave. The model is "multifractal."

Similarly, one can construct log-normal models by using exponentials of fBm. Fig. 6 shows a typical example with the same mean, variance and number of realizations as in Fig. 5; as expected, more severe sampling problems in the high values are apparent. Such multiplicative models are generically multifractal but still obey the stated conditions for (A.6) and are therefore 1-point scale-independent.

Before closing this technical appendix, we recall that there is another class—and indeed a considerably older one—of multiplicative scaling models in the turbulence and geophysical literatures. These models are not functions per se but measures (Dirac's *generalized* $\delta$-functions are in fact an extreme example). They generically have $\zeta(q) \equiv 0$ so, at $q = 2$, they do not obey the conditions of (A.6) and are therefore not 1-point scale-independent. As an instance, this class includes the "unbounded" limit ($H \to 0$) of Cahalan's bounded cascades in (A.10) that were first introduced by Meneveau and Sreenivasan [35] to account for the intense variability of the kinetic energy dissipation rate in strong turbulence.

Beyond their degenerate structure functions, these singular cascade models have interesting multifractal properties in their own right [6,10,11,28,35]. Although these properties are based on the smoothed versions of the measures defined exactly as in Eq. (A.1), they are out of the scope of this survey. It suffices to say that the auto-correlation functions of these scale-invariant models are weakly singular power-laws as parameterized in Eq. (59b): $\rho(\delta x) \propto \delta x^{-\mu}$ $(0 < \mu < 1)$. This case can be treated as in Section A.1 by formally taking $a < 0$ (to enforce a positive prefactor), setting $2H = -\mu$, and ignoring the constant term in (A.6). The outcome in (A.7) is that var$[f_r]$ has the same power-law dependence on $r$ with the prefactor positive becoming $-b = -a/(1-\mu/2)(1-\mu)$. This confirms the violation of 1-point scale-independence, as found in Section A.1 for $\delta$-correlated "white" noise.

Returning to the topic of the main text, direct transmission laws in fractal media with such strong (but not white-noise) variability in the small-scale limit turn out to be sub-exponential in spite of the lack of 1-point scale-independence, at least in a number of special cases [6,7].

## References


[1] Marchuk G, Mikhailov G, Nazaraliev M, Darbinjan R, Kargin B, Elepov B. The Monte Carlo methods in atmospheric optics. New York (NY): Springer, Berlin, 1980.

[2] Jensen JLWV. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math 1906;30:789–806.

[3] Feller W. An introduction to probability theory and its applications, vol. 2. New York (NY): Wiley, p. 1971.

[4] Davis A. Radiation transport in scale-invariant optical media, PhD thesis, Physics Department, McGill University, Montreal, Canada, 1992.

[5] Borovoi AG. Radiative transfer in inhomogeneous media. Dok Akad Nauk SSSR 1984;276:1374–8 (English version: Sov Phys Dokl 1984; 29 (6)).

[6] Lovejoy S, Watson B, Schertzer D, Brosamlen G. Scattering in multifractal media. In: Briggs L, editor. Proceedings of Particle Transport in Stochastic Media. La Grange Park (Il): American Nuclear Society, 1995. p. 750–60.

[7] Knyazikhin Y, Kranigk J, Myneni RB, Panfyorov O, Gravenhorst G. Influence of small-scale structure on radiative transfer and photosynthesis in vegetation canopies. J Geophys Res 1998;103:6133–44.

[8] Papoulis A. Probability, random variables, and stochastic processes. New York (NY): McGraw-Hill, 1965.

[9] Pomraning GC. Linear kinetic theory and particle transport in stochastic mixtures. Singapore: World Scientific, 1991.

[10] Muzy JF, Bacry E, Arnéodo A. The multifractal formalism revisited with wavelets. Int J Bifurcation Chaos 1994;4: 245–302.

[11] Marshak A, Davis AB, Wiscombe WJ, Cahalan RF. Scale-invariance of liquid water distributions in marine stratocumulus, Part 2—multifractal properties and intermittency issues. J Atmos Sci 1997;54:1423–44.

[12] Cahalan RF, Snider JB. Marine stratocumulus structure during FIRE. Remote Sens Environ 1989;28:95–107.

[13] Davis AB, Marshak A, Wiscombe WJ, Cahalan RF. Scale-invariance of liquid water distributions in marine stratocumulus, Part 1—spectral properties and stationarity issues. J Atmos Sci 1996;53:11538–58.

[14] Cahalan RF, Ridgway W, Wiscombe WJ, Bell TL, Snider JB. The albedo of fractal stratocumulus clouds. J Atmos Sci 1994;51:2434–55.

[15] Barker HW, Morcrette JJ, Alexander GD. Broadband solar fluxes and heating rates for atmospheres with 3D broken clouds. Quart J Roy Meteorol Soc 1998;124:1245–71.
[16] Davis AB, Marshak A. Multiple scattering in clouds: insights from three-dimensional diffusion/P1 theory. Nucl Sci Eng 2001;137:251–88.
[17] Goody RM. A statistical model for water-vapor absorption. Quart J Roy Meteor Soc 1952;78:165–9.
[18] Romanova LM. Radiative transfer in a horizontally inhomogeneous scattering medium. Izv Acad Sci USSR Atmos Ocean Phys 1975;11:509–13.
[19] Stephens GL, Gabriel PM, Tsay SC. Statistical radiative transport in one-dimensional media and its application to the terrestrial atmosphere. Transp Theory Statis Phys 1991;20:139–75.
[20] Kostinski AB. On the extinction of radiation by a homogeneous but spatially correlated random medium. J Opt Soc Am A 2001;18:1929–33.
[21] Landau LD, Lifshitz EM. Statistical physics. London (UK): Pergamon Press, 1958.
[22] Borovoi AG. On the extinction of radiation by a homogeneous but spatially correlated random medium: comments. J Opt Soc Am A 2002;19:2517–20.
[23] Kostinski AB. On the extinction of radiation by a homogeneous but spatially correlated random medium: review and response to comments. J Opt Soc Am A 2002;19:2521–5.
[24] Shaw RA, Kostinski AB, Lanterman DD. Super-exponential extinction of radiation in a negatively-correlated random medium. JQSRT 2002;75:13–20.
[25] Sobol IM. A primer for the Monte Carlo method. Boca Raton (Fl): CRC Press, 1994.
[26] Baker B. Turbulent entrainment and mixing in clouds: a new observational approach. J Atmos Sci 1992;49:387–404.
[27] Brenguier JL. Observations of cloud microstructure at the centimeter scale. J Appl Meteor 1993;32:783–93.
[28] Davis AB, Marshak A, Gerber H, Wiscombe W. Horizontal structure of marine boundary layer clouds from centimeter to kilometer scales. J Geophys Res 1999;104:6123–44.
[29] Kostinski AB, Shaw RA. Scale-dependent droplet clustering in turbulent clouds. J Fluid Mech 2001;434:389–98.
[30] Jeffery CA. Investigating the small-scale structure of clouds using the delta-correlated closure. Atmos Res 2001; 59–60:199–215.
[31] van Tiggelen BA, Lagendijk A, Wiersma DS. Radiative transfer of localized waves, a local diffusion theory. In: Soukoulis CM, editor. Photonic crystals and light localization in the 21th century. Dordrecht (Netherlands): Kluwer, 2001. p. 475–87.
[32] Mandelbrot BB. The fractal geometry of nature. San Francisco (CA): W.H. Freeman, 1982.
[33] Benassi A, Cohen S, Deguy S, Istas J. ANHA-Applied and Numerical Harmonic Analysis. In: Debnath L, editor. Benassi, et al., Chapter II, Self-similarity and Intermittency: Birkhäuser, 2003;361–375.
[34] Marshak A, Davis AB, Cahalan RF, Wiscombe WJ. Bounded cascade models as non-stationary multifractals. Phys Rev E 1994;49:55–69.
[35] Meneveau C, Sreenivasan KR. Simple multifractal cascade model for fully developed turbulence. Phys Rev Lett 1987;59:1424–7.