Geophysical Journal International

Geophys. J. Int. (2022) **228,** 816–825 https://doi.org/10.1093/gji/ggab376
Advance Access publication 2021 September 21
GJI Seismology

The relationship between cross correlations and Green's functions in ambient noise interferometry with Bayesian constraints

Victor C. Tsai and Korbinian Sager

Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, RI 02912*, USA. E-mail:* victor_tsai@brown.edu

Accepted 2021 September 16; in original form 2021 September 7

SUMMARY

Cross correlation of ambient seismic noise has begun to provide detailed images of the Earth that are not possible to attain using earthquake-based or active-source imaging techniques. However, the theoretical justification for the ambient noise correlation or seismic interferometry approach typically relies on an equipartition or diffuse field assumption that is not expected to be satisfied on Earth. Here we demonstrate that a Bayesian inference approach to calculating the expected cross correlation of seismic signals leads to an improved understanding of its relationship with the Green's function. The Bayesian derivation is found to exactly replicate the equipartition result under a global energy constraint, and is able to accommodate any number of other constraints that typically cannot be utilized in other approaches. With these stronger constraints, the cross correlation is found to deviate more and more strongly from the 'expected' Green's function, but nonetheless a specific prediction can be made regarding the expected cross correlation which can be used as a more reliable point of comparison with observations than the Green's function. The approach therefore provides a path forward for how to use seismic cross correlation data under more and more realistic conditions and improvements in knowledge, and should eventually help improve the accuracy of resulting images of the Earth.

Key words: Interferometry; Seismic noise; Theoretical seismology; Wave propagation; Seismic tomography.

1 INTRODUCTION

In the past two decades, there has been an explosion in the number of studies that use ambient seismic noise to image the Earth (Nakata *et al.* 2019). This growth in turn was spurred by the theoretical finding that the cross correlation of noise observed at two points can sometimes be related with the Green's function, or impulse response function, between the points (Eckart 1953; Aki 1957; Lobkis & Weaver 2001; Snieder 2004). The usual assumption made to relate the cross correlation to the Green's function is that the modes of the system are equipartitioned and uncorrelated (ergodic, Lobkis & Weaver 2001). However, it has since been realized that the Earth is very far from satisfying the assumptions needed for the theoretical equipartition relationship to hold (Snieder 2004; Lin *et al.* 2008; Tsai 2009), making it unclear whether the theoretical basis for the various noise correlation studies has merit. Furthermore, despite some progress in this direction (Tsai 2009; Harmon *et al.* 2010; Sager *et al.* 2020), it has been unclear how practitioners can best make use of the wealth of knowledge about the ambient noise field that may be available to improve the accuracy with which noise correlations can be used to infer Earth structure.

Bayesian inference theory has made major strides over the last half century, starting with the seminal work of Shannon (1948), where Shannon showed that the amount of information in a system can be characterized uniquely by what is now called the Shannon entropy. With the recognition that the thermodynamic Boltzmann/Gibbs entropy and pure informational entropy were closely related concepts, Jaynes (1957) showed that all of statistical mechanics could be reformulated using a Bayesian information theory approach rather than the standard equipartition approach of Boltzmann that is typically found in textbooks (Kittel & Kroemer 1980). As information theory developed further, Jaynes (2003) emphasized how a Bayesian statistical inference framework for considering the state of knowledge of a system has wide ranging applications in many fields. This type of Bayesian approach has begun to permeate seismology (e.g. Mosegaard & Tarantola 1995; Bodin & Sambridge 2009; Muir & Tsai 2020), reviving the old ideas of Jeffreys (1948) and highlighting their usefulness (Robert *et al.* 2009).

Here, we demonstrate how Bayesian inference theory can be used to derive the cross correlation under a variety of different constraints. With a global energy constraint, the cross correlation is found to be related to the Green's function in exactly the manner predicted by the equipartition result. Under other stronger constraints, the cross correlation is found to deviate from this expectation. A path forward for making better use of ambient seismic noise correlations is suggested which should lead to more accurate Earth imaging results in realistic situations that depart from the idealized equipartition assumptions that are not satisfied on Earth.

2 BRIEF SUMMARY OF EQUIPARTITION DERIVATION

In order to appreciate the differences between the Bayesian inference theory approach and the standard equipartition approach, we briefly outline the key arguments of the derivation for why equipartition of ambient noise leads to a relationship between the cross correlation and the Green's function. For further details, we refer the reader to Lobkis & Weaver (2001), Tsai (2010) and Fichtner & Tsai (2019).

The standard derivation begins with the assumption that the system satisfies a wave equation and therefore has normal modes that can generally be expressed as

$$u_n(\boldsymbol{x}, t) = s_n(\boldsymbol{x})\cos(\omega_n t + \phi_n), \tag{1}$$

where u_n is the nth mode displacement, $\boldsymbol{x}$ is the position vector, t is time, $s_n(\boldsymbol{x})$ is the mode shape, ω_n is the mode frequency and ϕ_n is the phase of the mode. A general oscillating solution can then be expressed as

$$u(\boldsymbol{x}, t) = \sum_{n=1}^{N} a_n u_n(\boldsymbol{x}, t) = \sum_{n=1}^{N} a_n s_n(\boldsymbol{x})\cos(\omega_n t + \phi_n) \tag{2}$$

for some choice of coefficients a_n (Dahlen & Tromp 1998). For a modal system such as this, the Green's function can be written as $G = \sum_n G_n$, where

$$G_n(\boldsymbol{x}, \boldsymbol{x}_s, t) = \begin{cases} \frac{s_n(\boldsymbol{x})s_n(\boldsymbol{x}_s)}{\omega_n}\sin(\omega_n t) & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}. \tag{3}$$

When all ω_n are distinct, it can be shown that the cross correlation

$$C_{AB}(t) \equiv \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} u(\boldsymbol{x}_A, \tau)u(\boldsymbol{x}_B, \tau + t)\mathrm{d}\tau \tag{4}$$

has no cross-mode terms because of temporal orthogonality ($\int \cos\omega_m t \cos\omega_n t \, \mathrm{d}t = 0$ when $\omega_m \neq \omega_n$) and the cross correlation is therefore given by (Tsai 2010; Fichtner & Tsai 2019)

$$C_{AB}(t) = \frac{1}{2}\sum_n a_n^2 s_n(\boldsymbol{x}_A)s_n(\boldsymbol{x}_B)\cos(\omega_n t). \tag{5}$$

If energy is equipartitioned, then $a_n = \alpha/\omega_n$, where α is a constant. Substituting this into eq. (5) and comparing with eq. (3) immediately demonstrates that

$$\frac{\mathrm{d}C_{AB}}{\mathrm{d}t} = -\frac{\alpha^2}{2}[G(\boldsymbol{x}_A, \boldsymbol{x}_B, t) - G(\boldsymbol{x}_A, \boldsymbol{x}_B, -t)]. \tag{6}$$

When some of the ω_n are equal (degenerate), the deterministic derivation above is incomplete since the orthogonality property required to derive eq. (5) is not satisfied. Instead, one must rely on the phases ϕ_n of those degenerate modes to be randomly sampled in time (ergodically) so that on average the correlation between different modes of the 'noise' field still have no contribution, and again lead to eq. (6), where now $u(\boldsymbol{x}, t)$ must be thought of as having sampled an ensemble of random phases (see Appendix A1). We note that the diffuse field framework (e.g. Lobkis & Weaver 2001) for obtaining a similar result also implicitly relies on degenerate modes having their energy equipartitioned and ergodic, so that the derivation is similar.

Before continuing to the Bayesian inference theory derivation, we point out a number of important characteristics of the equipartition result that we will later contrast with the Bayesian inference theory results. Perhaps most importantly, the key requirement that energy is equipartitioned between all of the modes is a very strict assumption that is not expected to hold in any Earth application. Not only are ambient seismic noise sources non-uniformly spatially distributed, with a strong bias towards the Earth's surface, but the assumption of ergodicity that is typically relied on to achieve equipartition (Chirikjian 2012) is known to fail for oscillatory systems that are weakly nonlinear (Arnold 1963) like wave propagation on the Earth. Thus, even at relatively high frequencies, where scattering and attenuation are significant, the length of time needed to reach an ergodic equipartitioned state is likely longer than the age of the universe due to the large number of modes (Jaynes 2003). Moreover, when modes are degenerate, one must explicitly rely on the ergodicity of the system to sample an ensemble of random phases in addition to attaining equal energies (Tsai 2010), so that the Earth even more clearly fails to satisfy the assumptions required for the derivation to hold.

3 BAYESIAN INFERENCE THEORY DERIVATION

Approaching the problem from an Bayesian inference theory standpoint can be framed by asking what knowledge is known about the physical system and how to best express that degree of probabilistic knowledge or ignorance. This Bayesian inference approach is now the standard in data analysis (Sivia 2006), but is not as common in other applications. In the statistical mechanics problem of determining the probabilities that gas molecules occupy certain states, this approach can lead to specific predictions about the macroscopic features like the ideal gas law from determining which probabilities are most likely, and is an alternative approach to the Boltzmann approach of assuming equal probabilities of states as a starting point (Jaynes 2003; Schekochihin 2020). For the case of interest here, of attempting to derive a relationship between point-wise cross correlations and the system's Green's function, we again hope to be able to derive a similar property (e.g. as eq. 6) with this more general alternative approach to the equipartition approach. This 'Bayesian' instead of 'frequentist' approach can seem intuitively less rigorous and more subjective, but there is abundant evidence that the approach is useful and often yields identical results (Jaynes 2003). More importantly, as will be discussed in later subsections, the Bayesian approach can be more easily generalized to a range of more realistic situations where certain constraints on the system are known.

Before performing any calculations, we briefly outline the main idea of the Bayesian inference theory approach. To express the probabilistic knowledge of a physical quantity like a cross correlation function, the expected value of that quantity is determined by integrating the quantity over its full probability distribution, accounting for all of the different possible states of the system. In this case, the states are all of the various possible oscillating solutions (see eq. 2) and can be identified by a set of amplitudes and phases. In contrast to the equipartition derivation, modes will not be required to have equal energies and instead all possible states are assumed to contribute, including oscillations with vastly different amounts of energy. If some constraints are known, this information can be used to further restrict which states are possible (or less probable) and thus restrict the integration to a subset of the full set of states or downweight various contributions. Later sections implement such constraints or weights, but it is important to note that none of the constraints used are as restrictive as an *a priori* equipartition constraint. Finally, we note that the results presented here require none of the additional formalism described by Jaynes (1957, 2003) besides the simple expected value calculation described above, but that an entropy perspective could have been taken and would have resulted in more sophisticated calculations that provided a deeper analogy with that work. Such a comparison is beyond the scope of this work.

3.1 No constraints

If no other information is known about the system, then all possible states of the system may be thought to be equally likely. Before we can calculate what this implies, however, we must decide what this statement implies specifically about knowledge of a_n and ϕ_n. One possibility would be that there is a uniform probability that a_n and ϕ_n take on any real value, or one could say that one has a uniform prior on those parameters. However, another possibility would be that the complex number $b_n \equiv a_n e^{i\phi_n}$ has a uniform probability in the complex plane $\mathbb{C}$, or that b_n has a uniform prior. In this case, both assumptions lead to similar results (differing only by a constant factor) and we demonstrate the uniform b_n choice here without elaborating on the subtlety of the choice of prior (see Jaynes 2003).

To calculate the expected value of the cross correlation, the Bayesian inference approach tells us to simply integrate over all possible cross correlations calculated from each possible state S within the entire probability distribution $P(S)$. We denote this probabilistic version of the cross correlation C^P_{AB}. With a uniform probability for b_n, the 2-mode case ($N = 2$) yields

$$\begin{aligned} C^P_{AB} &= \int_S P(S) C_{AB} \mathrm{d}S = \int_{b_1\in\mathbb{C}} \int_{b_2\in\mathbb{C}} c \cdot C_{AB} \mathrm{d}z_2 \mathrm{d}z_1 \\ &= \lim_{R\to\infty} \frac{1}{\pi^2 R^4} \int_0^R \int_0^{2\pi} \int_0^R \int_0^{2\pi} a_1 a_2 C_{AB} \mathrm{d}\phi_2 \mathrm{d}a_2 \mathrm{d}\phi_1 \mathrm{d}a_1, \end{aligned} \tag{7}$$

where c is a normalization constant (e.g., Sivia 2006), and the general case has a similar $2N$-dimensional integral over a_n and ϕ_n. In both cases, each contribution from the cross correlation can be simplified as in eq. (5) for the non-cross-mode terms, and the cross-mode contributions are of the form $a_m a_n s_m(\boldsymbol{x}_A) s_n(\boldsymbol{x}_B) \cos(\omega_n t + \phi_m - \phi_n)$. In this Bayesian inference approach, the cross-mode contributions disappear after integration over ϕ_n since the average value of $\cos(\phi_m - \phi_n)$ is zero (see Appendix A1), similar to how the cross-mode contributions cancel in the equipartitioned case. We note, though, that the reason for the integration is philosophically different, with it expressing an ensemble of ergodically attained states in Section 2 whereas it expresses an uncertainty of attained states in the present case. With the cross-mode cancellation, eq. (7) simplifies to

$$\begin{aligned} C^P_{AB} &= \lim_{R\to\infty} \frac{2}{R^4} \int_0^R \int_0^R a_1 a_2 \sum_{n=1}^{2} a_n^2 s_n^A s_n^B \cos(\omega_n t) \mathrm{d}a_2 \mathrm{d}a_1 \\ &= \lim_{R\to\infty} \sum_{n=1}^{2} \frac{s_n^A s_n^B \cos(\omega_n t)}{R^2} \int_0^R a_n^3 \mathrm{d}a_n \\ &= \lim_{R\to\infty} \frac{R^2}{4} \sum_{n=1}^{2} s_n(\boldsymbol{x}_A) s_n(\boldsymbol{x}_B) \cos(\omega_n t), \end{aligned} \tag{8}$$

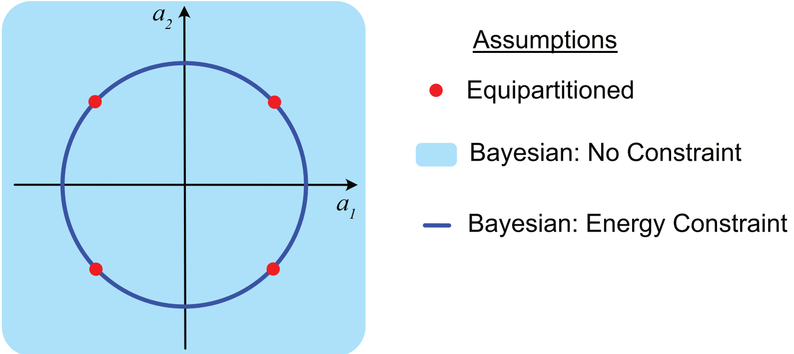

Figure 1. Schematic showing the differences in assumed accessible states between the standard equipartition approach and the Bayesian inference theory approach (with and without constraints). Example is for a 2-mode case where $\omega_2 = \omega_1$, and only amplitudes a_i are shown. All phases are accessible in all three cases.

where the abbreviation $s_n^A \equiv s_n(\boldsymbol{x}_A)$ was used. The generalization from the 2-mode case to the N-dimensional case is a straightforward application of the same simplifications and yields

$$\begin{aligned} C_{AB}^P &= \int_{\mathbb{C}^N} c \cdot C_{AB} \mathrm{d}\boldsymbol{z}^N = \lim_{R\to\infty} \frac{2^N}{R^{2N}} \int_{\mathbb{R}_+^N} C_{AB} \mathrm{d}a_n^N \\ &= \lim_{R\to\infty} \sum_{n=1}^{N} \frac{s_n^A s_n^B \cos(\omega_n t)}{R^2} \int_0^R a_n^3 \mathrm{d}a_n \\ &= \lim_{R\to\infty} \frac{R^2}{4} \sum_{n=1}^{N} s_n(\boldsymbol{x}_A) s_n(\boldsymbol{x}_B) \cos(\omega_n t), \end{aligned} \tag{9}$$

where $\mathbb{R}_+$ denotes the positive real numbers.

The final result in eq. (9) has obvious similarities to the equipartition result ($a_n = \alpha/\omega_n$ in eq. 5) but also significant differences. Firstly, properly taking the limit ($R \to \infty$) yields an infinite answer, which might have been anticipated since the average magnitude of a vector in the complex plane is infinite and the cross correlation scales with the vector magnitude squared. Ignoring this normalization issue (keeping R as a variable), comparison with eq. (3) leads to

$$C_{AB}^P = \frac{R^2}{4} \frac{\mathrm{d}}{\mathrm{d}t} \left[G(\boldsymbol{x}_A, \boldsymbol{x}_B, t) - G(\boldsymbol{x}_A, \boldsymbol{x}_B, -t) \right]. \tag{10}$$

This result is exactly equivalent (up to normalization) to that described by Tsai (2010) and Fichtner & Tsai (2019) when modal amplitudes (rather than energies) are equipartitioned. However, importantly, it should be noted that in this Bayesian inference version, there are many contributions from terms for which modal amplitudes are not equipartitioned (see Fig. 1). In the equipartitioned case, individual ensemble members could have had non-zero contributions, but it is an assumption that the ensemble average is exactly equipartitioned rather than the result of a specific calculation for the expected value as in the Bayesian approach. Viewed another way, the Bayesian approach provides the justification that the modal equipartition assumption is valid in this unconstrained case.

To summarize this section's results, investigating the probabilistic average value of the cross correlation for the most general case reveals a direct relationship between the correlation function and the inter-station Green's function. In the Bayesian approach there is no need to introduce specific *a priori* assumptions, for example on equipartitioned modal amplitudes, which is hard to justify in conventional derivations. Instead, it is the probabilistic value of the various contributions that favors certain contributions more than others when priors are relatively uninformative. The eq. (10) result may therefore best be thought of as being the most naive expectation, appropriate only when one has complete ignorance about which probabilistic contributions may be more likely. Including constraints shifts the nature of the most likely contributions, which we demonstrate in the following sections.

3.2 Global energy constraint

The formally infinite answer of the cross correlation in eq. (10) can be remedied in a number of different ways. Perhaps the simplest solution is to keep R finite, which has the interpretation of making the (prior) probability of b_n constant over a disk of radius R in the complex plane and setting the probability to zero for all other b_n. This case is in fact the best interpretation of the result in eq. (10) since all limits are then finite.

However, there is a more interesting remedy possible by adding the constraint that the total global energy is known or constrained. This approach is similar to the approach taken in the Bayesian information theory approach to statistical mechanics, where all states are possible subject to the constraint that the total thermodynamic energy of the system is known or measured (Jaynes 2003). In the present case, if the modes s_n are normalized such that $\int_{\boldsymbol{x}} s_n^2(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$, then the global energy constraint can be written as

$$U = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \int_{\boldsymbol{x}} \dot{u}^2(\boldsymbol{x}, t)\mathrm{d}\boldsymbol{x}\mathrm{d}t, \tag{11}$$

where this expression assumes constant density for simplicity and U is the total energy per unit density. A factor of 2 multiplies the kinetic energy to account for the equal amount of potential energy (Taylor 2005). Evaluating the integrals in eq. (11) then leads to

$$U = \frac{1}{2} \sum_{n=1}^{N} a_n^2 \omega_n^2, \tag{12}$$

where the cross-mode terms disappear due to the spatial orthogonality of modes $\int_{\boldsymbol{x}} s_m s_n \mathrm{d}\boldsymbol{x}$ for $m \neq n$. This constraint implies that the probability $P(S)$ of many states is zero and that only the states satisfying eq. (12) will retain their original (constant non-zero) contributions.

Evaluating the cross correlation integral like in eq. (7) given the eq. (12) constraint can be done in a number of different ways. First, performing the integrals over ϕ_n again cause the cross-mode contributions to disappear, leaving an N-dimensional integral over the a_n without cross-mode terms. In the 2-mode case, one can then think of the eq. (12) constraint as taking the now 2-D integral and reducing it to a 1-D integral through the 2-D space (since all other contributions are zero), where a_1 and a_2 are constrained to lie along the ellipse with an a_1-semi-axis of $\sqrt{2U}/\omega_1$ and an a_2-semi-axis of $\sqrt{2U}/\omega_2$. The average squared coordinate values (e.g. average values of a_1^2 and a_2^2) of an ellipse are simply one half of the squared semi-axis lengths (see Appendix A2), so that $(a_1\omega_1)^2$ and $(a_2\omega_2)^2$ both have an average value of U. A similar argument can be made in the general N-dimensional case, where the remaining N-dimensional integral is constrained to lie along an N-dimensional hyperellipsoid, reducing the remaining integral to an $(N - 1)$-dimensional integral. Moreover, much like in the 2-mode case, the hyperellipsoid constraint implies that $(a_n\omega_n)^2$ has an average value of $2U/N$ (see Appendix A3), so that the constrained probabilistic value of the cross correlation is

$$\begin{aligned} C_{AB}^{P,U} &= \int_{S,U} \sum_{n=1}^{N} \frac{1}{2} a_n^2 s_n(\boldsymbol{x}_A) s_n(\boldsymbol{x}_B) \cos(\omega_n t)\mathrm{d}S \\ &= \sum_{n=1}^{N} \frac{U}{N\omega_n^2} s_n(\boldsymbol{x}_A) s_n(\boldsymbol{x}_B) \cos(\omega_n t), \end{aligned} \tag{13}$$

where $\int_{S,U}$ denotes the probabilistic integral over all states with the constraint on U, and we remind the reader that the cross-mode terms disappear due to integration over ϕ_n. The Green's function relationship in the global energy constrained case is therefore

$$\frac{\mathrm{d}C_{AB}^{P,U}}{\mathrm{d}t} = -\frac{U}{N}[G(\boldsymbol{x}_A, \boldsymbol{x}_B, t) - G(\boldsymbol{x}_A, \boldsymbol{x}_B, -t)]. \tag{14}$$

Interestingly, this cross correlation/Green's function relationship is identical to the one derived when assuming equipartition of energy (with the association $U = N\alpha^2/2$) despite the fact that negligibly small contributions to the Bayesian probabilistic cross correlation have modes with energy equipartitioned (see Fig. 1). Instead, it is the probabilistic value of the various contributions that force the average values of $a_n^2\omega_n^2$ to be equal despite the majority of the contributions not satisfying this equality. Although there is negligibly small contribution from equipartioned states, these contributions are probabilistically more likely than other contributions, and can be viewed as a justification for equipartitioned energies to be expected probabilistically when the global energy is constrained. Again, the situation is much like in the statistical mechanics case, where the Boltzmann (equipartition) result and the Bayesian result with constrained U lead to identical macroscopic predictions (Jaynes 1957, 2003). However, the Bayesian inference approach has a distinct change in philosophy in how to achieve the result, and can be generalized more straightforwardly to cases with other more detailed constraints. We turn to these other cases now.

3.3 Local energy constraint

For seismology on the Earth, a global energy constraint is rarely available, due to the inability to place seismometers everywhere on and inside the Earth. Instead, it is much more common to simply have a local pointwise constraint on the energy at the location of the sensors since the energy observed at each sensor is finite and observed. In this section, we therefore evaluate the corresponding result for when the local energy density at A is constrained. The new constraint is now

$$U_A = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \dot{u}^2(\boldsymbol{x}_A, t)\mathrm{d}t, \tag{15}$$

where U_A is the energy per unit mass at A. Assuming that the number of modes N is large enough, we show in Appendix B that the cross correlation with the local energy constraint simplifies to

$$C_{AB}^{P,U_A} = \sum_{n=1}^{N} \frac{U_A}{N\omega_n^2} \frac{s_n(\boldsymbol{x}_B)}{s_n(\boldsymbol{x}_A)} \cos(\omega_n t). \tag{16}$$

Unlike in the global energy constraint case, it is more difficult to construct an equation that explicitly links eq. (16) to the Green's function (sum of eq. 3), due to the term-wise division by $s_n(\boldsymbol{x}_A)$. A term-wise comparison can be made (see eq. B3), but this relationship is of limited usefulness since generally the terms of the cross correlation cannot be separated (see Appendix B for a discussion of this).

The deteriorating relationship between the cross correlation and the Green's function as further constraints or more restrictive constraints are enforced is expected. Depending on the nature of the constraints, the deterioration may be systematic like in eq. (B3), or it can fail in a more dramatic fashion. Nonetheless, the cross correlation can be computed, providing a theoretical expectation that can be compared with observations, as highlighted by Fichtner & Tsai (2019). In subsequent sections, we show how the Bayesian inference approach taken here is able to incorporate any number of practical constraints that include but can differ from the directional or spatial noise source distributions discussed by previous authors (Lin *et al.* 2008; Tsai 2009, 2011; Harmon *et al.* 2010; Tromp *et al.* 2010; Fichtner & Tsai 2019).

3.4 Beamformer constraints

When the modes of the system are known to be propagating waves that can arrive from a variety of different directions, then new sets of constraints can be put upon the system that depend on the known directionality of the wavefield. For example, for high-frequency surface waves within a smoothly laterally varying medium (which the Earth sometimes satisfies), it is known that waves of any given mode branch can arrive from any azimuth, θ (Tromp & Dahlen 1992); and for high-frequency body waves within a smoothly varying 3-D medium (which again the Earth sometimes satisfies), waves can arrive at the surface from a half-hemisphere of directions (often specified by polar and azimuthal directions θ and φ). Here, for simplicity we consider 2-D scalar surface waves so that the direction is uniquely specified by the angle θ, but note that 3-D waves (specified by 2 angles), vector-oriented surface waves (like Rayleigh waves), or other dimensional problems could be analysed as well.

With the above assumptions, the wavefield as in eq. (2) can then be expressed as

$$u(\boldsymbol{x}, t) = \int_0^{2\pi} a(\theta) \cos[\omega_0 t - \boldsymbol{k} \cdot \boldsymbol{x} + \phi(\theta)] \mathrm{d}\theta, \tag{17}$$

where $a(\theta)$ is the amplitude from direction θ, $\boldsymbol{k} = (\cos\theta\,\hat{\boldsymbol{x}} + \sin\theta\,\hat{\boldsymbol{y}})\omega_0/v$ is the wavevector, v is the wave speed, the modes being considered are only those at a given frequency ω_0 along a single mode branch, and the expression holds for the region $\boldsymbol{x}$ over which the directionality assumption is valid. This therefore results in the same plane-wave setup that has been discussed by many authors (Cox 1973; Tsai 2009; Harmon *et al.* 2010).

In this plane-wave setup, Cox (1973) derived the general solution for the cross correlation when the coefficients $a(\theta)$ are arbitrarily determined but with uncorrelated phases (similar to as in Section 2), which implies

$$\begin{aligned} C_{\Delta x} &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u(\mathbf{0}, \tau) u(\Delta x \hat{\boldsymbol{x}}, \tau + t) \mathrm{d}\tau \\ &= \int_0^{2\pi} \frac{a(\theta)^2}{2} \cos[\omega_0(t - \Delta x \cos\theta / v)] \mathrm{d}\theta, \end{aligned} \tag{18}$$

where $C_{\Delta x}$ denotes the cross correlation between two points separated by Δx along the x-axis. When $a(\theta)$ is constant, this immediately simplifies to $C_{\Delta x} = \pi a^2 J_0(\omega_0 \Delta x/v) \cos(\omega_0 t)$, where J_0 is a Bessel function of order zero (Tsai & Moschetti 2010). This is the manifestation of eq. (6) for the equipartitioned, uncorrelated phase state given this directional wavefield assumption (Aki 1957).

We can now contrast this 'equipartition' derivation with the Bayesian approach. In this alternative, we can initially consider the case where amplitudes $a(\theta)$ are no longer assumed to be constant but are allowed to take on any value subject to the constraint that

$$2U/\omega_0^2 = \frac{1}{2\pi} \int_0^{2\pi} a^2(\theta) \mathrm{d}\theta, \tag{19}$$

and phases $\phi(\theta)$ are no longer assumed to be uncorrelated but instead all possible values are averaged over. This case is similar to the global energy constraint of Section 3.2 but where the number of modes is infinite and implies that the average value of $a^2(\theta)$ is $2U/\omega_0^2$. The probabilistic averaged cross correlation is given by

$$C_{\Delta x}^P = \frac{1}{Z} \int_0^{2\pi} \int_0^{2\pi} \int_{\mathrm{All}\ a, a'} \int_{\mathrm{All}\ \phi, \phi'} C_{\Delta x}^{\theta\theta'} \mathrm{d}\boldsymbol{\phi} \mathrm{d}\boldsymbol{a} \mathrm{d}\theta' \mathrm{d}\theta, \tag{20}$$

where $C_{\Delta x}^{\theta\theta'} = \frac{aa'}{2} \cos[\omega_0(t - \Delta t_2' + \Delta t_1) + \phi' - \phi]$, Δt_i is the source-to-station i travel time, $C_{\Delta x}^P$ denotes the cross correlation for 2 points separated by Δx evaluated using Bayesian inference theory and Z is a normalization factor related to the number of states. Just as before, the averages over ϕ and ϕ' result in the $\phi' \neq \phi$ and $\theta' \neq \theta$ terms disappearing. Ignoring normalization difficulties related to the infinite number of modes, this simplifies eq. (20) to

$$C_{\Delta x}^P = \int_0^{2\pi} \int_{\mathrm{All}\ a} \frac{a^2}{2} \cos[\omega_0(t - \Delta x \cos\theta / v)] \mathrm{d}\boldsymbol{a} \mathrm{d}\theta, \tag{21}$$

where still the constraint eq. (19) has not yet been imposed. Using the result that the average value of a^2 is $2U/\omega_0^2$ then simplifies the result to

$$C_{\Delta x}^{P,U} = \frac{2\pi U}{\omega_0^2} J_0\left(\frac{\omega_0 \Delta x}{v}\right)\cos(\omega_0 t) \tag{22}$$

which, again, is identical to the equipartitioned, uncorrelated result when the normalization factor is appropriately defined. This would be the appropriate expectation when no beamformer data other than a total energy constraint are available.

When specific beamformer data is available, then there is additional information about the likely values of $a(\theta)$ and they are therefore no longer expected to uniformly take on any value with constant probability. If the beamformer estimate of $a(\theta)$ is $\hat{a}(\theta)$ with uncertainty $\hat{\sigma}(\theta)$, then the appropriate probabilistic weight would be to take $P(a, \phi) = \mathcal{N}(a|\hat{a}, \hat{\sigma})$, where $\mathcal{N}(x|\mu, \sigma)$ is a Gaussian with mean $x = \mu$ and standard deviation σ. This specific probabilistic weighting would then account for the data misfit between the observed beamformer output and that of the modeled beamformer with the expected cross correlation being

$$C_{\Delta x}^{P} = \frac{1}{Z}\iiint_{\mathrm{All}\ a,a'}\int_{\mathrm{All}\ \phi,\phi'} P(\boldsymbol{a}, \phi) C_{\Delta x}^{\theta\theta'} \mathrm{d}\boldsymbol{\phi} d\boldsymbol{a} \mathrm{d}\theta' \mathrm{d}\theta. \tag{23}$$

To provide a concrete example of how observed beamformer data can be used to modify the expected noise cross correlation, we consider the simplest example of a single localized source at $\theta = \theta_0$, with beamformer output $\hat{a}(\theta) = \mathcal{N}(\theta|\theta_0, \sigma_\theta)$ (and unconstrained ϕ). If the beamformer uncertainty $\hat{\sigma}$ is large, then there will be many non-zero contributions to eq. (23), making the calculation computationally challenging. If, instead, $\hat{\sigma}$ is very small, then $a(\theta)$ would be nearly deterministically determined, that is $a(\theta) \approx \hat{a}(\theta)$, with all other contributions being zero. In this simplest case, then eq. (23) simplifies to

$$C_{\Delta x}^{P} \approx \int_{\theta_0-\pi}^{\theta_0+\pi} \frac{e^{-(\theta-\theta_0)^2/\sigma_\theta^2}}{4\pi\sigma_\theta^2} \cos[\omega_0(t - \Delta x \cos\theta/v)]\mathrm{d}\theta \tag{24}$$

which must then be numerically integrated. The only case that is analytically tractable is when σ_θ also approaches zero, in which case the source is a single plane wave with correlation $C_{\Delta x}^{P} \propto \cos[\omega_0(t - \Delta x \cos\theta_0/v)]$ as expected. In all other cases, the expected cross correlation must be numerically evaluated. When beamformer data is available, we recommend that practitioners make use of a calculation like that of eq. (24) to improve the accuracy of the expected ambient noise correlation rather than comparing correlation data directly with the Green's function, even if a full Bayesian estimate like that of eq. (23) is beyond what is computationally practical.

3.5 General constraints

In addition to allowing standard beamformer constraints, the Bayesian inference approach can be used to incorporate more general information as constraints or probabilistic weights. In principle, one can impose any constraint or probabilistic weight on $a(\theta)$, and the Bayesian computation has a clear path of simply evaluating the average only over states that satisfy the constraint or with the appropriate weight. For example, one could impose the constraint that $\int_{0^\circ}^{20^\circ} a(\theta)\mathrm{d}\theta = q\int_{75^\circ}^{95^\circ} a(\theta)\mathrm{d}\theta$ [which might be constrained based on a subset of discretized beamformer observations like those of Stehly *et al.* (2006) and Harmon *et al.* (2010)] by simply performing the same integral but throwing out all contributions to the average from $a(\theta)$ that do not satisfy the constraint. While an analytic result in such a case is intractable, a computational answer for the expected cross correlation is feasible by discretizing θ, a and ϕ into an appropriate number of bins related to the desired accuracy of the solution and computing by brute force. We note that one could also have constraints on the phases $\phi(\theta)$; for example if it were known that all sources within a certain azimuth range were correlated (i.e. had the same phase), then one could impose such a constraint that would force certain cross-mode terms to remain rather than cancelling out.

An outline of the general algorithmic answer for the cross correlation with directional information could be summarized as follows: (1) For every possible $a(\theta)$, $a^{'}(\theta^{'})$, $\phi(\theta)$ and $\phi^{'}(\theta^{'})$, evaluate $C_{\Delta x}^{\theta\theta'}$ as well as any expressions E_i needed for constraints of the form $E_i = 0$. (2) Check if constraints $E_i = 0$ are satisfied. If yes, keep in average; if not, remove from average. (3) Calculate $C_{\Delta x}^{P}$ as in eq. (23) by integrating over remaining states along with any desired additional probabilistic weights to account for a non-uniform $P(S) = P(a, \phi)$. Note that computational shortcuts can be made under certain conditions. For example, as in the evaluation of eq. (21), the cancellation of $\phi^{'} \neq \phi$ terms reduces the number of computations drastically. Since the most general version of the algorithm involves far too many evaluations to be feasible on modern computers, it remains essential to use such shortcuts (like in eq. 24) if cross correlation synthetics are to be produced computationally using this method.

From a computational standpoint, then, it is interesting to note that the problem of computing the expected cross correlation is more computationally tractable when either there are almost no constraints or when nearly all constraints of a certain category are known. In the no constraint case (or when only the global energy is known), the most naive algorithm would have been very computationally intensive to carry out, but the abundant symmetries allow the calculation to be simplified dramatically to a single analytic calculation (i.e., eq. 13). On the other extreme, if all amplitudes and phases were completely constrained, then the system would be completely deterministic and again the computation reduces to a single calculation. More interestingly, if the amplitudes are all constrained (e.g. all $a(\theta)$ are known) but all phases are unknown, again the computation reduces to a single calculation [i.e. eq. (21) without the integral over all $\boldsymbol{a}$]. The hardest computational case is when partial information exists for either a or ϕ, particularly when certain symmetries are broken that preclude the ability to analytically simplify the expressions prior to computation.

When the system is explicitly treated as a 2-D spatially extended medium (e.g. for surface waves, rather than as a directional wavefield as in Section 3.4), one can also apply the information theory approach in a similar manner. Without going into the details, the same approach taken in Section 3.4 shows that if all possible source possibilities are accounted for but with a global energy constraint, then an average over these states results in the same average cross correlation as for the equipartitioned, uncorrelated state (or equivalently, for a uniform noise source distribution, Tromp *et al.* 2010; Fichtner 2014). Also as above, additional constraints can now be put upon the system related to which sources are active (e.g. Sager *et al.* 2020) and how their phases may or may not relate to one another. Such constraints or probabilistic weights could be based on data misfits between any observed quantity and its associated model prediction for each hypothetical state, as in the beamformer case. Given that this case is even less computationally feasible to calculate than the average in Section 3.4, we do not expand on this idea further. If new approaches to minimizing the computational burden of this Bayesian information theory approach are found, then it would be worthwhile to revisit this case.

4 CONCLUSIONS

We have shown that a Bayesian inference approach can be used to determine the expected cross correlation in a modal system and that this average over many states generally yields the same result as when assuming equipartitioned, uncorrelated modes, particularly in regards to how the resulting cross correlation is related to the Green's function. The Bayesian approach is found to be more flexible in allowing for the evaluation of the cross correlation under less restrictive conditions than are usually assumed. In this case, the relationship between the cross correlation and the Green's function breaks down but the expected cross correlation can still be evaluated under more relaxed assumptions. The success of ambient noise correlation seismology appears to be more robustly underpinned by these less restrictive conditions than the general conditions that appear not to be satisfied by ambient noise on the Earth (Snieder 2004; Tsai 2009; Boschi & Weemstra 2015). This Bayesian approach therefore unifies previous work on the topic (generalizing the concept of stationary phase) and provides a path forward for ambient seismic noise practitioners who must have a theoretical expectation with which to compare observations against to make the most accurate use of the seismic noise correlation data. While in some cases the amount of computation needed to evaluate the cross correlation with this approach may be infeasibly large, certain symmetries in the system sometimes allow for reduction in computational time.

Finally, we note that the Bayesian approach represents a different philosophy from the equipartition approach, and it parallels the philosophy advocated by Jaynes (1957) for deriving statistical mechanics results using an information theory approach in contrast to the Boltzmann equipartition approach that is usually taken. While the apparent dependence on the naivety or sophistication of the practitioner or observer may at first seem odd, the power of the Bayesian approach and the many different ways it can be applied with constraints that are otherwise difficult to include suggest that it is a valuable tool that will eventually help improve our ability to infer Earth structure using ambient seismic noise correlations.

5 DATA AVAILABILITY

No data or codes were used in this work.

REFERENCES

Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthq. Res. Inst.,* **35,** 415–457.

Arnold, V.I., 1963. Proof of a theorem of A.N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian, *Uspekhi Mat. Nauk,* **18,** 13–40.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178,** 1411–1436.

Boschi, L. & Weemstra, C., 2015. Stationary-phase integrals in the cross correlation of ambient noise, *Rev. Geophys.,* **53,** 411–451.

Chirikjian, G.S., 2012. Statistical mechanics and ergodic theory, in *Stochastic Models, Information Theory, and Lie Groups,* ed. Benedetto, J.J., Vol. **2,** Birkhauser.

Cox, H., 1973. Spatial correlation in arbitrary noise fields with application to ambient sea noise, *J. acoust. Soc. Am.,* **54,** 1289–1301.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton Univ. Press.

Eckart, C., 1953. The theory of noise in continuous media, *J. acoust. Soc. Am.,* **25,** 195–199.

Fichtner, A., 2014. Source and processing effects on noise correlations, *Geophys. J Int.,* **197,** 1527–1531.

Fichtner, A. & Tsai, V.C., 2019. Theoretical foundations of noise interferometry, in *Seismic Ambient Noise,* eds Nakata, N., Gualtieri, L. & Fichtner, A., Chapter 4, pp. 109–143, Cambridge Univ. Press.

Harmon, N., Rychert, C. & Gerstoft, P., 2010. Distribution of noise sources for seismic interferometry, *Geophys. J. Int.,* **183,** 1470–1484.

Jaynes, E.T., 1957. Information theory and statistical mechanics, *Phys. Rev.,* **106,** 620–630.

Jaynes, E.T., 2003. *Probability Theory: the Logic of Science,* Cambridge Univ. Press.

Jeffreys, H., 1948. *Theory of Probability,* 2nd edn, Clarendon Press.

Kittel, C. & Kroemer, H., 1980. *Thermal Physics,* W.H. Freeman and Company.

Lin, F.C., Moschetti, M.P. & Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int.,* **173,** 281–298.

Lobkis, O.I. & Weaver, R.L., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.,* **110,** 3011–3017.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100,** 12 431–12 447.

Muir, J.B. & Tsai, V.C., 2020. Did Oldham discover the core after all? Handling imprecise historical data with hierarchical Bayesian model selection, *Seismol. Res. Lett.,* **91,** 1377–1383.

Nakata, N., Gualtieri, L. & Fichtner, A., 2019. *Seismic Ambient Noise,* Cambridge Univ. Press.

Robert, C.P., Chopin, N. & Rousseau, J., 2009. Harold Jeffreys's theory of probability revisited, *Stat. Sci.,* **24,** 141–172.

Sager, K., Boehm, C., Ermert, L., Krischer, L. & Fichtner, A., 2020. Global-scale full-waveform ambient noise inversion, *J. geophys. Res.,* **125,** e2019JB018644,.

Schekochihin, A.A, 2020. *Lectures on Kinetic Theory of Gases and Statistical Physics,* http://www-thphys.physics.ox.ac.uk/people/AlexanderSchekochihin/A1/2014/A1LectureNotes.pdf, Accessed Feb 2021.

Shannon, C.E., 1948. A mathematical theory of communication, *Bell. Syst. Tech. J.,* **27,** 379–423.

Sivia, D.S., 2006. *Data Analysis: A Bayesian Tutorial,* 2nd edn, Oxford Univ. Press.

Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E,* **69,** doi:10.1103/PhysRevE.69.046610.

Stehly, L., Campillo, M. & Shapiro, N.M., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.,* **111,** B10306, doi:10.1029/2005JB004237.

Taylor, J.R., 2005. *Classical Mechanics,* University Science Books.

Tromp, J. & Dahlen, F.A., 1992. Variational principles for surface wave propagation on a laterally heterogeneous Earth - I. Time-domain JWKB theory, *Geophys. J Int.,* **109,** 581–598.

Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.,* **183,** 791–819.

Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.,* **178,** 1555–1564.

Tsai, V.C., 2010. The relationship between noise correlation and the Green's function in the presence of degeneracy and the absence of equipartition, *Geophys. J. Int.,* **182,** 1509–1514.

Tsai, V.C., 2011. Understanding the amplitudes of noise correlation measurements, *J. geophys. Res.,* **116,** B09311, doi:10.1029/2011JB008483.

Tsai, V.C. & Moschetti, M.P., 2010. An explicit relationship between time-domain noise correlation and spatial autocorrelation (SPAC) results, *Geophys. J. Int.,* **182,** 454–460.

APPENDIX A: EVALUATION OF INTEGRALS AND CANCELLATION OF CROSS-TERMS

A1 Average value of $\cos(\theta_m - \theta_n)$

Here we demonstrate that the average value of $\cos(\theta_m - \theta_n)$ is zero when integrated over all θ_m and θ_n. This can be proven by a simple change of variables as

$$\begin{aligned}\langle\cos(\theta_m - \theta_n)\rangle &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\cos(\theta_m - \theta_n)\mathrm{d}\theta_m\mathrm{d}\theta_n\\ &= \frac{1}{4\pi^2}\int_0^{2\pi}\int_{-\theta_n}^{2\pi-\theta_n}\cos(\theta)\mathrm{d}\theta\mathrm{d}\theta_n\\ &= \frac{1}{4\pi^2}\int_0^{2\pi} 0\,\mathrm{d}\theta_n = 0,\end{aligned} \tag{A1}$$

where $\langle X\rangle$ expresses the average value of X, and $\theta = \theta_m - \theta_n$. It is a straightforward generalization that $\cos(\omega_n t + \phi_m - \phi_n)$ should then also have an average value of zero, regardless of ω_n or t, thus proving the claim needed to derive eq. (9).

A2 Average squared coordinate values on an ellipse

Here we demonstrate that the average squared coordinate values of an ellipse are one half of the squared semi-axis lengths. Without loss of generality, we prove that for the circle $x^2 + y^2 = R^2$, $\langle x^2\rangle = \langle y^2\rangle = R^2/2$. The elliptical result follows immediately by appropriate scaling of each coordinate. In the circular case, one can simply evaluate the average value integral as

$$\langle x^2\rangle = \frac{1}{2\pi}\int_0^{2\pi} x^2\mathrm{d}\theta = \frac{R^2}{2\pi}\int_0^{2\pi}\cos^2\theta\mathrm{d}\theta = \frac{R^2}{2}, \tag{A2}$$

where $x = R\cos\theta$. By symmetry, also $\langle y^2\rangle = R^2/2$ where $y = R\sin\theta$.

A3 Average squared coordinate values on a hypersphere

Here we demonstrate that the average squared coordinate values of a hypersphere are equal to R^2/N where R is the hypersphere radius and N is the dimension of the hypersphere. Without loss of generality, we consider the unit-hypersphere case where $R = 1$. The hypersphere can then be defined as $x_1^2 + ... + x_N^2 = 1$. Defining polar coordinates with respect to x_1, then $x_1 = \cos\theta$. The average value of x_1^2 over the hypersphere can then be expressed as

$$\langle x_1^2\rangle = \frac{1}{S_N(1)}\int_0^{2\pi}\cos^2\theta\cdot S_{N-1}(\sin\theta)\mathrm{d}\theta, \tag{A3}$$

where $S_N(R)$ is the surface area of the N-dimensional hypersphere of radius R. While the integral in the general N-dimensional case is difficult to evaluate directly, one can make use of the symmetry between x_1 and x_i to note that $\langle x_i^2\rangle = \langle x_1^2\rangle$ for all i so that applying the averaging operator to the hypersphere definition yields $N\langle x_1^2\rangle = 1$, or $\langle x_1^2\rangle = 1/N$. This can be checked directly by integration for any particular N, for example with the 3-D sphere case yielding $\frac{1}{4\pi}\int\cos^2\theta\cdot 2\pi\sin\theta\, d\theta = 1/3$. Thus, in N dimensions, the average squared value of any coordinate is $1/N$. The hyperellipsoid case again can be derived by scaling each coordinate appropriately. This proves the claim needed to derive eq. (13).

A4 Average value of $\frac{1}{N^2}\sum\sum\cos(\theta_m-\theta_n)$

Here we demonstrate that when N is large, the average value of $\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\cos(\theta_m-\theta_n)$ is significantly more likely to be zero than away from zero, or more precisely that the number of states contributing to a value within ϵ of zero is infinitely more likely than any other range of 2ϵ as $N\to\infty$. To prove this, we first note that the problem can be reduced to proving the same is true of the single sum of $\frac{1}{N}\sum_{n=1}^{N}\cos\theta_n$ without loss of generality. Secondly, without the $1/N$ normalization, this single sum can be thought of as finding the x value after summing N unit-vectors in the complex plane with uniformly distributed polar angle θ_n, or in other words an N-step random walk in the complex plane. This in turn is known to converge to a Gaussian distribution with variance $\sqrt{N}$ due to the central limit theorem. The result is therefore proven by realizing that the peak of the normalized Gaussian at $x=0$ becomes arbitrarily larger than the value anywhere else as $N\to\infty$. We note that when the step sizes vary, as in the case where the mode shape s_n is included, the same proof still works as long as there are sufficiently many contributions with comparable magnitudes (i.e. the sum is not dominated by just a few steps, or in other words that the effective N still approaches infinity). This proves the claim needed to derive eq. (16) when many modes are degenerate.

APPENDIX B: LOCAL ENERGY CONSTRAINT DERIVATION

In the 2-mode case, eq. (15) simplifies to

$$U_A=\frac{1}{2}[a_1^2\omega_1^2s_1^2(\boldsymbol{x}_A)+a_2^2\omega_2^2s_2^2(\boldsymbol{x}_A)]\\+\delta_{\omega_1\omega_2}a_1a_2\omega_1\omega_2s_1(\boldsymbol{x}_A)s_2(\boldsymbol{x}_A)\cos(\phi_2-\phi_1),\tag{B1}$$

where the Kronecker delta δ_{mn} is equal to zero when $m\neq n$ and equal to one when $m=n$. When $\omega_2\neq\omega_1$, this constrains a_1 and a_2 to lie along an ellipse with an a_1-semiaxis of $\sqrt{2U_A}/[\omega_1s_1(\boldsymbol{x}_A)]$ and an a_2-semiaxis of $\sqrt{2U_A}/[\omega_2s_2(\boldsymbol{x}_A)]$, much like in the global energy constraint case. Following the same logic as above, when none of the modes are degenerate ($\omega_m\neq\omega_n$ for all $m\neq n$), the general case is constrained to a hyperellipsoid which implies the result

$$C_{AB}^{P,U_A}=\sum_{n=1}^{N}\frac{U_A}{N\omega_n^2}\frac{s_n(\boldsymbol{x}_B)}{s_n(\boldsymbol{x}_A)}\cos(\omega_nt).\tag{B2}$$

Unlike in the global energy constraint case, the $\omega_m=\omega_n$ case is significantly more complicated due to the additional cross-mode terms in eq. (B1) compared to eq. (12). Thus, without additional assumptions, one cannot simplify the cross correlation analytically to achieve a result similar to eq. (B2). One path forward towards such a result can be made by imposing the further constraint that N is large enough that contributions with an average value of $\cos(\phi_m-\phi_n)$ near zero are significantly more likely than contributions with an average value away from zero (i.e., closer to ±1). A similar argument is made in standard statistical mechanics that leads to the principle of entropy maximization (Jaynes 2003) and we present a derivation in Appendix A4. With this extra constraint, regardless of the precise values of s_n, again the cross-mode terms become negligible so that eq. (B1) reduces to a_n lying along a hyperellipsoid, from which eq. (B2) follows. eq. (16) restates eq. (B2).

Also unlike in the global energy constraint case, it is more difficult to construct an equation that explicitly links this final expression (eq. 16) to the Green's function (sum of eq. 3), due to the term-wise division by $s_n(\boldsymbol{x}_A)$. However, there are at least two possible routes for still obtaining a relationship. One route is to be satisfied with term-wise comparisons, in which case we can immediately write

$$\frac{\mathrm{d}C_{ABn}^{P,U_A}}{\mathrm{d}t}=-\frac{U_A}{Ns_n^2(\boldsymbol{x}_A)}[G_n(\boldsymbol{x}_A,\boldsymbol{x}_B,t)-G_n(\boldsymbol{x}_A,\boldsymbol{x}_B,-t)],\tag{B3}$$

where C_{ABn}^{P,U_A} is the nth term of C_{AB}^{P,U_A}. However, this relationship is of limited usefulness since generally the terms of the cross correlation cannot be separated. The one instance in which the terms can be separated is if all the frequencies are different (i.e. there is no degeneracy between any modes), in which case a bandpass filtering operation as discussed in Tsai (2010) separates the various contributions. The other possibility for maintaining a relationship with the Green's function is if the magnitudes of s_n are identical at A, which in the 2-mode case means $s_1(\boldsymbol{x}_A)=s_2(\boldsymbol{x}_A)$. For certain modes, the same symmetries that lead to degeneracy (i.e. not falling into the previous category of being separable by filtering) imply that the magnitudes of s_n are also identical, which means this situation may not be as unlikely as originally might have been suspected. For example, in an Earth that is laterally homogeneous, fundamental-mode surface waves of a given frequency but traveling in different lateral directions all share the same magnitude for s_n (and also share the large N requirement for high enough frequencies). For this subset of modes, then $U_A/s_1^2(\boldsymbol{x}_A)$ can be factored out, leaving a similar cross correlation/Green's function identity as in eq. (14) with U replaced by $U_A/s_1^2(\boldsymbol{x}_A)$.