AMT Review of "Systematic deviation in wavelet-based covariance estimation and implications for eddy-covariance flux calculations"

August 23, 2026

General assessment

This manuscript examines the reconstruction of variance and covariance from continuous wavelet transform (CWT) coefficients and argues that a commonly used formulation based on Torrence and Compo (1998) can introduce systematic deviations associated with normalization and discretization choices. The authors derive an alternative estimator from the CWT inner-product identity and compare the resulting formulation with the Torrence–Compo approach using sinusoidal, white-noise, and pink-noise signals, followed by an application to eddy-covariance surface turbulence observations.

The topic is relevant to *Atmospheric Measurement Techniques*. Accurate reconstruction of second-order momentum is important in atmospheric signal analysis, and normalization issues in wavelet-based variance and covariance calculations deserve careful examination. The manuscript also has several positive attributes: the presentation is generally clear, the proposed formulation is explicitly connected to the CWT energy identity, and the numerical results suggest that differences between reconstruction conventions can be non-negligible. The reported deviations of approximately 10–15% are sufficiently large to merit attention in applications involving turbulent covariance and flux calculations since such level exceeeds the insturmental noises and retrieval errors.

However, I believe that several aspects of the manuscript require substantial clarification and additional analysis before its broader conclusions can be supported. In particular, the manuscript currently moves too fast between four distinct issues: (1) an exact continuous CWT inner–product identity; (2) its numerical discretization; (3) the statistical properties of a finite–sample covariance estimator; and (4) the physical interpretation of covariance as turbulent transport under nonstationary atmospheric conditions. These levels are related but are not equivalent.

My principal concern is therefore not with the motivation for reconsidering CWT normalization, which I find worthwhile, but with the statistical and physical claims subsequently built upon the reconstruction result.

I recommend **Major Revisions**.

Major Comments

1. Mathematical reconstruction identity versus statistical estimator

The central derivation begins from the CWT inner–product identity and produces discrete expressions intended to reconstruct variance and covariance. This is mathematically reasonable as a starting point. However, the manuscript then frequently describes the resulting quantity as a statistical estimator and makes statements concerning bias, robustness, convergence, and asymptotic consistency.

These concepts need to be distinguished more carefully.

An exact relation of the form

$$\langle x, y \rangle = \frac{1}{C_\psi} \iint W_x(a,b) W_y^*(a,b) \frac{db\, da}{a^2}$$

is an identity in the continuous L^2 framework. Practical implementation introduces finite temporal support, discrete translations, a finite range of scales, numerical quadrature in scale, boundary treatment, and a particular normalization of the wavelet coefficients. A statistical application introduces still further assumptions concerning the stochastic process generating the observations.

Consequently, the following sequence:

$$\text{continuous identity} \rightarrow \text{discrete approximation} \rightarrow \text{finite-sample estimator}$$

should not be treated as a single step.

The authors state, for example, that the proposed estimator preserves "asymptotic consistency," whereas the TC formulation may lack it. If "consistency" is intended in its statistical sense,

$$\widehat{\sigma_N^2} \xrightarrow{P} \sigma^2 \quad \text{as} \quad N \to \infty,$$

the assumptions under which this result is claimed should be explicitly stated and preferably demonstrated analytically. If the authors instead mean numerical convergence of a discretized reconstruction toward a known finite–record variance, the terminology should be modified accordingly.

At present these two meanings of convergence appear to be conflated.

2. Relationship between covariance in the time, Fourier, and wavelet domains should be made explicit

The manuscript would benefit from a more careful discussion of what quantity is actually being reconstructed.

For two zero–mean stationary signals, covariance at zero lag may be written in the time domain or equivalently through the cross-spectrum,

$$\operatorname{cov}(x, y) = R_{xy}(0) = \int S_{xy}(f)\, df,$$

with the physical covariance for real signals obtained from the real part of the cross–spectrum, i.e. the integrated cospectrum.

The proposed CWT relation should therefore be presented explicitly as another representation of this same global second-order quantity.

This clarification becomes particularly important later in the manuscript, where a wavelet representation is sometimes discussed as though it defined a fundamentally different covariance under nonstationary conditions. An exact or nearly exact representation of total covariance in another basis does not, by itself, change the statistical or physical definition of the quantity being reconstructed.

I suggest that the authors clearly distinguish between:

(1) reconstruction of total covariance;

(2) decomposition of covariance by scale and time; and

(3) identification of the portion of that covariance that represents atmospheric turbulent transport.

The present manuscript rigorously addresses primarily the first question.

3. The synthetic validation should use genuine Monte Carlo ensembles

The statistical validation is currently insufficient for the strength of the claims being made.

The manuscript states that signal parameters are varied over ten values, producing what is termed a "pseudo–ensemble," and results from different signal lengths and parameter combinations are subsequently summarized using boxplots.

Changing amplitude, frequency, variance, or sample length is not equivalent to generating independent realizations from the same stochastic process. In particular, pooling multiple values of N in a single distribution makes it difficult to separate sampling variability from changes in estimator performance with record length.

For white-noise and long–memory experiments, I recommend generating a sufficiently large number of independent realizations for every combination of process parameters and N. The manuscript should then report, as functions of N,

$$\mathrm{Bias}(\widehat{\sigma^2}),$$

$$\mathrm{Var}(\widehat{\sigma^2}),$$

and preferably

$$\mathrm{RMSE}(\widehat{\sigma^2}).$$

Confidence intervals on bias or normalized error would also be valuable.

This analysis would provide substantially stronger support for statements concerning bias and convergence than the current pooled boxplots.

4. The pink-noise experiment requires reconsideration

The treatment of pink noise is potentially problematic.

The manuscript constructs a signal with spectral density proportional to $1/f$ and interprets it as a representative long–range dependent process. It then studies convergence as the sample length N increases and attributes slow convergence to the low–frequency structure of the process.

A nominal $S(f) \propto 1/f$ process requires careful treatment because its variance depends critically on the low–frequency cutoff. For a finite sampled record, the lowest represented frequency typically depends on record duration. Thus, as N changes, the simulated low–frequency content and potentially the quantity regarded as the "true variance" may change as well.

The manuscript should state exactly how pink noise is generated, how the zero–frequency and lowest-frequency components are treated, and whether the underlying stochastic process remains statistically identical as N changes.

I would strongly encourage the authors to supplement or replace this test with a well-defined stationary long-memory process, such as fractional Gaussian noise, for which the spectral behavior and variance are theoretically controlled. Several values of the Hurst exponent could provide a much cleaner assessment of the effect of increasing long–range dependence.

This issue is especially important because the pink–noise experiment provides much of the evidence used to discuss asymptotic behavior.

5. Variance validation cannot simply substitute for covariance validation

The manuscript states that detailed benchmarking is performed only for variance because the covariance expression has the same mathematical structure and that the extension to covariance was verified but is not shown.

I do not find this sufficient for a study whose title and atmospheric application focus explicitly on covariance and eddy–covariance fluxes.

While variance and covariance share related reconstruction formulas, covariance introduces additional properties that variance cannot test: relative phase, frequency–dependent coherence, sign changes in the cospectrum, and cancellation between positive and negative contributions at different scales.

The authors should therefore construct bivariate stochastic signals with a known cross–spectrum,

$$S_{xy}(f) = \gamma_{xy}(f)\sqrt{S_{xx}(f)S_{yy}(f)}e^{i\phi(f)},$$

where coherence $\gamma_{xy}(f)$ and phase $\phi(f)$ are controlled.

This would allow direct assessment of whether the proposed method reconstructs:

- positive and negative covariance;
- broadband versus narrowband covariance;
- frequency-dependent coherence;
- phase-dependent cross-spectral structure; and
- cospectral cancellation.

Such tests are much closer to the actual problem of turbulent flux estimation than univariate variance reconstruction.

6. Reconstruction of total covariance should not be equated with turbulent flux under nonstationarity

This is my most important concern from the perspective of atmospheric turbulence.

The manuscript argues that wavelet–based approaches can relax the stationarity requirement of conventional eddy-covariance analysis and allow otherwise rejected data to be retained. **This conclusion is stronger than the analysis presented in the paper can presently support.**

Consider two measured quantities decomposed schematically as

$$w = w_T + w_M, \qquad c = c_T + c_M,$$

where the subscripts T and M denote turbulent and lower-frequency/mesoscale or nonstationary components. The total covariance contains

$$\overline{w'c'} = \overline{w_T c_T} + \overline{w_M c_M} + \overline{w_T c_M} + \overline{w_M c_T}.$$

A CWT reconstruction that conserves the total inner product can accurately reconstruct the sum of these contributions, but this does not by itself establish which contribution represents turbulent vertical transport.

Wavelet localization is certainly useful for examining this question, but localization and energy conservation alone do not provide a physical criterion for separating turbulent transport from mesoscale variability, trends, transitions, coherent motions, or other nonstationary contributions.

I therefore recommend substantially moderating the claim that the proposed normalization permits nonstationary periods to be incorporated directly into flux estimation. **The manuscript instead demonstrates a potentially improved reconstruction of total covariance and provides a framework from which subsequent scale- and time-dependent turbulent analyses could be developed.**

That distinction is important both mathematically and physically.

7. The interpretation of the nonstationary eddy–covariance results is not unique

Figure 3 is useful, but I believe the interpretation given to the nonstationary cases is incomplete.

Under stationary conditions, the proposed CWT reconstruction agrees closely with conventional EC covariance, whereas the agreement deteriorates substantially for the records classified as nonstationary. The authors interpret this deterioration primarily as evidence that conventional time–averaged covariance becomes unrepresentative under nonstationarity.

That is one possible interpretation.

An equally important possibility is that the wavelet reconstruction and conventional EC calculation cease to represent exactly the same physical component of the signal because low–frequency and transient contributions enter the integrated wavelet covariance.

The present experiment does not distinguish these hypotheses.

I suggest that the authors examine the scale–resolved cospectral/wavelet–covariance contributions for several representative stationary and nonstationary cases. Showing where in scale and time the additional covariance arises would greatly strengthen the atmospheric interpretation.

At present, Figure 3 establishes disagreement under nonstationarity but does not establish which estimate should be interpreted as the appropriate turbulent flux.

8. Morlet is only approximately analytic

The mathematical development for analytic wavelets assumes that the Fourier transform of the wavelet has support only at positive frequencies. Under that condition, the factor of two used in the reconstruction follows naturally from exclusion of the negative–frequency contribution.

However, the conventional Morlet wavelet with $\omega_0 = 6$, as used in the numerical experiments, is generally treated as approximately rather than strictly analytic.

Because the central argument of this manuscript concerns relatively modest but systematic normalization errors, the distinction between exact and approximate analyticity should not be ignored.

The authors should specify the exact implementation of the Morlet wavelet used in the calculations and quantify its negative–frequency leakage. If this contribution is negligible relative to the 10–15% deviations under discussion, showing this quantitatively should be straightforward and would remove ambiguity.

More generally, the manuscript should consistently distinguish "analytic" from "approximately analytic" wavelets.

9. The statement concerning two independent Morlet coefficient components is too general

The discussion attributes part of the improved performance of the Morlet wavelet to its complex coefficients having two statistically independent components and consequently a χ^2 distribution with two degrees of freedom.

This statement requires additional assumptions.

The real and imaginary components of complex wavelet coefficients are not necessarily statistically independent solely because the mother wavelet is complex. Such results can be established for particular classes of stationary Gaussian processes and approximately analytic wavelets, but they cannot automatically be generalized to arbitrary correlated or nonstationary geophysical signals.

The authors should either state the stochastic assumptions required for this argument or soften the explanation accordingly.

10. Boundary effects and extension using neighboring observations deserve greater attention

The manuscript recognizes boundary effects and extends individual 30-min records with observations from adjacent periods so that the scales of interest remain outside the cone of influence. This requires considerably more discussion in the context of nonstationarity.

For the Morlet wavelet, the extension is reported to require approximately five times more neighboring data than for DOG2.

If the target interval is specifically selected because it is nonstationary, extending the transform far into neighboring records can import information from meteorologically different regimes. Mathematically this may reduce edge artifacts, but statistically it means that the resulting wavelet coefficients within the nominal half–hour interval are influenced by data outside that averaging interval.

The authors should quantify the temporal extent of the required padding/extension in physical units and discuss the consequences for interpretation of rapidly evolving atmospheric conditions.

A sensitivity experiment using different extension procedures would be helpful.

Additional Comments

The manuscript would benefit from a clearer notation separating the continuous scale a, discrete scale a_j, Fourier frequency, and equivalent Fourier period. Because much of the paper concerns normalization, dimensional consistency should be explicitly checked and units associated with the wavelet coefficients should be stated.

The authors should also clarify whether the variance against which their reconstructed estimates are compared is the theoretical ensemble variance of the generating process or the conventional sample variance of each particular realization. These are different reference quantities for finite stochastic samples and lead to different interpretations of "error."

For the sinusoidal experiment, it would be useful to indicate how the sinusoidal frequency is positioned relative to the discrete wavelet scales. Spectral leakage and scale-grid mismatch may contribute to numerical reconstruction errors independently of the normalization issue.

The selection of $\delta j = 0.05$ is very fine. A sensitivity analysis to scale discretization would be informative because one objective of the paper is specifically to separate normalization bias from numerical quadrature error.

The manuscript states that the DOG2 residual white-noise bias arises from imperfect spectral partitioning and overlapping filter responses. This interpretation would be strengthened considerably by plotting the summed frequency responses of the discretized wavelet filters over the scales actually used in the analysis. Such a plot could directly demonstrate whether departures from an approximately constant partition of energy explain the observed residual bias.

Finally, because the paper emphasizes implications for eddy-covariance flux calculations, I encourage the authors to distinguish consistently between **covariance reconstruction** and **flux estimation**. Their observational analysis evaluates covariance and explicitly does not convert the result to flux units. Claims concerning operational flux improvement should therefore remain appropriately qualified.

Recommendation

The manuscript raises an important methodological issue and the derivation based on the CWT inner-product identity appears worthy of further investigation. However, the current work does not yet sufficiently distinguish mathematical reconstruction from statistical estimation, and the validation is not adequate to support several of the broader claims concerning estimator consistency and nonstationary turbulent fluxes.

In particular, direct covariance validation using controlled bivariate processes, genuine Monte Carlo experiments, a better-defined long-memory test, and a more cautious interpretation of nonstationary eddy-covariance measurements are needed.

I therefore recommend **Major Revision**. With these issues addressed, the manuscript could provide a useful contribution to wavelet-based analysis of atmospheric measurements.