Author comment: response to referee n°1

Systematic deviation in wavelet-based covariance estimation and implications for eddy-covariance flux calculations

Jonathan Bitton[1], Bernard Heinesch[1], Samuel Nicolay[2] and Catherine Charles[1]

[1]Biosystems Dynamics and Exchanges, TERRA Teaching and Research Center, Gembloux Agro-Bio Tech, University of Liege, Gembloux, Belgium
[2]Department of Mathematics, University of Liege, Liege, Belgium

Correspondence to: Jonathan Bitton (jonathan.bitton@uliege.be)

We thank all referees for their careful and constructive evaluation of our manuscript. Their comments have helped us improve the clarity, methodological robustness and overall presentation of the study.

This document provides a detailed, point-by-point response to all comments. For transparency, responses are organized as follows: an overall assessment, followed by general comments.

All modifications made in the manuscript are indicated at the end of each response and introduced by the symbol "⇨". References to sections, figures and tables in the revised manuscript are highlighted in blue.

Overall assessment

"This study revisits wavelet-based variance/covariance estimation for eddy-covariance flux calculations. The authors identify that the Torrence and Compo formulation introduces 10–15% systematic bias due to normalisation issues, and propose a corrected CWT-derived estimator. The theoretical derivation is rigorous, synthetic validation is comprehensive, and the EC data application demonstrates practical relevance. The work addresses an important methodological gap in atmospheric flux measurements. Overall, the paper is well structured and clearly written. However, several minor issues require attention before publication."

We sincerely thank the referee for the positive assessment and constructive suggestions. The comments helped us clarify the scope of the atmospheric application, improve the description of the numerical implementation and substantially strengthen the synthetic validation.

The principal revisions include:

- replacing the original pseudo-ensemble with genuine Monte Carlo experiments based on 200 independent bivariate realizations per configuration and record length;
- distinguishing reconstruction relative to the moments of each finite record from comparison with the theoretical moments of the generating process;
- adding direct covariance validation for positive, negative and nominally zero covariance;
- adding fractional Gaussian noise experiments with three Hurst exponents;
- adding prescribed cross-spectral experiments with frequency-dependent coherence and phase, including cospectral cancellation;
- analyzing the discrete wavelet-filter partition and the sensitivity to scale resolution;
- adding scale-resolved and time-resolved diagnostics for the atmospheric observations;
- evaluating sensitivity to day–night transitions and boundary treatment.

General comments

Comment n°1: "***The authors claim that the TC formulation exhibits a systematic bias of approximately 10–15%, whereas the proposed CWT-based estimator is essentially unbiased. For synthetic signals, this claim is well supported. However, in the EC application, the authors treat the classical time-averaged EC covariance as the reference benchmark. The classical EC estimator itself relies on Reynolds decomposition and strict stationarity assumptions and is not an absolute reference, particularly under non-stationary conditions. A clearer distinction should be drawn between theoretical truth in the synthetic experiments and the conventional EC estimate in the atmospheric application. The non-stationary results should be framed as diagnostics of multi-scale covariance structure rather than deviations from the true flux.***"

We agree with the referee and have revised the terminology and interpretation accordingly.

In the synthetic experiments, the reconstruction can be evaluated against two well-defined references: the variance or covariance calculated directly from each finite realization and, where available, the theoretical moment of the generating process. These experiments provide quantitative assessment of reconstruction bias. In the atmospheric application, the conventional 30 min covariance is instead an operational reference calculated from the same finite record. We agree that it should not be described as the theoretical truth or as an absolute representation of turbulent flux, particularly when the statistical properties evolve within the averaging interval.

We nevertheless distinguish between the applicability of the two decompositions under non-stationarity. Conventional 30 min EC covariance relies on a block-averaged Reynolds decomposition, whose interpretation becomes problematic when the mean and higher-order statistical properties evolve within the averaging interval. The CWT is better suited to resolving such contributions because it localizes covariance jointly in time and scale. However, this advantage does not by itself determine which upper period should separate turbulent from mesoscale transport.

In the revised manuscript, we added a scale-resolved analysis of the atmospheric records. Non-stationary records showed stronger and more variable contributions at intermediate and long periods, with cumulative covariance continuing to evolve more strongly toward the 30 min cutoff. Representative cross-scalograms showed persistent long-period structures, localized events and alternating positive and negative contributions.

We now reserve the terms "bias" and "reconstruction error" for the synthetic experiments with defined references. For the atmospheric application, we use "agreement", "difference", or "divergence" relative to the conventional 30 min covariance.

⇨ Changes in Manuscript: In the revised manuscript, we have:

i. distinguished theoretical process moments, finite-record moments and operational atmospheric references in Sect. 2.2.3 and Sect. 2.3;
ii. replaced language referring to deviations from the "true flux" with a comparison of covariance reconstructions and their scale composition in Sect. 3.4;

iii. added the scale-resolved interpretation of the EC–CWT divergence (Sect. 3.4.2) and separated covariance reconstruction from the physical selection of turbulent contributions (Sect. 4.4);
iv. moderated claims in Sect. 4 concerning the direct recovery of turbulent flux under non-stationary conditions.

Comment n°2: ***"Synthetic experiments show Morlet outperforms DOG2 in convergence and frequency localisation (page 10, Figure 2), suggesting theoretical superiority. Yet in the EC application under stationary conditions (Figures 3a, 3c) (now Fig. 4), DOG2 shows closer 1:1 agreement than Morlet—an apparent reversal not adequately explained by invoking DOG2's "stronger temporal localisation." Possible hypotheses: (i) DOG2's broad frequency response may compensate for the 2/3 inertial subrange slope; (ii) Morlet's wider temporal support may increase sensitivity to residual non-stationarities in "stationary" EC periods. Further investigation (e.g., residual spectral decomposition) would clarify whether this advantage is physically meaningful"***

We thank the referee for identifying this apparent inconsistency. The revised analyses show that the synthetic and atmospheric comparisons assess different aspects of performance.

The synthetic experiments evaluate numerical reconstruction for processes with controlled statistical and spectral properties. Morlet generally shows lower dispersion and more uniform finite-scale spectral coverage. A discrete filter-partition analysis confirms that its summed response is effectively constant at $\delta j = 0.05$. DOG2 exhibits a less uniform response and a small residual white-noise bias, predicted as –1.863 % and observed as −1.863 % in the Monte Carlo experiment.

The atmospheric comparison additionally depends on temporal support and boundary sensitivity. DOG2 has a more compact temporal support than Morlet and requires fewer external data for coefficients near the 30 min upper-period limit. A boundary-sensitivity analysis shows that, under non-stationary conditions, omitting external temporal support primarily affects the 10–30 min range and has a larger effect for Morlet. Reflected extension closely reproduces the original adjacent-data calculation for both wavelets, with greater record-to-record variability for Morlet under non-stationary conditions.

These results support the following hypotheses:

i. temporal localization and boundary sensitivity are plausible explanations for the closer DOG2 agreement in the selected atmospheric records. They do not provide evidence that DOG2's broad spectral response specifically compensates for an inertial-subrange (2/3) slope.
ii. Morlet may be more sensitive to residual non-stationary structures because of its broader temporal support. However, the present design does not isolate this effect independently of covariance magnitude, time of day or atmospheric regime and we therefore present it as a consistent interpretation rather than a demonstrated causal mechanism.

We now clarify that Morlet's better numerical performance in the controlled experiments does not imply that it must show the closest agreement with block covariance in every finite atmospheric record. Numerical partition accuracy and sensitivity to localized atmospheric structure are distinct properties.

⇨ Changes in Manuscript: To address this comment, we have:

i. added the quantitative discrete filter-partition results for Morlet and DOG2 in Sect. 3.3;
ii. added the boundary-treatment sensitivity analysis for the atmospheric records in Sect. 3.4;
iii. distinguished frequency-partition accuracy from temporal localization and boundary sensitivity in Sect. 4.3;
iv. removed any implication that performance in the controlled synthetic experiments uniquely predicts agreement with conventional EC covariance in individual atmospheric records.

Comment n°3: ***"In the TC formulation (page 3), the reconstruction factor is denoted as C_δ, while in the proposed formulation (page 4), the admissibility constant is denoted as C_ψ. Although the authors indicate that these are conceptually distinct – the former being an empirical discretisation factor and the latter the theoretical admissibility constant – their numerical values differ substantially for a given wavelet. This notational proximity is likely to cause confusion. An explicit statement is needed, where C_δ is first introduced, clarifying that it is not equivalent to C_ψ, but rather an empirical normalisation scalar arising from the discretisation scheme of Torrence and Compo (1998). A small table listing the numerical values of both constants for the two wavelets used would further aid clarity."***

We thank the referee for this comment. We agree that the similarity between the notations (C_ψ) and (C_δ) may cause confusion, although the two constants have distinct definitions and roles.

The admissibility constant C_ψ arises from the continuous CWT inner-product identity and depends on the selected wavelet and normalization convention. By contrast, C_δ is the empirical reconstruction factor introduced in the original Torrence and Compo formulation for their discrete reconstruction convention. It should therefore not be interpreted as equivalent, either conceptually or numerically, to C_ψ.

We have clarified this distinction and added a table listing the constants and reconstruction factors used for Morlet and DOG2. Nevertheless, we retained the original notation for both quantities. C_ψ follows the conventional notation used in mathematical developments of the CWT, whereas C_δ is the notation introduced in the original Torrence and Compo formulation. Preserving these established symbols facilitates comparison between our derivation and the corresponding mathematical and methodological references.

⇨ Changes in Manuscript: In the revised manuscript, we have:

i. clarified the differences between both constants in Sect. 2.1.4;
ii. added a table summarizing values of both constants as well as key parameters used in this study in Sect. 2.1.6.

Comment n°4: **"*On page 6, the authors state: "Stationary and non-stationary conditions were identified using the combined steady-state and integral turbulence characteristics test (SSITC), with SSITC=0 indicating stationary conditions and SSITC=2 indicating non-stationary conditions (Foken et al., 2004)." However, the threshold values or criteria used to assign these classes are not described. It would be helpful to specify, either here or in a supplementary section, the exact numerical thresholds employed for the steady-state test and the integral turbulence characteristics test, as different implementations of the SSITC use different tolerance levels"***

We agree that the classification criteria should be stated explicitly. We used the ICOS quality classes associated with the FC_SSITC_TEST flag, rather than defining study-specific thresholds. These classes are the following:

SSITC flag	ITC deviation	SS deviation	Data quality
0	< 30 %	< 30 %	stationary or highest-quality class
1	< 100 %	< 100 %	intermediate class, excluded from both groups
2	> 100%	> 100 %	non-stationary or lowest-quality class

If the two tests yield different classes, the combined flag retains the lowest-quality class. Records with a missing FC_SSITC_TEST flag were also excluded.

⇨ Changes in Manuscript: In the revised manuscript, we have revised Sect 2.3 to:

i. clarify that the FC_SSITC_TEST flag was obtained directly from ICOS adopted classes;
iii. add the ITC and steady-state deviation thresholds associated with ICOS classes.

Comment n°5: "*The reduction from 44 (DOG2) to 42 (Morlet) non-stationary periods (page 6) is modest given Morlet requires 5× longer extensions. It should be clarified whether this reflects data availability or other filtering criteria, and the edge-effect exclusion criteria for each wavelet should be briefly explained.*"

We thank the referee for requesting clarification.

The target half-hours were selected before applying either wavelet and were therefore common to the Morlet and DOG2 analyses. The smaller final Morlet sample therefore does not result from a different initial selection of data. It reflects the longer temporal support required by Morlet and the associated requirement for valid high-frequency observations over a wider surrounding interval. At the 30 min upper-period limit, DOG2 required approx. 19.3 min on either side of the target interval while Morlet required approx. 81.0 min on either side. The resulting ratio was approx. 4.2 rather than exactly five. Because input data were loaded in complete 30 min blocks, one neighboring half-hour was loaded on either side for DOG2 and three for Morlet, although the wavelet support did not use the complete outer blocks.

The processing details are the following:

For each quality class, we initially selected the 50 largest absolute covariance records with sufficient continuous data to cover the longer temporal support required by Morlet. For the stationary group, the target half-hour and the three preceding/following half-hours were all required to belong to the stationary class, providing a continuous sequence of seven stationary half-hours, or 3.5 h in total. For the non-stationary group, only the target half-hour was required to belong to the non-stationary class. The adjacent half-hours were required to contain valid observations but were not required to be non-stationary.

The high-frequency variable used for this selection had already undergone the standard quality-control procedures, including exclusion of unsuitable wind sectors and out-of-range values, instrument diagnostics based on the site logbook, absolute-limit and rate-of-change tests, a minimum raw-data completeness requirement of 30,000 out of 36,000 samples and despiking followed by linear interpolation across detected spikes. For the CWT calculations, we applied an additional and more restrictive completeness criterion over the full wavelet-specific input support, since wavelet coefficients are sensitive to isolated spikes. The Vickers and Mahrt (1997) spike-detection procedure was reapplied, detected spikes were counted as missing samples and records were excluded when more than 1 % of the samples required by the corresponding wavelet were unavailable. Remaining gaps were filled by linear interpolation.

Consequently, the reduction from the 50 initially selected records to 49 stationary records for both wavelets and to 44 DOG2 and 42 Morlet non-stationary records arose from this additional 99 % high-frequency completeness requirement. The smaller Morlet sample reflects its longer required temporal support and therefore its greater probability of encountering missing or rejected samples within the analyzed input interval.

These processing details were abbreviated in the original manuscript because the atmospheric application was intended as a methodological comparison between covariance calculations under the two quality regimes, rather than as an operational flux-quality assessment. We agree, however, that the selection and preprocessing procedure should be sufficiently documented for reproducibility. We have therefore summarized the main criteria in the Methods.

⇨ Changes in Manuscript: To address this point, we have substantially revised Sect. 2.3 to clarify the data selection and treatment procedure.

Comment n°6: **"*The term "pseudo-ensemble" (page 6) is non-standard. An ensemble typically means multiple realisations from the same parameter setting; here it appears to mean discrete parameter sampling (frequency, amplitude, variance) instead. For stochastic signals, a single realisation itself contains sampling variability, so the "pseudo-ensemble" mixes parametric and stochastic variability. Clarification is needed on whether each parameter combination corresponds to one or multiple realisations, and justification of the statistical representativeness.*"**

We agree and have replaced the original validation design.

For each combination of signal type, process parameters, record length and wavelet, the revised analysis now uses 200 independent bivariate realizations. For a mean performance metric, this number gives a Monte Carlo standard error equal to $\frac{1}{\sqrt{200}} \approx 0.071$ times the between-realization standard deviation, providing sufficient precision to distinguish the persistent reconstruction differences examined here. Record lengths range from 2^{10} to 2^{19}. The revised experiments include:

- sinusoidal pairs with controlled amplitude, period, random initial phase and relative phase;
- Gaussian white-noise pairs with controlled variance and correlation;
- finite-record pink-noise pairs;
- fractional Gaussian noise pairs with controlled variance, correlation and Hurst exponent.

The same simulated pair is analyzed with Morlet and DOG2, allowing paired comparison between wavelets. Bias, error variance and RMSE are calculated across independent realizations for each configuration and record length.

We no longer use the term "pseudo-ensemble". Variation among process configurations is reported separately from Monte Carlo variability among independent realizations.

⇨ Changes in Manuscript: To address this point, we have:

i. replaced the pseudo-ensemble with genuine Monte Carlo experiments using 200 independent realizations per configuration in Sect. 2.2.2;
ii. defined more precisely error metrics used throughout the study in Sect. 2.2.3;
iii. replaced the pooled pseudo-ensemble summaries with convergence curves and configuration-specific Monte Carlo summaries in Sect. 3.1 and Sect. 3.2.

Comment n°7: "*In Figure 3 (now Fig. 4), a legend is provided only in subpanel (d); subpanels (a)–(c) lack one. This hinders interpretation when comparing across conditions or wavelet types. A consistent legend should be added to all subpanels, or a single comprehensive legend placed outside the panel array.*"

We agree that the graphical conventions and legends were not sufficiently consistent across the original figures. We have therefore revised the style and legends of all figures.

Multi-panel figures use either a shared legend or complete panel-specific labeling, depending on which option provides the clearest presentation. The captions have also been expanded where necessary so that the methods, reference quantities, uncertainty ranges and auxiliary lines can be interpreted without referring repeatedly to the main text.

⇨ Changes in Manuscript: To address this point, we have revised all figures and captions for consistency and clarity.