The paper proposes an extension of the classic DTS/DTS-T criterion (Qin et al., 2013), adding wavelet-based frequency-domain validation, including band power enhancement, cross-layer coherence ≥0.5, physically consistent phase lag, non-stationarity. In the intention of the authors, this extension would reduce false positives in multi-parameter earthquake precursor detection. The method is applied to the Feb 6, 2023 Turkey doublet looking at different physical parameters across different Earth layers (i.e. MBT, SLHF, OLR, GPS/GIM-TEC, Swarm Ne/Te), reporting a coherent four-stage pre→co→post-seismic LCAI cascade and claims an 85% reduction in false anomalies relative to conventional DTS (F-score 0.90 vs 0.38). The work is ambitious and addresses important questions in earthquake precursor research. However, several issues require attention.

Weaknesses / potential incongruences

1. Single Case Study Validation: The metrics are evaluated exclusively on one event (the 2023 Turkey doublet). Over-fitting thresholds/weights to a single historical event is a significant risk, and true validation requires applying DTSF blindly across multiple region-wide historical datasets. This limitation should be more explicitly mentioned. Moreover, a comparison of the present results are made with other few works on the same doublet event, but one of the most complete is missing: comparison with Cianchini et al. (2024) results is encouraged (also updating Fig. S4).

2. Circularity in the DTS-vs-DTSF benchmark. The "true positives" used to compute both methods' performance (Sect. 4.2) were themselves identified via "retrospective analysis" of cross-layer phase consistency, i.e., using essentially the same coherence/phase logic that constitutes the DTSF frequency criterion. If the ground truth is defined by criteria resembling F2/F3, then comparing DTS against that ground truth is almost guaranteed to look bad, and DTSF's superior score is partly definitional rather than demonstrated. This should need an independent (e.g., blind, or physically-modelled) ground truth to be convincing.

3. Contradictory TEC anomaly descriptions. The paper describes GIM TEC anomalies as both positive and negative in ways that seem inconsistent. The spatial and temporal transitions between positive/negative anomalies are not well-explained. The paper states negative anomalies on Jan 27-30 are due to "E×B drift draining F2-peak Ne" while positive anomalies on Feb 4-5 are due to "AGW numbers pushing GIM-TEC towards positive values." This mechanistic switch is asserted without rigorous physical justification. In addition, consider whether GIM vs. GPS station observations might have different sensitivities (e.g. GIM-TEC vs ISTA station on Feb 6).

4. Numbers in Fig. 12 don't reconcile with the text. Text states DTS: 11 TP / 36 FP out of 47 → precision = 11/47 ≈ 23%; DTSF: 11 TP / 3 FP out of 14 → precision ≈ 79%. Figure 12 reports Precision = 62% (DTS) and 92% (DTSF), and separately plots "True Positive Rate" and "Recall" as distinct bars with identical values (73/89), normally these are the same quantity for a fixed set of actual positives. Please consider that the metric definitions need to be spelled out explicitly or there's a computational/plotting error.

5. Geomagnetic filtering inconsistency. The paper claims "geomagnetically quiet conditions" but the time period includes potentially significant solar activity. In section 2.3 authors say: “space weather was stable during January to February 6", while in section 3.4.2: “solar F10.7 index was high (179.7 SFU) till February 9". My concern is that F10.7=179.7 SFU is NOT “quiet conditions”.

6. Deviation threshold moved the wrong direction for a "stricter" method. DTSF lowers the deviation threshold from the conventional ±1.5σ (Qin et al., 2013) to ±1.4σ — i.e., *more* permissive at the D-stage — while the paper's rhetorical framing is that DTSF is more rigorous. That's fine if the frequency stage compensates, but it should be justified/sensitivity-tested rather than left as an unexplained parameter choice.

7. Equations 1-5: Weight justification and variable definitions. Equation (1–5) uses arbitrary weights ($w_1 = 0.25, w_2 = 0.50, w_3 = 0.25$). While the authors perform a sensitivity analysis showing low composite score variation, the initial assignment of weights remains heuristic.

Moreover, some variables in equations 1-5 should be defined/clarified (e.g. I image that k is the kurtosis, but this is never indicated). By the way, the non-stationarity criterion F3 raises questions.

a. Why is the threshold k > 0.25 considered "non-stationary"? For a Gaussian distribution, kurtosis = 3 (excess kurtosis = 0). A threshold of 0.25 (excess kurtosis?) or 0.25 (raw kurtosis?) is not physically justified.

b. The denominator 2.5 seems arbitrary — why does kurtosis need to reach 2.75 to achieve F3 = 1?

c. Does this criterion accidentally penalize perfectly valid Gaussian-distributed anomalies that happen to be stationary?

8. Multiple-comparisons / researcher-degrees-of-freedom exposure. Anomalies are searched across 5+ parameters, several stations, many days, and multiple period bands, with no stated correction for multiple testing. Given $\pm 1.4\sigma$ is not a very stringent threshold to begin with, some fraction of the reported "hits" (even the coherence-validated ones) could arise from chance co-occurrence, particularly since real geophysical/atmospheric fields are autocorrelated (this inflates effective false-alarm rates beyond nominal Gaussian expectations). This is exactly the kind of look-elsewhere concern that undermines single-case LAIC studies generally. By the way, other authors uses more stringent criteria, e.g. 2 sigma (Piscini et al., 2017)

9. Sub-hourly coherence claim exceeds stated data resolution. Sect. 3.5 reports OLR–TEC/SLHF–TEC coherence "0.55–0.70 in the 1–3 hour band" for Feb 6–7, immediately after acknowledging that AMSR-2 (daily) sampling cannot resolve sub-daily lags and that testing the 15–45 min AGW-transit prediction "would require ≤1 hour resolution" data not available for this study. If OLR/SLHF inputs are daily means (as stated in Table 1: "2°×2° avg., daily means"), a coherence estimate in a 1–3 h band is not physically extractable from that series, this looks like either a resolution mismatch/error or needs clarification of which higher-resolution proxy was actually used there.

10. Attribution to "seismic driver" vs. mere co-occurrence with geomagnetic disturbance not fully closed. The authors themselves flag (correctly) that the Feb 6 TEC signal may be conflated with Feb 7–9 spread-F/high F10.7 (179.7 SFU) contamination and say "further analysis is required": good caveat, but this uncertainty isn't propagated into the DTSF performance statistics for that case, even though Feb 6–8 rows are central to the reported "co-seismic MBT-TEC coherence 0.70–0.85" result.

11. Exponents in the Dobrovolsky radius strain.

In Section 2.4 (lines 114–115) and Figure 3 (Phase 2 box), the spatial radius equation for the lithosphere/coversphere has an additional exponent (-0.39) that is almost opposite to those of Atmosphere (+0.20) and Ionosphere (+0.54), producing a much smaller region. By the way, what about the boundary regions, i.e. between coversphere and atmosphere, and atmosphere and ionosphere? Are there just jumps on these values? But this would be unphysical. This should be briefly explained, even if there is a reference where this topic is central.

12. "Large Te changes and reduced Ne" (lines 232 and follow ones) . These two values are dependent for some physical reasons (collisions between ions and neutrals; thermal capacity of ionospheric plasma), normally in opposite phase, so when one increases the other decreases. So it is not surprising to find this opposite behaviour.

13. Generalizability. N=1 event. Turkey is explicitly called "near-ideal" coupling conditions (shallow focus, crystalline basement, dense instrumentation), reasonable self-awareness, but it does mean the 89%/8%

headline numbers in the abstract will likely be over-quoted out of context by readers who don't reach the caveat in Sect. 5.

14. Figure 5 chronological incongruence: The top-right plot title shows "January 26" while the left shows "January 22" and middle "January 27" (chronology label issue).

15. Terminology.
"Coversphere" is still not widely adopted in geophysics. Many readers may question its necessity unless it is carefully justified.

16. Missing Baseline Period Information

The paper uses 2013-2022 baselines for anomaly detection but doesn't fully describe the baseline methodology (Section 2.2). I have some c**oncerns**:

How are seasonal cycles handled? A 10-year baseline should account for solar cycle variations

The 15-day rolling baseline is mentioned but the 10-year climatology and 15-day baseline relationship is unclear

TEC has strong solar cycle dependence, was this corrected?

17. Swarm Data Interpretation (Section 3.4.3, Table 2). The Swarm satellite data interpretation needs more rigor.

Swarm satellites at 450-500 km altitude pass over the region only occasionally. How many passes were actually included? The paper doesn't specify.

The "negative Ne-positive Te correlation" is interpreted as "abnormal upward electric fields." However, this correlation can also arise from: Solar activity effects (which were present, F10.7 = 179.7); Natural ionospheric variability; Seasonal effects.

Table 2 shows many Swarm anomalies but the DOY/DAE columns don't clearly indicate which days had Swarm passes.

18. Missing Uncertainty Propagation

The paper reports results with high precision but doesn't propagate uncertainties.

Examples:

2.5 ± 0.3 days lag: where does the ±0.3 come from?

$C = 0.71$: what's the confidence interval?

89% detection rate: what's the uncertainty on this?

You can use bootstrap methods to estimate uncertainties on coherence values and lags and report 95% confidence intervals.

References

Cianchini, G. et al. (2024). The preparation phase of the 2023 Kahramanmaras (Turkey) major earthquakes from a multidisciplinary and comparative perspective. Remote Sensing 16, 2766.
Piscini A. et al. (2017). A Multi-parametric Climatological Approach to Study the 2016 Amatrice–Norcia (Central Italy) Earthquake Preparatory Phase, Pure and Applied Geophysics 174 (10), 3673-3688.