

# **Referee Report on “Long-Lead Stratospheric Precursors of European Vegetation Dynamics and Crop Yield”**

## **General comments**

This manuscript investigates links between winter stratospheric polar vortex (SPV) variability, European surface climate, vegetation phenology, vegetation hazards, and crop yields in northern France. The authors propose two pathways: a direct influence of winter SPV conditions through winter and early-spring climate, and an indirect pathway through the timing of the final stratospheric warming (FSW).

The study addresses an original and potentially important Earth-system question. Its cross-disciplinary scope is well suited to Earth System Dynamics, and the manuscript is generally well organized. The progression from stratospheric variability to surface climate, vegetation, and agricultural impacts is easy to follow. Figures 2 and 8 communicate the proposed mechanisms effectively, and the French breadbasket provides an interesting impact-oriented case study.

My main concern is that several conclusions are stronger than the evidence currently supports. The analysis is based primarily on composites from a 41-year record, with only 13–15 years in the main SPV categories and 7–9 years in some jointly conditioned categories. The proposed pathways are plausible but not statistically separated, the hazard-risk metric needs clarification, and the crop-yield analysis is mainly descriptive. In addition, the manuscript discusses long-lead prediction without directly evaluating forecast skill.

I believe these issues can be addressed through targeted additional analyses, clearer reporting, and more cautious interpretation. I therefore recommend major revision.

## **Major comments**

### **1. Separation of the proposed SPV pathways**

The two-pathway framework is central to the manuscript, but the present composites do not clearly distinguish the direct winter-SPV pathway from the pathway mediated by FSW timing. Winter SPV strength, SSW occurrence, FSW timing, the NAO, and jet-stream position are dynamically related. Consequently, the “SPV + FSW” composites may represent compound circulation states rather than independent FSW effects.

Could the authors test whether FSW timing explains additional vegetation or surface-climate variability after controlling for winter SPV strength? A feasible analysis would be a multiple regression or partial-correlation analysis including continuous winter SPV

strength, FSW date, and one or more tropospheric circulation indices. The authors could also compare models with and without FSW timing.

If the sample is insufficient for a formal pathway analysis, the manuscript should describe these as “proposed pathways” and distinguish clearly between directly demonstrated relationships and process-based interpretations. Figure 2 or Fig. 8 could use solid arrows for tested relationships and dashed arrows for hypothesized links.

## **2. Definition and interpretation of vegetation-hazard risk**

Equation (1) requires clarification. Are `days_sample` and `days_other` accumulated numbers of hazard days, mean annual numbers of days, or proportions of eligible days? Because the SPV groups contain different numbers of years, total counts would not be directly comparable.

The exposure period also varies because hazards are counted between GSO and GSP, whose timing is itself associated with SPV state. Could differences in hazard-day counts partly reflect differences in the length of this interval?

I suggest calculating the proportion of eligible GSO–GSP days classified as hazardous for each year and then comparing these annual proportions between SPV categories. Alternatively, a count model with the number of eligible days included as an exposure term could be used. Please report absolute hazard frequencies, relative changes, confidence intervals, and sample sizes for the French breadbasket. This would make statements such as “threefold greater risk” easier to interpret.

The definitions of soil-moisture deficit and excess should also be corrected and expanded. Deficit presumably means soil moisture below the 10th percentile, whereas excess means values above the 90th percentile. Please specify the climatological period and temporal window used to calculate these percentiles.

## **3. Statistical robustness and sensitivity to individual years**

The relatively short record and small conditioned samples are important limitations. Some SPV–FSW composites contain only seven or nine years, while significance is evaluated across many grid cells and variables at the 10% level. This raises concerns about sampling uncertainty, multiple testing, and sensitivity to individual events.

Could the authors provide leave-one-year-out results for the principal regional conclusions? In particular, do the French NDVI, hazard, and yield relationships remain when 2013 or 2016 is excluded? A simple supplementary figure showing the range of leave-one-out estimates would be informative.

I also recommend using bootstrap confidence intervals rather than relying only on

deviations expressed in bootstrap standard deviations. Some assessment of spatial field significance or false-discovery-rate control would help distinguish robust regional patterns from isolated significant grid cells. Sensitivity to alternative SPV thresholds, FSW categories, and detrending choices should be reported in the Supplement.

#### **4. Phenology method and correspondence between NDVI and crops**

The GSO threshold is written as  $0.1(\text{NDVI}_{\max} - \text{NDVI}_{\min})$ . Was the seasonal baseline  $\text{NDVI}_{\min}$  included in the actual calculation? If so, the equation should be corrected. If not, the authors should explain the rationale and assess how the omission affects GSO estimates.

Please also provide the smoothing parameters, explain how multiple threshold crossings and multi-peak seasonal cycles are handled, and clarify how growing seasons beginning in the previous calendar year are treated. Because the original NDVI observations are biweekly, the uncertainty associated with deriving daily phenological dates should be acknowledged.

A feasible robustness test would be to repeat the main analysis using 20% and 30% seasonal-amplitude thresholds. Do the French and Baltic hotspots remain under these alternative definitions?

Finally, the French NDVI region is a broad rectangular area, whereas the yield data represent particular crops and administrative departments. Could the analysis be repeated with a cropland mask or with NDVI averaged over the crop-producing departments? This would provide a more direct comparison between satellite vegetation activity and crop yield.

#### **5. Crop-yield inference and long-lead prediction claims**

The crop-yield results are potentially important but currently remain descriptive. Please explain how crops and departments are aggregated, whether harvested-area weighting is applied, which crops are classified as winter and spring crops, and whether each crop series is detrended separately.

Could the authors report formal differences between weak and strong SPV years, including effect sizes and confidence intervals? A permutation test would be feasible with the current sample. Crop-specific results and leave-one-year-out estimates would help determine whether the relationship is widespread or dominated by one crop or exceptional year.

The manuscript also refers repeatedly to predictability and lead times of up to five months, but no out-of-sample prediction is evaluated. A feasible test would be leave-one-year-out cross-validation for French summer NDVI and yield anomalies. SPV-based

predictions should be compared with climatology and, if possible, predictions based on the NAO or jet-stream position. This would answer whether SPV provides additional predictive information.

If a prediction experiment is outside the intended scope, the manuscript should consistently refer to “potential predictability” or a “candidate precursor.” Claims about operational early-warning applications and five-month lead time should then be presented as hypotheses for future work rather than demonstrated outcomes.

### **Minor comments**

1. Section 3.1: “SVP variability” should be “SPV variability.”
2. GSO terminology: “bud-bust” should be “bud burst.”
3. Hazard definition: “10thth” and “90thth” should be corrected.
4. Figure 5 caption: The criterion “significant  $\alpha < 0.1$ ” appears incorrect. The intended notation is presumably  $p < 0.1$ .
5. Section 3.4, around line 333: The reference to Fig. 4k appears to be intended as Fig. 6k.
6. Section 3.4, around line 337: The discussion of precipitation refers to Fig. 3i, but Fig. 3i presents cloud cover. Fig. 3f appears to be the relevant precipitation panel.
7. Section 3.3: “The results presented in this sections” should be “The results presented in this section.”
8. Figure 7 caption: “anoamlies” should be “anomalies.”
9. Conclusion: “with to 300% relative risk change” should be revised.
10. French regional definition: Please report the longitude boundaries in a consistent west-to-east order and explain why this rectangular region was chosen.
11. French GSO result: The abstract reports a delay of up to three weeks. Please clarify whether this is a regional mean or a local grid-cell maximum, and provide the corresponding uncertainty.
12. Surface soil moisture: Please justify the use of the 0–7 cm layer for cereal impacts. If possible, include a sensitivity test using deeper or root-zone soil moisture.
13. Frost versus yield response: Weak SPV years show greater frost-after-bud-burst risk but generally positive cereal-yield anomalies. Please discuss whether the frost signal affects different vegetation types or phenological stages.
14. Crop categories: The discussion alternates between greater vulnerability of spring

cereals and losses in winter crops. Please clarify the crop-specific interpretation and ensure consistency with Fig. 7.

15. Figure 1 significance: Because the SPV groups are defined from DJF-mean SPV strength, significant DJF differences are partly imposed by the classification. Please distinguish this expected separation from independent findings later in the season.

16. 2013 and 2016: These years are useful illustrations, but they should not be described as validation of the composite relationships.

17. Relative versus absolute risk: Please accompany percentage changes with absolute values, especially where baseline hazard frequency is low.

18. Mechanism schematic: Figure 8 could distinguish observed associations from mechanisms inferred from previous literature.

19. Data availability: Please provide versions and persistent links for all input datasets, not only the code repository.

20. Code availability: The repository should include the complete preprocessing and statistical workflow, including detrending, SPV classification, phenology extraction, bootstrapping, and crop aggregation.

21. Reference formatting: Several references contain duplicated DOI prefixes such as <https://doi.org/https://doi.org/>. Please check the reference list systematically.

22. Language: The manuscript would benefit from careful proofreading for subject–verb agreement, repeated words, punctuation after semicolons, and inconsistent singular/plural usage.

## **Recommendation**

This manuscript presents a novel and potentially valuable connection between stratospheric dynamics, vegetation phenology, and agricultural impacts. The conceptual framework is interesting and relevant to Earth System Dynamics. However, clarification of the hazard metric, additional robustness tests, stronger crop-yield inference, and better alignment between the predictive claims and the analyses are needed.

I recommend major revision. The study would be substantially strengthened if the authors could demonstrate which proposed relationships are statistically independent, show that the main results are not driven by individual years, and either validate the predictive value of SPV or frame it more explicitly as potential predictability.