Response to Referees

Manuscript egusphere-2026-2058

Influence of synoptic patterns (NAO vs. WeMO) on rainfall isotopic composition in SE Iberia: A machine learning approach

Stachnik et al., submitted to Weather and Climate Dynamics

Dear Editor,

Please convey our gratitude to both Referees for their detailed reports, which have considerably improved the manuscript. Below is the list of responses and changes relating to each of the points raised by the reviewers. Referees' comments are reproduced in blue and our replies follow in black; text quoted verbatim from the revised files is shown in a shaded box, labelled with its section and with the line range of the submitted manuscript that it revises. Note that "wet season" means ONDJFM and "dry season" AMJJAS, and that all new statistics use the production methodology of the manuscript.

Response to Referee #1

Major comment 1 - The negative NAO–δ^{18}O relationship requires a stronger physical explanation

> **The most important issue concerns the interpretation of the wet-season relationship between δ18O and NAO. The manuscript identifies NAO as the dominant predictor of δ18O variability during the wet season... However, the observed relationship is negative... This result appears difficult to reconcile with the classical understanding of Mediterranean hydroclimate... If precipitation amount is not responsible for the observed relationship, then the manuscript should investigate which physical processes underlying NAO variability are controlling δ18O... Could atmospheric temperature be the main driver? The varying contributions of frontal precipitation (under NAO-) vs. convective precipitation could play a role as well.**

We fully agree that a mechanistic explanation was missing, and we have addressed it directly with two new analyses that together identify the mechanism and exclude the two most obvious alternatives, amount and temperature.

(a) Temperature is not the driver.

Following the Referee's suggestion we added ERA5-Land monthly 2 m temperature over the network and tested whether the NAO signal is a temperature artefact. In the wet season the δ^{18}O–NAOi relationship is essentially unchanged when temperature is controlled for, and it remains significant when the full predictor set is partialled out:

Wet-season δ^{18}O–NAOi	Spearman ρ	p	n

zero order	-0.41	< 0.001	153
controlling for T	-0.37	< 0.001	153
controlling for T, WeMOi, precipitation, altitude	-0.30	< 0.001	145
$\delta^{18}O$–T controlling for NAOi	+0.16	0.05	153

Table R1. Partial Spearman correlations for wet-season $\delta^{18}O$. The NAO signal survives control for temperature and for the full predictor set, whereas the temperature signal does not survive control for NAO. Full table in the new Sect. S4.

In the dry season the situation reverses, with temperature dominant (partial $\rho = +0.67$, $p < 0.001$ after removing NAO) and NAO non-significant ($\rho = +0.08$, $p = 0.59$). Temperature therefore modulates but does not explain the wet-season signal.

(b) The sign is set by the thermal character of the air masses, not by the amount effect.

Because precipitation amount is uncorrelated with wet-season $\delta^{18}O$ ($\rho = -0.01$, $p = 0.87$), the remaining coherent driver is the thermal character of the precipitation the NAO delivers here. SE Iberia lies close to the southern node of the NAO dipole, where the temperature response is not a simple function of the index: Castro-Díez et al. (2002) showed that southern European winter temperature depends on the position of the NAO centres of action, so that opposite anomalies can arise under the same NAO sign and the linear correlation with a simple index approaches zero. We obtain exactly that, a weak but significant negative NAOi–temperature correlation over the network ($\rho = -0.23$, $p = 0.004$); this weak coupling, opposite in sign to the strong positive coupling of central Europe, sets the sign through the classical temperature effect.

Two independent tracers exclude the alternatives the Referee raises. A warm, proximal Mediterranean source for the enriched negative-NAO months would require d-excess to increase towards negative NAOi, whereas both d-excess and ^{17}O-excess increase towards positive NAOi ($\rho = +0.17$, $p = 0.031$ and $\rho = +0.27$, $p = 0.007$). Sub-cloud re-evaporation, the process behind the frontal versus convective contrast the Referee suggests, is likewise excluded for the wet season, where the dual-isotope slope stays close to equilibrium (7.5) and d-excess is unrelated to rainfall amount; it is instead dominant in the dry season, where the slope falls to 6.4 and d-excess correlates with $\delta^{18}O$ ($\rho = -0.79$) and with precipitation ($\rho = +0.69$).

We present this as the most consistent mechanism with our data rather than as an established regional result, and the revised text states it explicitly: confirmation will require longer series and wider spatial coverage. We note that the NAO signal in circum-Mediterranean rainfall $\delta^{18}O$ has been reported as weak (Baldini et al., 2008; Fischer and Mattey, 2012) and that no station of those compilations lies in SE Iberia, which is the gap this network fills.

One consequence of the absence of a wet-season amount effect is worth stating explicitly, and Sect. 4.3 now does so: in this sector cold-season rainfall $\delta^{18}O$ is a circulation tracer and not a rainfall-amount tracer, so the conventional reading of Iberian archives, in which lower $\delta^{18}O$ is taken to mean wetter conditions, does not compete with the NAO relationship reported here. At monthly resolution the amount link is absent across the network ($\rho = -0.01$, $p = 0.87$) and at every individual station ($|\rho| \leq 0.19$, none significant), whereas the same collectors give ρ between -0.86 and -0.94 at every station in the dry season. Resolving the individual events does not overturn this: the wet-season amount effect is then weak although formally significant (ρ = -0.12, p = 0.04, n = 314) against -0.62 for dry-season events, and on those same samples the $\delta^{18}O$-NAOi correlation is the stronger of the two (ρ = -0.21, p < 0.001).

(c) Additional check against an independent isotope model (not included in the manuscript).

As a further test that the sign is not an artefact of our network or of our six-year window, we compared it with Piso.AI (Nelson et al., 2021), a machine-learning model trained on European GNIP data that predicts monthly precipitation $\delta^{18}O$ anywhere in Europe and was never trained on our data, which are not part of GNIP. Over the months our record and the model share (the model ends in December 2020), Piso.AI reproduces a negative wet-season $\delta^{18}O$–NAO relationship at our coordinates on the same station-month rows (ρ = -0.36, p = 0.011) against ρ = -0.60 for the measurements. Over its full 1950-2020 record at this location the relationship is non-stationary and averages to zero, which is consistent with the limitation stated above. We report this experiment for completeness and have not added it to the manuscript, since it would require introducing a further methodology.

Changes in the manuscript. A new paragraph was added at the end of Sect. 4.2.1, and the supporting partial-correlation analysis is given in the new Sect. S4 of the Supplement.

Revised text (Sect. 4.2.1, new paragraph; ll. 494-523):

> The sign of this relationship deserves explicit mechanistic attention, because more positive NAO conditions are associated with more depleted $\delta^{18}O$ even though positive NAO phases bring drier winters to Iberia, so a conventional amount effect would predict the opposite. Two candidate explanations can be excluded on the basis of our data. [...]

Revised text (Sect. S4, Supplement):

> In the wet season the $\delta^{18}O$–NAOi correlation is essentially unchanged when temperature is controlled for and remains significant when the full predictor set is partialled out: $\rho = -0.41$ (zero order, p < 0.001), −0.37 controlling for temperature (p < 0.001), and −0.30 controlling for temperature, WeMOi, precipitation amount and altitude (p < 0.001). [...]

Major comment 2 - Atmospheric temperature is not considered

> **The second limitation of the study is the complete absence of atmospheric temperature among the predictors used in the statistical and machine-learning analyses. Temperature is widely recognized as one of the principal controls on precipitation isotopes throughout the Mediterranean region... Temperature should be included among the tested predictors, or the authors should provide a strong justification for excluding it.**

Agreed and done. We added ERA5-Land monthly 2 m air temperature, area-averaged over the network box and aligned to the monthly isotopic series, so the predictor set is now NAOi, WeMOi, precipitation amount, altitude and temperature. The correlation matrix, the PCA, the station-wise correlations and the Random-Forest and SHAP models were all regenerated with temperature included. The result strengthens our initial conclusions:

Model	n	R^2 without T	R^2 with T
$\delta^{18}O$ wet	145	0.48	0.52

$\delta^{18}O$ dry	52	0.60	0.61
d-excess wet	145	0.43	0.45
d-excess dry	52	0.35	0.37
^{17}O-excess wet	88	0.20	0.19
^{17}O-excess dry	37	0.40	0.40

Table R2. Monte-Carlo cross-validated R^2 before and after adding temperature. The improvement is modest and consistent; temperature does not destabilise any model.

In the wet season NAO remains the dominant control on $\delta^{18}O$ even in direct competition with temperature (permutation importance NAO = 0.42 versus T = 0.35), consistent with the partial-correlation result above. In the dry season temperature becomes a leading control alongside precipitation amount, exactly as expected for an amount- and temperature-dominated warm-season regime. Adding temperature therefore sharpens the seasonal contrast that is the core of the paper.

We also tested the East Atlantic (EA) pattern, prompted by the same reasoning, and retain it only as a supplementary robustness check: it is non-significant for wet-season $\delta^{18}O$ and, although significant in the dry season, strongly collinear there with the retained predictors. The interplay of EA and NAO over Iberia is documented by Sanchez-Lopez et al. (2016), already in our reference list, which we now cite explicitly for this purpose.

Changes in the manuscript. The predictor description in Sect. 2.2.6 was rewritten and the EA test added; the results in Sects. 3.2.2 and 3.2.6 were updated with the new numbers; the EA test is documented in the new Sect. S5. Munoz-Sabater et al. (2021) was added to the reference list for ERA5-Land.

Revised text (Sect. 2.2.6, predictor set and temperature source; ll. 267-288):

For each isotopic target and season, models were trained using five predictors: NAOi, WeMOi, monthly precipitation amount (pcp_sum), station altitude, and air temperature (T). [...]

Revised text (Sect. 3.2.2, correlations including temperature; ll. 313-331):

Atmospheric indices display season-dependent associations. NAOi is negatively correlated with $\delta^{18}O$ during the wet season ($\rho = -0.41$, $p < 0.001$) and weakly, non-significantly, positively correlated during the dry season ($\rho = +0.08$, $p = 0.59$). [...]

Revised text (Sect. 3.2.6, wet-season model; ll. 404-432):

During the wet season (ONDJFM), the model achieved $R^2 = 0.52 \pm 0.12$ and RMSE = 1.38 ± 0.22‰ (n = 145). The wet-season models use 145 complete cases rather than the 153 available monthly observations because the WeMO index is unavailable for December 2021, which removes the eight samples of that month (one per station). [...]

Revised text (Sect. S5, Supplement (EA test)):

The East Atlantic pattern was evaluated as a candidate predictor using the monthly EA index from the NOAA Climate Prediction Center, following the same Spearman methodology. It is not significantly correlated with wet-season $\delta^{18}O$ ($\rho = -0.02$, $p = 0.83$; and -0.13, $p = 0.11$ after controlling for temperature), i.e. it adds nothing to the wet-season signal on which the main conclusions rest. [...]

Major comment 3 - The sampling strategy and temporal aggregation require clarification

> **The manuscript does not clearly explain how the isotopic dataset was constructed. Section 2.2.1 (L174-175) states that sampling was conducted every 1-2 months, whereas Section 2.2.2 (L200) refers to "event-based precipitation samples" aggregated into monthly observations. The relationship between the raw samples, event-scale precipitation, and monthly weighted averages remains unclear.**

We agree this was ambiguous and have rewritten it. The collector is a bulk (cumulative) sampler operated on an event basis: after each precipitation episode the accumulated water was collected, quantified and sub-sampled, and the collector was cleaned and re-armed for the next episode. Because episodes can span several intermittently raining days, a sample corresponds to a precipitation event and not to a fixed calendar interval. Field visits occurred every 1-2 months on average and could therefore retrieve one or several event samples per site; this is what the sentence at L174-175 referred to. Event samples were then aggregated to the calendar month by precipitation-weighted averaging, following the standard GNIP/IAEA convention.

We verified this against the archived data: reconstructing the monthly values from the event file by precipitation weighting reproduces the monthly $\delta^{18}O$ series exactly. For two months in which the weighted and arithmetic means differ by more than 1 ‰, the archived monthly value matches the weighted mean to three decimals, confirming that precipitation-weighted monthly aggregation is what was used.

We also harmonised the sample counts, which were inconsistent between the Abstract and the Results. The correct figures are 448 samples collected, of which 406 yielded valid isotopic determinations, aggregated into 205 precipitation-weighted monthly observations (153 wet-season and 52 dry-season), of which 133 have triple-oxygen data. The "154 / 448" figures in the Abstract corresponded to an earlier data version and have been corrected.

On weighted versus arithmetic means (the Referee's related point at L201).

The Referee rightly notes that precipitation-weighted means are natural for recharge studies whereas arithmetic means might better represent atmospheric variability. We retain the precipitation-weighted monthly values and now justify that choice explicitly in the text, while acknowledging the limitation the Referee points to.

Changes in the manuscript. Sect. 2.2.1 was rewritten to separate sampling resolution from field-visit frequency and to clarify the paraffin handling; the justification for weighted means was added to Sect. 2.2.2; and the Abstract counts were corrected.

Revised text (Sect. 2.2.1, sampling description; ll. 168-198):

> The collectors operated as bulk samplers on an event basis: after each precipitation episode the accumulated water was collected, quantified and sub-sampled, and the collector was cleaned and re-armed. A sample therefore corresponds to a precipitation event (occasionally a short multi-day episode of intermittent rainfall) rather than to a fixed calendar interval. [...]

Revised text (Sect. 2.2.2, weighted versus arithmetic means; ll. 199-211):

> Precipitation-weighted monthly means were retained rather than arithmetic means for three reasons: the predictors used here (circulation indices, temperature) and the isotopic target are all defined at monthly resolution, so precipitation weighting provides the consistent scale for circulation–isotope attribution; it follows the GNIP/IAEA standard, ensuring comparability with the wider network and with previous [...]

Revised text (Abstract, corrected counts; ll. 15-42):

> After extending the WeMO index forward from 2020, when instrumental measurements end, to 2025 with a physically constrained XGBoost reconstruction based on regional sea-level pressure predictors (R^2 = 0.85 ± 0.05 in cross-validation over 2010–2020), we analyze 448 rainwater samples collected between 2017 and 2023, of which 406 yielded valid isotopic measurements, aggregated into 205 precipitation-weighted monthly [...]

Major comment 4 - Limitations of the trajectory analysis

> **The HYSPLIT analysis raises some methodological questions. I am not fully convinced that a single trajectory initialized at 12:00 UTC adequately represents the atmospheric pathways associated with an entire precipitation event. In addition, the selected integration time of 96 h appears relatively short compared to many isotope and moisture-source studies... Furthermore, no moisture uptake analysis was performed... How do the authors expect the inclusion of a moisture-source analysis to affect their interpretation?**

We agree these choices deserved testing rather than assertion, and we carried out a substantial sensitivity experiment. We repeated the entire back-trajectory campaign for eight initialization hours per day (00, 03, 06, 09, 12, 15, 18 and 21 UTC, the native 3-hourly resolution of GDAS1) and for three integration lengths (96, 168 and 240 h) at all three arrival heights: 52 days × 8 hours × 3 heights x 3 lengths = 3744 trajectories, re-clustered with the identical k-means and silhouette procedure.

(i) A single 12:00 UTC trajectory is representative at 96 h.

At the 96 h integration used in the manuscript, the eight start hours agree with the day-majority regime 88.7 % of the time, and the 12:00 UTC trajectory matches the day-majority regime for 93.6 % of day-by-height groups, so only 6.4 % would be re-labelled by choosing a different hour. At the representative 1500 m level the divergence is essentially zero. An independent four-hour campaign (00, 06, 12 and 18 UTC) run through the NOAA READY web service gives the same result (9.0 % change), confirming robustness to the sampling design.

(ii) Lengthening the trajectories does not change the interpretation.

The question the Referee raises is whether a longer integration would alter the transport-regime interpretation, so we tested that directly rather than only reporting reproducibility. Repeating the tercile stratification at each length shows that the relationship between circulation phase and regime frequency is preserved: the Mediterranean–Iberian regime remains concentrated in the negative WeMO tercile, and Atlantic pathways in the positive tercile, at 96, 168 and 240 h alike. The conclusion reported in the manuscript therefore does not depend on the integration length (Sect. S6).

What does change with length is how deterministic the label of an individual event is. Agreement among the eight start hours falls monotonically, as expected once trajectories are integrated beyond the approximately four-to-five-day limit of Lagrangian predictability:

Integration length	**Mean start-hour agreement**	**Regime change vs. 12:00 UTC**
96 h	88.7 %	6.4 %
168 h	78.2 %	19.9 %
240 h	75.3 %	25.0 %

Table R3. Determinism of the day-level regime assignment as a function of integration length, from the 3744-trajectory experiment. Referee #2 raised the same question and suggested testing a longer step of 6 to 10 days; 168 and 240 h correspond to 7 and 10 days. The regime structure itself, and its relationship to circulation phase, are preserved across all three lengths.

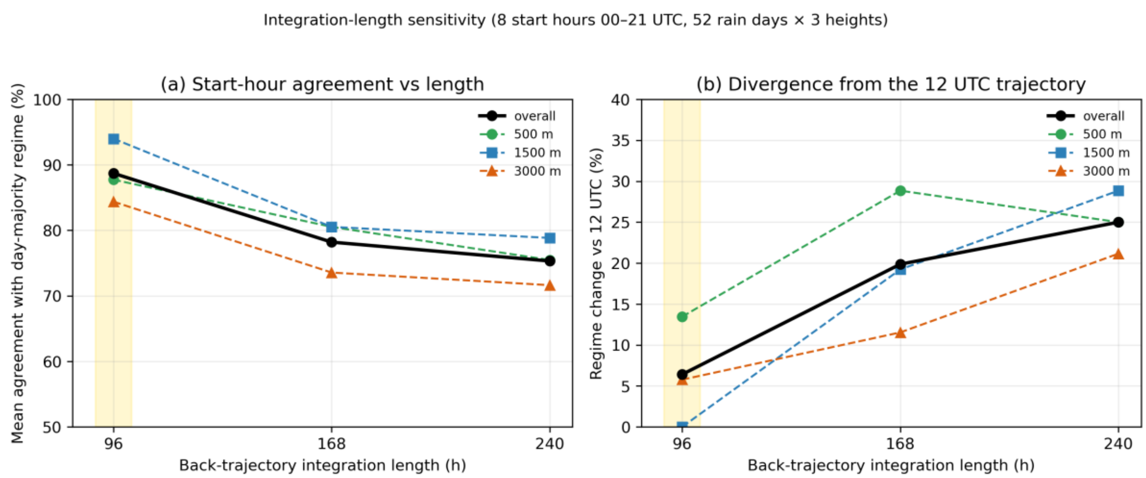

Figure R1 (new Fig. S11). Sensitivity of the transport-regime classification to trajectory design: mean agreement of the eight initialization hours with the day-majority regime, and the fraction of day-by-height groups whose majority regime differs from the 12:00 UTC trajectory, as a function of integration length, overall and by arrival height.

The number of transport regimes and the silhouette quality are essentially stable across lengths, so what lengthening the integration degrades is not the regime structure but the reproducibility of an individual event's label. We therefore retain the 96 h integration, at which a single daily trajectory remains a deterministic label for the event, while noting explicitly that the circulation-regime result does not depend on that choice. We also took the opportunity to verify that the result does not depend on the number of clusters requested, which is fixed a priori: recomputing the tercile stratification for two to six clusters leaves the relationship essentially unchanged (Sect. S6). The retained partitions are separated well beyond a shuffled null in five of the six season-by-height subsets; the exception is the wet-season subset at 500 m, from which we draw no conclusions, consistent with boundary-layer flow being more locally steered.

(iii) Moisture-uptake analysis and scope.

We agree that a Lagrangian moisture-uptake or source diagnostic would add valuable source attribution, and we now say so explicitly. Our method deliberately classifies the geometry of synoptic transport, in order to relate rain days to recurrent circulation regimes, rather than the quantitative moisture-source contribution; the two are complementary. A full uptake analysis would refine where moisture is taken up along each pathway but would not change the regime-frequency result that underpins our circulation interpretation.

Changes in the manuscript. Sect. 2.2.5 now states the scope of the method, the validation of the design and the justification of the number of clusters; the full experiment, the cluster-number sensitivity and the null-model comparison are documented in the new Sect. S6 with Fig. S11 and Tables S6 to S8. We also corrected a statement in Sect. 3.2.5: the dry-season 1500 m subset was described as collapsing to a single regime, whereas this follows from the number of clusters requested together with the small dry-season sample, and is not an absence of structure. The 3744 trajectories and the driver, clustering and validation scripts were added to the data repository.

Revised text (Sect. 2.2.5, scope of the method; ll. 236-266):

> The method deliberately characterizes the geometry of synoptic transport rather than quantifying moisture uptake along the pathway; a Lagrangian moisture-source diagnostic would refine where moisture is taken up and is a natural direction for future work, but it is complementary to, and would not alter, the regime-frequency results used here. Only dates with available isotope observations were considered.

Revised text (Sect. 2.2.5, validation of the 12:00 UTC / 96 h design; ll. 236-266):

> This design was validated by a sensitivity experiment in which the entire campaign was repeated for the eight initialization hours resolved by GDAS1 (00, 03, 06, 09, 12, 15, 18 and 21 UTC) and for three integration lengths (96, 168 and 240 h) at all three arrival heights, yielding 3744 trajectories that were re-clustered with the identical procedure (Sect. S6, Fig. S11). [...]

Revised text (Sect. S6, Supplement):

> Lengthening the integration degrades, rather than improves, the determinism of the day-level regime assignment: start-hour agreement falls from 88.7 % at 96 h to 78.2 % at 168 h and 75.3 % at 240 h, while the fraction of groups whose regime differs from the 12:00 UTC assignment rises from 6.4 % to 19.9 % and 25.0 % (Fig. S11). [...]

Major comment 5 - Physical interpretation of the transport-regime results

> **Several conclusions derived from the trajectory clustering are difficult to reconcile with the expected physical meaning of NAO. In particular, the manuscript repeatedly suggests that Atlantic transport regimes may be favoured under both NAO+ and NAO- conditions. This appears inconsistent with the classical interpretation of NAO, according to which the Atlantic advection is promoted during NAO- conditions and deserves further explanation.**

We agree the original wording overstated a one-to-one mapping and have reconciled it. The key point is a distinction of scale: the NAO index is monthly, whereas each transport regime is assigned to an individual rain day, so a month of a given phase contains rain days with a distribution of trajectory geometries. The tercile analysis therefore describes how the phase shifts the relative frequency of regimes, not a deterministic control of every event. A second point is that our "Atlantic" category spans more than one advection geometry, zonal and south-westerly, both Atlantic in origin, so its appearance under positive NAO reflects zonal, northward-displaced Atlantic flow rather than a contradiction of NAO dynamics. The start-hour sensitivity result above further shows that day-level regime labels are robust at 96 h, so the frequency statistics are meaningful.

Changes in the manuscript. The relevant passage in Sect. 4.2.1 was rewritten; the corresponding sentences flagged at L518-521 and L546 were revised in the same way.

Revised text (Sect. 4.2.1, regime frequency; ll. 494-523):

> These statistics describe how the circulation state shifts the relative frequency of regimes and not a deterministic control of individual events, since the indices are monthly whereas each transport regime is assigned to a single day with rainfall, so that a month of a given phase contains rain days with a distribution of trajectory geometries. [...]

Minor comments

The following minor points are addressed in the order raised. Where the change is a short reformulation we describe it; where it introduces new content we quote it.

- L58-61 (broader Mediterranean perspective). Agreed. The Introduction now situates the question within the wider Mediterranean, citing the suggested works.

  Revised text (Sect. 1, Introduction; ll. 43-110):

  > This question is part of a broader Mediterranean picture, in which recent work has emphasized that precipitation variability is dominated by high temporal variability rather than by monotonic trends (Vicente-Serrano et al., 2025) and is governed by atmospheric mechanisms that extend beyond the NAO alone (Luppichini et al., 2026), while isotope monitoring and speleothem studies from other Mediterranean sectors [...]

- L75 (Luppichini et al., 2026). Agreed; added where the NAO-Mediterranean precipitation relationship and its complexity are introduced, and again in the new mechanistic paragraph in Sect. 4.2.1, where it supports the argument that Mediterranean precipitation variability is not reducible to the NAO alone.
- L100 and throughout ("isotopy"). Agreed. The term has been replaced throughout with standard terminology (isotopic composition, precipitation isotopes, isotope variability), including in five section headings: Sects. 2.2, 2.2.4, 2.2.6, 3.2.4 and 4.2.3.
- L174 (citation and paraffin oil). Corrected to Natali et al. (2024). We also clarify that the paraffin oil was removed before analysis; see the quoted rewrite of Sect. 2.2.1 under major comment 3 above.
- Figure 1. On reflection we have not added a line to the panel-A map, and we explain why. Panel B is an idealized profile that orders the collector sites by elevation along the NW–SE gradient, rather than a geometric cross-section along a specific line; marking a precise transect on the map would therefore imply a spatial precision the profile does not represent. The eight collector sites are, however, already plotted on the panel-A map and labelled consistently (RC1–RC8) across both panels, and the panel-B caption now states explicitly that the profile runs from the north-westernmost site (RC8, 560 m a.s.l.) to the south-easternmost site (RC1, 1800 m a.s.l.) as plotted in panel A, so the correspondence between the two panels is unambiguous without adding a line.
- L193 (calibration; GISP). Clarified, following the Referee's point that GISP is normally used as an independent quality-control standard rather than as a calibration point.

- L197 (number of injections). Clarified: ten injections per sample, the first three discarded to remove memory effects, so reported values are averages of the last seven. We also report the standard-based precision from a MILIQ standard analysed alongside the samples.

  Revised text (Sect. 2.2.1, calibration and injections; ll. 168-198):

  > Triple-oxygen isotopes ($\delta^{17}O$, $\delta^{18}O$ and ^{17}O-excess) were normalized using a two-point V-SMOW–SLAP calibration following Schoenemann et al. (2013), which considers ^{17}O-excess = 0 for V-SMOW and SLAP, whereas $\delta^{2}H$ used a three-point V-SMOW–SLAP–GISP calibration. [...]

- L201 (weighted vs. arithmetic means). Addressed under major comment 3, with the justification quoted there.
- L308 ("monthly observations" ambiguous). Disambiguated by the rewrite of Sects. 2.2.1 and 2.2.2.
- Sect. 3.2.2 (NAO–$\delta^{18}O$ discussion). Expanded per major comment 1. We agree the positive $\delta^{18}O$–WeMO relationship is physically more consistent, and the two indices are now contrasted explicitly.
- L356-357 (opposite NAO–$\delta^{18}O$ sign between high- and low-altitude stations in summer). We agree this needed care, and we have now quantified it rather than describing it: none of the dry-season per-station NAOi correlations is individually significant, they rest on five to ten monthly values per station, and the signs are evenly split across the network (four positive, four negative). The passage is tempered accordingly, and the contrast is no longer interpreted.

  Revised text (Sect. 3.2.4; ll. 350-374):

  > None of these dry-season NAOi correlations is individually significant and they rest on few monthly values per station (five to ten), with signs evenly split across the network; the apparent altitudinal contrast is therefore not robust and is not interpreted further. Correlations with WeMOi range from −0.50 to 0.03. [...]

- L371-374 (dry-season precipitation vs. d-excess and ^{17}O-excess). We agree the positive relationship is physically plausible and now state the interpretation explicitly.

  Revised text (Sect. 3.2.4; ll. 350-374):

  > The positive dry-season association between precipitation amount and both excess parameters is physically plausible: under low ambient humidity, small and intermittent events undergo stronger sub-cloud evaporation, which lowers d-excess and ^{17}O-excess, whereas more intense convective rainfall reduces this effect (Froehlich et al., 2008; Natali et al., 2022).

- Sect. 3.2.5 (differences among 500, 1500 and 3000 m). A short explanation has been added.

  Revised text (Sect. 3.2.5; ll. 375-403):

  > The three arrival heights sample different layers of the flow: 500 m reflects boundary-layer transport that is more strongly steered by local topography, 1500 m represents low-free-tropospheric transport and is used as the reference level, and 3000 m samples free-tropospheric flow. The increasing number of retained regimes with height, and the minimal start-hour divergence at 1500 m (Sect. [...]

- L473 (d-excess seminal and recent studies). Agreed; Froehlich et al. (2008) and Natali et al. (2022) have been added and are also used for the sub-cloud interpretation quoted above.
- L501 (recent Mediterranean isotope studies). Agreed; Natali et al. (2023, 2025) added.

- L518-521 and L546 (Atlantic regimes under NAO+). Reconciled under major comment 5.
- L536 (reduced altitude effect in summer). Agreed; Natali et al. (2022) added, which reports a similar summer reduction of the isotopic altitude effect.
- L566-571. Agreed; Natali et al. (2021 and references therein) added as consistent Mediterranean context.
- L600 (future-climate implications). These rest on the wet-season NAO–$\delta^{18}O$ relationship, which is now mechanistically underpinned; Sect. 4.3 cross-references the new mechanism and the partial-correlation evidence, so the passage no longer relies on an unexplained correlation.
- L616. Agreed; Columbu et al. (2025) added as western-Mediterranean support for this interpretation.

Response to Referee #2

Major comment 1 - Sections 2.1 and 3.1 should be a separate publication or placed in the supplement

> **I understand the high interest in extending the WeMOi to 2021-2025, and that it required specific and non-trivial work. I believe this is an important matter, and that the method described by the authors is appropriate and useful. However I think Sections 2.1 and 3.1 should either be described in a separate publication or placed in the supplement, because they are far from the focus of the article on the drivers of precipitation water isotopes in Iberia.**

We agree that the technical details should not interrupt the flow of the paper and have restructured accordingly. We have only maintained a condensed account in the main text rather than moving the reconstruction entirely to the Supplement, for one decisive quantitative reason: the published WeMOi ends in December 2020, whereas 141 of our 205 monthly isotope observations, that is 69 %, postdate that endpoint. Every wet-season WeMO-isotope statistic in the paper therefore depends on the reconstructed index, and we believe the reader should be able to assess its origin and skill (10-fold cross-validated R^2 = 0.85 ± 0.05, RMSE = 0.41 ± 0.07) without leaving the main text. Relegating it entirely to the Supplement would place a key source of uncertainty affecting half of the paper's circulation analyses out of view. We also note the Referee's own assessment that the extension "may prove useful for a wide range of future studies in the Western Mediterranean"; a concise main-text account preserves its visibility for that purpose.

Changes in the manuscript. Sects. 2.1 and 3.1 were condensed to the motivation, a short outline of the method and the validation figure. The technical detail (predictor construction, station handling, hyperparameter search space and validation) moved to the new Sect. S2 and Table S5, and the discussion in Sect. 4.1 was shortened from five paragraphs to three. The mentions of the reconstruction in the Abstract (ll. 22-27) and in the Conclusions (ll. 633-640) were shortened in proportion, so that the space given to the WeMOi there matches the space it now occupies in the body.

Revised text (Sect. 2.1, motivation and pointer to the Supplement; ll. 112-166):

> This extension is a prerequisite for the present study rather than an ancillary result: the published WeMOi ends in December 2020, whereas 141 of our 205 monthly isotope observations (69 %) postdate that endpoint, so every wet-season WeMO–isotope statistic reported here depends on the reconstructed index. [...]

Revised text (Sect. S2, Supplement (predictors)):

> Predictors. Monthly mean sea-level pressure (SLP) was compiled for the AEMET San Fernando Observatory (Spain) and for four northern Italian stations of the ARPAV network (Malo, Venezia, Teolo, Cavallino) over 2010–2025. [...]

Revised text (Sect. S2, Supplement (model and validation)):

> Model and validation. A gradient-boosted regression tree model (XGBoost; Chen and Guestrin, 2016) was trained with a squared-error objective. Hyperparameters were selected by randomized search over sixty candidate configurations evaluated with 5-fold cross-validation, minimizing cross-validated mean squared error; the search space is given in Table S5. [...]

Major comment 2 - Sampling frequency (Sect. 2.2.1 and Table 2)

> **The sampling frequency of precipitation is not clear to me... In Section 2.2.1, I suggest clarifying the sampling frequency. In Table 2, I suggest adding the number of valid months sampled for each season.**

The Referee's second reading is correct, and we apologise for the ambiguity. The collector is a bulk sampler operated on an event basis, and the 1-2 month interval refers to field visits, each of which could retrieve one or several event samples. The rewritten Sect. 2.2.1 is quoted under Referee #1, major comment 3. Table 2 now reports the number of valid sampled months per station and season, as requested:

Station	Altitude (m)	Months (wet)	Months (dry)	Months (total)	Valid event samples
RC1	1800	17	6	23	41
RC2	1600	16	5	21	45
RC3	1450	17	6	23	50
RC4	1274	23	10	33	70
RC5	1024	19	7	26	52
RC6	790	20	8	28	52
RC7	600	19	5	24	46
RC8	560	22	5	27	50
Total		153	52	205	406

Table R4. Valid sampled months per station and season, now incorporated into Table 2.

Changes in the manuscript. Table 2 now includes the number of valid months per season alongside the number of samples collected at each site, and the X and Y columns were replaced by geographic longitude and latitude converted from the original ETRS89 / UTM zone 30N positions, as also requested in the Referee's minor comments. Table R4 below breaks the same network down by station, counting the 406 samples that yielded a valid isotopic measurement rather than the 448 collected.

Revised text (Table 2 caption; ll. 178-183):

> Lon and Lat are geographic coordinates (WGS84, decimal degrees) converted from the original ETRS89 / UTM zone 30N positions, and altitude is given in metres above sea level. The last two columns report the number of calendar months with valid isotopic determinations at each site, separated by season (wet: ONDJFM; dry: AMJJAS).

Major comment 3 - How are spatial and temporal drivers combined?

> **In most analyses of the article, starting in Section 2.2.4, I don't understand how spatial and temporal drivers of water isotopes are combined... When looking at correlation or PCA with regard to elevation, does it mean that the correlation is performed on the 8 stations for all the months of the considered season?... If this is the case, I do not think that spatial and temporal correlations can be fairly compared or combined. If not, can you clarify the method?**

The Referee has understood the design correctly, and we agree it needed to be stated explicitly and defended quantitatively. Our correlation matrices and models operate on the pooled station-month panel, in which each row is one station in one month (n = 153 wet, 52 dry). In this design the circulation indices and regional temperature vary only in time, altitude varies only in space, and precipitation varies in both. Altitude is included precisely so that the spatial dimension is represented explicitly rather than confounded, and this is also why the per-station analysis of Fig. 4 exists, in which altitude is constant and the correlations are purely temporal.

The legitimate concern is whether the pooled temporal correlations could be artefacts of mixing spatial and temporal variance. We have now tested this directly with a new analysis that collapses the spatial dimension entirely: the eight stations are aggregated into a single precipitation-weighted network-mean monthly series, and the key correlations are recomputed on this purely temporal series. Because months in which few stations reported are noisy estimates of the network mean, we show the result as a function of minimum monthly coverage.

Months with at least N stations	n months	δ^{18}O–NAOi ρ (p)	δ^{18}O–WeMOi ρ (p)
1 (all months)	24	-0.36 (0.087)	+0.04 (0.86)
4	20	-0.49 (0.029)	+0.19 (0.43)
6	17	-0.42 (0.090)	+0.41 (0.12)
8 (complete network)	11	-0.76 (0.007)	+0.55 (0.098)

Table R5. Network-mean (spatially collapsed) wet-season correlations as a function of monthly station coverage. For fully sampled months the purely temporal correlation (-0.76) is stronger than the pooled value (-0.41), indicating that pooling dilutes rather than inflates the signal.

A second, independent test uses the per-station correlations, each of which is a purely temporal statistic. In the wet season, δ^{18}O–NAOi is negative at all eight stations and δ^{18}O–WeMOi positive at all eight (exact two-sided binomial sign test, p = 0.008 in each case). In the dry season, δ^{18}O–precipitation is negative at all seven stations with sufficient data (p = 0.016) and δ^{18}O–temperature positive at all eight (p = 0.008), whereas δ^{18}O–NAOi splits four positive and four negative, consistent with the absence of a dry-season circulation control. Every headline conclusion of the paper therefore replicates station by station.

On the three methods (Sects. 3.2.2, 3.2.3 and 3.2.6).

We agree the complementary role of each tool should be explicit rather than implied. The correlation analysis establishes the bivariate associations and their significance; the Random Forest and SHAP analysis quantifies multivariate, potentially non-linear predictor importance with the predictors in direct competition, which bivariate correlations cannot do given the collinearity among indices; the PCA describes the covariance structure of the predictor space but adds no attribution beyond the other two. Following the Referee's suggestion we have moved the PCA to the Supplement.

Changes in the manuscript. Sect. 2.2.4 now describes the pooled design and the role of altitude and points to the new Sect. S7; a division-of-labour sentence was added at the start of Sect. 3.2; the PCA moved to Sect. S3 with Fig. S10 and Table S1, leaving a pointer in Sect. 3.2.3.

Revised text (Sect. 2.2.4, pooled station-month design; ll. 220-235):

> All correlations and models operate on the pooled station-month panel, in which each row is one station in one month (n = 153 wet and 52 dry). In this design the circulation indices and regional temperature vary only in time, altitude varies only in space, and precipitation varies in both; altitude is included explicitly so that the spatial dimension is represented rather than confounded. [...]

Revised text (Sect. 3.2, complementarity of the methods; ll. 304-432):

> The following analyses are complementary rather than redundant: the correlation analysis establishes the bivariate associations and their statistical significance, the PCA (Sect. [...]

Revised text (Sect. 3.2.3, pointer replacing the PCA section; ll. 332-349):

> A principal component analysis (PCA) of the predictor set was performed separately for each season to describe the covariance structure of the predictor space. Because it provides a description of predictor covariance rather than an attribution of isotopic variability, and its conclusions are reproduced by the correlation and Random-Forest analyses reported here, the PCA is presented in full in Sect. S3 (Fig. [...]

Revised text (Sect. S7, Supplement (network mean)):

> The wet-season $\delta^{18}O$–NAOi relationship survives this collapse and strengthens as coverage improves: $\rho = -0.36$ ($p = 0.087$) using all months, -0.49 ($p = 0.029$) for months with at least four reporting stations, and -0.76 ($p = 0.007$) for the eleven months in which the complete network reported. [...]

Revised text (Sect. S7, Supplement (per-station sign test)):

> A complementary test uses the per-station correlations of Fig. 4, each of which is a purely temporal statistic. In the wet season, $\delta^{18}O$–NAOi is negative at all eight stations and $\delta^{18}O$–WeMOi positive at all eight (exact two-sided binomial sign test, $p = 0.008$ in both cases). [...]

Major comment 4 - Significance of correlations

> **In addition to my comments above, the significance of correlations must be added next to each correlation coefficient in Section 3.2.2 and Section 3.2.4. I suppose that correlation coefficients as low as -0.01 (L319) are not significant. Significance should also be added in Fig. 3 and Fig. 5.**

Agreed and done. Figures 3 and 4 (the latter being the former Fig. 5) have been regenerated with significance markers on every cell, and the underlying p-value matrices are provided in the data repository. Every correlation quoted in the text now carries its p-value. The Referee is correct that the wet-season $\delta^{18}O$–precipitation correlation of -0.01 is not significant ($p = 0.87$); this non-significance is itself a central result, because the absence of a wet-season amount effect is what motivates the circulation-based interpretation. We have therefore made it explicit rather than leaving it implicit.

Revised text (Sect. 3.2.2, significance of the amount effect; ll. 313-331):

Precipitation amount shows no significant correlation with isotopic composition during the wet season (δ^{18}O–pcp_sum: ρ = −0.01, p = 0.87) and a strong, significant negative correlation during the dry season (δ^{18}O–pcp_sum: ρ = −0.74, p < 0.001). [...]

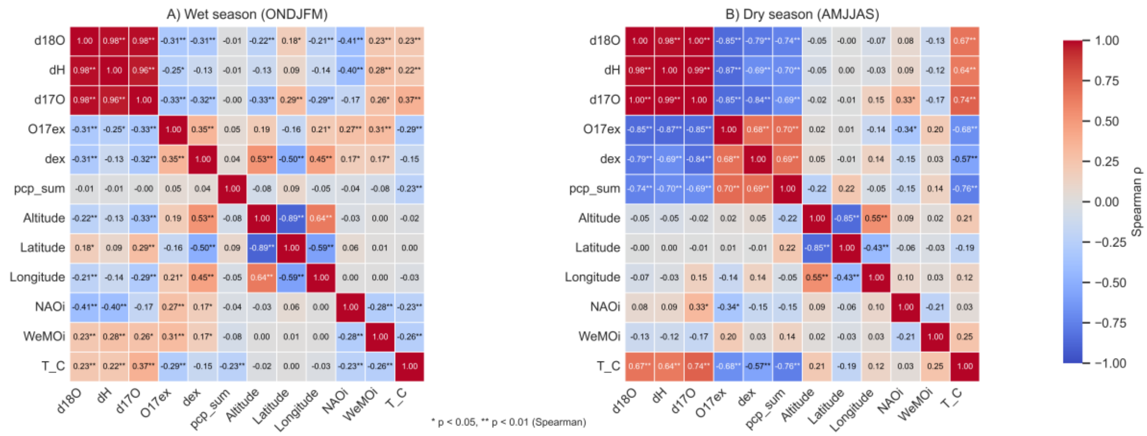

Figure R2 (revised Fig. 3). Seasonal Spearman correlation matrices, now including temperature and with significance markers (p < 0.05, ** p < 0.01).*

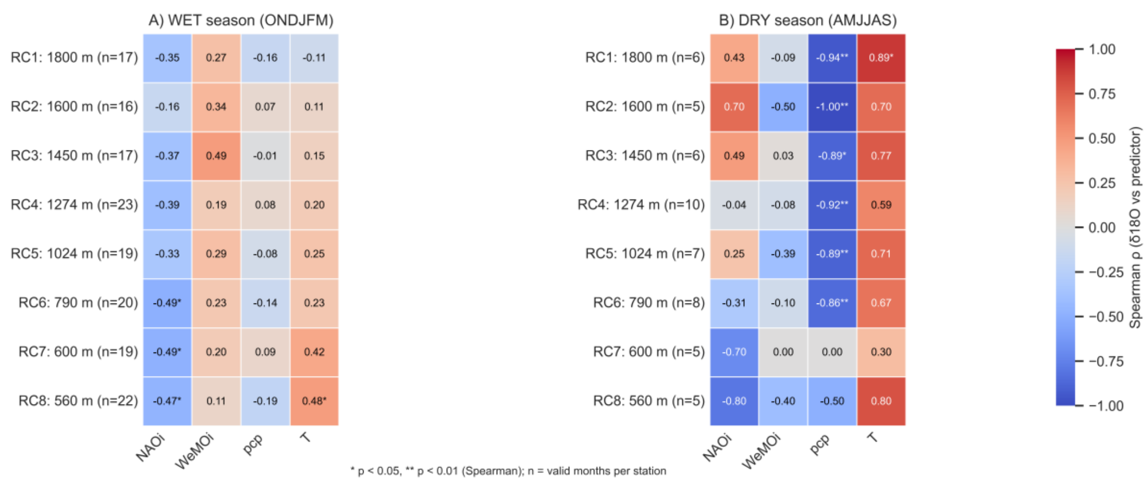

Figure R3 (revised Fig. 4, formerly Fig. 5). Station-wise correlations, now with significance markers and the number of valid months per station and season in the row labels.

Changes in the manuscript. Figs. 3 and 4 were regenerated with significance markers, and Fig. 4 additionally reports per-station sample sizes; Figs. S1 and S2 likewise. P-values were added to all correlations quoted in Sects. 3.2.2 and 3.2.4, and the captions state the significance convention.

Major comment 5 - The Discussion largely repeats the Results

> **In Section 4 (Discussion), a large part of Sections 4.1 and 4.2 repeats the results, which can be seen as a conclusion rather than a discussion. Then, the summary of these results is repeated again in Section 5 (Conclusion), in a shorter version. I suggest either combining the discussion and conclusion and merging the information... or removing the lengthy repetitions of results in the Discussion section... In either case, I suggest removing the repetitions and significantly shortening the "Discussion and Conclusion".**

We agree and have shortened both. The paragraphs that opened Sects. 4.2.1 and 4.2.2 by restating results now begin directly from the interpretive point they add; the WeMOi discussion in Sect. 4.1 was condensed from five paragraphs to three; and the Conclusions were compressed. We have kept the section structure and all interpretive content, removing repetition rather than substance, so that the argument remains traceable for readers who do not read the Results in full. Two illustrative rewrites:

Revised text (Sect. 4.2.1, opening (result restatement removed); ll. 494-523):

> The wet-season dominance of the NAO, with the WeMO as a secondary modulating mode, is expressed consistently across the network, while the comparatively weak correlations with monthly precipitation amount are consistent with wet-season $\delta^{18}O$ being organized primarily by synoptic moisture delivery and upstream rainout history rather than by month-to-month fluctuations in local totals under predominantly stratiform [...]

Revised text (Sect. 4.2.2, opening (result restatement removed); ll. 524-550):

> The dry-season shift toward precipitation-regime dominance is consistent with amount-effect behaviour under convective precipitation and strong event intermittency. [...]

Changes in the manuscript. Sects. 4.1, 4.2.1, 4.2.2 and 4.2.3 were trimmed of result restatements and the Conclusions compressed; the two PCA paragraphs of the Discussion were merged into a single sentence pointing to Sect. S3. No interpretive content was removed.

Major comment 6 - Sect. 4.3 and concrete examples of dry-season archives

> **I really appreciated Section 4.3. In my view, it could be the central part of the Discussion by significantly reducing Sections 4.1 and 4.2, as suggested above. Can you provide concrete examples of archival materials integrating dry-season conditions?**

We thank the Referee for this suggestion, which we have adopted: with Sects. 4.1 and 4.2 shortened, Sect. 4.3 now carries more of the weight of the Discussion. On the specific question, we now give concrete examples of archive types whose isotopic signal is biased toward dry-season conditions in semi-arid Iberia, and contrast them with the wet-season recharge archives.

Changes in the manuscript. Sect. 4.3 was expanded with the archive examples and their season-of-integration logic.

Revised text (Sect. 4.3; ll. 589-631):

> Concrete examples of such archives include lacustrine gypsum hydration water, which records lake-water composition during warm-season evaporative concentration and gypsum precipitation (Gázquez et al., 2020); pedogenic and land-snail carbonates, whose formation is biased toward the

> warm season and toward evaporatively enriched soil and body water (Breecker et al., 2009; Yanes et al., 2013); and tree-ring cellulose [...]

Minor comments

- Back-trajectory length (4 days is short; consider one longer step of 6 or 10 days instead of the shorter steps). This experiment was carried out in full and is reported under Referee #1, major comment 4: we tested 96, 168 and 240 h, that is 4, 7 and 10 days, exactly the range suggested. The circulation–regime relationship reported in the manuscript is reproduced at all three lengths, so the conclusion does not depend on this choice; 96 h is retained because the label of an individual event is more reproducible there (Table R3, Fig. R1, Sect. S6).
- L247 (the choice of 12:00 UTC should be adjusted to the most intense time of the event). The same experiment addresses this: across the eight 3-hourly initialization times the 12:00 UTC trajectory matches the day-majority transport regime in 93.6 % of day-by-height cases at 96 h, and essentially always at the representative 1500 m level. Re-timing each trajectory to the event's peak intensity would therefore change the regime label of at most a few % of events while introducing a heterogeneous, reanalysis-dependent sampling design.
- L29 and L267 (the SHAP acronym was not introduced). Introduced at first use in both the Abstract and Sect. 2.2.6, with the reference Lundberg and Lee (2017) added to the reference list.
- Table 2 (replace X, Y by lon, lat). Done; see the quoted caption under major comment 2 above.
- L242 (replace "of pcp > 0.1 mm"). Replaced by "of precipitation amount greater than 0.1 mm day^{-1}".
- L306-309 (should be merged into the Methods). Merged; the Results now open directly with the findings.
- L314-315 and L477-478 (the emphasis on the strong correlation among water isotopes seems overrated). We agree: the first-order co-variation of the three isotope ratios is expected, and this is precisely why the second-order parameters are analysed. The emphatic wording was removed and these correlations are now presented as a consistency check, with attention redirected to d-excess and ^{17}O-excess.

Revised text (Abstract, revised wording; ll. 15-42):

> As expected, bulk isotope ratios co-vary strongly ($\rho > 0.95$), so δ^{18}O is used to represent the dominant network-scale signal, while d-excess and ^{17}O-excess are analysed as second-order parameters.

- Fig. 5 (add a column with the number of valid months for each station and season). Done; the counts now appear in the row labels of the revised Fig. 4 (Fig. R3 above).
- L377-379 (repetition of the Methods section). Removed.
- Section 3.2.5 (what is a silhouette value?). A definition has been added at first use, with the reference Rousseeuw (1987).

Revised text (Sect. 3.2.5; ll. 375-403):

> The silhouette coefficient measures how similar each trajectory is to its own cluster relative to the nearest alternative cluster, ranging from −1 to +1, with higher values indicating better-separated clusters (Rousseeuw, 1987). Silhouette values range from 0.227 to 0.393 for wet-season subsets and from 0.325 to 0.410 for dry-season subsets where more than one regime is retained (Table S3).

- Repeated citations (Dansgaard, 1964; Gat, 1996; Pfahl and Sodemann, 2014 cited four times in Sects. 4.2.2 and 4.2.3; Pfahl and Sodemann cited 13 times overall). We audited the citation use, consolidated the repeated triplet and replaced several instances with more specific references appropriate to each statement (Froehlich et al., 2008; Natali et al., 2022; Risi et al., 2008).
- L623 (need not -> may not). Corrected.

Additional corrections identified by the authors during revision

While implementing the revisions we carried out several consistency checks of our own. The following corrections were not raised by either Referee but are reported here for completeness and transparency.

Item	Correction
Random Forest reference	The method was cited as Geurts et al. (2006), which is Extremely Randomized Trees, a different algorithm. Our implementation is scikit-learn's RandomForestRegressor, so the citation was corrected to Breiman (2001) in the text and the reference list.
Wet-season sample size (n = 145)	The multivariate models use 145 rather than 153 wet-season cases. This is a complete-case restriction: the WeMO index is unavailable for December 2021, which removes the eight samples of that month, one per station. This is now stated in Sect. 3.2.6.
Abstract sample counts	The Abstract reported 448 samples and 154 monthly observations, inconsistent with the 205 reported in the Results. Corrected to 406 valid measurements and 205 monthly observations (see major comment 3).
Duplicated phrase in Sect. 4.1	The sentence reporting the WeMOi model performance contained a duplicated fragment ("during cross-validation cross-validated performance"). Corrected.
Misplaced reference entry	The entry for Castro-Díez et al. (2002), added in this revision, had been inserted next to the in-text citation of Chazarra-Bernabé et al. (2018) in Sect. 2.2 instead of the reference list. It is now in its alphabetical position in the reference list.
Figure 5 (formerly Fig. 6)	The panels are now labelled A and B, matching the caption. The caption also states that the map title reports N = 36 while the caption reports n = 40 clustered trajectories, the difference being the four December 2021 days for which the WeMOi, and hence the tercile, is undefined.
Supplement front matter	The Supplement began directly with its first section. It now opens with the article title, the author list and the correspondence address, and its section headings read S1, S2, ... while cross-references from the main text keep the journal's Sect. S1 form.
Stale cross-reference in the Supplement	The caption of Table S2 referred to "Section 4.6", a section number that does not exist in the submitted version. Corrected to Sect. 2.2.5.
Typographical corrections	A missing preposition in Sect. 1 ("monitoring network Sierra de Segura") and a double space in the same section were corrected.

Revised text (Sect. 4.1, corrected sentence; ll. 434-467):

The close agreement between observed and modelled values, both in cross-validation (R^2 = 0.85 ± 0.05; RMSE = 0.41 ± 0.07) and in calibration (R^2 = 0.99; RMSE = 0.11), indicates that the regional

> pressure configuration underlying the original Padua–San Fernando dipole is effectively preserved when Padua is represented by a multi-station northern Italian proxy. [...]

References added to the manuscript in this revision

The following references were added to the reference list of the revised manuscript. They are quoted here in the journal's format as they appear in the revised reference list.

> Baldini, L. M., McDermott, F., Foley, A. M., and Baldini, J. U. L.: Spatial variability in the European winter precipitation δ^{18}O–NAO relationship: Implications for reconstructing NAO-mode climate variability in the Holocene, Geophys. Res. Lett., 35, L04709, https://doi.org/10.1029/2007GL032027, 2008.
>
> Breiman, L.: Random forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

In addition, Lundberg and Lee (2017) and Rousseeuw (1987) were added for the SHAP method and the silhouette coefficient respectively, and Sanchez-Lopez et al. (2016), already in the reference list, is now cited explicitly for the EA–NAO interplay. Geurts et al. (2006) was removed, as Extra-Trees is not used anywhere in the study.

Two further references are cited in this response but do not appear in the manuscript, and are given here for completeness:

> Geurts, P., Ernst, D., and Wehenkel, L.: Extremely randomized trees, Mach. Learn., 63, 3-42, https://doi.org/10.1007/s10994-006-6226-1, 2006.
>
> Nelson, D. B., Basler, D., and Kahmen, A.: Precipitation isotope time series predictions from machine learning applied in Europe, P. Natl. Acad. Sci. USA, 118, e2024107118, https://doi.org/10.1073/pnas.2024107118, 2021.

Data and code availability

The updated repository adds the ERA5-Land temperature series, the partial-correlation, EA-test and network-mean analysis scripts and their outputs, the 3744 trajectories of the HYSPLIT sensitivity experiment together with the driver and re-clustering scripts, and the regenerated figures with their p-value matrices. It will be released as a new version of the existing Zenodo record, so that the concept DOI cited in the manuscript remains valid.