Dear Reviewer #1,

Thank you very much for your thorough and constructive review. Your comments have been helpful to identify aspects that required clarification. We have addressed all your remarks below and in the manuscript. For clarity, we group related comments together when they point to the same underlying issue.

1. Description of the main measure (Rayleigh-wave ellipticity), DOP-E method, and presentation of ellipticity traces

Comments

"I didn`t find the main measure used in the paper well-explained. Is it the sum of the two larger eigenvalues of the cross-spectral matrix, normalized by the trace? Is it, or is it something else?"

"The sense of motion in a Rayleigh wave is also indicative of subsurface structure. Is that included in the measure or not? "

"What is the sensitivity to lateral dimensions? Is it sensitive up to a half wavelength away laterally? "

We have expanded the description of the Rayleigh-wave ellipticity (RWE) measure and the DOP-E workflow to clarify the definition of the observable and the associated processing steps. Specifically, we have rewritten section 3.2 to provide a clear definition of the ellipticity observable, its estimation from the spectral covariance matrix, and the associated processing steps. The revised text explicitly clarifies how ellipticity is defined, its lateral sensitivity, and the scope of the observable used in this study.

We also explicitly comment on the sense of motion (prograde/retrograde) in the Results and Discussion, noting that it is not included in the definition of the observable used here and is not critical for our temporal-monitoring focus.

Comment

How does behave for a nearly monochromatic wave?

For a monochromatic signal, ellipticity would be defined at a single frequency and only if the wavefield is dominated by Rayleigh-wave motion. This situation does not apply to the broadband ambient noise data analyzed in this study.

Comments

"What time windows are used? How is the stacking done?"

"What are min/max values the measure of ellipticity can assume?"

This has been clarified in Section 3.2. Ellipticity is estimated every 2 hours from a single 80-s analysis window at each sampling time; no temporal stacking is applied. Moreover, ellipticity is estimated from multiple polarized Rayleigh-wave signals detected within each 80-s window, and the reported value is taken as their median rather than a fixed min/max value.

Comment

Are you using a synthetic reference ellipticity calculation here, or not?

No, we do not use a synthetic reference ellipticity calculation in this study. As clarified in the manuscript, instead of applying the Dissimilarity Index (DSI) used in Seivane et al. (2024), we directly analyze the raw ellipticity variations as a function of frequency and time, without reference to a synthetic or baseline ellipticity model.

Comment

Why actually should the ellipticity measure be insensitive to variations in noise sources?

In this work, Rayleigh-wave ellipticity is derived using the DOP-E method, which estimates ellipticity directly from the polarization properties of particle motion at the station. In this framework, ellipticity is not a path-averaged quantity (as are far-field group or phase velocities), but a strictly local measure controlled primarily by the shallow elastic structure beneath and around the receiver. The particle motion of a Rayleigh wave describes a vertical ellipse, and ellipticity is defined as the ratio between the semi-minor and semi-major axes (or equivalently, horizontal and vertical components). As a consequence, the absolute amplitude of the wavefield cancels out in the ratio, so changes in source strength do not affect the ellipticity value. Moreover, DOP-E explicitly isolates the polarized Rayleigh-wave component of the ambient field at the station, which strongly reduces sensitivity to the spatial distribution and mechanism of the ambient noise sources. While other ways of estimating ellipticity can be more influenced by source characteristics, in our implementation the measure is therefore largely insensitive to reasonable variations in noise sources, provided that sufficiently polarized Rayleigh-wave energy is present.

Comment

“Neither the raw seismological data, nor the ellipticity traces show up in the paper anywhere. Don`t you think that this is missing? “

To address this point, we have added a new figure (Figure 3 in the new version) presenting representative raw Rayleigh-wave ellipticity observations, making the ellipticity time series and its frequency dependence directly visible to the reader.

2. Auxiliary datasets, tidal components, and interpretation of the 3 cpd signal

Comment

Was the same time window used for barometric data from ERA5, as for the ellipticity.

Yes. The ERA5 barometric pressure time series is sampled at the same temporal resolution as the ellipticity estimates, namely at 2-hour intervals. This is explicitly stated in the manuscript, in section Results and Discussion, where we indicate that “The RWE and driver series were synchronized at a common 2 h interval, segmented into overlapping windows of 60 days, tapered with a Hanning window, and averaged in the frequency domain.”

Comment

Which tidal component is used in Figure 2? Vertical displacement?

As correctly noted by the reviewer, the vertical component of the Earth tides is used for the spectral calculations shown in Figure 2. This information is explicitly stated in the figure caption.

Comment

The statement that the 3 cpd seismic component "represents a reliable diagnostic of pressure-driven processes" (line 448) seems too bold to me.

We acknowledge the reviewer’s concern regarding the wording of this statement. By “reliable diagnostic,” we do not imply that the 3 cpd component constitutes a universal or exclusive indicator of pressure-driven processes across all geological settings. Rather, in the specific context of La Palma island, atmospheric pressure is the only external forcing that exhibits a pronounced 3 cycles-per-day (3 cpd) component, whereas other tidal forcings do not display energy at this frequency.

Consequently, the observation of a clear 3 cpd signature in the seismic response of the shallow subsurface can be robustly attributed to pressure-driven processes rather than to other external forcings. To clarify this interpretation and avoid an overly general reading, we have slightly softened the wording in the manuscript as follows:

"Overall, our findings indicate that, in the La Palma island setting, the 3 cpd component constitutes a reliable diagnostic of pressure-driven processes in the shallow subsurface, with the potential to identify subtle diffusivity changes."

Comment

What line 217 say in comparison with the 2/day component "about 5 times smaller" "should say "..about 10 times smaller", right?

Yes, the reviewer is correct. This was a typographical error, and the statement has been corrected in the revised manuscript.

Comment

What type are the seismic stations? One of them is an accelerometer, ok, but what about the others?

This information has now been incorporated into the revised manuscript. The manuscript explicitly states: *"Both EHIG and TBT are equipped with broadband velocity sensors, whereas EXILP is an accelerometer. Its recordings were therefore converted to velocity using the instrument response to correctly estimate particle motion through covariance analysis (Schimmel & Gallart, 2004))."*

Comment

Unfortunately, there has been only a short time window of outgassing measurements at one station, and that does not really agree well with the gain measurements. Line 309 states a "moderate agreement" between the lines in Figure 4a and A2. It is obvious what making such a statement would require: averaging the gain data to annual values, and comparing that quantitatively with the CO2 release. What is the statistical significance of the comparison (quantitatively)?

We agree with the reviewer that a quantitative comparison is required to support the statement of a moderate agreement between the seismic gain variations and the CO_2 outgassing measurements. To address this point, we have added a new panel (Figure A2-b), in which the low-pass filtered 3 cpd gain at EHIG is averaged over time intervals defined by the available CO_2 release estimates and displayed together with their associated variability.

In addition, we computed Pearson correlation coefficients between the CO_2 release estimates and the corresponding period-averaged gain values shown in Figure A2-b (black, blue, and green curves). The resulting correlation coefficients are 0.54, 0.55, and 0.47, respectively, indicating a moderate positive correlation. These results quantitatively support the interpretation of a moderate agreement between the seismic gain variations and the CO_2 release estimates, consistent with the wording used in the manuscript.

3. Other comments and minor points

Comment

Line 313/314: that is pure speculation

We acknowledge the reviewer's concern regarding the speculative nature of this interpretation. To explicitly address this limitation, we have added the following sentence in line 315: *"Unfortunately, no direct records of gas emissions are available for this period to validate this interpretation."*

We nevertheless consider it appropriate to briefly comment on this time interval, as it corresponds to a pronounced increase in coherence within the 16-year time series. While the interpretation remains tentative, it is now clearly identified as such in the manuscript, and no definitive conclusions are drawn in the absence of independent gas emission measurements.

Comment

Line 313/314: The paper is extremely dense and not easy to read

We appreciate the reviewer's observation. We are aware that the previous version was dense, and similar concerns were also raised by Reviewer 2 regarding structure and readability. In response, we have substantially reorganized the manuscript to improve balance and clarity. In particular, the Study Area section is now divided into four focused subsections, and the Materials and Methods section has been restructured to clarify the logical progression from subdaily atmospheric forcing to long-term variability. We believe these changes significantly improve the overall readability of the paper.

Comment

The seasonal variation is real, but it is not clear which process causes that

We agree with the reviewer that the seasonal variation is clearly observed and that identifying its driving mechanism requires careful consideration. As discussed in the manuscript, we systematically compare the observed semi-annual modulation with other environmental variables available for La Palma, including volumetric soil water content, wind, and atmospheric pressure. While some of these variables may exhibit semi-annual variability, none of them display a pattern that is as regular, phase-consistent, and well defined as that of the solid Earth tides.

The close temporal correspondence between the semi-annual cycles observed at the three seismic stations and the solid Earth tides therefore provides the strongest explanation supported by the available data. On this basis, we argue in the manuscript that the detected semi-annual modulation is most consistently explained by solid tidal forcing, rather than by other environmental processes.

Comment

The anomaly is intriguing, but the other two stations don`t corroborate it.

We agree with the reviewer that the anomaly is not systematically observed at all stations. However, such a uniform response is not expected, given that diffuse degassing patterns are spatially heterogeneous and strongly site dependent as illustrated by the degassing maps of Padrón et al. (2015). This point is already addressed in the manuscript.

As discussed in the text and highlighted in Figure 4, the anomaly at EHIG station is characterized by only a slight increase in coherence during the eruption episode and in early 2021, together with a disruption of the otherwise regular annual phase seasonality. Importantly, as explicitly stated in the manuscript, these phase-lag perturbations are not accompanied by a corresponding increase in gain and are therefore not interpreted as evidence of enhanced gas accumulation. Instead, they are attributed to changes in the hydrological state of the vadose zone, most likely related to variations in soil moisture or soil permeability.

Comment

Line 460: "a slight increase is observed" seems to optimistic to me: to me that peak look just like noise.

We acknowledge the reviewer's concern regarding the interpretation of this feature. As already emphasized in the manuscript by the use of the term *"only"*, the increase in coherence at EHIG is weak and transient. To further clarify our interpretation, we have added a sentence stating that, given the limited temporal extent of this coherence change and the absence of a corresponding increase in gain before and during the eruption, this feature is not interpreted as evidence of intense gas degassing at this site.

Comment

The Appendix shows DSI. That requires a few words saying what it is (one minus the crocor at 0-lag).

We agree with the reviewer and have addressed this point by including a brief description of the Dissimilarity Index (DSI) in the caption of Figure A1, where it is defined as one minus the normalized zero-lag cross-correlation.

Minor Points

All minor points raised by the reviewer have been addressed in the revised manuscript.

Two exceptions are clarified below.

The first refers to the suggestion of replacing the term *HVSNR* by *HVSR*. We have not implemented this change, as the method is explicitly defined in the main text as the horizontal-to-vertical spectral noise ratio (HVSNR) method, following the terminology used throughout the manuscript. To ensure consistency and avoid ambiguity, we therefore retained the original notation.

The second exception concerns the comment related to Figure 5. While using the same frequency band for all stations would be preferable for consistency, this is not possible for station EXILP because a clear 3 cpd spectral peak is not observed in the 40–50 Hz band at this site. As shown in the spectrogram presented in the main manuscript (Figure 5a), the 3 cpd component at EXILP progressively weakens and eventually vanishes toward lower frequencies, becoming indistinguishable below approximately 70 Hz.

For completeness, we additionally computed the spectrogram restricted to the 40–50 Hz band (shown below). This representation confirms the absence of a well-defined 3 cpd spectral peak in this frequency range, demonstrating that applying the same band used for the other stations would not be physically meaningful for EXILP. The frequency band selected for this station in Figure 5 was therefore chosen to ensure that the analysis focuses on frequencies where the 3 cpd signal is actually present and can be robustly characterized.

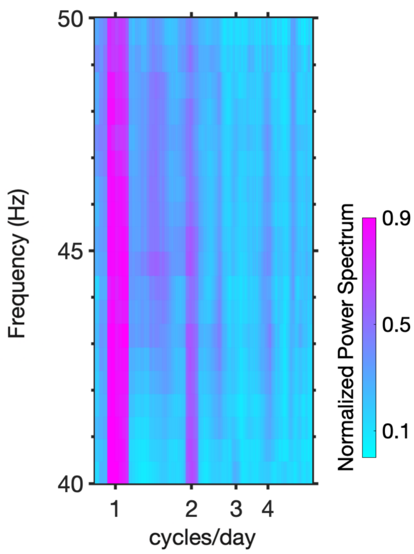

Figure R1. Normalized frequency–period Fourier spectrogram of the RWE series during 2021 (frequency range 40–50 Hz) at the EXILP station. Sliding windows of 30 days with 50% overlap were applied. The absence of a clear 3 cpd spectral peak in this frequency range explains why the same band used for other stations cannot be applied at EXILP.

We hope we've covered your comments and are willing to respond to any further questions and suggestions you may have,

Sincerely,

Helena Seivane and Martin Schimmel

Dear Reviewer #2,

We sincerely appreciate your careful and insightful review. Your suggestions have greatly contributed to clarifying the manuscript and strengthening its organization and presentation. Several of the points you raise were also independently noted by Reviewer 1, and we have revised the manuscript accordingly to address these issues.

Below, we respond to your comments in detail.

Comments on the structure, focus, and clarity of the manuscript

Comment

"I think the current structure of the paper is unbalanced and not ideal. Some sections are extremely long, while other, more important sections are rather short and confusing. Even after reading the paper several times, I found the main message somewhat difficult to follow. I think the authors should consider restructuring the paper and its focus."

In the revised version, we have substantially reorganized the paper to improve balance and readability. The Study Area section has been restructured into four focused subsections, and the Materials and Methods section has been rewritten to clarify the progression from subdaily atmospheric forcing to long-term variability. These changes are intended to better highlight the main message of the paper and to guide the reader through the argument in a clearer and more coherent way.

Comment

"Some aspects mentioned in the paper are not fully relevant to the study and its interpretation."

Although it was not fully clear to us which specific aspects were considered not directly relevant, we carefully revised the manuscript with this concern in mind. During the reorganization of Sections 2 and 3 we streamlined the text and made explicit why the remaining information is relevant to the objectives of the study. We hope that this revised version makes the focus of the paper clearer.

Comment

"If I understand correctly, the authors want to measure changes in degassing caused by atmospheric tides using RWE. While I found this idea very interesting, I became confused when the time scales switched back and forth between short-term Fourier analysis and long-term interpretation. I think the idea and the goal should be outlined more clearly."

We assume that degassing is governed by the internal dynamics of the volcanic system and not by atmospheric tides. In our study, atmospheric tides are used only as an external, repeatable forcing to probe changes in the vadose zone. Our focus is on how the presence and progressive accumulation of gas (previous to the final degassing to the atmosphere) modify the pressure transmission and seismic velocity response in the upper vadose zone, as recorded by RWE.

To clarify the treatment of time scales, in the revised manuscript we now explicitly separate:

1- the short-term subdaily response to atmospheric tides, and
2- the long-term evolution of this response under a gas-accumulation scenario.

These aspects are discussed in the new subsections 3.1.1 ("Short-term response to subdaily atmospheric tides") and 3.1.2 ("Long-term evolution under a gas-injection scenario"). This restructuring is intended to clarify the conceptual framework and the objectives of the study, and to make the link between short-term Fourier analysis and long-term interpretation more transparent.

Comments on the RWE methodology, tidal-band selection, and physical interpretation

Comment

"The authors use the RWE measure, which seems to be relatively new to the community (there are not many references to its application in monitoring). Since this is the key metric, I think the authors should explain it in more detail. They should explain how the processing is done and how the RWE is determined. They should also consider showing the original RWE series, as well as the processed ones"

As RWE is still relatively new as a monitoring observable, we agree that a more detailed description of how it is computed and processed is important. In the revised manuscript we have therefore substantially expanded Section 3.2. We now describe in detail the DOP-E workflow, including the window length, polarization analysis, mode-selection criteria, and the construction of RWE time series at each frequency.

We have also followed the reviewer's recommendation to show the original RWE series used in the analysis (Figure 3 in the revised version of the manuscript). We believe these additions clarify how RWE is determined and address this concern.

Comment

"Where is the high-frequency noise coming from? What are the sources of the noise? These questions should be discussed."

High-frequency noise in the analysed band is most likely dominated by anthropogenic sources (e.g., traffic and other cultural noise), as is typical at these frequencies. In the revised manuscript we now state this explicitly and emphasize that a key advantage of the DOP-E–based RWE measurement is its reduced sensitivity to noise source variability, which is particularly important in such high-frequency bands. For this reason, we briefly discuss the likely origin of the noise but do not attempt a detailed source characterization.

Comment

"Other things, such as a description of atmospheric tides and their origin, are missing. "

We thank the reviewer for this remark. A concise description of atmospheric tides, their main components, and key references was already included in the previous version and remains in the revised manuscript. We consider this level of detail appropriate, since our aim is not to review atmospheric-tide theory but to use these tides as an external forcing to probe subsurface changes. For clarity, we have slightly refined the wording in Section 3 to make this role more explicit.

Comment

"For me, it is unclear why the authors present Rayleigh wave ellipticity as a measure when V_p is more affected."

Rayleigh-wave ellipticity depends jointly on V_P, V_S, density, and near-surface layering. It is well established that, in many settings, ellipticity exhibits a higher sensitivity to V_S than to V_P. We fully agree with the reviewer on this general point.

In our case, however, the physical forcing that we investigate is barometric pumping in the vadose zone. This mechanism induces pressure oscillations that predominantly modify the bulk modulus of the vadose zone, and therefore mainly affect V_P, while producing comparatively minor changes in V_S. Our working hypothesis is thus that V_S remains approximately constant over the hourly timescales considered, whereas V_P varies in response to pressure-driven changes in pore-fluid conditions.

Under this hypothesis, the observed ellipticity variations can be attributed primarily to V_P changes, even though the observable itself is, in principle, sensitive to both V_P and V_S. We have clarified this reasoning in the revised manuscript, emphasizing that ellipticity does not become more sensitive to V_P, but rather that the dominant physical process in this study preferentially perturbs V_P**.**

Comment

"Why is the 8-hour tide used? For example, a 4-hour tide has a similar magnitude according to the spectrum. In general, the testing signal is quite weak. I'm wondering why the authors are not using 2cpd instead. "

The choice of the 8-hour atmospheric tide is motivated by its spectral isolation as a barometric forcing in La Palma, rather than by its absolute amplitude. As we now explain in Section 3.1.3, the 3 cpd component is clearly expressed in atmospheric pressure, while the corresponding 8-hour oceanic, solid Earth, and temperature tides are nearly absent, making this band a clean diagnostic of barometric forcing. In contrast, the 2 cpd band overlaps with strong semidiurnal solid Earth and ocean tides (S2, M2), so changes in amplitude or phase would be harder to interpret uniquely in terms of barometric pumping. For this reason, we prefer a weaker but well-isolated 8-

hour tide over a stronger but more ambiguous 2 cpd signal. This rationale, and the limitations of using the 2 cpd band in our setting, are now explicitly discussed in Section 3.1.3 and in the Discussion.

Comment

"While I am convinced by the author's claim that RWE is affected by the atmospheric tide, I am not so sure about the degassing part. I did not understand whether RWE is small or large, e.g. before an event, or what influence a greater release of gas has on RWE."

The intention of the manuscript is not to propose that degassing directly modifies RWE. Rather, we assume that gas released from depth first accumulates in the vadose zone before reaching the surface. This accumulation changes the physical state of the shallow subsurface (pore pressure, gas content, compressibility and hydraulic diffusivity), and therefore modifies how the medium responds to the externally imposed atmospheric-tide forcing. What we observe is not the degassing signal itself, but the evolution of the response of the vadose zone to the external forcing.

For this reason, our interpretation does not rely on whether RWE becomes "large" or "small" before an event. Instead, we examine how the amplitude, persistence, and phase of the 3 cpd, barometrically forced component evolve through time. In the revised manuscript we have clarified this conceptual framework and emphasize that RWE is used as an indicator of pressure-driven changes in the shallow subsurface associated with gas accumulation, rather than of instantaneous degassing rates.

Comment

"The anomalies identified by the authors appear to be of a similar magnitude to noise or other events. How did the authors classify the anomalies? "

In this manuscript the term "anomaly" is used in a contextual sense, not as the result of an automatic detection scheme. We refer to deviations from the long-term seasonal pattern of the 3 cpd response that are persistent and repeat across different attributes (e.g., coherence and phase lag). The features we highlight differ from the preceding decade of observations and therefore stand out from short-term noise fluctuations, even if their absolute amplitude is comparable.

Minor Points

"The authors claim that RWE is similar to dv/v, yet no comparison is made, for example on line 266."

RWE changes when the seismic velocities change, similar to dv/v. In both cases there exist ambiguities, i.e., it is not clear if changes in V_P or V_S or both cause the observed measurements. In that sense RWE has some similarities to dv/v.Thus, in the revised manuscript we clarify that, in our framework, RWE plays a role analogous to dv/v as a monitoring observable, in the sense that both are sensitive to temporal changes in seismic properties. However, RWE is not used as a direct quantitative proxy for dv/v. RWE has no units and an inversion is required to translate corresponding changes into dv/v with some ambiguities. Here, we only use

RWE as a practical tool to detect changes in the subsurface. A detailed joint comparison between RWE and dv/v is therefore beyond the scope of this study.

“I found the use of terms such as 'RWE' and 'RWE gain' confusing. What is the difference between them? Why are they not consistent?”

RWE refers to the ellipticity time series itself, whereas RWE gain denotes the magnitude of the transfer function between atmospheric pressure and RWE at 3 cpd, i.e., the amplitude of the RWE response relative to the barometric forcing. The two quantities are therefore related but not identical: RWE is the observable, while RWE gain characterizes how strongly it responds to the atmospheric tide by analyzing the evolution of a spectral component of the RWE measurement. Using the gain has the additional advantage of normalizing by the amplitude of the driver, so that an increase in RWE amplitude can be interpreted as a change in the medium response rather than simply as a result of stronger atmospheric tides. The definition of the gain is given in Equation (8).

“Why is the tidal spectrum suddenly being calculated when it's about pressure changes? “

The tidal spectrum is not introduced to replace atmospheric pressure, but to distinguish between the different forcings that can generate subsurface pressure changes. In coastal volcanic settings, periodic pressure variations may originate from atmospheric tides, ocean tides transmitted through hydraulically connected aquifers, and solid Earth tides. Since we do not have direct piezometric records, we compute the tidal spectra to assess the possible contribution of ocean and solid Earth tides and to justify that the 3 cpd signal we analyze can be attributed to barometric forcing rather than to tide-induced aquifer loading. This rationale is already introduced and motivated in the Introduction.

We hope that our revisions satisfactorily address your comments and we would be pleased to clarify any remaining questions

Yours Sincerely,

Helena Seivane and Martin Schimmel