Dear Paula Koelemeijer,

We thank you and both reviewers for reading and commenting on our software report and considering this manuscript for publication in Seismica.

We have considered all the comments in the following rebuttal letter with detailed responses to each point and modified the text accordingly.
The manuscript has been improved, especially the introduction. We believe that the description and scope of the WMSAN package are now clearer.

Our responses are colored in blue in the following. We put the reviewers' questions above for clarity.

Thank you again for your time handling this submission.

The authors

REVIEW A

Dear Authors,

I have reviewed your manuscript entitled "WMSAN Python Package: From Oceanic Forcing to Synthetic Cross-correlations of Microseismic Noise" and found the subject matter to be of interest. The presented work on the WMSAN Python package has the potential to be a valuable contribution to the ambient seismic noise community.

The paper presents a valuable contribution to the field of ambient seismic noise by modeling microseismic sources of ambient noise with the recent improvements in ocean wave models and computing synthetics of seismic spectrograms and cross-correlations for surface waves (Rayleigh) and body waves (P, SV).

However, the current manuscript requires some improvement in clarity and structure to effectively communicate the significance and impact of your work.

I believe that with some revisions addressing the points below, this manuscript has the potential to be a valuable contribution to the scientific literature. I encourage you to carefully consider these suggestions and revise the manuscript accordingly.

Thank you for taking the time to review our manuscript. We are pleased to hear that you found the article of interest for the ambient noise community. We will consider your remarks in the revised version of the manuscript and detail the modified points hereafter.

Line 48 - 53 : I expected that starting from line 51, there will be more introduction about seasonal pattern which has implications for cross correlation, but line 51 explains something more general about noise sources. I think it would be better if the presentation is modified.

The first three paragraphs in the introduction have been reordered to start with the generation mechanism of seismic waves and finish with cross-correlations and source distribution issues. Here is the new structure:

- ocean wave generation mechanism (primary, secondary, and hum)

- the use of such sources (but not only secondary microseisms) in ambient noise studies.

- seasonality of secondary microseismic sources and the need for a precise source description.

Line 54 : It is a bit strange to start the explanation of each noise source starting from a less understood phenomenon because the reader is not informed beforehand about the more understood phenomenon. I agree that hum and primary microseisms should have a smaller proportion in Introduction compared to secondary microseisms which are the focus of this study.

Thank you for this remark. We changed the order in which the different mechanisms are presented. As reviewer B also mentioned, this paragraph did not link hum and primary microseisms clearly, so the new order seems more natural.

Line 98 : It would be better if the reader is introduced to WW3 outputs in general, what are amplification coefficients and their influence in modeling in general? something like in line 157 that helps the reader understand WW3 outputs.

The beginning of this section has also been reordered to introduce WW3 in general before going into details on the modeling of secondary microseismic sources. Formerly line 157 has been added to the section 2 introductory summary and removed from section 3.

Line 132: It would be easier to understand the context if the applications of surface waves were mentioned, e.g. imaging and monitoring.

The sentence was modified accordingly: “Secondary microseismic ambient noise records are dominated by surface waves which are therefore widely used for imaging local to regional areas or monitoring (e.g., Sabra et al., 2005; Bensen et al., 2008; Lu et al., 2018)”. Lines 142-143.

Line 162: Ifremer should be written in block letters to maintain consistency.

True, it is in block letters everywhere now. Line 74, 108, 173.

Line 165: The authors mention several software packages that are commonly used in handling recorded seismic data. MSNoise and NoisePy may have some common purposes, but ObsPy does not seem to be specifically designed to explore ambient noise studies. Can the author explain more about how WMSAN stands as a software with a specific purpose that can overlap or integrate with other software in ambient seismic noise studies?
Does one need to use WMSAN before processing large volumes of cross-correlation with, e.g. MSNoise or NoisePy?

This paragraph indeed lacked clarity. We intend to highlight the fact that WMSAN provides only the modeling of seismic observables. If the user wishes to compare the outputs of WMSAN, other packages could be necessary to compute the corresponding observed data. We give some examples of Python packages (MSNoise, NoisePy, and broader Obspy) for the reader who is new to the field. Here is the modified paragraph:

“WMSAN only provides tools to model seismic data proxies, such as synthetic spectrograms, synthetic cross-correlation, or maps of the distribution of seismic sources. We do not provide tools to handle recorded seismic data, since other Python packages can be used for data processing such as ObsPy (e.g., Krischer et al., 2015) or, more specifically for ambient noise

studies, to calculate cross-correlations such as MSNoise or NoisePy (e.g., Lecocq et al., 2014; Jiang and Denolle, 2020).” Lines 177-181.

Line 168: The European Center for Medium-range Weather Forecasting requires an abbreviation.

Indeed, the acronym was added in parentheses after for more clarity.

Line 266: Between numeric values and units of measurement do not use a space, e.g. Line 264 ? Should be consistent and follow the convention.

Indeed, the manuscript was not consistent on this aspect.
We looked for the exact rule about space between a number and a unit and found a document by the *Bureau International des Poids et Mesures* that states:

“The language of science: using the SI to express the values of quantities

The value of a quantity is written as the product of a number and a unit. The number multiplying the unit is the numerical value of the quantity in that unit. A single space is always left between the number and the unit.”

Here is a link to the document:
https://web.archive.org/web/20210505230221/https://www.bipm.org/documents/20126/41483022/SI-Brochure-9-concise-EN.pdf/2fda4656-e236-0fcb-3867-36ca74eea4e3

The manuscript has been modified consistently, adding a space between a number and its unit.

Figure 4: The font size of the labels may be adjusted to make the display more consistent. There should be some space between the top and bottom figures for the figure subtitles.

The font and label size should be consistent now. We added some space between the top and bottom rows.

General questions:
In the Introduction (Line 83) and Conclusion (Line 293) it is stated that seismologists should not consider significant wave height as seismic noise amplitude and rather promote the use of WW3 spectral density of the pressure field at the sea surface. I think this is a strong point for this study.
Is it possible to explain or demonstrate the difference between these two assumptions and the implications for ambient noise studies?

Indeed, the significant wave height (Hs*)* is sometimes compared to noise sources inversion results. However, the secondary microseism mechanism depends on the oceanic wave direction (opposite wave trains) and frequency content (similar period of oscillations).

In Ardhuin et al. (2011), the difference between significant wave height and seismic sources of ambient noise is detailed. Here is in substance the main argument to use the spectral density of the pressure field at the sea surface.

The significant wave height is defined as $H_s = 4\sqrt{\{\int_0^\infty E(f)df\}}$ with $E(f)$ the power spectrum of the sea surface vertical displacement. Therefore, there is no information on the direction of the oceanic waves in the significant wave height.

On the contrary, the spectral density of the wave-induced pressure just below the sea surface F_p, in $Pa^2.m^2.s$ is computed as:

$$F_{p(r,2f)} = [2\pi]^2 [\rho_w g]^2 \, 2f \, E^2(r,f) \int_0^{\pi} M(r,f,\theta) M(r,f,\theta+\pi) d\theta$$

With $\int_0^{\pi} M(r,f,\theta) M(r,f,\theta+\pi) d\theta$ carrying the directional information of the sea state.

So, the significant wave height cannot be directly used as a secondary microseismic source proxy. We added the definition of the significant wave height in section 2.1 to highlight the difference between both equations. Line 124-127.

REVIEW B

Review for "WMSAN Python Package: From Oceanic Forcing to Synthetic Cross-correlations of Microseismic Noise", submitted to Seismica by Tomasetto et al.

This report describes a new software package (Python) which facilitates modeling realistic ambient seismic noise spectrograms and cross-correlations in the secondary microseism band. To do this, the tool enables users to download wave model hindcast data, in particular the equivalent pressure term giving rise to secondary microseism sources, to compute the force terms needed to synthesize ambient seismic noise spectrograms and cross-correlations, and to carry out the computation of cross-correlations and spectrograms themselves. Intermediate outputs and final outputs can be visualized thanks to the plotting capacities of the package. I am extremely happy about the availability of this tool, and I believe that it will be used by multiple people in the ambient noise community. It allows users to investigate source effects on ambient noise cross-correlations, which have been intensely debated. While the ocean wave hindcast data has been available since years, the physics of microseism generation is quite involved and the modeling of ambient noise cross-correlations technically cumbersome. Therefore, only few researchers deeply involved in the topic were previously using the hindcast data for ambient noise modeling. This tool opens their use for a wider part of the community. The manuscript is in general well written and I have no concerns from a methodological point of view. However, as the topic is complex, the report is not very straightforward to follow. I provide some questions and comments for clarification below. Most of them are minor.

Thank you for your careful reading and understanding of the manuscript and relevant remarks. We hereafter detail our answers point by point to each and hope the final draft will be clearer for future users.

Most important points:

1. The main products of the code from the point of view of ambient noise sources are, if I understand well, the source force amplitude proxy F_{prox} in Newton and the power spectrum of vertical displacement at the ocean bottom SDF. These are computed from the inputs of F_p (the pressure spectral density at the sea surface) and the bottom topography, depending on wave type.

Exactly.

I am quite confused about several things:

- The force proxy in equation 4 is integrated over frequency, so how is it related to the computation of spectrograms/cross-correlations? I guess the term S from line 228 is used for the cross-correlations?

  The force proxy, F_{prox}, as defined in Equation 4, is used to compute maps of the secondary microseism sources distribution. As you noticed, the term S line 228 is used for the cross-correlations input. The spectrograms are computed using the power spectrum of the vertical ground displacement F_{δ} line 333 in the Appendix. We believe Table 1, added after the Introduction, helps link the package outputs with the terms described in the article.

- What is the term that is shown in Figure 2, top panels (equivalent force)? Basically, for readers not very familiar with the physics, it is not easy to keep the overview of the various terms and their meaning (F_P, F_{prox}, S_{DF}, S). To make it easier, I would suggest the following:

  - including a bullet-point list in the introduction or the beginning of the theory section of all relevant terms with reference to the section where they are described.
  - Including the symbol (F_P, F_{prox}, S_{DF}, S), the equation number, the section where the term is described and the Figure where the term is shown in Figure 1. It could be very short (e.g.: "F_P; eq. 1; Fig x; Sect. y") and I think it would really help the reader a lot to link all the pieces and understand the capacities of the tool better.

  Thanks a lot for this suggestion; we believe it makes the available outputs and links between equations and figures much clearer. We added the table between the introduction and theory sections. We also included the format in which the term is stored, following your point 3.

2. Eq. 2, what is the index i (modes?) and why does it run from 1 to 4?

The i index is the mode for the Rayleigh surface wave amplitude response function as defined in Longuet-Higgins (1950). In this article, he computed the coefficient for the first (a priori dominant) 4 modes, i=1 corresponds to the fundamental mode. This was specififed lines 151-152.

3. This is a technical paper, so I suggest to briefly describe the file formats for all outputs with references for the formats (many formats have an associated publication, e.g. hdf5). It is implicit from Figure 1, a short sentence with the reference would be sufficient to make it more complete.

The format were added in the table mentioned in 1., also a short sentence is added to reference the two formats used (hdf5 and netCDF).

"The outputs are stored in NetCDF format (UCAR/Unidata) by default, or in HDF5 format for the starting model for noisi (The HDF Group)." Line

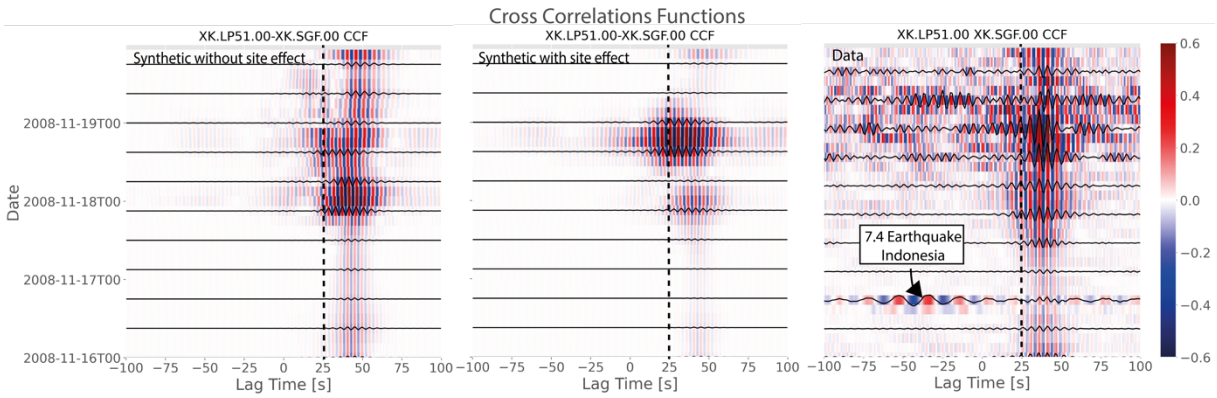
4. Line 230, "...but a first estimate of the variability of the cross-correlation" – does the variability here refer to the variability of the amplitude, or the waveform, or something else?
   Indeed the word variability can have several meanings here. We modified the sentence as follows:

"Therefore, we do not expect to retrieve the absolute amplitude of the real data cross-correlation, but rather a first estimate of the relative amplitude variability as a function of source distribution and frequency content." Line 248-250.

5. Section 3.3: How is the geographic area of the sources chosen? Would it be possible to show the comparison without the source site effect in the appendix?

   This example was chosen to be able to use the LAPNET network which was deployed and recorded from late 2007 to late 2009. The homogeneous nature of this craton was interesting to compare to modeling in a 1D model (ak135f here). We then selected a secondary microseismic event in the Northern Atlantic from Nishida and Takagi (2022) catalog. The centroid location was around 14°W 69°N, as Rayleigh waves are dominant at a distance below 2000 km from the station location we selected the area approximately within a 20° radius from the XK station locations. The extent of the area was also scaled so that the example takes less than an hour to run.

   Here a comparison of the synthetic cross-correlations with and without site effect:



The site effect seems to have an impact on the relative amplitude on November 19, 2008. Also, the arrival at 45s seems well retrieved without site effect, but the maximum amplitude is on November 18, instead of November 19 in the data.

The site effect takes into account the bathymetry, so we decided to present only the cross-correlation with site effect as it is by default in the package. Removing the site effect would need to put the amplification coefficient to one, which is not directly implemented. We added the comparison between each panel in the Appendix and referenced the Figure in the main text. Line 296.

6. Appendix table 1: Ideally, add the Figure number where the outputs of the respective notebooks appear in the manuscript.

   That's a fair point; it was added to the Table (now Table 2).

7. Appendix: It is a little bit confusing that SDF is not described in the main text, although it is one of the six outputs of the tool. What is the motivation for putting it in the appendix?

   We chose to put the synthetic spectrograms in the Appendix as it was a rephrasing of Ardhuin et al. (2011) and Stutzmann et al. (2012). Also, similar codes computing synthetic spectrograms taking into account bathymetry site effect and additional

propagation properties will soon be available. We therefore decided to emphasize the correlation modeling and source distribution maps.

Minor points:

Is there a period (.) after the title on purpose?

No, it got removed.

There are several sentences in the manuscript where symbols are used that were not yet introduced, e.g. in the first sentence of the abstract, period T. I suggest to introduce all symbols (even though many may seem obvious) close to where they are first used.

Line 39: compute a proxy of seismic observables: Is this not too technical for the non-technical summary?

Indeed, I rewrote the non-technical summary to make it more accessible:

"Continuous shakes of the ground recorded everywhere on Earth, called seismic ambient noise, show significant peaks in energy around 7s and 14s. These correspond to seismic waves originating from interactions between oceanic waves with themselves or with the sea floor at the coast respectively. Seismic ambient noise studies focus on retrieving information on the Earth's structure at different scales and depths. Knowing the dynamic of the source of seismic waves is crucial to extracting the physical characteristics of the sampled medium. Recent developments in oceanic wave modeling, through satellite and buoy data integration, have opened new opportunities for seismologists to understand recorded seismic traces. In this study, we introduce the Wave Model Sources of Ambient Noise (WMSAN, pronounced wam-san]) Python package to visualize the maps of ambient noise sources and compute the simulated seismic waveforms in an efficient, user-friendly fashion." Lines 24-34

Line 50, Fichtner 2014: It is a purely synthetic study which does not comment on seasonality, I suggest replacing it with one that looked at seasonal variations and/or real data, like Igel et al.

The reference to Igel et al. (2023) now replaced the reference to Fichtner (2014). Thank you. Line 63.

Line 55 ff: ("Nevertheless, ...") – if I understand well, Fabrice Ardhuin considers that the mechanism of the hum excitation is essentially the primary microseism mechanism, just with a different location in relation to the coast and different shape of bathymetry (deep slope vs. shallow slope/undulated topography). The way this paragraph is formulated does not make this clear, readers who are not familiar with the topic may believe that it is a different mechanism. I suggest to make it more clear that the mechanism is essentially the same as for the primary microseism. (although I may not be fully up to date on the latest state of knowledge)

This paragraph was reordered (see Review A second point). I made explicit the similarity between the hum and primary microseisms mechanism. Line 51.

Line 67: Matched-field processing: Igel et al. 2023 use MFP for the starting model but then proceed with noise source inversion, so it might be added as an additional method (matched field processing and noise source inversion). There is a study by Josefine Umlauft that uses purely matched field processing, but on small scale (https://academic.oup.com/gji/article-abstract/219/3/1550/5556535).

Thank you for the precision. We added “or match field processing followed by noise source inversion” to the text, line 67. As the study by Josefine Umlauft is at the local scale and in a different frequency band we thought it might confuse the reader at this stage in the introduction, so we didn’t add it. Thank you for bringing this reference to our knowledge.

Eq. 3: I would explicitly state that here, P/SV stands for either P or SV, to avoid any confusion.

The P/SV was replaced by $P \mid SV$ and explained the P or SV in the equation description.

Line 199: explain also that the equivalent force is shown too. What does it correspond to? Is it the frequency-integrated version of Eq. 1?

The equivalent force corresponds to: $2\pi\sqrt{\{\int F_p(f_s)df_s dA}$. So it is similar to $\mathrm{F_{prox}}$ without the amplification coefficient applied. I added a sentence to say that the equivalent force is shown on the top of the Figure.

“The top row presents the equivalent force which is defined similarly to without amplification coefficients applied: $2\pi\sqrt{\int F_p(f_s)df_s dA}$.” Line 215-218.

Line 202: “...seasonality of such sources is retrieved” but the example shows only January, I would stress that the distribution is typical for this season

The sentence was modified into: “Also, the distribution of noise sources is typical for this season, with stronger sources in the Northern Hemisphere from October to March, and stronger sources in the Southern Hemisphere from April to September.” Line 64-65.

Line 207: citation Ermert et al (2020), there are no inversion results in that study; inversion results would be available in Ermert et al. (2021) or (2017), the citation could also be dropped as these all refer to Earth hum.

The citation was removed, thank you for the precision.

Equation 5: I suggest to use a different symbol for the cross-correlation, e.g. \mathcal{C} because C is already used for the source site effect of Rayleigh waves.

Indeed it was confusing, thank you for the suggestion.

Line 234: some observations of repeating arrivals from microseisms are becoming available, for example: https://seismica.library.mcgill.ca/article/view/499 – but it is true that this has not been commonly observed.

I added the reference to the text to show that it has been observed at least once. This hypothesis might not be valid in some cases. Line 253.

Figure 4: Dates have to be shown on the y axis of panels c). If not, it becomes hard to follow the description. Maybe they were cut off by the layout of the journal.
Showing a few traces as lines underneath the 2-D plots could help to see the frequency difference better.

It is an error from my side, I included the dates and some of the waveforms as you suggested in the Figure 4 and Figure B.

Extra-knitpicky points: (these are only about style and language)

several places in the manuscript use ‘s to form a possession, like waves’ source, sources’ dynamic. I think this is quite unusual for inanimate nouns and most people would use either “... of the ...” or just “wave source, source dynamic” etc. (lines 33, 48, 117, 172, 190, 228, 268, 272)

modified

abstract lines 16 – 17: “..mechanisms....called primary and secondary microseismic peaks.” To be extra clear here, I suggest to change this to “...called primary and secondary microseism”.

changed

non-technical summary line 30: around 7s and 14 s *of period*?

added

Line 41: Ocean waves and extreme climate conditions – extreme weather? Assuming that these are relatively short-lived conditions

changed

Line 49: some early studies – why early? The latest one is from 2024, so that is very recent

True, removed early.

Line 55: strictly speaking, Nishida (2014) considers also resonance between atmosphere and solid Earth as a possible mechanism for very low-frequency hum

Thank you for the suggestion, I added this hypothesis in the list.Line 50.

Line 104 among them – which interact among themselves changed
Line 108 ...used for seismology applications by Ardhuin et al. (2011) / ...used for seismology applications (Ardhuin et al., 2011) changed

Line 155: visualize and compare ambient noise sources – this is a bit confusing because what is compared to observed data are models based on the WW3 noise sources, but not the source maps directly (models of PSD, correlations etc)
I modified the sentence to be more precise: "We introduce default examples to compare observed data to synthetic counterparts based on WW3 ambient noise source distribution.” Line 168-169.

Line 164 This package → WMSAN changed
Line 165 We don’t → We do not (more formal) changed
Line 177 thinner → finer changed
Line 178 (NOAA .., 2022) or: by NOAA (..., 2022) changed
Line 253 XK.LP51.00 and XK.SGF.00 which path → whose inter-station path? changed

Figure 3: What does z* stand for? Taking the complex conjugate? It may be clearer if this is mentioned explicitly

Indeed it refers to the complex conjugate, it was added explicitely on Figure 3.

Line 296 “if this tool doesn’t bring any significantly new...” → this sounds too modest, since it is a software report, it is maybe not expected to describe a new method. Why make this contribution seem smaller than what it actually is? It does answer a need in the community and quite a pressing one.

Thank you for this remark, we reformulated the sentence.Line 318.

“This tool answers a need in the community to apprehend source distribution of secondary microseisms.”