Reviews and responses for Nye et al. *Validation framework for semi-stochastic simulations in Cascadia earthquake early warning*

Review 1 (anonymous)

Testing EEW systems cannot be done by regionally-appropriate datasets since the range of target events are no-where near sufficient in the existing recorded archives, since large events occur rarely and and EEW-ready seismic networks are typically only very recently in operation, for years or at most decades. This paper demonstrates that EEW systems can be tested using synthetic waveforms, with application to the Cascadia region. They demonstrate this by analysis of synthetic seismograms, and the full play back of 6 synthetic events, using 2 algorithms, Elarms/EPIC and ONC-EW. They claim the performance is satisfactory to justify the use of this approach.

General Comments:

The paper is well written (though sometimes the language is a rather colloquial), but the structure and organisation needs more careful attention. With this impressive and long list of co-authors, this seems easy for them to address.

Citations are generally too sparse.

General idea that regionally-collected seismic data is important to EEW. But limited motivation of why this the case. A statement that synthetics can provide regionally-appropriate synthetics, aside from 'depth', needs to be strengthened to be convincing, especially since there is limited regional dependence, since only a 1-D velocity model is used, and site conditions are ignored.

The paper does not recognise the current implementation of ShakeAlert across the West Coast with a now more homogenous approach to algorithms and alerts - I know the paper may be have been prepared before this occurred some months ago, but this is now out of date and should be fixed.

It took me nearly 7-8 pages of text realise the only algorithms being tested are EPIC and ONC-EW. This absolutely crucial information could even be in the title! At least it should go in the introduction. Note since both algorithms appear close to identical in terms of style (with caveat ONC-EW also uses geodetic PGD for magnitude estimation) , this is not a complete test of how appropriate the simulation approach is to test EEW algorithms in Cascadia.

The authors repeatedly mention that seismic data from inertial sensors such as accelerometers cannot be used to used to discriminate magnitudes for very large events >M7. It is the case for algorithms such as EPIC that have a point source location and that rely on station magnitude estimates from short time windows of narrow band data. But it is not the case for algorithms that take advantage of spatial ground motion distributions, such as FinDer. Please fix this in all places.

The paper seems unnecessarily long. Significant Shortening of some of the analysis sections seems quite possible, eg 6.3.

After all the analysis, only providing 6 full scenarios seems 'light'.

Section Comments:

- 1. Introduction is too thin
  - Overview of the variety of methods that can be used to create synthetic algorithm missing.
  - Content of previous usage of synthetic algorithms in seismology is missing.
  - Monitoring history in region is not sufficiently addressed (few large events is mentioned, but nothing specific. No mention that roll-out of EEW-ready seismic network is really quite recent (is it? I think it is), so high quality data cannot be collected.)
  - Make clear that EPIC and ONC-EW are being tested only.
  - I do not know the details of ONC-EW vs ShakeAlert, but this paper does at all enlighten the reader on this, it would be useful to include this context in the introduction.
- 2. EEW section :
  - No mention of new geodetic algorithm in ShakeAlert
  - Current seismic network and EEW-ready roll-out is not properly described for either US or Canada.
  - ONC-EW algorithm is not described at all, and there is no citation. What is it?
  - I realise later there is a drip-drip of info given on how each algorithm works, but a clear summary should be given here. At least point out both are point source approaches using STA/LTA, event associators and station/network magnitude estimates.

    - Fig 1. : needs annotations. Plates, countries / states should be indicated. Since major cities are mentioned they should be shown. Please make clear with shape / colour the difference between events and seismic stations.
    - Description of EEW systems and networks is incomplete / out of date.
- 3. 3.2.1/3.2.2 - To be clear, synthetics are generated by combining 2 independent methods. A deterministic approach for <1Hz, and a stochastic approach >1Hz. For GNSS, on only the deterministic approach is used. For broadband, both are combined. If this understanding is correct, I suggest to clarify this in the text, which is unnecessarily confusing.
- 3.2.3 - the noise is ‘simple’ - overlapping events are avoided, which are typically scenarios causing most problems in EEW. This should be mentioned.

l. 76 small?

L 84 completely?

L 85 Oftentimes is colloquial

L 97 citation needed

L114,115 - odds -> liklihood

L117 be more specific, add citation.

L126 with similar convergence rates? Citetion, examples needed.

L 133 remove However.

L 152 - is FinDer slower? Citation.

L 166 - didn’t similar cables exist decades ago in Japan? Also now exist in Taiwan? Please make clear what is different with NEPTUNE from these.

L 171 fastest -> faster

L 173 misses? Or pulls on-shore?

L175 ‘as real time-series data were used’ - don’t understand.

L 182-185 - is this relevant?

L 199 what system?

L238-239 - really? Sounds like ShakeAlert is not serious about Cascadia. I know you don’t mean this, maybe rephrase

L239 cover? Rephrase. Lie in-between seismic sources and the cities?

L326 the **bandpassed** signal amplitude

L 331 EPIC, **one of 3** algorithms umbrellaed under…

L 335 resulting noisy? Rephrase

L 352 dampen -> lower

L360 that -> for

L351 alerts **and locations** that

L362-364 bigger problem is false locations!

L371 noise -> coda

L 380-381 smaller events have a weaker signal - not really true for some waveform metrics, when smaller is M6.5

L 384 spelling themselves.

Fig 3: events -> station detections. Consider changing to log scale on e,f - any late pick more than 10-20s is irrelevant to EEW, it's too late.

L405 contribute less -> not contribute

L410-412 not clear.

L 485 rephrase

Fig 4: use same y-scale for all, it's confusing to see patterns.

L634 single station seismic recordings. This sentence is not true, FinDer uses seismic recordings and produces non-saturating magnitudes

L641 'testing' Is this US/ShakeAlert centric? Please specify what this means?

L 651 different noise percentiles? This is not introduced in the generating seismogram section. Its used a lot subsequently. Please fix it.

Fig 9 - station colours are difficult to distinguish.

L 788. ShakeAlert? So both FinDer and EPIC? Or only EPIC?

L805-809 - this basic information is very late to introduce - it could go in the intro or at least the EEW section!

L 815 - ‘exactly’ is a bit too much.

L 818 2 ln units?

L 820 the delays are very significant and not remarked on. Event 65 is only seem at once time, which is not explained.

L 824-828 confusing, rephrase.

L 839 successful at detecting synthetic seismograms from large…

Reviewer 2: Martijn van den Ende

This manuscript considers the use of synthetically-generated data to complement the lack of strong-motion data to calibrate/test earthquake early warning systems in the Cascadia region, and describes offline tests of the two of those systems, EPIC and ONC-EW. The objective is to create a diverse data set of large earthquake recordings that is sufficiently realistic to act as a surrogate for real-world data. The study is quite extensive and the manuscript is certainly on the longer side, but the authors do a good job explaining the various components and results in a clear manner, which allowed me to read through the manuscript with relative ease. While I have some comments on the interpretation of the data and some of the choices that the authors have made, I do think that the study is relevant and that the manuscript is in good shape.

Since there was plenty of material to read, I also have plenty of comments, but none of these would affect the results or conclusions of the study. I hope that the authors will consider them for the next iteration of the manuscript.

Kind regards,

Martijn van den Ende

General comments:

1) Point clouds tend to draw the eye towards outliers, and obscure the highest point densities by occlusion (stacking 10 points would look the same as stacking 1000). Given how often the authors discuss trends, variance, and mean values, it would make a lot of sense to replace the point cloud representations with 2D kernel density plots (a simple example would be Python Matplotlib’s “hexbin” representation). Kernel density plots are a

much more faithful representation of the distribution of the data in 2D, and allow the reader to assess the local mean and variance of the distribution without being influenced by outliers and occlusion. The downside is that the colour of such plots would represent the point density, and so colours cannot be used to represent e.g. hypocentral distance or magnitude. I don't know if the authors already considered this type of visualisation, but I can imagine that it could improve the data representation in Figs. 3-7. This is especially true for Figs. 5, 6, and 7, where the colours do not encode information other than the data source (which seem to overlap for the most part).

Related to this point, I think that the clarity of Fig. 3 could be improved a bit. Panels (c) and (d) show the detected and missed events with different symbols, but the missed events overlay the detected events, and the detected events overlay themselves, to an extent that a large part of the panels reduces to a purple blob. I would experiment with separating the detected and missed events, so that at least these two classes can be correctly identified. For panels (e) and (f) I find the representation somewhat unintuitive. Currently, the colours carried by the transparent symbols blend into greenish blobs, and I wonder if the appearance is affected by the order in which the symbols are plotted (is the insertion ordered by magnitude or is it random?). I personally would have plotted magnitude versus distance, and then colour the symbols according to their misfit. In such a representation, the colours should start to trace out iso-intensity contours, with a cluster of high misfit in one corner representing the S-picks. Again, I don't know if the authors already considered this and opted for the current representation, and it may not improve the picture at all, but to me it would seem worth the effort.

2) A large part (if not the majority) of this study revolves around ground motions. I find it a little unsatisfactory to empirically calibrate the simulated attenuation against a different but likewise empirically calibrated ground motion model (GMM), which itself is calibrated against real-world recordings. Following this calibration, the calibrated simulations are then extensively compared with the calibrated GMMs, which may or may not be correctly representing the real data.

I am not at all a ground motions expert, and so I suppose that this is the best way to do things, but I can imagine that a reader (myself included) would feel a little cheated when the authors make claims regarding the accuracy of their approach. To their credit, the authors do acknowledge this in lines 747-756, and I fully understand that there is a severe lack of strong-motion data to directly compare against. I also cannot provide concrete suggestions to improve upon this approach, and so the authors should just largely ignore this comment. Perhaps one thing that can alleviate my unease is a brief discussion of the (quantified) uncertainties of the BCHydro / NGA-Sub GMMs, which should provide an indication of the inaccuracies that can be expected in real-world scenarios (basically expanding lines 747-756 with a quantitative justification).

Aside from this philosophical point, I think it would improve the clarity of the manuscript if the authors could describe the attenuation relation used by FakeQuakes. The authors describe the modifications that are necessary to better fit the GMMs, but they do not describe how attenuation is treated by FakeQuakes (unless I missed it). I looked-up the Melgar et al. (2016) reference that is cited in line 197, but a quick scan did not lead me to a description of the attenuation relation that is supposedly modified for this study (again, I might have missed it). I hope that the authors could provide a more self-contained description of this, so that an interested reader does not need to search the literature/GitHub codes to understand the foundation of this study.

3) This study is quite extensive and has many components (slip model, lo/hi-freq synthetic waveforms, modified attenuation relationship, conversion to PG[A/V/D] and MMI, STA/LTA detection, comparison with BCHydro / NGA-Sub, comparison with GA21, offline tests of EPIC and ONC-EW), and the reader has many things to keep in the back of their minds as the manuscript progresses. Perhaps it would help readers with a short memory, like myself, to include a summary diagram or flowchart that mentions the various components and tests that this study features, and how they relate to each other.

Specific comments:

- Lines 195-210: perhaps merging these two paragraphs and changing the order a bit could improve readability: “We curate a simulated dataset [...] using FakeQuakes [...]. Various methodologies exist [...]. Our objective is to assess [...]”.

- Section 3.2.2: the first paragraph explains why a deterministic approach wouldn’t be feasible, but it doesn’t describe the stochastic approach in the same level of detail as the low-frequency synthetics (Section 3.2.1). The second paragraph only describes parameter choices and justification. A few words describing the approach of generating the high-frequency synthetics would be appropriate.

- Line 301, “unusual responses”: what kind of “responses” are we talking about? Are these simply unidentified noises? If so, including them would make sense, as they would occur in a real-world EEW scenario.

- Lines 332-333: I presume that the STA/LTA parameters (0.05/5 and 1/5) are the window lengths in units of seconds?

- Lines 349-351: I think that many readers could wonder about the false-positive rate, which is just as important as the detection rate. Did the authors check this (by testing on just noise), or is there a general reference on the false-positive rate of EPIC/ONC-EW? This would be a bit outside of the scope (as it pertains to EPIC/ONC-EW specifically), but if the authors have a ready answer it would be good to mention it somewhere.

- Line 355, "relatively no bias": relative to what?

- Line 360: "recordings that detected events" = "recordings on which detections were triggered"

- Lines 379-381: the authors state here that the detector has larger residuals for moderate magnitude events, which they attribute to the lower SNR. I'm not sure I agree with this argument, since the waveform onset should be independent of the eventual magnitude of the event. At the time scale of a trigger (the STA of < 1 sec), the earthquake has not yet reached its final magnitude, and so all earthquakes should look the same in the eyes of a causal detector.

- Line 384: "themsevles" = "themselves"

- Lines 409-412: I actually have a hard time imagining a different way to combine signal and noise, other than what the authors adopted for this study. How did the EPIC operators do this summation? Did they pad the noise with zeros?

- Lines 419-421: I think it would be fair to also include a reference to Men-Andrin Meier's work (the 2021 GJI paper specifically), as the debate regarding earthquake determinism is far from settled (which this paragraph seems to allude to).

- Lines 451-452: this S-wave truncation is easily done with synthetics, but not in the real world. To allow the effect of SNR to propagate into the result, would it make sense to adopt a real-world methodology to isolate the P-wave? (fixed time window after first trigger, or theoretical S-wave arrival based on estimated hypocentre, or something else?)

- Lines 461-463: I don't quite understand the "1.5 log unit" selection criterion. The word "extending" suggests that previously the dataset was restricted to some arbitrary misfit value, and I don't see why a misfit selection criterion should be used (instead of a distance criterion).

- Lines 463-465: the apparent larger variance could simply be a visualisation artefact (as pointed out above); a non-Gaussian distribution with wide tails would appear similar to a Gaussian distribution with large variance. Histograms or kernel density estimates should clarify this.

- Lines 471-473: it would probably make sense to describe this (also) in section 3.2.2 (line 275).

- Line 558: did I understand correctly that Qc = 150 f^0 means that Qc simply assumes a constant value of 150? It would be good to explicitly mention this to avoid confusion.

- Line 562, “the frequency factor of Qs was changed from 0.6 to 0.8”: what is this frequency factor? I suppose that if the authors expand the description of the attenuation model, this question is automatically answered.

- Lines 577-578, “Recordings with the largest Rrup come from the smaller magnitude events”: shouldn’t this be the larger magnitude events (higher SNR)?

- Lines 579-580: are BCHydro / NGA-Sub not also 1D GMMs? They are calibrated against data from the 3D world, but in the end the spatial information is encoded by the hypocentral distance. In that case, discrepancies between the simulated data and the GMMs cannot be attributed to 3D structure.

- Lines 619-620: as for the attenuation model of FakeQuakes, I think it would help to briefly describe the scaling models of Worden et al. (2012) here (or in an appendix).

- Lines 699-701: unless I’m mistaken, if the signal and noise are independent random distributions, then the superposition of signal and noise should have a higher expected PGD than the PGD of the pure signal. If the GMMs are calibrated against real-world data that include noise, then the GMM-predicted PGD should be similar to the simulated PGD + 100% noise, and larger than the simulated PGD + 0% noise. Fig. 7 reports the opposite trend (as a function of magnitude = signal amplitude). What this figure seems to suggest instead, is that the authors greatly overestimate the amplitude of the noise that is included in the calibration of the GMMs, and that a value of around 5% seems to be a much more suitable value (giving a flat trend as a function of magnitude).

- Line 733: “regardless of [absolute] noise levels” (to avoid confusion with SNR).

- Lines 849-852: this is a valid argument for EPIC, but would it be possible to compare ground motions predicted by ONC-EW?

- Lines 873-878: the main motivation for using FakeQuakes is that deterministic models are too expensive to run at the required resolution. It is therefore a bit surprising that the authors suggest improving the study with a 3D velocity structure and 3D Earth waveform modelling, which brings us back to the motivation for using a simple 1D model in the first place. I certainly agree with the authors that their suggestions would make for a great leap forward, but these are contingent on improved numerical infrastructures or algorithms.

- Data and code availability: I would strongly recommend to give version numbers of the software that was used (the authors already noted that changes in the FakeQuakes software could be the source of discrepancies with previous studies), and ideally to also copy the relevant codes to a long-term archiving repository like Zenodo or Figshare (Seismica has a “community” on Zenodo). It is a real possibility that Microsoft will impose restrictions on public Github repositories in the future, or change the URLs that will result in broken links. Zenodo/Figshare will stamp the repository with a DOI and guarantee its long-term accessibility. The DOI of the “ONC & PNSN (2024)” reference does not resolve.

Review Round 2 – Comments from Reviewer 2:

I think that the authors put in a lot of work to improve the manuscript, and I agree with their decisions regarding visualisation and data presentation. I only have one major comment regarding attenuation (lines 1265-1317 of the tracked-changes document). I'm aware that this section was also present in the previous version of the manuscript, but somehow I didn't entirely pick up on it.

Regarding point (1): is the first arrival not always the wave packet that arrives first, regardless of whether it followed a crustal or Moho trajectory? Hence, if a sub-Moho phase arrives first, then "direct arrivals used to generate S-wave packets were taken as the first arriving ray" should use those phases? Does the waveform generation method not perform a summation over all S-arrivals, accounting for each phase's arrival time and attenuation separately (Eq. S3)?

Regarding point (2): what is the justification for abandoning the strong physical/geometric constraint of 1/R scaling? Sahakianet al. arbitrarily fixed this parameter to get a better fit, but acknowledge the trade-offs with other parameters, notably the intrinsic attenuation term. If I'm not mistaken, their "a5" term would be inversely proportional to Q_{s,eff} here (Eq. (5)), so a similar bias reduction could have been obtained with the more physical 1/R model and changing Qc a little bit. Since Qc is used as a free variable to reduce the bias, it can be arbitrarily adjusted to accommodate the more physical 1/R scaling. This is not meant as strong criticism (since in the end it should make little difference), but perhaps the authors could add one line to justify this (e.g. for consistency with previous work, or based on other theoretical considerations).

Regarding point (3): Perhaps it is my lack of expertise in attenuation modelling, but to my understanding this point discusses only anelastic (intrinsic) attenuation, with a potentially frequency-dependent quality factor Qc. The frequency dependence is omitted, so that Qc is constant, which is combined with the likewise constant Qs to produce another constant Q_{s,eff}. This effective attenuation term enters Eq. (S3), modelling the anelastic attenuation component. In this formulation, Qc and Qs are indistinguishable, and since a correction is made to keep the overall attenuation constant (L1315), I don't see the argument that the authors are trying to make. To account for the scattering component, an integral equation like Eq. (2) in Pujades et al. (1997) needs to be solved, where Qc and Qs are incorporated separately. Perhaps this is the case, in which case the authors could extend Supplementary Section S1 to explain the separate contributions of Qc and Qs. If this is not the case, and only anelastic/intrinsic attenuation is modelled, then I would omit this section to avoid confusion.

Besides this main point, I have a number of smaller comments. These line numbers refer to the tracked-changes document:

- L420: "through [a] 3D velocity structure"
- L607: thrtee = three
- L775: since the term "aseismic" carries a "slow fault slip" connotation, it is perhaps wiser to replace "aseismic triggering" by "non-event triggering", "false-positive triggering", or similar.
- L876, "Once an event is created": perhaps a better term would be "detected" or "confirmed"?
- L887: the placement of the reference to Meier et al. makes it appear as if they support the idea of "the rupture organizes into a slip pulse", which I don't think is the case (they argue the opposite). It would be better to put this reference in "(i.e., distinguishable within the first few seconds; Meier et al, 2021)".
- L971: patterns present [in] the Trugman
- L1103: hunders = hinders
- Eq. (4): the "PGD" in this equation refers to the GMM PGD right? (not the simulated PGD)
- L1145-1150: the phrasing here causes a little bit of confusion about whether the difference in sign convention between the present work and Melgar's is already accounted for. What does "they have a different implication" mean?
- L1210: broadband waveforms. [A] 0.075 Hz
- L1253: Is the R_rup defined here the same as the R_p defined in Eq. (3)? I suspect not, but it's better to ask than to guess
- L1397: sensitivite = sensitivity
- L1757: why is it not clear why the geodetic data did not contribute? Isn't this a synthetic test where all inputs are well-controlled and known?
- Fig. 9: if I understood correctly, the fault kinematic models include a time-dependent rupture evolution, and hence the "true magnitude" should evolve with time, rather than being a constant. This is just a small detail and I don't think it is critical for this study, but it is a real-world aspect that is typically included in offline EEW tests.
- L1839: conern = concern
- L1855: spport = support
- L1857: repsect = respect

Kind regards,

Martijn van den Ende

Dear Dr. Christie Rowe,

We want to thank you for your time and effort overseeing the review process and for the opportunity to respond to the reviews. We sincerely appreciate each of the reviewer's thoughtful comments, and we have taken the time to thoroughly address each one. Our manuscript has been substantially revised based on the reviewers' feedback. The overall findings have not changed, but the clarity of the manuscript and figures are significantly improved. Specifically, we have made the following large-scale changes:

1. *More detailed discussion of the EPIC and Ocean Networks Canada early warning algorithms in the introduction.*
2. *Clearer overview of the data processing and analyses provided by the flowchart in Figure 2.*
3. *Discussion of peak ground displacement has been moved from Section 6: Ground-Motion Intensity to Section 5: Magnitude Estimation as PGD is used to estimate earthquake magnitude rather than support shaking intensity estimates.*
4. *Most of the figures have been modified to improve reader interpretation.*
5. *The length of the manuscript has been shortened by moving some information to the supplement and by reducing overly detailed or unnecessary text.*

Please see our detailed responses to the review comments below.

Sincerely,

Tara Nye et al.

Reviewer A:

Testing EEW systems cannot be done by regionally-appropriate datasets since the range of target events are no-where near sufficient in the existing recorded archives, since large events occur rarely and EEW-ready seismic networks are typically only very recently in operation, for years or at most decades. This paper demonstrates that EEW systems can be tested using synthetic waveforms, with application to the Cascadia region. They demonstrate this by analysis of synthetic seismograms, and the full play back of 6 synthetic events, using 2 algorithms, Elarms/EPIC and ONC-EW. They claim the performance is satisfactory to justify the use of this approach.

General Comments:

The paper is well written (though sometimes the language is a rather colloquial), but the structure and organisation needs more careful attention. With this impressive and long list of co-authors, this seems easy for them to address.

We thank the reviewer for their feedback and have restructured the manuscript for clarity and conciseness. Revisions are described in more detail in our responses to the reviewer's list of comments.

Citations are generally too sparse.

We included additional citations in the manuscript, which are listed at the end of this document.

General idea that regionally-collected seismic data is important to EEW. But limited motivation of why this the case. A statement that synthetics can provide regionally-appropriate synthetics, aside from 'depth', needs to be strengthened to be convincing, especially since there is limited regional dependence, since only a 1-D velocity model is used, and site conditions are ignored.

We are a bit confused what the reviewer wanted us to address, but we think they were implying that the simulated ground motions are not necessarily "regionally appropriate" for Cascadia considering a simple 1D velocity model was used and shallow site conditions were not incorporated. The reviewer is correct that we did not account for full regional heterogeneity in our simulations. Considering the instrumental record for large Cascadia events is limited to non-existent, we would not be able to truly validate the regional appropriateness of the simulations even if we used a complex model. Simulated ground motions have been less validated in the literature, especially higher-frequency ground motions, compared to the rupture kinematics. Thus, our primary objective was to ensure the ground motions match global empirical relations for subduction zone earthquakes, rather than assess their regional appropriateness. We believe this is adequately addressed in the text.

The paper does not recognise the current implementation of ShakeAlert across the West Coast with a now more homogenous approach to algorithms and alerts - I know the paper may be have been prepared before this occurred some months ago, but this is now out of date and should be fixed.

We thank the reviewer for bringing this to our attention. This manuscript was originally submitted prior to the recent McGuire et al. (2025) publication, which details the current implementation of ShakeAlert. We have updated the introduction to reflect these changes. Please see the revised introduction paragraph on ShakeAlert below (modifications shown in bold):

*"ShakeAlert is **the** U.S. **EEW system, which expanded coverage from California to the entire west coast in 2020 (Kohler et al., 2020). The current version of ShakeAlert (V3; McGuire et al., 2025)** is supported by **three source characterization** algorithms**:** EPIC, FinDer**, and GFAST-PGD**. EPIC is a modified version of the ElarmS**-3** point source algorithm **(**Chung et al., 2019**; Kohler et al., 2018**), **which uses a short-term average/long-term average (STA/LTA) trigger for event detection. Once an event is detected, the earthquake magnitude and hypocenter location are estimated through source characterization algorithms using the first few seconds of the P-wave amplitude**. FinDer is a finite fault algorithm, which **uses evolving peak ground accelerations (PGA) to estimate the line source and characterize the rupture for moderate-to-large events** (Allen and Melgar, 2019; Böse et al., 2018, 2023). It is slower than EPIC but provides more accurate **magnitude** estimates and is used to update alerts as the rupture progresses **(Böse et al., 2022). GFAST-PGD is the newest algorithm and uses geodetic peak ground displacement (PGD) from Global Navigation Satellite System (GNSS) data to estimate event magnitudes for large earthquakes given a location estimate from seismic data (Crowell et al., 2016; Murray et al., 2023). The Solution Aggregator aggregates source info from the three algorithms, which is fed into EQinfo2GM to estimate shaking intensity (Thakoor et al., 2019).**"*

It took me nearly 7-8 pages of text realise the only algorithms being tested are EPIC and ONC-EW. This absolutely crucial information could even be in the title! At least it should go in the introduction. Note since both algorithms appear close to identical in terms of style (with caveat ONC-EW also uses geodetic PGD for magnitude estimation), this is not a complete test of how appropriate the simulation approach is to test EEW algorithms in Cascadia.

We thank the reviewer for their comment and agree that the algorithms should have been specified earlier on in the manuscript. Additionally, we never intended to present a complete test of the EEW algorithms in Cascadia. Rather, our focus was on validating simulations, and we wanted to show an example application of running the simulations through offline EEW tests. We recognize that the initial phrasing did not make this clear. We have addressed these concerns in the abstract:

*"We also **present an example** performance **test** of the **ShakeAlert EPIC and Ocean Networks Canada early warning** algorithms using **six of the** simulated events."*

We have also addressed these concerns in the last paragraph of Section 1:

*"In this study, we (1) generate a comprehensive dataset of kinematic slip models and associated waveforms for use in testing EEW, (2) evaluate these data based on the key components of EEW systems (i.e., earthquake detection, magnitude estimation, and ground-motion **intensity**) using the CSZ as a case study, and (3) **run** a subset of the simulations **through offline tests using the ShakeAlert EPIC and Ocean Networks Canada early warning algorithms.**"*

The authors repeatedly mention that seismic data from inertial sensors such as accelerometers cannot be used to used to discriminate magnitudes for very large events >M7. It is the case for algorithms such as EPIC that have a point source location and that rely on station magnitude estimates from short time windows of narrow band data. But it is not the case for algorithms that take advantage of spatial ground motion distributions, such as FinDer. Please fix this in all places.

We thank the review for their correction. They are right that FinDer does not saturate in the way that a point source algorithm like EPIC does. Although, FinDer estimates can be biased if the ground-motion model used does not adequately reflect ground motions in the region (Böse et al., 2015). We have corrected the manuscript text. See for example:

*"FinDer is a finite fault algorithm**, which uses evolving peak ground accelerations (PGA) to estimate the line source and characterize the rupture for moderate-to-large events** (Allen and Melgar, 2019; Böse et al., 2018, 2023)."*

The paper seems unnecessarily long. Significant Shortening of some of the analysis sections seems quite possible, eg 6.3.

We agree with their review that the original manuscript was a bit long and verbose. We made the following revisions to address this:

- Section 3 Stochastic Rupture Simulations: Details of user-defined simulation parameters have been moved from the main text to the supplement (Text S1).
- Section 5 Magnitude Estimation: Introduction shortened.
- Section 6.3 (now Section 5.2) Peak ground displacement: In-depth discussion of GNSS signal-to-noise ratio and PGD estimates was removed, and the corresponding figure (Figure 8) was moved to the supplement (Figure S4).
- Section 6.2 Modified Mercalli intensity: Section shortened.
- Section 8 Conclusions: Section significantly shortened.
- Throughout the manuscript, text was revised to be more concise where possible.

After all the analysis, only providing 6 full scenarios seems 'light'.

If the main purpose of this study was a full performance evaluation of the earthquake early warning algorithms, we would agree that 6 scenarios would appear insufficient. However, our main goal was to evaluate the simulations and provide an example of their implementation into offline tests as a proof of concept. The computational expense of running these simulations through both EPIC an ONC-EW makes testing a large subset of simulations beyond the scope of this study. We added the following statement for clarification:

"We limit the analysis to six scenarios due to the computational expense of running offline tests, but we recommend using the full dataset if conducting a comprehensive analysis."

Section Comments:

<u>Section 1. Introduction is too thin:</u>

Overview of the variety of methods that can be used to create synthetic algorithm missing.

We have carefully considered the reviewer's comment, but we do not find it necessary to add a section to the introduction discussing synthetic data methods. At the beginning of Section 3: Stochastic Rupture Simulations, we state:

*"We curate a simulated dataset for the CSZ consisting of 112 kinematic rupture scenarios and associated waveforms using FakeQuakes**, an earthquake simulation module from the MudPy repository** (Melgar et al., 2016; **Melgar et al., 2025**). Various simulation methodologies exist, ranging from simple stochastic models of high-frequency ground motions (e.g., Boore, 2003) to complex deterministic models involving numerical simulations of full waveform propagation through a 3D velocity structure (e.g., Komatitsch and Tromp, 1999; Faccioli et al., 1997; Mazzieri et al., 2013). FakeQuakes is a 1D semi-stochastic broadband simulation model, which provides more realistic ground motions compared to a simple stochastic model, yet is efficient enough to **support** large data generation. FakeQuakes is not the only existing 1D semi-stochastic model; however, it is preferred for this study, as opposed to other similar models (e.g., SCEC Broadband platform, Maechling et al., 2015), as it allows for P-wave generation (necessary for EEW) and has previously been tested in the Cascadia region (Melgar et al., 2016; Goldberg and Melgar, 2020; Small and Melgar, 2023)***.**

Content of previous usage of synthetic algorithms in seismology is missing.

We expanded our original statement that "earthquake simulations have been widely used in various applications" by providing several example applications and references. Please see the revised section as follows:

*"**E**arthquake simulations have been used in **a variety of** applications**, including investigating effects of Earth structure on rupture kinematics (e.g., Rodgers et al., 2019; Harris et al., 2021), improving seismic hazard assessments (e.g., Aagaard et al., 2001; Graves et al., 2011; Melgar, 2025; Poulos et al., 2025), and training EEW systems (Lin et al., 2021)**. **However,** validation of the models used to generate the data have focused primarily on rupture physics, and to some extent ground motions (e.g., Lin et al., 2021; Melgar et al., 2016; Small and Melgar, 2023)."*

Monitoring history in region is not sufficiently addressed (few large events is mentioned, but nothing specific. No mention that roll-out of EEW-ready seismic network is really quite recent (is it? I think it is), so high quality data cannot be collected.)

We thank the reviewer of their comment; however, we do not find it necessary to include a detailed summary of the monitoring history in the region as there has not been much significant seismicity during the instrumental period. Additionally, although public roll-out of EEW in this region is relatively recent, previously recorded earthquakes can be used in offline replay tests of EEW systems. Thus, the limiting factor is the regional seismicity rather than the timeliness of EEW roll-out. We added the following statements to the Section 2: EEW in the Pacific Northwest, which we hope addresses the reviewer's concerns.

"Historically, the CSZ has hosted large M 8–9 earthquakes, with the last major rupture occurring in 1700 (Satake et al., 2003; Walton et al., 2021; Wirth et al., 2021). Although the CSZ is capable of generating large earthquakes, it experiences very little background seismicity, other than intraslab earthquakes and a few patches along the megathrust offshore central Oregon (Tréhu et al., 2015; Wang and Tréhu, 2016)."

“ShakeAlert is the U.S. EEW system, which expanded coverage from California to the er [illegible] west coast in 2020 (Kohler et al., 2020).”

“Ocean Networks Canada (ONC) developed an EEW system for Southwestern British Columbia, Canada (hereafter, ONC-EW; Schlesinger et al., 2021) in a collaborative effort with Natural Resources Canada (NRCan), which has been operational since 2023 (McMartin, 2025).”

Make clear that EPIC and ONC-EW are being tested only.

We addressed this in our response to one of the reviewer’s general comments above.

I do not know the details of ONC-EW vs ShakeAlert, but this paper does at all enlighten the reader on this, it would be useful to include this context in the introduction.

We have addressed this concern in response to one of the comments under “Section 2. EEW section:”

Section 2. EEW section:

No mention of new geodetic algorithm in ShakeAlert

We did include a reference to G-FAST in our original submission. Please see the original text below:

“Three HR-GNSS early warning 640 algorithms have previously undergone testing— G-larmS (Grapenthin et al., 2014a, 2014b), G-FAST (Crowell et al., 2016), and BEFORES (Minson et al., 2014). G-FAST, which includes PGD-magnitude scaling as a key element, has been selected for implementation into ShakeAlert (Murray et al., 2023) and is active as of June 2024.”

We updated this section to include a more recent reference detailing the current state of practice for ShakeAlert (McGuire et al., 2025).

Current seismic network and EEW-ready roll-out is not properly described for either US or Canada.

This has been addressed in our response to other comments by this reviewer.

ONC-EW algorithm is not described at all, and there is no citation. What is it?

The ONC-EW algorithm is not as well documented in public literature as the ShakeAlert algorithms; however, we did include multiple citations in our original submission. Schlesinger et al. (2021) is the main publication summarizing the EEW system. Rosenberger (2010, 2019) and Rosenberger et al. (2018, 2019a, and 2019b) detail different components of the algorithms used. We agree that we did not thoroughly describe the ONC-EW system in the introduction. We moved some of the ONC-EW algorithm discussion to the introduction to help make this clearer early on.

“ONC-EW’s network configuration is leading edge as it comprises collocated High-Rate Global Navigation Satellite System (GNSS) and strong-motion sensors, which only exists for a handful of ***sites*** *around the world (Schlesinger et al., 2021).* ***This unique network configuration allows for***

onsite combination of high-frequency seismic data with unsaturated geodetic data via a Kalman filter, without the typical time latency transferring information between stations (Bock et al., 2011; Kalman, 1960; Melgar et al., 2013; Rosenberger et al., 2018; Smyth and Wu, 2007). The resulting un-biased displacement and velocity time-series are used for all steps in the EEW system workflow, which includes STA/LTA event detection and P-wave magnitude estimation."

I realise later there is a drip-drip of info given on how each algorithm works, but a clear summary should be given here. At least point out both are point source approaches using STA/LTA, event associators and station/network magnitude estimates.

We thank the reviewer for their comment. We have updated our descriptions of EPIC and ONC-EW in the introduction to provide a comprehensive description of the algorithms.

For ShakeAlert we revised the text as follows:

*"EPIC is a modified version of the ElarmS-**3** point source algorithm (Chung et al., 2019; **Kohler et al., 2018**), **which uses a short-term average/long-term average (STA/LTA) trigger for event detection. Once an event is detected, the earthquake magnitude and hypocenter location are estimated using the first few seconds of the P-wave amplitude.** FinDer is a finite fault algorithm, which us**es evolving peak ground accelerations (PGA) to estimate the line source and characterize the rupture for moderate-to-large events** (Allen and Melgar, 2019; Böse et al., 2018, 2023). It is slower than EPIC but provides more accurate **magnitude** estimates and is used to update alerts as the rupture **progresses (Böse et al., 2022). GFAST-PGD is the newest algorithm and uses geodetic peak ground displacement (PGD) from Global Navigation Satellite System (GNSS) data to estimate event magnitudes for large earthquakes given a location estimate from seismic data (Crowell et al., 2016; Murray et al., 2023). The Solution Aggregator aggregates source info from the three algorithms, which is fed into EQinfo2GM to estimate shaking intensity (Thakoor et al., 2019). Around 1400 seismic stations and 1100 GNSS stations can contribute to ShakeAlert at any given time (McGuire et al, 2025)."***

For ONC-EW, we revised the text as follows:

*"Ocean Networks Canada (ONC) developed an EEW system for southwestern British Columbia, Canada **(hereafter,** ONC-EW**; Schlesinger et al., 2021**) in a collaborative effort with Natural Resources Canada (NRCan)**, which has been operational since 2021 (McMartin , 2025)**. ONC-EW's network configuration is leading edge as it comprises collocated High-Rate Global Navigation Satellite System (GNSS) and strong-motion sensors, which only exists for a handful of **sites** around the world (Schlesinger et al., 2021). **This unique network configuration allows for onsite combination of high-frequency seismic data with unsaturated geodetic data via a Kalman filter, without the typical time latency transferring information between stations (Bock et al., 2011; Kalman, 1960; Melgar et al., 2013; Rosenberger et al., 2018; Smyth and Wu, 2007). The resulting un-biased displacement and velocity time-series are used for all steps in the EEW system workflow, which includes STA/LTA event detection and P-wave magnitude estimation."***

Fig 1. needs annotations. Plates, countries / states should be indicated. Since major cities are mentioned they should be shown. Please make clear with shape / colour the difference between events and seismic stations.

We thank the reviewer for their feedback. We revised Figure 1 to include annotations for the countries (U.S. and Canada), states (Washington, Oregon, and California), Vancouver Island, and the tectonic plates (Juan de Fuca, Explorer, and Gorda plates). In our original figure, the events were circles, and the stations were different shapes depending on the sensor type and operating networks. The events are now open stars to match the star used to indicate the epicenter in panel b. Additionally, we lightened the bathymetry/topography colors to distract less from the events and stations.

Description of EEW systems and networks is incomplete / out of date.

We thank the reviewer for bringing this to our attention. We have updated our description of the EEW systems to reflect current state of practice. For example, we now state:

"ShakeAlert is the ***U.S.*** *EEW system****, which expanded coverage from California to the entire west coast in 2020 (Kohler et al., 2020). The current version of ShakeAlert (V3; McGuire et al., 2025)*** *is supported by* ***three source characterization*** *algorithms****:*** *EPIC****,*** *FinDer,* ***and GFAST-PGD****."*

Please see Section 2, paragraphs 4–5 for full revisions.

<u>Section 3:</u>

3.2.1/3.2.2 - To be clear, synthetics are generated by combining 2 independent methods. A deterministic approach for <1Hz, and a stochastic approach >1Hz. For GNSS, on only the deterministic approach is used. For broadband, both are combined. If this understanding is correct, I suggest to clarify this in the text, which is unnecessarily confusing.

We streamlined the text for clarity as follows:

"A ***semi-****stochastic approach is used* ***to model broadband ground motions due to velocity model limitations and to reduce computational cost at higher frequencies. Small-scale CSZ velocity features are not well-resolved with limited offshore instrumentation. The low frequencies (< 1 Hz) are generated using the deterministic method described in the previous section, and the*** *high frequencies (> 1 Hz)* ***) are generated by summing the spectral response for each subfault assuming a random (i.e., stochastic) phase*** *(Graves and Pitarka, 2010, 2015):"*

3.2.3 - the noise is 'simple' - overlapping events are avoided, which are typically scenarios causing most problems in EEW. This should be mentioned.

We thank the reviewer for their comment, but we disagree that the noise is 'simple'. We combined the signal with representations of what real station noise looks like. We chose not to include noise segments with spurious signals because we did not want those artifacts to get mapped into our validation analyses, as our focus was on validating the signal. In the accompanying dataset (Ocean Networks Canada Society & Pacific Northwest Seismic Network, 2026), we include the raw simulated waveforms and the noise-combined waveforms. Users implementing our dataset are welcome to use the raw simulations and add complex noise for their analyses. The decision to include noise with other signals would affect station triggering but would likely not impact EEW performance as multiple stations would need to trigger for an event to be created. It's true that overlapping events are challenging for EEW, but to test this, we would need to create multi-event simulations. It would not be a realistic test case if only one station had noise contaminated by another earthquake. We added the following sentence at the end of the paragraph:

"We recognize that these scenarios do not include multi-events, such as in the case of foreshocks or aftershocks, which are a challenge for EEW systems. That is beyond the scope of this project, but could be an avenue for additional performance testing."

Line-by-line Comments:

l. 76 small?

We clarified the sentence as follows:

"Although varying in technical details, the first alert issued by most EEW systems is from information recorded in the initial P-waves, ***which are not strong enough to cause noticeable shaking****."*

L 84 completely?

We clarified the sentence as follows:

"Testing and validating EEW systems performance is critical because the algorithms are *fully automated and have no human interaction at any point in the processing chain, and because the consequences of missed or false alerts are significant.****"***

L 85 Oftentimes is colloquial

We removed the word "Oftentimes".

L 97 citation needed

We added the following citations: (Rodgers et al., 2019; Harris et al., 2021; Aagaard et al., 2001; Graves et al., 2011; Melgar, 2025; Poulos et al., 2025; Lin et al., 2021)

L114,115 - odds -> liklihood

We accepted the reviewer's suggestion.

L117 be more specific, add citation.

We revised the sentence as follows:

"The ***likelihood*** *of a very large (****M*** *~8)* ***partial margin*** *rupture* ***initiating*** *along the southern margin is around 37****–42****%* ***(Goldfinger et al., 2012). Such an event could still generate moderate-to-strong shaking (i.e. MMI V–VI) along the central Oregon coast and potentially as far north as the central Washington coast (Priest et al., 2014).****"*

L126 with similar convergence rates? Citation, examples needed.

Regardless of convergence rate, the Cascadia Subduction Zone is unusually quiet, with very few interface earthquakes having been recorded on seismic instruments (Tréhu et al., 2015; Wang et al., 2016). We added these references and revised the sentence as follows:

"Although the CSZ is capable of generating large earthquakes, it experiences very little background seismicity, other than intraslab earthquakes and a few patches along the megathrust offshore central Oregon (Tréhu et al., 2015; Wang and Tréhu, 2016)."

L 133 remove However.

We accepted the reviewer's suggestion.

L 152 - is FinDer slower? Citation.

FinDer is a little slower because it estimates fault rupture extent using evolving ground-motion intensities, rather than a simple point source event detection and magnitude estimate. Böse et al. (2022) stated "FinDer is typically 0.1–0.35 s slower for initial detections". We added this reference to the sentence.

L 166 - didn't similar cables exist decades ago in Japan? Also now exist in Taiwan? Please make clear what is different with NEPTUNE from these.

Although cabled observatories have existed in Japan, NEPTUNE is the first ***regional multi-node*** cabled observatory. It consists of an 800 km cable loop draped over the Juan de Fuca plate boundary and was developed to enhance real-time data acquisition and interdisciplinary scientific analysis of the data and subduction zone. We do not find it within the scope to go into significant detail about NEPTUNE in the manuscript, but we have clarified the sentence as follows:

"Another unique feature of the network configuration is the addition of five subsea strong-motion sensors connected to an underwater cabled network, NEPTUNE (North-East Pacific Time-series Undersea Networked Experiments), which drapes over the CSZ and is the world's first ***regional*** *multi-node cabled ocean observatory* ***(Barnes and Tunnicliffe, 2008)."***

L 171 fastest -> faster

We accepted the reviewer's suggestion.

L 173 misses? Or pulls on-shore?

We modified the text in this section, which involved removing the sentence this comment was referring to.

L175 'as real time-series data were used' - don't understand.

Our intent was to state that the magnitudes were limited in performance analyses by Hartog et al. (2016) because they used real data (rather than simulations). We modified this section and removed this reference, but we have updated the corresponding text as follows:

"Lux et al. (2014) evaluated the overall performance of ShakeAlert, and Babaie Mahani et al. (2023,2024) evaluated components of ONC-EW using recorded offshore earthquakes, but their datasets were limited to M < 7 events, most of which were not slab originating."

L 182-185 - is this relevant?

We find this reference relevant because it describes another example of earthquake early warning performance testing, for ONC-EW, but we revised this sentence to remove extraneous information. Please see our response to the previous comment for the updated text.

L 199 what system?

By "system", we were referring to the set of algorithms used for a particular EEW system. We clarified the sentence as follows:

"Our objective is to assess how well these waveforms perform at replicating reality with respect to key metrics needed for an EEW system, and subsequently ***test the performance of EEW systems using several example scenarios from the simulated dataset."***

L238-239 - really? Sounds like ShakeAlert is not serious about Cascadia. I know you don't mean this, maybe rephrase

We thank the reviewer for the comment. We revised the sentence as follows:

"The Washington and Oregon sites are a subset of ShakeAlert's network configuration, ***focusing on sites at greatest risk for shaking form slab-originating events."***

L239 cover? Rephrase. Lie in-between seismic sources and the cities?

We revised the sentence as follows:

"The selected stations ***include coverage of*** *the major cities landward of the CSZ coastline and are located within a few hundred* ***kilometers*** *of the trench."*

L326 the **bandpassed** signal amplitude

This is not always true. EPIC applies several highpass filters, but not a bandpass filter.

L 331 EPIC, **one of 3** algorithms umbrellaed under...

We revised the text and no longer include the phrase "one of 3 algorithms umbrellaed under".

L 335 resulting noisy? Rephrase

We revised the sentence as follows:

*“For **STA/LTA analyses** using EPIC parameters, we we filter the vertical component of the **noise-combined** acceleration broadband waveforms”*

L 352 dampen -> lower

We removed text from this section, including the sentence this comment was referring to.

L360 that -> for

We revised the sentence and no longer need “that” or “for”:

*“We next analyze **P-wave arrival detection latencies (δ_{arr}) for recordings that triggered the STA/LTA picker, as increased latencies bias location estimates and alert times.”***

L351 alerts **and locations** that

We believe the reviewer was referring to line 361 here. We modified the text to include mention of location biases. Please see the previous comment for the revised text.

L362-364 bigger problem is false locations!

We addressed this in our previous comment.

L371 noise -> coda

We removed this discussion and point the readers to Figure S3 in the supplement. In the figure caption, we state that the STA/LTA picker triggered on coda noise.

L 380-381 smaller events have a weaker signal - not really true for some waveform metrics, when smaller is M6.5.

At equivalent source-to-site distances, peak ground motions correlate positively with magnitude, thus a M6.5 should still exhibit lower peak ground motions than an event with a magnitude > 6.5. We removed “which is expected as smaller events have a weaker signal” from the sentence.

L 384 spelling themselves.

We made this correction.

Fig 3: events -> station detections. Consider changing to log scale on e,f - any late pick more than 10-20s is irrelevant to EEW, it’s too late.

We agree with the reviewer. We updated this figure to plot distance in log scale, and we also truncated the y-axis to only show records with latencies ≤ 5 s.

L405 contribute less -> not contribute

We accepted the reviewer's suggestion.

L410-412 not clear.

We removed this paragraph.

L 485 rephrase

We revised the sentence as follows:

"Regardless, we find that the simulations exhibit the expected trends in P_d evolution, where P_d increases with increased measurement time window ***and*** *larger magnitudes are better resolved when longer P-wave windows are available.* ***For time windows considered in EEW (~4–5 s), the simulations behave as expected."***

Fig 4: use same y-scale for all, it's confusing to see patterns.

We accepted the reviewer's suggestion.

L634 single station seismic recordings. This sentence is not true, FinDer uses seismic recordings and produces non-saturating magnitudes.

We are not sure which line the reviewer was referring to, as this section does not discuss FinDer.

L641 'testing' Is this US/ShakeAlert centric? Please specify what this means?

We removed the sentence mentioning other geodetic algorithms considered for ShakeAlert.

L 651 different noise percentiles? This is not introduced in the generating seismogram section. Its used a lot subsequently. Please fix it.

We thank the author for bringing this to our attention. We now address this in Section 3.2.3 Post processing. Please see the revised paragraph on GNSS noise below:

*"**We apply** synthetic noise **to** the simulated GNSS waveforms **using the** Melgar et al. (2020) **noise model**. GNSS noise is more variable over time compared to that of inertial instruments, thus one day of noise would not be an adequate representation of conditions over a long period of time. Moreover, the synthetic noise model from Melgar et al. (2020) was derived using data from the Cascadia region, rendering it a better average compared to a single day of real GNSS data. **We generate 10th, 50th, and 90th percentile noise time-series to simulate low-, median-, and high-noise scenarios."***

Fig 9 - station colours are difficult to distinguish.

We revised the figure to use more visually distinct colors for the different sets of stations.

L 788. ShakeAlert? So both FinDer and EPIC? Or only EPIC?

The study that we referenced here (Hartog et al., 2016) evaluated the performance of ELARMS-2 (a predecessor of EPIC), but they referred to the system as ShakeAlert because this was the only active algorithm used by ShakeAlert at the time. We revised the corresponding sentence as follows:

"Another known challenge ***for point source algorithms is that large events are sometimes split into*** *separate events (Hartog et al., 2016), which can significantly affect the issued alerts."*

L805-809 - this basic information is very late to introduce - it could go in the intro or at least the EEW section!

We agree, and we moved the discussion of the Kalman filter to the EEW. This was addressed in our response to this reviewer's Section 2. EEW set of questions.

L 815 - 'exactly' is a bit too much.

We removed "exactly" from the sentence.

L 818 2 ln units?

We revised the sentence as follows:

"PGD is on average ***two natural log*** *units larger than expected"*

L 820 the delays are very significant and not remarked on. Event 65 is only seem at once time, which is not explained.

There was a significant delay because the closest stations used for the ONC-EW offline tests were much farther from the rupture compared to the closest stations used for the EPIC tests. It is not clear why ONC-EW only reported one magnitude estimate. We added the following discussion:

"However, ONC-EW significantly under-reported the magnitudes of ***scenarios*** *65 and 96,* ***which*** *initiated south of their network and detection grid.* ***The initial magnitude estimates for these events were delayed due to the travel time to reach the ONC-EW network stations. For scenario 65, geodetic data did not contribute to event estimation, so only seismic P_d was considered. It is not clear why the geodetic data did not contribute nor why only one magnitude estimate was reported. This warrants further investigation by ONC. Even though scenario 96 was a full margin rupture, it ruptured bilaterally, and the southern extent of the rupture propagation was not captured by the limited ONC network coverage. Inclusion of more PNSN stations near the southern border of the detection grid could improve detection time and magnitude estimation, but the current ONC-EW algorithm is unable to support large networks."***

L 824-828 confusing, rephrase.

We revised this section as follows:

"We did not include earthquake location ***estimation as a component of the*** *validation tests* ***in the previous sections because location accuracy is highly sensitive to*** *network configuration, and* ***our dataset did not include the*** *full network of stations* ***operated*** *by EPIC and ONC-EW. However****, the offline tests show that*** *EPIC and OCN-EW* ***are mostly able*** *to resolve* ***simulation*** *epicenters within several kilometers****, even with the limited network configuration*** *(Figure 1****0****).* ***Both EEW systems*** *use a direct grid search algorithm,* ***and*** *ONC-EW additionally uses a linear least squares algorithm for epicenter location (Chung et al., 2019; Schlesinger et al., 2021)."*

L 839 successful at detecting synthetic seismograms from large…

We revised the text as follows:

"Overall, EPIC and ONC-EW ***perform well at characterizing large synthetic events considering the ruptures are offshore and are more complex than crustal events in California."***

Recommendation: Resubmit for Review

Reviewer D:

This manuscript considers the use of synthetically-generated data to complement the lack of strong-motion data to calibrate/test earthquake early warning systems in the Cascadia region, and describes offline tests of the two of those systems, EPIC and ONC-EW. The objective is to create a diverse data set of large earthquake recordings that is sufficiently realistic to act as a surrogate for real-world data. The study is quite extensive and the manuscript is certainly on the longer side, but the authors do a good job explaining the various components and results in a clear manner, which allowed me to read through the manuscript with relative ease. While I have some comments on the interpretation of the data and some of the choices that the authors have made, I do think that the study is relevant and that the manuscript is in good shape.

Since there was plenty of material to read, I also have plenty of comments, but none of these would affect the results or conclusions of the study. I hope that the authors will consider them for the next iteration of the manuscript.

Kind regards,

Martijn van den Ende

General comments:

1) Point clouds tend to draw the eye towards outliers, and obscure the highest point densities by occlusion (stacking 10 points would look the same as stacking 1000). Given how often the authors discuss trends, variance, and mean values, it would make a lot of sense to replace the point cloud representations with 2D kernel density plots (a simple example would be Python Matplotlib's "hexbin" representation). Kernel density plots are a much more faithful representation of the distribution of the data in 2D, and allow the reader to assess the local mean and variance of the distribution without being influenced by outliers and occlusion. The downside is that the colour of such plots would represent the point density, and so colours cannot be used to represent e.g. hypocentral distance or magnitude. I don't know if the authors already considered this type of visualisation, but I can imagine that it could improve the data representation in Figs. 3-7. This is especially true for Figs. 5, 6, and 7, where the colours do not encode information other than the data source (which seem to overlap for the most part).

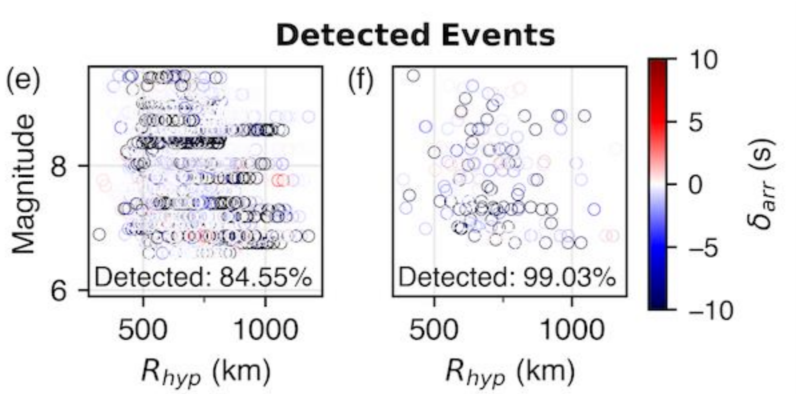
We thank the reviewer for their suggestion, and we agree that density plots would improve the data visualization. We updated Figures 5–7 to be hexbin plots rather than scatter plots. To support this, we chose to only show one GMM per plot as the inclusion of residuals for two different models did not add significant value.

We also slightly modified Figure 4. We made the decision to keep the scatter markers in this figure, but we did modify marker formats to help distinguish between symbols that overlap. Although, the changes did not make a significant difference.

Related to this point, I think that the clarity of Fig. 3 could be improved a bit. Panels (c) and (d) show the detected and missed events with different symbols, but the missed events overlay the detected events, and the detected events overlay themselves, to an extent that a large part of the panels reduces to a purple blob. I would experiment with separating the detected and missed events, so that at least these two classes can be correctly identified. For panels (e) and (f) I find the representation somewhat unintuitive. Currently, the colours carried by the transparent symbols blend into greenish blobs, and I wonder if the appearance is affected by the order in which the symbols are plotted (is the insertion ordered by magnitude or is it random?). I personally would have plotted magnitude versus distance, and then colour the symbols according to their misfit. In such a representation, the colours should start to trace out iso-intensity contours, with a cluster of high misfit in one corner representing the S-picks. Again, I don't know if the authors already considered this and opted for the current representation, and it may not improve the picture at all, but to me it would seem worth the effort.

We appreciate the reviewer's feedback on Figure 3. We modified panels (c) and (d) by removing the distance-dependent color scheme. All missed events are now red throughout the figure and detected events are blue. We attempted to modify panels (e) and (f) based on the reviewer's suggestion (see Figure R1 below). However, we do not find the new version to be informative. The velocity moveout with distance (i.e., S-wave picks) is no longer visible, and it is difficult to interpret the colors due to the large negative bias in the residuals. Most colors would be near white or black if we extended the colorbar to capture all residual values. We have decided to keep the original format but simplified the colors to improve readability. The magnitude-dependent color scheme in the bottom two panels did not add much value as there does not appear to be a significant trend with magnitude, so all markers are now blue (the color used for 'detected' throughout the new version of the figure.

Figure R1. Hypocentral distance (R_{hyp}) versus earthquake magnitude for events detected by the



STA/LTA algorithm using (left) EPIC and (right) ONC-EW parameters. Markers are colored by the true–detected arrival time residual (δ_{arr}).

2) A large part (if not the majority) of this study revolves around ground motions. I find it a little unsatisfactory to empirically calibrate the simulated attenuation against a different but likewise empirically calibrated ground motion model (GMM), which itself is calibrated against real-world recordings. Following this calibration, the calibrated simulations are then extensively compared with the calibrated GMMs, which may or may not be correctly representing the real data.

The reviewer brings up a valid argument. Generally, validating simulations against an empirical model is acceptable if the model was developed using data solely from the region of interest because it reflects the regional ground-motion trends. If the empirical GMM was developed using a multi-regional dataset, as are most ergodic GMMs, the median model may not adequately reflect ground-motion trends within a particular region (i.e., Cascadia). We updated the text as follows:

"Ideal circumstances would allow for direct comparison of simulated recordings with observed events in the region of interest to ensure unique regional conditions are captured. However, limitations imposed by the dearth of interface events in the CSZ require comparison of the simulated ground motions with estimates from empirical models. ***Empirical models are still a suitable validation benchmark, especially if the model was developed using solely data from the study region as they reflect regional ground-motion trends.*** *NGA-Sub was derived using global interface and intraslab recordings, including intraslab recordings from Cascadia."*

Additionally, we retain the following paragraph in the *Ground-motion summary* section:

"Considering the infrequency of large subduction zone earthquakes, ***GMMs*** *for such events are regressed on a sparse* ***global dataset, and*** *estimates from such models have both high epistemic uncertainty (poor resolution due to small number of events) and high aleatory uncertainty (regional variations not accounted for in the models). We recognize that ensuring our simulated data are reasonably modeled by such a GMM may result in poor representation of individual study regions. For example, Atkinson and Boore (2003) find that ground motions recorded at geologically similar sites from similar events in Cascadia and Japan exhibit amplitude differences of more than a factor of two. Nevertheless, we believe these generalized ground motions are sufficient for EEW testing purposes, and they can be further improved upon as more large earthquakes are recorded in the future."*

I am not at all a ground motions expert, and so I suppose that this is the best way to do things, but I can imagine that a reader (myself included) would feel a little cheated when the authors make claims regarding the accuracy of their approach. To their credit, the authors do acknowledge this in lines 747-756, and I fully understand that there is a severe lack of strong-motion data to directly compare against. I also cannot provide concrete suggestions to improve upon this approach, and so the authors should just largely ignore this comment. Perhaps one thing that can alleviate my unease is a brief discussion of the (quantified) uncertainties of the BCHydro / NGA-Sub GMMs, which should provide an indication of the inaccuracies that can be expected in real-world scenarios (basically expanding lines 747-756 with a quantitative justification).

We appreciate the reviewer's concerns, but unfortunately (as they already pointed out) there is no best solution to address this with the lack of observational data in the Cascadia region. In Figures 5–7, we plot the maximum standard deviation (i.e., aleatory uncertainty) for the NGA-Sub Cascadia region model. There is a distance-dependence to the model standard deviation, but the values are around 0.89 natural log units and 0.76 natural log units for PGA and PGV, respectively. The developers of NGA-Sub also calculated region-specific epistemic uncertainty (σ_ε, uncertainty due to incomplete dataset) for pseudo-spectral acceleration (PSA). We can approximate σ_ε for PGA as σ_ε for PSA at 0.01 s and σ_ε for PGV as σ_ε for PSA at 1.0s. For Cascadia, σ_ε would then be 0.35 and 0.24 ln units for PGA and PGV, respectively. As we already plot the GMM standard deviations in the ground-motion intensity figures, we do not believe that adding these numbers to the main text would strengthen the manuscript.

Aside from this philosophical point, I think it would improve the clarity of the manuscript if the authors could describe the attenuation relation used by FakeQuakes. The authors describe the modifications that are necessary to better fit the GMMs, but they do not describe how attenuation is treated by FakeQuakes (unless I missed it). I looked-up the Melgar et al. (2016) reference that is cited in line 197, but a quick scan did not lead me to a description of the attenuation relation that is supposedly modified for this study (again, I might have missed it). I hope that the authors could provide a more self-contained description of this, so that an interested reader does not need to search the literature/GitHub codes to understand the foundation of this study.

We thank the reviewer for the comment and agree that the clarity of the methodology would be improved by adding a description of the simulation attenuation model. We revised the section discussing changes to model attenuation as follows:

*"Lastly, **we add** an elastic attenuation term (Q_C) with the form $Q_{C,base} f^n$ (Farahbod et al., 2016; Pujades et al., 1996) **for modeling Q_S attenuation. Although this term allows for frequency dependence,** we **find $Q_c = 150 f^0$ (i.e., $Q_C = 150$) best reduces the attenuation bias. The path attenuation term (G_{ij} in Eq. 1; see Text S1 or Graves and Pitarka 2010) for more information) is computed using the effective S-wave attenuation ($Q_{S,eff}$)** (Pujades et al., 1997):*

$$Q_{S,eff} = \frac{1}{1/Q_S + 1/Q_c} \qquad \textbf{(R1)}$$

Anelastic attenuation values are reduced in the velocity model to account for the addition of the elastic attenuation term."

3) This study is quite extensive and has many components (slip model, lo/hi-freq synthetic waveforms, modified attenuation relationship, conversion to PG[A/V/D] and MMI, STA/LTA detection, comparison with BCHydro / NGA-Sub, comparison with GA21, offline tests of EPIC and ONC-EW), and the reader has many things to keep in the back of their minds as the manuscript progresses. Perhaps it would help readers with a short memory, like myself, to include a summary diagram or flowchart that mentions the various components and tests that this study features, and how they relate to each other.

We thank the reviewer for their comment, and we agree that the study includes many components that are difficult to keep track of. We have addressed this concern in two ways:

1. We focus on only one ground-motion model for each intensity measure. For PGA and PGV, we use NGA-Sub. For PGD, we use the PGD model from Goldberg and Melgar (2021) that was derived from observational data only. Figures 5–7 have been updated to reflect this.
2. We replaced Figure 2 with a flowchart summarizing the components of this study.

Specific comments:

- Lines 195-210: perhaps merging these two paragraphs and changing the order a bit could improve readability: "We curate a simulated dataset [...] using FakeQuakes [...]. Various methodologies exist [...]. Our objective is to assess [...]".

We thank the reviewer for their suggestion and agree that these section would benefit from restructuring. We have combined the two paragraphs as follows:

*"We curate a simulated dataset for the CSZ consisting of 112 kinematic rupture**s** and associated waveforms using FakeQuakes,* ***an earthquake simulation module from the MudPy repository*** *(Melgar et al., 2016**; Melgar et al., 2025**). Various simulation methodologies exist, ranging from simple stochastic models of high-frequency ground motions (e.g., Boore, 2003) to complex deterministic models involving numerical simulations of full waveform propagation through a 3D velocity structure (e.g., Komatitsch and Tromp, 1999; Faccioli et al., 1997; Mazzieri et al., 2013). FakeQuakes is a 1D semi-stochastic broadband simulation model, which provides more realistic ground motions compared to a simple stochastic model, yet is efficient enough to support large data generation. FakeQuakes is not the only existing 1D semi-stochastic model; however, it is preferred for this study, as opposed to other similar models (e.g., SCEC Broadband platform, Maechling et al., 2015), as it allows for P-wave generation (necessary for EEW) and has previously been tested in the Cascadia region (Melgar et al., 2016; Goldberg and Melgar, 2020; Small and Melgar, 2023)**. Our objective is to assess how well these waveforms perform at replicating reality with respect to key metrics needed for an EEW system, and subsequently test the performance of EEW systems using several example scenarios from the simulated dataset.**"*

- Section 3.2.2: the first paragraph explains why a deterministic approach wouldn't be feasible, but it doesn't describe the stochastic approach in the same level of detail as the low-frequency synthetics (Section 3.2.1). The second paragraph only describes parameter choices and justification. A few words describing the approach of generating the high-frequency synthetics would be appropriate.

We thank the reviewer for their comment, and we have updated this section to elaborate on the methodology. The discussion of user-defined parameters has been moved to the supplement (Text S1).

*"A **semi**-stochastic approach is used **to model broadband ground motions due to velocity model limitations and to reduce computational cost at higher frequencies. Small-scale CSZ velocity features are not well-resolved with limited offshore instrumentation. The low frequencies (< 1 Hz) are generated using the deterministic method described in the previous section, and the** high frequencies (> 1 Hz) **are generated by summing the spectral response for each subfault assuming a random (i.e., stochastic) phase** (Graves and Pitarka, 2010, 2015).*

$$A_i(f) = \sum_{j=1,M} C_{ij}\, S_i(f) G_{ij}(f) P(f) \qquad \text{(R2)}$$

***The spectrum resulting from slip on a single subfault (A_i) is a function of the radiation scale factor C_{ij}, source specrum S_i, path attenuation G_{ij}, and high-frequency decay P for subfault i, ray type j, and frequency f.**"*

- Line 301, "unusual responses": what kind of "responses" are we talking about? Are these simply unidentified noises? If so, including them would make sense, as they would occur in a real-world EEW scenario.

By "unusual responses", we were referring to significant noise spikes above the baseline level. As we mentioned in one of our responses to the other reviewer, we chose not to include noise segments with spurious signals because we did not want those artifacts to get mapped into our validation analyses. Our study focused on validating the signal itself. In the accompanying dataset (Ocean Networks Canada Society & Pacific Northwest Seismic Network, 2026), we include the raw simulated waveforms and the noise-combined waveforms. Users implementing our dataset are welcome to use the raw simulations and add complex noise for their analyses. The decision to include noise with other signals would affect station triggering but would likely not impact EEW performance as multiple stations would need to trigger for an event to be created. It's true that overlapping events are challenging for EEW, but to test this, we would need to create multi-event simulations. It would not be a realistic test case if only one station had noise contaminated by another earthquake.

We revised this sentence as follows:

"The noise streams are screened to remove potential earthquake signals or ***significant noise spikes****."*

- Lines 332-333: I presume that the STA/LTA parameters (0.05/5 and 1/5) are the window lengths in units of seconds?

Yes, the window lengths are in seconds. We revised the sentence as follows:

*"**EPIC** uses STA**(s)**/LTA**(s)**/threshold parameters of 0.005/5/20 on velocity time-series data."*

- Lines 349-351: I think that many readers could wonder about the false-positive rate, which is just as important as the detection rate. Did the authors check this (by testing on just noise), or is there a general reference on the false-positive rate of EPIC/ONC-EW? This would be a bit outside of the scope (as it pertains to EPIC/ONC-EW specifically), but if the authors have a ready answer it would be good to mention it somewhere.

False-positive detections are not uncommon at individual stations. In Lines 373–378 of the original manuscript, we stated

"The STA/LTA parameters are designed to be highly sensitive to ensure high detection success, which can lead to the algorithm triggering on noise spikes, even during seismically-quiet periods. In a real-time framework, the EEW workflow would not progress until a certain number of stations triggered within a reasonable geographical range. A minimum of four station triggers are required to send an alert for both ShakeAlert and ONC-EW (Chung et al., 2019; Schlesinger et al., 2021)."

We appreciate the reviewer's suggestion, but we do not find it informative to test the false-positive rate of the simulations, as the four-station-minimum criteria acts as a check point to prevent false alerts from being issued. This does not work 100% of the time, as was demonstrated by the false alert issued for a M5.9 event near Reno, Nevada on Dec. 5th 2025. Although it would be interesting to explore this further, this would be better suited for a paper assessing the performance of EEW, rather than our simulation validation-focused study.

- Line 355, “relatively no bias”: relative to what?

We clarified the sentence as follows:

“We ***do not observe a magnitude bias for missed detections****”*

- Line 360: “recordings that detected events” = “recordings on which detections were triggered”

We revised the sentence as follows:

“We next analyze the ***P-wave arrival detection latencies (δ_{arr}) for recordings that triggered the STA/LTA picker, as increased latencies bias location estimates and alert times.”***

- Lines 379-381: the authors state here that the detector has larger residuals for moderate magnitude events, which they attribute to the lower SNR. I’m not sure I agree with this argument, since the waveform onset should be independent of the eventual magnitude of the event. At the time scale of a trigger (the STA of < 1 sec), the earthquake has not yet reached its final magnitude, and so all earthquakes should look the same in the eyes of a causal detector.

We thank the reviewer for their comment. They are correct that event detection accuracy should be magnitude independent *if no other factors are considered*. We should have been more specific in the text and stated that at a given hypocentral distance (in particular at distances > 500 km), the largest residuals are observed for smaller events. We have since updated this figure to remove the magnitude-dependent color scheme as it did not add significant value. We also removed the sentence the reviewer was referring to.

- Line 384: “themsevles” = “themselves”

We made this correction.

- Lines 409-412: I actually have a hard time imagining a different way to combine signal and noise, other than what the authors adopted for this study. How did the EPIC operators do this summation? Did they pad the noise with zeros?

They did not add any “noise” to the signal, but rather padded it with zeros. The STA/LTA algorithm would trigger at the zero-pad/origin time boundary, which would not allow enough time to trigger again on the P-arrival. This paragraph has been removed from the manuscript.

- Lines 419-421: I think it would be fair to also include a reference to Men-Andrin Meier’s work (the 2021 GJI paper specifically), as the debate regarding earthquake determinism is far from settled (which this paragraph seems to allude to).

We added the reference to Meier et al. (2021) following the reviewer’s suggestion.

- Lines 451-452: this S-wave truncation is easily done with synthetics, but not in the real world. To allow the effect of SNR to propagate into the result, would it make sense to adopt a real-world methodology to isolate the P-wave? (fixed time window after first trigger, or theoretical S-wave arrival based on estimated hypocentre, or something else?)

We thank the reviewer for their suggestion. Although it would be possible to implement a "real-world" methodology for S-wave truncation, we did not want to map uncertainty in S-wave arrival times into our analyses, as our focus was on validation of the P-waves. In the accompanying dataset (Ocean Networks Canada Society & Pacific Northwest Seismic Network, 2026), we include the raw, unprocessed simulated waveforms. Users of our dataset are welcome to use the raw simulations and apply a more realistic S-wave truncation methodology.

- Lines 461-463: I don't quite understand the "1.5 log unit" selection criterion. The word "extending" suggests that previously the dataset was restricted to some arbitrary misfit value, and I don't see why a misfit selection criterion should be used (instead of a distance criterion).

We recognize this section was not written clearly. "1.5 log units" was not a selection criterion. What we intended to convey was that using our full dataset (i.e., records out to 1200 km), resulted in event means biased lower than the Golder and Melgar (2020) model. If we impose a distance limit like Trugman et al. (2019), our simulated Pd values closely match the model. We made significant revisions to this section and the accompanying Figure 4. We now only plot simulated event means that were calculated using records within 200 km epicentral distance. We include individual recordings for all distances, but they are colored by epicentral distance.

- Lines 463-465: the apparent larger variance could simply be a visualisation artefact (as pointed out above); a non-Gaussian distribution with wide tails would appear similar to a Gaussian distribution with large variance. Histograms or kernel density estimates should clarify this.

We thank the reviewer for their comment. In Figure R3 below, we plot the distribution of $\log10(Pd)_{10km}$ (demeaned) for observed and simulated recordings. For a couple time windows (e.g., 1 s and 20 s), there is a slight skew to the data. Overall, the data are normally distributed, thus we find it appropriate to say the variance is larger for simulated Pd compared to observed.

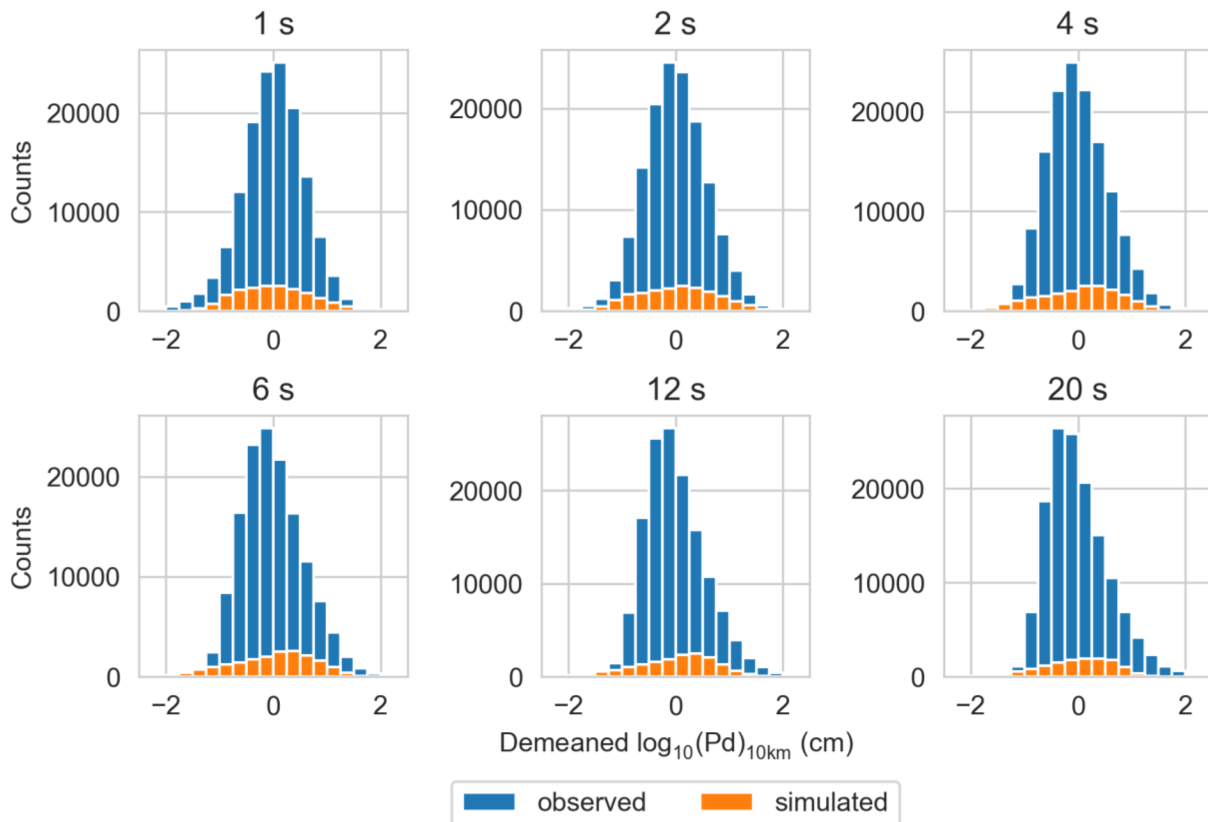


Figure R3. Histograms of p-wave displacements (Pd) corrected to 10km hypocentral distance (R_{hyp}) for time windows of 1, 2, 4, 6, 12, and 20 s. Blue: observed recordings; orange: simulated recordings. Simulated data include recordings out to 1200 km R_{hyp}.

- Lines 471-473: it would probably make sense to describe this (also) in section 3.2.2 (line 275).

We thank the reviewer for their suggestion, and we agree that a more detailed description of the matched filter process is warranted in section 3.2.2. We revised the two sections as follows:

Last paragraph of Section 3.2.2: *"The low- and high-frequency components are combined to form a full spectrum broadband waveform (generated for all seismic stations) by first double differentiating the low-frequency displacements to acceleration, after which the two components undergo a matched 4th order causal Butterworth filter.* ***Matched filters apply a lowpass filter to the double-integrated low-frequency displacement waveforms and a highpass filter to the high-frequency acceleration waveforms prior to summation. Following Nye et al. (2024b), we implement a two-frequency matched filter, with a 1Hz lowpass filter and a 0.1 Hz highpass filter, to reduce an artificial dip in the broadband spectra where the low- and high-frequency components are combined (Nye et al., 2024a,b).****"*

Last paragraph of Section 5.1: *"A key deviation in our methodology* ***is*** *the use of a tw-frequency matched filter for broadband simulation.* ***Goldberg and Melgar (2020) used a single fc for both the***

***lowpass and highpass filters. The implementation** of a two-frequency approach increase**s** energy in the simulated recordings **by reducing an artificial dip in the spectra (Nye et al., 2024a)**, likely **contributing** to the decreased magnitude saturation."*

- Line 558: did I understand correctly that Qc = 150 f^0 means that Qc simply assumes a constant value of 150? It would be good to explicitly mention this to avoid confusion.

The reviewer is correct that Qc simplifies to 150 in our model. We modified the corresponding sentences as follows:

*"Lastly, **we add** an elastic attenuation term (Q_C) with the form $Q_{C,base} f^n$ (Farahbod et al., 2016; Pujades et al., 1996**) for modeling Q_s attenuation. Although this term allows for frequency dependence**, we **find $Q_c = 150f^0$ (i.e., $Q_C = 150$) best reduces the attenuation bias.**"*

- Line 562, "the frequency factor of Qs was changed from 0.6 to 0.8": what is this frequency factor? I suppose that if the authors expand the description of the attenuation model, this question is automatically answered.

The discussion of this user-defined parameter has been moved to the supplement (Text S1), which includes a more detailed description of the attenuation model used in the simulations. Please see the text from the supplement below:

***"Attenuation Parameters:** The high-frequency amplitude spectra for individual subfaults (A_i) is modeled as a function of the radiation scale factor C_{ij}, source specrum S_i, path attenuation G_{ij}, and high-frequency decay P for subfault i, ray type j, and frequency f.*

$$A_i(f) = \sum_{j=1,M} C_{ij}\, S_i(f) G_{ij}(f) P(f) \qquad (S2)$$

The path attenuation term G_{ij} is modeled as an exponential decay dependent on the travel-time weighted Q values from the 1D model (Equation S3; Graves and Pitarka, 2010).

$$G_{ij}(f) = \frac{I_i(f)}{r_{ij}} exp\left[-\pi f^{1-x} \sum_{k=1,L} t_{ijk}/q_k\right] \qquad (S3)$$

Here, $I_i(f)$ represents gross impedance from the quarter-wavelength theory, r_{ij} is the total path length, x is the frequency-dependence of the attenuation, t_{ijk} is the travel-time, and q_k is the Q value for each subfault i, ray type j, frequency f, and velocity layer k. We use a frequency factor x equal to 0.8 opposed to the default value of 0.6. Atkinson and Macias (2009) determined the appropriate frequency factor for the Cascadia region to be nearly half that at 0.45, but they constrained this value using only four events and three stations. We find a value of 0.8 better reflects predicted ground-motion trends from the empirical Next Generation Attenuation Subduction ground-motion model (Parker et al., 2020)."

- Lines 577-578, "Recordings with the largest Rrup come from the smaller magnitude events": shouldn't this be the larger magnitude events (higher SNR)?

We understand why the reviewer would assume this. The reason is that R_{rup} is calculated in reference to the closest point on the fault. For the larger ruptures activating most of the slab, R_{rup} is short for most stations. The largest R_{rup} records correspond to stations located on the opposite end of the slab from where the smaller earthquake ruptured.

- Lines 579-580: are BCHydro / NGA-Sub not also 1D GMMs? They are calibrated against data from the 3D world, but in the end the spatial information is encoded by the hypocentral distance. In that case, discrepancies between the simulated data and the GMMs cannot be attributed to 3D structure.

We thank the reviewer for their comment, but we disagree with this statement. It is true that these GMMs are ergodic and rely on simple spatial parameters, but as the reviewer mentioned, they are developed using data which are controlled by the 3D subsurface. Therefore, the average effects of 3D structure are inherent in the model (e.g., the geometric spreading and attenuation parameters, as well as relevant reported path-related uncertainties). If one were to simulate ground motions using 1D and 3D velocity models, the simulation attenuation rates would likely differ (depending on how variable the 3D model is), which would influence the GMM residuals. Fadugba et al. (2024) demonstrated this by comparing simulations and GMM predictions for peak ground displacement.

- Lines 619-620: as for the attenuation model of FakeQuakes, I think it would help to briefly describe the scaling models of Worden et al. (2012) here (or in an appendix).

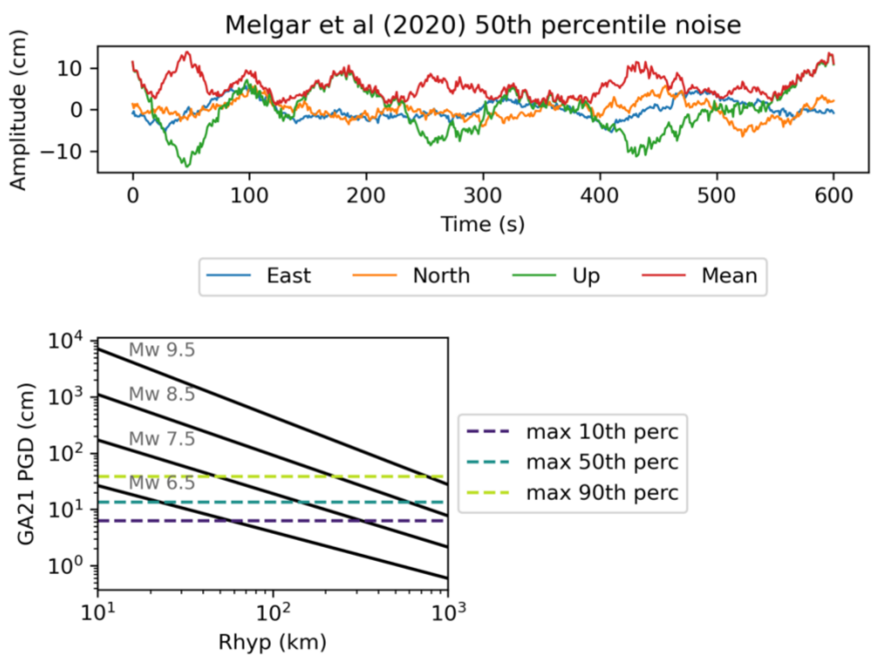
We added the following description of the Worden et al. (2012) model:

"MMI is a linear function of the logarithm of peak ground motion, logarithm of source-to-site distance, and earthquake magnitude."

- Lines 699-701: unless I'm mistaken, if the signal and noise are independent random distributions, then the superposition of signal and noise should have a higher expected PGD than the PGD of the pure signal. If the GMMs are calibrated against real-world data that include noise, then the GMM-predicted PGD should be similar to the simulated PGD + 100% noise, and larger than the simulated PGD + 0% noise. Fig. 7 reports the opposite trend (as a function of magnitude = signal amplitude). What this figure seems to suggest instead, is that the authors greatly overestimate the amplitude of the noise that is included in the calibration of the GMMs, and that a value of around 5% seems to be a much more suitable value (giving a flat trend as a function of magnitude).

The reviewer's understanding is generally correct in that PGD GMMs likely overestimate PGD as they are calibrated on data with real noise. The synthetic noise was added using the GNSS noise model from Melgar et al. (2020; cited in the paper). This model was developed using real GNSS observations spanning from California to Alaska (most of which encompasses our study area). The 50^{th} percentile model is their median model capturing the most likely noise behavior, so PGD + 50^{th} percentile noise should be most likely to match GMM estimates.

Records closest to the hypocenter are mostly unaffected by addition of noise because the unattenuated signal is larger than the noise floor. Records from the largest events (M ~8.75+) are also mostly unaffected by addition of noise because the signal is large, even at large hypocentral



distances. The record amplitudes that are most sensitive to the presence/absence of noise are those from lower magnitude events and/or recorded at larger distances.

Figure R2 below helps to illustrate this. In the top plot, we show 600 seconds of randomly generated noise using the Melgar et al. (2020) GNSS 50th percentile noise model. In the bottom plot, we show the Goldberg et al. (2021) PGD ground-motion model for Mw 6.5, 7.5, 8.5, and 9.5 along with the maximum amplitude of synthetic noise generated using different model percentiles. Even for moderately large events, there is a distance limit after which the noise biases the apparent PGD.

Figure R2. (top) Randomly generated noise using the Melgar et al. (2020) GNSS 50th percentile noise model. The mean is the Euclidean norm of the three components, the averaging method commonly used to obtain PGD from GNSS data. (bottom) Goldberg et al. (2021) PGD ground-motion model predictions for Mw 6.5, 7.5, 8.5, and 9.5. The dashed lines are the maximum amplitude of the three-component synthetic noise waveforms generated using different model percentiles.

- Line 733: “regardless of [absolute] noise levels” (to avoid confusion with SNR).

We removed the discussion of SNR and PGD and moved Figure 8 to the supplement (Figure S4).

- Lines 849-852: this is a valid argument for EPIC, but would it be possible to compare-ground motions predicted by ONC-EW?

ONC-EW does not estimate ground motions directly, but rather provides the earthquake source parameters (magnitude and location estimates), and the end-user calculates shaking intensity using their desired ground-motion models. We clarified this in the manuscript as follows:

“ONC-EW allows subscribers to run software code to estimate time left before shaking and MMI given their locations of concern (Schlesinger et al., 2021).”

- Lines 873-878: the main motivation for using FakeQuakes is that deterministic models are too expensive to run at the required resolution. It is therefore a bit surprising that the authors suggest improving the study with a 3D velocity structure and 3D Earth waveform modelling, which brings us back to the motivation for using a simple 1D model in the first place. I certainly agree with the authors that their suggestions would make for a great leap forward, but these are contingent on improved numerical infrastructures or algorithms.

This reviewer poses a valid argument. We added a statement mentioning that these could be avenues for improvement *IF* additional computational resources are available:

"there are avenues for further if additional computational resources are available. This could include use of a realistic 3D Earth structure (Fadugba et al., 2024), *addition of a coupling model (Small and Melgar, 2021),* ***increased*** *velocity resolution from the Cascadia2021 initiative (Ashraf et al., 20223), and* ***contributions*** *from upcoming Cascadia Region Earthquake Science Center (CRESCENT) efforts."*

- Data and code availability: I would strongly recommend to give version numbers of the software that was used (the authors already noted that changes in the FakeQuakes software could be the source of discrepancies with previous studies), and ideally to also copy the relevant codes to a long-term archiving repository like Zenodo or Figshare (Seismica has a "community" on Zenodo). It is a real possibility that Microsoft will impose restrictions on public Github repositories in the future, or change the URLs that will result in broken links. Zenodo/Figshare will stamp the repository with a DOI and guarantee its long-term accessibility. The DOI of the "ONC & PNSN (2024)" reference does not resolve.

We thank the reviewer for their thoughtful comment. We added a reference to the most recent version of MudPy, which is hosted on Zenodo (Melgar et al., 2025).

Recommendation: Revisions Required

References mentioned in review response but not in the manuscript

Böse, M., C. Felizardo, & T. H. Heaton (2015). Finite-Fault Rupture Detector (*FinDer*): Going Real-Time in Californian *ShakeAlert* Warning System. *Seismological Research Letters*, *86*(6), 1692–1704. doi: 10.1785/0220150154

References added to the manuscript

Aagaard, B. T., Hall, J. F., & Heaton, T. H. (2001). Characterization of near-source ground-motions with earthquake simulations. *Earthquake Spectra*, *17*(2), 177–207. doi: 10.1193/1.1586171

Allen, R. M. (2011). Earthquakes, early and strong-motion warning. In F. Wenzel & J. Zschau (Eds.), *Encyclopedia of earthquake engineering* (pp. 226–234). Springer.

Atwater, B. F. (1987). Evidence for great holocene earthquakes along the outer coast of Washington State. *Science (American Association for the Advancement of Science)*, *236*(4804), 942–944.

Barnes, C. R., and Tunnicliffe, V. (2008). Building the World's First Multi-Node Cabled Ocean Observatories (NEPTUNE Canada and VENUS, Canada): Science, Realities, Challenges and Opportunities. *In proceedings of OCEANS 2008 MTS/IEEE Kobe-Techno-Ocean*, Kobe, Japan, April 8–11, 2008. doi: 10.1109/OCEANSKOBE.2008.4531076

Boore, D. M. (2010). Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bulletin of the Seismological Society of America*, *100*(4), 1830–1835. doi: 10.1785/0120090400

Cooper, J. D. (1868). Earthquake indicator. *San Francisco Bulletin*, November 3, 1868.

Crane, S., Seywerd, H., Adams, J., Bird, A., Kolaj, M., & Perry, C. (2023). *National Earthquake Early Warning for Canada*. Paper presented at the Canadian Conference – Pacific Conference on Earthquake Engineering 2023, Vancouver, British Columbia, Canada.

Goldfinger, C., Nelson, C. H., Morey, A .E., Johnson, J. R., Patton, J., Karabanov, E., Gutierrez-Pastor, J., Eriksson, A. T., Gracia, E., Dunhill, G., Enkin, R. J., Dallimore, A., and Vallier, T. (2012). Turbidite event history—Methods and implications for Holocene paleoseismicity of the Cascadia subduction zone: U.S. Geological Survey Professional Paper 1661–F, available at http://pubs.usgs.gov/pp/pp1661/f

Graves, R. W., Aagaard, B. T., & Hudnut, K. W. (2011). The ShakeOut earthquake source and ground motion simulations. *Earthquake Spectra, 27*(2), 273–291. doi: 10.1193/1.3570677

Harris, R. A., Barall, M., Lockner, D. A., Moore, D. E., Ponce, D. A., Graymer, R. W., Funning, G., Morrow, C. A., Kyriakopoulos, C., & Eberhart-Philips, D. (2021). A geology and geodesybased model of dynamic earthquake rupture on the Rodgers Creek–Hayward–Calaveras fault system, California. *Journal of Geophysical Research: Solid Earth*, *126*, e2020JB02057. doi: 10.1029/2020JB020577

Kobayashi, M., Takemura, S., & Yoshimoto, K. (2015). Frequency and distance changes in the apparent P-wave radiation pattern: Effects of seismic wave scattering in the crust inferred from dense seismic observations and numerical simulations. *Geophysical Journal International*, *202*, 1895–1907. doi: 10.1093/gji/ggv263

Lux, A. I., Smith, D., Böse, M., McGuire, J. J., Saunders, J. K., Huynh, M., Stubailo, I., Andrews, J., Lotto, G., Crowell, B., Crane, S., Allen, R. M., Given, D., Hartog, R., Heaton, T., Husker, A., Marty, J., O'Driscoll, L., Tobin, H., McBride, S. K., & Toomey, D. (2024). Status and Performance of the ShakeAlert Earthquake Early Warning System: 2019–2023, *Bulletin of the Seismological Society of America*, *114*, 3041–3062. doi: 10.1785/0120230259

McGuire, J. J., Ulberg, C. W., Lux, A. I., Böse, M., Andrews, J. R., Smith, D. E., Crowell, B. W., Murray, J. R., Henson, I., Hartog, R., et al. (2025). ShakeAlert ® version 3: Expected performance in large earthquakes. *Bulletin of the Seismological Society of America*, 115(2), 553–561. doi: 10.1785/0120240189

McMartin, K. (2025). *Variations in earthquake detection across Ocean Networks Canada's early warning system* (Honours thesis). School of Earth and Ocean Sciences, University of Victoria.

Meier, M.-A., Ampuero, J.-P., Cochran, E. and Page, M. (2021). Apparent earthquake rupture predictability. *Geophysical Journal International*, *225*(1), 657–663. doi: 10.1093/gji/ggaa610

Melgar, D. (2025). *Consideration of rupture kinematics increases tsunami amplitudes in far-field hazards assessments*. EarthArXiv. doi: 10.31223/X5BQ61

Melgar, taranye96, Lin, T., Kong, Q., christineruhl, sheng6010, Adair M., Marfito, B., & LeVeque R. J. (2025). UO-Geophysics/MudPy: v1.4 (v1.4). Zenodo. doi: 10.5281/zenodo.15547301

Nye, T., Dybing, S. Melgar, D., & Sahakian, V. J. (2024a). The case of the missing frequencies: Reduction of an artificial spectral deficiency in semistochastic broadband simulation. Abstract presented at 2024 SSA Annual Meeting, 14–18 Apr.

Nye, T. A., Sahakian, V. J., & Melgar, D. (2024b). Validation of tsunami earthquake ground-motion simulations using near-field observational data: Rupture parameter constraints from the 2010 Mentawai event. *Seismica*, *3*(2), 1–19. doi: 10.26443/seismica.v3i2.1152

Poulos, A., Hirakawa, E., Parker, G., & Baltay, A. (2025). Orientation Dependence of Probabilistic Seismic Hazard Results From Physics-Based Simulations. Abstract presented at 2025 SSA Annual Meeting, 14–18 Apr.

Priest, G. R., Zhang, Y., Witter, R. C., Wang, K., Goldfinger, C., & Stimely, L. (2014). Tsunami impact to Washington and northern Oregon from segment ruptures on the southern Cascadia subduction zone. *Natural Hazards*, *72*, 849–870. doi: 10.1007/s11069-014-1041-7.

Rodgers, A. J., Pitarka, A., & Mccallen, D. B. (2019). The effect of fault geometry and minimum shear wavespeed on 3D ground-motion simulations for an Mw 6.5 Hayward Fault scenario earthquake, San Francisco Bay Area, Northern California. *Bulletin of the Seismological Society of America*, *109*(4), 1265–1281. doi: 10.1785/0120180290

Thakoor, K., Andrews, J., Hauksson, E., & Heaton, T. (2019). From earthquake source parameters to ground-motion warnings near you: The ShakeAlert earthquake information to ground-motion (eqInfo2GM) method, *Seismological Research Letters*, 90(3), 1243–1257. doi: 10.1785/0220180245

Tréhu, A. M., Braunmiller, J., Davis, E. (2015). Seismicity of the Central Cascadia Continental Margin near 44.5° N: A decadal view. *Seismological Research Letters, 86* (3), 819–829. doi: 10.1785/0220140207

Wang, K., and Tréhu, A. M. (2016). Invited review paper: Some outstanding issues in the study of great megathrust earthquakes—The Cascadia example. *Journal of Geodynamics*, 98, 1–18.