

Dear Reviewer,

We would like to sincerely thank you for the careful review of our manuscript “**Seismic and Tsunami Hazard Potential of the Negros–Sulu Megathrust, Philippines**” and for giving us the opportunity to submit a revised version. The comments have greatly improved the clarity and overall quality of the manuscript.

The comments have been adopted and carefully addressed, and the manuscript has been revised accordingly. Please find below the **responses (highlighted in yellow)** to the reviewers’ comments, and the **revised in-line texts and figure captions (highlighted in cyan)**. Line numbers in the responses and/or actions refer to revisions that appear in the untracked version of the revised manuscript:

General points

The manuscript "Seismic and Tsunami Hazard Potential of the Negros–Sulu Megathrust, Philippines" develops a suite of highly simplified tsunami scenarios to help us better understand tsunami hazards in the region. This is an important topic and of interest for readers of EGUSphere.

My main criticism of the manuscript is that the 18 modelled tsunami scenarios are very simplified, and it appears that they all produce waves that are smaller than some historically observed tsunamis produced by smaller magnitude earthquakes. This suggests that the simplicity of the source models is failing to characterise hazards in the region. This is an early study of the region, and I do not expect the models to be perfect. But currently the limitations are not clear enough - readers with an interest in the hazard may be misled. To strengthen the manuscript, I strongly encourage the authors to consider other scenarios with more compact high-slip rupture, which may better represent past events (and thus be more appropriate for hazard assessment). There should also be a clearer discussion of how the results compare with history, to help readers interpret the model uncertainty, which will guide both future work as well as appropriate use of results for hazard assessment.

Response/Action:

We agree that these models are simple as uniform slip has been used. For this work, we begin with simple approaches that will serve as baseline models to compare with future studies, e.g., stochastic approaches of slip variability, including those more compact high-slip ruptures. We believe that these simple models still provide valuable information for regional-scale identification of highly exposed coastal areas and are timely needed in this less-studied subduction zone. We agree that the limitations should be clearly stated, and in the revised manuscript, *Section 5.3 Limitations and Future Work* has been included so that readers are clearly informed of the limitations of these baseline models.

Beyond this core issue, there are other issues that will be easier to address:

-Many of the figures use colour schemes that make them hard to see clearly. Please use colour schemes with more variation, which better resolve small values. This includes Figs 5, 9, 10, 11 in the main text, and many figures in the supplementary.

Response/Action:

The color schemes in Figures 5, 9, 10, 11, S1–S5, and S8–S14 have been revised to have better visibility of the values as shown below:

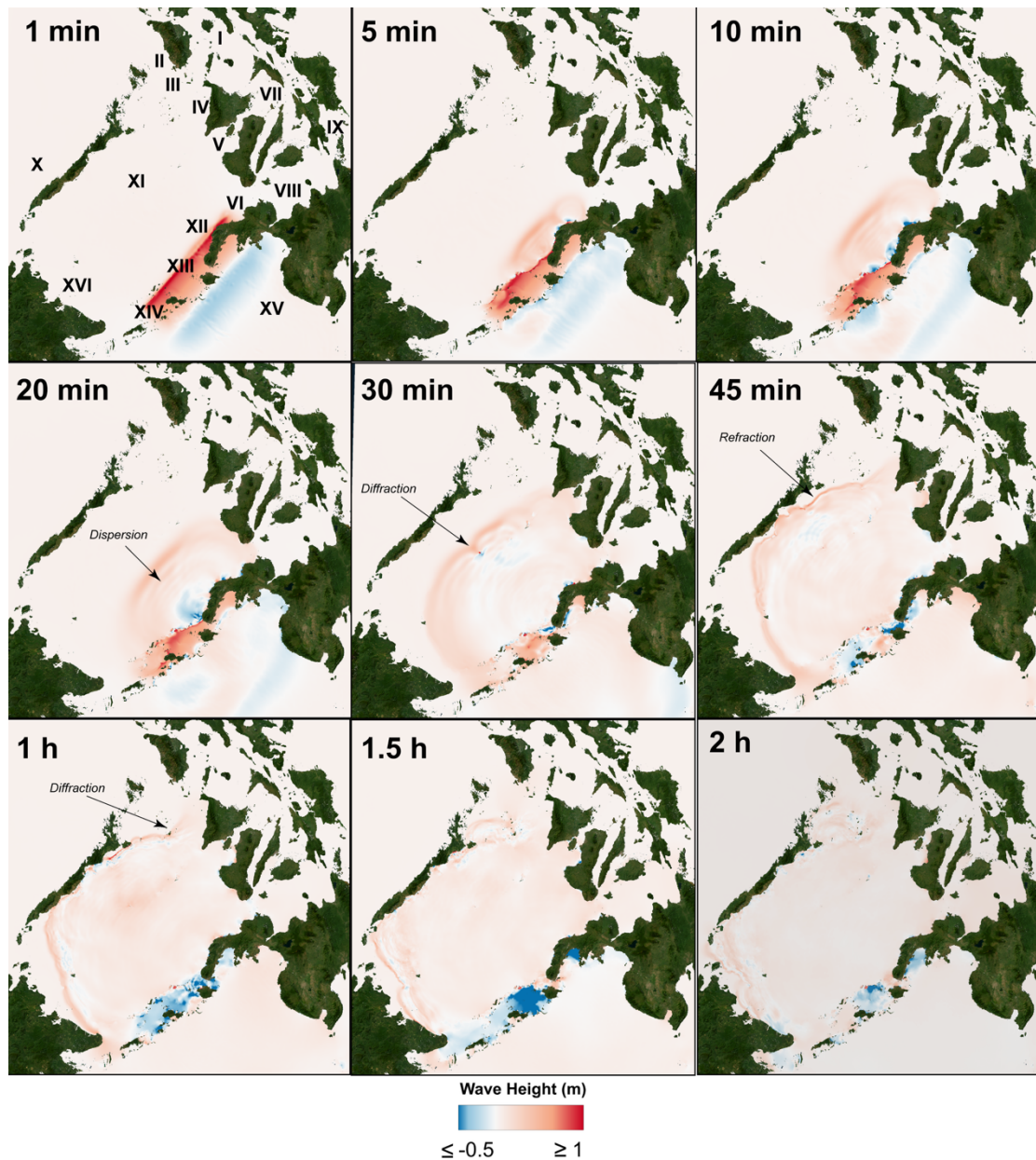


Figure 5. Tsunami wave (2 h) propagation from ST segment at 20° dip angle. Basemap from World Imagery.

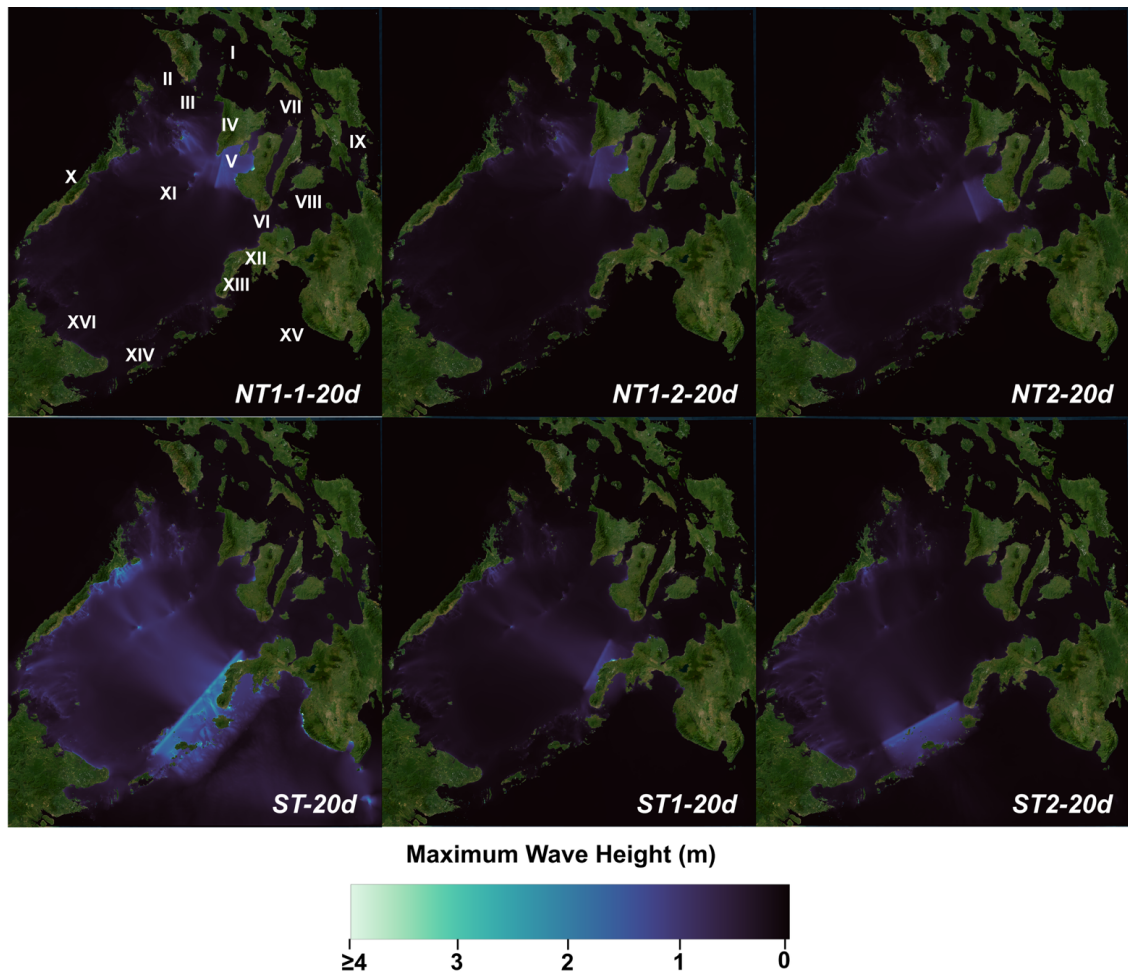


Figure 9. Maximum wave height (m) from the 4-h tsunami simulation of six segments with a megathrust dipping at 20°. Basemap from World Imagery.

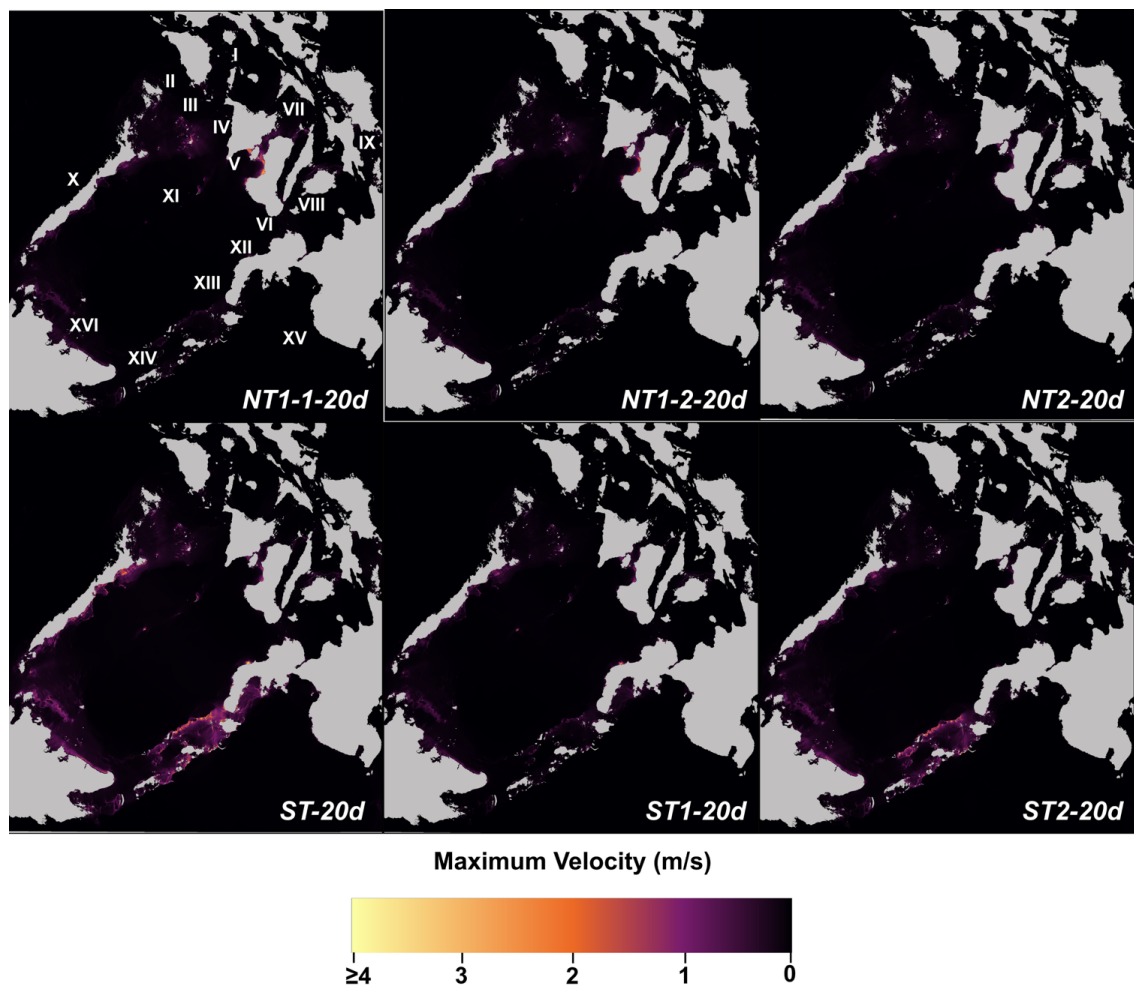


Figure 10. Maximum flow velocity (m/s) from the 4-h tsunami simulation of six segments with a megathrust dipping at 20°.

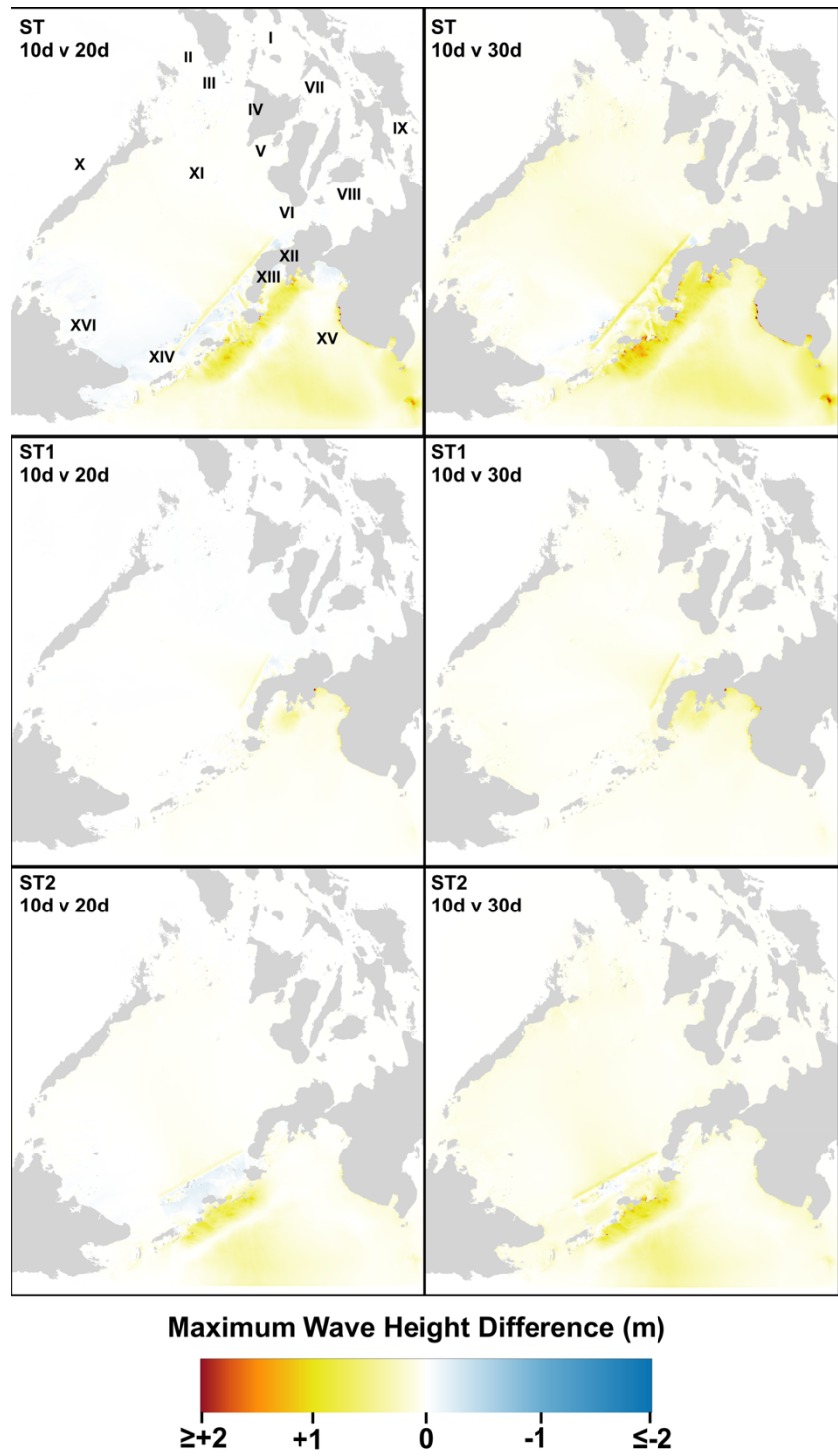


Figure 11. Maximum wave height difference from 10°-dip megathrust compared to that of 20° and 30° in the Sulu Trench segments.

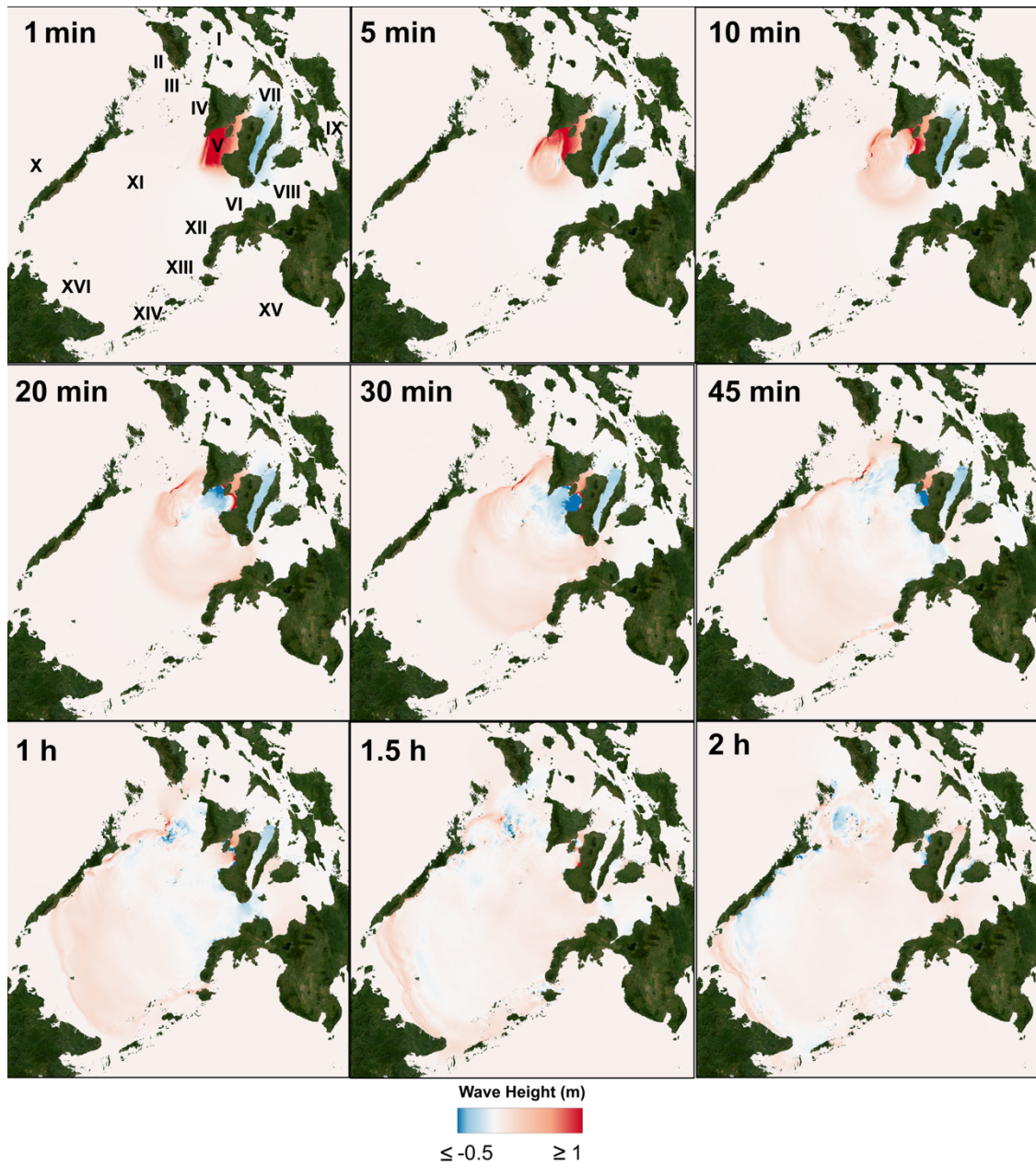


Figure S1. Tsunami wave (2 h) propagation from NT1-1 segment at 20° dip angle. Basemap from World Imagery.

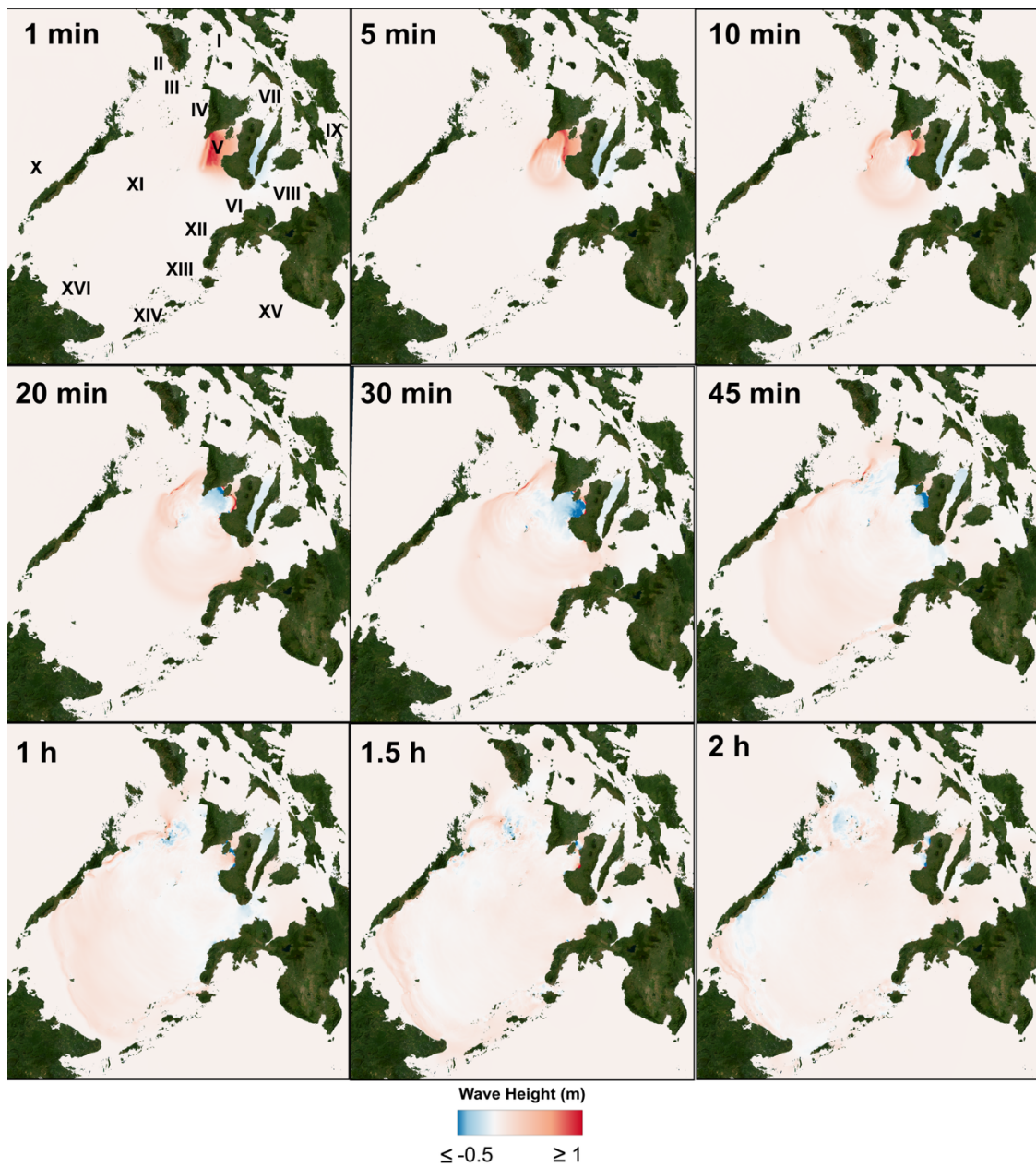


Figure S2. Tsunami wave (2 h) propagation from NT1-2 segment at 20° dip angle. Basemap from World Imagery.

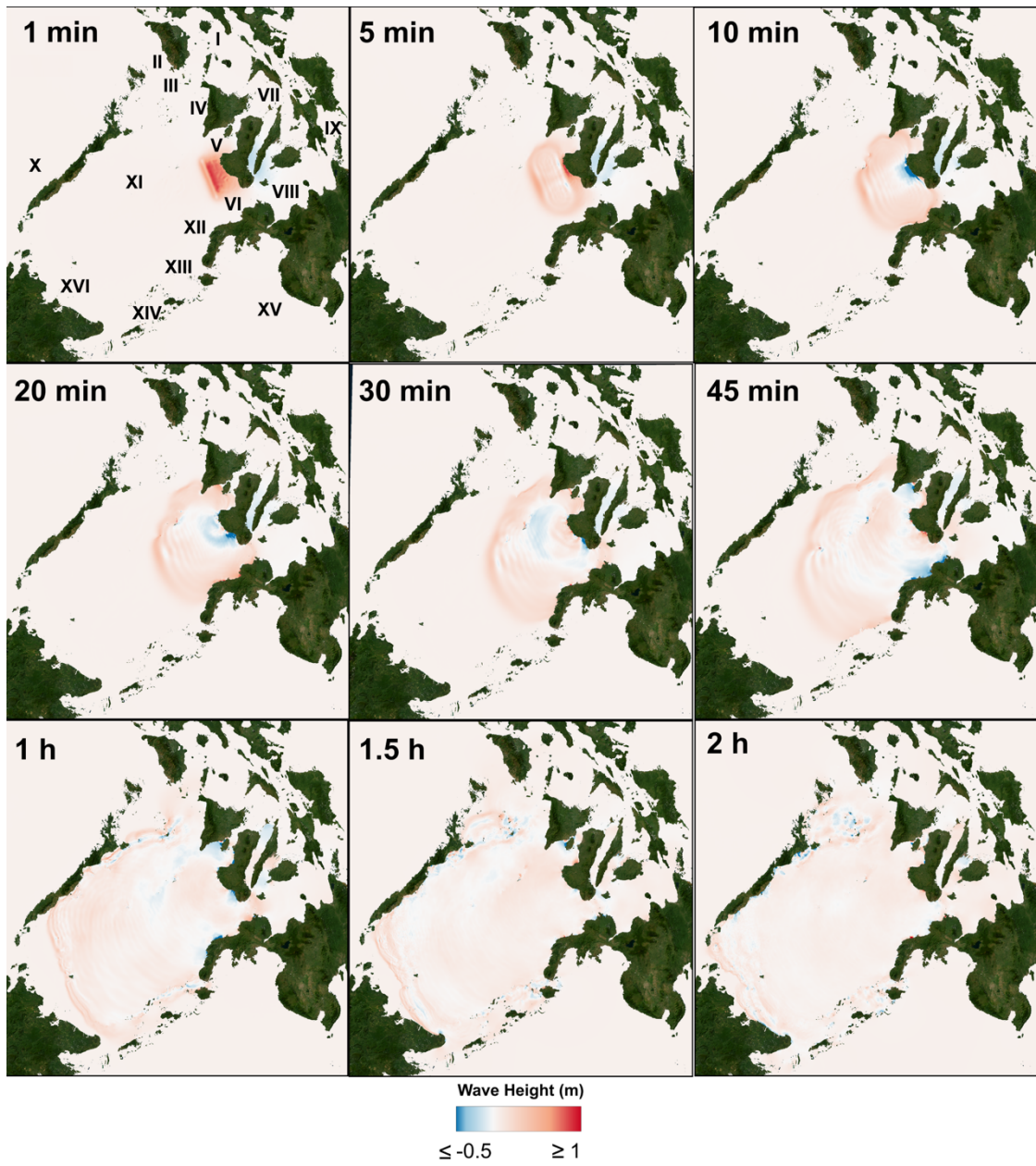


Figure S3. Tsunami wave (2 h) propagation from NT2 segment at 20° dip angle. Basemap from World Imagery.

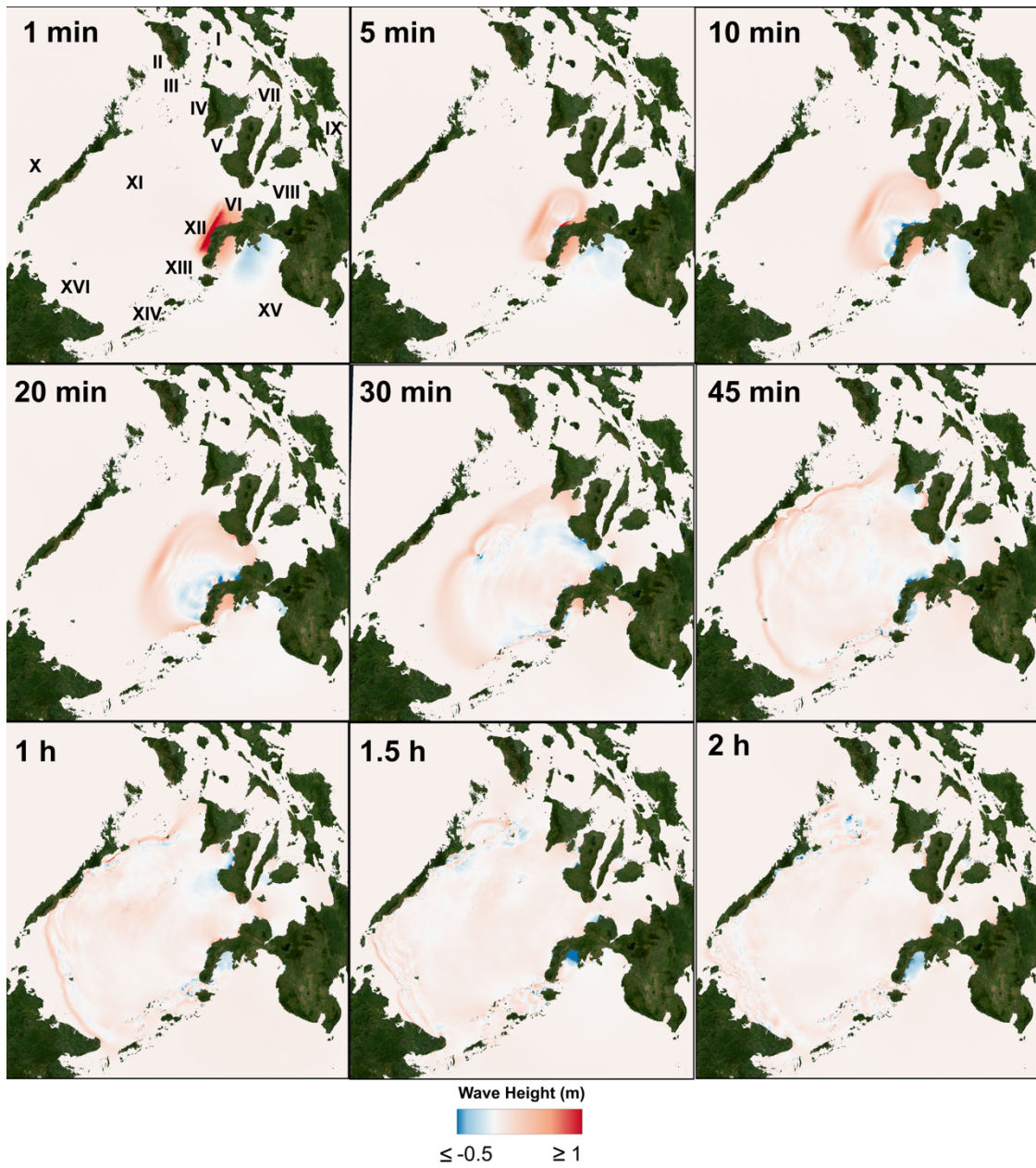


Figure S4. Tsunami wave (2 h) propagation from ST1 segment at 20° dip angle. Basemap from World Imagery.

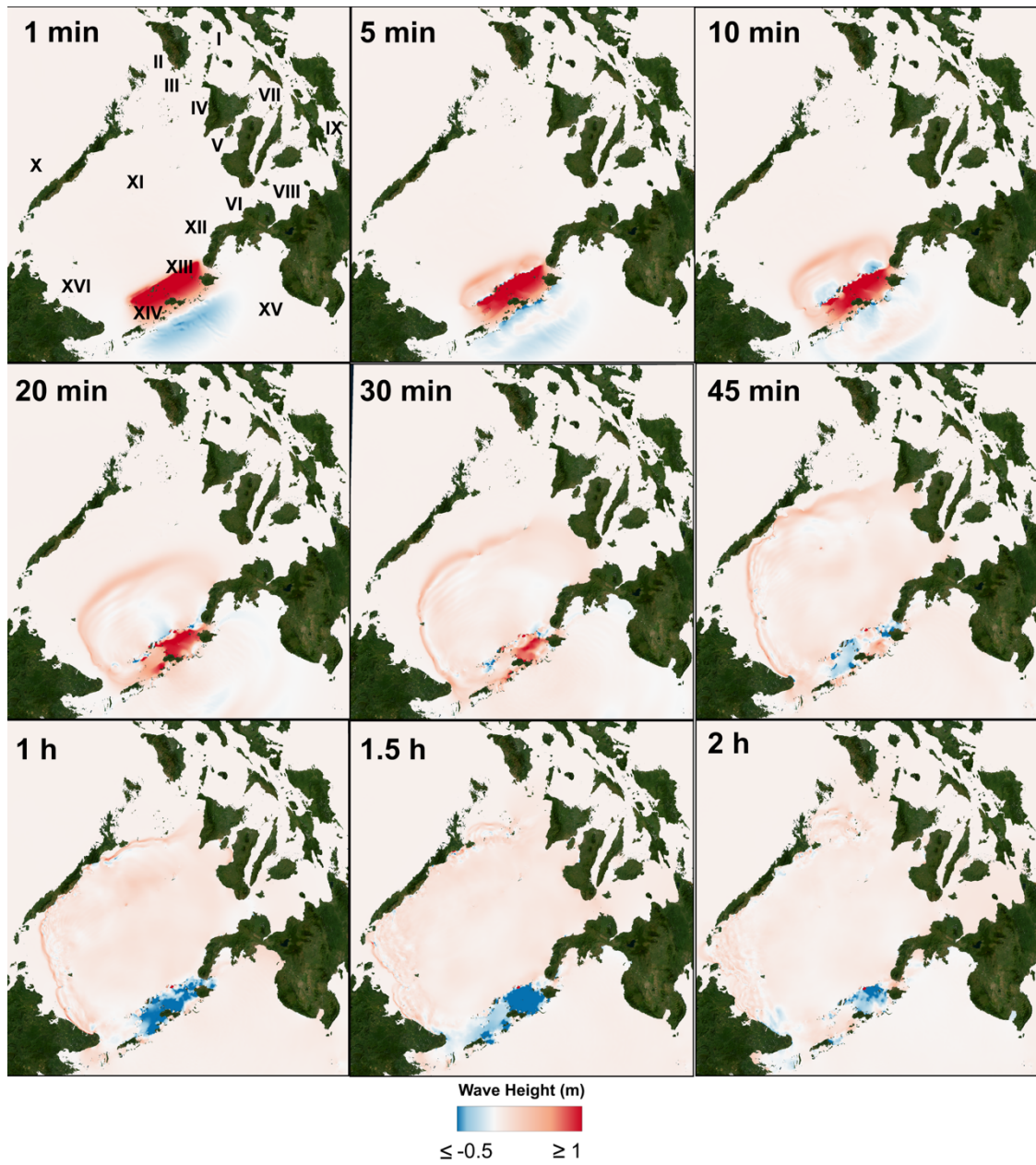


Figure S5. Tsunami wave (2 h) propagation from ST2 segment at 20° dip angle. Basemap from World Imagery.

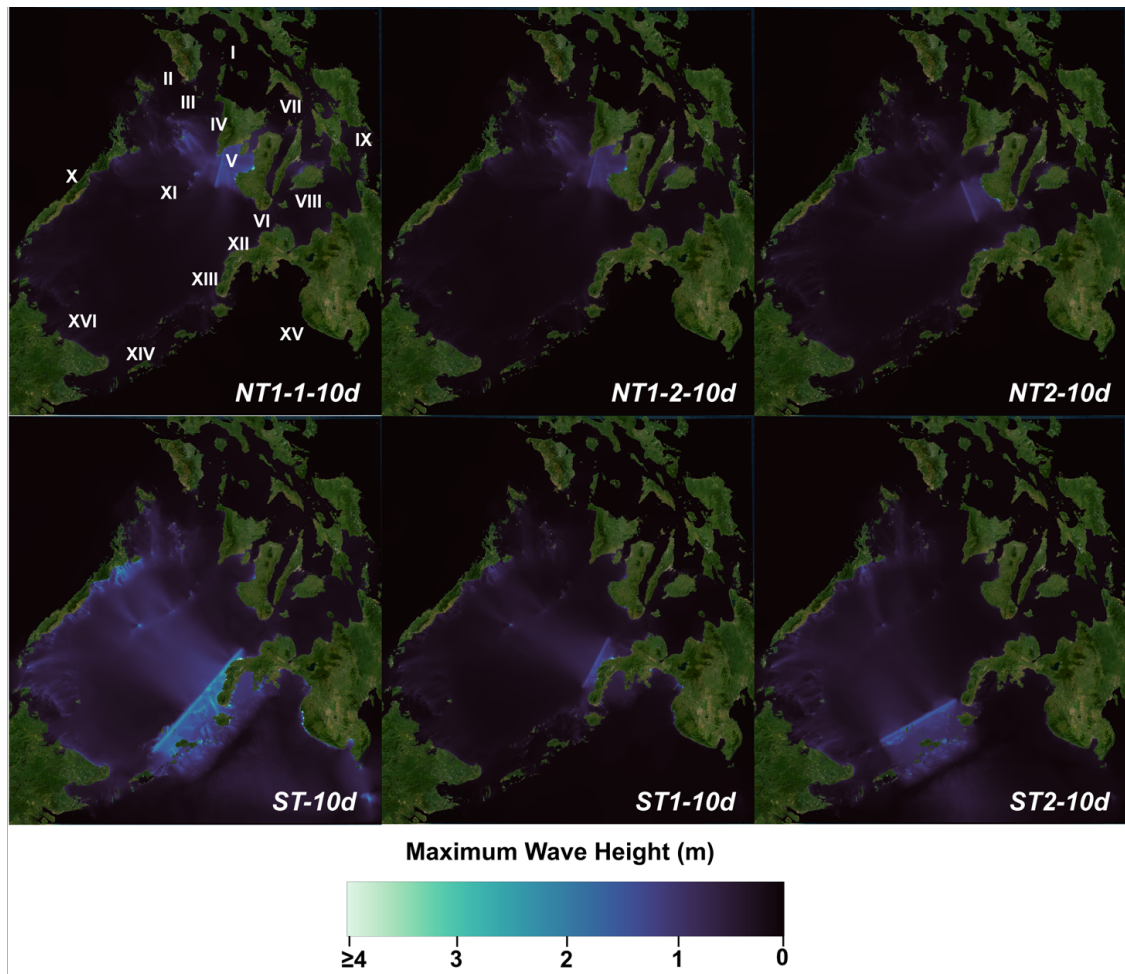


Figure S8. Maximum wave height (m) from the 4-h tsunami simulation of six segments with a megathrust dipping at 10°. Basemap from World Imagery.

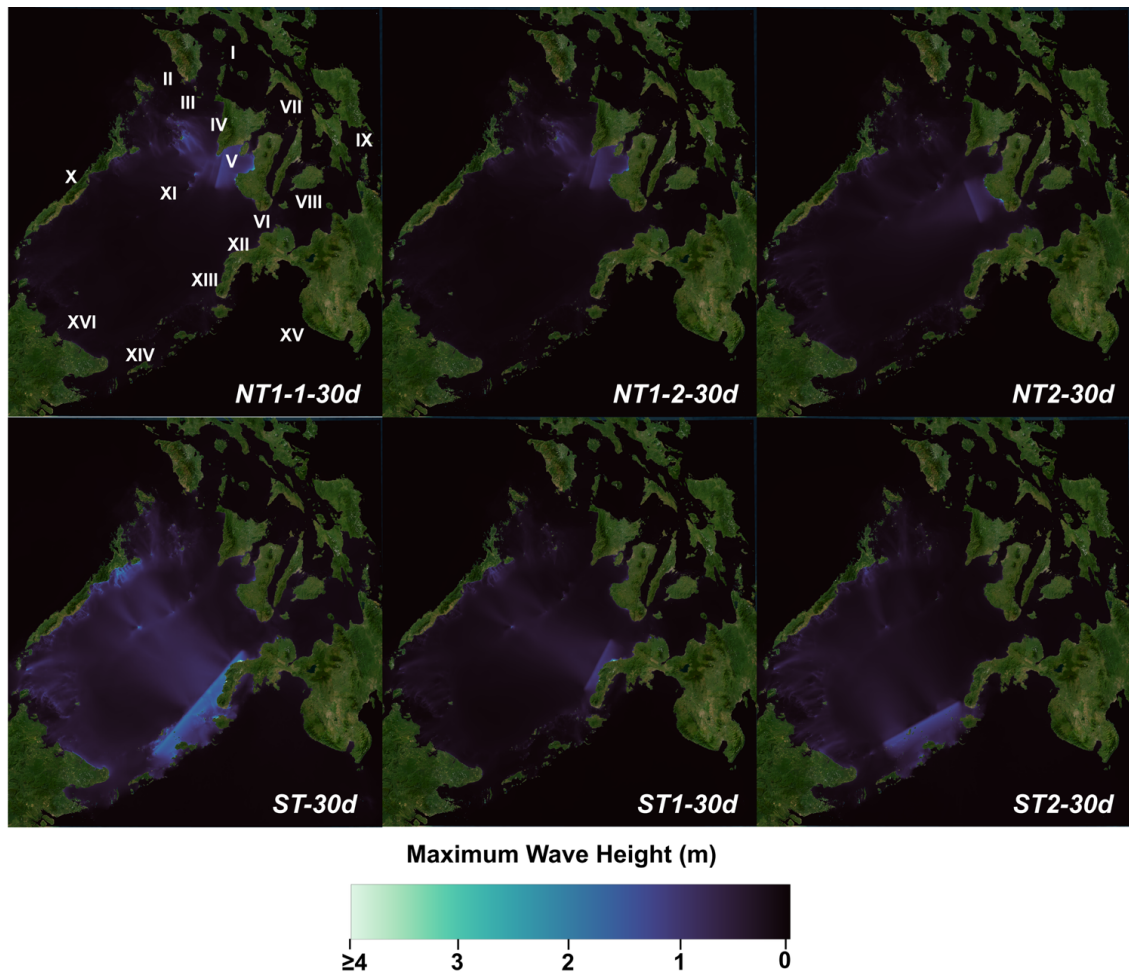


Figure S9. Maximum wave height (m) from the 4-h tsunami simulation of six segments with a megathrust dipping at 30°. Basemap from World Imagery.

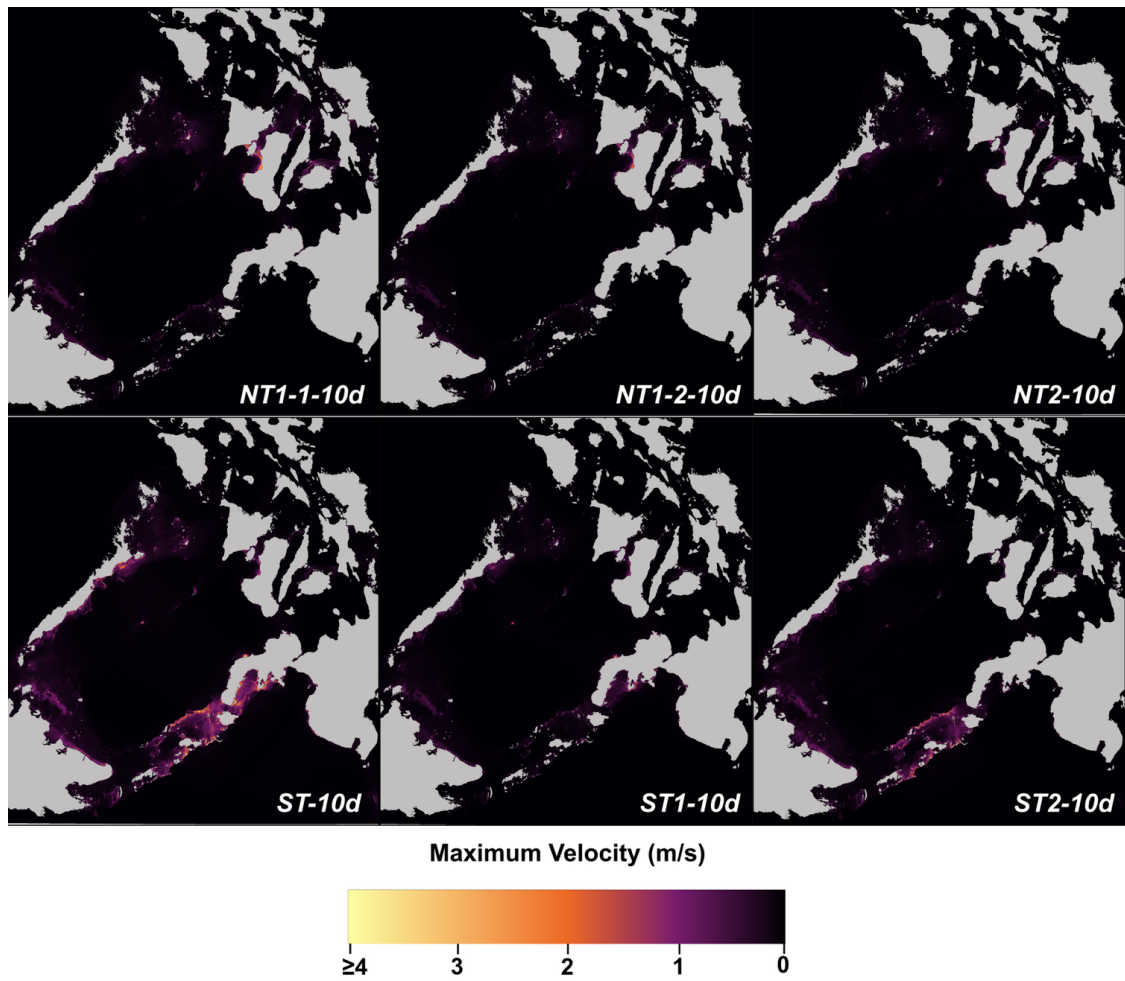


Figure S10. Maximum flow velocity (m/s) from the 4-h tsunami simulation of six segments with a megathrust dipping at 10°.

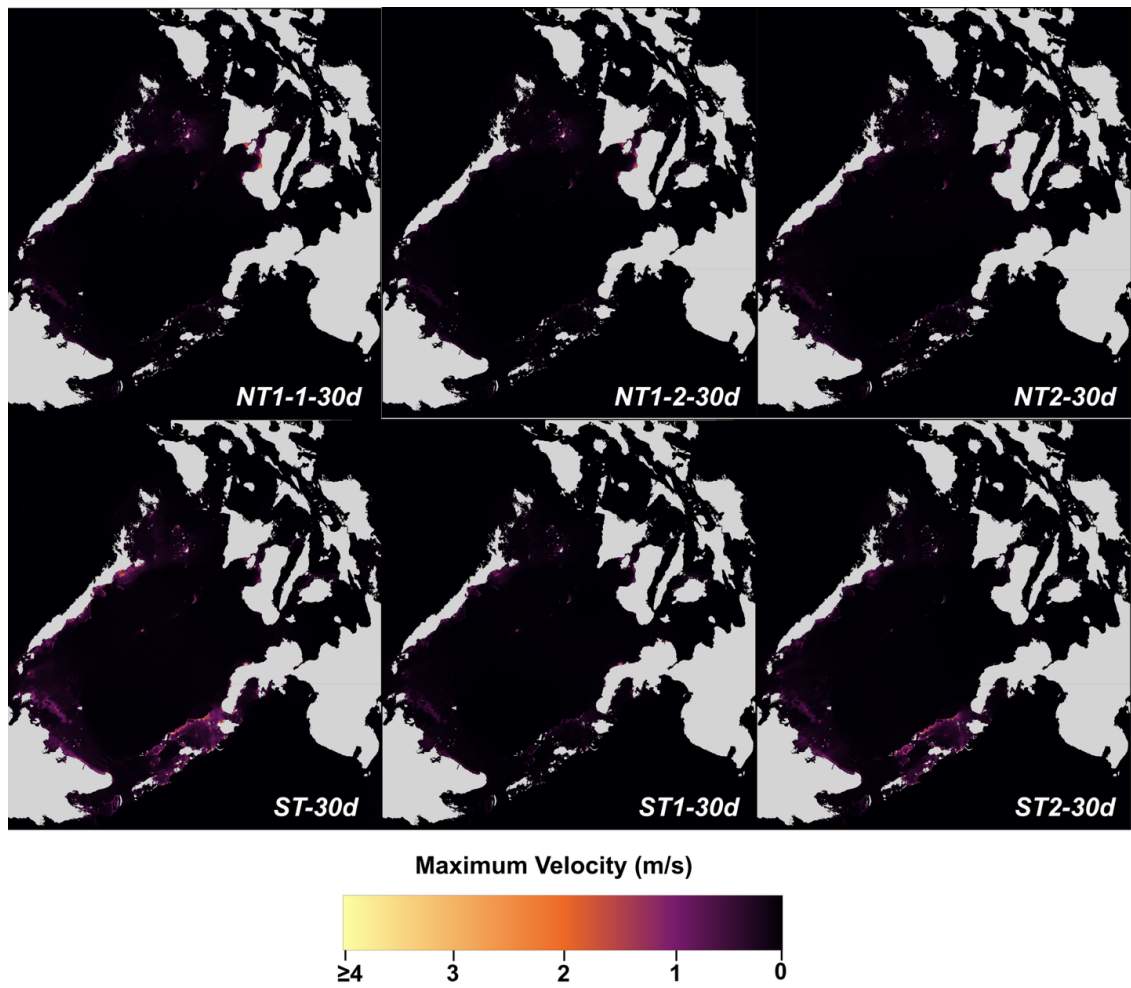


Figure S11. Maximum flow velocity (m/s) from the 4-h tsunami simulation of six segments with a megathrust dipping at 30°.

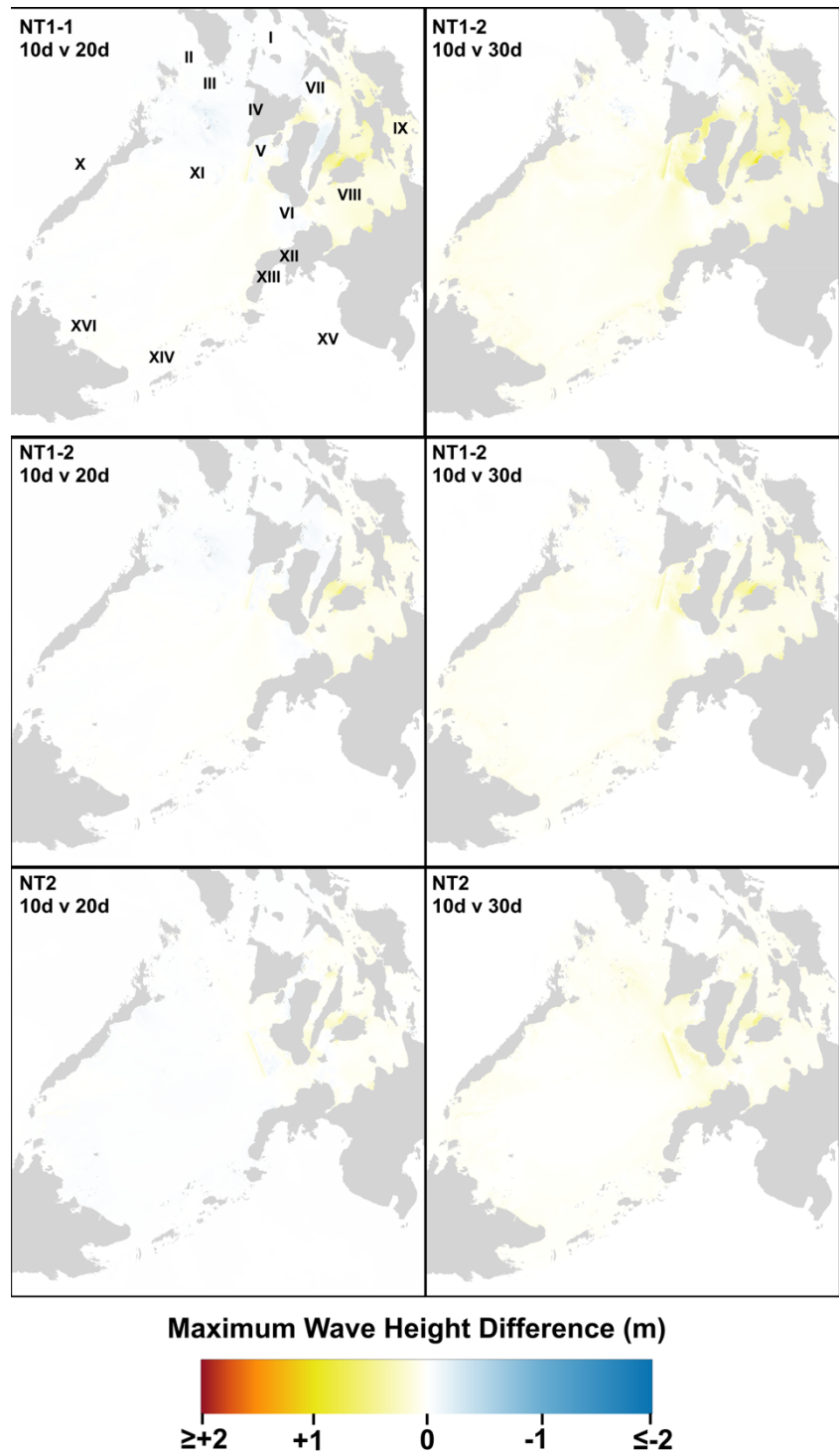


Figure S12. Maximum wave height difference from 10°-dip megathrust compared to that of 20° and 30° in the Negros Trench segments.

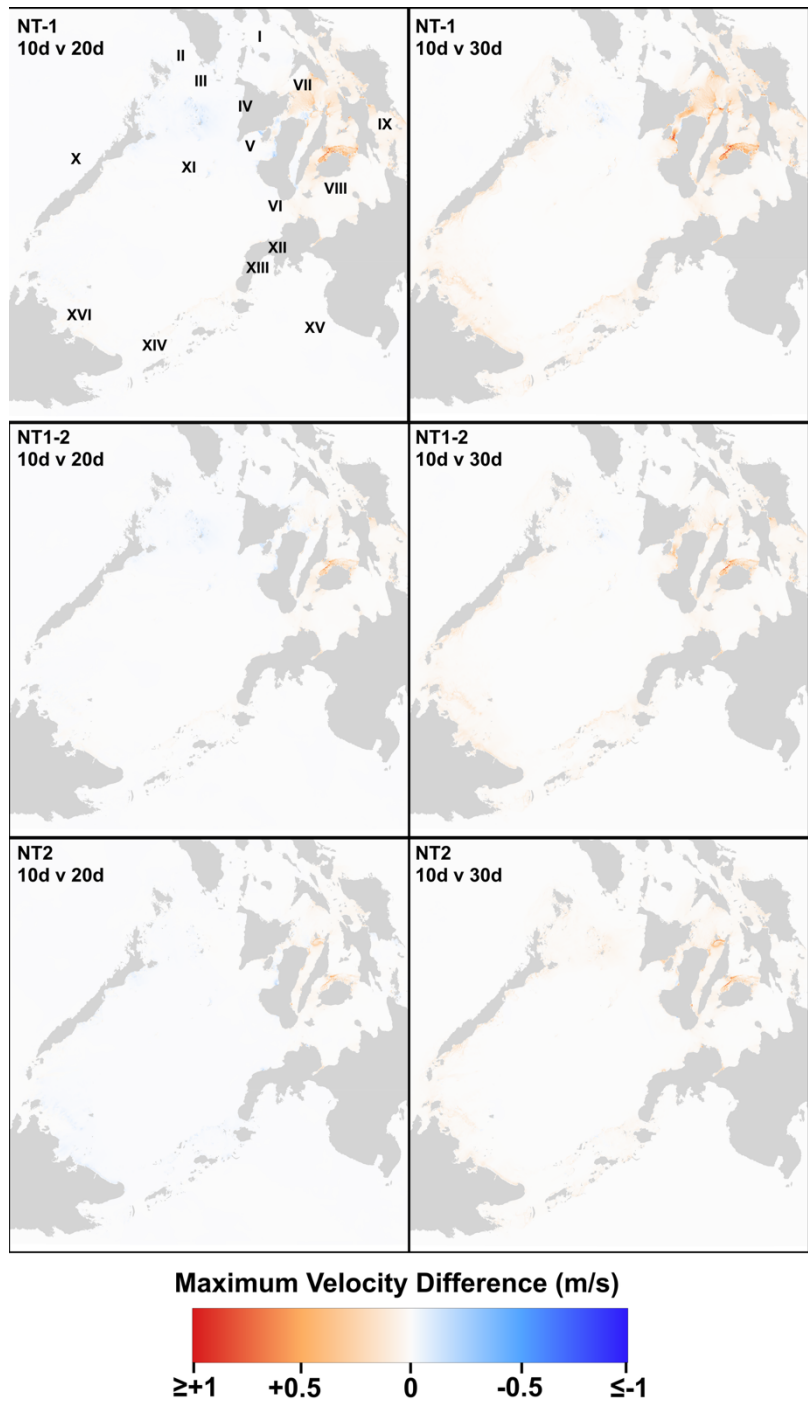


Figure S13. Maximum flow velocity difference from 10°-dip megathrust compared to that of 20° and 30° in the Negros Trench segments.

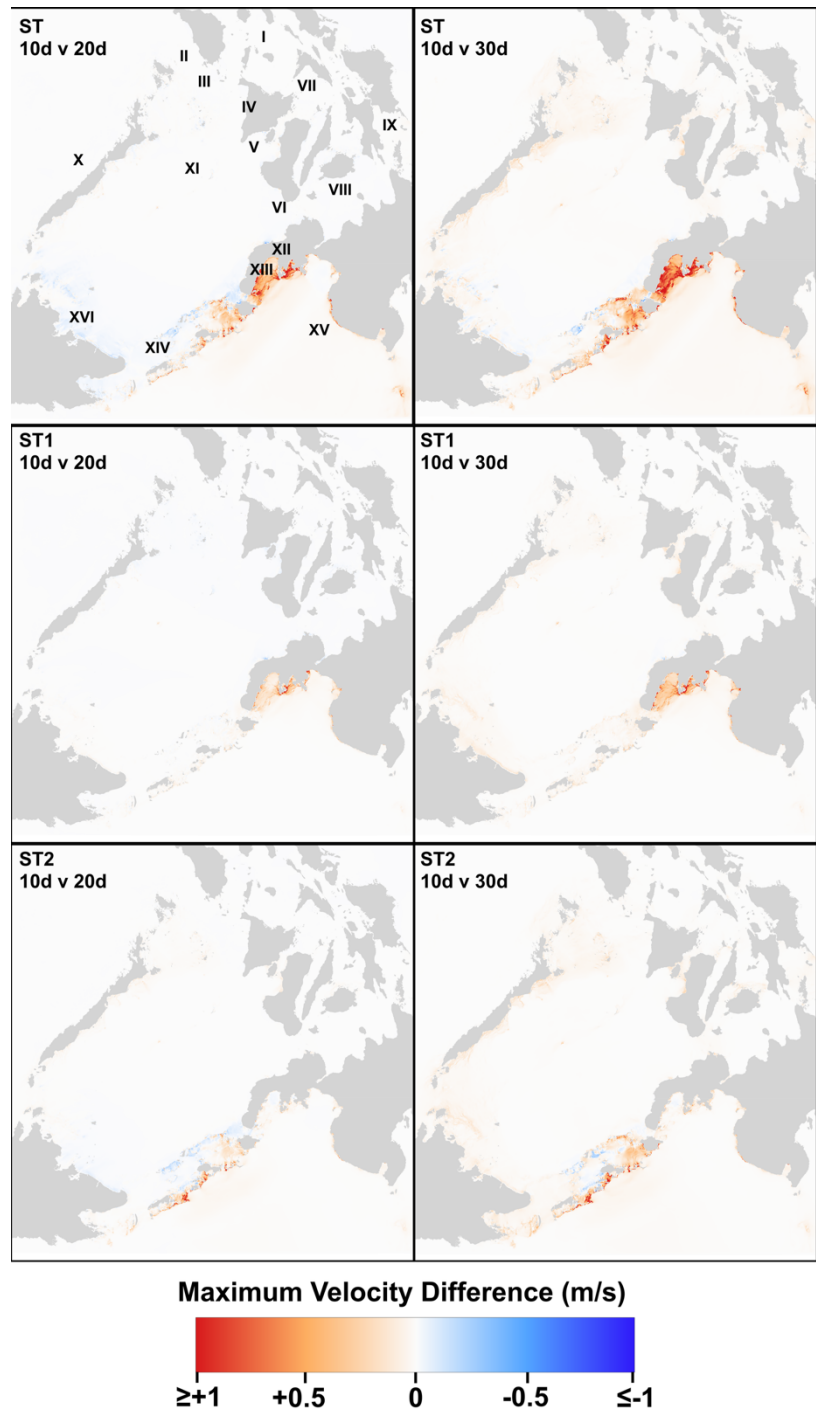


Figure S14. Maximum flow velocity difference from 10°-dip megathrust compared to that of 20° and 30° in the Sulu Trench segments.

- Although the manuscript is generally well written, I found some parts difficult to understand.

See the detailed comments for more information.

In my judgement addressing these issues will require major revisions before the manuscript is suitable for publication. But I think this is certainly achievable.

Detailed comments

L30: "...that underscore accurate nearshore bathymetry..." -- consider replacing with "that underscore the need for accurate nearshore bathymetry ..."

L31 -- instead of "exposure maps" consider just "maps", since usually 'exposure' implies an analysis of how the built environment could be affected.

L34: In recommending the use of modelling for policy, it will be very important to consider limitations and uncertainties [is the 6-8m peak really enough, given the historical observations?]

Response/Action:

Thank you for this clarification. We agree that the 6–8 m maximum wave height is more likely underestimated than overestimated. The abstract has been revised to state that these maximum wave heights and flow velocity are treated as minimum, as shown below:

The Negros–Sulu megathrust poses an imminent threat to coastal communities surrounding the Sulu Sea, but with limited information on past tsunamis, megathrust geometry, and locked segments, robust seismic and tsunami hazard assessments are hindered. To identify highly exposed coastal areas amid this knowledge gap, we estimated the Negros–Sulu megathrust source parameters using a structural-based segmentation and scaling-relation approach for tsunami modeling. A total of 18 sets of source parameters from six segments and three dip angle scenarios were considered with moment magnitude and average slip of Mw 8.0–8.9 and 1.68–6.23 m, respectively. Tsunami simulations were modeled in JAGURS, accounting for nonlinear shallow water equations, horizontal and vertical seafloor displacements, and Boussinesq dispersion effects. Coastal regions directly facing the segments have the highest exposures with <2 min arrival times, highlighting the major control of wave directivity and the need for rapid evacuation strategies. The Negros Trench generates up to 6 m wave height and 7 m/s flow velocity, while the Sulu Trench produces up to 8 m and 6 m/s. Coastal areas with ≥ 2 m wave heights typically exhibit a concave morphology with a nearshore width interquartile range of 2–4.5 km. At wider nearshore widths (>20 km), wave dissipation results in lower wave heights (<2 m) that underscore the need for accurate nearshore bathymetry in tsunami modeling. As average slip is used in these models, we highlight that the maximum wave height and flow velocity are likely underestimated and recommend treating them as minimum values. Nonetheless, the relative exposure levels of coastal areas from these results provide a significant basis for regional-scale tsunami risk mitigation strategies and for searching tsunami deposits.

L44: Notice these historical earthquakes, with magnitudes in the low 8s, had peak wave heights that exceed the largest modelled scenarios in this study. That suggests that the modelled scenarios are not accounting for the diversity of real earthquake-generated

tsunamis. Furthermore, globally, it is common for earthquakes with magnitudes in the mid 8s to produce waves much larger than modelled here (e.g. Cheung et al. 2022).

Response/Action:

We agree on the underestimation of these models. However, in 5.2 *Implications for Tsunami Hazard Potential*, we stated that these eyewitness accounts also pose uncertainties, especially for older events like the 1897 Ms 7.5 earthquake. Other rupture mechanisms (e.g., outer rise, tsunami earthquake, intraplate, upper-plate earthquakes) may also have occurred in these events that deviated from megathrust earthquakes. Lines 335–344 are shown below:

It should be noted, however, that these eyewitness accounts pose uncertainties and are only available in certain coastal areas that have been inhabited before these events. For events that occurred before the 1960s, the estimated earthquake magnitudes are based on reported damages or anecdotes. The magnitude of the 1897 Ms 7.5 earthquake, for example, may have been underestimated. This event may have also been a tsunami earthquake which generates large tsunami waves compared to its earthquake magnitude (e.g., Tanioka and Seno, 2001), or an earthquake-triggered submarine landslide also contributed to the basin-wide tsunamis as these submarine features are prevalent in this active margin (Nawanao & Ramos, 2023). As our models assumed blind megathrust ruptures, these historical events may also have been generated by other rupture mechanisms (e.g., outer rise, tsunami earthquake, intraplate, upper-plate earthquakes) with a different set of parameter estimations. With the scarcity of information, these historical accounts are still valuable to compare with tsunami models. What is consistent is that the coastal positions and configuration relative to the trench dictate where the highest exposures are identified, i.e., controlled by wave directivity.

L50-55: More care is needed when discussing interpretations of seismic gaps, coupling and slip deficits. It is not as straightforward as implied. At the global scale, there are sites where geodetic evidence suggests weak coupling even though there is significant evidence for tsunamis (e.g. Witter et al 2016, Debaeker et al 2026). It has been suggested that the neglect of viscoelastic effects in geodetic studies can lead to mistaken conclusions about the readiness of faults to rupture (Wang et al., 2021). Also, the recent Kamchatka 2025 earthquake may have slipped more than the deficit accumulated since the last large event (Liu et al. 2026). Thus, these concepts should not be blindly accepted.

Response/Action:

Thank you very much for pointing this out. This has been incorporated in lines 85–88:

However, even with relatively denser declustered seismicity, it should be noted that the rheological conditions and seismic behavior are not fully captured due to the complex factors affecting frictional heterogeneity (Chlieh et al., 2008; Fagereng and Beall, 2021; Fan and Zhao, 2021; Volpe et al., 2023; Yabe and Ide, 2018) and the short span of observations relative to the apparent hundred-to-thousand year recurrence of megathrust earthquakes.

L63: insert "and hydrodynamic complexities such as " before "nonlinear effects"

Response/Action:

This has been adopted in lines 63–66 in the revised manuscript.

Figure 1:

- In panel 'b' it is hard to see the coastline or otherwise locate the pixels. Consider adding the trench locations to make this easier.
- In panel 'a', what are the velocity vectors measured relative to?
- In panel A, the 'max wave height' results do not all appear to be on the coast. Are they reported at the earthquake location? Make sure the reader knows how to interpret them.

Response/Action:

The velocity vectors are relative to no-net-rotation reference frame, which is incorporated in the revised figure caption. The location of the points is the epicenter, while the colors are the eyewitness accounts of maximum wave heights. Figure 1 and caption were revised as shown below:

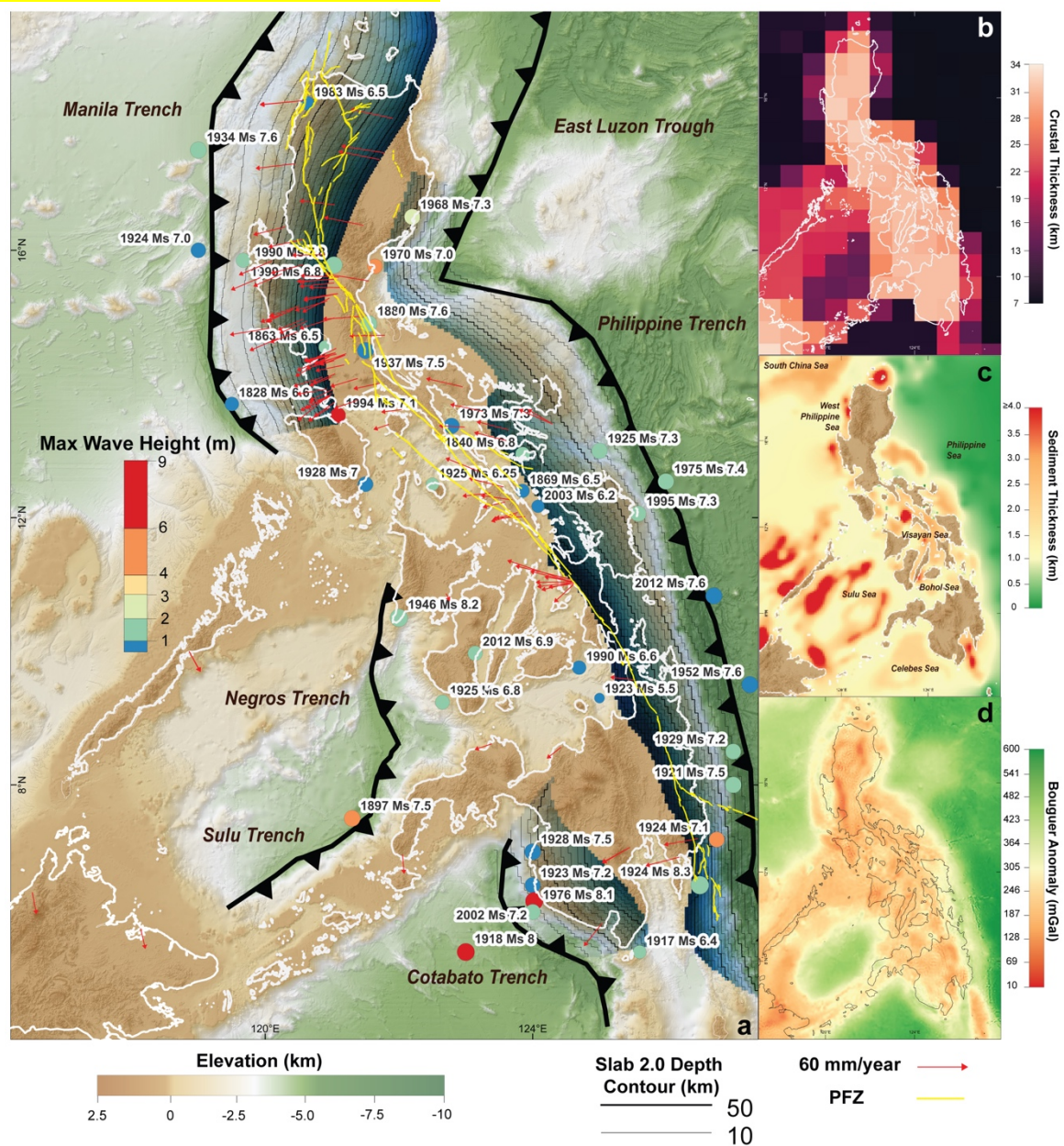


Figure 1. Philippine tectonic setting. (a) Estimated epicenters of positive tsunamigenic earthquakes and the associated eyewitness accounts of maximum wave heights in the affected coastal areas since 1589 (Bautista et al., 2012). The trace of the Philippine Fault Zone (PFZ) was adapted from PHIVOLCS. Depth contours of the Slab 2.0 model from Hayes et al. (2018). Arrows indicate the velocity vectors relative to no-net-rotation reference frame from inland GPS networks (Kreemer et al., 2014). Bathymetry and topography data from the 2024 General Bathymetric Chart of the Oceans (GEBCO) (GEBCO Compilation Group, 2024). (b) Crustal thickness (Laske et al., 2013), (c) sediment thickness (Straume et al., 2019), and (d) Bouguer anomaly (Balmino et al., 2012) across the archipelago.

L77: "where < 1.0 -- suggest to replace with "where smaller values ". Also consider replacing "are associated with a higher proportion of large earthquakes and stress accumulation" with "indicate that large earthquakes contribute more to the long-term seismic moment rate"

L78: "the evaluated number of declustered seismicity" -- suggest to replace with "the number of earthquakes in the declustered catalogue"

Response/Action:

Thank you for the suggestions. These have been adopted in lines 77–83:

Peñarubia et al. (2020) evaluated an empirical Gutenberg–Richter (GR) magnitude–frequency distribution (MFD) in these subduction zones. The b-values range from 0.691 to 1.724, based on the declustered seismicity catalog from the International Seismological Centre (ISC) and the Philippine Institute of Volcanology and Seismology (PHIVOLCS), where <1.0 indicates that large earthquakes contribute more to the long-term seismic moment rate (Nishikawa and Ide, 2014). In the study, the number of earthquakes in the declustered catalog in all trench segments was less than 300, with the majority of the segments having <50 earthquakes (Peñarubia et al., 2020).

L80-82: Note there will be substantial statistical uncertainty in the GR parameters when using small numbers of earthquakes. There will also be uncertainties about the appropriateness of the model appropriateness (even with many earthquakes).

Response/Action:

Thank you for this comment. We incorporated the GR limitations in lines 85–88:

However, even with relatively denser declustered seismicity, it should be noted that the rheological conditions and seismic behavior are not fully captured due to the complex factors affecting frictional heterogeneity (Chlieh et al., 2008; Fagereng and Beall, 2021; Fan and Zhao, 2021; Volpe et al., 2023; Yabe and Ide, 2018) and the short span of observations relative to the apparent hundred-to-thousand-year recurrence of megathrust earthquakes.

Equations 1 - 3:

- Eq 1 is using the 'unlimited width' Allen and Hayes (2017) regression for rupture area, while also limiting the width to 196 km. Allen and Hayes (2017) provide another regression relation for the case with a capped width, which is their preferred model. Why not use the 'capped width' regression relation, given that widths are capped in this study?

- Equation 2/3 provide both the average slip & the max slip. How is the max slip used to make the earthquake sources? This isn't clear. If both max and average slip are used,

there must be some spatial non-uniformity of the hypothetical slip models. That needs to be clearly described.

- I don't like the use of superscripts as footnotes to describe the standard deviation, because the meaning is not standardised and readers will easily get confused. If the uncertainty in the predictors is used (a very good idea) then why not include terms such as (for $\sigma = 0.266$) " $+ 0.266 \times \epsilon$ " and state that " ϵ " has a standard normal distribution. If M_w is predicted from the dimensions (instead of the converse) then provide that equation separately with its own predictive uncertainty. Alternatively, use a table like in Allen and Hayes (2017) to treat both cases.

Response/Action:

In these models, the width was calculated as a function of the depth of the seismogenic zone and the dip angle. The width for the 10° megathrust was capped at 196 km, not necessarily only due to the scaling equation used, but to account for the apparent saturation of global subduction zones, as the trigonometric width of 259 km for a 10° slab would likely be unrealistic. We used the area– M_w scaling equation rather than width– M_w to also account for the length of the megathrust. Furthermore, we used the linear area– M_w equation rather than the bilinear area– M_w equation as the former has a better Akaike information criterion (AIC) in the study. The former equation also produces higher M_w and thus a higher average slip, as we chose to be conservative of the M_w magnitude. These were included in the revised text in lines 140–149. The maximum slip was not used in the model but was included to serve as a future reference. To avoid confusion, we adopted your comment on removing the maximum slip in this manuscript. The uncertainty terms were also included in the text as suggested:

To account for the variability of length and width, the area (S) of the fault plane is used as input in the linear rupture-scaling equations (**Eqs. 1–2**) of Allen and Hayes (2017) to estimate the maximum M_w and the average (D_{ave}) slip (m). The linear area– M_w was preferred over the bilinear area– M_w scaling equation as the former produced slightly higher M_w , adopting a more conservative approach. A total of 18 sets of source parameters were constrained for elastic dislocation and tsunami modeling, with the ST (379 km) and NT1-1 (201 km) having the longest lengths (**Table 1**). Estimates of the moment magnitude from scaling equations range from M_w 8.0 to 8.9 (± 0.27), where the ST segment with the longest fault length has the largest magnitude. These earthquake magnitudes correspond to an average slip of 1.68–6.23 (± 0.21) m.

$$\log_{10} S = -3.63 + 0.96 \times M_w \quad (1)$$

$$\log_{10} D_{ave} = -5.05 + 0.66 \times M_w \quad (2)$$

General comment on Section 3.1: From the description that has been provided I cannot tell what the source parameters are (length, width, magnitude, mean slip), what the magnitude is, or how the maximum slip is used. Please be precise about *exactly* what was done to construct the scenarios, and remove any extraneous equations. It will help to move Table 1 up to this part of the text. Remove the maximum slip if it isn't actually being used (although I would prefer if the scenarios accounted for possible localised high slip). d

Another general comment on Section 3.1: Please carefully consider whether these source models are realistic, noting historical observations of earthquakes closer to Mw 8 with larger waves than most modelled scenarios (despite the latter having larger magnitudes). This could partly be due to the coarse resolution of the tsunami model at the coast. But I think it's very likely to reflect that uniform slip tsunami sources with average scaling-relation slip sometimes do a poor job of representing real tsunamis. In such cases compact uniform slip models or heterogeneous slip models often perform better (e.g. An et al., 2018; Davies, 2019, 2025). One reason is that often localised patches of high slip ("asperities") contribute disproportionately to the tsunami.

Response/Action:

This comment has been adopted. In order for the readers to directly find the source parameters mentioned, Section 4.1 was moved to Section 3.1, including Table 1.

3.1 Constraining Megathrust Fault Parameters

Simplified rectangular fault models were adapted for the megathrust with parameters length (km), width (km), strike ($^{\circ}$), dip angle ($^{\circ}$), depth from the seafloor (km), rake ($^{\circ}$), and slip amount (m). Insights from previous seismic tomography profiles (Fan et al., 2017; Fan and Zhao, 2019; Sevilla, 2011) on the geometry of the subducting slab reveal tapering of high p-wave velocity in the northern Negros Trench segment west of Panay Island, which provides limiting constraints on the length of the northern Negros Trench segment. Steepening of the subducting slab was also observed in the seismic tomography profiles around 100 km depth, although this is well beyond global seismogenic zones. Best-fit estimates on the dip of the shallow portion of the subducting slab derived from the tomography profiles within 50 km depth range from 9° to 30° . Focal mechanism solutions (ISC data from 1977 to 2023) of reverse-oblique rake values (10° to 170°) (Bondár et al., 2015; Di Giacomo et al., 2015, 2018; Storchak et al., 2013, 2015) also show a wide distribution of dip angles ranging from 9° to 90° (**Figure 2**). To account for this uncertainty of the slab dip, we used three dip angles for all source parameters at 10° , 20° , and 30° .

We utilized a structural-based segmentation in which the megathrust fault length is derived from the orientation variations of the frontal wedge (i.e., Nawanao & Ramos, 2023), and the width and bottom depth are limited by the seismogenic zone. These led to four main segments of the Negros–Sulu subduction zone, namely, NT1, NT2, ST1, and ST2. The NT1 and NT2 are marked by north-northeast and north-northwest orientations, respectively, while the ST1 and ST2 by north-northeast and northeast, respectively (**Figure 3a**). Due to the uncertainty of the northern extent of NT1 where p-wave velocity profiles of Fan et al. (2017) and Fan and Zhao (2019) indicate tapering of the slab northward, we included longer (NT1-1) and shorter (NT1-2) rupture lengths for NT1. For the Sulu Trench, an additional longer segment (ST) that combines ST1 and ST2 is also modeled due to their similar northeast trend. A total of six subfaults, i.e., NT1-1, NT1-2, NT2, ST, ST1, ST2, were considered for the tsunami modeling.

A shallow blind rupture is assumed for all scenarios with a top depth of 5 km and a rake of 90° (purely reverse slip). Based on the interquartile range (1.5IQR) of 207 reverse-oblique earthquakes (**Figure 2**), a threshold depth of 50 km was set as the

bottom depth of the seismogenic zone. The width of the megathrust is calculated using sine function from the top and bottom depth, with the given the dip angles, and we adopted a saturation limit of 196 km as observed in global subduction zones (Allen and Hayes, 2017; Skarlatoudis et al., 2016; Tajima et al., 2013). The dip angles 10°, 20°, and 30° correspond to 196 km (maximum width limit), 132 km, and 90 km, respectively. (**Figure 3**).

To account for the variability of length and width, the area (S) of the fault plane is used as input in the linear rupture-scaling equations (**Eqs. 1–2**) of Allen and Hayes (2017) to estimate the maximum M_w and the average (D_{ave}) slip (m). The linear area– M_w was preferred over the bilinear area– M_w scaling equation as the former produced slightly higher M_w , adopting a more conservative approach. A total of 18 sets of source parameters were constrained for elastic dislocation and tsunami modeling, with the ST (379 km) and NT1-1 (201 km) having the longest lengths (**Table 1**). Estimates of the moment magnitude from scaling equations range from M_w 8.0 to 8.9 (± 0.27), where the ST segment with the longest fault length has the largest magnitude. These earthquake magnitudes correspond to average slip of 1.68–6.23 (± 0.21) m.

$$\log_{10} S = -3.63 + 0.96 \times M_w \quad (1)$$

$$\log_{10} D_{ave} = -5.05 + 0.66 \times M_w \quad (2)$$

Regarding the maximum wave heights of eyewitness accounts, we added in lines 335–344) (*5.2 Implications for Tsunami Hazard Potential*) the uncertainties of these events and other mechanisms, deviating from megathrust rupture, that may have generated these historical earthquakes.

L150-155: The tsunami model is using low-quality elevation data, and its 15 arcsecond resolution corresponds to cell sizes of roughly 450 m. I think this will be OK to model offshore waves. But 450m resolution is insufficient for mapping inundation, and may miss sites with more localised coastal amplification. This should be mentioned.

Response/Action: We are limited by the availability of bathymetry data, as currently there are no available high-resolution nearshore bathymetry (<100 m depth) connecting inland. While another approach is to make the GEBCO bathymetry data finer, the quality of the depth values still relies on the depth from the original ~450-m resolution bathymetry. Adopting this comment in lines 372–374, we recommended gathering high-resolution and accurate nearshore bathymetry data to better capture local-scale wave amplification and deformation as shown below:

Gathering of high-resolution bathymetry data is necessary to capture the local-scale wave amplification and deformation. The accuracy of nearshore bathymetry connecting inland and offshore must also be assessed, as the nearshore width shown in this study has significant effects on the resulting maximum tsunami waves along the coasts.

L190 -- Both average and maximum slip are presented, but I cannot see how the maximum slip is being used to represent the sources. If it's not used, it shouldn't be mentioned. That said, I would strongly encourage consideration of alternative, more compact scenarios that have higher slip, which may better describe the large waves seen historically for earthquake magnitudes in the low 8s.

Response/Action:

This has been adopted. The maximum slip was removed in the manuscript. We highly considered the suggestion on alternative slip models, and in future studies, heterogeneous and more compact slip models will be conducted.

Table 2 isn't easy to interpret. For instance, it's hard to know whether the max/min differences reflect a significant part of the model, or just isolated differences. Also the standard deviations I think it would be better to provide figures in the supplementary material for each case, using a colour scheme with more variability near zero, so we can well see variations at both small and large scales.

Response/Action:

The statistics are based on the differences of the entire boundary between models and thus the minimum and maximum are isolated differences. The color schemes and raster extents of values in Figures 11, S12, and S13, were revised to better highlight the differences between models.

L318-328: This paragraph mentions how the simple tsunami source representations might cause over-or-under estimation of the tsunami size. I think it needs to be edited to mention how the maximum runups observed in historical events with magnitudes in the low 8s are sometimes larger than the modelled scenarios, clarifying that underestimation is very plausible.

Response/Action:

This has been adopted by adding lines 358–360 (*5.3 Limitations and Future Work*) in the paragraph:

With the bathymetry data used in this study and the lack of variable and asperity-based slip distribution when considering megathrust ruptures, we recommend that the maximum wave heights are treated as minimum and that the actual highest tsunami waves are more likely underestimated.

L329-330: "The use of single-fault models with uniform slip distributions in this study can serve as baseline information in light of the scarcity of geodetic data for finite-fault slip models." -- You don't need geodetic data to construct non-uniform slip models -- there is a huge literature on stochastic approaches -- see e.g. Geist (2002), Geist and Oglesby (2014). I think this statement should be removed.

L330: Here I think the text is misrepresenting the results of An et al (2018). The latter study shows that uniform slip models can work well if they are unusually compact (something also found in Davies (2019, 2025)). They don't provide support to more naive uniform slip models as used in this manuscript, which often underestimate tsunami wave heights (according to the latter studies).

Response/Action:

Thank you for pointing these out. These statements have been removed in the revised manuscript.

Figure 15: Please add any historical observations of runup to the plot, along with a label showing the earthquake magnitude. This will help make the model limitations clear, and can be integrated into the limitations discussion.

Response/Action:

Figure 15 has been revised, and paragraphs have been included in 5.2 *Implications for Tsunami Hazard Potential*, discussing the comparison between the tsunami models and historical accounts as shown below:

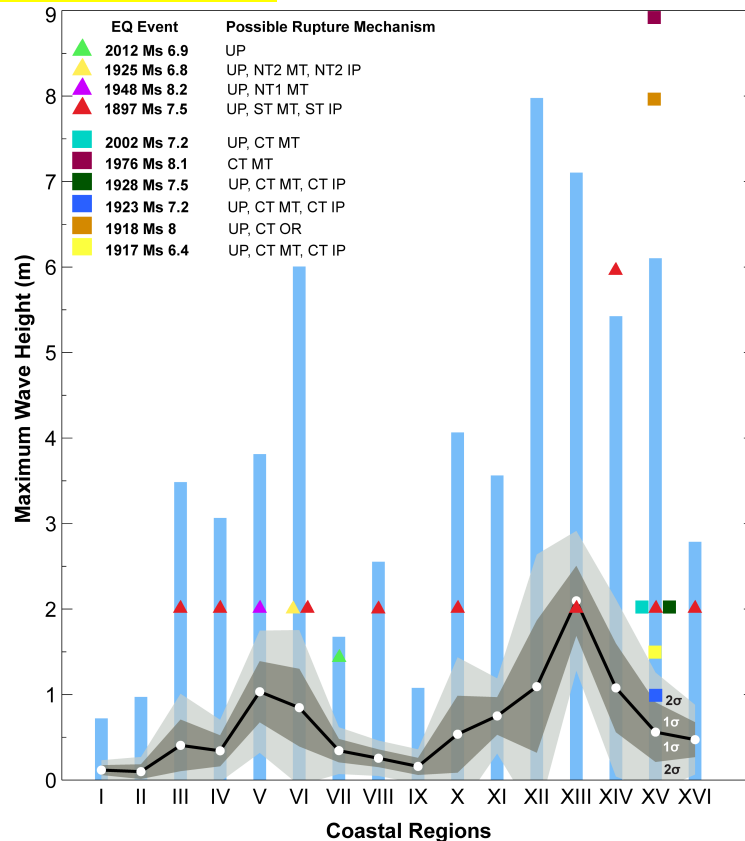


Figure 18. Comparisons of maximum wave heights among the 16 coastal regions from the models and eyewitness accounts (triangles and squares) of historical tsunamigenic earthquakes in the Sulu and Celebes Sea from Bautista et al. (2012). White circles and blue bars indicate the mean and maximum values, respectively, while the envelope indicates the 1σ and 2σ standard deviations. I, Visayan Sea; II, Mindoro; III, Coron–Cuyo; IV, West Panay; V, Panay Gulf; VI, Sipalay–Bayawan; VII, Tanon–Cebu Strait; VIII, Bohol Sea; IX, Leyte Gulf; X, Palawan; XI, Cagayancillo; XII, Sindangan–Labason; XIII, Siocon–Sibuco; XIV, Basilan–Sulu; XV, Moro Gulf; XVI, northeast Borneo. Rupture mechanisms: UP, upper plate; MT, megathrust; IP, intraplate; OR, outer rise. Subduction zones: NT, Negros Trench; ST, Sulu Trench; CT, Cotabato Trench.

We compared the maximum wave heights from the models to the eyewitness accounts of historical tsunamis in the Sulu and Celebes seas from Bautista et al. (2012). Although the source mechanisms of the majority of earthquakes were unknown, we can attribute them to megathrust, upper-plate, outer-rise, or intraplate ruptures, depending on the hypocenter when available and the position of the estimated epicenter relative to the trench, as labeled in Figure 15. The estimated epicenter of the 1897 Ms 7.5 event was along the Sulu Trench with a 2 m maximum wave height in the majority of the Sulu Sea basin (III, IV, VI, X, XIII, XV, XVI) and up to 6 m in coastal region XIV. In III, IV, VI, X, XIII, XV, and XVI, the highest

modeled tsunami wave in this study is slightly greater than the eyewitness accounts. However, in XIV where a 6 m wave height was estimated, the highest modeled tsunami wave is slightly lower at 5.4 m. Eyewitness accounts of the more localized tsunamis by the 2012 Ms 6.9, 1925 Ms 6.8, and 1948 Ms 8.2, reported up to 2 m wave height, which is lower than the highest modeled tsunami. In Moro Gulf (XV), six historical tsunamigenic earthquakes (2002 Ms 7.2, 1976 Ms 8.1, 1928 Ms 7.5, 1923 Ms 7.2, 1918 Ms 8, 1917 Ms 6.4) were generated along the Cotabato Trench (CT) with eyewitness accounts reporting tsunami wave heights ranging from 1 m to 9 m. In this region, the highest modeled tsunami from the Sulu Trench reached 6 m. This indicates that this coastal region is not only susceptible to the Cotabato subduction zone rupture but also to the Sulu Trench. In addition, prehistoric extreme wave deposits that may be recovered from this coastal region (e.g., Claro et al., 2021) may have also been deposited from Sulu megathrust rupture and tsunami.

It should be noted, however, that these eyewitness accounts pose uncertainties and are only available in certain coastal areas that have been inhabited before these events. For events before the 1960s, the estimated earthquake magnitudes are based on reported damages or anecdotes. The magnitude of the 1897 Ms 7.5, for example, may have been underestimated. This event may also have been a tsunami earthquake which generates large tsunami waves compared than its earthquake magnitude (e.g., Tanioka and Seno, 2001), or an earthquake-triggered submarine landslide also contributed to the basin-wide tsunamis as these submarine features are prevalent in this active margin (Nawanao & Ramos, 2023). As our models assumed blind megathrust rupture, these historical events may also have been triggered by other rupture mechanisms (e.g., outer rise, tsunami earthquake, intraplate, upper-plate earthquakes) with a different set of parameter estimations. With the scarcity of information, these historical accounts are still valuable to compare with tsunami models. What is consistent is that the coastal positions and configuration relative to the trench dictate where the highest exposures are identified, i.e., controlled by wave directivity.