Fingerprinting subduction margins using principal component analysis profiles: A data science approach to assessing earthquake hazard

Valerie A. Locher[1], Rebecca E. Bell[1], Parastoo Salah[1], Robert Platt[1], and Cédric M. John[2,1]

[1]Imperial College London, Department of Earth Science and Engineering, Royal School of Mines, London, SW7 2AZ, United Kingdom

[2]Queen Mary University of London, Digital Environment Research Institute, Empire House, 67-75 New Road, London, E1 1HH, United Kingdom

ABSTRACT

Giant earthquakes ($M_w \geq 8.5$) along subduction margins pose great hazards to coastal societies. While it is generally accepted that geological margin properties play a role, controls on giant earthquake occurrence remain undetermined. The long intermittence times of giant earthquakes and the comparatively short earthquake record obscure any correlations between margin properties and seismicity.

This work presents a new approach to relating margin properties to seismicity. We apply principal component analysis (PCA) to four margin properties to 'fingerprint' margins by assigning them a PCA profile, which we compare to giant earthquake occurrence. This approach reduces bias from the short earthquake record because seismicity is not used as a PCA input feature. Using kernel PCA, a nonlinear PCA variant, we find nonlinear patterns in margin properties and suggest that links between properties and seismicity are nonlinear, which helps explain why they have been hard to establish. PCA clusters identify "active and moderate" and "quiet and extreme" margins. We argue that margin segments with "quiet and

extreme" PCA profiles but no giant earthquakes since 1900 are as hazardous as those that have ruptured in giant earthquakes recently.

INTRODUCTION

Giant earthquakes ($M_w \geq 8.5$) along subduction margins, such as the 2004 Sumatra-Andaman, 2010 Chile, and 2011 Tohoku (Japan) earthquakes, pose serious hazards to coastal societies. The control mechanisms that initiate giant earthquakes remain uncertain, in part due to long earthquake recurrence times. Not all subduction margins have experienced giant earthquakes since 1900, the start of our well-documented seismic record (e.g. McCaffrey, 2008; Schellart & Rawlinson, 2013). This poses the question whether certain geological margin properties enable or inhibit large earthquakes (e.g. Ruff, 1989). Alternatively, perhaps all subduction margins can initiate large earthquakes, and the apparent lack thereof is simply due to the short earthquake record (e.g. Stein & Okal, 2007; McCaffrey, 2008).

Geological controls on subduction margin seismicity have been explored in some detail, both focusing on single margins (e.g. Wallace et al., 2009) and considering global data compilations of margin properties (e.g. Ruff, 1989; Schellart & Rawlinson, 2013). Early interpretations linked giant earthquake occurrence to fast-converging young plates, suggesting that plate density and convergence rate control seismicity (e.g. Ruff & Kanamori, 1980). However, this model could not explain the 2004 Sumatra and 2011 Tohoku earthquakes, which occurred in areas of slow convergence and old plate age, respectively (Stein & Okal, 2007, 2011). Further studies explored properties such as the seabed roughness of the incoming plate, which has been linked to intraplate coupling (e.g. Lallemand et al., 2018) or its bending angle, thought to influence seismogenic zone width and plate hydration (e.g. Nishikawa & Ide, 2015). Multiple studies linked high sediment thickness to giant earthquake occurrence, concluding that abundant sediment availability increases intraplate locking (e.g. Ruff, 1989; Heuret et al., 2012; Scholl et al., 2015). However, the 1952

Kamchatka (Russia) and 2011 Tohoku earthquakes (both M_w 9) challenge this hypothesis, given that they occurred in areas of fairly low sediment thickness (Scholl et al., 2015).

Based on these observations, many have concluded that seismicity is determined by a complex interplay of multiple factors (e.g. Wallace et al., 2009; Schellart & Rawlinson, 2013; Wirth et al., 2022). Recent studies thus examined connections between margin properties and maximum observed earthquake magnitude using multivariate statistics and regression approaches (e.g. Brizzi et al., 2018; Nakao et al., 2023). However, short measurement and historical records compared to giant earthquake intermittence times commonly lead to underestimated maximum magnitudes, meaning such models are trained on incomplete data.

We present an unsupervised approach to explore correlations between four properties (sediment thickness, relative plate velocity, and the subducting plate's dip angle and roughness) and maximum magnitude. We fingerprint margin segments by applying principal component analysis (PCA) to project the property data into a lower-dimensional orthogonal space where their location is described by their PCA profile. By excluding maximum magnitude data from the PCA input and using them only to infer seismic behaviours of different PCA profiles, we reduce bias resulting from the short earthquake record. We then utilize correlations between PCA profiles and maximum magnitude to assess giant earthquake potential for margins with no such events since 1900. Our results suggest that connections between margin properties and seismic behaviour are nonlinear.

METHODS

Data Preprocessing

We utilize a dataset encompassing sediment thickness, dip angle, roughness, and convergence rate for 1540 margin segments, each measuring 25 x 200 km, oriented with their long axis orthogonally to and centred on the subduction trench along margins (compiled by McLellan & Audet, 2020 for a study relating margin properties to slow slip). Roughness

values were derived from gravimetry data thus best represent oceanic basement roughness (see Smith, 2014, and Bassett and Watts, 2015, for context on determining roughness). Data pre-processing included imputing missing values, scaling the data (see the Supplemental Material[1]), and log-transforming roughness, dip, and sediment thickness values.

Using U.S. Geological Survey (1900 - 2023) and Global Historical Earthquake Catalogue (Albini et al., 2013) data, we assign each segment a maximum observed magnitude (Fig. 1) with a custom binning algorithm: we calculate rupture length for each earthquake using empirical relationships from Wells and Coppersmith (1994) and draw a circle of that diameter around the earthquake's epicentre; any segment within this circle is assigned the earthquake's magnitude. We recognise that rupture lengths and geometries are more complex, but detailed data are available only for few well characterised events, and thus cannot be used in our global study. Figure S9 in the Supplemental Material compares our rupture length estimates with finite fault rupture models for 63 events. Because all margins are longer than the largest rupture lengths (Fig. 9), we exclude margin length as a limiting factor in our analysis.

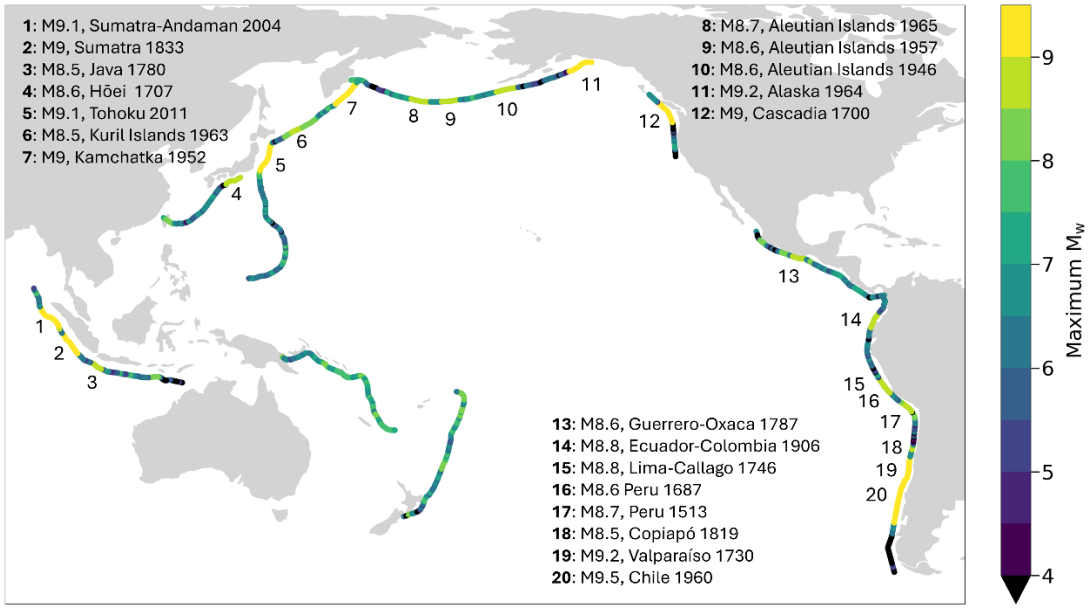


Figure 1: Map showing observed maximum earthquake magnitudes, with year, for considered subduction margins.

Principal Component Analysis

Principal component analysis (PCA) is a method to reorient data along orthogonal axes of highest variance, called principal components (PCs) (Jolliffe, 2002). Given that PCs are numbered in order of decreasing amount of explained variance in the original data, considering only the first few PCs (e.g. PC1 and PC2 here) can reduce a dataset's dimensionality while retaining most of its information. By applying PCA to margin property data, we observe margin segments in an interpretable two-dimensional (2D) space while considering multiple properties.

Kernel PCA (Schölkopf et al., 1998) uses kernels to map data to a higher-dimensional space before PCA, whereby different kernels result in different projections. This enables the detection of complex, nonlinear patterns that linear methods cannot capture. We generate a diverse set of projections by applying PCA and kernel PCA to the margin property data, experimenting with an assortment of kernels, including linear, polynomial, radial basis function (RBF), sigmoid, and cosine. The observed relationships are consistent across PCA types. WE here show projections from linear PCA and from RBF and cosine kernel PCA, selected because they capture different nonlinear behaviours. This is based on the kernels' differences in calculating the similarity of two points: the RBF kernel estimates the likelihood of two points sharing a Gaussian curve, while the cosine kernel measures similarity based on the angle between vectors (Schölkopf et al., 1998; Liu et al., 2004). The RBF kernel's parameter γ was set to its scikit-learn (Python library) default value of 1 / (number of features). Lower γ values yield projections similar to linear PCA, whereas higher γ values obscure the dataset's global structure. For a more detailed explanation of PCA and kernel PCA, see the Supplemental Material.

By plotting PC1 against PC2 and mapping the relative density of segments with maximum magnitudes (M_w) $\geq$ 8.5 as contours, we infer the possibility of giant earthquake

117 occurrence on any given margin despite the incomplete earthquake records. We observe that
118 using a maximum magnitude cutoff of 8 does not change our conclusions (see Fig. S7).
119 The code used is available at github.com/gems-val22/subduction_data_analytics. The
120 Supplemental Material includes pair plots and parallel coordinate plots for PC1-PC4 (Figs.
121 S2-S6) and variance explained per PC (Fig. S8).
122 **MARGIN FAMILIES IN THE PC SPACE**
123 Fig. 2 shows projections generated with linear PCA and with RBF and cosine kernel
124 PCA. We observe that in all projections, segments from the same margin cluster closely in the
125 PC space (Fig. 2A-2C), indicating that individual margins have distinct combinations of the
126 properties considered. These margin clusters appear to form "meta clusters" (broader clusters
127 including several margin clusters), which we interpret as "margin families" characterized by a
128 similar combination of properties. We calculate 2-D Wasserstein distances (Wasserstein,
129 1969) between each pair of margins to quantify proximity.
130 For example, the Cascadia (northwestern North America), Hikurangi (New Zealand),
131 and Nankai-Ryukyu (Japan) clusters belong to the same margin family, with similar PCA
132 profiles and low Wasserstein distances (see supplemental Fig. S10-12). This similarity in
133 properties is reflected in their seismic behaviour: all three margins show slow slip (e.g.
134 Rogers & Dragert, 2003; Douglas et al., 2005; Nishimura et al., 2013), and historical or
135 paleoseismic observations indicate occasional giant earthquake occurrence (Satake et al.,
136 1996; Clark et al., 2019; Fujiwara et al., 2020).
137 A distinct margin family at high PC1 values includes the Mariana (western Pacific
138 Ocean), Solomon, Vanuatu, and Tonga-Kermadec (southwestern Pacific Ocean) margins (Fig.
139 2 B-2C). We note these margins have oceanic overriding plates and no large island arcs. Ide
140 (2013) described them as "active and moderate", characterised by frequent rupture in
141 moderate-magnitude earthquakes. In contrast, "quiet and extreme" margins (at low PC1

values) show low background seismicity but occasionally rupture in high-magnitude earthquakes (Ide, 2013). This can be seen in Fig. 2G-2I, where "quiet and extreme" clusters contain most segments with assigned maximum magnitudes ≥ 8.5 (“extreme”) and < 4 (“quiet”). This supports the idea that margins capable of hosting giant earthquakes may experience prolonged quiescence, with the plate boundary locked as strain accumulates.

This delineation of “active and moderate” from “quiet and extreme” margins in PCA projections suggests that the difference in seismic behaviour is related to a combination of the four assessed properties (sediment thickness, roughness, dip angle, relative plate velocity). Density contour plots (Fig. 2D-2F) show that kernel PCA projections separate margin families into individual clusters more effectively than PCA projections, suggesting the relationship between a combination of margin properties and seismic behaviour is non-linear.

We use Adjusted Rand Index (ARI) scores (Hubert & Arabie, 1985) to quantify similarity between Kernel- and linear PCA projections’ clusters (ARI = 100% for identical projections). ARI scores of 57.8% (RBF kernel compared to linear) and 47.0% (cosine kernel compared to linear) confirm that the kernel methods capture different behaviours within the margin property dataset than linear PCA.

The consistent grouping of margin families across projections underscores the robustness of this fingerprinting approach. Given that the non-linear methods are particularly successful in distinguishing between margin families (Fig. 2), it appears that connections between margin properties and seismicity are nonlinear.

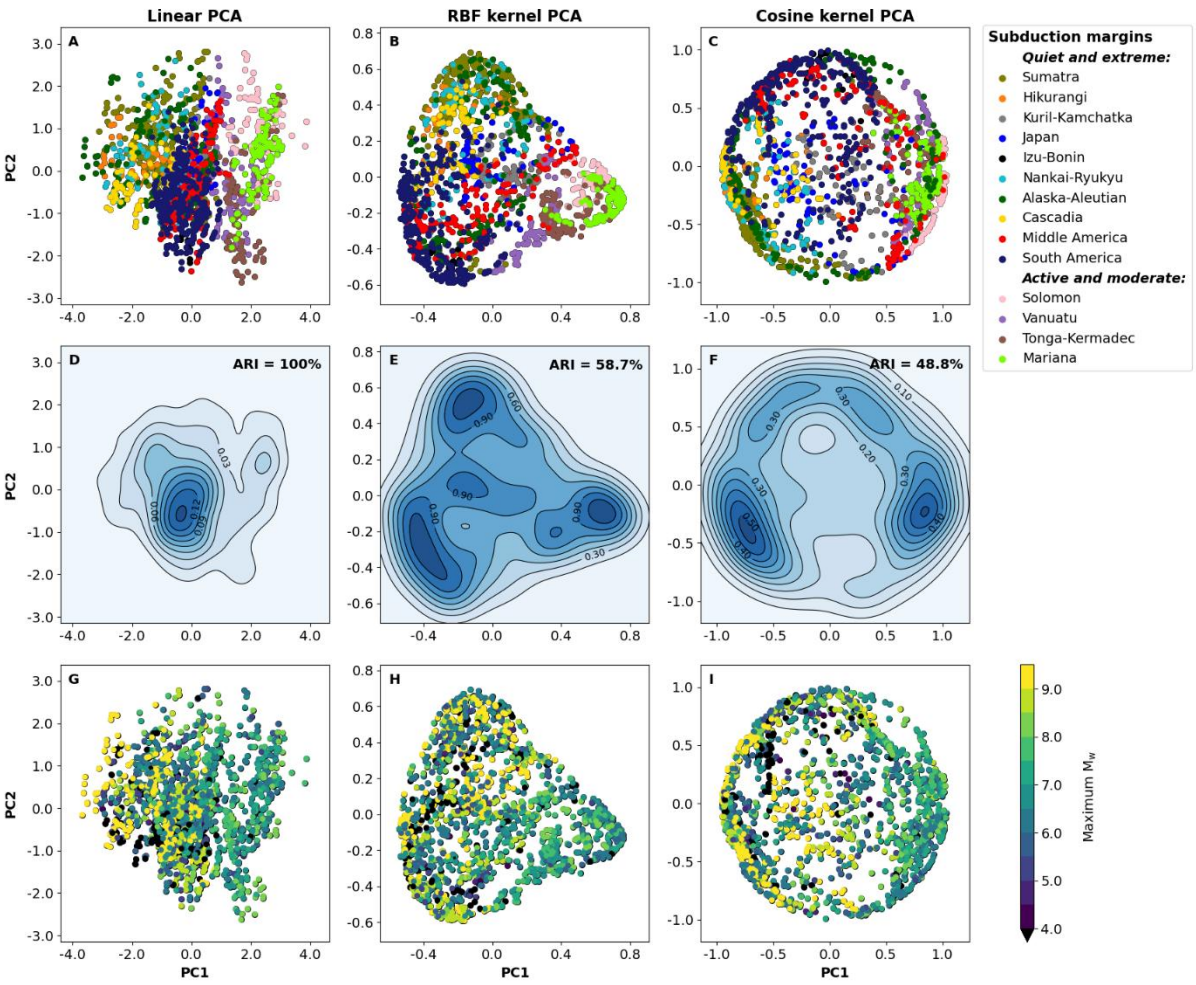


Figure 2: Principal component analysis (PCA) projections generated with linear PCA and with radial basis function (RBF) and cosine kernel PCA. **(A-C)** Projected subduction segments by subduction margin. **(D-F)** Density distribution of segments in the principal component (PC) space and adjusted Rand index (ARI) scores (Hubert & Arabie, 1985). Darker contours indicate higher density. **(G-I)** Projected segments by maximum observed magnitude.

IDENTIFYING MARGINS PRONE TO GIANT EARTHQUAKES

Having identified margin families with different properties and seismicity in the PC space, we utilise PCA profiles to assess the possibility of giant earthquake occurrence for margins with potentially underestimated maximum magnitudes. We consider a margin

capable of generating giant earthquakes if its PCA profile is similar to those of segments with high maximum magnitudes, indicating similar margin property combinations. Figure 3 shows the PC space distributions of high-maximum-magnitude segments as a density map, to which we compare the PCA profiles of margins with no recorded post-1900 $M_w \geq 8.5$ earthquakes (Hikurangi, Nankai-Ryukyu, Middle America [eastern Pacific Ocean], Solomon, Vanuatu, Mariana, Tonga-Kermadec, Izu-Bonin [western Pacific Ocean], and Cascadia). Of these, segments from the Hikurangi, Cascadia, Nankai-Ryukyu, Izu-Bonin, and Middle America margins plot in high-density regions, suggesting they may rupture in giant earthquakes. Historical and paleoseismic evidence for large earthquakes along the Cascadia (Satake et al., 1996), Hikurangi (Clark et al., 2019) and Nankai-Ryukyu (Fujiwara et al., 2020) trenches supports this assessment.

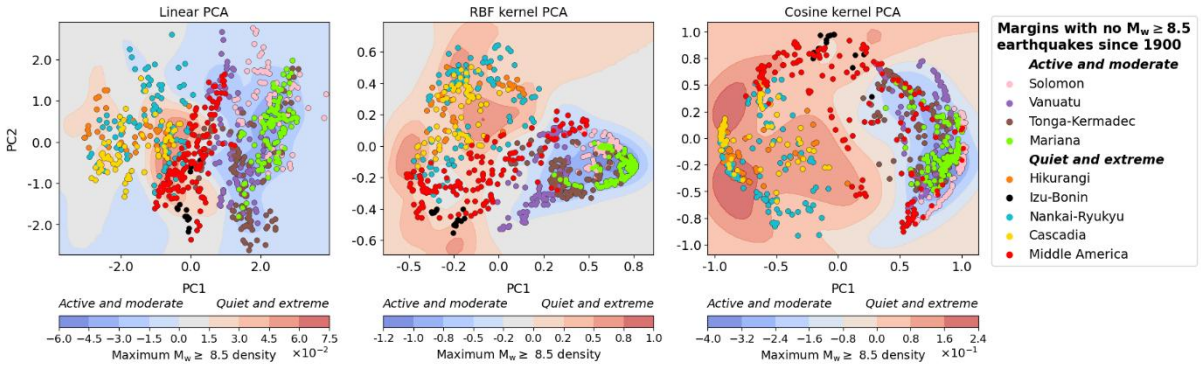


Figure 3: Principal component (PC) space relative density maps of subduction segments with assigned maximum magnitudes $M_w \geq 8.5$ (filled-in contours) and segment distributions for margins with no giant earthquakes in the instrumental record (scatter points) for projections generated using linear principal component analysis (PCA), and using radial basis function (RBF), and cosine kernel PCA.

HOW DO MARGIN PROPERTIES INFLUENCE SEISMICITY?

Fig. 2 and 3 show that a (nonlinear) combination of margin properties can distinguish between different seismic behaviours. To examine the individual margin properties' influence on the PCA projections, we calculate feature contributions for linear PCA (Fig. 4A) and plot the individual properties' distributions in the PC space for cosine kernel PCA (Fig. 4B-4E). Due to the inherent nonlinearity of kernel PCA, feature contributions cannot be calculated as for linear PCA. We cannot account for uncertainties because neither the McLellan and Audet (2020) dataset nor the underlying data include them.

Fig. 4 shows some general trends: "quiet and extreme" segments (at low PC1 values) appear to have a combination of high sediment thickness, shallow dip angles, slow relative plate velocities, and low roughness, while "active and moderate" segments are characterised by a combination of lower sediment thickness, steeper dip angles, faster convergence, and varying roughness. Observations of links between individual properties and seismicity supporting this assessment include Heuret et al. (2012) and Scholl et al. (2015) for sediment thickness, van Rijsingen et al. (2018) for roughness, and Muldashev and Sobolev (2020) for sediment thickness and dip angle. Physical explanations for these observations have been proposed: high coupling at smooth plate interfaces (from high sediment availability and low basement roughness; e.g., Ruff, 1989; van Rijsingen et al., 2018), low pore fluid pressure at seismogenic depth (due to shallow dip angles limiting bending-related fracturing of the subducting plate, and/or a thick sediment layer acting as a barrier to upwards fluid migration; e.g., Nishikawa and Ide, 2015; Seno, 2017), and a wide seismogenic zone (at shallow dip angles; e.g., Muldashev and Sobolev, 2020) may enable giant earthquakes. While additional properties may influence seismicity, we demonstrate here that our unsupervised approach, using this property set, provides valuable insights into margin seismicity.

Given the nonlinear connections between combinations of margin properties and seismic behaviour found in this work, we suggest abandonment of the search for linear correlations between individual properties and instead encourage study of margins through their PCA profiles. Our findings could be further developed with the establishment of global datasets including additional subduction properties, such as fluid presence or different-wavelength roughness, along with their uncertainties.

A Table: Variance explained and feature contributions for linear PCA.

Linear PCA	Variance explained	Sediment thickness	Dip angle	Relative plate velocity	Roughness
PC1	47.8%	-0.505	0.584	0.317	0.551
PC2	25.2%	0.523	-0.138	0.828	0.149

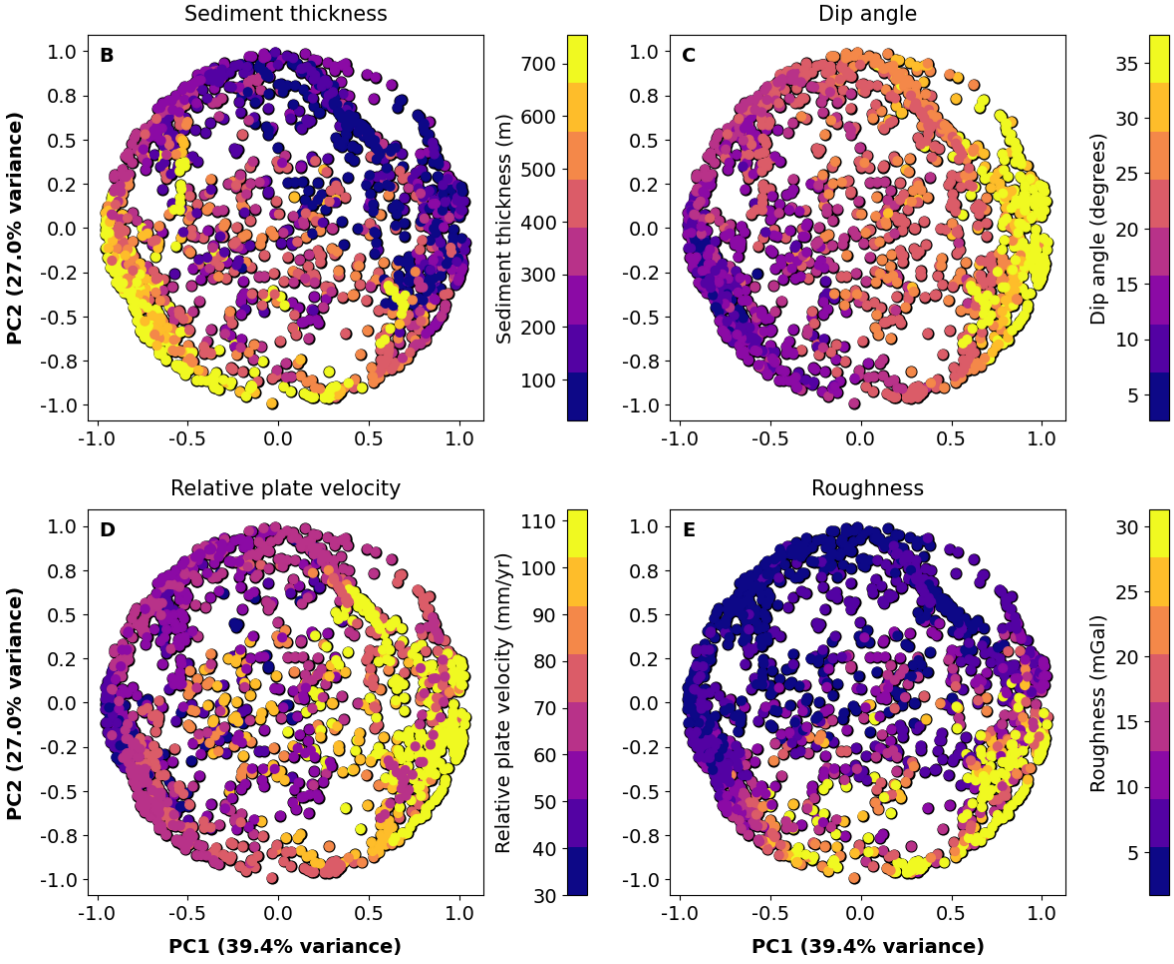


Figure 4: (A). Explained variance and feature importance for linear principal component analysis (PCA) projection. **(B-E)** Property distributions in principal component (PC) space for cosine kernel PCA.

CONCLUSION

We explore the relationship between four margin properties – the subducting plate's dip angle, roughness, sediment thickness, and convergence rate – and maximum earthquake magnitudes using (kernel) PCA projections. This approach avoids bias from underestimated maximum magnitudes resulting from the short earthquake record.

We observe a distinct cluster of low-maximum magnitude segments in the PC space and suggest it represents "active and moderate" margins, contrasted by "quiet and extreme" margins (as described by Ide, 2013). Using PCA projections to identify margins prone to giant earthquakes with no precedent in the instrumental record, we highlight the Hikurangi, Cascadia, Nankai-Ryukyu, Izu-Bonin, and Middle America margins.

Based on these (kernel) PCA projections, we suggest that seismic behaviour at subduction margins is describable as a nonlinear combination of margin properties. We find that a combination of high sediment thickness, low dip angles, low roughness, and low relative plate velocity is generally associated with higher maximum magnitudes, while the opposite is generally true for lower maximum magnitudes. We argue that shifting from searching for linear connections to studying margins in terms of their PCA profile opens new pathways for understanding subduction margin seismicity.

ACKNOWLEDGEMENTS

This work is part of Locher's M.S. and Ph.D. studies at Imperial College London, funded by the Department of Earth Science and Engineering's computational science and engineering scholarship for women. The code utilizes the Python library scikit-learn. The authors thank Xiaodong Yang, Alex Whittaker, Saskia Goes, and three anonymous reviewers for their valuable feedback.

[1]Supplemental Material. Extended explication of (kernel) PCA, supplementary figures and tables. Please visit https://doi.org/10.1130/GEOL.S.XXXX to access the supplemental material, and contact editing@geosociety.org with any questions.

REFERENCES

Albini, P., Musson, R., Gomez Capera, A., Locati, M., Rovida, A., Stucchi, M., & Viganò, D. (2013). Global historical earthquake archive and catalogue (1000-1903). *Pavia, Italy*.

Allen, T. I., & Hayes, G. P. (2017). Alternative rupture-scaling relationships for subduction interface and other offshore environments. *Bulletin of the Seismological Society of America*, *107*(3), 1240-1253.

Bassett, D., & Watts, A. B. (2015). Gravity anomalies, crustal structure, and seismicity at subduction zones: 1. Seafloor roughness and subducting relief. *Geochemistry, Geophysics, Geosystems*, *16*(5), 1508-1540.

Brizzi, S., Sandri, L., Funiciello, F., Corbi, F., Piromallo, C., & Heuret, A. (2018). Multivariate statistical analysis to investigate the subduction zone parameters favoring the occurrence of giant megathrust earthquakes. *Tectonophysics*, *728*, 92-103.

Clark, K., Howarth, J., Litchfield, N., Cochran, U., Turnbull, J., Dowling, L., Howell, A., Berryman, K., & Wolfe, F. (2019). Geological evidence for past large earthquakes and tsunamis along the Hikurangi subduction margin, New Zealand. *Marine Geology*, *412*, 139-172.

Douglas, A., Beavan, J., Wallace, L., & Townend, J. (2005). Slow slip on the northern Hikurangi subduction interface, New Zealand. *Geophysical Research Letters*, *32*(16).

Fujiwara, O., Goto, K., Ando, R., & Garrett, E. (2020). Paleotsunami research along the Nankai Trough and Ryukyu Trench subduction zones–current achievements and future challenges. *Earth-Science Reviews*, *210*, 103333.

Heuret, A., Conrad, C., Funiciello, F., Lallemand, S., & Sandri, L. (2012). Relation between subduction megathrust earthquakes, trench sediment thickness and upper plate strain. *Geophysical Research Letters*, *39*(5).

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, *2*, 193-218.

Ide, S. (2013). The proportionality between relative plate velocity and seismicity in subduction zones. *Nature Geoscience*, *6*(9), 780-784.

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Lallemand, S., Peyret, M., van Rijsingen, E., Arcay, D., & Heuret, A. (2018). Roughness characteristics of oceanic seafloor prior to subduction in relation to the seismogenic potential of subduction zones. *Geochemistry, Geophysics, Geosystems*, *19*(7), 2121-2146.

Liu, Q., Lu, H., & Ma, S. (2004). Improving kernel Fisher discriminant analysis for face recognition. *IEEE transactions on circuits and systems for video technology*, *14*(1), 42-49.

McCaffrey, R. (2008). Global frequency of magnitude 9 earthquakes. *Geology*, *36*(3), 263-266.

McLellan, M., & Audet, P. (2020). Uncovering the physical controls of deep subduction zone slow slip using supervised classification of subducting plate features. *Geophysical Journal International*, *223*(1), 94-110.

Muldashev, I. A., & Sobolev, S. V. (2020). What controls maximum magnitudes of giant subduction earthquakes? *Geochemistry, Geophysics, Geosystems*, *21*(9), e2020GC009145.

Nakao, A., Kuwatani, T., Ueki, K., Yoshida, K., Yutani, T., Hino, H., & Akaho, S. (2023). Regression analysis and variable selection to determine the key subduction-zone parameters that determine the maximum earthquake magnitude. *Earth, Planets and Space*, *75*(1), 1-12.

Nishikawa, T., & Ide, S. (2015). Background seismicity rate at subduction zones linked to slab-bending-related hydration. *Geophysical Research Letters*, *42*(17), 7081-7089.

Nishimura, T., Matsuzawa, T., & Obara, K. (2013). Detection of short-term slow slip events along the Nankai Trough, southwest Japan, using GNSS data. *Journal of Geophysical Research: Solid Earth*, *118*(6), 3112-3125.

Rogers, G., & Dragert, H. (2003). Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. *Science*, *300*(5627), 1942-1943.

Ruff, L., & Kanamori, H. (1980). Seismicity and the subduction process. *Physics of the Earth and Planetary interiors*, *23*(3), 240-252.

Ruff, L. J. (1989). Do trench sediments affect great earthquake occurrence in subduction zones? *Subduction Zones Part II*, 263-282.

Satake, K., Shimazaki, K., Tsuji, Y., & Ueda, K. (1996). Time and size of a giant earthquake in Cascadia inferred from Japanese tsunami records of January 1700. *Nature*, *379*(6562), 246-249.

Schellart, W. P., & Rawlinson, N. (2013). Global correlations between maximum magnitudes of subduction zone interface thrust earthquakes and physical parameters of subduction zones. *Physics of the Earth and Planetary Interiors*, *225*, 41-67.

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, *10*(5), 1299-1319.

Scholl, D. W., Kirby, S. H., von Huene, R., Ryan, H., Wells, R. E., & Geist, E. L. (2015). Great (≥ Mw8. 0) megathrust earthquakes and the subduction of excess sediment and bathymetrically smooth seafloor. *Geosphere*, *11*(2), 236-265.

Seno, T. (2017). Subducted sediment thickness and Mw 9 earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*(1), 470-491.

Smith, M. W. (2014). Roughness in the earth sciences. *Earth-Science Reviews*, *136*, 202-225.

Stein, S., & Okal, E. A. (2007). Ultralong period seismic study of the December 2004 Indian Ocean earthquake and implications for regional tectonics and the subduction process. *Bulletin of the Seismological Society of America*, *97*(1A), S279-S295.

Stein, S., & Okal, E. A. (2011). The size of the 2011 Tohoku earthquake need not have been a surprise. *Eos, Transactions American Geophysical Union*, *92*(27), 227-228.

U.S. Geological Survey. (1900 - 2023). *Earthquake Catalog*. Retrieved August 09, 2023 from https://earthquake.usgs.gov/earthquakes/search/

van Rijsingen, E., Lallemand, S., Peyret, M., Arcay, D., Heuret, A., Funiciello, F., & Corbi, F. (2018). How subduction interface roughness influences the occurrence of large interplate earthquakes. *Geochemistry, Geophysics, Geosystems*, *19*(8), 2342-2370.

Wallace, L. M., Reyners, M., Cochran, U., Bannister, S., Barnes, P. M., Berryman, K., Downes, G., Eberhart-Phillips, D., Fagereng, A., & Ellis, S. (2009). Characterizing the seismogenic zone of a major plate boundary subduction thrust: Hikurangi Margin, New Zealand. *Geochemistry, Geophysics, Geosystems*, *10*(10).

Wasserstein, L. N. (1969). Markov processes on countable product space describing large systems of automata. *Probl. Pered. Inform*, *5*, 64-73.

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, *84*(4), 974-1002.

Wirth, E. A., Sahakian, V. J., Wallace, L. M., & Melnick, D. (2022). The occurrence and hazards of great subduction zone earthquakes. *Nature Reviews Earth & Environment*, *3*(2), 125-140.

Supplemental Material

1. PCA and Kernel-PCA

Principal Component Analysis (PCA) (e.g. Jolliffe, 2002) is a method to project data into a different space, where the axes are the "Principal Components" (PCs), which represent the directions of the highest variance in a dataset. PCA calculates as many PCs as there are features (dimensions) in the original dataset and re-orients the data points into this new space. PCs are numbered in descending order of the proportion of variance of the original data distribution they capture, i.e. PC1 captures the most variance, PC2 the second most, etc.

In **Kernel-PCA** (Schölkopf et al., 1998), the data is projected into a higher-dimensional space using a kernel function before finding the PCs and reorienting the data points into the new space. This can allow for linear separation between clusters which are inseparable in the original space. For instance, considering two-dimensional data points, we can apply a function to add a third dimension: a point (x, y) will be transformed to (x, y, z) where $z = \phi(\mathrm{x}, \mathrm{y})$ is calculated as a function ϕ of the original data (x, y). Using non-linear functions for ϕ, such as Gaussian, polynomial, or trigonometric functions, we can introduce non-linearity when transforming the data into the higher-dimensional space. Therefore, once the transformed, higher-dimensional data is projected into the PC space, the orientation of the PCs and data distribution may capture non-linear patterns in the original data. Fig. S1 shows a conceptual illustration of these steps.

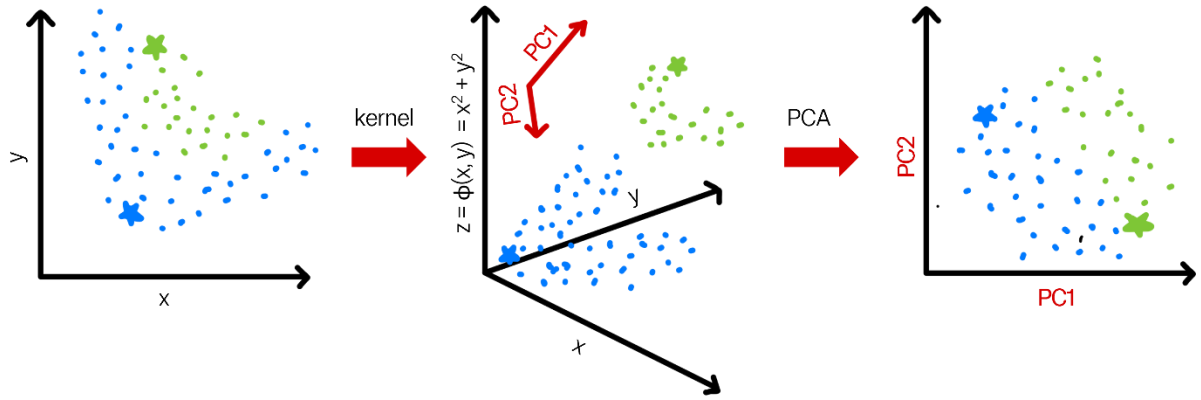


Fig. S1: Sketch illustrating the steps of Kernel-PCA. Here, the new dimension z is calculated as $\phi(\mathrm{x}, \mathrm{y}) = x^2 + y^2$.

However, transforming large, high-dimensionality datasets into even higher-dimensional spaces is computationally expensive. Instead, the "kernel trick" is used to avoid having to transform each data point by instead calculating the similarity between each combination of two data points. Kernels are functions that measure the similarity between two points x and y (described as vectors), for example:

- **Radial Basis Function (RBF) kernel (Gaussian):** $k(x, y) = \exp\left(\frac{||x-y||^2}{2\sigma^2}\right)$ where σ is a free parameter defining the width of the Gaussian curve; in scikit-learn, it is expressed as γ where $\gamma = \frac{1}{2\sigma^2}$
- **Cosine kernel** (Liu et al., 2004): $k'(x, y) = \frac{k(x,y)}{\sqrt{k(x,x)k(y,y)}}$ where k is a polynomial kernel
- **Polynomial kernel:** $k(x, y) = (x \cdot y)^d$ where d is the degree of the polynomial

The difference between these kernels is that they use different methods to calculate the similarity between two points. For instance, the RBF kernel calculates the similarity as the likelihood of a point x belonging to a Gaussian curve centered around another point y. Applying a cosine kernel is

comparable to calculating the cosine of the angle between the two points' vectors (Liu et al., 2004). Thus, for the cosine kernel's calculation of similarity, the direction in which points lie is more important than the Euclidean distance between them.

In conclusion, Kernel-PCA allows us to look for non-linear patterns in datasets, and that different kernels result in different PCA projections owing to the similarity calculation used.

2. Data Scaling

Scaling or standardizing data prior to PCA is critical. We use scikit-learn's built-in scaler functions, which are defined as:

- StandardScaler: $x_{scaled} = \frac{x - mean}{standard\ deviation}$
- RobustScaler: $x_{scaled} = \frac{x - median}{interquantile\ range}$ where the interquantile range describes the difference between the 25th and 75th quantiles
- MinMaxScaler: $x_{scaled} = \frac{x - minimum}{maximum - minimum}$, where the maximum and minimum are the largest and smallest feature values

For more information, see the documentation of scikit-learn scalers.

3. Supplemental figures and tables

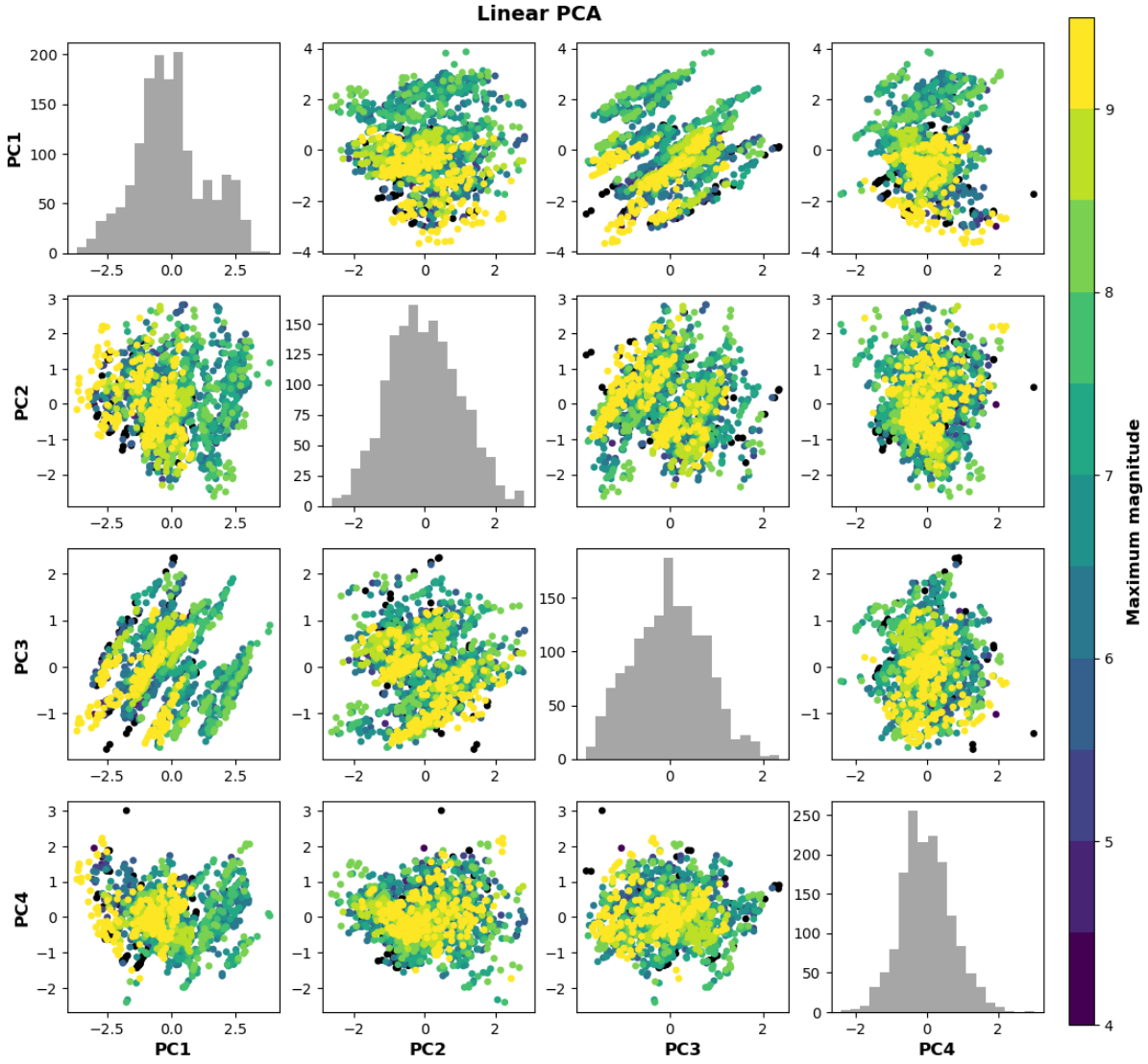


Fig. S2: Full pair plots for linear PCA, colour coded by maximum magnitude.

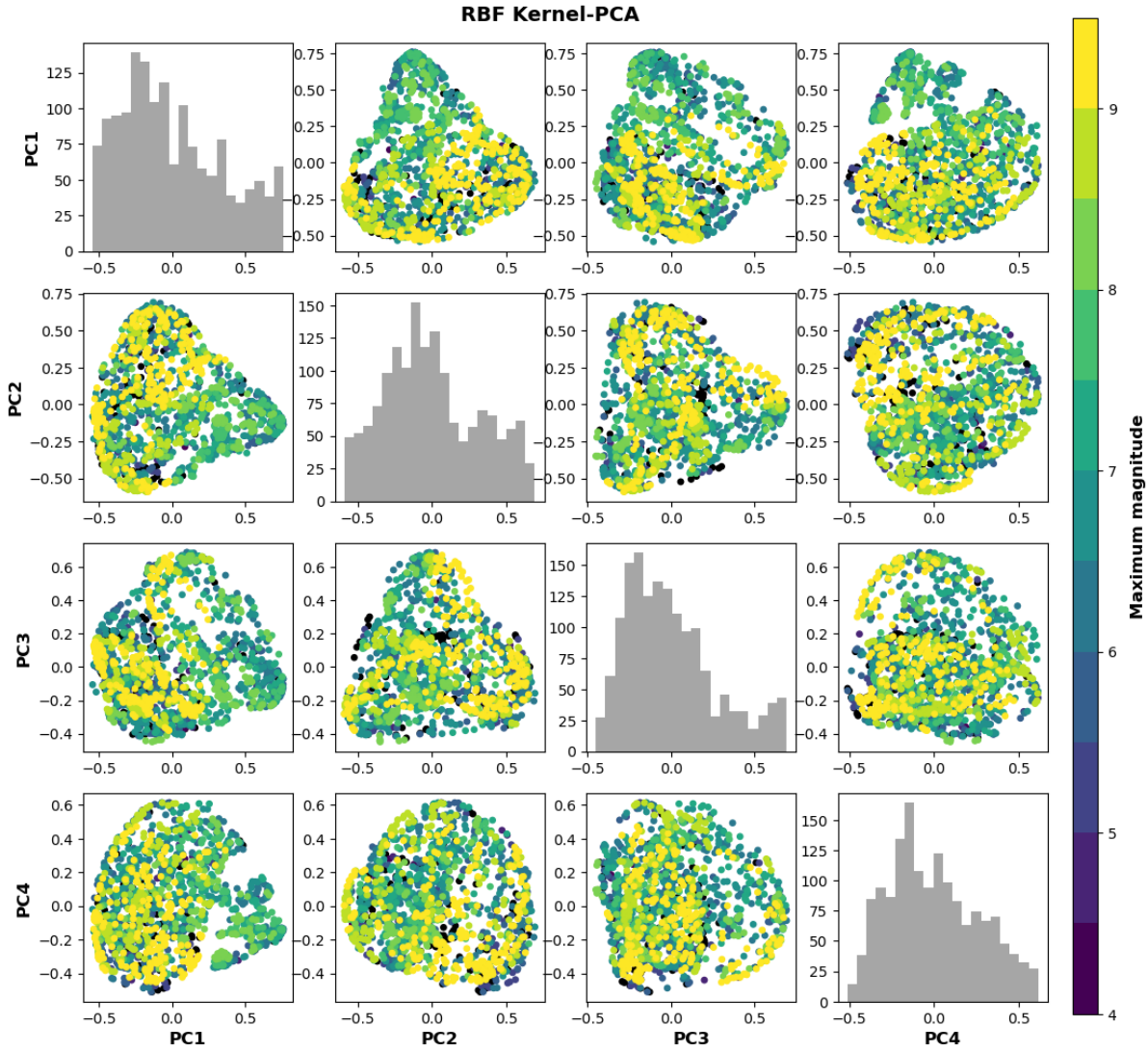


Fig. S3: Pair plots of PC1 through PC4 for RBF Kernel-PCA, colour coded by maximum magnitude.

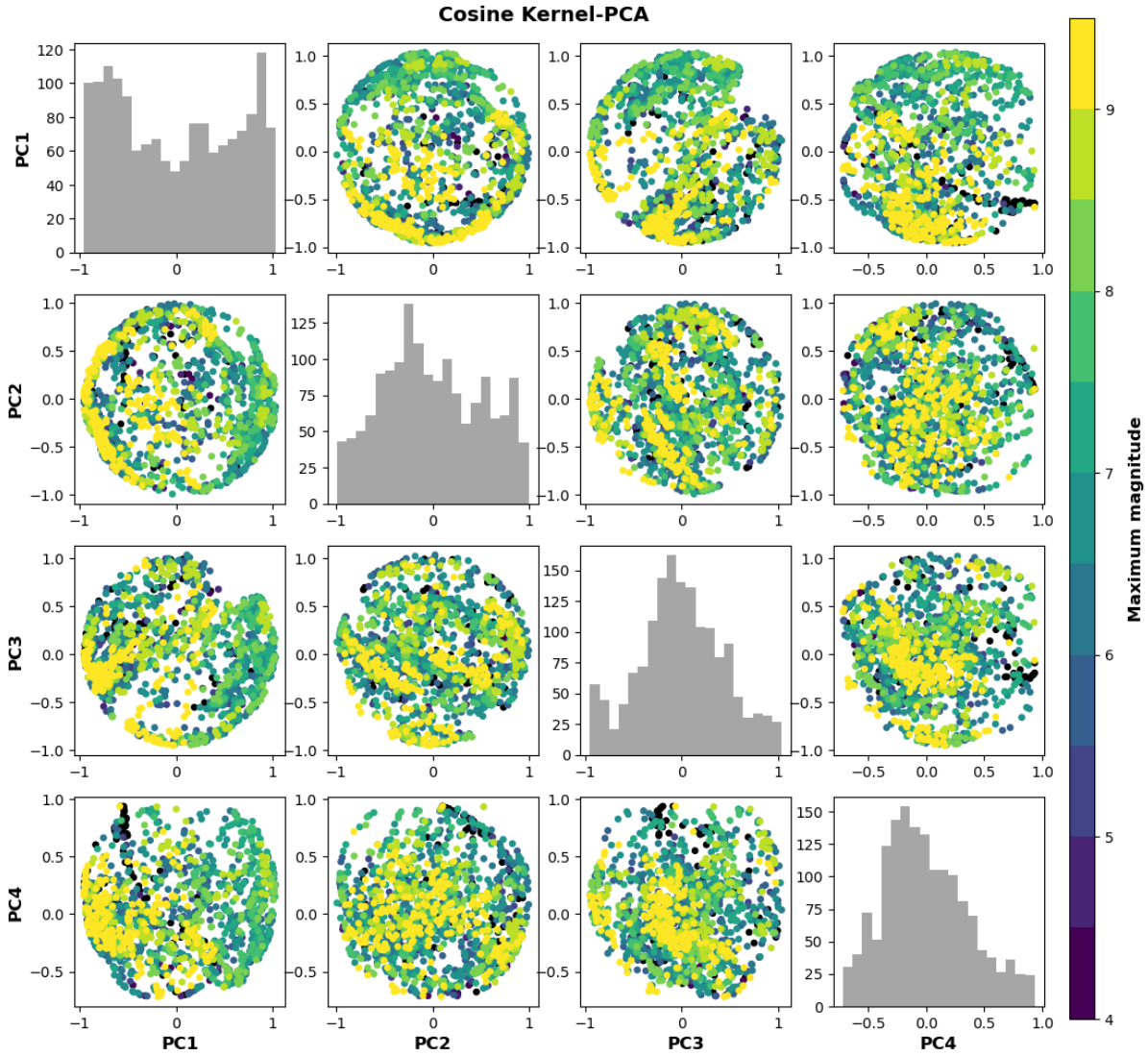


Fig. S4: Full pair plots for cosine Kernel-PCA, colour coded by maximum magnitude.

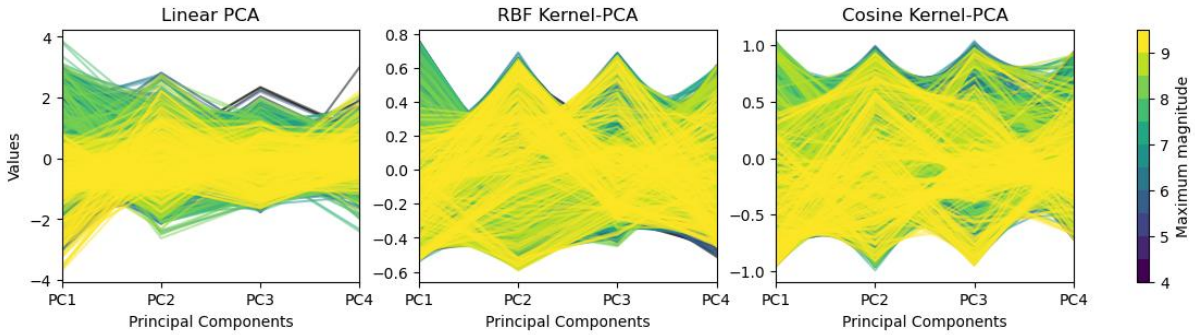
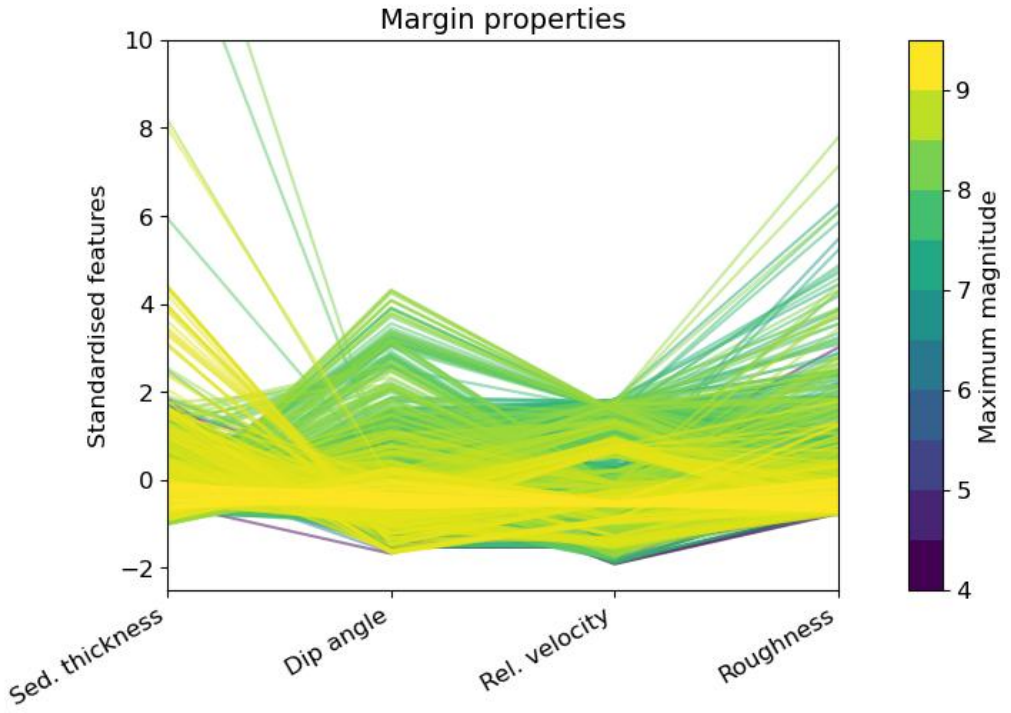


Fig. S5: Parallel coordinates plot for PC1 through PC4 for linear, RBF Kernel- and cosine Kernel-PCA.



401

402 **Fig. S6:** Parallel coordinates plot the original (standardized) margin properties.

403

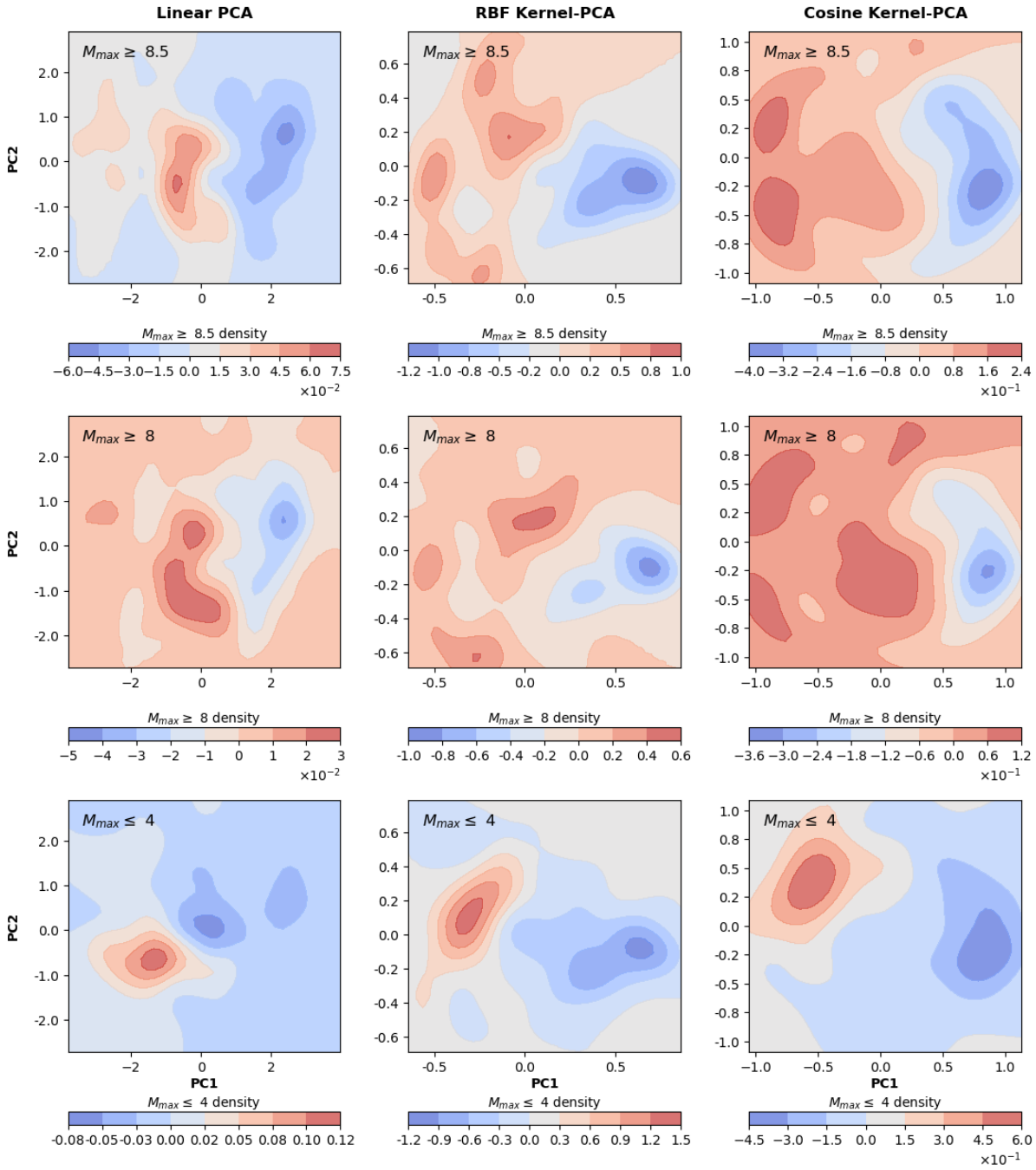
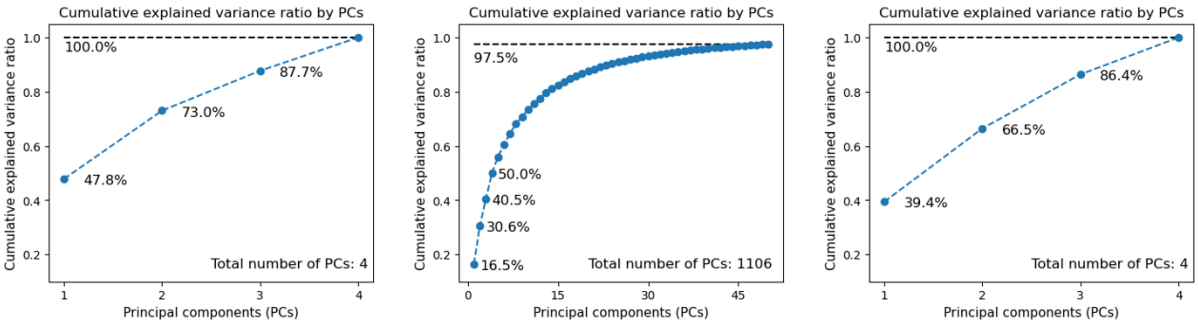


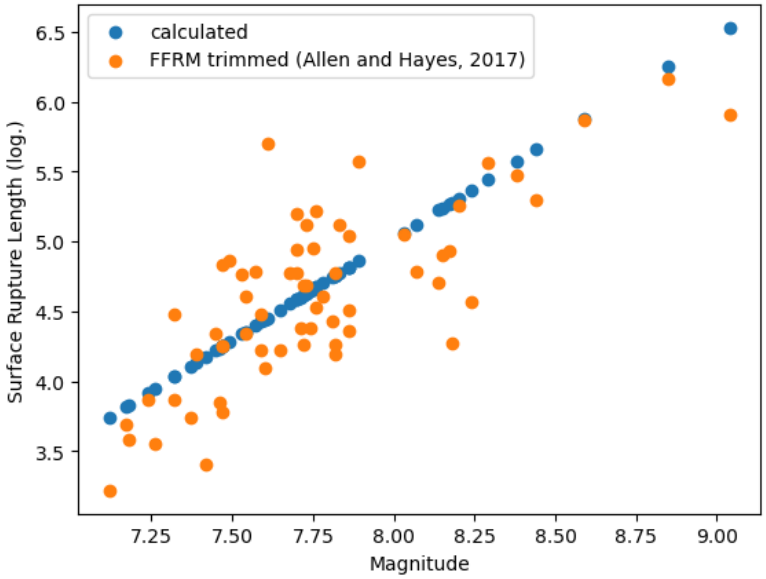
Fig. S7: Density difference maps for different magnitude cutoffs (≥ 8.5, ≥8, and ≤4) for linear PCA, RBF and Cosine Kernel-PCA. This illustrates a) that changing the magnitude threshold from 8.5 to 8 results in the same pattern, demonstrating our method is robust to this change, and b) that segments of maximum magnitude ≤ 4 ("quiet") project in the same area as segments of high (≥ 8.5 or ≥8) maximum magnitude ("extreme").



411

412 **Fig. S8:** Cumulative variance explained from linear PCA, RBF and cosine Kernel-PCA.

413



414

415 **Fig. S9:** Surface rupture lengths (logarithmic scale) as estimated from earthquake magnitudes (Wells
416 & Coppersmith, 1994) in our approach, compared to their corresponding finite fault rupture models
417 (using data compiled by Allen & Hayes, 2017). This shows the Wells and Coppersmith estimates are a
418 reasonable estimate for surface rupture length, validating our method.

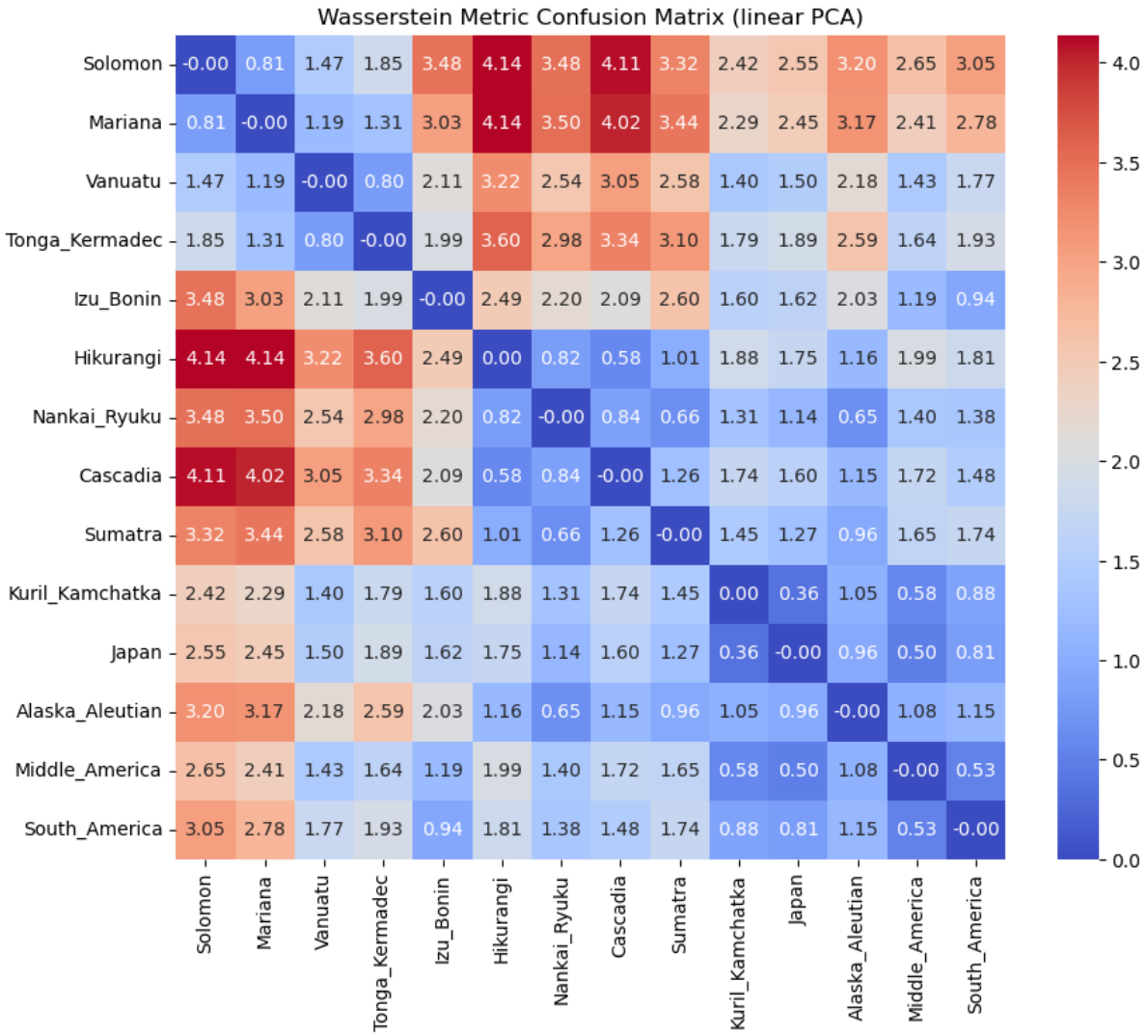


Fig. S10: Wasserstein distances calculated for margin clusters in the linear PCA projection. Low Wasserstein distances indicate that the PC space clusters of the two margins are close, reflecting similar PCA profiles and margin properties. In contrast, high Wasserstein distances suggest the clusters are far apart, indicating the two margins have distinct PCA profiles and property combinations.

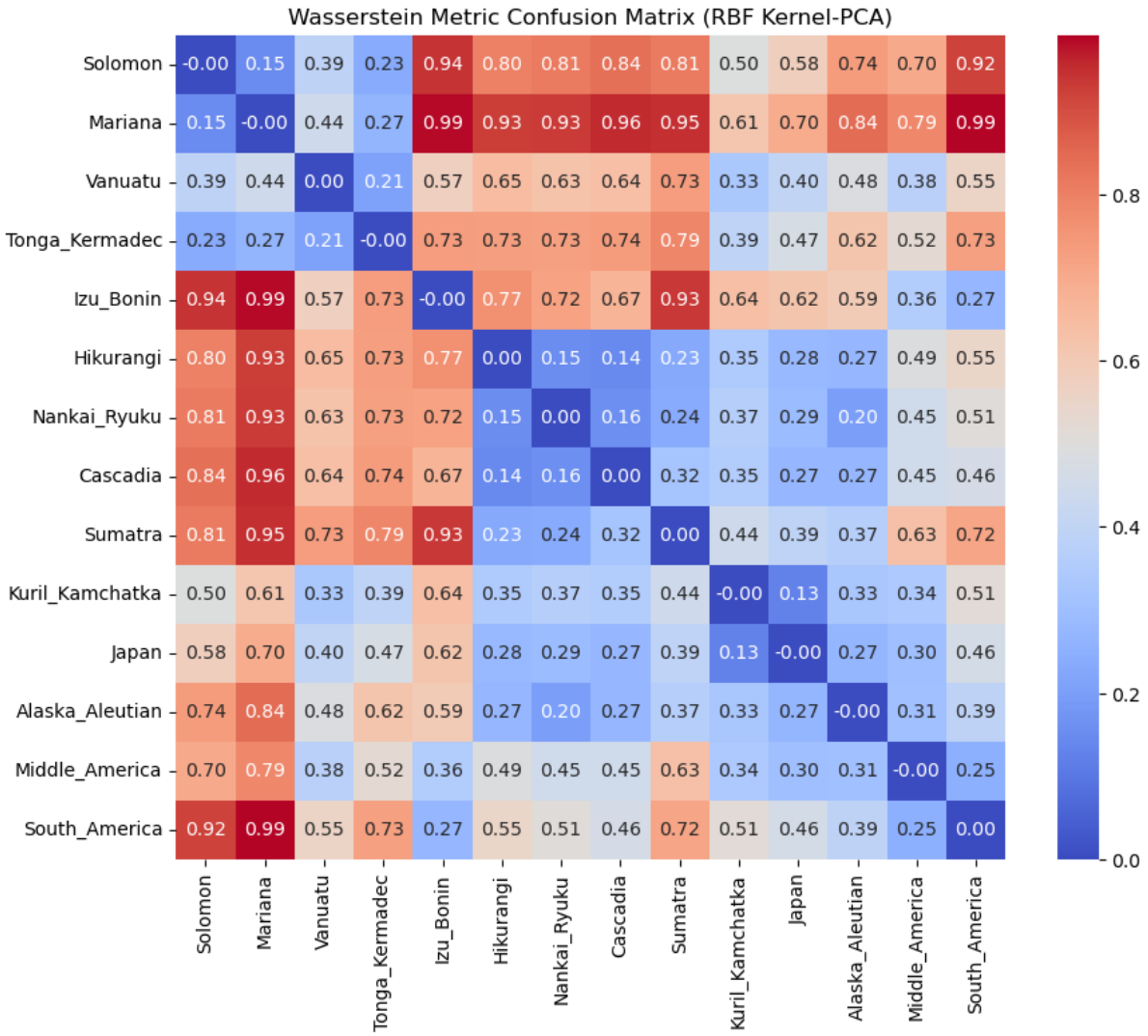


Fig. S11: Wasserstein distances calculated for margin clusters in the RBF Kernel-PCA projection. Low Wasserstein distances indicate that the PC space clusters of the two margins are close, reflecting similar PCA profiles and margin properties. In contrast, high Wasserstein distances suggest the clusters are far apart, indicating the two margins have distinct PCA profiles and property combinations.

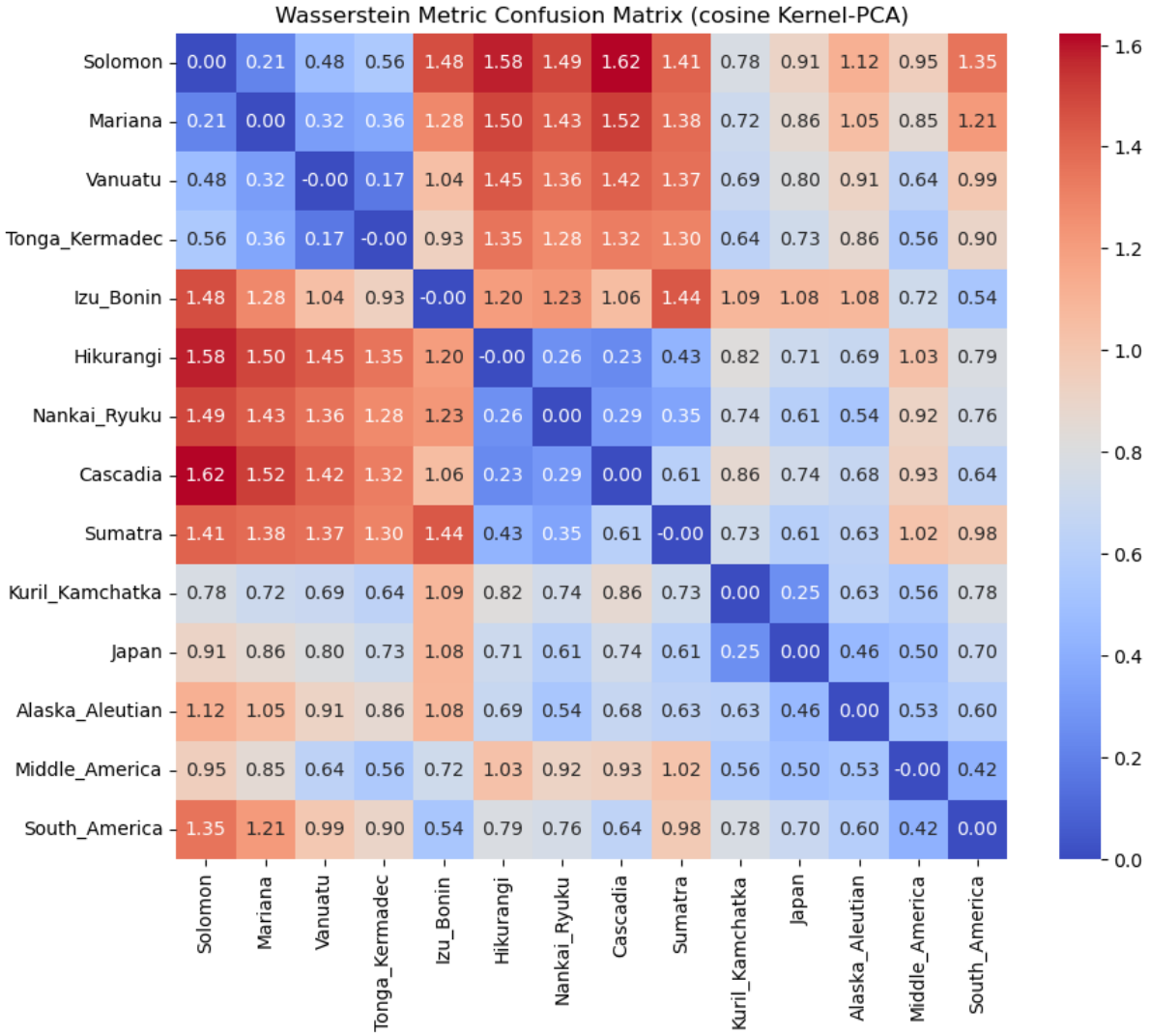


Fig. S12: Wasserstein distances calculated for margin clusters in the cosine Kernel-PCA projection. Low Wasserstein distances indicate that the PC space clusters of the two margins are close, reflecting similar PCA profiles and margin properties. In contrast, high Wasserstein distances suggest the clusters are far apart, indicating the two margins have distinct PCA profiles and property combinations.

References

Allen, T. I., & Hayes, G. P. (2017). Alternative rupture-scaling relationships for subduction interface and other offshore environments. *Bulletin of the Seismological Society of America*, *107*(3), 1240-1253.

Jolliffe, I. T. (2002). *Principal component analysis for special types of data*. Springer.

Liu, Q., Lu, H., & Ma, S. (2004). Improving kernel Fisher discriminant analysis for face recognition. *IEEE transactions on circuits and systems for video technology*, *14*(1), 42-49.

Schölkopf, B., Smola, A., & Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, *10*(5), 1299-1319.

Wells, D. L., & Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America*, *84*(4), 974-1002.