Supporting Information for "High-angle orientations of active conjugate faults in the Anza-Borrego shear zone, Southern California"

Xiaoyu Zou[1], Yuri Fialko[1], Andrew Dennehy[2], Alexander Cloninger[3,4],

Shabnam J. Semnani[5]

[1]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093, USA.

[2]Committe on Computational and Applied Mathematics, University of Chicago, Chicago, IL 60615, USA.

[3]Department of Mathematical Sciences, University of California San Diego, La Jolla, CA 92093, USA.

[4]Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA 92093, USA

[5]Department of Structural Engineering, University of California San Diego, La Jolla, CA 92093, USA.

Contents of this file

Corresponding author: Xiaoyu Zou, Institute of Geophysics and Planetary Physics, University of California, San Diego, 8800 Biological Grade, La Jolla, CA, 92093. (x3zou@ucsd.edu)

Text S1. Description of LINSCAN algorithm

LINSCAN is based on OPTICS (Ordering Points to Identify the Clustering Structure), a density-based clustering algorithm. OPTICS uses three parameters for clustering, ϵ, $MinPts$, and ξ. ϵ is the maximum distance to consider when confining a cluster, $MinPts$ describes the minimum number of points required to define a cluster, and ξ determines the minimum steepness to determine the local minimum of *reachability distance*, which constitutes the boundary of a cluster. OPTICS first searches for *core points* in the data, which is defined as a point with at least $MinPts$ of points found within its neighborhood of ϵ distance. For each of the *core point*, it defines a *core distance*, which is the distance between a *core point* and its $MinPts$-th closest point. Next, OPTICS calculates the *reachability distance* among all core points. The *reachability distance* between points a and b is either the distance between a and b or the *core distance* of a, whichever is larger. Points that don't have at least $MinPts$ of points found within its neighborhood of ϵ distance won't be classified as a *core point*, so their *core distance* and *reachability distance* are undefined. Points within a cluster have a low *reachability distance* to their most adjacent neighbors, so points with local minimum *reachability distance* are identified as one cluster.

Since OPTICS uses the Euclidean metric to define clusters, the shape of clusters only depends on their density structure and is not necessarily linear. LINSCAN keeps the basic method of OPTICS but replaces the Euclidean metric with a distance function derived from Kullback-Leibler Divergence, which is the measurement of similarity between two distributions. In this way, the clustering results will be more linear. In LINSCAN we

not only keep ϵ, $MinPts$, and ξ, but also introduce two additional parameters: *ecc_pts* and *threshold*. For each point, LINSCAN approximates its *ecc_pts* nearest neighbors as a normal distribution. Then, for clustering, we define the pairwise distance between $P = \mathcal{N}(\mu_P, \Sigma_P)$ and $Q = \mathcal{N}(\mu_Q, \Sigma_Q)$ to be:

$$\begin{aligned} D(P,Q) = &\frac{1}{2}||\Sigma_Q^{-\frac{1}{2}}\Sigma_P\Sigma_Q^{-\frac{1}{2}} - I||_F \\ &+\frac{1}{2}||\Sigma_P^{-\frac{1}{2}}\Sigma_Q\Sigma_P^{-\frac{1}{2}} - I||_F \\ &+\frac{1}{\sqrt{2}}||\mu_P - \mu_Q||_{\Sigma_Q^{-1}} \\ &+\frac{1}{\sqrt{2}}||\mu_P - \mu_Q||_{\Sigma_P^{-1}} \end{aligned} \tag{1}$$

where $||A||_F$ denotes the frobenius norm of the matrix A and

$$||x||_A = \sqrt{x^T A x} \tag{2}$$

denotes the elliptic norm defined by A for vector x and matrix A. This distance can be viewed as a low-order approximation of the symmetrized KL-divergence.

The *threshold* parameter represents the minimum correlation coefficient. All clusters with correlation coefficients lower than the *threshold* won't be included in the final result.

With all parameters being set, LINSCAN will go through each point in the dataset and label it either as a member of linear clusters with a sufficient correlation coefficient or an unqualified data point.

With synthetic data, LINSCAN is proven accurate in linear clustering. After testing LINSCAN on randomly generated synthetic data with noise, non-linear clusters, and linear clusters, we found that LINSCAN can correctly identify as many as 80% true linear cluster data points (Figure S1).

References

Fialko, Y., & Jin, Z. (2021). Simple shear origin of the cross-faults ruptured in the 2019 Ridgecrest earthquake sequence. *Nature Geoscience*, *14*, 513–518.

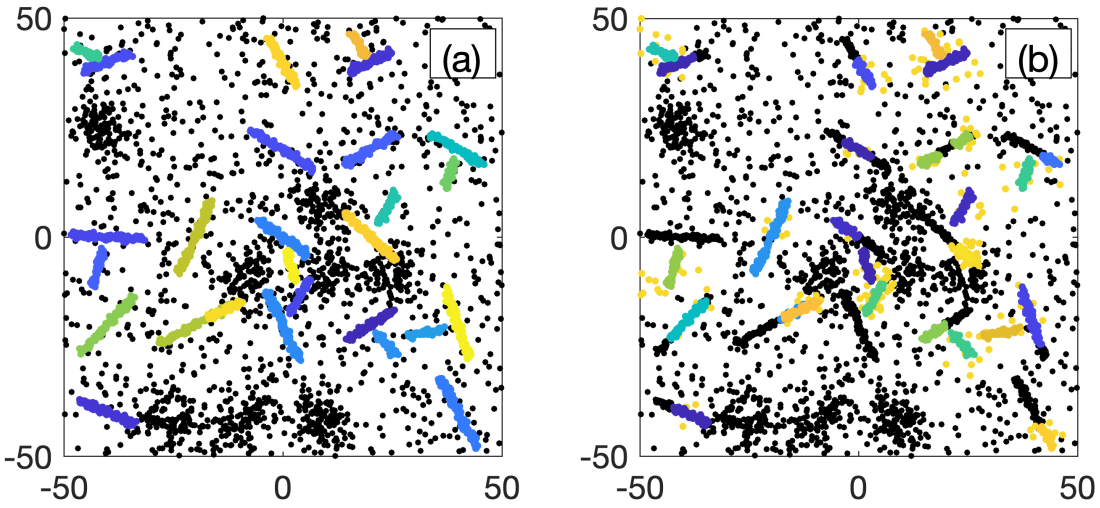

Figure S1. (a) A synthetic data set mimicking a distribution of earthquake epicenters. Black points represent "noise" (i.e., background seismicity, and irregular or quasi-isometric clusters). Linear clusters are denoted by sets of points having the same color (other than black). (b) LINSCAN classification: color (non-black) sets of points denote identified linear clusters, and black points represent the remaining data (i.e., data points that were not identified as belonging to a linear cluster).

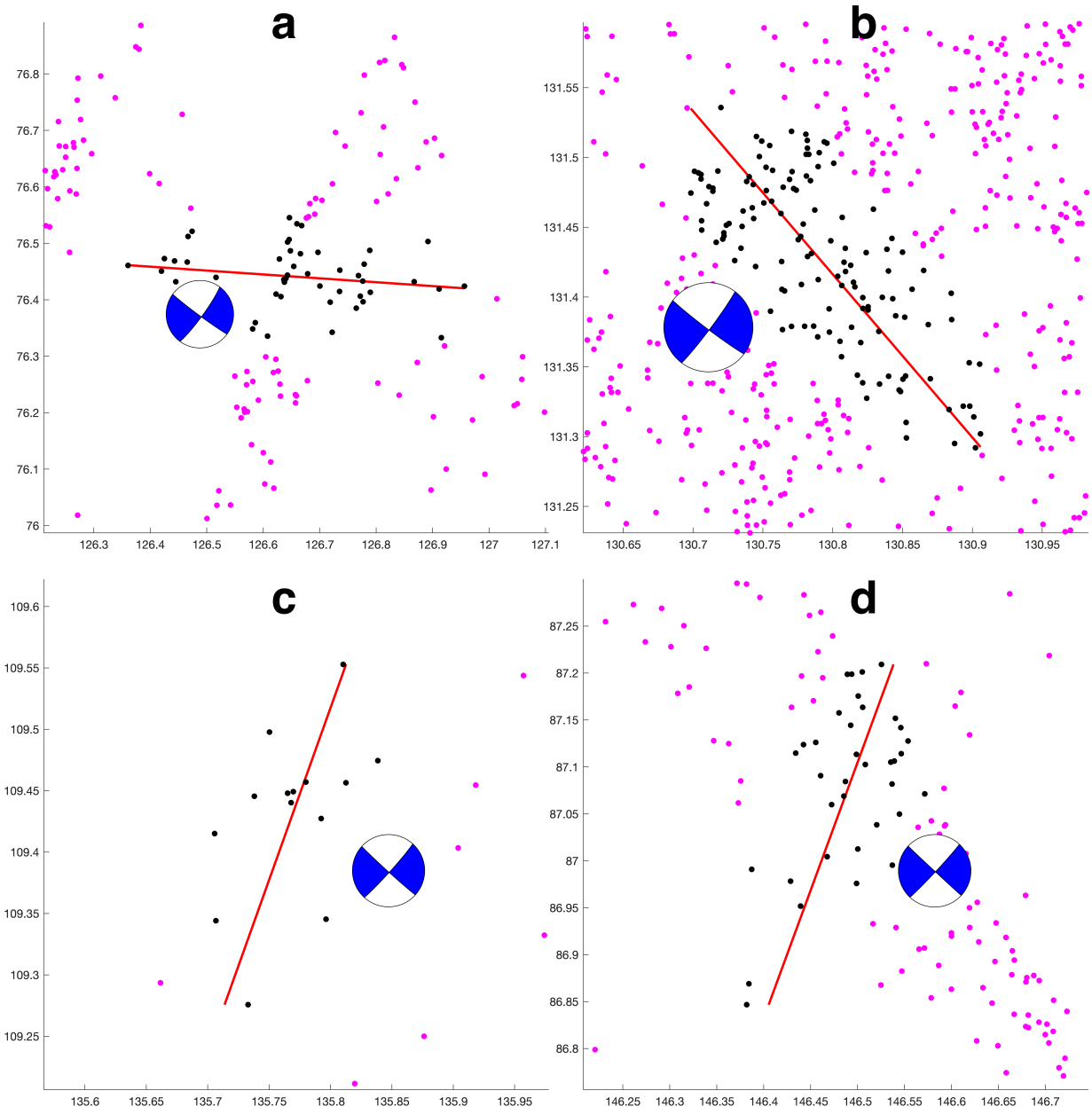

Figure S2. Examples of clusters selected by LINSCAN that failed to pass a visual quality check. Panels (a) and (d) illustrate clusters which contain smaller linear features that are not aligned with the overall trend. Panel (b) illustrates a selected cluster which is not obviously distinguishable from the background. Panel (c) illustrates a quasi-linear cluster with points that are unevenly and/or sparsely distributed.

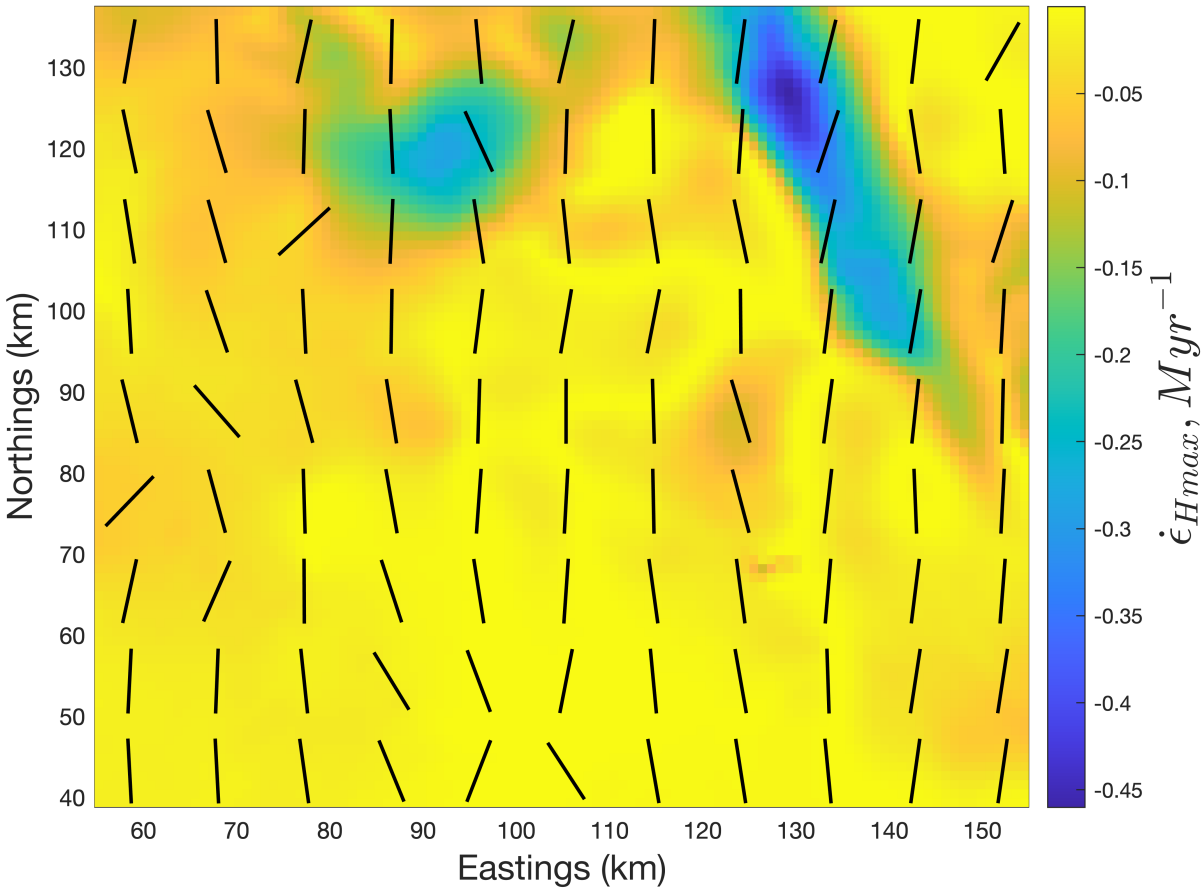

Figure S3. The magnitude (color) and orientation (black tick marks) of the maximum compressive strain rate in the study area calculated from the GNSS-derived secular velocities (for details, see Methods in Fialko & Jin, 2021).

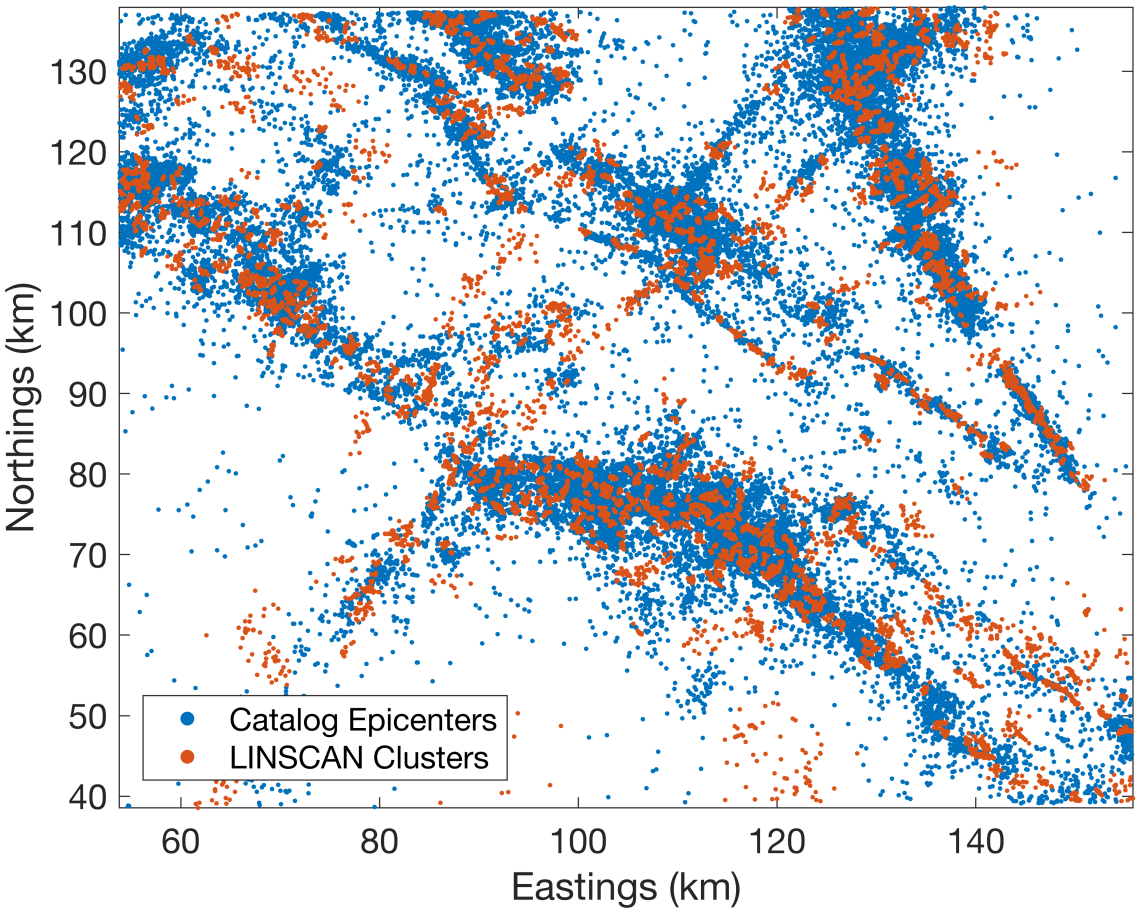

Figure S4. Catalog seismicity (blue dots, same as in Figure 1), and quasi-linear clusters (QLCs) of seismicity identified by the LINSCAN algorithm (orange dots). The total number of QLCs identified by LINSCAN is 1181.

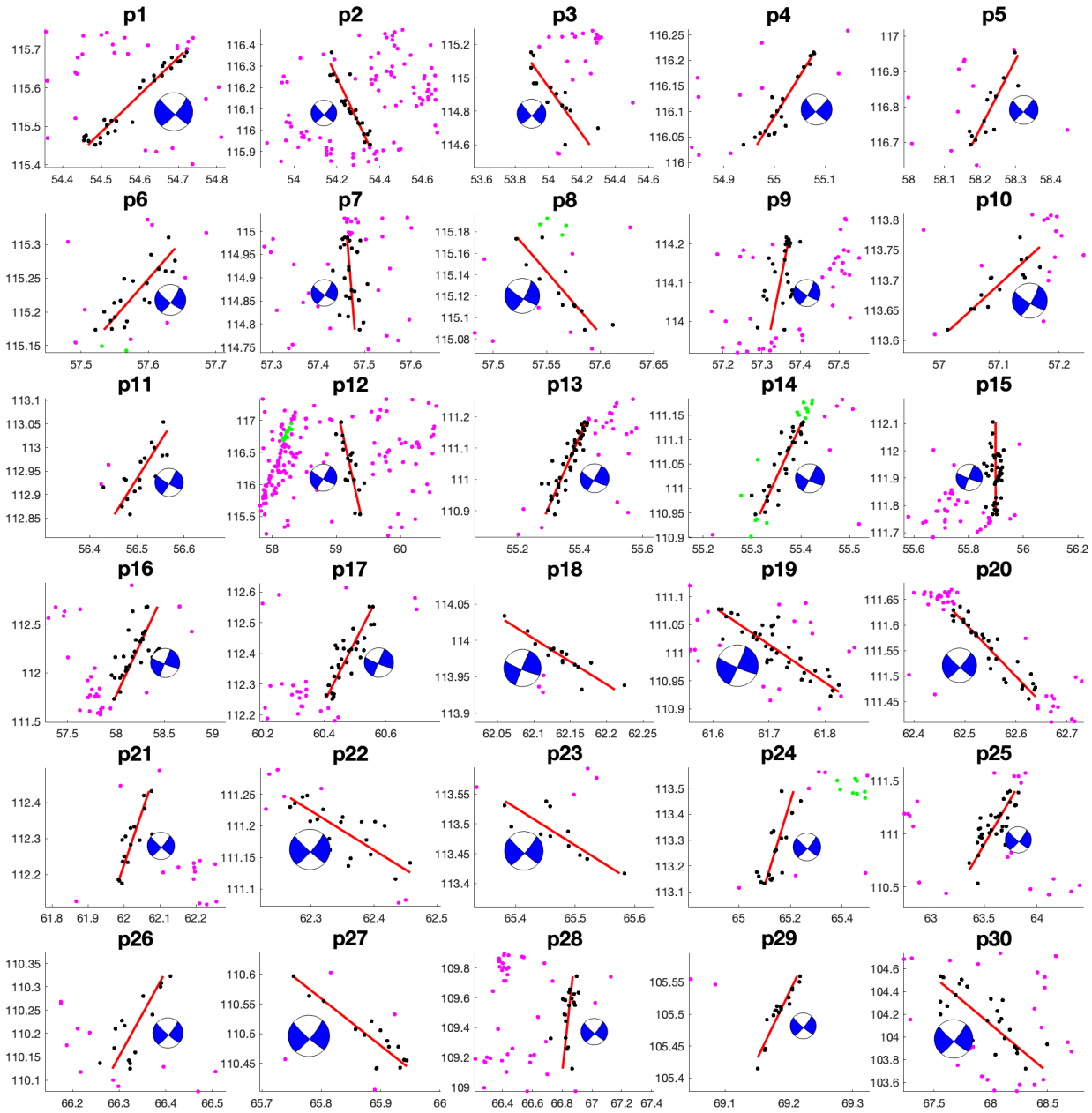

Figure S5. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

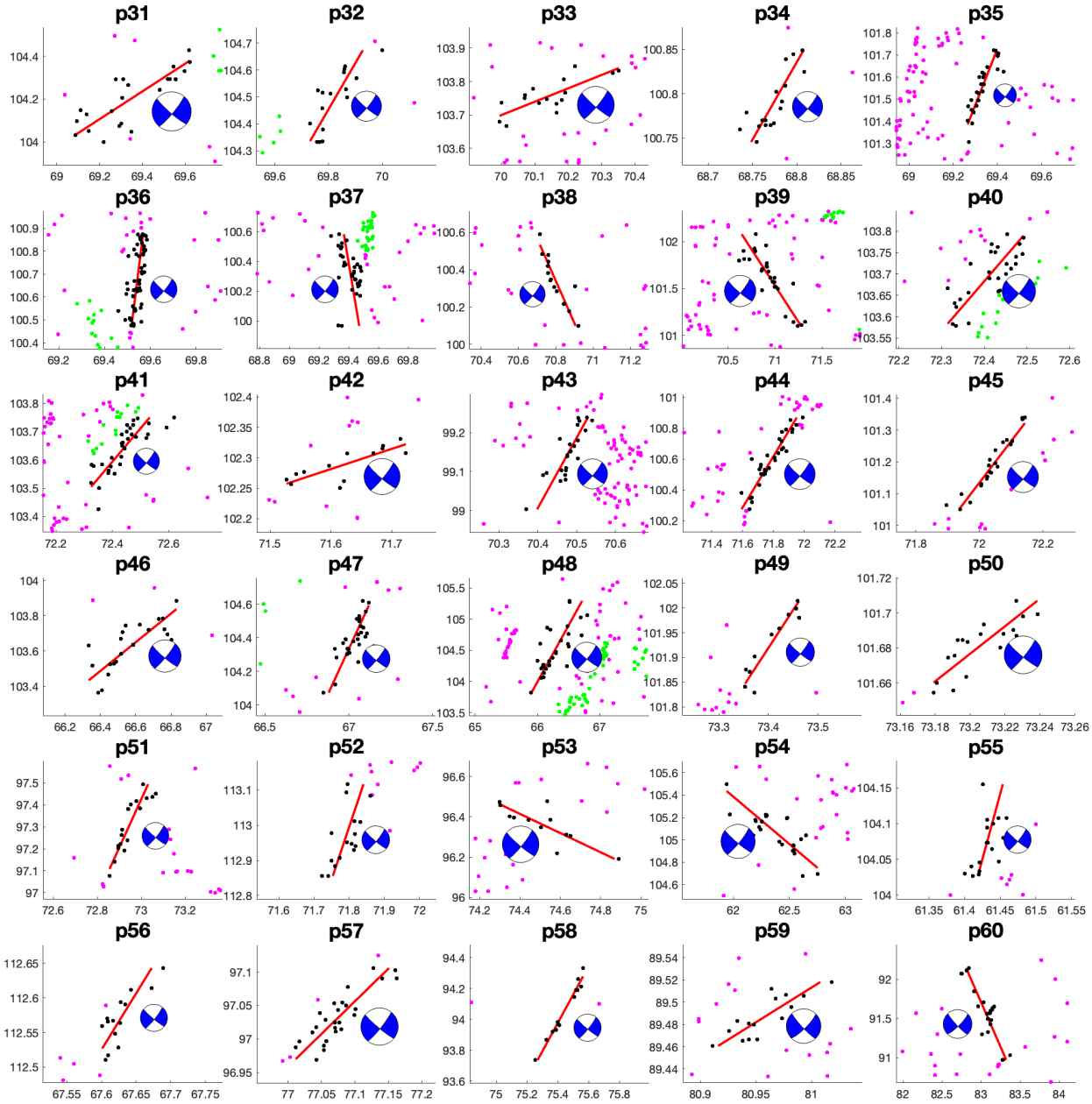

Figure S6. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

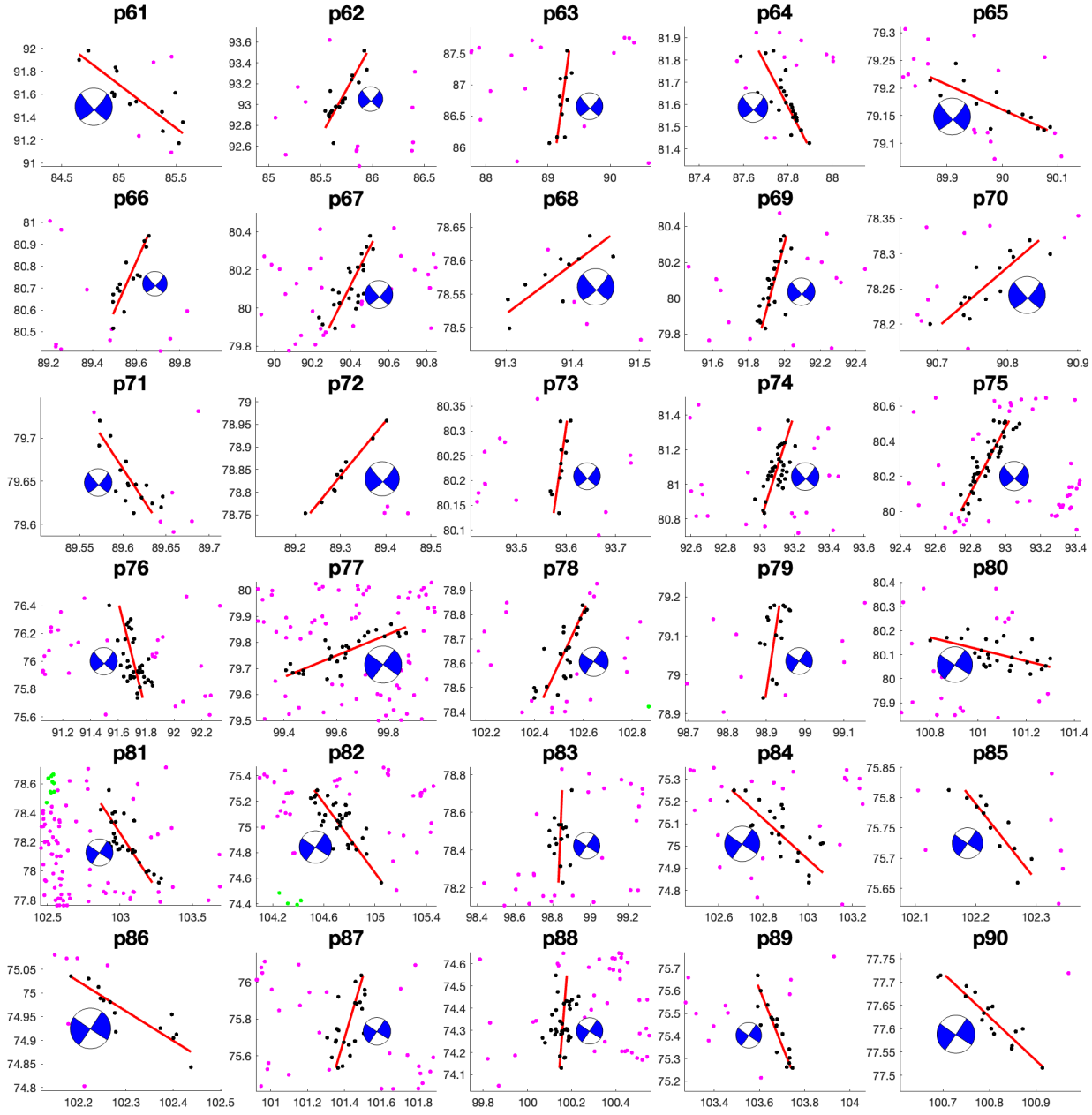

Figure S7. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

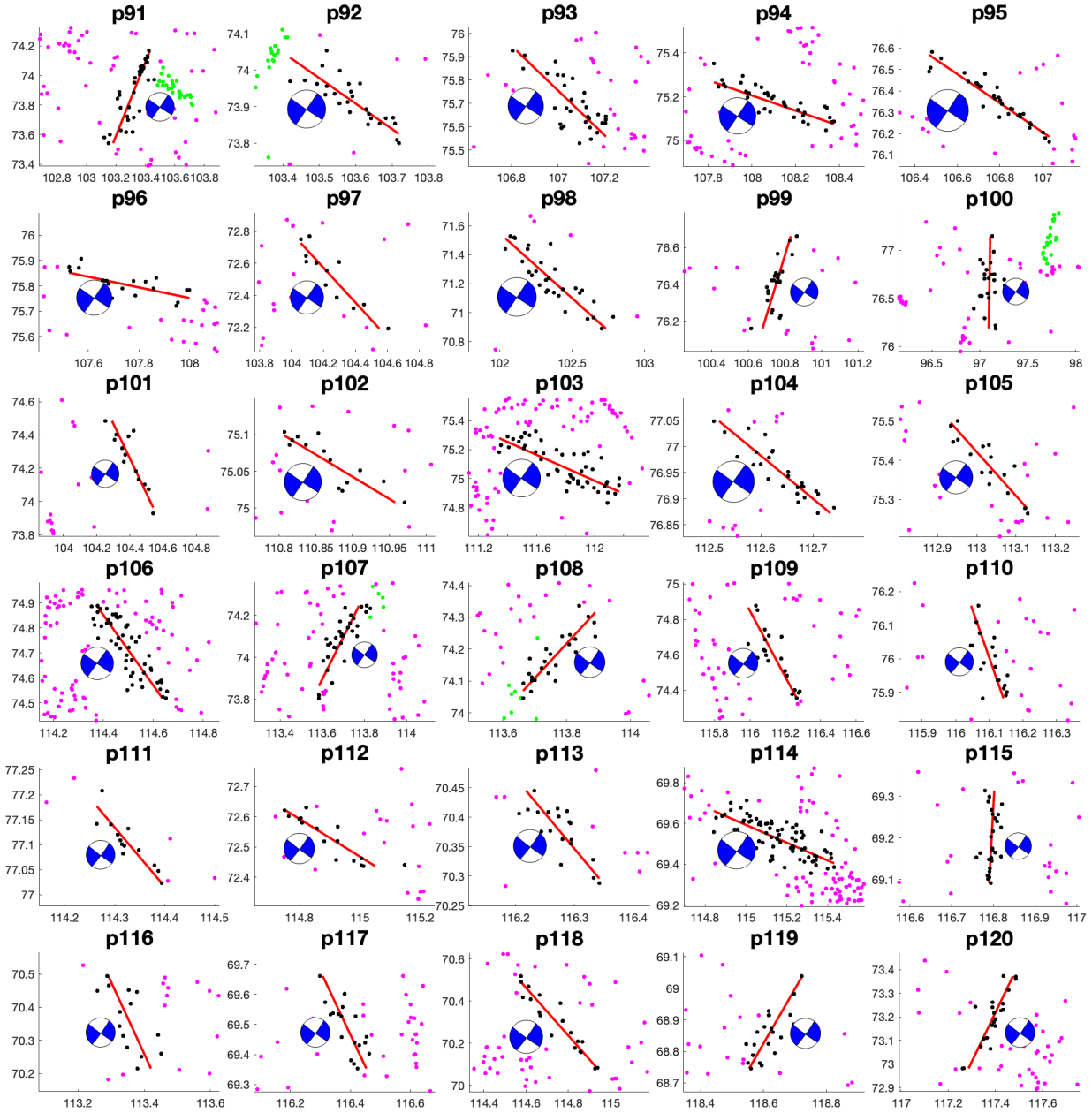

Figure S8. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

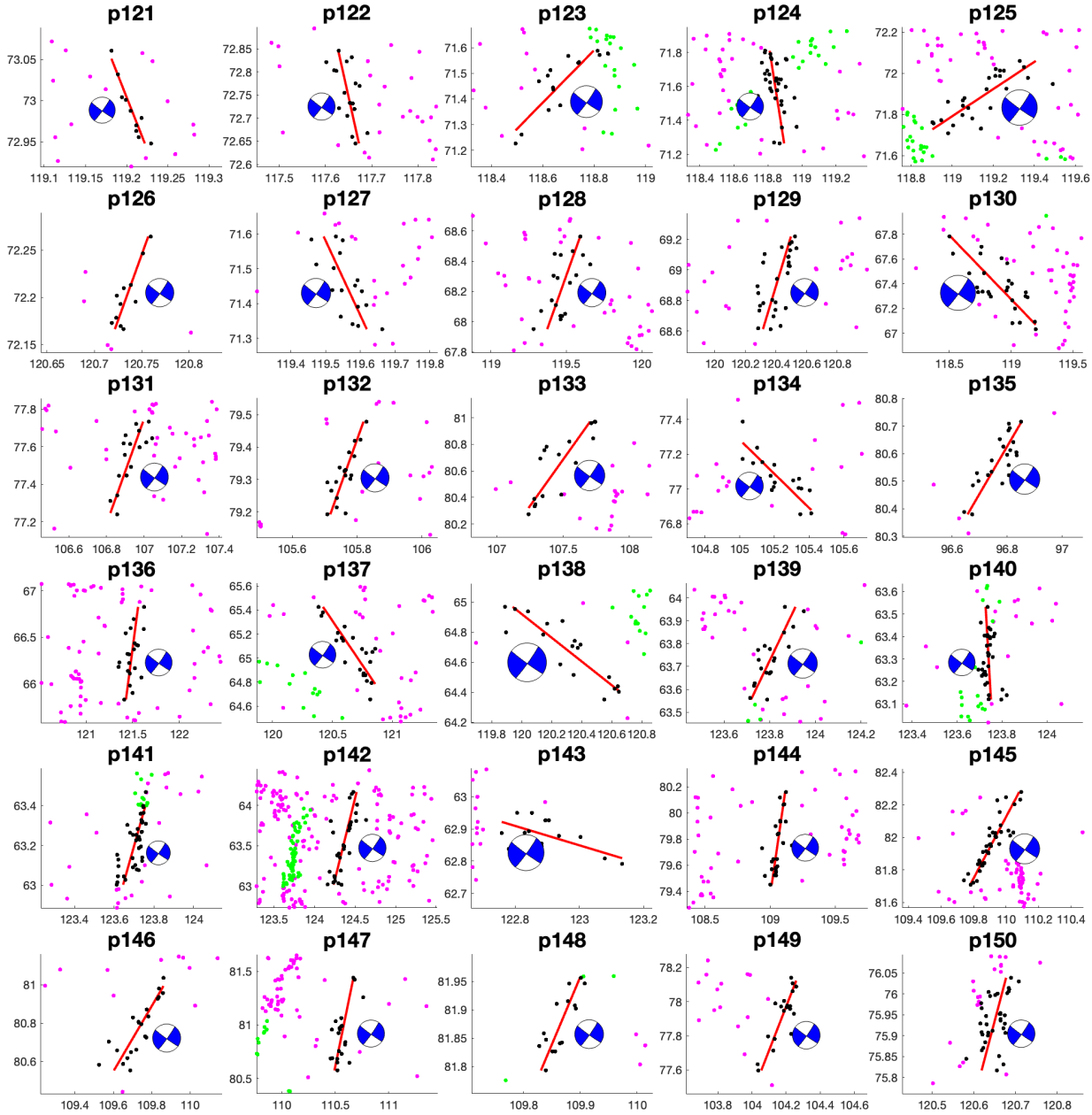

Figure S9. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

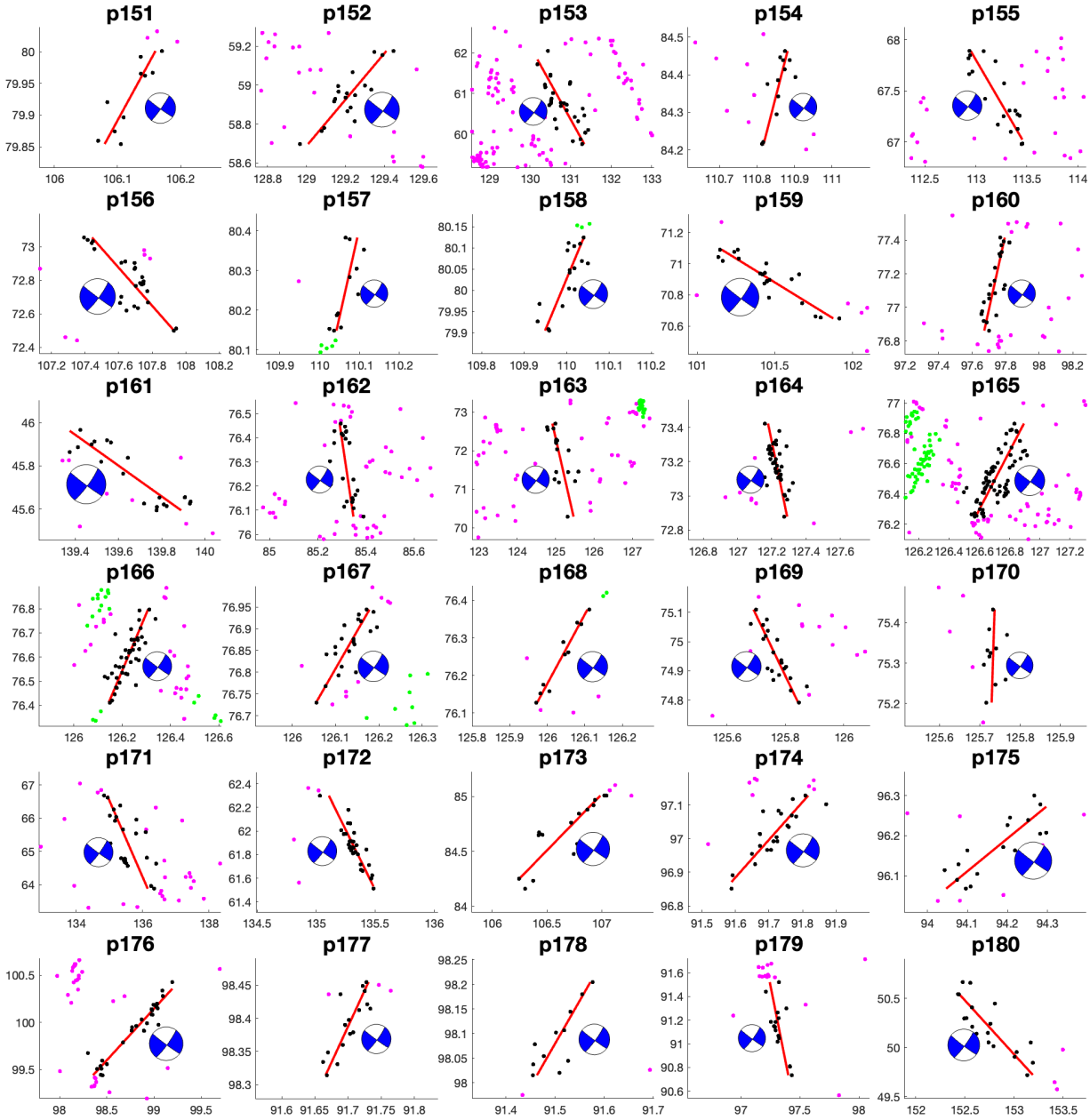

Figure S10. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

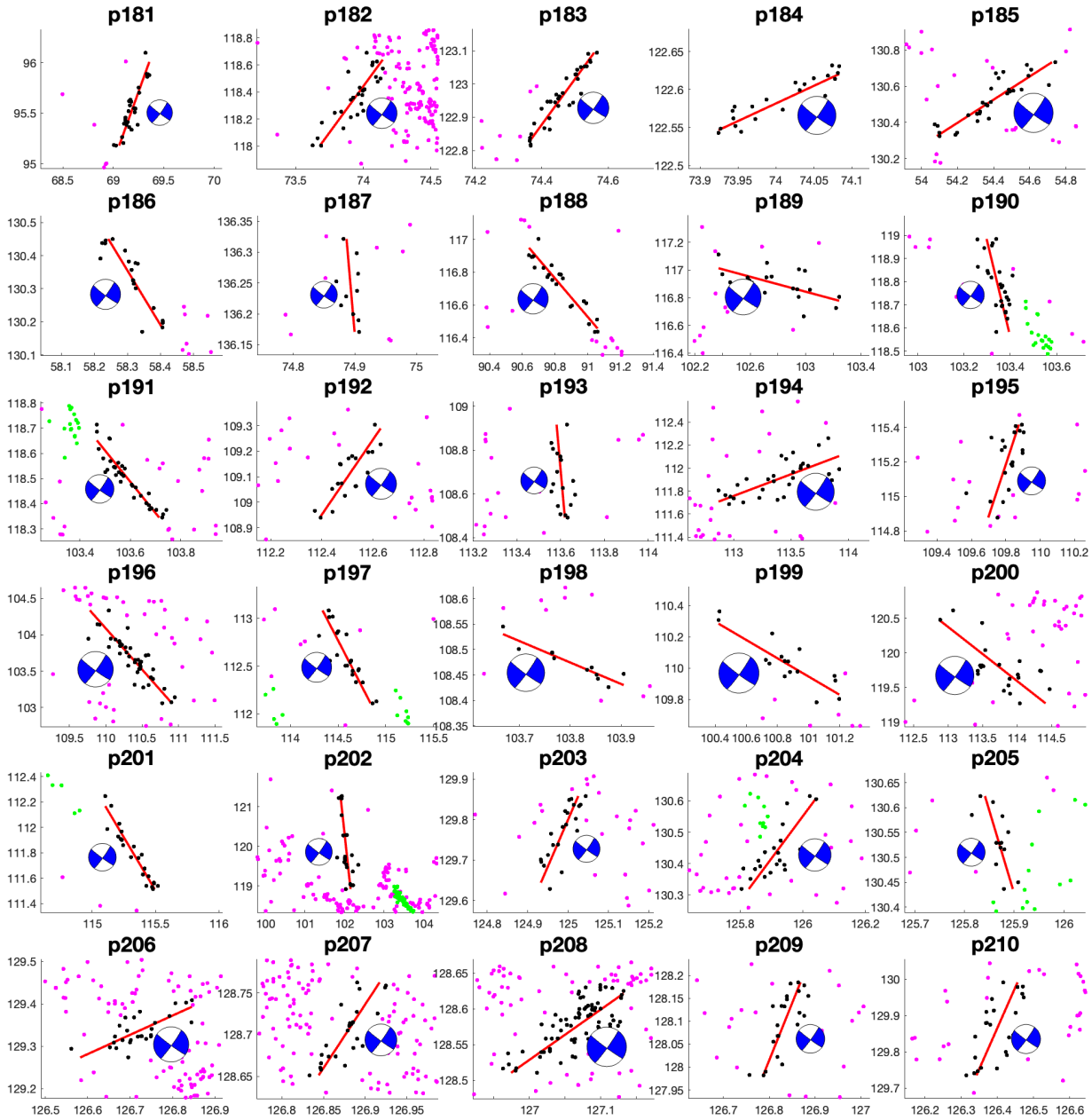

Figure S11. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

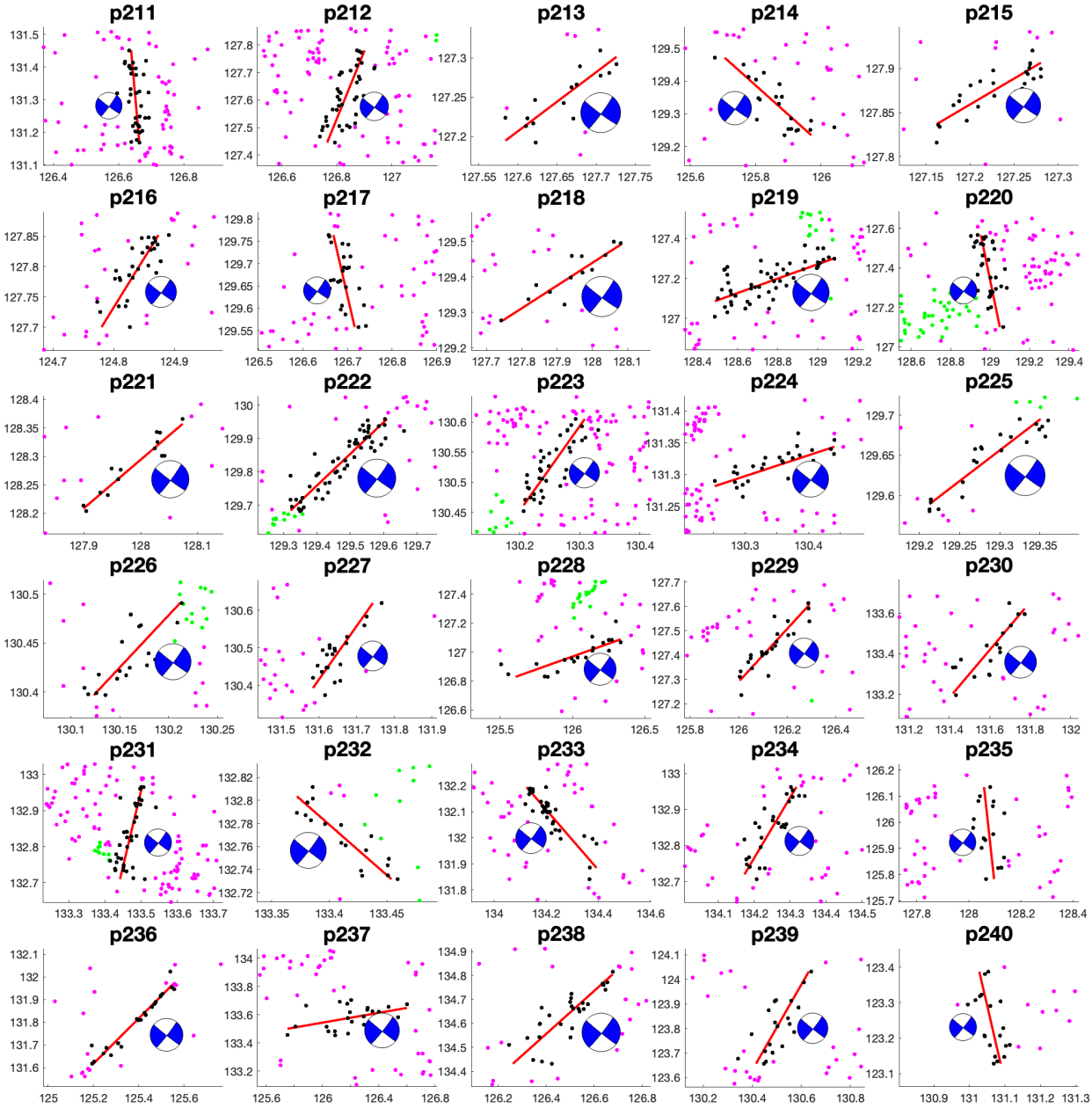

Figure S12. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

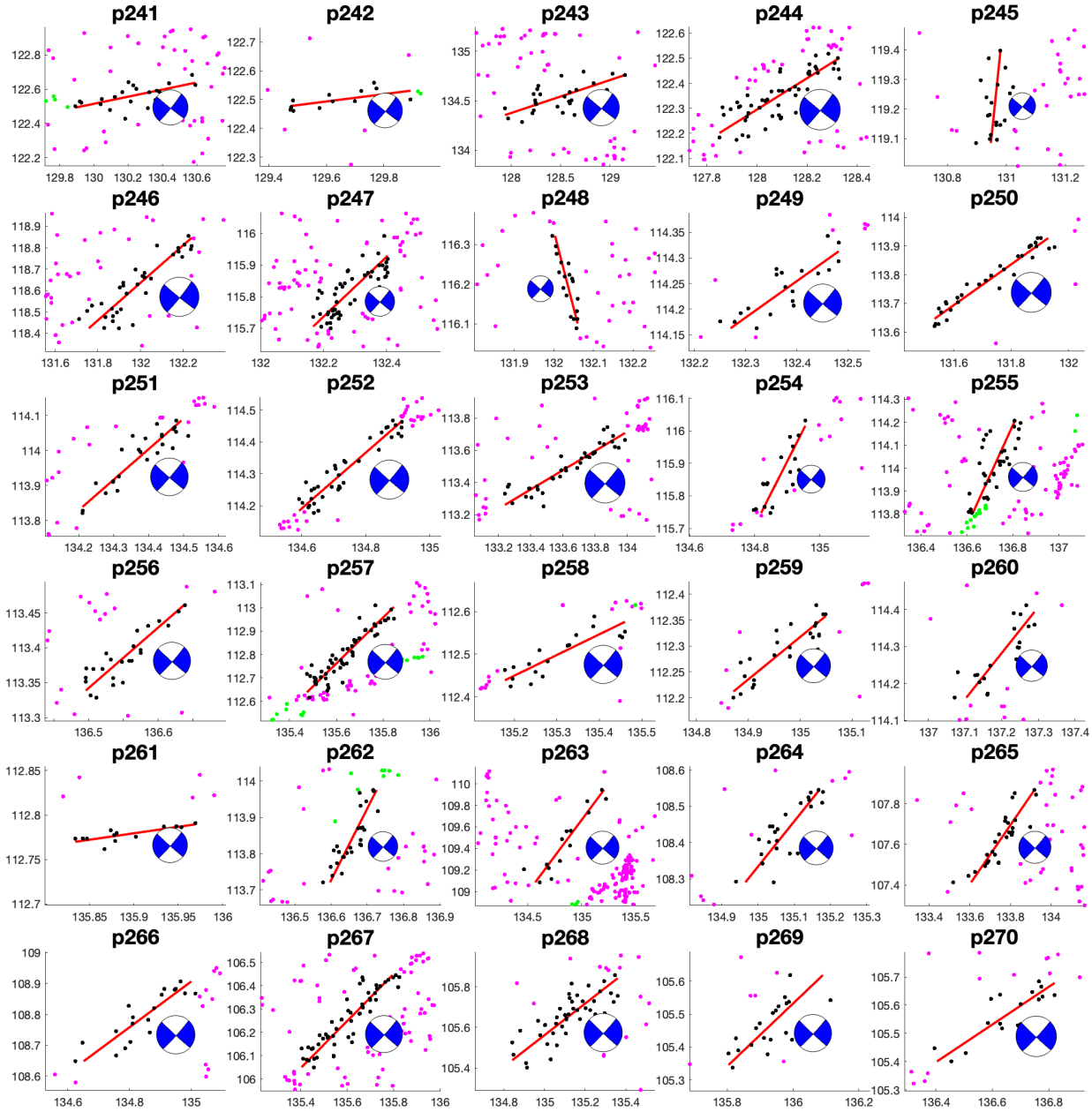

Figure S13. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

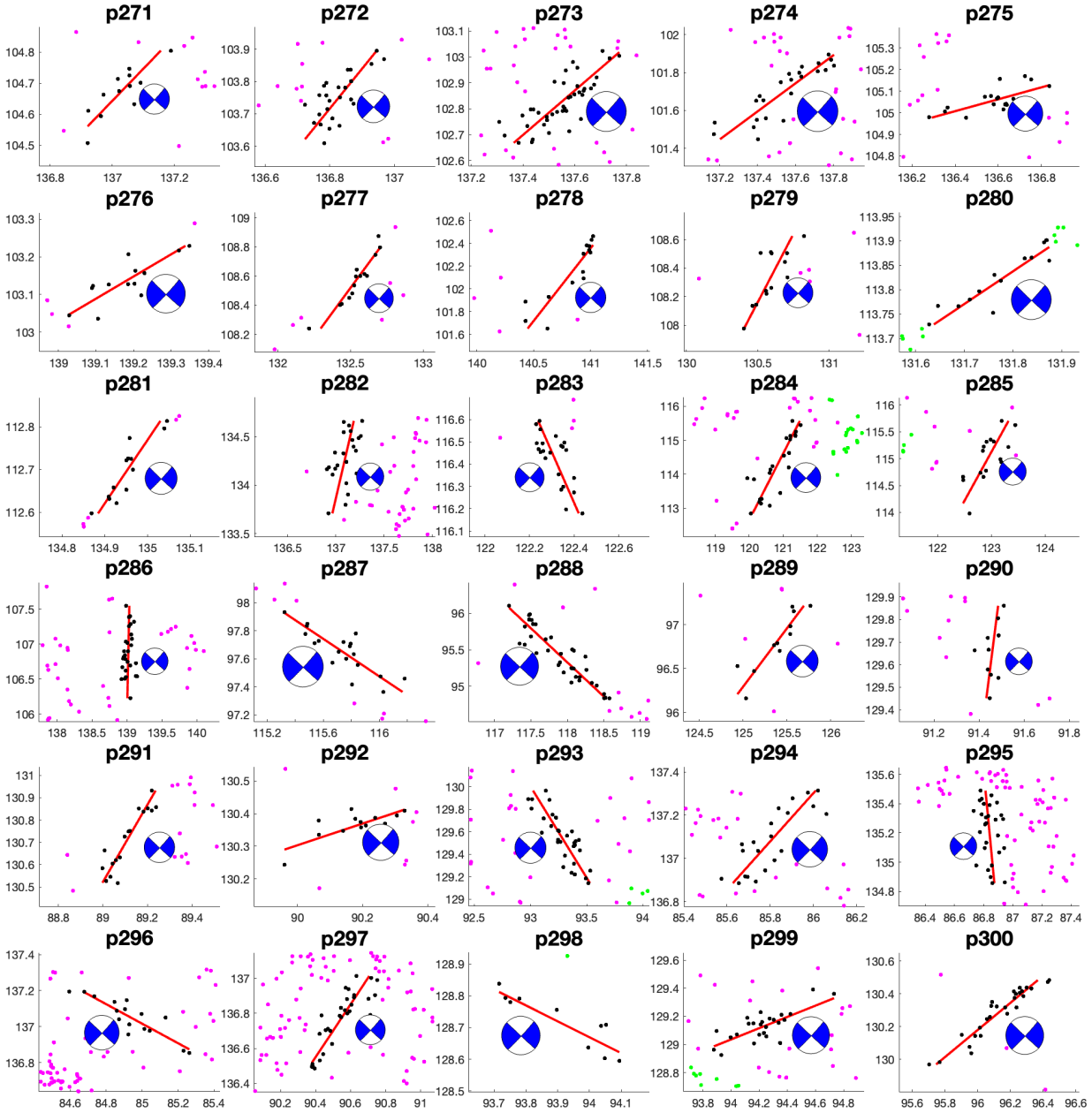

Figure S14. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

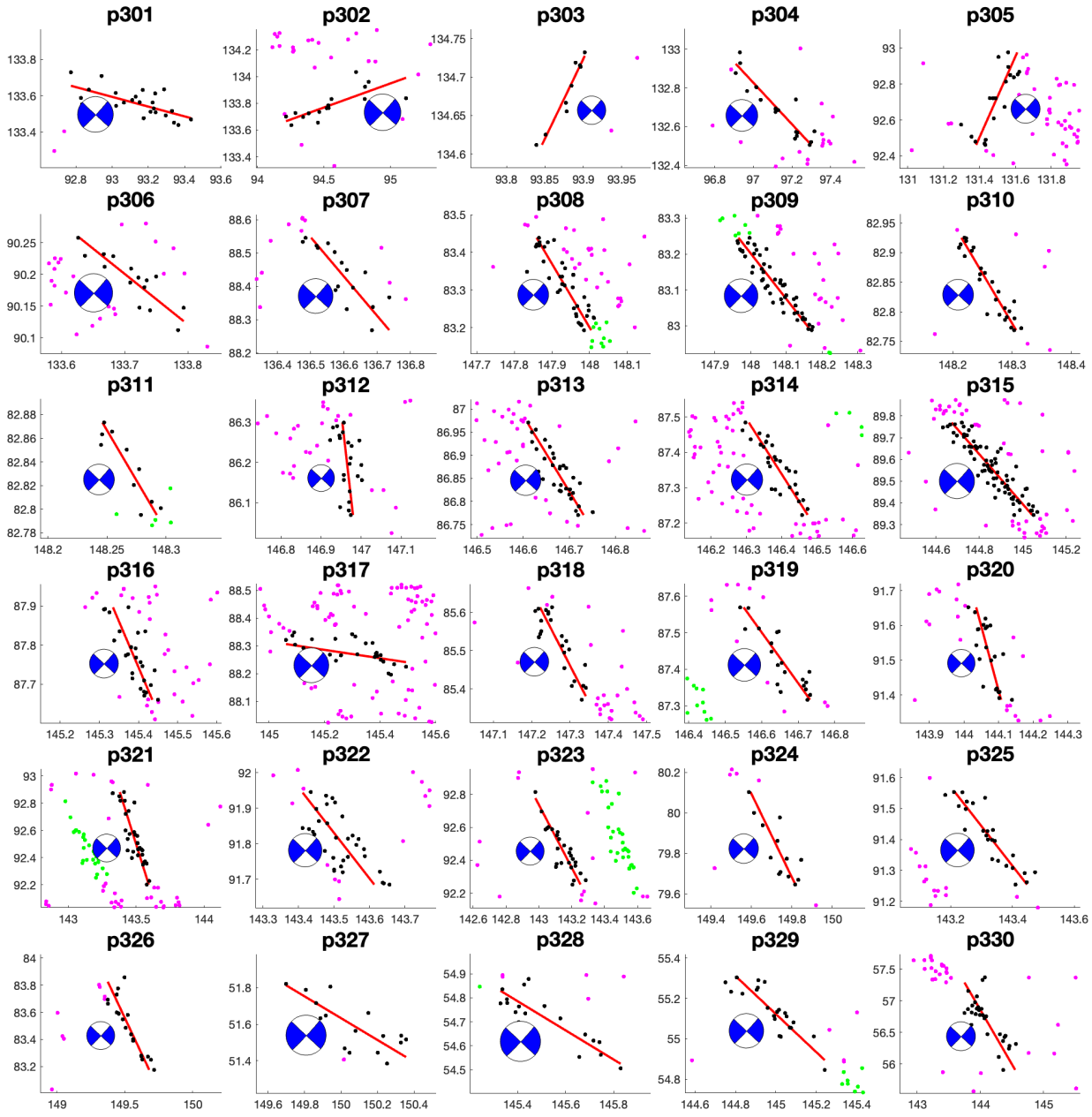

Figure S15. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.

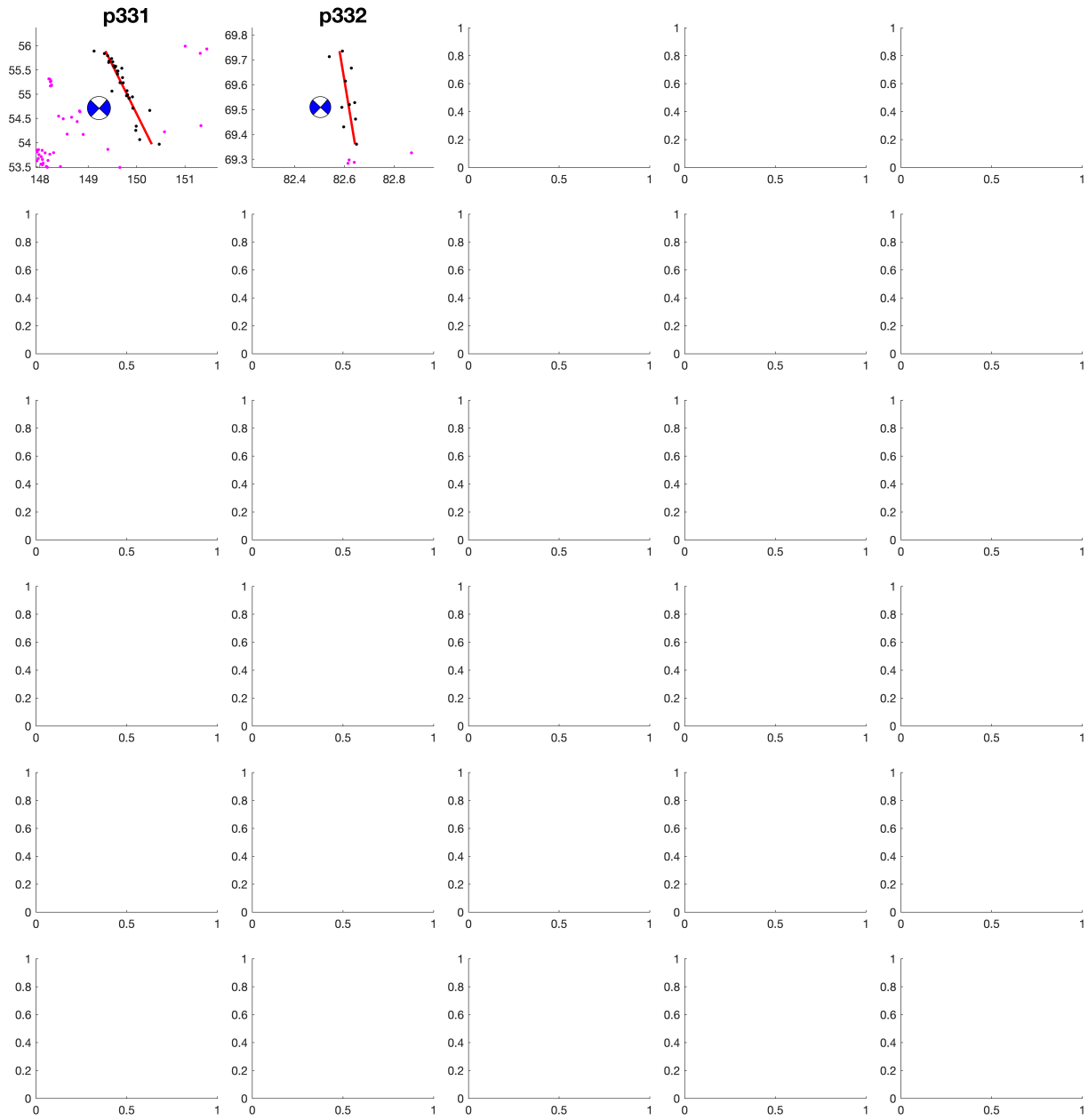

Figure S16. Close-up views of selected quasi-linear clusters. Each panel shows a quasi-linear cluster (black dots), along with a composite focal mechanism (blue beach ball). Solid red line denotes a best-fit line segment. Magenta dots denote background seismicity. Green dots denote other selected quasi-linear clusters.