Large-Scale Ambient Noise Cross-Correlation Across California using Cloud Computing

Chris D. Lin[1] (chrisdjlin@berkeley.edu), Weiqiang Zhu[2,3], Taka'aki Taira[3]
[1]National Taiwan University, [2]University of California, Berkeley, [3]Berkeley Seismological Laboratory

Large Data in California

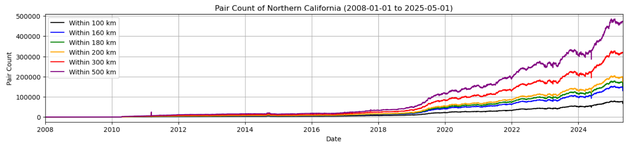

Figure 1. Daily number of available station pairs in northern California from January 1, 2008 to May 1, 2025, calculated within different interstation distance ranges (100–500 km).

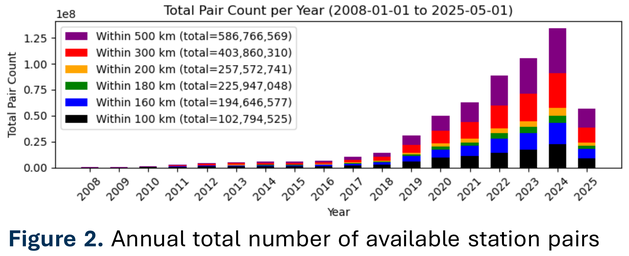

Figure 2. Annual total number of available station pairs in northern California from January 1, 2008 to May 1, 2025, shown for different interstation distance ranges (100–500 km). The cumulative totals for each distance range are indicated in the legend.

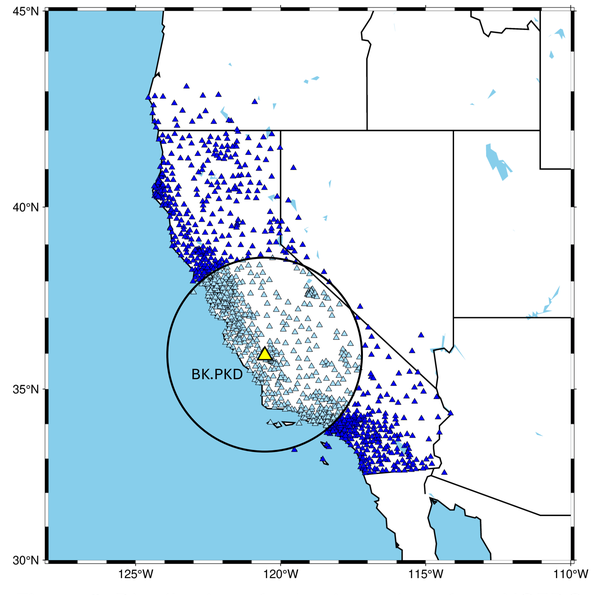

Figure 3. Distribution of seismic stations from NCEDC and SCEDC. Stations located within 300 km of BK.PKD (yellow triangle) are enclosed by the black circle, representing potential cross-correlation pairs with BK.PKD.

Data Application - Ambient Noise Tomography

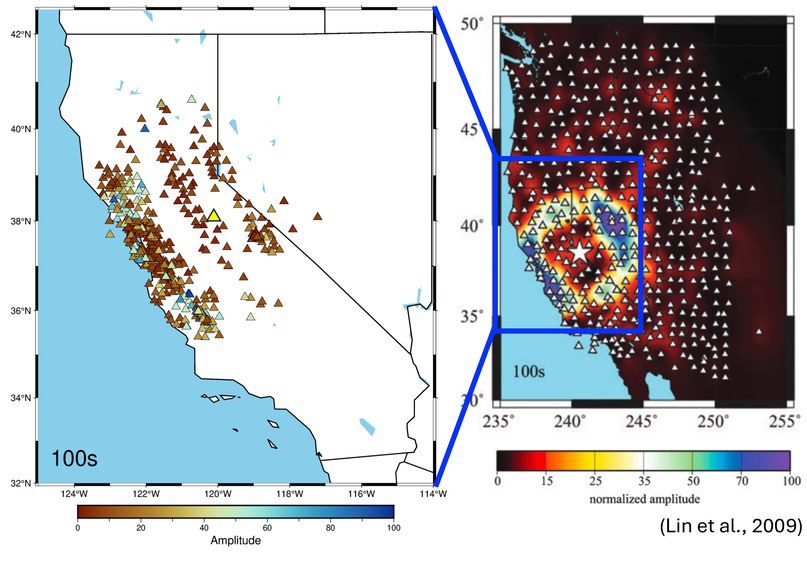

Figure 8. Amplitude of CCFs paired with NC.MLR (yellow triangle) at 100 s. Stations with higher amplitudes (blue) are located at similar distances from the source as shown in the reference map on the right (Lin et al., 2009).

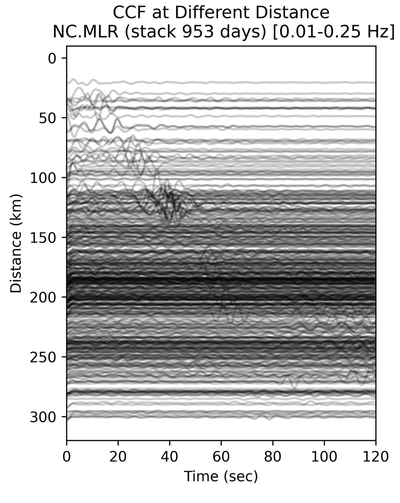

Figure 9. Stacked CCFs for NC.MLR (953 days, 0.01–0.25 Hz). Traces are shown by interstation distance, revealing surface-wave arrivals.

Ambient Noise Cross-Correlation

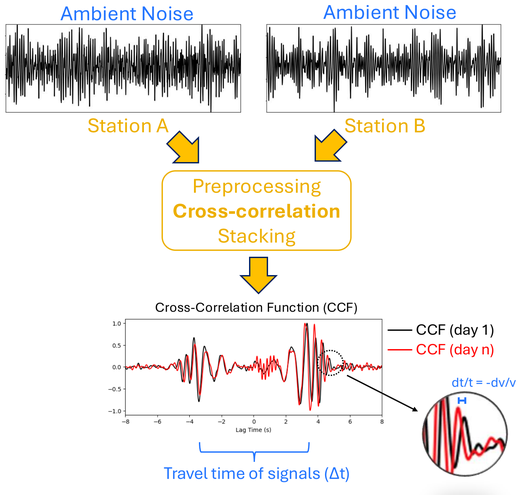

Figure 4. Schematic of ambient noise cross-correlation. Noise from two stations is cross-correlated to obtain the cross-correlation functions (CCFs), which provides surface-wave travel times for tomography and temporal shifts for seismic velocity change (dv/v).

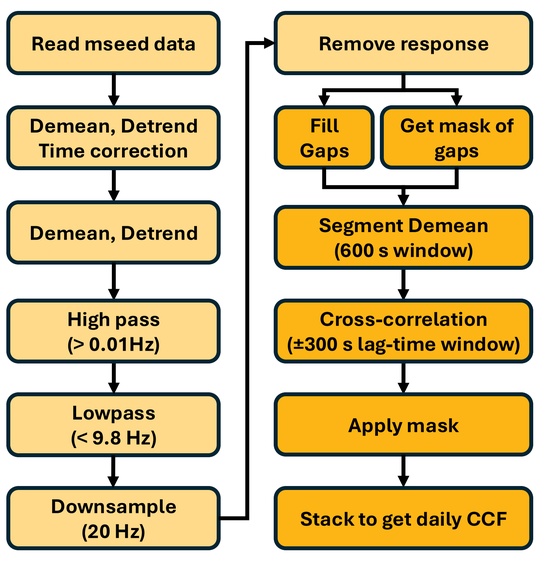

Figure 5. Workflow of preprocessing and cross-correlation. Steps highlighted in gold are parallelized and executed on the cloud.

Optimizing Stacking - Accuracy vs. Efficiency

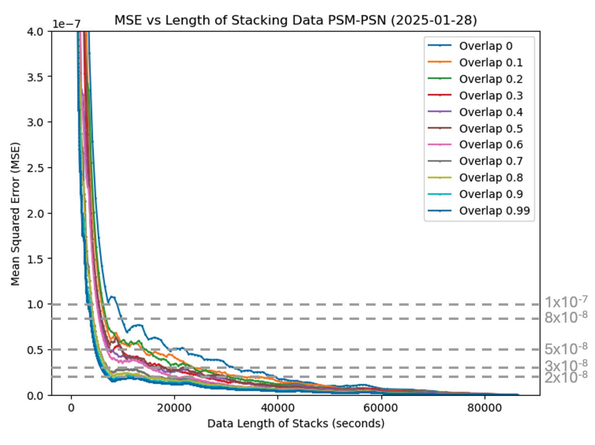

Figure 10. MSE decreases rapidly as the length of stacked cross-correlations increases. Different overlap ratios (0–0.99) show consistent convergence behavior, with higher overlap leading to slightly lower MSE at shorter stacking lengths.

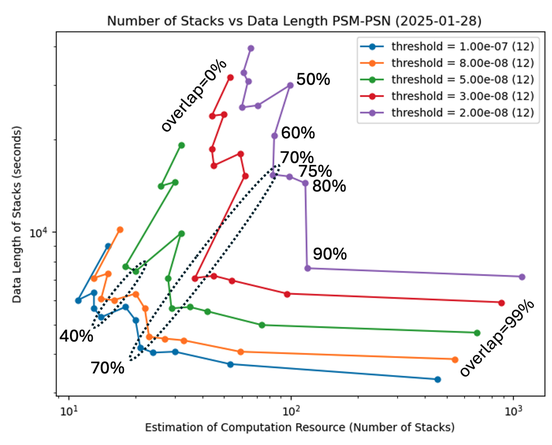

Figure 11. Trade-off between computation cost (number of stacks) and stacking length for different overlap ratios. Higher overlap ratios (≥70%) achieve lower MSE with shorter data lengths but require significantly more stacks (higher computational cost).

Cloud Computing

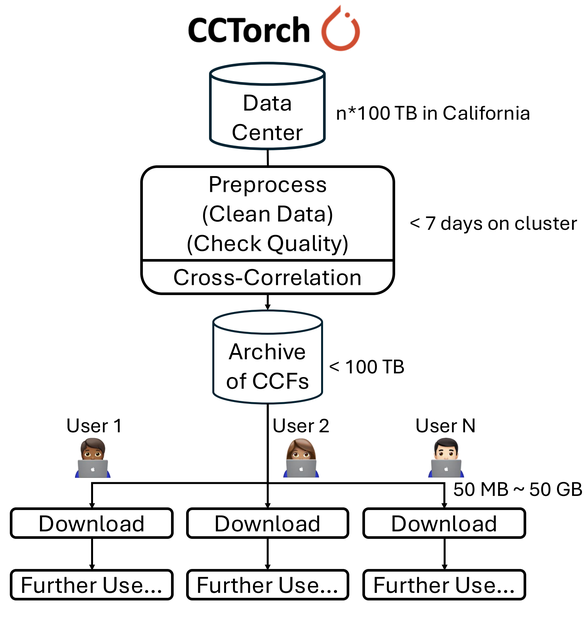

Figure 7. CCTorch workflow: large raw seismic datasets are preprocessed and cross-correlated into an archive of CCFs (<100 TB). Users can then download smaller subsets (50 MB–50 GB) for further studies.

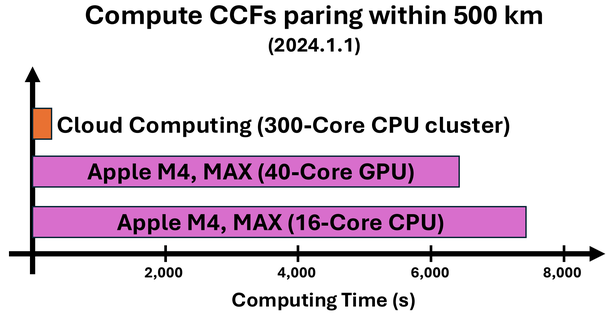

Figure 6. Computing times for CCFs within 500 km (2024-01-01). Cloud CPU clusters outperform local Apple M4 Max hardware, but GPU acceleration provides significant gains over CPU-only runs.

Benchmark

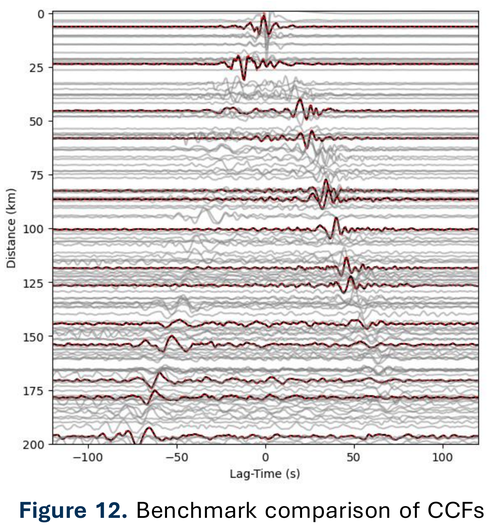

Figure 12. Benchmark comparison of CCFs from MSNoise (red dashed lines) and from the cloud-based package CCTorch (gray lines). The CCFs are computed for station pairs with BK.PKD within 200 km.

Summary

- We processed large-scale ambient noise CCFs across California using cloud computing from 2008 to 2025, greatly reducing redundant work for subsequent applications.
- The CCFs span a ±300 s lag-time window with a 20 Hz sampling rate and show strong consistency with results from MSNoise.
- At the same convergence threshold of CCFs, a 70% overlap provides an optimal balance, requiring less data length and computational resources.
- These data can be applied to construct 3D surface-wave velocity models, evaluate ground-motion amplification, and monitor seismic velocity changes.
- They provide valuable resources for time-lapse monitoring of fault zone structures, natural resources, and groundwater recharge systems.

References

- Fan-Chi Lin, Michael H. Ritzwoller, Roel Snieder, Eikonal tomography: surface wave tomography by phase front tracking across a regional broad-band seismic array, Geophysical Journal International, Volume 177, Issue 3, June 2009, Pages 1091–1110, https://doi.org/10.1111/j.1365-246X.2009.04105.x