The Journal of Open Source Software

PyTurbo_SF: An Adaptive Bootstrap Framework for Efficient Structure Function Analysis in Turbulent Flows

Adam Ayouche[1¶], **Baylor Fox-Kemper**[1], **Momme Hell**[2], **Brodie Pearson**[3], and **Cassidy Wagner**[3]

1 Department of Earth, Environmental, and Planetary Sciences (DEEPS), Brown University, Providence, RI, USA **2** Woods Hole Oceanographic Institution, Woods Hole, MA, USA **3** College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA ¶ Corresponding author

DOI: 10.21105/joss.09876

Software

- Review
- Repository
- Archive

Editor: Gabriele Bozzola

Reviewers:

- @NoraLoose
- @einaraz

Submitted: 01 October 2025
Published: 10 April 2026

Summary

Structure functions (SFs) are fundamental diagnostic tools in turbulence research that quantify spatial correlations between field differences at varying separation distances, revealing energy cascade characteristics and scaling laws (Frisch, 1995; Pope, 2000). PyTurbo_SF is a comprehensive Python package providing efficient, statistically rigorous structure function calculations for 1D, 2D, and 3D turbulent datasets through an innovative adaptive bootstrap framework.

The package addresses computational bottlenecks through power-of-2 spacing strategies, adaptive convergence monitoring, and memory-optimized algorithms. PyTurbo_SF delivers robust uncertainty quantification through bootstrap resampling and supports diverse (conditional) structure function types including longitudinal, transverse, scalar, advective, and energy flux functions (Pearson et al., 2021, 2024).

Applications span oceanographic time series and satellite measurements to high-resolution simulations, enabling consistent methodology across scales from laboratory to planetary systems. Despite its name evoking the broader turbulence field, PyTurbo_SF is specifically designed for structure function analysis — one of the most widely used diagnostic approaches in turbulence research.

Statement of need

Contemporary turbulence research relies on massive datasets from satellite missions, autonomous platforms, and high-resolution simulations. Traditional structure function calculations face severe limitations: computational intractability for large datasets, absence of uncertainty quantification, manual parameter tuning, and limited function types.

Existing tools address only subsets of these challenges. fastSF provides parallelized implementations but lacks advanced function types and uncertainty quantification (Sadhukhan et al., 2021). MATLAB toolkits are environmentally limited and lack comprehensive statistical frameworks (Fuchs et al., 2022). FluidSF (Wagner et al., 2025) is a related Python package supporting core structure function types, which we compare in detail below. Alternative approaches like FlowSieve cannot provide the scale-by-scale information structure functions uniquely deliver (Storer & Aluie, 2023).

There is a growing need for tools to analyze emerging datasets from NASA's SWOT satellite (Morrow et al., 2019) and next-generation atmospheric simulations generating terabyte-scale outputs. Similarly, flux tower networks (e.g., AmeriFlux, FLUXNET) generate extensive eddy-covariance time series of velocity and scalar turbulence that would benefit from standardized structure function diagnostics with uncertainty quantification. Recent advances in structure function theory, particularly advective structure functions (Pearson et al., 2021) and spectral flux estimation (Pearson et al., 2024), require frameworks handling both traditional and novel function types with statistical rigor.

PyTurbo_SF fills this gap by providing the first comprehensive, statistically robust framework that scales from small observational datasets to massive simulation outputs while delivering quantified uncertainties essential for scientific interpretation.

Software functionality

PyTurbo_SF implements the complete mathematical framework for structure function analysis, supporting functions of the form $S_n(r) = \langle |\phi(\mathbf{x}+\mathbf{r}) - \phi(\mathbf{x})|^n \rangle_\mathbf{x}$ where ϕ represents arbitrary field variables (velocity, scalars, derived quantities), $\mathbf{r}$ is the separation vector, n is the order, and $\langle \cdot \rangle_\mathbf{x}$ denotes spatial averaging. For velocity fields, the package computes **longitudinal** SFs using the velocity component parallel to the separation vector ($\delta u_L = \delta \mathbf{u} \cdot \hat{\mathbf{r}}$), **transverse** SFs using the perpendicular component ($\delta u_T = |\delta \mathbf{u} \times \hat{\mathbf{r}}|$), and **scalar** SFs for temperature, salinity, or other tracers. Cross-structure functions correlate different components (e.g., longitudinal-scalar), while **advective** SFs correlate velocity increments with advection increments to directly quantify energy flux (Pearson et al., 2021).

The core algorithmic breakthrough is adaptive bootstrap methodology increasing computational efficiency and statistical reliability. The algorithm employs power-of-2 spacings optimizing memory access patterns while providing optimal scale separation. Adaptive convergence monitoring dynamically allocates computational resources, eliminating manual parameter tuning while guaranteeing robust uncertainty estimates.

Performance optimization enables analysis of previously intractable datasets. Memory-efficient structures maintain peak usage at 2-5× base dataset size. Parallel processing provides near-linear scaling with available cores. Benchmark testing demonstrates O(NM log N log M) complexity for 2D data, enabling analysis of datasets with millions of grid points.

The package provides three main interfaces: `bin_sf_1d()` for time series, `bin_sf_2d()` for surface fields, and `bin_sf_3d()` for volumetric data. All functions automatically optimize computational strategies while supporting both isotropic and directional analysis. Applications extend beyond traditional oceanographic and atmospheric modeling contexts to include eddy-covariance field measurements, where high-frequency sonic anemometer time series of wind components and scalars can be analyzed using `bin_sf_1d()` for temporal structure functions.

Figures 1–4 demonstrate PyTurbo_SF applied across scientific domains, illustrating the package's versatility for both satellite observations and numerical simulations.

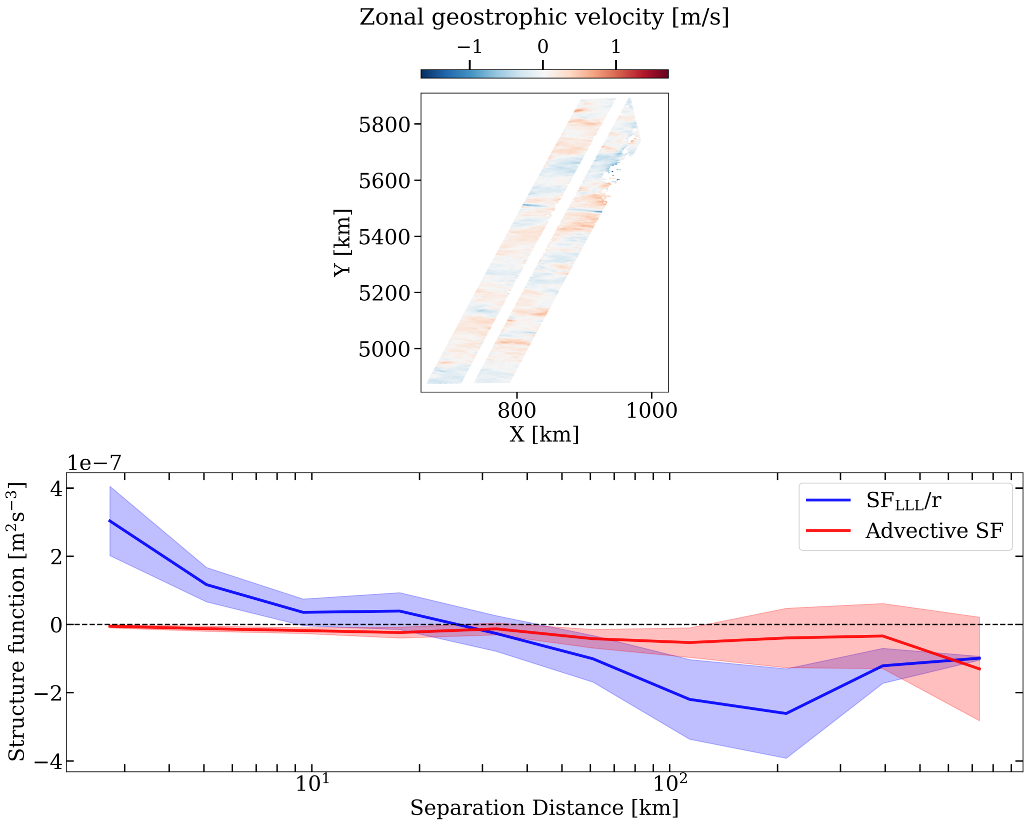

Figure 1: Analysis of NASA's Surface Water and Ocean Topography (SWOT) satellite altimetry data (Morrow et al., 2019) in the Gulf Stream region. **Upper panel:** Zonal geostrophic velocity field (u_g, m/s) derived from sea surface height gradients, mapped onto a local Cartesian coordinate system where X and Y represent eastward and northward distances in kilometers; the two diagonal swaths correspond to SWOT's twin Ka-band radar interferometer tracks. **Lower panel:** Isotropically-averaged structure functions computed using `get_isotropic_sf_2d()`: the third-order longitudinal SF divided by separation distance (SF_{LLL}/r, blue; `fun='longitudinal', order=3`) and the advective SF (ASF_V, red; `fun='advective'`), both in units of m^2 s^{-3}. Shaded regions indicate 95% bootstrap confidence intervals. These third-order SFs are proportional to spectral energy flux: positive values indicate forward cascade (energy transfer to smaller scales) while negative values indicate inverse cascade (energy transfer to larger scales). The sign change near 100 km separation distance reveals the transition between cascade regimes characteristic of quasi-geostrophic ocean turbulence.

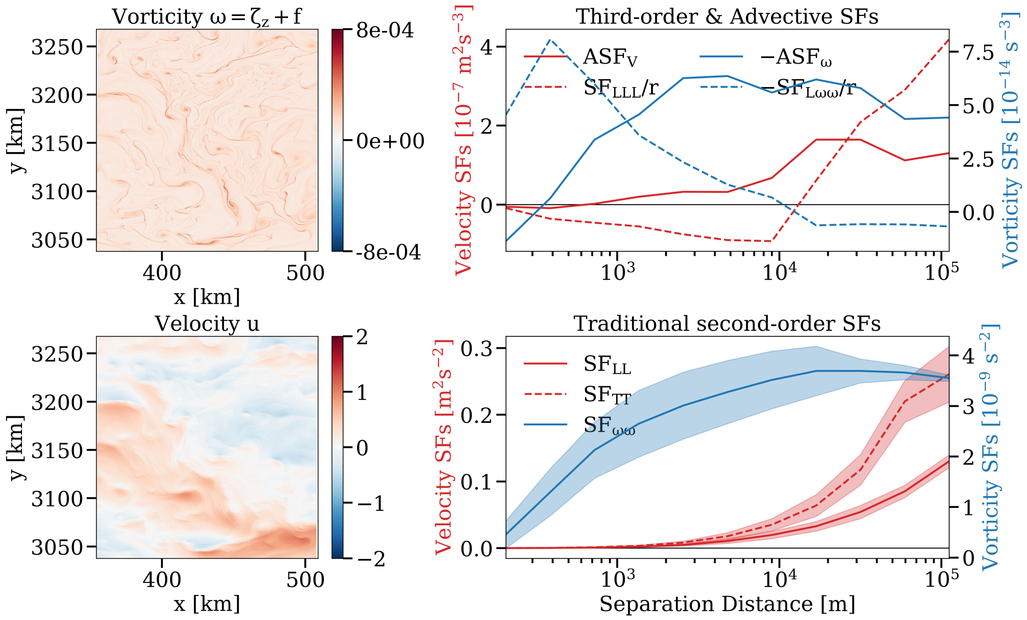

Figure 2: Analysis of Coastal and Regional Ocean COmmunity model (CROCO) output (Shchepetkin & McWilliams, 2005), a high-resolution regional ocean simulation. **Left panels:** Input fields over a 150 km × 250 km domain — absolute vorticity $\omega = \zeta_z + f$ (vertical component of relative vorticity plus Coriolis parameter; upper) and zonal velocity u (lower). **Upper right:** Third-order and advective SFs related to energy flux — velocity advective SF (ASF_V, solid red) and SF_{LLL}/r (dashed red) on left axis (m^2 s^{-3}); negative vorticity advective SF ($-\mathrm{ASF}_\omega$, solid blue) and $-\mathrm{SF}_{L\omega\omega}/r$ (dashed blue) on right axis (s^{-3}). **Lower right:** Traditional second-order SFs — longitudinal (SF_{LL}, solid red) and transverse (SF_{TT}, dashed red) velocity SFs on left axis (m^2 s^{-2}); scalar vorticity SF ($\mathrm{SF}_{\omega\omega}$, blue with shading) on right axis (s^{-2}). Shaded regions represent 95% bootstrap confidence intervals. All isotropic SFs computed using `get_isotropic_sf_2d()` with appropriate `fun` and `order` parameters.

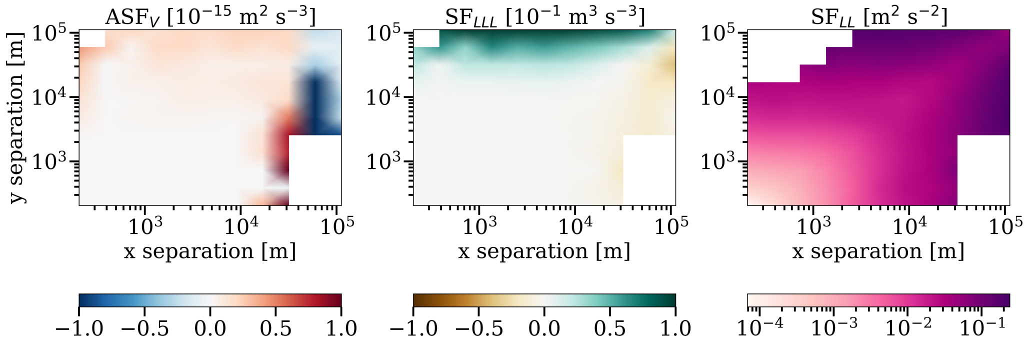

Figure 3: Two-dimensional (non-isotropic) structure functions from CROCO simulation computed using `bin_sf_2d()`, showing SF values as functions of separation distance in both x (zonal) and y (meridional) directions on logarithmic axes from 10^3 to 10^5 m. **Left:** Advective SF (ASF_V, m^2 s^{-3}). **Center:** Third-order longitudinal SF (SF_{LLL}, m^3 s^{-3}). **Right:** Second-order longitudinal SF (SF_{LL}, m^2 s^{-2}). Anisotropic features are evident: the SFs are not symmetric about the diagonal, indicating different turbulent energy transfer rates in zonal versus meridional directions, reflecting the influence of background flow and mesoscale eddy structures. White regions at large separations indicate insufficient data pairs for reliable statistics.

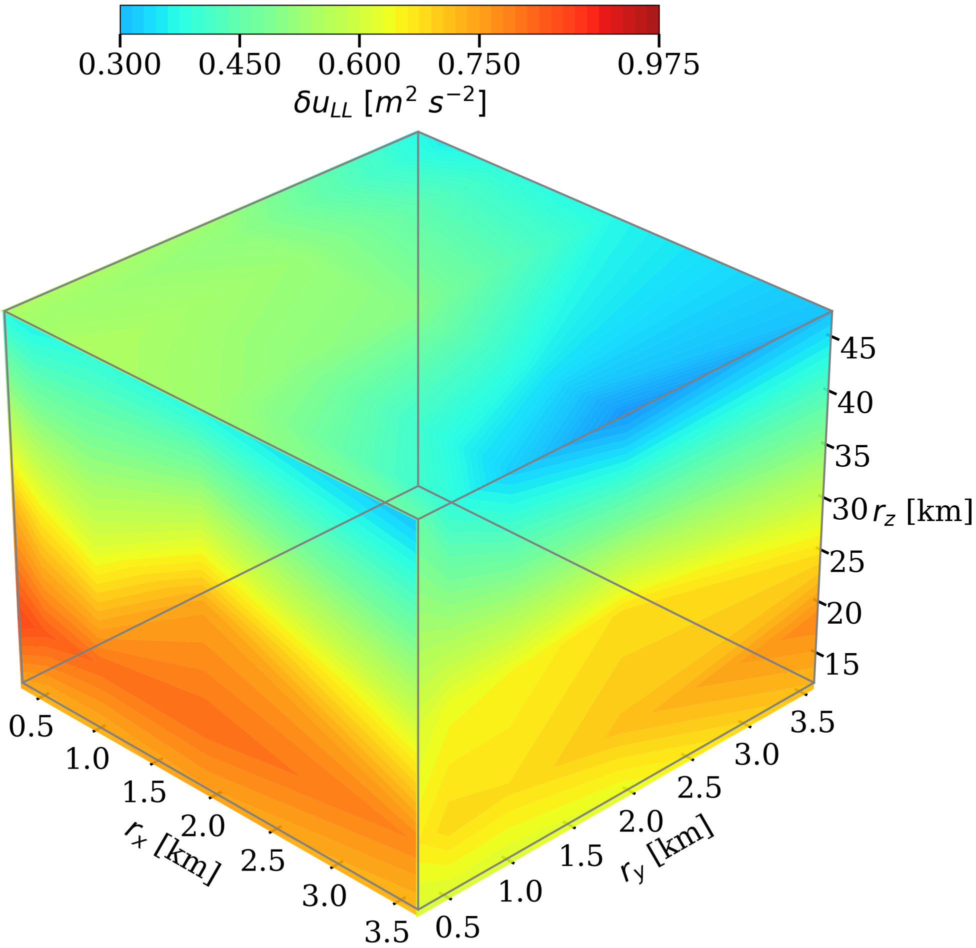

Figure 4: Three-dimensional structure function analysis of DYCOMS-II (Second Dynamics and Chemistry of Marine Stratocumulus) large eddy simulation output (Stevens et al., 2005), modeling the stratocumulus-topped marine boundary layer. The figure displays the second-order longitudinal SF (SF_{LL}, m^2 s^{-2}) computed using `bin_sf_3d()` with `fun='longitudinal'` and `order=2`. Axes represent separation distances: r_x and r_y (horizontal, 0–3500 m) and r_z (vertical, 0–45000 m). Color scale ranges from ~0.3 to ~0.9 m^2 s^{-2}. The SF increases with separation distance as expected for turbulent flows. The approximately symmetric structure in the horizontal (r_x–r_y) plane indicates isotropic horizontal turbulence, while different scaling in the vertical direction reflects anisotropy imposed by stable stratification at the boundary layer top.

Related work and scientific impact

PyTurbo_SF represents a significant advancement by uniquely combining comprehensive function types, adaptive bootstrap methodology, and optimized algorithms. While fastSF provides basic parallelized calculations (Sadhukhan et al., 2021) and MATLAB toolkits offer specific analyses (Fuchs et al., 2022), no existing software delivers the combination of statistical rigor, efficiency, and breadth required for contemporary turbulence research.

FluidSF (Wagner et al., 2025) is a related Python package for structure function calculations, supporting 1D/2D/3D data and core structure function types including longitudinal, transverse, scalar, and advective functions. At the time of this writing, FluidSF provides second- and third-order structure functions with a simplified isotropic averaging approach and outputs results as NumPy arrays. It offers a velocity-scalar blended cross-term type and a single transverse decomposition in 3D.

PyTurbo_SF is, to our knowledge, the first open-source package to combine arbitrary-order structure function computation, adaptive bootstrap uncertainty quantification with automatic convergence monitoring, and parallel processing in a single framework. It provides extensive cross-term structure function types (longitudinal-transverse, longitudinal-scalar, transverse-scalar, and scalar-scalar), full 3D transverse decomposition across all three coordinate planes (ij, ik, jk), conditional structure functions, exact spherical and polar binning for isotropic averaging, and spectral energy flux estimation via Bessel transform — all accelerated by power-of-2 spacing strategies and joblib-based parallelization. Results are returned as richly annotated xarray Datasets with embedded metadata, convergence diagnostics, and quality masks.

PyTurbo_SF's primary contributions are: (1) rigorous uncertainty quantification through adaptive bootstrap resampling with automatic convergence monitoring, (2) support for arbitrary-order structure functions essential for intermittency analysis, and (3) computational efficiency through parallelization and power-of-2 spacing strategies enabling analysis of large datasets.

The package enables application of recent theoretical developments, particularly advective structure functions providing direct energy flux measurements (Pearson et al., 2021) and spectral flux estimation methodologies (Pearson et al., 2024). These reveal energy pathways traditional approaches cannot capture, offering insights into cascade mechanisms in ocean and atmospheric turbulence.

Acknowledgements

This software package, and the contributions of AA and BFK, is based upon work supported by the US Department of Energy grant DE-SC0024572. BP and CW were supported by the National Science Foundation under Grants OCE-2023721 and OCE-2525055.

References

Frisch, U. (1995). *Turbulence: The legacy of A.N. Kolmogorov*. Cambridge University Press. https://doi.org/10.1017/CBO9781139170666

Fuchs, A., Kharche, S., Patil, A., Friedrich, J., Wächter, M., & Peinke, J. (2022). An open source package to perform basic and advanced statistical analysis of turbulence data and other complex systems. *Physics of Fluids*, *34*(10), 101801. https://doi.org/10.1063/5.0107974

Morrow, R., Fu, L.-L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., d'Ovidio, F., Farrar, J. T., Gille, S. T., Lapeyre, G., & others. (2019). Global observations of fine-scale ocean surface topography with the Surface Water and Ocean Topography (SWOT) mission. *Frontiers in Marine Science*, *6*, 232. https://doi.org/10.3389/fmars.2019.00232

Pearson, B., Pearson, J., & Fox-Kemper, B. (2021). Advective structure functions in anisotropic two-dimensional turbulence. *Journal of Fluid Mechanics*, *916*, A49. https://doi.org/10.1017/jfm.2021.247

Pearson, B., Wagner, C., Fox-Kemper, B., & Samelson, R. (2024). Estimating spectral fluxes in quasi-two-dimensional flows with advective structure functions and Bessel functions. *Journal of Physical Oceanography*. https://doi.org/10.1175/jpo-d-24-0211.1

Pope, S. B. (2000). *Turbulent flows*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840531

Sadhukhan, S., Bhattacharya, S., & Verma, M. K. (2021). fastSF: A parallel code for computing the structure functions of turbulence. *Journal of Open Source Software*, *6*(57), 2185. https://doi.org/10.21105/joss.02185

Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system

(ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, *9*(4), 347–404. https://doi.org/10.1016/j.ocemod.2004.08.002

Stevens, B., Moeng, C.-H., Ackerman, A. S., Bretherton, C. S., Chlond, A., Roode, S. de, Edwards, J., Golaz, J.-C., Jiang, H., Khairoutdinov, M., & others. (2005). Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Monthly Weather Review*, *133*(6), 1443–1462. https://doi.org/10.1175/MWR2930.1

Storer, B. A., & Aluie, H. (2023). FlowSieve: A coarse-graining utility for geophysical flows on the sphere. *Journal of Open Source Software*, *8*(84), 4277. https://doi.org/10.21105/joss.04277

Wagner, C. M., Lee, A., & Pearson, B. (2025). FluidSF: A Python package for calculating turbulent flow statistics. *Journal of Open Source Software*, *10*(114), 7873. https://doi.org/10.21105/joss.07873