TSE and CUDAHM

Statistical tools for analysis and modeling of astronomical time series and populations

Tom Loredo,[1] Jeff Scargle,[2] and Tamás Budavári[3]

[1]Cornell Center for Astrophysics and Planetary Science; [2]NASA Ames Research Center; [3]Dept. of Applied Math, Johns Hopkins University

TSE: Time Series Explorer

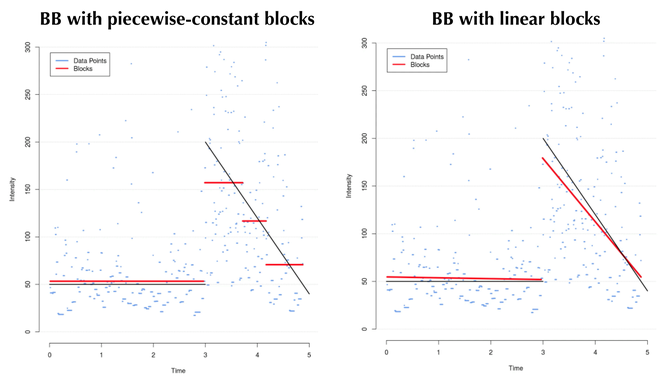
Work by Loredo & Scargle. This work is funded by NASA ADAP via grant NNX16AL02G.

Project scope

TSE is a project developing ***open-source software*** in Python and MATLAB for ***exploratory analysis and statistical modeling of astronomical time series***. The software implements a suite of time-tested ***conventional methods***, as well as ***new methods***. Here we highlight selected new methods.

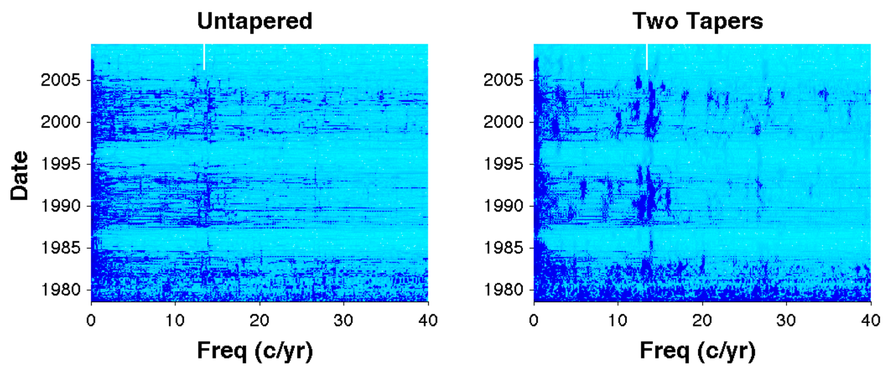
Bayesian blocks

The widely-used ***Bayesian Blocks*** (BB) algorithm for change point detection/adaptive histogramming has been generalized to accommodate ***diverse block shapes***, and accelerated to handle larger datasets.



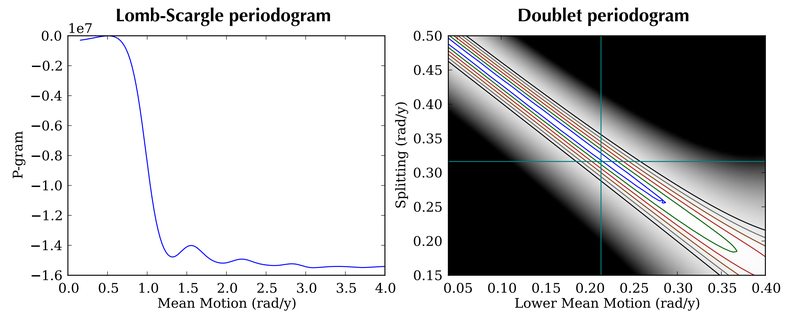
Time-resolved transforms

We are developing a suite of tools that manipulate the correlation function to ***quantify evidence for local structure*** in time series. We implement ***multitapering*** to control power spectrum leakage. Some tools are applicable to ***irregularly sampled data***, via use of the discrete correlation function. The example below shows ***time-frequency distributions*** for 33 yr of full-disk solar Ca II Kline emission index data from Sacramento Peak National Solar Observatory (Keil & Worden 1984), computed by Fourier transforming TSE autocorrelation functions of the data in a sliding time window. Multitapering significantly enhances the signal at the solar rotation period (apparent at 13.5 cy/yr).



New periodograms

We are developing ***multiplet periodograms*** capable of separating signals with similar periods that are not distinguishable in a standard periodogram or Lomb-Scargle periodogram, applicable to unevenly-sampled data. The figure below shows a generalized Lomb-Scargle periodogram and a *doublet periodogram* applied to simulated RV and astrometry data from a Jupiter+Saturn system; the RV data span 15 yr and the astrometry data span 10 yr; they are irregularly and asynchronously spaced. The two planetary periods (11.9 and 19.5 yr) cannot be resolved by a standard periodogram (left). The doublet periodogram (plotted vs. the lower period and a possible period splitting) resolves the two signals.



Parametric inference engine (PIE)

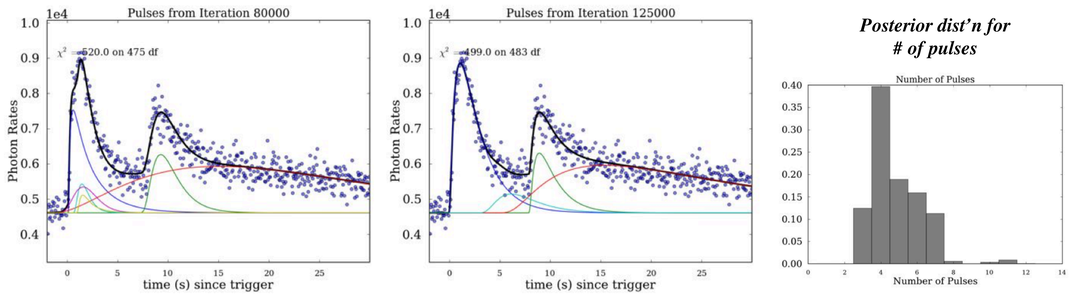
PIE is a flexible Python framework for modest-dimensional parametric modeling of time series and other data using three approaches: ***weighted least squares*** (e.g., $\chi 2$ minimization), ***maximum likelihood***, and ***Bayesian***. Users build an inference class by mixing inference and predictor classes. The framework enables moving between approaches by changing a mixin.

Currently supported inference tools include:

- ***Optimization*** (with the SciPy suite), including ***subspace optimization*** over a grid in the complementary subspace (e.g., for profile likelihood and profile posterior computation)
- ***Laplace approximation*** (with numerical Hessian)
- Bayesian inference via ***cubature***
- ***MCMC-based posterior sampling*** with multiple methods (single-chain & population-based)

Ex.: Bayesian droplet decomposition of GRB 931008 (BATSE #2571) into "Norris kernel" pulses.

Two representative decompositions from parallel thinning MCMC



CUDAHM: CUDA for Hierarchical Models

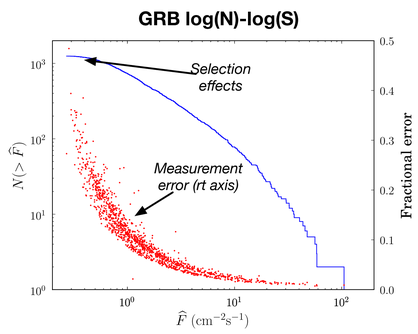
Work by Loredo & Budavari, with Brandon Kelly (UCSB) & Janos Szalai-Gindl (Eotvos U.) This work has been funded by the NSF via grants AST-1312903 (Cornell) & DMS-1127914 (SAMSI).

The problems CUDAHM solves

CUDAHM ***estimates distributions describing cosmic populations*** using data that have ***measurement errors*** and possibly also ***selection effects***.

Examples:

- ***Number-size distributions*** ("log(N)–log(S)") or number counts distributions from flux data
- ***Luminosity functions*** (possibly z-dependent) from flux and redshift data
- Bivariate/multivariate ***joint distributions*** (correlation studies) where all variables have measurement error



CUDA + HM

CUDA (Compute Unified Device Architecture): A parallel programming framework by Nvidia for general-purpose computing with ***graphics processing units*** (GPUs).

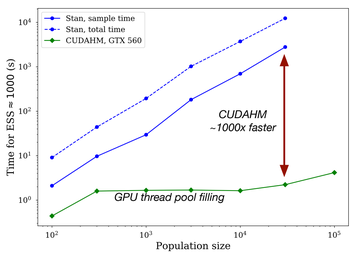
Hierarchical Bayesian model (HBM, aka probabilistic graphical model, Bayes net): A probabilistic model for data built by composition of simple stochastic components—e.g., describing a population, measurement error, selection effects. Replication (e.g., across a population) ⇒ hierarchical structure.

CUDAHM is a C++ framework for implementing HBMs on GPUs to enable ***cosmic demography scalable to large populations***.

Benchmark

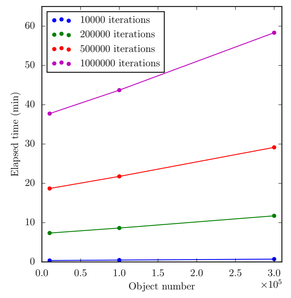
Normal-Normal model = A population modeled with a Gaussian distribution, with each member having measurements with noise described by separate Gaussian distributions. It's a standard textbook HBM.

We compare CUDAHM to *Stan* (http://mcstan.org/), a probabilistic programming language that produces a compiled C++ library implementing a state-of-the art Hamiltonian Monte Carlo algorithm for the user's model. Stan is easier to use than CUDAHM, but does not fully exploit the parallel structure of HBMs nor use GPUs. Shown is the time to produce 1000 effectively independent posterior samples, as a function of population size. Higher = Slower.



Demonstration

We simulated data from a smoothly truncated Schechter-function-like luminosity function, with errors and selection effects mimicking SDSS, and used CUDAHM to recover the luminosity function. Figure shows performance vs. population size and MCMC chain length. Parametric inference with models with several parameters, for catalogs with ~1e8 objects, should be achievable with computing times of order a day with a modest-sized GPU clusters.



See Szalai-Gindl et al., "A GPU-Accelerated Hierarchical Bayesian Method for Estimating the Luminosity Function from Flux Limited Observations with Photometric Noise," *Astron. & Comp.* (in press)