

Supplementary Materials for

Lidar reveals activity anomaly of malaria vectors during pan-African eclipse

Mikkel Brydegaard*, Samuel Jansson, Elin Malmqvist, Yeromin P. Mlacha, Alem Gebru, Fredros Okumu, Gerry F. Killeen, Carsten Kirkeby*

*Corresponding author. Email: mikkel.brydegaard@fysik.lth.se (M.B.); ckir@sund.ku.dk (C.K.)

Published 13 May 2020, *Sci. Adv.* **6**, eaay5487 (2020)
DOI: 10.1126/sciadv.aay5487

The PDF file includes:

Supplementary Methods
Figs. S1 to S4
Table S1
Legend for data file S1

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/20/eaay5487/DC1)

Data file S1

Methods

Field campaign

Field monitoring was carried out in the vicinity of Lupiro village during the middle of the dry season, August 31st to September 5th in 2016, including during the pan-African solar eclipse. There was no precipitation and little wind. The weather was relatively stable with patchy clouds. Light smoke from cooking and agriculture appears in the lidar signal but was removed by data processing. Temperature minima of ~19°C occurred at 6:00 and ~28°C maxima at 16:00. Tanzania is at UCT+3 (Coordinated Universal Time). At the site, during the campaign, sunrise occurred 6:34, solar noon occurred 12:32, sunset occurred 18:31 and true midnight was 0:32. On 1st Sep. moonrise was at 6:27 and moonset was at 18:40, with a back-lit new moon of 0% illumination at 392965 km distance. On the 4th Sep. moonrise was 8:34 and moonset was 21:00 with 9% illumination at 402322 km distance.

Lidar transect

The lidar transect was terminated on a black neoprene sheet at a distance of 596 m (8°22'44.93"S, 36°40'31.39"E, 306 m ASL). The beam propagated 3-5 meters above ground and intersected elevated footpaths at 85, 124, 152, 230, 357 and 450 m range. The instrument was deployed in a hut at the periphery of the Lupiro village (8°23'3.74"S, 36°40'26.66"E, 308 m ASL), overlooking the nearby agricultural patches, and was powered by a small 2kW motor-generator throughout the experiment, see *Fig.1* and *Fig.SF1*.

Lidar configuration

A lidar with Scheimpflug configuration with infinite focal depth^{29,37,40,46-49} was used. This enables very high sample rates for modulation analysis of remote insect targets^{31-33,50}. Monitoring was performed with an invisible near-infrared beam, transmitted from a 3W laser diode emitting at a wavelength of 808 nm with vertical polarization. The average reflectance of aerofauna, including *Anopheles*, in this spectral region is around 20%^{40,51,52}. The beam is expanded by a f600mm, ø120 mm refractor and focused into a 2.5x23.3 cm line at the distant termination target (toothpaste shaped beam).

With the F/5 beam expander in this study, the coupling efficiency is ~25%. Total emitted intensity is estimated to 400 mW CW or 457 µJ pulse energy. The intensity at the aperture is 3.5 mW/cm² CW or pulse energies of 4µJ/cm². Intensity at the termination is 17 mW/cm² CW and pulse energy 19 µJ/cm². The Maximum Permitted Energy (MPE) is 4 µJ/cm² and 808 nm and 300 µs for pulsed consideration and 20 mW/cm² at 808 nm and 300 µs for CW consideration. The beam is not considered eyesafe. The beam is inaccessible to the general public and has minimum altitude over the ground of 3 meters. A surveillance camera with near-infrared filter and sensitivity confirms the beam stability at the termination target at all times.

The backscattered light was collected by a f800mm, ø200mm Newtonian reflector telescope. The baseline transmitter-receiver separation was 814 mm and positioned vertically. After being transmitted through an 808 nm, 10 nm FWHM band-pass filter, light is detected with a linear 2048 pixel CMOS detector in Scheimpflug configuration at 45° tilt. The sensor was operated at a 3.5 kHz line rate and the laser was modulated on even- and odd exposures for sunlight subtraction throughout the day.

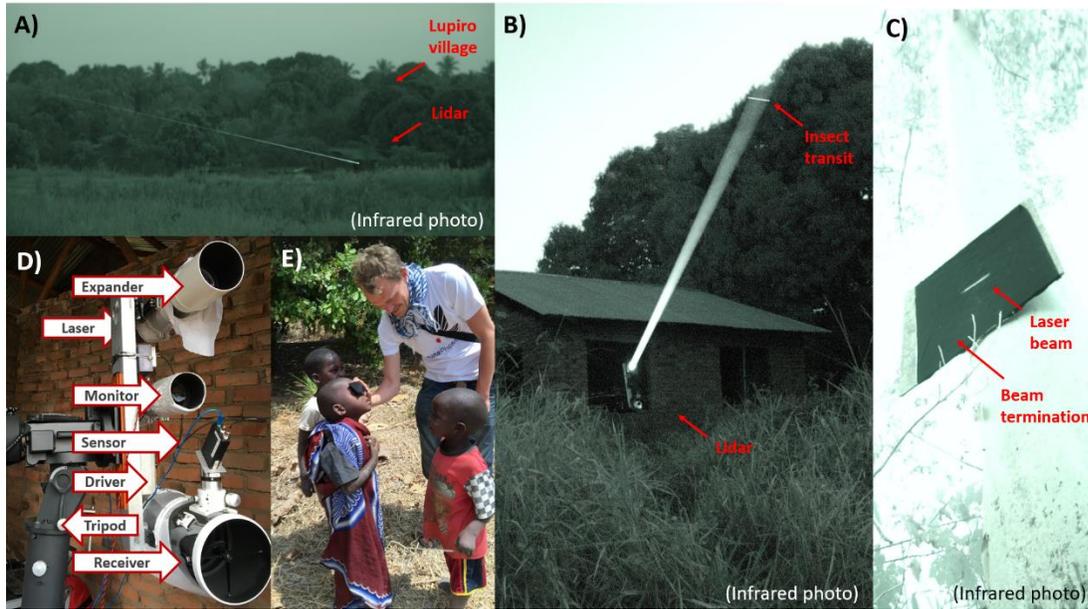


Fig. SF1. Scenes from the campaign captured by the first author. A) Panoramic view of the agricultural land towards Lupiro and the lidar. B) Insect leaving an infrared trace when transiting the laser beam. C) The beam termination on a neoprene-covered board at 596 m distance, installed 5 m above ground on a tree trunk. D) The lidar instrument and various components outlined. E) Outreach while the experiment is running.

Data, calibration and estimation of target modulation power spectrum.

Raw data was saved in frames with 35000 exposures spanning 10 s, each file 140 Mb. The entire data set was stored on several USB3 hard drives and was approximately 6 Tb. The dark or background exposures were identified by comparing even- and odd exposures, and hereafter subtracted (synchronized lock-in detection). In **Fig.2** three fractions of the raw data display three particular observations of active mosquitoes during the eclipse.

Range was calibrated based on triangulation and precise knowledge of the mechanical details of the lidar system components, global position coordinates (GPS) of both lidar and termination target and identification of termination echo pixel on the linear sensor. The range resolution deteriorates linearly by range and the precision is roughly 3%. The main source of this uncertainty is the beamwidth. For details, see^{29,53,54}.

Backscatter cross-sections is derived by backscattered light intensity. It is the product between geometrical projected area and reflectance, $\sigma_t = A_t * R_t$, where reflectance is relative a Lambertian white target. This quantity was calibrated by knowledge on the reflectance of the neoprene termination target (1.8% Lambertian reflectance at 808 nm), and estimation of the beam and pixel footprint at the target. Following this fixpoint for the backscatter cross-section, insect backscatter cross-sections were calibrated by back-extrapolation towards the lidar with the assumption of $1/r^2$ dependence and homogeneous atmosphere. For details, see⁴⁷. The quantity displayed in **Fig. SF3** is the time median during each observation. Mosquitoes constitute the smallest classes, and the values roughly correspond to previous estimates^{33,51}.

An alternative size measure, unrelated to reflectance, is the apparent size. This quantity is derived from the product of distance and opening angle (pixel spread). Remote small insects constitute point scatter sources, whereby the point spread function of the telescope is observed. However, larger insects at close range, constitute an opening angle in conjunction with the beamwidth. This causes observations to

spread over various pixels on the sensor. Since the distance is provided by the lidar, this opening angle can be converted to millimeters. For details on size, estimations see^{40,47}.

Data were divided into static signal and rare constituents by descriptive statistics and insect observations exceeding a signal-noise-ratio (SNR) of 2 were selected. This implies that the expectation value of static atmospheric range dependent echo is estimated by the temporal median for each 10 s frame. The static echo originates from aerosols, moisture, and molecules in the air. The static echo also reflects the range dependent overlap function of laser intensity illuminating each pixel footprint. To filter out occasional plumes from burning in agriculture and cooking the temporal median was following calculated from the entire measurement period. In analogy with the temporal median, the instrument noise was estimated by the temporal inter-quartile-range (IQR). The noise is predominantly read-out noise, but also contains dark-current, background-noise from sunlight and atmospheric turbulence. This noise is symmetric and displays a Gaussian distribution. This implies that the noise amplitude is 2.355 times the standard deviation and 3.491 times the IQR. Insect observations of $SNR > 2$, implies that the maximal signal strength from an insect observation exceeds 2 noise amplitude for the given range, for further details see⁴⁸. We counted 312191 observations throughout the study period, where we define an insect observation as a connected patch of pixels and exposures in the range-time frame. For the purpose of classifying insects and estimating their modulation power spectrum, we selected observations exceeding the most common beam transit time, which was 23 ms in this experiment (see **Fig.SF2**). Hereafter, 233660 observations (75%) were considered for further analysis.

The power spectral density was estimated by Welch's method in Matlab (MathWorks®, USA). We used a Gaussian time window of 23 ms FWHM to avoid side lobes. The window vector had 41 elements since an odd number of elements is needed for minimum width and symmetry. Modulation power was estimated in 41 frequency bins spaced 22 Hz between bin centers. The minimal observable frequency was 44 Hz (one period during 23 ms) and the Nyquist frequency was 875 Hz (half sampling rate after background subtraction and a quarter of the sensor line rate). We identified anomalies in the outer most frequency bins, presumably, this relates to frequency folding, bin centers and edges of the implementation of the Welch method. The outermost frequency bins were therefore excluded from the analysis. This left us with spectral power density estimates in 39 frequency bins from 65 Hz to 854 Hz (included in the data file).

We synthesized a 10 s sinusoid, coinciding with the central frequency bin center, and fed it through identical Gaussian window width based on most likely transit times of 23 ms. This yielded a spectral resolution of 83 Hz FWHM comparable to the Rayleigh resolution criterion. This resolution estimate is added to **Fig.3**.

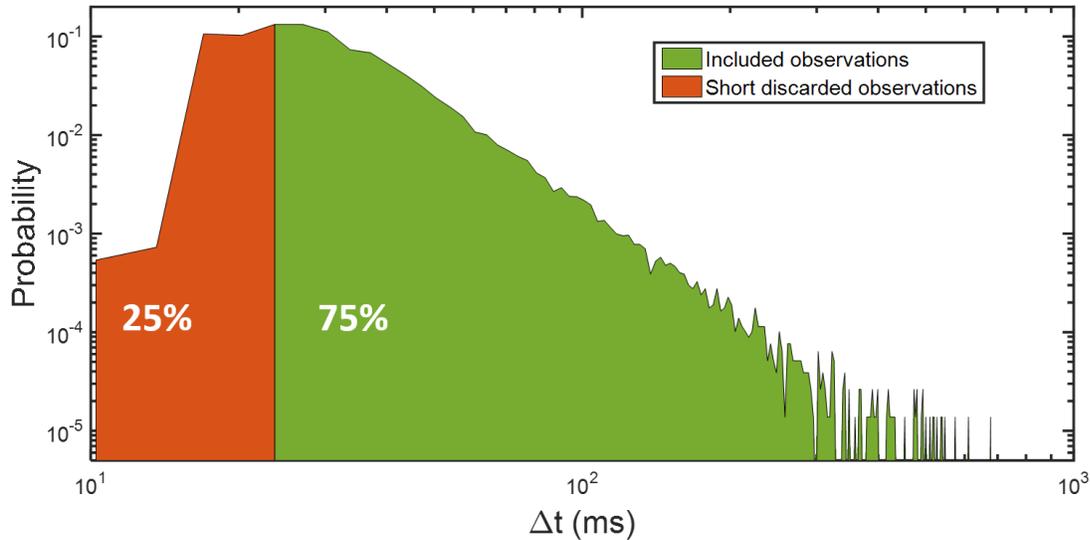


Fig. SF2. Histogram of all observed insect transit times. The most common transit time was 23 ms. Short observation was discarded and spectral analysis was applied to the remaining 75% of the observations.

Clustering and interpretation

Each modulation spectrum was normalized by a corresponding pink noise model (noise decrease with frequency). The models have the form; $\log(P_{noise}) = k_0 - k_1 f$, where P_{noise} is the power spectrum of adjacent empty data and the two model coefficients, k_0 and k_1 , is found by regression. Noise models were representative for each observation with respect to range and time of the day. The normalized modulation spectra of all observations were arranged in a 233660x39 matrix, powers were logarithmized and for the purpose of Hierarchical Cluster Analysis (HCA), Euclidean distance was calculated for all observation pairs in a 39-dimensional space (linkage). The euclidean distance of logarithmized values implies a logical conjunction operation, where simultaneous contents of all frequency bins in a cluster are similar. The pairwise linkage computation is demanding but was accomplished in Matlab (MathWorks®, USA) using the ‘ward’ and ‘savememory’ algorithm flags. We visualized the 20 first branches of the dendrogram (**Fig.4**), by plotting the centroid modulation spectra for each branch, using the within-group median spectrum, (**Fig. 4**). The spread was visualized by the within-group interquartile range (IQR). Details within the group spread were not interpreted. The identity of the clusters was primarily inferred by the fundamental frequency. The fundamental tones and their harmonics were subjectively added to the graphs by an experienced examiner. In some cases the fundamental tone was ambiguous because of sample folding, these cases were omitted from further interpretation. Mosquitoes, however, have a particularly high wing beat frequencies, and frequencies of the possible target sex and species are available from the literature^{31-35,55}. The associated mosquito species were selected among the species present in the area at the time of the experiment; see the survey in the next paragraph. The frequency ranges and FWHM were added to **Fig.3**. Coarse values for temperature shifts for morning and evening temperatures were estimated³⁴. Coarse estimations for frequency shift due to the payload for blood-fed and gravid mosquitoes were also estimated^{43,55,56}. The various shifts are indicated in **Fig. 3**. Apart from frequency content, some clusters in **Fig. 4** were labeled according to particular behavioral features such as the daily pattern (twilight/morning insects) or range distribution (far insects), as well as signal properties such as the backscatter cross-section (large) or the reflectance (bright), see **Fig.SF3**.

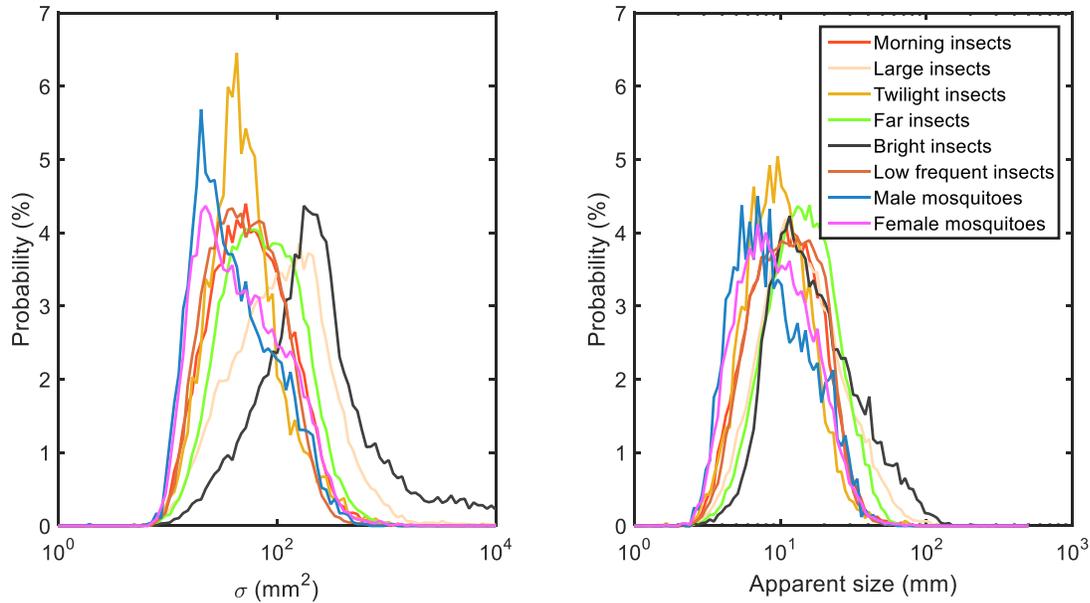


Fig. SF3. Size distributions for hierarchical clusters. A) Distribution of backscatter cross-section, σ , derived from signal intensity for each group. B) Corresponding apparent sizes derived from pixel spread. The two quantities displayed here is the median during each observation, for details see⁴⁰. The integral of each graph is 100%.

Previous laboratory studies of photonic mosquito classification³¹⁻³³, have reported high accuracies in the range of 80-95% using Naïve Bayes Classifiers (NBCs). However, these studies only include a few species and the accuracy diminishes when similar species are included in the studies. In this *in situ* study a high number of species can be expected to intercept the probe volume and the true identity of each observation is not available. In order to provide some clue on the possible overlap of the clusters in **Fig.4** independent Gaussian distributed modulation powers in the 39-dimensional space of frequency bins were assumed. Since Gaussian distributions range from minus infinity to plus infinity, and since, modulation powers are positive definite, then modulation power was logarithmized. This is in analogy with the logarithmic powers fed to the HCA in this study. The NBC tool in Matlab (MathWorks®, USA) was used to fit 39-dimensional Gaussian distributions to the logarithmized modulation powers of each of the 20 clusters in **Fig.4**. The same toolbox was used to evaluate the confusion matrix and Gaussian overlap between the clusters, see **Fig. SF4**. The estimated accuracy and overlap of the hierarchical clusters are seen in the right pane. Many of the clusters, which were interpreted similarly, display the largest confusion; therefore, the accuracy for the grouped clusters is better than accuracies for the individual clusters. One observation is that NBC performs worse than random for the cluster C4 (female mosquitos). It is concluded that clusters identified by the HCA are not necessarily Gaussian distributed, they do not necessarily have independent variables and they are not necessarily spherical or ellipsoids in the 39-dimensional frequency space. This is in accordance with our previous understanding³⁷; the frequency content and overtones from a single species and sex are governed by spherical functions according to the observation angles, therefore modulation powers across the frequency domain are neither independent nor linearly related, even for the same species and sex. **Fig. SF4** therefore also largely reflects differences between HCA and NBC rather than actual accuracy of the clustering.

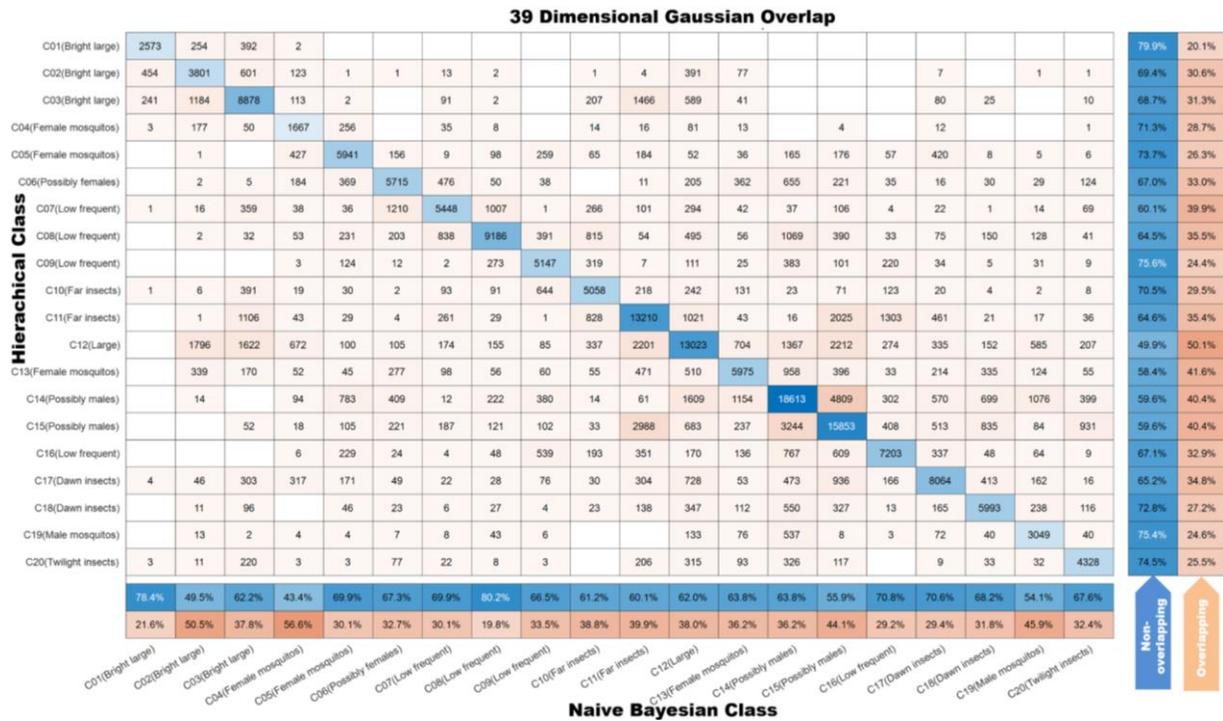


Fig. SF4. Confusion matrix and estimation of possible overlap between clusters from Hierarchical Cluster Analysis in Fig. 4 assuming Gaussian distribution and independent variables.

Comparisons to mosquitoes detected in homes

The indoor densities of the local malaria mosquitoes were measured throughout the course of the experiment using miniature CDC (Center for Disease Control, USA.) light traps. These traps were placed indoors to provide a daily level of host-seeking mosquitoes. Three houses from the village close to the lidar equipment (within ~50 m, at the beginning of the lidar transect) were selected. The traps were placed beside occupied bed nets on the foot side. Humans sleeping inside the bed nets acted as bait for human-seeking mosquitoes. The traps were turned on at 18:00 and emptied the following mornings at 6:30. In a small laboratory, located in the Lupiro village, the collected mosquito samples were identified to species level based on morphological characters following the criteria of Gillies and Coetzee⁵⁷. Specimens identified as members of the *An. gambiae* complex⁵⁸ or *An. funestus* group⁵⁹ were identified to species by Polymerase Chain Reaction (PCR). Blood meal analysis was performed by Enzyme-Linked Immunosorbent Assay (ELISA). Caught individuals were unfed but the results are not representative for free-flying mosquitoes in general and therefore not this analysis is considered in this study.

Nights 2016 Sampling hours 18:00-6:00	An. gambiae s.l	An. funestus	An. coustani	Culex spp	Mansonia spp	Coquilettidia spp
29.Aug-30.Aug	169	0	0	36	0	0
30.Aug-31.Aug	182	0	0	34	3	0
31.Aug-1.Sep	514	2	4	149	6	0
1.Sep-2.Sep	481	4	7	126	17	3
2.Sep-3.Sep	536	5	5	161	8	0
3.Sep-4.Sep	152	1	0	74	1	0
4.Sep-5.Sep	482	1	1	279	11	3
5.Sep-6.Sep	200	0	0	45	4	2
6.Sep-7.Sep	202	0	0	54	3	0
8.Sep-9.Sep	176	0	0	48	0	0
Total	3094 (74%)	13 (0.3%)	17 (0.4%)	1006 (24%)	53 (1.3%)	8 (0.2%)

Table ST1. Numbers of mosquito specimens caught by light traps per night. The bold line indicates the day of the eclipse. Specimens were identified to species or species complex based on morphological characters and the speciation for anophelins by PCR-analysis.

Supporting data.

The data required to reproduce the presented results are included in a Matlab file, *SDI* (MathWorks®, USA). The content is as follows.

Variable name	Unit	Size	Explanation
<i>t</i>	<i>Absolute time</i>	233660 obs	<i>Time stamp</i>
<i>dt</i>	<i>s</i>	233660 obs	<i>Transit time through beam</i>
<i>r</i>	<i>m</i>	233660 obs	<i>Range from lidar</i>
<i>sigma</i>	<i>mm²</i>	233660 obs	<i>Scattering cross section compared to Lambertian white</i>
<i>apSize</i>	<i>mm</i>	233660 obs	<i>Apparent size</i>
<i>F</i>	<i>Hz</i>	39 bins	<i>Frequency bins for modulation signature</i>
<i>P</i>	<i>dB</i>	233660 obs x 39 bins	<i>Matrix including power spectra for each observation</i>
<i>T</i>	<i>Class # 1..20</i>	233660 obs	<i>Classes from hierachical clustering</i>
<i>Z</i>	<i>Euclidean distance</i>	233659x3	<i>Linkage structure</i>
<i>clabel</i>	<i>Strings</i>	20 clusters	<i>Interpretations for each cluster</i>
<i>cf0</i>	<i>Hz</i>	20 clusters	<i>Estimated fundamental frequency for each cluster</i>
<i>fs</i>	<i>Hz</i>	<i>scalar</i>	<i>Sampling frequency</i>