Nonparametric Approach to Weak Signal Detection in the Search for Extraterrestrial Intelligence (SETI)

Anne D. Brooks[1][0000-0002-9163-656X] and Robert A. Lodder[2][0000-0001-6133-7561]

[1] Stetson University, DeLand, FL 32723, USA
[2]University of Kentucky, Lexington, KY 40536, USA
Lodder@uky.edu

Abstract. It might be easier for intelligent extraterrestrial civilizations to be found when they mark their position with a bright laser beacon. Given the possible distances involved, however, it is likely that weak signal detection techniques would still be required to identify even the brightest SETI Beacon. The Bootstrap Error-adjusted Single-sample Technique (BEST) is such a detection method. The BEST has been shown to outperform the more traditional Mahalanobis metric in analysis of SETI data from a Project Argus near infrared telescope. The best algorithm is used to identify unusual signals and returns a distance in asymmetric nonparametric multidimensional central 68% confidence intervals (equivalent to standard deviations for one D data that are normally distributed, or Mahalanobis distance units for normally distributed data of the dimensions). Calculation of the Mahalanobis metric requires matrix factorization and is order of d^3. Furthermore, the accuracy and precision of the best metric are greater than the Mahalanobis metric in realistic data collection scenarios (many more wavelengths available then observations at those wavelengths). An extension of the BEST to examine multiple samples (subclusters of data) simultaneously is explored in this paper.

Keywords: parallel algorithm, bootstrap, supernova, gamma ray burst, solar transit

1 Introduction

1.1 Scope of the Problem

SETI is at least a complex 5-dimensional problem. Five dimensions is a lot of space to search. The first three dimensions, length, height, and width, are the (X, Y, Z) spatial coordinates that everyone is used to in daily life. The fourth dimension is frequency or wavelength. The system must be listening at the right optical wavelength or microwave frequency in order to detect a signal. Time is the fifth dimension. In addition to looking in the right place, and listening at the right frequency, the system also must be listening when the signal comes in. Five dimensions is a lot of space to search, and this problem partially explains why finding signals has been so difficult.

Moreover, the rotation of planets and the revolution around stars means that transmitting and receiving antennas rarely line up. In fact, the drift scan transit time of high gain receivers and antennas on earth is usually on the order of seconds. Finally, the Doppler shift from planetary movement complicates signal averaging to increase signal-to-noise ratio. It is well-known that signal adds as n while random noise adds as the square root of n (where n=the number of times the signal is measured, or the signal integration time)[Downs, 2012]. This fact is used to increase the signal-to-noise ratio by a factor of the square root of n by averaging signals over time. However, the Doppler shift imposed on signals by planetary motion is enough to limit the averaging of signals because the signals are moving [SETI@Home, 2019]. Doppler shifts can be compensated for in software, and most SETI systems that work off-line are able to apply dozens or even hundreds of different Doppler shifts in the signal averaging process in order to enhance weak signals.

Detecting a signal is also a somewhat complex process. The statistical hypothesis tested in software tests the hypothesis that no intelligent signal is present against the alternative that a signal from ETI has been detected. Before signal collection can begin, every station must verify proper system operation with a signal generator or another celestial source. This procedure is generally repeated at the end of the data collection run. During a data collection run, an ETI signal should show:

1. Coherence not achievable by known natural emission mechanisms.
2. A signal intensity variation that is consistent with the known antenna pattern and the aiming coordinates (azimuth and elevation). A directional signal should drop in intensity when a directional antenna is moved away from the signal source.
3. A Doppler shift consistent with planetary motion (or the motion of a reasonable object in space, like a spaceship.) Satellites can be mistaken for ETI signal sources, but they show a Doppler shift that changes with their angular velocity.
4. Finally, before any signal detection can be announced, there must be a simultaneous detection with proper Doppler shifts at widely separated terrestrial coordinates. Ideally, this detection takes place at a station on the other side of the earth, where the same terrestrial sources of interference would not be present. Even an interfering signal from an airplane or satellite would be unlikely to hit two SETI telescopes on opposite sides of the planet at the same time. If the satellite was far enough out in space to hit two SETI telescopes on opposite sides of the planet at the same time, it would likely still show up at different celestial coordinates (right ascension and declination).

1.2 Reducing the Scope

One way to collapse the five-dimensional problem of signal detection into a more tractable problem is to use time and direction synchronizers in SETI. Reducing the dimensionality of the problem is possible with appropriate synchronizers that attract the attention of scientists. A synchronizer should be a big enough event to attract Galactic attention. Supernovas and gamma ray bursts fit the requirement (see Fig.

1)[Corbet, 2000]. For example, type IA supernovas are used as standard candles to measure the expansion of the universe. They are interesting to our scientists and are probably interesting to alien scientists for the same reasons. When one of these events occurs in the Milky Way galaxy or even in another faraway galaxy (for example, Supernova Refsdal is in a galaxy 9.3 billion light-years away from the Milky Way galaxy and earth, making it a good target for measuring the expansion of the universe), the light will eventually reach an alien planet on the way to earth. At that time, ETI will direct its transmitters in a direction roughly collinear to the received supernova light and away from the supernova. ETI will also probably direct its signal in a small cone so that the image of their Beacon appears to the side of the image of the supernova. The light from the alien Beacon and the supernova should arrive at the earth at the same time. In this way, ETI wishing to advertise the presence of a Galactic Internet could take advantage of a high-energy signal source to set four of the five variables in the SETI search space (the x, y, and z coordinates in space as well as the time coordinate), leaving only the frequency or wavelength variable to be determined.

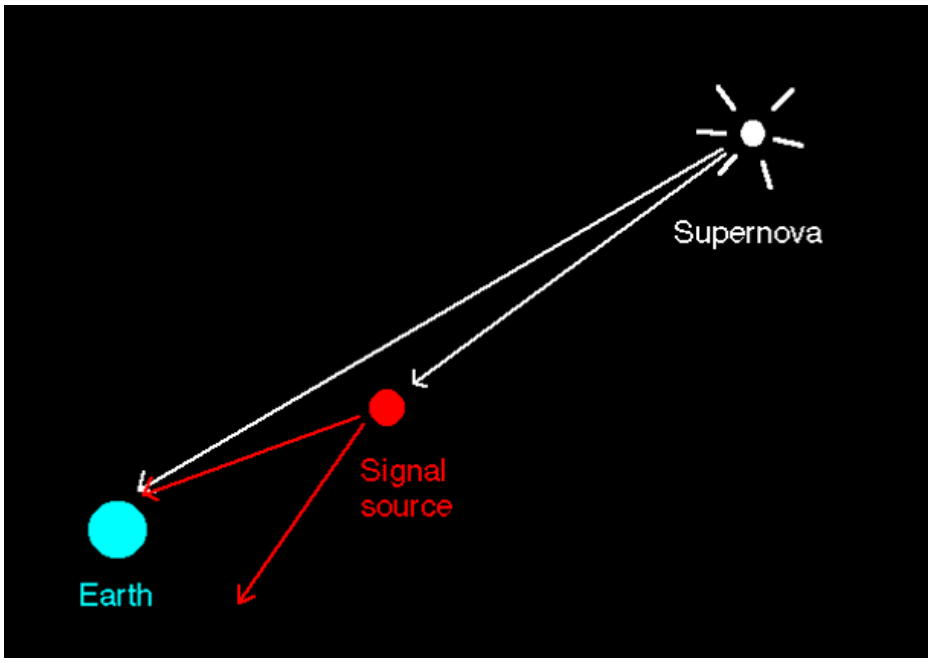

Fig. 1. The supernova SETI synchronizer strategy

Another possible SETI synchronizer is the solar transit. This synchronizer takes advantage of the solar eclipse. When the Earth passes in front of the sun, it blocks a small part of the sun's light. Potential observers outside our solar system might be able to detect the resulting dimming of the sun and study the Earth's atmosphere. This

transit method has helped to find most of the thousands of exoplanets known to exist today.

The last variable in the five-dimensional space, the frequency or wavelength, is worth special consideration. Most SETI research has been done in the microwave region since Frank Drake's Project Ozma in the early 1960s (Drake, 1963). Microwave SETI searches for pulses of electromagnetic radiation within the microwave portions of the spectrum. The beam is wider in the microwave region and thus targeting is not as much of an issue. Historically, microwaves have been viewed as better able to penetrate the atmosphere and thus more likely to be used for interstellar communications. However, developments in the last decade or two in adaptive optics have made arguments for microwave SETI far less convincing. Optical SETI searches for pulses of laser light in or near the portion of the light spectrum that is visible. The beam is narrow, enabling a higher power density to be directed toward a distant target. The use of near-infrared and infrared light in certain bands enables the energy to both escape the atmosphere and avoid being blocked by dust in the galactic disk.

The advent of adaptive optics [Beckers, 1993] has given great impetus to near-infrared and optical SETI. Adaptive optics was developed for destroying incoming ICBMs by President Reagan's Star Wars program. Originally telescope mirrors were very thick to keep them from bending and going out of focus when the telescope was angled in a new direction. A 1 to 6 ratio was commonly used in mirror construction (in other words, a 6-inch mirror had to be 1 inch thick, and a six-foot mirror had to be 1 foot thick) in order to be mechanically stable in all directions. A 10 m mirror built in this way would be huge and impractical. Optical telescopes began to approach the 10 m size when the mirrors were instead made very thin and lightweight and were connected to an array of electromechanical actuators to bend in the mirror back into shape when the telescope was moved to a new angle (an electronically deformable mirror). The Star Wars scientists realized that distortion caused by refraction in the Earth's atmosphere could be corrected by these electromechanical actuators if the actuators could be moved at high speed (e.g., 1000 times per second).

Adaptive optics use an artificial guide star directed toward a layer of ions above most of the Earth's atmosphere (and certainly above the turbulent part). This artificial guide star is created by a laser on the surface of the earth that excites fluorescence in the ions. This laser is placed next to the large telescope (i.e., 10 m mirror telescope) on earth and excites ions in the field of view of the telescope. A computer is then programmed to deform the mirror at ~1000 times per second to correct the shape of the artificial guide star, which is distorted by atmospheric turbulence, back to the original shape transmitted by the laser on earth. The distortions required to correct the original shape of the artificial guide star also correct all the other turbulence in the optical path.

This correction does more than take away the twinkle of the stars when light is received. The correction can also be used to take away the twinkle of a transmitted sig-

nal. When humans are ready to join the Galactic Internet, a small laser to excite ions above the atmosphere will join a METI (Messaging Extraterrestrial Intelligence) laser on an adaptive optics telescope. The smaller laser will create the artificial guide star, and the larger METI laser will be directed toward the deformable mirror. The mirrors deformations will then cause the refractions in the turbulent atmosphere to rebuild the transmitted light beam in the process of exiting the Earth's atmosphere, leading to a clean signal transmitted to a distant planet or spacecraft.

Infrared light includes wavelengths too long to be visible, from approximately 700 nanometers to about 1 mm in wavelength. Visible light seen by the human eye ranges over about 400 to 700 nanometers in wavelength. Light is called near-infrared or near-ultraviolet based upon its proximity to the visible portion of the spectrum. So, near-infrared light is the highest in energy and the shortest in wavelength of the infrared region, while near UV light is the longest in wavelength and the lowest in energy of the ultraviolet region.

Most cosmic dust particles are between a few molecules to 100 nm in size. Near-infrared light penetrates the Milky Way galaxy better than visible light because of reduced scattering. For example, the star-forming region G45.45+0.06 is visible from earth at 2200 nm but obscured by galactic dust at 1250 nm. Light scattering falls off as one over wavelength to the fourth power. In other words, doubling the wavelength reduces light scattering by a factor of 16. Infrared light would be better than near infrared light, except that infrared light is absorbed more by the atmosphere of the earth. In a recent paper in The Astrophysical Journal, two researchers at MIT argue that it might be easier for intelligent extraterrestrial civilizations to be found when they mark their position with a bright laser beacon [Clark, 2018]. Given the possible distances involved, however, it is likely that weak signal detection techniques would still be required to identify even the brightest SETI beacon.

2 Experimental

Modern microwave and near-infrared/optical systems now often incorporate a software-defined radio (SDR). An SDR is a radio communication system where components that have been typically implemented in hardware (e.g., mixers, filters, amplifiers, modulators / demodulators, detectors, etc.) are instead implemented by means of software on a personal computer or embedded system. While the concept of SDR is not new, rapidly evolving digital electronics render practical many processes which used to be only theoretically possible. This approach greatly reduces the cost of instrumentation and is the approach we have adopted for our microwave and infrared SETI telescopes (Project Argus station EM77to). The software defined radio acts as a very sensitive spectrum analyzer, displaying the Fourier transform of the signals present at the InGaAs detector. In this slide, the center detection frequency is set at 147.5 MHz. The SDR# software displays all signals between 146.3 and 148.7 MHz in

single spectrum (top) and waterfall (bottom) mode. A Fourier transform converts signals in the time domain to signals in the frequency domain. In other words, a sine wave with amplitude on the Y axis and the time on the X axis appears in a graph as a single spike following Fourier transformation. The single spike appears at the frequency of the sine wave in a graph that still has amplitude on the Y axis, but now frequency on the X axis. An inverse Fourier transform converts signals in the frequency domain back into signals in the time domain. The Fast Fourier Transform (FFT) simply refers to an efficient algorithm for performing a discrete Fourier transform on data.

Our group uses two near-infrared telescopes, a 6-inch Newtonian reflector with all gold first-surface optics, and a one-meter Fresnel refractor with an aluminum compound parabolic concentrator. Both telescopes use Dobsonian az-el mounts. The fully assembled Newtonian reflector telescope is shown in Fig. 2 with the near infrared detector installed in the eyepiece. The handle on the primary mirror is visible in the end of the telescope. The telescope can be programmed using an ordinary laptop computer. The software defined radio attaches to the computer through a USB port. The computer currently runs Microsoft Windows 10.

Fig. 2. This near-infrared telescope for SETI uses all gold first surface optics and a high speed InGaAs detector. An SDR connects the detector to the computer for data monitoring and collection.

The detectors are high speed InGaAs photodiodes and the photodiode signals feed into the SDRs through coaxial cable. SDR#, an open source spectrum analysis program for SDRs, is used as the GUI for the telescope and for data collection.

Fig. 3 shows the effect of FM radio interference on signals near 105.18 MHz. Because FM radio signals are modulated, they do not appear as narrow spikes (i.e., a delta function). Because the signals are terrestrial in origin, they also appear brighter and stronger than we would expect a SETI signal to appear. Neither of the FM radio sta-

tion signals appear at the center frequency, which is shown by the red vertical line in the upper graph. Another clue that these signals are not from deep space is the absence of Doppler shift caused by the motion of the earth. An actual signal from deep space would also include a Doppler shift from motion of the source of the signal.

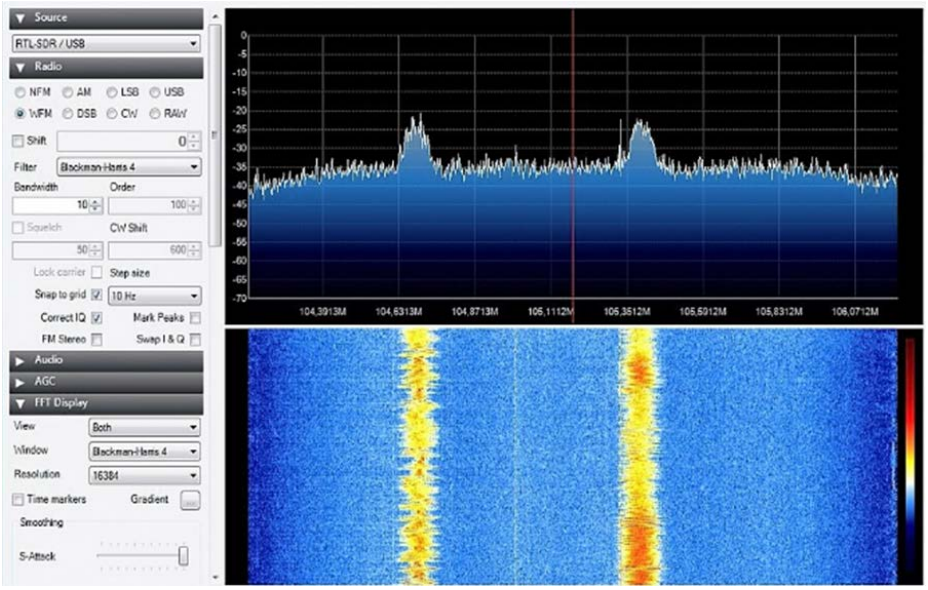

Fig. 3. An example of FM radio interference. The signals have a large bandwidth compared to a beacon. This red signal is not Doppler shifted, suggesting that it is terrestrial in origin. Good shielding will prevent this sort of problem.

Fig. 4 depicts a test using frequency modulated light pulses. The position of the pulses is varied by voice information. Side bands are seen around the central red signal. This red signal is not Doppler shifted, suggesting that it is terrestrial in origin.

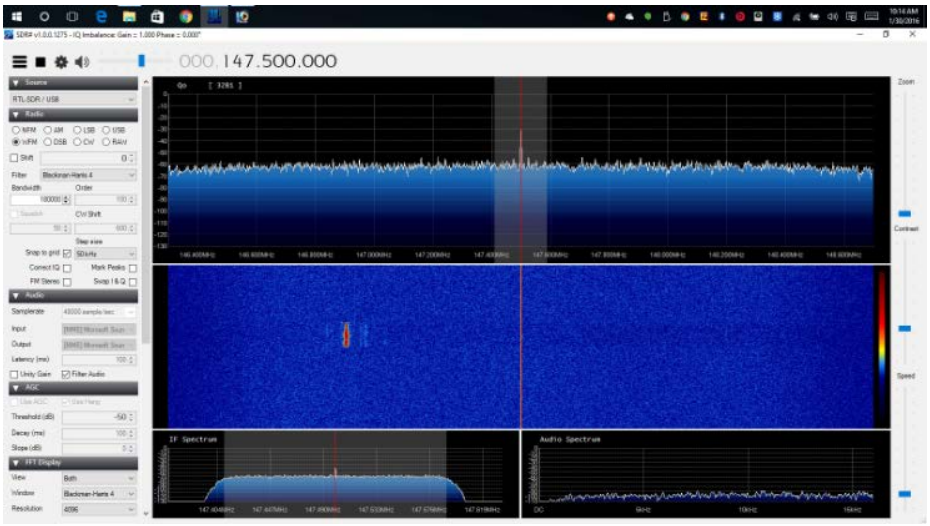

Fig. 4. A system test using frequency modulated light pulses.

In Fig. 5, the center frequency of 147.5 MHz is shown by the vertical red line and the delta function in the top graph. The weak, SETI-like signal appears as a vertical line to the left of the red line marking the center frequency of the receiver.

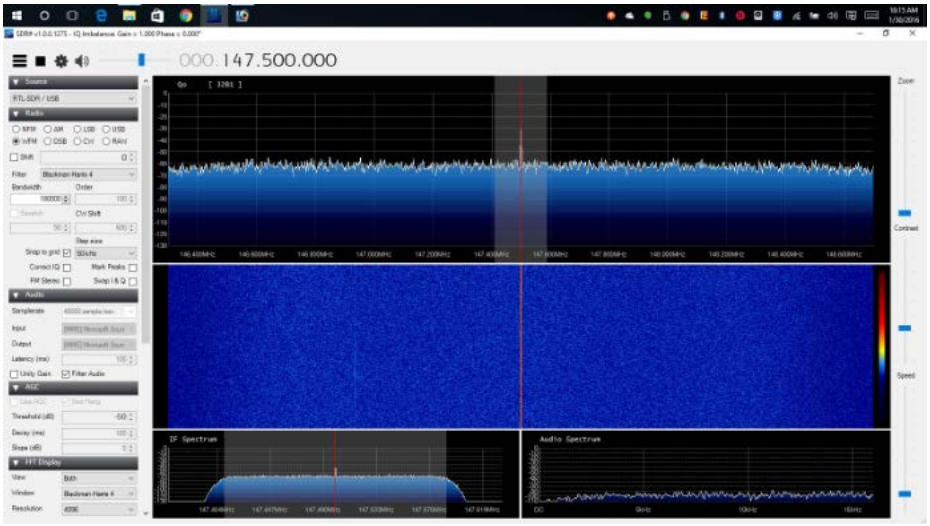

Fig. 5. A beacon-like signal with a more realistic intensity, but still lacking in Doppler shift

2.1 The Extended BEST for Subcluster Detection

The Bootstrap Error-adjusted Single-sample Technique (BEST) is a weak signal detection technique (Lodder, 1988)(the algorithm is summarized in the Appendix). The

BEST has been shown to outperform the more traditional Mahalanobis distance metric in analysis of SETI data from a Project Argus near-infrared telescope. The BEST algorithm is used to identify unusual signals, and returns a distance in asymmetric nonparametric multidimensional central 68% confidence intervals (equivalent to standard deviations for 1-D data that are normally distributed, or Mahalanobis distance units for normally distributed data of d dimensions). Calculation of the Mahalanobis metric requires matrix factorization and is $O(d^3)$. In contrast, calculation of the BEST metric does not require matrix factorization and is $O(d)$. Furthermore, the accuracy and precision of the BEST metric are greater than the Mahalanobis metric in realistic data collection scenarios (i.e., many more wavelengths available than observations at those wavelengths).

In near-infrared multivariate statistical analyses, ETI emitters with similar spectra produce points that cluster in a similar region of spectral hyperspace. These clusters can vary significantly in shape and size due to variation in signal modulation, bandwidth, and Doppler shift. These factors, when combined with discriminant analysis using simple distance metrics, produce a test in which a result that places a particular point inside a particular cluster (the training data are typically noise collected in a specific region of sky) does not necessarily mean that the point is actually a member of the cluster. Weak signal strength may be insufficient to move a data point beyond 3 or 6 SDs of a cluster. Instead, the point may be a member of a new, slightly different cluster that overlaps the first. This happens when the test data contain a weak artificial signal not present in the training noise. A new cluster can be shaped by factors like signal modulation, bandwidth, and Doppler shift. An extension added to part of the BEST can be used to set nonparametric probability-density contours inside spectral clusters as well as outside, and when multiple points begin to appear in a certain region of cluster-hyperspace the perturbation of these density contours can be detected at an assigned significance level. When we have more than a single point sample, it is possible that a larger sample of data points from the test set will produce a new cluster with a different mean and standard deviation that overlaps the training set. If we could collect a sufficiently large sample of these spectra, we might be able to detect a signal even inside the three standard deviation limit on single points from the training cluster center. To do this, the algorithm

- Integrates the training samples from the center of the training set outward, and
- Integrates the test samples AND the training samples combined from the center of the training set outward.

These two integrals are compared in a QQ plot. The detection of candidate ETI signals both within and beyond 3 SDs of the center of the noise training set is possible with this method. Using this technique, distinctive diagnostic patterns form in the QQ plots that are discussed below (see Fig. 6). These patterns have predictable effects on the correlation coefficient calculated from the QQ plots.

Fig. 6a depicts a pure location difference between the training set (noise) and the test set (noise and a signal). A pure location difference is the situation that might exist

when a fairly strong signal with no Doppler shift is detected. In this example, the two populations are identical except for their locations (centers). The shapes of the distributions have been arbitrarily selected to be circles (or hyperspheres in hyperspace of larger dimension) with the same standard deviation in all directions.

Fig. 6b illustrates the Cumulative Distribution Functions (CDFs) of the training set (blue) and test set (red) from (a). The x axis values represent the sorted normalized Euclidean distances of each point from the center of the training set.

The QQ plot of the Location Difference Only example in (a) is given in Fig. 6c. A correlation coefficient calculated for the QQ plot gives an indication of how well the two distributions (Training Set and Test Set) match. A correlation coefficient of r=1 indicates perfectly matching distributions, and no SETI signals when the algorithm is trained on galactic background noise. This QQ plot has a break in the line that would indicate the presence of a signal.

Fig. 6d. illustrates the effect of a Location Difference Only in (a) on the correlation coefficient of the QQ plot as the distance between the centers of the clusters increases. The horizontal line represents a confidence limit on the training set calculated with the use of cross validation samples. When the line drops below the confidence limit a signal has been detected.

A Scale Difference Only, in which the Training Set is larger than the Test Set, is depicted in Fig. 6e. This illustration shows a training set and a test set in hyperspace, the two population distributions share the same center, and the training set population distribution is larger in scale than the test set distribution. This situation would occur when there was no SETI signal present but the noise level in the receiver dropped.

Fig. 6f shows a QQ plot from a Training Set and Test Set that differ only in scale when the Training Set is larger than the Test Set as in (e). There are two bends in the plot that reduce the correlation coefficient calculated from the QQ plot.

The effect of a pure scale difference (when the Test Set is smaller than the Training Set) on the correlation coefficient calculated from the QQ plot as the scaling factor changes is shown in Fig. 6g. The x axis values represent the distance factor by which the test set is smaller in scale than the training set.

Fig. 6h shows a Scale Difference Only with the Training Set Smaller than the Test Set. Fig. 6h illustrates a training set and a test set in spectral hyperspace with the size relationship opposite to that just observed in Fig. 6e. The two population distributions share the same center, and the training set population distribution is smaller in scale than the test set population distribution. An increase in background noise or a modulated signal could cause this pattern to emerge in the data.

Fig. 6i reveals a QQ plot from the subcluster detection method corresponding to the pure scale difference situation in Fig. 6h. The test set is larger in scale than the training set, and a test set forms the lower line with the larger slope in the figure. The bend in the line is slight because the difference between the two set scales is only a factor of 2.5.

Fig. 6j. shows the effect of a pure scale difference (training set smaller than the test set) on the correlation coefficient calculated from a QQ plot as the distance scaling factor changes.

Fig. 6k illustrates the situation in which Simultaneous Location and Scale Differences exist and the Training Set is smaller than the Test Set. This is the situation commonly encountered when a signal is detected.

Fig. 6l. The QQ plot when a training set and a test set exhibit simultaneous location and scale differences, and the test set population distribution is larger in scale than the training set population distribution. There is both a bend and a break in the QQ plot line that lowers the correlation coefficient. The training set forms the lower line (blue) in the figure, and the test set forms the upper line (red).

Fig. 6m shows how the correlation coefficient is affected by changes in scaling factor and distance between the clusters when the Training Set is smaller than the Test Set. The highest line represents a test set that is a factor of 2 larger than the training set, the middle line a test set that is a factor of 5 larger than the training set, and the lowest line a test set: that is a factor of 10 larger than the training set. The horizontal line at the top of the graph is a 98% confidence limit. Only one test set crosses the 98% limit (meaning it is considered the same as noise), and that test set is a factor of two larger in scale than the training set, with the two set centers less than 0.5 standard deviation of the training set apart.

Fig. 6n. shows Simultaneous Location and Scale Differences with the Training Set Larger than the Test Set. A strong terrestrial signal could cause this effect.

Fig. 6o. shows the effect of simultaneous location and scale differences on the correlation coefficient calculated from a QQ plot when the test set is larger in scale than the training set. As in Fig. 6m, only when the test set is 2x the size of the training set is it ever identified as being the same as the training set, and then only when the two sets are about 0.1 SD of the training set apart. The test sets 5x and 10x the size of the training set are always identified as being different distributions (i.e., a signal is detected).

2.2 November/December 2018 Observations

Near-IR spectra from the vicinity of AT2018ivc, a supernova discovered in M77 on Nov. 24, 2018, were analyzed using the BEST subclustering method to identify unu-

sual signals. Observations were made on Nov. 26, 28, 29, 2018, and on Dec. 2 and 6, 2019 (2 Gb collected each night). All runs were negative. The collected data produced patterns similar to Fig. 6h with the horizontal line at the 99.9999% level. As usual, weather is the major limiting factor on data collection. Cloud cover and precipitation are problematic for optical and near-IR SETI methods.

3 Future Work and Conclusions

In the future our research will continue to focus on using the SETI synchronizer strategy based on supernovas and gamma ray bursts, and will introduce some solar transit synchronizer experiments. Planned upgrades to the microwave radio telescope system include the addition of a vacuum-sealed, liquid helium-cooled front end (low noise amplifier, mixer, and antenna probe). For some frequencies, it may be easier to omit the amplifier and send the antenna signal directly into the mixer to down convert it to a lower frequency, where lower noise gain is easier to achieve. An SIS mixer (Superconductor-Insulator-Superconductor) can introduce nonlinearity from quantum tunneling between the two superconductors, achieving low noise in the mixing process.

Collapsing the 5-dimensional SETI problem with synchronizers may never be proven the most fruitful approach to the search for extraterrestrial intelligent (ETI) life. Until we detect the first ETI (and in fact, many more) and know how those detections were made, it will be impossible to say with certainty what is the best approach. Until then, scientists need instrumentation and algorithms capable of collecting and processing increasingly large amounts of Big Data from their searches.

References

1. Beckers, J. M. Adaptive optics for astronomy - Principles, performance, and applications. Annual review of astronomy and astrophysics. Vol. 31 (A94-12726 02-90), p. 13-62 (1993).
2. Clark, J. R., & Cahoy, K.. Optical Detection of Lasers with Near-term Technology at Interstellar Distances. The Astrophysical Journal, 867(2), 97, (2018)
3. Corbet, R. H.. The Use of Gamma-Ray Bursts as Time and Direction Markers in SETI Strategies. In IAF, International Astronautical Congress, 51 st, Rio de Janeiro, Brazil. (2000)
4. Downs, R. 2012. Electronic Design. Understand the Tradeoffs of Increasing Resolution by Averaging. https://www.electronicdesign.com/print/51659. Last accessed 2019/03/06.
5. Drake, F. "How can we detect radio transmissions from distant planetary systems, Project Ozma. In “Interstellar Communication.”(AGW Cameron, ed.), Chapts. 16 and 17." (1963).
6. Lodder, R. A., Hieftje, G. M. Detection of subpopulations in near-infrared reflectance analysis. Applied spectroscopy, 42(8), 1500-1512 (1988).
7. SETI@Home 2019. More about signals. https://setiathome.berkeley.edu/ nebula/web/signals.php. Last accessed 2019/03/06.

4 Appendix

Method

A population **P** in a hyperspace **R** represents the universe of possible spectrometric samples (the rows of **P** are the individual samples, and the columns are the independent information vectors such as wavelengths or energies). **P*** is a discrete realization of **P** based on a calibration set **T**, chosen only once from **P** to represent as nearly as possible all the variations in **P**.

P* is calculated using a bootstrap process by an operation k(**T**). **P*** has parameters **B** and **C**, where C = E(**P**) (the group mean of **P**) and **B** is the Monte Carlo approximation to the bootstrap distribution [Lodder, 1988].

Given two data sets $\mathbf{P_1}^*$ and $\mathbf{P_2}^*$ with an equal number of elements *n*, it is possible to determine whether $\mathbf{P_1}^*$ and $\mathbf{P_2}^*$are drawn from the same population even if the distance between them is < 3 SDs (standard deviations). Quantile-quantile (QQ) plots and a simple correlation test statistic are used [Lodder, 1988].

$$\rho\left(\left\{\int_R \mathbf{P}_1^*\right\}, \left\{\int_R \mathbf{P}_1^*\right\} \bigcup \left\{\int_R \mathbf{P}_2^*\right\}\right)$$

A bootstrap method is employed to set confidence limits on ρ, the correlation coefficient. The central 68% confidence interval on ρ is also used to calculate σ_ρ, a distance in SDs that is sensitive to small differences in location and scale between $\mathbf{P_1}^*$ and $\mathbf{P_2}^*$. This approach to spectral analysis has significant advantages. More wavelengths can be used in the calibration than there are samples in the calibration set, without degrading the results. Full spectra can be used without down-weighting some of the information at certain wavelengths, reducing the possibility of missing something new that may appear in future samples. Also, the method is completely nonparametric, and the shape, scale, and skew of the spectral sample distributions do not affect the quality of the results.

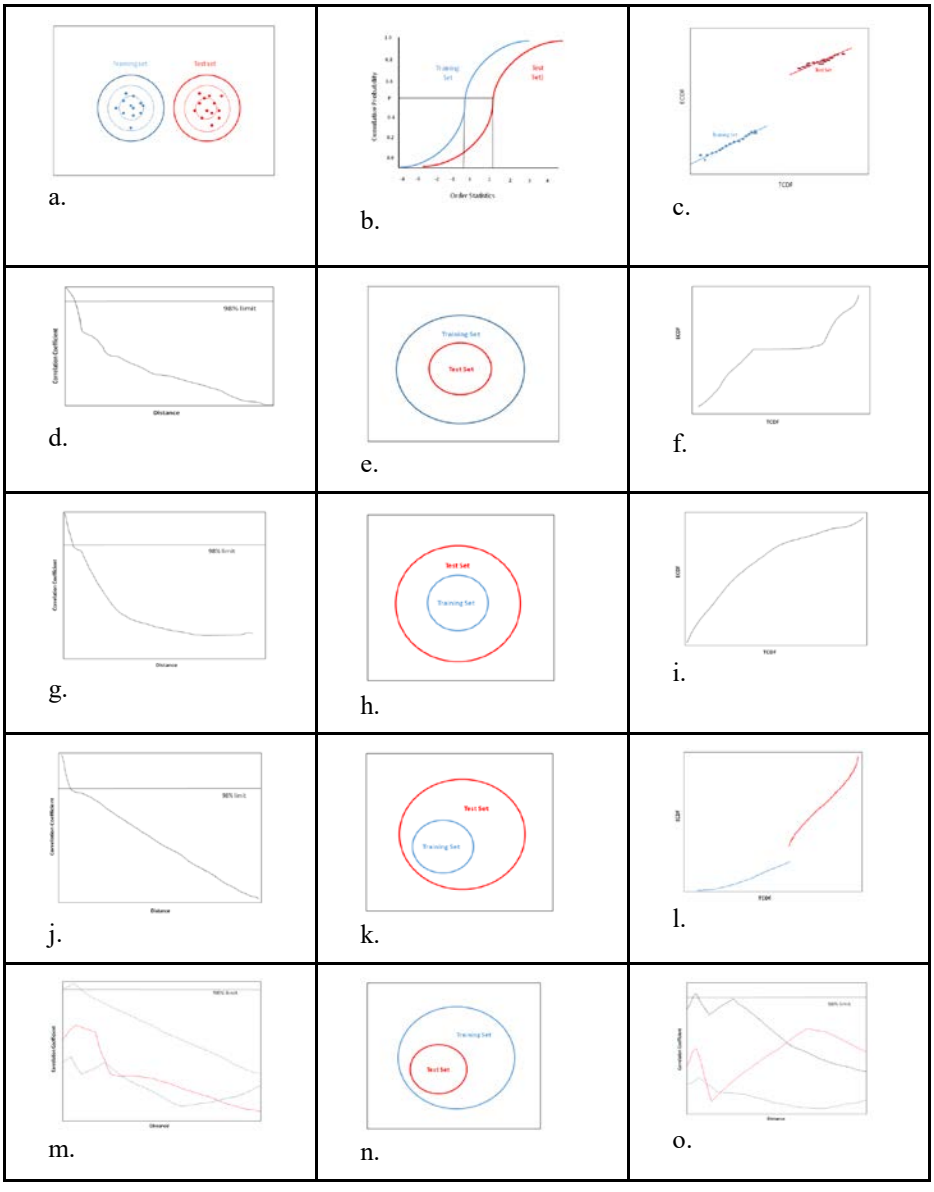

Fig. 6. Different patterns that can emerge in QQ plots as a result of different received signals in the training set and test set, along with the effect on correlation coefficient calculated from the QQ plot as the training set and test set vary in location and scale. Spectra at *d* wavelengths are represented as points in a *d*-dimensional hyperspace. The quantiles are integrals from the center of the Training Set outward in all directions. The first Empirical Cumulative Distribution Function (ECDF) comes from the Training Set alone (the system is typically trained on galactic noise), while the second ECDF contains points from the Training Set AND the Test Set.