

Boyer's Wharf

Dec. 8, 2003

**Jim & Holly Akenson
Taylor Ranch Field Station
HC 83
Cascade, ID 83611**

In my e-mail earlier last fall I promised to send additional documentation and a more detailed description of my evolving methodology. I'm finally sending it! I kept delaying writing about the direction I'm taking as it is still in the early stages and will evolve further. Rather than wait any longer I wrote a draft and with a date and version number should convey message that it is far from final. Share any of the material with Katie, or any others you deem appropriate. Most of my graphs are done in a software program that is old, and to my liking, but which prevent electronic distribution as few others can read those files. This mandates paper distribution.

Included in this package are copies of some tape recorder/AnaBat issues and Titley Electronics description of their ZCAIM-CF. While I firmly believe that the ZCAIM-CF is the best currently available technical solution my recommendation is tempered by the thoughts that Taylor Ranch has a limited budget and that the unit's cost in US \$ has increased considerably since my purchase. I purchased when the AU\$ was only US\$0.56 versus a current value of US\$0.736

Also included is a print of Simon Jolly's home page which provides links to several useful bat information sources.

Happy Holidays,

Jim Morris

AnaBat File Analysis
(Version 1.0, 12/07/2003)

Introduction and disclaimer

The following is my personal approach to the analysis of AnaBat files. I must emphasize that these are views that have not had any expert or peer review and as such may not have merit. Recall also that I am new to the subject, am not a biologist but rather a retired electrical engineer who spent most of his life researching and developing microwave and fiber optic components. My approach has been to initially do an independent examination, develop the approach, and then follow up with evaluation and criticism by experts in the field. Readers can help by pointing out the flaws in my reasoning. As will be obvious later in this document this is an interim report on work in progress and much of the work is not yet complete.

Critique of some existing method

As indicated in my earlier e-mail, copy attached, one approach to species identification is to compare the graphical and extracted parameters with the examples provided in the library sold by O'Farrell. I see two weaknesses in this approach. First one is comparing to a single example of the species. This therefore does not allow for the probable variations within that species due to age, gender, geographic location, or activity. Secondly this approach assumes that the researcher extracted pulses and parameter values from the recording of the unknown exactly as was done for the library reference. The selection criterion for pulses in the library is unspecified. The criteria used by the researcher trying to make the identification is unlikely to match that used for the reference library file. This method is best suited to use by meticulous experts using only reference library parameters they personally generated.

A second approach, described in an excellent on-line tutorial <<http://members.ozemail.com.au/~jollys/>> uses extracted parameters values (from any method) and compares them to reference parameter values of multiple species using a statistical software package and then calculates the probability that the unknown matches one of the reference species. The tutorial appears to use an older (circa 1998) base version of SPSS software without any extensions. This is excellent software but rather expensive (the current base version costs about \$1,150 and the additional regressions and advanced extensions cost \$700 each). For the version used by the author of the tutorial I suggest the following methodology weakness and limitations. While the mathematics is solid it assumes, as in the method above, that each species have unique parameter values and that the unknown's parameter values were extracted using matching analysis criteria. No allowance is made for parameter variances. That seems to be in conflict with my own work, described below, that indicates both large parameter variances and variance values that vary widely between species. Additionally the tutorial's software analysis seems to assume that the measured parameters are all uncorrelated. Again my intuition and initial data indicate some parameter correlation.

My parameter data base development

My parameter data base is developed from the AnaBat files available from the library maintained by the University of New Mexico <<http://talpa.unm.edu/batcall/>> . I've downloaded the available files of those species I have reason to suspect might be found in either Idaho or Pennsylvania. I'm using separate data bases for the two states in which I have interest. No AnaBat files were available for several of the species of interest but may become available at a later date. For the species *Eptesicus fuscus* where statistically sufficient numbers of UNM files were recorded in both eastern and western states I've included in the Pennsylvania data base only those recorded in the East (New York) and only those recorded in western states in the Idaho data base. This was done to safeguard against possible regional dialects as have been cited by Australian researchers.

All files were analyzed using "Analyze for Windows 95" software using the default preferences. While one may debate the relative merits of the several available analysis software packages I selected Analyze for Windows 95 since it applies a consistent, user independent, analysis criteria and thereby provides results that can be duplicated by other researchers. Not all AnaBat files downloaded from the UNM contained data meeting the Analyze for Windows 95 minimum criteria for analysis. The number of western species files from UNM and the number files with one or more pulses meeting the Analyze for Windows 95 minimum default criteria are shown, by species, on the attached spreadsheet. All pulses successfully analyzed for each species were concatenated into single files and the mean and standard deviation of each noted. These values are also shown on the spreadsheet as are the pulse totals. I apologize to readers for my use of common names for species but I find it convenient and often facilitate shorter file names. The parallel effort on the Pennsylvania species is in process using, unfortunately, less than statistically meaningful numbers of files for many of the species.

Analysis of the parameter data base

Prior to analysis of any data it is advisable to examine that data and the nature of the distributions.

Initial probability plots of several parameters on a few species indicate that most parameters are approximately normally distributed. This is useful since it allows use of readily available software and well-documented data handling procedures. It is most interesting to note that much improved normal distributions are generated by taking the log of the parameter values measured. One such example is shown for the pulse duration of *Eptesicus fuscus* recorded by Gannon in New York State. Note that the fitted value of r^2 is considerably improved for the logarithmic case. For the parameters of *Eptesicus fuscus* recorded by Gannon in New York State I find the log of the extracted parameter produces a more normal distribution for Pulse Duration, Maximum Frequency, Mid Frequency, End Frequency, and Curvature. Results for Start and End Slopes are inconclusive. Should this prove typical for all or most species I expect to use the log values for species identifications.

As can be seen on the spreadsheet the mean values of many species are similar and their distributions overlap. Plotting ideal normal distributions based on these means and standard deviations illustrates one issue that makes species identification challenging. Attached is a plot of the distributions of End Frequency for ten of the western species. Note the considerable overlap. An end frequency of 23 KHz it is highly likely to be from six of the species, possibly from one species and unlikely to be from the three remaining species.

Correlation between parameters is also of interest. Again using the parameter data for the species *Eptesicus fuscus* recorded in New York by Gannon from the UNM files each parameter (except curvature) is plotted against each other parameter and linear regressions examined. Curvature is not included since the granularity of the data extracted from Analyze for Windows 95 precludes its use. While there is considerable scatter in the data several pairs of parameters suggest some correlation. Four of the paired parameter combinations have linear regression r^2 values of 0.8 or higher. Should the correlation of parameter values prove typical in all or specific species then the identification statistical approach must take this into account.

Proposed unknown analysis methodology

My plan for species identification by analysis of AnaBat files is to initially use the Analyze for Windows 95 software with the preferences set exactly the same (at the default values) as was done in creating the data base. Each pulse of each file will then be tested against the data base statistics (means and standard deviations) for all likely species matches and the difference expressed in terms of number of standard deviations of that reference species. The averages of the results of the tests for that pulse against each reference species statistics would be noted. A weighted average may also be implemented giving higher weight to key parameters (end frequency for example) and reduced weight to correlated parameters or parameters suspected to be species independent. Once averages have been determined for all pulses in the unknown species' AnaBat file the averages of the averages will be used to determine the most likely matches. Note it is unclear at this time if logs (or some other function) of parameter values will be used.

The approach can be followed without the use of expensive SPSS or equivalent software. Commonly available spreadsheet programs should be adequate. Finally it must be recognized that definitive species determination by echolocation recording analysis may not always be possible and that physical capture and identification may sometimes be necessary for species verification.

Att.

Copy Sept. 2003 e-mail,
spreadsheet, probability plots,
Western bat distributions,
correlation study

Jim Morris
jemorris@ieee.org

BATANALZ.DOC

To: tayranch@direcpc.com
From: Jim Morris <jemorris@ieee.org>
Subject: Bat detection & identification
Cc:
Bcc:
Attached:

Holly,

It was great to again visit Taylor Ranch (my 5th visit) this summer and to see you and Jim. We spoke briefly about bat studies and I'd like to extend that discussion. Separately I'll be mailing some documentation that does not lend itself to email.

As I mentioned I purchased an AnaBat detector shortly after my visit in Aug. 2002. These detectors are normally used with a ZCAIM (Zero Crossing Analysis Interface Module) and some type of recording device. Katie Gillies used both a tape recorder and a laptop. Tape recorders tend to be problematic as you can read in the mailed material. Titley Electronics accordingly introduced a new ZCAIM in 2002 that records directly to removable memory chips (the same Compact Flash Memories used in many digital cameras) which eliminates the need for field tape recorders. For a full description see <<http://www.titley.com.au/anabatcf.htm>>. These new ZCAIM-CF Modules do a better job of recording and provide numerous additional functions (example simultaneously recording input from a GPS receiver). It is Titley's ZCAIM-CF that I purchased and have been using. Its cost is slightly greater than the earlier models but does not require one to buy a tape recorder.

Recording bat echolocation vocalizations is the easy part. Analysis is where time and effort is most needed. From what I've read there are two basic approaches to species identification from AnaBat recordings.

One approach is to analyze each recording using any one of the several similar software packages (Analook, Analyse for DOS, Analyze for Windows 95, etc.) and then compare screen images and parameters with a reference voucher recording such as contained in the library sold by O'Farrell. I believe that's the method Katie was using during my visit to Taylor.

A second approach is a more statistical method. I will not try to fully describe that approach since you can access an excellent tutorial and worked example online <<http://members.ozemail.com.au/~jollys/>>. In this method parameters are extracted as in the method above but entered into a statistical package (from SPSS) and compared to mean values for several species to produce specific probabilities that unknown belong to each of the voucher species. Mathematically this is solid statistics.

I'm looking to try to extend that methodology. Method One, above, compares an unknown to a single voucher example. Method TWO, above, rigorously compares each parameter to the mean values of that parameter for several species.

As I see it the various parameters of a given species probably vary considerably for a variety of causes, including the bat's age, gender, geographical region, activity etc. For example my initial

look at the parameter fmax for *Eptesicus fuscus* shows a mean of 46.01kHz with a standard deviation of 13.3kHz. The variance of a particular parameter is also likely to be different for different species. For that reason I want to include both the mean values for the voucher files but also the expected variance.

I'll try to write a detailed outline of my approach as it evolves so as to allow for peer review. I'll need guidance, as I'm not a trained biologist. My first order of business is creating a database of reference AnaBat voucher calls for the species that I hope to identify. The greater number of calls in that data base the better statistically (but more work for me). I'd like to include all species found or likely to be found at Taylor Ranch. I've got one listings of bat species likely to be found in Idaho but would appreciate your input (or Katie's if available) as to the species you suspect, or know, reside in your area. The Univ. of New Mexico has online a wonderful database of AnaBat files many of which I've downloaded. I will need more, however, as not all species are included and for some species the numbers are not statistically significant. Once I know all the species of interest I plan to solicit files other researchers may have using the Batline message board.

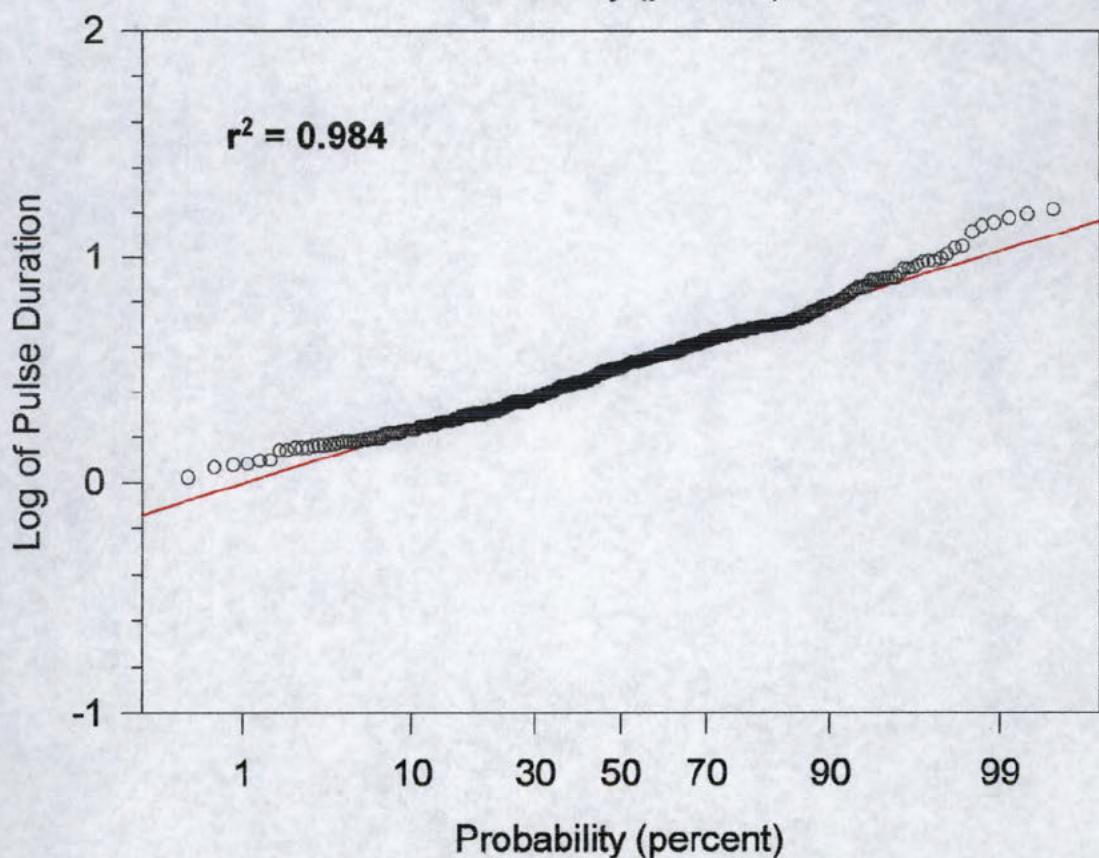
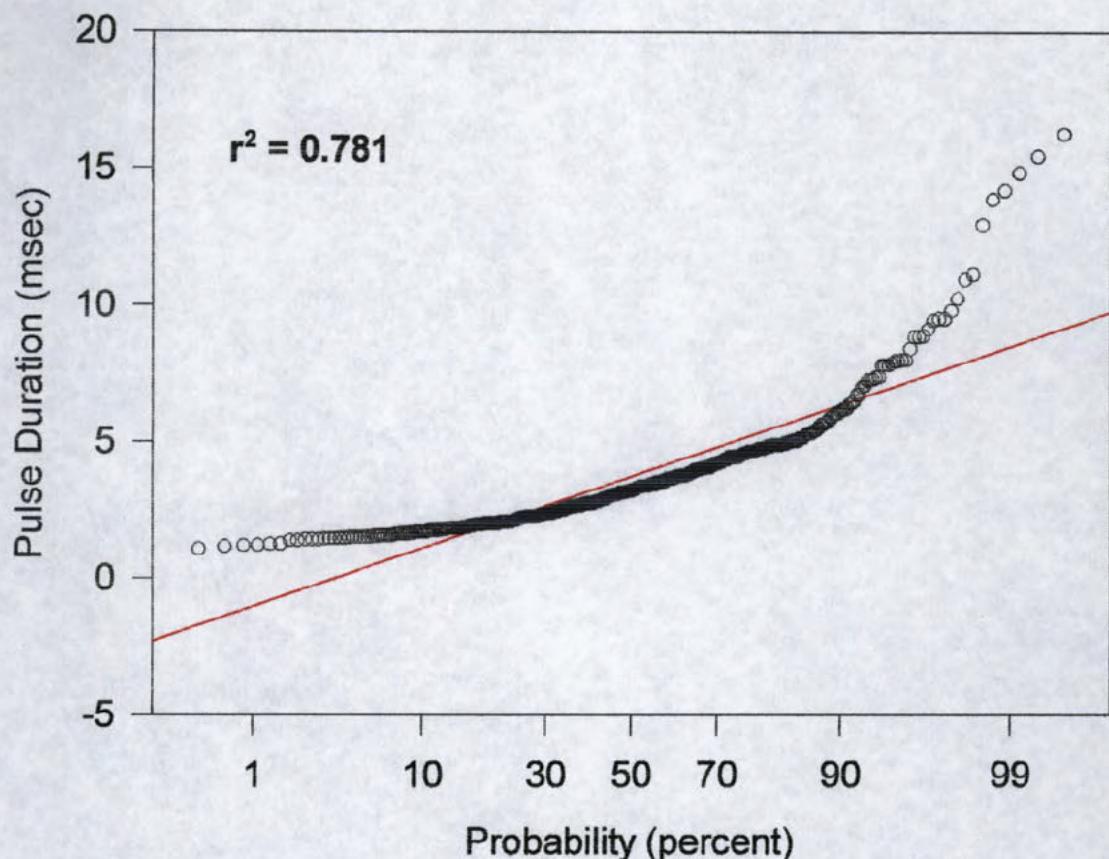
I continue to entertain thoughts of flying in to Taylor in the late winter or early spring both to see the area in another season and to pursue a possible joint effort on bat detection and identification.

Jim Morris

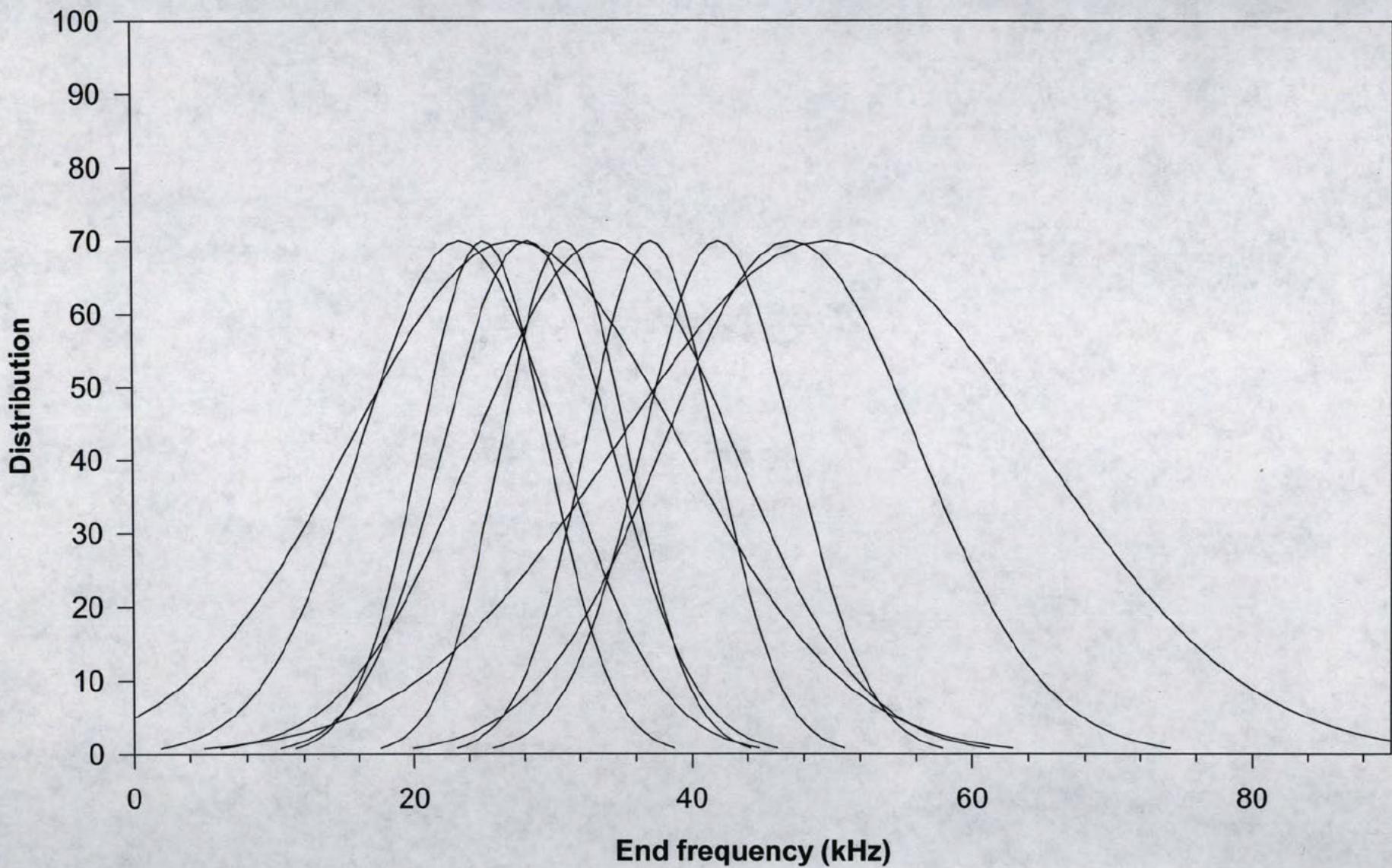
Species (possibly found in ID)	Raw UNM AnaBat Files	Number files from Analyze for Windows 95	Number pulses analyzed	Duration (mSec) Mean	Duration Std.Dev	Max.Freq. (KHz) Mean	Max.Freq. Std.Dev	Avg.Freq. (KHz) Mean	Avg.Freq. Std.Dev	End Freq (KHz) Mean	End Freq Std.Dev	Starting slope Mean	Starting slope Std.Dev	End slope Mean	End slope Std.Dev	Curvature Mean	Curvature Std.Dev
Big Brown Bat (Western files)	310	308	6563	7.725	3.498	41.511	11.990	32.036	7.001	28.231	5.879	-3.363	4.142	-1.066	2.449	2.668	1.289
California Myotis	69	69	1994	2.364	1.326	80.690	21.937	60.941	15.190	49.649	14.641	-20.122	14.788	-10.057	10.513	1.970	1.264
Fringed Myotis	50	50	1884	3.580	1.726	62.324	15.079	38.596	7.585	23.354	7.026	-16.375	7.945	-8.881	5.403	1.641	0.672
Hoary Bat	27	14	300	7.220	3.386	41.070	10.031	29.947	5.493	25.080	4.474	-4.574	4.049	-1.885	3.516	2.806	1.591
Little Brown Bat	18	18	901	4.746	2.334	55.615	12.456	43.077	6.132	37.024	4.546	-6.189	3.607	-2.443	1.869	1.992	0.958
Long-eared Myotis	41	39	730	3.393	2.812	67.836	18.879	45.499	9.651	33.668	9.019	-17.781	12.095	-8.371	8.513	2.193	1.712
Long-legged Myotis	23	22	491	3.196	1.993	66.383	17.552	51.248	11.746	42.852	10.190	-14.242	12.190	-6.915	8.697	2.071	1.330
Pallid bat	27	27	515	4.841	2.754	49.405	10.991	34.816	10.720	27.272	11.655	-8.518	6.421	-3.473	3.486	2.152	1.496
Silver-haired bat	0	0	0														
Spotted bat	0	0	0														
Western Pipistrelle	248	90	920	4.87	3.46	57.46	9.94	49.42	6.18	46.46	6.12	-4.66	3.99	-1.26	2.31	2.94	1.59
Western Small-footed Myotis	67	64	1673	2.946	1.607	64.428	14.481	48.578	5.759	41.835	5.271	-12.319	8.109	-4.041	4.194	2.353	1.404
Townsend's Big-eared Bat <i>Plecotus townsendii</i>	0	0	0														
Townsend's Big-eared Bat <i>Corynorhinus townsendii</i>	3	3	29	2.106	0.447	48.817	9.450	39.018	3.930	30.866	4.367	-10.174	5.862	-7.224	3.668	1.214	0.266
Yuma Myotis	214	211	7534	3.017	1.529	74.583	15.563	57.579	8.415	47.055	8.882	-12.937	8.301	-7.496	6.872	1.623	0.734

Probability Plots of Big Brown Bat Pulse Duration

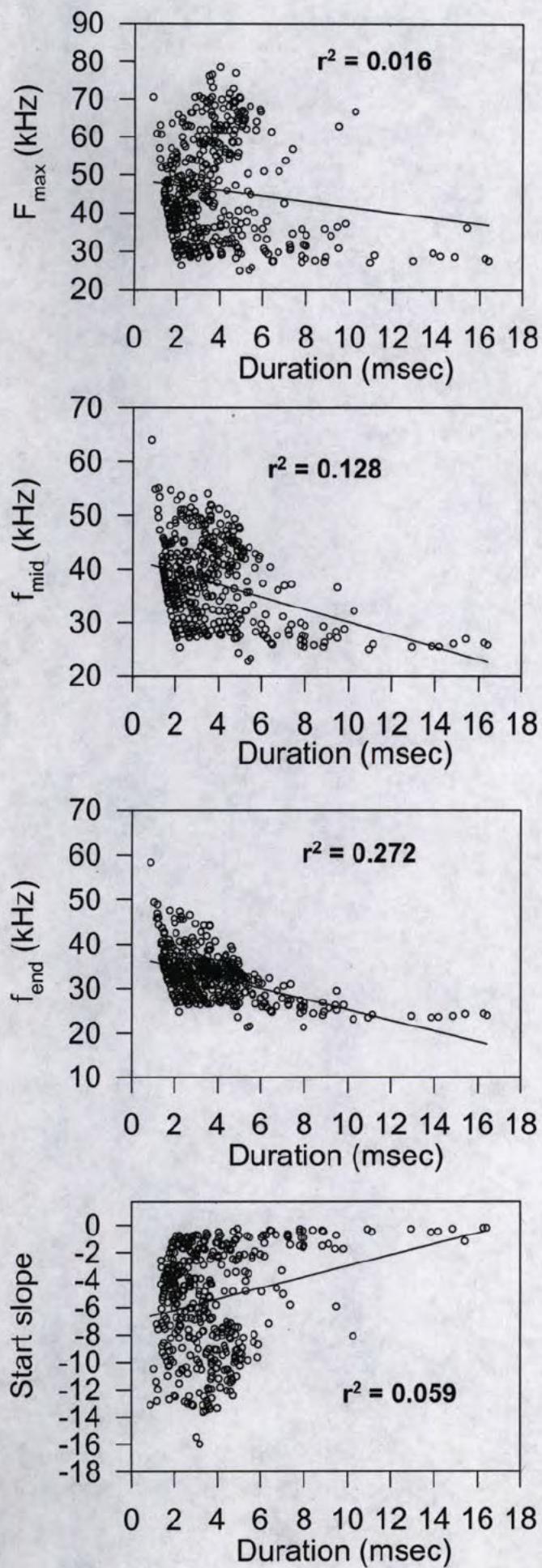
From 429 pulses extracted using Analyze for Windows from 110 of 160 Univ. New Mexico AnaBat files recorded in New York by Gannon



**Idealized Gaussian Distributions of Echolocation End frequency
for ten species of Western North American Bats**



BATFEND.SPW



Pulse Parameter Correlation Study *(Eptesicus fuscus)*

From 429 pulses extracted using Analyze for Windows on 110 of 160 Univ. New Mexico AnaBat files recorded in New York by Gannon

