Correcting Milagro Rates for Environmental Conditions

Senior Undergraduate research project for James Ledoux

Milagro Rate Variation

- Milagro has a high count rate, ~ 10^6 s^{-1} in high-threshold scaler mode.
- It depends on environmental conditions such as temperature and pressure as well as instrumental factor, e.g., pond depth. The exact dependence in not well known.
- To use Milagro as a cosmic ray detector, we must eliminate the dependence on environmental conditions (as well as correcting for instrumental effects).
- Principal Component Analysis (PCA) is a method to do this.

Principal Component Analysis

- When applied to a multi-dimensional data set, Principal Component Analysis can find new basis vectors (the Principal Components) that organize the variations that linearly depend on the original data.
- Milagro scaler rates can thus be combined with environmental parameters to reduce the variation in the background rate of the detector by correlating the rates with linear combinations of the environmental data.

Principal Component Analysis

- To perform a PCA, we first compute a covariance matrix for the data set. For a data set with 3 dimensions (*x*,*y*,*z*),

$$Cov(x,y,z) = \begin{bmatrix} Var(x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Var(y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Var(z) \end{bmatrix}$$

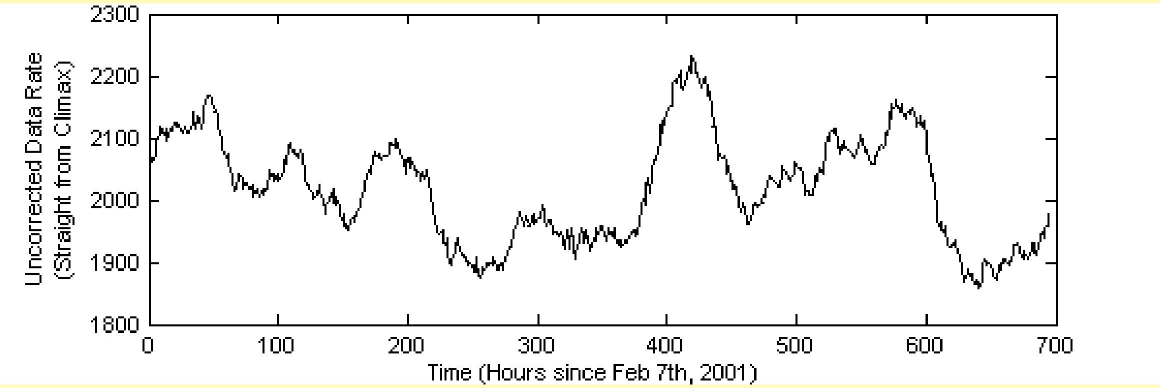
- $Var(x) = \sigma_x^2 = Cov(x,x)$ and is a measure of the spread.
- $Cov(x,y) = Cov(y,x) = \sum (x-\mu_x)(y-\mu_y)$ and is a measure of the correlation between the two dimensions.

Principal Component Analysis

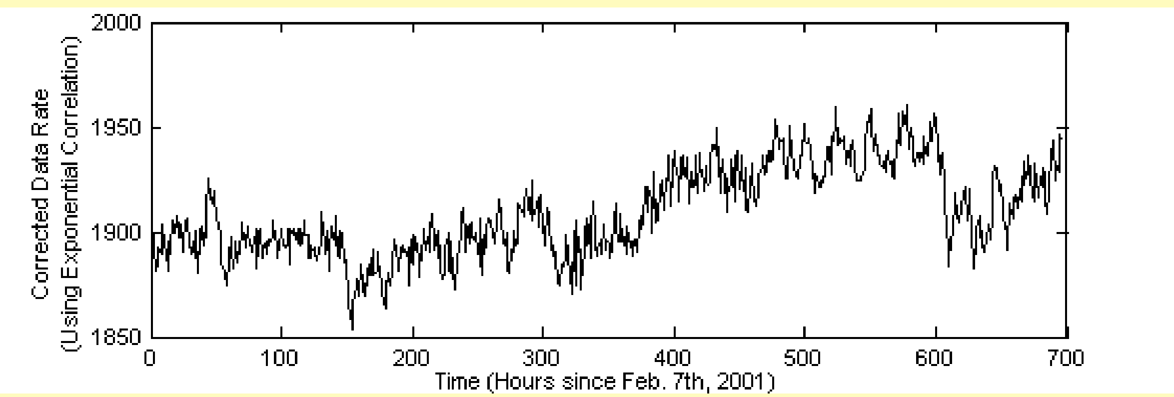
- Each data point is a vector in the n-d data space.
- Must first diagonalize the covariance matrix by finding the eigenvectors of the covariance matrix.
- Since the matrix is symmetric ($Cov(i,j) = Cov(j,i)$), the eigenvectors are automatically orthogonal to one another.
- The data set can then be rotated into these new dimensions, so the variations depend on the coordinates of the new basis vectors, in order of the eigenvalues.

PCA Example – Climax Data

Uncorrected Climax Data (U)



Corrected Climax Data (C)

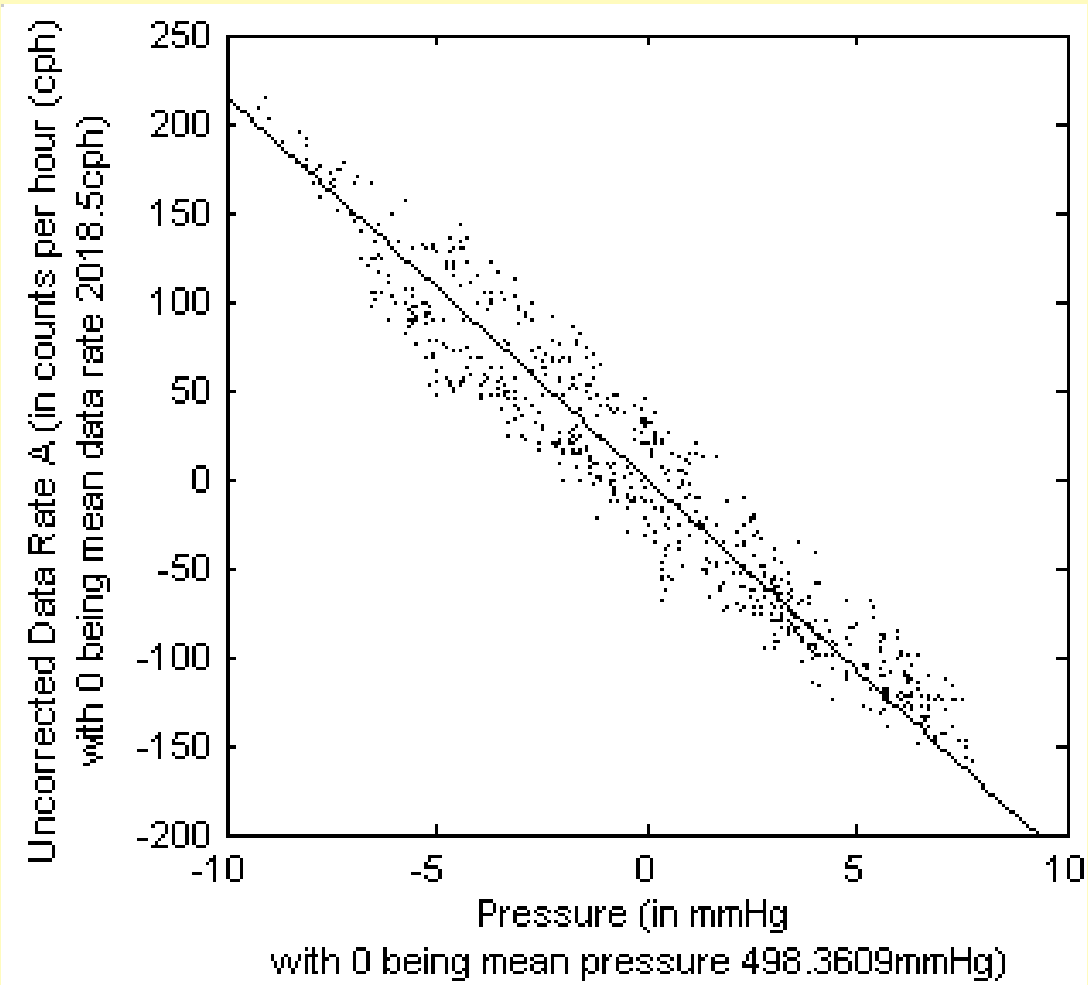


Time (Feb 7th, 2001 to Feb 28th 2001)

A PCA was performed on the Climax Neutron Monitor data, that varies only with barometric pressure (exponentially).

Using PCA we can estimate this variation.

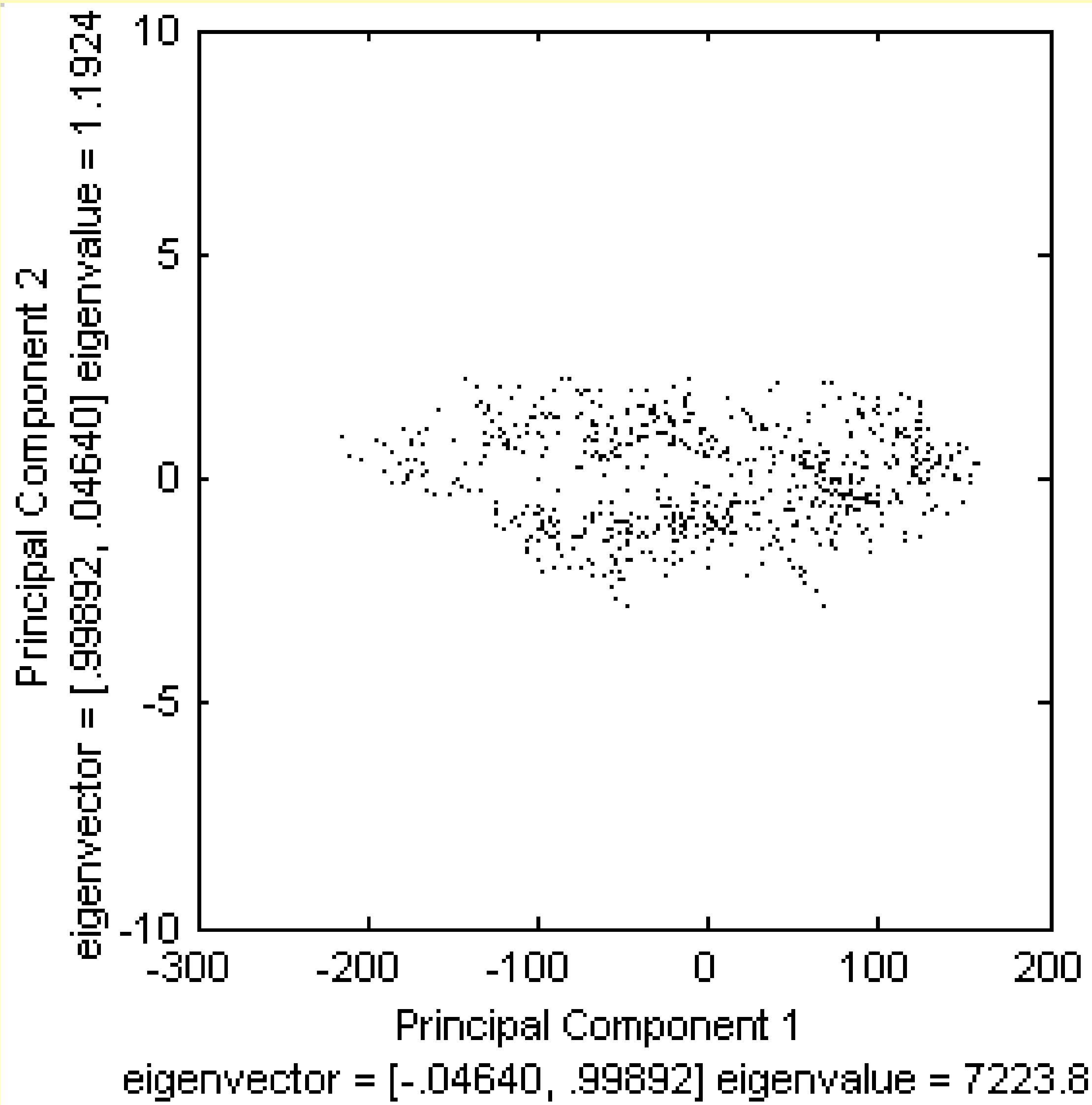
PCA Example – Climax Data



Plotting uncorrected rate vs. pressure uncovers a correlation between the two variables.

The first principal component is shown.

PCA Example – Climax Data

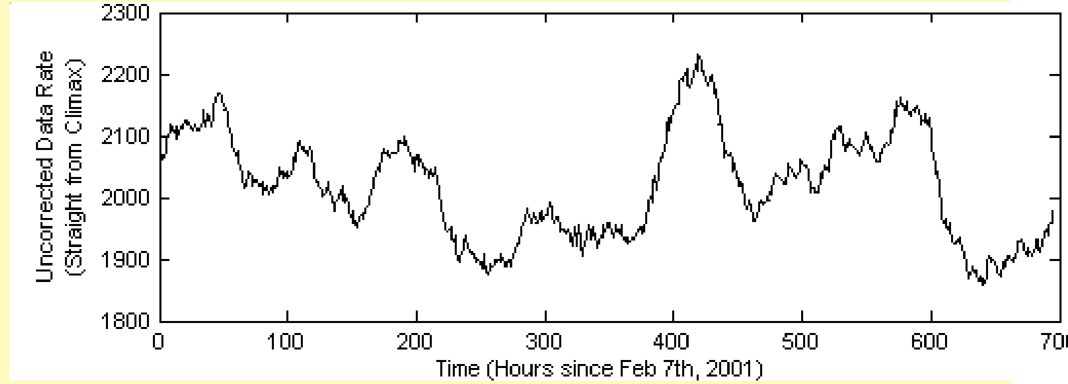


Rotating the data set into its principal components leaves most of the variation accounted for in the first principal component, but the **second** principal component contains the variation beyond that due to pressure, in this case heliospheric.

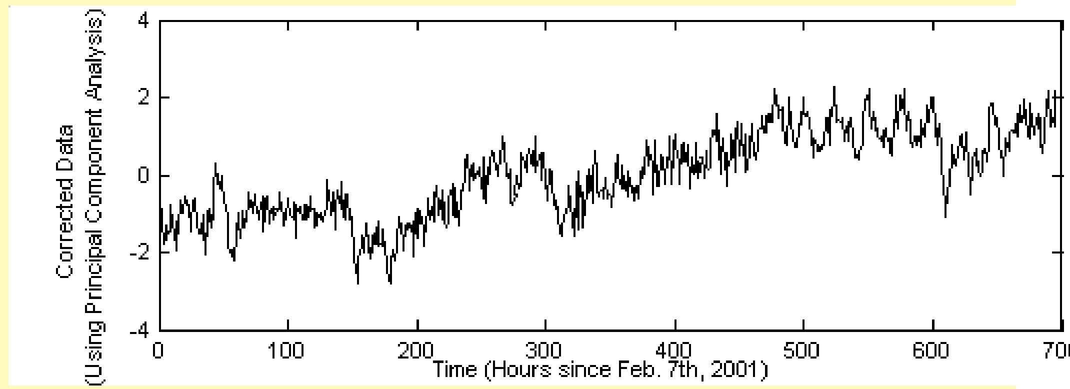
Plotting this last principal component vs. time, we mimic the corrected rate obtained using the standard corrections.

This shows how the method produces a correction factor when it is unknown.

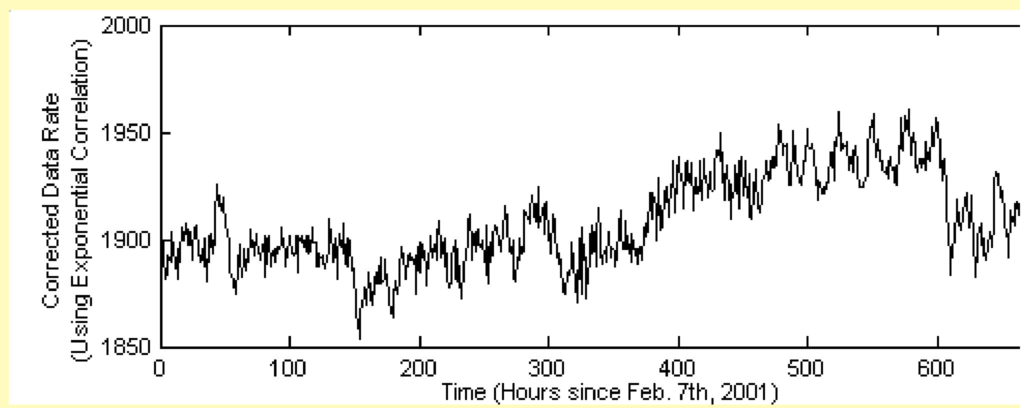
Uncorrected Climax Data



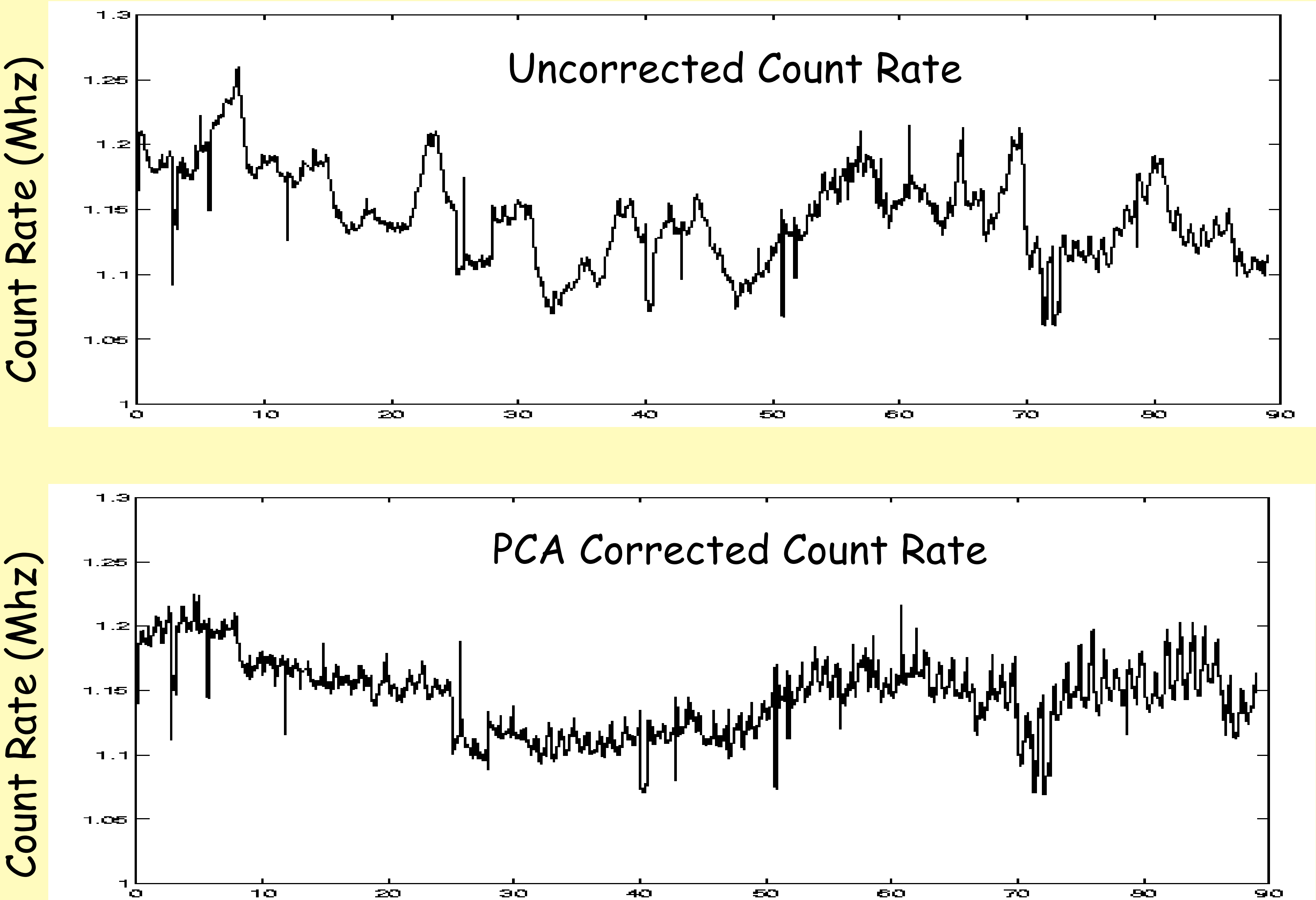
PCA Corrected Climax Data



Exponentially Corrected Climax Data



Applying PCA to High Threshold Scaler Data



T_i, T_o, & P

Time (Days Since Feb 1st, 2001)

Conversion formula derived from Principal Component Analysis

$$\text{CorrectedRate} = \text{OldRate} + 0.131 \times p(\text{in}_{Hg}) + 0.0012 \times T_i - 0.0022 \times T_o$$

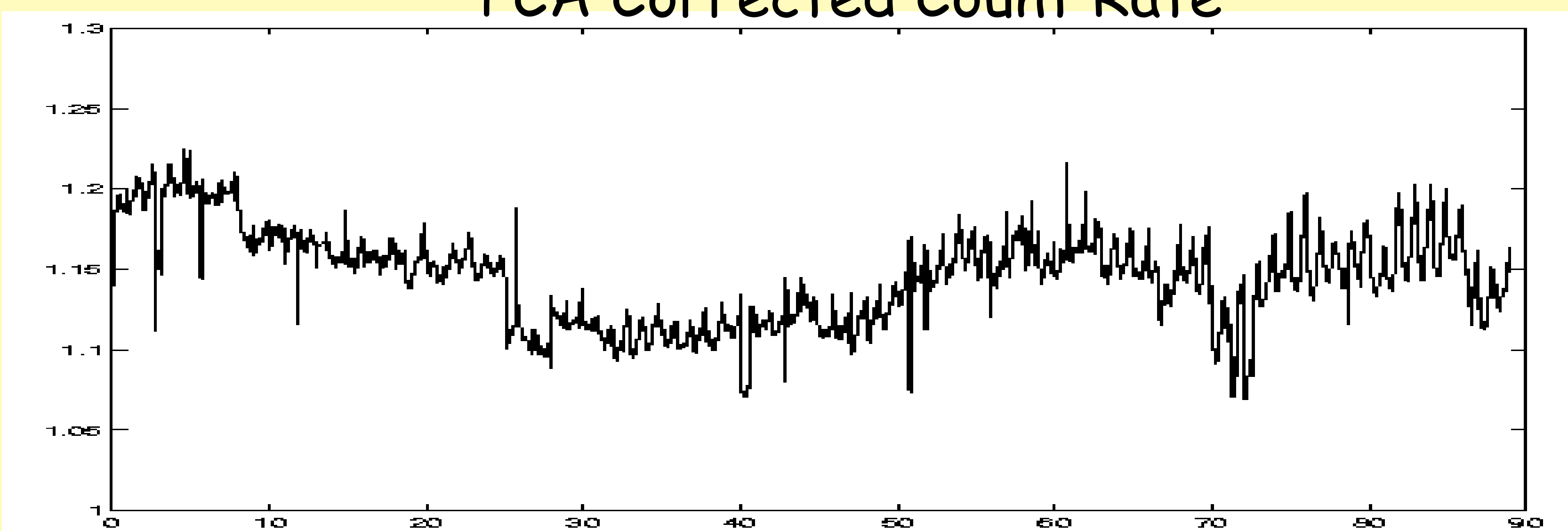
***Other parameters such as pond depth, number of dead tubes not included.**

Results

- The background is significantly flatter, with many features disappearing when correlated with T and p.
- However, the diurnal fluctuation becomes large during April 2001.
- There are other spikes in the data associated w/ dead PMTs.
- Some features correlated with those in Climax data

PCA Corrected Count Rate

Count Rate (Mhz)

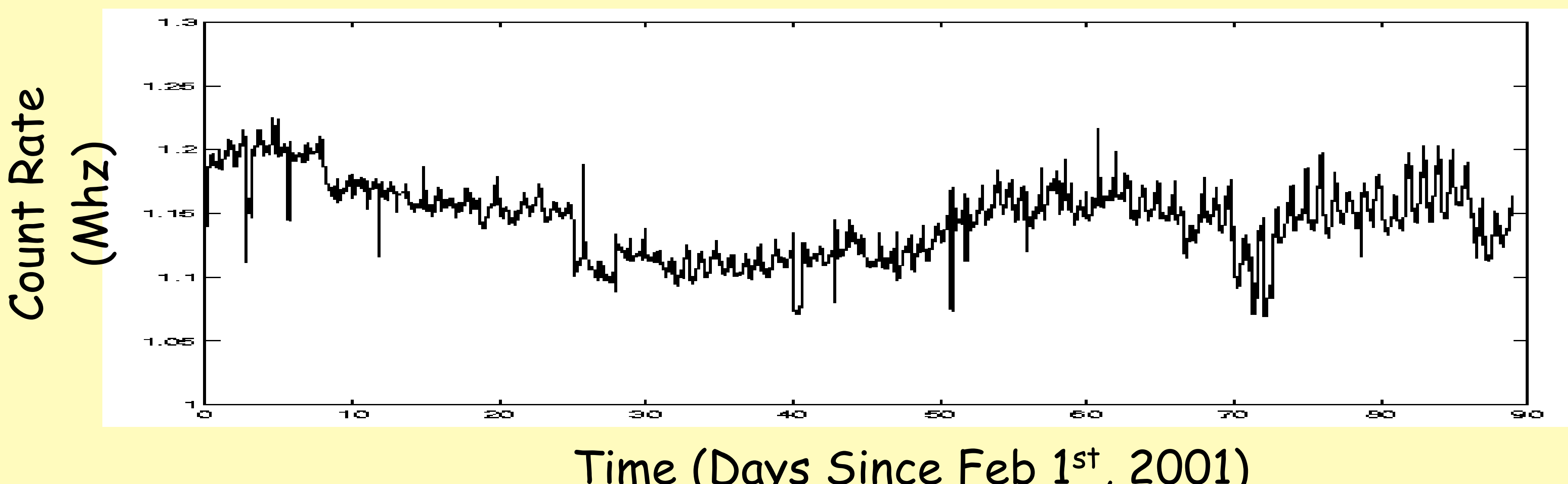


Time (Days Since Feb 1st, 2001)

Conclusions

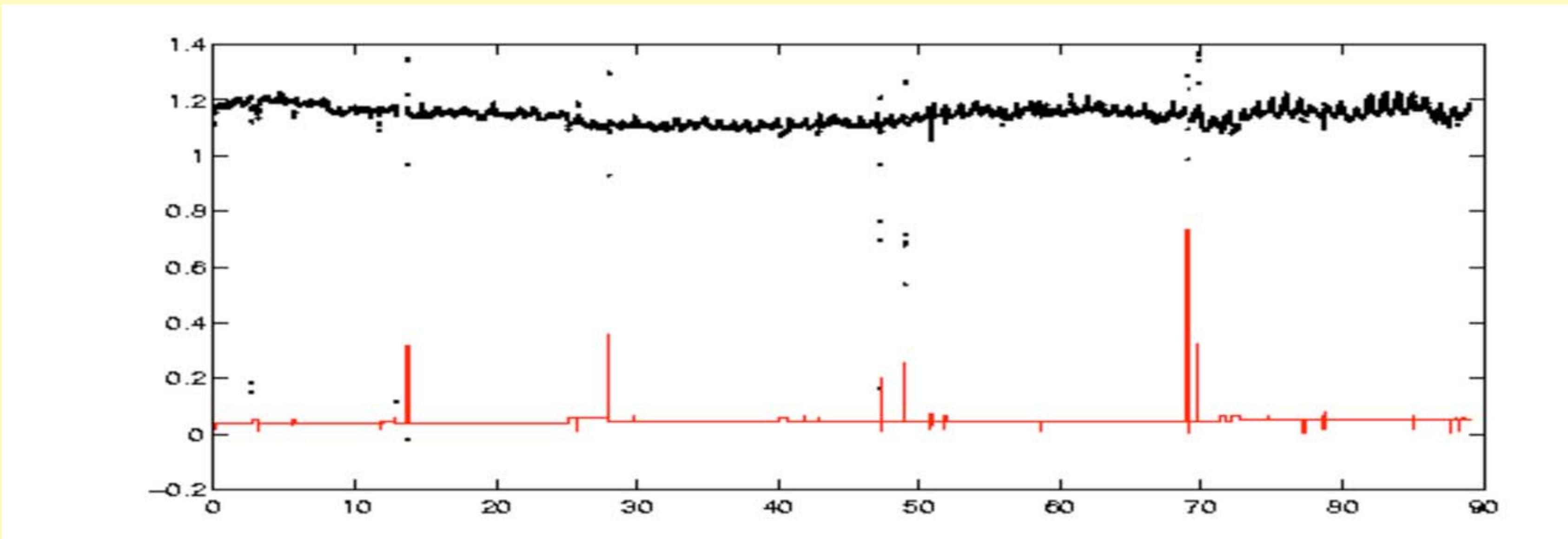
- The PCA was performed on February data but was applied to the warmer months' data, through April.
- Dead PMTs affect the count rate for high scaler threshold. Other data channels of Milagro (such as 2-fold coincidence, etc.) may be more stable.
- Incorporating data from high altitude weather measurements may further reduce fluctuations.

PCA Corrected Count Rate



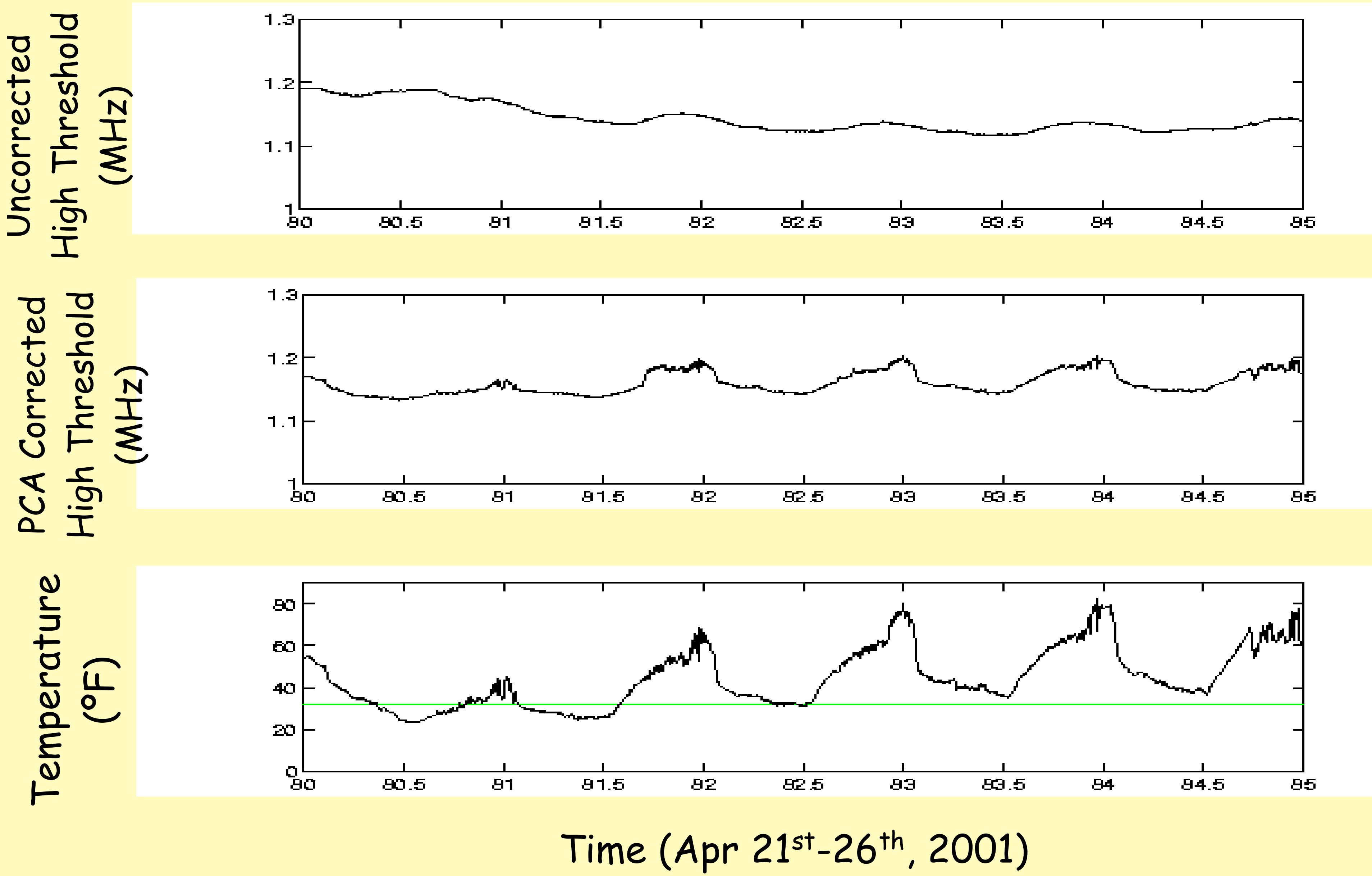
Dead PMT Effect

PCA High Threshold (MHz)
# of Dead Tubes / 1000



Time (Days Since Feb 1st, 2001)

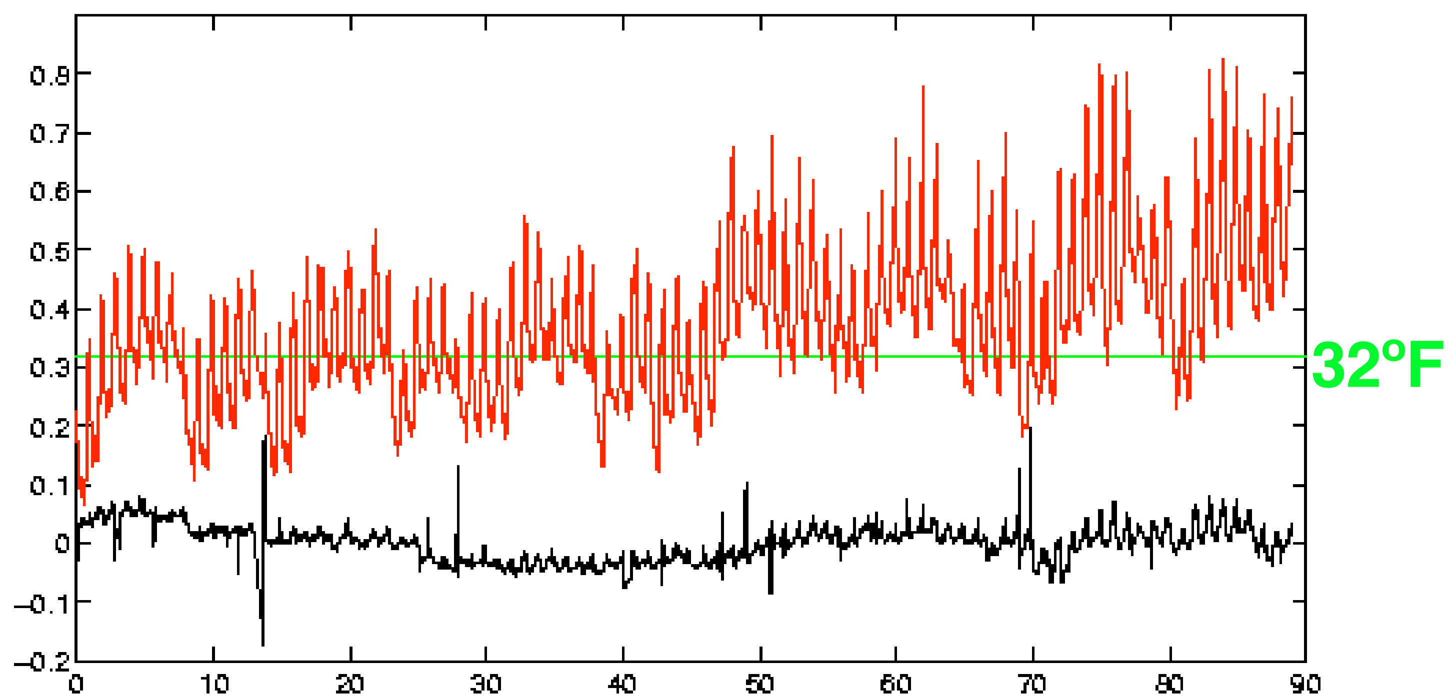
PCA Magnifies Daily Modulation (?)



Temperature and PCA'd High Thresh

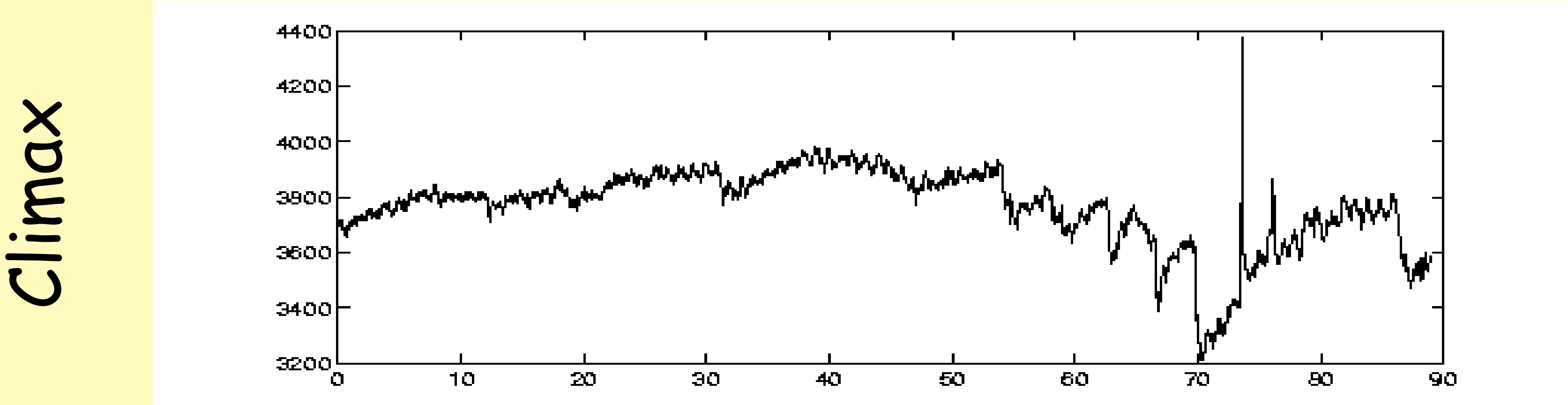
Significant levels of freezing present in Feb 2001, when PCA was performed, but not in Apr 2001.

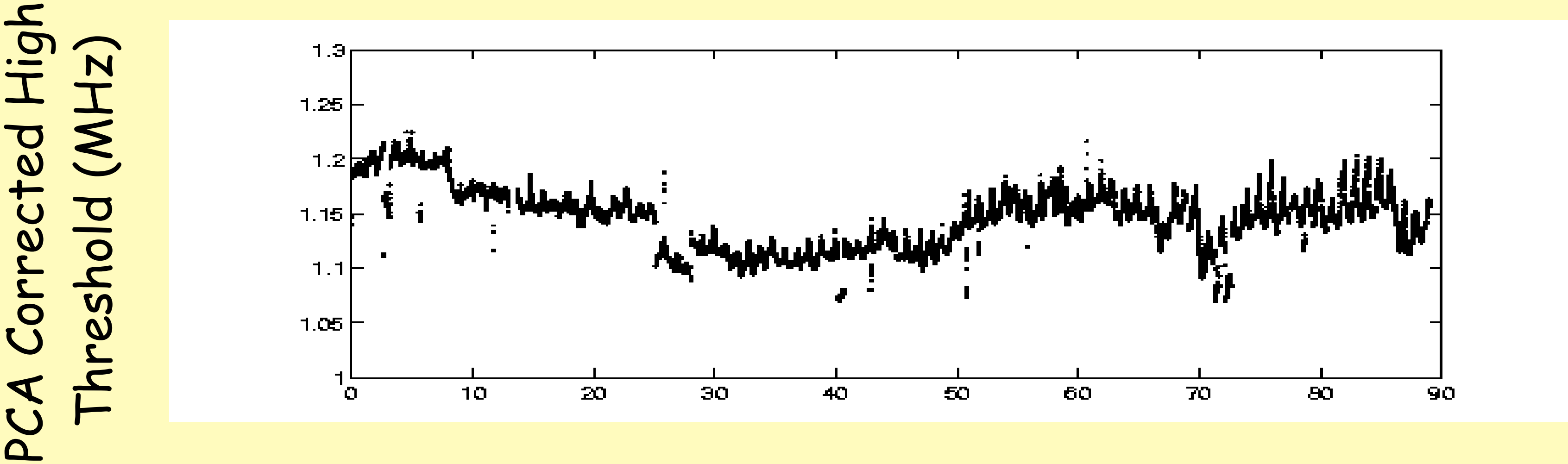
PCA High Threshold (MHz from mean)
Temperature (°F/100)



Time (Days Since Feb 1st, 2001)

Climax and PCA'd High Threshold



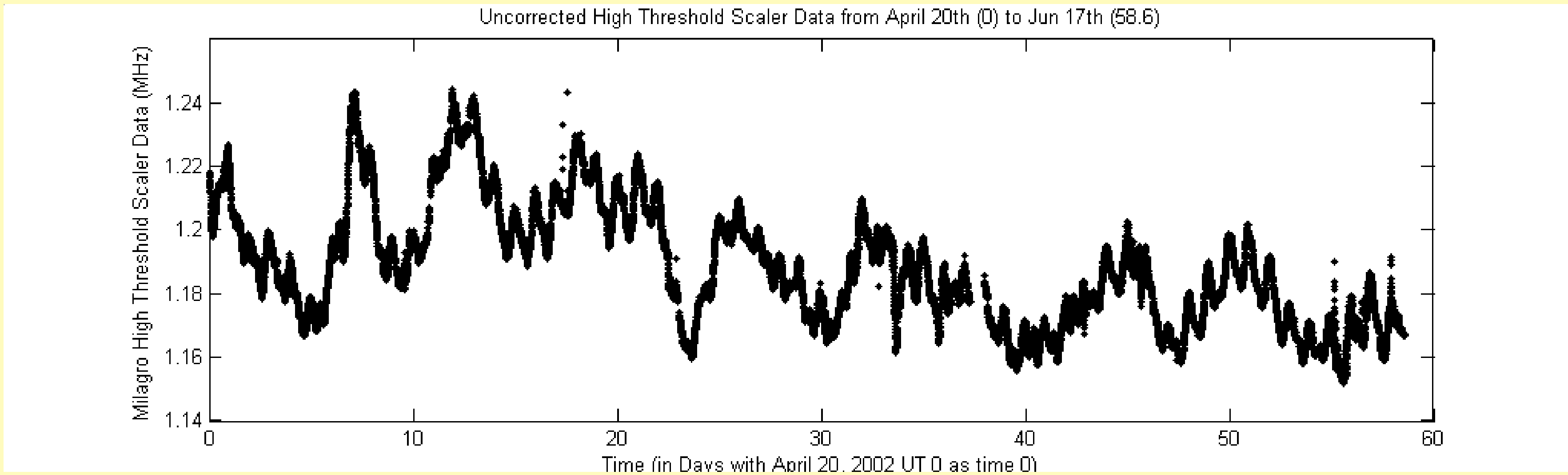


Time (Days Since Feb 1st, 2001)

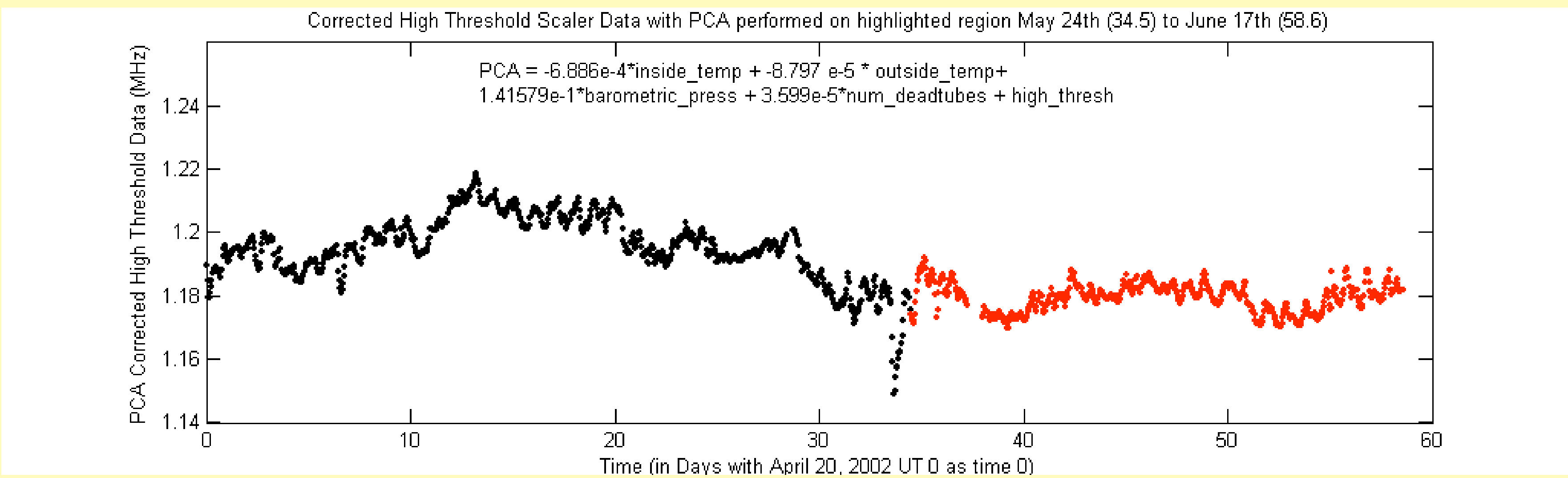
Two Forbush decreases detected in corrected data.

PCA Applied to H-T Scaler Data
April-June 2002

Raw



Corrected

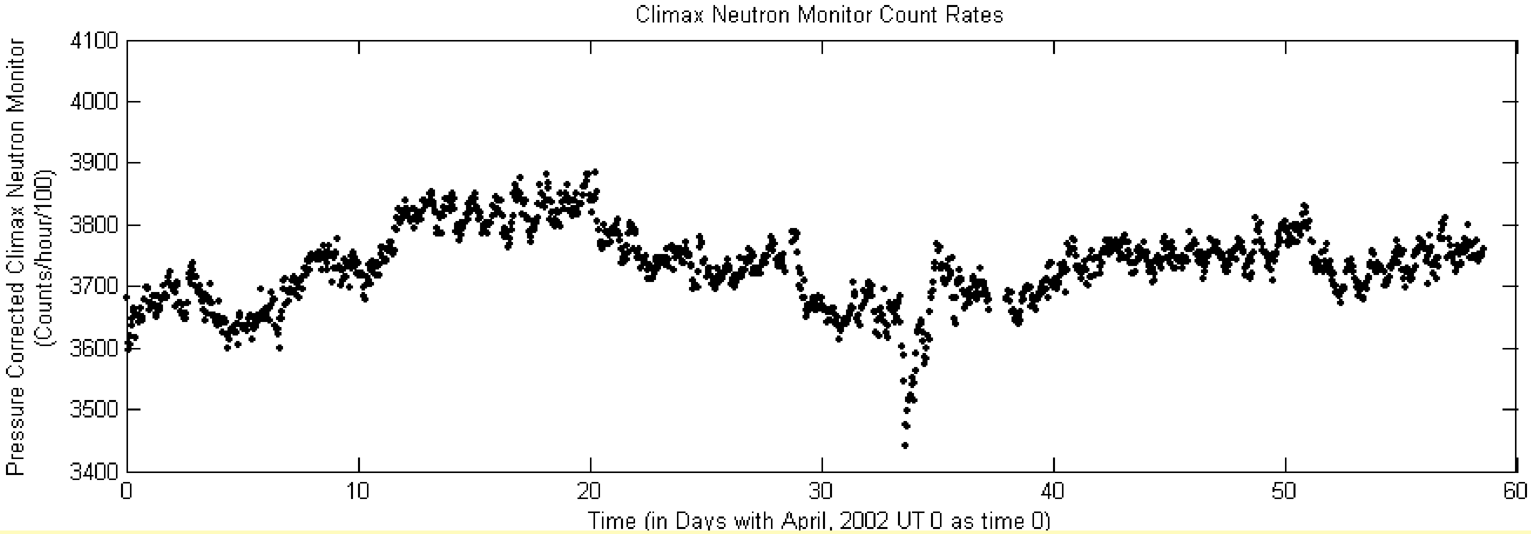


Forms basis vectors

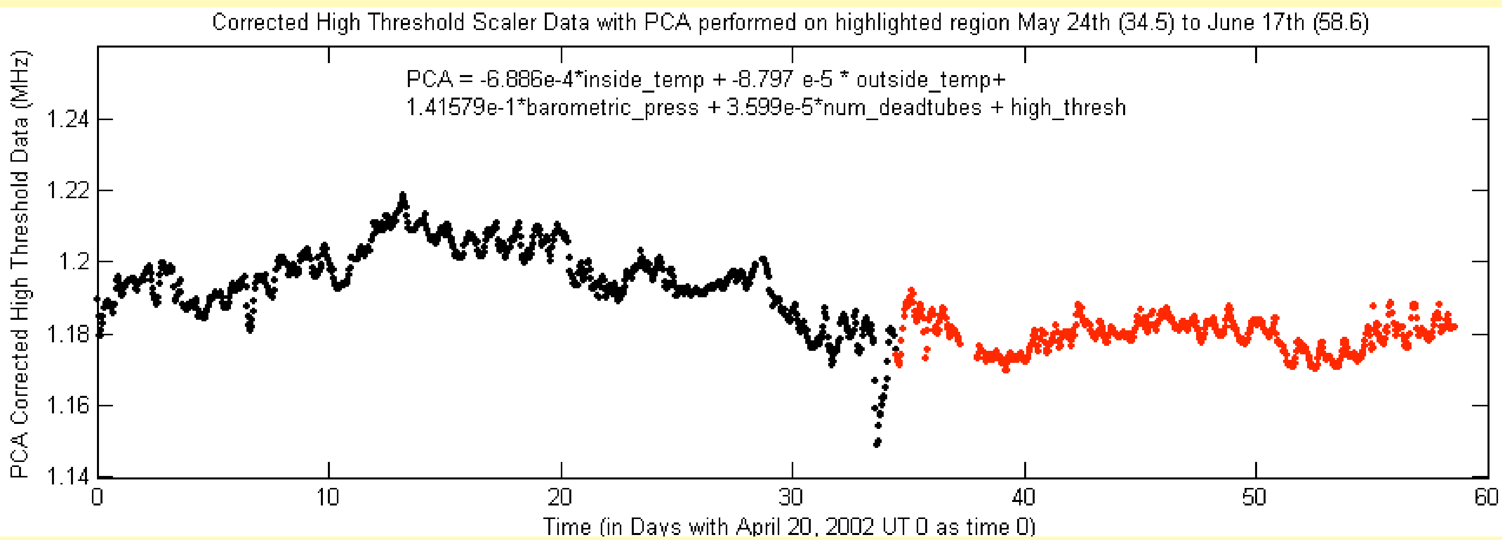
*No pond depth

PCA Applied to H-T Data
April-June 2002

Climax



Corrected H-T



Better corrections for newer data

Results so far

Data Used:

- Inside temperature
- Outside temperature
- Ground-level barometric pressure
- # of dead tubes

Not used:

- pond depth
- high altitude pressure (temp.)

1. **Good agreement with Climax data.**
2. **Care required in picking interval to characterize background.**
3. **Stability in tube count important.**
4. **Will apply to other data channels.**