Using R to find Gamma Ray Bursts
R.Giannitrapani and F.Longo
SLAC, 12-13 February 2004

Yeah, I found another language to learn ..

GLAST people will be happy ...

Dipartimento di Fisica
Università degli Studi di Udine
Sezione INFN Trieste-Udine
Don’t be fooled by the title, this is mainly a talk on my experience with the statistical analysis tool $\textbf{R}$; GRB finding is just a case study, see next three talks for serious things.

A big thank to $\textbf{M. Frailis}$ (Udine), $\textbf{N. Omodei}$ (Pisa) and $\textbf{R. Rando}$ (Padova) for helpful discussion and suggestions.

Slides made in ConTeXt, edited in emacs, rendered in PDF. Calvin and Hobbes by Bill Watterson.
I know what you are thinking: **oh no, yet another tool to learn !!!** ... But

- Really easy

- In some way similar to **Matlab, Octave** (maybe also IDL?)

- It does not replace science tools, it can be used as a
  
  - FITS viewer (like fv or DS9)
  
  - tool to try in a fast way new algorithms and ideas to implement later in another way
  
  - graphics tools to produce plots, histograms, maps etc etc
  
  - statistics calculator
  
  - much more ...
Why R?

- A statistical tool and language targeted to data analysis and exploration; it is mostly an implementation of S language
- Open source
- Multiplatform (Linux, Windows, MacOS)
- Many out of the box functionalities and a real simple array oriented language to do your own
- Big user community and lot of already available extra packages (from wavelet to decisional trees, from NN to clustering algorithms and more)
- You can extend it in Fortran, C and C++ (and in some way also in Python and Ruby) .. useful for performance boost and to interface to external data format.
DC1 Postcards

For DC1 we need an interface to FITS files

- Quite easy to build a dynamic library loadable by R using CFITSIO
- It is not yet a generic FITS interface, it is targeted to DC1 format
- At the moment it reads just the ENERGY, TIME, RA and DEC and fill a data frame, that is an etherogeneous R table
- Ask me privately if you are interested in this interface code
Here is an example of loading the FITS file and create a summary (the interface is with **Emacs**, but there are others if you dislike Ctrl-stuff)

```r
startup.R
setwd("c:/MyR/")
source("c:/MyR/GlastDC1.R")
source("c:/MyR/Count.R")
source("c:/MyR/testdll.R")

sky <- readDC1("DC1AllSky-FT1-1.fits")
summary(sky)
```

<table>
<thead>
<tr>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>5.371e+00</td>
<td>4.042e-04</td>
<td>-89.794</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>5.279e+01</td>
<td>1.013e+02</td>
<td>-34.458</td>
</tr>
<tr>
<td>Median</td>
<td>9.762e+01</td>
<td>2.058e+02</td>
<td>-4.009</td>
</tr>
<tr>
<td>Mean</td>
<td>4.636e+02</td>
<td>1.922e+02</td>
<td>-2.633</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2.167e+02</td>
<td>2.775e+02</td>
<td>28.059</td>
</tr>
<tr>
<td>Max.</td>
<td>4.879e+07</td>
<td>3.600e+02</td>
<td>89.832</td>
</tr>
</tbody>
</table>

---

**ONYATTL**

Why R?

DC1 Postcards

Finding GRB

Conclusions
I than wrote some small useful functions that help looking at the data; for example here it is a count map of the first day of DC1.
An alternative to count map is given by **kernel smoothed density map** techniques.
I selected a detail of the anticenter region and show its smooth map with contour plot (out of the box)
Ant here it is a 3D perspective view of the same anticenter region
The spectrum (not exactly) can be extracted easily and plotted in log-log scale.
Finding GRB

Let’s start by looking at the histogram of counts per seconds; there are “evident” outliers that can be used as GRB candidates.
The distribution of counts can be easily plotted ...
... and also the related **boxplot** that can be used to determine outliers in a distribution based on the quantiles.

```

Counts boxplot
```

---

**SLAC - 12-13 February 2004**
From this we can try to extract the outliers of the counts distribution as possible GRB candidates; in that case they are the seconds with more than 13 counts. This is a too small thresholds, outliers should be take with respect to a Poissons distribution (see later);
We can then extract all the photons that contribute to the outliers and plot a smoothed counts map.

Smooth counts map for quantile outliers photons
To enhance the result we can multiply the quantile threshold by a constant (as in the WIKI page). A better method is by comparison of the counts distribution with a Poisson distribution (with the same mean) by looking at the **qq-plot**.
For comparison here it is the same qq-plot for the second day (where no evident GRB are occurring)
Another way to see at this is to plot the function \( F(x) = \text{card}\{y : C(y) > x\} \) (with \( C(x) \) the counts distribution) for the DC1 data of the first day and for a Poisson distribution with the same mean.
Again here it is the same plot for the second day
Here is the time counts histogram for the extracted photons
And here there is the counts map
Another way to see the outliers is looking at the (smoothed) counts map for (RA, TIME) coordinates (or for (DEC,TIME)); doing this for all the photons gives
While for the outliers photons gives
Its more evident by looking to it in 3D perspective mode
Finally here there is a 3D scatter plot of the outliers photons
At this point we have a list of candidate GRB from this (naive) temporal analysis; to find “real” GRB we surely need also a spatial analysis on which we are working at the moment (as a rough estimate we are just studying the RA and DEC distribution in search of spatial clusters for the time intervals identified). Anyway a temporal analysis like the one presented has lots of problems; one is that we are analyzing the full sky, and this means that we are for sure missing faint gamma ray bursts that are not global outliers, but only local ones. To solve this one can constrain the temporal analysis on small spatial regions (see next talks and the WIKI pages of Nicola Omodei and David Band) or on small temporal regions (smaller than a full day).

<table>
<thead>
<tr>
<th>Time</th>
<th>RA</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3000</td>
<td>200</td>
<td>-32</td>
</tr>
<tr>
<td>27211</td>
<td>260</td>
<td>-16</td>
</tr>
<tr>
<td>43255</td>
<td>147</td>
<td>33</td>
</tr>
<tr>
<td>47274</td>
<td>???</td>
<td>??</td>
</tr>
<tr>
<td>71387</td>
<td>220</td>
<td>-35</td>
</tr>
<tr>
<td>75437</td>
<td>92</td>
<td>56</td>
</tr>
<tr>
<td>83510</td>
<td>200</td>
<td>-32</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>RA</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>176748</td>
<td>133</td>
<td>62</td>
</tr>
<tr>
<td>215701</td>
<td>252</td>
<td>27</td>
</tr>
<tr>
<td>220441</td>
<td>134</td>
<td>-3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>RA</th>
<th>DEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>410280</td>
<td>237</td>
<td>40</td>
</tr>
</tbody>
</table>

Day 1     Day 3     Day 5
To conclude, it is interesting to see what happens if one applies the cuts (CALIB VERSION) to the data and look for GRB; from the temporal counts histograms comparison it is possible to see that the outlier at 47274, one of the more evident, disappear completely (the others are reduced, but still present). This is confirmed by the spatial analysis around that instant that does not show any evident clustering.
Conclusions

Much ado about nothing .. (??)

Few concluding remarks ...

▷ Lots of things can be done in R

▷ GRBs search was an excuse to learn such a tool

▷ We are just R hobbyists at the moment, but the learning curve is quite smooth.

▷ In the future we are planning to develop more astrophysics related functions and (maybe) an interface to some indexed photons database from R (work with Marco Frailis in Udine)

▷ Give it a try, it’s free

http://www.r-project.org/
http://www.fisica.uniud.it/~riccardo/research