MAXIMUM LIKELIHOOD METHOD APPLIED TO COMPTEL SOURCE RECOGNITION AND ANALYSIS

H. de Boer[2], K. Bennett[4], H. Bloemen[2], J.W. den Herder[2],
W. Hermsen[2], A. Klumper[2], G. Lichti[1], M. McConnell[3],
J.Ryan[3], V. Schönfelder[1], A.W. Strong[1], C. de Vries[2]

[1]*Max Plank Institut für Extraterrestische Physik, D-8046 Garching, FRG*
[2]*Space Research Leiden, P.O. Box 9504, NL-2300 Leiden, NL*
[3]*Institute for the Study of Earth, Oceans and Space,*
University of New Hampshire, Durham NH, USA
[4]*ESA/ESTEC Space Science Department, NL-2200 Noordwijk, NL*

1. INTRODUCTION

As the instrumental resolution of high-energy astronomical experiments increases, the observer is confronted with a 'space of outcomes' (hereafter *dataspace*) with at most a few events per bin (particularly if time resolution comes into play). This implies that the sought signal is not only contaminated by additive noise components (*e.g.* instrumental, earth's atmosphere), but is also masked by relatively large intrinsic statistical fluctuations. In the case of the 1-30 MeV imaging telescope COMPTEL, the number of counts per bin is of the order 1, if the full resolution is to be explored.

The canonical approach to data analysis in such a situation is to model the probability distribution (pdf) of the measured quantities, based on knowledge of the instrumental response and photon intensities. Such models may contain free parameters (say θ) which one wants to constrain by the experimental result (in our case *e.g.* source position or flux). For any such set of parameters, one may calculate the probability of the measurement at hand. If one considers this probability as a function θ for a given experimental result, the appropriate name is likehood function and is denoted by $L(\theta)$. Given two hypotheses H_0 and H_1 (which may differ only in parameter values), the likelihood ratio $L(H_0)/L(H_1)$ is to be interpreted as the degree to which the data support H_0 against H_1 (Edwards, 1972). It is evident that the likelihood ratio is a convenient statistic for composite hypothesis testing and parameter estimation. In this paper we shall overview some of its properties and its application to γ-ray astronomy.

Data Analysis in Astronomy IV, Edited by V. Di Gesù *et al.*
Plenum Press, New York, 1992

In section 2 we briefly outline the interpretation of test-statistics in general, and discuss how the likelihood ratio method (LRM) can be applied in practice . We briefly describe the COMPTEL dataspace in section 3, and present some preliminary results of the method for COMPTEL (based on Monte Carlo simulations of the dataspace) in section 4.

2. BASIC THEORY

In general, the observer postulates a model of reality, which summarizes his *a priori* knowledge. This is translated into a probability distribution function (pdf) for the anticipated events encountered in the experimental set-up at hand. In astronomy the straightforward approach is to make an input intensity model map (composed of a finite number of well defined components) and fold it with the point-spread function of the telescope. A particular set of parameters defining the model map constitute a simple hypothesis, H.

It is customary to divide dataspace into bins for computational convenience and to allow for certain well-known *distribution-free* tests. Although this implies a small loss of information, it will not significantly affect parameter determination as long as the coordinates defining the dataspace (*e.g.* time, angular position, energy) are binned according to the corresponding instrumental resolution. For example, if the number of events per bin is larger than ~ 10, we may apply the minimum χ^2 test (see *e.g.* Lampton *et al.*, 1976) for parameter estimation, tacitly assuming a gaussian distribution of the number of events per bin.

Generally, the model will predict a certain continuum intensity for the number of photons per time unit per bin, say $\lambda_i(\theta)$ where i denotes the bin number. If the events have no "memory" then the probability of finding n_i counts after integration time T is distributed like $P(n_i) = e_i^{n_i} \exp(-e_i)/n_i!$, where $e_i = \lambda_i T$. If the probability for bin j is independent of the probability for bin i ($\forall i,\ j$), we can straightforwardly assign a likelihood to a dataspace of n bins under a given hypothesis H:

$$L(\{n_i\}|H) = \prod_{i=1}^{n} e_i^{n_i} \exp(-e_i)/n_i! \qquad (1)$$

We tend to express most believe in those hypotheses which maximize L. Such a maximum will be denoted by $\hat{L}$, and the corresponding (maximum likelihood-) estimates of the involved parameters θ by $\hat{\theta}$. It is here that parameter estimation and model testing (following the terminology of Cash (1979)) seemingly go separate ways. In principle, model testing may incorporate the determination of L under various functional *shapes*, *e.g.* power law versus exponential energy density spectrum. Parameter estimation concentrates on the variation of L (or another statistic) with θ under an accepted model "shape", thereby introducing a 'Bayesian preference' towards the anticipated models (Eadie *et al.*, 1971). However, in the practical cases encountered the Bayesian approach leads to a confusion between model and parameter estimation. To illustrate this, consider a pdf of the shape $\sum_k a_k S_i$, where the a_k are the amplitudes of various shapes S_k. The S_k may be rooted in physically different processes (*e.g.* pulsar radiation versus diffuse emission) and themselves contain other free parameters. One tends to consider that estimate of a_k which maximizes L as the "best"

estimate, say $\hat{a}_k$. From a non-Bayesian point-of-view we cannot truly derive the probability distribution of L for we do not know the true a_k and so we cannot assign a statistical significance to the result. However, one may derive the pdf of the $\hat{a}_k$ for supposedly true a_k^s, $\hat{P}(\hat{a}_k|a_k^s)$ and accept only those a_k^s for which $\hat{P}$ is larger then a certain value (defining the confidence interval for a_k^s). In the above example this literally means generating confidence intervals on independent models, thus blurring the distinction between model and parameter estimation. The conservative approach to the decomposition problem is to use Occam's razor: starting out from the most simple hypothesis that one still has believe in (H_0) towards more and more complex hypotheses. The LRM, although computationally slow for large dataspaces, is ideal to do the job in the case of a small number of events per bin (non-gaussian distribution) and simultaneously allows for model testing and parameter estimation. Let the general hypothesis H_g involve p undetermined parameters, $\theta = (\theta_1, \ldots, \theta_p)$. Suppose that the true values of q parameters are θ_q^t (implicitly assuming that the functional shape of H_g is appropriate). Although these values are unknown, we formally introduce a "sub-hypothesis" H_s which is H_g with the former q-parameter values set to their true values. The likelihood ratio $R(q)$ is defined by

$$R(q) = \hat{L}(\{n_i\}|H_g)/\hat{L}(\{n_i\}|H_s) \tag{2}$$

Obviously, $R \geq 1$ because H_g includes the most likely H_s. The theorem of Wilks (1938,1963) establishes that $\lambda = 2\log R(q)$ will adopt a χ^2 probability distribution with q degrees of freedom as $\sum_i n_i \to \infty$. For simple hypothesis testing, the λ statistic provides the most powerful test. So the procedure will be:

a. State the most simple hypothesis (H_0) which one has confidence in (in the example above for instance, put $a_k = \delta_{kl}$ so that all but model component l are excluded). This fixes q parameters of the general hypothesis H_g.

b. Next state a more complex hypothesis (H_1) which incorporates H_0 and which specifies r parameters of H_g, $f = q - r > 0$.

c. Calculate $R(f) = \hat{L}(\{n_i\}|H_1)/\hat{L}(\{n_i\}|H_0)$. If H_0 is true, λ will have a χ_f^2-pdf so that it is highly unlikely to find a value much in excess of $\sim 2f$. If we reject H_0 when $\lambda > \lambda_c$ the confidence level is $P(\chi_f^2 < \lambda_c)$.

d. If we reject H_0, we generate confidence intervals on the maximum-likelihood estimates of the f parameters by putting $\tilde{H}_0 = H_1(\hat{\theta}_f + \delta\theta_f)$ and $\tilde{H}_1 = H_1(\hat{\theta})$. If r of the $\delta\theta_f$ are non-zero, $\tilde{H}_0$=true implies $\tilde{\lambda} = 2\log L(\{n_i\}|\tilde{H}_1)/\log L(\{n_i\}|\tilde{H}_0)$ has a χ_r^2-pdf and we will accept those $\delta\theta_f$ for which it is not too unlikely to find $\hat{\theta}$ as maximum likelihood estimates.

The above procedure again illustrates the full equivalence of model and parameter estimation from the likelihood perspective. Obviously, having accepted H_1 the procedure may be repeated (with more complex hypotheses) if there is a physical reason to do so and if simultaneously a statistically significant improvement is achieved.

The LRM was succesfully applied in the analysis of the observations made by the γ-ray telescope COS-B ($E \approx 50\,\mathrm{MeV} - 5\,\mathrm{GeV}$). At first it was used to confirm the detection of extragalactic γ-ray sources (Pollock *et al.*, 1981). It later proved to be a convenient method for studying the properties of the galactic diffuse γ-ray emission

(Lebrun *et al.*, 1983; Bloemen *et al.*, 1986; Strong *et al.*, 1988; Bloemen 1989) and the superimposed point-like sources (Pollock *et al.*, 1985).

However, the COS-B dataspace could be described in 2 dimensions so that the input sky image was of the same dimension as the dataspace. In the case of COMPTEL, 3 dimensions are in principle required due to the nature of the measured quantities (see below). The complexity of the resulting dataspace makes a study of the likelihood method results based on simulations desirable.

3. COMPTEL DATASPACE

The Compton telescope is described in detail elsewhere in these proceedings (Schönfelder *et al.*; Diehl *et al.*). The instrument utilizes the most efficient γ-matter interaction process at few MeV energies for light nuclei, namely Compton scattering. This scattering takes place in the first layer of scintillator detectors D1 (the 'lense') after which the scattered photon is absorbed in an underlying layer of detectors D2 (the 'film'). The Compton scattering is that part of the telescope response which allows for imaging, as it is direction sensitive. However, because for unpolarized photons the corresponding cross-section depends only on the scatter angle and not on its azimuth, the reconstruction of the underlying image is not straightforward.

If the energy deposit in D1 is E_1 and in D2 is E_2, Comptel's formula for the scatter angle $\bar{\phi}$ reads

$$\cos\bar{\phi} = 1 - mc^2\left(\frac{1}{E_2} - \frac{1}{E_\gamma}\right) \quad \text{where} \quad E_\gamma = E_1 + E_2 \tag{3}$$

(m is the electron rest mass). The photon interaction positions inside the detectors are combined into a vector $\vec{r}$: the direction into which the photon has been scattered. The cone centered on $\vec{r}$ with opening angle $2\bar{\phi}$ is the collection of possible photon arrival directions. If we map the sky in an arbitrary spherical coordinate system, denoted by (χ, ψ) (which could be for example equivalent to (l, b)), the cone projects as a circle centered on the direction $(\chi(\vec{r}), \psi(\vec{r}))$ with angular radius $\bar{\phi}$. However, for every source a range of $\bar{\phi}$ will occur (distributed according to the Klein-Nishina cross-section) and to exploit this degree of freedom $\bar{\phi}$ is added as the third dimension of the dataspace in which each photon may now be described by $(\chi, \psi, \bar{\phi})$. If we take the origin of (χ, ψ) somewhere near the COMPTEL pointing axis, then for the photons of interest we can use the 'locally flat approximation' (LFA): if the source is at (χ_0, ψ_0) and not much more than $\sim 10°$ from the pointing axis, $\bar{\phi} \approx \sqrt{(\chi - \chi_0)^2 + (\psi - \psi_0)^2}$. In the LFA, the dataspace response to a source is an event-cone with apex at (χ_0, ψ_0), running at an angle of 45° with the (χ, ψ)-plane. Because of the finite resolution in $\vec{r}$ (event location within the modules) and in $\bar{\phi}(E_1, E_2)$, the cone transforms into a mantle. For each energy E_γ this mantle defines the PSF, and is denoted by $f(\chi, \psi, \bar{\phi}; \chi_o, \psi_o, E_\gamma)$. In the LFA the PSF dependency on the image coordinates is only through $(\chi - \chi_0, \psi - \psi_0)$. The effective area for COMPTON scattering may be calculated for a given pointing direction and photon arrival direction (χ', ψ'), say $A(\chi', \psi'; E_\gamma)$. Because of the finite dimensions of the instrument only a fraction of the scattered photons is absorbed in D2 with geometrical absorption probability $g(\chi, \psi)$. The expected number of events for integration time T invoked by a *mono-energetic* sky intensity distribution $I(\chi', \psi')$

can be written as

$$\tilde{e}(\chi,\psi,\bar{\phi}) = g(\chi,\psi)\int d\chi'\,d\psi' I(\chi',\psi')A(\chi',\psi')Tf(\chi,\psi,\bar{\phi};\chi',\psi') \qquad (4)$$

where explicit reference to energy has been dropped. If we denote a dataspace bin by d, a sky pixel by s and the exposure AT by X, the discretized form of (4) becomes

$$\tilde{e}(d) = g(d)\sum_{s} f(d,s)I(s)X(s) \qquad (5)$$

Additional dataspace structures, such as due to background lines (activated within the instrument) and random coincidence photons, can in general not be cast in the form of equation (5). If we assume that we can use a time averaged shape for this contribution (B) in reducing data of a given exposure, the final expectation value becomes $e(d) = \tilde{e}(d) + a_B B(d)$, where a_B can be adopted as a free parameter. Complicated dataspace selections, which must exclude most of the events arriving from the earth's atmosphere, can in principle be incorporated in the matrix g, so that the dataspace description is not altered. If all detectors are active and no explicit dataspace selections are required, g does not depend on $\bar{\phi}$ and is given by the function displayed in figure 1. The dataspace response to an on-axis point source at 6.13 MeV is given in figure 2 (based on the emperically derived PSF (Strong, 1990)). Note that the probability density is still significant at large $\bar{\phi}$, so that a typical dataspace for a single pointing contains contributions of a sky image of the size $\sim 150° \times 150°$.

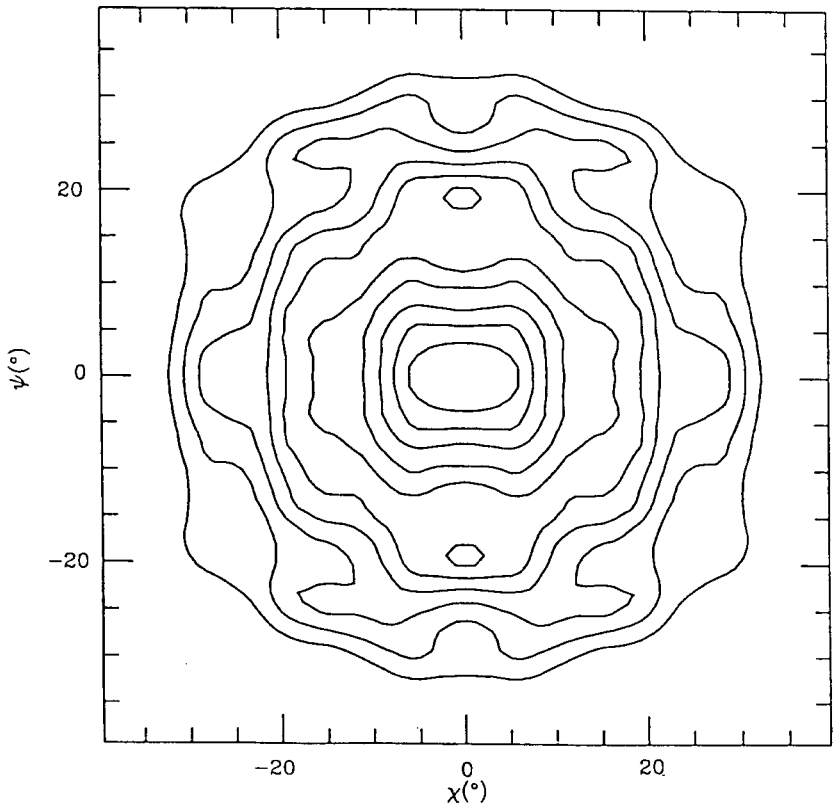

Fig. 1. The variation of the geometry function g with (χ,ψ) when COMPTEL points towards $(\chi,\psi) = (0,0)$. The fluctuations reflect the positions of the 7 D1 and 14 D2 detectors (see Diehl *et al.*, this volume). Contour levels at $n \times 0.48$, $n = 1,\ldots,10$.

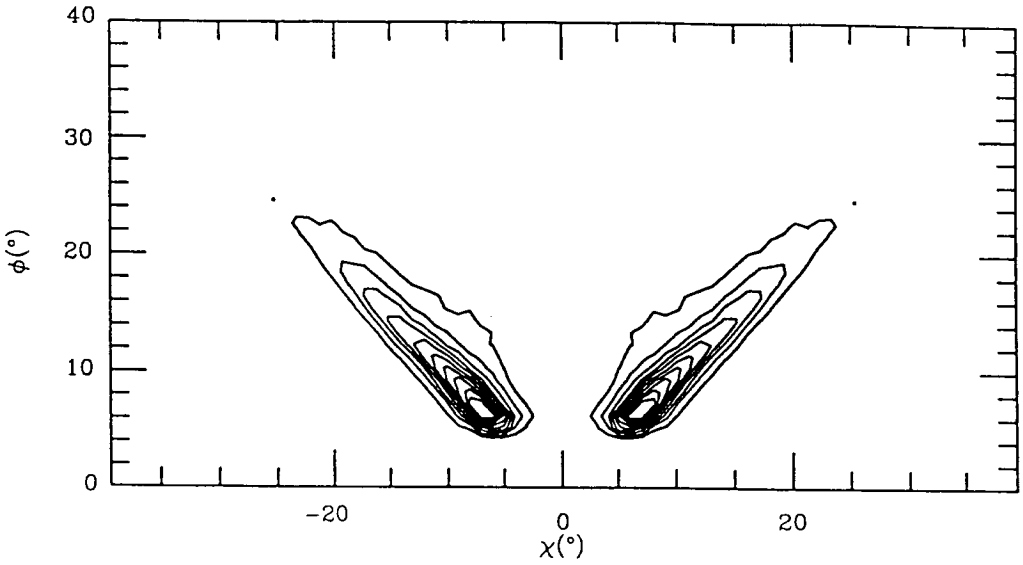

Fig. 2. A cut along $\psi = 0$ of the dataspace response ($e(\chi, \psi, \bar{\phi})$ in the text) for an on-axis source at 6.13 MeV. Contour levels at $n \times 0.16$, $n = 1, \ldots, 10$.

4. RESULTS

We simulated dataspaces for arbitrary model intensities, because calibration data have limited use for the verification of the applied LRM to flight data: the signal-to-noise ratio is large and the sources do not appear as ideal point sources because of their finite distance (*e.g.* Strong *et al.*, this volume). Furthermore we want to be able to control our input image completely. The simulation is based on equation (5), however without the LFA (see section 3). The LRM starts from an *a priori* prediction of the dataspace event density due to galactic diffuse and instrumental emission, say M_1. We test for the presence of a source and if significant, add it to the model. We can summarize this by, using the notation of the previous sections, writing the expectation per dataspace bin as

$$e(d) = \begin{cases} S_1 M_1(d) & \text{if } H_0 \\ S_1 M(d) + S_2 f(d, \tilde{s}) g(d) & \text{if } H_1 \end{cases} \qquad (6a).$$

S_1 is the scaling of the background distribution, S_2 is the source strength (proportional to the exposure) and $\tilde{s}$ is the assumed source position (χ_0, ψ_0). If we calculate λ for each $\tilde{s}$, we obtain a likelihood ratio map $\lambda(\tilde{s})$. If there is no source, then the $\max_{\tilde{s}} \lambda(\tilde{s}) \equiv \lambda_{\max}$ is distributed in a classical interpretation as χ^2_3 so that we have a 99% confidence detection if $\lambda_{\max} > 11.3$. If we accept H_1, confidence levels can be generated on both the source position and the source flux. A source which is significant 'beyond reasonable doubt' will be added to the background model at the most likely position (say $\hat{s}$), with its flux as a free scaling parameter.

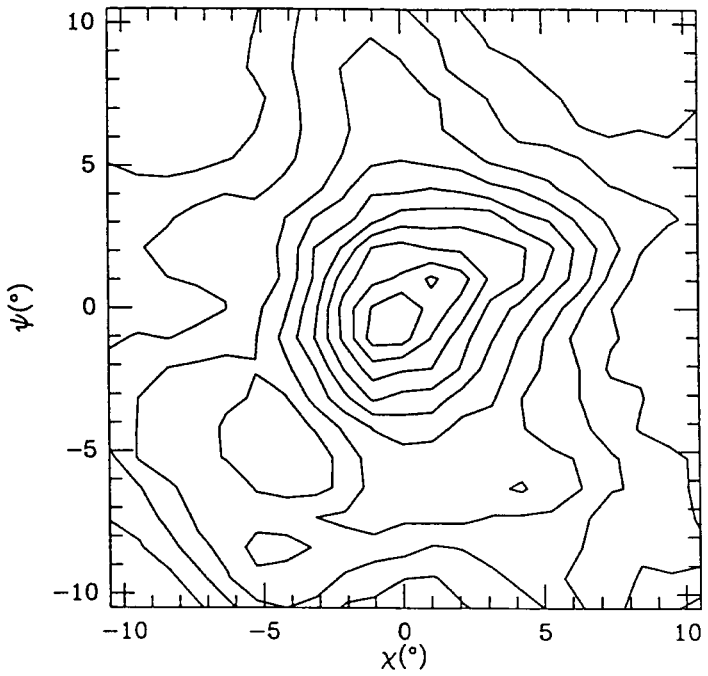

Fig. 3. A sample λ map for a single simulation; contour levels start at 8.8 with steps of 7.8. For details, see text.

The updated hypotheses become:

$$e(d) = \begin{cases} S_1 M_1(d) + S_2 M_2(d) & \text{if } H_0 \\ S_1 M_1(d) + S_2 M_2(d) + S_3 f(d,\tilde{s}) g(d) & \text{if } H_1 \end{cases} \qquad (6b),$$

where $M_2(d) = g(d) f(d, \hat{s})$.

The simulations were done with the emperically determined PSF at 6.13 MeV, and counts are typical for the integrated energy range of 3 to 12 MeV. The background is estimated mainly from balloon-flights (see *e.g.* Schönfelder *et al.*, 1980) and probably comprises about 10^5 photons for the quoted energy range and a full observation period ($\sim 4 \times 10^5$ sec effective integration time). The dataspaces used for the likelihood results are $79° \times 79° \times 40°$ in $(\chi, \psi, \bar{\phi})$ with 1° bins along each dimension. The likelihood ratio map displayed in figure 3 is for a simulated on-axis source (corresponding to (χ, ψ)=(0,0)), with about 1100 source counts whereas the $\sim 10^5$ background counts are distributed as they would be for an isotropic sky ($I_B(s) =$ constant). About 6000 source counts would approximate the number expected for the Crab total emission from 3 to 12 MeV for a single pointing. Ofcourse, since the data are not consistent with H_0, the distribution of λ in these examples is dictated by the source position and strength instead of by χ^2 statistics. The formal resolution obtained, for 1% source counts on about 10^5 counts in total, is about 2° if we adopt a 99% confidence level. The resolution quickly improves with the number of source counts. For instance, for 6% source counts it is $\sim$ 0.3°. For such a well-resolved source, the relative error in flux is characteristically less than 0.05.

A likelihood ratio map for a simulation of 2 sources, each containing about 3000 counts on an isotropic background of the same strength as above, is given at left in figure 4. Note that the simulated events exhibit statistical fluctuations, so that the λ-map is not symmetrical with respect to the $\chi = 0$-axis. The sources are separated by 6 degrees, both are 3 degrees off-axis.

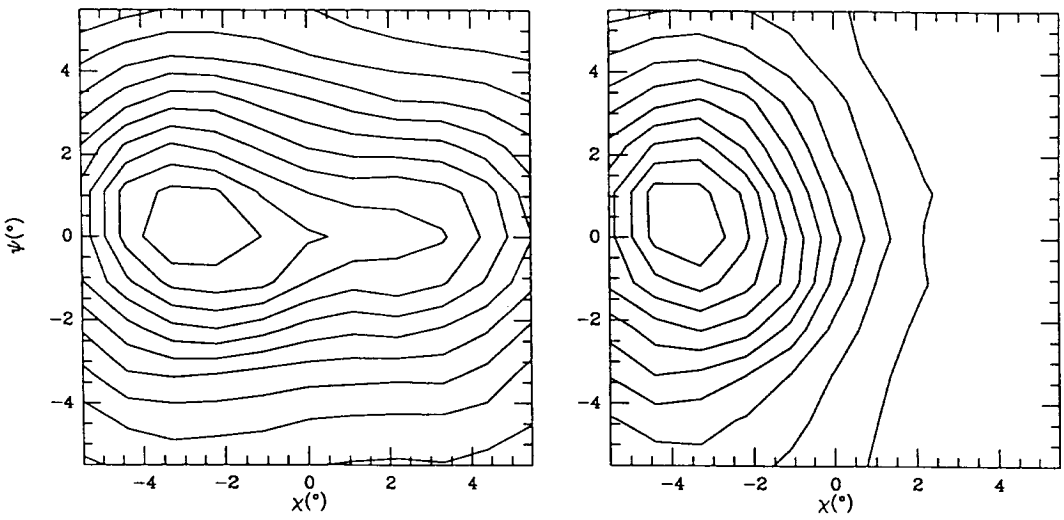

Fig. 4. *Left*: λ-map for 2 sources, at (χ, ψ)=(-3,0) and (3,0) respectively; levels start at 127, increment=35. *Right*: λ-map for the same 2 sources, but with one source included in the background model; levels start at 44, increment=26. Details are in the text.

The applied model hypotheses are given by equation (6a), which we know is wrong for this case, but still there is a clear indication of extended emission which is not consistent with the expectation for a single localized source.

Suppose we knew about a potential γ-ray source near the $\chi = +3°$ position. This would lead us to update the model to that given by equation (6b), with $\tilde{s} = (3, 0)$. We then find the remaining source with the proper number of counts (2800 ± 125) and at the right position, as we can see from the map at right shown in figure 4.

If we take the simulated data of figure 3, and apply the 4.43 MeV PSF as f in (6a), we find that the likelihood ratios have dropped characteristically by about 30%, but the formal angular resolution and source flux are consistent with the results described above. We therefore conclude that at least in this energy range, the derived source parameters are relatively insensitive to energy.

5. CONCLUSIONS

The success of previous applications of the LRM in γ-ray astronomy invites a similar approach to COMPTEL data. To verify if the method works and converges for the latter data, we presented results based on simulations of the COMPTEL dataspace. These results also allow us to anticipate the kind of likelihood ratios we may expect for realistic source detections. This is useful because in general we only have a very global *a priori* model M_B and we do not accurately know the behaviour of the likelihood ratio statistic in the presence of additional unknown sources. The simulated results in principle can tell us, for a given total number of counts, how a λ value corresponds to a signal-to-noise ratio, so that we have a reference model for interpretating flight-data.

REFERENCES

Bloemen, J.B.G.M., Strong, A.W., Blitz, L., Cohen, R.S., Dame, T.M., *et al.*, 1986, *Astron. Astrophys.* **154**, 25

Bloemen, J.B.G.M., 1989, in : *Annual Review of Astronomy and Astrophysics*, 469-516

Cash, W., 1979, *Astrophys. J.* **228**, 939

Eadie W.T., Drijard, D., James, F.E., Roos M., and Sadoulet, B., 1971: "Statistical Methods in Experimental Physics", North-Holland Publ. Comp., Amsterdam.

Edwards, A.W.F., 1972: "Likelihood", C.U.P.

Lampton, M., Margon, B., and Bowyer, S., 1976, *Astrophys. J.* **208**, 177

Lebrun, F., Bennett, K., Bignami, G. F., Bloemen, J.B.G.M., Buccheri, R., *et al.*, 1983, *Astrophys. J.* **281**, 634

Pollock, A.M.T., Bignami, G.F., Hermsen, W., Kanbach, G., Lichti, G.G., *et al.*, 1981, *Astron. Astrophys.* **94**, 116

Pollock, A.M.T., Bennett, K., Bignami, G.F., Bloemen, J.B.G.M., Buccheri, R., *et al.*, 1985, *Astron. Astrophys.* **146**, 352

Schönfelder, V., Graml, F., and Penningsfeld, F.P., 1980, *Astrophys. J.* **240**, 350

Strong, A.W., 1990, COMPASS internal report COM-AL-MPE-RES-016 (algorithm description empirical PSF generation)

Strong, A.W., Bloemen, J.B.G.M., Dame, T.M., Grenier, I., Hermsen, W., *et al.*, 1988, *Astron. Astrophy.* **207**, 1

Wilks, S.S., 1938, *Ann. Math. Stat.* **9**, 60

Wilks, S.S., 1963, *Mathematical Statistics*, Princeton University Press, Princeton.