# 利用复杂性科学领域的概念和方法来分析高能物理实验的数据

科研菜鸟
http://www.sciencenet.cn/u/sanshiphy/

从2002年至2005年4月份，孟大中教授在华中师范大学主持了一系列研究。这些研究借鉴了复杂性科学领域中的统计方法，通过对高能物理实验数据的分析，发现了一些令人印象深刻的现象，这些现象与宏观世界中复杂性科学领域所出现的现象有相似之处，从而使我们相信高能微观的世界中的所观察到的某些现象是可以从一种全新的角度，也即是研究宏观世界复杂性系统所发展起来的概念和方法来理解的。关于这方面已经发表的成果主要有：

A）Liu Qin & Meng Ta-chung, Phy. Rev. **D69**, 054026

B）Liu Qin & Meng Ta-chung, Phy. Rev. **D72**, 014011

C）兰洵，钱宛燕，蔡勖，高能物理和核物理，**Vol.30**, 94.

本文就是对这些成果做一些简略的介绍，希望能对大家的研究工作有所启发和帮助。

## 动机和目的

上个世纪七八十年代以来，人们开始关注宏观世界所展现的各种复杂性和非线性现象，随后发现了许多普遍存在的规律，并因此提出了混沌理论和分形理论等一系列理论。这些理论在研究宏观系统中经常出现的复杂多变的起伏现象时，取得了巨大的成功，同时也揭示出：

一方面，复杂多变的起伏中也存在规律，它的出现具有深刻的动力学原因。但是传统上，人们却往往把这些复杂多变的起伏看作“噪声”，也就是说把起伏看作是由于外界的各种随机或偶然性的因素所产生，与系统的动力学机制没有关系，因而常常被忽略掉。

另一方面，各种复杂多变的起伏所表现出的规律具有普适性，相同的规律经常在截然不同的系统中被观察到。

在高能碰撞强子产生的数据中也存在着复杂多变的起伏现象，比如说末态强子的快度分布。那么，在微观世界中所表现出来的起伏现象与宏观世界中的起伏现象有着相似的规律吗？如果有，它产生的动力学机制是什么？这种动力学机制与宏观复杂系统中产生复杂多变的起伏现象的机制相同吗？

另外大家知道，研究高能碰撞强子产生机制的传统做法可以被总结为“三步法”【1】:

第一步，将入射强子（一个自由强子或处于核内的一个束缚态强子）描述为一大群部分子（夸克、反夸克和胶子），其中每一个部分子都带有这个强子的纵动量的一部分（被称为Feynman- x ，往往简单地记做x ）。对于各种部分子的x的分布都已经被很好地参数化了，具体细节在相关文献中很容易找到。

第二步，人们想象射弹强子中的某些部分子被靶强子中的某些部分子所散射，他们根据由给定的QCD 拉氏量所导出的Feynman 规则发生相互作用。在“某些反应机制比另外一些反应机制更加重要”和“微扰方法可用”的假设之下，人们可以计算出相应的Feynman 图。这些计算的细节已经被解决了，在相关文献中也很容易找到。

第三步，人们假设被观察到的末态强子与在第二步中所提到的被散射的部分子是直接相关的，他们之间的关系通过“碎裂函数”来参数化地描述。所有可能种类的部分子的碎裂函数都已经被定下来了，这些细节也很容易在相关文献中找到。

对于这样一个“三步法”，如何去进行客观、科学的评价，我们认为，回忆

下列两个事实是十分重要的。

其一，到目前为止，夸克、反夸克与胶子没有被直接观测到，并且根据QCD和禁闭理论，他们是永远不可能被直接观测到的。

其二，微扰方法只能被用在动量转移非常之大，以至于相应的QCD 耦合常数小于1 的散射过程中，而在强子产生过程中，占压倒性多数的都是一些软过程，即动量转移是相当低的。

这些事实都告诉我们，要想对上述的“三步法”中的每一步分别地做一个实验上的检查是不可能的。由于不能得到实验的支持，因此，人们所能够做的唯一选择是，要么全盘接受，要么弃之一旁，另辟蹊径。针对这种情况，我们不禁要问：如果要找出高能强子产生过程的关键特征，我们需要“三步法”中所提到的那些细节信息吗？我们需要这么多的假设和可调参数吗？我们能不能利用不依赖具体物理模型的统计方法来分析实验可直接测量的数据，从而得到强子产生机制的有用信息？

综合上述观点，我们的研究“从系统地收集、整理和分析高能碰撞实验数据着手，总结规律，用适合于研究复杂系统的概念和方法来探测强子化的动力学机制，确切地说是从第一手资料，即强子产生实验结果（数据）出发，通过有效的统计方法进行数据分析，总结出客观的物理图像，并以此为基础进一步地探讨强子产生和强子化的动力学机制”

**稳定性，平稳性和标度无关性（Stability, Stationarity, Scaling）**

起伏现象的研究中最有名的一个理论是“随机行走”理论。提起这个理论，大家都会想起爱因斯坦，为解释布朗运动现象，他于 1905 年提出了这个理论。

但事实上，一名年轻的法国学生 Bachelier 早于爱因斯坦，于 1900 年研究经济学价格起伏的现象中首次提出该理论。Bachelier 认为，假设 $X(t)$ 是在时刻 $t$ 时的价格，则价格在时间 $T$ 内的变化

$$L(t,T) \equiv X(t+T) - X(t)$$

是独立平稳的高斯型随机过程，也就是说在不同时刻的 $L(t,T)$ 之间相互独立，并且满足同一高斯型分布。很长一段时间内，大家认为 Bachelier 的高斯型假设是合理的。直到上世纪 60 年代，Mandelbrot 通过大量的数据分析对 Bachelier 的理论进行了修正，他指出

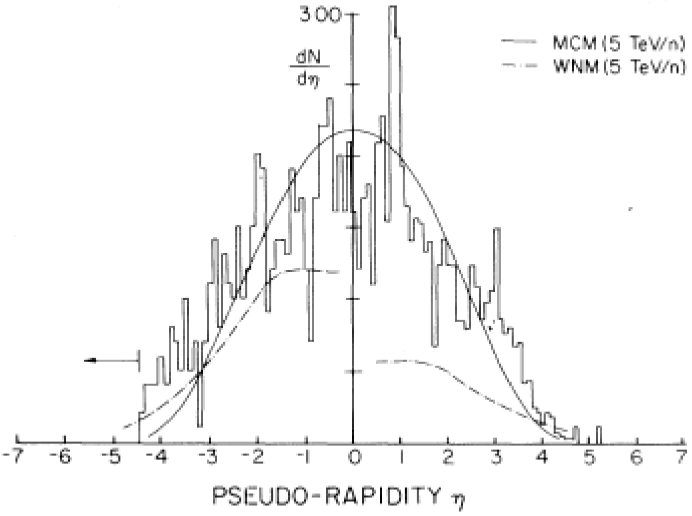
$$L_M(t,T) \equiv \ln X(t+T) - \ln X(t)$$

是独立平稳的稳定随机过程，也就是说在不同时刻的 $L_M(t,T)$ 之间相互独立，并且满足同一稳定型分布。【5】【6】

孟大中和刘勤（参见文献【1】，以下简记为 ML1）利用 Mandelbrot 等人的数据分析方法，研究了 JACEE 实验组探测宇宙线时获得的两个有名的事例【4】：每核子 4TeV 下 Si+AgBr（记为 JACEE1，见图 1）和每核子 100TeV 下 Ca + C（或 O）（记为 JACEE2，见图 2）产生的末态强子的快度分布。这两个事例的碰撞能量比实验室所能获得的能量高许多倍，除此之外，它们还因具有很高的多重数和明显的起伏而倍受关注。类似于 Mandelbrot 的做法，ML1 定义了

$$L(\eta,\Delta\eta) \equiv \ln\frac{dn}{d\eta}(\eta+\Delta\eta) - \ln\frac{dn}{d\eta}(\eta) \qquad (1)$$

并在 $L(\eta,\Delta\eta)$ 关于 $\eta$ 是独立同分布的随机变量的假设下，从三方面（稳定性，平稳性，标度无关性）对 $L(\eta,\Delta\eta)$ 进行了分析研究。下面我将一一介绍他们的研究结果。




图一

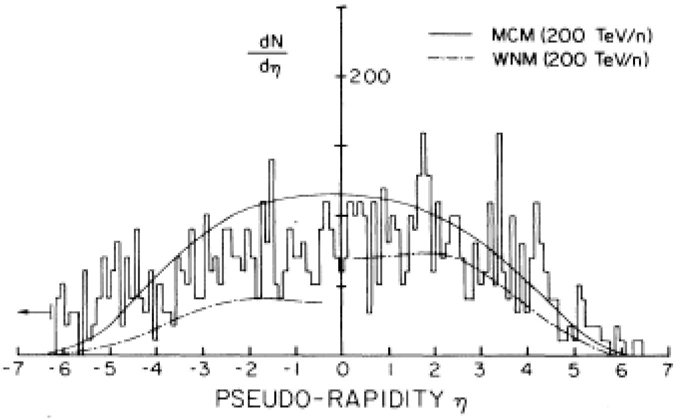



FIG. 3. The CMS pseudorapidity distribution of charged particles in the Ca+C event. Curves are the same as in Fig. 1 except for the energy.

图二

稳定性（stability）分析：分析$L(\eta,\Delta\eta)$是不是稳定型随机变量，主要分三步进行分析。

第一步：计算$L(\eta,\Delta\eta)$样本平均值$\overline{L}_n(\Delta\eta)$

$$\overline{L_n}(\Delta\eta)=\frac{1}{n}\sum_{i=1}^{n} L(\eta_i,\Delta\eta).$$

其中，对于 JACEE1 和 JACEE2 的最小快度区间均为$\Delta\eta=0.1$，分析结果见图 3（JACEE1）和图 4（JACEE2）。我们可看到两个事例的平均值都趋为零。

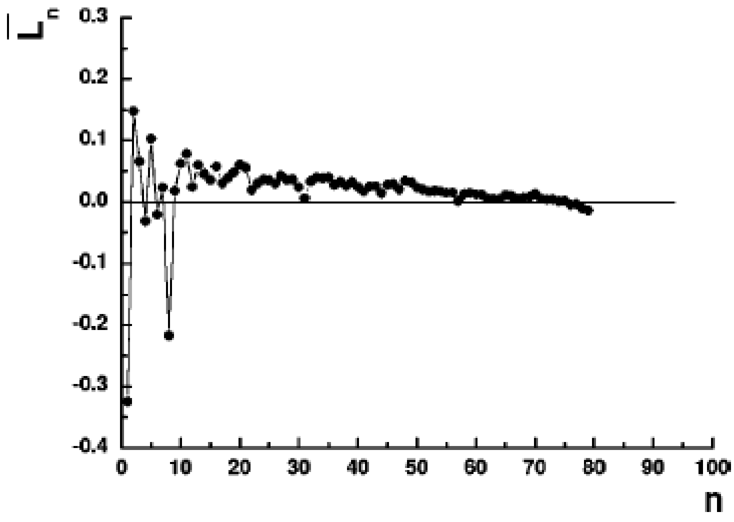



图 3

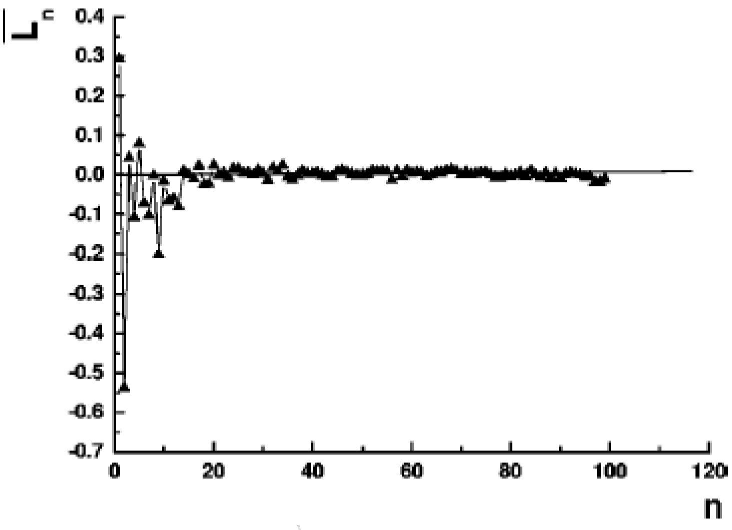



图 4

第二步：将 $L(\eta,\Delta\eta)$ 按正负分为两组，对正的一组计算右尾分布 $P(L>L^{+})$ ，对负的一组计算左尾分布 $P(L<L^{-})$ ,其中 P 表示几率 ,例如 $P(L>L^{+})$ 表示 $L>L^{+}$ 的几率有多大。计算结果见图 5（JACEE1）和图 6（JACEE2）。图中为方便和清晰起见，将 $P(L<L^{-})$ 关于 L = 0 做了镜像反演和平移。

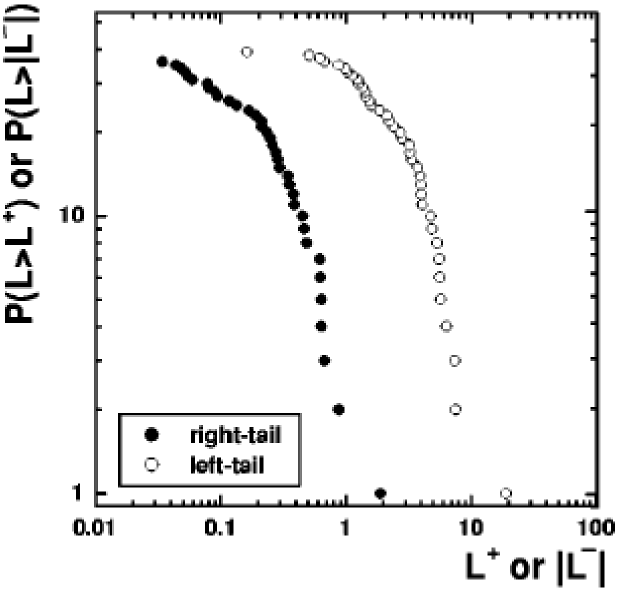

P(L>L+) or P(L>|L−|)
10
1
right-tail
left-tail
0.01
0.1
1
10
100
L+ or |L−|

图 5

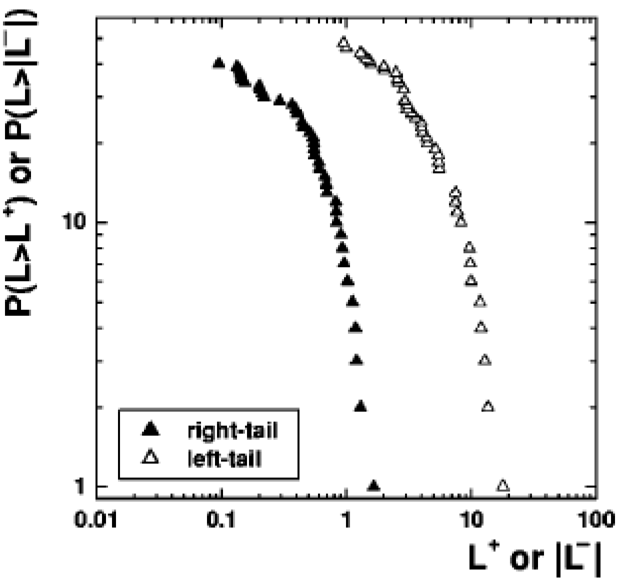

P(L>L+) or P(L>|L−|)
10
1
right-tail
left-tail
0.01
0.1
1
10
100
L+ or |L−|

图 6

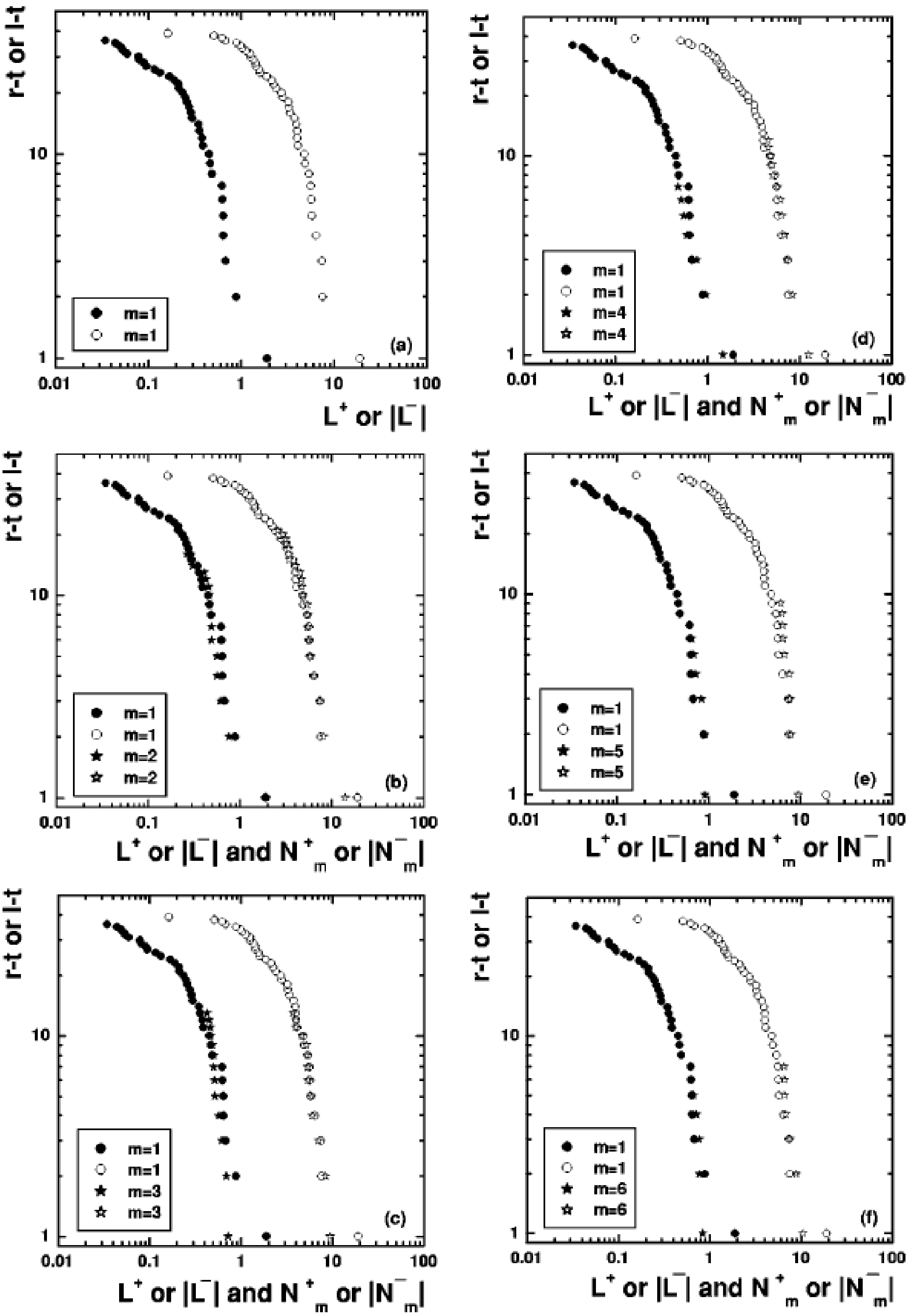

r-t or l-t
m=1
m=1
(a)
L+ or |L−|
m=1
m=1
m=2
m=2
(b)
m=1
m=1
m=3
m=3
(c)
m=1
m=1
m=4
m=4
(d)
m=1
m=1
m=5
m=5
(e)
m=1
m=1
m=6
m=6
(f)
L+ or |L−| and N+m or |N−m|
0.01
0.1
1
10
100

图 7

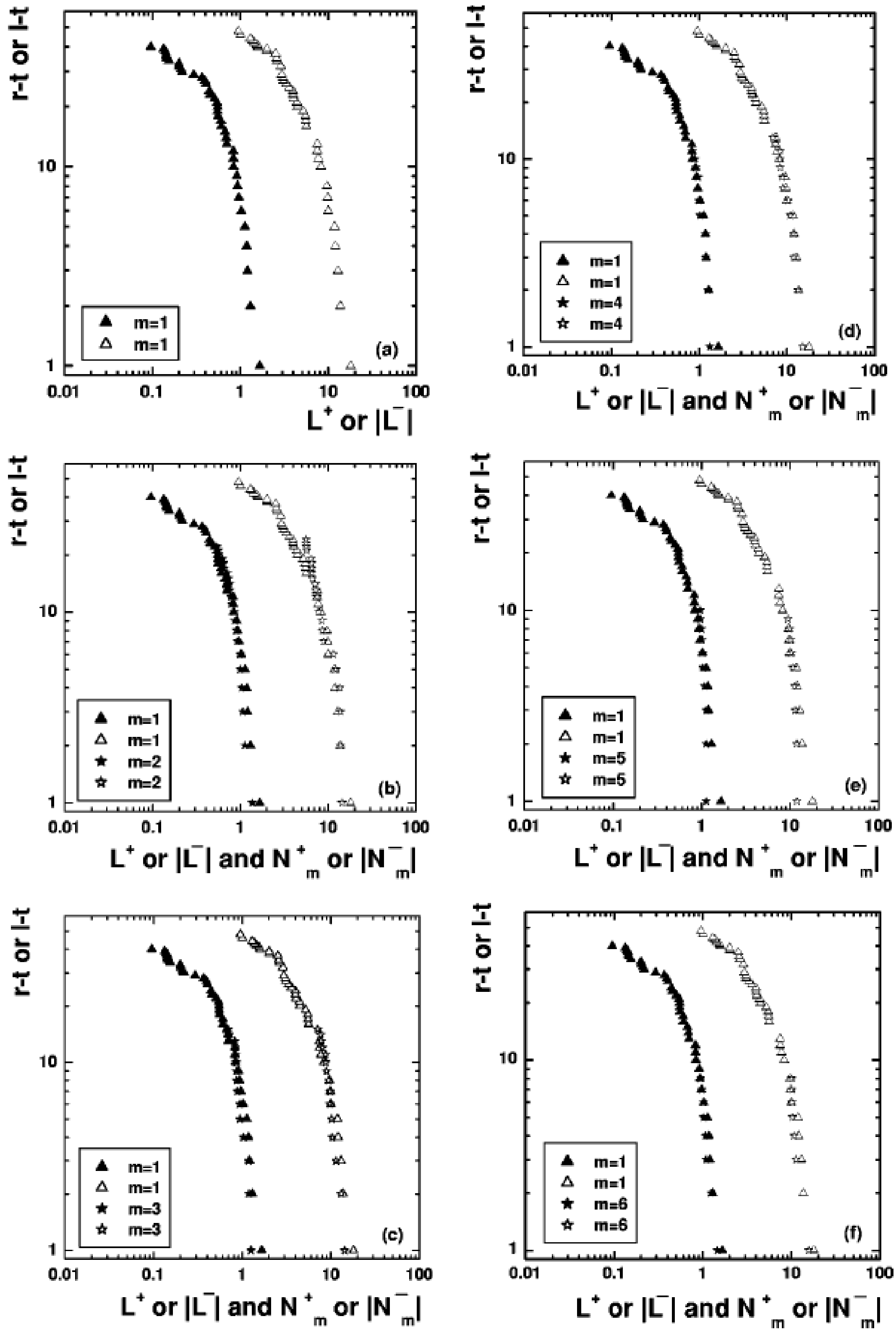

r-t or l-t
L+ or |L−|
L+ or |L−| and N+m or |N−m|
m=1
m=2
m=3
m=4
m=5
m=6
(a)
(b)
(c)
(d)
(e)
(f)
0.01
0.1
1
10
100

图 8

第三步：定义

$$
\begin{aligned}
S_m &= \sum_{i=1}^{m} L(\eta_i, \Delta\eta) \\
&= L(\eta, m\Delta\eta) \\
&= \ln\frac{dN}{d\eta}(\eta + m\Delta\eta) - \ln\frac{dN}{d\eta}(\eta)
\end{aligned}
$$

按照稳定随机变量的定义，为了验证 $L(\eta,\Delta\eta)$ 是稳定型随机变量，必须验证：

$$c_m^{-1}(S_m - \gamma_m) \overset{d}{=} L \qquad (2)$$

其中，“$\overset{d}{=}$”表示具有相同的随机分布。另外，$c_m = m^{1/\alpha}$，$0 < \alpha \le 2$，$\gamma_m$ 是任意一个实数，m 是所有大于 1 的整数。验证结果见图 7( JACEE1 )和图 8( JACEE2 )，当然我们不可能验证所有 m 大于 1 的情况，不过从验证结果看，m 直到 6，( 2 ) 式都满足的很好，这就充分的说明了 $L(\eta,\Delta\eta)$ 是稳定型随机变量。

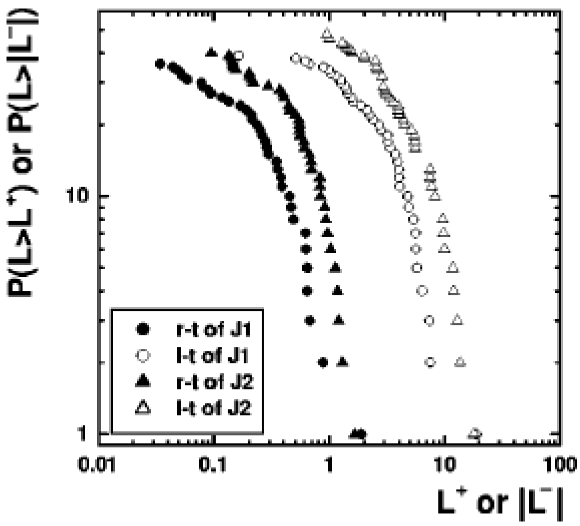



图 9

平稳性（Stationarity）分析：验证从不同的时间（或时间段）内观测到的快度分布中得到的 $L(\eta,\Delta\eta)$ 是否具有相同的统计性质。ML1 比较了 JACEE1 和 JACEE2 的左尾分布和右尾分布（见图 9），分析的结果很令人意外，尽管这两个事例具有不同的碰撞能量以及不同的靶粒子和射弹粒子，并且发生在不同的时

间，但是它们的统计性质非常相似。

标度无关性( Scaling )分析 对于稳定型随机变量 数学上可以证明 ,当$\alpha = 2$时，该稳定性随机变量就是大家通常熟悉的高斯型随机变量，具有有限的平均值和方差。当$\alpha \neq 2$时，稳定型随机变量不具有有限的方差。我们已经前面验证了 JACEE 的两个事例的$L(\eta, \Delta\eta)$具有稳定型分布的特征，那么他们到底是方差有限的高斯分布还是方差无限的稳定型分布呢？当然，无论 JACEE 满足那种分布，我们总是可以计算其样本方差，这个值总是有限的。通过对 JACEE 样本方差的分析能不能帮助我们回答这个问题呢？

ML1 中定义$L(\eta, \Delta\eta)$的样本方差：

$$S_n^2(\Delta\eta) = \frac{1}{n-1}\sum_{i=1}^{n}[L(\eta_i, \Delta\eta) - \overline{L}_n(\Delta\eta)]^2 \qquad (3)$$

根据（3）式，JACCE1 的结果见图 10，JACEE2 的结果见图 11，从这两张图中都没有看出 JACEE 的样本方差趋于一个定值，这说明了 JACEE 很可能是方差无限的稳定分布。ML1 又根据图 10 和图 11 的结果分析了$S_n^2$的频率分布（未

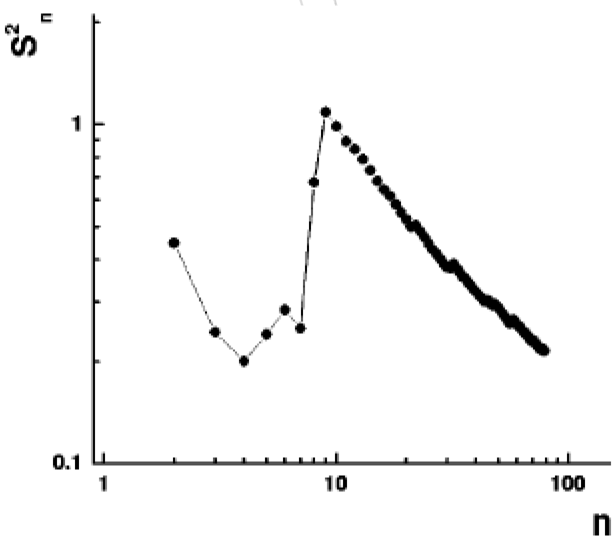



图 10

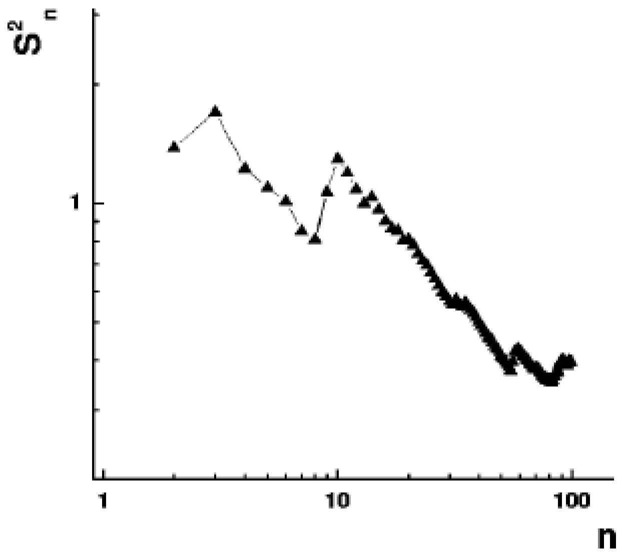



图 11

归一）$P(S_n^2 > s_n^2)$，分析结果见图 12（JACEE1）和图 13（JACEE2）。为了比较，ML1 又给出了与 JACEE 样本数相同的高斯型随机变量的分析结果。从图中可以清晰的看出 JACEE 的样本方差频率分布与高斯型样本有着很大的不同，这说明了 JACEE 满足方差无限的稳定型分布。

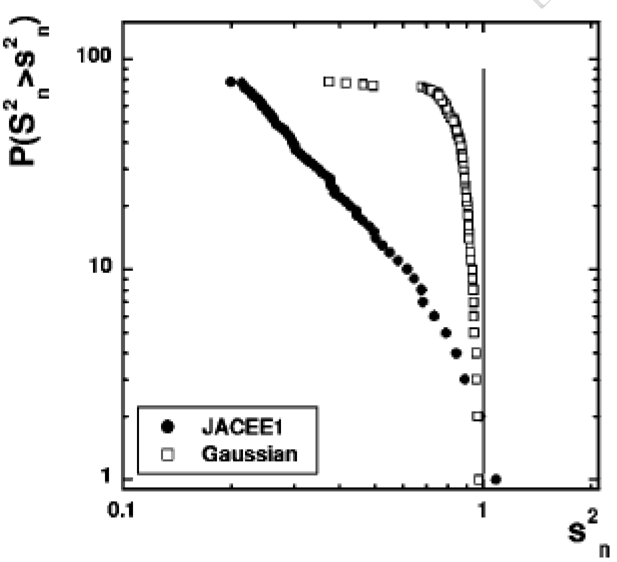



图 12

我们在回头看看图 12 和 13，发现 JACEE1 和 JACEE2 的频率分布在双对数图上近似是一条直线，这表明 JACEE 事例的 $S_n^2$ 的频率分布满足幂次率，这个性质M L 1 称为标度无关性（Scaling）。

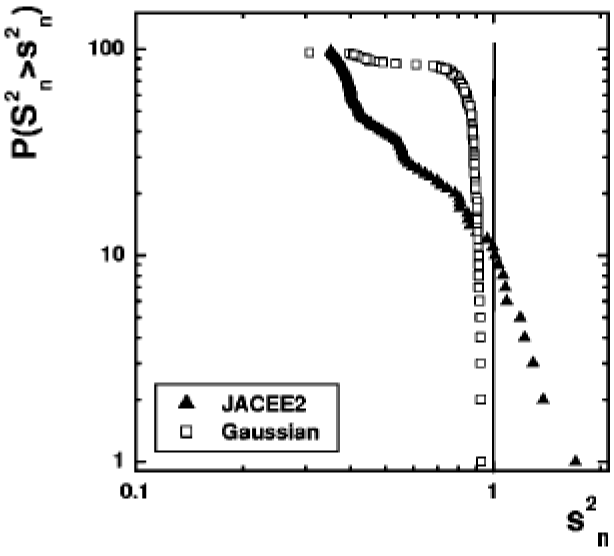



图 13

以上分析表明了 JACEE 事例中的起伏满足稳定性，平稳性以及标度无关性。对与其它的事例，特别是在对撞机上末态强子的快度分布是否也具有上述性质？兰洵对 STAR（$\sqrt{S_{NN}} = 200GeV$ 的 Au + Au）和 EMU01（200AGeV 的 $^{32}S + Au$）中的高多重数和大起伏的事例进行了同样的分析，发现 STAR 事例中有 61%是非高斯分布，EMU01 事例中有 84%是非高斯分布，这些非高斯事例都具有标度无关性【3】，但是平稳性并不普遍存在。

## 全局统计依赖性（Global Statistical Dependence）

全局统计依赖性也称长程统计依赖性（Long - run Statistical Dependence），高能碰撞产生的末态强子中是否存在全局统计依赖性还是一个没有解决的问题，人们曾进行了各种各样的努力来解决这个问题，但都是像“三步法”那样依赖于具体细节的描述，以及需要大量的假设和输入参数（具体的调研情况及对各种方法精彩的评述，请参看文献【2】，以下简记为 ML2）。那么我们能不能寻找一种方法，它能有效的分析高能碰撞末态强子起伏现象中存在的全局统计依赖性，并且满足下列特性：

其一，不需要知道基本组成成分及其之间相互作用的所以动力学细节；

其二，不需要利用微扰 QCD 或唯象的模型来做计算；

人们分析数据中存在什么样的统计依赖性，主要是利用谱分析的方法（Spectral analysis）。但是 Mandelbrot 指出，对于远离高斯的情况，谱分析的方法会失效【7 - 9】。上一节我们分析了 JACEE 快度分布中的 $L(\eta, \Delta\eta)$ 是方差无限的稳定型分布，那么我们所寻找的用来分析统计依赖性的统计方法，还应具有下列特性：不仅能够适用于统计独立和具有短程统计依赖性的情况，还要能适用于长程统计依赖性的情况。

ML2 中建议了一种满足这些特征的统计方法——R/S 分析，这种分析方法起源于 Hurst，他是一位地球物理学家。为了根据往年的河流流量数据来设计一个既不会干涸也不会溢出的理想水库，他发明了一种统计方法来分析每年河流流量记录，并因此发现了一种奇特的规律，后人称之为 Hurst 定律。更有意思的是他和其他的人随后又将这种统计方法用在其它的领域，比如说经济学中价格的起伏，太阳黑子数目的变化等等，都发现了有 Hurst 定律存在【10】。Hurst 定律为什么会普遍的存在？它产生的机制是什么？这些都还是有待解决的问题。Mandelbrot 随后根据大量的数值分析结果指出，Hurst 发明的统计方法可以用来分析统计依赖性，并且是普遍适用的。Mandelbrot 将 Hurst 的原始统计方法做了数学化和系统化的处理，处理后的统计方法被称为 R/S 分析【8】。

为形象起见，我们以 Hurst 原始表述为例来介绍 R/S 分析。设有一段持续时间为 $\tau$ 的年度流量记录 $\xi(t)$，设计一个水库，每年平均排水量为：

$$\langle \xi \rangle_{t_0,\tau} = \frac{1}{\tau} \sum_{t=t_0}^{t_0+\tau} \xi(t),$$

这样一个水库每年的存水量为：

$$X(t_0, t, \tau) = \sum_{u=t_0}^{t_0+t} \{\xi(u) - \langle \xi \rangle_{t_0,\tau}\}.$$

那么在这一段时间$\tau$内，水库的最大存水量和最小存水量之差，即为所应设计的水库的容量为：

$$R(t_0, \tau) = \max_{0 \le t \le \tau} X(t_0, t, \tau) - \min_{0 \le t \le \tau} X(t_0, t, \tau).$$

R 再除以方差

$$S(t_0, \tau) = \left( \frac{1}{\tau} \sum_{t=t_0}^{t_0+\tau} \{\xi(t) - \langle \xi \rangle_{t_0,\tau}\}^2 \right)^{1/2}$$

就得到了一个无量纲的量 R/S，Hurst 发现对于很多自然现象都有

$$\frac{R}{S}(t_0, \tau) \sim \tau^{H(t_0)}$$

这个规律被称为 Hurst 定律。

再对初始值进行平均，得到

$$\langle R/S(t_0, \tau) \rangle_{t_0} \sim \tau^H$$

R/S 分析指出：

H = 0.5，不存在全局统计依赖性，包括方差存在或者不存在的统计独立，马尔科夫或短程统计依赖的随机过程；

0.5<H<1，存在全局统计依赖性，对应着长程统计正关联；

0<H<0.5，存在全局统计依赖性，对应着长程统计负关联。

全局统计依赖性的程度取决于 H 偏离 0.5 的程度，H 偏离 0.5 越远，全局统计依赖性就越强。

ML2 用来对快度分布做 R/S 分析的公式如表一所示，为了便于比较，表一

中重新给出了上述计算年度流水量 R/S 的公式。

表一

| 排水量 $\xi(t)$ | 快度分布 $\ln\frac{dN}{dy}(y)$ |
|---|---|
| $\langle\xi\rangle_{t_0,\tau}=\frac{1}{\tau}\sum_{t=t_0}^{t_0+\tau}\xi(t),$ | $\left\langle\ln\frac{dN}{dy}\right\rangle_{y_i,Y}=\frac{1}{Y}\sum_{y=y_i}^{y_i+Y}\ln\frac{dN}{dy}(y).$ |
| $X(t_0,t,\tau)=\sum_{u=t_0}^{t_0+t}\{\xi(u)-\langle\xi\rangle_{t_0,\tau}\}.$ | $X(y_i,y,Y)=\sum_{u=y_i}^{y_i+y}\left\{\ln\frac{dN}{dy}(u)-\left\langle\ln\frac{dN}{dy}\right\rangle_{y_i,Y}\right\},$ |
| $R(t_0,\tau)=\max_{0\le t\le\tau}X(t_0,t,\tau)-\min_{0\le t\le\tau}X(t_0,t,\tau).$ | $R(y_i,Y)=\max_{0\le y\le Y}X(y_i,y,Y)-\min_{0\le y\le Y}X(y_i,y,Y)$ |
| $S(t_0,\tau)=\left(\frac{1}{\tau}\sum_{t=t_0}^{t_0+\tau}\{\xi(t)-\langle\xi\rangle_{t_0,\tau}\}^2\right)^{1/2}$ | $S(y_i,Y)=\left(\frac{1}{Y}\sum_{y=y_i}^{y_i+Y}\left\{\ln\frac{dN}{dy}(y)-\left\langle\ln\frac{dN}{dy}\right\rangle_{y_i,Y}\right\}^2\right)^{1/2}$ |
| $\frac{R}{S}(t_0,\tau)\sim\tau^{H(t_0)}$ | $\frac{R}{S}(y_i,Y)\sim Y^{H(y_i)},$ |

ML 的分析结果如图 14 所示。图 14 中的（a）和（b）以及（c）和（d）分别对应不同的初始快度 $y_i$ 的分析结果。ML 也还对其它初始快度做了分析，结果就像图 14 所展示的那样，Hurst 定律对不同的初始快度都成立，并且指数 H 不依赖于初始快度 $y_i$.

兰洵按上述方法对STAR和EMU01的数据进行了分析【3】,所得到的结论是：Hurst定律普遍成立，EMU01的指数H都在0.9左右，STAR的指数都在0.6左右，也就是说这两个实验组的末态带电强子的快度分布具有全局统计依赖性。

## 结语

让我们回顾一下前面所讲的孟大中教授等人所做的工作：一是根据 Bacheliei 和 Mandelbrot 等人在经济学领域所采用的分析方法，分析了 JACEE 实验组末态强子快度分布的记录，指出两个连续快度区间强子数的相对变化是非高斯的稳定性分布，并且满足平稳性和标度无关性。二是根据 Hurst 和 Mandelbrot 等人在水利工程学中所采用的分析方法，同样分析了 JACEE 实验组末态强子快度分布的记录，指出其中存在全局统计依赖性。兰洵对利用 STAR 和 EMU01 的数据验证了上述许多结论是普遍成立的。

孟大中教授等人的研究结果表明，复杂性领域的概念和方法是可以有效的用来分析高能物理的实验数据的。那么在复杂性领域，还有那些方法和概念可以被用来分析高能数据，以从中提取更多的有用信息？这是一个值得探索的问题，我们曾经设想过一些研究方案，可参考文献【11】【12】。另外，他们的研究结果也提出了一个值得深思的问题：我们该如何理解这些普遍存在的规律？传统的理论方法能不能解释它们呢？

从我个人的观点来看，孟教授等人的上述研究也还存在着值得商榷的地方。比如说对 JACEE 数据的 R/S 分析，大家再回过头来看看 JACEE 的两个事例，粗看起来，它们的形状像两个大三角形。我对三角形数据做了 R/S 分析，发现 Hurst 定律满足的很好，并且指数 H 约在 0.9 左右。这是不是告诉我们，利用 R/S 分析 JACEE 数据得到的结论是平庸的，它只不过说明了 JACEE 的快度分布像个大三角？事实上，R/S 不是万能的，有些作者讨论过它的局限性，至于它何时适用的问题，据我所知到目前为止并没有一个明确的说法。因此在应用 R/S 分析时应慎重，否则会得到错误的或平庸的结论。另外，我们关心的是，除了能动量守恒以外，末态强子中是否存其它机制造成的长程统计依赖。R/S 分析能不能将能

动量守恒造成的长程统计依赖与其它机制区分开来呢？这也是一个需要深入思考的问题。

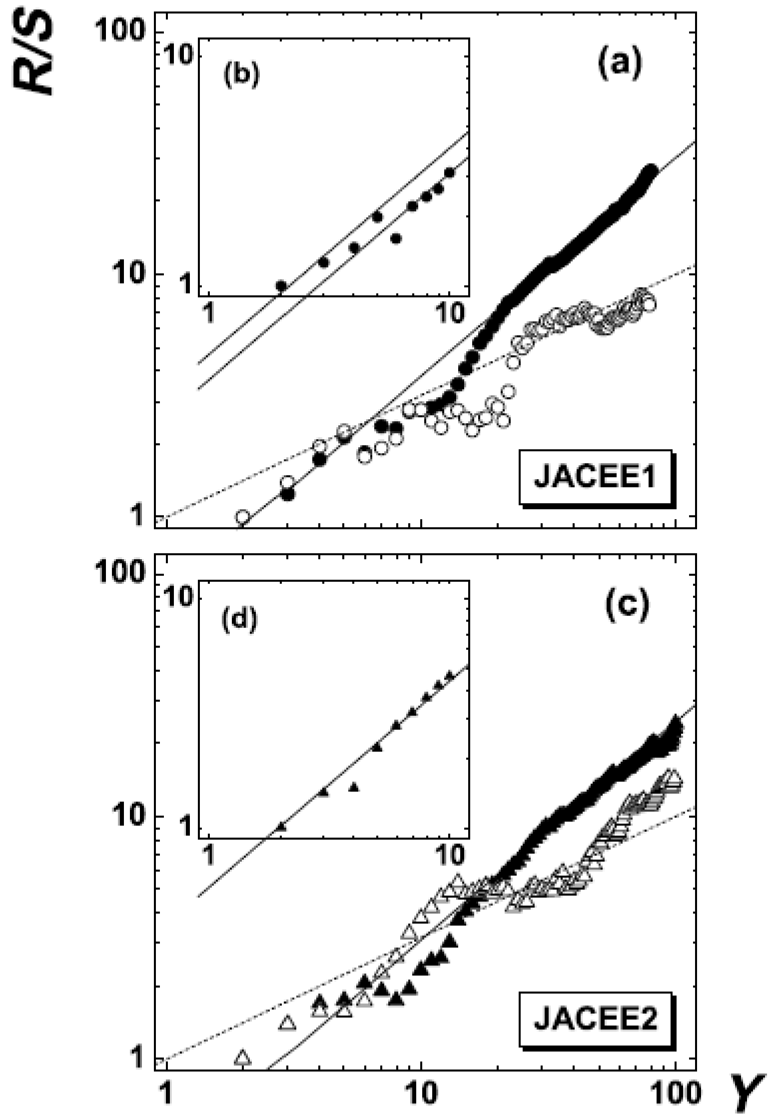



图 14



## 参考文献

【1】Liu Qin & Meng Ta-chung, Phy. Rev. **D69**, 054026

【2】Liu Qin & Meng Ta-chung, Phy. Rev. **D72**, 014011

【3】兰洵，钱宛燕，蔡勖，高能物理和核物理，**30**, 94

【4】JACEE Collaboration，T.H. Burnett *et al.*, Phys. Rev. Lett. **50**, 2062

【5】B.B. Mandelbrot, J. Business **36**, 394

【6】B.B. Mandelbrot, J. Business **40**, 393

【7】B. B. Mandelbrot, *Fractals and Scaling in Finance*（Springer, New York, 1997）, and the references given there in.

【8】B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*（Springer, New York, 2002）, and the references given there in.

【9】B. B. Mandelbrot, *The Fractal Geometry of Nature*（Freeman, San Francisco, 1999）

【10】J. Feder, *Fractals* （Plenum Press, New York, 1988）

【11】兰洵，浅谈分形，多重分形及其在物理中的应用，华中师范大学2004年学士论文

【12】刘磊，“分形时间”及其对物理研究的启发，华中师范大学2004年学士论文