数据分析

董绪强

中科院高能所

校园宇宙线观测联盟技术组

COCO2022

校园宇宙线观测暑期学校

感受国之重器，从宇宙线观测开始

7月23-24日 北京

主办单位

校园宇宙线观测联盟

支持单位

中国科学院高能物理研究所 Institute of High Energy Physics Chinese Academy of Sciences

国科大关心下一代工作委员会

中国科学院大学教育基金會 University of Chinese Academy of Sciences Education Foundation

创益家 HOME

国家高能物理科学数据中心 National HEP Data Center

2022/7/25

原初宇宙线的事例率

原初宇宙线的“烫衣抛弹”多频繁

讲授人：董绪强

CCO2022

来自ICD2021_Booklet

目录/CONTENTS

1. 宇宙线
2. 事例率
3. 数据处理
4. 总结与思考

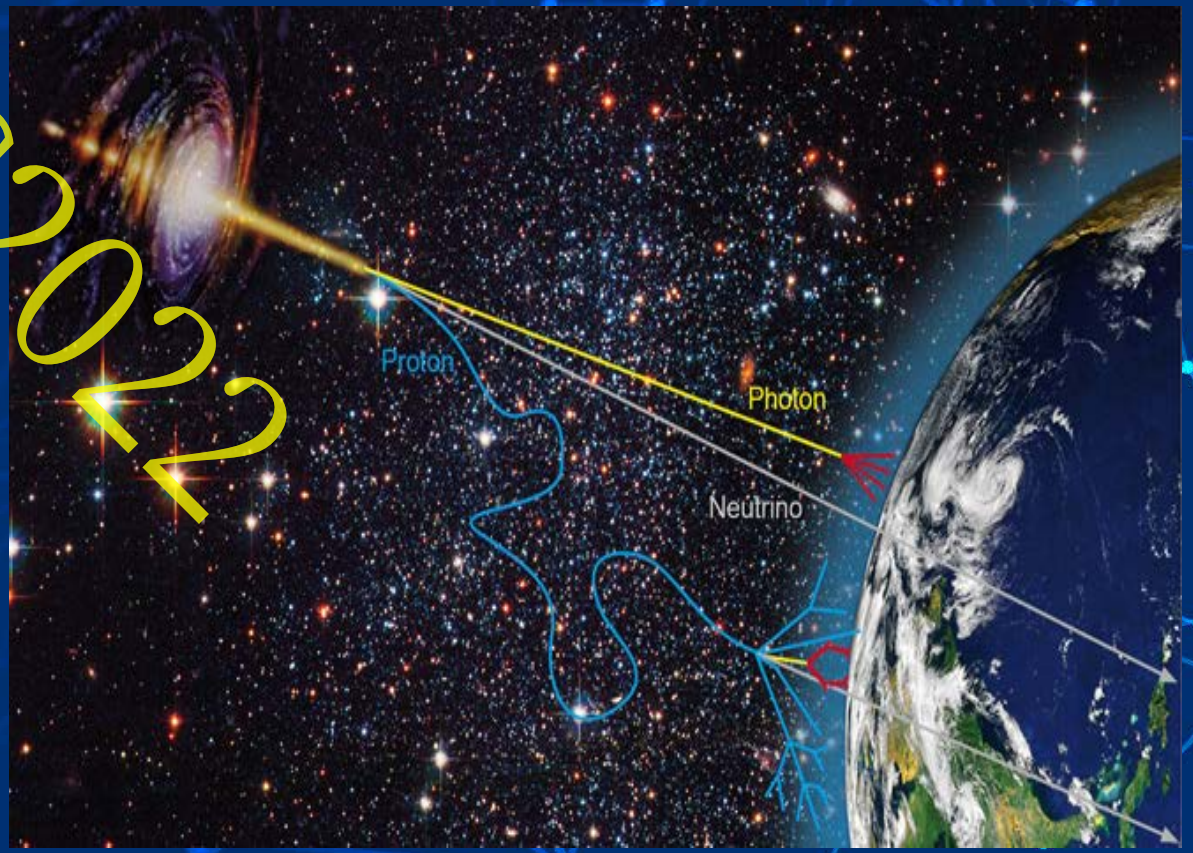
CCOC2022

这样的“烫衣抛弹”多久被我们探测到一次呢？

宇宙线

- 来自遥远天体“抛射”的高能粒子流

CCOC2022



事例率（频率）介绍

- 事例率：单位时间探测到的原初宇宙线粒子数，单位：HZ
- 探测器有一定概率可以探测到粒子，因此粒子的探测是随机事件

- 统计学：
  - 随机事件服从泊松分布
  - 随机事件的时间间隔服从指数分布 $P(t) = \frac{1}{\lambda} e^{-\lambda t}$

$$P(N(t) = n) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}$$

事例率介绍——泊松分布

- 有n个粒子到达探测器，探测器平均探测到λ个粒子，探测概率就为p=λ/n，探测不到为1-p=1-λ/n
- 由二项分布，探测到k个粒子的概率为

$$C_n^k p^k (1-p)^{n-k} \longrightarrow \frac{n!}{(n-k)!\times k!}\times(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k}$$

$$= \frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}\lambda^k\frac{1}{k!}(1-\frac{\lambda}{n})^{n-k}$$

$$= 1(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})\lambda^k\frac{1}{k!}(1-\frac{\lambda}{n})^{n-k}$$

$$= 1(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})\lambda^k\frac{1}{k!}\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}$$

$$1(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n}) \to 1$$

$$(1-\frac{\lambda}{n})^k \to 1^k \to 1$$

$$(1-\frac{\lambda}{n})^n = (1+\frac{-\lambda}{n})^n \to e^{-\lambda}$$

$$= 1(1-\frac{1}{n})(1-\frac{2}{n})\dots(1-\frac{k-1}{n})\lambda^k\frac{1}{k!}\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \approx 1\times\lambda^k\frac{1}{k!}\frac{e^{-\lambda}}{1} = \frac{\lambda^k e^{-\lambda}}{k!}$$

此时λ为事例率

$$C_n^k p^k (1-p)^{n-k}$$

2022/7/25

事例率介绍——指数分布

- 假设小于t的时候没有探测到事例，那么

$$P(X < t) = P(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

- 所以我们可以先得到时间的分布，通过指数函数拟合，就可以得到事例率了

COCO2022

数据处理——数据结构(重建后)

探测器相对触发时间

触发探测器个数

A	B	C	D	E	F	G
t0 /ns	t1 /ns	t2 /ns	t3 /ns	t4 /ns		Nhit
96.2	94.0	89.3	92.6	0.0		4
97.0	100.6	0.0	0.0	95.8		3
0.0	0.0	92.5	91.4	99.0		3

N	O	P	Q	R	S	T
trg_num		l	m	phi0	phi	theta
36582		-0.217141	-0.274075	-0.900788	4.0423876	0.3572166
36583		0.2877614	-0.392067	-0.93764	5.3453597	0.507892
36584		0.0907706	-0.916393	-1.472066	4.8109336	1.1703256

事例序列

重建参数

方位角

天顶角

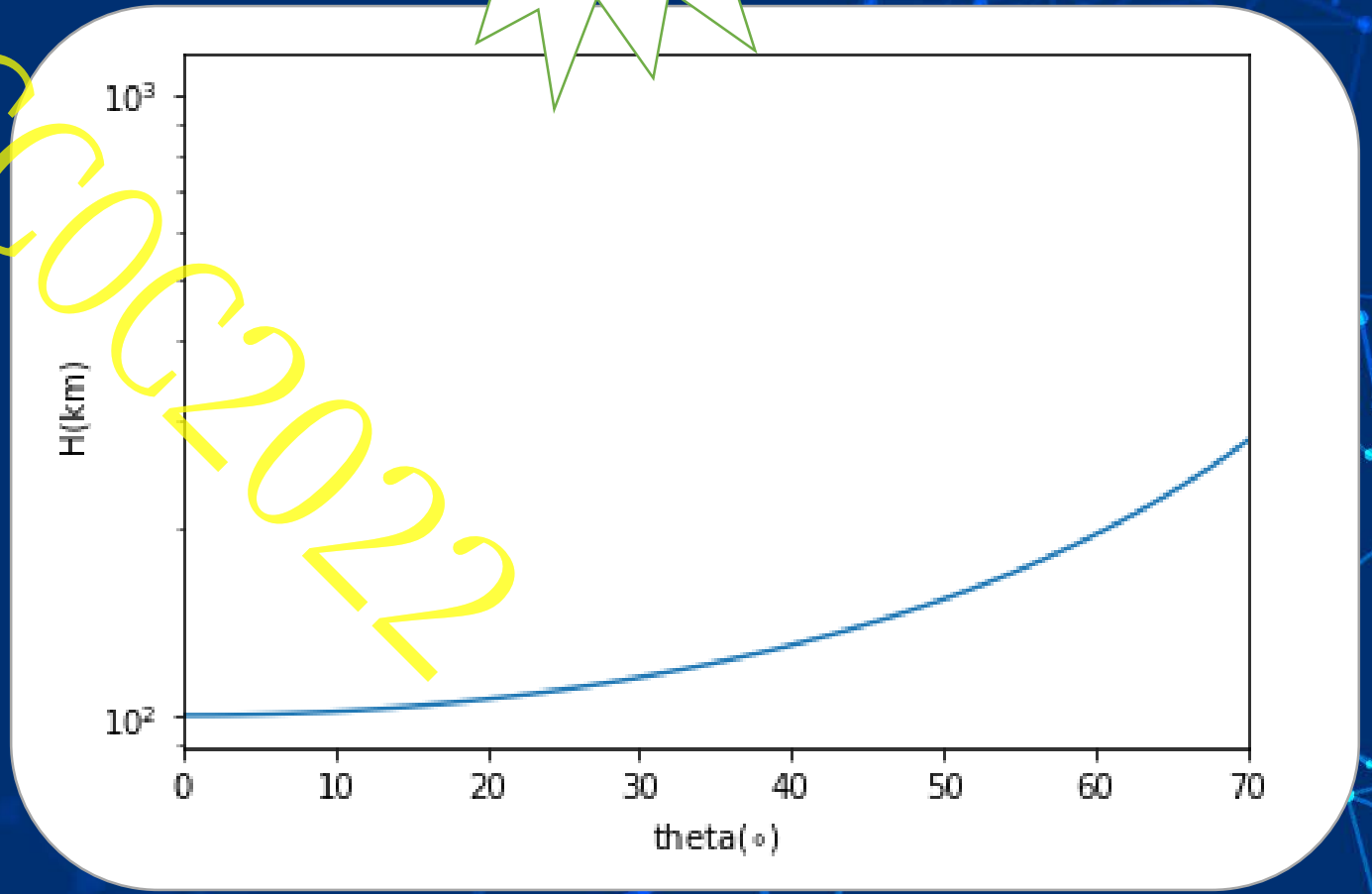
X
Time
59574.35679

约化儒略日时间（mjd）/天

数据处理——数据筛选

- 天顶角小于50°
  - 假设地球大气厚100km
  - 天顶角等于50°的时候是天顶角0°大气厚度的1.5倍
  - 60°为2倍
- 触发探测器个数大于1个
  - 防止偶然事件发生

这张图怎么来的？



CCOC2022

数据处理——数据筛选

- 将弧度制天顶角变成角度制天顶角$\Theta = \theta * 180/3.1415926$
- 将时间变成单位为s，t=（T-59574）*24*3600

T	U
theta	theta/degree
0.3572166	20.46700575

X	Y
Time	time/s
59574.35679	30826.41146
59574.35683	30830.17111

- 按照数据筛选条件筛选数据
  - 先用if语句设置筛选条件，若符合则为1，不符合则为0
    - =IF(AND(G2>1,U2<50),1,0)
  - 利用excel中的筛选，将返回值为0的数据筛选出去，保留返回值为1的数据
    - 总共23520个事例，筛选后剩20563

theta/degree	cut
20.46700575	1
29.10007128	1
67.05472037	0
25.05872392	1
69.07037303	0
19.9358671	1
21.85777939	1

数据处理

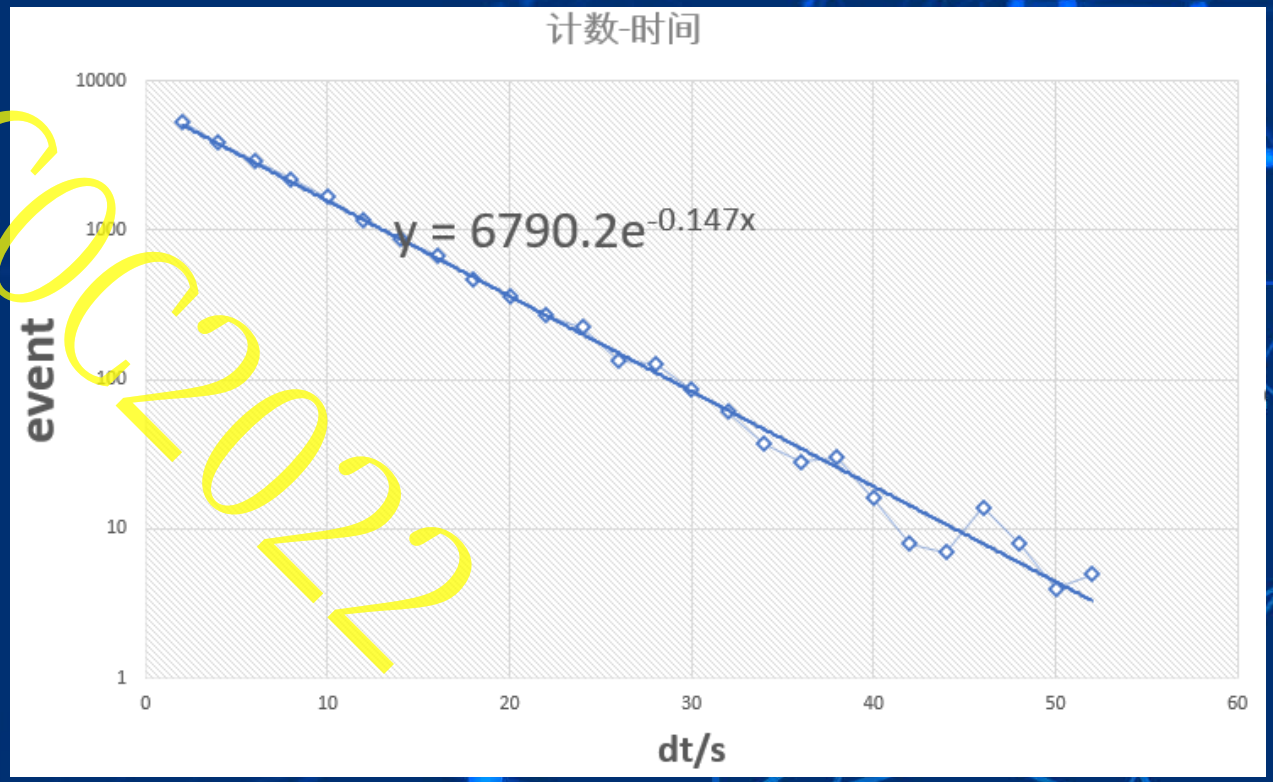
- 计算两个事例之间的时间间隔
  - 计算最大和最小时间间隔
  - 每2s分一个bin
  - 去除事例数为0的bin
- 画直方图
  - 利用frequency函数统计每个bin里的事例数
    - =FREQUENCY(B3:B20565,C2:C34)
    - 选定区域 Ctrl+Shift+Enter
  - 选择区域插入直方图拟合得到结果

	time/s	dt/s	bins	事例数
2	30826.411		2	5338
3	30830.171	3.7596526	4	3829
4	30840.285	10.113742	6	2928
5	30852.602	12.316985	8	2189
6	30862.937	10.335177	10	1676
7	30863.052	0.1151108	12	1152

数据处理

- 利用指数拟合直方图并得到结果
- 与计算结果对比

最大时间间隔	65.5746
最小时间间隔	0
总事例数	20564
总时间	138628
频率	0.14834
拟合频率	0.147



COCO2022

总结与思考

- 总结：
  - 对比事例率计算结果和拟合结果，我们可以得到宇宙线的事例率约为0.147HZ
- 思考：通过以上数据处理过程，我们还能做哪些实验？
  - 不同天顶角宇宙线的频率有多少？有什么差别？
  - 不同方向宇宙线的频率有多少？差别又如何？
  - 将探测器布满地球，宇宙线的频率有多少？

2022

教学

- **【教学目标】**
  - 知识目标：理解数据的结构、分类、筛选的概念和作用；
  - 技能目标：掌握使用excel软件完成数据分析过程
- **【教学重点】**
  - 理解泊松分布的概念，以及指数分布
  - 数据结构与事例率概念的理解，数据筛选条件的理解
  - 熟练使用excel中的if和frequency函数
- **【教学难点】**
  - 理解计数率的概念
  - 理解什么是泊松分布
  - 不同天顶角宇宙线穿过大气深度的计算
  - 对软件的使用，由点及面从小探测器推广到整个地球。

COCO2022

教学

- **【教学方法】**
  - 任务驱动、讲授教学、示范教学、图文结合、实际操作
- **【课前准备】**
  - 多媒体教室，前面课程复习。
  - 马玲玲老师讲述宇宙线的流强的概念
  - 左雄老师宇宙线重建的概念
- **【课前准备】**
  - 复述前面课程讲的数据重建过程，得到事例与时间信息。
  - Excel表格的使用，多媒体教室，将数据下发给学生。

ECOC2022

谢谢

t0 /ns	t1 /ns	t2 /ns	t3 /ns	t4 /ns	Nhit	探测器编号	trg_num	I	m	phi0	phi	theta	Time
96.2	94.0	89.3	92.6	0.0	4	X /m	36582	-0.21714	-0.27408	-0.90079	4.042388	0.357217	59574.35679
97.0	100.6	0.0	0.0	95.8	3	Y /m	36583	0.287761	-0.39207	-0.93764	5.34536	0.507892	59574.35683
0.0	0.0	92.5	91.4	99.0	3		36584	0.090771	-0.91639	-1.47207	4.810934	1.170326	59574.35686
96.0	0.0	0.0	90.6	94.8	3		36585	0.238586	-0.34996	-0.97242	5.310584	0.437357	59574.35695
95.0	102.9	0.0	109.8	99.6	4		36586	0.442635	0.822476	1.077078	1.077078	1.205505	59574.35697
106.2	0.0	108.6	104.5	0.0	3		36588	0.322413	-0.11095	-0.33142	5.951579	0.347947	59574.35709
0.0	0.0	97.8	100.5	101.4	3		36590	-0.22075	-0.2998	-0.9361	4.077699	0.38149	59574.35721
93.4	89.2	0.0	94.1	92.3	4		36591	-0.314	0.060306	0.189749	2.951851	0.325449	59574.35721
94.4	0.0	0.0	94.6	95.9	3		36594	0.230616	0.011652	0.050483	0.050483	0.233013	59574.35738
0.0	85.8	91.3	97.8	0.0	3		36597	-0.51273	0.363875	0.617194	2.524406	0.679916	59574.35764
124.4	131.0	128.5	122.8	125.5	5		36598	0.490564	-0.13462	-0.26782	6.015176	0.533674	59574.35766
85.5	94.1	0.0	0.0	88.6	3		36599	0.681903	0.16196	-0.23318	6.049816	0.776618	59574.35767
117.5	113.6	0.0	107.9	111.0	4		36602	-0.30066	-0.62504	-1.12245	4.264047	0.766471	59574.35797
91.4	79.7	0.0	83.7	0.0	3		36603	-0.92442	-0.50827	0	3.1416	0	59574.35806
106.2	120.7	103.3	104.6	106.2	5		36604	0.521249	-0.62634	-0.87672	5.406281	0.952494	59574.3581
96.2	96.0	91.0	0.0	0.0	3		36605	-0.01211	-0.32709	-1.53379	4.675393	0.33346	59574.3582
0.0	0.0	94.3	88.2	92.6	3		36606	0.479473	-0.17636	-0.35246	5.930541	0.536206	59574.35821
93.1	96.7	0.0	0.0	95.4	3		36607	0.282732	0.062438	0.21735	0.21735	0.293751	59574.35847
0.0	0.0	102.8	112.1	110.9	3		36608	-0.73258	-0.456	-0.55678	3.698375	1.040996	59574.35847
118.5	110.2	0.0	0.0	115.5	3		36609	-0.6515	0.150209	0.2266	2.915	0.732312	59574.35855
0.0	103.0	102.3	94.9	0.0	3		36610	0.583253	-0.04763	-0.08148	6.201515	0.625121	59574.35859
104.5	0.0	106.3	112.7	0.0	3		36611	-0.50599	0.539309	0.817262	2.324338	0.832349	59574.35863
90.5	93.0	0.0	0.0	97.4	3		36612	0.201144	0.748282	1.308197	1.308197	0.886469	59574.35864
91.1	87.9	84.5	0.0	0.0	3		36613	-0.25617	-0.21812	-0.70534	3.846945	0.343153	59574.35873
8.2	0.0	0.0	95.2	5.9	3		36614	-7.22036	5.710923	0	3.1416	0	59574.35881
97.4	94.4	98.8	97.8	92.8	5		36615	-0.07987	0.157399	1.101237	2.040363	0.177432	59574.35885
102.1	94.7	97.3	105.7	118.8	5		36616	-0.62598	0.205284	0.316887	2.824713	0.719202	59574.35888
97.8	0.0	0.0	90.4	94.9	3		36619	0.126507	-0.48654	-1.31642	4.966585	0.526741	59574.35929
0.0	0.0	108.5	91.4	87.9	3		36620	1.351336	1.582227	0	0	0	59574.35954
93.5	88.4	0.0	0.0	93.9	3		36621	-0.39876	0.387464	0.771036	2.370564	0.589566	59574.35962
103.9	0.0	0.0	94.7	94.4	3		36622	-0.76972	-0.60621	-0.66711	3.808712	1.369329	59574.35963