

偏度和峰度在粒子谱研究中的应用

刘延刚

(气象科学研究院)

提 要

本文通过引进两个统计学参数:偏度 S 和峰度 K ,提出了〈1〉实际谱是否可用 Γ -谱近似的定量化验证方法;〈2〉在实际谱可用 Γ -谱近似的条件下,一个定量分析实际谱偏离指数谱 $N(D) = N_0 \exp(-\lambda D)$ 的指标——偏离系数;〈3〉依据偏离系数的值,对粒子谱型的一种分类方法。

一、引言

降水粒子谱的研究,对遥感技术、光及无线电波传输、云降水化学、雷达定量测量云和降水、云模式的设计,特别是对自然界中的成云致雨机制的研究,都具有一定的意义。不同的谱型可能对应不同的物理机制。Marshall和Palmer[1](1948)给出了一个平均的雨滴谱形式: $N(D) = N_0 \exp(-\lambda D)$,其中 $N_0 = 0.08 \text{ cm}^{-4}$ 为常数, $\lambda = 41 R^{-0.21} \text{ cm}^{-1}$ (其中 R 是降水强度,单位是 $\text{mm} \cdot \text{hr}^{-1}$)。后来许多学者研究发现 N_0 不是常数,是一个与降水强度有关的量。如Gun和Marshall[2](1958)通过分析融滴谱得到 $N_0 = 3.8 \times 10^3 R^{-0.87}$, $\lambda = 25.5 R^{-0.48}$ 。随着工作的进一步深入,人们发现以下几点:

1. 谱参数 N_0 和 λ 不仅是降水类型的函数,而且在同一次降水过程中, N_0 和 λ 也是随时间变化的[3]。

2. 指数分布形式往往是在粒子尺度 D 大于某一临界值 D^* 时才成立[4]。

3. 降水粒子谱时常偏离指数分布,用

Γ -函数 $N(D) = \alpha_0 D^{\alpha_1} \exp(-\alpha_2 D)$ 其中 α_0 、 α_1 、 α_2 为谱参数, D 为粒子直径, N 表示浓度)近似实际谱更具有普遍性和适用性[5]。

但迄今为止,在实际应用和理论研究中,多是基于粒子谱服从指数分布形式的假设。这与实际情况是有偏差的;有些讨论实际谱偏离指数谱的工作,也一般是直观的定性的,没有一个定量的指标;在运用 Γ -谱代替指数谱来近似实际谱的工作中,主要也是通过直观观察法,在数学上很少有验证。

另一方面,云和降水粒子谱的形成过程是一典型的随机过程,谱分布具有典型的统计性质。选取具有统计学意义的参数来研究谱,有利于进一步深入研究谱的演变和成因。

二、偏度和峰度的引入

偏度 S (Skewness)和峰度 K (Kurtosis)是两个典型的统计量,在一般的数理统计中,是用来研究分布的正态性的。其定义如下:设 $f(r)$ 表示分布的密度函数, $\bar{r}$ 表示数学期望,则:

$$S = \frac{\int (r - \bar{r})^3 f(r) dr}{\left[\int (r - \bar{r})^2 f(r) dr \right]^{3/2}} \quad (1)$$

$$K = \frac{\int (r - \bar{r})^4 f(r) dr}{\left[\int (r - \bar{r})^2 f(r) dr \right]^2} - 3 \quad (2)$$

对于正态分布 $S = K = 0$; 对于其它情况, 如图 1 所示, $S > 0$ 表示正偏分布, $S < 0$ 表示负偏分布; K 的大小则反映了分布在峰值附近的集中程度。

引用到谱研究中, S 和 K 的表达式如下:

$$S = \frac{\int (D - \bar{D})^3 \frac{N(D)}{N_T} dD}{\left[\int (D - \bar{D})^2 \frac{N(D)}{N_T} dD \right]^{3/2}} \quad (3)$$

$$K = \frac{\int (D - \bar{D})^4 \frac{N(D)}{N_T} dD}{\left[\int (D - \bar{D})^2 \frac{N(D)}{N_T} dD \right]^2} - 3 \quad (4)$$

式中 $N_T = \int N(D) dD$ 代表总浓度。

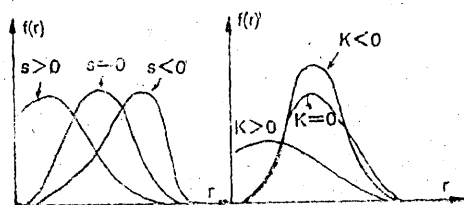


图 1 S、K 示意图

三、指数分布的偏度和峰度

在现有的粒子谱研究和应用中, 一般是采用指数谱来近似实际谱的, 为此首先推导出指数分布的偏度和峰度。

$$\mu = \int D^m N(D) dD \quad m = 0, 1, 2, \dots \quad (5)$$

$$\mu^m = \int D^m \frac{N(D)}{N_T} dD \quad m = 1, 2, 3, \dots \quad (6)$$

分别表示谱 $N(D)$ 的 m 次矩量和频率谱 $\frac{N(D)}{N_T}$ 的 m 次矩量。根据 S 和 K 的定义, 可以推出下面的表达式:

$$S = \frac{\mu_3 - 3\mu_1 \cdot \mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}} \quad (7)$$

$$K = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3 \quad (8)$$

对于指数分布, $N(D) = N_0 \exp(-\lambda D)$, 所以

$$\begin{aligned} \mu_m &= \frac{\int D^m N_0 \exp(-\lambda D) dD}{\int N_0 \exp(-\lambda D) dD} \\ &= \frac{m!}{\lambda^m} \quad m = 1, 2, 3, \dots \end{aligned} \quad (9)$$

所以有:

$$\mu_1 = \frac{1}{\lambda}, \quad \mu_2 = \frac{2}{\lambda^2}, \quad \mu_3 = \frac{6}{\lambda^3}, \quad \mu_4 = \frac{24}{\lambda^4}$$

将其代入 (5) 式和 (6) 式得:

$$S_{EX} = \frac{\frac{6}{\lambda^3} - 3 \cdot \frac{1}{\lambda} \cdot \frac{2}{\lambda^2} + 2 \frac{1}{\lambda^3}}{\left(\frac{2}{\lambda^2} - \frac{1}{\lambda^2} \right)^{3/2}} = 2$$

$$K_{EX} = \frac{\frac{24}{\lambda^4} - 4 \cdot \frac{1}{\lambda} \cdot \frac{6}{\lambda^3} + 6 \left(\frac{1}{\lambda} \right)^2 \cdot \frac{2}{\lambda^2} - 3 \frac{1}{\lambda^4}}{\left(\frac{2}{\lambda^2} - \frac{1}{\lambda^2} \right)^2} - 3$$

$$= 6$$

(其中, S_{EX} , K_{EX} 分别表示指数分布的偏度和峰度), 由此可见, 指数分布的偏度和峰度为常数。

四、 Γ -分布的偏度和峰度

许多观测表明, 用 Γ -分布 $N(D) = \alpha_0 D^{\alpha_1-1} \exp(-\alpha_2 D)$ 来近似粒子谱更切合实际, 更具普遍性(指数谱是 Γ -谱在 $\alpha_1 = 0$ 时的特例), 那么 Γ -分布的情况如何? 将式 $N(D) = \alpha_0 D^{\alpha_1-1} \exp(-\alpha_2 D)$ 代入 (6) 式中可得到:

$$\begin{aligned} \mu_m &= \frac{\int D^m \alpha_0 \cdot D^{\alpha_1-1} \cdot \exp(-\alpha_2 D) dD}{\int \alpha_0 D^{\alpha_1-1} \cdot \exp(-\alpha_2 D) dD} \\ &= \frac{\Gamma(m + \alpha_1 + 1)}{\Gamma(\alpha_1 + 1)} \cdot \frac{1}{\alpha_2^m} \end{aligned} \quad (10)$$

($\Gamma(\cdot)$ 为 Γ -函数)

所以有:

$$\begin{cases} \mu_1 = \frac{\Gamma(\alpha_1 + 2)}{\Gamma(\alpha_1 + 1)} \cdot \frac{1}{\alpha_2} \\ \mu_2 = \frac{\Gamma(\alpha_1 + 3)}{\Gamma(\alpha_1 + 1)} \cdot \frac{1}{\alpha_2^2} \\ \mu_3 = \frac{\Gamma(\alpha_1 + 4)}{\Gamma(\alpha_1 + 1)} \cdot \frac{1}{\alpha_2^3} \\ \mu_4 = \frac{\Gamma(\alpha_1 + 5)}{\Gamma(\alpha_1 + 1)} \cdot \frac{1}{\alpha_2^4} \end{cases}$$

将以上各式代入 (7) 式得:

$$S_{GA} = \frac{2}{\sqrt{1 + \alpha_1}} \quad (11)$$

$$K_{GA} = \frac{6}{1 + \alpha_1} \quad (12)$$

其中, S_{GA} 和 K_{GA} 分别表示 Γ -分布的偏度和峰度。显然, Γ -分布的偏度和峰度仅与 α_1 有关, 而与 α_2 无关。

五、两个判断方法的确立

从以上分析可以得到以下几点: (1) 指数分布的偏度和峰度俱为常数: $S_{EX} = 2$, $K_{EX} = 6$; (2) Γ -分布的偏度和峰度只与 α_1 有关: $S_{GA} = \frac{2}{\sqrt{1 + \alpha_1}}$, $K_{GA} = \frac{6}{1 + \alpha_1}$ 。从

表达式 $N(D) = \alpha_0 D^{\alpha_1} \exp(-\alpha_2 D)$ 中可以看出, 如果实际谱可用 Γ -谱来近似, 那么 α_1 的大小恰好说明了实际谱偏离指数谱的程度。另一方面, 在用 Γ -谱来近似实际谱的研究论文中, 一般仅是用直观观测方法, 没有一个数学验证方法, 从以上分析可以看出, 对 Γ -分布而言, 其偏度和峰度存在关系式: $K_{GA} = \frac{3}{2} S_{GA}^2$ 。由此可将实测谱的特

征通过计算 S 和 K 来反映。为更清楚起见, 引进新参数:

$$C_1 = \frac{S_{GA}}{S_{EX}} = \frac{1}{\sqrt{1 + \alpha_1}} \quad (13)$$

$$C_2 = \frac{K_{GA}}{K_{EX}} = \frac{1}{1 + \alpha_1} \quad (14)$$

据此, 我们可以得到以下两点:

(1) 对 Γ -谱而言, 关系式 $C_2 = C_1^2$ 成立, 因此可通过计算实际谱的 C_1 和 C_2 来验证实际谱是否可用 Γ -谱近似。

(2) 在实际谱可用 Γ -谱来近似的条件下, 参数 C_1 和 C_2 的大小可用来作为实际谱偏离指数程度的指标, 将其定义为偏离系数 1 和偏离系数 2。依据 C_1 或 C_2 , 可将 Γ -分布大致分为以下三类, 如图 2 所示。

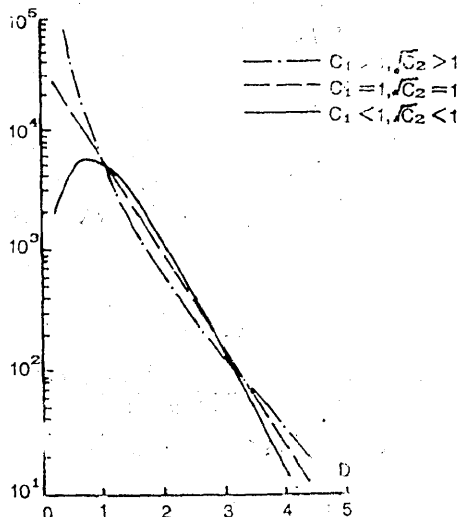


图 2 分类示意图

- A. 超指数分布 $N(D) = \alpha_0 D^{\alpha_1} \exp(-\alpha_2 D)$ $C_1 > 1, C_2 > 1$, 对应 $\alpha_1 < 0$
- B. 指数分布 $N(D) = \alpha_0 \exp(-\alpha_2 D)$ $C_1 = 1, C_2 = 1$, 对应 $\alpha_1 = 0$
- C. 次指数分布 $N(D) = \alpha_0 D^{\alpha_1} \exp(-\alpha_2 D)$ $C_1 < 1, C_2 < 1$, 对应 $\alpha_1 > 0$

故此, 可以通过计算 C_1 和 C_2 , 来确定粒子谱的基本情况; 据 C_1 或 C_2 的大小可判断实际谱偏离指数谱的程度, C_1 和 C_2 的值与 1 差别越大, 实际谱偏离指数谱越厉害。

六、具体应用

1. 一般方法

在实际观测中, 常常是给出某一直径 D 对应的粒子浓度 N , 在降水粒子谱的分析研究中, 一般是直接点绘出谱分布图, 定性的观察分析, 在粒子谱资料很多时, 这种

直观分析很不方便甚至不太可能,并且对于谱型只能停留在定性上。从以上的理论分析中,我们可以看出,可首先利用诸观测值 $\{D_i, N_i\}$ 计算偏度 S 和峰度 K , 然后根据 C_1 和 C_2 的值, 来分析研究粒子谱。具体步骤为:

(1) 据实测值 $\{D_i, N_i\}$ 计算频谱矩量 μ_m

$$\mu_m = \frac{\sum_{i=1}^n D_i^m N_i}{\sum_{i=1}^n N_i} \quad (m=1, 2, 3, \dots, n \text{ 为样本数})$$

(2) 将 $\mu_m (m=1, 2, 3, 4)$ 代入下式计算 S 和 K

$$S = \frac{\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3}{(\mu_2 - \mu_1^2)^{3/2}}$$

$$K = \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{(\mu_2 - \mu_1^2)^2} - 3$$

(3) 计算偏离系数 C_1 和 C_2

$$C_1 = \frac{S}{2}$$

$$C_2 = \frac{K}{6}$$

(4) 据 $C_2 = C_1^2$ 验证实际谱是否可用 Γ -谱近似。

(5) 判断粒子谱型 (以 C_1 为例)

$C_1 > 1$, 超指数谱

$C_1 = 1$, 指数谱

$C_1 < 1$, 次指数谱

并且 C_1 偏离 1 越大, 实际谱偏离指数谱的程度越大。

2. 个例分析

为了具体说明该方法的应用, 深入分析粒子谱的演变特性, 本文计算了 1983 年 5 月 24 日、5 月 25 日、5 月 26 日及 6 月 1 日发生在西安地区的 4 次稳定性降水的 C_1 和 C_2 , 依据以上提出的方法作了如下分析。

(1) 点聚图分析

在实际的降水过程中, 尽管有人发现用 Γ -谱来近似实际谱较用指数谱近似要合理, 但多是依据对粒子谱图的直观观察, 没有准确的数学依据。从以上分析中, 我们得到: 关系式 $C_2 = C_1^2$ 对 Γ -分布成立。因此, 除了直观观察外, 可以通过分析 C_1 和 C_2 的关系, 来验证 Γ -谱近似假设的合理性。本文分别将上述 4 次稳定性降水的粒子谱作了计算, 并分别作了点聚图分析, 样品总数为 266。为了清楚起见, 本文采取了分析 $\sqrt{C_2}$ 和 C_1 的点聚图, 对 Γ -分布而言, 应用 $\sqrt{C_2} = C_1$ 。如图 3 所示。

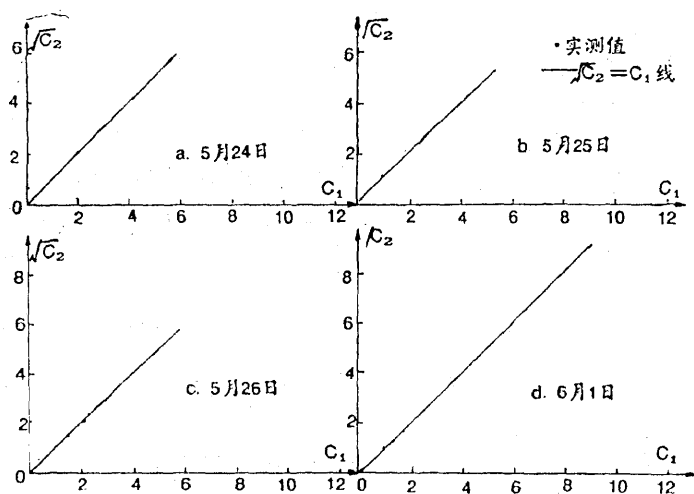


图 3 C_1 、 $\sqrt{C_2}$ 点聚图

从图中可以看出,对4次降水而言,粒子谱所对应 C_1 和 C_2 ,基本上是满足关系式 $C_2 = C_1^2$ ($\sqrt{C_2} = C_1$)。这一方面验证了对这4次降水过程而言, Γ -谱近似假设的合理性;另一方面,从图中可以看出落在点 $\sqrt{C_2} = C_1 = 1$ 的样品几乎没有,这说明实际的粒子谱很少满足指数谱近似,进而说明从形态上讲常规的指数谱假设不太合理。还有一点需要指出的是:从图中可以看出,对4次不同的降水过程而言, $\sqrt{C_2}$ 和 C_1 的点聚图极为相似,基本上可用 Γ -谱线 $\sqrt{C_2} = C_1$ 近似,这是否可作进一步的推论: Γ -谱基本上可作为实际粒子谱的近似,尚需进一步论证。对于点聚图与线 $\sqrt{C_2} = C_1$ 的偏离,可能与截断谱有关,因为式 $C_2 = C_1^2$ 是在 D 从0到 ∞ 的条件下推导的,而实测谱均有上界和下界。另外,从图4可以看出, $\sqrt{C_2}$ 和 C_1 的时间变化基本上是同步的,这也是 Γ -谱假设合理性的一种验证。

(2) 粒子谱的时间演变特征

通过(1)的分析,我们可以看出,对

这4次降水过程而言,粒子谱基本上可用 Γ -谱近似。那么,在每次降水过程中,粒子谱型是怎样变化的?粒子谱偏离指数谱的情况如何?是否有和前人的工作有不同的发现?为了说明这些问题,将4次降水过程的 C_1 和 $\sqrt{C_2}$ 的时间变化线绘于图4。从图中可以看出:(a)对这4次降水过程而言,虽然 C_1 和 $\sqrt{C_2}$ 随时间的变化特征有差异,但仍有其共同之处: $\sqrt{C_2}$ 和 C_1 的变化是同步的;且它们都在1的附近振荡。这说明在这4次稳定性降水中,粒子谱是绕指数谱振荡的,指数谱可能是降水过程的平衡态谱;(b)在振荡过程中, C_1 和 $\sqrt{C_2}$ 的振幅有变化,有时存在突变点,这些突变点可能与降水类型的转换有关;(c)4次降水过程持续时间都在10小时左右,但 C_1 和 $\sqrt{C_2}$ 随时间的变化情况有所不同,这很可能与降水过程有关;(d)对4次降水而言 C_1 的总平均值为0.99, $\sqrt{C_2}$ 的总平均值为1.47,与指数分布的1很接近,这说明平均谱很接近指数谱。回顾前人的工作,本文所得到的

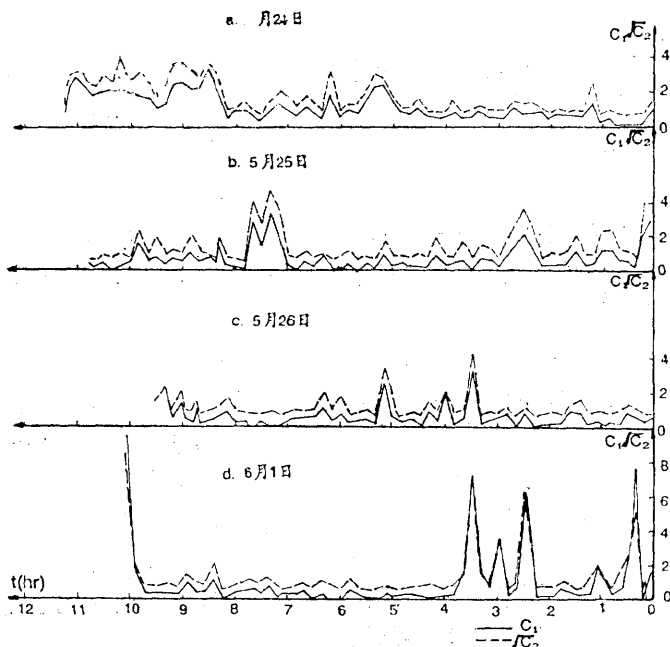


图4 C_1 、 $\sqrt{C_2}$ 时间变化图

结果与之非常接近; Marshall和Palmer是在平均了大量的雨滴谱资料, 得出著名的M-P分布的; Waldevogal假设在降水过程中, 粒子谱服从指数分布 $N(D) = N_0 \exp(-\lambda D)$, 进而发现了 N_0 的突变现象。对 Γ -谱讲, 可能反映在参数 C_1 和 C_2 或 α_1 上; Brown[6] (1989)通过数值计算发现粒子谱在演变过程中, 趋近于平衡态谱, 而且这种趋近不是单调的, 而是振荡性的; 同时又指出, 在自然条件下, 很少有足够的时间使实际谱演变到平衡态谱。但是以上工作都是基于指数分布的。通过本文的分析, 进一步发现: (a) 瞬时的粒子谱一般可用 Γ -谱近似, 在演变过程可历经超指数谱或次指数谱; (b) 从形态学来讲, 很难达到指数谱, 指数谱可能是过程的平衡态谱; (c) 粒子谱趋近平衡态的方式不是单调的, 在其演变过程中, 具有振荡特性; (d) 粒子谱型的振荡振幅各有差异, 有时可能出现突变现象, 这很可能与降水类型有关。

七、结论和讨论

本文通过理论分析和推导, 引进了两个统计学参数: 偏度和峰度, 利用这两个参数, 得到一个验证粒子谱服从 Γ -分布的方法; 提出了一个在 Γ -谱假设条件下, 判断实际谱偏离指数谱程度的定量指标——偏离系数 C_1 或 C_2 ; 找到了一个依据该指标分类粒子谱的方法, 将谱大致分为超指数分布(SE)、指数分布(E)、和次指数分布(UE)三大类。利用所提出的方法, 分析了4次稳定性降水的粒子谱演变情况, 结果发现: (1) 实际谱基本上可用 Γ -谱近似, 用指数谱来近似瞬时粒子谱不太合理; (2) 在降水过程中, 实际谱具有振荡特征可历经超指数谱或次指数谱, 很难达到指数谱, 指数谱可能是过程的平衡态谱; (3) 实际谱在演变过程中, 有时存在突变现象。

需要指出的是, 本文虽然提出了一个新的方法, 并利用所提出的方法, 通过对4次降水个例的分析有些新的发现, 但仍然是初步的, 确切地讲所得到的结论仅对4次个例而言, 或者说是一种假设, 是否可作一般性的推广, 尚需进一步验证。

另外, 如何利用对S和K的分析, 来研究云和降水过程的物理性质, 将两者有机的联系起来? 如何考虑实际谱的截断效应? 如何将该方法推广应用到诸如雷达气象、数值模拟等与粒子谱有关的领域, 还需深入探讨。

致 谢

本文是在游来光副研究员的悉心指导下完成的, 郭恩铭副研究员等也曾提过宝贵意见。

参 考 文 献

- (1) J.S. Marshall & W.M. palmer, The Distribution of Raindrops with size, J. Met. 5, 165—166, 1948.
- (2) Gun, K.L.S & J. S. Marshall, The Distribution With size of Aggregates Snowflakes J. Met, 15, 452—461, 1958.
- (3) A. Waldvogal, The No Jump of Raindrop spectra, J.A.S.31, 1067—1078, 1974.
- (4) Houze, R.A. Jr & P.V. Hobbs, Size Distributions of Precipitation particles in Frontal Clouds, J.A.S 36, 156—162, 1979.
- (5) P.T. Willis, Functional Fit to Some observed Drop Size Distributions and Parameterization of Rain, J.A.S. 41, 1648—1661, 1984.
- (6) P.S. Brown, JR., Coalescence and Breakup-Induced Oscillations in The Evolution of The Raindrop Size Distribution, J. A.S.46, 1186—1192, 1989.

The application of skewness and kurtosis to the studies on particle distribution

Liu Yangang

(Academy of Meteorological Science)

Abstract

Two statistical parameters, skewness and kurtosis are introduced to investigate particle distribution. A quantitative method is formulated to testify the fitting goodness to Γ -spectra. A quantitative criterion named 'Deviation Coefficient' is presented to express the natural deviation from exponential spectra under the Γ -spectra assumption and according to the criterion a classification is developed.