

線形多変量解析を用いた おうし座分子雲-1 における星間分子スペクトル線サーベイ データのシグナル/ノイズ分類 (第2報)

神部順子^a、大石雅寿^b、長嶋雲兵^{c*}

^a 大東文化大学外国語学部, 〒175-8571 板橋区高島平 1-9-1、

^b 国立天文台、〒三鷹市大沢 2-21-1

^c 産業技術総合研究所グリッド研究センター, 〒110-0015 台東区上野 6-9-3 住友不動産上野ビル
8号館 7階

E-mail: u.nagashima@aist.go.jp

1. はじめに：本研究では、天文データマイニングの一步として、国立天文台野辺山にある電波望遠鏡で取得した 8-50GHz のスペクトル線サーベイデータを従来の線形多変量解析技法（主成分分析、クラスター分析、判別分析）を用いて自動分類することを試みた。本研究の最終目的は、発達の著しい計算機・情報処理技術を活用して、研究者を機械的作業から開放し、「思考過程」に専念できるよう補助する機能を実現するものである。本報告では 2002 年秋季大会で発表したデータ（発表番号 313）より属性値をさらに厳選し再計算したものを報告する。

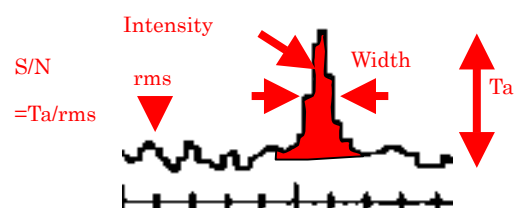


Figure 1. Peak Description Parameters

2. データと解析プログラム：属性値（周波数、アンテナ温度、半値幅、強度(Figure 1)) によって特徴づけられたピーク値の総数は 677 であり、これらはすでにノイズと信号の分類が行われている。主成分分析、判別分析には、Excel 統計 2000[1]を用いた。クラスター分析には、Excel 太閤[2]を用いた。

3. 主成分分析の結果：全データを用いた主成分分析を行った。第一主成分の寄与率は 56.6%、第二主成分の寄与率は 22.4%である。第三主成分の寄与率は、17.9%、第四主成分の寄与率は、2.2%であった。第一主成分と第二主成分を主軸とする、677 個の主成分得点を用いた K-L プロットを Figure 2 に示した。シグナル◇とノイズ○は、K-L プロットでは重なりが大きく明確な分類は不能である。シグナルは第一主成分および第二主成分の 0 付近に集中しているが、第一主成分に沿って -5.0~15.0 の範囲でまんべんなく分布しており、逆にノイズは第一主成分 0 付近に分布している。ノイズは第二主成分の影響

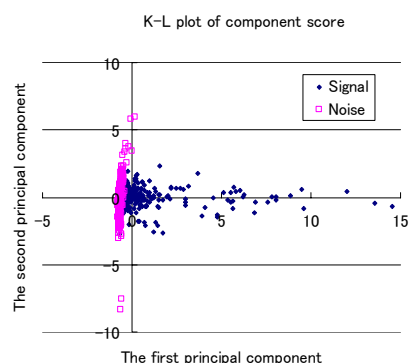


Figure 2 K-L plot of data on the 1st and 2nd principal components

[1] 多変量解析プログラム Excel 統計、http://www.esumi.co.jp/products_info/toukei/toukei_1.html.

[2] アンケート分析プログラム Excel 太閤、http://www.esumi.co.jp/products_info/taiko/profile_1.html.