

ひさき衛星が取得した惑星大気・プラズマの観測データに対する Lasso を用いた時系列解析

鈴木 文晴 [1]; 岡田 真人 [1]; 吉岡 和夫 [2]; 益永 圭 [3]; 吉川 一朗 [4]
[1] 東大・新・複雑; [2] 東大・理; [3] 東大・理
; [4] 東大・理・地惑

Time Series Analysis Using Lasso of Atmospheric and Magnetospheric Data Obtained by HISAKI

Fumiharu Suzuki[1]; Masato Okada[1]; Kazuo Yoshioka[2]; Kei Masunaga[3]; Ichiro Yoshikawa[4]
[1] Complexity Science, Univ. of Tokyo; [2] The Univ. of Tokyo; [3] Univ. Tokyo; [4] EPS, Univ. of Tokyo

<http://www.astrobio.k.u-tokyo.ac.jp/yoshikawa/>

Thanks to recent progress in technology related to observation instruments, we can obtain various data on planetary science. In such a privileged situation, it should be pointed out that we should use analysis methods suitable for dealing with a large number of data.

For example, in the periodic analysis, the intensity of many frequency components can be obtained by prolonging the observation period and improving the temporal resolution. However, it is inappropriate to treat all of them as meaningful signals in terms of physical interpretation. Furthermore, when fitting by using the least squares method such as Fourier transform to derive frequency components, it is impossible to prevent 'overfitting' to noise-containing data. So frequency components are erroneously estimated.

To solve the above problem, 'sparse modeling' used in the analysis of big data becomes a breakthrough. This method can extract only important variables after handling a large number of data. It is very effective against the above problem.

In this study, we show the results of analyzing by applying 'Lasso' (Least Absolute Selection and Shrinkage Operator) [Tibshirani et al., 1996] to the optical observation data of HISAKI. Lasso is a mathematical method which is modeling in which a 'penalty term' is added to the least squares method. Lasso has the advantage that it can extract important variables and prevent overfitting. Lasso is widely used in information science and natural sciences such as medicine and astronomy, but it is not used in data analysis of planetary science. In this presentation, we introduce that Lasso gives better interpretive results than conventional methods.

近年、国内外の観測機器によって惑星科学に関する様々なデータが蓄積されている。このような恵まれた状況では、多数のデータの取り扱いに最適な解析方法の必要性が議論されるべきである。

例えば周期解析では、観測期間や時間分解能の向上により多数の周期成分が得られるようになるが、その全てを意味のある信号として取り扱うことは物理的解釈の上で不適切である。さらに、従来手法である離散フーリエ変換のように最小二乗法を用いたフィッティングによる周期成分の導出では、誤差を含むデータへの「過剰適合」を防げない。そのため結果として得られる偽の周期成分を排除できない危険性がある。

上記の問題に対しては、情報科学分野でビッグデータの解析に用いられるスパースモデリングが突破口となる。これは多数のデータを同等に扱った上で重要な変数のみを抽出する手法であり、上記のような惑星科学衛星のデータ解析における問題に対して非常に効果的であると考えられる。

本研究では、ひさき衛星の光学観測データに対しスパースモデリングの一種である“Lasso” (Least Absolute Selection and Shrinkage Operator) [Tibshirani et al., 1996] を適用して解析した結果を示す。この数理手法は、フィッティング手法として一般的に用いられる最小二乗法に特定の「罰則項」を加えたモデリングであり、過剰適合の防止や重要な変数の抽出が可能であるという利点を持つ。Lasso は情報科学だけでなく医学や天文学などの自然科学分野でも広く用いられているが、惑星観測衛星のデータ解析に適用された例はない。Lasso を用いることで、従来手法に比較して解釈性に優れ、物理現象の説明に役立つ結果が得られた。