A data-driven model of nucleosynthesis with chemical tagging in a lower-dimensional space

GAゼミ@20191101 松野

Casey, A. et al. (2019) ApJ accepted, arxiv: 1910.09811

Abstract 化学組成のスペースで次元削減を行うとともに次元削減後の空間でのclustering 解析を行う。GALAH で17の元素すべてに対し組成が測定された、2566個の星に対しこの手法を適用したところ、6次元空間への次元削減が好まれた。サンプルはclustering 解析から3つのグループへと分けられたが、いずれのグループも位置・運動空間上で構造を持たなかった。

Goal chemical tagging / 化学組成空間にいくつ独立な情報（元素合成サイト）があるかを明らかにする。

Originality すべての元素組成が独立であるという仮定を置かずに化学組成空間でclustering 解析をかけた。

Logic

これまでの化学組成空間でのclustering 解析はすべての元素組成が独立な情報を持つとしてきた。
しかし例えば Mg やCa はα元素とまとめられるように、元素合成サイトの数は限られているのでこれは必ずしも正しくない。
次元削減を行いつつ低次元空間でclustering 解析を行うことでこの問題を解決する。

$$\mathbf{X} = \boldsymbol{\mu} + \mathbf{LS} + \mathrm{e}$$ （Fig.1）

X: abundance data (*D* elements x *N* stars)
L: latent factor spaceと abundance space の変換 (*D* elements x *J* latent factors)
S: latent factor space での各データ点の座標 (*J* latent factors x *N* stars)
*J*次元, *K* components のGaussian mixture
e: 観測ノイズ

以下の式を最大化して各パラメータを求める。(Y = X-μ)

$$f(\mathbf{Y};\boldsymbol{\Psi}) = \sum_{k=1}^{K} \pi_k \phi(\mathbf{Y}; \mathbf{L}\boldsymbol{\xi}_k, \mathbf{L}\boldsymbol{\Omega}_k\mathbf{L}^{\intercal} + \mathrm{diag}(\mathbf{D}))$$

MML を使って最適な*K* と*J* を見つける。（Figure 5）
Latent factor space で混合ガウスモデル を適用しているので、各データ点がどのcluster に属するかという情報も同時に得られる。

GALAH のデータで2566星に対する17の元素の組成に適用したものがFigure 6 & 7。

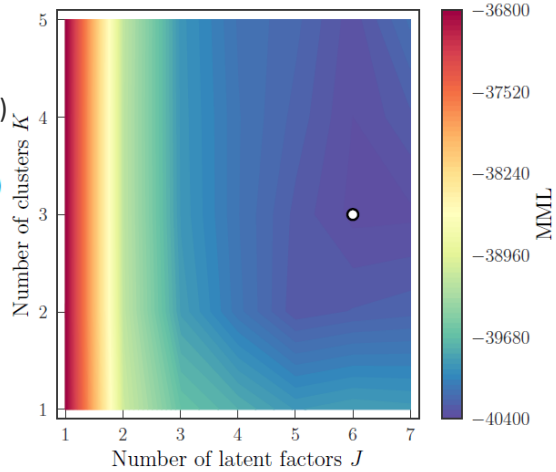


Fig. 5 MMLによる最適な*K*と*J*の探索。

Comment 著者もコメントしているように各Latent factor が直交するという条件や混合ガウスモデルを採用しているという点があるので突っ込んだ解釈は危険だが、各元素の組成が独立でないという問題に対する解決策の第一歩。

スペクトルからLatent factor を直接推定できるようにトレーニングすると、低分散分光観測でも個々の元素の組成ベースで議論するより精度が出そう。

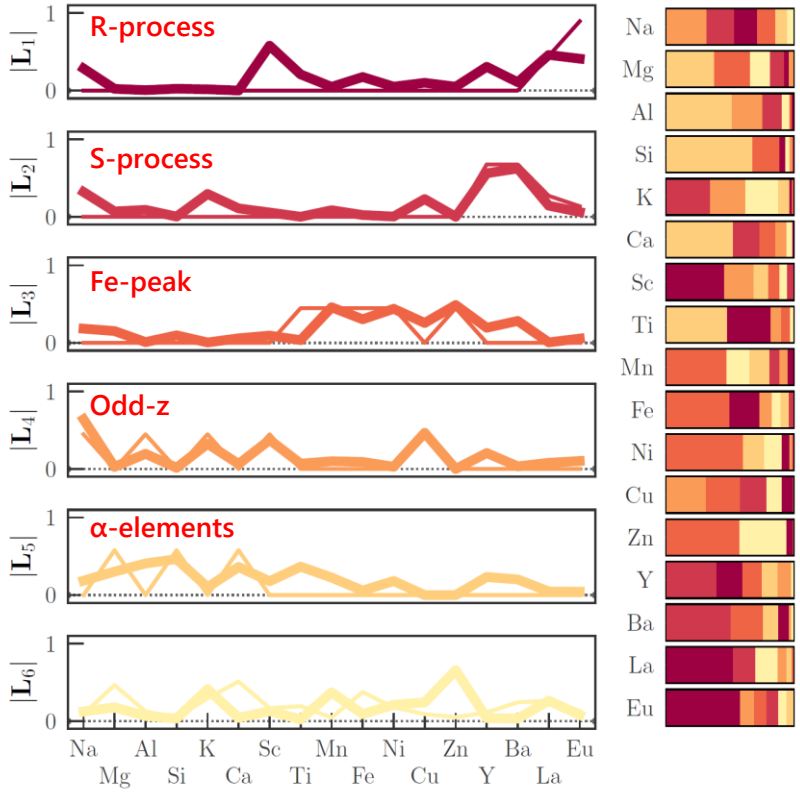


Fig. 6 Latent factor。数値が大きいほどそのfactorにその元素が寄与していることを示す。

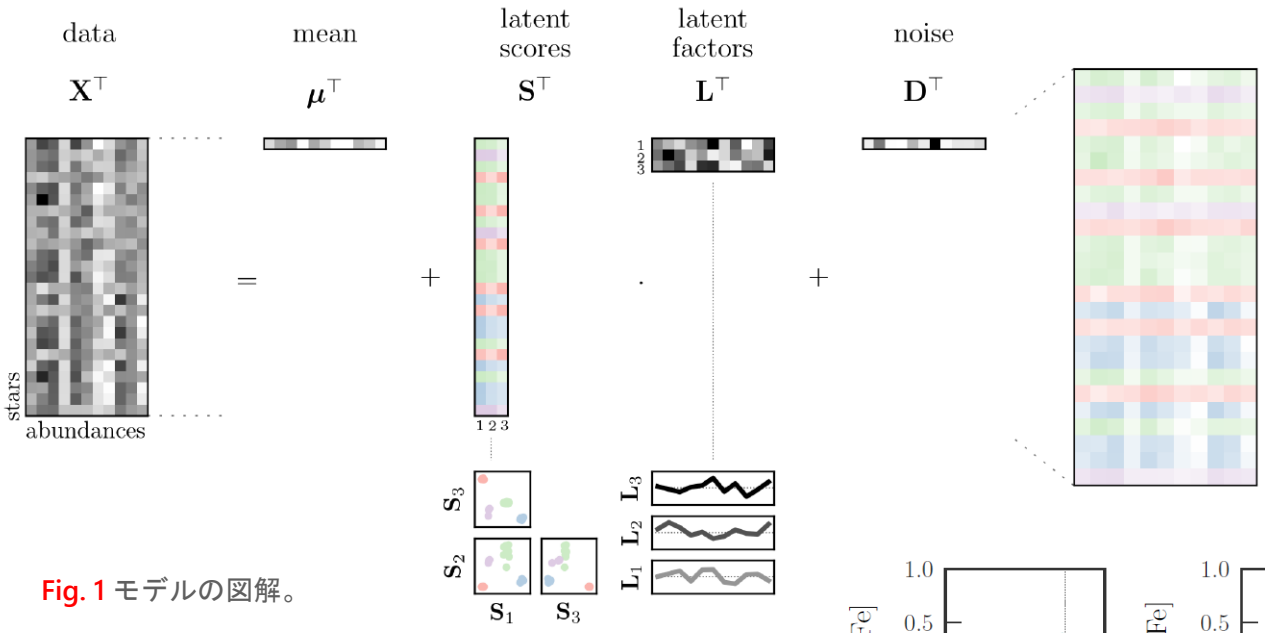


Fig. 1 モデルの図解。

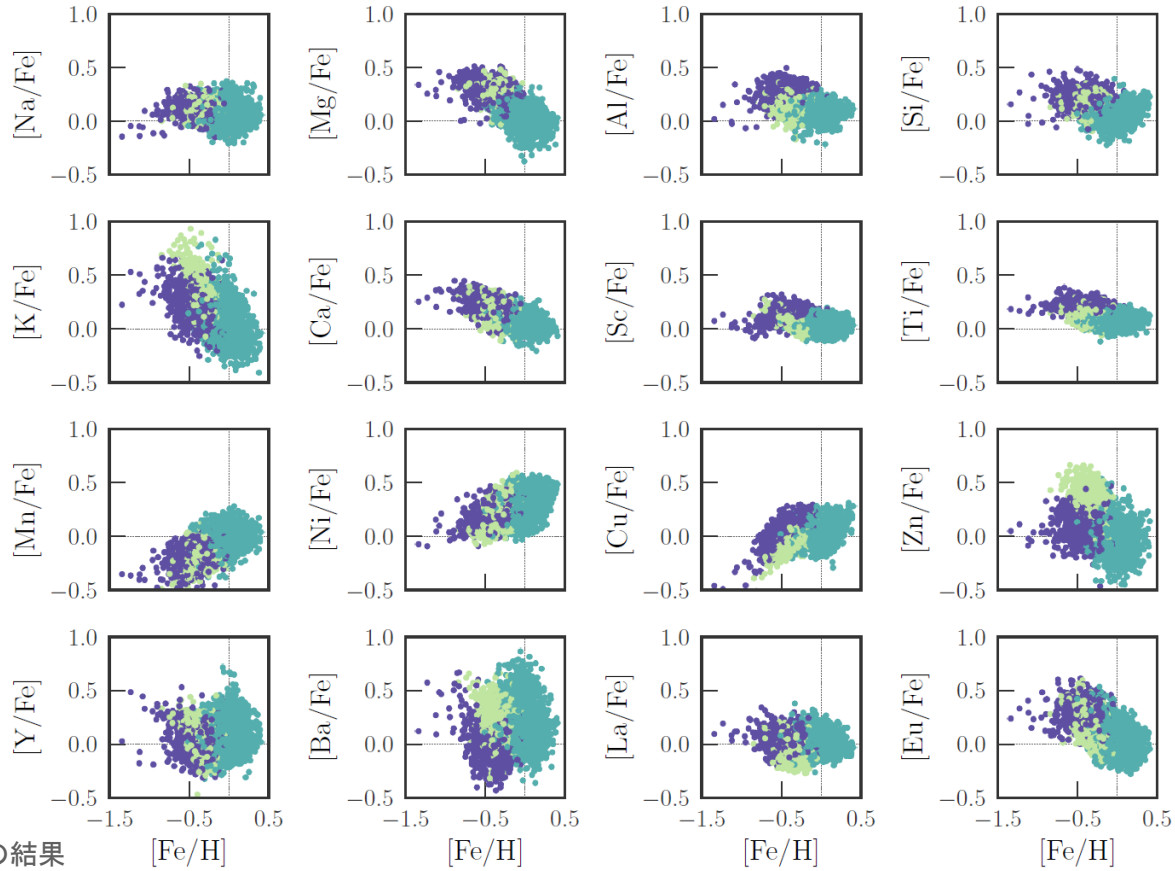


Fig. 7 clustering 解析の結果