

地震動シミュレーションデータの自動分類

Automatic classification of the ground-motion simulation data

赤木 翔* 早川 俊彦* 下野 五月*

Sho Akagi, Toshihiko Hayakawa, Satsuki Shimono

近年、計算機の発達に伴い、多様な震源モデルを想定した多数の地震動シミュレーションが実行可能となった。防災・減災に活用することを目的として、南海トラフや相模トラフにおける巨大海溝型地震のシミュレーションが実施され、結果が蓄積されている。我々は、これらのシミュレーションデータから地震動分布の推定手法の高度化等に活用できる情報を抽出することを目的として、膨大なシミュレーションデータを機械学習により自動的に分類する手法を開発した。本稿では南海トラフの地震動シミュレーションデータを例として、手法の適用結果と応用への展開について述べる。

In recent years, many ground-motion simulations assuming various earthquake source models became possible with the development of computers. Many megathrust earthquakes on Nankai Trough or Sagami Trough are simulated for the purpose of utilizing for disaster prevention and disaster mitigation. We developed automatic classification methods for the big simulation data by machine learning methods to extract the information that improve the estimation methods for earthquake distribution. We describe the application results for Nankai Trough simulation data and the next deployment of the method.

1. まえがき

近年、スーパーコンピュータを始めとする計算機の発達により、地震や津波を再現するシミュレーションを複数並行して高速に実行できるようになった。多様な断層の破壊過程を想定した地震の想定シナリオ（以下、地震シナリオ）に対して広範囲にわたる地震動シミュレーションが実施され、多数のシミュレーション結果が蓄積されている（例えば、参考文献(1)）。これらのシミュレーション結果には、防災上重要な地震ハザードの情報や地震動分布の推定手法の高度化に応用できるような地震動の性質等、防災・減災に広く活用できる有用な情報が含まれている。しかし、1つの地震シナリオのシミュレーション結果は数万点に及ぶ評価地点で計算される地震動指標値からなるため、多数の地震シナリオに共通している性質や各評価地点での地震動の特徴を調査することは容易ではない。

我々は、国立研究開発法人防災科学技術研究所（以下、防災科研）と協力し、地震動シミュレーションの結果から防災・減災に有用な情報を効率的に取り出すことを目的として膨大なシミュレーションデータを概観可能な程度の数のグループに自動的に分類する手法を開発した⁽²⁾。

本稿では、面的に類似した地震動分布を持つ地震シナリオ同士を同一のグループに分類する手法と、多数の地震シナリオを通して類似した地震動が観測される評価地点同士を同一のグループに分類する手法の2つの手法を紹介する。2つの手法は共に、機械学習の代表的な手法である主成分分析とK-means法を組み合わせで構成した。本稿では南海トラフの巨大海溝型地震を想定した地震動シミュレーションデータを例として、2つの手法の適用結果とその特徴及び結果に対する考察を述べる。

2. 地震動シミュレーションデータ

本稿で扱う南海トラフの地震動シミュレーションデータについてその概要を説明する。

防災科研がパッケージ化した地震波伝播シミュレーションツールであるGMS⁽³⁾では、多様な震源モデルの破壊過程を表現するため、震源域、破壊開始点、浅部アスペリティ位置等が設定できる。防災科研では南海トラフの巨大海溝型地震を想定して計369個の地震シナリオの計算を実施した。震源域の計算パラメータとして、南海トラフ震源域全体を分割した計18個の小領域から15パターンの組み合わせを作成した。また、小領域の境界線の交点を10パターンの破壊開始点とした。南海トラフ

*つくば事業部 第四技術部

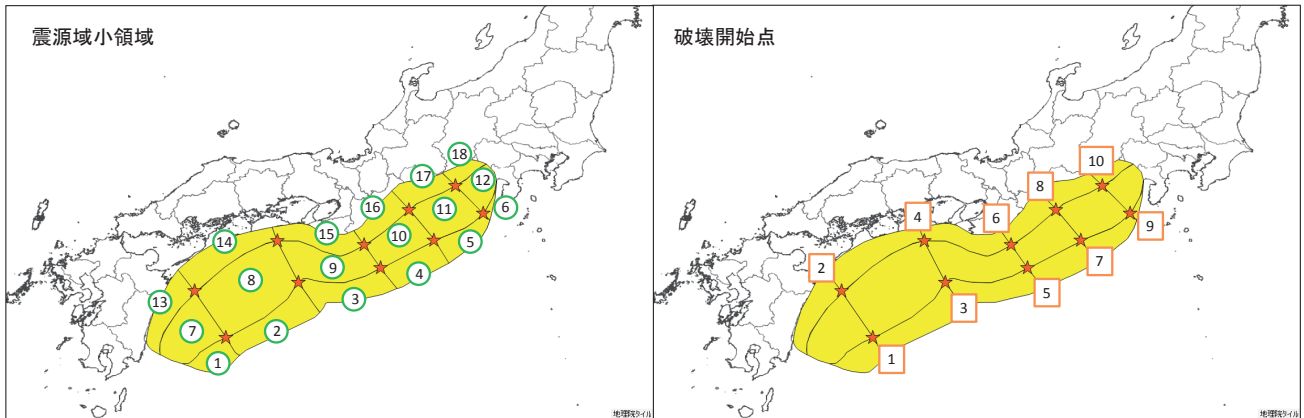


図1 南海トラフ地震シナリオの震源域の小領域と破壊開始点

震源域の小領域と破壊開始点を図1に示す。図1で★は破壊開始点を表す。また、左図の丸囲み数字は震源域の小領域に対して、右図の四角囲みの数字は破壊開始点に対して割り振った番号をそれぞれ示している。地震動の評価範囲は九州地方、四国地方及び本州の東北地方南部までを含む範囲であり、計算結果を出力する評価地点の数は77,609点にのぼる。

3. 基礎技術

開発手法の基礎となる主成分分析と K-means 法について概要を説明する。

3.1 主成分分析

主成分分析(例えば、参考文献(4))は、 M 次元 ($M > 1$)の入力データに対してデータをよく表現する P ($P \ll M$)個の方向ベクトルを見つけ、データの各サンプルを P 個の方向ベクトルで張る空間に射影することで次元削減されたデータを得る手法である。方向ベクトルは射影されたデータ点の分散が大きくなるように選ぶ。分散の大きさが p 番目 ($p = 1, 2, \dots, P$)となる方向ベクトル $\mathbf{u}_p \in \mathbb{R}^M$ を第 p 主成分ベクトルと呼ぶ。 i 番目のサンプル $\mathbf{x}^i \in \mathbb{R}^M$ の第 p 主成分ベクトルへの射影 α_p^i を第 p 主成分スコアと呼ぶ。入力データの各サンプル \mathbf{x}^i は、 p 番目までの主成分ベクトルの線形和で近似される。

$$\mathbf{x}^i \approx \sum_{p=1}^P \alpha_p^i * \mathbf{u}_p \quad (1)$$

元のデータをよく表現する次元 P を決める方法として、累積寄与率による方法がある。寄与率とは、ある次元の分散を全ての次元の分散の総和で除算した値である。累積寄与率は第1次元からある次元までの寄与率の和であり、次元 p の分散を λ_p としたとき、次元 P の累積寄与率 r_p は以下の式で与えられる。

$$r_p = \frac{\sum_{p=1}^P \lambda_p}{\sum_{p=1}^M \lambda_p} \quad (2)$$

累積寄与率 r_p が1.0に近いほど、次元削減により失われる情報量が少なくなる。

入力データの各サンプルを P ($P \ll M$)個の主成分スコアに変換することで、解析対象となるデータの次元が削減され、計算量が大きく減少する。また、主成分ベクトルを調べることで、データ全体の性質を説明するようなデータの特徴を確認することができる。

3.2 K-means 法

K-means 法(例えば、参考文献(5))は、データサンプル間の距離に基づき、入力データを事前に決められた数のグループ(クラスター)に自動的に分類する手法である。サンプル \mathbf{x}^i と、 \mathbf{x}^i が属するクラスター C_k ($k = 1, 2, \dots, K$)の重心 $\boldsymbol{\mu}_k$ との距離の二乗の総和

$$\sum_{k=1}^K \sum_{\mathbf{x}^i \in C_k} \|\mathbf{x}^i - \boldsymbol{\mu}_k\|^2 \quad (3)$$

を最小化するように各サンプルが所属するクラスターを繰り返し更新し、クラスター C_k ($k = 1, 2, \dots, K$)を最適化する。

クラスターの数 K は、クラスター内のデータのばらつきが小さいこと及びクラスター間の距離が離れていることを基準にして決定する。

4. 地震シナリオの分類

面的に類似した地震動分布を持つ地震シナリオ同士を同一のグループに分類する手法について、手法の解析フローと適用結果及び結果に対する考察を記述する。

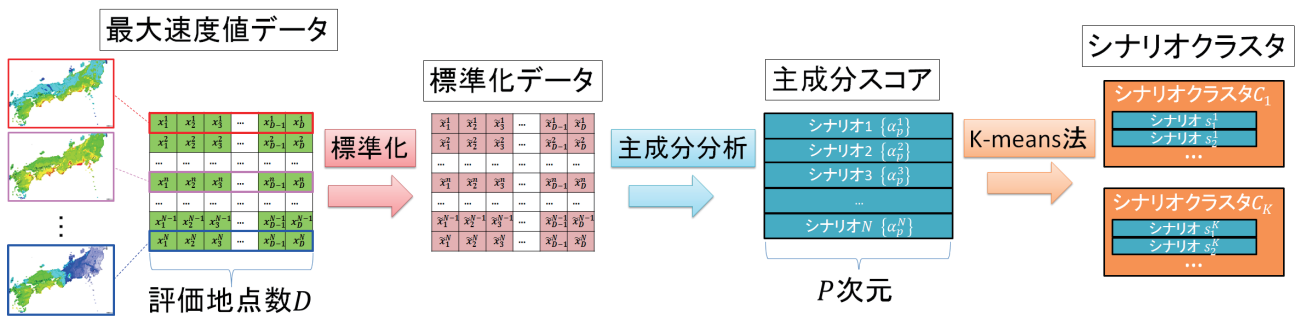


図2 地震シナリオの分類の解析フロー

4.1 地震シナリオの分類手法

地震シナリオを分類することを目的として、主成分分析と K-means 法を組み合わせた手法を開発した。主成分スコアを入力とした K-means 法を行うことで、高次元なデータである地震シナリオを低次元空間において分類することが可能となる。主成分分析を組み合わせることは結果の可視化においても有効である。地震シナリオ等、面的な広がりを持つ高次元のデータは分類結果やデータ点同士の関係性を可視化することに困難が伴うが、低次元の主成分スコアであれば分類されたクラスター内のデータ点を一覧で表示し、それぞれの特徴を調べることも容易である。

地震シナリオに対する分類の解析フローを図2に、処理の順序とその説明を以下に示す。

- (1) 長周期地震動シミュレーションにより、 D 個の評価地点で計算された最大速度値 (cm/s) からなる D 次元実数値ベクトルをひとつのサンプルとするデータを入力とする。入力データのサンプル数はシナリオ数 N となる。
- (2) N 個のサンプルのそれぞれに対して全評価地点での平均値が 0、分散が 1 となるように線形変換を施す。このような変換をデータの標準化と呼ぶ。標準化を行うことにより、振幅の大きさにのみ依存することなく、地震動の空間的な分布の特徴を捉えた

分析が可能となる。

- (3) 標準化された分布データに対して主成分分析を行い、主成分ベクトルと主成分スコアを得る。主成分の次元数は累積寄与率の値から決定する。
- (4) 主成分スコアを K-means 法によって K 個のクラスターに分類する。

以上の手順により、面的に類似した地震動分布を持つ地震シナリオ同士を同一のグループに分類する。

4.2 南海トラフシミュレーションデータへの適用

南海トラフの 369 シナリオの地震動シミュレーションデータを使用して、地震シナリオの分類を行った。

主成分の次元数 P は、累積寄与率が 0.8 を越える次元とし、 $P=9$ と定まった。この結果、 $D=77,609$ 次元のデータが $P=9$ 次元の主成分スコアとして分析可能となった。主成分ベクトルの各要素を地図上にプロットした結果を図3に示す。各主成分ベクトルはノルム 1 に正規化されている。 $p=1,2$ の主成分ベクトルは評価地点全体の大域的な傾向を表す分布となっているのに対し、高次の主成分ベクトルは空間方向に高周波な成分の特徴を表現していることがわかる。例えば $p=1$ の主成分ベクトル u_1 からは、評価範囲の西側で最大速度値が大きくなっている場合に東側では最大速度値が小さくなる傾向があることが読み取れる。

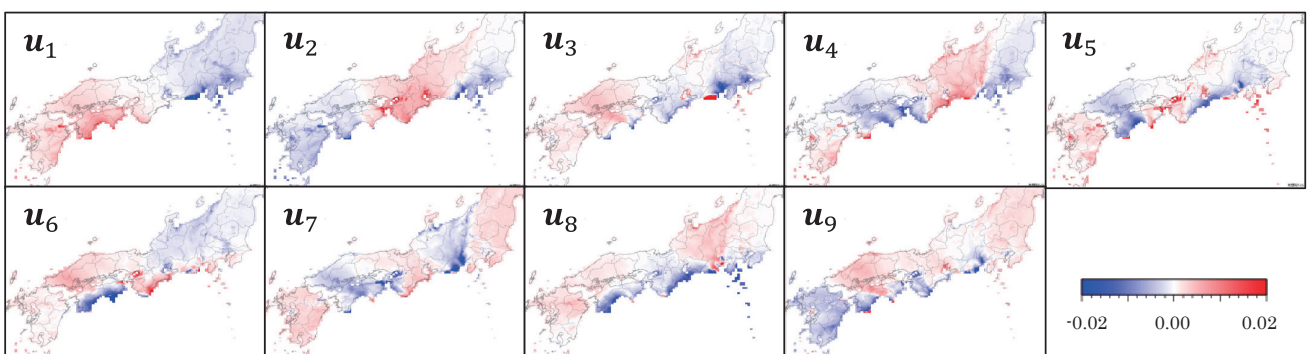


図3 主成分ベクトルの可視化

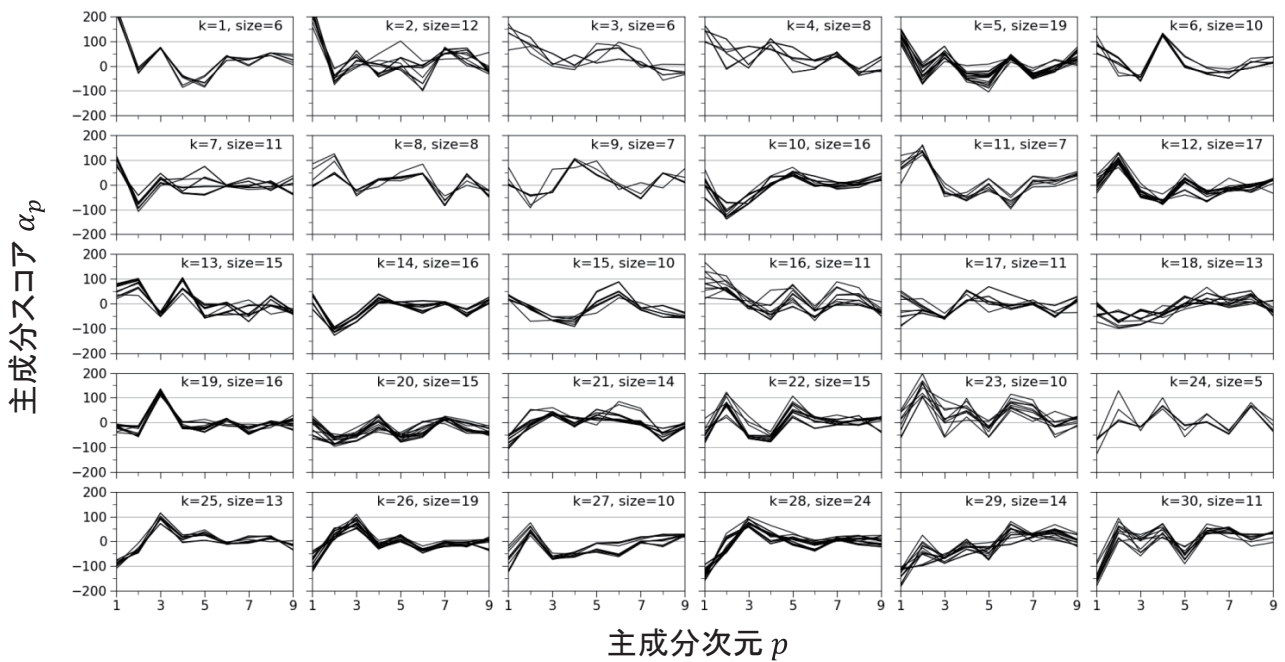


図4 地震シナリオの分類の結果

クラスタ数を $K=30$ として主成分スコアによる地震シナリオの分類を実施し、クラスタごとに主成分スコアをプロットした結果を図4に示す。各パネル内の k はクラスタ番号を、 $size$ はクラスタに含まれる地震シナリオ数を示している。クラスタ番号は第1主成分スコアのクラスタ内平均値の降順から決定した。主成分スコアは平均値が0となるように変換してプロットしている。主成分スコアの分布が似通っている地震シナリオで各クラスタが構成され、また異なるクラスタ間では主成分スコアの特徴が異なっていることがわかる。 $K=30$ よりクラスタ数が少ない場合では、クラスタ内の主成分スコアのばらつきが大きくなる傾向が見られた。

4.3 考察

各クラスタに分類された地震シナリオの主成分スコアとシナリオのパラメータの関係について考察する。

分類された地震シナリオの特徴を調べるため、クラスタ内の代表地震シナリオの最大速度値分布と主成分スコア及びクラスタに分類された地震シナリオに設定された震源域・破壊開始点パラメータの頻度分布の一例を図5、図6に示す。なお、クラスタ内の代表シナリオはランダムに選別した。クラスタ $k=1$ (図5、1段目) とクラスタ $k=30$ (図5、2段目) は、クラスタ内の第1主成分スコア平均値が最も大きいクラスタと最も小さいクラスタである。クラスタ内の代表シナリオの最大速度値分布においても、最大振幅が出ている地域がそれぞれ

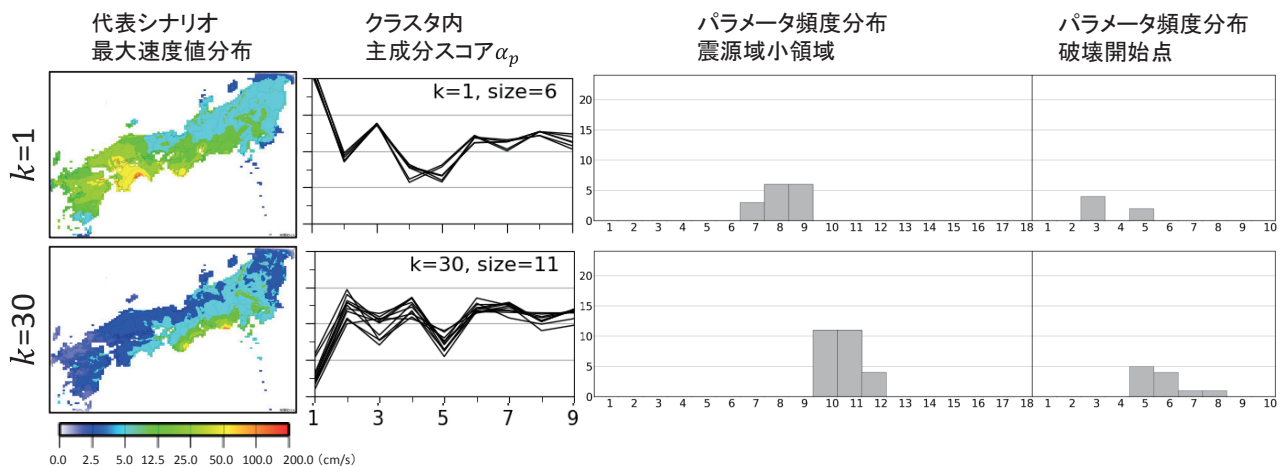


図5 クラスタの代表地震シナリオ最大速度値分布と主成分スコア及びシナリオパラメータの頻度分布 ($k=1, 30$)

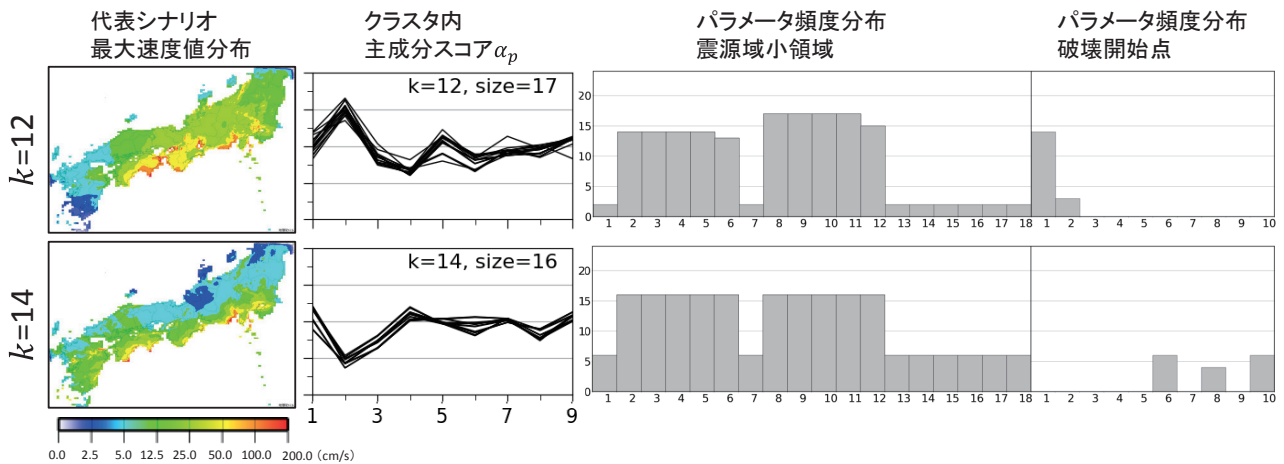


図6 クラスターの代表地震シナリオ最大速度値分布と主成分スコア及びシナリオパラメータの頻度分布 ($k=12, 14$)

主成分ベクトル \mathbf{u}_1 の大小に対応していることがわかる。クラスター内の震源域小領域パラメータ頻度分布に着目すると、クラスター $k=1$ は南海トラフ震源域の西側に位置する小領域7、8、9からなるシナリオのみであるのに対し、クラスター $k=30$ は南海トラフ震源域の東側に位置する小領域10、11、12からなるシナリオのみで構成されていることがわかる。これらのシナリオでは震源域の位置が最大速度分布に大きく寄与していることがわかる。

クラスター $k=12$ (図6、1段目) とクラスター $k=14$ (図6、2段目) は、第1主成分スコアは似通った値となっているが、第2主成分スコアが大きく異なっている。クラスター内のパラメータ頻度分布について確認する。震源域小領域においては、両クラスターが類似した地震シナリオから構成されていることがわかる。一方で破壊開始点においては、クラスター $k=12$ が南海トラフ震源域の西側に位置する破壊開始点1、2のシナリオのみからなるのに対し、クラスター $k=14$ は南海トラフ震源域の北部から北東部に位置する破壊開始点6、8、10のシナリオのみからなっている。これらの地震シナリオでは破壊開始点の位置に従い、地震動の指向性の効果により振幅が大きくなる地域が決まっているといえる。

以上の例のように、主成分スコアとシナリオパラメータには関係性があり、地震シナリオの分類によって得られたクラスター情報を基にしてその関係性を抽出できることがわかった。

開発手法の今後の展開として、シナリオパラメータと主成分スコアの関係性を定量的に評価することができれば、任意のシナリオパラメータに対応する地震動分布を主成分ベクトルの線形和として空間的に密に推定することが可能となる。

5. 評価地点の分類

多数の地震シナリオを通して似通った地震動が観測される評価地点同士を同一のクラスターに分類する手法の解析フローと適用結果及び結果に対する考察を記述する。

5.1 評価地点の分類手法

地震シナリオの分類と同様に、主成分分析とK-means法を組み合わせた手法を構築した。評価地点分類の解析フローを図7に示す。処理の手順を以下に示す。

- (1) ある評価地点に対して、 N 個のシナリオで計算された最大速度値 (cm/s) からなる N 次元実数値ベクトル

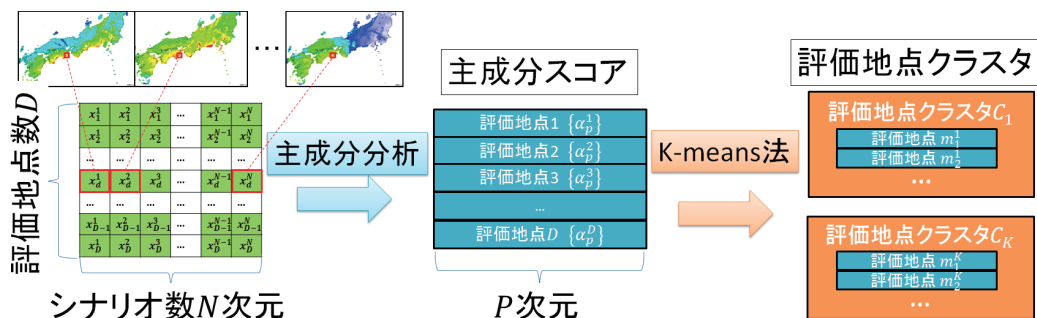


図7 評価地点の分類の解析フロー

をひとつのサンプルとみなしたデータを入力とする。評価地点の分類では、サンプル数は評価地点数 D となる。

- (2) 入力データに対して主成分分析を行い、主成分ベクトルと主成分スコアを得る。主成分の次元数 P は累積寄与率の値から決定する。
- (3) 主成分スコアを K -means 法によって K 個のクラスターに分類する。

以上の手順により多数の地震シナリオを通して似通った地震動が観測される評価地点同士の分類を実現する。

5.2 南海トラフシミュレーションデータへの適用

地震シナリオの分類と同様に、南海トラフの海溝型地震 369 シナリオの地震動シミュレーション結果を使用した。評価地点の分類においては、入力データのサンプル数は $D = 77,609$ 、各サンプルの次元数は $N = 369$ である。

主成分分析の次元数 P は累積寄与率が 0.8 を越える次元とし、 $P = 3$ と定まった。クラスター数は $K = 10$ とした。評価地点の所属クラスターを地図上にプロットした結果を図 8 に示す。同一クラスターに分類された評価地点を同じ色で塗り分けている。

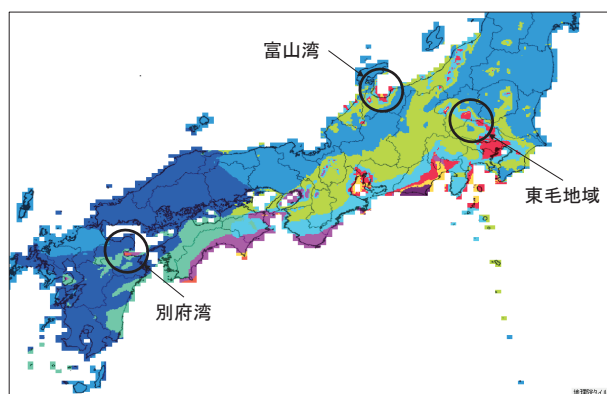


図 8 評価地点の分類の結果

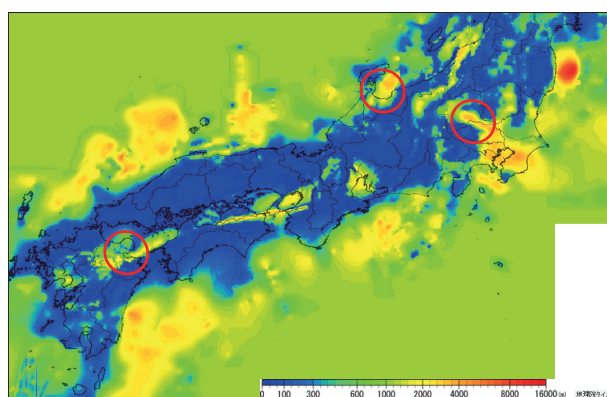


図 9 速度構造モデルの速度層下面深度分布 ($V_s = 1500$ m/s 下面深さ)⁽⁶⁾

5.3 考察

図 8 から、評価地点クラスターはおおむね震源域からの位置関係に基づき、空間的に連続した地点から形成されていることがわかる。一方で、群馬県東毛地域や富山湾と別府湾の沿岸部等では、太平洋沿岸地域のクラスターと同じクラスターに分類されている特徴的な評価地点の分布が確認できる。評価地点の分類結果に対し、地震動シミュレーションの計算設定である速度構造モデルの下面深度分布 ($V_s = 1500$ m/s 下面深さ、図 9)⁽⁶⁾ と比較をとると、特徴的な評価地点が分布している地域で深度が深くなっていることが確認できる。これらの地域では速度構造モデルに従い、各地震シナリオにおいて震源域との位置関係のみに依らず振幅が大きくなるため、南海トラフ震源域に近い評価地点と同じクラスターに分類されたと考えられる。

多数の地震シナリオで似通った地震動となる評価地点に自動分類を適用し、得られた評価地点クラスターが震源域からの位置関係や面的な計算設定である速度構造モデルに基づいて形成されていることを確認した。開発した手法の応用への展開として、評価地点クラスターの中で代表地点を定めてクラスター全体を代表地点と同一視することで、少数の地点での地震動指標値を回帰により推定する手法を面的な地震動指標値分布の推定に発展させることができると考えられる。また、評価地点クラスターとシナリオパラメータとの関係性を抽出する手法⁽⁷⁾を、本稿の開発手法で得られた評価地点クラスターに対しても適用することで、評価地点クラスターを形成する要因をより詳細に分析できる。

6. むすび

広範囲で面的に計算された地震動シミュレーションデータに対して、地震シナリオ及び評価地点をそれぞれ分類する手法を開発し、南海トラフの海溝型地震を想定した多数の長周期地震動シミュレーション結果に対して適用した。シナリオパラメータや地下構造モデルと分類されたクラスターとの関係性について分析し、得られた結果を地震動分布の推定手法の高度化に適用できる可能性が示唆された。今後は防災・減災に資することを目的とし、地震動の曝露人口や建物被害等のリスク評価値を指標として地震シナリオや評価地点の分析を進めていくことが重要な課題となる。

地震動シミュレーションの計算設定パラメータ及び計算結果のデータを利用させていただき、解析手法等についてご指導・ご助言いただいた防災科学技術研究所の関係者様に感謝いたします。本稿に掲載した地図の作成には、国土地理院地図 (<http://cyberjapandata.gsi.go.jp>)

を使用しました。本稿に掲載した図の一部は、GMT⁽⁸⁾を使用して作成しました。

参考文献

- (1) Maeda, T., Iwaki, A., Morikawa, N., Aoi, S., Fujiwara, H. : Seismic-Hazard Analysis of Long-Period Ground Motion of Megathrust Earthquakes in the Nankai Trough Based on 3D Finite-Difference Simulation, *Seismological Research Letters*, **87**, No.6, 1265 ~ 1273 (2016)
<https://doi.org/10.1785/0220160093>
- (2) Maeda, T., Fujiwara, H., Hayakawa, T., Shimono, S., Akagi, S. : Cluster Analysis of Long-Period Ground-Motion Simulation Data with Application to Nankai Trough Megathrust Earthquake Scenarios, *Journal of Disaster Research*, **13**, No.2, 254 ~ 261 (2018)
<https://www.fujipress.jp/jdr/dr/dsstr001300020254/>
- (3) 青井 真, 早川 俊彦, 藤原 広行 : 地震動シミュレータ : GMS, *物理探査*, **57**, 651 ~ 666 (2004)
- (4) Hotelling, H. : Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, **24**, No.6, 417 ~ 441 (1933)
- (5) MacQueen, J. B. : Some Methods for Classification and Analysis of Multivariate Observations, *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 281 ~ 297, University of California Press (1967)
- (6) 防災科学技術研究所 : J-SHIS 地震ハザードステーション (2019)
<https://doi.org/10.17598/nied.0010>
- (7) Maeda, T., Fujiwara, H. : Seismic Hazard Visualization from Big Simulation Data: Cluster Analysis of Long-Period Ground-Motion Simulation Data, *Journal of Disaster Research*, **12**, No.2, 233 ~ 240 (2017)
<https://www.fujipress.jp/jdr/dr/dsstr001200020233/>
- (8) Wessel, P., Smith, W. H. F. : New version of the Generic Mapping Tools released, *Eos, Transactions, American Geophysical Union*, **76**, No.33, 329 (1995)

執筆者紹介

赤木 翔

2015年入社。つくば事業部第四技術部所属。地震・津波防災分野の解析に従事。

早川 俊彦

1998年入社。つくば事業部第四技術部所属、博士（理学）。地震・津波防災分野の解析・システム開発に従事。

下野 五月

2005年入社。つくば事業部第四技術部所属、技術士（応用理学部門）。主として鉄道事業者を対象とした地震防災分野に携わる。