

# ロバスト主成分分析に基づく 沖合津波観測点配置の最適化

## Optimization of Offshore Gauge Configuration with Robust Principal Component Analysis

藤田真粹<sup>1)</sup> 野村怜佳<sup>2)</sup> 大竹雄<sup>3)</sup> 越村俊一<sup>2)</sup> 森口周二<sup>2)</sup> 寺田賢二郎<sup>2)</sup>

Saneiki Fujita, Reika Nomura, Yu Otake, Shunichi Koshimura, Shuji Moriguchi, Kenjiro Terada

<sup>1)</sup>東北大学大学院 工学研究科 (〒 980-8572 宮城県仙台青葉区荒巻字青葉 468-1, E-mail: saneiki.fujita.s8@dc.tohoku.ac.jp)

<sup>2)</sup>東北大学 災害科学国際研究所 (〒 980-8572 宮城県仙台青葉区荒巻字青葉 468-1)

<sup>3)</sup>東北大学大学院 工学研究科 (〒 980-8579 宮城県仙台青葉区荒巻字青葉 6-6-06)

In this study we present an optimization scheme to configure offshore wave gauges using machine learning and sparse sensing techniques. Specifically, the robust principal component analysis is employed for extracting low-dimensional spatial mode from a pre-computed tsunami scenario dataset. According to the low-rank spatial mode that is dominant for the whole tsunami propagation, the optimal sparse gauges' configuration is determined. As a demonstration, the optimal observation gauges off the coast of Shikoku is detected for 1564 possible Nankai Trough earthquakes. The forecast capability of the detected gauges' configuration is examined within the framework of the previous tsunami scenario detection.

**Key Words :** Gauge Configuration Optimization, Robust Principal Component Analysis, Tsunami Forecasting

### 1. 緒言

巨大地震とそれに伴う津波は沿岸地域に甚大な被害を及ぼす。近年、津波被害低減を目的に、沖合での津波観測データを利用したリアルタイムリスク評価手法が提案されている [1]。こうした手法は津波の観測情報を直接的に使用することから、観測地震動データを使用する従来手法に対して、予測不確実性がより小さいという利点がある。一方、国土全体を対象としたリアルタイムリスク評価の実現のためには、広範囲かつ密な津波観測網の導入が望ましいが、計測器の設置や運用に要する費用が問題となる。そのため、スパースな配置でありながら、襲来する津波の情報を効率的に観測できる観測網の設計が必要となる。すなわち、リアルタイム予測に求められる「即時性」を満足するような、津波発生直後の沖合での微小な海面変動に敏感な観測網を設計することが求められる。

本研究ではロバスト主成分分析 (RPCA: Robust Principal Component Analysis) [2] による低ランク構造の特徴量抽出と、スパースセンサーの最適配置手法 [3] を用いた津波観測点配置の最適化を提案する。RPCA は主成分分析 (PCA: Principal Component Analysis) を拡張した手法であり、局所的もしくは偶発的に生じる外れ値を除去した低ランク構造を同定し、データに対して支配的な特徴量の抽出を可能にする。これにより、沿岸部において局所的に増幅した波高成分を外れ値として除去し、沖合での津波伝播を低ランク構造として扱うことが可能となる。検証として、南海トラフ地震・津波を想定し、RPCA と PCA それぞれを用いた最適化によって四国沖合での観測点配置を決定し、比較を行っ

た。具体的には最適化により特定された少数の観測点における観測情報を入力とし、津波シナリオ同定に基づくリスク予測 [1] の精度を比較することで、本手法の優位性を確認する。

### 2. ロバスト主成分分析による沖合観測点配置の最適化手法

#### (1) スパース観測点配置の最適化

本研究では、スパースセンシングのためのセンサー位置最適化手法 [3] に従い、 $n$  点の候補位置での波高情報の復元を可能とする  $p$  ( $< n$ ) 点の組み合わせを特定することを観測点配置の最適化と位置付ける。 $n$  点の候補観測点位置での時刻  $t$  における波高  $\mathbf{x}_t \in \mathbb{R}^n$  は、空間モード行列  $\Phi \in \mathbb{R}^{n \times n}$  を用いて、次式のように表すことができる。

$$\mathbf{x}_t = \Phi \mathbf{a}_t \approx \Phi_r \mathbf{a}_t \quad (1)$$

ここで、 $\Phi_r \in \mathbb{R}^{n \times r}$ 、 $\mathbf{a}_t \in \mathbb{R}^r$  はそれぞれ、 $r$  次元まで次元縮約された空間モード行列と係数ベクトルである。式 (1) から  $n$  次元の波高情報は、次元縮約空間モード行列  $\Phi_r$  を介することで  $r$  次元の状態ベクトル  $\mathbf{a}_t$  で表現できる。

いま、 $n$  個の候補観測点での波高  $\mathbf{x}_t$  と  $p$  ( $< n$ ) 個の少数観測点における波高  $\mathbf{y}_t \in \mathbb{R}^p$  をダウンサンプリング行列  $\mathbf{C} \in \mathbb{R}^{p \times n}$  により次のように関連付けることを考える。

$$\mathbf{y}_t = \mathbf{C} \mathbf{x}_t + \mathbf{w} \approx \mathbf{C} \Phi_r \mathbf{a}_t + \mathbf{w} \quad (2)$$

この式において、 $\mathbf{w} \in \mathbb{R}^p$  は観測情報に含まれる誤差などのノイズベクトルであり、少数の観測点位置を決定するダウンサンプリング行列  $\mathbf{C}$  は次式のように書き表さ

れる。

$$\mathbf{C} = [\mathbf{e}_{\gamma_1} \quad \mathbf{e}_{\gamma_2} \quad \cdots \quad \mathbf{e}_{\gamma_p}]^T \quad (3)$$

ここで、 $\gamma_i \in \{1, \dots, n\}$  は選定される観測点番号であり、 $\mathbf{e}_{\gamma_i} \in \mathbb{R}^n$  は  $\gamma_i$  番目の要素を 1, その他の要素を 0 とした  $n$  次元ユークリッド空間の標準基底である。そして、式 (2) の逆問題を考えると、 $r$  次元の状態量ベクトル  $\mathbf{a}_t$  を、 $p$  次元のスパース観測情報  $\mathbf{y}_t$  から推定することができる。

$$\hat{\mathbf{a}}_t = (\mathbf{C}\Phi_r)^\dagger \mathbf{y}_t \quad (4)$$

ここで、 $(\bullet)^\dagger$  は  $\bullet$  の一般化逆行列である。推定された状態ベクトル  $\hat{\mathbf{a}}_t$  を式 (1) に代入することで、 $p$  次元観測量に基づく  $n$  次元の全候補位置での波高推定が可能となる。

すなわち、推定結果  $\hat{\mathbf{a}}_t$  と真値との誤差  $\epsilon = \mathbf{a}_t - \hat{\mathbf{a}}_t$  が最小化されるような配置である場合、 $p$  点から  $n$  候補点の情報を高精度に復元することができているといえる。よって本研究では、次式のように推定誤差の共分散行列の行列式を最小化するという目的関数を設定する。なお、観測ノイズ  $\mathbf{w}$  についてはガウス分布に従うもの ( $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ ) と仮定した。

$$\min_{\mathbf{C}} \det [\text{Var}(\mathbf{a}_t - \hat{\mathbf{a}}_t)] = \min_{\mathbf{C}} \det [(\mathbf{C}\Phi_r)^T \mathbf{C}\Phi_r]^{-1} \quad (5)$$

このような組み合わせ最適化問題を効率的に解く手法として、列ピボット付き QR 分解に基づく貪欲法 [3] が提案されている。これに倣い、本研究でも次式のような行列展開を適用する。

$$(\Phi_r \Phi_r^T) \mathbf{C}^T = \mathbf{Q}\mathbf{R} \quad (6)$$

ここで、 $\mathbf{Q} \in \mathbb{R}^{n \times n}$  は直交行列であり、 $\mathbf{R} \in \mathbb{R}^{n \times n}$  は上三角行列である。また、 $\mathbf{C} \in \mathbb{R}^{n \times n}$  は  $\mathbf{R}$  の対角成分を降順に並び替えるための置換行列である。逐次的に  $\mathbf{R}$  の対角成分を最大化する行列  $\mathbf{C}$  を特定することにより、最適化目的関数における貪欲的な行列式の最小化が達成される。

## (2) ロバスト主成分分析による低ランクな沖合観測情報の抽出

主成分分析 (PCA) は教師無し機械学習の一種であり、次式のようにデータ行列  $\mathbf{X}$  に特異値分解を適用することにより、特徴量抽出と次元縮約を可能とする手法である。

$$\mathbf{X} = \Phi \Sigma \mathbf{V}^T \approx \Phi_r \Sigma_r \mathbf{V}_r^T \quad (7)$$

ここで、データ行列  $\mathbf{X}$  は、式 (1) における波高ベクトル  $\mathbf{x}_t$  を列方向に並べた次式のものとして定義される。

$$\mathbf{X} = \begin{bmatrix} | & & | \\ \mathbf{x}_{t=1} & \cdots & \mathbf{x}_{t=m} \\ | & & | \end{bmatrix} \in \mathbb{R}^{n \times m} \quad (8)$$

式 (7) における  $\Phi$  と  $\mathbf{V}$  は、それぞれ左特異ベクトルと右特異ベクトルを列方向に並べた行列であり、 $\Sigma$  は特異値を対角成分に降順に並べた行列で、 $\Sigma_r$ ,  $\Phi_r$ ,  $\mathbf{V}_r$  は

それぞれ次元縮約した行列で、大きい順に  $r$  番目までの特異値とそれに対応する  $r$  個の特異ベクトルのみが格納されている。先述のように、本研究における最適化は、PCA により抽出・次元縮約された空間モード行列  $\Phi_r$  に応じて観測点位置の選定を行う。しかし、データの分散を参照する PCA は外れ値の存在に敏感である。そのため、津波波高データに対して PCA を適用した場合、抽出される特徴量は波高の振幅が大きい沿岸部での観測点情報に強く依存してしまい、即時予測性に優れた沖合観測点の情報が反映されにくいという課題が生じてしまう。

このような課題を克服するため、本研究ではロバスト主成分分析 (RPCA) [2] を適用することで沿岸部での情報を効果的に除去し、沖合で伝播する波に対して支配的な特徴量の抽出を目指す。RPCA では、対象とするデータ行列  $\mathbf{X} \in \mathbb{R}^{n \times m}$  が、秩序構造を格納した低ランク行列  $\mathbf{L} \in \mathbb{R}^{n \times m}$  と外れ値を格納したスパース行列  $\mathbf{S} \in \mathbb{R}^{n \times m}$  の和で構成されているものとし、データ行列  $\mathbf{X}$  から各行列  $\mathbf{L}$ ,  $\mathbf{S}$  の同定を目的とする。すなわち、沿岸の高振幅な観測情報を  $\mathbf{S}$ 、沖合の観測情報成分を  $\mathbf{L}$  として可算分解することを可能にする。

この同定は、次式のように低ランク行列  $\mathbf{L}$  のランクとスパース行列  $\mathbf{S}$  の L0 ノルムの最小化問題解くことで達成される。

$$\min_{\mathbf{L}, \mathbf{S}} \text{rank}(\mathbf{L}) + \|\mathbf{S}\|_0 \quad \text{subject to } \mathbf{L} + \mathbf{S} = \mathbf{X} \quad (9)$$

ただし、行列のランク最小化問題と L0 ノルム最小化問題はともに非凸最適化問題に属する。そのため、実際にはそれぞれを凸最適化問題へと緩和した核ノルムと L1 ノルムの最小化問題を解くことにより、近似的に行列の低ランク性とスパース性を満たす行列を同定する。

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{subject to } \mathbf{L} + \mathbf{S} = \mathbf{X} \quad (10)$$

ここで、 $\lambda$  は重みパラメータであり、対象行列  $\mathbf{X} \in \mathbb{R}^{n \times m}$  の次元から  $\lambda = 1/\sqrt{\max(n, m)}$  により設定される。また、 $\|\bullet\|_*$  は核ノルムであり、 $\bullet$  の特異値の総和により算出される。このような等式制約のある凸最適化問題は拡張ラグランジュ法 [4] により容易に解くことができる。

データ行列  $\mathbf{X}$  の代わりに、低ランク行列  $\mathbf{L}$  に対して特異値分解を適用し、次式のように特徴量の抽出と次元縮約を行う。

$$\mathbf{L} = \Phi^L \Sigma^L \mathbf{V}_r^{L^T} \approx \Phi_r^L \Sigma_r^L \mathbf{V}_r^{L^T} \quad (11)$$

外れ値に依存しない低ランク行列  $\mathbf{L}$  に対応する特徴量  $\Phi^L$  は、式 (7) における PCA の特徴量  $\Phi$  に対して、より支配的な低次の空間モードを抽出することが可能となる。すなわち、沿岸の高振幅な観測情報  $\mathbf{S}$  を除去した、沖合の観測情報成分  $\mathbf{L}$  によってのみ、観測点位置最適化 (式 (5)) を考えることができる。

## 3. 検証例題

RPCA による低ランク成分の特徴量抽出及び観測点配置最適化の有効性を四国沖を対象として例証する。まず、南海トラフ地震・津波を想定し、地震規模と発生位

置の組み合わせから作成した 1564 件の仮想的な地震・津波シナリオ [5] のうち、ランダムに選定した 1414 件を学習シナリオとして用い、PCA と RPCA により抽出された特徴量から観測点の最適配置を同定する。残る 150 件をテストシナリオとし、それぞれの最適観測点で得られる情報を入力としたリアルタイム津波予測 [1] を実施する。データ行列  $X$  は TUNAMI-N2[6] による津波シミュレーションから得られる 1414 件のシナリオにおける  $n=134$  の観測点 (図-2) での波高の時刻歴データを集約したものとする。すなわち、各学習シナリオについて式 (8) により構成される行列  $X^i$  を列方向に並べることにより、1414 件の学習シナリオを集約したデータ行列  $X$  を次式のように定義する。

$$X = \begin{bmatrix} X^{i=1} & \dots & X^{i=1414} \end{bmatrix} \quad (12)$$

データ行列  $X$  に対して、PCA と RPCA をそれぞれ適用し、求められた特異値  $\sigma_i$  を次式に代入することから算定した、 $i$  番目の空間モードの寄与率を累積して図-1 に示す。

$$c(i) = \frac{\sigma_i}{\sum_{j=1}^n \sigma_j} \quad (13)$$

図-1 より、RPCA により抽出した空間モード (赤点) は PCA (青点) と比較して、低次の成分がより高い寄与を有していることが確認できる。実際、RPCA による抽出の場合、 $r=7$  までの空間モードの累積寄与率は約 82 %であるが、PCA の場合、データ行列の 80 %を説明するためには  $r=46$  までの空間モードを使用する必要がある。このことから、データ行列から外れ値を除去したことで、支配的な低次の特徴量抽出が達成できているといえる。

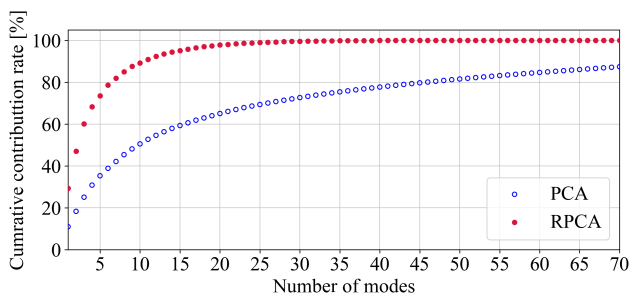


図-1 空間モードの累積寄与率

四国沖に設置した 134 点の候補観測点から、次元縮約後の空間モード  $\Phi_{r=7}$ 、 $\Phi_{r=7}^L$  をそれぞれ用いて最適化された  $p=7$  の観測点配置を図-2 に示す。PCA 空間モード  $\Phi_{r=7}$  を用いた最適化 (青点) では、選定された観測点が沿岸部に集中していることが確認できる。一方で、RPCA 空間モード  $\Phi_{r=7}^L$  に基づく最適配置 (赤点) では沿岸部への集中傾向が緩和され、沖合から沿岸まで網羅的に観測点が選定されていることがわかる。この結果から、RPCA を用いることにより、沿岸部での高振幅成分を外れ値とみなし、沖合での低振幅な低ランク成分の特徴量を抽出することで、より合理的な最適化が実現できるといえる。

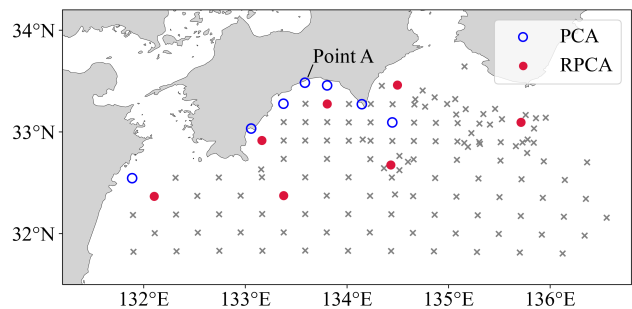


図-2 7 点の観測点配置の最適化結果

図-2 に示す二つの最適配置について、津波予測に対する有用性の検証を実施する。150 件の検証用シナリオにおける、 $p=7$  の観測点での波高情報 10 分間分を入力とし、ベイズ更新に基づく最尤シナリオ同定から津波波高予測を行う。150 件のうち 1 件のテストシナリオについて、高知市沿岸部 (図-2 Point A) での津波波形とその予測結果を図-3 に示す。PCA による最適観測点を用いた予測結果 (青線) に対して、RPCA を使用した場合の予測 (赤線) は時系列データのトレンドや振幅などがテストシナリオの波形 (黒線) に類似していることがわかる。加えて、Point A (図-2) における各テストシナリオの観測波高と予測波高の最大値を、全 150 件についてプロットしたものを図-4 に示す。どちらの予測結果においても正の相関傾向が見られるものの、RPCA により最適化した観測網 (図-4 (b)) の方が全てのテストシナリオに対してより精度の高い予測をしていることが確認できる。以上の結果から、RPCA により抽出された支配的な低次の空間モードと、それを参照した少数の最適観測点は、津波予測において有用であるといえる。

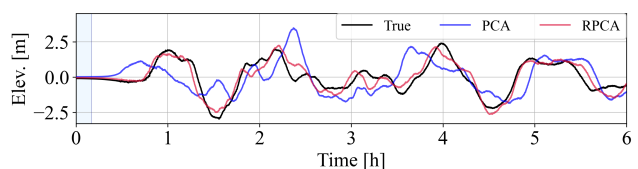


図-3 高知市沿岸観測点 (図-2 Point A) における津波波形の予測結果

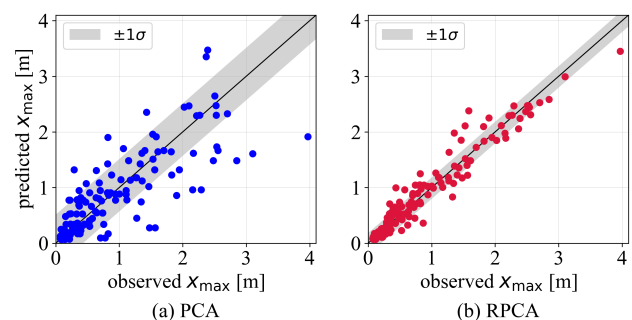


図-4 150 件のテストシナリオに対する観測波高と予測波高の最大値の比較

#### 4. 結言

本研究では、ロバスト主成分分析 (RPCA) とスパースセンシング技術を適用して、津波リスク予測のための観測点位置最適化手法を提案した。具体的には、津波波高データから沖合の観測波高情報が強く反映される低ランク構造の抽出を行い、得られた特徴量に基づいた観測点配置の最適化を行った。四国沖を対象とした検証例題からは、RPCA に基づく最適化により、従来のPCA による最適化では沿岸部に集中してしまっていた観測点を、沖合から沿岸まで網羅的に配置させることが可能であることを確認した。また、南海トラフ地震・津波を想定した即時津波予測を実施したところ、RPCA により算定される7点の最適観測点は、PCA による最適観測点配置より高精度な津波波高予測が可能であることを示した。以上の結果から、提案する手法が即時的な津波予測において有用であると確認できた。

#### 参考文献

- [1] Reika Nomura, Saneiki Fujita, Joseph M Galbreath, Yu Otake, Shuji Moriguchi, Shunichi Koshimura, Randall J LeVeque, and Kenjiro Terada. Sequential bayesian update to detect the most likely tsunami scenario using observational wave sequences. *Journal of Geophysical Research: Oceans*, Vol. 127, No. 10, p. e2021JC018324, 2022.
- [2] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, Vol. 58, No. 3, pp. 1–37, 2011.
- [3] Krithika Manohar, Bingni W Brunton, J Nathan Kutz, and Steven L Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Control Systems Magazine*, Vol. 38, No. 3, pp. 63–86, 2018.
- [4] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.
- [5] 野村怜佳, 藤田真粹, 大竹雄, 森口周二, 越村俊一, 寺田賢二郎, 橋詰正広. 模擬津波動態とベイズ更新によるリアルタイム津波リスク評価手法. 日本計算工学会論文集, Vol. 2022, pp. 20220003–20220003, 2022.
- [6] Fumihiko Imamura, Ahmet Cevdet Yalciner, and Gulizar Ozyurt. Tsunami modelling manual. *UN-ESCO IOC international training course on Tsunami Numerical Modelling*, 2006.