

令和 4 年 6 月 21 日現在

機関番号：17301

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K11861

研究課題名（和文）時空間データのスパースモデリングと実データへの応用による現象理解

研究課題名（英文）Sparse modeling of spatio-temporal data and phenomenon understanding by application to real data

研究代表者

西井 龍映（Nishii, Ryuei）

長崎大学・情報データ科学部・教授

研究者番号：40127684

交付決定額（研究期間全体）：（直接経費） 3,300,000 円

研究成果の概要（和文）：本研究の目的は時空間実データに対して統計モデルを開発・評価し、当該現象の特徴を把握することである。3年間の研究期間では種々の実データについての統計モデリングを行った。以下は論文として発表した主な研究成果である。

1) 色認識空間における楕円の推定論文, 2) 植物表現型推定におけるコンピュータビジョンのレビュー論文, 3) スパース回帰分析と回帰分析（単行本）, 4) 時空間データの回帰分析における効率的変数選択法, 5) 高解像度のハイパースペクトル画像にもとづく教師なし土地利用被覆推定, 6) 植物の表現型に関する遺伝子や環境要因の特定, 7) 太陽風が地震の引き金の一つであることの実証分析。

研究成果の学術的意義や社会的意義

太陽風が地震に影響しているという仮説を、地震頻度や太陽風に関する9次元の物理計測値から検証し、マグニチュード6より小さい地震頻度に影響することを示した。このように異質なビッグデータを考察することで、新しい知見が得られることの好例である。またオオムギの収穫量等の表現形を、遺伝子タイプや生育方法のうち重要な変数を選んでモデル化することにより、穀物の増産や炭素固定に貢献できる。さらに電力売買において、明日の電力使用が不足せず、かつコストの意味で効率な予測方法を研究した。

研究成果の概要（英文）：The purpose of this study is to develop and evaluate statistical models for real space-time data and to understand the characteristics of the phenomenon. During the three-year research period, statistical modeling was performed on various actual data. The following are the main research results.

・ Estimation of ellipses in color recognition space, ・ Computer vision review paper in plant phenotypic estimation, ・ Sparse regression analysis and pattern recognition (text book), ・ Efficient variable selection in regression analysis of spatio-temporal data, ・ Unsupervised land use cover detection based on high-resolution hyperspectral images, and ・ Identification of genes and environmental factors related to plant phenotypes, and ・ Empirical analysis that solar wind is one of the triggers of earthquakes, were performed.

研究分野：応用統計学

キーワード：時空間データ解析 スパースモデリング ネットワーク推定 深層学習 非対称損失 精密農業 カラーマッチング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

ビッグデータから新しい知を獲得することの重要性が語られて久しい。代表者は学術分野や企業活動から得られるビッグデータの解析を通して当該分野に貢献すること、およびその過程で現れた統計的・機械学習的問題を考察するという研究スタイルを続けてきた。異種のデータ間の関連を数理モデルにより探索すると、新しい知の発見につながることを強調したい。

本研究は代表者・分担者との協働により、学術研究や企業活動から得られる現実の時空間データに対して、統計モデルを適応・改良し、当該現象およびモデルの特徴を把握することを目的とする。さらに課題解決の過程で得られた新たな課題を考察する。本課題では(1) 統計学や機械学習における基礎的問題、および特に研究の重点をおく応用分野(2)、(3)を考察する。研究概要は以下の通りである。

- (1) 環境・社会から得られる時空間データの数理モデリングと効率的な推定法の開発
- (2) 植物のゲノム情報と表現型の関連解析による表現型推定法の開発
- (3) 環境リモートセンシングデータと経済社会データとの融合による新しい知の発見

### 2. 研究の目的

1. で述べた3通りの課題について、次の具体的研究目的を設定した。

#### (1) スパースな時空間モデルの基礎研究

太陽風と地震の関連解析から、機械学習の手法であるガウスカーネルに基づくサポートベクター回帰での説明変数の選択に関する問題が喚起された。そこで深層学習によるモデリングを考察し、より高精度な予測を考察する。また課題(3)の時空間回帰モデルの母数の逐次推定では、周辺との空間や時間依存性のため、計算負荷の高い固有値計算が必要となる。この推定および変数選択の効率化について研究する。

#### (2) 植物のゲノム情報による表現型予測

植物の形質(開花や結実、環境への適応性、種子の量など)は、ゲノム上にコードされる遺伝子(遺伝子要因)と光や温度といった環境変化(環境要因)の相互作用によって決定される。また、植物の形質のばらつき(表現型)は、発芽からの生長過程における相互作用の積み重ねとして顕在化する。しかし、ゲノム上の遺伝子は数万にのぼり、季節や生長状態によって遺伝子の働きも大きく変化する。そこでスパースモデリングは、表現型の説明変数となる遺伝子や環境要因の特定に有効であり、精密農業や作物の育種を効率良く進めるために有用な情報の抽出に貢献することを目的とする。

#### (3) リモートセンシングデータと経済社会データとの融合

世界各国の民主化の程度は、周辺国の民主化の程度やGDP(国内総生産)に依存することが知られている。このようなデータのモデル化のため、空間依存性や時間依存性を取り入れた複数のモデルが提案されている。ただ時空間回帰モデルのモデル推定は、2つの依存性のために高速での実行は難しい。ここでは説明変数の選択が高速に行える手法を提案し、実証分析を行うことを目的とする。

### 3. 研究の方法

研究の主な道具としては、統計的な時空間モデルである。また非線形度が高い回帰データでは、深層学習を含む機械学習によるモデルも併用した。また変数選択のためには、AIC (Akaike Information Criterion) 等の情報量規準や L1 罰則 (回帰係数の絶対値和による罰則) 等を用いてスパースモデリングを適応した。

共同研究の進め方：代表者・分担者とは主にオンライン会議で相互に密接に情報を交換した。得られた研究成果は国内外のシンポジウム、学会でオンラインあるいは口頭で発表し、学術雑誌で発表した。

### 4. 研究成果

論文として発表した主な研究成果の概要をそれぞれ記入する。

(1) 色彩科学では、人には色のズレがあっても同じ色と見なす視覚特性があり、同じ色と認識する境界は Lab 表色系では楕円体になることが知られている (カラーマッチング)。一方で 19 色の色中心と楕円体の形を決定する正定値行列は、DuPont dataset として公開されている。この教師データに基づき、色に応じた楕円体を予測する 2 種類の統計的方法論を提案した。いずれの手法も当該分野において先駆的研究であり、今後の色楕円研究のベンチマークとなり得る研究成果である。

(2) 植物の表現型の変化 (例：開花、出穂) を特定するために、コンピュータービジョンを使用して画像やビデオから有用な情報を抽出することは重要な手法になっている。深層学習ベースの技術によって強化された画像解析手法、多重分光画像による植物の生育と生理反応に関連するデータセット、作物育種や精密農業のための大規模なフィールド表現型分析による自動化された植物表現型分析の最近の発展についてレビューした。

(3) 時空間データの回帰分析モデルでは、周辺からの影響を推定するために固有値計算や逐次計算が求められる。ここでは 2 段階推定法により最尤推定量の近似計算を高速に行う手法を提案した。またインドネシアの州ごとの経済指標の時空間回帰分析の実証研究を行い、最適な説明変数の組み合わせを探索する手法を提案した。

(4) ハイパースペクトル画像 (HSI) における教師なし特徴抽出について考察し、オートエンコーダー (AE) ネットワークに基づいて HSI からスペクトル-空間特徴を抽出する新しいアプローチを提案した。提案手法では、通常の AE ベースの特徴抽出器は隣接する入力との依存関係を考慮している。これにより、AE が提供する特徴量で表現された滑らかな判別境界を生成することができた。

(5) 回帰分析において、回帰係数は通常は最小二乗法のように対称な損失関数の最小化で推定される。本論文では、非対称損失の期待値を最小化する推定法を提案した。このとき損失の分散も同時に小さくすることができることを示した。本手法は電力市場からの電力調達からヒントを得たものである。

(6) マルチオミクス解析に基づいて植物の健康状態をモニタリングするための分析技術の最近の進歩と、複数のオミクス分野から得られた異種のデータを統合して、目的の形質に関連する有益な因子を特定するための戦略についてレビューした。さらに、3次元表現型、植物の根の表現型、埋め込み型/注入型センサー、手頃な価格の表現型デバイスなど、さまざまな手段で植物の成長を非侵襲的かつ継続的にモニタリングすることを可能にする新しいフェノミクス技術を紹介した。

(7) 空間データに対する回帰モデルのパラメータ推定について考察する。対象となる変数を正確に予測するためには、隣接するセルからの影響を取り入れる必要があることはよく知られて

いる．ここでは、次の2段階の方法を提案する． 独立性を仮定して、説明変数の最適な非線形回帰関数を求める． 推定された関数によって目標値を予測する． 中心と近傍の予測値との線形回帰に基づく補正により、最終的な予測値を得る．実際のデータに適用したところ、非常によく機能し、説明変数の近隣効果の違いを明確にすることに成功した．なお、本手法は一般的な設定でも適用可能である．

(8) 地震を誘発する要因としては、潮汐力、降雨、人工貯水池の建設、採掘、化石燃料の抽出などが文献に記載されている．ここでは太陽風が地震のトリガの一つであるという仮説を検証するため、全球で観測されるマグニチュード 5.9以下の地震頻度を目的変数、説明変数として時間遅れを持つ地震頻度とした自己回帰型非線形モデルを推定した．さらに説明変数として太陽風の速度、動圧、温度等の9次元物理量を追加した回帰モデルは、自己回帰モデルより有意に予測能力が上がることを示した．モデルとしては深層学習を用いた．地震頻度や物理量の時間遅れに対するTCN: Temporal Convolutional Networkを適応することにより、テストデータで一日後のマグニチュード 4-4.9の地震頻度を相関0.65と予測できた．

(9) 空間で観測された Spatial Panel Dataの典型的な回帰モデルは Spatial Durbin Modelで与えられる．モデルには空間的隣接関係を表す既知行列、および各小地域の標高や人口のような環境的・社会的説明変数が用いられる．モデルの母数推定は容易ではなく、また変数選択が重要で求められる．ここでは逐次選択により最適な変数組を求める手法を開発した．

(10) 植物の形質は、ゲノム上にコードされる遺伝子要因、光や温度といった環境要因の相互作用によって決定される．また植物の表現型は、発芽からの生長過程における相互作用の積み重ねとして顕在化する．しかしゲノム上の数万にのぼる遺伝子は、季節、生長状態、環境ストレスによって遺伝子の働きやネットワークは変動し、タンパク質やホルモン、その他の代謝物といった高次の分子ネットワークも大きく変化する．ここではスパースモデリングによる変数選択により表現型の説明変数となる遺伝子や環境要因の特定を行った．

(11) 今までの研究とりまとめとして「スパース回帰分析とパターン認識」と題した単行本を共著出版した．すでにデータ解析の標準的手法となったスパース回帰分析、判別分析、深層学習、サポートベクターマシン、ランダムフォレストなどをRコードとともに解説した．特に深層学習はRのパッケージ keras で実装し、具体例も豊富で実用性も高いと評価された．

## 5. 主な発表論文等

〔雑誌論文〕 計15件（うち査読付論文 15件／うち国際共著 7件／うちオープンアクセス 9件）

1. 著者名 Leilei Li, Hong Gu, Ryosuke Kikuyama, Ryuei Nishii, Pan Qin	4. 巻 12
2. 論文標題 Investigation of Relation between Solar Activity and Earthquakes with Deep Learning Method	5. 発行年 2021年
3. 雑誌名 International Journal of Geosciences	6. 最初と最後の頁 704 - 713
掲載論文のDOI（デジタルオブジェクト識別子） 10.4236/ijg.2021.128040	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 M. N. Esfahani, K. Inoue, K. H. Nguyen, H. D. Chu, Y. Watanabe, A. Kanatani, D.J. Burritt, K. Mochida, L.-S. P. Tran	4. 巻 44
2. 論文標題 Phosphate or nitrate imbalance induces stronger molecular responses than combined nutrient deprivation in roots and leaves of chickpea plants	5. 発行年 2021年
3. 雑誌名 Plant Cell Environ	6. 最初と最後の頁 574 - 597
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/pce.13935	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Shun Tamaki, Keiichi Mochida, Kengo Suzuki	4. 巻 19
2. 論文標題 Diverse Biosynthetic Pathways and Protective Functions against Environmental Stress of Antioxidants in Microalgae	5. 発行年 2021年
3. 雑誌名 Plants	6. 最初と最後の頁 1 - 17
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/plants10061250	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 E. Okamura, K. Ohtaka, R. Nishihama, K. Uchida, A. Kuwahara, K. Mochida, M. Yokota Hirai	4. 巻 478
2. 論文標題 Diversified amino acid-mediated allosteric regulation of phosphoglycerate dehydrogenase for serine biosynthesis in land plants	5. 発行年 2021年
3. 雑誌名 Biochemical Journal	6. 最初と最後の頁 2217 - 2232
掲載論文のDOI（デジタルオブジェクト識別子） 10.1042/BCJ20210191	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 N. Loudya, P. Mishra, K. Takahagi, Y. Uehara-Yamaguchi, K. Inoue, L. Bogre, K. Mochida, E. Lopez-Juez	4. 巻 22
2. 論文標題 Cellular and transcriptomic analyses reveal two-staged chloroplast biogenesis underpinning photosynthesis build-up in the wheat leaf	5. 発行年 2021年
3. 雑誌名 Genome Biology	6. 最初と最後の頁 1 - 30
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13059-021-02366-3	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 S. Koda, F. Melgani and R. Nishii	4. 巻 17
2. 論文標題 Unsupervised spectral-spatial feature extraction with generalized autoencoder for hyperspectral imagery	5. 発行年 2020年
3. 雑誌名 IEEE Geoscience and Remote Sensing Letters	6. 最初と最後の頁 469-473
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/LGRS.2019.2921225	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 N. Yamaguchi, Y. Yamaguchi and R. Nishii	4. 巻 -
2. 論文標題 Minimizing the expected value of the asymmetric loss function and an inequality for the variance of the loss	5. 発行年 2020年
3. 雑誌名 Journal of Applied Statistics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/02664763.2020.1761951	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 K. Mochida, R. Nishii and T. Hirayama	4. 巻 61
2. 論文標題 Decoding plant-environment interactions that influence crop agronomic traits	5. 発行年 2020年
3. 雑誌名 Plant and Cell Physiology	6. 最初と最後の頁 1408-1418
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/pcp/pcaa064	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Keiichi Mochida, Alexander E Lipka, Takashi Hirayama	4. 巻 61
2. 論文標題 Exploration of Life-Course Factors Influencing Phenotypic Outcomes in Crops.	5. 発行年 2020年
3. 雑誌名 Plant & Cell Physiology	6. 最初と最後の頁 1381-1383
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/pcp/pcaa087	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Ryo Harada, Toshihisa Nomura, Koji Yamada, Keiichi Mochida, Kengo Suzuki	4. 巻 8
2. 論文標題 Genetic Engineering Strategies for Euglena gracilis and Its Industrial Contribution to Sustainable Development Goals: A Review.	5. 発行年 2020年
3. 雑誌名 Frontiers in Bioengineering and Biotechnology	6. 最初と最後の頁 790-790
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/fbioe.2020.00790	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 K. Mochida, S. Koda, K. Inoue, T. Hirayama, S. Tanaka, R. Nishii and F. Melgani	4. 巻 8 (1)
2. 論文標題 Computer vision-based phenotyping for improvement of plant productivity	5. 発行年 2019年
3. 雑誌名 GigaScience	6. 最初と最後の頁 1-12
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/gigascience/giy153	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 T. Nakamoto, R. Nishii and S. Eguchi	4. 巻 1950002
2. 論文標題 Predicting precision matrices for color matching problem	5. 発行年 2019年
3. 雑誌名 International Journal of Mathematics for Industry	6. 最初と最後の頁 1-10
掲載論文のDOI (デジタルオブジェクト識別子) 10.1142/S2661335219500023	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Hirayama T, Saisho D, Matsuura T, Okada S, Takahagi K, Kanatani A, Ito J, Tsuji H, Ikeda Y, Mochida K	4. 巻 pcaa046
2. 論文標題 Life-course monitoring of endogenous phytohormone levels under field conditions reveals diversity of physiological states among barley accessions	5. 発行年 2020年
3. 雑誌名 Plant Cell Physiologists	6. 最初と最後の頁 1-32
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/pcp/pcaa046	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Nomura T, Inoue K, Uehara-Yamaguchi Y, Yamada K, Iwata O, Suzuki K, Mochida K	4. 巻 11
2. 論文標題 Highly efficient transgene-free targeted mutagenesis and single-stranded oligodeoxynucleotide-mediated precise knock-in in the industrial microalga <i>Euglena gracilis</i> using Cas9 ribonucleoproteins	5. 発行年 2019年
3. 雑誌名 Plant Biotechnology Journal	6. 最初と最後の頁 2032-2034
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/pbi.13174	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Wang P, Yamaji N, Inoue K, Mochida K, Ma JF.	4. 巻 11
2. 論文標題 Plastic transport systems of rice for mineral elements in response to diverse soil environmental changes	5. 発行年 2019年
3. 雑誌名 New Phytologist	6. 最初と最後の頁 156-169
掲載論文のDOI (デジタルオブジェクト識別子) 10.1111/nph.16335.	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計9件 (うち招待講演 2件 / うち国際学会 5件)

1. 発表者名 持田恵一、金俊植、高萩航太郎、金谷麻加、井上小楨、上原由紀子、清水みなみ、最相大輔、井藤純、服部公央亮、岡田聡史、池田陽子、梅崎太造、辻寛之、平山隆志
2. 発表標題 Diversity of developmental trajectories in barley illustrated through deep phenotyping ディープフェノタイピングによるオオムギの成長トラジェクトリの多様性の解明
3. 学会等名 第63回日本植物生理学会年会
4. 発表年 2022年



1 . 発表者名 Gigih Fitrianto, Ryuei Nishii, Shojiro Tanaka
2 . 発表標題 Attesting Interlinkages between Income Inequality and the Covariates by Variable Selection for Sustainable Regional Development: A Case Study of Palm Oil Industry in Sumatra Island, Indonesia
3 . 学会等名 Hiroshima International Conference on Peace and Sustainability ( 国際学会 )
4 . 発表年 2022年

1 . 発表者名 R. Nishii, P. Qin and R. Kikuyama
2 . 発表標題 Solar activity is one of triggers of earthquakes with magnitudes less than 6
3 . 学会等名 Proc. IEEE International Geoscience and Remote Sensing Symposium ( 国際学会 )
4 . 発表年 2020年

1 . 発表者名 R. Nishii and S. Tanaka
2 . 発表標題 Two stage estimation procedure of non-linear regression functions for spatially-dependent data
3 . 学会等名 Proc. IEEE International Geoscience and Remote Sensing Symposium ( 国際学会 )
4 . 発表年 2020年

1 . 発表者名 S. Tanaka, R. Nishii and G. Fitrianto
2 . 発表標題 Two stage estimation procedure for spatial regression models and model selection
3 . 学会等名 IGARSS 2019 ( 国際学会 )
4 . 発表年 2019年

1. 発表者名 G. Fitrianto, S. Tanaka and R. Nishii
2. 発表標題 Analysis of regional economic growth against crisis
3. 学会等名 ISI World Statistics Congress (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 持田 恵一、高萩 航太郎、上原 由紀子、井上 小楨、金谷 麻加、清水 みなみ、最相 大輔、平山 隆志
2. 発表標題 野外生育環境におけるオオムギ成長過程の多様性
3. 学会等名 第61回日本植物生理学会年会
4. 発表年 2020年

1. 発表者名 持田 恵一
2. 発表標題 オオムギのライフコース研究
3. 学会等名 第136回日本育種学会講演会 (招待講演)
4. 発表年 2019年

1. 発表者名 江口 真透, 中本 尊元, 西井 龍映
2. 発表標題 正定値行列の一般化平均 - 色認知空間上の楕円予測 -
3. 学会等名 統計関連学会連合大会
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 梅津 佑太, 西井 龍映, 上田 勇祐	4. 発行年 2020年
2. 出版社 講談社	5. 総ページ数 208
3. 書名 スパース回帰分析とパターン認識	

〔産業財産権〕

〔その他〕

長崎大学 情報データ科学部 教員紹介 <a href="https://www.idsci.nagasaki-u.ac.jp/staff/staff1.html">https://www.idsci.nagasaki-u.ac.jp/staff/staff1.html</a> 理化学研究所 環境資源研究センター バイオ生産情報研究チーム <a href="http://www.csr.s.riken.jp/jp/labs/birt/index.html">http://www.csr.s.riken.jp/jp/labs/birt/index.html</a> 田中章司郎のウェブページ <a href="http://www.shojiro-tanaka.net/">http://www.shojiro-tanaka.net/</a>
---

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	田中 章司郎  (Tanaka Sojiro)  (00197427)	広島経済大学・メディアビジネス学部・教授   (35402)	
研究分担者	持田 恵一  (Mochida Keiichi)  (90387960)	国立研究開発法人理化学研究所・環境資源科学研究センター・チームリーダー   (82401)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------

イタリア	University of Trento			
インドネシア	Gadjah Mada University			