

機械学習による局地気象予報手法の開発

Development of method of spot weather forecast using machine learning

吉兼 隆生^{*1}
Takao Yoshikane

芳村 圭^{*1}
Kei Yoshimura

^{*1} 東京大学生産技術研究所
Institute of Industrial Science, The University of Tokyo

Local weather forecasting is a very challenging issue because it is technically very difficult and there are a lot of problems with numerical model for now. In this study, we found the linkage between the observation and simulated regional patterns and then estimate the local precipitation, temperature, and surface wind recognizing the regional distribution patterns, which is corresponded to the weather patterns, using machine learning (A regression model of support vector machine). As the result, the model biases were drastically reduced by the innovative approach. This method could greatly contribute to realize the local weather forecast.

1. はじめに

1.1 数値シミュレーションの限界

今いる場所の天気はどうなるのか。天気の変化が自然災害や経済活動に直結するため、古くから局地気象予報の実現が期待されてきた。近年、計算機性能の飛躍的な向上に伴い、数値シミュレーションによる気象予測精度も大きく向上してきた。しかし、局地気象予報については多くの問題があり、未だに実現できていない。局地気象に大きな影響を与えるメソスケール気象現象のメカニズムに未解明の部分が多く、複雑な現象を適切にモデル化できていないことが主な原因として挙げられる。数 km の解像度では複雑な地形に対応したプロセスを再現することは難しく、モデルバイアスの大きな原因となる。実際の局地気象は、地形だけでなく、土地利用分布、海陸分布の影響を受けており、降水過程に大きな影響を与える。降水活動が他の気象現象(地上風など)に影響を与えて、局地気象をより複雑にする。降水過程など物理プロセスについても未解明の部分が多く、複雑な現象を適切に再現するには至っていない。

数値モデルの高解像度化が解決方法の1つと考えられるが、現状では技術的な難しさに加えて膨大な計算機資源を必要とするため、実用化にはかなりの時間を要するかもしれない。また、高解像度化で本当に複雑な局地気象を予測できるのかもよく分かっていない。自然現象は極めて複雑であり、それらを完全に理解してモデル化できるのか不明な点も多い。(自然を理解してモデル化するのは自然科学分野において目標であり、研究の主流であるが)。何れにしても、数値シミュレーションには限界があり、数値シミュレーションのみに依存せず局地気象予測を行う方法を見つけ出す必要がある[1]。

数値シミュレーションを行う以前は、経験的に知られていた天気の変化パターンにから局地気象を予測してきた[2]。「観天望気」とも呼ばれており、現在も船舶の出航の判断などに利用されている[3][4][5]。複雑なプロセスやメカニズムは理解できていないが、もし明らかに何らかの規則性(パターン)があるならば、それに関係した局地気象を予測することは可能だろう。このような経験則に基づく気象予測も過去に多く試みられている。例えば、地上気圧パターンなどから予測する方法などがあるが、十分な

精度が得られているとは言い難い。局地気象は、様々な局地的な要因の影響を受けて複雑にかつ短時間で大きく変化する特性があり、広域での気圧パターンとの関係性がそれほど大きくないためと推測される。予測精度を高めるには局地気象と強い関係性をもつパターンを認識する必要があるだろう[5]。

1.2 局地気象予測における機械学習の可能性

2018 年度人工知能学会での発表では、数値シミュレーションによる降水パターンを認識して局地降水を推定した[6]。広域での降水変化パターンが観測とよく一致しており、数値シミュレーションと実際の地形の違いが、降水分布の違いを形成したと仮定し、機械学習を用いて実証した。この方法により、高い精度で局地降水予測が可能であることを示した。(図1)

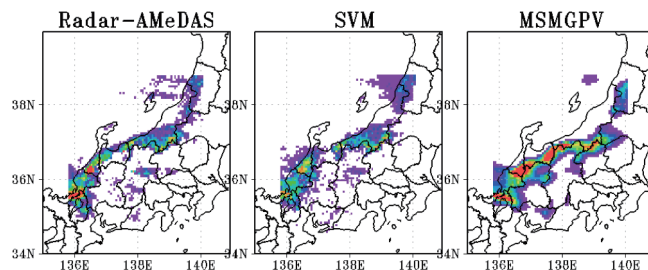


図1 2015 年 1 月 17 日12時の降水量分布. 観測(左), 機械学習(中央), 数値シミュレーション(右)。

数値シミュレーションでは、直接局地気象を予測することは不可能だが、広域での特性は現実をよく反映している。つまり、大規模スケールでの気象現象(季節風や低気圧など)に伴う天気パターンに領域での降水分布が対応しており、結果的に領域内のシミュレーションと観測のそれぞれの降水分布パターンがお互いに強く関係する。(分布特性自体を大きく異なる)(図2)。つまり、局地降水を推定するためには、天気パターンに対応し、かつ観測の領域での降水パターンとの関係性が高い降水パターンを再現可能な数値シミュレーションが不可欠となる。それには、領域内のメソスケール現象の特性を再現できる解像度を持つ数値モデル(メソスケールモデル)が必要となる[7]。解像度が粗すぎると、観測と関係性を持つ分布パターン自体を再現できない。領域の設定も重要になる。領域が狭すぎると、分布パターンを認識できず(機械学習では領域全体で晴れているのか局所的に晴れているだけなのか識別できない)、局地降水予測

に大きな誤差を生じる。つまり、シミュレーションによる降水分布が機械学習での特徴量であり、特徴量を適切に設定することが局地降水を推定する上で最も重要となる。

降水に限らず、その他の気象要素についてもメソスケール現象の影響を受けているため、この手法が降水以外の局地気象予測にも適用可能であると推測される。本研究では、この仮説を検証するため、地上風、気温、湿度、降雪などアメダスにより観測される要素について手法を適用して、評価を行った。また、手法の拡張性を調査するために、衛星データを用いた降水分布推定の有効性についても調査を実施した。

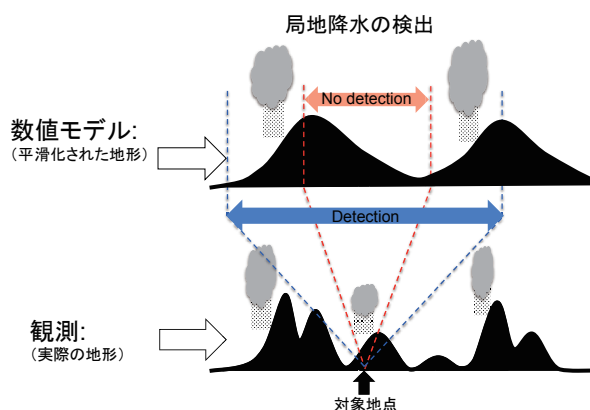


図2 数値シミュレーションと観測による降水分布の違い。対象地点での降水を推定するために必要な情報（数値シミュレーションでの降水分布）の範囲を設定。

2. 方法

機械学習にはサポートベクターマシンの回帰モデル (SVR) を用いた。SVR は統計解析ソフト R のパッケージソフトウェアである e1071 を利用した[8]。

入力には3時間ごとのアメダスと気象庁メソスケールモデルによる解析値 (MSMGPV) [7] の地上風、気温データを用いた。学習期間は、テスト期間を除く 2011 年～2018 年までの対象月の前後 3 ヶ月間（例えば、対象月が 7 月の場合、6 月～8 月）とした。各地点での観測値（地上風と気温）と 0.06 度四方（約 5km 四方）の格子に対して観測地点に最も近い格子を中心とした 30x20 格子の領域での数値シミュレーションによる地上風と気温をそれぞれ SVM により学習させた（図3）。（地上風を推定するときは、地上風の東西および南北風を特徴量とし、東西風、南北風を正解データとしてそれぞれ別に推定した。気温の推定では、特徴量、正解データ共に地上気温データを用いた。）

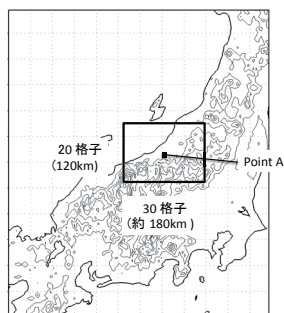


図2 機械学習で適用した入力データの範囲。数値シミュレーションデータ 30 x 20 格子と特徴量、その中心にあたる地点での観測値を正解データに設定。

3. 予測結果

SVR(SVM)による 2018 年 7 月の熊谷 (アメダス) での気温の時系 (3時間ごと) を図 4 に示す。2018 年 7 月 23 日は熊谷で歴代最高気温 41.1℃を記録している。数値シミュレーションでは全体的に過小評価傾向があるが、変動パターンはよく再現した。機械学習による推定では、過小評価傾向がほぼなくなり、定量的にも観測値とよく一致した。しかし、歴代最高気温を記録した 23 日の日最高気温の推定は難しく過小評価であった。

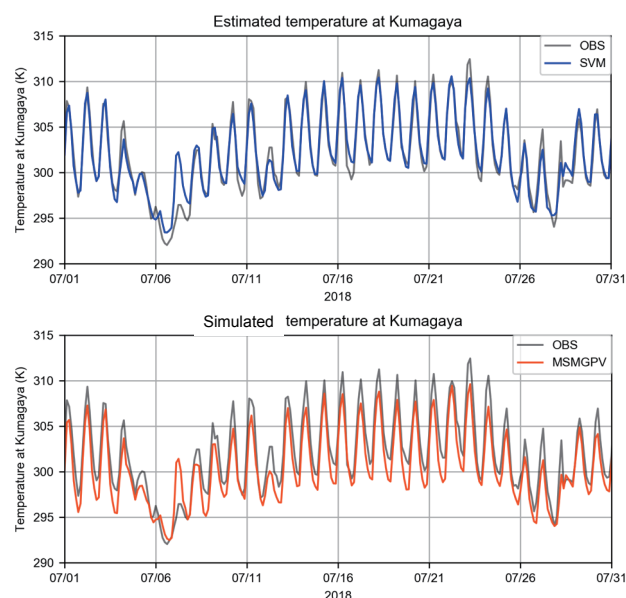


図4 2018 年 7 月の熊谷 (アメダス) での観測 (灰色)、機械学習 (青色)、数値シミュレーション近傍格子 (赤色) の気温時系列。23 日に歴代最高気温 41.1℃を記録。

図5は、富士山山頂付近 (3775m) での気温の時系列を示す。数値シミュレーションでは地形の平滑化の影響で、実測値との誤差が大きくなる。一方で、機械学習による推定値では、観測値と比較的よく一致しており、変動特性もよく似ていた。

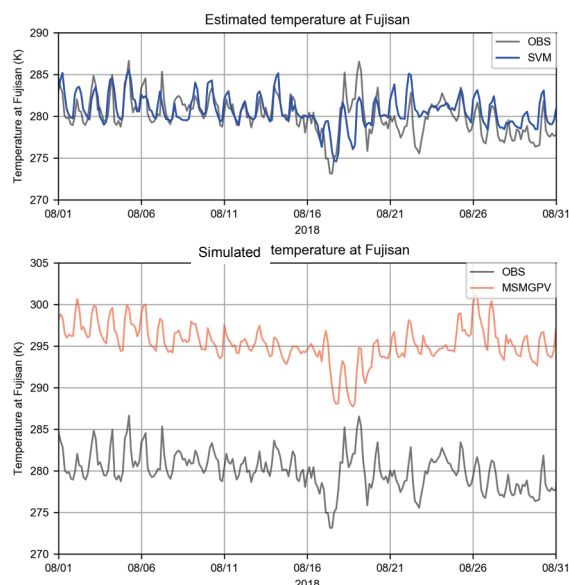


図5 富士山山頂付近での 2018 年 8 月の気温時系列。

次に、地上風のバイアスが顕著になる冬季の富山平野での地上風に注目し、機械学習で推定を行った。図6は北陸地域での観測(左)、数値シミュレーション(中央)、機械学習(右)の地上風分布(月平均値と北西季節風時)を示す。今回、機械学習による推定では富山(アメダス)での観測値を用いた。冬季の富山平野では、北西季節風時に南風がよく観測される。このメカニズムについては、実際にはよく分かっていない。北アルプスが障壁となって迂回風が関係する可能性の他、日本海上での暖かい海面水温による暖気と内陸での寒気の温度差による重力流の形成による可能性も考えられる。数値シミュレーションでは、月平均値で見ると富山平野内部で南西風が弱いながら再現されているが、北西季節風が強まるときは、ほとんど再現されない。一方で、アメダスでは、どちらの場合も南風が観測されており、顕著なバイアスが確認できる。機械学習により地上風を推定した結果、観測と変動パターンがよく一致しており、北西季節風時の南風も推定された(図7)。

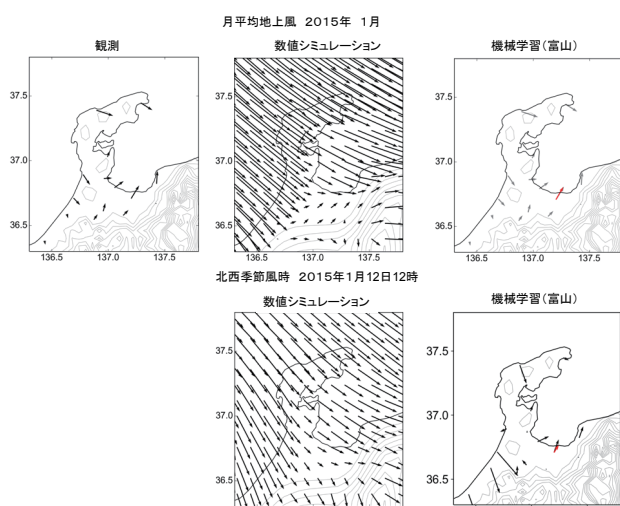


図6 北陸地域での2015年1月の月平均値と北西季節風時の地上風。赤色の矢印が富山(アメダス)での機械学習による推定値。

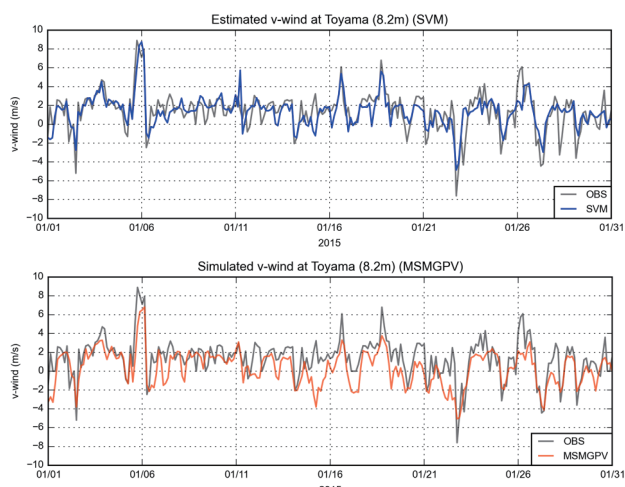


図7 2015年1月の富山(アメダス)での観測(灰色)、機械学習推定値(青色)、数値シミュレーション(赤色)の南北風の時系列。

4. 考察

数値シミュレーションでは、地形が平滑化されるため、山岳地域の気温のバイアスを避けることが難しい。本手法による推定値がバイアスを低減に成功し、観測値と変動特性がよく一致していたことは、気温についても天気パターンに対応した領域での気温分布パターン(数値シミュレーション)が観測値を推定するための情報を有していたことを意味する。気温は、低気圧や季節風などの大規模スケールの気象システムだけでなく、局地循環やフェーンなどの局地的な影響を大きく受けている。数値シミュレーションでは、領域内の細かい気温分布は異なるものの、領域での天気パターンに対応した気温特性をよく再現していることを示している。局地での気温は、天気パターンによる影響だけでなく、付近の土地利用や建物の影響も大きいと推測されるが、それらの局所的な影響も天気パターンに何らかの形で連動していると推測される。数値シミュレーションでは局所的な要因を詳細に再現するだけの性能は有しておらず局地気温の予測は難しいが、本手法を用いることによりモデル化されていないプロセスの効果を反映して予測推定することが可能になる。ただし、図4で示したように、過去に例のない事象の再現は難しく、過小評価になる。

地上風についても同様のことが言えるだろう。数値シミュレーションでは、北西季節風時に富山平野で形成する南風を精度よく再現することができない。解像度が不十分、あるいは物理プロセスに問題があると考えられる。ただし、天気パターンに対応した領域での地上風分布の特性はよく再現しており、かつ観測点での地上風パターンとも連動していることから、機械学習による推定が高い精度で実現したと推測される。

図8は昨年度発表時に示した機械学習による降水分布推定値(中央)であるが、若干過小評価傾向が見られるものの観測と同様に沿岸域に沿って降水の形成が確認できる。北陸地域では冬季に雷がよく観測される。雷の発生は、寒ブリ漁とも時期が重なることから「鯰起こし」とも呼ばれている。観測される降水帯とも何らかの関係があると考えられるため、局地降水の予測は社会的にも重要になる。一方で、数値シミュレーションでは再現が難しく、山岳で降水が多くなるモデルバイアスが顕著に見られる。本手法を用いることにより、数値シミュレーションの予報値を活用しつつ、局地気象予測が可能になるだろう。

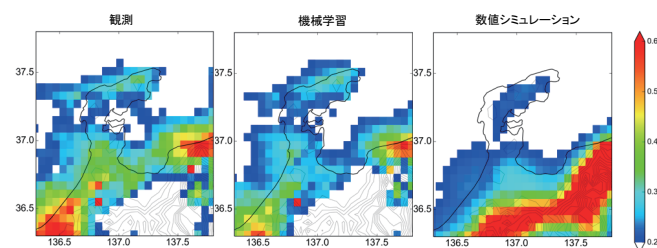


図8 2015年1月の北陸地方における観測(左)、機械学習による推定値(中央)、数値シミュレーション(右)。観測は雨量解析値[9]を使用。成果の一部は2018年度人工知能学会で発表。

5. まとめと課題

カオス現象(初期値依存性)により、気象予報ではアンサンブル手法等を用いた確率予報が不可欠である。本手法による局地気象予報においても、個々の積雲対流、メソスケール気象現

象の発達状態までは予測できないため、定量的に正確に予報することは不可能である。一方で、本手法でバイアスを低減することにより、局地気象予報(確率予報)の実現に大きく前進したと考えている。基本的に数値シミュレーションは仮想世界であり、現実世界とは大きく異なる。数値モデルは人が理解できるプロセスで構成されており、再現される現象も限定される。自然の複雑さを完全に理解して全てをモデル化することは容易ではない。当然ながら数値シミュレーションには本来は評価できない、あるいは評価してはならない結果も含まれており、大きな誤差や誤解を生じる原因になる。数値シミュレーションで再現されるもののうち、何が現実を反映したものを正しく評価する必要がある。機械学習によるアプローチは、数値シミュレーションの結果を検証し、気象研究を大きく発展させるためのポテンシャルを有している。

機械学習と数値シミュレーションによる組み合わせは、お互いの長所を活かすことにより、従来の方法では不可能であった問題の解決が可能になるだろう。特に複雑さを扱う自然科学分野にとって極めて強力な手法になると期待される。

さらに高精度の予測情報を得るには、最適な特徴量を見つけ出す必要がある。理論に基づいた手法の開発およびその検証を継続して実施することにより大きく改善できるだろう。その一方で、デープラーニング[10]による手法の発展も無視できない。人には複雑すぎて発見が難しい規則性(パターン)も多くあると想定される。デープラーニングにより未知の規則性の発見が容易になるかもしれない。人工知能技術は、従来の研究のあり方を大きく変革し発展する端緒となり得るだろう。

謝辞

本研究の一部は、統合的気候モデル高度化研究プログラム(MEXT)、統合的データ解析システム(DIAS)、宇宙航空研究開発機構(JAXA)、内閣府総合科学技術・イノベーション会議の戦略的イノベーション創造プログラム(SIP)「国家レジリエンス(防災・減災)の強化」(管理法人:防災科研)によって実施されました。

参考文献

- [1] Rummukainen, Markku. State-of-the-art with regional climate models. Wiley Interdisciplinary Reviews: Climate Change 1.1, 82-96. 2010.
- [2] Kimura R, Numerical weather prediction. Journal of Wind Engineering and Industrial Aerodynamics. 90.12-15, 1403-1414, 2002.
- [3] 山田力: 小型船舶安全運航のための観天望気を入れた気象海象の話, 関東小型船安全協会, 2007, <http://www.shoankyo.or.jp/kisyoku/pdf/khn2.pdf>.
- [4] Acharya, S., Presage biology Lessons from nature in weather forecasting. Indian Journal of Traditional Knowledge, 10, 114-124. 2011.
- [5] Met Office: Red sky at night and other weather lore, 2017, <https://www.metoffice.gov.uk/learning/learn-about-the-weather/how-weather-works/red-sky-at-night>.
- [6] 吉兼隆生, 芳村圭. 機械学習による局地気象予報の試み, 人工知能学会全国大会論文集 2018 年度人工知能学会全国大会 (第 32 回) 論文集. 一般社団法人 人工知能学会, 2018.
- [7] 気象庁: メソモデル・局地モデル. 2018. <http://www.jma.go.jp/jma/kishou/known/whitep/1-3-6.html>.
- [8] David Meyer et al.: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, 2017. <https://cran.r-project.org/web/packages/e1071/e1071.pdf>
- [9] 気象庁: 解析雨量. 2018. <http://www.jma.go.jp/jma/kishou/known/kurashi/kaiseki.html>.
- [10] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton: Deep learning. nature 521.7553, 436. 2015.