

機械学習を用いた降雨の直前予測

北海道大学 大学院理学院 宇宙理学専攻

野田 明羅

(学籍番号：20213064)

2022 年 12 月 14 日

目 次

1	イントロダクション	1
1.1	極端降雨気象	1
1.2	機械学習を用いた降雨の予測に関する研究	5
1.3	本研究の目的	9

概 要

ゲリラ豪雨や台風に伴う降雨帯、線状降水帯といった極端降雨気象の発生や降水量を正確に予測することは現代の最新の技術をもってしても難しい。これらの降雨領域の狭さや発達時間の速さが、予測の困難さの原因になっている。近年、計算機の処理速度の向上やビッグデータの登場の恩恵を受けて、機械学習技術の発展が著しい。加えて、気象学分野においても、機械学習を用いた手法が盛んに研究されている。降雨のような時空間データを学習するための機械学習モデルで最も有名なものの1つとして、Shi *et al.*[2015] によって提案された ConvLSTM が挙げられる。本論文内で降雨のレーダーエコーデータを用いて学習した ConvLSTM モデルが従来手法よりも高い精度で降雨を予測したことが示された。さらに Aifang *et al.*[2020] では降雨に加え風の情報も学習データに加えることで、予測の難しい局所的な対流が原因と考えられる降雨の分布変化を予測できる可能性を示した。しかしこれらの機械学習モデルの多くはその予測構造がブラックボックス化しており、予測結果と背景気象場との関連性を直接的に考察することが依然として難しい。入力パラメータを増やしたことでモデル内部の学習状態がどのように変化して予測が改善したのか説明ができないのである。そして Lin *et al.*[2020] らは自己注意機構 (Vaswani *et al.*[2017]) を用いて ConvLSTM を拡張した Self-Attention ConvLSTM を提案した。同論文内では、Self-Attention ConvLSTM が ConvLSTM や他の時空間予測の機械学習モデルよりも高い精度で予測できたことが示された。自己注意機構とは与えられたデータのどの部分に注目すればよいかを学習するための機構である。空間データの場合、どの領域に注目すべきかを学習する。またその注目度合を可視化できる。自己注意機構はブラックボックス化していた学習状態を明らかにできる手法として近年大きく注目を集めている。Self-Attention ConvLSTM を用いた降雨の予測の研究は未だ少ない。さらに極端降雨の特徴でもある発達領域の狭いローカライズな降雨のデータを用いた研究はまだなされていない。加えて入力のパラメータを変化させた場合に Self-Attention ConvLSTM モデル内部の学習状態がどのように変化することで精度が向上したのかを検証・考察した研究はまだない。先行研究例では、雨だけでなく他のパラメータも同時に学習させることで予測精度が改善したと報告されている。したがってローカライズな降雨において、入力パラメータの変化がモデルの学習状態や予測精度にどのような影響を与えるのか、Self-Attention ConvLSTM を用いて検証することは非常に重要である。我々の研究グループは、2017 年から P-POTEKA と呼ばれる自動気象観測装置の導入を進めてきた。豪雨やそれに伴う洪水の被害が多発しているマニラ (フィリピン) の首都圏に現在 35 個の P-POTEKA が設置されていて、降水量や気温・気圧・湿度・風速・風向を 1 分毎に観測している。これにより、発

達領域の狭く時間変化の激しい降雨現象を捉えるのに適した高いデータを得ることができるようになった。P-POTEKA のデータに対してガウス過程回帰を用いてデータを内挿することで、時間雨量に加え気温・湿度・風向・風速の時空間データを作成した。2020/04～2020/11 の期間で、300 以上の降雨イベントを含むデータセットを作成した。上記のデータを用いて、降雨と共に学習させるパラメータを変化させた異なる Self-Attention ConvLSTM モデルを用意して、一時間後（10 分間隔で 6 ステップ）までを予測した。観測地点における実測値と予測値の平均平方二乗誤差（RMSE）を未学習の降雨イベントに対して計算し予測性能を評価した。さらに学習させるパラメータの種類を変えたことで、自己注意機構の注意度合がどのように変化するか検証した。これにより背景の気象パラメータ同士の関連付けがモデル内部でどのように行われているか考察した。（結果は今後記載）

1 イン트로ダクション

1.1 極端降雨気象

日本において地上気象観測は古くから行われている。近年では自動気象データ収集システム (AMeDAS) や、C バンドレーダー・X バンドレーダーの全国配備が進んでいる。これにより気温・気圧などの気象パラメータと降水分布がほぼリアルタイムに観測されるような態勢が確立されている。しかしながらこのような最先端の気象観測網をもってしても、ゲリラ豪雨や線状降水帯、さらには台風に伴う豪雨などといった極端気象現象の予測は依然として難しい。気象庁のホームページ https://www.jma.go.jp/jma/kishou/known/yougo_hp/kousui.html では降雨の内、「著しい災害が発生した顕著な大雨現象」と豪雨としている。一方、集中豪雨を「同じような場所で数時間にわたり強く降り、100mm から数百 mm の雨量をもたらす雨」と説明している。これらのような極端降雨現象は、最先端の技術でも予測が大変難しいことで知られている。集中豪雨の中でも、稀にしか発生しないような大雨は極端豪雨・極端降雨気象などと呼ばれる。近年における状況を Masaki[2020] の発表をもとにまとめる。気象庁における統計データ [1] では、1 時間降水量が 50、80、200、400mm 以上の降雨の年間発生件数の、1975 年以降の変化傾向はいずれにおいても増加トレンドにあることが示されている。(図 1)

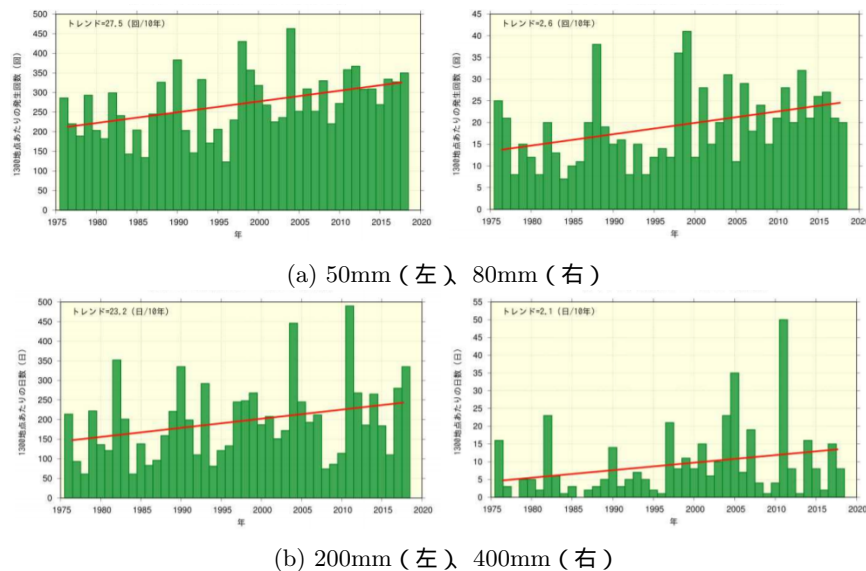


図 1: 1976-2018 年期間の全国のアメダスの 1 時間降水量 50mm、80mm (a)、200mm、400mm (b) 以上の年間発生回数の変化。棒グラフは全国のアメダスによる観測値を 1300 地点あたりに換算した値、直線は長期変化傾向を表す。

Fumikai *et al.*[2006] は、国内 51 地点の 104 年間 (1901 ~ 2004 年) の日降

水量の資料から日降水量の降水階級、100mm 以上の日数、年最大降水量、年間の上位 100 事例の発生頻度等の様々な大雨の尺度を用いて経年変化を調べた。いずれの尺度においても大雨の数は増加傾向にあることがわかる。

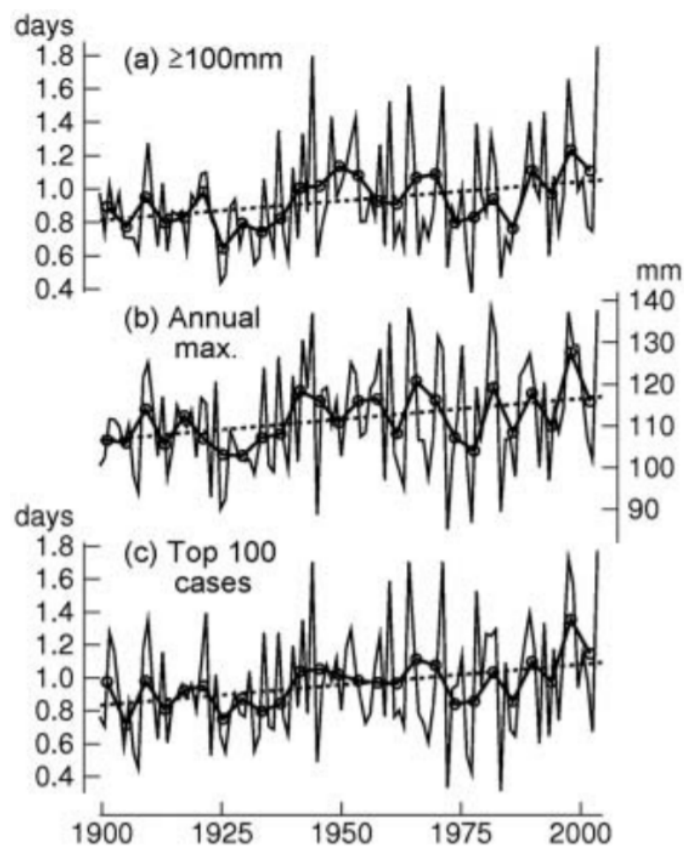


図 2: 1901-2004 年期間の極端豪雨の経年変化：100mm 以上の日数、年最大降水量、累年の上位 100 事例の発生頻度の時系列を示す。横軸は年。

日本における極端豪雨の事例のうち、甚大な被害を起こした現象については気象庁が名称を定めている。(表 1)

表 1: 平成以降に気象庁が定めた気象現象からの豪雨事例（気象庁ホームページ: https://www.jma.go.jp/jma/kishou/known/meishou/meishou_ichiran.html より改変）

名称	期間・現象等	「地域独自の名称等」、主な被害
平成 16 年 7 月新潟・福島豪雨	平成 16 年 7 月 17 日～13 日	「7.13 新潟豪雨」。
平成 16 年 7 月福井豪雨	平成 16 年 7 月 17 日～18 日	福井県の浸水害・土砂災害等。
平成 18 年 7 月豪雨	平成 18 年 7 月 15 日～24 日	「平成 18 年 7 月鹿児島県北部豪雨」。 諏訪湖（長野県）周辺の土砂災害、浸水害、天竜川（長野県）の氾濫等。
平成 20 年 8 月末豪雨	平成 20 年 8 月 26 日～31 日	名古屋市・岡崎市（愛知県）の浸水被害等。
平成 21 年 7 月中国・九州北部豪雨	平成 21 年 7 月 19 日～26 日	「平成 21 年 7 月 21 日豪雨」、「山口豪雨災害」
平成 23 年 7 月新潟・福島豪雨	平成 23 年 7 月 27 日～30 日	五十嵐川・阿賀野川（新潟県）の氾濫等。
平成 24 年 7 月九州北部豪雨	平成 24 年 7 月 11 日～14 日	「熊本広域大水害」、「7.12 竹田市豪雨災害」。八女市（福岡県）、竹田市（大分県）の土砂災害・洪水被害、矢部川（福岡県）の氾濫等。
平成 26 年 8 月豪雨	平成 26 年 7 月 30 日～8 月 26 日	「広島豪雨災害」、「8.20 土砂災害」、「2014 年 8 月広島大規模土砂災害」、「丹羽市豪雨災害」、「2014 高知豪雨」。
平成 27 年 9 月関東・東北豪雨	平成 27 年 9 月 9 日～11 日	「鬼怒川水害」。鬼怒川（茨城県）・渋井川（宮城県）の氾濫等。
平成 29 年 7 月九州北部豪雨	平成 29 年 7 月 5 日～6 日	朝倉市・東峰村（福岡県）・日田市（大分県）の洪水・土砂災害等。
平成 30 年 7 月豪雨	平成 30 年 6 月 28 日～7 月 8 日	「西日本豪雨」。広島県・愛媛県の土砂災害、倉敷市真備町（岡山県）の洪水害など、広域的な被害。
令和元年房総半島台風	令和元年 9 月（台風 15 号）	房総半島を中心とした各地で暴風等による被害。台風「ファクサイ」。
令和元年東日本台風	令和元年 10 月（台風 19 号）	東日本の広い範囲における記録的な大雨により大河川を含む多数の河川氾濫等による被害。台風「ハギビス」
令和 2 年 7 月豪雨	令和 2 年 7 月 3 日～31 日	「熊本豪雨」。西日本から東日本の広範囲にわたる長期間の大雨。球磨川（熊本県）などの河川氾濫や土砂災害による被害。

平成 16 年以降、ほぼ毎年のように極端豪雨が発生しており近年極端豪雨が増加していることがわかる。特に、令和 2 年 7 月豪雨では大規模な線状降水

帯が発生し甚大な被害を与えたことは記憶に新しい。近年この「線状降水帯」が原因の極端豪雨の発生が増加しており、そのメカニズムや予測研究が盛んに行われている。このような極端豪雨が増加している原因の1つとして、地球温暖化があげられている。気象庁が毎年発表している業務評価レポート [2] では、該当年の降水短時間予報の精度が公表されている。2019 年度の予測精度は 0.52 であった。降水短時間予報は、1 時間 10mm 以上を超えるやや強い降水に対する降水量の予測である。降水短時間予報が 2 時間後から 3 時間後までの 1 時間に降ると予測した 1km 格子ごとの降水量を、5km 格子ごとの値に平均した値を予測値として、同様に 1km 格子ごとの解析雨量を 5km 格子ごとの値に平均した値を実況値とした。これらの予測値と実況値の合計が 20mm 以上の場合に、予測値と実況値のうち大きな方を分母として比を計算、5km 格子ごとに計算した比を日本の陸上付近で 1 年にわたって平均した値が精度となる。このことから、今もなお、最新の観測・予測技術をもってしても、極端豪雨を含めた強い雨を正確に予測することは難しいことが示されている。

極端豪雨の予測が難しい主な原因として、Sato[2020] の中では以下のように説明されている。

“(前略)長時間の時間スケールにおいては、エネルギー収支によって降水量は制約されており、水蒸気が大気中で凝結する際に放出する潜熱と大気中の放射冷却のバランスによって説明される。(中略)しかしこのようなバランスは、長時間の広領域で平均した関係であるため、個々の降水イベント、とくに極端豪雨に関してはこのようなバランスは成り立たない。(中略)集中豪雨は、一般に数十 km スケールのメソ対流システム (Meso-scale Convective System : MCS) によってもたらされる。これらは台風や温帯低気圧、梅雨の前線といった千 km スケールの相関規模の擾乱の内部で発生し、さらに、より大きな 1 万 km スケールの大気循環 (モンスーン循環やハドレー循環等) の大規模場の変化がこれらの総観規模の擾乱の変調をもたらす。” [Sato, 2020, p.144]

すなわち、極端豪雨はその空間的かつ時間的なスケールが小さいために予測を難しくしているといわれている。気象観測の先進国である日本であっても予測は困難であり、気象観測インフラがより脆弱な東南アジア域では、極端気象・極端豪雨の予測はさらに困難な問題となる。加えて毎年のように甚大な災害を引き起こしており、これからは社会インフラや人間活動を守る対策として、さらに SDGs(Sustainable Development Goals) の観点からも、極端気象の予測は急務の課題となっている。

1.2 機械学習を用いた降雨の予測に関する研究

近年、コンピューターの処理速度の向上やデータのデジタル化による扱えるデータの巨大化などの要因に後押しされ、機械学習の分野の発展は大変著しい。この機械学習の技術を用いて上記の極端降雨現象の予測に対する新しいアプローチの模索が行われている。

降雨のような時空間データを学習させるための機械学習モデルとして Shi *et al.*[2015] によって ConvLSTM(Convolutional Long-Short Term Memory) が発表された。ConvLSTM は空間的な特徴を学習させるための機械学習モデルである CNN(Convolutional Neural Network) と、時間的な変化を学習させるためのモデルである LSTM(Long-Short Term Memory) を組み合わせた機械学習モデルである。具体的な計算式は以下の通りである。

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \odot C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \odot C_{t-1} + b_f) \quad (2)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \odot C_t + b_o) \quad (4)$$

$$H_t = o_t \odot \tanh(C_t) \quad (5)$$

さらに ConvLSTM の内部構造の概念図以下に示す。

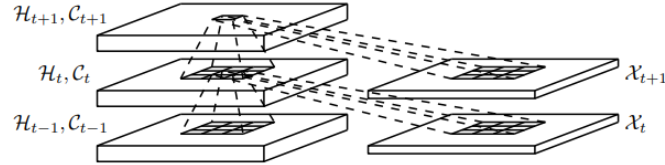


Figure 2: Inner structure of ConvLSTM

図 3: ConvLSTM の内部構造の概念図。Shi *et al.*[2015] Figure2 より引用。

図 3 の $X_t \cdot C_t \cdot H_t$ はそれぞれ入力・セル状態・隠れ層の出力を表しており、上記の数式における $X_t \cdot C_t \cdot H_t$ と対応している。入力は各時刻における入力される空間データを示す。セル状態は過去の状態に関する情報を蓄積する働きをする。このセル状態が時系列データの効率的な学習を可能としている。計算式 (1) の i_t は入力ゲートと呼ばれ、入力 X_t と 1 つ前の時刻の出力 H_{t-1} に対して畳み込み演算 (*) が施された結果と 1 つ前の時刻のセル状態 C_{t-1} が重み付けされた結果の和をシグモイド関数を用いて正規化している。さらに計算式 (2) の f_t は忘却ゲートと呼ばれ入力ゲートと同様の計算が

行われるが、重み W が異なっている。計算式 (3) の右辺第 1 項は先ほど求めた忘却ゲートを用いて過去のセル状態を調整している。右辺第 2 項は入力ゲートを用いて入力 X_t と 1 つ前の時刻の出力 H_{t-1} の和を調整している。計算式 (3) は最終的にこれらを用いてセル状態を更新していることがわかる。⊙ はアダマール積を示す。計算式 (4) の o_t は出力ゲートと呼ばれており、計算式 (5) で最終的な出力 H_t を求める際に更新されたセル状態 C_t を調整する働きをしている。

Shi *et al.*[2020] の論文内では降雨のレーダーエコー画像を入力として ConvLSTM モデルを学習させ一定時間後の降雨の分布と強度の予測を行った。その結果を図 2 に示す。

Table 2: Comparison of the average scores of different models over 15 prediction steps.

Model	Rainfall-MSE	CSI	FAR	POD	Correlation
ConvLSTM(3x3)-3x3-64-3x3-64	1.420	0.577	0.195	0.660	0.908
Rover1	1.712	0.516	0.308	0.636	0.843
Rover2	1.684	0.522	0.301	0.642	0.850
Rover3	1.685	0.522	0.301	0.642	0.849
FC-LSTM-2000-2000	1.865	0.286	0.335	0.351	0.774

図 4: Shi *et al.*[2015] Table2 より引用。

図 2 の通り、従来手法の一つである ROVER(Real-time Optical flow by Variational methods for Echoes of Radar) を比較した結果、ConvLSTM モデルは従来手法よりも高い精度で降雨の分布と強度を予測できた。ROVER1・2・3 はモデルの初期化条件を 3 つのパターンに分けた場合のそれぞれの ROVER モデルである。FC-LSTM(fully connected LSTM) は LSTM のみを用いて予測を行うモデルである。Rainfall-MSE は予測値と実測値との Mean Squared Error(平均 2 乗誤差)、CSI・FAR・POD はそれぞれ Critical Success Index・False Alarm Ratio・Probability of Detection を示す。0.5mm/h 以上の降雨が観測 OR 予測された場合に 1、それ以外の場合を 0 としそれぞれ CSI・FAR・POD を計算した。

この論文が発表されて以来、ConvLSTM を用いた降雨予測の研究が盛んに行われた。Su *et al.*[2020] の研究では ConvLSTM に降雨のレーダーエコー画像データだけでなく東西・南北風成分を同時に学習させることで予測のより困難な局所的な対流が原因の降雨の強度や分布の変化を予測できる可能性が示された。本論文の学習プロセスの概要図を図 4 に示す。

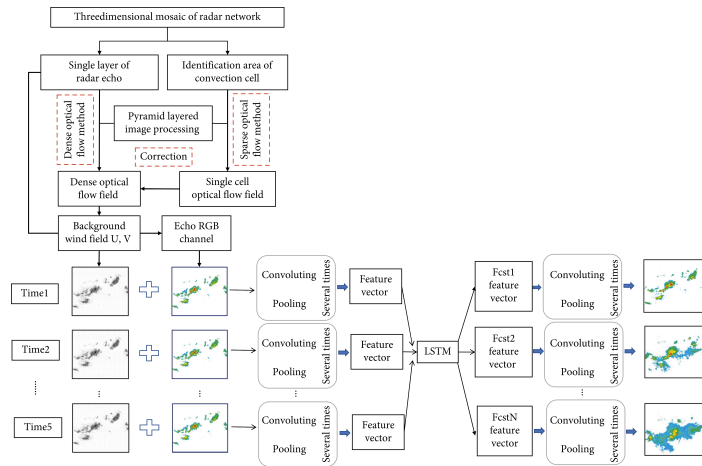


図 5: Su *et al.*[2020] Figure2 より引用。

以下の図 5 は Su *et al.*[2020] の予測結果の 1 つである。

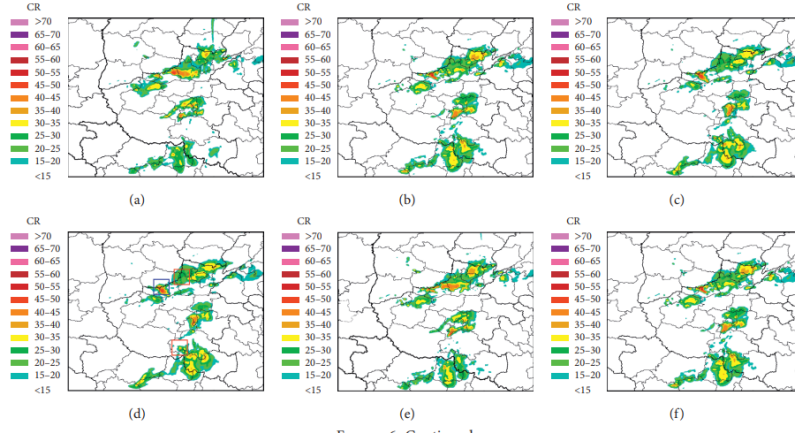


FIGURE 6: Continued.

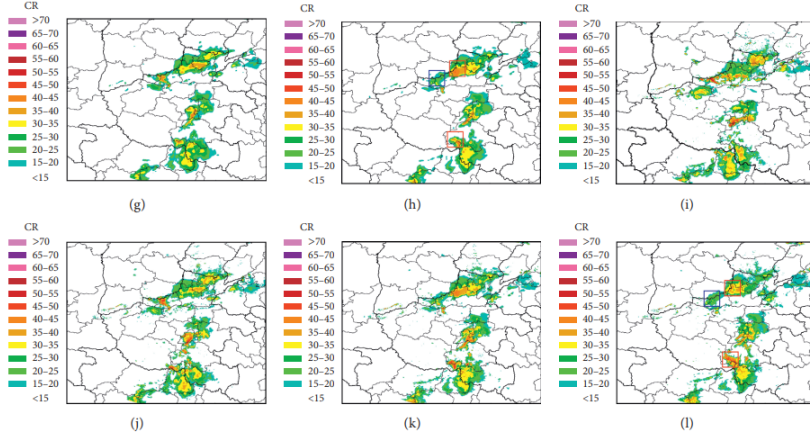


図 6: Su *et al.*[2020] Figure6 より引用。従来手法であるオプティカルフローでの 30 分・60 分・90 分・120 分後の予測がそれぞれ (a)・(b)・(c)・(d)。ConvLSTM での 30 分・60 分・90 分・120 分後の予測がそれぞれ (e)・(f)・(g)・(h)。降雨の 30 分・60 分・90 分・120 分後の実測画像がそれぞれ (i)・(j)・(k)・(l)。

図 5 の従来手法と ConvLSTM 予測結果 (d), (h) と降雨の実測画像 (l) には赤枠と青枠でそれぞれ ConvLSTM で予測でき、従来手法で予測できなかった部分とその逆の部分が示されている。このケースのみならず多くの場合において雨だけでなく風と共に学習させた ConvLSTM の方が降雨が強い部分をより多く予測できた結果となった。降雨とそれに関連する気象パラメータと一緒に学習させることで、従来手法では予測が難しかった降雨が予測できる可能性が示された。

さらに、ConvLSTM の発表をベースにして様々な時空間データを学習させ

るための機械学習モデルが提案された。PredRNN (Wang *et al.*[2017b]) は水平方向だけでなく鉛直方向成分を持つ時空間データを扱うために新しいセル状態を導入した。これによって空間のダイナミクスに関する情報をより効率的に学習可能となった。さらに PredRNN++ (Wang *et al.*[2018b]) では

そして Self-Attention ConvLSTM (Lin *et al.*[2020]) が発表され、ConvLSTM や他の派生モデルよりも高いベンチマークスコアを記録した。このモデルでは Self-Attention 機構を導入したことで Convolution 処理では扱えなかったデータ全域に渡る特徴を捉えることができるようになった。さらに過去の Self-Attention 機構の状態を新しく追加したメモリーモジュールで保管・更新することで時系列に適したモデルとして拡張した。

1.3 本研究の目的

参考文献

- [1] Shi, X.; Chen, Z.; Wang, H.; Yeung, D.-Y.; Wong, W.-K.; and Woo, W.-c. 2015. Convolutional lstm network: A machine learning approach for precipitation nowcasting. NIPS 2015, 802–810.
- [2] Lin Z.; Li M.; Zheng Z.;, Cheng Y.; and Yuan C. 2020. Self-Attention ConvLSTM for Spatiotemporal Prediction. Association for the Advancement of Artificial Intelligence
- [3] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need., NIPS, 2017.
- [4] Su A.; Li H.; Cui L.; and Chen Y. 2020. A Convection Nowcasting Method Based on Machine Learning. Hindawi.
- [5] Hubel D.H. and Wiesel T.N. 1962. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. The Journal of Physiology.
- [6] Hochreiter S. and Schmidhuber J. 1997. Long Short-Term Memory. Neural Computation.
- [7] Ko C.M.; Jeong Y.Y.; Lee Y.M.; and Kim B.S. 2020. The Development of a Quantitative Precipitation Forecast Correction Technique Based on Machine Learning for Hydrological Applications. MDPI.
- [8] 気象庁. 2018. 気候変動監視レポート 2018 世界と日本の気候変動および温室効果ガスとオゾン層等の状況. 気象庁.

- [9] 気象庁. 2020. 気象庁業務評価レポート（令和 2(2020) 年度版）. 気象庁.
- [10] 佐藤正樹;. 2020. 近年における降雨状況の実態：極端豪雨は増えているか 水環境学会誌 第 43 巻（A）第 5 号 pp.142 147. 公益社団法人 日本水環境学会.
- [11] Fumikai F.; Nobuo Y.; and Kenji K. 2006. Long-Term Changes of Heavy Precipitation and Dry Weather in Japan(1901-2004). Meteorological Society of Japan.