

機械学習を用いた高精度地下水位予測モデルの開発 ～これまでにない汎用性の高い地下水位予測～

海城高等学校

1. はじめに

地下水は水質が良好で水温が安定しているため、古くから人々の生活を支えてきた。しかし、高度経済成長期には地下水利用の増大とともに地下水位が低下し、地盤沈下や湧水の枯渇、塩水化など様々な問題を引き起こした。地下水は生活になくてはならない重要な資源であるが、同時に見えない資源でもある。

湧水が地下水によって形成された貴重な水環境であるとともに、地域住民に癒しを与え、コミュニティーの交流の場となっている。湧水は地域にとって絶やしてはいけない存在なのである。しかし、不浸透域の拡大などに起因する地下水位の低下により湧水はその数を大幅に減らしており、東京都の調査では過去10年で68カ所の湧水が姿を消した。

地下水位の低下による湧水の減少の問題に対処するためには予測をすることが非常に重要である。予測手法としてタンクモデルやランダムフォレストを用いた手法が提案されてきた。菅原(1972)は河川流量を予測するために工学モデルであるタンクモデルを使用した。このモデルは地質構造をタンクと見立て、河川流域における水の流量を予測するモデルであり、地形にかなった合理的なモデルである。しかしながら、タンクモデルの作成には入手困難である様々な項目のデータや専門家の知識を必要とする。また、長野ら(2017)は手取川扇状地域において機械学習の手法であるランダムフォレストを用いて予測を行った。タンクモデルに比べ、使用しているデータ項目は少ないが、適用が2つの井戸のみと限定的であり、また精度も高くない。さらに、これらのモデルは前日までのデータから翌日の地下水位しか予測できないため、実用上このモデルを適用することによる効果は小さい。

そこで本研究では上記の問題を解決するために、機械学習の手法の1つであるTransformerを用いたモデルを作成した。そして、従来のモデルとは違いモデルを直近の観測データを使用せずに予測できるようにする

ことで、実用面においての課題も解決した。さらに機械学習モデルは入力に対して非常に柔軟で、他の地域で適用する際にもパラメーターを1つ変更するだけで学習が可能であり、専門家の知識も必要としない。

ここで本研究の成果及び地下水位予測技術への貢献についてまとめる。

- 1) 提案したモデルを福井県大野市と石川県手取川扇状地域において適用し、実験において予測の精度が非常に高いことを示した。
- 2) 地下水位の観測データと気象庁が公開するオープンデータのみを予測に用いることで、より容易にモデルの作成をすることができることを示した。
- 3) 直近の観測データを使用せずに予測できるモデルを使用することで、開発したモデルが実用上非常に効果的であることを示した。
- 4) モデルが入力データのどの部分を重視しているかを数値化することで、モデルの予測プロセスを可視化した。
- 5) 機械学習を自然科学の分野に適用した。実験の結果、今までの工学モデルよりも非常に扱いやすいことがわかった。
- 6) 本研究で得られた結果は開発したモデルが農業・工業分野や地下水保全の一助となりうることを示した。

2. 使用データ

本研究では石川県手取川扇状地域及び福井県大野市の気象データ、地下水位の観測データを用いた。各種データは予測モデルが地下水位変動の傾向を掴むために使用する「学習用データ」と、学習を行った複数のモデルから最終的な予測モデルの選択に用いる「検証用データ」。そして選択したモデルを評価する「テスト用データ」に分けた。また、予測結果において示す予測、評価指標は全て「テスト用データ」における性能評価である。各地点のデータの期間や気象データの詳細については表1にまとめた。

表1 データセットの詳細

地域	学習期間	地点	検証期間	テスト期間	気象台	気象データ項目
石川県手取川扇状地	1965-2015(51年)	末広	2016-2017 (2年間)	2018-2019 (2年間)	金沢	日平均気温、降水量の日合計 1時間降水量の日最大、日照時間 日最深積雪、降雪量の日合計 (6項目)
	1973-2015(43年)	北市				
	1974-2015 (42年)	安吉,太平寺,井関,福増,赤井浅 赤井深,千代野西浅,千代野西深				
	1997-2015 (19年)	鞍月浅,鞍月中,鞍月深				
	1976-1992 (17年)	藤江浅,藤江中,藤江深				
福井県大野市	1993-2016 (24年)	下五条,猪島,木本原上,木本原中,木本原下	1993-1994 (2年間)	1995-1996 (2年間)	大野	
	2006-2016 (11年)	中据	2017-2018 (2年間)	2019-2020 (2年間)		
	2009-2016 (8年)	上据				

2.1 気象データ

気象データは気象庁が公開しているオープンデータを用いた。手取川扇状地域、大野市のどちらにおいても日平均気温、降水量の日合計、1時間降水量の日最大、日照時間、日最深積雪、降雪量の日合計の6項目を使用した。

2.2 地下水位データ

石川県手取川扇状地域の地下水位データは石川県生活環境部環境政策課より提供して頂き、福井県大野市のデータは大野市くらし環境部水環境・水循環課より提供していただいた。学習期間は観測期間に応じて地点により異なる。

3. 適用地域

次に各観測地域の地理的特徴についてまとめる。

手取川は白山を源流とし、山麓部から日本海との間に過去の手取川によって運搬された砂礫で構成された手取川扇状地を形成している。帯水層は3つに大別され、地下水は工業や農業等幅広く用いられ、冬季には地下水を用いた消雪機器も稼働する。石川県の工業、農業、商業などの諸産業が立地し、県の中心的基盤をなす地域である。

大野盆地は九頭竜川上流部に位置し、九頭竜川、真名川、清滝川、赤根川の一級河川が南から北へと流れている。帯水層は5つに分類され全体的に南から北に向かって流れており、木本扇状地地下水系と、真名川が深く関与する真名川地下水系に2分され、両者は互いに影響を及ぼし合っている。地下水は工業や農業に加え、水道水源としても用いられている。

4. Architecture

本研究では私たちはNext Day PredictionとDynamic Predictionの予測方法を用いた二つの種類のモデルを提案する。最初にモデルを構成する各種Moduleについて説明していく。またモデルの大まかな

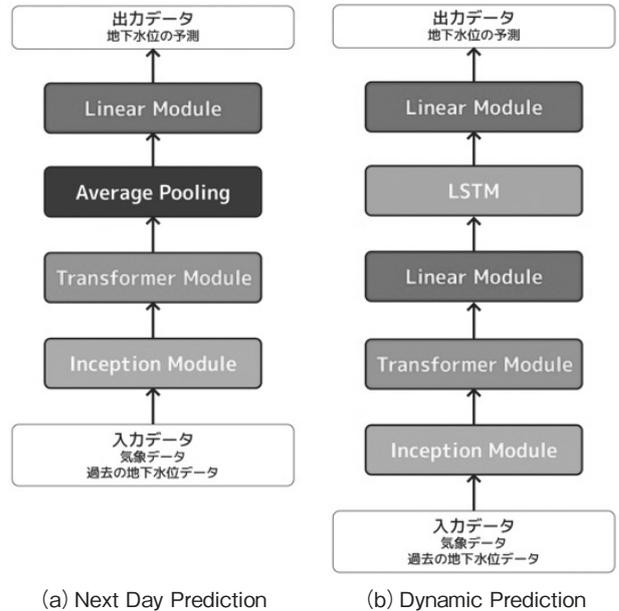


図1 モデルの構造

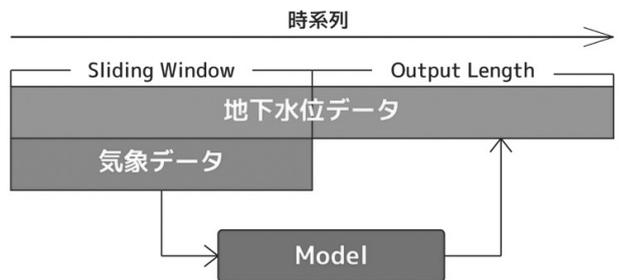


図2 予測方法

構造、予測方法については図1, 2にまとめた。

4.1 Module

- 1) Inception Module このModuleでは入力データに対して4つの異なるサイズのフィルター (1, 3, 5, 7) をかけることで、効率よくデータに内在するさまざまな依存関係を抽出する。フィルターはそれぞれ128枚ずつ用いた。
- 2) Transformer Module Inception Moduleによって処理されたデータはTransformer Moduleに受け渡される。ここではデータにおける時系列的特徴を抽出する。Transformerは本来機械翻訳のモデルとして提案されたが、時系列データにおいて

高い精度を発揮することがわかっている。

- 3) Linear Module Linear Moduleでは入力データに対して、重みをかけ、バイアスを加える。本モデルにおいてこのModuleは主にデータの次元を変える際に用いる。
- 4) Average Pooling 本Moduleでは行列における特定の次元を平均することでデータの次元を削減する。
- 5) LSTM これは複数の日数の予測を同時に予測するためModuleである。尚、このモデルでは予測日数の次元方向に関して両方の方向に隠れ特徴量を伝播させた。

4.2 予測方法

予測方法において、予測に用いる過去の学習期間の日数をsliding window,同時に予測する日数をoutput lengthとする。本研究では、翌日の地下水位を予測するNext Day Predictionと、数日後の地下水位を予測するDynamic Predictionの2つのモデルを提案する。

1) Next Day Prediction

この予測方法ではsliding window分の気象データと観測データを用いて次の日の地下水位データを予測する。最も基本的であり、長野らも同じような予測方法を用いた。しかしこの予測手法は実用上、効果が小さいという課題がある。

2) Dynamic Prediction

実用上の課題を解決するために、Dynamic Predictionを提案する。この予測方法はLSTM Moduleを用いることで、共通の入力で、将来の地下水位を複数日同時に予測することができる。またNext Day Predictionはこの予測方法においてoutput lengthを1とした場合の予測方法であると捉えることができる。

4.3 予測過程の可視化

本研究ではモデルがどのデータに注目しているかを可視化するために、Grad-CAMという手法を用いてヒートマップ(可視化グラフ)を作成した。この手法は誤差逆伝播法を用いて計算された勾配の値がそのデータの重要性を表すという仮定のもと提案された手法である。勾配の値はデータの標準偏差に大きく影響を与えるため、データの項目ごとに標準化をおこなった。絶対値が大きいほど他と比べて重要度が高いまたは低いことを示しており、値と予測に正の相関がある場合、値が相対的に大きいところは他よりも重視している

部分で小さい部分は重視していない部分を示す。また負の相関の場合、その逆となる。

5. データ処理方法

ここでは学習において行った処理や設定について説明する。

まず、学習するにあたって安定性と地点間で、同じスケールで性能を比較するため、入力と出力の地下水位データは全て学習データにおける平均と標準偏差で標準化した。

また、モデルの性能評価として、MSE (Mean Squared Error)とR2値を用いた。

MSEは実測と予測の誤差を表す指標で小さいほど良い性能であるといえる。MSEはデータの標準偏差によって変化してしまうが、今回、予測対象は全て標準化しているので全て同じスケールである。R2値は回帰タスクにおいて、予測の精度がどれほど良いかを示す値である。これはデータの標準偏差に影響を受けない。1以下の全ての実数を取りうるが1に近いほど性能が良いことを表す。

モデルの構造の部分でsliding windowとoutput lengthの二つを示したが今回の実験では全てにおいてsliding windowを28、output lengthを7とした。

6. 結果

ここでは各地点におけるモデルの予測結果と可視化の結果について示す。

Dynamic Predictionにおいては予測結果から[1, 2, ..., output_length]のそれぞれについてMSEとR2値を計算した(式(1), (2))。表においては[3, 5, 7]のみを記載した。

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (\hat{y}_i: Prediction, \bar{y}: Mean, y_i: Measurement) \quad (1)$$

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\hat{y}_i: Prediction, y_i: Measurement, N: The Number of Data) \quad (2)$$

6.1 石川県手取川地域

最初に先行研究である長野ら(2017)と比較するために、データの項目以外は共通で学習を行い精度比較

表2 先行研究との比較

モデル	安吉	北市
長野ら(2017年)	0.0063	0.0032
本研究	0.0034	0.0023

した(表2)。先行研究ではデータを標準化していないため、本研究でもこの部分のみ標準化しなかった。それによりデータの分散が非常に小さいため、R2値の使用は適切でないため、比較においては用いなかった(式(1))。MSEで比較すると実験の結果本研究の方がランダムフォレストを用いた先行研究に比べて精度が高いことが示された。

次に全ての地点において、予測を行なった(表3)。

Next Day Predictionにおいてはほとんど地点でR2値が0.8を上回る結果となり非常に高精度で予測できていることがわかる。またグラフでも非常に高い精度で予測できている。しかし、学習データの短い一部の地点におい

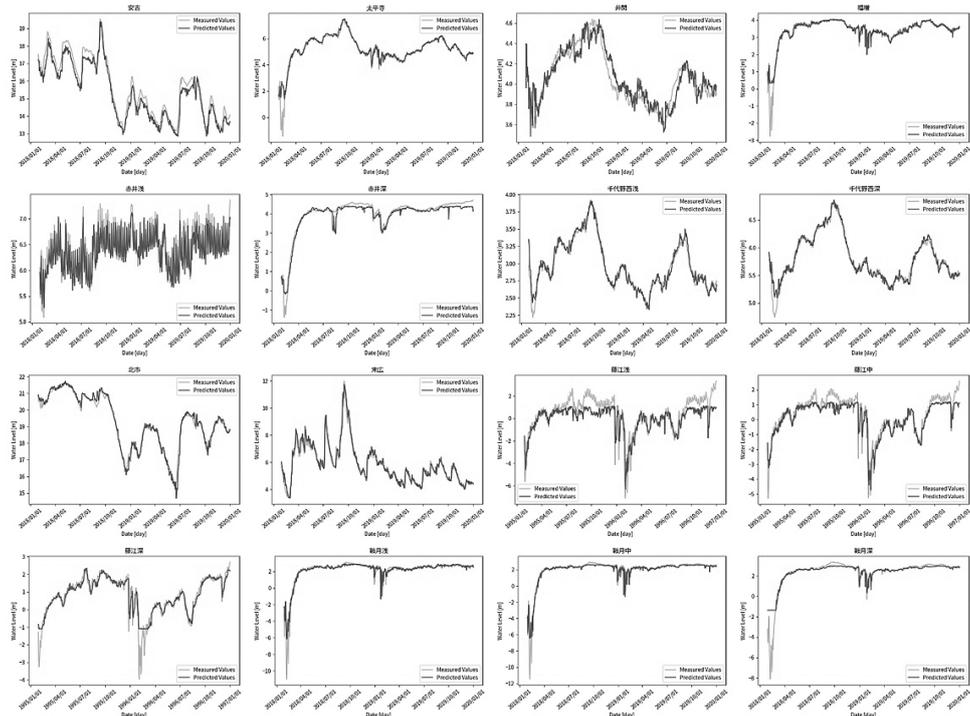
ては他と比べて精度が低くなる傾向が見られる(グラフ1)。

Dynamic Predictionにおける結果は地点によって精度が大きく変わる。Next Day Predictionと同様に、学習データが少ない地点ほど精度が下がる傾向がある。またグラフにおいても大きな変動を予測できていない地点がある。しかしそれを除けばどの地点においても、実際の実用として用いることができるほどの性能であることがわかる(グラフ2)。

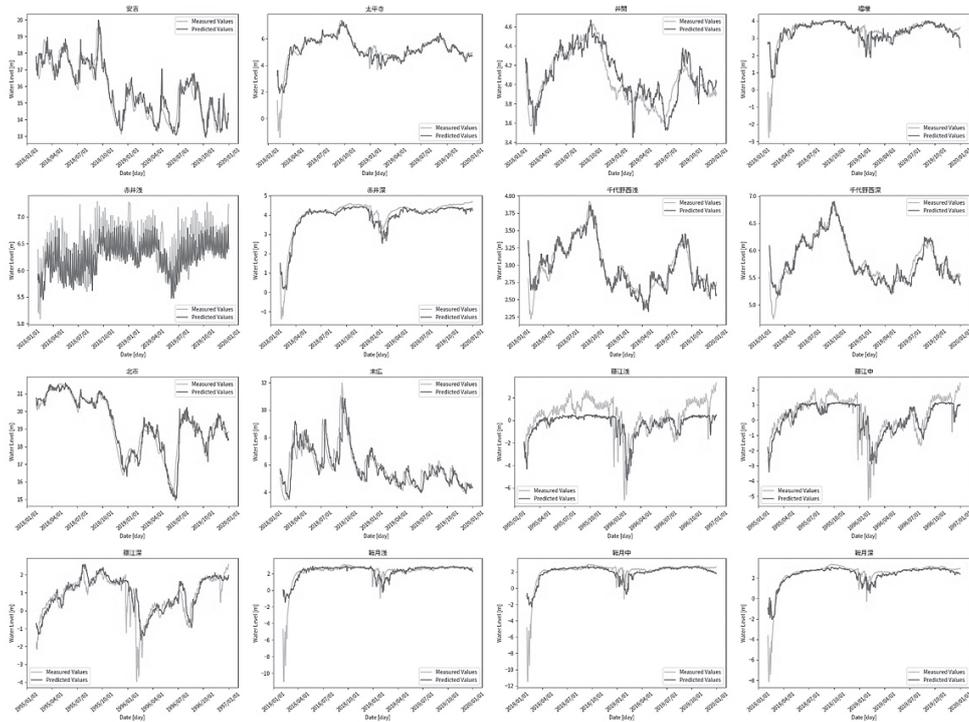
モデルが重視している部分を可視化したヒートマップ(グラフ5)ではどの項目においても予測する日に近いほど値の絶対値が大きい。つまりこの地域においては直近の気象状況がすぐに地下水位に影響することを示

表3 石川県手取川扇状地域 予測スコア

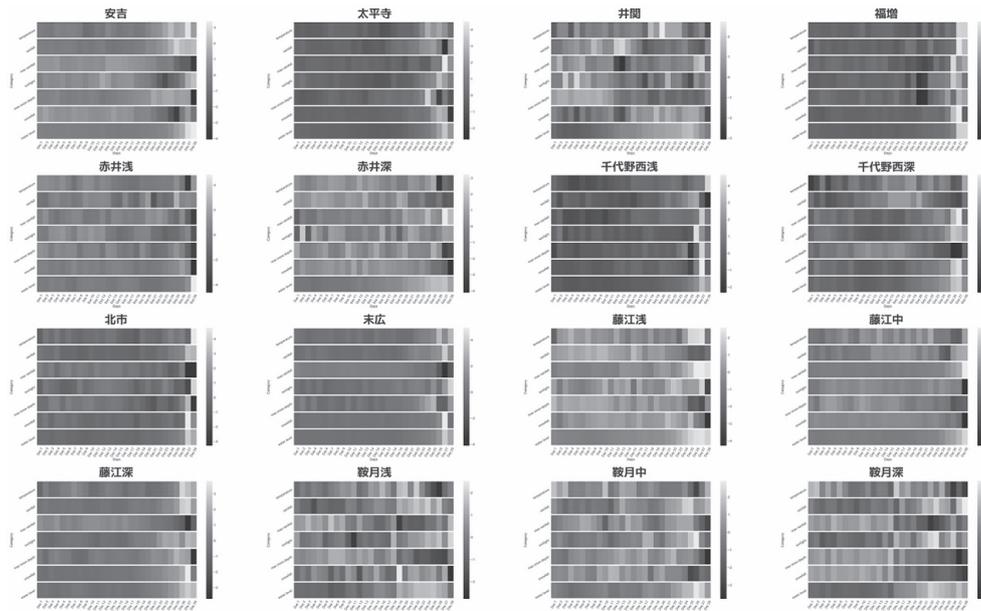
地点	Next Day Prediction		Dynamic Prediction					
			Day3		Day5		Day7	
	MSE	R2	MSE	R2	MSE	R2	MSE	R2
安吉	0.031	0.939	0.022	0.957	0.039	0.925	0.065	0.874
太平寺	0.066	0.887	0.078	0.865	0.098	0.827	0.115	0.793
井関	0.002	0.843	0.003	0.800	0.003	0.764	0.004	0.734
福増	0.112	0.875	0.237	0.729	0.314	0.628	0.428	0.484
赤井浅	0.086	0.855	0.279	0.520	0.348	0.395	0.395	0.323
赤井深	0.052	0.944	0.059	0.934	0.080	0.907	0.111	0.867
千代野西浅	0.005	0.963	0.011	0.922	0.014	0.901	0.018	0.868
千代野西深	0.015	0.956	0.026	0.923	0.032	0.907	0.038	0.887
北市	0.024	0.976	0.032	0.967	0.055	0.946	0.088	0.913
末広	0.019	0.954	0.066	0.842	0.098	0.764	0.137	0.671
藤江浅	0.165	0.694	0.249	0.521	0.266	0.470	0.381	0.228
藤江中	0.137	0.848	0.214	0.755	0.260	0.687	0.311	0.617
藤江深	0.172	0.859	0.335	0.720	0.392	0.661	0.441	0.608
鞍月浅	0.162	0.898	0.680	0.554	0.709	0.511	0.723	0.492
鞍月中	0.132	0.922	0.554	0.660	0.624	0.597	0.674	0.557
鞍月深	0.496	0.770	0.496	0.762	0.569	0.713	0.585	0.697



グラフ1 石川県手取川扇状地域 予測グラフ [Next Day Prediction]



グラフ2 石川県手取川扇状地域 予測グラフ [Dynamic Prediction] (全てDay7の予測である)



グラフ5 石川県手取川扇状地域 可視化グラフ [Dynamic Prediction]
(項目は上から日平均気温、降水量の日合計、1時間降水量の日最大、日照時間、日最深積雪、降雪量の日合計、地下水位)

している。

6.2 福井県大野市

予測結果は表4の通りである。

Next Day Predictionにおいては全ての地点においてMSEが0.03以下かつR2値が0.96以上という非常に高い精度を示した(グラフ3)。

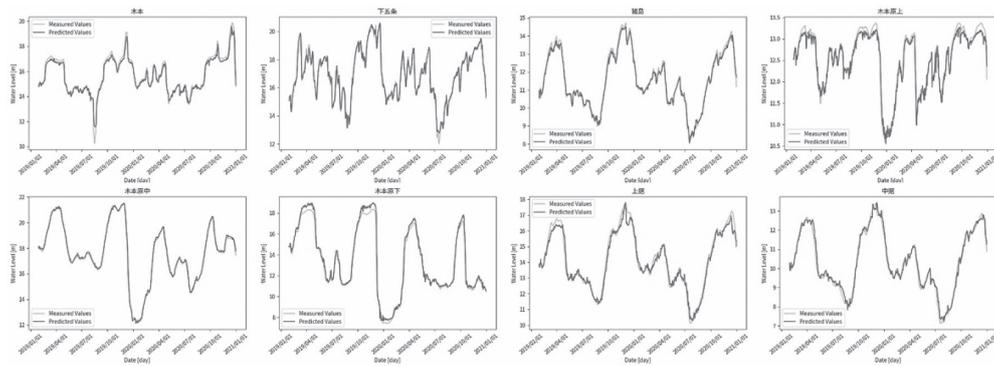
またDynamic Predictionにおいても下五条と木本原上以外の地点では全てR2値が0.8近くという非常に高い精度を示している。また、下五条と木本原上はス

コアにおいては悪いが、グラフでは実用に資するほどの精度で予測できていることがわかる(グラフ4)。

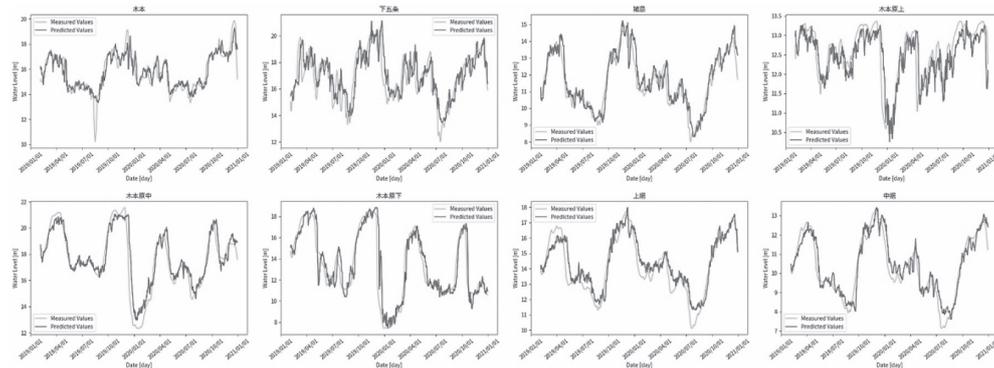
次にモデルが重視しているデータを可視化したヒートマップ(グラフ6)を見ていく。石川県と同様に、直近の気象状況がすぐに地下水位に影響することを示している。また精度に低い地点において、重視している部分が精度の良い地点に比べ分散しており、ここからモデルが正確に気象状況との関係の学習が不十分であることが推測できる。

表4 福井県大野市 予測スコア

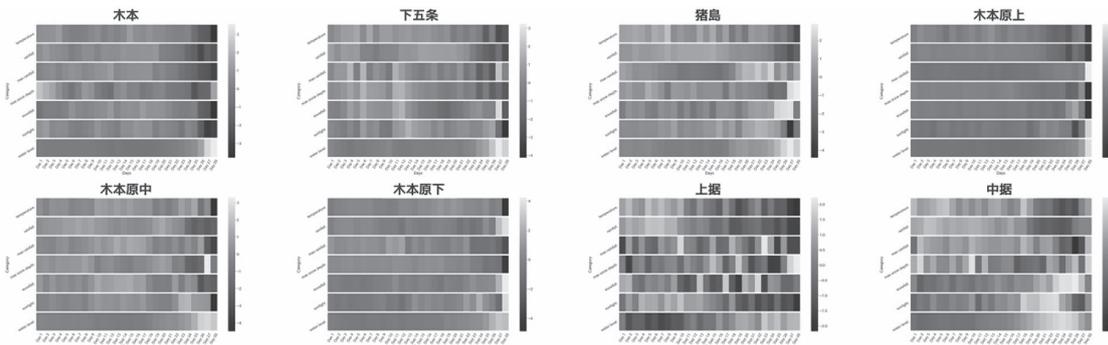
地点	Next Day Prediction		Dynamic Prediction					
	MSE	R2	Day3		Day5		Day7	
			MSE	R2	MSE	R2	MSE	R2
木本	0.038	0.966	0.121	0.890	0.171	0.844	0.223	0.797
下五条	0.006	0.985	0.049	0.872	0.096	0.749	0.140	0.635
猪島	0.006	0.990	0.043	0.932	0.056	0.913	0.084	0.869
木本原上	0.007	0.969	0.038	0.833	0.057	0.748	0.081	0.643
木本原中	0.005	0.997	0.074	0.946	0.114	0.917	0.168	0.876
木本原下	0.021	0.984	0.093	0.932	0.150	0.890	0.219	0.840
上据	0.024	0.973	0.090	0.902	0.100	0.891	0.131	0.858
中据	0.011	0.984	0.042	0.938	0.050	0.925	0.068	0.899



グラフ3 福井県大野市 予測グラフ [Next Day Prediction]



グラフ4 福井県大野市 予測グラフ [Dynamic Prediction] (全てDay7の予測である)



グラフ6 福井県大野市 可視化グラフ [Dynamic Prediction]

(項目は上から日平均気温、降水量の日合計、1時間降水量の日最大、日照時間、日最深積雪、降雪量の日合計、地下水位)

7. 本モデルの利用

ここでは本モデルの活用方法について説明していく。

結果の部分でも示したように本モデルは非常に高い精度で予測できていることがわかる。また、先行研究にはできなかった数日先の予測においても非常に高い精

度である。この結果からここでは農業面の活用、地下水保全施策の策定における活用方法について提案する。

7.1 農業・工業分野での活用

農業用水や工業用水として地下水を使用する地域は

少なくない。また、地下水は一度汲み上げ過ぎてしまうと回復が難しくなるため、適切な量を使用する必要がある。しかしながら、その量を科学的に推定するには専門的な知識や多くのデータを必要とし、コスト面や技術面において簡単ではない。そのため、使用量の決定は曖昧に決められてきた。そこで、本研究で提案したモデルを使用すれば少ないデータで、専門的な知識を必要とすることなく将来の地下水位の予測ができる。それゆえ、今までのデータに基づく合理的な制限や使用量をもとに地下水利用をすることができる。

7.2 地下水保全施策の策定

環境省水・大気環境局の『地下水保全ガイドライン』では地下水保全について「将来の地下水の状態の予測は、水収支モデルを用いたシミュレーション計算を行って地域の現況の水収支の再現、環境変化を考慮した将来予測の順に行うことが多く、この結果をもとに水循環や地下水環境・機能を保全し、利用するための具体的な方策を検討する。」と述べられており、地下水位の予測は地下水保全施策の策定において重要な意味を持つ。しかしこれも農業と同様に、合理的な規制を策定するには非常にコストがかかる。しかし、本研究で示したように提案したモデルは非常に精度が高い。それだけでなく、入力データを仮想のものとするだけで、環境変化を踏まえた将来のシミュレーションとしても使用することができる。

8. 結論

地下水位の変動予測は地下水の保全において非常に重要であるにもかかわらず、従来の手法ではデータの収集や作成自体に非常にコストがかかるだけでなく、使用用途も非常に限定的であった。そこで本研究ではこれらの課題を解決することができる高精度の機械学習モデルを開発した。実験により従来の機械学習の手法よりも精度が良いことを示しただけでなく、Dynamic Predictionという手法を導入することで、より先の地下水位も予測することが可能になった。これにより、従来の手法ではできなかった、農業・工業分野での活用や、地下水保全施策の策定に使用することができる実用上非常に効果的なモデルを開発することができた。

9. 謝辞

本研究は海城中学・高等学校の山田直樹先生には外部との連絡をはじめ、様々なお力添えを頂きました。また、

石川県生活環境部環境政策課、大野市くらし環境部水環境・水循環課の方々には突然ご連絡をしたにもかかわらず快く協力していただき、本研究において使用した貴重な長期間にわたる地下水位の実測データを提供して頂きました。また、東京都土木支援・人材育成センター技術支援課 國分邦紀氏、中山俊雄氏には本研究の研究方針や予測モデルの実社会での活用方法について様々なご意見を頂きました。本研究を行うにあたり多くの方のご指導、ご支援にこの場を借りて御礼申し上げます。

10. 参考文献

- (1) 菅原正巳『流出解析法』1972年、p.p.1~257
- (2) 長野 峻介, 野村 和哉, 藤原 洋一, 田中健二, 高瀬 恵次, 一恩 英二『ランダムフォレストを用いた手取川扇状地における地下水位変動解析』2017年、p.p.1~2
- (3) 渡部景隆, 山崎良雄『扇状地の水理地質—手取川扇状地を例として (I)』1974年
- (4) 環境省 水・大気環境局 土壤環境課 地下水・地盤環境室『地下水保全ガイドライン』2016年
- (5) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention Is All You Need. 2017
- (6) Sepp Hochreiter, Jürgen Schmidhuber. Long Short-term Memory. 1997
- (7) Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. Going Deeper with Convolutions. 2014
- (8) Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. 2019
- (9) Sergey Ioffe, Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. 2015
- (10) Diederik P. Kingma, Jimmy Ba. Adam: A Method for Stochastic Optimization. 2014