

ニューラルネットワークを用いた山形県内の気温 および降雪量の予測実験*

高野 哲夫*¹

1. はじめに

東北・北陸地方の日本海側の地域において、冬季の気象情報に対する関心は高い。例えば、夜間の降雪量および気温の予測は、道路の除雪作業や凍結防止剤の散布を行う上で重要な判断材料となる。高野 (2013) では、新潟県中越地方の例を紹介した。

また、産業界では統計学や機械学習を応用した「データサイエンス」の発展が目覚ましい。最近では気象庁による「気象データアナリスト育成講座」の認定 (<https://www.jma.go.jp/jma/kishou/shinsei/wda/index.html> 2022.02.18閲覧) もスタートしている。今後、統計学や機械学習を用いた気象データ分析やその活用の場がさらに広がるものと期待される。

さて、筆者はこれまで機械学習 (ニューラルネットワーク) を用いて山形県・新潟県における局地気象の解析に取り組んできた。まず高野 (2015) ではサクラ開花日の予測を試みた。続いて、山形県の冬季降水域および気温分布 (高野 2016)、新潟県の冬季降水域 (高野 2018) の解析を行った。さらに、全球数値予報モデル (以下、GSM) の予報値 (以下、GPV) を基に山形県内の降雪量および気温を予測するモデルを構築し、予測実験を試みた (高野 2020a, b, 高野 2021a, b)。なお、これらの研究で使用したモデルは全て3層構造のフィードフォワード型ニューラルネットワークである。

そこで本研究では、新たに4層構造のニューラル

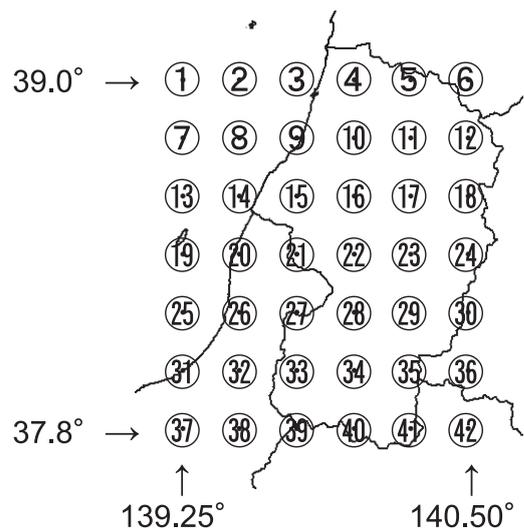
ネットワーク、並びに重回帰分析に基づくモデルを構築し、3種類のモデルによる気温および降雪量予測の二乗平均平方根誤差 (以下、RMSE) を比較した。

2. 使用するデータ

ニューラルネットワークによる機械学習、および重回帰分析のため、GSMGPV と地上気象観測から構成されるデータベース (気温・降雪量の2種類) を構築した。以下に使用したソースデータの概要を述べる。

2.1 GSMGPV

山形県付近に対応する東西方向6×南北方向7の計42格子点 (第1図) における地上気温 (時別値) および降水量 (前1時間) を使用した。対象期間は2016~2022年の各1~2月、初期時刻は9時 (日本時間) のGPVを使用した。



第1図 山形県付近のGSM格子点 (計42点)。

* The experiment on prediction temperature and snowfall in Yamagata prefecture with using the neural network

*¹ Tetsuo TAKANO, 株式会社 SnowCast.
takano@snowcast-web.com

© 2022 日本気象学会

2.2 地上気象観測

山形県内15地点（第2図）の気温（時別値）、降雪量（前1時間の積雪深差）および降水量（前1時間）を使用した。対象期間は2016～2022年の各1～2月である。気温の観測値は鶴岡を除く14地点、降雪量・降水量の観測値は鶴岡を除く14地点を使用した。

データベースの構築に際して、気温の欠測が1地点以上ある時刻のデータは対象から除去した。また、降雪量の欠測が1地点以上ある場合は、気温および降水量から当該降雪量を推定し、これを補完した（後述4.2による）。ただし、気温または降水量も欠測の場合は、当該時刻のデータを対象から除去した。

3. 気温の予測実験

まず、地上気温 GPV を基に山形県内14地点の気温を予測するモデルを構築し、予測実験を行った。

3.1 説明変数と目的変数

本研究では、モデルの入力に相当する変数を説明変数、出力に相当する変数を目的変数と表記する。

この実験において、説明変数は地上気温 GPV、目的変数は山形県内14地点の気温である。過去（2016～

2021年の各1～2月）の気温データベースを基に、3種類のモデルを構築し、将来（2022年1～2月）の気温の予測を試みた。ここで、対象時間帯は初期時刻の当日19時～翌朝6時とした。

3.2 ニューラルネットワークを用いたモデル

3.2.1 3層モデルの構成

3層モデルは入力層—中間層（1層）—出力層から構成されている（第3図）。

素子数は入力層42、中間層28、出力層14であり、中間層の活性化関数はシグモイド関数、出力層の活性化関数は線形関数を適用した。また、学習方法は誤差逆伝搬法を用いた。

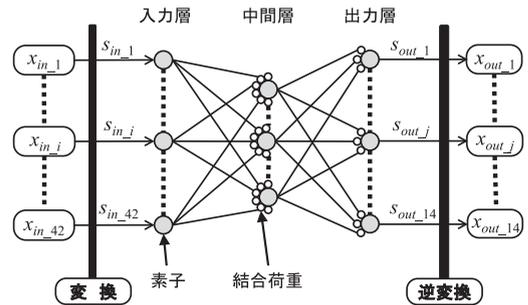
3.2.2 4層モデルの構成

4層モデルは入力層—中間層（2層）—出力層から構成されている（第4図）。ここで、中間層の多層化に際しては、熊沢（1998）および涌井・涌井（2017）を参考にした。

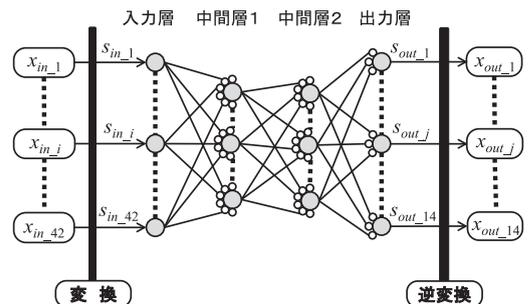
素子数は入力層42、中間層は順に28と21、出力層14であり、中間層の活性化関数はいずれもシグモイド関数、出力層の活性化関数は線形関数を適用した。また、学習方法は誤差逆伝搬法を用いた。



第2図 山形県内の観測点（気温・降雪量・降水量）。鶴岡は気温、楡引は降雪量・降水量を使用（降水量は降雪量が欠測時の推定のみで使用）。



第3図 3層モデルの構成。変換は(1) (5)式、逆変換は(2) (6)式による。



第4図 4層モデルの構成。変換は(1) (5)式、逆変換は(2) (6)式による。

3.2.3 モデル内部のレベル変換

中間層素子の活性化関数にシグモイド関数を用いるため、モデル内部では説明変数・目的変数の値を 0 ～ 1 の実数の形で処理する必要がある。

そこで、説明変数 x_m を内部入力値 s_{in} に変換し、また内部出力値 s_{out} を目的変数 x_{out} に逆変換する仕組みを設けた (第 3 図, 第 4 図)。この変換・逆変換には、格子点・観測点毎の過去の値の平均 m および標準偏差 σ を用いて、次式 (1) (2) を適用するものとした。これは、 $x = m - 3\sigma$ を $s = 0$ 、 $x = m + 3\sigma$ を $s = 1$ に各々対応させたものである。格子点・観測点毎の平均値、標準偏差を第 1 表および第 2 表に示す。

$$s_{in,i} = \frac{x_{in,i} - (m_i - 3\sigma_i)}{6\sigma_i} \tag{1}$$

$$x_{out,j} = (m_j - 3\sigma_j) + 6\sigma_j s_{out,j} \tag{2}$$

ここで、 $x_{in,i}$ 、 m_i 、 σ_i 、 $s_{in,i}$ は格子点 i の説明変数、過去の平均値および標準偏差、内部入力値、また、 $x_{out,j}$ 、 m_j 、 σ_j 、 $s_{out,j}$ は観測点 j の目的変数、過去の平均値および標準偏差、内部出力値。

3.2.4 機械学習と予測

第 5 図には機械学習と予測のフローを示す。機械学習では、過去の説明変数と目的変数 (教師データ) の対で構成される気温データベース (訓練用) を基に機械学習を行う。その後の予測では、将来の GPV に基づく説明変数をモデルに適用し、目的変数を得る。

3.3 重回帰分析を用いたモデル

ニューラルネットワークによるモデルと比較するために、重回帰分析に基づくモデル (以下、重回帰モデル) も構築した。第 6 図には重回帰モデルの構築と予測のフローを示す。

格子点 i の説明変数を x_i 、係数を a_i 、切片を a_0 、さらに $x_0 = 1$ とおくと、観測点 j の目的変数 y_j は次式 (3) で表される。

$$y_j = \sum_{i=0}^{42} a_i x_i \tag{3}$$

3.4 予測結果の比較

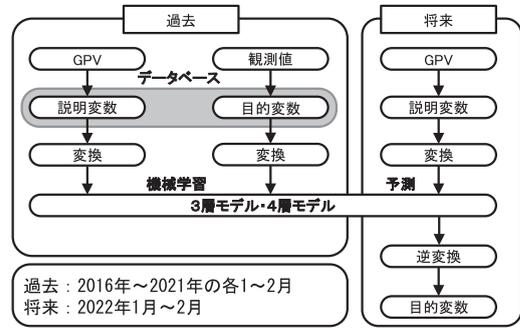
3 層モデル、4 層モデル、および重回帰モデルによる RMSE の比較を第 7 図に示す。14 地点中 9 地点で重回帰モデルの値が最も低く、予測精度が良好であった。また、3 層モデルと 4 層モデルの RMSE を比較す

第 1 表 格子点毎の平均・標準偏差 (気温) および最大値 (降雪量)。格子点番号は第 1 図に基づく。

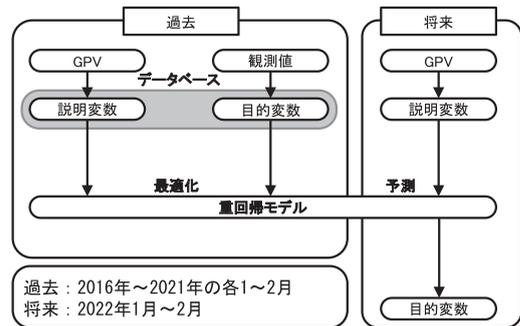
格子点	気温		降雪量
	平均 (°C)	標準偏差 (°C)	最大値 (cm)
①	4.5	2.7	8.3
②	4.5	2.6	8.5
③	2.9	2.6	17.3
④	-0.9	3.1	33.2
⑤	-2.8	3.6	36.5
⑥	-3.9	3.9	29.7
⑦	4.8	2.7	8.0
⑧	4.4	2.5	10.6
⑨	1.7	2.8	26.0
⑩	-1.2	3.5	38.6
⑪	-2.7	3.9	35.0
⑫	-3.8	3.9	26.3
⑬	5.0	2.6	7.9
⑭	3.6	2.5	18.6
⑮	-1.0	3.2	36.2
⑯	-2.9	3.7	37.5
⑰	-3.0	4.0	30.7
⑱	-3.4	4.0	22.4
⑲	5.0	2.5	8.9
⑳	1.9	2.8	27.7
㉑	-2.8	3.5	42.3
㉒	-4.1	3.8	36.6
㉓	-3.3	4.0	21.9
㉔	-3.1	3.8	24.4
㉕	4.8	2.5	10.8
㉖	0.6	3.1	30.3
㉗	-3.0	3.8	37.2
㉘	-3.7	4.1	27.4
㉙	-3.3	4.1	26.1
㉚	-2.8	3.6	28.9
㉛	3.1	2.8	37.4
㉜	-0.7	3.4	46.5
㉝	-3.2	3.9	41.1
㉞	-3.4	4.1	30.0
㉟	-3.1	3.9	29.1
㊱	-2.2	3.5	29.0
㊲	0.5	3.5	61.9
㊳	-2.1	3.8	59.1
㊴	-3.6	4.0	44.4
㊵	-4.0	3.9	31.7
㊶	-3.3	3.6	32.2
㊷	-1.7	3.4	27.9

第2表 観測点毎の平均・標準偏差(気温)および最大値(降雪量). 鶴岡は気温のみ, 楡引は降雪量のみを対象.

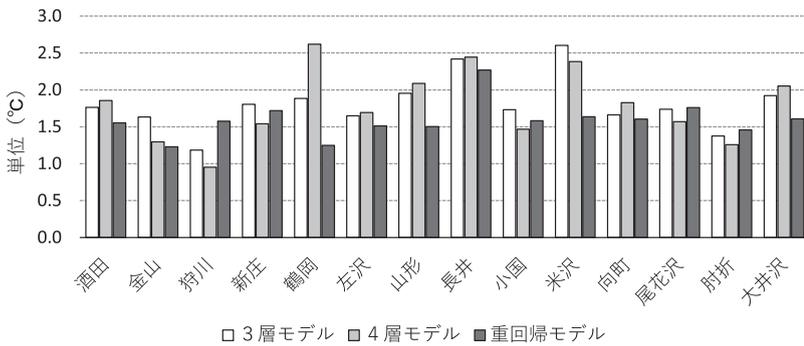
観測点	気温		降雪量
	平均 (°C)	標準偏差 (°C)	最大値 (cm)
酒田	2.0	2.8	32
金山	-1.6	2.6	21
狩川	0.6	2.5	110
新庄	-1.0	2.7	32
鶴岡	1.7	2.7	—
楡引	—	—	40
左沢	-2.0	2.7	25
山形	-0.3	2.9	21
長井	-1.0	3.4	29
小国	-0.5	2.9	29
米沢	-1.4	3.2	28
向町	-1.5	2.9	31
尾花沢	-1.3	2.8	35
肘折	-1.9	2.9	56
大井沢	-2.6	3.4	30



第5図 機械学習と予測のフロー. 将来の目的変数は予測結果, 過去の目的変数は教師データに相当.



第6図 重回帰モデルの構築と予測のフロー. 将来の目的変数は予測結果に相当.



第7図 RMSEの比較(気温).

ると, 7地点で4層モデルの方が低かった.

予測値と実況値の対応例として, 第8図に鶴岡に關する4層モデルと重回帰モデルの結果を示す. これは予測値の階級を行方向, 実況値の階級を列方向に取り, 両者の各組み合わせが出現した回数を集計したものである. この回数が対角線上に分布するほど予報精

度は高いと言える.

第8図の分布によると, (a) 4層モデルでは実況値に比べて予測値の方が高い傾向が顕著であった. 一方, (b) 重回帰モデルでは概ね対角線付近に分布しており, より適切に予測されていたことが判る. 他の地点でも程度の差はあるが, 同様の傾向が見られた.

今後の課題としては, (1) (2) 式で用いるパラメータの設定や中間層素子数の仕様が挙げられる.

4. 降雪量の予測実験

続いて, 地上気温と降水量のGPVから推定した積算降雪量(以下, 推定降雪量)を基に, 山形県内14地

点の積算降雪量を予測するモデルを構築し、予測実験を行った。

4.1 説明変数と目的変数

この実験において、説明変数は推定降雪量、目的変数は山形県内14地点の積算降雪量である。過去（2016～2021年の各1～2月）の降雪量データベースを基に、3種類のモデルを構築し、将来（2022年1～2月）の12時間降雪量の予測を試みた。ここで、対象時間帯は初期時刻の当日18時～翌朝6時とした。

4.2 推定降雪量の生成

現状、GSMでは降雪量GPVは提供されていない。そこで、地上気温と降水量のGPVから推定降雪量を生成する（第9図）。地上気温 T に次の近似式(4)を適用して雪水比 $r(T)$ を求め、前1時間降水量 P との積 $Pr(T)$ を前1時間の推定降雪量とする。

$$r(T) = \frac{a}{1 + \exp\{b(T+c)\}} \tag{4}$$

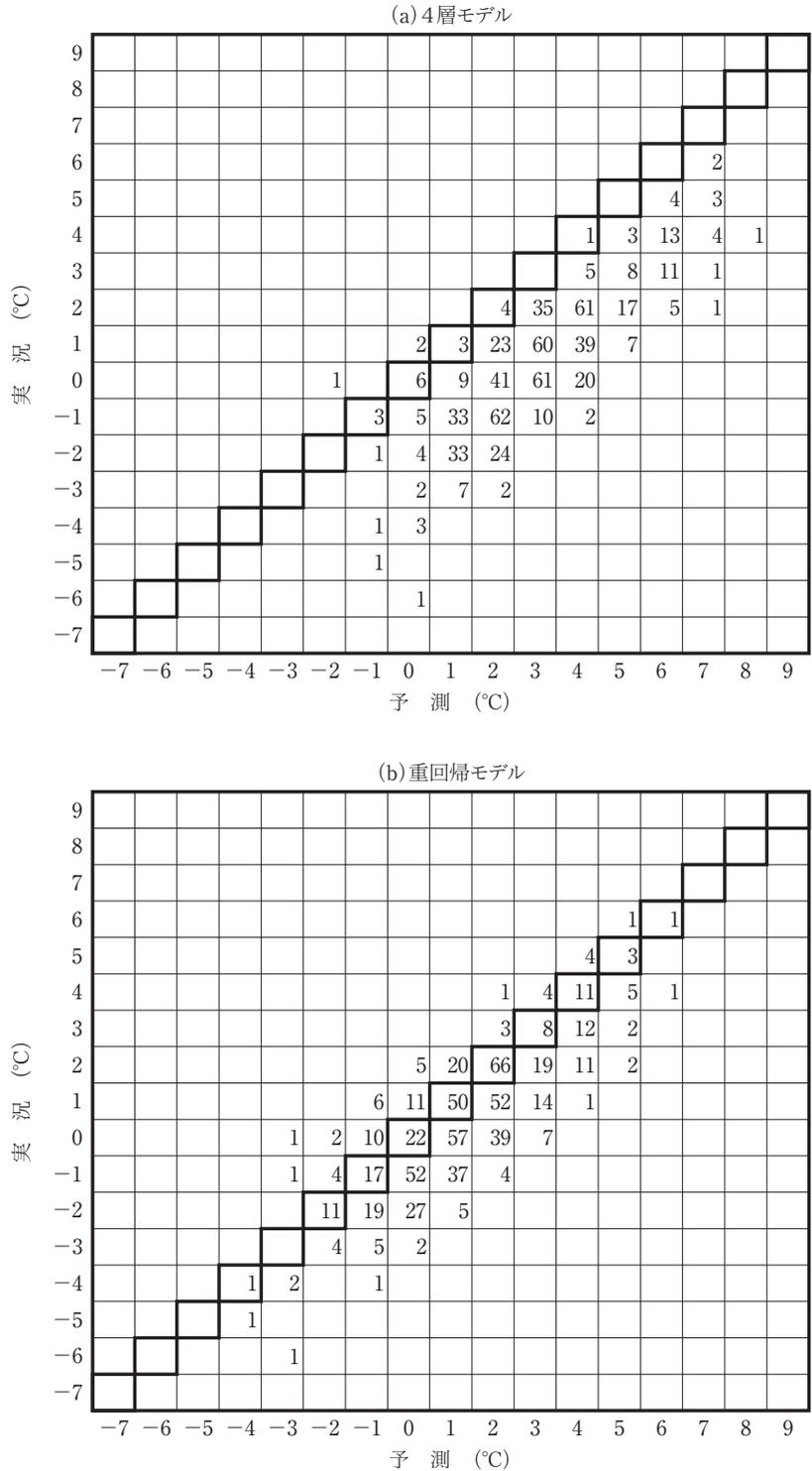
ここで、 $a=2.08$, $b=1.49$, $c=-0.33$ (高野 2021a)。

初期時刻の当日19時から翌朝6時まで各時刻の前1時間の推定降雪量を算出し、その12時間の積算値を推定降雪量(説明変数)とした。

4.3 ニューラルネットワークを用いたモデル

4.3.1 3層モデルと4層モデルの構成

気温の場合と同様、3層モデルと4層モデルの2種類を



第8図 予測と実況の対応(気温:鶴岡)。(a) 4層モデル, (b) 重回帰モデル。予測・実況共に四捨五入した整数値を基に階級を判定。

構築した。各モデルの構造は、先述の3.2.1, 3.2.2と同様である。

4.3.2 モデル内部のレベル変換

気温の予測実験と同様、説明変数 x_{in} を内部入力値 s_{in} に変換し、また内部出力値 s_{out} を目的変数 x_{out} に逆変換する仕組みを設けた (第3図, 第4図)。この変換・逆変換には、格子点・観測点毎の過去の値の最大値 x_{max} を用いて、次式 (5) (6) を適用するものとした。これは、 $x=0$ を $s=0$, $x=x_{max}$ を $s=1$ に各々対応させたものである。格子点・観測点毎の最大値を第1表および第2表に示す。

$$s_{in,i} = \frac{x_{in,i}}{x_{max,i}} \tag{5}$$

$$x_{out,j} = x_{max,j} s_{out,j} \tag{6}$$

ここで、 $x_{in,i}$, $x_{max,i}$, $s_{in,i}$ は、格子点 i の説明変数、過去の最大値、内部入力値、また、 $x_{out,j}$, $x_{max,j}$, $s_{out,j}$ は、観測点 j の目的変数、過去の最大値、内部出力値。

4.3.3 3層モデルと4層モデルの構成

先述の3.2.4と同様、第5図のフローに従って降雪量データベース (訓練用) を基に機械学習と予測を行った。

4.4 重回帰分析を用いたモデル

先述の3.3と同様、第6図のフローに従って重回帰モデルを構築した。

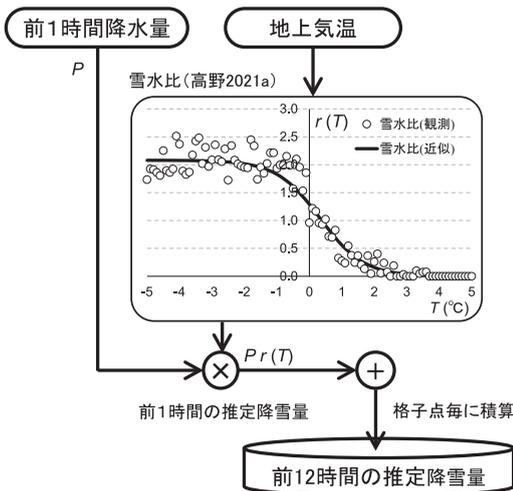
4.5 予測結果の比較

3層モデル, 4層モデル, および重回帰モデルによるRMSEの比較を第10図に示す。14地点中9地点で4層モデルの値が最も低く、予測精度が良好であった。また、3層モデルと4層モデルのRMSEを比較すると、10地点で4層モデルの方が低かった。

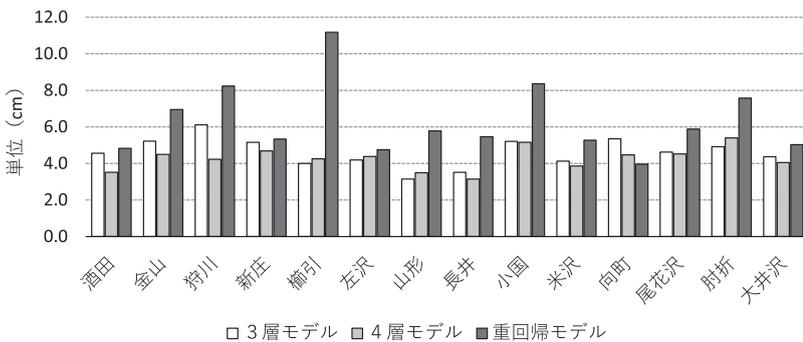
予測値と実況値の対応例として、第11図に楯引に関する4層モデルと重回帰モデルの結果を示す。これは第8図と同様に、予測値と実況値の階級の各組み合わせが出現した回数を集計したものである。

第11図の分布によると、(b) 重回帰モデルでは実況値より予測値の方が過大となる事例が目立った。一方、(a) 4層モデルではより対角線に沿って分布する傾向が見られた。他の地点でも程度の差はあるが、同様の傾向が見られた。これは、ニューラルネットワークの関数近似能力が良く現れたものと考えられる。

引き続き、事例検証を重ねて行きたい。



第9図 推定降雪量の生成.

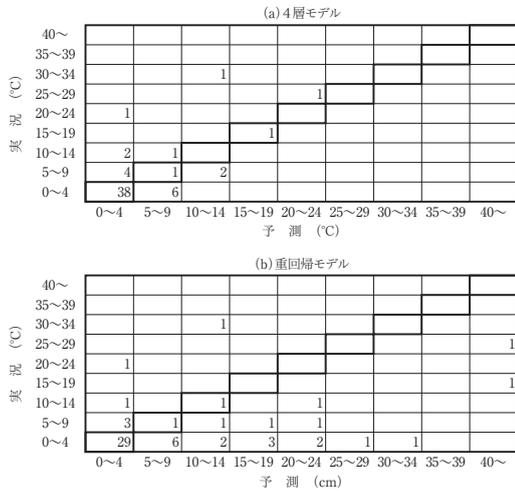


第10図 RMSEの比較 (降雪量).

5. おわりに

本研究では、3層および4層構造のニューラルネットワーク、並びに重回帰分析に基づく予測モデルを構築し、気温および降雪量の予測誤差を比較した。この結果、降雪量予測におけるニューラルネットワーク活用の可能性が見出された。一方、気温予測では改善の余地が多く残されている。

今後も引き続き、過去デー



第11図 予測と実況の対応（降雪量：楯引）。(a) 4層モデル、(b) 重回帰モデル。予測・実況共に四捨五入した整数値を基に階級を判定。また、予測値が負の値を示す場合は0 cmに補正。

タベースのさらなる充実，モデル内部のレベル変換の改善，中間層素子数の適正化，さらにCNN（Convolutional Neural Network：畳み込みニューラルネットワーク）の適用も視野に入れて検討したい。

謝 辞

研究や教育を本務としない環境にあって本研究を進めるに当たり，株式会社 SnowCast 代表取締役杉浦聡氏より御支援を頂きました。

また，日本気象学会の春季・秋季大会の場において，参加者の皆様より多くのコメントやアドバイスを頂き

ました。数々の議論を通じて，研究の方向性や課題解決のヒントを得ることができました。

以上，関係者の皆様に感謝申し上げます。

参 考 文 献

熊沢逸夫，1998：電子情報通信工学シリーズ 学習とニューラルネットワーク。森北出版，182pp。

高野哲夫，2013：新潟県中越地方における降雪量と時間帯の関係—除雪作業支援の観点から—。天気，60，591-596。

高野哲夫，2015：ニューラルネットワークによるサクラ開花日の学習・予測実験。天気，62，807-812。

高野哲夫，2016：ニューラルネットワークを用いた山形県内の冬季降水域・気温分布の解析。日本気象学会2016年度春季大会講演予稿集，P406。

高野哲夫，2018：ニューラルネットワークを用いた新潟県内の冬季降水域の解析。日本気象学会2018年度秋季大会講演予稿集，D461。

高野哲夫，2020a：GSM 地上とニューラルネットワークを用いた山形県内の降雪量解析の試み。日本気象学会2020年度春季大会講演予稿集，P108。

高野哲夫，2020b：GSM 地上を用いた山形県内における降雪量ニューロ・モデルの開発。日本気象学会2020年度秋季大会講演予稿集，SP4-3。

高野哲夫，2021a：GSM 地上を用いた山形県内における降雪量ニューロ・モデルの開発（第3報）。日本気象学会2021年度春季大会講演予稿集，WF-23。

高野哲夫，2021b：GSM 地上を用いた山形県内における冬季気温ニューロ・モデルの開発。日本気象学会2021年度秋季大会講演予稿集，WF-17。

涌井良幸，涌井貞美，2017：ディープラーニングがわかる 数学入門。技術評論社，240pp。