

深層学習手法を用いた積雪深推定

熊本大学工学部土木建築学科 学生会員 玉置 慎

熊本大学 くまもと水環境・減災研究教研究センター 正会員 石田 桂

熊本大学自然科学教育部 学生会員 横尾 和樹

熊本大学自然科学教育部 学生会員 永里 起義

熊本大学自然科学教育部 学生会員 坂口 大珠

1. はじめに

豊かな自然に恵まれている日本は、その一方で自然災害が多い国としても知られている。その中でも特に雪害において、日本は国土の半分以上が豪雪地帯に指定されており約 2000 万人もの人々が豪雪地帯で生活を営んでいる。豪雪地帯において、降雪量は水資源に直接的に影響を与える。そのため、積雪・降雪の面的な状況の把握が重要となるが、観測点がまばらにしか存在していない、標高の高い場所がない等、空間分布が十分ではないのが現状である。これらの理由より、積雪量のモデリングおよび積雪深の推定はその地域の重要な課題の一つであると言える。近年、深層学習手法の一つである LSTM (Long Short-Term Memory)¹⁾を時系列モデリングに適用して精度の良い結果を出した事例が増えている。本研究では LSTM を用いて、複数の観測点のデータセットで学習を行うことでモデルを作成し、石狩川流域の他の観測点にモデルを適用して積雪深の推定を行い、一地点のみで学習し、積雪深の推定を行った場合の結果と精度の比較を行う。

2. 適用

本研究では対象地に LSTM を用いて推定を行う。積雪深モデルを構築するための入力には日単位の降水量データと気温データを用い、対象とした地点における積雪深を出力とする。LSTM の学習過程における重みとバイアスの初期値に起因するランダム性を考慮し、学習を 50 回行った。その 50 回の学習のうち最も精度が高いものをモデルとして用いた。本研究では、この訓練されたモデルに学習、検証に用いていない他地点の入力データを与えて出力された推定積雪深を実際の観測積雪深と比較し、精度を調査する。

3. データセット

3.1 対象地

積雪期間に多くの降雪量が観測され、豊富な観測データがあることから、本研究の対象地として石狩川流域を選定した。石狩川は、北海道中西部を流れ日本海へ注ぐ日本三大河川の 1 つであり、流域面積は 14,330km²、幹川流路延長は 268km、流域内人口は 312 万 5 千人の一級河川である。

表 1. ハイパーパラメータの組み合わせ

ハイパーパラメータ	
IDL (入力データの長さ)	50
HSL (Hidden State Layer)	75
バッチサイズ	1024
バッチ要素抽出方法	shuffle
Patience	30
Loss function	MSE
オプティマイザー	Adam

3.2 データ

気温データは、ERA5 (European Centre for Medium-Range Weather Forecasts Reanalysis v5)²⁾から取得し、標高で補足を行い日平均気温として用いた。ERA5 は ECMWF によって作成された全球レベルの大気情報の再解析データセットである。水平解像度は 25km×25km であり、1979 年から現在までのデータが公開されている。降水量データは、APHRODITE (Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation of water resources)³⁾から取得し、石狩川の流域平均値に変換をして流域平均日降水量として用いた。APHRODITE は観測データを元にした日降水量グリッドデータであり、空間解像度は 5km である。日本全域においては 1900 年から 2015 年の日降水量が提供されている。任意の地点の積雪深データは気象庁から取得した。

3.3 実装

本研究では、石狩川流域の観測点のうち、11 地点の入力、出力データを学習に用い、さらに別の 4 地点を検証に、そしてまた異なる 4 地点をテストに用いた。これらの地点の選出は地点ごとの標高の組み合わせが均等になるように行った。また期間としては、それぞれ 1988 年から 2015 年までのデータを用いている。そして精度を比較するためにテストとして用いている 4 地点と同じ地点でそれぞれ、一地点を学習期間(1988 年から 2004

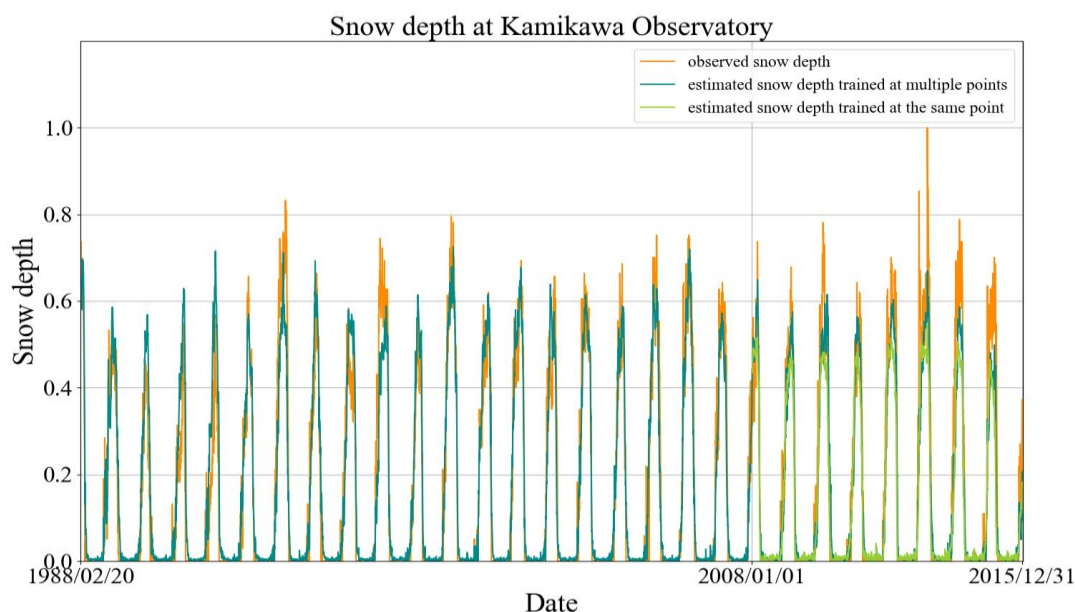


図. 1 上川観測点の観測積雪深と推定積雪深の比較

年),検証期間(2005 年から 2007 年),テスト期間(2008 年から 2015 年)に分けて学習を行い,比較を行った。

各モデルの精度を比較するための評価指標として,Nash-Sutcliffe efficiency (NSE) ⁴⁾を使用した。NSE は負の無限大から 1 で表され,値が 1 に近づくほどモデルの精度がよいとされる。また,ハイパーパラメータについてはそれぞれ数値を変化させて学習を行い,最も精度が良いとしたものを用いている。そのハイパーパラメーターの組み合わせについては,表 1 に示している。

4. 結果・考察

今回テストとして用いた 4 地点の内の一つである上川観測所の観測積雪深(1988 年～2015 年), 11 地点を学習データとして推定を行った推定積雪深(1988 年～2015 年),上川観測点のデータを学習期間,検証期間,テスト期間に分けて推定を行った場合の推定積雪深(2008 年～2015 年)を図 1 に示す。X 軸は日付を, Y 軸には積雪深を正規化した値を示している。

4.1 観測積雪深と推定積雪深の比較

図 1 の観測積雪深と,本研究で 11 地点のデータを学習して推定した積雪深との比較を行うと, 一年ごとの大まかな変動傾向を捉えた積雪深の推定ができていると言える。また,精度評価に用いた NSE の値も 0.9112 という値を示しており高い精度であると言える。そのため,入力データとして用いた 2 つのデータは積雪に密接に関係していると考えられる。しかし大まかな変動傾向は捉えられていても各年のピークとなるような積雪深の推定には誤差が生じている。

4.2 各条件で推定した積雪深の比較

図 1 の 11 地点のデータを学習して推定した積雪深と上川観測点のテスト期間の推定積雪深の比較を行う。尚,比較に関しては上川観測点のみを入力データに用いた推定の場合,対象としている期間が 2008 年以降の

ため,それぞれの推定結果の比較は 2008 年以降のものとする。2008 年以降の推定に関してそれぞれの NSE の比較を行うと,1 地点のデータのみで推定した場合の NSE 値は 0.8663,11 地点のデータを学習して推定を行なった場合の NSE 値は 0.8937 となっており,NSE の値から見ても推定精度が向上していると言える。これは 11 地点の入力データと出力データの関係性をそれぞれ LSTM が学習したことで,テストに用いた地点の入力データの特徴に近い学習結果から推定ができたからと考えられる。

5. まとめ

本研究では,深層学習手法の 1 つである LSTM を用いて,複数の観測点のデータを入力データに用いてモデルを作成し,観測積雪深や一地点のみを入力としたモデルとの比較を行なった。本研究で作成したモデルはテストとして用いた 4 地点において,1 地点のみを入力とした場合の精度に近く,高い精度での推定ができていた。そのため,他の観測点でも同じように高い精度で推定することが可能となる,その地域で汎用的なモデルの作成ができたと言える。

6. 参考文献

- 1) Hochreiter, Sepp, and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9.8: 1735-1780, 1997 (LSTM) networks. Hydrol. Earth Syst. Sci., 22.11: 6005-6022., 2018
- 2) Hersbach, Hans, et al. "The ERA5 global reanalysis." Quarterly Journal of the Royal Meteorological Society 146.730 (2020): 1999-2049.
- 3) Kamiguchi, K. Arakawa, O. Kitoh, A. Yatagai, A. Hamada, A. Yasutomi, N. Development of APHRO_JP, the first Japanese high resolution daily precipitation product for more than 100 years. Hydrological Research Letters, 4: 60-64 2010.
- 4) Nash, J. Eamonn; SUTCLIFFE, John V. River flow forecasting conceptual models part I-A discussion of principles. Journal of hydrology, 10.3: 282-290, 1970