

딥러닝 기반 GNSS 천정방향 대류권 습윤지연 추정 연구

Estimation of GNSS Zenith Tropospheric Wet Delay Using Deep Learning

임수현¹⁾ · 배태석²⁾

Lim, Soo-Hyeon · Bae, Tae-Suk

Abstract

Data analysis research using deep learning has recently been studied in various field. In this paper, we conduct a GNSS (Global Navigation Satellite System)-based meteorological study applying deep learning by estimating the ZWD (Zenith tropospheric Wet Delay) through MLP (Multi-Layer Perceptron) and LSTM (Long Short-Term Memory) models. Deep learning models were trained with meteorological data and ZWD which is estimated using zenith tropospheric total delay and dry delay. We apply meteorological data not used for learning to the learned model to estimate ZWD with centimeter-level RMSE (Root Mean Square Error) in both models. It is necessary to analyze the GNSS data from coastal areas together and increase time resolution in order to estimate ZWD in various situations.

Keywords : Deep Learning, Global Navigation Satellite System, Zenith Tropospheric Wet Delay, Meteorological Data

초 록

최근 딥러닝을 활용한 데이터 분석 연구가 다양한 분야에서 진행되고 있다. 본 논문에서는 딥러닝 모델인 MLP (Multi-Layer Perceptron)와 LSTM (Long Short-Term Memory) 모델을 통해 ZWD (Zenith tropospheric Wet Delay)을 추정함으로써 딥러닝을 활용한 GNSS (Global Navigation Satellite System) 기반 기상 연구를 수행하였다. 딥러닝 모델은 기상 데이터와 천정방향 대류권 총 지연, 건조지연을 통해 추정된 ZWD로 학습되었고, 학습에 사용되지 않은 기상 데이터를 학습된 모델에 적용하여 두 모델에서 센티미터 수준의 RMSE (Root Mean Square Error)로 ZWD 결과를 산출하였다. 추후 해안지역의 GNSS 데이터를 함께 사용하고 시간 해상도를 높여 다양한 상황에서 ZWD가 추정될 수 있도록 추가적인 연구가 수행될 필요가 있다.

핵심어 : 딥러닝, 범지구항법위성시스템, 천정방향 대류권 습윤지연, 기상 데이터

Received 2021. 01. 25, Revised 2021. 02. 08, Accepted 2021. 02. 19

1) Member, Dept. of Geoinformation Engineering, M.S. student, Sejong University (E-mail: lsooh0612@gmail.com)

2) Corresponding Author, Member, Dept. of Geoinformation Engineering, Professor, Sejong University (E-mail: baezae@sejong.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

현대 사회에서 기상 예측은 인류의 생활 및 산업에 큰 영향을 주는 중요한 정보이다. 최근 기후변화로 예상치 못한 이상 기후들이 발생하면서 기상 예측을 위한 다양한 기술들이 사용되고 있다. 그중 빅데이터와 AI (Artificial Intelligence)를 이용한 연구들이 활발하게 진행되고 있는데, 미국의 다국적 기업인 구글은 AI와 기계학습을 이용한 기상 예측 모델인 Nowcast를 개발하여 분석 시간 단축과 더불어 최대 6시간 이후의 날씨를 예측하는 연구성과를 발표하였다(Jason, 2020). 국내 기상청에서도 기상 빅데이터와 AI를 융합하여 인공지능 기상 예보 프로그램인 알파웨더를 한국과학기술원과 함께 공동 개발 중이다(KMA, 2020). 기상 분야뿐만 아니라 공간정보 분야에서도 GNSS 기반의 기상 및 강수량을 예측하기 위한 연구가 국내외로 다수 수행되고 있는데, GNSS 자료를 통해 산출되는 값 중에서 수증기와 관련 있는 대류권 지연과 가강수량을 이용한 GNSS 기반 기상 연구가 주로 진행되고 있다.

GNSS 대류권 지연은 위성 신호가 대류권을 통과할 때 비산란성(non-dispersive) 매질인 중립 대기층에 의한 신호 굴절로 위성파 수신기 사이의 기하학적 거리가 실제보다 길어지는 현상을 일컫는다(Hofmann-Wellenhof *et al.*, 2001). 중립 대기층은 건조한 대기 성분과 수증기로 이루어지는데, 구성 성분의 굴절률에 따라 대류권 지연도 건조지연, 습윤지연으로 나누어진다. 일반적으로 건조지연의 경우 경험적 모델로 추정하고 습윤지연은 총 지연에서 건조지연을 뺀 값으로 결정한다(Bevis *et al.*, 1992). 국내에서 Kim and Bae (2015)는 GNSS 대류권 지연 추정값을 통해 계산한 가강수량을 분석하여 강수량 전조현상 예측에 적용하는 연구를 수행하였으며, Nam and Song (2019)은 세 가지 대류권 건조지연 모델을 통해 GNSS를 이용한 가강수량을 산출하여 비교 분석을 수행하였다. 최근 AI와 딥러닝(deep learning)에 관한 관심이 늘어나면서 이러한 GNSS 기반 기상 연구도 딥러닝을 적용하여 다양하게 수행되고 있다. Katsougiannopoulos and Pikridas (2009)는 GPS 관측소의 위치 정보와 시간 데이터를 MLP (Multi-Layer Perceptron) 모델의 입력 인자로 사용하여 ZTD (Zenith tropospheric Total Delay)를 센티미터 수준의 정확도로 예측하였다. Kim and Bae (2017)는 기상데이터와 LSTM 모델을 활용한 강수량 예측 연구를 수행하였다. 2020년에는 ANN (Artificial Neural Network) 모델을 통해 6시간 후의 천정방향 대류권 습윤지연(ZWD)을 예측하는 연구가 이루어졌

다(Selbesoglu, 2020).

기존 칼만 필터를 이용한 추정법은 추정값에 영향을 주는 노이즈(noise)의 공분산 행렬의 초기값을 경험적으로 설정하여야 하고, 이는 추정 결과가 사용자의 숙련도에 영향을 받을 수 있음을 나타낸다. 딥러닝을 통한 추정법도 경험적으로 모델 내부 변수를 조정하는 과정이 필요하다. 그러나 칼만 필터와 달리 추정값의 노이즈를 직접적으로 다루지 않는다는 점에서 차이가 존재한다. 또한 국내에서 딥러닝을 통한 가강수량 추정 연구는 수행되었으나, 가강수량을 계산하기 전 단계에서 필요한 ZWD 추정 연구는 현실점에서 이루어지지 않은 것으로 파악된다. 본 논문은 딥러닝 기반의 ZWD 추정 연구로서 정밀질대측위(PPP: Precise Point Positioning) 방식을 이용하여 산출한 천정방향 대류권 지연과 기상 데이터를 통해 딥러닝 모델을 학습하고 ZWD를 추정하였다. 딥러닝 모델로는 MLP와 LSTM을 사용하였으며, ZHD (Zenith tropospheric Hydrostatic Delay) 계산을 위해 경험적 건조지연 모델을 사용하였다.

2. GNSS 습윤 지연 추정

2.1 천정방향 대류권 습윤지연

천정방향 대류권 습윤지연은 Eq.(1)과 같이 표현할 수 있으며, 계산을 위해서는 천정방향 대류권 총 지연과 건조지연이 필요하다.

$$ZWD = ZTD - ZHD \quad (1)$$

여기서 ZTD는 GNSS 자료처리를 통해 추정할 수 있다. 본 연구에서는 GNSS 데이터 통합센터에서 사용할 데이터인 수월 상시관측소(SUWN)의 2015년 일일 RINEX (Receiver INdependent EXchange) 형식 데이터를 수집하였다(GDC, 2020). 대류권 지연은 대기 중 수분에 영향을 많이 받는 값이기 때문에 그 영향을 최소화하고자 최근 몇 년간 평균 연강수량 중에서 가장 낮은 값을 기록한 2015년 데이터를 사용하였다. 자료처리에 필요한 위성궤도, 시계오차, DCB (Differential Code Biases)와 같은 외부 데이터를 다운로드한 뒤 과학기술용 자료처리 소프트웨어인 Bernese GNSS Software V5.2를 통해 PPP 방식으로 자료처리를 수행하여 시간 해상도가 1시간인 ZTD를 추정하였다. GNSS 데이터 처리과정은 Kim and Bae (2012, 2018), Bae (2018)을 참고하였다. ZHD는 경험적 건조지연 모델을 통해 계산이 가능한데, 본 연구에서는 다양한 건조지연 모델 중 GNSS 기상 기술 분야에서 가장 많이 사용되는 Saastamoinen 모델을 사용하였다(Nam and Song, 2019;

Saastamoinen, 1973). Saastamoinen 모델은 다음 식과 같으며, 건조지연 단위는 mm이다.

$$ZHD = \frac{(2.2779 \pm 0.0024) \cdot P}{1 - 0.00266 \cdot \cos 2\phi - 0.00026 \cdot h} \quad (2)$$

where P is the surface pressure (hPa), ϕ is the latitude and h is the ellipsoidal height (km).

건조지연 계산에 필요한 표면 압력 데이터는 기상청에서 운영하는 날씨누리를 통해 수원 지역의 1시간의 시간 해상도를 갖는 데이터를 수집하였으며(KMA, 2020), 수원 관측소의 위도, 타원체고 정보를 통해 천정방향 대류권 건조지연을 계산하였다. GNSS 자료처리를 통해 추정된 ZTD와 경험적 모델로 계산한 ZHD를 이용하여 최종적으로 딥러닝 모델 학습에 필요한 ZWD를 추정하였다.

2.2 딥러닝 모델

Fig. 1은 연구 과정을 도식화한 순서도이다. 본 연구에서는 MLP와 LSTM 두 종류의 딥러닝 모델을 이용하여 대류권 습윤지연을 추정하였다. 모델 학습에 사용된 데이터는 온도, 상대 습도, 운량, 수증기압(WVP: Water Vapor Pressure)의 기상 데이터와 ZWD이다. 학습 데이터의 관측 기간은 전체 데이터의 80%에 해당하는 2015년 1월부터 9월까지이며 1시간 간격의 데이터를 사용하였다. 전체 데이터에서 학습 데이터를 제외하여 모델을 통한 ZWD를 추정하였다.

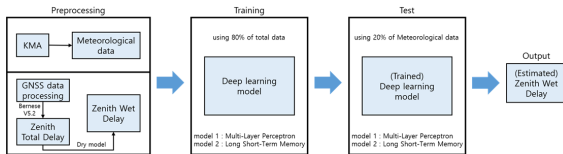


Fig. 1. Flow chart

2.2.1 MLP

MLP는 입력층과 출력층 사이에 은닉층(hidden layer)을 두어 비선형적으로 분리되는 데이터에 대한 학습이 가능하도록 고안된 모델이다. 본 연구에서는 Fig. 2와 같이 입력층과 출력층 사이에 3개의 은닉층을 추가한 MLP 모델에 기상 데이터(온도, 상대 습도, 운량, 수증기압)와 ZWD 데이터를 입력하여 학습을 진행하였다.

데이터마다 범위가 다르므로 학습 전 모든 입력 데이터의 범위를 [0, 1]로 변환시켰으며, 학습 데이터를 시간 순서대로 입력하지 않고 무작위로 넣어 시간의 영향을 고려하지 않고 기상 데이터를 통해 ZWD를 추정할 수 있도록 모델을 설계하였다. MLP 모델의 학습은 300회 반복하여 수행하였으며, 학습 과정에서 손실 그래프를 확인하며 내부 설정을 조정하였다.

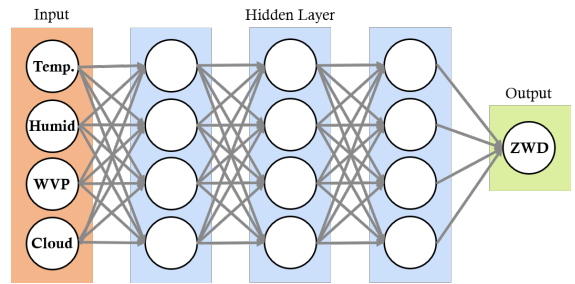


Fig. 2. Example of MLP model structure

2.2.2 LSTM

LSTM은 RNN (Recurrent Neural Network)의 한 방법으로 시계열 데이터의 학습과 예측에 좋은 성능을 내는 새로운 처리 방식이다. Fig. 3과 같은 LSTM 구조는 RNN과 같이 입력과 출력 사이 신경망이 재귀하는 구조와 더불어 4가지 레이어(삭제, 입력, 갱신, 출력)의 재귀를 통한 정보 이동을 제어하여 이전 시점이 증가하더라도 gradient가 사라지는 문제를 최소화한다(Olah, 2015; Seo, 2019).

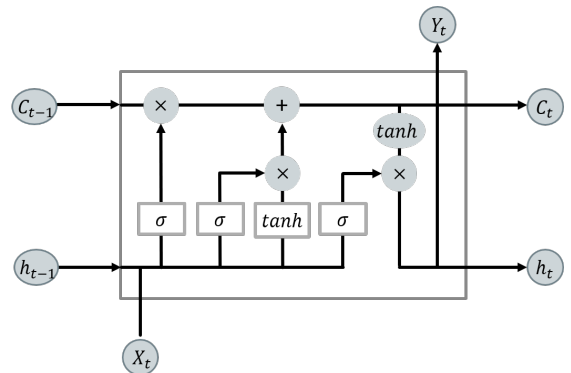


Fig. 3. Recurrent of LSTM (Olah, 2015)

본 연구에서 사용한 LSTM 모델은 여러 LSTM층과 단일 신경망으로 구성된 Stacked LSTM 모델이며, 학습 데이터는 MLP 모델과 동일한 기간의 데이터를 사용하고 과거 12시간의 데이터를 통해 이어지는 다음 1시간의 ZWD를 추정한다

록 모델을 설계하였다. MLP 모델과 동일하게 모델 학습 과정에서 데이터 종류별 최소-최대값을 [0, 1]로 맞추어 300회 반복 학습을 수행하였다.

3. 실험 및 결과

Fig. 4는 딥러닝 모델 학습을 위해 준비한 기상 데이터와 ZWD 데이터를 [0, 1] 범위로 스케일링(scaling)하여 나타낸 것이다. 습도와 운량의 경우 하루에도 크게 변화하는 값이기 때문에 범위를 축소시켰을 때 시각적으로 유의미하게 보이지 않으나, 온도와 수증기압의 경우 ZWD와 비슷한 형태의 그래프가 그려지는 것을 확인할 수 있다. 기상 데이터의 경우 모든 날짜의 데이터가 존재하지만, ZWD는 데이터 수집 과정에서 수원 상시관측소의 RINEX 파일이 없는 33일을 제외하고 플랫폼하였다(Fig. 4). 1월-9월 데이터로 학습을 진행한 모델에 10월-11월 기상 데이터를 입력하여 해당 기간의 ZWD를 추정하였으며, 기상 데이터로 추정한 ZWD와 ZTD, ZHD를 통해 계산된 ZWD를 이용하여 RMSE를 산출하였다.

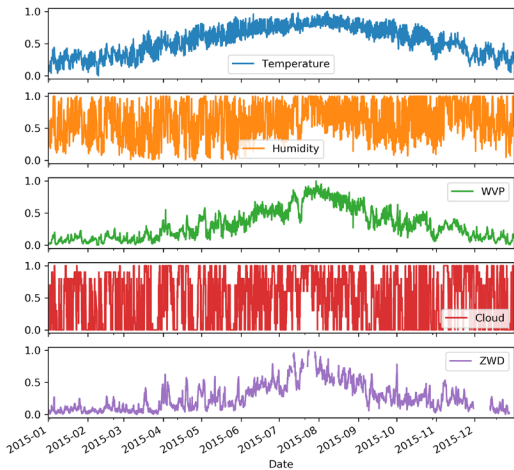


Fig. 4. Scaled data of Meteorology and ZWD

3.1 MLP 결과

MLP 모델 은닉층 내부 뉴런 수의 경우 처음에는 5, 10, 15로 작은 값을 5씩 늘려가며 조정하였으나 결과가 크게 좋아지지 않아 30, 60으로 늘린 후 다양한 조합을 만들어 진행하였다. 결과가 크게 변동이 없는 것을 확인한 뒤에 100, 150, 300의 조합으로 테스트를 진행하여 뉴런 수가 훨씬 큰 경우의 결과를 확인하였다. Table 1은 그 중 추정된 ZWD의 RMSE가 작은 순으로 4개를 선정하여 나타낸 표이다. 대부분의 경우

RMSE 3cm 수준의 결과가 산출되었는데, 뉴런 수가 100보다 큰 경우에는 오히려 RMSE가 커졌으며, 10 정도로 설정할 경우 33mm, (5, 15, 10)의 조합으로 가장 작은 31.6mm의 결과를 얻을 수 있었다. RMSE가 두 번째로 작은 (60, 30, 10) 조합 결과 그래프와 비교하면 10월 말경 발생한 큰 변화 부분을 추정할 때 (5, 15, 10) 조합의 경우 125mm, (60, 30, 10) 조합은 150mm에 가깝게 추정한 것을 통해 뉴런의 수가 변동이 큰 데이터를 추정할 때 영향을 미치는 것으로 판단된다(Fig. 5). Fig. 6은 추정된 ZWD의 오차를 나타낸 것으로 변화가 큰 구간에서의 오차는 상대적으로 (60, 30, 10) 조합에서 작지만 전반적으로 (5, 15, 10) 조합의 오차 변동 크기가 작기 때문에 추정 정확도가 더 높게 계산된 것으로 파악된다.

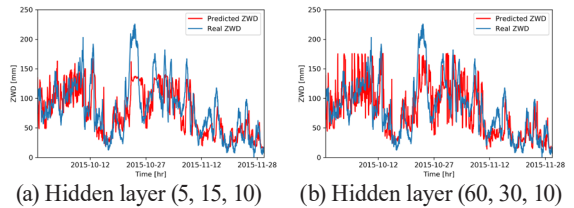


Fig. 5. Estimated ZWD using MLP model

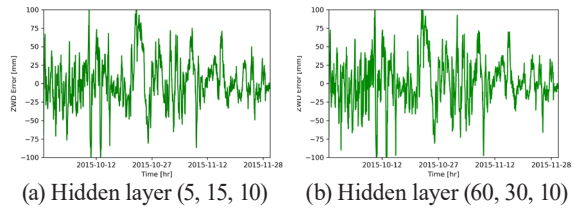


Fig. 6. Error of ZWD using MLP model

3.2 LSTM 결과

Table 2는 Stacked LSTM 모델의 LSTM 층별 뉴런과 sequence에 따른 ZWD 추정값 RMSE를 정리한 표이다. LSTM 모델의 경우 세 번째 층을 추가하지 않아도 대부분의 경우 2cm 수준의 RMSE가 산출되었으며, 다음 시간을 추정하기 위해 사용되는 과거 데이터 기간인 sequence에 따라 크게 변동하는 것을 확인하였다. sequence를 6시간, 12시간, 24시간으로 변경하며 모델을 학습시켰을 때 12시간에서 RMSE가 가장 작았다. 이는 어떤 시점의 ZWD를 추정할 때 시점 직전 12시간 데이터를 통해 추정할 경우에 추정 정확도가 가장 높음을 의미한다. MLP 모델과 마찬가지로 추정된 ZWD를 그래프로 나타내었고, LSTM 층 뉴런 또한 많을수록 추정값에 변동이 큰 것을 확인할 수 있었다(Figs. 7 and 8).

Table 1. ZWD estimation result by MLP model

	Number of neurons			RMSE [mm]
	Hidden layer 1	Hidden layer 2	Hidden layer 3	
1	5	5	5	33.91
2	5	15	10	31.60
3	60	30	10	34.40
4	100	300	150	36.78

Table 2. ZWD estimation result by LSTM model

	Number of neurons			Sequence	RMSE [mm]
	LSTM layer 1	LSTM layer 2	LSTM layer 3		
1	6	6		12	29.09
2	6	15		12	28.03
3	6	15	3	6	30.70
4	6	15	3	12	27.73
5	6	15	3	24	39.09
6	10	30	20	12	26.19

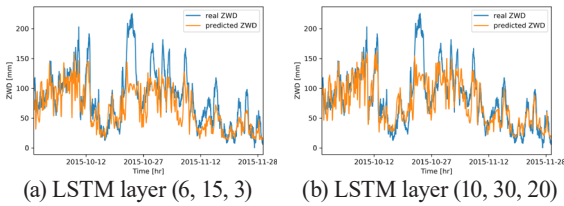


Fig. 7. Estimated ZWD using LSTM model

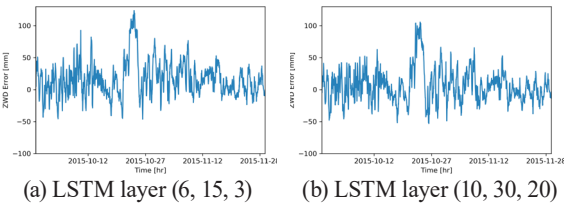


Fig. 8. Error of ZWD using LSTM model

4. 결론

본 논문에서는 딥러닝 모델인 MLP와 LSTM 모델을 사용하여 천정방향 대류권 습윤지연을 추정하였다. 딥러닝 모델을 학습시킬 데이터로 온도, 상대습도, 운량, 수증기압의 기상 데이터와 대류권 습윤지연 데이터를 사용하였다. 학습에 사

용되지 않은 기상 데이터를 입력하여 MLP 모델에서 RMSE 31.6mm, LSTM 모델에서 RMSE 26.2mm 수준으로 천정방향 대류권 습윤지연을 추정하였다. Selbesoglu (2020)는 2개의 관측소의 6종류 연간 데이터를 통해 ANN을 학습하고 ZWD를 추정하여 1cm 수준의 RMSE를 얻었다. 본 연구와는 사용한 모델과 데이터의 종류가 다르므로 결과를 비교하기에 다소 어려움이 있다. 하지만 본 연구에서 사용된 학습 데이터보다 3배정도 많은 데이터를 학습하여 산출된 결과라는 것에 초점을 맞추어보면, 더 많은 데이터를 학습에 사용할 경우 추정 정확도가 향상될 것으로 사료된다. 각 모델의 은닉층, LSTM 층 뉴런과 추정된 ZWD의 오차를 통해 뉴런이 많아질 수록 추정값 변동이 커진다는 것을 확인하였으며, 뉴런의 수가 너무 적거나 클 경우 추정 정확도가 떨어지므로 적절한 개수를 정하여 학습을 진행하는 것이 중요하다. MLP 모델과는 달리 LSTM 모델의 경우 과거 데이터에 대한 기억 기간을 조정하는 sequence에 대한 영향도 받으므로 학습 과정에서 뉴런 개수와 함께 고려해야 한다. 딥러닝을 통한 추정 시 학습 데이터의 종류 및 모델의 형태와 모델 내부 조정요소에 따라 추정 정확도에 차이가 있을 수 있음을 밝힌다.

GNSS ZWD의 경우 추정 전 GNSS 자료처리를 통해 ZTD를 추정하고 경험적 모델로 ZHD를 계산해야 하여 데이터를 취득하는 데 많은 시간이 소요된다. 본 연구는 GNSS 데이터와 기상 데이터 그리고 딥러닝을 이용하여 대류권 습윤지

연을 센티미터 수준의 정확도로 추정하였다. 이를 통해 딥러닝 모델의 세부적인 조정을 통해 정확도가 충분히 향상된다면 ZWD 추정을 위해 수행해야 하는 과정과 소요 시간을 줄일 수 있다는 가능성을 제시하였다. 또한 ZWD가 합리적인 수준으로 추정이 가능할 경우 GNSS 기반 기상 연구를 위한 데이터 수집 목적으로 사용될 수 있을 것으로 기대된다. 대류권 지연은 대기 중 수증기에 영향을 많이 받아 기상 변동이 클 경우 추정값과 차이가 크게 발생하며, 우리나라의 기상은 계절별로 비슷한 변화를 보이는 특성이 있다. 본 연구는 내륙지역의 한 예시만을 나타내어 지역적인 특성의 영향을 분석하는데 어려움이 있다. 또한 일년의 데이터를 사용하여 추정과정에서 계절적인 변화를 고려하지 못한 점을 한계로 들 수 있다. 향후에는 내륙지역보다 기상 변동이 큰 해안지역의 데이터를 함께 사용하고 대류권 지연의 연간 특성을 고려하기 위해 장기간의 데이터를 이용한 연구가 추가적으로 진행할 필요가 있다고 사료된다.

감사의 글

본 논문은 해양수산부 국가R&D사업 “지상기반 센티미터급 해양 정밀 PNT 기술개발”로 수행된 연구 결과입니다 (20200451).

References

- Bae, T.-S. (2018), Accuracy analysis of GNSS-based public surveying and proposal for work processes, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 36, No. 6, pp. 457-467.
- Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R., and Ware, R.H. (1992), GPS meteorology-remote sensing of atmospheric water vapor using the Global Positioning System, *Journal of Geophysical Research*, Vol. 97, No. D14, pp. 15787-15801.
- GDC (2020), Global Navigation Satellite System Integrated Data Center home page, <http://gnssdata.or.kr> (last date accessed: 20 November 2020).
- Hofmann-Wellenhof, B., Lichtenegger, H., and Collins, J. (2001), *GPS Theory and Practice*, Springer-Verlag Wien New York, Wien, Austria, 382p.
- Jason, H. (2020), Using machine learning to Nowcast precipitation in high resolution, *Google AI Blog*, <https://ai.googleblog.com/2020/01/using-machine-learning-to-nowcast.html> (last date accessed: 10 November 2020).
- Katsougiannopoulos, S. and Pikridas, C. (2009), Prediction of zenith tropospheric delay by multi-layer perceptron, *Journal of Applied Geodesy*, Vol. 3, No. 4, pp. 223-229.
- Kim, S.-K. and Bae, T.-S. (2012), Long-term analysis of tropospheric delay and ambiguity resolution rate of GPS data, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 30, No. 6-2, pp. 673-680.
- Kim, J.S. and Bae, T.-S. (2015), Comparative analysis of GNSS precipitable water vapor and meteorological factors, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 33, No. 4, pp. 317-324.
- Kim, H.-U. and Bae, T.-S. (2017), Preliminary study of deep learning-based precipitation prediction, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 35, No. 5, pp. 423-429.
- Kim, S.-K. and Bae, T.-S. (2018), Long-term GNSS analysis for local geodetic datum after 2011 Tohoku earthquake, *The Journal of Navigation*, Vol. 71, No. 1, pp. 117-133.
- KMA (2020), Korea Meteorological Administration home page, <http://www.kma.go.kr/home/index.jsp> (last date accessed: 10 November 2020).
- Nam, J.Y. and Song, D.S. (2019), Retrieval biases analysis on estimation of GNSS precipitable water vapor by tropospheric zenith hydrostatic models, *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 37, No. 4, pp. 233-242.
- Olah, C. (2015), Understanding LSTM Networks, Colah's Blog, <http://colah.github.io/posts/2015-08-Understanding-LSTMs> (last date accessed: 23 December 2020).
- Saastamoinen, J. (1973), Contribution to theory of atmospheric refraction, *Bulletin Géodésique*, Vol. 107, pp. 13-34.
- Selbesoglu, M.O. (2020), Prediction of tropospheric wet delay by an artificial neural network model based on meteorological and GNSS data, *Engineering Science and Technology, an International Journal*, Vol. 23, No. 5, pp. 967-972.
- Seo, Y.-M. (2019), *Analysis of Prediction Accuracy of Fine Dust Concentration for Seoul Region Using LSTM Model*, Master's thesis, Sejong University, Seoul, Korea, 77p.