

Preface

Hi reader of this proposal!

When reading this pre-proposal for the thesis ring, I would highly appreciate feedback on the Problem Description, the Research Questions (as I'm not sure how to differentiate those from the objectives) and the Method section. For the Method section, I'm particularly interested in hearing how the explanation of Neural Networks is received (subsections 5.1 and 5.2). Thanks in advance!

Improving rainfall rate predictions via Commercial Microwave Link signals in Sri Lanka using Deep Transfer Learning

Ludo Diender

MSc thesis proposal

Hydrology and Quantitative Water Management Group, Wageningen University

1 Problem description

Having ample and correct precipitation data is important for a plethora of applications, including flood warnings, agriculture, river safety, shipping routes and many more (Chwala & Kunstmann, 2019). In urban areas, an even higher spatial and temporal resolution of rainfall is needed, due to the complex urban hydrology system (Overeem, Leijnse, & Uijlenhoet, 2011). Due to their high density in populated areas, Commercial Microwave Links (CML) can help in retrieving precipitation data. In the past 25 years, CML have been recognized as a valuable opportunistic method to measure rainfall (Leijnse, Uijlenhoet, & Stricker, 2007) (Ruf, Aydin, Mathur, & Bobak, 1996). CML are back-haul links used by telecommunication companies to transfer information from one telecommunication station to the next using microwave signals. The links' signals get attenuated by rainfall by means of scattering and absorption. This attenuation is measured and stored by the telecommunication companies for monitoring purposes and can be used to retrieve path-averaged rainfall rates. In data scarce areas, where little precipitation is measured, CML have proven to be an excellent addition to precipitation data (Overeem et al., 2021; Doumounia, Gosset, Cazenave, Kacou, & Zougmore, 2014; Diba, Samad, Ghimire, & Choi, 2021).

The first studies on the use of CML signals to retrieve rainfall rates were done by using a specific Power-Law (PL) to relate the attenuation of the signal and the rainfall rate (Overeem et al., 2011; Leijnse et al., 2007). This method, which includes a wet-dry classification, baseline estimation, wet antenna attenuation estimation and finally a rainfall rate retrieval, has yielded good results in multiple studies (de Vos, Overeem, Leijnse, & Uijlenhoet, 2019; Graf, Chwala, Polz, & Kunstmann, 2020; Fencel, Dohnal, Rieckermann, & Bareš, 2017). Recently methods have shifted away from this PL algorithm and used a more data-driven approach in the form of neural networks of different architectures. Neural networks, or Deep Learn-

ing, are a subpart of Machine Learning that is inspired by the neural structure of the brain. A neural network consists of different layers of nodes (neurons) that are connected to each other. Similar to Machine Learning, a neural network is able to learn on its own, without intervention from a researcher. Studies have been performed in Sweden, Israel (Habi, 2019), Germany (Polz, Chwala, Graf, & Kunstmann, 2020), South Korea and Ethiopia (Diba et al., 2021) on the use of such data-driven networks in relating CML signals to rainfall rates. Previous studies have shown that data-driven models can be more accurate, less time-demanding and more robust in estimating rainfall rates compared to the PL method (Polz et al., 2020; Pudashine et al., 2020). Neural networks are not a novelty in predicting rainfall (French, Krajewski, & Cuykendall, 1992), but the application to CML data has recently been rising in popularity.

One of the disadvantages of using data-driven methods like neural networks, apart from the black-box characteristics inherent to the method, is the dependency on a large training data set. In areas with less or little available training data, transfer learning provides the opportunity to adapt an already existing model with a certain structure to do a slightly different task (Tan et al., 2018). The concept of transfer learning is based on the way humans learn. When humans learn a new task, we do not start from scratch. We use previous knowledge and skills to quickly adapt to the new task. When learning to ride a motorcycle, it helps if you already know how to ride a bike. Concepts like balance and steering can be transferred from the latter task to the former, thus speeding up the learning process. The technique exploits the availability of data in the source domain and is able to transfer that knowledge to the target domain. It does so by relaxing the underlying assumption that training and test data for a Machine Learning model should be independent and identically distributed (Weiss, Khoshgoftaar, & Wang, 2016). A conceptual picture of transfer learning can be seen in figure 1 Although used quite often in dif-

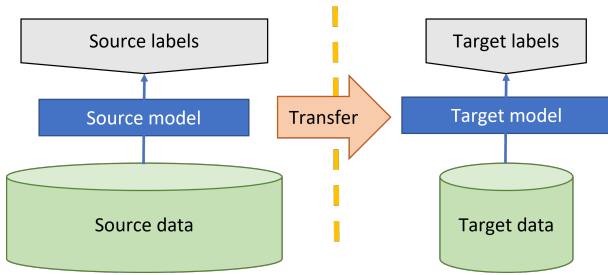


Figure 1: Concept of transfer learning. Adapted from (Sarkar, 2018)

than the currently used interpolation methods? In other words, is it worth the effort?

3 Field site and data

This research will make use of datasets in two different countries, with a slightly different structure.

ferent applications (Zhuang et al., 2021), the concept of transfer learning has not yet been used to improve the precipitation estimation using CML in data-scarce areas.

Recent research focused on the use of CML data to measure rainfall in tropical regions, more specifically Sri Lanka (Overeem et al., 2021) and Brazil (Gaona, Overeem, Raupach, Leijnse, & Uijlenhoet, 2017a). The relatively small amount of reference rain gauges in Sri Lanka especially made this research more challenging compared to well-equipped countries like the Netherlands (Overeem, Leijnse, & Uijlenhoet, 2013). Both of the two studies mentioned above (Sri Lanka and the Netherlands) are based on the PL method. There have not been any efforts yet to analyze the potential of CML for rainfall retrieval using data-driven methods for neither the Netherlands nor Sri Lanka.

2 Research objectives and questions

The questions asked in this research are the following: 1) How does a neural network perform on Dutch CML data in retrieving rainfall rates? By training, testing and validating a neural network on the Dutch CML data, this research aims to find out if such a data-driven approach yields better results compared to more traditional approaches.

2) How does transfer learning improve the use of CML in Sri Lanka for measuring precipitation? As transfer learning has the potential to improve the learning curve of a neural network significantly, this question is aimed to find out if applying the concept of transfer learning is useful and worth it. Is there a significant benefit to this method that would make it preferable over other more traditional methods?

3) What is the potential of the use of neural networks for 2D interpolation of rainfall maps? A neural network architecture can be used to detect unknown patterns in rainfall distributions and use those to improve the interpolation of rainfall maps. Is this significantly better

3.1 The Netherlands

The Dutch data used in this research is similar to previous research (Overeem, Leijnse, & Uijlenhoet, 2016). The first research question will make use of the CML data from NEC and NOKIA microwave links, operated by T-Mobile. Both datasets span the period of 14 January 2011 to 30 July 2013, roughly 2,5 years. The data contains 3101 links, but as shown before (Overeem et al., 2016), not all of these links will be proven useful or available. The total number of actual links used will therefore be lower, depending on the quality. Received Signal Level (RSL) is stored at 15-minute intervals, at which the minimum and maximum signal level over the time period are stored. The power resolution of the data is 1 dB (NOKIA) and 0.1 dB (NEC). Apart from the dynamic time signal data, static metadata is included as well. This includes coordinates of the start and end point of the link, distance between these points and frequency per link.

As a reference, a gauge-adjusted radar rainfall dataset is used (freely available as "Radar precipitation climatology" via <http://climate4impact.eu>) with a resolution of 0.9 km². The radar is adjusted with the use of both the automated and manual rain gauge network operated by the KNMI (Royal Dutch Meteorological Institute).

3.2 Sri Lanka

The Sri Lankan data is similar to previous research (Overeem et al., 2021). The CML data in Sri Lanka spans 3,5 months (12 September 2012 to 31 December 2012) and similar to the Dutch data, is logged once every 15 minutes (minimum and maximum RSL). The dataset consists of on average 1140 link paths. Static data include start and end point of the link, link length and average frequency. Although the data is available throughout the whole country of Sri Lanka, the network density around the biggest city Colombo is significantly higher. PICTURE OF THE NETWORK IN SRI LANKA AND THE NETHERLANDS

4 Methods

4.1 Neural networks

To conduct this research, I will make use of a neural network. Neural networks are part of deep learning. An input signal is transferred through different layers of the model. Every layer consists of different neurons (or nodes) that are able to extract features from the input signal. Every node consists of a weighted sum of the complete input signal. The weights that are given to the different values in this input signal, together with a bias that is added to this weighted sum, make up the parameter set of the neuron. The subsequent layers extract combinations out of the previous layer, until finally the network ends up with a prediction with a certain probability. By applying a loss function, the network can learn how to improve, by changing the different parameters in all neurons. The to-be-extracted features are unknown beforehand to the researcher. The network itself learns which features are interesting, separable and are helpful in classifying the specific signal. This is the cornerstone of machine learning/deep learning.

In this research I will use a Long Short Term Memory (LSTM) architecture for my neural network, which is part of a Recurrent Neural Networks (RNN). RNN a characterized by the recurrent use of the same bit of network, to keep on improving the prediction. The network has one or two layers, depending on the complexity of the features and the signal. The signal is then processed multiple times by the same two layers. RNN are designed for sequential data like the time signals that are used in this research. A disadvantage of RNN is the risk of vanishing or exploding gradients, due to the lack of memory in the network. LSTM resolve this issue by adding a memory to the model by using different gates to combine old and new data on every recurrent step of the network. The use of LSTM to create a network for CML data has been demonstrated before (Habi, 2019; Diba et al., 2021; Pudashine et al., 2020). The networks used in those studies will serve as a source of inspiration for this research. Other types of neural networks have been proposed as well (Polz et al., 2020), but as RNN (and specifically LSTM) are designed to work with sequential data, those are preferred in this research.

4.2 Training, testing and transferring

The Dutch dataset as described in section 3 will be split up in two sets. One is dedicated to training the model and the other one will be used for testing. The training data is passed through the LSTM, the loss is calculated for each training sample and the network 'learns' (is updated) by

using backward propagation. Backward propagation is a machine learning technique where the final layers are updated first and the updates move backwards through the model.

After training for a certain number of epochs (evolutions) the model is tested using the test data. The performance on the testing data is the most important and determines the performance of the model overall.

Once the LSTM performs well, transfer learning will be applied, to overhaul the information and knowledge gained by the model to quickly adapt to Sri Lankan data. By removing the outer layer of the model, the feature extraction part remains and can be used for a different dataset with different characteristics as well. The transferred model will, similarly to the first model, be trained on a subset of the Sri Lankan dataset. Afterwards, it will be tested on the remaining data to evaluate the potential of the transferred model.

Finally, both the trained model on the Dutch CML and the transferred model on the Sri Lankan data will be compared to the current performance of CML algorithms (Overeem et al., 2011, 2021), which are bundled in the openly available R-package RAINLINK. (SOURCEEEEE). During this comparison, focuspoints will be the overall match with the reference dataset and the performance during periods with rather low or high rainfall rates. Low rainfall rates have been a challenge in previous CML research, as the noise in the CML signal is harder to distinguish from small rain-events (Uijlenhoet, Overeem, & Leijnse, 2018). The performance during periods of higher rainfall is expected to be similar, as the signal attenuation from large rain-events is quite clear.

4.3 2-D interpolated rainfall maps

CNN's have been proposed to deal with spatial patterns in rainfall data (Sadeghi et al., 2019). Already 20 years ago, advances have been made in using neural networks for spatial interpolation of environmental variables (Rigol, Jarvis, & Stuart, 2001)). In this research, a similar neural network structure will be proposed to interpolate the rainfall data obtained from CML. The interpolated map will be compared to a reference map created by using a gauge-adjusted radarproduct. These neural networks need terrain information next to the data from the CML. The resulting interpolated maps are compared to more commonly used interpolation methods to see if the extra effort of using neural networks yields maps that more closely resemble the reference map.

5 Timetable

Adapt the Table below to make it specific for your project (or make your own). Set deadlines for the products. Be as specific as possible: mention when you will collect which data /do which model runs / write which parts of the report. It often helps to link activities and products to your sub-questions. A specific planning can help later on to see if you are on schedule or that you should e.g. shorten a certain data-processing step or stop calibrating your model, so you have enough time to do the analyses and answer your research questions. It often helps to link the tasks to the methodology (and therefore to the research questions). Specify special conditions: are you planning to take courses, vacation, etc.

References

- Chwala, C., & Kunstmann, H. (2019). Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges [Journal Article]. *WIREs Water*, 6(2), e1337. doi: <https://doi.org/10.1002/wat2.1337>
- de Vos, L. W., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2019). Rainfall estimation accuracy of a nationwide instantaneously sampling commercial microwave link network: Error dependency on known characteristics [Journal Article]. *Journal of Atmospheric and Oceanic Technology*, 36(7), 1267-1283. doi: 10.1175/JTECH-D-18-0197.1
- Diba, F., Samad, M., Ghimire, J., & Choi, D. (2021). Wireless telecommunication links for rainfall monitoring: Deep learning approach and experimental results [Journal Article]. *IEEE*, 9, 11. doi: 10.1109/ACCESS.2021.3076781
- Doumounia, A., Gosset, M., Cazenave, F., Kacou, M., & Zougmore, F. (2014). Rainfall monitoring based on microwave links from cellular telecommunication networks: First results from a west african test bed [Journal Article]. *Geophysical Research Letters*, 41(16), 6016-6022. doi: <https://doi.org/10.1002/2014GL060724>
- Fencl, M., Dohnal, M., Rieckermann, J., & Bareš, V. (2017). Gauge-adjusted rainfall estimates from commercial microwave links [Journal Article]. *Hydrol. Earth Syst. Sci.*, 21(1), 617-634. (HESS) doi: 10.5194/hess-21-617-2017
- French, M. N., Krajewski, W. F., & Cuykendall, R. R. (1992). Rainfall forecasting in space and time using a neural network [Journal Article]. *Journal of Hydrology*, 137(1), 1-31. doi: [https://doi.org/10.1016/0022-1694\(92\)90046-X](https://doi.org/10.1016/0022-1694(92)90046-X)
- Gaona, M., Overeem, A., Raupach, T., Leijnse, H., & Uijlenhoet, R. (2017a). Rainfall retrieval with commercial microwave links in sao paulo, brazil [Journal Article]. *Atmos. Meas. Tech.*, 11(7), 11. doi: 10.5194/amt-11-4465-2018
- Graf, M., Chwala, C., Polz, J., & Kunstmann, H. (2020). Rainfall estimation from a german-wide commercial microwave link network: optimized processing and validation for 1 year of data [Journal Article]. *Hydrol. Earth Syst. Sci.*, 24(6), 2931-2950. (HESS) doi: 10.5194/hess-24-2931-2020
- Habi, V. (2019). *Rain detection and estimation using recurrent neural network and commercial microwave link* (Thesis).
- Leijnse, H., Uijlenhoet, R., & Stricker, J. N. M. (2007). Hydrometeorological application of a microwave link: 2. precipitation [Journal Article]. *Water Resources Research*, 43(4). doi: <https://doi.org/10.1029/2006WR004989>
- Overeem, A., Leijnse, H., Leth, T., Bogerd, L., Priebe, J., Tricarico, D., ... Uijlenhoet, R. (2021). Tropical rainfall monitoring with commercial microwave links in sri lanka [Journal Article]. *Environmental Research Letters*, 16. doi: 10.1088/1748-9326/ac0fa6
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2011). Measuring urban rainfall using microwave links from commercial cellular communication networks [Journal Article]. *Water Resources Research*, 47(12). doi: <https://doi.org/10.1029/2010WR010350>
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2013). Country-wide rainfall maps from cellular communication networks [Journal Article]. *Proceedings of the National Academy of Sciences*, 110(8), 2741-2745. doi: 10.1073/pnas.1217961110
- Overeem, A., Leijnse, H., & Uijlenhoet, R. (2016). Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network [Journal Article]. *Atmos. Meas. Tech.*, 9(5), 2425-2444. (AMT) doi: 10.5194/amt-9-2425-2016
- Polz, J., Chwala, C., Graf, M., & Kunstmann, H. (2020). Rain event detection in commercial microwave link attenuation data using convolutional neural networks [Journal Article]. *Atmos. Meas. Tech.*, 13, 18.
- Pudashine, J., Guyot, A., Petitjean, F., Pauwels, V. R. N., Uijlenhoet, R., Seed, A., ... Walker, J. P. (2020). Deep learning for an improved prediction of rainfall retrievals from commercial microwave links [Journal Article]. *Water Resources Research*, 56(7), e2019WR026255. doi: <https://doi.org/10.1029/2019WR026255>
- Rigol, J. P., Jarvis, C. H., & Stuart, N. (2001). Ar-

Table 1: Schedule of the project.

- tificial neural networks as a tool for spatial interpolation [Journal Article]. *International Journal of Geographical Information Science*, 15(4), 323-343. doi: 10.1080/13658810110038951
- 370 Ruf, C. S., Aydin, K., Mathur, S., & Bobak, J. P. (1996). 35-ghz dual-polarization propagation link for rain-rate estimation [Journal Article]. *Journal of Atmospheric and Oceanic Technology*, 13(2), 419-425. doi: 10.1175/1520-0426(1996)013<0419:Gdpplf>2.0.Co;2
- 375 Sadeghi, M., Asanjan, A. A., Faridzad, M., Nguyen, P., Hsu, K., Sorooshian, S., & Braithwaite, D. (2019). Persiann-cnn: Precipitation estimation from remotely sensed information using artificial neural networks-convolutional neural networks [Journal Article]. *Journal of Hydrometeorology*, 20(12), 2273-2289. Retrieved from <GotoISI>://WOS:000499430000001 (Jr1yt Times Cited:30 Cited References Count:65) doi: 10.1175/Jhm-D-19-0110.1
- 380
- 385 Sarkar, D. (2018, November 14, 2018). *A comprehensive hands-on guide to transfer learning with real-world applications in deep learning* [Web Page]. Retrieved from <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>
- 390
- 395 Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning [Conference Proceedings]. In (p. 270-279). Springer International Publishing.
- Uijlenhoet, R., Overeem, A., & Leijnse, H. (2018). Opportunistic remote sensing of rainfall using microwave links from cellular communication networks [Journal Article]. *WIREs Water*, 5(4), e1289. doi: <https://doi.org/10.1002/wat2.1289>
- 400
- Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning [Journal Article]. *Journal of Big Data*, 3(1), 9. doi: 10.1186/s40537-016-0043-6
- 405
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... He, Q. (2021). A comprehensive survey on transfer learning [Journal Article]. *Proceedings of the IEEE*, 109(1), 43-76. doi: 10.1109/JPROC.2020.3004555
- 410