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

In the past 25 years, Commercial Microwave Links (CML) have been recognized as a valuable opportunistic method to measure rainfall (Leijnse, Uijlenhoet, & Stricker, 2007) (Ruf, Aydin, Mathur, & Bobak, 1996). The links' signals get attenuated by rainfall by means of scattering and absorption. This attenuation is measured and stored by telecom companies for monitoring purposes and can be used to retrieve rainfall rates. Even more so in data scarce areas, where little precipitation is measured, CML's 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). Their density in populated areas makes them especially useful for urban hydrology where high spatial and temporal resolution of precipitation data is required (Overeem, Leijnse, & Uijlenhoet, 2011). In general, having ample and correct precipitation data is important for flood warnings, agriculture, river systems, shipping routes and many more (Chwala & Kunstmann, 2019).

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; Fencl, 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. 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 such 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 is 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 sligthly different task (Tan et al., 2018). The technique exploits the availability of data in the source domain and is able to transfer the 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). Although used quite often in different applications (Zhuang et al., 2021), the concept of transfer learning has not yet been used to improve the precipitation estimation using CML's 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 algebraic PL method. There have not been any efforts yet to analyze the potential of CML's for rainfall retrieval using data-driven methods for neither the Netherlands nor Sri Lanka.

2 Research objectives

The objectives of this research are to:

- train, test and validate a neural network on CML data in the Netherlands to obtain precipitation rates
- apply the concept of transfer learning to make the model able to obtain precipitation rates for Sri Lanka
- create 2D interpolated rainfall maps using the neural network architecture
- compare the data-driven approach to previous approaches.

With these four objectives, this research will give an indication of the potential of transfer learning and data-driven methods for areas with little reference data like Sri Lanka.

3 Research questions

The research questions all relate to the objectives mentioned above.

- How does a neural network perform on Dutch data?
- How does transfer learning improve the use of CML's in Sri Lanka as precipitation measurements?
- What is the potential of the use of NN for 2D interpolation of rainfall maps
- Does a neural network outperform previous methods like the power-law algorithm in retrieving rainfall rates from CML's?

4 Field site and data

I will write this section later, skip this for now when reading! Making use of Aart's dataset of the Netherlands and the provided data for Sri Lanka. Combination makes the research come together. Explain the type of signal data that I got (quantity, quality, metadata)

5 Methods

5.1 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's 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's are designed for sequential data like the time signals that are used in this research. A disadvantage of RNN's is the risk of vanishing or exploding gradients, due to the lack of memory in the network. LSTM's 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's to create a network for CML data has been demonstrated before (Habi, 2019; Diba et al., 2021; Pudashine et al., 2020). Other types of neural networks have been proposed as well (Polz et al., 2020), but as RNN's are designed to work with sequential data, those are preferred in this research.

5.2 Training, testing and transferring

The Dutch dataset as described in section 4 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, number yet to be determined) 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.

5.3 2-D interpolated rainfall maps

This section still needs some extra attention, please skip this for now CNN's have been proposed to deal with spatial patterns in rainfall data (Sadeghi et al., 2019). Could use radar images to learn the patterns of the Netherlands and subsequently use the CML data to recognize these patterns and create a rainfall map for the whole country using these CNN's. Might not result in anything different than the maps that are already made using the IDW or Nearest Neighbour techniques.

5.4 Comparison to PL algorithm

This section still needs some extra attention, please skip this for now

6 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

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 [Jour-

Table 1: Schedule of the project.

nal 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

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

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

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 `<Go to ISI>://WOS:000499430000001` (Jr1yt Times Cited:30 Cited References Count:65) doi: 10.1175/Jhm-D-19-0110.1

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.

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

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