Atmos. Meas. Tech. Discuss., 8, C2643–C2651, 2015
www.atmos-meas-tech-discuss.net/8/C2643/2015/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network" *by* A. Overeem et al.

Anonymous Referee #1

Received and published: 19 August 2015

SUMMARY: A method, a data set and a code (in the statistical programming language R) for the estimation of country-wide rainfall fields from commercial microwave links in the Netherlands is provided.

RECOMMENDATION: reject and resubmit

I don't think this paper merits publication. It is poorly written and structured, contains no new ideas, concepts or methods and is essentially reheated content from previous publications. The only new part is the R code that the authors share and the small data sample that comes with it. The code works but does nothing groundbreaking. It is highly specific to the Netherlands and contains many hard-coded empirically estimated parameters and thresholds that make it hard to transfer to other locations. I should also

point out that the code does not comply with the general guidelines for programming in R and has not been tested properly. I strongly support and encourage data transmission and code sharing within the scientific community. But I seriously doubt that this contribution is going to be useful. I invite the authors to rethink their approach and resubmit a new draft that takes into account the numerous comments below. Most importantly, I think the authors should take the time to properly test their code and evaluate their methods before distributing them to the rest of the community.

GENERAL COMMENTS:

GC1: Most of Section 3 is just reheated content from previous publications. What's the point of repeating the description of a methodology that has been introduced in earlier papers? Please refer to the literature, point out the differences and be more concise!

GC2: The methodology contains many arbitrary thresholds and empirically estimated quantities. The dry/wet classification alone relies on 4 empirical thresholds (15 km, -0.7 dB/km, -1.4 dB and 2 dB). There's also 2 for the baseline estimation (2.5 and 6 h), 2 for the rain rate retrieval (Aa and alpha) and many others for the spatial interpolation. The paper reads like a "cookbook" rather than a rigorous and scientifically tested procedure.

GC3: The employed interpolation technique (i.e., ordinary kriging) relies on strong stationarity and isotropy assumptions that are clearly violated by the data (look for example at Figure 9). The authors admit this limitation in the conclusions, but propose no alternative method.

GC4: The authors make little efforts to test their code, e.g., using different input parameters, datasets and sampling resolutions. There's clearly more work to be done here.

GC5: The fact that the MWL data alone can not be used to derive rainfall maps is a serious limitation. According to the authors, one of the goals of the paper is to "promote the application of rainfall monitoring using microwave links in poorly gauged

regions around the world". But if rain gauges are necessary to compute climatological variograms, how is this supposed to transfer to poorly gauged regions? This needs to be discussed in more detail. More generally, I think the code should include at least one alternative method for the estimation of rainfall maps based solely on the link data. If this turns out to be unfeasible, the paper should focus on how MWLs can be used to improve rainfall estimates, rather than promoting their use independently of gauges and radar.

GC6: The isotropic variogram model is clearly not adequate for the rainfall accumulation field shown in Figure 9 which exhibits clear anisotropy along the NE direction. So the question is why do you use it at all?

GC7: Section 4 (conclusions) still introduces new pieces of information that can hardly be qualified as "conclusions". It would probably make more sense to separate this section into two parts: a discussion and a conclusion.

GC8: Given the tremendous difficulties and the numerous uncertainties involved in the retrieval and mapping methods, wouldn't it make more sense to go back to simpler aspects and more specific questions that can be answered precisely rather than jumping intermediate steps and distributing something that has not been properly tested?

SPECIFIC COMMENTS:

SC1: p.8192, l.10: Remove the global scale. Continental scale already seems ambitious enough. The global scale is not realistic because there are no MWLs over the oceans.

SC2: p.8192, ll.20-22: please provide some references to recent satellite products (e.g., GPM IMERG), together with a discussion of their advantages and limitations with respect to MWLs.

SC3: p.8192, ll.22-23: "This calls for alternative and complementary sources of rainfall information". Ok, it's always nice to have more data, but what about the quality and

consistency? Rain gauges, radars and satellites are specifically designed to measure rainfall. They come with high standards and are operated by qualified national agencies. MWLs are operated by private telecommunication companies, with little to no interest in sharing their data. They are not designed to measure rainfall, are poorly maintained and at risk of being replaced by fiber optics. So what is your main argument in favor of MWLs? What I am saying is that it would be a good idea to be more specific about the type of applications that could benefit from MWL rainfall estimates and about the reasons why the authors think we should promote such a technology. To give some perspective: the latest GPM-IMERG rainfall product now covers most of the globe at resolutions in the order of 10km and 30min. This includes remote areas and the ocean where there are no MWLs.

SC4: p.8193, ll.23-24 "Note that the link data are utilised in a stand-alone fashion to obtain rainfall information": I guess one can debate the meaning of stand-alone. But to me, it seems that the method heavily relies on previous parameters computed from rain gauges and radar. If this is supposed to be the next step towards a continental application, then the authors should address more carefully how they intend to estimate and transfer the model parameters to places with little or no rain gauge or radar coverage.

SC5: p.8194, l.9 and others: write transmitted/transmission power instead of "transmit power".

SC6: p.8194, ll.17-18: "The majority of the provided links does not exist anymore due to network renewal and deployment of underground fiber optical cable networks": What is the point of promoting a technology that is disappearing? Is there so much to gain?

SC7: p.8195, section 2.1.2: instead of throwing in all the variables in a random order, you might want to give them in a more organized way, by grouping them into categories: For example physical properties (frequency, polarization), geographical information (link identifier, coordinates of antennas, path length, altitude) and measure-

ments (Pmin/Pmax, time/date, . . .).

SC8: p.8196, section 3.1: point 2, i.e., the numbering of the time intervals, is not clear. Please revise.

SC9: p.8197, l.3: the exponent b has not been formally defined at this point of the paper. Please introduce it properly or refer to the point in the text where it is defined.

SC10: p.8197, l.7: a better way would be to check if the overlapping measurements are equal before removing them. If they are all equal to each other, you can keep one of them and delete the others.

SC11: p.8198, section 3.2: You may want to consider a more specific name for the dry/wet classification method than the "link approach". After all, most of these classification techniques use the link data. I suggest something like "nearest neighborhood" dry/wet classification.

SC12: p.8199, l.1: "meant as a successful illustration". This does not make any sense. Please reformulate.

SC13: p.8199, l.7, l.22: Why 6 h?

SC14: pp.8199-8200: shouldn't the thresholds in steps 6 and 7 depend on the frequency of the links and the temporal resolution? Moreover, how sensitive are the results to these thresholds?

SC15: p.8200, l.23: why 2.5h?

SC16: p.8206, l.22: "indispensable" is probably too strong in this context. Please rephrase.

SC17: p.8207, ll.3-14: you should also consider the possibility that a, b, alpha might not be adequate for this event.

SC18: p.8209, l.9: "and D is the duration (h)". The duration of what?

SC19: p.8209, l.9: The nugget is basically the variance at zero distance. No, the variance at zero distance is equal to zero. Please rephrase.

SC20: p.8210, l.7: "After it has been explained how the data". Bad English. Please reformulate.

SC21: p.8210, l.12: "and must be the same as used in program". Bad English. Please reformulate.

SC22: p.8210, ll.22-23: "In general the link network is able to correctly determine the spatial rainfall patterns". Yes, but how much of this "correct" spatial pattern actually comes from the gauge-derived variogram?

SC23: p.8214, ll.22-24: Yes indeed. This also means that, as a stand-alone solution, MWLs are not that great. So what is your strategy for applying the method in regions with no rain gauge or radar data?

SC24: p.8215, ll.8-16: This paragraph conveys a very strong negative message. It essentially states that (a) the interpolation method is not adequate (b) not transferable, (c) based on poor mathematical assumptions that are obviously violated and (d) the method needs to be substantially improved. This is not very encouraging. Yet it does not seem to bother the authors too much.

TYPOS (this is not an exhaustive list):

p.8192, l.19: there is a missing parenthesis.

p.8194, l.21: the location of the links

p.8195, l.26: The programs explicitly employ

p.8196, l.5: Overeem at al. (2013) used a gauge-adjusted

p.8197, l.4: with microwave frequency

p.8200, l.16: Now that the rainy and non-rainy . . .

p.8206, l.7: Overeem (2011) optimised

p.8206, l.10: They compared daily link-based . . .

p.8209, ll.17-18: corresponds to the same row

p.8214, l.1: Preferably a rain gauge or a disdrometer should be available near the antenna.

R-CODE:

Here are some general suggestions that could help improve the code: (1) use shorter variable names (2) Wrap long lines (3) Use the assignment operator ← instead of = (4) Avoid hard coding, i.e., put the important input parameters in a separate config.R file, then source this file at the beginning of the main script. In this way, users can modify the variables more easily without having to modify the main code.

Preprocessing_Linkdata.R

ll.40-41, l.44, l.54, ll.70-71: please put the definitions of MinFrequency, MaxFrequency, Timestep, PERIOD, X_middle and Y_middle into a separate parameter input file. Same for the projection information on l.178. If a user wants to change these values, they need to be clearly separated from the rest of the code.

l.46: what happens if the user changes the time step and 60/Timestep is not an integer anymore? Same for NrStepsDay on l.48 and in the other scripts.

l.64: range = seq(from=2,to=2,by=1) what's this good for? And why is it hard-coded?

ll.130-131: a (slightly) more efficient way of doing this would be: selection = which(ID==LINK_ID[i]) cond = timestep[selection]

ll.134-135 (consistency checks): a better way would be to check if the overlapping measurements are equal before removing them. If they are all equal to each other, you can keep one of them and delete the others.

WetDryClassification_LinkApproach.R

l.36: please put the definition of Timestep into a separate parameter input file.

ll.43-36: the factors 2, 6 and 2.5 should be input parameters and not hard-coded.

ll.143-155 these multiple assignments are very confusing. This is bad programming practice.

ll.164-167: Here you risk ending up with vectors of different lengths. Please check if this is ok.

l.176: distance_limit should be defined in a separate input parameter file.

ll.301-302: the thresholds -0.7 and -1.4 should be in a separate input parameter file.

RainfallRetrieval_Links.R

l.29: please put the definition of Timestep into a separate parameter input file.

l.63: the threshold -32.5 should not be hard-coded but in a separate parameter input file.

ll.81-92: This might lead to vectors of different lengths depending on where the missing values are. For example, x_start and y_start might not correspond to each other anymore. Check if that's ok.

l.98: Please put Aa in a separate input parameter file. Same for alpha on l.104

l.99: some parentheses might be a good idea: cond_max = {Am_max > Aa}

l.110 why is there a hard-coded 0.25 here? Shouldn't this be Timestep/60?

Interpolation_Rainfall_Intensities.R

l.35, ll.67-68: please put the definitions of Timestep, X_middle, Y_middle and the projection information on l.74 into a separate parameter input file.

ll.149-154: Instead of hard-coding the variogram, you might want to create a function for it. You can put the definition at the beginning of the code or even better, in a separate input file.

l.162: Same here. You should define a function that creates the model to be passed to krige, instead of hard-coding it.

Interactive comment on Atmos. Meas. Tech. Discuss., 8, 8191, 2015.