Journal of Hydrology 603 (2021) 126909

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Rainfall retrieval using commercial microwave links: Effect of sampling strategy on retrieval accuracy

Jayaram Pudashine [a,*], Adrien Guyot [a], Aart Overeem [b,c], Valentijn R.N. Pauwels [a], Alan Seed [d], Remko Uijlenhoet [e], Mahesh Prakash [f], Jeffrey P. Walker [a]

[a] Department of Civil Engineering, Monash University, Victoria, Australia
[b] Hydrology and Quantitative Water Management Group, Wageningen University and Research, the Netherlands
[c] R&D Observations and Data Technology, Royal Netherlands Meteorological Institute, the Netherlands
[d] Griffith University, School of Engineering and Built Environment, Brisbane, Australia
[e] Department of Water Management, Delft University of Technology, the Netherlands
[f] Data 61, CSIRO, Victoria, Australia

ARTICLE INFO

This manuscript was handled by Marco borga, Editor-in-Chief, with the assistance of Francesco Marra, Associate Editor

Keywords:
Cellular communication networks
Opportunistic sensing
Rainfall
Microwave links
Remote sensing

ABSTRACT

This study presents the first evaluation of using commercial microwave link (CML) data for rainfall measurements in Australia, with the test site being the greater Melbourne Metropolitan area. More than 100 CMLs with microwave frequency ranging between 10 and 40 GHz have been used for the rainfall retrieval. The 15-minute received signal levels (RSLs) for each CML based on two sampling strategies (average and minimum/maximum) collected for 2 years provided a unique dataset to compare performances of rainfall retrievals. The open source algorithm RAINLINK was used for deriving rainfall from the 15-minute RSL data. From two years of data, a subset of 30 rainy days distributed across this period were used for calibrating the RAINLINK parameters, with the remaining data used for validation. For this study, only path-averaged rainfall intensities were validated based on a gauge-adjusted radar product serving as the reference. The result of the wet-dry classification showed that the minimum and maximum RSL data performed better, with lower probability of false detection and higher Matthews correlation coefficient than average RSL data. For the rainfall retrieval, both datasets showed similar correlation with the gauge adjusted radar product. However, based on other statistics (RMSE, bias and CV) minimum and maximum RSL data outperformed average for the rainfall retrieval. Overall, this study highlights the robust accuracy of commercial microwave links for rainfall retrieval while using only minimum and maximum RSL data.

1. Introduction

Accurate and timely rainfall information is crucial for real-time flood forecasting and various agricultural management activities. Wherever available, dense rain gauge networks combined with operational weather radars are currently considered as the most reliable source of temporal and spatial rainfall estimates at ground level (Russell et al., 2010). However, deployment of such infrastructure for national coverage is costly, and most low-to middle-income countries are not able to afford such equipment. Moreover, there are a number of limitations with this rainfall measurement approach. For example, conventional rain gauges provide only discrete point observations which may be biased due to wind and splash effects(Muller and Kidd, 2006). While radar rainfall products provide integrated observations with a large spatial coverage, they do not directly measure rainfall close to the ground. Instead, their measurements are based on an indirect measurement of reflected microwave energy from hundreds of meters above the ground making them subject to large uncertainties in their retrieval. These include hardware calibration, beam broadening, attenuation due to rain, ground clutter, anomalous propagation, and wind effects (Doviak, 1984; Joss et al., 1990; Germann et al., 2006; Berne and Krajewski, 2013). Alternatively, satellites provide a nearly global spatial coverage of rainfall, but they do not provide a direct measurement of rainfall near the ground either. Rather, they provide an indirect estimate based on cloud top temperatures or emitted or reflected microwave energy. Thus these data still need to be calibrated using ground

* Corresponding author.
E-mail address: jayaram.pudashine@monash.edu (J. Pudashine).

https://doi.org/10.1016/j.jhydrol.2021.126909
Received 15 October 2020; Received in revised form 25 August 2021; Accepted 30 August 2021
Available online 14 September 2021

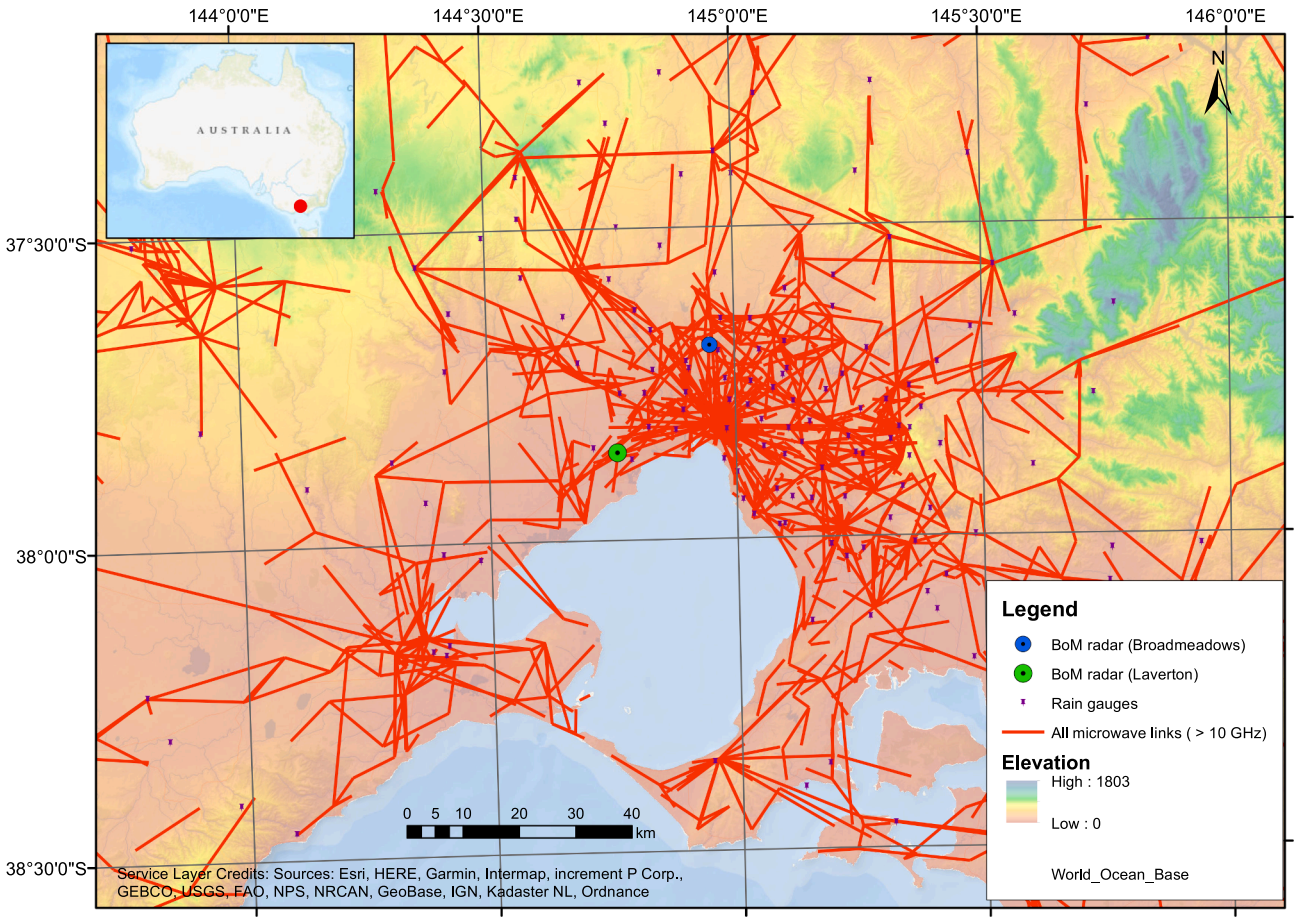

Fig. 1. Location of commercial microwave links for greater Melbourne with a frequency above 10 GHz. The density of such links decreases from the central business district towards the outer suburbs. There are a total of 2,290 unique microwave links in this region (Source: Australian Communication and Media Authority (ACMA) database, 2019).

measurements, which limits the stand-alone use of satellite products for operational applications (Kucera et al., 2013).

In contrast to the traditional methods outlined above, commercial microwave links (CMLs) operated by mobile network operators (MNOs) have proven to be a complementary source of rainfall information (Messer, 2018; Uijlenhoet et al., 2018; Chwala and Kunstmann, 2019). A CML includes a transmitter on one backhaul tower and a receiver on another backhaul tower. Depending on the number of transmitters and receivers on these towers, a CML could include a single link or multiple links at the same or different frequencies and/or polarisations, known as "sub-links". The commonly called "duplex links" are a CML that includes two sub-links permitting two-way data transmission. In addition, redundant links may be installed to back up the existing operational link during a failure. The primary objective of these CMLs is to provide telecommunication services, but the information they collect to maintain their network could be utilized for retrieving rainfall rates (Messer et al., 2006; Leijnse et al., 2007).This technique of rainfall measurement has been proven useful in areas where there are no other sources of rainfall information, but also in densely populated cities where the existing infrastructure does not provide reliable means of observation (Overeem et al., 2011; Pastorek et al., 2019). In such cities, it is usually problematic to find the appropriate location to install rain gauges among high-rise buildings according to the official World Meteorological Organisation requirements, with the high-rise buildings also creating ground clutter for weather radars. The CMLs which are widespread in urban areas typically transmit the microwave signals over distances of a couple of hundred meters to a few tens of kilometres, at tens of meters above the ground, thus providing path-integrated rainfall information close to the ground (Overeem et al., 2016b; Gazit and Messer, 2018).

CML-derived rainfall estimates are based on the fact that the transmitted signal is attenuated as it passes through the rain medium. This attenuation in the signal is more pronounced at higher frequency bands, *e.g.* it typically becomes significant above 10 GHz. This phenomenon was widely studied by the telecommunication engineering community in the early 1960 s to design an optimal spacing between microwave towers for efficient and reliable communication (Hogg, 1968). Later, various experimental studies using microwave links showed that inversion of the technique could be used for rainfall retrieval (Atlas and Ulbrich, 1977; Giuli et al., 1991; Christopher et al., 1996; Mello et al., 2002; Holt et al., 2003). However, application of this technique was limited until the early 2000 s, when Messer et al. (2006) and Leijnse et al. (2007) concomitantly demonstrated the use of CML signal attenuation for rainfall measurement. This was a major breakthrough toward demonstrating the potential to use the more than 4 million commercial microwave links in the world (Ericsson, 2017) for rainfall monitoring purposes. Subsequently, this technique gained popularity with feasibility and validation studies undertaken for a variety of locations around the world including: Brazil (Rios Gaona et al., 2015), Burkina Faso (Doumounia et al., 2014), Czech Republic (Fencl et al., 2013; Fencl et al., 2017), Germany (Chwala et al., 2012; Chwala et al., 2016; Smiatek et al., 2017; Graf et al., 2020), Israel (Messer et al., 2006; Goldshtein et al., 2009), Italy (Roversi et al., 2020) , The Netherlands (Leijnse et al., 2007; Overeem et al., 2011; 2013; 2016b; de Vos et al., 2019), Pakistan (Sohail Afzal et al., 2018) and Switzerland (Bianchi et al., 2013).

These validation studies have been conducted based on a few links to

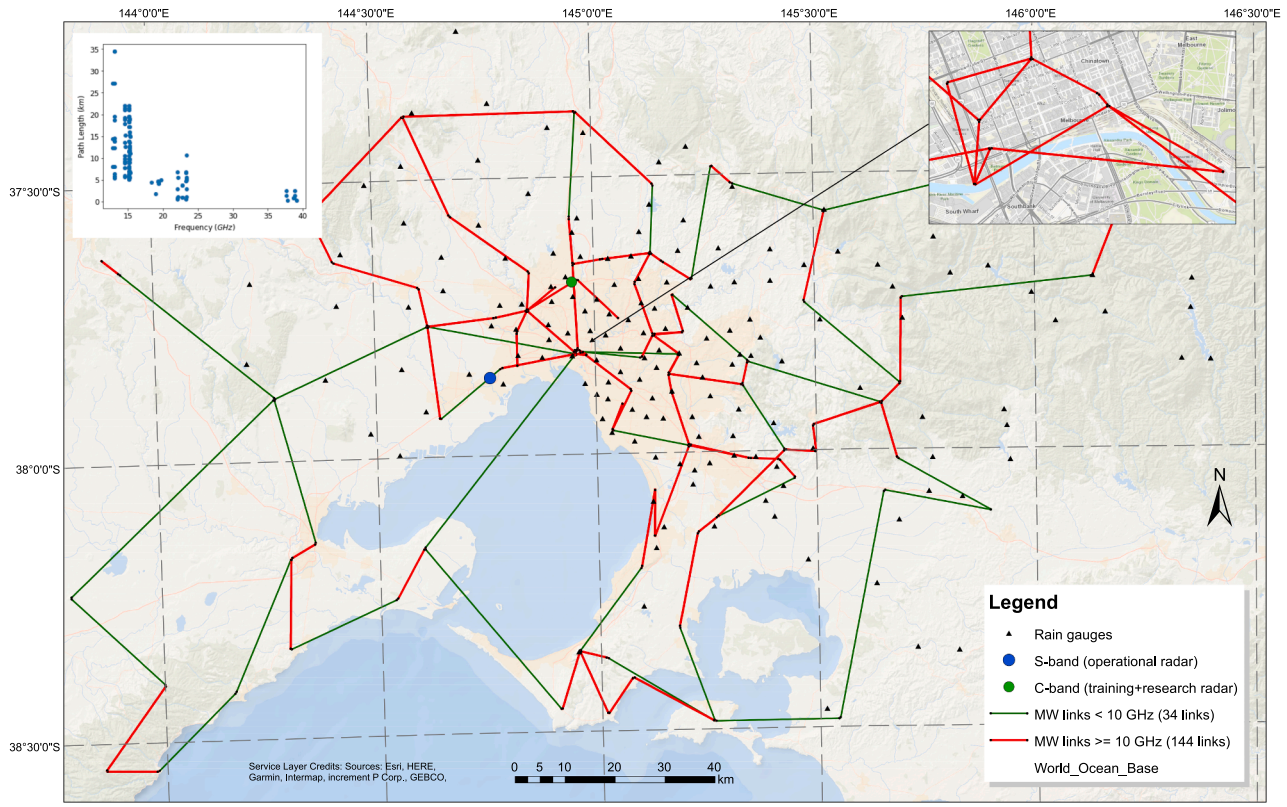

Fig. 2. Locations of commercial microwave links from one of the operators for greater Melbourne. Red lines indicate microwave links having a microwave frequency greater than 10 GHz (which were used for this study). Green lines indicate microwave links with frequencies lower than 10 GHz (which were not used for this study). On the top left is a plot of the microwave frequency of the microwave links f (GHz) against the path length L (km) for the 144 sub-links used for this study. A small subset on the top right shows the microwave links at the Melbourne Central Business District. There were 80 unique link paths including both 64 duplex and 16 single links. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a couple of thousand links covering an entire country such as The Netherlands (Overeem et al., 2013) and Germany (Graf et al., 2020). The temporal resolution of such CML rainfall estimates typically varies from a few seconds to 15 min, with most telecommunication operators sampling the received signal level (RSL) at 10 Hz but storing it at a much coarser temporal resolution. In most studies, 15-minute minimum and maximum RSL data, as stored operationally by the MNO's network management systems, were used for rainfall retrieval (Leijnse et al., 2007; Goldshtein et al., 2009; Overeem et al., 2011; 2016b; Rios et al., 2017). There have been a few studies using 1-min and even higher temporal resolution, up to a second, instantaneous RSL data for rainfall estimation (Doumounia et al., 2014; Chwala et al., 2016). Overeem et al. (2016b) evaluated 2.5 years of microwave link rainfall estimates for the Netherlands with more than 3000 microwave links (using 15-minute minimum–maximum sampling) against gauge-adjusted radar rainfall data, showing a relative underestimation of 9% for 15-min interpolated rainfall maps with a 74 km^2 resolution. However, the interpolated hourly rainfall map using CMLs outperformed automatic rain gauges compared with gauge adjusted radar data. Similarly, Chwala et al. (2012) used 1-minute averaged RSL data recorded with data loggers for five microwave links, showing a good correlation between link and radar-derived rainfall.

Some of the MNOs also provide instantaneous RSL (periodic snapshots) data over the 15-minutes: de Vos et al. (2019) compared the performance of instantaneous versus minimum and maximum RSL data for The Netherlands. Even though this comparison was based on data from two different periods, each having a different network, the use of minimum and maximum sampled data outperformed the instantaneous 15-minute data. Similarly, average sampling of the received signal level over the 15-minute interval is also common for telecommunication operators in some parts of the world, but this has not been evaluated against the widely used minimum and maximum RSL strategy. Accordingly, this study tests this alternative strategy, while demonstrating for the first time the capability of rainfall retrieval using CML signal attenuation data in the Australian continent

To date, there has not been a study evaluating the errors introduced by the minimum–maximum sampling as opposed to average sampling. This study explores the capability of rainfall retrieval using CML signal attenuation data for the Greater Metropolitan area of Melbourne, the second largest city in Australia, with a population of 4.48 million (Australian Bureau of Statistics, 2016). This study compares the performance of rainfall retrieval using two commonly sampled datasets for the same period, where data based on minimum–maximum and average sampling from the same link paths are compared. A total of 135 microwave links are used, covering approximately 2 years of data. These CML data were stored by the network monitoring system (NMS) every 15-minute based on10 Hz sampling data. This 15-minute data includes the minimum, maximum and average over 15-min intervals with the constant transmitted power.

2. Study area, data and methods

2.1. Description of the study area

The study area covers the greater Melbourne region in the Australian state of Victoria. This region has a temperate oceanic climate (Cfb, Köppen-Geiger classification), with an annual average rainfall (based on 29 years of rainfall data from 1990 until 2018 for 73 stations) varying from 500 mm in the west of Melbourne to 1400 mm in the Dandenong ranges towards the eastern part of the city, with a standard deviation of

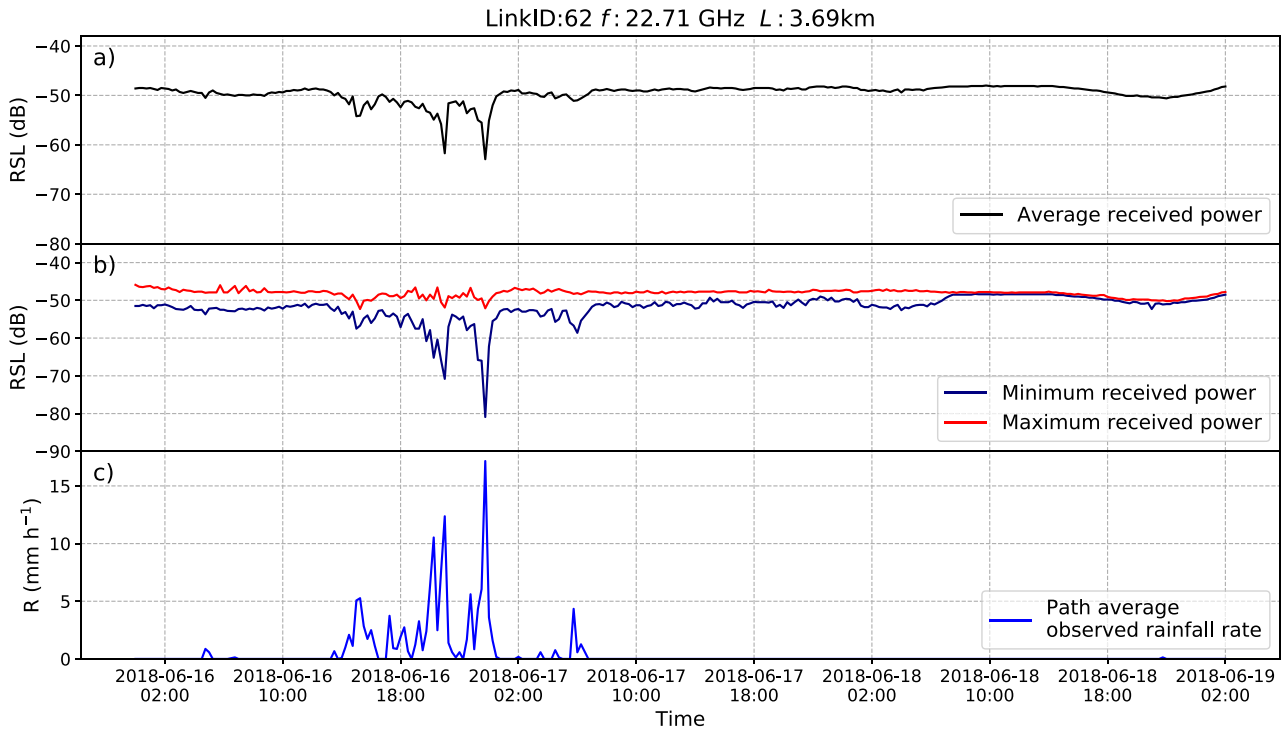

Fig. 3. Time series of a rainfall event on the 16th to 19th June 2018 showing: (a) 15-min average received signal level; (b) 15-min minimum and maximum received signal levels; and, (c) Path-averaged observed rainfall rate along the microwave link.

175 mm. Most of the rainfall occurs during southern hemisphere winter (June, July and August) and spring (September, October and November). On average, there are 110 days each year with at least 1 mm of rainfall. The average temperature (based on the same period with 21 stations) of the study area varies between 18 °C and 24 °C for summer and varies between 6 °C and 12 °C for winter season. The elevation of the study area ranges from sea level to 1803 m. Fig. 1 shows the study area with location of radar, rain gauge, and the microwave links above 10 GHz from all telecommunication operators. The density of these microwave links ranges from 0.27 to 2.15 km per km^2.

2.2. Data

2.2.1. Commercial microwave link data

Received signal level (RSL) data from one of the telecommunication providers were collected for this study within a radius of approximately 200 km around the Melbourne central business district. Data from a total of 178 sub-links (64 duplex links and 50 single links) for the period ranging from 15 July 2017 to 31 July 2019 were collected as shown in Fig. 2. This dataset contains minimum, maximum and average received signal level (RSLs) over 15-min intervals with a resolution of 0.1 dB, based on a 10 Hz sampling rate. Hereafter, the sampling strategy providing minimum and maximum RSLs over 15-minutes is named "MinMax" and the strategy providing average RSLs over the 15-minute interval is named "Average". These 178 links represent the data collected by only one of the MNOs in the area. The total density of links within this area (all operators) is quite high as shown in Fig. 1. These 178 links have frequencies ranging from 6 to 39 GHz with path lengths ranging from 0.2 km to 57 km. Among the 178-microwave links, only links with frequencies above 10 GHz (144 links) were retained for the present study. Among these selected 144 links, 9 links were horizontally polarized while all other links were vertically polarized. The majority (128 links) were duplex (transmitter and receiver at both ends) with only 16 links being single channel, thus forming 80 unique link paths. The transmitting powers of all retained links were constant through time.

Fig. 3 shows timeseries of 15-minute Average and MinMax RSL for a selected event. In both cases, the received power levels decreased proportionally with the observed rainfall rate, however the power level still fluctuated during the dry periods. This fluctuation is seen more in the minimum and maximum compared with the average power level. The values of the minimum and maximum power levels over the 15-minute intervals obviously reached more extreme values (larger maximum and lower minimum). Since the RSL was sampled at 10 Hz, the minimum and maximum are the extremes of a distribution of values or the outlier of that distribution (Pudashine et al., 2020).

Among the 144 retained links, 138 contained more than 90% of the maximum available data, equivalent to a total of 15,722 h of data. The other 6 links had more than 70% available data, equivalent to a total of 12,748 h. Among this dataset, a subset of 30 rainy days spread over the 2-year period was used for calibration of the rainfall retrieval algorithm parameters. This subset was chosen in such a way that it includes a maximum number of available microwave links and total wet time intervals and maximum rainfall intensity on a daily basis. The details of this subset are presented in Appendix 1; the remaining data were used for validation which includes 128 rainy days.

2.2.2. Weather radar

An S-band weather radar operated by the Bureau of Meteorology collected data over the study area during the 2 years of this study. This radar was located at Laverton (37°51′36″ S, 144°45′36″ E), 44 m above sea level. A gauge-adjusted radar data product named Rainfields (Seed et al., 2008) was obtained from the Bureau of Meteorology with a spatial resolution of 0.5 km × 0.5 km and a temporal resolution of 5 min. *Rainfields* is a comprehensive framework, which provides real-time quality-controlled quantitative precipitation for the operational Australian weather radar network. This framework follows a series of quality control measures including removal of ground and sea clutter, interferences, bright band correction and partial beam blocking. This filtered observation is then converted to surface rainfall maps estimating the reflectivity at the earth surface using a three-dimensional kriging interpolation technique (Wesson and Pegram, 2006), which was then

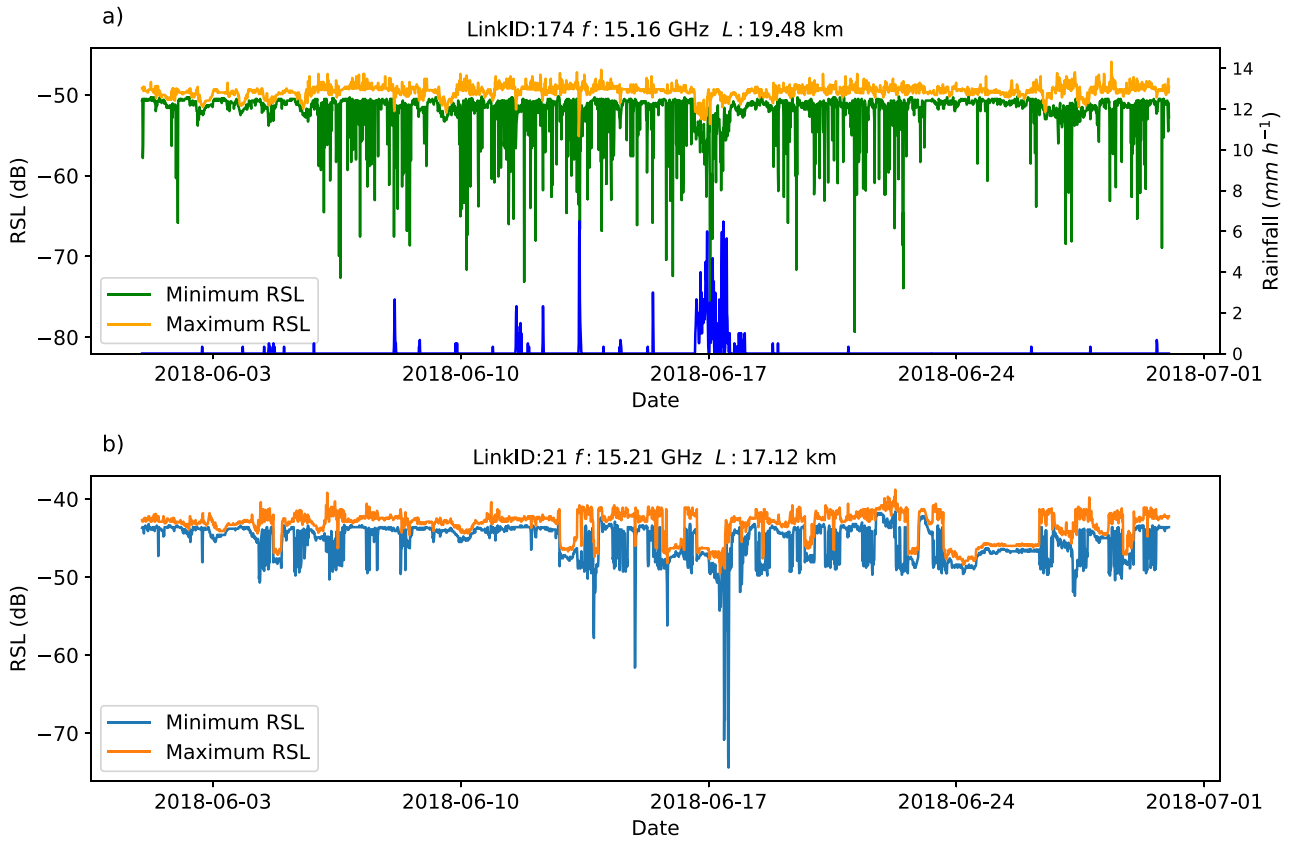

Fig. 4. Example of some suspicious microwave link data: (a) There is noise observed in the minimum (green) and maximum (yellow) RSL; (b) There is a sudden drop in the received power level even if the transmitted power remains constant. Blue lines indicate observed rainfall rates from *Rainfields*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

converted to rainfall rates at ground-based on static Z-R relationship and finally correcting bias the bias by using near-real-time rain gauge information (Seed et al., 2008). For the real-time rain gauge observation, a total of 261 rain gauges were used for this radar. For this analysis, *Rainfields* (gauge-adjusted radar) data were used as a reference to compare with the rainfall retrievals from CMLs. The path-average rainfall rates were calculated based on the weights of the intersecting CML paths for each radar pixel.

3. Methodology

3.1. Preliminary data processing and quality check

The CML dataset was delivered by the operator in two separate files: one with 15-min RSL data for all the links stored daily, and the other with the corresponding metadata. These files were received on a monthly basis at the end of each month. Metadata included the location of transmitter and receiver nodes, the elevation of the antennas, the assigned microwave frequencies (including frequency bandwidth), the polarization of the signal, path lengths and the IP addresses of each transmitter and receiver. Using the IP address as a unique identifier in the CML data and metadata, merged RSL data with necessary fields (frequency, latitude, longitude, and polarization) were prepared for further processing.

Among all 144 links, there were nine links that showed some suspicious behaviour in the data. Three distinct behaviours were identified in these links: (a) presence of noise in the dataset; (b) a sudden drop in the signal level during dry periods; and (c) a gradual increase/decrease in the signal level. Examples are shown in Fig. 4. These suspicious links were excluded from further analysis manually. As in this case, there were only a limited number of links; thus, it was possible to verify this manually. However, for large scale studies with thousands of links, an automated statistical analysis as proposed by Graf et al. (2020) could be used.

3.2. Use of RAINLINK

After pre-processing the data, the freely available RAINLINK package developed by Overeem et al. (2016a) was used for retrieving rainfall rates. Originally, RAINLINK was designed for handling minimum and maximum RSL data with a constant transmitting power. de Vos et al. (2019) describe the pre-processing necessary to handle instantaneous received signal level data, where the transmitted power is allowed to vary. Studies employing average RSL data have not been published to date. This study is therefore the first to use average RSL data in RAINLINK for retrieving rainfall. Here average and minimum/maximum RSL data were processed separately. Further details of the RAINLINK package can be found in Overeem et al. (2016a).

The RAINLINK package includes the following processing steps:

1) **Pre-processing of the data**. Duplicated link identifiers, identifiers with inconsistent metadata, and links with frequencies outside the range 10–40 GHz are excluded from the analysis.
2) **Dry/wet classification.** Rainy periods are identified based on spatial correlation. When at least half the nearby links (default radius of 15 km) experienced a drop in the minimum or average signal level, the time interval was considered as “wet”. This radius was increased to 20 km (in contrast to the default value of 15 km) based on the spatial distribution of the microwave links in the Melbourne metropolitan area. This drop in the signal level was calculated based on the difference between the RSL data as compared with the maximum value of the link over the previous 24-hour period, both as

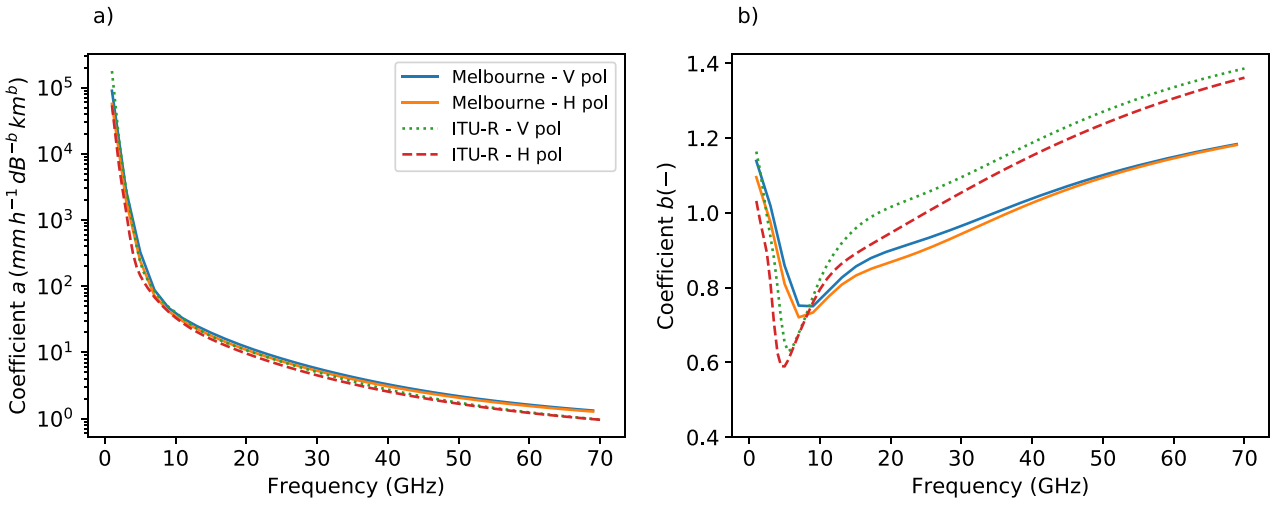

Fig. 5. (a) Coefficients a and (b) exponent b of the power-law relation between R and k for both horizontally and vertically polarized signals for frequencies ranging from 1 to 70 GHz. The values recommended by the International Telecommunication Union, Radio communication (ITU-R, 2005) for computing specific attenuation for given rain rates for world wide application are shown in dotted and dashed lines.

Table 1
List of variables used for calibration of RAINLINK. Here, the alpha coefficient provides the weightage between the minimum and maximum rainfall. The parameters a and b used in this relationship are based on a local disdrometer data.

Variable description	Symbol	Unit
Threshold median	Q_{mP}	dB
Threshold median per unit length	Q_{mPL}	dB km $^{-1}$
Wet antenna attenuation	A_a	dB
Alpha coefficient	α	–
Prefactor of rainfall-attenuation relationship	a	mm h^{-1} dB^{-b} km^{b}
Exponent of rainfall-attenuation relationship	b	–

a difference and difference divided by the path length. If the median of all nearby links was less than the predefined threshold median $Q_{mP}(dB)$ for the difference and $Q_{mPL}(dBkm^{-1})$ for the difference per kilometre link path, the link was labelled as "wet" for that interval.

3) **Reference signal level.** Based on the moving median of the signal level during the previous 24-hour dry period the reference signal level (P_{ref}) was determined. The difference between this reference level signal and actual signal level provided the attenuation due to rainfall.
4) **Outlier removal.** Based on a filter that relies on the principle of the rainfall distribution over space (Overeem et al., 2016a) outliers were removed. This filter discards the time interval of a link for which the cumulative difference between its specific attenuation and that of the surrounding links (default radius of 15 km, although for this study increased to 20 km) over the 24 h becomes lower than the outlier filter threshold value.
5) **Attenuation level.** The corrected minimum ($P_{Cor/min}$), maximum ($P_{Cor/max}$) and average ($P_{Cor/avg}$) attenuation level was calculated for each time interval as:

$$P_{Cor/min} = \begin{cases} P_{min} \ if \ \text{wet AND} \ P_{min} < P_{ref} \\ P_{ref} \ \text{otherwise} \end{cases} \quad (1)$$

$$P_{Cor/max} = \begin{cases} P_{max} \ if \ P_{Cor/min} < P_{ref} \ \text{AND} \ P_{max} < P_{ref} \\ P_{ref} \ \text{otherwise} \end{cases} \quad (2)$$

$$P_{Cor/avg} = \begin{cases} P_{avg} \ if \ \text{wet AND} \ P_{avg} < P_{ref} \\ P_{ref} \ \text{otherwise} \end{cases}, \quad (3)$$

where P_{min}, P_{max} and P_{avg} are the raw minimum, maximum and average attenuation, respectively.

6) **Wet antenna attenuation.** A constant wet antenna attenuation (A_a) was deducted from the corrected total attenuation. This attenuation was later divided by the path length to obtain the specific attenuation (k).
7) **Rainfall estimation.** The rainfall rate R was calculated from k using the power law equation proposed by Olsen et al. (1978):

$$R = ak^{b}, \quad (4)$$

where the values of the parameters a and b in Eq. (4) were derived for Melbourne using data obtained from an OTT PARSIVEL[1] optical disdrometer, as described in Guyot et al. (2019), and shown in Fig. 5. Extinction cross sections were estimated based on the T-matrix method developed by Mishchenko and Travis (1994) using a python interface developed by Leinonen (2014) being the most comprehensive and computationally efficient method for the calculation of electromagnetic scattering of particles of arbitrary shape. This was later used to derive the specific attenuation, which were related to rainfall rates using the power law model based on the least squares method. For the case of using MinMax, the weighing factor α was used to calculate the average rainfall from R_{min} and R_{max}.

$$R = \alpha.R_{max} + (1-\alpha).R_{min} \quad (5)$$

However, for the case of using average sampling, this α parameter was obviously not required.

3.3. Calibration of the RAINLINK parameters

RAINLINK has 14 parameters for rainfall retrieval, including those related to wet-dry classification, reference signal determination, outlier filter, wet antenna attenuation and rainfall retrieval using the power law. The optimal values of these parameters are likely to differ between different climatic conditions and microwave link networks (*e.g.* regarding sampling strategy, spatial link density, and resolution of the RSL data), thus it is recommended to calibrate the most important parameters on a subset of the data (de Vos et al., 2019). For this study, based on a sensitivity analysis using one month of dataset, the three parameters Q_{mP}, Q_{mPL} and A_a for both the Average and MinMax dataset were identified as most important for the overall rainfall retrieval. Additionally, α parameter was also identified as the most important for MinMax RSL data. For the sensitivity analysis, the cost function proposed by de Vos et al. (2019) was used, which includes the Probability of Detection (POD), Probability of False Alarm (POFA) or False Alarm Ratio, Coefficient of variation(CV), percentage bias and correlation.

Table 2
Confusion matrix based on gauge adjusted radar and CML.

		Gauge adjusted radar (Rainfields). Wet (0)	Dry (1)
CML	Wet (0)	True Positive (TP)	False Positive (FP)
	Dry (1)	False Negative (FN)	True Negative (TN)

TP: True positive (both R_{link} and R_{radar} detect rainfall),
TN: True negative (both R_{link} and R_{radar} show no rainfall),
FP: False positive (R_{link} detects rainfall but R_{radar} shows no rainfall),
FN: False negative (R_{link} shows no rainfall but R_{radar} detects rainfall).

Besides these, two additional parameters a and b are required; based here on local drop size distribution data from Guyot et al. (2020). More information on these parameters is provided in Table1. A subset of 30 rainy days spread over the 2-year period was selected for the calibration of the RAINLINK parameters. This subset was selected such that there is high availability of RSL data from microwave links during the selected periods, also ensuring that data from different seasons were included. This subset represents a total rainfall amount of 390 mm, ranging from 3 mm to 33 mm per day. The details of the calibration dataset are shown in Table 6 in Appendix 1; the remaining data were used for validation which includes 128 rainy days.

3.3.1. Parameters for the average RSL

The optimised values of the three most important parameters Q_{mP}, Q_{mPL} and A_a for the average RSL data have been identified using the optimization procedure in de Vos et al. (2019). The detail of the calibration procedure is described later in this section. The radius for nearby links was increased from 15 km to 20 km, as the density of the links was lower for the present case, as compared to The Netherlands. Other parameters besides these were kept at their default values in the RAINLINK package. Based on the calibration dataset, hourly rainfall estimates were calculated for the various combinations of values of Q_{mP}, Q_{mPL} and A_a. Here, hourly rainfall estimates were considered to minimize the sampling error caused by the measurement lag in the radar, due to its providing measurements aloft that usually take a couple of minutes for the rainfall to reach the earth's surface. Accordingly, Q_{mP} was varied from −2.5 to −0.1 dB, Q_{mPL} was varied from −2.0 to −0.1 dB km^{-1} and A_a was varied from 0.5 to 3 dB (steps of 0.1 for all parameters). The sensitivity of each of these parameters are shown in Fig. A1. The results obtained for each of the combinations were evaluated against the gauge-adjusted radar product, also accumulated to path-averaged hourly values. The path-averages were calculated based on weights of the intersecting CML paths for each radar pixel.

3.3.2. Parameters for minimum/ maximum RSL

For the minimum and maximum RSL data, an additional parameter α is required in the optimization process of the rainfall retrieval (Overeem et al., 2016a). The value of α was varied between 0.10 and 0.50 (with steps of 0.01). The other three parameters, including wet antenna attenuation A_a, were obtained in a similar manner as for the average RSL. Even though the A_a is physically related to the types and materials of the antenna cover it is expected to behave non-linearly because of dependence on rain rate and temporal variations within a time interval. This implies that the optimal value will vary between sampling strategies and that A_a also has a negative correlation with the value of α. Thus, the optimized value of A_a from Average may not be entirely suitable for application in MinMax."

3.4. Performance metrics

The overall performance of the path-average rainfall retrieval was assessed based on a series of evaluation criteria, covering the two main steps in the rainfall retrieval: (a) Wet-dry classification and; (b) Rainfall retrieval.

(a) Wet-dry classification

This provides a measure of how well the link observations correctly estimate the occurrence of rainfall. The following criteria (de Vos et al., 2019; Graf et al., 2020) were used to assess the performance of the classification based on the confusion matrix as shown in Table 2.

1) **The probability of detection (POD)** provides a measure of proportion of actual wet periods that are identified by both the CML and the radar. In this case, POD is defined as the percentage of wet periods identified using the nearby link approach when both R_{link} and R_{radar} detect rainfall. The POD is given as:

$$\mathrm{POD} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\%. \tag{6}$$

The POD value ranges from 0 to 100%, with 100% being a perfect score and 0% being the worst.

2) **The Probability of False Alarm (POFA)** provides a measure of the proportion of the identified wet periods that are incorrectly identified (Barnes et al., 2009). This is also knownas the False Alarm Ratio (FAR) and is used in reporting the performance of dry-wet classification (de Vos et al., 2019). The POFA is given as:

$$\mathrm{POFA} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}} \times 100\%. \tag{7}$$

Similarly, the POFA value also ranges from 0 to 100 %, but with 100% being the worst score and 0% being the best score.

3) **The Probability of False Detection (POFD)** provides a measure of the proportion of the actual dry periods that are incorrectly identified as being wet. This is also known as the False Alarm Rate. The POFD is given as:

$$\mathrm{POFD} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TN}} \times 100\%. \tag{8}$$

POFD value also ranges from 0 to 100%, but with 100% being the worst score and 0% being the best score.

4) **The Matthews correlation coefficient (MCC)** provides a measure of quality of the binary classification (wet-dry in the case of classification for CMLs) (Matthews (1975). This is considered as one of the best ways to report the result of the confusion matrix:

$$\mathrm{MCC} = \frac{\mathrm{TP}\times\mathrm{TN} - \mathrm{FP}\times\mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}. \tag{9}$$

The MCC value ranges from 0 to 1, with 1 being the best score and 0 being the worst.

5) **The Error rate (ERR) or misclassification rate** provides the performance measure of binary classification based on the misclassification from both positive and negative classes and is calculated as:

$$\mathrm{ERR} = \frac{(\mathrm{FP}+\mathrm{FN})}{(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN})}. \tag{10}$$

Similarly, ERR ranges from 0 to 1 with 0 being the best score and 1 being the worst score.

(b) Rainfall retrieval

This set of evaluation parameters provides a measure of how well the CML-derived rainfall relates to the reference rainfall depths (in this case the gauge-adjusted radar).

Table 3
Calibration results for a selection of four of the RAINLINK parameters and comparison with values obtained for The Netherlands. The default parameters for MinMax data were −1.4 dB, −0.7 dB km^{-1}, 2.30 dB and 0.33 for Q_{mP}, Q_{mPL}, A_a, and α respectively.

Dry/ wet classification parameters		Rainfall retrieval parameter	
Threshold Median, Q_{mP}(dB)	Threshold Median L,Q_{mPL} (dB km^{-1})	Wet antenna attenuation, A_a (dB)	Alpha,α
Average data			
−0.7	−0.2	1.6	–
MinMax data			
−1.50	−0.40	1.4	0.29
Instantaneous RSL data (de Vos et al., 2019)			
−0.6	−0.4	1.4	–

Table 4
Performance criteria for the calibration period for Average and MinMax.

Dataset	Dry/ Wet classification				Rainfall retrieval		
	POD	POFD	POFA	MCC	Relative Bias (%)	CV	ρ
Average	68.25	10.25	35.45	0.47	0.20	1.41	0.67
MinMax	64.62	7.38	30.39	0.55	−0.68	1.12	0.68

1) The **Pearson correlation coefficient (ρ)** provides the correlation between rainfall depths measured by the link R_{link} and the gauge-adjusted radar (R_{radar}). It is given as:

$$\rho = \frac{cov(R_{link}, R_{radar})}{std(R_{link})std(R_{radar})}, \quad (11)$$

where *cov* (x, y) is the covariance between × and y and std(x) is the standard deviation of x. ρ values range from 0 to 1, with 1 being the best and 0 the worst performance.

2) The coefficient of variation (***CV***) provides a measure of the dispersion of data points between the rainfall intensity derived by the link (R_{link}) and the gauge-adjusted radar (R_{radar}). It is given as:

$$CV = \frac{std(R_{res})}{\overline{R}_{radar}}, \quad (12)$$

where $R_{res} = R_{link} - R_{radar}$ and $\overline{R}_{radar}$ is the mean of the gauge adjusted radar data. The smaller the CV the better the performance.

3) The relative bias provides the average error between the rainfall intensity measured by the link R_{link} and the gauge-adjusted radar (R_{radar}). It is given as:

$$\frac{\overline{R}_{res}}{\overline{R}_{radar}} \times 100\%, \quad (13)$$

where, $\overline{R}_{res}$ is the mean of the residual. Similarly, values closer to 0 are better, however positive values indicate overestimation and negative values indicate underestimation compared with the reference.

4. Results

4.1. Calibration

Table 3 shows the calibration results for the four most important RAINLINK parameters. Two parameters, threshold median (Q_{mP}) and threshold median per unit length (Q_{mPL}), are related to the wet-dry classification while the remaining two, wet antenna attenuation (A_a) and Alpha (α), are related to rainfall retrievals. For Average, both Q_{mP} and Q_{mPL} are less negative compared to the default values of RAINLINK while the A_a value is also lower compared to MinMax, but higher compared to instantaneously sampled data. For MinMax, the threshold median is slightly higher, but the threshold median per unit length is less negative than the default value. This indicates that the time interval is more likely to be classified as wet, with the threshold median per unit length being closer to 0 than the default value, which is also the case for

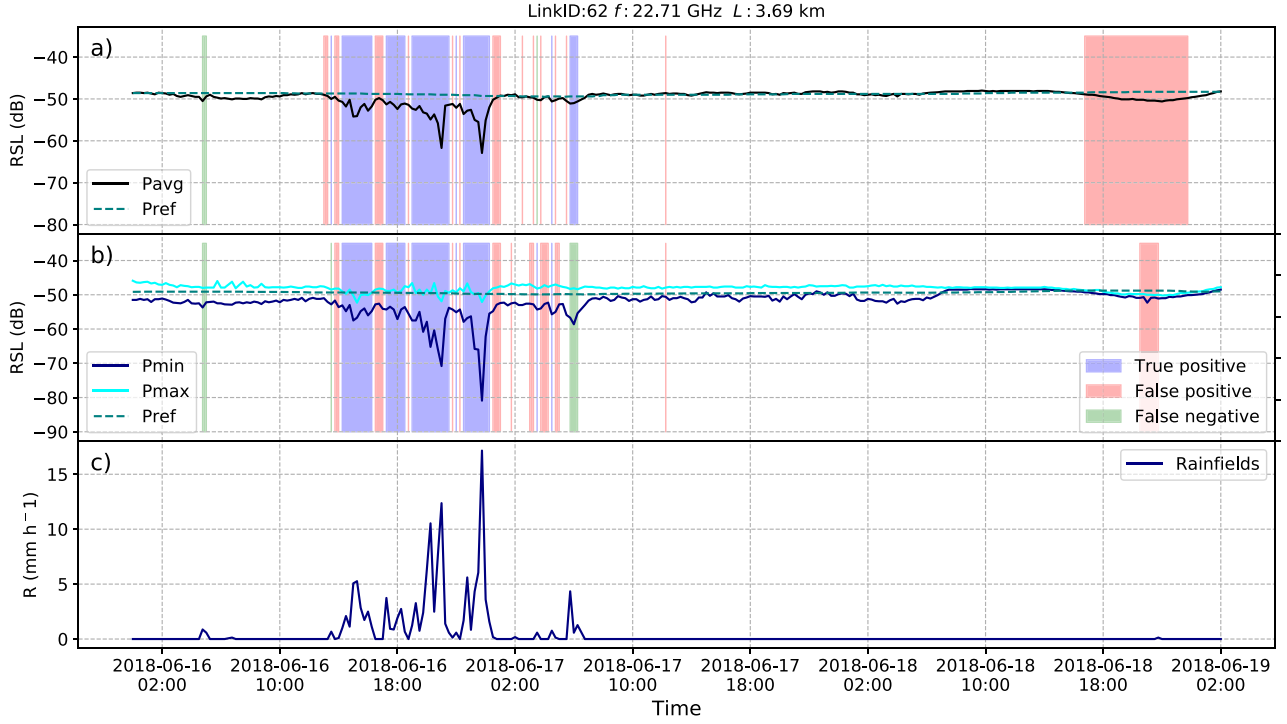

Fig. 6. Time-series of (a) average RSL; (b) minimum and maximum RSL; and (c) gauge adjusted radar rainfall intensities for a selected rainfall event for LinkID 62. The wet-dry classification using the calibrated parameters is shown as shaded colours.

Table 5
Performance of wet-dry classification for 15-min average and 15-min minimum/maximum RSL data.

Data set	Threshold for time interval to be wet											
	0 mm h^{-1}				0.1 mm h^{-1}				0.5 mm h^{-1}			
	POD	POFD	POFA	MCC	POD	POFD	POFA	MCC	POD	POFD	POFA	MCC
Average	64.34	8.76	42.63	0.38	68.05	8.94	45.25	0.38	77.91	9.57	50.57	0.35
MinMax	53.58	3.70	36.23	0.42	57.34	3.84	39.01	0.46	68.41	4.35	45.78	0.45

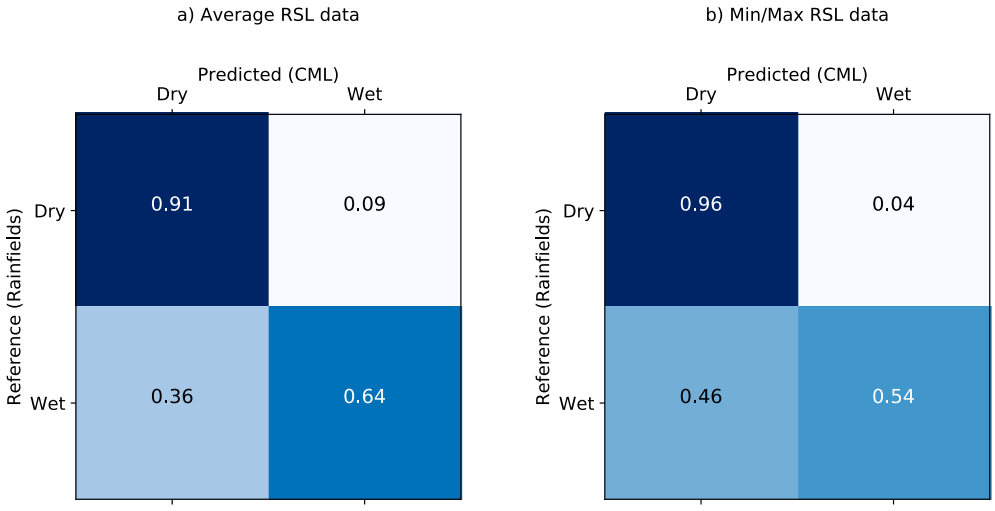

Fig. 7. Normalized confusion matrix for wet-dry classification for (a) Average data; and (b) MinMax data.

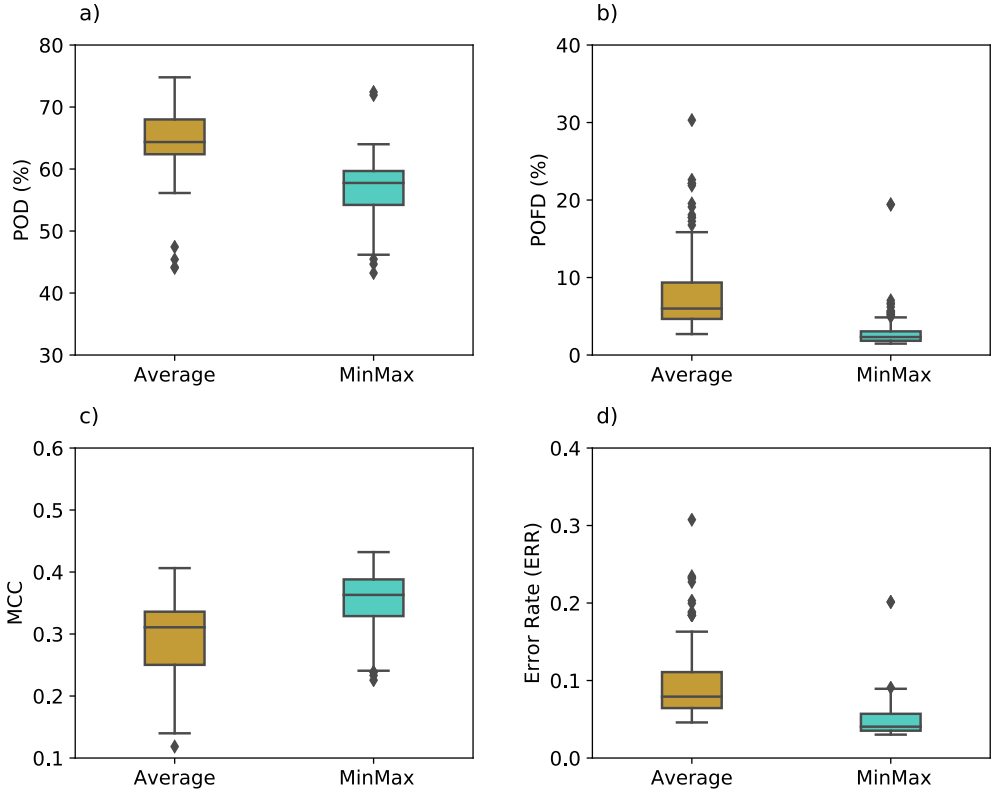

Fig. 8. Box plots showing the performance criteria for the wet-dry classification using the nearby links methods with: (a) Probability of detection (POD); (b) Probability of False Detection (POFD); (c) Matthew correlation coefficient (MCC); and (d) Error rate (ERR) for Average and MinMax data.

the threshold median for Average.

Table 4 shows the performance for the calibration dataset for both the Average and MinMax data presented at an hourly timescale. For dry/wet classification, MinMax had a better performance when compared to Average data, although its POD value was lower. For the rainfall retrieval, both datasets showed similar performance for ρ. MinMax had a

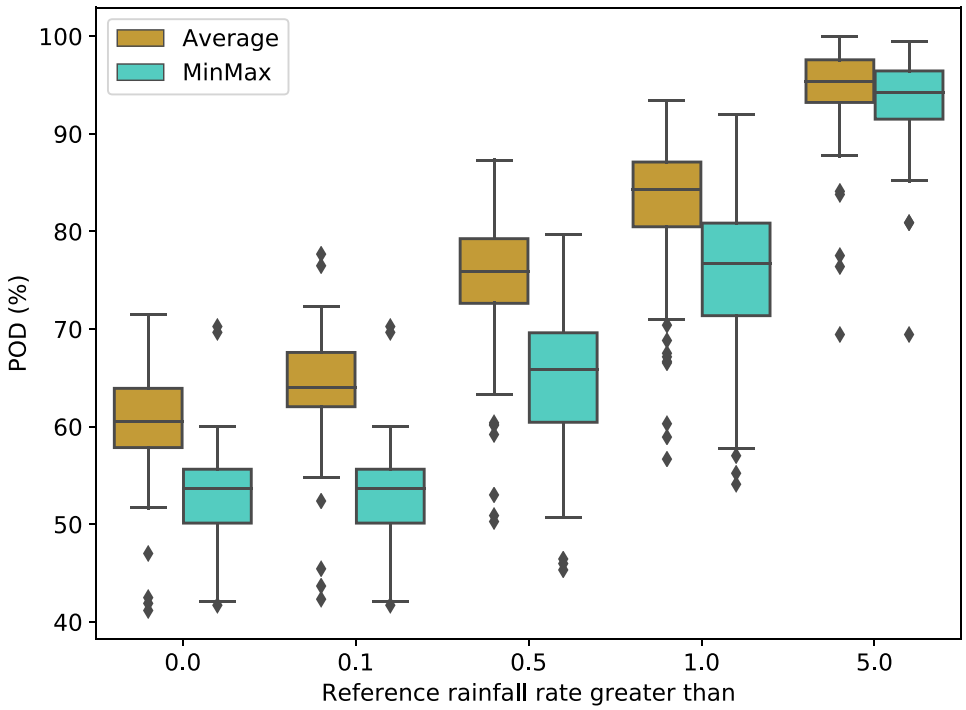

Fig. 9. Probability of detection (POD) based on a reference rainfall rate. Here five thresholds (0, 0.1, 0.5, 1.0 and 5.0-mm h^{-1}) for the gauge adjusted radar data are used to filter the data, thus this result provides POD result for only wet intervals.

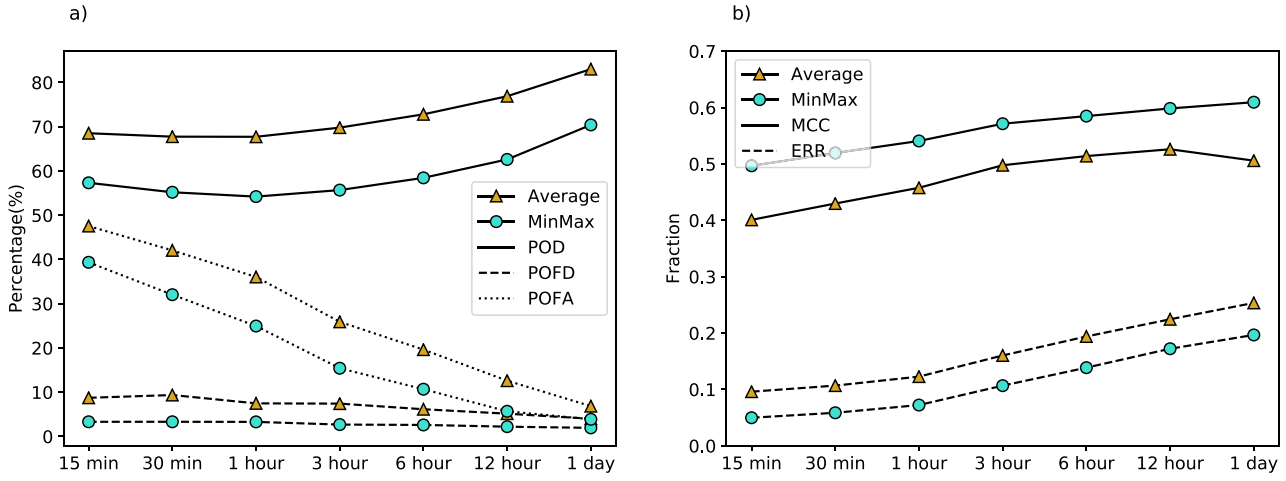

Fig. 10. Validation criteria showing the ability of CMLs to detect rainy periods against gauge-adjusted radar data; (a) POD, POFD and POFA; and. (b) Matthew correlation coefficient (MCC) and Error rate (ERR). All four parameters were calculated for Average and MinMax for various accumulation intervals, using a threshold of 0.1 mm h^{-1} to detect rain occurrence.

small negative bias compared to the almost unbiased Average data, and had a much lower value for CV. Hence, MinMax resulted in the best overall performance.

5. Validation result

After processing the dataset (excluding the data used for calibration), the number of data points containing results for the two sampling strategies (MinMax and Average) were of different sizes, because the outlier filter used in processing the raw data removed different time intervals for specific links from MinMax and Average data. Thus, in order to make a fair comparison, the time intervals with available data for both strategies were retained. Also, to note that all the performance evaluation was based on the path-average rainfall depths against the reference.

5.1. Performance of wet-dry classification

Fig. 6 shows the time series of RSL with corresponding rainfall intensities for a selected event, together with the wet-dry classification for one of the microwave links. For both the Average and MinMax data, most of the wet periods with higher rainfall intensities were identified correctly as shown by the true positives. However, some time intervals with low rainfall intensities during a wet period were classified as dry, *i.e.* false negatives. There were instances where the time intervals are incorrectly classified as wet even though there was no rain observed on the ground, indicated by the false positives. These false positives were observed during time-intervals where the RSL dropped below the

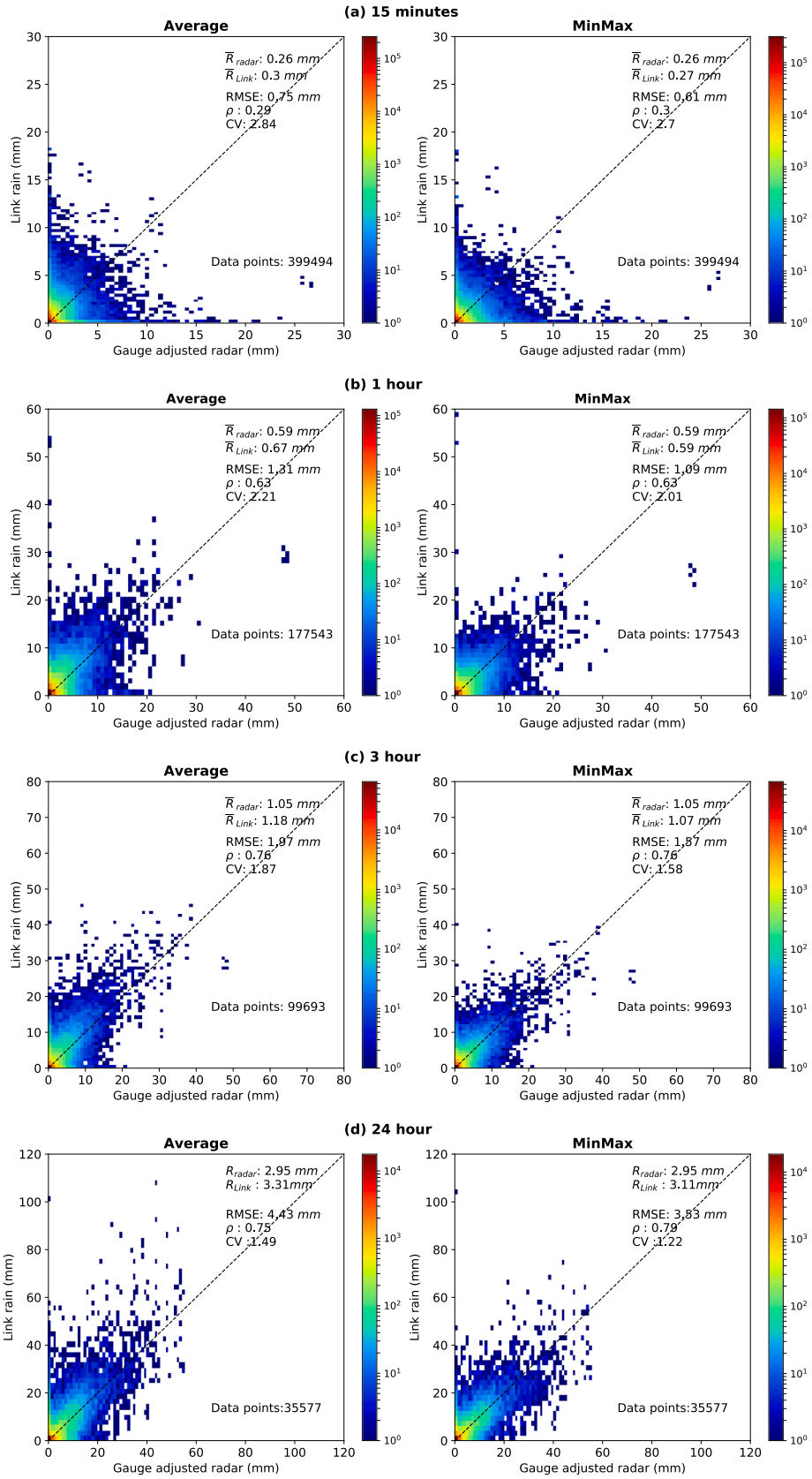

Fig. 11. Validation of path-average CML-rainfall against gauge-adjusted radar rainfall. Scatter density plot of link-derived rainfall with radar over intervals of: (a) 15 min; (b) 1 h; (c) 3 h; and (d) 24 h. This plot includes both false positives and false negatives only leaving out the dry-dry cases.

Table 6
Validation of 15-min and 1-hour accumulation link-derived rainfall against gauge-adjusted radar rainfall (reference) on a seasonal basis.

Dataset	15-minute			1-hour		
	Relative bias (%)	CV	ρ	Relative bias (%)	CV	ρ
Summer (Dec, Jan, Feb)						
Average	15.75	2.87	0.29	15.80	1.92	0.69
MinMax	9.05	2.85	0.32	9.05	1.90	0.69
Autumn (Mar, Apr, May)						
Average	12.73	2.81	0.30	12.71	2.60	0.52
MinMax	13.68	2.65	0.29	13.70	2.32	0.51
Winter (Jun, Jul, Aug)						
Average	9.56	2.52	0.22	8.78	2.21	0.53
MinMax	3.60	1.38	0.23	3.62	1.90	0.56
Spring (Sep, Oct, Nov)						
Average	7.92	2.71	0.32	8.12	1.90	0.71
MinMax	7.26	2.58	0.33	7.51	1.72	0.72

baseline signal level due to reasons other than rain. For this event, Average data correctly identified wet intervals with a POD of 90% and slightly lower POD of 84% for the MinMax data. Also, for the Average data 18% and for the MinMax only 9% of the dry periods were incorrectly classified as wet. Similarly, 10% and 16% of the wet intervals were missed by the Average and MinMax data respectively.

Table 5 shows the performance summary of the wet-dry classification for three threshold values used to distinguish each of the time intervals between wet and dry periods based on the gauge-adjusted radar data for all the dataset. Average data showed a higher probability of detection (64%) as compared to MinMax (54%). Also, 9% and 4% of the dry periods were incorrectly classified as wet based on the Avg and MinMax data, respectively (see Fig. 7). For all thresholds, Average data showed a higher probability of detection (POD) when compared with the MinMax data. However, at the same time it also showed a much POFD and POFA, meaning there were a larger number of time intervals that were misclassified as wet as compared with MinMax. Based on the MCC values, for all the threshold values MinMax data outperformed the Average data.

As a further investigation of the performance for wet-dry classification, Fig. 8 shows the box plot for four different statistics (POD, POFD, MCC and ERR with the wet-dry threshold of 0.1 mm h^{-1}). Considering all four statistics, three of the values showed better performance of MinMax compared with Average data. Although the POD value was higher for Average, there were a higher number of both positive and negative misclassifications, which is reflected in the higher value of ERR. The wider range of all four statistics suggests that some of the links

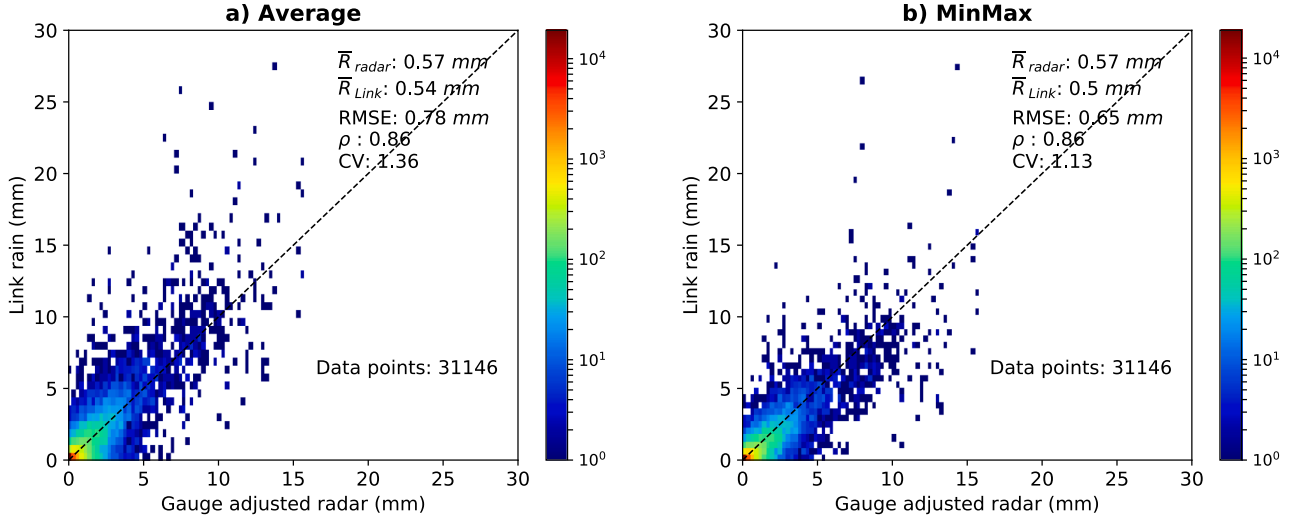

Fig. 12. Scatter density plots of path-average CML-rainfall against gauge-adjusted radar rainfall for 342 rainfall events.

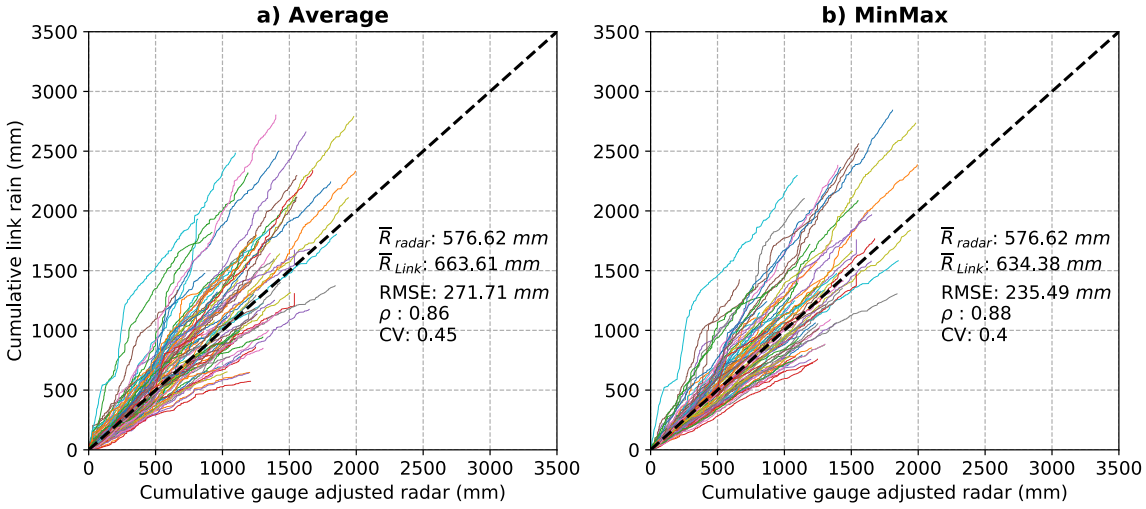

Fig. 13. Double mass curves all links: (a) Link-derived cumulative rainfall using Average data versus cumulative gauge-adjusted radar rainfall; and (b) Link-derived cumulative rainfall using MinMax data versus cumulative gauge-adjusted rainfall data.

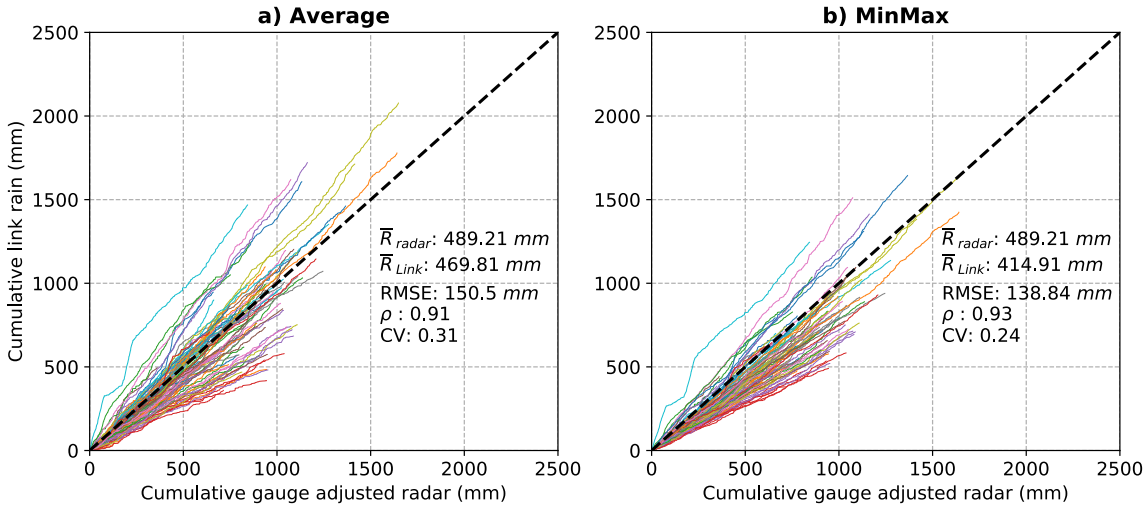

Fig. 14. Double mass curves for all links during only the wet period (no false alarm included) for: (a) Average, and (b) MinMax data.

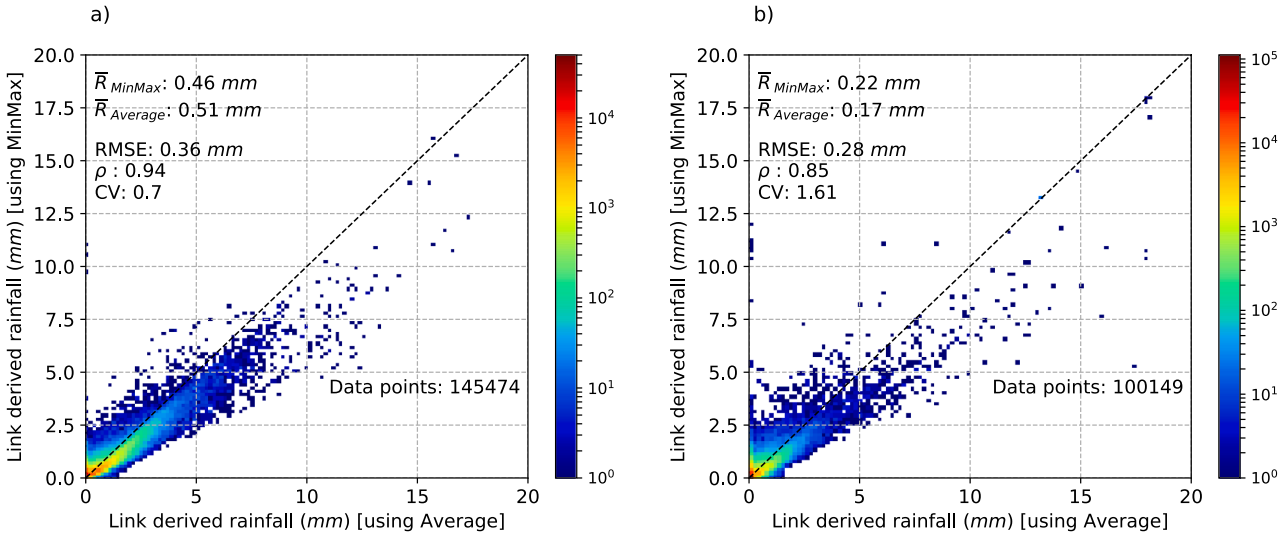

Fig. 15. Comparison of link-derived rainfall retrievals using 15-min Average vs MinMax RSL data for: (a) wet periods only; and, (b) all time intervals with false alarm only.

were performing poorly. Graf et al. (2020) obtained similar statistics in their study for Germany, where they found the median MCC value of 0.47, which is slightly higher than that obtained for MinMax data here. In their case, they used data with a temporal resolution of 1 min for the wet-dry classification.

In order to further investigate the performance of the wet-dry classification for different rain rates, the POD was calculated for each of the different classes of data exceeding certain threshold rain rates from the reference data as shown in Fig. 9. The probability of detecting rain increases with higher rain rates, reaching more than 90 % when 5 mm h^{-1} was used as a threshold rain rate rather than a lower rainfall rate. This suggests that the correct detection rate of wet periods is better under more intense rainfall conditions.

Similarly, Fig. 10 shows the accuracy of the links in determining rainfall occurrence for various accumulation intervals. Here, a threshold of 0.1 mm h^{-1} was used to distinguish wet from dry periods using the gauge-adjusted radar data. Although the POD increased and both POFD and POFA decreased as the considered intervals became longer, the error rate simultaneously increased. This is due to a decrease in the relative proportion of true negative (TN) values compared with the lower accumulation interval. The increase in MCC values suggests that the performance increases for longer accumulation intervals for both Average and MinMax data.

5.2. *Performance for the rainfall retrievals*

Fig. 11 shows the comparison of link-derived rainfall with gauge-adjusted radar data for 15-min, 1-hour, 3-hour and 1-day accumulations for both MinMax and Average. The accuracies of the link-derived rainfall increase for longer durations for both sampling strategies. This

can be seen by the decrease in the value of CV (for Average RSL 2.84 to 1.49 and for MinMax 2.70 to 1.22), and the increase in the value of the correlation coefficient (for

Average 0.29 to 0.75 and MinMax 0.30 to 0.79). A systematic overestimation in link-derived rainfall estimates with respect to gauge-adjusted radar data was found for all accumulation intervals for both sampling strategies. The lower values of RMSE and CV for MinMax, the smaller overestimation and generally similar values for the correlation coefficient indicate that it outperforms Average.

In addition, Table 6 shows the results for the performance of the rainfall retrieval on a seasonal basis. For both sampling strategies (Average and MinMax), link-derived rainfall is overestimated for all four seasons, with the largest overestimation occurring during the Summer and Autumn. This larger magnitude of overestimation is mainly attributed to precipitation events with the higher intensity of rainfall during these two seasons. The performance in terms of bias and CV was better for winter and spring for both sampling strategies compared with the other two seasons. This result differs from the result presented by Graf et al. (2020) for the winter season in Germany. They obtained the lowest performance during the winter month (with the highest overestimation and higher CV values of 16.04) due to the presence of both mixed and solid precipitation which is not the case for this study. Further, this overestimation during other seasons was most likely due to dew formation on the antenna covers as solid/melting snow does not occur in the study area. Similarly, for all seasons link-derived rainfall corresponded well to the validation data for hourly accumulations. The best performance in terms of correlation coefficient and CV was found in the Spring followed by the Summer. Based on the two sampling strategies, MinMax had a lower overestimation and a better CV value compared with Average data.

To have a better understanding of the performance of CML rainfall retrievals, event-based results were also analysed for CML-derived and gauge-adjusted radar rainfall data as shown in Fig. 12. Here, a rainfall event was defined as a rain period separated by a 1-hour or longer rain-free period and having each 15-min time interval with minimum rainfall rate of 0.1 mm h^{-1}. There were altogether 342 such rainfall events with periods lasting from 45 min to 29.25 h. Compared to the results presented in Fig. 11, correlation coefficients were significantly higher, with values of 0.86 for both Average and MinMax data. In terms of relative bias Average was closer to the reference gauge-adjusted radar dataset compared with MinMax. However, other statistics (RMSE and CV) were lower for MinMax, showing the better performance.

In order to further investigate the continuous performance of link-derived rainfall estimation, double mass curves between link-derived and gauge-adjusted radar rainfall are shown as accumulation plots in Fig. 13. Intervals where either link or radar had missing data were excluded. Most of the links rainfall were passing through the 1:1 black line indicates a good agreement between the link-derived rainfall and the gauge-adjusted radar reference data. In overall comparison, both Average and MinMax showed a positive bias of about 15 %. However, 37 links for Average and 40 links for MinMax showed a mean negative bias of 19.30% and 17.34%, respectively. There were almost similar numbers of links (11 links for Average and 10 links for MinMax) for both RSL data types showing overestimation above 50% and up to 135%.

The results presented in Fig. 13 for the rainfall retrieval include both timesteps with the false-positives and false-negatives. As there was a significant difference in the false alarm rate between the two datasets (Average and MinMax), to further investigate the performance for only wet periods, the double mass curves for all links are plotted in Fig. 14 by including only time intervals which have non-zero rainfall for both the links and the radar. In terms of bias, Average data showed better performance with a negative bias of 3.96% compared with MinMax data showing a stronger negative bias of 15.18 %. For a higher number of links (35 links), there was an overestimation based on Average data compared with only 18 links for MinMax data. However, other statistics (RMSE, ρ, CV) suggest that MinMax performed slightly better compared with average RSL.

Fig. 15 provides more insights regarding the performance of rainfall retrieval by the two sampling strategies. The average link-derived rainfall using the MinMax data was lower than the Average RSL (Fig. 12(a)) for time intervals where it was wet according to both sampling datasets (not involving the reference). Furthermore, almost all heavy rainfall depths were underestimated by the MinMax data. In addition, Fig. 12(b) shows the density scatter plot of link-derived rainfall during all the periods when there was a false alarm for both sampling strategies when compared to the reference. This result suggests that the mean rain rate obtained from MinMax data was higher compared with Average data even though the heavy rainfall depths were lower. So, on average overestimations during false alarms were higher for MinMax, but higher for Average in the case of larger rainfall depths.

6. Discussion

6.1. Optimized parameters for RAINLINK

Among the parameters used in RAINLINK, only the four most sensitive parameters were optimized for this Melbourne dataset, specifically the threshold median(Q_{mP}), the threshold median per unit length (Q_{mPL}), wet antenna attenuation (A_a) for Average RSL data and the alpha (α) for MinMax RSL data. In addition, two parameters, *a* and *b,* were obtained based on local disdrometer data from the study of Guyot et al. (2020), and are the most critical for rainfall retrieval of all the parameters. In the absence of these parameters, one needs to use the generalized values from the ITU recommendations (ITU-R, 2016) or, when available, those from other studies from a similar climate. $Q_{mP}$and Q_{mPL}, which are related to dry/wet classification, were obtained separately for the Average and MinMax datasets. For Average RSL data, Q_{mP} and Q_{mPL} were found to be −0.7 dB and −0.2 dB km^{-1} (no reference is available for comparison using the Average RSL data). For the MinMax RSL data, Q_{mP} and Q_{mPL} were found to be -1.50dB and −0.40 dB km^{-1}, which is less negative compared with a similar dataset for The Netherlands. This means using the new parameter values enabled corresponding time intervals to be classified as wet with a lower deviation in median and median per unit length values. The average POD and POFA for the MinMax reported in this study were in the similar range to the values for RAINLINK reported by de Vos et al. (2019).

The optimized value of parameter A_a was found to be 1.6 dB and 1.4 dB for the Average and MinMax data respectively, and was within the range of values suggested by Overeem et al. (2011) (1.2–1.9 dB) for MinMax data. However, for the MinMax data, Overeem et al. (2013) used A_a = 2.3 dB, which is higher than the value obtained for the Melbourne dataset. In their case, A_a and α was optimized for the rainfall after determining the parameters for the wet-dry classification separately and is therefore different than the case presented herein as both wet-dry and rainfall were optimized together. Also, the weighing factor (α = 0.29) for the minimum and maximum attenuation for the Melbourne dataset was slightly lower than obtained by Overeem et al. (2013).

There are other parameters in RAINLINK (like the minimum number of available links, the period over which the reference level has to be determined, the minimum number of hours that should be dry in the preceding period, the outlier filter threshold, and the radius for finding

nearby links) which have not been altered from the original values of RAINLINK. These parameters are more likely to remain constant and have less dependency on the dataset and climatology of the study area. However, the radius for finding the nearby links was increased from 15 km to 20 km, as the density of links in the employed dataset for the Melbourne metropolitan area was low compared to other studies in The Netherlands.

6.2. Effect of sampling type on overall rainfall retrievals

For the wet-dry classification, the Average RSL showed a higher POD when compared with the MinMax RSL data, but at the same time the POFD and POFA values were also higher for the Average RSL data. This means that the Average RSL missed less rain, but the higher POFD shows that often false alarms are provided during dry periods, whereas the MinMax RSL strategy indeed has a lower POD but also a lower POFD. This is mainly because for each of the time intervals, the MinMax RSL contains additional information of 15-minute data characteristics as opposed to the Average RSL data. Similar result has been reported by de Vos et al. (2019) for the MinMax data from the Netherlands. They obtained a POD of 38% and a POFA of 35% for a threshold of 0 mm, a POFA of 50% and a POFA of 40% for a threshold of 0.1 mm. Also, taking into account the MCC and ERR, the MinMax data outperformed the Average RSL strategy for the wet-dry classification.

For the rainfall retrieval, the Average data performed similarly to the MinMax data. In a few cases, considering relative bias, Average data performed slightly better. However, other statistics, such as the CV and RMSE support the MinMax sampling strategy. Thus, in the overall comparison, the MinMax outperformed Average data. The main reason behind this is that the MinMax data had better wet-dry classification compared with Average data. Even though for some cases results favours Average data, for the overall rainfall retrieval process both wet-dry classification and rainfall retrieval are necessary steps. There are no other studies to date that compare the performance of Average versus MinMax data, but the MinMax, the performance is quite similar to the results of Overeem et al. (2016b).

6.3. Limitations of using gauge-adjusted radar data as a reference

In this study, gauge-adjusted radar data were considered as truth, even though radar rainfall estimates are not a direct rainfall measurement close to the ground. This is especially so in this study, where *Rainfields* interpolated the radar volume at 1 km height, is used as a reference. This product is therefore better compared with other similar studies and hence the representativeness error would be smaller as well. Importantly, rainfall measured aloft takes a couple of minutes before the rainfall reaches the earth's surface, and thus advection of rainfall and changes in rainfall intensity between the radar volume and the Earths' surface are other issues associated with the radar rainfall measurements. Thus, these differences in sampling compared to ground-based sensors can lead to considerable differences in observations that could be interpreted as an error in the CML estimates herein. As CMLs provide a path-integrated rainfall measurement, which is compared with the mean rainfall derived from radar cells that overlap the link path, this could also add some errors in the comparison. This is also because there are representativeness errors in radar data at such short time and length scales. Thus, it is also important to present daily results.

7. Conclusion

This study presents rainfall retrievals over the greater Melbourne Metropolis using 135 commercial microwave links operating at frequencies ranging from 10 to 40 GHz and path lengths of 0.2 km to 25 km. This study is the very first to conduct a comparison of rainfall retrieval using CML data with two different sampling strategies over the same link paths (average and min/max RSL data over 15 min). For this, the RAINLINK package was used and a new set of parameters was derived for rainfall retrieval across Melbourne. For the wet-dry classification using the nearby link approach, Average data was found to perform better, with a higher POD and a similar POFA compared with the MinMax data. Other statistics including MCC and ERR suggested that the use of MinMax data achieved fewer false alarms. While the Average data had similar rainfall retrieval performance to using MinMax data when compared for different accumulation intervals, MinMax data provided the best performance based on statistics (relative bias, CV, RMSE and ρ) and double mass curve.

This study used the current optimization approach based on a cost function that considers both dry/wet classification and rainfall retrieval using the specific weights for each of the performance metrics. However, some of the performance metrics are correlated with each other. As an example, a lower POFA could be achieved with a lower POD, but an increased POFA might be stronger associated with a positive bias than increased POD. Thus, for future studies, it is recommended to separate the optimization for dry/wet classification and rainfall retrieval as a two-step process in Wolff et al. (2021).

As the primary purpose of CMLs is to provide reliable communication services rather than contribute to rainfall monitoring, utilising available data that are not optimised for this purpose will remain challenging. Understanding the feasibility and improvement in rainfall measurement from CML datasets is crucial for future operational use of the data available from mobile network operators.

This study has validated the use of more than 100 CMLs data for rainfall retrieval in the Melbourne metropolitan area, making it the very first large-scale study in Australia. Here, gauge-adjusted radar data were used as reference data even though there are limitations of such a radar-derived product. Thus, it is therefore recommended to compare with nearby gauge data for the link performance. Furthermore, it is recommended that CML rainfall estimates be used as a complementary source of information where there are no radars or alternative rainfall measurement instrument. At the city scale, CML-derived rainfall estimates can also help overcome data gaps due to radar clutter from high-rise buildings and absence of traditional rain gauges. Accordingly, there is great opportunity to combine the three datasets into a "merged product". Moreover, studying the performance of different sampling strategies in detail with a dedicated microwave link data is expected to provide greater insight into the optimal sampling strategy. This may also contribute to an improved rainfall retrieval algorithm.

8. Software and model codes

Rainfall retrieval was undertaken using the R package RAINLINK version 1.2, which is freely available on GitHub (https://github.com/overeem11/RAINLINK). All other data processing and plotting was done using Python.

9. Data availability

The CML data is not publicly available due to a non-disclosure agreement. Other data are available upon request.

CRediT authorship contribution statement

Jayaram Pudashine: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Adrien Guyot:**

Methodology, Software, Writing - review & editing, Supervision. **Aart Overeem:** Methodology, Software, Writing - review & editing. **Valentijn R.N. Pauwels:** Writing - review & editing, Supervision. **Alan Seed:** Writing - review & editing. **Remko Uijlenhoet:** Methodology, Writing - review & editing. **Mahesh Prakash:** Writing - review & editing. **Jeffrey P. Walker:** Writing - review & editing, Supervision, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

An Australian Research Council Discovery Project (Grant number DP160101377) supported this project. The authors also acknowledge the Commonwealth Scientific and Industrial Research Organisation (CSIRO) for the top-up scholarship for the first author of his study. The authors would like to thank Motorola, the Bureau of Meteorology (BoM), and Melbourne Water for providing data. The authors declare no conflict of interest.

Appendix

Table 6
(Table 7)

Table 7
Dataset used for calibration of RAINLINK parameters.

Day	Time steps with rain	Total rain (mm)	Maximum rainfall intensity (mm/hr)
2017–08-06	34	10.64	4.40
2017–08-07	32	7.87	3.34
2017–08-15	33	8.87	6.17
2017–08-18	51	9.53	2.86
2017–08-26	28	13.65	8.47
2017–09-05	33	12.61	7.12
2017–09-07	38	10.35	3.76
2017–09-08	36	5.85	2.28
2017–09-13	33	8.21	3.92
2017–09-15	51	30.13	19.64
2018–06-08	35	12.64	5.25
2018–06-15	22	6.74	6.07
2018–06-16	48	30.19	7.44
2018–07-07	57	19.55	5.03
2018–07-08	70	15.42	3.66
2018–08-07	31	9.80	4.08
2018–08-12	39	6.06	2.27
2018–08-18	51	18.52	6.57
2018–08-19	38	7.59	3.99
2018–08-21	38	4.29	2.35
2018–10-09	37	7.25	4.11
2018–10-16	24	8.85	6.99
2018–10-17	21	4.20	3.72
2018–10-19	27	13.35	6.87
2018–11-07	32	9.66	5.17
2018–11-20	31	32.12	17.69
2018–11-21	24	13.14	8.97
2018–11-22	64	33.66	9.22
2018–11-23	58	17.93	4.07
2018–12-20	26	3.18	0.94

Table 8
Comparison of performance of rainfall retrieval using default and calibrated RAINLINK parameters (The parameters a and b for both RAINLINK default and calibrated were based on the local disdrometer data).

	RAINLINK default parameter	RAINLINK calibrated parameters
POD	27.20	50.10
POFD	5.41	3.90
MCC	0.30	0.45
Correlation coefficient	0.29	0.30
Percent bias	−80.99	−1.50

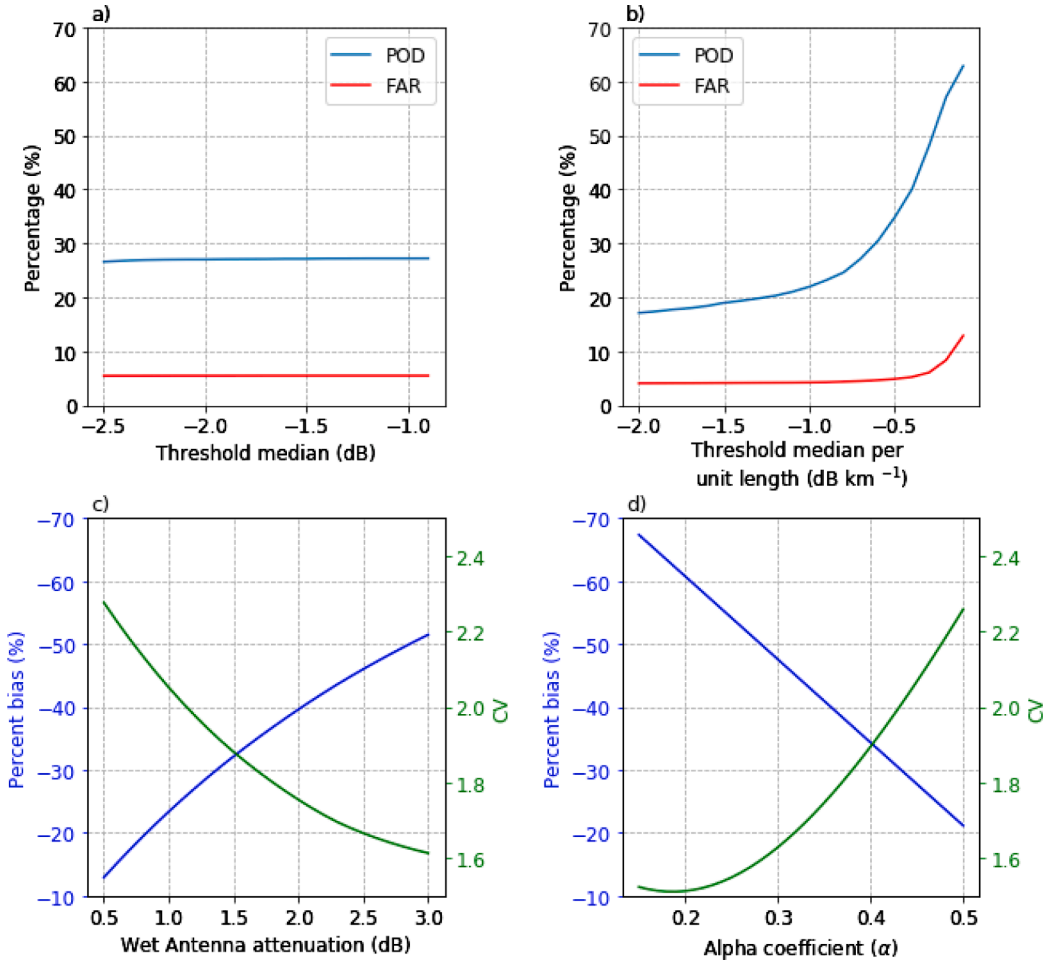

Fig. A1. Sensitivity analysis using MinMax RSL data for (a) Threshold median; (b) Threshold median per unit length; (c) Wet antenna attenuation; and (d) Alpha coefficient.

References

Australian Bureau of Statistics: https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/2GMEL?opendocument, access: 16/04/2020, 2016.

Atlas, D., Ulbrich, C.W., 1977. Path-and Area-Integrated Rainfall Measurement by Microwave Attenuation in the 1–3 cm Band. J. Appl. Meteorol. 16, 1322–1331. https://doi.org/10.1175/1520-0450(1977)016<1322:PAAIRM>2.0.CO;2.

Barnes, L.R., Schultz, D.M., Gruntfest, E.C., Hayden, M.H., Benight, C.C., 2009. CORRIGENDUM: False Alarm Rate or False Alarm Ratio? Weather Forecas. 24, 1452–1454. https://doi.org/10.1175/2009waf2222300.1.

Berne, A., Krajewski, W.F., 2013. Radar for hydrology: Unfulfilled promise or unrecognized potential? Adv. Water Resour. 51, 357–366. https://doi.org/10.1016/j.advwatres.2012.05.005.

Bianchi, B., Rieckermann, J., Berne, A., 2013. Quality control of rain gauge measurements using telecommunication microwave links. J. Hydrol. 492, 15–23. https://doi.org/10.1016/j.jhydrol.2013.03.042.

Christopher, S.R., Kultegin, A., Savyasachee, M., Justin, P.B., 1996. 35-GHz Dual-Polarization Propagation Link for Rain-Rate Estimation. J. Atmos. Oceanic Technol. 13 (2), 419–425.

Chwala, C., Gmeiner, A., Qiu, W., Hipp, S., Nienaber, D., Siart, U., Eibert, T., Pohl, M., Seltmann, J., Fritz, J., Kunstmann, H., 2012. Precipitation observation using microwave backhaul links in the alpine and pre-alpine region of Southern Germany. Hydrol. Earth Syst. Sci. 16, 2647–2661. https://doi.org/10.5194/hess-16-2647-2012.

Chwala, C., Keis, F., Kunstmann, H., 2016. Real-time data acquisition of commercial microwave link networks for hydrometeorological applications. Atmos. Meas. Tech. 9, 991–999. https://doi.org/10.5194/amt-9-991-2016.

Chwala, C., Kunstmann, H., 2019. Commercial microwave link networks for rainfall observation: Assessment of the current status and future challenges, Wiley Interdisciplinary Reviews. Water 6 (2). https://doi.org/10.1002/wat2.2019.6.issue-210.1002/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. J. Atmos. Oceanic Technol. 36, 1267–1283. https://doi.org/10.1175/jtech-d-18-0197.1.

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. Geophys. Res. Lett. 41 (16), 6016–6022. https://doi.org/10.1002/2014GL060724.

Doviak, R. J.: Doppler radar and weather observations, In: Zrnić, D. a. S., Orlando, Fla. (Eds.) Academic Press, Orlando, Fla., 1984.

Ericsson: Ericsson microwave outlook: trends and needs in the microwave industry, 2017.

Fencl, M., Dohnal, M., Rieckermann, J., Bareš, V., 2017. Gauge-adjusted rainfall estimates from commercial microwave links. Hydrol. Earth Syst. Sci. 21 (1), 617–634. https://doi.org/10.5194/hess-21-617-201710.5194/hess-21-617-2017-supplement.

Fencl, M., Rieckermann, J., Schleiss, M., Stransky, D., Bares, V., 2013. Assessing the potential of using telecommunication microwave links in urban drainage modelling, Water science and technology : a journal of the International Association on Water. Pollution Research 68, 1810–1818. https://doi.org/10.2166/wst.2013.429.

Gazit, L., Messer, H., 2018. Advancements in the Statistical Study, Modeling, and Simulation of Microwave-Links in Cellular Backhaul Networks. Environments 5, 75. https://doi.org/10.3390/environments5070075.

Germann, U., Galli, G., Boscacci, M., Bolliger, M., 2006. Radar precipitation measurement in a mountainous region. Q. J. R. Meteorolog. Soc. 132 (618), 1669–1692. https://doi.org/10.1256/qj.05.190.

Giuli, D., Toccafondi, A., Gentili, G.B., Ferni, A., 1991. Tomographic reconstruction of rainfall fields through microwave attenuation measurments. J. Appl. Meterol. 30, 17.

Goldshtein, O., Messer, H., Zinevich, A., 2009. Rain Rate Estimation Using Measurements From Commercial Telecommunications Links. IEEE Trans. Signal Process. 57 (4), 1616–1625. https://doi.org/10.1109/TSP.2009.2012554.

Graf, M., Chwala, C., Polz, J., Kunstmann, H., 2020. Rainfall estimation from a German-wide commerical microwave link network : Optimized processing and validation for 1 year of data. Hydrol. Earth Syst. Sci. 24, 2931–2950. https://doi.org/10.5194/hess-24-2931-2020.

Guyot, A., Pudashine, J., Protat, A., Uijlenhoet, R., Pauwels, V.R.N., Seed, A., Walker, J. P., 2020. Effect of disdrometer type on rain drop size distribution characterisation: a new dataset for south-eastern Australia. Hydrol. Earth Syst. Sci. 23, 4737–4761. https://doi.org/10.5194/hess-23-4737-2019.

Hogg, D.C., 1968. Millimeter-Wave Communication through the Atmosphere: The known and unknown features in propagation of short radio waves are discussed. Science (New York, N.Y.) 159 (3810), 39–46. https://doi.org/10.1126/science:159.3810.39.

Holt, A. R., Kuznetsov, G. G., and Rahimi, A. R. 2003.Comparison of the use of dual-frequency and single-frequency attenuation for the measurement of path-averaged rainfall along a microwave link, Iee P-Microw Anten P, 150, 315-320, https://doi.org/10.1049/ip-map:20030616.

ITU-R: 2005. Recommendation ITU-R P.838-3.

ITU-R: 2016. Attenuation by atmospheric gases, Geneva, 276-211.

Joss, J., Waldvogel, A., Collier, C.G., 1990. In: Radar in Meteorology. American Meteorological Society, Boston, MA, pp. 577–606. https://doi.org/10.1007/978-1-935704-15-7_39.

Kucera, P.A., Ebert, E.E., Turk, F.J., Levizzani, V., Kirschbaum, D., Tapiador, F.J., Loew, A., Borsche, M., 2013. Precipitation from Space: Advancing Earth System Science. Bull. Am. Meteorol. Soc. 94 (3), 365–375. https://doi.org/10.1175/BAMS-D-11-00171.1.

Leijnse, H., Uijlenhoet, R., Stricker, J.N.M., 2007. Rainfall measurement using radio links from cellular communication networks. Water Resour. Res. 43 (3) https://doi.org/10.1029/2006WR005631.

Leinonen, J., 2014. High-level interface to T-matrix scattering calculations: architecture, capabilities and limitations. Opt. Express 22, 1655–1660. https://doi.org/10.1364/OE.22.001655.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phase lysozyme. BBA 405, 442–451.

Mello, L.A.R.d.S, Costa, E., Souza, R.S.L.d., 2002. Rain attenuation measurements at 15 and 18 GHz. Electron. Lett. 38, 197–198. https://doi.org/10.1049/el:20020105.

Messer, H., 2018. Capitalizing on Cellular Technology—Opportunities and Challenges for Near Ground Weather Monitoring. Environments 5, 73. https://doi.org/10.3390/environments5070073.

Messer, H., Zinevich, A., and Alpert, P.: Environmental Monitoring by Wireless Communication Networks, Science, 312, 713-713, https://doi.org/10.1126/science.1120034, 2006.

Mishchenko, Michael I., Travis, Larry D., 1994. T-matrix computations of light scattering by large spheroidal particles. Opt. Commun. 109 (1-2), 16–21. https://doi.org/10.1016/0030-4018(94)90731-5.

Olsen, R., Rogers, D.V., Hodge, D.B., 1978. The aRb relation in the calculation of rain attenuation. IEEE Trans. Antennas Propag. 26, 318–329. https://doi.org/10.1109/TAP.1978.1141845.

Overeem, A., Leijnse, H., Uijlenhoet, R., 2011. Measuring urban rainfall using microwave links from commercial cellular communication networks. Water Resour. Res. 47 (12) https://doi.org/10.1029/2010WR010350.

Overeem, A., Leijnse, H., Uijlenhoet, R., 2013. Country-wide rainfall maps from cellular communication networks. PNAS 110 (8), 2741–2745. https://doi.org/10.1073/pnas.1217961110.

Overeem, A., Leijnse, H., Uijlenhoet, R., 2016a. Retrieval algorithm for rainfall mapping from microwave links in a cellular communication network. Atmos. Meas. Tech. 9, 2425–2444. https://doi.org/10.5194/amt-9-2425-2016.

Overeem, A., Leijnse, H., and Uijlenhoet, R.: Two and a half years of country-wide rainfall maps using radio links from commercial cellular telecommunication networks, Water Resources Research, n/a-n/a, https://doi.org/10.1002/2016WR019412, 2016b.

Pastorek, Jaroslav, Fencl, Martin, Rieckermann, Jörg, Bareš, Vojtěch, 2019. Commercial microwave links for urban drainage modelling: The effect of link characteristics and their position on runoff simulations. J. Environ. Manage. 251, 109522. https://doi.org/10.1016/j.jenvman.2019.109522.

Rios Gaona, M.F., Overeem, A., Leijnse, H., Uijlenhoet, R., 2015. Measurement and interpolation uncertainties in rainfall maps from cellular communication networks. Hydrol. Earth Syst. Sci. 19, 3571–3584. https://doi.org/10.5194/hess-19-3571-2015.

Rios, G.M.F., Overeem, A., Raupach, T.H., Leijnse, H., Uijlenhoet, R., 2017. Rainfall retrieval with commercial microwave links in São Paulo, Brazil. Atmos. Meas. Tech. 1–21 https://doi.org/10.5194/amt-2017-287.

Roversi, G., Alberoni, P.P., Fornasiero, A., Porcù, F., 2020. Commercial Microwave Links as a tool for operational rainfall monitoring in Northern Italy. Atmos. Meas. Tech. Discuss. 1–30, 2020. https://doi.org/10.5194/amt-2020-43.

Russell, B., Williams, E.R., Gosset, M., Cazenave, F., Descroix, L., Guy, N., Lebel, T., Ali, A., Metayer, F., Quantin, G., 2010. Radar/rain-gauge comparisons on squall lines in Niamey, Niger for the AMMA. Q. J. R. Meteorolog. Soc. 136 (S1), 289–303. https://doi.org/10.1002/qj.548.

Seed, A., Leahy, C., Duthie, E., and Chumchean, S.: Rainfields : The Australian Bureau of Meteorology System for Quantitative Precipitation Estimation, and It's Use in Hydrological Modelling, Proceedings of Water Down Under 2008, Modbury, SA, 2008, 661-670,.

Smiatek, Gerhard, Keis, Felix, Chwala, Christian, Fersch, Benjamin, Kunstmann, Harald, 2017. Potential of commercial microwave link network derived rainfall for river runoff simulations. Environ. Res. Lett. 12 (3), 034026. https://doi.org/10.1088/1748-9326/aa5f46.

Sohail Afzal, M., Shah, S.H.M., Cheema, M.J.M., Ahmad, R., 2018. Real time rainfall estimation using microwave signals of cellular communication networks: a case study of Faisalabad, Pakistan. Hydrol. Earth Syst. Sci. Discuss. 1–20 https://doi.org/10.5194/hess-2017-740.

Uijlenhoet, Remko, Overeem, Aart, Leijnse, Hidde, 2018. Opportunistic remote sensing of rainfall using microwave links from cellular communication networks, Wiley Interdisciplinary Reviews. Water 5 (4). https://doi.org/10.1002/wat2.2018.5.issue-410.1002/wat2.1289.

Wesson, S.M., Pegram, G.G.S., 2006. Improved radar rainfall estimation at ground level. Nat. Hazards Earth Syst. Sci. 6, 323–342. https://doi.org/10.5194/nhess-6-323-2006.

Wolff, W., Overeem, A., Leijnse, H., Uijlenhoet, R., 2021. Rainfall retrieval algorithm for commercial microwave links: stochastic calibration. Atmos. Meas. Tech. Discuss. 1–22, 2021. https://doi.org/10.5194/amt-2021-34.