**Atmospheric
Measurement
Techniques**

Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals

Z. Wang[1]**, M. Schleiss**[2]**, J. Jaffrain**[2]**, A. Berne**[2]**, and J. Rieckermann**[1]

[1]The Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[2]Laboratoire de Télédétection Environnementale (LTE), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland

Correspondence to: Z. Wang (zhe.wang@eawag.ch)

Abstract. A Markov switching algorithm is introduced to classify attenuation measurements from telecommunication microwave links into dry and rainy periods. It is based on a simple state-space model and has the advantage of not relying on empirically estimated threshold parameters. The algorithm is applied to data collected using a new and original experimental set-up in the vicinity of Zürich, Switzerland. The false dry and false rain detection rates of the algorithm are evaluated and compared to 3 other algorithms from the literature. The results show that, on average, the Markov switching model outperforms the other algorithms. It is also shown that the classification performance can be further improved if redundant information from multiple channels is used.

1 Introduction

Precipitation is an important component of the Earth's water cycle and needs to be accurately measured. So far, several techniques have been proposed to measure rainfall with different spatial and temporal resolutions, ranging from traditional point measurements from rain gauges to observations from weather radars and satellites. Each of these techniques has its advantages, but also its limitations (Sevruk, 1999; Upton et al., 2005; Germann et al., 2006).

Recently, microwave links (MWL), which are commonly used in telecommunication networks for wireless data transmission, have been suggested as a novel tool to monitor rainfall in urban areas (Messer et al., 2006; Leijnse et al., 2007c). The main idea behind this technique is to relate the rain-induced signal attenuation to the path-averaged rain rate along the considered link. The potential of this technique has been demonstrated using microwave links specifically designed for rainfall estimation (Ruf et al., 1996; Rahimi et al., 2003; Holt et al., 2003; Upton et al., 2005; Krämer et al., 2005) and commercial microwave links operated by telecommunication companies (Messer et al., 2006; Zinevich et al., 2009). Note that, in addition to estimating rain rates, MWL can also be used to measure evaporation (Leijnse et al., 2007a) and water vapour (David et al., 2009). In fact, MWL nicely complement traditional rainfall sensors because they provide rain rate measurements (near the ground level) at an intermediate scale between point measurements from rain gauges and weather radars with sampling volumes up to several km^3. The fact that MWL networks can be very dense can also be used to improve rain rate estimates using spatial interpolation techniques (Zinevich et al., 2008).

A very important issue that needs to be addressed prior to rainfall estimation using MWL is the so-called baseline estimation problem (Rahimi et al., 2003; Leijnse et al., 2007c). It consists of identifying and separating the attenuation which occurs during dry periods from the rain-induced attenuation (which is the quantity of interest in most applications). Depending on the link characteristics, this problem can be very difficult (Upton et al., 2005). Dry-weather signal attenuations can exhibit significant variability caused, for example, by changes in water vapour, wind effects on the antennas, birds or insects crossing the beam, losses during transmission or reception, interferences, wet-antenna and multi-path effects (Zinevich et al., 2010). Moreover, attenuation measurements are often quantized, which introduces additional variability in the process. It is only after the attenuation baseline has

been properly estimated for each time step that the corresponding path-averaged rain rate can be retrieved. A very important step in this procedure is the ability to identify the dry and rainy periods using solely the measurements of the MWL. The measured attenuation levels during the dry periods can then be used to better estimate the attenuation baseline during rainy periods.

So far, various techniques have been suggested to solve this identification problem (see Sect. 2 for a detailed description). The simplest of them uses a global attenuation threshold, i.e., all the periods for which the path-integrated attenuation (PIA hereinafter) is above a given threshold are considered rainy and vice versa (Leijnse et al., 2007b). A slightly more sophisticated procedure was suggested by Schleiss and Berne (2010) who proposed to use a threshold on the temporal variability of the PIA. A more complex algorithm by Reller et al. (2011) investigates the possibility to identify dry and wet periods using a Bayesian approach based on Factor Graphs (Loeliger et al., 2010). The underlying idea is that the variations of the baseline are learned during dry weather and propagated to wet periods. Although it is more sophisticated and flexible, the algorithm also requires the choice of a subjectively estimated threshold for classification. Finally, special algorithms have been developed for situations where signals from dual-frequency MWLs or simultaneous PIA's from two communication channels are available. For example, Holt et al. (2003) and Rahimi et al. (2003) proposed a method based on the assumption that the correlation between the attenuations from two different frequencies tends to be higher during rainy periods. Such methods, however, also rely on empirically adjusted thresholds for the classification, which is not necessarily optimal.

In this article, a new classification algorithm based on Markov switching models is introduced. It is based on a simple state-space model and has the advantage of not relying on any empirically estimated threshold parameters. Also, the proposed algorithm can be easily generalized to multivariate inputs, i.e., inputs from different channels or frequencies. A real-world application of the algorithm (see Sect. 4) shows that it performs better than other existing techniques and that its performance can be improved if multiple channel inputs are considered. The proposed algorithm thus improves the signal processing of MWL data and helps estimating better attenuation baselines required for accurate rainfall retrieval.

This article is structured as follows: Sect. 2 describes some of the existing classification methods and introduces the Markov switching model. In Sect. 3, the experimental setup used to quantify the performances of the classification algorithms is described. Section 4 evaluates and compares the performances of the different algorithms for two very different datasets. Possible improvements of the algorithm are then discussed in Sect. 5. The conclusions are given in Sect. 6.

2 Methods

This section briefly describes some of the existing methods proposed in the literature to identify dry and rainy periods using single channel MWL attenuation measurements. Then, a new classification method based on Markov switching models is introduced and discussed in detail.

2.1 Existing algorithms

Three popular classification methods have been chosen: the simple threshold method (ST), the moving window method (MW) and the Factor Graph (FG). The simple threshold algorithm (Leijnse et al., 2007b) is straightforward and computationally efficient. It uses a global threshold on the path-integrated attenuation to distinguish between the dry and the rainy periods. Each time period for which the PIA is above the threshold is classified as rainy, and vice versa.

$$\text{Decision rule for ST:} \begin{cases} \text{rainy if } A_t > a_0 \\ \text{dry if } A_t \leq a_0 \end{cases} \quad (1)$$

where A_t [dB] denotes the path-integrated attenuation at time t and a_0 [dB] is a given threshold value. The method has shown to produce good results in practical applications and can be applied in real-time. Finding the optimal detection threshold a_0 is, however, difficult. Moreover, the performance (in terms of false dry and rain detections) of this algorithm can be very sensitive to the value of the threshold. Finally, this method is only appropriate for datasets for which the dry-weather attenuation is more or less constant. This is not always the case as can be seen in the right panel of Fig. 2. In some situations, the dry-weather attenuation exhibits clear daily cycles and a strong temporal drift in the PIA, possibly due to changes in temperature between day and night and hardware instabilities. Obviously, the simple threshold is not appropriate for such types of signals and alternative classification methods have been suggested.

A slightly more complex approach which may be better suited for non-stationary dry-weather attenuations has been proposed by Schleiss and Berne (2010). Their method, hereinafter referred to as the moving window algorithm, is based on the assumption that the temporal variability of the PIA is small and bounded during dry weather. On the other hand, rainy periods are characterised by larger signal fluctuations. Hence, each time period is classified according to the following decision rule:

$$\text{Decision rule for MW:} \begin{cases} \text{rainy if } S_{W_t} > \sigma_0 \\ \text{dry if } S_{W_t} \leq \sigma_0 \end{cases} \quad (2)$$

where S_{W_t} [dB] represents the local (temporal) variability of the signal attenuation for a moving window $[t - w, t]$ and σ_0 [dB] is a rain detection threshold estimated using one of the two approaches described in Schleiss and Berne (2010). The moving window algorithm is also computationally efficient and can be applied in real-time to non-stationary time

series of attenuations. However, finding the optimal detection threshold σ_0 can be very difficult without appropriate calibration data over extended periods of time. Moreover, one of the main disadvantages of the moving window algorithm is its inability to separate light rain from dry periods because both signals exhibit similar variability.

The Factor Graph algorithm proposed by Reller et al. (2011) can also be applied to non-stationary MWL signals, but does not require large datasets for model calibration. A Factor Graph is a particular type of graphical model, with applications in Bayesian inference, which computes marginal distributions through the sum-product message passing algorithm (Kschischang et al., 2001). More specifically, the Factor Graph algorithm models the attenuation baseline during dry weather using a line model whose parameters can vary slowly over time together with periodicity constraints. In this, it assumes a smoothly varying baseline and, where the signal exceeds a certain threshold, the algorithm identifies that the system enters another state. The Factor Graph algorithm possesses several advantages, as it can deal with irregular time series and not only identifies dry and rainy periods, but simultaneously estimates the baseline and, thus, delivers the rain-induced attenuation. However, it also relies on several tuning parameters that need to be estimated subjectively prior to the classification into dry and rainy periods.

In the following, a new algorithm for the identification of dry and rainy periods based on MWL attenuation measurements is introduced. It uses Markov switching models to estimate the state of the system (i.e., dry or rainy).

2.2 Univariate Markov switching model (MSU)

A Markov switching model combines dynamic linear system behaviour with a Markov process, which models the transitions between different states. It belongs, similarly to the Factor Graph, to a very general class of so-called state-space models. Such models are commonly used to model a change in behaviour with respect to different regimes. The regimes themselves can be related to certain events, often stochastic, such as a financial crisis or changes in government policy. Practical applications of such models can be found (among others) in the fields of Economics (Hamilton, 1989) and Physics (Yue and Han, 2005; Metzner et al., 2007). Markov switching models have also been used in weather generators to model rainfall patterns (Weiss, 1964).

For simplicity, the details of the algorithm are only given for the univariate case, i.e., a single channel input. The multivariate case is briefly described at the end of this section. For more details on Markov switching models, the reader is referred to Hamilton (1989, 1990) and Kim (1994).

Note that Rayitsfeld et al. (2011) proposed a similar approach based on a hidden Markov model with a slightly different implementation. They did, however not compare their method with previously proposed classification techniques.

The underlying assumption of the Markov switching algorithm is that the magnitude and the variability of the PIA are fundamentally different during dry and rainy periods. During dry periods, the PIA mildly fluctuates around a given value, while for rainy periods it is much larger and variable. This additional variability is caused by the scattering and absorption of the transmitted signal by the raindrops along the path of the link. Hence, it should be possible to identify two fundamentally different states of the system (dry/rainy) from the different behaviour of the PIA. For example, the following, very simple model can be used to describe the data:

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period} \\ \mu_1 + \varepsilon_1 & \text{for every rainy period} \end{cases} \qquad (3)$$

where A_t [dB] represents the path-integrated attenuation at time t, μ_0 [dB] and μ_1 [dB] represent the average value of the attenuation during dry and rainy periods. The noise terms ε_0 [dB] and ε_1 [dB] are assumed to be independent Gaussian random variables with zero mean and standard deviations given by σ_0 [dB] and σ_1 [dB]. The transitions between the dry and the rainy periods are modelled using a stationary hidden random variable $S_t \in \{0,1\}$ where

$$S_t = \begin{cases} 0 & \text{for every dry period} \\ 1 & \text{for every rainy period} \end{cases} \qquad (4)$$

The unconditional probability of the system being in the dry state is denoted by $p_0 = \mathbf{Pr}(S_t = 0) = 1 - p_1$. Combining Eqs. (3) and (4), it is possible to write A_t using a single expression given by

$$A_t = \mu_{S_t} + \varepsilon_{S_t} \qquad (5)$$

with 5 model parameters $\Theta = (\mu_0, \mu_1, \sigma_0, \sigma_1, p_0)$. The maximum likelihood technique is then used to infer the optimal model parameters for a given set of observations $\{A_t = a_t\}$:

$$\hat{\Theta} = \arg\max \{l(\Theta)\} \qquad (6)$$

where the log-likelihood function $l(\Theta)$ is given by

$$l(\Theta) = \sum_t \log\left[\sum_{k=0}^1 f_k(a_t, \Theta) p_k(a_t, \Theta)\right] \qquad (7)$$

with

$$f_k(x, \Theta) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \qquad k = 0, 1$$

and

$$p_k(x, \Theta) = \frac{p_k f_k(x, \Theta)}{p_0 f_0(x, \Theta) + (1 - p_0) f_1(x, \Theta)} \qquad k = 0, 1$$

where $f_k(x, \Theta)$ denotes the probability density for a given state k and $p_k(x, \Theta)$ the associated state probability (for a

given set of model parameters Θ). The maximization of $l(\Theta)$ is performed using a standard Newton-type algorithm. In order to be valid, the solution must satisfy some simple conditions. Specifically, one must have $1 > p_0 > 0$, $\sigma_1 > \sigma_0 > 0$ and $\mu_1 > \mu_0 > 0$. Once $\hat{\Theta}$ has been estimated, the classification into dry and rainy periods can be easily derived from the estimated state probabilities $p_0(a_t, \hat{\Theta}) = 1 - p_1(a_t, \hat{\Theta})$. In absence of any prior information and two states, dry and rainy, this leads to a threshold of $\frac{1}{2}$.

$$\hat{s_t} = \begin{cases} 0 & \text{if } p_0(a_t, \hat{\Theta}) > \frac{1}{2} \\ 1 & \text{else} \end{cases} \tag{8}$$

Note that it is also possible to choose another threshold depending on the relative cost associated to each of the classification errors. One of the advantages of the Markov switching model is that it can be easily generalized to include multivariate inputs from different channels or frequencies.

2.3   Multivariate Markov switching model (MSM)

Telecommunication microwave links are usually operated using multiple channels such as two directions, frequencies or polarizations. This redundant information can be used to improve the classification performance.

In the multivariate case with N channels, the attenuation at time step t is given by a vector $A_t = (A_t^{(1)}, ..., A_t^{(N)})$ where

$$A_t^{(j)} = \mu_{S_t}^{(j)} + \varepsilon_{S_t}^{(j)} \qquad \forall j = 1, ..., N. \tag{9}$$

Note that the vector of model parameters is now significantly longer and given by $\Theta = (\mu_0^{(1:N)}, \mu_1^{(1:N)}, \sigma_0^{(1:N)}, \sigma_1^{(1:N)}, p_0)$, that is, $4N + 1$ variables to estimate. The major difference with respect to the univariate case concerns the difficulty to estimate the joint densities $f_0(A_t, \Theta)$ and $f_1(A_t, \Theta)$, although significant simplifications occur if the channels are assumed independent. While this is certainly not the case for rainy periods, it is, at least, reasonable during dry periods (which usually represent the majority of all the periods). In the absence of any further information, a pragmatic solution, therefore, consists in assuming that all channels are independent and that the log-likelihood function is given by

$$l(\Theta) = \sum_t \log \left[\sum_{k=0}^{1} \prod_{j=1}^{N} f_k^{(j)}(a_t^{(j)}, \Theta) \, p_k(a_t, \Theta) \right] \tag{10}$$

where

$$f_k^{(j)}(x, \Theta) = \frac{1}{\sqrt{2\pi(\sigma_k^{(j)})^2}} \exp\left(-\frac{(x - \mu_k^{(j)})^2}{2(\sigma_k^{(j)})^2} \right) \quad k = 0, 1$$

and

$$p_k(a_t, \Theta) = \frac{p_k \prod_{j=1}^{N} f_k^{(j)}(a_t^{(j)}, \Theta)}{\sum_{i=0}^{1} p_i \prod_{j=1}^{N} f_i^{(j)}(a_t^{(j)}, \Theta)}$$

Table 1. Longitude, latitude, height and frequencies of the installed microwave link.

Parameters	Site 1: Dübendorf	Site 5: Wangen
Longitude	8°37′43.10″ E	8°38′16.26″ E
Latitude	47°24′4.80″ N	47°25′0.25″ N
Height	436 m a.m.s.l.	486 m a.m.s.l.
Freq. (horizontal)	38′657.5 MHz	37′397.5 MHz
Freq. (vertical)	38′650.5 MHz	37′390.5 MHz
Length	1.85 km	

Maximizing $l(\Theta)$ yields, similarly to the univariate case, the maximum likelihood estimate $\hat{\Theta}$. The classification into dry and rainy periods can then be derived from the estimated state probabilities $p_k(a_t, \hat{\Theta})$. Possible extensions to correlated attenuation values, at least during rainy periods, and more general expressions for the joint density $f_1(A_t, \Theta)$ will not be discussed. For simplicity, only the independent case is presented in Sect. 4.

3   Experimental set-up

The experimental site is located in Dübendorf, near the city of Zürich, Switzerland (see Fig. 1). It consists of a 1.85 km long commercial dual-polarization microwave link, 5 disdrometers and 3 rain gauges placed approximatively at equal distances along the path of the link. The dataset is complemented by climatic and meteorological data from two weather stations. The experiment is designed to investigate different aspects of rainfall monitoring using microwave links in the context of a humid continental climate, such as the retrieval of path-averaged rain rates, the influence of the drop size distribution, the characteristics of dry-weather attenuation and wet-antenna effects. In particular, the horizontally and vertically polarized signals could be used to retrieve the effective drop size distribution along the link path.

Microwave link

The installed microwave link is an "Ericsson Mini-link TN ETSI", a widely used system in commercial telecommunication applications. The MWL is operated at about 38 GHz in a dual-polarization set-up with the specific characteristics given in Table 1. For more redundancy, the link provides measurements on 4 different channels (2 polarizations and 2 directions). In its original configuration, the link only records the transmitted and received powers every 15 min. This is clearly not sufficient for accurate rainfall monitoring at scales relevant for modern hydrological and meteorological applications. Therefore, a stand-alone data logging application using the SNMP protocol has been developed and implemented to record the power measurements in much shorter intervals (see Appendix). For the purposes of this project, a 4 s
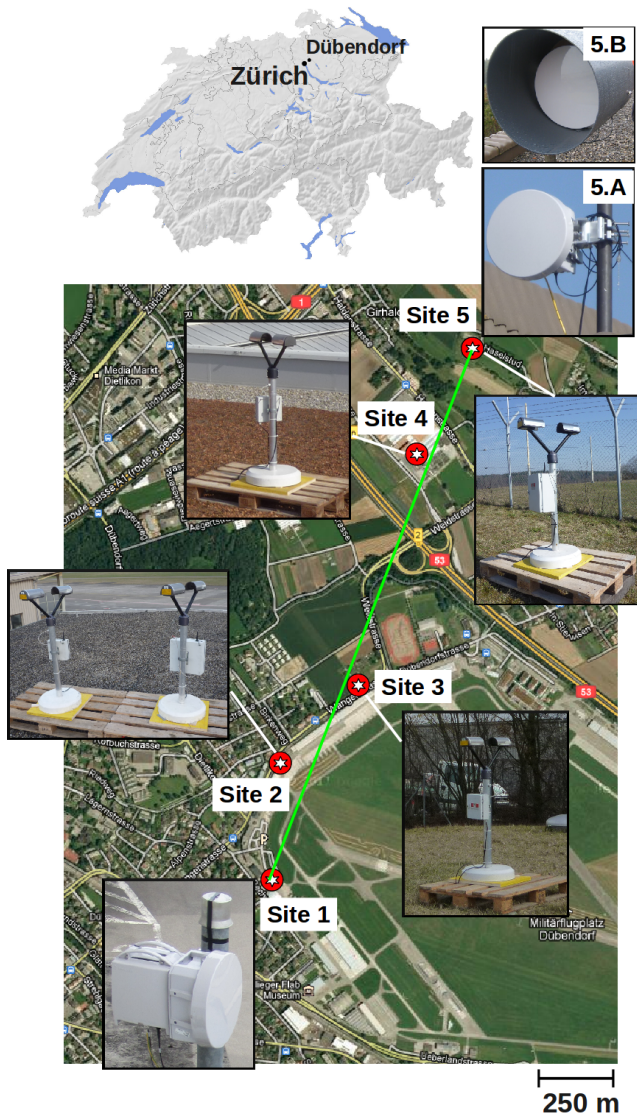
Fig. 1. Experimental set-up deployed in Dübendorf, Switzerland. The Disdrometers are located at sites 2 (2 collocated stations), 3, 4 and 5. The rain gauges at sites 2, 4 and 5. Pictures 5.A and 5.B show the MWL (at site 5) with (respectively without) the rain shields.

temporal resolution has been chosen because it was a good trade-off between a high sampling resolution and a limited amount of missing data (caused by occasional long response times of the radio equipment).

The experiment started in March 2011 and is continuing into the first half of 2012. It is divided into two parts. During the first part of the project, i.e., until 10 October 2011, the antennas of the link were fully exposed to the rain. Consequently, during rainy periods, a thin film of water was formed on the surface of the antennas, causing additional attenuation in the order of several dB (Kharadly and Ross, 2001; Leijnse et al., 2008). Preliminary data analysis suggest that the antennas remain wet for some time after the rain has stopped. An-

tenna drying seems to fundamentally depend on the weather conditions and lasted up to several hours for some cases. In the second part of the experiment, i.e., after the 10 October 2011, the antennas were shielded from rain using plastic shields specifically designed for this experiment (see 5.A in Fig. 1). Visual inspection of the antennas proved that these shields effectively protect the surface of the antennas, even during strong rainfall and moderate wind speeds.

In addition to wet-antenna effects, the experimental set-up also revealed unexpected fluctuations in the transmitted power levels. According to the manufacturer, the received power is measured with an accuracy of 0.1 dB and the transmitted power with an accuracy of 1 dB. Additional measurements of the transmitted power using a power meter showed that the transmitted power was accurate within a range of approximatively 0.35 dB over a period of 11 days, for temperatures between 7 °C and 23 °C and relative humidities between 37 % and 100 %. This is confirmed by independent measurements collected in the laboratory, with (more or less) constant temperatures and humidities and for which the uncertainty in the transmitted power was found to be 0.3 dB.

3.1 Disdrometers and rain gauges

5 optical disdrometers of type Parsivel (1st generation, manufactured by OTT) have been deployed at 4 different sampling locations (sites 2–5) along the 1.85 km path of the link (see Fig. 1). For more details on the principle of these optical disdrometers, see Löffler-Mang and Joss (2000). All the disdrometers are designed to be autonomous in terms of power supply and data transmission (Jaffrain et al., 2011). They provide measurements of particle sizes and velocities at a 30-s temporal resolution. Note that sampling point 2 is equipped with two collocated disdrometers in order to quantify the measurement uncertainty associated with Parsivel disdrometers (Jaffrain and Berne, 2011). The 4 sampling locations have been chosen as a trade-off between a regular distribution of the instruments, the distance to the path of the link, line of sight for data transmission between the different instruments and minimum probability of disturbance and vandalism.

In addition to the 5 disdrometers, 3 tipping-bucket rain gauges from Précis Mécanique (model 3029) have been deployed at sampling locations 2, 4 and 5. The tipping-bucket rain gauges have a catching area of 400 cm^2 and are connected to data loggers that record the tipping time with an accuracy of 0.1 s. One tip corresponds to 0.1 mm of rain. Note that the 3 rain gauges are not transmitting the data in real time. The collected data are used to check the calibration of the disdrometers and to identify possible biases between the sensors.

3.2 Additional data

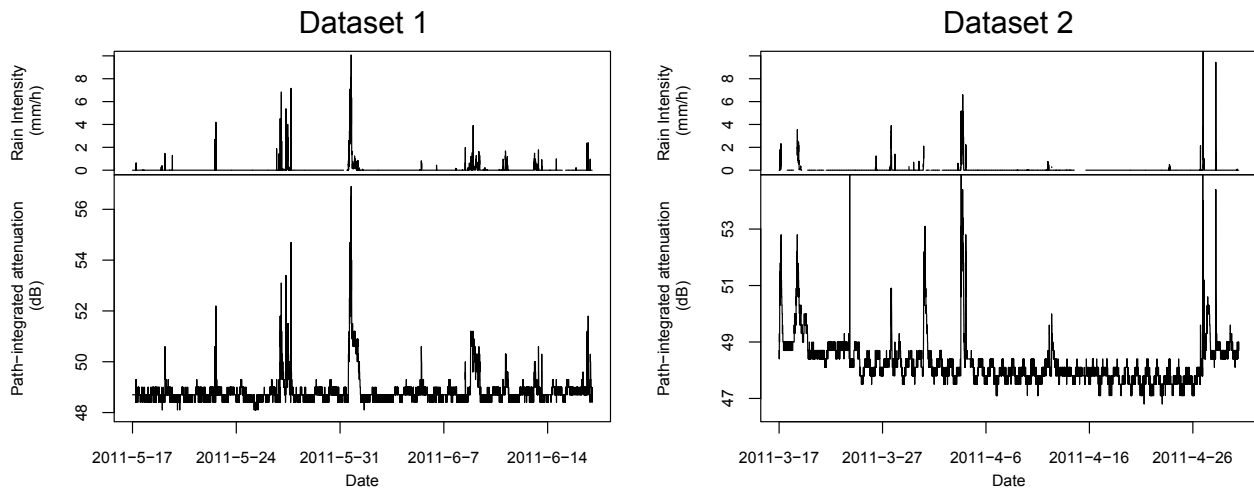The rainfall measurement network is complemented by operational radar data provided by MeteoSwiss. Processed maps

Fig. 2. Rain rate [mm h^{-1}] and path-integrated attenuation [dB] for dataset 1 (stationary) and 2 (non-stationary). For better illustration, only the attenuation of channels 1 (dataset 1), and 4 (dataset 2) are shown. The time is given in UTC.

of rain rate and radar reflectivities are available at a spatial resolution of 1×1 km^2 and a temporal resolution of 5 min. In addition, meteorological and climatic data (e.g., temperature, relative humidity, pressure, wind speed and wind direction) are collected using a MIDAS IV weather station (manufactured by Vaisala) located at the airport in Dübendorf. The MIDAS IV system collects data from two sensors situated at both ends of the runway. The temporal resolution depends on the considered parameter and can vary between 3 and 60 s.

3.3 Originality

Several other studies involving simultaneous measurements of microwave links, rain gauges, disdrometers and weather radar can be found in the literature. Rincon and Lang (2002) proposed a method to estimate the drop size distribution from the measurements of a dedicated 2.3 km, dual-frequency research link and validated their results using 6 rain gauges and a single 2D video disdrometer placed along the path of the link. Rahimi et al. (2003) used a 23.3 km, dual-frequency research link with 22 rain gauges and radar data. However, only 4 or 5 rain gauges were reasonably close to the considered link. More recently, Leijnse et al. (2007c) used a 4.89 km, 27 GHz research link with 6 rain gauges placed along the path of the link. Finally, Zinevich et al. (2010) compared the rain estimates from 23 commercial microwave links with 5 nearby rain gauges. The experimental set-up presented above is original because it combines attenuation measurements from a dual-polarization commercial microwave link with a sufficiently dense network of disdrometers to accurately estimate the path-averaged DSD. This provides a platform to develop and validate new methods for rainfall retrieval using MWL and to evaluate their respective performances as outlined above. In particular, it can be used to investigate if the redundancy between the different channels and polarizations

can be used to improve the rain rate estimates. Furthermore, it might also be of interest to radio engineers concerned with better predictions of rain-induced attenuation and MWL simulation methods (Paulson, 2002; Callaghan et al., 2008). It is intended to make the data publicly available for download from a web-platform after the end of experiment.

3.4 Selected datasets

Two datasets have been selected from the experimental observational record to evaluate the performance of the algorithms described in Sect. 2 under fundamentally different conditions. A visual illustration of these datasets is given in Fig. 2. Note that for a better visibility, the attenuation measurements are only shown for one channel. The first dataset covers the period between 17 May 2011 and 12 June 2011 and is representative of a (more or less) constant dry-weather attenuation baseline (hereinafter referred to as the stationary case). This period is also characterised by small variations in the PIA during dry weather. The second dataset covers the period between 17 March 2011 and 26 April 2011 and illustrates a very different behaviour (hereinafter referred to as the non-stationary case). This period is characterised by a highly-variable attenuation baseline with a strong temporal drift and daily cycles in the PIA, due to changes in temperature and humidity. A preliminary analysis of the current observational record suggests that the non-stationary cases represent a non-negligible amount (about 10–20 %) of all the time periods and must therefore be considered carefully.
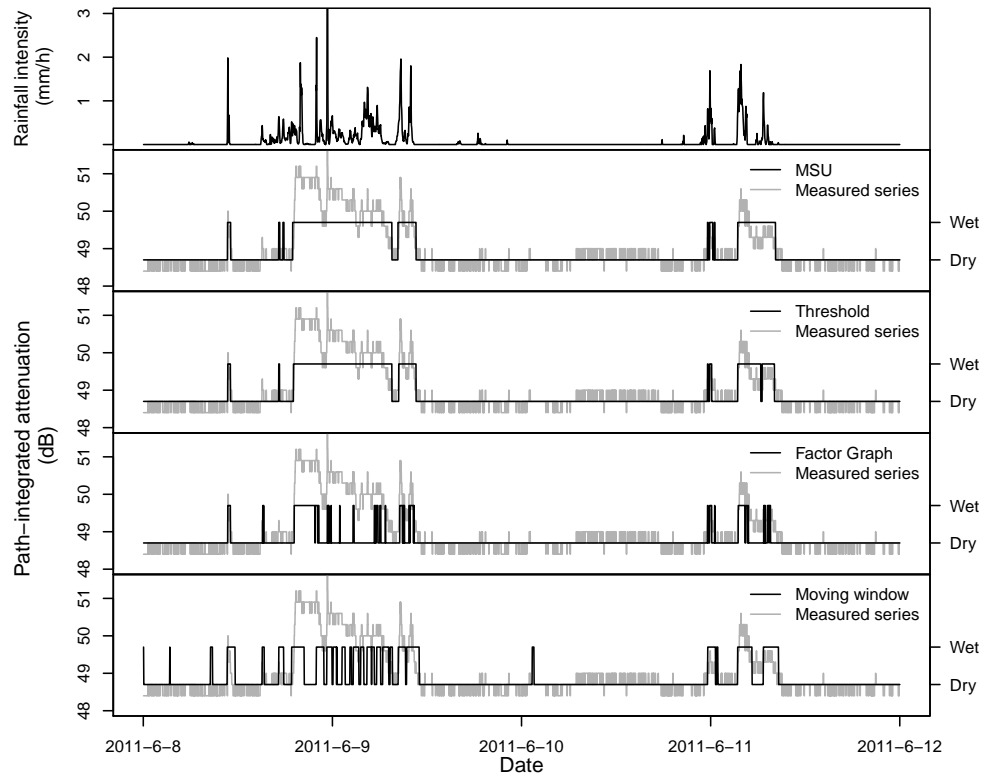
Fig. 3. Illustration of the classification performances for the univariate Markov switching algorithm (MSU), the simple threshold, the Factor Graph and the moving window on a subset of dataset 1 (stationary case). Displayed are the observations from channel 1. The time is given in UTC.

4   Results and model comparison

4.1   False rain and dry detections

The performances of the algorithms described in Sect. 2 are evaluated and compared using two criteria:

$$\text{type I error:} \quad \frac{\text{\#dry periods classified as rainy}}{\text{\#dry periods}}$$

$$\text{type II error:} \quad \frac{\text{\#rainy periods classified as dry}}{\text{\#rainy periods}}$$

In other words, type I errors correspond to false rain detections and type II errors to false dry detections. A perfect classification algorithm has 0 type I error and 0 type II error. In practical applications, however, both types of errors are usually competing against each other, i.e., if the type I error decreases, the type II increases and vice versa. Finding an optimal trade-off between both errors is difficult and depends on the underlying application and the cost associated to each type of error. However, this goes far beyond the scope of this paper and will not be addressed here.

For comparison purposes, it is assumed that the path-averaged rain rate measured by the 5 disdrometers along the path of the link (see Sect. 3) is representative of the "true"

weather state. If the path-averaged rain rate is greater than zero, the period is considered rainy. Otherwise, it is supposed to be dry. In order to analyse the sensitivity of the results with respect to this rain-detection threshold, a slightly higher rain detection threshold of $0.1\,\text{mm}\,\text{h}^{-1}$ is also considered. All periods for which the path-averaged rain rate is smaller than $0.1\,\text{mm}\,\text{h}^{-1}$ are considered dry and vice versa. The value of $0.1\,\text{mm}\,\text{h}^{-1}$ was chosen as a threshold because it approximatively corresponds to the hardware-induced measurement uncertainty of 0.1 dB in the path-integrated attenuation (ITU-R P.838-3, 2005). In other words, rainy periods with rain rates smaller than $0.1\,\text{mm}\,\text{h}^{-1}$ cannot be distinguished from dry periods because of the uncertainty in the power measurements. Finally, note that because the disdrometer data are provided at a 30-s temporal resolution, the corresponding MWL data (at a 4-s temporal resolution) are averaged at 30-s prior to the analysis. Periods for which one of the instruments was not working are not considered for the comparison.

4.2   Stationary dry-weather attenuation baseline

The results for the first dataset (stationary case) are shown in Table 2. To better illustrate important details, a small subset of dataset 1 (a 5-days period between 8 and 12 June 2011) is plotted in Fig. 3. The univariate Markov switching model
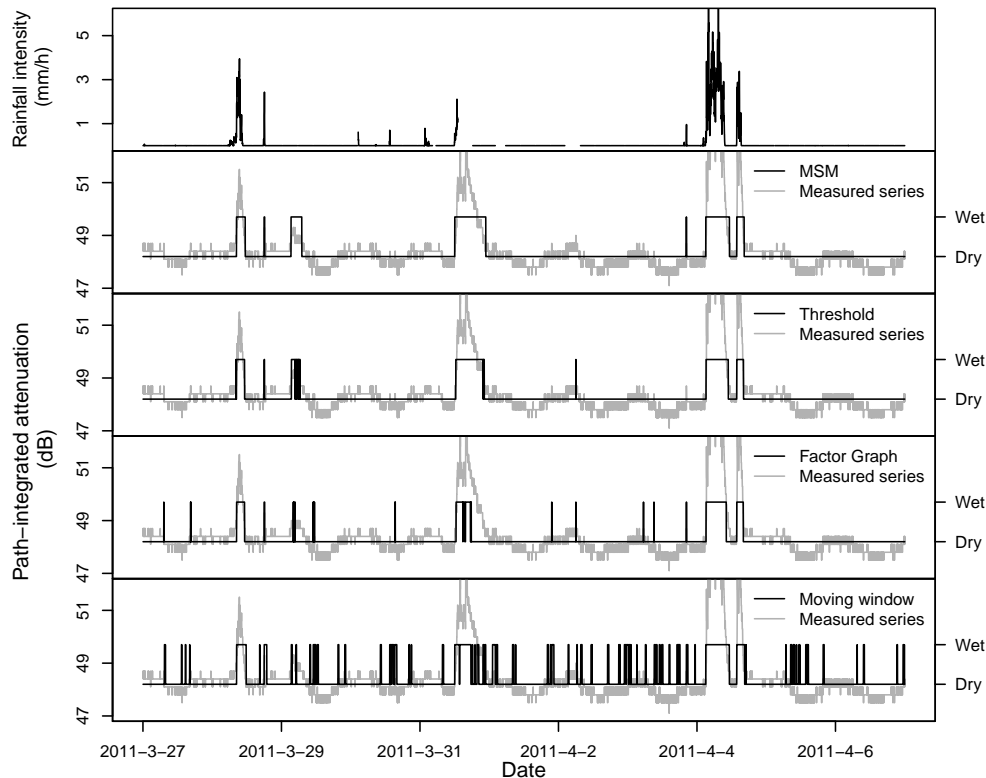
Fig. 4. Illustration of the classification performances of the algorithms on a subset of dataset 2 (non-stationary case). Displayed are the observations from channel 4. Note that the results of the MSU, which did not converge for this channel, have been replaced by the results of the MSM.

Table 2. Classification performances (in percentages) for the simple threshold (ST), the moving window (MW), the Factor Graph (FG), the univariate Markov switching (MSU) and the multivariate Markov switching (MSM) algorithms for dataset 1 (stationary case). For the univariate algorithms, the value given in the table corresponds to the average classification performance for all 4 channels. In parentheses the associated standard deviation. Note that no model parameters could be fitted for the MSU algorithm on channel 2.

Models	rain detection threshold $0\,\mathrm{mm\,h^{-1}}$		rain detection threshold $0.1\,\mathrm{mm\,h^{-1}}$	
	type I error	type II error	type I error	type II error
ST	2.74 (0.88)	23.07 (4.44)	4.52 (0.98)	13.06 (4.26)
MW	12.35 (0.46)	39.04 (2.44)	14.43 (0.14)	11.84 (1.78)
FG	12.00 (2.13)	38.78 (5.75)	12.80 (2.04)	27.07 (6.06)
MSU	2.11 (0.10)	23.32 (2.50)	3.89 (0.18)	13.28 (2.53)
MSM	2.46	21.97	4.27	11.82

(MSU) clearly produced the best classification performances among the univariate models, closely followed by the simple threshold method. The good performance of the simple threshold algorithm is explained by the fact that the dry-weather attenuations over this time period are (more or less) constant with very low fluctuations. The moving window and the Factor Graph, on the other hand, have significantly higher values of type I and type II errors. This can be partially explained by the fact that these models rely on pre-defined threshold parameters which were not necessarily optimal over the considered time period. For example, it is pos-

sible to decrease the type II error rate in the moving window algorithm by increasing the value of σ_0. This will, however, also result in an increased type I error rate. Additional tests with different threshold parameters confirmed that the moving window algorithm produces, on average, less reliable classifications than the threshold and the Markov switching algorithm. Not surprisingly, the multivariate Markov switching model outperformed all the other univariate models in terms of type I and type II errors. Its false rain/dry detection rates are 2.46 % and 21.97 % for a rain detection threshold of $0\,\mathrm{mm\,h^{-1}}$ and 4.27 % and 11.82 % for a rain detection

Table 3. Classification performances for dataset 2 (non-stationary case). Same format than in Table 2. Note that no model parameters could be fitted for the MSU algorithm on channels 2 and 4.

Models	rain detection threshold 0 mm h^{-1}		rain detection threshold 0.1 mm h^{-1}	
	type I error	type II error	type I error	type II error
ST	7.30 (2.74)	48.91 (2.35)	7.51 (2.69)	17.72 (3.57)
MW	11.54 (0.28)	65.74 (0.77)	11.97 (0.34)	17.54 (2.16)
FG	1.80 (0.47)	61.81 (1.40)	1.93 (0.46)	34.78 (2.57)
MSU	1.83 (0.43)	56.10 (2.17)	2.06 (0.47)	27.09 (2.27)
MSM	3.24	50.55	3.53	20.08

threshold of 0.1 mm h^{-1}(not shown). This confirms the intuitive idea that the state of the system can be estimated more accurately using 4 channels rather than 1. The improvement is, however, only minor because the univariate Markov switching model already produced good and similar classifications for all the considered channels (except for channel 2 for which no valid model parameters could be fitted). The fact that the univariate Markov switching model provides realistic classifications can also be seen in Fig. 3, which shows the estimated states (dry/rainy) for all the considered algorithms. A qualitative evaluation suggests that the best classifications are obtained for the threshold method and the univariate Markov switching model (MSU). The classifications obtained using the Factor Graph and the moving window algorithm are not satisfactory. Both the Factor Graph and the moving window produce considerable false dry detections. The moving window algorithm also produces some false rain detections at the beginning of the period. Clearly, the threshold parameters (which were subjectively estimated for the entire dataset) are not optimal for this period.

4.3 Non-stationary dry-weather attenuation baseline

The results for the second dataset are shown in Table 3. As for the first dataset, the classification performances is illustrated in Fig. 4, where the results are plotted for a 11-days subset from 27 March 2011 to 7 April 2011. The first point to notice is that all the considered models have a very high rate of type II errors (about 50–60 % for the first rain detection threshold and 20–35 % for the second rain detection threshold). This is due to the large variability of the attenuation baseline during dry periods, which makes it difficult for the models to separate dry periods from light rainfall. Consequently, more rainy periods are classified as dry. This is also confirmed by the low type I error rates, meaning that very few dry periods are actually classified as rainy. The "best" average performance (among the univariate algorithms) is again obtained for the univariate Markov switching model and the simple threshold method, although these two models do not have the same type I,II error rates. As can be seen in Table 3, the threshold method produces less false dry detections but

more false rain detections. The moving window algorithm has the highest rate of type II errors (65.74 % on average for a rain detection threshold of 0 mm h^{-1}), but most of these false dry detections correspond to very light rain rates. This is indicated by the fact that, for a rain detection threshold of 0.1 mm h^{-1}, which essentially removes very light rainfall, a much lower type II error rate of 17.54 % is obtained. In fact, for the higher rain detection threshold, the moving window algorithm performs similarly to the simple threshold and the univariate Markov switching model. Again, the multivariate Markov switching algorithm outperformed (on average) the univariate algorithms in terms of false dry and rain detections. In particular, it is worth mentioning that no valid model parameters could be fitted for the univariate Markov switching model for channels 2 and 4 whereas the multivariate Markov switching model (using all 4 channels) was still able to provide valid parameter estimates for all channels. The threshold method and the multivariate Markov switching algorithm (MSM) produce very good and similar results for this time period. The classifications obtained using the Factor Graph and the moving window algorithm do not look very good. In particular, the strong variability in the attenuation baseline causes the moving window algorithm to produce a large amount of false rain detections. This problem could be (partially) solved by considering a lower detection threshold σ_0 for this time period, but there is currently no easy way of doing this automatically in the absence of any control data from nearby weather stations.

5 Discussion and possible developments

The Markov switching model proposed in Sect. 2 already provides good results at a reasonable computational cost. It remains, however, very simple in its formulation and does not exploit the full potential of state space models. As a possible extension, the performance of an autoregressive state space model of order 1 was also investigated. Although more elaborate, the autoregressive model of order 1 only showed little improvement in performance compared to the much simpler AR(0) model. Because it is computationally more expensive and more difficult to fit autoregressive state space models,
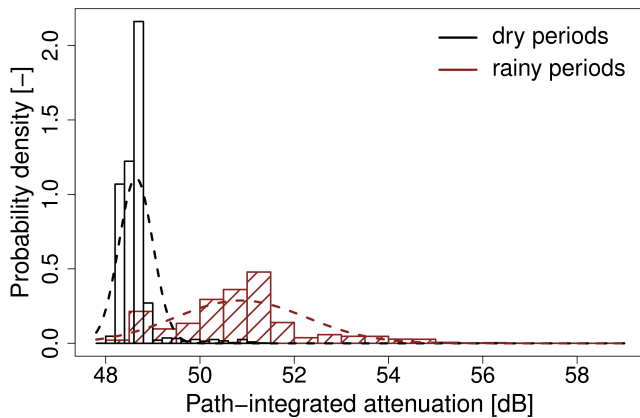
Fig. 5. Empirical probability density functions of attenuation values for dry and rainy periods (for dataset 1). The dashed lines represent the fitted densities of a Gaussian distribution with same mean and variance as the samples. The dry and rainy periods are derived from the disdrometer data.

the AR(0) was preferred for practical applications. Next, the authors investigated how the Gaussian error assumption in Eq. (3) affects the dry/wet classification performance. It is well known that the distribution of rain rate values (and consequently path-integrated attenuation) is skewed and closer to a log-normal distribution than to a Gaussian distribution. An alternative model formulation with non-Gaussian error structure was therefore considered:

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period} \\ \mu_0 + \varepsilon_0 + \varepsilon_1 & \text{for every rainy period} \end{cases} \quad (11)$$

where ε_0 is a Gaussian random variable with zero mean and standard deviation σ_0 and ε_1 is a positive random variable with log-normal distribution representing the rain-induced attenuation. The major drawback of such a formulation is that it has no analytical expression for the conditional density of A_t knowing $S_t = 1$. It is, at the expense of additional computation costs, however still possible to fit this model using numerical approximations. Surprisingly, the more complex and physical error structures did not improve the classification performance significantly. The reason for this can be seen in Fig. 5 which shows the probability density functions of attenuation values for dry and rainy periods. The sample distributions are not exactly Gaussian, but the fact that the tails of the distributions are not correctly reproduced is not critical with respect to the classification problem. In fact, the optimal classification threshold which is at the intersection between the two empirical probability density functions (i.e., about 49 dB) is very close to the threshold derived from the Gaussian model (i.e., the intersection between the two Gaussian densities). Similar results are obtained for all channels and all considered datasets and show that the Gaussian error assumption is not critical with respect to the classification problem.

It is important to note that the Markov switching models suggested here are general switching autoregressive models, which might not perfectly represent the structural patterns observed in each and every MWL dataset. This also holds true for the other models. However, as these models are applied in a classification context, there is usually more concern about overfitting. Overfitting would be problematic where the model sticks too closely to the data and reproduces irrelevant details, which impairs the capabilities to predict future observations. In our case, however, we do not use the ground truth on dry and wet periods from the disdrometers for training and only compare the performance after classification.

Another fundamental problem that needs to be addressed in future studies concerns the problem of the wet antenna effects on the classification into dry and wet periods. Most commercial microwave links do not have shielded antennas. Consequently, they experience some additional attenuation due to a thin water film formed on the surface of the antennas. This effect can be in the order of several dB and must be taken into account when estimating dry and rainy periods, especially during and immediately after a given rain event where the antenna can stay wet for several hours. Future investigations could consider two different states for dry periods, depending on the state of the antenna.

$$S_t = \begin{cases} 0 & \text{for every dry period with dry antenna} \\ 1 & \text{for every dry period with wet antenna} \\ 2 & \text{for every rainy period (with wet antenna)} \end{cases} \quad (12)$$

In this case, a possible attenuation model could be given by

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period with dry antenna} \\ \mu_1 + \varepsilon_1 & \text{for every dry period with wet antenna} \\ \mu_2 + \varepsilon_2 & \text{for every rainy period} \end{cases} \quad (13)$$

It must be noted, however, that such a model might be poorly identifiable, i.e., the parameters and states can not be identified without ambiguity because of the uncertainty affecting the power measurements and because of the strong dependence between the model parameters.

6 Conclusions

In this article, a new algorithm based on a Markov switching model has been introduced to classify attenuation measurements from commercial microwave links into dry and rainy periods. The performance of the algorithm has been evaluated using real data from a new and original experimental setup and compared to 3 existing classification methods. The results show that the Markov switching algorithm performs well and that its classification performance can be increased if multiple channel inputs are considered. Clearly, this is a big advantage compared to other univariate algorithms from the literature which cannot be generalized easily to the multivariate case. The fact that the Markov switching model does

not require any empirically estimated threshold parameters is also of advantage.

The experimental set-up described in Sect. 3 provides a unique platform from which various aspects of rainfall retrieval using MWL can be investigated. For example, it is now possible to rigorously evaluate and compare the classification performances of the different algorithms presented in Sect. 2, which is difficult based on single gauges, which are usually not directly under the MWL beam. The potential applications and scientific value of this experiment go, however, far beyond the simple application presented in this article. Future studies will, for example, investigate the effect of wet antenna bias on retrieved rain rates, explore how attenuation of orthogonal polarizations can be used to retrieve the effective drop size distribution (DSD) along the link path, and the possibility to use multiple channels in order to improve the accuracy of the rain rate estimates.

Appendix A

SNMP Programming on Mini-Link

The microwave link used in this experiment is an "Ericsson Mini-link TN ETSI", a widely used platform in commercial telecommunication applications. However, it is not a dedicated remote sensing device, which required the development of a custom software for high-frequent data logging. As the Mini-link is relatively inexpensive and widely used, details of our solution might be of use to others and facilitate future studies.

The Mini-link can provide some management information through a software called Mini-link Craft. In the case of rainfall estimation, the major parameters of interest are the transmitted and received powers. However, in its initial configuration, these values are only provided at a 15 min temporal resolution. This is clearly not enough considering the temporal and spatial dynamics of rainfall. Consequently, a simple network management protocol (SNMP) has been implemented to query the transmitted and received powers using a much higher temporal resolution.

SNMP is an Internet-standard protocol for managing devices on IP networks. It allows network management systems to monitor the conditions of network devices. Three SNMP versions can be distinguished: the initial implementation (SNMP v1) and its revised versions SNMP v2, SNMP v3 which offer enhanced security for Internet communications. In our case, a Windows/C++ library called SNMP++ (Mellquist, 1997), was used to query the MWL data. The SNMP for the Mini-link uses three key software components, as shown in Fig. A1.

SNMP manager: the client software running on the administrator's computer.

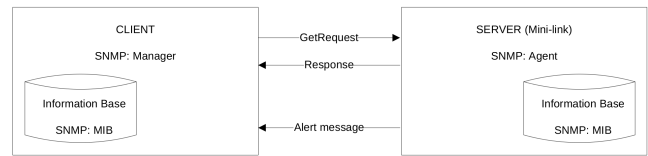*SNMP agent:* the server software running on the Mini-link.

Fig. A1. SNMP client/server concept to communicate with and manage the Mini-Link.
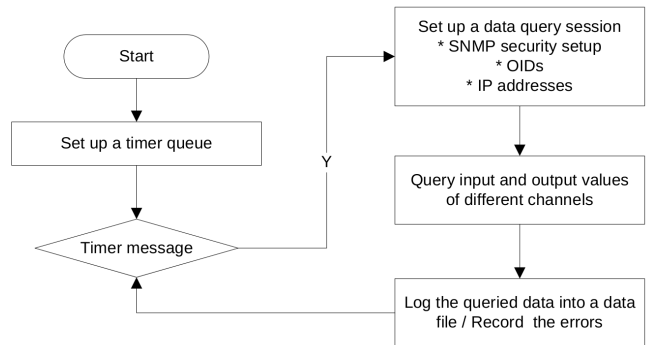

Fig. A2. Flowchart for data acquisition and logging.

Management information base (MIB): the MIB is a virtual database, i.e., a hierarchically arranged collection of information that lists all objects that can be accessed via SNMP reading/writing operations. Each object has a unique object identifier (OID).

Communication is usually initiated by the manager who sends a read/write request to the agent via the command *GetRequest*. The agent reads/writes the desired values from/to the local information base, and sends a response via the *Response* command together with a status information. If the manager and the agents are connected through the Internet, IP addresses need to be assigned for remote communications. The SNMP manager and agents can also be connected locally through a USB connection.

The Ericsson Mini-link provides a large number of object identifiers for system management purposes. An SNMP manager can query the Mini-link OIDs for specific information. For example, the OID "1.3.6.1.4.1.193.81.3.4.3.1.3.1.1.0" denotes the transmitted power and "1.3.6.1.4.1.193.81.3.4.3.1.3.1.10.0" denotes the received power. If there are more than one channel in the Mini-link system, each channel is identified using a number, e.g., "2129920257". The transmitted power for this channel is then identified using the OID "1.3.6.1.4.1.193.81.3.4.3.1.3.1.1.2129920257". Because OIDs can be very long and complicated, the MIB translates each OID into a human readable form. To conveniently retrieve OIDs, Net-SNMP and MG-SOFT proved to be useful tools.

Once the OID for each parameter and channels have been identified, the implementation of the data acquisition and logging is straight forward. A flowchart illustrating the procedure is shown in Fig. A2. For our application, the retrieved data are organised into daily files. Each file contains the date (dd/mm/yy), time (HH:MM:SS) in UTC and the transmitted/received powers (in dB) for all the considered channels. The path-integrated attenuation is then derived by subtracting the received power from the transmitted one. The data acquisition software can be made available on request.

References

Callaghan, S., Boyes, B., Couchman, A., Waight, J., Walden, C., and Ventouras, S.: An investigation of site diversity and comparison with ITU-R recommendations, Radio Sci., 43, RS4010, doi:10.1029/2007RS003793, 2008.

David, N., Alpert, P., and Messer, H.: Technical note: Novel method for water vapour monitoring using wireless communication networks measurements, Tech. Rep. 7, 2009.

Germann, U., Galli, G., Boscacci, M., and Bolliger, M.: Radar precipitation measurement in a mountainous region, Q. J. Roy. Meteor. Soc., 132, 1669–1692, doi:10.1256/qj.05.190, 2006.

Hamilton, J.: A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 457, 357–384, 1989.

Hamilton, J.: Analysis of time series subject to changes in regime, J. Econometrics, 45, 39–70, doi:10.1016/0304-4076(90)90093-9, 1990.

Holt, A. R., Kuznetsov, G. G., and Rahimi, A. R.: Comparison of the use of dual-frequency and single-frequency attenuation for the measurement of path-averaged rainfall along a microwave link, IEE Proc-H, 150, 315–320, 2003.

ITU-R P.838-3: Specific attenuation model for rain for use in prediction methods, International Telecommunication Union Radiocommunication Recommendations, 2005.

Jaffrain, J. and Berne, A.: Experimental quantification of the sampling uncertainty associated with measurements from Parsivel disdrometers, J. Hydrometeor., 12, 352–370, doi:10.1175/2010JHM1244.1, 2011.

Jaffrain, J., Studzinski, A., and Berne, A.: A network of disdrometers to quantify the small-scale variability of the raindrop size distribution, Water Resour. Res., 47, W00H06, doi:10.1029/2010WR009872, 2011.

Kharadly, M. M. Z. and Ross, R.: Effect of wet antenna attenuation on propagation data statistics, IEEE T. Antenn. Propag., 49, 1183–1191, 2001.

Kim, C. J.: Dynamic linear models with Markov-switching, J. Econometrics, 60, 1–22, 1994.

Krämer, S., Verwon, H.-R., and Redder, A.: Improvement of X-band radar rainfall estimates using a microwave link, Atmos. Res., 77, 278–299, 2005.

Kschischang, F. R., Frey, B. J., and Loeliger, H.-A.: Factor graphs and the sum-product algorithm, IEEE Trans. Inf. Theory, 47, 498–519, 2001.

Leijnse, H., Uijlenhoet, R., and Stricker, J. N. M.: Hydrometeorological application of a microwave link: 1. Evaporation, Water Resour. Res., 43, W04416, doi:10.1029/2006WR004988, 2007a.

Leijnse, H., Uijlenhoet, R., and Stricker, J. N. M.: Hydrometeorological application of a microwave link: 2. Precipitation, Water Resour. Res., 43, W04417, doi:10.1029/2006WR004989, 2007b.

Leijnse, H., Uijlenhoet, R., and Stricker, J. N. M.: Rainfall measurement using radio links from cellular communication networks, Water Resour. Res., 43, W03201, doi:10.1029/2006WR005631, 2007c.

Leijnse, H., Uijlenhoet, R., and Stricker, J. N. M.: Microwave link rainfall estimation: effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation, Adv. Water Resour., 31, 1481–1493, 2008.

Loeliger, H. A., Dauwels, J., Ju, J., Korl, S., Ping, L., and Kschischang, F. R.: The factor graph approach to model-based signal processing, Proc. of the IEEE, 95, 1295–1322, 2010.

Löffler-Mang, M. and Joss, J.: An optical disdrometer for measuring size and velocity of hydrometeors, J. Atmos. Oceanic Technol., 17, 130–139, 2000.

Mellquist, P.: SNMP++: An Object-Oriented Approach to Developing Network Management Applications, Prentice Hall PTR, 1997.

Messer, H., Zinevich, A., and Alpert, P.: Environmental monitoring by wireless communication networks, Science, 312, 713–713, doi:10.1126/science.1120034, 2006.

Metzner, P., Horenko, I., and Schuette, C.: Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time, Phys. Rev. E, 76, doi:10.1103/PhysRevE.76.066702, 2007.

Paulson, K. S.: The spatial-temporal statistics of rain rate random fields, Radio Sci., 37, 2002.

Rahimi, A. R., Upton, G. J. G., Holt, A. R., and Cummings, R. J.: Use of dual-frequency microwave links for measuring path-averaged rainfall, J. Geophys. Res., 108, 4467, doi:10.1029/2002JD003202, 2003.

Rayitsfeld, A., Samuels, R., Zinevich, A., Hadar, U., and Alpert, P.: Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system, Atmospheric Research, 104–105, 119–127, doi:10.1016/j.atmosres.2011.08.011, 2011.

Reller, C., Loeliger, H., and Diaz, J.: A model for quasi-periodic signals with application to rain estimation from microwave link gain., in: Proc. 19th European Signal Proc. Conf. (EUSIPCO), Barcelona, Spain, 2011.

Rincon, R. S. and Lang, R. H.: Microwave link dual-wavelength measurements of path-average attenuation for the estimation of drop size distribution and rainfall, IEEE T. Geosci. Remote Sens.,

40, 760–770, 2002.

Ruf, C. S., Aydin, K., Mathur, S., and Bobak, J. P.: 35-GHz dual-polarization propagation link for rain-rate estimation, J. Atmos. Oceanic Technol., 13, 409–425, 1996.

Schleiss, M. and Berne, A.: Identification of dry and rainy periods using telecommunciation microwave links, IEEE Geosci. Remote Sens. Lett., 7, 611–615, doi:10.1109/LGRS.2010.2043052, 2010.

Sevruk: Adjustment of tipping-bucket precipitation gauge measurements, Atmos. Res., 42, 237–246, 1999.

Upton, G. J. G., Holt, A. R., Cummings, R. J., Rahimi, A. R., and Goddard, J. W. F.: Microwave links: the future for urban rainfall measurements?, Atmos. Res., 77, 300–312, 2005.

Weiss, L. L.: Sequences of wet or dry days described by a Markov chain probability model, Mon. Weather Rev., 92, 169–176, 1964.

Yue, D. and Han, Q.: Delay-dependent exponential stability of stochastic systems with time-varying delay, nonlinearity, and Markovian switching, IEEE T. Autom. Control, 50, 217–222, 2005.

Zinevich, A., Alpert, P., and Messer, H.: Estimation of rainfall fields using commercial microwave communication networks of variable density, Adv. Water Resour., 31, 1470–1480, doi:10.1016/j.advwatres.2008.03.003, 2008.

Zinevich, A., Alpert, P., and Messer, H.: Frontal rainfall observation by commercial microwave communication network, J. Appl. Meteor. Climate, 48, 1317–1334, 2009.

Zinevich, A., Messer, H., and Alpert, P.: Prediction of rainfall intensity measurement errors using commercial microwave communication links, Atmos. Meas. Tech., 3, 1385–1402, doi:10.5194/amt-3-1385-2010, 2010.