

Responses to comments posted by Referee #1

We thank the first referee for reviewing our article and providing his fruitful feedbacks. They identify some unclear issues and help to improve the presentation of our research. In the following, we answer to all of the comments one by one. The Referee comments are in blue.

The location of the specific author's changes in the marked-up manuscript version (track changes) are given in green.

Comment 1: The paper, in general, is quite well written and well structured, and for sure will be interesting to the readers of HESS. Measures of model quality based on information theory add an additional facet to the problem of model assessment and model selection (which traditionally are based mainly on estimates of mean (squared) model error assuming deterministic models). The conducted experiments are comprehensive, very well designed and presented. Illustrative material is adequate. Conclusions are based on evidence from experiments. It can be considered for publication, however there are some issues outlined below which require attention before publication can be recommended.

Reply 1: We thank the referee for his generally positive evaluation of our manuscript.

Comment 2: On the references to earlier literature. Immediately after the authors pose the objectives of this work they write: "Comparable approaches have been suggested by Sharma and Mehrotra (2014) and Thiesen et al. (2018)", but in the rest of the paper you do not cite these authors again, and don't mention what are

the differences of your approach and the one taken in the papers mentioned. It is suggested to present briefly the essence of the approaches already published, and to explain the advances made in this paper, and formulate its novelty w.r.t. earlier work.

Reply 2: Agreed. We will do so in a revised version of the manuscript.

Author's change: P 3, L 4

Comment 3: The same can be said about a citation of Yang et al. 2017 or Kirstetter et al. 2015 which is said to be "a similar approach". (It also remains unclear, what this this approach is similar to. Have these researchers also use probabilistic Z-R relationship instead of the deterministic one?)

Reply 3: Agreed. We will do so in a revised version of the manuscript.

Author's change: P 3, L 9

Comment 4: A general comment: it is suggested, if any reference to earlier work is given, to specify what was done in that work, its main conclusion, and what does this mean for this work.

Reply 4: Agreed. See Replies 2 and 3. We will do so in a revised version of the manuscript. We will briefly describe the results and applied methods of the works cited. In addition we will more explicitly show how our manuscript compares and contrasts to the previously published and cited work.

Author's change: P 3

Comment 5: P 2, L 27: the aim of the paper is formulated: “aim of this paper to suggest and apply a framework which comprises expressing relations among data directly by empirical discrete probability distributions (dpd’s), and measuring the strength of relations and remaining uncertainties with measures from Information Theory.” However the paper title says that you want to estimate precipitation (QPE). In my view these are two different objective (albeit related). It is also unclear what is “estimation of precipitation” exactly - is it deterministic, or probabilistic? It is therefore suggested to formulate the objectives clearer, and to relate them to the title, and to the section 1.1. “design of experiments”.

Reply 5: Agreed. In a revised version of the manuscript, we will explain more clearly that with the paper we follow indeed two, albeit related objectives: Present the information-based and probabilistic approach to estimation/prediction in general, and show how it works for QPE in particular. To further comment on this: With our approach we do not want to provide a deterministic, single-valued rainrate, but promote the use of probabilistic QPE, which adequately reflects the (considerable) intrinsic uncertainties related to radar-based QPE. The history of radarmeteorology shows that the most important part is the quality control and the error handling. With our approach we want to provide a probable rain rate value range and distribution, which for users has the added value of knowing intrinsic uncertainties compared to a single-valued QPE value. This is similar to ensemble approaches in operational weather forecasting. By knowing a predictive distribution rather than a single value the user will make

better-informed decisions, especially in the case of extreme events. And if the user requires a single-valued statement: a distribution can always be collapsed to a single value, e.g. by calculating the mean or mode, while this is not possible in the opposite direction.

Author's change: P 3, L 16

Comment 6: In Sec 1 there is a block of text, before 1.1, with a number. Suggest to give it a title and number it as well.

Reply 6: Agreed. We will do so in a revised version of the manuscript.

Author's change: P 2, L 6

Comment 7: P4, L11-13: Eq 1 uses \log_2 but in the text you mention other options. Please coordinate better. L18: definition of information was about an "event", and on L18 you switch to "signal" (indeed this latter is what the most work on information theory traditionally use). I suggest to think of a more consistent terminology for this paper, or explain how event and signal relate to each other.

Reply 7: Agreed. We will move the first sentence after Eq 1 up, to make clear that while several units for information exist, we will stick to [bit]. Also, in a revised version of the manuscript we will introduce both terms 'event' and 'signal' once to explain that both refer to 'the outcome of a random experiment, i.e. drawing randomly from a known distribution', but then will stick to 'signal' throughout the text.

Author's change: P 4, L 31 – P 5, L 4

Comment 8: P5 L5: terms expected information and expected uncertainty — here I would use quotes around the terms. L10: why Entropy starts with is the capital letter, and information not? Suggest to be consistent. L15: While → while

Reply 8: P5L5: Agreed, will do. L10: Agreed, we will use both 'Entropy' and 'Information' with capital letters throughout the text. L15: Agreed, will do.

Author's change: Quotes: P 5 L 27 - 'Information' with capital letter changed throughout the text (e.g. P 5, L 11, 12, 13, 14, 17, 19, 27, 31) – While -> while (P6 L 6)

Comment 9: L27: please introduce what is set Y. Are you guessing X, or you are guessing a realisation of X, i.e. x_i ?

Reply 9: As explained in the sentence starting in L28, Y refers to available (a priori known) data used to guess an unknown target quantity of interest, X. To do so, both the general relation between X and Y must be known from a joint learning data set $\{X,Y\}$ with paired data x_1,y_1 x_2,y_2 etc. ; and a particular observation of Y (i.e. y) to guess the particular, related realisation of X (i.e. x). We agree that subscript 'i' here may lead to confusion and will distinguish in a revised version of the manuscript the set from a particular value coming from that set by capital and small letters only, e.g. X, x. Also, we will illustrate these things in more detail with an example using hydrometeorological data (i.e. we will elaborate the sentence starting in L28).

Author's change: P 6, L 18 – P 7, L 2

Comment 10: L32: you write $Y=y_i$, but in Eq 3 it us just y

Reply 10: Please also see our reply to Comment 9: We agree that subscript 'i' here may lead to confusion and will distinguish in a revised version of the manuscript the set from a particular value coming from that set by capital and small letters only, e.g. X, x .

Author's change: P 6, L 24 – P 7, L 4

Comment 11: Overall comment on section 2.1: this is a brief introduction to information and entropy, and it uses terminology common for I.T. textbooks. It will be clear to those who know it, but I am not sure all is clear to those who have not used information theory before. It is also somewhat different form the terminology used in the rest of the paper. What is "event"? What is a set $X=\{x_1...x_n\}$? Is X a sample from the (same?) distribution? Is its pdf known? Is X a time series as well? An example from hydrometeorology would have helped a lot - assuming, if I understand correctly, that X is some random variable related to rainfall (is its distribution assumed to be known?), and x_i are its realisations (is this right?).

Reply 11: The referee's interpretation of X, x_i and the general procedure is correct. We agree that we need to better explain these things in the text and will do so according to our suggestions in Reply's 9 and 10.

Comment 12: P6 L26: “true distribution and a model thereof” - it is first time these terms are used, and a problem of building a model of the true distribution is mentioned. If indeed it is the problem to be solved in this paper, I would suggest to present it earlier.

Reply 12: Good point. To express more clearly what we mean by 'true distribution' and 'a model thereof', we will introduce these terms in the previous paragraph about Cross Entropy (after the sentence starting in L13) and add an illustrative example.

Author's change: P 7, L 13 - 29

Comment 13: P7 L3: 2.2. is Methods, but 2.1 was also methods, right? Of Information theory does not belong to Methods? Subsection is entitled: “Data-based models and predictions, information-based model evaluation”, so it is also about measures (since “evaluation” is based on measures) - but 2.1 had also "measures for information theory”. Suggestion: to coordinate the titles of various sections to prevent overlaps.

Reply 13: Good point. We will retitle the subsection, to prevent apparent overlaps. Our new title for subsection 2.2 (P7 L3) will be “Modelling and evaluation strategy”).

Author's change: P 8, L 5

Comment 14: Reading this subsection, I was expecting to find the “data-based models”, but could not... It is suggested to clarify what is meant by these and to present them, or not to use this term. (When seeing the term “data-based, or “data-driven” models, I would expect to see a model build on data, using statistical or machine learning techniques and able to make a prediction of a variable (predictant) based on several other variables (predictors).)

Reply 14: Thanks for the comment, which indicates we need to explain better what we mean by 'data-based model' in this paper: The model is simply the empirical, multivariate, discrete frequency (or probability when normalized) distribution derived from all available data. Its input is values for all predictors, and it returns a conditional (on the values of the predictors) distribution of the predictant. We will explain this better in section 2.2 of a revised version of the manuscript.

Author's change: P 9, L 3 -17

Comment 15: L17: “we lose the information about the absolute and relative position of the data tuples in the data set” - unclear. (Do you mean we lose the time stamp of each data tuple (since this are time series)? If so, this is of course always true when a time series is represented as a pdf.)

Reply 15: Yes, it means we lose the information contained in the time stamp (i.e. absolute position and relative order of the data), and this not specific to our

approach but always true if a time series is mapped to a pdf. We will explain this in more detail in a revised version of the manuscript.

Author's change: P 8, L 19 - 20

Comment 16: P8 L1: "statement about the target value" - I don't see a statement about a "value" (value meaning a real valued estimate of the target), just pdf.

What is presented on (a very useful and informative) Fig 1 is estimation of predictive uncertainty conditional on the model output (being the range [-2, -2.3]).

This uncertainty is estimated based on ALL data across the whole time domain.

(Please see e.g. papers by Todini on predictive uncertainty, and I would suggest to add at least one of them.) This is not a model that predicts the target value *for a given time moment*. (But such model what would nice to have of course.)

Reply 16: Indeed the 'value' statement in our case does not contain a single value, but a pdf, which to us is a joint statement about the predicted target value AND at the same time its predictive uncertainty. If for any reason the receiver of such a prediction cannot use the pdf directly, then its two components (value and uncertainty) can easily be extracted by e.g. calculating its mean (or mode) and the variance (or entropy). Please see also our related Reply 5. The referee is right in stating the uncertainty is based on ALL data across the whole time domain, or, more specific, based on all data with indistinguishable (in the light of the chosen binning) predictor values (in Fig. 1 predictor values in range [-2,-2.3]). So from a predictor value point of view, these predictive situations are all the same, and should hence also lead to the same values of the predictant. In reality they don't, because the predictors don not represent all factors influencing the predictant's

value, so what we get is a predictive distribution rather than a single value. Thank you for pointing us at the related work of Ezio Todini, we will do a literature review and will cite related literature in a revised version of the manuscript.

Author's change: P 9, L 9 - 14

Comment 17: P17 L25: "to build better models by simply adding more predictors, which according to the Information Inequality (Eq. 4) never hurts." – well, in theory... If we assume that we are building a predictive model (e.g. predicting R on the basis of radar data), in practice more predictors typically means more complex models (more parameters to calibrate/train), and this could be a problem due to the following. (1) You may break the balance between the amount of available data and number of parameters to train, and (2) more model parameters means increasing the dimension of the search space, and it could mean that during training there is a higher chance to be stuck in a local minimum (e.g. if MLP neural networks are used). So there are good practical reasons to avoid having too many predictors, and (3) more complex model may overfit and not be accurate on cross-validation sets.

Reply 17: We completely agree with the referee's reasons against using more and more predictors and that overfitting should be avoided. This holds for parametric models (models with a predefined structure plus parameters estimated with a training data set) as described by the referee, but also for the 'conditional pdf' approach we use. While in our case there are no parameters to train, the choices we need to make are the type and number of predictors and the related binning.

In Figs 3 and 4, we show the relation between number of predictors and required data to adequately 'train' (in our case it's rather 'construct') the model, which is essentially a great number of split sample tests for different sizes of the learning data set. We suggest including in a revised version of the manuscript a more extended discussion about the potential problems of adding more and more predictors along the lines of the referee comment and our reply.

Author's change: P 20, L 23 - 25

Comment 18: P18 L1: I am not sure I understand how overfitting relates to curse of dimensionality by Bellman.

Reply 18: We agree, this is misleading. For brevity and clarity, we suggest to restrict the discussion to 'overfitting' here.

Author's change: P 20, L 23

Comment 19: To summarize the most important comments (and there are more above):

1. The title promises "Quantitative precipitation estimation using a data-based approach". Perhaps I missed something, but I have not found a data-based model that makes a quantitative estimation of rainfall using radar data. Yes, analysis of distributions of radar data help to provide informative uncertainty estimates of precipitation (and assess the overall information gain) but in my view it is not what is typically understood by QPE (deterministic, and possibly with uncertainty estimates on top).

2. Suggested to present clearly innovation w.r.t. to earlier work.
3. Introduction: suggested to formulate the objectives clearer, and to relate them to the six experiments.
4. On the ways to measure the quality of predictive models. What does real practice need? In essence, the probability distribution-based measures (including the ones using the information theory) give an estimate how far are distributions from each other, rather than measuring the distance between time series, i.e. how far are the deterministic model predictions from (ground truth) observations (on average, or for particular important time moments). In my view it is important to stress this difference. So decision on what measures to use depends on the needs of practitioners and the corresponding goal setting. I would suggest that the best is use both since they measure different things.
5. As far as I understand, the authors used All data to estimate the information gain (and hence to “build” a model allowing for QPE). In other words, there is no data left for cross validation and for testing. If so, this does not correspond to the theory of data-based model building. Is this right?

Reply 19

1. We hope we could suitably answer this point in replies 5 and 16.
2. We hope we could suitably answer this point in replies 2, 3 and 4
3. We hope we could suitably answer this point in reply 5.
4. We agree that quality measures should be selected with the intended use of the model in mind, and established quality measures expressing distances in the units of the data (e.g. Nash-Sutcliffe efficiency, bias, RMSE etc.) are clearly valuable. Problems arise in the case of multivariate data and models: How to combine

distance measures expressed in different units? Possible solutions involve normalization of each distance measure to $[0,1]$ scales or choice of a weighted combination scheme. Using quality measures in the probability domain however, such as the information measures we apply, has the advantage of being independent of the units of the data, which facilitates comparison and combination. Also, both Cross Entropy and Kullback-Leibler divergence as presented in the text can be used as distance measures between a single-valued observation/ground truth and a probabilistic prediction. In this particular case, one of the two distributions (the observation) is simply a distribution with $p=1$ for a single bin, and $p=0$ for all others. Kullback-Leibler divergence then becomes a measure of likelihood of the observation given the model.

5. We indeed used all data to construct the 'perfect models'. However, in order to evaluate the effect of overfitting, we used a very large number of modified split-sampling tests as explained in section 2.2 (sampling strategy), 3.2 and displayed in Figs 4 and 5. An example: Fig 4, green line, sample size 2000: The cross entropy is 1.75. This value was obtained as follows: From all available data, 2000 value pairs of RR0 and dBZ1500Rad were randomly chosen and used to construct the multivariate, discrete pdf's which manifest our model. This model was then applied to ALL data (including the 2000 for training) and Cross Entropy was calculated to the 'perfect model' based on all data. The entire procedure was repeated 500 times. The value in the Figure shows the mean of all 500 Cross Entropies. This is not a classical split sample test where calibration and validation data set are completely separate. However, the advantage we see in our approach is that we can express total uncertainty as an additive combination of two effects: Uncertainty due to the limited information content of the predictor about the predictant (which is Cross Entropy using all data, i.e. Cross Entropy at the right end

of the plot), and uncertainty because we use a model constructed from only a limited data set. We suggest that while this is not a standard way of model evaluation, it offers useful ways to look at magnitude and sources of uncertainty. In this context, please also see our replies 16 and 17.

Editorial comments

Comment 20: Minor editing of English may be desirable, e.g.:

- Besides rain gauges with its own limitations: its → their

Reply: Agreed. We will change this in a revised version of the manuscript

Author's change: P 2 , L 8

- Radar data have among other been used for urban hydrology → Among other data sources, radar data have been used in urban hydrology

Reply: Agreed. We will change this in a revised version of the manuscript

Author's change: P 2, L 10

- its use relies on some sometimes more, sometimes less justified assumptions → its use relies on some assumptions, which are sometimes justified and sometimes not,

Reply: Agreed. We will change this in a revised version of the manuscript

Author's change: P 2, L 17

- Much work has since then been done → Much work has since been done

Reply: Agreed. We will change this in a revised version of the manuscript

Author's change: P 2, L 26

- framework which comprises expressing relations among data directly by →
framework which would use relationships between data expressed as the

Reply: Agreed. We will change this in a revised version of the manuscript

Author's change: P 3, L 1

- I suggest shorten some long sentences, and to split long paragraphs (e.g. the first one in the Introduction).

Reply: Agreed. We will change this in a revised version of the manuscript

We thank the referee for the editing suggestions. Besides this, there will be a final copy editing by HESS before the paper is published.

Malte Neuper and Uwe Ehret

Responses to comments posted by Referee #2

We thank the second referee for reviewing our article and providing his fruitful feedbacks. In the following, we answer to all of the comments one by one. The Referee comments are in blue.

Comment 1: Page 2, line 24: It seems to me that the work of Vignal et al (1999) and (2000) only addresses the variability of Z from high altitude to the ground but not directly the variability of the Z-R relationship along the vertical.

Reply 1: We agree. To make things clearer, we will better explain the content of Vignal et al. (1999) and (2000) and will add another reference which more closely addresses the variability of the Z-R relationship to the ground: Peters, G., Münster, H., Clemens, M., and Wagner, A., (2005): Profiles of raindrop size distributions as retrieved by microrain radars. J. Appl. Meteor., 44, 1930–1949, <https://doi.org/10.1175/JAM2316.1>.

Author's change: P 2, L 27 - 29

Comment 2: Page 12, line 11: the reflectivity factors of the MRR, due to the short wavelength used, can also sometimes be greatly attenuated, with consequences on the retrieved DSD.

Reply 2: Good point. It is true that one has to clarify this important possible influence on the results. We will include a short discussion of this potentially significant source of uncertainty in a revised version of the manuscript.

Author's change: P 14, L 13

Comment 3: Page 12, line 11: why is the 3.5 moment used and not (the most commonly used) 3.67? Could you add a reference?

Reply 3: Thanks for indicating this point. Actually we calculated the rain rate by using the 3.67th moment. In a first version however we used the 3.5th moment of the drop size distribution for the rain rate by using the approximation for the fall velocity $v(D) \approx c \cdot D^{0.5}$, which we took from the script of a radar meteorology class. We missed correcting the text accordingly, but will do so in a revised version.

Author's change: P 14, L 10

Comment 4: Page 17, Line 11: A 15.3% reduction in uncertainty when RR0 is conditioned by dBZ1500Rad seems quite low, doesn't it? This proves that radar reflectivity is one of the most important sources of information but that there is a high variability in the Z-R relationship that limits entropy reduction. This explanation is suggested later (experiment 3) but could perhaps already be mentioned here. I think we are also suffering here from the effects of PVR and perhaps even attenuation for high intensities.

Reply 4: We agree. It makes sense to mention the explanation - suggested later in experiment 3 - already in this section. Also, as suggested, we will in a revised version of the manuscript point out that while the radar reflectivity factor is one

of the most important sources of information, there remains a high variability in the Z-R relationship that limits entropy reduction.

Author's change: P 20, L 4

Comment 5: Page 24, Line 3-5: I understand that $DKL=3.43$ bits and 5.04 correspond to the application of the Z-R Marshall Palmer relationship and that $DKL=2.69$ and 4.30 correspond to the optimized Z-R relationship. But I don't understand what differentiates 3.43 from 2.69 (and 5.04 from 4.30): what are the differences in terms of condition applied?

Reply 5: It is correct that $DKL = 3.43$ and Cross Entropy (total uncertainty) = 5.04 bit result from the application of the (deterministic) Marshall-Palmer relationship (red line in Fig. 6), and $DKL = 2.69$ and Cross Entropy = 4.30 bit result from applying a deterministic Z-R relationship fitted to the available data (black line in Fig. 6). So the difference between 3.43 and 2.69 (and 5.04 and 4.30) comes from using a general Z-R relation (derived from the Z-R data pairs Marshall and Palmer had available) in the first and a fitted-to-the-available data Z-R relation in the second case. So while the Marshall-Palmer relation may be better overall (if used over a large variety of data), the custom-made Z-R relation quite expectedly works better when applied directly to the data it was built from. To us, it was rather surprising how good the Marshall-Palmer relation performed compared to the custom-made relation: It really seems to be a good default.

Comment 6: Page 26, Line 22: after "attributed to the ambiguous relationship between radar reflectivity and rain rate", I think that "due to the natural variability of Drop Size Distribution" may be added.

Reply 6: This is definitely correct. We will add this statement.

Author's change: P 28, L 23

Comment 7: Section 3.6: Why kriging (often used method) was not considered for the raingauge approach?

Reply 7: We agree that it would be interesting to compare our results to established interpolation methods such as kriging. However, performance comparison of interpolation methods was not our main focus in this section. Rather, the goal was to show that our databased approach easily allows joint use of several information sources without many additional assumptions, and secondly to demonstrate that information content is a function of distance between stations. In order to keep the study short, we suggest keeping this section as it is. As a side note: We are currently exploring the potential of data- and information based spatial interpolation in depth, comparing it to a range of established interpolation methods. We hope to submit a related manuscript by mid-2019.

Comment 8: The publication of Cecinate et al (2017) cited on page 2 does not appear in the list of references.

Reply 8: This is right. Thank you for detecting this missing reference. We will add it in a revised version of the manuscript.

Author's change: P 2, L 11 (it was a wrong spelling of the name)

Comment 9: Section 1.1, page 3, is a little "orphaned". It may not be necessary to distinguish this paragraph from the introduction. Its content could appear just before the announcement of the plan (between lines 10 and 11 on page 3).

Reply 9: Another Referee also commented on this section, to which we already replied (Reply 6 to referee #1). In a revised version of the manuscript we will give the block of text before 1.1 a number and subtitle title as well, such maintain the same level of text hierarchy throughout.

Author's change: P 2 , L 6

Comment 10: Table 1: I propose to replace "[min, max]" by "[center of the min bin, center ofmax bin]" or something like that.

Reply 10: We agree, that this would be a more robust value range and not so vulnerable to error prone outliers. On the other hand, we already cut down the outliers and fulfilled a thorough data quality analysis, so that we think, that the total value range is already quite robust and directly shows the total range of possible values.

Comment 11: Figures 4, 5, 6 and 7: I think that the texts of the legends should be enlarged.

Reply 11: We agree that the legends are relatively small. We will ask the HESS typesetting team whether they recommend larger legends for the final layout of the paper and will do so if recommended.

Comment 12: Figure 6: Wouldn't the figure in a log-log frame be more readable and give less weight to very weak reflectivities?

Reply 12: We agree that a log-log display would put more emphasis on the higher reflectivities and rain rates, which are commonly of higher interest than the low ones. However, we would like to keep the figure in its current form as the non-transformed data better reveal the nonlinear power-law shape of the Z-R relation as mentioned in the figure caption and in the text (page 23 lines 10 pp).

Comment 13: Table 3: I propose to replace "RR0" by "RR0Dis" to highlight that disdrometer is used (as for dBZ0Dis).

Reply 13: RR0 does not just contain rain rates measured by the disdrometers, but also by standard rain gauges at Useldange, Roodt and Reichlange (please see the paragraphs 'Distrometer data' and 'Rain gauge data' in section 2.3, and column 'description' in table 1). For this reason, we think the general term 'RR0' is more appropriate than 'RR0Dis'.

Yours sincerely,

Malte Neuper and Uwe Ehret

Quantitative precipitation estimation with weather radar using a data- and information-based approach

Malte Neuper¹, Uwe Ehret¹

¹Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology - KIT, Karlsruhe, Germany

5 *Correspondence to:* Uwe Ehret (uwe.ehret@kit.edu)

Abstract. In this study we propose and demonstrate a data-driven approach in an information-theoretic framework to estimate precipitation quantitatively. In this context predictive relations are expressed by empirical discrete probability distributions directly derived from data instead of fitting and applying deterministic functions as it is standard operational practice. Applying a probabilistic relation has the benefit to provide joint statements about rain rate and the related estimation uncertainty. The information-theoretic framework furthermore allows integration of any kind of data considered useful and explicitly considers the uncertain nature of quantitative precipitation estimation (QPE). With this framework we investigate the information gains and losses associated with various data and practices typically applied in QPE. To this end we conduct six experiments using four years of data from six laser optical disdrometers, two micro rain radars MRR, regular rain gauges, weather radar reflectivity and other operationally available meteorological data from existing stations. Each experiment addresses a typical question related to QPE: *First*, we measure the information about ground rainfall contained in various operationally available predictors. Here weather radar proves to be the single most important source of information, which can be further improved when distinguishing radar reflectivity – ground rainfall relationships (Z-R relations) by season and prevailing synoptic circulation pattern. *Second*, we investigate the effect of data sample size on QPE uncertainty using different data based predictive models. It shows that the combination of reflectivity and month of the year as a double predictor model is the best trade-off between robustness of the model and information gain. *Third*, we investigate the information content in spatial position by learning and applying site-specific Z-R relations. The related information gains are only moderate and especially lower than when distinguishing Z-R relations according to time of the year or synoptic circulation pattern. *Fourth*, we measure the information loss when fitting and using a deterministic Z-R relation, as it is standard practice in operational radar based QPE applying e.g. the standard Marshall-Palmer relation, instead of using the empirical relation derived directly from the data. It shows that while the deterministic function captures the overall shape of the empirical relation quite well, it introduces additional 60% of uncertainty when estimating rain rate. *Fifth*, we investigate how much information is gained along the radar observation path, starting with reflectivity measured by radar at height, continuing with the reflectivity measured by a MRR along a vertical profile in the atmosphere and ending with the reflectivity observed by a disdrometer directly at the ground. The results reveal that considerable additional information is gained by using observations from lower elevations by avoiding information losses caused by ongoing microphysical precipitation processes from cloud height to ground. This underlines both the importance of vertical corrections for accurate

10
15
20
25
30

QPE and of the required MRR observations. In the *sixth* experiment we evaluate the information content of radar data only, rain gauge data only and a combination of both as a function of the distance between the target and predictor rain gauge. The results show that station-only QPE outperforms radar-only QPE up to a distance of 7 to 8 km and that radar-and-gauge QPE performs best, even compared to radar-based models applying season or circulation pattern.

5 1 Introduction

1.1 Approaches to Quantitative Precipitation Estimation (QPE)

Quantitative precipitation estimation in high temporal and spatial resolution and high quality are important prerequisites for many hydro-meteorological design and management purposes. Besides rain gauges with ~~their~~s own limitations (Huff, 1970, Nešpor and Sevruc 1999, Nystuen 1999, Yang et al. 1999), weather radar plays an increasingly important role for QPE:

~~Among other data sources, R~~ radar data have ~~among other~~ been used for urban hydrology (Thorndal et al. 2017, ~~Cecinati et al. 2017~~Cecinati et al. 2017a, Wang et al. 2015), hydrological analysis and modelling (Bronstert et al. 2017, Rossa et al. 2005), real-time QPE (Germann et al. 2006), rainfall climatology (Overeem et al. 2009), rainfall pattern analysis (Kronenberg et al. 2012, Ruiz-Villanueva et al. 2012) and rainfall frequency analysis (Goudenhoofd et al. 2017). For a comprehensive overview on radar theory and applications, see Battan (1959a,b), Sauvageot (1992), Doviak and Zrnic, (1993), Rinehart (1991), Fabry (2015) or Rauber and Nesbitt (2018).

While the advantage of weather radar is that it provides 3-dimensional observations in high spatial and temporal resolution and large coverage, unfortunately its use relies on some ~~assumptions, which are sometimes justified and sometimes not. It is further -sometimes more, sometimes less justified assumptions and is~~ hampered by considerable error and uncertainty arising from measuring the radar reflectivity factor Z (herein after referred to as reflectivity) instead of rain rate R, measuring at height instead of at the ground, and many other factors such as ground clutter, beam blockage, attenuation, second-trip echoes, anomalous beam propagation and bright band effects—. For a good overview on sources of errors see Zawadzki,(1984) or Villarini and Krajewski (2010).

In this paper we will focus on the first two aspects, namely Z-R relation and vertical profile of reflectivity. Typically, the Z-R relation is expressed by a deterministic exponential function of the form $Z = a \cdot R^b$ fitted to simultaneous observations of Z at height and R at the ground, the most common being the Marshall-Palmer relation (Marshall and Palmer 1948). Much work has since ~~then~~ been done to acknowledge the strongly non-linear and time-variant nature of this relation (Lee and Zawadzki 2005, Cao et al. 2010, Adirosi et al. 2016), ~~the variability of the reflectivity from high altitude to the ground (Joss and Waldvogel, 1990) and correcting f~~or the vertical profile of reflectivity (Vignal et al. 1999, Vignal et al. 2000). ~~The transition of the Z-R relations -and its transition~~ from at height to the ground ~~(was investigated by Peters et al. (2005) Vignal et al. 1999, Vignal et al. 2000)~~. Much effort has also been spent on ways to improve QPE by combining radar with other sources of information such as rain gauges (Goudenhoofd and Delobbe 2009, Wang et al. 2015) or numerical weather prediction (Bauer et al. 2015) and to quantify uncertainty of radar-based QPE (Cecinati et al. 2017b).

In this context, it is the aim of this paper to suggest and apply a framework which would use relationships between data expressed as comprises expressing relations among data directly by empirical discrete probability distributions (dpd's), and would measure measuring the strength of relations and remaining uncertainties with measures from Information Theory. Comparable approaches have been suggested by Sharma and Mehrotra (2014) and Thiesen et al. (2018): The former use an information-theoretic approach to formulate prediction models for cases where physical relationships are only weakly known but observational records are abundant; the latter emulate expert-based classification of rainfall-runoff events in hydrological timeseries by constructing dpd's from large sets of training data-

In particular, we investigate the effect of applying a purely data-based, probabilistic Z-R relation instead of a deterministic function fitted to the data. A similar probabilistic QPE approach was conducted for example by Kirstetter et al. (2015), by computing probability distributions of precipitation rates modelled from the conditional distribution of precipitation rate for a given precipitation type and radar reflectivity factors on the basis of a one-year long data sample in the United States. But in contrast to their approach with a simple theoretical model we use an information-theoretic framework. Similar work was also done by This holds true the work of Yang et al. (2017), who developed a new relationship converting the vertical profile of reflectivity (VPR) to rain rates by using a terrain-based weighted random forest method. (see Yang et al. 2017 or Kirstetter et al. 2015 for a similar approach). The potential advantage of applying a probabilistic relation is that it yields joint statements of both the value of R and the related estimation uncertainty. That is with our approach we do not want to provide a deterministic, single-valued rainrate, but promote the use of probabilistic QPE, which adequately reflects the (considerable) intrinsic uncertainties related to radar-based QPE. The history of radar meteorology shows that the most important part is the quality control and the error handling. With our approach we want to provide a probable rain rate value range and distribution, which for users has the added value of knowing intrinsic uncertainties (included in the systematic and random parts of the error (Kirstetter et al., 2015)) compared to a single-valued QPE value. This is similar to ensemble approaches in operational weather forecasting. By knowing a predictive distribution rather than a single value the user will make better-informed decisions, especially in the case of extreme events such as flash flood forecasting. And if the user requires a single-valued statement: a distribution can always be collapsed to a single value, e.g. by calculating the mean or mode, while this is not possible in the opposite direction. We further test the potential of various operationally available observables such as synoptic Circulation Patterns (CP), convective and other meteorological indices (retrieved from rawinsonde data), ground meteorological ground variables and season indicators to distinguish typical Z-R relationships. The idea is to improve QPE by applying Z-R relationships tailored to the prevailing hydro-meteorological situation as expressed by the predictors. Lastly, we use a comprehensive data set from four years of 1-h data available from one C-band weather radar, two vertical micro rain radars (MRR), six laser beam disdrometers and six rain gauges set up in the 288 km² catchment of the Attert, Luxembourg (Fig. 2) to evaluate the information gains about ground rainfall when moving from measuring at height to measuring closer to the ground.

The remainder of the paper is structured as follows: In the next section, we briefly present the experiments carried out in the paper. In section 2, we give a short overview on concepts and measures from Information Theory (section 2.1) and on the

methods (section 2.2) and data (section 2.3) used in the study. In section 3, we present and discuss the results of all experiments; our conclusions are presented in section 4.

1.21 Design of experiments

We conduct altogether six experiments: In *Experiment 1*, we investigate the information about ground rainfall contained in various predictors, such as weather radar observations alone or in combination with additional, operationally available hydro-meteorological predictors. In *Experiment 2*, we investigate the effect of limited data on the uncertainty of ground rainfall estimation for various data-based models. In *Experiment 3*, we examine to which degree the empirical Z-R relationship in the 288 km² test domain varies in space, and what minimum data sets are required to support the use of site-specific Z-R relations. In *Experiment 4*, we evaluate the effect of functional compression by measuring the information loss when using the deterministic Marshall-Palmer relationship instead of the empirical, 'scattered' relationship between Z and R as contained in the data. In *Experiment 5*, we investigate information gains along the radar path, i.e. when we use observations of the reflectivity measured increasingly close to the ground. We start with observations from weather radar at height, continue with observations from MRR along a vertical profile and finally use disdrometer observations of the reflectivity at the ground. In the last *Experiment 6*, we compare two methods of QPE, radar-based and rain gauge-based, using information measures and explore the benefits of merging them.

2 Data and methods

2.1 Concepts and measures from Information Theory

Since its beginnings in communication theory and the seminal paper of Claude Shannon (Shannon, 1948), Information Theory (IT) has developed into a scientific discipline of its own, with applications ranging from Meteorology (Brunsell 2010) and Hydrology (Pechlivanidis et al. 2016, Gong et al. 2014, Loritz et al. 2018) to Geology (Wellmann and Regenauer-Lieb 2012) any many others. Information Theory has been proposed as one important approach to advance catchment hydrology to deal with predictions under change (Ehret et al. 2014). A good overview on applications in Environmental and Water Engineering is given in Sing (2013), Cover and Thomas (1991) provide a very accessible yet comprehensive introduction to the topic.

Please note that while the concepts of IT are universal and apply to both continuous and discrete data and distributions, we will for the sake of clarity and brevity restrict ourselves to the latter case and work with discrete (binned) probability distributions throughout all experiments.

Information

The most fundamental quantity of IT, ~~I~~information $I(x)$, is defined as the negative logarithm of the probability p of an event or signal x (see Eq. 1). In this context the terms 'event' and 'signal' can be used interchangeably to refer to the outcome of a

random experiment, i.e. a random draw from a known distribution. and depending on the base of the logarithm it is measured in [nat] for base e, [hartley] for base 10, or [bit] for base 2.

$$I(x) = -\log_2 p(x) \quad (1)$$

Depending on the base of the logarithm, Information it is measured in [nat] for base e, [hartley] for base 10, or [bit] for base

2. We will stick to unit bits here as it is the most commonly used (especially in the computer sciences) and because it offers some intuitive interpretations. Information in the context of IT has a fundamentally different meaning than in colloquial use, where it is often used synonymously with 'data'.

Information can be described as 'the property of a signal that effects a change in our state of belief about some hypothesis (Nearing et al., 2016, section 2.3). This definition has important implications: Firstly, in order to quantify the Information content of a signal, we have to know (or at least have an estimate) of its occurrence probability *a priori*. Secondly, missing Information can be seen as the distance between our current state of belief about something and knowledge, which establishes a link between the concepts of uncertainty and Information. Information is carried by data and interestingly, the Information content of the same data reaching us can be different depending on our prior state of belief: If we already have knowledge prior to receiving the data, its Information content can only be zero and in all other cases it can at most be as large as our prior state of belief. Information can therefore be interpreted as a measure of surprise: The less probable an event is, the more surprised we are when it occurs, and the more informative are the data revealing this to us. The dependency of Information content to our prior state of belief is expressed by the prior probability p we assign to the particular event.

Compared to working with probabilities, using its log-transforms, Information, has the welcome effect that counting the total Information flown by a sequence of events is additive, which is computationally more convenient than the multiplicative treatment of probabilities.

Information Entropy

Information Entropy $H(X)$, or simply Entropy is defined as the expected or average value of Information (see Eq. 2) of a specific value or bin of a data set $X = \{x_1, x_2, \dots, x_n\}$. This means that if Information is the additional insight gained from the disclosure of the outcome of a *single* random experiment, then Entropy is the average additional insight if the experiment is repeated *many* times, thus independent, identically distributed (iid) sampling of the entire underlying probability density function (pdf). Again, we can use the terms 'expected Information' and 'expected uncertainty' as flip-side expressions of the same thing (more details on the interpretation of Entropy as uncertainty are presented by Weijts, 2011, chapter 3).

$$H(X) = \mathbf{E}[I(x)] = - \sum_{x \in X} p(x) \log_2 p(x) \quad (2)$$

In Eq. 2, $I(x)$ indicates Information and E is the expected value.

While Information is a function of the occurrence probability of a particular outcome only, Entropy is influenced by and a measure of the shape of the entire pdf. A pdf with only a single possible state of probability $p = 1$ therefore has an Entropy of zero: There will be no surprise from the disclosure of any random draw as we already know the result in advance. If on the other hand the probability mass is spread evenly over the entire value range, Entropy will be maximized. Uniform

distributions therefore serve as maximum Entropy (minimum artificially added Information) estimates of unknown pdfs. Like the variance of a distribution, Entropy is a measure of spread, but there are some important differences: while variance takes into account the values of the data and is expressed in (squared) units of the underlying data, Entropy takes into account the probabilities of the data and is measured in bit. Variance is influenced by the relative position of the data on the measure scale and dominated by values far from the mean; Entropy is influenced by the distribution of probability mass and dominated by large probabilities. Some welcome properties of Entropy are that it is applicable to data that cannot be placed on a measure scale (categorical data), and that it allows comparison of distributions from different data due to its generalized expression in bit.

Conditional Entropy

So far we know that Entropy is as a measure of the expected Information of a single distribution. We could also refer to this as a measure of self-Information, or Information we have about individual data items when knowing the data distribution.

If we do not only have a data set $X = \{x_1, x_2, \dots, x_n\}$ of single data items available, but joint data sets of paired data items $X, Y = \{x_1 y_1, x_2 y_2, \dots, x_n y_n\}$, then an obvious question to ask is 'What is the benefit of a priori knowing a realisation y coming from -Y when we want to guess a particular realisation x coming from X?'. In practice, this situation appears each time we want to make predictions about a quantity of interest x (e.g. ground rain rate) by exploiting available, related data (e.g. reflectivity), with the prerequisite that we know the general relation between the two data sets. ~~In the following, we will refer to the former as the target, to the latter as predictor(s).~~

Expanding the definition of Entropy, Conditional Entropy $H(X|Y)$ is defined as the probability-weighted (expected) Entropy of all distributions of X conditional on the prior knowledge of $Y = y_i$ (see Eq. 3).

$$H(X|Y) = \sum_{y \in Y} p(y) H(X|Y = y) = - \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2 p(x|y) \quad (3)$$

An example: Suppose we want to predict ground rainfall (our target X) based on radar reflectivity measurements (our predictor Y). The general relation between the two can be derived from a time series of joint observations and can be expressed by their joint distribution. This empirical relation is assumed to be both invariant, i.e. valid also for predictive situations, and representative, i.e. a good approximation of the 'true' relation constructed from a hypothetical, infinitely long data set. Now, if in a particular predictive situation (i.e. at a particular point in time) we know the particular reflectivity

value, then our best prediction of the corresponding rainfall value is the distribution of all past rainfall observations measured in combination with this particular reflectivity value.

If X and Y are completely independent, then prior knowledge of y when guessing x from a data pair (x, y) does not help at all. In this case each conditional distribution $p(X|Y = y_i)$ equals the marginal distribution $p(X|Y)$ and the Conditional Entropy $H(X|Y)$ is exactly equal to the unconditional Entropy $H(X)$. While such a situation is clearly not desirable, it also provides us with the useful insight that even if we apply a completely useless predictor, we can never make worse guesses than if we ignored it altogether. This is called the Information Inequality (Eq. 4), or simply 'Information can't hurt'. For the complete proof, see Cover and Thomas (2006) p. 28.

$$H(X) \geq H(X|Y) \quad (4)$$

Cross Entropy

Entropy and Conditional Entropy measure information contained in the shape of distributions and the underlying data. Their calculation depends on prior knowledge of a the pdf of from these data and implies the assumption that this pdf is invariant and representative. In practice, this condition is not always fulfilled, either because we construct the pdf from limited data or because the system generating the data is not invariant. In these cases and we work with approximation estimates, or models, of the pdf instead, which - This means we estimate the information attached to a signal based on imperfect premises, which we in turn pay for with increased uncertainty. Cross Entropy $H_{pq}(p||q)$ as defined in Eq. (5) quantifies exactly this.

$$H_{pq}(p||q) = - \sum_{x \in X} p(x) \log_2 q(x) \quad (5)$$

In Eq. (5), p is the true probability distribution of the data, q the assumed-to-be-true distribution. This means we calculate the information content of a particular signal based on it's a priori known, but only approximate occurrence probability $q(x)$. However, as we draw from the real distribution p , this particular signal really occurs with probability $p(x)$. The larger the mismatch between p and q , the more additional questions we have to ask. In the best case, if premises are correct and q is identical to p , H_{pq} will be equal to H . In the worst case H_{pq} will go to infinity. An accessible and comprehensive treatment of this topic can be found in Weijs (2011), chapter 3.5.

Kullback-Leibler divergence

If both the Entropy of a distribution and the Cross Entropy between the true distribution and an approximation model (a model) thereof are known, we can separate these two components of total uncertainty: Uncertainty due to the shape of the true distribution, and uncertainty because we do not know it exactly. The latter is measured by the Kullback-Leibler

divergence $D_{KL}(p||q)$ (see Eq. 6 and Kullback and Leibler 1951). It can either be calculated by the probability-weighted difference of true and assumed-to-be-true occurrence probabilities, or by subtracting the Entropy from the Cross Entropy.

$$D_{KL}(p||q) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{q(x)} = H_{pq}(p||q) - H(p) \quad (6)$$

5 2.2 [Modeling and evaluation strategy methods](#)

Data-based models and predictions, [Information-based model evaluation](#)

Suppose we have a data set of many repeated, joint observations of several variates (data tuples), e.g. a time series of joint observations of R and Z. For each data tuple we can consider their values as coordinates in a multi-dimensional space, whose number of dimensions equals the number of variates. Let us further suppose we choose, separately for each variate, a value range and a strategy to subdivide it into a finite set of bins. The first can be based on the observed range of data or on physical considerations, the latter on the resolution of the raw data or on user requirements about the resolution at which the data should be evaluated or predictions should be made. In this context, many methods for choosing an optimal binning strategy have been suggested (e.g. Knuth 2013), but still the most popular and straightforward method that at the same time introduces minimal side [Information](#) is uniform binning, i.e. splitting of the value range into a finite number of equal-width bins. Throughout this paper, we will stick to uniform binning.

With range and binning chosen, we can map the data set into a multivariate, discrete frequency distribution as shown in Fig. 1, left plot. Normalization with the total number of data tuples yields the corresponding multivariate, discrete probability distribution (dpd). Note that the mapping comes at the price of a certain [Information](#) loss: Firstly, we lose the [Information](#) about the absolute and relative position of the data tuples in the data set ([e.g. for time series data, we discard the time stamp and the time order of the data](#)), secondly we lose any [Information](#) about the data tuple values at resolutions higher than the bin width. Nevertheless, the way how probability mass is spread within the coordinate space is an indication of its structuredness, and its Entropy is a direct measure thereof. High values of Entropy indicate that probability mass is widespread, which also means we are highly uncertain when guessing a particular data tuple randomly drawn from the dpd. On the other side, if all probability mass is concentrated in a single bin, Entropy will be zero and we can with zero uncertainty predict the values of a data tuple randomly drawn from the dpd.

So far we used Entropy to measure uncertainty if we'd have to guess all values of a data tuple randomly drawn from an a priori known distribution. But often we do not just know the distribution a priori, but also the values of parts of the data tuples (e.g. we know Z and the empirical Z-R relation and want to predict R). In this case, our uncertainty about the target variate R given the predictor variate Z and the relation among the two can be measured by Conditional Entropy, which we know from Eq. 4 is always smaller or equal the unconditional Entropy of the target. In Fig. 1, left plot, the conditional

frequency distribution of the target X given predictor Y is shown in red, in the right plot, also in red, the related conditional probability distribution and, in blue, the unconditional probability distribution of the target.

In short, we can consider a dpd constructed from a data set as a *minimalistic simple predictive model*, which, given some values for the predictor variates provides a probabilistic prediction of the target variate. It is minimalistic in the sense that it involves only a small number of assumptions (representativeness of the dpd) and user decisions (choice of the binning scheme). Compared to more common modeling approaches, where relations among data are expressed by deterministic equations, the advantage of a such a dpd-based model and prediction is that it yields – it provides a joint probabilistic predictions in the form of conditional dpd's, which include a statement about the target value and the uncertainty associated with that statement, – expressed by the spread of the predictive dpd of the target. Also, as dpd-based modeling does not involve strong regularizations, it reduces the risk of introducing wrong information. When building a standard model, choosing a particular deterministic equation to represent a data-relation, which is a common form of regularization, involves the risk of introducing bias, and it ignores predictive uncertainty. In this context, Todini (2007) provides a very comprehensive discussion of the relative merits and problems of data-based models and models relying on additional information in the form of physical knowledge, and the role and treatment of predictive uncertainty. And as long as we remain with our predictions within the domain of the data we used for constructing the model, our uncertainty statements are accurate, i.e. they are neither over nor underconfident.

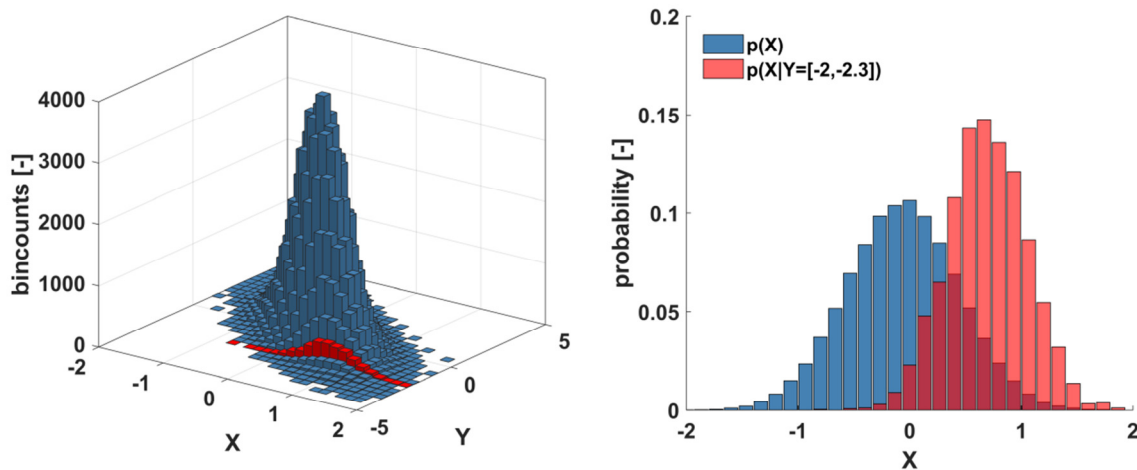


Fig.1. Illustration of a bivariate conditional probability distribution as a simple data-based predictive model. Left: Joint histogram of target X and predictor Y (blue). Conditional histogram of target X given predictor values from the interval $[-2,-2.3]$ (red). Right: Marginal (unconditional) probability distribution of target X (blue) and conditional probability distribution of target X given predictor values from the interval $[-2,-2.3]$ (red).

Things are different, though, if we apply the model to a new situation, i.e. if we construct the model from one data set and use it with predictor data of another under the assumption that the predictor-target relation expressed by the model also holds for the new situation. This is clearly only the case if the learning and the application situation are identical, in most cases

however, the situations will differ and we will pay for this inconsistency with additional uncertainty. We can measure this uncertainty by Kullback-Leibler divergence (Eq. 6) between the dpd we use as a model and the true dpd of the application situation. Total predictive uncertainty is then expressed by Cross Entropy (Eq. 5) and can be calculated as the sum of the predictive uncertainty given a perfect model (expressed by Conditional Entropy), which is limited by the Information content of the predictors about the target, and the inadequacy of the model (expressed by Kullback-Leibler divergence).

We will apply this approach to construct and analyse predictive models throughout the paper, and use it to learn about the Information content of various predictors about our target, ground rainfall R , and learn about the Information losses when applying predictive models that were e.g. constructed from limited data sets, constructed at different places, or simplified by functional compression. Note that expressing relations among predictor and target data by dpd's is not limited to data-based models, in fact functional relationships of any kind (e.g. the Marshall-Palmer Z-R relation) can be expressed in a dpd. In that sense, the framework for model building and testing together with the measures to quantify uncertainty we use here are universally applicable.

Benchmark models and minimum model requirements

Expressing predictions by probability distributions, and expressing uncertainty as 'Information missing' as described in the previous sections has the advantage that we can build default models providing lower and upper bounds of uncertainty, which we can then use as benchmarks to compare other models against. The smallest possible uncertainty occurs if the predictive distribution of the target is a Dirac, i.e. the entire probability mass is concentrated in a single bin. In such a case, irrespective of the number of bins covering the value range, the Entropy of the distribution is zero: $H_{Dirac} = 0$. If such a case occurs, we know we have applied fully informative predictors and a fully consistent model. Things are more interesting when we want to formulate upper bounds: The worst case occurs if we use a model which is unable to provide a prediction for a set of predictors in the application case. This happens when the particular situation was never encountered in the learning data set, but appears during application. In that case, $p(x)$ in Eq. 6 will be non-zero, $q(x)$ will be zero, and Kullback-Leibler divergence will be infinite, indicating that the model is completely inadequate for the application situation. In this case total uncertainty will also be infinite, no matter how informative the predictors are. Infinity as an upper bound for uncertainty is not very helpful, though, and we can do better than that: If the above case occurs, we can argue that the model is inadequate just because it learned from a limited data set, and since we require the model to provide a prediction for all predictive situations, we can allocate a small but non-zero probability to all bins of the model. In this paper, we did so using the minimally invasive maximum Entropy approach suggested by Darscheid et al. (2018). Note that this was only required in Experiments 2, 3 and 4, when we calculated Kullback-Leibler divergence between a reference and models based either on very small samples (Experiments 2 and 3) or deterministic relations (Experiment 4).

Given that we successfully avoid infinite Kullback-Leibler divergence, the worst thing to happen is that our predictions are completely uninformative, i.e. we provide a uniform distribution across the entire value range of the target. In this case, the Entropy of the distribution is equal to the logarithm of the number of bins: $H_{uniform} = \log_2(\text{number of bins})$. This means

that for all cases where we are sure that infinite Kullback-Leibler divergences will not occur, we can provide an upper limit of total uncertainty which is dependent on the required resolution of our predictions.

Finally, for the special cases where we know that the model we use is perfect (typically because we apply it to the same situation it was constructed from), we know that Kullback-Leibler divergence is zero, and total uncertainty equals Conditional Entropy. In this case, we can state another upper bound for uncertainty: If at worst the predictors we use are completely uninformative, Conditional Entropy will be equal to the Unconditional Entropy of the target distribution (see Eq. 4).

We will use these lower and upper bounds for total uncertainty throughout the experiments to put the performance of the tested models into perspective.

Sampling strategy

In Experiments 2 and 3 we investigate the effect of limited sample size, i.e. the information loss (or uncertainty increase) if we construct a model not from a full data set, but from subsets drawn thereof. This corresponds to the real-world situation of building models from available, limited data. For the sake of demonstration we assume in the experiments that a long and representative reference data set is available for evaluation. While this is clearly not the case in real-world situations, we can get answers from such experiments for practically relevant questions such as 'what is a representative sample, i.e. how many observations are required until a model build from the data does not further change when adding more observations?' or 'should we build a model from locally available but limited data, or should we use a model learned elsewhere, but from a large data set?'

Throughout the experiments, we apply simple random sampling without replacement to take samples from data sets. In order to reduce effects of chance, we repeat each sampling 500 times, calculate the results for each sample and then take the average.

2.3 Data

This study uses data from various sources collected in a four-year period (Oct 1, 2012 – Sep 30, 2016) within the research project CAOS (Catchments As Organised Systems). For more detail on project goals and partners see Zehe et al. (2014). QPE was an important component of CAOS, and to this end a focus was laid on measuring rain rates and drop size distributions (DSD) by six laser optical disdrometers, two vertical pointing K-band micro rain radars, standard rain gauges and weather radar. Table 1 provides an overview on these and additionally used data.

Study area and hydroclimate

The project was conducted in the Attert basin, which is located in the central western part of the Grand Duchy of Luxembourg and partially in Eastern Belgium (see Figure 2) with a total catchment area of 288 km². The landscape is orographically slightly structured, with a small area underlain by sandstones in South, with heights up to 380 m a.s.l, a wide

area of sandy marls in the centre part, in which the main Attert river flows from west to east, and an elevated region in the North which is part of the Ardennes massif and reaches elevations up to 539 m a.s.l.

The study area is situated in the temperate oceanic climate zone (Cfb according to the Köppen–Geiger classification, Köppen and Geiger 1930). Precipitation is mainly associated with synoptic flow regimes with a westerly wind component and amounts to around 850 mm (Pfister et al., 2005 and Pfister et al., 2000), ranging from 760 mm in the centre to about 980 mm in the north-western part (Faber, 1971) due to orography and luv-lee effects. Especially during intensive rain events within a synoptic north-westerly flow regime, the main part of the Attert region lies in the rain shadow of the Ardennes, whereas the outermost north-western part is excluded from this lee effect and still within the region of local rain enhancement by upward orographic lift (Schmithüsen, 1940). This also partly explains local differences in the annual precipitation regime: While in the elevated regions of the north-west, the main rain season is (early) winter, in the central lowlands convective summer rainfall also provides a substantial contribution. Interestingly, the latter mainly occurs along two main storm tracks: One north-eastern starting in the south-west corner of Luxembourg and touching the study area in the east; one north-eastern starting in Belgium and touching the study area in the west.

In our study, we investigated both the effects of location and season on drop-size distributions and ultimately on QPE.

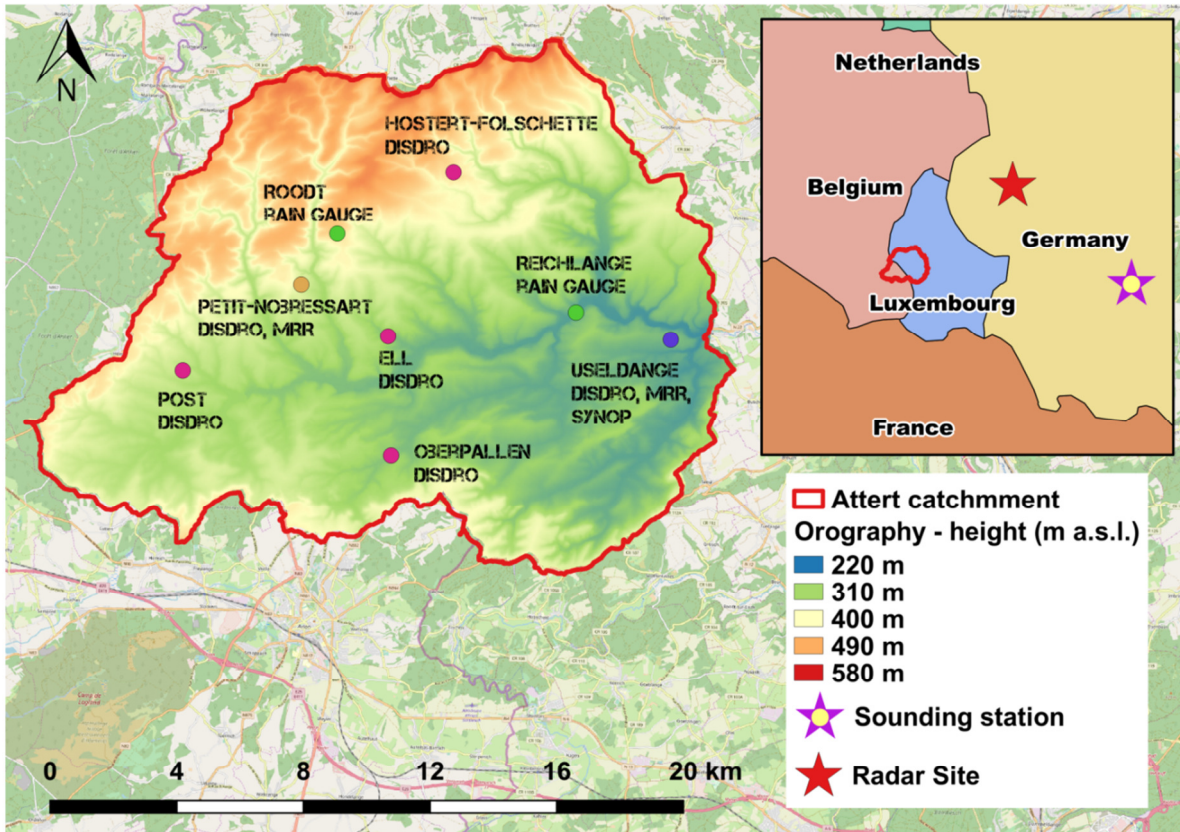


Fig. 2 The position of the Attert catchment in Luxembourg and Belgium with superimposed Orography (in m a. s.l.) and the locations of the MRR's, disdrometers, rain gauges, the synoptic station (ASTA) and radar site (Neuheilenbach) and the rawinsonde launching (sounding) station (Idar-Oberstein- WMO-ID 10618) as well as the scale and orientation of the small-scale map.

5

Radar Data

We used 10-minute reflectivity data from a single pol C-Band Doppler radar located in Neuheilenbach (see Figure 2), and operated by the German Weather Service (DWD). The raw volume data set has an azimuthal resolution of 1° and 500 m radially, the latter a mean of the 125 m range resolution (short pulse). The antenna's -3dB-beamwidth is 1°. The distance of the study area to the radar site is between 40 and 70 km, which renders high-resolution data and avoids cone of silence issues. The raw data were filtered by static and Doppler clutter filters and bright band correction (Hannesen, 1998) but no attenuation corrections were applied. Second trip echoes (Bücker, 2010) as well as obvious anaprop echoes were also removed, as especially the latter are prominent in this area during fall and spring (Neuper, 2009). From the corrected data we

15

constructed a pseudo PPI (plan position indicator) dataset at 1500 m above ground and, to make the data compare- and combinable to all other data used, took hourly averages.

MRR Data

5 We also used drop size spectra measured at 1500 meter and 100 meter above ground by two vertical pointing K-band METEK micro rain radars (MRR) (Löffler-Mang et al. 1999; Peters et al. 2002) located at the sites Useldange and Petit-Nobressart (Figure 2). We operated the MRRs at 100 vertical meters and 10 seconds temporal resolution, but for reasons of storage and processing efficiency did all further processing on 1-minute aggregations thereof. The raw Doppler spectra were transformed to drop size distributions via the drop size – fall velocity relation given in Atlas et al (1973). From the drop size
10 distributions the rain rate and the reflectivity were calculated using the 3.5th-67th and the 6th statistical moments of the drop size distributions (as in Doviak and Zrinic, 1993, Ulbrich and Atlas, 1998 and Zhang et al., 2001). In doing so we assumed the vertical velocity of the air to be negligible, although this may sometimes play a significant role (Dotzek and Beheng 2001). Furthermore the intensity of the electromagnetic waves of the MRR is attenuated on the propagation path by different processes. Due to the short wavelength, the attenuation by water vapour is relatively strong. This process is neglected in this study. Attenuation due to rain can also be significant, especially at moderate and higher rain rates, if higher altitudes (with a long beam propagation path) are considered. Without considering the attenuation by rain a height dependent underestimation of the rain rates would be retrieved. To this point we corrected for attenuation assuming liquid rain with applying a step wise path integrated rain attenuation following the manufacturer indications. As before, we converted the data to 1-hour values by averaging (reflectivity) and summation (rain sum).
15
20

Disdrometer Data

We also deployed six second generation OTT Particle Size and Velocity (PARSIVEL², Löffler-Mang and Joss, 2000) optical disdrometers in the study area to measure drop size distributions at ground level (Figure 2) in 1-minute resolution. Two were located at the same sites as the MRRs (Useldange and Petit-Nobressart). The others were placed such as to both capture the
25 hydroclimatic variations in the study area and to cover it as uniformly as possible. We applied a quality control to the raw data as described by Friedrich et al. (2013 a, b), converted the filtered data to drop-size concentrations per unit air volume to make them comparable to the weather radar and MRR data, then converted them to reflectivity and rain rate using the 3.5th and 6th statistical moments of the drop size distributions and finally took 1-hour averages and sums thereof.

30 *Rain gauge data*

Next to the rain rate retrieved from the disdrometer data, we also used additional observations from standard tipping bucket rain gauges at sites Useldange, Roodt and Reichlange (Figure 2). Quality-controlled rain gauge data from Useldange and Roodt were provided by the Administration des services techniques de l'agriculture (ASTA), who operates a nationwide network of surface weather stations for agricultural guidance. Raw rain gauge data at Reichlange were provided by the

Hydrometry Service Luxembourg. We applied plausibility checks to all these data, eliminated questionable data, cases with solid precipitation (based on the output of the hydrometeor classification algorithm of the disdrometers) and finally took 1-hour sums.

5 *Additional predictors*

In addition to direct observations of precipitation and drop-size spectra, we collected a number of operationally available data to test their value as additional predictors for QPE. We selected standard meteorological in-situ observations such as 2 m temperature, relative humidity, zonal and meridional wind speeds, season indicators such as month and tenner-day of the year as well as synoptic indices such as convective available potential energy (CAPE) and classified circulation pattern. The latter two are operationally provided by the German weather service. All predictors are listed in Table 1.

The in-situ observations were taken from ASTA station Useldange (Figure 2), and like all other observations underwent additional quality filtering and 1-hour aggregation. The CAPE values are based on rawinsonde data obtained from station Idar-Oberstein (WMO-ID: 10618) located about 90 km east of the study area with a temporal resolution of 6h. The soundings were downloaded from the University of Wyoming homepage (see *data availability*) and checked for contamination by convection as described in Bunkers et al. (2006). The CAPE values were calculated as surface based CAPE using the virtual temperature (Vasquez 2017). Selecting CAPE as a descriptor for raindrop size distribution followed the reasoning that updraft strength has an influence on the size of the raindrops (Seifert and Beheng 2006). We used a base-10 logarithm transform of CAPE to assure high resolution at lower CAPE values and at the same time sufficient population of the higher classes.

Classified circulation patterns: The German Weather Service provides daily an objective classification of the prevailing circulation pattern over Europe into one of 40 classes (Dittmann et al. 1995). It is based on numerical weather analysis from one of the operational forecast models of DWD at 12 UTC and considers wind direction, cyclonality (high or low pressure influence), and atmospheric humidity. Selecting the circulation as a predictor was based on the perception that it should contain information about both the prevailing precipitation process (stratiform or convective) and about the presence and origin of air masses, which could influence drop size spectra due to the presence or absence of characteristic cloud condensation nuclei and ice nucleation nuclei.

Binning choices

Using all these data in the methods as described in section 2.2 requires binning. Our general binning strategy was to cover the entire data range, divide it into as few bins as possible to keep bin populations high, but at the same time use enough bins to resolve the shape of the data distributions. We used uniform binning whenever possible, but for strongly skewed data such as reflectivity and rain rate we applied irregular binning. For the latter, we defined the edges of the first bin such as to cover the range from the smallest possible value (0 mm) and 0.2 mm, which is the detection limit of most tipping bucket rain collectors. So essentially, this bin covers all cases of non-zero but irrelevant rain. For the last bin, we set the upper edge such

that it still covered rainfall based on the largest observed reflectivity value transformed to rain rate by the standard Marshall-Palmer Z/R relation ($Z = 200 R^{1.6}$), which was 119.9 mm/h. For the bins in between, we increased bin width with rain rate (see footnote to Table 1) to acknowledge both the increasing uncertainty and sparsity of high rainfall observations. With the binning for rain rate fixed, we applied the Marshall-Palmer relation in reverse to get the corresponding bin edges for reflectivity.

Data filter

Rainfall is an intermittent process, and quite expectedly most of our 4-year series contained zeroes for rainfall. If we had used this complete data set for analysis, the results would have been dominated by these dry cases, which however was not what we were interested in. We therefore applied a data filter to select only the hydro-meteorologically relevant cases with measurements from all available stations and at least two rain gauges showing rainfall ≥ 0.5 mm/h. Additionally, cases with solid precipitation were excluded using the output from the disdrometers' present weather sensor software. Altogether 11984 data sets passed this 'minimum precipitation' filter, which amounts to almost 17 months of hourly data.

In Figure 3, for the filtered data binned probability distributions of the most important variables are shown.

Table 1. Summary of the raw data used in the experiments: Description, summary statistics, binning.

Symbol	Unit	Value range*	Binning**	Description
RR0	[mm/h]	[0, 86.98]	[+] 18	Rain rate at ground level, measured at 8 sites by disdrometers and rain gauges
dBZ1500Rad	[dBZ]	[-99, 56.28]	[++] 18	Decibel of reflectivity, measured 1500 m above ground at 8 sites by weather radar
dBZ1500MRR	[dBZ]	[-99, 56.28]	[++] 18	Decibel of reflectivity, measured 1500 m above ground at 2 sites by MRR
dBZ100MRR	[dBZ]	[-99, 56.28]	[++] 18	Decibel of reflectivity, measured 100 m above ground at 2 sites by MRR
dBZ0Dis	[dBZ]	[-99, 56.28]	[++] 18	Decibel of reflectivity, measured at ground level at 6 sites by Disdrometer
logCAPE	[log J/kg]	[0, 11.3]	[0:1:12] 13	Logarithm of the surface based convective available potential energy (CAPE)

GWLo	[-]	[1, 40]	[1:1:40] 40	Circulation pattern according to German Weather Service objective classification
TA2	[°C]	[-15.2, 36.4]	[-20:5:40] 13	Air temperature, measured 2 m above ground at station Useldange
RH2	[%]	[18.1, 100]	[0:10:100] 11	Air relative humidity, measured at 2 m above ground at station Useldange
U10, V10	[m/s]	[-13.7, 19.1]	[-14:2:20] 18	Wind velocity in East-West (u) and North-South (v) direction, measured 10 m above ground at station Useldange
ToY	[-]	[1, 36]	[1:1:36] 36	Tenner-day of the year
MoY	[-]	[1, 12]	[1:1:12] 12	Month of the year
Statnum	[-]	[1,8]	[1:1:8] 8	Station number ***

* [min, max]

** [center of first bin, uniform bin width, center of last bin] number of bins

*** 1: Ell, 2: Hostert-Folschette, 3: Oberpallen, 4: Petit-Nobressart, 5: Post, 6: Useldange, 7: Reichlange, 8: Roodt

+ Irregular binning with edges [-0.1 0.19 1 2 4 6 8 10 15 20 25 30 40 50 60 70 90 110 130]

5 ++ Irregular binning with edges [-100 11.5 23 27.8 32.6 35.5 37.5 39 41.8 43.8 45.4 46.6 48.6 50.2 51.5 52.5 54.3 55.7 56.8]

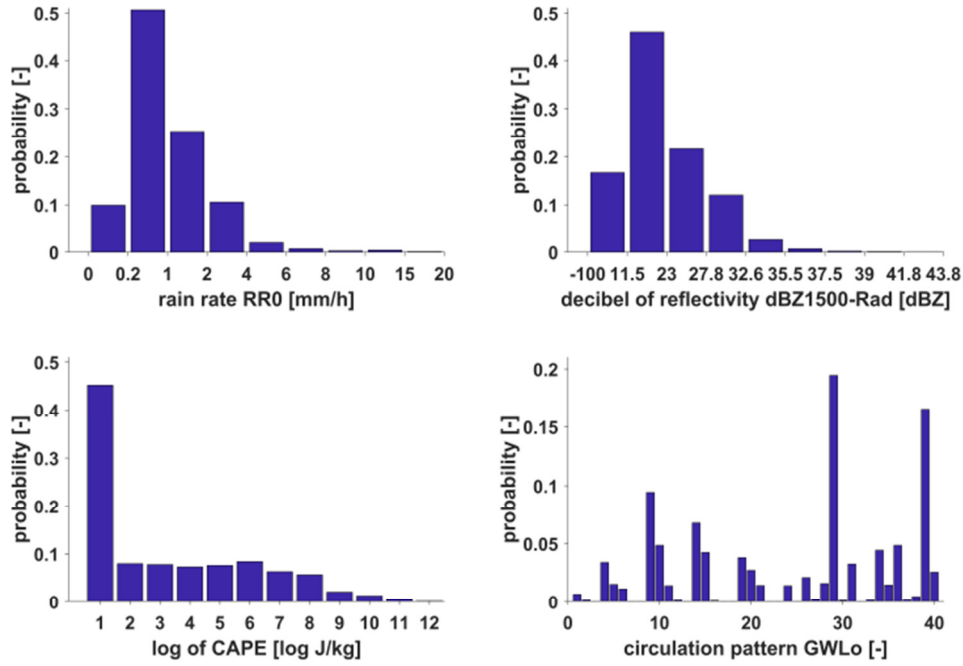


Fig.3. Discrete probability distributions of the most important variables used in the experiments (filtered data set). The description of variables and binning can be found in Table 1. X-axes for rain rate and reflectivity are truncated at the right to avoid display of empty bins.

5

3 Results and discussion

3.1 Experiment 1: Information in various predictors

In this experiment we explore the information content about ground rainfall R in various predictors (see Table 1). We selected the predictors with the constraint of being operationally available at any potential point of interest, which applies most importantly to reflectivity measured by weather radar, but also to the predictors we assumed to be spatially invariant within the test domain: convective available potential energy (as surface based CAPE), circulation pattern, air temperature and humidity, wind, and season. We excluded reflectivity measurements by MRR and disdrometer, as these are usually only available at a few locations.

We used ground rainfall observations from the eight rain gauges in the test domain as target data, filtered the raw data with the 'minimum precipitation' filter and created models by various predictor combinations. We measured the usefulness of the predictors with Entropy and Conditional Entropy (Eq. 2 and Eq. 3). The results are shown in Table 2.

Table 2. Entropy of the target (RR0) and the benchmark distribution (RR0 uniform), Conditional Entropies of the target given one or several predictors, non-normalized and normalized. Underlying data are from the filtered data set. Predictors are ordered by descending Conditional Entropy.

case	Predictor(s)	(Conditional) Entropy	Relative ***
		H(X), H(X Y) [bit]	(Conditional) Entropy [%]
1	RR0 uniform	4.17*	219
2	RR0	1.90**	100
3	dBZ1500Rad	1.61	84.7
4	dBZ1500Rad, RH2	1.57	82.6
5	dBZ1500Rad, U10	1.57	82.6
6	dBZ1500Rad, V10	1.57	82.6
7	dBZ1500Rad, TA2	1.55	81.6
8	dBZ1500Rad, logCAPE	1.55	81.6
9	dBZ1500Rad, MoY	1.54	81.1
10	dBZ1500Rad, GWLo	1.52	80
11	dBZ1500Rad, ToY	1.47	77.4
12	dBZ1500Rad, MoY, GWLo	1.30	68.4
13	dBZ1500Rad, ToY, GWLo	1.19	62.6

5 * unconditional Entropy of the benchmark uniform distribution

** unconditional Entropy of the target

*** $H(\text{RR0}|\text{predictor(s)}) / H(\text{RR0}) * 100$

10 There are two upper benchmarks we can use to compare the different QPE models against: Unconditional Entropy if we know nothing but the binning of the target and use a uniform (= maximum Entropy) distribution for prediction (case 1 in Table 2), and the unconditional Entropy of the observed distribution of the target (case 2 in Table 2), which we used as a reference here. The difference between the two is considerable (119%), which means that merely knowing the true distribution of RR0 is already a valuable source of [I](#)nformation.

The next important source of Information are radar reflectivities: If we use it as a single predictor (case 3), uncertainty is reduced by 0.28 bit or 15.3% compared to RR0. Note that this approach directly applies the reflectivity data provided by the radar, no side Information was added nor existing Information in the data destroyed by applying an additional Z-R relation. These results shows on the one hand nicely, that radar reflectivity is one of the most important sources of Information. On the other hand - since the reduction in uncertainty seems-is to-be-quiteet low when RR0 is conditioned by dBZ1500Rad - it seems, that the high variability in the Z-R relationship limits entropy reduction. The reason for this may include the effects of the vertical profile of reflectivity and-even attenuation at high rain rates.

Using each of the other predictors separately did not reduce uncertainty much (not shown), we therefore only show results for the cases where they were applied as 2-predictor models in combination with radar reflectivity (cases 4 to 11). If we compare the relative Conditional Entropies of their predictions to those of the radar-only model, we can see that neither the ground meteorological observations nor CAPE contained much additional Information (cases 4 to 8). Instead, the three most informative models (cases 9 to 11) either distinguish the relation between R and Z by circulation pattern or by season, which corresponds to the operational practice of many weather services to use one Z-R relation for summer and one for winter.

Based on these results, we built and evaluated 3-predictor models only with combinations of these relatively informative predictors (cases 12 and 13). The Information gain from using three predictors in combination is considerable, and altogether uncertainty reductions down to 68.4 % (case 12) and 62.6 % (case 13) compared to the 'target-distribution-only case' (case 2) were achieved. The benefit of applying a season- and circulation pattern dependent relation between R and Z also becomes obvious if we compare case 3 (the radar-only model) to cases 12 and 13: In the first case, uncertainty is reduced by 16.3 %, in the latter by 22.1 %.

An obvious conclusion from these findings would be to build better models by simply adding more predictors, which according to the Information Inequality (Eq. 4) never hurts. In fact, when learning from limited data sets, adding enough predictors will in the end result in perfect models with zero predictive uncertainty. However, there is a catch in this known as overfitting-or 'Curse of Dimensionality' (Bellman 1957): In order for a model to be robust in the sense of 'being only weakly sensitive to the presence or absence of particular observations in the learning data set' and 'performing well not only within the training data set, but also on data unseen during training', it must be supported by an adequate number of learning data, and this number grows exponentially with the number of predictors included in the model. -So instead of adding more predictors, we will explore the robustness of our models in the next experiment.

3.2 Experiment 2: The effect of sample size

The data base and filter used here are identical to the previous experiment, so a set of overall 11984 joint observations of the target (rain rate at the ground) and predictors (radar reflectivity, circulation pattern, tenner-day and month of the year) were available. From these data we built and tested altogether four predictive models as described in Experiment 1: A single predictor model applying radar reflectivity only, two double predictor models applying radar reflectivity and tenner-day or month of the year, and a triple predictor model using radar reflectivity, month of the year and circulation pattern. The

difference to Experiment 1 is that now we do not only apply the entire data set but also randomly drawn samples thereof to build the model (see section 2.2 for an explanation of the sampling strategy). Each model is then applied to and evaluated against the full data set. In this case, total predictive uncertainty is measured by Cross Entropy (Eq. 5), which is the sum of Conditional Entropy of the target given the predictors for the full data set and Kullback-Leibler divergence of the sample-based model and the model built from the full data set (see section 2.2).

The results are shown in Fig. 4 as a function of sample size. As in Experiment 1, we included the benchmark uncertainties for applying a maximum Entropy model (red horizontal line, case 1 in Table 2) and a zero-predictor model (black horizontal line, case 2 in Table 2) to put the other models into perspective.

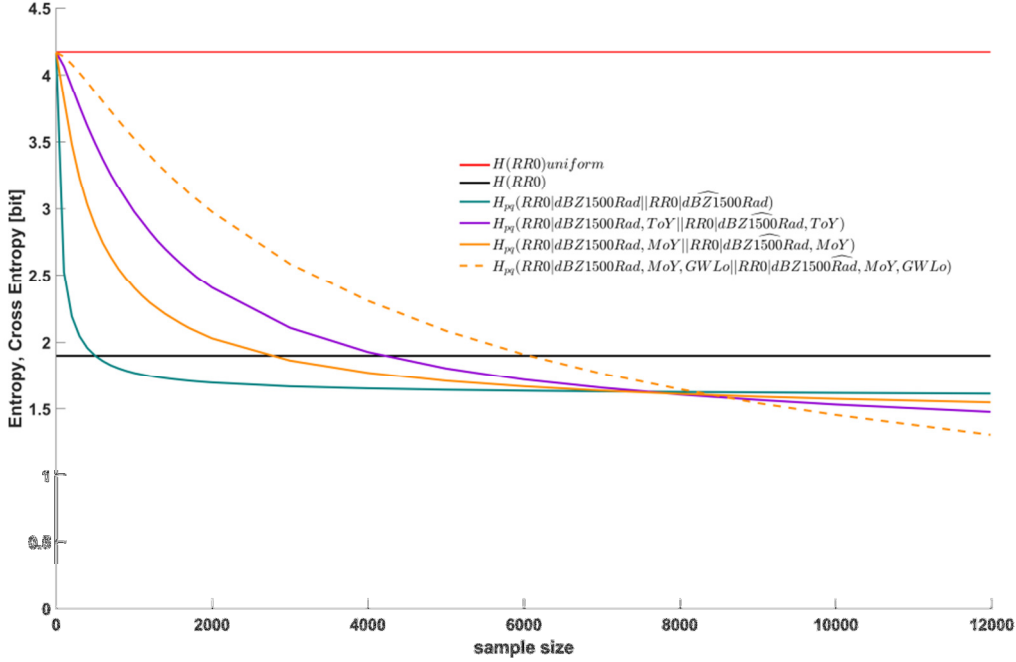


Fig. 4. Entropy of the unconditional target distribution (filtered data set, black line), Entropy of the benchmark uniform distribution (filtered data set, red line), Cross Entropies between conditional distributions of the target given one, two and three predictors of the filtered data set and samples thereof (green, purple, yellow and dashed yellow lines). The 'hat' symbol indicates a sample.

At the right margin of Figure 4, Cross Entropies are shown for the case when the sample comprises the entire data set. In this case, Kullback-Leibler divergence is zero, and total uncertainty equals Conditional Entropy. This is the same situation as in Experiment 1 and the values correspond to those in Table 2 for cases 3, 9, 10 and 12. The triple-predictor model (dashed yellow line) quite expectedly outperforms the double predictor models (solid purple and yellow lines), which in turn outperform the single-predictor model (solid green line). Things look differently, though, if only samples are used for learning. Consider the single-predictor model: For very small sample sizes close to zero, the Information content of the sample is close to zero, and hence predictive uncertainty is close to that of the ignorant maximum Entropy model (red line)

and considerably higher than that of the zero-predictor model. However, when increasing the size of the sample just a little, its information content quickly rises and Cross Entropy drops. In fact, when learning the relation between radar reflectivity and ground rainfall from only about 2000 joint observations, the model is almost identical to a model learned from the full data set of 11984 joint observations: Kullback-Leibler divergence is almost zero and Cross Entropy is almost as low as for the model built from the full data set. From this we can conclude that the full data set contains considerable redundancy, which in turn implies that we can build robust single-predictor models from the available data.

We can interpret the lines in Fig. 4 as learning curves, or more specifically they represent the information about the target contained in samples of different sizes. If we now consider the two double-predictor models and the triple-predictor model and compare it to the single-predictor model, we can see that the more predictors we add, the slower the learning rates become and the longer learning takes (the curve inclinations are lower for small sample sizes, but remain non-zero for larger samples). For the two double-predictor models this means that samples larger than about 8000 (about two thirds of the data set) are required before their total predictive uncertainty fall below that of the single-predictor model. For the triple-predictor model, even samples larger than about 8500 are required, and interestingly the model continues to learn even for very large sample sizes (the yellow dashed line is still inclined even for large samples).

The learning behaviour of the models, which differs with the number of predictors used, is a manifestation of the Curse of Dimensionality, and visual examination of learning curves of different models as plotted in Fig. 4 allows two choices: For a given sample size, we can choose the best (= least uncertain) model; for a given size data set, we can choose the model with the best trade-off between performance and robustness. For the latter choice, we can establish selection criteria such as 'a model qualifies as robust if it learns from at most two thirds of the available data at least 95% of what it can learn from the entire data set' and then choose the best model satisfying this criterion. From the models displayed here, according to our subjective choice, the double-predictor model using a month-specific relation between radar reflectivity and ground rainfall provides the best trade-off between performance and robustness.

3.3 Experiment 3: Site-specific Z-R relations

In this experiment we investigate the information content in spatial position by learning and applying site-specific relations between radar reflectivity and ground rain rate (in the previous experiments we applied them in a spatially pooled manner). We used data filtered with the the 'minimum precipitation' filter again, so a set of overall 11984 joint observations of ground rain rate and radar reflectivity were available. We included spatial information by simply using the ID number of each station ('Statnum' in Table 1) as an additional predictor, which means that we built and applied relations between reflectivity and rain rate at the ground specifically for each station in the test domain (see Fig. 2). As in the previous experiment we evaluated the predictive performance of these models as a function of sample size with Cross Entropy. The results are shown in Fig. 5.

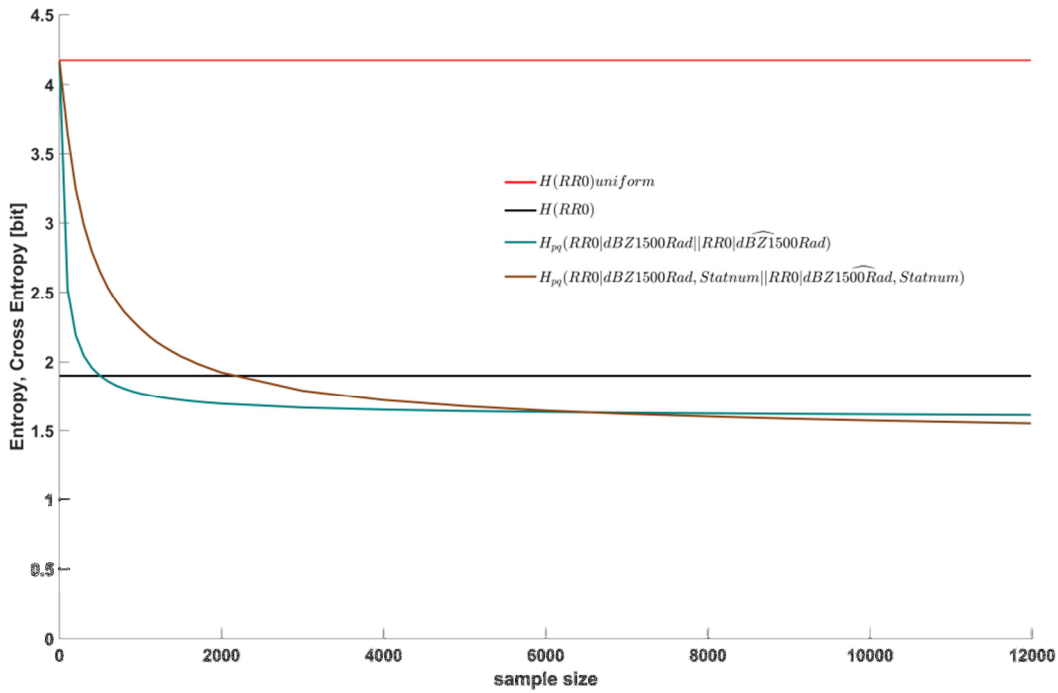


Fig. 5. Entropy of the unconditional, target distribution (filtered data set, black line, same as in Figure 4), Entropy of the benchmark uniform distribution (filtered data set, red line, same as in Figure 4), Cross Entropies between conditional distributions of the target of the filtered data set and samples thereof: Single predictor model applying dBZ1500Rad (green line, same as in Figure 4), double predictor model applying dBZ1500Rad & station number (brown line). The 'hat' symbol indicates a sample.

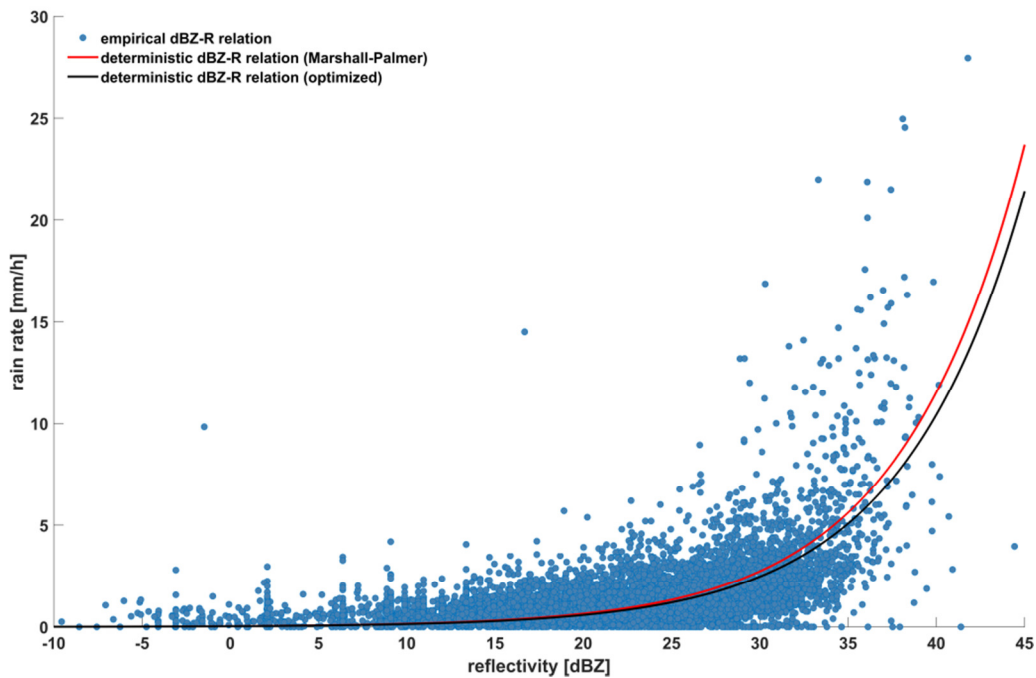
The red and black horizontal lines are the same as in Fig. 4 and as before represent the benchmark unconditional Entropies of the target for a maximum Entropy uniform distribution and the observed distribution. The green line was also included in Fig. 4: It represents the Conditional Entropy of the single-predictor model applying radar reflectivity only. The brown line shows the performance of the two-predictor model including station ID. The overall pattern is similar to Experiment 2: Adding a predictor reduces total uncertainty if the full data set is used for learning (at the right margin, the brown line lies below the green), but higher-predictor models require more data for learning (the green line descends slower and longer than the brown). Here, the double predictor model outperforms the zero-predictor and the single-predictor model only for samples larger than 2200 and 6500, respectively.

Overall, the Information gain of using site-specific Z-R relations is moderate (Cross Entropy of the radar-only model for the full data set is 1.61 bit, and 1.55 bit for the site-specific model) and lower than when distinguishing Z-R relations according to time or circulation pattern (cases 9, 10 and 11 in Table 2). This is not very surprising if we consider the extent of the test domain: The largest distance between two stations is 9.5 km (between stations Oberpallen and Useldange), and the largest elevation difference is 182 m (between stations Reichlange and Roodt). Across these relatively small distances, it appears reasonable that Z-R relations do not differ substantially. However, this could be different when working in larger domains or

in domains with hydro-meteorologically distinctly different subdomains, such as lowlands and mountain areas or different climate zones (as shown for example in Diem 1968).

3.4 Experiment 4: The effect of functional compression

5 | In this experiment we evaluate the effect of functional compression by measuring the I information loss when using a deterministic function to express the relation between radar reflectivity and ground rain rate instead of the empirical relation derived from the data. As before, we used all joint observations of radar reflectivity and ground rainfall passing the 'minimum precipitation' filter. Each data pair is shown as a blue dot in Fig. 6, and we can see that there exists a strong, positive and non-linear relation between them. We already made use of this relation in Experiment 1 when we built a single-
10 predictor model using reflectivity to estimate ground rainfall. Comparing cases 2 and 3 in Table 2 we can see that prior knowledge of reflectivity indeed contains valuable I information, reducing total uncertainty by 15.3 % (100 % to 84.7 %).



15 **Fig. 6.** Relation between target RR0 and predictor dBZ1500Rad. Empirical relation as given by the filtered data set (blue dots). Deterministic power-law relation according to Marshall and Palmer, 1948 (red line); deterministic power-law relation with optimized parameters $a = 235$ and $b = 1.6$ (black line).

Let us suppose we would not have been in the comfortable situation to have joint observations of target and predictor to construct a data-based model, or suppose it would take too much storage or computational resources to either store or apply such a model. In such cases, it could be reasonable to approximate the 'scattered' relation as contained in the data either by a

deterministic function gained from other data or a deterministic function fitted to the data. In fact this is standard practice. Expressing a data-relation by a function drastically reduces storage space, is easy to apply, and preserves the overall relation among the data. However, what we lose is **I**nformation about the strength of that relation as expressed by the scatter of the data. Instead, when applying a deterministic function we claim that the predictive uncertainty of the target is zero.

5 | Our aim here is to quantify the **I**nformation loss associated with such deterministic functional compression. Applying a deterministic model is in principle no different than using a model learned from a subset of the data as we did in Experiment 2: We use an imperfect model, and this will result in *additional* uncertainty which we can measure via Kullback-Leibler divergence between the true data relation and the model we apply (see Eq. 6). Total predictive uncertainty is then measured
10 | by Cross Entropy (Eq. 5) as the sum of Conditional Entropy (the uncertainty due to incomplete **I**nformation of the predictor about the target) and Kullback-Leibler divergence (see section 2.2).

For demonstration purposes we applied two typical deterministic Z-R relations of the form $Z = a \cdot R^b$. The first is the widely used Marshall-Palmer relation (mostly attributed to Marshall and Palmer, 1948 as for example in Battan 1959b and Sauvageot 1992, but where this particular formula actually comes from is not known to the authors) with parameters $a = 200$ and $b = 1.6$. The Marshall-Palmer relation is often used as a default model if no better options or local data are available. For
15 | the second, we assumed the local reflectivity and rain rate observations to be available and used them to optimize parameter a by minimizing the root mean square error (RMSE) between the observed rain rate and the modelled rain rate using reflectivity observations. Following the recommendations by Hagen (2003) we only varied a and kept b constant at 1.6. In the end, the optimized $a = 235$ was not far from its Marshall-Palmer pendant and reduced RMSE only slightly from 1.18 mm/h (Marshall-Palmer relation) to 1.17 mm/h. Apparently, the default Marshall-Palmer relation already nicely fit our data.

20 | In Fig. 6, the two deterministic Z-R functions are plotted as red (Marshall-Palmer) and black (optimized Z-R relation) lines. Both capture the overall shape of the empirical Z-R relation quite well, except for high reflectivities where they tend to underestimate the observed rain rates. Applying such a deterministic Z-R relation to a given reflectivity observation is straightforward and yields a prediction of the related rain rate. However, from such a single-valued prediction we cannot infer the related predictive uncertainty, and the best we can do is to additionally provide the model's RMSE as a proxy.

25 | As described above, we additionally used our **I**nformation-based approach and calculated the conditional Kullback-Leibler divergence between the predictive distributions given by the empirical (= perfect) and the deterministic (= imperfect) models for all available data. From Eq. 6 we see that when comparing a reference distribution p to a model distribution q , a situation can occur where the latter is zero but the former is not, which means that an event is contained in the reference data that, according to the model, can never occur. In such a case Kullback-Leibler divergence will be infinite, branding the model as
30 | completely inadequate. However, this verdict may seem too hard, e.g. if the model shows otherwise good agreement with reality, if we have reason to believe that the mismatch only occurred from a lack of opportunity due to a small data set rather than due to a principal mismatch of model and reference or, as in our case, if we know that we use a deterministic approximation instead of the real data relation. In such cases, infinite divergence can be avoided by padding empty bins of

the model distribution with small but non-zero probabilities, which we did by applying the minimally invasive maximum Entropy approach suggested by Darscheid et al. (2018).

For our data, Conditional Kullback-Leibler divergence for the Marshall-Palmer model was 3.43 bit, and 2.69 bit for the optimized model. Added to the Conditional Entropy of the empirical Z-R relation, this resulted in total predictive uncertainties of 5.04 and 4.30 bit, respectively. In terms of relative contributions, this means that 68% (Marshall-Palmer) and 62.5% (optimized model) of total uncertainty are due to deterministic functional compression. This is quite considerable, and even more so if we compare these results to the two benchmark cases in Table 2 (cases 1 and 2): Even the default and safe-side model of applying a uniform distribution (case 1) involves smaller predictive uncertainty than the deterministic models. This seems counter-intuitive at first when recalling the good visual agreement of the empirical and deterministic Z-R relations in Fig. 6. The reason for such large Kullback-Leibler divergences are the relatively high binning resolution for rain rate (Table 1) in combination with the way Kullback-Leibler divergence is computed: Probability differences are calculated bin by bin and irrespective of probabilities in neighbour bins. This means that even a small over- or underestimation of a model, with an offset of the main probability mass by just one bin compared to observations results in large divergence, which can in the end even exceed that of a prudent model spreading probability mass evenly over the data range. For the data used in this experiment, we assume that the agreement between deterministic model predictions and observations would quickly increase when coarse-graining the binning. This would be an interesting question to pursue in a future study, for now we restrict ourselves to the main conclusion of this experiment: As long as learning about and application of data-relations is carried out in the same data set, compression of probabilistic data-relationships to deterministic functions will invariably increase uncertainty about the target. However for the forward case, i.e. cases where there are no data available for learning but predictions are nevertheless required, application of robust deterministic relations capturing the essential relation between available predictors and the target is nevertheless useful.

3.5 Experiment 5: Information gains along the radar path

In this experiment we explore how the **I**nformation content about ground rainfall in reflectivity observations is related to the measurement position along a vertical profile above the rain gauge, and how it is related to the measurement device. To this end, we used data from two sites, Petit-Nobressart and Useldange (see Fig. 2), where a range of reflectivity observations along a 1500 m vertical profile starting at ground level was available: disdrometer observations of reflectivity and rain rate at ground level, MRR reflectivity observations in 100 meter resolution between 100 and 1500 m above ground, and observations from C-band weather radar at 1500 m. With the goal in mind to provide guidance for the layout of future QPE sensor networks we addressed the following questions: 'Are MRR observations at 1500 m more informative than weather radar observations taken at the same height?'; 'How much **I**nformation is gained if we use near-surface instead of elevated MRR observations, thus omitting the influence of the vertical profile of reflectivity?'; and finally 'How much **I**nformation is lost if we measure reflectivity instead of rainfall at ground level?'

Differently from the previous experiments, we were restricted to two (instead of eight) sites with MRR's installed. This made application of the standard 'minimum precipitation' filter (see section 2.3) inappropriate. So instead of filtering the raw data by at least two rain gauges showing rainfall ≥ 0.5 mm/h, we now applied this threshold separately to each of the two sites. At Petit-Nobressart altogether 1241 data tuples passed the filter, at Useldange 612 (Useldange was put into operation a year after Petit-Nobressart).

As in Experiment 1, we used Entropy and Conditional Entropy to measure the information content of the available predictors and added the Entropy of both a uniform and the observed distribution of ground rainfall as benchmarks. The results are shown in Table 3. As the results for the two stations are similar, we mainly discuss Petit-Nobressart in the following, moving from the remotest and presumably least informative predictor to the closest.

Table 3. Separately for sites Petit-Nobressart and Useldange: Entropy of the target (RR0) and the benchmark distribution (RR0 uniform), Conditional Entropies of the target given various predictors along the radar path, non-normalized and normalized. Underlying data are from the filtered data set. Predictors are ordered by decreasing distance to the target.

case	Predictor(s)	Location Petit-Nobressart		Location Useldange	
		(Conditional) Entropy H(X), H(X Y) [bit]	Relative *** (Conditional) Entropy [%]	(Conditional) Entropy H(X), H(X Y) [bit]	Relative *** (Conditional) Entropy [%]
1	RR0 uniform	4.17*	242	4.17*	250
2	RR0	1.72**	100	1.67**	100
3	dBZ1500Rad	1.45	84.3	1.41	84.4
4	dBZ1500MRR	1.48	86.0	1.37	82.0
5	dBZ100MRR	1.27	73.8	1.30	77.8
6	dBZ0Dis	1.14	66.3	1.14	68.3

* unconditional Entropy of the benchmark uniform distribution

** unconditional Entropy of the target

*** $H(\text{RR0}|\text{predictor}) / H(\text{RR0}) * 100$

Weather radar data, even if they are measured at far range and at height contain considerable information about ground rainfall: Using them as predictors reduced uncertainty by 15.7 % (100% to 84.3%) compared to the benchmark Entropy of

the target distribution (Table 3, cases 2 and 3). This is comparable to the outcomes of Experiment 1 based on all sites (15.3%, Table 2, same cases). We expected MRR observations taken at the same site and elevation to be more informative than their weather radar pendant, because the MRR signal path and with it the potential for signal corruption is considerably shorter. This is however not clearly evident from the results: At Useldange, using MRR instead of weather radar data additionally reduced uncertainty by only 2.4% (84.4% to 82.0%, Table 3, cases 3 and 4), and at Petit-Nobressart they were even less informative (86.0% instead of 84.4%, Table 3, same cases). These results should be interpreted with some care due to the relatively limited data base, however it seems safe to conclude that there is no large difference in the information content of MRR and weather radar observations taken at height.

So why take the extra trouble to operate an MRR? Its benefit shows when moving down to elevations inaccessible by weather radar. Changes in drop size distribution along the pathway of rainfall from cloud to ground can be considerable, and the closer to the ground the observation is taken the stronger its relation to ground rainfall: Using MRR data from the downmost bin (≤ 100 m above ground) reduced uncertainty by 12.2% compared to using the uppermost bin (86.0% to 73.8%, Table 3, cases 4 and 5). This underlines the importance of VPR correction when using weather radar data (Vignal et al. 1999, Vignal et al. 2000). Based on these results, an obvious next step would be to derive VPR corrections from the two MRR's and investigate the information gain when applied to other sites in the domain; however for the sake of brevity we leave this for future studies.

Further information gains can be achieved when measuring reflectivity directly at the ground: Using the disdrometer measurements further reduced uncertainty by 7.5% (73.8% to 66.3%, Table 3, cases 5 and 6), compared to the MRR observations. This gain is not only due to using ground observations, which completely excludes any negative VPR effects, but also because predictor and target data were observed at the same spot and by the same sensor. Seen from this perspective, it is even surprising that nevertheless considerable uncertainty remains (66.3% of the benchmark uncertainty, Table 3, cases 6 and 2) which must be attributed to the ambiguous relationship between radar reflectivity and rain rate ~~as measured by the disdrometer~~ due to the natural variability of drop size distribution.

3.6 Experiment 6: QPE based on Radar and rain gauge data

In this final experiment we compare two methods of QPE, radar-based and rain gauge-based and additionally explore the benefits of jointly tapping both sources of information. As we were not dependent on the availability of MRR data as in the previous experiment, we could again make use of the full eight-site set of observations filtered with the 'minimum precipitation' filter.

We used the same data-based approach of constructing empirical dpd's as predictive model like in all previous experiments; the only difference between the three tested QPE models was the type of predictor used: For the radar-based QPE, we used weather radar observations at 1500 m above ground in a single-predictor model to predict ground rainfall at the same site. This is the same approach as applied in Experiment 1. For the rain gauge-based QPE, we built a single-predictor model based on a straightforward approach comparable to leave-one-out cross validation: Each of the eight available stations (see

Fig. 2 and Table 1) was once used as a target station, and observations from each of the remaining seven stations were separately used as predictors. This way we could calculate the Information content in the predictor as a function of the distance between stations. Eight stations render altogether 56 unique station pairings, for our stations the minimum, average, and maximum distances were 1.9, 7.2, and 15.3 km, respectively. For later plotting, we binned the results in seven distance classes of two km width each and took averages within each bin. This is comparable to using range bins when calculating a semivariogram.

For the joint QPE model we extended the approach used for rain gauge interpolation to a double-predictor model: Again each of the eight available stations was once used as a target station, but now not only observations from each of the remaining stations were (separately) used as predictors, but additionally radar observations measured at height above the target station. This is a typical approach when merging radar and rain gauge data for QPE: We use data from rain gauges observed at a horizontal, and radar data observed at a vertical distance from the target.

As we build and compared in this experiment models with different numbers of predictors, and as the models for each particular target station were built from a subset of the data only, it could be worthwhile to explore the additional uncertainty due to the effect of sample size here as in Experiments 2 and 3. However, for this particular experiment we found it more useful for the reader (and us) to discuss results as a function of distance between stations, as it provides a link to the large body of literature on spatial rainfall structure analysis and station-based rainfall interpolation. The results are shown in Fig. 7.

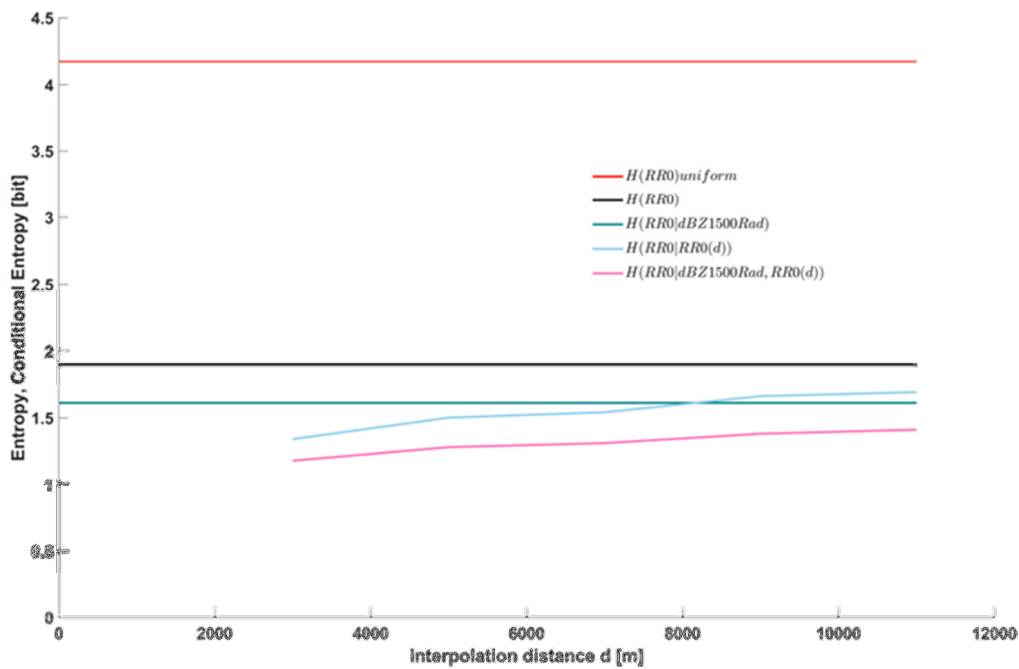


Fig. 7. Entropy of the unconditional, target distribution (filtered data set, black line, same as in Figure 4), Entropy of the benchmark uniform distribution (filtered data set, red line, same as in Figure 4), Conditional Entropy of the target given reflectivity as predictor (green line), Conditional Entropy of the target given station rain rate observations as a function of interpolation distance (light blue line), Conditional Entropy of the target given reflectivity and rain rate at stations as predictor, as a function of interpolation distance (pink line). Underlying data are from the filtered data set.

Just as in Fig. 4 and Fig. 5, we included the benchmark unconditional Entropies of the target for a maximum Entropy uniform distribution (red line) and the observed distribution of all stations put together (black line) to put the results into perspective. As the radar-only model (green line) is independent of any interpolation distance, simply because it does not make use of any station data, it plots as a horizontal line. Its conditional Entropy (1.61 bit) corresponds to the right-end (= full-size sample) value of the green line in Fig. 4, and to case 3 in Table 2.

For the QPE model based on rain gauge observations (light blue line), interpolation distance however does play a role: As to be expected, the smaller the distance between the target and the predictor station, the higher the information content of the predictor and the smaller Conditional Entropy, i.e. the blue lines rises from left to right. For small distances between two and four km, Conditional Entropy is 1.34 bit, which is lower than for the radar-only QPE. If we take the unconditional Entropy of the target again as a reference as in Experiment 1 (Table 1, case 2), station-only QPE for short distances reduces uncertainty to 70.5%, radar-only to 84.7%. For large distances however this order is reversed, and station-only QPE reduces uncertainty only to 1.69 bit or 88.9% of the reference. The break-even point at which both methods perform equally well lies at a station distance of about 8 km. This means that if we were asked to choose one of the two models for QPE, the best choice would be to use station interpolation for all targets within less than about 8 km from the nearest station, and radar-QPE for all other.

The above results indicate that each of the two QPE methods has its particular strengths. In other words they add non-redundant information, and therefore we can expect some benefits when joining them in a two-predictor model. This is indeed the case if we take a look at the related Conditional Entropies plotted as pink line in Fig. 7: While it resembles the plot of the station-only model in overall shape and trend, it is shifted downward: Conditional Entropy for small distances is now reduced to only 1.18 bit (62.1% of the reference), and even for the largest distances (1.41 bit = 74.2% of the reference) it is still smaller than both the station-only and the radar-only values. Interestingly, in the combined model the radar observations add information even for small station distances. We assume that this can be attributed to convective rainfall situations, where strong spatial rainfall heterogeneity occurs, unresolvable even by rain gauges only a few kilometres apart.

In Experiment 1 we also built and tested two-predictor models. The best among them, using radar data and tenner-days of the year for predictors (Table 1, case 11), reduced Conditional Entropy to 77.4% of the reference. While this is a considerable improvement, it is surpassed by the radar-rain gauge model even for the largest station distance. Obviously, ground rainfall and radar observation contain non-redundant information which, used together, have a large potential to improve QPE.

4 Summary and conclusions

Reliable QPE is an important prerequisite for many hydro-meteorological design and management purposes. In this context we pursued two aims with this paper: The first was to suggest and demonstrate a probabilistic framework based on concepts of Information Theory, in which predictive relations are expressed by empirical discrete probability distributions directly derived from data. The framework allows integration of any kind of data deemed useful and explicitly acknowledges the uncertain nature of QPE. The second aim was to investigate the information gains and losses associated with various data and practices typically applied in QPE. For this purpose we conducted altogether six experiments using a comprehensive set of data comprising four years of hourly-aggregated observations from weather radar, vertical radar (MRR), disdrometers, rain gauges, and a range of operationally available hydro-meteorological observables such as large-scale Circulation Patterns, ground meteorological variables and season indicators.

In *Experiment 1*, we measured the information about ground rainfall contained in various operationally available predictors with Entropy and Conditional Entropy. Weather radar proved to be the single-most important source of information, which could be further improved by distinguishing Z-R relationships by season and prevailing circulation pattern.

In *Experiment 2*, we tested the robustness of QPE models developed in the previous experiment by measuring the additional uncertainty due to limited learning data with Kullback-Leibler divergence and Cross Entropy. The main lesson learned here was that this added uncertainty is strongly dependent on the number of predictors used in the model, and that for unfavourable constellations (multiple-predictor models learning from small samples) this effect quickly dominates total uncertainty: This is the well-known Curse of Dimensionality. For the data set used in this study, we found a double-predictor model using a month-specific relation between radar reflectivity and ground rainfall to provide the best trade-off between performance and robustness.

In *Experiment 3*, we examined to which degree the empirical Z-R relationship in the test domain varies in space, and what size data sets are required to support the use of site-specific Z-R relations. Overall, the information gain of site-specific Z-R relations was smaller than distinguishing them according to season or circulation pattern as we did in the previous experiments. Moreover, applying site-specific Z-R relations instead of a single, site-independent one means adding another predictor to the QPE model, which in turn means that we increase the Curse of Dimensionality. In fact, more than half of the available data had to be used for learning the site-specific model before the additional effort paid off. The conclusion here was to stick to a site-independent relation, which also seems reasonable from a meteorological point of view due to the relatively small extent of the test domain (288km²).

In *Experiment 4*, we evaluated the effect of functional compression, i.e. the information loss when replacing the data-inherent probabilistic relationship between reflectivity and rain rate by a deterministic functional approximation. We measured the additional uncertainty with Kullback-Leibler divergence, which adds to the uncertainty caused by the incomplete information of the predictor about the target measured by Conditional Entropy. We used two standard power-law Z-R relations: The well-known Marshall-Palmer relation, and a modified version thereof with optimized function

parameters. For both, the added uncertainty was in the order of 60% of total uncertainty. While these numbers may vary with the chosen binning resolution, compression of probabilistic data-relationships to deterministic functions will invariably increase uncertainty as long as learning and applying of data-relations is carried out in the same data set.

5 In *Experiment 5*, we investigated Information gains along the radar path using weather radar, MRR and disdrometer data from two sites. The main insights here were that the Information content of radar reflectivity measured at height (1500 m above ground) does not differ much between weather radar and MRR, but considerable additional Information is gained by using observations from lower elevations (100 m above ground), thus avoiding Information losses by changes in drop size spectra along the pathway of precipitation from cloud to ground. This underlines both the importance of VPR corrections for accurate QPE and of the required MRR observations. Despite these Information gains, considerable uncertainty remained 10 even when using as predictor reflectivity observations taken at the same spot and by the same device (disdrometer) as the target variable (rain rate) itself. This indicates a somewhat ambiguous relationship between radar reflectivity and rain rate as measured by the disdrometer, which could potentially be sharpened by taking into account precipitation type.

In the last *Experiment 6*, we built QPE models based on radar data only, rain gauge data only and a combination thereof and evaluated their Information content as a function of the distance between the target and predictor rain gauge. Comparing the 15 first two revealed that a separation distance of ~ 8 km marks a transition: For smaller distances, gauge-based QPE is superior, for larger distances radar-based QPE. Combining the two in a double-predictor QPE model however not only outperformed both of them for all separation distances, but also competing radar-based double-predictor models applying season or circulation pattern. For rain gauge networks comparable in coverage and density to the one used in this study, rain gauge and radar data apparently contain useful, non-redundant Information which should be jointly exploited.

20 We would like to emphasize that the results from these experiments are partly contingent on the choice of the data filter: The 'minimum precipitation' (at least two rain gauges with rainfall ≥ 0.5 mm/h) filter we applied excluded all cases where rainfall occurred at only a single station, and it also excluded all cases of widespread but very low-intensive rain.

Quantitative statements about the Information content of particular predictors or the relative performance of competing QPE models may differ according to these user choices. However, we would like to point out that all of the different experiments 25 in this study could be formulated, conducted and evaluated in a single framework and relied on a single property, Information. It is therefore not so much the particular results we want to emphasize here, but rather the probabilistic, data- and Information based framework we applied. By its probabilistic concept, it explicitly acknowledges the uncertain nature of QPE, by expressing probabilities in terms of Information it facilitates both interpretation and computation. In the framework, predictive relationships are directly derived from data and expressed as discrete probability distributions. The 30 advantage is that this avoids as much as possible introduction of unwanted side Information, e.g. by parametric choices, and that it avoids deletion of existing Information e.g. by data transformation or lossy compression. Altogether, this facilitates tracking sources and sinks of Information. These advantages however come at a price: Learning robust data-based relations requires a considerable amount of available target and predictor data, and applying them for predictions is computationally more expensive than using deterministic functions.

Data availability: The reflectivity and rain rate data measured by the six disdrometers and two MRR's is published via the GFZ Data Service repository (Neuper and Ehret 2018). The rain gauge data of the station at Roodt and the rain gauge data, the 2 m temperature data, the relative humidity, and the wind data of the station at Useldange is publicly available from the Administration des services techniques de l'agriculture ASTA at <http://www.agrimeteo.lu/> (last access: 04. December 2018). The rain gauge data of the station at Reichlange was provided upon request by the Hydrometry Service Luxembourg (<https://www.inondations.lu/> ; last access: 04. December 2018). The Sounding Data from which CAPE was calculated is publicly available from the University of Wyoming at <http://weather.uwyo.edu/upperair/sounding.html> (last access: 04. December 2018). The data of the objective classification of the prevailing circulation pattern over Europe Model is publicly available from the German weather service at <https://www.dwd.de/DE/leistungen/wetterlagenklassifikation/wetterlagenklassifikation.html> (last access: 04. December 2018). The raw reflectivity volume radar data from the single pol C-Band Doppler radar located in Neuheilenbach is available upon request from the German Weather Service (DWD). The code used to conduct all analyses in this paper is publicly available at <https://github.com/KIT-HYD/Radar-based-QPE-using-information-measures>.

15

Author contribution UE and MN directly contributed to the conception and design of the work, collected and processed the various data sets and conducted a thorough data quality control, did the data analysis and interpretation and wrote the manuscript.

20 *Acknowledgements.* We acknowledge support by Deutsche Forschungsgemeinschaft DFG and Open Access Publishing Fund of Karlsruhe Institute of Technology (KIT). This research contributes to the "Catchments As Organized Systems" (CAOS) research group funded by the Deutsche Forschungsgemeinschaft DFG.

We gratefully acknowledge the fruitful cooperation and joint operation of the meteorological observation network within CAOS with Hans-Stefan Bauer and Volker Wulfmeyer.

25 We gratefully acknowledge the providers of meteorological data used in this study: The German Weather Service DWD for radar and classified circulation pattern data, the Administration des services techniques de l'agriculture ASTA for rain gauge and climate data, and the Hydrometry Service Luxembourg for rain gauge data.

Further on, we especially thank Jean-François Iffly from the Luxembourg Institute of Science and Technology (LIST) for his extensive commitment finding the right locations to put up the instruments and his outstanding help in maintaining them. We further thank Kirsten Elger from GFZ Helmholtz Centre Potsdam for her patient help with the data publication at the GFZ repository, our colleagues Ralf Loritz, Sibylle Haßler, Conrad Jackisch and Erwin Zehe for fruitful discussions on data analysis and Jan Handwerker from the KIT Institute of Meteorology and Climate Research for help concerning radar meteorology issues. Last but not least we thank the rest of the CAOS team for the warm-hearted project meetings and discussions.

Competing interests: The authors declare that they have no conflict of interest.

References

- Adirosi, E., Volpi, E., Lombardo, F., and Baldini, L.: Raindrop size distribution: Fitting performance of common theoretical models, *Advances in Water Resources*, 96, 290-305, <https://doi.org/10.1016/j.advwatres.2016.07.010>, 2016.
- Atlas, D., Srivastava, R., and Sekhom, R.: Doppler radar characteristics of precipitation at vertical incidence. *Rev. Geophys. Space Phys.*, 11, 1-35, <https://doi.org/10.1029/RG011i001p00001>, 1973.
- Battan, L. J.: *Radar Meteorology*, The University of Chicago Press, Chicago, 1959a.
- Battan, L. J.: *Radar Observations of the Atmosphere*, The University of Chicago Press, Chicago, 1959b.
- 10 Bauer, H.-S., Schwitalla, T., Wulfmeyer, V., Bakhshaii, A., Ehret, U., Neuper, M., and Caumont, O.: Quantitative precipitation estimation based on high-resolution numerical weather prediction and data assimilation with WRF - a performance test, 2015, <https://doi.org/10.3402/tellusa.v67.25047>, 2015.
- ~~Bellman, R. E.: *Dynamic Programming*, Princeton University Press, Princeton NJ, 1957.~~
- 15 Bronstert, A., Agarwal, A., Boessenkool, B., Fischer, M., Heistermann, M., Kohn-Reich, L., Moran, T., and Wendi, D.: The Braunsbach Flashflood of Mai 29th, 2016 – Origin, Pathways and Impacts of an Extreme Hydro-Meteorological Event. Part 1: Meteorological and Hydrological Analysis, *Hydrologie und Wasserbewirtschaftung*, 61, 150-162, https://doi.org/10.5675/HyWa_2017,3_1, 2017.
- Brunsell, N.: A multiscale information theory approach to assess spatial–temporal variability of daily precipitation, 165-172 pp., 2010.
- 20 Bückle, J: *Korrektur von Second Trip Echos in Radardaten*, Diploma thesis, Institute of Meteorology and Climate Research, University of Karlsruhe, 2010.
- Bunkers, M. J., Johnson, J. S., Czepyha, L. J., Grzywacz, J. M., Klimowski, B. A., and Hjelmfelt, M. R., An observational examination of long-lived supercells. Part II: Environmental conditions and forecasting. *Wea. Forecasting*, 21(5), 689-714, <https://doi.org/10.1175/WAF952.1>, 2006
- 25 Cao, Q., Zhang, G. F., Brandes, E. A., and Schuur, T. J.: Polarimetric radar rain estimation through retrieval of drop size distribution using a bayesian approach, *J. Appl. Meteorol. Climatol.*, 49, 973-990, <https://doi.org/10.1175/2009jamc2227.1>, 2010.
- Cecinati, F., de Niet, A. C., Sawicka, K., and Rico-Ramirez, M. A.: Optimal temporal resolution of rainfall for urban applications and uncertainty propagation, *Water*, 9, <https://doi.org/10.3390/w9100762>, 2017^a.
- 30 Cecinati, F., Rico-Ramirez, M. A., Heuvelink, G. B. M., and Han, D.: Representing radar rainfall uncertainty with ensembles based on a time-variant geostatistical error modelling approach, *Journal of Hydrology*, 548, 391-405, <https://doi.org/10.1016/j.jhydrol.2017.02.053>, 2017^b.

- Cover, T. and Thomas, J.: Elements of information Theory. John Wiley & Sons, New York, 1991.
- Darscheid, P., Guthke, A., and Ehret, U.: A maximum-entropy method to estimate discrete distributions from samples ensuring nonzero probabilities, *Entropy*, 20, 601, DOI: 10.5445/IR/1000085954, 2018.
- Diem, M.: Zur Struktur der Niederschläge III – Regen in der arktischen, gemäßigten und tropischen Zone, *Arch. Met. Geoph. Biokl., Ser. B*, 16, 347-390, <https://doi.org/10.1007/BF02243180>, 1968.
- 5 Dittmann, E., Barth, S., Lang, J. and Müller-Westermeier, G.: Objektive Wetterlagenklassifikation, *Berichte des Deutschen Wetterdienstes (DWD)*, 197, Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, 1995.
- Dotzek, N. and Beheng, K. D.: The influence of deep convective motions on the variability of Z–R relations, *Atmos. Res.*, 59, 15-39, [https://doi.org/10.1016/S0169-8095\(01\)00107-7](https://doi.org/10.1016/S0169-8095(01)00107-7), 2001.
- 10 Doviak, R. J. and Zrnich, D. S.: Doppler Radar and Weather Observations. Academic Press, San Diego, 1993.
- Ehret, U., Gupta, H. V., Sivapalan, M., Weijers, S. V., Schymanski, S. J., Blöschl, G., Gelfan, A. N., Harman, C., Kleidon, A., Bogaard, T. A., Wang, D., Wagener, T., Scherer, U., Zehe, E., Bierkens, M. F. P., Di Baldassarre, G., Parajka, J., van Beek, L. P. H., van Griensven, A., Westhoff, M. C., and Winsemius, H. C.: Advancing catchment hydrology to deal with predictions under change, *Hydrol. Earth Syst. Sci.*, 18, 649-671, <https://doi.org/10.5194/hess-18-649-2014>, 2014.
- 15 Faber, R: Climatologie du Grand-Duché de Luxembourg. - Publication du Musée national d'histoire naturelle et de la Société des naturalistes luxembourgeois, Graphic Center Bourg-Bourger, Luxembourg, 1971.
- Fabry, F.: Radar-meteorology: Principles and practice, Cambridge University Press, 2015.
- Friedrich, K., Higgins, S., Masters, F. J., and Lopez, C. R.: Articulating and stationary PARSIVEL disdrometer measurements in conditions with strong winds and heavy rainfall, *J. Atmos. Oceanic Technol.*, 30(9), 2063-2080, <https://doi.org/10.1175/JTECH-D-12-00254.1>, 2013a.
- 20 Friedrich, K., Kalina, E. A., Masters, F. J., and Lopez, C. R.: Drop-size distributions in thunderstorms measured by optical disdrometers during VORTEX2. *Mon. Wea. Rev.*, 141(4), 1182-1203, <https://doi.org/10.1175/MWR-D-12-00116.1>, 2013b.
- Germann, U., Galli, G., Boscacci, M., and Bolliger, M.: Radar precipitation measurement in a mountainous region, *Quarterly Journal of the Royal Meteorological Society*, 132, 1669-1692, <https://doi.org/10.1256/qj.05.190>, 2006.
- 25 Gong, W., Yang, D., Gupta, H. V., and Nearing, G.: Estimating information entropy for hydrological data: One-dimensional case, *Water Resources Research*, 50, 5003-5018, <https://doi.org/10.1002/2014wr015874>, 2014.
- Goudenhoofdt, E., Delobbe, L., and Willems, P.: Regional frequency analysis of extreme rainfall in Belgium based on radar estimates, *Hydrol. Earth Syst. Sci.*, 21, 5385-5399, <https://doi.org/10.5194/hess-21-5385-2017>, 2017.
- Goudenhoofdt, E., and Delobbe, L.: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, *Hydrology and Earth System Sciences*, 13, 195-203, <https://doi.org/10.5194/hess-13-195-2009>, 2009.
- 30 Hagen, M., and Yuter, S. E.: Relations between radar reflectivity, liquid-water content, and rainfall rate during the map sop, *Quarterly Journal of the Royal Meteorological Society*, 129, 477-493, <https://doi.org/10.1256/qj.02.23>, 2003.
- Hannesen, R.: Analyse konvektiver Niederschlagssysteme mit einem C-Band Dopplerradar in orographisch gegliedertem Gelände. Dissertation, Institute of Meteorology and Climate Research, University of Karlsruhe, 1998.

- Huff, F.A.: Sampling Errors in Measurement of Mean Precipitation, *J. Appl. Meteor.*, 9, 35-44, [https://doi.org/10.1175/1520-0450\(1970\)009<0035:SEIMOM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0035:SEIMOM>2.0.CO;2), 1970.
- [Joss, J. and Waldvogel, A., Precipitation measurement and hydrology. In: Atlas D. \(eds\) Radar in Meteorology. Amer. Meteor. Soc., 577-606, 1990](#)
- 5 Kirstetter, P.E., Gourley, J.J., Hong, Y., Zhang, J., Moazamigoodarzi S., Langston C. and Arthur, A.: Probabilistic Precipitation Rate Estimates with Ground-based Radar Networks, *Water Resources Research*, 51, 1422–1442, <https://doi.org/10.1002/2014WR015672>, 2015.
- Knuth, K.: Optimal Data-Based Binning for Histograms. arXiv:physics/0605197v2 [physics.data-an], 2013.
- Köppen, W. and Geiger, R.: *Handbuch der Klimatologie: Band 1, Teil C: Das geographische System der Klimate*, Gebrüder
10 Borntraeger, Berlin, 1930.
- Kronenberg, R., Franke, J., and Bernhofer, C.: Classification of daily precipitation patterns on the basis of radar-derived precipitation rates for Saxony, Germany, *Meteorologische Zeitschrift* 21(5):475-486, <https://doi.org/10.1127/0941-2948/2012/0343>, 2012.
- Kullback, S., and Leibler, R. A. (1951): On information and sufficiency, *Ann. Math. Stat.*, 22(1), 79-86, 1951.
- 15 Lee, G., and Zawadzki, I.: Variability of drop size distributions: Time-scale dependence of the variability and its effects on rain estimation, *J. Appl. Meteorol.*, 44, 241-255, <https://doi.org/10.1175/JAM2183.1>, 2005.
- Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E.: On the dynamic nature of hydrological similarity, *Hydrol. Earth Syst. Sci.*, 22, 3663-3684, <https://doi.org/10.5194/hess-22-3663-2018>, 2018.
- Löffler-Mang, M., Kunz, M., and Schmid, W.: On the performance of a low-cost k-band doppler radar for quantitative rain
20 measurements, *Journal of Atmospheric and Oceanic Technology*, 16, 379-387, [https://doi.org/10.1175/1520-0426\(1999\)016<0379:otpoal>2.0.co;2](https://doi.org/10.1175/1520-0426(1999)016<0379:otpoal>2.0.co;2), 1999.
- Löffler-Mang, M., and Joss, J.: An optical disdrometer for measuring size and velocity of hydrometeors, *Journal of Atmospheric and Oceanic Technology*, 17, 130-139, [https://doi.org/10.1175/1520-0426\(2000\)017<0130:aodfms>2.0.co;2](https://doi.org/10.1175/1520-0426(2000)017<0130:aodfms>2.0.co;2), 2000.
- 25 Marshall, J. S., and Palmer, W. M.: The distribution of raindrops with size, *Journal of Meteorology*, 5, 165-166, 1948.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijis, S. V.: A philosophical basis for hydrological uncertainty, *Hydrological Sciences Journal*, 61, 1666-1678, <https://doi.org/10.1080/02626667.2016.1183009>, 2016.
- Nešpor, V., and Sevruk, B.: Estimation of Wind-Induced Error of Rainfall Gauge Measurements Using a Numerical
30 Simulation. *J. Atmos Ocean.Tech.*, 16, 450–464, [https://doi.org/10.1175/1520-0426\(1999\)016<0450:EOWIEO>2.0.CO;2](https://doi.org/10.1175/1520-0426(1999)016<0450:EOWIEO>2.0.CO;2), 1999.
- Neuper, M.: Anomalous propagation: Examination of ducting conditions and anaprop events in SW-Germany. Seminar work, Institute of Meteorology and Climate Research, University of Karlsruhe, 2009.

- Neuper, M. and Ehret, U.: CAOS rain rate and reflectivity data set of 6 disdrometers and 2 micro rain radars at 3 different heights at 6 stations in the Atert catchment, Luxembourg from Oct 2012 - Sept 2016. GFZ German Research Centre for Geosciences. <http://doi.org/10.5880/FIDGEO.2018.070>, 2018
- 5 Nystuen, J.: Relative Performance of Automatic Rain Gauges under Different Rainfall Conditions. *J. Atmos Ocean.Tech.*, 16, 1025–1043, [https://doi.org/10.1175/1520-0426\(1999\)016<1025:RPOARG>2.0.CO;2](https://doi.org/10.1175/1520-0426(1999)016<1025:RPOARG>2.0.CO;2), 1999.
- Overeem, A., Holleman, I., and Buishand, A.: Derivation of a 10-year radar-based climatology of rainfall, *J. Appl. Meteorol. Climatol.*, 48, 1448-1463, <https://doi.org/10.1175/2009jamc1954.1>, 2009.
- Pechlivanidis, I. G., Jackson, B., McMillan, H., and Gupta, H. V.: Robust informational entropy-based descriptors of flow in catchment hydrology, *Hydrological Sciences Journal*, 61, 1-18, <https://doi.org/10.1080/02626667.2014.983516>, 2016.
- 10 Peters, G., Fischer, B., and Andersson, T.: Rain observations with a vertically looking Micro Rain Radar (MRR). *Boreal Env. Res.* 7: 353–362, 2002.
- [Peters, G., Münster, H., Clemens, M., and Wagner, A.: Profiles of raindrop size distributions as retrieved by microrain radars. *J. Appl. Meteor.*, 44, 1930–1949, https://doi.org/10.1175/JAM2316.1., 2005](https://doi.org/10.1175/JAM2316.1)
- 15 Pfister, L., Wagner, C., Vansuypeene, E., Drogue, G. and Hoffmann, L.: In C. Ries (Ed.), *Atlas climatique du grand-duché de Luxembourg*, Luxembourg: Musée National d'Histoire Naturelle, Société des naturalistes luxembourgeois, Centre de Recherche Public — Gabriel Lippmann, Administration des Services Techniques de l'Agriculture, 2005.
- Pfister, L., Humbert, J. and Hoffmann, L.: Recent trends in rainfall – runoff characteristics in the Alzette River basin, Luxembourg, *Climatic Change*, 45, 323–337, <https://doi.org/10.1023/A:1005567808533>, 2000.
- Rauber, R. M. and Nesbitt, S. L.: *Radar Meteorology: A First Course*, John Wiley & Sons, Hoboken NJ, 2018
- 20 Rinehart, R.: *Radar for Meteorologists*. Univ. North Dakota, Grand Forks, 58206-6124, US, 1991.
- Rossa, A., Bruen, M., Frühwald, D., Macpherson, B., Holleman, I., Michelson, D., and Michaelides, S.: Cost 717 use of radar observations in hydrological and nwp models, *European Co-Operation in the Field of Scientific and Technical Research*, 292, 2005.
- Ruiz-Villanueva, V., Borga, M., Zoccatelli, D., Marchi, L., Gaume, E., and Ehret, U.: Extreme flood response to short-duration convective rainfall in South-West Germany, *Hydrol. Earth Syst. Sci.*, 16, 1543-1559, <https://doi.org/10.5194/hess-16-1543-2012>, 2012.
- 25 Sauvageot, H. R.: *Radar Meteorology*. Artech House, Boston, 1992.
- Schmithüsen, J.: *Das Luxemburger Land - Landesnatur, Volkstum und bäuerliche Wirtschaft*. Verlag von S. Hirzel, Leipzig, 1940.
- 30 Seifert, A., and Beheng, K. D.: A two-moment cloud microphysics parameterization for mixed-phase clouds. Part 2: Maritime vs. continental deep convective storms. *Meteorol Atmos Phys* 92.1-2, 67-82, <https://doi.org/10.1007/s00703-005-0113-3>, 2006.
- Shannon, C. E.: A mathematical theory of communication, *Bell Syst. Tech. J.*, 27, 623–656, 1948.

- Sharma, A., and Mehrotra, R.: An information theoretic alternative to model a natural system using observational information alone, *Water Resources Research*, 50, 650-660, <https://doi.org/10.1002/2013WR013845>, 2014.
- Sing, V.: *Entropy Theory and its Application in Environmental and Water Engineering*, Wiley-Blackwell, 2013.
- Thiesen, S., Darscheid, P., and Ehret, U.: [Identifying rainfall-runoff events in discharge time series: A data-driven method based on information theory](https://doi.org/10.5194/hess-23-1015-2019), *Hydrol. Earth Syst. Sci.*, 23, 1015-1034, [10.5194/hess-23-1015-2019](https://doi.org/10.5194/hess-23-1015-2019), 2019.
- ~~Thiesen, S., Darscheid, P., and Ehret, U.: [Identifying rainfall-runoff events in discharge time series: A data-driven method based on Information Theory](https://doi.org/10.5194/hess-2018-427), *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2018-427>, in review, 2018.~~
- Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., ten Veldhuis, M. C., Arnbjerg-Nielsen, K., Rasmussen, M. R., and Molnar, P.: Weather radar rainfall data in urban hydrology, *Hydrology and Earth System Sciences*, 21, <https://doi.org/10.5194/hess-21-1359-2017>, 2017.
- Todini, E.: [Hydrological catchment modelling: Past, present and future](https://doi.org/10.5194/hess-11-468-2007), *Hydrol. Earth Syst. Sci.*, 11, 468-482, [10.5194/hess-11-468-2007](https://doi.org/10.5194/hess-11-468-2007), 2007.
- Ulbrich, C.W. and Atlas, D.: [Rainfall Microphysics and Radar Properties: Analysis Methods for Drop Size Spectra](https://doi.org/10.1175/1520-0450(1998)037<0912:RMARPA>2.0.CO;2), *J. Appl. Meteorol.*, 37, 912-923, [https://doi.org/10.1175/1520-0450\(1998\)037<0912:RMARPA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1998)037<0912:RMARPA>2.0.CO;2), 1998.
- Vasquez, T.: *Instability, Skew-T & Hodograph Handbook*, Weather Graphics Technologies, Garland TX, 2017
- Vignal, B., Galli, G., Joss, J., and Germann, U.: Three methods to determine profiles of reflectivity from volumetric radar data to correct precipitation estimates, *J. Appl. Meteorol.*, 39, 1715-1726, <https://doi.org/10.1175/1520-0450-39.10.1715>, 2000.
- Vignal, B., Andrieu, H., and Creutin, J. D.: Identification of vertical profiles of reflectivity from volume scan radar data, *J. Appl. Meteorol.*, 38, 1214-1228, [https://doi.org/10.1175/1520-0450\(1999\)038<1214:IOVPOR>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<1214:IOVPOR>2.0.CO;2), 1999.
- Villarini, G., and Krajewski, W. F.: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall, *Surveys in Geophysics*, 31, <https://doi.org/10.1007/s10712-009-9079-x>, 2010.
- Wang, L. P., Ochoa-Rodríguez, S., Onof, C., and Willems, P.: Singularity-sensitive gauge-based radar rainfall adjustment methods for urban hydrological applications, *Hydrol. Earth Syst. Sci.*, 19, 4001-4021, <https://doi.org/10.5194/hess-19-4001-2015>, 2015.
- Weijis, S.: *Information theory for risk-based water system operation*, PhD thesis, Faculty of Civil Engineering & Geosciences, section Water Resources Management, TU Delft, The Netherlands, Delft, 210 pp., 2011.
- Wellmann, J. F., and Regenauer-Lieb, K.: Uncertainties have a meaning: Information entropy as a quality measure for 3-d geological models, *Tectonophysics*, 526, 207-216, <https://doi.org/10.1016/j.tecto.2011.05.001>, 2012.
- Yang, D., Elomaa, E., Tuominen, A., Aaltonen, A., Goodson, B., Gunther, T., Golubev, V., Sevruk, B., Madsen, H., and Milkovic, J.: Wind-induced Precipitation Undercatch of Hellman Gauges., *Nordic Hydrology*, 30, 57-80, <https://doi.org/10.2166/nh.1999.0004>, 1999.

Yang, X. B., Kuang, Q. M., Zhang, W. S., and Zhang, G. P.: A terrain-based weighted random forests method for radar quantitative precipitation estimation, *Meteorological Applications*, 24, 404-414, <https://doi.org/10.1002/met.1638>, 2017.

[Zhang, G., Vivekanandan, J. and Brandes, E.: A method for estimating rain rate and Drop size Distributions from polarimetric measurements, *IEEE Transactions on Geoscience and remote sensing*, 39, 830-841, 2001.](#)

5 Zawadzki, I.: Factors affecting the precision of radar measurements of rain. Preprints, 22nd Int. Conf. on Radar Meteorology, Zurich, Switzerland, Amer. Meteor. Soc., 251–256, 1984

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: Hess opinions: From response units to functional units: A thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, *Hydrol. Earth Syst. Sci.*, 18, 4635-4655, <https://doi.org/10.5194/hess-18-4635-2014>, 2014.