Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Scientific Report MeteoSwiss No. 92

Geostatistical methods for hourly radar-gauge combination: An explorative, systematic application at MeteoSwiss

Rebekka Erdin

ISSN: 1422-1381

Scientific Report MeteoSwiss No. 92

Geostatistical methods for hourly radar-gauge combination: An explorative, systematic application at MeteoSwiss

Rebekka Erdin

Recommended citation:
Erdin, R.: 2013, Geostatistical methods for hourly radar-gauge combination: An explorative, systematic application at MeteoSwiss, *Scientific Report MeteoSwiss,* **92,** 58 pp.

Editor:
Federal Office of Meteorology and Climatology, MeteoSwiss, © 2013

MeteoSwiss
Krähbühlstrasse 58
CH-8044 Zürich
T +41 44 256 91 11
www.meteoschweiz.ch

Abstract

Gridded quantitative precipitation estimates (QPE) of high spatial and temporal resolution are increasingly required, for instance for hydrological flood forecasting and nowcasting. The combination of observations from both remote-sensing radar data and in-situ rain-gauge measurements is a popular means to produce QPEs. The idea is to incorporate into a common scheme their high spatial resolution and good local accuracy, respectively. Geostatistics provides a frequently used stochastic framework for such a combined radar-gauge gridding.

This study presents a systematically applicable approach for a real-time combination of radar and rain gauges. It is based on two popular geostatistical combination methods: kriging with external drift (KED) and ordinary kriging of radar errors. Data transformation is applied (trans-Gaussian kriging) to enhance the compliance of precipitation data with model assumptions. To deal with the often very sparse data inherent to a systematic application on short time scales, climatological parameters for the classical parametric variogram and an alternative non-parametric correlogram estimation are compared. A systematic application over a full year for the hourly time scale and the domain of Switzerland is performed. To assess the QPE characteristics at ungauged locations, cross-validation is applied. The results are analyzed with respect to both, the accuracy of the deterministic best estimate (point estimate) and the probabilistic estimate (point estimate and related uncertainty estimate).

Both combination methods show a considerable improvement compared to single sensor estimates (radar and interpolated rain gauges). And this improvement is robust throughout different geographical regions, seasons and precipitation situations. The skill of the combination varies considerably depending on the accuracy and representativeness of the radar and rain-gauge measurements. As a result there are marked seasonal and regional variations of the skill common to all methods and their technical flavors. Kriging with external drift is found superior with respect to bias and mean error, ordinary kriging of radar errors with respect to wet-dry distinction and representation of intense precipitation. The two covariance estimation approaches show comparable performance concerning their point estimates, but a clear advantage of the classical parametric variogram estimation concerning their reliability of the probabilistic estimate.

Contents

1 Introduction

Anyone who says sunshine brings happiness has never danced in the rain
Anonymous

There is an increasing demand for accurate gridded quantitative precipitation estimates (QPE) with high spatial and temporal resolution for many applications. The scales of interest are often as challenging as single kilometers and hours. On the one hand, gridded QPE are used in meteorology and climatology. Applications are amongst others climatological analyses of precipitation characteristics (e.g. DeGaetano and Wilks, 2008; Overeem et al., 2009; Wüest et al., 2009) or single events (e.g. MeteoSchweiz, 2006), the assimilation of precipitation data in numerical weather models (NWP) (e.g. Jones and Macpherson, 1997; Leuenberger and Rossa, 2007), the validation of numerical weather models (e.g. Weusthoff et al., 2010; Zimmer and Wernli, 2011) and nowcasting (e.g. Pereira Fo. and Crawford, 1999). On the other hand, spatial precipitation analyses are increasingly used by other user groups as input data for models of natural systems. Hydrological flood forecasting in medium and small scale catchments, (e.g. Young et al., 2000, Zappa et al., 2008, Viviroli et al., 2009; Zhang et al., 2011), for instance, requires high-resolution QPE grids. This application is especially important during heavy precipitation events, as responsible authorities rely on precise runoff forecasts in order to take accurate flood intervention measures. In agriculture, spatial precipitation data are, for instance, employed in crop-modeling (e.g. Westcott et al., 2005). For many applications, spatial precipitation analyses should be available in real-time, i.e. shortly after the end of the accumulation period. This offers a particular challenge to involved measurement systems and data processing procedures.

There are three established measurement platforms to observe precipitation: rain gauges, radar and satellites. Rain gauges measure precipitation directly, whereas radar and satellite are remote sensing instruments. Satellite observations are often the only source of information in tropical regions and over oceans. However, their accuracy is inferior to estimates from dense rain-gauge networks operated over most mid-latitude land areas (e.g. Rudolf et al., 1996; Rubel and Rudolf, 2001).

Although rain gauges do exhibit measurement errors (Neff, 1977), they are comparatively accurate in absolute values at their point location. Their spatial representativeness, however, is strongly limited by the density of the network (e.g. Villarini et al., 2008; Frei et al., 2008). Radar, on the contrary, offers a gridded and extensive coverage with a relatively high spatial resolution, but its QPE are often imprecise due to a variety of uncertainties caused for example by the variability of the Z-R relation, ground clutter, visibility problems and vertical variability of precipitation (Villarini and Krajewski, 2010). In mountainous regions, radar measurements are additionally hindered by beam blockage and intensified ground clutter (Germann et al., 2006).

In this setting, the idea to join the high spatial resolution of radar with the local accuracy of rain gauges in a combined QPE naturally emerges. In a first step, radar-based QPE are often adjusted to long term gauge measurements with the help of static correction maps based on climatological global and local radar-gauge comparisons (Brandes, 1975; Gjertsen et al., 2004; Tabary 2007). At MeteoSwiss, static global and local gauge-adjustment on a yearly basis is part of the post-processing of radar data and has been shown to increase the accuracy of radar-based QPE (Germann et al., 2006).

Apart from this static gauge-adjustment, various approaches to combine radar and rain gauge data in a dynamic way (i.e. separately for each time step) have been proposed. At MeteoSwiss, the project CombiPrecip, a part project of the NCCR Climate III, was launched in 2009 with the objective to develop a prototype method for a dynamic radar rain-gauge combination. Statistical methods applicable in real-time should be implemented, analyzed and, where required, adapted for the sub-daily time-scale and the territory of Switzerland. With regard to a subsequent operationalization, the procedure for this dynamic combination should be applicable systematically, i.e. in any precipitation situation, and produce robust improvements compared to single sensor estimates (radar QPE, interpolated rain gauges).

Table 1 lists a selection of examples of dynamic radar-gauge combination, making no claim to be complete. It includes pioneer studies (Krajewski 1987; Creutin 1988; Seo 1998) as well as a variety of recent applications and developments. The purpose is to give an overview on the classes of existing combination approaches and to illustrate differences in technical choices. The first column of Table 1 groups combination methods according to their interpolation concept: On the one hand, there are combination methods based on deterministic interpolation concepts such as distance weighting (DW) (Bartels et al., 2004; DeGaetano and Wilks, 2008) or splines (Garcia-Pintado et al., 2009; Schneider and Steinacker, 2009). On the other hand, a wide range of applications based on stochastic interpolation concepts has been proposed. These have the advantage of offering not only point estimates, but also estimates of the related uncertainties (i.e. probabilistic estimates) and estimates of the spatial covariance structure. This allows to follow the uncertainty of the input data in application models, for instance by employing simulated precipitation ensembles as input for hydrological modeling (e.g. Zappa et al., 2008; Germann et al., 2009). Many of the stochastic combination methods are based on classical geostatistics, where the spatial covariance is modeled as a function of the lag. In the second column of Table 1, these geostatistical combination methods are further classified by distinguishing three different methodological approaches: Conditional merging/ordinary kriging radar errors (OKRE), kriging with external drift (KED) and co-kriging. Other stochastic combination concepts use a Bayesian framework (e.g. Todini, 2001; Fuentes et al., 2008) or explicitly account for intermittency (Seo, 1998; Fuentes et al., 2008) or temporal dependencies (Seo and Breidenbach, 2002; Sideris et al., accepted).

The fourth and fifth columns of Table 1 indicate differences in technical choices (covariance modeling and data transformation) among the geostatistical combination methods. The covariance structure is traditionally modeled using a parametric variogram function (P), the parameters are fitted by using the (radar and gauge) information at rain gauge locations. Alternatively, a non-parametric correlogram estimation (NP) has been suggested (Velasco-Forero et al., 2009; Schiemann et al., 2011), where the covariance structure is derived from the radar field with the help of some approximation steps (see later, section 3.2). This allows to account for possible anisotropy and an

estimation of the covariance structure for each time step separately even if precipitation is very sparse. Geostatistical methods are based on the assumptions of second-order stationarity and Gaussian distribution and precipitation data often does not comply with these assumptions. Data transformation (square root (SR), logarithmic (L) or variable (V)) has been applied in order to achieve better compliance with model assumptions (Schuurmans, 2007; Verworn and Haberlandt, 2011; Erdin et al., 2012). It has been shown that data transformation can improve not only the compliance with model assumptions, but also the reliability of the probabilistic estimate, i.e. improve the accuracy of the estimation uncertainty (Erdin et al., 2012).

Table 1: Review over a selection of studies on real-time radar-gauge combination with classifications of underlying interpolation concept, combination method, variogram estimation, data transformation, accumulation period (time scale), topography and type of evaluation.

Concept	Method				Authors, Year	Vario-gram	Data Trans-formation	Time scale	Topo-graphy	Appli-cation
	OKRE/Cond. Merging	KED	Co-Kriging	Others		Parametric Non-Parametric	None Square Root Log Variable	Daily Hourly 10min	Flat Moderate Mountainous	Conceptual Selected events Systematic
Deter-ministic				x	Bartels et al. 2004	-	V	H	Md	SE
				x	DeGaetano and Wilks, 2008	-	N	D	Md	Sy
				x	Garcia-Pintado et al., 2009	-	V	H	Md	SE
				x	Schneider and Steinacker, 2009	-	N	D	Mt	SE
Stochastic			x		Krajewski, 1987	P	N	D	F	Co
			x		Creutin et al., 1988	P	N	D	Md	SE
				x	Seo, 1998	P	N	H	Md	Sy
				x	Todini, 2001	P	N	H	Mt	SE
				x	Seo and Breidenbach, 2002	P	L	H	Md	Sy
	x				Sinclair and Pegram, 2005	P	N	-	-	Co
		x			Haberlandt, 2007	P	N	H	Md	SE
	x				Jurcyk et al.,2007	P	N	H	Md	Sy
		x	x		Schuurmans et al., 2007	P	SR	D	F	SE
				x	Fuentes et al., 2009	P	L	10 m	Md	SE
	x	x			Goudenhoofdt & Delobbe, 2009	P	N	D	Md	SE
		x			Velasco-Forrero et al., 2009	NP	N	H	Md	SE
		x			Schiemann et al., 2011	NP	N	H	Mt	Sy
		x			Verworn and Haberlandt, 2011	P	L	H	Md	SE
		x		x	Xie et al., 2011	P	N	H	Md	SE
			x		Yeung et al., 2011	P	N	H	Md	Sy
		x			Erdin et al., in press	P	N/SR/V	H	Mt	Sy
		x	x		Sideris et al., accepted	P	SR	H	Mt	SE

The last three columns of Table 1 address some application characteristics (time scale, topography and evaluation). Many QPE applications (i.e. hydrological flood forecasting) call for high temporal resolution. However, shorter accumulation times imply more zeros (dry rain gauges) and stronger

skewness (large small scale variability, single precipitation maxima) in the data. The sparsity of data and the difficulties to comply with geostatistical model assumptions are thus intensified for short time scales. Moreover, the rain gauge representativeness error decreases nonlinearly with the aggregation period (Villarini et al., 2008). This calls for a sensible choice of time scale to ensure both, utility and feasibility. The selected studies apply radar-gauge combination on daily or hourly time scale. Fuentes et al. (2009) consider even two exemplary 10-min accumulations. Sideris et al. (accepted), include a 10-min accumulation and show that the results for this short time period are degraded compared to hourly accumulations due to the fact that discrepancies between radar and rain gauge measurements are larger for such short aggregation periods. Another distinguishing feature is the topographic character of the application area. Topography influences radar-gauge combination twofold: On the one hand, the spatial variability of precipitation is expected to be increased in complex terrain. This aggravates the representativity problem of the rain gauges. On the other hand, radar measurements are obstructed in mountainous terrain, leading to additional uncertainties in the radar data. The methodologies for radar-gauge combination are evaluated either conceptually (using simulated situations and data) or by the application to real precipitation data. A majority of studies applying the combination to real precipitation data uses selected events for their evaluations. Such events are usually situations with intense and relatively widespread precipitation, supplying ample data for model parameter estimation. In contrast, a systematic application at short time scales (daily or even hourly) needs to be capable of dealing with any precipitation situation, which poses a challenge to combination methods: Data is often very sparse due to the intermittent nature of precipitation, leading to unrobust or even impossible estimation of model parameters. Fall back solutions such as predefined climatological parameters (Seo, 1998; Seo and Breidenbach, 2002; Jurczyk et al., 2007; Yeung et al., 2011) or non-parametric correlogram estimation (Velasco-Forero et al., 2009; Schiemann et al., 2011) are hence required for a systematic application.

This report documents a part of the work that has been achieved in the course of the part project CombiPrecip at MeteoSwiss. It presents an application of dynamic radar-gauge combination on the hourly time scale in Switzerland, a region with mountainous and complex topography. To this end, two popular geostatistical combination methods applicable in real-time, kriging with external drift (KED) and ordinary kriging of radar errors (OKRE), are employed. Data transformation is applied (i.e. we use a trans-Gaussian framework), because a precursor study (Erdin et al., 2012) suggests better compliance with model assumptions and improved reliability of the probabilistic estimate thereby. A systematic application of both combination methods over a full year (2008) is performed. This setting results in particular challenges: Data is often very sparse (hourly time scale, systematic application) and the uncertainty of measurements is large (mountainous topography). Two alternatives to cope with the sparsity of data are compared: predefined climatological parameters for the classical parametric approach and non-parametric correlogram estimation as suggested by Velasco-Forero et al. (2009). For the evaluation of the combined QPE, cross-validation for the full systematic application is performed and different error measures for the point estimate and the probabilistic estimate are calculated to assess the characteristics of the gridded QPE. Single sensor QPE (radar and interpolated rain gauges) serve as a reference and the combination methods are compared with each other. To gain a detailed understanding of the potential and limitations of the combination approaches, analyses with respect to different regions and seasons and the influence of technical choices, such as data transformation level and fall back solutions for situations with sparse data, are discussed.

The report is organized as follows: The setting of the study and the data employed are described in section 2. Applied geostatistical combination methods, specific technical settings and evaluation procedures are introduced in section 3. Section 4 presents and discusses the results of our application and section 5 summarizes the findings of the study and draws conclusions. Appendix A supplies an explanation on a transformation-dependent behavior of the overall bias observed in our application.

2 Data

The study region is Switzerland, with an area of roughly 41'000 km^2 and a rather challenging topography with altitudes ranging from below 200 m to more than 4000 m (see Fig.1). The high mountain range of the Alps divides the northern part of the country (Swiss Plateau, moderately hilly) from the South (moderately mountainous). The medium-rise mountains of the Jura border the Swiss Plateau to the northwest.

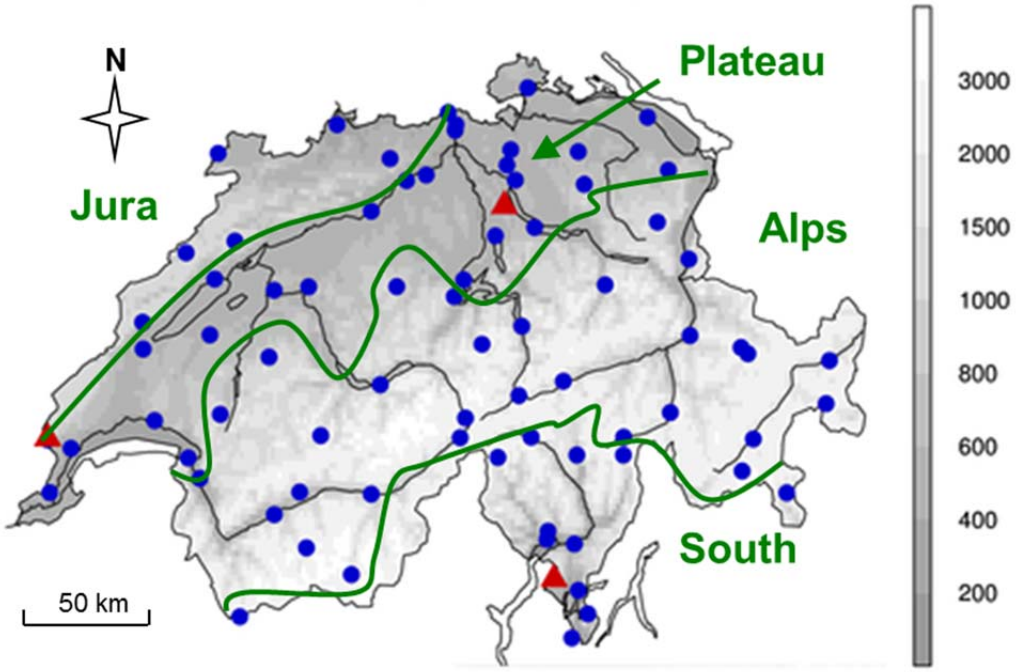

Figure 1: Study area and operational real-time precipitation network of MeteoSwiss in 2008: 75 automated rain gauges (blue dots) and 3 C-band radars (red triangles). Topography (grey shadings, in meters above sea level) and 4 geographic regions of Switzerland (green division lines and labels).

The gauge network for our analyses consists of 75 automated rain gauges (tipping buckets) of the SwissMetNet (SMN) operated by MeteoSwiss (MeteoSwiss, 2010) in the considered time period (blue dots in Fig.1). It covers all Switzerland with a typical inter-gauge distance of 25 km. Hourly accumulated precipitation by gauges is provided in mm with a discretization of one decimal place. The gauge data employed in this study have passed thorough manual quality control (Musa et al., 2003), but no correction for systematic biases (Neff, 1977).

For the radar data, we use hourly composites from the 3 C-band radars operated by MeteoSwiss in 2008 (red triangles in Fig.1) on a grid of 1 x 1 km. These composites constitute a further advancement of the Swiss radar QPE described in Germann et al. (2006), termed NASS. The conversion from reflectivity to rain rate is based on a fixed Z-R-relation and the post-processing includes clutter filters, visibility and vertical profile correction (Germann and Joss, 2002). Additionally, there is a static global and local bias adjustment based on climatological radar-gauge comparisons. The hourly composites are derived by averaging the original 5-min radar composites. This limited temporal sampling (average of 12 instantaneous rain rates) occasionally manifests in unrealistic patterns in the hourly composite, especially when small-scale rainfall systems move fast. Schemes to mitigate these effects by a temporal smoothing have been suggested (as e.g. Fabry et al., 1994; Tabary et al., 2007). However, experiments (not shown) showed an ambiguous influence of such a technique depending on the precipitation situation and only small potential improvements, even in an extreme case with fast moving hail storms for our application. We therefore use no further time aggregation of the radar data here. Note that rain gauge measurements are considered as ground truth in this study, i.e. they serve as a local reference to assess the accuracy of radar and other QPE.

The study is based on a systematic application of methods throughout the year 2008. From this total of 8784 hourly accumulations, 3375 (38%) are dry, i.e. all gauges record dry conditions (0 mm). In 1621 hours (18%), 10 or more gauges register precipitation ≥ 0.5 mm, referred to as ‘ample sampling’ in the following. In the remaining 3788 (43%) hours, there are gauges with positive precipitation, but only 9 or less record ≥ 0.5 mm, referred to as ‘poor sampling’ hereafter.

Beside this systematic evaluation over one year, a set of 12 test events from the years 2005-2009 has been defined. These test cases cover 11-24 hours each and represent various types of precipitation situations that presumably influence the performance of combination methods: different seasons, heavy and light precipitation, stratiform and convective regimes, fast-moving and slow-moving cells. Moreover, they include situations with accurate and less accurate radar estimates (compared to gauge measurements). The idea is to have a relatively compact but broadly representative testbed, for experiments that are computationally very demanding. The total of 254 hourly accumulations of these 12 test cases includes 18 dry hours, 153 hours of ample sampling and 83 hours of poor sampling. Figure 2 and Figure 3 show exemplary hourly accumulations of two of these test cases. These examples will be employed later on (in section 3) to illustrate properties of the combination methods and single sensor references. The test case of July 23, 2009 (Fig. 2) shows a fast-moving, harmful hail storm affecting mainly the Swiss Plateau. It represents an ample sampling case (29 gauges record ≥ 0.5 mm), with high spatial variability and locally intense precipitation (up to 14.5 mm). Radar significantly overestimates the extent and intensity of precipitation in this example case, as is evident by comparison to the station measurements. The test case of June 11, 2008 (Fig. 3) shows isolated convective precipitation at the northern slope of the Alpine ridge in Central Switzerland. It represents a poor sampling case (5 gauges record ≥ 0.5 mm) with high variability on small spatial scales. Radar shows relatively accurate QPE for this example. However, there is a slight overestimation by radar in the northern part of the wet region, whereas the precipitation at the southernmost wet rain gauge is missed.

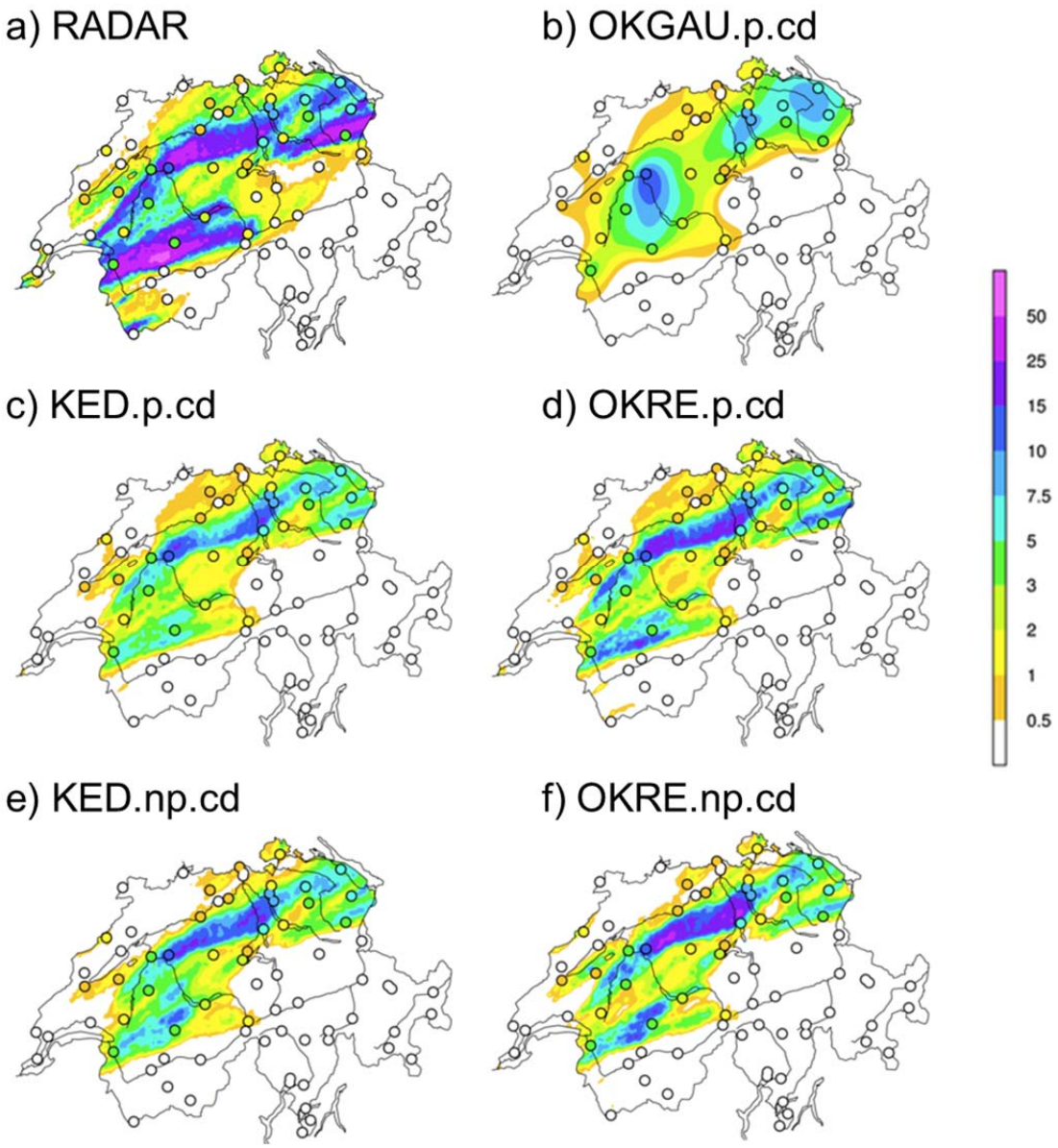

Figure 2: Example of an ample sampling case: July 23, 2009, 15:00- 16:00 UTC. Hourly precipitation estimates (point estimates) by (a) RADAR, (b) OKGAU,(c) parametric and (e) non-parametric KED and (d) parametric and (f) non-parametric OKRE. Case-dependent transformation, estimated from the data. Circle fillings represent the station measurements and are the same in all panels.

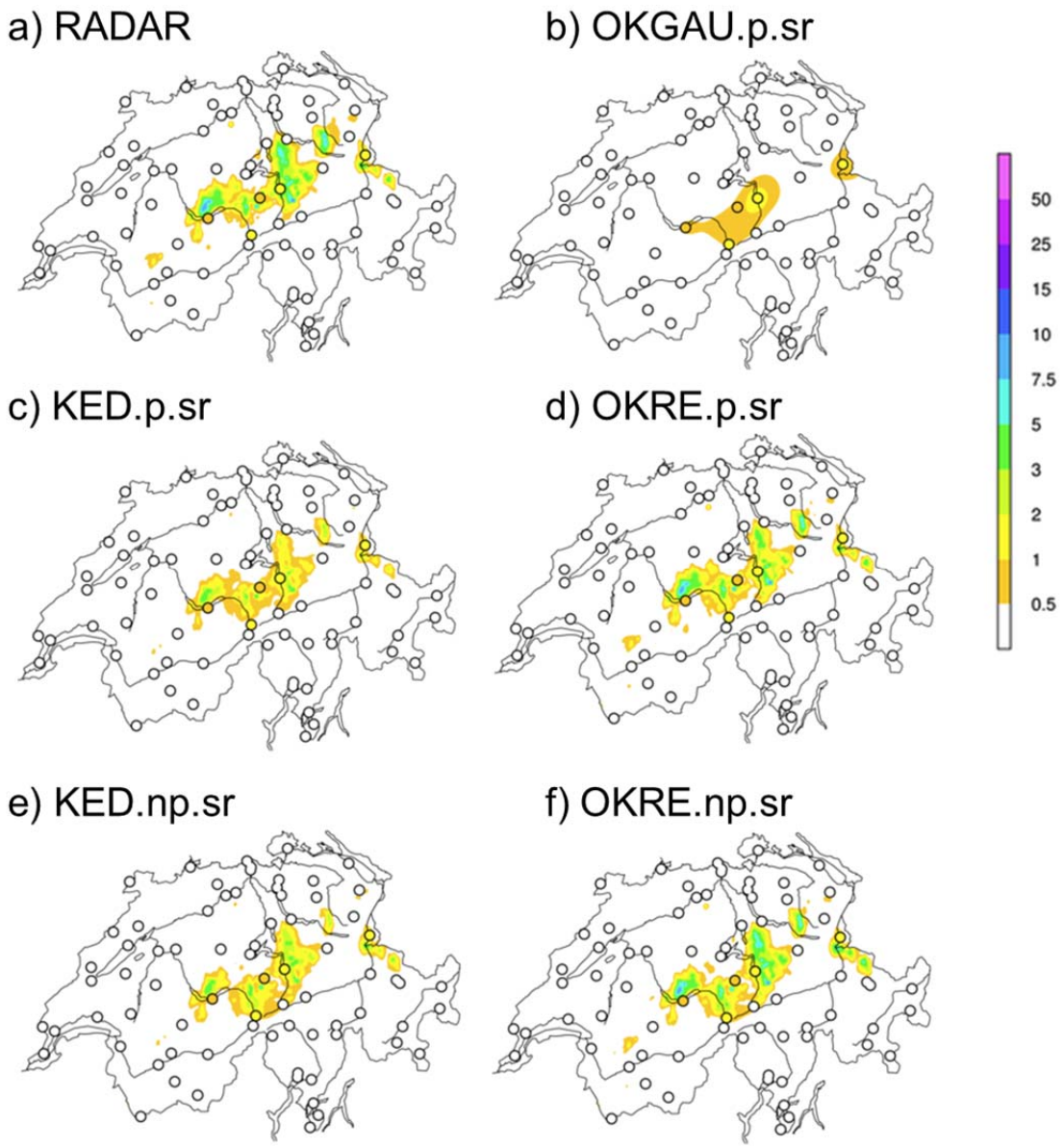

Figure 3: Example of a poor sampling case: June 11, 2008, 23:00- 24:00 UTC. Hourly precipitation estimates by (a) RADAR, (b) OKGAU ,(c) parametric and (e) non-parametric KED and (d) parametric and (f) non-parametric OKRE. Prescribed square root transformation.

3 Methods

In classical geostatistics, spatial observations are perceived as one single realization of a multivariate random variable, which is assumed to be spatially correlated, second-order stationary and Gaussian. The covariance structure – hence assumed to depend on the lag-distance only in an isotropic setting – is traditionally modeled by a parametric semi-variogram function. The geostatistitical interpolation method kriging produces the best linear unbiased predictor (BLUP) conditional on the observations at any location in the domain. Kriging yields the kriging predictor, a point estimate (best estimate), but also the kriging variance, an estimate of the uncertainty, which depends on the spatial covariance and the local density and geometry of the observations. Due to the Gaussian assumption, a Gaussian probabilistic estimate at any target location can thus be constructed. Thorough introductions in geostatistics can be found in the literature (e.g. Cressie, 1993; Schabenberger and Gotway, 2005; Diggle and Ribeiro Jr., 2007).

The following subsections cover the specific methods and techniques employed in this study. The concepts of the two geostatistical radar-gauge combination methods, the estimation procedures for model parameters and the data transformation in our trans-Gaussian framework are introduced in the first three subsections. Our choices for the fall-back procedures required in a systematic application and technical details such as radar-gauge matching are presented in the fourth subsection. The last two subsections introduce the single sensor QPEs used as a reference and the validation procedures.

Table 2 will be repeatedly referred to in the following subsections. It gives an overview on the combination methods and configurations analyzed in this study and on the single sensor references used for comparison. The acronyms listed will be used throughout analyses to concisely define the configuration of combination method/reference, covariance modeling and transformation setting.

Table 2: List and acronyms of references and combination methods applied in this study, with corresponding setting of variogram estimation and transformation setting. The last three columns list the climatological parameters for the case dependent transformation and the parametric variogram estimation in hours with poor sampling (<10 gauges with <0.5mm).

Acronym	*Method*	*Covariance*		*Transformation*		*Climatological parameters*		
		Param.	Non-param.	Square root $\lambda = 0.5$	Case-dependent $\lambda = \lambda_e$	λ	Nugget-sill-ratio	Effective range
RADAR	Single sensor QPE by radar (Reference)	-	-	-	-	-	-	-
OKGAU.p.sr	Single sensor QPE by ordinary kriging of gauges (OKGAU) (Reference)	x		x		0.5	0.03	126km
OKGAU.p.cd		x			x	0.2	0.03	126km
KED.p.sr	Combination of radar and gauges by kriging with external drift (KED)	x		x		0.5	0.2	75km
KED.p.cd		x			x	0.2	0.2	75km
KED.np.sr			x	x		0.5	-	-
KED.np.cd			x		x	0.2	-	-
OKRE.p.sr	Combination of radar and gauges by ordinary kriging of radar error (OKRE)	x		x		0.5	0.28	108km
OKRE.p.cd		x			x	0.2	0.28	108km
OKRE.np.sr			x	x		0.5	-	-
OKRE.np.cd			x		x	0.2	-	-

3.1 Geostatistical combination methods

In this study we employ two popular geostatistical methods for radar-gauge combination. In both, the rain gauges are regarded as (erroneous) measurements of true precipitation, whereas radar is comprised as auxiliary information on the spatial distribution. At this point, the difference in spatial support between the two measurement device should be emphasized: The surface of a rain gauge measures 200 cm^2 and can therefore be perceived as a point measurement compared to the radar and target grid size of 1 km^2. However, the gauge measurement can also be perceived as an estimate of the areal precipitation at the corresponding grid pixel, though affected with random errors. This alternative perception justifies the combination and allows the required radar-gauge matching which will be discussed further below (see section 3.4).

The first method is known as kriging with external drift (KED). KED consists of a deterministic and a stochastic model part:

$$\mathrm{Prec}_{i,j} = \underbrace{\alpha + \beta * \mathrm{Radar}_{i,j}}_{\substack{\mathrm{Trend} \\ \mathrm{Deterministic-Part}}} + \underbrace{Z_{i,j}}_{\substack{\mathrm{Random-Process} \\ \mathrm{Stochastic-Part}}} \quad (1)$$

Where Prec stands for precipitation, i and j are the spatial indices of a target grid cell, α and β are the coefficients of the trend model and Z denotes a Gaussian random process. The deterministic part (trend) in KED is hence a linear model with the radar value of the grid cell of interest as predictor. The stochastic part employs a Gaussian random process (Z) to model local deviations from this trend model.

The second method is termed ordinary kriging of radar errors (OKRE) in this study. The idea of OKRE is to interpolate the observed radar errors at gauge locations (determined by local radar-gauge comparison for details see section 3.4) by ordinary kriging (OK). The resulting radar error field is subsequently subtracted the from the original radar field to obtain the actual precipitation estimate:

$$\mathrm{Prec}_{i,j} = \mathrm{Radar}_{i,j} - \left(\underbrace{\underset{\sim}{\alpha}}_{\substack{\text{Trend} \\ \text{Deterministic-Part}}} + \underbrace{Z_{i,j}}_{\substack{\text{Random-Process} \\ \text{Stochastic-Part}}} \right) \qquad (2)$$

The deterministic model part of OKRE consists of a constant, namely the mean of all observed radar errors. Analogous to KED, the stochastic part employs a Gaussian random process (Z) to model the local deviations from this constant trend field. The comparison of Equation 1 and Equation 2 shows that OKRE can be seen as a special case of KED with $\beta = 1$ and inversed sign of ($\alpha + Z$).

KED and OKRE have both been employed and analyzed previously for radar gauge combination (see Tab 1). However, applications differ in technical details and choices (e.g. data transformation, parameter estimation). The method termed OKRE in this study is similar to a combination approach suggested by Sinclair and Pegram (2005) known as conditional merging, which is the reason for pooling them in Table1 . Both approaches use the unaltered ($\beta = 1$) radar pattern for the spatial structure in between the rain gauges (which are regarded as the truth and, hence, conditioned on). The difference between the two methods lies in the variable to be interpolated: in conditional merging, the radar values at rain gauge locations are interpolated (and then compared to the original radar field, resulting in a field of 'interpolation error' which is added to interpolated gauge field), whereas in OKRE the radar error at gauge locations is interpolated as described above. If these two variables had identical covariance structures, conditional merging and OKRE would lead to exactly the same precipitation field. The other way around, differences in covariance structure between radar and radar error (as typically expected) lead to differences in the results by conditional merging and OKRE.

Figure 2 and Figure 3 show exemplary gridded QPE by the two combination methods KED and OKRE (with different configurations of covariance estimation and data transformation see section 3.2 and 3.3) and the single sensor references (see section 3.5). They illustrate how combination methods typically incorporate the spatial pattern of radar. At the same time, the local radar bias is widely corrected by the conditioning to gauge values. The visual comparison between the two combination methods shows a slightly smoother pattern and lower precipitation values for KED, compared to a more rugged pattern (similar to the original radar) and higher precipitation amounts for OKRE. These differences are characteristic, as the inspection of a variety of other cases (not shown) confirms. They can be explained by the fact that the radar coefficient in KED is usually between zero and one (see Appendix Fig. A.4), leading to an attenuation of the influence of radar in KED compared to the unaltered inclusion in OKRE.

3.2 Estimation

Both geostatistical combination models introduced in the preceding subsection (3.1) require the estimation of model parameters. On the one hand, the deterministic model part includes one (OKRE) or two (KED) trend parameters. On the other hand, the stochastic part requires a model for the covariance structure of the Gaussian random process. In this study, two different covariance modeling approaches are employed and compared.

The first is a classical, parametric approach, where we assume an isotropic, exponential variogram function with a nugget. This function describes the spatial covariance of the random process as a function of the lag distance based on three parameters: the range, defining the maximum distance of the spatial correlation, the nugget, defining the spatial variability at unresolved, small spatial scales as well as measurement errors and the sill, defining the total variance of the random process. The choice of this variogram function is motivated by its relative simplicity (three parameters) with regard to the often very sparse data in our application context. Moreover, preliminary experiments (Erdin, 2009) have shown no large impact of the choice of variogram function. Trend and variogram parameters are estimated simultaneously by restricted maximum likelihood estimation (REML). This fully data driven estimation procedure does not require subjective choices and avoids biases in variogram parameters by conventional maximum likelihood estimation (MLE) incurred due to the finiteness of the sample.

The second is a non-parametric approach based on fast fourier transform (FFT) suggested by Velasco-Forero et al. (2009) and applied and modified by Schiemann et al. (2011). Here, the covariance structure is estimated by a non-parametric correlogram based on the radar field. However, as the covariance structure of the radar error (OKRE) or of the residuals from the linear trend model (KED) is needed, a field of true precipitation would in principle be required to compare the radar field to. Obviously, this is not available and hence needs to be approximated. This is solved by an iterative procedure, starting with an ordinary kriging of rain gauges using the radar field for the non-parametric correlogram which is taken as the first approximation for true precipitation. For a detailed description of the procedure refer to Velasco-Forero et al. (2009) and Schiemann et al. (2011). Experiments with different numbers of iteration steps (not shown) suggest a considerable improvement up to the second iteration, but only slight changes (improvements or even retrogressions) for more iterations. Therefore, two iterations are performed in our application for this study (compared to no iteration by Velasco-Forero et al. (2009) and one iteration by Schiemann et al. (2011)).

Throughout this study, the acronyms of combination methods and single sensor references include *p.* whenever the parametric and *np.* whenever the non-parametric covariance modeling approach has been employed (see Tab. 2). In Figure 2 and Figure 3 , parametric and non-parametric KED and OKRE are shown for an exemplary visual comparison. In the example with high anisotropy in the precipitation pattern (July 2009, Fig. 2), notable differences between the two covariance modeling approaches are observed for both combination methods, whereas in the more isotropic example (June 2008, Fig. 3), there are only small differences. However, in both examples, the influence of the combination method is larger than the influence of the covariance modeling approach.

3.3 Data transformation

Classical geostatistics are based on the assumptions of second-order stationarity and Gaussian distribution of the spatial variable under consideration. Radar errors (OKRE) or the residuals of a linear radar-gauge relation (KED) often do not comply with these assumptions: They usually scale with precipitation amount leading to a non-stationary variance (small in dry areas, large in regions with intense precipitation) and skewed and/or long-tailed distributions. In a previous study (Erdin et al., 2012), a Trans-Gaussian framework has been suggested to mitigate these violations of model assumptions. It has been shown that by suitably transforming radar and gauge data, the compliance with model assumptions and the reliability of the probabilistic estimate can be improved. The point estimate has been shown to be relatively insensitive to data transformation as long as excessive transformations are avoided. Motivated by these findings, the trans-Gaussian framework is pursued in the present study: Radar and gauge data are transformed using the Box-Cox transformation family (Box and Cox, 1964):

$$Y^* = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(Y) & \lambda = 0 \end{cases} \qquad (3)$$

where Y and Y* are the untransformed and transformed variable respectively, and λ is the transformation parameter. The Box-Cox transformation is essentially a power transformation with lambda representing the transformation exponent.

Two transformation levels are applied and compared in our application. The first is a prescribed square root transformation (λ = 0.5) denoted by *sr* in the acronym (Tab. 2). The second is a case-dependent (variable) transformation, where λ is estimated from the data ($\lambda = \lambda_e$), denoted by *cd* in the acronym (Tab. 2). λ_e is estimated by MLE (Maximum Likelihood estimation), optimizing for a Gaussian distribution of radar errors/residuals. A beta-prior distribution with positive probability density on [0.2, 1.5] is imposed to this MLE to avoid λ below 0.2, because positively biased point estimates have been found for such small values of λ in the precursor study (Erdin et al., 2012). In our systematic application over the year 2008, λ_e have been found on [0.2, 0.68] with a large majority of values close to the lower bound (mean λ_e for KED 0.24, for OKRE 0.25).

After the application of the kriging on transformed scale, the resulting probabilistic estimates (uniquely defined by the kriging predictor, the kriging variance and the Gaussian assumption) are back-transformed to obtain an estimate at precipitation scale. We use a discretized probabilistic estimate (199 equidistant quantiles) and the inverse transformation function to this end. The back-transformed best estimate (point estimate) is then computed by the mean of the back-transformed discretized probabilistic estimate.

An example of gridded QPE (i.e. fields of the back-transformed point estimate) of combination methods employing the case-dependent (*cd*) transformation can be found in Figure 2. Figure 3 shows an example application of the prescribed square root (*sr*) transformation.

3.4 Fall back procedures and technical details

The major challenge of the presented application setting of geostatistical radar-gauge combination is posed by the sparsity of rain-gauge samples: In a systematic application at the hourly time scale, most of the 75 available rain gauges are usually dry (e.g. in 82% of all hours in 2008, there are less than 10 gauges with positive precipitation ≥ 0.5 mm, see section 2). This makes a robust estimation of parameters (transformation, trend and variogram) delicate or even impossible and, hence, calls for fall back solutions.

For the parametric covariance estimation, predefined climatological values for the nugget-sill ratio and range, (see Tab. 2) are employed in all hours with poor sampling (< 10 wet gauges (≥0.5 mm)). These climatological parameters are motivated by precursor experiments (not shown): Pooled variograms for all hours with poor sampling have been calculated, considering the current hour together with up to 5 past time-steps to obtain at least 10 wet gauges in the pooled sample. The parameters of these pooled estimated variograms showed highly unrobust behavior leading to problematic effects on the gridded QPE. For a more solid and reliable solution, the medians of these estimated pooled parameters are taken as climatological values here. We employ the same climatological variogram parameters for both transformation settings (*sr* and *cd*), as experiments showed no considerable differences. Note that the climatological variogram parameters define the range and the nugget-sill ratio of the random process, but not the magnitude of the sill and nugget. To obtain this magnitude, the variance of the residuals (KED) or radar errors (OKRE) at gauge locations in the current hour is taken as an estimate of the total sill. The trend parameters are calculated by ordinary least squares (i.e. without accounting for the spatial correlation of residuals) in hours with poor sampling. We are aware that using these predefined parameters is a strong simplification of the model which might not capture the storm characteristics of hourly precipitation adequately in some cases. However, more complex fall back solutions (e.g. several sets of climatological parameters for different storm types) would require at least one additional data-driven decision, which would again contribute to robustness problems. Therefore, the strong simplification to only one set of climatological parameters is chosen here. An example of the described fall back procedure based on climatological variogram parameters can be found in Figure 3, panels (b), (c) and (d). It shows a case of poor sampling: Five rain gauges register precipitation of at least 0.5 mm. Note that only the parametric variogram estimation relies on the predefined climatological parameters, whereas the non-parametric correlogram estimation can be applied unaltered even in poor sampling cases, because the covariance is modeled based on the radar data.

The transformation parameter for the case-dependent transformation (λ_e) is set to 0.2 for all hours with poor sampling for both, the parametric and the non-parametric approach. Experiments (not shown) revealed that λ_e sticks to this lower bound of the imposed prior distribution as a rule in hours with sparse data.

Beside this reduction of model parameters for hours with poor sampling, solutions for totally dry hours and for situations where the algorithms fail are required with respect to an operational application. In dry hours (all gauges register 0 mm), obviously, the kriging model cannot be employed for KED and OKGAU (a single sensor reference method, see the following section 3.5). Precipitation is set to zero at every location in these cases (even if radar may indicate rainfall at some points in the domain). Similarly, OKRE cannot be employed in cases where the radar error at all gauge locations

is zero. This occurs typically only in dry hours, when radar and rain gauges measure 0 mm at all gauge locations. The radar error field is then set to zero, and hence the unaltered radar field is taken as precipitation estimate in these cases. Note that this includes the possibility of positive precipitation in between the gauges. Failure of the algorithm can occur due to a lack of data (e.g. only 1 gauge with positive precipitation) or due to optimization problems in the MLE (occurs only very rarely, roughly in 0.2% of all cases in 2008). We choose to set KED and OKRE to the unaltered radar field in these cases. Note that these solutions for dry and failure cases offer only point estimates, no probabilistic estimates. Since the focus of this report is to understand the performance of the combination methods, only hours where an application (with or without fall back procedures) is possible are included in our analyses. Careful analyses and possibly a reassessment of failure solutions is therefore recommended if an operational implementation is aspired.

As discussed above (see 3.1), radar and rain gauges differ in spatial support: rain gauges can be perceived as point measurements, whereas radar provides an areal average of precipitation over a grid cell, in our case 1 km^2. This difference in spatial support accounts for a part of the difference between radar and gauge measurements due to the gauge representativeness error (e.g. Ciach and Krajewski, 1999). Nevertheless, the two precipitation measurements can be matched, as required for the trend estimation, if the gauge is perceived as an (erroneous) point measurement of the true areal precipitation in the radar pixel. The matching is, hence, done by nearest neighbors, i.e. the radar value of the pixel where the gauge is located in is taken for the radar value at the gauge location. The random error by the gauge is accounted for by the nugget parameter in the geostatistical model. Note that the nugget contains not only this representativeness error due to unresolved small scale variability, but also other uncertainties associated with gauge measurements (i.e. measurement errors).

The radar product employed in this study (see section 2) exhibits very small positive values in some pixels, most of them are presumably artifacts. To prevent a degradation of combined QPEs by these spurious values, all radar values below 0.05 mm are set to zero in our application. This corresponds to the discretization level of the rain gauge (0.1mm).

All computations are done with the statistical software R (R Development core Team, 2009, R: A Language and Environment for Statistical Computing) and build on the geostatistical R-package geoR (Ribeiro Jr., 2001, geoR: a package for geostatistical analysis)

3.5 Single sensor references

The combined QPE are compared to single sensor QPE by radar and rain gauges to analyze the added value by the combination.

For the single sensor reference by radar (acronym RADAR, see Tab. 2), the radar product employed for the combination methods (NASS, described in section 2) is taken.

For the single sensor reference by rain gauges (acronym OKGAU, see Tab. 2), a gridded interpolation of the 75 real-time gauges is produced. The interpolation is done by ordinary kriging (OK). The technical setting (data transformation, variogram estimation, estimation procedures) for this ordinary kriging is identical with the procedures for the combination methods (see preceding subsections). This should help to assess the added value by including radar information without

disturbing influences of technical choices. Note that only the parametric variogram estimation is possible for OKGAU, as the non-parametric correlogram estimation by definition includes radar information.

Examples of QPE fields by the single sensor references are displayed in Figure 2 and Figure 3, subfigures (a) and (b). In the example of July 2009 (Fig. 2), RADAR shows a considerable overestimation of precipitation amount and spread. This illustrates the typical shortcoming of single sensor QPE by radar, namely the low local accuracy. In the example of June 2008, radar estimates are more accurate. However, there is a slight overestimation of precipitation in the northern part of the wet region. The precipitation at the southernmost wet rain gauge is missed by the radar. This illustrates how precipitation at the ground can be invisible for the radar because of shielding effects. OKGAU, in contrast, produces relatively unbiased precipitation fields. However, the very smooth fields with the circular structures around stations (typical for isotropic ordinary kriging) seem unrealistic. They illustrate how the local structure of precipitation is often completely lost in single sensor QPE by rain gauges, reflecting their typical shortcoming, namely their spatial sparcity leading to a low spatial resolution.

3.6 Validation

To analyze and compare the performance of combination methods, full leave-one-out cross-validation is performed. This means that gauge *i* is excluded and the parameter estimation and kriging are run with the set of all remaining (*n-1*) gauges to produce an precipitation estimate at the location of gauge *i.* This is repeated for each of the available rain gauges and, thus, cross-validated estimates are obtained for all gauges and time steps. For the single sensor reference RADAR, the radar value of the pixel where the gauge is located in is taken as precipitation estimate at the gauge location. The estimates by cross-validation (and by radar) are compared to the rain gauge measurements for the validation of combination methods and single sensor references. I.e. gauge measurements are regarded as an approximation of true precipitation at their location in this study, which is motivated by their comparatively high local accuracy.

For the analysis of the point estimate, the following set of error measures is calculated from the pooled point estimates by cross-validation. The idea is to assess different characteristics of the performance of a method. First of all, we are interested in a possible systematic error of a QPE. To this end, we assess *BIAS*, defined as the ratio between the total estimated precipitation and the total observed precipitation (over all time steps and locations). The ratio is expressed in log-scale (dB), as is frequent practice in radar meteorology. A perfect estimate produces a *BIAS* of 0 dB, positive values of *BIAS* reflect an overestimation and negative values of *BIAS* an underestimation of total precipitation amount. For instance, a *BIAS* of 1 dB corresponds to an overestimation by a factor of 1.26, 3 dB to a factor 2. To assess the ability of a QPE to distinguish between dry and wet areas, the Hanssen-Kuipers discriminant (*HK*) is calculated (see Peirce, 1884), using a separation threshold of 0.5 mm. *HK* corresponds to the probability of detection (of wet observations) minus the false alarm ratio (for wet). *HK* = 1 is reached by a perfect estimate, whereas *HK* = 0 is equivalent to a random estimate without any skill. For the overall performance, we consider a new error measure, the relative mean root squared error (*Rel MRTE,* see Eq. 5). The *Rel MRTE* is similar to a root mean squared error, but adapted to the skewed nature of precipitation. Instead of the untransformed estimates and

observations, the square root transformed values are taken in order to account for the fact that intense precipitation is not predictable with the same absolute accuracy as light precipitation or dry values. The *Rel MRTE* is based on the mean root squared error (*MRTE*) defined as:

$$MRTE = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{est_i} - \sqrt{obs_i}\right)^2 \quad (4)$$

Where *est* stands for the precipitation estimate (by cross-validation, or by radar), *obs* denotes the observation of the gauge and *i* is the index over *n* observation-estimate pairs under consideration. A perfect QPE would produce a *MRTE* of 0. For the *Rel MRTE*, this *MRTE* is normalized by the mean root squared deviation (*MRTD,* Eq. 6) in order to mitigate the dependency on the precipitation situation:

$$Rel.MRTE = \frac{MRTE}{MRTD} \quad (5)$$

With

$$MRTD = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{obs_i} - \overline{\sqrt{obs}}\right)^2 \quad (6)$$

A *Rel MRTE* of 0 corresponds to a perfect estimate, whereas a *Rel MRTE* of 1 would be obtained by predicting the average precipitation at all locations. Note that despite the square root transformation of precipitation values and the normalization with the observation variance, *Rel MRTE* cannot be compared one-to-one between different samples (as most error measures).

Geostatistical combination methods produce not only a kriging estimate (best estimate, point estimate), but also an estimate of the related uncertainty, the kriging variance. Together with the assumption of a Gaussian distribution on transformed scale, a Gaussian probability densitiy function (PDF) at any target location can thus be constructed and subsequently be back-transformed. This leads to (positively skewed) PDFs for precipitation at any target location, termed probabilistic estimate in this study. To assess the reliability of these probabilistic estimates, we use a procedure similar to the Talagrand diagram (or rank histogram) in ensemble verification (Talagrand et al., 1997; Hamill, 2001): The idea is to compare the probabilistic estimate to the actually observed distribution of rain gauge measurements. To this end, the frequency of gauge observations to fall in between two quantiles (say Q_{p1} and Q_{p2}) of the probabilistic estimate by cross-validation is computed. This frequency is then compared with the expected frequency in a fully reliable system (i.e. p_2-p_1). Let us consider a numerical example to illustrate this procedure: Assume that 10% of all rain gauge measurements in a certain application are found between the 5% quantile ($Q_{0.05}$) and the 10% quantile ($Q_{0.10}$) of their pertinent probabilistic estimate by cross-validation. The observed frequency in the inter-quantile bin between the 5% and the 10% quantile would thus be 0.1. In a fully reliable system, however, only 5% (p_2-p_1, i.e. 0.10 – 0.05) of all observations would be found in this inter-quantile bin. The ratio between the observed and the expected frequency in this example would thus be 2. The reliability assessment applied in this study now consists of calculating this ratio of the observed and expected frequency for a range of predefined inter-quantile bins covering the entire PDF of the probabilistic estimate. A method with perfectly reliable probabilistic estimates would result in ratios of 1 for all interquantile-bins. Values below 1 indicate too few and values above 1 indicate

too many observations in the respective interquantile-bin. Note that the mean of the frequency-ratios is one by definition, since all observations have to fall any inter-quantile bin.

4 Results

This section presents and discusses the results of our application of geostatistical radar-gauge combination for the hourly time scale and the territory of Switzerland.

The first three subsections are based on the systematic application of the two combination methods (KED and OKRE) throughout the year 2008. Note that this extensive application includes both transformation settings (*sr* and *cd* in the acronyms), but solely parametric variogram estimation (*p* in the acronyms). The two single sensor estimates (RADAR and OKGAU) serve as references to assess the added value by the combination. The analyses are based on the cross-validation results of 4774 hours with at least one gauge with positive precipitation (> 0mm). To ensure the comparability between all applied QPE methods, hours where (at least) one method was forced to use the failure solution (see 3.4) have been excluded from the analyses. This is only the case for a few single hours and, hence, should not affect the validity of the presented results. The first subsection focuses on the performance of the point estimate. To this end, different error measures (see 3.6) are analysed with respect to sampling conditions, transformation setting, geographical regions and seasonal variations. In the second subsection, the sensitivity of the method performance on precipitation intensity is discussed. The third subsection addresses the assessment of the probabilistic estimate, in general and for intense precipitation amounts. This includes a discussion on the influence of the combination method and transformation setting on the reliability.

The fourth subsection compares the two covariance modelling approaches – the classical parametric variogram estimation and the non-parametric correlogram estimation. The analyses of this last subsection, assessing the performance of the point and the probabilistic estimate, are based on an application to a set of 12 test cases (see sections 2 and 4.4).

4.1 Point estimates by parametric methods

Figure 4 shows barplots of the error measures *BIAS*, *HK* and *REl MRTE* for combination methods and single sensor references for different sampling conditions: The leftmost group of bars (*all*) are the error measures calculated from the full sample of the systematic application (all 4774 hours). The central group of bars *(>=10 wet stats*) is based on all hours with ample sampling (1552 hours, see section 2). The group to the right *(<10 wet stats*) refers to the error measures for all hours with poor sampling (3222 hours). This stratification compares the performance of applied QPE methods for situations, where the full set of parameters can be estimated by MLE (ample sampling) with situations, where climatological parameters have to be applied (poor sampling). All QPE obtain considerably better error measures in the ample sampling cases than in the poor sampling cases. For *BIAS* and *HK*, this difference is particularly large for the geostatistical methods (OKGAU, KED

and OKRE), and less pronounced for RADAR. With respect to the overall performance (*Rel MRTE*), however, there is a similar deterioration in cases with poor sampling for all QPE methods. The quality of QPE by radar-gauge combination is, hence, strongly depending on the sampling condition. This implies that the comparison of different QPE methods is only meaningful for consistently comparable sampling conditions. In this first analysis of error measures (Fig. 4), OKGAU is included with both transformation settings. In the following analyses of the error measures, only OKGAU.p.sr is pursued for sake of clarity.

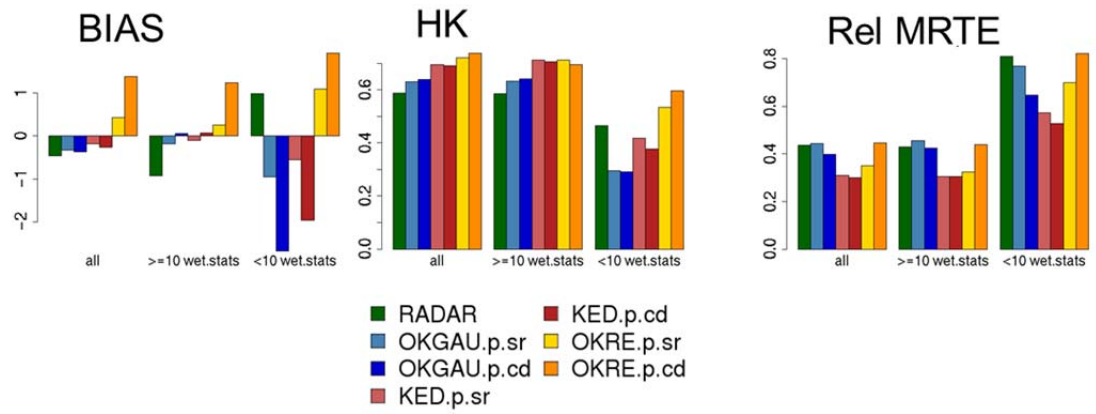

Figure 4: Error measures of the systematic application over the year 2008, based on 1552 hours with ample sampling (>=10 wet.stats) and 3222 hours with poor sampling (<10 wet.stats). Comparison between RADAR and OKGAU, KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

To analyze possible regional characteristics of radar-gauge combination, error measures stratified for four geographical regions of Switzerland (cf. Fig. 1) are displayed in Figure 5. The four geographical regions are derived by a pooling of 12 climatic regions commonly used at MeteoSwiss (Oswald et al., 2010). As one would expect, RADAR shows better error measures in the Swiss Plateau (moderate topography, central region: coverage by more than one radar) than in Alps (high mountains) and in the Jura and South (rather mountainous, border regions: coverage mainly by only one radar). The two combination methods reduce these regional differences in performance to a certain extent, but still show the same pattern of error measures.

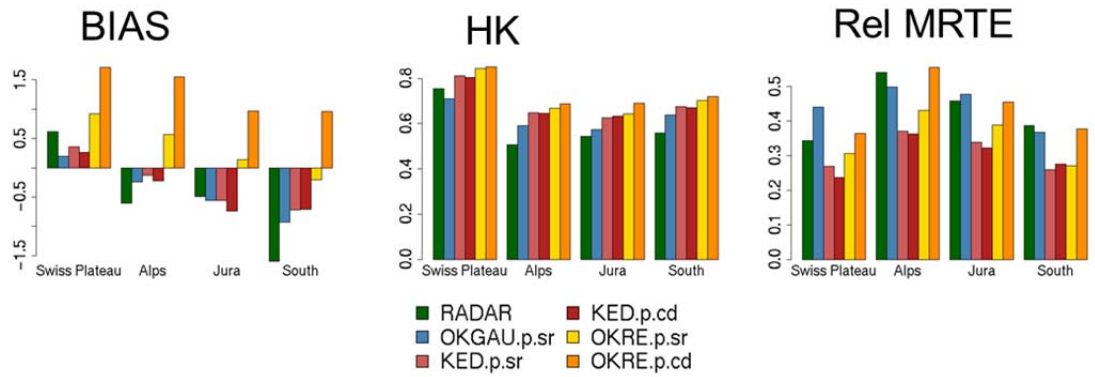

Figure 5: Error measures of the systematic application over the year 2008, based on 4774 hours and stratified for four geographical regions of Switzerland (see Fig. 1). Comparison of RADAR, OKGAU with prescribed square root transformation and KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

Seasonally stratified error measures are shown in Figure 6. In general, there are only small seasonal variations. The combination methods and RADAR show slightly better error measures in summer and autumn than in spring and winter. This seasonal pattern is not observed for OKGAU. We suggest that this seasonal differences can be explained by the predominance of convective precipitation in summer, opposed to predominantly stratiform precipitation in other seasons: Convective precipitation systems are characterized by more small scale variability of precipitation patterns. This leads to a relative advantage of RADAR due to its good spatial coverage. Moreover, convective precipitation systems reach higher altitudes which mitigates visibility problems by RADAR.

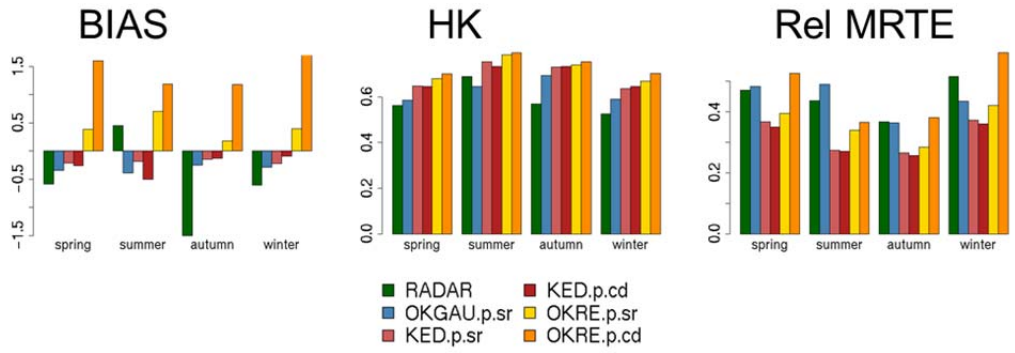

Figure 6: Error measures of the systematic application over the year 2008, based on 4774 hours and stratified for seasons (MAM, JJA, SON, DEF). Comparison of RADAR, OKGAU with prescribed square root transformation and KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

Based on the stratified analyses in Figure 4, Figure 5 and Figure 6 several comparisons are discussed in the following:

The comparison between combination methods and single sensor references shows an improvement of QPE by the combination compared to both, RADAR and OKGAU. This improvement is found for ample and poor sampling cases, for all regions and all seasons.

The comparison between the two combination methods shows a superiority of KED with respect to accuracy of total precipitation amount (*BIAS*) and mean error (*Rel MRTE*), and a superiority for OKRE with respect to wet-dry distinction (*HK*) in all regions and seasons. A pronounced difference in *BIAS*, especially for the poor sampling cases and the case-dependent transformation (*cd*), can be observed: there is an underestimation (negative *BIAS*) of total precipitation by KED and an overestimation (positive *BIAS*) by OKRE. This peculiarity of the two examined combination methods with respect to *BIAS* is discussed in more detail in the appendix (see Appendix A.1). OKRE can be considered as a special case of KED with a radar coefficient $\beta = 1$ (see section 3.1). In our application, the estimated radar coefficients in KED are typically between 0 and 1 (see Appendix, Fig. A.4). This means that there is usually an attenuation of the radar pattern in KED producing smoother QPE fields, compared to the more rugged fields by OKRE due to the one-to-one inclusion of radar. This characteristic difference can be observed in the illustrative example fields in Figure 2 and Figure 3. It provides an explanation for the described relative strengths of the two combination methods: The smoother KED yields a better mean error, whereas the more rugged OKRE performs better in wet-dry distinction.

The comparison between the prescribed square root (*sr*) and the estimated, case-dependent (*cd*) transformation shows no large influence on the point estimate in general. However, there is one exception, namely a much stronger positive *BIAS* in OKRE by the case-dependent transformation. This difference in overestimation is less pronounced in the Swiss Plateau and in summer than in the other regions and seasons. Note that a strong *BIAS* also leads to large mean errors (*Rel MRTE*). The difference in *Rel MRTE* between the two transformation settings of OKRE is presumably attributed to this effect. For the poor sampling cases, a similar effect is found for KED but with reversed signs: Here, the case-dependent transformation produces a stronger negative *BIAS* than the prescribed square root transformation. An explanation and discussion of this transformation-dependent behavior of *BIAS* can be found in the appendix (see Appendix A.1).

What is the influence of radar quality on the performance of combination methods? The stratified evaluations in Figure 5 and Figure 6 illustrate that radar quality is particularly high in the Swiss Plateau (good visibility due to relatively flat topography) and in summer (good visibility due to typically high vertical extent of precipitation systems). For instance, radar performs better in this region and season than OKGAU, but comparable or worse in other regions and seasons. The combination methods seem to profit from this high radar quality: The improvement by the combined QPE compared to OKGAU is particularly large in the Swiss Plateau and in summer.

4.2 Sensitivity on precipitation intensity

The accuracy and representation of intense precipitation is of particular interest for several applications of gridded QPE, such as hydrological flood forecasting. This section presents an

analysis on the performance of KED and OKRE with respect to observed intense precipitation above three predefined thresholds: 1 mm, 2 mm and 5 mm (hourly totals). Figure 7 shows the error measures *BIAS* and *MRTE* calculated from all cross-validation estimates for hours and gauges where intense precipitation above these thresholds has been measured. We prefer to look at *MRTE* instead of *Rel MRTE* here, because the MRTD (see Eq. 6) of these subsamples conditioned on a precipitation threshold provides no intuitive meaning.

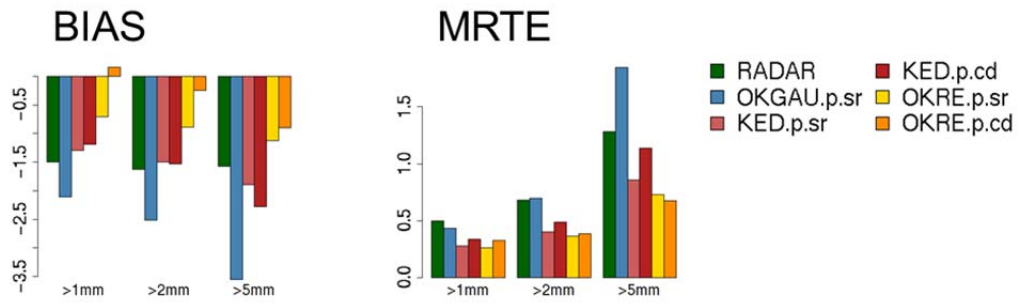

Figure 7: Error measures of the systematic application over the year 2008, for observations exceeding 1mm (26673 observations, 7.6%), 2mm (12391 observations, 3.5%) and 5mm (2211 observations, 0.6%). Comparison of RADAR, OKGAU with prescribed square root transformation and KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

The negative values of *BIAS* indicate a systematic underestimation of total precipitation amounts for high intensity rainfalls by all QPE methods. And this underestimation increases with increasing precipitation threshold. These characteristics are expected, because of the conditioning on a precipitation threshold (a best estimate underestimates the largest and overestimates the smallest observations on average). The slightly positive value of *BIAS* for observations above 1 mm by OKRE.p.cd is the result of compensatory effects of the usually considerable positive *BIAS* by OKRE.p.cd (see 4.1 and Appendix A.1) and the underestimation because of the conditioning. The *MRTE* shows a decrease in accuracy with increasing precipitation threshold. To some extent, this decrease in *MRTE* is caused by the increase in *BIAS* (a mean error can be decomposed into systematic error and random error). Another part of this decrease can presumably be attributed to the fact that *MRTE* is based on square root transformed observations and estimates (see Eq. 4), whereas estimated transformation parameters λ_e to stabilize the variance for hourly precipitation are presumably rather somewhere between a logarithmic ($\lambda = 0$) and a square root ($\lambda = 0.5$) transformation (Erdin et al., 2012). Or the other way around: The observed increase in *MRTE* corroborates the supposition that suitable values of λ_e are typically below 0.5.

The comparison of combination methods and single sensor references in Figure 7 shows a better performance, i.e. less underestimation (*BIAS*) and smaller mean errors (*MRTE*) by the combined QPE. When comparing the two combination methods with each other, OKRE provides consistently better results than KED. On the one hand, the less negative values of *BIAS* by OKRE can be the result of compensatory effects: We have seen a tendency of OKRE to overestimate precipitation in general (see 4.1), which could mitigate the general underestimation by the conditioning to intense precipitation. On the other hand, this could also be partly attributed to the one-by-one inclusion of radar in OKRE compared to the attenuated inclusion of radar in KED (see 3.1): Intense precipitation

values are often local maxima in the precipitation field. If the radar detects such a precipitation maximum in between the gauges, its representation is expected to be better in OKRE due to its unattenuated inclusion of radar. There seems to be no clear influence by the transformation on the accuracy of intense precipitation: The two transformation settings produce comparable results in general. However, the transformation-dependence of *BIAS* by OKRE is still observed in Figure 7, but at shifted scale (to negative overall values) because of compensatory effects as discussed in the preceding paragraph.

4.3 Probabilistic estimates by parametric methods

In the following, we assess the reliability of the probabilistic estimates by the geostatistical combination methods. This means that we are interested not only in the best estimate (point estimate), but also in the estimate of its uncertainty. The single sensor reference RADAR cannot be included for these analyses as it provides no information about the uncertainty. The characteristics of the probabilistic estimate by hourly KED, in particular its sensitivity to data transformation, has been the focus of a precursor study (Erdin et al., 2012). Therefore, the topic is only touched briefly here and the reader is referred to the mentioned publication for more details.

In Figure 8 the relative frequencies of rain-gauge measurements to fall in between two quantiles of the probabilistic estimate are shown for 14 predefined inter-quantile bins (for details see section 3.6). This reliability assessment by relative frequencies is most easily explained by a numerical example: The second (from left) inter-quantile bin in Figure 8 ranges from the 0.5% quantile (0.005 cumulative probability) to the 2.5% quantile. In a fully reliable system, 2% of all observations would fall in this inter-quantile bin. This is the expected frequency. The observed frequency, i.e. the number of all observations falling in this inter-quantile bin of their pertinent cross-validated probabilistic estimate with respect to the total number of observations in the sample, is now compared to this expected frequency. Relative frequencies above 1 indicate an observed frequency of more than 2%, relative frequencies below 1 an observed frequency below 2% for this exemplary inter-quantile bin. Note, that the relative frequencies (y-axis in Fig. 8) are plotted on logarithmic scale for comparability of positive and negative deviations of the horizontal line at 1 labelling a perfectly reliable system. Quantiles of the probabilistic estimate below 0.05 mm have been set to zero previously to the calculation of the relative frequencies. This corresponds to the discretization level of the rain gauges (0.1 mm) and helps to avoid irrelevant but confusing influences at the lower end of the probabilistic estimate (if the lowermost quantiles of a probabilistic estimate at locations of dry rain gauges are very small, but not exactly zero, the observation would always fall below the lowermost quantile).

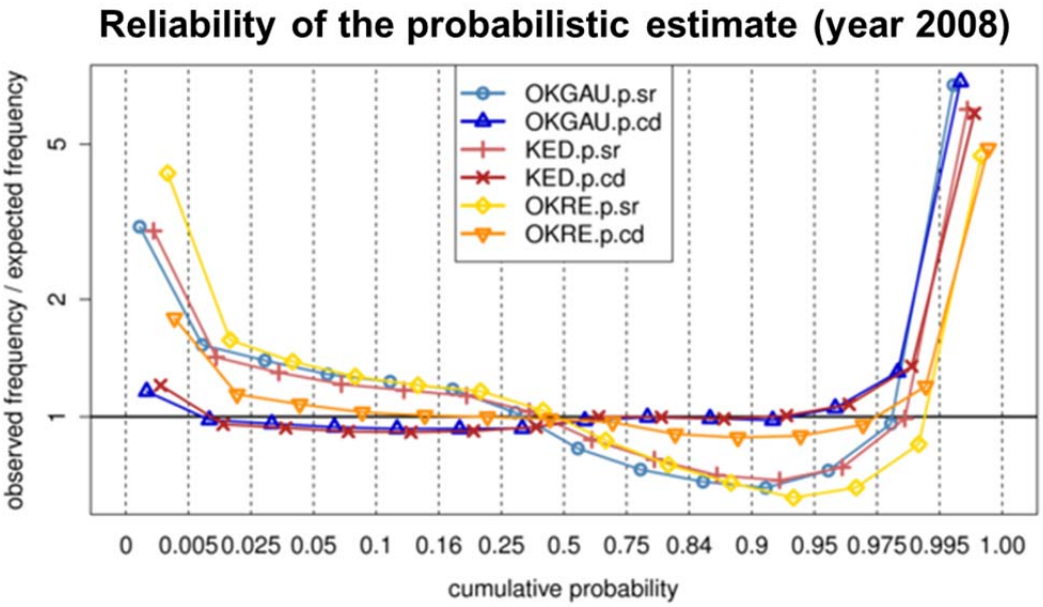

Figure 8: Frequency of rain-gauge measurements to fall in inter-quantile bins of the pertinent cross-validated probabilistic estimate (cumulative probability, x-axis). Frequencies are expressed relative to the expected frequency in a perfectly reliable system (y-axis, log-scale). Bins of the distribution are defined by the 0.5%, 2.5%, 5%, 10%, 16%, 25%, 50%, 75%, 84%, 90%, 95%, 97.5%, 99.5% quantiles (dashed vertical lines). Results from the systematic application over the year 2008 (4774 hours) comparing OKGAU, KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

Figure 8 shows u-shaped curves for all QPE methods. I.e. there are too many observations in the extreme and too few observations in the intermediate inter-quantile bins. The PDF of the probabilistic estimate is, hence, in general too narrow and too many observations are falling outside this distribution (below the 0.5% or above the 99.5% quantile). We can say that there is a general overconfidence by both combination methods and OKGAU. However, there are differences between the six examined QPE methods. The strongest influence is found for the transformation setting: The case-dependent transformation produces considerably more reliable probabilistic estimates than the square root transformation throughout the PDF, except for the two uppermost inter-quantile bins with similar performance for both transformation settings. The influence of the method, on the other hand, is much smaller: The relative frequencies of KED, OKRE and OKGAU with the same transformation setting are very similar in all intermediate inter-quantile bins. However, KED (and OKGAU) has a better reliability in the lowermost, OKRE a better reliability in the uppermost inter-quantile bin. An explanation for this could be the typically sharper pattern by OKRE due to its unattenuated inclusion of radar (see 3.1). This leads to smaller point estimates at dry and larger estimates at wet locations than KED. Owing to the applied data transformation, the uncertainty estimate is positively dependent on precipitation intensity (Erdin et al., 2012). OKRE is, hence, supposed to produce more narrow probabilistic estimates at dry locations, but wider probabilistic estimates at very wet locations than KED. As a consequence, more observations would fall below the lowermost quantile (dry or almost dry cases) and less above the uppermost quantile (intense precipitation) for OKRE than for KED.

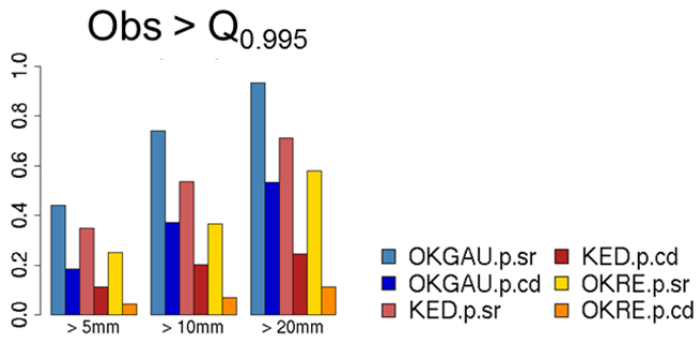

Figure 9: Fraction of observations of intense precipitation considered as extremely unlikely (falling above the 99.5% quantile, i.e. p=0.005) by the pertinent probabilistic estimate in the systematic application over the year 2008. This includes 2211 observations > 5mm, 363 observations > 10mm and 45 observations > 20mm. Comparison of OKGAU, KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

For practical applications (e.g. hydrological flood forecasting), a good reliability of the probabilistic estimate is particularly important with respect to intense precipitation amounts. Figure 9 shows the fraction of observed intense precipitation above three threshold (>5 mm, >10 mm and >20 mm) falling above the 99.5% quantile of their pertinent cross-validated probabilistic estimate. The notion is that these observations are considered as “extremely unlikely” by the probabilistic estimates ($p<0.005$). Nevertheless, they have been measured. One could therefore say, these observations are “missed” by the probabilistic estimate. In an unconditioned sample, a perfectly reliable estimate would result in a fraction of 0.005. The conditioning to a threshold justifies larger fractions of misses (not straightforward to quantify). Excessive fractions of misses, however, considering a large part or even the majority of actually observed intense precipitation amounts as extremely unlikely, are certainly undesirable.

The observed fractions in Figure 9 increase with increasing threshold and range from 4.3% (OKRE.p.cd, >5mm) to 93.3% (OKGAU.p.sr, >20mm).The transformation setting has a considerable influence on the fraction of observed intense precipitation considered as extremely unlikely: The case-dependent transformation yields much better (lower) fractions, than the square root transformation. The negligibly small influence of the transformation setting on the reliability in the uppermost inter-quantile bin observed in Figure 8 is therefore only valid up to moderate precipitation amounts. As moderate (and smaller) precipitation amounts constitute the largest part of the sample, they dominate the overall frequencies shown in Figure 8. There is a consistent hierarchy of the three QPE methods for all thresholds: OKGAU yields the largest fractions, followed by KED, whereas OKRE produces the smallest fractions. Our interpretation of this hierarchy is that radar helps to improve the reliability here, because intense precipitation amounts are often spatial maxima of precipitation, which cannot be seen by gauges in a cross-validation mode. The superiority of OKRE over KED is then consistent, as KED includes usually an attenuated radar pattern compared to the one-by-one inclusion by OKRE. However, part of this superiority of OKRE could again be a side effect of its tendency to overestimate precipitation in general (see preceding subsections 4.1 and 4.2).

4.4 Parametric vs. non-parametric modeling of covariance structure

For the comparison of the parametric and non-parametric covariance estimation (see 3.2), an application to a set of 12 representative test cases (see section 2) is employed. Full cross-validation for the non-parametric setting is computationally very demanding, making an application over a full year unfeasible. (However, the calculation of the QPE field without cross-validation is comparable with the parametric variogram estimation in terms of computational costs.) The pooled test cases consist of 222 hourly accumulations, 144 with ample sampling and 78 with poor sampling conditions (this is a larger fraction of ample sampling than in the systematic application, owing to the selection of interesting precipitation events).

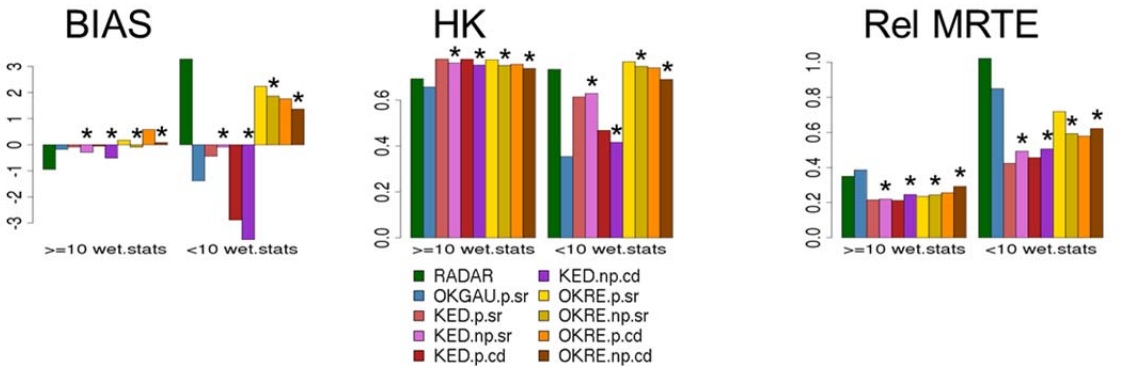

Figure 10: Error measures of the systematic application over a set of 12 test cases with 144 hours with ample sampling (>=10 wet.stats) and 78 hours with poor sampling (<10 wet.stats). Comparison of RADAR, OKGAU with prescribed square root transformation and parametric and non-parametric KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*). Non-parametric applications are labelled with stars.

In Figure 10, the error measures based on the pooled point estimates by cross-validation over all 12 test cases are displayed. The grouped barplots show stratified error measures for ample sampling (>=10 wet.stats) and poor sampling (<10 wet.stats) hours for the following QPE methods: KED and OKRE using parametric and non-parametric covariance modeling and both transformation settings. All QPE based on non-parametric correlogram estimation are labeled with stars for better lucidity. RADAR and OKGAU.p.sr are included as single sensor references as in the preceding subsections. The combined QPE show a stable improvement compared to the single sensor references, which is consistent with the results of the broader application (see 4.1). The comparison between the parametric and the non-parametric covariance estimation approach shows a similar performance. However, the parametric variogram estimation obtains slightly better error measures in general. And this difference is found not only for ample sampling cases, where estimation of all model parameters is possible, but also for the poor sampling cases, where the parametric approach relies on climatological parameters. This comes as a surprise, since we hypothesized a relative advantage of the non-parametric approach for the poor sampling cases due to its unrestricted applicability in all precipitation situations. Based on these findings, we assume that the difference in performance between ample and poor sampling cases observed in the broader application (see Fig. 4) is mainly owing to the challenging precipitation situation per se, and not primarily to the fact that climatological parameters need to be employed for the poor sampling cases. The comparison between the two combination methods exhibits the same characteristics as in the year-long application: Better

performance of KED with respect to mean error (*Rel MRTE*) and better performance of OKRE with respect to wet-dry distinction (*HK*). And a distinct pattern of *BIAS* for poor sampling cases with an underestimation of precipitation by KED and an overestimation by OKRE. Only a small influence of the transformation setting on the point estimate is observed. However, in poor sampling hours, the case-dependent transformation setting produces a pronounced negative *BIAS* and more errors in the wet-dry distinction than the square root transformation in KED. An explanation for this transformation dependent behavior can be found in the appendix (see Appendix A.1.3). Figure 10 compares combined QPE configurations varying in three dimension: combination methods, covariance modeling and transformation setting. It can be seen that in our application setting, the choice of combination method (KED or OKRE) and transformation (*sr* or *cd*) have a stronger impact on the point estimate than the choice of covariance modeling approach (*p* or *np*).

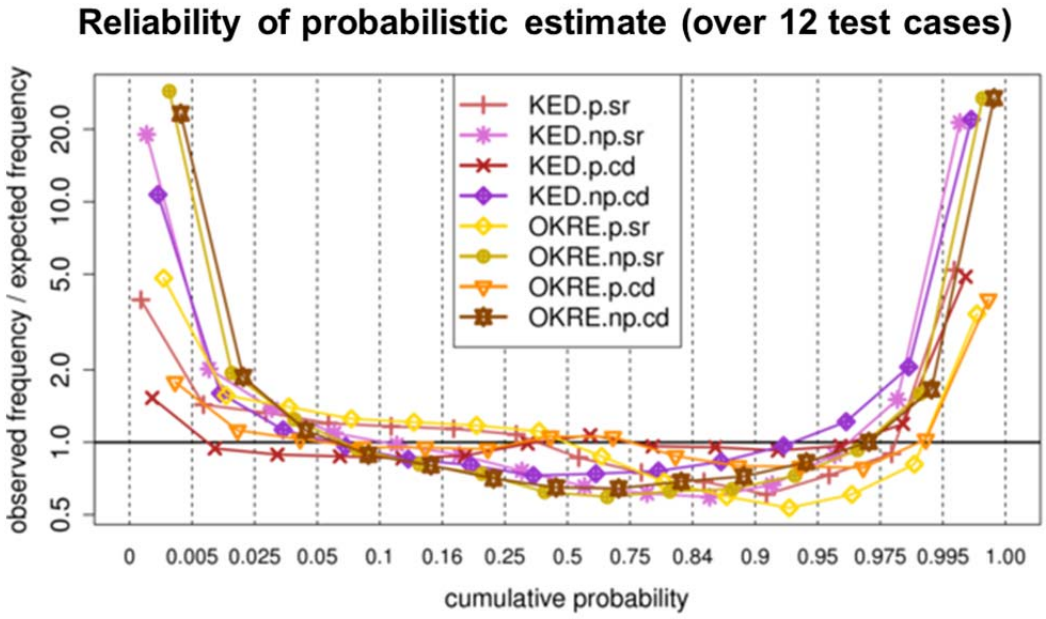

Figure 11: Frequency of rain-gauge measurements to fall in inter-quantile bins of the pertinent cross-validated probabilistic estimate (for detailes see legend Fig. 8). Results from the systematic application over a set of 12 test-cases (222 hours) comparing parametric and non-parametric KED and OKRE with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

We now turn to comparing the parametric variogram and the non-parametric correlogram estimation with respect to the reliability of the probabilistic estimate. Figure 11 displays the relative frequencies of observations to fall in between predefined inter-quantile bins and Figure 12 the fractions of observed intense precipitation considered as extremely unlikely by the pertinent cross-validated probabilistic estimates (for details see preceding section 4.3, Fig. 8 and Fig. 9). Frequencies and fractions based on the pooled results of all test cases are shown for KED and OKRE with both covariance modeling approaches and both transformation settings. Again, a general overconfidence (U-shape in Fig. 11 and relatively large fractions in Fig. 12, cf. 4.3) by all QPE methods can be observed. However, there is a pronounced difference between the two covariance modeling approaches: The probabilistic estimates by methods employing parametric variogram estimation are much more reliable than those by methods employing non-parametric correlograms (note the log

scale in Fig. 11!). Schiemann et al. (2011) found a bias in the estimated sill by the non-parametric approach. Although this is corrected in our application according to their suggestion, the estimated variance of the random process is possibly still too small. This would provide an explanation for the strong overconfidence by methods employing non-parametric correlogram estimation. Among the parametric methods, the same pattern than in the broader systematic analysis (see section 4.3 and Fig. 8) is observed. Namely, a better reliability by the case-dependent transformation than by the prescribed square root transformation and a superiority of OKRE compared to KED for the uppermost inter-quantile bin in general and with respect to intense precipitation (see Fig. 12). The non-parametric methods, in contrast, produce a pronounced U-shape and unrealistically large fractions of missed intense precipitation for both combination methods and transformation settings. With respect to the reliability of the probabilistic estimate, the influence of the covariance modeling approach is larger than the influence of the choice of combination method or transformation setting.

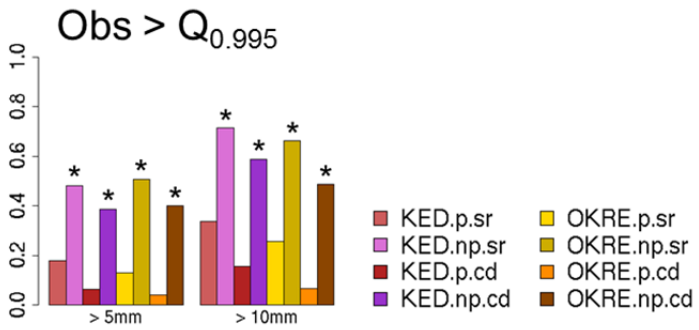

Figure 12: Fraction of observations of intense precipitation considered as extremely unlikely (falling above the 99.5% quantile, i.e. p=0.005) by the pertinent probabilistic estimate in the systematic application over a set of test cases. This includes 691 observations > 5mm and 136 observations > 10mm. Comparison of parametric and non-parametric KED and OKRE with two different transformation settings each: λ=0.5 (*sr*) and $\lambda = \lambda_e$ (*cd*). Non-parametric applications are labelled with stars.

In summary, the comparison between the non-parametric correlogram and the parametric variogram estimation shows no large differences with respect to the point estimate in the examined framework. However, a much better reliability of the probabilistic estimate by the parametric than by the non-parametric methods is found. And this is even the case for situations with poor sampling conditions, where an advantage of non-parametric correlogram estimation from radar could have been expected.

5 Conclusions

This study presents a systematically applicable framework for hourly real-time combination of radar and rain gauge data. It is based on two existing geostatistical methods (KED and OKRE), which are applied in a trans-Gaussian setting to mitigate the violation of model assumptions. Climatological parameters for the classical parametric variogram and an alternative non-parametric correlogram estimation are employed to deal with situations when only very few stations are hit by rainfall leading to robustness problems in classical parameter estimation. A systematic application of the two combination methods over a full year (2008) for the domain of Switzerland is performed. Full leave-one-out cross-validation is applied to analyze the performance of the combined spatial precipitation analyses. The added value of the combination is assessed by comparing the combination to its parents: the radar composite and interpolated rain gauge measurements. On the one hand, the estimated precipitation by the combination methods is analyzed with respect to technical choices and regional and seasonal patterns. On the other hand, also the pertinent uncertainty estimates are assessed on their reliability to reflect the true deviations of observations from the precipitation estimates.

A general improvement by both combination methods is found compared to the spatial estimates of precipitation based on information from either radar or rain-gauges only. And this improvement is stable throughout several stratified analyses: for hours with abundant and with sparse sampling of precipitation by the gauges, for different geographical regions and in all seasons. Situations with relatively high accuracy of the radar data have been found beneficial for the performance of the combination methods. These findings are in line with the results of two related studies performed recently at MeteoSwiss: The comparison of the two geostatistical combination methods presented here with a more sophisticated approach including one past time step in the model (Sideris et al., accepted) for the hourly time scale by Keller (2012) and the systematic application of KED for the daily time scale by Vogel (2013).

The accuracy of the combination methods, in general, strongly depends on the precipitation situation. For instance, on the spatial variability, the sampling by rain gauges and the capture by radar of the precipitation patterns. This leads to marked differences in accuracy in different regions and seasons which are much larger in extent than differences between the two combination methods or between different technical options (data transformation, covariance estimation). Nevertheless, some characteristic differences between the two combination methods have been found in the comparison with each other: kriging with external drift (KED) is better in terms of bias and mean error and ordinary kriging of radar errors (OKRE) is better in terms of wet-dry distinction and representation of intense precipitation. This can be explained by the fact that radar information is usually attenuated in KED (radar coefficient typically between 0 and 1), but always included one-to-one in OKRE. The

precipitation fields are, hence, smoother by KED, favoring mean error and bias, and more sharp and rugged by OKRE, favoring wet-dry distinction and precipitation maxima.

In general, no large influence of the data transformation setting on the spatial precipitation analyses has been found. Nevertheless, with respect to the accuracy of the overall precipitation amount (bias), there are characteristic transformation-dependent patterns, namely a larger overestimation in OKRE and a larger underestimation in KED by the case-dependent transformation than by the prescribed square root transformation. This is presumably due to a transformation-parameter-dependent inflation of the uppermost fraction of the precipitation estimate in the back-transformation process (see Appendix A.1). In consequence, a prescribed square root transformation might be preferred over a data-driven transformation if exclusively best estimates (point estimates) of precipitation are of interest. However, the case-dependent transformation leads to considerably more reliable quantifications of the inherent analysis uncertainties than the prescribed square root transformation, which systematically underestimates the extent of the uncertainty (overconfidence). This corroborates the findings of our precursor study (Erdin et al., 2012), where a data driven, case-dependent transformation has been found important for the reliability of the probabilistic information.

The reliability of the uncertainty (probabilistic estimate) is important for applications which aim at incorporating the uncertainty of the input data. Thereby they can obtain a probabilistic output for their part. One way to pursue the uncertainty of the spatial precipitation analyses in application models is to use precipitation ensembles as model input, for instance for hydrological modeling (e.g. Zappa et al., 2008; Germann et al., 2009). A preliminary attempt to construct such precipitation ensembles from geostatistical radar-gauge combination by KED at the daily time scale has been recently undertaken at MeteoSwiss (Vogel, 2013). The promising results of these experiments, together with the good comparability of KED for the hourly and the daily time scale, suggest that it would be worthwhile to test the construction of hourly precipitation ensembles from radar-gauge combination. Therein, the temporal consistence of the ensemble should additionally be considered, since, in contrast to the daily time scale, the temporal correlation of precipitation is important on the hourly time scale (e.g. Rakovec et al., 2012).

The comparison between the parametric variogram and the non-parametric correlogram estimation shows a fairly comparable performance of the spatial precipitation analyses, even in situations when only a few rain gauges sample precipitation and the parametric version therefore relies on prescribed parameters. The uncertainty estimate, however, is found to be much more reliable if the classical parametric variogram estimation is applied. There is a strong underestimation of the uncertainty (overconfidence) by the non-parametric approach for the presented application in Switzerland. However, the non-parametric approach has only been proposed and analyzed recently and further developments might lead to more competitive applications.

In summary, both analyzed geostatistical methods can be recommended for automatic real-time combination of radar and rain gauges at the hourly time-scale, also in mountainous regions as Switzerland. If the main focus lies in a good delineation of wet and dry areas and a good representation of precipitation maxima, OKRE may be slightly preferred. If small mean-errors and a good accuracy of the overall estimated precipitation amount are important, KED may be more recommended. The classical parametric variogram estimation with predefined climatological parameters for hours with sparse data is preferable over the non-parametric approach in the

presented framework. As concerns data transformation, a prescribed square root transformation is adequate, if only deterministic information on precipitation (point estimates) is of interest. If also probabilistic information is desired, we advise to opt for a data-driven, case-dependent transformation. Our analyses suggest that the quality of the radar data is crucial for a high accuracy of the spatial precipitation analyses by radar-gauge combination. The combination, hence, does not render an amplification of the radar network or further developments in the radar post-processing redundant. On the contrary, improvements in the accuracy of radar data by such measures are expected to have a beneficial effect on the combined spatial precipitation analyses.

Beside the added value by the combination of radar and rain gauge data, our analyses demonstrate also shortcomings of the geostatistical approach in the presented framework: On the one hand, there is a remaining overconfidence of the probabilistic estimate, even though the reliability can be substantially improved by an adequate (case-dependent) data transformation. The intermittent nature of precipitation and dependencies of the uncertainty other than the modeled dependency on precipitation intensity (cf. Erdin et al., 2012) are supposable reasons for this overconfidence. On the other hand, the case-dependent transformation produces biases in the total estimated precipitation amount. The choice of transformation setting might thus result in a trade-off between the point and the probabilistic estimate. This is probably reflecting that, even though made more compliant by data transformation, precipitation data does not really match the assumptions of geostatistical models (cf. Erdin et al., 2012). Therefore, we suggest further advancements in statistical combination of radar and rain-gauge data outside the classical geostatistical framework. For instance we would expect beneficial results by approaches explicitly accounting for intermittency (e.g. Seo, 1998; Fuentes et al., 2008) and/or temporal dependencies (e.g. Fuentes et al., 2008; Sideris et al., accepted). The aspiration should be to find models better suitable for precipitation data and thereby improve both, the deterministic and the probabilistic information on precipitation.

Abbreviations

BUF	Rain gauge operated by MeteoSwiss in Buffalora (Grisons)
Est	Estimate
FFT	Fast Fourier Transform
KED	Kriging with external drift
MLE	Maximum Likelihood Estimation
Obs	Observation
OK	Ordinary kriging
OKGAU	Ordinariy kriging of rain gauges
OKRE	Ordinary kriging of radar errors
PDF	Probability density function
QPE	Quantitative precipitation estimate
RMLE	Restricted Maximum Likelihood Estimation
SCU	Rain gauge operated by MeteoSwiss in Scuol (Grisons)
UTC	Coordinated universal time

References

Bartels, H., E. Weigl, T. Reich, P. Lang, A. Wagner, O. Kohler, and N. Gerlach, 2004: Zusammenfassender Abschlussbericht zum Projekt RADOLAN.

Box, G. E. P., and D. R. Cox, 1964: An Analysis of Transformations. *J. Roy. Stat. Soc.*, **26,** 211-252.

Brandes, E. A., 1975: Optimizing Rainfall Estimates with the Aid of Radar. *J. Appl. Meteor.*, **14,** 1339-1345.

Ciach, G., and W. Krajewski, 1999: On the estimation of radar rainfall error variance. *Advances in Water Resources*, **22,** 585-595.

Cressie, N. A. C., 1993: *Statistics for Spatial Data.* Wiley, New York, 900 pp.

Creutin, J. D., G. Delrieu, and T. Lebel, 1988: Rain Measurement by Raingage-Radar Combination: A Geostatistical Approach. *J. Atmos. Oceanic Technol.*, **5,** 102-115.

DeGaetano, A. T., and D. S. Wilks, 2008: Radar-guided interpolation of climatological precipitation data. *Int. J.Climatol.***,** DOI: 10.1002/joc.1714.

Diggle, P. J., and P. J. Ribeiro Jr., 2007: *Model-based Geostatistics.* Springer Science + Business Media, New York, 228 pp.

Erdin, R., 2009: Combining Rain Gauge and Radar Measurements of a Heavy Precipitation Event over Switzerland: Comparison of geostatistical methods and investigation of important influencing factors. *Veröffentlichungen der MeteoSchweiz,* **81,** 108pp.

Erdin, R., C. Frei, and H. R. Künsch, 2012: Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *J. Hydrometeor.*, **13,** 1332-1346.

Fabry, F., A. Bellon, M. R. Duncan, and G. L. Austin, 1994: High-resolution rainfall measurements by radar for very small basins: the sampling problem reexamined. *J. Hydrol.*, **161,** 415-428.

Frei, C., U. Germann, S. Fukutome, and M. Liniger, 2008: Möglichkeiten und Grenzen der Niederschlagsanalyse zum Hochwasser 2005.

Fuentes, M., B. Reich, and G. Lee, 2008: Spatial-Temporal Mesoscale Modeling of Rainfall Intensity Using Gage and Radar Data. *Ann. Appl. Stat.*, **2,** 1148-1169.

Garcia-Pintado, J., G. Barbera, M. Erena, and V. Castillo, 2009: Rainfall estimation by rain gauge-radar combination: A concurrent multiplicative-additive approach. *Water Resour. Res.*, **45,** -.

Germann, U., and J. Joss, 2002: Mesobeta profiles to extrapolate radar precipitation measurements above the Alps to the ground level. *J. Appl. Meteor.*, **41,** 542-557.

Germann, U., G. Galli, M. Boscacci, and M. Bolliger, 2006: Radar precipitation measurement in a mountainous region. *Quart. J. Roy. Meteor. Soc.*, 1669-1692.

Germann, U., M. Berenguer, D. Sempere-Torres and M. Zappa, 2009: REAL – Ensemble radar precipitation estimation for hydrology in a mountainous region. *Quart. J. Roy. Meteor. Soc.,* **135,** 445-456.

Gjertsen, U., M. Salek, and D. B. Michelson, 2004: Gauge adjustment of radar-based precipitation estimates in Europe. *European Conference on radar in Meteorology and Hydrology (ERAD)*, Visby, Sweden, 7-11.

Goudenhoofdt, E., and L. Delobbe, 2009: Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrol. Earth Syst. Sci.*, 195-203.

Haberlandt, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol.*, **332,** 144-157.

Hamill, T. M., 2001: Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon. Wea. Rev.*, **129,** 550-560

Jones, C. D., and B. Macpherson, 1997: A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorol. Appl.*, **4,** 269-277.

Jurczyk, A., K. Osrodka, and J. Szturc, 2007: Research studies on improvement in real-time estimation of radar-based precipitation in Poland. *Meteor. Atmos. Phys.*, DOI 10.1007/s00703-00007-00266-00703.

Keller, D., 2012: Evaluation and comparison of radar-rain gauge combination methods: Scientific basis for selecting a new real-time precipitation estimation method for Switzerland, Master Thesis ETH Zurich at MeteoSwiss, 68 pp.

Krajewski, W. F., 1987: Cokriging Radar-Rainfall and Rain Gage Data. *J. Geophys. Res.*, **92,** 9571-9580.

Künsch, H. R., A. Papritz, C. Schwierz, and W. Stahel, 2011: Robust estimation of the external drift and the variogram of spatial data. *ISI 58th Statistics Congress of the International Statistical Institute*.

Leuenberger, D., and A. Rossa, 2007: Revisiting the latent heat nudging scheme for the rainfall assimilation of a simulated convective storm. *Meteor. Atmos. Phys.*, 195-215.

MeteoSchweiz, 2006: Starkniederschlagsereignis August 2005, 63 pp pp.

MeteoSwiss, 2010: SwissMetNet Newsletter Nr. 7, *www.meteoswiss.admin.ch*: http://www.meteoswiss.admin.ch/medialib/documents/swissmetnet.Par.0001.File.tmp/smnnewsletter7de.pdf

Musa, M., E. Grüter, M. Abbt, C. Häberli, E. Häller, U. Küng, T. Konzelmann and R. Dösegger, 2003: Quality control tools for meteorological data in the MeteoSwiss data warehouse system. *Proc. ICAM/MAP 2003*, May 19th - 23rd 2003, Brig, Switzerland.

Neff, E. L., 1977: How much rain does a rain gage gauge? *J. Hydrol.*, **35,** 213-220.

Oswald, M., M. Perl, S. Pfahl, 2010: Regionen der Schweiz: Übersicht zu den bei MeteoSchweiz verwendeten Regionseinteilungen. *MeteoSwiss internal document*, 48 pp.

Overeem, A., T. Buishand, and I. Holleman, 2009: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.*, **45,** -.

Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, **4,** 453-454.

Pereira Fo, A., and K. Crawford, 1999: Mesoscale precipitation fields. Part I: Statistical analysis and hydrologic response. *Journal of Applied Meteorology*, **38,** 82-101.

R Development core Team, 2009: R Foundation for Statistical Computing.

Rakovec, O., P. Hazenberg, P. J. J. F. Torfs, A. H. Weerts and R. Uijlenhoet, 2012: Generating spatial precipitation ensembles: impact of temporal correlation structure. *Hydrol. Earth Syst. Sci.*, **16,** 3419-3434.

Ribeiro Jr., P. J., and P. J. Diggle, 2001: geoR: a package for geostatistical analysis. *R-NEWS*, **1,** 15-18.

Rubel, F., and B. Rudolf, 2001: Global daily precipitation estimates proved over the European Alps. *Meteorologische Zeitschrift*, **10,** 407-418.

Rudolf, B., H. Hauschild, W. Rüth, and U. Schneider, 1996: Comparison of raingauge analyses, satellite-based precipitation estimates and forecast model results. *Adv. Space Res.*, **18,** 53-62.

Schabenberger, O., and C. A. Gotway, 2005: *Statistical Methods for Spatial Data Analysis.* Chapman & Hall/CRC, 488 pp.

Schiemann, R., R. Erdin, M. Willi, C. Frei, M. Berenguer, and D. Sempere-Torres, 2011: Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland. *Hydrol. Earth Syst. Sci.*, **15,** 1515-1536.

Schneider, S., and R. Steinacker, 2009: Utilisation of radar information to refine Precipitation fields by a variational approach. *Meteor. Atmos. Phys.*, **103,** 137-144.

Schuurmans, J. M., M. F. P. Bierkens, E. J. Pebesma, and R. Uijlenhoet, 2007: Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar. *J. Hydrometeor.*, **8,** 1204-1224.

Seo, D., and J. Breidenbach, 2002: Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J. Hydrometeor.*, **3,** 93-111.

Seo, D.-J., 1998: Real-time estimation of rainfall fields using radar rainfall and rain gage data. *J. Hydrol.*, **208,** 37-52.

Sideris, I. V., M. Gabella, R. Erdin, and U. Germann, accepted: Real-time radar-raingauge merging using spatiotemporal co-kriging with external drift in the alpine terrain of Switzerland. *Quart. J. Roy. Meteor. Soc.*

Sinclair, S., and G. Pegram, 2005: Combining radar and rain gauge rainfall estimates using conditional merging. *Atmos. Sci. Lett.*, **6,** 19-22.

Tabary, P., 2007: The new French operational radar rainfall product. Part I: Methodology. *Weather Forecast.*, **22,** 393-408.

Talagrand, O., R. Vautard, and R. Strauss, 1997: Evaluation of probabilistic prediction systems. *ECMWF Workshop on Predictability*, Shinfield Park, Reading, United Kingdom, ECMWF, 1-25.

Todini, E., 2001: A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements. *Hydrol. Earth Syst. Sci.*, **5(2),** 187-199.

Velasco-Forero, C. A., D. Sempere-Torres, E. F. Cassiraga, and J. J. Gomez-Hernandez, 2009: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Adv. Water Resour.*, **32,** 986-1002.

Verworn, A., and U. Haberlandt, 2011: Spatial interpolation of hourly rainfall - effect of additional information, variogram inference and storm properties. *Hydrol. Earth Syst. Sci.*, **15,** 569-584.

Villarini, G., and W. Krajewski, 2010: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics*, **31,** 107-129.

Villarini, G., P. Mandapaka, W. Krajewski, and R. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *Journal of Geophysical Research-Atmospheres*, **113**.

Viviroli, D., M. Zappa, J. Schwanbeck, J. Gurtz, and R. Weingartner, 2009: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland - Part I: Modelling framework and calibration results. *J. Hydrol.*, **377,** 191-207.

Vogel, R., submitted: Quantifying the uncertainty of spatial precipitation analyses by merged radar-gauge observation ensembles. *Scientific Report MeteoSwiss,* **XX**, nn pp.

Westcott, N., S. Hollinger, and K. Kunkel, 2005: Use of real-time multisensor data to assess the relationship of normalized corn yield with monthly rainfall and heat stress across the central United States. *J. Appl. Meteor.*, **44,** 1667-1676.

Weusthoff, T., F. Ament, M. Arpagaus, and M. Rotach, 2010: Assessing the Benefits of Convection-Permitting Models by Neighborhood Verification: Examples from MAP D-PHASE. *Mon. Wea. Rev.*, **138,** 3418-3433.

Wüest, M., C. Frei, A. Altenhoff, M. Hagen, M. Litschi, and C. Schär, 2009: A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *Int. J. Climatol.*, DOI: 10.1002/joc.2025.

Xie, H., X. Zhang, B. Yu, and H. Sharif, 2011: Performance evaluation of interpolation methods for incorporating rain gauge measurements into NEXRAD precipitation data: a case study in the Upper Guadalupe River Basin. *Hydrol. Process.*

Yeung, H.-Y., C. Man, S.-T. Chan, and A. Seed, 2011: Application of radar-raingauge co-kriging to improve QPE and quality control of real-time rainfall data. *Weather Radar and Hydrology*, Exeter, UK.

Young, C., A. Bradley, W. Krajewski, A. Kruger, and M. Morrissey, 2000: Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting. *J. Hydrometeor.*, 241-254.

Zappa, M., and Coauthors, 2008: MAP D-PHASE: real-time demonstration of hydrological ensemble prediction systems. *Atmos. Sci. Lett.***,** 80-87.

Zhang, Y., S. Reed, and D. Kitzmiller, 2011: Effects of Retrospective Gauge-Based Readjustment of Multisensor Precipitation Estimates on Hydrologic Simulations. *J. Hydrometeor.*, **12,** 429-443.

Zimmer, M., and H. Wernli, 2011: Verification of quantitative precipitation forecasts on short time-scales: A fuzzy approach to handle timing errors with SAL. *Meteorologische Zeitschrift*, **20,** 95-105.

Acknowledgement

First of all, I would like to cordially thank Christoph Frei for his supervision, support and friendship during my work for CombiPrecip. I enjoyed the countless inspiring discussions we shared and his scientific expertise, precision and ingenuity have been truly instructive.
Secondly, I owe my gratitude to Prof Hansruedi Künsch from ETH Zurich for his patient and supportive guidance of our research project. Furthermore, I would like to thank all the persons at MeteoSwiss that have contributed to CombiPrecip and the work presented in this study in one way or another. Especially Ioannis Sideris, Reinhard Schiemann, Marco Willi, Urs Germann, Marco Gabella, Denise Keller, David Masson, Raphaela Vogel, Mark Liniger and Christof Appenzeller.

Appendix 1: Impact of Transformation on *BIAS*

Particular characteristics of all kriging methods (KED, OKRE and OKGAU) depending on the transformation setting have been observed with respect to overall precipitation amount (*BIAS*) in this study (cf. section 4). In the following subsections, we seek understanding of this behavior and describe the suppositionally underlying inflation mechanism.

A.1.1 Inflation effects by back-transformation

Our analyses suggest that the observed overestimation by OKRE (positive *BIAS*) and underestimation by KED and OKGAU (negative *BIAS*) (see section 4) are based on the same underlying mechanism. This mechanism consists of a transformation-dependent inflation effect of the uppermost fraction of a kriging estimate at a specific location, and a transformation-dependent shrinkage effect of the lowermost fraction accordingly. The general mechanism of this inflation/shrinkage effect is introduced by an artificial example in following.

The black curve in Figure A.1 illustrates the relations between a box-cox-transformed (λ = 0.2, y-axis) and the untransformed (λ = 1, x-axis) scale for an artificial example. Constant fractions of 2 out of 7 units at the lower and upper end at each scale respectively are sketched: in red for the transformed and in blue for the untransformed scale. This simulated experiment demonstrates the inflation (or shrinkage) effect by back-transformation, if one follows the red lines from the transformed y-axis back to the untransformed x-axis. If the uppermost 2/7 at transformed scale are back-transformed, they are heavily inflated with respect to the total amount at back-transformed (i.e. untransformed) scale: this fraction now ranges from 1 to 5.4 (i.e. accounts for almost 6 out of 7 units!) and is, hence, much larger than the respective fraction at untransformed scale (blue, ranging from 4 to 5.4). The opposite happens to the lowermost fraction of 2/7: Because of the very steep rise of the relation between the untransformed and the transformed scale close to the origin, this fraction shrinks to an illegibly small value (0.0102) when back-transformed. It hence shrinks dramatically compared to the 2/7 at untransformed scale sketched in blue (1.4).

Note that through the process of transformation, the opposite effect takes place: The lowermost fractions of a range are inflated, whereas the uppermost fractions are shrinked. This can be retraced by following the red lines in Figure A.1 in the opposite direction than above: from the untransformed x-axis to the transformed y-axis.

With this background of the effects of back-transformation on fractions at the upper and lower end of a range, we now turn to the impact of these effects on the point estimates of precipitation by our kriging techniques.

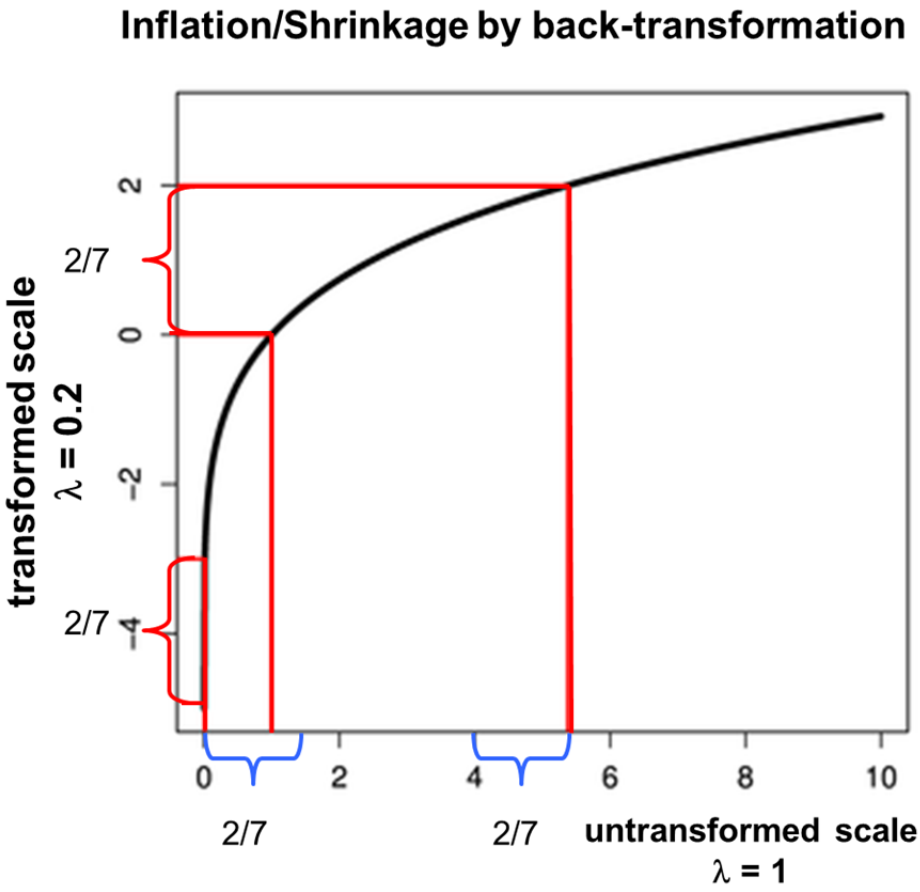

Figure A.1: Sketch to illustrate the inflation/shrinkage effect of fixed proportions by back transformation. A transformed scale with λ=0.2 is plotted against the untransformed scale (λ=1) and fixed proportions of 2/7 at the uppermost and lowermost end of the interval are compared for transformed and then back-transformed scale (red) and for the untransformed scale (blue).

A.1.2 Positive *BIAS* of OKRE

In the results of this study, an overestimation of total precipitation amount by OKRE (pos *BIAS*) has been observed. And this overestimation is considerably larger for the case-dependent transformation setting (*cd*) than for the prescribed square root transformation (*sr*) throughout all stratified analyses (see Figs. 4, 5 and 6). Various analyses with the data of the systematic application over the year 2008 to investigate this behavior have been undertaken (not shown here). These analyses suggest that this transformation-dependent, positive *BIAS* in OKRE is caused by the inflation of overestimated interpolated radar errors at wet locations. The overestimation of these radar error is caused by adjacent precipitation missed or severely underestimated by radar. In the following, this assumptive mechanism is explained by an example and corroborated by quantitative analyses.

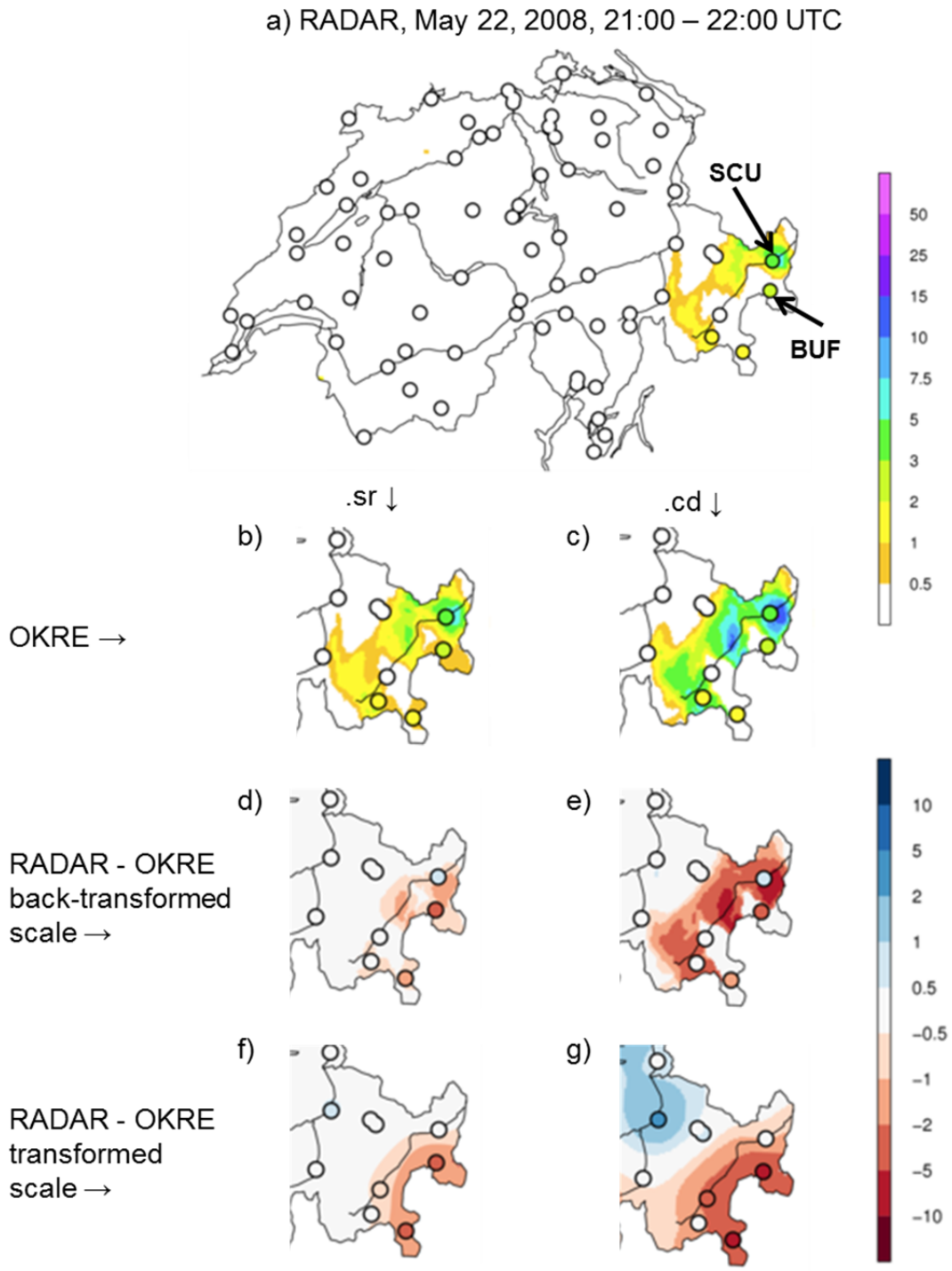

Figure A.2: Examplary case May 22, 2008, 21:00- 22:00 UTC: Hourly accumulated precipitation estimates [mm] by a) RADAR b) OKRE.p.sr c) OKRE.p.cd, radar errors [mm] d) (RADAR – OKRE.p.sr) e) RADAR – OKRE.p.cd and transformed radar errors [mm'] f) RADAR' – OKRE.p.sr' g) RADAR' – OKRE.p.cd'. The black arrows in subfigure (a) indicate the location of the rain gauge at Scuol (SCU) and Buffalora (BUF), respectively.

Figure A.2 shows an example case to illustrate the inflation effect supposedly leading to a general overestimation of precipitation by OKRE. Isolated precipitation in the mountainous region of Grisons

in eastern Switzerland is observed in this exemplary hour of May, 22, 2008. This precipitation observed by four rain gauges is measured rather accurately at some locations (e.g. gauge SCU), but missed completely at others (e.g. gauge BUF) by RADAR (subfigure (a)). Subfigures (b) and (c) show the gridded QPE by OKRE with the prescribed square root ($\lambda = 0.5$) and the case-dependent ($\lambda = 0.2$, in hours with sparse sampling cf. Tab. 2) transformation. It can be seen that there is an overestimation of precipitation around gauge SCU in both QPE fields which is not found in the original radar information. And this overestimation is considerably more pronounced for the setting with the stronger transformation (*cd*). Subfigures (f) and (g) show the interpolated radar errors at transformed scale for *sr* and *cd* respectively. It can be seen that the influence of the negative radar error (red colors) observed at BUF is much larger for *cd* than for *sr*. In contrast, the slightly positive radar error (blue colors) at SCU becomes negligible in *cd* and is fully attributed to nugget (i.e. the allowed deviation from the local trend). This is explained by the fact that the inflation at the lower end of the scale by transformation (reciprocally to the shrinkage effect of the lowermost fractions at back-transformation, see preceding subsection A.1.1) is stronger for smaller transformation parameters λ. Subfigures (d) and (e) show the fields of radar error (i.e. the difference between RADAR and OKRE) at back-transformed scale. It can be seen that the generally negative radar error (red colors) found on the transformed scale (subfigures (f) and (g)) has been inflated in wet regions (SCU) but shrinked in dry regions (BUF). This can be understood as follows: as in OKRE, the radar error is subtracted from the radar field before the back-transformation takes place, errors in wet regions get to the uppermost part of the range under consideration, whereas errors at dry regions (according to the radar) are found at the lowermost part. The shrinkage/inflation-effect by back-transformation as described in the preceding subsection (A.1.1), hence can explain this observed inflation of radar errors in wet regions and shrinkage of radar errors in dry regions. The inflated negative radar errors at wet locations correspond to the observed overestimation of precipitation in subfigures (b) and (c). Again, the smaller transformation parameter λ, the stronger is the described effect. This explains the difference in overestimation between *cd* ($\lambda = 0.2$) and *sr* ($\lambda = 0.5$).

Table A.1: Comparison of *BIAS* by OKRE with two different transformation settings: prescribed square root (*sr*) and case-dependent (*cd*) for the pooled hours with and without missed precipitation by radar (radar = 0mm, gauge > 0mm).

		BIAS		
		OKRE.p.sr	OKRE.p.cd	RADAR
Pooled hours with	Missed precip	0.31	1.40	-0.90
	No missed precip	2.33	1.52	3.15

The mechanism suggested to produce the observed transformation dependent positive *BIAS* in OKRE is, hence, the presence of missed (or severely underestimated) precipitation by radar close to wet areas seen by radar. Table A.1 shows a comparison of *BIAS* for all hours of the systematic application in 2008 where radar missed at least one wet gauge and all hours where there is no such missed precipitation by radar. It shows a larger positive *BIAS* by OKRE.p.cd than by OKRE.p.sr if there is any missed precipitation by radar. RADAR itself shows a negative *BIAS* for these pooled

hours, which is reasonable, since the missing of actually wet gauges should produce a negative *BIAS* per se, and could furthermore indicate a general tendency for visibility problems and hence underestimation of precipitation. For the pooled hours without such missed precipitation by radar, in contrast, the *BIAS* by OKRE.p.cd is smaller than the *BIAS* by OKRE.p.sr. RADAR has an even more positive *BIAS* in these hours than the two OKRE settings, which shows how the combination method successfully reduces this radar bias to a certain extent. These quantitative results show that the observed mechanism in the illustrative example in Figure A.2 is able to explain the transformation-dependent differences in *BIAS* by OKRE in general.

Furthermore, this hypothesis is also in line with other observations in this study. In Figure 5, where stratified evaluations for different geographical regions of Switzerland are shown, the smallest difference in *BIAS* between OKRE.p.cd and OKRE.p.sr are found in the Swiss Plateau. We expect that precipitation is missed less often by radar (better visibility) in the Swiss Plateau than in the more mountainous regions (Alps, Jura and South). In the seasonally stratified analyses (Fig. 6) the smallest *BIAS*-difference between the two transformation settings is observed in summer. Precipitation is expected to be missed less often in summer due to the suppositional higher altitude of the mainly convective precipitation systems than in the colder seasons with mainly stratiform precipitation.

But what about an opposite effect? Erroneous radar echoes in dry regions close to a wet location should be expected to cause a severe underestimation of precipitation at the wet location (due to an inflation of the positive radar error produced by the nearby overestimation of precipitation at the dry location). In fact, this is observed. However, overall the overestimation by OKRE due to adjacent missed precipitation by radar outweighs the underestimation due to adjacent erroneous radar echoes. This is another example of how different compensatory effects can cancel each other out in the overall *BIAS*, observed repeatedly in this study. The reason for the predominance of overestimation due to adjacent missed precipitation by radar is suggested to be threefold: First, adjacent missed precipitation to wet locations is more frequent than adjacent erroneous radar echoes (47% compared to 42%). Second, the magnitude of the missed precipitation is on average of larger magnitude (0.38 mm compared to 0.28 mm). And third, the inflation effect is expected to be stronger in the case of missed precipitation, because the kriging term at wet locations is added to the radar value (i.e. a negative radar error is subtracted) in these cases not subtracted from it. Thereby, it gets located higher-up on the scale, which produces a stronger inflation effect (see A.1.1).

We expect the mountainous topography of Switzerland to increase the incidence of missed precipitation by radar close to wet locations by radar (visibility problems of the radar, mountain and valley stations close to each other). We could imagine, that the overall *BIAS* in regions of flatter topography would therefore be less positive and less transformation-dependent than in our application here.

A.1.3 Negative *BIAS* of KED in poor sampling cases

Additionally to this positive *BIAS* by OKRE, another transformation-dependent behavior has been observed in the analyses of this study: total precipitation is systematically underestimated (negative *BIAS*) by KED in sparse sampling hours, and this underestimation is more severe for the stronger case-dependent transformation (*cd*) than for the prescribed square root transformation (*sr*) (cf. Fig. 4 and Fig. 10). Note that this underestimation is similarly observed for the single sensor reference

method OKGAU. Our suggestion for the mechanism causing this underestimation will be explained for OKGAU first in the following, as it is better comprehensible. The behavior of KED will then be discussed based on these explanations.

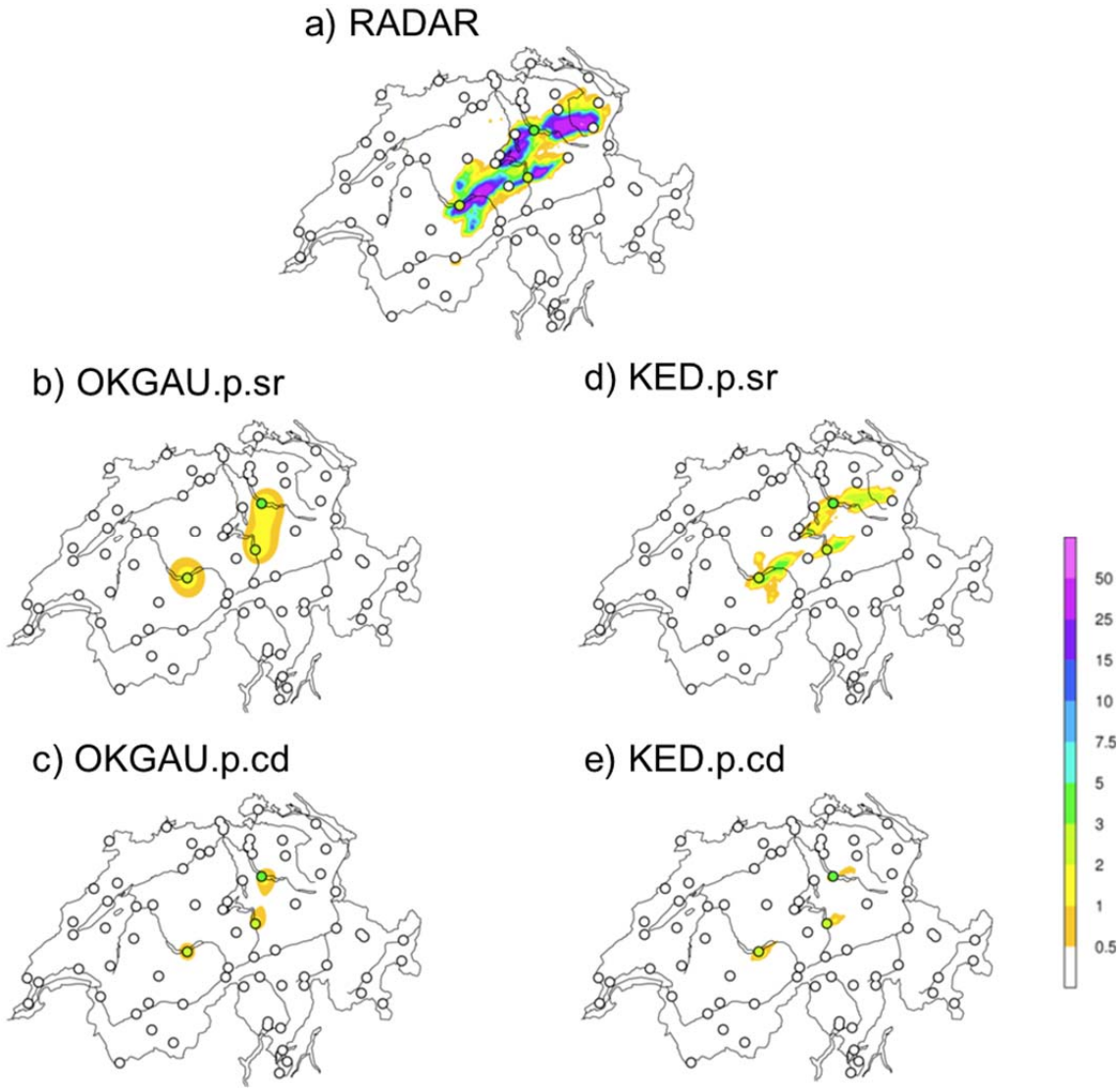

Figure A.3: Example case June 25, 2008, 16:00- 17:00 UTC: Hourly precipitation estimates by a) RADAR and (b), (c) OKGAU and (d), (e) KED with two different transformation settings each: prescribed square root (*sr*) and case-dependent (*cd*).

Analyses with the data of our systematic application over one full year suggest that the observed underestimation in sparse sampling hours is caused by an inflation (cf. inflation effect, section A.1.1) of the nugget. This mechanism is illustrated in the following by an exemplary hour of June 25, 2008 displayed in Figure A.3. There are only three wet gauges (≥0.5mm) and RADAR shows a severe overestimation of precipitation amount and extent in this hour (subfigure (a)). Subfigures (b) and (d) show an underestimation of precipitation by OKGAU.p.sr ($\lambda = 0.5$) and OKGAU.p.cd ($\lambda = 0.2$). However, this underestimation is considerably more pronounced for *cd* than for *sr*: There are lower estimated precipitation amounts, and the extent of the wet areas is smaller. In such a situation with

only a few isolated wet gauges, the precipitation estimate very close to one of these wet gauges corresponds to the value of gauge minus the nugget. Hence, the fraction of this subtracted nugget is located at the uppermost part of the considered range. We have seen at the beginning of this section (A.1.1), how fractions at the uppermost edge get inflated in the process of back-transformation. As we use a constant nugget/sill ratio for all cases with climatological variogram parameters in our application (see Tab. 2), the nugget on transformed scale constitutes a constant fraction of the total value of the rain gauge under consideration for both transformation settings. This constant fraction is more heavily inflated by the back-transformation with the smaller transformation parameter λ = 0.2 (*cd*) than with λ = 0.5 (*sr*). And this larger nugget-fraction at back-transformed scale allows the QPE field to deviate more from the gauge measurements close to an isolated wet rain gauge and hence produces smaller point estimates there.

Note, that at a dry rain gauge close to a wet gauge, the estimate close to the dry gauge is positive, because of the allowed nugget, accordingly. However, at back-transformation, this nugget shrinks to a negligible value, because it is situated at the lowermost edge of the considered range (nugget is added to 0mm measured by the dry gauge). This overestimation at dry rain gauges, hence, cannot cancel out the underestimation at the wet rain gauges at the back-transformed scale, as it does on transformed scale (kriging is bias free by definition!).

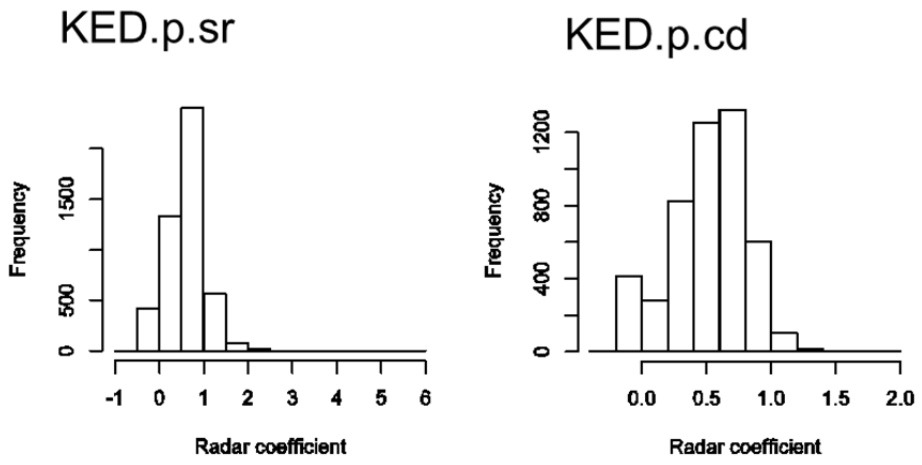

Figure A.4: Histograms of radar coefficients in the systematic application of KED over the year 2008 for with two different transformation settings: prescribed square root (*sr*) and case-dependent (*cd*).

Figure A.4 shows histograms of the distribution of the radar coefficient β (see Eq. 1) in KED in our systematic application over the year 2008 for both transformation settings. It can be seen that the radar coefficient is typically between 0 and 1 for both, KED.p.sr and KED.p.cd. Note that *sr* shows a slightly larger spread of coefficients. This is in line with the findings of our precursor studies on the influence of transformation (see Erdin et al, 2012). If β = 1, KED corresponds to OKRE (cf. 3.1). If β = 0, the trend in Eq. 1 no longer is influenced by the radar, and KED corresponds to OKGAU. If β is found between 0 and 1, as usually the case, a mixture of the effects observed for OKGAU and OKRE is expected in KED. This is in fact observed: In Figure A.4 KED shows a similar underestimation as OKGAU due to the inflation of the nugget at back-transformation. The radar coefficients are β = 0.27

for KED.p.sr and β = 0.31 for KED.p.cd, hence a strong this similarity of KED and OKGAU is what we expect.

Further analyses (not shown) have revealed that the inflation of the radar error in situations with adjacent missed precipitation by radar (as described for OKRE, see preceding section A.2) occurs in KED too. As expected, it is weaker than in OKRE and the scale of overall *BIAS* is shifted to negative values. This shift to negative values is another example of the superposition of different effects (inflation of the nugget as observed in OKGAU superposes the inflation of the radar error at wet locations when adjacent missed precipitation by radar is present as observed in OKRE) in the overall *BIAS*. In the results of our systematic application, these processes can be traced: Figure 4 shows that the transformation-dependent negative *BIAS* in OKGAU is similarly found in KED, but less pronounced. The seasonal analyses (Fig. 6) show that the *BIAS* by KED.p.cd is only stronger negative than the *BIAS* by KED.p.sr in summer (and slightly in spring). This corroborates our hypothesis, since a higher frequency of small isolated precipitation cells is expected in summer, when precipitation is predominantly convective.

We have seen that the negative *BIAS* in KED seems to be caused by the prescribed climatological parameters, namely a (too) large nugget. A recently suggested approach for robust trend and variogram parameter estimation (Künsch, 2011) could possibly mitigate the underestimation of KED (and OKGAU), as it has been shown to estimate substantially smaller nuggets in simulated examples.

MeteoSchweiz
Krähbühlstrasse 58
CH-8044 Zürich

T +41 44 256 91 11
www.meteoschweiz.ch

MeteoSchweiz
Flugwetterzentrale
CH-8060 Zürich-Flughafen

T +41 43 816 20 10
www.meteoswiss.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno Monti

T +41 91 756 23 11
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genève 2

T +41 22 716 28 28
www.meteosuisse.ch

MétéoSuisse
Chemin de l'Aérologie
CH-1530 Payerne

T +41 26 662 62 11
www.meteosuisse.ch