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. 94

Evaluation and comparison of radar-rain gauge combination methods

Denise Keller

ISSN: 1422-1381

Scientific Report MeteoSwiss No. 94

Evaluation and comparison of radar-rain gauge combination methods

Scientific basis for selecting a new real-time precipitation estimation method for Switzerland

Denise Keller

Recommended citation:

Keller, D: 2013, Evaluation and comparison of radar-rain gauge combination methods. *Scientific Report MeteoSwiss,* **94**, 84 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

Applications in a wide range of disciplines require quantitative precipitation analyses at high spatial and temporal resolution with near real-time availability. To meet these needs, MeteoSwiss has developed, in an internal research project (CombiPrecip), three prototypes for generating km-scale, hourly precipitation analyses by real-time combination of radar and rain gauge information. The present study evaluates and compares these methods systematically in order to provide the scientific basis for the implementation of an operational system for such analyses at MeteoSwiss.

The three combination methods are all based on geostatistical concepts and include, in the order of increasing model complexity: Ordinary Kriging of Radar Errors (OKRE), Kriging with External Drift (KED) and Co-kriging with External Drift (CED). The evaluation encompasses a detailed visual inspection for specific test episodes and a quantification of the methods' accuracy (cross-validation errors) in a systematic application for one year (2008).

All three combination methods were found to produce smaller errors than precipitation estimates from radar alone or from rain gauges alone. Seasonally and regionally varying biases in the radar-based estimates are reduced, but not entirely removed. The increased model flexibility of KED and CED is found to be of advantage for the estimation of intensities in wet areas. The distinction between wet and dry areas is, however, better for OKRE. In general, KED and CED yield very similar estimates and error scores, except in hours when limitations in sampling require a fallback to prescribed model parameters. This study suggests that a careful definition of fallback procedures may be as decisive as the modelling concept. A discussion of illustrative examples provides insight into the potential and limitations of geostatistical radar-rain gauge combination in Switzerland.

Acknowledgments

First, I would like to address special thanks to PD Dr. Christoph Frei for offering me this interesting master thesis at MeteoSwiss. He gave me the great opportunity to write my thesis apart from ETH and to gain important insight into applied science. Furthermore I would like to address special thanks to Dr. David Masson. He was my closest contact person and I have really appreciated his great support. Together with my two supervisors I had numerous interesting and inspiring discussions, which I really liked and appreciated. Additionally, I also would like to thank Rebekka Erdin. She was always willing to answer my questions concerning geostatistics and her methods. Her advice and support were very helpful. Finally, special thanks to all colleagues from MeteoSwiss who have supported me in any way during my work, notably Mark Liniger, Ioannis Sideris, Urs Germann and Marco Gabella.

Apart from the mentioned people, I would like to thank my family and friends, who have encouraged and supported me in numerous ways during my studies and my master thesis.

Contents

List of Figures

List of Tables

1 Introduction

Monitoring precipitation using real-time quantitative precipitation estimates (QPE) is highly relevant for various fields of research and applications such as risk management (Overeem et al., 2009) or hydrological forecasting (Zhang et al., 2011). This is particularly true for Switzerland, where precipitation variability is large both in time and in space but where accurate and high-resolution QPE is limited by complex topography.

The Swiss Federal Institute of Meteorology and Climatology (MeteoSwiss) operates three weather radars to monitor precipitation over Switzerland in real-time. While radars are currently the best device to measure precipitation fields at high-resolution in real-time, different sources of errors affect the quality of the radar product (Germann et al., 2006). The awareness of these limitations and the pressing need for accurate and high-resolution precipitation estimates have led scientists at MeteoSwiss to question whether a more advanced solution to the radar-only product could improve current precipitation estimates.

In the context of the NCCR (National Center of Competence in Research) III Climate Project, MeteoSwiss has launched a subproject called CombiPrecip. The goal of CombiPrecip is to improve current real-time QPE by applying geostatistical interpolation techniques, such as kriging (e.g. Gandin, 1965; Matheron, 1965). The idea behind the application of kriging in CombiPrecip is to combine the high accuracy of rain gauge and the high spatial resolution of radar data in order to improve the precipitation estimates on the ground. Therefore, kriging methods are expected to correct the radar errors while maintaining the spatial and temporal structures captured by the radar.

While classical interpolation methods weight gauge measurements as function of the distance to a target location (e.g. Shepard, 1984), kriging uses the spatial covariance structure of the precipitation field derived from gauge measurements. Consequently, kriging is expected to provide more accurate precipitation estimates than simple deterministic interpolation methods.

By evaluating the performance of two different radar-gauge combination methods for the daily time scale, Erdin (2009) has shown that geostatistical combination methods bring added value over single sensor analyses (i.e. radar or rain gauge only) in mountainous regions and in cases of abundant rainfall. As a consequence, MeteoSwiss decided to test combination methods based on kriging in order to monitor precipitation operationally at the hourly time scale and in near real-time for Switzerland. Several candidates exist among kriging techniques and three variants of radar-gauge combination methods have been developed by two teams of MeteoSwiss Zurich and MeteoSwiss Locarno.

In order to better understand the properties of these new products, the present study systematically compares and evaluates the potentials and limitations of the different techniques. On the one hand, the evaluation relies on visual comparison, on the other hand on statistical error scores. A testbed composed of twelve precipitation events occurring between 2005 and 2009 as well as the entire year 2008 is considered. The twelve precipitation events were selected because of their challenging precipitation configurations, such as small-scale precipitation systems, anisotropic fields or a small number of wet gauges. The present evaluation addresses the following questions:

- What is the added value of combination over single sensor analyses?
- What are the differences between the methods and what are their causes?
- Are there external factors influencing the quality of the methods?
- Is there an overall best method?

The outcome of this thesis will serve as an objective basis for the choice of the new real-time precipitation monitoring system at MeteoSwiss. Eventually, the operational implementation of this system shall be completed by the end of 2012.

This master thesis is structured as follows: the data are described in chapter 2. The rain gauge, the radar measurements, the disaggregated precipitation dataset and the specification of the target grid are introduced. The need of data transformation is explained in chapter 3 together with an overview of the different geostatistical methods. Special efforts are devoted to explaining in detail the evaluation techniques, since they build the basis of a large part of the analysis. The results as well as their interpretation are exposed in the fourth chapter. Finally, chapter 5 will conclude this thesis.

2 Data

MeteoSwiss currently operates two devices to measure precipitation in Switzerland, namely weather radars and rain gauges. In this chapter, the principles of these two devices are presented. In addition, another dataset based on the disaggregation of daily measurements into hourly precipitation fields is also introduced and will serve as an alternative reference dataset later.

The aim of CombiPrecip is to develop a new real-time precipitation monitoring system that provides hourly precipitation estimates for Switzerland with a spatial resolution of 1 x 1 km^2. In this study, the recorded precipitation amounts of the gauges as well as those of radar were aggregated to hourly precipitation sums.

2.1 Weather Radars

Weather radars are active remote sensing instruments measuring precipitation indirectly by sending out electromagnetic pulses in the microwave frequency domain. Hydrometeors, such as raindrops, hailstones or snowflakes scatter the radar beam back to the radar, providing information about the objects encountered (Meischner, 2004). The characteristics of the hydrometeors, such as their density and size distribution, determine the intensity of the backscattered echoes. This signal is recorded by the radar sensor as reflectivity Z ($mm^6\ m^{-3}$), which can be related to the rain rate R (mm h^{-1}) thanks to an expression derived from the Rayleigh scattering (see Joss et al., 1998):

$$Z = 316 \times R^{1.5}$$

MeteoSwiss operates three meteorological C-band radars located at the mountaintop of the Albis, La Dôle and Monte Lema (Fig. 2.1). They scan the atmosphere with frequencies between 0.5 and 1 GHz, at 20 elevations from -0.3° to 45°, a time resolution of 5 minutes and up to a range of 230 km (Germann et al., 2006). The final product is provided in nominal resolution of 1 x 1 km^2. Because of their high spatial and temporal resolution, radars are used to monitor precipitation in real-time. Despite the high spatial and temporal resolution, instrumental retrieval and limitations affect the reliability of radar estimates. Especially in mountainous regions beam shielding and ground clutter are severe issues. Even though sophisticated algorithms are implemented to correct for these effects (e.g. vertical profile correction), they can lead to errors in the precipitation estimates. Additionally, radar tends to overestimate precipitation in cases where melting snow or hailstones are detected and to over- or

underestimate precipitation when the size of the hydrometeors changes beneath the available elevation of the radar (Germann et al., 2006).

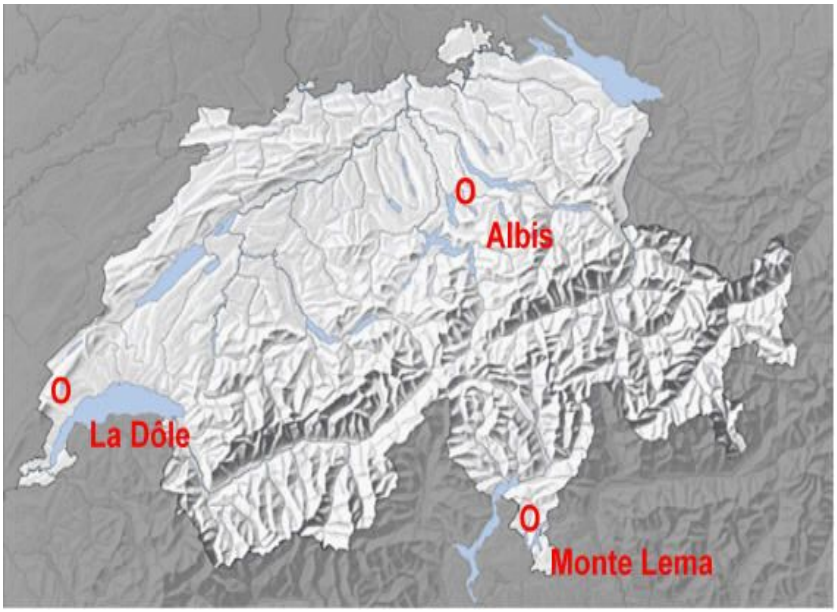


Figure 2.1: Position of the three Swiss weather radars operated by MeteoSwiss at the time of the study (Germann et al., 2006).

2.2 Rain gauges

Rain gauges provide point measurements of precipitation rate over a collecting area of typically 200 cm^2 (Konzelmann et al., 2007). MeteoSwiss maintains a rain gauge network of approximately 450 manual or automated stations, which covers a large portion of the country. Fig. 2.2 shows the 75 automated stations (SwissMetNet, status 2008) measuring precipitation every 10 minutes. The typical distance between two stations is 15-20 km (Wüest et al., 2010). In order to capture the highly variable precipitation fields over mountainous regions, the station density in the Alps is slightly larger than in the Swiss Plateau.

Compared to the radar resolution, the observational network cannot capture the complex spatial structure of the precipitation field. In terms of accuracy rain gauge measurements are primarily affected by systematic errors (Neff, 1977) of which the magnitude increases with wind speed and decreases with drop size. In Switzerland these biases vary between 4% in flatland in summer and 30% at high altitudes in winter (Sevruk, 1985). Nevertheless, the gauge measurements are in general more accurate than the radar measurements (Joss et al., 1998). In this study they are considered as estimates of the true precipitation rate at the station location. They serve as a complementary device to correct biases and errors of the weather radars. The 75 SwissMetNet stations shown in Fig. 2.2 were used to construct and evaluate the precipitation fields based on the radar-gauge combination.

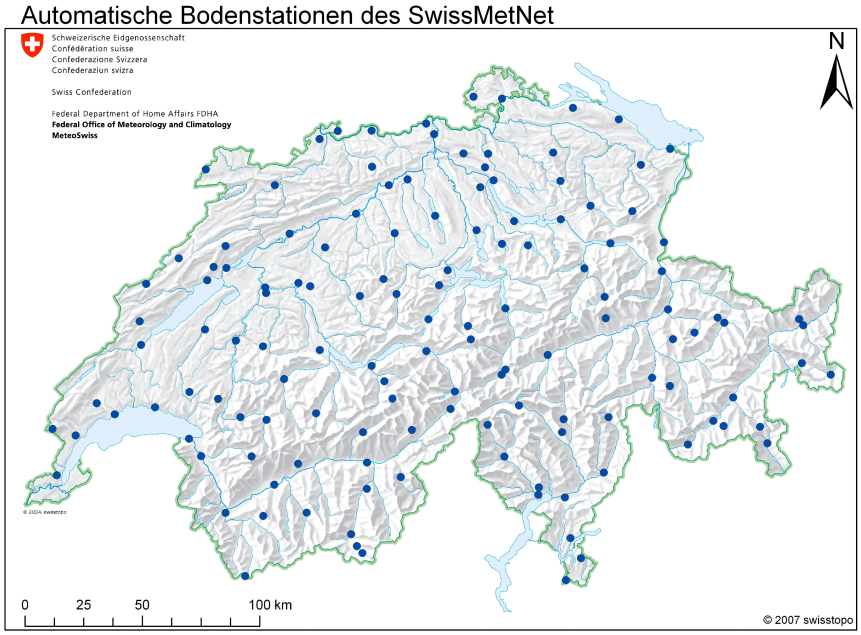


Figure 2.2: Position of 75 automatic SwissMetNet rain gauge stations (status 2008) measuring precipitation every 10 minutes (MeteoSwiss, 2010).

2.3 Time-disaggregated dataset

The disaggregated hourly precipitation estimate is an alternative dataset in addition to automatic gauge network and radar QPE. It is used as an additional reference when comparing the performance of different geostatistical interpolation techniques. The accuracy of the disaggregated dataset is high due to the combined use of a denser station network (approximately 450 rain gauge station) at the daily time scale together with the hourly radar product (Wüest et al. 2010). The daily precipitation rate $R_d(\mathbf{x})$ measured at the gauge location $\mathbf{x}$ is used to correct the bias of the daily radar precipitation estimates $E(\mathbf{x})$. The hourly radar precipitation estimate $E(\mathbf{x},t_i)$ is corrected by the factor calculated above and yields the hourly disaggregated precipitation amount:

$$R(\mathbf{x},t_i) = \frac{R_d(\mathbf{x})}{\sum_{i=1}^{24} E(\mathbf{x},t_i)} E(\mathbf{x},t_i)$$

The disaggregation method was developed primarily for climatological applications and for model validation, but is not appropriate for real-time precipitation monitoring since daily precipitation sums are needed. Consequently, the disaggregated dataset will serve here for comparison purposes only and gives an idea about the performance of a dataset used in climatology. Because there are more station measurements available at daily resolution, the performance of the disaggregation dataset is expected to be better than the near real-time analyses.

3 Methods

The first part of the method section introduces the fundamentals of geostatistical interpolation in order to give the reader an overview and some general understanding. More explanation about the mathematical details can be found in textbooks like Webster and Oliver (2007), Wackernagel (1995) or Schabenberger and Gotway (2005). The applied geostatistical interpolation methods used in this study are presented then in the second part of this chapter. Furthermore, the used data transformation and its necessity when applying geostatistical interpolation methods to precipitation fields are explained in the third part. The validation techniques and skill scores used to evaluate the methods are presented in the last section of this chapter.

3.1 Fundamentals of geostatistical interpolation

Besides interpolation methods based on inverse distance (e.g. Shepard 1984), more sophisticated techniques based on a geostatistical framework exist. The aim of a geostatistical interpolation is to predict the realization z of a random field Z at any unsampled location $\boldsymbol{x}$ on a target grid, conditional on the surrounding observations $z(\boldsymbol{x}_i)$ $(i= 1…N)$, where N is the total number of stations. Therefore, the spatial characteristics of a random field Z is the key aspect of any geostatistical interpolation method and can be quantified by the covariance between the locations $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:

$$C\left[Z(\mathbf{x}_1), Z(\mathbf{x}_2)\right] = E\left[\left\{Z(\mathbf{x}_1) - \mu_1(\mathbf{x}_1)\right\}\left\{Z(\mathbf{x}_2) - \mu_2(\mathbf{x}_2)\right\}\right]$$

where $E[.]$ is the expected value operator and μ_1 and μ_2 are the mean values of Z at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. In other words, the spatial covariance is determined by the expected value of the product of the deviations from the corresponding means μ_1 and μ_2 at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. A major concern is that only a single realization is available at time t ~~and~~ therefore not only the true values μ_1 and μ_2 are unknown, but also the covariance itself. Two assumptions concerning the field can greatly simplify this problem. First, the assumption of stationarity and second, the assumption of weak-stationarity.

A random field Z is said to be stationary if all its finite-dimensional distributions are invariant under arbitrary spatial translations. In particular, all statistical moments of Z do not depend on the location $\boldsymbol{x}$. Since Z is considered as a Gaussian random field, its entire distribution at a given location $\boldsymbol{x}$ is fully determined by its mean μ and its variance σ. Therefore, the assumptions of weak-stationary are sufficient to describe the Gaussian random field Z completely. This implies that the covariance

structure of a two-dimensional field depends only on the lag distance $\boldsymbol{h}$ between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ rather than on the absolute positions of them. Therefore we can re-express the covariance as a function of the lag vector $\boldsymbol{h}$ between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:

$$C(\mathbf{h}) = C\left[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})\right] = E\left[\left\{Z(\mathbf{x})\right\}\left\{Z(\mathbf{x}+\mathbf{h})\right\} - \mu^2\right]$$

If $C(\boldsymbol{h})$ only depends on the norm of $\boldsymbol{h}$, the field is said to be isotropic. Otherwise, it is anisotropic.

Although the covariance expression is of a central importance it is only defined for second-order stationary random functions. In addition an estimate of the mean μ of the random field Z is required. In order to overcome these major drawbacks of the covariance expression, Matheron (1965) replaced the aboce expression by the semi-variance $\gamma(\boldsymbol{h})$, which is a more general formulation and like the covariance only lag dependent:

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[(Z(\mathbf{x}+h) - Z(\mathbf{x}))^2\right]$$

Under weak-stationary conditions the semi-variance and the covariance are related through:

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$$

Matheron's hypothesis is a central aspect of geostatistical interpolation that leads to assign weights to the surrounding observations according to the lagged covariance structure of the field.

The relation between the lag distance $\boldsymbol{h}$ and the semi-variance γ is represented by a variogram, as shown for example in Fig. 3.1. Depending on the variance structure of the field, several parametric models e.g. exponential, spherical, circular or a power model can describe the variogram. In this study, all the methods are based on an exponential variogram model which produced the best results.

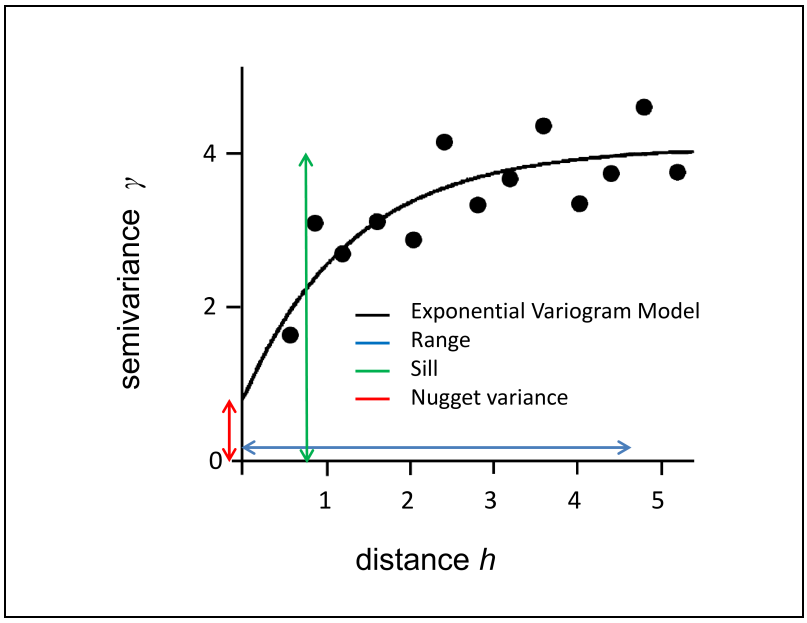


Figure 3.1: Example of an empirical variogram.

A variogram model is composed of three elements, namely the nugget variance, the sill and the range (see Fig. 3.1). These elements give us important information about the spatial variability of the field. The semi-variance at h=0 is called the nugget. It quantifies the subgrid variability of the field and possible measurement errors. A semi-variance equal to zero at h=0 implies that the field varies smoothly in space with no discontinuities. The sill describes the a priori variance of the process and is reached (in the case of an exponential model) at 95% of a lag distance called the effective range. Observations that are separated by larger distance lags can be considered independent (i.e., they do not influence each other). Fig. 3.1 shows that small lag distances are associated with small semi-variance values. Equivalently, the covariance is high at small lags. The inverse is true for larger lags. Two different precipitation fields and their corresponding empirical and fitted variograms are shown as illustration in Fig. 3.2.

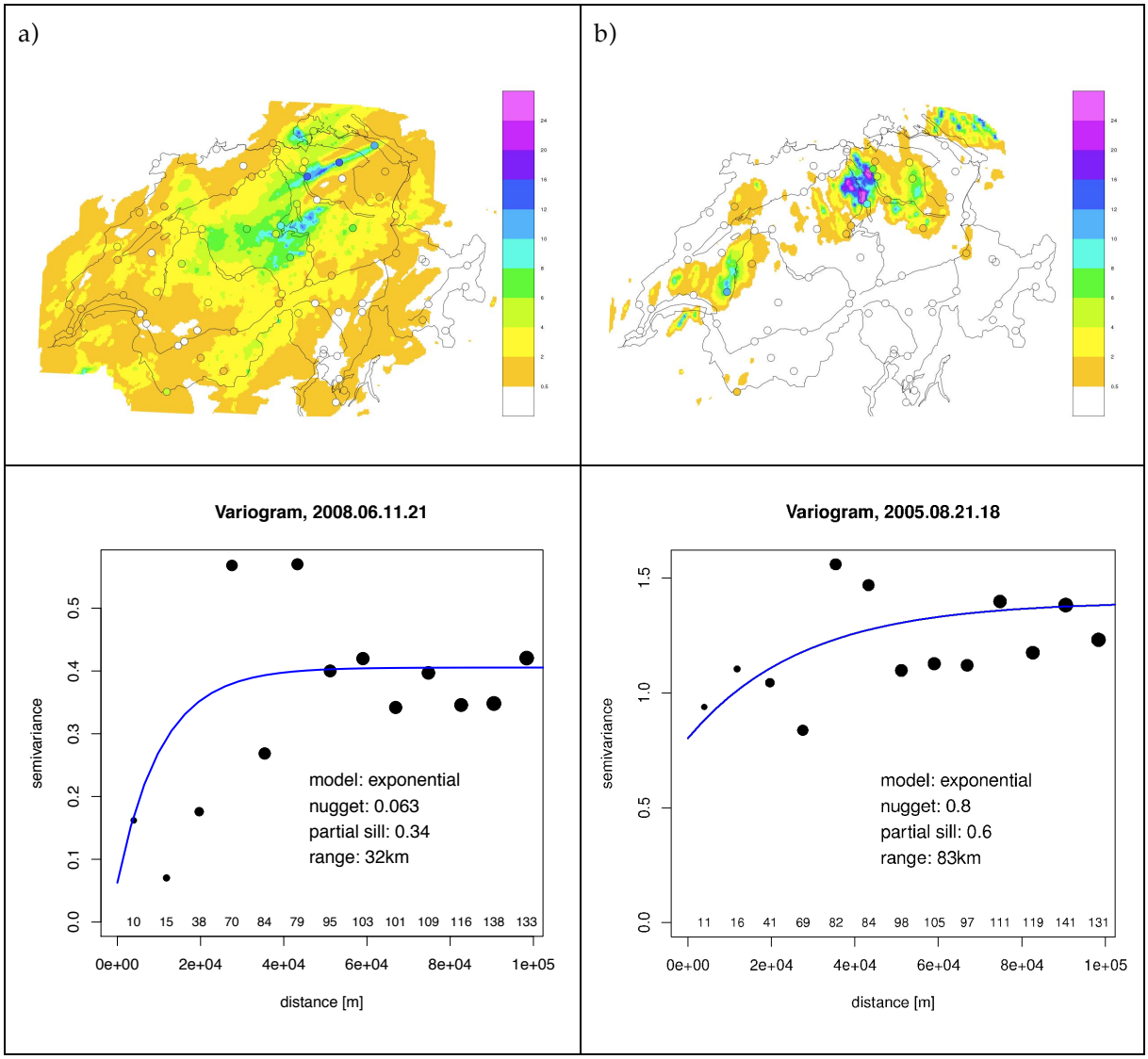


Figure 3.2: Examples of smooth (a) and rugged (b) precipitation fields and their corresponding variogram.

The field in Fig. 3.2a varies more smoothly in space, whereas the field in Fig. 3.2b varies highly in space and shows discontinuities. These features are well reflected in the corresponding variograms. A large range and a small nugget indicate smoothness of a field. In contrast, a small range and large nugget indicate large short-range fluctuations. A central feature of geostatistical interpolation methods is the importance given to the information contained within the range, while observations

lying outside the range are less relevant for the prediction. In case of a smoothly varying field as displayed in Fig. 3.2a, nearer observations to a target grid-point carry more weights than distant observations. For a field with large fluctuations as illustrated in Fig. 3.2b, discontinuities are large and the range is small. Consequently, the relation between two near located observations is weak and the weights of the observation relatively to the target point are small.

The semi-variogram described above is now used to determine the weights λ_i attached to the sourrounding stations $z(x_i)$ in order to estimate the realisation z at an unsampled location $\boldsymbol{x}$. Geostatistical interpolation methods weight the information from the stations $z(x_i)$ in such a way that the prediction variance is minimal and the weights sum up to one. The generic form of a linear spatial interpolation method can be written as:

$$z(\mathbf{x}) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

3.2 Geostatistcal interpolation methods used in this study

Although rain gauges provide accurate point measurements of a continuous varying field, they cannot capture the high spatial variability of precipitation because of their coarse resolution and limited spatial representativity. Since many applications require accurate QPE on a target grid, different institutions have developed combination techniques based on geostatistical interpolation, such as kriging as pioneered by Gandin (1965) and Matheron (1965).

Since kriging takes the spatial covariance structure of the random field into account, it is able to represent the high spatial variation of the precipitation field more adequately than other methods would. Furthermore, kriging methods are able to estimate the prediction uncertainty of an interpolated field. These are the reasons why kriging is a popular framework for radar-gauge combination. Three different kriging methods were considered in this intercomparison, namely ordinary kriging of radar errors (OKRE), kriging with external drift (KED) and universal co-kriging with external drift (CED). All these methods consist of a deterministic and a stochastic part. The deterministic part represents a first approximation of the field, which is either a constant averaged value or a linear regression model (often referred as the trend). Formally the deterministic part represents a spatially varying expected value of the random Gaussian field Z introduced previously. Of course, a stochastic field shows deviations from this first approximation. These deviations from the mean value represent the stochastic part and are estimated using kriging.

3.2.1 Ordinary kriging of radar errors

Ordinary kriging of radar errors (OKRE) interpolates the radar errors over the entire target grid. In order to get the precipitation field back, the predicted interpolated error field is subtracted from the radar field. The radar error E at the rain gauge locations $\boldsymbol{x}_i$ ($i=1\ldots N$) is defined as the difference between the radar value R at $\boldsymbol{x}_i$ and the observed gauge value G at $\boldsymbol{x}_i$ (note: R and G are square root transformed values as described in section 3.3). The deterministic part in OKRE is a constant value defined as the mean radar error E_{mean} over all N stations:

$$E_{mean} = \frac{1}{N}\sum_{i=1}^{N}\left[R(\mathbf{x}_i) - G(\mathbf{x}_i)\right] = \frac{1}{N}\sum_{i=1}^{N}E(\mathbf{x}_i)$$

The mean radar error serves in OKRE as a trend or first approximation of the error field. Of course, the precipitation field varies in space and exhibits deviations from this trend. These deviations represent the stochastic part of the model. It consists of the sum of the weighted differences between the radar error at the station *i* and the mean radar error. In OKRE, the predicted radar error $E(\mathbf{x})$ at an unsampled location $\mathbf{x}$ is expressed by:

$$E(\mathbf{x}) = E_{mean} + \sum_{i=1}^{N} \lambda_i \left[E(\mathbf{x}_i) - E_{mean} \right]$$

OKRE is the simplest model of the considered combination techniques. It has the smallest number of parameters (four in total, one for the trend, three for the variogram). In our application of OKRE, the parameters are estimated by the restricted maximum likelihood principle, which makes direct use of the underlying stochastic concept and is more robust and bias free (especially for small samples) compared to estimation via an empirical variogram (see Diggle and Ribeiro, 2007). OKRE was developed and applied previously for daily radar-gauge combination in Switzerland by Erdin (2009) and further studied in hourly real-time QPE by Erdin et al. (2012).

3.2.2 Kriging with external drift

In kriging with external drift (KED), the deterministic component is a linear transformation of the radar field. The trend surface is completed with a stochastic part ε such that the complete model of the precipitation Z at location $\mathbf{x}_i$ is given by:

$$Z(\mathbf{x}_i) = \alpha + \beta R(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i)$$

Here α and β are respectively the intercept and the slope of the linear relation with radar $R(\mathbf{x}_i)$ and $\varepsilon(\mathbf{x}_i)$ are the residuals representing the deviation of the gauge value G at $\mathbf{x}_i$ from the trend surface:

$$\varepsilon(\mathbf{x}_i) = G(\mathbf{x}_i) - (\alpha + \beta R(\mathbf{x}_i))$$

The stochastic part of KED consists of an interpolation of the residuals and is described by a weighted sum of the differences between the gauge G observations and collocated trend. Thus, the precipitation at any unsampled location $Z(\mathbf{x})$ is estimated by the combination of the trend with the modelled residuals:

$$Z(\mathbf{x}) = \alpha + \beta R(\mathbf{x}) + \sum_{i=1}^{N} \lambda_i (\varepsilon(\mathbf{x}_i))$$

Compared to OKRE, KED is more complex in model formulation as it involves a spatially varying trend. KED has one more parameter (the radar coefficient β) and hence allows for additive and multiplicative correction of radar errors. As with OKRE, KED is applied with square root transformed data here (see section 3.3) and uses the robust restricted maximum likelihood procedure for parameter estimation. An early version of the KED procedure used in this study was developed in Erdin (2009). A detailed study of the effect of transformation with KED is given in Erdin et al. (2012).

Compared to Co-kriging with external drift (see section 3.2.3) KED transforms radar values smaller than 0.05 mm/h to 0 mm/h.

3.2.3 Co-kriging with external drift

Co-Kriging is a variant of geostatistical models that exploits additional secondary information in the stochastic model part. Secondary information can be useful to increase the accuracy of an interpolation provided that information has a strong relation to the primary variable.

The radar-gauge combination method termed as "Co-Kriging with external drift (CED)" in this study is a new proposal by Ioannis Sideris, Urs Germann and Marco Gabella from MeteoSwiss Locarno. Apart from the instantaneous measurements at the target time *t*, this method also includes radar and gauge measurements from time *t-1* (i.e. one hour before) as secondary information in a co-kriging framework. The model can be formulated as:

$$Z(\mathbf{x}_i) = \alpha + \beta R_t(\mathbf{x}_i) + \varepsilon(\mathbf{x}_i \mid G_t, R_t, G_{t-1}, R_{t-1})$$

Where the stochastic model part ε is now conditioned on the measurements (gauge and radar) from two successive hours. An optimal linear estimate of ε is obtained by linear combination of the conditioning variables, where the coefficients are determined from the spatial covariance of the primary variable (as with OKRE and KED) plus from the spatial cross-covariance of the secondary variables (*t-1*) with the primary variable (*t*). Hence apart from the semi-variogram, CED also parameterises two cross-covariograms.

The ultimate idea of CED is to make a better use of available data at time *t* by including information from *t-1*. The CED model is even more complex than KED. The number of parameters increases by 6 (3 for each cross-covariogram) compared to KED. It can be expected that this increase in model sophistication is rewarded if there is a high cross-correlation between two time steps (secondary information is influential) and if the increase in parameter number is compensated by the additional information (samples) at time *t-1*.

As with OKRE and KED, CED also uses square root transformed measurements. But unlike OKRE and KED, CED estimates model parameters via non-linear least squares from empirical variograms. More details about CED can be found in Sideris et al. (2011).

3.2.4 Ordinary kriging of gauges

Besides the temporal disaggregation dataset (section 2.3), ordinary kriging of rain gauges (OKGAU) only will serve as a reference method for illustrating and quantifying the added value of the combination techniques. The ordinary kriging approach was applied to square root transformed gauge measurements.

OKGAU uses the mean of all gauge values G_{mean} as a first approximation of the precipitation field. The stochastic part consists of the sum of the weighted differences between the gauge G at location $\boldsymbol{x}_i$ and the mean gauge value. The prediction of precipitation at any unsampled location is given by:

$$Z(\mathbf{x}) = G_{mean} + \sum_{i=1}^{N} \lambda_i (G(\mathbf{x}_i) - G_{mean})$$

3.3 Data transformation

The idea of geostatistical interpolation methods is based on the one hand on random Gaussian fields and on the other hand on minimizing the prediction variance (quadratic errors). However, if we apply geostatistical interpolation methods to precipitation data we have to consider two issues. Firstly, that precipitation fields do not meet the characteristics of a random Gaussian field since it is bounded by zero and strongly positively skewed. And secondly, using the concept of minimal prediction variance, high precipitation values can negatively influence the quality of the interpolation. Goudenhoofdt and Delobbe (2009) showed that applying data transformation to precipitation data can greatly overcome these issues.

In order to mitigate the influence of high precipitation values and to approximate a random Gaussian field, we transformed the data beforehand using the Box-Cox transformation (see e.g. Wilks, 2006):

$$Y(z) = \begin{cases} \frac{z^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \ln(z) & \lambda = 0 \end{cases}$$

where each value z is a realization of a Gaussian random variable Z representing the precipitation. The transformation of the data by a power λ yields a more symmetric probability density function (pdf). Following Erdin et al. (2012), the methods tested in this study have selected the square root transformation (λ=½), which brings the data closer to the normal distribution without reducing the quality of the interpolation. After the interpolation is calculated on transformed data, the field is finally quantilewise back-transformed in precipitation values using the inverse transform:

$$z(Y) = \begin{cases} \sqrt[\lambda]{Y * \lambda + 1} & \lambda \neq 0 \\ \exp[Y] & \lambda = 0 \end{cases}$$

3.4 Fallback solutions

In order to reproduce the hourly precipitation field adequately, the kriging methods require an accurate estimation of the parameters of the trend and the variogram model. However, a reliable estimation of these parameters is only possible when enough wet gauges are available in one hour. A gauge is denoted as wet when it records precipitation intensity larger or equal than 0.5 mm/h. In cases where not enough wet gauges exist, the methods are not able to estimate reliable parameters and are therefore forced to use an alternative technique. In the following we term this the "fallback" solution. For example, the fallback variogram can use predefined parameters instead of variogram parameters estimated from the data.

Since the two teams developed their methods independently, the three combination methods use different settings of the fallback solutions. On the one hand they use different requirements on the minimal numbers of wet gauges needed to start the estimation of the variogram parameters. And on the other hand, they also differ in the fallback default variogram parameters. Some of these parameters are determined from empirical studies while others were arbitrarily chosen. Tab. 3.1 summarizes these differences.

Table 3.1: Variogram parameter specification for the different methods.

		Variogram parameters for Fallback solution	
	Critical number of wet gauges	Nugget-sill ratio	Range (km)
OKRE	10	0.28	108
KED	10	0.2	75
CED	5	0.1	15

As OKRE and KED need at least 10 wet gauges to estimate the variogram parameters, these methods use more often the predefined variograms as fallback solution compared to CED, which tries to fit a variogram model as soon as five wet gauges are available. Note also the large difference in the default range parameters. The small range of CED implies that a smaller number of surrounding gauges are included for kriging in regions where the target is close to the stations. In regions that are far from the stations, the stochastic component relaxes towards a constant. The implication of this small range will be discussed in the result section.

The performance of the fallback solution will be evaluated separately in order to identify its impact on the overall performance of a method. If the predefined variogram parameters are appropriate, adopting the fallback solution can in principle be more profitable than relying on non-robust parameters. Note that in some rare cases with very few wet gauges, it was not possible to fit a variogram model to the sample data for KED and OKRE. In these situations, the radar field was used as an ultimate fallback.

3.5 Evaluation methods

3.5.1 Cross-validation

Cross-validation is a valuable technique to assess the performance of models when no test dataset is available. The idea is to exclude one observational data point from the dataset and to fit the model to the remaining data points (see e.g., Wilks, 2006). Then, the precipitation value at the excluded location is estimated and compared to the observation, yielding the cross-validation error, which is defined as the difference between the predicted and the observed value. This procedure is applied at each station location and at each time step, yielding a cross-validation table, which becomes the starting point for the evaluation and comparison of the methods. Note that this kind of validation technique has extensively been used to evaluate radar-rain gauge combination methods (e.g. Seo, 1998; Haberlandt, 2007).

The rain gauge measurements are assumed to be the true precipitation values. Although rain gauges are affected by several errors (Sevruk, 1985), these are generally smaller compared to those of radar. Nevertheless, the cross-validation error will involve a component that emanates from the differences in spatial representativity between the validation gauge (local) and the pixel estimate (regional), and hence cannot be related to a deficiency (error) of the method. In the following we assume that this representativity discrepancy is similar between all methods.

3.5.2 Error scores

Error scores summarize cross-validation errors by a single scalar value, allowing an easier interpretation of the results. Since a single error score is not able to represent all performance aspects, different scores are necessary to cover several features of a method. Four different scores are used in this thesis: the bias, the relative mean root transformed error, the Hanssen-Kuipers discriminant and the scatter. The bias and the Hanssen-Kuipers discriminant are well known and often used scores, whereas the relative mean root transformed error was especially created within the CombiPrecip project. The scatter was developed by German et al., (2006).

Bias

Commonly, the bias measures the systematic error of a method and indicates the mean error of a method. It is usually defined as the difference between the average prediction and the average observation (Wilks, 2006). In this study, the definition of bias is different and corresponds to the ratio between the sum of n predicted values (pred) and the sum of n observed values (obs):

$$BIAS = 10 \times \log_{10}\left(\frac{\sum_{i=1}^{n} pred_i}{\sum_{i=1}^{n} obs_i}\right)$$

As prevalent in radar meteorology the bias is calculated with logarithmic transformed ratios. The unit of bias is thus in decibel (dB). If the sum of the predicted precipitation values is the same as the sum of the observed values, the bias is per definition zero, meaning that the systematic error is zero. A positive bias indicates an overestimation of the precipitation values by the method, whereas a negative bias implies an underestimation of the precipitation values. A drawback of the bias is that positive and negative errors can compensate, which calls for additional measures of random error.

Relative mean root transformed error

The relative mean root transformed error (REL.MRTE) uses root-transformed data in order to mitigate the importance of large precipitation values. It incorporates aspects of the root-mean squared error (RMSE) with other aspects of the traditional bias and is defined as the ratio between the mean root transformed error (MRTE) and the mean root transformed deviation (MRTD):

$$REL.MRTE = \frac{MRTE}{MRTD} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{pred_i} - \sqrt{obs_i}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{obs_i} - \sqrt{\overline{obs}}\right)^2}$$

The MRTE measures the mean error between the predicted and the observed values and bares some similarity with RMSE. Squaring the errors avoids the compensation of negative and positive errors. MRTD is a measure for the mean deviation of the observations from the observational mean. The MRTE is normed by MRTD to ensure comparability between samples and to simplify the interpretation of the results.

The values of REL.MRTE are between 0 and +∞. An error score of zero indicates that the error spread of a method vanishes and the method is therefore able to reproduce perfectly the station values. In cases where the score reaches the critical value of one, the prediction error is as large as the mean deviation of the transformed observation. As a consequence, the use of the interpolation method

brings no added value, because a constant value (the mean) through the entire field would lead to a similar error score.

So far we have introduced two error scores that assess how close the predicted QPE are compared to the gauge observations. However, several applications are more interested in the distinction between wet and dry conditions rather than in the precipitation quantity. Consequently, it is important to assess the ability of a method to distinguish between wet and dry conditions. A suitable score assessing this characteristic of a method is the Hanssen-Kuipers Discriminant.

Hanssen-Kuipers Discriminant

Considering that the realization of a random variable Z is dichotomous, i.e. either wet or dry, the prediction-observation pair can be classified according a 2 x 2 contingency table as shown in Tab. 3.2. A gauge precipitation value implies wet conditions when it is larger or equal to 0.5mm/h. This threshold has been used since precipitation values smaller than 0.5 mm/h are probably not reliable due to the gauge uncertainty (Sevruk, 1985), and representativity differences.

Table 3.2: Contingency table

		Observation		Sum
		Wet	Dry	
Prediction	Wet	*a*	*b*	*a+b*
	Dry	*c*	*d*	*c+d*
Sum		*a+c*	*b+d*	*N*

The Hanssen-Kuipers Discriminant (HK) consists of two error measurements, namely the probability of detection (POD) and the false alarm ratio (FAR), which are built from the component of the 2 x 2 contingency table.

The probability of detection (POD) measures the ability of a method to predict wet events correctly and is defined as the ratio of the correct predicted wet events (*a*) and all wet observed events (*a+c*):

$$POD = \frac{a}{a+c}$$

The false alarm ratio (FAR) is defined as the ratio of incorrectly wet predicted events (*b*) and all dry observations (*b+d*). It measures how often a method wrongly predicts wet events:

$$FAR = \frac{b}{b+d}$$

Subtracting FAR from POD yields the Hanssen-Kuipers Discriminant (HK) (see, e.g. Wilks, 2006) that assesses the ability of a method to distinguish between wet and dry events. The HK reaches values between minus one and one.

$$HK = POD - FAR = \frac{ad - bc}{(a+c)(b+d)} \qquad -1 \le HK \le 1$$

For convenience, the HK skill score is transformed into an error score referred as HK.ERR:

$$HK.ERR = 1 - HK$$

The HK.ERR of a perfect model is equal to zero, meaning that the POD is equal to one and the FAR equal to zero. A HK.ERR of one means that the POD is equal to the FAR and therefore the model does not bring an added value. The same is true for HK.ERR scores larger than one where the FAR is larger than the POD.

Scatter

A complementary error score is the scatter (German et al., 2006). In contrast to the other error scores, it evaluates only wet locations, where precipitation is actually observed and predicted, and hence incorporates only values larger or equal to 0.5 mm/h. Scatter is a measure for the spread of the error distribution and is based on the cumulative error distribution function (CEDF). Fig. 3.3 shows the CEDF of radar where the x-axis displays the errors and the y-axis the contribution of the corresponding errors to the total rainfall.

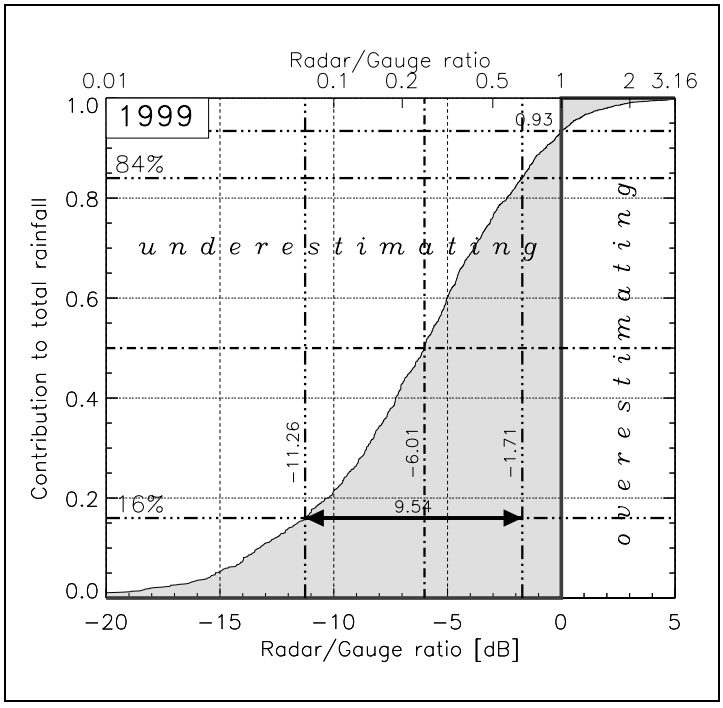


Figure 3.3: Cumulative error distribution function (CEDF) of radar (German et al., 2006).

Scatter can be calculated by measuring half the distance between the 16% and 84% percentile of the CEDF and is measured in decibel. A score near zero indicates a small error spread.

$$SCATTER = \frac{1}{2}\left(CEDF_{84} - CEDF_{16}\right)$$

Since scatter trims the upper and lower 16% percentile of the CEDF, it is a robust error measure mitigating the influence of outliers. However, scatter has no information about systematic errors, since by construction, the error spread is quantified around the median error.

3.5.3 Validation datasets

Twelve test cases (see Tab. 3.3) and the whole year 2008 are considered for the validation and evaluation of the combination methods. The test cases include different precipitation events between

2005 and 2009 in Switzerland and consist each of 24 successive hours. They have been selected because of their challenging and varying precipitation configuration, like small-scaled precipitation systems, anisotropic fields, attenuation of radar and small number of wet gauges (see Tab. 3.3). A detailed description of the twelve test cases can be found in Willi et al. (2011).

The twelve test cases have been investigated by visual comparison primarily in order to reveal the potentials and limitations of the combination methods. In addition to the twelve test cases, the combination methods were evaluated for the entire year 2008. Given the larger amount of data over one year, the dataset 2008 was stratified, hour by hour, as shown in Fig. 3.4 to tackle the following questions:

- How do the methods perform in dry hours?
- How does the fallback solution influence the performance of a method?
- How do the interpolation methods vary over different regions and seasons?

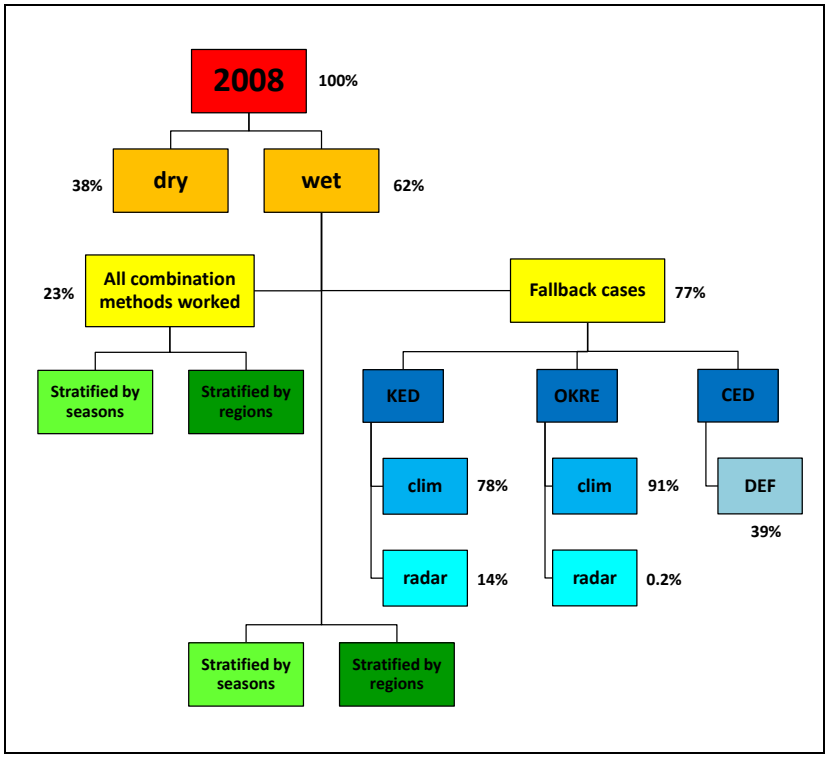


Figure 3.4: Stratification of year 2008 into subsets of hours.

Each box in Fig. 3.4 indicates a subset of hours from the whole data set 2008. The percentage displayed beside the boxes denotes the numbers of hours belonging to the corresponding data subset relative to the hierarchically superior subset. The term "dry" refers to hours where all the rain gauges over Switzerland recorded values smaller than 0.1mm/h. The term "wet" indicates hours where at least one rain gauge measured a value larger than 0.1mm/h. This threshold corresponds to the minimal precipitation value that is trusted when reading gauge measurements.

As shown in Fig. 3.4, the whole dataset 2008 is first split into dry and wet subsets. The wet subset is further stratified by cases where all methods worked simultaneously and by situations where at least one method had to use the fallback solution. The investigation of cases where all methods worked simultaneously ensures that the methods are evaluated without the influence of the fallback solutions. Furthermore, the fallback cases were stratified by methods and their corresponding fallback solution.

Additionally, the wet cases as well as the situations where all methods worked simultaneously were further stratified by regions and seasons.

3.6 Software

The statistical computation software R is used for the statistical analysis as well as the graphics. R is part of the Free software Foundation's GNU project (R Development Core Team, 2011). KED and OKRE rely on the R package geoR (Ribeiro and Diggle, 2001) and CED on gstat (Pebesma, 2004). For the present analysis packages developed by Rebekka Erdin and Christoph Frei at MeteoSwiss were used.

Table 3.3: Summary of the twelve test cases.

Case	Title	Precipitation characteristic	Radar bias	Anisotropy	Error source	Variability	Challenges
1	August 2005 floods	Stratiform	-	Observed		Strong	Correction of radar bias
2	Widespread heavy rainfall	Stratiform	-		Beam shielding, attenuation		Attenuation
3	An active cold front crossing the Alps	Convective overall	-	Observed		Strong	
4	Slow-moving small-scale rainfall cells - flooding during EURO 08	Convective isolated	-		Beam blocking, spatial support of radar	Strong	Extreme small-scale variation, Parameter estimation
5	Thunder- and hailstorms	Convective isolated	+	Observed	Hailstones		Anisotropy, Parameter estimation
6	Extended wet-spell in southern Switzerland	Convective overall	-	Observed	Limited visibility of radar, small number of station in high-intensity area	Strong	High local variability between North and South
7	Convection with below-cloud evaporation	Convective isolated	+		Below cloud evaporation		Absence of wet gauges
8	Unstable air along the southern Alpine slope	Convective overall				Strong	
9	Fast moving heavy hailstorms	Convective overall	+	Observed	Beam shielding, wet radome		Anisotropy
10	Local and intense showers	Convective isolated	+		Ground clutter	Strong	Small- scaled precipitation structure
11	Local showers and residual clutter	Convective isolated	+		Ground clutter		Ground clutter
12	Orographic lifting south of the Alps and cold front crossing	Stratiform	-				

4 Results

The focus of the first part of the results lies in the visual examination of the precipitation fields that reveals potentials and limitations of radar and the three combination methods. In the second part of the results, the combination methods are quantitatively investigated using the error scores applied over the entire year 2008. The result section is completed by further investigations at the daily and the annual time scale and finally, the role of time persistence is analysed and discussed.

4.1 Visual Evaluation for test cases

The twelve test cases, consisting of 24 succeeding hours each, have been selected because of their challenging precipitation configuration. In this section we reproduce a selection of example cases to gain some visual understanding of the combination methods, to highlight how successfully radar errors are corrected, but also to illustrate some of their limitations.

4.1.1 Correction of radar errors

A visual inspection of the precipitation fields during the test case events reveals that the combination methods are able to correct radar errors. This correction is particularly efficient in situations where the radar strongly over- or underestimates precipitation intensities, as shown in Fig. 4.1 for an episode of widespread heavy rainfall over Switzerland (test case 02).

The filled circles represent the gauge locations and the filling colour the measured precipitation intensity in mm/h. During this precipitation event, the radar was affected by beam shielding and attenuation. As a consequence, the radar underestimates precipitation intensities up to a factor 5 in comparison to the gauges, as it is apparent in the western part of Switzerland and around Interlaken in central Switzerland. In contrast to the radar, the three combination methods produce precipitation fields that are much closer to the intensities observed at the gauges and hence are less biased. Note that this correction takes into account the regional variation of radar errors that tend to be larger in the western parts of the country compared to central Switzerland.

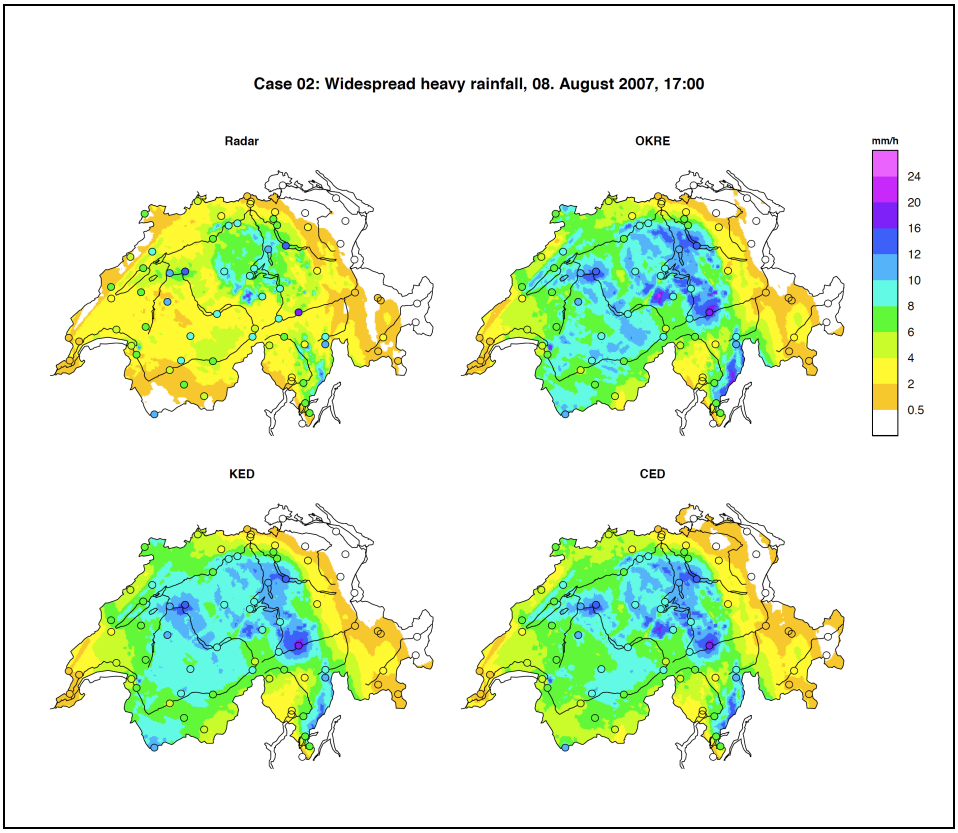


Figure 4.1: Precipitation field (mm/h), 08. August 2007, 17:00 UTC.

Quantitatively, the improvement of the combination methods over radar is also apparent in Fig. 4.2 showing the bias calculated from the cross-validation table over all 24 hourly steps of this event. The strong underestimation of the radar is reflected by a large negative bias that amounts to almost 2.5 dB, whereas the correction due to the gauge component of the combination methods is reflected by bias values near zero.

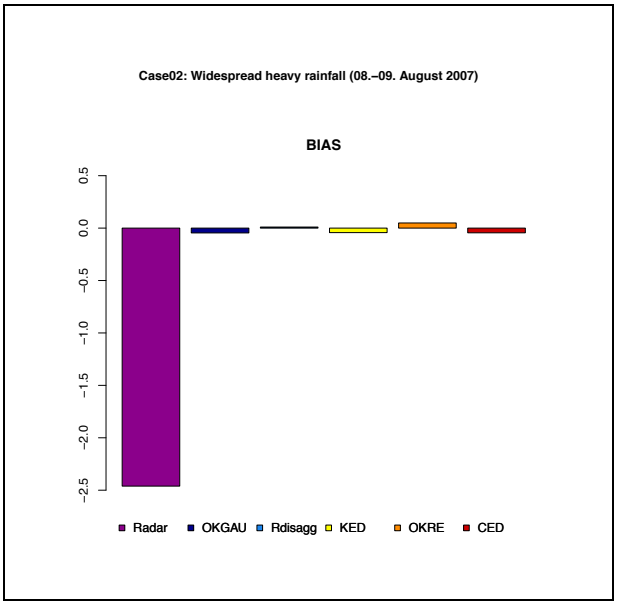


Figure 4.2: Bias (in dB) for test case 02: Widespread heavy rainfall (08. - 09.08.2007).

The combination methods are also able to correct a strong positive bias, as illustrated in Fig. 4.3, which shows a fast moving heavy hailstorm (test case 09). In this case, the strong radar overestimation is a direct consequence of deep convection and the occurrence of hail.

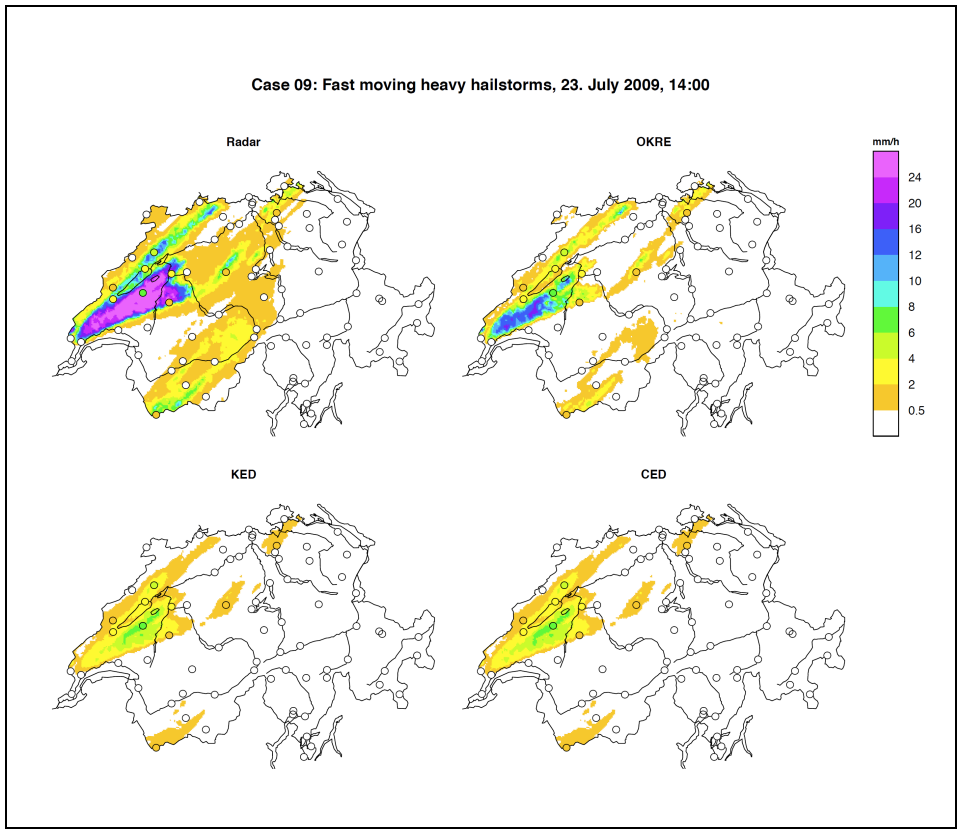


Figure 4.3: Precipitation field (mm/h), 23. July 2009, 14:00 UTC.

The radar bias is particularly apparent in the western part of Switzerland. In this region radar overestimates precipitation intensities up to a factor 6. In spite of this large bias, KED and CED are able to correct this overestimation adequately, while OKRE remains affected by the radar overestimation. Nevertheless, the three combination methods produce a precipitation field that matches the gauge values more precisely.

The improvement of the combination methods over the radar is quantitatively shown in Fig. 4.4. The examination of the bias confirms that the methods are able to efficiently correct the positive radar error. Nevertheless, OKRE is not able to reduce the positive bias of radar as efficiently as KED and CED. This feature is inherent to OKRE and is related to the comparative simplicity of this method, allowing for additive correction of the radar field only (in transformed space).

At this point, the correction of strong radar errors appears to be a particular advantage of using the combination methods. Note that the correction is not limited to overall error but also its local variation. For example, the large radar bias over the Swiss Plateau (Fig. 4.1) is corrected while the predicted precipitation remains identical to the original values in the eastern part of Switzerland. In addition, the combination methods are able to correct the radar biases even when the number of wet gauges is quite small, e.g. less than 10 in Fig. 4.3. Despite these positive aspects, each combination method is affected by its specific limitations that reduce the quality of the QPE. The following subsection presents some typical limitations attached to the combination methods.

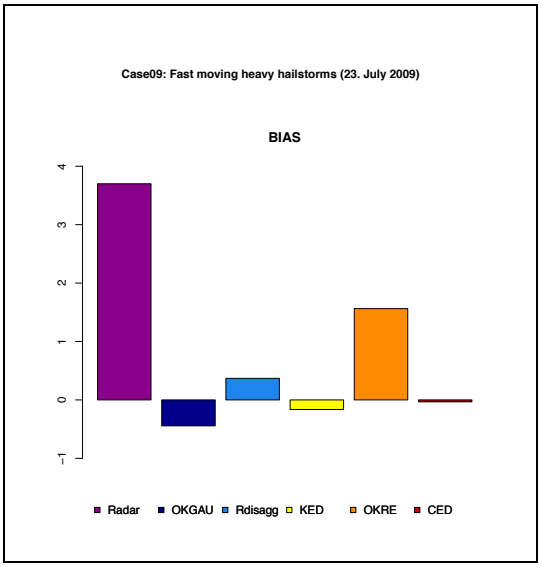


Figure 4.4: Bias (in dB) for test case 09: Fast moving heavy hailstorm (23.July 2009).

4.1.2 Limitations of the combination methods

Since the radar strongly influences the field generated by OKRE, some artefacts typically due to the radar, like strong overestimation or ground clutter, remain visible in the field of OKRE. An illustrative example is shown in Fig. 4.5 (test case 11).

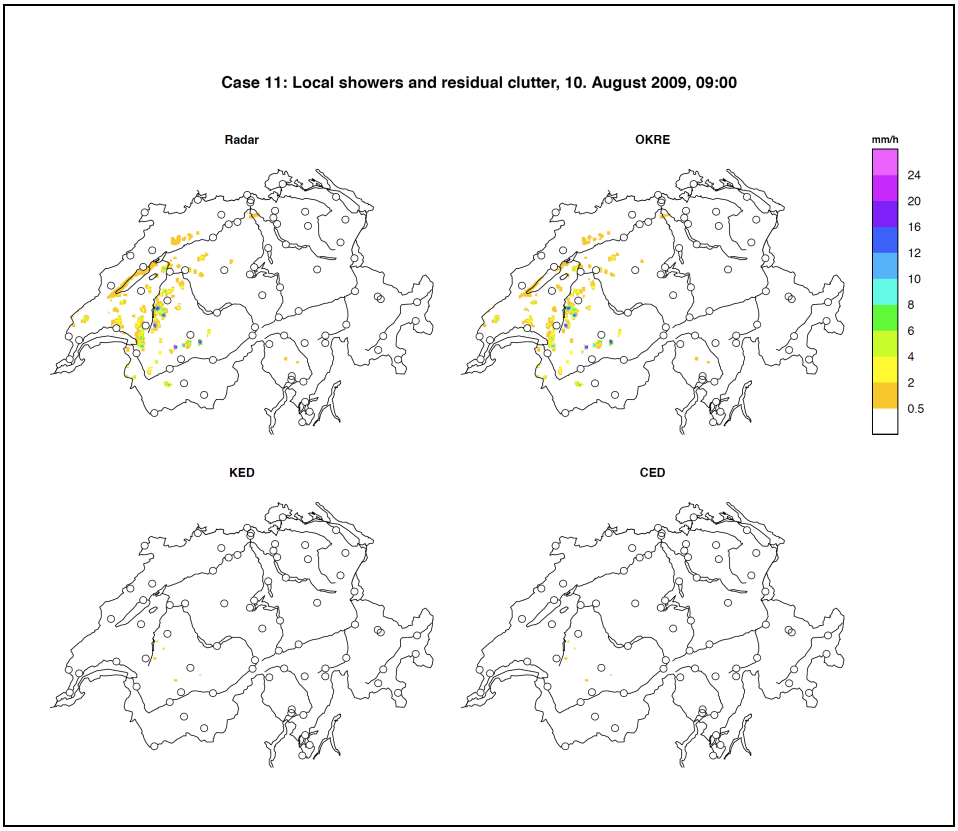


Figure 4.5: Precipitation field (mm/h), 10. August 2009, 09:00 UTC.

The residual ground clutters create fictive small-scaled precipitation systems, where in fact no precipitation is observed. In this case, the influence of radar errors on the OKRE prediction is obvious. While KED and CED are able to remove these artefacts, radar artefacts remain visible in the field of OKRE. The reason why OKRE is different from KED and CED lies in the formulation of the kriging model: while KED and CED correct the radar information by a trend parameter β, OKRE incorporates the radar field as-is (without re-scaling) and is therefore less capable of weakening radar artefacts.

Another limitation of the combination methods is that occasionally precipitation is predicted at locations where neither radar nor the gauges recorded any rainfall, as illustrated in Fig. 4.6. The examination of all the test cases revealed that both KED and CED were affected by this problem, even though CED more frequently and strongly.

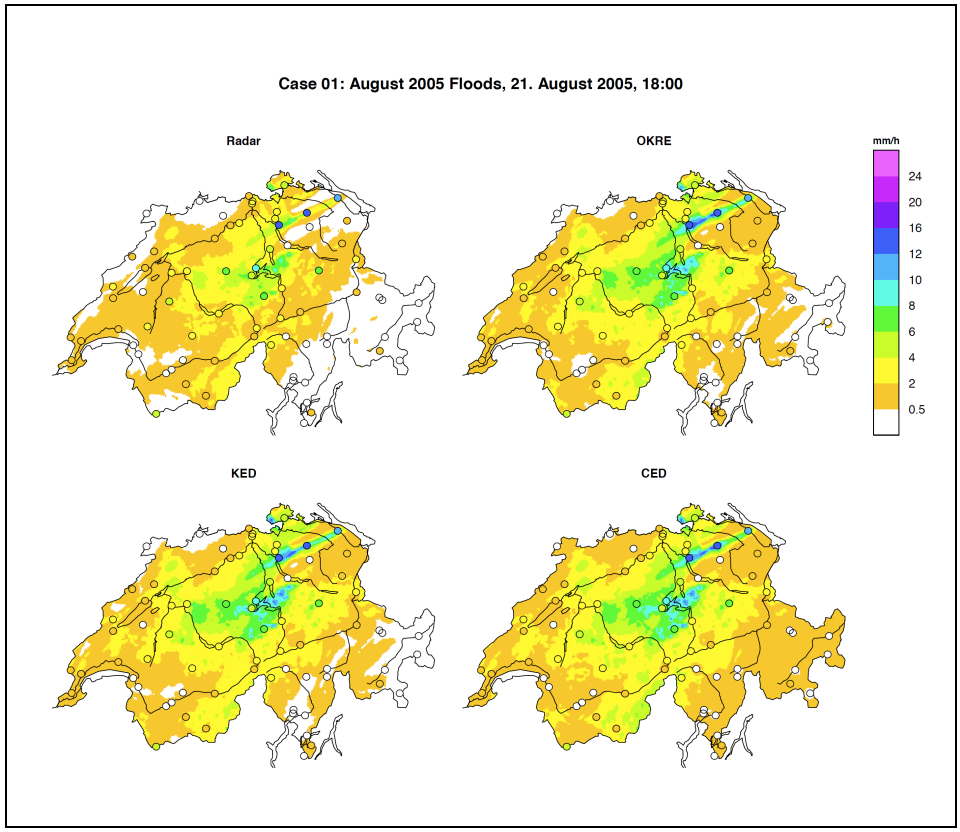


Figure 4.6: Precipitation field (mm/h), 21. August 2005, 18:00 UTC.

Fig. 4.6 shows an hour of a severe flooding event in 2005. Notice that despite the fact that neither radar nor the gauges recorded any precipitation in eastern and south-eastern part of Switzerland, CED predicted rainfall in these regions. In this case, we suspect that the differences between KED and CED are due to differences in the choice of the variogram parameters.

Another limitation is that in cases where the correlation between the radar field and the gauges is small, e.g. rain cells are detected by radar but missed by surrounding gauges, the influence of the radar field on those of the combination methods is weak. Fig. 4.7 to Fig. 4.9 illustrate three succeeding hours of a precipitation event characterized by local and intense showers (test case 10).

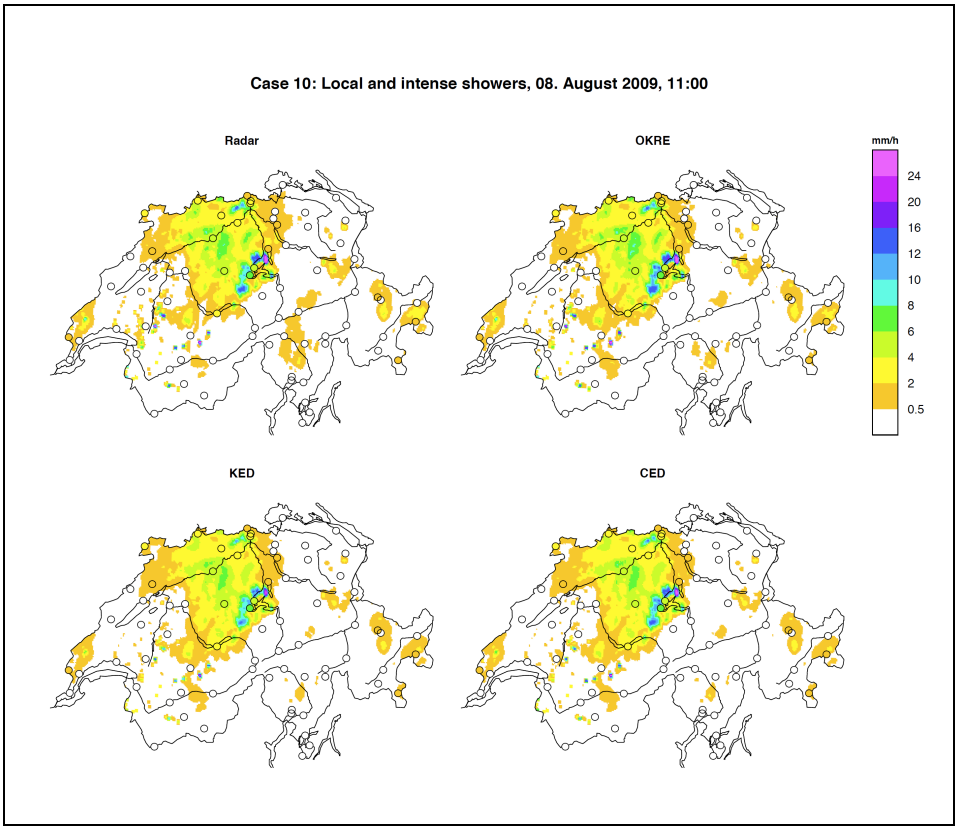


Figure 4.7: Precipitation field (mm/h), 08. August 2009, 11:00 UTC.

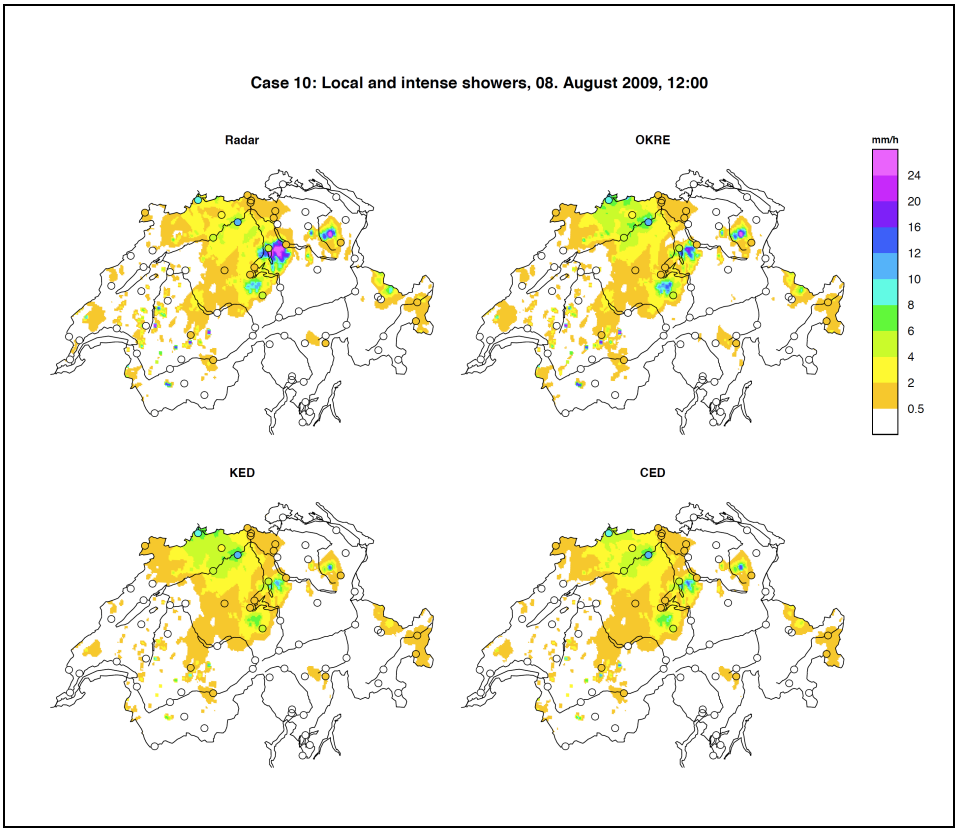


Figure 4.8: Precipitation field (mm/h), 08. August 2009, 12:00 UTC.

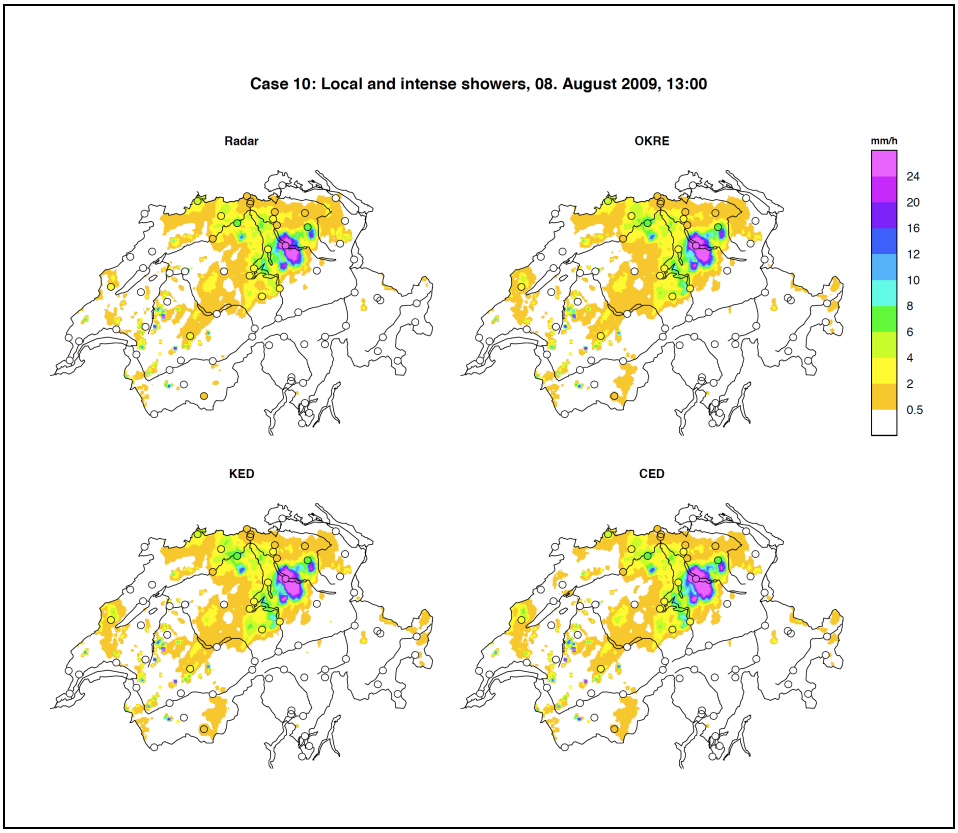


Figure 4.9: Precipitation field (mm/h), 08. August 2009, 13:00 UTC.

As illustrated in Fig. 4.7, radar detects near the lake of Zug an intense rain cell missed by the surrounding gauges. As a consequence, the correlation between the radar field and the gauges is small and therefore the combination methods KED and CED down-weight the radar, resulting in a strong reduction of this cell's rainfall intensity. During the next two hours (Fig. 4.8 and Fig. 4.9) this cell further intensifies in the radar and moves towards lake of Zurich, where finally in the last hour a gauge recorded the intense rainfall. For this hour, the combination methods KED and CED consequently take into account the radar field with a higher coefficient and the previously down-weighted cell becomes more intense in the predicted field. Therefore, when small rain cells that are only detectable by radar occur, the combination methods KED and CED can reveal a distorted temporal evolution. Note that CED, even though incorporating information from the previous hours could not provide remedy for this.

4.1.3 Additional remarks and summary

The inspection of the twelve test cases has shown that KED and CED generally produce precipitation fields very close to each other. Differences are much larger between OKRE and the other two combination methods. The largest differences between KED and CED were found in those rare cases when KED was not converging numerically and had to fall back to radar. An example hour when this happened is displayed in Fig. 4.10.

Our visual intercomparison also revealed that choosing a climatological variogram instead of an explicit estimation does not have obvious detrimental effects. Fig. 4.11 depicts a case where KED and OKRE used a prescribed variogram (i.e. fallback) while CED was estimating. The resulting fields are very similar.

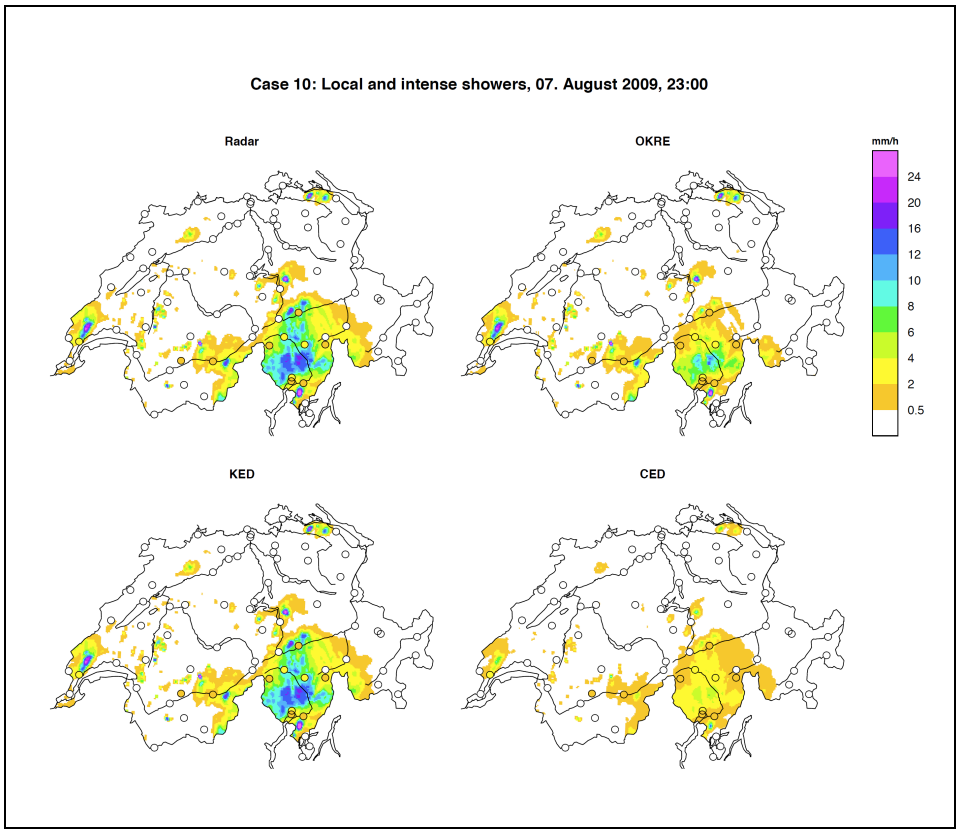


Figure 4.10: Precipitation field (mm/h), 07. August 2009, 23:00 UTC.

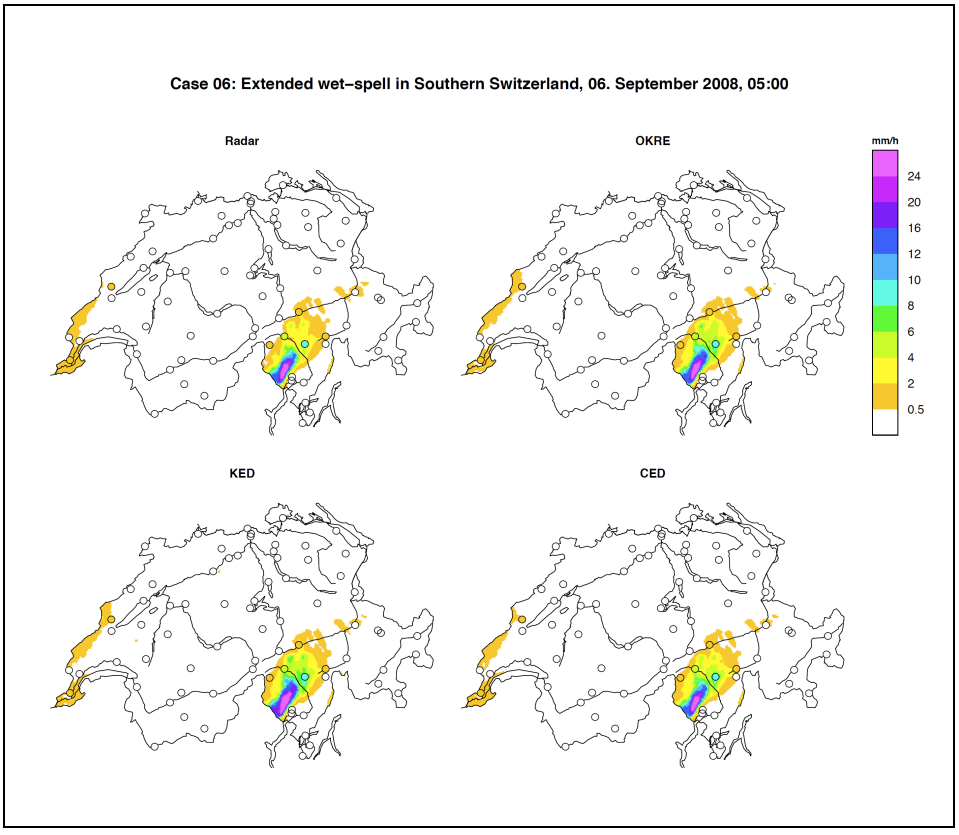


Figure 4.11: Precipitation field (mm/h), 06. September 2008, 05:00 UTC.

In summary, the visual investigation of the twelve test cases revealed that the three combination methods are able to correct strong and regionally varying radar errors and even in situations where only a few gauges are wet. Nevertheless, each combination method shows limitations that result in differences in the field structures. Unlike KED and CED, OKRE has a stronger imprint of radar and ground clutter is less well removed. We also found that KED and CED (as compared to OKRE) have a tendency to predict precipitation in areas where radar and gauges are dry. The visual impression is that this occurs more frequently with CED. The quality of the fields generated by the combination methods is also doubtful in situations where the correlation between the radar field and the gauges is low. Concerning the field structure, KED and CED are more similar than OKRE.

So far, the analysis of selected test cases gave some information about the potential and the limitations of the methods. In complement to the visual inspection of precipitation fields, the error scores were calculated for the test cases to quantify the differences between the methods. Since each test case only includes a maximum of 24 hours, the error scores are uncertain because of the small sample size. This is the reason why error scores were systematically calculated for the entire year 2008. The results of this analysis are presented in the next section.

4.2 Quantitative evaluation of one year

Because different characteristics of the methods need to be systematically examined, the year 2008 was stratified according to the flow chart showed in Fig. 4.12. In contrast to the test cases, this evaluation focuses on statistical error scores. Firstly, the performance of the methods as function of the season or the regions is investigated. Secondly, a further analysis focuses on the impact of the fallback solution on the overall performance on one side, and limited to dry hours on the other side.

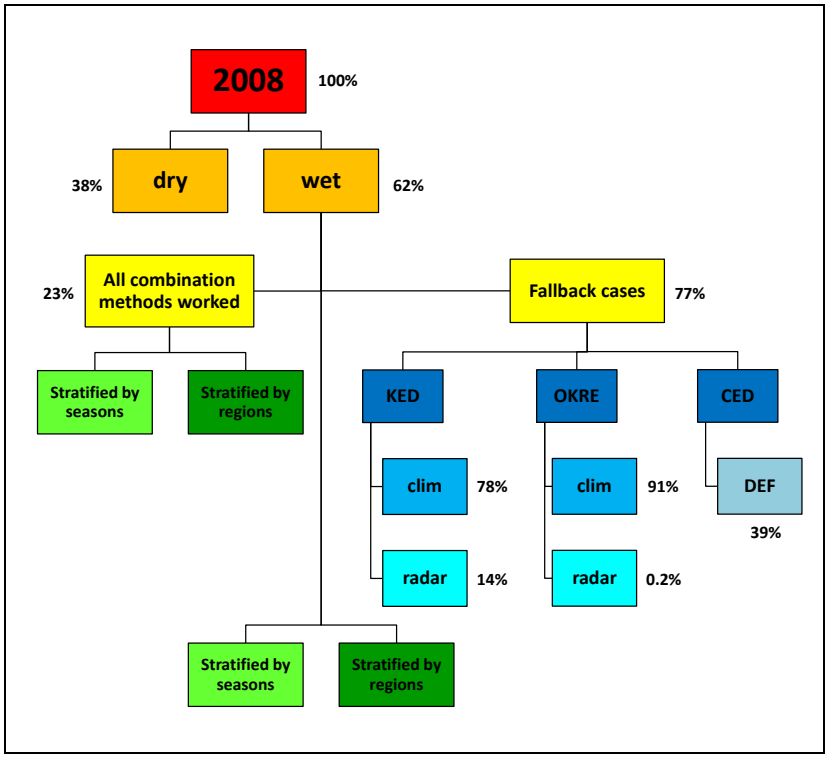


Figure 4.12: Stratification of hours in year 2008 used for the systematic evaluation. "Wet" denotes hours when at least one gauge reported precipitation (>0.1 mm/h). "All methods worked" are those "wet" hours when none of the methods had to fall back. "Fallback cases" are those hours when at least one of the methods had to fall back.

4.2.1 Wet hours stratified by seasons

As shown in Fig. 4.12, the wet subset was stratified by the four seasons. We have chosen to discuss winter and summer where the largest contrasts are observed between the methods and between the prevailing rainfall situation. In winter, precipitation is characterized by low stratiform cloud systems covering large areas. In contrast, precipitation in summer is characterized by local convective systems induced by radiation, evolving and decaying rapidly. More precipitation is generally recorded in summer than in winter.

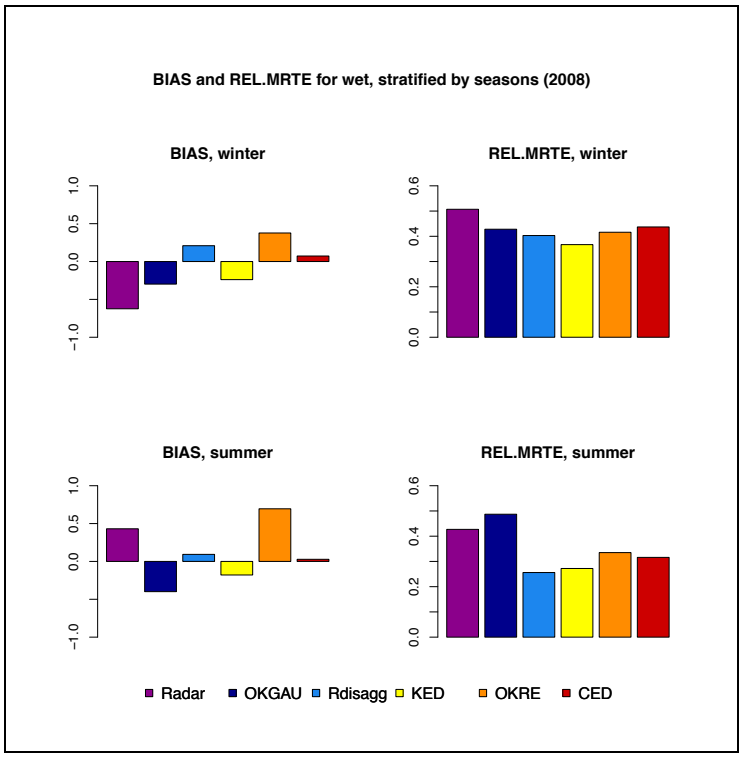


Figure 4.13: Bias and REL.MRTE in winter (top) and in summer (bottom)

As indicated in Fig. 4.13, the radar tends to underestimate precipitation in winter and to overestimate in summer. The underestimation of the radar in winter is a direct consequence of low stratiform precipitation systems that are not always fully visible for radar. In contrast the occurrence of deep convection, often associated with an abundance of hail, leads to an overestimation in summer. Concerning the correction of the radar bias, all methods are able to reduce the radar bias in winter and also in summer except for OKRE. CED shows the smallest bias of all methods.

Beside the bias, the added value of the combination methods over the radar is also confirmed by the reduction of the REL.MRTE in both seasons. The strongest reduction of the REL.MRTE is accomplished by KED. Compared to the combination methods, OKGAU is a very crude interpolation method yielding smooth precipitation fields. In situations where precipitation varies highly in space as in summer, this smoothing effect is very strong, resulting in a large random error. Overall, this seasonal comparison indicates that CED performs best in bias, while KED outperforms all other combination methods in terms of REL.MRTE.

We now turn to the HK.ERR to analyse the ability of the methods to distinguish between wet and dry conditions. Fig. 4.14 shows the HK.ERR and scatter for summer and winter. According to Fig. 4.14, the combination methods bring also added value in the HK.ERR and the scatter. In addition, all methods

perform better in summer than in winter and OKRE is the best method in terms of HK.ERR. Note also that all combination methods show considerable improvement over OKGAU. The analysis of the scatter reveals only little difference between the methods and between the seasons. It is interesting to see that the combination methods attain scores that are comparable to the benchmark "Rdisagg", even though the latter makes use of many more rain gauge measurements. This is particularly the case in winter.

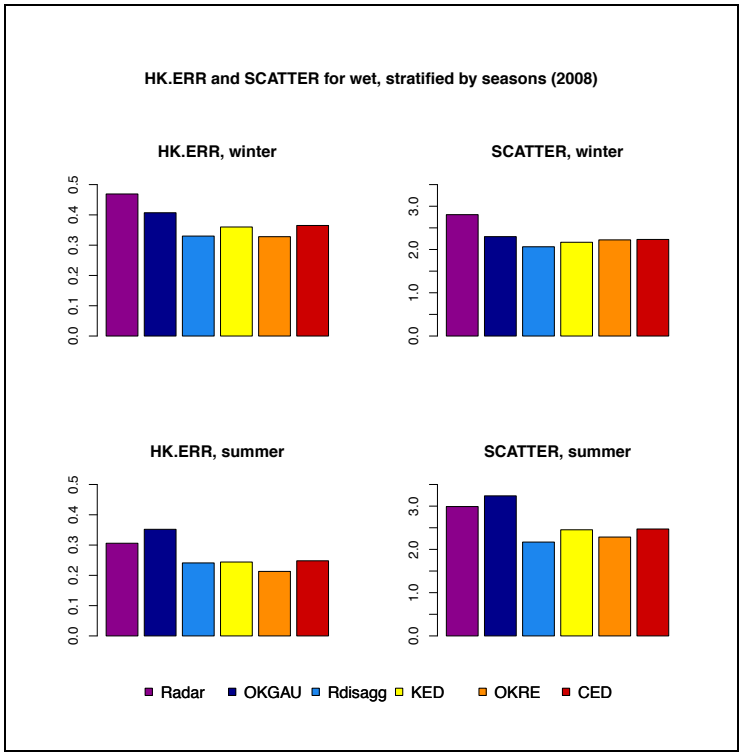


Figure 4.14: HK.ERR and scatter in winter (top) and in summer (bottom).

4.2.2 Wet hours stratified by regions

The analysis of the error scores continues with regional subsets where we analyse the Swiss Plateau and the Alps. We have chosen these two regions for comparison, since most distinct differences of the methods are found between these two regions. As shown in Fig. 4.15, the radar bias is positive over the Swiss Plateau and negative in the Alps. Notice that all the combination methods are able to correct the underestimation of radar in the Alps, while at least part of the overestimation over the Swiss plateau remains. We argue that since the radar underestimates precipitation in a complex topography such as the Alps, the entire Swiss radar field is in a first step uniformly corrected by an amplifying factor. As a consequence, while the underestimation in the Alps is adjusted, the positive bias over the Swiss Plateau increases. This mechanism is particularly large for OKRE (and to a lesser extend for CED) where a positive bias in the Alps exists. Finally, the only method able to reduce the positive radar bias over the Swiss Plateau is KED. In addition, note that the disaggregated dataset shows the strongest radar bias reduction for both regions. This is not surprising because radar enters in this method only via its temporal evolution not via its quantitative values.

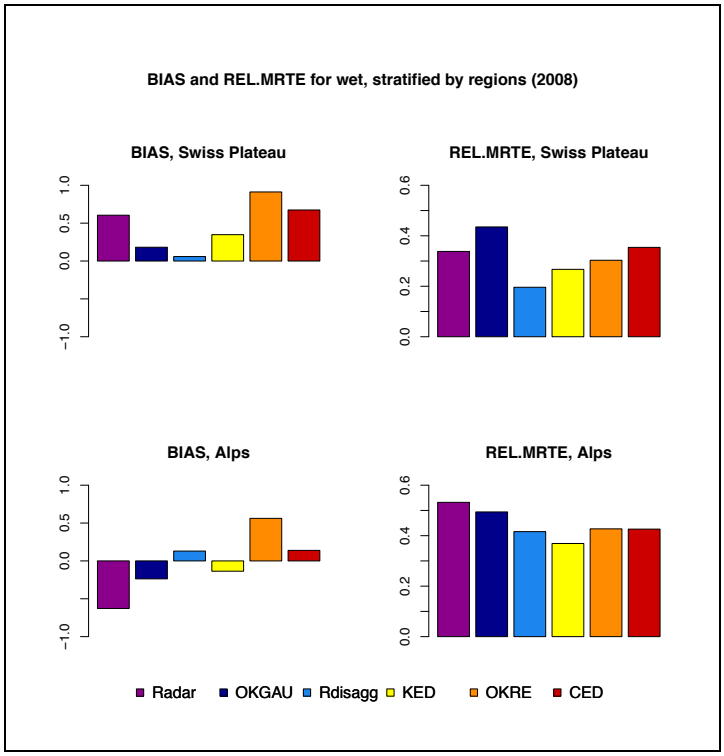


Figure 4.15: Bias and REL.MRTE for the Swiss Plateau (top) and the Alps (bottom)

A glance on the remaining scores REL.MRTE, HK.ERR and scatter in Fig. 4.15 and Fig. 4.16 shows that the error scores are smaller over the Swiss Plateau than in the Alps. This is consistent with a spatial correlation structure being more complex in mountainous regions and therefore larger uncertainties of precipitation estimates. Finally, the largest differences across the methods are found for the REL.MRTE over the Swiss Plateau, where KED outperforms all other combination techniques. Note also that all combination methods outperform the disaggregated dataset in terms of HK.ERR. We suggest that for a good performance in HK.ERR the inclusion of radar information is generally an advantage for a method. Since the disaggregated dataset is less influenced by the spatial radar structure than the combination methods it is not able to reach the same performance. Finally there are little differences across the methods with regard to HK.ERR and scatter.

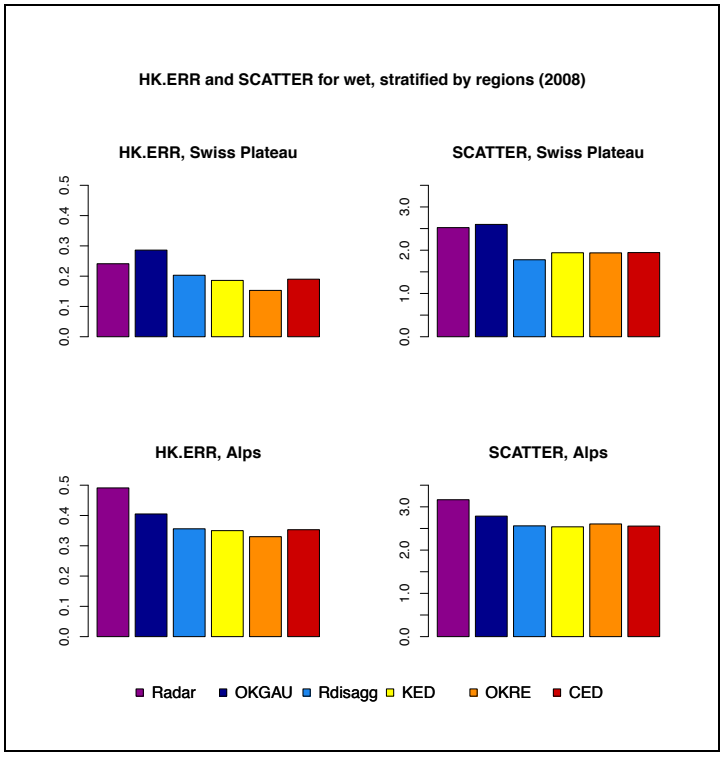


Figure 4.16: HK.ERR and scatter for the Swiss Plateau (top) and the Alps (bottom).

4.2.3 Regular calculation versus fallback solutions

So far, the dependence of the performance in the wet subset has been investigated in terms of seasons and regions. Keep in mind that an hour is considered as wet as soon as a single station records a value larger than 0.1 mm/h. Hence, the wet subset studied above also included hours when a method cannot estimate the empirical variogram and needs a fallback solution. In order to understand the optimal potential of the combination methods, the wet dataset is separated in two complementary subsets: the hours where all methods could simultaneously estimate the variogram, and the hours when at least one method needs a fallback solution (see Fig. 4.12). 77% of all wet hours are fallback situations while only in 23% of all wet hours the methods could estimate the model parameters regularly from the data. Here we study differences between regular and fallback operation.

Before investigating the fallback solutions we start with the results of the entire wet dataset of 2008, as presented in Fig. 4.17. For the entire wet dataset, radar is affected by a negative bias. We assume that this underestimation is a consequence of strong underestimation in the Alps and during winter. Fig. 4.17 shows that all combination methods are able to correct the negative radar bias and to reduce the error spread of the radar. Because it does not weight the radar observations, OKRE is unable to reduce the radar error in the same magnitude as KED or CED. CED outperforms KED and OKRE in terms of bias but is queuing behind in terms of REL.MRTE. Even though the radar demonstrates already good ability in distinguishing between wet and dry, the other methods are able to improve this performance.

Figure 4.17: Error scores for wet hours (2008).

Even though the REL.MRTE and the scatter relate to some measure of random error, Fig. 4.17 shows that the REL.MRTE and the scatter differ in the ranking of the combination methods. While the REL.MRTE shows distinct relative differences, the scatter does not. We explain this divergence by the different prominence of small and medium precipitation intensities in these error scores. While the scatter discards observational and predicted values smaller than 0.5 mm/h, the REL.MRTE considers the entire spectrum of precipitation intensities. When calculating the REL.MRTE with a dataset that leaves small precipitation values out, as shown in Fig. 4.18, the relative differences between the methods decreases while the scores increases.

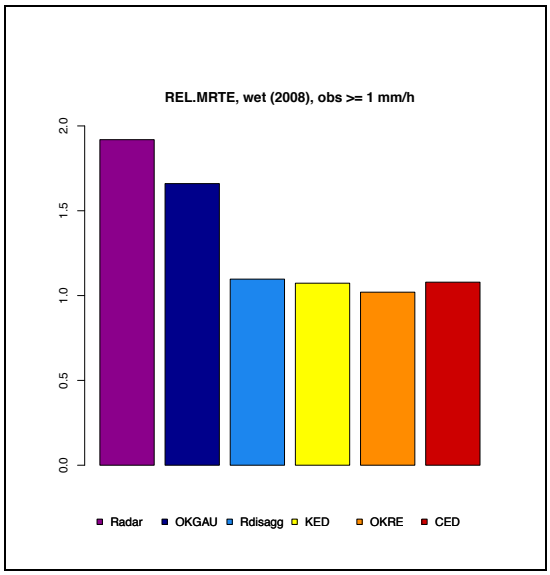


Figure 4.18: REL.MRTE for wet case 2008, when leaving out station values smaller than 1 mm/h.

Comparing Fig. 4.17 and Fig. 4.18 reveals that KED shows the strongest error increase relative to the other combination methods. Hence KED shows better ability to estimate small precipitation values than the other combination methods.

We focus now on a subset of 2008 where all combination methods were able to estimate the variogram parameters from the data, i.e. when 10 or more rain gauges were wet. Note that this corresponds in fact to only 23% of the wet dataset. In this case, Fig. 4.19 shows that the radar is affected by a negative bias that is almost twice as large in magnitude than all other subsets studied above. The improvement of the combination methods over the radar is apparent in all error scores. The ranking of the combination methods within the error scores is similar to the entire wet dataset (compare with Fig. 4.17). Yet, the relative difference between the methods is more contrasted in the set of data considering all the wet hours in 2008 than for the subset where all the method estimated the variogram. This fact stresses the importance of the fallback solutions for the overall performance.

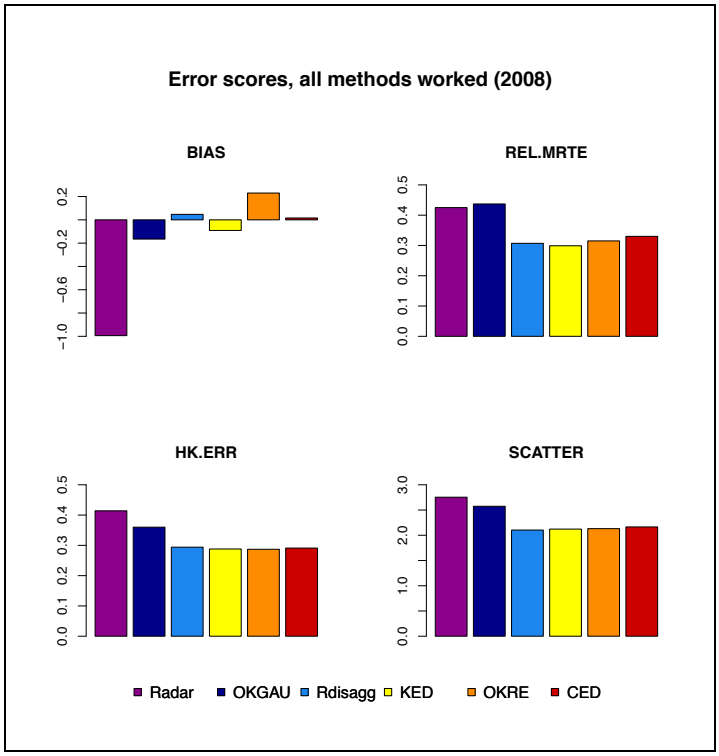


Figure 4.19: Error scores when all methods worked simultaneously.

As a matter of facts, in 77% of the wet hours, at least one combination method will use a fallback solution (see Fig. 4.12). The frequent use of fallback solutions primarily by OKRE and KED indicates that the performance of a method does not only depend on the kriging model itself but also on the appropriate choice of the fallback solution. Consequently, the combination methods and their corresponding fallback solution are examined more in details in the following sections. Notice that each individual method is analysed separately, ignoring whether any other combination method is estimating its variogram parameters or uses a fallback solution.

KED uses a climatological variogram

In fallback conditions, KED uses either a climatological variogram with predefined variogram parameters (78% of the time) or simply takes the radar field as-is (14% of the time). More details on the decision scheme can be found in section 3.4.3. The remaining hours (8%) corresponds to situation where KED is able to estimate the variogram parameters but some other method is not. The scores corresponding to climatological variogram subset of KED are presented in Fig. 4.20.

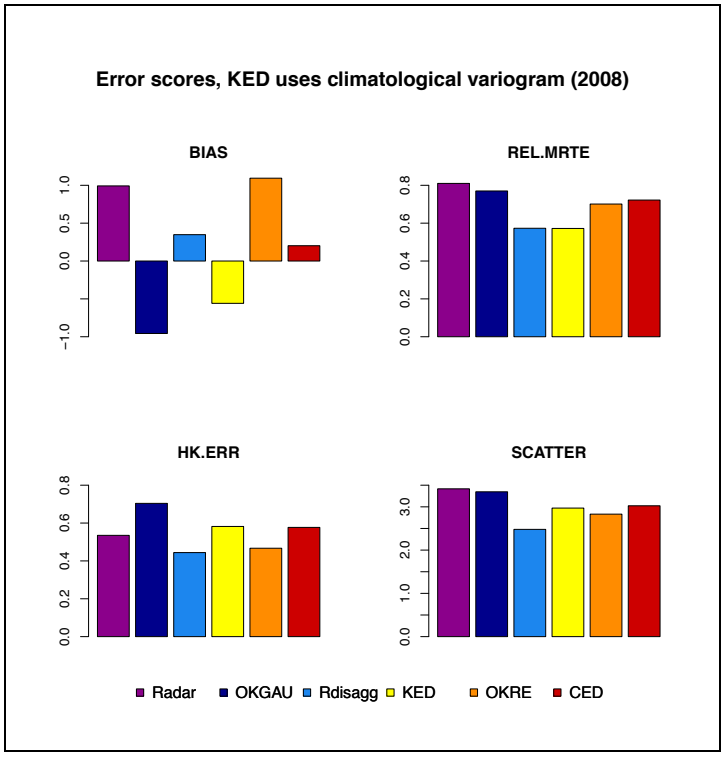


Figure 4.20: Error scores when KED uses a climatological variogram as fallback solution.

Despite the fact that a climatological variogram is suboptimal compared with an empirically estimated variogram, KED has smaller errors than radar except for HK.ERR. Moreover, KED is better than other methods when considering REL.MRTE. Notice that KED uses twice as much the fallback solution compared to CED and hence the subset considered here includes many hours where CED went for regular estimation. This illustrates that the more frequent decision for fallback together with suitable default variogram parameters did not have a detrimental influence compared to CED. In contrast, the more frequent decision for fallback did help to reduce effects from difficult sampling conditions: The differences in REL.MRTE between KED and CED are larger in all wet hours than in hours when all methods could properly estimate (compare figures 4.17 and Fig. 4.19). This illustrates that a good fallback choice plays an important role for the performance of a combination method.

KED uses the radar field

When the algorithm of KED fails to fit a variogram for whatever reasons, the radar field is taken as prediction. Note that this occurs in about 6% of all hours in 2008 (see Fig. 4.12). In this case, it is not surprising that the performance of the radar and KED are almost identical (see Fig. 4.21). The small differences are due to KED transforming radar values smaller than 0.05 mm/h into 0 mm/h. Therefore, the negative bias of radar is enhanced in KED while in contrast the error spread is slightly smaller for KED than for radar. Surprisingly, even when KED needs the radar as fallback solution, it outperforms OKRE and CED in REL.MRTE. Moreover, in this subset of hours CED reaches the critical score of REL.MRTE equal to one, indicating a low explanatory ability not better than the station gauge variability. Nevertheless, OKRE and CED are able to improve the radar errors when considering HK.ERR and the bias.

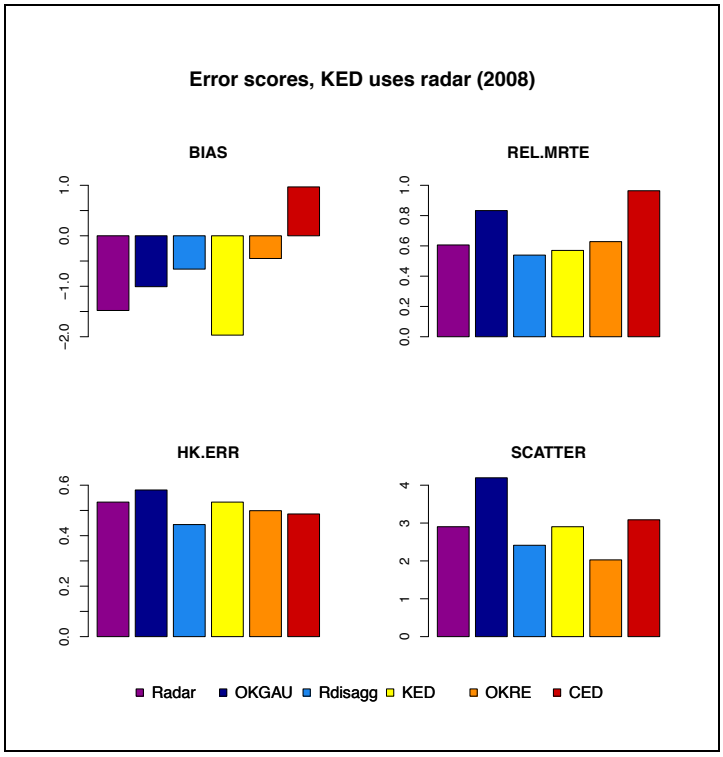


Figure 4.21: Error scores when KED uses the original radar information as fallback solution.

OKRE uses a climatological variogram

In 91% of all fallback cases, OKRE uses a climatological variogram with predefined variogram parameters. In about 0.2% of the cases, OKRE uses the radar as alternative fallback solution. Since this fraction is too small, an analysis is not conducted. The remaining fraction corresponds to the hours when OKRE is able to estimate the variogram parameters from the data and is of course not relevant to the study of the fallback role. As shown in Fig. 4.22, the main feature of OKRE during fallback is a large overestimation of the same order as the radar.

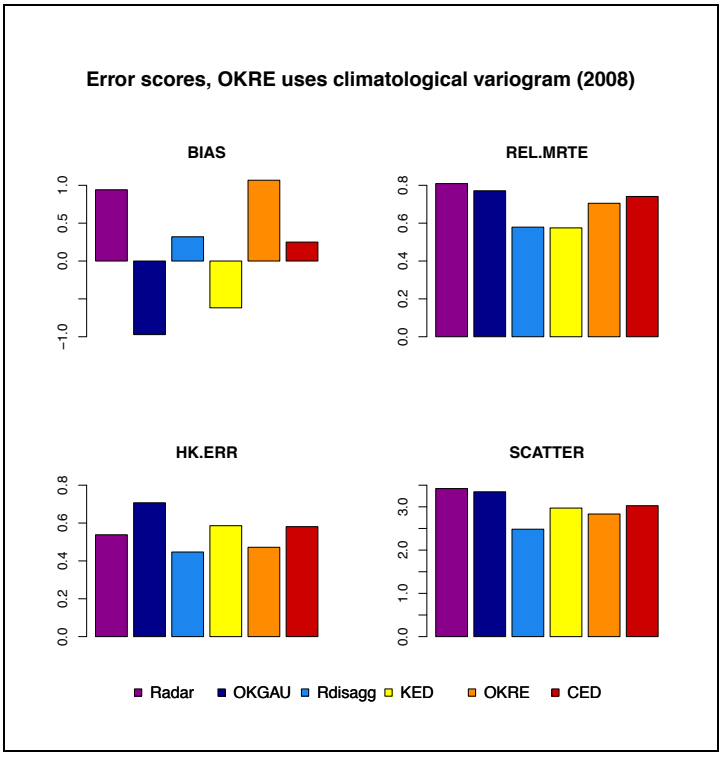


Figure 4.22: Error scores when OKRE uses a climatological variogram as fallback solution.

CED uses a default variogram

In contrast to KED and OKRE, CED estimates the variogram parameters as soon as 5 wet gauges are available (versus 10 wet gauges for KED and OKRE). This is the reason why CED only uses half as often the fallback solution than KED and OKRE. In addition, CED uses in the fallback solution a variogram with a range of 15 km and a nugget-sill ratio of 0.1. The choice of these parameters implies that the spatial autocorrelation of the residual field is restricted to 15 km. Compared to KED this range is much smaller and consequently the predicted field of CED is more influenced locally, in the immediate vicinity of stations. Since the range of CED is smaller than the typical distance between two stations, it is assumed that the field prediction in-between stations will relax to the trend.

The large radar bias and the large radar score in REL.MRTE shown in Fig. 4.23 suggest challenging precipitation situations. During these hours, CED is able to correct the radar bias and to decrease the error spread. Nevertheless, CED reaches the critical value of REL.MRTE equal to one, implying no additional ability compared with the variance of the gauge stations. In addition, CED cannot beat the radar HK.ERR score (but this feature is also shared with KED). Another interesting result is that KED outperforms all other methods in terms of REL.MRTE and scatter, even in these challenging precipitation situations.

At the hourly time scales, sampling conditions that do not allow for completely data driven parameter estimation are frequent, and hence the criteria when to relax to default parameters and the setting of those is seen to influence the overall accuracy of the methods. Even though KED and OKRE opt for more frequent fallback to prescribed variogram parameters, their accuracy is not inferior to that of CED, except that OKRE has a larger positive bias, but this is a general feature of this method. Moreover, KED is seen to outperform CED particularly when both methods are falling back on prescribed settings. This is particularly evident in REL.MRTE. It suggests that the prescriptions of KED (larger range, larger nugget-sill ratio) are more suitable.

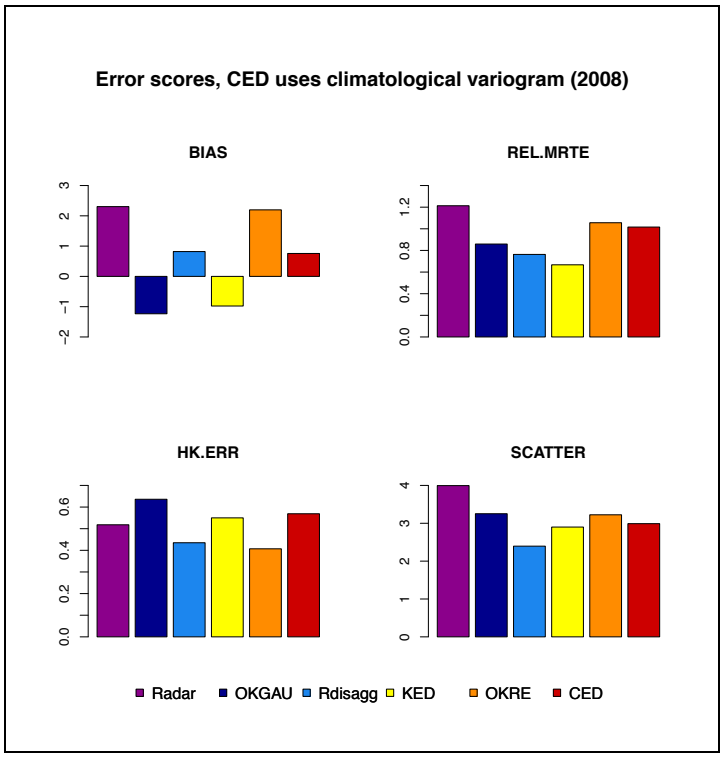


Figure 4.23: Error scores when CED uses a variogram with predefined parameters as fallback solution.

4.2.4 Dry cases

So far, only results referring to wet hours are presented. Since the dry cases represents 38% of all hours during 2008, they deserve a special attention. How the methods predict precipitation in dry cases may sound paradoxical. Nevertheless, as long as the radar measures precipitation, methods might still provide non-zero predictions. Here we use the mean deviation of the predicted values from 0 mm/h to quantify this. Results are shown in Fig. 4.24. As expected, OKGAU shows no deviation from 0 mm/h because the interpolation of dry gauges yields a dry field. KED has also no deviations since its algorithm is programmed to set the predicted precipitation to zero as soon as all gauges are dry. Despites the fact that the errors of the remaining methods are small, the annual sum of, e.g. CED errors, account up to 2.5% of the annual precipitation amount, calculated from the station values.

What precipitation intensities are responsible for these errors? The density distribution of the errors is therefore examined in detail for each method. Fig. 4.25 to Fig. 4.27 show that the radar, OKRE and CED have different deviation spectra. The small precipitation intensities for the radar and OKRE are the main component of the total error. In contrast, CED predicts more often rain in a higher precipitation category than radar and OKRE. There may be two reasons for this: Firstly an extrapolation in time of wet conditions due to the inclusion of information from previous (wet) hours, and secondly the different treatment (absence of filtering) of very small radar intensities. As a consequence, the error distribution of CED results in the largest mean deviation for the dry hours subset.

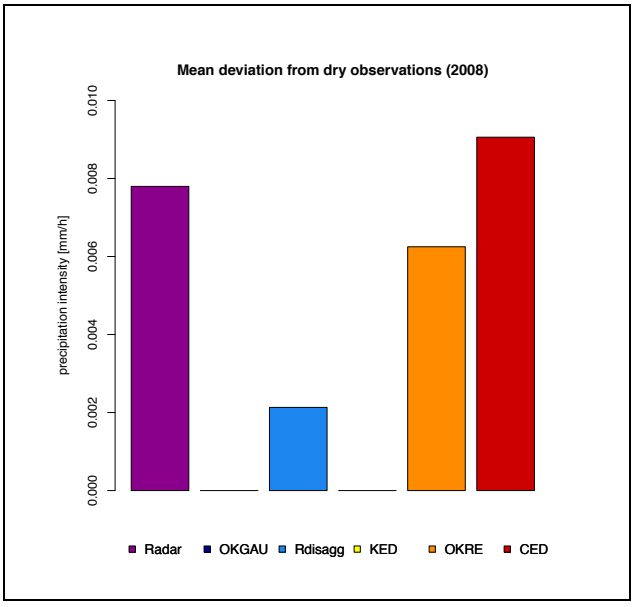


Figure 4.24: Mean deviation from dry observations for 2008.

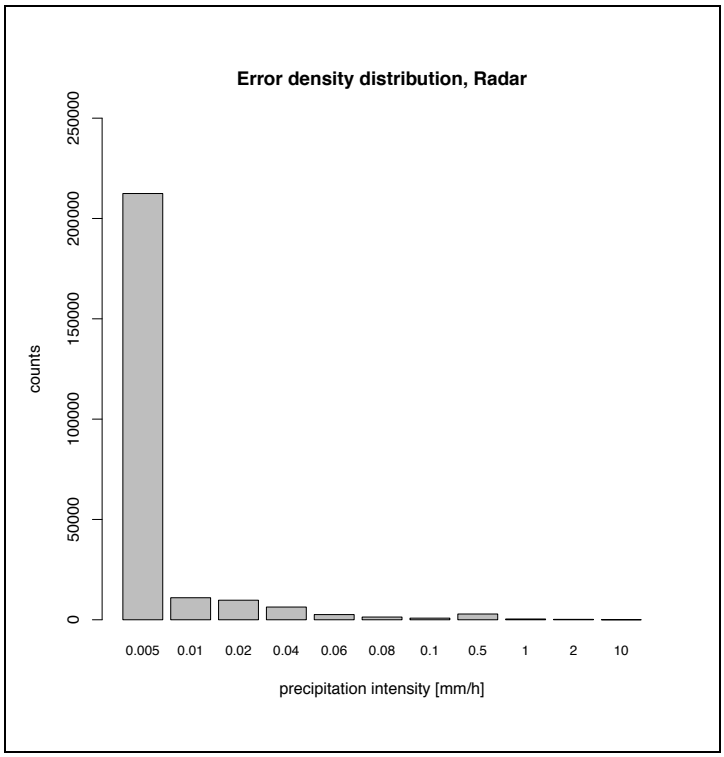


Figure 4.25: Error density distribution for radar during dry hours in 2008.

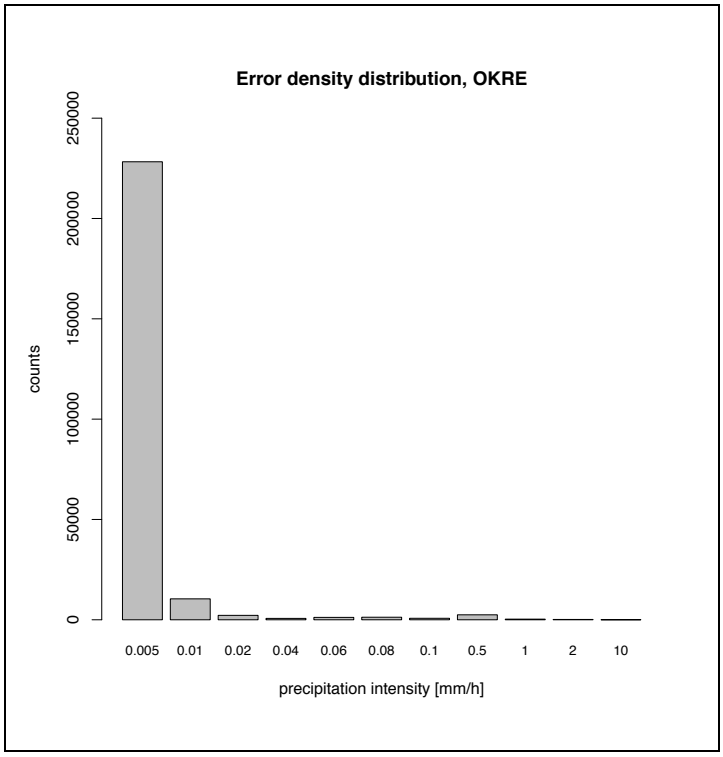


Figure 4.26: Error density distribution for OKRE during dry hours in 2008.

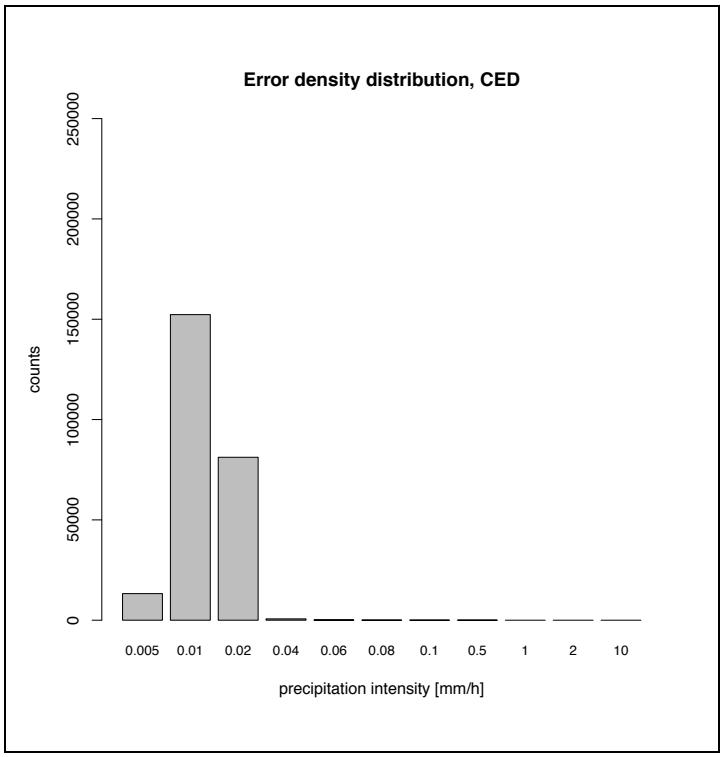


Figure 4.27: Error density distribution for CED during dry hours in 2008.

We conclude this section by summing up the findings of the analysis of the year 2008 at the hourly time scale. Independently of the season or the region, the combination methods bring an added value compared with the radar. Generally, the combination methods worked better in summer than in winter and better for the Swiss Plateau than for the Alps. The examination of the fallback solutions revealed that the main differences between the combination methods are due to the fallback solutions and that the performance of the combination methods particularly depends on the appropriate choice of the fallback parameters and criteria. Additionally, note that KED always outperforms the other methods in REL.MRTE, even when it uses the fallback solution. The analysis of the dry hours has illustrated that differences in the mean deviation are a direct consequence of how the methods handle small radar values. Because CED uses previous hours for information it can predict small rain rates where no precipitation is in fact observed.

4.3 Further investigations

4.3.1 Comparison at daily time scale

So far, the results for the year 2008 were focused on the performance of the combination methods at the hourly time scale. In order to evaluate the combination methods more thoroughly with a much larger and independent validation dataset, the evaluation is repeated for daily precipitation sums. For this analysis the daily gauge observations (extracted from data warehouse of MeteoSwiss), consisting of approximately 450 stations, are compared to daily accumulations of the combination methods at corresponding nearest grid points. Since the network working at the daily scale includes many more gauges than that of the automatic stations, more reliable and representative error measures are expected. For this analysis the hourly disaggregated dataset is termed as gridded stations, since this dataset was re-aggregated to daily fields. Here it serves as a primary reference.

The results of the daily precipitation sums are shown in Fig. 4.28 and reveal that the combination methods bring added value over radar also at the daily time scale. Concerning the bias, CED reaches a score similar to the gridded stations, which clearly outperforms all other methods. A striking feature seen in the REL.MRTE and in the HK.ERR is a clear contrast between the methods. Even OKGAU outperforms the radar for the distinction between dry and wet. At the daily time scale it seems that the gauge measurements provides better information about the field variability than the radar. The same is true for the REL.MRTE. For HK.ERR and REL.MRTE Fig. 4.28 shows that the larger the influence of the radar on a method is the larger are the errors. Since the gridded stations, conducted from the disaggregated dataset and in Fig. 4.28 termed as gridding, contain the daily gauge observations, it is not independent of the validation dataset. Nevertheless, the predicted field generated by gridded stations is a valuable reference method in terms of what can be achieved by the combination methods at best. However, the field of the gridded stations is not error-free since the observational value is compared to the nearest grid point of the gridding.

The error scores of Fig. 4.28 revealed that a simple ordinary kriging of gauges performs in terms of REL.MRTE and HK.ERR similar than KED and even better than OKRE and CED. Moreover, it reveals that at daily time scale radar alone cannot compensate the large information gain available from the daily rain gauge network.

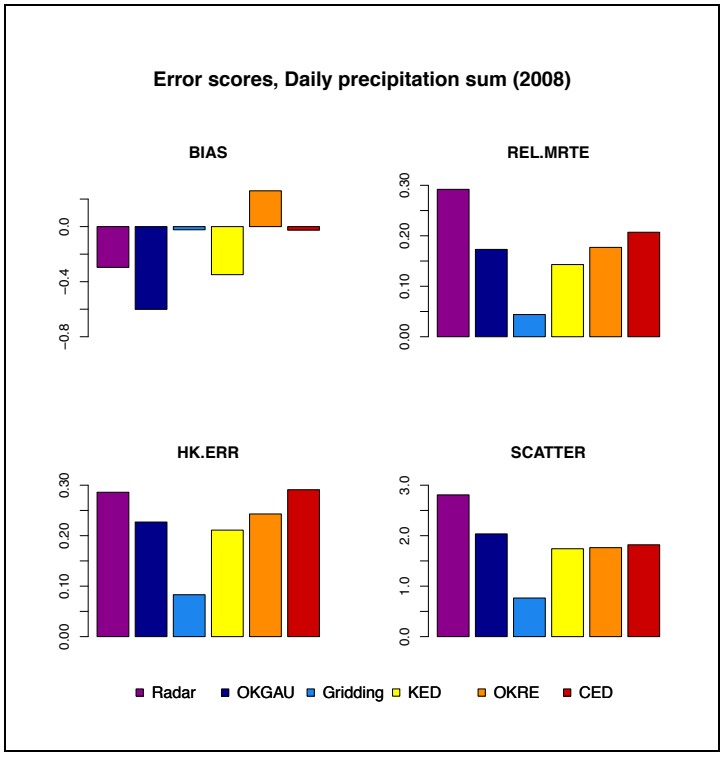


Figure 4.28: Error scores, daily precipitation sums for 2008.

4.3.2 Field comparison of annual precipitation sums

This section begins with an analysis of the fields of annual precipitation sums. Then the differences between the predicted field and the gridded station values are analysed and compared.

Fig. 4.29 reveals two regions with a particular regime: the Valais, in the south-western part of Switzerland and the Engadin, located in south-eastern part. On the one hand, these regions are remote valleys where the visibility of radar is restricted by mountains. On the other hand, both regions are inner alpine dry valleys characterized by smaller precipitation sums. Therefore the radar shows substantially smaller precipitation sums in these regions compared with the combination methods. A positive result is the ability of the combination methods to correct the radar underestimation. Note also that KED is dryer than CED over the Swiss Plateau. This result is consistent with KED having often a negative bias and CED a positive bias.

We now turn to the field differences between the methods and the gridded station values as shown in Fig. 4.30. Consistent with the previous result, this analysis reveals a strong radar overestimation over the Swiss Plateau and an underestimation in the Alps. Overall, the largest correction of this regionally varying bias is achieved by KED. Note a striking feature in the eastern part of canton Ticino where strong overestimation close to strong underestimation occurs. This region shows strong gradient in the radar bias, which is related to beam shielding effects on the Lema radar. The existence of several rain gauges in the shielded area (see Fig. 2.2) lets combination methods to compensate the radar underestimation. However, the extrapolation of this correction into the area that is not shielded results in a remarkable overestimation here.

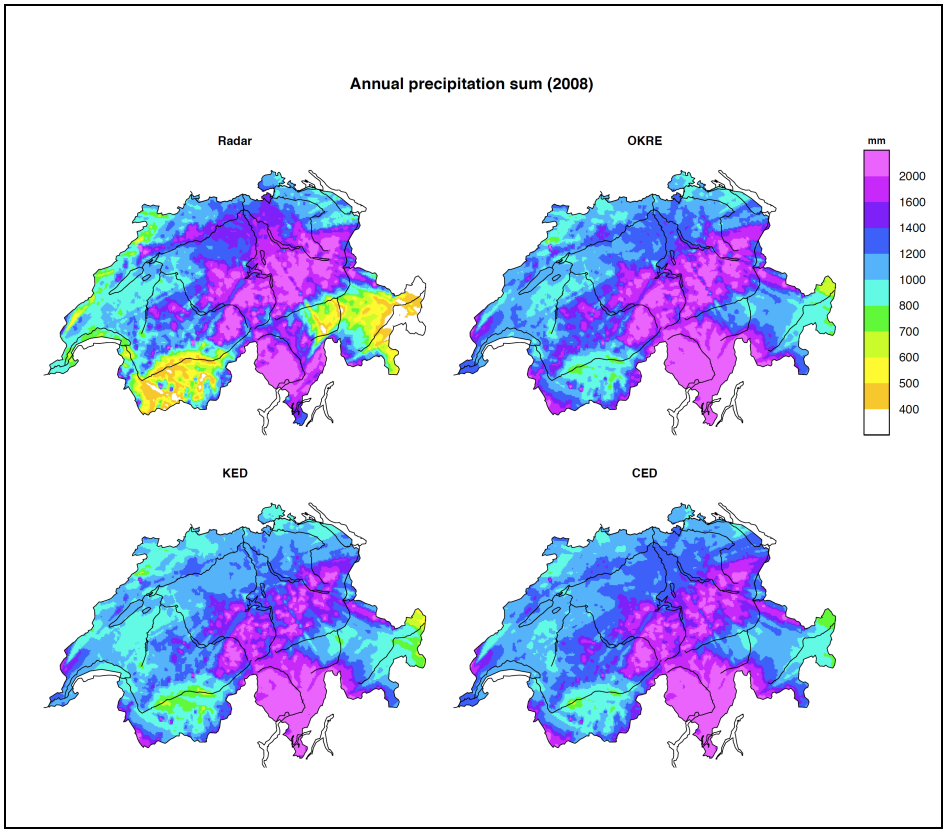


Figure 4.29: Annual precipitation sum (2008), as derived by accumulating hourly fields of radar and the combination methods.

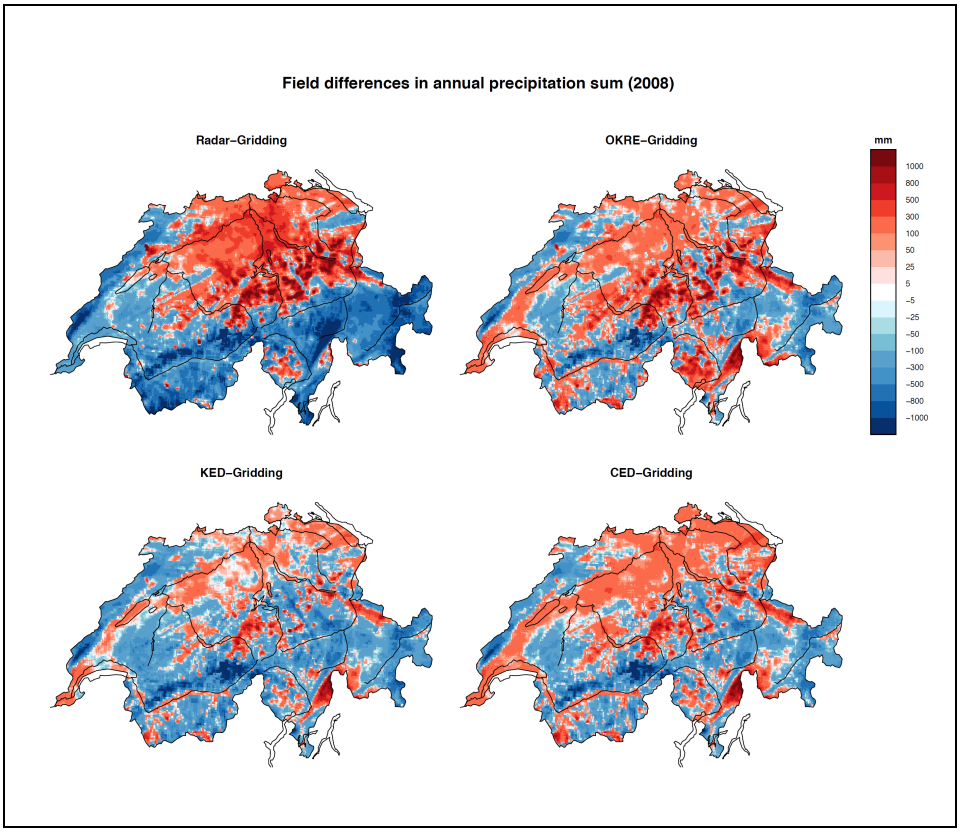


Figure 4.30: Differences of the annual precipitation sum (2008) between the radar / combination methods and the gridded station values.

4.3.3 Time persistence

As explained in the method section, CED incorporates information from the previous hour. Because the results presented so far did not substantially differ from those of KED, the added value of the time persistence compared with KED is studied more in detail in this section. For this purpose we stratify the evaluation dataset (year 2008) according to the temporal correlation of succeeding hours. The information flow from the previous hour in CED depends on the time correlation of residuals and hence it can be expected that CED can take advantage of its time feature especially during periods of high persistence in the precipitation field.

Figures 4.31 and 4.32 illustrate fields with high respectively low persistence. Fig. 4.31 illustrates a high persistent case of a slow-evolving precipitation episode while Fig. 4.32 represents a rapid-evolving episode associated with limited persistence. As a measure of persistence, the squared Pearson correlation between the radar field at the hour *t* and the previous hour *t-1* is computed for the entire year 2008 (wet hours only and when all methods could estimate the variogram). In order to limit the effect of outlying values, the correlation is calculated for root-transformed precipitation data. The frequency distribution of the correlation coefficient R^2 is shown in Fig. 4.33.

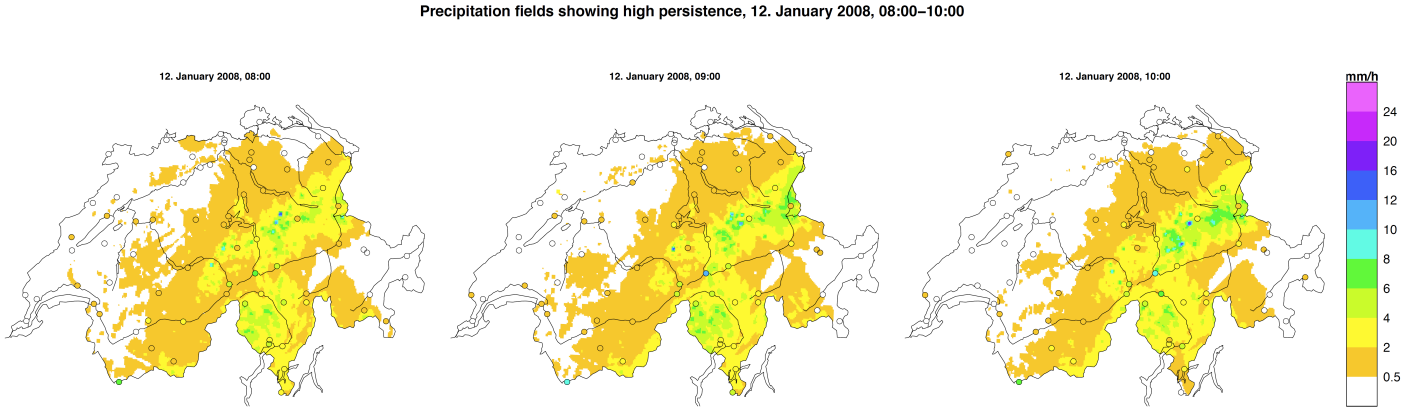


Figure 4.31: Precipitation fields showing high time persistence.

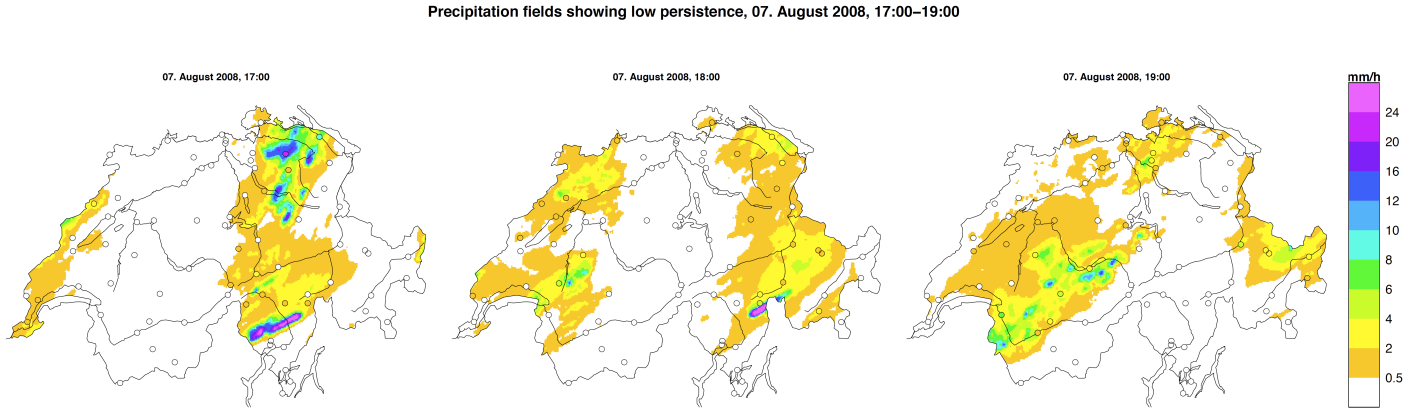


Figure 4.32: Precipitation fields showing low time persistence.

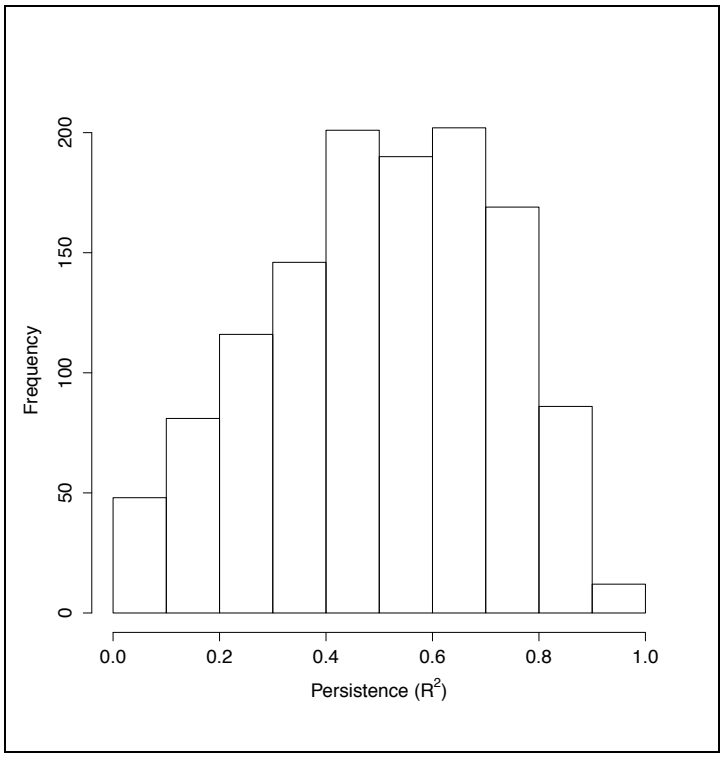


Figure 4.33: Frequency distribution for lag-1 hour correlation R^2.

The distribution of the lag-1 hour correlation is negatively skewed with squared correlation coefficient often smaller than 0.5, implying that persistent episodes happen less often than cases with weak persistence.

The scatterplot of the persistence versus the difference in REL.MRTE between CED and KED is shown in Fig. 4.34. Each circle stands for a particular hour. Positive values on the x-axis indicate a larger REL.MRTE score for CED than for KED, meaning that CED shows larger error spread than KED.

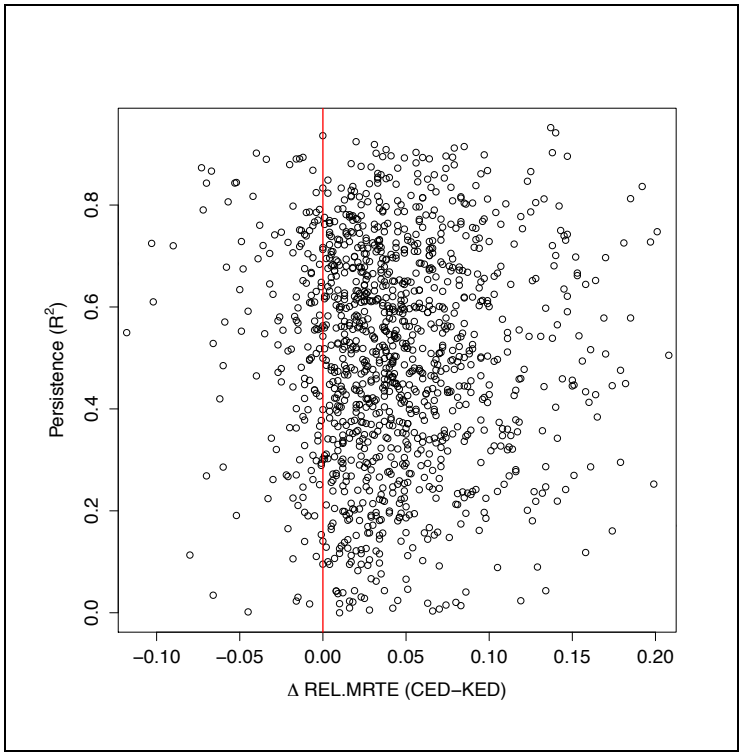


Figure 4.34: Error spread difference between CED and KED as a function of the persistence.

The difference between CED and KED is often positive indicating the overall better performance of KED compared to CED. More importantly, Fig. 4.34 does not show any systematic dependence between persistence and performance improvement in terms of REL.MRTE. One can argue that the calculation of the REL.MRTE for single hours including only 75 stations does not guarantee robust results. Therefore, persistence was further investigated by calculating the error scores for different persistence categories. The results are shown in Tab. 4.1 where each error score is analysed. As the positive values indicate, CED produces larger errors than KED in each category and for all error scores.

Table 4.1: Differences in error scores for the persistence quantiles

	Differences in error scores between CED and KED			
Persistence quantile	**BIAS**	**REL.MRTE**	**HK**	**SCATTER**
0% -25%	0.126	0.031	0.000	0.067
25%-50%	0.129	0.033	-0.008	0.006
50%-75%	0.118	0.032	-0.001	0.063
75%-100%	0.056	0.028	-0.004	0.058

As a consequence, the specific added value of CED by using information from the past hours could not be demonstrated, at least at the hourly time scale.

5 Discussion and Conclusions

Preliminary experiments to combine rain gauge with radar information in Switzerland were first undertaken at the daily time scale (Erdin, 2009). These revealed that geostatistical interpolation methods substantially bring added value to monitor precipitation at daily time scale. Whether combination methods are also profitable at the hourly time scale was yet open: the small number of automatic stations together with the higher variability of precipitation make the estimation of parameters, especially those of the variogram challenging. Nevertheless, early results of the CombiPrecip project showed that even at the hourly scale, combination methods yield reasonable results. The present study confirms this impression from a thorough inspection of many cases and it demonstrates that the combination methods improve the quality of radar systematically and are suitable tools for the production of hourly precipitation analyses in near real-time.

Visual examination of many precipitation fields in variable conditions showed that the more flexible formulations of KED and CED compared to OKRE are of advantage. They allow for both additive and multiplicative (at transformed scale) correction of radar errors, while structures related to multiplicative radar errors often clearly remain visible in the results from OKRE. Nevertheless, all three methods successfully corrected global and regional errors of the radar, and in many of the studied cases theses corrections were found to be substantial, e.g. a factor of 2 to 5 in cases of prominent summer convection. Nevertheless, cases are found where KED and CED had estimates of radar coefficients much smaller than one. Such cases were characterized by conditions where the correlation between radar and gauge data were very small. In these cases the estimated precipitation fields show less small-scale structure than in the radar and in the estimates of OKRE, because the trend field was very smooth. Altogether, the higher flexibility is nevertheless a valuable feature.

Our intercomparison has also revealed some general limitations of the combination methods: Firstly, some of the small-scale error patterns of radar are inherited by the combination. This is particularly the case for ground clutter and beam shielding structures. Secondly, the estimates of radar error can vary substantially from one hour to the next and this alters the general picture of temporal evolutions in the precipitation field. These limitations were found to be equally prominent in KED and CED, despite the latter uses information from the past. This limitation will require further attention before combination results are used for nowcasting.

Nevertheless, kriging at the hourly time scale has a price: since less wet stations are generally available compared to the daily case, efficient fallback solutions have to be implemented. The quantitative investigation of the combination methods revealed that the main differences between the combination methods are due to different fallback solution settings. We have therefore discovered

that the performance of a combination method does not solely depend on the method philosophy, but also critically relies on an appropriate choice of the default variogram parameters for the fallback solution and how frequently this solution is needed. Furthermore, the quantitative analysis of the four error scores revealed that KED and CED show an improvement over radar particularly in the bias as well as in REL.MRTE. The radar shows a good ability to distinguish between wet and dry and this positive feature is successfully inherited by the combination methods, especially by OKRE. Nevertheless, KED and CED seldom improve the HK beyond the radar performance. Especially in situations where small isolated precipitation systems occur, radar clearly outperforms the combination methods. Generally, we can say that radar is at the hourly time scale an equivalent choice to the other methods to distinguish wet from dry areas. In addition, the visual as well as the qualitative comparison of hourly precipitation fields revealed little differences between KED and CED. Even if KED is a less sophisticated technique than CED, we observed no situation where KED is inferior to CED, except for the bias. Even in this case, note that the bias is an error measure where errors can cancel each other and this feature is particularly evident for long time periods such an entire year and the whole region.

The results presented in this study rely on data and methods that underlie technical as well as conceptual changes and refinements. On the one hand MeteoSwiss has launched a project called Rad4Alp to expand and to improve the radar network. In near future the Swiss radar network will consist of five weather radars with more sophisticated instruments that will be able to distinguish better between rain, snow and hail. In addition the SwissMetNet including automated rain gauges is currently being expanded. These developments will be beneficial to the combination methods too. On the other hand the developed combination techniques undergo a continuous improvement processes e.g. adjustment of data transformation and optimizing of fallback solutions. Consequently, the results, presented in this study, refer to the technical and conceptual standard of August 2011. Further improvement of the combination methods and of the precipitation monitoring devices can result in another model performance.

Overall, the systematic evaluation and the cross-comparison presented in this thesis was an important step in identifying and understanding the strengths and limitations of the combination methods. Finally, this analysis serves also as a scientific reference helping MeteoSwiss for the choice of a future hourly precipitation monitoring system.

Literature

Diggle P.J, Ribeiro Jr. P.J. 2007. Model-based Geostatistics. *Springer, 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**: 108 pp.

Erdin R, Frei C, Künsch H.R. 2012. Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *Submitted*.

Gandin L.S. 1965. Objective analysis of meteorological fields. *Israel Program for Scientific Translation, Jerusalem*: 242 pp.

Germann U, Galli G, Boscacci M, Bolliger M. 2006. Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society* **132**: 1669-1692.

Goudenhoofdt E, Delobbe L. 2009: Evaluation of radar-gauge merging methods for quantitative precipitation estimates. *Hydrology and Earth System Sciences* **13**: 195-203

Haberlandt U. 2007. Geostatistical interpolation of hourly precipitation from rain-gauge and radar for a large-scale extreme rainfall event. *Journal of Hydrology* **332**: 144–157.

Joss J, Schädler B, Galli G, Cavalli R, Boscacci M, Held E, Bruna G, Kappenberger G, Nespor V, Spiess R. 1998. Operational Use of Radar for Precipitation Measurements in Switzerland. *Vdf Hochschulverlag AG ETH Zurich*: 134 pp.

Konzelmann T, Wehren B, Weingartner R. 2007. Niederschlagsmessnetze. *Hydrological Atlas of Switzerland*, HADES: 20. Available from University of Bern, Plate 2.1: Bern, Switzerland.

Matheron G. 1965. Les variables régionalisées et leur estimation.

Meischner P. 2004. Weather radar: Principles & advanced applications. *Springer, Berlin, Heidelberg*: 337 pp.

MeteoSwiss. 2010. SwissMetNet: Ein Messnetz für die Zukunft! *Bundesamt für Meteorologie und Klimatologie MeteoSchweiz*. 8 pp.

Neff E.L. 1977. How much rain does a rain gage gage? *Journal of Hydrology* **35**: 213-220.

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

Pebesma E. 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geoscience* **30**: 683-691.

R Development Core Team. 2011. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL: http://www.R-project.org.

Ribeiro Jr. P.J, Diggle P.J. 2001. geoR: A package for geostatistical analysis. *R-News* **1(2)**: 15-18.

Schabenberger O, Gotway C.A. 2005. Statistical methods for spatial data analysis. *Chapman & Hall, CRC*: 488 pp.

Seo D-J. 1998. Real-time estimates of rainfall fields using radar rainfall and rain gauge data. *Journal of Hydrology* **208**: 37–52.

Sevruk B. 1985. Systematischer Niederschlagsmessfehler in der Schweiz. *Der Niederschlag in der Schweiz. Bericht der Arbeitsgruppe "Niederschlag" der Hydrologischer Kommission der SNG **31**:* 65–75.

Shepard D. 1984. Computer mapping: The SYMAP interpolation algorithm. In *Spatial Statistics and Models.* D. Reidl Publishing Company, Dordrecht: 133–145.

Sideris I, Gabella M, Germann U. 2011. Real-time spatiotemporal combination of radar and rain gauge measurements in Switzerland. *9th Swiss Geoscience Meeting 2011, Zurich.*

Wackernagel H. 1995. Multivariate geostatistics; An introduction with applications *Springer-Verlag:* 387 pp.

Webster R, Oliver M.A. 2007. Geostatistics for Environmental Scientists, *John Wiley & Sons, Ltd:* 315 pp.

Wilks D. 2006. Statistical Methods in the Atmospheric Sciences. *Elsevier:* 627 pp.

Willi M, Gabella M, Germann U, Frei C. 2011. CombiPrecip - Test Cases: Documentation of selected precipitation events for evaluating radar-gauge combination methods obtained within CombiPrecip, subproject of NCCR Climate III at MeteoSwiss, *Internal Document of MeteoSwiss.*

Wüest M, Frei C, Altenhoff A, Hagen M, Litschi M, Schär C. 2010. A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *International Journal of Climatology* **30**: 1764-1775.

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

Appendix

Error scores

Error scores for wet (2008), Winter				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	-0.624	0.507	0.469	2.805
OKGAU	-0.298	0.428	0.407	2.297
Rdisagg	0.209	0.403	0.330	2.063
KED	-0.239	0.367	0.360	2.167
OKRE	0.377	0.416	0.328	2.222
CED	0.073	0.437	0.365	2.232

Error scores for wet (2008), Summer				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	0.431	0.427	0.306	2.990
OKGAU	-0.399	0.487	0.352	3.237
Rdisagg	0.093	0.256	0.241	2.169
KED	-0.179	0.272	0.244	2.453
OKRE	0.695	0.335	0.213	2.285
CED	0.028	0.316	0.248	2.471

Error scores for wet (2008), Swiss Plateau				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	0.605	0.338	0.241	2.522
OKGAU	0.182	0.435	0.286	2.596
Rdisagg	0.059	0.196	0.203	1.779
KED	0.348	0.267	0.186	1.941
OKRE	0.912	0.303	0.153	1.938
CED	0.674	0.354	0.190	1.945

Error scores for wet (2008), Alps				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	-0.628	0.532	0.491	3.164
OKGAU	-0.237	0.494	0.405	2.785
Rdisagg	0.130	0.416	0.356	2.560
KED	-0.136	0.369	0.350	2.538
OKRE	0.562	0.427	0.330	2.604
CED	0.139	0.426	0.353	2.555

Error scores for wet (2008)				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	-0.495	0.429	0.408	2.876
OKGAU	-0.328	0.438	0.364	2.726
Rdisagg	0.101	0.309	0.291	2.164
KED	-0.200	0.306	0.302	2.264
OKRE	0.417	0.345	0.277	2.239
CED	0.070	0.364	0.305	2.294

Error scores for wet (2008), all methods worked				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	-0.994	0.425	0.414	2.754
OKGAU	-0.165	0.437	0.360	2.574
Rdisagg	0.047	0.307	0.294	2.103
KED	-0.091	0.299	0.288	2.122
OKRE	0.230	0.315	0.287	2.131
CED	0.016	0.330	0.291	2.165

Error scores for wet (2008), KED uses climatological variogram				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	0.992	0.810	0.535	3.416
OKGAU	-0.957	0.770	0.704	3.347
Rdisagg	0.348	0.573	0.444	2.479
KED	-0.559	0.572	0.582	2.971
OKRE	1.094	0.701	0.467	2.831
CED	0.202	0.722	0.577	3.024

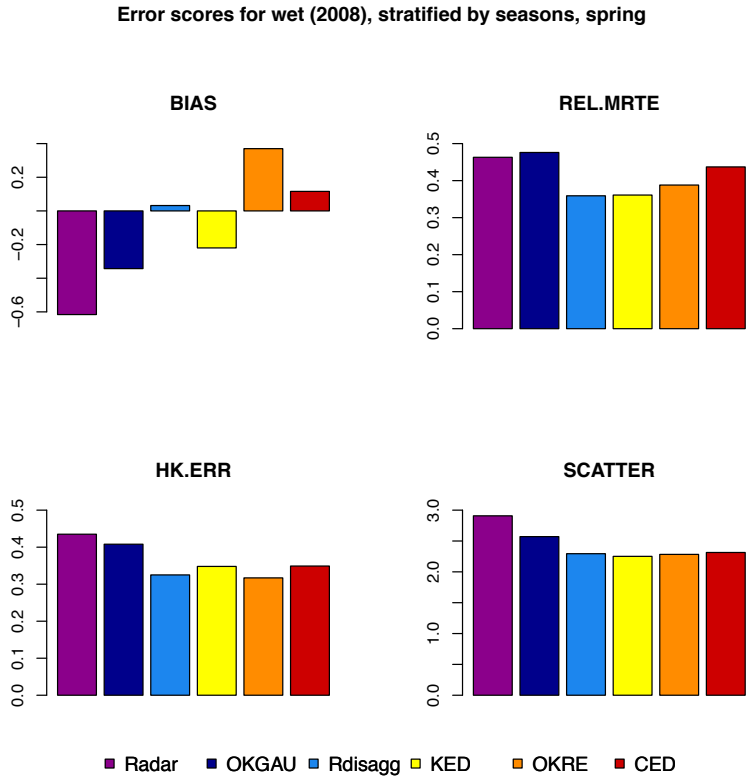
Error scores for wet (2008), KED uses radar				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	-1.480	0.606	0.533	2.903
OKGAU	-1.008	0.833	0.581	4.195
Rdisagg	-0.660	0.539	0.444	2.413
KED	-1.968	0.570	0.533	2.903
OKRE	-0.450	0.628	0.499	2.027
CED	0.968	0.964	0.486	3.086

Error scores for wet (2008), OKRE uses climatological variogram				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	0.941	0.809	0.538	3.422
OKGAU	-0.972	0.771	0.707	3.347
Rdisagg	0.320	0.579	0.447	2.483
KED	-0.618	0.575	0.586	2.971
OKRE	1.067	0.705	0.472	2.834
CED	0.250	0.741	0.581	3.026

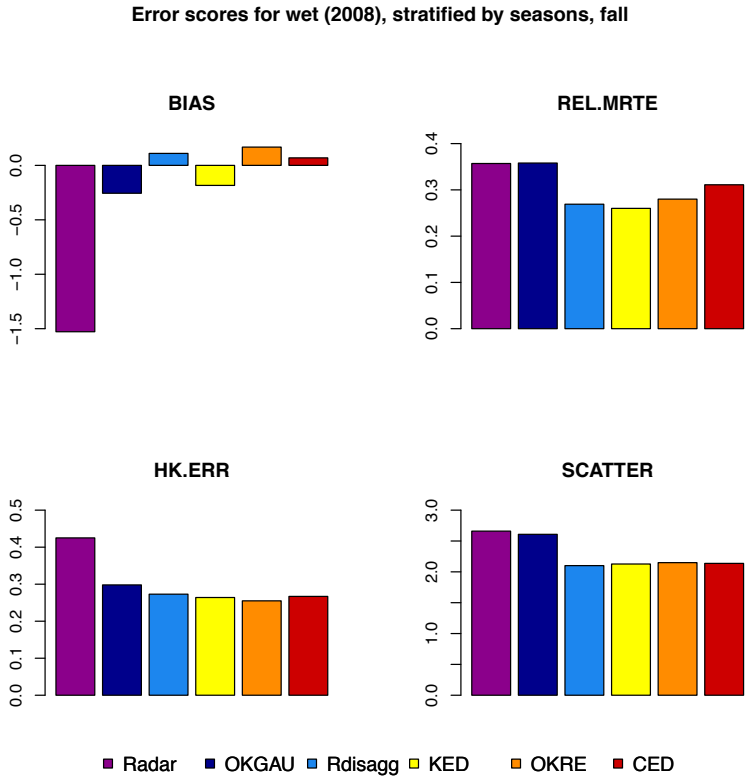
Error scores for wet (2008), CED uses default variogram				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	2.304	1.213	0.518	3.995
OKGAU	-1.234	0.859	0.636	3.252
Rdisagg	0.820	0.763	0.435	2.396
KED	-0.978	0.667	0.550	2.900
OKRE	2.197	1.056	0.407	3.224
CED	0.757	1.016	0.569	2.990

Error scores for daily precipitation sum (2008)				
	Bias [dB]	**REL.MRTE**	**HK.ERR**	**Scatter [dB]**
Radar	-0.296	0.292	0.286	2.808
OKGAU	-0.601	0.173	0.227	2.035
Rdisagg	-0.023	0.044	0.083	0.765
KED	-0.349	0.143	0.211	1.741
OKRE	0.260	0.177	0.243	1.763
CED	-0.026	0.207	0.291	1.819

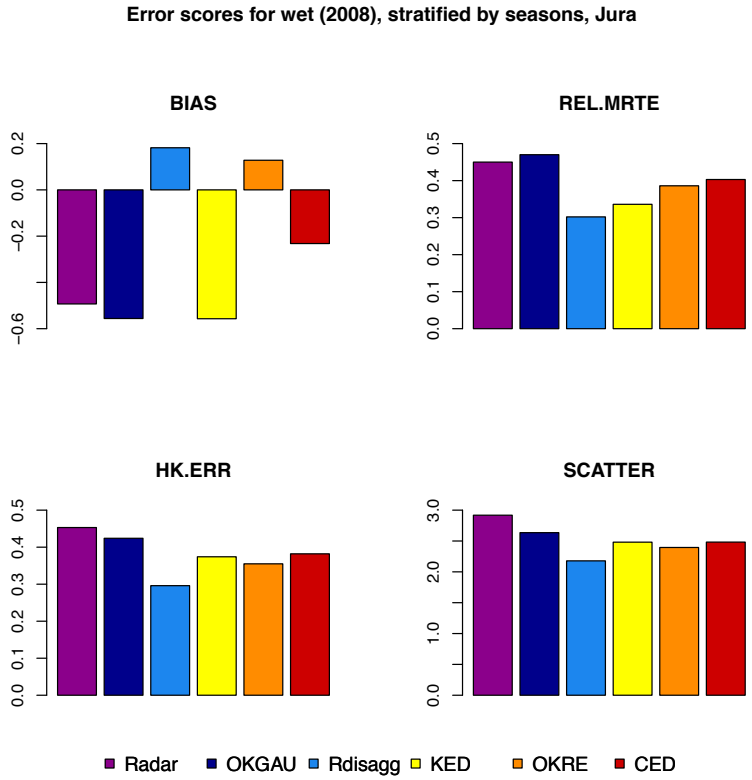
Error scores for wet (2008), Spring				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	-0.616	0.463	0.435	2.908
OKGAU	-0.343	0.476	0.408	2.571
Rdisagg	0.032	0.359	0.325	2.294
KED	-0.220	0.361	0.348	2.251
OKRE	0.370	0.388	0.317	2.283
CED	0.116	0.437	0.349	2.314



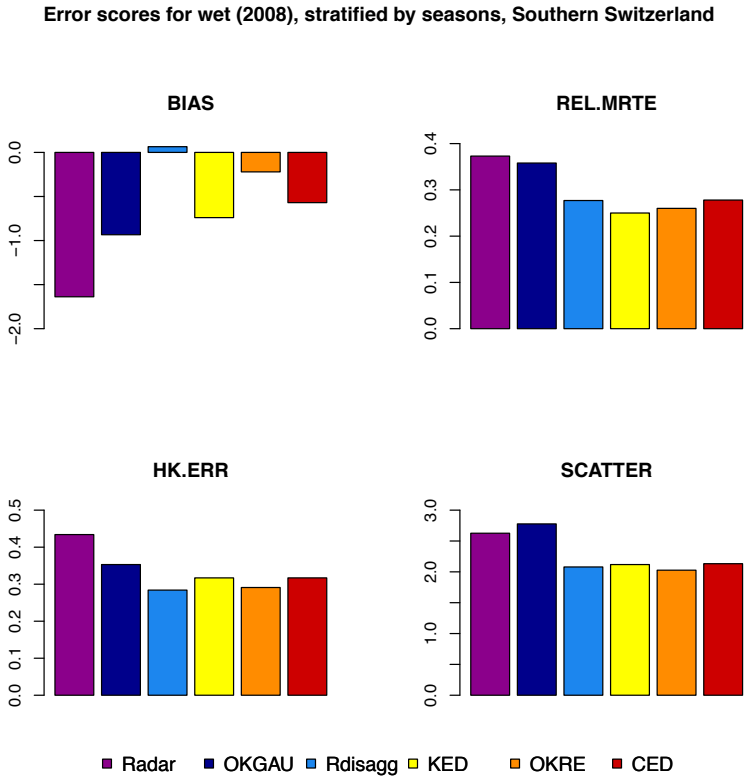
Error scores for wet (2008), Fall				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	-1.528	0.357	0.425	2.660
OKGAU	-0.256	0.358	0.298	2.608
Rdisagg	0.110	0.269	0.273	2.101
KED	-0.184	0.260	0.264	2.127
OKRE	0.168	0.280	0.255	2.149
CED	0.069	0.311	0.267	2.137



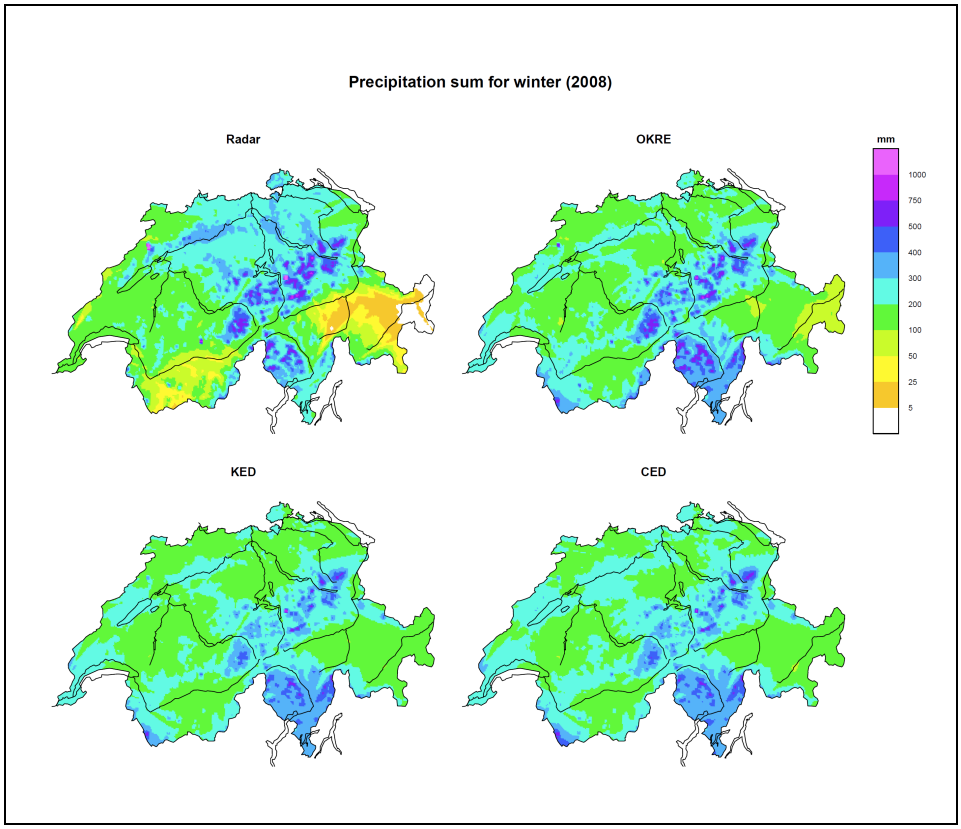
Error scores for wet (2008), Jura				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	-0.493	0.450	0.453	2.918
OKGAU	-0.556	0.470	0.424	2.635
Rdisagg	0.182	0.302	0.296	2.178
KED	-0.557	0.336	0.374	2.481
OKRE	0.128	0.386	0.355	2.395
CED	-0.232	0.403	0.382	2.482

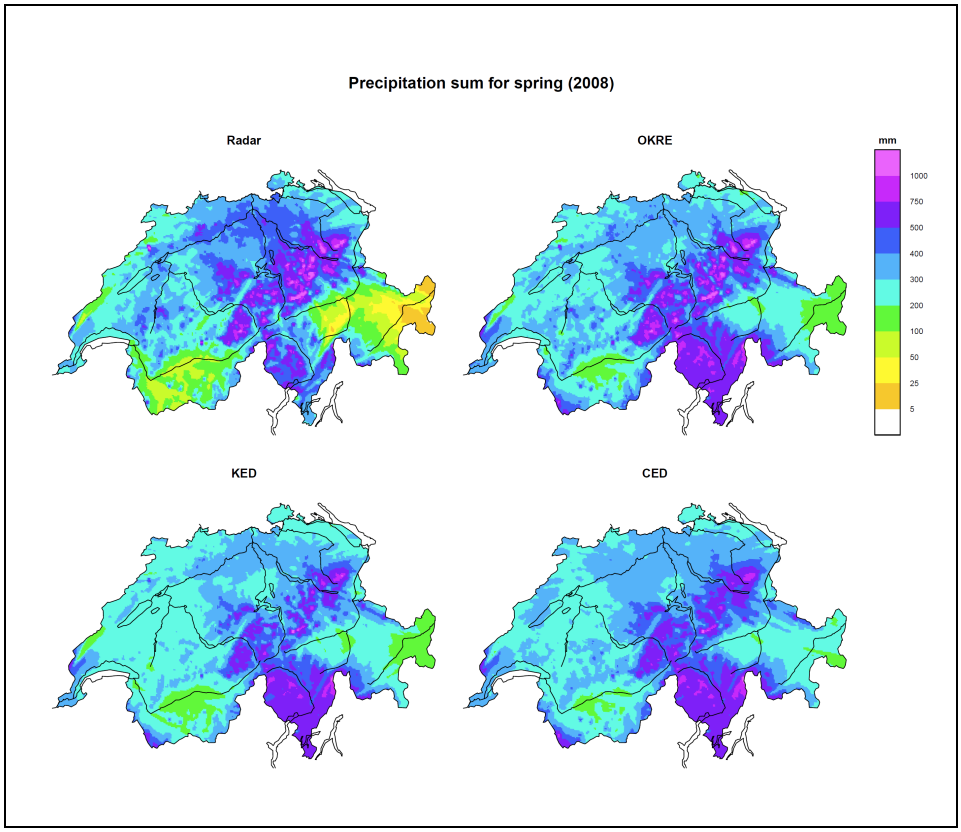


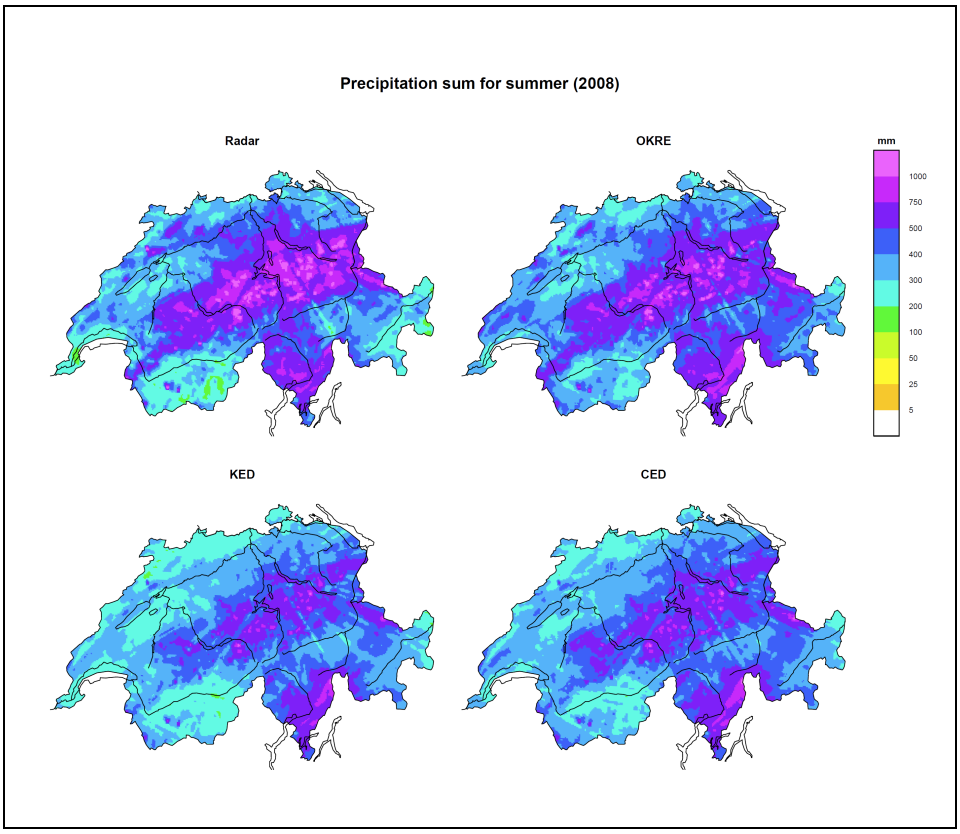
Error scores for wet (2008), Southern Switzerland				
	Bias [dB]	REL.MRTE	HK.ERR	Scatter [dB]
Radar	-1.638	0.373	0.434	2.626
OKGAU	-0.934	0.358	0.353	2.777
Rdisagg	0.065	0.277	0.284	2.079
KED	-0.740	0.250	0.317	2.118
OKRE	-0.221	0.260	0.291	2.027
CED	-0.570	0.278	0.317	2.132

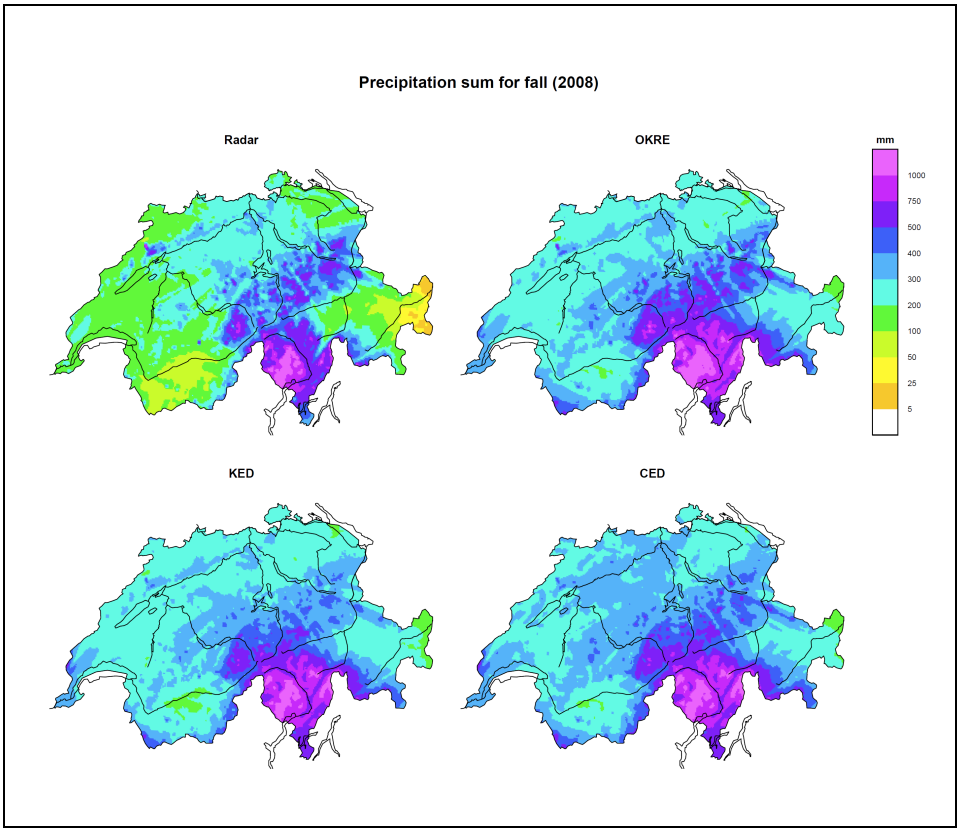


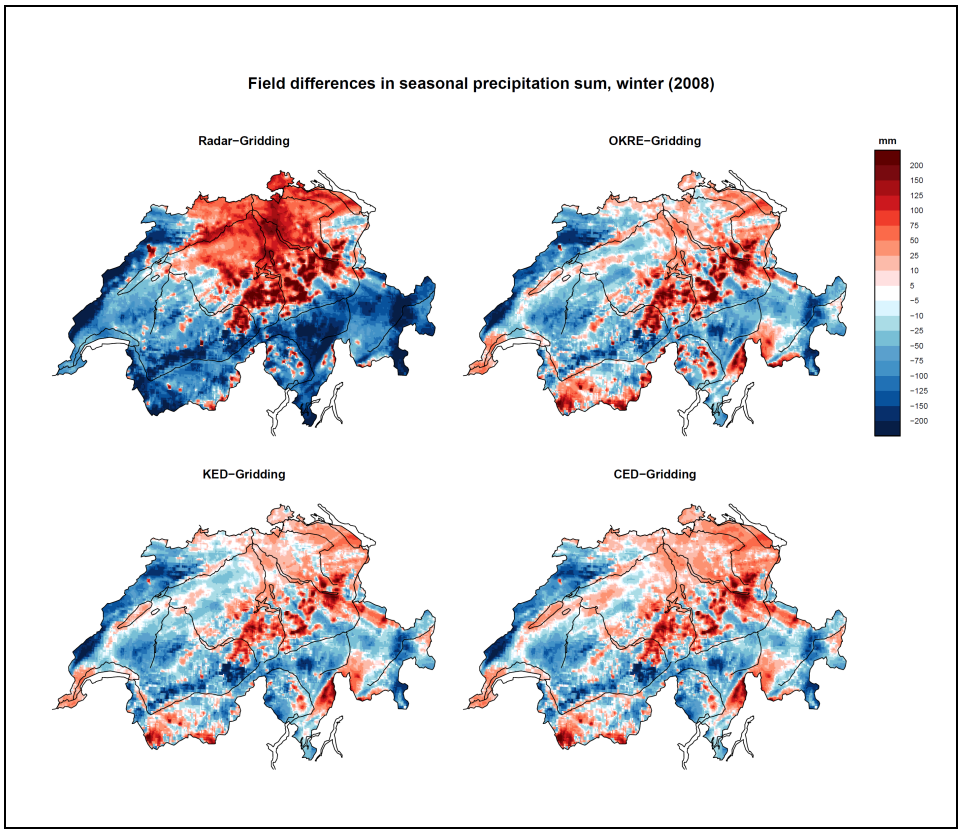
Seasonal Fields

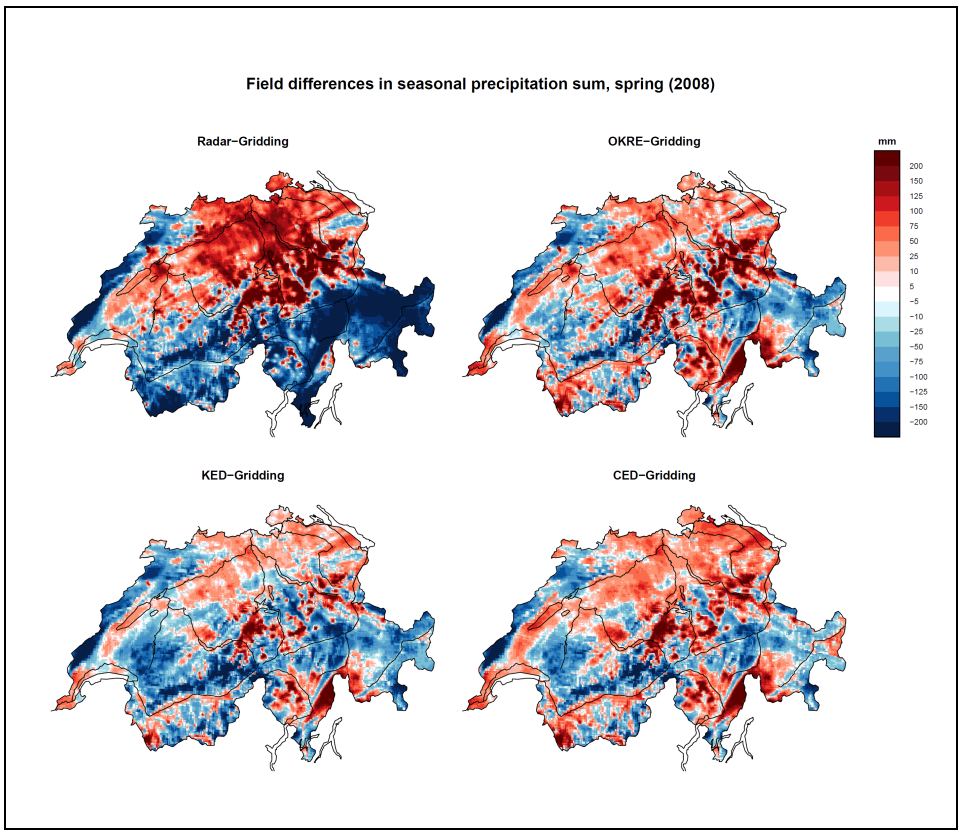


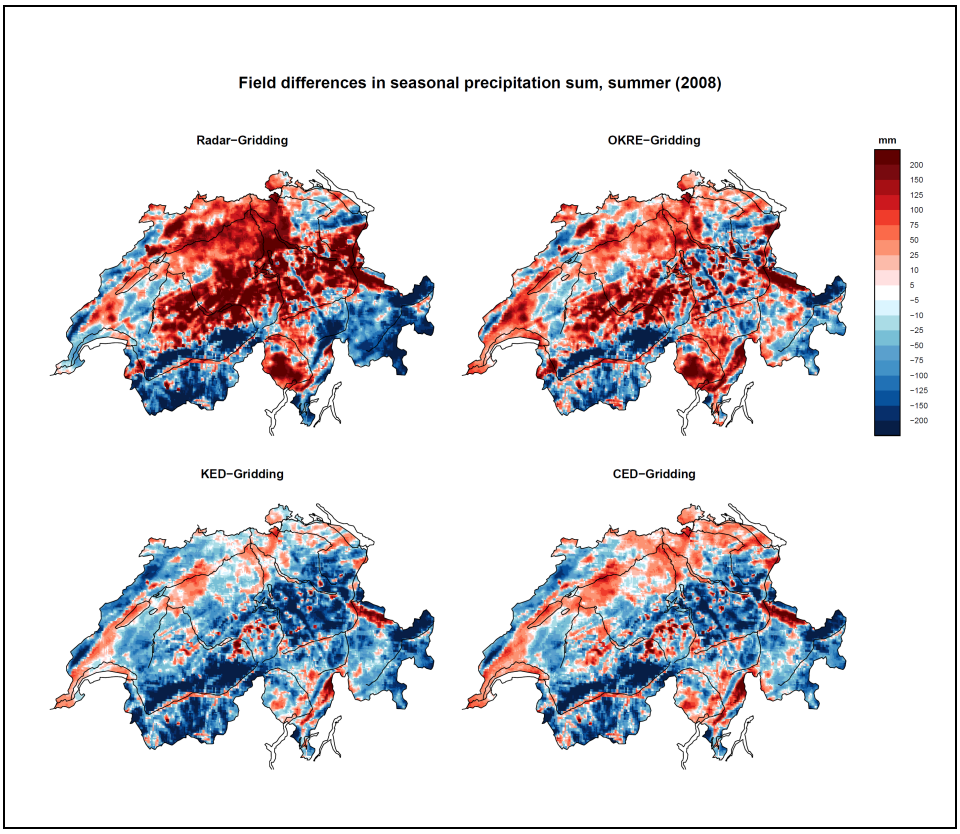


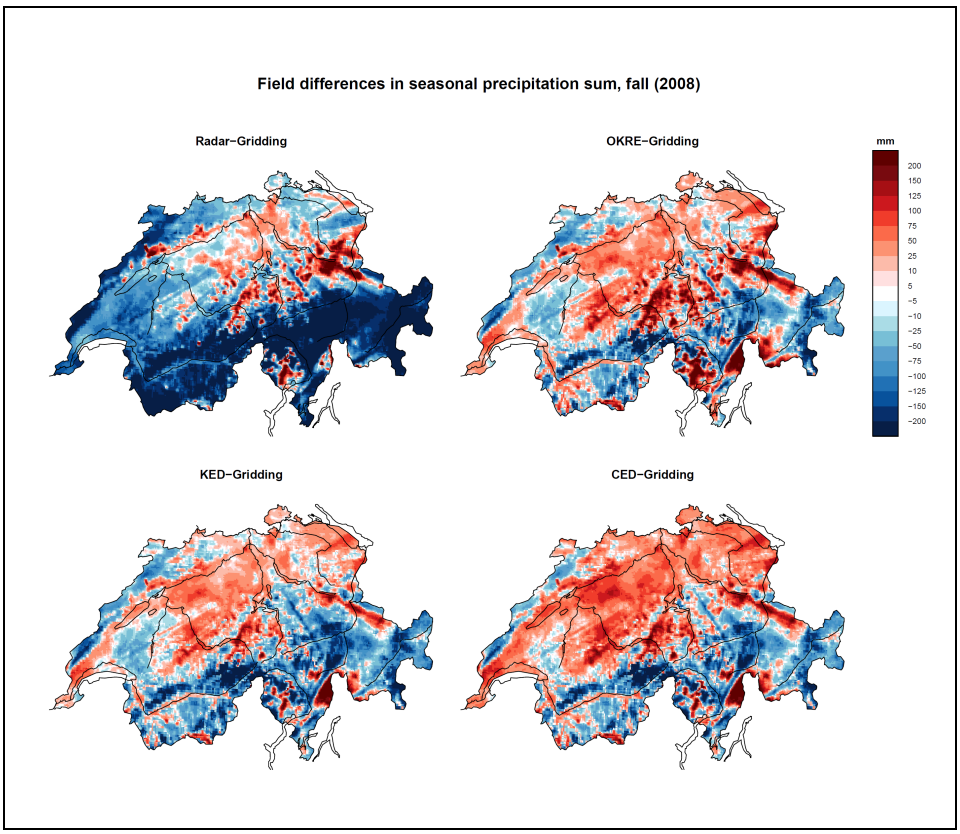

Precipitation sum for summer (2008)
Radar
OKRE
mm
1000
750
500
400
300
200
100
50
25
5
KED
CED


Precipitation sum for fall (2008)
Radar
OKRE
mm
1000
750
500
400
300
200
100
50
25
5
KED
CED


Field differences in seasonal precipitation sum, winter (2008)
Radar−Gridding
OKRE−Gridding
KED−Gridding
CED−Gridding
mm
200
150
125
100
75
50
25
10
5
−5
−10
−25
−50
−75
−100
−125
−150
−200


Field differences in seasonal precipitation sum, spring (2008)
Radar−Gridding
OKRE−Gridding
KED−Gridding
CED−Gridding
mm
200
150
125
100
75
50
25
10
5
−5
−10
−25
−50
−75
−100
−125
−150
−200


Field differences in seasonal precipitation sum, summer (2008)
Radar−Gridding
OKRE−Gridding
KED−Gridding
CED−Gridding
mm
200
150
125
100
75
50
25
10
5
−5
−10
−25
−50
−75
−100
−125
−150
−200


Field differences in seasonal precipitation sum, fall (2008)
Radar−Gridding
OKRE−Gridding
KED−Gridding
CED−Gridding
mm
200
150
125
100
75
50
25
10
5
−5
−10
−25
−50
−75
−100
−125
−150
−200

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