Quarterly Journal of the Royal Meteorological Society *Q. J. R. Meteorol. Soc.* 140: 1097–1111, April 2014 DOI:10.1002/qj.2188

RMetS
Royal Meteorological Society

Real-time radar–rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland

I. V. Sideris,* M. Gabella, R. Erdin and U. Germann
Meteoswiss, Locarno Monti, Switzerland

*Correspondence to: I. V. Sideris, Meteoswiss, Via ai monti 146, CH-6605 Locarno Monti, Switzerland.
E-mail: Ioannis.Sideris@meteoswiss.ch

The problem of the optimal combination of rain-gauge measurements and radar precipitation estimates has been investigated. A method that attempts to generalize well-established geostatistical techniques, such as kriging with external drift, is presented. The new method, besides allowing spatial information to be incorporated into the modelling and estimation process, also allows temporal information to be incorporated. This technique employs temporal data as secondary co-kriged variables. The approach can be considered both straightforward and practical as far as design and programming aspects are concerned. Co-kriging with external drift leads to significant improvements in the results compared with typical radar estimates. It also seems to be more advantageous than kriging with external drift in modelling stability terms. Evidence is provided showing the advantages of co-kriging with external drift modelling over kriging with external drift. The difference becomes particularly pronounced for non-robust input data. The theoretical background and mathematical structure of the method is demonstrated. The method has been applied to four events, three during summer and one during winter, that took place over the complex Swiss orography. It is shown that cross-validation skill scores improve when the aggregation period of the input data increases from ten minutes to one hour. This improvement can be attributed to the increasing robustness of the input data with the period of aggregation. Moreover, a straightforward disaggregation scheme, which starts from hourly precipitation maps, produced by means of the aforementioned geostatistical technique and generating precipitation estimates at a temporal resolution of five minutes, is proposed.

Key Words: kriging; geostatistics; meteorological radar; quantitative precipitation estimation; precipitation field variability

Received 23 July 2012; Revised 23 April 2013; Accepted 29 April 2013; Published online in Wiley Online Library 31 July 2013

1. Introduction

Rainfall is a dynamical process that involves complex spatio-temporal evolution and variability. It is usually treated as a stochastic process with certain characteristics, such as spatial and temporal correlations. These characteristics can be derived through the analysis of direct and/or indirect measurements of precipitation depths. Precipitation depth is the total amount of rainfall that reaches the ground over a certain area during a certain period of time and it is usually monitored by rain-gauges and radars. These two devices complement each other in a promising fashion: (a) rain-gauges provide accurate measurements, but their networks are sparse; (b) radars provide widespread and dense coverage, but there is a strong variability in the relationship between reflectivity and rainfall intensity; and (c) radar and rain-gauge rainfall data are, in general, fairly well correlated in space and time (Creutin *et al.*, 1987). The idea of combining rain-gauges with radar measurements in order to improve the accuracy of rainfall maps emerged naturally, since since such maps are important for a variety of hydrological and meteorological applications.

Several researchers have employed various deterministic or statistical methods for radar–rain-gauge merging (Brandes, 1975; Creutin *et al.*, 1987; Krajewski and Georgakakos, 1987; Seo *et al.*, 1990; Gabella *et al.*, 2000; Seo and Breidenbach, 2002; Gabella and Notarpietro, 2004; Haberlandt, 2007; Schuurmans *et al.*, 2007; Salek and Novak, 2008; Cole and Moore, 2009; Erdin, 2009; Velasco-Forero *et al.*, 2009), such as the following: (a) computation of a constant multiplicative calibration factor (deterministic); (b) deterministic interpolation of the gauge to radar ratio (deterministic); (c) statistical approaches based on multivariate analysis; (d) radar–rain-gauge probability distribution analysis; (e) geostatistical estimators; and (f) Bayesian methods. Numerous applications of these tools have appeared in the literature, all of which provide valuable contributions to common knowledge regarding the problem in question.

Moreover, a useful comparison between several interpolation methods, including geostatistical ones, can be found in Mitas and Mitasova (1999).

In this work, a spatio-temporal extension of geostatistical interpolation techniques is presented, which attempts to provide enhanced stability in terms of modelling, suitable mainly for operational applications. In particular, a scheme is proposed in which the precipitation information in hand is interpolated using four different kinds of input: (a) the rain-gauge measurements available for the time period of interest; (b) the radar estimates available for the time period of interest; (c) the rain-gauge measurements available for the time period preceding the current one; and (d) the radar estimates available for the time period preceding the current one. Rain-gauge and radar information from previous time states introduces temporal characteristics into the scheme.

Although rain-gauge measurements are assumed to be fairly accurate, they are seldom free of error (Sevruk and Lapin, 1993). For example, in strong winds, undercatch by rain-gauges may exceed 20% (Nespor and Sevruk, 1999). Moreover, tipping bucket gauges have been known to underestimate rainfall for intense events, compared with weighing bucket gauges. Rain-gauges provide direct point measurements of rainfall depth that has accumulated continuously over a time period of interest. Precipitation fields are characterized by high temporal and spatial variability. The point character of rain-gauge measurements and the sparcity of operational rain-gauge networks, e.g. due to the high cost of telemetred data collections (Creutin *et al.*, 1987), make these measurements inadequate to provide sufficient information on the variability of precipitation fields. A carefully designed network may help by improving the computation of statistical quantities such as the covariances of measurements (Moore *et al.*, 2000). However, the high altitudes that characterize regions like the Swiss alpine terrain introduce additional complications as far as installing rain-gauges at all the desirable locations is concerned.

The rain-gauge representativeness error is especially important in the context of combining radar and rain-gauge data, since the resolution of the two devices is spatially different by several orders of magnitude. Many studies have investigated the issue. Zawadzki (1975) investigated the subject from a theoretical point of view and showed that the smoothing effect originated through convolution between the antenna radiation pattern and the radar reflectivity field. In more practical terms, Kitchen and Blackall (1992) showed that at an hourly scale in England the root-mean-square error (RMSE) due to the spatio-temporal differences in sampling mode is as large as 150% of the mean rainfall. Ciach and Krajewski (1999) introduced the error-variance separation method in order to assess the uncertainty associated with the area–point resolution difference (an approach that was also used and discussed in several other articles, e.g. Villarini *et al.*, 2008; Seo and Krajewski, 2011). They estimated that, at hourly scales, the assumption of true areal rainfall estimation using rain-gauges accounts for approximately half of the overall discrepancies between radar and rain-gauges. In Florida, Habib and Krajewski (2002) estimated that the spatial sampling error accounts for approximately 40–80% of the overall radar–gauge disagreement. However, the assessment of this uncertainty typically requires the employment of dense rain-gauge networks, including some rain-gauges located very close to each other. This is motivated by the need to have some reliable estimates of the 'true rainfall' taking place in very small-scale regions. In the same context, Gires *et al.* (2011) employed downscaled realizations to quantify rainfall variability and concluded that small-scale rainfall variability should be taken into account in sewer network management or to improve rainfall data resolution by implementing high-resolution radars such as *X*-band ones.

Radars provide averages of backscattered energy from precipitation particles in elevated volumes. Radar rainfall estimates are indirect but are also spatially dense and provide continuous monitoring of precipitation. The main drawback is that large observation errors may occur (Zawadzki, 1984; Villarini and Krajewski, 2010), which can be attributed to numerous meteorological, topographic and technical factors, such as the following: (a) variations of the raindrop size distribution both within and between the clouds; (b) vertical air movements; (c) incomplete and/or non-uniform beam filling (Germann and Joss, 2004); (d) attenuation and beam blockage (Creutin *et al.*, 1987); (e) beam broadening with distance and range degradation (Gabella *et al.*, 2011); (f) vertical change of the radar echo (the profile) caused by the growth and transformation of the precipitation (Germann and Joss, 2002); and (g) miscalibration of the electronic components (Seo *et al.*, 1990). Moreover, radar measurements of averaged backscattered power introduce a space-smoothing effect over a cell/pixel. This is due to the convolution between the antenna radiation pattern and the reflectivity field. They also introduce a variable time lag with respect to the overall precipitation that reaches the ground, since they represent a volume at some height over the ground (Zawadzki, 1975; Jordan *et al.*, 2000).

Because of the aforementioned problems, discrepancies arise between radar and rain-gauges. These become more prominent for smaller precipitation accumulation time (Villarini *et al.*, 2008). It is a common practice to employ aggregations over at least one hour in order to mitigate the effect (Zawadzki, 1975).

In this article, an attempt has been made to produce rain-gauge-adjusted estimates of precipitation on a lattice with spatial resolution equal to that of the radar rainfall estimation map of Switzerland ($1\,km^2$). The proposed modelling employs rain-gauge measurements and the corresponding radar estimates at rain-gauge locations. An inherent, although not necessarily sound, assumption is made that each rain-gauge measurement and its corresponding radar estimate (at the rain-gauge location) refer to the same precipitation average over an area of $1\,km^2$. It is assumed that the rain-gauge precipitation value is close to the mean value of that pixel. This is not necessarily true, especially when the microscale variability is of a significant size, which might occur during strong convection. Moreover, it is also assumed that radar estimate errors are spatially correlated, which is a rather reasonable assumption (Zawadzki, 1984; Joss and Lee, 1995; Berenguer and Zawadski, 2008a; Berenguer and Zawadzki, 2008b; Germann *et al.*, 2009). Therefore, adjacent pixels are expected to be characterized by similar errors. It could therefore be expected that, although the radar estimates are erroneous, the trends and the outlines they imply should be indicative of the actual precipitation process.

The basic characteristics of the precipitation monitoring devices employed are described in section 2. The mathematical formulation of the geostatistical scheme is presented in section 3. The main results are presented and discussed in section 4.

2. Description of available data

The aim of the research activity was to derive automatic, real-time, spatio-temporally high-resolution estimates of the precipitation field over Switzerland, which are more accurate than radar-only rainfall estimates. Two types of information have been used (the account refers to the 2005–2009 period).

(1) Rain-gauge precipitation measurements. There are 445 rain-gauges in Switzerland, 75 of which are automatic, tipping-bucket and provide measurements every 10 minutes. 370 are manual and provide daily measurements. The number of automatic rain-gauges has been continuously increasing since the period under investigation. The rain-gauge measurements are subject to certain errors but they are considered fairly reliable. The disadvantage of these measurements is that in a statistical context they are point measurements, each of which represents a spatially very limited area. Intuitively, their representativeness should

depend on the involved meteorology: although in stratiform precipitation conditions such measurements may be well-representative, when convective cells emerge in the weather pattern then rain-gauge measurements may be misleading due to significant microvariability.

Rain-gauge measurements, as well as other meteorological variables, are subject to extensive quality control (by MeteoSwiss) before they are used as an input of the combination application. This involves the following tests (Grüter *et al.*, 2003): (i) limit tests (comparison with physical and climatological limits); (ii) variability tests (tests of maximum and minimum acceptable variability during a specified time interval); (iii) interparameter consistency tests (values of variables measured at the same time and place cannot be inconsistent with each other); and (iv) spatial consistency tests (between nearby stations).

Since the aim of the research was to produce precipitation estimates in real time, only the 75 automatic rain-gauges were used, although the measurements from the more numerous group of manual rain-gauges could be used in post-processing validation schemes. The automatic rain-gauge network has been designed to cover the entire Swiss region almost uniformly. The typical distance between two rain-gauges is about 24 km. Measurable, non-zero rainfall is often only recorded in a small subset of these automatic rain-gauges, a fact that makes modelling challenging. The accuracy of correlation structures such as variograms may be seriously impaired when the set of non-zero precipitation measurements is very limited and/or the measurements are characterized by serious errors.

(2) Radar backscatter measurements: Switzerland has been equipped with three Doppler radars since 1994. Several procedures are followed to assure the high quality of radar data: automatic hardware calibration and monitoring, clutter removal, correction for radar visibility, vertical profile of reflectivity (VPR) correction and others. In such a complex orography region, real-time precipitation measurement, by means of radar, requires appropriate system design and tailored solutions. The elimination of ground echoes and the correction of errors caused by shielding of the radar beam by mountain ranges are particularly difficult. The article by Germann *et al.* (2006), which appeared in this journal, presents the Swiss solution. Radar precipitation estimates are generated every 5 minutes and a combination of the three radar images is used to produce a final composite raster, the size of which is 620×620 pixels (each pixel representing a spatial range of $1\,\text{km}^2$). The spatial coverage of the radars is extensive but occasionally subject to significant errors. Radar estimates are produced through a much more complex process than rain-gauge measurements and cannot be as reliable. However, the information they produce regarding the gradients and outline of the precipitation field is of utmost importance and their widespread instantaneous coverage and representativeness renders them indispensable.

Here, we have attempted to combine rain-gauge and radar measurements to produce a precipitation map of Switzerland with a resolution equal to the resolution of the radar raster ($1\,\text{km}^2$ per pixel). Precipitation maps will be presented for a box of size 360×230 pixels, which covers a geographic area of $82\,800\,\text{km}^2$. This includes the entire Swiss region ($41.285\,\text{km}^2$) as well as some areas belonging to neighbouring countries.

3. Mathematical formulation

3.1. Co-kriging with external drift

Precipitation intensity is characterized by temporal continuity. This can be detected easily visually, but also from correlations between precipitation fields observed in successive time states of sufficiently short time-lengths. These correlations can be incorporated into a geostatistical model in several ways. The most comprehensive approach is to construct a full spatio-temporal geostatistical model. Such modelling is characterized by a covariance function that involves both the spatial and temporal dimensions and has been described in several investigations over the last decade (Cressie and Huang, 1999; Cesare *et al.*, 2001; De Iaco *et al.*, 2001; Gneiting, 2002). Such complex spatio-temporal models were unnecessary for present purposes. Moreover, it would be difficult to automate these models for real-time applications. It was more practical to introduce temporal information into a typical kriging with external drift (KED) scheme through co-kriging. The mathematical formulation is described hereafter.

The aim was to estimate the precipitation depth p_0 at a certain spatial location (x_0, y_0) over the aggregation period T_a. In order to model precipitation as a geostatistical process, the following four vectors need to be employed.

(1) The rain-gauge measurement vector(as the primary variable):

$$\mathbf{g}(T_\text{a}) = (g(\mathbf{s}_1, T_\text{a}), g(\mathbf{s}_2, T_\text{a}), \ldots g(\mathbf{s}_N, T_\text{a})), \qquad (1)$$

where N is the number of available rain-gauge measurements over the given aggregation period T_a and $\mathbf{s}_i = (x_i, y_i)$ is the spatial-coordinate vector of a rain-gauge with index i.

(2) The radar precipitation estimates at rain-gauge locations over the aggregation period T_a:

$$\mathbf{r}(T_\text{a}) = (r(\mathbf{s}_1, T_\text{a}), r(\mathbf{s}_2, T_\text{a}), \ldots r(\mathbf{s}_N, T_\text{a})). \qquad (2)$$

(3) The rain-gauge measurements over the aggregation period T_b, preceding aggregation period T_a (for example, if the aggregation period T_a of interest is the interval (9:00,10:00] then the aggregation period T_b represents the interval (8:00,9:00]):

$$\mathbf{g}(T_\text{b}) = (g(\mathbf{s}_1, T_\text{b}), g(\mathbf{s}_2, T_\text{b}), \ldots g(\mathbf{s}_M, T_\text{b})). \qquad (3)$$

The length, M, of this vector is usually (but not necessarily) equal to N.

(4) The radar precipitation estimates over the aggregation period T_b:

$$\mathbf{r}(T_\text{b}) = (r(\mathbf{s}_1, T_\text{b}), r(\mathbf{s}_2, T_\text{b}), \ldots r(\mathbf{s}_M, T_\text{b})). \qquad (4)$$

Moreover, during the estimation process, the map of the radar estimates over the aggregation period T_a is used.

Two inherent assumptions are made here. Firstly, the fact that the radar and rain-gauge data are characterized by spatial resolutions of different size has been ignored; it is assumed that both describe a common physical quantity, the precipitation depth at spatial blocks of a size equal to the spatial resolution of the radar ($1\,\text{km}^2$). Therefore, each rain-gauge measurement is considered to represent a precipitation average over a much larger area than the actual physical size monitored by the rain-gauge itself. Secondly, it is assumed that the precipitation estimates made by the radar at a certain altitude above ground at each spatial block of size $1\,\text{km}^2$ reflect the actual precipitation depth of this same block on the ground. Therefore, there is no in or out 'leakage' of precipitation from or to neighbouring blocks.

It is possible to combine all the aforementioned information into a single geostatistical scheme using co-kriging with external drift (CED). More details about the procedure are given hereafter. According to the general geostatistical description (see, for instance, Kyriakidis and Journel, 1999), a random process $Z(\mathbf{s}, t)$

can be modelled as a decomposition into two parts, a deterministic mean function $m_Z(\mathbf{s}, t)$ that reflects average or trend component variability and a stochastic stationary residual component $R(\mathbf{s}, t)$ with a zero mean, which models higher frequency fluctuations about that trend, so that

$$Z(\mathbf{s}, t) = m_Z(\mathbf{s}, t) + R(\mathbf{s}, t) \quad \forall\, \mathbf{s} \in D \subseteq \mathbb{R}^2;\ t \in T \subseteq \mathbb{R}, \quad (5)$$

where s represents the spatial coordinates and t represents time, while $\mathbb{R}$ is the set of real numbers. The sets D and T that index the random function Z are its parameter sets and they can be discrete or continuous. Non-stationary modelling is implemented by making the mean dependent on space, time or both, which is equivalent to introducing a trend into the model.

As far as the present problem is concerned, the precipitation depth in question (time period T_a) is decomposed as

$$p(\mathbf{s}, T_a) = m_p(\mathbf{s}, T_a) + R(\mathbf{s}, T_a), \quad (6)$$

while the precipitation depth from time period T_b is accordingly decomposed as

$$p(\mathbf{s}, T_b) = m_p(\mathbf{s}, T_b) + R(\mathbf{s}, T_b). \quad (7)$$

When external drift is employed into a geostatistical scheme, the trend $m_Z(\mathbf{s}, t)$ (Eq. (5)) is modelled as a linear function of a smoothly varying external variable $y(\mathbf{s}, t)$ (Goovaerts, 1997), in our case the radar estimates:

$$m_Z(\mathbf{s}, t) = a_0(\mathbf{s}, t) + a_1(\mathbf{s}, t) y(\mathbf{s}, t). \quad (8)$$

This trend is not estimated through a regression process prior to the kriging but instead is inherent in the kriging estimation process.

The autocorrelations of and cross-correlation between the stochastic parts $R(\mathbf{s}, T_a)$ and $R(\mathbf{s}, T_b)$ are described by their covariance matrices $\mathbf{C}_{aa}$, $\mathbf{C}_{bb}$ and $\mathbf{C}_{ab}$, respectively. Geostatistics considers the problem of minimization of the error between the predicted and true values by imposing that the expected value of this error is equal to zero (best linear unbiased prediction or BLUP: Goovaerts, 1997). This consideration leads to the following system of equations (Myers, 1982; Stein *et al.*, 1991; Ver Hoef and Cressie, 1993):

$$\begin{pmatrix} \mathbf{C}_{aa} & \mathbf{C}_{ab} & \mathbf{F}_a & \mathbf{O} \\ \mathbf{C}_{ba} & \mathbf{C}_{bb} & \mathbf{O} & \mathbf{F}_b \\ \mathbf{F}'_a & \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}'_b & \mathbf{O} & \mathbf{O} \end{pmatrix} \begin{pmatrix} \lambda_a \\ \lambda_b \\ \mu_a \\ \mu_b \end{pmatrix} = \begin{pmatrix} \mathbf{c}_a \\ \mathbf{c}_b \\ \mathbf{f}'_a \\ \mathbf{0} \end{pmatrix}, \quad (9)$$

where $\mathbf{C}_{aa}$ is the $N \times N$ covariance matrix between the residuals at the monitored locations over the aggregation period T_a; $\mathbf{C}_{bb}$ is the $M \times M$ covariance matrix between the residuals at the monitored locations over the aggregation period T_b; $\mathbf{C}_{ab}$ is the $N \times M$ cross-covariance matrix between the residuals at the monitored locations over the aggregation period T_a to the residuals at the monitored locations over the aggregation period T_b; $\mathbf{F}_a$ and $\mathbf{F}_b$ are the trend matrices over the aggregation periods T_a and T_b, with $\mathbf{F}'_a = [1, \mathbf{Y}_a]$ and $\mathbf{F}'_b = [1, \mathbf{Y}_b]$ ($\mathbf{Y}_a$ and $\mathbf{Y}_b$ are the radar values at the rain-gauge locations) in accordance with Eq. (8); $\mathbf{c}_a$ and $\mathbf{c}_b$ are the covariance matrices between the unmonitored location and the monitored locations over the aggregation periods T_a and T_b; $\mathbf{f}_a$ is the trend at that location over the aggregation periods T_a with $\mathbf{f}'_a = [1, \mathbf{y}_a]$ ($\mathbf{y}_a$ is the radar value at the estimation location); λ_a and λ_b are the weights of the kriging predictor; μ_a and μ_b are the Lagrange multipliers vectors associated with the minimization process.

The solution of this system provides the weights of the linear kriging predictor and in this manner the expected value of the precipitation depth can be computed at the unmonitored location. The estimator equation is given as

$$\hat{p}(\mathbf{s}, T_a) = \sum_{i=1}^{N} \lambda_{ai} g(\mathbf{s}_i, T_a) + \sum_{j=1}^{M} \lambda_{bj} g(\mathbf{s}_j, T_b). \quad (10)$$

For comparison purposes, the typical KED system is given by

$$\begin{pmatrix} \mathbf{C}_{aa} & \mathbf{F}_a \\ \mathbf{F}'_a & \mathbf{O} \end{pmatrix} \begin{pmatrix} \lambda_a \\ \mu_a \end{pmatrix} = \begin{pmatrix} \mathbf{c}_a \\ \mathbf{f}'_a \end{pmatrix}, \quad (11)$$

and its estimator equation is

$$\hat{p}(\mathbf{s}, T_a) = \sum_{i=1}^{N} \lambda_{ai} g(\mathbf{s}_i, T_a). \quad (12)$$

It should be obvious that CED is a generalization of KED in which the additional co-kriged variables introduce a spatio-temporal coupling into the solver.

3.2. Variogram analysis

Three sample variograms, two direct and one cross-variogram, were computed for each aggregation period of each of the events studied. The two direct variograms characterize the autocorrelations of the residuals corresponding to the aggregation periods T_a and T_b. The cross-variogram refers to the cross-correlations between the two sets of residuals. The theoretical variograms are fitted employing the package 'gstat' (Pebesma, 2012b), coded in R (R Development Core Team, 2012). The general mathematical formula that defines the variogram is given by Oliver (2010) and Pebesma (2012a):

$$\gamma_{kl}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z_k(s_i) - Z_k(s_i + h))(Z_l(s_i) - Z_l(s_i + h)), \quad (13)$$

where k and l signify different variables. For a single direct variogram, this formula simplifies to

$$\gamma_k(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z_k(s_i) - Z_k(s_i + h))^2. \quad (14)$$

In general, a relation between variogram and covariance exists for stationary random functions with finite variances under a symmetry condition satisfied when $C_{kl}(h) = C_{kl}(-h)$ (Myers, 1982; Künsch *et al.*, 1997) and is expressed as

$$\gamma_{kl}(h) = C_{kl}(0) - C_{kl}(h). \quad (15)$$

The sample variograms were fitted by means of the exponential model:

$$\gamma(h) = K_0[1 - \delta(h)] + K(1 - \mathrm{e}^{-h/r}) \qquad h \geq 0, \quad (16)$$

where K_0 represents the nugget variance, while the Kronecker delta $\delta(h)$ is 1 for $h = 0$ and 0 for $h > 0$. K is the partial sill and r is the variogram range.

Variogram modelling is central to any kriging estimation process. A spatial analyst usually monitors the fitting process and may introduce their input manually. However, the proposed algorithm aims at being used in real-time automatic applications where human intervention is not a possible option. Scenarios can be encountered in which the input data are limited and/or subject to significant errors. In such cases, the sample variogram may be ill-defined and consequently the automatic fitting of

the theoretical variogram may end up being problematic (for example it may be singular, have very high nugget, etc.). These scenarios are not the rule but the exception. However, because of the criticality of the issue and the need to find a solution with an operational character, an addition was introduced to the typical coding in order to produce a variogram fitting that would lead to acceptable results even in such cases.

A straightforward process has been established to remedy the situation, which involves computing the underlying sample variograms for many different sets of values of the involved parameters. The parameters that describe the sample variograms are (Pebesma, 2004) (a) the cut-off distance (the spatial separation distance up to which point pairs are included in the variogram estimates), (b) the bin size (the width of the subsequent distance intervals into which data-point pairs are grouped for variogram estimates) and (c) the parameter dX, which allows pooled estimation to be conducted (that is, it only allows variogram pairs of data points $Z(s_1), Z(s_2)$ to be included in the calculation when $||x(s_1) - x(s_2)|| < \mathrm{d}X$ with $x(s_i)$) are the values of regressors (radar values) at that location). The proposed code attempts several combinations of these three parameters until the fitting manages to provide an estimated variogram with a realistically low nugget as well as a nugget-to-sill ratio. This scheme usually operates smoothly without this addition, i.e. using the default set of parameter values. However there are some cases in which the data are such that the fitted variogram becomes problematic. It is in these cases that the aforementioned technique proves useful, as it manages to provide stabler alternatives than the default results. In the analysis that follows, graphs will be used to demonstrate the benefits of this method.

Several authors have attempted to introduce the moving windows technique in geostatistical schemes by calculating several local variograms (e.g. Haas, 1990; Harris *et al.*, 2010). That would not be especially practical for an algorithm that aims to be operational and for a process, such as precipitation, that is characterized by strong intermittence. This is because, taking into consideration the already small spatial size of Switzerland and the small total number of available rain-gauges, the number of wet rain-gauges within each window would be rather small, a fact that may often lead to ill-defined, unintelligible, variograms.

Nonparametric variogram fitting techniques (Velasco-Forero *et al.*, 2009) suggest a different philosophy regarding variogram modelling. Although these techniques are useful in general, it would be non-trivial to introduce non-parametric modelling into the proposed spatio-temporal scheme, which requires the simultaneous fitting of three variograms. Moreover, the technique described here may be considered more transparent and practical regarding identification of problems. For example, issues with variogram fitting can be directly visualized; moreover it may be easier to identify points in the variogram that are clearly outliers and have to be excluded. Other methods such as multiquadric estimators (e.g. Cole and Moore, 2009) circumvent variogram estimation, but Borga and Vizzaccaro (1997) suggested that they are comparable to kriging for network densities greater than $0.022\,\mathrm{gauge\,km^{-2}}$, while at lower densities (as in this work) kriging is more accurate.

Investigations are presented in the results section using five different modelling and estimation processes: (a) original radar raster data (no rain-gauge data); (b) ordinary kriging (ORK) of only rain-gauge data (no radar data); (c) rain-gauge–radar combination by means of KED, using default parameter choices for the fitting of the direct variogram; (d) rain-gauge–radar combination by CED using default parameter choices for the fitting of the three variograms (two direct and one cross-variogram); and (e) rain-gauge–radar combination by means of co-kriging with external drift using the aforementioned extensive capabilities of variogram fitting, referred to as 'Co-kriging with External Drift with Extensive capabilities' (CED.e), i.e. searching for the optimal variogram bin size, cut-off distance and pooling threshold to produce the most intelligible variogram possible.

3.3. Transformation

Both the radar estimates and rain-gauge measurements have been transformed using the square-root transformation, i.e. the geostatistical process was applied to the square roots of the rain-gauge and radar scores. It has been shown that this transformation provides data that approach the Gaussian distribution (Schuurmans *et al.*, 2007; Erdin, 2012) and it can therefore be used to produce exceedance probabilities from kriging predictions (Goovaerts, 1997). The square-root transformation offers the advantage of having an analytical back-transformation formula for the mean and the variance derived from the moment-generating function of the Gaussian distribution (Gregoire *et al.*, 2008; Wackerly *et al.*, 2008), which is given by

$$E(Y^2) = \mu^2 + \sigma^2, \tag{17}$$

and,

$$\mathrm{Var}(Y^2) = 4\mu^2\sigma^2 + 2\sigma^4, \tag{18}$$

where μ is the mean and σ is the standard deviation of the square-root-transformed kriging prediction at a certain location. This mean value is actually the kriging prediction at the location in question and, to the extent that the transformed data are Gaussian, the standard deviation represents the standard deviation of the Gaussian distribution at that location. The back-transformation of the quantiles takes place directly through the squaring of the corresponding transformed ones. The square of the mean transformed value corresponds to the median of the untransformed distribution. The analytical back-transformation is an important point for operational, real-time applications in which the processing time plays a role. Although other more complex transformations can offer certain advantages over the square-root transformation, they may depend on human decisions regarding the proper choice of the involved parameters and/or they may require a significant back-transformation processing time when an estimate is needed for a large number of locations (as generally is the case for a precipitation fields: Erdin *et al.*, 2012). For these reasons it was decided that the square-root transformation was the best choice for the present operational-mode-oriented application.

3.4. Skill scores and cross-validation

The cross-validation of the methods has involved the usual leave-one-out scheme. At rain-gauge locations, one measurement is left out each time and the kriging estimate for that location is obtained using the remaining available measurements. It should be pointed out that during the cross-validation of each measurement located at (x_0, y_0), with corresponding to time T_a left out, the measurement at the same location (x_0, y_0) corresponding to time T_b was also left out. Therefore, no information pertaining to the present or previous time at the location in question is included in the estimation process. Several skill scores were employed to assess the performance of the results. These skill scores involve precipitation estimates, which, in our context, can be either radar-only or radar–rain-gauge combination derived and precipitation (ground) references, which are the rain-gauge measurements.

(a) In order to assess systematic errors, reference was made to the overall bias on a logarithmic (decibel) scale, which is defined as

$$\mathrm{Bias} = 10\log_{10}\frac{\sum_{i=1}^{N}\hat{y}_i}{\sum_{i=1}^{N}y_i}, \tag{19}$$

where $\hat{y}_i$ refers to the estimated value and y_i to the observed one. In the present context, bias describes the overall

agreement between the precipitation estimates and the precipitation references. The bias is averaged over the whole space–time window of interest of the sample using all available gauges in the country. A positive (negative) bias in dB denotes an overall radar overestimation (underestimation).

(b) The mean-root-transformed error (MRTE: Erdin, 2011), which is a generalization of the root-mean-square error (the mean-square error of the square-root-transformed variable):

$$\mathrm{MRTE} = \frac{1}{N}\sum_{i=1}^{N}(\sqrt{\hat{y}_i} - \sqrt{y_i})^2. \tag{20}$$

(c) The median absolute deviation (MAD), which is here defined as the median of the absolute values of the differences between the precipitation estimates and the precipitation reference values (Erdin, 2011):

$$\mathrm{MAD} = \mathrm{median}(|\hat{y}_i - y_i|). \tag{21}$$

(d) The Hanssen–Kuipers discriminant (HK: Hanssen and Kuipers, 1965; Woodcock, 1980):

$$V = \frac{AD - BC}{(A + B)(C + D)}. \tag{22}$$

The meanings of A, B, C and D are given in Table 1. The HK is a skill score that is used to assess the value of a 'yes' or 'no' forecast and it gives successes and failures equal weight. It was used to assess to what extent the estimation of 'yes' or 'no' precipitation was correct or false, compared with the ground-measured reference values. The threshold used to determine whether a measurement was considered 'yes rain' was 0.5 mm.

(e) Scatter, which is defined as half the distance between the 16% and 84% quantiles of the error distribution (Germann *et al.*, 2006). Scatter refers to the spread of estimate–reference ratios when all estimates and reference values are pooled together. The definition of scatter is closely connected to the selected error distribution from a hydrological (end-user) and radar-meteorological (operational remotely sensed samples of the spatio-temporal variability of the precipitation field) perspective. The error distribution is expressed as the cumulative contribution-to-total rainfall (hydrologist's point of view, y-axis) as a function of the estimate–reference ratio (radar-meteorologist's point of view, x-axis). An example of the error distribution is shown in figure 7 of Germann *et al.* (2006).

From the radar-meteorological point of view, the multiplicative nature of the error prevails over the additive one. For example, water on the radome, a wrong calibration constant or a bad estimate of the profile all result in a multiplicative error (i.e. a factor) rather than an additive error (i.e. a difference). This is why two of the scores used in this article (bias and scatter) are expressed as ratios in dB. A 3 dB scatter, for instance, means that radar–gauge ratios vary by a factor of 2. If the bias is zero, it is interpreted as follows: the radar-derived estimate lies within a factor of 2 of the gauge estimate for 68% of rainfall, while the uncertainty is larger for the remaining 32%. Scatter, as defined above, is a robust measure of the spread. It is insensitive to outliers for two reasons. Firstly, each radar–gauge pair is weighted by its contribution to the total rainfall (y-axis of the cumulative error distribution). An ill-defined large ratio that results from two small values, e.g. 2 mm/10 mm (about 7 dB), describes an irrelevant event from a hydrological point of view and only receives small weight. Secondly, if the distance between the 16% and the 84% percentiles is considered then the tails of the error distribution are not overrated. Another important advantage of the spread measure is that it is unaffected by the bias error and thus provides a more complementary view of the error in the estimates. The above definition of scatter is thus a better measure of the spread than the less resilient standard deviation.

Table 1. Definition of elements A, B, C and D in the Hanssen–Kuipers discriminant.

	Forecast	
Observed	Yes	No
Yes	A	B
No	C	D

3.5. *Description of events*

3.5.1. *Event 1: 21–22 August 2005 (from 1200 to 1200 UTC)*

In August 2005, Switzerland was hit by one of the most devastating precipitation events in its modern history (Willi *et al.*, 2011). Several days of continuous heavy rainfall caused severe flooding, landslides and debris flows. The event mainly affected the northern side of the Swiss Alpine main crest, including the Swiss Plateau and the Jura mountain range. Parts of western Austria and southern Germany were also affected. There were six fatalities and damage was estimated at about three billion CHF, which is the largest amount estimated for damage since the systematic recording of storm damages was first introduced in 1972 (Bezzola and Hegg, 2007). The event was characterized by outstanding 48 h precipitation sums, which underlined its sustained heavy rainfall nature. Rainfall accumulations exceeded the 50 year return levels at many stations (Frei *et al.*, 2008). The rainfall event lasted from 19–23 August, with the largest amounts measured on 21 and 22 August. On 21 August, the comparison between radar and station-derived precipitation amounts revealed substantial quantitative differences: radar underestimations were evident over both flat and mountainous regions. The overall bias of the daily totals (radar/gauges) that resulted was of the order of −2.2 dB. The overall bias for the totals on the previous (following) day was of the order +0.2 (−3.4) dB. Despite the large quantitative differences, the spatial patterns of radar-derived precipitation estimates and those from a dense network of approximately 450 gauges read daily were similar. The event analyzed in this article refers to 24 h of the 5 day event.

3.5.2. *Event 2: 8–9 August 2007 (from 1000 to 1000 UTC)*

Between 6 and 9 August 2007, a long period of sustained intense rainfall affected almost all of Switzerland and caused flooding along several medium to large rivers in the Jura and Swiss Plateau (Willi *et al.*, 2011). The flood plains along the Aare were affected in particular. The event triggered lively political discussions about flood intervention and river-flow management. Damage amounted to 380 million CHF. A meteorological analysis of this event is given in Schmutz *et al.* (2008) and the hydrological impacts are described in Bezzola and Ruf (2009). The main rainfall activity occurred on 7–8 August: for both days the radar underestimated precipitation with respect to the gauges. On 7 August, the hourly bias amounts were generally between −1 and −2 dB; on 8 August, they were usually between −2 and −3 dB. The event analyzed here refers to 24 h of the 4 day event.

3.5.3. *Event 3: 30 November 2009 (from 0400 to 1600 UTC)*

The 3 day event from 29 November to 1 December showed all the typical features of a heavy precipitation event on the southern slope of the Alps. The cold front that moved towards the Alps from the west was linked to a deep depression over England. Southwesterly currents ahead of the system caused orographic lifting south of the Alps for about 24 h and this led to the first precipitation. As the cold front approached the Alps, the intensity of precipitation increased and this resulted in a total of 100–150 mm of water in 48 h, with the snow line being at around 1000 m asl. The backing of the currents caused an orographic

depression over the western Po basin and the Genoa Gulf. As far as the northern side of the Alps is concerned, on 29 November the south foehn caused gusts of up to 114 km h^{-1} at one station and up to 171 km h^{-1} at higher altitudes. 12 h of this event have been studied (30 November). That day the radar underestimation was remarkable, in terms of both average and median precipitation: the daily BIAS was approximately −3 dB.

3.5.4. *Event 4: 7–8 August 2009 (from 1900 to 1900 UTC)*

The event of 8 August 2009 was a typical summer convective event: several cells were observed starting from late morning (diurnal cycle). Southwesterly upper-level winds were relatively weak, so that the cells were almost stationary, causing large precipitation amounts that were spatially very confined. The region between the Zug and Zurich lakes was affected the most. The main radar–gauge combination challenge in this case was the small spatial structure and sparse sampling of the rainfall field by the rain-gauges (even greater than the usual undersampling!).

4. Results

4.1. *Description of the process*

The exploratory analysis and the estimation procedure are described here in some detail for the first of the four analyzed events, in order to provide the reader with an idea of the problem and the process that was followed. The rest of the events will be used to demonstrate the main results of this work. Event 1 covered a period of 24 h: 21–22 August 2005 (from 1200 to 1200 UTC). It is useful to track the difference between the two sensors – rain-gauges and radar – hour by hour to obtain some insight into the event. Setting the rain-gauge precipitation measurements as a reference, an exploratory analysis of the data has revealed that there has been a significant underestimation of the precipitation by the radar. This is evident in the probability density function graphs for the rain-gauge and radar data at rain-gauge locations (Figure 1). It is visually discernible that the radar estimates lie closer to the origin while the rain-gauge measurements extend to larger values. The rain-gauge rug plots make it easy to detect several sizeable rain-gauge measurements, which are usually absent in the corresponding radar densities and rug plots. The rug plots refer to the small vertical lines seen just underneath the probability density functions in Figure 1. Each line represents one data value occurrence. Their purpose is to give the reader an idea of the location of individual data, since this is sometimes difficult to see from the probability density function alone (especially for outliers).

For example, it is possible to observe the difference between the two densities where the heavy precipitation that was detected by the rain-gauges was not (accordingly) detected by the radar. Similarly, it is possible to look at the corresponding scatter plots in which the 45° line has been interposed. It can be seen that the radar values underestimate precipitation most of the time, an effect that is more pronounced when more intense rainfall

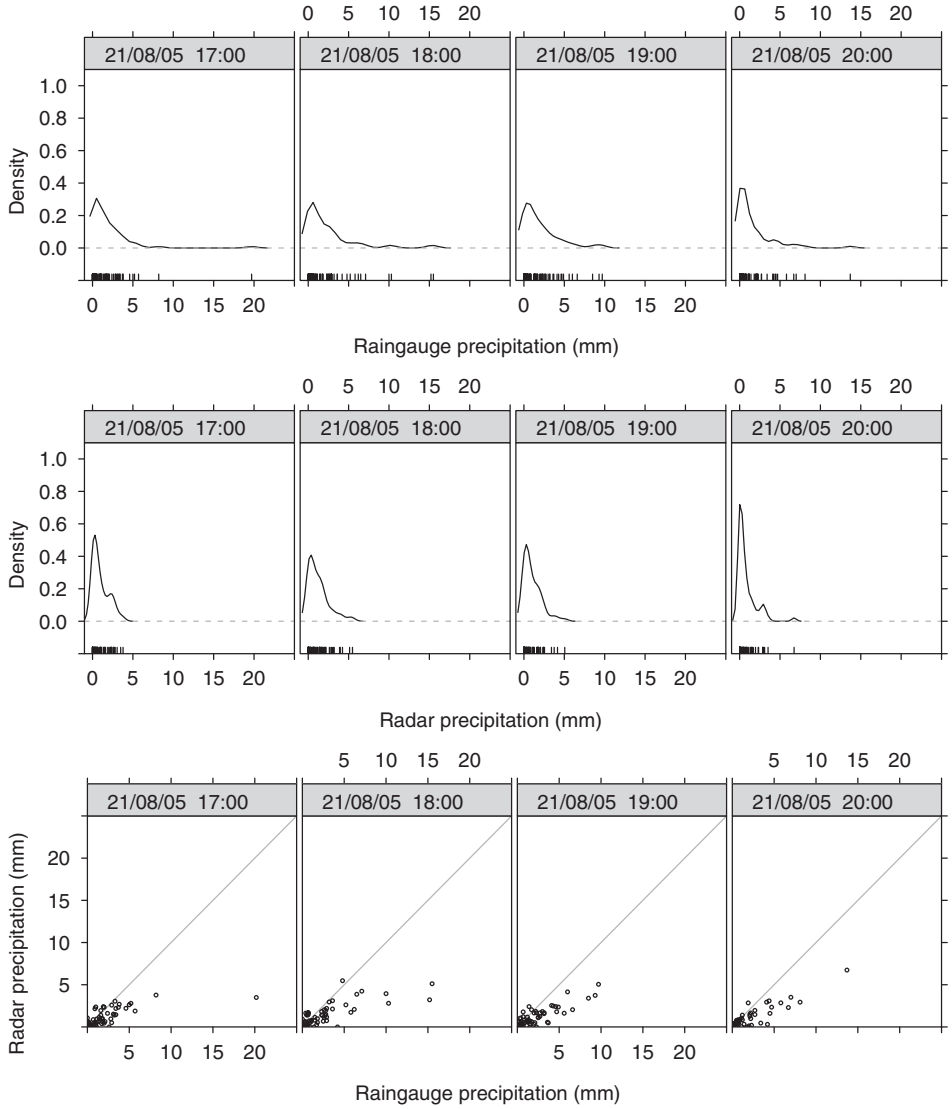

Figure 1. Rain-gauge measurement probability density function (first row), radar estimate probability density function (second row) and rain-gauge–radar scatterplot (third row) for four consecutive hours of event 1 (21–22 August 2005).

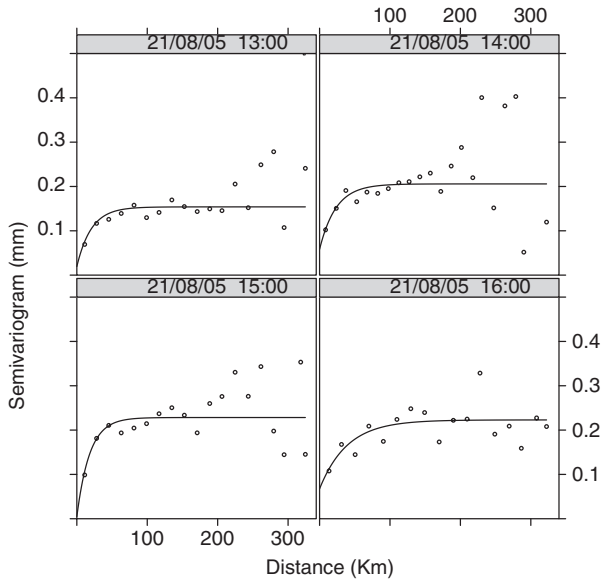

Figure 2. Sample of four direct variograms for four hours of event 1 (21–22 August 2005). Each of these variograms was fitted simultaneously with two other variograms – one direct and one cross-variogram – in the context of the spatio-temporal co-kriging process.

takes place. Unsurpisingly, the radar/rain-gauge bias for this 24 h sample was −2.52 dB.

The proposed geostatistical method was used to operate on the radar raster and adjust the value at each pixel, by applying a model constructed from rain-gauge data and radar data at the rain-gauge locations. This technique is based on the stochastic processes framework and relies on correlations between the square-root-transformed data, as described earlier in the mathematical formulation section. The next step is to compute the sample variograms for each hour and fit them using the (exponential) theoretical model given by Eq. (16). Co-kriging with external drift relies on the simultaneous fitting of two direct variograms and one cross-variogram. The fitting process employs routines from the R package 'gstat' Pebesma (2004, 2012b). It starts by fitting the model on the direct sample variogram of the residuals of aggregation period T_a (Eq. (6)) using weighted least-squares with weights $N(h)/h^2$, where $N(h)$ is the number of point pairs per variogram bin and h is the distance from the origin. The process then uses this fitted model as an origin to fit the three variograms (two direct and one cross-variogram) simultaneously.

A direct sample variogram of the residuals and the corresponding fitted theoretical variograms for four successive hours is presented in Figure 2. It should be mentioned here that variograms occasionally do not appear to have common-sense, intelligible behaviour when the correlations in the short range are larger than the correlations in the longer ranges. As previously mentioned, this behaviour is a result of a combination of causes, such as the involved errors (especially those of the radar), the support difference between the two monitoring devices, the limited representativeness of the rain-gauges and the small number of wet rain-gauges, but also the weather pattern microvariability, which may not be detected due to a lack of rain-gauges as they have not been positioned sufficiently close to each other. Moreover, the fitting process may occasionally yield singular models. To mitigate such effects, a nugget effect was allowed and the scheme described in section 3.2 was also employed to help improve the intelligibility of the variograms and to resolve singularity issues. This solution, although empirical, has been shown to function satisfactorily.

The original radar rainfall field and the rain-gauge-adjusted field that is the output of the present kriging application (CED.e) can be seen in Figure 3. Circles have been superimposed at the automatic rain-gauge locations. The colour of each circle represents the precipitation measured by the rain-gauge at each location. It can be observed that there is much better agreement between the estimated precipitation and the rain-gauge values in the rain-gauge-adjusted image. This results from the fact that the rain-gauge measurements were taken as the reference or primary variable within the kriging mechanism. It should also be observed that, since a nugget effect has been introduced, the precipitation estimates may differ from the rain-gauge measurements in the rain-gauge surroundings. Such a difference is usually small and reflects the variance associated with the nugget value of the model. This variance accounts for measurement errors but also for the effect of microvariability, which cannot be captured sufficiently by the existing measurements due to inadequate sampling.

The per-hour bias and MRTE of each of the four events can help us visualize the behaviour of CED.e better. In Figures 4, 5, 6 and 7 it is possible to see how CED.e produces improved results compared with the radar-only rainfall estimates. As far as bias is concerned, the improvement is both significant and consistent for all cases. The radar underestimates precipitation by 2–3 dB in cases 1–3. The presented scheme improves this picture dramatically by lowering the bias to an almost zero value. In general, this is accompanied by improvement in MRTE. In case 4, a clear improvement can also be observed regarding bias. However it can be seen that this is accompanied by an improvent in MRTE for some hours only, while for the rest, radar seems to behave slightly better. This can be attributed to two special characteristics of this particular event. Firstly, the number of rain-gauges that measured an appreciable quantity of precipication is small; only 10 rain-gauges measured precipitation more than 0.5 mm h^{-1}. Secondly, several intense convective cells were present. In situations like this, it is often difficult or even impossible to capture the characteristics of

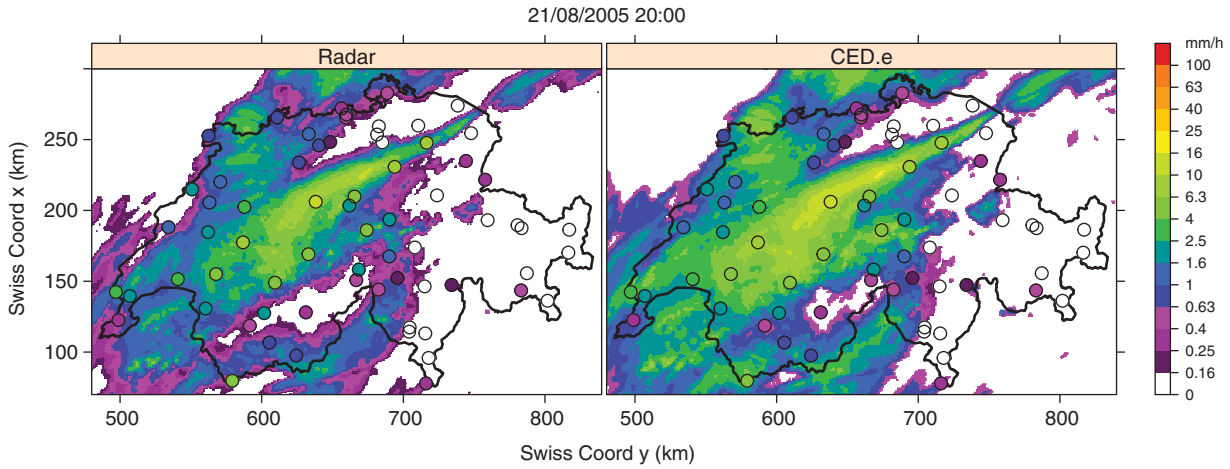

Figure 3. Event 1 (21–22 August 2005): comparison between the radar and CED.e hourly precipitation field. The small circles are colour-coded according to the rain-gauge measurement.

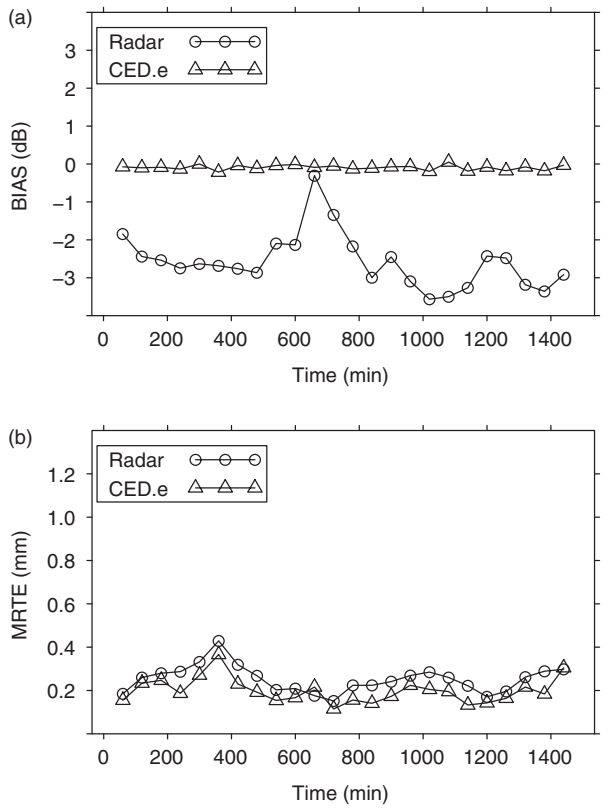

Figure 4. Event 1 (21–22 August 2005): hourly comparison of the bias and MRTE scores derived from the radar and CED.e estimates.

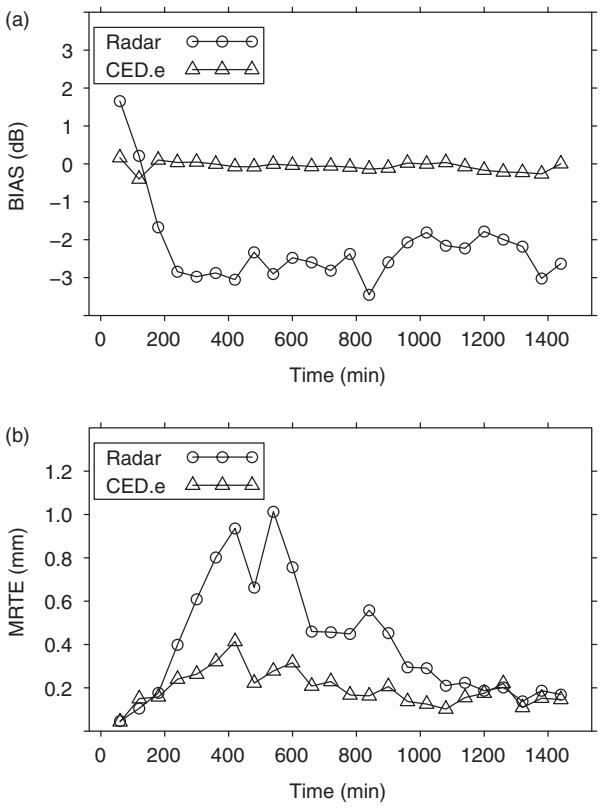

Figure 5. Event 2 (8–9 August 2007): hourly comparison of the bias and MRTE scores derived from the radar and CED.e estimates.

the underlying stochastic process sufficiently well, since the probability of having a sufficient number of rain-gauges very close to the maxima of the cells is relatively small. This may make computational structures, like the sample variogram, insufficiently representative of the situation and render the subsequent modelling problematic, a fact that is naturally reflected in some of the skill scores. Later, in the conclusions section (section 5), mention will briefly be made of a concept the authors are currently working on, which may serve (to a certain extent) as a remedy in cases like this.

4.2. Effects of input aggregation period

Geostatistical methods are expected to be more successful when the involved variables are characterized by high spatial (or spatio-temporal) correlations. In principle, it is possible to apply kriging to short-period aggregated data, for example for 10 minutes. However, such short time-scales are often characterized by low correlations between radar estimates and rain-gauge measurements, an effect that may result in an unacceptable outcome and consequently cause problems in estimation process operations. The importance of this point has been recognized and commented on since the 1970s, when Zawadzki reported that 'the radar–rain-gauge comparison is acceptably good when total amounts over several hours of rainfall are considered. Instantaneous comparisons show great discrepancies...' (Zawadzki, 1975). In this sense, and within our context, by 'non-robust' we mean that there are inconsistent data in terms of radar/rain-gauge comparisons, probably related to large errors and/or insufficient representativeness of the input. For these reasons, we opted for the employment of hourly aggregated input for the estimation mechanism. The authors believe that, in the context of real-time no-human-intervention applications, hourly input aggregations provide an acceptable balance between very high temporal resolution, but non-robust, aggregations and very robust but low temporal resolution input data.

The radar/rain-gauge discrepancies are generated by both radar and rain-gauge measurement errors and rain-gauge representativeness issues, described in the introductory section. Regarding

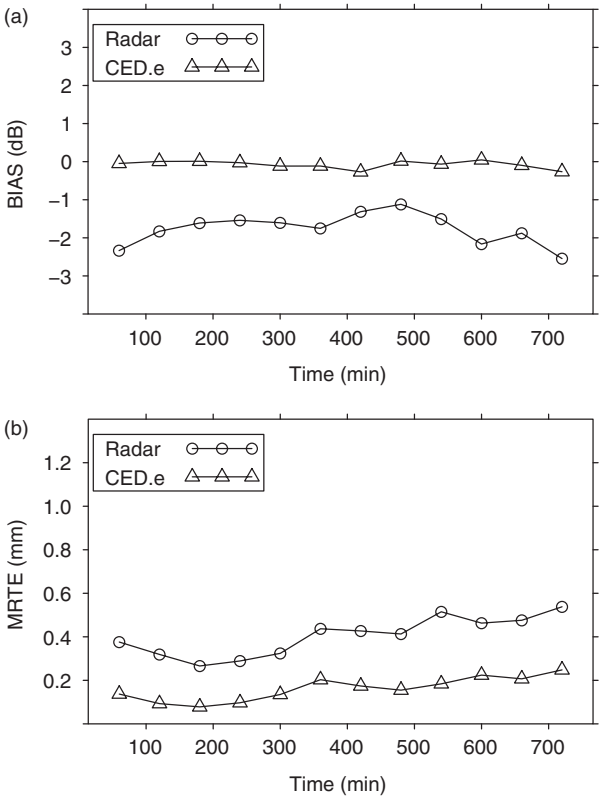

Figure 6. Event 3 (30 November 2009): hourly comparison of the bias and MRTE scores derived from the radar and CED.e estimates.

the latter, it has been shown that the spatial sampling error declines nonlinearly with accumulation time. It decreases very fast for time-scales of 1–60 min, but this decrease slows down considerably for time-scales between 60 min and 1 day (figure 7 of Villarini *et al.*, 2008). This is also qualitatively consistent with our choice of aggregation time. However, these results are more difficult to

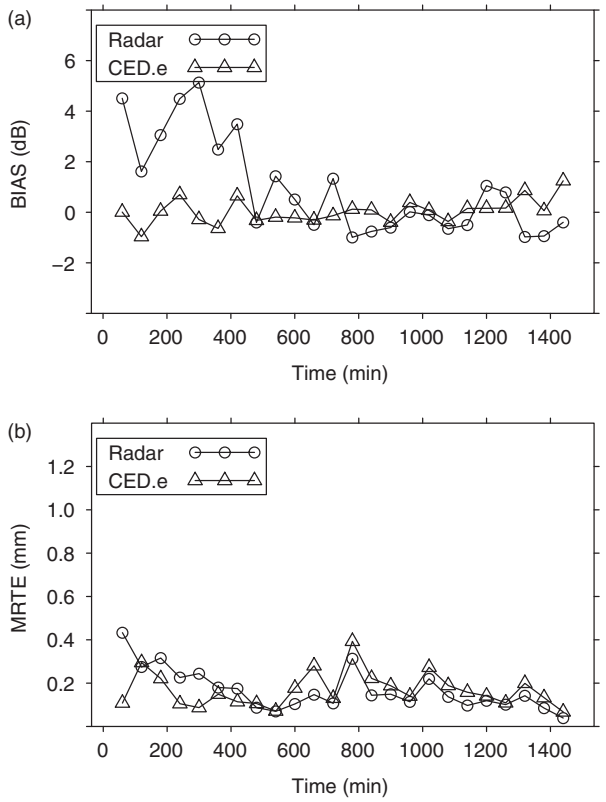

Figure 7. Event 4 (7–8 August 2009): hourly comparison of the bias and MRTE scores derived from the radar and CED.e estimates.

generalize accurately in an absolute quantitative way, especially due to the particular topography that characterizes Switzerland.

To demonstrate that the results do in fact deteriorate for a decrease in the aggregation period of the input data, cumulative skill scores were computed for progressively increasing input aggregation periods for the four investigated events. In particular, CED.e was applied for input data with input aggregations of 10, 20, 30 and 60 min. In order to construct a fair basis for comparison, all the CED.e estimates were aggregated to 60 min. For example, when using 10 min aggregated input CED.e was applied six times per hour, once for each of the six 10 min segments of that hour, then the six kriging estimates were aggregated to produce hourly estimates. This process was adopted for all of the aggregations and the skill scores were computed for these final aggregated estimations. This approach avoids biases associated with comparing precipitation data from different time-scales and comparing sets with different numbers of elements. The results can be seen in Table 2. It is possible to observe that, while bias and scatter remain relatively stable as the aggregation period of the input is increased, there is generally a sizeable improvement in terms of MRTE, MAD and HK. This demonstrates that for processes such as kriging, which depend to a great extent on coupled radar–rain-gauge data, there is a quantitatively consistent dependence of the output on the aggregation period input, a result that can be considered a by-product of the aforementioned observation by Zawadzki, which it thus supports.

4.3. *Comparison between KED and CED*

It was expected that CED.e would show some advantages over other estimators due to the introduction of temporal (in addition to spatial) data into its mechanism. This was found to be true to some extent. Both KED and CED.e were found to behave better than ORK for hourly aggregated input and to show an improvement over radar rainfall estimates. These results were expected and are reported in Table 3. It is worth noting that ORK was always able to mitigate the bias and hence significantly reduce the average error. As far as the dispersion of the error (see e.g. the scatter, MRTE and MAD scores) and wet versus dry discrimination (see e.g. the HK score) are concerned, ORK has shown improvement over radar estimates only in three events (out of four), while it has failed for the most important one: the devastating flood of August 2005.

The CED.e and KED skill scores are similar for input aggregation equal to 60 min; the comparison between these two schemes requires further investigation in order to identify whether they are characterized by any significant differences. Both techniques were therefore applied to different input aggregation periods, that is 10, 20, 30 and 60 min. The final estimates were aggregated for each case to produce hourly values, as mentioned in section 4.2. Focus is given here to bias to demonstrate its behaviour and the differences between radar, KED and CED.e.

In Figure 8 it can be seen that, for short-time aggregated input where the input is expected to be less and less robust, the direct variogram fitting associated with KED occasionally results in dramatic failures, which can be seen on sudden discontinuities and which appear as spikes in the graphs. On the other hand, co-kriging with external drift with a default set of parameters (CED) behaves much more satisfactorily, while CED.e seems to operate without any problems and shows remarkable stability. This behaviour clearly suggests that a failure of the modelling process, in particular the fitting of the variogram, has taken place and it is plausible to assume that this is directly connected to non-robustness in terms of correlations, i.e. there are inconsistent

Table 2. CED.e skill scores for all events with varying aggregation period input.

Event	Method	Aggr. (min)	Bias (dB)	MRTE (mm)	MAD (mm)	HK	Scatter (dB)
1	CED.e	10	−0.06	0.236	0.445	0.532	2.26
1	CED.e	20	−0.08	0.217	0.416	0.592	2.19
1	CED.e	30	−0.08	0.209	0.408	0.612	2.13
1	CED.e	60	−0.09	0.199	0.379	0.632	2.1
2	CED.e	10	−0.02	0.218	0.479	0.68	1.84
2	CED.e	20	−0.03	0.206	0.458	0.721	1.83
2	CED.e	30	−0.04	0.207	0.443	0.735	1.82
2	CED.e	60	−0.05	0.196	0.42	0.738	1.85
3	CED.e	10	0.01	0.16	0.452	0.212	1.68
3	CED.e	20	0	0.157	0.43	0.188	1.63
3	CED.e	30	−0.01	0.158	0.445	0.222	1.67
3	CED.e	60	−0.08	0.161	0.432	0.277	1.65
4	CED.e	10	−0.06	0.221	0.184	0.8	2.57
4	CED.e	20	−0.06	0.199	0.159	0.812	2.59
4	CED.e	30	−0.01	0.187	0.141	0.817	2.60
4	CED.e	60	−0.05	0.17	0.103	0.829	2.59

Table 3. Radar, ORK, KED and CED.e skill scores for aggregation period input of 60 min.

Event	Method	Aggr. (min)	Bias (dB)	MRTE (mm)	MAD (mm)	HK	Scatter (dB)
1	Radar	60	−2.53	0.252	0.295	0.672	2.12
1	ORK	60	−0.06	0.405	0.663	0.319	3.02
1	KED	60	−0.06	0.198	0.383	0.632	2.07
1	CED.e	60	−0.09	0.199	0.379	0.632	2.1
2	Radar	60	−2.56	0.408	0.582	0.718	2.45
2	ORK	60	0.04	0.234	0.483	0.699	1.82
2	KED	60	−0.06	0.196	0.428	0.749	1.80
2	CED.e	60	−0.05	0.196	0.42	0.738	1.85
3	Radar	60	−1.74	0.404	0.792	0.531	2.39
3	ORK	60	0.23	0.163	0.523	0.047	1.68
3	KED	60	0.05	0.164	0.482	0.153	1.73
3	CED.e	60	−0.08	0.161	0.432	0.277	1.65
4	Radar	60	0.9	0.169	0.009	0.813	2.76
4	ORK	60	0.16	0.475	0.312	0.516	5.50
4	KED	60	0.03	0.184	0.118	0.81	2.55
4	CED.e	60	−0.05	0.17	0.103	0.829	2.59

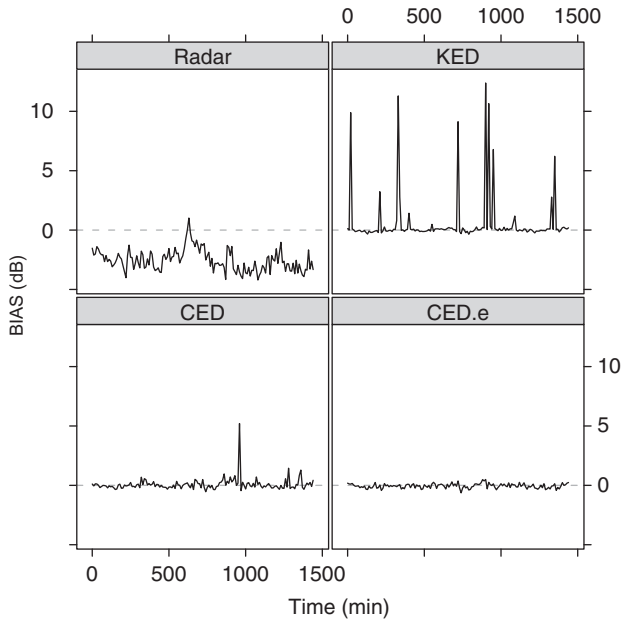

Figure 8. Bias comparison. The panels show the bias estimated using four different methods for event 1 (21–22 August 2005) and 10 min aggregation period. Intense discontinuities appear in the case of (default) KED. One intense discontinuity appears in the case of default CED (at time = 950 min). CED.e manages to keep the bias very close to zero. The time aggregation used here is 10 min. [Correction added 24 February 2014 after original online publication: the labeling of the y-axis has been corrected to 'BIAS (dB)'.]

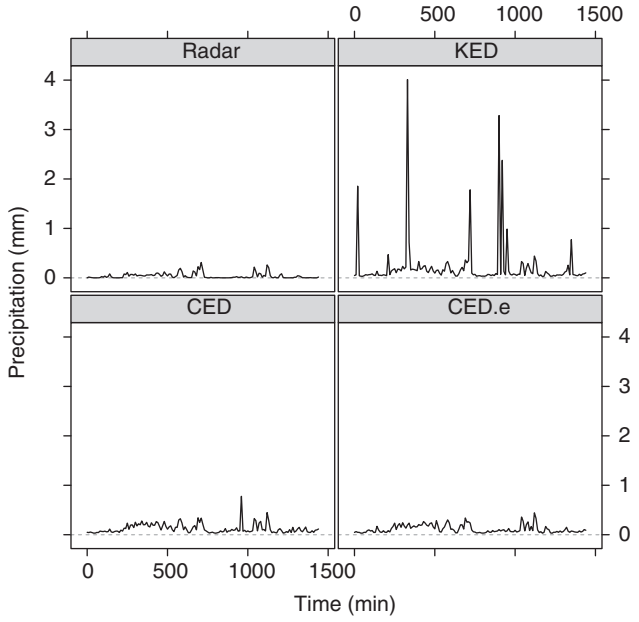

Figure 9. Temporal continuity comparison. The panels show the precipitation depths estimated using four different methods at a single pixel for event 1 (21–22 August 2005) and 10 min aggregation period. The discontinuities reflect problems in the modelling process, as also evident in Figure 8. CED.e preserves the continuity observed in the radar time series.

data in terms of radar/rain-gauge comparison, probably related to large errors and/or insufficient representativeness of the input. In an attempt to investigate the reasons for such behaviour in more detail, the sample variograms were examined closely in cases in which the KED direct variogram fitting process failed. Such variograms were found to be clearly unintelligible and to be characterized by rather erratic behaviour, especially for short ranges, which justifies why well-tested robust regression schemes such as the one used in 'gstat' (iteratively reweighted least-squares: Cressie, 1993; Pebesma, 2004) failed. However, the present investigation suggests that, in such harsh terrains, CED and CED.e have in fact a much better chance of succeeding, mainly and primarily because of the employment of temporal information (and consequently the fitting of three instead of one variogram) and then because of the employment of the extensive variogram-fitting techniques mentioned in section 3.1.

The effect of such fitting failures in terms of estimation and continuity of the outcome can be seen in the precipitation-estimate plots of a single pixel during an entire event (Figure 9). Referring to one pixel only does not limit the generality of the argument: similar behaviour can be observed at each pixel, since the erroneous modelling affects the estimate of the entire precipitation map. In these plots it is possible to observe how the modelling failures (which are also in the aforementioned bias discontinuities; see Figure 8) translate in the rainfall map. While the radar time series and the CED.e scheme appear to have a continuous precipitation evolution, the default KED scheme shows unacceptably high precipitation values while the default CED scheme seems to be problematic only on one occasion (a spike can be observed at time = 950min). In a real-time no-human-intervention operational environment, such failures would have an impact on the reliability of the application.

Although the aforementioned discontinuities do not necessarily appear in each and every event, it is not clear how to define an aggregation threshold above which they stop emerging, since such effects seem to depend on the quality/robustness of the data at hand and can ultimately be case-dependent. The described effect was most pronounced in events 1 and 4, while it had much

Table 4. Radar, KED and CED.e skill scores for aggregation period input of 10 min.

Event	Method	Aggr. (min)	Bias (dB)	MRTE (mm)	MAD (mm)	HK	Scatter (dB)
1	Radar	10	−2.53	0.252	0.295	0.672	2.12
1	KED	10	1.57	0.52	0.668	0.423	2.61
1	CED.e	10	−0.06	0.236	0.445	0.532	2.26
2	Radar	10	−2.56	0.408	0.582	0.718	2.45
2	KED	10	−0.03	0.227	0.482	0.678	1.84
2	CED.e	10	−0.02	0.218	0.479	0.68	1.84
3	Radar	10	−1.74	0.404	0.792	0.531	2.39
3	KED	10	0.13	0.18	0.575	0.071	1.78
3	CED.e	10	0.01	0.16	0.452	0.212	1.68
4	Radar	10	0.9	0.169	0.009	0.813	2.76
4	KED	10	3.76	0.945	0.217	0.646	3.33
4	CED.e	10	−0.06	0.221	0.184	0.8	2.57

Table 5. Radar, KED and CED.e skill scores for aggregation period input of 20 min.

Event	Method	Aggr. (min)	Bias (dB)	MRTE (mm)	MAD (mm)	HK	Scatter (dB)
1	Radar	20	−2.53	0.252	0.295	0.672	2.12
1	KED	20	0.03	0.219	0.406	0.588	2.10
1	CED.e	20	−0.08	0.217	0.416	0.592	2.19
2	Radar	20	−2.56	0.408	0.582	0.718	2.45
2	KED	20	−0.08	0.21	0.443	0.718	1.8
2	CED.e	20	−0.03	0.206	0.458	0.721	1.83
3	Radar	20	−1.74	0.404	0.792	0.531	2.39
3	KED	20	−0.01	0.166	0.506	0.141	1.68
3	CED.e	20	0	0.157	0.43	0.188	1.63
4	Radar	20	0.90	0.169	0.009	0.813	2.76
4	KED	20	0.83	0.306	0.183	0.722	2.63
4	CED.e	20	−0.06	0.199	0.159	0.812	2.6

smaller, or even no, consequence in events 2 and 3. Discontinuities even appeared for longer than 10 min aggregated input, especially for event 4, which was characterized by a small number of wet rain-gauges. This is intuitively understandable: variogram modelling can be expected to become gradually more challenging as the number of available wet rain-gauges decreases, since such a decrease often affects the intelligibility of the variogram. The results for 10 and 20 min aggregated input can be seen in Tables 4 and 5. It is evident in these tables that the skill scores between default KED and CED.e show a considerable difference for events 1 (10 min) and 4 (10 and 20 min), due to the aforementioned effects.

Input data may occasionally be non-robust. In the present context, the robustness of the input data depends on many factors besides meteorological ones, such as maintenance, quality-control schemes, calibration, density of the rain-gauge network, etc. When it comes to the critical issue of automatic no-human-intervention modelling and the subsequent estimation processes, spatio-temporal CED seems to provide a more stable scheme than KED, despite the fact that more parameters need to be estimated for CED than for KED. In fact, this work suggests that, in the present context at least, a process that deals with non-robust data may benefit considerably when a more complex spatio-temporally-oriented system of parameter estimation is involved.

4.4. 5 min precipitation maps

A question naturally arises: how can one produce rain-gauge–radar combination precipitation maps with smaller than hourly resolution? Such maps may be necessary, for instance, for nowcasting applications, which generally require short time-scale input that represents 10 or even 5 min precipitation fields. Although geostatistical combination methods may, in principle, be applied to short-time aggregated input data, such attempts do not always provide accurate results, as previously mentioned. Moreover, modelling stability is not ensured.

A straightforward method is proposed to tackle this problem: kriging is performed for hourly aggregated input and the resulting estimation is disaggregated into rasters that have time resolution 5 min. In particular, the following algorithmic scheme is adopted: (a) kriging is performed with hourly aggregated input (as described in the previous sections); (b) the fraction of precipitation for each of the 12 5 min segments of the whole hour is computed at every pixel location; this is achieved by dividing, for each pixel, the 5 min radar precipitation map by the hourly radar precipitation map; and (c) this fraction is then multiplied by the hourly kriging estimate of that pixel (as described in the first step). In this way a series of 12 values is produced for each pixel of the precipitation map, each of which corresponds to a 5 min segment of the hour in question. The process is described mathematically as

$$\hat{p}_{5_i} = \hat{p}_{60}\frac{r_{5_i}}{r_{60}} \qquad i \in [1, 12], \tag{23}$$

with,

$$r_{60} = \sum_{i=1}^{N=12} r_{5_i}, \tag{24}$$

where $\hat{p}_{60}$ represents the hourly kriging estimation, r_{5_i} the 5 min radar estimates for segment i of the hour in question, r_{60} the hourly aggregated radar value and $\hat{p}_{5_i}$ the resulting 5 min kriging-based rainfall estimate. The resulting images of this process can be seen in Figure 10.

This approach is straightforward and computationally inexpensive when vectorization techniques are employed and

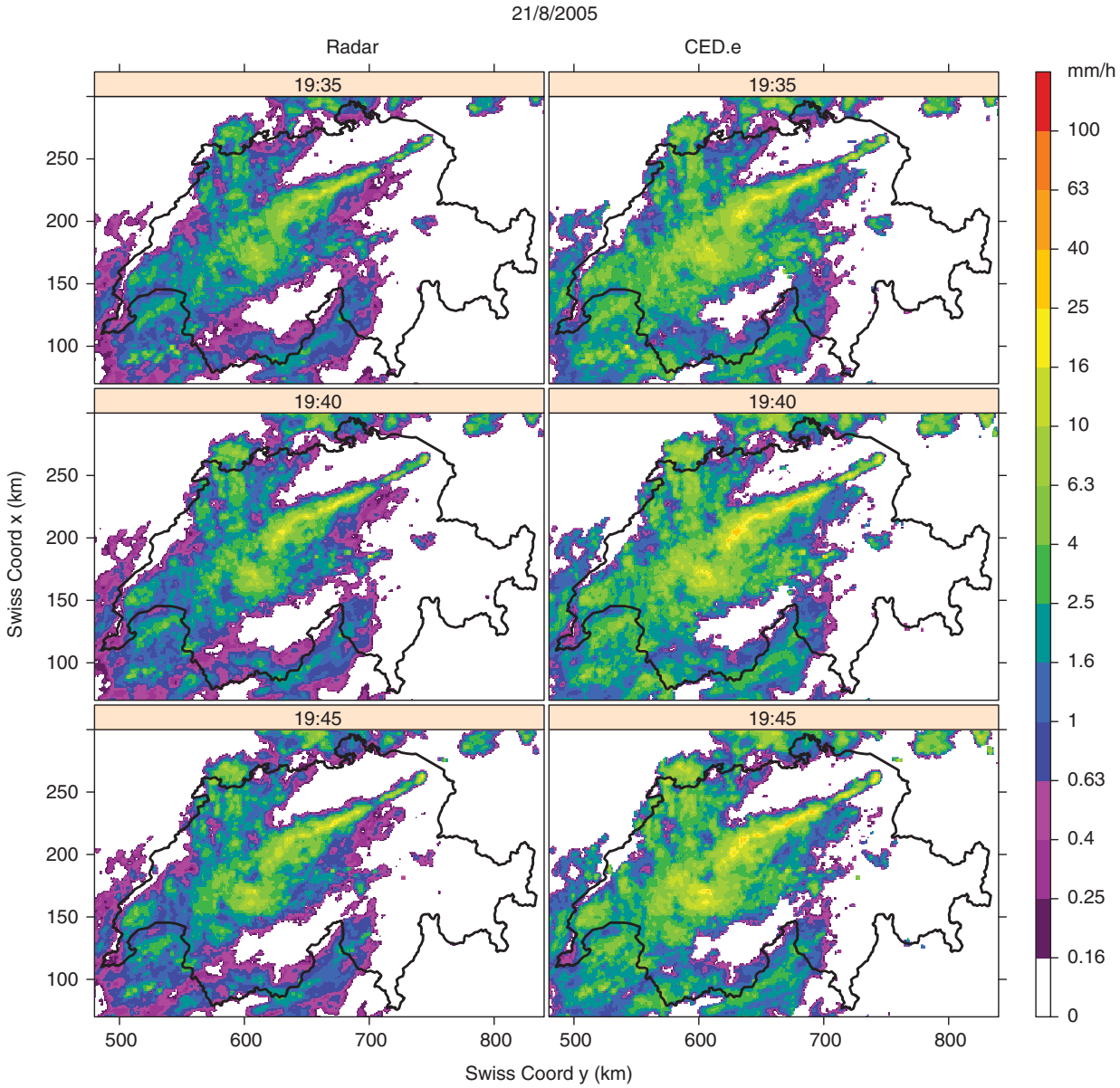

Figure 10. Sequential panels with pictures of 5 min precipitation fields. The colour scale is equivalent to an hourly rain rate (in practice each 5 min raster has been multiplied by 12 to produce hourly equivalent values.)

it follows a scheme that was proposed for daily to hourly disaggregations (Paulat *et al.*, 2008; Wüest *et al.*, 2009). It involves the reasonable assumption that the radar estimate errors are equipped with sufficient temporal continuity (therefore the overestimation or underestimation behaviour and corresponding errors of pixels do not change erratically within the period of one hour). However, this approach involves two simplifications/assumptions: (a) the fact that the radar error depends on the amount of precipitation is not taken into account; and (b) it is assumed that the fractions of precipitation at each location are representative. This may not be the case for locations that the radar cannot monitor with sufficient accuracy. However, as a first approximation this technique seems to operate reasonably well. Moreover, it is computationally fast enough to be integrated smoothly within a production line of an operational environment. It has to be emphasized that, within the framework of an operational environment, this disaggregation does not involve a delay of one hour. In such an environment the update time is different from the aggregation time, e.g. hourly images may be produced every 10 min; thus, although the whole last hour can be disaggregated into 12 segments, only the last two 5 min grids are actually new, while the first 10 segments have been computed by disaggregations of previous times. This makes it suitable for providing data to hydrological models for forecasting flash floods.

5. Conclusions

The geostatistical scheme presented in this article (CED.e) relies on an established and well-tested technique (KED) but it expands it by introducing and coupling temporal data with spatial data through co-kriging. The described approach is conceptually simple, practical in terms of design and coding and relatively straightforward to reproduce, since it is based on easily accessible open-source libraries. Moreover, it is equipped with additional mechanisms to avoid modelling pitfalls when the input data are limited and/or subject to significant errors. This addition was motivated by the automatic character required for the present application. Besides, since the proposed technique has the aim of serving as a real-time operational tool in Switzerland, it was important to assure that the code operated efficiently in terms of computational speed and reliability. Significant time lags are unacceptable in a real-time operational environment, since results have to be produced quickly to be useful (as input) for applications such as nowcasting or hydrological simulations.

The benefits of CED.e, compared with radar or with ORK, can easily be seen. The improvement is evident in literally all skill scores (Table 3). To establish the extent to which CED.e and KED differ, the methods were applied in aggregation period 10–30 min data. Such data are commonly not robust

in the aforementioned sense, i.e. they are often characterized by discrepancies in terms of correlations between radar estimates and rain-gauge measurements. In this context, CED.e shows superior stability to KED.

This result is important for two reasons. Firstly, if CED.e is considered simply as a general geostatistical method, which can be used in applications completely different from the ones presented here, its stability may prove to be a critical issue. It is not unrealistic for a geostatistical application to have to deal occasionally with non-robust input. For instance, it may not be possible to produce long time-scale aggregated input easily (as is possible with precipitation fields). In this context, CED.e provides a practical and straighforward remedy for the possible modelling pitfalls of KED. Secondly, in the context of QPE, time-scales during which data become sufficiently robust and KED can therefore perform satisfactorily are not always given. This seems to depend on several factors, including technicalities like how well-set, calibrated and maintained the monitoring devices are, as well as the meteorology and topography conditions. The fact that data produced by the Swiss monitoring system start losing robustness for aggregation periods of less than one hour does not establish a principle of an input-aggregation threshold of satisfactory robustness. Such a threshold may be difficult to establish and probably changes case-by-case. Moreover, it may depend on the particular characteristics of the network and monitoring devices available in each country. Since this threshold is difficult to define, the stability advantage of CED.e over KED becomes a desirable feature, especially in operational environments.

One additional advantage of CED.e emerges from its mathematical formulation. CED.e has the ability to incorporate additional point measurements as co-kriged secondary variables. In the present work, temporal data were co-kriged, but there is no reason why other interesting possibilities should not be considered. For example in Switzerland, besides the automatic Swiss rain-gauge network, which is characterized by high-level quality control, there are also secondary rain-gauge networks of lower quality, a fact which, in principle, prohibits them from serving as primary variables. However, the rainfall information from such networks may prove important, especially in cases of significant microvariability, since these networks cover large parts of the Swiss area. Although a more thorough quality check of such measurements would be preferred, this may not be possible within the available time limits or may not even be possible for a variety of technical or financial reasons. CED.e has the ability to incorporate such information into its mechanism not as primary but as secondary data. In this sense, it provides the following scheme: all information is incorporated but the low-quality parts are given less importance/weight. The authors are investigating such possibilities and the results will appear in a future work.

The second main contribution of this article is the introduction of a simple, but efficient, method to produce very high temporal resolution precipitation maps, based on maps produced for longer time-scales using kriging. This method can be considered a first approximation for maps with spatio-temporal resolution equal to that of the radar. It relies on the fractions of precipitation corresponding to the radar maps with 5 min temporal resolution. Assuming that the errors in radar measurements remain consistent within a period of one hour, it is not unrealistic to expect that, although the absolute radar estimations are subject to significant errors, the fractions of precipitation over 5 min segments are representative to a satisfactory degree. The introduction of maximum temporal resolution rain-gauge data (10 min) would probably complicate the algorithm unnecessarily, since rain-gauge data may be problematic in terms of representativeness for resolutions that high. This technique can be considered an alternative to applying a geostatistical application to precipitation input aggregations less of than one hour. Such an application would be impaired, due to input-robustness problems, a problem that has been discussed extensively throughout this article. However, it would be interesting to quantify the aforementioned arguments in a future work and explore how the output of our disaggregation schemes compares with the results produced by applying kriging for very short-time aggregated input, e.g. 10 min.

Finally, two observations of practical interest should be made. Firstly, in general kriging behaves well as an interpolation process, while (unsurprisingly) it fails one way or another as an extrapolation process. This is typical of interpolation schemes; the results deteriorate progressively as the distance from the interpolation regime increases (the area covered by the primary variable locations – the rain-gauges in this article). Such problems can and do cause significant estimation artefacts, especially when the input data are subject to significant errors. Secondly, due to the limited number of rain-gauge measurements, the choice of rain-gauge installation locations and the complexity of the meteorology situation, the microvariability of the precipitation is not always captured in a satisfactory manner. A typical example is event 4, which is rich in convective cells. In such cases there are usually not enough rain-gauges measuring sizeable amounts of rainfall. Moreover, the ones in which precipitation can be observed are rarely located close to the maxima of the convective cells. For this reason, kriging may miss or significantly depreciate the precipitation depth in the neighbourhood of the convective cells.

The aforementioned problems can limit the practical use of a geostatistical scheme to a great extent, even when it behaves well in many other respects. Some first tests suggest that it is possible to mitigate the effect of both problems significantly by employing 'virtual rain-gauges'. The concept of virtual rain-gauges involves introducing into the geostatistical scheme radar or adjusted-radar estimates at certain pixels of the radar raster as rain-gauge measurements. Such an introduction, although delicate to implement, seems to make the extrapolation artefacts disappear almost entirely. Moreover, it provides a method for effectively manipulating important local information such as that associated with convective cells.

Acknowledgements

The authors thank their colleagues in MeteoSwiss, especially the members of the CombiPrecip project, for their support and contributions. This work has been made possible thanks to the Swiss projects CombiPrecip-NCCRIII and Rad4Alp.

References

Berenguer M, Zawadzki I. 2008a. A study of the error covariance matrix of radar rainfall estimates in stratiform rain. *Weather and Forecasting* **23**: 1085–1101.

Berenguer M, Zawadzki I. 2008b. A study of the error covariance matrix of radar rainfall estimates in stratiform rain. Part II: Scale dependence. *Weather and Forecasting* **24**: 800–811.

Bezzola GR, Hegg C. 2007. '*Ereignisanalyse Hochwasser 2005, Teil 1 – Prozesse, Schaden und erste Einordnung*', Umwelt-Wissen, 0707, 215. Bundesamt fur Umwelt BAFU, Eidgenossische Forschungsanstalt WSL: Bern, Switzerland.

Bezzola GR, Ruf W. 2009. '*Ereignisanalyse Hochwasser August 2007. Analyse der Meteo- und Abussvorhersagen; vertiefte Analyse der Hochwasserregulierung der Jurarandgewasser*', Umwelt-Wissen, 0927, 209. Bundesamt fur Umwelt: Bern, Switzerland.

Borga M, Vizzaccaro A. 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *J. Hydrol.* **195**: 160–171.

Brandes EA. 1975. Optimizing rainfall estimates with the aid of radar. *J. Appl. Meteorol.* **14**: 1339–1345.

Cesare LD, Myers D, Posa D. 2001. Product–sum covariance for space–time modelling: an environmental application. *Environmetrics* **12**: 11–23.

Ciach GJ, Krajewski WF. 1999. On the estimation of radar rainfall error variance. *Adv. Water Resources* **22**: 585–595.

Cole SJ, Moore RJ. 2009. Distributed hydrological modelling using weather radar in gauged and ungauged basins. *Adv. Water Resources* **32**: 1107–1120.

Cressie N. 1993. *Statistics for Spatial Data, Revised Edition*. John Wiley and Sons Inc.: New York, NY.

Cressie N, Huang H-C. 1999. Classes of nonseparable stationary covariance functions. *J. Amer. Stat. Assoc.* **94**: 1330–1340.

Creutin JD, Dekrieu G, Lebel T. 1987. Rain measurement by raingage–radar combination: a geostatistical approach. *J. Atmos. Oceanic Technol.* **5**: 102–114.

De Iaco S, Myers D, Posa D. 2001. Space–time analysis using a general product-sum model. *Statistics & Probability Letters* **52**: 21–28.

Erdin R. 2009. '*Combining rain gauge and radar measurements of a heavy precipitation event over Switzerland*', Veroeffentlichungen der MeteoSchweiz. MeteoSwiss: Zürich; 110pp.

Erdin R. 2011. '*Evaluation guidelines of CombiPrecip*', Internal Report. MeteoSwiss: Zürich; 10pp.

Erdin R. 2012. '*Geostatistical methods for hourly radar–gauge combination: An explorative, systematic application at MeteoSwiss*,' Scientific Report No. 92. MeteoSwiss: Zürich.

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

Frei C, Germann U, Fukutome S, Liniger M. 2008. '*Moglichkeit und Grenzen der Niederschlagsanalyse zum Hochwasser 2005*', Arbeitsbericht der MeteoSchweiz Nr. 221, Swiss Federal Office of Meteorology and Climatology MeteoSwiss: Zürich; 21.

Gabella M, Notarpietro R. 2004. Improving operational measurement of precipitation using radar in mountainous terrain. *IEEE Geosci. Remote Sensing Lett.* **1**: 78–83.

Gabella M, Joss J, Perona G. 2000. Optimizing quantitative precipitation estimates using a noncoherent and a coherent radar operating on the same area. *J. Geophys. Res.* **105**: 2237–2245.

Gabella M, Morin E, Notarpietro R. 2011. Using TRMM spaceborne radar as a reference for compensating ground-based radar range degradation: Methodology verification based on rain gauges in Israel. *J. Geophys. Res.* **116**: 1–14.

Germann U, Joss J. 2002. Mesobeta Profiles to Extrapolate Radar Precipitation Measurements above the Alps to the Ground Level. *J. Appl. Meteorol.* **41**: 542–557.

Germann U, Joss J. 2004. Operational measurement of precipitation in mountainous terrain. In *Weather Radar: Principles and Advanced Applications, Series Physics of Earth and Space Environment* Meischner P (ed). Springer-Verlag: Berlin, Germany; pp 52–77.

Germann U, Gianmario G, Boscacci M, Bolliger M. 2006. Radar precipitation measurement in a mountainous region. *Q. J. R. Meteorol. Soc.* **132**: 1669–1692.

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

Gires A, Schertzer D, Tchiguirinskaia I, Lovejoy S, Onof C, Maksimovic C, Simoes N. 2011. Impact of small scale rainfall uncertainty on urban discharge forecasts. *Weather Radar and Hydrology* **400**: 400–406.

Gneiting T. 2002. Nonseparable, stationary covariance functions for space–time data. *J. Amer. Stat. Assoc.* **97**: 590–600.

Goovaerts P. 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: Oxford.

Gregoire T, Lin Q, Boudreau J, Nelson R. 2008. Regression estimation following the square-root transformation of the response. *Forest Science* **54**: 597–606.

Grüter E, Abbt M, Häberli C, Haller E, Küng U, Musa M, Konzelmann T, Dössegger R. 2003. '*Quality control tools for meteorological data in the MeteoSwiss data warehouse system*', Internal report, MeteoSwiss: Zürich; 11pp.

Haas T. 1990. Lognormal and moving window methods of estimating acid deposition. *J. Amer. Stat. Assoc.* **85**: 950–963.

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

Habib E, Krajewski WF. 2002. Uncertainty analysis of the TRMM ground-validation radar-rainfall products: Application to the TEFLUN-B field campaign. *J. Appl. Meteorol.* **1975**: 558–572.

Hanssen A, Kuipers W. 1965. On the relationship between the frequency of rain and various meteorological parameters. *Meded. Verhand.* **81**: 2–15.

Harris P, Charlton M, Fotheringham AS. 2010. Moving window kriging with geographically weighted variograms. *Stochastic Environmental Research and Risk Assessment* **24**: 1193–1209.

Jordan PW, Seed AW, Austin GL. 2000. Sampling errors in radar estimates of rainfall. *J. Geophys. Res.* **105(D2)**: 2247–2257.

Joss J, Lee R. 1995. The application of radar–gauge comparisons to operational preciptation profile corrections. *J. Appl. Meteorol.* **34**: 2612–2630.

Kitchen M, Blackall R. 1992. Representativeness errors in comparisons between radar and gauge measurements of rainfall. *J. Hydrol.* **13**: 33.

Krajewski WF, Georgakakos KP. 1987. Cokriging radar-rainfall and rain gauge data. *J. Geophys. Res.* **21**: 764–768.

Künsch H, Papritz A, Bassi F. 1997. Generalized cross-covariances and their estimation. *Mathematical Geology* **29**: 779–799. Available at http://www.springerlink.com/index/Y5853832M082M581.pdf.

Kyriakidis PD, Journel AG. 1999. Geostatistical space–time models: A Review. *Mathematical Geology* **31**: 651–684.

Mitas L, Mitasova H. 1999. Spatial interpolation. In *Geographical Information Systems: Principles, Techniques, Management and Applications, Vol. 1*, Longley P, Good-child M, Maguire D, Rhind D (eds). Wiley: London; pp 481–492.

Moore RJ, Jones DA, Cox DR, Isham VS. 2000. Design of the HYREX rain-gauge network. *Hydrol. Earth System Sciences* **4**: 523–530.

Myers D. 1982. Matrix formulation of co-kriging. *Mathematical Geology* **14**: 249–257.

Nespor V, Sevruk B. 1999. Estimation of wind-induced error of rainfall gauge measurements using a numerical solution. *J. Atmos. Oceanic Technol.* **16**: 450–464.

Oliver M. 2010. *Geostatistical Applications for Precision Agriculture*, First edn. Springer: Berlin; 169pp.

Paulat M, Frei C, Hagen M, Wernli H. 2008. A gridded dataset of hourly precipitation in Germany: Its construction, climatology and application. *Meteorologische Zeitschrift* **17**: 719–732.

Pebesma E. 2004. Multivariable geostatistics in s: the gstat package. *Computers and Geoscience* **30**: 683–691.

Pebesma E. 2012a. Gstat User's Manual. Available online at http://www.gstat.org/manual/.

Pebesma E. 2012b. Package gstat. Available online at http://cran.r-project.org/web/packages/gstat/.

R Development Core Team. 2012. *R: A language and environment for statistical computing, reference index version 2.2.14*. R Foundation for Statistical Computing: Vienna, Austria. ISBN 3-900051-07-0. Available at http://www.R-project.org.

Salek M, Novak P. 2008. 'Experience gained by five years of the utilization of the radar-based quantitative precipitation estimation at the Czech Hydrometeorological Institute'. In *Proceedings of ERAD, Helsinki, Finland*; pp 251–256.

Schmutz C, Arpagaus M, Clementi ML, Frei C, Fukutome S, Germann U, Liniger M, Schacher F. 2008. '*Meteorologische Ereignisanalyse des Hochwassers 8. bis 9. August 2007*, Arbeitsbericht der MeteoSchweiz Nr. 222, Swiss Federal Office of Meteorology and Climatology MeteoSwiss: Zürich; 30.

Schuurmans JM, Bierkens MFP, Pebesma EJ, Uilenhoet R. 2007. Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar. *J. Hydrometeorol.* **8**: 1204–1224.

Seo DJ, Breidenbach JP. 2002. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *J. Hydrometeorol.* **3**: 93–111.

Seo BC, Krajewski WF. 2011. Investigation of the scale-dependent variability of radar-rainfall and rain gauge error correlation. *Adv. Water Resources* **34**: 152–163.

Seo DJ, Krajewski WF, Bowles DS. 1990. Stochastic interpolation of rainfall data from rain gages and radar using cokriging. 1. Design of Experiments. *Water Resources Research* **26**: 469–477.

Sevruk B, Lapin M (eds). 1993. 'Precipitation measurement and quality control', *Proc., Int. Symp. on Precipitation and Evaporation, Vol. 1*. Slovak Hydrometeorological Institute: Bratislava, Slovakia.

Stein A, van Eijnsbergen AC, Barendregt LG. 1991. Cokriging nonstationary data. *Mathematical Geology* **23**: 703–719.

Velasco-Forero CA, Sempere-Torres D, Cassiraga EF, Jaime Gómez-Hernández J. 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Adv. Water Resources* **32**: 986–1002.

Ver Hoef J, Cressie N. 1993. Multivariable spatial prediction. *Mathematical Geology* **25**: 219–240.

Villarini G, Krajewski WF. 2010. Review of the different sources of uncertainty in radar-based estimates of rainfall. *Surv. Geophys.* **31**: 107–129.

Villarini G, Mandapaka PV, Krajewski WF, Moore RJ. 2008. Rainfall and sampling uncertainties: a rain gauge perspective. *J. Geophys. Res.* **113**(D11): 1–12.

Wackerly DD, Mendenhall W, Scheaffer RL. 2008. *Mathematical Statistics With Applications*. Cengage Learning.

Wüest M, Frei C, Altenhoff A, Hagen M, Litschi M, Shär C. 2009. A gridded hourly precipitation dataset for Switzerland using rain-gauge analysis and radar-based disaggregation. *Int. J. Climatol.* **30**: 1764–1775.

Willi M, Gabella M, Germann U, Frei C. 2011. '*CombiPrecip test cases*', Internal Report. MeteoSwiss: Zürich.

Woodcock F. 1980. Hannsen and Kuipers discriminant related to the utility of yes/no forecasts. *Mon. Weather Rev.* **109**: 172–173.

Zawadzki I. 1975. On radar–rain-gauge comparison. *J. Appl. Meteorol.* **14**: 1430–1436.

Zawadzki I. 1984. Factors affecting the precision of radar measurements of rain. In *22nd Conf. on Radar Meteorology, Zurich*. Amer. Meteorol. Soc.: Boston, MA; pp 251–256.

