Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

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

MeteoSwiss

Scientific Report MeteoSwiss No. 103

On the Man-Machine Supervision of Precipitation Ensemble Forecast

Alexandre Grandchamp, Lionel Moret & Daniel Cattani

ISSN: 1422-1381

Scientific Report MeteoSwiss No. 103

On the Man-Machine Supervision of Precipitation Ensemble Forecast

Alexandre Grandchamp, Lionel Moret & Daniel Cattani

Recommended citation:
Grandchamp A., Moret L. and Cattani D. : 2019, On the Man-Machine Supervision of Precipitation Ensemble Forecast, *Scientific Report MeteoSwiss*, **103**, 38 pp.

Editor:
Federal Office of Meteorology and Climatology
MeteoSwiss
Operation Center 1
CH-8058 Zürich-Airport
T +41 58 460 91 11
www.meteoswiss.ch

Abstract

The production of localized and accurate meteorological forecast is of primary importance for the prediction of natural hazards, the management of civil engineering project, the direction of aeronautic services or in the organization of outdoor events. The highly varying landscape over Switzerland together with the inherent complex dynamics of the atmosphere makes that task an extremely difficult goal in general, and especially at the scale of a few hours and of a few kilometers squared. The considerable development of the numerical models have led to the daily production of several detailed simulations delivering approximations for the future three-dimensional state of the troposphere over the country. Every three hours a complete forecast is available for the following 33 hours and at the scale of 1 km and, moreover, every twelve hours an ensemble of 21 complete forecasts is available for the following 120 hours at the scale of 2 km. The latter is used to estimate the corresponding prediction uncertainty from a reference simulation (unperturbed). Despite these significant achievements, numerical simulations are still unable to deliver reliable representations of small scale events and can be subject to systematic biases in specific meteorological situations. The human expertise has consequently still a crucial role to play in the analysis and the validation of model outputs to ensure the constant quality of daily forecasts and of specific warnings. This report aims then to discuss operational methods for that *Man-Machine supervision* in the context of two-dimensional precipitations fields. We describe how the ensemble of fine scale simulations can be corrected in order to provide local precipitation forecast which are consistent with the regional forecaster expertise while preserving the complex physically coherent patterns computed by the models. More precisely, the proposed multi-scale statistical framework is made of three main parts: first an upscaling step is applied for each region, then a quantile-mapping method trained on the model behaviour is used, and finally a downscaling step returns the corrected ensemble of two-dimensional precipitations fields. This method is then illustrated by the computation of local threshold fuzzy probabilities. These examples demonstrate that many spatial features of the original precipitation ensemble are preserved (such as orographic pattern) while the overall local event probabilities are modified according to the given forecaster values. We finally conclude by discussing potential improvements of the proposed methods using *super-ensembles* which would use simultaneously different available models, physical fields and climatic archives.

Contents

Contents

1 Introduction

1.1 Motivations

Over the past decades the developments in computer performances and in the refinement of the physical models have largely improved the overall meteorological forecast quality. However, the production of automated and accurate prediction at the scale of a few days and of a few kilometers remains extremely challenging (see section 1.2 for more detail). The numerical simulations are not only still unable to deliver reliable representations of small scale events, but are also inherently highly sensitive to modeling error (such as the physical parametrization or landscape discretization) which generally leads to low confidence forecast. The human expertise has consequently still a crucial role to play in the analysis and the validation of model outputs to ensure the constant quality of daily forecasts and of specific warnings. We are then facing the following situation. On the one hand, numerical simulations reproduce the overall physical behavior as well as the detailed complex phenomena on a relatively fine scale grid; however, they can exhibit unrealistic features and lead, in some cases, to inaccurate forecast. On the other hand, human forecaster have a meteorological expertise for each different regions which improves -at least on average- the forecast quality at a the scale of hundreds of kilometer squared and of a day; however, it is very difficult to adjust accordingly the detailed dynamics of the atmospheric variables in a physically coherent way. This report discusses the design of a numerical method to achieve that *Man-Machine supervision* in the context of precipitation fields. More precisely, we expose how the prescription of a regional and deterministic accumulation value can be used to build a corrected and local probability of precipitation quantities based on the small scale ensemble model COSMO-E.

1.2 Ensemble forecast from numerical models

Detailed weather forecasts are obtained through numerical simulations of the atmospheric dynamics over a three dimensional grid. These computations are not only very intensive but also have to deal with many subtle and crucial aspects, such as: initial and boundary conditions, physical parametrization, orography resolutions, surface type (such as forest, water area, cities, etc.) and unresolved small scale phenomena. All these various intrinsic uncertainties are commonly observed to have significant impacts on the meteorological forecast, even for integration times of a few days. The future state of the atmosphere can be -or should be- consequently considered as a statistical outcome among the ensemble of possible deterministic trajectories consistent with the known model uncertainties, as explained for instance by the European Center for Medium-Range Weather Forecasts (ECMWF): *So a*

probability forecast [...] explicitly reminds the user that there is always a forecast uncertainty which should be [...] taken into account when making any practical use of the forecast. In fact even "deterministic" forecasts are in reality probability forecasts in disguise (ECMWF web site). Even if the governing physical equations are somehow explicit, and even if only highly idealized situations are considered, it is extremely difficult to get closed form expressions which describe the statistical physics of the atmospheric dynamics, as discuss for instance in *Gleeson* (1970); *Majda and Wang* (2006). The fast evolution of the computing power over the last decades has nevertheless led to the development of several promising sampling techniques, see for instance *Leith* (1974); *Palmer et al.* (1993); *Anderson* (1996); *Marsigli et al.* (2005); *Doucet and Johansen* (2009); *van Leeuwen* (2010), which make the production of accurate probabilistic meteorological forecast a realistic goal.

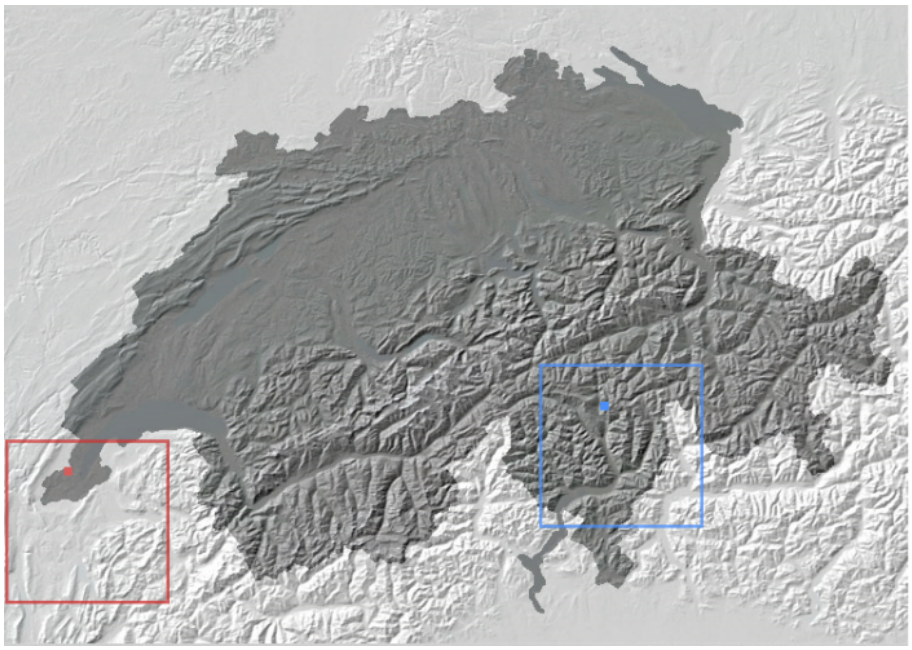
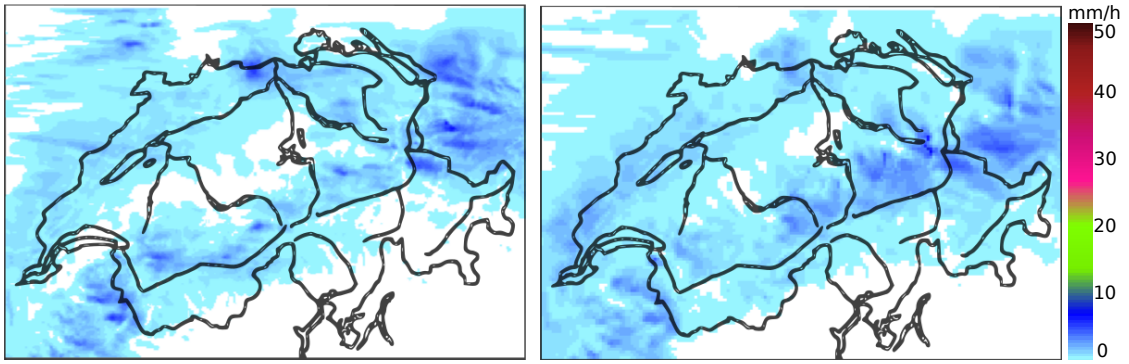


Figure 1: *Representation of the considered computational domain over Switzerland. The squares depict specific areas that are discussed in the present study (around Geneva in red and around Comprovasco in blue).*

The ECMWF run twice a day an Earth model using an horizontal grid of about 9 km of resolution, called *High Resolution Model*, which provides meteorological forecast for the ten following days. In addition, to take care of the multiple uncertainties, the *Ensemble model* is a set of 50 perturbed models that are computed on a sub-grid of 16 km to estimate the evolution of the atmospheric state distribution as a function of lead time with respect to an unperturbed model, called *control run*. Despite all that computational effort, the overall accuracy of local weather forecast remains relatively low over specific regions (such as the Alps for instance) in part due to the incapacity of the numerical models to resolve the necessary physical details.

The Consortium for Small-scale Modeling (COSMO) aims to improve and develop the numerical modeling at fine scale but for a limited area. Designed for the Switzerland area, the *COSMO-1* model has a resolution of about 1.1 km and runs every three hours to provide a deterministic forecast for the next 33 hours. The *COSMO-E* model is analogous to the global Ensemble model, as it delivers a distribution of likely meteorological states over the country, while it is made of a set of 21 trajectories on a 2.2 km mesh. It runs every twelve hours and has 120 hours of lead time. The COSMO-E grid



(a) **(b)**

Figure 2: *COSMO hourly precipitation fields (stratiform event), 02.03.2017,* $t = 4:00$ *UTC: (a) COSMO-1 (run 03:00) and (b) COSMO-E (control run 00:00)*

is exactly comprised in the COSMO-1 grid and has therefore half the number of mesh points. In this study we will consider the domain depicted[1] in figure 1, the square areas will be specifically studied and are containing the meteorological station GVE (red square) and COM (blue square). Figures 2a and 2b show respectively precipitation field outputs from COSMO-1 and COSMO-E (control run) in a stratiform event for the same forecast time. It corresponds to the passage of a night cold front. In that example, we observe the typical swiss orographic effects (over the Jura and the Alps for instance), as shown in figure 1, as well as a nice correspondence between the two forecast fields. For comparison, figures 3a and 3b show the hourly precipitation predicted by COSMO-1 and COSMO-E (control run) for a convective event. In the evening of that day, strong storm systems were expected over the Geneva region. The poorer correspondence in that case is explained by the fact that the lead time is larger and that convective situations are much more sensitive to uncertainties. More specifically, that control run is forecasting heavy precipitations (about 45 mm/h in average) to reach the Geneva airport between 16:00 and 17:00 UTC. Figure 4 illustrates the time dynamics of hourly precipitation predicted by the COSMO-E control run for a region being a square of 60km $\times$ 60km containing the station GVE (black square): that region is drawn with a red square in figure 1. Each of these models have its own strength and weakness, an optimal combination of their output is a promising way to enhance ensemble forecast over Switzerland.

[1] The present map has been taken from the following website www.ornitho.ch

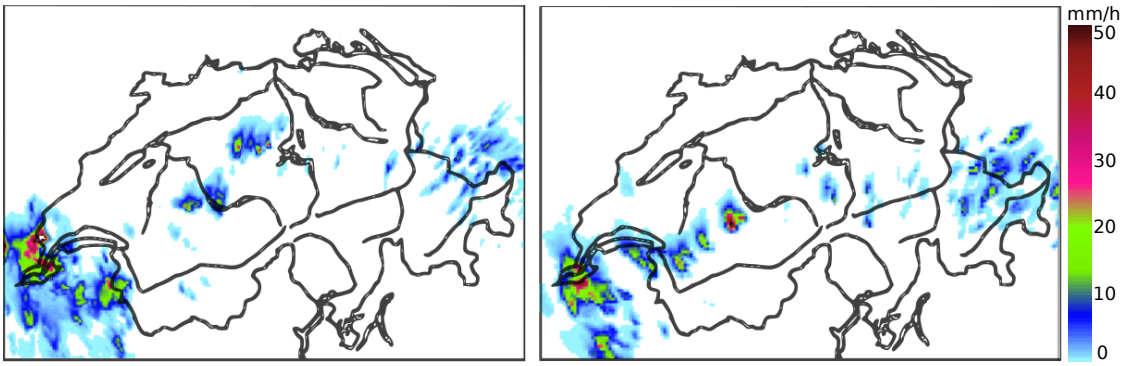


Figure 3: *COSMO hourly precipitation fields (convective event), 31.05.2017,* $t = 17:00$ *UTC. Panel (a) shows the COSMO-1 forecast (run 03:00 UTC) and panel (b) the COSMO-E forecast (control run 00:00 UTC)*

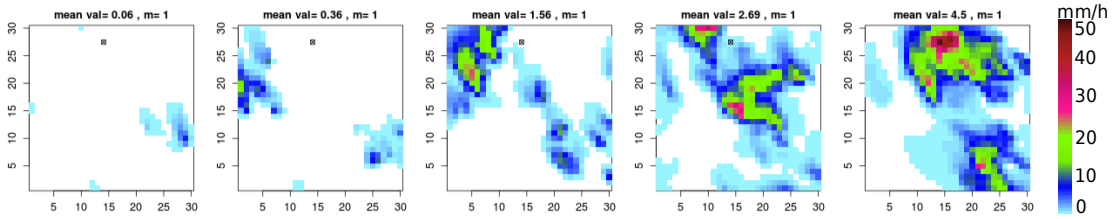


Figure 4: *Example of a hourly regional precipitation field output from COSMO-E . Panels from left to right corresponds respectively to the forecast fields for* $t = 13:00$ *,* $t = 14:00$*, ...,* $t = 17:00$ *UTC (31.05.2017, control run 00:00 UTC). The area is a* 60 *km* $\times$ 60 *km square (depicted in red in figure 1) and the black dot corresponds to the station GVE.*

1.3 Man-Machine perspectives

Interestingly, in many meteorological contexts, human expertise can still improve the numerical prediction accuracy at the scale of a region and several hours by correcting the corresponding model output values. More precisely, nowadays in the MeteoSwiss weather forecast centers, the forecaster team is editing -at least twice a day- regional forecast values for the next seven days (D+1, D+2, ..., D+7) for several time periods and regions. In that report we will consider the first five days. The first two days are divided into 6-hour periods while the last three days are single 24-hour periods, and, accordingly, the country is divided into 27 regions for the lead time D+1 and D+2, and only 11 regions for D+3, D+4 and D+5. The fine regions have been defined to reflect meteorological heterogeneity of Switzerland, mainly due to its strongly varying landscape. However, since the forecast uncertainty is growing for larger lead times, bigger region as well as longer time periods are used at scale of a few days. Figure 5 illustrates the improvements, using COMFORT scores from *Cattani et al.* (2016), on precipitation quantities, sunshine durations and temperature made by the MeteoSwiss forecaster team with respect to deterministic operational numerical models at a regional scale and averaged over all regions and all forecast day over the year 2016. It has to be understood that these relative scores provide a comparison between the raw data coming from the model and the forecaster values *assisted by the model*

data, and not the forecaster alone. In other words this table shows an average quantitative improvement obtained by a man-machine supervision at a regional scale. An enhancement of the precipitation forecast accuracy is specifically observed.

Forecast Time	D+1	D+2	D+3	D+4	D+5	D+6	D+7
Precipitation	3	2	3	4	5	6	5
Sunshine duration	4	2	1	2	2	2	2
Temperature. Min.	1	1	0	0	0	-1	0
Temperature. Max.	-1	0	0	-1	-1	0	1

Figure 5: *Average improvements using COMFORT scores made by the MeteoSwiss forecaster team with respect to the corresponding numerical model predictions.*

The aim of this report is to discuss how the forecaster insights can be explicitly used together with the numerical model outputs in a post-processing framework to improve the forecast quality and its uncertainty quantification, not only at the regionial level but also at a finer spatial scale. More precisely, we are mainly interested in the improvement of the precipitation field statistics and in the induced forecast uncertainty. Despite the importance for many purposes of an accurate forecast of precipitation locations and quantities, that statistical modeling process remains nowadays a challenging issue.

A brief introduction to the notation and key concepts to study ensemble models of precipitation forecast is done in section 2. Section 3 presents the overall discussion about the methods to modify the simulated precipitation field ensembles in order to satisfy a set of desired conditions. More precisely, we first present in section 3.1 how the distribution of regional precipitation values can be adjusted based on a given external value to match a prescribed median and standard deviation. Then, in section 3.2, we expose a downscaling method (in time and space) to transform consistently each precipitation field belonging to the modified ensemble in order to mimic as much as possible *true* model outputs. A complete process to correct precipitation ensembles according to large scale conditions, while keeping fine scale details, is consequently proposed. An application to fuzzy threshold analysis is done in section 4 to demonstrate the post-processing method. These examples illustrate the local difference in *forecast uncertainty* between the original and modified precipitation ensemble.

2 Precipitation Ensembles

As anticipated, this report discusses methods to modify regional precipitation field ensembles in order to increase the detailed forecast quality. We present in that section some minimal notations as well as the necessary statistical concepts that are needed to define suitable ensemble corrections.

A typical ensemble model precipitation output is of the form

$$\{q^{(m)}(x,t)\,|\,x\in\mathcal{X},\ t\in\mathcal{T}\ \text{and}\ m=1,2,...,M\} \tag{1}$$

where $x=(x_1,x_2)$ stands for some two-dimensional coordinates on the numerical grid $\mathcal{X}$, t denotes some lead time among the list of integration time $\mathcal{T}$, m is indexing the various trajectories belonging to the ensemble, and finally $q^{(m)}(x,t)$ is a corresponding amount of precipitation. For instance, for a COSMO-E simulation run over the domain depicted in figure 1 we have roughly $\mathcal{X}=\{1,2,..,360\}\times\{1,2,..,240\}$ and $M=21$ trajectories, whereas for the ECMWF ensemble model we have roughly $\mathcal{X}=\{1,2,..,90\}\times\{1,2,..,60\}$ and $M=51$. In this study the data that are hourly cumulated precipitation over the 5 next days, which gives the lead time set $\mathcal{T}=\{1,2,...,120\}$. An *ensemble* of the form of (1) refers then to an indexed set of time dependent two-dimensional fields. More generally, if $X\subset\mathcal{X}$ denotes a specific region and $T\subset\mathcal{T}$ denotes a specific time period, the ensemble of regional precipitation forecast over that time period is then of the form

$$\{q^{(m)}(x,t)\,|\,x\in X,\ t\in T\ \text{and}\ m=1,2,...,M\}. \tag{2}$$

At a coarser scale, the precipitation forecast can then, for each index m, either be regarded as *time period average*

$$q^{(m)}(x,T)=\frac{1}{|T|}\sum_{t\in T}q^{(m)}(x,t) \tag{3}$$

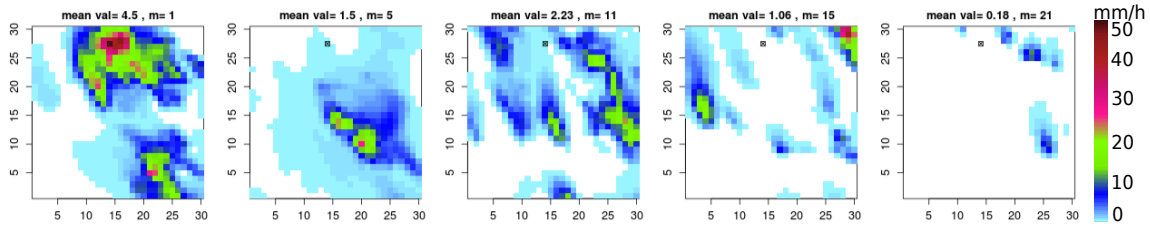
where $|T|$ denotes the period length of T, or as *regional average*

$$q^{(m)}(X,t)=\frac{1}{|X|}\sum_{x\in X}q^{(m)}(x,t) \tag{4}$$

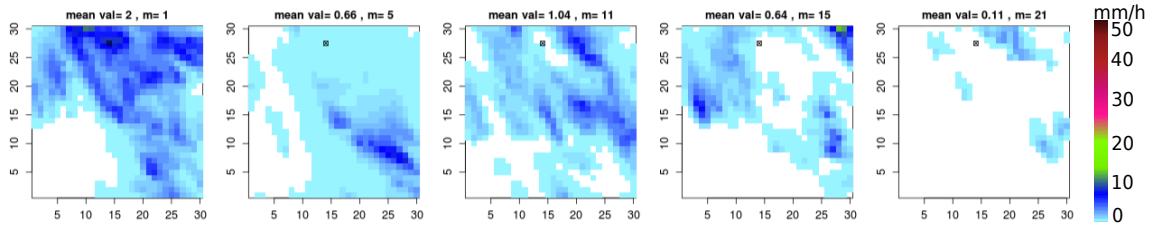
where $|X|$ stands for the region area X, or as *regional and time period average*

$$q^{(m)}(X,T)=\frac{1}{|X||T|}\sum_{x\in X,t\in T}q^{(m)}(x,t). \tag{5}$$

In order to appreciate the relative *forecast uncertainty* of the precipitation field predicted in the last panel of figure 4, we show in figure 6a the control run output (m=1) together with 4 other corresponding



(a)



(b)

Figure 6: *Example of COSMO-E regional ensemble for five selected members (31.05.2017, run 00:00): (a) hourly precipitation fields at 17:00 UTC and (b) six-hour mean precipitation fields (12:00 to 18:00 UTC). The area is a* 60 *km* $\times$ 60 *km square (depicted in red in figure 1) and the black dot corresponds to the station GVE.*

perturbed members (m= 5,11,16,21). For comparison, figure 6b shows the same regional data but where the precipitation in each pixel has been averaged over six consecutive hours. The global or regional ensemble (1) and (2) induce therefore, for each fixed time step $t \in \mathcal{T}$ or time period $T \subset \mathcal{T}$, an *image ensemble* made of two-dimensional fields. For any of the average quantities (3) to (5), and for a fixed point x or region X and a fixed time step t or period T, we introduce the *empirical cumulative distribution function* $F(\bar{q})$ which counts the relative number of instances lower than a given threshold value $\bar{q}$ and which is defined as

$$F(\bar{q}) = \frac{\#\{q^{(m)} \leq \bar{q}\}}{M}, \tag{6}$$

where $\#\{\,.\,\}$ stands for the number of elements in the set $\{\,.\,\}$. The value $1 - F(\bar{q})$ measures consequently the model uncertainty of observing a value beyond the threshold $\bar{q}$. Reciprocally, the *empirical quantile function*

$$F^{-1}(\alpha) = q^{[\alpha]} \tag{7}$$

returns the threshold value as a function of the relative number of instances, i.e. $F(q^{[\alpha]}) = \alpha$, and is obtained by linear interpolation of the closest available values. The *median value* is defined as $F^{-1}(0.5) = q^{[0.5]}$.

Alternatively, for each grid point x or region X, an ensemble model provides then M realizations of a time dependent *point-wise time series*, or *regional time series*. Figure 7 illustrates the difference between the hourly regional value $q^{(m)}(X,t)$ and a hourly point-wise value $q^{(m)}(x,t)$ of the ensemble run shown in figure 6a. For each time steps the quantiles $\alpha = 0, 0.2, 0.4, 0.6, 0.8, 1$ are computed using (6) and the corresponding interval are plotted. Each interval contains the same number of members by definition and, consequently, their relative length variations illustrate the overall member spread. Darker colors aim to emphasize higher relative density of the member spreads, where the word *density* is here referring to the ratio of the number of members per interval to the specific interval length. That representation has been chosen in order to provide a detailed graphical description of the generally

complex precipitation distribution in time and space. It emphasizes for instance the non-negativity of the data, their skewness or the possible sharp variations in time. The plots in figure 7 stress that the regional and point-wise precipitation values can have, in principle, very different behaviors. It can also be observed that there is not a high model confidence in heavy precipitation at the GVE station shown in the last panel of figure 4, which is a typical situation for such strong convective events.

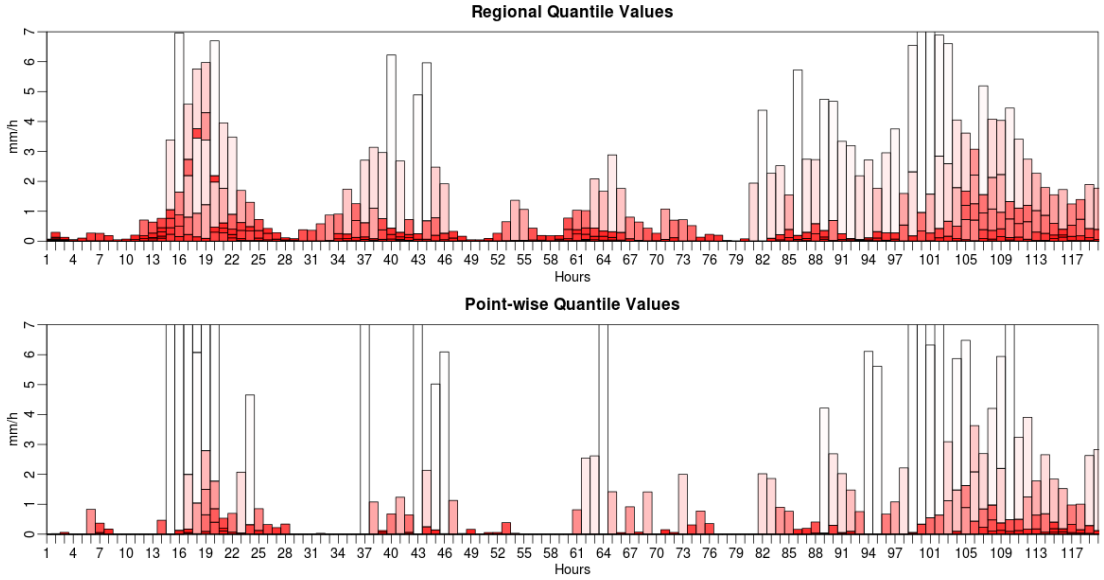


Figure 7: *Regional (60km $\times$ 60km) and point-wise (station GVE) precipitation quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ as model outputs (31.05.2017, run 00:00 UTC). The color gradient illustrates the density of COSMO-E members within each bar.*

For comparison, figures 8 and 9 illustrate respectively the statistics at the scale of one hour and six hours for a precipitation event with a relatively strong stratiform component predicted within the COSMO-E model. A South-Westerly flow is bringing a perturbed weather over a few consecutive days. In the southern part of the Alps, strong orographic precipitation are observed. Each period is represented by a bar centered around its starting time. The point coordinates corresponds to the station of COM and the region is again a square of 60km $\times$ 60km.

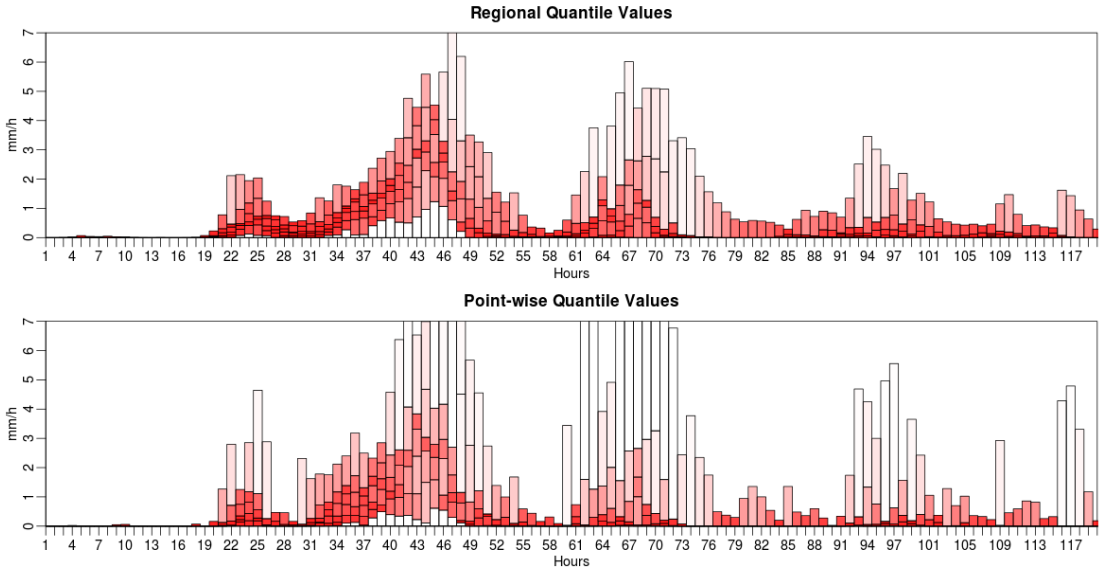


Figure 8: *Regional (60km $\times$ 60km) and point-wise (station COM) hourly precipitation quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ as model outputs (10.05.2017, run 00:00 UTC).*

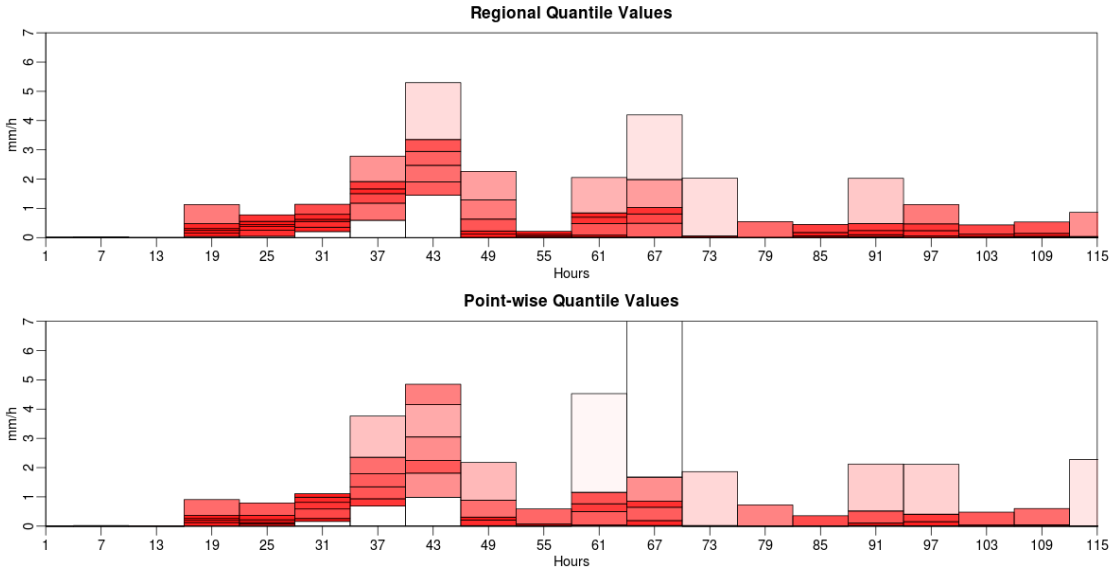


Figure 9: *Regional (60km $\times$ 60km) and point-wise (station COM) six hour precipitation (average value) quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ as model outputs (10.05.2017, run 00:00 UTC).*

Histograms and cumulative distributions for a fixed region and time period provide also an interesting complementary viewpoint. Figure 10 shows the plot corresponding to the six hour regional data in figure 9 for the time periods $T = 25 - 30$, $T = 31 - 36$, ..., $T = 49 - 54$. The arrows denotes the median value for each periods. Figure 11 shows the cumulative distribution plot corresponding to the histograms in figure 10. The circle denotes the value of the control run.

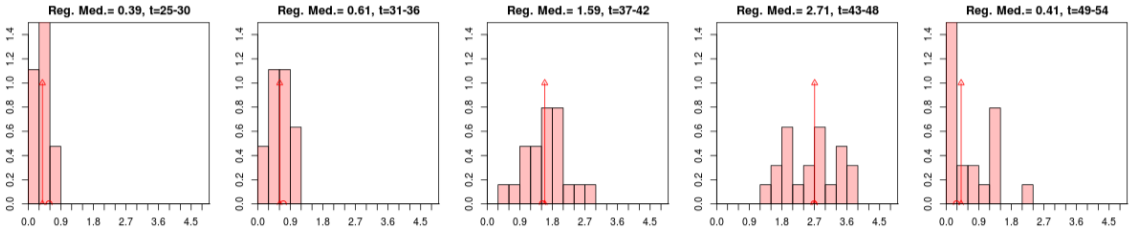


Figure 10: *Histogramms of the six-hour average regional (60km $\times$ 60km) precipitation values for the time periods $T = 25 - 30$, $T = 31 - 36$, ..., $T = 49 - 54$ (10.05.2017, run 00:00 UTC). The arrows shows the median value for each period.*

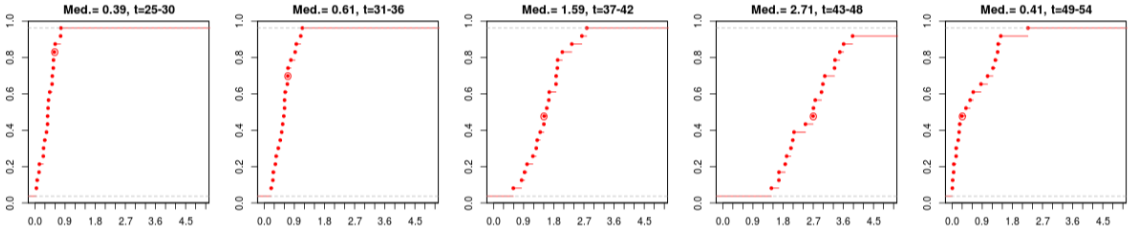


Figure 11: *Cumulative distribution of the six-hour average regional (60km $\times$ 60km) precipitation values for the time periods $T = 25 - 30$, $T = 31 - 36$, ..., $T = 49 - 54$ (10.05.2017, run 00:00 UTC). The circle denotes the value of the control run.*

3 Ensemble Correction

Motivated by the results of table 5 presented in section 1, we would like to asses whether the numerical model forecast accuracy, together with its uncertainty quantification, can be improved by using the human expertise provided by the forecasters. More precisely, we aim to define a method to correct each pixel value of each member in an ensemble model, like a COSMO-E run for instance, in order to obtain an modified ensemble which is, on one hand, consistent at a regional scale with the forecaster value and, on the other hand, which preserves the fine details as well as the uncertainty spread predicted by the numerical models.

A) The first step is to define what it means for an ensemble to be consistent with a regional value. The working hypothesis will be that *for each region and for each time period the forecaster value* $q_F(X,T)$ *equates the median value of the data* in the form (5), i.e. the median of the data averaged over the corresponding region and time period.

B) The second step is to correct pixel by pixel the precipitation ensemble to get high resolution fields satisfying the desired consistency constraint. The main challenge is here to preserve as much as possible the physical characteristics of the precipitation fields predicted by the numerical models. Among many specific features of precipitation fields, using radar imaging for instance *Pegram and Clothier* (2001); *Bowler et al.* (2006); *Niemi et al.* (2014); *Willeit et al.* (2015); *Nerini et al.* (2017), one could cite the following points: only non-negative values, strong spatial correlations, strong spatial heterogeneity, strong anisotropy, strong non-Gaussian distribution (in space and time) and strong non-linear dynamics. The statistical modelling of precipitation data requires then the use of innovative solutions since most of the classical hypotheses are broken (despite an extensive literature existing on time series analysis, filtering techniques and geostatistics: see for instance *Fieguth* (2010)). In that post-processing framework, the goal is to design a method which is moreover not extremely computationally expensive.

The overall post-processing scheme proposed in this report to meet conditions A) and B) is the depicted in figure 12. The first *upscaling* step is described in (5) and leads to a regional value ensemble $q^{(m)}(X,T)$. Then, a *quantile mapping* is applied to adjust the desired median value and to deliver a corrected ensemble of regional values $q_*^{(m)}(X,T)$. Finally, this set of corrected regional values is used in a *downscaling* step to build a ensemble of corrected precipitation fields $q_*^{(m)}(x,t)$. The correction method depends then on the implementation choices made both for the quantile mapping and for the upscaling and downscaling steps.

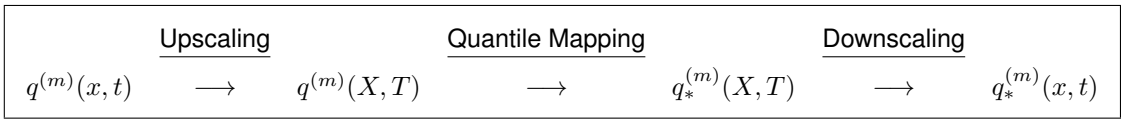


Figure 12: *Scheme of the precipitation ensemble correction*

3.1 Quantile mapping

Quantile mapping is commonly used to adjust statistical distributions with respect to an external piece of information, see for instance *Piani et al.* (2010); *Erdin et al.* (2012); *Veenhuis* (2013). We discuss in this section the definition of a transformation $q^{(m)}(X,T) \mapsto q_*^{(m)}(X,T)$ such that the *corrected ensemble* $q_*^{(m)}(X,T)$ has the property that its median value, denoted by $q_*^{[0.5]}(X,T)$, equates a given forecaster value q_{F}.

3.1.1 Method

A short discussion on scales has to be done before the development of the method. According to the example presented in section 2, the value of 1 mm/h seems to be a convenient scale to represent precipitation quantities and it will be then chosen as working unit in the following presentation. An important question is then the definition of dry events with respect to that scale. In this report we use a tolerance value of $tol = 0.01$ mm/h and any value less than *tol* is considered as vanishing.

As explained in section 1, the forecaster is editing the spatial mean value over a region and over a time period, which is *only* little information regarding the amount of parameters needed to characterize a precipitation ensemble. Consequently, it exists many ways to correct the ensemble values in order to fulfill the desired condition $q_*^{[0.5]}(X,T) = q_{\mathsf{F}}$, such as pure additive or pure multiplicative mapping for example. However, we also would like to produce a corrected ensemble which looks *realistic*, or in other words which *could have been a true model output*. As it can be observed in figure 10, the median values seem to be correlated to the spread of the histograms. In what follows we will use $q^{(m)}$ and σ instead of $q^{(m)}(X,T)$ and $\sigma(X,T)$ for a shortened notation. As an example figure 13 shows the log-linear fit between observed median values $q^{[0.5]}$ and observed ensemble standard deviations σ for the regional data in figure 9 and for several time periods. More precisely, for each fixed region and for each fixed time period a sliding window of half-size Δt hours is used to produce $(2\Delta t + 1)$ data points of the form $(q^{[0.5]}, \sigma)$ and then a least-square fit for a power law of the form

$$\sigma \sim C_1 \left(q^{[0.5]}\right)^{C_2} \tag{8}$$

is computed. In the example shown in figure 13 a value $\Delta t = 2$ has been chosen and 5 data points are then used to do the fit (8). The counterintuitive trend observed for the period $t = 43 - 48$ can be understood by looking at figure 8: the standard deviation is growing even if the median value is decreasing. The high quality of the fit (8) demonstrates that nor a pure additive nor pure multiplicative map can actually be a consistent modeling choice. Moreover, for each region and for each time period, the coefficients C_1 and C_2 can be used to extrapolate a *realistic* ensemble spread σ_* from a given median value.

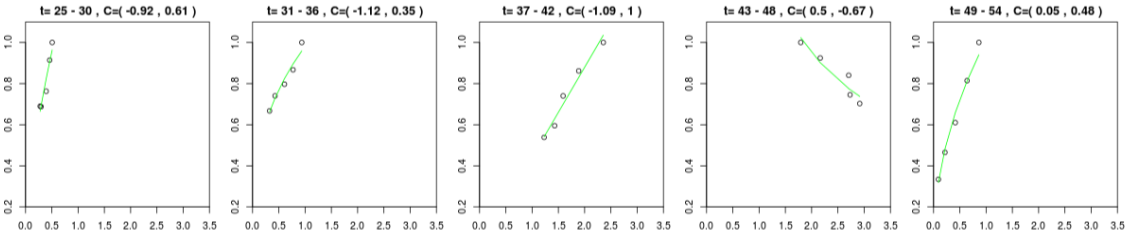


Figure 13: *Log-linear fit between ensemble median (horizonal axis) and standard deviation (vertical axis) for six-hour average regional (60km $\times$ 60km) precipitation values (10.05.2017, run 00:00 UTC). Data are drawn with circles and the fit with a continuum line. For a better view, the maximum standard deviation has been normalized within each plots.*

We therefore propose to the following **quantile mapping**:

$$q^{(m)} \mapsto q_*^{(m)} = \left(\frac{q^{(m)}}{q^{[0.5]}} \right)^{\frac{\lambda_*}{\lambda}} q_{\mathsf{F}} \tag{9a}$$

with

$$\lambda_* = \sqrt{\ln\left(1 + \sqrt{1 + \frac{4\sigma_*}{{q_{\mathsf{F}}}^2}}\right) - \ln(2)} \tag{9b}$$

$$\sigma_* = C_1 {q_{\mathsf{F}}}^{C_2} \tag{9c}$$

and where $q^{[0.5]}$ stands for the median value of the original ensemble $\{q^{(m)}(X,T)\}$, where λ denotes the standard deviation of the ensemble $\{\ln q^{(m)}(X,T)\}$, and the coefficients C_1 and C_2 are obtained through the fit in (8).

The formulas appearing in (9) can be understood as follows. First we shall consider the ensemble $\{\ln q^{(m)}\}$ instead of $\{q^{(m)}\}$ since their quantile values are the same[2] and the data set $\{\ln q^{(m)}\}$ can now be modified without the non-negativity constraint. The values $\{q^{(m)}\}$ close to zero are problematic and will be replaced by *tol* $= 0.01$ mm/h. Intermittency regime should consequently be treated separately. Then, a simple computation shows that for any real numbers $a > 0$ and b the ensemble

$$\left\{ \ln q_*^{(m)} = \frac{a}{\lambda} \left(\ln q^{(m)} - \ln q^{[0.5]} \right) + b \right\} \tag{10}$$

admits b as median values and a as standard deviation. The choice $b = \ln q_{\mathsf{F}}$ implies then that the median value of $q_*^{(m)}$ condition is exactly q_{F} for any $a > 0$. It remains then to choose a value for the standard deviation of the ensemble $\left\{ \ln q_*^{(m)} \right\}$ such that the standard deviation of $\left\{ q_*^{(m)} \right\}$ is given by σ_* in (9c). In order to achieve that condition, we make the assumption that the data $\left\{ q_*^{(m)} \right\}$ are nearly following a *log-normal distribution* and that leads[3] to the choice $a = \lambda_*$ defined in (9b). Expression (9a) is then deduced from (10). In other words, the quantile mapping (9) precisely enforces the desired median value, but only enforce approximatively the value of standard deviation σ_*: which is itself only an extrapolation given by the log-linear fit (8). We note however that the relative error on the value of σ_* caused by the use of λ_* is only of a few percents. The two limiting cases $q^{[0.5]} = 0$ and $q_{\mathsf{F}} = 0$ are singular and have to be discussed separately. In order to regularize the mapping (9) for these two cases the current strategy is simply to replace vanishing values by the tolerance *tol* $= 0.01$ mm/h. The

[2] which is true because $\ln()$ is a strictly increasing function.
[3] More precisely, formula (9a) is deduced from the analytical inversion of $\sigma_* = (\exp\{{\lambda_*}^2\} - 1) \exp\{q_{\mathsf{F}} + {\lambda_*}^2\}$.

proposed algorithm takes then the form

$$
\begin{cases}
\textbf{For } \text{a region } X \ \& \ \textbf{For } \text{a period } T \\
\textbf{For } m = 1, ..., M \\
\quad \textbf{Get } q^{(m)}(X,T) \text{ using (5)} \\
\textbf{Compute } \text{the coefficients } C_1 \text{ and } C_2 \text{ using (8)} \\
q^{[0.5]} = \max(q^{[0.5]}, tol) \\
q_{\mathsf{F}} = \max(q_{\mathsf{F}}, tol) \\
\textbf{For } m = 1, ..., M \\
\quad q^{(m)} = q^{(m)}(X,T) \\
\quad \textbf{Compute } q_*^{(m)} \text{ using (9)} \\
\quad\quad \textbf{If}(q_*^{(m)} < tol) \textbf{ then } q_*^{(m)} = 0 \\
\quad \textbf{Return } q_*^{(m)}(X,T) = q_*^{(m)}
\end{cases}
\tag{11}
$$

and leads then to an explicit quantile mapping for ensemble the regional value $\{q^{(m)}(X,T)\} \mapsto \{q_*^{(m)}(X,T)\}$. Figure 14 and 15 illustrate the result of that transformation on the ensemble depicted in figure 10 and 11. The original median $q^{[0.5]}$ and the new value $q_*^{[0.5]} = q_{\mathsf{F}}$ are provided in the subtitle for each panel. The last column exhibits the most striking case where $q^{[0.5]} = 0.41$ mm/h and $q_*^{[0.5]} = q_{\mathsf{F}} = 2$ mm/h for the time period $t = 49-54$. Log-normal approximations has been drawn using continuous lines for comparison in the cumulative distribution plots in figure 15. Figures 16 shows the regional value distribution drawn (first row) in figure 9 with its corresponding transformation according to the quantile mapping in (11). We observe in particular the deformation of the ensemble value corresponding to the period $T = 49 - 54$, as illustrated in figures 14 and 15. The gap appearing in figures 16 for the period $T = 68 - 54$ corresponds to a vanishing value from the forecaster, i.e. $q_{\mathsf{F}} = 0$.

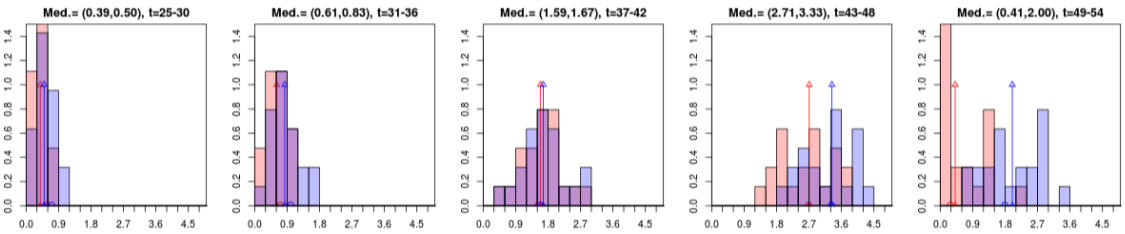


Figure 14: *Histogramms of the six-hour average regional (60km $\times$ 60km) precipitation values for the time periods $T = 25 - 30$, $T = 31 - 36$, ..., $T = 49 - 54$ (10.05.2017, run 00:00 UTC): original (red) and modified (blue) ensemble. The arrows shows the median value for each time period.*

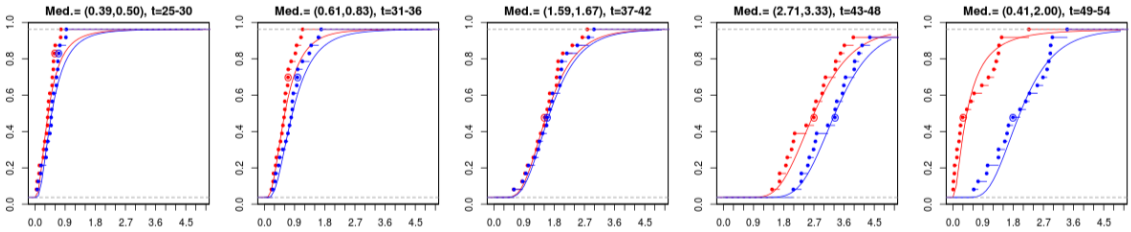


Figure 15: *Cumulative distribution of the six-hour average regional precipitation values for the time periods $T = 25 - 30$, $T = 31 - 36$, ..., $T = 49 - 54$ (10.05.2017, run 00:00 UTC): original (red) and modified (blue) ensemble. The circles denote the value of the control run. The continuous lines shows the corresponding log-normal distribution.*

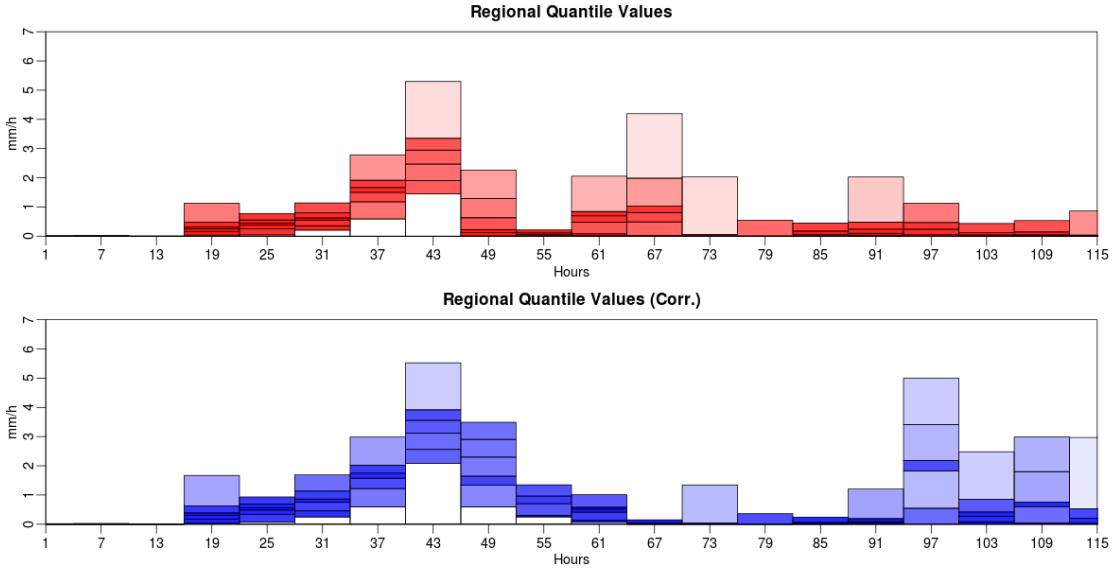


Figure 16: *Regional (60km $\times$ 60km) original (red) and modified (blue) six-hour average precipitation quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ (10.05.2017, run 00:00 UTC). The color gradient illustrates the density of members within each bar.*

3.1.2 Remarks

i . It is unclear if the log-normal distribution provides the *best* modelling assumption for the values $q_*^{(m)}$ and many other choices can, in principle, be considered *Alexandersson* (1985); *Shoukri et al.* (1988); *Piani et al.* (2010); *Reath et al.* (2016); *Wenger* (2016). However, the log-normal distribution has the property that its median and standard deviation can be explicitly expressed, and in a relatively simple way, throughout its shape parameters. A generalization of that case is given by the use of the *Box-Cox power transform* instead of the $\ln()$ map but no explicit expressions analogous to (9b) seems to be available and a lower bound constraint needs to be used, see for instance *Erdin et al.* (2012).

ii . The log-linear fit to provide a consistent standard deviation for the modified ensembles is used in order to reflect the underlying *model uncertainty* specific to the ensemble run under consideration. It would be interesting to analyze the stability (or not) of the fitted coef C1 and C2 with respect to time and domain size. An interesting alternative would be to deliver a new standard deviation to maximize the *model reliability* with respect to the actual probability of an event. This option is promising but requires an analysis which goes beyond this report.

iii . Quantile mapping methods are designed to modify a value distribution under some constraints. In the context of operational forecasting, its effect on extreme values (tail) has to be assessed. In the example presented in figure 15 the corrected extreme values are higher than the original ones, however the opposite behaviour has equally be observed in other studied cases. More precisely, according to modelling equations 9, if the prescribed median is significantly higher than the original value the corrected distribution will have more probable extreme values (heavier tail), and reciprocally if the prescribed median is significantly lower than the original value. The modelling quality of the corrected distribution with respect to meteorological archives remains however to be assessed.

iv . It would be interesting to modify the proposed method to take care of the extreme cases which are going from $q^{[0.5]} = 0$ to $q_{\mathsf{F}} > 0$ and from $q^{[0.5]} > 0$ to $q_{\mathsf{F}} = 0$ in a more satisfactory way. The current algorithm (11) basically expands the noisy (close to vanishing) values in the first case and, conversely, shrinks the overall ensemble values in the second case. The use of regional meteorological archives could be a way to improve the uncertainty quantification, as well as the probabilistic reliability, in that post-processing framework.

v . The method has been illustrated on square region for simplicity. However, it has to be emphasized that this is **not** a limiting assumption since any arguments and formulas hold for any region geometry.

3.2 Space and Time Downscaling

It has been exposed in section 3.1 how to get a modified ensemble of precipitation data averaged over a region and over a time period $\{q_*^{(m)}(X,T)\}$. We present in this section a method to modify each values

in a precipitation fields of the form $\{q^{(m)}(x,t) \,|\, x \in X,\, t \in T\}$ to get a new field $\{q_*^{(m)}(x,t) \,|\, x \in X,\, t \in T\}$ which satisfies that its space and time upscaling (5) equates the modified regional mean $q_*^{(m)}(X,T)$ for each fixed member m, region X and period T. Both the time and space fine scale corrections presented here have strong connexions to commonly used method in the study of precipitation fields, see *Niemi et al.* (2014); *Nerini et al.* (2017) for instance.

3.2.1 Method

The proposed method has two steps and is applied for each member: first, a time downscaling is performed at the regional scale and, then, a space downscaling is applied to modify each pixel values. Figure 17 depicts these detailed steps in an analogous way to figure 12. Similarly to the quantile mapping, we aim that the transformed precipitation fields $q^{(m)}(x,t)$ look *realistic*, i.e. do not exhibits non-physical features such as orography and meteorological situation inconsistency.

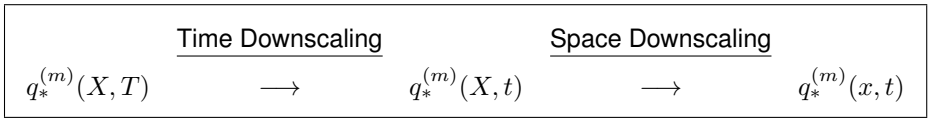


Figure 17: *Scheme of the downscaling processes*

For each m, the modified hourly regional values $\{q_*^{(m)}(X,t) \,|\, t \in T\}$ are defined by

$$q_*^{(m)}(X,t) = \frac{q_*^{(m)}(X,T)}{q^{(m)}(X,T)} q^{(m)}(X,t) \text{ for all } t \in T \tag{12}$$

where $q^{(m)}(X,T)$ and $q_*^{(m)}(X,T)$ stand respectively for the original and modified regional values over the period T. The time dependent values $q_*^{(m)}(X,t)$ have the property that their time average (3) is the desired value $q_*^{(m)}(X,T)$. According to the examples we have studied so far, it does not seem that the multiplicative transformation (12) can induce inconsistent results. The only issue is when $q^{(m)}(X,T) = 0$: currently, since no other external information is added to that ensemble correction process, we define $q_*^{(m)}(X,t) = q_*^{(m)}(X,T)$ when $q^{(m)}(X,T)$ is vanishing. Figure 18 shows the application of the time downscaling to the data in figure 16.

Concerning the space downscaling, a simple multiplicative transformation is unlikely to give physically consistent results, as it can be observed in figures 6a for instance. In analogy to the observed correlation between the ensemble median value and standard deviation presented in figure 13 when discussing the quantile mapping, the correlation between the spatial regional mean $q^{(m)}(X,t)$ and the spatial regional standard deviation $\Sigma^{(m)}(X,t)$ of hourly data has been studied. In what follows we will then use the notations $q(x) = q^{(m)}(x,t)$, $q(X) = q^{(m)}(X,t)$ and $\Sigma(X) = \Sigma^{(m)}(X,t)$ to have a more synthetic presentation. A similar fitting to (8) is proposed to model the correlation between the spatial mean and standard deviation over the region X, i.e.

$$\Sigma(X) \sim c_1 q(X)^{c_2} \tag{13}$$

for some coefficients $c_1 > 0$ and c_2. More precisely we look, for each members, at the $(2\Delta + 1)$ neighboring values obtained using a time sliding window of half-size Δt. Figure 19 shows the obtained

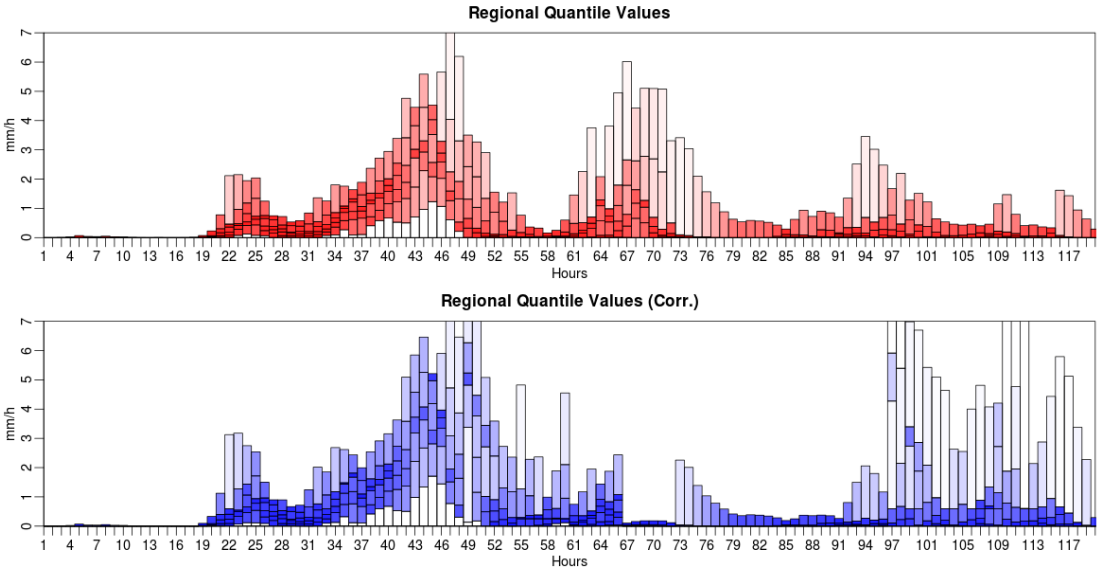


Figure 18: *Regional (60km $\times$ 60km) original (red) and modified (blue) hourly precipitation quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ (10.05.2017, run 00:00 UTC).*

approximations. Even if the fitting quality is sometimes less good than in figure 13 the coefficients c_1 and c_2 can be used to extrapolate the numerical trend between the spatial spatial mean and standard deviation.

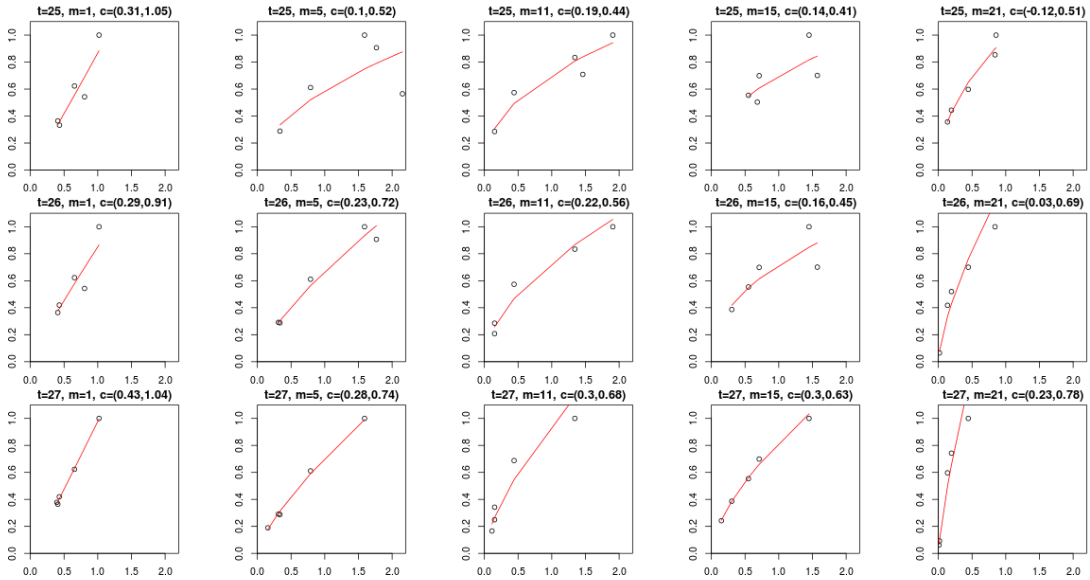


Figure 19: *Log-linear fit between spatial mean (horizontal axes) and spatial standard deviation (vertical axes) for hourly regional precipitation values, $t = 25, 26, 27$ and $m = 1, 5, 11, 15, 21$ (31.05.2017, run 00:00 UTC). For a better view, the maximum standard deviation has been normalized within each plots.*

The modification of precipitation fields constrained to a new regional mean $q_*(X)$ and a new spatial regional standard deviation $\Sigma_*(X)$ is discussed for instance in *Niemi et al.* (2014); *Nerini et al.* (2017) in the context of statistical analysis of precipitation properties. The simplest approach is to define the

modified field $q_*(x)$ by

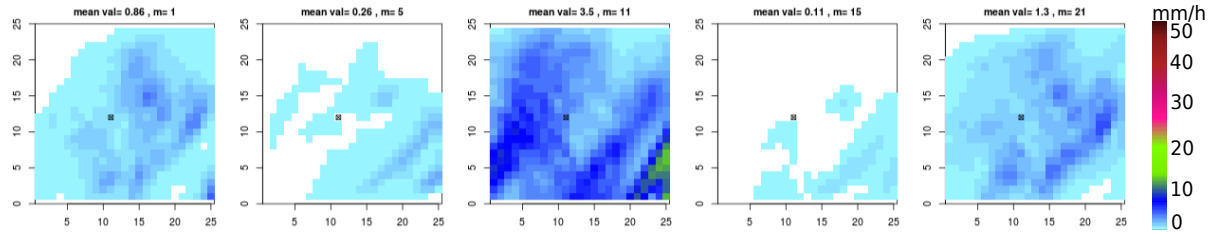
$$q_*(x) = \frac{\Sigma_*(X)}{\Sigma(X)} [q(x) - q(X)] + q_*(X) \text{ for all } x \in X \tag{14}$$

which can be refereed as the *shift-and-scale method*. The case $\Sigma(X) = 0$ only appears for fully dry fields, the formula (14) is then formally replaced by $q_*(x) = q_*(X)$ in that case. We would like to stress that (14) **is not** the formula proposed in the *Data4Web* manual *Matter* (2016) to do the spatial downscaling. The main issue with the shift-and-scale method is to ensure the non-negativity of the precipitation field $q_*(x)$ at any point $x \in X$. The strategy proposed in *Niemi et al.* (2014); *Nerini et al.* (2017) is to apply (14), then to replace all negative values by 0's and to iterate that procedure a few times, which results in a non-negative field whose mean and standard deviation are approximately $q_*(X)$ and $\Sigma_*(X)$. In our context, we are mainly interested to match the hourly regional mean value $q_*(X,t)$ obtained through the quantile mapping (11) and time downscaling (12), whereas the standard deviation modification is only a modeling decision made in order to get fields as consistent as possible with model outputs. We consequently propose the following algorithm for **space and time downscaling**

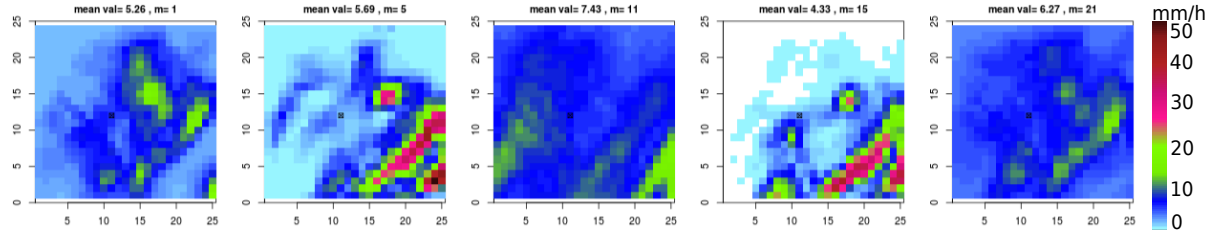
Compute $q_*^{(m)}(X,T)$ for $m = 1, ..., M$ using (11)
For $m = 1, ..., M$ & **For** $t \in T$
  Compute $q_*^{(m)}(X,t)$ using (12)
  $q(X) = q^{(m)}(X,t)$
  $q_*(X) = q_*^{(m)}(X,t)$
  $\Sigma(X) = \Sigma^{(m)}(X,t)$
  Compute the coefficients c_1 and c_2 using (13)
  $\Sigma_*(X) = c_1 {q_*}^{c_2}$
  For $x \in X$
    Compute $q_*(x)$ using (14)
  While$(\min q_*(x) < 0)$
    $\Sigma_*(X) = 0.95 * \Sigma_*(X)$
    For $x \in X$
      Compute $q_*(x)$ using (14)
  Return $q_*^{(m)}(x,t) = q_*(x)$ (15)

which basically adjusts for each precipitation field the standard deviation value obtained by the log-linear fit until the non-negativity constraint is satisfied. We note that only a few iterations are necessary in practice. The main difference with the shift-and-scale method application in *Niemi et al.* (2014); *Nerini et al.* (2017) is the fact that the spatial standard deviation is a non-linear function of the spatial mean value in that case. Figures 20a and 20b illustrate the application of the algorithm (15) on 5 regional hourly precipitation fields: the control run and 4 other members. We observe that each of these field is transformed specifically and that the large modifications are induced by the relatively large difference between the original and modified quantile distributions for the time period $T = 49 - 54$, as depicted in figure 15 for instance. However, many spatial features are conserved through the ensemble correction

such as for example the different behaviors in the North-West and South-East corner. Note that the South-West to North-East orientation of the precipitation isolines in the South-East corner corresponds with the position of the Valle Mesoclina. More remarkably, dry areas are also conserved even if the overall regional value is increased. For comparison figures 21a and 21b show the six-hour average field for the period $T = 49 - 55$. Figures 22 and 23 illustrate the induced time dependent distribution for the average precipitation values at the station COM for hourly and six hour period respectively. The green line shows the corresponding measured value at that station. For comparison we also show in figures 24a and 24b an example where the the forecaster value q_F is lower than the one originally predicted by the model. Interestingly, the regional value is decreased not only by changing the non-vanishing pixels but also by adding new dry areas.



(a)



(b)

Figure 20: *Example of hourly regional ensemble correction, control run and members* $5, 11, 15$ *and* $21, t = 49$ *UTC (10.05.2017, run 00:00): (a) original fields and (b) modified fields. The area is a* 60 *km* $\times$ 60 *km square and the black dot corresponds to the station COM.*

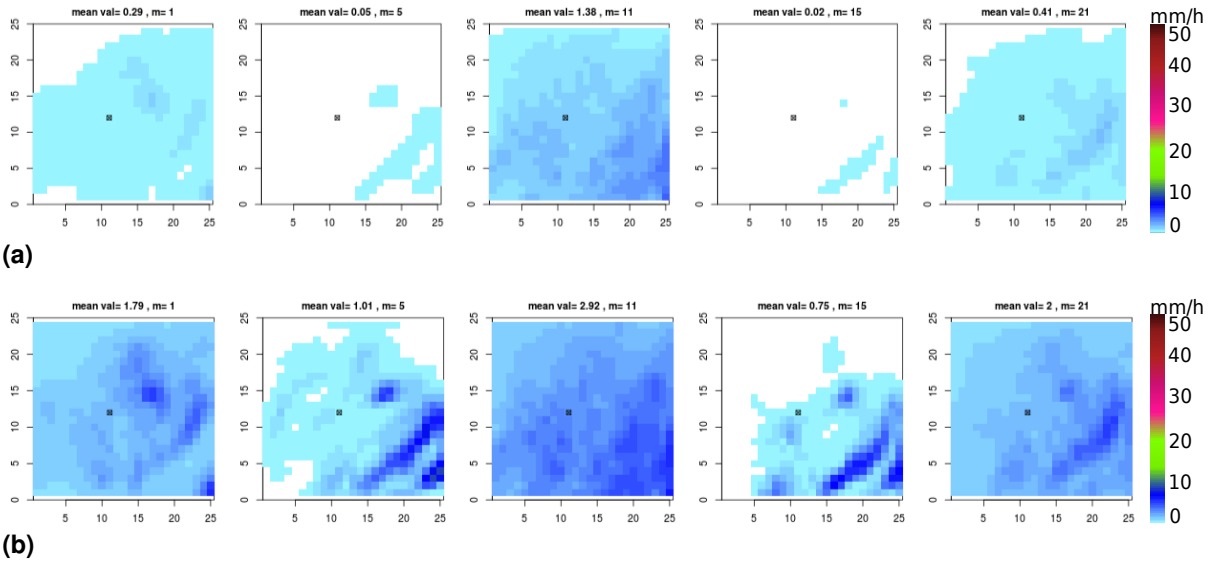


Figure 21: *Example of six-hour average regional ensemble correction, control run and members* $5, 11, 15$ *and* 21*,* $T = 49 - 55$ *UTC (10.05.2017, run 00:00): (a) original fields and (b) modified fields. The area is a* 60 *km* $\times$ 60 *km square and the black dot corresponds to the station COM.*

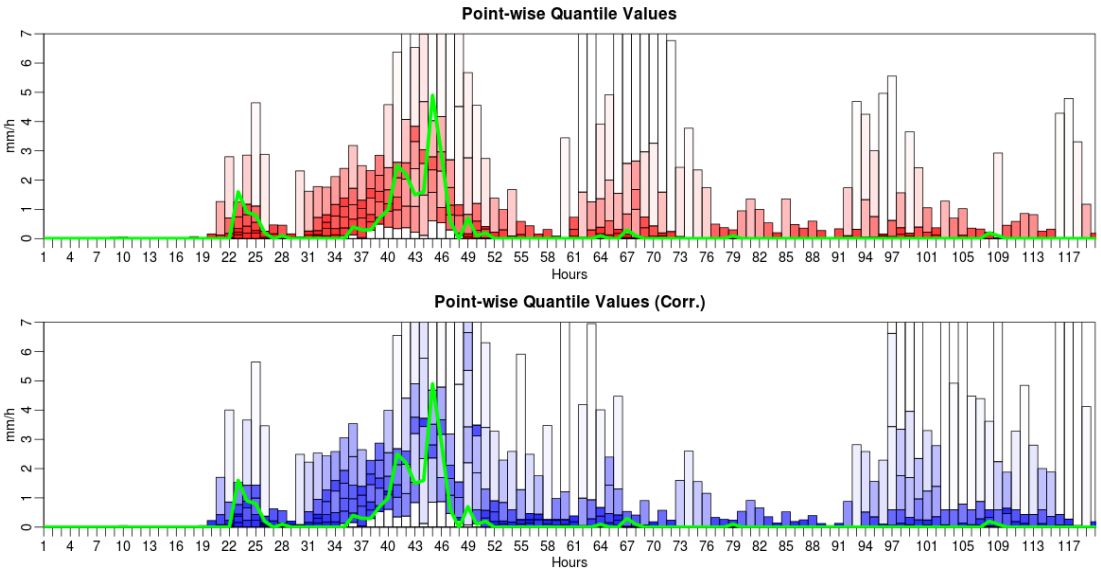


Figure 22: *Point-wise (station COM) original (red) and modified (blue) hourly precipitation quantiles* $\alpha = 0, 0.2, ..., 0.8, 1$ *(10.05.2017, run 00:00 UTC). The green line shows the corresponding measured value.*

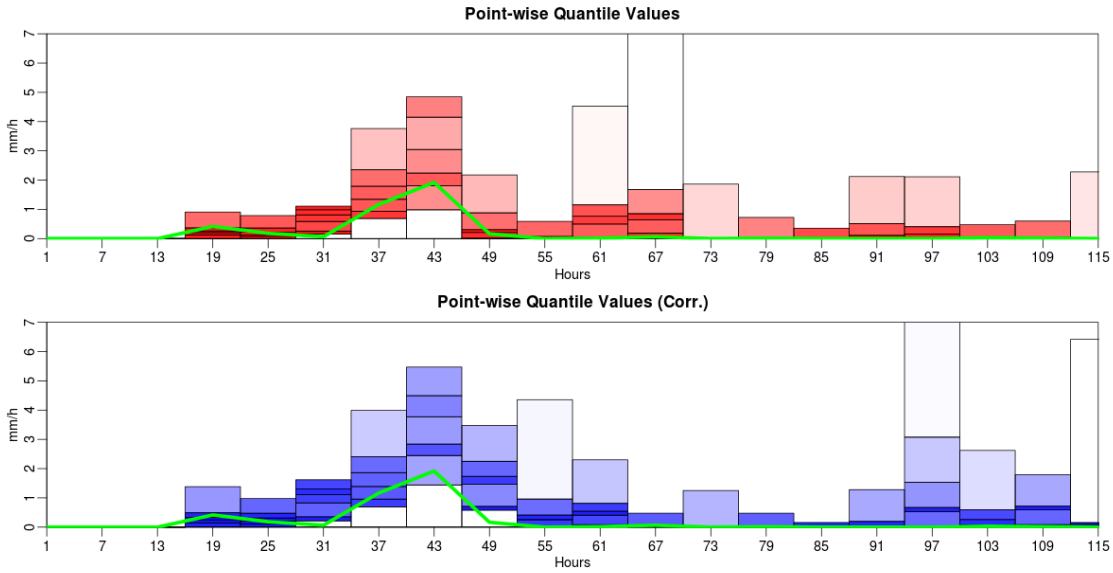


Figure 23: *Point-wise (station COM) original (red) and modified (blue) six-hour average precipitation quantiles $\alpha = 0, 0.2, ..., 0.8, 1$ (10.05.2017, run 00:00 UTC). The green line shows the corresponding measured value.*

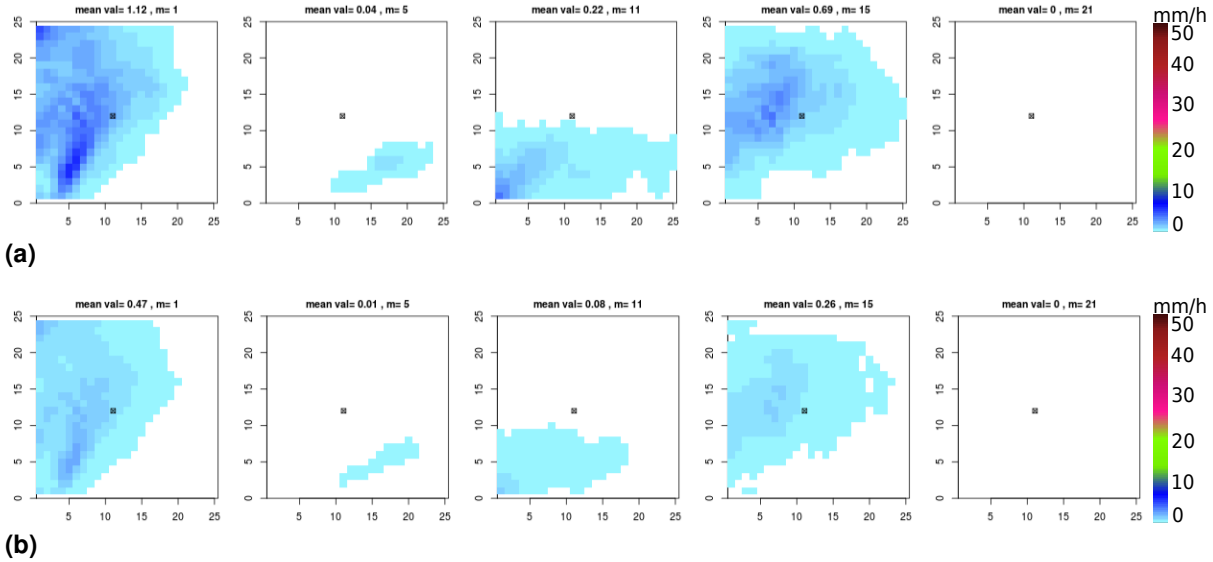


Figure 24: *Example of hourly regional ensemble correction, control run and members $5, 11, 15$ and 21, $t = 94$ UTC (10.05.2017, run 00:00): (a) original fields and (b) modified fields. The area is a 60 km $\times$ 60 km square and the black dot corresponds to the station COM.*

3.2.2 Remarks

- The proposed method to get modified precipitation fields uses log-linear fit of the spatial standard deviation in order to keep as much as possible the physical features predicted by the numerical model. A promising way to improve that point and to produce realistic fields with a given spatial mean value is to use *analogue situations data* based on meteorological archives. A very interesting related work has been done by I. Sideris at MeteoSwiss based on radar precipitation data[4].
- An other approach would be to respect at least observed spatial statistical features and generate *stochastic precipitation fields* with the correct regional value $q_*^{(m)}(X,T)$ and with realistic regional spatial correlations[5] *Nerini et al.* (2017), based on meteorological archives for instance.
- As it has been noticed, the downscaling formulas (12) and (14) **differ** from the corresponding ones in *Data4Web*. These choices are closer to commonly used methods in the field of radar precipitation analysis *Niemi et al.* (2014); *Nerini et al.* (2017). The proposed expressions provide then also alternative deterministic formulas that can be used in the operational setting, to unbiased COSMO-1 for instance. More precisely in *Data4Web* the regional value are also modified in time but the method uses additive corrections instead of multiplicative. Then, regarding the space distribution of the value, the algorithm is purely multiplicative (no spatial variance modification) but uses individual mesh point altitude to build non-vanishing field on originally dry area. See *Matter* (2016) for more detail.
- The time downscaling formula (12) can be improved for instance by using as well the past and future forecaster values, which would provide a smoother time evolution of the modified ensemble values and therefore avoid discontinuities like in figure 18.
- An important remaining question concerns the validation of the proposed method and whether the model uncertainty quantification and/or whether the probabilistic reliability are improved.

[4] Private communication: Ioannis.Sideris@meteoswiss.ch
[5] For more detail: Daniele.Nerini@meteoswiss.ch and Loris.Foresti@meteoswiss.ch

4 Application to Threshold Analysis

Ensemble models are designed to quantify numerical forecast uncertainty for non-linear systems. A direct and interesting application is *threshold analysis* which estimates the percentage of members which overcome a given value or, conversely, estimates the quantile values for a given certainty level. This analysis can be done over many length scales (pixel, sub-region, etc.) and many time scales (hour, a couple of hours, etc.). It is generally observed that coarser scales leads to less sharp but more reliable conclusions, see for instance in *Ebert* (2008), which can be explained by the intrinsic difficulty for numerical models to predict precipitation events at a specific mesh point and specific time step. For each mesh point $x \in X$ and for each time period T we then propose to extend the empirical distributive distribution $F(\bar{q})$ in (6) to

$$F(\bar{q}; U(x), \omega) = \frac{\#\{q^{(m)}(x', T_\tau) \leq \bar{q} \,|\, x' \in U(x), \tau \in \omega\}}{M|U(x)||\omega|}. \tag{16}$$

which counts the relative number of precipitation values (averaged on the period T) which are lower than the threshold value $\bar{q}$, but on a *spatial neighborhood* $U(x)$ of the point of interest x and a *time neighborhood* $\{T_\tau = T + \tau \,|\, \tau \in \omega\}$ of the period of interest T. The time window is typically chosen as a symmetric interval $\omega =]-\Delta t, \Delta t[$. The space and time location of an event is then relaxed with that definition. Reciprocally, the inverse map

$$F^{-1}(\alpha; U(x), \omega) = q^{[\alpha]} \tag{17}$$

returns the corresponding quantile value for the certainty threshold α. The use of the functions (16) and (17) is here called *fuzzy threshold analysis*. In order to illustrate the ensemble correction method presented in section 3.2, we will use as neighborhood $U(x)$ a square of 10 km $\times$ 10 km centered at the mesh point x, which is a 5 pixels $\times$ 5 pixels square for the COSMO-E grid, and a time window $\omega =]-1, 1[$ hours. In other words, an event is occurring *at* x if it occurs at roughly 5 km from x and within one hour in the past or the future. Many other more subtle -and probably more meaningful- definitions of the set $U(x)$ and ω can be made, but we believe that these choices are sufficient to discuss the use of fuzzy threshold analysis in that report. Figure 25 shows the application to the data appearing in figure 22 of the corresponding fuzzy empirical distribution (16). We observe that time dependent distribution features are slightly smoothed, while the main trends regarding the median and standard deviation values are kept. Figure 26 illustrates the relative forecast probability obtained using that threshold fuzzy analysis to have a six-hour cumulated value within the interval $0-1$ mm/6h, $1-3$ mm/6h, $3-5$ mm/6h, $5-10$ mm/6h, or greater than 10 mm/6h for the fields drawn in figures 21a and 21b. In that example, the median regional value has been strongly increased, as shown in figure 15 for the time period $T = 49 - 54$. We observe that the application of the ensemble correction method

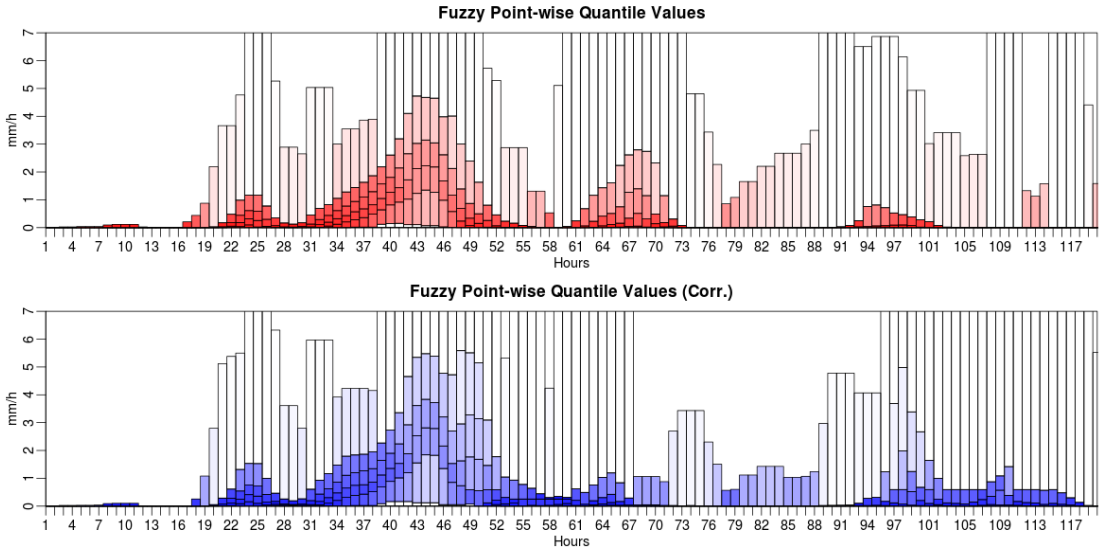


Figure 25: *Example of fuzzy point-wise (station COM) original (red) and modified (blue) hourly precipitation quantiles* $\alpha = 0, 0.2, ..., 0.8, 1$ *(10.05.2017, run 00:00 UTC).*

induces globally higher certainty of precipitation value above 5 mm/6h, whereas the original model output is predicting mainly values less than that threshold value. However, the downscaling algorithm (15) is shown to conserve many interesting heterogeneous features in space: such as the North-West and South-East difference, as well the location of the Valle Mesoclina, discussed in figure 21a and 21b. Conversely, the function (17) can be used to compute for each mesh point the quantile values

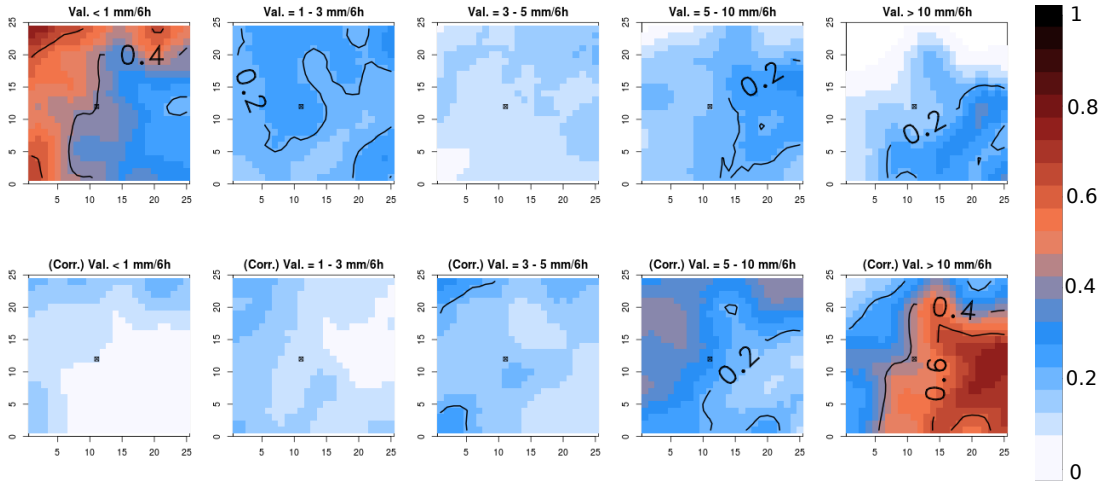


Figure 26: *Example of fuzzy threshold analysis for the original ensemble (top row) and the corrected ensemble (bottom row),* $T = 49 - 54$ *(10.05.2017, run 00:00 UTC). The colors illustrate the probability to have a value in the corresponding range. The contour lines denote the probability value* $0.2, 0.4, 0.6$ *and* 0.8*. The black square corresponds to the station COM.*

corresponding to a given range of probability. These values corresponds to threshold under which the required model certainty is reached. Figure 27 shows the obtained results for the ensembles used in figure 26. We observe for instance that the six-hour precipitation value over the station COM is expected to be within 0.5 mm/6h and 15 mm/6h with a probability of 70% for the original ensemble

(see top row in figure 27), whereas it is expected to be within 4 mm/6h and 25 mm/6h for the modified ensemble and for the same certainty level (see bottom row in figure 27).

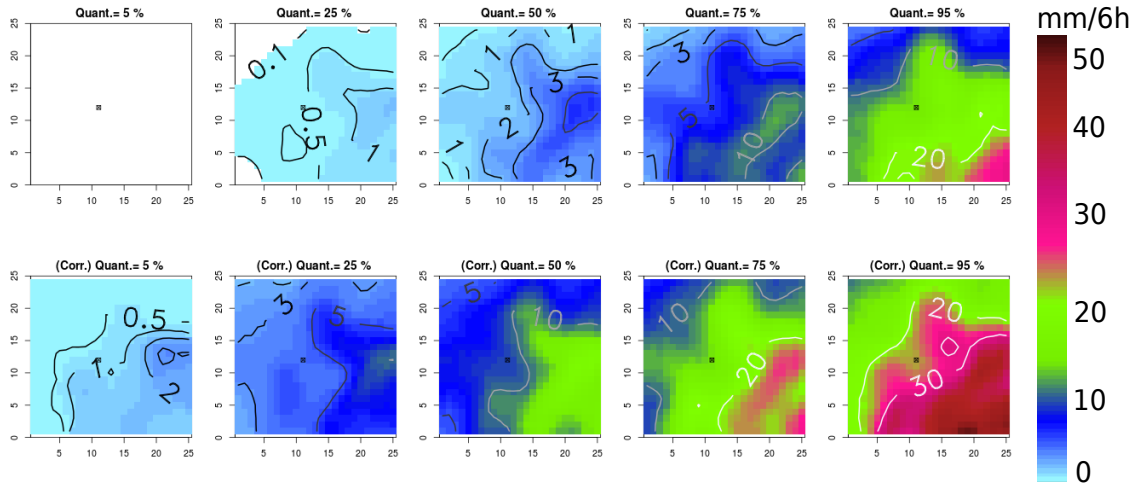


Figure 27: *Example of fuzzy threshold analysis for the original ensemble (top row) and the corrected ensemble (bottom row),* $T = 49 - 54$ *(10.05.2017, run 00:00 UTC). The colors illustrate the corresponding quantile value. The contour lines denote the value* $0.1, 0.5, 1, 2, 3, 5, 10, 20, 30$ *and* 50 *mm/6h. The black square corresponds to the station COM.*

To conclude that section on fuzzy threshold analysis, we briefly expose the case of the heavy convective precipitations over the station GVE presented in figures 4, 6a and 6b. Figure 28 illustrate the original and modified quantile distributions. Non-vanishing median values have been set by the forecaster for time period $T = 13 - 18$ and $T = 19 - 24$. Figure 29 shows the fuzzy threshold analysis for the period $T = 13 - 18$ for original and modified ensemble data. We also observe in that example that the proposed ensemble correction method changes the overall regional value distribution while it keeps many of the spatial features of the original precipitation ensemble. The GVE station is for instance close to two sub-regions in the modified ensemble where more than 10 mm/6h is expected with a certainty of about 50%, which only has a certainty of about 30% in the original ensemble model. Alternatively, according to figure 30, the modified ensemble predicts precipitation values less than 20 mm/6h with a certainty level of 75% (see bottom row in figure 30) for the GVE station, while the original ensemble predicts value less than 10 mm/6h for the same certainty level (see top row in figure 30).

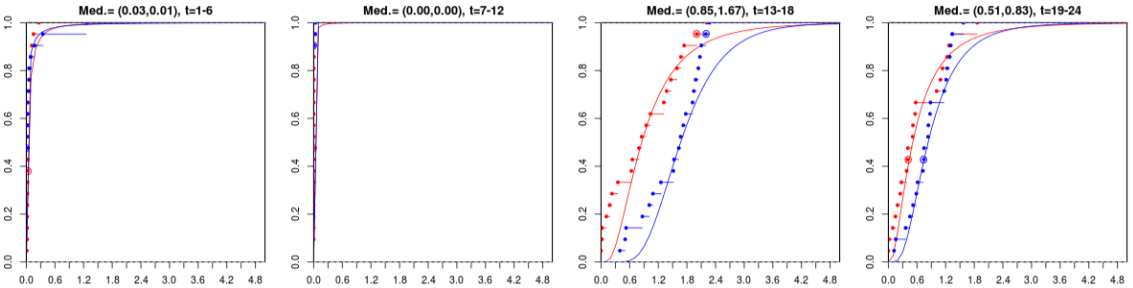


Figure 28: *Cumulative distribution of the six-hour average regional (Station GVE) precipitation values (31.05.2017, run 00:00 UTC): original (red) and modified (blue) ensemble. The circles denote the value of the control run.*

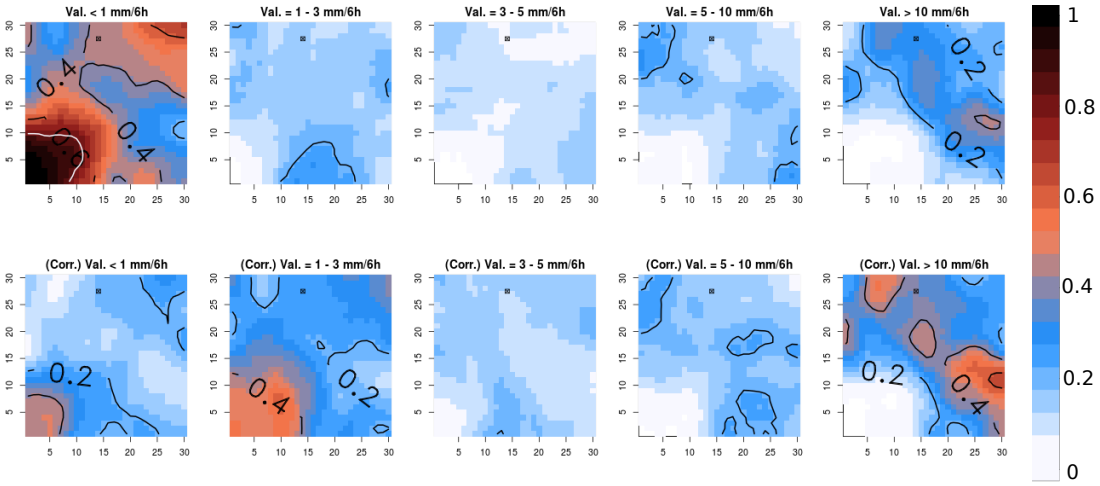


Figure 29: *Example of fuzzy threshold analysis for the original ensemble (top row) and the corrected ensemble (bottom row),* $T = 13 - 18$ *(31.05.2017, run 00:00 UTC). The colors illustrate the probability to have a value in the corresponding range. The contour lines denote the probability value* $0.2, 0.4, 0.6$ *and* 0.8*. The black square corresponds to the station GVE.*

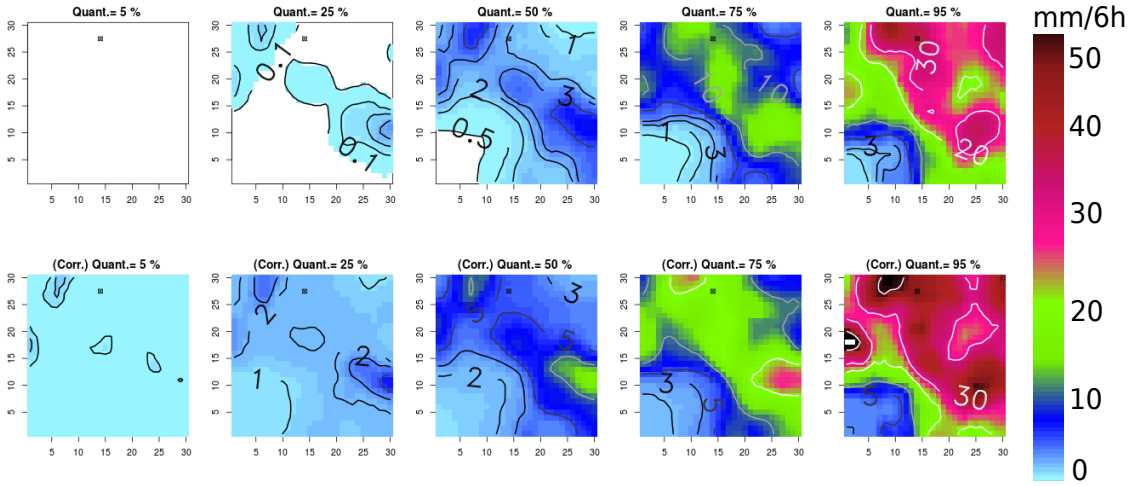


Figure 30: *Example of fuzzy threshold analysis for the original ensemble (top row) and the corrected ensemble (bottom row),* $T = 13 - 18$ *(31.05.2017, run 00:00 UTC). The colors illustrate the corresponding quantile value. The contour lines denote the value* $0.1, 0.5, 1, 2, 3, 5, 10, 20, 30$ *and* 50 *mm/6h. The black square corresponds to the station GVE.*

5 Conclusion and Future Improvements

Ensemble model provide not only detailed precipitation forecast both in time and space but also some measure of the corresponding uncertainty. That powerful and expensive tool is however subject to specific biaises, due to unmodeled phenomena or unresolved small scale events, that the human expertise can compensate (at least in average). The aim of that report was then to present a post-processing framework, together with an explicit method, in which the feasibility of that man-machine supervision is demonstrated. As discussed in detail in section 3, the proposed method is first made of upscaling step followed by a quantile mapping which enforces the median of the regional ensemble values. Then a downscaling step is applied which modifies the precipitation values at each mesh point and for each member. Two important modeling hypotheses concern the explicit log-linear fits between the ensemble median and standard deviation value in one hand, and between the spatial mean and standard deviation in the other hand. They are assumed in order to mimic model outputs as much as possible. This method leads then to the definition of a modified ensemble made of detailed precipitation field whose statistics are consistent with regional forecaster value. An application to fuzzy threshold analysis is then presented in section 4 and it suggests that many of the spatial features of the original precipitation ensemble are preserved. In order to improve the accuracy of the uncertainty quantification the space and time location definition of an event is relaxed to a neighborhood of the point and of the period of interest. This process leads to smoother distributions both in space and time, while preserving important characteristics regarding model uncertainties.

A proper quantitative analysis is needed to assess whether the proposed man-machine supervision framework can improve the (statistical) reliability of local precipitation forecast. A first step would be to try to compare the observed values and the outputs of the proposed method for an ensemble of representative stations and representative meteorological events. That task rises many complications, such as: the miss-match between numerical mesh-points (every 2 km) and the real station locations, the inherent measure errors at the stations (CombiPrecip modified data could help here for instance), the fluctuation of the forecaster reliability or the constant evolution of the meteorological models. Despite these difficulties, there is various reasons to believe that such a man-machine approach can participate in the improvement of the overall forecast quality. On the one hand, numerical models provide detailed simulations of both space and time precipitation distribution based not only on fundamental physical principles but also on global (averaged) a posteriori verifications. However for some given meteorological event, i.e. some specific meteorological situations and some specific area, the model predictions are proved to be inaccurate (see the example presented in figure 4). On the other hand, the human forecaster has been shown to be able to improve the model output reliability based on its own expertise of specific meteorological events, at least on short time scale and for regions averaged values (see table in figure 5). A rigorous and quantitative assessment remains then to be done in order

to validate the proposed method.

On the theoretical side, substantial improvements can be done at least in two places. Firstly, since several different simulations are available every day, the size of the original ensemble should be increased by considering together different models as well as consecutive runs: multiple outputs from European Ensemble, European High-Resolution, COSMO-E and COSMO-1 could be grouped to form a *super-ensemble*. The obvious advantage of that strategy is the creation of an ensemble which contains simultaneously long range forecast -but with a coarse spatial resolution- and fine detailed simulations -but with short lead time. This super-ensemble corresponds then truly to set of data that are available to the forecaster and which are used to make a quantitative prediction. One the top of these data one should of course add the set of climatological archives as well as his or her meteorological expertise. Figure 31 illustrates that situation.

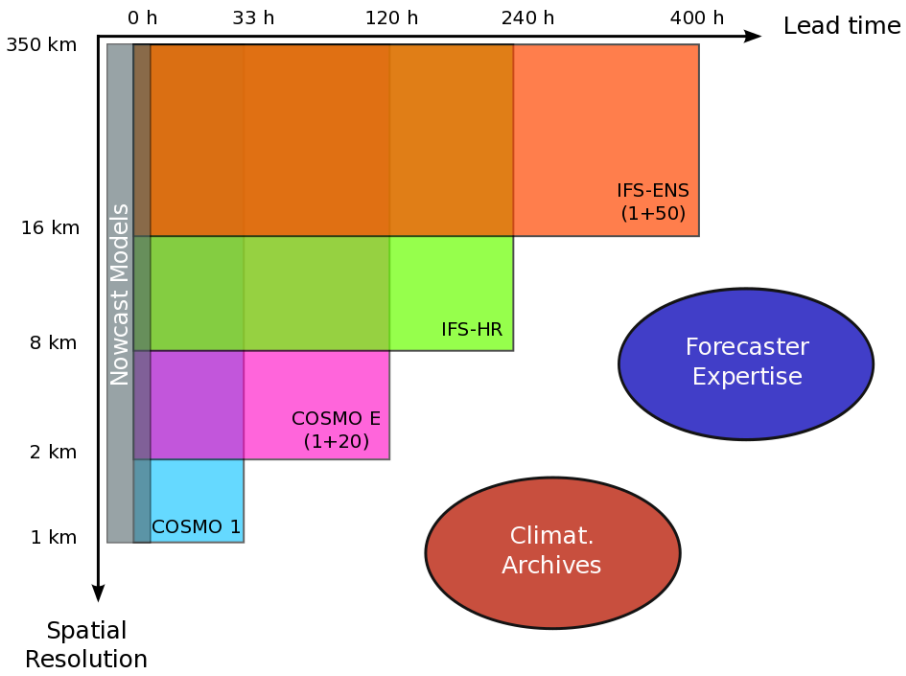
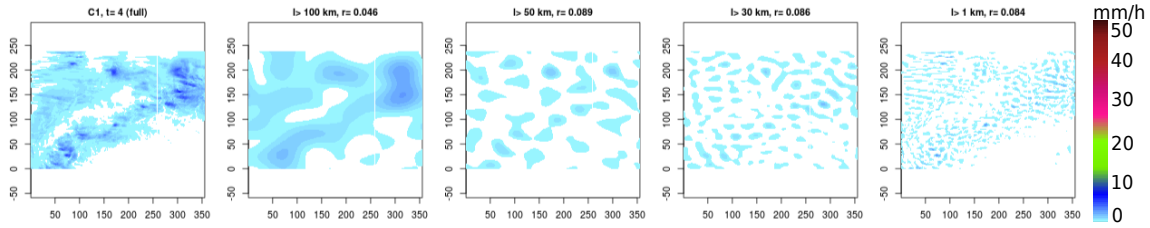


Figure 31: *Sketch of a multi-model approach*

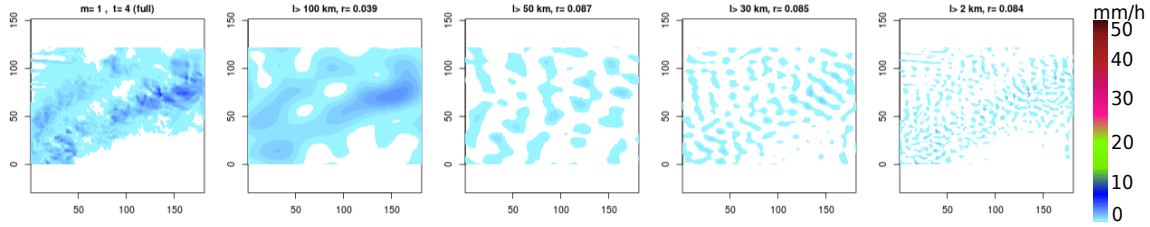
We are however facing the important technical difficulty which is that the data are somehow inconsistent: different number of mesh points (see figures 2a and 2b), different spatial resolution of an event (see figure 3a and 3b), different initialization time, etc. A promising way to solve that issue is the use of a spatial spectral decomposition *Harris et al.* (2001); *Ahrens* (2003); *Paschalis et al.* (2013); *Niemi et al.* (2014); *Nerini et al.* (2017) of the time dependent precipitation field ensemble to obtain a hierarchy of the spatial mode statistics. Very interesting related studies are done using two dimensional Fourier decompositions to develop multi-resolution radar based nowcast model, see for instance *Seed et al.* (2013); *Foresti et al.* (2016), or to bridge these nowcast models with COSMO-1 for instance [6]. Such spectral decompositions of spatial data have the property that each class of modes (ordered by a characteristic length scale) can be compared irrespective to the model they come from. Figures 32a and 32b show respectively that decomposition for the stratiform precipitation events predicted by the

[6] Private communication: Daniele.Nerini@meteoswiss.ch and Loris.Foresti@meteoswiss.ch

COSMO-1 model in figure 2a and by the COSMO-E model in figure 2b. In each row the right panel corresponds to the full and original field. Then from right to left the mode characteristic length scale l (chosen as the half wave length) are: $l > 100$ km, $100 > l > 50$ km, $50 > l > 30$ km, $30 > l > 10$ km and $30 > l > l_{\text{min}}$, where $l_{\text{min}} = 1$ km for COSMO-1 and $l_{\text{min}} = 2$ km for COSMO-E . The exact same procedure is also applied to the convective event in figures 3a and 3b, and the results are shown in figures 33a and 33b. As expected, in the stratiform case, most of the precipitation values are encoded through the large scale modes (second panels from the right), whereas in the convective case a strong signal is also present in the very fine scales (left panels), which is explained by the specific physical structure of the convective cells. This precipitation signature has been quantified by the scaling exponent of the corresponding *Fourier power spectrum* *Niemi et al.* (2014); *Willeit et al.* (2015); *Nerini et al.* (2017). In the frequency space any available precipitation fields (different observations, models or runs) can then be easily compared according to a given lead time and given characteristic length scales. That transformation leads then to the definition a super-ensemble by the application of any kind of mixing method (statistical weight, multi-scale nudging, etc.).



(a)

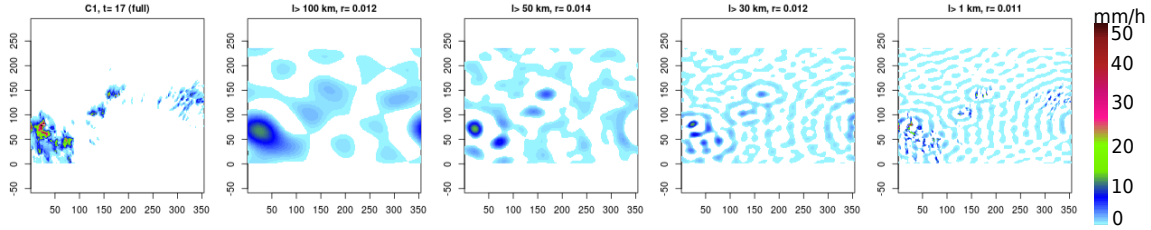


(b)

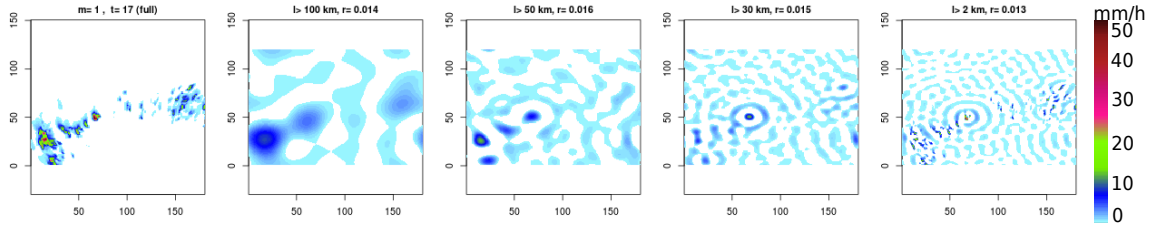
Figure 32: *Illustration of the Fourier decomposition of a precipitation field (02.03.2017, 4:00 UTC): (a) COSMO-1 (run 03:00) and (b) COSMO-E (run 00:00). From right to left the mode characteristic length scale l (half wave length) are: $l > 100$ km, $100 > l > 50$ km, $50 > l > 30$ km, $30 > l > 10$ km and $30 > l > l_{min}$, where $l_{min} = 1$ km for COSMO-1 and $l_{min} = 2$ km for COSMO-E .*

Secondly, the modeling choices made all along the development of the method should be tested against observed data (corresponding CombiPrecip fields for instance) to conclude whether the *model reliability* has been increased and if a better definition of the modified ensemble distribution is required. In particular it seems clear that the only use of the forecaster value is very restrictive and cannot lead generally to the definition of realistic statistics. A interesting framework would then be to add computer based knowledge, learned from models outputs or meteorological archives, to build physically consistent precipitation fields and time dependent distributions as a function of the forecaster inputs. The detailed spatial and time dependent features would be handle automatically (in a deterministic or stochastic setting) whereas the forecaster could still edit the field values at a regional scale. More generally, one could also use several other forecaster values such as thunderstorm switch, minimum and maximum daily temperature and sun durations, and their corresponding model outputs, to obtain

realistic meteorological state ensemble. As a conclusion, a regional man-machine supervision method



(a)



(b)

Figure 33: *Illustration of the Fourier decomposition of a precipitation field (31.05.2017, 17:00 UTC): (a) COSMO-1 (run 03:00) and (b) COSMO-E (run 00:00). From right to left the mode characteristic length scale l (half wave length) are: $l > 100$ km, $100 > l > 50$ km, $50 > l > 30$ km, $30 > l > 10$ km and $30 > l > l_{min}$, where $l_{min} = 1$ km for COSMO-1 and $l_{min} = 2$ km for COSMO-E .*

has been presented for the statistical description of precipitation events predicted by numerical ensemble models and an application to fuzzy threshold analysis has been discussed. The proposed method *only* uses the forecaster mean regional values over given time periods and try to deliver corrected ensemble which are not only consistent with that sequence of values, but also with model outputs. Several improvements can be considered by using more available information. Figure 34 illustrates a possible improved post-processing scheme. First, the spectral decomposition of the various model outputs would be used to build a super-ensemble of numerically comparable atmospheric states over the Switzerland domain. The relative weights between the different models can typically be tuned using machine learning type algorithms. The obtained field statistics could be then be corrected by region according to climatic archives and to the forecaster expertise, using for instance methods similar to the one presented in that report. Finally, that Man-Machine supervised super-ensemble would be used to forecast local observable, such as threshold probability.

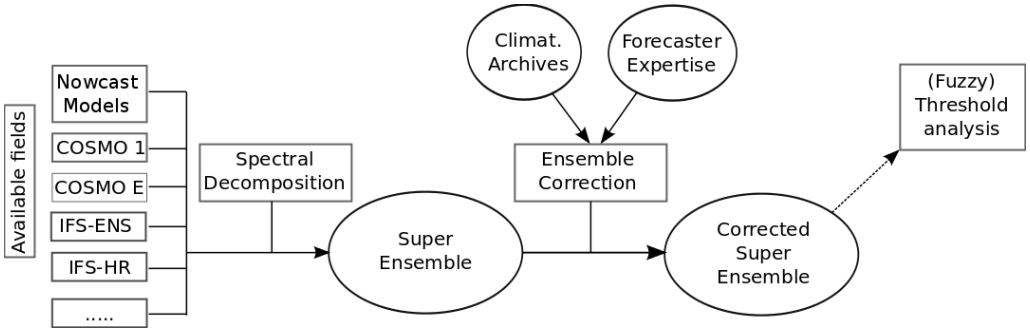


Figure 34: *Sketch of an improved super-ensemble correction scheme*

List of Figures

References

Ahrens, B. (2003), Rainfall downscaling in an alpine watershed applying a multiresolution approach, *Journal of Geophysical Research: Atmospheres*, *108*(D8).

Alexandersson, H. (1985), A simple stochastic model of the precipitation process, *Journal of climate and applied meteorology*, *24*(12), 1285–1295.

Anderson, J. L. (1996), A method for producing and evaluating probabilistic forecasts from ensemble model integrations, *Journal of Climate*, *9*(7), 1518–1530.

Bowler, N. E., C. E. Pierce, and A. W. Seed (2006), Steps: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP, *Quarterly Journal of the Royal Meteorological Society*, *132*(620), 2127–2155.

Cattani, D., A. Faes, M. Giroud, and M. Matter (2016), COMFORT: Continuous MeteoSwiss Forecast Quality Score, *Météosuisse*.

Doucet, A., and A. M. Johansen (2009), A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of nonlinear filtering*, *12*(656-704), 3.

Ebert, E. E. (2008), Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework, *Meteorological applications*, *15*(1), 51–64.

ECMWF web site (2017), European Center for Medium-Range Weather Forecasts (ECMWF), modeling and prediction, `http://www.ecmwf.int/en/research/modeling-and-prediction`.

Erdin, R., C. Frei, and H. R. Künsch (2012), Data transformation and uncertainty in geostatistical combination of radar and rain gauges, *Journal of Hydrometeorology*, *13*(4), 1332–1346.

Fieguth, P. (2010), *Statistical image processing and multidimensional modeling*, Springer Science & Business Media.

Foresti, L., M. Reyniers, A. Seed, and L. Delobbe (2016), Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in belgium, *Hydrol. Earth Syst. Sci*, *20*, 505–527.

Gleeson, T. A. (1970), Statistical-dynamical predictions, *Journal of Applied Meteorology*, *9*(3), 333–344.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit (2001), Multiscale statistical properties of a high-resolution precipitation forecast, *Journal of Hydrometeorology*, *2*(4), 406–418.

Leith, C. (1974), Theoretical skill of monte carlo forecasts, *Monthly Weather Review*, *102*(6), 409–418.

Majda, A., and X. Wang (2006), *Nonlinear dynamics and statistical theories for basic geophysical flows*, Cambridge University Press.

Marsigli, C., F. Boccanera, A. Montani, and T. Paccagnella (2005), The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlinear Processes in Geophysics*, *12*(4), 527–536.

Matter, M. (2016), Data4Web, guide de l'utilisateur (version 1.2), *Météosuisse*.

Nerini, D., N. Besic, I. Sideris, U. Germann, and L. Foresti (2017), A non-stationary stochastic ensemble generator for radar rainfall fields based on the short-space fourier transform, *Hydrology and Earth System Sciences*, *21*(6), 2777.

Niemi, T. J., T. Kokkonen, and A. W. Seed (2014), A simple and effective method for quantifying spatial anisotropy of time series of precipitation fields, *Water Resources Research*, *50*(7), 5906–5925.

Palmer, T., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia (1993), Ensemble prediction, in *Proceedings of the ECMWF Seminar on Validation of models over Europe*, vol. 1.

Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando (2013), A stochastic model for high-resolution space-time precipitation simulation, *Water Resources Research*, *49*(12), 8400–8417.

Pegram, G., and A. Clothier (2001), High resolution space-time modelling of rainfall: the string of beads model, *Journal of Hydrology*, *241*(1), 26–41.

Piani, C., J. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over europe, *Theoretical and Applied Climatology*, *99*(1-2), 187–192.

Reath, J., J. Dong, and M. Wang (2016), Improved parameter estimation of the log-logistic distribution with applications, *Computational Statistics*, pp. 1–18.

Seed, A. W., C. E. Pierce, and K. Norman (2013), Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme, *Water Resources Research*, *49*(10), 6624–6641.

Shoukri, M., I. Mian, and D. Tracy (1988), Sampling properties of estimators of the log-logistic distribution with application to canadian precipitation data, *Canadian Journal of Statistics*, *16*(3), 223–236.

van Leeuwen, P. J. (2010), Nonlinear data assimilation in geosciences: an extremely efficient particle filter, *Quarterly Journal of the Royal Meteorological Society*, *136*(653), 1991–1999.

Veenhuis, B. A. (2013), Spread calibration of ensemble MOS forecasts, *Monthly Weather Review*, *141*(7), 2467–2482.

Wenger, D. (2016), Data4Web : Distributions de probabilités des précipitations et calibration par le prévisionniste, *Météosuisse*.

Willeit, M., R. Amorati, A. Montani, V. Pavan, and M. S. Tesini (2015), Comparison of spectral characteristics of precipitation from radar estimates and cosmo-model predicted fields, *Meteorology and Atmospheric Physics*, *127*(2), 191–203.

Acknowledgments

The authors are grateful to the Swiss Federal Office of Meteorology and Climatology *MeteoSwiss* for the financial and material contribution to this study, and more specifically to Christof Appenzeller, Head Analysis and Forecasting, and Isabelle Bey, Head Regional Center West.

We would like to especially acknowledge Dominique Stussi, Olivier Duding, Lionel Peyraud, Christophe Salamin, Pierre Eckert and Lionel Fontannaz for their expertise and valuable comments regarding operational forecasting of precipitations in a regional meteorological office. Special thanks goes to Loris Foresti, Daniele Nerini and Ioannis Sideris for their precious advice concerning the numerical modelling of precipitation fields and the methods developed in this study.

A special gratitude goes to the Alexis Berne for his detailed review of this report.

MeteoSchweiz
Operation Center 1
CH-8044 Zürich-Flughafen
T +41 58 460 99 99
www.meteoschweiz.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno Monti
T +41 58 460 97 77
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genve 2
T +41 58 460 98 88
www.meteosuisse.ch

MétéoSuisse
Chemin de lArologie
CH-1530 Payerne
T +41 58 460 94 44
www.meteosuisse.ch