EPFL

hydrique.

École Polytechnique Fédérale de Lausanne

Master thesis

Integration of radar data in rainfall-runoff models for flash flood forecasting

---

Author:
Matthieu Rykner

Supervisors:
Prof. Alexis Berne - EPFL, LTE
Dr. Frédéric Jordan, Dr. Raphaël Mutzner - Hydrique Ingénieurs

February-August 2022

Remerciements

The frontpage picture represents a flood occurring in the Ticino river in Giornico the 8th July 2021.

Contents

List of Figures

List of Tables

Nomenclature

SwissMetNet Rain gauge network of *MeteoSwiss*

AZC *MeteoSwiss* raw radar QPE with hourly timestep

BVE Rain gauge network of *Bern* canton

$CCDF$ Complementary Cumulative Distribution Function - percentage of exceedance function

$COSMO1-E$ Medium lead time COSMO NWP

$COSMO2-E$ Large lead time COSMO NWP

$COSMO$ COnsortium for Small scale MOdelling - European consortium providing COSMO NWPs

$CPCH$ *CombiPrecip* - *MeteoSwiss* radar-rain gauge QPE

FAR Rate of false alarms

$GR3$ Ground storage model

HIT Rate of event detection

$inca0$ First timestep of INCA at +10 minutes

$INCA$ Integrated Nowcasting through Comprehensive Analysis - *MeteoSwiss* radar QPF

MAE Mean Absolute Error

$Nash$ Nash-Sutcliffe coefficient

NWP Numerical Weather Prediction

oasi Rain gauge network of *Ticino* canton

QPE Quantitative Precipitation Estimate

QPF Quantitative Precipitation Forecast

Radar4Infra Project of which this Master Thesis is part

RS *Routing System* - rainfall-runoff model used at *Hydrique Ingénieurs*

RZC *MeteoSwiss* raw radar QPE with 5-minute timestep

$SWMM$ Storm Water Management Model

Abstract

Flash flood prediction and dynamic management is a rising field of research due to the increasing frequency of flood events and the inability of permanent physical measures to remove all flooding risks. The prediction of flash floods usually rely on rainfall-runoff models associated with radar data. The latter have a high spatial resolution and show good performances at short lead time, particularly important for flash flood forecasting.

In that scope, the *Radar4Infra* project aims at developing the combined use of rainfall-runoff models and radar Quantitative Precipitation Forecasts (QPFs) in operational flash flood forecasting. The analysis of radar products provided by *MeteoSwiss* shows that they are generally in agreement with rain gauges but produce higher precipitation estimates (QPEs) at high intensity and in summer. This feature allows to better catch intense precipitations that may result in flooding. However raw radar QPEs are bad in winter and in altitude. Hence a merging of radar with rain gauges (called *CombiPrecip*) is necessary to provide better estimations. Some issues still exist in *CombiPrecip* with a significant underestimation bias in some locations far from the rain gauge network.

Based on these observations, several rainfall-runoff models are developed in two catchments: one mountainous basin (Emosson) and one suburban-rural basin (Zurich-Altbach). Precipitation products are first de-biased before being introduced in the rainfall-runoff model RS *Routing System*. Re-forecast simulations are then performed with the developed models to assess the interest of radar data and to investigate some features improving performances. In particular, it is shown that the modelling of part of the catchment area as impermeable surface is critical to the good predictions of flash floods. A model is notably developed in Zurich-Altbach where the impervious area is increased at high intensity, following similar model modifications proposed in the literature. This modification helps to forecast the main flash flood events occurring in Zurich-Altbach. Nevertheless, the developed forecasts are subject to some intense false alarms that need to be removed.

Finally, an alert system is designed with the goal of proposing the best detection performances with the lower false alarm rates. Appropriate alert systems based on redundancy or on multiple rainfall-runoff models allow to move closer to that goal by removing the most intense false alarms. It is thus recommended to invest time in designing appropriate alert systems. It is also suggested to turn to ensemble approaches to cope with the uncertainty inherent to radar QPFs.

Keywords: flash-floods, radar nowcasting, radar-rain gauge merging, rainfall-runoff models, rainfall intensity, alert system

Résumé

La prédiction et la gestion dynamique des crues éclair constitue un champ de recherche en expansion rapide. En effet, la fréquence et l'intensité des crues éclair augmentent, notamment sous l'effet du changement climatique. Les mesures physiques permanentes ne permettent par ailleurs pas d'atteindre le risque zéro. Dans la majeure partie des cas, la prévision des crues éclair se base sur l'utilisation combinée de modèles pluie-débit et de prévisions de précipitation radar. Ces dernières ont une grande résolution spatiale et de bonnes performances à faible horizon, ce qui est particulèrement important dans le cas des crues éclair.

C'est dans cette perspective qu'a été lancé le projet *Radar4Infra*, dont le but est de développer la prévision des crues éclair à l'aide des radars. Une analyse des produits radar fournis par *MéteoSuisse* est ainsi réalisée. Elle montre que ceux-ci sont généralement en accord avec les pluviomètres de *MéteoSuisse* bien que de plus fortes précipitations soient en général observées l'été et lors des évènements intenses. Ce point est particulièrement intéressant pour la prévision des crues éclair, causées par des précipitations intenses. Toutefois, les précipitations radar pures sont moins bonnes l'hiver et en altitude. Pour cette raison, il est nécessaire de les corriger *via* les pluviomètres de *MéteoSuisse*. Les nouvelles estimations radar-pluviomètres, nommées *CombiPrecip*, présentent de meilleures performances mais ont toujours un biais de sous-estimation dans certaines régions loin des pluviomètres.

A partir de ces observations, plusieurs modèles pluie-débit sont développés dans deux bassins versants : l'un montagneux (Emosson) et l'autre périurbain-rural (Zurich-Altbach). Les produits radar sont tout d'abord débiaisés avant d'être introduits dans le modèle pluie-débit RS *Routing System*. Des prévisions a posteriori sont ensuite effectuées en simulant les processus opérationnels, en partant des modèles précédemment mis en place. Cela permet de mettre en évidence certaines pistes d'amélioration des modèles. En particulier, la nécessité d'introduire des zones imperméables dans la modélisation est mise en évidence. Un modèle incluant une variation de la surface imperméable du bassin versant à fortes intensités est notamment développé à Zurich. Ce modèle est inspiré par un modèle similaire développé en France. Grâce à cette modification, il est désormais possible de simuler à l'avance les évènements les plus intenses ayant lieu à Zurich-Altbach. Néanmoins, ces prévisions sont sujettes à de fortes fausses alarmes qu'il sera nécessaire de faire disparaître.

Un système d'alerte est finalement mis en place, dans l'objectif de réduire les fausses alarmes tout en maintenant les meilleures capacités de détection des évènements intenses. A Zurich, les meilleurs systèmes utilisent la redondance des prévisions ou le couplage de plusieurs modèles pour éliminer les fausses alarms les plus importantes. Cette étude permet de mettre en évidence l'importance d'investir du temps dans l'élaboration d'un système d'alerte adapté. Enfin, il est suggéré de se tourner désormais vers les approches ensemblistes pour gérer l'incertitude inhérente aux prévisions radar.

Mots-clés: crues éclair, prévisions radar, nowcasting, couplage radar-pluviomètres, modèles pluie-débit, intensité de précipitation, systèmes d'alerte

Chapter 1

Introduction

In 2021, OCHA[a] recorded 432 disastrous events in its Emergency Event Database[1]. Of these 432 events, 223 where flood events, making it the first type of disaster all around the world. Deadly floods were observed in India, China, Afghanistan, Germany, Belgium, etc. Floods ranked also first in death toll (4143 deaths) and second in economic damages (74.4 billion USD). Climate change is expected to increase the frequency of floods, with more intense precipitations, notably at high latitudes[2].

Switzerland is not an exception. As described in Figure 1.1, it is largely affected by floods, and notably by flash floods – due notably to its complex orography. Flash floods are floods occurring in catchments with small response time where storms are generally the main cause of flooding.
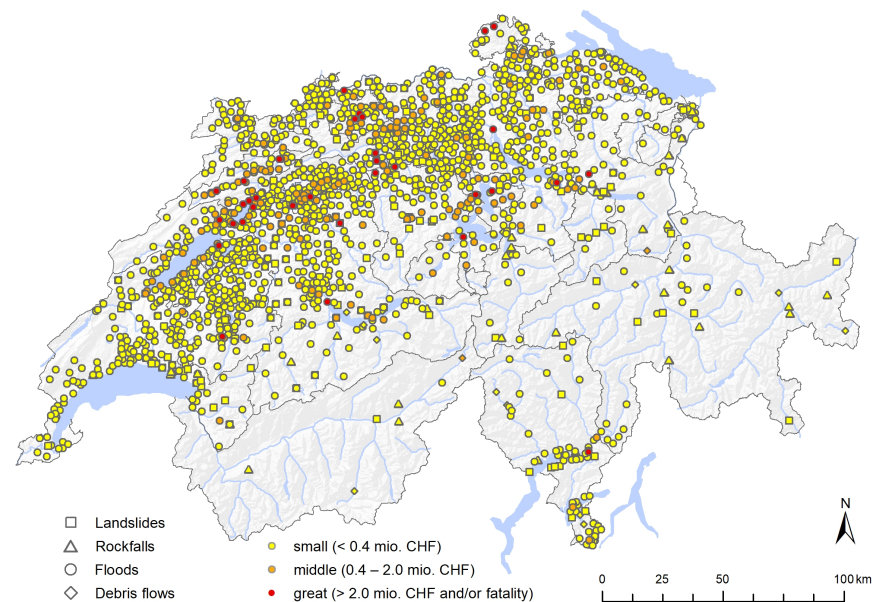
Figure 1.1: Map of the major flood events occurring in 2021 in Switzerland with associated economic damages. Taken from the WSL webpage[3].

A striking example of recent flash flood in Switzerland happened in Cressier in 2021. Due to intense precipitations (28.3 mm in 20 minutes, 40 mm in 40 minutes), to the local topography (a steep-sided valley)[4] and to a significantly low permeable area[5], a severe flash flood occurred. Some images of the damages due to the flash flood are shown in Figure 1.2. Photos were taken from the website

---

[a]United Nations Office for the Coordination of Humanitarian Affairs, in charge of coordinating the global emergency response

`cressier-avant-apres.ch` which proposes a virtual exposition of photographies taken just after the event and one year later in the same place.

(a) Rue des Saint-Martin - after the event   (b) Rue des Saint-Martin - one year later

(c) Rue de Neuchâtel - after the event   (d) Rue de Neuchâtel - one year later

(e) Rue des Saint-Martin - after the event   (f) Rue des Saint-Martin - one year later

(g) Caveau de l'entre-deux-lacs   (h) Chemin du vieux-moulin   (i) Chemin du vieux-moulin

Figure 1.2: Several photos of the damages due to the flash flood events from 22ndJune 2021 in Cressier. Photos are taken from the virtual exposition "Cressier avant-après" and were taken by Alexandre Witschi.

Despite the severe consequences of floods, the World Meteorological Organization (WMO) recalled that "complete protection from flooding is rarely a viable goal"[6,7]. Hence flood protection measures cannot be based only on permanent physical measures such as structural infrastructures (dams,

channels, flood plains, etc.) or interdiction of human activities in flood-prone areas. It is indeed necessary to prepare emergency measures in case of higher events. These measures could be either temporary protections (flood gates, barriers, sandbags) or temporary evacuation of people and of the most valuable goods (for example museum collections). However, to decide these emergency measures it is necessary to anticipate the arrival of floods. Flood forecasting is thus an essential feature of flood management.

Nevertheless, while floods in large basins can be predicted with only observed discharges and precipitations with quite significant time available for action, flash floods are much more complicated to predict. With smaller response time, precipitation forecasts (QPFs) and discharge prediction with rainfall-runoff models become, for example, necessary. The constant research of systems increasing the anticipation time has made flash flood forecasting an active field of research. This field also owns the specificity of requiring significant links between the academical field and the public authorities, notably via meteorological agencies.

It is in this context that the *Radar4Infra* project has been launched one year ago. Supported by companies such as *Alpiq* or *Hydrique Ingénieurs*, federal agencies and schools such as *MeteoSwiss*, *WSL* or *EPFL* and local administrations such as the cantons of Basel and Valais and the city of Lausanne, it is a good example of the common work required between all these actors. It aims at improving the prediction of flash floods triggered by intense precipitations, by using radar nowcasting products provided by *MeteoSwiss*.

Radar nowcasting tools provide spatially-distributed forecasts of precipitation with high spatial resolution. They are also largely better than the commonly used Numerical Weather Predictions (NWPs) at short lead times, notably from 0 to 6 hours, which are the usual response times of catchments subject to flash floods. Thus, they should in principle significantly improve the performances of currently used forecasting tools. This report aims at introducing radar data in the rainfall-runoff models developed by *Hydrique Ingénieurs*, at confirming their interest in operational forecast and at pointing the main features to care about. In particular, its goal is to answer the following questions:

- Do *MeteoSwiss* radar products improve the estimation and forecast of precipitation, and if yes for which features and with which limitations ?
- Which techniques allow to take the most of radar data to improve the prediction of discharge ?

To answer these questions, the report is organized as follows:

- Chapter 2. A literature review is first conducted to highlight the main issues and ways of improvements developed and used by the research community.
- Chapter 3. The methods used during this project are presented.
- Chapter 4. An analysis of radar products provided by *MeteoSwiss* is performed to evaluate the main differences between radar and rain gauges products.
- Chapter 5. Radar data are introduced in rainfall-runoff models in two catchments (Emosson and Zurich) to evaluate their added value.
- Chapter 6. Finally, several alert systems are developed to highlight the necessity of carefully choosing the moment when an alert must be issued.

This report focuses on flash flood prediction and alerts. Other features, such as volume predictions at short lead times, have also been studied during this Master Thesis, as they can also be improved by radar data and may be interesting for some partners of the project. These more exhaustive analyses are proposed in Appendix.

Chapter 2

Literature review

2.1 Introduction to flash flood forecasting

To predict flash floods, several research topics are widely investigated. Hapuarachchi *et al.*[8] have proposed a comprehensive review of the main steps in flash flood forecasting:

1. The improvement of input data is first required. Notably the introduction of radar precipitation estimates and forecasts (QPEs and QPFs) improves the streamflow prediction performances for small catchments (Section 2.2).

2. Streamflow forecast models are another active field of research which will be presented in Section 2.3.

3. Finally, the design of an alarm system and the decision-making process are central aspects for the mitigation of flash floods impacts (Section 2.4).

Ensemble systems are finally a promising way to improve flash flood forecasting. Though they are not the focus of this Master Thesis, they will be quickly presented in Section 2.5, for sake of completeness.

2.2 Radar acquisition and forecasting

2.2.1 Introduction

Radar (or RAdio Detection And Ranging) is used to call all detection technologies based on the reflection of radio waves. Its principle was discovered in 1886 by Hertz[9] who demonstrated the theory of Maxwell that electromagnetic waves can be reflected by objects, as light. Radar interest to detect objects was first demonstrated in the beginning of the 1900s by Hulsmayer in the case of ships detection. However it only started to be developed in the 1930s and during the world war II as a way to detect enemy planes. One of the issue encountered at that moment was the disturbance of radar by clouds and rainfall cells. This feature was studied, after the war, contributing to the development of meteorological radars. The first meteorological radars were useless military radars converted into weather ones like the first installed in the USA in Washington D.C. in March 1947.

In Switzerland, the first weather radar was installed in 1959 on La Dôle[10]. The radar system then evolved with three consecutive generations before the construction of the fourth generation of radars in the 2010s. Weather radars are mostly used to detect precipitation cells. Compared to rain gauges they have the advantages of providing information on a large area rather than on just one point. However they are usually associated with more bias and uncertainties.

Figure 2.1: Photo of the new *MeteoSwiss* radar in Weissfluhgipfel, put in operation in 2016 in the Grisons. Taken from the *MeteoSwiss* website[11].

2.2.2 Quantitative Precipitation Estimates (QPEs)

Principles of radar precipitation measurements

The main objective of weather radar is the detection of precipitations, also called Quantitative Precipitation Estimates (QPEs). To do so weather radar emits an electromagnetic signal with a constant frequency in a determined direction (usually a pulsed signal). This signal is then scattered by objects and notably by water droplets, snow flakes or hail. The scattered signal is finally recorded and amplified by the radar. The time between the emission and the reception of the signal gives access to the distance between the precipitation and the radar. The main input variable for radars is the reflectivity which corresponds to the amount of power received by the radar divided by a reference power. A post-processing of the reflectivity map is then required to convert it to a precipitation intensity map.

The first generation of weather radars used mostly the reflectivity information to access the 2D horizontal precipitation intensity map. However more recent radars are now able to provide much more information on the precipitation field. First, current radars are most of the time Doppler radars: they are able to measure the frequency of the scattered signal and hence to estimate the speed and direction of precipitation.

Moreover, while most old radars are single polarisation radars with the same antenna emitting and receiving the signal, new radars are mostly dual-polarisation radars, consisting of an emitter and a separate receiver, as the one tested in Montana in 1981[12]. This allows to gain a lot of information and notably to provide information on the vertical distribution of precipitation. Indeed the signal direction can now be controlled in both vertical and horizontal directions giving access to a 3D precipitation map. The NEXRAD (NEXt generation RADar) network in the USA operated by the National Weather Service (NWS) is an example of such dual-polarisation radar networks and has demonstrated the improvements brought by this technology in the case of Iowa precipitation events[13]. Dual polarisation radars are also able to provide information on the size, shape and hence

nature of precipitation. Indeed they are able to capture variables such as the circular depolarization ratio (CDR) which gives information on the anisotropy of the particle shape[12,14,15]. Correlation coefficient (CC, measure of the similarity between particles), differential reflectivity (ZDR, measure of the ratio between horizontal and vertical reflectivity, giving information on the shape of particles) or specific differential phase (KDP, measure of the phase change ratio in horizontal and vertical directions, indicating where droplets are bigger and more concentrated) can also be obtained by dual-pol radars[16]. From these measures it is possible to determine the nature of precipitation (rain, snow, hail, etc.), the rain/snow limit[17], the intensity of precipitation or the particle size distribution[18].

Finally radars can provide other information, such as the wind speed[19–21]. Such measurement requires however the presence of reflecting particles such as droplets, debris or even insects[22]. The operation mode can also be adapted to observe smaller particles and determine the wind speed (in clear-air mode)[23]. Wind speed at sea surface can similarly be measured by radars (usually airborne or satellite radars)[24]. Radars can also be used to detect thunderstorms[25,26] or tornadoes notably at night when visual detection is not possible[27–29] (Figure 2.2). They can finally be used in the study of longer-term phenomena such as the solar cycles[30,31].
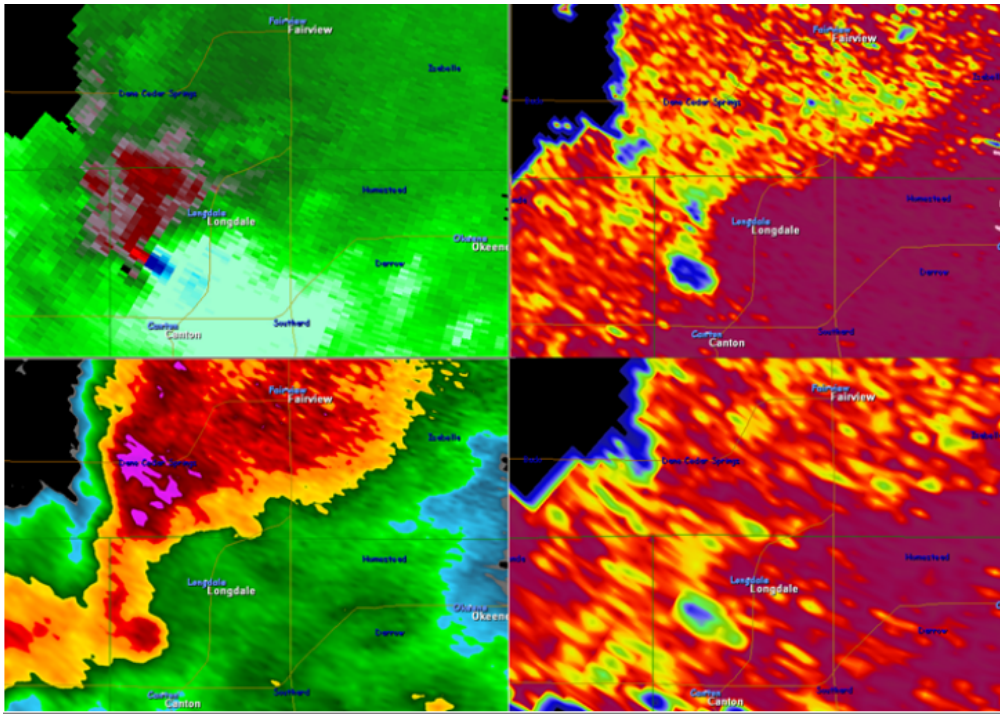
Figure 2.2: Base reflectivity image (lower left), storm-relative velocity (upper left) and dual pol cross-correlation data (CC, taken for different elevation angle, right) during a storm event occurring in Kentucky. A tornado with debris is observed in the bottom left of each image, characterized by a duet of strong velocities in red and blue in the storm-relative velocity image. Right images show that the tornado has low CC values (in blue) due to the presence of debris of various sizes. Taken from the NWS website[28].

Beyond single-pol and dual-pol radars, radars are usually split in several classes according to their frequency range[32]. The distribution of radar by classes over the globe is displayed in Figure 2.3a taken from Saltikoff *et al.*[33]. The smaller frequency radars are called L-band radars and use frequency between 1 and 2GHz. Their wavelength is large, the signal is nearly not attenuated and hence the distance range is big. However L-band radars are not able to detect small particles. They are also complex, big and expensive. They are mostly used for clear-air studies. S-band radars use the next trench of frequencies (2-4GHz) and are a little bit smaller. They are nevertheless still

expensive and large (with dish size up to 7-8m). This is a very commonly used class of radars, notably for the Chinese[34] and American[35] radar networks. C-band radars are also widely used and deployed as they are simpler and cheaper. Their range is smaller as signal attenuation plays a more important role but they are still particularly adapted to medium countries such as European ones. They are massively deployed, notably in Europe, but are also used in the Chinese and American networks. All five *MeteoSwiss* radars are C-band radars[11,36,37]. However C-band radars are still complex system to build and operate, while their cost is still significant. Thus X-band radars are also largely deployed, notably to complement existing networks in smaller (usually mountainous) areas[38,39]. The mix between X-band and C-band radars, though being more complex to operate than a one class network[38], may provide more insights, notably for severe events[25,40]. Finally the smaller radars are K-band radars.
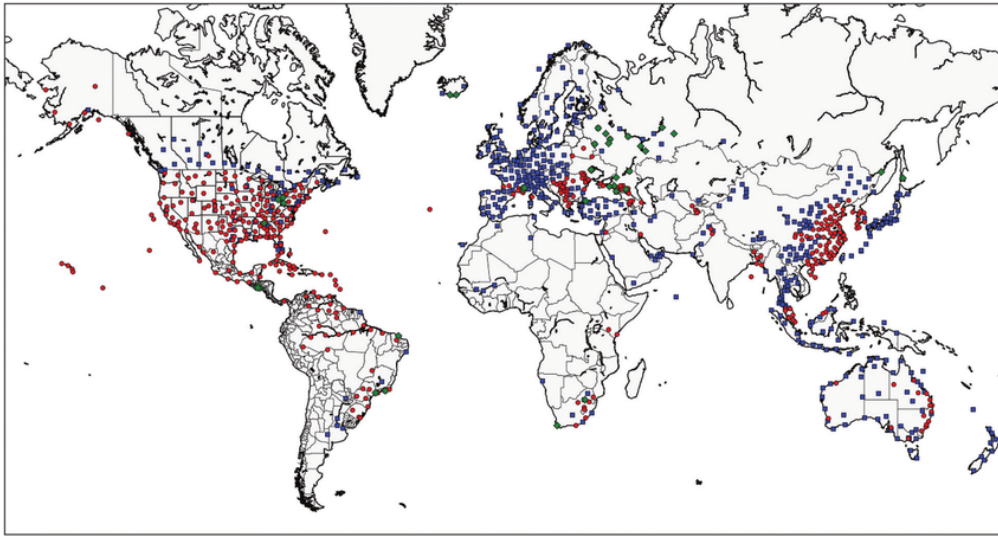
(a)
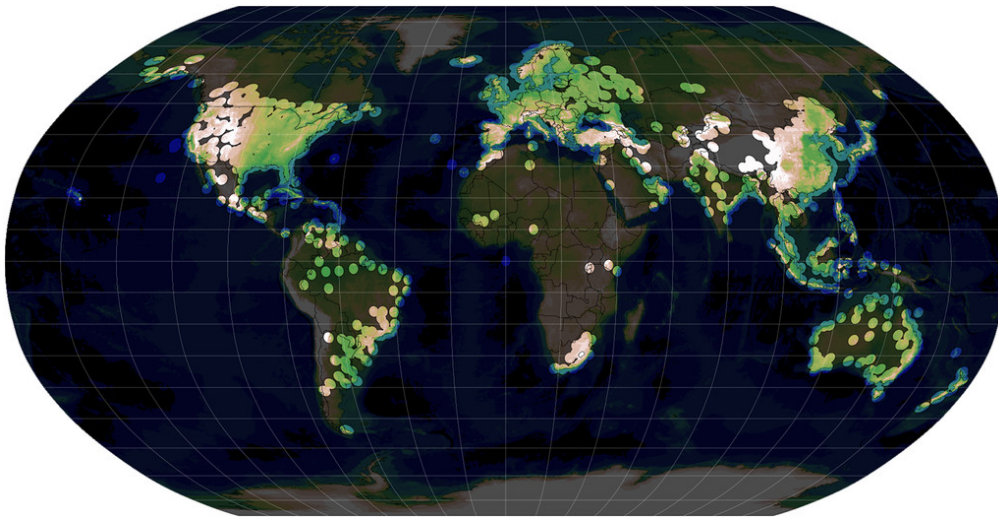
(b)

Figure 2.3: Coverage maps of the Earth by radars. (a) Distribution of radars by band class in 2016 (S-band: red, C-band: blue, X-band: green). Chinese and Russian locations of radars are approximate. Taken from Saltikoff *et al.*[33]. (b) Radar coverage of the world. Taken from Saltikoff *et al.*[41].

Radar coverage is nowadays highly variable from one place of the world to another. Indeed, as illustrated by Figure 2.3b, while USA, Europe, Eastern China and Australia-New Zealand are nearly completely covered by radar, only few radars are installed in the rest of the world. Hence most of publications related to radars are focused on these developed countries.

QPEs in complex orography

Unlike most of the USA and China, Switzerland is not a flat, homogeneous country. It has a quite complex orography, marked by the four big regions: the *Alps* in the South, the mountainous *Pre-Alps* in the center, the almost flat *Plateau* in the North and the mountains of the *Jura* in the North-Western border. Hence the coverage of Switzerland by radars imposes new challenges due to its mountainous areas. These challenges are notably summarized by Germann *et al.*[42]. They can basically be split between two categories:

- Practical issues due to the difficulties to access some locations. The choice of the location of radars is hence determinant and must include these practical aspects.

- Scientific issues due to the location of radars, to the nature of precipitation in altitude and to the impacts of the ground. The major issues are terrain shielding (hiding certain areas to the radar), ground cluttering (non-precipitating echoes), signal attenuation, snow and melting snow contamination, vertical variation of the reflectivity or vertical variation of the cell sizes.

The choice of the location of radars is hence determinant. It should be as high as possible to cover as much area as possible but must also not be too difficult to reach. The hardware should also be protected against severe and varying meteorological conditions as the system needs to be stable and to operate remotely and continuously (24/7). Finally the number of radars is essential to cover the widest zone possible and to improve the data quality.

On the scientific side, in addition to the location of the radars, a post-processing of data with specific algorithms is required. The improvements of current algorithms constitutes a significant field of research since 20 years[43–46]. To counter cluttering and non-precipitating echoes (Figure 2.4), *MeteoFrance* notably developed fuzzy logic techniques to identify and remove cluttering more efficiently than with a simple mean clutter removal[45,46]. Fornasiero *et al.*[47] tackled both ground cluttering and terrain shielding issues by using previous radiosonde observations to better model the propagation trajectories of the radar beam. Improvements of radar data processing step can also pass by the use of different input variables. For example, Vulpiani *et al.*[48] used the specific differential phase (KDP) rather than the usual reflectivity map to estimate precipitation in two Italian regions including Alpine South Tyrol. Their method performed better than the common reflectivity-based technique except during winter storms where contamination by phase-shifting snow degraded the performances. Similarly, Cremonini and Bechini[49] overcame attenuation issues in extreme events in the case of the Piemonte region – where complex orography is coupled with strong Mediterranean rainfall events – by relying on the phase of the signal rather than on the reflectivity. In the same way, Montopoli *et al.*[50] compared several algorithms based on KDP, single-polarisation reflectivity (Zhh) and dual-pol differential reflectivity (Zdr). They demonstrated than an algorithm based on the combination of the former three performs up to 25% better than algorithms based on only one or two of the previous ones. The research of more accurate post-processing algorithms is thus an active field of research with direct implications for operational applications, notably for *MeteoSwiss*[51–53].

Today radars are largely used in complex orography. The major Alpine countries have notably developed different networks, with the 5 C-band radars of the Rad4Alp Swiss project[11,42,54], the C-band dual-pol Austrian radar installed in 2007[55], the mixed C and X-band French radar network (with X-band radars used to fill the gaps in the Alps)[38,39,56] and the numerous Alpine Italian radars[57]. Similarly a large number of mountainous regions in developed countries are covered by radars and are coping with the issues presented before: in the USA[58], in Taiwan[59], South Korea[60], China[61,62], New Zealand[63]... Some radars are also starting to be operated in underdeveloped or developing countries such as Nepal[64] or Ecuador[65]. A particular case of complex orography radars is finally constituted by islands or coastal mountainous regions where radars need to be positioned at sea level to capture atmospheric phenomena above the sea. Ground shielding becomes a huge problem in those cases (for example in New Zealand, Taiwan, South Korea or Pacific islands).
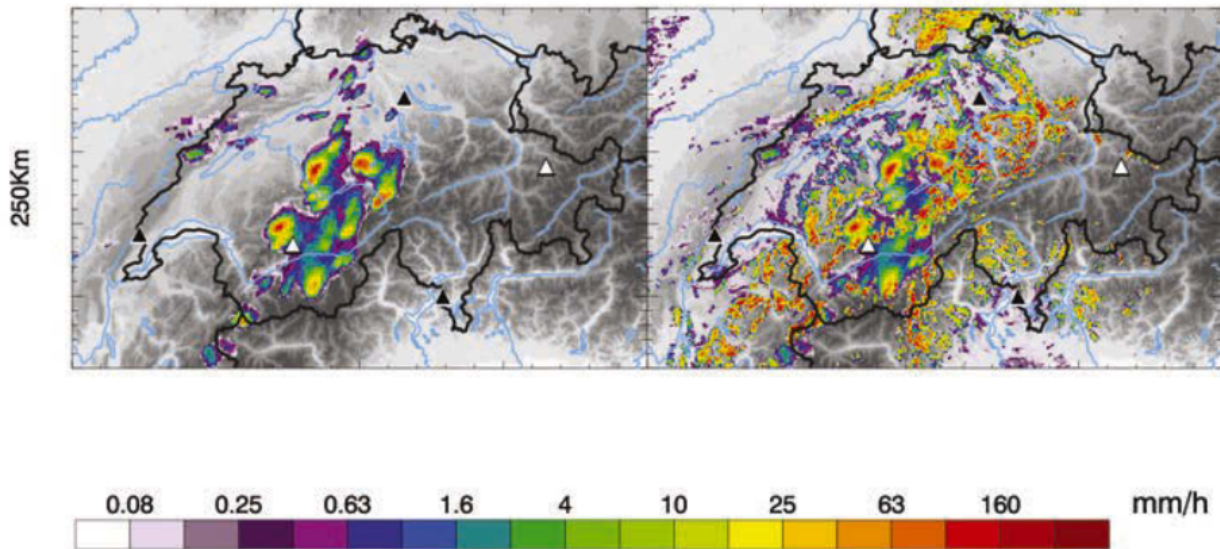
Figure 2.4: Two Swiss radar images with (left) and without (right) radar clutter suppression algorithm. Taken from Germann *et al.*[36].

QPEs in urban hydrology

Urban hydrology is another essential field for the Switzerland case as a certain number of extreme runoff events happen in urban centres due to the imperviousness of soils and to the high concentration of people and goods. To predict, and later prevent, urban floods a precise knowledge of precipitation is necessary. However rain gauges are point-located and cannot provide sufficiently precise data for urban hydrology. Concerning usual radar products, their gridscale is of usually 1 km at least, which could be sufficient or not depending on the size of the catchment, on the storm characteristics and on the type of simulation run – peak runoff response upstream of the basin will require a coarser radar grid than downstream[66]. Hence it could be necessary to develop other radar products with shorter-range, higher resolution possibilities. Special algorithms are also needed for such application.

Radar-rain gauge merging

While radar QPEs provide precipitation data on a large spatial grid, they are usually less precise than rain gauges. In order to correct the errors from radar, rain-gauges and radar data can be merged to create more reliable QPEs. Such approaches have been widely developed since the first use of weather radars. The correction of bias can be performed with a variety of methods from simple use of multiplicative factors to more complex geostatistical or probabilistic approaches.

In the case of multiplicative factors, they can be either constant in time but spatially dependent[67], constant in space but timely dependent – with the use of mean field bias[68,69], Kalman filter[70], multiquadric surface[71] – or dependent on multiple other parameters. For example, Gabella *et al.*[43] proposed a correction of bias based on a multiple linear regression with altitude (HG), height of radar visibility above the rain gauge (HV) and distance from the gauge to the radar (D).

Geostatistical approaches for the merging of radar and rain gauge data have been largely investigated[72–79]. Most of the times the authors use kriging techniques. Using ordinary kriging (OK), it is first possible to interpolate rain gauge data over space[76]. Radar data can later be incorporated in the kriging process, either as an external drift (KED)[75,76] or as another random variable with cokriging. Cokriging could itself be split between ordinary cokriging (OCK)[72–74], ordinary collocated

cokriging (OCCK)[76], universal (UCK) or disjunctive (DCK) cokriging[74], or cokriging with external drift (CKED)[78]. These kriging methods differ by the way the secondary random variable (radar measurements) is incorporated.

Finally probabilistic approaches (mainly *Probability matching method* - PMM) were also used sometimes though less often than geostatistical ones[80,81]. Bayesian approaches are however constituting an interesting approach to account for uncertainty in both rain gauge and radar measurements[82].

In the case of Switzerland, *MeteoSwiss* developed, in addition to the raw radar data, a merging tool with radar and rain gauges data called *CombiPrecip*. This tool is decribed by Sideris *et al.*[78] and consists in a co-kriging with external drift. The model uses the two assumptions that the point-rain gauge measurements are close to the precipitation value around the rain gauge (the rain gauge area-point variance is small) and that radar errors are spatially correlated (adjacent radar grid points have similar bias). Four variables are then used in the co-kriging: the rain gauge and radar data on the considered period of time and the rain gauge and radar data on a period of time preceding the period of interest. The results appear to have better cross-validation scores when the aggregation period goes from ten minutes to an hour. Hence the final *CombiPrecip* product proposed by *MeteoSwiss* has an original timestep of one hour. A disaggregation of this product with a timestep of 5 minutes is also available[83]. The *CombiPrecip* products do not include any altimetric gradient in the calculations. Concerning the performances of *CombiPrecip*, they have been validated on 132 independent rain gauge stations, over the period 2012-2018 and for several aggregation times by Barton *et al.*[84]. Four properties were evaluated there: the characterization of wet/dry events, the characterization of extreme events, the bias and the error on the mean precipitation. Results concluded that the co-kriging merging of radar and rain gauges largely improved all performances.
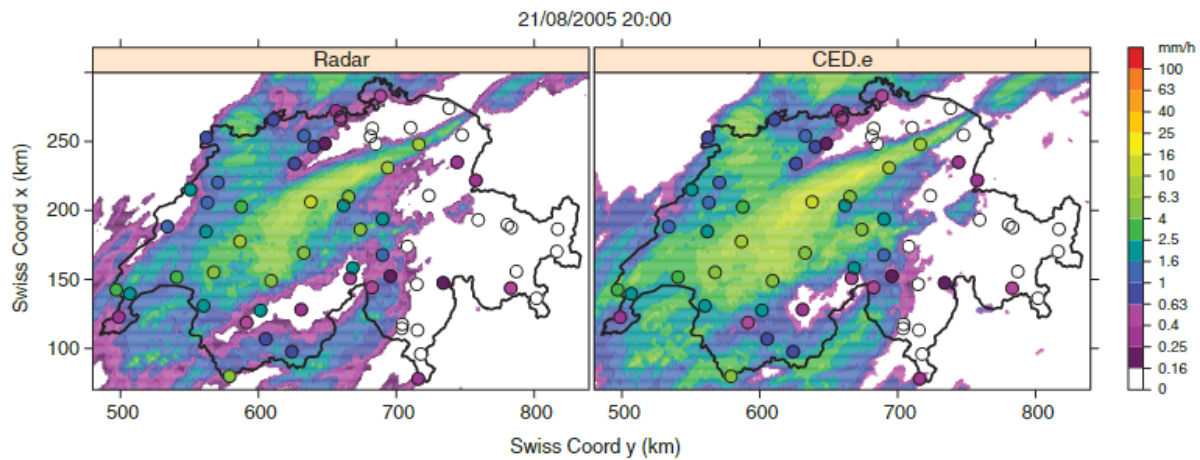
Figure 2.5: Raw radar image (left) and *CombiPrecip* data after co-kriging with rain-gauge (right). Taken from Sideris *et al.*[78].

The merging of radar QPEs with rain gauges usually improves the performances of QPEs. Nevertheless it can also introduce some errors and remove part of the information brought by radars. Indeed rain gauges are also subject to a bench of uncertainties. Notably systematic errors are usually important with an underestimation of precipitation reaching the ground that could be as high as 30%[85]. High intensities are also usually underestimated. Such systematic errors can be in part automatically corrected by instruments. The WMO listed several sources of errors and uncertainties usually encountered with rain gauges in its 2018 report[10,85]: errors due to wind that could be up to 10% (up to 50% for snow), errors due to wetting loss on the walls of the collector or when the collector is empty (up to 15% in summer), errors due to evaporation (0-4%), error due to blowing and drifting snow, errors due to the splashing of water, mechanical and sampling errors or random errors of the instruments. In the Swiss case the snow and wind uncertainties will be particularly

present notably in altitude: rain-gauges are usually less reliable there. A last uncertainty comes from the point-area variance between localized gauge measurements and real areal precipitations[86], especially in convective situations, typically occurring in summer. Radar can thus bring new information and it is important to prevent radar data from being completely covered by rain gauges. Hence *MeteoSwiss* manually gave more importance and freedom to radar data in summer in the *CombiPrecip* tool. In the view of most meteorologists, radar data are the reference data on which rain gauges are evaluated. However when radar products are introduced they need to be evaluated and rain gauges are then needed as reference.

Radar QPEs and performance assessment

To evaluate the performances of radar QPEs, rain gauges are generally used as reference data. This evaluation is performed either for distinct events[87,88] or for continuous data spanning over months or years[89]. Such methods will be used in this report to estimate the quality of *MeteoSwiss* QPEs both for events (using HIT/FAR metrics) and for continuous data (using cumulated precipitation metrics). However these methods assume rain gauges are perfect data which is not always the case. Thus, other techniques of performance measurements have been developed to take the rain gauge uncertainties into account. An example is provided by Anagnostou *et al.*[86] which proposes a method to estimate rain gauge area-point variance and then better estimate the radar to rain gauge error.

A last interest of radar QPEs resides in the possibility to use them as reference on which to compare Numerical Weather Predictions (NWPs). Indeed rain gauges cannot be used as references as they are localized while NWPs aim at providing a spatial forecast. Such comparison between NWPs and rain field obtained by radar have for example been led in Czech Republic to assess the performances of the COSMO NWP for two extreme events[90].

2.2.3 Quantitative Precipitation Forecasts (QPFs)

Once QPEs are available from radars, they can be used to develop new forecasting tools in addition to usual NWPs. All these products range in the category of the Quantitative Precipitation Forecasts (QPFs).

Techniques used to create QPFs

Two types of QPFs are based on radar data:

- Models based on storm advection (or Lagrangian persistence).
- Models intending to estimate precipitation cells growth and decay.

The Lagrangian persistence is the simplest way to use radar QPEs to predict future precipitations. It is based on the continuity idea that precipitation cells currently present somewhere will still be present in another place in the future with similar intensities. The idea behind Lagrangian persistence was described notably by Zawadski *et al.*[91]. It starts from a simple comparison to Eulerian persistence where the precipitation intensity (denoted i) in location (x, y) and at time $t + \Delta t$ can be estimated as $\hat{i}(x, y, t + \Delta t) = i(x, y, t)$ (where the hat denotes an estimate). Future precipitations are here only estimated with current precipitation. Every forecast model must have better performances than Eulerian persistence. Similarly, the Lagrangian persistence estimates that the intensity in the rainfall cell will be constant. However this time the cell may move in space. Hence the future precipitation is now estimated to be:

$$\hat{i}(x, y, t + \Delta t) = i(x - \alpha_0, y - \beta_0, t)$$

α_0 and β_0 are spatial lags that can be chosen by different ways depending on the algorithm used. They can either be computed by successive radar images analysis[91,92] – searching the maximum

cross-correlation displacements – or being chosen *via* wind measurements. The Lagrangian persistence is thus similar to the simulation of rainfall (or storm) advection process without decay or growth of the cells. Despite its simplicity and high accuracy at very short lead time, the Lagrangian persistence wears several defaults and has notably quickly degrading performances at higher lead times (few hours) when convective cells can appear or disappear. Studies on the relationship between lead time and spatial scale and resolution of the radar data have been developed to assess this aspect[93].

The low lead time of Lagrangian persistence in highly convective regions with quickly varying cells led to the development of conceptual models based on the growth and decay of rain cells. Nevertheless modelling these phenomena is quite challenging and has not given good results until recently. A first example (based on satellite data rather than radar data) has for example been proposed by Walker *et al.*[94]. It is a quite complex model and still raise a high rate of false alarms. More recently, Han *et al.* and Mecikalski *et al.* have demonstrated that growth and decay models could be improved by the use of respectively Machine Learning[95] and blending with NWPs[96]. However Lagrangian persistence is still the most used radar QPFs in operational fields.

To introduce growth and decay notions in Lagrangian persistence QPFs, it is also possible to rely on other NWPs. This is notably what is done by *MeteoSwiss* in its *CombiPrecip* products[92] where localized growth and decay factors are added to the Lagrangian persistence. These growth and decay factors are determined from historical NWPs data suing simple linear regressions. Other approaches based on Machine Learning have been followed, for example by Foresti *et al.*[97].

Blending process with Numerical Weather Predictions (NWPs)

To improve the forecasting skills of Lagrangian persistence and to counter the quick decrease of its accuracy with time, a blending of radar QPFs with NWPs is most of the time performed in operational circumstances. It allows to combine the good prediction at very short-time of Lagrangian persistence with the good NWPs predictability at higher lead time. Indeed NWPs have usually poor performances at very short lead time due to imprecise determination of the initial state of the atmosphere, downscaling difficulties and time needed to compute NWPs[92,98]. The accuracy and complementarity of NWPs and radar QPFs is illustrated by Figure 2.6. The research of a good blending scheme is thus another source of intense focus.
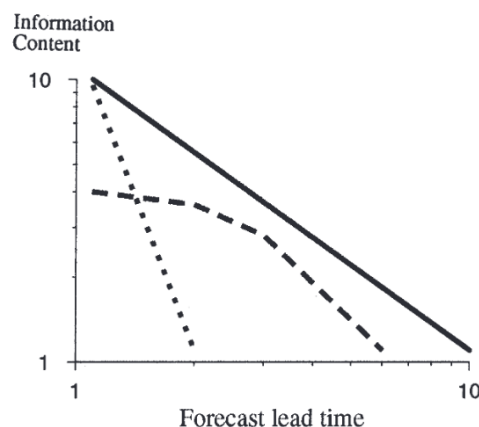
Figure 2.6: Evolution of the information contained in NWPs (dashed line) and radar QPFs (dotted line) with lead time. The maximum theoretically reachable information is given by the solid line limit. Radar QPFs are able to give a nearly perfect information at very-short lead time but their loss of information increases very quickly unlike NWPs. Taken from Golding[99].

Usually the blending scheme consists in a simple linear scheme where weights are given to the radar QPF and to the NWP. The weights vary linearly with time from a situation where only the radar QPF is considered to a situation where all the forecast is based on the NWP. Such scheme was used by *MeteoSwiss* – and other meteorologial agencies in Central Europe – in the INCA product, with only storm advection before 2h of lead time and a decrease of its importance until 6h of lead time when all the forecast is based on NWP[100] (ALADIN in the case of Austria, COSMO for Switzerland). With a slightly more complex approach, the NIMROD forecasting tool in United Kingdom incorporates radar data in their older NWP forecast. This time the weights are determined by long-term forecast error statistics[99] (the system is slightly more complex as it also includes clouds and visibility analysis and some growth and decay aspects). More recently, Lin *et al.*[101] compared several blending schemes in the case of Taiwan extreme events, with notably a linear blending scheme called ExAMP. They showed that this scheme had the best performances.

Recently, more complex blending schemes have been proposed. Notably, in Switzerland, *MeteoSwiss* investigated a blending based on a bayesian approach[102]. Both the NWPs and Lagrangian persistence are here seen as probability distributions with a certain uncertainty rather than deterministic forecasts. An Ensemble Kalman Filter (EnKF) can thus be set up by using NWPs and Lagrangian ensembles. Today this system is not operationally provided but COSMO NWPs ensemble is already available (with 21 members) while INCA radar QPF ensemble is expected to be available soon.

The current nowcasting product from *MeteoSwiss*, called *NowPrecip*, includes a variety of new state-of-the-art features[92]. It was implemented in 2019 and is based on an ensemble approach. Based on each member of the COSMO ensemble[103–105] it generates a new nowcasting member. The Lagrangian persistence is complemented with localized growth and decay factors of rain cells based on historical localized COSMO growth and decay rates. This process does not predict the birth of new cells. Some spatially-correlated noise and localization adjustment noise are also introduced. Finally the blending scheme between Lagrangian persistence and COSMO depends on the Pearson coefficient of the COSMO member: the better are COSMO performances, the quicker COSMO will be used in *NowPrecip*. In Sideris *et al.*[92] the input QPF is the raw radar data. However the input QPF used by *MeteoSwiss* is now *CombiPrecip* which integrates rain gauge data (see Section 2.2.2).

Performance assessment

The performances of radar QPFs need to be assessed once developed. This should be done for all products, blended or not. Usually this would be done by comparison to radar observations and for some events only as in Lin *et al.*[101]. More complete investigations on the sources of QPFs errors have also been performed. This is the case of Ebert and McBride who split the errors of radar QPFs into three categories[106]: location errors (due to a wrong estimation of the rainfall displacement), rain volume errors (errors on the spatially integrated intensity of precipitation) and pattern errors (due to changes in the shape of the convective event). They demonstrated that location errors are the major source of errors for most precipitation events while intensity errors are dominant in extreme events.
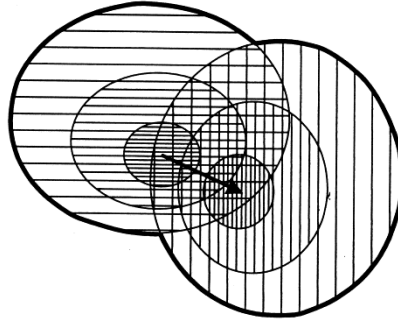
Figure 2.7: Scheme of the different types of errors of a Lagrangian persistence radar QPF. Observation and forecast are represented respectively with horizontal and vertical hatching. Location errors (arrow) and pattern errors (major axis: horizontal for observation *vs* vertical for forecast) can be noticed. Intensity errors do not seem to be present. Taken from Ebert and McBride[106].

Other assessments of the QPFs performances can also be made with rain gauges as reference. This is notably the case in Werner and Cranston where the NIMROD nowcasting British system is evaluated[88]. This allows also to evaluate the observed radar data as will be done later in this report for *MeteoSwiss* products. The interest of Werner and Cranston article resides also in the introduction of new evaluation metrics. Indeed, they introduce two metrics based on accumulated rainfalls over several events:

- The RMSf: $RMSf = exp\left[\left(\frac{1}{N}\sum_{i=1}^{N}\left(ln\frac{R_i}{G_i}\right)^2\right)^{\frac{1}{2}}\right]$ where N is the number of events considered, R_i is the cumulated volume of precipitation observed or forecasted and G_i is the cumulated volume of precipitation obtained by rain gauges. The RMSf should ideally be equal to 1. It is a way to evaluate the precision of radar data.

- The bias: $Bias = \frac{1}{N}\sum_{i=1}^{N}\left(R_i - G_i\right)$. It is a way to evaluate the systemic error of radar data.

Such analysis are interesting as they are based only on some important events. Events are here simply determined via a pre-determined threshold level.

2.2.4 Satellite data

The focus of this section was only on radar data. However it is interesting to notice that similar procedures (notably for QPFs production and further blending scheme) can be applied to infrared satellite data. A model of growth and decay of rain cells was notably presented above[94] while Lagrangian persistence has been used with satellite data[107].

Compared to radar, satellites can also provided other information such as remotely-sensed basin parameters[8] that could be useful for hydrological modelling. They are notably able to estimate evapotranspiration rates[108], snow cover[109], soil moisture[110] or glacier extent evolution[111].

Finally spaceborne radars (radar satellites) can also be used[112,113]. They present the advantages of removing all shielding issues of ground-based radars and of working at constant ranges (about 400-420km – on the contrary ground-based radars must work between very short ranges and large ranges). However they are not always available and present new resolution issues when their angle of incidence become large. Both technologies can complement each other.

2.3   Rainfall-runoff models

To alert the public authorities and the population in case of flash flood occurrence, the knowledge of the precipitation intensity is not always sufficient. Indeed a variety of parameters affect the runoff discharge generated by rainfall such as the localization of intense rainfall, the state of the soil, the characteristics of the basin, the presence of snow, etc. Rainfall-runoff models are thus necessary to convert precipitation into discharge.

A large variety of rainfall-runoff models have been investigated and developed since the development of the Rational Method by Mulvaney in 1851[114]. They can be separated according to four major features:

- The approach followed[115,116]: deterministic or probabilistic as described in Figure 2.8.

- The modelling structure[117]: empirical (or data-driven, based on black-box statistical relationships between inputs and outputs), conceptual (based on simplified components and equations) or physical (based on physical equations and real characteristics of a catchment). Deterministic models can fall in either category while probabilistic models are very often empirical models which do not include any physical phenomena. Some conceptual or physical models could however also be probabilistic.

- The spatial structure: either lumped (the basin is treated as a single homogeneous unit), semi-distributed (the basin is split into sub-basins treated as homogeneous units) or distributed (the basin is regularly split into a grid of cells treated as homogeneous units) as illustrated in Figure 2.9.

- The time structure: continuous or event-based models[118]. Most of the models are continuous but some event-based models will be described in Section 2.3.5.
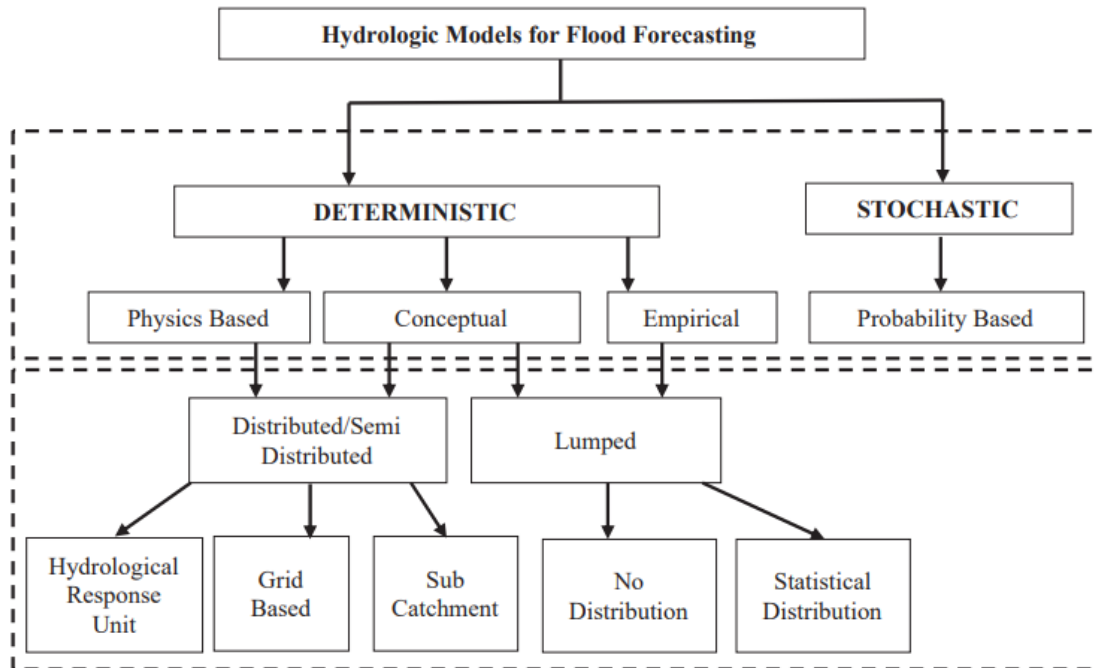
Figure 2.8: Classification of rainfall-runoff models. Taken from Jain *et al.*[116] and World Meteorological Organization (WMO) flood forecasting report[6].

The modelling and the spatial structures of rainfall-runoff models are not independent: empirical models are usually lumped, conceptual models are lumped, semi-distributed or distributed, while physical models are semi-distributed or distributed. Each type of model has advantages and drawbacks favouring their use in one case or another.
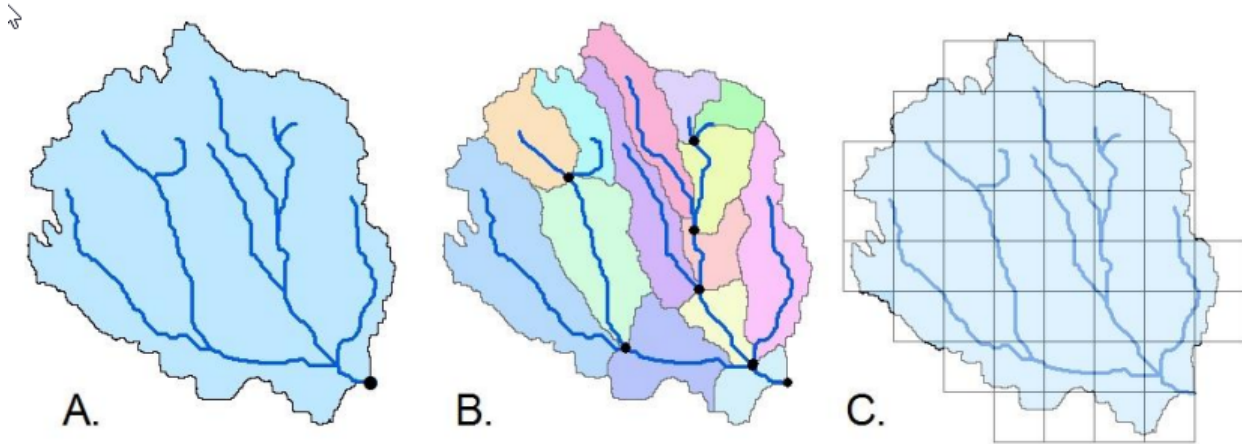
Figure 2.9: Representation of the three types of rainfall-runoff model spatial structures: A. lumped - B. Semi-distributed - C. Distributed. Taken from Sitterson *et al.*[117].

2.3.1 Data-driven models

Empirical or data-driven models are the simplest models which are mostly used when data on the physical processes are lacking. These models usually need input and output data covering a large period of time[6]. They are often referred as black-box models as they do not model physical processes. They can be divided into two main categories[116]: stochastic models based on regression or Bayesian approaches, and non-linear models with notably neural networks and fuzzy logic models.

Stochastic models have been widely used before for seasonal forecast. Models based on linear regressions have notably been used, for their simplicity, to predict the amount of available water with lead time up to several months. Garen[119] proposed some key rules to build a good seasonal forecast, such as the use of only data known at forecast time, the use of cross-validation, principal component regression (PCR) or systemic search of optimal combinations of input variables. Later, Pagano *et al.*[120] used accumulated precipitation and Snow Water Equivalent (SWE) predictors with z-score pre-normalization to set up a daily-updated seasonal forecast of water supply in Western United States. The use of historical data can also be included to improve seasonal forecasts[121].

More complex probabilistic data-driven models have also been investigated more recently, though less actively than other data-driven models. As linear regression models, they aim at describing the relationship between two quantities (here mostly rainfall and runoff) in a probabilistic framework. Variables are here described as probability distributions. As an example, Abdollahi *et al.*[122] proposed to use *Copula* functions to build an event-based rainfall-runoff model. Other approaches using other data than just rainfall can also be found in the literature such as the Probability Distributed Model (PDM) developed by Moore[123,124].

Since the development of bigger and faster computers, neural networks have become more and more used in rainfall-runoff models. The first to propose the use of neural networks was Daniell in 1991[125]. Multi-Layer Perceptrons (MLP) were then largely used in rainfall-runoff models[126–128]. To improve performances in the particular case of rainfall-runoff models, where memory and persistence effects play a major role – with the existence of numerous storage units: snow, soil, glaciers, etc. – Recurrent Neural Networks (RNN) – notably Long-Short Term Memory (LSTM) – have been intensively investigated recently[129–133]. Neural networks can be fed with a variety of input data: localized data

(rain-gauges, temperatures, etc.), spatially distributed data such as radar and satellite[129,131,134] or even results of conceptual or physical rainfall-runoff models in so-called hybrid models[133,135,136].

Finally models based on the fuzzy set theory developed by Zadeh[137] using 'IF-THEN' principle relating input and output variables have been developed recently[138–140]. The overall principles of these models are described by Shrestha *et al.*[141].

2.3.2 Conceptual models

Unlike empirical models, conceptual models try to describe in a simple way the main components of the hydrological cycle. Conceptual rainfall-runoff models are the most widespread models in operational forecast and flood management, as they combine a rational description of physical phenomena with a still reasonable complexity. Conceptual models are usually based firstly on lumped models which can then be converted into semi-distributed or even distributed models by adding up similar lumped models with different parameters representing sub-basins – for semi-distributed models – or cells in a grid – for distributed models.

A first type of conceptual rainfall-runoff model is the unit hydrograph method developed by Sherman in 1932[142] and later Snyder in 1938[143], where only rainfall data in a gauged catchment are needed. This method is based on the historical relationship between the precipitations and the resulting outflow. For a standard precipitation of 1 cm, it is possible to draw an hydrograph representing the evolution of discharge with time. Unit hydrograph models can be developed into semi-distributed and distributed approaches to model flash floods at the level of a complete catchment. For example, by first building the unit hydrographs of several sub-catchments and then summing their contributions for known precipitation, Sharma *et al.*[144] and Goñi *et al.*[145] have been able to estimate correctly the peak discharge of several events respectively in Lower Tapi basin in North-Western India and in the Oiartzun watershed in Northern Spain. Similarly Khaleghi *et al.*[146] used different techniques to generate unit hydrographs and reproduce several events in Northern Iran. Recently, Wang *et al.*[147] used unit hydrographs in a distributed approach. They modelled unit hydrographs for each cell in a spatial grid before simulating the transport of water from cell to cell using geomorphic data to estimate the velocity field. However unit hydrograph models are not widely developed in operational forecast.

The most common models used in operational forecast are conceptual models based on a description of different storage reservoirs and of their connections. Reservoirs can be soil reservoirs (with one or several reservoirs described with different parameters), groundwater reservoirs, snowpack, glaciers, overland interception reservoir (such as vegetation intercepting rain and snow), channels and rivers, lakes and artificial reservoirs, etc. A calibration of the parameters with known runoff data is then required. An example of such a model is described in Figure 2.10. The list of conceptual models used in operational forecast includes: the semi-distributed GR (*Génie Rural*) family (with notably the GR4j model) from France[148,149], *LISFLOOD*[150,151] the distributed model from the *European Flood Awareness System* (*EFAS*)[152], *RS Routing System* developed at *EPFL*[153,154] – and used notably by the *CREALP*, *Hydrique Ingénieurs* or the French *Service de Prévision des Crues* (SPC) of Northern Alps – *HBV* the Swedish lumped/semi-distributed hydrological model[155] in operation notably in Sweden[156], in Russia with the OpenForecast tool[157] or in the Swiss part of the Rhine river[158], etc. A comparison between several large-scale rainfall-runoff models for operational purposes (including conceptual and physical models) was carried out in 2016 by Kauffeldt *et al.*[159].
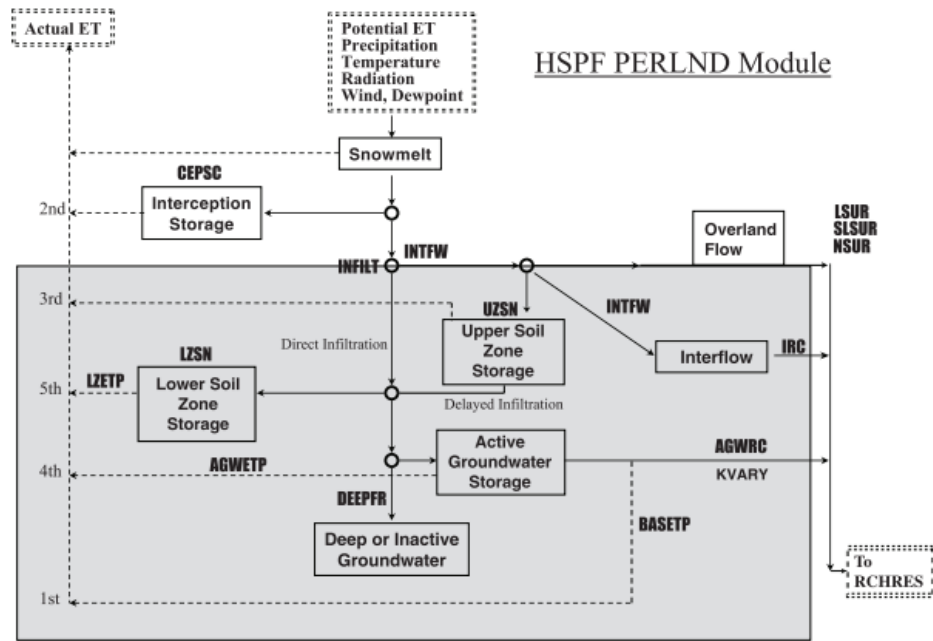
Figure 2.10: Scheme of processes and reservoirs of the conceptual HSPF model from Atkins *et al.*[160]

In conceptual models, the propagation of flow between the reservoirs and between the sub-catchments (for semi-distributed or distributed models) is made through hydraulic or hydrological routing[6]. Hydrological routing can be either level-pool reservoir routing (the discharge leaving a reservoir is determined by the level of the reservoir only) or distributed storage routing in the case of rivers. Hydraulic routing is used in open channels to model the wave propagations. In each cases several models of routing can be used, as described in Figure 2.11. Routing is a major component notably for semi-distributed models such as *Routing System* the rainfall-runoff model used by *Hydrique Ingénieurs.*
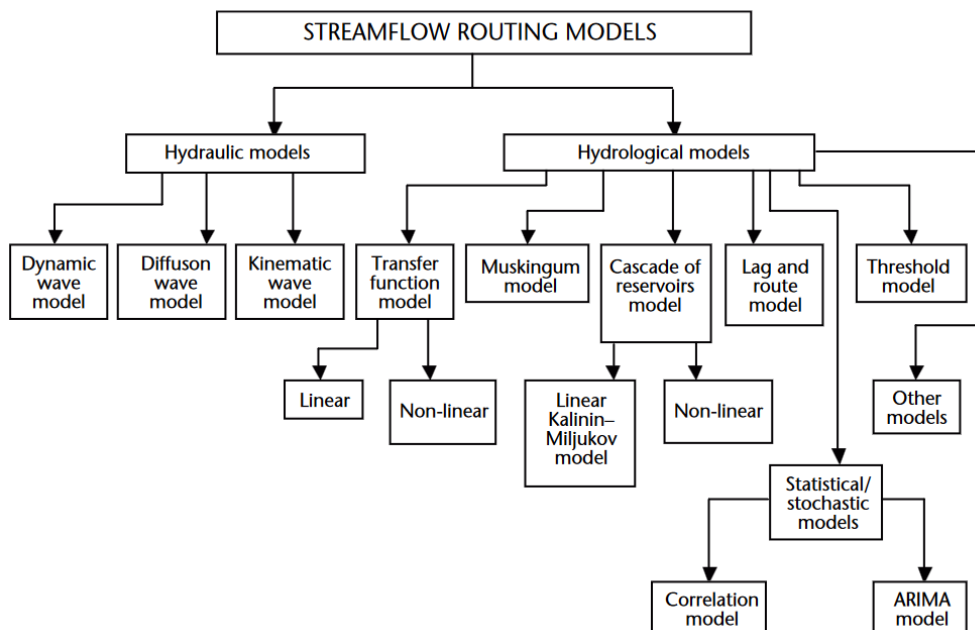
Figure 2.11: Classification of the different routing models used in rainfall-runoff models. Taken from the WMO 2011 report[6].

Conceptual models are aimed at performing in operational forecast, while being as simple as possible in order to be understandable by stakeholders. It is thus essential to choose carefully the physical processes that will be modelled in a conceptual model. These processes depend on the location, size and flood response of the catchment as can be seen in Figure 2.12.

Interaction between basin size, physical process and flood response

Type of basin	Physical process						
	Wind	Infiltration	Rainfall intensity	Runoff	Propagation	Tide and surge	Water table
Urban		X	XXX	XXX	X		
Upper basin		XX	XX	XXX	X		
Long river	X		X	XX	XXX		X
Estuary	XXX				XX	XXX	
Aquifer	X			X		X	XXX

XXX: dominant effect   XX : direct effect   X: minor effect

Figure 2.12: Major physical processes occurring on different types of catchments. Taken from the WMO 2011 report[6].

2.3.3   Physical models

Physical models aim at modelling the most precisely possible most physical phenomena occurring along the hydrological cycle, from hydraulic flow in channels to evapotranspiration and glacier behaviours through fluid flow in porous media. Such models are complex and computationally expensive but present the advantage of working with only few data. They are for example the best models to use in ungauged catchments. Though some of the physical models are semi-distributed, a large number of them are distributed as they do not rely on any sub-basin division (only geomorphological data are usually needed). The incorporation of distributed data such as radar is thus easier in these models.

The frontier between conceptual and physical models is porous and some models can be sometimes classified either as conceptual or physical. The frontier is often defined by the need, or not, of a calibration process: normally, physical models do not require a calibration. However they suppose the knowledge of the real value of some parameters, and thus most physical models are finally calibrated to improve their performances. *LISFLOOD* or *RS Routing System* are two examples of operational forecasting models that are usually classified as conceptual models as they require a calibration but that use similar equations to physical models.

2.3.4   Sources of uncertainty and error in rainfall-runoff models

The errors and the associated uncertainties of rainfall-runoff models come from various processes. The World Meteorological Organization (WMO) listed them in its 2011 special report on flood forecasting[6]:

- Model errors due to forgotten or not perfectly modelled physical processes.
- Parameter errors due to the bad knowledge or calibration of parameters.
- Boundary condition errors due to bad choices of boundary conditions and notably to wrong geometric representation in physical models.

- Initial condition errors due to a wrong estimation of the initial conditions (for example initial height of reservoirs in conceptual models).

- Observation errors of the measurement instruments. This class of errors encompasses input (e.g., precipitation) and output (discharge) errors.

Sources of errors can compensate each other. The calibration process can notably allow parameter errors to compensate other sources of errors. Calibration can be performed manually or automatically using different types of algorithms such as downhill simplex method[161], annealing[162] or genetic algorithms[163]. An automatic calibration is usually more suited for highly-conceptual models with few parameters and is more complicated when the number of parameters, the non-linearity and the discontinuity of the model and the number of local minima increase. For such models, a more complex global optimization is often used with notably the shuffled complex evolution method (SCE-UA)[164,165] which is however longer to converge[166]. To evaluate the performances of an automatic calibration, cross-validation schemes are usually performed. Cross-validation could be performed by splitting the set of events or the period in several subsets but also by cutting the basins into several sub-basins with some being used in the calibration and other in validation[167,168].

To correct errors, an updating process with the latest measures (also called Data Assimilation or DA) can be performed during operational forecasting and is usually yielding significant improvements. The WMO stated that non-updating models are "generally considered to be inefficient in the context of real-time modelling"[6] as they do not use the real-time flow measurements to improve their forecasting skills. Non-updating forecasts may then continuously drift from the measures while an update with the real-time measures will avoid such drift. Event errors in forecast can be classified in three categories: amplitude (or volume), phase and shape errors, as represented in Figure 2.13. Updating aims at reducing these errors. Tangara[169] demonstrated that at short lead time, even badly calibrated models can be saved by the updating process. Nevertheless the calibration and the choice of the model is still the main process to avoid errors at longer lead time. Updating methods can be split between updates of the parameters[169–172] (adaptive calibration) and updates of the initial conditions[169,173–175]. Updating can also be more simply applied as a post-processing of only the final runoff with no modification of the parameters or state variables[176].
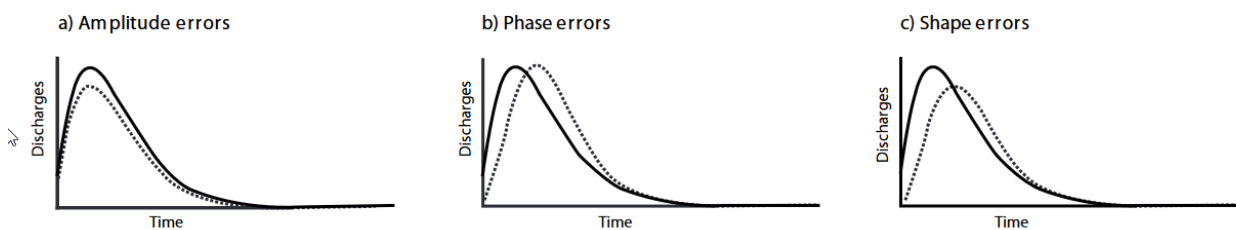
Figure 2.13: Description of the three fit errors in operational flood forecast. Taken from Serban and Askew[177] and the WMO 2011 report[6].

Several techniques can be used in the updating procedure. Wöhling *et al.*[173] used an updating procedure of a distributed *HBV* model[155] where state variables (notably the levels of the upper runoff reservoir) are updated iteratively to minimize the observed-simulated difference ΔQ. They chose to use a spatially-constant factor to modify these levels. Silvestro *et al.*[175] proposed an updating method designed for operational use in Italy with various advices. Their first advice is to only use updating (or post-processing) techniques when "highly reliable" runoff gauges are available. Rather than relying only on instantaneous reference runoff they also recommend to use time-windows in their correction schemes to avoid erroneous observed data. Their updating process is based on two steps. First, a Direct Insertion Method is used to correct water levels. The idea is to exactly invert the equation linking water level and runoff with a corrected value for runoff

(based on observed runoffs over a time window). Secondly they iteratively correct the root-zone soil moisture in order to minimize the runoff bias. In its PhD dissertation, Borup[174] investigated a computationally-efficient method to update water levels ensuring the model stability (to observed discharge errors) without excessively dampening the correction. To do so he used linear regressions to assess the evolution of the last hourly runoff errors and convert it into reservoir level errors[178].

More complex updating techniques rely on Kalman Filters (KF) or Ensemble Kalman Filters (EnKF) to cope with observed data uncertainties[79,170,179,180]. Kalman filter assumes the forecasting process (i.e. the model) is a stochastic process with Gaussian noise. KF aims at finding the best estimation. It is described hereafter with the notations from Liu and Gupta[181]. The prediction and updating steps are described in this stochastic framework by respectively prediction equations (2.1) and (2.2) and update equations (2.3) and (2.4). The former ones introduce the Kalman gain K. By inverting Equation (2.4), K can be obtained and reintroduced in Equation (2.3) to obtain the updated vector of state variables x_{k+1}^+. In these equations x_k represents the vector of state variables at timestep k. Exponents $(-)$ indicate a non-updated state and $(+)$ an updated state. M_k is the matrix representing the prediction step (the model) allowing to go from state $k-1$ to state k. θ represents the parameters of the model and u_k the input data (*e.g.*, precipitation). P_k is the error covariance matrix between the state variables x_k, and Q_k is the error covariance of the model added at each prediction step. Finally H_k is the observation operator at step k while $d_k = z_k - x_k^-$ is the difference between the observation and the simulation.

$$x_{k+1}^- = M_{k+1}(x_k^+, \theta, u_{k+1}) \tag{2.1}$$

$$P_{k+1}^- = M_{k+1}P_k^+ M_{k+1}^T + Q_{k+1} \tag{2.2}$$

$$x_{k+1}^+ = x_{k+1}^- + K_{k+1}d_{k+1} \tag{2.3}$$

$$P_{k+1}^+ = P_{k+1}^- - K_{k+1}H_{k+1}P_{k+1}^- \tag{2.4}$$

Because KF presents the difficulty to be valid only for linear or linearised problems, EnKF was developed to better update non-linear models by performing KF for an ensemble of random input perturbations[181]. Other techniques such as Particle Filtering or Variational Data Assimilation[181] are also aiming at coping with observation and simulation uncertainties. However all these models are more complex than simple empirical updating schemes which are already performing well. Hence, Moore[7] recommended to use the former one in operational flood forecasting. In *RS Routing System*, the updating technique is an empirical iterative method aiming at minimizing the volume ratio between observed and simulated runoff over a chosen time window[79]. It is hence nearly not susceptible to isolated wrong observed runoff data.

On the post-processing side, Ho and Lee[176] proposed to assume their model was well forecasting the difference in runoff. As described by Figure 2.14, they proposed to use the last measure plus the simulated difference in runoff as forecast. Their updating algorithm performed well. Similarly *Hydrique Ingénieurs* is starting to use a post-processing based on polynomial splines to connect observation and forecast.
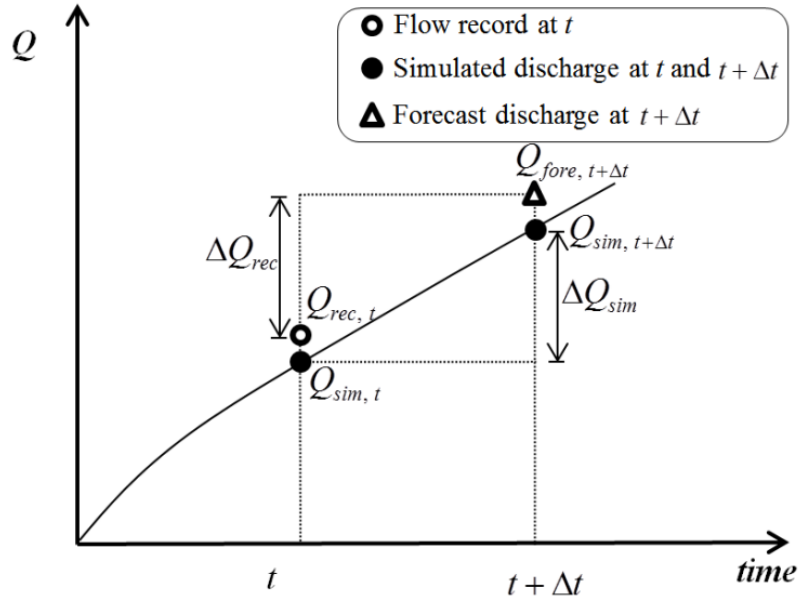
Figure 2.14: Description of the simulated difference based updating algorithm proposed by Ho and Lee[176].

2.3.5 Event-based and improved continuous models for flash floods

Most operational models for flood prediction are based on well-known continuous models – *RS*, *HBV*, *LISFLOOD*, *GRJ...* – used for other applications such as seasonal forecasting, hydroelectricity trading, assessment of the effects of regulation measures or land-use, evaluation of the impacts of climate change, etc. However some models have been specifically developed for flood and flash flood forecasting. It is sometimes necessary to develop such models as other models are often either too simple and not able to accurately predict floods or too complex and not usable in real-time operations. Flash-flood-specific models are usually event-based, distributed models making the best of radar and satellite distributed data. According to the WMO[6], the main issue of these models reside in the lack of sufficient data for their calibration and the high sensitivity to initial conditions. All these models are usually based on the same architecture:

1. Knowledge or computation of spatially-distributed precipitations.
2. Determination of distributed precipitation excess.
3. Transformation of precipitation excess into surface runoff *via* notably empirical Hortonian processes[182,183] or physical Green-Ampt schemes[183].
4. Addition of simplified base flow.
5. Channel and reservoir routing and hydrographs combination at confluence points.

Several examples of flash flood event models exist. The MARINE model[184] (*Modélisation de l'Anticipation du Ruissellement et des Inondations pour des évéNements Extrêmes*) developed by the French National Flood Forecasting Service of Haute-Garonne, is an event-based, flash flood distributed model composed of an hydrological rainfall-runoff module and an hydraulic module solving the *Saint-Venant* equations. This model is designed only for real-time operational flash flood predictions. It is simple enough to be run in a few minutes which is essential in flash flood operational forecasting as mentioned by Montz and Gruntfest[185]. The hydrological module uses the input precipitation from radar data, the nature and utilisation of the soils and the DEM of the terrain to supply flood hydrographs all along the main river to the hydraulic module, which propagates

the flood in the main river. Furthermore, evapotranspiration, interception or lateral underground transfers are neglected as their are not affecting significantly the flood volume and timing for small duration-high volume flash floods, which is supported by previous researches[186–188]. The infiltration and runoff are modelled from precipitation distributed inputs by a Hortonian process[183], though a Green-Ampt model[183] may provide better results. The main difficulty of this model resides in the calibration process as only very few flood data are available prior to the chosen flash floods. Hence it consists mostly of a pre-determination of parameters rather than a real optimisation based on numerous observed data. The number of parameters of the model is reduced as much as possible (three to six parameters). This model is performing quite well, as illustrated in Figure 2.15, and could be improved by a better knowledge of the hydraulic conductivity of the soil and of its initial humidity.
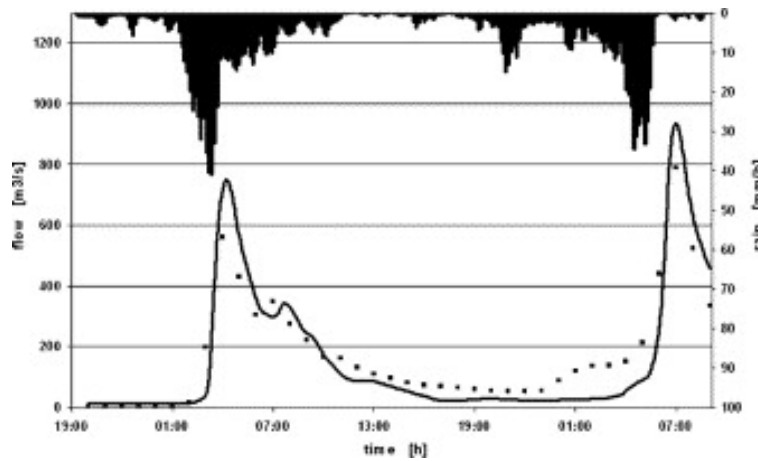
Figure 2.15: Results of the MARINE forecast of the flash flood from 3 to 5$^{\text{th}}$October 1995 in Corbès, France (solid line) compared to the observed discharge (dots). Taken from Estupina Borrell *et al.*[184].

Similarly the Runoff Burroughs Event Model (RORB) developed by Laurenson and Mein[189] for application in Australia includes the same principles of hydrographs estimation and routing for event simulation. It is open-source and the latest version from 2012 is available online.

Multiple comparisons between event-based and continuous models can be found in the literature. Their conclusions diverge and appear to be highly dependent on the model and on the considered catchment, with several studies concluding event-based models performed better[190–192] and several others concluding the contrary[193–195]. It is probable that depending on the model, catchment, quantity of data, etc. one of them will outperform the other. Notably event-based models have the advantage to provide a specific calibration of the parameters for extreme events with the drawback of being calibrated on only few data unlike continuous models.

In between continuous and event-based models, some continuous models can be used with added features to improve flash-flood forecasting. The first idea could be to calibrate the model specifically on flood events. Such approach has been followed by De Silva *et al.*[191] which used the parameters of an event-based model to feed a continuous model. Similarly, Hossain *et al.*[192] used a continuous model with calibration on only events as their event-based model. In the case of this project, though the *Routing System* model used at *Hydrique Ingénieurs* is a continuous model, the calibration procedure will rely both on continuous indicators and on event-based indicators calculated for discharges higher than a threshold.

Besides an event-specific calibration, other features have been sometimes added for flash-flood prediction. For example, Peredo *et al.*[196] have recently tried to improve the skills of the semi-distributed *GRSD* model (part of the French *GR* family) during winter and spring floods in the Aude catchment. In this scope they have modified the proportion η of the net rainfall P_n (after accounting for

evapotranspiration) that is directly leaving as surface runoff without being infiltrated and stored in the soil. In the original *GRSD* model, which is described in Figure 2.16, this factor was determined only by the square of the filling factor of the soil storage (Equation 2.5). Here the authors add a term depending on the intensity of precipitations to increase this factor for high intensity rainfall. With this modified factor (Equation 2.6), more water bypass the soil reservoir and is converted into direct runoff during high rainfall events. Such modifications allow to better model the events and notably the timing of the peak discharge.

$$\eta_{ini} = \frac{P_r}{P_n} = \left(\frac{S}{X_1}\right)^2 \tag{2.5}$$

$$\eta_{new} = \frac{P_r}{P_n} = \frac{\left(\frac{S}{X_1}\right)^2 + 1 - exp\left(-\frac{i}{i_0}\right)}{2 - exp\left(-\frac{i}{i_0}\right)} \tag{2.6}$$
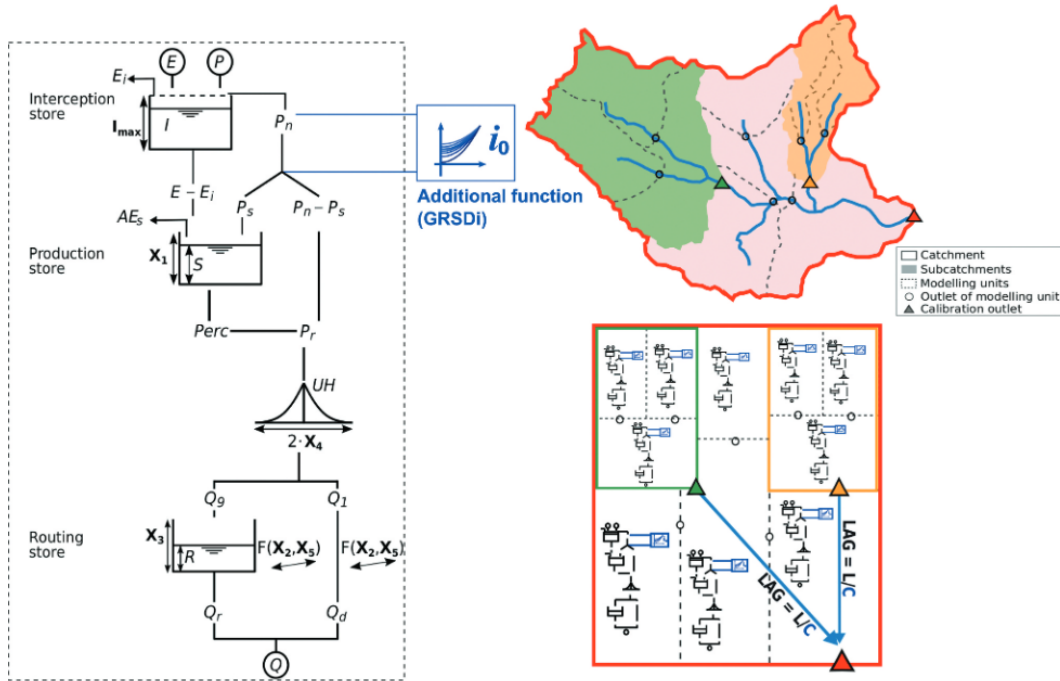
Figure 2.16: Scheme of the modified GRSD model. The infiltration rate P_s/P_n is modified to be smaller for high intensity of precipitation. Taken from Peredo *et al.*[196].

2.4 Flash flood alarm systems

Once a discharge prediction is produced, it needs to be analysed and transferred to the public and to decision-makers. Indeed providing a single discharge value does not bring any information if it does not come with reference information, risk and probability assessment, damage and action costs, etc. Hence alarm systems are developed to treat the raw forecasted discharge and convert it into usable information. Alert systems development is an active field of research, largely connected to operational purposes.

2.4.1 Class of alarm systems

The World Meteorological Organization[6] lists four categories of alarm systems by order of complexity:

- Threshold-based flood alerts. These systems are based only on observed discharges. The measures are compared to reference threshold discharge to generate some levels of alerts. The determination of reference discharge is here essential and will be discussed in Section 2.4.3. These systems require the discharge to be highly persistent so that the discharge in the next hours can be easily predicted from the last measures and will evolve slowly enough to let authorities act. It is mostly adapted to long response time catchments, but is also used when simplicity is required (*e.g.*, basins with too few locations/periods of floods where the investment does not worth it, under-developed regions with less resources for complex alarm systems, etc.).

- Flood forecasting alert systems. Here the forecasted discharge are compared to reference threshold discharges. Rainfall-runoff models of various types can be used with various input data (see Section 2.3), depending on the catchment characteristics and resources available. For example in small response time catchments it is essential to use forecast rainfall products while in large response time basins it could be enough to use only rainfall observations to save money and resources.

- Vigilance maps. The previous flood forecasting alert systems can be replicated in a large number of locations to generate a map of alerts. Several levels of alert can then be defined. This is for example the case of the French operational flood alarm system *Vigicrues*[197].

- Inundation forecasts. The forecasted discharges are used to generate a map of areas that will be at risk of being flooded. This requires a digital elevation model (DEM) and a hydro-dynamic (or hydrological) level-and-flow model. A large number of historical data are also needed as these systems are highly sensitive, notably in plain areas. The distinction between flood forecasting (*Prévisions de crue* in French or *Hochwasser* in German) and inundation forecasting (*Prévisions d'inondations* in French, *Überschwemmungen* in German) is essential to evaluate the areas at risk of damage during a forecasted event.

A fifth category can be added and consist in the meteorological alarm systems based only on observed or forecasted rainfall in catchments where the hydrology is mostly based on response to rainfall.

2.4.2 Meteorological alarm systems

The simplest alarm systems are meteorological alarm systems. Here, only current information on precipitations are necessary. One example of such system is the French system APIC[197,198] (*Avertissement Pluies Intenses à l'échelle des Communes*). It is an operational alarm system sending SMS and e-mail warnings to registered towns and cities when precipitations exceed historical thresholds. Two levels of warnings are generated: for precipitation levels for a 10 years response time ("high

flood risk") or for a 50 years response time ("very high flood risk"). Precipitations are measured using radar and rain gauge measurements and are provided only on places with high-quality QPEs[199]. Information are also provided to the public *via* the APIC website[200].

The Flash Flood Guidance (FFG)[201] is a more complex system based on the determination of a threshold runoff R generating a flooding flow Q_p. Alerts are launched when the accumulated precipitation over a given period of time exceeds the threshold runoff. Hence, rather than forecasting the discharge in the river, the alert system uses only the value of recorded precipitations. Q_p and R are connected *via* the unit hydrograph theory following Equation 2.7, where q_{pR} is the unit hydrograph peak and A the catchment area.

$$Q_p = q_{pR}RA \tag{2.7}$$

Q_p is determined from historical data, as explained in Section 2.4.3, while q_{pR} is obtained from the unit hydrograph theory quickly described in Section 2.3.2. Several methods based on historical data exist for the determination of q_{pR}[201]. Notably GIS and DEM information allow to reduce the uncertainty associated to q_{pR}. Measured soil moisture values can also be integrated in q_{pR} determination. Once Q_p and q_{pR} are chosen, R is determined from Equation 2.7 and alerts are generated when accumulated rainfalls exceed its value. The FFG is today widely used for its simplicity and its development is notably supported by WMO in several under-developed countries all around the world, some being highly vulnerable to flash floods[202].

More complex meteorological alarm systems can finally be based on precipitation forecasts and not only on precipitation observations. These systems are more complex to design and subject to more uncertainty but they allow to increase the anticipation time of the alert which is determinant to effectively protect the goods and people. Such complex systems are however not largely developed worldwide and QPFs are mostly used in human-controlled warning system such as in the warnings systems of *MeteoSwiss* and *MeteoFrance*. In both cases there is no automatic systems based on QPFs. QPFs could be useful both for long lead time and large catchment where they help to activate protection measures, and for smaller lead time and smaller catchments where evacuation orders can be given in case of a high risk of flash floods on towns at risks.

2.4.3 Determination of reference discharges

Reference threshold discharges are needed to generate alerts in hydrologically-based alarm systems. Their appropriate choice is determinant. The cases of gauged and ungauged catchments need to be studied separately.

For gauged catchment, the classical procedure is to use historical data as reference discharge. Several methods exist and are associated with different risks and consequences[201]:

- First, a conservative approach to limit damages is to use the bankfull discharge. This definition is physically-based but is conservative as a bankfull discharge is usually not enough to cause damages. The bankfull discharge can be estimated from Manning-Strickler formula[201].

- Another solution is to use historical flows for given return periods. This solution is statistical as it is based on discharge observations and as flows for long return periods have not always already been observed. To estimate flows for given return periods, the Gumbel law on extreme values is used[203]. According to Carpenter *et al.*[201] and Henderson[204], flows for return periods of one to two years are usually well correlated to fullbank discharge. The two-year return period discharge is usually higher than the bankfull one. This method allows to design an alarm systems with several levels of alert (for several return periods) as in Vigicrues Flash in France (two levels of alerts for two return periods). This method has the drawback that it is not associated to damage assessment: in some places, damages can occur for small return periods while in others they will not occur even at high return periods[205].

The choice between a conservative or a larger discharge alert threshold is important to limit the cost of potential damages but also of protective measures such as sandbags, evacuations or mobilization of civil protection agents. The trade-off between damages and alert costs is well illustrated by utility curves such as the one from Figure 2.17 taken from the WMO's 2011 report[6].
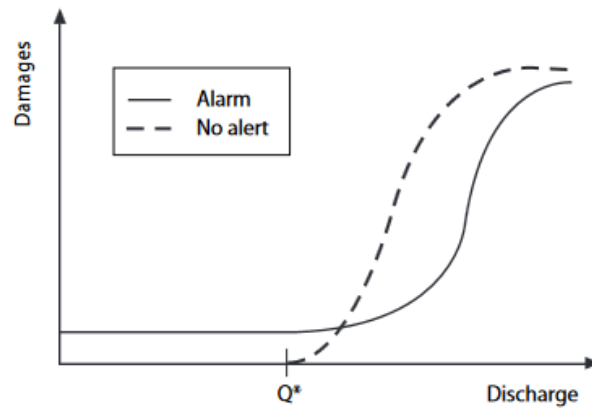
Figure 2.17: Example of a utility function for a flood alert decision. The dashed line representes the cost of potential damages depending on the real discharge and the solid line represents the costs occurring when an alert is launched. The latter are composed of inevitable damages and of costs associated with protection measures. Taken from the WMO 2011 report[6].

In ungauged catchments, the bankfull dicharge is still an available solution to fix discharge threshold as it does not need the knowledge of historical discharges. Another solution could be to estimate the discharge during major flood events by post-observation of damages, satellite data or interviews with citizens. An example of such field inquiry has been performed by Javelle *et al.*[199] for a flash flood on the Argens catchment in Southern France. However an easier solution is to run simulation with a large number of scenarii to estimate the simulated distribution of discharges and consequently the simulated discharges for given periods of time. Younis *et al.*[206] used 12 years of meteorological data in the Cévennes-Vivarais region of Southern France to run the LISFLOOD model and generate a cumulative distribution function of simulated discharge. From this distribution they extracted four reference thresholds for alert: *severe* alert for the highest simulated discharge, *high* alert for the discharge exceeded 1% of the time, *medium* alert for the one exceeded 2% of the time and *low* alert for the one exceeded 3% of the time. Similarly, in a case study in Ticino, Alfieri *et al.*[207] used historical meteorological data in a re-forecast process with the European Flood Alert System (EFAS). They later extracted the annual simulated discharge maxima and define three threshold levels: *medium* alert for the mean of annual maxima, *high* alert for the 5-year return period simulated discharge, and *severe* alert for the 20-year return period simulated discharge. These techniques have the advantage of eliminating part of the model bias: if simulated discharges are always twice as big as real (unknown) discharges, alert will still be generated correctly.

Finally in both gauged and ungauged catchments, alert systems could be generated according to the value of other physical quantity than discharge. The following quantities, observed or forecasted, alone or in combination, could be used:

- Discharge or discharge increase rate
- Water level or rate of rise of level at significant locations
- Rainfall intensity or cumulated level (meteorological alert system)
- Catchment wetness conditions

Complex alarm systems can then be developed by combining several of these quantities in a given case study. The design of such alert system needs to be developed with local authorities, taking

into account the local context and the potential damages[6]. In the case of a complex alert system, several stages of alerts (moderate, high, severe, etc.) are usually provided.

2.4.4 Management of uncertainty

The main issue of alert systems is the high uncertainty associated with weather forecasts and rainfall-runoff models. The risks and stakes of false alarms or missed events are high and need to be managed carefully. Several solutions have been proposed to cope with uncertainty. First the design of several warning stages allows to inform authorities at an early stage when uncertainty could be high. Indeed, severe events are not often observed largely in advance but can very often be forecasted with smaller intensities at large lead time. Hence a moderate stage of alert can often mean that there is a risk to reach a higher stage of alert. This is way to inform authorities that a severe event *may* happen.

Uncertainty can also be introduced *via* ensembles of forecasts. These ensembles could be produced either by using QPF ensembles – or artificially modified QPFs when the QPF is deterministic – before the use of a rainfall-runoff model or by using several rainfall-runoff models. Alfieri *et al.*[207] used the European Flood Awareness System (EFAS) based on a meteorological ensemble forecast (ECMWF) and on the LISFLOOD rainfall-runoff model to create an ensemble of discharge forecasts. Probability distribution can then be extracted from the ensemble results allowing to evaluate the uncertainty of the model at each timestep. An example of forecasted discharge with confidence interval is displayed in Figure 2.18.
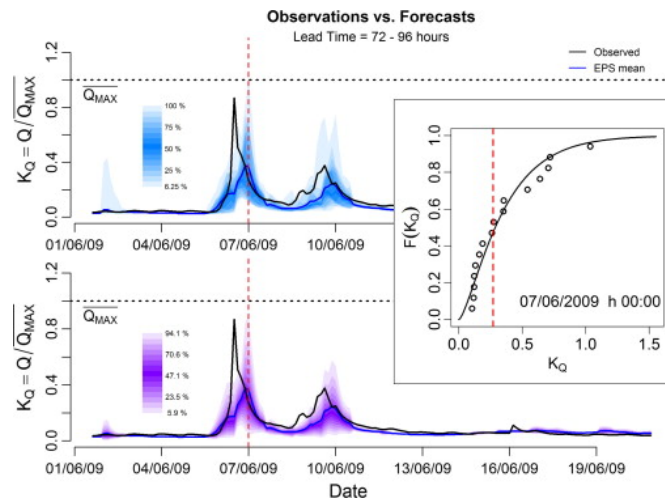
Figure 2.18: Example of a forecasted discharge with confidence interval. The upper and lower plots are respectively for lead times of 72h a 96h. A cumulative distribution plot of discharges is also provided for one timestep in the right plot. Taken from the Alfieri *et al.*[207].

Finally, uncertainty could be introduced by analysing several consecutive forecasts, called *lagged forecasts*. Bartholmes *et al.*[208] demonstrated that the use of persistent alert systems where a threshold must be exceeded in two consecutive forecasts for an alert to be launched, largely improved performances, for both deterministic or ensemblist forecasts. Alfieri *et al.*[207] used *lagged forecasts* to double the number of members of their ensemble. They showed that *lagged forecasts* are well adapted to ensemblist approaches. Decisions can for example be taken on the level of probability of exceeding a threshold in several consecutive forecasts. However *lagged forecasts* imply a loss of time required to act due to the wait time between two alarm forecasts. This is very detrimental for flash floods on small response time catchments.

Once uncertainty is integrated into an alert system, it is very important to communicate with public authorities to form them on the notion of uncertainty, as pointed by the WMO[6] and by

Berthet *et al.*[209]. Indeed, a very critical step is the transformation of a probabilistic alert into deterministic actions by authorities. Actions must be taken at the right moment knowing the risks of acting or not-acting. Berthet *et al.*[209], after noticing that the level of uncertainty is imposed and not chosen and that its reduction is difficult and limited, proposed a series of recommendations for the communication and management of uncertainty. The idea beyond these recommendations is that uncertainty must not hinder the decision process.

- The production of forecasts and the decision process must be dissociated. Indeed the hydrometeorological forecast is an element among others, which is, by definition, uncertain. Uncertainty increases with lead time and for extreme events, which, unfortunately, are the most important events. Hence other elements such as the knowledge of damage hazards and costs or the planning of mitigation and prevention actions are features to take into account in the decision process.

- Uncertainty must be well communicated. The most largely spread solution is to provide a range of values or a confidence interval rather than a single deterministic value.

- Decision makers must be formed to manage uncertainty. Rules and tools helping the decision must be set up before the occurrence of an event. The WMO also stressed the importance of training by dedicating one chapter of its 2011 report to this problematic[6].

- Automatic uncertainty estimation processes are required. The estimation of uncertainty cannot be empirical as forecasters are frequently changing.

2.4.5 Official alert systems in several countries

Flood and flash flood alert systems have been developed worldwide, notably on regions susceptible to these events. However their step of development and characteristics are highly variable and dependent on the economic development of the country with more operational, automatic and country-covering systems in developed countries. Some examples of alert systems are quickly described below.

Europe

The European Flood Awareness System (EFAS) is an alert system covering the European continent and operational since 2012[116]. It is based on the ECMWF NWP ensemble (European Centre for Medium-range Weather Forecasts)[a] fed into the LISFLOOD rainfall-runoff model[207,208,210]. Early flood warnings are then analysed and transmitted to the hydrological services participating in the development of EFAS. Its ensemblist character allows to introduce uncertainty in the alert systems.

Switzerland

In Switzerland, the flood vigilance is the role of the Federal Office of Environment (OFEV). The alert system is based on current observations of discharge at discharge stations. Five alert levels are used depending on the discharge return period: No danger, limited danger (above Q_2), significant danger (above Q_{10}), high danger (above Q_{30}) and very high danger (above Q_{100}). Small adaptations are made when more knowledge on the potential risks and damages are known. The map is available at `https://www.hydrodaten.admin.ch/fr/messstationen_gefahren.html`.

France

France has developed three meteorological and hydrometeorological alert systems to cover catchments of different sizes and response times[197]:

- APIC (*Avertissement Pluies Intenses à l'échelle des Communes*, since 2011). APIC is a meteorological alert system based on observations only and proposed by *MeteoFrance*. Measured precipitations are compared to historical data to generate two levels of alert (*high* for a return

---

[a]which is used to generate the COSMO NWP that will later be used in this project.

period of 10 years and *severe* for a return period of 50 years). Precipitations are measured by radar and rain gauges with the AIGA system. SMS and e-mails warnings are sent directly to the registered local public authorities and a map is available for the public on the Internet (`https://apic-vigicruesflash.fr`).

- *Vigicrues*: a non-automatic system of flood prevision and surveillance for major rivers and catchments (high response time). 23 000 km of the 270 000 km of French rivers are covered by *Vigicrues*. It is updated twice a day at 10h and 16h and more frequently during flood events. Forecasts are based on several distinct hydrological models depending on the region and are performed by regional SPCs (*Services de prévision des crues*). All forecasts are centralised by the SCHAPI (*Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations*) and available at `https://www.vigicrues.gouv.fr`.

- *Vigicrues Flash* (since 2017): an automatic system of flood forecast for the remaining small ungauged rivers. The distributed rainfall-runoff model GRD, developed by the INRAE, is fed with AIGA data to provide information to local authorities[199]. As it is focused on small response time, this system provides mostly alerts for flash floods. Alerts are available at `https://apic-vigicruesflash.fr`.

United Kingdom

United Kingdom alert system is under the responsibility of each country. Hence two systems have been developed[211]. In Scotland, the FEWS (Flood Early Warning System) is based on radar-rain gauge data fed into the G2G physical-conceptual distributed rainfall-runoff model[212]. Flood alerts are available on the Internet (`https://floodforecastingservice.net`). The Scottish system is interesting as Scotland is constituted mostly of small catchy basins.

In England and Wales, flood alerts are generated by the National Flood Forecasting System of the Environment Agency. Catchments have usually higher response times than in Scotland hence alerts are mostly provided at regional level.

United States

The flood alert is part of the missions of the NOAA (National Oceanic and Atmospheric Agency) and of the National Weather Service. It is based on the HEFS (Hydrologic Ensemble Forecast Service) and provides forecasts with lead times higher than 6 hours. It is not a unified model as a variety of rainfall-runoff models are used such as *Snow-17*, *SAC-SMA* or models based on the unit hydrograph approach[116].

Flood alerts are generated by several distinct systems. Flash Flood Guidance (FFG) is notably used[199] but other alert systems based on discharge forecasts can also be used locally. A website, available to the public, provides the latest cumulated precipitations on thousands of rain gauges and their respective precipitation FFG thresholds (`https://www.weather.gov/serfc/ffg`).

India

Flood forecasting, which is of central importance in India, is supported by the Central Water Commission (CWC). Forecasts (level or discharge) are produced for a network of 199 stations covering the whole country. Models depend on the basin but most of them are statistical models[116].

Nepal

Some operational forecasting systems have also been developed in countries under development such as Nepal. The Nepali flood forecasts are available at `http://www.hydrology.gov.np`. It indicates the risk of floods and the state (rise or fall) of the discharge. It is based on the WRF (Weather Rainfall Forecast) NWP fed into a semi-distributed rainfall-runoff model MIKE[116].

2.5 Improvement of flash flood forecasting with ensembles

The most promising way to improve flood forecasting and alarm systems resides probably in ensemble approaches. Indeed, the main issue in operational flood forecasting is the uncertainties arising mostly at three levels in the process: uncertainties of the meteorological observations and predictions, uncertainties of the initial conditions, and errors introduced by rainfall-runoff models[213]. For this reason, ensembles is a very active field of research in meteorology and flood forecasting. A large review of ensemble techniques for streamflow forecasting has recently been published by Troin *et al.*[213].

2.5.1 Types of ensemble approaches

In discharge forecasting, ensemble approaches cover a large variety of models. Figure 2.19 describes the different types of ensemble approaches. They can first be separated in three main families, depending on the type of input data used:

- Statistics-based streamflow prediction (SBSP) which do not use any weather data but only historical streamflow data. They rely usually on data-driven rainfall-runoff models such as artificial neural networks, autoregressive models, fuzzy logics models, etc. One or several models can be used to generate a probabilistic ensemble of predictions.

- Ensemble Streamflow Prediction (ESP) using only historical observations, either meteorological or streamflow data. One or several hydrological models can be used. These techniques are usually well suited for seasonal and long-range streamflow predictions, where the weather uncertainty dominates other sources of uncertainty. The most famous ESP is the Extended Streamflow Prediction proposed by Day[214]. In this approach, the assumption that historical meteorological events are representative of future events is made. One recent example of use of this technique was performed by Olsson *et al.*[215] for seasonal forecasting of spring-flood in Sweden.

- Ensemble Prediction Systems (EPS) based on observations and forecasts. NWPs are added to the previous data in these ensemble systems. They are the more interesting for flash flood predictions and small response time catchments.
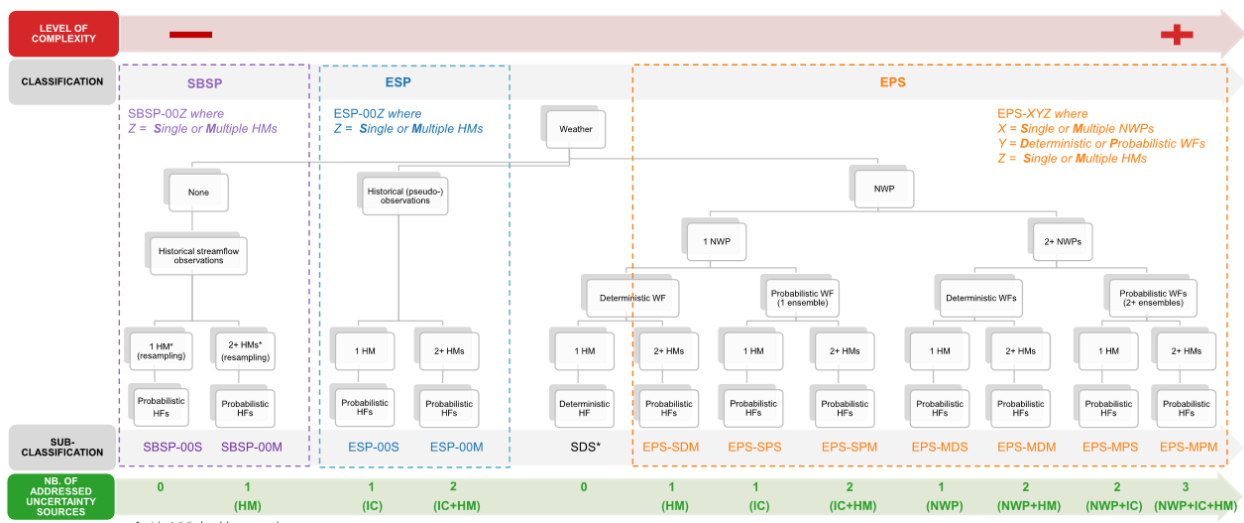
Figure 2.19: Classification of the different types of ensemble models for streamflow prediction. Taken from Troin *et al.*[213].

The EPS can themselves be split between several families, depending on the number of hydrological models used (HM), and on the number and nature – probabilistic or deterministic – of NWPs. The case where one rainfall-runoff model is used with one deterministic NWP is called SDS (Single Deterministic NWP Single HM). It is the only system which is not probabilistic. In this report, all the analyses will be based on an SDS approach as only the median member of *MeteoSwiss* NWPs are available. The other systems, all probabilistic, are listed below, from the simplest to the most complex:

- EPS-SDM (Single Deterministic NWP Multiple HMs): several models are used with only one deterministic weather forecast as input.

- EPS-SPS (Single Probabilistic NWP Single HMs): only one model and one NWP is used but the NWP is an ensemble of weather forecast members. This is the main type of ensemble systems used worldwide[213].

- EPS-SPM: several models are used for one probabilistic NWP.

- EPS-MDS and EPS-MDM: one or several models are used with several distinct deterministic NWPs.

- EPS-MPS: several probabilistic NWPs are fed into the same hydrological model.

- EPS-MPM: this class of systems is the most complex where several probabilistic NWPs are fed into several rainfall-runoff models. It is the full-blown multi-model multi-ensemble concept. Only very few studies have been performed with such system, usually at a long range and lead time, as it is time and computationally costly. One example is the Pan-European multimodel seasonal system developed by Wanders *et al.*[216].

2.5.2 Main features used to reduce uncertainty

Ensemble systems allow to introduce new features to reduce the uncertainty of a streamflow forecast. All these features are similar to those used for ensembles of weather forecasts. Three main families exist:

- Data assimilation. It refers to all techniques used to integrate the input data into a rainfall-runoff model. For example the *update* of initial conditions which is strongly recommended by the WMO[6] is a type of data assimilation. Several data assimilation techniques exist such as Kalman filters, variational methods, particle filtering, etc.[213]

- Aggregation of forecasts. Once a hydrological model has been used, the different forecasts produced need to be combined, with the goal of reducing the final uncertainty. Aggregation techniques can be either based on constant or dynamic deterministic weighting method, or on probabilistic combination into a density forecast. The blending of radar Lagrangian persistence and COSMO NWP performed by *MeteoSwiss* (Section 2.2.3) is an example of dynamic deterministic weighting for weather forecasts that can be copied for streamflow forecasts.

- Post-processing. It covers all techniques used to reduce errors, bias or dispersion of the ensemble members in the end of the forecasting process. It could notably be based on regression or Bayesian approaches.

Each of these three aspects constitutes a deep research field where specific reviews have been published. The propagation of uncertainty through the ensemble models is rendered possible by these features. It constitutes a significant field of research notably in Switzerland[217].

In the case of flash flood forecasting, a good example of the use of ensembles is developed by Alfieri *et al.*[207] in Southern Switzerland (see Section 2.4.4).

Chapter 3

Methodology and data

3.1 MeteoSwiss products

MeteoSwiss, the Swiss meteorological agency, provides several meteorological products, for both observation and prediction. In this project, precipitation, temperature and radiation will be needed. For the latter two, measures are taken from the *SwissMetNet* network of meteorological stations while forecasts are provided in the COSMO1-E (temperature) and COSMO2-E (radiation) NWPs that will be described hereafter. For precipitations, several products will be used and studied.

3.1.1 Quantitative Precipitation Estimates (QPEs)

Quantitative Precipitation Estimates measure precipitation either at localized points (rain gauges) or all over a given area (gridded radar products). *MeteoSwiss* provides several QPEs for Switzerland.

3.1.1.1 Rain gauges

MeteoSwiss has developed *SwissMetNet*, an automatic network of standardized meteorological stations covering the whole Switzerland[218]. The network contains nearly 200 rain gauge stations. Precipitation data are provided with 10 minutes or one hour timestep. The *SwissMetNet* stations can also provide several other meteorological quantities such as temperature and radiation, depending on their type.

3.1.1.2 Raw radar QPE

From 2011 to 2016, the Swiss radar network has been modernized and completed. It is now constituted of 5 C-band radars covering the different regions: La Dôle, Albis, Monte Lema and the new Plaine Morte and Weissfluhgipfel radars. The precipitation field resulting from the combination of measurements in these 5 radars is available in less than 60 seconds[11,37]. It is aggregated over 60 minutes and called AZC. Data are available since 2014 but changes on the network happened until 2016 with the introduction of the Weissfluhgipfel radar. The spatial resolution of AZC is 1 km.

The AZC product is also available in a 5-minute disaggregated form[83]. In that case it is called RZC. In this project, the smallest timestep that will be used is 10 minutes. Hence, when needed, the RZC product coming directly from *MeteoSwiss* will be aggregated over 10 minutes.

3.1.1.3 *CombiPrecip*

MeteoSwiss is also proposing a radar-rain gauge merging product called *CombiPrecip* (also denoted CPCH). It is the result of a co-kriging with external drift between the AZC product and rain gauges from the *SwissMetNet* network[78]. The process to generate *CombiPrecip* from the raw radar product is described in Section 2.2.2. It has evolved over time but the data used in this project

have been regenerated according to the latest algorithm. The process of generation of CPCH is thus homogeneous. However the number of stations used in the kriging step has risen from 181 stations in 2014 to between 262 and 269 stations since 2017. As for AZC, the spatial resolution of CPCH is 1 km. CPCH also exists both in its original 1h-timestep form and in a disaggregated form with 5-minute timestep[83]. In this project, the latter will be aggregated into a 10min-timestep QPE.

3.1.1.4 inca0

Finally, during this project, the first timestep of the *MeteoSwiss* radar QPF INCA will be often used as observation. It is not a real observation as it is taken from a forecast but it may be quite close to it while giving information on potential biases of the INCA QPF. It is also expected to be close to CPCH as it is based on it since 2017. It is available at *Hydrique Ingénieurs* since August 2018 with a 10-minute timestep. It will be denoted inca0.

3.1.2 Quantitative Precipitations Forecasts (QPFs)

This report uses mainly two types of precipitation forecasts (QPFs): a radar QPF (INCA) and COSMO, the usual Numerical Weather Prediction (NWP) from *MeteoSwiss.*

3.1.2.1 COSMO NWP

The COnsortium for Small scale MOdelling (COSMO)[105] is a European research consortium, created in 1998, aiming a developing a non-hydrostatic limited-area atmospheric model. It is used both for research and operational purposes and notably develops several Numerical Weather Prediction products. Today, two ensemblist NWPs are available with an hourly timestep:

- COSMO1-E is constituted by 11 members with a 1.1 km spatial resolution. Its computation timestep is 10 seconds and it is updated 8 times a day (every three hours). It has a usual lead time of 33h, except for the forecast issued at 3h UTC which has a 45h lead time. COSMO1-E also provides temperature forecasts which are the ones used in this project. COSMO1-E replaced the deterministic COSMO1 NWP, updated twice a day, in the summer 2020. As the focus of this report is only on deterministic forecasts, the median member of the COSMO1-E will be used during its ensemblist period and the deterministic COSMO1 before that.

- COSMO2-E, with a higher lead time but which is less precise, is constituted by 21 ensemble members with a spatial resolution of 2.2 km and a computation timestep of 20 seconds. It is updated twice a day at noon and midnight UTC and has a 120h lead time (five days). In this project, the median member of COSMO2-E will only be used for the radiation measurements when required (*ie.* in the case of the Emosson catchment in Section 5.2).

3.1.2.2 INCA

INCA (Integrated Nowcasting through Comprehensive Analysis)[100] is the radar QPFs provided by *MeteoSwiss.* It is also ensemblist but only the deterministic median member is provided by *MeteoSwiss* up to now. It results from the blending of COSMO QPFs with the storm advection radar QPF, as described in Section 2.2.3. Its available timestep is 10 minutes, its spatial resolution 1 km and its lead time 6 hours. It is updated every 10 minutes which makes it very interesting for flash flood predictions. In hourly analyses, INCA will be aggregated over 1h. INCA data are based on the archives of forecasts received by *Hydrique Ingénieurs* since 2018. Therefore, they are not homogeneous as several upgrades of INCA happened during the period of the archives (notably with the introduction of *NowPrecip* in 2019[92]). In this project Inca is only providing precipitation forecasts and will be complemented by COSMO1-E for temperature and COSMO2-E for radiation.

3.2 RS Routing System

3.2.1 History

Routing System, hereafter called RS, is the rainfall-runoff model developed and run by *Hydrique Ingénieurs*. It has first been developed in the *Laboratoire de Construction Hydraulique* (LHC) of the *Ecole Polytechnique Fédérale de Lausanne* (EPFL). It has since evolved independently at the CREALP (*RS Minerve*) and at *Hydrique Ingénieurs*.

RS was initially created in the LHC in 2000[219] to model the routing of floods in hydraulic networks. It included an hydrological rainfall-runoff model (GR3), a flood routing module adapted for routing in catchments, rivers or reservoirs and modules to add discharges at confluences. In 2007, during the MINERVE[a] project[154], new features, such as snow and glacier modules, were added to create *Routing System II*. All modules were developed during the CRUEX[b], SWURVE[c] and MINERVE projects.

3.2.2 General features

RS is a semi-distributed physical-conceptual model. It is mainly constituted of two kinds of modules: hydraulic modules for the routing of water and hydrological modules. Catchments are first split in several sub-catchments, as illustrated in Figure 3.1a which represents the four sub-catchments of the Emosson model and their links. Each sub-catchment is itself cut into hydrological parts that are transferring water downstream to routing modules. Figure 3.1b describes some of the hydraulic modules present in a sub-catchment. Here the water, coming from the hydrological models 'Tri1.1' is routed through a kinematic wave river section. A diversion then splits it between two junction modules so that a part can be compared to the measured discharge, in the '*UP*' comparator. The comparator allows to calibrate the parameters of the model. Other hydraulic modules are available such as reservoirs, spillways, river routing modules through St Venant or Muskingum-Cunge models, turbines or pumps.

The hydrological parts of a sub-catchment are split between several altitude bands that generate the initial runoff. Usually altitude bands covered 300 m of altitude (0-300 m, 300-600 m, etc.). Bands are separated between glacial areas and non-glacial areas. Figure 3.1c provides an example of altitude bands organisation with three altitude bands, two with glacier and one without it. The hydrological processes follow the GSM-SOCONT model (Glacier Snow Melt - Soil CONTribution model)[220] which is illustrated by Figure 3.1d. Meteorological variables are first interpolated on the altitude bands in the virtual station modules. In landscape (non-glacial) areas, the precipitation i and the temperature T are then sent to the snow module which allows to determine the input flow to the soil GR3 module. At the outlet of the soil module infiltrated water is sent to the river outlet while runoff water is transferred first to the SWMM model. In glacial areas, inflow water from the snow module is running into the glacier module. Part of the flow is then leaving directly to the river outlet while the rest is sent to the GR3 module of the landscape area. The main features and parameters of these modules will be described in the next section.

---

[a]Model for the prevision and management of floods by preventive operations on hydroelectric reservoirs
[b]Modelling of extreme floods in Alpine catchments
[c]Study of climate change impacts on Alpine catchments

(a)

(b)

(c)

(d)
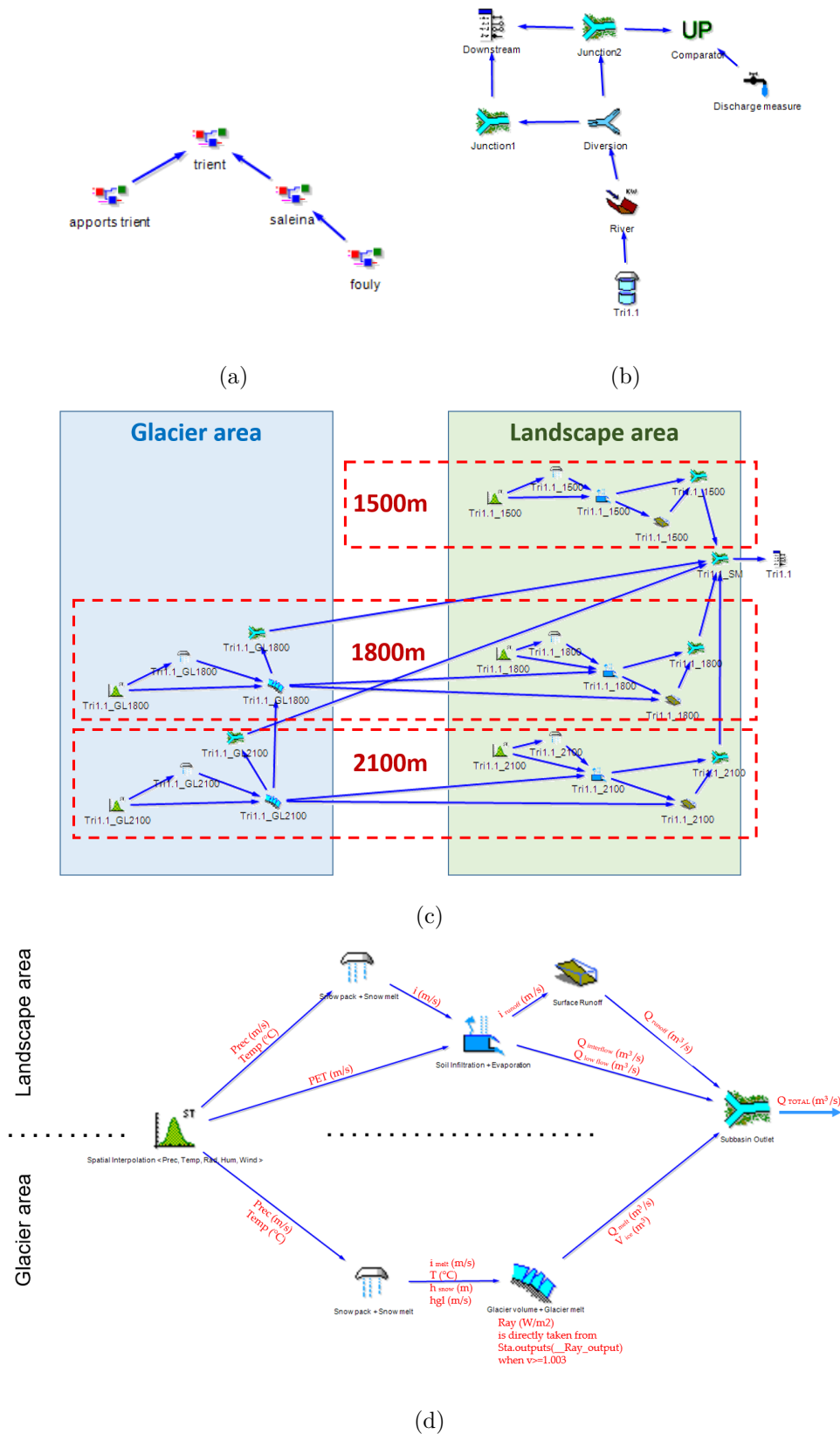
Figure 3.1: General organisation of a *Routing System* model: 3.1a. division of the catchment into several sub-catchments - 3.1b. example of sub-catchment with diversion, junctions, kinematic wave river routing and comparator modules - 3.1c. example of three altitude bands (1500m only snow, 1800m snow and glacier and 2100m snow and glacier) with the cascade of modules (virtual station, snow, glacier, junctions, SWMM and GR3 modules) - 3.1d. Scheme of an altitude band with the GSM-SOCONT cascade of processes (from *Hydrique Ingénieurs* internal documentation).

3.2.3   Modules and parameters

3.2.3.1   Virtual station

Meteorological data (precipitation, temperature and radiation) are interpolated in each altitude band in the virtual stations. Weights are given to each input data station. They can be either manually chosen or calculated according to the distance between the real station and the altitude band. Meteorological data can be corrected globally and according to the altitude, thanks to the parameters listed in Table 3.1. Concerning the altimetric gradient of temperature, it has usually a smaller absolute value in wet conditions than in dry conditions. Hence a critical precipitation intensity is introduced to separate the two weather regimes. Once precipitation, temperature and radiation are determined, the evapotranspiration can be estimated and transmitted to the soil GR3 model.

Table 3.1: Main parameters of the virtual station.

Parameter	Unit	Description
Virtual station		
coeffI	-	Multiplicative precipitation coefficient
coeffT	°C	Additive temperature correction
coeffETP	-	Multiplicative evapotranspiration coefficient
gradI	m/s/m	Altimetric gradient of precipitation
gradT (Dry/Wet)	$°C\,m^{-1}$	Altimetric gradient of temperature - values for dry and wet weather
iCrGradTHum	$mm\,h^{-1}$	Critical precipitation intensity between dry and wet weather

3.2.3.2   Snow band

The snow band converts the input precipitation into a virtual precipitation which is later transferred to the soil or to the glacier models. It uses the height of the snowpack as state variable and updates it. The main parameter of the snow model is the rate of snow melt which is expressed as the height of snow that is melt in a day and for each degree celsius above $0\,°C$. The snow melt can also be due to radiation which is controlled by the *SRF* parameter[d].

Table 3.2: Main parameters of the snow band.

Parameter	Unit	Description
Snow band		
An	mm/°C/d	Snow melt rate due to temperature
SRF	mm/W/d	Snow melt rate due to radiation

3.2.3.3   Glacier

The glacier is modelled by two main reservoirs: the glacier itself and a snow reservoir describing the snow cover of the glacier. When the snow and ice are melting they are not directly transferred to rivers. Hence the coefficient *KGL* (resp. *KN*) represents the fraction of the water from glacier melting (resp. snow melting) that is transferred to rivers. The parameters *Agl* and *SRF* control the rate of melt of the glacier due to respectively temperature and solar radiations. A slow reservoir,

---

[d]although this parameter will mostly be put to 0 in this project.

with characteristics similar to the fast reservoir of a GR3 model, can also be added (its parameters are not significantly modifying the simulations).

Table 3.3: Main parameters of the Glacier model.

Parameter	Unit	Description
Glacier model		
Agl	mm/°C/d	Glacier melt rate due to temperature
KN	-	Transfer coefficient of the snow reservoir
KGL	-	Transfer coefficient of the glacier reservoir
Ray	-	Multiplicative coefficient of radiation
SRF	mm/W/d	Glacier melt rate due to radiation
Slow reservoir		Similar characteristics to the GR3 fast reservoir

3.2.3.4 GR3 soil model

The GR3 model is used to describe the hydrological processes occurring in the soil. It is composed of two reservoirs (one fast, one slow) controlled by water level state variables. A scheme of the GR3 main processes is provided by Figure 3.2. The input equivalent precipitation can either flow directly as surface runoff or be infiltrated into the GR3 fast reservoir which water level takes values between 0 and *Hmax*. Hence water can leave the GR3 model by three processes:

- Water can first flow into the river without entering the fast reservoir. If $h < Hmax$, part of the water is infiltrated (Equation 3.1) and the rest leaves directly as surface runoff. Moreover when $h > Hmax$ (fast reservoir full), all the incoming water is sent to the river. In both cases, the surface runoff first passes by the SWMM model described in the next section.

- The water that has been infiltrated (according to Equation 3.1), can then leave the fast reservoir and flow directly to the river. This inter-flow *iFast* is controlled by a transfer coefficient K and an exponent *aK* according to Equation 3.2.

- Finally, water can also be sent to a slow reservoir and later flow directly to the river according to an expression similar to the fast reservoir outflow. As for the glacier model, the slow reservoir is not mandatory and its parameters are not influencing a lot the models, notably the high discharges.

$$i_{inf} = i_{eq} \left(1 - \frac{h}{Hmax} \right)^{aInf} \tag{3.1}$$

$$i_{fast} = K \left(\frac{h}{Hmax} \right)^{aK} \tag{3.2}$$

The main parameters are described in Table 3.4.

Table 3.4: Main parameters of the GR3 model.

Parameter	Unit	Description
GR3		
Hmax	m	Maximum water level of the GR3 reservoir
K	-	Transfer coefficient of the GR3 reservoir
aK	-	Exponent applied at the outlet of the GR3 reservoir
aInf	-	Exponent applied to the infiltrated flow
Slow reservoir		Similar characteristics to the fast reservoir
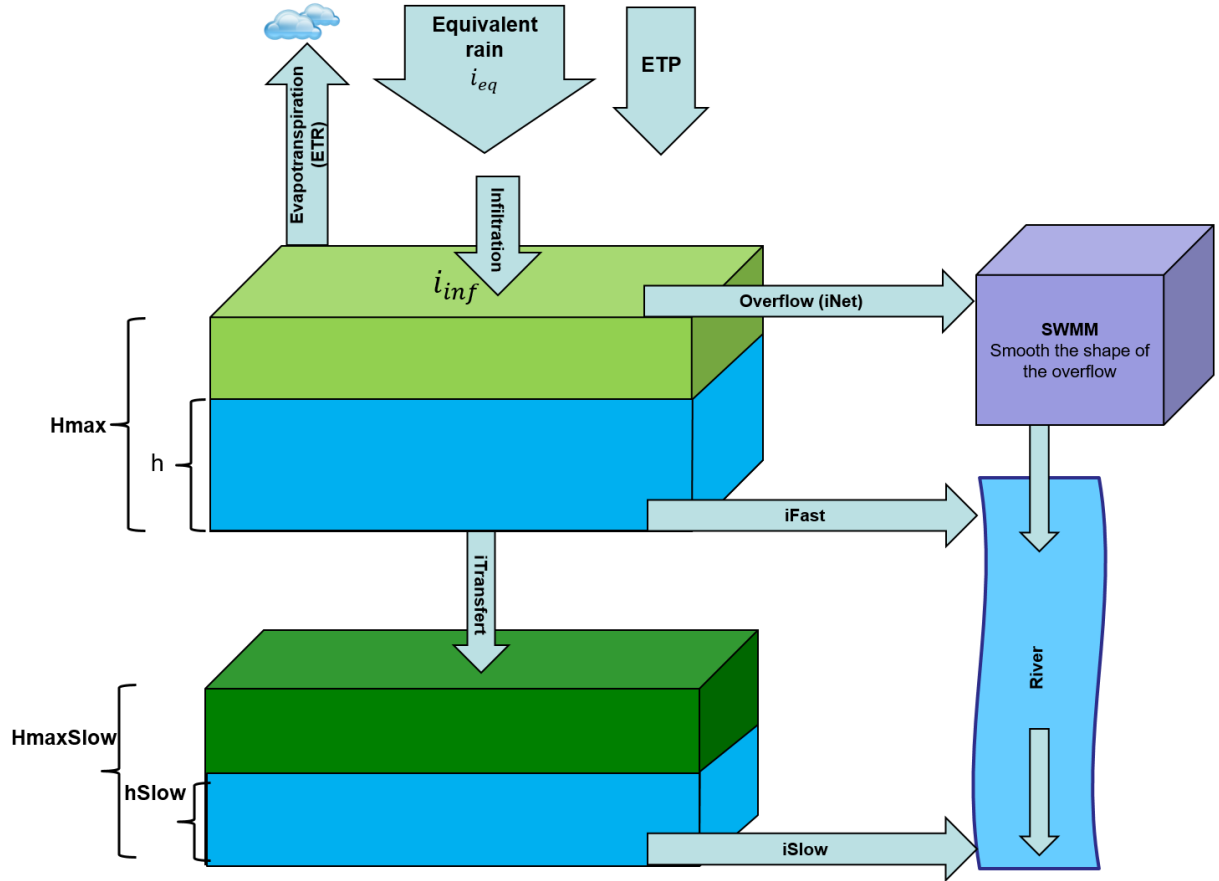
Figure 3.2: Scheme of the reservoirs and main processes of the GR3 model.

3.2.3.5 SWMM model

The SWMM model (Storm Water Management Model) is used to smooth the direct surface overflow from the GR3 model. It is based on Manning-Strickler's equation and hence is characterized by a Manning coefficient K_s. Moreover, the total area of the altitude band can be divided into a normal permeable area and an impervious area. The proportion of impermeable area is controlled by the coefficient Imp. On the impervious zone, infiltration is null and water is flowing directly into the SWMM model with a higher Manning coefficient than on the permeable zone. Hence the Imp parameter also affects the GR3 model.

Table 3.5: Main parameters of the SWMM model.

Parameter	Unit	Description
SWMM		
K_s	$\mathrm{m}^{1/3}\,\mathrm{s}^{-1}$	Strickler coefficient
Imp	-	Fraction of impervious surface

3.2.4 Calibration of the model

The first step on each catchment is to build a model including all physical phenomena of importance for the discharge estimation. Once this is done, the parameters of the model need to be optimized. This optimization is performed sub-catchment by sub-catchment, from upstream to downstream. Only sub-catchments where a reliable discharge measurement is available are used. For each sub-catchment, each parameters of each module can be modified.

To realize a good calibration, the focus should be on several distinct elements of the discharge curve including: the base level, the seasonal variations, the daily amplitudes – due to snow and glacier melt – and their variations, the shape, timing and levels of peak discharges, etc. To guide the calibration several indicators are also provided. In this report we focused mostly on obtaining:

- A *Volume Ratio* close to 1. The volume ratio corresponds to the ratio between the cumulated volumes of the simulation and of the measure. It ensures that the water balance of the observed discharge is respected in the simulation.

- The best possible *Nash-Sutcliffe* values (later called Nash). The Nash is the mean squared error of the simulation, as expressed in Equation 3.3 where Q_{obs} and Q_{sim} represent respectively the observed and simulated discharges and $\overline{x}$ represents the mean of quantity x. A Nash value of 1 represents a perfect simulation. In the calibration process, we will use either global Nash or Nash values computed only for observed discharges higher than a chosen threshold. Such indicator allows to focus more on large discharges than on small ones, less interesting for the prevision of floods. It allows to include event-based characteristics in the model by calibrating the parameters mostly for high discharge events. *Log Nash-Sutcliffe* values have not been used in this project as they tend to give less importance on higher discharges.

$$Nash = 1 - \frac{\sum_{t=t_0}^{t_{end}} \left(Q_{obs}(t) - Q_{sim}(t)\right)^2}{\sum_{t=t_0}^{t_{end}} \left(Q_{obs}(t) - \overline{Q_{obs}}\right)^2} \tag{3.3}$$

The model can then be evaluated thanks to several metrics described in Section 3.3.2.3.

3.2.5 Hindcast

Once a model is calibrated and evaluated on continuous data, it is important to simulate the processes used in operational mode with forecasted data, in order to assess its benefices on possible warning systems. Figure 3.3 illustrates the two steps of a hindcast process:

- First, from the continuous control simulation used to calibrate the model (in orange - past - and purple - future - in Figure 3.3), an update of the initial conditions is performed. This step allows to come closer to the runoff observed at river gauges and known in real time (red curve). It will be called hereafter the *update* step and provides a continuous set of updated initial conditions.

- Knowing the initial conditions at each moment, a hindcast process – also called historical re-forecast – is performed. Precipitation, temperature and radiation forecasts are fed into the model with the initial conditions. At every timestep, a forecast of predicted discharges is created over a given lead time. At the end of this process numerous small datasets are available (one per forecast). In Figure 3.3, three different forecasts (blue, yellow and green) are displayed at two consecutive timesteps.

During the *update* step, initial conditions are iteratively corrected so that the ratio between the cumulated volume simulated and observed in the last 24h equals 1. Depending on the model the update period can also be smaller than 24h. The *update* step is highly recommended by the WMO[6] in operational forecast and has already been demonstrated to be essential in all catchments in the first intermediate report of the *Radar4Infra* project[221]. Hindcasts will be done for several different input forecasts:

- Perfect forecasts produced by feeding the model with observed input data for the next hours. Of course these forecasts cannot be used in operational mode as the "future" observed data are not available at the time of the forecast. They allow to estimate the model performances if QPFs and temperature forecasts were perfect. In this project perfect forecasts will be produced using either CPCH, AZC (or RZC) or rain-gauge data.

- Non-filtered forecasts based either on INCA QPF – with COSMO1-E and COSMO2-E forecasts for respectively temperature and radiation – or COSMO1-E QPF.

- Filtered forecasts. A filtering process will be performed to connect the previous forecasts to the last runoff data measured by river gauges. The filter starts from the last observations, uses the same initial slope and connects the forecast at a chosen filtering time using a cubic spline. The *update* and *post-processing* steps are illustrated together in Figure 3.4.
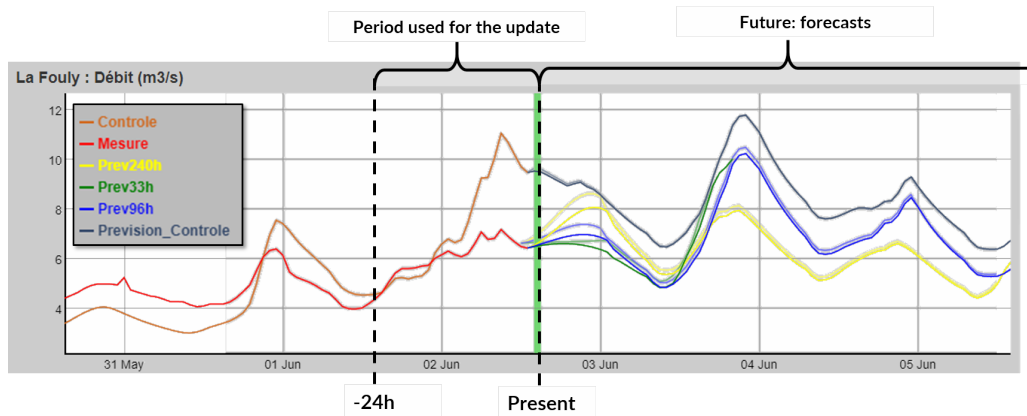
Figure 3.3: Example of simulation curves. Present time is marked by the green vertical line. In the past, the measured discharge is in red and the control simulation (without update) is in orange. The control simulation using precipitation forecasts (but still without update) is in purple. Several forecasts are provided for different lead times (in blue, yellow and green). Thanks to the update step, the forecasts are a lot closer to the measured discharge than the non-updated control simulation.
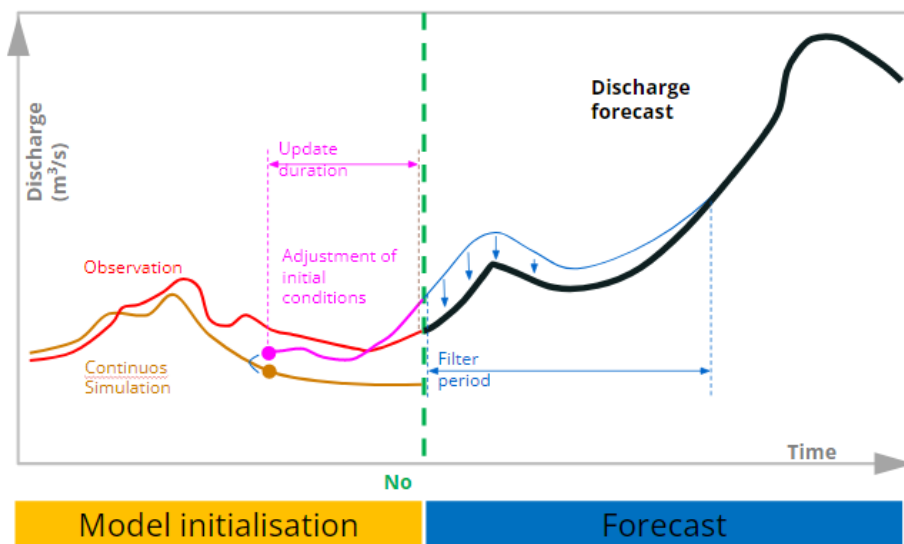
Figure 3.4: Update and post-processing: the model (orange) initial conditions are first corrected to get closer to the observation (red). Updated discharge is in pink. A forecast is then produced (blue) and post-processed to be linked to the observation resulting in the black forecast.

3.3 Analyses and indicators

3.3.1 Radar benchmark: Measurement skills at rainfall stations

In a first part of this project, the radar QPEs provided by *MeteoSwiss* will be evaluated to determine what are their characteristics, potential biases and interest for flash flood rainfall-runoff models. This benchmark will be mostly performed at *MeteoSwiss* rain gauge locations as the *SwissMetNet* rain gauges will be used as reference.

3.3.1.1 Data

The three radar QPEs provided by *MeteoSwiss* and presented in Section 3.1 will be compared:
- AZC, the raw radar data, aggregated with an hourly timestep.
- *CombiPrecip* (also called CPCH), the radar-rain gauge merging product.
- inca0, the first timestep of the INCA QPFs, corresponding to precipitation forecasted for the first 10 minutes.

The reference rainfall measurements are the rain gauge data provided by *MeteoSwiss*. The three radar QPEs are provided for a grid of points covering all Switzerland (and part of the neighbouring countries). The spatial resolution is 1km*1km. To be compared to rain gauges, the gridded QPEs are interpolated at the stations using the 1, 4 or 9 closest points with either equal weights (average) or squared inverse distance weights. Except in Section 4.6 (where it will be 10 minutes), the timestep used in the analysis is one hour.

3.3.1.2 Periods and stations analysed

181 stations are considered for the analysis. The map of these stations is shown in Figure 3.5. In Section 4.6, where an analysis with a 10-minute timestep will be performed, only 145 stations will be used as 10-minutes data are not available for the other stations.
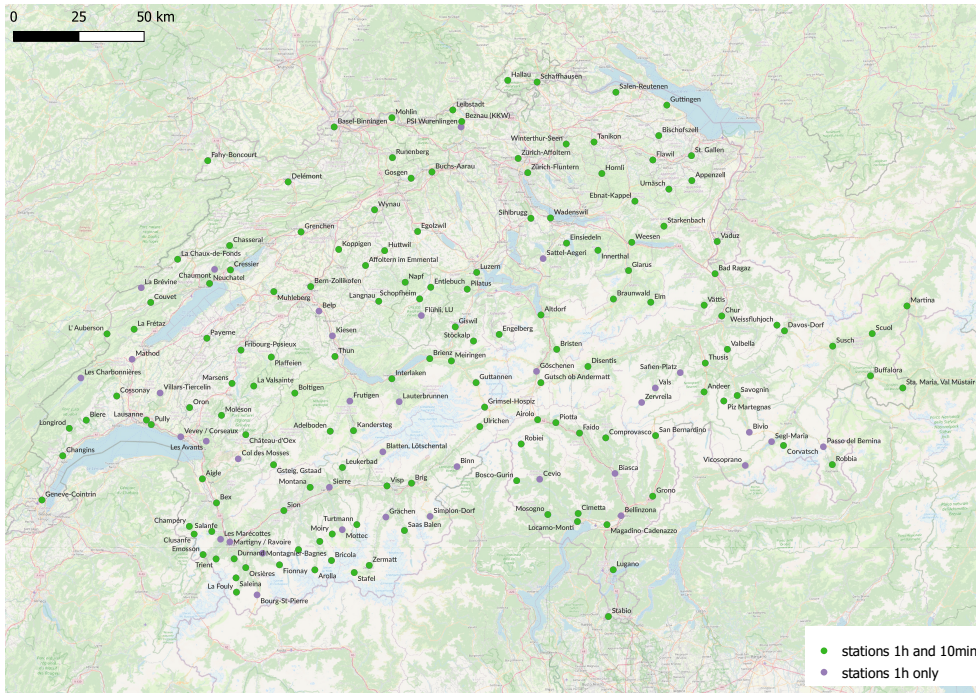
Figure 3.5: Stations included in hourly analyses. The 145 stations available also for the 10-minute analysis are coloured in green.

The period of analysis lasts from 17/07/2018 00h00 to 28/09/2021 00h00 local time. It corresponds to the largest period where the data were already available for all three products. A longer period of time lasting from 2014 to 2021 will be used in Section 4.5 to compare the CPCH and AZC performances evolution over years.

3.3.1.3 Analyses

The analysis of radar QPEs will be split into six sections:

- Section 4.2: As different interpolation techniques can be used to estimate the radar precipitation intensity at rain gauge stations, a sensitivity analysis of the interpolation method will first be conducted to choose the interpolation method for the following sections.
- Section 4.3: The performances of the three radar QPEs will be evaluated at the scale of Switzerland.
- Section 4.4: The best product will then be used to observe if its performances depend on the location of the stations (altitude or region of Switzerland).
- Section 4.5: The historical evolution of radar performances will be assessed to determine a period of time for hydrological calibration.
- Section 4.6: The performances of radar data for a 10-minute timestep will be evaluated.
- Section 4.7: A simple validation of the performances of the co-kriging far from the *SwissMetNet* stations of *MeteoSwiss* will also be performed thanks to cantonal rain gauges in Bern and Ticino cantons.

3.3.1.4 Indicators

Various indicators, detailed hereafter, will be used in each of the aforementioned sections:

- HIT-FAR scores which provide information on the agreement between radar or rain gauge QPEs for the detection of events.
- Mean precipitation analysis, allowing to detect systemic biases and their seasonal dependency.
- Complementary cumulative distribution functions (CCDF) which give information on the bias related to the intensity of precipitation.

Most of the analyses will be based on the average of the previous metrics over a set of stations. However it is also possible to compute these metrics for a given station alone.

HIT-FAR scores

The HIT score represents the proportion of events observed by the reference data (here rain gauges) that is also observed by the "simulated" data (here radar QPEs). It is comprised between 0 and 1 and its best value is 1. Conversely, the FAR score corresponds to the proportion of events observed in radar data that are not observed in the rain gauges. It is also between 0 and 1 and should be equal to 0 for a perfect agreement between radar and rain gauges. HIT and FAR scores can be obtained from the table of contingency (Table 3.6) with Equations 3.4 and 3.5 respectively:

$$HIT = \frac{N_a}{N_a + N_b} \tag{3.4}$$

$$FAR = \frac{N_c}{N_a + N_c} \tag{3.5}$$

Table 3.6: Table of contingency for HIT and FAR scores calculation.

		Event simulated	
		Yes	No
Event observed	Yes	hit: N_a	missed: N_b
	No	false alarm: N_c	no event: N_d

Events are determined by precipitation threshold, meaning an event is a moment where the precipitation intensity is higher than a given threshold. Several HIT-FAR scores for different precipitation thresholds will be computed and presented in a common graph.

To determine the number of hit N_a and false alarms N_c, a tolerance will be used, as described in Figure 3.6. Tolerances of 10% and 30% will be used. The rainfall intensity thresholds are 0,0.5,1,2,3,4,5,6,7,8,9,10,15,20,25,30,35,40,45,50 in mm h^{-1}. The HIT/FAR scores are computed by event, meaning that events exceeding the threshold are first separated before being described as reference, simulated, hit or nofar events (as in Figure 3.6). An event is delimited thanks to a base to peak separation method. In the present case, a ratio of 80% is used, meaning the event is constituted by all the timesteps around the event peak where the precipitations are higher than 80% of the peak value. Additionally two consecutive events must be separated by at least 4 h. To be considered a hit, an event must exceed the threshold at the same time that the reference event exceeds its threshold. **The HIT-FAR metric provides information on the number of events observed simultaneously by radar or rain gauge products or missed by one of them.**

In the case of the 10-minute timestep analysis, a temporal tolerance is added. An event can be considered a hit if it occurs between 40 minutes before the reference event and 20 minutes after it. The time of each event is the time when it reaches the threshold. This temporal tolerance is not included in hourly analyses.
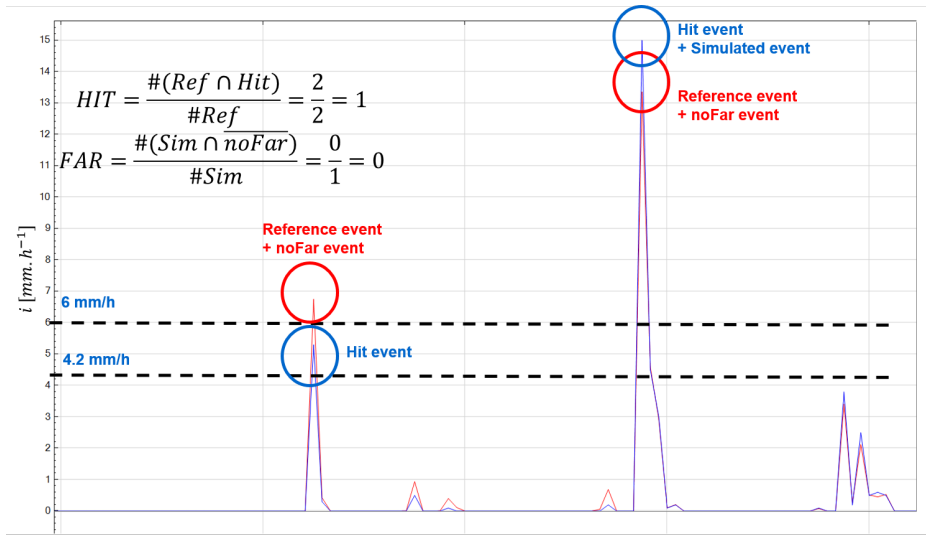
Figure 3.6: Determination of hit and far events. The rainfall intensity threshold is 6 mm h^{-1} with a tolerance of 30%. The reference data (rain gauge) are in red and the simulated data (from radar) are in blue. Reference and simulated events are the events exceeding the 6 mm h^{-1} threshold while a simulated event (respectively a reference event) is a noFar event (resp. a hit event) if the reference data (resp. simulated data) exceeds the 4.2 mm h^{-1} threshold corresponding to a 30% tolerance.

Mean precipitation analysis

In the mean precipitation analysis, the average hourly precipitations for each month will be computed and plotted. In this report, the ratio between the mean precipitation for the radar data (AZC, CPCH and inca0) and the reference data (rain gauge) will be used as it removes the seasonal variations of precipitation intensity. **This analysis allows to detect systemic bias and their seasonal dependency.**

Complementary cumulative distribution functions (CCDF)

For each precipitation intensity, the percentage of time this intensity is exceeded is computed. Only time steps where at least one of the four product (CPCH, AZC, rain gauges and inca0) contains non null precipitations are considered. In order to facilitate the analysis, the ratio between the CCDF of each radar product and the one of rain gauges will be plotted for each intensity value. CCDF (probability of exceedance) is used instead of cumulative distribution function (CDF, probability of non-exceedance) as its ratio shows more differences between the different products. **This metric provides information of bias related to the intensity of precipitation.**

3.3.2   Improvements of rainfall-runoff models with radar data

Knowing the performances of radar QPEs – which also gives information on the radar QPF INCA as it is linked to inca0 and CPCH – it is possible to develop rainfall-runoff models taking the most of radar data. Notably, radar QPFs distributed character and their good skills at low lead time may help to forecast flash floods. Flash flood rainfall-runoff models will thus be set up using *Routing System* for two different catchments. For each of them the focus will first be on calibration of models with historical continuous observations. In a second time operational-like re-forecasts of historical storm events will be simulated.

3.3.2.1   Data

All *MeteoSwiss* data will be used in this part of the report. To introduce localized precipitation products in RS (rain gauges), each virtual station of each altitude band interpolates the data from the closest stations – which can already be quite far. On the contrary, for gridded products, they are introduced directly in the virtual station of the band they belong to, making the most of the spatial information included in these products.

Two catchments, with different characteristics will be studied:

- A mountainous catchment (Emosson - Section 5.2). This catchment is part of the *Radar4Infra* project and important for *Alpiq*, a partner of the project.
- A countryside catchment in the Swiss plateau, with some impermeable soil areas (Zurich airport - Section 5.3).

The quickly responding urban catchment of Lausanne will not be studied as it is already based on INCA.

3.3.2.2   Analyses

For each of the two catchments, the analysis will be split in two parts

1. A continuous calibration and evaluation of the RS model over a multi-year period will first be developed.
2. Major events will then be re-forecasted with the calibrated RS model. The prediction performances of the forecasts will then be analysed.

These two steps are linked and RS continuous models may be modified and improved when the re-forecast (or hindcast) appears not to be good enough.

3.3.2.3   Indicators

The two previous steps of the analysis require distinct indicators.

Continuous data performance analysis
 For the continuous analysis the metrics used are the HIT/FAR scores, already presented in Section 3.3.1.4. The discharge thresholds will depend on the catchment. Two types of discharge HIT/FAR scores can be computed depending on the type of catchment. For non-mountainous catchment with no seasonal behaviour (Zurich), the HIT and FAR scores are computed by event, meaning that events exceeding a threshold will first be separated, before being described as reference, simulated, hit or nofar events (as in Figure 3.7). As hourly precipitation, to be considered a hit, a discharge event must exceed the threshold within a given temporal window around the reference event. This time window will usually be of 40 minutes before and 20 minutes after the reference event. For hourly timestep the temporal tolerance will thus be null.

For mountainous catchments (Emosson) where the snowpack and the glaciers play a determinant role creating a strong seasonal curve – where daily variations of the discharge are smaller than its seasonal variation – the event separation is however difficult. Daily HIT-FAR scores are thus used. They are described in Appendix D.1.
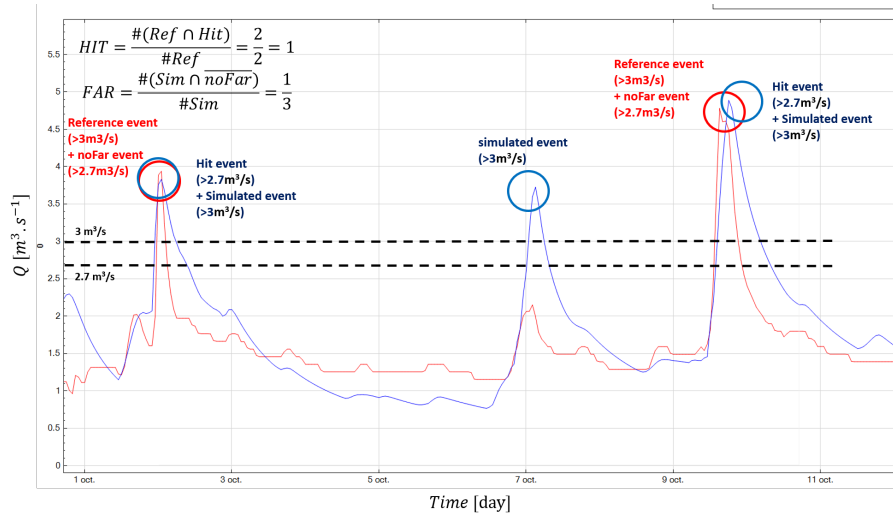
Figure 3.7: Determination of hit and far events. The discharge threshold is $3\,\mathrm{m}^3\,\mathrm{s}^{-1}$ with a tolerance of 10%. The reference data (river gauge) is in red and the simulated data from RS is in blue. Reference and simulated events are the events exceeding the $3\,\mathrm{m}^3\,\mathrm{s}^{-1}$ threshold while a simulated event (resp. a reference event) is a noFar event (resp. a hit event) if the reference data (resp. simulated data) exceeds the $2.7\,\mathrm{mm}\,\mathrm{h}^{-1}$ threshold corresponding to a 10% tolerance.

Forecast performance analysis

Hindcast performance analysis will be based only on a **selection of events** as the hindcast process is time-costly as it produces one dataset every 10 minutes. To assess the forecast performances this report will lean on alert HIT-FAR scores. These scores evaluate the quality of an alert system. An alert is supposed to be launched for a given threshold if **any** forecast predicts that the threshold will be reached within the **first 90 minutes** of the forecast. Points located later on the forecast are not used to generate alerts as the focus is on very-short lead times. An event is considered to be a hit event if it has been observed **in the forecasts provided between 90 minutes and 20 minutes before** the measurement crossing of the threshold – forecasts with 10 minutes or 0 minute anticipation are considered to be provided too late. The predicted time of the crossing must also **be strictly less than 20 minutes after** the observed threshold crossing (temporal tolerance). A predicted event is a far event if the reference does not reach the threshold value **between the moment the alert is issued and 90 minutes after the predicted crossing**. As before, a discharge tolerance of 10% could additionally be applied. A description of some alert HIT-FAR computations is provided in appendix in Figure A.1. The computation process of hit and far scores is also described in Figure A.2. Only the events crossing the threshold will be considered in the HIT-FAR computation (if a reference event is always bigger than the threshold it will not be considered).

The focus of this report and of this Master Thesis is to improve the generation of flash flood alerts. However, it is also interesting to know the future inflow cumulated volumes, notably for hydroelectricity purposes. This is the case in Emosson. Analyses of cumulated volumes are out of the scope of this report but are essential for some partners of the *Radar4Infra* project such as *Alpiq*. Hence volume analyses will be provided in Appendices D and E.

It needs to be noted that distinct HIT-FAR scores (*e.g.*, daily, by event, for alerts, for cumulated volumes, etc.) cannot be compared between them. Indeed each HIT-FAR score is designed for a different purpose. For example alert scores inform on the performances of a given alert system based on a chosen model. On the contrary cumulated volume scores inform on the ability of the system to predict volumes, daily scores determine the performances on the prediction of daily maxima, etc. The usual levels of HIT/FAR scores in each cases are largely different and largely dependent on the basin and its characteristics. A 80% HIT score for daily discharge in Emosson will be poor performance while it will be good for alerts.

3.4 Characteristics of the catchments

3.4.1 Emosson

3.4.1.1 Presentation of the catchment

The Emosson basin is located in the Western part of the Valais, close to the French border. The catchment is illustrated by Figure 3.8 together with the meteorological stations used in the model. The part of interest is here the sub-basins where water is diverted to the Collecteur Est (the longer pipe lying from South-East to North-West). The total area of this catchment is $108\,\mathrm{km}^2$ covered notably by $27\,\mathrm{km}^2$ of glacier. The catchment has also a complex orography with altitudes ranging from $1548\,\mathrm{m}$ (Trient intake) to $3898\,\mathrm{m}$ (Aiguille d'Argentière). The approximate response time of the whole basin is between 4 and 6 hours but some of the sub-basins have smaller response time. Annual precipitations vary between $1400\,\mathrm{mm}$ and $2000\,\mathrm{mm}$ according to the Swiss hydrological atlas website[222].

Five discharge stations are available before the Collecteur Est: La Fouly, Saleina, Trient, Collecteur Est and Val Ferret intakes. The calibration of the model is made on the four first ones[e]. The discharge data are provided by *Alpiq*. Concerning the meteorological data, except in Section D.2.1, three temperature stations will be used: Les Marécottes, Montagnier-Bagnes (both provided by *MeteoSwiss*) and Champéry (provided by *Meteogroup* via *MeteoSwiss*). The Grand-Saint-Bernard station will only be used for the radiation measurements. A total of 16 rain gauges are used in the models based on rain gauges.

Finally the period of calibration will run between the 1st October 2018 and the 1st October 2021. Figure 3.9 displays the discharge between 2017 and 2021 at Emosson-Collecteur Est (in red) and La Fouly (blue). The seasonal behaviour is here clear and hence **HIT/FAR will be calculated daily** (see Section 3.3.2.3).
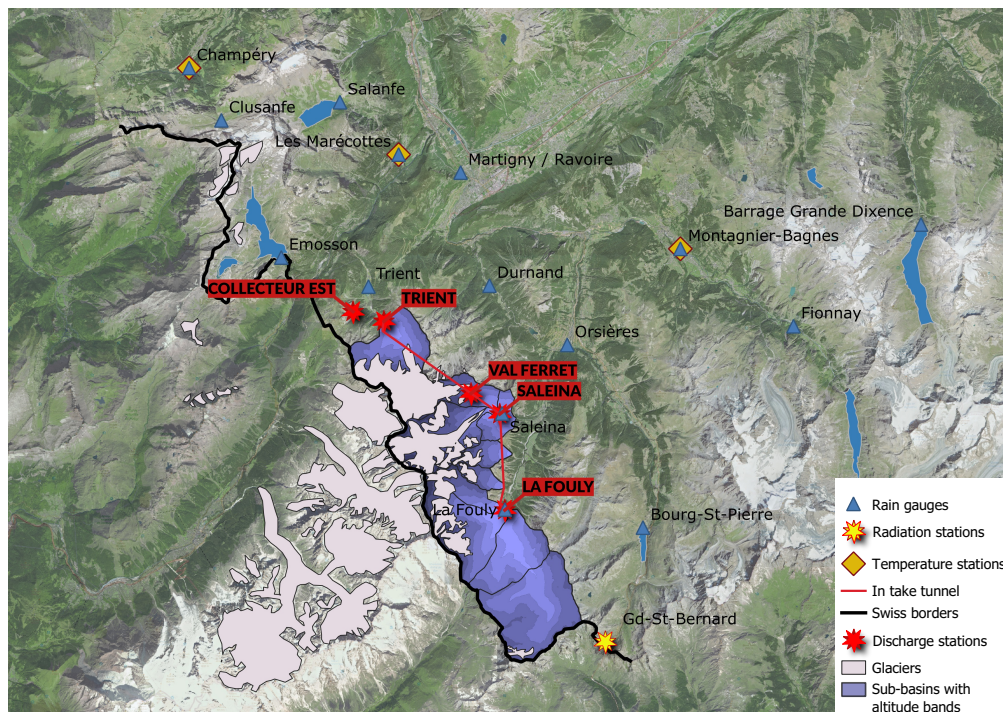
Figure 3.8: Description of the Emosson catchment and of the used meteorological stations.

---

[e]Val Ferret discharge curve has the same shape as Emosson-Collecteur Est with just lower values.
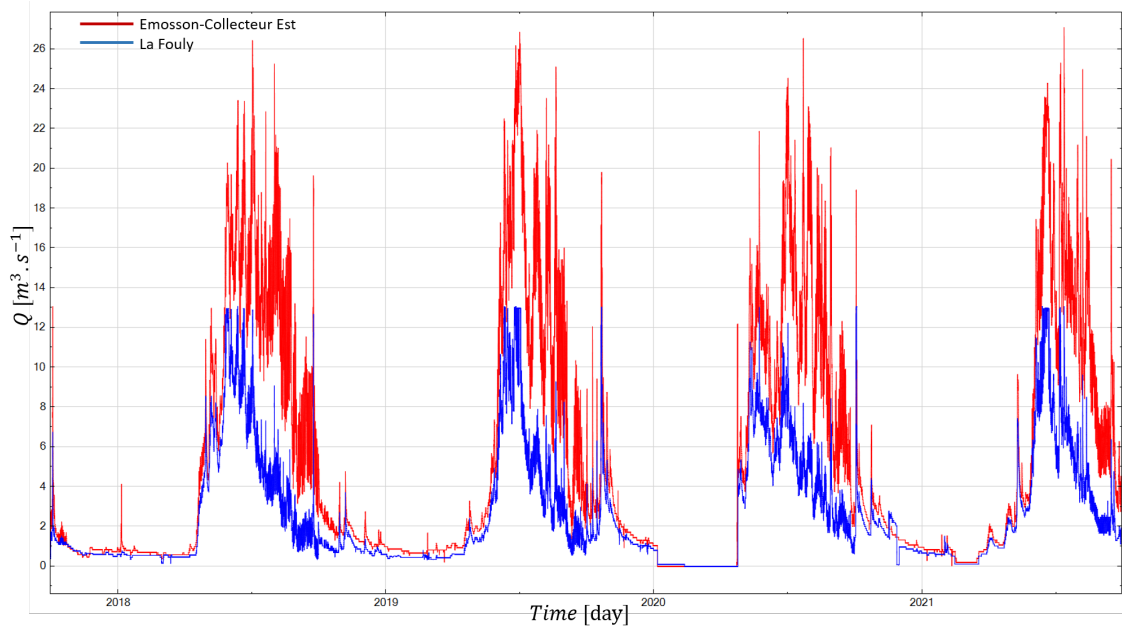
Figure 3.9: Discharges measured at Emosson-Collecteur Est (red) and La Fouly (blue) over the period 01/10/2017 to 01/10/2021.

The *Radar4Infra* project in Emosson aims both at triggering flash flood alarms and at estimating the cumulated volumes generated in these events. Hence, **both alert HIT-FAR and cumulated volume scores** will be analysed (see methodology in Section 3.3.2.3 and Appendix D.1.2). Only the main results are presented in the body of this report. More detailed results are available in Appendix D.

3.4.1.2   Precipitation analysis in Emosson

The performances of CPCH and AZC are first assessed in terms of precipitation for this specific basin. Figures 3.10 and 3.11 display respectively the HIT-FAR scores of AZC, CPCH and inca0 products for the Emosson and La Fouly stations and the monthly mean precipitation analysis for the stations of Emosson, La Fouly, Saleina and Trient. The HIT/FAR analysis shows that CPCH is performing very well in Emosson. On the contrary the raw radar AZC has mostly low HIT scores which points to an underestimation of precipitation there. This underestimation is confirmed by Figure 3.11. AZC FAR scores are also worse than CPCH ones. In La Fouly, CPCH is still performing a lot better than AZC, but its performances decrease quickly when precipitation intensity increases. For most events with small precipitation rates, CombiPrecip is hence expected to be quite similar to rain gauges. However differences may appear for some sub-catchments at higher intensities.

Concerning the mean precipitations by month, CPCH respects the rain gauge cumulated volume everywhere except in Saleina where it is underestimated. AZC mean precipitation ratio are more variable along the year (notably in Emosson and Saleina) and are globally underestimated. Finally, inca0 seems to be globally consistent with CPCH values but with a nearly constant underestimation bias. Indeed inca0 curves are all similar to CPCH ones with a reduction of precipitation between 5 and 20% for inca0. All these analyses are limited to comparison at stations locations where CPCH is interpolated to rain gauges. **The behaviour of CPCH can differ far from the rain gauges locations.**
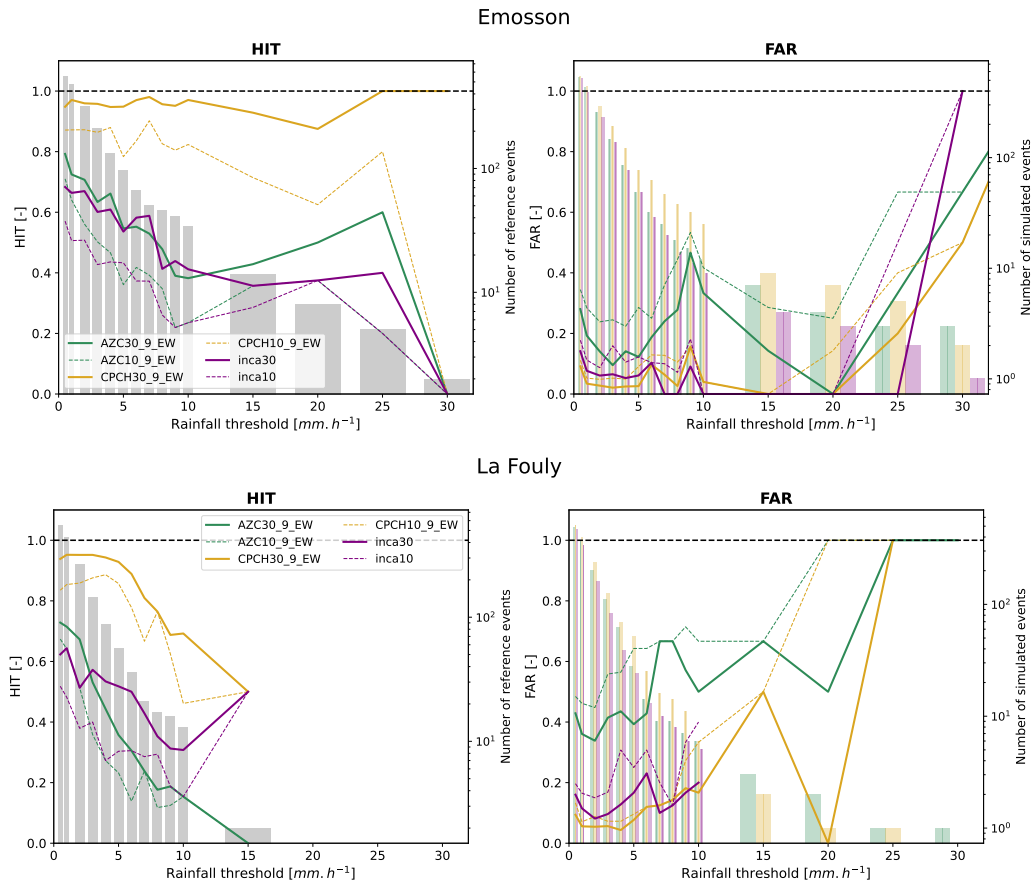
Emosson

Figure 3.10: HIT/FAR values for AZC, CPCH and inca0 as simulated data (rain gauge as reference) for the stations of Emosson and La Fouly. Two tolerances (10% – dashed lines – and 30% – solid lines) are used. Radar grid points are interpolated at rain gauge stations using the average of the 9 closest points. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score). The covered period of time is 17/07/2018-28/09/2021.
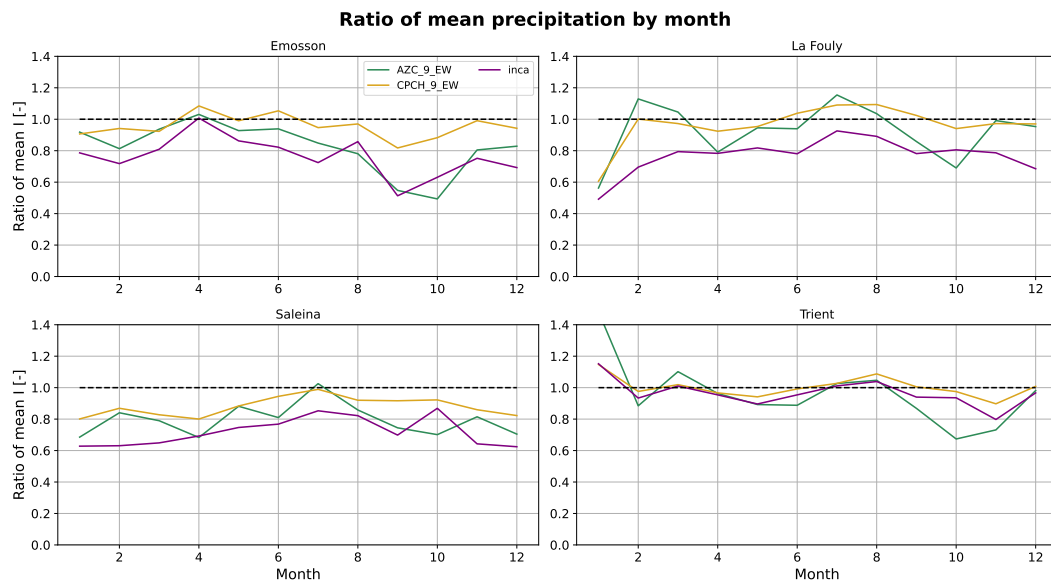
Figure 3.11: Ratio of AZC, CPCH and inca0 monthly mean precipitation over rain gauge mean precipitation for four stations of the Emosson catchment for the period 17/07/2018-28/09/2021.

3.4.2   Zurich-Altbach catchment

3.4.2.1   Presentation of the catchment

The Zurich airport basin is located in the North of Zurich. Water is flowing in the Altbach river from Bassersdorf and Nürensdorf to the airport. A discharge gauge is located at the crossing between the Altbach river and the Schaffhauserstrasse, close to the Kloten airport entrance where the Altbach river is tunnelled through the airport. The catchment is illustrated by Figure 3.12. The rain gauge stations from Zurich-Kloten and Winterthur/Seen are used for precipitation while Zurich-Kloten is used for temperatures – Nürensdorf station data are not available in real time for an operational use. The total area of the catchment is $22.6\,\mathrm{km}^2$ with mostly cropland (50%) and forest (20%) but also some impermeable areas in the cities and near the airport. The catchment orography is rather flat (at least compared to Emosson) with altitudes ranging from $436\,\mathrm{m}$ (Zurich Kloten Schaffhauserstrasse intake) to $613\,\mathrm{m}$ (Brütten). The approximate response time of the whole basin is in the order of 1 hour. Annual precipitations are close to $1400\,\mathrm{mm}$ according to the Swiss hydrological atlas website[222].
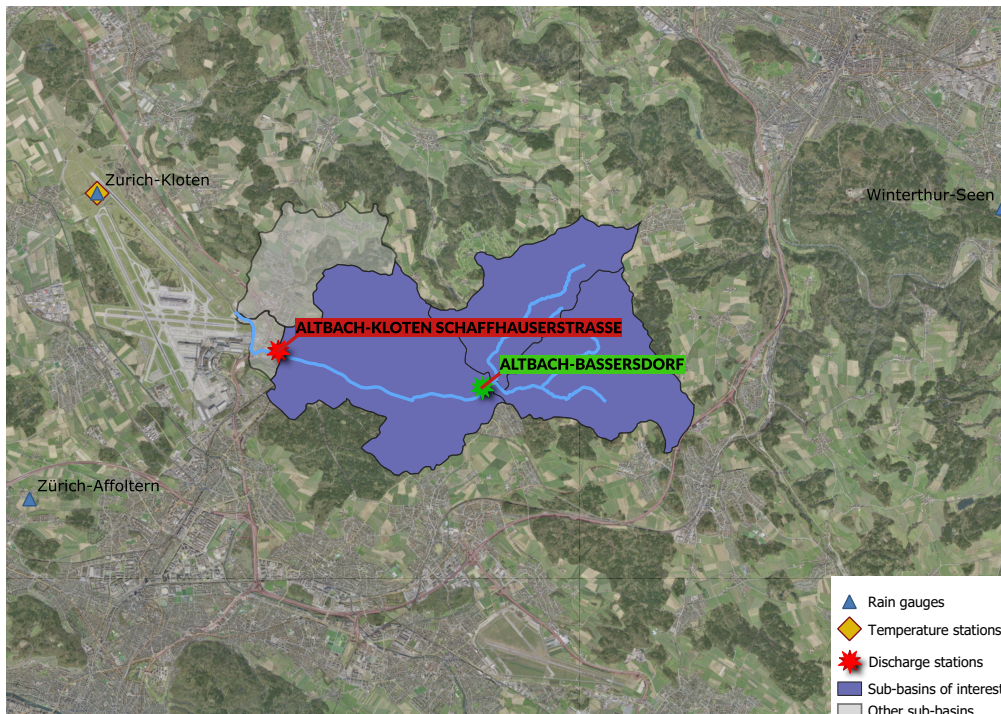
Figure 3.12: Description of the Zurich catchment. The Altbach-Kloten discharge will be used in all the models of this section while the Altbach-Bassersdorf discharge station will only be used in Section 5.3.3.

The discharge data at Zurich Kloten-Schaffhauserstrasse are provided by the canton of Zurich. No radiation data are needed as the physical processes are limited to infiltration, runoff, soil moisture storage and other soil processes. Snow plays only a minor role in winter. The introduction of impermeable areas in the model is however necessary to model flash floods correctly. The period of calibration will run between the 1st September 2018 and the 1st September 2021.

Figure 3.13 displays the discharge between 2017 and 2021 at Zurich-Kloten Schaffhauserstrasse. No clear seasonal behaviour is observed and hence **HIT/FAR will be calculated by event** (see Section 3.3.2.3).
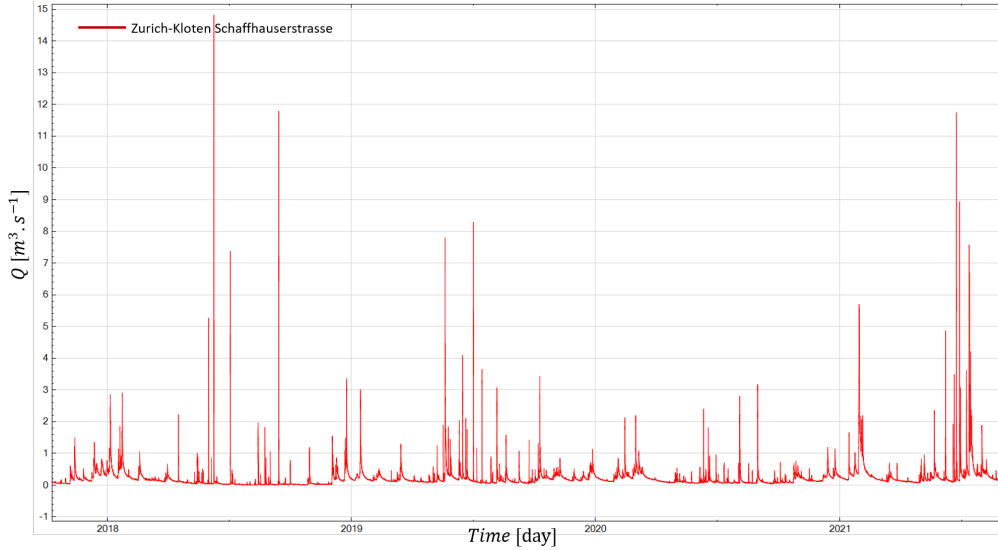
Figure 3.13: Discharge measured at Zurich-Kloten over the period 01/10/2017 to 01/10/2021.

The goal of the modelling in the Zurich-Altbach catchment is to provide flash flood alerts to evacuate the Zurich airport commercial zone. The threshold level for evacuation has been fixed at $20 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$. Hence the focus will be mostly on high flash flood events in the following sections. Notably, the most important indicators are the **HIT-FAR alert scores** which simulate and evaluate the real alerts that will be issued. Scores for cumulated volumes will not be presented as, unlike in hydroelectric reservoirs such as Emosson, the flood volumes are not important. However, they could be interesting for similar catchments where the focus is on hydroelectricity. Hence, a volume analysis is proposed in Appendix E.

3.4.2.2 Precipitation analysis in Zurich

Figures 3.14 and 3.15 provide respectively the monthly mean precipitation analysis and the HIT/FAR scores for the two rain gauge stations of the basin. Compared to Emosson, *CombiPrecip* gives precipitation estimates similar to the rain gauges as both the mean precipitation ratio and the HIT scores are close to 1 while the FAR scores are close to 0. Performances of raw radar AZC are however quite bad with a strong seasonal behaviour marked by more intense spring rains than those measured by rain gauges in both Kloten and Winterthur and an overall underestimation of precipitation in Winterthur. inca0, as at the global Swiss level, follows a seasonal behaviour similar to CPCH but with an underestimation bias between 10% and 20%. This bias will need to be corrected in the rainfall-runoff model.
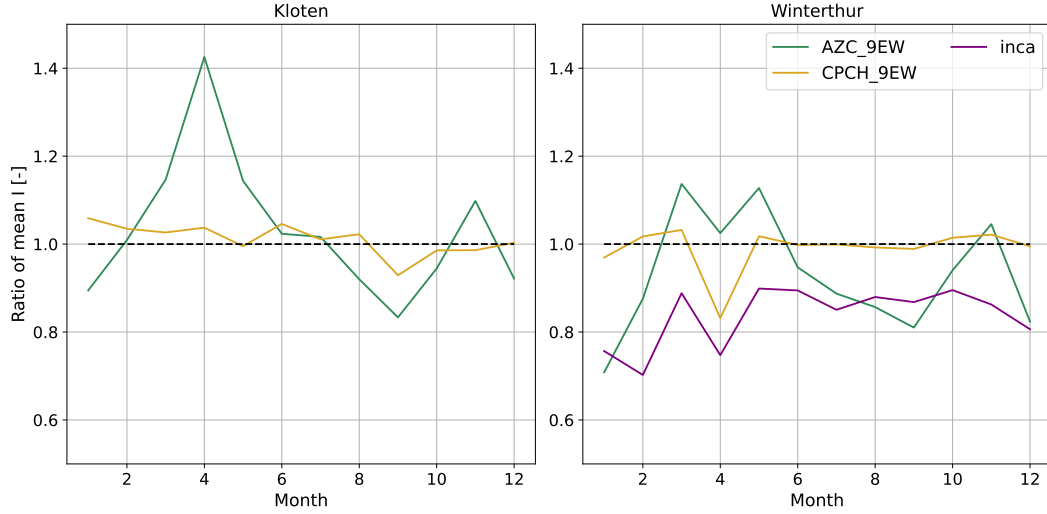
Figure 3.14: Ratio of AZC, CPCH and inca0 monthly mean precipitation over rain gauge mean precipitation for the two stations of the Zurich catchment for the period 17/07/2018-28/09/2021. inca0 data have not been interpolated for the whole period in Kloten and are not represented.
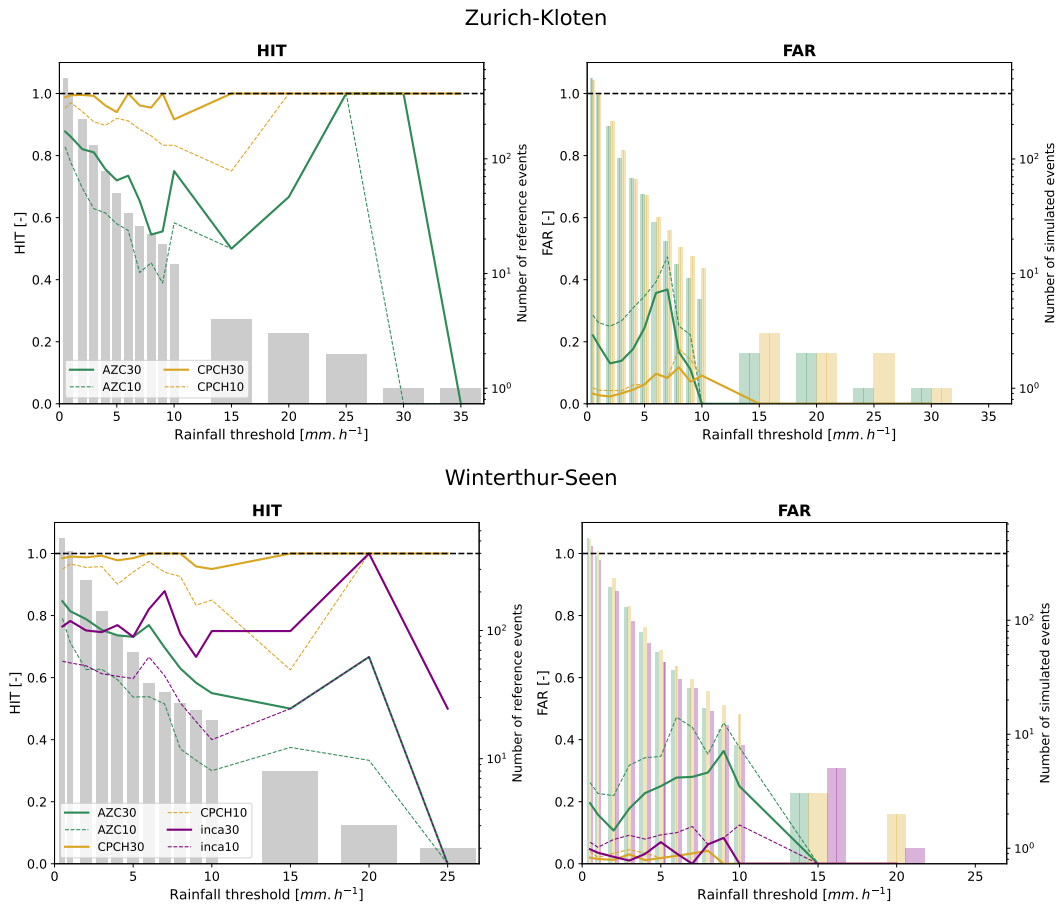

Figure 3.15: HIT/FAR values for AZC and CPCH (rain gauge as reference) for the stations of Kloten and Winterthur-Seen. Two tolerances (10% – dashed lines – and 30% – solid lines) are used. Radar grid points are interpolated at rain gauge stations using the average of the 9 closest points. The covered period of time is 17/07/2018-28/09/2021. Scores for inca0 are also provided for Winterthur.

Chapter 4

Radar measurement assessment

4.1 Introduction

In order to evaluate the interest of radar in the improvement of flood flow nowcasting, comparing the input radar data to rain gauge measurements is a necessary first step. This analysis will mainly be performed at *MeteoSwiss* stations locations.

Different interpolation techniques can be used to estimate the precipitation intensity at the rain gauge stations from radar data. They will be compared in a first section. In a second section, the performances of several radar products will be assessed at the scale of Switzerland. The best product will be used in a third section to observe if its performances depend on the location of the stations. In a fourth part, the historical evolution of radar performances will be assessed to determine a period of time for hydrological calibration. Finally the performances of radar input data for different timesteps will be compared.

A simple validation of the performances of the co-kriging far from the *SwissMetNet* stations of *MeteoSwiss* will also be performed thanks to cantonal rain gauges in Bern and Ticino cantons.

4.2 Comparison between different interpolation techniques

An analysis based on raw radar data (AZC) and *CombiPrecip* data (CPCH) is first done to assess the importance of the interpolation techniques used to determine the radar rain value interpolated at each station. Five methods have been used:

- The closest radar point is used for each station.

- The 4 closest radar points are used for each station and each point is given a weight proportional to the squared inverse of its distance to the station.

- The 4 closest points are used for each station and an average of these point data is computed (equal weight for each point).

- The 9 closest points are used for each station with distance squared inverse weights.

- The 9 closest points are used for each station with equal weights.

The analysis is performed only for four stations in the Emosson catchment. Emosson is a mountainous catchment where precipitation can vary very much spatially. Hence if the differences between the different interpolation techniques are small in Emosson, it is reasonable to assume they are small everywhere[a]. The response times of the sub-catchments of Emosson are also diverse with

---

[a]Such verifications were performed in Zurich, confirming this assumption.

some small response time catchments (1-2h). Moreover Emosson is a catchment monitored by *Alpiq* which participates in the *Radar4Infra* project. Mean precipitation analysis (for the four stations) and HIT/FAR analysis (only for the Emosson station) are provided respectively in Figures 4.1 and 4.2a (for AZC) and 4.2b (for CPCH).
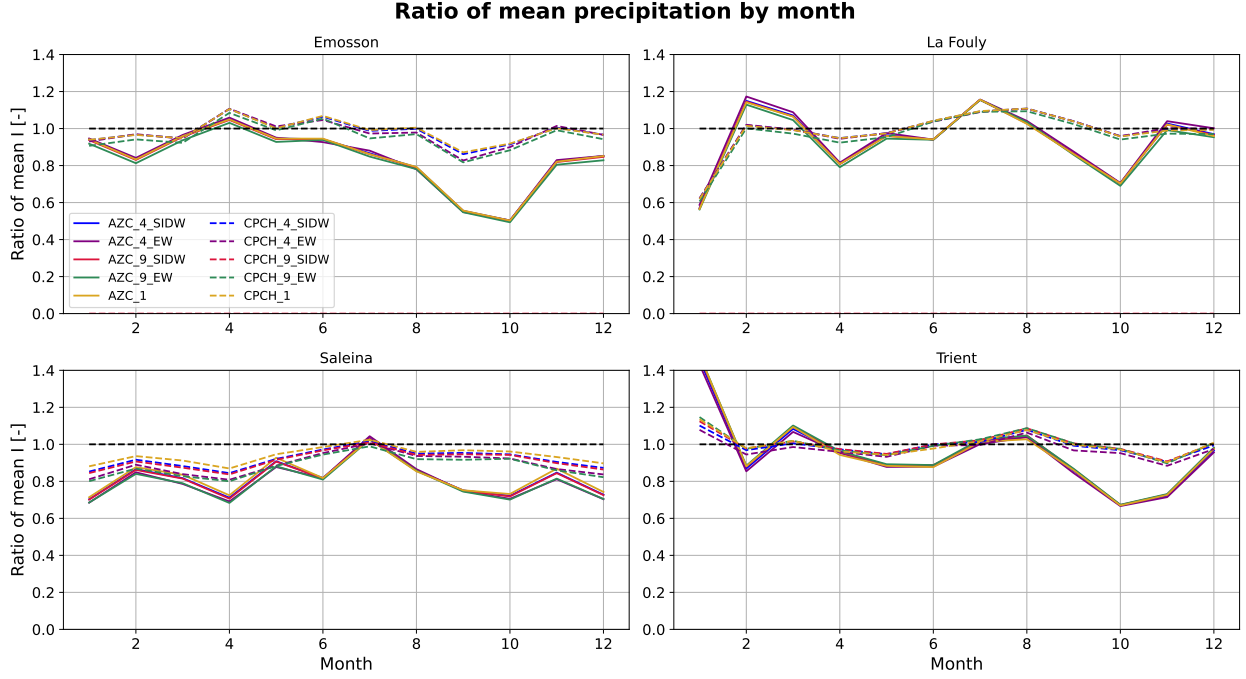
Figure 4.1: Monthly mean precipitation ratio between AZC and rain gauge data (solid lines) or CPCH and rain gauge data (dashed lines) for different interpolations. The number of radar grid points considered for each interpolation is provided in the legend. Equal and squared inverse distance weights are respectively designated as *EW* and *SIDW*. The covered period of time is 17/07/2018 0h-28/09/2021 0h.

Figure 4.1 does not show any clear difference between the AZC curves, all five interpolations provide similar results for all stations. The curves for CPCH are also quite close but show some slight differences. Indeed the 4 or 9 points interpolation with equal weights have usually smaller ratios diverging by up to 2%. Conversely the 1-point interpolation has sometimes slightly higher ratios.

Concerning the HIT and FAR scores, Figures 4.2a and 4.2b show nearly no differences in both HIT and FAR scores for small precipitation thresholds (below $10 \, \mathrm{mm\,h^{-1}}$). For higher thresholds, where the number of events becomes small, some differences can appear, mostly when the tolerance is low (10%). With a 10% tolerance the 4 and 9 point-averaged interpolations have slightly lower HIT scores for CPCH. Hence the HIT/FAR and mean precipitation analysis underscore some slight differences between the different interpolation techniques for CPCH, notably for high thresholds. **However, as these differences are small, in the following parts, the average of the 9 closest points will be used**. This decision is taken in order to be in agreement with the standard of *MeteoSwiss*.
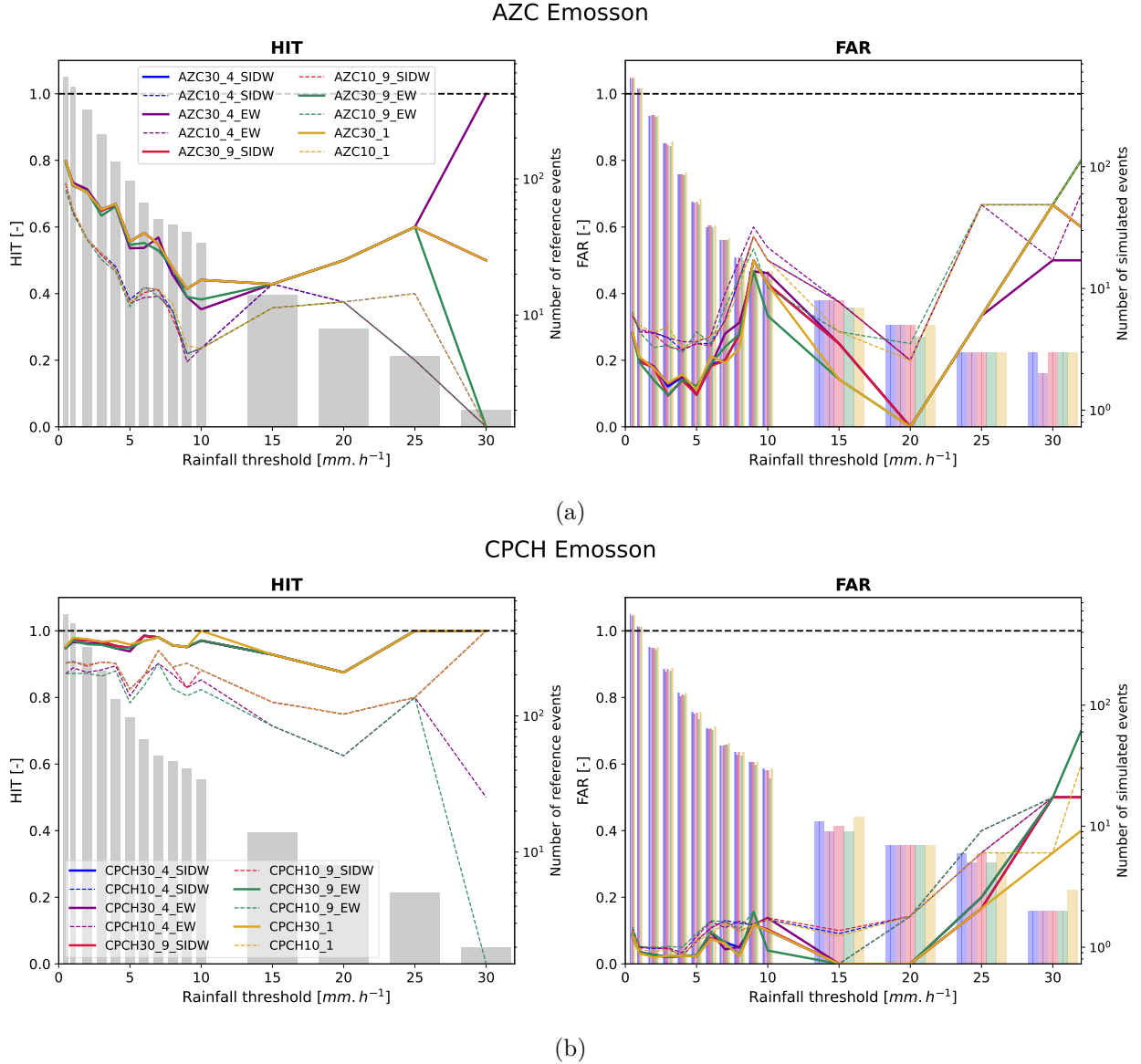
Figure 4.2: HIT/FAR values with AZC (Figure 4.2a) or CPCH (Figure 4.2b) as simulated data and rain gauge as reference data for the station of Emosson. Two tolerances (10% – dashed lines – and 30% – solid lines) are used. The number of radar grid points considered for each interpolation is provided in the legend. Equal and squared inverse distance weights are respectively designated as *EW* and *SIDW*. The right axis (bar plot) provides the number of reference events (for the HIT score) and simulated events (for the FAR score). The covered period of time is 17/07/2018 0h-28/09/2021 0h.

4.3 Comparison between raw radar data, *CombiPrecip* and inca0

In this section AZC, CPCH and inca0 data are compared using the three metrics previously described. These metrics are averaged over all available stations from Figure 3.5. The radar gridded data will be interpolated on rain gauge stations using the average of the 9 closest points.

4.3.1 HIT FAR analysis

Figure 4.3 provides the HIT and FAR scores averaged over all stations for the period 2018-2021. It illustrates the role of tolerance, which allows to largely increase the HIT and FAR scores – by $10 - 15\%$ – meaning $10 - 15\%$ of events are well observed, but with intensities between 10 and 30% higher or lower than in rain gauges. It is also clear that the co-kriging of raw AZC radar data with rain gauges in the *CombiPrecip* product largely improves both HIT and FAR scores (by 20% for both tolerances and both HIT and FAR scores). Hence it will probably be better to use the CPCH product as precipitation input of hydrological model rather than the raw radar data.

Concerning inca0, it performs badly in HIT score (slightly lower than AZC) and quite well in FAR score, which probably highlights a tendency to underestimate precipitation in the INCA forecast.

The *CombiPrecip* performances with a 30%-tolerance are:

- HIT values above 90% for intensities below $30\,\mathrm{mm\,h^{-1}}$.

- FAR values around 20% until $20\,\mathrm{mm\,h^{-1}}$, increasing for higher intensities.
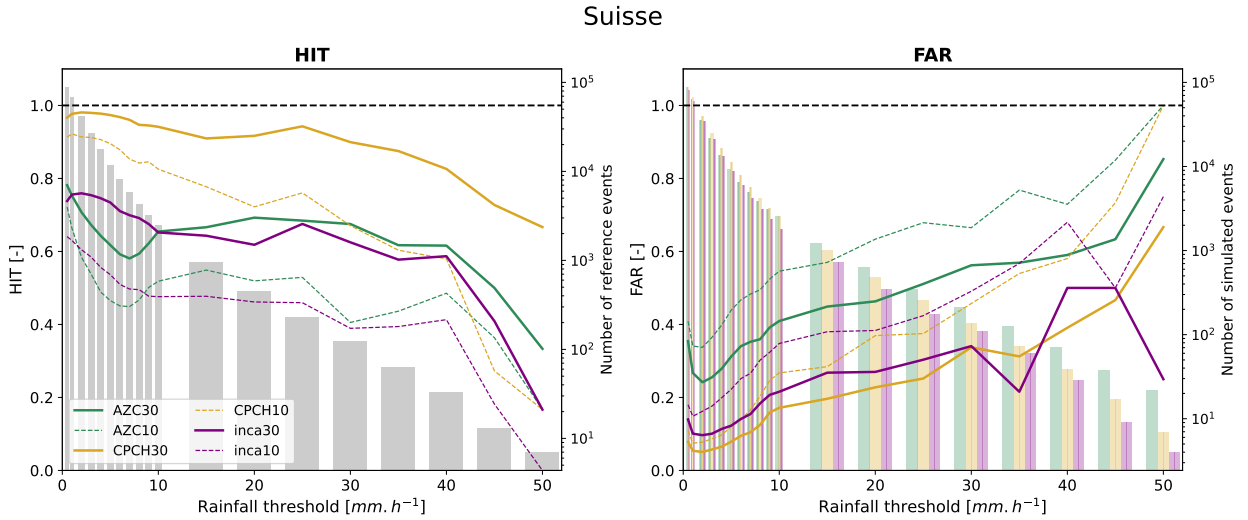
Figure 4.3: HIT/FAR mean values of AZC, CPCH and inca0 for the period from 17/07/2018 0h to 28/09/2021 0h (local time). Two tolerances (10% – dashed lines – and 30% – solid lines) are used. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).

4.3.2 Month average rainfall analysis

Figure 4.4 provides the monthly mean precipitation measured by radar for the three products provided by *MeteoSwiss* and the corresponding ratios with rain gauge measurements for the period 2018-2021. Precipitations are averaged over all the stations.

AZC precipitations are largely underestimated in winter, amounting to only 70% of the rain gauge measurements some months. Conversely they are overestimated in summer, from June to August, with a peak at $115 - 120\%$ of the observed rain gauge data. This seasonal phenomenon is also present for CPCH though largely smaller, with only 5% of underestimation in winter and 9% of overestimation in summer. This was expected as CPCH data are interpolated at stations with rain gauge data. This interpolation seems to be quite effective though not perfect.

The winter underestimation may be explained by the radar difficulties to observe snow precipitations. This will need to be confirmed by observing performances of stations depending on the altitude. Nevertheless, it needs also to be noted that rain gauges are less performing during snow precipitations. Concerning the summer overestimation, it may be due to the rain gauges inability to catch intense summer events. Indeed the gauge mean precipitation displayed in Figure 4.4-left shows a small depletion in July while radar data displays a peak at this time. Radars are probably better at observing localized peak precipitation during storms.

Concerning inca0 data, as expected with the HIT/FAR analysis, they are usually underestimated, resulting in rainfall volumes approximately 10% lower than those from *CombiPrecip*. The overall shape is otherwise similar to the shape of the CPCH product. It is probable that a volume bias correction has been applied to the previous INCA data to create the new *CombiPrecip* product.
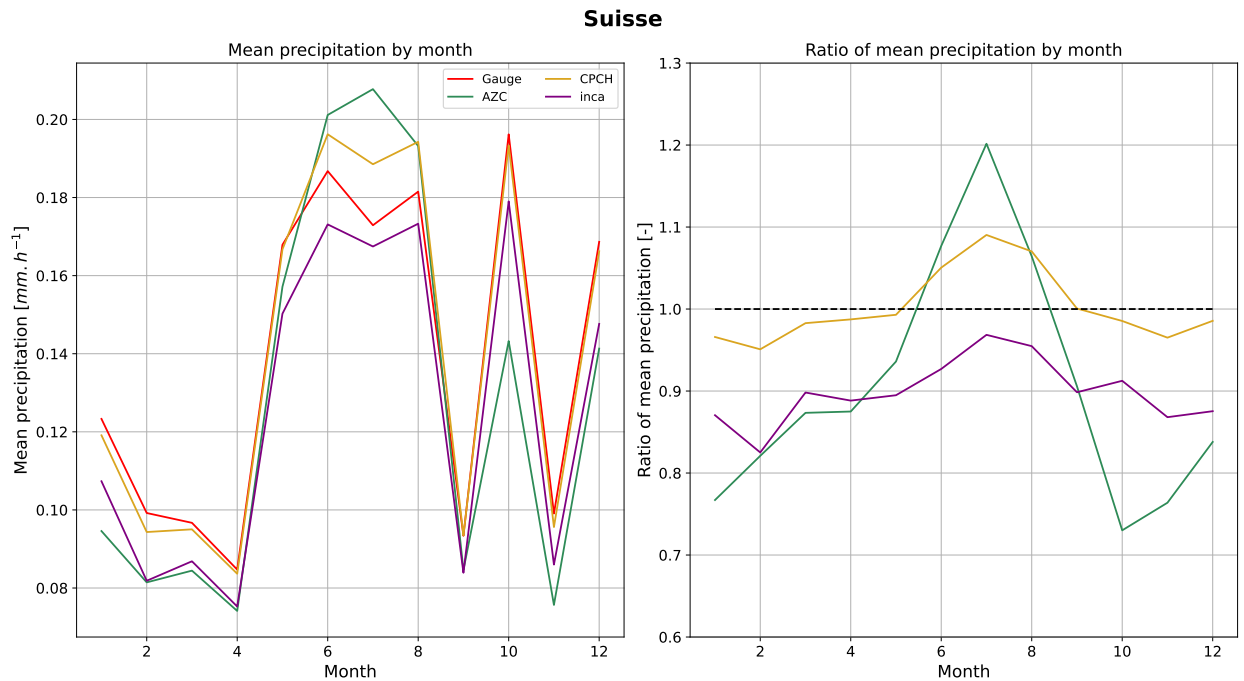
Figure 4.4: Left: Monthly mean precipitation averaged over all stations from rain gauge, AZC, CPCH and inca0 data for the period 17/07/2018 0h-8/09/2021 0h (local time). Right: Ratio of AZC, CPCH and inca0 mean precipitation over rain gauge mean precipitation.

4.3.3 Cumulative distribution functions

Complementary cumulative distribution functions (CCDF) of the four products are now computed and plotted in Figure 4.5-left. For each product and for a given precipitation intensity, this plot provides the percentage of time this intensity is exceeded. Only times when there is non-null precipitations in a least one of the four products are considered. The curve is very similar to cumulative distribution functions (CDF) from the literature[223]. As the differences between the four curves are small, the ratio between the radar curves and the rain gauge one are plotted in Figure 4.5-right.
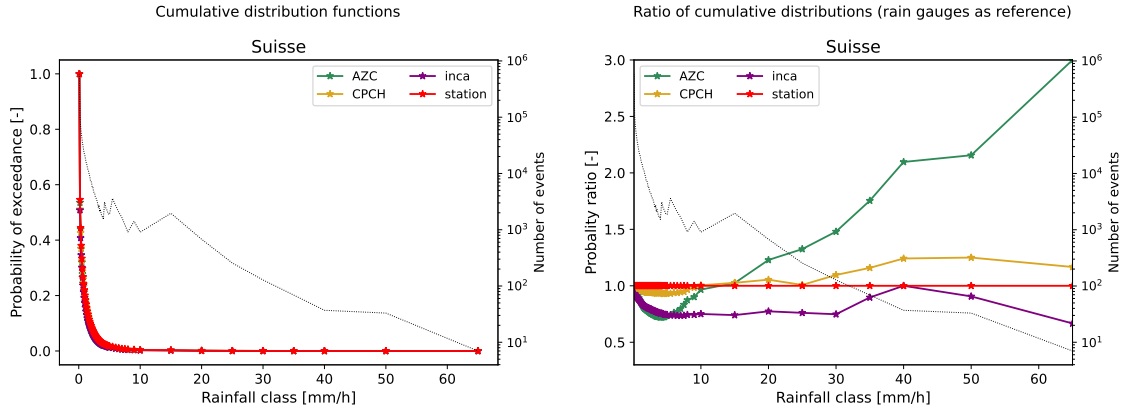
Figure 4.5: (left) CCDF of the four products (rain gauge, CPCH, AZC and inca0) over the period 17/07/2018-28/09/2021. (right) Ratio of the CCDF of CPCH, AZC and inca0 over the CCDF of rain gauges. The right axis displays the number of events observed in CPCH.

Table 4.1: Percentage of precipitation timesteps for each product for the 181 stations over the period 17/07/2018-28/09/2021.

Rain gauge	CPCH	AZC	inca0	One of them
13.00%	22.23%	15.80%	18.75%	23.76%

This plot shows a clear dependency of QPEs values to the precipitation intensity. Overall, while **small events are underestimated in radar products** (lower probability of exceedance at low intensities), **high events are overestimated above** $15 \, \mathrm{mm \, h^{-1}}$ **in radar products** (there are more high events in radar products). This trend is particularly important in raw radar AZC data with twice more events exceeding $40 \, \mathrm{mm \, h^{-1}}$ over the period and a significant underestimation of events below $10 \, \mathrm{mm \, h^{-1}}$. This last point is probably due to the surfacic character of radar data which are taken over a $9 \, \mathrm{km^2}$ area (for the chosen interpolation method) rather than on only few centimetres as in rain gauges: more localized precipitation can then be observed and averaged over $9 \, \mathrm{km^2}$. Hence, **raw radar data are more dispersed** over the range of precipitation than rain gauges. The same behaviour is observed for *CombiPrecip* but with **a largely smaller dispersion**. Finally inca0 seems, as observed previously, to follow the CPCH trend with a **constant underestimation bias**. Hence, even at high intensities, the probability of exceedance of inca0 is smaller than the one of rain gauges which is not the case for CPCH.

Concerning the occurrence of precipitation, Table 4.1 provides the percentage of timesteps where each product contains non-null precipitation. It confirms that more events are observed in radar than in rain gauges. However there is largely more events observed by CPCH than by AZC. Indeed, CPCH includes both rain gauges and raw radar events: most events (93.6%) observed in at least one of the products are observed by *CombiPrecip*. Finally there is a small decrease in the number of events in inca0, probably for two reasons: because some events are dying between CPCH and the inca0 forecast at 10 minutes, and because of the steady underestimation bias of inca0.

4.4 Geographical dependency of *CombiPrecip* performances

Until now, only the average situation for the whole Switzerland has been analysed while it is reasonable to believe radar performances may be dependent on the location. In this section, the geographical behaviour of the *CombiPrecip* performances is investigated. The focus is first on regional dependency before turning to the altitude dependency of radar performances.

4.4.1 Performance assessment for several regions

Switzerland stations were split between 6 regions: *Plateau*, *Jura*, *Préalpes*, *Valais*, *Tessin* and *Grisons*. The number of stations in each region lies between 13 in *Jura* and 54 in *Plateau* for the period 2018-2021 (Table 4.2).

Table 4.2: Number of stations by region and altitude.

By region		By altitude	
Region	Number of stations	Altitude	Number of stations
Plateau	54	0-500m	46
Valais	38	500-1000m	60
Préalpes	34	1000-1500m	43
Grisons	24	1500-2000m	29
Tessin	18	2000m and more	15
Jura	13		
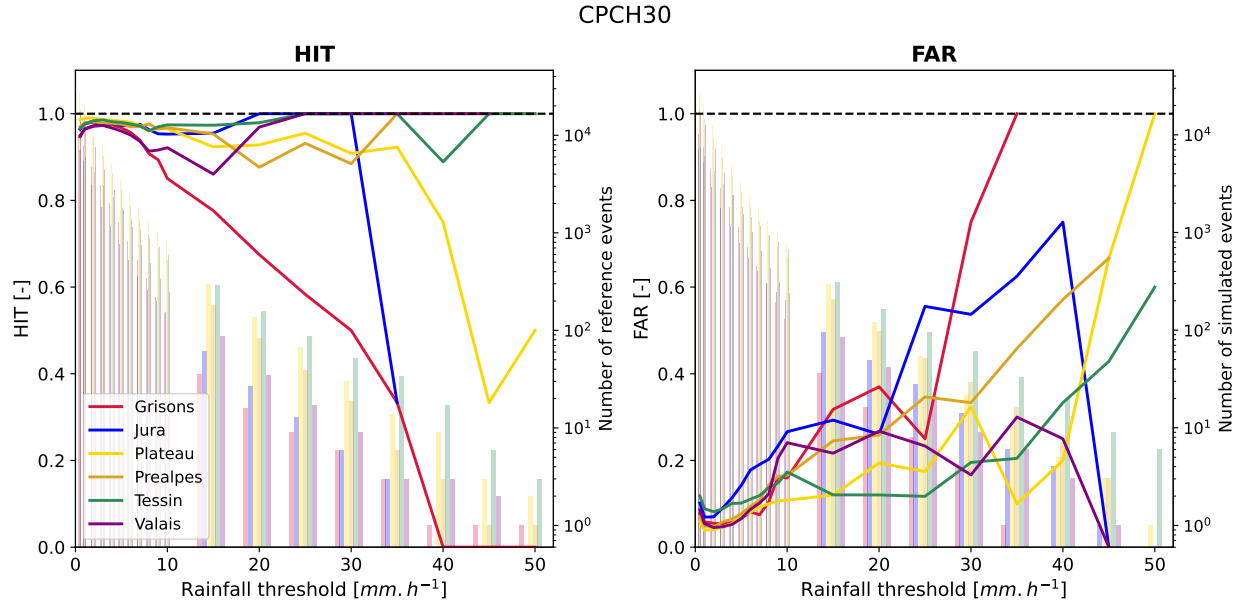
Figure 4.6: HIT/FAR regional mean values of CPCH for the period from 17/07/2018 0h to 28/09/2021 0h (local time) with a tolerance of 30%. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).

Figure 4.6 displays the HIT and FAR scores according to the region of Switzerland where the station is located. For the small intensities (below $10\,\mathrm{mm\,h^{-1}}$) all regions are performing well in HIT score (above 95%). However for higher intensities the HIT value of the *Grisons* region starts to decrease quickly and steadily. The number of events is still significant (37 events at $15\,\mathrm{mm\,h^{-1}}$ and still 9 at $25\,\mathrm{mm\,h^{-1}}$ where the HIT is below 60%) and similar to other regions where the HIT score is still very high (notably in the *Jura* and *Valais*). All other regions keep HIT scores higher than 0.85-0.9.
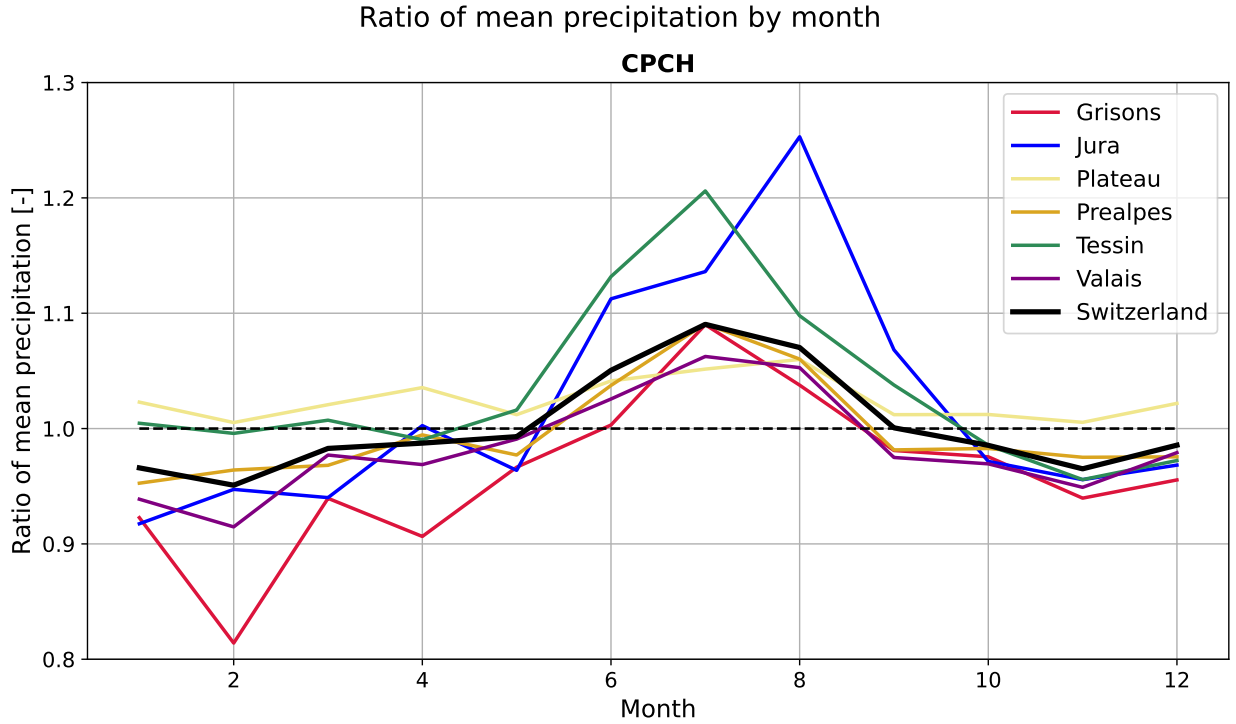
Figure 4.7: Month precipitation ratio between CPCH data and rain gauge data in the period from 17/07/2018 0h to 28/09/2021 0h. Both CPCH and rain gauge data are averaged by regions.

Concerning the FAR score it remains low everywhere for small precipitations but increases at high precipitations in *Grisons*, *Jura* and to a smaller extent *Préalpes*. It probably underlines a difficulty to obtain good quality radar data in the *Grisons* while, in the *Jura*, the high precipitations are quite overestimated compared to rain gauge, as both HIT and FAR scores are high. This last point does not mean that radar data are of poor quality in the *Jura* but rather that they are able to catch very localized high intensity rainfall very frequent in *Jura* in summer. The discrepancies between the regions are mostly due to radar performances, as the observations above are still valid for the AZC product (with poorer global performances), which HIT-FAR scores are provided in Appendix B.1.1.

The hypothesis of an overestimation of precipitation in the *Jura* in summer is confirmed by Figure 4.7 where the mean precipitation ratios are displayed month by month for each region. In the case of *Jura* it is clear that rainfall are higher in radar data than in gauge stations in summer while being smaller in winter. This general feature, observed for the whole Switzerland (Figure 4.4) is present everywhere with different intensities: the bias is low in the *Plateau*, precipitations are mostly underestimated in winter in *Valais* and *Grisons*, while they are mostly overestimated in the *Tessin* in summer. These features were globally already present in the raw radar data (Appendix B.1.1) and were only reduced in *CombiPrecip*.

This seasonal behaviour is further confirmed and explained by Figure 4.8 which presents the CCDF ratios in the different regions of Switzerland. Indeed, as observed in Figure 4.4 (left), precipitations are more important in summer than in winter in Switzerland. Figure 4.8 illustrates that there is an **important bias at high intensities** where the probability of having events higher than $40\,\mathrm{mm\,h^{-1}}$ over one hour can be up to 2.5 times higher in CPCH than in rain gauges in the *Jura* or *Préalpes*. These events are occurring in summer and explain the seasonal tendency of CPCH. Similarly to the mean precipitation analysis, this tendency is more significant in *Jura* and in *Tessin* but also in *Préalpes*. Conversely, the *Plateau* shows a strong agreement between CPCH and rain gauge data, with even a small underestimation of radar at high intensities. Finally, in *Grisons*, at all intensities, events are usually higher in rain gauge data than in radar ones, supporting observations
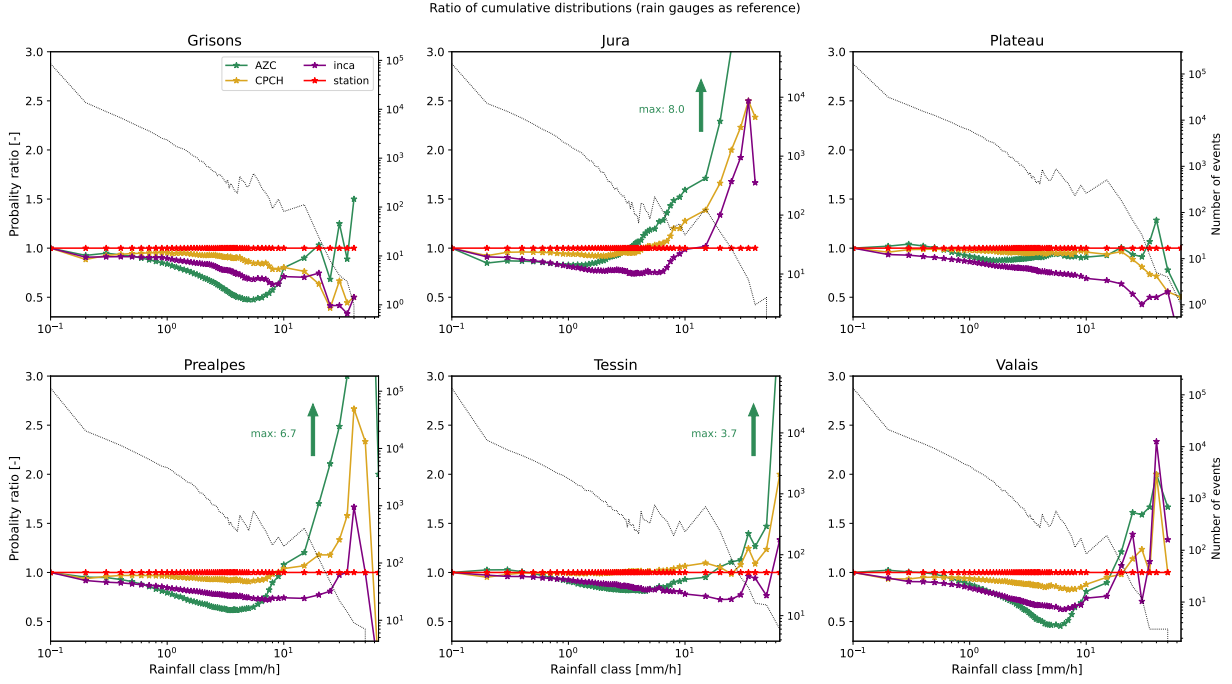
Figure 4.8: Ratio of the CCDF of CPCH, AZC and inca0 over the CCDF of rain gauges in the different regions of Switzerland between 2018 and 2021. A logarithmic scale is used for the x-axis. The right axis displays the number of events observed in CPCH.

from Figure 4.7. At low intensity, there is a higher probability of observing small events with the radar everywhere, as explained before for the whole Switzerland. For the other products, the trend observed in CPCH is stronger for AZC, while inca0 curve is still steadily shifted downward from CPCH curve.

In summary, **CPCH data are close to rain gauge ones in the *Plateau* while enduring an important seasonal behaviour in *Jura* and *Tessin* with high intensity summer events overestimated in CPCH compared to rain gauges**. Radar data are also disagreeing with rain gauges in *Grisons*. To complement this study, an observation of variations of these scores with altitude is proposed in the next section.

4.4.2   Performance assessment according to altitude

Similarly to the previous section, Switzerland stations have been split into 5 categories depending on their altitude: $0 - 500$m, $500 - 1000$m, $1000 - 1500$m, $1500 - 2000$m and 2000m and higher. Numbers of stations for each altitude class are provided in Table 4.2.

The first observation is that the higher is the altitude the lower are the HIT/FAR performances. Notably, below $1500\,\mathrm{m}$, the HIT scores are higher than 0.9-0.95 for all precipitation thresholds (except at very high intensities where only few events are recorded). Conversely, even at low precipitations, higher stations have worse performances. Notably below $10\,\mathrm{mm\,h^{-1}}$ the curves are almost perfectly ordered by altitude with better HIT for lower altitudes. For higher intensities the HIT scores of altitude comprised between 1500 and 2000m are around 85% only, while the stations higher than 2000m have HIT values going down quickly to 0.

Similarly, the FAR values of high-altitude stations (mostly above 2000m) increase quickly with the rainfall threshold, while lower stations have quite good performances even at high thresholds. Notably the stations below 1000m have FAR scores remaining below a 0.3-value until thresholds as high as $40\,\mathrm{mm\,h^{-1}}$ where only few events are present. These observations were already present in AZC – which curves are provided in Appendix B.1.2.

CPCH30
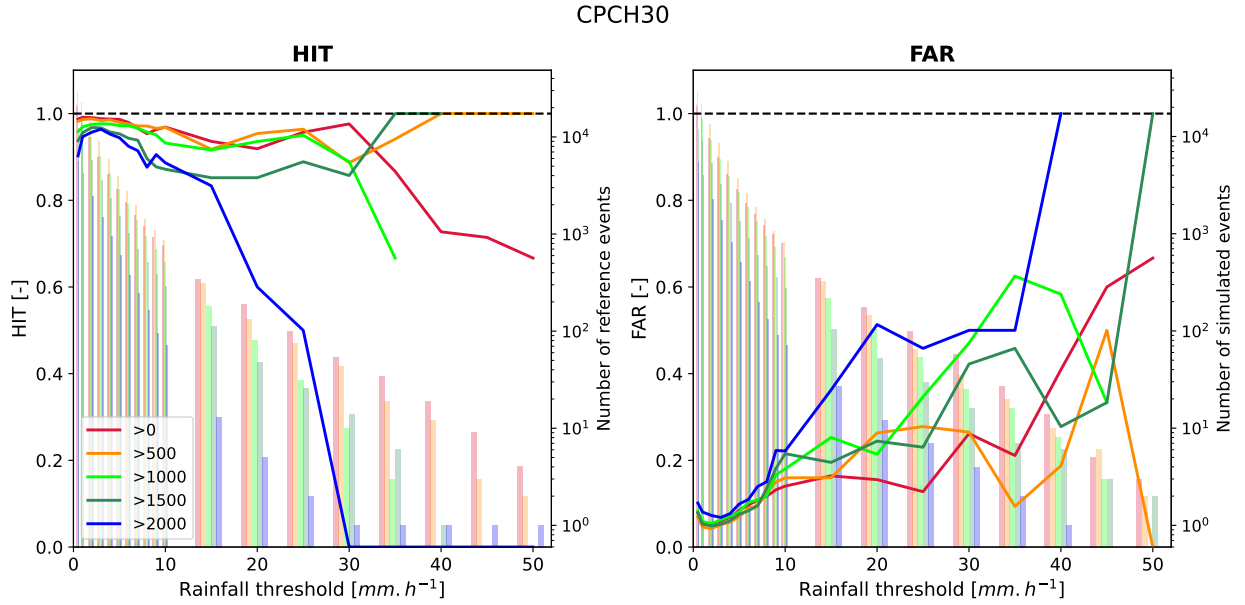
Figure 4.9: HIT/FAR scores of CPCH aggregated by station altitudes for the period from 17/07/2018 0h to 28/09/2021 0h (local time) with a tolerance of 30%. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).

The mean precipitation ratio plot of Figure 4.10 is even clearer with **curves ordered according to their altitude: higher stations are linked to an underestimation of precipitation by radar while lower stations are characterized by an overestimation all over the year. The seasonal variations of the bias are also more significant above** 1000m**. All these features were already present in the AZC product which curves are presented in Appendix B.1.2. Compared to regions, altitude seems to be a better predictor of the bias and may be a better way to de-bias the data.** Such behaviour could be partially explained by the nature of precipitation: at high altitude and mostly in winter, precipitation falls as snow which is more difficult to observe with radar. It could also be due to the worse quality of rain gauge reference data in altitude as rain gauges are known to perform worse at high altitudes due to snow and foehn issues.

Finally, Figure 4.11 presents the ratios of the *CombiPrecip* complementary cumulative distribution function (CCDF) over the one of rain gauges. The curves are still ranked by altitude at low intensities with **more small events observed in high altitudes** (meaning an underestimation of small events). There is also **more high events in CPCH than in rain gauges at high altitudes** (>1000 m). However here the curves are not perfectly ordered as the curve for altitudes between 1000 m and 1500 m increases before the two curves for higher altitudes. One explanation is that, as observed in Figure 4.8, **the bigger overestimation of CPCH at high intensity occurs in *Jura* which altitudes are mostly met in the** $1000 - 1500$ m **range.** All these features were already present in the raw radar AZC – with more small and high intensity events – which curves can be found in Appendix B.1.2.

Ratio of mean precipitation by month
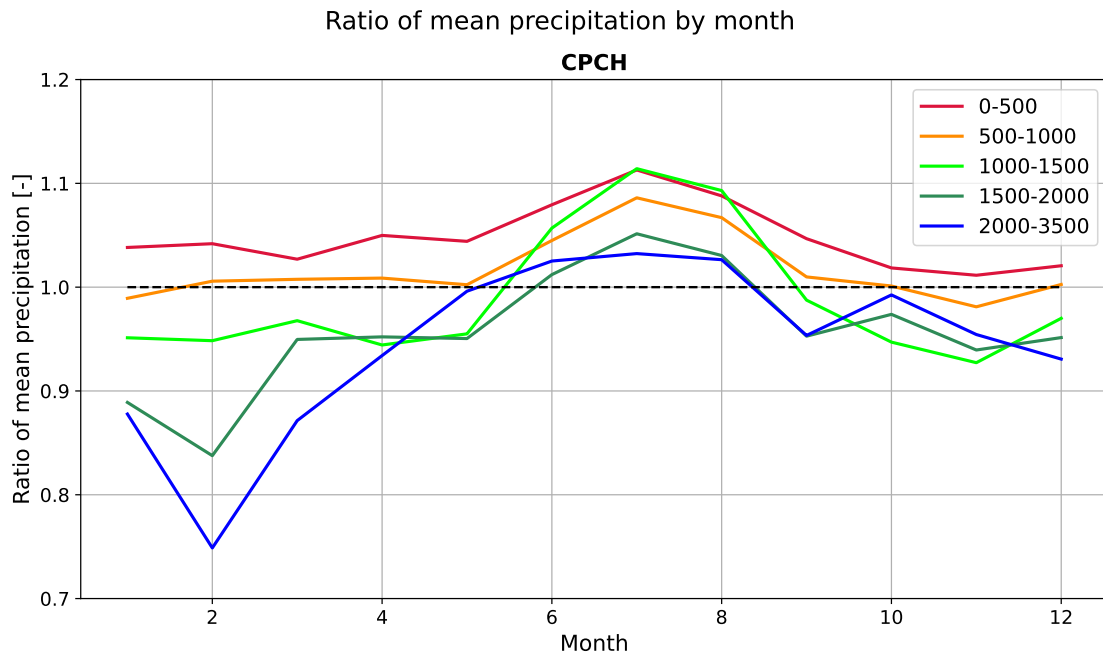
Figure 4.10: Month precipitation ratio between CPCH data and rain gauge data in the period from 17/07/2018 0h to 28/09/2021 0h (local time). Both CPCH and rain gauge data are averaged by altitude.

Ratio of cumulative distributions (rain gauges as reference)
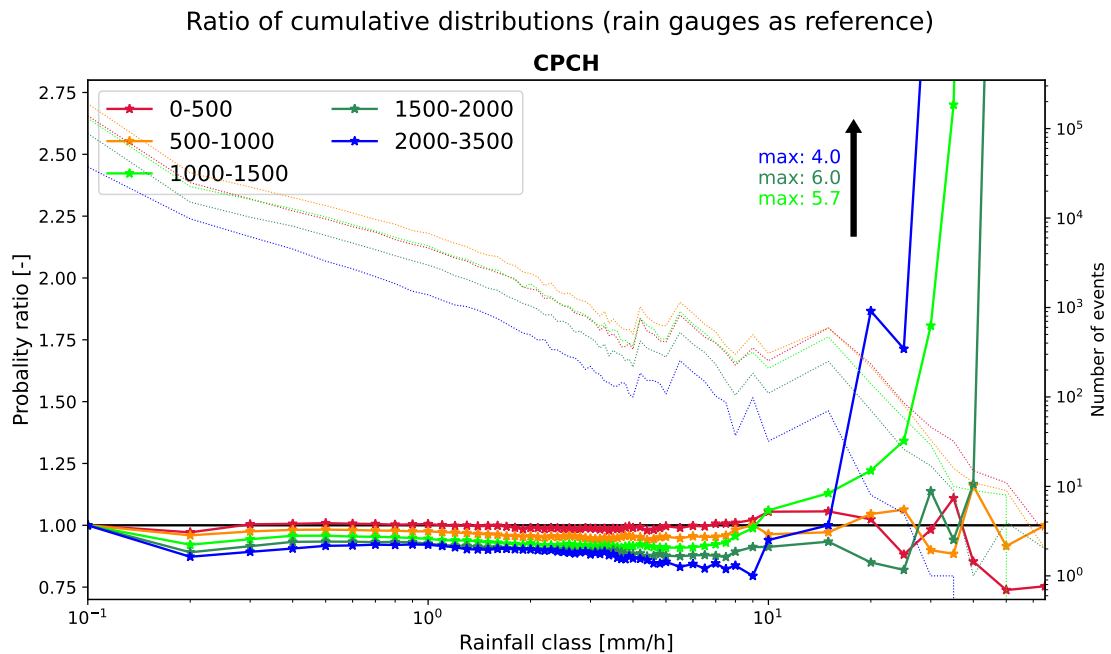
Figure 4.11: Ratio of the CCDF of CPCH over the CCDF of rain gauges for different altitudes between 2018 and 2021. A logarithmic scale is used for the x-axis. The right axis displays the number of events observed in CPCH.

4.5 Evolution of *CombiPrecip* performances

Until now the period of focus covered approximately 3 years between 2018 and 2021. However AZC and CPCH data are available since 2014. Hence it is possible to study the evolution of the performances of both AZC and CPCH in order to determine the widest period of study on which data are reliable. In this section, four different periods of time will be studied:

- 2014-2015: from 01/01/2014 2h local time to 01/01/2016 1h local time.
- 2016-2017: from 01/01/2016 2h local time to 01/01/2018 1h local time.
- 2018-2019: from 01/01/2018 2h local time to 01/01/2020 1h local time.
- 2020-2021: from 01/01/2020 2h local time to 31/12/2021 0h local time.
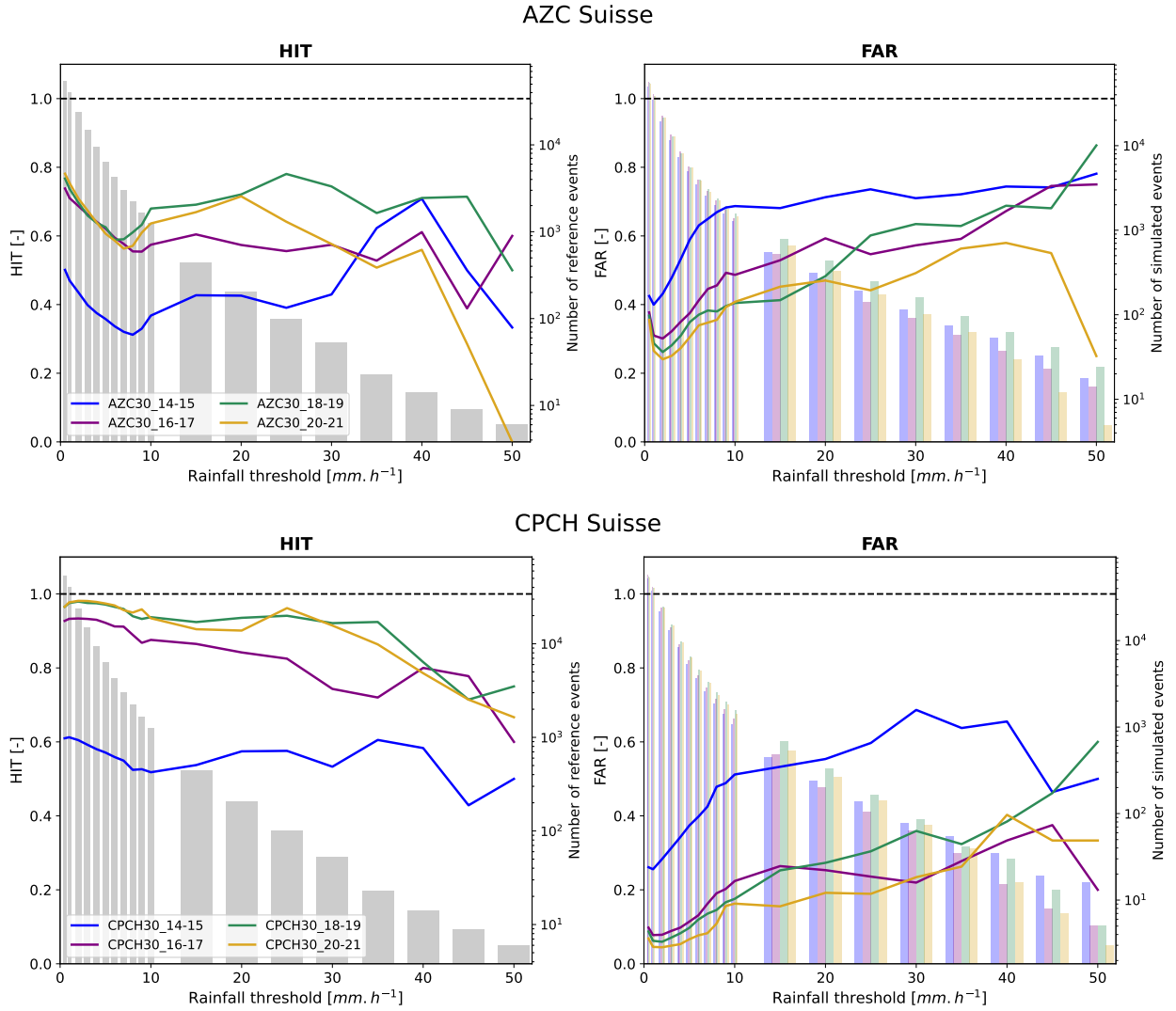
Figure 4.12: HIT/FAR scores of AZC and CPCH products for the whole Switzerland over 4 distinct periods of time. A tolerance of 30% is chosen. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).

Figure 4.12 displays the HIT and FAR scores for both AZC (top) and CPCH (bottom) for these four periods. The scores in 2014-2015 were bad for both AZC (HIT below 50%, FAR above 60%) and CPCH (HIT below 60%, FAR above 50%). These performances were largely improved in 2016-2017 and the HIT scores were further improved in 2018-2019 for both AZC and CPCH. Since then performances are constant. However the performances improvement was slightly higher for

CPCH than for AZC and could be as high as a 30% improvement in both HIT and FAR. These improvements have been observed for all regions and all altitudes though the differences observed in the previous section for different altitudes has also been reduced over time.

Two explanations can be proposed for this improvement and they probably both play a role:

- On the one hand the radar data acquisition was improved along the period with notably two new radars put into operation: Plaine Morte in 2014 and Weissfluhgipfel in 2016. The post-processing of data has also probably been improved, notably to better estimate the precipitation in altitude and to counter shadowing issues.

- On the other hand the *CombiPrecip* interpolation may have been improved with the addition of new rain gauge stations used as reference for this interpolation. Notably in 2014 181 stations were used for the kriging process (Section 3.3.1.1)[78], while since 2017 between 262 and 269 stations are used.

For the assessment and improvement of nowcasting using radar data, the period of study should hence focus on a large enough period where the radar data are as reliable as possible. **Thus we will use later periods comprised between 01/10/2017 and 01/10/2021 for the calibration of hydrological models.**
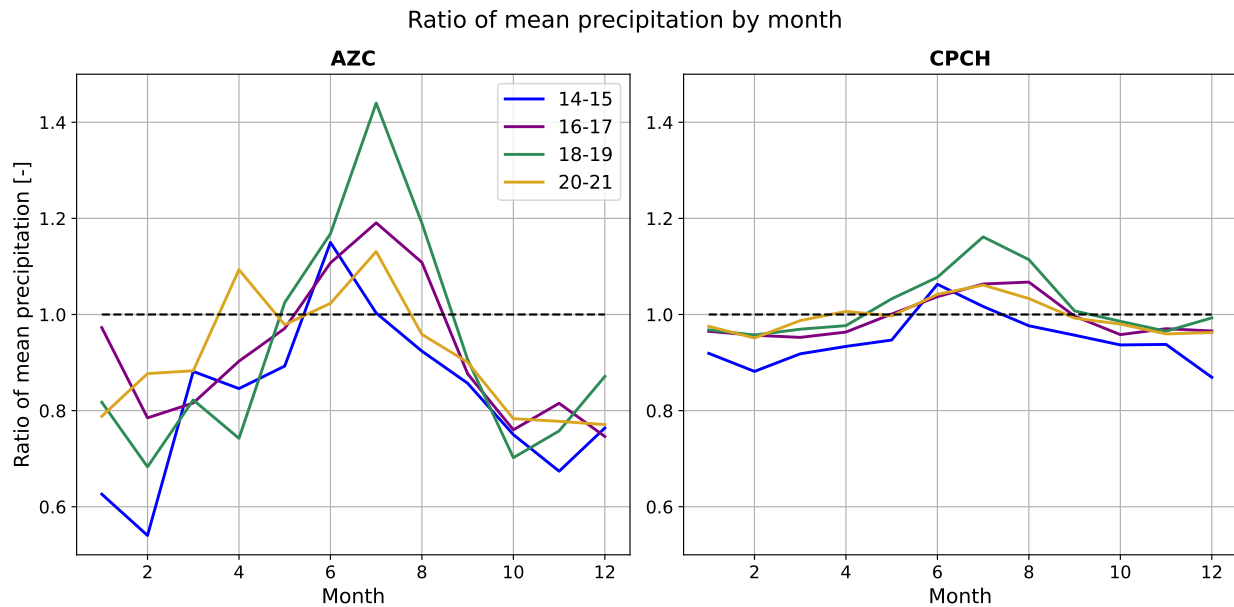
Figure 4.13: Month precipitation ratio between AZC and rain gauge (left) or between CPCH and rain gauge (right) for the four studied periods of time.

On the mean precipitation side provided by Figure 4.13, the precipitation were underestimated in 2014-2015 (for both AZC and CPCH). 2016-2017, 2018-2019 and 2020-2021 have similar shapes and values in winter but differ in summer (2018-2019 being more overestimated) which is probably due to different meteorological conditions. It is for example possible that in 2018 and 2019 radar has been able to catch more huge precipitations that the rain gauges had difficulties to observe.

4.6   Performances for a timestep of 10 minutes

In this section, the performances of the previous products with 10-minute timesteps will be analysed. As volume analysis is not dependent of the timestep, only HIT/FAR scores will be studied.

4.6.1   Impact of the different interpolation techniques

The timestep may have an influence on the differences between several interpolation techniques. To verify if the statements from Section 4.2 are still valid, a HIT-FAR analysis for different interpolation techniques is performed for both RZC and CPCH. Results are provided in Figure 4.14.
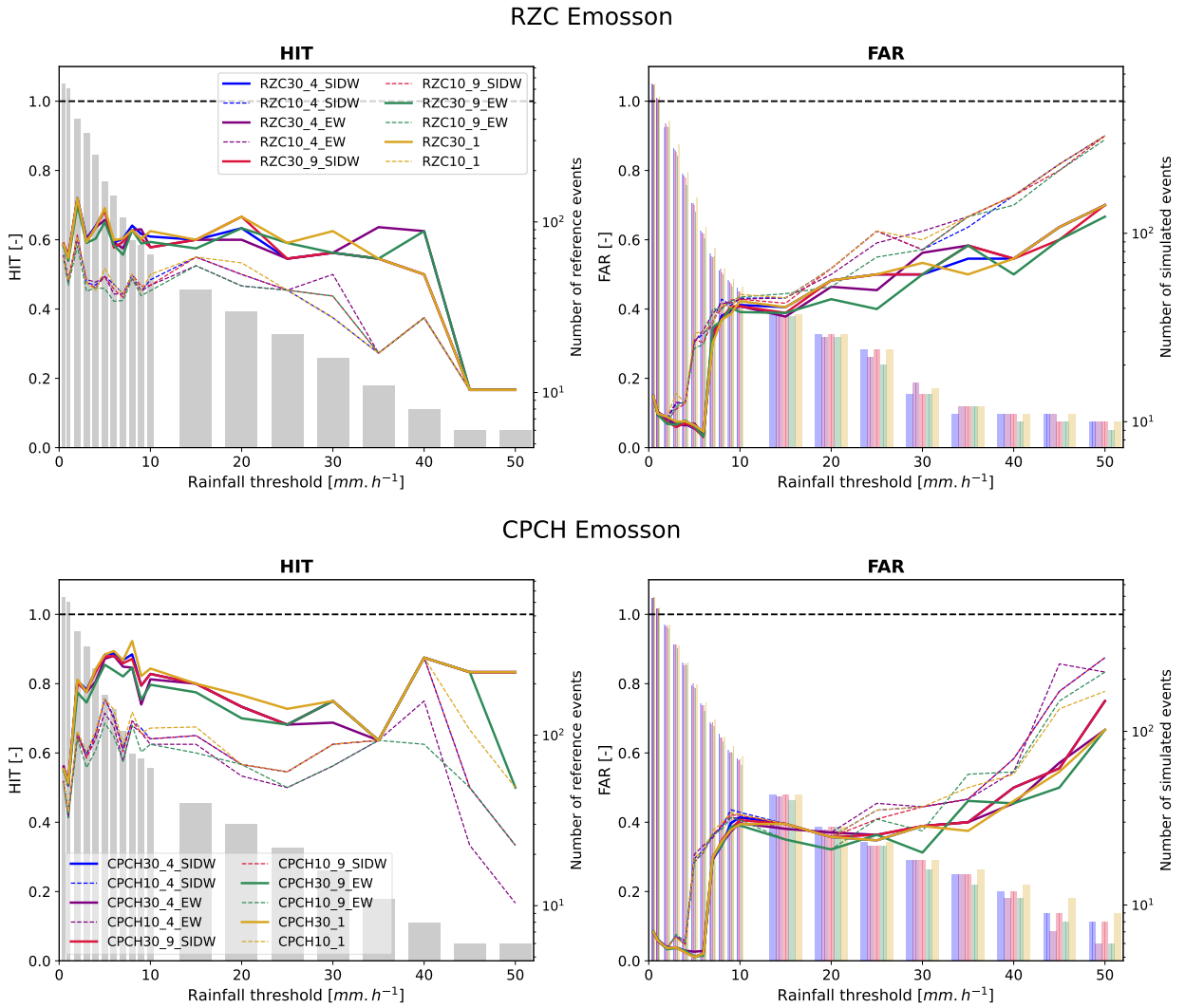
Figure 4.14: HIT/FAR values with RZC or 10-minute CPCH for the station of Emosson (reference: rain gauge). Tolerances of 10% – dashed – and 30% – solid – are used. The interpolation used in each case is provided in the legend. The right axis provides the number of reference events or simulated events. The period of time is 17/07/2018 0h-28/09/2021 0h.

Once again the differences are small both in HIT and FAR scores. However, as in Section 4.2, it is also possible to notice that for both RZC and CPCH and for both tolerances, the best HIT and FAR scores are obtained for the 1-point interpolation. Nevertheless, as results are similar to Section 4.2, the following analysis will continue to use the average of the 9 closest points which is the *MeteoSwiss* standard.

4.6.2 Results

As for the hourly timestep analysis, a first assessment of the performances of raw radar (RZC), *CombiPrecip* (CPCH) and inca0 products is performed. Results are displayed in Figure 4.15.

At first glance, both HIT and FAR scores are smaller than for a 1-hour timestep (Figure 4.3). This is due to a volume underestimation of 6 to 8% which was observed between 10-minute and 1-hour products. According to *MeteoSwiss*, such discrepancy is due to the encoding process of the data. Here, the *CombiPrecip* product is once again the best product both in FAR and HIT scores. The performance increase compared to raw radar RZC data is however smaller than for an hourly timestep, being around 15-20% in FAR and 5-20% in HIT. The RZC is even slightly better in HIT at the highest intensities for a tolerance of 10%. Moreover **the *CombiPrecip* HIT scores are largely smaller than for an hourly timestep** shrinking from 0.95 to 0.8 (for a tolerance of 30%) at low thresholds. **The RZC and inca0 scores are a lot closer to their values for an hourly timestep**. This behaviour probably originates from the way *CombiPrecip* is disaggregated: **the disaggregation is based on the 5-minute and 1-hour raw radar** product, as explained by Barton *et al.*[83] – INCA is not a disaggregation product, having a natural timestep of 10 minutes.
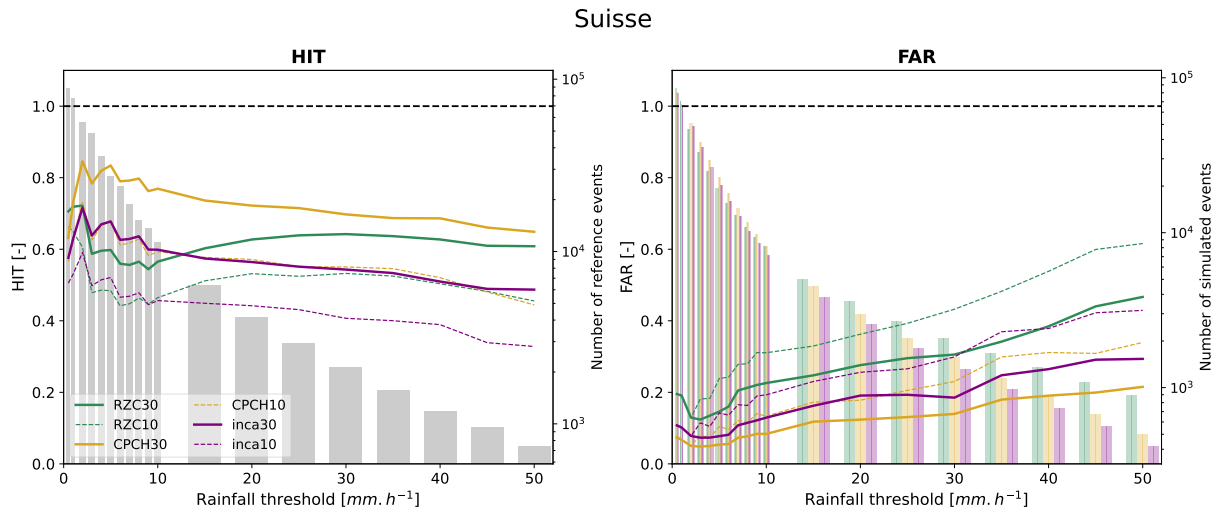
Figure 4.15: HIT/FAR mean values of RZC, CPCH and inca0 for the period from 17/07/2018 0h to 28/09/2021 0h (local time) for a timestep of 10 minutes. Two tolerances (10% – dashed lines – and 30% – solid lines) are used. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).

Concerning the spatial dependencies of *CombiPrecip* performances, Figures 4.16 and 4.17 provide respectively the scores according to the region or to the altitude of each station.

The regional dependency is globally similar to the 1-hour timestep product (Figure 4.6) but with worse HIT and better FAR scores: *Grisons* is performing poorly while *Jura* has good HIT but bad FAR (intensity overestimation). However, *Valais* is also performing quite poorly. Such difference between 10-minute and 1-hour data is probably due to the stations included in the 1-hour analysis and not in the 10-minute one. Indeed, *Valais* is the region were most of the lost stations are included (10 stations) and these stations are globally located at low altitudes (4 of the 10 *Valais* stations below 1000 m have been excluded while none of the 6 stations above 2000 m has been excluded). As performances decrease with altitude it is normal that the performances are decreasing more in *Valais* than in other regions.

Concerning altitude, the patterns observed in Figure 4.9 are still present with better performances at low altitudes. Once again this could be due either to more difficulty to obtain good radar precipitation data at high altitudes or to the difficulty to have reliable rain gauge data there.
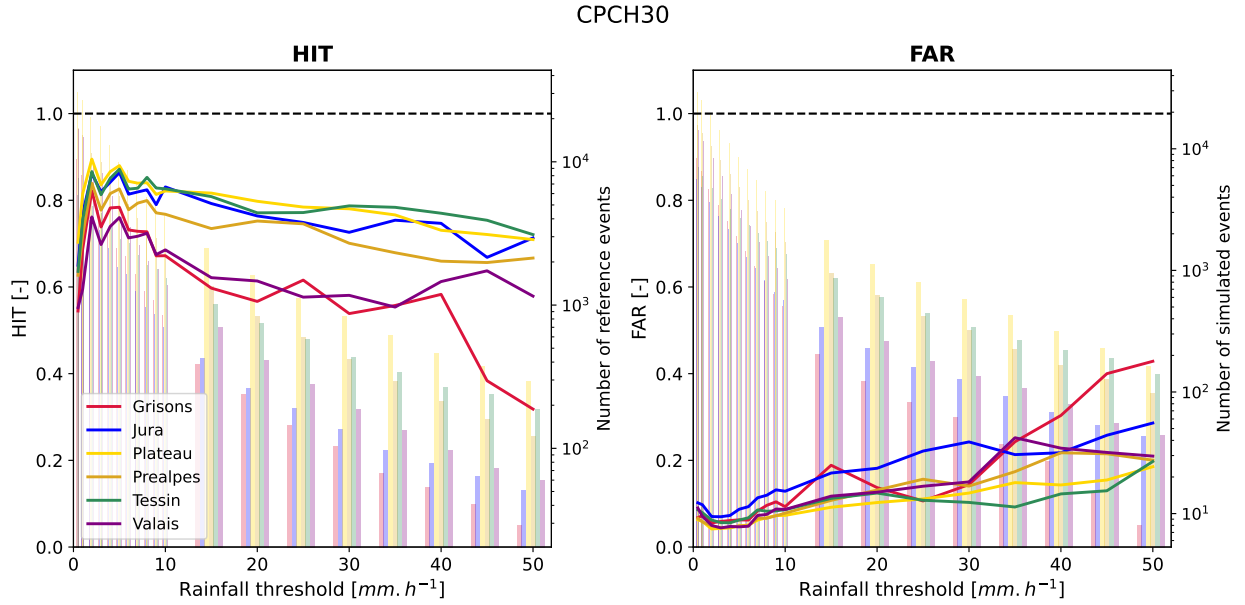
Figure 4.16: HIT/FAR regional mean values of CPCH (with 10-minute timestep) for the period from 17/07/2018 0h to 28/09/2021 0h (local time) with a tolerance of 30%. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).
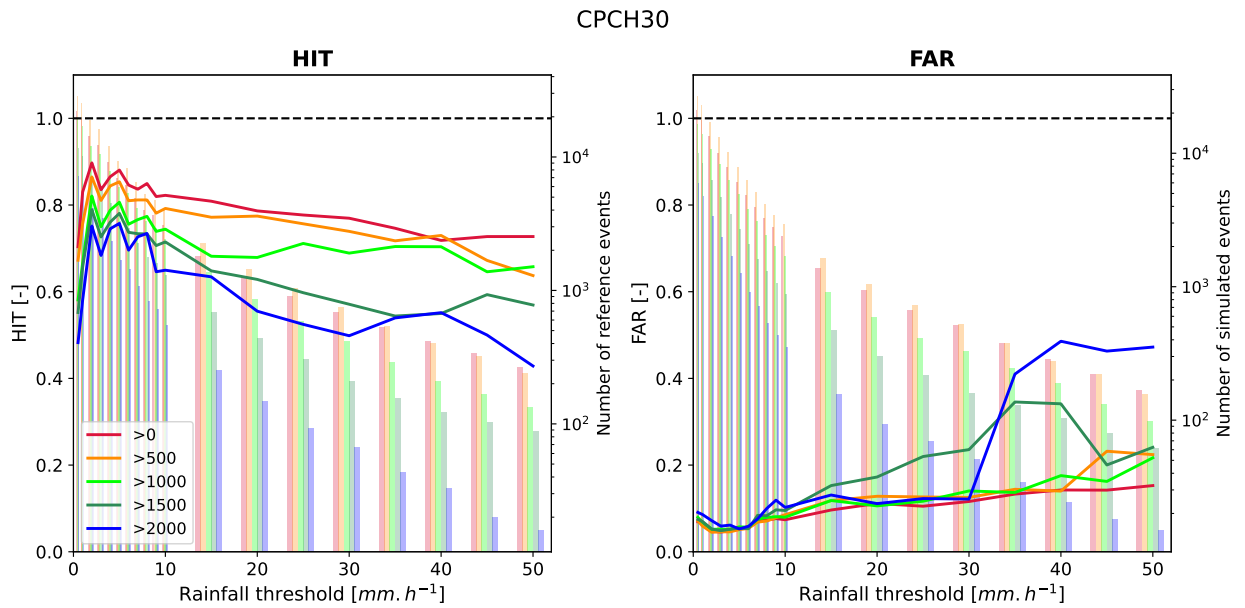
Figure 4.17: HIT/FAR scores of CPCH (10-minute timestep) aggregated by altitudes for the period 17/07/2018 0h-28/09/2021 0h (local time) with a 30% tolerance. The right axes provide the number of reference events (for the HIT score) and simulated events (for the FAR score).

4.7   Performances far from the stations: validation

The previous sections allowed us to characterize the performances of the radar products, and notably of *CombiPrecip*, at *MeteoSwiss* stations. We have notably observed the freedom given to *CombiPrecip* in summer which allowed higher peak precipitations notably in altitudes and in the *Jura*, while *CombiPrecip* leads usually to closer results to the rain gauge measurements at lower altitude and in spring and autumn. However, this analysis is limited to the *SwissMetNet* rain gauge stations, which are the one used in the co-kriging process. Hence, far from these stations, while no differences are expected in the raw radar product AZC, differences might appear in CPCH. In this section, the performances of the co-kriging far from the stations will be analysed. To do so, two independent rain gauge networks are used: the one from Bern canton (BVE network - Public work and transportation direction[b]) and the one from Ticino canton (oasi - *Osservatorio Ambientale della Svizzera Italiana*[c]). They constitute reliable and accurate networks of rain gauges not used in the co-kriging generation of *CombiPrecip*. The rain gauges from the Bern canton are provided by OTT and Lambrecht.

4.7.1   Previous studies

A previous validation has been performed by Barton *et al.*[84] using 132 stations from several independent networks. Four indicators were used to conclude that *CombiPrecip* outperforms the raw radar AZC for all performances: characterization of dry/wet events, of extreme events, mean precipitation estimation, or mean error. Concerning the mean precipitation analysis, they used the *Bias* as indicator which is the logarithm of the volume ratio used in this report (Equation 4.1 with $\overline{i_{rad}}$ and $\overline{i_{RG}}$ the radar and rain gauge precipitation averaged over a given period of time). In the present validation, the indicator will hence also be the *Bias* to be able to compare the results. The analysis period will run over 3 years from 2018 to 2021 while Barton *et al.* used a 7-year period from 2012 to 2018. The timestep of data is 60 minutes which is one of the aggregation period tested by Barton *et al.*

$$Bias = 10 log_{10}\left(\frac{\overline{i_{rad}}}{\overline{i_{RG}}}\right) \tag{4.1}$$

4.7.2   Presentation of the cantonal rain gauge networks
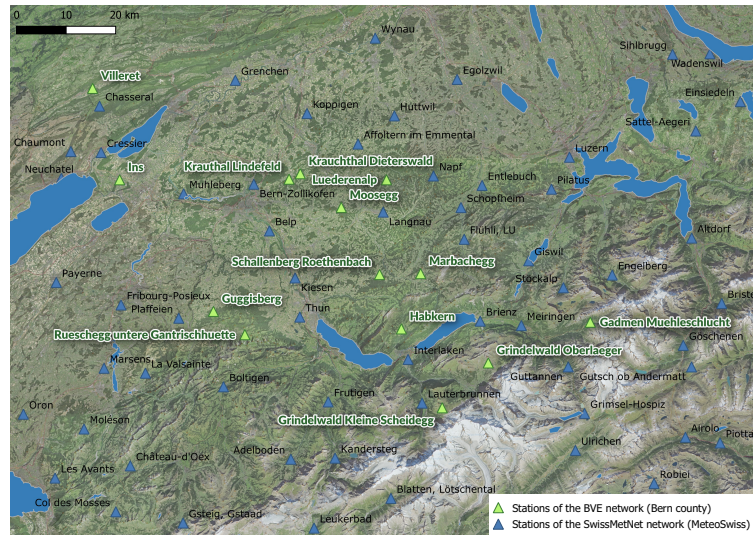
Figure 4.18: BVE network of rain gauges used in the validation of radar performances.

---

[b]https://www.map.apps.be.ch/pub/synserver?project=a42pub_hydromn&view=Ansicht_Messstationen%20Meteo

[c]https://www.oasi.ti.ch/web/dati/selezione-avanzata.html

The Bern canton covers a wide range of terrains from the Jura to the foot of the Alps. The altitudes of the 14 BVE stations used in this section range between 450 m (Ins) and 2060 m (Grindelwald Kleine Scheidegg). The locations of the 14 stations are displayed in Figure 4.18. Compared to the whole Switzerland, as illustrated by Figure 4.20-left, the overall shape is quite similar with slightly more precipitation in summer due to the influence of the Jura. In this figure, it is also possible to observe that the precipitations are in average slightly higher in the BVE network that in the *SwissMetNet* stations from the Bern region. This precipitation bias is mostly constant though slightly higher in summer.

The oasi network, which is presented in Figure 4.19, is denser than the BVE one (19 stations used) and cover a smaller zone. Nevertheless it contains stations with very different characteristics, some being in altitude (*e.g.*, Biasca at 1405 m) and others being closer to sea level (*e.g.*, Moleno at 255 m). Compared to the Bern canton, the precipitations in Ticino are however quite different from the rest of Switzerland, as shown in Figure 4.20-right, with very high precipitations in October. The two networks are similar in Ticino although, as in Bern, the *SwissMetNet* rain gauges have smaller overall precipitations than the oasi network.
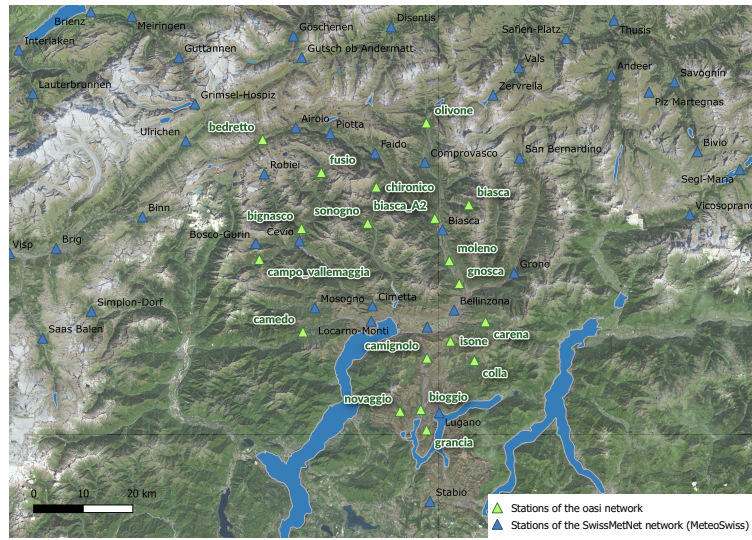
Figure 4.19: oasi network of rain gauge stations used in the validation of radar performances.
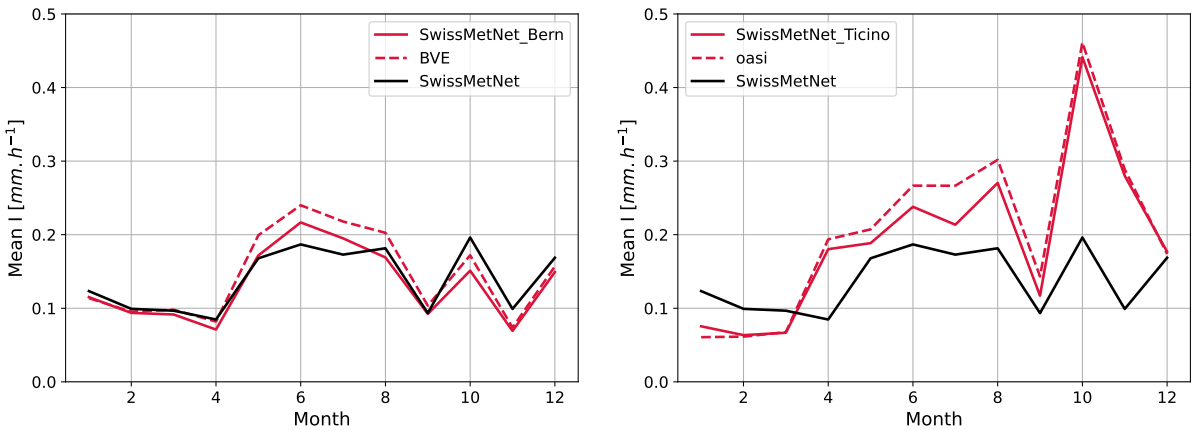
Figure 4.20: Monthly mean precipitation in the periods 2018-2021 for the *SwissMetNet* global network (black solid line), the *SwissMetNet* rain gauges in the region of interest (solid red) and the cantonal networks (red dashed). Left: Bern canton - BVE. Right: Ticino canton - oasi.

4.7.3 Co-kriging performances in the two networks

Figure 4.21 provides the radar mean precipitation biases in the two regions for both the cantonal network and the *SwissMetNet* network. Other analyses (mean monthly precipitation and HIT-FAR scores) are provided in Appendix B.2. The first observation is that in both cases **the raw radar AZC bias is similar between the cantonal network and *SwissMetNet***. While there is a strong underestimation bias in winter, there is nearly no bias in summer in Bern and a strong AZC overestimation in Ticino. However, in the oasi network, the AZC bias is more negative than the *SwissMetNet* one. This is most probably due to the slight global overestimation of the oasi network compared to *SwissMetNet*.

(a) Bern region

(b) Ticino region
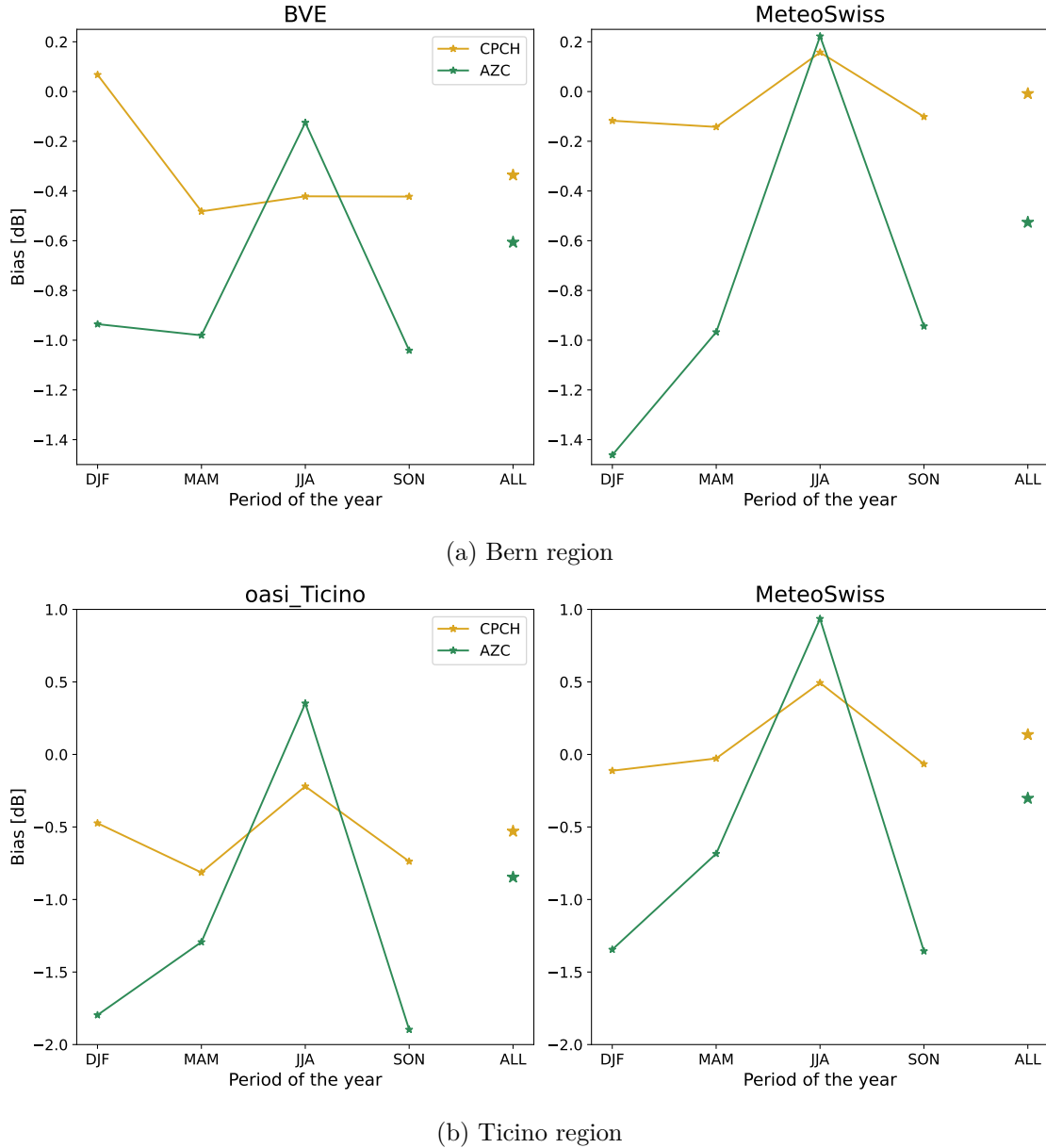
Figure 4.21: Seasonal and global (ALL) volume biases of the AZC and CPCH radar products in the different networks. (top-left) BVE network. (top right) *SwissMetNet* rain gauges of the Bern region. (bottom left) oasi network. (bottom right) *SwissMetNet* rain gauges of the Ticino region.

To observe the influence of the co-kriging process close and far from *SwissMetNet* rain gauges, **the differences between AZC and CPCH biases** need to be analysed. In the Ticino case, the

average bias difference is only slightly larger in the *SwissMetNet* rain gauges (around 0.4-0.5 in both cases). Moreover, the seasonal variations of the bias are similar in both networks, with a higher positive bias in summer and a rather constant bias the rest of the year. This slightly positive summer bias was desired by *MeteoSwiss* which estimates that radar data are more appropriate than rain gauges in summer. The main difference between oasi and *SwissMetNet* is that all biases are shifted down by approximately 0.5-0.6 for the oasi network, due to higher precipitations recorded by its rain gauges. Hence, **CombiPrecip allows to reduce the bias nearly similarly all over Ticino.**

On the contrary in the Bern region, while the co-kriging process clearly affects and reduces the CPCH bias, high co-kriging differences can be observed close to the BVE or *SwissMetNet* stations. Indeed, while the co-kriging has shifted the *SwissMetNet* radar global bias from -0.6 to 0, it has only reduced it from -0.6 to -0.3 for the BVE network. Moreover, the seasonal shape of this bias is not similar in both cases: while the positive summer bias is observed in *SwissMetNet*, it is not present in the BVE network. In winter, radar is also over-estimated by the co-kriging far from the *SwissMetNet* stations. Hence the co-kriging seems not to be working as well in the Bern canton as in the Ticino canton. **A de-biasing may thus be required in the rainfall-runoff models as CombiPrecip, similarly to AZC, is underestimated far from the stations.**

4.7.4   Comparison to the literature

In both the Bern and Ticino cases, the present analysis agrees with Barton *et al.*'s[84] conclusion that CPCH outperforms AZC. However this performance increase appears not to be as good as in this article, notably for the BVE network. Barton *et al.*'s results is reproduced in Figure 4.22 (for the bias indicator). The point of interest is the last point of each curve corresponding to 60-minute aggregation timestep.

The AZC bias shape is similar in all cases with the higher positive bias in summer, followed by spring. The levels of this bias are however different which is due to their high variations depending on the region of interest – differences in levels are for example high between Ticino and Bern cantons, even on *SwissMetNet*. For the CPCH bias, it is close to 0 all along the year in Barton's analysis which is also the case in Ticino and in *SwissMetNet*, but not in BVE. The summer peak in CPCH bias is not present in Barton's article, but this is probably because this peak is regionally dependent and mostly present in Ticino and Jura (both covered by Bern and Ticino cantons, but not by other cantonal networks from Barton's analysis). Hence, **the main difference between Barton's work and the present report is the case of the Bern canton**. It would be interesting to see if the poorer co-kriging performances were already present in Bern in Barton's analysis.
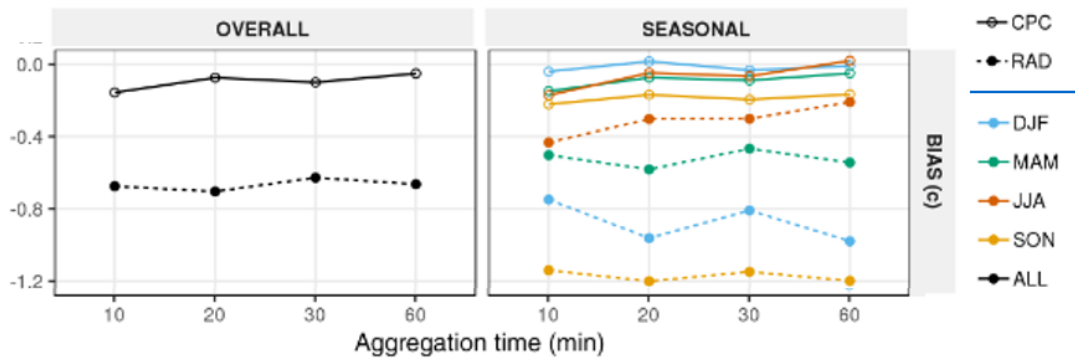
Figure 4.22: Seasonal bias of CPCH (*CPC*) and AZC (*RAD*) compared to the rain gauge for both 132 rain gauge stations independent of *CombiPrecip*. Taken from Barton *et al*[84].

4.7.5 Parameters influencing the loss of performances

To observe if co-kriging performances are modified differently in distinct stations, AZC and CPCH HIT-FAR are represented together stations by stations. This is done both for the cantonal rain gauge networks and for the *SwissMetNet* stations. Such images are provided in Figures 4.23 (Bern canton) and 4.24 (Ticino canton). Increase or decrease of performances are coloured respectively in green and red.

In the Bern canton, the co-kriging has largely improved the performances between AZC and CPCH in all the *SwissMetNet* stations (all stations are largely coloured in green in Figure 4.23b). On the contrary, **in all the BVE stations** (Figure 4.23a), the green area is very small: the co-kriging has only very slightly improved the performances. All stations have similar curves, whatever their altitude or distance to *SwissMetNet*. Hence these parameters, or any other, are not influencing or explaining the loss of performances of the co-kriging far from the stations: **the loss of co-kriging performances is homogeneous all over the region.** The distance to *SwissMetNet* below which the co-kriging is more efficient is thus probably smaller than the smallest distance of the BVE rain gauges (4 km for Villeret).

In Ticino, while the co-kriging is very efficient everywhere in the Ticinese *MeteoSwiss* network (Figure 4.24b), it is less efficient in the oasi network (Figure 4.24a). The performances of CPCH are however much more increased compared to the BVE network – the green zones are bigger. The loss of performances can also be partly explained here by the bias observed between the two rain gauge networks. Concerning the parameters influencing the co-kriging performances, some small differences appear between stations, with better increase of performances in some (Biasca, Novaggio, Grancia) and smaller in others (Isone, Bedretto, Camedo). However these differences are small and **no clear pattern due to either altitude or distance to *SwissMetNet* is observed.**

The influence of altitude or distance to the closest *SwissMetNet* rain gauge does not appear to be present, which has been confirmed when grouping stations by altitude or distance classes. Such analyses are summarized in Appendix B.2.2.

(a) BVE network

(b) *SwissMetNet* network in the Bern region
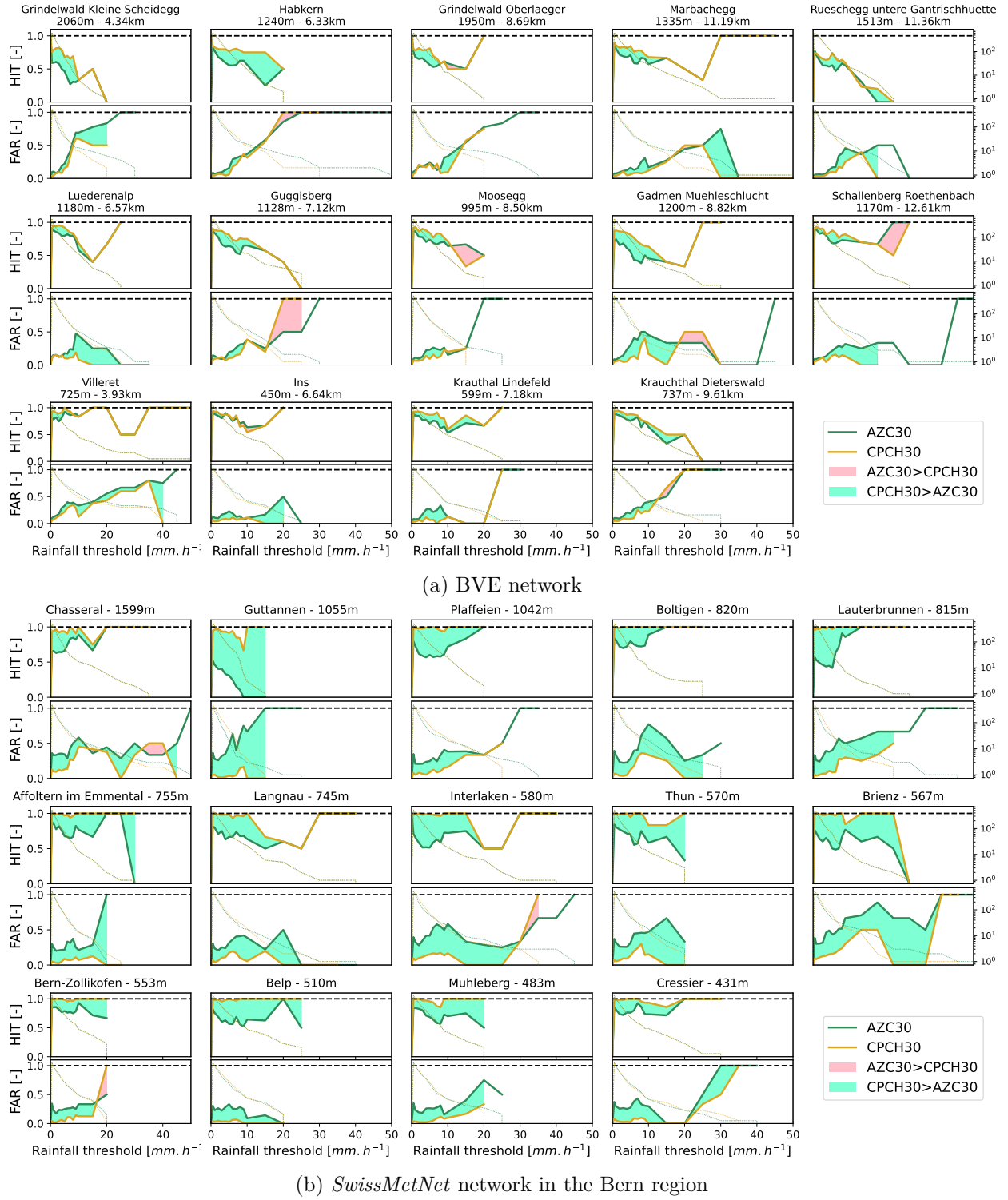
Figure 4.23: Differences in HIT and FAR scores between the AZC and CPCH products for several rain gauge stations. The improvement of performances due to co-kriging is coloured in green (higher CPCH HIT or lower FAR). The higher stations are on the top of the image and the closest to *SwissMetNet* are on the left (for BVE).

(a) oasi network

(b) *SwissMetNet* network in the Ticino region
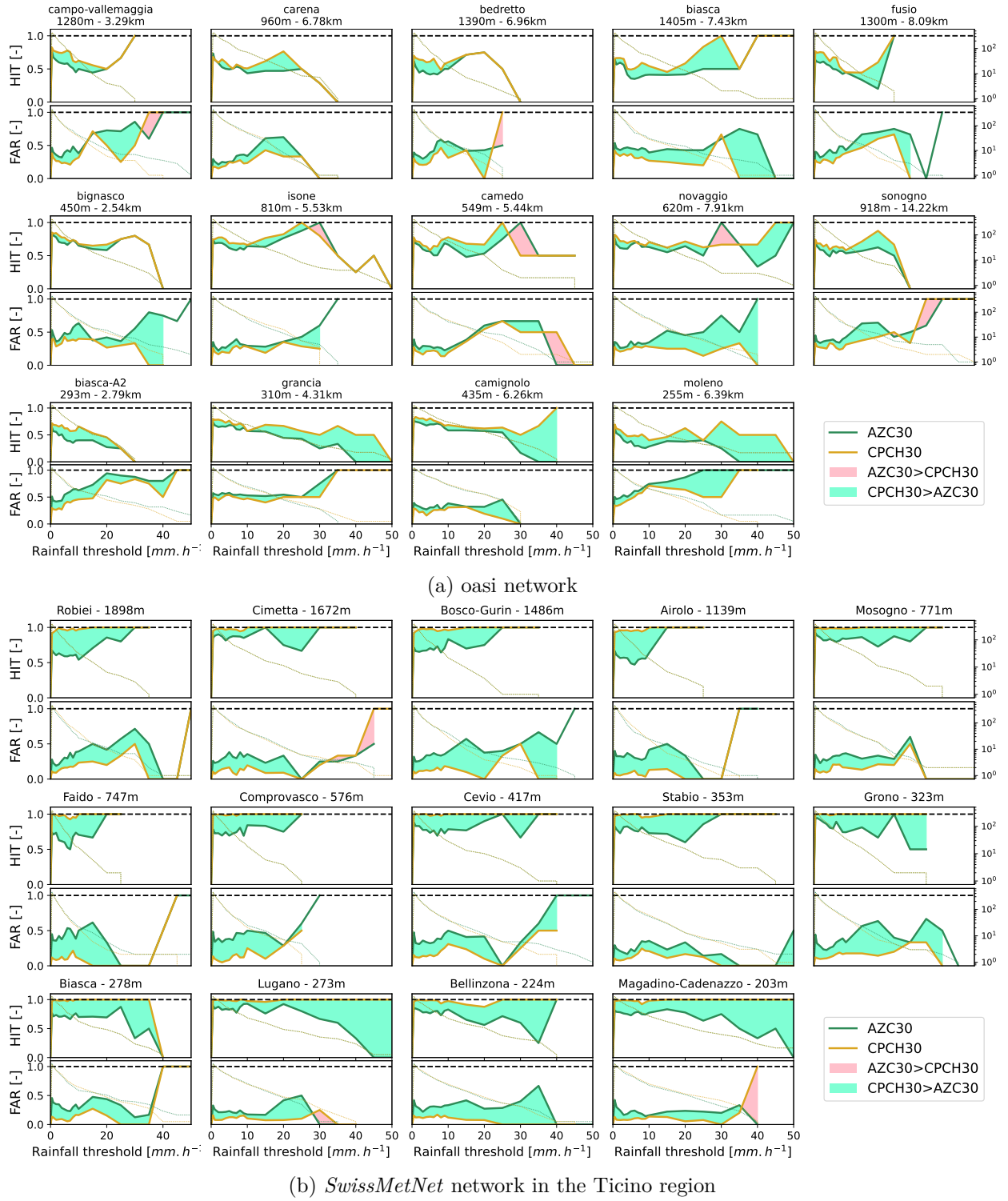
Figure 4.24: Differences in HIT and FAR scores between the AZC and CPCH products for several rain gauge stations. The improvement of performances due to co-kriging is coloured in green (higher CPCH HIT or lower FAR). The higher stations are on the top of the image and the closest to *SwissMetNet* are on the left (for oasi).

4.8 Conclusion

In this chapter, the performances of three radar products provided by *MeteoSwiss* have been evaluated. An overview of the performances and the direction of volume bias according to the product, period and location is proposed in Table 4.3. While the raw radar data (AZC) are **not sufficient to be used** in an hydrological model, the *CombiPrecip* product (CPCH) consisting of the raw data interpolated *via* a kriging process with rain gauge measurements present **largely better results on the *MeteoSwiss* network**. The first timestep of the INCA forecast, based on *CombiPrecip*, has intermediate performances linked with a steady underestimation bias compared to *CombiPrecip*.

However all three products performances depend both on the period of the year (**underestimation of volume in winter, overestimation in summer**), of the intensity of precipitation (**underestimation at low intensity, overestimation at high intensity**), and of the spatial localisation. **Performances are usually better at low altitudes**. The radar data have also evolved over time and had bad performances before 2016, due to lack of spatial coverage of radar units. It will hence be necessary to work **only on a reduced period of time starting in 2017**. The timestep of the data is also an important features as smaller timestep data have **slightly worse performances**. This point will need to be taken into account when hydrological models will be calibrated with a 10-minute timestep.

A caveat of the benchmark proposed here is that it only compares radar and rain gauge data at station sites. However gridded radar data bring much more information as they provide precipitation data on a complete grid with spatial resolution of 1km. Such grid should be used in hydrological model and bring better precipitation information. However, far from the *SwissMetNet* stations used in the co-kriging process, the performances of *CombiPrecip* may deteriorate. This has been observed notably in the Bern canton and it might be the case elsewhere. **A de-biasing factor may then be needed in the rainfall-runoff models**. Nevertheless the co-kriging process is still improving the raw radar performances far from the *MeteoSwiss* rain gauges.

Moreover in this analysis we have considered the rain gauge data as references. Nevertheless they are not perfect measurements notably in some altitude stations, where precipitation are often taking the form of snow and where wind can be important introducing more uncertainty in the precipitation measurements[85].

Now that radar performances have been assessed, their data can be plugged into hydrological models to improve nowcasting performances catchment by catchment. The framework developed here should be used at catchment scale prior to the calibration of a hydrological model in order to know the performances of radar data for the given catchment.

Table 4.3: Overview of performances of CPCH and AZC, close to the *SwissMetNet* stations, depending on altitude, period of the year or class of intensity.

Product	Altitude	Intensity - Season	Volume compared to rain gauges	HIT-FAR performances
CPCH	Low	Summer	↑	
		Winter	-	
		High intensity	↑↑	++
		Low intensity	-	+++
	High	Summer	↑	
		Winter	↓↓	
		High intensity	↑↑	-
		Low intensity	↓↓[a]	++
AZC		Summer	↑↑	
		Winter	↓↓↓	
		High intensity	↑↑↑	- -
		Low intensity	↓[a]	-
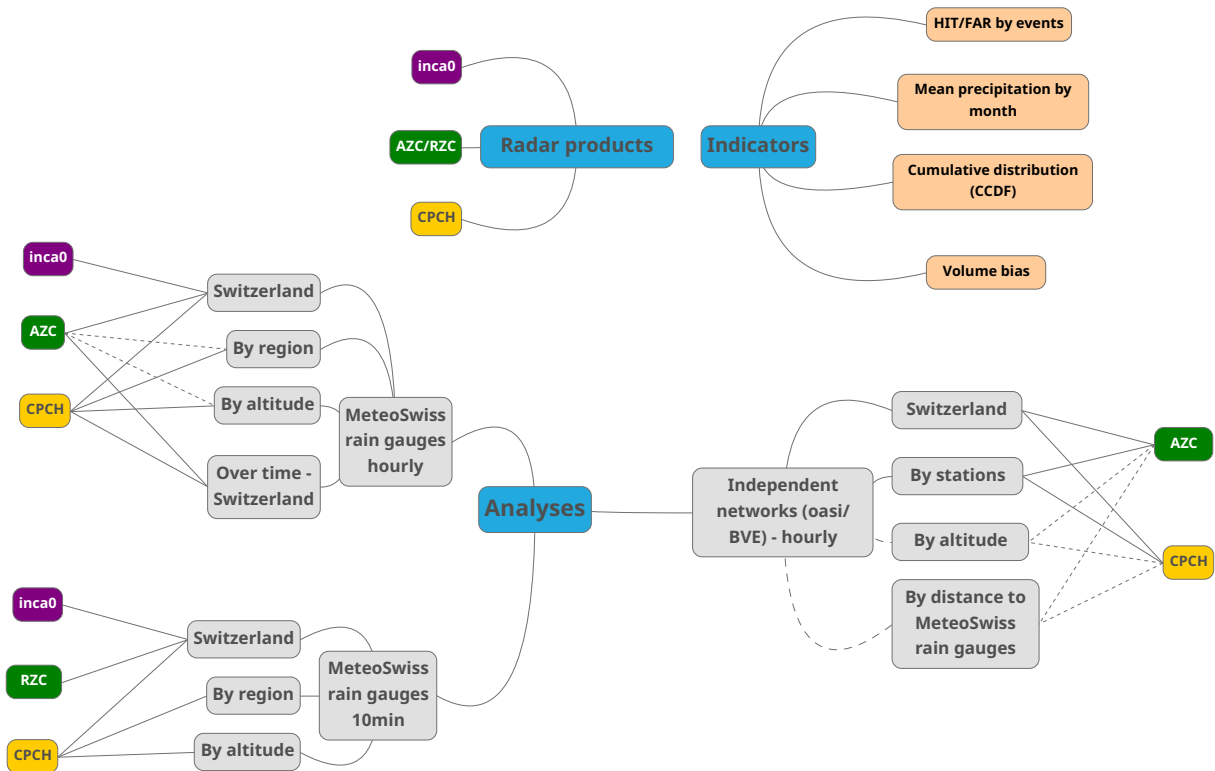
Figure 4.25: Overview of analyses performed in this chapter. Dashed lines represent analyses proposed in Appendix B.

---

[a]more small events

Chapter 5

Integration of radar measurements in rainfall-runoff models for flash floods

5.1 Introduction

A precipitation benchmark has previously been performed between radar and rain gauge measures provided by *MeteoSwiss*. Once we know how the radar observations perform compared to rain gauge data, it is possible to use them to feed *Routing System* (RS), the semi-distributed rainfall-runoff model developed by *Hydrique Ingénieurs*.

Radar data can improve performances thanks to their spatial availability and to their good forecasting skills at low lead time. This chapter will analyse what radar can bring to the forecasting skills of hydrological models for the hydrological basins of Emosson and Zurich-Altbach. For each catchment the focus will first be on observations with an historical calibration of hydrological models before turning to re-forecasts of historical storm events with radar QPFs. The results of several simulations, based either on radar or rain gauges will be compared.

5.2 Emosson

The Emosson catchment is a mountainous basin with a glacial behaviour. It is composed of several sub-basins with different characteristics and sizes and thus constitutes a good candidate to observe the improvements brought by radar data in the rainfall-runoff model. The main results obtained in Emosson are summed up in this section and detailed results can be found in Appendix D.

5.2.1 Calibration of the Emosson model

The Emosson model is first calibrated. Several steps have been required to do so. The main results are summed up here and detailed in Appendix D.

The current operational model, based on rain gauges only, is first updated by changing the temperature station from Aigle to a combination of Montagnier-Bagnes, Les Marécottes and Champéry. Coupled with a new, better calibration of the model, it allows to increase HIT scores and reduce FAR scores by 8% for discharges higher than $15\,\mathrm{m}^3\,\mathrm{s}^{-1}$.

Radar data are then integrated. Two solutions can be used to integrate them. First, in a mathematical approach, it is possible to **calibrate the RS model for each type of input precipitation used** (*i.e.* rain gauges, CPCH, AZC, etc.). This mathematical solution allows to remove some of the errors and uncertainties linked to the input data. A second simpler method, based on the fact that physical phenomena are not affected by the input data, **assumes that the calibration does not need to be changed between rain gauges and radar data**. In that case, only a de-biasing of radar data is needed. Both approaches were developed and compared. **Although the**

first approach improves performances by few percent, it is probably better to invest time in improving the modelling of physical phenomena rather than in calibrating the model for each input data. In both cases, radar data must be de-biased by a factor 1.5 compared to rain gauge data. Indeed radar data (both AZC and CPCH) appear to be biased far from the rain gauge stations, while it was nearly unbiased close to them (Section 3.4.1.2).

Following these conclusions, the model calibrated for rain gauges is improved by adding new features. A dependence of the snow and glacier melt with radiation – and not only temperature – is first introduced in the model to improve the seasonal shape. Transfer coefficients of the glacier and the snow models (KGL and KN) are also tuned to better model the diurnal variations of discharge. More importantly, impermeable areas are introduced in the SWMM object. Depending of the sub-basins, between 0% and 20% of the total area are considered as impermeable area. On this area, water will not infiltrate in the ground but will directly be returned as run-off. The fraction of impermeable area is a parameter which does not affect the general behaviour of discharge but which increases the peak discharge during rainfall events. An example of response to an event in La Fouly sub-basin with and without impermeable area is proposed in Figure 5.1. **The addition of impermeable area is able to increase the peak discharge during precipitation events**, allowing to be closer to the measured peak discharge (in red). The shape of simulated events is also sharper though not enough compared to the reference. However, as observed on the 7[th]October, the drawback is more false alarm events.
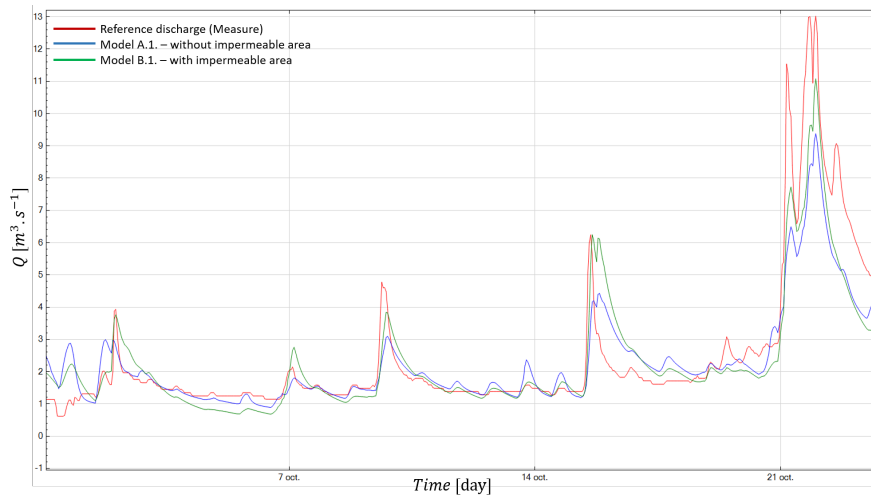
Figure 5.1: Five consecutive events occurring in early October 2019 in La Fouly. The red curve is the reference discharge measured at La Fouly, the blue one is the model without impermeable areas while the green curve has 10% of the sub-basin modelled as impermeable area.

The new model for rain gauge QPE is then de-biased for AZC and CPCH by multiplying them by 1.5 compared to rain gauges. The three models show improved performances compared to the models without impermeable areas. Moreover, the improvements are higher for AZC and CPCH, demonstrating the better resilience to input data of the new model: **the physics behind the model is probably better now. The de-biasing procedure appears to work better with the new models and is now a reliable but simple method**, that allows to obtain similar or better performances with CPCH radar data than with rain-gauges.

Finally, models are tested with 10-minute timestep data. While performances are slightly worsened in La Fouly, they are slightly improved in the smaller catchments of Saleina and Trient. Globally **10-minute data do not affect the models and can be used in place of hourly data without any difficulty.**

5.2.2 Hindcast and event analysis

The 10-minute timestep model with impermeable areas and CPCH input data is then used to perform an update of the model initial conditions. Several runoff forecasts are then produced to analyse the evolution of forecast performances:

- Three perfect forecasts based on observed precipitation data from rain gauges, CPCH or RZC. They differ from continuous simulations by the use of the *update* step before the simulation. These forecasts are only available in re-forecast mode (and not in operational mode). They are called perfect forecasts because they use perfect QPFs (*i.e.* observations).

- A semi-perfect forecast where precipitation are from INCA QPF and temperature and radiation are based on observations. This forecast is only available in re-forecast mode.

- The INCA forecast based on INCA precipitation and COSMO temperature and radiation.

- The COSMO forecast based on COSMOs NWPs.

- Filtered forecasts based on INCA and COSMO previous forecasts and filtered with splines to connect runoff observations and forecasts.

In this section, only the main results are presented. The analysis is based on events observation and alert HIT-FAR scores. Cumulated volume analysis is important for hydroelectricity purposes but is not the main focus of this report. Hence cumulated volume results are provided in Appendix D.3.

Concerning the perfect forecasts, despite slightly better performances at very short lead time for *CombiPrecip*, **nearly no differences are observed between the CPCH and rain gauge input data.** The raw radar data (RZC) have poorer performances but are sometimes more reactive than CPCH and rain gauges. Raw radar QPF is probably not a tool to use alone in operational forecast as its perfect forecast is already quite unreliable. However **it could be interesting in a multi-model alarm system as it provides sometimes alerts before CPCH and rain-gauge perfect forecasts**. In this project raw radar QPF is nevertheless not available and the analysis will thus not be further followed. Detailed results on perfect forecasts are provided in Appendix D.3.2.

5.2.2.1 Introduction of INCA QPF

INCA forecast are then introduced in the model. First, a semi-perfect forecast with measured temperature and predicted precipitation is set up (in blue in the following figures). A true forecast with temperature predictions provided by COSMO1-E is then simulated (in green). Results for a selection of events are provided in Figures 5.2 for Collecteur Est and 5.3 for two events in La Fouly. Alert HIT-FAR scores are also provided by Figure 5.4. The volume analysis can be found in Appendix D.3.3.

Between the perfect *CombiPrecip* forecast (in gold) and the semi-perfect forecast (in blue) – where precipitation are based on INCA forecast – **there is a clear loss of performances** with more disperse forecasts and some missed events (*e.g.*, the second part of the event of 20/10/2019). **The tendency leans globally more towards false alarms**. In La Fouly, similar patterns are observed but the INCA semi-forecast is also much more reactive, being able to double its value in just one hour. **It seems that INCA data help to catch the surges of discharges in this small catchment** – which is supported by the higher alert HIT scores. **The drawback is more false alarms**. In the Collecteur Est, differences are much smaller and bot HIT and FAR alert scores are lower. These differences between La Fouly and Collecteur Est can be explained by two aspects:

- **The size of the catchment**. INCA errors are more easily integrated in a large catchment where they can compensate. For example, if a convective cell is forecasted in the wrong place, it is probable that it will still be located in the catchment area in a large basin. However, in a small basin it can quickly be wrongly outside/inside the catchment which will largely affect performances and notably create false alarms.

- **The nature of the measured discharge**. In La Fouly, the discharge is measured **after a diversion** with a fixed maximum discharge of $13\,\mathrm{m^3\,s^{-1}}$. To maximize the performances of the continuous model, it is probable that the calibration made in La Fouly has been rather overestimated than underestimated. Indeed, as simulated discharges are then cut at a threshold value of $13\,\mathrm{m^3\,s^{-1}}$, and as the reference discharge is in general closer to this value than to 0 (due to the glacial seasonal behaviour), slightly overestimating the discharge leads usually to better scores than the contrary. This was nevertheless not made consciously.

Figure 5.2: CPCH perfect forecast (gold), INCA semi-perfect forecast (blue) and INCA forecast (green) for six events in Collecteur Est. The measure is in red.

The comparison between the semi-perfect forecast and the real forecast, where temperatures are taken from COSMO1-E, enlightens the role of temperature in the rainfall-runoff model. **Forecasts for summer events are only slightly affected by the use of COSMO1-E temperature predictions. On the contrary spring and autumn forecasts are largely affected by it**. This shows the strong model dependency to temperature in these periods where the rain-snow limit is determinant for the discharge estimation. Summer discharges are less affected as the rain-snow limit is high enough and the only impacting factor is the snow reservoir level which only depends on observed temperatures. Concerning the direction of the changes, the real INCA forecast is sometimes less reactive (02/10/2020) and sometimes more reactive (28/10/2018) than the semi-

perfect forecast. The uncertainty on temperature is thus random and not systematic, which is further confirmed by the alert HIT-FAR scores which do not display clear differences between the semi-perfect and the real INCA forecasts.

(a) 20/10/2019

(b) 22/07/2020

Figure 5.3: CPCH and INCA forecasts for two events in La Fouly.

Figure 5.4: Alert HIT and FAR scores of CPCH perfect forecast (gold), INCA semi-perfect forecast (blue) and INCA forecast (green) at Collecteur Est and La Fouly. A tolerance of 10% is applied on dashed lines while no tolerance is used for solid lines. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

5.2.2.2 Comparison with COSMO forecast and post-processing effects

Figures 5.5 and 5.6 display a comparison between INCA and COSMO forecasts, respectively in Collecteur Est and La Fouly. Alert HIT-FAR scores are provided in Figure 5.7 for the two forecasts with and without post-processing while Figure 5.8 shows filtered forecasts for some events in Collecteur Est. The post-processing allows to connect the forecasts to the latest measured data (see Section 3.2.5). The intermediate report of the *Radar4Infra* project[221] demonstrated in the case of Emosson that the filtering process improved a lot the performances, notably at short lead time. A filter period of 3h is used. The volume analyses can be found in Appendices D.3.4 and D.3.5.

(a) 28/10/2018

(b) 27/08/2019

(c) 20/10/2019

(d) 22/07/2020

(e) 28/08/2020

(f) 02/10/2020
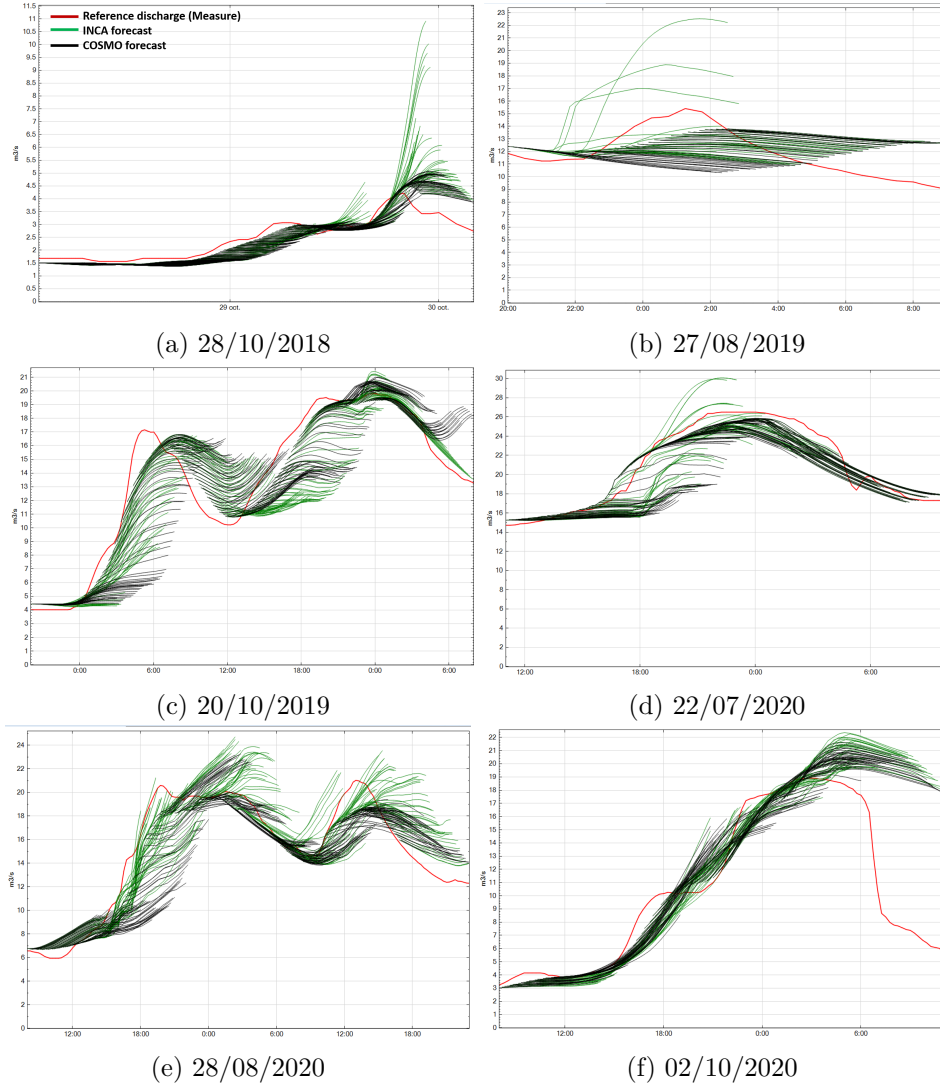
Figure 5.5: INCA (green) and COSMO (black) forecasts for six events in Collecteur Est.

(a) 20/10/2019

(b) 22/07/2020
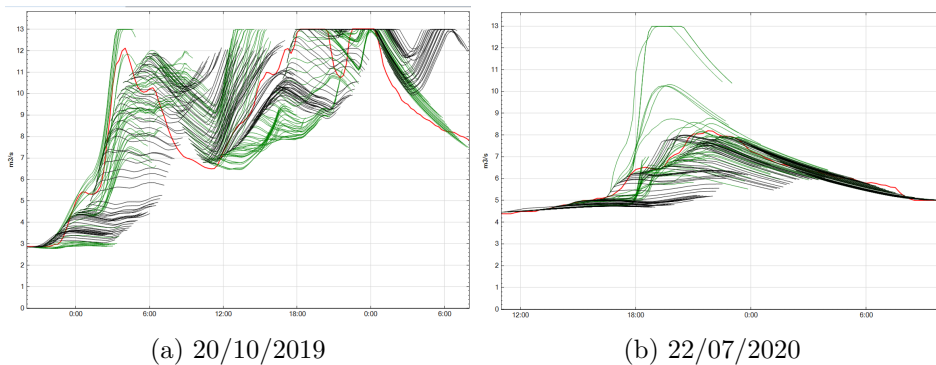
Figure 5.6: INCA and COSMO forecasts for two events in La Fouly.

In the events, differences between INCA and COSMO are globally small but two trends can be underlined. First **INCA is usually more reactive than COSMO.** The growths of discharge are observed sooner (*e.g.*, during the 20/10/2019 and 28/08/2020 events). Secondly, INCA is **more susceptible to produce false alarms** with the striking example of the 28/10/2018 event. As in the previous section, this behaviour is amplified in La Fouly. The first event (20/10/2019) has a very good INCA forecast (notably in its first part) while the second one (28/08/2020) displays a strong false alarm. These two events are much less reactive in Collecteur Est. **The higher reactivity of INCA, notably in La Fouly, is further confirmed by alert scores**. HIT and FAR scores are higher for INCA than for COSMO in Collecteur Est and, more significantly, in La Fouly.
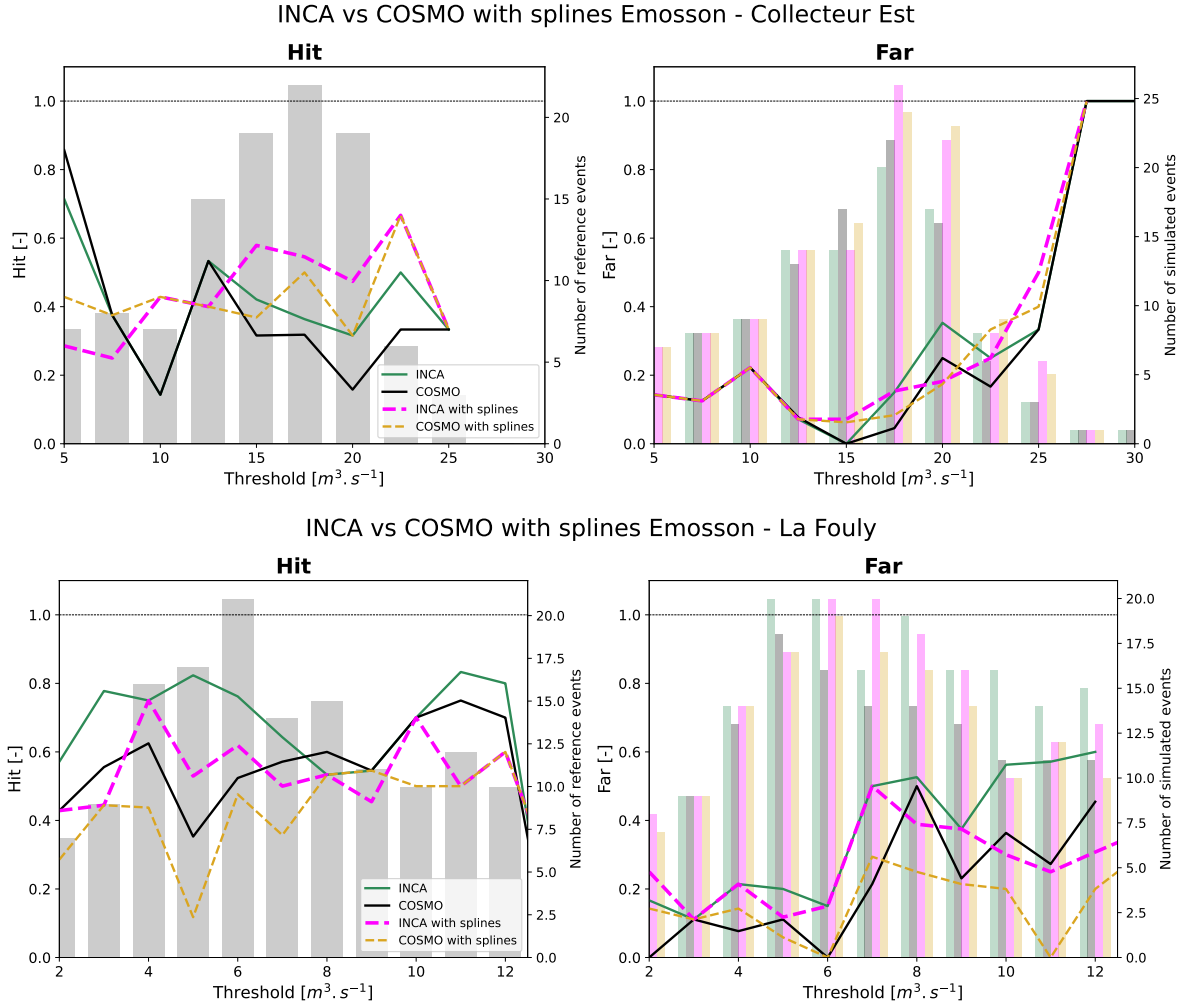
Figure 5.7: Alert HIT and FAR scores of INCA (green) and COSMO (black) forecasts at Collecteur Est and La Fouly. The magenta and gold dashed curves are respectively for INCA and COSMO forecasts once filtered. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

(a) 28/10/2018

(b) 27/08/2019

(c) 20/10/2019

(d) 22/07/2020

(e) 28/08/2020

(f) 02/10/2020
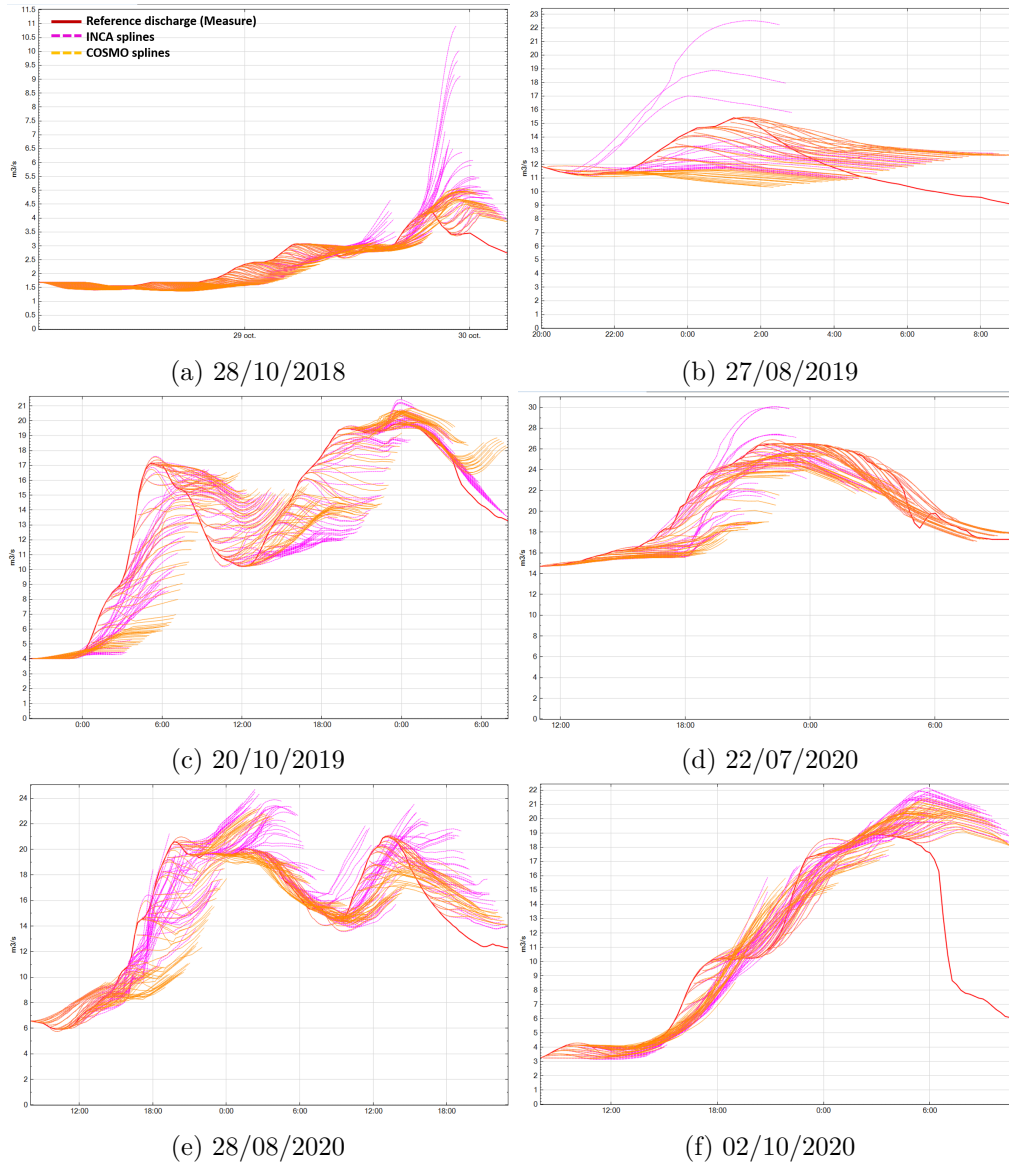
Figure 5.8: INCA (magenta) and COSMO (gold) forecasts with spline filters for six events in Collecteur Est. The measure is in red.

In Collecteur Est, events are better forecasted when applying splines notably when the forecasts were not reliable (*e.g.*, COSMO the 20/10/2019, both forecasts the 27/08/2019). Alert HIT scores are also largely improved for both INCA and COSMO. Despite smaller differences between INCA and COSMO, INCA seems to still be more reactive. The filtering step is also not able to remove the false alarms as in the case of the 28/10/2018. However, in La Fouly, HIT scores are worsened. This is probably due to the better pre-filtering performances observed there (around 70% of alert HIT for INCA compared to 40% in Collecteur Est). The filtering step is probably more efficient for big catchments with smaller reactivity than for smaller catchments. Indeed, it does not use the first timesteps of the runoff forecasts where INCA has added very localized information allowing to catch the discharge surges. **Filtering appears to be useful in the large Emosson catchment with 4-6h response time but detrimental in the smaller basin of La Fouly.**

5.2.3 Conclusion

The Emosson catchment is a rather complicated catchment where a **variety of different processes** are at stake: glaciers and snow cycles, underground slow or fast storages, direct runoff on impermeable areas during intense precipitations, water interceptions by vegetation or glaciers, interactions of previous phenomena, etc. Its location in altitude and its complex orography constitute other interesting properties suggesting radar data may improve rainfall-runoff models and predictions. Finally, it contains several sub-basins of different scales allowing to investigate the impact of INCA radar forecast on different locations.

This complex framework allows to develop a method to calibrate a *Routing System* model adapted to radar gridded data. A **de-biasing of precipitation input data** is first demonstrated to be essential to obtain good rainfall-runoff models. Though a complete calibration of the model according to the nature of the input precipitations leads to slightly improved results, it has been demonstrated that investing time on the modelling of the physical processes allows to get similar results between rain gauge and *CombiPrecip* data. The final model is thus robust to any change of input type. On a simpler catchment, it is reasonable to think that such a process will still be reliable as a calibration can only be performed on a smaller number of parameter.

In operational mode, several steps need to be performed once the model is set up. First an *update* of the initial conditions has already been proven to be necessary[6,221]. With this step, compared to COSMO, **INCA seems to introduce more reactivity on the prediction of flash events, notably in the small-response time catchment of La Fouly**. The other side of the coin is **increased false alarm rates**. Reflections on the requirements and design of an alarm systems will need to be done to balance HIT and FAR scores. Raw radar forecasts, not available in this project, could also be useful in a multi-model alarm systems thanks to their ability to detect some events not observed in other products. However their overall performances are too low to build an alarm system relying only on them.

The temperature forecast plays an important role in spring and autumn which highlights the necessity of an operational monitoring of the model notably during these periods. Finally, the post-processing appears to improve the alert performances for intermediate response time catchments (Emosson-Collecteur Est, 4-6h) while being detrimental in smaller basins (La Fouly, 1-2h). Post-processing will need to be investigated further as it probably needs to be adapted for small basins (smaller filtering period for example).

Cumulated volumes, which are analysed in detail in Appendix D.3, are **nearly only affected by the post-processing step, which largely improves its predictions at small lead times**. However, they are nearly not affected by filtering for larger lead times. When forecasts are needed for cumulated volume predictions rather than for flash-flood alerts, a post-processing of output data will be required. The type of input data (rain-gauge, CPCH or RZC QPEs or INCA or COSMO QPFs) is nearly not modifying volume scores.

5.3 Zurich airport-Altbach basin

The Altbach basin, close to Zurich, is a countryside basin with several small urban portions. It is described in Section 3.4.2. HIT-FAR scores by events will be used for the assessment of continuous models while alert HIT-FAR scores will be studied in the hindcast process.

5.3.1 Calibration of the Zurich model

5.3.1.1 Calibration and de-biasing of precipitations for four types of input data

Due to the short response time of Zurich catchment, it is necessary to work directly with 10-minute data. The model for Zurich-Altbach is first calibrated with rain gauge stations. A de-biasing process is then used to obtain the CPCH and RZC models (which are identical) and the inca0 model that will be used in the hindcast process. The inca0 model is slightly modified compared to CPCH-RZC model, due to the underestimation observed in inca0 precipitation intensity (Figure 3.14 for Winterthur). On the contrary the CPCH/RZC models have the same *coeffI* as the rain gauge model (0.9), as the average precipitation ratio between CPCH/RZC and rain gauges is close to 1 in Zurich. The main parameters of the model are provided in Appendix C.2.1.

The Nash scores of the four models are provided below in Table 5.1. The volume ratios between simulated and measured discharges are also provided. They show that inca0 and RZC could have still be de-biased a little bit more to reach a volume ratio closer to 1. However we prefer to keep the higher Nash scores rather than biding to volume ratios exactly equal to 1. In any case, all these volumetric biases are small. With Nash scores, the CPCH model appears to be the best model. At this point, radar data seem to improve quite significantly the performances at all discharges. Raw radar (RZC) data are however performing poorly everywhere. Finally the change from *CombiPrecip* to inca0 is nearly not affecting the performances of the model.

Table 5.1: Nash scores of Zurich-Altbach models. The best scores are coloured in green. Volume ratios are also provided.

Model	Nash for $Q >$				Volume ratio
	0	$0.5 \, \mathrm{m^3 \, s^{-1}}$	$1 \, \mathrm{m^3 \, s^{-1}}$	$1.5 \, \mathrm{m^3 \, s^{-1}}$	
Rain gauges	0.65	0.53	0.36	0.09	1.06
CPCH	0.72	0.61	0.43	0.2	1.01
RZC	0.56	0.35	0.07	-0.29	0.94
inca0	0.71	0.59	0.39	0.17	0.94

As no seasonal behaviour is observed in the reference discharge from Figure 3.13, HIT and FAR scores are computed by event as explained in Section 3.3.2.3. The results are provided in Figure 5.9. It is important to note that these scores cannot be compared to those in Emosson as they are not computed in the same way. However it is also possible to notice that significant events (higher than $3 \, \mathrm{m^3 \, s^{-1}}$ are most of the time not well simulated.

The best model in HIT/FAR is the one using rain gauge data. Indeed, despite slightly more false alarm at low discharges, it is able to catch some of the few high discharge events that other models are not able to catch. Although this is due to only a small number of events (only 9 events are higher than $5 \, \mathrm{m^3 \, s^{-1}}$ and 4 than $8 \, \mathrm{m^3 \, s^{-1}}$), this is significant as the goal of this project is to improve the flash flood forecast performances of the model. The worst model is, as could be expected, the RZC model with significantly higher FAR and lower HIT values. inca0 and CPCH are performing nearly as well with slightly better performances for CPCH, probably due to the 10 minutes of delay in inca0 values (inca0 values correspond to the first timestep of the forecast, predicted at +10 minutes).

In conclusion, the model fed with **rain gauge data seems to be the one performing the best for flash flood prediction**, though the *CombiPrecip* model is better at low discharges and for Nash scores.
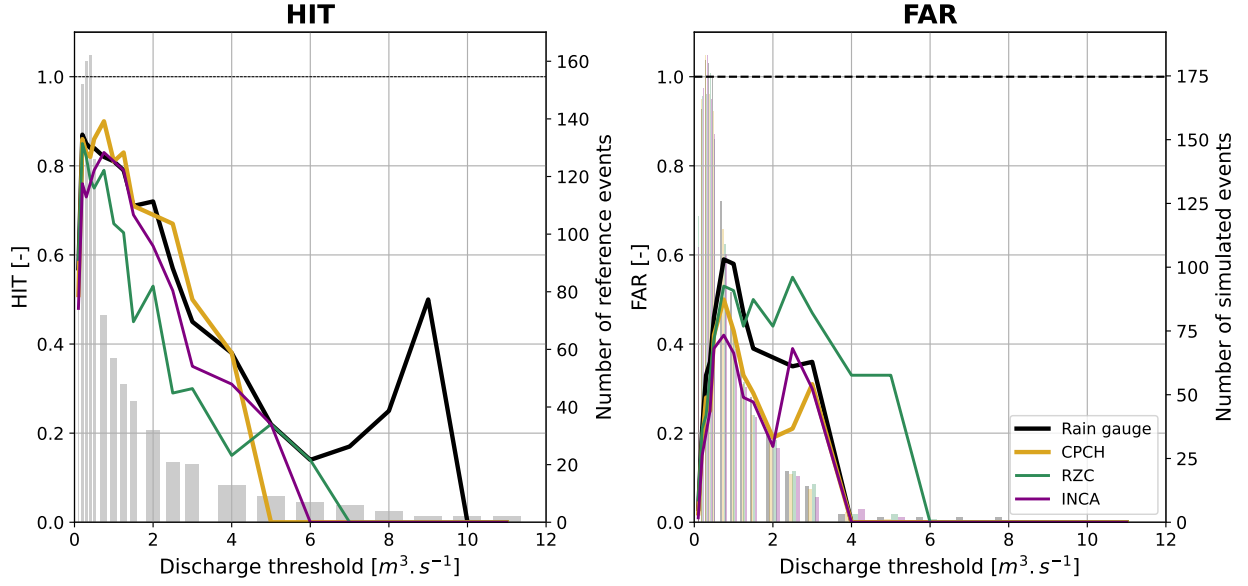
Figure 5.9: HIT/FAR scores of the discharge from the four models in Zurich. HIT and FAR are computed by events with a tolerance of 10%. A temporal tolerance of 20 minutes is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

5.3.1.2 Seasonal debiasing

Figure 5.10a displays the evolution of the volume ratio along the year for the four previous continuous simulations. Though small, the volume ratios all seem to show seasonal bias with too high volumes in February and March and too small volumes in summer and mostly autumn. Hence a seasonal adjustment of precipitation inputs could be an interesting improvement to investigate.

Three seasonal de-biased models are thus created for the CPCH-inputs. The de-biasing is now made month by month rather than with a constant factor all over the year. The goal is to reach the HIT performances of the rain-gauge model without creating too much FAR. Table 5.2 provides the de-biasing coefficients used for each models. These coefficients are multiplied by *coeffI* which is still equal to 0.9 for CPCH models. The volume ratios obtained in the three models (and in the initial CPCH and rain gauge stations ones) are plotted in Figure 5.10b. The first debiased model is designed by taking approximately the inverse of the volume ratio observed in the initial CPCH model. However, as observed with the red curve, it does not result in volume ratios equal to 1, probably because if more precipitations are added in month i it could mean more discharge in month $i+1$ and not only in month i, depending on the ground conditions. A second, simpler, debiasing model is thus set up. Here, debiasing factors are chosen to be more constant over the seasons (high in summer, low in winter). In the volume ratio plot (blue curve), it results in too big volume ratios notably in summer. Thus an intermediate model 3 is created with smaller summer debiasing adjustment factors.

Nash scores of the three models (and of the initial CPCH one) are provided in Table 5.3. HIT and FAR scores are finally plotted in Figure 5.11. The Nash scores are usually similar (rather smaller than bigger) than the initial CPCH models: their is no clear increase in performances at this stage. Their volume ratios are also globally larger than initially. HIT scores have been slightly increased, notably at high discharges in quite the same amplitude for the three models. However they do not

reach the performances of rain gauge stations for the few very high discharge events and the cost is an increase in FAR notably at high discharges. The three models are hence proposing different trade-off between HIT and FAR scores but are not able to really improve CPCH performances. Hence, **in the Zurich case, it seems that a seasonal de-biasing is not significantly useful and that the rain gauge model is the best one**. The *update* of initial conditions of the model will be performed with this rain-gauge model and input data.
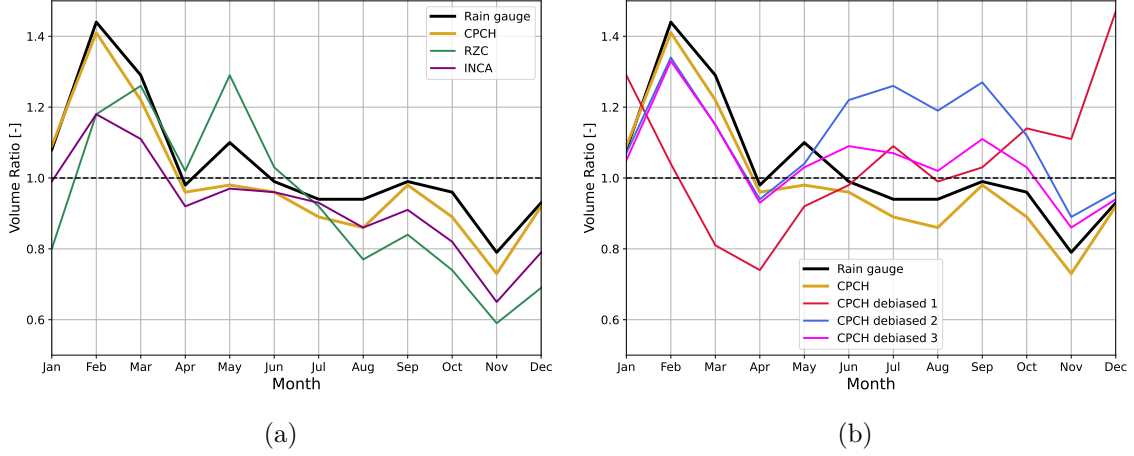
Figure 5.10: Seasonal evolution of volume ratios (simulated divided by observed). Figure 5.10a: four initial models from Section 5.3.1.1. Figure 5.10b: three debiased models for CPCH input data.

Table 5.2: Monthly de-biasing adjustment factors.

Model	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
Debiased 1	0.9	0.6	0.8	1	1	1.05	1.1	1.1	1	1.1	1.25	1.1
Debiased 2	0.95	0.95	0.95	1	1.05	1.2	1.2	1.2	1.2	1.05	1	0.95
Debiased 3	0.95	0.95	0.95	1	1.05	1.1	1.1	1.1	1.1	1.05	1.05	0.95

Table 5.3: Nash scores of Zurich-Altbach models with seasonal de-biasing. The best scores are coloured in green. Volume ratios are also provided.

Model	Nash for $Q >$				Volume ratio
	0	$0.5\,\mathrm{m^3\,s^{-1}}$	$1\,\mathrm{m^3\,s^{-1}}$	$1.5\,\mathrm{m^3\,s^{-1}}$	
CPCH	0.72	0.61	0.43	0.2	1.01
CPCH debiased 1	0.69	0.58	0.45	0.26	1.08
CPCH debiased 2	0.65	0.52	0.35	0.16	1.13
CPCH debiased 3	0.71	0.61	0.46	0.27	1.06
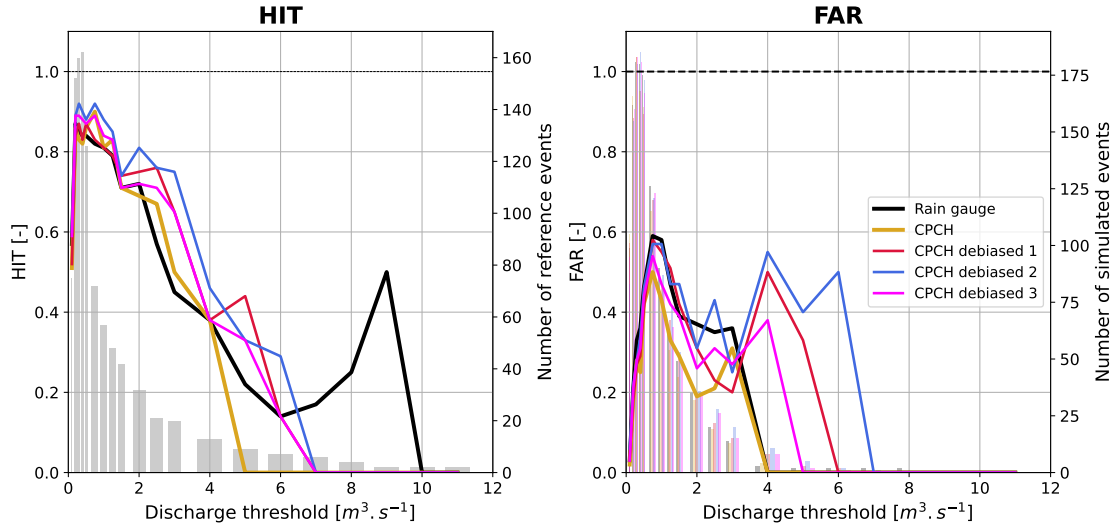
Figure 5.11: HIT/FAR scores for the discharge from the three seasonally-debiased models in Zurich. Scores for the initial CPCH and rain-gauge stations models are also provided. HIT and FAR are computed by events with a tolerance of 10%. The temporal tolerance is put at 20 minutes. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

5.3.2 Hindcast and event analysis

The *update* of the initial conditions is performed with rain-gauge model and data as they provide the best HIT performances without FAR at high discharge. Several forecasts are then re-forecasted:

- Four perfect forecasts: rain-gauge stations, CPCH, RZC and inca0.
- The INCA forecast (with COSMO1-E forecast for temperature).
- A COSMO forecast to highlight radar QPFs interest.
- Filtered forecasts (COSMO and INCA).

As the goal of the Zurich-Altbach project is to generate alerts for flash floods, the cumulated volume analysis is not the focus of this report. Results are however provided in Appendix E.

5.3.2.1 Perfect forecasts

Figure 5.12 displays the results of the re-forecast for the four perfect forecasts on eight events representative of the diversity of situations. Some events are flash flood events with high precipitation intensity (13/09/2018, 23/06/2021 and 28/06/2021) and sharp discharge increase and decrease, other endure small but steady precipitations over several days (20-21/05/2019) and some are intermediate cases with medium precipitations over a rather small time (13/02/2020, 10/06/2020 or 18-19/06/2021). Finally the event from January 2021 is a rain-on-snow event. Alert HIT-FAR scores are then presented in Figure 5.13 for the four perfect forecasts with a 10% tolerance (dashed lines) or without tolerance (solid lines).

Usually all events are simulated similarly by the rain gauge, CPCH and inca0 perfect forecasts: the events of 20-21/05/2019 and 10/06/2020 are not observed by any perfect forecast and the other ones are partially observed (13/09/2018 and the three events of June 2021) or well-observed (13/02/2020 and 28-30/01/2021). **All models seem to have difficulties to describe the rapid discharge responses of the sharp events**. However, some differences can be observed between the perfect forecasts, with the rain gauge ones being better for some events (*e.g,* on the 13/02/2020 or 23/06/2021) and worse for others (*e.g.,* on the 13/09/2018, 18-19/06/2021 or 28/06/2021). **The anticipation time is usually similar** for these three perfect forecasts. Inca0 and CPCH have moreover very similar behaviours with only small but systematic differences with slightly higher inca0 previsions.

Alert HIT and FAR scores highlight **the similarity of the three previous perfect forecasts**: all curves are close one to the other. However, some mild HIT differences point towards slightly better performances for inca0 and CPCH perfect forecasts than for rain gauges. Concerning the addition of tolerance, it can be observed that dashed curves are quite close to the solid curves: the majority of missed events is missed by more than 10%.

As in Emosson, the RZC perfect forecast behaviour is more chaotic, with more false alarms (28-30/01/2021) or less reactivity (13/09/2018, 20-21/05/2019 or 13/02/2020). However, unlike in Emosson, the tendency leans more towards underestimation of discharges than false alarms, as reflected by the HIT scores. FAR scores are on the contrary similar to CPCH. Due to this last observation, **raw radar data can be expected to be less useful in a multi-model alert system than in Emosson, as they are giving less early alarms**. This could be due to the good reliability of the rain gauge network near Zurich and to the subsequent good reliability of the co-kriging process in *CombiPrecip*. These observations will need to be evaluated and confirmed in future studies with the raw radar QPFs.

(a) 13/09/2018

(b) 20-21/05/2019

(c) 13/02/2020

(d) 10/06/2020

(e) 28-30/01/2021

(f) 18-19/06/2021

(g) 23/06/2021

(h) 28/06/2021
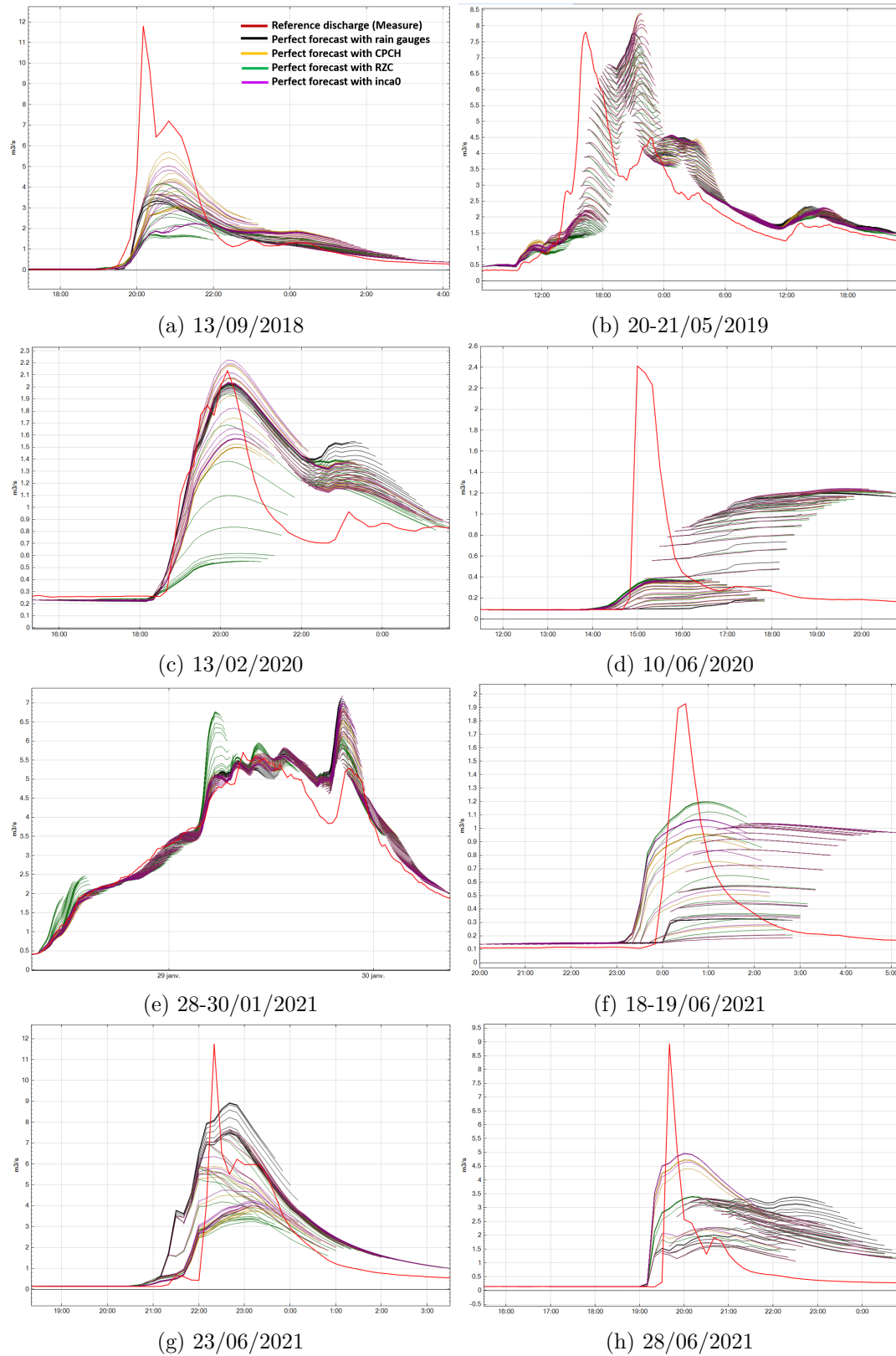
Figure 5.12: Rain-gauges (black), CPCH (gold), RZC (green) and inca0 (purple) perfect forecasts for eight events in Zurich-Altbach discharge station. The measure is in red.
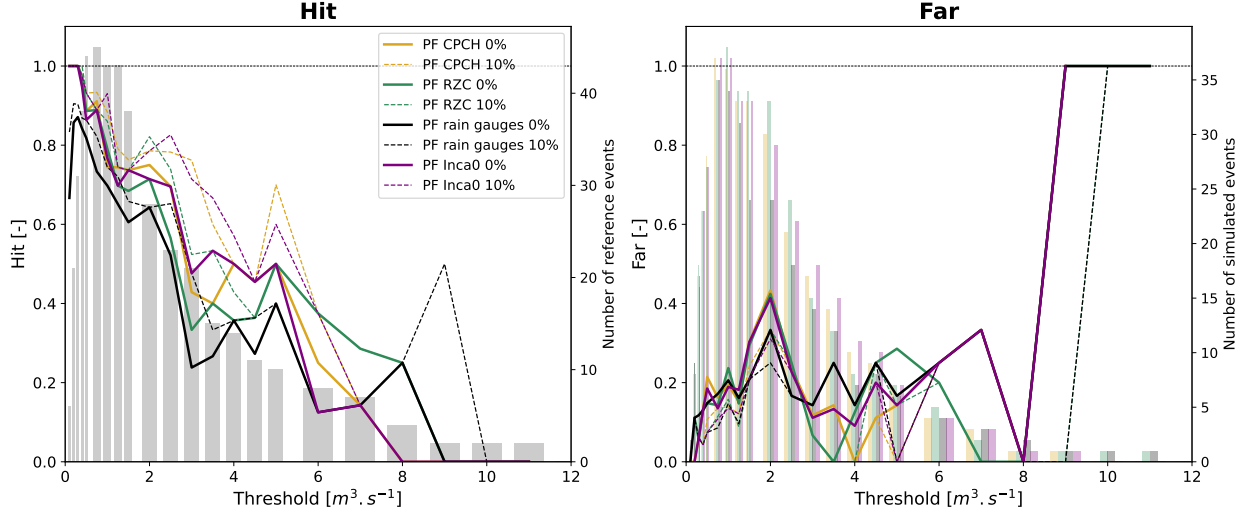
Figure 5.13: Alert HIT and FAR scores of the four perfect forecasts at Zurich-Altbach discharge station. A tolerance of 10% (dashed lines) or no tolerance (solid lines) is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

5.3.2.2 INCA and COSMO forecasts

In Zurich, snow is only present in few events in January. Hence, it is not useful to create a semi-perfect forecast with perfect temperature previsions and real INCA precipitation previsions as the temperature data will not affect the results[a]. In this section, the CPCH perfect forecast (in golden), the INCA forecast (in green) and the COSMO forecast (in black) are compared. Results for 8 events are displayed in Figure 5.14 while alert HIT-FAR are provided by Figure 5.16.

In most events, a clear lack of performances can be observed between the perfect *CombiPrecip* forecast and the INCA forecast. It results either in **missed or underestimated events** (all the first part of the 28-30/01/2021 event and the majority of the forecasts during the 13/09/2018 and the three June 2021 events), in **delayed observations** (notably on the 28/06/2021) or in **false alarms** (in the end of the missed 2021/05/2019 event and very clearly for two forecasts during the 18-19/06/2021 small event). Notably for the 18-19/06/2021 event, one forecast is predicting incredibly high discharges (up to $14\,\mathrm{m^3\,s^{-1}}$) higher than the most important observed events. This is due to a very high small convective cell observed in some INCA forecasts during this event. Figure 5.15 provides the INCA precipitation forecasts for several lead times over one of the upper sub-basins of the catchment. In some of the forecasts some heavy thunderstorms are forecasted explaining the high false alarm observed in the simulation. It is probable that the cell was very localized and finally crossed a neighbouring catchment instead of the Altbach basin. Such observation underlines **what can bring an ensemble precipitation product which would help to simulate several discharge scenarios and launch an alarm containing the discharge intensity of the event and the probability related to this discharge** (and not only a forecasted discharge that can finally be ten times higher – or ten times lower – than the observed discharge).

This tendency to false alarm is observed in Figure 5.16 where CPCH and INCA HIT have similar values (slightly higher for INCA) while FAR are largely higher at large discharges in INCA. Notably several false alarms are generated above $9\,\mathrm{m^3\,s^{-1}}$ (three at $9\,\mathrm{m^3\,s^{-1}}$, two at $10-12\mathrm{m^3\,s^{-1}}$ and still one at $13-14\mathrm{m^3\,s^{-1}}$). Nevertheless, the INCA forecast seems, as in Emosson, to be **more reactive than the *CombiPrecip* perfect forecast**. This is confirmed by the slightly higher HIT scores of the INCA forecast.

---

[a]This has been checked and the only impact of temperature is on the 28-29/01/2021 which is a rain-on-snow event.

(a) 13/09/2018

(b) 20-21/05/2019

(c) 13/02/2020

(d) 10/06/2020

(e) 28-30/01/2021

(f) 18-19/06/2021

(g) 23/06/2021

(h) 28/06/2021
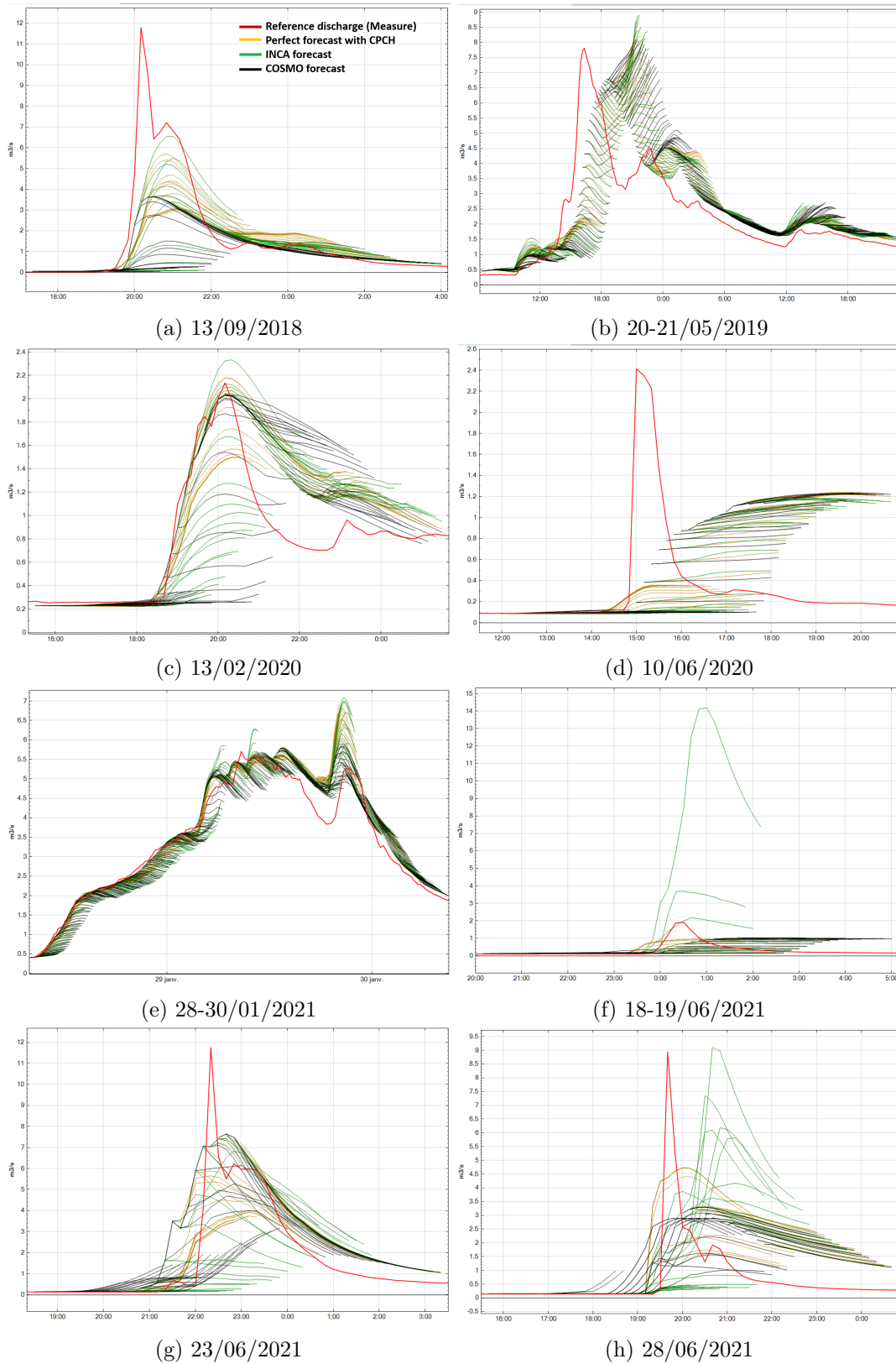
Figure 5.14: CPCH perfect forecast (gold), INCA semi-oerfect forecast (blue) and INCA forecast (green) for eight events in Zurich-Altbach discharge station. The measure is in red.
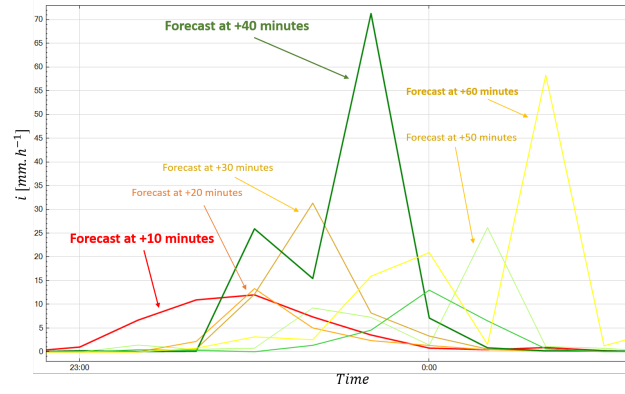
Figure 5.15: INCA forecasted rainfall intensities over one of the upper sub-basin of the Zurich-Altbach catchment during the event of the 18/06/2021. Forecasts for several lead times are given. At 23h10 and 23h20 a convective cell with huge intensities is forecasted to pass in the basin at respectively 23h50 and 0h20 (with 40 and 60 minutes of anticipation). It results in two high discharge predictions in Figure 5.14 made at 23h00 and 23h20.

Concerning the COSMO forecast, the events display the poor performances of the COSMO model for flash flood forecasting. At the exception of the small event of the 13/02/2020, COSMO is usually underestimating the peak discharge of the events. COSMO can sometimes also miss an event that INCA will forecast, such as on the 18-19/06/2021. It results in lower alert HIT scores than INCA. However, as explained above, INCA is much more susceptible to false alarms which is also observed in the FAR scores. COSMO is notably not generating any false alarm higher than $8\,\mathrm{m}^3\,\mathrm{s}^{-1}$.
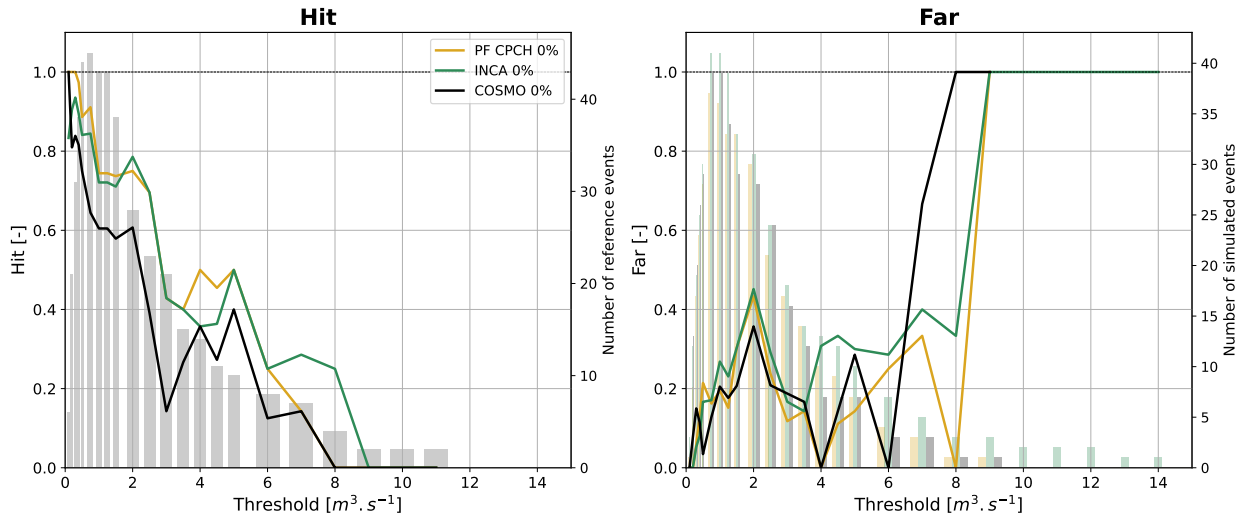
Figure 5.16: Alert HIT and FAR scores of the three previous forecasts at Zurich-Altbach discharge station. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

5.3.2.3 Improvements of forecast performances with post-processing

Splines were applied to the COSMO and INCA forecasts with a filtering period of three hours. The results for eight events are displayed in Figure 5.18. Alert scores are provided in Figure 5.17.

In Zurich, splines with a filtering period of three hours generate huge false alarms, as observed in most events. Moreover they erase all differences between INCA and COSMO. Such observations could have been expected as the filtering process forget all values of the forecasts located before the filtering period. In Zurich most of the important predictions are located at very short lead time both in COSMO and INCA and hence the post-processing is relying only on forecast points where differences between INCA and COSMO are nearly insignificant. Modifying the filtering period could improve the forecast by allowing the filter to connect the observations to more significant predictions at shorter lead time. **Nevertheless, for the flash flood events rising in one or two time steps, the filter cannot work properly as it relies in the initial slope of the model. Whatever the filtering period, the interpolation will start with a steep increase while the real discharge will already be decreasing**, as in the 13/09/18 event.

Unsurprisingly, FAR scores are largely increased. Discharges up to $36\,\mathrm{m^3\,s^{-1}}$ are forecasted while the measured maximum is just below $12\,\mathrm{m^3\,s^{-1}}$. HIT scores are also decreased. **For the design of an alert system with sharp events, post-processing of forecasts with splines does not appear to be suitable**. Splines may be useful either in larger catchment with higher response time or for the estimation of flood volumes.
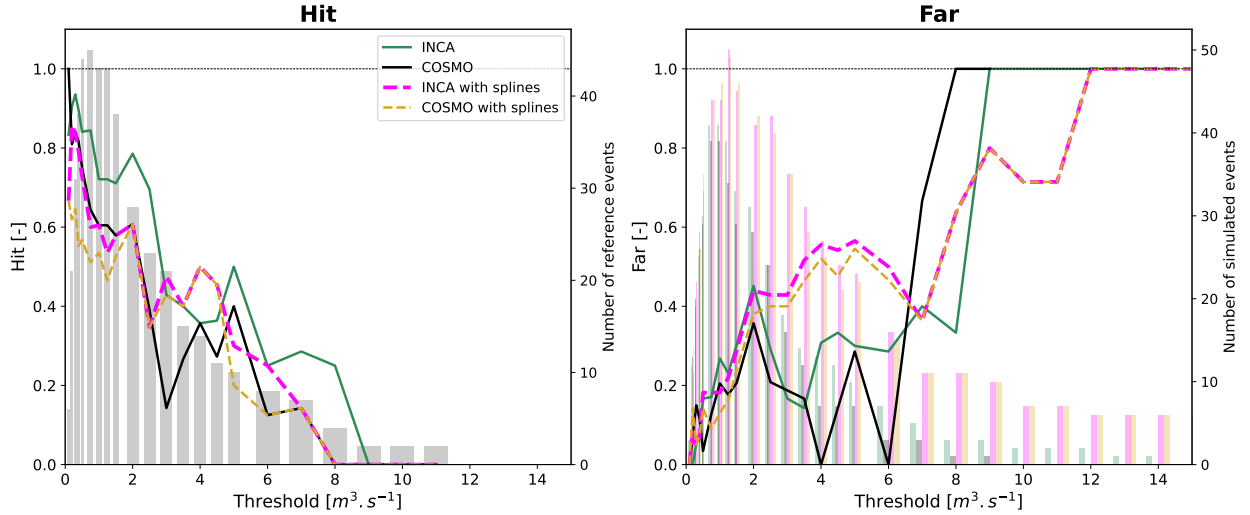
Figure 5.17: Alert HIT and FAR scores of non-filtered forecasts (INCA: green- - COSMO: black) and filtered forecasts (INCA: magenta - COSMO: gold) at Zurich-Altbach discharge station. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

(a) 13/09/2018

(b) 20-21/05/2019

(c) 13/02/2020

(d) 10/06/2020

(e) 28-30/01/2021

(f) 18-19/06/2021

(g) 23/06/2021

(h) 28/06/2021
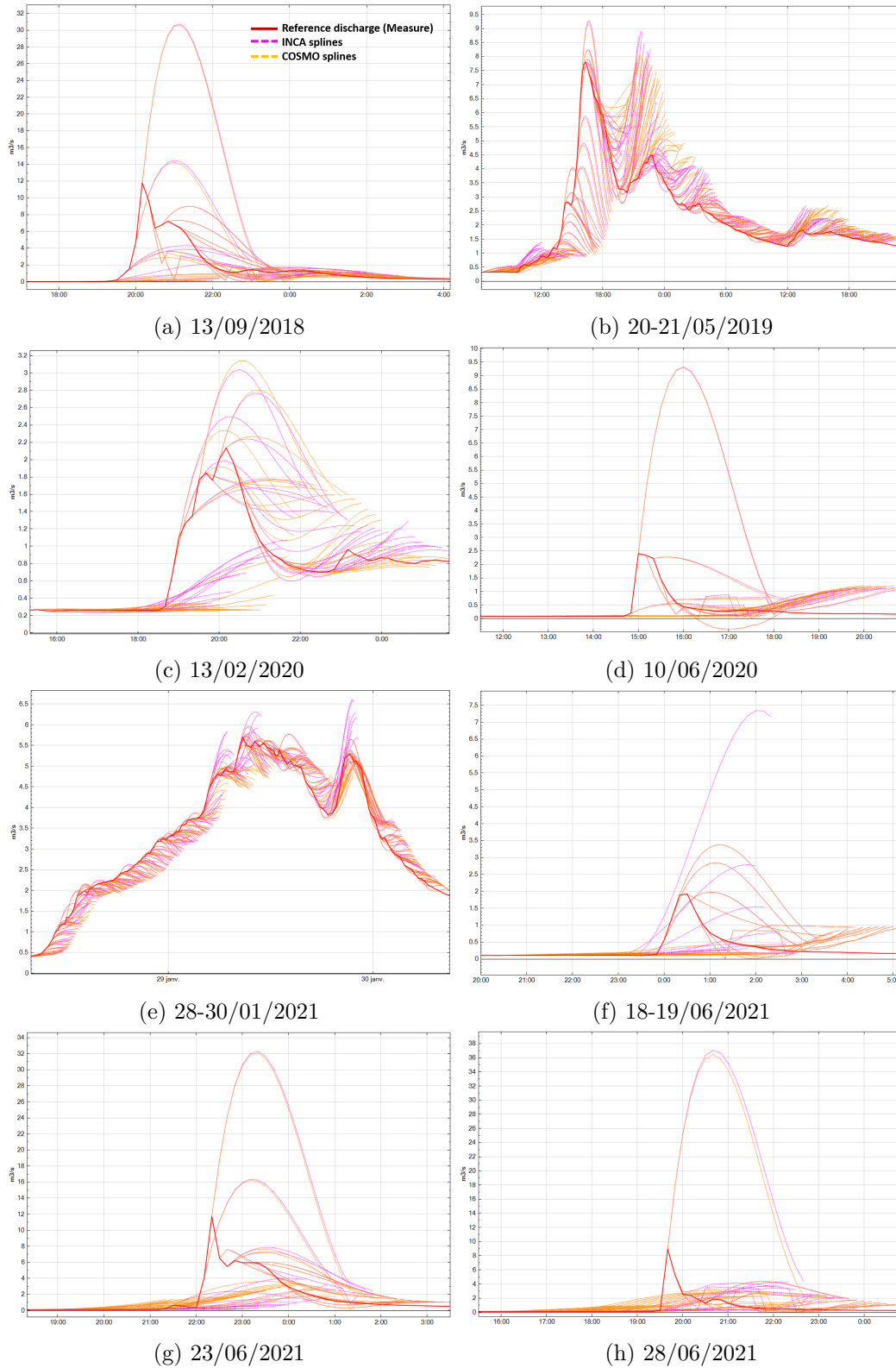
Figure 5.18: INCA (magenta) and COSMO (golden) forecasts after being filtered with splines for eight events in Zurich-Altbach discharge station. The measure is in red.

5.3.3 Adaptation of the continuous model

5.3.3.1 Modifications brought to the model

The previous models have displayed difficulties to announce the higher discharge events, with alert HIT values going down quite quickly (below 0.5 since $3\,\mathrm{m}^3\,\mathrm{s}^{-1}$) and high FAR scores. These events are unfortunately the most important ones, as the goal of the modelling in the Zurich-Altbach catchment is to provide flash flood alerts to evacuate the Zurich airport commercial zone. The threshold level for evacuation has been fixed at $20\,\mathrm{m}^3\,\mathrm{s}^{-1}$. Post-processing also appeared not to be suitable in the case of Zurich catchment high events.

The difficulties in this catchment arise from the dual nature of events. Indeed, some events are characterized by a slow discharge increase due to either snow melting or lasting small intensity precipitations (for example the event of January 2021), while others are marked by a quick and sharp discharge increase followed by a slightly slower decrease. For example, on the 13/09/2018, the discharge is passing from $1\,\mathrm{m}^3\,\mathrm{s}^{-1}$ to $12\,\mathrm{m}^3\,\mathrm{s}^{-1}$ in only 20 minutes and directly decreases afterwards. It is hence complex to simulate these two types of events with one model.

Moreover, a careful observation of sharp events reveals the existence of two peaks. After a first sharp and high peak, a second lower and smoother peak is observed 20 to 40 minutes later. As observed in Figure 5.19, this double-peak phenomenon is not present in the discharge measured by the Bassersdorf discharge gauge, located in the middle of the Altbach catchment. This station had not been used in the calibration of the first models. The presence of two peaks is thus probably due to different travel times of water in the different sub-basins. This could also be due to the path of precipitation cells which are usually passing from West to East in the opposite direction of the water flow. Hence the downstream catchment responds **quickly and sharply** while rivers are **delaying and smoothing** the response of the upper catchment (above Bassersdorf), creating the two peak process of Figure 5.19.

(a) 13/09/2018   (b) 23/06/2021
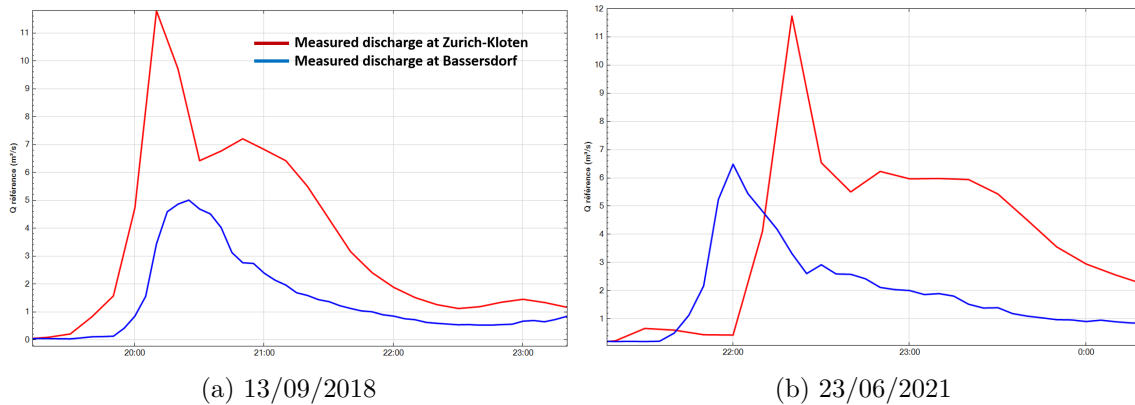
Figure 5.19: Comparisons of the discharge measured at Zurich-Kloten and at Bassersdorf for two flash flood events.

For that reason, and because the continuous models have proven in Figure 5.9 to be improvable, new continuous *RS* models are set up. Several modifications are brought to the model:

- **Some parameters are differentiated between the upstream and downstream sub-catchments** to make emerge the two peaks of discharge – and, as a consequence, to better catch the real levels of the discharge. For example, in the downstream sub-catchment, the soil storage (GR3) maximum level is increased. More water can then be stored in this catchment allowing to accelerate the decrease of discharge after a sharp flash flood event and to better split the two peaks. This could be explained by the higher fraction of forest present there. This

differentiated calibration is allowed by the knowledge of the Bassersdorf discharge measures which help to calibrate the upper sub-basins. The parameters that have been modified are listed in Appendix C.2.2.

- **River sections are added in several places to delay the peaks of discharge and to split the two peaks from the two sub-catchments**. Rivers are separated in several sections as it helps to numerically simulate the wave propagation. For the same total river length, splitting the river in several sections allows to slightly increase the discharge delay while slightly decreasing the peak attenuation. This is a numerical effect due to the increase of the number of degrees of freedom of the wave model. The location and number of river sections is provided in Appendix C.2.2.

- The timestep of computation is decreased from 10 minutes (600 seconds) to 150 seconds. This modification does not affect most of the processes as the precipitation will not be different during 4 consecutive timesteps. Nevertheless, it is useful to improve the simulation of the wave propagation in rivers by increasing the temporal resolution of this process. The drawback is an increase of the computation time.

- Finally, in order to improve the timing, peak value and shape of the few sharp high events which are usually under-estimated, **the parameter Imp, representing the fraction of impermeable area, is modified**. Following Peredo *et al.*[196], the influence of the intensity of precipitation is introduced in the infiltration model. This modification is described below and is applied to the downstream sub-basin in one model. Another model is set up without this modification for comparison purposes.

The modification of the model to strengthen the response to intense precipitations – and to improve its timing – is inspired from Peredo *et al.*[196]. While Peredo *et al.* played directly on the infiltration rate from their production store (see Section 2.3.5), the modification is here introduced in the impermeable fraction. This is easier to do in the code and has shown better preliminary results. Imp is hence modified according to Equation 5.1 which is illustrated by Figure 5.20. It varies with a sigmoid function between an initial value Imp_0 for no precipitation[b] and a maximum value Imp_{max} for high intensities. The parameters $i_{1/2}$ and i_0 represent respectively the half intensity at which $Imp(i_{1/2}) = \frac{Imp_{max}+Imp_0}{2}$ and where the slope is maximum, and the sharpness of the transition from Imp_0 to Imp_{max}. All these parameters have been calibrated to obtain the best continuous simulation: Imp_0 and Imp_{max} are calibrated to obtain respectively good low intensity events and good high intensity sharp events; $i_{1/2}$ is modified to fix the limit between the two types of events and i_0 is chosen to have an appropriately sharp transition.

$$Imp = Imp_0 + \frac{Imp_{max} - Imp_0}{1 + exp\left(-\frac{i-i_{1/2}}{i_0}\right)} \tag{5.1}$$

This process models the fact that, at high intensity, a higher fraction of water is directly flowing and cannot be infiltrated even on normally permeable areas. It could also model other phenomena such as waste water treatment overflow. The addition of this impermeable fraction variation allows to better split the two discharge peaks during sharp events while keeping good results for other events.

---

[b]The value at $i = 0$ is not rigorously equal to Imp_0 but is very close to it.
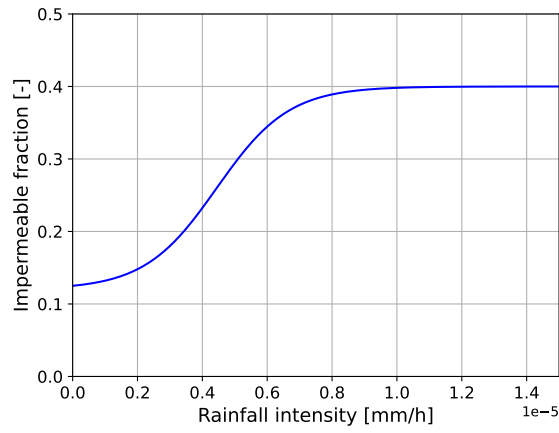
Figure 5.20: Fraction of impermeable area in the downstream sub-catchment according to the intensity of precipitation.

(a) 13/09/2018

(b) 01-02/07/2019

(c) 23/06/2021

(d) 28/06/2021
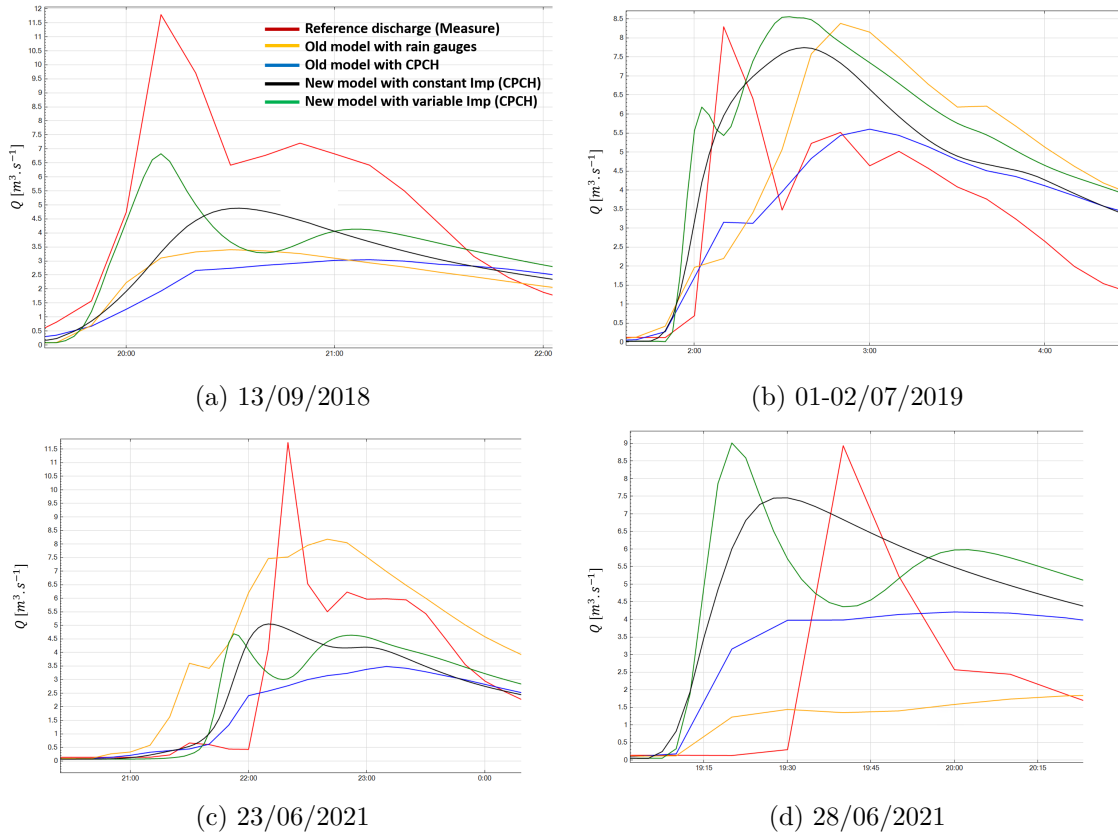
Figure 5.21: Comparisons of the best old and new continuous models for the four major sharp events. The measure is in red and the old models are in yellow (stations) and blue (CPCH). The new CPCH models are in black (constant impermeable area) and green (variable impermeable area).

The results of the two new models (with or without variable Imp) are provided by Figure 5.21 for the four main flash flood events – which are also corresponding to the four higher discharges measured in the period. The old CPCH and rain gauge models are also provided for comparison. In each of these four events, and for both new models, both the timing, the peak level and the peak shape have been improved compared to the previous CPCH model and in most cases to the rain gauges model. The double peak appears clearly in the 13/09/2018 and 23/06/2021 events with a smoother and lower peak. However, it does not appear in 2019 due to a first peak being too late.

Finally for the 28/06/2021 event the double peak appears but is not present in the measure. This is due to an advance of the first peak. **The model is still very sensitive to precipitation input data and a small shift in time or space can change the shape of the simulated discharge**. Concerning the differences between the model with or without variable Imp, the shape and timing is usually better in the model with variable Imp, notably for the first event. However this is not as determinant as all the other modifications proposed.

Nash and HIT-FAR scores are respectively displayed in Table 5.4 and Figure 5.22. As proposed in the literature[191,192] the approach has been more focused on events than on the whole continuous period – even if it was not a calibration only on events. As a consequence, the Nash on the whole period and for discharges below small thresholds have been worsened. However at higher discharges the Nash value has been improved. Similarly the HIT scores has been improved mostly at high discharges, reaching more than 50% up to $6\,\mathrm{m}^3\,\mathrm{s}^{-1}$. The event of the 23/06/2021 is however still not estimated as well as with rain gauge data. The drawback is a FAR increase. However this increase is largely reasonable, notably compared to the old model with rain gauge data. Finally the differences between the model with or without variable Imp are small: Nash values are similar and the model without it seems only slightly better in HIT at low discharges. A new model with rain gauges as input data has also been set up and show poorer performances than the CPCH models. Indeed, the old model with rain gauges was already high on the main peaks. On this model, there is no double peak in the main events as it is not possible to create them with only two rain gauge stations.

Figure 5.22: HIT/FAR scores for the discharge from the three new models and the two best old models in Zurich. HIT and FAR are computed by events with a tolerance of 10% and a temporal tolerance of 20 minutes. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

Table 5.4: Nash scores of Zurich-Altbach new models. The best scores are coloured in green. Volume ratios are also provided.

Model	Nash for $Q >$				Volume ratio
	0	$0.5\,\mathrm{m^3\,s^{-1}}$	$1\,\mathrm{m^3\,s^{-1}}$	$1.5\,\mathrm{m^3\,s^{-1}}$	
Old rain gauges	0.65	0.53	0.36	0.09	1.06
Old CPCH	0.72	0.61	0.43	0.2	1.01
New CPCH - variable Imp	0.65	0.56	0.41	0.27	1.01
New CPCH - constant Imp	0.66	0.56	0.42	0.28	1
New rain gauges - var Imp	0.47	0.34	0.18	-0.09	1.06

As the calibration has been mostly focused on only few flash flood events – although the other smaller events have been observed in order not to worsen their simulation – a validation on the year 2017-2018 can be performed. Indeed *CombiPrecip*, *MeteoSwiss* and *HydroZurich* data are available between the 1st September 2017 and the 1st September 2018. Two major flash flood events occurred in this period and are represented in Figure 5.23. Nash and HIT-FAR scores are also provided in Table 5.5 and in Figure 5.24. It can be observed that the events were completely missed by the two old models and notably by the rain gauge one. They are now a lot better simulated while the small events have not been worsened according to the HIT-FAR and Nash scores. The model with variable Imp is also clearly better on the 08/08 event. However this is only for one event and it is not possible to say that this model will always be better than the one with constant Imp. Nevertheless, once again, the shape of the peaks is better on the former. **The validation confirms the initial observations and prove the necessity to introduce the new features listed above**.

(a) 08/06/2018

(b) 03/07/2018
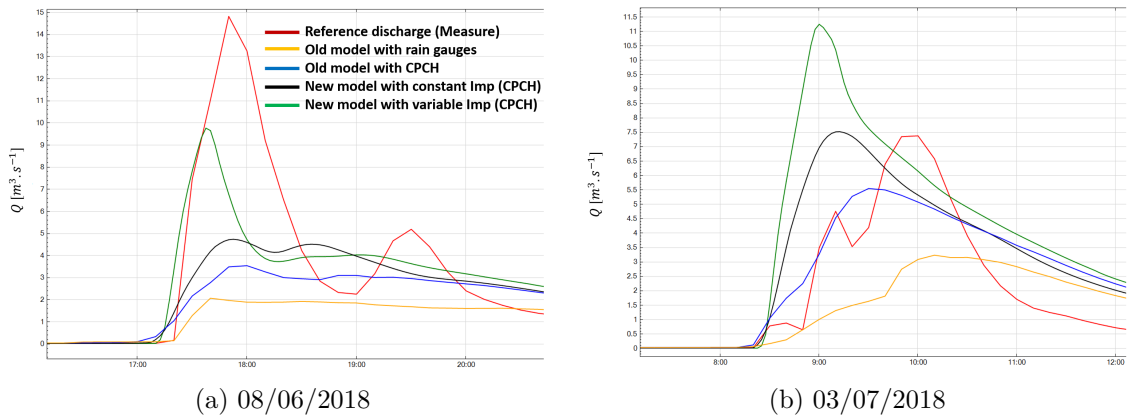
Figure 5.23: Comparisons of the best old and new continuous models for the two major sharp events of the validation period. The measure is in red and the old models are in yellow (stations) and blue (CPCH). The new CPCH models are in black (constant impermeable area) and green (variable impermeable area).

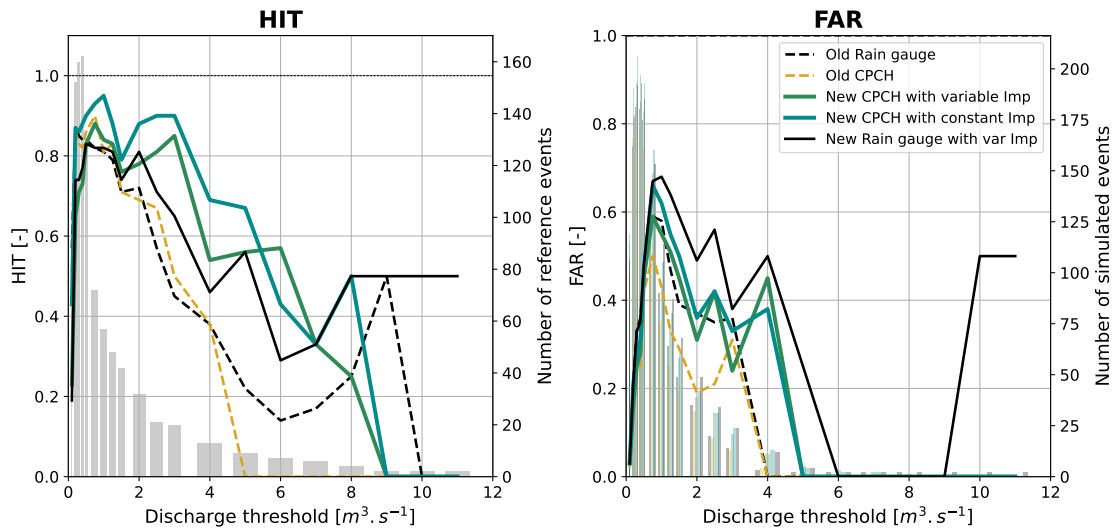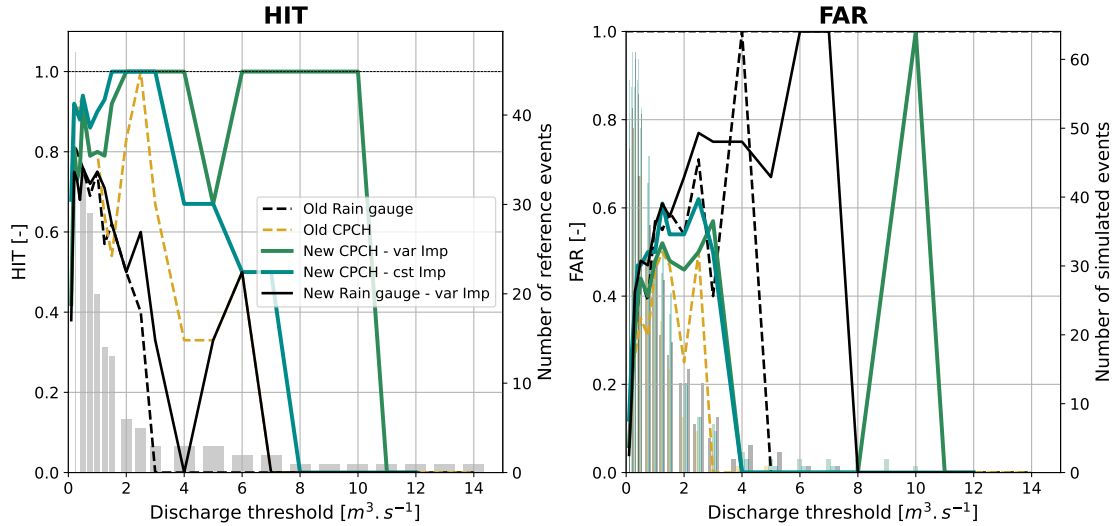Figure 5.24: Validation HIT/FAR scores for the discharge from the three new models and the two best old models in Zurich. HIT and FAR are computed by events with a tolerance of 10% and a temporal tolerance of 20 minutes. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

Table 5.5: Validation Nash scores of Zurich-Altbach new models. The best scores are coloured in green. Volume ratios are also provided.

Model	Nash for $Q >$				Volume ratio
	0	$0.5\,\mathrm{m^3\,s^{-1}}$	$1\,\mathrm{m^3\,s^{-1}}$	$1.5\,\mathrm{m^3\,s^{-1}}$	
Old rain gauges	0.44	0.03	-0.16	-0.31	1.19
Old CPCH	0.62	0.42	0.34	0.25	1.19
New CPCH - var Imp	0.56	0.35	0.36	0.34	1.18
New CPCH - cst Imp	0.58	0.35	0.35	0.28	1.17
New rain gauges - var Imp	0.33	-0.15	-0.33	-0.4	1.19

Despite their slight FAR increase, the new models are clearly better than the previous one. **The differentiation of the upper and lower sub-basins and the introduction of rivers to delay the response have proven to be important features to better model the sharp events**. It needs to be noted that there is nearly no differences between these models and the old one for the other longer events. **The addition of a variation of the impermeable area fraction Imp appears not to be determinant in the scores of the models but helps to better catch the shape of flash flood events and is probably helping to better model the physical processes** occurring in the Zurich-Altbach catchment. For this reason, the model where Imp varies with the precipitation intensity will be mostly used in hindcast – though a hindcast with INCA and the model with constant Imp will also be tested.

5.3.3.2 Hindcast: Perfect forecasts with the new model

Perfect forecast simulations are first run with the new model with variable *Imp*. An *update* step is performed before with separated update of the state variables in the upper and lower sub-basins. Results are provided for six events in Figure 5.25. The list of events has been slightly modified to include the sharp event of the 1stJuly 2019 and to remove most of the slow events which show no differences between models. Only the main rain-on-snow event from January 2021 is kept. All perfect forecasts appear to be improved compared to Section 5.3.2.1, notably concerning the shape and the timing of the peaks. The rain gauge perfect forecast is usually providing high false alarms which is explained by the already high values obtained with the old model: the new model has been mostly adapted for CPCH and INCA. In this model, spatial information provided by the radar has been used to create a double peak in the discharge. However such spatial information is not available with the only two rain gauges used in Zurich (Kloten and Winterthur), which are outside the catchment. Hence the two peaks from the two sub-catchments tend to occur simultaneously leading to a peak overestimation.

(a) 13/09/2018

(b) 01-02/07/2019

(c) 28-30/01/2021

(d) 18-19/06/2021

(e) 23/06/2021

(f) 28/06/2021
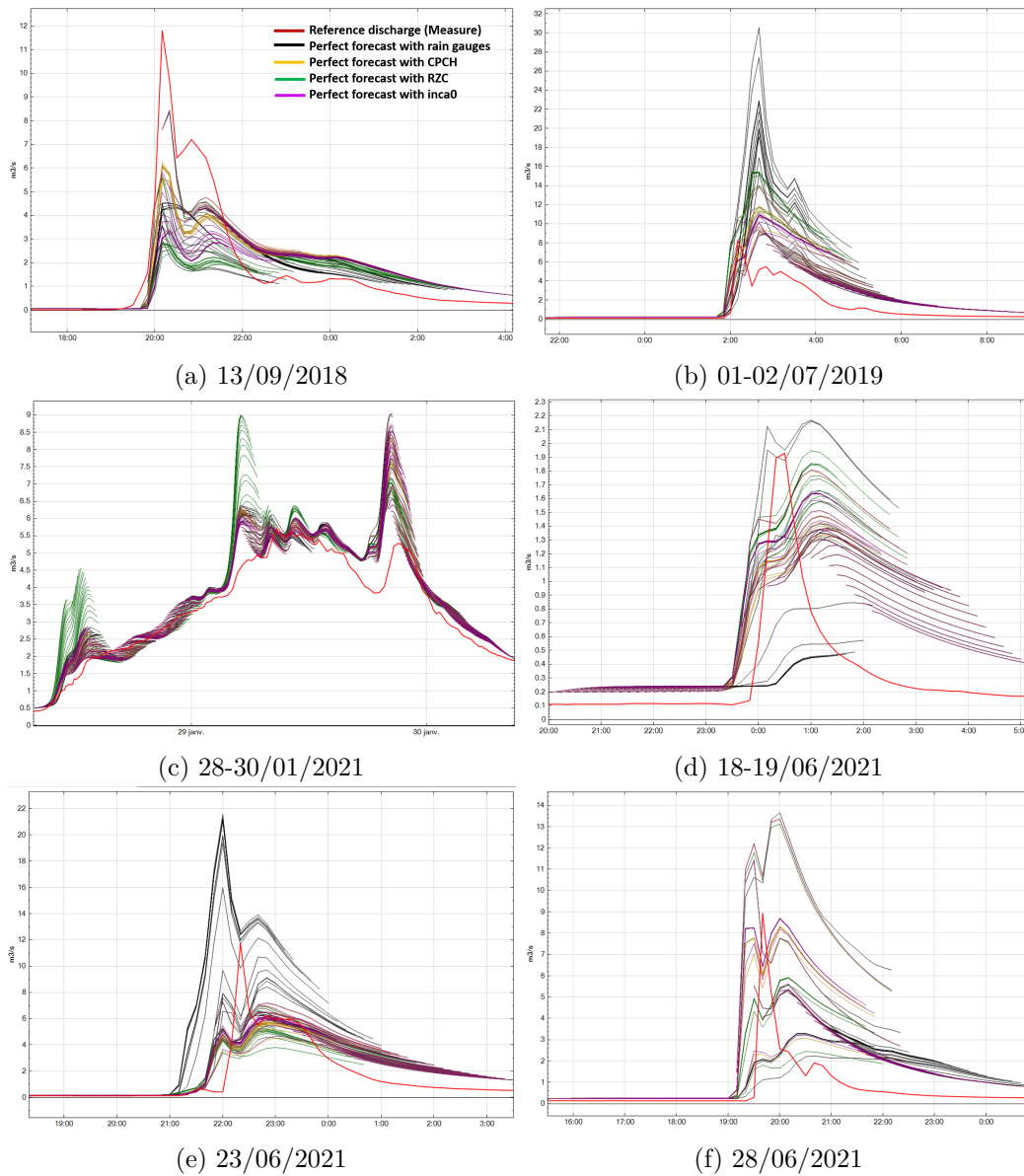
Figure 5.25: New perfect forecasts for six events in Zurich-Altbach discharge station: CPCH (golden), RZC (green), stations (black) and inca0 (purple). The measure is in red.
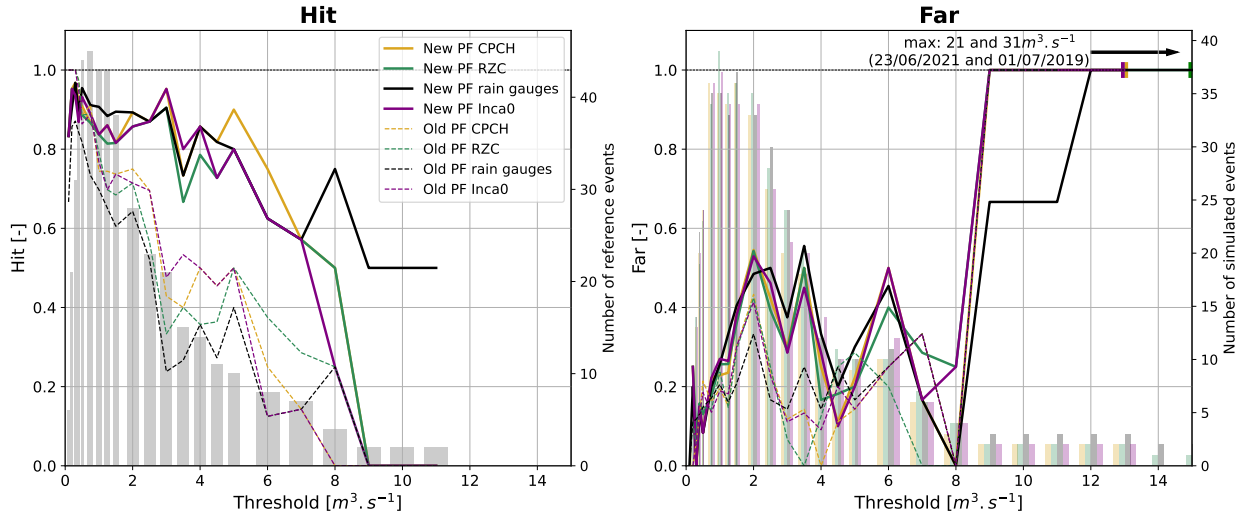
Figure 5.26: Alert HIT and FAR scores of the new perfect forecasts: CPCH (golden), RZC (green), rain gauge stations (black) and inca0 (purple). No tolerance is applied. The old perfect forecasts are represented with dashed lines.

Figure 5.26 provides the alert HIT and FAR for the four perfect forecasts. The perfect forecasts with the old model are also displayed with dashed lines. **HIT scores are largely improved for all models.** As with the previous model, performances are however similar between the four perfect forecasts. The drawback for these increased HIT is a slight increase in FAR. However, the increase in HIT is largely better than the FAR worsening. **The new model is hence much more adapted for an alert system.** The bigger problem resides in the few false alarms at very high discharges. However all these alarms occur at lead times equal to 70, 80 or 90 minutes. **It is probable that alarms will need to be observed in consecutive forecasts (*lagged forecasts*[207]) or at a smaller lead time before launching an evacuation order.**

5.3.3.3 Hindcast: Old model *vs* new models

The two models (with/without variable *Imp*) are now run with the INCA QPF and compared to the old INCA forecast. As for the perfect forecasts, an *update* step is performed before, but here state variables are modified similarly in the upper and lower sub-basins. Figure 5.27 displays the discharge forecasts for the selection of six events. **For all these events, the forecasts have usually been improved.** However, as observed in the continuous simulations and in the perfect forecasts, the model increases the number and intensity of false alarms. This is particularly the case for the false alarms of the 01/07/2019, 18/06/2021 and 28/06/2021 events. False alarms are also, logically, larger for the model with variable impermeable area than for the one without it. Indeed, if INCA provides a false alarm – due to a wrong estimation of the position, timing or intensity of storm cells – the impermeable area will be wrongly increased which will multiply the intensity of the false alarm: **while the model with variable impermeable is better modelling the processes occurring in the catchment, it is more sensitive to precipitation data and can generate bigger false alarms in case of wrong precipitation forecast.**
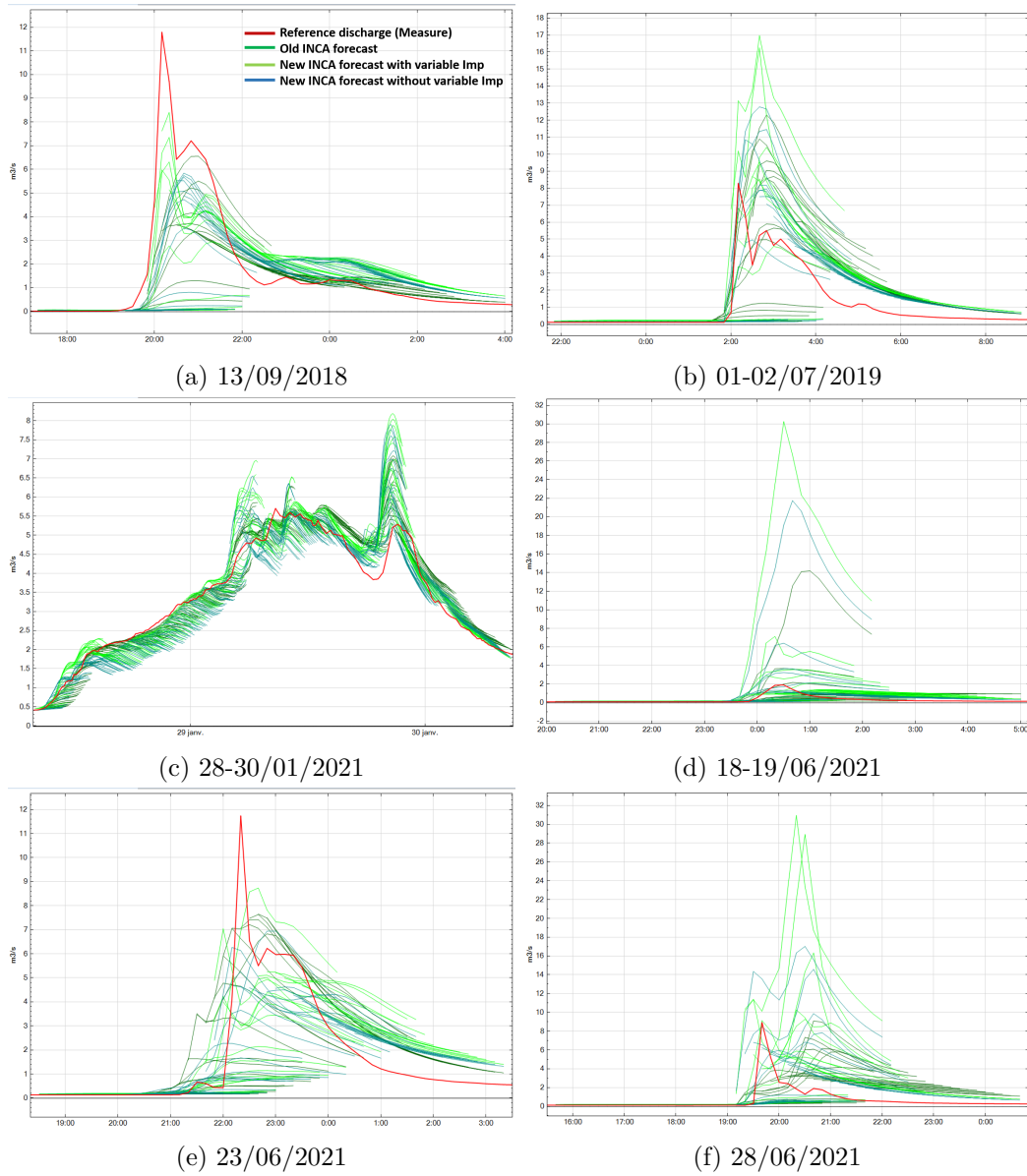
Figure 5.27: Old (darkgreen) and new, with or without variable impermeable area (resp. lime or cyan), INCA forecasts for six events in Zurich-Altbach discharge station. The measure is in red.

Both observations are confirmed by the HIT-FAR alert scores from Figure 5.28. Notably, alert HIT scores are largely improved with the new modifications of the model and notably with the variable impermeable. As expected, the drawback is an increase of false alarms, very reasonable at low intensity. Two gigantic false alarms are generated for the 18 and 28 June 2021 events. However these false alarms are generated 80 or 90 minutes before the event. At smaller lead time, the false alarm rate will probably largely be reduced[c].

---

[c]The construction of a better alert system will be discussed in Section 6
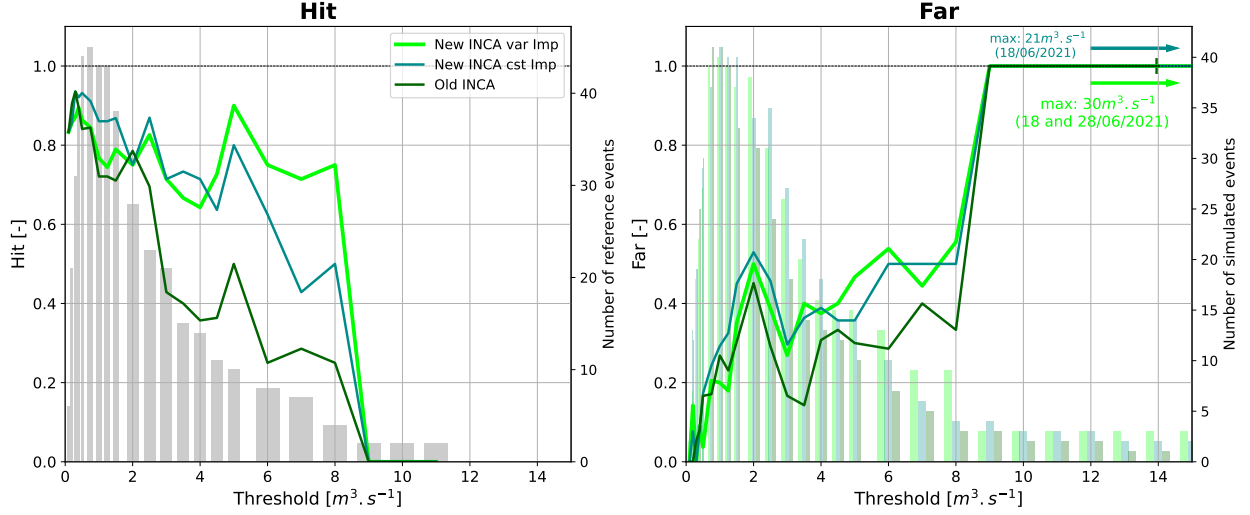
Figure 5.28: Alert HIT and FAR scores of the old (darkgreen) and the new INCA forecasts. The new forecast with constant *Imp* is in cyan while the one with variable *Imp* is in lime. No tolerance is applied.

5.3.3.4 Hindcast: Influence of the *update* step

The impact of the *update* step can also be evaluated. Indeed it has been found to be essential in the first report of the *Radar4Infra* project[221], notably for long events such as those occurring in Emosson or the rain-on-snow events of Zurich. However, for sharper flash flood events, the *update* step may worsen the forecasts. To evaluate the importance of an *update* step the model with variable *Imp* is run with the INCA forecast with or without *update* step. Two kinds of *update* are tried: the update of the initial state variables can be performed either with constant correction factors over the catchment (*update 1 zone* - as in Section 5.3.3.3) or with distinct correction factors for the upstream and downstream sub-catchments (*update 2 zones*, as for the perfect forecasts in Section 5.3.3.2).
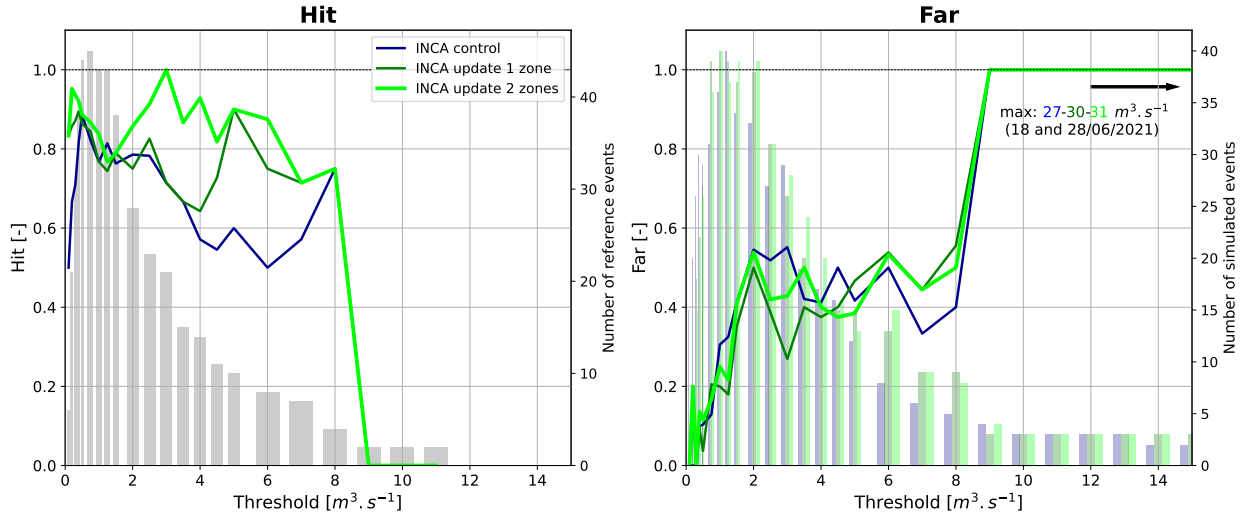
Figure 5.29: Alert HIT and FAR scores of the new INCA forecast with an *update* step similar on both zones (darkgreen), different in both zones (lime) or without *update* step (blue). No tolerance is applied.

(a) 13/09/2018

(b) 01-02/07/2019

(c) 28-30/01/2021

(d) 18-19/06/2021

(e) 23/06/2021

(f) 28/06/2021
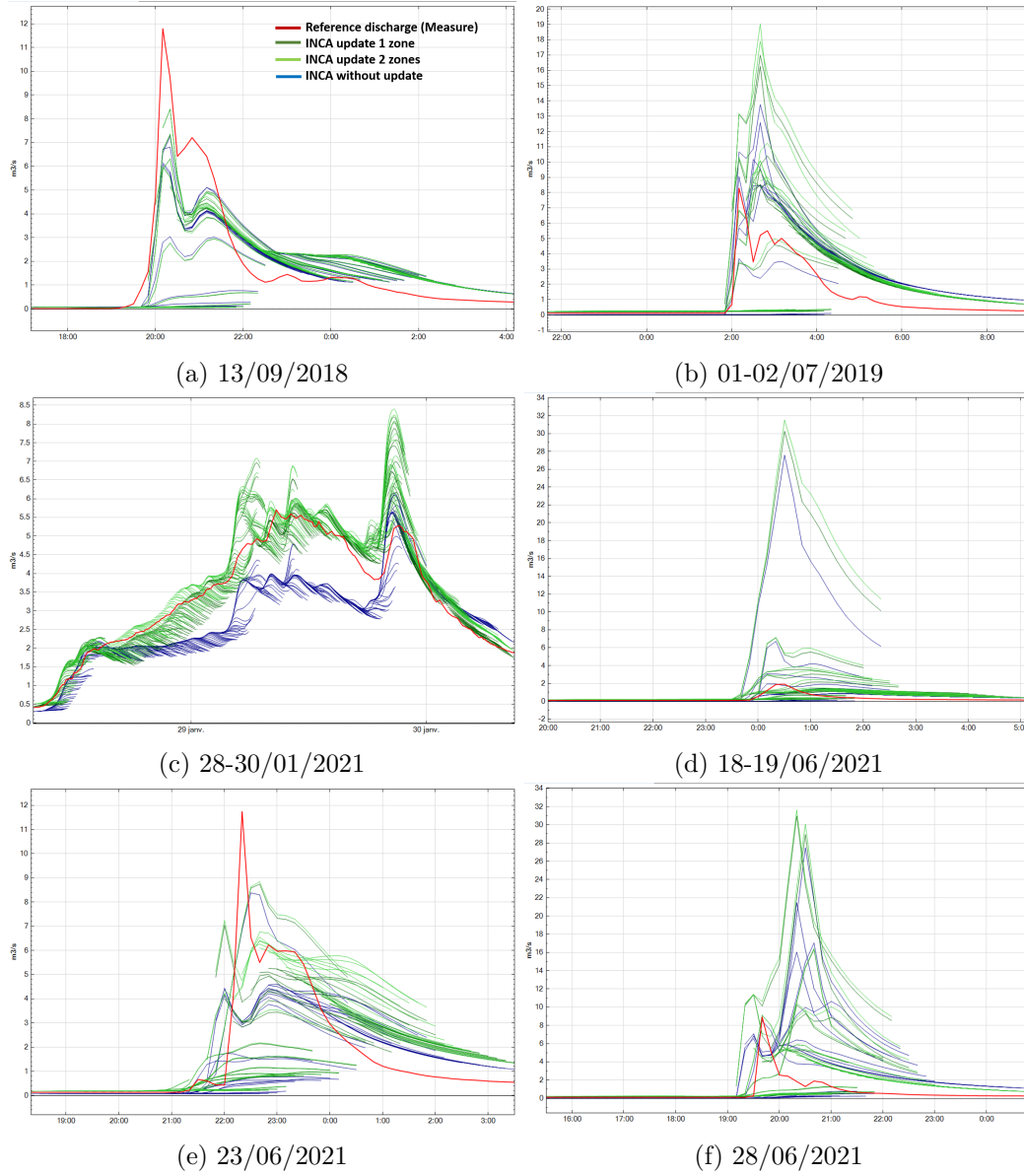
Figure 5.30: New INCA forecast with variable impermeable area for six events in Zurich-Altbach discharge station. The blue curve has no *update* while a similar (resp. a different) *update* is performed on the two sub-basins in the darkgreen (resp. lime) curve. The measure is in red.

Figure 5.30 provides the results for the six events while HIT-FAR scores for alert are presented in Figure 5.29. For each of the sharp events, the *update* step does not worsen the prediction. It usually does not affect significantly the discharge but when it does so, it generates higher and earlier runoff increases. This is confirmed in the alert HIT-FAR where HIT scores are increased compared to the control forecast (with no update) and where the maximum discharge is increased in FAR. Differences between the two *update* techniques are small but the one with two independent zones appear to be slightly better in HIT and in the events. Moreover, concerning the longer events, as observed in the first report of the *Radar4Infra* project[221], the *update* step is essential. The rain-on-snow event is completely missed without *update*. **Hence the *update* step does not appear to be worsening the sharp flash flood events while it is essential for slower events**. It will be used with independent corrections in the downstream and upstream basins until now[d].

---

[d]and it has already been used in the two previous subsections.

5.3.3.5 Hindcast: INCA *vs* COSMO

In this last section, the decrease or increase of performances due to the input QPF is evaluated. The CPCH perfect forecast, the INCA forecast and the COSMO forecast are run with the new model with variable *Imp* (after an *update* step). Events are displayed in Figure 5.32 while alert HIT-FAR scores are provided in Figure 5.31. The INCA QPF introduces a reactivity that is not present in COSMO. Hence, although most of the events are observed both in COSMO and INCA, the predictions is quicker and more reactive with INCA. However, it implies also higher false alarms, notably the 01/07/2019, the 18/06/2021 and the 28/06/2021. These observations are confirmed by the alert HIT-FAR scores.

Concerning the expected loss of performances between the perfect CPCH forecast and the INCA forecast, it is not so significant. It is mostly observed with the higher false alarms and the dispersion of forecasts in INCA. Nevertheless, INCA is even sometimes able to catch some events better and with more reactivity than CPCH. This is confirmed by the good alert HIT scores of the INCA forecast.

The new model, notably with the introduction of a variable impermeable area and with the use of an *update* step, when used with the INCA QPF allows to largely increase the hit performances for alerts. However, it also generates tremendous false alarms that will need to be removed. Two ideas can be developed to do so:

- First, it could be useful to use ensembles to introduce and propagate the uncertainty of the INCA QPF into the rainfall-runoff model. Several scenarii will then be available and a new alert systems can be used to make the most of these scenarii. Ensembles can be generated by translation in the time or space, by modifications of the discharge value (random noise, modifications of peak levels) or by the use of several distinct rainfall-runoff models such as the one developed here. *MeteoSwiss* will also continue to develop the INCA product with an ensemblist approach: INCA ensemble members will be interesting in such approach.

- Secondly, without ensembles, it is also possible to modify the way alerts are generated. The rules developed to compute alert HIT-FAR scores can be modified by using smaller lead times for example. This approach will now be developed.
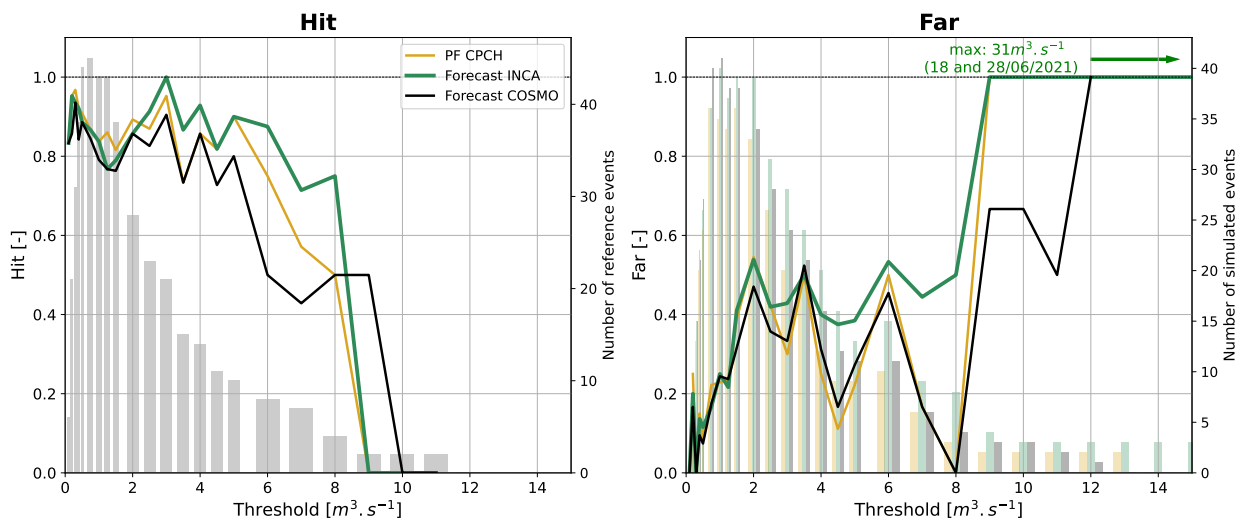
Figure 5.31: Alert HIT and FAR scores of the new perfect CPCH (golden), INCA (green) and COSMO (black) forecasts.

(a) 13/09/2018

(b) 01-02/07/2019

(c) 28-30/01/2021

(d) 18-19/06/2021

(e) 23/06/2021

(f) 28/06/2021
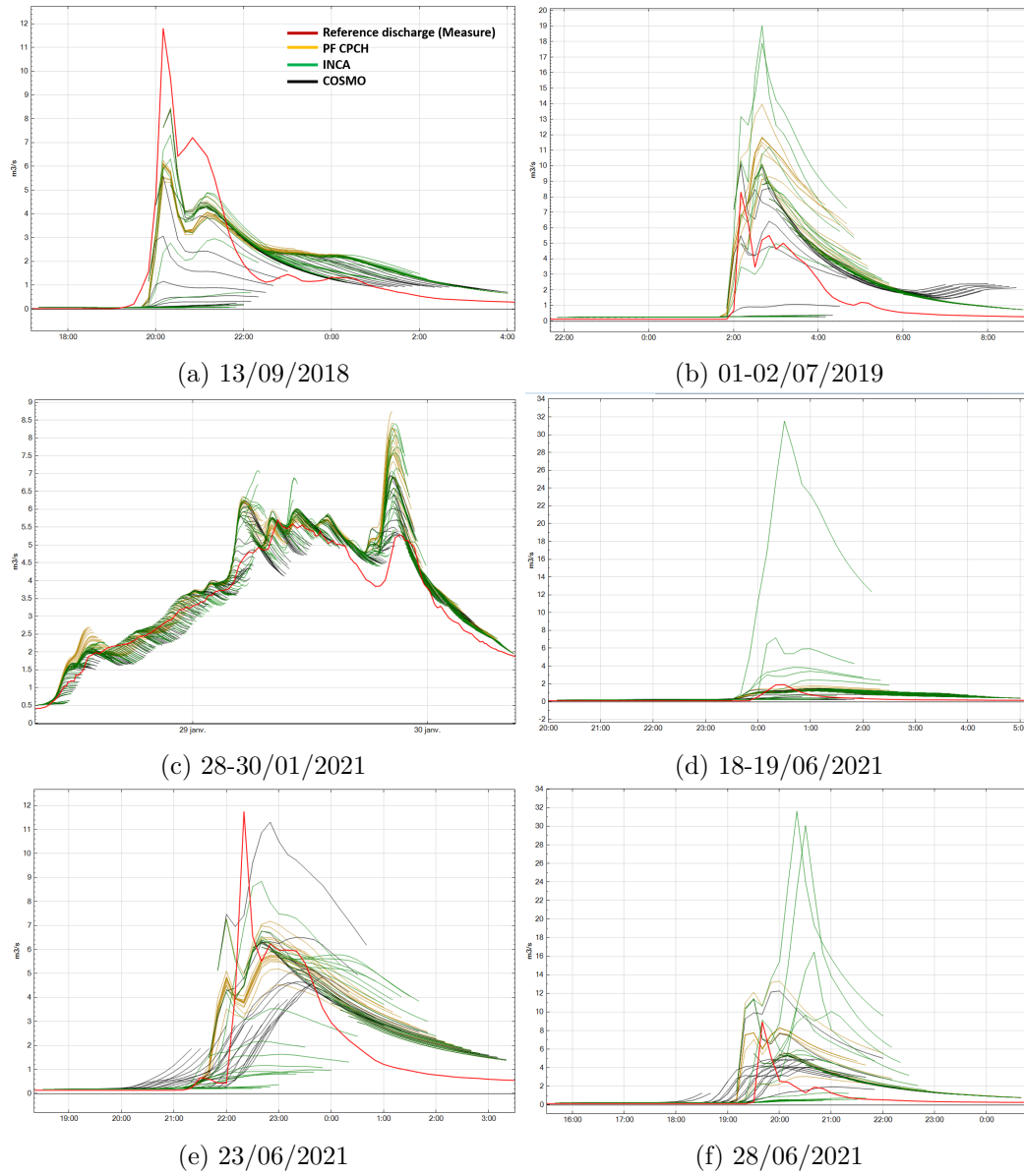
Figure 5.32: Hindcast with variable impermeable area for six events in Zurich-Altbach discharge station: perfect *CombiPrecip* forecast (golden), INCA forecast (green) and COSMO forecast (black). The measure is in red.

5.3.4 Conclusion

The Zurich-Altbach catchment is characterized by two distinct types of response to rainfall. In some cases the discharge response is slow and steady, resulting in events relatively harmless and easy to predict. However, the Altbach river also endures flash floods where the discharge increases in one or two timesteps only, which can cause severe damages, notably for the Kloten airport downstream. To improve the predictions of these flash flood events, the Zurich-Altbach RS model was improved in two steps.

In a first step, the operational model was re-calibrated by focusing more on sharp events. The results highlight the interest of using the INCA radar QPF in the rainfall-runoff model compared to using the COSMO NWP. **The INCA forecast is notably much more reactive to flash floods than COSMO**. However, the better prediction performances of INCA are also linked to more false alarms due to wrong predictions of the location, timing and intensity of storm events. Finally, a post-processing based on the initial slope of the observed discharge appears to be detrimental and to only lead to huge false alarms while erasing all differences between COSMO and INCA.

In a second step, by observing that the continuous simulations still had difficulties to simulate sharp events, a new model was set up. It focuses mostly on flash flood events and is inspired by the literature. It first allows to model the double peak shape observed in most sharp events and which is probably due to the existence of two sub-catchments and to the usual meteorological evolution of convective cells from downstream to upstream (West to East). Modelling this double peak shape allows to get closer to the observed discharge. Moreover, to better catch the most intense flash flood events, a variation of the infiltration rate at high intensity is introduced *via* the impermeable fraction parameter *Imp*. The new model demonstrates improved performances both in continuous simulations and in operational re-forecast. **The hit performances of an alert system based on this model are now acceptable. However some very high false alarms are also issued** due to the combination of a highly responsive model and of an uncertain INCA QPF. To remove these false alarms it is possible to modify the rules for issuing alerts (Section 6) or to introduce and evaluate the uncertainty of the model, for example by using ensemble predictions.

A summary of all the tests performed in the Emosson and Zurich catchments, and of the main conclusions, is provided next page in Figure 5.33
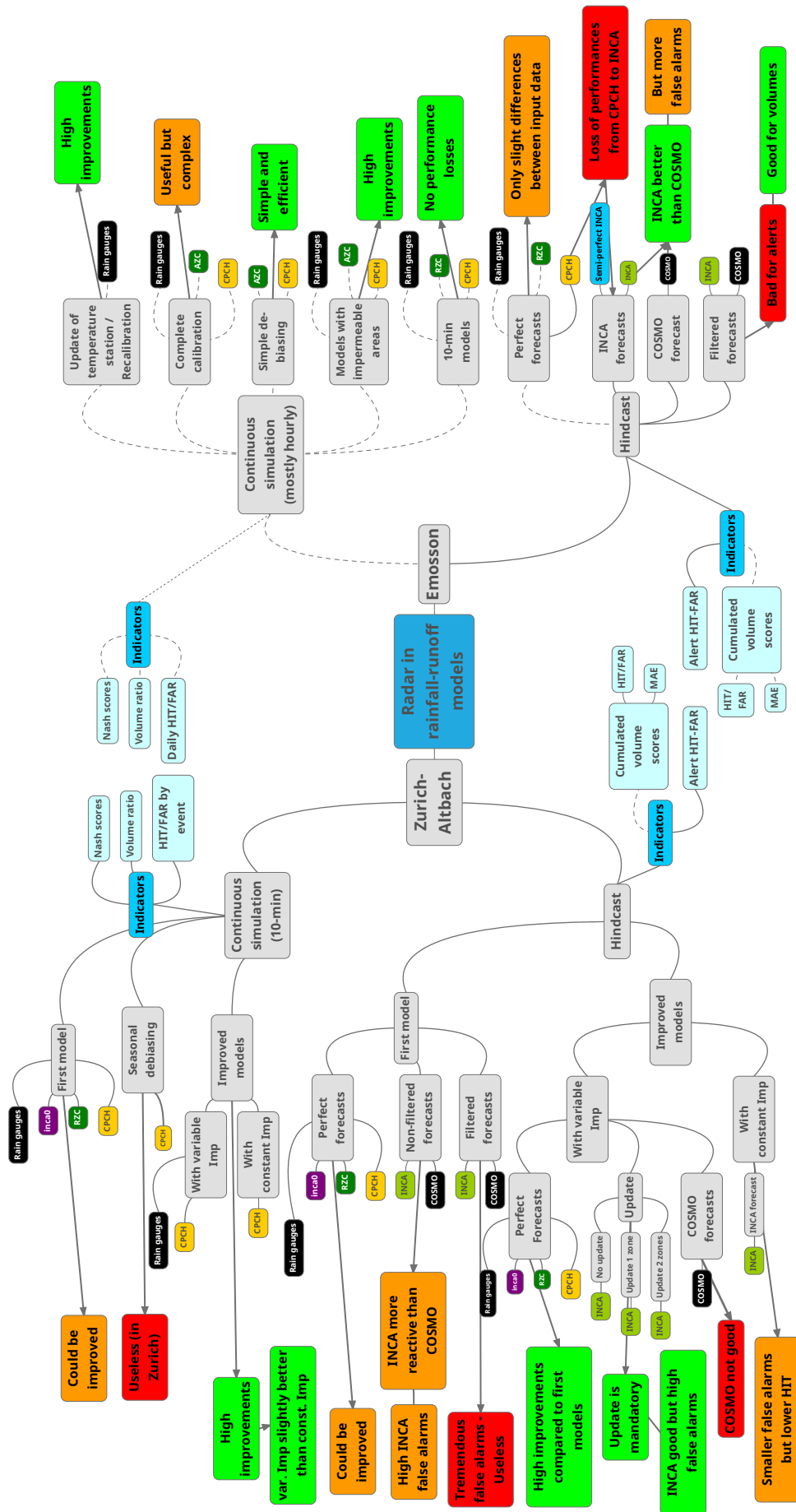
Figure 5.33: Overview of continuous and hindcast models developed and of the main conclusions. Dashed lines represent results detailed in Appendices D and E.

Chapter 6

Design of an alert system

6.1 Introduction

The careful design of an alert system is necessary to tap the full potential of a rainfall-runoff prediction. Indeed, as mentioned by the WMO[6], "*there is a crucial balance to be struck between a precautionary approach and an unwillingness to issue warnings for fear of these being "wrong"*". False alarms, which are inevitable, must be minimized so that flood warnings are not triggered too frequently.

The goal of the present section is to underscore the importance of investing time in the design of alert systems while proposing and investigating some approaches. To do so, the Zurich airport project is chosen as case study. The issue with the current alert system (from Figure 5.28), is the few false alarms observed at very high intensities – notably the 01/07/2019, 18/06/2021 and 28/06/2021 with respectively 19.0, 31.5 and 31.6 $\mathrm{m^3\,s^{-1}}$ predicted. As the system aims at issuing alerts for very high discharges (higher than 20 $\mathrm{m^3\,s^{-1}}$), they must be reduced or removed, without decreasing HIT performances.

In this section, the two best rainfall-runoff models of the Zurich catchment will be used (see Section 5.3.3 for details):

- The INCA model using a variable Imp parameter. Hereafter, it will be called **model V** (as variable) for clarity.

- The INCA model using a constant Imp parameter, which will be called **model C** (as constant).

Several strategies will be investigated:

- Section 6.2: first, we will try to **reduce the maximum lead time** where the INCA forecasts are considered. Alert system based on one of the two rainfall-runoff models will be tested.

- Section 6.3: then, the introduction of **lagged forecasts**[207] will be investigated to reduce false alarms. Several systems with lagged forecasts will be tested. Only model V will be used.

- Section 6.4: finally, the two rainfall-runoff models V and C will be **combined to develop multi-model alert systems**. This option can be coupled with the use of lagged forecasts. It is a first incursion in the world of ensembles with a simple example of EPS-SDM (several models for one deterministic QPF, see Section 2.5).

6.2 Decrease of the lead time

To remove the high false alarms, considering they are usually predicted at lead times between +60 and +90 minutes, a first idea is to reduce the lead time of the forecast. Until now, alarms were issued if events were observed in the first 90 minutes of a forecast. In this section, this time will be reduced to 80 minutes, 70 minutes, and so on. As before, if an alert is provided with less than 20 minutes of anticipation (compared to the observed event) it is considered as missed because it does not give enough time for authorities to act. Furthermore, an event is still not considered a false alarm if a real event happens before the predicted event or within the first 90 minutes following it.

6.2.1 Model V

HIT-FAR scores are provided in Figure 6.1 for several maximum lead times, with a single-model alert system based on model V. In some cases, a curve may be hidden by others if they have identical values. As curves are drawn by ascending lead times, the above curve is always the one with larger lead time. For example the HIT curves at 70 and 80 minutes are hidden by the curve at 90 minutes. Fortunately, the more restrictive is an alert system the lower are its HIT scores. This is also the case in general for FAR scores[a]. To observe the level of the high false alarms, Table 6.1 provides the maximum discharge predicted for the three high false alarm events – the three events predicted above $15\,\mathrm{m}^3\,\mathrm{s}^{-1}$, *i.e.* the 01/07/2019, 18/06/2021 and 28/06/2021 events.

When the maximum lead time is decreased, less false alarms are produced. The FAR number is reduced only slightly but steadily until 50 minutes, before dropping significantly at 40, 30 and 20 minutes. HIT scores are not affected at 80 and 70 minutes, are slightly dropping at 60, 50 or 40 minutes before being largely affected by smaller lead times. Hence, in order to keep good detection performances, lead times strictly below 40 minutes must be avoided, while lead times of 80 or 90 minutes are useless. As observed in Table 6.1, maximum lead times of 60 or 70 minutes are also still producing false alarms above $20\,\mathrm{m}^3\,\mathrm{s}^{-1}$. Thus, for this simple single-model alert system, **a lead time of 40 or 50 minutes is recommended.** It will however result in small action time available to public authorities. Better systems are needed.

Figure 6.1: Alert HIT-FAR scores for the single alert system based on model V with different maximum lead times ranging from 20min to 90min. All forecasts between 20min and the maximum lead time are considered to generate alerts.

---

[a]As the FAR is the ratio between false alarms and alarms, if a correct alarm is removed the FAR score can be increased even if less alarms are produced.

Table 6.1: Maximum predicted discharge – in $m^3\,s^{-1}$ – of the three high false alarm events for several maximum lead times.

Maximum lead time	01/07/2019	18/06/2021	28/06/2021
20 minutes	13.2	1.3	11.4
30 minutes	15.8	5.1	10.6
40 minutes	19.0	11.4	13.5
50 minutes	19.0	16.8	15.5
60 minutes	19.0	24.7	23.9
70 minutes	19.0	31.5	31.6
80 minutes	19.0	31.5	31.6
90 minutes	19.0	31.5	31.6
Observed discharge	8.0	2.0	9.0

6.2.2 Model C

Similarly, Figure 6.2 and Table 6.2 provide respectively the HIT-FAR scores and the maximum discharge alert generated for several maximum lead times based on model C. The observations and conclusions from model V are still valid here. However, compared to model V, the FAR scores are not really better while the HIT scores have largely been reduced. The interest of this model lies in the strong reduction of the three high false alarms. Indeed, the maximum discharges are reduced by approximately 50% with model C. Hence a lead time of 60 minutes appears now to be suitable with this model. However, as the loss of detection performances in model C is quite important, **it seems preferable to work first with model V**.
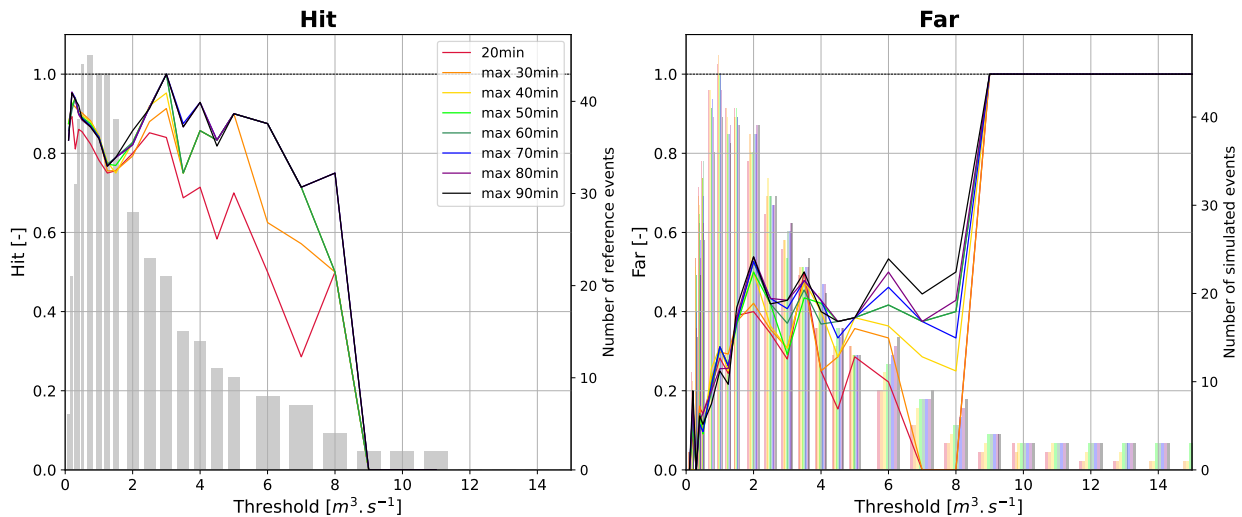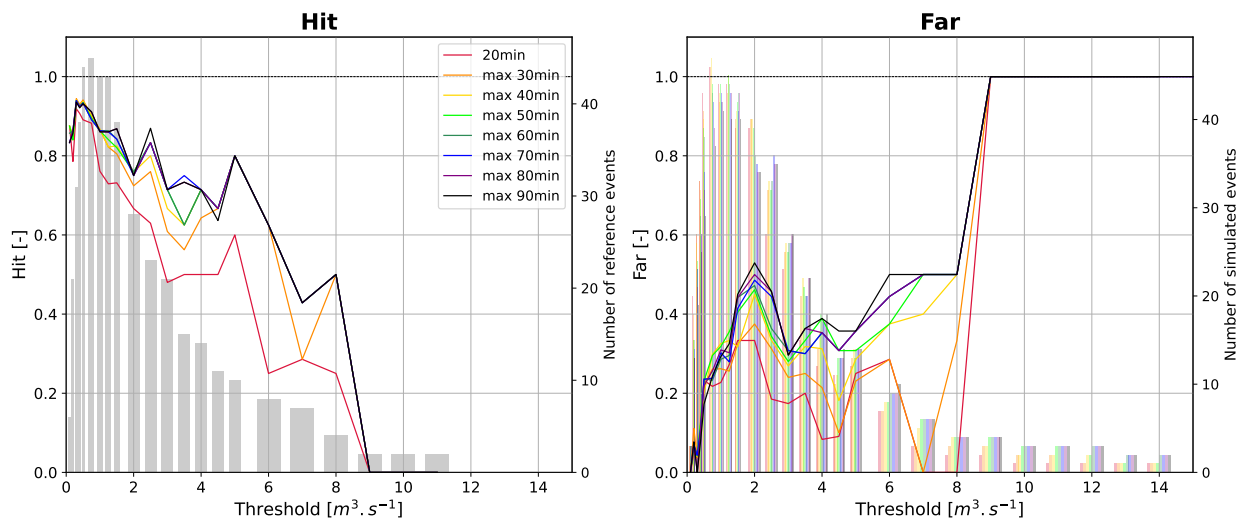

Figure 6.2: Alert HIT-FAR scores for the single alert system based on model C with different maximum lead times ranging from 20min to 90min. All forecasts between 20min and the maximum lead time are considered to generate alerts.

Table 6.2: Maximum predicted discharge – in $\mathrm{m^3\,s^{-1}}$ – of the three high false alarm events for several maximum lead times.

Maximum lead time	01/07/2019	18/06/2021	28/06/2021
20 minutes	9.9	1.4	14.2
30 minutes	11.4	3.5	14.2
40 minutes	12.4	8.4	14.2
50 minutes	12.8	11.1	14.2
60 minutes	12.8	14.2	14.2
70 minutes	12.8	19.2	16.3
80 minutes	12.8	21.8	17.0
90 minutes	12.8	21.8	17.0
Observed discharge	8.0	2.0	9.0

6.3   Introduction of redundancy: lagged forecasts

Following the literature review – Bartholmes *et al.*[208], Alfieri *et al.*[207] – the introduction of redundancy in the alert system may help to reduce false alarms. In this section, alert systems with lagged forecasts are designed based on model V. Several alert triggering rules are investigated.

6.3.1   Consecutive lagged forecasts

As a first step, two consecutive threshold crossings can be required to issue an alert. It means, at each time, the smaller maxima of the last two forecasts is taken as the maximum level expected. The rules to determine hit and far have otherwise not been modified. The lagged forecast (the second one) must be issued at least 20 minutes before the real event occurs. HIT-FAR scores are provided in Figure 6.3. The value of the three potential high false alarms are provided in Table 6.3.

The observations from Figure 6.1 are still present with nearly no HIT differences from 50-60 minutes to 90 minutes, and with a large decrease of HIT at lower lead times. However, FAR scores are now smaller and less dependent of the lead time, allowing to use a longer lead time. While FAR scores at high lead times have been improved compared to Figure 6.1, the drawback is a significant reduction of HIT scores. Lagged forecasts appear to be an interesting approach, as the major false alarm from the 18$^{\text{th}}$June 2021 has been removed – unlike the other two false alarms, this one is not linked to any significant event in the measure. However, **a warning system with less restrictions is needed to maintain better HIT performances**. The minimum anticipation time required for authorities (20 minutes in this study) may explain part of the loss of performances, as some events are now predicted too late to be considered hit.
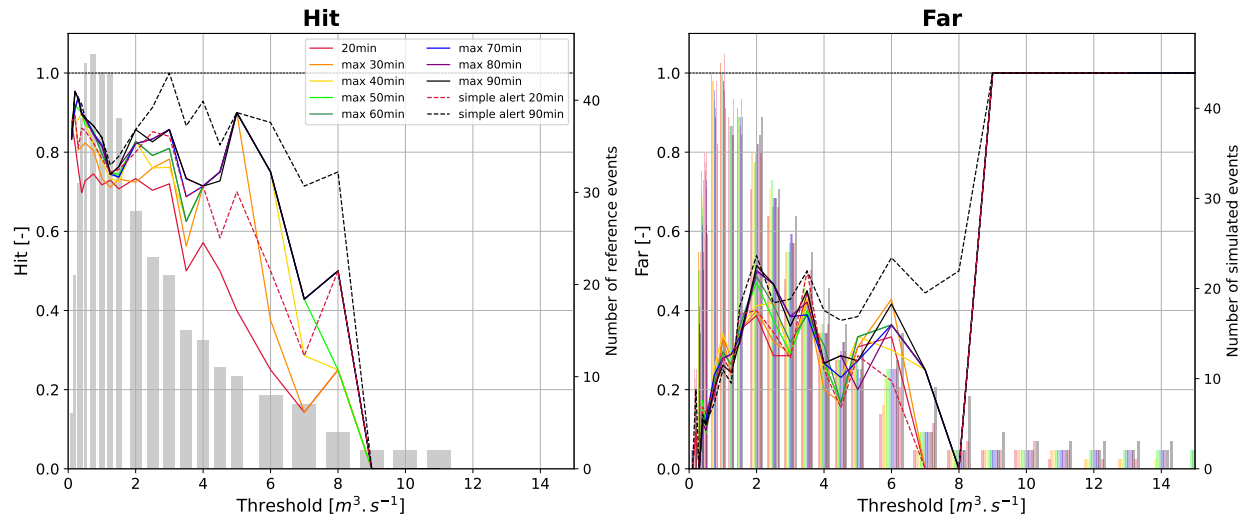
Figure 6.3: Alert HIT-FAR scores using lagged forecasts for several maximum lead times. The lead time in the legend corresponds to the lead time of the second forecast (when the alert is finally issued). The two dashed curves correspond to the extremal curves from Figure 6.1.

Table 6.3: Maximum predicted discharge – in $m^3 s^{-1}$ – of the three high false alarm events for several maximum lead times, using lagged forecasts.

Maximum lead time	01/07/2019	18/06/2021	28/06/2021
20 minutes	10.3	1.1	11.4
30 minutes	13.2	2.1	11.4
40 minutes	14.5	3.0	11.4
50 minutes	17.9	3.0	15.1
60 minutes	17.9	3.3	23.1
70 minutes	17.9	3.5	30.1
80 minutes	17.9	3.7	30.1
90 minutes	17.9	3.9	30.1
Observed discharge	8.0	2.0	9.0

6.3.2 Lagged forecasts without the consecutiveness rule

In the previous section, lagged forecasts have been required to be consecutive. However it is possible to be less restrictive and issue an alert if at least two forecasts, consecutive or not, display a threshold crossing. The two forecasts must be issued within a reasonable time (chosen here to be equal to the maximum lead time). The last forecast must still be issued at least 20 minutes before the real event occurrence. HIT-FAR scores for this case are provided in Figure 6.4. The value of the three potential high false alarms are provided in Table 6.4.

Compared to the previous section with consecutive lagged forecasts, the model presented here have slightly better HIT scores at high discharges (6-7 $\mathrm{m^3\,s^{-1}}$). FAR are also slightly increased but mostly for the 80 and 90 minutes lead time curves. Concerning the high false alarm events, the 18/06/2021 event is still largely reduced as it does not go above $7.2\,\mathrm{m^3\,s^{-1}}$. **Differences with consecutive lagged forecasts are thus small but can be important to reduce the anticipation time of an alert**. Such time is essential to allow public authorities to act.
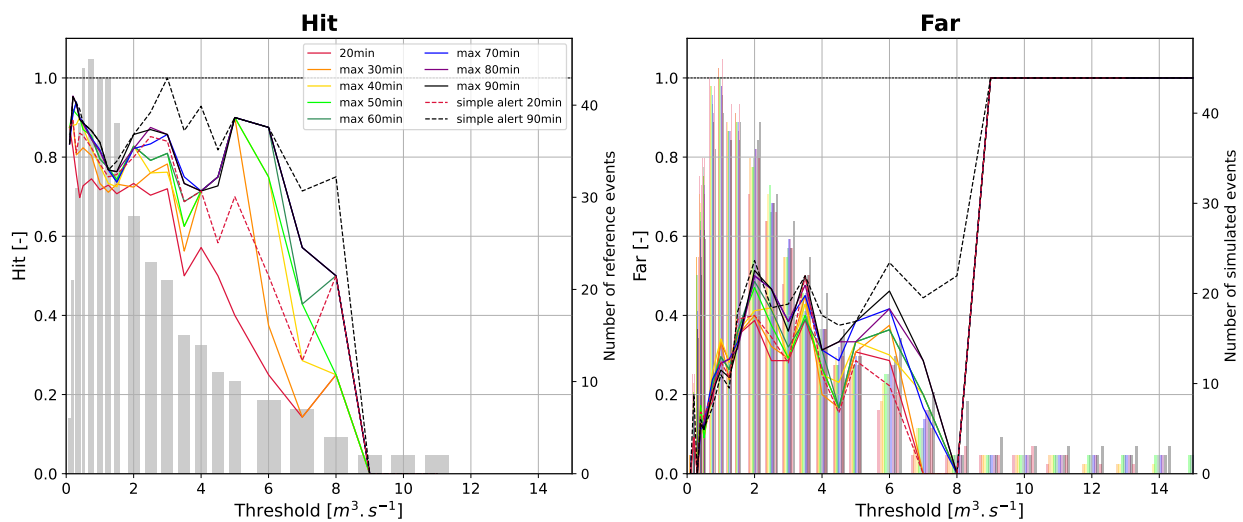
Figure 6.4: Alert HIT-FAR scores using lagged forecasts – possibly non-consecutive – for several maximum lead times. The maximum lead time corresponds to the maximum lead time of the second forecast.

Table 6.4: Maximum predicted discharge – in $\mathrm{m^3\,s^{-1}}$ – of the three high false alarm events for several maximum lead times, using lagged forecasts – possibly non-consecutive.

Maximum lead time	01/07/2019	18/06/2021	28/06/2021
20 minutes	10.3	1.1	11.4
30 minutes	13.2	2.1	11.4
40 minutes	14.5	3.0	11.4
50 minutes	17.9	3.0	15.1
60 minutes	17.9	3.3	23.1
70 minutes	17.9	6.5	30.1
80 minutes	17.9	7.2	30.1
90 minutes	17.9	7.2	30.1
Observed discharge	8.0	2.0	9.0

6.3.3   Mix between lagged forecasts and normal forecasts

A solution to reduce the time of action lost with lagged forecasts is to mix lagged and normal forecasts depending on the predicted time before the event. For example, if an event is predicted in 40 minutes or before, an alert will directly be issued. However, if the event is predicted to occur in 50 minutes or more, the alert will need to be confirmed in another forecast before launching an alert. The cut-off anticipation time can thus be another variable of the alert system. As alert systems with lagged forecasts at 90 minutes have shown an interesting false alarms reduction at high discharges, all the next alert systems will have a maximum lead time of 90 minutes. To increase HIT without increasing FAR, lagged forecasts will be introduced only strictly above the cut-off lead time (40 minutes in the previous example). Different cut-off are tested in Figure 6.5 and Table 6.5.

All scores are located between those of the system using only lagged forecasts (dashed red) and those of the one using only direct alerts (dashed black). HIT scores are not increased when using a cut-off time of 20 minutes. Hence, the cut-off must be at least equal to 30 minutes. Moreover, FAR scores are only really dropping for cut-off time below 40 minutes. Notably, for a cut-off of 30 minutes, all alerts issued above $8\,\mathrm{m^3\,s^{-1}}$ correspond to events effectively occurring. This cut-off corresponds also to the moment when the 18/06/2021 event is reduced below $10\,\mathrm{m^3\,s^{-1}}$. **A mixed system based on direct alerts until 30 minutes and lagged alerts above 40 minutes appears to be the best system up to now** with nearly the best HIT scores and a reduced number of false alarms.
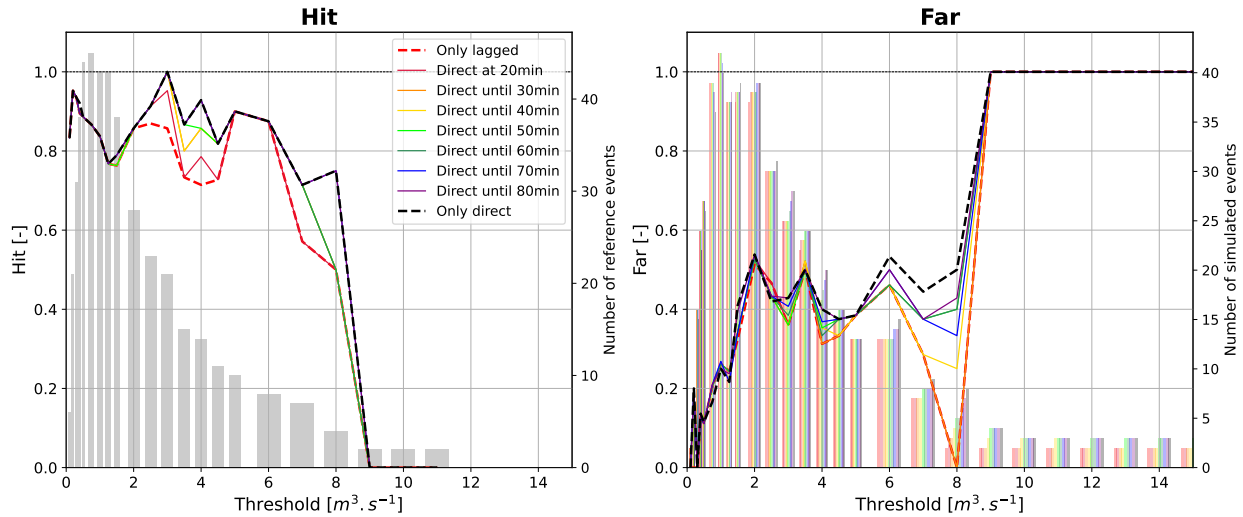
Figure 6.5: Alert HIT-FAR scores using lagged forecasts (possibly non-consecutive). If an event is predicted before the cut-off in the legend, an alert is issued without waiting for a lagged forecast.

Table 6.5: Maximum predicted discharge – in $\mathrm{m^3\,s^{-1}}$ – of the three high false alarm events for several cutoff lead times, using lagged forecasts for higher prediction times.

Direct alert until...	01/07/2019	18/06/2021	28/06/2021
Only lagged forecasts	17.9	7.2	30.1
20 minutes	17.9	7.2	30.1
30 minutes	17.9	7.2	30.1
40 minutes	19.0	11.4	30.1
50 minutes	19.0	16.8	30.1
60 minutes	19.0	24.7	30.1
70 minutes	19.0	31.5	31.6
80 minutes	19.0	31.5	31.6
Only direct alerts	19.0	31.5	31.6
Observed discharge	8.0	2.0	9.0

6.3.4 Mix between double lagged forecasts and normal forecasts

Similarly, when the time before the prediction is large enough, the alert can be confirmed not once but also twice to further reduce false alarms. Indeed an event observed in three forecasts – possibly non-consecutive – is usually not a false alarm. The different alert systems developed until now with a 90-minute lead time are displayed together in Figure 6.6 and Table 6.6. The alert systems from Sections 6.2.1, 6.3.1 and 6.3.2 are displayed with dashed lines respectively in black, golden and red. The mixed alert system from Section 6.3.3 **with a cut-off of 30 minutes** is in green while a similar system using **double lagged forecasts** rather than simple lagged forecasts is displayed in purple.

HIT scores are not affected by the use of two lagged forecasts rather than only one to confirm predictions above 40 minutes. Indeed the purple and green curves are identical. Both are very close to the initial direct alert system: nearly no additional events are missed. Meanwhile the FAR of the double mixed system (in purple) are the smallest obtained until now. Moreover, for the first time, all three high false alarms have been strongly decreased. No false alarm will be issued above $17\,\mathrm{m^3\,s^{-1}}$ with this alert system. **Such mixed system between direct alerts and lagged forecasts allow to obtain very good performances both in HIT and FAR**.
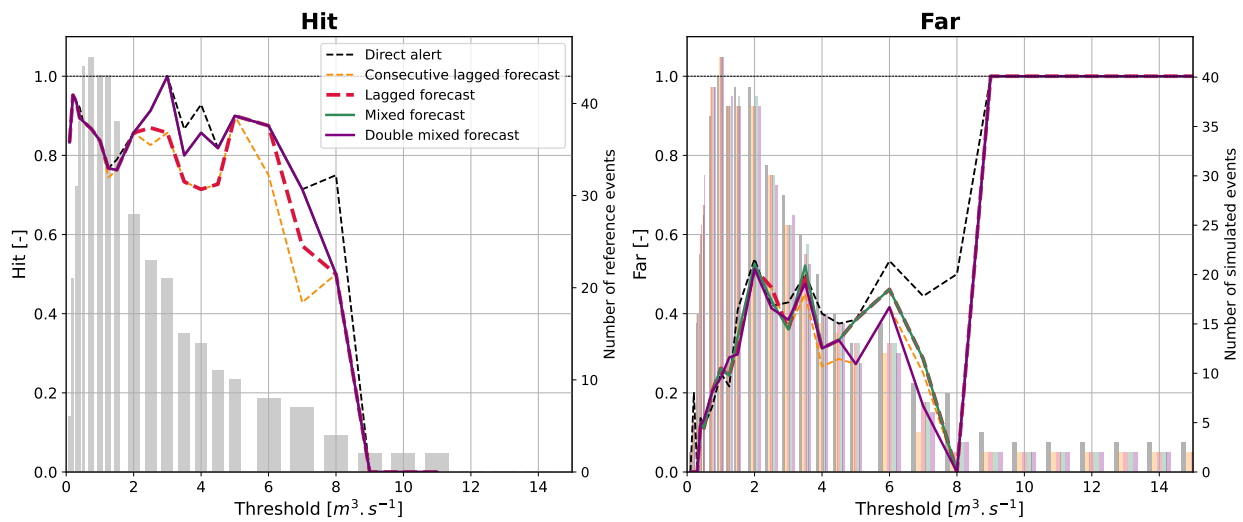
Figure 6.6: Alert HIT-FAR scores for alert systems based on model V with 90 minute of lead time. Alerts can be generated with one forecast only (black dashed), two consecutive forecasts (dashed golden), two lagged forecasts possibly non-consecutive (dashed red) or mixed systems where alerts are instantaneously issued before 30 minutes but need to be confirmed once (green) or twice (purple) above 40 minutes.

Table 6.6: Maximum predicted discharge – in $\mathrm{m^3\,s^{-1}}$ – of the three high false alarm events for several cutoff lead times, using several single-model alert systems with 90 minutes lead time.

Alert system	01/07/2019	18/06/2021	28/06/2021
Direct alert	19.0	31.5	31.6
Consecutive lagged forecasts	17.9	3.9	30.1
Lagged forecasts	17.9	7.2	30.1
Mixed forecast	17.9	7.2	30.1
Double mixed forecast	15.8	5.1	16.5
Observed discharge	8.0	2.0	9.0

6.4 Multimodel

Two distinct attempts to make the most of the two INCA models are developed. The goal is to obtain HIT scores close to those of model V while avoiding its high false alarms.

In the first model (in golden in Figure 6.7) an alert is directly launched if the two forecasts predict simultaneously a threshold crossing. The need for a redundancy between the two models allow to reduce some of the main false alarms as observed in Table 6.7. However false alarms are still numerous and the HIT scores are closer to model C than to the better model V.

Hence a second more complex system is developed (in purple in Figure 6.7). In this system, alerts are directly launched when model V predicts an alert in the next 30 minutes. However, if a threshold crossing is observed with more anticipation (at least 40 minutes) it needs to be confirmed by three other lagged forecasts either in model C or V. Here the results are largely improved compared to the two single model alert systems: HIT scores are close to those of model V while false alarms are reduced everywhere and notably at high intensity. Nevertheless, this system does not appear to be better than the single-model system "mixed double" (in dashed green) from Section 6.3.4.
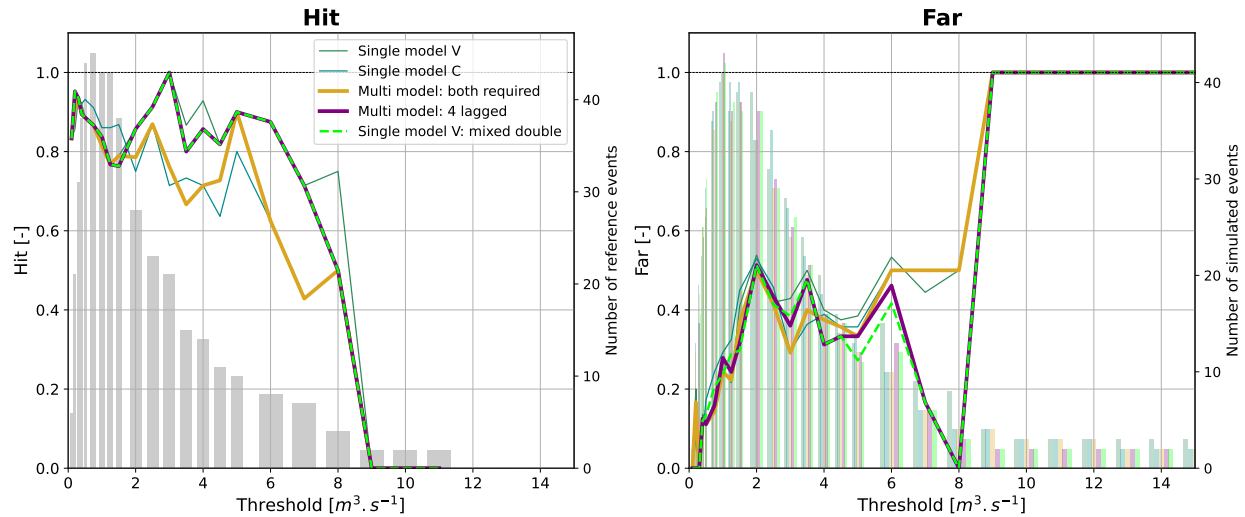
Figure 6.7: Alert HIT-FAR scores for several single- and multi-model alarm systems. The green and blue lines represent the single-model alert systems without lagged forecasts for models V and C respectively. The dashed lime curve is for the mixed single-model system based on model V (direct alert until 30 minutes, double lagged forecasts for higher lead times). Finally the gold and purple curves provide the scores of multi-model alert systems, respectively without lagged forecasts and with 4 forecasts required (on either of the two models).

Table 6.7: Maximum predicted discharge – in $m^3\,s^{-1}$ – of the three high false alarm events for several single- and multi-model alarm systems.

Alert system	01/07/2019	18/06/2021	28/06/2021
Single-model V	19.0	31.5	31.6
Single-model C	12.8	21.8	17.0
Single-model V - mixed double	15.8	5.1	16.5
Multi-model: both required	12.8	21.8	17.0
Multi-model: 4 lagged	15.8	6.4	14.6
Observed discharge	8.0	2.0	9.0

6.5 Best model

Thanks to the appropriate design of an alarm system, it is now possible to obtain the best detection HIT scores while avoiding most false alarms. Two systems have reached this goal. Both are **instantaneously issuing alerts if an event is observed by model V, 30 minutes or less after the forecast issue**. If the model predicts events later, **it requires a confirmation**. The two models differ on the type of confirmation required. One is based on a multi-model approach while the other one relies only on model V. Being simpler, the latter **is considered to be the best system** to implement, as of now.

The alert system could be further improved by developing more rainfall-runoff models. These models need probably to have more differences to be interesting in a multi-model approach. The direct use of QPEs or QPFs can also help to improve the alert system. However, **the main limitation of the alert system is due to precipitation forecasts**. Hence the most promising way of improvement is probably to turn to **ensembles** and probabilistic approaches. Another significant limitation of the system is the **lack of data**. Indeed only three years of INCA were available with a maximum observed discharge of $11.5\,\mathrm{m^3\,s^{-1}}$ in the period, while the evacuation threshold has been set at $20\,\mathrm{m^3\,s^{-1}}$.

Finally, it is important to point that the choice of the system **needs to be developed in agreement with public authorities**, notably to decide where to put the trade-off between false alarms and missed events.

Chapter 7

Conclusion

The determination of reliable discharge forecasts is an essential component of operational flood management. In short response time catchments, prone to flash floods, the use of radar QPFs becomes necessary to increase the spatial resolution and the quality of very short lead time forecasts. In this project, radar data from *MeteoSwiss* were introduced in the rainfall-runoff model RS.

MeteoSwiss radar QPEs appear to be **mostly in agreement with rain gauges**. However some notable differences persist. Concerning the raw radar AZC, precipitation are largely **underestimated in winter and at small intensities and overestimated in summer and at high intensities** compared to rain gauges. While the winter underestimation is problematic, the summer overestimation highlights the ability of radar to catch small size precipitation cells not recorded by rain gauges. Moreover, *CombiPrecip* (CPCH), the radar-rain gauge merging QPE, largely improves precipitation estimation performances in winter. It results in a **very good agreement with rain gauges in winter and a slight and interesting overestimation in summer**. CPCH performances are however limited by two issues. First, CPCH performances, similarly to AZC, **degrade with altitude**. Secondly, in some regions of Switzerland, far from the *SwissMetNet* rain gauges used in the radar-rain gauge merging process, CPCH performances may become similar to AZC. **A de-biasing factor of up to 1.5 may be required to use radar products in place of rain gauges in rainfall-runoff models**. It was for example necessary in the mountainous Emosson basin while useless in the Zurich-Altbach catchment, which is at a lower altitude, and closer to several rain gauge stations.

Once precipitation data are de-biased, they can be introduced in rainfall-runoff models. To make the most of their spatial resolution, radar QPEs and QPFs are **introduced directly in the smaller geographical unit they belong to** – altitude bands in the case of RS. Models are then calibrated. During the calibration process, **time should be mostly used to improve the modelling of physical phenomena**. Well-calibrated models have proven to be quite resilient to a change of input data, which means it is not necessary to completely modify the calibration when using different types of QPEs. Among the physical phenomena that should be investigated for flash flood forecasting, **a particular focus must be given to infiltration**. Notably, the fraction of impervious surface of the total basin – due to either natural (rocks) or artificial (roads, buildings) soil sealing – is a critical parameter of the calibration. Its introduction in the Emosson catchment largely improves the continuous model, while **its modification at high precipitation intensities in Zurich allows to catch the most severe flash flood events** that were previously missed.

In operational forecast, it is confirmed that **the update step is highly recommended**, even for sharp flash flood events. **Post-processing, notably with cubic splines, is detrimental** for the issue of flash flood alerts in small response time catchments, but appears interesting for cumulated volume estimation (critical quantity for hydroelectricity purposes) or in larger basins. When replacing radar QPEs by radar QPFs (*i.e.* INCA), a **clear loss of performances** is observed

both in Emosson and Zurich-Altbach. However, compared to the COSMO NWP, INCA provides **significantly better performances with notably a good reactivity**. The two main issues of INCA discharge forecasts are the **small alert anticipation time and the high false alarms**. The latter are due to incorrect localization of precipitation cells inside of the catchment. This problem, which may be reduced by the use of ensembles, is mostly present in small-size catchments such as La Fouly and Zurich-Altbach. Another solution to mitigate both difficulties is to design an appropriate alert system, by using forecast redundancy or multi-model approach. The **necessity to invest time in the design of the alert system** has been demonstrated in this project.

This project needs to be continued to further improve operational discharge forecasts with radar data. Notably, **ensembles of radar QPFs appear to be an interesting approach**. It is an emerging field of research and seems to be **adapted to the present situations, notably in Zurich-Altbach**. Ensembles may be provided by *MeteoSwiss* in the INCA product or may be created from the median scenario used in this report. In the latter case, ensembles can be produced by spatial or temporal modifications, by the addition of random noise, or by the use of several hydrological models. An important focus will then need to be put on the integration of ensemble results into a deterministic (or probabilistic) alert system. In a world that is becoming more and more prone to severe natural hazards, as stated by the WMO[6], it is necessary that all actors of flood management integrate uncertainty at all steps of the process, from precipitation forecasts to final decisions and actions, passing by discharge predictions and alert issue.

Appendix A

HIT-FAR for alerts - computation details

Figure A.1 describes with some examples the HIT and FAR computation for alerts. Events (both realized and announced) are considered as distinct events if there is a gap of 90 minutes between the moment when the first event is passing below the threshold and the moment when the second event is passing above it. As explained in Section 3.3.2.3, for the hit computation, only the 90 first minutes of the forecast generated between 90 minutes and 20 minutes before the observed crossing are considered. If any of them predicts a crossing between 90 minutes before and strictly less than 20 minutes after the observed crossing, the event is hit. The hit computation scheme is provided in Figure A.2a. For the FAR score, an announced event from a forecast is not a far if the reference is crossing the threshold between the moment the alarm is launched and +90 minutes after the predicted time of crossing. The process is described in Figure A.2b.

(a)

(b)

(c)
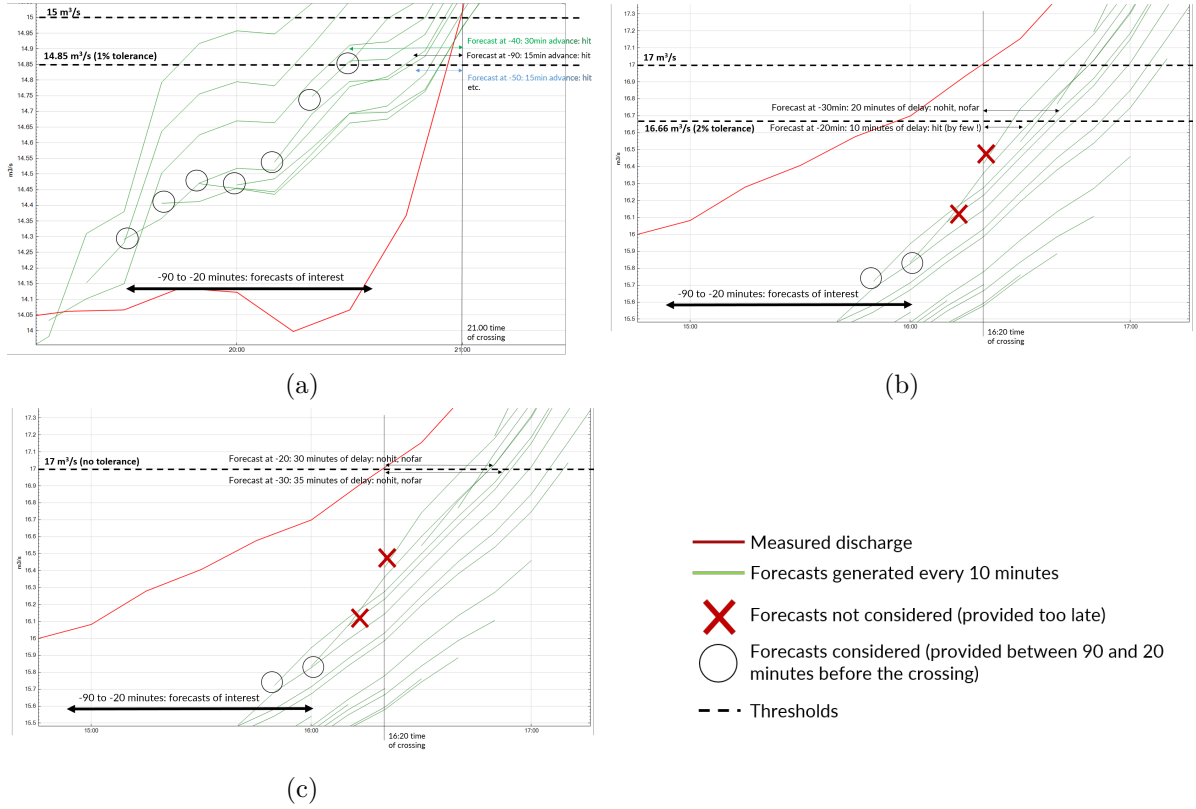
Figure A.1: Computation of alert HIT-FAR scores. Events are characterized by their time of threshold crossing. To be a hit, an event must be forecasted to be between 90min before and 20min after the event. The forecast must also be provided between 90 and 20min before the event (circles). Forecasts provided afterwards are provided too late (red crosses). An alarm is not a false alarm if an event occurs between 90min before and 90min after the predicted event. In case A.1a, the 8 forecasts between 90 and 20min before the event (21h) are considered. Each of them predict a crossing of the threshold (with tolerance) before 21h so it is a hit event. In case A.1b the forecast at -20min predicts the event just 10 minutes after the real event (16h20). Thanks to both temporal and discharge tolerances, as one of the forecast produces a good alarm, this event is a hit. However, in case A.1c where no tolerance is applied on the discharge, the event is predicted too late and is considered to be missed. It is not a far event as an event is observed after the prediction time and before the predicted event.

(a)

(b)
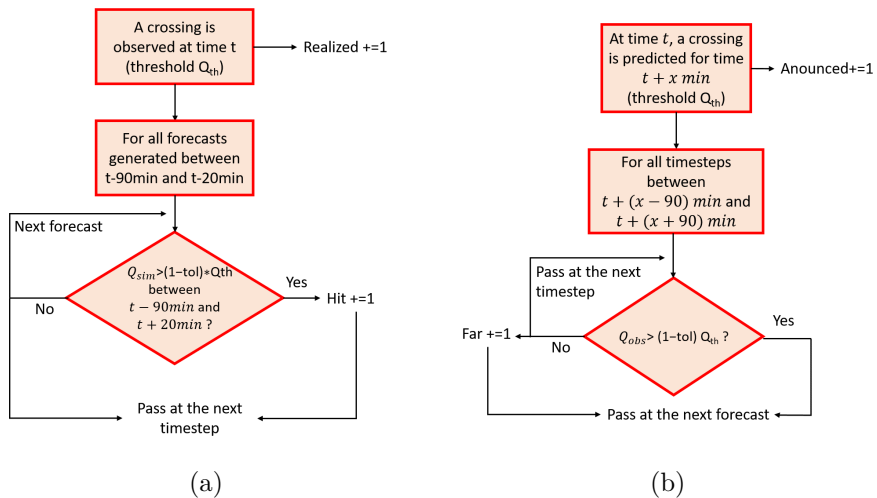
Figure A.2: Schemes of computation followed for determination of hit (left) and far (right) events.

Appendix B

Additional figures for the assessment of radar products

B.1   AZC regional performances

B.1.1   Regional dependency

The results of the raw radar AZC product according to the region of Switzerland are provided below in Figures B.1 (HIT-FAR scores) and B.2 (mean monthly precipitation). The ratios of CCDF were provided in Figure 4.8. The observations from Section 4.3 and 4.4.1 are still valid.
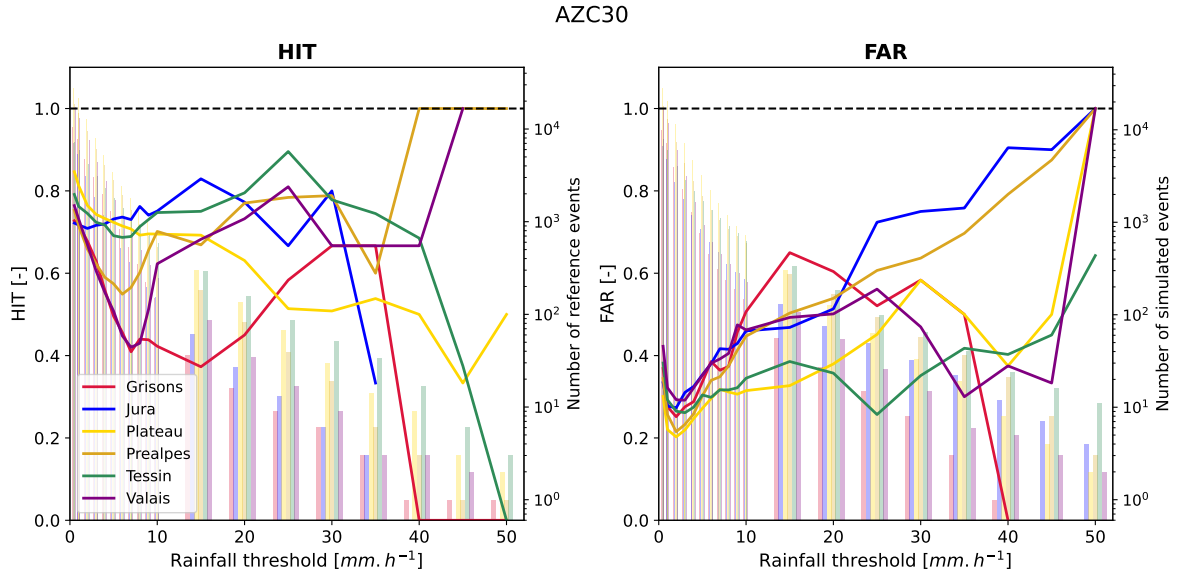
Figure B.1: HIT/FAR regional mean values of AZC for the period from 17/07/2018 0h to 28/09/2021 0h (local time) with a tolerance of 30%. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).
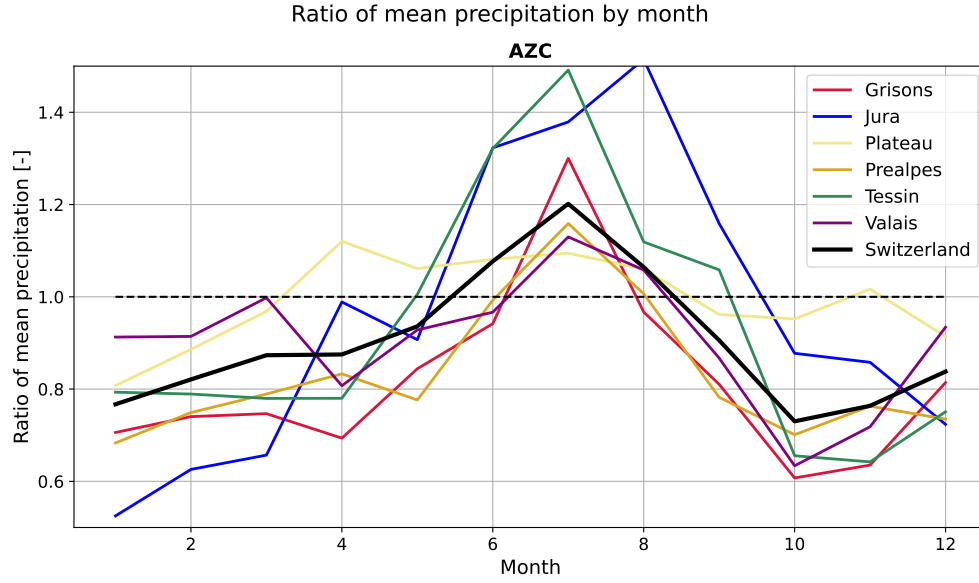
Figure B.2: Month precipitation ratio between AZC data and rain gauge data in the period from 17/07/2018 0h to 28/09/2021 0h (local time). Both AZC and rain gauge data are averaged by regions.

B.1.2 Altitude dependency

The results of the raw radar AZC product ordered by class of altitudes are provided below in Figures B.3 (HIT-FAR scores), B.4 (mean monthly precipitation) and B.5 (cumulative distribution functions). The observations from Section 4.3 and 4.4.2 are still valid with curves ordered by altitude – higher performances, lower bias and seasonal behaviour, more small and high intensity events at low altitude – and with poorer performances and higher seasonal bias in AZC.
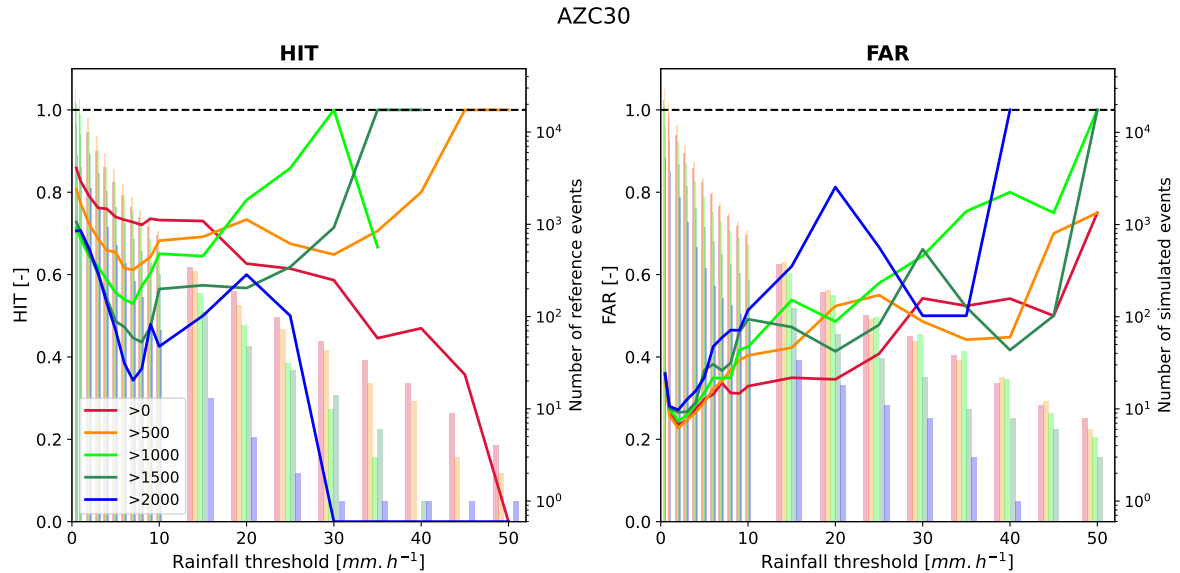
Figure B.3: HIT/FAR scores of AZC aggregated by station altitudes for the period from 17/07/2018 0h to 28/09/2021 0h (local time) with a tolerance of 30%. The right axis provides the number of reference events (for the HIT score) and simulated events (for the FAR score).
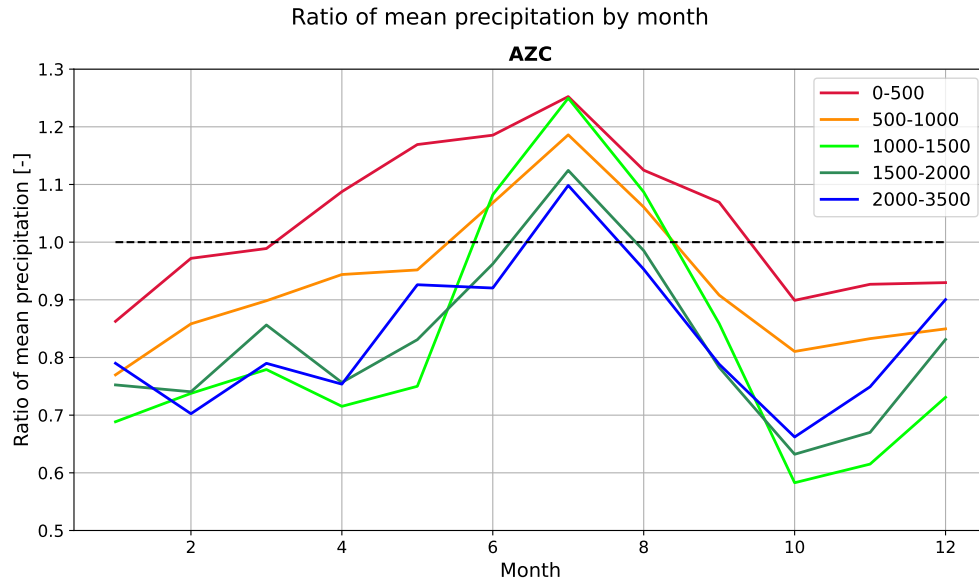
Figure B.4: Month precipitation ratio between AZC data and rain gauge data in the period from 17/07/2018 0h to 28/09/2021 0h (local time). Both AZC and rain gauge data are averaged by altitude.
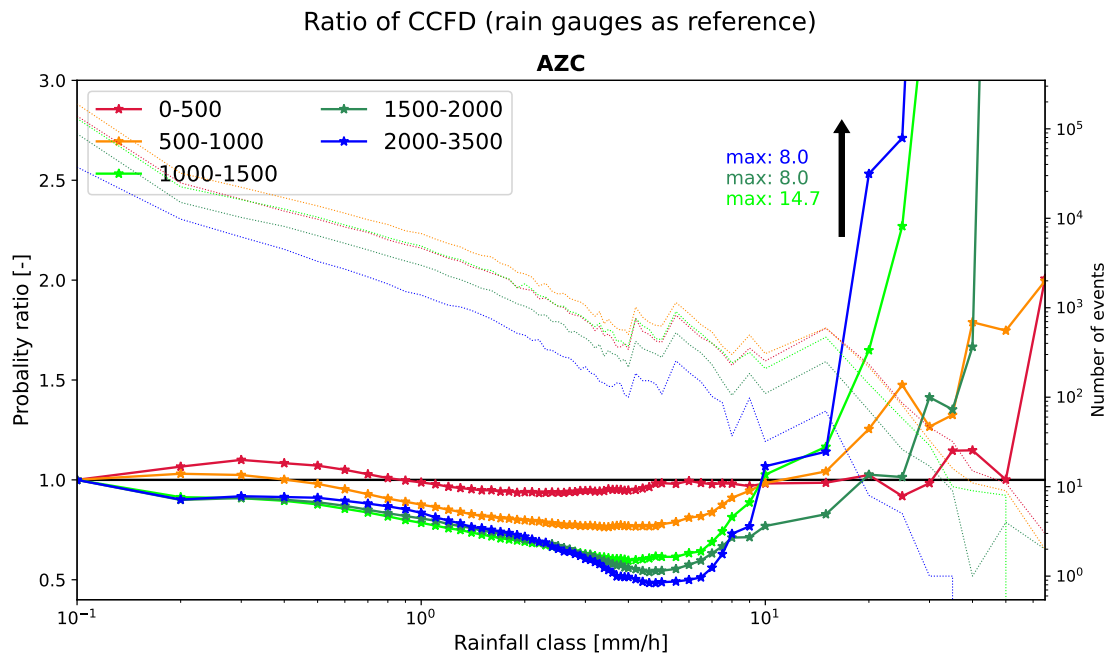
Figure B.5: Ratios of the CCDF of AZC over the CCDF of rain gauges for different altitudes between 2018 and 2021. A logarithmic scale is used for the x-axis.

B.2 Co-kriging performances far from *SwissMetNet* stations

B.2.1 Overall performances

The classic analyses of this report are provided for the two networks in this section. HIT-FAR scores are provided in Figure B.6 while monthly mean precipitations are provided in Figure B.7.

Concerning the HIT and FAR scores, while they were largely better for CPCH than for AZC in the *SwissMetNet* stations, they are nearly similar between them (very slightly better for CPCH) in the BVE network. The co-kriging is however more efficient in Ticino where CPCH is clearly improved both in HIT and FAR, compared to AZC.

For the mean monthly precipitation, the analysis joins the bias analysis presented in Section 4.7.3. In Bern, while AZC has similar shapes in both networks, CPCH and inca0 are significantly different. On the contrary, the shapes of CPCH (and inca0) are pretty similar in the oasi and *SwissMetNet* networks in Ticino.

(a) Bern region - BVE network

(b) Ticino region - oasi network
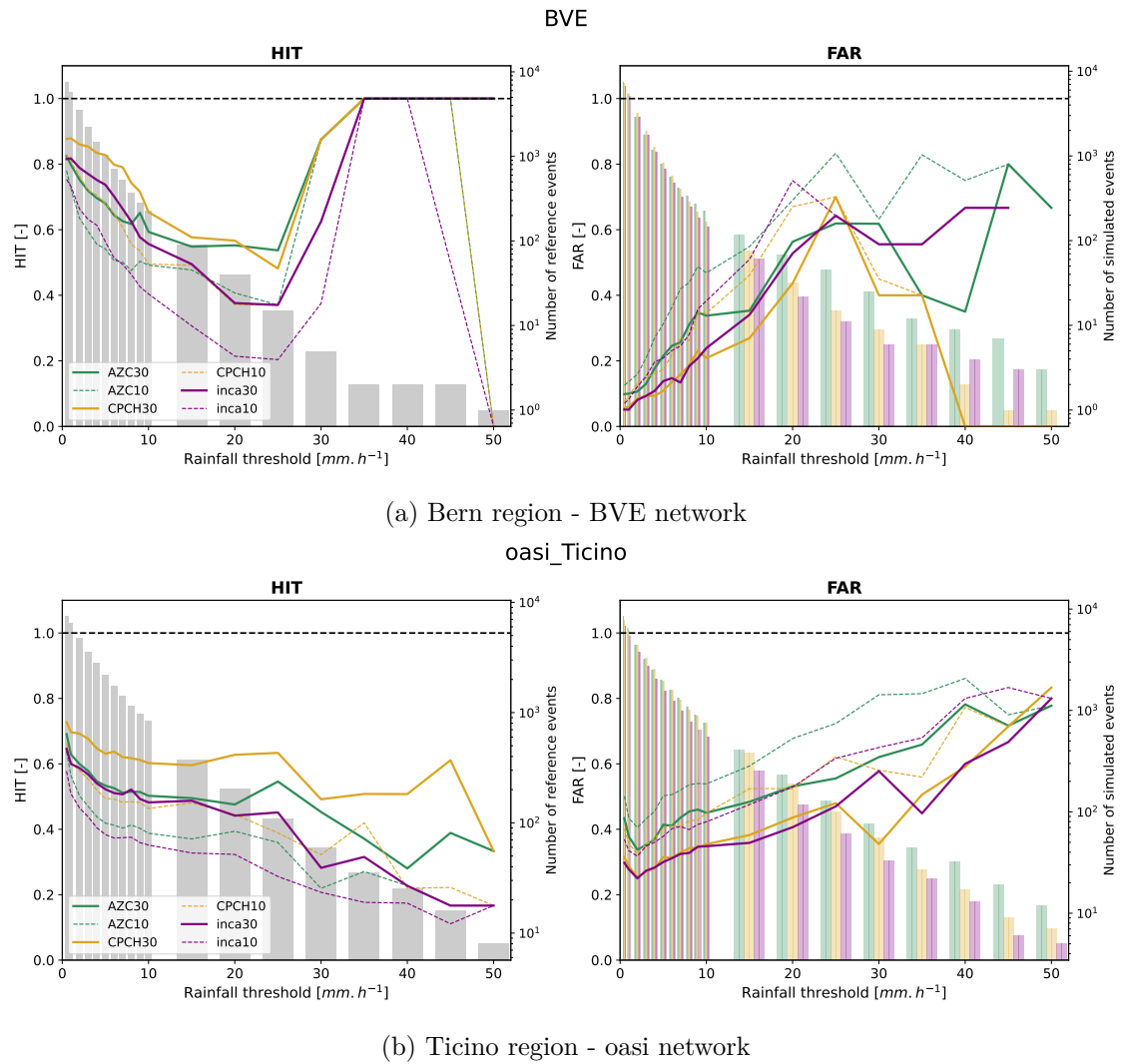
Figure B.6: HIT/FAR scores of AZC, CPCH and inca0 for the period from 2018-2021 for the two cantonal networks. Two tolerances (10% – dashed lines – and 30% – solid lines) are used.

(a) Bern region - BVE network

(b) Bern region - *SwissMetNet*

(c) Ticino region - oasi network

(d) Ticino region - *SwissMetNet*
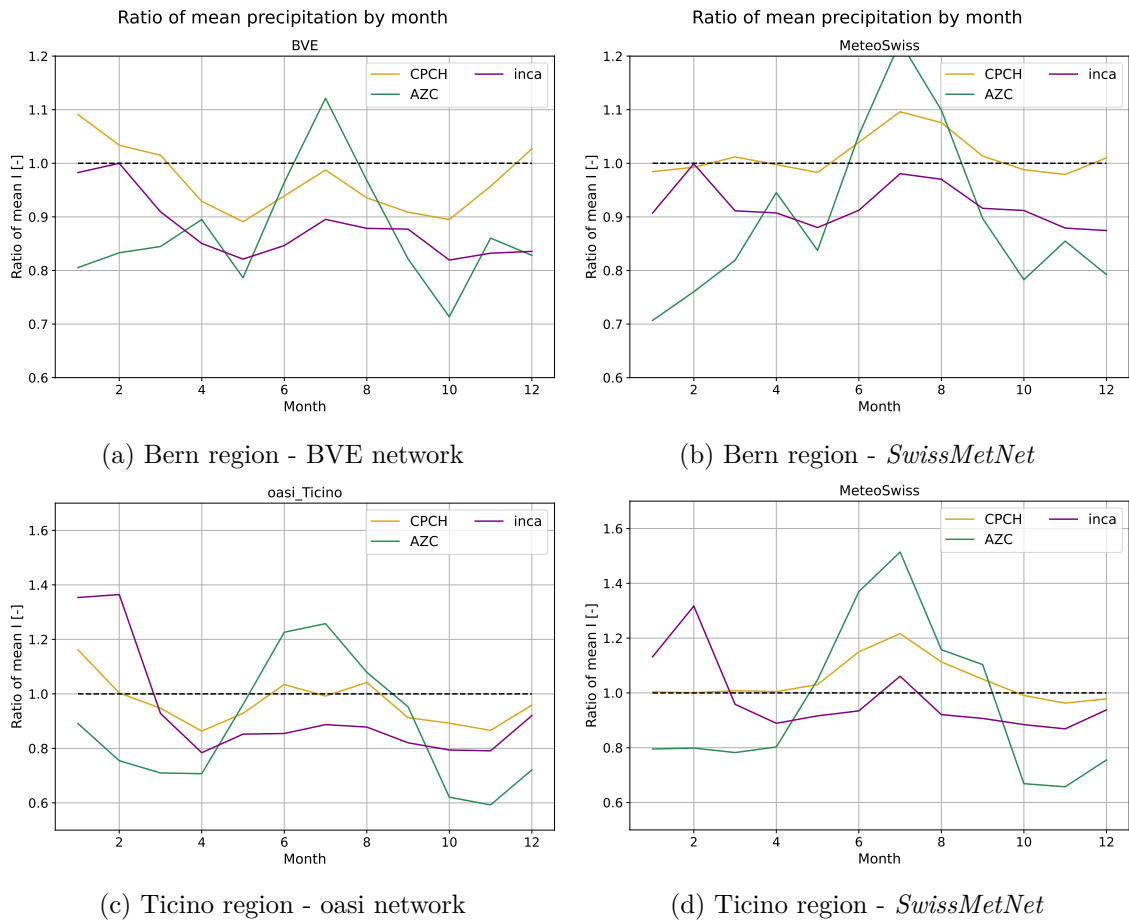
Figure B.7: Radar mean monthly precipitation analysis in Bern (top) and Ticino (bottom) cantons for the cantonal network (left) and for *SwissMetNet* (right). The period of analysis is 2018-2021.

B.2.2 Influence of altitude and distance on co-kriging performances

The 14 stations of the BVE network and the 19 stations of the oasi network can be ordered according to their distance to the closest *MeteoSwiss* station or to their altitude. Table B.1 provides the number of rain gauges in each class. HIT-FAR and mean monthly precipitation analyses, according to the distance and altitude, are provided by Figures B.8 to B.11. All differences between the AZC and CPCH curves appear to be similar for each altitude and for each distance.

Table B.1: Number of BVE and oasi stations by altitude and distance to the closest *SwissMetNet* station.

	BVE		**oasi**	
	Class	Number of stations	Class	Number of stations
Altitude	0-1000m	5	0-500m	7
	1000-1500m	6	500-1000m	6
	>1500m	3	>1000m	6
Distance	0-5km	2	0-4km	4
	5-7.5km	5	4-6km	4
	7.5-10km	4	6-8km	8
	>10km	3	>8km	3

(a) Distance to *SwissMetNet*
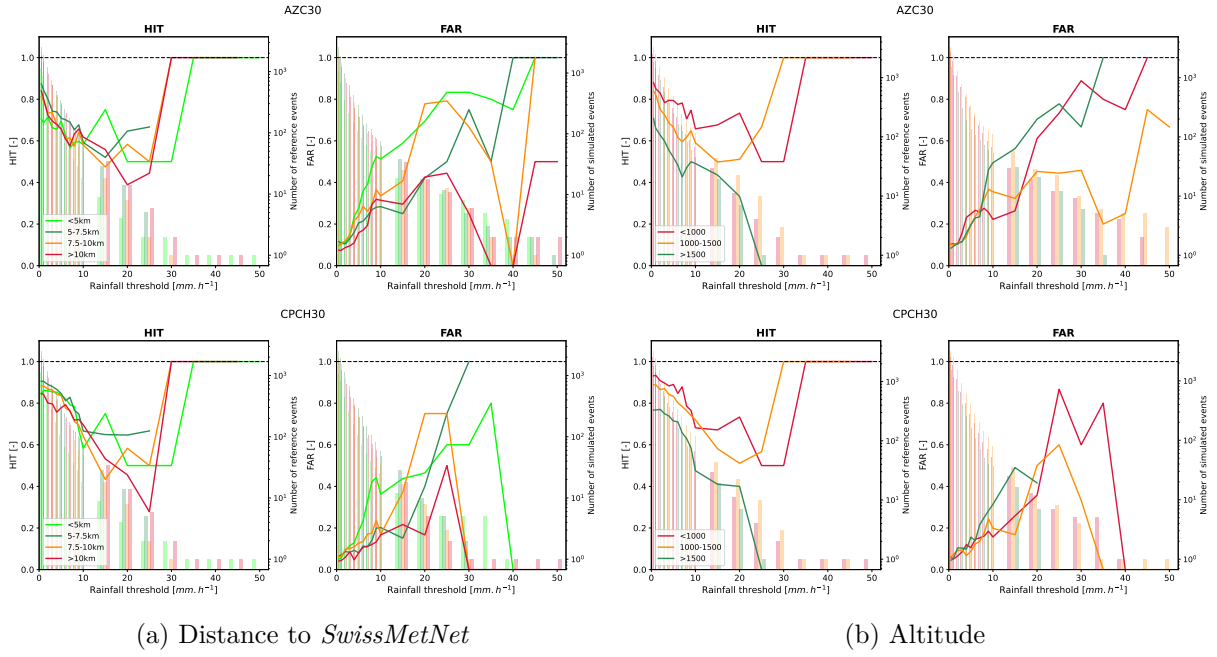
(b) Altitude

Figure B.8: BVE HIT/FAR scores of AZC or CPCH between 2018 and 2021 according to distance to the *SwissMetNet* network (left) or to altitude (right). A tolerance of 30% is applied. Improvements between AZC and CPCH are similar at all distances and altitudes.

(a) Distance to *SwissMetNet*
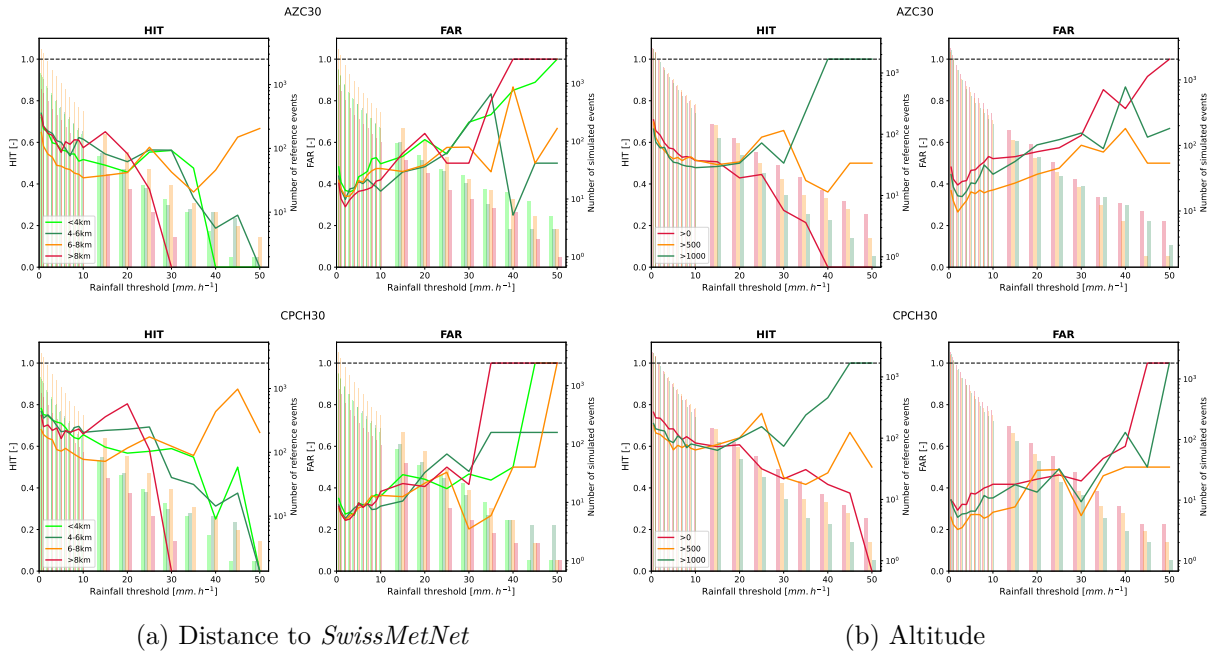
(b) Altitude

Figure B.9: oasi HIT/FAR scores of AZC or CPCH between 2018 and 2021 according to distance to the *SwissMetNet* network (left) or to altitude (right). A tolerance of 30% is applied. Improvements between AZC and CPCH are similar at all distances and altitudes.
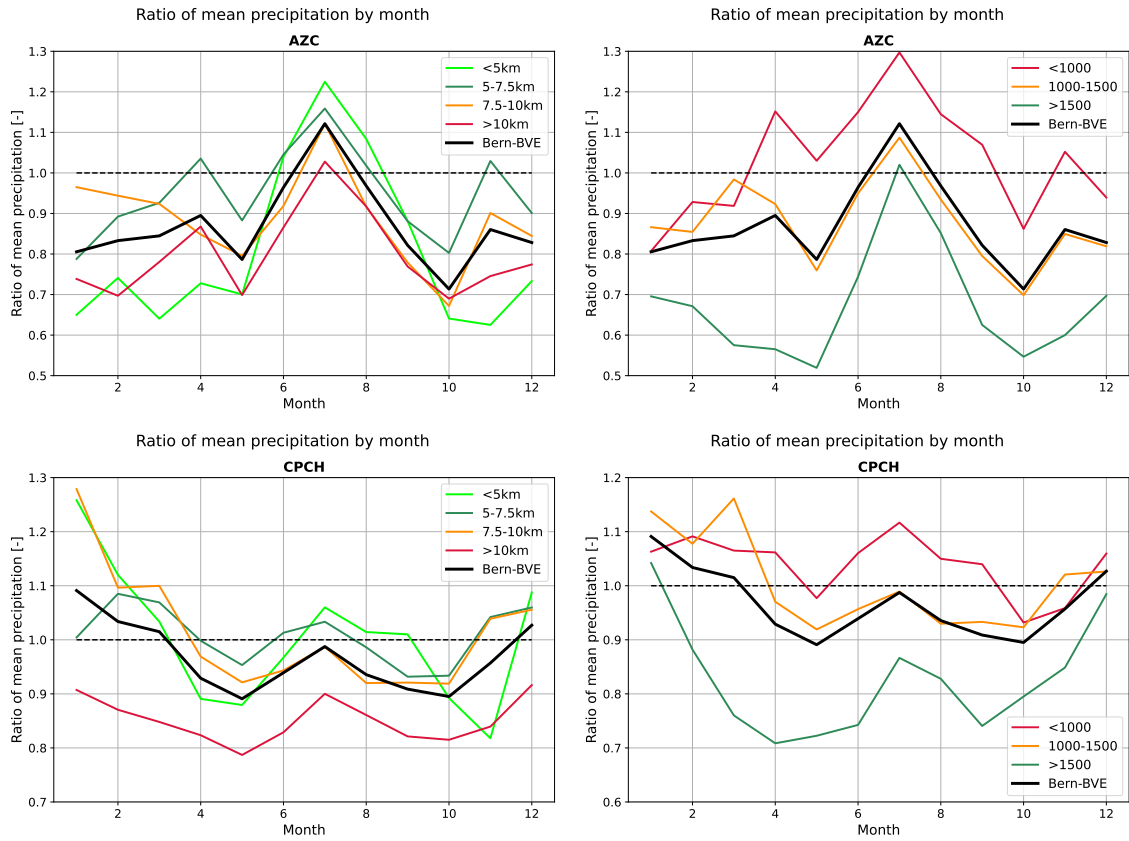
Figure B.10: Month precipitation ratio between AZC or CPCH and rain gauges from the BVE network according to the distance to the *SwissMetNet* network (left) or to the altitude (right).
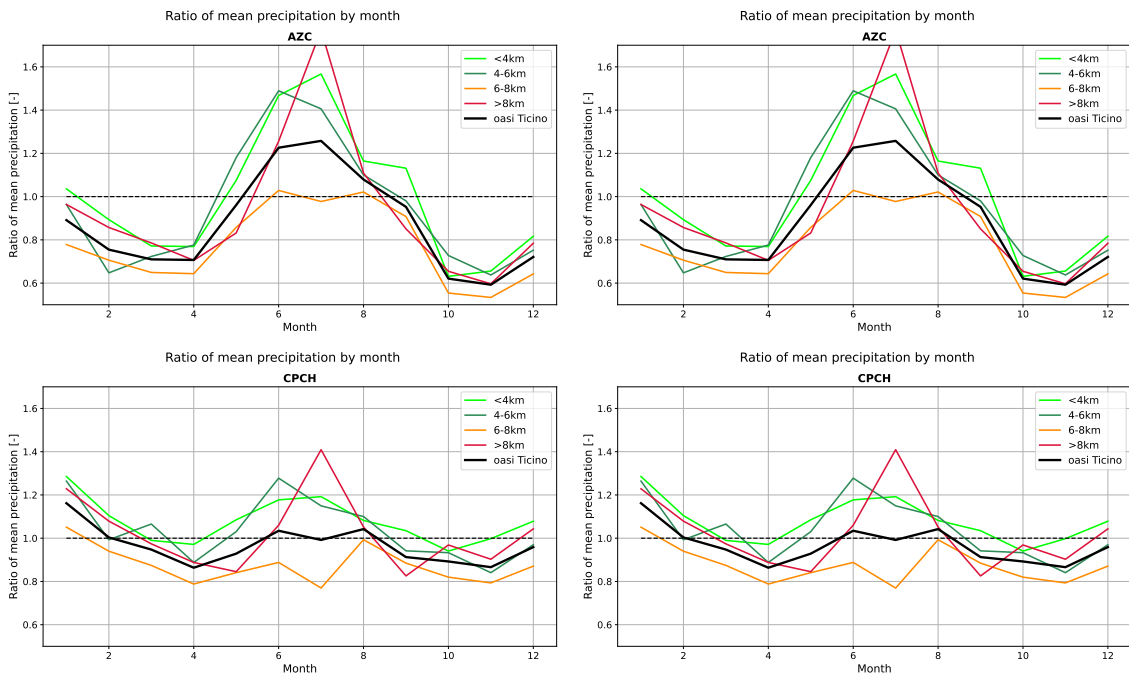
Figure B.11: Month precipitation ratio between AZC or CPCH and rain gauges from the oasi network according to the distance to the *SwissMetNet* network (left) or to the altitude (right).

Appendix C

Parameters of RS models

C.1 Emosson

C.1.1 Update of the model with new temperature stations

The main parameters of the RS models used in Section D.2.1 are provided in Tables C.1 and C.2. The former contains the parameters that are identical in all four models while the latter contains the parameters that have been modified.

For the slow reservoirs, these tables only report their presence or absence. Their parameters are not detailed as they do not impact strongly the models. Similarly, the snow melt parameter An has been calibrated by altitude bands. However we only report here its minimal and maximal values in the lower altitude bands which are the one playing the main role in snow melt discharge.

Table C.1: Parameters of the Emosson models not modified for the four models from Section D.2.1.

Parameter	Unit	La Fouly	Saleina	Trient	Coll. Est
Virtual station					
coeffI	-	1	0.75	1	
coeffETP	-	1			
gradT (Dry/Wet)	°C m^{-1}	-0.0054/-0.006	-0.0054/-0.005	-0.006/-0.005	-0.0054/-0.0054
iCrGradTHum	mm h^{-1}	3			
SWMM					
Area imp	-	0			
K_s	m$^{1/3}$ s^{-1}	Tab. C.2	0.8		0.7
GR3					
Hmax	m	0.3	0.25	0.3	
K	-	0.003			
aK	-	5	6		5
aInf	-	2	3		2
Slow reservoir		Tab. C.2	✗		
Glacier					
KN	-	0.1	0.05	0.2	
KGL	-	Tab. C.2	0.3		
Slow reservoir		✓(slight variations)	✓		

Table C.2: Parameters of the Emosson RS models modified between the four models from Section D.2.1.

Parameter	Sub-basin	Aigle	Les Marécottes	Montagnier	Montagnier-LM
Virtual station					
$gradI$ [m/s/m]	La Fouly	0.4	0.32	0.23	0.5
	Saleina	0.45	0.5	0.28	
	Trient	0.3	0.5	0.2	
	Coll. Est	0.5			
$coeffT$ [°C]	La Fouly	0.4	-0.8	-0.7	-0.45
	Saleina	-0.4	-1.25	-0.7	
	Trient	-0.3	-1.2	-1	
	Coll. Est	0			
SWMM					
K_s [m$^{1/3}$s^{-1}]	La Fouly	0.7	0.8	0.7	
GR3					
Slow reservoir	La Fouly	✓			✗
Glacier					
Agl [mm/°C/d]	La Fouly	6		4	5
	Saleina	2.8	2.7		2.4
	Trient	4.5	3	3.8	
	Coll. Est	4.5			
KGL [-]	La Fouly	0.2	0.3		0.2
Snow					
An [mm/°C/d]	La Fouly	$\sim$ 0.5-2.5	$\sim$ 1-2.5		
	Saleina	$\sim$ 1-2.5		$\sim$ 0.8-1.5	$\sim$ 0.5-1.2
	Trient	$\sim$ 1-2			
	Coll. Est	$\sim$ 1.5-2.5			$\sim$ 1-2

C.1.2 Integration of radar data

Tables C.3 and C.4 provide the parameters of models A.1. (rain gauge input), A.2. (CPCH), A.4. (AZC) and B.1. (rain gauge). The former reports the parameters that have not been modified in the four models while the latter reports the modified parameters. For the other models, model A.3. (CPCH) and A.5. (AZC) are **based on model A.1. with two modifications**: *coeffI* is equal to 1.4 and *gradI* is null in both models and for all four sub-catchments. Similarly models B.2. (CPCH) and B.3. (AZC) are **based on B.1. with the same two modifications** of *coeffI* and *gradI*. Finally 10-minute timestep models B.4., B.5. and B.6. are **based respectively on models B.1., B.2. and B.3.** with the critical precipitation for humid temperature gradient *iCrGradTHum* passing from $3\,\mathrm{mm\,h^{-1}}$ to $2\,\mathrm{mm\,h^{-1}}$.

In the last model B.1. impermeable areas (in the SWMM model) and radiation effects (snow and glaciers) have been introduced in La Fouly and Trient.

Table C.3: Parameters of the Emosson RS models not modified between the models A.1., A.2., A.4. and B.1. from Sections D.2.2 and D.2.3.

Parameter	Unit	La Fouly	Saleina	Trient	Coll. Est
Virtual station					
coeffETP	-	1			
gradT (Dry/Wet)	$^\circ\mathrm{C\,m^{-1}}$	See Tab. C.4	-0.0054/-0.005	See Tab. C.4	-0.0054/-0.0054
iCrGradTHum	$\mathrm{mm\,h^{-1}}$	3			
SWMM					
Imp	-	See Tab. C.4	0	See Tab. C.4	0
K_s	$\mathrm{m^{1/3}\,s^{-1}}$	0.7	0.8		0.7
GR3					
Hmax	m	0.3	0.25	0.3	
K	-	0.003			
aK	-	5	6		5
aInf	-	2	3		2
Glacier					
Slow reservoir		✓(slight variations)			
Ray	-	See Tab. C.4	0	See Tab. C.4	0
SRF	mm/W/d	See Tab. C.4	0	See Tab. C.4	0
Snow					
SRF	mm/W/d	See Tab. C.4	0		

Table C.4: Parameters modified between models A.1., A.2., A.4. and B.1.

Parameter	Sub-basin	A.1. (rain-gauges)	A.2. (CPCH)	A.3. (AZC)	B.1. (rain-gauges)
Virtual station					
$coeffI$ [-]	La Fouly	0.9	1.4		0.9
	Saleina	0.9	1.3	1.2	0.9
	Trient	0.9	1.3	1.25	0.9
	Coll. Est	0.9	1.3	1.5	0.9
$gradI$ [m/s/m]	La Fouly	0.5	0		0.5
	Saleina	0.3	0		0.3
	Trient	0.35	0		0.35
	Coll. Est	0.5	0		0.5
$coeffT$ [°C]	La Fouly	0		0.1	0
	Saleina	-0.7	-0.4	-0.2	-0.7
	Trient	-1	-0.5	-0.2	-1
	Coll. Est	0	0.5		0
$gradT$ (D/W) $°C\,m^{-1}$	La Fouly	-0.0054/-0.006			-0.0054/-0.0054
	Trient	-0.006/-0.005			-0.006/-0.0038
SWMM					
Imp [-]	La Fouly	0			0.1
	Trient	0			0.2
GR3					
Slow reservoir	La Fouly	✗	✓		
	Saleina	✗	✓		✗
	Trient	✓			
	Coll. Est	✗	✓		✗
Glacier					
Agl [mm/°C/d]	La Fouly	7.5	5.9		2
	Saleina	1.9	1.7		1.9
	Trient	4.4	3.5		3.2
	Coll. Est	6.5			
KN [-]	La Fouly	0.1	0.05	0.2	0.02
	Saleina	0.05			
	Trient	0.05			0.07
	Coll. Est	0.2	0.3		0.2
KGL [-]	La Fouly	0.2	0.1	0.3	0.07
	Saleina	0.3	0.1	0.15	0.3
	Trient	0.1		0.15	0.2
	Coll. Est	0.5	0.2		0.5
Ray [-]	La Fouly	0			1.8
	Trient	0			0.25
SRF [mm/W/d]	La Fouly	0			35
	Trient	0			8
Snow					
An [mm/°C/d]	La Fouly	∼ 0.8-2.2	∼ 1.5-2.5	∼ 1-2	∼ 0.4-1.6
	Saleina	∼ 0.3-1.2	∼ 1-1.2		∼ 0.3-1.2
	Trient	∼ 1.5-2	∼ 0.6-2	∼ 0.3-2	∼ 0.8-1.7
	Coll. Est	∼ 0.3-1.5	∼ 0.8-1.5		∼ 0.3-1.5
SRF [mm/W/d]	La Fouly	0			15

C.2   Zurich

C.2.1   First models

The parameters for Zurich catchment are provided in Table C.5. As the calibration process is only based on a calibration on stations and a de-biasing for radar inputs, all models are similar except for *coeffI*, which is slightly larger in inca0, and *gradI* which needs to be put to 0 for radar input data.

Table C.5: Parameters of the Zurich RS models from Section 5.3.1.

Parameter	Unit	Rain-gauge	CPCH	RZC	INCA0
Virtual station					
coeffI	-		0.9		0.95
gradI	m/s/m	0.3		0	
coeffT	°C		0		
gradT (Dry/Wet)	°C m^{-1}		-0.0054/-0.00378		
iCrGradTHum	mm h^{-1}		3		
coeffETP	-		2		
SWMM					
K_s	m$^{1/3}$ s^{-1}		0.3		
Imp	-		0.1		
GR3					
Hmax	m		0.4		
K	-		0.003		
aK	-		5		
aInf	-		2		
Slow reservoir			✓		
Snow					
An - 600-900m	mm/°C/d		2.5		

C.2.2   Improved models

Table C.6 provides the parameters that have been modified in the two new models. Figure C.1 displays the schemes of the three models (old model and 2 new models with/without variation of the impermeable area *Imp*) with notably the river sections that have been added. These river sections have similar characteristics, and have lengths comprised between 500 and 700 m. Between the two new models, there is two differences:

- The impermeable area is constant and equal to Imp_0 in both sub-basins in the model with constant impermeable area while it follows Equation C.1 in the lower sub-basin in the model with variable impermeable area.

- More rivers are added in the model with variable impermeable area as the response is accelerated at high intensity. This response needs to be delayed to match the observed discharge.

$$Imp = Imp_0 + \frac{Imp_{max} - Imp_0}{1 + exp\left(-\frac{i - i_{1/2}}{i_0}\right)} \tag{C.1}$$

Table C.6: Parameters modified in the new Zurich RS models from Section 5.3.3.

Parameter	Unit	Previous value	Upstream catchment (Bassersdorf	Downstream catchment
SWMM				
K_s	$\mathrm{m}^{1/3}\,\mathrm{s}^{-1}$	0.3	0.8	0.3
Imp_0	-	0.1	0.1	0.12
Imp_{max}	-	0.1	0.1	0.4
$i_{1/2}$	$\mathrm{mm}\,\mathrm{h}^{-1}$	-	-	16
i_0	$\mathrm{mm}\,\mathrm{h}^{-1}$	-	-	4
GR3				
Slow reservoir		✓	slight modifications	slight modifications
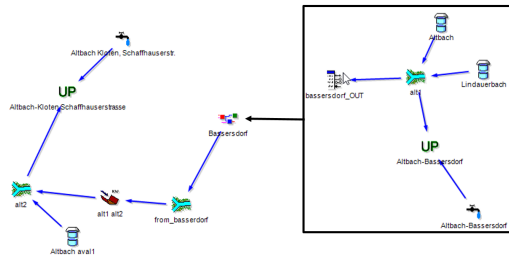
(a) Old models
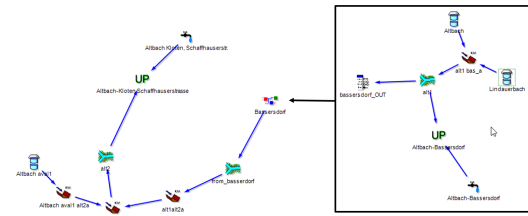
(b) New model with constant impermeable area
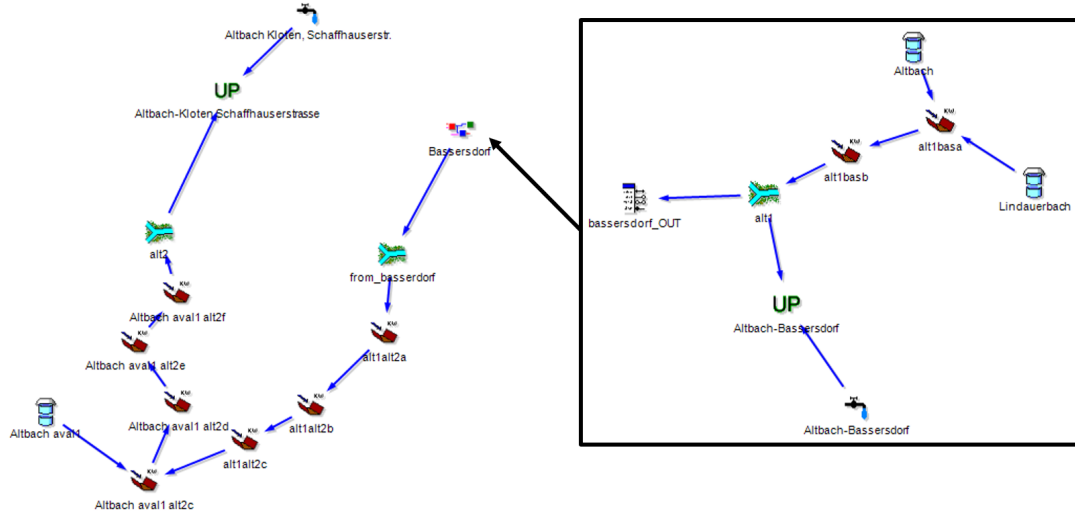
(c) New model with variable impermeable area

Figure C.1: Schemes of the old model and of the two new models (with/without variable impermeable area). River sections have been added in the new model to delay the response. They are modelled *via* a kinematic wave model (KW).

Appendix D

Integration of radar data in the Emosson RS model

This chapter provides the detailed results of the study developed on the Emosson basin. A synthesis of these results is proposed in Section 5.2.

The Emosson catchment is a mountainous basin with a glacial behaviour. As explained in Section 3.4.1, daily HIT-FAR scores will be used to assess the performances of continuous models while alert HIT-FAR scores and cumulated volume indicators will be studied in the hindcast process.

D.1 Method adaptations for mountainous catchments

D.1.1 Continuous HIT-FAR scores

For mountainous catchments (Emosson) where the snowpack and the glaciers play a determinant role creating a strong seasonal curve – where daily variations of the discharge are smaller than its seasonal variation – the event separation is difficult. Indeed discharge will be higher than small thresholds during a prolonged period (e.g., whole summer) and this period will be considered as only one event (or more if some data are missing...). One solution could be to subtract a seasonal component to the total discharge. Nevertheless such method has not been developed here and daily HIT and FAR scores will be computed instead. These scores are computed by taking the daily maximum values and comparing them between simulated and reference data. Time is not anymore important: if the maximum of the reference discharge occurs at 1h and the maximum of the simulated discharge the same day has the same value but at 23h it will still count as a hit.

A tolerance of 10% is still applied on the value but with a different mechanism for the FAR computation. For FAR by events, the chosen threshold is applied to the simulated discharge and the event is a false alarm if the reference discharge is below 90% of this threshold. On the contrary, a 110%-threshold is applied to the simulated discharge and the daily event is a far if the reference is below 100% of this threshold. The daily HIT-FAR scores mechanism is presented in Figure D.1. For the HIT computation there is no differences in the management of tolerance between both cases.
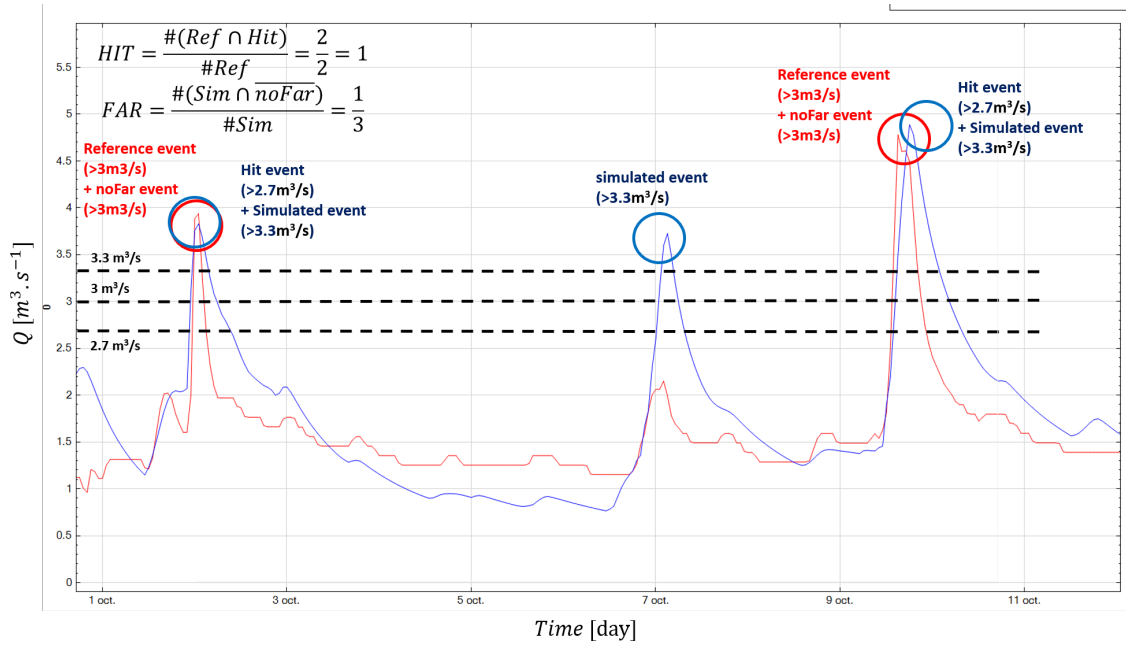
Figure D.1: Determination of hit and far events for daily analysis. The discharge threshold is $3\,\mathrm{m^3\,s^{-1}}$ with a tolerance of 10%. The reference data (river gauge) are in red and the simulated data are in blue. Reference events are the events (from the reference) exceeding the $3\,\mathrm{m^3\,s^{-1}}$ threshold while simulated events are those (from the simulation) exceeding the $3.3\,\mathrm{m^3\,s^{-1}}$ threshold (10% tolerance). A simulated event (resp. a reference event) is a noFar event (resp. a hit event) if the reference data (resp. simulated data) exceeds the $3\,\mathrm{mm\,h^{-1}}$ threshold (resp. $2.7\,\mathrm{mm\,h^{-1}}$).

D.1.2 Volume analysis

In Emosson, the knowledge of the future inflow cumulated volume is important for the reservoir management. Hence the performances of the models for volume predictions will also be evaluated. Two indicators will be used:

- HIT-FAR scores. They will be applied for cumulated volume over a given period of a forecast, usually 60 or 360 minutes. Usual HIT/FAR by event are obtained on these continuous datasets (same process as in the previous section) considering the maximum of the event (the exact timing of the event is not considered).

- The Mean Absolute Error (MAE) of the cumulated volumes will also be computed month by month. Its formula is provided by Equation D.1 with n the number of timesteps, V_{obs} the observed cumulated volume and V_{sim} the simulated one. It notably gives information on the seasonal dependency of forecast errors.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|V_{sim}(t_i) - V_{obs}(t_i)| \tag{D.1}$$

D.2 Calibration of the Emosson model

D.2.1 Update of the model with new temperature stations

As a first step, the RS model for Emosson basin is updated to use newly available temperature stations. Historically, the Emosson model of *Hydrique* was only fed by Aigle and Grand-Saint-Bernard (GSB) stations for respectively temperature and precipitation. With the integration of the Valais precipitation network in 2015, the GSB precipitation was discarded. On the temperature side new temperature stations have been recently integrated to the *MeteoSwiss* network: Les Marécottes has been integrated in 2015 and Montagnier-Bagnes in 2019 (but data are available since 2017). It could be interesting to improve the current model by replacing the Aigle station by those new stations which are closer to the basin. Therefore, different models, based on different sets of temperature stations, are developed and compared:

- Aigle: A model based only on the temperature from Aigle.
- Les Marécottes: A model based on the temperature from Les Marécottes and Champéry (the latter provided by *Meteogroup*).
- Montagnier: A model based on the temperature from Montagnier-Bagnes and Champéry.
- Montagnier-LM: A model combining the temperature data from Les Marécottes, Montagnier-Bagnes and Champéry.

All the models use rain gauge data as inputs and are calibrated independently. The main parameters of the models are provided in Appendix C.1.1. Table D.1 displays the Nash scores for each of the four sub-catchments on which the calibration is performed. Nash values for high reference discharges only are also proposed. Compared to the old operational model (which uses Aigle as temperature input), excepted in La Fouly, Nash scores have been largely improved. These values show that the best model uses Montagnier-Bagnes, Les Marécottes and Champéry as temperature inputs.

Table D.1: Nash scores of RS models for different temperature inputs. Best scores for each sub-basin are coloured in green.

	La Fouly		Saleina		Trient		Collecteur Est	
$Q >$	0	$3\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$5\,\mathrm{m^3\,s^{-1}}$
Old model	0.92	0.76	0.86	0.33	0.87	0.32	0.92	0.72
Aigle	0.88	0.63	0.88	0.44	0.89	0.53	0.93	0.76
Les Marécottes	0.88	0.65	0.88	0.47	0.9	0.59	0.94	0.79
Montagnier	0.9	0.69	0.87	0.4	0.9	0.6	0.95	0.82
Montagnier-LM	0.9	0.75	0.89	0.49	0.9	0.6	0.95	0.83

Moreover Figure D.2 displays the values of daily HIT and FAR discharge scores at the Collecteur Est for each model, along with the HIT and FAR scores of the old model (based on Aigle temperature data). The HIT and FAR scores of the four new models are all close to each other and are better than the old model. As observed during the calibration, the model Montagnier-LM is globally performing the best with the second highest HIT scores and the second lowest FAR scores. Hence, **for the next analyses, all models will be based on temperature data from Montagnier-Bagnes, Les Marécottes and Champéry.**
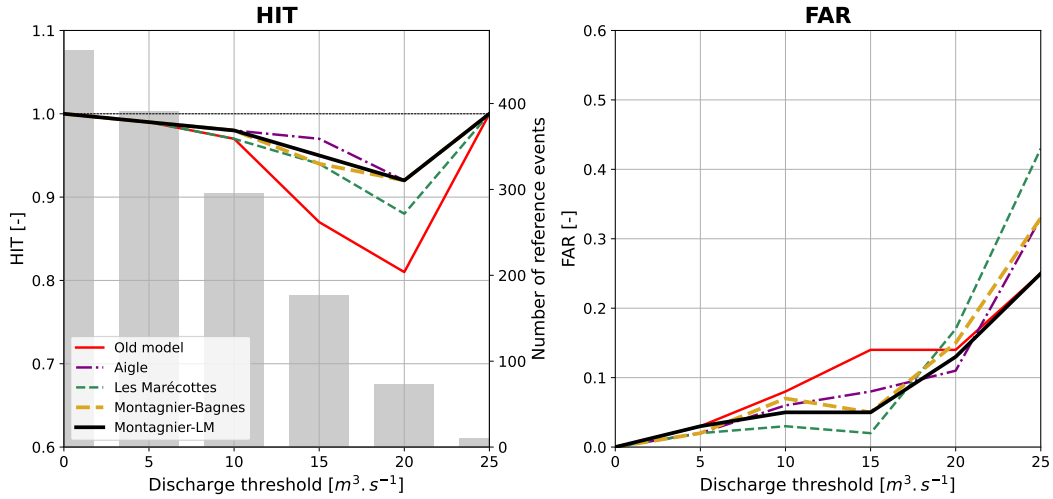
Figure D.2: Daily discharge HIT/FAR scores at Collecteur Est for the models with different temperature input data. The old model scores are displayed in red and a tolerance of 10% is used. The vertical axis has been stretched to better observe the differences. The right axis provides the number of reference events for the HIT score. The covered period of time is 01/10/201801/10/2021.

D.2.2 Integration of radar data in the hydrological model: Simple de-biasing or complete calibration ?

To take advantage of the spatial information provided by radar data, CPCH and AZC gridded data need to be introduced in the RS hydrological models. Each grid point will be introduced directly in the altitude band to which it belongs. However several strategies can be used to adapt the model to the input data:

- A complete calibration of the model parameters could be realized to reach the best Nash scores possible. This requires an important time as a calibration is needed for each input data: rain gauge stations, CPCH and AZC.

- A simple de-biasing could be introduced more simply. The idea is to calibrate the model only once, for example for the rain-gauge stations, and then to only adapt one parameter when using CPCH or AZC: *coeffI* a multiplicative factor of correction applied to the precipitation input data[a]. The de-biasing of precipitation inputs is also performed during the complete calibration.

Physically, changing the parameters of the model for each type of precipitation input does not make a lot of sense as the physical phenomena do not depend on the precipitation. However all precipitation data, including rain gauge data, are subject to uncertainties and errors. Rain gauges are also available only locally and need to be interpolated on the basins. Changing the physical parameters of the model can help to artificially remove some of these errors. Notably, the calibration can reduce seasonal bias by putting more weight on certain seasonal phenomena than on others (*e.g.*, snow melt, glacier melt, intense summer precipitation, etc.). Such a seasonal bias has notably been observed for AZC (Figure 3.11). The comparison between simple de-biasing and complete calibration is in reality a comparison between different approaches: a physical approach where the physics is similar in all model and a mathematical approach where the goal is only to optimize some indicators.

---

[a]A second parameter, the altimetric coefficient of precipitation *gradI*, must also be modified and put to 0 as radar data are provided at ground surface. It is used, for rain-gauge input data, to interpolate the precipitation in altitude where the intensity is increased. Such feature is probably lacking in radar data which may explain the CPCH bias far from stations. It may be needed to introduce it during the co-kriging interpolation process.

To investigate the interest of calibrating the model for each type of input rather than only de-biasing it, several models are set up:

- **A.1.** A reference model using rain gauge stations as inputs and calibrated for it.
- **A.2.** A model using CPCH gridded data as inputs and calibrated for it.
- **A.3.** A model using CPCH gridded data as inputs with a simple de-biasing. Except *coeffI*[b] the other parameters are identical to model A.1.
- **A.4.** A model using AZC gridded data as inputs and calibrated for it.
- **A.5.** A model using AZC gridded data as inputs with a simple de-biasing. Except *coeffI*[b] the other parameters are identical to model A.1.

The multiplicative factor *coeffI* required to de-bias radar data is chosen to respect the yearly cumulated precipitation observed. These data are provided by the the Swiss hydrological atlas website[222]. To do so, a coefficient of 0.9 is applied to rain gauge inputs in model A.1. Meanwhile, a coefficient of 1.4 is applied to both CPCH and AZC models (A.2-5.) to obtain the right amounts of precipitation. It demonstrates that, while CPCH is **nearly unbiased close to rain gauge stations** (Section 3.4.1.2), it is **as biased as AZC far from them**. This coefficient is similar between specifically-calibrated models and de-biased-only models: specifically-calibrated models A.2. and A.4. require an important de-biasing of input data that the calibration is not able to remove alone. The other parameters are provided in Appendix C.1.2.

Table D.2 provides the Nash scores (global and above a threshold) for the four sub-basins. These scores are of course higher for specifically-calibrated models (A.2. and A.4.) than for de-biased-only models (A.3. and A.5.) where no further steps are performed to improve them. However, the differences are small in Trient and in the outlet of the catchment in Collecteur Est.

Table D.2: Nash scores of models A.1. to A.5. in Emosson basin. When the scores of specifically-calibrated models are more than 3% higher than those of de-biased-only models, the last ones are coloured in red.

	La Fouly		Saleina		Trient		Collecteur Est	
$Q >$	0	$3\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$5\,\mathrm{m^3\,s^{-1}}$
A.1.	0.9	0.73	0.91	0.62	0.92	0.68	0.95	0.84
A.2.	0.93	0.78	0.92	0.64	0.91	0.66	0.95	0.83
A.3.	0.88	0.66	0.9	0.54	0.9	0.63	0.94	0.82
A.4.	0.88	0.64	0.9	0.55	0.87	0.47	0.92	0.75
A.5.	0.86	0.58	0.88	0.44	0.87	0.47	0.92	0.75

Concerning the HIT-FAR scores, on which no model has been calibrated, they are displayed in Figure D.3 for Emosson-Collecteur Est and La Fouly. As observed for the precipitation inputs (Figure 3.10), AZC models performances are poorer than rain-gauge and CPCH ones in both sub-basins: the errors on raw radar input data are transferred through the rainfall-runoff model. CPCH models are also performing slightly worse than the rain-gauge model A.1. in Emosson-Collecteur Est, with model A.2. having smaller HIT scores and model A.3. having higher FAR scores.
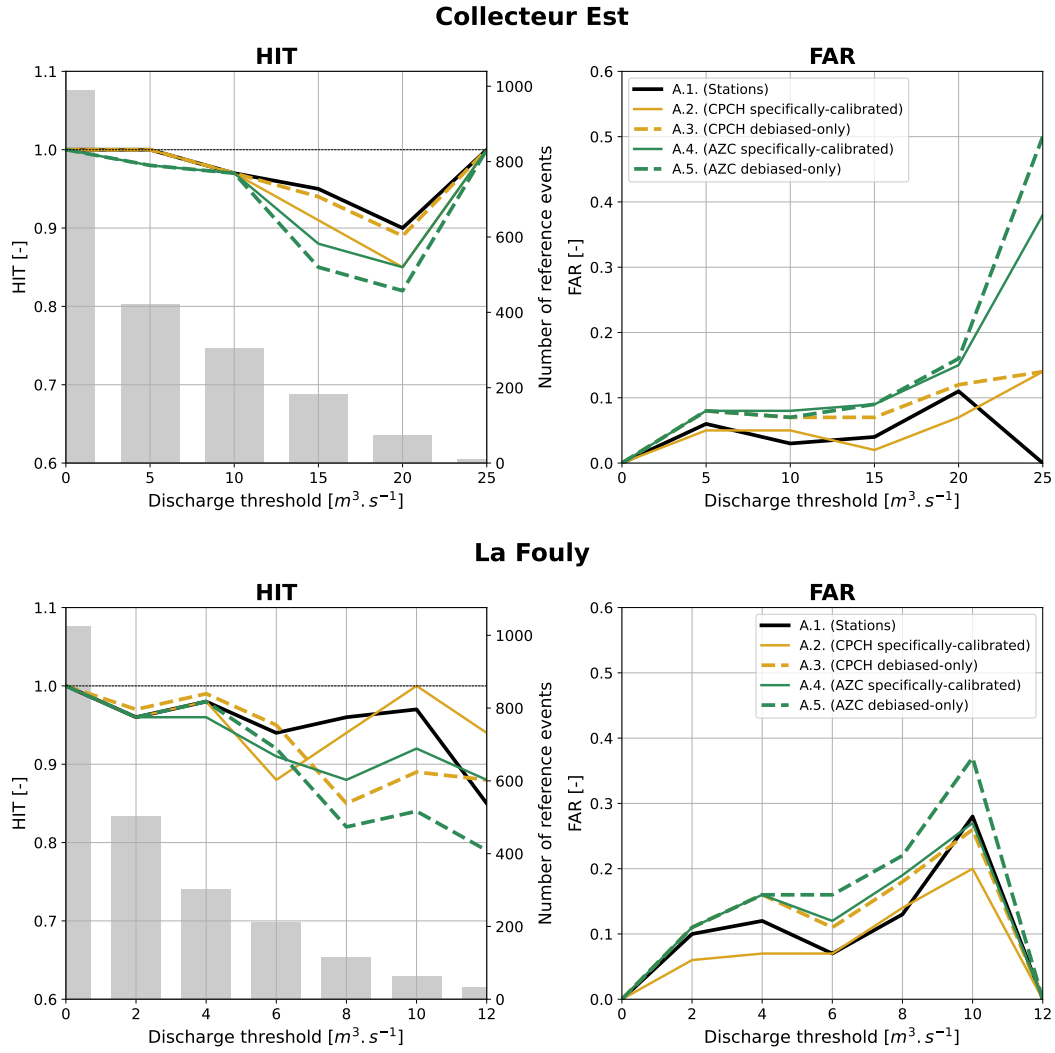
---

[b]and *gradI*

Figure D.3: Daily discharge HIT/FAR values of models A.1. to A.5. for Collecteur Est (top) and La Fouly (bottom). The vertical axes have been stretched to better observe the differences. The right axis provides the number of reference events for the HIT score. Solid lines represent specifically-calibrated models and dashed lines de-biased-only models.

For AZC, the calibration process seems to bring small improvements of few percent at high discharges both in HIT and FAR and for both catchments. The only bigger difference is in Emosson-Collecteur Est in HIT where differences reach nearly 10%. No differences are observed at small discharges. For CPCH however, in Emosson-Collecteur Est, though FAR scores are higher for the de-biased-only model A.3., the latter also as better HIT scores. It hence seems to perform as well as the specifically-calibrated model A.2. with a slightly different trade-off between HIT and FAR. Its performances are nevertheless poorer in La Fouly where the calibration appears to have improved performances by 5 to 10% at high discharges.

Although a complete specific calibration seems to **improve by few percent** the performances of radar-input models, it is **time-costly** and does probably not worth to invest time in it. **A simpler de-biasing procedure provides already good performances** with HIT scores above 0.85-0.9 even at high discharges and FAR scores below 0.2. **The best way to improve the model is thus probably more to integrate new features in the models for rain-gauge inputs and then to de-bias it for radar data.**

D.2.3 Best model

An improvement of the rain gauge model is developed with some new features. A dependence of the snow and glacier melt with radiation – and not only temperature – is first introduced in the model to improve the seasonal shape, notably between the end of July and October. As the peaks of temperature and radiation occur at different times, a new degree of freedom is available to model the melt of glaciers. The characteristics of the slow reservoir of the glacier are also modified. Finally, transfer coefficients of the glacier and the snow models are tuned to better model the diurnal variations of discharge.

Impermeable areas are then introduced in the SWMM object (*Storm Water Management Model*). Depending of the sub-basins, between 0% and 20% of the total area is considered as impermeable area. On this area water will not infiltrate in the ground but will directly be returned as run-off. The fraction of impermeable area *Imp* is a parameter which does not affect the general behaviour of discharge but which increases the peak discharge during rainfall events. An example of response to an event in La Fouly sub-basin with and without impermeable area is proposed in Figure D.4. **The addition of impermeable area is able to increase the peak discharge during precipitation events**, allowing to be closer to the measured peak discharge (in red). The shape of simulated events is also sharper though not enough compared to the reference. However, as observed on the 7thOctober, the drawback is more false alarm events.
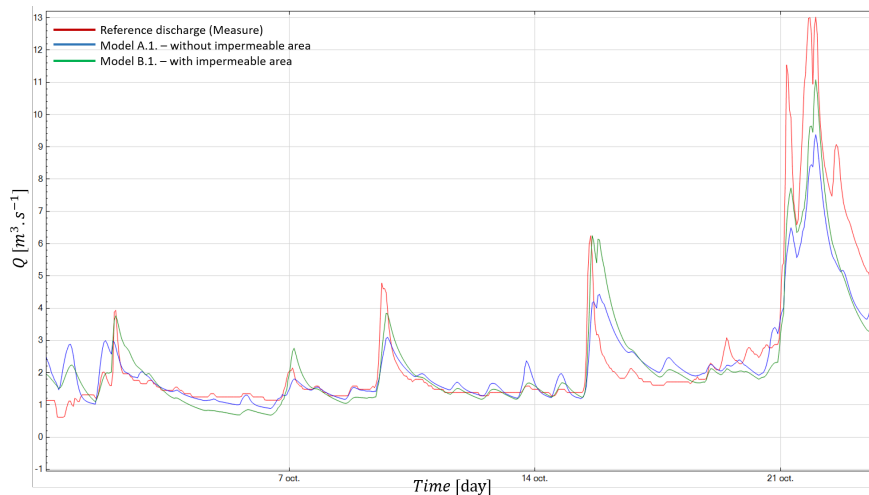

Figure D.4: Five consecutive events of early October 2019 in La Fouly. The red curve is the reference discharge measured at La Fouly, the blue one is from model A.1. with no impermeable areas and the green one is from model B.1. with 10% of the sub-basin modelled as impermeable area.

With these elements, a better model, called **B.1.**, calibrated for the rain-gauge stations is obtained. Its parameters are provided in Appendix C.1.2. A de-biasing is then performed for CPCH (model **B.2.**) and AZC (**B.3.**) with the same *coeffI* factors than in Section D.2.2. Nash values for the three models are proposed in Table D.3. These scores show that, despite focusing on improving the shape and peak discharge of events, the global model for the rain-gauge stations (B.1.) has been improved. Moreover improvements are even larger for CPCH and AZC, demonstrating that the new model is more resilient to changes of input data: **the physics behind the model is probably better now**.

To observe if these global improvements result in boosted flood modelling skills, daily HIT and FAR scores are displayed in Figure D.5. The performances of the rain-gauge model B.1. have been slightly improved in La Fouly while they are similar to model A.1. in Emosson-Collecteur Est[c].

---

[c]The only worsening point is observed for FAR at high discharges (25 m^3 s^{-1} corresponding to 27.5 m^3 s^{-1} with the tolerance – see Section D.1 and Figure D.1.) where 3 new events – underestimated in the previous model A.1. – have just past over the threshold during the summer 2021 in addition to the 4 real events.

Moreover, improvements in the calibration of the rain-gauge model B.1. result mostly in strong and steady improvements of the HIT and FAR scores for de-biased CPCH and AZC models B.2. and B.3. both in Emosson-Collecteur Est and in La Fouly, notably at high discharges. **The de-biasing procedure appears to work better with the new model B.1. De-biasing is now a reliable but simple method**, that allows to obtain similar or better performances with CPCH radar data than with rain-gauge stations data. **These models, and notably the model B.2. for CPCH gridded data, can now be used in re-forecast analysis**.

Table D.3: Nash scores of models B.1. to B.3. in Emosson. Values are coloured in green if higher than in the corresponding models of the previous section (A.1., A.3. and A.5.) and in red if lower.

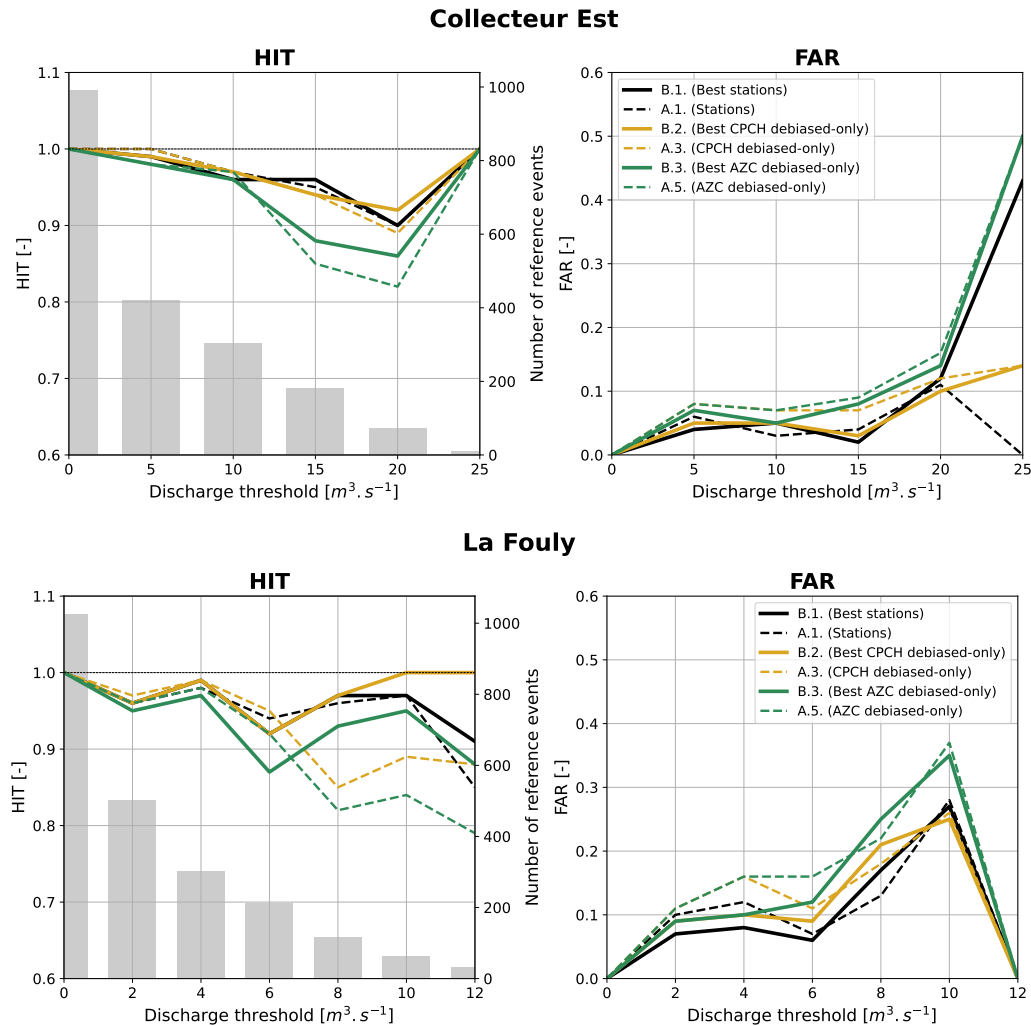|	La Fouly		Saleina		Trient		Collecteur Est	
$Q >$	0	$3\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$5\,\mathrm{m^3\,s^{-1}}$
B.1.	0.91	0.74	0.91	0.62	0.92	0.67	0.95	0.84
B.2.	0.9	0.71	0.9	0.54	0.91	0.61	0.94	0.83
B.3.	0.87	0.63	0.88	0.44	0.88	0.5	0.93	0.77

Figure D.5: Daily discharge HIT/FAR values of models B.1. to B.3. (in solid lines) for Collecteur Est (top) and La Fouly (bottom). The vertical axes have been stretched to better observe the differences. The right axis provides the number of reference events for the HIT score. Dashed lines provide scores of models A.1., A.3. and A.5. as displayed in Figure D.3.

D.2.4   10-minute timestep models

To pave the way for re-forecast analysis with 10-minute timestep forecast data, it is finally necessary to use the previous models with 10-minute data. Radar data were obtained by aggregating over 10 minutes the 5-minute-timestep products provided by *MeteoSwiss*. RS hydrological models are similar to 1-hour timestep models[d]. Three models **B.4.** to **B.6.** are used, corresponding respectively to models B.1. to B.3. Computation is done with a timestep of 10 minutes, but results are saved all hours for comparison purposes with previous simulations. Nash values are displayed in Table D.4 and HIT-FAR scores in Figure D.6.

Globally, while performances are worse in La Fouly both for Nash, HIT and FAR scores, they are better in the smaller catchments of Saleina and Trient and are similar in the outlet in Emosson-Collecteur Est – where both HIT and FAR scores are reduced, resulting in a different trade-off between them. Such behaviour is not unexpected as Trient and Saleina basins being smaller[e], they have also smaller response times and **10-minute data can provide new information**. On the contrary in La Fouly, **10-minute data bring more error and uncertainty than new useful information**. It needs to be noted that a small underestimation bias of up to 4-5% has been observed in the cumulated volumes of CPCH and AZC disaggregated precipitations compared to the original 1h-data. A different (higher) de-biasing may have brought better results notably in La Fouly.

Table D.4: Nash scores of models B.4. to B.6. in Emosson basin. Values are coloured in green if higher than in hourly models (B.1-3.) and in red if lower.

	La Fouly		Saleina		Trient		Collecteur Est	
$Q >$	0	$3\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$1.5\,\mathrm{m^3\,s^{-1}}$	0	$5\,\mathrm{m^3\,s^{-1}}$
B.4.	0.9	0.72	0.91	0.63	0.92	0.68	0.95	0.84
B.5.	0.9	0.71	0.9	0.55	0.91	0.63	0.94	0.82
B.6.	0.87	0.62	0.88	0.45	0.88	0.51	0.93	0.77

---

[d]The only change consists in a diminution of the *ICrGradTHum* parameter. This parameter is a precipitation threshold above which the altimetric gradient of temperature is changed from a dry value to a smaller wet value as temperature increases when it rains. It needs to be diminished as precipitation are more frequently in the dry zone with 10-minute data.

[e]The discharge gauges in Saleina and Trient are only measuring the water inflow from their sub-basin and not the ones from upstream sub-basins. The sum of all sub-basins discharges corresponds approximately to the discharge measured in Emosson-Collecteur Est.

Collecteur Est
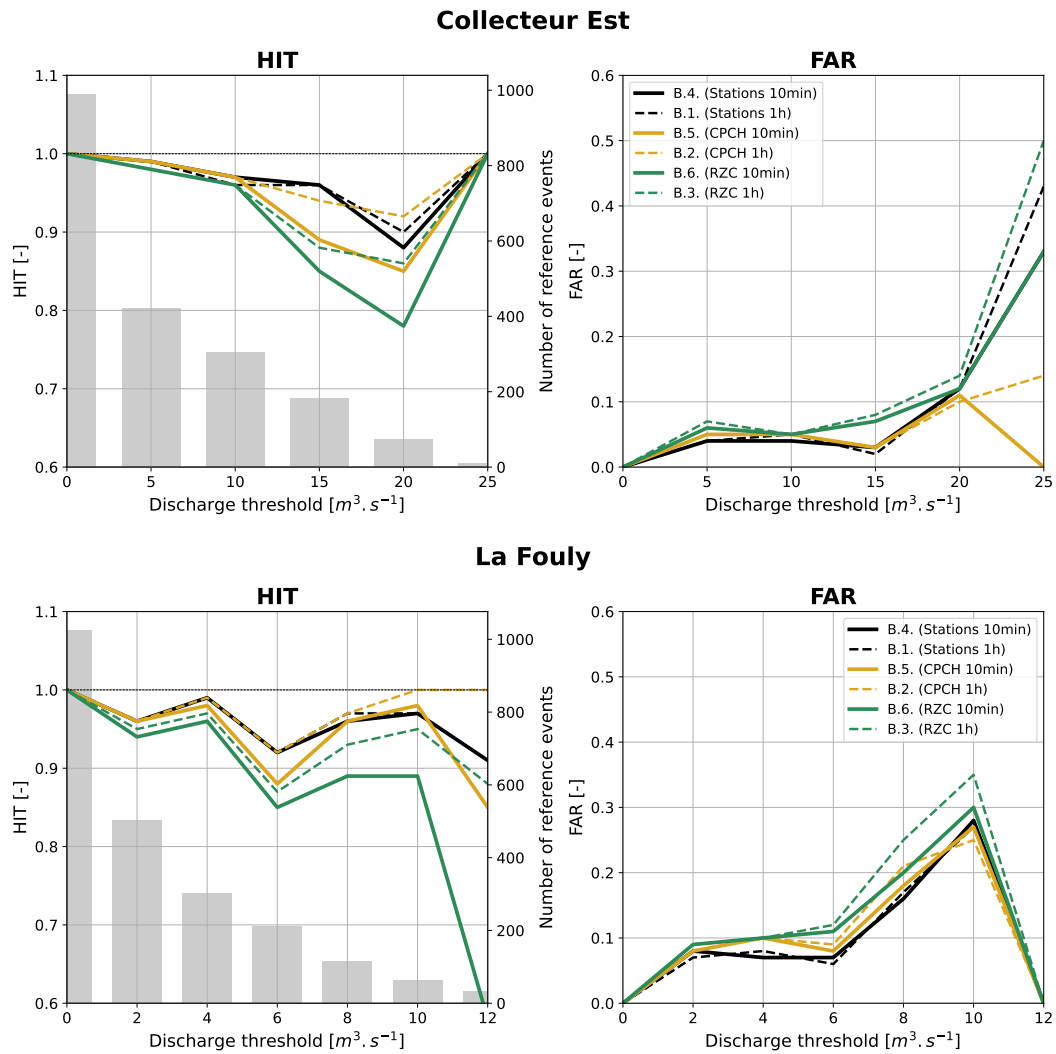

La Fouly

Figure D.6: Daily discharge HIT/FAR values of 10-minute timestep models B.4. to B.6. (in solid lines) for Collecteur Est (top) and La Fouly (bottom). The vertical axes have been stretched to better observe the differences. The right axis provides the number of reference events for the HIT score. Dashed lines provide scores of models B.1. to B.3. as displayed in Figure D.5.

D.3 Hindcast and event analysis

D.3.1 Hindcast process

An update of the model initial conditions is performed based on model B.5. (CPCH with 10-minute timestep) as this model performs as well as the rain-gauge based model. Several runoff forecasts are then produced:

- Three perfect forecasts based on observed precipitation data from rain gauges, CPCH or RZC. They differ from continuous simulations by the use of the *update* step before the simulation. These forecasts are only available in re-forecast mode (and not in operational mode). They are called perfect forecasts because they use perfect QPFs (*i.e.* observations).
- A semi-perfect forecast where precipitation data are from INCA QPF and temperature and radiation are based on observations. This forecast is only available in re-forecast mode.
- The INCA forecast based on INCA precipitation and COSMO temperature and radiation.
- The COSMO forecast based on COSMOs NWPs.
- Filtered forecasts based on INCA and COSMO previous forecasts and filtered with splines to connect runoff observations and forecasts.

The analysis will be based on events observation, alert HIT-FAR scores and 360-minute cumulated volume indicators, except for the effect of post-processing where 60-minute cumulated volume indicators will also be studied[f].

D.3.2 Perfect forecasts comparisons

The three perfect forecasts are compared hereafter. HIT and FAR scores for the alerts and for the cumulated volumes over 360 minutes are provided respectively in Figures D.9 and D.10 both in Emosson-Collecteur Est and La Fouly. For alert HIT-FAR scores, they are computed only on the events crossing the discharge threshold: if an event is always higher than the threshold it is not considered. The MAE curves for the cumulated volumes over 360 minutes are also provided in Figure D.11. Some events representative of the diversity of events are also displayed in Figures D.7 and D.8 respectively for the Collecteur Est and in La Fouly.

The differences between CPCH and rain-gauge perfect forecasts are small in the events. The same events are usually well-predicted or badly-predicted (*e.g.*, first phase of 28/10/2018 event and 27/08/2019) in both perfect forecasts. However, the CPCH perfect forecast is usually more reactive in the first timestep, as it can be observed in the event of the 28/10/2018 and in the three presented events of 2020. This is also observed in La Fouly. In Emosson-Collecteur Est, this behaviour results in slightly higher alert HIT scores while FAR scores are similar. However, nearly no differences can be observed in the cumulated volumes HIT-FAR scores while the mean absolute error of the CPCH cumulated volumes over 6 hours is slightly worse than for rain gauge inputs in June and July. Hence **at very short lead time (less than 90 minutes), *CombiPrecip* seems to improve the forecast** in the Collecteur Est while for larger lead time rain gauges are apparently better. **However differences are tiny**. The alert HIT scores for all products are also significantly low, usually below 50%. However while applying a 10% tolerance on the discharge level, HIT scores are largely improved which demonstrates that events are probably a little bit underestimated /seen too late but not completely missed.

In La Fouly, while cumulated volume HIT-FAR scores do not show clear differences between models, CPCH seems to be slightly less reliable than rain gauges in alert. The MAE curve is also displaying worse CPCH performances than rain gauges in autumn (October and November). This is probably due to the underestimation of CPCH in winter compared to rain gauges.

Finally, the RZC perfect forecast seems not to perform as well as the other two. Indeed it is often too reactive (22/07/2020) or not reactive enough (20/10/2019, 28/08/2020 and 02/10/2020

---

[f]60-minute volume indicators do not show any differences between forecasts except with post-processing[221].

for example). This brings higher alert FAR scores and worse cumulated volumes prediction in Emosson-Collecteur Est (significant in the MAE curve). Raw radar QPF is probably not a tool to use alone in operational forecast as its perfect forecast is already quite unreliable. However **it could be interesting in a multi-model alarm system as it provides sometimes alerts before CPCH and rain-gauge perfect forecasts**. In this project raw radar QPF is nevertheless not available and the analysis will thus not be further followed. In La Fouly, surprisingly, RZC cumulated volume prediction performances appear to be better, with a lower MAE notably in the end of summer and in autumn. It is difficult to interpret this observation but it is perhaps due to the high errors of rain gauges in altitude (due to the wind and the snow)[85].

(a) 28/10/2018 (b) 27/08/2019 (c) 20/10/2019

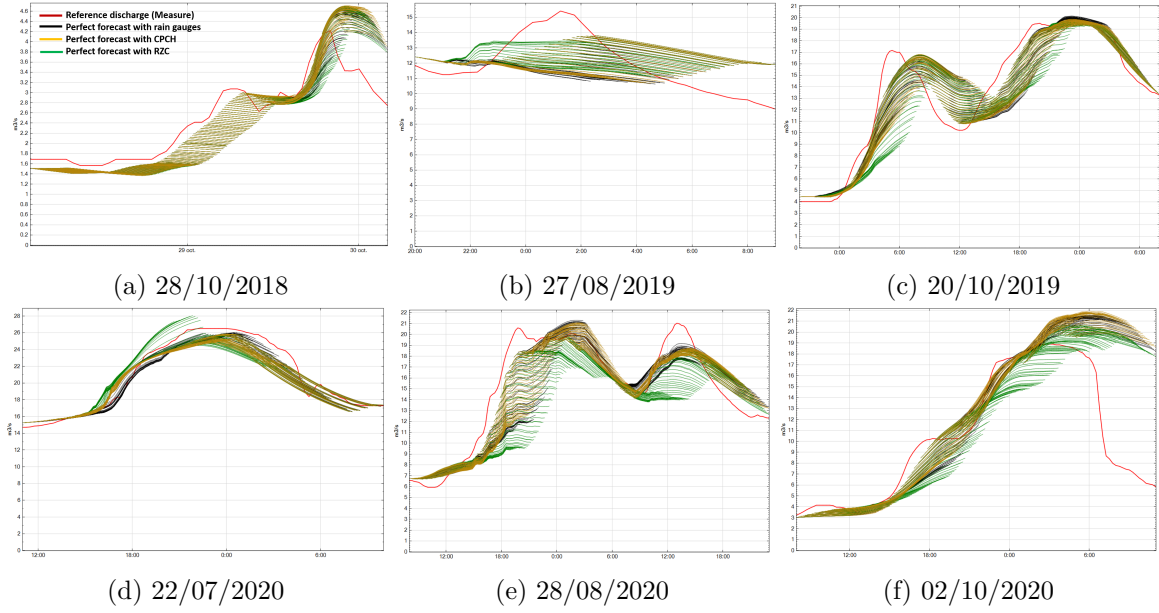(d) 22/07/2020 (e) 28/08/2020 (f) 02/10/2020

Figure D.7: Perfect forecasts for six events in Emosson-Collecteur Est. The measure is in red and three perfect forecasts are proposed: rain-gauge (black), RZC (green), CPCH (gold).

(a) 28/10/2018 (b) 27/08/2019 (c) 20/10/2019

(d) 22/07/2020 (e) 28/08/2020 (f) 02/10/2020
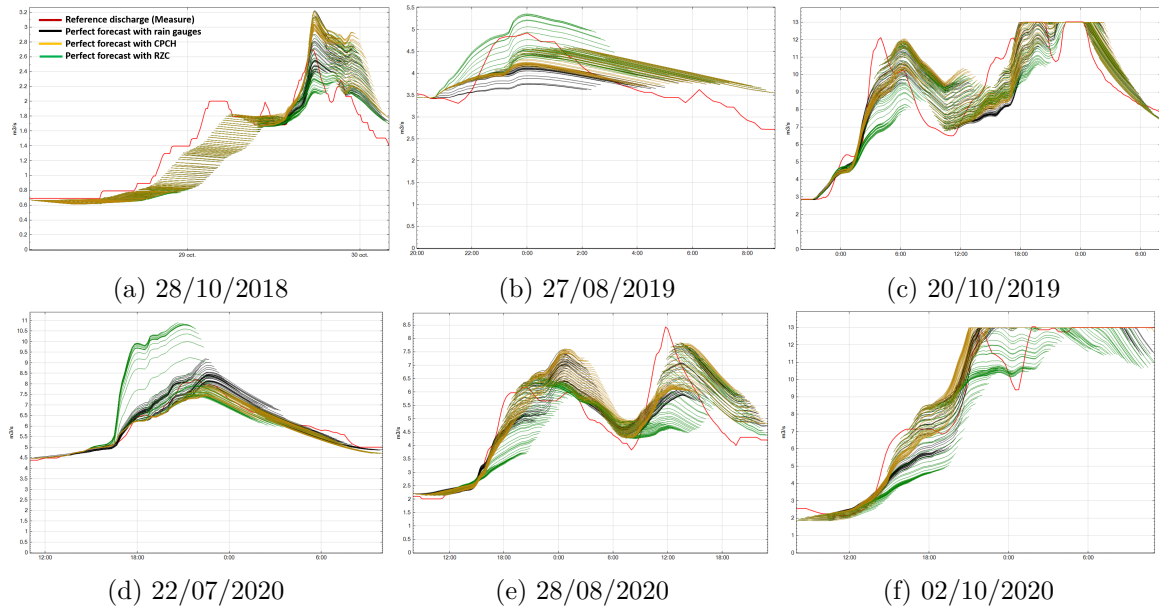
Figure D.8: Perfect forecasts for six events in La Fouly. The measure is in red and three perfect forecasts are proposed: rain-gauge (black), RZC (green) and CPCH (gold).

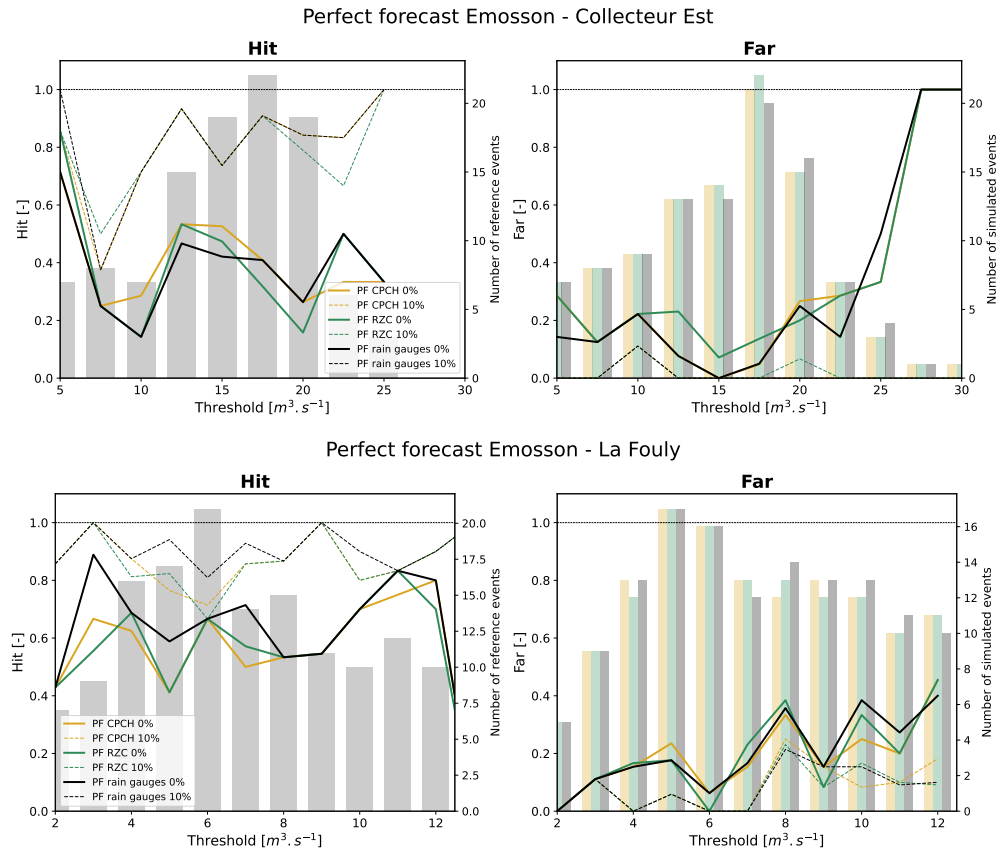Figure D.9: Alert HIT and FAR scores of three perfect forecasts at Collecteur Est and La Fouly. A tolerance of 10% (dashed lines) or no tolerance (solid) is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.
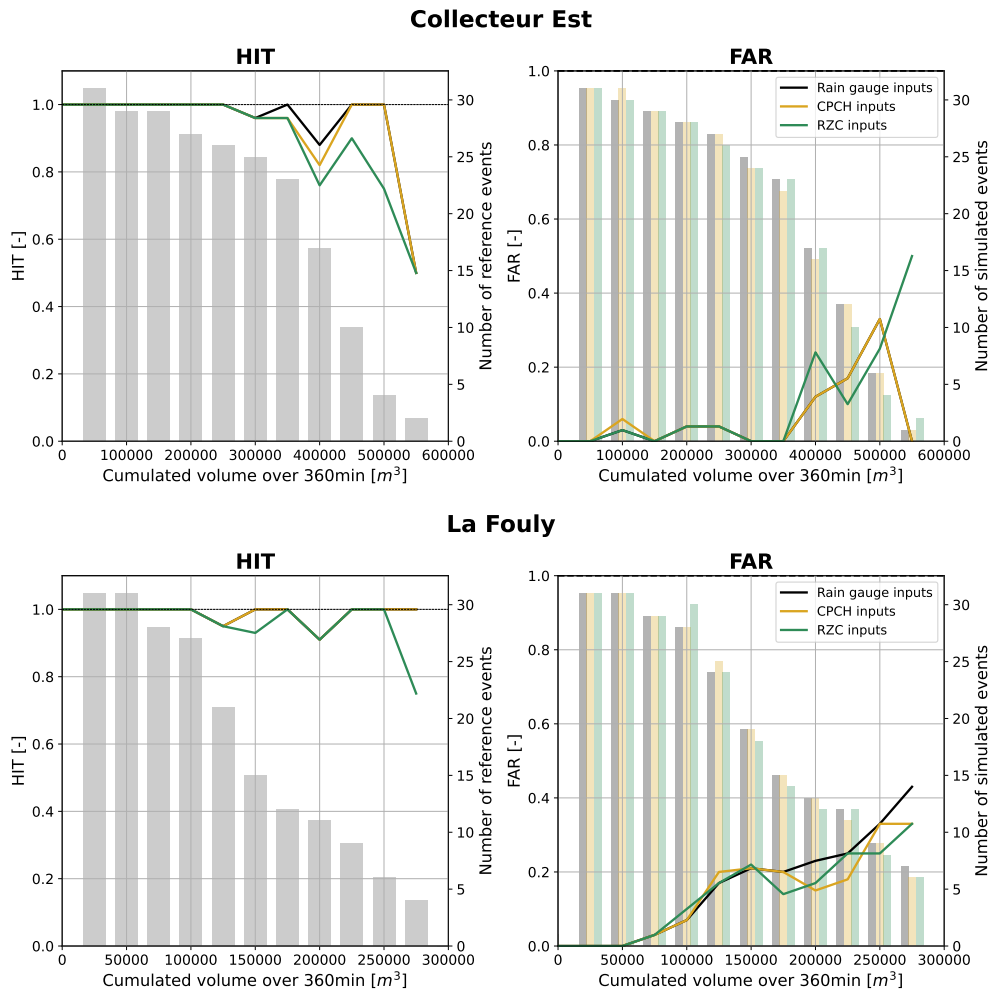
Figure D.10: HIT-FAR scores for the cumulated volume over 360 minutes of the three perfect forecasts at Collecteur Est and La Fouly. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.
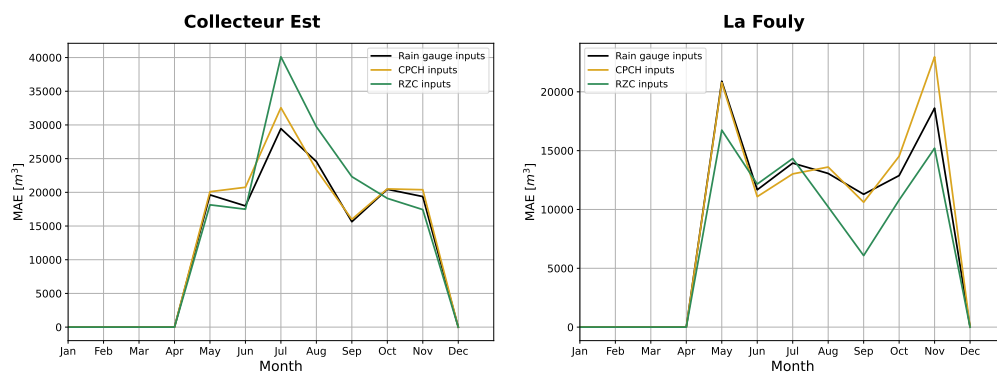

Figure D.11: MAE of cumulated volumes over 360 minutes in Collecteur Est and La Fouly.

D.3.3 INCA forecast

INCA forecast is now used to predict precipitations. A semi-perfect forecast with measured temperatures and predicted precipitations (in blue) and a true forecast with temperature predictions provided by COSMO1-E (in green) are simulated. Results for a selection of events are provided in Figures D.12 for Collecteur Est and D.13 for La Fouly. Alert HIT-FAR scores and 360min-cumulated volumes indicators are also provided by Figures D.14, D.15 and D.16.

(a) 28/10/2018 (b) 27/08/2019 (c) 20/10/2019

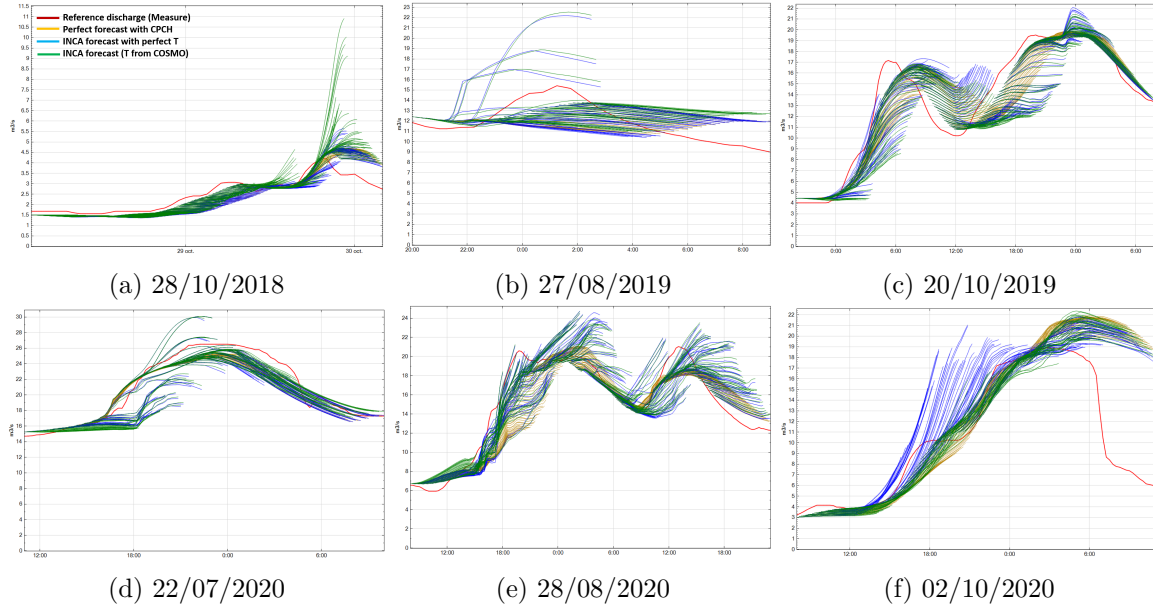(d) 22/07/2020 (e) 28/08/2020 (f) 02/10/2020

Figure D.12: CPCH perfect forecast (gold), INCA semi-perfect forecast (blue) and INCA forecast (green) for six events in Emosson-Collecteur Est. The measure is in red.
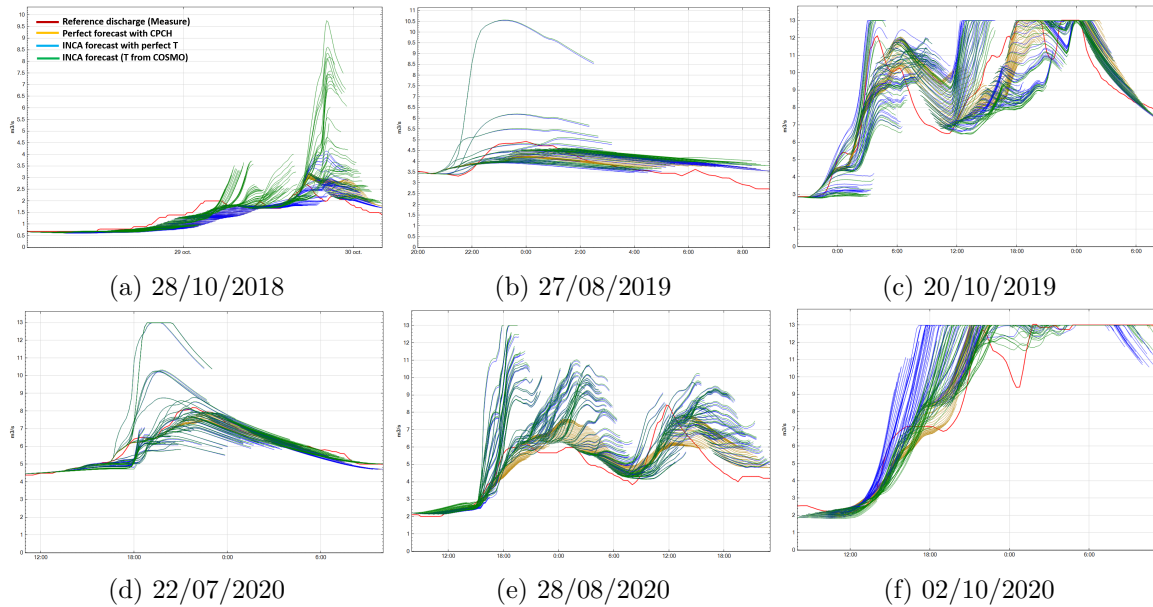

(a) 28/10/2018 (b) 27/08/2019 (c) 20/10/2019

(d) 22/07/2020 (e) 28/08/2020 (f) 02/10/2020

Figure D.13: CPCH and INCA forecasts for six events in La Fouly.

Between the perfect *CombiPrecip* forecast (in gold) and the semi-perfect forecast where precipitation are based on INCA forecast (in blue), **there is a clear loss of performances** with more disperse forecasts and some missed events (*e.g.*, the second part of the event of 20/10/2019). **The tendency leans globally more towards false alarms**. In La Fouly, similar patterns are observed but the INCA semi-forecast is also much more reactive, being able to double its value in just one hour. **It seems that INCA data help to catch the surges of discharges in this small catchment** – which is supported by the higher alert HIT scores. **The drawback is more false alarm**, as confirmed by the alert and cumulated volume FAR scores. In the Collecteur Est, differences are much smaller but still noticeable notably in the 360 minutes-cumulated volumes scores. These differences between La Fouly and Collecteur Est can be explained by two aspects:

- **The size of the catchment**. INCA errors are more easily integrated in a large catchment where they can compensate. For example, if a convective cell is forecasted in the wrong place, it is probable that it will still be located in the catchment area in a large basin. However, in a small basin it can quickly be wrongly outside/inside the catchment which will largely affect performances, and notably create false alarms.

- **The nature of the measured discharge**. In La Fouly, the discharge is measured **after a diversion** with a fixed maximum discharge of $13 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$. To maximize the performances of the continuous model, it is probable that the calibration made in La Fouly has been rather overestimated than underestimated. Indeed, as simulated discharges are then cut at a threshold value of $13 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$, and as the reference discharge is in general closer to this value than to 0 (due to the glacial seasonal behaviour), slightly overestimating the discharge leads usually to better scores than the contrary. This was nevertheless not made consciously.

The comparison between the semi-perfect forecast and the real forecast, where temperatures are taken from COSMO1-E, enlightens the role of temperature in the rainfall-runoff model. **Forecasts for summer events are only slightly affected by the use of COSMO1-E temperature predictions. On the contrary spring and autumn forecasts are largely affected by it**. This shows the strong model dependency to temperature in these periods where the rain-snow limit is determinant for the discharge estimation. Summer discharges are less affected as the rain-snow limit is high enough and the only impacting factor is the snow reservoir level which only depends on observed temperatures. Concerning the direction of the changes, the real INCA forecast is sometimes less reactive (02/10/2020) and sometimes more reactive (28/10/2018) than the semi-perfect forecast. The uncertainty on temperature is thus random and not systematic, which is further confirmed by the alert HIT-FAR scores which do not display clear differences between the semi-perfect and the real INCA forecasts. Concerning the volumes cumulated over 360 minutes, they are slightly better predicted by the complete INCA forecast with notably smaller MAE at both stations in autumn and spring.
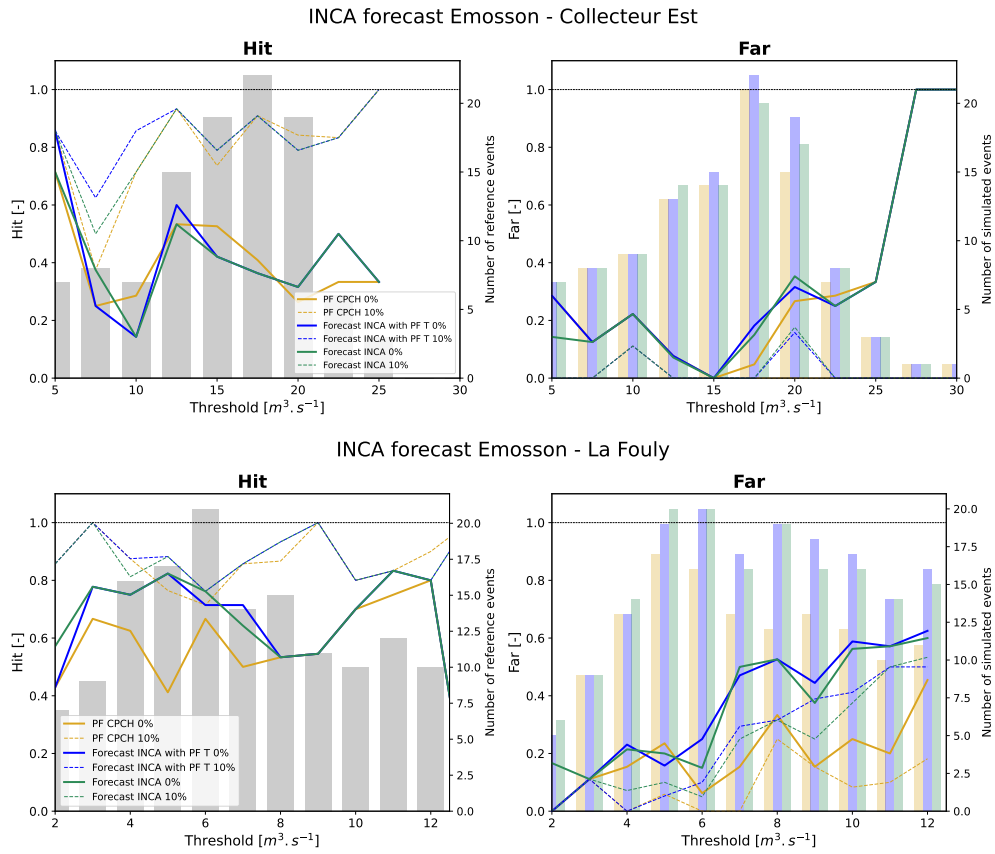
Figure D.14: Alert HIT and FAR scores of CPCH perfect forecast (gold), INCA semi-perfect forecast (blue) and INCA forecast (green) at Collecteur Est and La Fouly. A tolerance of 10% is applied on dashed lines while no tolerance is used for solid lines. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

Collecteur Est

Figure D.15: HIT-FAR scores for the cumulated volume over 360 minutes of the three previous forecasts at Collecteur Est and La Fouly. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

Figure D.16: MAE of cumulated volumes over 360 minutes in Collecteur Est and La Fouly.

D.3.4   INCA *vs* COSMO

Figures D.17 and D.18 display comparisons between INCA and COSMO forecasts for a selection of events respectively in Collecteur Est and La Fo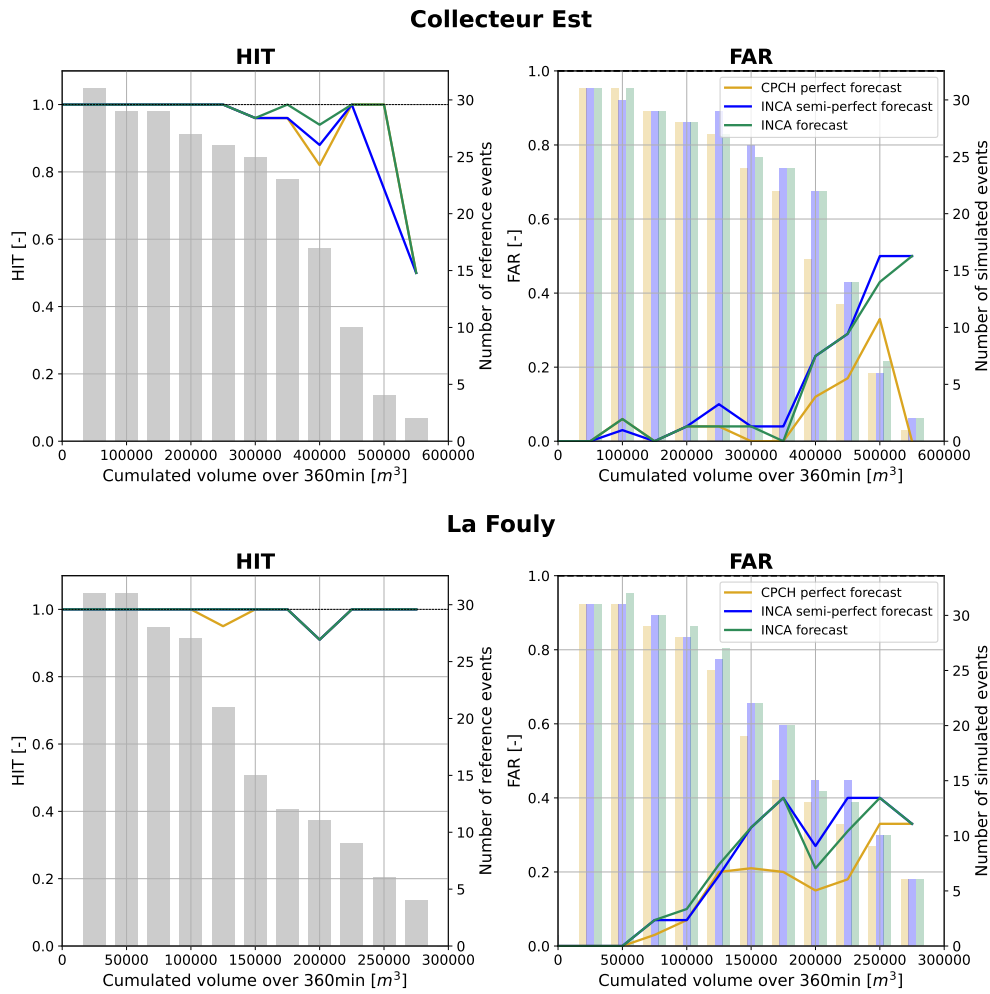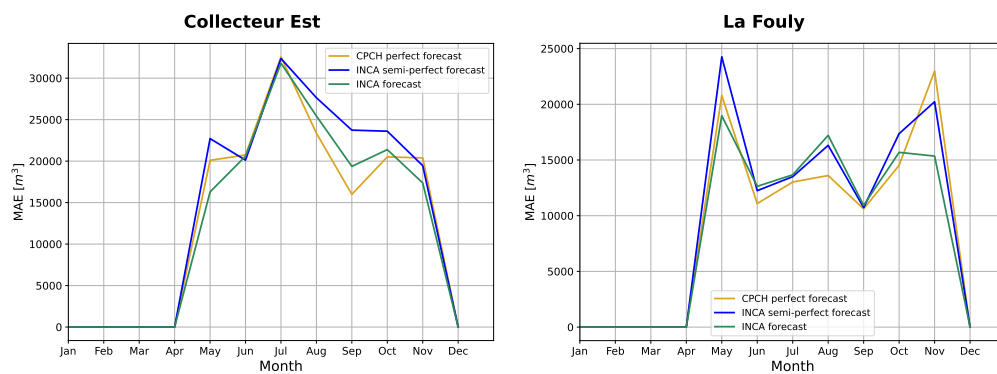uly. Alert HIT-FAR scores are provided in Figure D.19 for the two forecasts (INCA in green, COSMO in black, the other two curves are the filtered forecasts developed in the next Secion D.3.5). 360min-cumulated volume indicators are finally illustrated by Figures D.20 and D.21.

In the events, differences between INCA and COSMO are globally small but two trends can be underlined. First, **INCA is usually more reactive than COSMO**. The growths of discharge are usually observed sooner (*e.g.*, during the 20/10/2019 and 28/08/2020 events). This confirms the similar observations made between CPCH and rain gauges perfect forecasts. Secondly, **INCA is more susceptible to produce false alarms** with the striking example of the 28/10/2018 event. As in the previous section, this behaviour is amplified in La Fouly, with some very good INCA forecasts (*e.g.*, first part of 20/10/2019) and some strong false alarms (*e.g.*, 28/08/2020). These events are much less reactive in Collecteur Est. **This higher reactivity of INCA, notably in La Fouly, is further confirmed by the alert scores**. HIT and FAR scores are higher for INCA than for COSMO in Collecteur Est and, more significantly, in La Fouly. Similarly cumulated volumes over 360 minutes are overestimated in INCA resulting in higher FAR and MAE in La Fouly. In Collecteur Est, this increase in HIT/FAR scores is on the contrary resulting in better MAE.

(a) 28/10/2018   (b) 27/08/2019   (c) 20/10/2019

(d) 22/07/2020   (e) 28/08/2020   (f) 02/10/2020
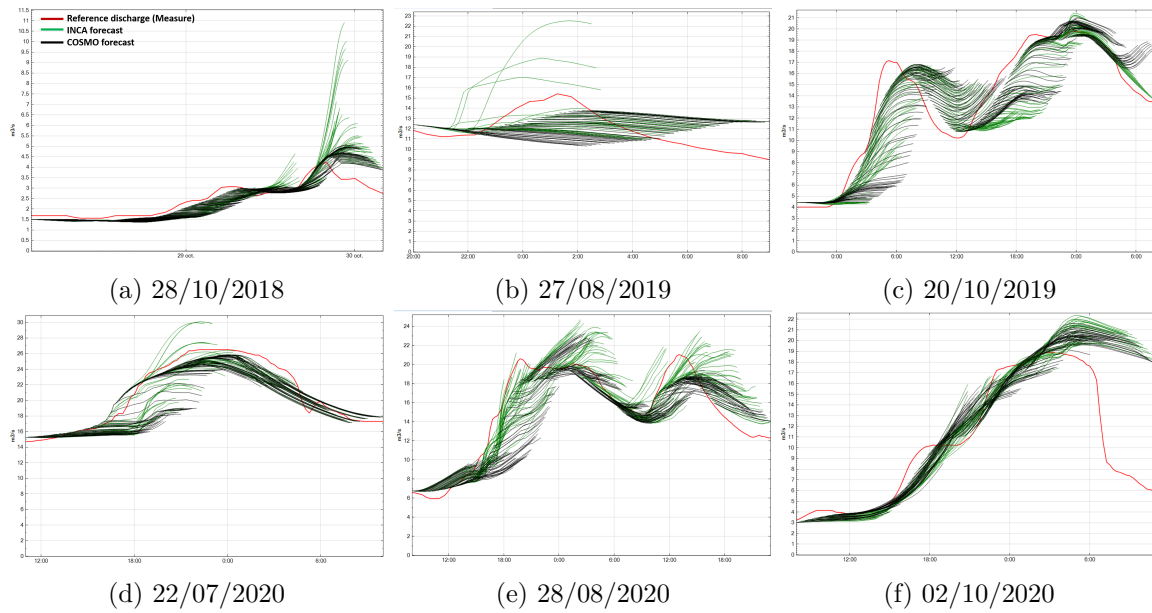
Figure D.17: INCA (green) and COSMO (black) forecasts for six events in Collecteur Est. The measure is in red.
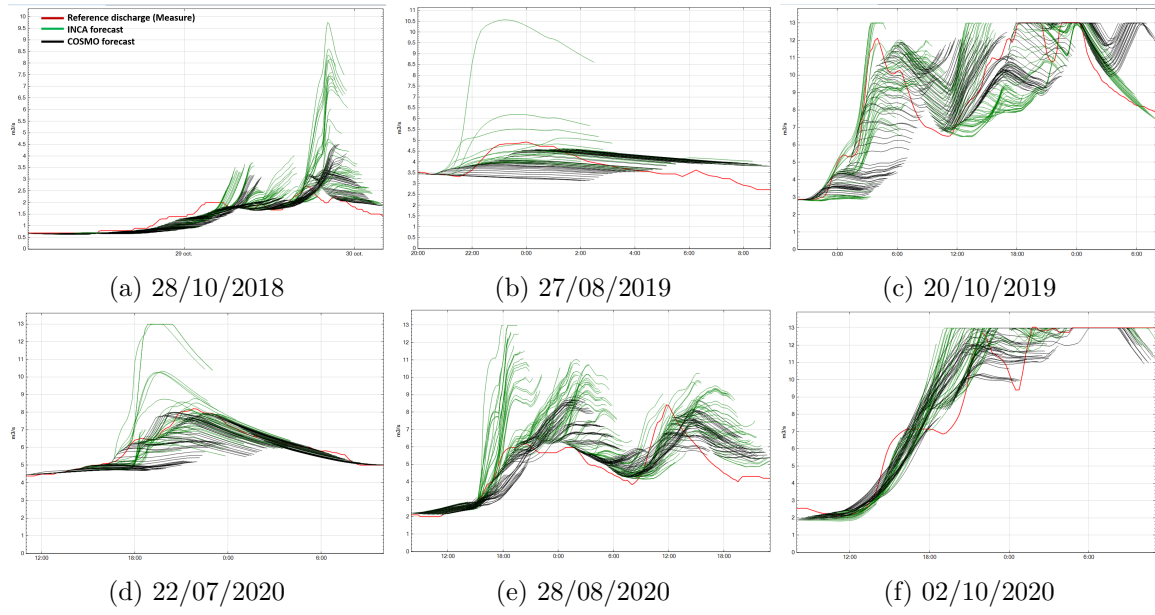
Figure D.18: INCA and COSMO forecasts for six events in La Fouly.
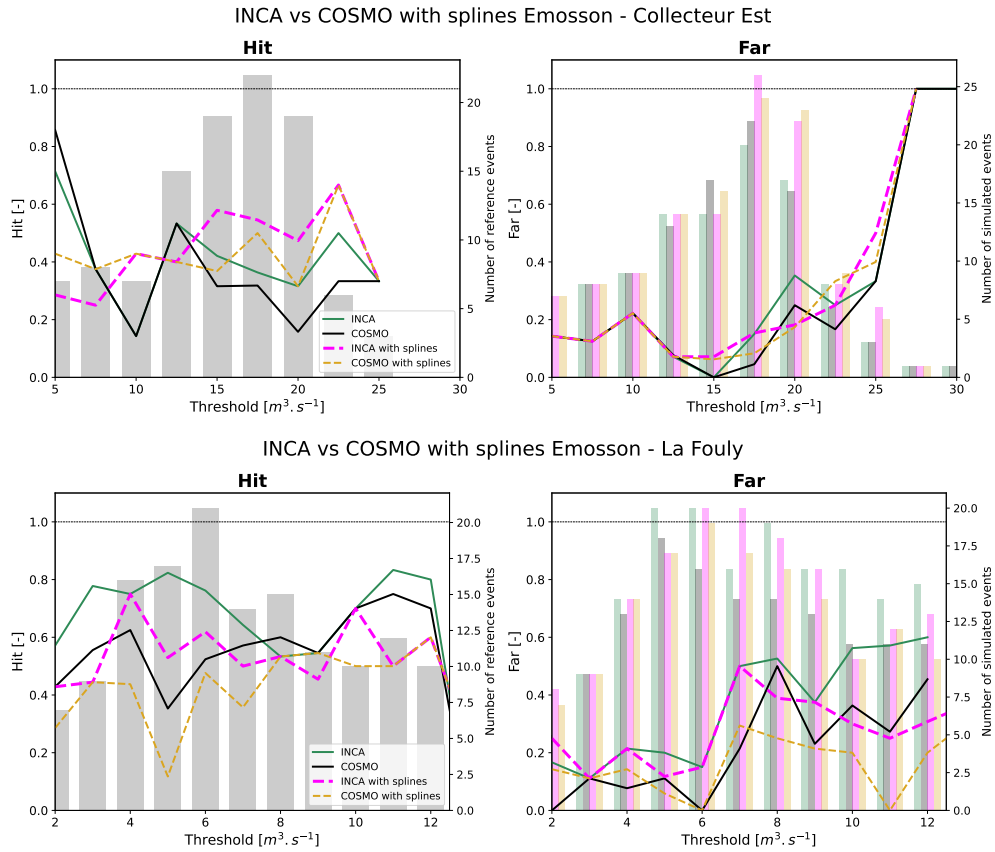
Figure D.19: Alert HIT and FAR scores of INCA (green) and COSMO (black) forecasts at Collecteur Est and La Fouly. The magenta and gold dashed curves are respectively for INCA and COSMO forecasts once filtered (see next section). No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.

Collecteur Est
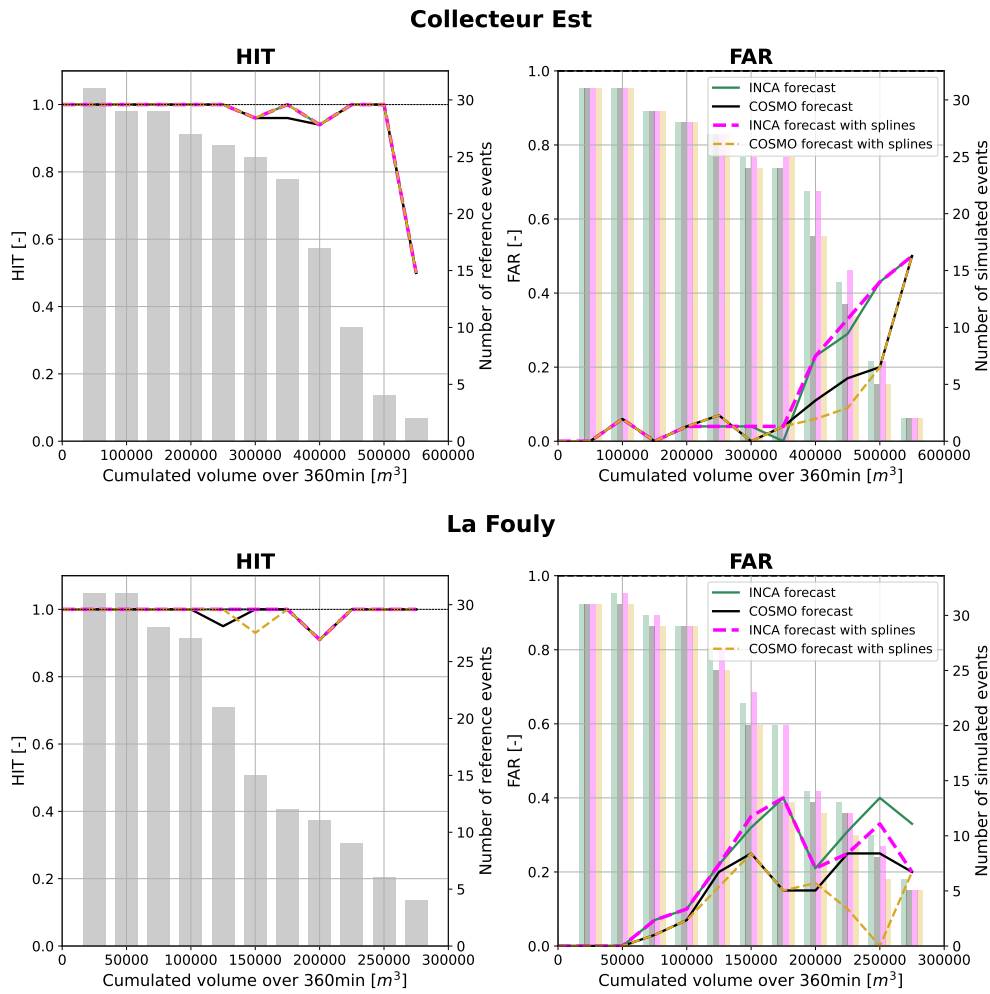
La Fouly

Figure D.20: HIT-FAR scores for the cumulated volume over 360 minutes of the four previous forecasts at Collecteur Est and La Fouly. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.
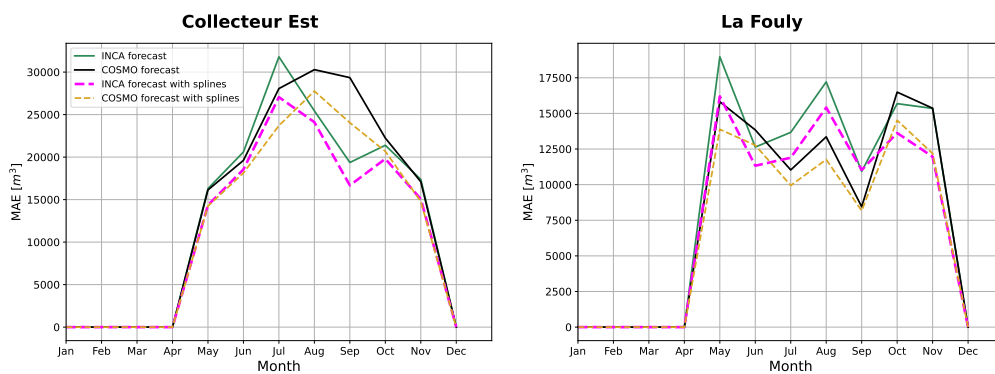
Figure D.21: Mean absolute error of the cumulated volumes over 360 minutes in Collecteur Est and La Fouly.

D.3.5 Effect of post-processing with splines filtering

A post-processing of the INCA and COSMO forecasts can be performed as discussed in Section 3.2.5. It allows to connect the previous forecasts to the latest measured data which usually improves the forecast. In the first intermediate report of the *Radar4Infra* project[221], it was demonstrated that in the case of Emosson the filtering process improved a lot the performances, notably at short lead time. A filter period of 3h is chosen for Emosson. Event analyses in the Collecteur Est and La Fouly are displayed respectively in Figures D.22 and D.23. A comparison between non-filtered and filtered INCA forecasts is also proposed in Collecteur Est in Figure D.24.

Alert and 360min-cumulated volume indicators for non-filtered and filtered INCA and COSMO forecasts were provided in Figures D.19, D.20) and D.21. 60-min cumulated volume HIT-FAR scores (for Collecteur Est) and MAE (for both stations) are finally given respectively by Figures D.25 and D.26.

| (a) 28/10/2018 | (b) 27/08/2019 | (c) 20/10/2019 |
| (d) 22/07/2020 | (e) 28/08/2020 | (f) 02/10/2020 |
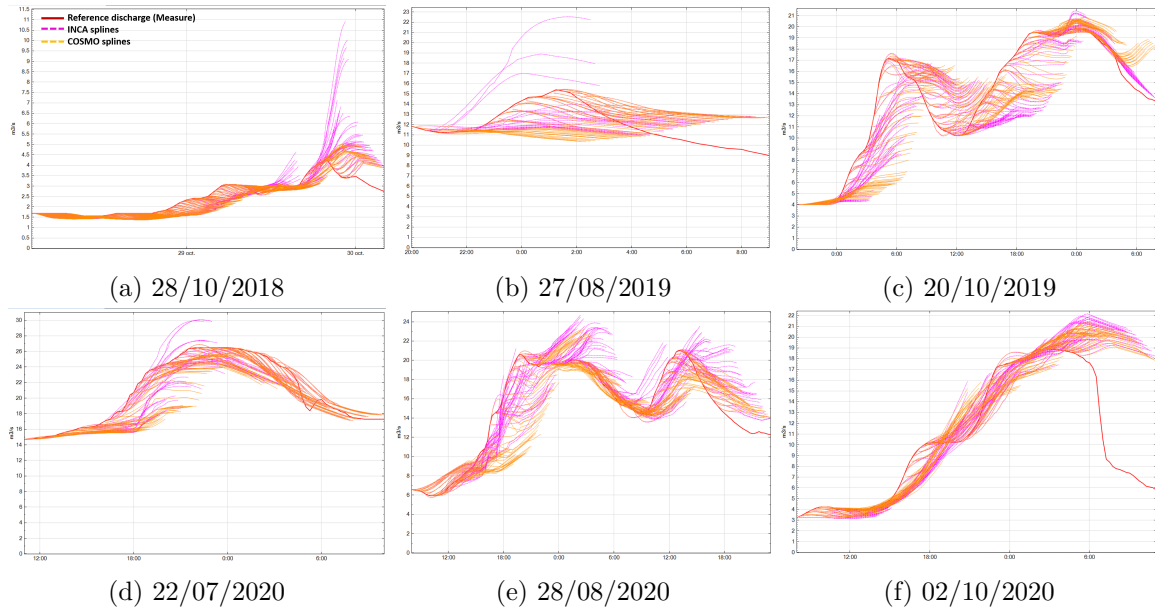
Figure D.22: INCA (magenta) and COSMO (gold) forecasts with spline filters for six events in Collecteur Est. The measure is in red.

(a) 28/10/2018   (b) 27/08/2019   (c) 20/10/2019

(d) 22/07/2020   (e) 28/08/2020   (f) 02/10/2020
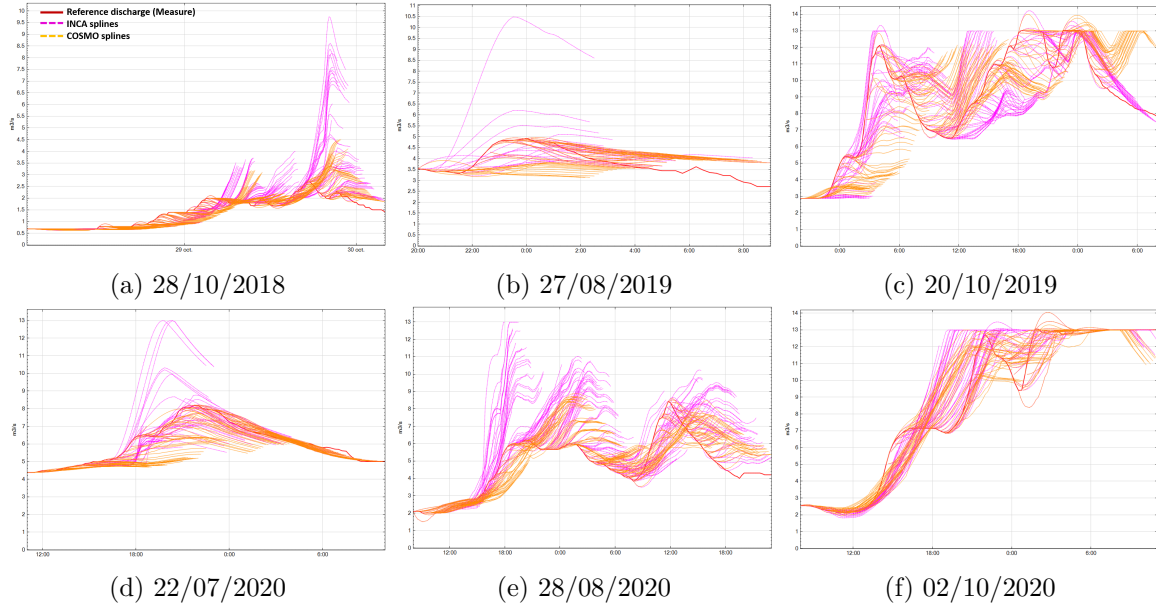
Figure D.23: INCA (magenta) and COSMO (gold) forecasts with spline filters for six events in La Fouly. The measure is in red.

(a) 28/10/2018   (b) 27/08/2019   (c) 20/10/2019

(d) 22/07/2020   (e) 28/08/20   (f) 02/10/2020
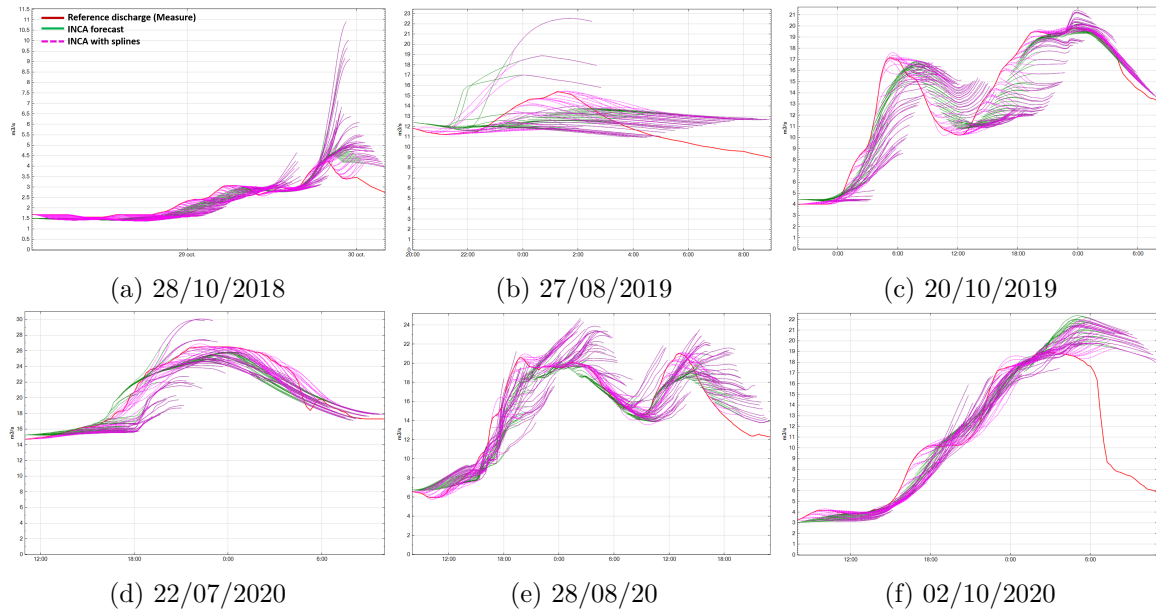
Figure D.24: INCA filtered (magenta) and non-filtered (green) forecasts for six events in Collecteur Est. The measure is in red.

In Collecteur Est, the events are better forecasted when applying splines notably when the forecasts were not reliable (*e.g.*, COSMO the 20/10/2019, both forecasts the 27/08/2019). Nevertheless INCA seems to still be more reactive than COSMO. The filtering step is also not able to remove the false alarms as in the case of the 28/10/2018. Events in La Fouly show similar patterns with, as for pre-processed forecasts, a bigger reactivity of INCA than in Collecteur Est. However, while alert HIT scores are largely improved in Collecteur Est (for both INCA and COSMO), they are worsened in La Fouly. This is probably due to the better pre-filtering performances observed there (around 70% of alert HIT for INCA compared to 40% in Collecteur Est). The filtering step is probably more efficient for big catchments with smaller reactivity than for smaller catchments. Indeed, it does not use the first timesteps of the runoff forecasts where INCA has added very localized information

allowing to catch the discharge surges. **Filtering appears to be useful in the large Emosson catchment with 4-6h response time but detrimental in the smaller basin of La Fouly.**

Concerning the volume analyses, **as observed in the intermediate *Radar4Infra* report, the impact of post-processing at large lead time (360 minutes) is not significant** (nearly no changes in HIT-FAR scores, small diminution of the Mean Absolute Error). **However for a smaller lead time, the impact is huge**. The cumulated volume over 60 minutes provides, for example, perfect HIT scores in Collecteur Est when filtering is applied. Similarly the MAE is more than halved in both catchments. Such feature is quite logical as, unlike the non-filtered forecasts, the filtered forecasts are directly based on the latest observed discharges for the first timesteps. The discharge being a continuous quantity, the last points of observation usually give a first hand good prediction of what will happen.
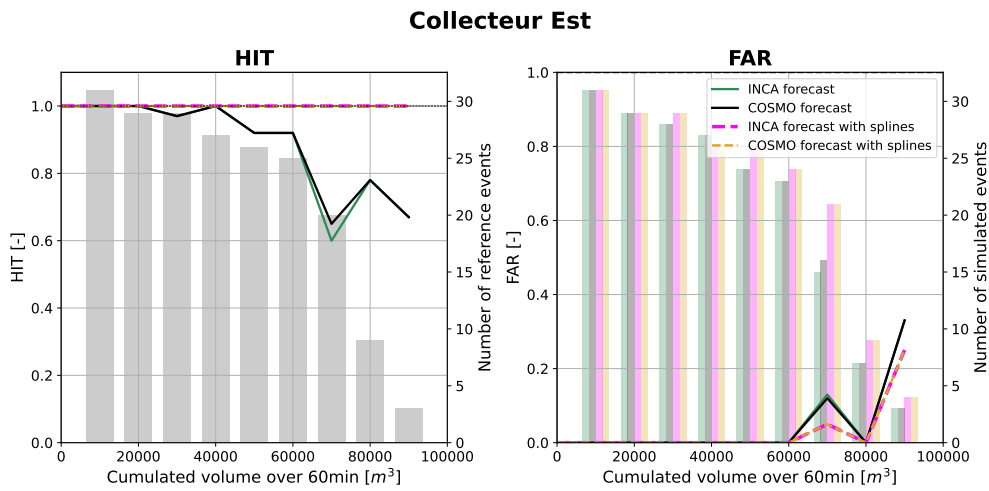

Figure D.25: HIT-FAR scores for the cumulated volume over 60 minutes of INCA and COSMO, filtered and non-filtered forecasts in Collecteur Est. No tolerance is applied. The right axes provide the number of reference (resp. simulated) events for the HIT (resp. FAR) score.
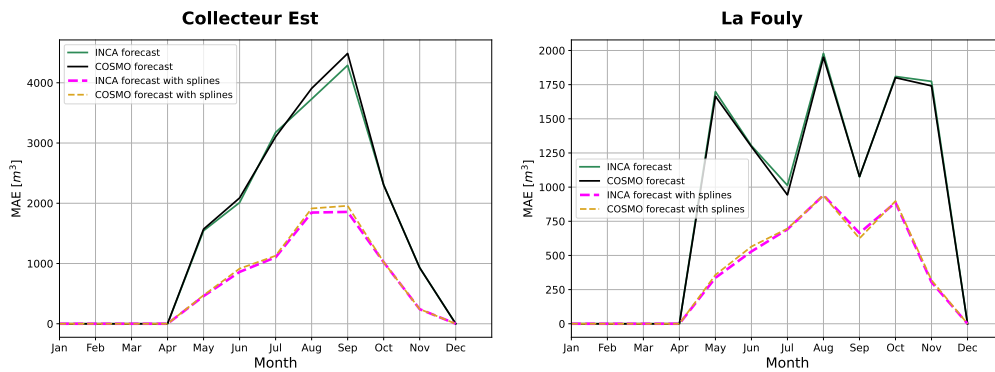

Figure D.26: Mean absolute error of the cumulated volumes over 60 minutes in Collecteur Est and La Fouly.

D.4 Conclusion

The Emosson catchment is a rather complicated catchment where **a variety of different processes** are at stake: glaciers and snow cycles, underground slow or fast storages, direct runoff on impermeable areas during intense precipitations, water interceptions by vegetation or glaciers, interactions of previous phenomena, etc. Its location in altitude and its complex orography constitute other interesting properties suggesting radar data may improve rainfall-runoff models and predictions. Finally, it contains several sub-basins of different scales allowing to investigate the impact of INCA radar forecast on different locations.

This complex framework allows to develop a method to calibrate a *Routing System* model adapted to radar gridded data. **A de-biasing of precipitation input data** is first demonstrated to be essential to obtain good rainfall-runoff models. Though a complete calibration of the model according to the nature of the input precipitations leads to slightly improved results, it has been demonstrated that investing time on the modelling of the physical processes allows to get similar results between rain gauge and *CombiPrecip* data. The final model is thus robust to any change of input type. On a simpler catchment, it is reasonable to think that such a process will still be reliable as a calibration can only be performed on a smaller number of parameter.

In operational mode, several steps need to be performed once the model is set up. First an *update* of the initial conditions has already been proven to be necessary[6,221]. With this step, compared to COSMO, **INCA seems to introduce more reactivity on the prediction of flash events, notably in the small-response time catchment of La Fouly**. The other side of the coin is **increased false alarm rates**. Reflections on the requirements and design of an alarm systems will need to be done to balance HIT and FAR scores. Raw radar forecasts, not available in this project, could also be useful in a multi-model alarm systems thanks to their ability to detect some events not observed in other products. However their overall performances are too low to build an alarm system relying only on them.

The temperature forecast plays an important role in spring and autumn which highlights the necessity of an operational monitoring of the model notably during these periods. Finally, the postprocessing appears to improve the alert performances for intermediate response time catchments (Emosson-Collecteur Est, 4-6h) while being detrimental in smaller basins (La Fouly, 1-2h). Postprocessing will need to be investigated further as it probably needs to be adapted for small basins (smaller filtering period for example).

Cumulated volumes are **nearly only affected by the post-processing step, which largely improves its predictions at small lead times**. However, they are nearly not affected by filtering for larger lead times. When forecasts are needed for cumulated volume predictions rather than for flash-flood alerts, a post-processing of output data will be required. The type of input data (raingauge, CPCH or RZC QPEs or INCA or COSMO QPFs) is nearly not modifying volume scores.

Appendix E

Volume analysis of forecasts in Zurich-Altbach

The analysis of cumulated volumes in Zurich is much less interesting and important than in Emosson as the goal of the *Radar4Infra* project in Zurich-Altbach catchment is to provide alerts for flash flood events. In the Emosson catchment it was an essential feature for the management of reservoir levels. In this appendix, the most interesting results for 180-minute cumulated volumes are presented. Results for 60 minutes are similar but with largely smaller differences between forecasts.

E.1 Perfect forecasts

For the predictions of cumulated volumes over 180 minutes, the perfect forecasts are not very sensitive to the input data (CPCH, RZC, rain gauges, or inca0) as illustrated by the HIT-FAR scores of Figure E.1. CPCH and inca0 are slightly better in HIT while rain gauges have smaller FAR.

Figure E.1: HIT and FAR scores for the cumulated volumes over 180 minutes of the four perfect forecasts at Zurich-Altbach discharge station. No tolerance is applied.

E.2 INCA forecasts

The introduction of the INCA forecast generates a loss of performances for 180-minute volumes compared to the CPCH perfect forecasts (Figure E.2). This loss of performances is only marked by higher FAR. HIT are similar. Concerning COSMO, volume HIT and FAR scores are higher in INCA than in COSMO which demonstrates a tendency to overestimation. Notably the increase of FAR is largely more significant than the increase in HIT. **INCA is not a good QPFs in Zurich for the determination of flash flood volumes**.
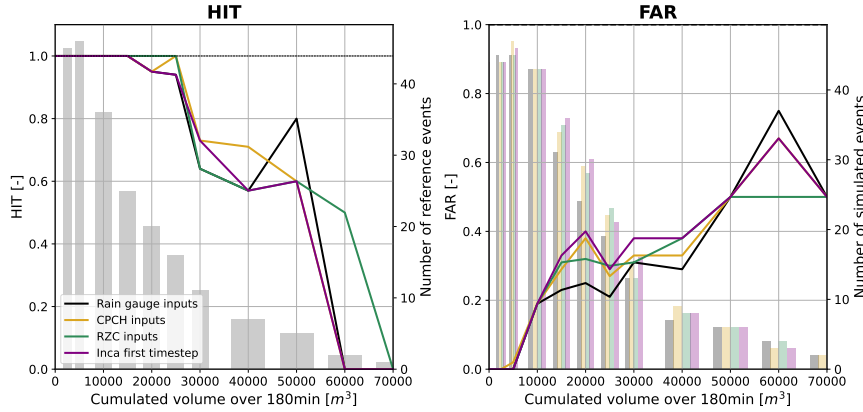
Figure E.2: HIT and FAR scores for the cumulated volumes over 180 minutes of the three previous forecasts at Zurich-Altbach discharge station. No tolerance is applied.

E.3 Effects of the post-processing

Post-processing results in higher volume estimations (more HIT and FAR). It also removes all differences between the INCA and COSMO forecasts: the input data is not important anymore. Overall the Mean Absolute Error (Figure E.4) is slightly reduced for 180 minutes. For a smaller lead time (60 minutes), as observed in Emosson, the post-processing allows to largely improves the estimation. **Hence, due to its high false alarm rate, post-processing is to be banned for 180-minute volume estimation but is interesting for 60-minute volumes.** The last observation is probably due to the fact that post-processing allows to link the forecast to the last observed data.
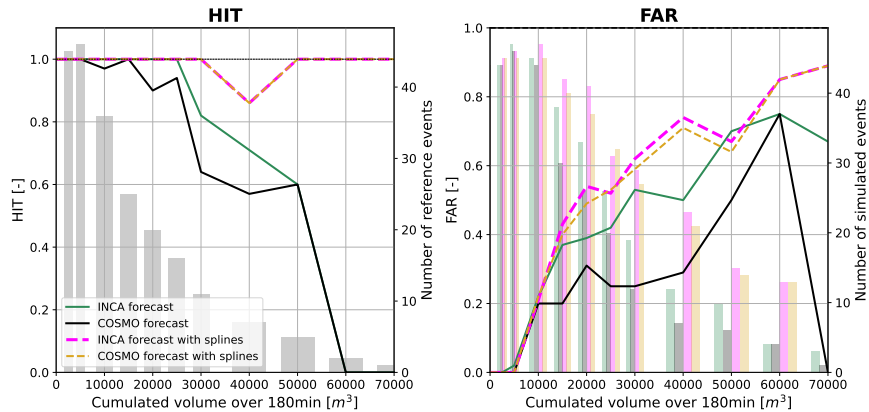
Figure E.3: HIT and FAR scores for the cumulated volumes over 180 minutes of the non-filtered INCA (green) and COSMO (black) forecasts at Zurich-Altbach discharge station. Filtered forecasts scores are also provided in magenta (INCA) and gold (COSMO). No tolerance is applied.
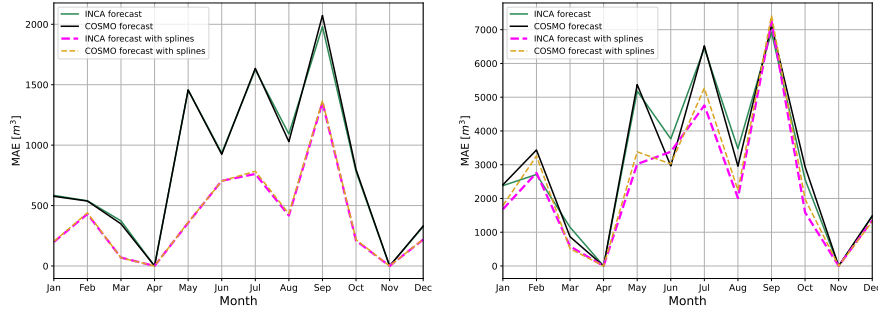
Figure E.4: Mean absolute error of the cumulated volumes over 60 and 180 minutes in Zurich-Altbach.

E.4 Volume estimation with the new models

In this part of the appendix, the cumulated volumes over 60 minutes will be presented as they result in higher differences between the old and new models than the cumulated volumes over 180 minutes. This is most probably due to the additional reactivity introduced in the new model (with variation of Imp and work on the timing of the peaks).

The main observation is that the old model without all the added features was probably better adapted for the estimation of cumulated volumes. Indeed, in Figure E.5, HIT scores are similar in the old and new forecasts but there is much more false alarms in the new models. Hence the new models are not improving the estimation but are only wrongly overestimating the already too high events: **to estimate the cumulated volume over a flash flood event, the old model appears to be more suited.**
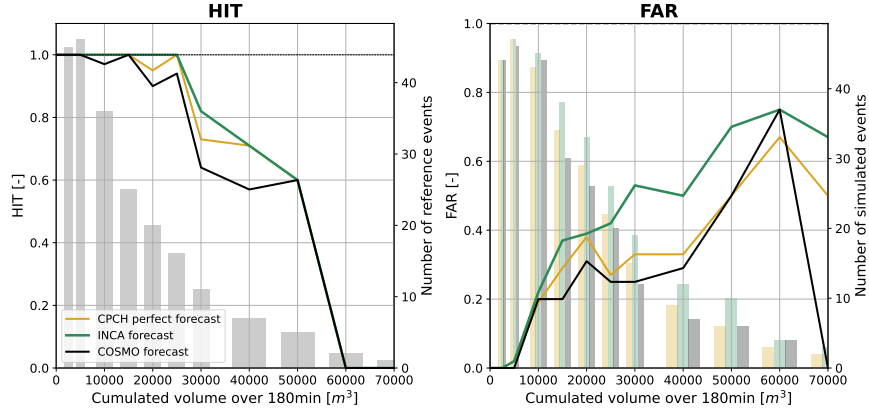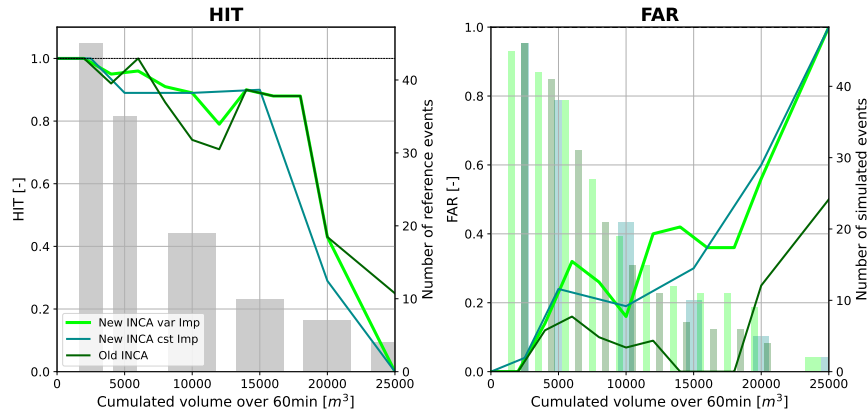

Figure E.5: HIT and FAR scores for the cumulated volumes over 60 minutes for the old and the new INCA forecasts at Zurich-Altbach discharge station. No tolerance is applied.

E.5 *update* step

In the first report of the *Radar4Infra* project[221], the *update* step has been found to be essential notably for the estimation of volumes. Figure E.6 demonstrates that it is still the case in the Zurich catchment case as HIT scores are largely improved when an *update* step is performed, both for 60 and 180-minute volumes. The *update* with distinct correction factors in the lower and upper sub-basins is also providing better results than the one with similar correction factors over the whole catchment.
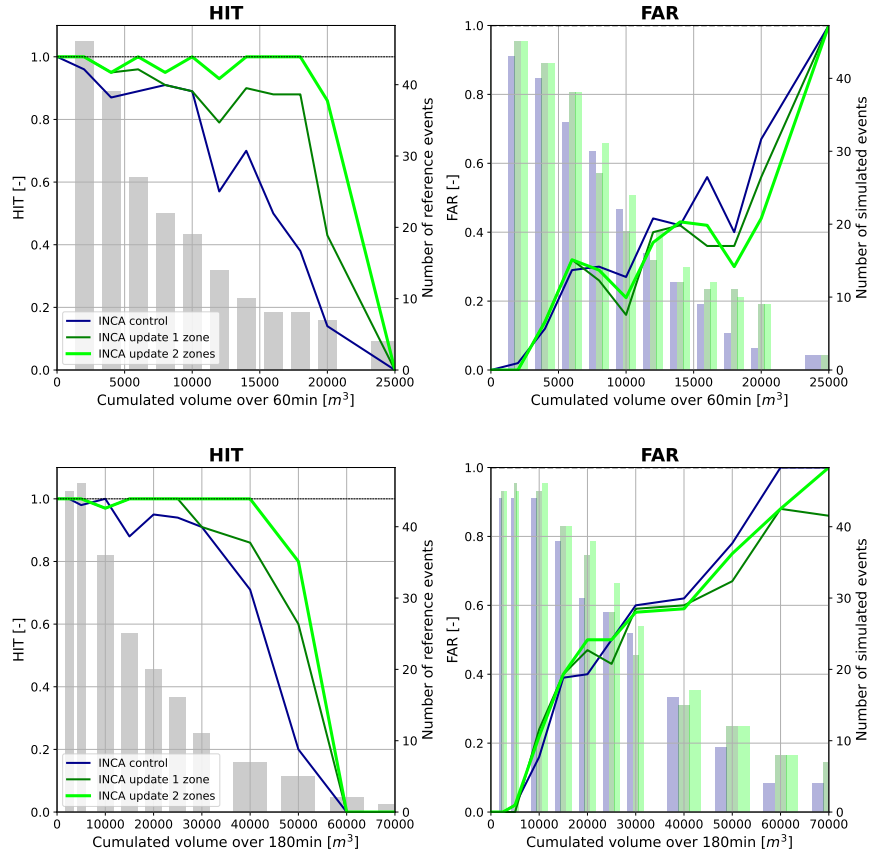
Figure E.6: HIT and FAR scores for the cumulated volumes over 60 (top) and 180 (bottom) minutes for the new INCA forecast with variable *Imp* with or without *update* step at Zurich-Altbach discharge station. No tolerance is applied.

E.6 Overview of results for volume estimation in Zurich

To estimate cumulated volumes, the INCA forecast is probably worse that the COSMO forecast as it tends to generate a high number of false alarms.

An *update* step has been shown to be necessary. Moreover the post-processing step increases the number of false alarms independently of the input data. It may be useful in the estimation of short lead times cumulated volumes.

Finally the new model with higher reactivity (with or without variable *Imp*) worsen the volumes estimation.

Bibliography

1. *2021 Disasters in Numbers - World | ReliefWeb* `https://reliefweb.int/report/world/2021-disasters-numbers` (2022).
2. *Climate Change Widespread, Rapid, and Intensifying – IPCC — IPCC* `https://www.ipcc.ch/2021/08/09/ar6-wg1-20210809-pr/` (2022).
3. *Floods and Landslides in 2021: Wet Summer Resulted in Highest Damage Costs since 2007 - WSL* `https://www.wsl.ch/en/news/2022/06/floods-and-landslides-in-2021-wet-summer-resulted-in-highest-damage-costs-since-2007.html` (2022).
4. *Retour Sur l'orage de Cressier - MétéoSuisse* `https://www.meteosuisse.admin.ch/home/actualite/meteosuisse-blog.subpage.html/fr/data/blogs/2021/6/orages-cressier.html` (2022).
5. *Evaluation de La Crue Du 22 Juin 2021 Dans Le Bassin Versant Du Ruhaut à Cressier (Ct. NE)* Mandat (Scherrer AG, Hydrologie und Hochwasserschutz for Neuchatel canton, Reinach, 2021).
6. *Manual on Flood Forecasting and Warning* (World Meteorological Organization, Geneva, 2011).
7. Moore, R. J., Bell, V. A. & Jones, D. A. Forecasting for Flood Warning. *Comptes Rendus Geoscience* **337,** 203–217 (2005).
8. Hapuarachchi, H. a. P., Wang, Q. J. & Pagano, T. C. A Review of Advances in Flash Flood Forecasting. *Hydrological Processes* **25,** 2771–2784 (2011).
9. Skolnik, M. Radar: From Hertz to the 21st Century. *IEEE Antennas and Propagation Society Newsletter* **30,** 13–18 (1988).
10. *From Weather Observations to Atmospheric and Climate Sciences in Switzerland* (eds Saskia, W. & Furger, M.) (vdf Hochschulverlag AG an der ETH Zürich, 2016).
11. *Le Réseau Suisse de Radars Météorologiques - MétéoSuisse* `https://www.meteosuisse.admin.ch/home/systemes-de-mesure-et-de-prevision/atmosphere/le-reseau-suisse-de-radars-meteorologiques.html` (2022).
12. Pasqualucci, F., Bartram, B. W., Kropfli, R. A. & Moninger, W. R. A Millimeter-Wavelength Dual-Polarization Doppler Radar for Cloud and Precipitation Studies. *Journal of Climate and Applied Meteorology* **22,** 758–765 (1983).
13. Seo, B.-C., Dolan, B., Krajewski, W. F., Rutledge, S. A. & Petersen, W. Comparison of Single- and Dual-Polarization–Based Rainfall Estimates Using NEXRAD Data for the NASA Iowa Flood Studies Project. *Journal of Hydrometeorology* **16,** 1658–1675 (2015).
14. Ryzhkov, A., Zhang, P., Cao, Q., Matrosov, S., Melnikov, V. & Knight, M. Measurements of Circular Depolarization Ratio with the Radar with Simultaneous Transmission / Reception, 10 (2014).
15. Ryzhkov, A. *et al.* Estimation of Depolarization Ratio Using Weather Radars with Simultaneous Transmission/Reception. *Journal of Applied Meteorology and Climatology* **56,** 1797–1816 (2017).
16. US Department of Commerce, N. *Dual-Pol Products* `https://www.weather.gov/jan/dualpolupgrade-products` (2022).
17. US Department of Commerce, N. *NWS Doppler Radar Dual Pol - Rain vs. Snow* `https://www.weather.gov/lmk/nws_radar_dualpol_rainsnow` (2022).
18. Johnston, P. E., Williams, C. R. & White, A. B. Rain Drop Size Distributions Estimated from NOAA Snow-Level Radar Data. *Journal of Atmospheric and Oceanic Technology* **39,** 353–366 (2022).
19. Browning, K. A. & Wexler, R. The Determination of Kinematic Properties of a Wind Field Using Doppler Radar. *Journal of Applied Meteorology and Climatology* **7,** 105–113 (1968).
20. Wood, V. T. & Brown, R. A. Single Doppler Velocity Signature Interpretation of Nondivergent Environmental Winds. *Journal of Atmospheric and Oceanic Technology* **3,** 114–128 (1986).
21. Gao, J., Droegemeier, K. K., Gong, J. & Xu, Q. A Method for Retrieving Mean Horizontal Wind Profiles from Single-Doppler Radar Observations Contaminated by Aliasing. *Monthly Weather Review* **132,** 1399–1409 (2004).

22. Achtemeier, G. L. The Use of Insects as Tracers for "Clear-Air" Boundary-Layer Studies by Doppler Radar. *Journal of Atmospheric and Oceanic Technology* **8,** 746–765 (1991).

23. *Understanding Weather Radar | Weather Underground* `https://www.wunderground.com/prepare/understanding-radar` (2022).

24. Ross, D. & Jones, W. L. On the Relationship of Radar Backscatter to Wind Speed and Fetch. *Boundary-Layer Meteorology* **13,** 151–163 (1978).

25. Grazioli, J. *et al.* An Adaptive Thunderstorm Measurement Concept Using C-Band and X-Band Radar Data. *IEEE Geoscience and Remote Sensing Letters* **16,** 1673–1677 (2019).

26. Ibrahim, I., Kopp, G. & Sills, D. Retrieving Wind Speed and Direction from WSR-88D Single- Doppler Measurements of Thunderstorm Winds. *American Association for Wind Engineering (AAWE) Workshop* (2021).

27. Brown, R. A., Lemon, L. R. & Burgess, D. W. Tornado Detection by Pulsed Doppler Radar. *Monthly Weather Review* **106,** 29–38 (1978).

28. US Department of Commerce, N. *NWS Doppler Radar Dual Pol - Tornado Debris* `https://www.weather.gov/lmk/nws_radar_dualpol_tordebris` (2022).

29. Chen, H. & Chandrasekar, V. *Real-Time Tornado Detection and Wind Retrieval with High-Resolution X-band Doppler Radar Network* in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (2016), 2150–2153.

30. Gabella, M., Huuskonen, A., Sartori, M., Holleman, I., Boscacci, M. & Germann, U. Evaluating the Solar Slowly Varying Component at C-Band Using Dual- and Single-Polarization Weather Radars in Europe. *Advances in Meteorology* **2017,** e4971765 (2017).

31. Gabella, M. & Leuenberger, A. Dual-Polarization Observations of Slowly Varying Solar Emissions from a Mobile X-Band Radar. *Sensors* **17,** 1185 (5 2017).

32. *Radar Bands* `https://www.everythingweather.com/weather-radar/bands.shtml` (2022).

33. Saltikoff, E. *et al.* The Threat to Weather Radars by Wireless Technology. *Bulletin of the American Meteorological Society* **97,** 151014141531002 (2015).

34. Min, C. *et al.* Coverage of China New Generation Weather Radar Network. *Advances in Meteorology* **2019,** e5789358 (2019).

35. *Next Generation Weather Radar* National Centers for Environmental Information (NCEI). `http://www.ncei.noaa.gov/products/radar/next-generation-weather-radar` (2022).

36. Germann, U., Boscacci, M., Gabella, M. & Sartori, M. Peak Performance: Radar Design for Prediction in the Swiss Alps. *Meteorological Technology International,* 42–45 (2015).

37. Germann, U., Nerini, D., Sideris, I., Foresti, L., Hering, A. & Calpini, B. Real-Time Radar - A New Alpine Radar Network. *Meteorological Technology International,* 88–92 (2017).

38. Beck, J. & Bousquet, O. Using Gap-Filling Radars in Mountainous Regions to Complement a National Radar Network: Improvements in Multiple-Doppler Wind Syntheses. *Journal of Applied Meteorology and Climatology* **52,** 1836–1850 (2013).

39. Yu, N., Gaussiat, N. & Tabary, P. Polarimetric X-band Weather Radars for Quantitative Precipitation Estimation in Mountainous Regions. *Quarterly Journal of the Royal Meteorological Society* **144,** 2603–2619 (2018).

40. Mahale, V. N., Brotzge, J. A. & Bluestein, H. B. The Advantages of a Mixed-Band Radar Network for Severe Weather Operations: A Case Study of 13 May 2009. *Weather and Forecasting* **29,** 78–98 (2014).

41. Saltikoff, E. *et al.* An Overview of Using Weather Radar for Climatological Studies: Successes, Challenges, and Potential. *Bulletin of the American Meteorological Society* **100,** 1739–1752 (2019).

42. Germann, U. *et al.* Weather Radar in Complex Orography. *Remote Sensing* **14,** 503 (3 2022).

43. Gabella, M., Joss, J. & Perona, G. Optimizing Quantitative Precipitation Estimates Using a Noncoherent and a Coherent Radar Operating on the Same Area. *Journal of Geophysical Research: Atmospheres* **105,** 2237–2245 (2000).

44. Dinku, T., Anagnostou, E. N. & Borga, M. Improving Radar-Based Estimation of Rainfall over Complex Terrain. *Journal of Applied Meteorology and Climatology* **41,** 1163–1178 (2002).

45. Berenguer, M., Sempere-Torres, D., Corral, C. & Sánchez-Diezma, R. A Fuzzy Logic Technique for Identifying Nonprecipitating Echoes in Radar Scans. *Journal of Atmospheric and Oceanic Technology* **23,** 1157–1180 (2006).

46. Gourley, J. J., Tabary, P. & du Chatelet, J. P. A Fuzzy Logic Algorithm for the Separation of Precipitating from Nonprecipitating Echoes Using Polarimetric Radar Observations. *Journal of Atmospheric and Oceanic Technology* **24,** 1439–1451 (2007).

47. Fornasiero, A., Bech, J. & Alberoni, P. P. Enhanced Radar Precipitation Estimates Using a Combined Clutter and Beam Blockage Correction Technique. *Natural Hazards and Earth System Sciences* **6,** 697–710 (2006).

48. Vulpiani, G., Montopoli, M., Passeri, L. D., Gioia, A. G., Giordano, P. & Marzano, F. S. On the Use of Dual-Polarized C-Band Radar for Operational Rainfall Retrieval in Mountainous Areas. *Journal of Applied Meteorology and Climatology* **51,** 405–425 (2012).

49. Cremonini, R. & Bechini, R. Heavy Rainfall Monitoring by Polarimetric C-Band Weather Radars. *Water* **2,** 838–848 (4 2010).

50. Montopoli, M., Roberto, N., Adirosi, E., Gorgucci, E. & Baldini, L. Investigation of Weather Radar Quantitative Precipitation Estimation Methodologies in Complex Orography. *Atmosphere* **8,** 34 (2 2017).

51. Schneebeli, M., Grazioli, J. & Berne, A. Improved Estimation of the Specific Differential Phase Shift Using a Compilation of Kalman Filter Ensembles. *IEEE Transactions on Geoscience and Remote Sensing* **52,** 5137–5149 (2014).

52. Gabella, M. Variance of Fluctuating Radar Echoes from Thermal Noise and Randomly Distributed Scatterers. *Atmosphere* **5,** 92–100 (1 2014).

53. Gabella, M. On the Use of Bright Scatterers for Monitoring Doppler, Dual-Polarization Weather Radars. *Remote Sensing* **10,** 1007 (7 2018).

54. Germann, U., Figueras, J., Gabella, M., Hering, A., Sideris, I. & Calpini, B. Alpine Weather Radar: Triggering Innovation. *Meteorological Technology International,* 62–65 (2016).

55. Kaltenboeck, R. New Generation of Dual Polarized Weather Radars in Austria, 6 (2012).

56. I Ventura, J. F. & Tabary, P. The New French Operational Polarimetric Radar Rainfall Rate Product. *Journal of Applied Meteorology and Climatology* **52,** 1817–1835 (2013).

57. Bechini, R. & Cremonini, R. The Weather Radar System of North-Western Italy: An Advanced Tool for Meteorological Surveillance. *Copernicus GmbH,* 400–404 (2002).

58. Chen, H., Cifelli, R. & White, A. Improving Operational Radar Rainfall Estimates Using Profiler Observations Over Complex Terrain in Northern California. *IEEE Transactions on Geoscience and Remote Sensing* **58,** 1821–1832 (2020).

59. Liou, Y.-C., Chang, S.-F. & Sun, J. An Application of the Immersed Boundary Method for Recovering the Three-Dimensional Wind Fields over Complex Terrain Using Multiple-Doppler Radar Data. *Monthly Weather Review* **140,** 1603–1619 (2012).

60. Tsai, C.-L., Kim, K., Liou, Y.-C., Lee, G. & Yu, C.-K. Impacts of Topography on Airflow and Precipitation in the Pyeongchang Area Seen from Multiple-Doppler Radar Observations. *Monthly Weather Review* **146,** 3401–3424 (2018).

61. Gou, Y., Ma, Y., Chen, H. & Wen, Y. Radar-Derived Quantitative Precipitation Estimation in Complex Terrain over the Eastern Tibetan Plateau. *Atmospheric Research* **203,** 286–297 (2018).

62. Gou, Y., Ma, Y., Chen, H. & Yin, J. Utilization of a C-band Polarimetric Radar for Severe Rainfall Event Analysis in Complex Terrain over Eastern China. *Remote Sensing* **11,** 22 (1 2019).

63. Gray, W. R. & Seed, A. W. The Characterisation of Orographic Rainfall. *Meteorological Applications* **7,** 105–119 (2000).

64. Talchabhadel, R. *et al.* Weather Radar in Nepal: Opportunities and Challenges in a Mountainous Region (2020).

65. Bendix, J. *et al.* RadarNet-Sur First Weather Radar Network in Tropical High Mountains. *Bulletin of the American Meteorological Society* **98,** 1235–1254 (2017).

66. Thorndahl, S. *et al.* Weather Radar Rainfall Data in Urban Hydrology. *Hydrology and Earth System Sciences* **21,** 1359–1380 (2017).

67. Brandes, E. A. Optimizing Rainfall Estimates with the Aid of Radar. *Journal of Applied Meteorology and Climatology* **14,** 1339–1345 (1975).

68. Smith, J. A. & Krajewski, W. F. Estimation of the Mean Field Bias of Radar Rainfall Estimates. *Journal of Applied Meteorology and Climatology* **30,** 397–412 (1991).

69. Anagnostou, E. N., Krajewski, W. F., Seo, D.-J. & Johnson, E. R. Mean-Field Rainfall Bias Studies for WSR-88D. *Journal of Hydrologic Engineering* **3,** 149–159 (1998).

70. Chumchean, S., Seed, A. & Sharma, A. Correcting of Real-Time Radar Rainfall Bias Using a Kalman Filtering Approach. *Journal of Hydrology* **317,** 123–137 (2006).

71. Cole, S. J. & Moore, R. J. Distributed Hydrological Modelling Using Weather Radar in Gauged and Ungauged Basins. *Advances in Water Resources. Weather Radar and Hydrology* **32,** 1107–1120 (2009).

72. Krajewski, W. F. Cokriging Radar-Rainfall and Rain Gage Data. *Journal of Geophysical Research: Atmospheres* **92,** 9571–9580 (1987).

73. Creutin, J. D., Delrieu, G. & Lebel, T. Rain Measurement by Raingage-Radar Combination: A Geostatistical Approach. *Journal of Atmospheric and Oceanic Technology* **5,** 102–115 (1988).

74. Seo, D.-J., Krajewski, W. F., Azimi-Zonooz, A. & Bowles, D. S. Stochastic Interpolation of Rainfall Data from Rain Gages and Radar Using Cokriging: 2. Results. *Water Resources Research* **26,** 915–924 (1990).

75. Haberlandt, U. Geostatistical Interpolation of Hourly Precipitation from Rain Gauges and Radar for a Large-Scale Extreme Rainfall Event. *Journal of Hydrology* **332,** 144–157 (2007).

76. Schuurmans, J. M., Bierkens, M. F. P., Pebesma, E. J. & Uijlenhoet, R. Automatic Prediction of High-Resolution Daily Rainfall Fields for Multiple Extents: The Potential of Operational Radar. *Journal of Hydrometeorology* **8,** 1204–1224 (2007).

77. Velasco-Forero, C. A., Sempere-Torres, D., Cassiraga, E. F. & Jaime Gómez-Hernández, J. A Non-Parametric Automatic Blending Methodology to Estimate Rainfall Fields from Rain Gauge and Radar Data. *Advances in Water Resources. Weather Radar and Hydrology* **32,** 986–1002 (2009).

78. Sideris, I. V., Gabella, M., Erdin, R. & Germann, U. Real-Time Radar–Rain-Gauge Merging Using Spatio-Temporal Co-Kriging with External Drift in the Alpine Terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society* **140,** 1097–1111 (2014).

79. Foehn, A. T. *Radar-Rain Gauge Merging and Discharge Data Assimilation for Flood Forecasting in Alpine Catchments* (EPFL, Lausanne, 2019). 214 pp.

80. Calheiros, R. V. & Zawadzki, I. Reflectivity-Rain Rate Relationships for Radar Hydrology in Brazil. *Journal of Applied Meteorology and Climatology* **26,** 118–132 (1987).

81. Rosenfeld, D., Wolff, D. B. & Amitai, E. The Window Probability Matching Method for Rainfall Measurements with Radar. *Journal of Applied Meteorology and Climatology* **33,** 682–693 (1994).

82. Todini, E. A Bayesian Technique for Conditioning Radar Precipitation Estimates to Rain-Gauge Measurements. *Hydrology and Earth System Sciences* **5,** 187–199 (2001).

83. Barton, Y., Sideris, I., Germann, U. & Martius, O. A Method for Real-time Temporal Disaggregation of Blended Radar—Rain-gauge Precipitation Fields. *Meteorological Applications* **27** (2019).

84. Barton, Y., Sideris, I. V., Raupach, T. H., Gabella, M., Germann, U. & Martius, O. A Multi-Year Assessment of Sub-Hourly Gridded Precipitation for Switzerland Based on a Blended Radar—Rain-gauge Dataset. *International Journal of Climatology* **40,** 5208–5222 (2020).

85. World Meteorological Organization. *Guide to Meteorological Instruments and Methods of Observation.* (World Meteorological Organization, Geneva, Switzerland, 2018).

86. Anagnostou, E. N., Krajewski, W. F. & Smith, J. Uncertainty Quantification of Mean-Areal Radar-Rainfall Estimates. *Journal of Atmospheric and Oceanic Technology* **16,** 206–215 (1999).

87. Smith, J. A., Baeck, M. L., Meierdiercks, K. L., Miller, A. J. & Krajewski, W. F. Radar Rainfall Estimation for Flash Flood Forecasting in Small Urban Watersheds. *Advances in Water Resources. Recent Developments in Hydrologic Analysis* **30,** 2087–2097 (2007).

88. Werner, M. & Cranston, M. Understanding the Value of Radar Rainfall Nowcasts in Flood Forecasting and Warning in Flashy Catchments. *Meteorological Applications* **16,** 41–55 (2009).

89. Gabella, M., Bolliger, M., Germann, U. & Perona, G. Large Sample Evaluation of Cumulative Rainfall Amounts in the Alps Using a Network of Three Radars. *Atmospheric Research* **1–4,** 256–268 (2005).

90. Rezacova, D., Sokol, Z. & Pesice, P. A Radar-Based Verification of Precipitation Forecast for Local Convective Storms. *Atmospheric Research* **83,** 211–224 (2007).

91. Zawadzki, I., Morneau, J. & Laprise, R. Predictability of Precipitation Patterns: An Operational Approach. *Journal of Applied Meteorology and Climatology* **33,** 1562–1571 (1994).

92. Sideris, I. V., Foresti, L., Nerini, D. & Germann, U. NowPrecip: Localized Precipitation Nowcasting in the Complex Terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society* **146,** 1768–1800 (2020).

93. Venugopal, V., Foufoula-Georgiou, E. & Sapozhnikov, V. Evidence of Dynamic Scaling in Space-Time Rainfall. *Journal of Geophysical Research: Atmospheres* **104,** 31599–31610 (1999).

94. Walker, J. R., MacKenzie, W. M., Mecikalski, J. R. & Jewett, C. P. An Enhanced Geostationary Satellite–Based Convective Initiation Algorithm for 0–2-h Nowcasting with Object Tracking. *Journal of Applied Meteorology and Climatology* **51,** 1931–1949 (2012).

95. Han, H. *et al.* Detection of Convective Initiation Using Meteorological Imager Onboard Communication, Ocean, and Meteorological Satellite Based on Machine Learning Approaches. *Remote Sensing* **7,** 9184–9204 (7 2015).

96. Mecikalski, J. R., Williams, J. K., Jewett, C. P., Ahijevych, D., LeRoy, A. & Walker, J. R. Probabilistic 0–1-h Convective Initiation Nowcasts That Combine Geostationary Satellite Observations and Numer-

ical Weather Prediction Model Data. *Journal of Applied Meteorology and Climatology* **54,** 1039–1059 (2015).

97. Foresti, L., Sideris, I. V., Nerini, D., Beusch, L. & Germann, U. Using a 10-Year Radar Archive for Nowcasting Precipitation Growth and Decay: A Probabilistic Machine Learning Approach. *Weather and Forecasting* **34,** 1547–1569 (2019).

98. Simonin, D., Pierce, C., Roberts, N., Ballard, S. P. & Li, Z. Performance of Met Office Hourly Cycling NWP-based Nowcasting for Precipitation Forecasts. *Quarterly Journal of the Royal Meteorological Society* **143,** 2862–2873 (2017).

99. Golding, B. W. Nimrod: A System for Generating Automated Very Short Range Forecasts. *Meteorological Applications* **5,** 1–16 (1998).

100. Haiden, T., Kann, A., Wittmann, C., Pistotnik, G., Bica, B. & Gruber, C. The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region. *Weather and Forecasting* **26,** 166–183 (2011).

101. Lin, H.-H. *et al.* Multi-Weather Evaluation of Nowcasting Methods Including a New Empirical Blending Scheme. *Atmosphere* **11,** 1166 (11 2020).

102. Nerini, D., Foresti, L., Leuenberger, D., Robert, S. & Germann, U. A Reduced-Space Ensemble Kalman Filter Approach for Flow-Dependent Integration of Radar Extrapolation Nowcasts and NWP Precipitation Ensembles. *Monthly Weather Review* **147,** 987–1006 (2019).

103. Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M. & Reinhardt, T. Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities. *Monthly Weather Review* **139,** 3887–3905 (2011).

104. *COSMO-NExT - Le Nouveau Système de Modèles Numériques Pour Les Prévisions Du Temps - Météo-Suisse* https://www.meteosuisse.admin.ch/home/recherche-et-collaboration/collaboration-internationale/cosmo.subpage.html/fr/data/projects/2012/cosmo-next.html (2022).

105. *Cosmo Public Area* http://www.cosmo-model.org/ (2022).

106. Ebert, E. E. & McBride, J. L. Verification of Precipitation in Weather Systems: Determination of Systematic Errors. *Journal of Hydrology* **239,** 179–202 (2000).

107. Grose, A., Smith, E., Chung, H.-S., Ou, M.-L., Sohn, B.-J. & Turk, J. Possibilities and Limitations for Quantitative Precipitation Forecasts Using Nowcasting Methods with Infrared Geosynchronous Satellite Imagery. *Journal of Applied Meteorology - J APPL METEOROL* **41,** 763–785 (2002).

108. Moletto-Lobos, I., Mattar, C. & Barichivich, J. Performance of Satellite-Based Evapotranspiration Models in Temperate Pastures of Southern Chile. *Water* **12,** 3587 (2020).

109. Aalstad, K., Westermann, S. & Bertino, L. Evaluating Satellite Retrieved Fractional Snow-Covered Area at a High-Arctic Site Using Terrestrial Photography. *Remote Sensing of Environment* **239,** 111618 (2020).

110. Peng, J. *et al.* A Roadmap for High-Resolution Satellite Soil Moisture Applications – Confronting Product Characteristics with User Requirements. *Remote Sensing of Environment* **252,** 112162 (2021).

111. Kordzakhia, G., Shengelia, L., Tvauri, G., Tsomaia, V. & Dzadzamia, M. Satellite Remote Sensing Outputs of the Certain Glaciers on the Territory of East Georgia. *The Egyptian Journal of Remote Sensing and Space Science. Research and Innovation for Sustainable Soil Management* **18,** S1–S7 (1, Supplement 1 2015).

112. Gabella, M., Morin, E. & Notarpietro, R. Using TRMM Spaceborne Radar as a Reference for Compensating Ground-Based Radar Range Degradation: Methodology Verification Based on Rain Gauges in Israel. *Journal of Geophysical Research* **116** (2011).

113. Gabella, M., Duque, D. & Notarpietro, R. Comparing Meteorological Spaceborne and Ground-Based Radars: Optimal Satellite Overpass Distance from a Ground-Based Radar Site. *International Journal of Remote Sensing* **33,** 322–330 (2012).

114. Mulvaney, T. J. *On the Use of Self-Registering Rain and Flood Gauges in Making Observations of the Relations of Rain Fall and Flood Discharges in a given Catchment.*

115. Beven, K. *Rainfall-Runoff Modelling: The Primer* (2012).

116. Jain, S. K. *et al.* A Brief Review of Flood Forecasting Techniques and Their Applications. *International Journal of River Basin Management* **16,** 329–344 (2018).

117. Sitterson, J., Knightes, C., Parmar, R., Wolfe, K., Avant, B. & Muche, M. An Overview of Rainfall-Runoff Model Types. *undefined* (2018).

118. Daniel, E. B., Camp, J. V., LeBoeuf, E. J., Penrod, J. R., Dobbins, J. P. & Abkowitz, M. D. Watershed Modeling and Its Applications: A State-of-the-Art Review. *The Open Hydrology Journal* **5** (2011).

119. Garen, D. C. Improved Techniques in Regression-Based Streamflow Volume Forecasting. *Journal of Water Resources Planning and Management* **118,** 654–670 (1992).

120. Pagano, T. C., Garen, D. C., Perkins, T. R. & Pasteris, P. A. Daily Updating of Operational Statistical Seasonal Water Supply Forecasts for the Western U.S. *JAWRA Journal of the American Water Resources Association* **45,** 767–778 (2009).

121. Nilsson, P., Uvo, C., Landman, W. & Nguyen, T. Downscaling of GCM Forecasts to Streamflow over Scandinavia. *Hydrology Research* **39.1,** 17–26 (2008).

122. Abdollahi, S., Akhoond-Ali, A. M., Mirabbasi, R. & Adamowski, J. F. Probabilistic Event Based Rainfall-Runoff Modeling Using Copula Functions. *Water Resources Management* **33,** 3799–3814 (2019).

123. Moore, R. J. The Probability-Distributed Principle and Runoff Production at Point and Basin Scales. *Hydrological Sciences Journal* **30,** 273–297 (1985).

124. Moore, R. J. The PDM Rainfall-Runoff Model. *Hydrology and Earth System Sciences* **11,** 483–499 (2007).

125. Daniell, T. M. *Neural Networks—Applications in Hydrology and Water Resources Engineering* in. International Hydrology & Water Resources Symposium (1991).

126. Hsu, K.-l., Gupta, H. V. & Sorooshian, S. Artificial Neural Network Modeling of the Rainfall-Runoff Process. *Water Resources Research* **31,** 2517–2530 (1995).

127. Solomatine, D., See, L. & Abrahart, R. in *Practical Hydroinformatics: Computational Intelligence and Technological Developments in Water Applications* (eds Abrahart, R. J., See, L. M. & Solomatine, D. P.) 17–30 (Springer, Berlin, Heidelberg, 2008).

128. Abrahart, R. *et al.* Two Decades of Anarchy? Emerging Themes and Outstanding Challenges for Neural Network River Forecasting. *Progress in Physical Geography* **36,** 480–513 (2012).

129. Lin Hsu, K., Gupta, H. V. & Sorooshian, S. *Application of a Recurrent Neural Network to Rainfall-Runoff Modeling: Proceedings of the 1997 24th Annual Water Resources Planning and Management Conference* in (1997), 68–73.

130. Kumar, N. D., Raju, S. K. & Sathish, T. River Flow Forecasting Using Recurrent Neural Networks. *Water Resources Management* **18,** 143–161 (2004).

131. Chiang, Y.-M. & Chang, F.-J. Integrating Hydrometeorological Information for Rainfall-Runoff Modelling by Artificial Neural Networks. *Hydrological Processes* **23,** 1650–1659 (2009).

132. Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks. *Hydrology and Earth System Sciences* **22,** 6005–6022 (2018).

133. Tian, Y., Xu, Y.-P., Yang, Z., Wang, G. & Zhu, Q. Integration of a Parsimonious Hydrological Model with Recurrent Neural Networks for Improved Streamflow Forecasting. *Water* **10,** 1655 (11 2018).

134. Smith, J. & Eli, R. N. Neural-Network Models of Rainfall-Runoff Process. *Journal of Water Resources Planning and Management* **121,** 499–508 (1995).

135. Rajurkar, M. P., Kothyari, U. C. & Chaube, U. C. Modeling of the Daily Rainfall-Runoff Relationship with Artificial Neural Network. *Journal of Hydrology* **285,** 96–113 (2004).

136. Farfán, J. F., Palacios, K., Ulloa, J. & Avilés, A. A Hybrid Neural Network-Based Technique to Improve the Flow Forecasting of Physical and Data-Driven Models: Methodology and Case Studies in Andean Watersheds. *Journal of Hydrology: Regional Studies* **27,** 100652 (2020).

137. Zadeh, L. A. Fuzzy Sets. *Information and Control* **8,** 338–353 (1965).

138. Luchetta, A. & Manetti, S. A Real Time Hydrological Forecasting System Using a Fuzzy Clustering Approach. *Computers & Geosciences* **29,** 1111–1117 (2003).

139. Yu, P.-S. & Chen, S.-T. Updating Real-Time Flood Forecasting Using a Fuzzy Rule-Based Model/Mise à Jour de Prévision de Crue En Temps Réel Grâce à Un Modèle à Base de Règles Floues. *Hydrological Sciences Journal* **50,** null–278 (2005).

140. Bae, D.-H., Jeong, D. M. & Kim, G. Monthly Dam Inflow Forecasts Using Weather Forecasting Information and Neuro-Fuzzy Technique. *Hydrological Sciences Journal* **52,** 99–113 (2007).

141. Shrestha, B. P., Duckstein, L. & Stakhiv, E. Z. Fuzzy Rule-Based Modeling of Reservoir Operation. *Journal of Water Resources Planning and Management* **122,** 262–269 (1996).

142. Sherman, L. K. Streamflow from Rainfall by the Unit-Graph Method. *undefined* (1932).

143. Snyder, F. F. Synthetic Unit-Graphs. *Eos, Transactions American Geophysical Union* **19,** 447–454 (1938).

144. Sharma, S. Snyder Unit Hydrograph and GIS for Estimation of Flood for Un-Gauged Catchments in Lower Tapi Basin,India. *Hydrology: Current Research* **06** (2015).

145. Goñi, M., López, J. J. & Gimena, F. N. Geomorphological Instantaneous Unit Hydrograph Model with Distributed Rainfall. *CATENA* **172,** 40–53 (2019).

146. Khaleghi, M. R., Gholami, V., Ghodusi, J. & Hosseini, H. Efficiency of the Geomorphologic Instantaneous Unit Hydrograph Method in Flood Hydrograph Simulation. *CATENA* **87,** 163–171 (2011).

147. Wang, W.-c., Zhao, Y.-w., Chau, K.-w., Xu, D.-m. & Liu, C.-j. Improved Flood Forecasting Using Geomorphic Unit Hydrograph Based on Spatially Distributed Velocity Field. *Journal of Hydroinformatics* **23,** 724–739 (2021).

148. Perrin, C., Michel, C. & Andréassian, V. Improvement of a Parsimonious Model for Streamflow Simulation. *Journal of Hydrology* **279,** 275–289 (2003).

149. Coron, L., Thirel, G., Delaigue, O., Perrin, C. & Andréassian, V. The Suite of Lumped GR Hydrological Models in an R Package. *Environmental Modelling & Software* **94,** 166–171 (2017).

150. For the Protection and Security of the Citizen (Joint Research Centre), I., Roo de, A., Knijff van der, J. & Burek, P. *LISFLOOD, Distributed Water Balance and Flood Simulation Model: Revised User Manual 2013* (Publications Office of the European Union, LU, 2013).

151. Van Der Knijff, J. M., Younis, J. & De Roo, A. P. J. LISFLOOD: A GIS-based Distributed Model for River Basin Scale Water Balance and Flood Simulation. *International Journal of Geographical Information Science* **24,** 189–212 (2010).

152. *Home | Copernicus EMS - European Flood Awareness System* https://www.efas.eu/en (2022).

153. Jordan, F. Modèle de Prévision et de Gestion Des Crues: Optimisation Des Opérations Des Aménagements Hydroélectriques à Accumulation Pour La Réduction Des Débits de Crue (2007).

154. Garcia Hernandez, J., Jordan, F., Dubois, J., Boillat, J.-L. & Schleiss, A. Routing System II - Modélisation d'écoulements Dans Des Systèmes Hydrauliques (2007).

155. Bergström, S. & Forsman, A. Development of a Conceptual Deterministic Rainfall-Runoff Model. *Hydrology Research* **4,** 147–170 (1973).

156. Pechlivanidis, I., Bosshard, T., Spångmyr, H., Lindström, G., Gustafsson, D. & Arheimer, B. Uncertainty in the Swedish Operational Hydrological Forecasting Systems (2014).

157. Ayzel, G. & Abramov, D. OpenForecast: An Assessment of the Operational Run in 2020–2021. *Geosciences* **12,** 67 (2 2022).

158. *FOEN, Bern* https://www.iksr.org/en/topics/floods/flood-warning-and-forecasting-centres/foen-bern (2022).

159. Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P. & Thielen, J. Technical Review of Large-Scale Hydrological Models for Implementation in Operational Flood Forecasting Schemes on Continental Level. *Environmental Modelling & Software* **75,** 68–76 (2016).

160. Atkins, J. T., Wiley, J. B. & Paybins, K. S. *Calibration Parameters Used to Simulate Streamflow from Application of the Hydrologic Simulation Program-FORTRAN Model (HSPF) to Mountainous Basins Containing Coal Mines in West Virginia* USGS Numbered Series 2005-5099 (2005).

161. Al-Asadi, K., Abbas, A. & A. Hamdan, A. Optimization of the Hydrological Tank Model by Downhill Simplex Method. *International Journal of Civil Engineering* **18** (2020).

162. Hassan, A., Tügel, F., Özgen-Xian, I., Molkenthin, F. & Hinkelmann, R. *Applying Optimization Algorithms to Calibrate a Shallow Water Model for Rainfall-Runoff Simulations* in (2021).

163. Preis, A. & Ostfeld, A. A Coupled Model Tree–Genetic Algorithm Scheme for Flow and Water Quality Predictions in Watersheds. *Journal of Hydrology* **349,** 364–375 (2008).

164. Duan, Q., Sorooshian, S. & Gupta, V. Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models. *Water Resources Research* **28,** 1015–1031 (1992).

165. Zhang, C., Wang, R.-b. & Meng, Q.-x. Calibration of Conceptual Rainfall-Runoff Models Using Global Optimization. *Advances in Meteorology* **2015,** e545376 (2015).

166. Jeon, J.-H., Park, C.-G. & Engel, B. A. Comparison of Performance between Genetic Algorithm and SCE-UA for Calibration of SCS-CN Surface Runoff Simulation. *Water* **6,** 3433–3456 (11 2014).

167. Klemeš, V. Operational Testing of Hydrological Simulation Models. *Hydrological Sciences Journal* **31,** 13–24 (1986).

168. Javelle, P., Fouchier, C., Arnaud, P. & Lavabre, J. Flash Flood Warning at Ungauged Locations Using Radar Rainfall and Antecedent Soil Moisture Estimations. *Journal of Hydrology. Flash Floods: Observations and Analysis of Hydrometeorological Controls* **394,** 267–274 (2010).

169. Tangara, M. *Nouvelle Méthode de Prévision de Crue Utilisant Un Modèle Pluie-Débit Global* These de doctorat (Paris, EPHE, 2005).

170. Szilagyi, J. State-Space Discretization of the Kalinin-Milyukov-Nash-Cascade in a Sample-Data System Framework for Streamflow Forecasting. *Journal of Hydrologic Engineering* **8,** 339–347 (2003).

171. Goswami, M., O'Connor, K. M., Bhattarai, K. P. & Shamseldin, A. Y. Assessing the Performance of Eight Real-Time Updating Models and Procedures for the Brosna River. *Hydrology and Earth System Sciences* **9,** 394–411 (2005).

172. Zhang, X. & Liu, P. A Time-Varying Parameter Estimation Approach Using Split-Sample Calibration Based on Dynamic Programming. *Hydrology and Earth System Sciences* **25,** 711–733 (2021).

173.   Wöhling, T., Lennartz, F. & Zappa, M. Technical Note: Updating Procedure for Flood Forecasting with Conceptual HBV-type Models. *Hydrology and Earth System Sciences Discussions* **10,** 783–788 (2006).

174.   Borup, M. *Real Time Updating in Distributed Urban Rainfall Runoff Modelling* (Department of Environmental Engineering, Technical University of Denmark (DTU), Kgs. Lyngby, 2014).

175.   Silvestro, F., Ercolani, G., Gabellani, S., Giordano, P. & Falzacappa, M. Improving Real-Time Operational Streamflow Simulations Using Discharge Data to Update State Variables of a Distributed Hydrological Model. *Hydrology Research* **52,** 1239–1260 (2021).

176.   Ho, J.-Y. & Lee, K. T. Grey Forecast Rainfall with Flow Updating Algorithm for Real-Time Flood Forecasting. *Water* **7,** 1840–1865 (5 2015).

177.   Serban, P. & Askew, A. J. Hydrological Forecasting and Updating Procedures. *IAHS Publications,* 57–369 (1991).

178.   Borup, M., Grum, M. & Mikkelsen, P. S. Real Time Adjustment of Slow Changing Flow Components in Distributed Urban Runoff Models, 9 (2011).

179.   Komma, J., Blöschl, G. & Reszler, C. Soil Moisture Updating by Ensemble Kalman Filtering in Real-Time Flood Forecasting. *Journal of Hydrology* **357,** 228–242 (2008).

180.   Sun, L., Seidou, O., Nistor, I. & Liu, K. Review of the Kalman-type Hydrological Data Assimilation. *Hydrological Sciences Journal* **61,** 2348–2366 (2016).

181.   Liu, Y. & Gupta, H. V. Uncertainty in Hydrologic Modeling: Toward an Integrated Data Assimilation Framework. *Water Resources Research* **43** (2007).

182.   Horton, R. E. An Approach Toward a Physical Interpretation of Infiltration-Capacity. *Soil Science Society of America Journal* **5,** 399–417 (1941).

183.   Parnas, F. E. Å., Abdalla, E. M. H. & Muthanna, T. M. Evaluating Three Commonly Used Infiltration Methods for Permeable Surfaces in Urban Areas Using the SWMM and STORM. *Hydrology Research* **52,** 160–175 (2021).

184.   Estupina Borrell, V., Chorda, J. & Dartus, D. Prévision des crues éclair. *Comptes Rendus Geoscience* **337,** 1109–1119 (2005).

185.   Montz, B. E. & Gruntfest, E. Flash Flood Mitigation: Recommendations for Research and Applications. *Global Environmental Change Part B: Environmental Hazards* **4,** 15–22 (2002).

186.   Thirriot, C. Rainwater Infiltration and Runoff. *La Houille Blanche* **66,** 259–265 (1980).

187.   Beven, K. On Subsurface Stormflow: An Analysis of Response Times. *Hydrological Sciences Journal* **27,** 505–521 (1982).

188.   Cosandey, C. & Robinson, M. *Hydrologie continentale Armand Colin* **4** (Association des Géographes de l'Est, 2000).

189.   Laurenson, E. M. *Hydrograph Synthesis by Runoff Routing* (E.M.Laurenson, Syd, 1962).

190.   Chu, X. & Steinman, A. Event and Continuous Hydrologic Modeling with HEC-HMS. *Journal of Irrigation and Drainage Engineering* **135,** 119–124 (2009).

191.   De Silva, M. M. G. T., Weerakoon, S. B. & Herath, S. Modeling of Event and Continuous Flow Hydrographs with HEC–HMS: Case Study in the Kelani River Basin, Sri Lanka. *Journal of Hydrologic Engineering* **19,** 800–806 (2014).

192.   Hossain, S. A Comparison of Continuous and Event-Based Rainfall-Runoff (RR) Modelling Using EPA-SWMM. *Water* **11,** 611 (2019).

193.   Hoes, O. & Nelen, F. *Continuous Simulation Or Event-basedModelling To Estimate Flood Probabilities?* in (2005).

194.   Calver, A., Stewart, E. & Goodsell, G. Comparative Analysis of Statistical and Catchment Modelling Approaches to River Flood Frequency Estimation. *Journal of Flood Risk Management* **2,** 24–31 (2009).

195.   Kemp, D. J. & Hewa, G. *An Investigation into the Efficacy of Australian Rainfall and Runoff 2016 Procedures in the Mount Lofty Ranges, South Australia* in *Hydrology and Water Resources Symposium (HWRS 2018): Water and Communities* (2018), 407–428.

196.   Peredo, D., Ramos, M.-H., Andréassian, V. & Oudin, L. Investigating Hydrological Model Versatility to Simulate Extreme Flood Events. *Hydrological Sciences Journal* **67,** 628–645 (2022).

197.   Piotte, O. *et al.* The Early Warning Service for Heavy Rainfalls and Flash Floods. *La Houille Blanche* **106,** 75–84 (2020).

198.   *Un Nouveau Guide Du CEPRI Sur La Prévision et l'anticipation Des Crues et Des Inondations - Bienvenue Sur Le Site Du CEPRI* https://www.cepri.net/actualites/items/Guide_prevision.html (2022).

199. Javelle, P., Demargne, J., Defrance, D., Pansu, J. & Arnaud, P. Evaluating Flash-Flood Warnings at Ungauged Locations Using Post-Event Surveys: A Case Study with the AIGA Warning System. *Hydrological Sciences Journal/Journal des Sciences Hydrologiques* **59** (2013).

200. Météo-France. *Avertissement Pluies Intenses Communes - Vigicrues Flash* http://apic-vigicruesflash.fr (2022).

201. Carpenter, T. M., Sperfslage, J. A., Georgakakos, K. P., Sweeney, T. & Fread, D. L. National Threshold Runoff Estimation Utilizing GIS in Support of Operational Flash Flood Warning Systems. *Journal of Hydrology* **224,** 21–44 (1999).

202. *Flash Flood Guidance System with Global Coverage (FFGS)* https://public.wmo.int/en/projects/ffgs (2022).

203. Demargne, J., Javelle, P., Organde, D., Garandeau, L. & Janet, B. Integrating High-Resolution Precipitation Nowcasts for Improved Flash Flood Warnings. *La Houille Blanche* **105,** 13–21 (2019).

204. Henderson, F. M. *Open Channel Flow* (Macmillan, New York, 1966).

205. De Saint-Aubin, C., Garandeau, L., Janet, B. & Javelle, P. A New French Flash Flood Warning Service. *E3S Web of Conferences* **7** (eds Lang, M., Klijn, F. & Samuels, P.) 18024 (2016).

206. Younis, J., Anquetin, S. & Thielen, J. The Benefit of High-Resolution Operational Weather Forecasts for Flash Flood Warning. *Hydrology and Earth System Sciences* **12,** 1039–1051 (2008).

207. Alfieri, L., Thielen, J. & Pappenberger, F. Ensemble Hydro-Meteorological Simulation for Flash Flood Early Detection in Southern Switzerland. *Journal of Hydrology* **424–425,** 143–153 (2012).

208. Bartholmes, J. C., Thielen, J., Ramos, M. H. & Gentilini, S. The European Flood Alert System EFAS – Part 2: Statistical Skill Assessment of Probabilistic and Deterministic Operational Forecasts. *Hydrology and Earth System Sciences* **13,** 141–153 (2009).

209. Berthet, L., Gaume, É., Piotte, O., les groupes de travail du Conseil scientifique et technique du SCHAPI & du chantier d'intérêt commun SCHAPI – SPC Estimation des incertitudes de prévision. Evaluer et communiquer les incertitudes associées aux prévisions hydrologiques pour mieux partager l'information. *La Houille Blanche* **102,** 18–24 (2016).

210. Thielen, J., Bartholmes, J., Ramos, M.-H. & de Roo, A. The European Flood Alert System – Part 1: Concept and Development. *Hydrology and Earth System Sciences* **13,** 125–140 (2009).

211. Werner, M., Cranston, M., Harrison, T., Whitfield, D. & Schellekens, J. Recent Developments in Operational Flood Forecasting in England, Wales and Scotland. *Meteorological Applications* **16,** 13–22 (2009).

212. Cranston, M. *et al.* Countrywide Flood Forecasting in Scotland: Challenges for Hydrometeorological Model Uncertainty and Prediction, 6.

213. Troin, M., Arsenault, R., Wood, A. W., Brissette, F. & Martel, J.-L. Generating Ensemble Streamflow Forecasts: A Review of Methods and Approaches Over the Past 40 Years. *Water Resources Research* **57,** e2020WR028392 (2021).

214. Day, G. N. Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management* **111,** 157–170 (1985).

215. Olsson, J., Uvo, C. B., Foster, K. & Yang, W. Technical Note: Initial Assessment of a Multi-Method Approach to Spring-Flood Forecasting in Sweden. *Hydrology and Earth System Sciences* **20,** 659–667 (2016).

216. Wanders, N. *et al.* Development and Evaluation of a Pan-European Multimodel Seasonal Hydrological Forecasting System. *Journal of Hydrometeorology* **20,** 99–115 (2019).

217. Zappa, M. *et al.* Propagation of Uncertainty from Observing Systems and NWP into Hydrological Models: COST-731 Working Group 2. *Atmospheric Science Letters* **11,** 83–91 (2010).

218. Suter, S., Konzelmann, T., Mühlhäuser, C., Begert, M. & Heimo, A. *SwissMetNet–the New Automatic Meteorological Network of Switzerland: Transition from Old to New Network, Data Management and First Results* in. Proceedings of the 4th International Conference on Experiences with Automatic Weather Stations (4th ICEAWS). **24** (2006).

219. Dubois, J., Boillat, J. & Schleiss, A. Routing System – Modélisation Du Routage de Crues Dans Des Systèmes Hydrauliques à Surface Libre. *undefined* (2000).

220. Schaefli, B., Hingray, B., Niggli, M. & Musy, A. A Conceptual Glacio-Hydrological Model for High Mountainous Catchments. *Hydrology and Earth System Sciences* **9,** 95–109 (2005).

221. Jordan, F. & Legrosdidier, S. *Radar4Infra: Annual Report N°1 2021* 1 (Hydrique Ingénieurs, 2021).

222. *HYDROmaps* https://hydromaps.ch/#fr/10/46.1794/7.1548/bl_hds--b02_b0200_alpine_precip_y_1971_2008v1_0$0/NULL (2022).

223. Progênio, M. F. & Blanco, C. J. C. Cumulative Distribution Function of Daily Rainfall in the Tocantins–Araguaia Hydrographic Region, Amazon, Brazil. *Natural Resource Modeling* **33,** e12264 (2020).