



**HAL**  
open science

# New Statistical Methods for Precipitation Bias Correction Applied to WRF Model Simulations in the Antisana Region (Ecuador)

María Belén Heredia, Clémentine Junquas, Clémentine Prieur, Thomas Condom

► **To cite this version:**

María Belén Heredia, Clémentine Junquas, Clémentine Prieur, Thomas Condom. New Statistical Methods for Precipitation Bias Correction Applied to WRF Model Simulations in the Antisana Region (Ecuador). *Journal of Hydrometeorology*, 2018, 19 (12), pp.2021-2040. 10.1175/JHM-D-18-0032.1 . hal-01971425

**HAL Id: hal-01971425**

**<https://hal.science/hal-01971425>**

Submitted on 28 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **New statistical methods for precipitation bias correction applied to WRF**  
2 **model simulations in the Antisana Region (Ecuador)**

3 María Belén Heredia\*, Clémentine Junquas<sup>†</sup>, Clémentine Prieur<sup>‡</sup>, Thomas Condom<sup>§</sup>

4 \**Corresponding author address:* ETNA, Irstea Grenoble, Saint-Martin-d'Hères, France

5 E-mail: maria-belen.heredia@irstea.fr

6 <sup>†</sup>Université de Grenoble Alpes, IRD, CNRS, IGE, F-38000 Grenoble, France

7 <sup>‡</sup>Université de Grenoble Alpes, CNRS, Inria Project/Team AIRSEA, Grenoble, France

8 <sup>§</sup>Université de Grenoble Alpes, IRD, CNRS, IGE (UMR 5001), F-38000 Grenoble, France

## ABSTRACT

9 The Ecuadorian Andes are characterized by a complex spatiotemporal vari-  
10 ability of precipitation. Global circulation models do not have sufficient hor-  
11 izontal resolution to realistically simulate the complex Andean climate and  
12 in situ meteorological data are sparse; thus, a high-resolution gridded precip-  
13 itation product is needed for hydrological purposes. The region of interest  
14 is situated in the center of Ecuador and covers three climatic influences: the  
15 Amazon basin, the Andes and the Pacific coast. Therefore, regional climate  
16 models are essential tools to simulate the local climate with high spatiotempo-  
17 ral resolution; this study is based on simulations from the Weather Research  
18 Forecasting (WRF) model. The WRF model is able to reproduce a realis-  
19 tic precipitation variability in terms of the diurnal cycle and seasonal cycle  
20 compared to observations and satellite products; however, it generated some  
21 nonnegligible bias in the region of interest. We propose two new methods for  
22 precipitation bias correction of the WRF precipitation simulations based on in  
23 situ observations. One method consists of modeling the precipitation bias with  
24 a Gaussian process metamodel. The other method is a spatial adaptation of the  
25 cumulative distribution function transform approach, called CDF-t, based on  
26 Voronoï diagrams. The methods are compared in terms of precipitation occur-  
27 rence and intensity criteria using a cross-validation leave-one-out framework.  
28 In terms of both criteria the Gaussian process metamodel approach yields bet-  
29 ter results. However, in the upper parts of the Andes ( $>2000$  m), the spatial  
30 CDF-t method seems to better preserve the spatial WRF physical patterns.

## 31 **1. Introduction**

32 The Andes Cordillera forms a natural orographic barrier along the western coast of the South  
33 American continent, causing a complex spatiotemporal distribution of precipitation (e.g., Garreaud  
34 1999; Espinoza et al. 2009). The spatial precipitation distribution is characterized by strong ele-  
35 vational gradients, with the eastern and western sides of the Andes exhibiting higher precipitation  
36 values than the high-elevation mountain peaks where the climate is relatively dry (see Fig. 1; e.g.,  
37 Bendix and Lauer 1992). We distinguish three different climate regions in Ecuador: the Pacific  
38 coast, the Andes and the Amazon. Each side of the Andes is influenced by different atmospheric  
39 processes. The western plains of Ecuador are strongly influenced by the sea surface temperature  
40 variability of the Pacific Ocean. For instance the occurrence of ENSO (El Niño Southern Oscil-  
41 lation) events on an interannual timescale produces strong temperature and precipitation anomalies  
42 and significant socioeconomic issues (e.g., Rossel et al. 1999; Vuille et al. 2000; Rabatel et al.  
43 2013; Vicente-Serrano et al. 2017). In the eastern part of the Andes, the moisture mainly comes  
44 from the Atlantic Ocean and water recycling through evapotranspiration over the humid Amazo-  
45 nian rainforest plains. In the Andes the interannual precipitation variability is influenced by both  
46 tropical Pacific and Atlantic sea surface temperature anomalies (e.g. Vuille et al. 2000; Espinoza  
47 et al. 2011). On the seasonal timescale, the precipitation variability is very complex and can be  
48 characterized by one or two rainfall seasons. On the Pacific coast, one rainfall season is generally  
49 described (e.g., Bendix and Lauer 1992; Vicente-Serrano et al. 2017) whereas two rainfall seasons  
50 are observed in most parts of the Andes (e.g., Bendix and Lauer 1992; Vicente-Serrano et al. 2017)  
51 and in the Amazon plains of Ecuador (e.g., Laraque et al. 2007; Espinoza et al. 2009) and these  
52 rainfall seasons occur from March to May, and from October to December. At the regional scale,  
53 these two periods correspond to the two annual transition phases of the American monsoon cy-

54 cle, between the mature phases of the North American monsoon system (June to August) and the  
55 South American monsoon System (December to February; e.g., Vera et al. 2006). However, there  
56 are large disparities at the local scale (e.g., Laraque et al. 2007), due to local atmospheric processes  
57 associated with the complex orography of the Andes. The slope of the eastern part of the Andes  
58 is also characterized by the presence of local maximum precipitation values called “hotspots” (Es-  
59 pinoza et al. 2015), and in these regions the elevational gradients are nonlinear, with the maximum  
60 values situated between 500 and 2000 m. Thus, the spatiotemporal variability of precipitation is  
61 quite complex in this area, making it challenging to characterize with statistical models.

62 The Antisana glacier culminates at approximately 5760 m, and is located close to the Amazon  
63 slope on the eastern side of the Ecuadorian Andes. Quito, the capital of Ecuador, is situated ap-  
64 proximately 50 km further west closest to the Pacific side of the Andes. The Antisana region is  
65 an important water reserve for the population (Chevallier et al. 2011; Hall et al. 2012; Basantes-  
66 Serrano 2015; Buytaert et al. 2017; Pouget et al. 2017). The water resources in this region depend  
67 in part on the Antisana glacier, whose mass balance is influenced by several factors, including  
68 precipitation variability, (e.g., Favier et al. 2004; Sicart et al. 2011). Recently, a dry trend has been  
69 identified in the western Amazon during the last decades, including in the Ecuadorian Amazon,  
70 and is particularly strong during austral winter (Espinoza et al. 2009). However, the station den-  
71 sity in the Andes is low relative to the complexity of the topography, so the spatial distribution of  
72 precipitation is poorly understood (Buytaert et al. 2006; Rollenbeck and Bendix 2011; Manz et al.  
73 2017). Precipitation in the highest elevation zones is particularly uncertain, as there are few sta-  
74 tions located above 3500 m (see Fig. 1 and Table 1). Thus, to understand how the water resources  
75 of this region might change in the future, an essential first step is to establish a spatially complete  
76 picture of current-day precipitation.

77 In the Andes, global circulation models (GCMs) do not have sufficient horizontal resolution to  
78 realistically simulate the complex Andean climate (IPCC 2013). For this reason, regional climate  
79 models (RCMs) are essential for simulating the local climate with high spatiotemporal resolution.  
80 In this study the Weather Research Forecasting (WRF) model is used. Several previous studies  
81 have used the WRF model in the Andes, including, the works developed by Ochoa et al. (2014),  
82 Ochoa et al. (2016), Mourre et al. (2016), Junquas et al. (2017). Mourre et al. (2016) and Ochoa  
83 et al. (2016), compared WRF simulations to rainfall products derived from satellite products and  
84 in situ stations in the Peruvian Andes and in the Ecuadorian Andes, respectively. Whereas the  
85 WRF model is able to reproduce a realistic precipitation variability in terms of the diurnal cycle  
86 and seasonal cycle compared to observations and satellite products, these studies have also iden-  
87 tified quantitative precipitation biases in the Andes, in terms of intensity (precipitation amounts)  
88 and occurrence (rainy/non-rainy days). Thus, before using WRF outputs in climate impact stud-  
89 ies, the application of bias correction methods of the simulated precipitation is crucial (Vrac and  
90 Friederichs 2015).

91 In the Andes the orographic gradients play an important role on the atmospheric processes. The  
92 WRF model is able to reproduce two different spatial-scale mechanisms associated with the pre-  
93 cipitation distribution (e.g., Ochoa et al. 2014; Mourre et al. 2016; Junquas et al. 2017): local-scale  
94 (e.g., valley and mountain winds) and synoptic-scale (e.g. low-level jet east of the Andes) circula-  
95 tion. The three previously defined climate regions in Ecuador (Pacific coast, Andes, and Amazon)  
96 are differently affected by these processes. Therefore, it is reasonable to think that the precipi-  
97 tation biases simulated by the WRF model could also be affected differently by these different  
98 atmospheric processes in each climate region. Thus, it is crucial to develop different statistical  
99 methods taking into account this particular climate distribution, by focusing on the spatial precip-  
100 itation bias distribution.

101 Our main objective in this study is to statistically correct the WRF outputs of precipitation at the  
102 daily timescale, during the two-year period in the Antisana region (2014-2015). Considering the  
103 unique climate characteristics of the region and the few observations, we decided to develop new  
104 methods by adapting statistical tools from the literature. The first method consists of modeling the  
105 precipitation bias with a Gaussian process. This approach is also known as kriging in geostatistics  
106 and takes into account the spatial statistical structure of a variable of interest. Several studies have  
107 been developed to correct the precipitation bias based on Gaussian process models. For example,  
108 Hanchoo Wong et al. (2012) developed a bias correction of radar rainfall based on the kriging ap-  
109 proach in Thailand, Müller and Thompson (2013) performed a bias adjustment of satellite rainfall  
110 in Nepal, they used kriging to interpolate precipitation from in situ measures, and Mourre et al.  
111 (2016) performed a precipitation interpolation based on kriging using as external drift the WRF  
112 simulation in the Cordillera Blanca (Peru). In Ecuador, the kriging method was already tested as a  
113 spatial interpolation method on the Pacific coast (Ochoa et al. 2014) and in the highlands (Buytaert  
114 et al. 2006) with in situ stations. They showed that using kriging interpolation with elevation as  
115 the external drift significantly improved the performance of the method in these regions. In our  
116 study, the novelty of our approach is to apply kriging to the daily precipitation bias instead of  
117 the precipitation amount, as is classically done. We will show that this adaptation is particularly  
118 useful in regions where different precipitation regimes coexist, as is the case in our region with the  
119 Amazon and Andean climates.

120 The second approach generalizes the quantile-quantile method (e.g., Déqué 2007) and is based  
121 on the cumulative distribution function transform (hereafter CDF-t) with Singularity Stochastic  
122 Removal approach developed by Vrac et al. (2016). The probabilistic approach “cumulative dis-  
123 tribution function-transform” (hereafter CDF-t) has been used in many applications, including  
124 correction of the punctual daily wind speed and regional downscaling (e.g., Michelangeli et al.

125 2009; Vrac and Vaittinada 2017). This approach has also been applied to correct the biases of dif-  
126 ferent atmospheric variables; such as temperature, precipitation and relative humidity (e.g., Colette  
127 et al. 2012; Vrac et al. 2012). Vrac et al. (2016) proposed a modification of the CDF-t method for  
128 bias correction, specifically designed for precipitation, called “Singularity Stochastic Removal”  
129 (hereafter SSR). The motivation for developing an approach specialized for precipitation is be-  
130 cause of its particular property in terms of a large number of zeros (non-precipitation events) in a  
131 daily time step. The principal advantage of this approach is that it allows us to correct biases while  
132 avoiding separating the correction in terms of occurrence (number of rainy days) and intensity of  
133 precipitation (quantity of precipitation). Previously, the SSR approach has been used to correct  
134 heat waves over France, as implemented by Ouzeau et al. (2016), and in a multivariate quantile  
135 mapping bias correction context to correct surface meteorological variables from regional climate  
136 model outputs across a North American domain (Cannon 2017).

137 The CDF-t is a variant of the quantile-mapping technique, which consists of mapping a model  
138 output  $x$  with cumulative distribution function (CDF)  $F_X$ , to its corresponding observation  $y$  with  
139 CDF  $F_Y$ , through a function  $T$  (Piani et al. 2010; Vrac et al. 2016). More precisely, considering  
140  $T = F_{Y^{-1}} \circ F_X$ , where  $F_Y^{-1}$  is the generalized inverse of  $F_Y$ , thus we obtain  $y = T(x)$  in the sense  
141 that  $F_Y = F_{T(X)}$  ( $y$  is distributed as  $T(x)$ ).

142 Then,  $T$  can be modeled either parametrically or nonparametrically, and estimated from the data.  
143 If the data are stationary and consist of  $n$  independent realizations of  $x$  (*resp.*  $y$ ), then  $T$  can be  
144 estimated by  $F_{Y,n}^{-1} \circ F_{X,n}$  with  $F_{X,n}$  (*resp.*  $F_{Y,n}$ ) representing the empirical CDF of  $x$  (*resp.*  $y$ ). In  
145 that case, the procedure is known as the empirical mapping procedure.

146 Usually, the CDF-t approach is used to correct model predictions for future periods. We propose  
147 in this paper a spatial adaptation of the CDF-t approach from a point scale correction to a correction  
148 on any grid point, partitioning the region of interest using a Voronoi diagram of the stations (see



149 Section 3 for more details). Voronoï diagrams, also known as Thiessen polygons, have been widely  
150 used in meteorological applications. As for example in (Buytaert et al. 2006), spatial interpolation  
151 of precipitation with Thiessen polygons in the south Ecuadorian Andes is performed. In (Ly et al.  
152 2011), spatial interpolation is performed in the Ourthe and Ambleve catchments in Belgium.

153 This paper is organized as follows. In Section 2 we present the data used in the study and the  
154 WRF simulation characteristics. In Section 3 we describe the new methods of precipitation bias  
155 corrections. We analyze the results and the intercomparison between them in Section 4. Finally,  
156 we summarize the main results and conclude in Section 5.

## 157 **2. Data**

### 158 *a. In situ data*

159 We use daily data from 26 in situ meteorological stations with elevations that range from 1110  
160 to 4812 m, during the 2014-2015 period. All of the stations with the exception of station 26 were  
161 installed and are managed by the National Service of Meteorology and Hydrology of Ecuador  
162 (INAMHI). The stations from the INAMHI are of a tipping bucket type, and the highest is station  
163 17 at 4009 m. The INAMHI data quality is routinely controlled, using the standard procedures in  
164 use by Met services worldwide. Based on in situ observations, Francou et al. (2004) determined  
165 the snowfall/rainfall limit at 4900 m close to the snout of the Ecuadorian glaciers. This elevation  
166 corresponds to a temperature threshold equal to  $0.5^{\circ}$ . All the stations from the INAMHI network  
167 are situated below 4009 m, so we do not observe snowfall at the INAMHI' stations (see Table 1).  
168 Station number 26, belonging to the SNO GLACIOCLIM, is situated at 4812 m. Snowfall is fre-  
169 quent at this altitude, and some care must be taken to reduce the uncertainty of the measurement.  
170 First, the gauge should be adapted to measure any type of hydrometeor (solid or liquid). Second

171 the problem of undercatch, principally caused by wind must be addressed. In the present study,  
172 we used data issued from Geonor gauge; this type the gauge is a weighting device specifically  
173 designed to measure all the hydrometeor types and is suitable for both solid and liquid precip-  
174 itation. To reduce the problem of under catch principally caused by wind effects, we use the  
175 correction proposed by Forland et al. (1996), depending on the air temperature and wind velocity.  
176 The detailed procedure for the data treatment is provided in Wagnon et al. (2009). At the regional  
177 scale for the whole Andean zone defined in the study, the snowfalls are not very important if one  
178 considers the surface of the ground located higher than 4900 m (less than 1% of the total area).

179 Figure 1 shows a map with the locations of the stations. The study area is divided into three  
180 regions corresponding to three regions of Ecuador (see Section 1): the region located on the Pacific  
181 coast side (hereafter Pacific coast) formed by stations 2, 22 and 23; the Amazon formed by stations  
182 19 and 25; and the Andes, formed by the remaining ones (21 stations). Most stations are located  
183 in the Andes (81%), with 11% on the Pacific coast and 8% in the Amazon. Table 1 presents  
184 a description of the location and accumulated precipitation for the period 2014-2015 for each  
185 meteorological station. The meteorological stations located in the Amazon registered the highest  
186 total precipitation values (with total precipitation greater than 6000 mm in the two years).

187 Because very few in situ stations were available in this region, we included two stations (num-  
188 bers 12 and 18) situated very close to the limits of the domain (less than 4km in latitude) and at  
189 the same elevation. Because the main idea of this study is to test bias correction methodologies,  
190 we decided to include these two stations for these tests, by assigning their corresponding model  
191 grid latitudes as 0.01 to avoid the boundary zone of the model (see Section 2c). Originally, sta-  
192 tion number 12 was situated at the latitude 0.05 and station number 18 was situated at the latitude  
193 0.03, corresponding to 4km and 2km from the model limit, respectively. We performed statisti-  
194 cal analysis (not shown) that confirmed that these precipitation timeseries of the WRF-1km grid

195 points are significantly correlated to the corresponding in situ stations timeseries in terms of oc-  
196 currence and intensity, even considering some km lags, highlighting that precipitation variability  
197 is homogeneous in this small region.

#### 198 *b. CHIRPS satellite product*

199 Satellite-based rainfall estimates such as CHIRPS (Climate Hazard Group 1981; Funk et al.  
200 2015) provide an opportunity for a wide range of hydrological applications, from water resource  
201 modeling to monitor of extreme events, such as droughts and floods. CHIRPS is a continental  
202 rain data set that combines satellite and rain gauges data with a spatial resolution of  $0.05^\circ \times 0.05^\circ$ .  
203 CHIRPS uses the global cold cloud duration (CCD) as a thermal infrared method to estimate  
204 the global precipitation. Then, the product TRMM-3B42 V7 is used to calibrate the precipita-  
205 tion estimated by the global CCD. Finally, gauge stations are used to calibrate the estimations of  
206 precipitation (Paccini et al. 2018). Recent studies note that, at daily time steps or for arid envi-  
207 ronments, important biases exist in these rainfall estimations (Herold et al. 2017; Paredes-Trejo  
208 et al. 2017). Furthermore, Bai et al. (2018) used the CHIRPS product in mountainous regions in  
209 China and concluded that the ability of CHIRPS to detect snowfall was limited. More generally,  
210 this product has known biases, including underestimation of extreme precipitation events (Funk  
211 et al. 2015). In our study, the use of the precipitation satellite product CHIRPS for the period of  
212 2014-2015 allows for a graphical evaluation of the corrected gridded precipitation products. In-  
213 deed, this product provides good spatial patterns at seasonal or annual scales (Zambrano-Bigiarini  
214 et al. 2017). Thus, we use this dataset for a spatially complete qualitative comparison, but only in  
215 an approximate sense.

216 *c. WRF simulation and its biases*

217 The WRF model version 3.7.1 (Skamarock et al. 2008) is used to simulate high-resolution pre-  
218 cipitation for the period 2014-2015 in the studied region. The model is nonhydrostatic and uses  
219 a terrain-following vertical coordinate (sigma). The WRF model is established with 4 one-way  
220 nested domains (27 km, 9 km, 3 km, and 1 km; see Fig. 2). The outer domain is forced by the  
221 NCEP-FNL reanalyses ( $1^\circ \times 1^\circ$ ). The simulation outputs of the innermost domain (1km  $\times$  1km)  
222 are used for this study. The in situ data for each station are compared with the closest 1km grid  
223 point of the WRF simulation. As mentioned in Section 2a, for two stations (numbers 12 and 18),  
224 the closest inner-domain gridpoint was considered to avoid the northern lateral boundary zone of  
225 the model (5 gridpoints of specified and relaxation zone; see Fig. 2d). The four domains are con-  
226 figured with 30 sigma levels in the atmosphere, and the top model is configured at 50hPa, as it was  
227 already used in previous studies in the tropical Andes (Junquas et al. 2017; Moya-Álvarez et al.  
228 2018). The output time resolutions are 6 h, 3 h, 3 h, and 1 h for the first, second, third and fourth  
229 domain, respectively.

230 Some options for the dynamical and physical parameterizations were previously tested to pro-  
231 vide better precipitation results in the region of interest (not shown). The chosen parameterizations  
232 are described as follows. We use the Yonsei University scheme (Hong et al. 2006) as the plane-  
233 tary boundary layer option, with a wind topographic correction for the complex surface terrain  
234 (Jiménez and Dudhia 2012), that has already been used in previous studies using the WRF model  
235 in the Andes (Mourre et al. 2016; Junquas et al. 2017). The Microphysical parameterization is  
236 from Lin et al. (1983), and the cumulus scheme is from Grell and Dévényi (2002). Preliminary  
237 tests have been performed with other parameterizations, and this configuration was chosen be-  
238 cause the precipitation bias in the Andes was less pronounced (not shown). We decided to employ

239 the cumulus parameterization in the four domains because in our tests, the convection-permitting  
240 experiment (no cumulus scheme activated at 3 km and 1 km) exhibit the greatest bias with a precip-  
241 itation overestimation of more than 300% in the Andes compared to station data (not shown). This  
242 result confirms the results of a recent paper that did not find precipitation improvements using con-  
243 vection permitting in WRF forecasting simulations in the Peruvian Andes region (Moya-Álvarez  
244 et al. 2018).

245 As the surface model, we use the Noah multi-physics model with a snow option (snf\_opt=2;  
246 Niu et al. 2011; Yang et al. 2011) as previously tested in the Cordillera Blanca in Peru (Mourre  
247 et al. 2016). The longwave and shortwave radiation options are RRTM (Mlawer et al. 1997)  
248 and Dudhia scheme (Dudhia 1989), respectively. The surface layer parameterization is MM5  
249 similarity (Paulson 1970). We used the SRTM (Shuttle Radar Topography Mission; Farr et al.  
250 2007) digital elevation model instead of the USGS (United States Geological Survey) data as  
251 topographic forcing, as suggested by preliminary studies.

252 We compared the in situ observations and the WRF simulations and found that they are biased  
253 (see Figure 3). The mean bias per station is  $1.89 \text{ mm day}^{-1}$  during the two years with, a minimum  
254 of  $0.04 \text{ mm day}^{-1}$  (achieved at station 17) and a maximum of  $9.72 \text{ mm day}^{-1}$  (station 2). During  
255 2014, the mean relative bias is an underestimation of 20%, its maximum underestimation is 80%  
256 (station 2) and the maximum overestimation is registered at station 6 (47%). During 2015, the  
257 mean relative bias is an underestimation of 42%, with a maximum underestimation of 85% (station  
258 2), and the maximum overestimation is 23%, registered at station 13.

259 The biases are more evident in the Amazon, where underestimations of approximately 8.20 and  
260  $6.96 \text{ mm day}^{-1}$  are obtained for stations 19 and 25. The biases of the 2014 and 2015 periods are  
261 slightly different because during 2014, there is strong overestimation of the simulated precipitation  
262 at some stations of the Andes (stations 3, 6 and 18), in contrast to 2015, when underestimation are

263 obtained for most of the stations (except for 13 and 17). It is clear from these figures that the spatial  
264 bias variability strongly depends on the period under consideration. The spatial distribution of the  
265 bias in 2015 (Fig. 3b) appears more homogeneous than the one in 2014 (Fig. 3a). This contrast is  
266 explained by different local influences of atmospheric processes on the interannual variability in  
267 the region of the Andes. The interannual variability is part of the complexity of the spatiotemporal  
268 precipitation distribution in the Ecuadorian Andes. Note that some biases identified in the WRF  
269 simulations could potentially be caused by errors in the in situ observations.

### 270 **3. Bias correction methods**

271 Two methods for bias correction are adapted and analyzed in this study. The first one is to model  
272 the WRF bias with a Gaussian process model, also known as kriging, and the second one is a time  
273 series preprocessing and spatial adaptation of the CDF-t method. The methods are described in  
274 this section (parts a and b), and we present the criteria used to evaluate the performance of the two  
275 approaches that are used (part c) in the results section (Section 4).

#### 276 *a. Gaussian process modeling*

277 The first method implemented is to model the WRF biases using a Gaussian process model;  
278 Figure 4 presents the flowchart of our method. In the following, we define bias as follows:

$$\text{BIAS} = \text{WRF simulation} - \text{Observation.} \quad (1)$$

279 Then, at in each point where there is no observation, we obtain a prediction of the bias ( $\widehat{\text{BIAS}}$ )  
280 and we compute the predicted precipitation ( $\widehat{\text{Precip.}}$ ) value as follows:

$$\widehat{\text{Precip.}} = \text{WRF simulation} - \widehat{\text{BIAS.}} \quad (2)$$

281 We refer to the work of Marrel et al. (2008) for a presentation of Gaussian process modeling  
 282 (also see the work of Oakley and O’Hagan (2002)). Consider that  $n$  observations of a phenomenon  
 283 are registered at  $n$  different locations (for example, the bias precipitation registered in  $n$  stations  
 284 of the region under study). We consider in the following that each observation  $y(x)$  is registered  
 285 at point  $x = (x_1, x_2) \in \mathbb{R}^2$  (the coordinates of  $x$  correspond to the longitude and latitude of the  
 286 station), endowed with the usual Euclidean distance. The set of points where the observations are  
 287 collected is denoted by  $x_s = (x^{(1)}, \dots, x^{(n)})$  with  $x^{(1)}, \dots, x^{(n)} \in \mathbb{R}^2$  (in our study, each  $x$  corresponds  
 288 to a station). The set of observations of the phenomenon is denoted by  $y_s = (y^{(1)}, \dots, y^{(n)})$  with  
 289  $y^{(i)} = y(x^{(i)})$ . The Gaussian process modeling consists of representing  $y(x)$  as a realization of a  
 290 random function  $Y(x)$  such that:

$$Y(x) = f(x) + Z(x) + U(x), \quad (3)$$

291 where  $Z(x)$  is a centered stationary Gaussian process;  $U(x)$  represents the noise in the obser-  
 292 vations and is a centered stationary Gaussian process with a diagonal covariance structure; and  
 293  $f(x)$  is a deterministic function that represents the tendency, also known as the external drift, lin-  
 294 ear combinations of longitude, latitude and elevation are commonly used. More generally, it is  
 295 constructed as a finite linear combination of  $k$  elementary functions:

$$f(x) = \sum_{j=0}^k \beta_j f_j(x) = F(x)\beta \quad (4)$$

296 where  $\beta = (\beta_0, \dots, \beta_k)^T$  is the regression parameter vector and  $F(x) = (f_0(x), \dots, f_k(x))$ . The  
 297 function  $f(x)$  allows the addition of an external drift into the modeling, and this is advantageous  
 298 because it allows a nonstationary global modeling framework; in other words the variable  $Y$  does  
 299 not need to be stationary but the variable  $Z$  is assumed to be stationary.

300 The Gaussian centered process  $Z(x)$  has the following a covariance function:

$$\text{Cov}(Z(x), Z(u)) = K(x - u) = \sigma^2 R(x - u), \quad (5)$$

301 where  $x, u \in \mathbb{R}^2$  (in our application,  $u$  also corresponds to the coordinates longitude and latitude  
302 of a station),  $\sigma^2$  is the variance of  $Z$ , and  $R$  is its correlation function. The process  $Z$  is stationary  
303 because it is considered that its correlation function depends only on the difference between  $x$  and  
304  $u$ .

305 In this study, we used the Matérn covariance functions because they are stationary and com-  
306 monly used in spatial statistics studies due to their flexibility (Paciorek and Schervish 2006); and  
307 they are defined follows:

$$K(x, u) = \frac{1}{\Gamma(\nu)2^{\nu-1}} \left[ \frac{\sqrt{2\nu}}{\kappa} |x - u| \right]^\nu K_\nu \left( \frac{\sqrt{2\nu}}{\kappa} |x - u| \right), \quad (6)$$

308 where  $K_\nu$  is the modified Bessel function of second kind of order  $\nu > 0$ ,  $\kappa$  is a positive param-  
309 eter that represents the characteristic length scale and  $\Gamma$  is the Gamma function (Rasmussen and  
310 Williams 2005). The Euclidean distance, written as  $|x - u|$ , is used.

311 The aim of Gaussian process modeling is to estimate the prediction of  $Y$  for a new grid point  
312  $x^*$ . In our study, Gaussian process modeling is applied to estimate the bias at grid points at which  
313 there is no station. In our application, first, the bias is computed for annual averages to assess  
314 the accuracy of four models constructed by the combination of three commonly used drifts (lon-  
315 gitude; latitude and elevation) and to choose one of them. Here-after, they are referred to as the  
316 GP model with drift longitude, latitude and elevation (GP+longitude+latitude+elevation), the GP  
317 model with drift longitude and latitude (GP+longitude+latitude), the GP model with drift longi-



318 tude (GP+longitude) and the GP model with drift elevation (GP+elevation). Then, we computed  
319 the daily bias using the GP with the selected drift to obtain a corrected daily precipitation product.

### 320 *b. Spatial adaptation of the CDF-t method*

321 Historically, the CDF-t method has been applied as a statistical downscaling method and to  
322 correct future time series from GCMs outputs. In our study, the CDF-t method aims at relating  
323 CDFs of a climate variable (here the precipitation) from WRF simulation outputs to the CDF of  
324 this variable from the in situ observation. However, instead of applying the correction over future  
325 time series, we adapt the method to correct the gridpoints of the domain, even where there is no  
326 associated observation. We call this approach a spatial adaptation of the CDF-t method. The main  
327 idea is to partition the region under study (see Fig. 1) into neighboring sub-regions, in such a  
328 manner that every subregion contains a station. We are going to assume that the precipitation  
329 biases in these subregions behave similarly.

330 To define the subregions, we divide the region using a partition based on Voronoï diagrams.  
331 This method is a simple way to define subregions, that is, applicable to any mountain region with  
332 few complete in situ precipitation time series, as in our case. In addition, as there is no spatial  
333 smoothing, it has the advantage of conserving the spatial coherence of the physical processes  
334 simulated by the WRF model inside each subregion. Another advantage of using Voronoï diagrams  
335 is their simplicity and low computational cost, which allow them to be used with large volumes of  
336 data.

337 At a given station  $s$ , let us denote  $X_t^s$  the model simulation at time  $t$  and  $Y_t^s$  its corresponding  
338 observation. The time series under study are nonstationary and autocorrelated, hence the standard  
339 empirical mapping cannot be used directly (see Section 1). Indeed, we performed the usual statisti-  
340 cal hypothesis testing procedures to detect nonstationarity: the Kwiatkowski Phillips Schmidt Shin

341 test (Kwiatkowski et al. (1992); KPSS where  $H_0$  : The time series is stationary), and the Mann-  
342 Kendall test (Mann (1945); Kendall (1948);  $H_0$  : The time series do not have a monotonic trend).  
343 The p-value results of the KPSS and Mann-Kendall tests are less than 0.1 and 0.05, respectively,  
344 for all the observed and simulated time series, meaning that the time series are nonstationary due  
345 to a unit root (autocorrelation close to 1) and dependent. It is thus necessary to perform differen-  
346 tiation and subsampling. More precisely, we applied the following preprocessing: we calculated  
347  $\Delta X_t^s = X_t^s - X_{t-1}^s$  and  $\Delta Y_t^s = Y_t^s - Y_{t-1}^s$  to stationarize the time series, and we used subsampling to  
348 eliminate the autocorrelation. The manner in which we performed subsampling is the following:  
349 as the autocorrelation length was estimated to  $k = 2$ , we skipped one observation out of two.

350 As already mentioned, the main issues of bias correction for precipitation data is the treatment  
351 of the rainfall occurrences. To solve this issue Vrac et al. (2016) proposed changing the null  
352 precipitation data for a uniform distribution. In our case, we corrected the differentiated time  
353 series of precipitation, thus we adapted the SSR to our framework. More precisely, we performed  
354 the following steps on our data:

355 **Step 1.-** Determinate a threshold  $\theta$  such that:

$$\theta = \min \left( \inf_{t \geq 1, |\Delta X_t^s| \neq 0} \{|\Delta X_t^s|\}, \inf_{t \geq 1, |\Delta Y_t^s| \neq 0} \{|\Delta Y_t^s|\} \right) \quad (7)$$

356 **Step 2.-** Each time  $\Delta X_t^s = 0$  (*resp.*,  $\Delta Y_t^s = 0$ ), we simulate a value  $v$  from the uniform distri-  
357 bution  $U[-\theta, \theta]$  and we replace  $\Delta X_t^s$  (*resp.*,  $\Delta Y_t^s$ ) with the sampled value.

358 Such a step avoids separating the correction of the occurrences from the one of the intensities.  
359 (Vrac et al. 2016).

360 **Step 3.-** Nonparametrically estimate the mapping  $F_{\Delta Y^s}^{-1}(F_{\Delta X^s})$  using e.g., the R package de-  
 361 veloped by Vrac (2015) (see also Michelangeli et al. (2009)). The mapping will be denoted  
 362 by  $\hat{T}^s$  in the following.

363 In this paper, we do not aim at correcting the bias for future predictions, but we want to correct  
 364 the bias at any grid point where no observation is available.

365 Therefore, we construct a Voronoï diagram based on seeds composed with the stations. For each  
 366 station (seed) there is a corresponding region consisting of all points closer to that seed than to any  
 367 other. In this manner, we obtain as many regions as the initial number of stations, let us say  $\mathcal{S}$ .  
 368 For  $s = 1, \dots, \mathcal{S}$ , we construct following Step 3 a mapping  $\hat{T}^s$  from time series  $X_t^s$  and  $Y_t^s$ . We then  
 369 assume that the mapping is constant on each Voronoï cell. We then proceed with the following  
 370 steps:

371 **Step 4.-** At any grid point, let us consider the closest station  $s$ . We consider the time series  
 372  $\Delta Z_t$ , where  $Z_t$  denotes the WRF simulation at time  $t$ . If the grid point coincides with station  
 373  $s$ , then  $Z_t = X_t^s$ . We apply the following bias correction:

$$V_t = Z_{t-1} + \hat{T}^s(\Delta Z_t) \quad (8)$$

374 **Step 5.-** The bias corrected data  $V_t$  lower than  $\theta$  are set to 0. This step allows us to recover  
 375 the correct occurrence of 0 precipitation.

376 As an illustration of the procedure, Figure 5 shows for 3 stations (one for each region) the orig-  
 377 inal time series ( $X_t$  and  $Y_t$ ), the differentiated ones ( $\Delta X_t$  and  $\Delta Y_t$ ) and the CDFs of the observation,  
 378 simulation and CDF-t correction (more details are presented in Section 4b).

379 *c. Evaluation criteria to compare the two approaches*

380 To compare the accuracy of the rainfall products created by these two methods (Gaussian pro-  
381 cess modeling and spatial CDF-t approach), we have computed various criteria concerned with  
382 occurrences (number of rainy/non-rainy days) and intensity of precipitation (precipitation quan-  
383 tity). These criteria are commonly used in the literature; for example, they were used in the works  
384 of Maussion et al. (2011), Ochoa et al. (2014), Moure et al. (2016) and, Vrac et al. (2016).

385 CRITERIA RELATED TO THE OCCURRENCE

386 A day is considered as a “rainy day” if its daily precipitation value is greater than 1 mm day<sup>-1</sup>.  
387 Note that other threshold values were tested, but the best agreement between the WRF model  
388 and in situ observations was obtained with 1 mm day<sup>-1</sup> (not shown). In the following, several  
389 measures that depend on the following four major parameters are used:

390 **True Positive (TP):** Rainy day identified by WRF as a rainy day.

391 **True Negative (TN):** Non-rainy day identified by WRF as a non-rainy day.

392 **False Positive (FP):** Non-rainy day identified by WRF as a rainy day.

393 **False Negative (FN):** Rainy day identified by WRF as non-a rainy day.

394 The false alarm rate (FAR) is defined as the incorrect number of rainy days simulated divided  
395 by the total number of rainy days simulated:

$$FAR = \frac{\#FP}{\#FP + \#TP} \quad (9)$$

396 The probability of detection (POD) is defined as the ratio between the number of rainy days  
397 simulated correctly and the total number of rainy days observed:

$$POD = \frac{\#TP}{\#TP + \#FN} \quad (10)$$

398 The probability of false detection (PODF) is the ratio between the number of rainy days incor-  
 399 rectly simulated to the number of non-rainy days of the observation:

$$PODF = \frac{\#FP}{\#FP + \#TN} \quad (11)$$

400 And finally, the Heidke skill score (HSS) is calculated as:

$$HSS = \frac{S - S_{ref}}{1 - S_{ref}}, \quad (12)$$

401 where  $S = \frac{\#TP + \#TN}{n}$  and  $S_{ref} = \frac{(\#TP + \#FP)(\#TP + \#FN) + (\#FP + \#TN)(\#FN + \#TN)}{n^2}$ . It could be interpreted  
 402 as the ability of the simulation to be better or worst than a random simulation. A perfect product  
 403 should have a FAR value of 0, a POD value of 1, a 0 PODF value and an HSS value of 1.

#### 404 CRITERIA RELATED TO THE INTENSITY

405 The following criteria are used to evaluate gridded products accuracy in terms of intensity: the  
 406 Kolmogorov-Smirnov test (KS) is a nonparametric test to compare two distributions; the maximal  
 407 difference between them is calculated. The Spearman correlation coefficient, the root mean square  
 408 error (RMSE) and the mean bias are computed. It is important for the precipitation also to know  
 409 the percentage of data that is greater than the 0.95 percentile of the observation and in the case  
 410 of a good precipitation product, it should be close to 5% (here-after referred as Q<sub>95</sub>). Finally, the  
 411 predictivity squared correlation coefficient Q<sub>2</sub> is computed. It measures the predictive ability of

412 the statistical model.

$$\text{mean bias} = \frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i), \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{x}_i - x_i)^2}, \quad (14)$$

$$Q_2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (\bar{x} - x_i)^2}, \quad (15)$$

413 where  $\hat{x}_i$  is the prediction of the precipitation (using one of the approaches described before)  
 414 at station  $i$ ,  $x_i$  is the observed precipitation at the same station  $i$  and  $\bar{x}$  is the observed mean.  
 415 These criteria should be computed on a set of stations independent from the ones used to learn  
 416 the statistical model. However, we used all the stations to train the model (Gaussian Process or  
 417 CDF-t); thus (13), (14) and (15) will be computed by cross-validation in the following. The leave-  
 418 one-out cross-validation consists of splitting the data into two groups: a group composed with all  
 419 the stations except one, which is used as learning sample, and another group whose sole element  
 420 is the remaining station, on which the model is validated. Then, the procedure is averaged on all  
 421 such leave-one-out splits. For example, for  $Q_2$ :

$$1 - \frac{\sum_{k=1}^n (x_k - \hat{x}_k^{-(k)})^2}{\sum_{k=1}^n (\bar{x} - x_k)^2} \quad (16)$$

422 where  $\hat{x}_k^{-(k)}$  is the prediction at station number  $k$ , when the model is trained by the  $n - 1$  remain-  
 423 ing stations.

## 424 4. Results

425 The principal results that we obtained are presented in this section. Subsections a and b are  
 426 devoted to the results for the GP modeling and spatial CDF-t. In subsection c we present the

427 intercomparison between both approaches. All the analysis and methods implementation were  
428 performed in R (R Core Team 2015).

#### 429 *a. Gaussian process modeling*

430 We implemented the GP models using the R package `gstat` developed in (Pebesma 2004; Gräler  
431 et al. 2016). We evaluated the four GP models to select an external drift using a cross-validation  
432 leave-one-out framework. Table 2 presents the cross-validation results for the four corrected pre-  
433 cipitation gridded products. All of the four proposed GP models exhibit better results than the  
434 uncorrected WRF outputs in terms of the criteria of Section 3c (mean, bias, RMS and correla-  
435 tion; see Table 2). However, in general, the GP+longitude+latitude model obtains the best results  
436 in terms of all of the criteria (bias, RMSE, correlation and  $Q_2$ ). In terms of predictability ( $Q_2$ ),  
437 the GP+longitude+latitude model exhibits the highest values, but the GP+elevation model values  
438 are not significantly different. The GP+longitude+latitude+elevation model yields the lowest pre-  
439 dictability values. Thus, this last model most likely overfits the data, whereas more parsimonious  
440 models have better predictive ability.

441 Analyzing the two years separately, it is found that the predictive ability is better in 2015 for  
442 the four models. However, for some criteria the values are not significantly different for each  
443 year, such as for RMSE values for GP+longitude and for GP+elevation. In addition, for the  
444 GP+elevation model the mean bias is higher for 2015 than for 2014. Longer periods are nec-  
445 essary to adequately analyze the choice of the external drift parameters on the results, which is  
446 beyond of the scope of this study considering that we only have available data spanning a two-  
447 year period. Therefore, we chose the GP+longitude+latitude model and used a Matérn covariance  
448 function to correct the daily precipitation by using separate daily variograms described in (Gräler  
449 et al. 2012) because this model yields the best results for both years of analysis. Figure 6a shows

450 the mean daily precipitation of the gridded products WRF and CHIRPS, and the cross-validation  
451 results of the GP compared to the mean daily precipitation of the station. Their respective linear  
452 regression lines are drawn. The  $R^2$  value of the linear regression of WRF is 0.38, that of GP is  
453 0.62 and that of CHIRPS is 0.70, which means that the results of the cross-validation of GP are  
454 better than WRF. Figures 7a,b,c show the accumulated precipitation of WRF, the GP correction in  
455 cross-validation and the precipitation registered at the three stations (Fig. 7a shows Pacific coast  
456 station 22; Fig. 7b shows Andes station 26, and Fig. 7c shows Amazon station 25). At the Pacific  
457 coast and the Andes stations, the corrections yield an overestimation of the precipitation (see Fig.  
458 7a and Fig. 7b), and at the Amazon station, the correction increases the precipitation to correct the  
459 underestimation simulated by WRF (see Fig. 7c).

#### 460 *b. Spatial CDF-t*

461 The procedure described in Subsection 3b is applied. The Voronoï diagram is calculated (see  
462 Fig.10f) and maps of mean of daily precipitation are presented in Fig. 10 (the stations), 10b  
463 (CHIRPS), 10c (WRF), 10d (GP) and 10e (CDF-t). The Voronoï diagram borders are marked in  
464 the Amazon due to the inhomogeneous distribution of the stations and also high underestimated  
465 precipitation in this region (for example, approximately  $3000 \text{ mm year}^{-1}$  at station 25). On the  
466 Pacific coast, the border of the polygon associated with station 22 is marked because it has recorded  
467 higher precipitation values. On the contrary, the polygon borders around the Andes are not visible  
468 in most of the cases because the biases in the Andes were quite homogeneous (see Figures 3a and  
469 3b). Therefore, in the Andes, the spatial CDF-t approach yields realistic results, by conserving  
470 the precipitation physical gradients simulated by WRF. A homogeneous station distribution could  
471 increase the accuracy of the method by taking into account more physical variables in addition  
472 to geometrical properties. Figure 6b shows the mean daily precipitation of the gridded products



473 WRF, CHIRPS, and CDFt and their linear regression lines. The  $R^2$  coefficient of CDFt (0.89) is  
474 better than that of WRF (0.38). Figures 7d,e,f shows the accumulated precipitation of WRF, the  
475 spatial CDF-t correction and the precipitation registered in the observation of three regions stations  
476 (Fig. 7d shows Pacific coast station 22; Fig. 7e, shows Andes station 26; and Fig. 7f Amazon  
477 station 25). At the Pacific coast station, the WRF simulation and its correction are similar; there is  
478 a slight increase in the precipitation in the correction to obtain a value closer to the observation (see  
479 Fig. 7d). The correction applied to the Andes station is also slight because the biases registered at  
480 these stations are low (see Fig. 7e). The correction for the Amazon station is more evident due to  
481 the high underestimation obtained by WRF (see Fig. 7f).

### 482 *c. Intercomparison between the two methods*

483 After analyzing separately the implementation of the spatial CDF-t approach and the GP cor-  
484 rection methods, we now present the an intercomparison between these different bias corrections  
485 using cross-validation leave-one-out. We use the criteria from the Subsection 3c to compare the  
486 two correction approaches (GP and spatial CDF-t) and WRF. The GP model used for these results  
487 is GP+latitude+longitude, as it was shown to outperform the other GP models tested in Table 2.

488 The criteria related to the occurrence are shown in Figure 8. The spatial CDF-t method yields  
489 results similar to WRF in terms of the FAR (mean of 0.47 and 0.45, respectively) and PODF criteria  
490 (spatial CDF-t has a mean of 0.21 and WRF 0.23); meanwhile the GP result is worst (mean of 0.53  
491 in FAR and 0.51 in PODF). The HSS results are similar for the three spatial products (WRF has a  
492 mean of 0.23, spatial CDF-t has 0.21 and GP has 0.22), and the HSS criterion is more stable for  
493 GP since its variance is less than those of the other products (GP has a standard deviation of 0.09,  
494 spatial CDF-t has 0.12, and WRF has 0.11, see Fig. 8d). The POD criterion is highly improved by  
495 GP (0.79 versus 0.49 for the Spatial CDF-t).

496 The results related to the precipitation intensity are shown in Figure 9. The KS criterion is  
497 improved with the spatial CDF-t (a mean of 0.21 versus 0.38 for GP) but the results exhibit a high  
498 variability (a standard deviation of 0.2, and GP has a standard deviation of 0.15). The RMSE  
499 criterion is similar for the two products (spatial CDF-t has a mean of 7.78, GP has 7.16, and WRF  
500 has 7.78). However, on the contrary, the Spearman correlation (GP has a mean of 0.38, versus 0.24  
501 for spatial CDF-t) and  $Q_{95}$  (GP has a mean value of 0.05, versus 0.04 for spatial CDF-t) values are  
502 slightly improved in GP.

503 The CHIRPS daily mean map is displayed in Figure 10b. Because of well-known quantitative  
504 biases in the tropical Andes (up to 80%; e.g., Espinoza et al. (2015)), we use only this data to visu-  
505 ally compare the spatial precipitation patterns. When visually comparing both corrected products,  
506 it seems that the GP model (Fig. 10d) is more similar to the satellite than the spatial CDF-t (Fig.  
507 10e) in the Andes, mainly due to the sharp discontinuities at the polygon borders on the eastern  
508 slope. However, in the Amazon the GP model shows a zone of maximum precipitation in the  
509 south-east of the domain that is not observed in the satellite data. However, given that, the satellite  
510 data are biased and there are no data in this part of the region, this result could be uncertain. A  
511 strong gradient of precipitation is evident in the eastern slope of the Andes in both the GP model  
512 and the satellite data. This gradient depends on the elevation and the presence of local atmospheric  
513 valleys processes (e.g., Egger et al. (2005); Junquas et al. (2017)). Previous studies have found  
514 that the WRF model is able to reproduce some local valley processes in the tropical Andes (e.g.,  
515 Mourre et al. (2016); Junquas et al. (2017)). Therefore, it is important to take into account that  
516 such WRF spatial patterns should be preserved in a bias correction method. This orographic limit  
517 is visually well represented with the GP method, compared with the satellite. Whereas the spatial  
518 CDF-t is visually unrealistic on the eastern slope of the Andes due to the polygon limits, in the An-  
519 des above 2000 m it seems to be able to conserve the spatial patterns of WRF. In addition, the CDF

520 of the WRF Antisana gridpoint is very similar to the Antisana station CDF (Fig. 5h), and the rela-  
521 tive bias is very weak (Fig. 3). We then expect, that in this particular region, no large quantitative  
522 bias correction should be applied. However, whereas the spatial CDF-t clearly exhibits very little  
523 quantitative correction in this region, the GP model exhibits increased precipitation, generating an  
524 overestimation compared to the observations (see Fig. 7a and Fig. 7b). The spatial CDF-t method  
525 seems then to be adapted to the upper parts of the Andes (above about 2000 m), where relatively  
526 low precipitation values dominate compared to the Amazon precipitation. In the contrast, it is not  
527 recommended to use the spatial CDF-t in regions where strong precipitation gradient exists.

## 528 **5. Conclusions and future work**

529 The aim of this study was to correct the WRF simulation precipitation biases in the studied  
530 region. Then, the final gridded products of precipitation will be used as external forcing data for  
531 hydrological and glaciological models to understand water resources and glaciers evolution in the  
532 Andes. Therefore, two methods of precipitation bias correction were explored and adapted: the  
533 first one consisted of modeling the daily WRF biases through Gaussian process (GP) models, and  
534 the second one was based on a spatial and time series adaptation of the CDF-t method developed  
535 by Michelangeli et al. (2009) and Vrac et al. (2016).

536 First, four GP models were proposed by using four combinations of external drifts (generally  
537 used in studies of this type, including latitude, longitude and elevation variables) to model the an-  
538 nual accumulated bias during the years 2014 and 2015. The accuracy of the GP models was tested  
539 in a cross-validation leave-one-out framework. Based on four criteria (Bias, RMSE, Correlation  
540 and  $Q_2$ ), the best model was GP with drift longitude and latitude. Thus, we chose this model to  
541 correct the daily precipitation by using separate daily variograms as it is described in (Gräler et al.  
542 2012).

543 We employed the SSR method with a time series adaptation to obtain the CDF estimations  
544 and a spatial adaptation to obtain the correction in the region. The methods were compared in  
545 terms of criteria related to the occurrence (FAR, POD, PODF, and HSS) and criteria related to  
546 the intensity (mean bias, Spearman correlation, KS, RMSE, Q2, and Q<sub>95</sub>). Compared with the  
547 WRF product, the spatial CDF-t approach did not exhibit significant changes, whereas the GP  
548 model correction increased the daily rain number and the total accumulated mean, improving  
549 (or worsening) significantly some intensity (occurrence) statistical scores. In terms of spatial  
550 distribution, when considering the entire WRF domain, including the three climate regions (Pacific  
551 coast, Andes, and Amazon), the GP correction yields a more realistic distribution than the spatial  
552 CDF-t, because of the marked polygon borders induced by this second method. However, at  
553 local scale in the Andes, the spatial CDF-t method seems to be more similar to the original WRF  
554 patterns.

555 In the Andes, the orography is an important factor that influences precipitation. Whereas, the GP  
556 model with elevation drift seems to be a good choice for mountainous regions, it was not found to  
557 be the best GP model considering our statistical scores. This could be because the majority of our  
558 observational data are from the high elevations of the Andes, above 2000 m. This result shows that  
559 above this limit, the spatial precipitation pattern is more complex than a simple orographic gradi-  
560 ent. Previous studies working with the WRF model in tropical Andes regions have demonstrated  
561 the importance of both local mountain winds and synoptic conditions (e.g., Moure et al. 2016;  
562 Junquas et al. 2017). In our study, the spatial CDF-t appears to be a bias correction method with a  
563 strong capacity for conserving the original spatial precipitation pattern (only considering the An-  
564 des above 2000 m). Therefore, depending on the bias characteristics of the WRF simulation, the  
565 region of study, and the intended application for the final product, one method or the other should  
566 be used for bias correction. If the bias correction is to be applied in a large region including various

567 climate characteristics with strong biases, the GP method would be recommended. Otherwise, if  
568 the region is a reduced domain with a relative uniform synoptic climate characteristics but strong  
569 influences of local atmospheric processes well represented by the model, the spatial CDF-t method  
570 would be preferred.

571 There is still work to be performed on the methods here presented to increase their accuracy.  
572 Thus, the perspectives of this study are the following: (i) to deeply analyze the implementation of  
573 stationary tests for a GP model, (ii) to develop the spatial CDF-t approach for a more complex spa-  
574 tialization strategy, including more than geometrical properties, as is the case for the Voronoï dia-  
575 gram. One alternative to the Voronoï diagram could be the use of a functional clustering meethod  
576 as in (Antoniadis et al. 2012) where a curve-based clustering is used to reduce the data dimen-  
577 sion for constructing a metamodel for West African monsoon. The functional clustering method  
578 has the advantage of taking into account time-point correlations of time series spatial data (Anto-  
579 niadis et al. 2012). However, available data with longer time series would be necessary to perform  
580 such an analysis in the Antisana region. These techniques could be further improved by defining  
581 climate subregions with the same climate characteristics. Unfortunately, such a subregion classi-  
582 fication would require a longer time-period and a more homogeneous in situ station distribution  
583 that what is available now. The spatial CDF-t method could also be tested and improved in other  
584 regions of the tropical Andes with a similar spatial climate complexity but with a different tem-  
585 poral variability, such as regions of the Peruvian or Bolivian Andes where only one precipitation  
586 season occurs during the year. Since the CDF-t method was originally developed for correcting  
587 future predictions, this method could be adapted to correct future simulations.

588 *Acknowledgments.* Within the CDP-Trajectories framework, this work is supported by the  
589 French National Research Agency in the framework of the “Investissements d’avenir” program

590 (ANR-15-IDEX-02). The first author M. B. H. was funded by the IRD program LMI-GREATICE  
591 and the OSUG@2020 labex. The simulations presented in this paper were performed using the  
592 Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported  
593 by the Rhône-Alpes region (GRANT CPER07\_13 CIRA), the OSUG@2020 labex (reference  
594 ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the pro-  
595 gramme Investissements d’Avenir supervised by the Agence Nationale pour la Recherche. The  
596 authors thank INAMHI (Ecuador), Luis Maisincho, SNO GLACIOCLIM and Antoine Rabatel for  
597 the in situ stations data. The authors also thank Gérémy Panthou (IGE) and Théo Vischel (IGE)  
598 for useful discussions.

## 599 **References**

- 600 Antoniadis, A., C. Helbert, C. Prieur, and L. Viry, 2012: Spatio-temporal metamodeling for  
601 West African monsoon. *Environmetrics*, **23** (1), 24–36, doi:10.1002/env.1134, URL <https://hal.archives-ouvertes.fr/hal-00551303>, special Issue: Spatio-Temporal Stochastic Modelling  
602 (METMAV).  
603
- 604 Bai, L., S. Chunxiang, L. Li, Y. Yang, and J. Wu, 2018: Accuracy of CHIRPS satellite-rainfall  
605 products over mainland china. *Remote Sensing*, **10** (3), 362, doi:10.3390/rs10030362, URL  
606 <https://doi.org/10.3390/rs10030362>.
- 607 Basantes-Serrano, R., 2015: Evolution of glaciers in the Ecuadorian Andes since the 1950s and its  
608 contribution to the study of the climate change in the inner tropics. Theses, Université Grenoble  
609 Alpes, URL <https://tel.archives-ouvertes.fr/tel-01219778>.
- 610 Bendix, J., and W. Lauer, 1992: Die Niederschlagsjahreszeiten in Ecuador und ihre klimadynamis-  
611 che interpretation. *Erdkunde*, **46** (118-134).

612 Buytaert, W., R. Celleri, P. Willems, B. D. Bièvre, and G. Wyseure, 2006: Spatial and temporal  
613 rainfall variability in mountainous areas: A case study from the south ecuadorian andes. *Journal*  
614 *of Hydrology*, **329 (3-4)**, 413–421, doi:<https://doi.org/10.1016/j.jhydrol.2006.02.031>, URL <http://www.sciencedirect.com/science/article/pii/S0022169406001144>.  
615

616 Buytaert, W., S. Moulds, L. Acosta, B. D. Bièvre, C. Olmos, M. Villacis, C. Tovar, and K. M. J.  
617 Verbist, 2017: Glacial melt content of water use in the tropical Andes. *Environmental Research*  
618 *Letters*, **12 (11)**, 114 014, URL <http://stacks.iop.org/1748-9326/12/i=11/a=114014>.

619 Cannon, A. J., 2017: Multivariate quantile mapping bias correction: an N-dimensional  
620 probability density function transform for climate model simulations of multiple variables.  
621 *Climate Dynamics*, 1–19, doi:10.1007/s00382-017-3580-6, URL <http://dx.doi.org/10.1007/s00382-017-3580-6>.  
622

623 Chevallier, P., B. Pouyaud, W. Suarez, and T. Condom, 2011: Climate change threats to environ-  
624 ment in the tropical Andes : glaciers and water resources. *Regional Environmental Change*, **11**,  
625 S179–S187, doi:10.1007/s10113-010-0177-6, URL [http://www.documentation.ird.fr/hor/fdi:](http://www.documentation.ird.fr/hor/fdi:010052664)  
626 [010052664](http://www.documentation.ird.fr/hor/fdi:010052664).

627 Climate Hazard Group, 1981: CHIRPS satellite product. [Online; accessed 27-May-2017], [http://](http://chg.geog.ucsb.edu/data/chirps/)  
628 [chg.geog.ucsb.edu/data/chirps/](http://chg.geog.ucsb.edu/data/chirps/).

629 Colette, A., R. Vautard, and M. Vrac, 2012: Regional climate downscaling with prior statistical  
630 correction of the global climate forcing. *Geophysical Research Letters*, **39 (13)**, doi:10.1029/  
631 2012GL052258, URL <http://dx.doi.org/10.1029/2012GL052258>, 113707.

632 Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthro-  
633 pogenic scenario: Model results and statistical correction according to observed values. *Global*

634 *and Planetary Change*, **57**, 16–26, doi:10.1016/j.gloplacha.2006.11.030.

635 Dudhia, J., 1989: Numerical Study of Convection observed during the Winter Monsoon  
636 Experiment Using a Mesoscale Two-Dimensional Model. *Journal of the Atmospheric*  
637 *Sciences*, **46 (20)**, 3077–3107, doi:10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2,  
638 URL [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2), [https://doi.org/10.1175/1520-0469\(1989\)046<3077:NSOCOD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2).

640 Egger, J., and Coauthors, 2005: Diurnal Circulation of the Bolivian Altiplano. Part I: Observations.  
641 *Monthly Weather Review*, **133 (4)**, 911–924, doi:10.1175/MWR2894.1, URL <https://doi.org/10.1175/MWR2894.1>,  
642 <https://doi.org/10.1175/MWR2894.1>.

643 Espinoza, J. C., S. Chavez, J. Ronchail, C. Junquas, K. Takahashi, and W. Lavado, 2015: Rainfall  
644 hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations  
645 with large-scale atmospheric circulation. *Water Resources Research*, **51 (5)**, 3459–3475, doi:  
646 10.1002/2014WR016273, URL <http://dx.doi.org/10.1002/2014WR016273>.

647 Espinoza, J. C., J. Ronchail, J. L. Guyot, G. Cochonneau, F. Naziano, W. Lavado, and P. Vauchel,  
648 2009: Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia,  
649 Colombia, and Ecuador). *International Journal of Climatology*, **29 (11)**, 1574–1594.

650 Espinoza, J. C., J. Ronchail, J. L. Guyot, C. Junquas, P. Vauchel, W. Lavado, G. Drapeau, and  
651 R. Pombosa, 2011: Climate variability and extreme drought in the upper Solimes River (western  
652 Amazon Basin): Understanding the exceptional 2010 drought. *Geophysical Research Letters*,  
653 **38 (13)**, doi:10.1029/2011GL047862, URL <http://dx.doi.org/10.1029/2011GL047862>, 113406.

654 Farr, T. G., and Coauthors, 2007: The Shuttle Radar Topography Mission. *Reviews of Geophysics*,  
655 **45 (2)**, doi:10.1029/2005RG000183, URL <http://dx.doi.org/10.1029/2005RG000183>, rG2004.



- 656 Favier, V., P. Wagnon, J. P. Chazarin, L. Maisincho, and A. Coudrain, 2004: One-year measure-  
657 ments of surface heat budget on the ablation zone of Antizana Glacier 15, Ecuadorian Andes.  
658 *Journal of Geophysical Research*.
- 659 Forland, E., and Coauthors, 1996: Manual for operational correction of Nordic precipitation data.  
660 Rep 24/96. Oslo Norwegian Meteorological Institute.
- 661 Francou, B., M. Vuille, V. Favier, and B. Cáceres, 2004: New evidence for an  
662 ENSO impact on low latitude glaciers: Antizana 15, Andes of Ecuador. *Jour-  
663 nal of Geophysical Research: Atmospheres*, **109** (D18), doi:10.1029/2003JD004484,  
664 URL [https://agupubs.  
665 onlinelibrary.wiley.com/doi/abs/10.1029/2003JD004484](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003JD004484), [https://agupubs.  
onlinelibrary.wiley.com/doi/pdf/10.1029/2003JD004484](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2003JD004484).
- 666 Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations: a new  
667 environmental record for monitoring extremes. *Scientific Data* 2, (150066), doi:[http://doi.org/  
668 10.1038/sdata.2015.66](http://doi.org/10.1038/sdata.2015.66).
- 669 Garreaud, R., 1999: Multiscale Analysis of the Summertime Precipitation over the Central  
670 Andes. *Monthly Weather Review*, **127** (5), 901–921, doi:10.1175/1520-0493(1999)127<0901:  
671 MAOTSP>2.0.CO;2, URL [https://doi.org/10.1175/1520-0493\(1999\)127<0901:MAOTSP>2.0.  
672 CO;2](https://doi.org/10.1175/1520-0493(1999)127<0901:MAOTSP>2.0.CO;2), [https://doi.org/10.1175/1520-0493\(1999\)127<0901:MAOTSP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0901:MAOTSP>2.0.CO;2).
- 673 Gräler, B., E. Pebesma, and G. Heuvelink, 2016: Spatio-Temporal Interpolation using gstat. *The R  
674 Journal*, **8**, 204–218, URL [https://journal.r-project.org/archive/2016-1/na-pebesma-heuvelink.  
675 pdf](https://journal.r-project.org/archive/2016-1/na-pebesma-heuvelink.pdf).
- 676 Gräler, B., M. Rehr, L. Gerharz, and E. Pebesma, 2012: Spatio-temporal analysis and interpolation  
677 of PM10 measurements in Europe for 2009. *ETC/ACM Technical Paper*, **8**, 1–29.

- 678 Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing  
679 convection combining ensemble and data assimilation techniques. *Geophysical Re-*  
680 *search Letters*, **29** (14), 38–1–38–4, doi:10.1029/2002GL015311, URL [https://agupubs.](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL015311)  
681 [onlinelibrary.wiley.com/doi/abs/10.1029/2002GL015311](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002GL015311), [https://agupubs.onlinelibrary.wiley.](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002GL015311)  
682 [com/doi/pdf/10.1029/2002GL015311](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002GL015311).
- 683 Hall, M., and Coauthors, 2012: Los peligros volcánicos asociados con el Antisana. Corporación  
684 Editora Nacional IG-EPN IRD.
- 685 Hanchoo Wong, R., U. Weesakul, and S. Chumchean, 2012: Bias correction of radar rainfall esti-  
686 mates based on a geostatistical technique. *ScienceAsia*, **38**, 373–385.
- 687 Herold, N., A. Behrangi, and L. V. Alexander, 2017: Large uncertainties in observed daily pre-  
688 cipitation extremes over land. *Journal of Geophysical Research: Atmospheres*, **122** (2), 668–  
689 681, doi:10.1002/2016JD025842, URL [https://agupubs.onlinelibrary.wiley.com/doi/abs/10.](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025842)  
690 [1002/2016JD025842](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025842), <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025842>.
- 691 Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A New Vertical Diffusion Package with an Ex-  
692 plicit Treatment of Entrainment Processes. *Monthly Weather Review*, **134** (9), 2318–2341,  
693 doi:10.1175/MWR3199.1, URL <https://doi.org/10.1175/MWR3199.1>, [https://doi.org/10.1175/](https://doi.org/10.1175/MWR3199.1)  
694 [MWR3199.1](https://doi.org/10.1175/MWR3199.1).
- 695 IPCC, 2013: *Summary for Policymakers*, book section SPM, 130. Cambridge University Press,  
696 Cambridge, United Kingdom and New York, NY, USA, doi:10.1017/CBO9781107415324.004,  
697 URL [www.climatechange2013.org](http://www.climatechange2013.org).
- 698 Jiménez, P. A., and J. Dudhia, 2012: Improving the Representation of Resolved and Unresolved  
699 Topographic Effects on Surface Wind in the WRF Model. *Journal of Applied Meteorology and*

700 *Climatology*, **51** (2), 300–316, doi:10.1175/JAMC-D-11-084.1, URL <https://doi.org/10.1175/JAMC-D-11-084.1>, <https://doi.org/10.1175/JAMC-D-11-084.1>.

702 Junquas, C., K. Takahashi, T. Condom, J.-C. Espinoza, S. Chavez, J.-E. Sicart, and T. Lebel,  
703 2017: Understanding the influence of orography on the precipitation diurnal cycle and  
704 the associated atmospheric processes in the central Andes. *Climate Dynamics*, doi:10.1007/  
705 s00382-017-3858-8, URL <https://doi.org/10.1007/s00382-017-3858-8>.

706 Kendall, M., 1948: *Rank correlation methods*. Griffin, London.

707 Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin, 1992: Testing the null hypothesis of station-  
708 arity against the alternative of a unit root: How sure are we that economic time series have a unit  
709 root? *Journal of Econometrics*, **54** (1-3), 159–178, URL <https://EconPapers.repec.org/RePEc:eee:conom:v:54:y:1992:i:1-3:p:159-178>.

711 Laraque, A., J. Ronchail, G. Cochonneau, R. Pombosa, and J. L. Guyot, 2007: Heterogeneous Dis-  
712 tribution of Rainfall and Discharge Regimes in the Ecuadorian Amazon Basin. *Journal of Hy-*  
713 *drometeorology*, **8** (6), 1364–1381, doi:10.1175/2007JHM784.1, URL <https://doi.org/10.1175/2007JHM784.1>, <https://doi.org/10.1175/2007JHM784.1>.

715 Lin, Y.-L., R. D. Farley, and H. D. Orville, 1983: Bulk Parameterization of the Snow Field in a  
716 Cloud Model. *Journal of Climate and Applied Meteorology*, **22** (6), 1065–1092, doi:10.1175/  
717 1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, URL [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2), [https://doi.org/10.1175/1520-0450\(1983\)022<1065:BPOTSF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2).

718 0.CO;2,  
719 0.CO;2.

720 Ly, S., C. Charles, and A. Degré, 2011: Geostatistical interpolation of daily rainfall at catchment  
721 scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium.

722 *Hydrology and Earth System Sciences*, **15** (7), 2259–2274, doi:10.5194/hess-15-2259-2011,  
723 URL <http://www.hydrol-earth-syst-sci.net/15/2259/2011/>.

724 Mann, H. B., 1945: Nonparametric tests against trend. *Econometrica*, **13** (3), 245–259, URL  
725 <http://www.jstor.org/stable/1907187>.

726 Manz, B., S. Páez-Bimos, N. Horna, W. Buytaert, B. Ochoa-Tocachi, W. Lavado-Casimiro,  
727 and B. Willems, 2017: Comparative Ground Validation of IMERG and TMPA at Variable  
728 Spatiotemporal Scales in the Tropical Andes. *Journal of Hydrometeorology*, **18** (9), 2469–  
729 2489, doi:10.1175/JHM-D-16-0277.1, URL <https://doi.org/10.1175/JHM-D-16-0277.1>, <https://doi.org/10.1175/JHM-D-16-0277>.  
730

731 Marrel, A., B. Iooss, F. Van Dorpe, and E. Volkova, 2008: An efficient methodology for modeling  
732 complex computer codes with Gaussian processes. *Computational Statistics & Data Analysis*,  
733 **52** (10), 4731–4744.

734 Maussion, F., D. Scherer, R. Finkelburg, J. Richters, W. Yang, and T. Yao, 2011: WRF simulation  
735 of a precipitation event over the tibetan plateau, China an assessment using remote sensing and  
736 ground observations. *Hydrology and Earth System Sciences*, **15** (6), 1795–1817, doi:10.5194/  
737 hess-15-1795-2011, URL <http://www.hydrol-earth-syst-sci.net/15/1795/2011/>.

738 Michelangeli, P. A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Ap-  
739 plication to wind cumulative distribution functions. *Geophysical Research Letters*, **36** (11), doi:  
740 10.1029/2009GL038401.

741 Mlawer, E., S. Taubman, P. Brown, M. Iacono, and S. Clough, 1997: Radiative transfer for inho-  
742 mogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of*  
743 *Geophysical Research: Atmospheres*, **102**, 16, doi:10.1029/97JD00237.

- 744 Mourre, L., T. Condom, C. Junquas, T. Lebel, J. E. Sicart, R. Figueroa, and A. Cochachin, 2016:  
745 Spatio-temporal assessment of WRF, TRMM and in situ precipitation data in a tropical moun-  
746 tain environment (Cordillera Blanca, Peru). *Hydrology and Earth System Sciences*, **20** (1), 125–  
747 141, doi:10.5194/hess-20-125-2016, URL <http://www.hydrol-earth-syst-sci.net/20/125/2016/>.
- 748 Moya-Álvarez, A. S., D. Martínez-Castro, J. L. Flores, and Y. Silva, 2018: Sensitivity Study  
749 on the Influence of Parameterization Schemes in WRF\_ ARW Model on Short- and Medium  
750 Range Precipitation Forecasts in the Central Andes of Peru. *Advances in Meteorology*, **2018**,  
751 doi:<https://doi.org/10.1155/2018/1381092>.
- 752 Müller, M. F., and S. E. Thompson, 2013: Bias adjustment of satellite rainfall data through stochas-  
753 tic modeling: Methods development and application to Nepal. *Advances in Water Resources*, **60**,  
754 121–134.
- 755 Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparam-  
756 eterization options (Noah-MP): 1. Model description and evaluation with local-scale mea-  
757 surements. *Journal of Geophysical Research: Atmospheres*, **116** (D12), n/a–n/a, doi:10.1029/  
758 2010JD015139, URL <http://dx.doi.org/10.1029/2010JD015139>, d12109.
- 759 Oakley, J., and A. O’Hagan, 2002: Bayesian inference for the uncertainty distribution of  
760 computer model outputs. *Biometrika*, **89** (4), 769–784, doi:10.1093/biomet/89.4.769, URL  
761 [+http://dx.doi.org/10.1093/biomet/89.4.769](http://dx.doi.org/10.1093/biomet/89.4.769), /oup/backfile/content\_public/journal/biomet/89/4/  
762 10.1093/biomet/89.4.769/2/890769.pdf.
- 763 Ochoa, A., L. Campozano, E. Sánchez, R. Gualán, and E. Samaniego, 2016: Evaluation of down-  
764 scaled estimates of monthly temperature and precipitation for a Southern Ecuador case study.  
765 *International Journal of Climatology*, **36**, 1244–1255, doi:10.5194/hess-18-3179-2014.

- 766 Ochoa, A., L. Pineda, P. Crespo, and P. Willems, 2014: Evaluation of TRMM 3B42 precipitation  
767 estimates and WRF retrospective precipitation simulation over the Pacific-Andean region of  
768 Ecuador and Peru. *Hydrol. Earth Syst. Sci.*, **18**, 3179–3193, doi:10.5194/hess-18-3179-2014.
- 769 Ouzeau, G., J.-M. Soubeyroux, M. Schneider, R. Vautard, and S. Planton, 2016: Heat waves  
770 analysis over France in present and future climate: Application of a new method on the EURO-  
771 CORDEX ensemble. *Climate Services*, **4**, 1 – 12, doi:https://doi.org/10.1016/j.cliser.2016.09.  
772 002, URL <http://www.sciencedirect.com/science/article/pii/S2405880716300309>.
- 773 Paccini, L., J. C. Espinoza, J. Ronchail, and H. Segura, 2018: Intraseasonal rainfall variabil-  
774 ity in the Amazon basin related to largescale circulation patterns: a focus on western Ama-  
775 zonAndes transition region. *International Journal of Climatology*, **38** (5), 2386–2399, doi:  
776 10.1002/joc.5341, URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5341>, <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5341>.
- 778 Paciorek, C. J., and M. J. Schervish, 2006: Spatial modelling using a new class of nonstationary  
779 covariance functions. *Environmetrics*, **17** (5), 483–506, doi:10.1002/env.785, URL <http://dx.doi.org/10.1002/env.785>.
- 781 Paredes-Trejo, F. J., H. Barbosa, and T. Lakshmi Kumar, 2017: Validating CHIRPS-based satel-  
782 lite precipitation estimates in Northeast Brazil. *Journal of Arid Environments*, **139**, 26 – 40,  
783 doi:https://doi.org/10.1016/j.jaridenv.2016.12.009, URL [http://www.sciencedirect.com/science/](http://www.sciencedirect.com/science/article/pii/S014019631630235X)  
784 [article/pii/S014019631630235X](http://www.sciencedirect.com/science/article/pii/S014019631630235X).
- 785 Paulson, C. A., 1970: The Mathematical Representation of Wind Speed and Temper-  
786 ature Profiles in the Unstable Atmospheric Surface Layer. *Journal of Applied Me-*  
787 *teorology*, **9** (6), 857–861, doi:10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2,

788 URL [https://doi.org/10.1175/1520-0450\(1970\)009<0857:TMROWS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2), [https://doi.org/](https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2)  
789 [10.1175/1520-0450\(1970\)009<0857:TMROWS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2).

790 Pebesma, E. J., 2004: Multivariable geostatistics in S: the gstat package. *Computers and Geo-*  
791 *sciences*, **30**, 683–691.

792 Piani, C., G. Weedon, M. Best, S. Gomes, P. Viterbo, S. Hagemann, and J. Haerter, 2010:  
793 Statistical bias correction of global simulated daily precipitation and temperature for the ap-  
794 plication of hydrological models. *Journal of Hydrology*, **395** (3), 199 – 215, doi:[https://](https://doi.org/10.1016/j.jhydrol.2010.10.024)  
795 [doi.org/10.1016/j.jhydrol.2010.10.024](https://doi.org/10.1016/j.jhydrol.2010.10.024), URL [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S0022169410006475)  
796 [S0022169410006475](http://www.sciencedirect.com/science/article/pii/S0022169410006475).

797 Pouget, J.-C., D. Proano, A. Vera, M. Villacis, T. Condom, M. Escobar, P. Le Goulven, and  
798 R. Calvez, 2017: Modélisation glacio-hydrologique et gestion des ressources en eau dans  
799 les Andes équatoriennes : l'exemple de Quito. *Hydrological Sciences Journal = Journal*  
800 *des Sciences Hydrologiques*, **62**, 431–446, doi:[10.1080/02626667.2015.1131988](https://doi.org/10.1080/02626667.2015.1131988), URL [http://](http://www.documentation.ird.fr/hor/fdi:010069651)  
801 [www.documentation.ird.fr/hor/fdi:010069651](http://www.documentation.ird.fr/hor/fdi:010069651).

802 R Core Team, 2015: *R: A Language and Environment for Statistical Computing*. Vienna, Austria,  
803 R Foundation for Statistical Computing, URL <https://www.R-project.org>.

804 Rabatel, A., and Coauthors, 2013: Current state of glaciers in the tropical andes: a multi-century  
805 perspective on glacier evolution and climate change. *The Cryosphere*, **7** (1), 81–102, doi:[10.](https://doi.org/10.5194/tc-7-81-2013)  
806 [5194/tc-7-81-2013](https://doi.org/10.5194/tc-7-81-2013), URL <https://www.the-cryosphere.net/7/81/2013/>.

807 Rasmussen, C. E., and C. K. I. Williams, 2005: *Gaussian Processes for Machine Learning (Adap-*  
808 *tive Computation and Machine Learning)*. The MIT Press.

- 809 Rollenbeck, R., and J. Bendix, 2011: Rainfall distribution in the Andes of southern Ecuador de-  
810 rived from blending weather radar data and meteorological field observations. **99**, 277–289.
- 811 Rossel, F., P. Le Goulven, and E. Cadier, 1999: Repartition spatiale de linfluence de IENSO sur  
812 les precipitations annuelles en equateur. *Revue des Sciences de l'Eau*, (**12**), 183–200.
- 813 Sicart, J., R. Hock, P. Ribstein, M. Litt, and E. Ramirez, 2011: Analysis of seasonal variations  
814 in mass balance and meltwater discharge of the Tropical Zongo Glacier by application of a  
815 distributed energy balance model. *J. Geophys. Res.*, **116**, doi:10.1029/2010JD015105.
- 816 Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF Version  
817 3. Tech. rep., National Center for Atmospheric Research, 1–113 pp.
- 818 Vera, C., and Coauthors, 2006: Toward a Unified View of the American Monsoon Systems. *Jour-*  
819 *nal of Climate*, **19** (**20**), 4977–5000, doi:10.1175/JCLI3896.1, URL [https://doi.org/10.1175/](https://doi.org/10.1175/JCLI3896.1)  
820 [JCLI3896.1](https://doi.org/10.1175/JCLI3896.1), <https://doi.org/10.1175/JCLI3896.1>.
- 821 Vicente-Serrano, S. M., and Coauthors, 2017: The complex influence of enso on droughts in  
822 ecuador. *Climate Dynamics*, **48** (**1**), 405–427, doi:10.1007/s00382-016-3082-y, URL [https://](https://doi.org/10.1007/s00382-016-3082-y)  
823 [doi.org/10.1007/s00382-016-3082-y](https://doi.org/10.1007/s00382-016-3082-y).
- 824 Vrac, M., 2015: *Statistical downscaling through CDF-transform*. URL [https://cran.r-project.org/](https://cran.r-project.org/web/packages/CDFt/index.html)  
825 [web/packages/CDFt/index.html](https://cran.r-project.org/web/packages/CDFt/index.html), r package version 1.0.1.
- 826 Vrac, M., P. Drobinski, A. Merlo, M. Herrmann, C. Lavaysse, L. Li, and S. Somot, 2012:  
827 Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty as-  
828 sessment. *Natural Hazards and Earth System Sciences*, **12** (**9**), 2769–2784, doi:10.5194/  
829 [nhess-12-2769-2012](http://www.nat-hazards-earth-syst-sci.net/12/2769/2012/), URL <http://www.nat-hazards-earth-syst-sci.net/12/2769/2012/>.



830 Vrac, M., and P. Friederichs, 2015: Multivariate-Intervariable, Spatial, and Temporal-Bias Cor-  
831 rection. *Journal of Climate*, **28** (1), 218–237, doi:10.1175/JCLI-D-14-00059.1.

832 Vrac, M., T. Noel, and R. Vautard, 2016: Bias correction of precipitation through Singularity  
833 Stochastic Removal: Because occurrences matter. *Journal of Geophysical Research: Atmo-*  
834 *spheres*.

835 Vrac, M., and P. Vaittinada, 2017: Influence of Bias Correcting Predictors on Statistical Down-  
836 scaling Models. *Journal of Applied Meteorology and Climatology*, **56** (1), 5–26, doi:10.1175/  
837 JAMC-D-16-0079.1, URL <http://dx.doi.org/10.1175/JAMC-D-16-0079.1>, <http://dx.doi.org/10.1175/JAMC-D-16-0079.1>.

838

839 Vuille, M., R. S. Bradley, and F. Keimig, 2000: Climate Variability in the Andes of Ecuador  
840 and Its Relation to Tropical Pacific and Atlantic Sea Surface Temperature Anomalies. *Jour-*  
841 *nal of Climate*, **13** (14), 2520–2535, doi:10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;  
842 2, URL [https://doi.org/10.1175/1520-0442\(2000\)013<2520:CVITAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;2), [https://doi.org/10.1175/1520-0442\(2000\)013<2520:CVITAO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2520:CVITAO>2.0.CO;2).

843

844 Wagnon, P., M. Lafaysse, Y. Lejeune, L. Maisincho, M. Rojas, and J. P. Chazarin, 2009: Under-  
845 standing and modeling the physical processes that govern the melting of snow cover in a tropical  
846 mountain environment in Ecuador. *Journal of Geophysical Research: Atmospheres*, **114** (D19),  
847 doi:10.1029/2009JD012292, URL [https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012292)  
848 [2009JD012292](https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012292), <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012292>.

849 Yang, Z.-L., and Coauthors, 2011: The community noah land surface model with multipa-  
850 rameterization options (noah-mp): 2. evaluation over global river basins. *Journal of Geo-*  
851 *physical Research: Atmospheres*, **116** (D12), n/a–n/a, doi:10.1029/2010JD015140, URL <http://dx.doi.org/10.1029/2010JD015140>, d12110.

852

853 Zambrano-Bigiarini, M., A. Nauditt, C. Birkel, K. Verbist, and L. Ribbe, 2017: Temporal and  
854 spatial evaluation of satellite-based rainfall estimates across the complex topographical and  
855 climatic gradients of Chile. *Hydrology and Earth System Sciences*, **21** (2), 1295–1320, doi:  
856 10.5194/hess-21-1295-2017, URL <https://www.hydrol-earth-syst-sci.net/21/1295/2017/>.

857 **LIST OF TABLES**

858 **Table 1.** Description of meteorological in situ stations. Total precipitation during the  
859 2014-2015 period at each meteorological station and total precipitation in [mm]  
860 simulated by WRF at 1 km resolution. Note that for station 12 and 18, we indi-  
861 cate the associated grid-point model used for the bias correction computations.  
862 43

863 **Table 2.** Cross-validation leave-one-out results of annual precipitation for the four Gaus-  
864 sian Process models over the WRF bias proposed with four diferent drifts The  
865 criteria are calculated for the 2014 and 2015 periods, separately. . . . . 44

866 TABLE 1. Description of meteorological in situ stations. Total precipitation during the 2014-2015 period at  
867 each meteorological station and total precipitation in [mm] simulated by WRF at 1 km resolution. Note that for  
868 station 12 and 18, we indicate the associated grid-point model used for the bias correction computations.

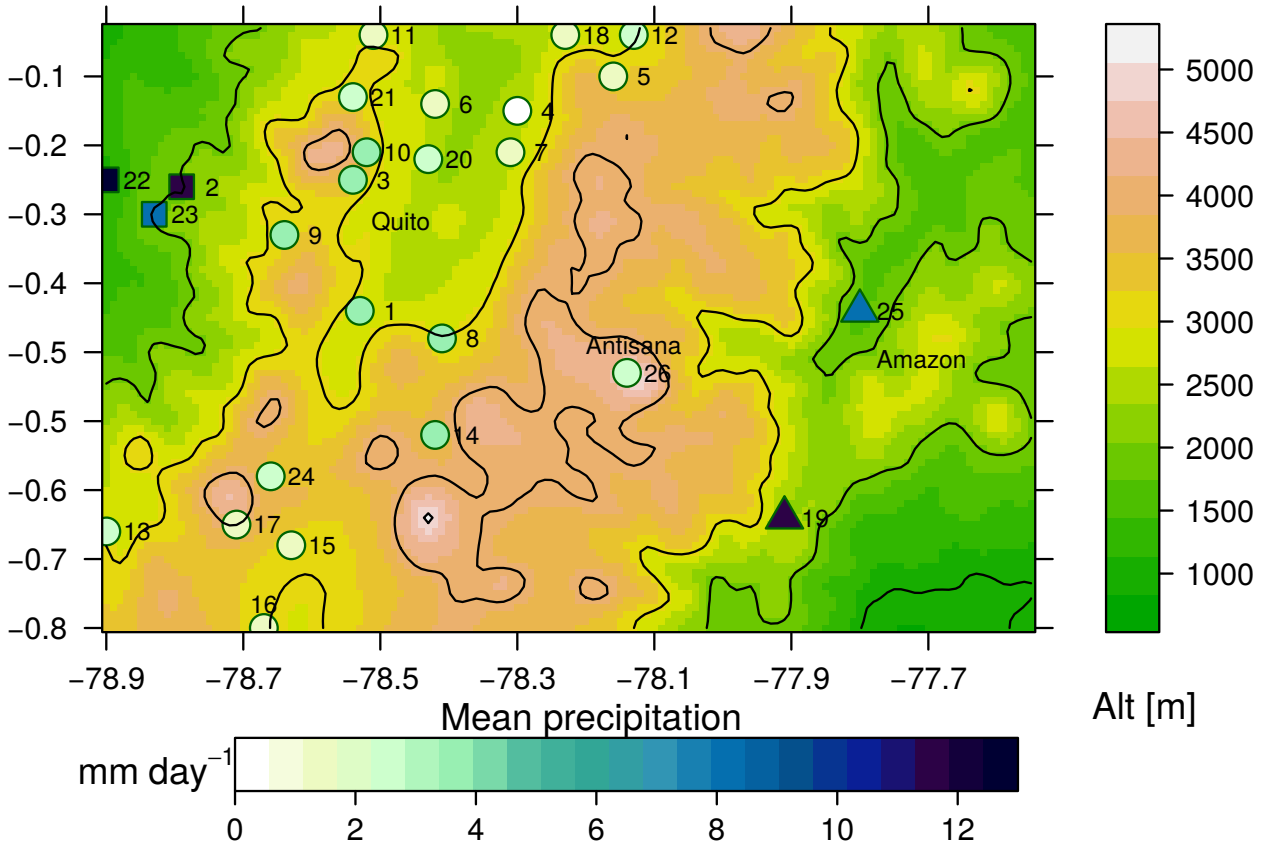
Number	Elev. [m]	Lon.	Lat.	P. Obs. [mm]	P. WRF [mm]
The Pacific coast					
2	1985	-78.78	-0.21	8656	1568
22	1110	-78.90	-0.21	8954	6892
23	2028	-78.82	-0.25	6132	2140
The Andes					
1	2843	-78.53	-0.39	2447	1475
3	3447	-78.54	-0.20	2510	2478
4	2530	-78.30	-0.10	769	659
5	3317	-78.17	-0.06	835	972
6	2625	-78.42	-0.10	831	932
7	2576	-78.32	-0.16	1221	562
8	3021	-78.42	-0.43	2403	1992
9	3276	-78.63	-0.28	2886	1082
10	3498	-78.52	-0.16	2602	2368
11	2880	-78.51	0.00	995	623
12*	2949	-78.14	0.00	1763	1835
13	2930	-78.89	-0.70	1694	2105
14	3705	-78.43	-0.56	2695	753
15	3157	-78.63	-0.72	1506	757
16	3035	-78.66	-0.83	962	839
17	4009	-78.70	-0.68	1203	1229
18*	2828	-78.23	0.00	1080	1327
20	2487	-78.43	-0.18	1699	948
21	3218	-78.54	-0.09	1824	695
24	3528	-78.66	-0.62	1989	690
26	4812	-78.15	-0.47	2255	2062
The Amazon					
19	2390	-77.93	-0.67	8276	2297
25	1700	-77.82	-0.39	6261	1186

869 TABLE 2. Cross-validation leave-one-out results of annual precipitation for the four Gaussian Process models  
 870 over the WRF bias proposed with four different drifts. The criteria are calculated for the 2014 and 2015 periods,  
 871 separately.

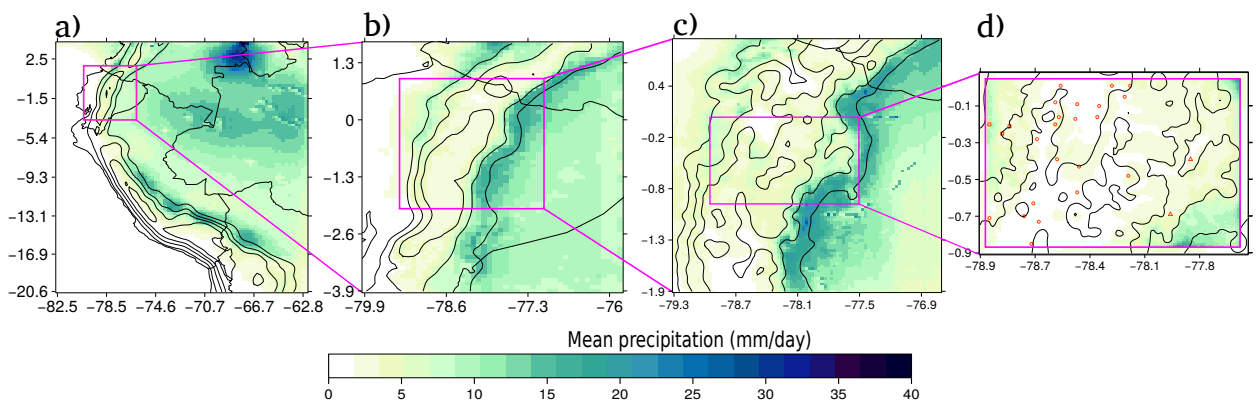
	2014				2015			
	Bias	RMSE	Correlation	Q <sub>2</sub>	Bias	RMSE	Correlation	Q <sub>2</sub>
WRF	1.77	2.89	0.59		2.19	3.67	0.65	
GP+longitude+latitude+alt.	1.71	2.31	0.71	0.44	1.56	2.14	0.80	0.65
<b>GP+longitude+latitude</b>	<b>1.48</b>	<b>2.14</b>	<b>0.76</b>	<b>0.56</b>	<b>1.31</b>	<b>2.08</b>	<b>0.84</b>	0.71
GP+longitude	1.50	<b>2.14</b>	<b>0.76</b>	0.49	1.32	2.11	<b>0.84</b>	0.64
GP+elevation	1.68	2.30	0.72	<b>0.56</b>	1.72	2.29	0.78	0.70

## LIST OF FIGURES

872		
873	<b>Fig. 1.</b>	Map of the region under study and INAMHI stations. The stations of each region (Pacific coast, Andes and Amazon) are marked with three geometric figures: rectangles for the Pacific coast, circles for the Andes and triangles for the Amazon. The color figures indicate the mean daily precipitation [mm day <sup>-1</sup> ] during the 2014-2015 period. The contour interval is 1000 m. Note that for station 12 and 18, we indicate the associated grid-point model used for the bias correction computations. . . . . 46
874		
875		
876		
877		
878		
879	<b>Fig. 2.</b>	Mean daily precipitation [mm day <sup>-1</sup> ] during the 2014-2015 period of the four nested domains of the WRF simulation a) WRF 27 km, b) WRF 9 km, c) WRF 3 km and d) WRF 1 km. Figure d) shows the exact WRF domain, the WRF domain without the lateral boundary zone (purple box) and the stations (red dots). Note that for station 12 and 18, we indicate the associated grid-point model used for the bias correction computations. . . . . 47
880		
881		
882		
883		
884	<b>Fig. 3.</b>	WRF precipitation relative biases for the a) 2014 and b) 2015 (% of relative error). Note that for station 12 and 18, we indicate the associated grid-point model used for the bias correction computations. . . . . 48
885		
886		
887	<b>Fig. 4.</b>	Flowchart for the two approaches a) Spatial CDF-t and b) Gaussian process model. . . . . 49
888	<b>Fig. 5.</b>	Time series treatment for a Pacific coast station (station 22), an Andes station (Antisana station 26), and Amazon station (station 25). a) Original time series ( $X_t$ in red and $Y_t$ in blue) for the Pacific coast station ( <i>resp.</i> , e and i, for the Andes station and the Amazon station), b) differentiated observed time series $\Delta Y_t$ ( <i>resp.</i> , f and j, for the Andes station and the Amazon station), c) differentiated simulated time series $\Delta X_t$ ( <i>resp.</i> , g and k, for the Andes station and the Amazon station) and d) CDFs of observation (obs.), simulation (Simu.) and CDF-t correction (Corr.) ( <i>resp.</i> , h and l, for the Andes station and the Amazon station). . . . . 50
889		
890		
891		
892		
893		
894		
895	<b>Fig. 6.</b>	Scatter plots of gridded product mean daily precipitation (CHIRPS, WRF) and a) CDF-t and b) GP results using cross-validation versus the stations mean daily precipitation [mm day <sup>-1</sup> ]. . . . . 51
896		
897	<b>Fig. 7.</b>	The accumulated precipitation [mm] of the WRF, GP ( <i>resp.</i> , CDF-t) and the in situ stations of three regions. GP is evaluated in cross-validation leave-one-out. a) Pacific coast (station 22) for GP ( <i>resp.</i> , d for CDF-t) b) Andes (station 26) for GP ( <i>resp.</i> , d for CDF-t) c) Amazon (station 25) for GP ( <i>resp.</i> , d for CDF-t). . . . . 52
898		
899		
900		
901	<b>Fig. 8.</b>	Boxplots of criteria related to precipitation occurrence (rainy/non-rainy events) for three gridded products: WRF, Spatial CDF-t and GP (Gaussian Process model with drift longitude and latitude), using a cross-validation leave-one-out framework. a) FAR criterion (ideal value 0), b) POD criterion (1), c) PODF criterion (0) and d) HSS criterion (1). . . . . 53
902		
903		
904		
905	<b>Fig. 9.</b>	Boxplots of criteria related to precipitation intensity for three gridded products: WRF, spatial CDF-t and GP (Gaussian Process model with drift longitude and latitude) using a cross-validation leave-one-out framework. a) KS, b) RMSE, c) Spearman correlation and d) $Q_{95}$ . . . . . 54
906		
907		
908		
909	<b>Fig. 10.</b>	Mean of daily precipitation maps [mm day <sup>-1</sup> ] during 2014-2015 and Voronoï diagram. a) Observations in situ, b) CHIRPS, c) WRF simulation, d) GP model, e) Spatial CDF-t and f) Voronoï diagram. The stations are represented by green circles. Note that for station 12 and 18, we indicate the associated grid-point model used for the bias correction computations. . . . . 55
910		
911		
912		



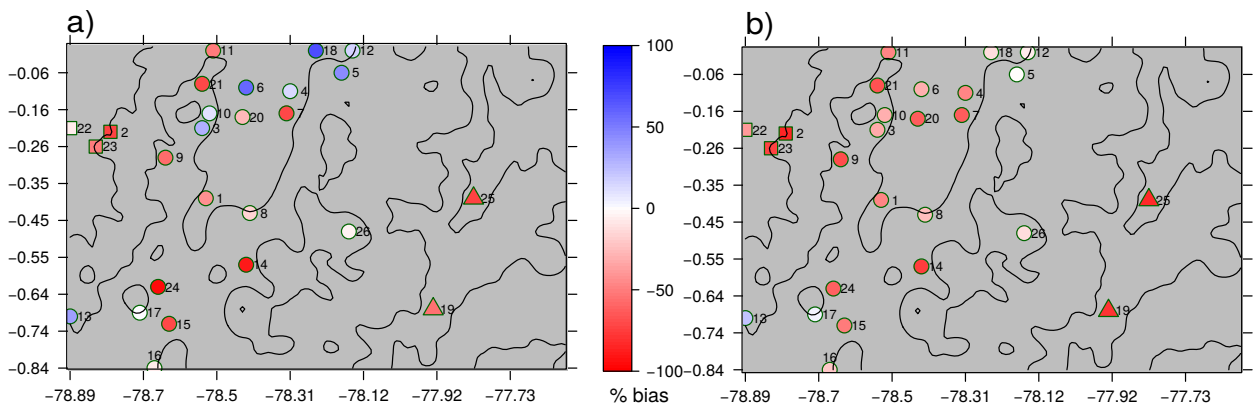
913 FIG. 1. Map of the region under study and INAMHI stations. The stations of each region (Pacific coast,  
 914 Andes and Amazon) are marked with three geometric figures: rectangles for the Pacific coast,  
 915 Andes and triangles for the Amazon. The color figures indicate the mean daily precipitation [mm day<sup>-1</sup>]  
 916 during the 2014-2015 period. The contour interval is 1000 m. Note that for station 12 and 18, we indicate the associated  
 917 grid-point model used for the bias correction computations.



918 FIG. 2. Mean daily precipitation [ $\text{mm day}^{-1}$ ] during the 2014-2015 period of the four nested domains of the  
 919 WRF simulation a) WRF 27 km, b) WRF 9 km, c) WRF 3 km and d) WRF 1 km. Figure d) shows the exact  
 920 WRF domain, the WRF domain without the lateral boundary zone (purple box) and the stations (red dots). Note  
 921 that for station 12 and 18, we indicate the associated grid-point model used for the bias correction computations.

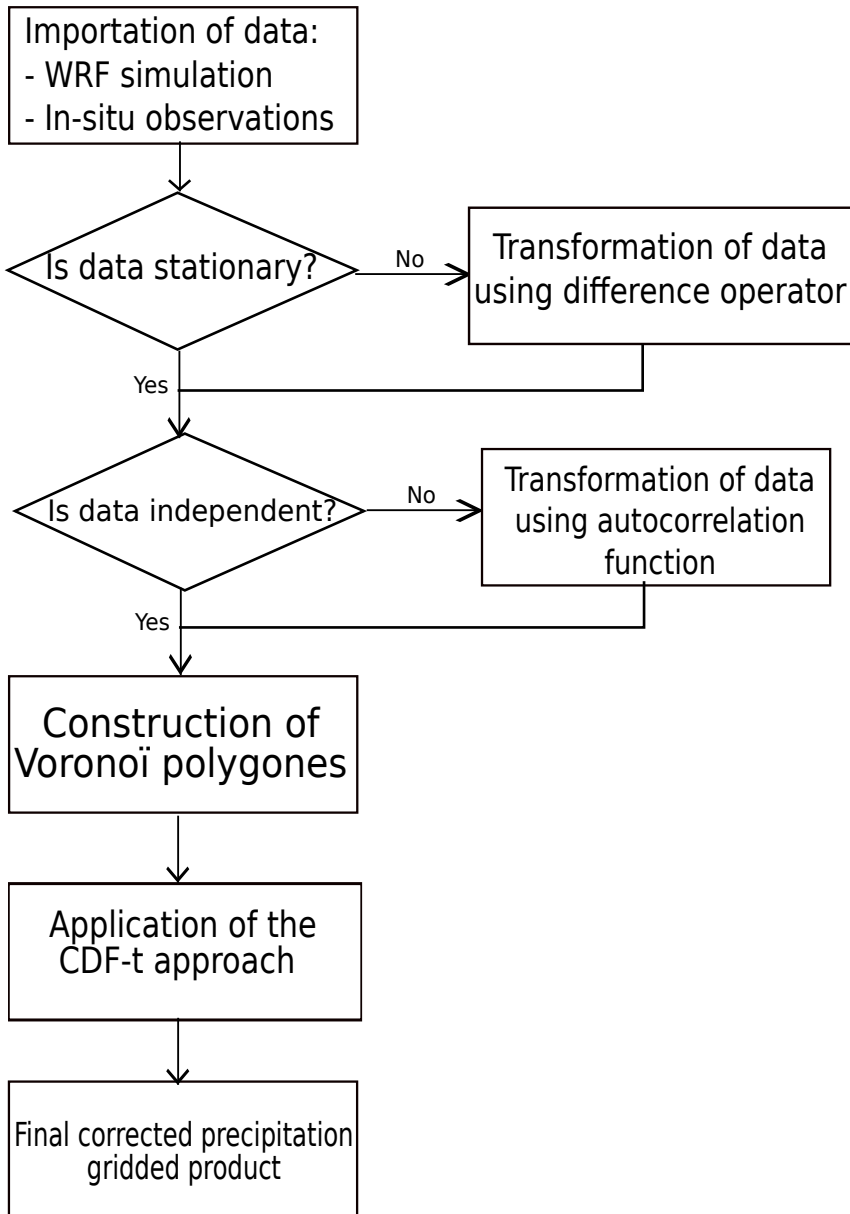
922





923 FIG. 3. WRF precipitation relative biases for the **a)** 2014 and **b)** 2015 (% of relative error). Note that for  
 924 station 12 and 18, we indicate the associated grid-point model used for the bias correction computations.

a)



b)

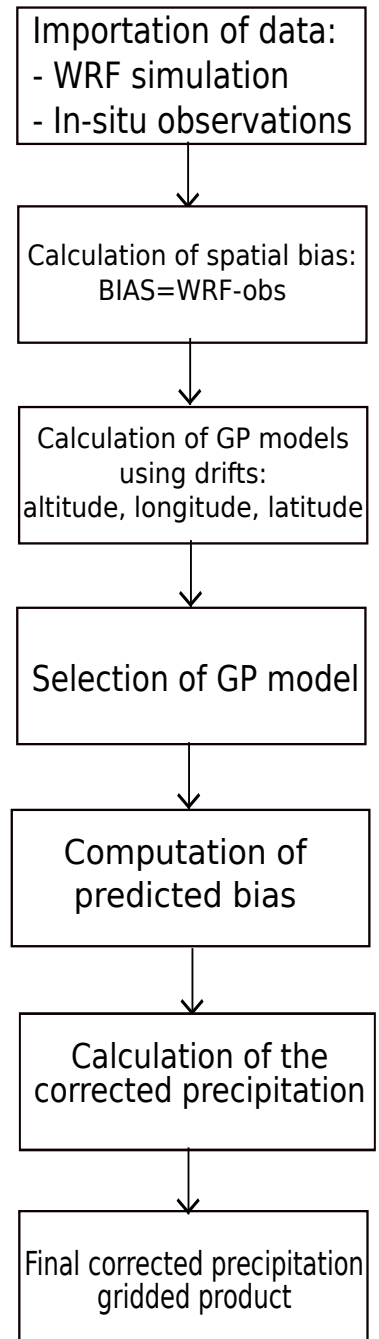
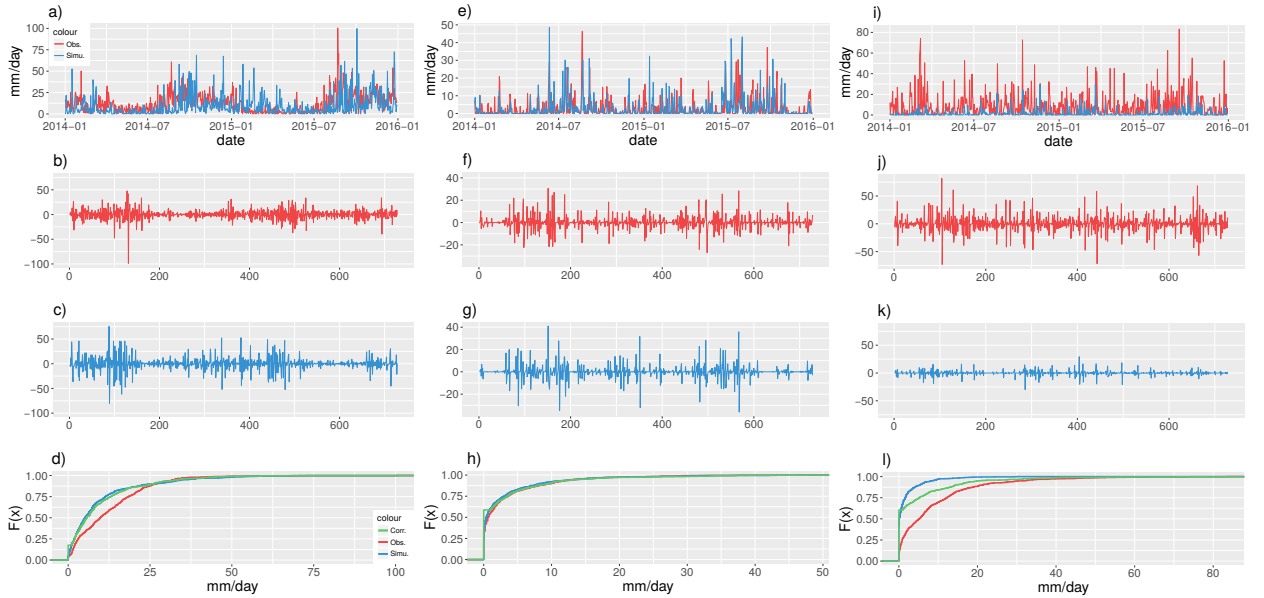
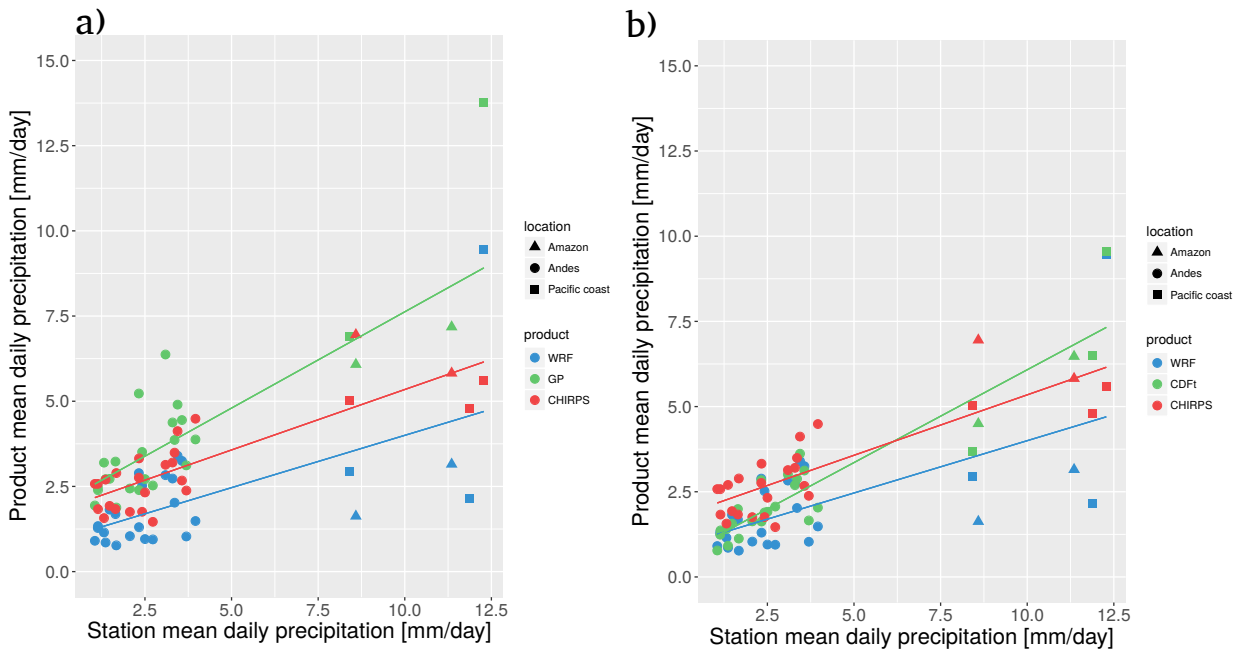


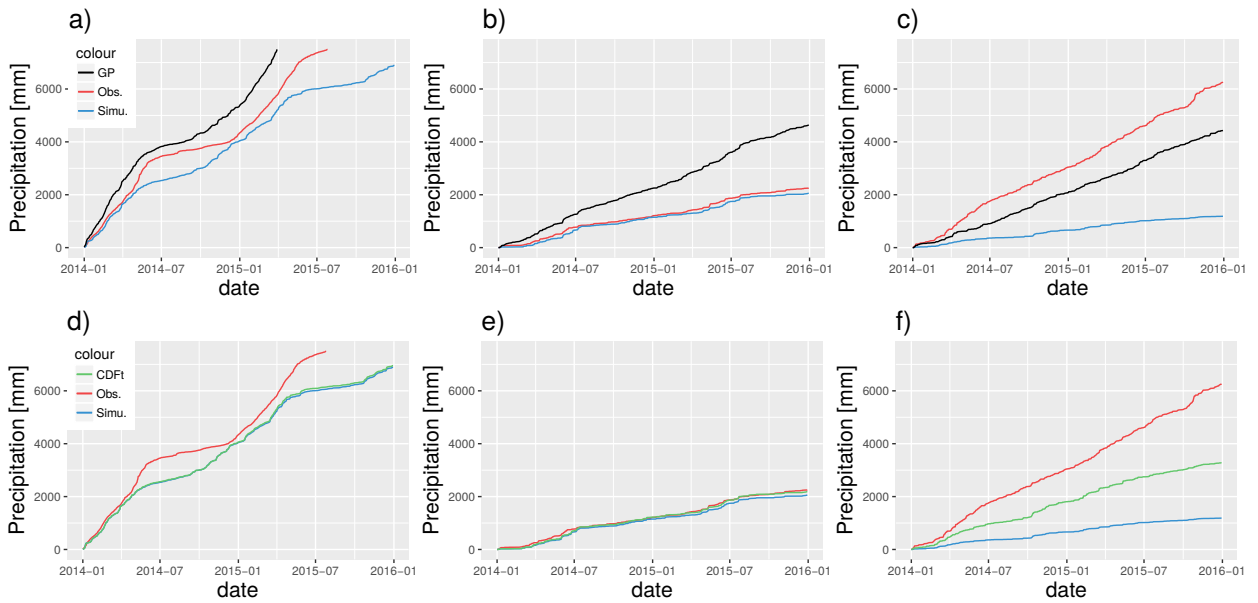
FIG. 4. Flowchart for the two approaches a) Spatial CDF-t and b) Gaussian process model.



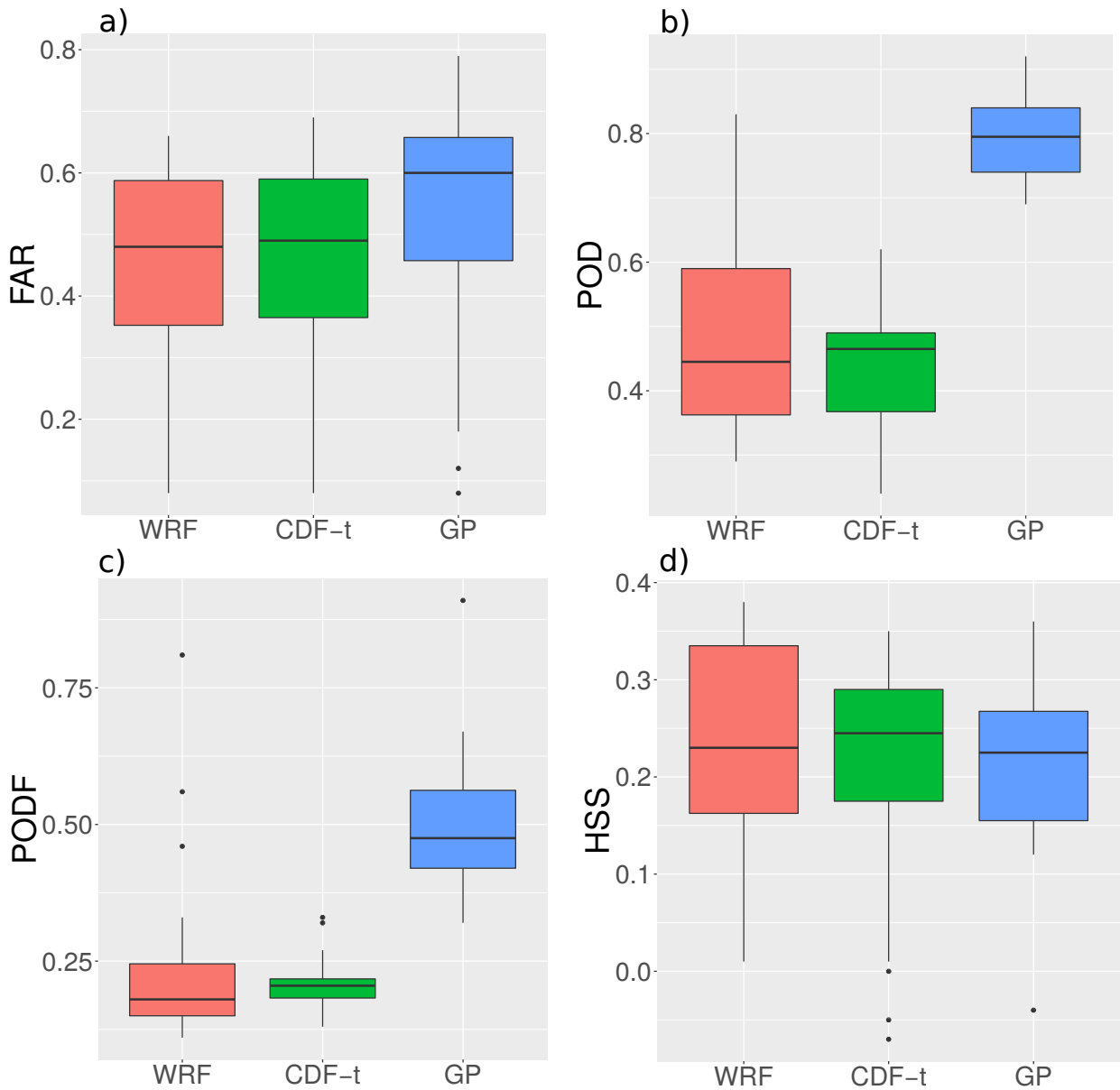
925 FIG. 5. Time series treatment for a Pacific coast station (station 22), an Andes station (Antisana station 26),  
 926 and Amazon station (station 25). **a)** Original time series ( $X_t$  in red and  $Y_t$  in blue) for the Pacific coast station  
 927 (*resp.*, **e** and **i**, for the Andes station and the Amazon station), **b)** differentiated observed time series  $\Delta Y_t$  (*resp.*, **f**  
 928 and **j**, for the Andes station and the Amazon station), **c)** differentiated simulated time series  $\Delta X_t$  (*resp.*, **g** and **k**,  
 929 for the Andes station and the Amazon station) and **d)** CDFs of observation (obs.), simulation (Simu.) and CDF-t  
 930 correction (Corr.) (*resp.*, **h** and **l**, for the Andes station and the Amazon station).



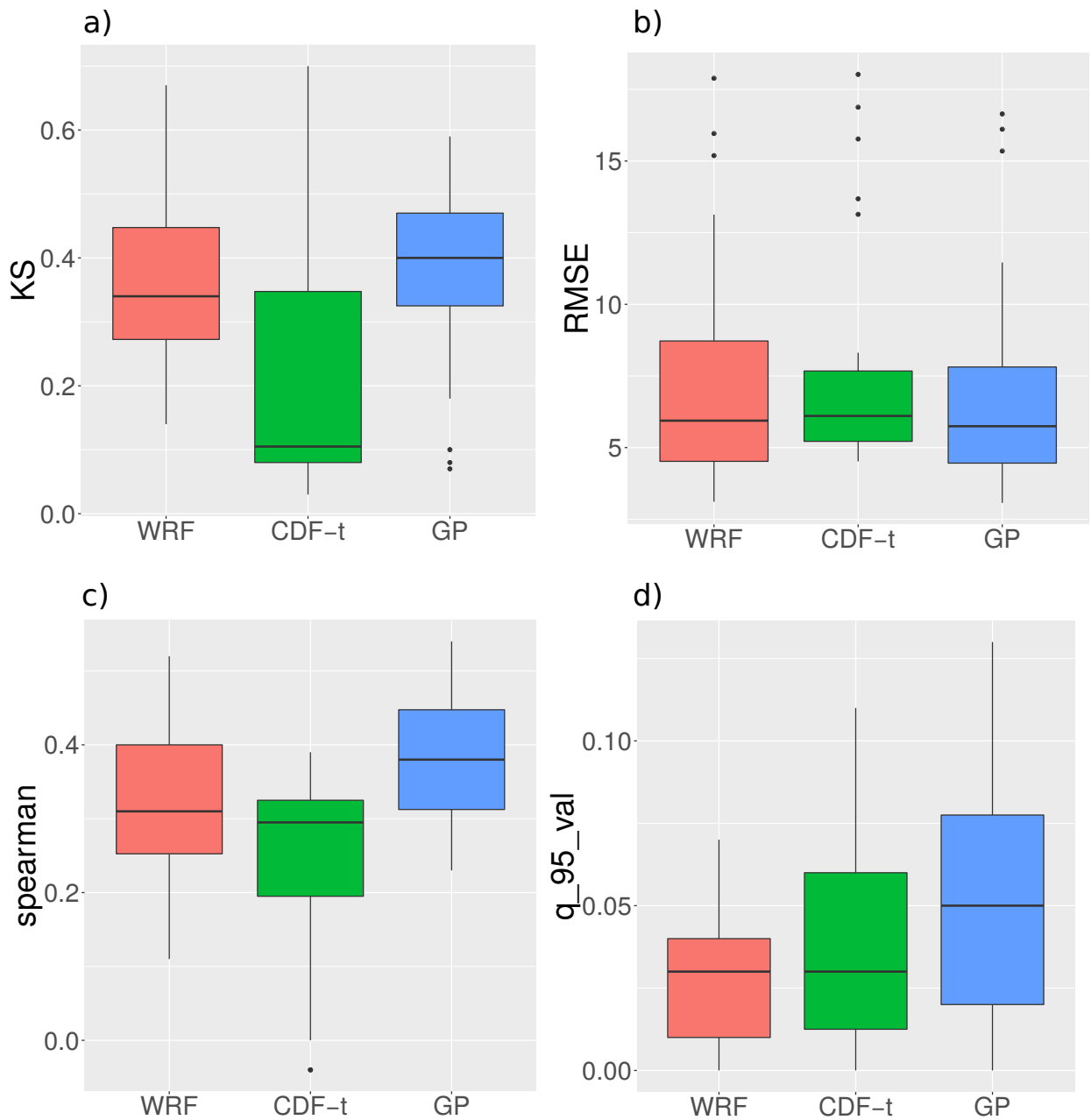
931 FIG. 6. Scatter plots of gridded product mean daily precipitation (CHIRPS, WRF) and a) CDF-t and b) GP  
 932 results using cross-validation versus the stations mean daily precipitation [mm day<sup>-1</sup>].



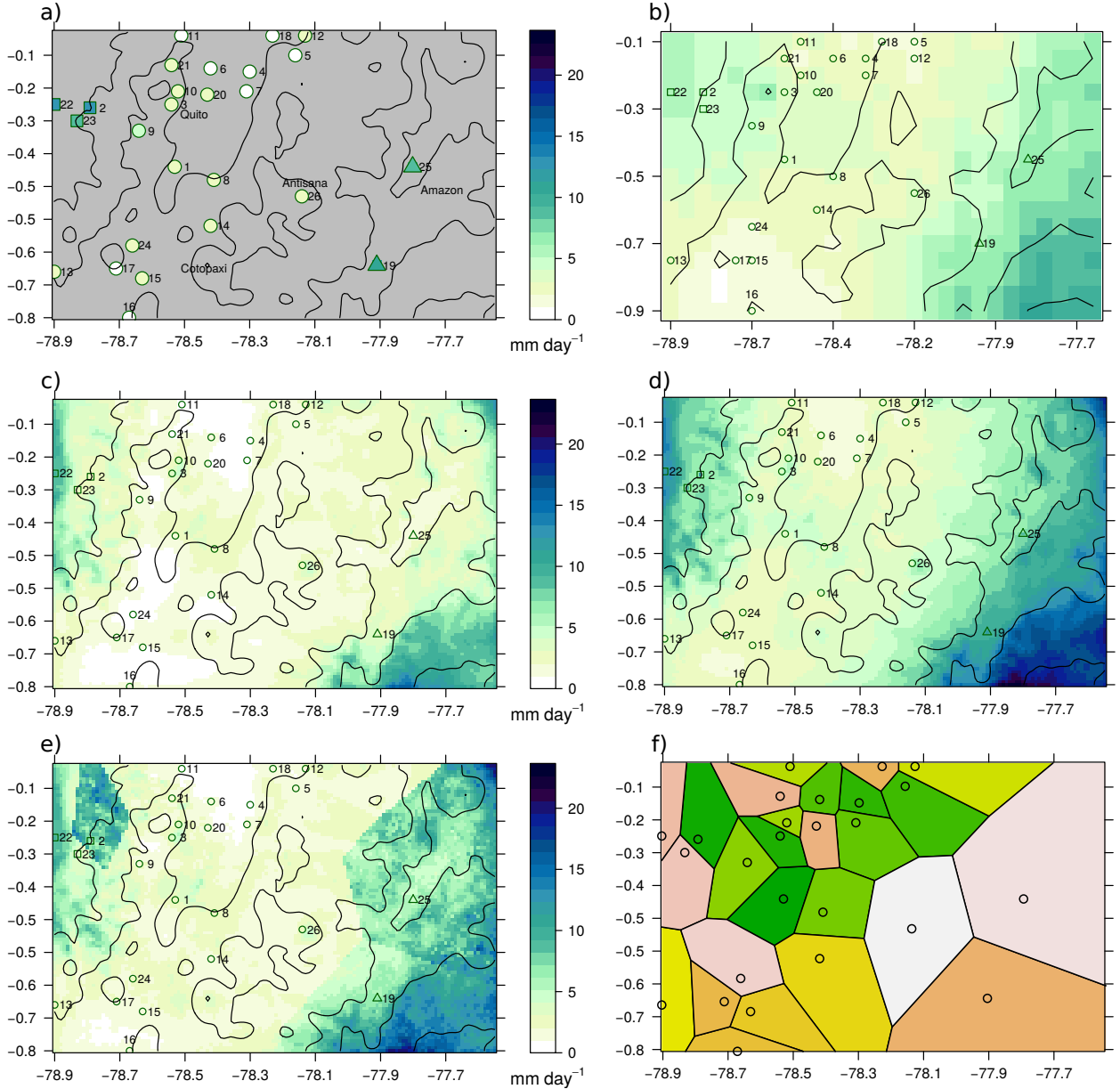
933 FIG. 7. The accumulated precipitation [mm] of the WRF, GP (*resp.*, CDF-t) and the in situ stations of three  
 934 regions. GP is evaluated in cross-validation leave-one-out. **a)** Pacific coast (station 22) for GP (*resp.*, **d** for  
 935 CDF-t) **b)** Andes (station 26) for GP (*resp.*, **d** for CDF-t) **c)** Amazon (station 25) for GP (*resp.*, **d** for CDF-t).



936 FIG. 8. Boxplots of criteria related to precipitation occurrence (rainy/non-rainy events) for three gridded  
 937 products: WRF, Spatial CDF-t and GP (Gaussian Process model with drift longitude and latitude), using a  
 938 cross-validation leave-one- out framework. **a)** FAR criterion (ideal value 0), **b)** POD criterion (1), **c)** PODF  
 939 criterion (0) and **d)** HSS criterion (1).



940 FIG. 9. Boxplots of criteria related to precipitation intensity for three gridded products: WRF, spatial CDF-  
 941 t and GP (Gaussian Process model with drift longitude and latitude) using a cross-validation leave-one-out  
 942 framework. **a)** KS, **b)** RMSE, **c)** Spearman correlation and **d)** Q<sub>95</sub>.



943 FIG. 10. Mean of daily precipitation maps [ $\text{mm day}^{-1}$ ] during 2014-2015 and Voronoi diagram. **a)** Obser-  
 944 vations in situ, **b)** CHIRPS, **c)** WRF simulation, **d)** GP model, **e)** Spatial CDF-t and **f)** Voronoi diagram. The  
 945 stations are represented by green circles. Note that for station 12 and 18, we indicate the associated grid-point  
 946 model used for the bias correction computations.