



High-Resolution Reanalysis of Daily Precipitation using AROME Model Over France

STÉPHANE VAN HYFTE

PATRICK LE MOIGNE

ERIC BAZILE

ANTOINE VERRELLE

AARON BOONE

*Author affiliations can be found in the back matter of this article

ORIGINAL RESEARCH
PAPER



STOCKHOLM
UNIVERSITY PRESS

ABSTRACT

Among the various meteorological variables, precipitation is one of significant interest, especially for hydrological studies. However, obtaining a reliable precipitation data set is a difficult challenge as precipitation can be very discontinuous in space and time. In this study, a method to obtain a high resolution precipitation reanalysis over France is purposed based on a study from 01/01/2016 to 31/12/2018. The French operational regional model Application de la Recherche à l'Opérationnel à Mésos-Echelle (AROME) is combined with precipitation observations, which have been quality controlled, using an optimum interpolation data assimilation algorithm. To use this technique, some hypotheses have to be verified, such as the Gaussian distribution of the innovations. Since the precipitation distribution is highly asymmetric, a Box-Cox transformation is applied to both background and observations to work with variables which behave like Gaussian variables. Then, the background and observation standard deviation errors are determined thanks to the semi-variogram technique, which provides daily values. Results show that the Box-Cox transformation provides better scores for light precipitation and has the same quality as the reference – analysis in the physical space – for high precipitation amounts.

CORRESPONDING AUTHOR:

Stéphane Van Hyfte

CNRM, Université de Toulouse,
Météo-France, CNRS, Toulouse,
France

stephane.van-hyfte@meteo.fr

KEYWORDS:

precipitation analysis;
Gaussian anamorphism;
optimum interpolation; bias
correction; surface reanalysis

TO CITE THIS ARTICLE:

Van Hyfte, S, Le Moigne, P,
Bazile, E, Verrelle, A and
Boone, A. 2023. High-
Resolution Reanalysis of Daily
Precipitation using AROME
Model Over France. *Tellus A:
Dynamic Meteorology and
Oceanography*, 75(1): 27–49.
DOI: [https://doi.org/10.16993/
tellusa.95](https://doi.org/10.16993/tellusa.95)

1 INTRODUCTION

The ability to produce a reliable precipitation data set is very important for hydrological studies as precipitation is the input variable with the most significant impact in hydrological models (Carrera et al., 2010). Moreover, events with high rainfall rates are often related to high societal and economical impacts. In France, high rainfall events known as “Mediterranean episodes” can be associated with flash flood events and can cause a great deal of damage. In addition, it is necessary to have a good representation of the precipitation amounts over a large period for climatological studies. High quality rainfall data sets are complicated to obtain due to the spatially discontinuous nature of the precipitation. Since the 90’s, several projects have been initiated in order to evaluate the atmospheric state over a long time period. For such studies, the reanalysis method is of particular interest as it combines information contained in an a priori state of the atmosphere, called background, and in the available observations which then leads to a good representation of the atmospheric state even in areas where the observations are sparse. As a result, many meteorological centers produced different sets of global reanalyses, such as the European Center for Medium-range Weather Forecast (ECMWF) with ERA-Interim (Dee et al., 2011) and most recently ERA5 (Hersbach et al., 2020) or the Japan Meteorology Agency (JMA) with JRA25 (Onogi et al., 2007) and JRA55 (Kobayashi et al., 2015). Then, meteorological centers produced regional reanalysis for which they can concentrate on a limited area, allowing a better resolution. The higher resolution of regional reanalyses allows a better representation of local events (Kaiser-Weiss et al., 2019) such as flash floods, convection, local wind or meteorological effects due to orographic features. European Reanalysis and Observations for Monitoring (EURO4M) was the first regional reanalysis in Europe, and it provided data over Europe from 1989 to 2013 at a 22 km resolution (Dahlgren et al., 2016). Uncertainties in Ensemble Regional ReAnalysis (UERRA) is a project created thanks to a cooperation of 12 meteorological centers and provides a deterministic (Ridal et al., 2017) reanalysis from 1961 to 2015 and a 20-member ensemble (Jerney et al., 2015) reanalysis from 1979 to 2016 by using the global reanalysis ERA-Interim as the Lateral Boundary Conditions (LBC) coupled with the Unified Model (UM) as the background for both analyses. Within the UERRA project, a first precipitation reanalysis was released at a 5.5 km resolution over Europe. To produce this surface reanalysis, the atmospheric reanalysis was downscaled to 5.5 km to be used as background, which leads to a data set at a high resolution as it was run offline and so it used less computing power. However, this surface reanalysis was obtained using a simple technique based on Optimum Interpolation (OI) in physical space. As a result,

this surface reanalysis suffers from some analysis errors, such as an underestimation of snow in mountainous areas. Under the Copernicus program, another European regional reanalysis project is in production which is called Copernicus European Regional ReAnalysis (CERRA), using the more recent global reanalysis of the ECMWF ERA5 as a LBC. This regional reanalysis uses satellite data and in situ observations and is directly used as a background for a surface reanalysis of two-meter temperature and relative humidity at a 5.5 km resolution. Other surface reanalyses have been created over the last few years. For example, the Système d’Analyse Fournissant des Renseignements Adaptés à la Nivologie (SAFRAN) system (Durand et al., 1993) has been developed to produce a surface reanalysis over France (Vidal et al., 2010) which has been used extensively as input for land surface models for various purposes such as avalanche risk forecasting or water resource monitoring.

Data assimilation (DA) methods combine observations with a background field, generally coming from a Numerical Weather Prediction (NWP) model. DA contains many hypotheses about the variable characteristics, such as the Gaussian distribution of the variable or even the knowledge of the background and observation errors. If the first hypothesis can be applicable for global variables, it is rarely the case for very local ones such as precipitation, as this variable is very asymmetric, with a great proportion of light precipitation compared to rare high precipitation events. This characteristic of the precipitation makes the analysis of this variable very complicated. As a consequence, the precipitation analysis is a problem often studied by many meteorological centers. A common solution to the problems related with the precipitation is the use of a data transformation to map a variable into a new one having a Gaussian distribution. Many transformations have been tested over time, such as the logarithm (Mahfouf et al., 2007; Rodríguez-Puebla et al., 2001), the Gaussian anamorphism (Amezcuca and Van Leeuwen, 2014; Devers et al., 2020; Lien et al., 2013; Lussana et al., 2020), and the Box-Cox transformation (Box and Cox, 1964; Cecinati et al., 2017; Erdin et al., 2012; Ishak and Ahmad, 2018; Lespinas et al., 2015; Tippet et al., 2007). However, the use of a data transformation, such as the Box-Cox transformation induces a significant bias (Evans, 2013; Fortin et al., 2015) which can induce large errors in hydrological models (Carrera et al., 2010). The background field errors, coming from the hypotheses assumed in the modeling systems and the observation errors (coming from measurement errors, the more common and known one being the undercatchment of precipitation due to the wind effect, (Sevruk, 1996)), have also to be determined. A common and easy solution is the use of fixed errors, but this is not optimal as precipitation suffers from more measurement errors when falling in a solid form or when the winds are stronger. Many studies have been done to determine the optimal values

of the observation and background errors. They have proposed different methods to allow the determination of the background and observation errors, such as the diagnostic a posteriori (Desroziers et al., 2005), the maximum likelihood method (Dee and Da Silva, 1999) or even the semi-variogram technique (Lespinas et al., 2015, Touma et al., 2018, Wang et al., 2019).

In this study, a method to analyze precipitation at a high resolution is proposed. The domain is over France at a 1.3 km resolution, which is used since it is the resolution of the Application de la Recherche à l'Opérationnel à Mésos-Echelle (AROME) operational weather forecast model. In order to give precipitation a near-Gaussian distribution, the Box-Cox transformation is applied to both the background and the observations. This transformation also allows the use of the semi-variogram technique to determine daily background and observation errors. As this transformation induces a negative bias for the analyzed state, a bias correction method is used and presented herein. In the first section, the DA system used in this study is presented. The data transformation method is also presented followed by a presentation of the semi-variogram technique which was used to determine the background and observation errors. The bias correction applied to the transformed back into physical space data is also detailed. The second section presents the different scores used in this study for assessing the benefit and the potential of the proposed method. Statistical scores are calculated over the whole period to have a global evaluation, and they are also calculated for different seasons (winter, spring, summer and autumn). A convective case is also studied to evaluate the impact of the Box-Cox transformation compared to the analysis obtained with the existing surface analysis system used for the UERRA surface reanalysis (MESCAN) for situations with high precipitation amounts. Finally, section 4 deals with the conclusions and the perspectives of this work.

2 EXPERIMENTAL SETUP AND METHODS

This section presents the different methods used to produce the precipitation analysis. First, the analysis system MESCAN is presented, including the background and the observations used in this study. The data transformation is then presented followed by the presentation of a geostatistical method allowing the daily determination of the background and observation errors. Then, the bias induced by the data transformation is presented and a bias correction formula is proposed.

2.1 THE MESCAN DATA ASSIMILATION SYSTEM

In this study, the analysis is calculated with an OI algorithm by combining information contained in the observations and in a background state with the MESCAN system developed by Soci et al. (2016) for the UERRA

project. The analysis state is found by determining the analysis variance error minimum using a Bayesian approach. By using the background vector, $\mathbf{x}^b$, the observation vector, $\mathbf{y}^o$, the observation operator, H , which interpolates the background values to the observation space by taking into account the dynamic of the NWP model and can be non-linear, the analysis state x_i^a at a grid point i is defined as:

$$x_i^a = x_i^b + \sum_{j=1}^p w_{ij} [y_j^o - H(x^b)_j] \quad (1)$$

In Eq. (1), p is the number of observation points used to calculate the analysis at the grid point i , and w_{ij} are the weights attached to the innovations $d_j = y_j^o - H(x^b)_j$. The weights are obtained by solving the linear system:

$$\sum_{j=1}^p w_{i,j} (R_{j,k} + B_{j,k}) = B_{i,k} \quad (2)$$

where R stands for the matrix of the observation error covariances, B stands for the matrix of the background error covariances and k is between 1 and p . p has been set to 16 historically so that each model grid point is surrounded by a minimum number of observations, but not too large for computational cost issues. Under the hypothesis that only horizontal background errors are considered, background errors are homogeneous and isotropic and coefficients of the matrix B are modelled by a second-order auto-regressive function (Mahfouf et al., 2007) defined by Eq. (3) where r_{ij} is the distance between the points i and j , L is the characteristic horizontal scale and σ_b is the standard deviation of the background error. As in the regional reanalysis UERRA, σ_b is set to 13 mm in this study even if the use of the (Desroziers et al., 2005) diagnosis shows that this value is likely too high and is dependent on the season, with higher values during summer and autumn and lower values in winter and spring (not shown here), which is in agreement with the results obtained in Soci et al. (2016) in which the use of the a posteriori diagnosis gave estimated values of the standard deviation of observation and background errors for December ($\sigma_o = 2.35$ mm and $\sigma_b = 4.49$ mm) and for June ($\sigma_o = 4.19$ mm and $\sigma_b = 6.98$ mm). Under the hypothesis of uncorrelated observation errors, the matrix of the observation error covariances, R , is diagonal and the coefficients are defined by Eq. (4) where σ_o stands for the standard deviation of observation errors and δ_{ij} stands for the Kronecker delta (i.e. $\delta_{ij} = 1$ if $i = j$ and is null otherwise):

$$B_{i,j} = \sigma_b^2 (1 + \frac{r_{i,j}}{L}) \exp(-\frac{r_{i,j}}{L}) \quad (3)$$

$$R_{i,j} = \delta_{i,j} \sigma_o^2 \quad (4)$$

Different instruments can be used to measure precipitation. The most common one is the tipping bucket. Therefore, the measure can suffer from many error

sources. The most common and the most documented one results from the undercatchment due to the wind. Moreover, it has been shown that observation errors are greater when the precipitation rate is high (L'Ecuyer and Stephens, 2002). Then, to take this effect into account, a variable σ_o which is formulated following Eq. (5) and used in the European reanalysis UERRA, is considered in this study. It is expressed as

$$\sigma_o = \begin{cases} 0.001 & \text{if } RR_{24h} = 0 \text{ mm} \\ 0.7 + 0.1RR_{24h} & \text{if } 0 < RR_{24h} \leq 50 \text{ mm} \\ 5.7 & \text{if } RR_{24h} > 50 \text{ mm} \end{cases} \quad (5)$$

This equation gives a great deal of confidence to null observations, and the higher the precipitation rate, the less confidence is given in measurements. The limit for cumulative precipitation over 24h (denoted RR_{24h} hereafter) of 50 mm is fixed to be close to the ratio $\alpha = (\frac{\sigma_o}{\sigma_b})^2$ of the operational configuration of SAFRAN ($\sigma_o = 5 \text{ mm}$ and $\sigma_b = 13 \text{ mm}$) which is used for surface reanalysis at Météo-France. A small value of σ_o for null precipitation (0.001 mm) is used to avoid overestimation of light precipitation produced by the model in the UERRA reanalysis. Such use of a variable formula of σ_o has also been tested in previous studies and is operational in the MESAN system used for the HIRLAM EURO4M surface reanalysis (Landelius et al., 2016), but with a greater slope in the linear relationship between σ_o and precipitation values (0.2) and without a limit for high precipitation values. However, if the use of a variable such as σ_o shows better results in general, it highlights some drawbacks such as an over-confidence in rain gauge with problems such as undercatchment for example. Therefore, the a posteriori diagnosis applied by precipitation class proved that the error of the background standard deviation also depended on the precipitation rate (not shown here). Fixed value ($\sigma_b = 13 \text{ mm}$) is considered in this study as it is used in the UERRA reanalysis, which is the reference in surface reanalysis over Europe.

2.2 BACKGROUND DATA

As explained before, the OI algorithm combines information from a background with observations. The background field is very important for long meteorological reconstructions as it brings some information in areas where the observational data set is sparse (Donat et al., 2014). Generally, the background come from a numerical weather forecast model, except for the SAFRAN precipitation analysis in which the background is a climatological field. In this study, the French regional operational model AROME (Seity et al., 2011) at a resolution of 1.3 km (Brousseau et al., 2016) has been chosen. This is a non-hydrostatic model, which explicitly resolves deep convection and uses a cycled 3DVar analysis scheme to analyze the

state of the atmosphere at a hourly frequency and an OI scheme to analyze the two meter air temperature and humidity at a three hours frequency, and it produces the initial state for the next forecast. Regional models aim at giving better information for local scale events, such as flash floods, which often occur in the South East of France during the fall. The precipitation analysis is run in an offline mode, meaning that the operational precipitation forecasts of AROME are used as background fields in the offline analysis. To match the period of the observed precipitation, the outputs of AROME from forecast lead time +6h to +30h have been used, which correspond to 06UTC, day D to 06UTC, day D+1, as in the operational version of MESCAN used at Météo-France. In this study, the background used is referenced as AROME-1.3 km.

2.3 OBSERVATION DATA

Observations used as inputs for the MESCAN system are checked by a quality control method used at Météo-France and developed by the department in charge of climatology at Météo-France. This quality control determines stations where null (*resp. high*) values are observed while the majority of neighbouring stations do not register null (*resp. high*) values. Thus, two kriging methods were used to estimate the rain gauge based values using neighbouring stations. The rain gauge value is judged as doubtful if the difference between this estimated value and the real value is too large. In this study, for stations where there are two or more observations, only the observation where the quality control shows no doubtful flags is considered. In the case that more than one observation is flagged as good, only the first observation is taken even if another solution can be the creation of a “super” observation by taking the average between the different observations as used in (Mahfouf et al., 2007). In this way, the observation data set reached around 2500 observations each day.

2.4 GAUSSIAN TRANSFORMATION

As explained in the introduction, UERRA precipitation reanalysis is obtained thanks to an OI algorithm applied in the physical space. One of the important hypothesis of the OI algorithm is the Gaussian distribution of the innovations. However, as precipitation has a very asymmetric distribution, this hypothesis is, in general, not valid. The use of a data transformation to approach a Gaussian distribution is a common method to overcome this problem. Many transformations have been tested for several studies and it appears that the Box-Cox transformation is the best adapted for precipitation (Legates and Willmott, 1990), and it is also rather easy to implement. As a consequence, in this study, a Box-Cox transformation of variable y (Eq. (6)) is used to calculate the analysis. The Box-Cox transformation is defined

by the tuple $\lambda = (\lambda_1, \lambda_2)$, hereafter called the Box-Cox parameters. The variable in the transformed space is defined as:

$$y^{(\lambda)} = BC^{(\lambda)}(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{if } \lambda_1 \neq 0 \\ \log(y + \lambda_2) & \text{if } \lambda_1 = 0 \end{cases} \quad (6)$$

where y is the original variable in the physical space, and $y^{(\lambda)}$ is the Box-Cox transformation of y . When the Box-Cox transformation is not defined (i.e. when $\lambda_1 = 0$ and $y + \lambda_2 \leq 0$ or $\lambda_1 \neq 0$ and $y + \lambda_2 < 0$), $BC^{(\lambda)}(y)$ is set to 0. It is important to remark that the Box-Cox transformation can not be applied to negative values of $y + \lambda_2$ when $\lambda_1 \neq 1$. Different studies have been conducted to optimize the Box-Cox parameters (Erdin et al., 2012; Ishak and Ahmad, 2018). Moreover, it can be seen that for $\lambda = (\lambda_1 \rightarrow 0, \lambda_2 = 1)$, the Box-Cox transformation approaches the logarithm transformation used in (Mahfouf et al., 2007).

However, the analysis in the transformed space does not bring a valuable information and can not be used directly by downstream application such as hydrology as hydrological models need physical precipitation values as input. Therefore, it is necessary to define the analysis in the physical space so the Eq. (6) has to be inverted. As the Box-Cox transformation is reversible, it is possible to define the real inverse Box-Cox transformation with Eq. (7) as

$$Y = (BC^{(\lambda)})^{-1}(y^{(\lambda)}) = \begin{cases} (y^{(\lambda)}\lambda_1 + 1)^{\frac{1}{\lambda_1}} - \lambda_2 & \text{if } \lambda_1 \neq 0 \\ \exp(y^{(\lambda)}) - \lambda_2 & \text{if } \lambda_1 = 0 \end{cases} \quad (7)$$

As for the Box-Cox transformation, it is considered that $(BC^{(\lambda)})^{-1}(y^{(\lambda)}) = 0$ when $\lambda_1 \neq 0$ and $y^{(\lambda)}\lambda_1 + 1 < 0$. It is easy to demonstrate that $Y = y$.

In order to obtain an innovation distribution as close as possible to a Gaussian distribution, a study to determine the optimal values of the Box-Cox parameters, and especially of λ_1 has been done. In this study, the transformed innovation is defined as $y^{(\lambda)} = BC^{(\lambda)}(y^o) - BC^{(\lambda)}[H(x^b)]$. The transformed innovation is not defined as $y^{(\lambda)} = BC^{(\lambda)}[y^o - H(x^b)]$ because the input of the Box-Cox transformation has to be positive (see the previous paragraph) which is not always the case for the variable $y^o - H(x^b)$. It is, in fact, the main goal of the parameter λ_2 . Therefore, this parameter has to be calculated every day and is, in fact, highly variable every day (not shown here). Thus, (Fortin et al., 2015) has shown that the intermittent nature of precipitation is only kept by considering $\lambda_2 = 0$ after the use of the bias correction (see section 2.6). As a result, in the following of this study, λ_2 is set to 0. As it is mainly the asymmetric property of the precipitation which make it a non-Gaussian variable, only $\lambda = (\lambda_1, 0)$ where $\lambda_1 \in [0, 1]$ are considered. More precisely, only $\lambda_1 = \frac{1}{k}, k \in \mathbb{N}^*$ are considered for reasons explained in

section 2.6. To determine optimal value of λ_1 , Quantile-Quantile-plots (QQ-plots) of the variable $y^{(\lambda)}$ are calculated for many values of λ_1 , which express the true quantiles of the variable compared to the quantiles of the variable if it had a Gaussian distribution. These plots give a line and a cloud of points for which we can calculate the correlation factor. The more this correlation factor is close to 1, the more $y^{(\lambda)}$ has a Gaussian distribution.

In this study, two different methods are used to determine the optimal value of λ_1 . The first method gives a time series of optimal values of λ_1 (Figure 1a). To obtain this time series, the correlation factor is calculated daily for each value of λ_1 . The daily optimal value of λ_1 is obtained by considering the λ_1 value corresponding to the higher value of the correlation factor (Figures 1a,b and c). The second method gives the λ_1 values which are used in this study. The correlation factor is calculated each day for each λ_1 value and then, the mean value of this correlation factor over the whole period (Figure 1d) is examined : if this mean correlation factor is superior to 0.85, then it is considered that the λ_1 value can be used for the OI analysis.

In Figure 1a, the time series of running average over 150 days – to avoid noise as optimal λ_1 values are highly variable from one day to another – of optimal values of λ_1 is plotted. It can be seen that optimal values of λ_1 have a seasonal variation with greatest values during autumn and winter and lower ones during spring and summer, with the mean value over the whole period being $\lambda_1 = 0.28$. This is coherent with the type of precipitation during winter, which is essentially stratiform and thus is characterized by large scale features and as a consequence, the precipitation transformation is less useful than during summer, when convective events often occur (which can be very located events). It can be seen that λ_1 values are lower during spring 2017, which can be explained by the fact that this spring was a very dry one compared to the other springs considered in this study (total mean observed precipitation are 3.04 mm for spring 2016, 2.35 mm for spring 2017 and 3.41 mm for spring 2018). The time series of running average over 150 days of correlation factors corresponding to the optimal λ_1 time series are plotted in Figure 1b and the distribution of correlation factor over the study period is plotted in Figure 1c. It can be seen that the majority of the coefficients are between 0.9 and 1, which proves the ability of the Box-Cox transformation to give a Gaussian distribution to the innovations. An important correlation between λ_1 values and the correlation factor should be noted : the more the values of λ_1 are small (resp. high), the more the correlation factor is small (resp. high).

The mean value of the correlation factor for each λ_1 value is shown in Figure 1d. It should be noted that for $\lambda_1 = 0$, the logarithmic transformation used in (Mahfouf et al., 2007) ($\lambda_1 = 0$ and $\lambda_2 = 1$) is considered in this study. It can be seen that the correlation factor allows a maximum

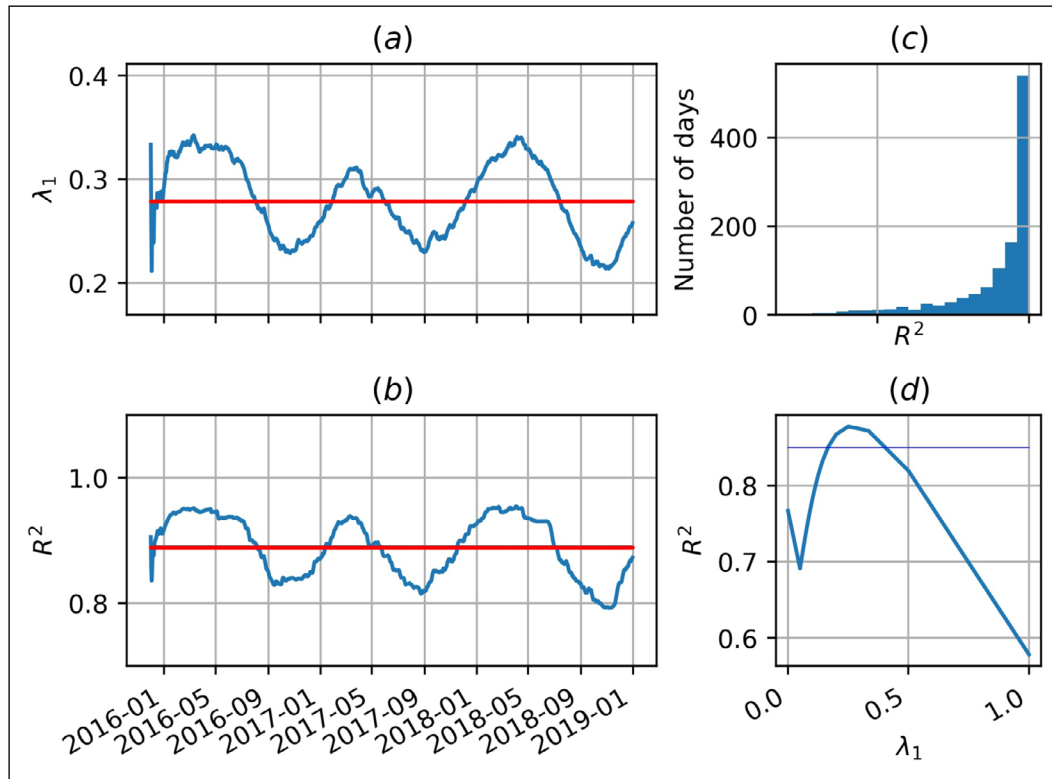


Figure 1 Determination of optimal values of λ_1 over the whole period : (a) is the time series of optimal λ_1 values (in blue) and its mean value (in red), (b) is the associated correlation factor, with daily values (in blue) and its mean value (in red), (c) is the distribution of the correlation factor over the whole period and (d) are the mean correlation factors over the whole period for each λ_1 value.

value for $\lambda_1 = \frac{1}{4}$ and is higher than 0.85 for $\lambda_1 = \frac{1}{3}$. From this figure, two important conclusions can be made : the minimum of the correlation factors is obtained for $\lambda_1 = 1$, which represents the physical space so this result proves the importance of Box-Cox transformation to give a Gaussian distribution for innovations. Thus, it can be seen that for $\lambda_1 = 0$ (logarithmic transformation), there is an increase of the correlation factor, which reinforces the use of logarithm transformation for precipitation analysis in previous studies.

From these results, the Box-Cox parameters $\lambda = (\frac{1}{4}, 0)$ and $\lambda = (\frac{1}{3}, 0)$ were used to calculate the precipitation analysis. A comparison between the scores of the analyses compared to independent observations (see section 3.1) show that the transformed back into physical space using the Box-Cox transformation with $\lambda = (\frac{1}{3}, 0)$ works better than using the Box-Cox transformation with $\lambda = (\frac{1}{4}, 0)$. Indeed, the bias - which refers here as the bias of the model (or the analysis) compared to observation data set, following Eq.(27) - (resp. RMSE) of the analysis using $\lambda = (\frac{1}{3}, 0)$ is -0.11 mm (resp. 2.38 mm) while the bias (resp. RMSE) of the analysis using $\lambda = (\frac{1}{4}, 0)$ is -0.13 mm (resp. 2.47 mm). As a result, $\lambda = (\frac{1}{3}, 0)$ was retained in the following of this study.

2.5 GEOSTATISTICAL DETERMINATION OF THE BACKGROUND AND OBSERVATION ERRORS

To estimate statistical errors for the observations and the background, (Desroziers et al., 2005) used an iterative

method, meaning the analysis has to be run more than one time. As this method is relatively expensive, another method called the semi-variogram technique is used in this study to determine these errors. This technique is used in the Canadian Precipitation Analysis (CaPA) system (Lespinas et al., 2015) and it relies upon the fact that two observations are more correlated if they are near than if they are far away from one another. Then, the semi-variogram permits the calculation of the correlation between two points separated by a distance, h . In this section, the different steps leading up to the determination of σ_o^{BC} (observation standard deviation error in Box-Cox space) and σ_b^{BC} (background standard deviation error in Box-Cox space) are described.

The daily experimental semi-variogram per distance class is defined as:

$$\gamma_e(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (8)$$

where $N(h)$ is the number of observations taken into account for each distance class, Z stands for the variable considered (here the transformed innovation defined in the previous section), $Z(x_i)$ is the innovation at an observed station i , and $Z(x_i + h)$ is the innovation at a station separated from the previous one by a distance of h .

In order to have enough observations in each distance class, the number of pairs used, $N(h)$, and the sum of the innovation differences are calculated over a period of

several days $N(d)$. Then, the semi-variogram used in this study is defined by:

$$\overline{\gamma_e}(h) = \frac{1}{\sum_{j=1}^{N(d)} 2N(h)_j} \sum_{j=1}^{N(d)} \sum_{i=1}^{N(h)} [Z(x_i)_j - Z(x_i + h)_j]^2 \quad (9)$$

where j relates the fact that it is calculated on the j^{th} day. A study (not shown here) has been done to determine the best value of the number of days to take into account in order to have enough observations in each distance class, but, also to prevent a signal from being smoothed too much. As a result, a period of 30 days is chosen in this study.

The use of semi-variograms implies that precipitation field is assumed to be spatially stationary, which is not really a good assumption as different mountainous areas play an important role in France such as the Pyrenees, the Alps, the Corsican mountains and the Massif Central. Wang et al. (2019) tuned Eq.(8) to take into account the elevation. Even if it has been shown that in complex terrain, the tuned equation provides better results, the use of classical semi-variogram equation (Eq. (8)) is used in this study.

Daily theoretical semi-variograms are obtained using daily experimental semi-variograms obtained with Eq. (9). Both semi-variograms are plotted in Figure 2a. The theoretical semi-variogram is obtained by fitting an experimental semi-variogram. σ_o^{BC} correspond to the theoretical semi-variogram value for the lower distance, σ_b^{BC} is the asymptotic value of theoretical semi-variogram and L is the distance for which 75% of the asymptotic value is obtained (Figure 2b).

Two different methods have been tested in this study. First, all observed stations are used to calculate the semi-variogram (Method 1). Then, a new sample has been used by considering only stations where both observation and background are strictly greater than zero to calculate the semi-variogram (Method 2), because precipitation measurement errors only occur when there

is precipitation. For method 2, between 40% and 80% (resp. 20% and 70%) of the stations are not considered for observations (resp. background) for the semi-variogram calculation. Results are shown in Figure 3 for the Box-Cox transformation with $\lambda = (\frac{1}{3}, 0)$. It can be seen that both methods provide similar values of σ_o^{BC} during the whole period while some differences can be pointed out for σ_o^{BC} values. Indeed, it can be seen that even if σ_o^{BC} values are similar for both methods, σ_o^{BC} values are, in general, higher (resp. lower) during summer (resp. winter) for the Method 2. Higher values of σ_o^{BC} are related to periods when there is a lot of null observed precipitation. When these null precipitation observations are deleted in the data set, the daily mean observed precipitation is then highly increased. These results are in agreement with the variable formula of σ_o purposed for MESCAN in the UERRA project (the weaker the precipitation, the more observation can be trusted, see Eq. (5) and section 2.1). From now, values obtained with Method 2 will be considered in the MESCAN analysis (Figure 4). It can be seen that there is a seasonal variation of σ_o^{BC} with higher (resp. lower) values during summer (resp. winter). This is due to the fact that winter contains, in general, stratiform precipitation with a large amplitude compared to local structures which occurs during summer, which can be more complicated to be well-predicted since the scales considered can be lower than the model resolution. On the other hand, σ_o^{BC} values are very consistent during the study period and there is not a visible seasonal variation even if summer is associated with higher σ_o^{BC} values (especially during summer 2017 and summer 2018). These results are consistent with those obtained by (Wang et al., 2019) and (Soci et al., 2016).

2.6 BIAS CORRECTION

2.6.1 Origins of the bias

In order to correct the bias induced by the data transformation and its inverse transformation, it is important to understand the origins of this bias. In this part, only the case $\lambda_1 \neq 0$ has been considered. As

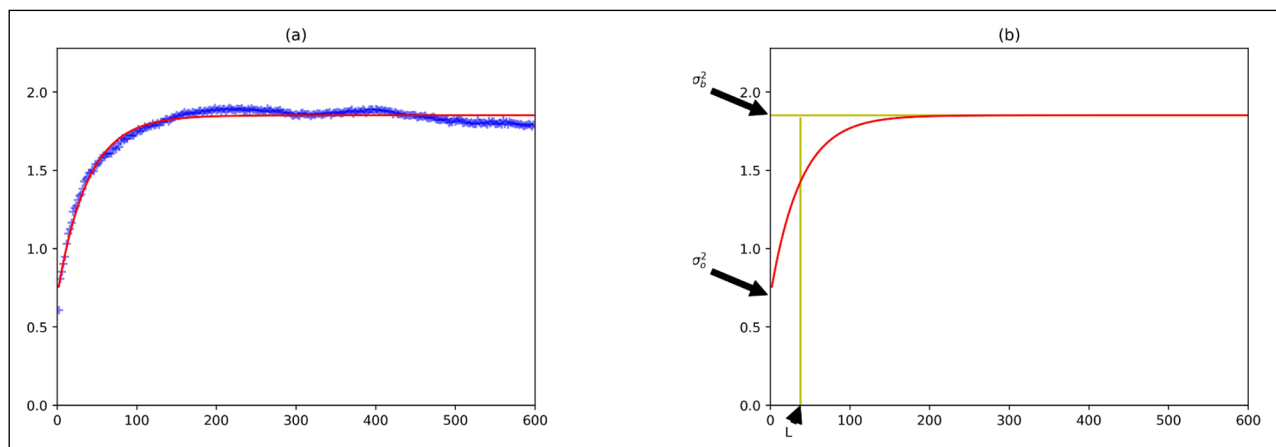


Figure 2 Experimental (blue cross) and theoretical (red line) semi-variograms (a) and determination of σ_o^{BC} , σ_b^{BC} and L (b).

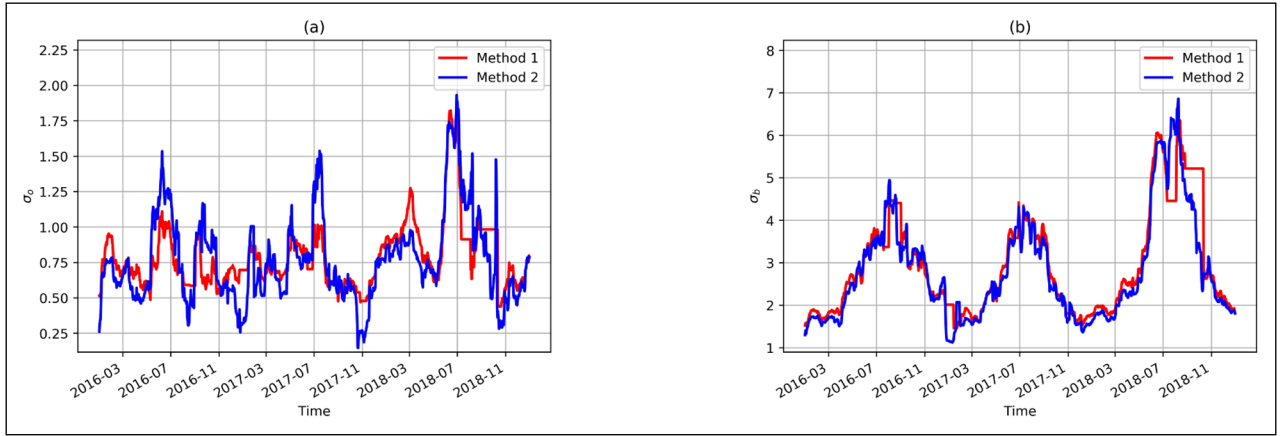


Figure 3 Daily values of σ_o^{BC} (a) and σ_b^{BC} (b) for the two methods.

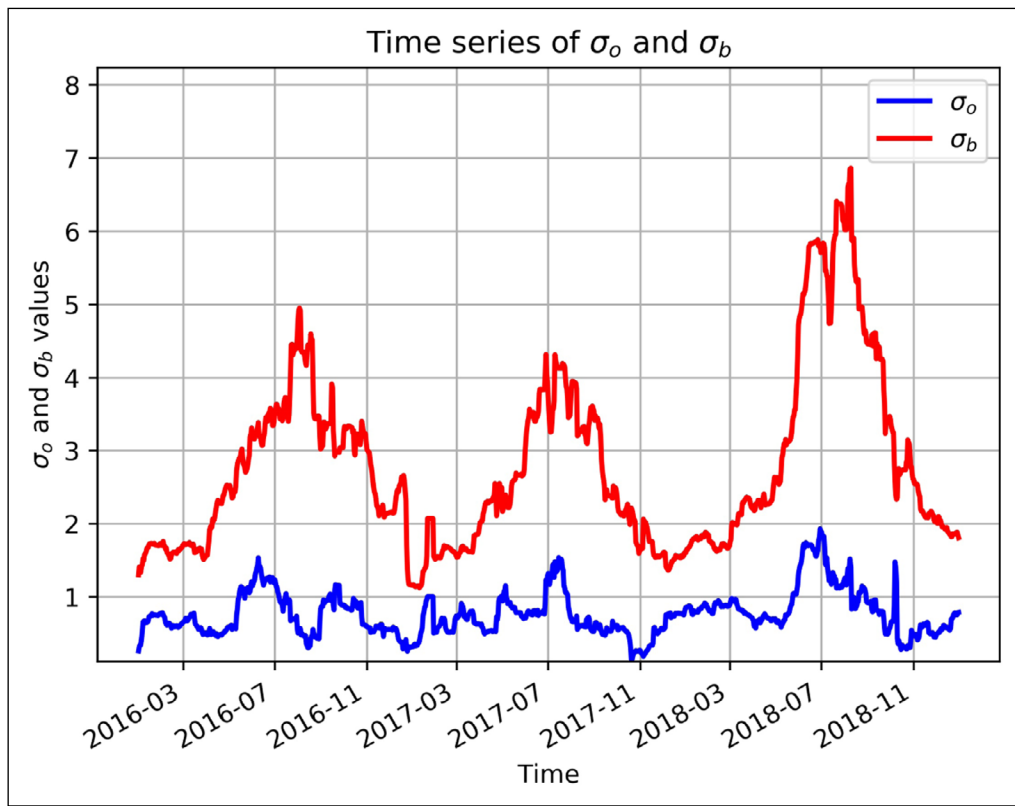


Figure 4 Daily values of σ_o^{BC} and σ_b^{BC} for the Box-Cox transformation with $\lambda = (\frac{1}{3}, 0)$.

explained before, both the Box-Cox transformation (Eq. (6)) and its inverse function (Eq. (7)) are non-linear transformations. The analysis state in the transformed space, $(x_i^a)_T$, is calculated from:

$$(x_i^a)_T = BC^{(\lambda)}(x_i^b) + \sum_{j=1}^p (w_{ij})_T \left\{ BC^{(\lambda)}(y_j^o) - H[BC^{(\lambda)}(x^b)]_j \right\} \quad (10)$$

where $(w_{ij})_T$ stands for the weights calculated in the transformed space. In this study, $H(x^b)_j$ is obtained by considering the nearest model point from the observation point j . As a consequence, one can deduce:

$$H[BC^{(\lambda)}(x^b)]_j = BC^{(\lambda)}[H(x^b)_j] \quad (11)$$

It can be seen from Eq. (10) that values transformed by the Box-Cox transformation are added. However, the Box-Cox transformation is defined only for positive inputs and the results interval is bounded to the left:

$$BC^{(\lambda)}: [-\lambda_2, +\infty[\rightarrow \left[\frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}, +\infty[\right.$$

As a result, the inverse Box-Cox transformation is defined as:

$$(BC^{(\lambda)})^{-1}: \left[\frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}, +\infty[\rightarrow [-\lambda_2, +\infty[\right.$$

Then, if we consider $x, y \in [-\lambda_2, +\infty[$, then $BC^{(\lambda)}(x) + BC^{(\lambda)}(y) \in [2 \frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}, +\infty[$. For now, it can be deduced that inverse Box-Cox transformation can not

always be applied to added values transformed by the Box-Cox transformation. Therefore, in order to use any values, the Box-Cox transformation has been extended as:

$$y^{(\lambda)} = BC^{(\lambda)}(y) = \begin{cases} \frac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{if } y \in [-\lambda_2, +\infty[\\ 0 & \text{if } y \in]-\infty, -\lambda_2[\end{cases} \quad (12)$$

It is important to note the discontinuity of the Box-Cox transformation at 0. Indeed, for $y < 0$, $BC^{(\lambda)}(y) = 0$ while $BC^{(\lambda)}(0) = \frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}$. As a consequence, by considering $\lambda_2 = 0$ and $\lambda_1 = \frac{1}{k}$, $k \in \mathbb{N}^*$, then $BC^{(\lambda)}(0) = -k$. Then, the larger the value of k , the more the singularity in 0 for the Box-Cox transformation is large.

In the same way, inverse Box-Cox transformation has been extended as:

$$y^{(\lambda)} = BC^{(\lambda)}(y) = \begin{cases} \left(y^{(\lambda)}\lambda_1 + 1\right)^{\frac{1}{\lambda_1}} - \lambda_2 & \text{if } y^{(\lambda)} \in \left[\frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}, +\infty\right[\\ 0 & \text{if } y^{(\lambda)} \in]-\infty, \frac{\lambda_2^{\lambda_1} - 1}{\lambda_1}] \end{cases} \quad (13)$$

A very simple example to understand the effect of the non-linearity of the Box-Cox transformation in the analysis process is explained here. We consider a fixed background value x_i^b (here we will consider the background value to be 0 mm, 5 mm, 10 mm and 200 mm). If only one observation

is used to correct the background, the analysis state in physical space (denoted x_i^a hereafter) is defined as follows:

$$x_i^a = x_i^b + y_j^o - H(x_j^b) \quad (14)$$

whereas in transformed space, the analysis state (noted $x_{i,T}^a$ hereafter) is defined as:

$$x_{i,T}^a = \left(BC^{(\lambda)}\right)^{-1} \left\{ BC^{(\lambda)}(x_i^b) + BC^{(\lambda)}(y_j^o) - BC^{(\lambda)}[H(x_j^b)] \right\} \quad (15)$$

where λ is fixed to $(\frac{1}{3}, 0)$ in this part.

We consider a fixed value of $y_j^o = 3 \text{ mm}$ and $H(x_j^b)$ values vary to correct the background from -2.5 mm to 2.5 mm (increment values, i.e. $y_j^o - H(x_j^b)$ in Eq. (14)) in physical space. Therefore, in the case in which $x_i^b = 0 \text{ mm}$, only positive increments are considered (from 0 mm to 3 mm) in order to consider positive analysis values. Figure 5 shows the values of x_i^a and $x_{i,T}^a$ for background values of 0 mm, 5 mm, 10 mm and 200 mm ((a), (b), (c) and (d) respectively). It can be seen on this figure that if the background is null, the analysis calculated in the Box-Cox space will always underestimate the precipitation. Then, if the background is not null, the analysis in the Box-Cox space will underestimate the precipitation when the increment is negative and it will overestimate the precipitation when the increment is positive.

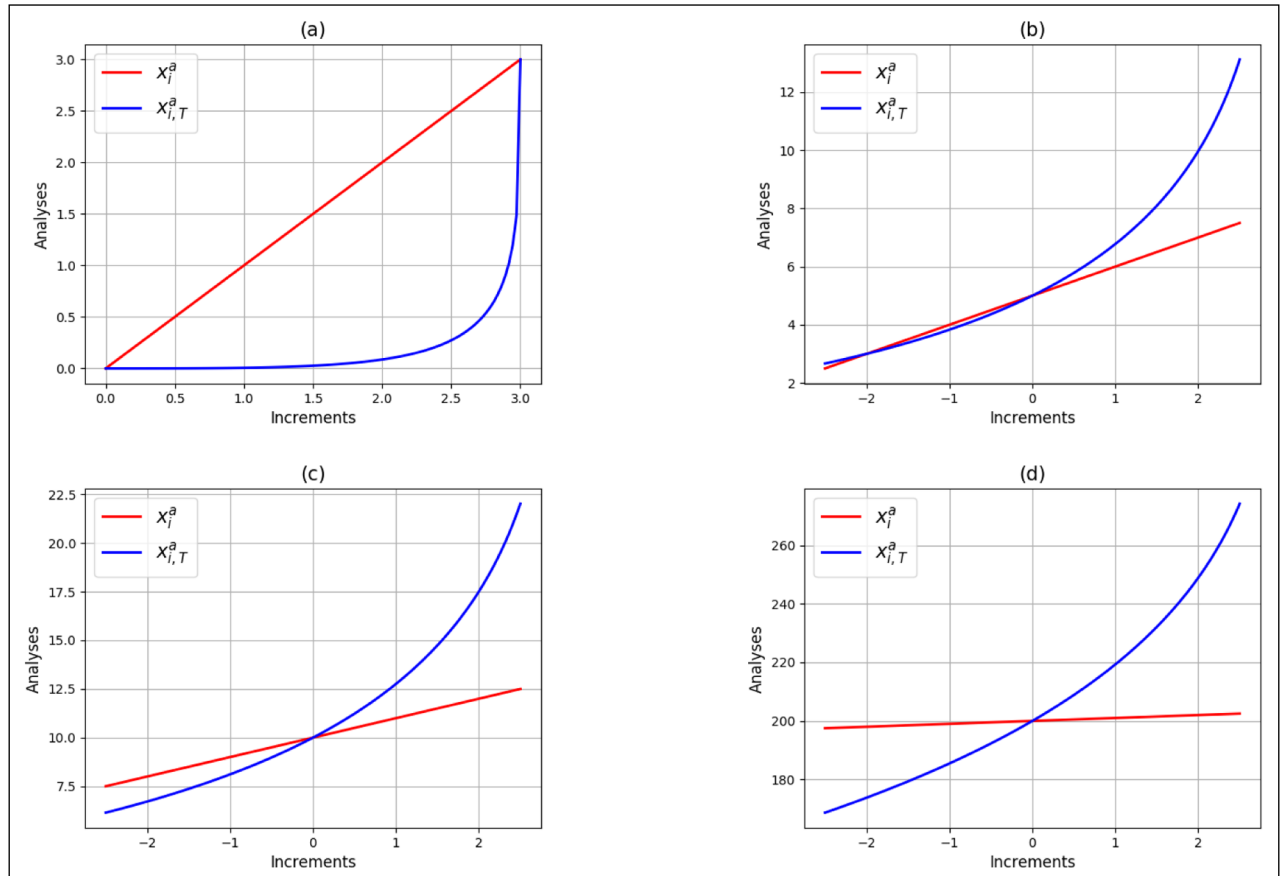


Figure 5 Put on evidence of the bias induced by the Box-Cox transformation. The red line stands for the analysis wanted (x_i^a) and the blue line stands for the analysis obtained after the Box-Cox transformation ($x_{i,T}^a$) for background values fixed at 0 mm (a), 5 mm (b), 10 mm (c) and 200 mm (d).

Another explanation of the bias can be found in Fletcher and Zupanski (2006). In Box-Cox space, background and observation errors are normally distributed. As a result, the analysis error is also normally distributed, and symmetric around the mean. Therefore, mean, median and mode are the same. As a result, the best estimate (analysis) can be chosen as the median, the mean or the mode, which makes no difference. However, inverse Box-Cox transformation has to be applied to the analysis to obtain precipitation in the physical space. After this step, the precipitation is no more normally distributed. As a result, the choice of the mean, median or mode will have an impact on the analysis, and can be another source of bias.

2.6.2 Bias correction

Using a transformed variable is a very efficient method for resolving the non-Gaussian distribution of the variable, but it also introduces a bias when moving back to physical space due to the non-linear aspect of the transformation (see the previous section). Avoid solving the induced bias can be a source of big error in the analysis and this type of systematic error can be detrimental for hydrological applications. For example, it is possible to run the precipitation analysis in logarithmic space with MESCAN but since the bias correction is not addressed, this option was not used in the UERRA surface reanalysis. CaPA uses the Box-Cox transformation and Evans (2013) developed a bias correction formula. In this study, a slightly transformed CaPA bias correction formula is used.

The derivation of the bias correction formula is given in the Appendix C. To determine this formula, a Taylor series decomposition of Eq. (7) is used. An important remark is that Eq. (7) can be derived for order n , $\forall n \in \mathbb{N}^* \setminus \{1\}$, using a recursive relation:

$$\left[(BC^{(\lambda)})^{-1} \right]^{(n)} (y^{(\lambda)}) = \prod_{i=1}^{n-1} (1 - i\lambda_1) (y^{(\lambda)} \lambda_1 + 1)^{\frac{1}{\lambda_1} - n} \quad (16)$$

It can be deduced from Eq. (16) that if $\lambda_1 = \frac{1}{k}$ with $k \in \mathbb{N}^*$ then $\forall l > k$ so that

$$\left[(BC^{(\lambda)})^{-1} \right]^{(l)} (y^{(\lambda)}) = 0 \quad (17)$$

As the Taylor decomposition is truncated, it is more accurate to have $\lambda_1 = \frac{1}{k}$, $k \in \mathbb{N}^*$ to neglect as few terms as possible for the bias correction, with the lowest value of k possible. Thus, for a normal variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then, $\forall n \geq 1$, $E[(X - \mu)^{2n+1}] = 0$ and $E[(X - \mu)^{2n}] = \frac{(2n)!}{2^n n!} \sigma^{2n}$.

So, by considering that $y^{(\lambda)} \sim \mathcal{N}[(x^a)^{(\lambda)}, (\sigma_a^{BC})^2]$ where σ_a^{BC} stands for the standard deviation of the analysis error in the transformed space and is defined by (Daley, 1993):

$$\left[(\sigma_a^{BC})^2 \right]_j = (\sigma_b^{BC})^2 - \sum_{i=1}^N (w_{i,j})_T (b_i)_T \quad (18)$$

Then it is easy to demonstrate that:

$$\forall n \in \mathbb{N}, E \left[\left[y^{(\lambda)} - (x^a)^{(\lambda)} \right]^{2n+1} \right] = 0 \quad (19)$$

$$\forall n \in \mathbb{N}^*, E \left[\left[y^{(\lambda)} - (x^a)^{(\lambda)} \right]^{2n} \right] = \frac{(2n)!}{2^n n!} (\sigma_a^{BC})^{2n} \quad (20)$$

Thus, for $k = 2$ or $k = 3$, Eq. (6) is equivalent to a square-root or a cubic-root transformation, respectively. Then, for $\lambda_1 = \frac{1}{3}$, the corrected analysis values will be defined by:

$$\begin{aligned} (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right]_{\text{corr}} &= (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \\ &+ (\sigma_a^{BC})^2 \frac{1}{2} (1 - \lambda_1) \left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \lambda_2 \right]^{1-2\lambda_1} \end{aligned} \quad (21)$$

For $\lambda_1 = \frac{1}{4}$, the corrected analysis values will be defined by:

$$\begin{aligned} (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right]_{\text{corr}} &= (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \\ &+ (\sigma_a^{BC})^2 \frac{1}{2} (1 - \lambda_1) \left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \lambda_2 \right]^{1-2\lambda_1} \\ &+ (\sigma_a^{BC})^4 \frac{1}{8} (1 - \lambda_1) (1 - 2\lambda_1) (1 - 3\lambda_1) \\ &\left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \lambda_2 \right]^{1-4\lambda_1} \end{aligned} \quad (22)$$

The correction for the analysis calculated in the logarithm space is:

$$\begin{aligned} (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right]_{\text{corr}} &= (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \\ &+ (\sigma_a^{BC})^2 \frac{1}{2} \left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \lambda_2 \right] \end{aligned} \quad (23)$$

$$\begin{aligned} (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right]_{\text{corr}} &= (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \\ &+ (\sigma_a^{BC})^2 \left(1 - \frac{(\sigma_a^{BC})^2}{(\sigma_b^{BC})^2} \right) \frac{1}{2} (1 - \lambda_1) \\ &\left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \lambda_2 \right]^{1-2\lambda_1} \end{aligned} \quad (24)$$

It can be seen from Eq. (18) that σ_a^{BC} is a spatially variable coefficient, meaning that there are some places where the analysis is more accurate. As this parameter is used as a coefficient product in the Eq. (21) and Eq. (22), it can be deduced that the better the analysis, the less precipitation is added owing to the bias correction. As stated in (Fortin et al., 2015), if the analysis is null, it is expected that the bias correction does not bring any precipitation, which implies the choice of $\lambda_2 = 0$. The bias correction formula used in this study is therefore slightly different from the one used in CaPA. Indeed, Eq. (24) is used in the CaPA analysis for which a factor $1 - \frac{(\sigma_a^{BC})^2}{(\sigma_b^{BC})^2}$ is accounted for avoid adding precipitation when no observation contributes to the analysis (i.e. when $(\sigma_a^{BC})^2 = (\sigma_b^{BC})^2$). In this study, it is decided to remove this coefficient because it is not derived from the mathematical demonstration of bias correction.

3 RESULTS

3.1 TOOLS FOR EVALUATION

Using the same quality control as that used to produce the observations used as input to the MESCAN system, 930 observations have been picked to make a data set for the evaluation. This data set is independent of the data set used as input for MESCAN analysis. The locations and the altitudes of these observations are shown in Figure 6. Even if the number of observations is high, some areas are not well-covered by this data set, such as the mountainous areas (Pyrenees, Alps, Massif Central and Corsican mountains).

Analysis results have been evaluated over the period from 01/01/2016 to 31/12/2018 in order to be able to use the operational outputs from AROME-1.3 km. The evaluation is done at the observations' location corresponding to the circles in Figure 6(a). As a consequence, analysis outputs are interpolated to the observation locations using a 16-point inverse distance weighted mean interpolation. The use of 16 points to interpolate analyses and background to observation points was motivated to avoid penalizing them if a precipitating event is detected but poorly located.

Different scores were calculated such as the bias, the Root-Mean Square Error (RMSE) which give spatial and general information, the Heidke Skill Score (HSS) and the Frequency Bias Index (FBI) which give information

on analysis accuracy and bias frequency for different precipitation classes. The definition of these scores are given in Appendix B. In order to evaluate the analysis, the number of dry days, defined as the days where $RR_{24h} \leq 0.1$ mm produced by the analysis and by the observations are also compared.

In this study, MESCAN with variable σ_0 (Eq. (5)) is considered as the reference and outputs from the analysis made in the Box-Cox space with $\lambda = (\frac{1}{3}, 0)$ will be evaluated.

3.2 RESULTS OVER THE WHOLE PERIOD

First, the scores were calculated over the whole period from 01/01/2016 to 31/12/2018. In Table 1, the bias and the RMSE of the different experiments are shown. This table also provides information about the difference of the accumulated daily precipitation averaged over France between the observations and the different experiments (see Figure 7) at the end of the period. The background AROME-1.3 km exhibits a very small bias which is in fact the result of a seasonal bias which can be seen in the Figure 7 and will be discussed in the next section. It is found that each analyses underestimates the precipitation. Indeed, the reference largely underestimates the precipitation over the whole period with a constant bias. The dry bias induced by the Box-Cox transformation compared to observations is shown in Table 1. This table also illustrates the efficiency of the bias correction proposed by Eq. (21),

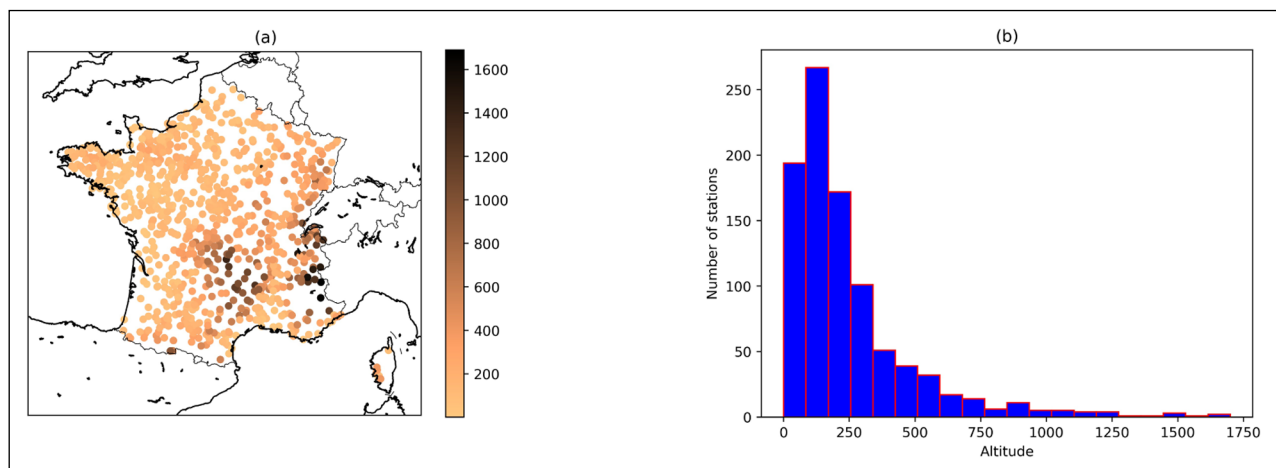


Figure 6 Spatial coverage of the evaluation observations' altitudes (m) (a) and number of stations per altitude (b).

EXPERIMENT		BIAS (mm)	RMSE (mm)	DIFFERENCE AT THE END OF THE PERIOD
AROME1-3 km		$1.73 \cdot 10^{-2}$	3.85	+18.91 mm (+0.71%)
Reference		-0.17	2.54	-180.91 mm (-6.84%)
Box-Cox	Back in physical space	-0.11	2.38	-118.63 mm (-4.48%)
	Correction (Eq.(21))	$-5.41 \cdot 10^{-2}$	2.40	-59.24 mm (-2.24%)
	Correction (Eq. (24))	$-5.67 \cdot 10^{-2}$	2.39	-62.14 mm (-2.35%)

Table 1 Values of the scores for the different experiments.

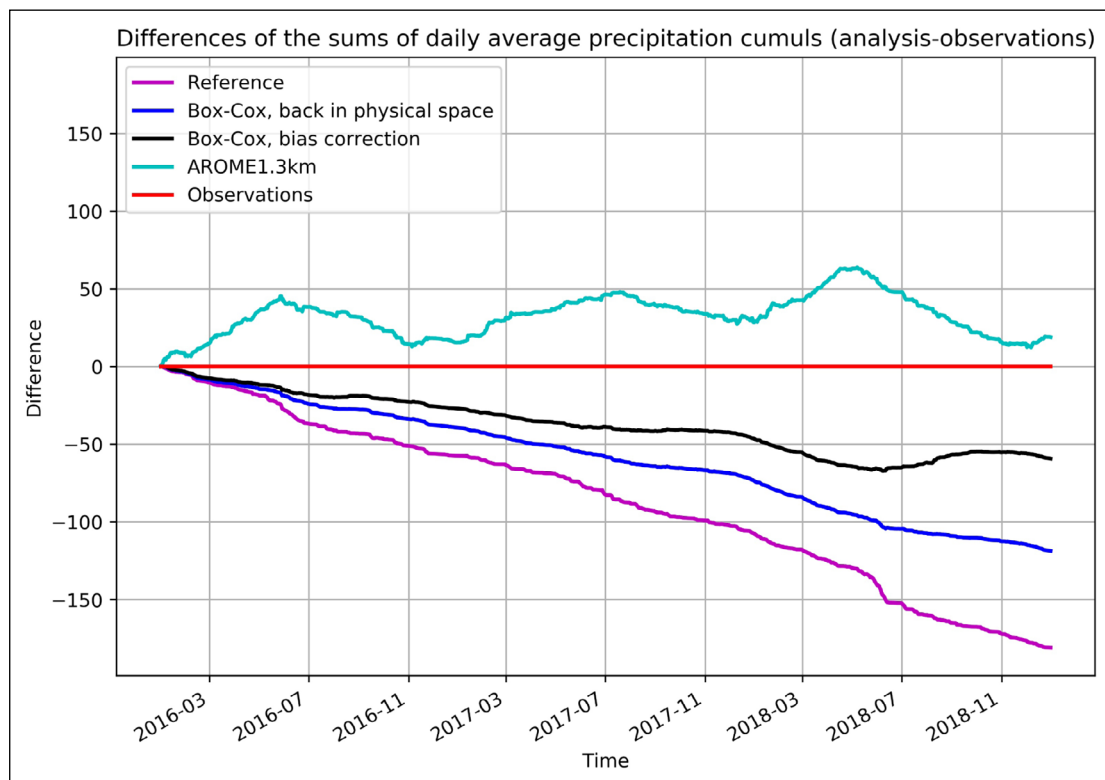


Figure 7 Differences of the accumulated daily precipitation averaged over France between the observations and the different experiments.

which is well adapted here as the analysis after the bias correction is slightly biased compared to the analysis transformed back into physical space. It should be noted that the use of Eq. (24) causes less precipitation than the use of Eq. (21), even if the RMSE of the experiment is slightly better. An important impact can be seen on June 2018, where the underestimation of the reference is higher while the dry bias of the Box-Cox experiment after the bias correction is lower. Therefore, the analysis is useful as it well reduces the RMSE of the background which is relatively high. The Box-Cox transformation is efficient since it reduces the RMSE values compared to the reference. On the other hand, an important result shown here is the slightly higher value of the RMSE after the bias correction compared to the value of transformed back in physical space, meaning that even if the bias induced by the Box-Cox transformation is well corrected, more random errors are made.

The spatial coverage of the bias is plotted in the Figure 8 which highlights that the dry bias of the reference, previously discussed is a result of a general dry bias over the France even if the reference tends to overestimate precipitation in the north-west of France. The comparison of the background to the observations shows an overestimation of the precipitation in northern France, and an underestimation in southern France. Thus, the background brings too much precipitation in the plains compared to mountainous regions. The dry bias induced by the Box-Cox transformation in the analysis process is also shown in this figure. Indeed, as was the case for the reference, the Box-Cox analysis tends to

underestimate the precipitation over the majority of the evaluation stations. The lower value of the bias of the Box-Cox analysis in Table 1 is explained by lower values of the underestimation at evaluation stations in general. The impact of the bias correction is largely visible because there are more stations where the analysis after correction overestimates the precipitation. Analysis done with the Box-Cox transformation does not reduce the humid bias in the north-west of France nor the dry bias in the Massif Central, but it corrects well the dry bias in Corsica. It is interesting to note here that the bias values of the analysis after correction are, in general, higher than values computed with a transformed back into physical space.

The spatial coverage of the RMSE is shown in Figure 9 for the reference along with the percent improvement for the other experiments compared to the reference. It can be seen that the RMSE values are low over the majority of France, even if there are high values at some stations in western France. Higher values can be seen in southern France, especially where the “cevenol” episodes occur. The background RMSE is better only at three stations : where the RMSE of the reference is high and in the Alps (around the Mont Blanc), where there are fewer observations available to integrate in the MESCAN analysis. The Box-Cox transformation generally improves the RMSE values over the whole country, especially in the north-western part of France. The analyses after the transformed back into physical space and after the bias correction have almost the same RMSE. The global improvement of RMSE values thanks to the Box-Cox transformation is consistent with the RMSE values in Table 1.

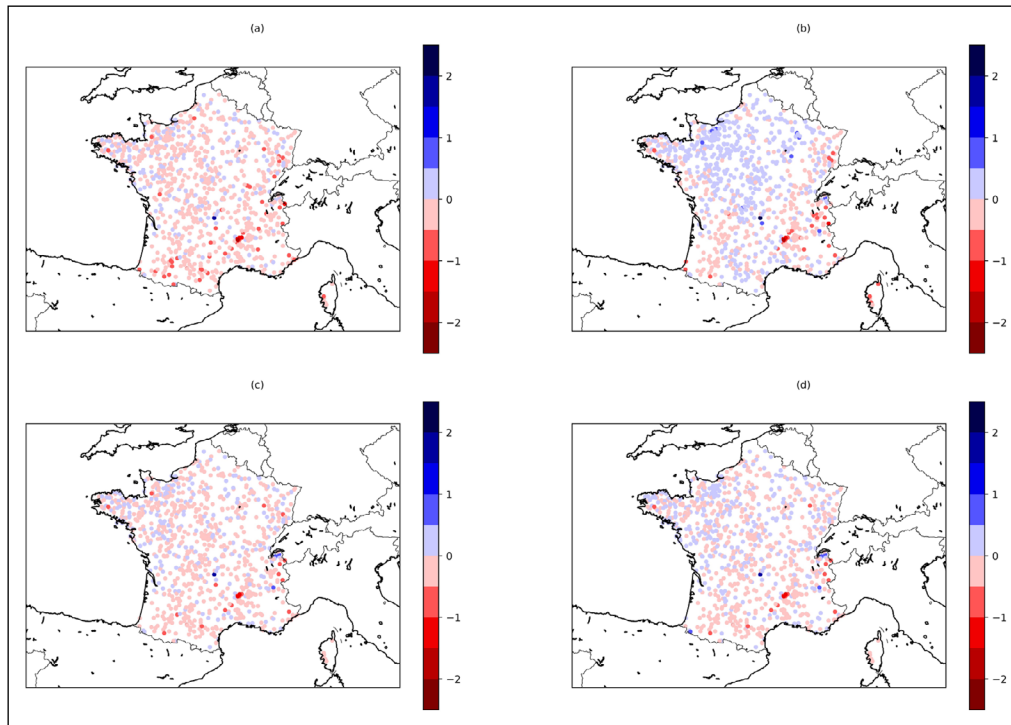


Figure 8 Bias over the whole period for the reference (a), the background (b), the Box-Cox experiment with a simple back in physical space (c) and after bias correction (d). Blue points mean an overestimation while red ones mean an underestimation.

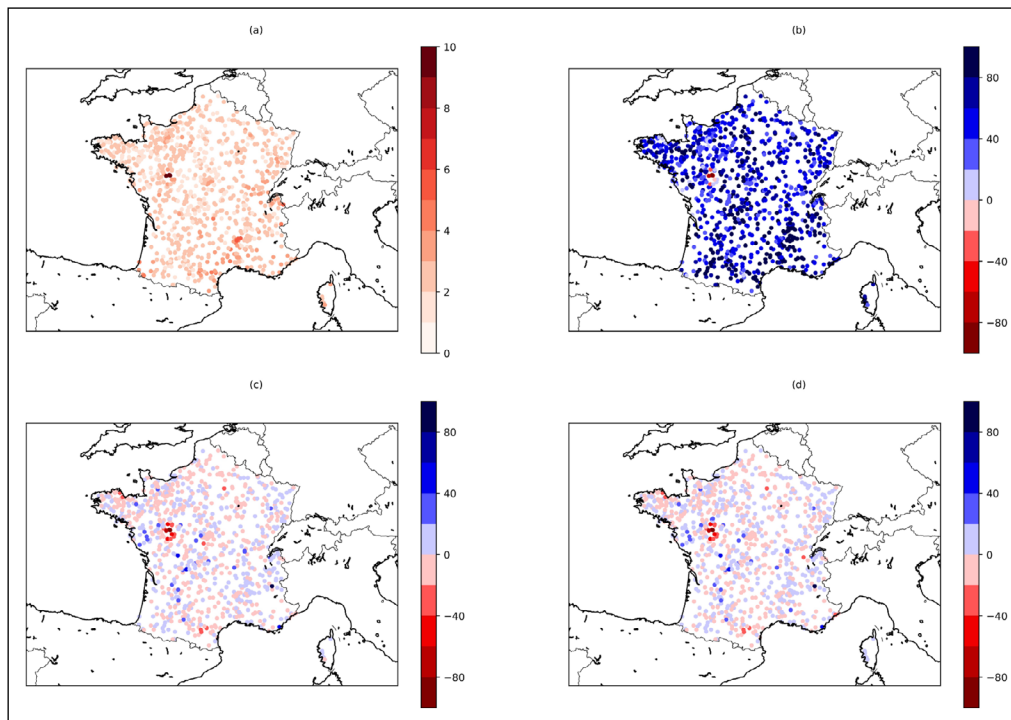


Figure 9 RMSE over the whole period for the reference (a) and percent of improvement for the background (b), the Box-Cox experiment with a simple back in physical space (c) and after bias correction (d). Blue points mean worse RMSE while red ones mean better ones.

If the bias and the RMSE reveal general information, it is of interest to know the behavior of the analyses per precipitation class. The HSS and the FBI are used for that purpose and shown in Figure 10. It can be seen that the AROME-1.3 km HSS decreases with the precipitation classes, except for the first class ($RR_{24h} \geq 0.2 \text{ mm}$). As for the RMSE, every analyses increases the

background scores. Thus, the Box-Cox transformation increases the scores for light values of precipitation (up to 5 mm). For higher values, the scores are very close to the reference values. As a result, the Box-Cox transformation associated with the bias correction allows a better representation of light and moderate rain than the reference. Moreover, it must be noted that HSS

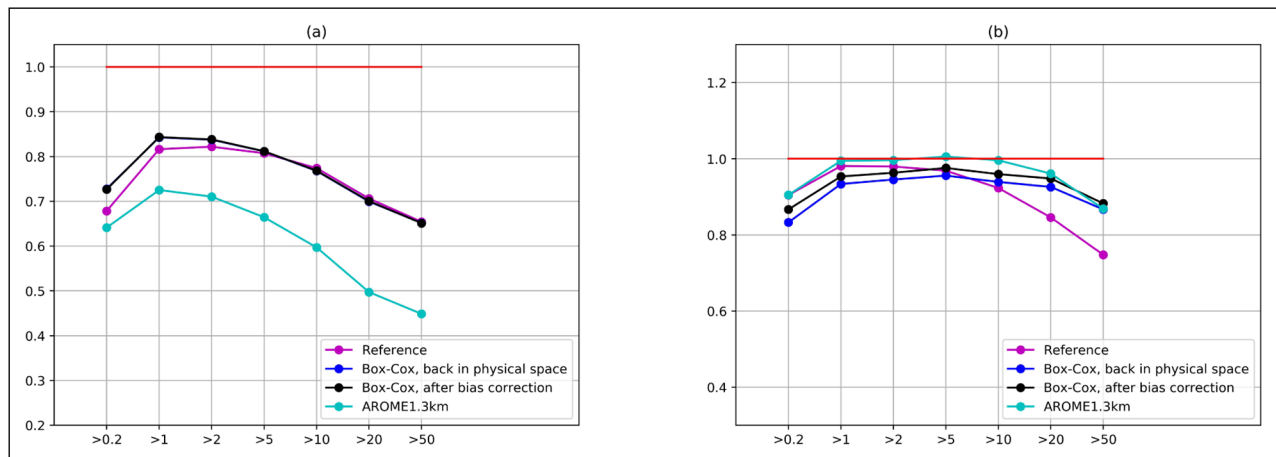


Figure 10 HSS (a) and FBI (b) for the different experiments.

values of the analysis transformed back into physical space are very close than HSS values of the analysis after bias correction. The FBI shows that the global underestimation of the analyses (see last paragraph) is a result of an underestimation for each precipitation class. AROME-1.3 km is slightly biased for each precipitation class, but again, this low bias is a result of a seasonal bias. It appears that for low and moderate precipitation classes (up to 5 mm), the reference is less biased than the Box-Cox experiment, but for high precipitation amounts, the reference significantly underestimates precipitation while the Box-Cox experiment is slightly more biased than the background. FBI scores show the efficiency of Eq. (21) to correct the dry bias induced by the Box-Cox transformation in the MESCAN analysis. Finally, even if the Box-Cox transformation produces better results in terms of HSS, it is slightly more biased than the reference for light and moderate rain. For higher precipitation amounts, the Box-Cox transformation is equivalent to the reference, but it is less biased. The improvement made by the Box-Cox transformation for light rain is coherent with results obtained by (Lespinas et al., 2015).

Another important aspect of the meteorological data used in hydrological studies is the number of dry days. In Table 2, the total number of dry days over the whole period and for every station is shown for each experiment. The definition of a dry day is done based on the precipitation measurement resolution: it includes the stations where $RR_{24h} < 0.1$ mm. The correlation is obtained by comparing the number of dry days for each evaluation station. The background AROME-1.3 km is close to the observations and the number of dry days is relatively well correlated with the observations. On the other hand, the reference underestimates the number of dry days, which is one of the common drawbacks of the analysis procedure. Analysis with the Box-Cox transformation increases the number of dry days and improves the correlation too. The bias correction removes some dry days, which is an expected result as the bias correction aims to add precipitation. Therefore,

EXPERIMENT		DRY DAYS	R
Observations		491 155	1.00
AROME-1.3 km		416 145	0.64
Reference		386 309	0.60
Box-Cox	Back in physical space	445 145	0.66
	Correction	373 206	0.66

Table 2 Number of dry days for the observations, the experiments and their correlation R .

it is important to note that even if the bias correction underestimates the number of dry days, the correlation with the observations is the same as for the experiment transformed back into physical space.

The spatial coverage of the difference between the number of dry days for the different experiments and between the number of dry days for the observations is represented in Figure 11. The reference underestimates the number of dry days over the entire country since there are only some stations in the Pyrenees, in the center of France and in the north where the reference overestimates the number of dry days. The background AROME-1.3 km also tends to underestimate the number of dry days over the majority of France, but there are more stations where it overestimates the number of dry days compared to the reference. In general, the magnitude of the overestimation or the underestimation of the number of dry days is lower for AROME-1.3 km than the reference. Even if the Box-Cox transformation generally underestimates the number of dry days over France, this underestimation is generally lower than the reference. Thus, there are more stations where the Box-Cox transformation overestimates the number of dry days. As a consequence of the goal of the bias correction, the analysis after bias correction underestimates the number of dry days over France with a greater amplitude. Moreover, in terms of the number of dry days, the analysis after bias correction and the reference are very close.

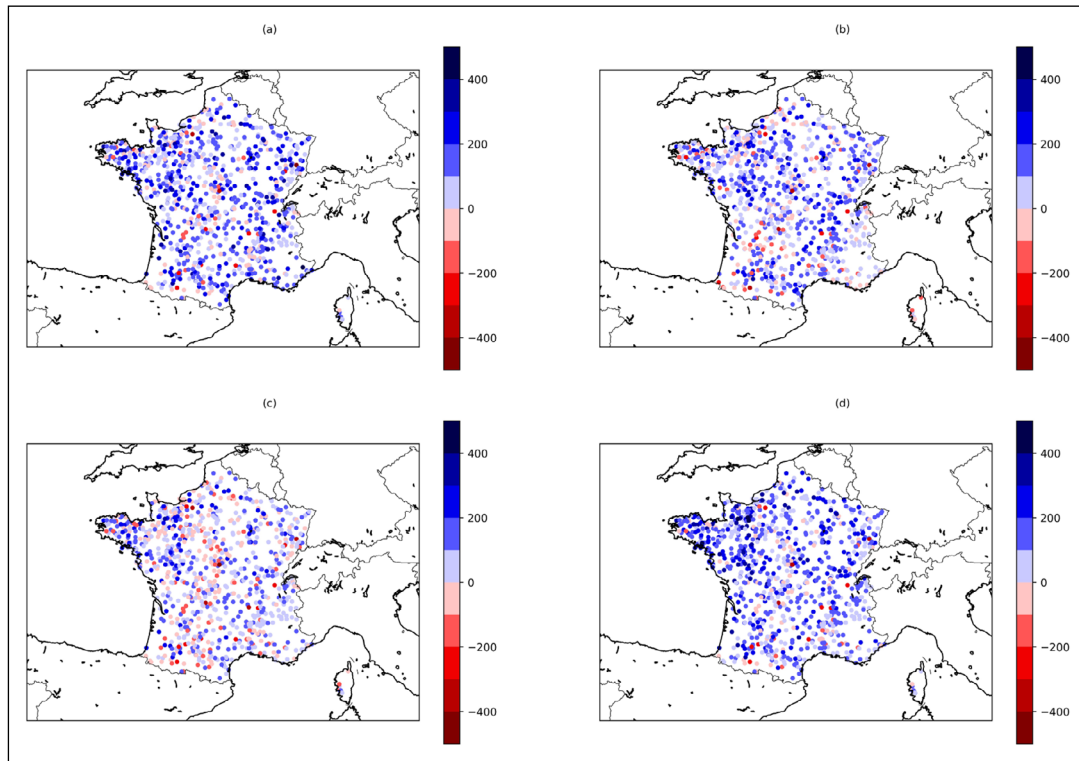


Figure 11 Difference of the number of dry days between observations and the analyses during the whole period for the reference (a), the background (b), the Box-Cox experiment with a simple back in physical space (c) and after bias correction (d). Blue (resp. red) points represent an underestimation (resp. overestimation) of the number of dry days.

EXPERIMENT		DJF	MAM	JJA	SON
AROME1-3 km		0.17	0.19	-0.15	-0.14
Reference		-0.14	-0.19	-0.21	-0.12
Box-Cox	Back in physical space	-0.13	-0.13	$-9.97 \cdot 10^{-2}$	$-7.15 \cdot 10^{-2}$
	Correction	-0.11	$-9.20 \cdot 10^{-2}$	$8.62 \cdot 10^{-2}$	$-2.44 \cdot 10^{-2}$

Table 3 Seasonal bias for the different experiments (mm).

3.3 SEASONAL RESULTS

Since the experiments were done over the period from 01/01/2016 to 31/12/2018, it is possible to compute seasonal scores by considering 3 winters (DJF), springs (MAM), summers (JJA) and autumns (SON). The bias and the RMSE for each season are shown in Table 3 and Table 4, respectively. It can be seen that the low bias of AROME-1.3 km over the whole period is the result of a wet bias in winter and spring and a dry bias in summer and autumn. The reference underestimates the precipitation at each season with the largest amplitude in summer and spring meaning that the analysis underestimates the convective precipitation. The Box-Cox transformation also introduces a bias compared to observations during winter and spring, but it is not biased during summer and autumn. This bias is corrected owing to Eq. (21) in spring but not during winter. Thus, each analysis improves the RMSE value of the background with higher values during the summer. It can be seen that the Box-Cox transformation improves the reference RMSE values with

EXPERIMENT		DJF	MAM	JJA	SON
AROME1-3 km		3.18	3.95	4.37	3.79
Reference		2.06	2.81	2.93	2.24
Box-Cox	Back in physical space	1.94	2.43	2.88	2.15
	Correction	1.94	2.43	2.91	2.16

Table 4 Seasonal RMSE for the different experiments (mm).

a greater amplitude during spring. Therefore, in summer, the Box-Cox transformation shows a lower improvement, and it can be seen that it is during this season that the correction is not well adapted since it increases the RMSE value. An explanation can come from the role of σ_b^{BC} in the bias correction. Indeed, this parameter is obtained thanks to the semi-variograms which represent an approximate distribution of precipitations.

The HSS for each season is plotted in Figure 12. For each season, every analysis has a better HSS scores than

the background. Thus, the Box-Cox transformation always has a better HSS for light precipitation (up to 5 mm) and similar HSS values for greater precipitation amounts. The improvement of the HSS from the analysis calculated in the Box-Cox space for light rain is higher during summer and spring. Since convective events often occur during summer, the HSS is lower this season for each

experiment, especially for high precipitation. FBI values for the different seasons are plotted in Figure 13. For every season, every analysis underestimates each class of precipitation. On the other hand, the bias correction works well, especially during the summer when the bias correction introduces considerable amounts of precipitation compared to the analysis transformed back

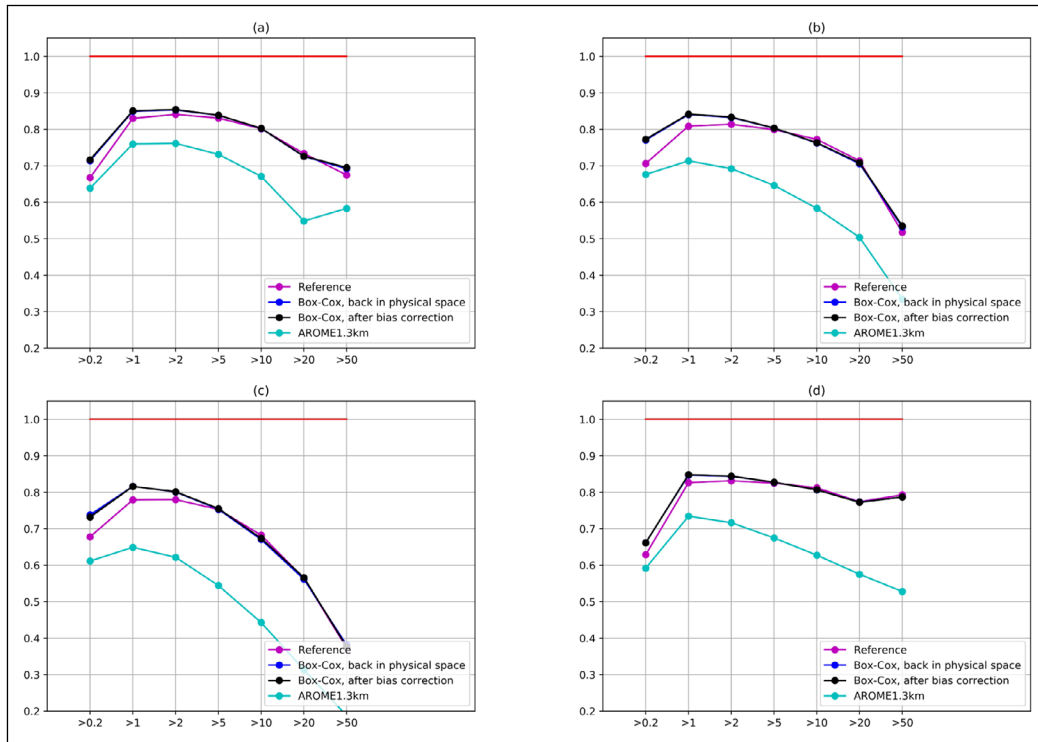


Figure 12 HSS for the different experiments during winter (a), spring (b), summer (c) and autumn (d).

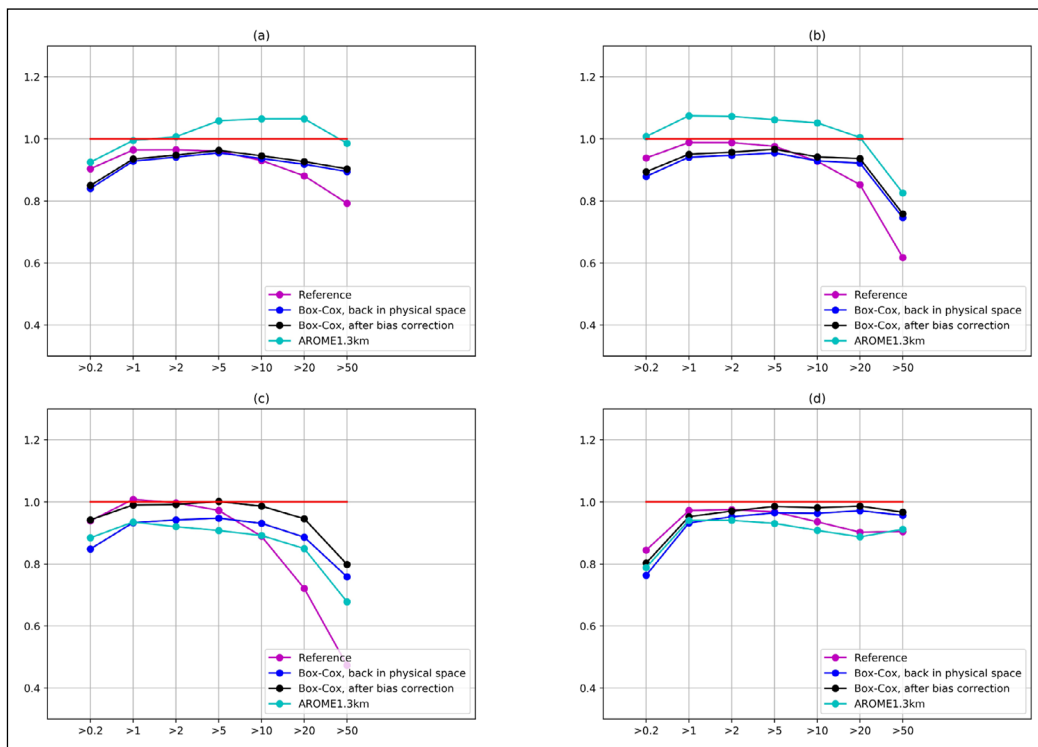


Figure 13 FBI for the different experiments during winter (a), spring (b), summer (c) and autumn (d).

in physical space. For every season, the analysis in the Box-Cox space is more biased than the reference for light rain (up to around 5 mm in general) and is less biased for heavy precipitation.

The total number of dry days for each season and for every station for each experiment are shown in Table 5. The reference underestimates the number of dry days for every season. On the other hand, AROME-1.3 km also underestimates the number of dry days every season, but this underestimation is lower during summer and autumn. The Box-Cox transformation improves the number of dry days, but the bias correction is worse than the reference in summer and autumn. Therefore, the correlation between the number of dry days at each station is better for the Box-Cox transformation than for the reference, and it is better than the background in winter and in autumn.

3.4 CASE STUDY

In this section, the case of a convective event which occurred on 16/09/2016 is studied. The background AROME-1.3 km simulation of the event is shown in Figure 14(a), where the rectangle plotted in black defines the area in which bias and RMSE of the different experiments were recalculated. In this area, 172 independent observations were used for the calculation of the scores. A zoom over the study area is shown in Figure 14(b), where the observations used for the evaluation are plotted as stars, and the observations used as input for the MESCAN analysis are represented by triangles. All the

available observations (observations used for evaluation and input observations) are also shown in Figure 14(c). The event is well represented by AROME-1.3 km but the location of the maximum precipitation of AROME-1.3 km is to the north of the observed maximum. This is caused by the fact that AROME-1.3 km does not simulate two convective precipitation bands but only one.

The impact of the precipitation analysis is shown in Figure 15, in which the difference between the reference analysis (cf section 3.1) and the background AROME-1.3 km is plotted (a). Red (*resp.* blue) on the map indicates that the reference produces more (*resp.* less) precipitation than the background. In Figure 15(b) is plotted the percent of added precipitation by the reference compared to the background. As for Figure 15(a), red (*resp.* blue) on the map indicates that the reference produces more (*resp.* less) precipitation than the background. On both figures, the observations used for the evaluation are plotted as stars, and the observations used as input for the MESCAN analysis are represented by triangles. The reference gives a more local representation of the convective event and cuts some areas where there is no precipitation compared to the background. As a consequence, there are two clear precipitation bands generated by the reference. On the other hand, the precipitation maximum is lower for the reference (145.85 mm) than for the background (155.09 mm). Comparison between the two figures show that the percent of added precipitation of the reference compared to the background is higher in regions where observed

Experiment		DJF		MAM		JJA		SON	
		Dry Days	R	Dry Days	R	Dry Days	R	Dry Days	R
Observations		92855	1.00	120889	1.00	156684	1.00	120727	1.00
AROME-1.3 km		66946	0.64	95485	0.61	143439	0.69	110275	0.55
Reference		69617	0.60	96545	0.54	127395	0.64	92752	0.51
Box-Cox	Back in phys.space	80973	0.70	109431	0.61	144184	0.69	110557	0.59
	Correction	73000	0.70	98434	0.60	113290	0.67	88482	0.59

Table 5 Number of seasonal dry days for the observations, the experiments and their correlation *R*.

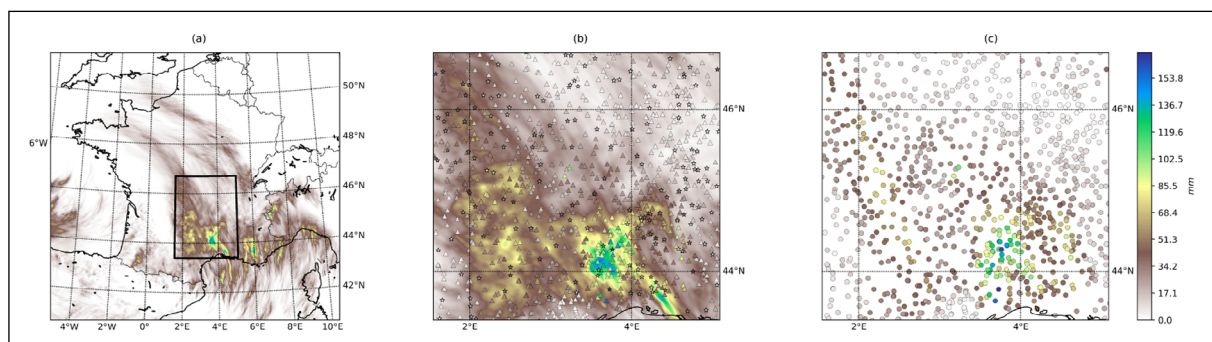


Figure 14 Convective event of the 16/09/2016 modelled by AROME-1.3 km over the whole domain (a) and over the study area (b). Stars corresponds to evaluation observations and triangles for observations used as input of the MESCAN analysis. All the available observations over the study area (c).

precipitation are, in general, relatively low. This result exhibits the positive impact of the precipitation analysis, as it produces more precipitation than the reference in areas where the background underestimates precipitation.

The percent of added precipitation of the reference compared to the transformed back into physical space analysis (*resp. the analysis after the bias correction*) is shown in Figure 16(a) (*resp. Figure 16(b)*). Red (*resp. blue*) on the map indicates that the reference produces more (*resp. less*) precipitation than the transformed back into physical space analysis or the analysis after the bias correction. The percent of added precipitation of the analysis after bias correction compared to the transformed back into physical space analysis is shown in Figure 16(c). Red (*resp. blue*) on the map indicates that the analysis after the bias correction produces more (*resp. less*) precipitation than the transformed back into physical space analysis. The reference, the transformed back into physical space analysis and the analysis after the bias correction are very close. Major differences are visible at the north eastern part and at the south western part of the domain. The Box-Cox transformation adds more precipitation than the reference in the high precipitation areas and less precipitation in the bands where there is less precipitation. By considering the difference between the analysis after bias correction

and the transformed back into physical space analysis, it can be seen that the bias correction does not add a lot of precipitation compared to the transformed back into physical space analysis except locally. Therefore, it can be noticed that the bias correction adds more precipitation, especially in the precipitation band in the western part of the domain, whereas adding precipitation at the location of the convection has little effect.

The area in which bias and RMSE are calculated and shown in Table 6 is plotted in Figure 14. Scores over the whole domain for the day are also calculated. Every analysis exhibits an important dry bias over the whole domain. The Box-Cox transformation decreases this bias over the whole domain compared to the reference. Over the study area, every experiment is more biased than

EXPERIMENT		BIAS (mm)	RMSE (mm)
AROME-1.3 km		-0.08 (-1.61)	13.71 (20.61)
Reference		-0.91 (-3.32)	7.49 (13.85)
Box-Cox	Back in physical space	-0.55 (-2.36)	7.56 (14.09)
	Correction (Eq. (21))	-0.41 (-2.18)	7.56 (14.05)

Table 6 Bias and RMSE values for the different experiments on 16/09/2016 over the whole domain and over the study area (between parenthesis).

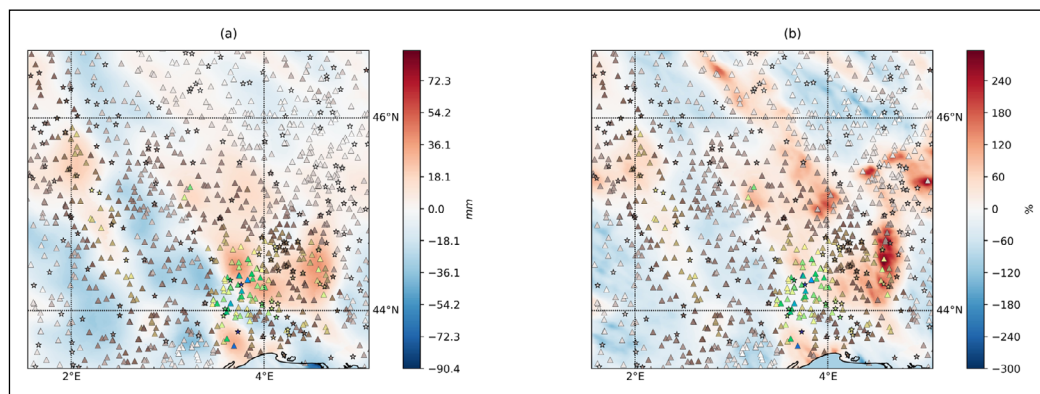


Figure 15 Difference of the reference precipitation analysis and the background AROME-1.3 km during the convective event of the 16/09/2016 over the study area (a) and percent of added precipitation of the reference compared to the background AROME-1.3 km over the study area (b).

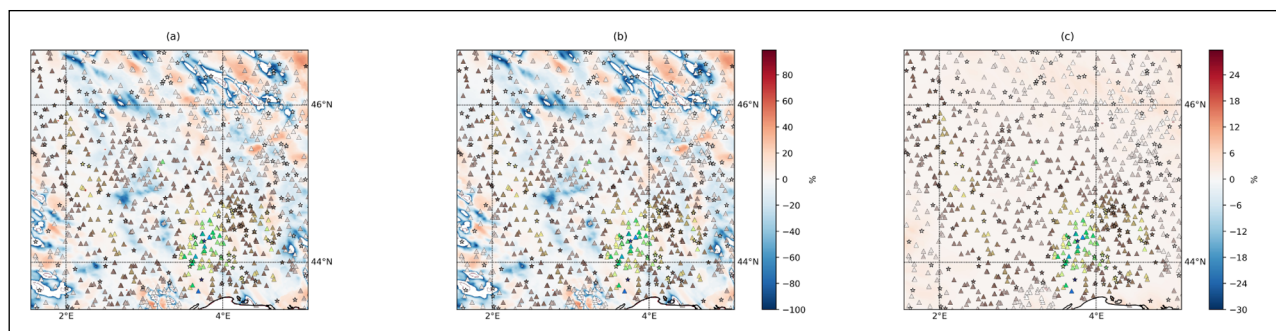


Figure 16 Percent of added precipitation of the transformed back into physical space analysis (a) and analysis after bias correction (b) compared to the reference and percent of added precipitation of the analysis after bias correction compared to the transformed back into physical space analysis (c).

over all of France. The analysis calculated in the Box-Cox space is less biased than the reference and it appears that the bias correction is effective in this case. Over all of France or over the study area, AROME-1.3 km is less biased than every analysis. In addition, the AROME-1.3 km RMSE is greater than every analysis RMSE values. Over all of France, the reference has a lower RMSE value and RMSE values are the same for transformed back into physical space analysis and for the analysis after bias correction. Over the study area, the RMSE value for the analysis after bias correction is lower than the RMSE value for the transformed back into physical space analysis. These results prove that the RMSE values of the Box-Cox transformation for convective events are very close to the reference but the significant bias exhibited by the reference is well-reduced by the transformation.

4 CONCLUSION AND DISCUSSION

In this article, a method for precipitation analysis included within a reanalysis project is presented. The non-linear Box-Cox transformation is used to calculate the analysis. Such a transformation allows us to work with Gaussian variables. A study about the optimal parameters of the Box-Cox transformation $\lambda = (\lambda_1, \lambda_2)$ highlights that $(\frac{1}{3}, 0)$ or $(\frac{1}{4}, 0)$ are the best tuned parameter values to use. By comparing scores of the two experiments using these two configurations, $\lambda = (\frac{1}{3}, 0)$ has been chosen to run the analysis. This study assesses the use of this configuration of the Box-Cox transformation, which is also used in the CaPA system. Background and observation standard deviation errors are daily determined thanks to the semi-variograms technique, which is found to be a real improvement compared to the other configurations where those parameters are fixed. The use of the Box-Cox transformation introduces a dry bias compared to observations, especially for high precipitation amounts, which is in agreement with the results obtained by (Carrera et al., 2010). A bias correction is applied in this study (Eq. (21)) which is derived from the equation purposed in (Evans, 2013). Precipitation added at the daily scale is, in general, low (section 3.4), but over the whole period, the bias correction has a significant role as the important dry bias over the entire period is well-reduced. Therefore, the use of this bias correction slightly increases the errors made by the analysis, especially in summer. An important result shown here is that such a configuration improves the HSS for light rain compared to the reference, and it has the same results for high precipitation but is less biased. A study of a convective event has shown the efficiency of the Box-Cox transformation over high-precipitation events as the very dry bias of the reference is well corrected by the Box-Cox transformation.

In this study, precipitation forecasts using AROME-1.3 km accumulations over 24h are used but as the quality of

the background is an important element of the analysis scheme, the use of accumulations from 3h forecasts could be used to define the background.

To assess the benefit of the method described in this study over a long time period, the use of the Box-Cox transformation should be tested with an observation data set similar to the one in the mid-19th century (which is less dense than the current one). It is assumed here that the background errors are isotropic, meaning that only the horizontal aspect is considered. Therefore, since precipitation is highly variable as a function of the orography, a vertical component can be added in Eq. (3). Even if such a configuration does not have any impact on the analysis in the plains, it has been shown to improve the analysis in the mountainous areas. Moreover, the use of a single semi-variogram over the whole area can be a source of error (Wang et al., 2019). An interesting perspective is to calculate local semi-variograms over a certain period (monthly for example). This can introduce a local information, especially in mountainous areas, where the stationary hypothesis is not verified by the precipitation. In this study, analyses are compared to independent observations which provide important information about the efficiency of the method. Therefore, such an evaluation does not give enough information over the mountainous areas, which are of great interest concerning water resource management in France. Hydrological model simulations should be run in order to have an indirect evaluation to assess the efficiency of the method over mountainous areas. It has been shown in this study that the optimal values of λ_1 seem to have a seasonal variation. A further study by using variable values of λ_1 would be potentially interesting.

APPENDICES

APPENDIX A – ACRONYMS

AROME: *Application de la Recherche à l'Opérationnel à Mésos-Echelle*

CaPA: *Canadian Precipitation Analysis*

CERRA: *Copernicus European Regional ReAnalysis*

DA: *Data Assimilation*

ECMWF: *European Center for Medium-range Weather Forecasts*

ERA: *European ReAnalysis*

EURO4M: *European Reanalysis and Observations for Monitoring*

FBI: *Frequency Bias Index*

HSS: *Heidke Skill Score*

JMA: *Japan Meteorological Agency*

JRA: *Japan ReAnalysis*

LBC: *Lateral Boundary Conditions*

NWP: *Numerical Weather Prediction*

OI: *Optimum Interpolation*

QQ-plots: *Quantile-Quantile-plots*

RMSE: *Root-Mean Square Error*

SAFRAN: *Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie*

UERRA: *Uncertainties in Ensemble Regional ReAnalyses*

UM: *Unified Model*

APPENDIX B – SCORES

Scores are defined using the following contingency table:

		OBSERVATIONS	
		Yes	No
Model	Yes	a	b
	No	c	d

From this table, the total number of events can be defined as $n = a+b+c+d$ and $R = \frac{(a+b)(a+c)+(b+d)(c+d)}{n}$.

The Heidke Skill Score (*HSS*) is calculated compared to a persistence (generally the precipitation the next day), and its values range between -1 and 1 , where negative values mean that the model is worse than persistence, positive values mean that the model is better than persistence, and 0 means that both persistence and model are equivalent. The optimal value of this score is 1 . The *HSS* is defined by

$$HSS = \frac{a+d-R}{n-R} \quad (25)$$

The Frequency Bias Index (*FBI*) shows the bias of the model for the different precipitation classes. Its values range between 0 and $+\infty$. A non-biased model gives a value of 1 and values under (resp. above) 1 occur when the model underestimates (resp. overestimates) the precipitation. The *FBI* is defined by:

$$FBI = \frac{a+b}{a+c} \quad (26)$$

Noting O as the observation, F the model and N the number of observations, the bias and the Root-Mean Square Error (*RMSE*) can be defined by, respectively,

$$b = \frac{1}{N} \sum_{i=1}^N (F_i - O_i) \quad (27)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2} \quad (28)$$

APPENDIX C – DEMONSTRATION OF THE BIAS CORRECTION

The Box-Cox transformation used here which has OI conditions (Gaussian distribution of the innovation), it can be written:

$$((x^a)^{(\lambda)} - (y^o)^{(\lambda)}) \sim \mathcal{N}(0, \sigma_{inn}^{BC}) \quad (29)$$

where σ_{inn}^{BC} stands for the variance of the innovation in the Box-Cox space, $(x^a)^{(\lambda)}$ and $(y^o)^{(\lambda)}$ stand for the analysis and

the observations in the transformed space, respectively. Then, the Taylor decomposition of the inverse Box-Cox transformation around $(x^a)^{(\lambda)}$ can be made:

$$(BC^{(\lambda)})^{-1}(y^{(\lambda)}) = (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] + \sum_{i=1}^{\infty} \left[(BC^{(\lambda)})^{-1} \right]^{(i)} \left[(x^a)^{(\lambda)} \right] \frac{[y^{(\lambda)} - (x^a)^{(\lambda)}]^i}{i!} \quad (30)$$

Then, the expectancy can be applied to the Eq. (30) to obtain

$$\begin{aligned} E \left[(BC^{(\lambda)})^{-1}(y^{(\lambda)}) \right] &= E \left[(BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \right] \\ &+ \sum_{i=1}^{\infty} E \left[\left[(BC^{(\lambda)})^{-1} \right]^{(i)} \left[(x^a)^{(\lambda)} \right] \frac{[y^{(\lambda)} - (x^a)^{(\lambda)}]^i}{i!} \right] \\ &= (BC^{(\lambda)})^{-1} \left[(x^a)^{(\lambda)} \right] \\ &+ \sum_{i=1}^{\infty} \frac{1}{i!} \left[(BC^{(\lambda)})^{-1} \right]^{(i)} \left[(x^a)^{(\lambda)} \right] E \left[[y^{(\lambda)} - (x^a)^{(\lambda)}]^i \right] \end{aligned} \quad (31)$$

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Stéphane Van Hyfte

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Patrick Le Moigne  orcid.org/0000-0001-7993-0056

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Eric Bazile  orcid.org/0000-0001-8120-9970

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Antoine Verrelle  orcid.org/0000-0003-2984-6483

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

Aaron Boone  orcid.org/0000-0003-1901-7024

CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

REFERENCES

- Amezcu, J** and **Van Leeuwen, PJ**. 2014. Gaussian anamorphosis in the analysis step of the EnKF: a joint state-variable/observation approach. *Tellus A: Dynamic Meteorology and Oceanography*, 66(1): 23493. DOI: <https://doi.org/10.3402/tellusa.v66.23493>
- Box, GE** and **Cox, DR**. 1964. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2): 211–243. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1964.tb00553.x>. DOI: <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Brousseau, P, Seity, Y, Ricard, D** and **Léger, J**. 2016. Improvement of the forecast of convective activity from the AROME-France system. *Quarterly Journal of the Royal*

- Meteorological Society*, 142(699): 2231–2243. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2822>. DOI: <https://doi.org/10.1002/qj.2822>
- Carrera, ML, Bélair, S, Fortin, V, Bilodeau, B, Charpentier, D and Doré, I.** 2010. Evaluation of snowpack simulations over the Canadian Rockies with an experimental hydrometeorological modeling system. *Journal of Hydrometeorology*, 11(5): 1123–1140. DOI: <https://doi.org/10.1175/2010JHM1274.1>
- Cecinati, F, Wani, O and Rico-Ramirez, MA.** 2017. Comparing approaches to deal with non-Gaussianity of rainfall data in kriging-based radar-gauge rainfall merging. *Water Resources Research*, 53(11): 8999–9018. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR020330>. DOI: <https://doi.org/10.1002/2016WR020330>
- Dahlgren, P, Landelius, T, Källberg, P and Gollvik, S.** 2016. A high-resolution regional reanalysis for Europe. Part 1: Three-dimensional reanalysis with the regional High-Resolution Limited-Area Model (HIRLAM). *Quarterly Journal of the Royal Meteorological Society*, 142(698): 2119–2131. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2807>. DOI: <https://doi.org/10.1002/qj.2807>
- Daley, R.** 1993. Atmospheric data analysis. Number 2. Cambridge university press.
- Dee, DP and Da Silva, AM.** 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Monthly Weather Review*, 127(8): 1822–1834. URL https://journals.ametsoc.org/view/journals/mwre/127/8/1520-0493_1999_127_1822_mleofa_2.0.co_2.xml. DOI: [https://doi.org/10.1175/1520-0493\(1999\)127<1822:MLEOFA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<1822:MLEOFA>2.0.CO;2)
- Dee, DP, Uppala, SM, Simmons, AJ, Berrisford, P, Poli, P, Kobayashi, S, Andrae, U, Balmaseda, MA, Balsamo, G, Bauer, P, Bechtold, P, Beljaars, ACM, van de Berg, L, Bidlot, J, Bormann, N, Delsol, C, Dragani, R, Fuentes, M, Geer, AJ, Haimberger, L, Healy, SB, Hersbach, H, Hólm, EV, Isaksen, I, Källberg, P, Köhler, M, Matricardi, M, McNally, AP, Monge-Sanz, BM, Morcrette, J-J, Park, B-K, Peubey, C, de Rosnay, P, Tavolato, C, Thépaut, J-N and Vitart, F.** 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656): 553–597. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828>. DOI: <https://doi.org/10.1002/qj.828>
- Desroziers, G, Berre, L, Chapnik, B and Poli, P.** 2005. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613): 3385–3396. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.108>. DOI: <https://doi.org/10.1256/qj.05.108>
- Devers, A, Vidal, J-P, Lauvernet, C, Graff, B and Vannier, O.** 2020. A framework for high-resolution meteorological surface reanalysis through offline data assimilation in an ensemble of downscaled reconstructions. *Quarterly Journal of the Royal Meteorological Society*, 146(726): 153–173. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3663>. DOI: <https://doi.org/10.1002/qj.3663>
- Donat, MG, Sillmann, J, Wild, S, Alexander, LV, Lippmann, T and Zwiers, FW.** 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *Journal of Climate*, 27(13): 5019–5035. URL <https://journals.ametsoc.org/view/journals/clim/27/13/jcli-d-13-00405.1.xml>. DOI: <https://doi.org/10.1175/JCLI-D-13-00405.1>
- Durand, Y, Brun, E, Merindol, L, Guyomarc'h, G, Lesaffre, B and Martin, E.** 1993. A meteorological estimation of relevant parameters for snow models. *Annals of Glaciology*, 18: 65–71. DOI: <https://doi.org/10.3189/S0260305500011277>
- Erdin, R, Frei, C and Künsch, HR.** 2012. Data transformation and uncertainty in geostatistical combination of radar and rain gauges. *Journal of Hydrometeorology*, 13(4): 1332–1346. URL https://journals.ametsoc.org/view/journals/hydr/13/4/jhm-d-11-096_1.xml. DOI: <https://doi.org/10.1175/JHM-D-11-096.1>
- Evans, A.** 2013. *Investigation of enhancements to two fundamental components of the statistical interpolation method used by the Canadian Precipitation Analysis (CaPA)*. PhD thesis, University of Manitoba. URL <https://mspace.lib.umanitoba.ca/xmlui/handle/1993/22276>
- Fletcher, SJ and Zupanski, M.** 2006. A data assimilation method for log-normally distributed observational errors. *Quarterly Journal of the Royal Meteorological Society*, 132(621): 2505–2519. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1256/qj.05.222>. DOI: <https://doi.org/10.1256/qj.05.222>
- Fortin, V, Roy, G, Donaldson, N and Mahidjiba, A.** 2015. Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology*, 531: 296–307. URL <https://www.sciencedirect.com/science/article/pii/S0022169415005624>. DOI: <https://doi.org/10.1016/j.jhydrol.2015.08.003>
- Hersbach, H, Bell, B, Berrisford, P, Hirahara, S, Horányi, A, Muñoz-Sabater, J, Nicolas, J, Peubey, C, Radu, R, Schepers, D, Simmons, A, Soci, C, Abdalla, S, Abellan, X, Balsamo, G, Bechtold, P, Biavati, G, Bidlot, J, Bonavita, M, De Chiara, G, Dahlgren, P, Dee, D, Diamantakis, M, Dragani, R, Flemming, J, Forbes, R, Fuentes, M, Geer, A, Haimberger, L, Healy, S, Hogan, RJ, Hólm, E, Janisková, M, Keeley, S, Laloyaux, P, Lopez, P, Lupu, C, Radnoti, G, de Rosnay, P, Rozum, I, Vamborg, F, Villaume, S and Thépaut, J-N.** 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730): 1999–2049. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803>. DOI: <https://doi.org/10.1002/qj.3803>
- Ishak, NAM and Ahmad, S.** 2018. Box-Cox optimal parameter estimation for multiple regressions with homoscedasticity. In *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2016)*, pages 1047–1054. Springer Singapore. DOI: https://doi.org/10.1007/978-981-13-0074-5_103

- Jermeý, P, Davie, J, Mahmood, S and Renshaw, R.** 2015. Development of ensemble variational data capability and report demonstrating ensemble uncertainty products. *UERRA deliverable D2.1*.
- Kaiser-Weiss, AK, Borsche, M, Niermann, D, Kaspar, F, Lussana, C, Isotta, FA, van den Besselaar, E, van der Schrier, G and Undén, P.** 2019. Added value of regional reanalyses for climatological applications. *Environmental Research Communications*, 1(7): 071004. DOI: <https://doi.org/10.1088/2515-7620/ab2ec3>
- Kobayashi, S, Ota, Y, Harada, Y, Ebata, A, Moriya, M, Onoda, H, Onogi, K, Kamahori, H, Kobayashi, C, Endo, H, et al.** 2015. The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, 93(1): 5–48. DOI: <https://doi.org/10.2151/jmsj.2015-001>
- Landelius, T, Dahlgren, P, Gollvik, S, Jansson, A and Olsson, E.** 2016. A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Quarterly Journal of the Royal Meteorological Society*, 142(698): 2132–2142. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2813>. DOI: <https://doi.org/10.1002/qj.2813>
- L'Ecuyer, TS and Stephens, GL.** 2002. An estimation-based precipitation retrieval algorithm for attenuating radars. *Journal of Applied Meteorology and Climatology*, 41(3): 272–285. URL https://journals.ametsoc.org/view/journals/apme/41/3/1520-0450_2002_041_0272_aebpra_2.0.co_2.xml. DOI: [https://doi.org/10.1175/1520-0450\(2002\)041<0272:AEBPRA>2.0.CO;2](https://doi.org/10.1175/1520-0450(2002)041<0272:AEBPRA>2.0.CO;2)
- Legates, DR and Willmott, CJ.** 1990. Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, 10(2): 111–127. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3370100202>. DOI: <https://doi.org/10.1002/joc.3370100202>
- Lespinas, F, Fortin, V, Roy, G, Rasmussen, P and Stadnyk, T.** 2015. Performance Evaluation of the Canadian Precipitation Analysis (CaPA). *Journal of Hydrometeorology*, 16(5): 2045–2064. DOI: <https://doi.org/10.1175/JHM-D-14-0191.1>
- Lien, G-Y, Kalnay, E and Miyoshi, T.** 2013. Effective assimilation of global precipitation: Simulation experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1): 19915. DOI: <https://doi.org/10.3402/tellusa.v65i0.19915>
- Lussana, C, Nipen, TN, Seierstad, IA and Elo, CA.** 2020. Ensemble-based statistical interpolation with Gaussian anamorphosis for the spatial analysis of precipitation. *Nonlinear Processes in Geophysics Discussions*, 2020(1): 1–43. URL <https://npg.copernicus.org/articles/28/61/2021/>. DOI: <https://doi.org/10.5194/npg-28-61-2021>
- Mahfouf, J, Brasnett, B and Gagnon, S.** 2007. A Canadian Precipitation Analysis (CaPA) project: Description and preliminary results. *Atmosphere-Ocean*, 45(1): 1–17. DOI: <https://doi.org/10.3137/ao.v450101>
- Onogi, K, Tsutsui, J, Koide, H, Sakamoto, M, Kobayashi, S, Hatsushika, H, Matsumoto, T, Yamazaki, N, Kamahori, H, Takahashi, K, et al.** 2007. The JRA-25 reanalysis. *Journal of the Meteorological Society of Japan. Ser. II*, 85(3): 369–432. DOI: <https://doi.org/10.2151/jmsj.85.369>
- Ridal, M, Olsson, E, Unden, P, Zimmermann, K and Ohlsson, A.** 2017. HARMONIE reanalysis report of results and dataset. *UERRA deliverable D2.7*.
- Rodríguez-Puebla, C, Encinas, AH and Sáenz, J.** 2001. Winter precipitation over the Iberian peninsula and its relationship to circulation indices. *Hydrology and Earth System Sciences Discussions*, 5(2): 233–244. URL <https://hal.archives-ouvertes.fr/hal-00304599>. DOI: <https://doi.org/10.5194/hess-5-233-2001>
- Seity, Y, Brousseau, P, Malardel, S, Hello, G, Bénard, P, Bouttier, F, Lac, C and Masson, V.** 2011. The AROME-France convective-scale operational model. *Monthly Weather Review*, 139(3): 976–991. URL <https://journals.ametsoc.org/view/journals/mwre/139/3/2010mwr3425.1.xml>. DOI: <https://doi.org/10.1175/2010MWR3425.1>
- Sevruk, B.** 1996. Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, 42(1–4): 237–246. URL <https://www.sciencedirect.com/science/article/pii/0169809595000666>. DOI: [https://doi.org/10.1016/0169-8095\(95\)00066-6](https://doi.org/10.1016/0169-8095(95)00066-6)
- Soci, C, Bazile, E, Besson, F and Landelius, T.** 2016. High-resolution precipitation re-analysis system for climatological purposes. *Tellus A: Dynamic Meteorology and Oceanography*, 68(1): 29879. DOI: <https://doi.org/10.3402/tellusa.v68.29879>
- Tippett, MK, Barnston, AG and Robertson, AW.** 2007. Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles. *Journal of climate*, 20(10): 2210–2228. URL <https://journals.ametsoc.org/view/journals/clim/20/10/jcli4108.1.xml>. DOI: <https://doi.org/10.1175/JCLI4108.1>
- Touma, D, Michalak, AM, Swain, DL and Diffenbaugh, NS.** 2018. Characterizing the spatial scales of extreme daily precipitation in the United States. *Journal of Climate*, 31(19): 8023–8037. URL: <https://journals.ametsoc.org/view/journals/clim/31/19/jcli-d-18-0019.1.xml>. DOI: <https://doi.org/10.1175/JCLI-D-18-0019.1>
- Vidal, J-P, Martin, E, Franchistéguy, L, Baillon, M and Soubeyroux, J-M.** 2010. A 50-year high-resolution atmospheric reanalysis over France with the SAFRAN system. *International Journal of Climatology*, 30(11): 1627–1644. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.2003>. DOI: <https://doi.org/10.1002/joc.2003>
- Wang, Y, Chen, J and Yang, D.** 2019. Bayesian assimilation of multiscale precipitation data and sparse ground gauge observations in mountainous areas. *Journal of Hydrometeorology*, 20(8): 1473–1494. URL https://journals.ametsoc.org/view/journals/hydr/20/8/jhm-d-18-0218_1.xml. DOI: <https://doi.org/10.1175/JHM-D-18-0218.1>

TO CITE THIS ARTICLE:

Van Hyfte, S, Le Moigne, P, Bazile, E, Verrelle, A and Boone, A. 2023. High-Resolution Reanalysis of Daily Precipitation using AROME Model Over France. *Tellus A: Dynamic Meteorology and Oceanography*, 75(1): 27–49. DOI: <https://doi.org/10.16993/tellusa.95>

Submitted: 05 May 2022 **Accepted:** 24 August 2022 **Published:** 05 January 2023

COPYRIGHT:

© 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Tellus A: Dynamic Meteorology and Oceanography is a peer-reviewed open access journal published by Stockholm University Press.

