SHORT COMMUNICATION

The SAFRAN daily gridded precipitation product in Tunisia (1979–2015)

Yves Tramblay1 | Haifa Feki2,3 | Pere Quintana-Seguí4 | José A. Guijarro5

1HydroSciences Montpellier (Univ. Montpellier, CNRS, IRD), Montpellier, France
2Ecole supérieure des ingénieurs de Medjez el beb, Tunisia
3Laboratoire GREEN TEAM, Université de Carthage, Tunisia
4Ebro Observatory, Ramon Llull University – CSIC, Roquetes, Spain
5Agencia Estatal de Meteorología (AEMET), Balearic Islands Office, Spain

Correspondence
Email: yves.tramblay@ird.fr

Abstract
A new high-resolution (5 km) gridded daily precipitation dataset for Tunisia between 1979 and 2015 is introduced. This product combines 960 rain gauges with the SAFRAN analysis to produce the precipitation gridded data. A validation approach on two different datasets reveals that the SAFRAN analysis outperforms other standard interpolation methods such as Inverse Distance, Nearest Neighbors, Ordinary Kriging or Residual Kriging with altitude. When compared to EOBS, a widely used gridded dataset over Europe, a strong negative bias in EOBS precipitation is found. However due to the aridity and the low density of rain gauges in south Tunisia, results in this region must be analyzed with care. The SAFRAN product could be useful for various purposes such as climate model evaluation, climate studies, hydrological modelling to support the planning and management of surface water resources in Tunisia.

KEYWORDS
Interpolation, Precipitation, SAFRAN, Tunisia

1 | INTRODUCTION

Middle East and North Africa (MENA) region is subject to water scarcity (Ragab and Prudhomme, 2002) and more than 60% of the population lives in areas of high water stress compared to a global average of about 35%. Droogers et al. (2012) mentioned that, in present time, the average per capita water availability in MENA region is slightly above the physical water scarcity limit at about 1,076 m³/yr compared to the world average of about 8,500 m³/yr. In particular, water resources in Tunisia are identified by their scarcity, low quality, poor distribution and seasonal distribution (Ben Zaied and Binet, 2015).

The 4th assessment report of the IPCC (IPCC, 2012) projects strong changes in climate across the MENA region. Climate change is expected to increase water stress through various mechanisms including reduced precipitation, intensifying rainfall variability and rising temperature. However, the problem of water scarcity in Tunisia, according to Haddadin (2001), is not only solely based on the availability of the resource but also driven by human pressure that is increasing water stress. Several studies (Ragab and Prudhomme, 2002, Schilling et al., 2012, Tramblay et al., 2018), mentioned that by 2050, North Africa is expected to have reduced rainfall amounts of 20 to 25% less than the present mean value. A recent study et Zittis (2017) using various existing gridded datasets, showed that the long term trends in the Middle-East and North Africa (MENA) region is indicating an overall drying since the beginning of the twentieth century mainly, over the Maghreb region. They also noted that the different data sources have statistically significant differences in the distribution of monthly precipitation for about 50% of the domain.

Precipitation studies are mostly carried out based on gridded precipitation data such as the EOBS (Haylock, 2008), CRU (Harris et al., 2014), CPC (Chen et al., 2008) or GPCC (Schamm et al., 2014) datasets. This type of data is necessary for local climate studies, climate change monitoring at regional scale, validation of regional climate models (RCM) and impact
models (Haylock, 2008). Several gridded precipitation products have been developed for countries such as Spain in Spain (Herrera et al., 2012), SAFRAN in France (Quintana-Seguí et al., 2008, Vidal et al., 2010) and Spain (Quintana-Seguí et al., 2016). For the whole Euro-Mediterranean region, the EOBS dataset is probably the most employed and provides daily high-resolution (25 km to 10 km) gridded precipitation. Yet, these gridded datasets are widely used for climate studies but in regions with data scarcity they can introduce a significant uncertainty (Romera et al., 2015, Prein and Gobiet 2017, Zittis, 2017, Kotlarski et al., 2017). As noted by several authors, the Euro-Mediterranean domain is covered by an uneven station density, and the use of this dataset could be problematic in particular when looking at extreme precipitation (Flaounas et al., 2012, Herrera et al., 2012, Turco et al., 2013, Fantini et al., 2016). As for Tunisia, and also other Maghreb countries, the EOBS dataset contains only a very small number of stations (Haylock et al., 2008). This is also the case for other large-scale observational databases (CRU, CPC, GPCC etc.) in which the North African countries are largely under-represented. Beside gridded datasets based on interpolated precipitation, a growing number of satellite-based precipitation products are also becoming available with almost a global coverage and relying on different sensors (Kummerow et al., 1998; Mehta and Yang, 2008, Dhiba et al., 2017). Several of these products have been successfully validated in the Mediterranean context (Nastos et al., 2013, Tramblay et al., 2016) and also merged with gauge and reanalysis datasets such as in the MSWEP product (Beck et al., 2017).

In Tunisia, rainfall is highly variable both temporally and geographically, while surface water is a very important resource for agricultural activities and consumption. In this context, there is a need for country-scale information systems to analyze and mitigate climate change impacts but also to develop regional or basin-scale surface modeling to improve resources management. Country-scale reliable precipitation data is currently lacking to better estimate the spatial variability of precipitation extremes (Dhiba et al., 2017) or to validate the most recent climate models (Bargaoui et al., 2014, Fathalli et al. 2018). The goal of the present study is to develop a gridded data set of precipitation in Tunisia with the whole rain gauge monitoring network of the country. Different interpolation methods are first compared and the SAFRAN analysis is implemented over the whole Tunisian territory. Then, the high-resolution gridded dataset produced is compared with a state-of-art daily precipitation product; the EOBS database (Haylock et al., 2008).

2 | DATASETS

The full rain gauges database of the Direction Générale des Resources en Eau (DGRE) of Tunisia containing over 2000 stations has been processed. The spatial coverage has a higher density in the North of Tunisia, where are located most the dams and reservoirs of the country. A global quality check has been performed; since several stations were lacking information’s about their locations or had long periods of missing data. Only the stations with at least 5 complete years between the years 1979 to 2015 have been considered for subsequent analysis, corresponding to 960 stations. The number of stations with available data each year between 1979 and 2015 is shown on Figure 1 together with the altitude of the stations. It can be seen that the number of available stations is gradually increasing until 2007 then is dropping to about 100 stations between 2007 and 2015. For the stations with no metadata, we used the historical publications of the DGRE, the Annuaire Hydrologiques de Tunisie, (see one example here for the year 2007/2008: http://www.hydrosciences.fr/sierem/produits/biblio/annales/TN/2007-2008.pdf) available for several years and containing the data, maps and station information.

In addition to station data, we used the EOBS database (Haylock et al., 2008) to obtain different precipitation indices to be compared with the SAFRAN product. These indices are computed on an annual basis and include the highest 1-day precipitation amount (RX1day), the number of wet days (R1mm) and the total precipitation due to wet days (PRCPTOT).

3 | METHODS

3.1 | Quality check and homogeneity tests

The precipitation dataset was checked for quality and homogeneity by means of the R package Climatol v.3.1 (Guijarro, 2017, 2018). As this software works better with whole years and the dataset comprised data until August 2015, the homogeneity test is applied to the period 1979–2014. Due to the high variability of daily rainfall, especially in arid climates, it is not possible to check the homogeneity of the series at the daily time step. Therefore, the homogenization was performed on monthly totals calculated from the daily data for stations having at least 20 years of data. The procedure implemented in Climatol consists in applying the Standard Normal Homogeneity Test (SNHT, Alexandersson, 1986) to the series of anomalies, that is, differences between observed data and their estimates from nearby stations, both in normalized form by dividing each data by their corresponding average. This procedure led to the detection of 35 break-points (changes in the mean of the anomaly series) with the SNHT test in 30 series. For these stations, the dates of the break-points were then used to split the daily series into separate homogeneous sub-periods and only the longest ones have been kept.
Outliers were also checked in the anomaly series, but beside a few obvious errors (i.e., negative precipitation or daily precipitation above 1,000 mm/day, for example) no daily data were rejected because the few outliers found were considered feasible in the frame of an arid climate with frequent isolated precipitations of convective origin.

3.2 Interpolation of rain gauge precipitation

The interpolation of the rain gauge data is performed in the present study with different methods. Deterministic methods such as Inverse distance weighting (ID) and the nearest neighbor (NN) are considered. In addition, the geostatistical method of Ordinary Kriging (OK) is also considered (Goovaerts 2000, Feki et al., 2017a). The variograms required for the OK method are fitted automatically with a spherical variogram model for each time step when rain is measured at least in 3 stations, otherwise ID interpolation is performed (Tramblay et al., 2016). The spherical variogram model is convenient for precipitation (Feki et al., 2012), which is not a spatially continuous field like temperature, since it provides a value of the de-correlation distance (Lebel et al., 1987).

In order to take the influence of elevation into account in the interpolation (Feki et al., 2012), the residual kriging (RK) method (Hengl et al., 2007) is implemented in addition to OK. It is mathematically equivalent to universal kriging or kriging with external drift methods, but RK allows the separate interpretation of the two interpolated components, the regression and the residuals, and use of a broader range of regression techniques (Feki et al., 2012, Feki and Slimani, 2015). The RK approach combines a regression model between precipitation and altitude with the spatial interpolation of the residuals of the regression. At each time step, for each location \( i \) the estimate of precipitation \( z \) can be computed as:

\[
z(i) = m(i) + e(i)
\]

Where \( m(i) \) is the regression model estimate and \( e(i) \) the spatially interpolated residual of the regression.

An ordinary least square (OLS) regression cannot be used here since the constraint of heteroscedasticity of the residuals would not be fulfilled: the elevation for neighboring stations is very likely spatially correlated. Therefore generalized least squares (GLS) should be used instead of OLS models (Hengl et al., 2007), since they do not rely on such a constraint. Here a GLS model is considered for the relationship between rain gauge and elevation data (Tramblay et al., 2016). For each day, if the correlation between the precipitation measured at the rain gauges and altitude is significant at the 10% confidence level, a GLS model is built and the residuals of the model are estimated by simple kriging with known mean (0). The variogram for the residuals is estimated for each time step, in a similar way as for the OK interpolation method explained above. By comparison, the EOBS dataset rely on a three step procedure: first the monthly mean values are interpolated with thin plate splines to define the underlying spatial trend, second the monthly anomalies are interpolated via kriging and third the anomalies are applied to monthly mean values to obtain the final result (Haylock et al., 2008).

3.3 The SAFRAN analysis

SAFRAN (“Système d’Analyse Fournissant des Renseignements Adaptés à la Nivologie”) is a high-resolution atmospheric analysis system developed at Météo France (Durand et al., 1993), based on an optimal-interpolation algorithm. It had been initially designed to provide atmospheric forcing data in mountainous areas for avalanche hazard forecasting (Durand et al., 1999) and it was then extended to France at the 8 km spatial resolution (Quintana-Seguí et al.; 2008, Vidal et al. 2010) for hydro-meteorological applications (Habets et al., 2008). Later, SAFRAN was applied in Spain at the 5 km spatial resolution (Quintana-Seguí et al., 2016), where it was shown that its precipitation fields are comparable to those of Spain02 (Herrera et al., 2012, Quintana-Seguí et al., 2017), which is based on Ordinary Kriging and thin plate splines interpolation methods that takes the orography into account (Herrera et al., 2012).
SAFRAN performs its analysis in two steps. First, the analysis is performed on climatological homogeneous zones. These zones, which in this case have a mean area of ~700 km², should have weak horizontal gradients of the studied variables. Catchments of Tunisia were obtained from the HydroSHEDS data (Lehner and Grill, 2013) based on a grid resolution of 15 arc-seconds (approximately 500 m at the equator). Watersheds in this database were delineated in a consistent manner at different scales, following the topological concept of the Pfafstetter coding system (de Jager and Vogt, 2010). Here the layer corresponding to the level 8 have been retrieved, to obtain the zones for the SAFRAN analysis. For some large catchments, the SAFRAN zones were manually drawn using the relief and our knowledge of the local climate as guiding information. The SAFRAN zones for Tunisia are shown in Figure 2. SAFRAN calculates one value of the studied variable, precipitation in this case, for each vertical level of 300 m. of the zone, using the available meteorological stations, mainly within the zone (but it can, if necessary, use information of neighboring stations). In a second step, the values are interpolated to a regular 5 km grid, considering the vertical gradients. This resolution has been chosen since it is appropriate for climatological or hydrological analysis and also to be comparable to the other SAFRAN products in France and Spain. As a result, within each zone, the values of precipitation for each grid point are different only if the altitudes of the grid points are different.

3.4 Validation framework

To validate the different interpolation methods (OK, RK, NN and ID) and the SAFRAN analysis, two different validation samples are randomly generated with the following constraints:

1. A minimum of 50% of complete years during the period 1979–2015
2. A minimum distance between two stations of two times the average distance between stations
Two samples of 103 stations, corresponding to approximately 10% of the 960 stations available, have been generated with these criterions. There are named thereafter as “validation sample a” and “validation sample b”. The different interpolation methods are then applied without these validation samples. The, for each station of the validation samples, the relative bias and the Spearman correlation coefficient between daily rainfall amounts, daily rainfall occurrence from the interpolated data and the observed data are computed.

**FIGURE 4** Box-plots of the relative bias, correlation of daily amounts and correlation of daily occurrence for the two validation samples (a and b). ID: Inverse-distance, OK: Ordinary Kriging, RK: Residual kriging, NN: Nearest neighbors, SA: SAFRAN [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE 5** Maps of the relative bias between observed and interpolated precipitation for the two validation samples a and b [Colour figure can be viewed at wileyonlinelibrary.com]
4 | RESULTS

4.1 | Comparison of interpolation methods on the two validation samples

The quality check and the homogeneity test performed on the whole station database led to select 960 stations with at least 5 years of data between 1979 and 2015, shown in Figure 3. These 960 stations have been used in the different interpolation methods and to build the SAFRAN analysis. From Figure 4, it can be seen that the efficiency of the different methods is not dependent on the validation sample, with similar results obtained with the two samples. On average, the OK method seems to perform slightly better, in terms of relative bias and correlation than the other spatial interpolations methods. However, the different interpolation methods provide very similar estimates, due to the high spatial density of observed precipitation notably for the northern part of Tunisia. It is worth to observe also that the use of elevation as a covariate in the RK approach does not improve the efficiency of this method by comparisons with the other interpolation schemes. On average, the SAFRAN products provide the lowest bias and the highest correlations on daily amounts and occurrence. When considering the spatial distribution of the results, it can be seen than SAFRAN and the NN method provide the most consistent estimations across Tunisia,
while the other methods have a very strong bias in the southern stations (Figure 5). For correlation (Figure 6), for some stations located in the North there is a high correlation close to one, but for the other regions the correlation patterns seem less organized than the bias observed in Figure 5. Overall, there is a clear North/South distinct behavior, with degraded performances in southern stations, due to the lower density of stations but also more arid conditions than in the North. Similar results have been found in Feki et al., (2017a). This is exemplified in Figure 7; south of 35°N the relative bias in validation for both samples is very high and for most stations is exceeding 100%. Therefore the use of spatially interpolated data in these regions is not recommended.

4.2 Comparison between SAFRAN and EOBS

The comparison has been performed between the Rx1day, R1mm and PRCPTOT indices computed from EOBS and SAFRAN annually, the inter-annual means of the indices between 1979 and 2015 are compared by computing a relative difference (Figure 8). It must be noted that the EOBS dataset contains only 13 stations for Tunisia (Tabarka, Bizerte, Tunis, Kelibia, Jenbdouba, Kairouan, Monastir, Gafsa, Sfax, Gabes, Djerba, Remada) as shown in Figure 3. This implies very smooth interpolation surfaces, not taking into account the regional differences due to orography or local climate characteristics. This is particularly true for PRCPTOT and R1mm indices, with the spatial gradient from Northwest to Southwest is clearly underestimated. On average over the whole country, the relative bias of EOBS compared to SAFRAN is −47% for precipitation totals (PRPTOT) and −59% for the number of wet days (R1mm). For annual maximum precipitation, there is a more complex picture. First, the areas with the higher precipitation intensities do no show a clear spatial organization in both datasets indicating the strong spatial variability of extreme precipitation. This result is confirmed by Feki et al. (2017b) concerning North Tunisia with station-based data. It is also worth noting that the “patchy” look of the SAFRAN map for Rx1day is caused by the interpolation process over homogenous spatial unit (see Figure 2) and the same behavior has been reported for SAFRAN over France or Spain. Secondly, the areas with high values of Rx1day in the South must be interpreted with care since the density of stations is low in these areas and these patterns might just be caused by the interpolation scheme and do
not necessarily represents reality. The relative bias of EOBS for Rx1day is $-17\%$ compared to SAFRAN.

5 CONCLUSIONS

We introduced in the present paper a new high-resolution (5 km) gridded precipitation dataset for Tunisia. It is the first product of this kind, to our knowledge, that covers one of the countries of the Middle East and North Africa region, which could be useful for various purposes such as climate model evaluation, climate studies and hydrological modelling. Our work will allow the scientific community to exploit such data for research purposes. A validation experiment has been conducted and it was found that the SAFRAN analysis outperforms other standard interpolation methods on two different validation samples with this Tunisian dataset. We also compared our SAFRAN dataset to EOBS, which is a very widely used dataset in particular in the Mediterranean region. The results show that EOBS has a strong negative bias, which exaggerates the aridity of the country. As climate models are often validated or bias-corrected with this EOBS dataset, this negative bias could influence the results over Tunisia and have strong impacts on water resources applications. We should not forget that in semi-arid and arid areas, soil moisture plays a very important role in climate processes and negative biases in precipitation will imply negative biases in soil moisture, with impacts in land-surface temperature, evapotranspiration, runoff generation, etc. Yet a note of caution must be provided about the uncertainties in the South of Tunisia, due to aridity and the low density of stations that does not guarantee robust spatial interpolation estimates.

ACKNOWLEDGEMENTS

This work is a contribution to the HYdrological cycle in The Mediterranean EXperiment (HyMeX) program, through INSU-MISTRALS support. The authors wish to thank the Direction Generale des Ressources en Eau of Tunisia for granting access to their database. The SAFRAN database is available for research applications on the HyMeX database at the address: https://doi.org/10.14768/mistrals-hymex.1514.

ORCID

Yves Tramblay © https://orcid.org/0000-0003-0481-5330

REFERENCES

Bargouzi, Z., Tramblay, Y., Lawin, E. and Servat, E. (2014) Seasonal precipitation variability in regional climate simulations over Northern basins of Tunisia. International Journal of Climatology, 34, 235–248.

Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B. and de Roo, A. (2017) MSWEP: 3-hourly 0.25_ global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci., 21, 589–615.

Ben Zaied, Y. and Binet, M.-E. (2015) Modelling seasonality in residential water demand: the case of Tunisia. Applied Economics, 47 (19), 1983–1996. https://doi.org/10.1080/00036846.2014.1002896.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Wayne Higgins, R. and Janowiak, J.E. (2008) Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res., 113, D04110. https://doi.org/10.1029/2007JD009132.

de Jager, A.L.L. and Vogt, J.V. (2010) Development and demonstration of a structured hydrological feature coding system for Europe. Hydrological Sciences Journal, 55(5), 661–675. https://doi.org/10.1080/02626667.2010.490786.

Dhiba, S., Mannaerts, C.M., Bargouzi, Z., Retsois, V. and Maathuis, B.H.P. (2017) Evaluating the MSG satellite Multi-Sensor Precipitation Estimate for extreme rainfall monitoring over northern Tunisia. Weather and Climate Extremes, 16, 14–22.

Durand, Y., Brun, E., Méridol, L., Guyomarc’h, G., Lesaffre, B. and Martin, E. (1993) A Meteorological Estimation of Relevant Parameters for Snow Models. Ann. Glaciol., 18, 65–71.

Durand, Y., Giraud, G., Brun, E., Méridol, L. and Martin, E. (1999) A Computer-Based Simulation Using Snowpack Structures as a Tool for Regional Avalanche Forecasting. J. Glaciol., 45, 469–484.

Droogers, P., Immerzeel, W.W., Terink, W., Hoogeveen, J., Bierkens, M.F.P., van Beek, L.P.H. and Debele, B. (2012) Water resources trends in Middle East and North Africa towards 2050. Hydrology Earth Syst. Sci., 16, 3101–3114.

Fantini, A., Raffaele, F., Torma, C., Bacer, S., Coppola, E., Giorgi, F. and Verdecchia, M. (2016) Assessment of multiple daily precipitation statistics in ERA-interim driven Med-CORDEX and EURO-CORDEX experiments against high resolution observations. Clim Dyn., 51, 877–900. https://doi.org/10.1007/s00382-016-3453-4.

Fathallhi, B., Pohl, B., Castel, T. and Safi, M.J. (2018) Errors and uncertainties in regional climate simulations of rainfall variability over Tunisia: a multi-model and multi-member approach. Clim Dyn., 52, 335–361. https://doi.org/10.1007/s00382-018-4150-2.

Feki, H., Slimani, M. and Cudennec, C. (2017a) Geostatistically based optimization of a rainfall monitoring network extension. Case of the climatically-heterogeneous Tunisia, Hydrology Research 2017 Apr, 48(2), 514–541. https://doi.org/10.2166/nh.2016.256.

Feki, H., Ouachani, R. and Dridi, S. (2017b) Spatio-temporal variability of extremes precipitation indices in Tunisia, 10th HYMEX Workshop, 4–July 7, 2017. Spain: Barcelona.

Feki, H. and Slimani, M. (2015) A comparison of three geostatistical procedures for rainfall network optimization. Hydrological Sciences and Water Security: Past, Present and Future, IAHS Publ, 366, 166–167. https://doi.org/10.5194/ptiaws-366-166-2015.

Feki, H., Slimani, M. and Cudennec, C. (2012) Incorporating elevation in rainfall interpolation in Tunisia using geostatistical methods. Hydrolog Sci J, 57, 1294–1314.

Flaounas, E., Drobinski, P., Borga, M., Calvet, J., Delrieu, G., Morin, E., Tartari, G. and Toffolon, R. (2012) Assessment of gridded observations used for climate model validation in the Mediterranean region: the HyMeX and MED-CORDEX framework. Environ. Res. Lett., 7, 024017. https://doi.org/10.1088/1748-9326/7/2/024017.

Goovaerts, P. (2000) Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology, 228, 113–129.
