Manuscript reference number: hess-2016-597 - Response to Anonymous referee #1

We would like to thank the referee for the helpful comments. Below we provide our response to the reviewer's comments and describe the modifications made to address them. In black font we include the reviewer's comments, and in blue is our response.

Manuscript entitled "Comparing Intensity–Duration–Frequency (IDF) curves derived from CMORPH and radar rainfall estimates over the Eastern Mediterranean" has been reviewed. In the paper, the authors compared IDF curves from radar and CMORPH in different climatic regions. They found that radar shows thicker tail distributions than CMORPH across the Eastern Mediterranean. The authors do not present a new tool. However, the manuscript is well organized. I divided my review into general comments and specific ones.

We would like to thank the referee for the detailed review.

This paper aims at advancing knowledge on the use of remote sensed data for rainfall frequency analysis in the form of IDF curves. The intent of this study is not to present a new "tool" for IDF estimation, but explore the use of remotely sensed precipitation from high-resolution satellite precipitation products vs. the more standard gauge-adjusted ground radar-rainfall fields. The use of standard tools for IDF estimation is crucial to avoid methodological issues masking out the results. To our knowledge this is the first study in which at-site IDF curves derived from different gridded datasets (from completely different sensors; with and without gauge-based adjustment) are compared. This approach strongly reduces the (spatial) scale issues related to the comparison of point and areal estimates, that hampered the research so far and potentially opens the way to apply remote sensing for deriving IDFs in data poor areas.

In order to make this aspect clearer, we modified the title of the manuscript to: "*Intensity–Duration–Frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the Eastern Mediterranean*" and revised introduction as follows: "*Since quantitatively accurate information is essential for design applications and the derivation of IDF curves based on historical records does not require short latency in the data, these studies made use of gauge-adjusted products and assessed the accuracy of the IDF curves derived from remote sensing datasets using rain gauge curves as a reference. However, this approach neglected two important aspects. First, early warning systems, e.g. for flash floods (Borga et al., 2011; Villarini et al., 2010; Borga et al., 2014), landslides/debris flows (Tiranti et al. 2014; Borga et al., 2014; Segoni et al., 2015) or heavy rain (Panziera et al. 2016), need to operate in real-time and rely on short-latency remote sensed measurements. In these situations, calculating the frequency of near real-time estimates using IDF curves derived from gauge-adjusted data could provide misleading results. It is therefore useful to analyse the characteristics of IDF curves derived from non-adjusted rainfall data, which are expected to represent the frequencies of near real time estimates. Second, areal IDFs provided by remote sensing instruments are expected to differ from point IDFs (Peleg et al., 2016a), and the use of different records (i.e. different samples of the climate) introduces further differences. No exact match between remote sensing and rain gauge IDFs should be expected a priori [...] To the authors' knowledge this is the first study in which at-site IDF curves derived from different gridded remote sensing datasets are compared*". [this will be recalled in the answer to reviewer's second comment].*

General comments:

1) The Koppen-Geiger classification used are not mentioned or described in section 2.1.

Thank you for pointing this out. We introduced the classification in section 2.1 as: *"Three climatic regions, Mediterranean, semiarid and arid, can be identified in the area, corresponding to the Csa, BSh and BWh Koppen-Geiger definitions, respectively (Peel et al., 2007). The criteria used to define the classes are reported in Table 1. In this study, we follow the classification by Srebro et al. (2011)."*

2) The lack of precision on the applied data leads to misunderstand some results: if the authors are using adjusted CMORPH gauge, what's the reason for analysis the unadjusted CMORPH data (in the same spatiotemporal resolution). I think there is something missing to clarify the methodology.

Thank you for raising this important issue – also shared by reviewer 2. As underlined in the general response, our introduction failed to emphasize the importance of using both the gauge-adjusted and un-adjusted versions of the satellite product (CMORPH, is this case). The gauge adjustment is expected to improve satellite estimates on average at the cost of having this improved product with several days of latency, which does not allow the use in real time applications. Furthermore, we would like to mention that there are regions on earth where there are no ground based sensors to adjust the satellite precipitation datasets. Hence, the natural question by the reviewers: 'why are you using the non-adjusted product in an application that is based on historical data rather than real time estimates?' The answer lies in the different requirements of the applications that make use of IDF curves: hydrologic design needs optimal quantitative accuracy, so evaluating unadjusted data is important for demonstrating uses of satellite precipitation in ungauged areas; early warning systems need to correctly identify the frequency of near real-time estimates. We updated text in the introduction to better motivate our study as mentioned above in the general response to the reviewer, and we recalled the concept in the conclusions, adding a new point: *"Comparison of HRC-IDF and CHRC-IDF against radar-IDF show consistent patterns of correlation and dispersion, and different biases. This means that gauge-adjustment influences the magnitude rather than the space-time organization of annual extremes and suggests that HRC-IDF can potentially be used to estimate the frequencies of CMORPH estimates in near real time early warning systems."*

In addition to this, it is interesting to note that the performance of gauge-adjustment on extremes is affected by the different measurement scales of remote sensing instruments and rain gauges. Moreover, the impact of gauge-adjustment over ungauged areas is rarely quantified, especially for extremes and potentially introduces non homogeneity in space due to different gauge density data.

[We group here two comments dealing with the same topic].

3) How is spatial (upscaled) data thought to affect precipitation (mainly IDF for return periods)? It is not clear to the readers without background knowledge of the appropriate upscaling method over the study area. Does an adopted upscaling approach perform better than some other techniques (this may not be necessary, but it should be discussed at least)?

7) Is there any physical link that could impacts on estimations due to different spatial resolution?

Thank you for raising these questions. The relation between point and areal extreme rainfall is the subject of research for a long time. As explained in the introduction, the reason for this interest is that the classic

approach derives IDF curves from point data (gauges) while for many applications the areal information (e.g. in a catchment scale) is required. The classic tool for converting point to areal information is the Areal Reduction Factor (ARF) method, already mentioned in the manuscript. We think that discussion on ARF methods is beyond the scope of this paper since it focuses on areal gridded information from remote sensing datasets: no upscaling method is used, no ARF is provided. However, we added text in the introduction to improve this aspect, also providing a useful reference: "*In principle, areal reduction factors may depend on a number of factors, such as geographic location, characteristics of the examined catchment, analysed duration and period, season, meteorological conditions, and others (Svensson and Jones, 2010). Their derivation is thus hampered by many sources of uncertainty.*"

4) Any thoughts on the reason why upscaling data is only found significant on the shape parameter (Section 4.1.1)? Maybe I am missing something in this part, but that is my understanding

Upscaling data significantly affects the location and scale parameters ("The location and scale parameters, as expected, consistently decrease as the spatial and temporal aggregation scales are increased and are not reported here"). The impact of spatial aggregation on mean and dispersion of the GEV distribution is related to the spatial smoothing of the extreme rainfall fields. This causes a decrease of the location and scale parameters. As in the case of ARFs, this decrease depends on a number of factors, which are difficult to isolate, and whose investigation falls out of the scope of this study. Conversely, the impact of spatial aggregation on the shape parameter can be easily assessed with the data in our possession and provides useful insights to the community. We revised the sentence to make it clearer: "*The location and scale parameters consistently decreased as the spatial and temporal aggregation scales increased. This is an expected effect, caused by the smoothing of rainfall fields operated by the spatial averaging, therefore results are not reported in this paper. Conversely, it is interesting to analyse the shape parameter.*".

In general, we believe that improving our knowledge on the interpretation of areal precipitation information is key for the future research in the field.

5) On the same subject, what is the CMORPH spatial upscaling (downscaling) impacts on precipitation?

We thank the reviewer for the question. The rationale for using weather radar data, rather than CMORPH data, for this analysis is twofold: (a) CMORPH estimates, even if provided at ~8 km resolution, are derived from coarser resolution PMW retrievals, so the relatively-small scale spatial structures are difficult to be captured with this product; (b) radar data allows to explore a broader range of scales.

Following the suggestion by the reviewer, we repeated the same analyses (starting from 8 km gridsize) using the CHRC data. The decrease of shape parameters with gridsize is less marked, but confirmed. Conversely, shape parameters slightly increased with duration: they are almost uniform for 8 km gridsize (between 0.14-0.16) and increase from ~0.02 (1h) to ~0.14-0.15 (12h-24h) for 64 km. The variations are less marked than in the radar case, since the shape parameters from CMORPH are already low. We however think that CHRC provides less accurate information, due to the point (a) above.

6) Why overload figures with features with the insufficient discussion in the text (e.g., Lack of enough discussion for location and scale parameters or for a different time interval in Fig 4)?

We thank the reviewer for this comment. We revised Fig. 2 showing the distribution of the scale parameters normalized over the corresponding location parameters (dispersion normalized over the

mean). The text in section 4.1 is subsequently revised as follows: "*Location parameters from HRC (CHRC) estimates are smaller (larger) than the ones from radar over Mediterranean climate and over the sea, meaning that extreme values from HRC (CHRC) are in general lower (higher) than radar extreme values while in semiarid and arid climates HRC and CHRC generally identify higher parameters than the radar (i.e. higher extreme values). Differences in the location parameters can be associated to the bias between extreme values in the datasets. The scale parameters are normalized over the corresponding location parameters to appreciate the relative differences. Normalized scale parameters from HRC and CHRC are similar and lower than the ones derived from radar. Normalized scale parameters, together with their variability, tend to increase when moving from sea to Mediterranean, semiarid and arid climates. The drier climate, the larger the dispersion of the GEV distribution. A slight increase of the normalized scale parameters with duration can be noticed in the HRC/CHRC data*".

6b) The figures axis could be improved if the authors want to make it clear to the reader. Also, I suggest the same size intervals in Y-axis (e.g., Fig 2&4).
[We moved here one of the specific comments of the reviewer that concerned the same issue: It is also suggested to present the parameter variation against the time by a linear representation that could help the readers to grasp the message from the figures.]
Thanks for the suggestion. The figure was revised. A linear scale on the duration axis is not recommended due to the almost logarithmically equi-spaced differences between the examined durations. Instead we used a logarithmic scale in the duration axis in order to improve the information provided by the figure and included the x-axis labels in all panels. Caption was then updated accordingly: "*The parameters for different products are represented around the corresponding duration so the logarithmic scale in x-axes should be interpreted accordingly*"

7) [Moved above]. We kindly refer to the reply to comment 3.

8) Some unnecessary materials: in my opinion, it is completely unnecessary to have Figure 9, when it is already well-known that the uncertainties increase for longer return period.
[We move here one of the specific comments of the reviewer that concerned the same issue: - Why Fig 9 presented? It is well known that confidence in a return level decreases rapidly when the period is more than about two times the length of the original data set. Fig 9 could be removed.]
Uncertainty increases with return period. As underlined by the reviewer, this is well known and expected. However, our objective is to quantify the uncertainty due to the use of a reduced record for different climates and different record lengths. In fact, the information provided by Fig. 9 and 10 is more than "uncertainties increase for longer return period": it includes (i) quantitative information (ii) for different climates and (iii) for different record lengths. Fig. 9 only reports 3 examples among many, but we think is it useful for the reader to better understand the message underlying Fig. 10.

9) Finally, I would like to draw your attention to a somewhat old, but still relevant, review on methods to IDF and DDF curves and their uncertainties that might have some valuable lessons in it: Aart Overeem, Adri Buishand and Iwan Holleman, Journal of Hydrology, 2008, 348, pp 124-134

4

We thank the reviewer you for the suggestion. The suggested paper, actually still up-to-date, helped in this revision process and is now cited in the new version of the manuscript.

Specific Comments:

Page 5- Line 15:
- I would expect to introduce parameters and explain a little bit how they were computed.
- It is not clear which method used for estimating the GEV distribution parameters by maximum likelihood.
- The assumptions regarding the convenience of the GEV to represent the AMS distribution through the Fisher-Tippett theorem, as well as calculation procedures, would be useful to present in the Appendix.
We thank the reviewer for these comments. We would like to point out that we used an established extreme values distribution (GEV) and derived the distribution parameters using one of the most commonly used methods (maximum likelihood). Details on these methods can be found in textbooks (e.g. Coles, 2001) and in the references provided in the text. Our contribution does not advance knowledge on this aspect.
As the reviewer suggested, the parameters need to be introduced in the main text. Therefore, we updated text in section 3.1 as follows: *"The GEV distribution is a three parameters extreme values distribution used worldwide to model rainfall extremes (e.g. Fowler and Kilsby, 2003; Gellens, 2002; Koutsoyiannis, 2004; Overeem et al., 2008). It is described by the location, scale and shape parameters, representing mean, dispersion and skewness of the distribution, respectively. [...] At-site GEV parameters (i.e. pixel by pixel) were identified using the maximum likelihood method (MATLAB statistics toolbox)."*
As suggested, Appendix was revised to include information on the convergence of extreme values distributions to the GEV: *"Under hypotheses on the regularity of the tail of the distribution, the Fisher-Tippet theorem demonstrates that GEV distribution is the only possible limit distribution for the extreme values of independent and identically distributed random variables."*

- The bootstrap was applied to assess this uncertainty of record length. This method considers only the uncertainty for the return periods (it's my understanding). What about the uncertainty of the estimation of the GEV parameters (i.e. sampling errors)?
The MATLAB statistics toolbox used in this study provides the confidence interval of the derived GEV parameters. However, the uncertainty of estimation of the GEV parameters is not the subject of this study. Overeem et al. (2008) and Overeem et al. (2009) have proposed the bootstrapping of the data series used for the GEV fit to identify the uncertainty related to the fit. Perhaps this approach used in the past motivates the question by the referee.
In sections 3.3 and 4.4 of this study, we are assessing the uncertainty related to the record length, i.e. to the under-sampling of rainfall climatology. To do so, as mentioned by the reviewer, we devised the bootstrap approach based on long records of rain gauge data, that are here assumed to represent the variability of extreme rainfall climatology (section 3.3: *"We assumed the records of rain gauge data to be a complete sample of the climatology of extremes for return periods comparable to the remotely sensed data record length"*). The bootstrapped approach consists in three steps: (i) randomly sampling the AMS, (ii) deriving the GEV parameters and (iii) deriving the IDF values for selected return periods. The record

length-related uncertainty in the GEV parameters is implicitly calculated (as 5-95th quantile interval) as a necessary step of the bootstrapping.

Page6- Line 12:
- Figure 2: Estimates of the GEV parameters could be present by Box-Whisker plot that obtained from 999 bootstrap samples, as well for 25th and 75th percentiles of the bootstrap samples.
We think the reviewer misunderstood Fig. 2. The figure shows the 25-75th quantile interval of the set of parameters obtained from pixels characterized by the examined climate, i.e. the 25-75th quantile interval among the number of pixels reported in Table 1. No bootstrap is performed here. We prefer to keep the quantile interval bars rather than using a box plot (i) because the number of elements in each climatic class is different and (ii) for consistency with Fig. 3 and with the use of quantile intervals we adopted throughout the paper.

It is also suggested to present the parameter variation against the time by a linear representation that could help the readers to grasp the message from the figures.
[Moved above]. We kindly refer to comment 6b above for more information.

- How are different climatic regions thought to affect precipitation (location and scale parameters)? It should be discussed at least (regarding figure 2).
Thank you for making this point. We revised section 2.1 by inserting the following statement: "*Important gradients have been reported also for the climatology of extreme rainfall. Low return period intensities were found to be scaled with the mean annual precipitation. Conversely, the more arid the climate is, the more skewed the extreme value distribution is, with long return period intensities for arid areas being higher than the corresponding values for semiarid and Mediterranean areas, especially for short durations (Ben Zvi, 2009; Marra and Morin, 2015).*"

Page 6-
- The location and scale parameters are not addressed in details. It is not clear to the reader without background knowledge of the GEV df.
Thank you for making this point. We added text to recall the meaning of the parameters and to be more explicit: "*We recall here that the scale, location and shape parameters provide a measure of the mean, dispersion and skewness of the underlying distribution, respectively. Location parameters from HRC (CHRC) estimates are smaller (larger) than the ones from radar over Mediterranean climate and over the sea, meaning that extreme values from HRC (CHRC) are in general lower (higher) than radar extreme values while in semiarid and arid climates HRC and CHRC generally identify higher parameters than the radar (i.e. higher extreme values). Differences in the location parameters can be associated to the bias between extreme values in the datasets. The scale parameters are normalized over the corresponding location parameters in order to appreciate the relative differences. Normalized scale parameters from HRC and CHRC are similar and lower than the ones derived from radar. Normalized scale parameters, together with their variability, tend to increase when moving from sea to Mediterranean, semiarid and arid climates. The drier climate, the larger the dispersion of the GEV distribution. A slight increase of the normalized scale parameters with duration can be noticed in the HRC/CHRC data.*"

Page 8-

- I am totally lost. What's the link between section 4.4 with previous sections?

This section presents results on the quantification of uncertainty related to the record length of remote sensing datasets, introduced in section 3.3. To make things clearer we added a short introductory sentence to the section: "*In this section, we present the results of the bootstrap sampling of long rain gauge records used to quantify the uncertainty related to the record length of remote sensing datasets. The uncertainty presented here is the component related to the under-sampling of rainfall climatology due to the use of short data records and is quantified as the 5–95th quantile interval of the bootstrap sampling.*"

Page 12- Line 13:

- Sorry if I missed this, how is the coverage of upscaled radar and CMORPH pixels? Are they exactly the same? Please, explicitly specify their spatial coverage.

Page 12 actually contains references, so we are not sure we understood what the reviewer is referring to (page 7?). We think that the reviewer can be either asking about each "upscaled" radar and CMORPH pixel, information that can be found in section 3.1 (page 5, lines 11-12): "*In order to have radar and satellite data on a common grid suitable for the comparison, the full archive of hourly radar data was remapped by spatially averaging the 1×1 km2 radar pixels to the corresponding ~8×8 km2 CMORPH pixels*" or about the areal coverage of the comparison, which is reported in section 3.2 (page 5, lines 24-27): "*The comparison is extended over an analysis domain defined excluding the pixels that are known to be not reliable. In particular, pixels located closer than 27 km or farther than 185 km from the radar or behind the hilly region are excluded due to insufficient reliability of the radar data, and pixels located in proximity of major lakes are excluded due to false rainfall signals in the CMORPH rainfall estimates*".

Page 18-

- What's the dashed line in Fig 2 for shape parameter?

The dashed lines showed the "0" value. We propose to remove it in order to make the figure clearer.

Page 20-

- Appendix A. The legend and description in caption do not show similar things (check).

We guess this comment also refers to both Fig. 4 and 6. We apologize for this; the captions will be updated.

Page 25-

- Why Fig 9 presented? It is well known that confidence in a return level decreases rapidly when the period is more than about two times the length of the original data set. Fig 9 could be removed.

[Moved above]. Please, see our response to comment 8 above.

References

Borga, M., E. N. Anagnostou, G. Blöschl, and J. D. Creutin. 2011. Flash Flood Forecasting, Warning and Risk Management: The HYDRATE Project. Environmental Science and Policy 14(7): 834–44

Borga M., M. Stoffel, L. Marchi, F. Marra, M. Jacob 2014: Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. Journal of Hydrology, 518 (2014), 194–205. http://dx.doi.org/10.1016/j.jhydrol.2014.05.022

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.

Overeem, A., Buishand, A., Holleman, I., 2008. Rainfall Depth-Duration-Frequency Curves and Their Uncertainties. J. Hydrol. 348(1–2): 124–34. http://dx.doi.org/10.1016/j.jhydrol.2007.09.044

Panziera L., M. Gabella, S. Zanini, A. Hering, U. Germann and A. Berne, 2016. A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland. Hydrol. Earth Syst. Sci., 20, 2317–2332. http://dx.doi.org/10.5194/hess-20-2317-2016

Segoni, S., Battistini, A., Rossi, G., Rosi, A., Lagomarsino, D., Catani, F., Moretti, S., and Casagli, N.: Technical Note: An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds, Nat. Hazards Earth Syst. Sci., 15, 853-861, http://dx.doi.org/10.5194/nhess-15-853-2015

Svensson, C., and D. A. Jones. 2010. Review of Methods for Deriving Areal Reduction Factors. J. Flood Risk Management 3 3: 232–45. http://dx.doi.org/10.1111/j.1753-318X.2010.01075.x

Tiranti, Davide et al. 2014. "The DEFENSE (Debris Flows triggEred by Storms - Nowcasting System): An Early Warning System for Torrential Processes by Radar Storm Tracking Using a Geographic Information System (GIS)." Computers and Geosciences 70: 96–109. http://dx.doi.org/10.1016/j.cageo.2014.05.004

Villarini, Gabriele et al. 2010. "Towards Probabilistic Forecasting of Flash Floods: The Combined Effects of Uncertainty in Radar-Rainfall and Flash Flood Guidance." Journal of Hydrology 394(1–2): 275–84. http://dx.doi.org/10.1016/j.jhydrol.2010.02.014

Manuscript reference number: hess-2016-597 - Response to RC2

We would like to thank the reviewer for his/her to-the-point comments. We provide here our response and modifications introduced to the manuscript to address the reviewer's comments and improve the general presentation of the study. Below, in black text font is the reviewer's comments, and in blue is our response.

The manuscript "Comparing Intensity–Duration–Frequency curves derived from CMORPH and radar rainfall estimates over the Eastern Mediterranean" represents a good contribution to the assessment of satellite-based rainfall estimations in areas where their impact could be potentially relevant. The authors have done a thorough analysis over a sufficient amount of time and thus the results are significant.
The manuscript certainly fits the journal scope and is suggested for publication after the following points are considered by the authors:
We would like to thank the reviewer for his/her positive review.

1) The reason why both versions (gauged and ungauged) of CMORPH are used is not sufficiently explained and detailed in the text.
Thank you for raising this important issue – also shared by reviewer 1. Our introduction failed to emphasize the importance of using both the gauge-adjusted and un-adjusted versions of the satellite product (CMORPH, is this case). The gauge adjustment is expected to improve satellite estimates on average at the cost of having this improved product with several days of latency, which does not allow the use in real time applications. Furthermore, we would like to mention that there are regions on earth where there are no ground based sensors to adjust the satellite precipitation datasets. Hence, the natural question by the reviewers: 'why are you using the non-adjusted product in an application that is based on historical data rather than real time estimates?' The answer lies in the different requirements of the applications that make use of IDF curves: hydrologic design needs optimal quantitative accuracy, so evaluating unadjusted data is important for demonstrating uses of satellite precipitation in ungauged areas; early warning systems need to correctly identify the frequency of near real-time estimates. We updated text in the introduction to better motivate our study: *"[...] early warning systems, e.g. for flash floods (Borga et al., 2011; Villarini et al., 2010; Borga et al., 2014), landslides/debris flows (Tiranti et al. 2014; Borga et al., 2014; Segoni et al., 2015) or heavy rain (Panziera et al. 2016), need to operate in real time and rely on short-latency remote sensed measurements. In these situations, calculating the frequency of near real-time estimates using IDF curves derived from gauge-adjusted data could provide misleading results. It is therefore useful to analyse the characteristics of IDF curves derived from non-adjusted rainfall data, which are expected to represent the frequencies of near real time estimates."*, and we recalled the concept in the conclusions, adding a new point: *"Comparison of HRC-IDF and CHRC-IDF against radar-IDF show consistent patterns of correlation and dispersion, and different biases. This means that gauge-adjustment influences the magnitude rather than the space-time organization of annual extremes and suggests that HRC-IDF can potentially be used to estimate the frequencies of CMORPH estimates in near real time early warning systems."*
In addition to this, it is interesting to note that the performance of gauge-adjustment on extremes is affected by the different measurement scales of remote sensing instruments and rain gauges. Moreover,

9

the impact of gauge-adjustment over ungauged areas is rarely quantified, especially for extremes and potentially introduces non homogeneity in space due to different gauge density data.

2) Did the authors check how many of the gauges are used in the production of the gauge-adjusted version of CMORPH in the area? This could have an impact in the analysis, maybe small. However, a couple of words on the subject need to be included.

Thank you for the question. The gauge-adjusted CMORPH product is using limited gauge data from the region, which may be also used in the adjustment of the radar-rainfall dataset. From Fig. 1d in Chen et al. (2008), the number of CPC gauges in the study area should be around ~12. We have contacted the developers of the gauge-adjusted CMORPH product for more specific information, but we have not received a response on our inquiry yet. We will mention this aspect in section 2.3: "*The gauge-adjusted CMORPH product is using data from ~12 gauges in the region (Chen et al., 2008), which may also be used in the adjustment of the radar-rainfall dataset.*"

3) While it is true that gauges are not available over the sea (!) the authors should also spend a few words on the fact that normally satellite-based estimations are far better over the sea surface. This partially contradicts their results. Just pay attention to this important fact. See, for example: Kidd, C., and V. Levizzani, 2011: Status of satellite precipitation retrievals. Hydrol. Earth Syst. Sci., 15, 1109-1116 for a discussion on the various methods of rain estimation from satellite.

We would like to thank the reviewer for this suggestion. We think that our results and text do not contradict his/her point. The results show that the CC between satellite and radar IDF maps over the sea is low. The text associates this to the absence of rain gauges over the sea. Since (i) gauge-adjustment is a crucial part of the radar quantitative precipitation estimation and (ii) similar CC patterns for HRC and CHRC are observed over land, we think that satellite datasets are expected to provide more accurate information on the spatial distribution of IDFs over the sea. We believe there has been a slight misunderstanding on the interpretation of this sentence, so we revised text to improve its clarity, also included a reference to the suggested paper: "*As pointed out above, gauge adjustment is only weakly impacting the space-time organization of CMORPH extreme estimates, while it is a crucial step in radar quantitative precipitation estimation. This observation, together with the increased reliability of satellite based estimations over the sea (Kidd and Levizzani, 2011), suggests that spatial distribution of IDF values indicated by satellite products should be considered more accurate.*"

4) A fine combing of the English is suggested since many imperfections are detected throughout the text.
English language has been re-checked and improved, thank you.

Minor point: The caption of Fig. 4 does not match the real colors used in the figure. It appears that the authors have used two different version prior and after a change they did while writing.
The reviewer guessed correctly. We apologize for the inconvenient, that affected also Fig. 6. The captions have now been updated.

References

Borga, M., E. N. Anagnostou, G. Blöschl, and J. D. Creutin. 2011. Flash Flood Forecasting, Warning and Risk Management: The HYDRATE Project. Environmental Science and Policy 14(7): 834–44

Borga M., M. Stoffel, L. Marchi, F. Marra, M. Jacob 2014: Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. Journal of Hydrology, 518 (2014), 194–205. http://dx.doi.org/10.1016/j.jhydrol.2014.05.022

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

Panziera L., M. Gabella, S. Zanini, A. Hering, U. Germann and A. Berne, 2016. A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland. Hydrol. Earth Syst. Sci., 20, 2317–2332. http://dx.doi.org/10.5194/hess-20-2317-2016

Segoni, S., Battistini, A., Rossi, G., Rosi, A., Lagomarsino, D., Catani, F., Moretti, S., and Casagli, N.: Technical Note: An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds, Nat. Hazards Earth Syst. Sci., 15, 853-861, http://dx.doi.org/10.5194/nhess-15-853-2015

Tiranti, Davide et al. 2014. "The DEFENSE (Debris Flows triggEred by Storms - Nowcasting System): An Early Warning System for Torrential Processes by Radar Storm Tracking Using a Geographic Information System (GIS)." Computers and Geosciences 70: 96–109. http://dx.doi.org/10.1016/j.cageo.2014.05.004

Villarini, Gabriele et al. 2010. "Towards Probabilistic Forecasting of Flash Floods: The Combined Effects of Uncertainty in Radar-Rainfall and Flash Flood Guidance." Journal of Hydrology 394(1–2): 275–84. http://dx.doi.org/10.1016/j.jhydrol.2010.02.014

11

Intensity–Duration–Frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the Eastern Mediterranean

~~Comparing Intensity Duration Frequency curves derived from CMORPH and radar rainfall estimates over the Eastern Mediterranean~~

Francesco Marra[1], Efrat Morin[1], Nadav Peleg[2], Yiwen Mei[3], Emmanouil N. Anagnostou[3]

[1]Institute of Earth Sciences, Hebrew University of Jerusalem, 91904, Israel
[2]Institute of Environmental Engineering, Hydrology and Water Resources Management, ETH Zurich, Switzerland
[3]Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA

Correspondence to: Francesco Marra (marra.francesco@mail.huji.ac.il)

Abstract. Intensity–Duration–Frequency (IDF) curves are widely used to quantify the probability of occurrence of rainfall extremes. The usual rain gauge based approach provides accurate curves for a specific location, but uncertainties arise when ungauged regions are examined or catchment scale information is required. Remotely sensed rainfall records, e.g. from weather radars and satellites, are recently becoming available, providing high resolution ~~information on rainfall extremes~~ estimates at regional or even global scales: their uncertainty and implications on water resources applications urge to be investigated. This study compares IDF curves from radar and satellite (CMORPH) estimates over the Eastern Mediterranean (covering Mediterranean, semiarid and arid climates) and quantifies the uncertainty related to their limited record on varying climates. We show that radar identifies thicker tail distributions than satellite, in particular for short durations, and that the ~~shape parameters~~ tail of the distributions depends on the spatial and temporal aggregation scales. The spatial correlation between radar-IDFs and satellite-IDFs is as high as 0.7 for 2–5 years return period and decreases with longer return periods, especially for short durations. The uncertainty related to the use of short records is important when the record length is comparable to the return period (~50%, ~100% and ~150% for Mediterranean, semiarid and arid climates, respectively). The agreement between IDF curves derived from different sensors on Mediterranean and, to a good extent, semiarid climates, demonstrates the potential of remote sensing datasets and instils confidence on their quantitative use for ungauged areas of the Earth.

1 Introduction

Intensity–Duration–Frequency (IDF) curves are widely used in hydrological design and as decision support information in flood risk and water management (Watt and Marsalek, 2013). They are distribution functions of rain intensity maxima conditioned on duration and allow linking the characteristics of a rainfall event to the probability of its occurrence (Chow et al., 1988; Eagleson, 1970). Derivation of IDF curves generally relies on historical precipitation data and consists in fitting an

extreme value distribution to ~~the peak annual~~extreme rainfall values. Rain gauges, providing long records of precipitation data, are traditionally used to estimate IDF curves at the gauge locations. Nevertheless, owing to the sparseness of gauge networks worldwide (Kidd et al., 2016), ~~the rain gauge~~this approach raises ~~some~~ important issues when~~ever~~ design applications require information at the catchment scale. Moreover, the representativeness of gauge derived IDFs decreases moving away from the

5 gauge location and no or very sparse information is available for the many ungauged locations of the Earth.

In the last decades a body of research has been devoted to these issues. Areal reduction factors and design storms assume homogeneity ~~of the climatology~~ of rainfall extremes climatology ~~allowing~~ to adapt the point IDF values estimated by rain gauges to wider areas, such as ~~a~~ catchments, basing either on the climatology of the region (Sivapalan and Bloschl, 1998), on spatial precipitation observations (Allen and De Gaetano, 2005; Wright et al., 2013; Paixao et al, 2015) or stochastic model

10 simulations (Peleg et al., 2016a, 2016b). In principle, areal reduction factors may depend on a number of factors, such as geographic location, characteristics of the examined catchment, analysed duration and period, season, meteorological conditions, and others (Svensson and Jones, 2010). Their derivation is thus hampered by many sources of uncertainty. Regionalization and interpolation techniques continue assuming spatial homogeneity to derive IDF curves for ungauged locations (Ceresetti et al., 2012). However, homogeneity of rainfall extremes is a weak assumption when wide areas are

15 considered or when, even at the small scales, the data record length is limited, especially in the case of short durations (Peleg et al., 2016a).

Remote sensing ~~precipitation estimates~~instruments, such as ~~those derived from~~ weather radars or satellites ~~observations~~, provide high spatiotemporal resolution (i.e. 1–10 km and 5–60 min), distributed, regional or even global rainfall estimates. These datasets, allow~~ing~~ capturing dynamics and variability of rainfall extremes at scales that cannot be represented by rain

20 gauges (Amitai et al., 2011; Chen et al., 2013; Marra et al., 2016; Panziera et al., 2016; Tapiador et al., 2012) and permit to overcome the issues related to the conversion of point to areal information. Given the quantitative uncertainty related to remote sensing precipitation products (Berne and Krajewski, 2013; Mei et al. 2014; Stampoulis et al., 2013) and their limited records, unable to provide adequate samples of the climatology, the possibility of using such datasets for rainfall frequency analysis has only recently started to be explored., Main focus of these studies was the assessment of~~mainly assessing~~ their potential

25 (Eldardiry et al., 2015), for ~~analysing~~ poorly instrumented regions (Awadallah et al., 2011; Nastos et al., 2013) ~~or~~ and areas characterized by strong climatic gradients (Marra and Morin, 2015). Few authors, so far, provided quantitative IDF curves from remote sensing instruments, using some kind of regionalization approach (Endreny and Imbeah, 2009; Overeem et al., 2009; Wright et al., 2013; Panziera et al., 2016).

Since quantitatively accurate information is essential for design applications and the derivation of IDF curves based on

30 historical records does not require short latency in the data, these studies made use of gauge-adjusted products and assessed the accuracy of the IDF curves derived from remote sensing datasets using rain gauge curves as a reference. However, this approach neglected two important aspects. First, early warning systems, e.g. for flash floods (Borga et al., 2011; Villarini et al., 2010; Borga et al., 2014), urban floods (Yang et al., 2016), landslides/debris flows (Tiranti et al. 2014; Borga et al., 2014; Segoni et al., 2015) or heavy rain (Panziera et al. 2016), need to operate in real-time and rely on short-latency remote sensed

measurements. In these situations, calculating the frequency of near real-time estimates using IDF curves derived from gauge-adjusted data could provide misleading results. It is therefore useful to analyse the characteristics of IDF curves derived from non-adjusted rainfall data, which are expected to represent the frequencies of near real time estimates. Second, areal IDFs provided by remote sensing instruments are expected to differ from point IDFs (Peleg et al., 2016a), and the use of different records (i.e. different samples of the climate) introduces further differences. No exact match between remote sensing and rain gauge IDFs should be expected a priori. Moreover, evaluating unadjusted data is important for demonstrating the value of satellite products in ungauged areas.

The aim of this study is to advance knowledge on the use of remote sensing precipitation estimates for rainfall frequency analysis. IDF curves computed from different datasets, namely the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center morphing (CMORPH) technique (Joyce et al. 2004), the gauge–adjusted CMORPH (Xie, et al., 2011) and the ground based C-Band Shacham weather radar archive (Marra and Morin, 2015), are compared over the Eastern Mediterranean using common spatiotemporal scales and records lengths. The effect of spatial and temporal aggregation of rainfall is analysed. The uncertainty related to the use of short records in the climatic contexts characterizing the study region is assessed using long records of rain gauge observations. To the authors' knowledge this is the first study in which at-site IDF curves derived from different gridded remote sensing datasets are compared.

The study area and the rainfall datasets are presented in section 2. Section 3 describes the methods used for computing and comparing satellite and radar IDF curves and for estimating the uncertainty related to the record length. In Sect. 4 the results of the study are presented and discussed. Section 5 provides the conclusions and suggestions for the future use of remote sensing datasets for rainfall frequency analysis.

2 Study area and data

2.1 Study area

The present study is focused on the Eastern Mediterranean, in particular on the area covered by the Shacham weather radar, shown in Fig. 1. The orography presents a longitudinal organization that, going east from the Mediterranean Sea, encounters a coastal plain, a hilly region (up to ~800 m a.s.l.), the Jordan rift valley (~ -400 m a.s.l.) and the Jordan Plateau (~1000 m a.s.l.). Mountains in the north of the study area raise up to 2800 m a.s.l.

The area is characterized by varying climatic conditions (Srebro et al., 2011). Three climatic regions, Mediterranean, semiarid and arid, can be identified in the area, corresponding to the Csa, BSh and BWh Koppen-Geiger definitions, respectively (Peel et al., 2007). The criteria used to define these classes are reported in Table 1. In this study, we follow the classification by Srebro et al. (2011). Mediterranean climate characterizes the north-western coastal region and the northern part of the Jordan Plateau. Arid climate characterizes the south-western and southern portions of the area, and the Jordan rift valley. These two regions are separated by a strip characterized by semiarid climate that can sometimes be very narrow (Fig. 1, Table 1). Climatic gradients are important with mean annual precipitation exceeding 1000 mm yr^{-1} in the northern part of the area and dramatically

14

dropping from 600 to 100 mm yr^{-1} in just 25 km distance along the narrow strip west of the hilly region (Alpert and Shafir, 1989; Goldreich, 1994). The summer is dry, with almost no rain from June to August. Rainfall over the Mediterranean areas is generally brought by cold fronts and postfrontal systems, associated with mid latitude cyclones (Goldreich, 2003; Peleg and Morin, 2012), and by Syrian low and active Red Sea Trough, occasionally correlated with flash floods, in the semiarid and arid climates (Dayan et al., 2001; Dayan and Morin, 2006; Kahana et al. 2002; de Vries et al., 2013). On rare occasions the Subtropical Jet, or Tropical Plume, brings widespread rainfall over the whole region (Dayan and Morin, 2006; Kahana et al., 2002; Rubin et al., 2007; Tubi and Dayan, 2014; Dayan et al., 2015). Important gradients have been reported also for the climatology of extreme rainfall. Low return period intensities were found to be scaled with the mean annual precipitation. Conversely, the more arid the climate is, the more skewed the extreme value distribution is, with long return period intensities for arid areas being higher than the corresponding values for semiarid and Mediterranean areas, especially for short durations (Ben Zvi, 2009; Marra and Morin, 2015).

2.2 Radar data

The Shacham weather radar is a C-Band (5.35 cm wavelength) non-Doppler instrument, operational since the late 1980s. Observations from this radar have been archived from October 1990 to Mar 2014, providing a unique 23-year record and have been extensively used for climatologic and hydrological studies (e.g. Morin et al. 1995, 2001, 2009; Morin and Gabella 2007; Peleg and Morin 2012; Peleg et al. 2013, 2016a; Yakir and Morin 2011).

-The Shacham radar record has recently been reanalysed for rainfall frequency analysis (Marra and Morin 2015). Radar quantitative precipitation estimation is obtained combining physically based correction algorithms and quantitative adjustments based on the comparison with rain gauge measurements. The procedure is discussed in details in Marra and Morin (2015) and included: (i) checking of antenna pointing, (ii) ground clutter filtering, corrections for the effects of (iii) wet radome attenuation, (iv) attenuation due to the propagation of the radar beam, (v) beam blockage and (vi) vertical variations of reflectivity, together with a (vii) hail filter. A two-step bias adjustment was then applied combining a yearly range dependent and an event based mean field bias adjustments based on comparison with quality checked rain gauge data. Radar rain rates exceeding 150 mm h^{-1} have been set to this cap value in order to avoid contaminations due to hail. The readers are referred to Marra et al. (2014) and Marra and Morin (2015) for an extensive description of the correction procedures and the quantitative assessment of the archive.

This study is based on the hourly radar archive. Each hourly radar product is created when at least 60% of radar scans are available during the 1 h time interval. This quality check on the data availability ensures a good coverage of the examined time interval and allows exploring durations longer than the ones analysed by Marra and Morin (2015).

2.3 Satellite data

The satellite precipitation products selected for this study are the high resolution CMORPH (HRC) and the its gauge-adjusted CMORPH version (CHRC) available from NOAA CPC. The two satellite products, with a resolution of 0.073º/half-hourly,

15

offer high resolution, quasi global, long term records of satellite ~~based~~ rainfall estimates (available, as at November 2015, for 16 years: from January 1998 to December 2013). CMORPH integrates multiple satellite based microwave rain estimates (Ferraro et al., 1997; 2000; Kummerow et al., 2001) in space and time using motion vectors derived from infrared images (Joyce et al. 2004). The newly available gauge-adjusted CMORPH applies daily gauge-adjustment using estimates from

5   ~30,000 gauges worldwide (Xie, et al., 2011). The gauge-adjusted CMORPH product is using gauge data from ~12 gauges in the region (Chen et al., 2008), which may be also be used in the adjustment of the radar-rainfall dataset. The ~~CMORPH~~ half-hourly original data has been aggregated to hourly intervals to match the radar archive resolution.

2.4 Rain gauge data

The small scale representativeness of rain gauge measurements makes them not suitable for a large scale quantitative

10   assessment of remote sensing products (Gires at al., 2014; Peleg et al., 2016a). Here, we take advantage of their long records to empirically quantify the uncertainty in rainfall frequency analysis related to the use of short records in different climatic conditions.

We identified 11 rain gauges, operated by the Israeli Meteorological Service, among the ones already used by Marra and Morin (2015). Rain gauge data is available as storm maxima values extracted from digitized charts for durations up to four hours

15   (before 2000) or automatic measurements with 10 min (from 2000 up to 2005) and 1 min (from 2006 on) resolution. This limits our analysis of rain gauge data to durations shorter than ~~4 h~~four hours. When possible, rain gauge data have been aggregated into hourly blocks to be as consistent as possible with the remote sensing datasets. Stationary of the records has been tested at 0.1 significance level (Phillips and Perron, 1988). The selected rain gauges belong to different climatic regions with five, four and two rain gauges available for Mediterranean, semiarid and arid climates, respectively (Fig. 1). The length

20   of the analysed records ranges between 30 and 67 years and is at least double of the satellite datasets record (16 years) with just one exception (30 years) belonging to the semiarid climatic region.

3 Methods

3.1 Derivation of Intensity–Duration–Frequency curves

In this study, the Generalized Extreme Value (GEV) distribution (Appendix A) is used to fit the Annual Maxima Series (AMS)

25   of average rain intensities observed over 1, 3, 6, 12 and 24 h durations. The use of AMS ensures independency of the elements of the series and, rather than peak over threshold series, is suitable for this study because it does not require the definition of thresholds, problematic operation on highly variable climatic conditions and~~,~~ potentially undermining the interpretation of the ~~results obtained from~~comparison between different datasets. The GEV distribution is a three parameters extreme values distribution used worldwide to model rainfall extremes (Fowler and Kilsby, 2003; Gellens, 2002; Koutsoyiannis, 2004;

30   Overeem et al., 2008). It is described by the location, scale and shape parameters, representing mean, dispersion and skewness of the distribution, respectively. The GEV distribution ~~and it~~ has been shown to fit the AMS better than other distributions,

such as the Pearson type III or the generalized logistic (Alila, 1999; Kysely and Picek, 2007; Perica et al., 2013; Schaefer, 1990). It's being widely ~~It has recently been~~ used for rainfall frequency analyses based on remote sensed data (Eldardiry et al., 2015; Marra and Morin, 2015; Overeem et al., 2009; Paixao et al., 2015; Panziera et al. 2016; Peleg et al, 2016a), owing to its ability to include all the three asymptotic extreme value types (Gumbel, Fréchet and Weibull) into one (Katz et al., 2002).

5   In order to have radar and satellite data on a common grid suitable for the comparison, the full archive of hourly radar data was remapped by spatially averaging the 1×1 km^2 radar pixels to the corresponding ~8×8 km^2 CMORPH pixels. We identified the AMS of calendar years for the examined durations using a moving window with 1 h jumps. Even if using hydrologic years is more natural, we chose to use calendar years to exploit the full CMORPH record, available from January 1998 to December 2013. In order to focus on a unique set of IDFs for each dataset, all the available rainfall estimates have been used, even if data

10   from another source was missing for the same storm, so that no co-occurrence of the annual maxima is imposed. There are, in fact, potential situations in which radar or rain gauges missed storms due to technical problems or power issues (Morin et al., 1998; Ben Zvi, 2009; Marra and Morin, 2015). ~~The~~ At-site GEV parameters (i.e. pixel by pixel) were identified ~~fit was operated~~ using the maximum likelihood method (MATLAB statistics toolbox). Situations for which the fitting was not successful due to convergence problems or excessive number of iterations were discarded. Less than 0.7% of the cases were discarded for

15   this reason. The IDF values were calculated for a set of five return periods (2, 5, 10, 15 and 25 years).

The effect of spatial and temporal aggregation of rainfall estimates, key issue when dealing with remote sensing instruments, is analysed spatially aggregating the original radar record (23 years, 1×1 km^2) on grid sizes gradually increasing from 2×2 to 64×64 km^2 and for durations from 1 to 24 h using moving windows with 1 km and 1 h jumps.

3.2 Comparison between Intensity–Duration–Frequency maps

20   IDF maps obtained from the satellite precipitation datasets (HRC and CHRC) are compared to the ones obtained from the radar archive during corresponding years (16 years: 1998–2013). The comparison is extended over an analysis domain defined excluding the pixels that are known to be not reliable. In particular, pixels located closer than 27 km or farther than 185 km from the radar or behind the hilly region are excluded due to insufficient reliability of the radar data, and pixels located in proximity of major lakes are excluded due to false rainfall signals in the CMORPH ~~rainfall~~ estimates (e.g. Guo et al., 2015).

25   The number of pixels analysed for each climatic region is reported in Table 1. A limited number of cases with problematic data, providing unrealistic outcomes, could still be found in all the products. In order to avoid single problematic situations dominating the results, we analysed the distribution of the GEV parameters over the whole study area and excluded from the comparison the pixels for which any of them was outside of the 1–99th quantiles interval of the corresponding distribution.

Three widely used non-dimensional, normalized metrics are selected to compare radar-IDF and satellite-IDF maps: Correlation

30   Coefficient (CC), measuring the spatial correlation of the maps; multiplicative bias (Bias), measuring the mean quantitative agreement of the maps; and Normalized Standard Difference (NSD), measuring the variability of the residuals of the normalized maps. Additional information on the metrics is provided in Appendix B.

3.3 Uncertainty related to the record length

We assumed the records of rain gauge data to be a complete sample of the climatology of extremes for return periods comparable to the remotely sensed data record length. Synthetic records of rain gauge data were created by randomly sampling years, without replacement, from the full rain gauge record, and the corresponding IDF values were calculated. We focused on synthetic records of 10, 15, 20 and 25 years and bootstrapped the operation 999 times for each rain gauge and for each synthetic record length. The 5–95th quantile interval of the obtained distributions was used to measure of the uncertainty related to the record length (Overeem et al., 2008).

4 Results and discussion

4.1 Comparison between satellite- and radar- derived GEV parameters

The distribution of the GEV parameters (25–75th quantile intervals and median over the whole study area) derived from radar and satellite datasets in the three climatic regions and over the Mediterranean Sea (sea, from here on) are presented in Fig. 2. We recall here that the scale, location and shape parameters provide a measure of the mean, dispersion and skewness of the underlying distribution, respectively. Scale and location Location parameters from HRC (CHRC) estimates are smaller (larger) than the ones from radar over Mediterranean climate and over the sea, meaning that extreme values from HRC (CHRC) are in general lower (higher) than radar extreme values while in semiarid and arid climates HRC and CHRC generally identify higher parameters than the radar (i.e. higher extreme values). Differences in the location parameters can be associated to the bias between extreme values in the datasets. The scale parameters are normalized over the corresponding location parameters to appreciate the relative differences. Normalized scale parameters from HRC and CHRC are similar and lower than the ones derived from radar. Normalized scale parameters, together with their variability, tend to increase when moving from sea to Mediterranean, semiarid and arid climates. The drier climate, the larger the dispersion of the GEV distribution. A slight increase of the normalized scale parameters with duration can be noticed in the HRC/CHRC data. On semiarid and arid climates radar generally identifies lower values.

The shape parameters are mostly greater than zero, suggesting, in line with previous studies (Katz et al., 2002; Papalexiou and Koutsoyiannis, 2013), that type II extreme values distribution should be considered for the area. Both radar and satellite products derive high shape parameters for arid climate that decrease when moving to semiarid and to Mediterranean climates, thus suggesting that drier regions are characterized by thicker tailed distributions. Over the sea, satellite shape parameters are close to zero, while radar shape parameters tend to be higher. Increasing duration, radar-derived parameters are decreasing, while satellite-derived parameters tend to increase, becoming comparable for durations longer than ~6 h. In general, the shape parameters derived from radar are higher than the ones derived from CMORPH so that, in arid climate, the radar dataset includes few cases with a shape parameter greater than 0.5 and, in single occasions, even greater than 1. This is likely associated to the measurement uncertainty of radar estimates and to the record intermittency due to radar shutdowns (Marra and Morin,

2015). In fact, the rainfall regime of arid climate is characterized by short and intense events (Karklinsky and Morin, 2006), so that gaps in the record can cause the missing of important storms leading to particularly low estimates of some annual maxima and, consequently, to thicker-tailed distributions. It should also be noted that missing of short duration extremes by CMORPH due to the overpasses frequency of microwave satellites could contribute to the differences observed between

5 CMORPH and radar.

4.1.1 Effect of spatial and temporal aggregation

The location and scale parameters~~, as expected,~~ consistently decrease~~d~~ as the spatial and temporal aggregation scales ~~are~~ increased. This is an expected effect, caused by the smoothing of rainfall fields operated by the spatial averaging, is expected, ~~and~~ therefore results are not reported ~~here~~in this paper. ~~Conversely, It~~it is ~~thus~~ interesting to analyse the shape parameter~~s~~.

10 The distributions of the shape parameters derived from the full radar record aggregated on grid sizes increasing from 2×2 to 64×64 km^2 and on durations increasing from 1 to 24 h are presented in Fig. 3. The shape parameters consistently decrease both with spatial and temporal aggregation. This means that the smoothing effect due to the spatial and temporal aggregation of rainfall measurement depends on the return period, and is more pronounced for longer return periods. This relates to a non-homogeneity of the scales of rainfall extremes with return period: the more extreme an event is (i.e. the longer its return period),

15 the more localized it is expected to be in both space and time. ~~Moreover, the magnitude~~ Magnitude and ~~the~~ combination of spatial and temporal effects are shown to depend on the climate, confirming that the spatiotemporal scales of rainfall extremes are highly dependent on the climatic conditions.

4.2 Satellite-IDF and radar-IDF

A visual comparison between IDF curves derived from rain gauges and from the co-located radar and satellite pixels is

20 presented in Fig. 4. Corresponding data periods (16 yr, 1 h block) are used over the three climatic regions (Mediterranean–En Hahoresh, semiarid–Beer Sheva and arid–Sedom; Fig. 1) for 1, 6 and 24 h durations. The reported cases, discussed in Marra and Morin (2015), are known to have good radar visibility, with the exception of Sedom that, being farther from the radar and behind the hilly region, can be subject to overshooting of the radar beam. Radar reproduces better the skewness of the IDF curves and radar-IDFs are, with the exception of the arid case, within the rain gauge confidence interval. Note that these results

25 represent the local scale; while interpreting them, one should take into account the different scales of rain gauges (point scale) and remote sensing datasets (~8×8 km^2, in this case) and the natural variability that extreme rainfall presents when relatively short records are used (Gires et al., 2014; Peleg et al., 2016a).

Examples of IDF maps for 3 h duration from HRC, CHRC and radar products are presented in Fig. 5. It is shown that the spatial variability of IDF values increases with return period. Noteworthy, HRC-IDFs are lower than the corresponding CHRC-

30 and radar-IDFs, while the CHRC-IDFs seem to be larger than radar-IDFs for 2 years return period and comparable for 25 years return period. A quantitative analysis of the differences between HRC/CHRC-IDF maps and radar-IDF maps is provided in the next section.

4.3 Comparison between satellite-IDF and radar-IDF

A comprehensive quantitative comparison between radar-IDF and CMORPH-IDF maps is presented in Fig. 6. In general, very similar patterns of CC and NSD are observed for HRC and CHRC while substantially different Bias patterns are observed. This confirms the point made by Mei et al. (2016) that daily gauge adjustment influences the magnitude rather than the space-time organization of annual extremes.

The CC, measuring the spatial correlation of the IDF maps, is as high as ~0.70–0.76 for 2 and 5 years return periods and decreases when longer return periods are examined, especially for shorter durations (~0.27–0.29 for 1 h, ~0.40–0.55 for 24 h). CHRC has higher CC with radar than HRC. Bias results show that CHRC tends to overestimate while HRC tends to underestimate relative to radar IDFs. For short durations, the Bias for both satellite datasets (gauge adjusted and unadjusted) strongly decreases with increasing return period. As the duration increases, this trend diminishes so that, for longer durations, the Bias is almost consistent for all return periods. This reflects what observed for the shape parameters: the higher shape parameters estimated from radar for short durations lead the radar IDF curves to be thicker-tailed, i.e. to predict larger intensities for longer return periods. Moreover, this suggests that, for sufficiently long durations, it is possible to identify a factor linking radar-IDFs to satellite-IDFs (~0.7 for HRC, ~1.2 for CHRC at 24 h). The NSD patterns are very similar and tend to increase with return period (up to ~0.75 for 25 years, 1 h) and to decrease with duration (~0.55–0.62 for 25 years, 24 h).

Figure 7 and 8 show the metrics calculated between HRC/CHRC-IDF and radar-IDF maps for the three climatic regions and over the sea. The spatial correlation between radar- and CHRC-IDF maps is larger, with respect to HRC-IDF maps, for all the climatic regions and is very high over arid areas/short durations and semiarid areas/long durations. On the contrary, CC between HRC-IDF and radar-IDF maps is very low for long durations (~0.05 for 25 years, 24h), probably due to the very low values measured by CMORPH when no gauge adjustment is applied.

Note that the CC between both HRC/CHRC-IDFs and radar-IDFs is low over the sea, especially for shorter durations, and that the difference becomes less important for longer durations. This is not coming as a surprise since no gauge data is available for the adjustment of satellite or radar data over the sea. As pointed out above, gauge adjustment is only weakly impacting the space-time organization of CMORPH extreme estimates, while it is a crucial step in radar quantitative precipitation estimation. This observation, together with the increased reliability of satellite based estimations over the sea (Kidd and Levizzani, 2011), suggests that spatial distribution of IDF values indicated by satellite products should be considered more accurate.

Overestimation of CHRC-IDF maps with respect to radar-IDF maps is more marked for Mediterranean and semiarid climates, with a factor ~1–1.6. For durations longer than 3 h, the Bias for Mediterranean and semiarid climates for both HRC and CHRC shows almost no trend with return period, meaning that the trend observed in the general case is due to arid climate and sea area, where the differences between radar- and HRC/CHRC- shape parameters is larger (Fig. 2). This confirms that radar and HRC/CHRC almost agree in the identification of the skewness of IDF curves over Mediterranean and semiarid climates for durations larger than 3h. NSD over arid and semiarid areas is larger than for Mediterranean climates and sea.

4.4 Uncertainty related to the record length

In this section, we present the results of the bootstrap sampling of long rain gauge records used to quantify the uncertainty related to the record length of remote sensing datasets. The uncertainty presented here is the component related to the under-sampling of rainfall climatology due to the use of short data records and is quantified as the 5–95th quantile interval of the bootstrap sampling. The uncertainty for two example cases is presented in Fig. 9. ~~The uncertainty related to the use of short record lengths is presented, for two example cases, in Fig. 9.~~ The figure reports the 5–95th quantile interval of the bootstrap sampling of 10, 15, 20 and 25 years of data out of the whole record of the three cases shown in Fig. 4. As expected, uncertainties become important when the record length is similar, or smaller, than the estimated return period. Arid climates are characterized by larger uncertainties, especially when short records and short durations are examined; this is probably due to the low number of rain events per year, and to the thicker-tail characteristic of arid IDF curves. Short records are shown to be more likely overestimating, rather than underestimating, the IDF values.

Figure 10 shows the relative uncertainty (width of the 5–95th quantile interval of the 999 bootstrap sampling repetitions normalized over the long record rain gauge-IDF value) as a function of the ratio between return period and record length. Uncertainties for 1 h and 3 h durations are comparable, with the uncertainty for 3 h duration being smaller. This suggests that time aggregation potentially decreases part of the issues related to the use of short records. Uncertainty is larger for return periods longer than the record length, especially for short durations and drier climates. For the Mediterranean climate, the uncertainty is generally lower than 50% when the return period is shorter than the record length and reaches up to ~125% for return periods more than twice the data record length. In the semiarid climate, the uncertainty is shown to be larger, exceeding 100% (30–100%) even for return periods comparable to the data record length and exceeding 200% for return periods more than twice the data record length. In the arid climate, the uncertainty is even larger, reaching 150% for return periods equal to the data record length and exceeding 250% for return periods more than twice the data record length.

5 Conclusions

This study compared the use of rainfall estimates from a ground based C-band weather radar and from a high-resolution satellite precipitation product, CMORPH (HRC) and its gauge-adjusted version (CHRC), for the identification of Intensity–Duration–Frequency curves. IDF curves were computed using the above products over the Eastern Mediterranean (Mediterranean, semiarid and arid climates and over the sea) and the uncertainties due to the limited record length of the remote sensing datasets were quantified basing on long records of rain gauge measurements. Our findings can be summarized as follows:

1. The shape parameters of the generalized extreme values distribution, as derived from radar, HRC and CHRC, are mostly greater than zero; drier climates are characterized by higher shape parameters, suggesting that thicker-tail distributions better describe rainfall extremes of drier areas. In general, the shape parameters derived from radar are higher than the ones from CMORPH, especially for arid climate and over the sea.

2. The shape parameter tends to decrease when rainfall estimates are aggregated in space and/or time. The effect is related to a non-homogeneity of spatial and temporal scales of rainfall extremes with return period. This non-homogeneity depends on the climatic conditions.

3. The spatial correlation coefficient between corresponding radar-IDF and HRC/CHRC-IDF maps is between 0.70 and 0.76 for short return periods, but decreases with increase in return period, especially for short durations. In general, for both HRC and CHRC the correlation is higher in arid climate for durations up to 3 h and in Mediterranean and semiarid climates for longer durations (6–24 h). Low correlations are observed over the sea.

4. HRC-IDFs and CHRC-IDFs are, respectively, lower and higher than radar-IDFs. In both cases the observed Bias decreases with return period, especially for short durations and arid climate. For longer durations and Mediterranean/semiarid climates, the decreasing trend almost disappears so that the Bias can be considered independent from the return period (~0.7 for HRC, ~1.2 for CHRC for 24 h).

4.5. Comparison of HRC-IDF and CHRC-IDF against radar-IDF show consistent patterns of correlation and dispersion, and different biases. This means that gauge-adjustment influences the magnitude rather than the space-time organization of annual extremes and suggests that HRC-IDF can potentially be used to estimate the frequencies of CMORPH estimates in near real time early warning systems.

5.6. The uncertainty related to the use of short records becomes important when the record length is shorter or comparable to the examined return period. This is particularly true for drier climates and shorter durations, with potential uncertainty of ~50%, ~100% and ~150% for return periods comparable to the record length in Mediterranean, semiarid and arid climates, respectively.

Rainfall frequency analysis by means of remote sensing rainfall estimates remains a challenging task, especially when dry climates are explored. Nevertheless, the agreement between IDF curves derived from different sensors on Mediterranean and, to a good extent, semiarid climates, demonstrates their potential for the description of small scale spatial patterns of IDF curves and instils confidence on their quantitative use for ungauged areas of the Earth. Spatial and temporal aggregation of rainfall information represent viable ways to take advantage of remote sensing datasets and decrease the uncertainties related to the derived IDF curves. In particular, remote sensed rainfall archives can provide important information when 2–10 years return periods and 12–24 h durations are requested, scales that are relevant for both flood risk management (e.g. issuing of warning) and hydrological design (e.g. sewer systems design, large scale drainage planning).

Appendix A. GEV distribution

The GEV cumulative distribution function can be written as (Coles, 2001):

$$
\begin{cases}
F(I; \mu, \sigma, \kappa) = \exp\left\{-\left[1 + \frac{\kappa}{\sigma}(I - \mu)\right]^{-\frac{1}{\kappa}}\right\} & \text{for } \kappa \neq 0 \\
F(I; \mu, \sigma, \kappa) = \exp\left\{-\exp\left[-\frac{1}{\sigma}(I - \mu)\right]\right\} & \text{for } \kappa = 0
\end{cases}
\tag{A.1}
$$

22

where I is the average intensity over a given duration and μ, σ and κ are respectively the location, scale and shape parameters of the distribution. The shape parameter is directly related to the tail of the extreme values distribution: low (high) shape parameters are related to lower (higher) probability of having large extremes. When the shape parameter is lower than 0 (type III distribution), an upper limit to the extreme values is expected. When the shape parameter is greater (type II) or equal (type I distribution) to 0, no upper limit to the values is expected. In particular, when the shape parameter is greater than 0.5, the standard deviation of the distribution is infinite and when it is greater than 1, the mean of the distribution is infinite. Under hypotheses on the regularity of the tail of the distribution, the Fisher-Tippet theorem demonstrates that GEV distribution is the only possible limit distribution for the extreme values of independent and identically distributed random variables.

Appendix B. Comparison metrics

Correlation Coefficient (CC) measures the spatial correlation of the derived maps. It is calculated as:

$$CC = \frac{\text{Cov}(\boldsymbol{r}, \boldsymbol{s})}{\text{StdDev}(\boldsymbol{r}) \cdot \text{StdDev}(\boldsymbol{s})} \tag{B.1}$$

where r and s refer to the radar-IDF maps and satellite-IDF maps (HRC or CHRC), respectively. It ranges between -1 and 1 with positive (negative) values indicating positive (negative) correlation.

Multiplicative *Bias* measures the mean quantitative agreement of two maps. It is calculated as:

$$\text{Bias} = \frac{\frac{1}{N}\sum_i^N s_i}{\frac{1}{N}\sum_i^N r_i} \tag{B.2}$$

where r_i and s_i are the radar-IDF and satellite-IDF estimates respectively, and i is the index of a pixel within the N pixels composing the maps. It may assume any positive value with 1 indicating perfect mean quantitative agreement, value higher (lower) than 1 indicating that the mean satellite estimate is higher (lower) than the mean radar estimate.

Normalized Standard Difference (NSD) measures the standard deviation of the residuals of the normalized maps and is calculated as:

$$\text{NSD} = \sqrt{\frac{1}{N}\sum_i^N \left(\frac{s_i}{\hat{s}} - \frac{r_i}{\hat{r}}\right)^2} \tag{B.3}$$

where $\hat{s}$ and $\hat{r}$ are the mean values of satellite and radar maps. NSD may assume any value ≥ 0.

Acknowledgments

References

Alila, Y., 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. J. Geophys. Res. 104 (D24), 31645–31655. http://dx.doi.org/10.1029/1999JD900764

Allen, R.J., DeGaetano, A.T., 2005. Considerations for the use of radar-derived precipitation estimates in determining return intervals for extreme areal precipitation amounts. J. Hydrol. 315 (1–4), 203–219. http://dx.doi.org/10.1016/j.jhydrol.2005.03.028.

Alpert, P., Shafir, H., 1989. Meso c-scale distribution of orographic precipitation: numerical study and comparison with precipitation derived from radar measurements. J. Appl. Meteor. 28, 1105–1117. http://dx.doi.org/10.1175/1520-0450(1989)028<1105:MSDOOP>2.0.CO;2

Amitai, E., Petersen, W.A., Llort, X., Vasilof, S., 2011. Multi-platform comparisons of rain intensity for extreme precipitation events. IEEE Trans. Geosci. Remote Sens. 99, 1–12. http://dx.doi.org/10.1109/TGRS.2011.2162737

Awadallah, A., ElGamal, M., ElMostafa, A., ElBadry, H., 2011. Developing intensity–duration–frequency curves in scarce data region: an approach using regional analysis and satellite data. Engineering 3 (3), 215–226. http://dx.doi.org/10.4236/eng.2011.33025

Ben-Zvi, A., 2009. Rainfall Intensity-Duration-Frequency Relationships Derived from Large Partial Duration Series. J. Hydrol. 367 (1–2), 104–14. http://dx.doi.org/10.1016/j.jhydrol.2009.01.007

Berne, A., Krajewski, W.F., 2013. Radar for hydrology: unfulfilled promise or unrecognized potential? Adv. Water Resour. 51, 357–366. http://dx.doi.org/10.1016/j.advwatres.2012.05.005

Borga, M., Anagnostou, E. N. , Blöschl, G., Creutin., J.D., 2011. Flash Flood Forecasting, Warning and Risk Management: The HYDRATE Project. Environmental Science and Policy 14 (7), 834–44

Borga M., M. Stoffel, L. Marchi, F. Marra, M. Jacob 2014: Hydrogeomorphic response to extreme rainfall in headwater systems: flash floods and debris flows. J. Hydrol., 518 (2014), 194–205. http://dx.doi.org/10.1016/j.jhydrol.2014.05.022

Ceresetti, D., Ursu, E., Carreau, J., Anquetin, A., Creutin, J.D., Gardes, L., Girard, S., Molinié, D., 2012. Evaluation of classical spatial-analysis schemes of extreme rainfall. Nat. Hazards Earth Syst. Sci. 12, 3229–3240. http://dx.doi.org/10.5194/nhess-12-3229-2012.

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

Chen, S., Hong, Y., Cao, Q., Kirstetter, P.E., Gourley, J.J., Qi, Y., Zhang, J., Howard, K., Hu, J., Wang, J., 2013. Performance evaluation of radar and satellite rainfalls for Typhoon Morakot over Taiwan: are remote-sensing products ready for gauge denial scenario of extreme events? J. Hydrol. 506, 4–13. http://dx.doi.org/10.1016/j.jhydrol.2012.12.026

Coles, S., 2001. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London.

Chow, V.T., Maidment, D.R., Mays, L.W., 1988. Applied Hydrology. McGraw-Hill.

Dayan, U., Morin, E., 2006. Flash flood–producing rainstorms over the Dead Sea: a review. Geol. Soc. Am. Special Papers 401, 53–62. http://dx.doi.org/10.1130/2006.2401(04)

Dayan, U., Ziv, B., Margalit, A., Morin, E., Sharon, D., 2001. A severe autumn storm over the middle-east: synoptic and mesoscale convection analysis. Theor. Appl. Climatol. 69, 103–122. http://dx.doi.org/10.1007/s007040170038

Dayan, U., Nissen, K., and Ulbrich, U., 2015. Review Article: Atmospheric conditions inducing extreme precipitation over the eastern and western Mediterranean. Natural Hazards and Earth System Sciences, 15(11), 2525–2544. http://doi.org/10.5194/nhess-15-2525-2015

Eagleson, P.S., 1970. Dynamic Hydrology. McGraw-Hill, New York, p. 462

Eldardiry, H., Habib, E., Zhang, Y., 2015. On the use of radar-based quantitative precipitation estimates for precipitation frequency analysis. J. Hydrol. 531 (Part 2), 441–453. http://dx.doi.org/10.1016/j.jhydrol.2015.05.016.

Endreny, T.A., Imbeah, N., 2009. Generating robust rainfall intensity–duration–frequency estimates with short-record satellite data. J. Hydrol. 371 (2009), 182–191. http://dx.doi.org/10.1016/j.jhydrol.2009.03.027

Ferraro, R. R., 1997: SSM/I derived global rainfall estimates for climatological applications. J. Geophys. Res., 102, 16715-16735.

Ferraro, R. R., F. Weng, N. C. Grody and L. Zhao, 2000: Precipitation characteristics over land from the NOAA-15 AMSU sensor. Geophys. Res. Ltr., 27, 2669-2672.

Fowler, H.J., Kilsby, C.G., 2003. A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. Int. J. Climatol. 23, 1313–1334. http://dx.doi.org/10.1002/joc.943

Gellens, D., 2002. Combining regional approach and data extension procedure for assessing GEV distribution of extreme precipitation in Belgium. J. Hydrol. 268, 113–126. http://dx.doi.org/10.1016/S0022-1694(02)00160-9.

Gires, A., Tchiguirinskaia I., Schertzer D., Schellart A., Berne A., and Lovejoy S., 2014. Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data. Atmos. Res. 138, 125-138

Goldreich, Y., 1994. The spatial distribution of annual rainfall in Israel – a review. Theor. Appl. Climatol. 50, 45–59. http://dx.doi.org/10.1007/BF00864902

Goldreich, Y., 2003. The Climate of Israel, Observations, Research and Applications. Springer, New York, p. 298

Guo, H., Chen, S., Bao, A., Hu, J., Gebregiorgis, A.S., Xue, X., Zhang, X, 2015. Inter-Comparison of High-Resolution Satellite Precipitation Products over Central Asia. Remote Sens. 2015, 7, 7181-7211. http://dx.doi.org/10.3390/rs70607181

Joyce, R. J., J. E. Janowiak, P. A. Arkin, P. Xie, 2004. CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. J. Hydrometeorol. 5, 487–503. http://dx.doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Kahana, R., Ziv, B., Enzel, Y., Dayan, U., 2002. Synoptic climatology of major floods in the Negev Desert, Israel. Int. J.

5   Climatol. 22, 867–882. http://dx.doi.org/10.1002/joc.766.

Karklinsky, M., & Morin, E. (2006). Spatial characteristics of radar-derived convective rain cells over southern Israel. Meteorologische Zeitschrift, 15(5), 513-520. http://dx.doi.org/10.1127/0941-2948/2006/0153

Katz, R. W., M. B. Parlange, and P. Naveau (2002) Statistics of extremes in hydrology. Adv. Water Resour., 25, 1287–1304

Kidd, C., A. Becker, G. Huffman, C. Muller, P. Joe, G. Skofronick-Jackson, and D. Kirshbaum, 2016: So, how much of the

10   Earth's surface is covered by rain gauges? Bull. Amer. Meteor. Soc., in press. http://dx.doi.org/10.1175/BAMS-D-14-00283.1

Kidd, C., & Levizzani, V., 2011: Status of satellite precipitation retrievals. Hydrol. Earth Syst. Sci., 15, 1109-1116. http://dx.doi.org/10.5194/hess-15-1109-2011

Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrol. Sci. J. 49 (4), 591–610. http://dx.doi.org/10.1623/hysj.49.4.575.54430

15 Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D-B Shin, and T. T. Wilheit, 2001: Evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. J. Appl. Meteor., 40, 1801-1820.

Kysely, J., Picek, J., 2007. Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. Climate Res. 33, 243–255

20 Marra, F. and E. Morin 2015: Use of radar QPE for the derivation of Intensity–Duration–Frequency curves in a range of climatic regimes. J. Hydrol. 531 (2015) 427–440. http://dx.doi.org/10.1016/j.jhydrol.2015.08.064

Marra, F., E.I. Nikolopoulos, J.D. Creutin, M. Borga 2014. Radar rainfall estimation for the identification of debris-flow occurrence thresholds. J. Hydrol. 519 (2014) 1607–1619. http://dx.doi.org/10.1016/j.jhydrol.2014.09.039

Marra, F., E.I. Nikolopoulos, J.D. Creutin, M. Borga 2016: Space–time organization of debris flows-triggering rainfall and its

25 effect on the identification of the rainfall threshold relationship. Journal of Hydrology, 541 (2016) 246–255. http://dx.doi.org/10.1016/j.jhydrol.2015.10.010

Mei, Y., E.I. Nikolopoulos, E.N. Anagnostou, D. Zoccatelli, M. Borga, 2016. Error analysis of satellite precipitation-driven modeling of flood events in complex alpine terrain, Remote Sens., 8, p. 293. http://dx.doi.org/10.3390/rs8040293

Morin, E., Gabella, M., 2007. Radar-based quantitative precipitation estimation over Mediterranean and dry climate regimes.

30 J. Geophys. Res. 112, D20108. http://dx.doi.org/10.1029/2006JD008206

Morin, E., Enzel, Y., Shamir, U., Garti, R., 2001. The characteristic time scale for basin hydrological response using radar data. J. Hydrol. 252, 85–99. http://dx.doi.org/10.1016/S0022-1694(01)00451-6

Morin, E., Jacoby, Y., Navon, S., Ben-Halachmi, E., 2009. Towards flash-flood prediction in the dry Dead Sea region utilizing radar rainfall information. Adv. Wat. Resour. 32, 1066–1076. http://dx.doi.org/10.1016/j.advwatres.2008.11.011

Morin, J., Rosenfeld, D., Amitai, E., 1995. Radar rain field evaluation and possible use of its high temporal and spatial resolution for hydrological purposes. J. Hydrol. 172, 275–292. http://dx.doi.org/10.1016/0022-1694(95)02700-Y

Nastos P.T., J. Kapsomenakis, K.C. Douvis, 2013. Analysis of precipitation extremes based on satellite and high-resolution gridded data set over Mediterranean basin. Atmos. Res. 131 (2013) 46–59.

5 Overeem, A., Buishand, A., Holleman, I., 2008. Rainfall Depth-Duration-Frequency Curves and Their Uncertainties. J. Hydrol. 348(1–2): 124–34. http://dx.doi.org/10.1016/j.jhydrol.2007.09.044

Overeem, A., Buishand, T.A., Holleman, I., 2009. Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. Water Resour. Res. 45(10), W10424. http://dx.doi.org/10.1029/2009WR007869

Paixao, E., M. M. Q. Mirza, M. W. Shephard, H. Auld, J. Klaassen and G. Smith, 2015. An integrated approach for identifying
10 homogeneous regions of extreme rainfall events and estimating IDF curves in Southern Ontario, Canada: incorporating radar observations. J. Hydrol. 528, 734–750. http://dx. doi.org/10.1016/j.jhydrol.2015.06.015

Panziera L., M. Gabella, S. Zanini, A. Hering, U. Germann and A. Berne, 2016. A radar-based regional extreme rainfall analysis to derive the thresholds for a novel automatic alert system in Switzerland. Hydrol. Earth Syst. Sci., 20, 2317–2332. http://dx.doi.org/10.5194/hess-20-2317-2016

15 Papalexiou, S. M. and D. Koutsoyiannis, 2013. Battle of Extreme Value Distributions: A Global Survey on Extreme Daily Rainfall. Water Resour. Res. 49(1): 187–201. http://dx.doi.org/10.1029/2012WR012557

Peel M.C., B.L. Finlayson and T.A. McMahon, 2007. Updated world map of the Koppen-Geiger climate classification. Hydrol. Earth Syst. Sci., 11, 1633–1644, www.hydrol-earth-syst-sci.net/11/1633/2007/

Peleg, N., Morin, E., 2012. Convective rain cells: radar-derived spatiotemporal characteristics and synoptic patterns over the
20 eastern Mediterranean. J. Geophys. Res. 117, D15116. http://dx.doi.org/10.1029/2011JD017353

Peleg, N., Ben-Asher, M., Morin, E., 2013. Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network. Hydrol. Earth Syst. Sci. 17, 2195–2208. http://dx.doi.org/10.5194/hess-17-2195-2013

Peleg, N., F. Marra, S. Fatichi, A. Paschalis, P. Molnar, P. Burlando, 2016a: Spatial variability of rainfall at radar subpixel
25 scale. J. Hydrol., in press. http://dx.doi.org/10.1016/j.jhydrol.2016.05.033

Peleg, N., Blumensaat, F., Molnar, P., Fatichi, S., and Burlando, P., 2016b: Partitioning spatial and temporal rainfall variability in urban drainage modelling, Hydrol. Earth Syst. Sci. Discuss., http://dx.doi.org/10.5194/hess-2016-530, in review.

Perica, S. et al., 2013. Precipitation-Frequency Atlas of the United States, NOAA Atlas 14, Volume 9, Version 2: Southeastern States, Silver Spring, Maryland, 2013: U.S. Department of Commerce, National Oceanic and Atmospheric Adminstration,
30 National Weather Service.

Phillips, P., and P. Perron., 1988. Testing for a Unit Root in Time Series Regression. Biometrika, 75, 335–346.

Rubin, S., Ziv, B., Paldor, N., 2007. Tropical plumes over eastern North Africa as a source of rain in the Middle East. Mon. Weath. Rev. 135 (12), 4135–4148. http://dx.doi.org/10.1175/2007MWR1919.1

Schaefer, M.G., 1990. Regional analyses of precipitation annual maxima in Washington State. Water Resour. Res. 26 (1), 119–131. http://dx.doi.org/10.1029/WR026i001p00119

Segoni, S., Battistini, A., Rossi, G., Rosi, A., Lagomarsino, D., Catani, F., Moretti, S., and Casagli, N.: Technical Note: An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds, Nat. Hazards Earth Syst. Sci., 15, 853-861, doi:10.5194/nhess-15-853-2015, 2015.

Srebro, Haim, and T. Soffer. The New Atlas of Israel: The National Atlas. Jerusalem: Survey of Israel and The Hebrew University of Jerusalem, 2011.

Sivapalan, M., Bloschl, G., 1998. Transformation of point rainfall to areal rainfall: intensity–duration–frequency curves. J. Hydrol. 204, 150–167. http://dx.doi.org/10.1016/S0022-1694(97)00117-0.

Stampoulis, D., E. N. Anagnostou, and E. I. Nikolopoulos, 2013. Assessment of high-resolution satellite-based rainfall estimates over the Mediterranean during heavy precipitation events. J. Hydrometeorol. 14(5), 1500-1514. http://dx.doi.org/10.1175/JHM-D-12-0167.1

Svensson, C., and D. A. Jones. 2010. Review of Methods for Deriving Areal Reduction Factors. J. Flood Risk Management 3 3: 232–45. http://dx.doi.org/10.1111/j.1753-318X.2010.01075.x

Tapiador, F.J., Turk, F.J., Petersen, W., Hou, A.Y., García-Ortega, E., Machado, L.A.T., Angelis, C.F., Salio, P., Kidd, C., Huffman, G.J., de Castro, M., 2012. Global precipitation measurement: methods, datasets and applications. Atmos. Res. 104–105, 70–97. http://dx.doi.org/10.1016/j.atmosres.2011.10.021

Tiranti, D., R. Cremonini, F. Marco, A. R. Gaeta and S. Barbero 2014. "The DEFENSE (Debris Flows triggEred by Storms - Nowcasting System): An Early Warning System for Torrential Processes by Radar Storm Tracking Using a Geographic Information System (GIS)." Computers and Geosciences 70: 96–109. http://dx.doi.org/10.1016/j.cageo.2014.05.004

Tubi, A., U. Dayan, 2014: Tropical Plumes over the Middle East: Climatology and synoptic conditions. Atmos. Res. 145-146, 168-181. http://dx.doi.org/10.1016/j.atmosres.2014.03.028

Villarini, G., W. Krajewski, A. A. Ntelekos, K. P. Georgakakos and J. A. Smith, 2010. Towards Probabilistic Forecasting of Flash Floods: The Combined Effects of Uncertainty in Radar-Rainfall and Flash Flood Guidance. J. Hydrol. 394(1–2): 275–84. http://dx.doi.org/10.1016/j.jhydrol.2010.02.014

de Vries, A.J., Tyrlis, E., Edry, D., Krichak, S.O., Steil, B., Lelieveld, J., 2013. Extreme precipitation events in the Middle East: dynamics of the Active Red Sea Trough. J. Geophys. Res. Atmos. 118, 7087–7108. http://dx.doi.org/10.1002/jgrd.50569

Watt, E., J. Marsalek, 2013: Critical review of the evolution of the design storm event concept. Can. J. Civ. Eng., 40, 105–113. http://doi.org/10.1139/cjce-2011-0594

Wright, D. B., Smith, J. A., Villarini, G., and Baeck, M. L., 2013. Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. Journal J. of Hydrol.ogy, 488, 150–165. http://doi.org/10.1016/j.jhydrol.2013.03.003

Xie, P., S.-H. Yoo, R. J. Joyce, and Y. Yarosh (2011), Bias-corrected CMORPH: A 13-year analysis of high-resolution global precipitation, EGU General Assembly, Eur. Geosci. Union, Vienna, Austria, 4 – 8 April.

Yakir H. and E. Morin, 2011. Hydrologic response of a semi-arid watershed to spatial and temporal characteristics of convective rain cells. Hydrol. Earth Syst. Sci., 15, 393–404. http://dx.doi.org/10.5194/hess-15-393-2011.

Yang, T.-H., Hwang, G.-D., Tsai, C.-C. and Ho, J.-Y.: Using rainfall thresholds and ensemble precipitation forecasts to issue and improve urban inundation alerts. Hydrol. Earth Syst. Sci., 20, 4731-4745, http://dx.doi.org/10.5194/hess-20-4731-2016, 2016

5

Table 1: Koppen-Geiger classification (Peel et al., 2007; Srebro et al., 2011) and number of pixels analyzed for each climatic region. Note that: MAT = Mean Annual Temperature; MAP = Mean Annual Precipitation; T_{hot} = Temperature of the hottest month; T_{cold} = Temperature of the coldest month; P_{ds} = Precipitation of the driest month in summer; P_{ww} = Precipitation of the wettest month in winter.

Climatic region	Koppen-Geiger definition	Koppen-Geiger criteria	Number of pixels (fraction of the total)
Mediterranean	Temperate, Dry summer, Hot summer (Csa)	$T_{hot} \geq 22°C$ $0°C < T_{cold} < 18°C$ $P_{ds} < 40$ mm $P_{ds} < P_{ww} / 3$	318 (20.1%)
Semiarid	Arid, steppe, hot (BSh)	$10 \times MAT \leq MAP \leq 20 \times MAT$ $MAT \geq 18°C$	404 (25.5%)
Arid	Arid, desert, hot (BWh)	$MAP < 10 \times MAT$ $MAT \geq 18°C$	225 (14.2%)
Sea	-	-	635 (40.1%)

5
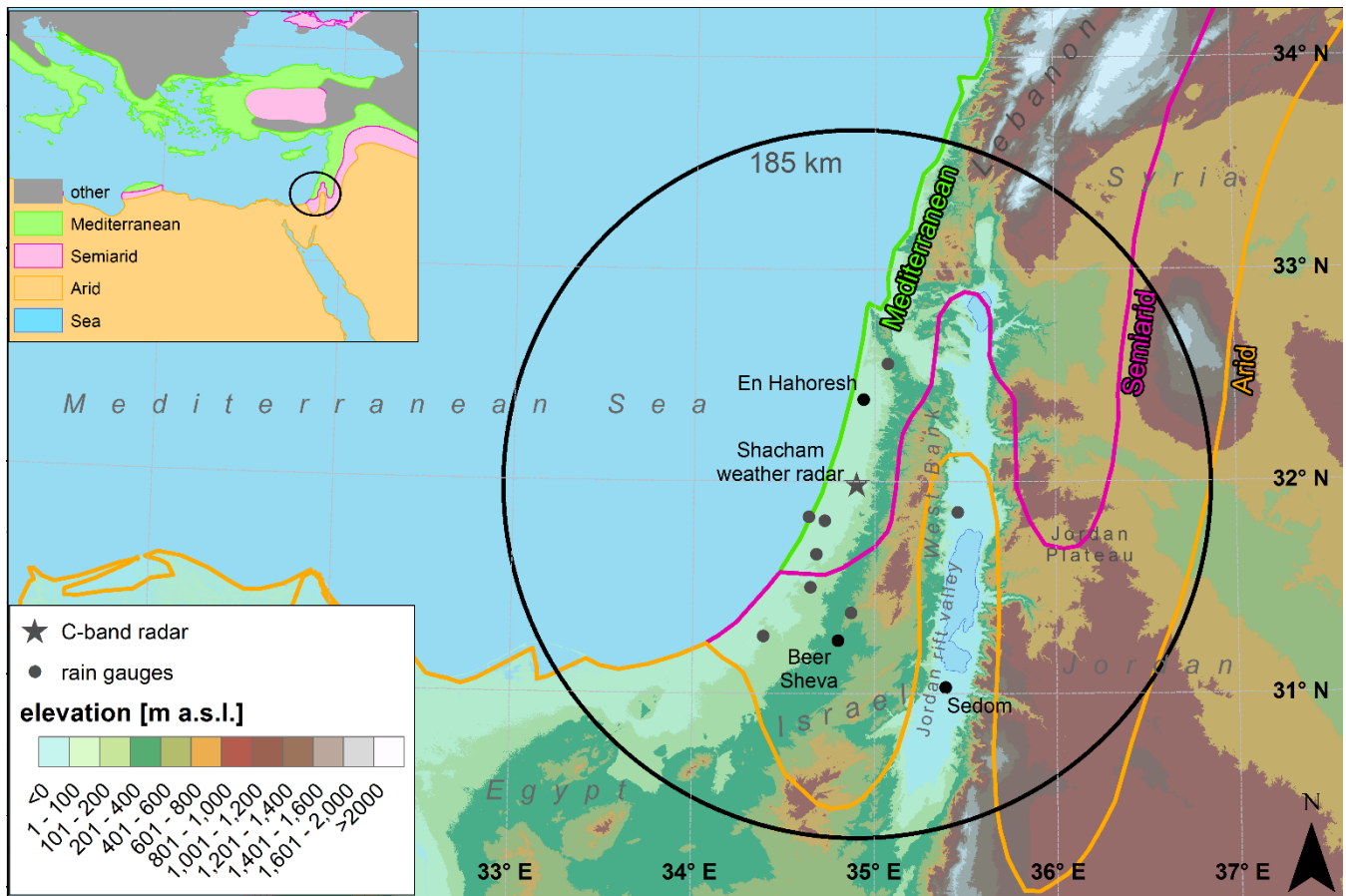
Figure 1: Map of the study area showing terrain elevation, climatic classification from Shahar et al. (2008), location of the radar and rain gauges used in the study. The small map shows the location of the study area within the Eastern Mediterranean.
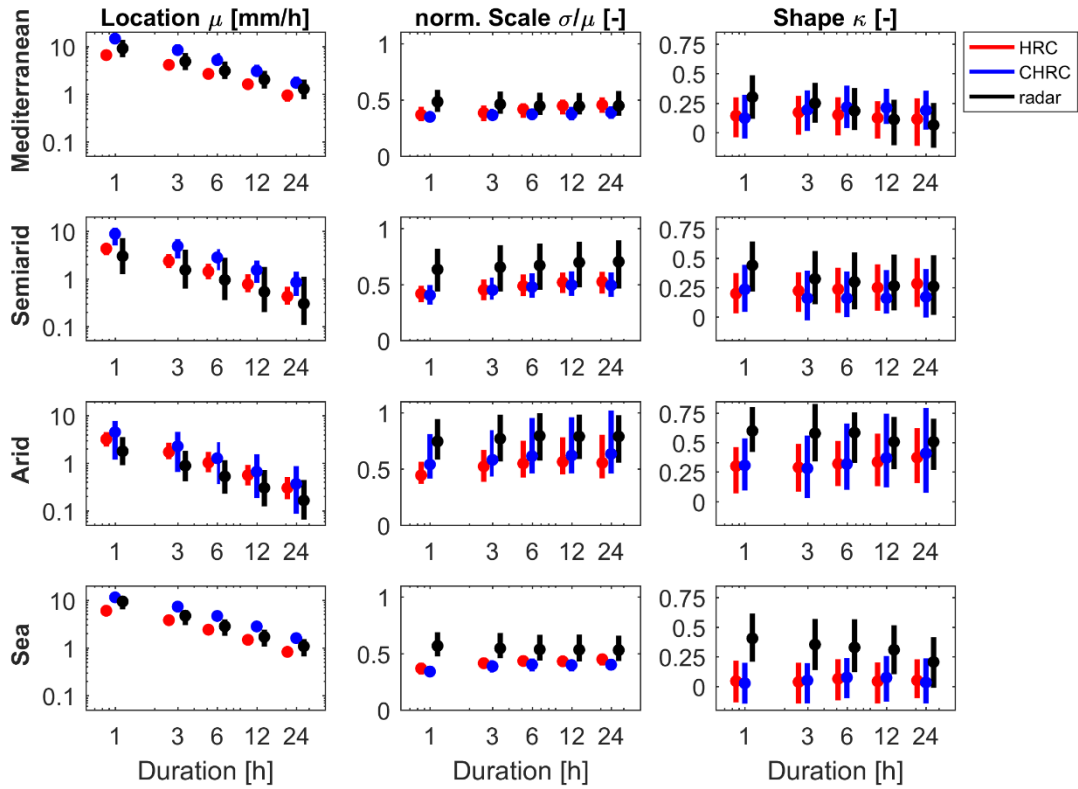
Figure 2: Distribution (median and 25th-75th quantiles) of the GEV parameters derived from satellite (HRC and CHRC) and radar datasets. Note that scale parameters are normalized over the corresponding location parameters. The parameters for different products are represented around the corresponding duration so the logarithmic scale in x-axes should be interpreted accordingly.

5
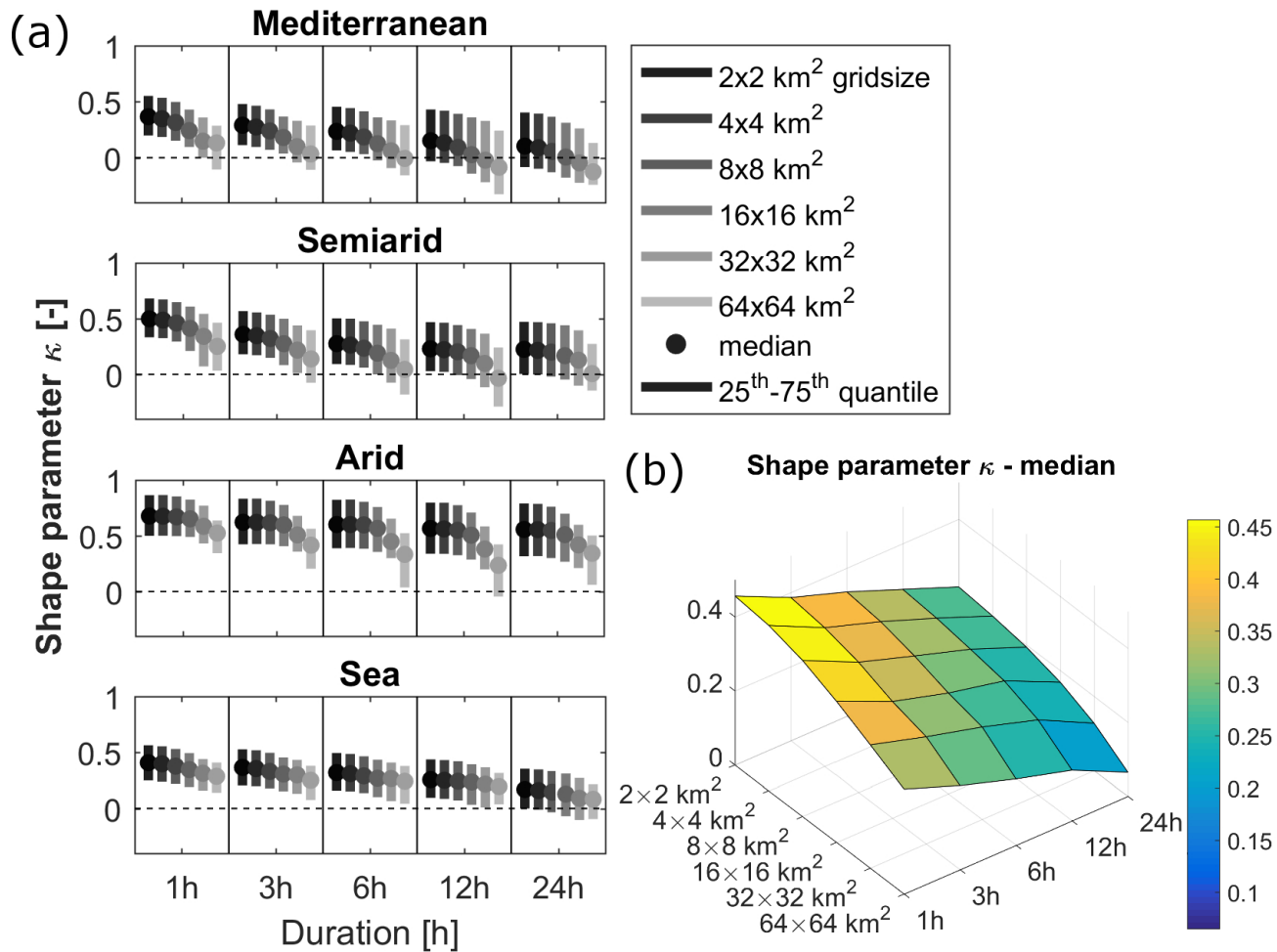
Figure 3: (a) Distribution (median and 25th -75th quantiles) of the shape parameters derived aggregating radar estimates on grid sizes increasing from 2×2 to 64×64 km2. **(b)** Median of the shape parameters (all climates) derived aggregating radar estimates as a function of the spatial and of the temporal aggregation scales.

5
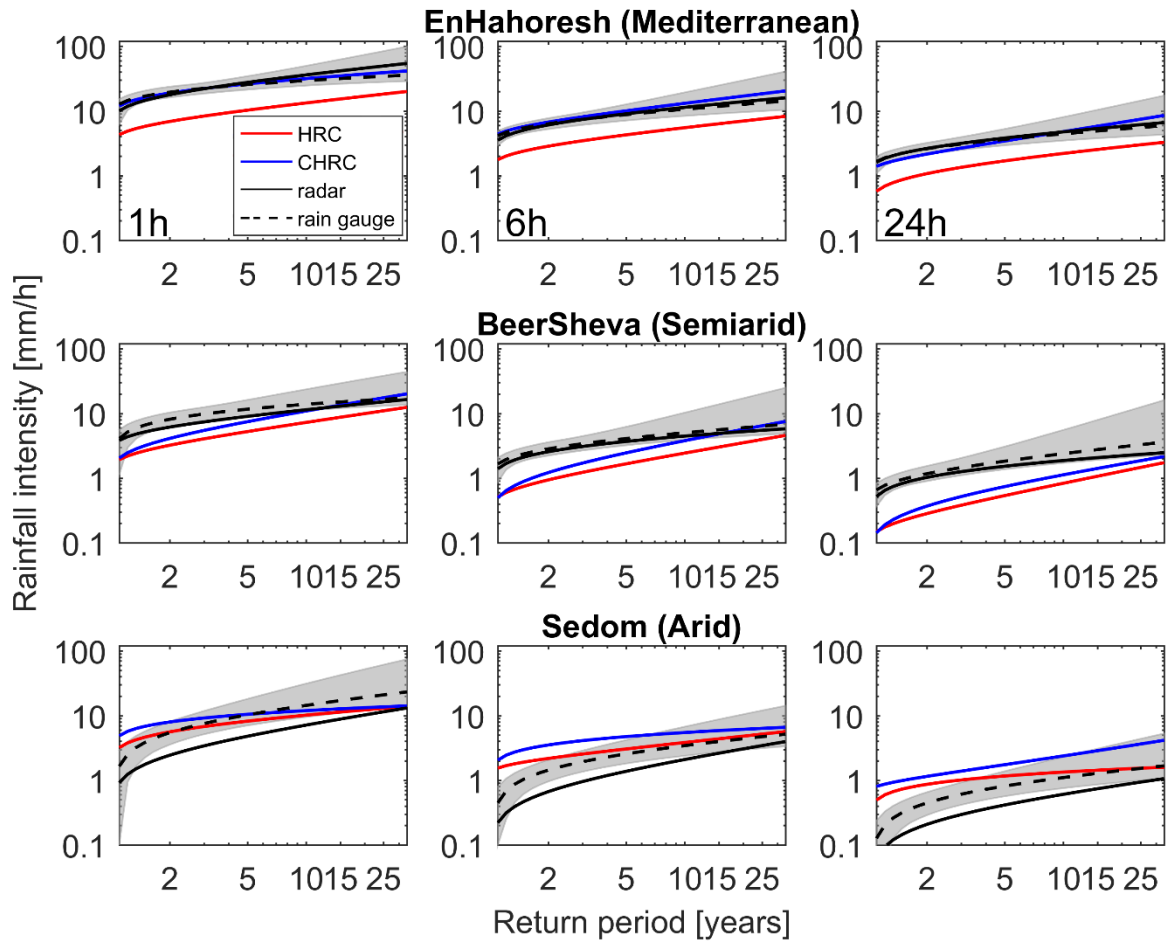
Figure 4: Visual comparison of the annual maxima and of the IDF curves derived for the examined 16 years (1998-2013) from HRC (red), CHRC (~~green~~blue), radar (~~solid~~ black) and rain gauge (~~blue~~dashed black)~~.~~ The shaded area represents the 95 % confidence interval of the rain gauge IDF. ~~for three~~ Three example locations in Mediterranean (En Hahoresh), semiarid (Beer Sheva) and arid (Sedom) climates and for 1, 6 and 24 h durations are shown (see gauge location in Fig. 1). ~~The shaded area represents the 95 % confidence interval of the rain gauge IDF.~~
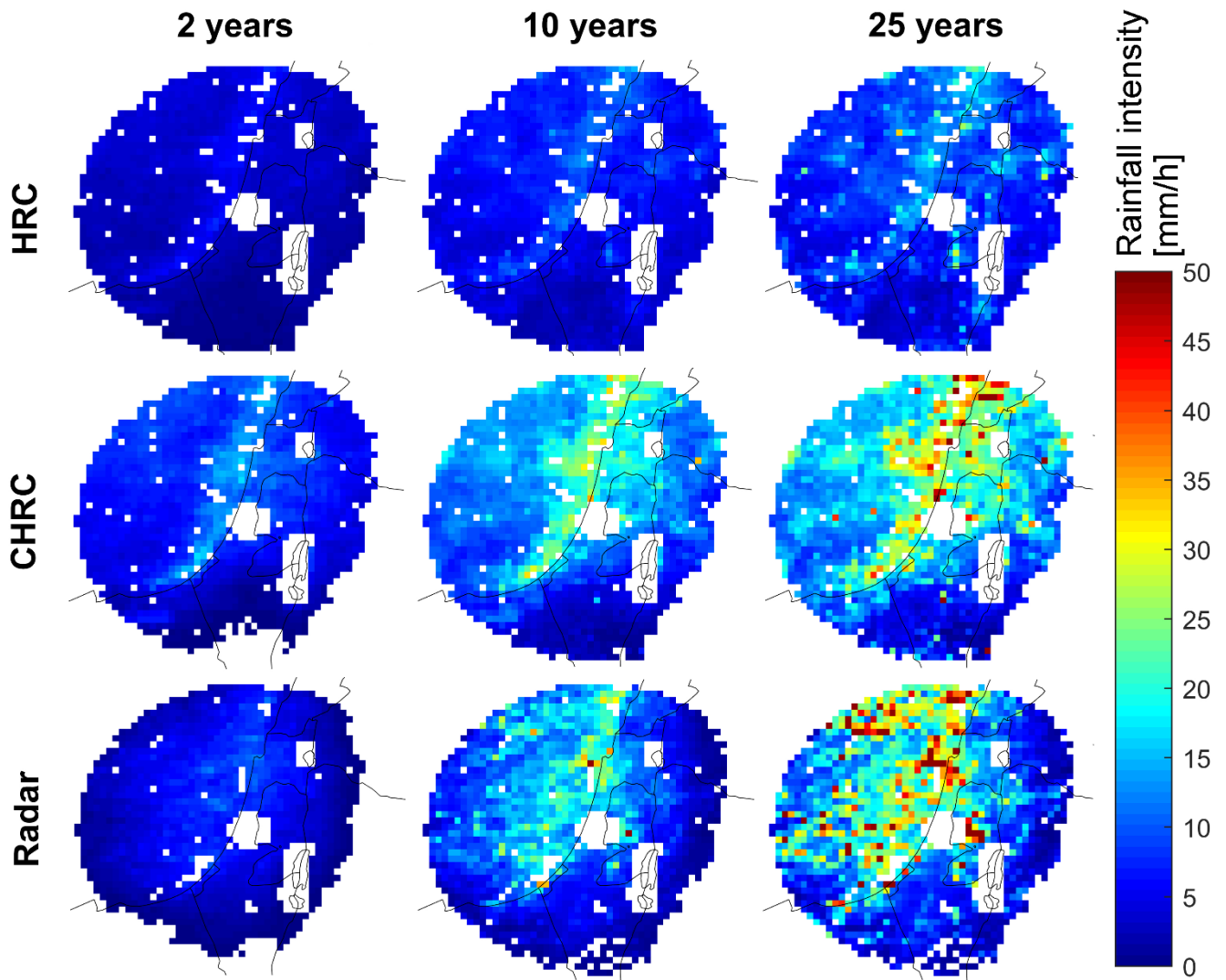
Figure 5: Example of IDF maps for 3 h duration from HRC, CHRC and radar for 2, 10 and 25 years return periods. Only the pixels included in the analyses are shown.
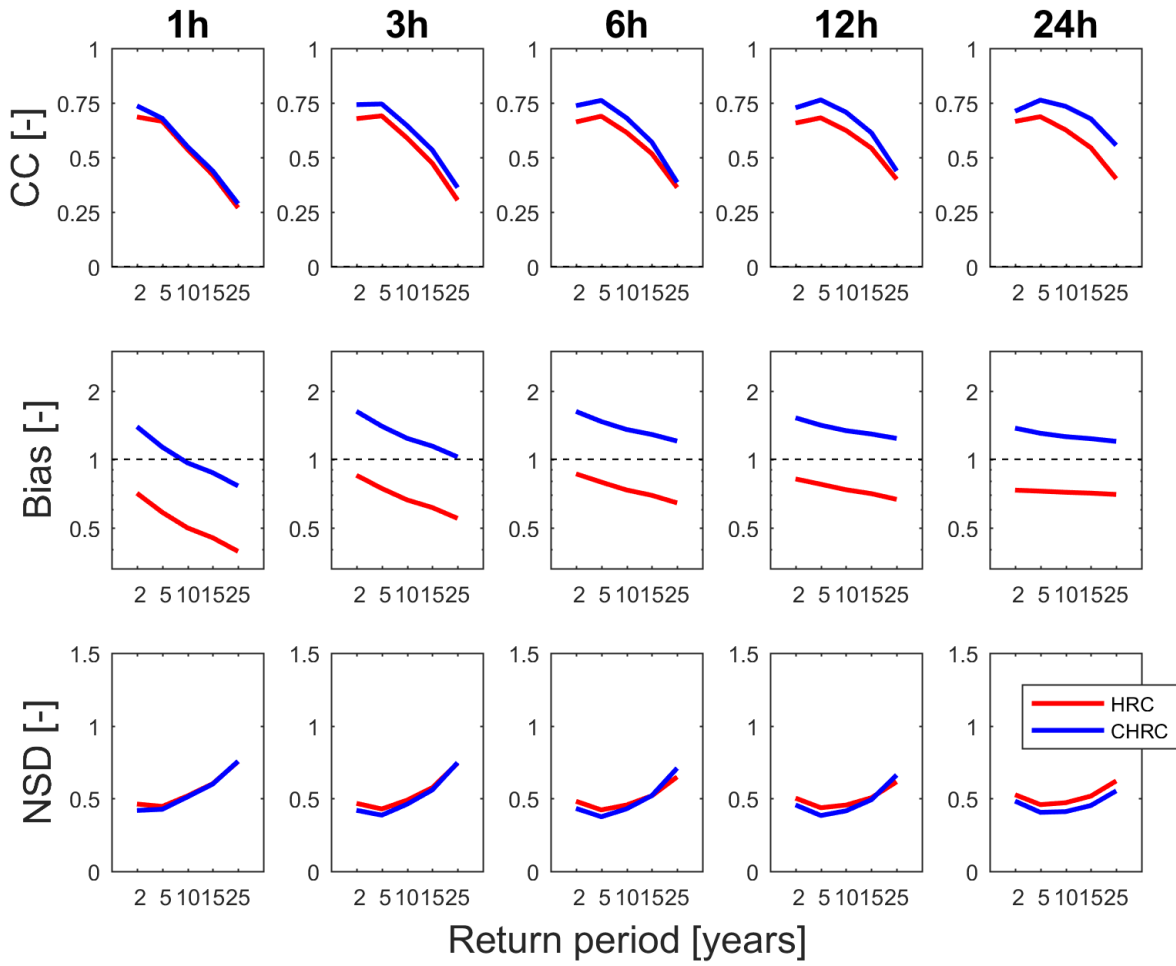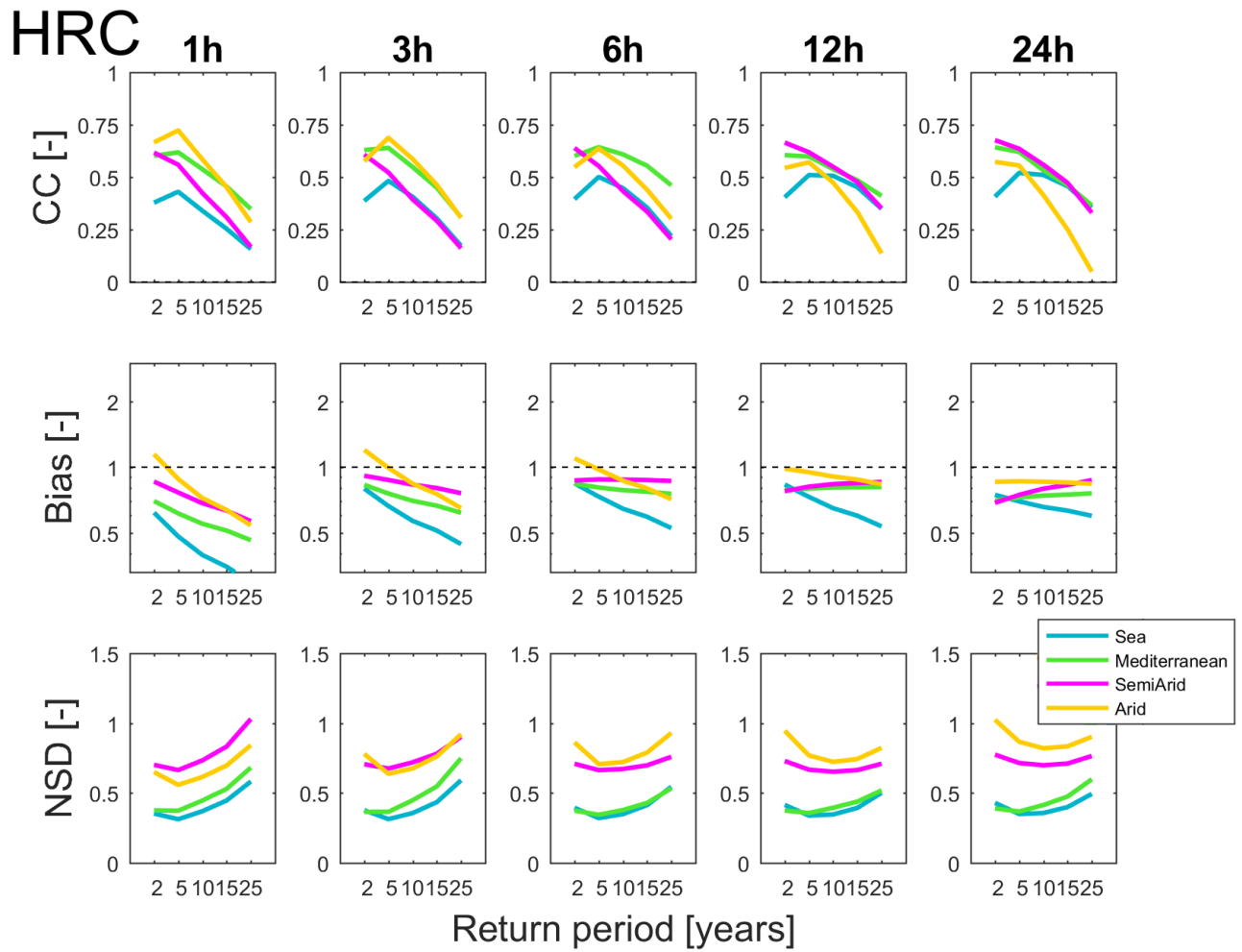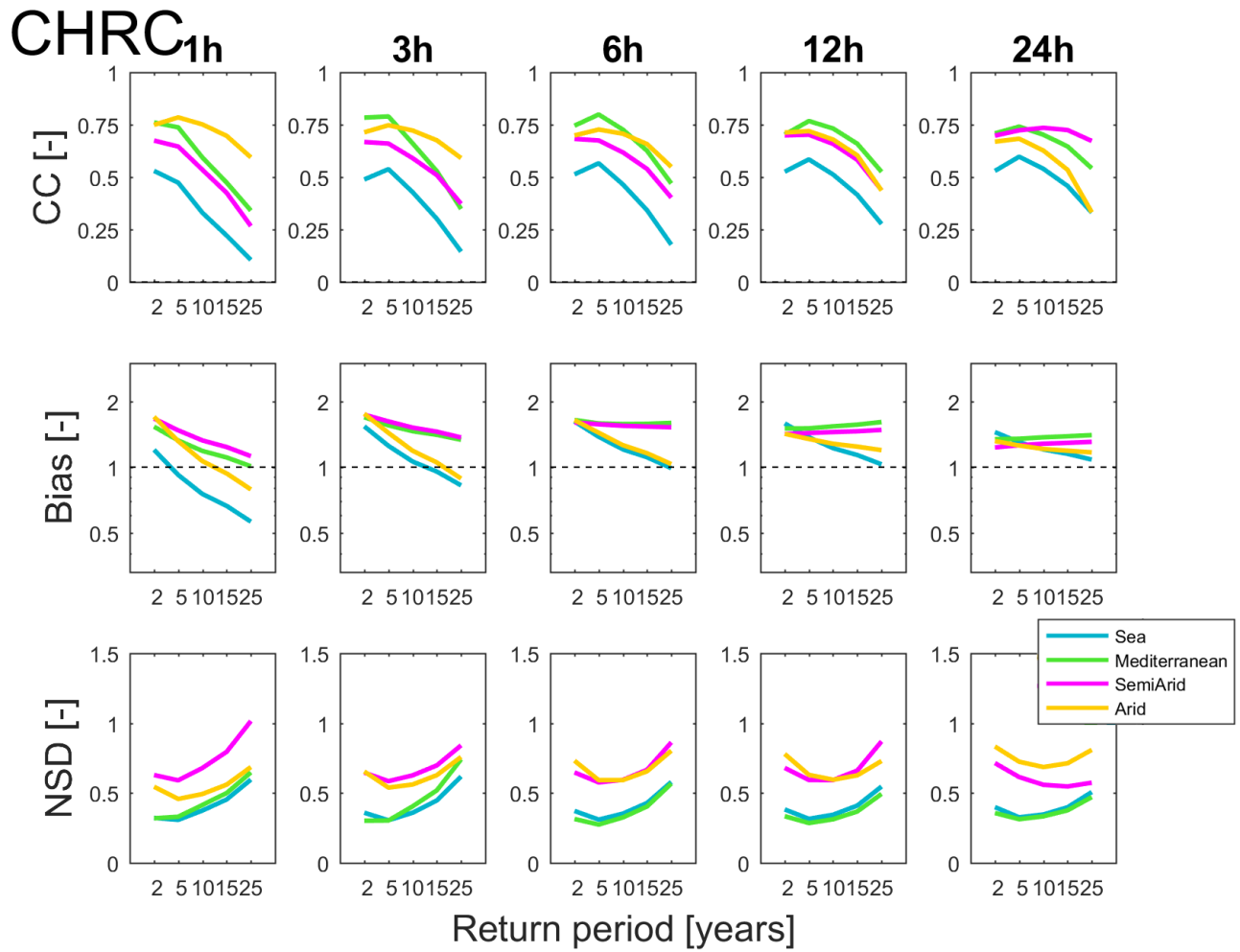
Figure 6: Comparison of IDF values between HRC and radar (red) and CHRC and radar (~~green~~blue). The first row of panels shows the CC for different durations, the second row the Bias and the third row the NSD.

Figure 7: Comparison of IDF values between HRC and radar for different climatic regions. Blue lines represent sea areas while green, pink and orange lines represent Mediterranean, semiarid and arid climates respectively. The first row of panels shows the CC for different durations, the second row the Bias and the third row the NSD.

Figure 8: Comparison of IDF values between CHRC and radar for different climatic regions. Blue lines represent sea areas while green, pink and orange lines represent Mediterranean, semiarid and arid climates respectively. The first row of panels shows the CC for different durations, the second row the Bias and the third row the NSD.
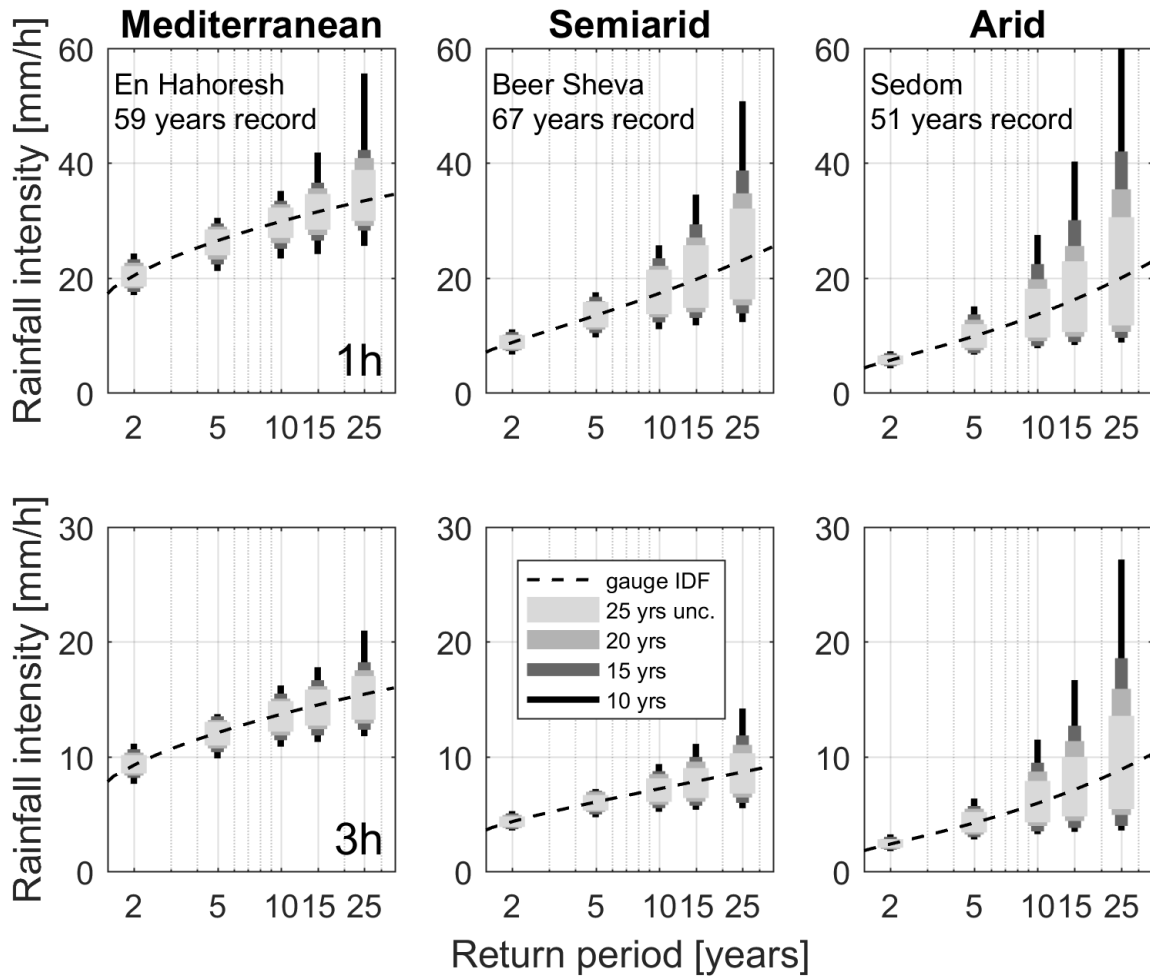
5

Figure 9: Uncertainty related to the record length for 1 and 3 h durations for three example cases in Mediterranean (En Hahoresh), semiarid (Beer Sheva) and arid (Sedom) climates. The dashed lines show the IDF curves from the full records (59, 67 and 51 years, respectively) and the vertical bars show the width of the 5-95th quantile interval of the 999 bootstrap sampling repetitions for record lengths of 10, 15, 20 and 25 years. Width and light of the colour of the bars increase with the record length.
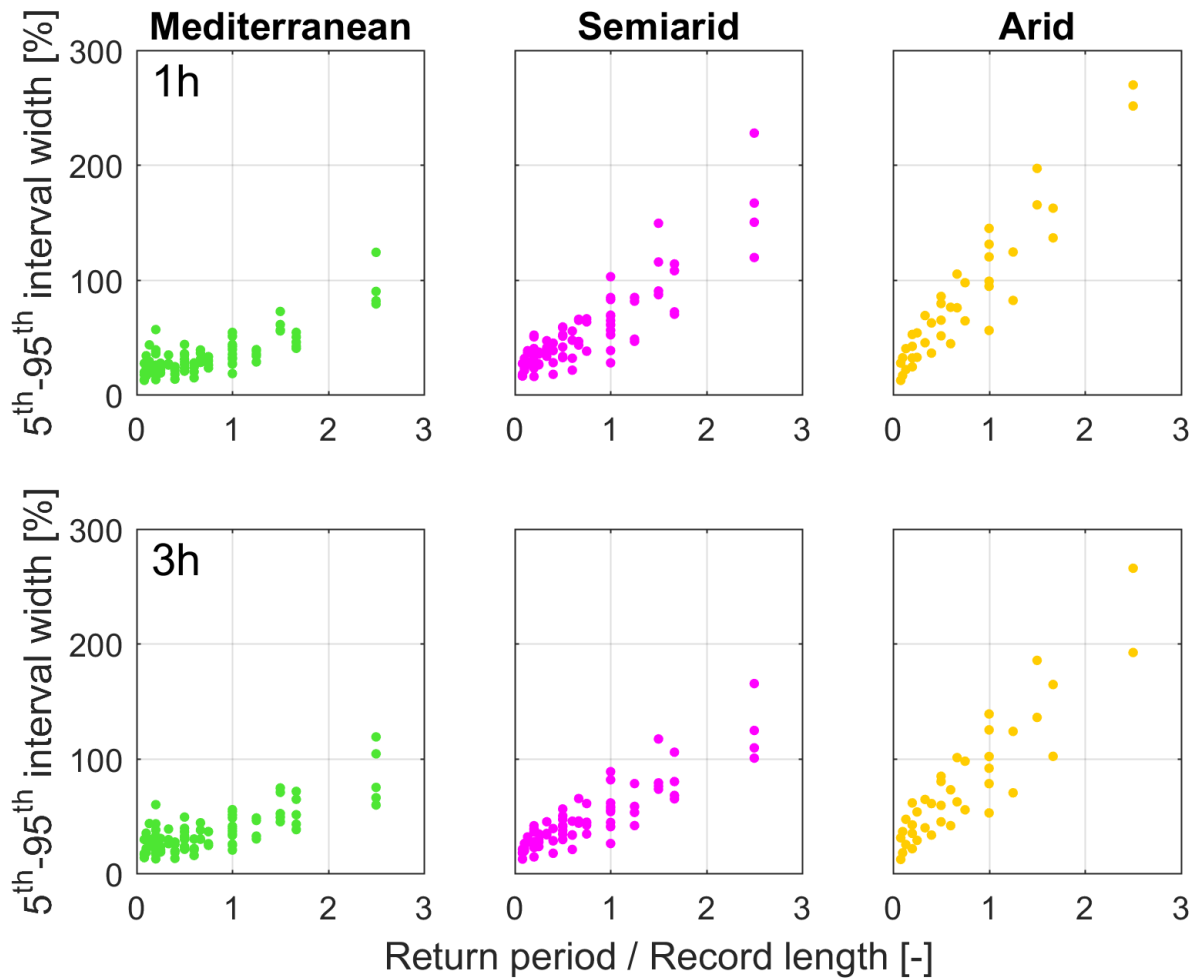
Figure 10: Relative uncertainty (width of the 5-95th quantile interval of the 999 bootstrap sampling repetitions normalized over the rain gauge-IDF value) plotted against the ratio between the estimated return period and the record length. The first row of panels shows the results for 1 h duration, the second row for 3 h. Columns of panels show results for each climatic region.

5