Latin American Data in Science, v.1, n.1
datainscience.com.br
ISSN: 2763-9290

Data analysis reveals that extreme events have increased the flood inundations in the Taquari River Valley, Southern Brazil.

Recebido: 07/06/2021 | Aceito: 08/07/2021 | Publicado: 26/07/2021
https://doi.org/10.53805/lads.v1i1.20

Francisco M. W. Tognoli[*1]; Sabrina D. Bruski[1], Thiago P. Araújo[1,2]

ABSTRACT

Flood inundations represented 43.5% of the world's deaths caused by natural disasters in 2019 and have caused uncountable damages in cities and human settlements in the last years, especially in Brazil. The dataset comprises the records of the Encantado´s pluviometric station, a municipality located beside the margin of the Taquari River in southern Brazil, which comprises the rainfall time series (n = 36,466) over 78 years, from April 1943 to December 2020. Complementary datasets also include the annual volume of precipitation per year and the level reached by the Taquari River during 44 flood inundations since 1941. The number of events is subsampled because only 32 years have the complete record of the river level. Three of the five major flood inundations at Encantado occurred after 2001, and the more severe flood recorded the maximum level of the Taquari River (20,27 meters) on July 8th, 2020. Thirty-four percent of all flood inundations in the city were recorded between 2011 and 2020. The months of July to October record 70% of all the events, but there is no record of floods in February and December throughout the data series. The human occupation of the floodplain has been fast in the last decades, and most of the urban area has a potential risk of being affected by flood inundations. Moreover, extreme rainfall events and flood events have been more frequent in the last 30 years. Therefore, this database can contribute as a starting point for developing predictive models and verifying a possible correlation of floods with extreme events and global climatic changes.

Keywords: Climate change; Sustainable Development Goals; Urban planning; Territorial management; Resilient cities.

PRIOR PUBLICATIONS

BRUSKI, S. D; TOGNOLI, F. M. W.; ARAUJO, T. P. Geotecnologias no contexto das cidades mais resilientes: zoneamento de áreas de risco a inundações como ferramenta de planejamento urbano. Engenharia Urbana em Destaque, 1(1): 30-57, 2020.

[1] PPGEO-UNISINOS, VizGEO Sci Data Lab, São Leopoldo, Brasil. ftognoli@unisinos.br
[2] A&T Geotecnia e Meio Ambiente, Guaíba, Brasil.

DATA IMPORTANCE

- Fluvial inundations are natural processes that historically affect populations living near rivers or other water bodies. However, people migrating to the cities promote inordinate urbanization and potentialize the risk of natural disasters. In 2019, 43.5% of the deaths caused by natural disasters in the world were a consequence of flood inundations (EMDAT, 2021). Flood events at Encantado increased in the last 30 years and more studies are necessary to unveil if these events result of an atypical rainy climatic cycle or if they are a consequence of the global climatic changes expressed as extreme rainfalls;
- The long temporal data series record is fundamental to understanding the pattern of rain periods through time and identifying the main controls of the flood inundations in the studied area. This database is a starting point for developing classification and prediction models that will help authorities and the Civil Defense empower citizens to deal with emergencies during natural disasters;
- We know that floods are not dependent on the precipitation that occurred only in the site of interest. Many different attributes of the entire hydrographic basin control these events, as presented by Kurek (2016). Therefore, we intended to analyze a long pluviometric time series and extract useful information to better understand these events;
- The territorial management and urban planning should include temporal data series analysis to plan urban expansion, reduce natural disasters, and become cities more resilient. The ONU´s Agenda 2030 has 17 Sustainable Development Goals (SDG), including risk reduction and mitigation of natural disasters as part of SDG 11, which intends to make cities and human settlements inclusive, safe, affordable, resilient, and sustainable.

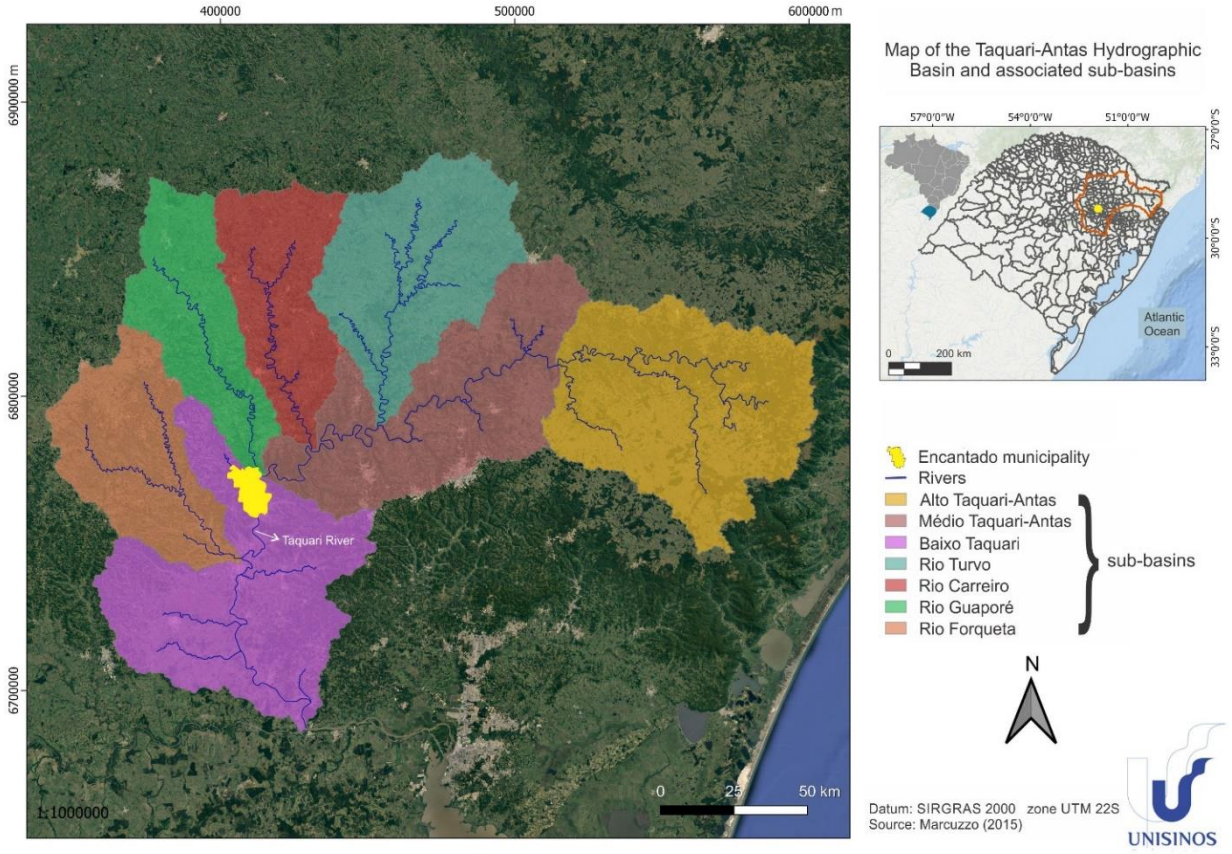
MATERIALS AND METHODS

Encantado is a municipality of 22,000 inhabitants situated in the center-east of Rio Grande do Sul state, south of Brazil (29°14'S, 51°52'W). It is part of the Taquari-Antas Hydrographic Basin (TAHB), a region with 119 municipalities distributed in 39.346 km^2 and with many industrial parks and growing urbanization in the last decades (OLIVEIRA et al., 2018; BRUSKI et al., 2020). The municipality is located in the upper domain of the Baixo Taquari-Antas Valley (Fig. 1), fed by the drainage network that flow from the high-altitude terrains of the TAHB (Alto Taquari-Antas, Médio Taquari-Antas, Rio Turvo, Rio Carreiro and Rio Guaporé) toward the Guaíba Lake, approximately 150 km far from Encantado (MARCUZZO, 2018; OLIVEIRA et al., 2018).

The raw data came from the Hidroweb system, a tool present in the National Water Resources Information System (SNIRH) and available at https://www.snirh.gov.br/hidroweb/serieshistoricas (ANA, 2001). The Encantado´s pluviometric station is identified with the code 2951010 and had data downloaded as a .csv file. After converting to a .xlsx format, the file was analysed to identify the lack of measurements (missing values) and check if days without precipitation had cells filled correctly with zero. This analysis revealed that the exported spreadsheet had included only months with the data record. Therefore, it had included all the missing months to facilitate the visualization. A detailed inspection was also performed on the last days of each month. Considering that months have 28, 29, 30, or 31 days, the database was updated in terms of cell filling and missing values for months with 28 and 29 days (February), 30 days (April, June, September, November), and 31 days (January, March, May, July, August, October and December). The weeks considered the different number of days following the number of days in each month. Therefore, months with 28, 29, 30, and 31 days considered the subdivisions as follow:

- 28 days: week 1 = 01-07, week 2 = 08-14, week 3 = 15-21, week 4 = 22-28
- 29 days: week 1 = 01-07, week 2 = 08-14, week 3 = 15-21, week 4 = 22-29
- 30 days: week 1 = 01-08, week 2 = 09-15, week 3 = 16-23, week 4 = 24-30
- 31 days: week 1 = 01-08, week 2 = 09-15, week 3 = 16-23, week 4 = 24-31

Figure 1. Location map of the Taquari Antas Hydrographic Basin, sub-basins and the Encantado municipality in southern Brazil. Modified from Bruski et al. (2020).



For the annual calculation of the statistic parameters, it was used all data available in the dataset. The year 1961 was not considered for calculations because there is only one month of measurements. The years of 1943, 1952, 1953, 1985, 2006, 2007, 2009, 2010, 2015, 2016, 1017, 2019 and 2020 have missing values in some months. Although they are part of the dataset, we must consider that almost all the statistic values from the annual analysis are underestimated, especially in the winter season.

A Python script named script_Encantado_LADS.ipynb (see Supplementary Material) run three .csv files using the pandas, seaborn, matplotlib, numpy, sklearn, and plotly libraries to convert the spreadsheet into a dataframe, inspect the statistics of each attribute, perform data cleaning, count missing values, calculate new attributes, and generate histograms, scatter plots, and Pearson´s correlation coefficient. One specific step calculated high outliers of precipitation based on the attribute "Montly_total_vol" using the formula: *Vmax = Q3 + (1.5*(Q3-Q1)*, being Q1 and Q3 the first and third quartiles, respectively. The output provided a list of anomalous rainfalls that was compared with the maximum flood inundations along the time series to verify if there is a correlation between monthly precipitation and the flood events at Encantado.

DATA DESCRIPTION

The analysis of the precipitation along daily to annual time series is fundamental to understand the pluviometric behavior through time (Fig. 2). We reinforce that the annual pluviometric records are underestimated, as indicated in the section Material and Methods. Therefore, the values of the mean, standard deviation, and quartiles are approximate. The period between 1943 and 1957 had precipitation below the historical mean. The interval 1943-1990 has 12 years with precipitation below the first quartile (Q1), whereas the interval 1991-2020 has possibly only one year (2004) because all the other years below Q1 are those with missing values into the dataset (2006, 2010, 2016). The comparison of the last two intervals also reveals that 18 years had precipitation above the mean and median in the interval 1943-1990, whereas the number to the interval 1991-2020 was 20. Considering the third quartile (Q3), the comparison shows 8 years for the interval 1943-1990 and at least 7 years for the interval 1991-2020 with precipitation above Q3. It shows unequivocally that the precipitation in the last 30 years is proportionally higher than the 48 first years of the time series (Fig. 2). Only the years 1966 and 2002 represents high outliers in the time series.

All the attributes related to the rainfall volume had calculated their statistic parameters. The mean daily rainfall is 38.7 mm, and the maximum volume of precipitation into 24h was 135.0 mm. The mean of rainy days is 7.5, and the maximum number of rainy days in a month was 20. The monthly data analysis revealed average precipitation of 116.9 mm, with a maximum of 382 mm. The annual rainfall values are 1359.5 mm (mean) and 2265.7 mm (maximum) (Table 1). The accumulated precipitation of a fortnight can be 81.5% of the monthly total, whereas one week can present 61.5% of the month´s rainfall.

Figure 2. Annual volume of rainfall along the time series 1943-2020. The statistic parameters Q1, mean, median, and Q3 indicate that the last 30 years (1991-2020) had proportionally more volume of precipitation than the first 48 years (1943-1990). Note the * 2020 had a very high volume of precipitation but the data record is not complete.

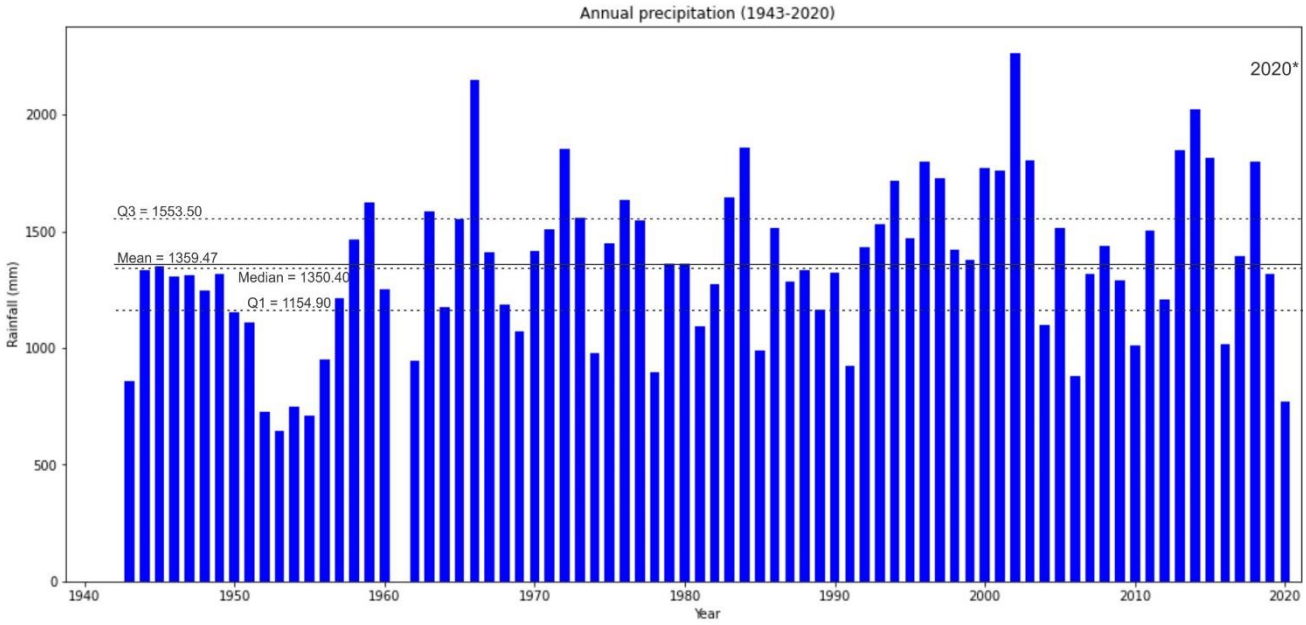


Table 1. Summary of the statistics of the main attributes. 50% represents the median, 25% and 50% represent the first and third quartiles. The * indicates that the minimum annual volume is underestimated because some years have missing monthly values.

	Daily_max_vol	Rain_ days	Monthly_ total_vol	Annual_ vol	Vol_ fortnight1	Vol_ fortnight2	Vol_ week1	Vol_ week2	Vol_ week3	Vol_ week4
count	898	898	898	77	898	898	898	898	898	898
mean	38.764	7.536	116.942	1359.5	56.461	60.481	29.054	27.407	31451	29.030
std	22.213	3.379	69.222	342.07	46.210	46.745	30.723	32.576	33.763	32.533
min *	0	0	0	644.3	0	0	0	0	0	0
25%	23.525	5	66.950	1154.9	22.500	25.525	5.200	1	4.7	3.225
50%	33.200	7	106.700	1350.4	45.300	53.450	21.350	17.050	22	20.400
75%	50.700	10	155.300	1553.5	76.975	86.375	40.375	39.200	48.075	41.8
max	135	20	382	2265.7	311.500	301.700	190	220.500	235.2	183

A logic function allowed selecting all the outliers related to the anomalous rainfall into the time series. The value 287.82 mm is the cutoff obtained by calculating the Vmax using the attribute "Montly_total_vol." Twenty-two months had precipitation higher than the cutoff into the 78 years of analysis, but the maximum inundation of July 2020 does not have data available on the Hidroweb system. Seven of these months had flood inundations recorded at Encantado, according to Peixoto & Lamberty (2019) and Bruski et al. (2020) (Table 2).

The number of flood inundations in the intervals of 1943-1990 and 1991-2020 is the same (22). However, the number of events in the last decade (2011-2020) is 66.7% higher than the former maximum (1951-1960) and 375% higher than the decade 2001-2010 (Fig. 3).

Figure 3. Distribution of the flood events at Encantado in the last 78 years.

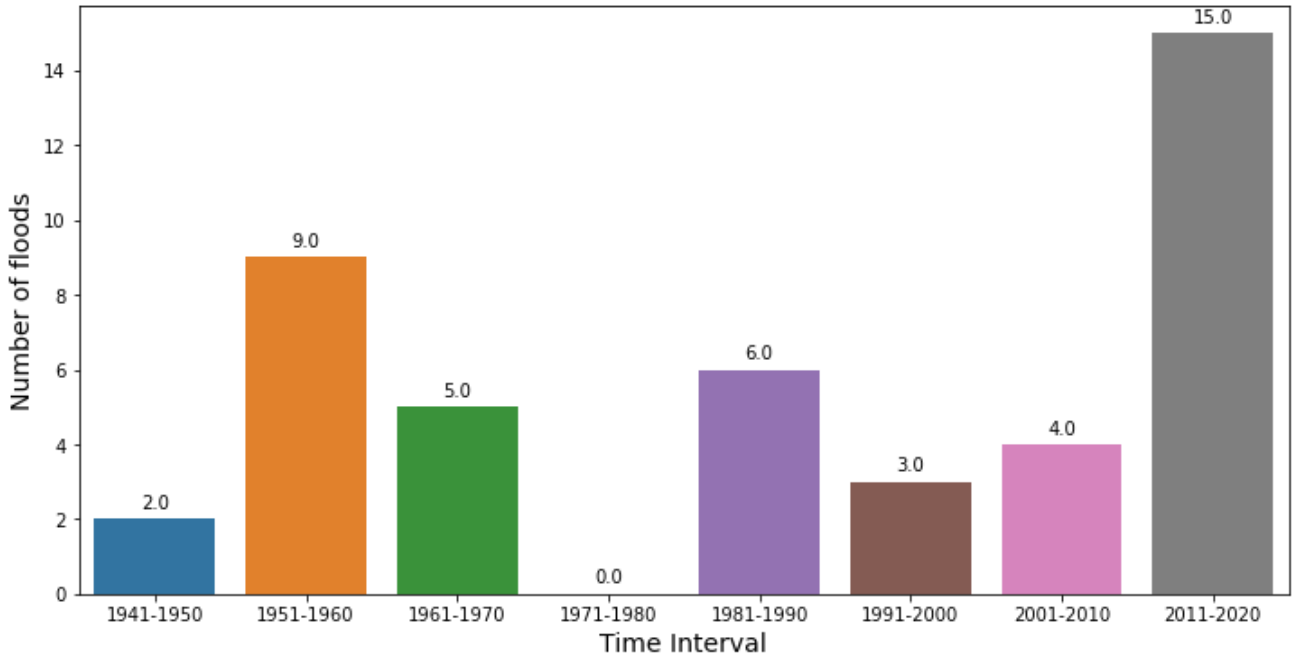


Table 2. List of months with high outliers of precipitation based on the monthly total rainfall volume. The cutoff of this attribute is 287.82 mm. The * indicates that the month of the maximum level of the Rio Taquari (July 2020) does not have pluviometric record.

	Date	Month	Daily_ max_vol	Rain_ days	Monthly_ total_vol	Flood inudation (Peixoto & Lamberty, 2019)	Flood inundation (Bruski et al., 2020)
*	01/07/2020	July	not informed			----	Yes
15	01/10/2019	October	53.4	12	319.2	No	No
51	01/10/2016	October	107.8	11	322.6	Yes	Yes
63	01/10/2015	October	81.0	14	293.5	No	Yes
89	01/08/2013	August	62.2	11	312.9	No	Yes
134	01/11/2009	November	90.3	11	318.3	No	No
183	01/10/2005	October	56.4	13	296.5	No	No
205	01/12/2003	December	100.2	10	334.3	No	No
219	01/10/2002	October	103.2	19	382.0	No	No
279	01/10/1997	October	72.1	18	349.2	No	No
299	01/02/1996	February	40.8	19	314.4	No	No
315	01/10/1994	October	61.0	13	324.4	No	No
388	01/09/1988	September	83.2	14	357.8	No	Yes
439	01/06/1984	June	64.3	13	291.1	No	No
545	01/08/1975	August	75.0	15	298.5	No	No
583	01/06/1972	June	85.0	15	307.2	No	No
598	01/03/1971	March	109.0	9	330.8	No	No
640	01/09/1967	September	61.2	10	342.7	Yes	Yes
649	01/12/1966	December	52.0	9	292.0	No	No
652	01/09/1966	September	91.0	7	311.5	No	No
900	01/01/1946	January	69.2	13	302.8	Yes	Yes
919	01/06/1944	June	102.1	12	294.8	No	No
					Vmax= 287.82		

A statistical analysis of the attributes related to the 44 flood inundations that occurred at Encantado, considering the high outliers and the entire time series of 1943-2020 (Table 3; Fig. 4), shows that:

- Only 14.0% of the flood inundations are directly related to the high outliers (> Vmax) of the attribute Monthly_total_vol (Table 3A), but 76.7% of the floods show a direct relationship with the rain volume of the third quartile (> Q3) (Table 3B).
- A mean of 18.6% of the floods is related to high outliers of the attributes Vol_fortnight1 and Vol_fortnight2 (Table 3A), but this value increases to 55.8% if considered the rain volume third quartile (Table 3B).
- 16.3% of the floods show a direct relationship with the high outliers of the attributes Vol_week1 to Vol_week4 (Table 3A); however, 44.2% of the flood events correlate with rain precipitations higher than the third quartile (Table 3B).

A statistical analysis of the attributes related to the 44 flood inundations at Encantado, considering the high outliers and subdividing the time series into two intervals, 1943-1990 and 1991-2020, shows that:

- The number of floods in the intervals of 1943-1990 and 1991-2020 is the same (3), based on the monthly data (Fig. 4).

- Considering the fortnights, the number of events of the interval 1943-1990 (4) is slightly higher than those of 1991-2020 (3) (Fig. 4).
- However, if based on the weekly data, the number of flood events after 1991 (17) is 41.6% higher than the events that occurred in the interval 1943-1990 (12) (Fig. 4).

Table 3. Temporal relationships of the 44 flood inundations since 1943. A, B) Database showing the date of the floods, the river level, and the monthly, fortnightly, and weekly precipitations. In A) and B), a logic function informs if the flood has a direct relationship with the precipitation higher than the Vmax (A) or the third quartile (Q3) (B).

A)

									> Vmax = Outlier						
Date	**River level**	**Mtv**	**Vf1**	**Vf2**	**Vw1**	**Vw2**	**Vw3**	**Vw4**	**Mtv**	**Vf1**	**Vf2**	**Vw1**	**Vw2**	**Vw3**	**Vw4**
00/07/2020	20.27	-	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
00/10/2018	11.62	182.10	57.40	124.70	54.10	3.30	0.00	124.70	No	No	No	No	No	No	Yes
00/10/2017	11.35	165.00	133.8	31.2	32.4	101.4	0	31.2	No	No	No	No	Yes	No	No
00/06/2017	14.43	29.70	29.7	0	29.7	0	0	0	No	No	No	No	No	No	No
00/05/2017	11.69	202.70	35.5	167.2	8.1	27.4	9.4	157.8	No	No	No	No	No	No	Yes
00/10/2016	17.34	322.60	20.9	301.7	0	20.9	235.2	66.5	Yes	No	Yes	No	No	Yes	No
00/07/2016	14.98	73.30	11.5	61.8	10.2	1.3	36	25.8	No	No	No	No	No	No	No
00/10/2015	15.95	293.50	196.6	96.9	107.1	89.5	96.6	0.3	Yes	Yes	No	Yes	No	No	No
00/07/2015	11.27	269.80	189.7	80.1	13.8	175.9	75.8	4.3	No	Yes	No	No	Yes	No	No
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
00/01/1946	17.70	302.80	118.8	184	12.5	106.3	69.4	114.6	Yes	No	Yes	No	Yes	No	Yes
	Vmax=	**287.82**	**158.68**	**177.65**	**93.14**	**96.5**	**113.13**	**99.66**	0.140	0.163	0.209	0.140	0.186	0.140	0.209
	Q3=	**155.30**	**76.98**	**86.38**	**40.38**	**39.2**	**48.08**	**41.8**	0.140	0.186			0.163		

B)

									> Q3						
Date	**River level**	**Mtv**	**Vf1**	**Vf2**	**Vw1**	**Vw2**	**Vw3**	**Vw4**	**Mtv> Q3?**	**Vf1**	**Vf2**	**Vw1**	**Vw2**	**Vw3**	**Vw4**
00/07/2020	20.27	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xxx
00/10/2018	11.62	182.10	57.40	124.70	54.10	3.30	0.00	124.70	Yes	No	Yes	Yes	No	No	Yes
00/10/2017	11.35	165.00	133.8	31.2	32.4	101.4	0	31.2	Yes	Yes	No	No	Yes	No	No
00/06/2017	14.43	29.70	29.7	0	29.7	0	0	0	No	No	No	No	No	No	No
00/05/2017	11.69	202.70	35.5	167.2	8.1	27.4	9.4	157.8	Yes	No	Yes	No	No	No	Yes
00/10/2016	17.34	322.60	20.9	301.7	0	20.9	235.2	66.5	Yes	No	Yes	No	No	Yes	Yes
00/07/2016	14.98	73.30	11.5	61.8	10.2	1.3	36	25.8	No	No	No	No	No	No	No
00/10/2015	15.95	293.50	196.6	96.9	107.1	89.5	96.6	0.3	Yes	Yes	Yes	Yes	Yes	Yes	No
00/07/2015	11.27	269.80	189.7	80.1	13.8	175.9	75.8	4.3	Yes	Yes	No	No	Yes	Yes	No
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
00/01/1946	17.70	302.80	118.8	184	12.5	106.3	69.4	114.6	Yes	Yes	Yes	No	Yes	Yes	Yes
	Vmax=	**287.82**	**158.68**	**177.65**	**93.14**	**96.5**	**113.13**	**99.66**	0.767	0.581	0.535	0.395	0.512	0.419	0.465
	Q3=	**155.30**	**76.98**	**86.38**	**40.38**	**39.2**	**48.08**	**41.8**	0.767	0.558			0.442		

A composite Pearson´s chart comparing the time series of 1943-1990 and 1001-2020 points out to a more expressive influence of the monthly, fortnightly, and weekly precipitation in the river level during flood events. The last 30 years present a higher Pearson´s coefficient for the time interval of 1991-2019 if compared with the same attributes for the interval of 1943-1990 (Fig. 5).

Dataset

The Encantado_annual_rainfall.csv is a file with 2 columns, Year and Volume(mm), used to plot the rainfall variation in the period 1943-2020.

The second file is the Encantado_rain_data.csv, the most complete dataset with 933 lines and 48 columns which includes all the data associated with each month from April 1943 to January 2021. The columns record, from left to right, Station Code, Data Type, Date (always indicated as the first day of each month), Month, Measurer, the maximum daily volume in millimeters (Daily_max_vol), the day of maximum volume (Day_max_vol), the number of rainy days into the month (Rain_Days), the monthly total volume of precipitation (Monthly_total_vol), the annual volume of precipitation (Annual_vol) and its respective year (Year)(both always indicated in the December line), the volume of the first and second fortnights (Vol_fortnight1 and Vol_fortnight2, respectively), the weekly volumes (Vol_week1, Vol_week2, Vol_week3, and Vol_week4, respectively), and the daily volumes (Day 01 until Day 28, 29, 30 or 31, respectively)(Table 4).

Figure 4. Statistic variation of the precipitation data and number of events (n) separated by time intervals. X indicates the number of events associated with high outliers; y and z indicate the number of events in each time interval. Mtv = Monthly total volume. Vf1 and Vf2 = Fortnightly volume. Vw1 to Vw4 = Weekly volume.

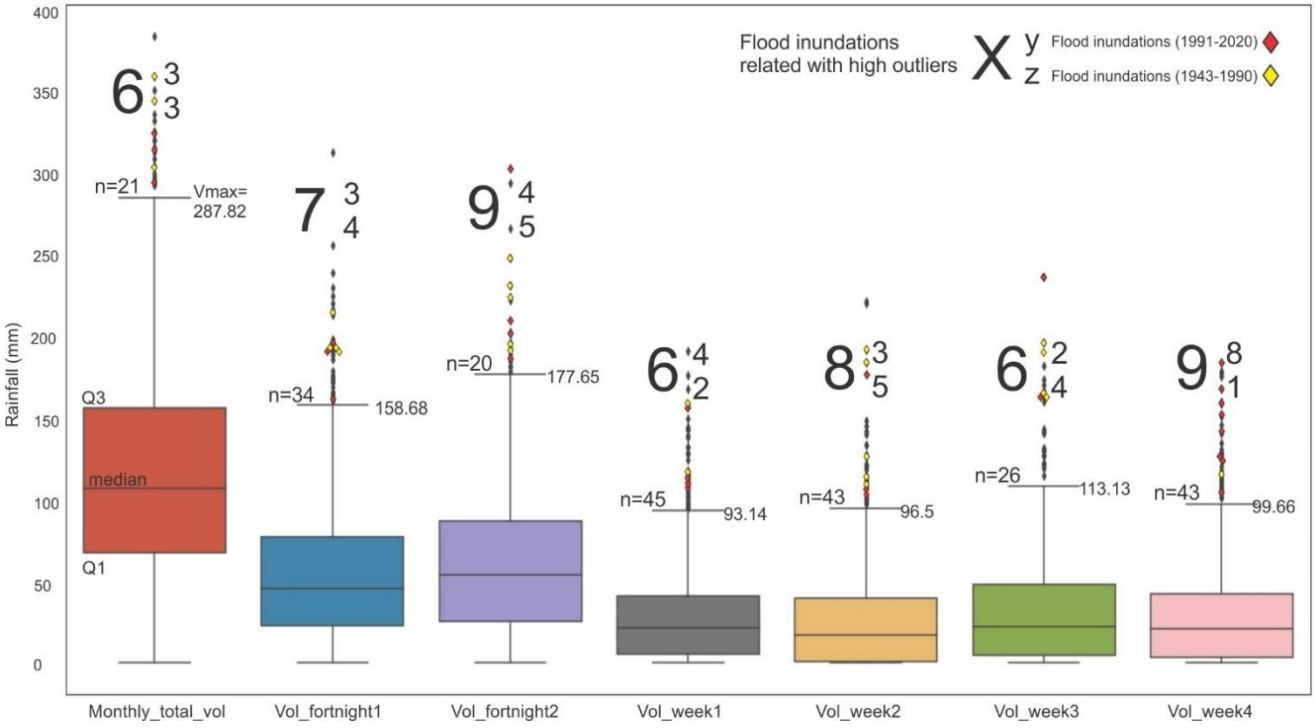


Figure 5. Composite Pearson´s determination coefficient showing the increase of the weekly mean as a control of the river level if compared the time intervals of 1943-1990 and 1991-2019. Compare the Vw_mean and the River level (m) cells for both time intervals. The maximum flood inundation of 2020 was excluded of the analysis because some data are missing.

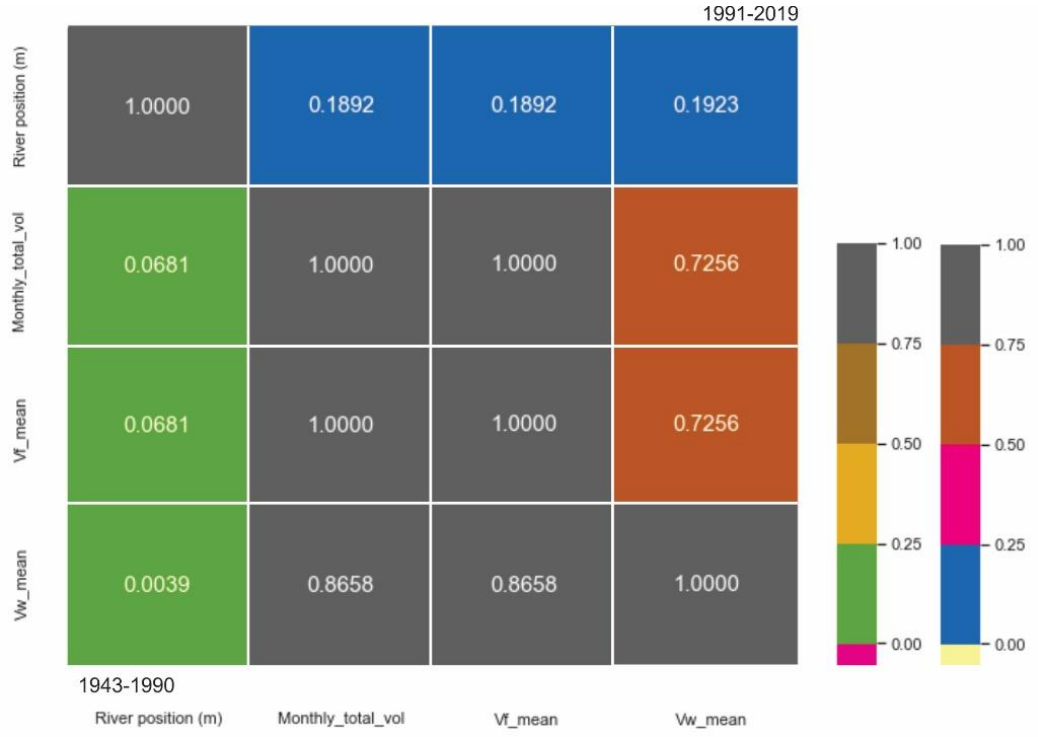


Table 4. Structure of the database showing the attributes.

	Station_ Code	Data_ Type	Date	Month	Measurer	Daily_max_ vol	Day_max _vol	Rain_ days	Monthly_ total_vol	Annual _vol	Year	...	Day22	Day23	...
0	2951010	Raw	01/01/2021	January	Pluviometer	47.7	17	12	223.5	NaN	NaN	⋯	6.4	0	⋯
1	2951010	Raw	01/12/2020	December	Pluviometer	35.8	3	4	91.3	769.8	2020	⋯	0	0	⋯
2	2951010	Raw	01/11/2020	November	Pluviometer	43.5	19	5	78.2	NaN	NaN	⋯	0	0	⋯
3	2951010	Raw	01/10/2020	October	Pluviometer	24.9	27	6	62	NaN	NaN	⋯	0	0	⋯
4	2951010	Raw	01/09/2020	September	Pluviometer	29.7	18	9	130.9	NaN	NaN	⋯	0	0	⋯

The third file combined the weekly, fortnightly, and monthly attributes recorded by the station 2951010 with the Taquari River´s level in periods of flood inundations and generated the file Encantado_rain_flood.csv. This file included calculations of the high outliers derived from the monthly, fortnightly, and weekly rain precipitation data. These new columns used the Vmax of each attribute to compare with the values of all the cells and a logic function determined if the value is an outlier (Yes) or not (No).

SUPPLEMENTARY MATERIALS

Encantado_annual_rainfall.csv
Encantado_rain_data.csv
Encantado_rain_flood.csv
script_Encantado_LADS.ipynb

ACKNOWLEDGEMENTS

The authors thank the Brazilian Geological Survey (CPRM), Agência Nacional das Águas (ANA) and Defesa Civil for providing data and information about the flooding events in Encantando-RS. FMWT thanks the financial support given by the Brazilian Council for Scientific and Technological Development (CNPq – proc. 311204/2017-1).

REFERENCES

ANA. Hidroweb 3.1.1: Séries históricas de estações. Available in: <https://www.snirh.gov.br/hidroweb/serieshistoricas>. Accessed in: 05/25/2021.

BRUSKI, S. D.; TOGNOLI, F. M. W.; ARAÚJO, T. P. Geotecnologias no contexto das cidades mais resilientes: zoneamento de áreas de risco a inundações como ferramenta de planejamento urbano. Engenharia Urbana em Debate, 1(1), 30-57, 2020.

EMDAT. OFDA/CRED International Disaster Database. Université Catholique de Louvain, Brussels, 2021. Available in: <https://ourworldindata.org/natural-disasters#natural-disasters-kill-on-average-60-000-people-per-year-and-are-responsible-for-0-1-of-global-deaths>. Accessed in: 07/05/2021.

MARCUZZO, F. F. N. Mapas e opções de divisão territorial do estado do Rio Grande do Sul por bacias hidrográficas. Congresso Brasileiro de Geologia, 49, 2018, Rio de Janeiro. Anais do 49º Congresso Brasileiro de Geologia. Porto Alegre: Sociedade Brasileira de Geologia, 2018. Available in: <http://rigeo.cprm.gov.br/jspui/bitstream/doc/19906/2/cbg2018_mapas_opcoes_poster.pdf>. Accessed in: 25/05/2021.

KUREK, R. K. M. Análise das Inundações no Vale do Taquari/RS como Subsídio à Elaboração de um Modelo de Previsão. 2016, 178p. Dissertação de Mestrado em Engenharia Civil, Centro de Tecnologia, Universidade Federal de Santa Maria, Santa Maria. Available in: <https://repositorio.ufsm.br/bitstream/handle/1/7908/KUREK%2c%20ROBERTA%20KARINNE%20MOCVA.pdf?sequence=1&isAllowed=y>. Accessed in: 25/05/2021.

OLIVEIRA, G. G. et al. Caracterização espacial das áreas suscetíveis a inundações e enxurradas na Bacia Hidrográfica do Rio Taquari-Antas. Geociências, 37(4), 849-863, 2018. DOI: https://doi.org/10.5016/geociencias.v37i4.12312.

PEIXOTO, C. A. B.; LAMBERTY, D. Setorização de áreas em alto e muito alto risco a movimentos de massa, enchentes e inundações: Encantado – Rio Grande do Sul. Ministério de Minas e Energia, 23p. 2019. Available in: <http://rigeo.cprm.gov.br/jspui/handle/doc/20097>. Accessed in: 25/05/2021.