

Reproducibility review of: Extreme heat alerts and impacts across Mozambique 2016 - 2022: gathering evidence from media articles

Eftychia Koukouraki 

2023-05-26



This report is part of the reproducibility review at the AGILE conference. For more information see <https://reproducible-agile.github.io/>. This document is published on OSF at <https://doi.org/10.17605/osf.io/eu8kw>. To cite the report use

Koukouraki, E. (2023, April 14). Reproducibility review of: Extreme heat alerts and impacts across Mozambique 2016 - 2022: gathering evidence from media articles. <https://doi.org/10.17605/OSF.IO/EU8KW>

Reviewed paper

Pereira Marghidan, C., van Aalst, M., Blanford, J., Maure, G., and Marrufo, T.: Extreme heat alerts and impacts across Mozambique 2016 - 2022: gathering evidence from media articles, AGILE GIScience Ser., 4, 11, <https://doi.org/10.5194/agile-giss-4-11-2023>, 2023

Summary

The paper elaborates on the data collection about extreme heat alerts in Mozambique from different media sources and subsequently overlays the collected datasets. After manual data collection, the data were gathered and analysed in spreadsheets and then shared via Github under CC BY 4.0. For this reproducibility review, we attempted to reproduce Figures 1-4 and Table 2. Since the data processing steps were not detailed enough in the README files provided by the corresponding author, in most cases we managed to do a verification of the results, rather than a reproduction. For this reason, the reproduction of this paper can be considered only **partially successful**.

Reproducibility reviewer notes

Data sharing

The materials that were used for the elaboration of this study were made publicly available under CC BY 4.0 via the following Github repository: https://github.com/carolinamarghidan/AGILE_MZ_Heat. The repository consists of a README with links to the data files that are necessary to reproduce the figures of the paper.

Dataset 1 is a spreadsheet that compiles and analyses temperature related records from several media sources that can be found online (Twitter, Facebook, Club of Mozambique, Rádio Moçambique, O País, and Televisão de Moçambique). The spreadsheet contains data about: Date of post, Location (district), Province, Maximum temperature recorded in media, Maximum temperature from weather station, Impacts recorded (eg. on agriculture), Other information about weather conditions of the heat wave, Duration in days, Recommendations for citizens, Link to online article and Source outlet. There are also three more spreadsheets that were shared upon the reproducibility committee's request for more detailed information about how the figures of the paper were produced: Hwalert_figure2.xlsx (Heat alerts over time - Figure 1), Hwalert_figure1_3.xlsx (Heat alerts per province - Figures 2 and 3) and agile_fig4.xlsx (Maputo and Cabo Delgado weather stations - Figure 4). Based on these spreadsheets, we were able to verify the connection of the figures with the data, but there were not enough instructions to produce the figures on our own based only on the initial dataset that was provided (Dataset 1). The correspondence author included a README to each one of these spreadsheets, which serves more like a short description, rather than a data transformation instruction sheet.

The README also points to two external data portals: 1) The Humanitarian Data Exchange portal for the shapefile with the administrative boundaries of Mozambique (<https://data.humdata.org/dataset/cod-ab-moz>), and 2) the National Centers for Environmental Information of the United States for the Weather station recorded Daily maximum temperatures <https://www.ncei.noaa.gov/access/search/datasearch/daily-summaries>. In the first portal we downloaded the file named *moz_adm_20190607b_SHP.zip* (accessed on March 6th, 2023). In the second portal, after clicking *Explore datasets* we filled the values "maximum temperature" for *What* and "Mozambique" for *Where*, as instructed by the README. From the automatically generated results by the portal, we selected the option *Search Data* of the *Global Historical Climatology Network - Daily (GHCN-Daily), Version 3* dataset. Then we downloaded the .csv files for the provinces Maputo, Xai-Xai, Panda, Inhambane, Chimoio, Beira, Tete, Quelimane, Nampula, Lichinga, and Pemba, as instructed by the README. These .csv files comprise the so called **Dataset 2**, uploaded also to the corresponding OSF component.

After the reproducibility review, the Github repository was forked and archived under the official Reproducible AGILE Github organization: https://github.com/reproducible-agile/AGILE_MZ_Heat.

For the reproduction of Figure 2, the correspondence author shared with the reproducibility committee one .mxd file (Map Exchange Document) and one shapefile that were necessary over e-mail.

Software used

For the reproduction of Figure 2, we used ArcGIS Pro 2.7.2, and for the verification of Figures 1, 3 and 4 we used Microsoft Excel of the Office Professional Plus 2019 suite. Both applications were running on Windows 10 Education operating computer.

Reproduced figures

All the reproduced figures are listed below.

Figure 1

The reproduced figure is quite similar to the one included in the paper, but not exactly the same. There are some differences of aesthetic nature, such as the font and the color scheme, that appears slightly

lighter on the reproduced figure. The original figure does not have a title. These differences are of minor importance. Furthermore, some of the plotted data for the heatwave impacts (on agriculture, health and livestock) are positioned differently in the reproduced figure. The connection of the impact representations to the data of the spreadsheet could not be verified. The rest of the data are in accordance with the original figure. This figure was extracted from the spreadsheet Hwalert_figure2.xlsx.

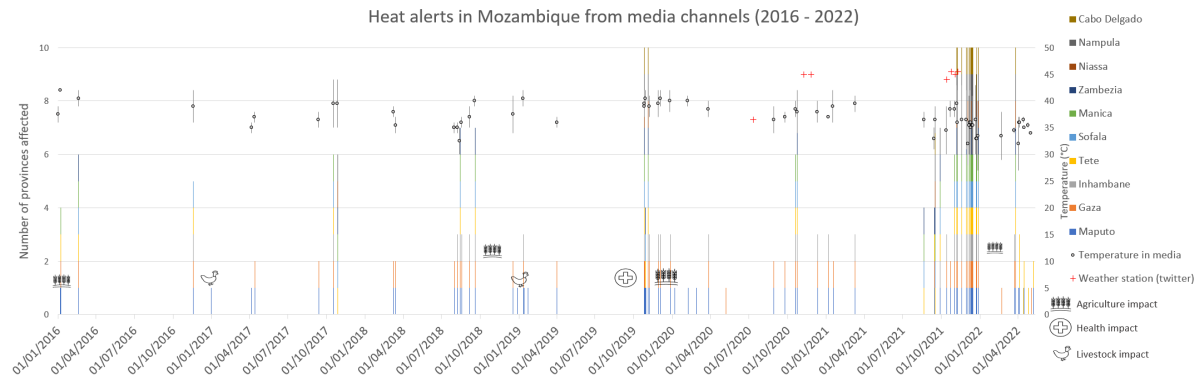


Figure 1: Corresponds to Figure 1 of the paper

Figure 2

The reproduced figure has a different title and also two additional subtitles for the layers shown in the legend. Both differences are considered of minor importance, since especially the latter could possibly be configured in the GUI of ArcGIS Pro. The two figures make use of different basemaps, which results in obvious (but consistent) differences in the color scheme of the sea and different citation on the bottom right corner of the figure. For the reproduction of this figure we had to connect three different files in ArcGIS Pro.

Figure 3

The reproduced figure demonstrates the same plotted data as the original one. The differences are minor and consist of the missing title, the fonts and the ranges of the axes. This figure was extracted from the spreadsheet Hwalert_figure1_3.xlsx.

Figure 4

This figure has two parts, 4a (upper - concerning Cabo Delgado) and 4b (lower - concerning Maputo city). The extracted figures from spreadsheet agile_fig4.xlsx seem to depict the same data as the original, as far as we could judge with the low quality of the image that was embedded in the paper. The differences are minor and include an additional legend under Figure 4a, different date format in the x-axes and missing numbering for the subfigures (a and b). We were not able to verify the connections of the figures with the data in the spreadsheet as we did for the rest of the figures, possibly due to a “Source not found” error in Microsoft Excel.

Reproduced table

Apart from the figures, we reproduced Table 2 of the paper, which gathers information about extreme temperatures in Mozambique from a specific Twitter account (@extremetemps - <https://twitter.com/extremetemps>). We used the *Advanced Search* feature of Twitter (<https://twitter.com/search-advanced>), filling in the queries “Mozambique” for *All of these words*, “@extremetemps” for *From these accounts*, “1/1/2016” for *From* and “1/1/2022” for *To* Dates (excluding replies of the account to tweets of other accounts). The steps for this reproduction were assumed based on the methodology described in the paper. The data collection on our side happened on March 2nd, 2023. We were able to verify 6 out of

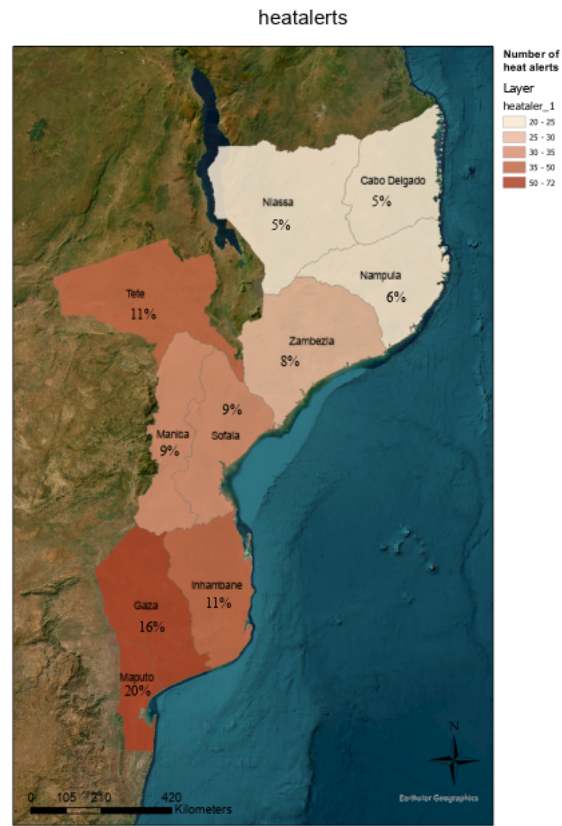


Figure 2: Corresponds to Figure 2 of the paper

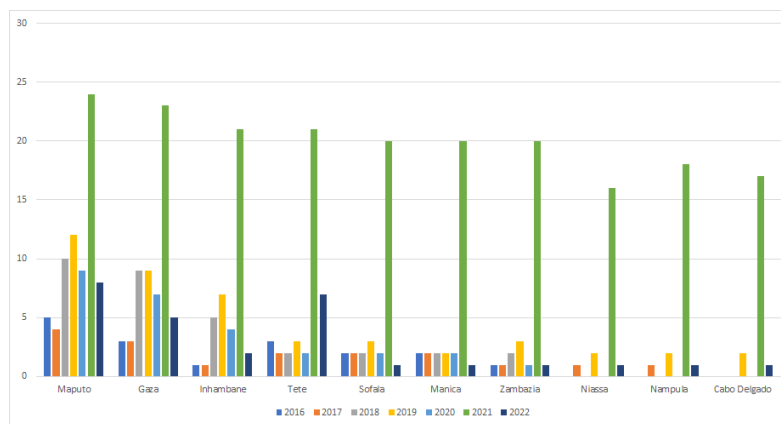


Figure 3: Corresponds to Figure 3 of the paper

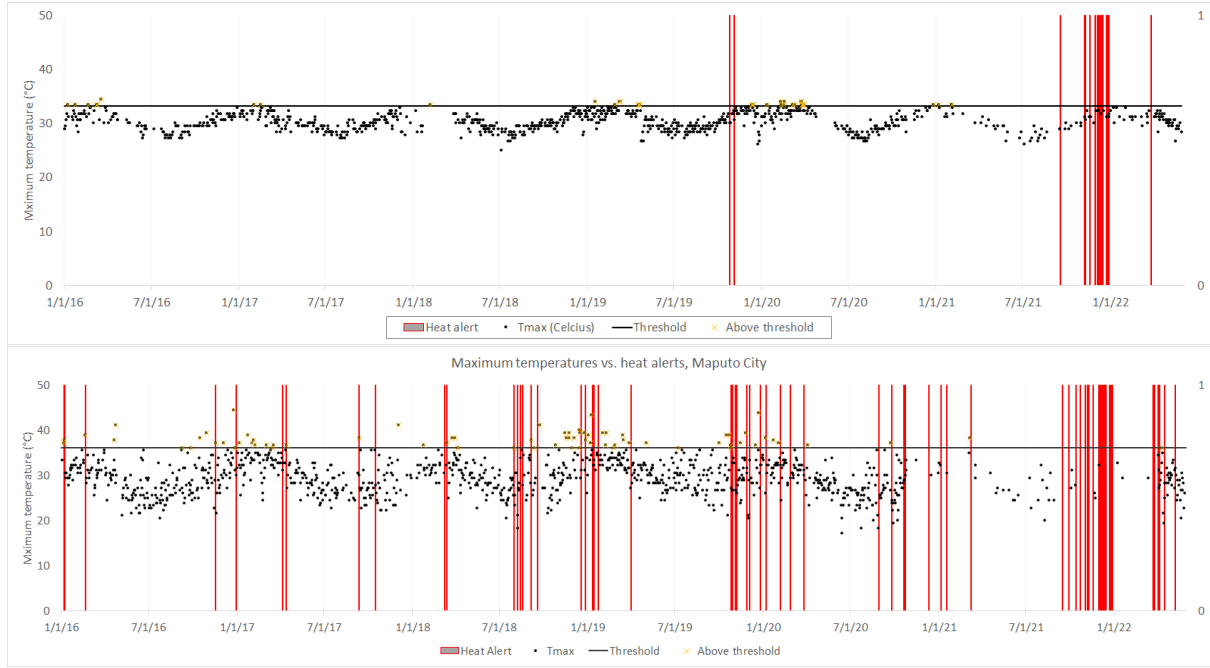


Figure 4: Corresponds to Figure 4 of the paper

7 entries of the table and we informed the correspondence author accordingly, who answered that this entry would be eventually be removed from the final version of the manuscript. This should also affect Figure 1, as it plots data from Twitter. The rest of the records are in accordance with the ones reported in the paper with regards to the date and the side notes. We identified a small numerical difference for the temperature record of 03/11/2021 and we could not verify the geographical subregion for the record of 11/07/2020. The Table can be seen below:

Table 1: Corresponds to Table 2 of the paper

Date	Location	Maximum temperature (°C)	Notes
11/07/2020		36.5	new national July Tmax
08/11/2020		45	
25/11/2020	Northern Mozambique	45	
24/10/2021	Tete	45.5	Southern hemisphere highest
03/11/2021		44.2	
09/11/2021	Tete	45.5	Southern hemisphere highest

Communication with the author

To achieve such level of reproducibility for this paper, multiple rounds of communication with the correspondence author were required. Although the Github looked much simpler at the beginning of the reproducibility review, the correspondence author was very cooperative and willing to share the necessary information with the reproducibility committee upon request. This resulted to a considerably improved Github README that will hopefully be helpful to future data reuse and methodology replications.

Reproduction efforts

The reproducibility review took more than four full days (expanded to several weeks), which is much longer than expected. The figure reproduction requirements were not clear from the beginning, the instructions for (re)producing them were lacking or even misleading at times. This resulted to several reproduction efforts that eventually failed. After communicating the failed reproduction results to the

correspondence author, they helped us by providing more clues for eventually verifying (and partially reproducing) Figures 1-4 and Table 2. Furthermore, not all datasets were mentioned in the first version of neither the *Data and Software Availability* section nor the Github README.