WEPPCLIFF: A command-line tool to process climate inputs for soil loss models

Ryan P. McGehee[1, 2], **Dennis C. Flanagan**[1, 2], **and Puneet Srivastava**[3]

1 Purdue University **2** USDA ARS National Soil Erosion Research Laboratory **3** Auburn University

DOI: 10.21105/joss.02029

Software

- Review
- Repository
- Archive

Editor: Arfon Smith

Reviewers:

- @kbarnhart
- @marianadobre
- @baharmon

Submitted: 02 December 2019
Published: 11 May 2020

Summary

A key driver of erosion, and arguably the most labor-intensive input to erosion models, especially the Water Erosion Prediction Project model (WEPP, Flanagan & Nearing, 1995), is climate. Even the simplest erosion models require climate inputs which are not always straightforward in their calculation. The complexity of these inputs dissuades potential users from creating their own inputs from observed data, which are almost always better than their simulated counterparts, especially when used to force erosion models for historical periods. For example, the Revised Universal Soil Loss Equation version 2 (RUSLE2, *User's reference guide*, 2013) and WEPP are two of the leading soil erosion prediction models in the world today, but both still rely on climate inputs that are generated from outdated data, tools, or both.

It appears that the lack of sufficient software in this field has resulted in models being used with suboptimal climate data inputs, which can, if left unaddressed, stifle scientific discovery and advancement. RUSLE2, as currently applied commonly within the United States, relies on climate data from 1960-1999. This data is not only 20 years out of date, but the years utilized had fewer observation stations and the stations themselves used a mix of older technologies, some of which have documented quality issues that impact usability and reliability (Hollinger, Angel, & Palecki, 2002; McGehee & Srivastava, 2018; *User's reference guide*, 2013). WEPP on the other hand, just had its climate database updated to the years 1974-2013 (Srivastava, Flanagan, J. R. Frankenberger, & Engel, 2019), but those observed values are only used to create parameter files for CLImate GENerator (CLIGEN, Nicks, Lane, & Gander, 1995) which is then used to create simulated climate files for WEPP. Some internal investigations at the National Soil Erosion Research Laboratory (NSERL) have identified the climate inputs to both models as a significant reason for differences in their predictions of soil loss under otherwise identical simulations.

In short, the erosion modeling community is in need of software that can at least shoulder most of the burden when preparing climate inputs for soil loss models. Therefore, we are proposing a WEPP CLImate File Formatter (WEPPCLIFF) program to begin addressing this need.

`WEPPCLIFF` is an R-based command line tool which was originally designed to prepare 'breakpoint format' climate file inputs for WEPP (Figure 1) that has been extended to perform other general functions such as storm separation, quality checking, gap filling, and erosion index (EI) calculations (climate inputs for the USLE family of models). The program is also provided with extensive accompanying documentation which covers a range of topics including most notably: installation, syntax, input, output, and examples.

The Journal of Open Source Software

```
0.0
  1   1   1
  Station: KMQE
Latitude Longitude Elevation (m) Obs. Years   Beginning Year  Years Simulated
 42.2167  -71.1167     199.00         19          2000              18
Observed monthly ave max temperature (C)
  0.9   2.2   6.0  13.2  18.7  22.6  26.5  25.9  22.0  15.1   9.2   3.5
Observed monthly ave min temperature (C)
 -6.6  -5.7  -1.8   3.6   9.3  14.1  17.6  17.4  13.7   7.5   2.3  -3.3
Observed monthly ave solar radiation (Langleys/day)
176.4 248.3 337.8 489.9 586.8 608.2 655.4 568.7 456.4 314.5 206.3 155.9
Observed monthly ave precipitation (mm)
106.7 128.7 128.0 105.3  95.8  88.6  78.6  84.0  97.1 121.8  97.1 124.6
 Day Month Year  Breaks  T-Max  T-Min   Rad   W-Vel  W-Dir  T-Dew
                   (#)    (C)    (C)   (L/d)  (m/s)  (Deg)   (C)
  1    1  2000     0      7.6   -3.6  201.9    7.1  223.8   -3.3
  2    1  2000     2     10.6    2.9  142.6    7.1  207.0    4.2
10.833   0.00
11.333   0.25
  3    1  2000     4     16.9    5.6  152.9    7.5  198.3    7.1
17.317   0.00
18.300   0.25
18.317   0.26
19.300   0.51
  4    1  2000    94     15.5    4.4  153.5    9.2  153.8    8.8
 0.650   0.00
 1.150   0.14
 1.583   0.37
 1.850   0.44
 2.083   0.57
 2.850   0.76
 3.950   0.76
 4.950   1.02
 8.467   1.02
 9.467   1.27
13.950   1.27
14.950   1.52
17.417   1.52
17.917   1.71
18.083   1.92
18.217   2.09
18.650   2.46
18.850   2.69
18.983   2.95
19.067   3.21
```

Figure 1: An example 'breakpoint format' climate input file for WEPP, which was prepared by WEPPCLIFF. More information on the format of WEPP climate inputs can be obtained from its documentation at: https://www.ars.usda.gov/ARSUserFiles/50201000/WEPP/usersum.pdf

Advanced Functionality

After the initial core functions of WEPPCLIFF had been developed, several more advanced functions were added to increase the utility of WEPPCLIFF (and to enhance its ability to deliver high quality climate inputs to WEPP and potentially other models as well). Each of the more impactful features are discussed below with figures when applicable. All of the following figures were produced with WEPPCLIFF standard plots.

Two of these features are more general and are used to refine the time series inputs for all variables supported in WEPPCLIFF (i.e. quality checking and gap-filling). WEPPCLIFF quality checking uses a combination of a) hard physical limits, b) adaptive physical rate change limits, and c) adaptive 'spiking', 'streaking', 'sticking', and frequency (return period) analyses to determine potentially faulty or erroneous data. These analyses are performed for precipitation and non-precipitation variables separately. An example quality checked input using default parameters is provided in Figure 2. At the time of publication, a new frequency analysis method based on a generalized extreme value model (climextRemes, Paciorek, 2019) was being incorporated in WEPPCLIFF source code; therefore, that functionality was not active for WEPPCLIFF at publication.

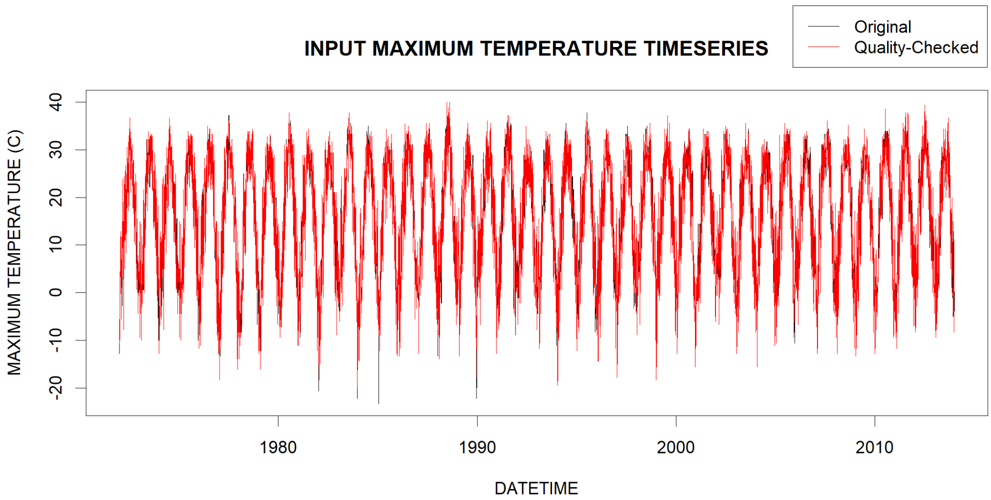

Figure 2: An example quality checked input timeseries (daily resolution) of maximum temperature data for a typical midwest climate station using default quality checking parameters.

Gap-filling in WEPPCLIFF is based on a multiple imputation model (MICE, van Buuren & Groothuis-Oudshoorn, 2011) which was configured to utilize all available data (within the same time series and from correlated multivariate relationships to other variables) to inform gap-filling of the missing time series. Figure 3 shows the same quality checked weather station from Figure 2 before and after gap-filling (using default parameters for the 'quick impute' option). At the time of publication, there was still some need to tweak each of these functions so that they would work seamlessly with the many different data types (temporal resolutions and formats) and diverse climate conditions (severe weather, droughts, nonstationarity, etc.). These functions will be refined as new information becomes available about how to improve them. Figures 4 and 5 show examples of how varied data missingness, seasonality, magnitude, variability, etc. can be for any two given stations. All other quality checking and gap-filling features were the original work of the WEPPCLIFF lead-developer.

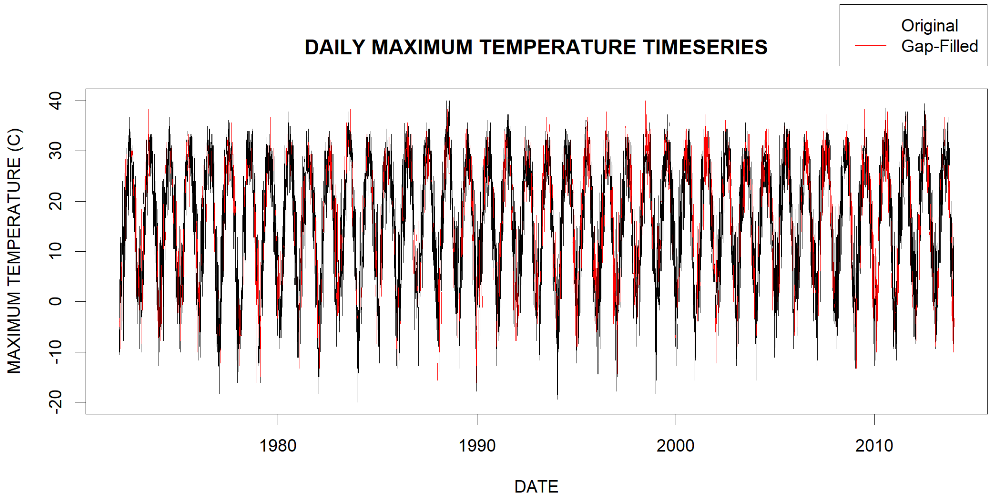

Figure 3: An example quality checked output timeseries (daily resolution) of maximum temperature data for the same station using default gap filling parameters for the 'quick impute' option.

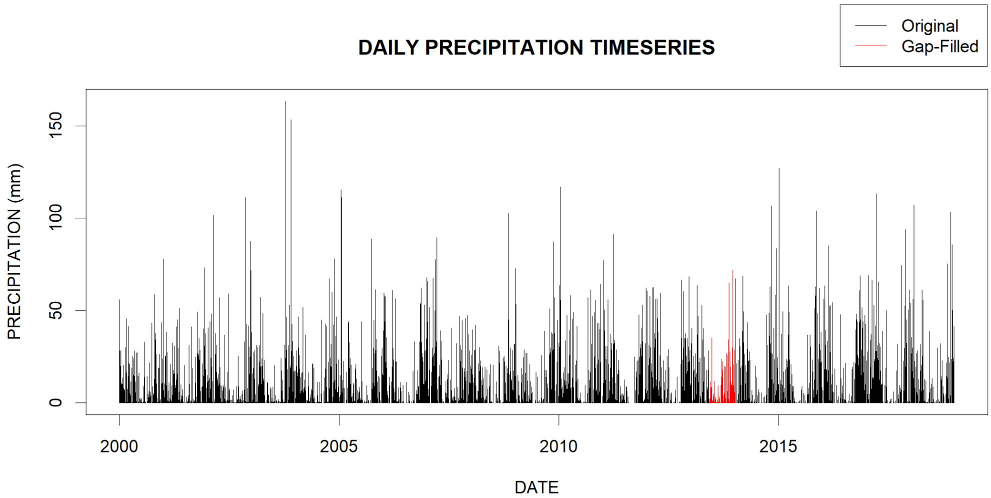

Figure 4: An example gap-filled precipitation time series with heavier, more seasonal rainfall from a longer time period with limited missingness.

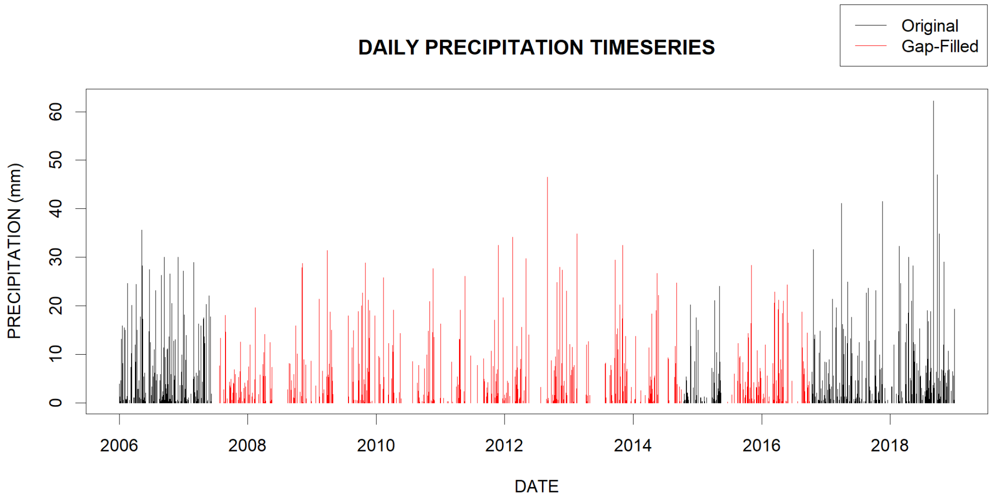

Figure 5: An example gap-filled precipitation time series with lighter, more variable rainfall from a shorter time period with rampant missingness.

The other two advanced features are specific to soil loss modeling: storm separation and erosion index (EI) calculation. These features provide a critical service from WEPPCLIFF enabling the production of empirical erosivity (more commonly known as R, R-factor, or EI) model inputs from the same time series inputs to physical models. Therefore, the results from WEPPCLIFF can be used to prepare identical climate inputs for empirical and physical models (e.g. WEPP and RUSLE2) so that their various outputs can be compared. The definition of a storm was based on soil loss modeling literature, which utilizes a time and depth threshhold to determine breaks between storms. These are directly specifiable in WEPPCLIFF, and they essentially control how many different storms (with the same I30) should be used to calculate EI. The EI calculation in WEPPCLIFF follows all the rules for various methods in erosion modeling literature, but most importantly, WEPPCLIFF allows the user to specify their own energy equation (a valid R expression based on breakpoint intensity input) for the exploration of new kinetic energy calculation methods and their impact on EI.

Existing Software

Historically, when providing climate inputs to WEPP, there were essentially two options. One must either a) use CLIGEN to prepare a daily 'tp-ip' format input file or b) manually prepare a 'breakpoint' format input file. The first option was the preferred method for providing climate inputs to WEPP, and it was originally the only way to prepare inputs for simulation with WEPP. The second option was added later in WEPP's development as a way to provide the highest resolution and best quality precipitation data for WEPP simulations. This enabled higher quality results primarily for single-storm simulation, but it was impractical for continous simulation for even a small period of time. `WEPPCLIFF` pushes this capability to the extreme by providing the capability to prepare relatively vast (format-intensive) climate inputs to force WEPP for extended continuous simulation while still preserving the most valuable components of breakpoint data (discussed in more detail later). Details on these two input formats can be found in either CLIGEN, WEPP, or WEPPCLIFF documentation.

The only other related software for providing climate inputs to soil loss models, of which we are aware, is a publicly distributed executable program called Rainfall Intensity Summarization Tool (RIST, Dabney & Justice, 2019). RIST is currently maintained by the USDA ARS National Sedimentation Laboratory (NSL) where it can be used as a tool to prepare climate inputs for RUSLE, WEPP, SWAT, and AnnAGNPS. However, the WEPP inputs for example, are only provided in the 'tp-ip' file format rather than the 'breakpoint' file format which means that only the intensity-at-peak factor, time-to-peak factor, and daily precipitation amount and duration are actually provided to WEPP to fit a double exponential storm distribution and finally generate breakpoints for erosion calculations. Therefore, much of the observed data is being discarded when providing 'observed climate inputs' to WEPP using RIST. WEPPCLIFF, on the other hand, maintains the original precipitation breakpoints in its climate file output thereby providing more valuable climate inputs to WEPP.

Validation

WEPPCLIFF was developed from and validated by comparing to several different works including: three Agriculture Handbooks (AH282 (Wischmeier & Smith, 1965), AH537 (Wischmeier & Smith, 1978), and AH703 (Renard, Foster, Weesies, McCool, & Yoder, 1997)), two peer-reviewed articles (McGehee & Srivastava, 2018; McGregor & Mutchler, 1995), and RIST (Dabney & Justice, 2019). The comparison with RIST is of particular interest since this was completed as part of a massive WEPP-RUSLE2 model comparison project between the NSERL and NSL ARS laboratories. This project resulted in several important findings related to climate representations in both models (WEPP and RUSLE2), their accompanying input generation softwares (CLIGEN, RIST, and now WEPPCLIFF), and the impacts on erosion predictions. An example comparison related to WEPPCLIFF found that RIST (April 2019 release version) erroneously calculated kinetic energy and RUSLE2 erosion indices to be 26% and 21% too low, respectively. RIST was updated with more correct calculations in May 2019 (there were still differences of about 3% for the stations compared at the time of this writing), but RIST has been actively used since at least 2011 with several scientific works, and this error may have impacted several or all of those works. More comprehensive summaries of findings from this comparison project are expected to be published in scientific literature in the near future.

References

Dabney, S., & Justice, V. (2019). Rainfall intensity summarization tool (rist) v3.99. Retrieved from https://www.ars.usda.gov/southeast-area/oxford-ms/national-sedimentation-laboratory/watershed-physical-processes-research/research/rist/rist-rainfall-intensity-summarization-tool/

Flanagan, D. C., & Nearing, M. A. (Eds.). (1995). *USDA water erosion prediction project hillslope and watershed model documentation.* West Lafayette, IN: USDA ARS.

Hollinger, S. E., Angel, J. R., & Palecki, M. A. (2002). *Spatial distribution, variation, and trends in storm precipitation characteristics associated with soil erosion in the united states* (Contract report CR No. 2002-08). Champaign, IL: Illinois State Water Survery.

McGehee, R. P., & Srivastava, P. (2018). Benchmarking reliable erosion indices from quarter-hour station data for climate studies in the southeastern united states. *Soil and Water Conservation*, *73*(4), 363–376. doi:10.2489/jswc.73.4.363

McGregor, K. C., & Mutchler, C. K. (1995). Erosivity index values for northern mississippi. *Transactions of the American Society of Agricultural Engineers*, *38*, 1039–1047. doi:10.13031/2013.27921

Nicks, A. D., Lane, L. J., & Gander, G. A. (1995). Chapter 2: Weather generator. In D. C. Flanagan & M. A. Nearing (Eds.), (Vols. USDA Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation, NSERL Report No. 10). West Lafayette, IN: USDA ARS.

Paciorek, C. J. (2019). ClimextRemes: Tools for analyzing climate extremes. Retrieved from https://cran.r-project.org/package=climextRemes

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K., & Yoder, D. C. (1997). *Predicting soil erosion by water: A guide to conservation planning with the revised universal soil loss equation (rusle)*. Washington, DC: USDA ARS.

Srivastava, A., Flanagan, D. C., J. R. Frankenberger, & Engel, B. A. (2019). Updated climate database and impacts on wepp model predictions. *Soil and Water Conservation*, *74*(4), 334–349. doi:10.2489/jswc.74.4.334

User's reference guide: Revised universal soil loss equation, version 2. (2013). Washington, DC: USDA ARS.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1–67. doi:10.18637/jss.v045.i03

Wischmeier, W. H., & Smith, D. D. (1965). *Predicting rainfall erosion losses from cropland east of the rocky mountains*. Washington, DC: USDA ARS.

Wischmeier, W. H., & Smith, D. D. (1978). *Predicting rainfall erosion losses*. Washington, DC: USDA ARS.