Detecting marine heatwaves

Robert W. Schlegel, Eric C. J. Oliver, Alistair J. Hobday, Albertus J. Smit

25 May 2018

Abstract

It is now known that marine heatwaves (MHWs) have been increasing in duration and intensity globally. It is therefore necessary that the detection of these events in all areas of the ocean be made possible. This includes, but is not limited to, areas with time series that are shorter than the prescribed 30 years, missing large amounts of random or consistent data, or have been collected with instrumentation from which the appropriate meta-data were not maintained. The best practices for how to deal with these issues have been investigated and outlined in detail here. Additionally, the use of alternative climatologies is investigated and the benefits of differing methods is discussed. This is all worked out within the framework of specific case studies, noting the pitfalls inherent in this field of research.

# Introduction

The phenomenon: what marine heatwaves are, why they are a problem, and what has environmental impacts have taken place as a result of them?

The context: framing problem i.t.o. previous work around in the area, and in particularly w.r.t. atmospheric heatwaves. Atmospheric heatwaves were defined with regards to human health concerns; marine heat waves resulted as the result of marine ecological impacts. The marine heat wave definition.

What is known about marine heat waves: a quick, one-paragraph synopsis of events globally; where, comparison of metrics, consequences.

Problem statement: what are some of the remaining problem we face when quantifying marine heat

waves? There are problems around the limitation of the data themselves, such as:

*Duration of the data series.* The duration of a time series does not affect the detection of marine heat waves *per se*, but it does affect the creation of a daily climatology relative to which the events can be detected. A climatology serves two main roles (WMO, 2017). First, it serves as a ‘benchmark’ relative to which past and future measurements can be compared, and against which anomalies can be calculated. Second, they also reflect the typical conditions likely to be experienced at a particular place at particular time. The WMO technical guidelines (WMO, 2016) stipulate that daily climatologies (which they call ‘climate normals’) must be based on the most recent 30-year period up until a year ending with 0 (currently 1981-2010). Currently most studies on marine heat waves have used the globally comprehensive NOAA daily optimum interpolated SST (dOISST v2; Reynolds et al., 2007) that combines *in situ* and satellite data on a 0.25° spatial grid. These data are updated in near real time and start on 1 September 1981. Due to the fact that the dOISST time series only starts in September 1981, the climatological base period is 1982-2011. One limitation of this data set stems from its coarse resolution—the 0.25° grid dimension limits the use of the data set closer to coastal areas, where coastal physical processes introduce large amounts of additional variance into the SST field, causing it to deviate from that of the dOISST data nearest to the coast (refs.). Many gridded daily SST products are available that may make them more useful closer to the coast (but see Smit et al., 2013; etc.) due to their finer grid size (1-4km, e.g. MODIS AQUA, MODIS TERRA, AVHRR Pathfinder v5.2, G1SST, MUR, etc.; refs.), but the data do not yet cover a full 30 year period.

*Missing values*. Some of the aforementioned data products (mention which…) also suffer from ‘gappiness’ that results from NAs being introduced due to cloud cover, unsuitable sea states, etc., which become more prevalent at smaller scales, particularly nearer the coast. Some products smooth out these influences, but this results in excessively smooth SST fields that mask some of the small-scale spatial variation in surface temperatures. Other times they rely on blending with data from other products, which may have its own suite of consequences.

*Low temporal resolution.* About data sets that come in weekly or monthly temporal resolution. These may be useful when daily data have too many NAs (e.g. AVHRR Pathfinder, MODIS, and MERIS data). Can we use weekly and monthly data? What has been done along these lines?

If we find that we can use shorter time series to detect events in, do they show the same events as the ones found in the dOISST data? If so, are the same Threshold, 2 × Threshold, 3 × Threshold, and 4 × Threshold events present in all of the data sets? (I think that correspondence will be better for the larger events, but the idea is to produce some sensitivity analysis or something). We can provide guidelines about which suitable shorter time series data can/should be used for MHW detection, and how to select the best climatology creation method.

There’s also unknowns around the best settings to use in the MHW algorithm. There are many default settings, but under some circumstances users might need to deviate from these. When? Why? What will the consequences be?

Under some situations users have asked to be able to insert their own baselines and climatologies. What are the consequences, and how does one best do this?

# Material and Methods

## Data

### Reference time series

Focus on the three time (dOISST v2; Reynolds et al., 2007) series included with the packages: Western Australian, NW Atlantic, and Mediterranean (henceforth reference time series).

### Shorter duration, higher resolution gridded data

It might be useful to show that in regions where events (at a certain threshold) can be detected in the dOISST data, that they also are present in the higher-res, shorter duration SST products. The we can show that in some scenarios the hi-res, short time series additionally capture some events that are not present in the OISST data due to its coarse spatial grid size.

Maybe insert a table here that lists all of the data sets that might be potentially useful, if we can overcome some of their limitation; in the table, also mention each product’s limitations (e.g. too short, too many NAs, etc.).

### Remotely sensed data

### Reanalysis data

### *In situ* data

## Assessing the effect of time series duration

* Systematically reducing time series length of the three reference time series—i.e., from 30-year to 20-year and 10-year.
* Bootstrapping (I have used 100 resamples, but this can be adjusted upwards if necessary).

### Standard climatologies (i.e. 11-day windowHalfWidth + 31-day smoothPercentile)

* Also assess the effect of systematic varying windowHalfWidth and smoothPercentile and studying the outcomes for the three time series lengths.
* Measurement metrics:
  + for each day-of-year (doy) in the climatology, calculate the SD of the climatological means of the 100 bootstrapped samples;
  + for each doy, calculate the RMSE of the boostrapped means relative to the true climatology (i.e. the one produced from the 30-year long time series);
  + correspondence of detected events when using climatologies calculated from reduced time series vs. when using the full duration time series climatologies.

### Fourier transform climatologies

* As in 2.3.1.

### Analysis of short-duration, high resolution gridded SSTs

* Comparison of detected events:
  + compare reference time series vs. other co-located SST data;
  + compare in special conditions where events may be expected, but are not present in the dOISST data due to constraints resulting from it not being of high enough resolution; e.g. in upwelling regions, embayments, etc.

## Assessing the effect of random missing data

## Assessing the effect of non-random missing data

## Using alternative climatologies

## Best practices

After the investigation into the aforementioned topics has been completed, a series of best practices for dealing with these issues may be discussed.

(RWS: Ideally these could also be retroactively worked into the languages to provide them as options in the code for users.)

# Results

## R vs Python

A host of analyses were performed to ensure that the language used in the following sensitivity tests could be performed in either distribution of the MHW detection algorithm. This included comparisons of the default outputs, how changing the arguments effected the default outputs, as well as a comparison of the other functionality provided between the two languages. It was found that while some style differences exist between the added functionality of the languages, the core climatology outputs are identical to within < 0.001 per measurement. This established that results obtained with either language are comparable.

## Assessing the effect of time series duration

## Assessing the effect of random missing data

## Assessing the effect of non-random missing data

## Alternative climatologies

# Discussion

A brief mention of existing similar packages, and how what heatwaveR and the python version do things differently.

## Time series length

## Missing values

## Best practices

### Technical

### Scientific

## Pitfalls

# Conclusions

# References

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22), 5473-5496.

Smit, A. J., Roberts, M., Anderson, R. J., Dufois, F., Dudley, S. F., Bornman, T. G., ... & Bolton, J. J. (2013). A coastal seawater temperature dataset for biogeographical studies: large biases between in situ and remotely-sensed data sets around the coast of South Africa. *PLoS One*, *8*(12), e81944.

WMO (2016). Technical Regulations, Basic Documents No. 2, Volume I – General Meteorological Standards and Recommended Practices (WMO-No. 49). 2015 edition, updated in 2016. Geneva.

WMO (2017). WMO Guidelines on the calculation of Climate Normals (WMO-No. 1203), pp. 1-18.