Effects of Natural Variability of Seawater Temperature, Time Series Length, Decadal Trend and Instrument Precision on the Ability to Detect Temperature Trends

Robert Schlegel* and Albertus Smit

Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, Republic of South Africa

**Corresponding author address:* Robert Schlegel, Department of Biodiversity and Conservation Biology, University of the Western Cape, Bellville, Republic of South Africa.

E-mail: 3503570@myuwc.ac.za

ABSTRACT

In South Africa 129 *in situ* temperature time series of 1 to 43 years are used for investigations of the thermal characteristics of coastal seawater. They are comprised of temperature recordings at precisions ranging from 0.5 °C to 0.001 °C and collected with handheld thermometers or underwater temperature recorders (UTRs). Using the naturally occurring range of seasonal signals, variability and temperature trends for 84 of these time series, the length, decadal trend and data precision of each time series were systematically varied before fitting generalized least squares (GLS) models to study the effect these variables have on trend detection. We determined that the variables contributing most to accurate trend detection in decreasing order are: the length of the time series, the decadal trend, variance, amount of missing data and the precision of the measurements. We found that time series at least 20 years in length may be used tentatively for climate change research, but that time series >30 years in length are preferable. The implication is that long-running thermometer time series in this dataset, and others around the world, are more useful for decadal scale climate change studies than the shorter, more precise UTR time series. It is important to note that due to the nature of the dataset used in this study, instrument drift was not able to be quantified.

1. Introduction

The roughly 3,000 km of South Africa's coastline is bordered by the Benguela and Agulhas currents (*e.g.* **??**), which, in combination with other nearshore processes, affect the country's marine coastal ecosystems (**?**). A thorough understanding of these coastal processes is provided by several physical variables, of which temperature is one of the main determinants (*e.g.* **???**). In order to ensure a true representation of organisms' biological thermal limits, nearshore temperatures must be accurately recorded and monitored. Some sources warn of the pitfalls in doing so *RWS: Add references here showing which sources say using SST for the coast is inappropriate*, and a study by **?** showed that SST data have a warm bias as large as 6 °C when compared to coastal *in situ* data. Nevertheless, a widespread approach in coastal ecological research is to use satellite and/or model-generated temperature data as a representation of the sea surface temperature (SST) along coastlines (*e.g.* **???**), because apparently the dangers of applying gridded SSTs to the coast are not widely known or in many places in the world there simply are no suitable *in situ* coastal temperature time series available. It is for this reason that we strongly recommended the use of *in situ* data to support research conducted within 400 m from the shoreline.

Where records of *in situ* coastal seawater temperature do exist, the reliability of many of these datasets that could be used in place of the remotely-sensed SST data remains to be verified. Users of SST data benefit from it being refined through a number of well documented validation and quality control processes (*e.g.* **???**), whereas the standards and methods with which local *in situ* data from a single dataset are collected and refined may differ greatly. For example, there are currently seven organizations and/or governmental departments (hereafter referred to as bodies) contributing coastal seawater temperature data to the South African Coastal Temperature Network (SACTN). These bodies use different methods and instruments to collect their data as no national standard has been set. One consequence of this methodological disparity is that two thirds of the data were sampled with hand-held thermometers that are manually recorded at a data precision of 0.5 °C, as opposed to the current generation of Underwater Temperature Recorders (UTRs) with an instrument precision of down to 0.001 °C. If these *in situ* data are to be used together *in lieu* of the satellite-based SST data, it is important that the characteristics of the contributing data sources are understood in terms of their ability to yield useful, reliable and accurate long-term measurements for use in climate change studies.

This prompted us to examine the 129 *in situ* time series that comprise the SACTN. The range of measurement precisions and statistical characteristics of this dataset were used to guide a series of enquiry-driven analyses into the suitability of the time series to yield statistically significant and accurate assessments of decadal temperature change. The length, decadal trend and data precision of each time series were adjusted in a systematic manner, and forms the core of our analyses. Our aim was to assess the effect that each of these variables has on the ability of a model to produce a robust estimate of time series decadal trend. The effect gaps in the time series may have on the fitting of models was also investigated as many of the time series used here have some missing data scattered throughout, which is unavoidable for a 20+ year time series that is sampled by hand by a single technician at each site.

The study provides a better understanding of some of the determinants of a time series that are influential in the detection success of decadal trends in coastal ocean temperature time series.

2. Methods

a. Data Sources

Our study lies within the political borders of South Africa's coastline. The location of each point of collection appears in Figure 1. Of these 129 time series, 43 are recorded with UTRs and the other 86 with hand-held mercury thermometers. The oldest currently running time series began on January 1st, 1972; there are 11 total time series that started in the 70s, 53 more started in the 80s, 34 began in the 90s, 18 in the 00s and 13 in the current decade.

The data are collected using two different methods and a variety of instruments. Hand-held mercury thermometers (which are being phased out in favor of alcohol thermometers or electronic instruments) are used in some instances at the shoreline, and represent seawater temperatures at the surface. At other places, predominantly along the country's east coast, data are collected with glass thermometers from small boats at the location of shark nets along the coast (**?**). Whereas both types of thermometers allow for a measurement precision of 0.1 °C, the recordings are written down at a precision of 0.5 °C. Data at other localities are collected using delayed-mode instruments that are permanently moored shallower than 10 m, but generally very close to the surface below the low-water spring tide level.

Over the last 40+ years the electronic instruments used to measure coastal seawater temperatures have changed and improved. The previous standard was the Onset Hobo UTR with a thermal precision of 0.01 °C. The new standard currently being phased in is the Starmon Mini UTR. These devices have a maximum thermal precision of 0.001 °C $\pm$ 0.025 °C (http://www.star-oddi.com/). Of the 43 UTR time series in this dataset, 30 were recorded at a precision of 0.001 °C for their entirety, five UTR time series include older data that were recorded at a precision of 0.01 °C or

0.1 °C and so have been rounded down to match this level of precision. Eight additional UTR time series have older data that were recorded at a precision of 0.1 °C.

The thermometer data are recorded manually and saved in an aggregated location at the head offices of the collecting bodies. UTRs are installed and maintained by divers and data are retrieved at least once annually. These data are digital and are downloaded to a hard drive at the respective head offices of the collecting bodies.

b. Data Management

Each of the seven bodies contributing data to this study have their own method of data formatting. Steps are being taken towards a national standard as we move towards replacing all the thermometer recordings with UTR devices; however, as of the writing of this article, one does not yet exist. Data from each organization were formatted to a project-wide comma-separated values (CSV) format with consistent column headers before any statistical analyses were performed. This allowed for the same methodology to be used across the entire dataset, ensuring consistent analysis. Before analysing the data they were scanned for any values above 35 °C or below 0 °C. These data points were changed to `NA`, meaning ‘not available’, before including them in the SACTN dataset.

All analyses and data management performed in this paper were conducted with R version 3.3.1 (2016-06-21) (**?**). The script and data used to conduct the analyses and create the tables and figures in this paper may be found at https://github.com/schrob040/Trend_Analysis.

Any time series with a temporal precision greater than one day were averaged into daily values before being aggregated into the SACTN. A series of additional checks were then performed (*e.g.* removing long stretches in the time series without associated temperature recordings) and time series shorter than five calendar years or collected at depths greater than 10 m were removed.

At the time of this analysis, this useable daily dataset consisted of 84 time series, consisting of 819,499 days of data; these data were then binned further to the 26,924 monthly temperature values available for use in this study.

c. Systematic Analysis of Time Series

We used the 84 time series simply for their variance properties (comprised of seasonal, inter-annual, decadal and noise components), which reflect that of the thermal environment naturally present along the roughly 3,000 km of South African coastline. Linear trends that may have been present in each time series were removed prior to the ensuing analysis by applying an ordinary least squares regression and keeping the detrended residuals as anomaly time series. In doing so we avoided the need to simulate a series of synthetic time series, whose variance components may not have been fully representative of that naturally present in coastal waters. These detrended anomaly time series (henceforth simply called 'time series') represent a range of time scales from 72 to 519 months in duration.

To each of the 84 time series we artificially added linear decadal trends of 0.00 °C to 0.20 °C dec^{-1}. In other words, we now had time series that captured the natural thermal variabilities around the coast, but with their decadal trends known *a priori*. The range of decadal trends was selected based around the global average of 0.124 °C from **?** and used in **?**. Furthermore, in order to represent the instrumental precision of the instruments used to collect these time series, we rounded each of these (84 time series $\times$ 5 decadal trends) to four levels of precision: 0.5 °C, 0.1 °C, 0.01 °C and 0.001 °C. Consequently, we had a pool of 1,680 time series with which to work.

To gain further insight into the effect of time series length on trend detection, each time series was first shortened to a minimum length of 5 years, starting in January so that the timing of the seasonal signal for each time series would be equitable. After fitting the model (see *Time Series Model*, below) to all 1,680 of the shortened time series, the next year of data for each time series was added and the models fitted again. This process was iterated until the full length of each time series was attained. For example, if a time series consisted of 12 full years of data, it would require 160 models (8 iterations of increasing length $\times$ 5 decadal trends $\times$ 4 levels of precision); similarly, 720 models would be applied to a 40 year time series. Considering the 84 time series available, the total number of individual models required to capture each combination of variables quickly increased to 36,220.

In order to deal with `NA`s present in some of the time series, we initially replaced these with linearly interpolated temperature values. It turned out that this was a terrible idea because doing so resulted in artificially increasing the goodness of fit of the detected trend: the degree to which this 'improvement' occurs is proportional to the amount of interpolation applied and to the size of the linear decadal trend added (see Appendix A). The analysis presented here therefore proceeded with non-interpolated data only.

Our approach of fitting models to each of the semi-artificial time series that we generated allowed us to study the effect that the relevant variables (time series length, natural variability, added slope and level of measurement precision) has on the ability of the time series model to faithfully detect the decadal thermal trend, which was known *a priori*. The primary results of interest in these analyses were the significance (p-value) of the model fit, the accuracy of the decadal trend determined by the GLS model as well as the error associated with the trend estimate.

d. Time Series Model

The selection of the appropriate model can greatly influence the ability to detect trends **?**. Two broad approaches are widely used in climate change research (**?**). The first group of models estimates linear trends, and although linearity may not reflect reality (*i.e.* trends are very frequently non-linear), these models do provide the convenience of producing an easy to understand decadal trend (*e.g.* 0.106 °C dec^{-1}; **??**). The other group accommodates non-linear trajectories of temperature through time by the use of higher-degree polynomial terms or non-parametric smoothing splines, but the inconvenience comes from not being able to easily compare models among sites (**??**). Both groups of models can accommodate serially correlated residuals, which is often the cause for much criticism due to their effect on the uncertainty of the trend estimates (**??**). For example, Generalized Least Squares (GLS; yielding estimates of linear trends) and Generalized Additive Mixed Models (GAMM; non-linear fitting with no trend estimate provided) can both capture various degrees of serial autocorrelation (**??**). Although our exploratory analysis assessed two parameterizations of each of the model groups, we opted to proceed here with a GLS equipped with a second-order autoregressive AR(2) correlation structure fitted using Restricted Maximum Likelihood (REML; **?**):

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

where the lag-2 autocorrelated residuals are given by

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + w_t$$

and the white noise series is

$$w_t \sim \text{i.i.d. } N(0, \sigma^2)$$

This is similar to that of the IPCC, although the latter uses an AR(1) error term (**?**). Another difference from the IPCC approach is that we nested the autoregressive component within year.

This modeling approach allowed us to assess how various properties of the detrended data sets would affect the models' ability to detect trends – in other words, by comparing the estimates of the trends themselves and how these deviate from the known trend.

3. Results

The residuals for the base 84 deterended time series may be seen in Figure 2. From these detrended time series the length, decadal trend and precision variables were systematically manipulated as explained in the methods. It was found that the important variables affecting the accuracy of the slope detected by the GLS model, in decreasing order, were: i) time series length; ii) the size of the added decadal trend; iii) initial SD of the time series (after detrending but prior to adding artificial slopes); iv) the amount of `NA`; and iv) measurement precision. These variables influenced the model fits in a systematic manner.

As would be expected, the size of the decadal trend estimated by the GLS increases in direct proportion to the decadal trend which we added and therefore knew *a priori*. What is especially noteworthy in this analysis is that time series of longer duration more often result in trend estimates converging with the actual trend than those of shorter length (Figure 3). This effect is most evident from around 30 years. Furthermore, how well the estimated model trend converges with the actual trend is also very visible in the standard error (SE) of the trend estimate (Figure 4): models fitted to short time series will always have modeled trends with larger SE compared to longer ones. The strength of this correlation is $r = 0.56$ ($p < 0.001$) and it remains virtually unchanged as the added decadal trend increases. The p-value of the fitted models also vary in relation to time series duration and to the steepness of the added decadal trend (Figure 5). It is usually the longer time series equipped with steeper decadal trends that are able to produce model fits with estimated trends that are statistically significant. Note, however, that this p-value tests the null hypothesis that the estimated trend is no different from $0\,^{\circ}\mathrm{C}\ \mathrm{dec}^{-1}$ at $p \leq 0.05$, and *not* that the slope is not different from the added trend. Taken together, these outcomes show that although our GLS model can very often result in trend estimates that *approach* the true trend, it is seldom that those estimates are statistically significant in the sense that the estimated trends differ statistically from $0\,^{\circ}\mathrm{C}\ \mathrm{dec}^{-1}$.

The variance of the detrended data is another variable that can affect model fitting, but its only systematic influence concerns the SE of the trend estimate. Here, it acts in a manner that is entirely consistent across all *a priori* trends (Figure 6). What we see is that as the variance of the data increases (represented here as standard deviation, SD) the SE of the slope estimates increases too. Moreover, it does so disproportionately more for time series of shorter duration. Again, as we have seen with the estimated trend that converges to the true trend around 30 years, so too does the initial SD of the data cease to be important in time series of around 3 decades in length.

The number of `NA`s permitted in any of our time series was limited to 15% per time series. Twenty-five of the 84 time series have fewer than 1% `NA`. An additional 45 time series have up to 5% `NA`, 10 have up to 10% `NA` and 4 have up to 15% `NA`. The mean number of `NA` for the data is 2.65%. The relationship between %`NA` and the *p*-value of the models is shown in Figure 7. At 2.5% or fewer `NA` their presence does not have any discernible effect on resultant *p*-values. Progressively increasing the number of `NA`s above 5%, however, leads to a drastic improvement of models fitted to series with no or gently increasing decadal trends (these generally have very large *p*-values indicative of very poor fits, perhaps due to the presence of a very weak signal), and a significant deterioration of models fitted to data with steep decadal trends (for these data, the model generally fits better at low numbers of `NA`s, as suggested by the greater number *p*-values that approach 0.05). In other words, the inclusion of missing values results in time series with no added decadal trend to veer away from $0\,^{\circ}\mathrm{C}\ \mathrm{dec}^{-1}$ towards a situation where they may erroneously appear to display a trend. On the other hand, time series that do indeed have decadal trends tend to produce fits that are not significantly different from 0 °C dec^{-1}.

Regarding the effect that the level of measurement precision has on the GLS models, we see in Figure 8 that decreasing the precision from 0.001 °C to 0.01 °C has an undetectable effect on any differences in the modeled trends. The Root Mean Square Error (RMSE) between the slopes estimated from 0.001 °C and 0.01 °C data is 0.001. The correspondence between the slopes estimated for data reported at 0.5 °C compared to that at 0.001 °C decreases to a RMSE of 0.03.

The effect of decreasing data measurement precision from 0.001 °C to 0.5 °C has almost no appreciable effect on any of the measures of variance presented in this study. The effect of measurement precision on the accuracy of the modeled slope, however, becomes very pronounced going from 0.1 °C to 0.5 °C. This effect is larger on smaller decadal trends. For example, at a trend of 0.05 °C dec^{-1}, the deviation from the true value of models fitted to data with a precision of 0.1 °C is negligible; however, the accuracy of the fitted model on data recorded at a precision of 0.5 °C with a real trend of 0.05 °C dec^{-1} is 10.81% different on average (*i.e.* given a slope of 0.05 °C dec^{-1} the model detects slopes of 0.055 40 °C dec^{-1}). This accuracy of the models improves to an average difference of 6.44% with a slope of 0.10 °C dec^{-1}, 2.24% at 0.15 °C dec^{-1} and decreases slightly to 2.30% at 0.20 °C dec^{-1}. A precision of 0.5 °C always provides clearly less accurate modeled trends than at higher precisions; however, the current analysis did not highlight one precision that consistently provides the most accurate estimate of the trends. This may however become determinable in an analysis of synthetic data with variance structures that are manipulated in a more consistent manner.

As the actual time series used to generate the data for this study are predominantly greater than 300 months in length and recorded at a data precision of 0.5 °C, we would be remiss not to investigate the interaction between the increase in accuracy provided by a lengthy time series, against the decrease caused by a data precision of 0.5 °C. In other words, at what point does a model fitted to a longer time series, with less precise measurements (*e.g.* those taken by thermometers and reported at a precision of 0.5 °C), become as accurate as a time series with more precise measurements (*e.g.* UTRs)? Figure 8 shows how varied the modeled trends become when a precision of 0.5 °C is used, and we see here that when these low resolution time series have a shallow slope of 0.05 °C dec^{-1}, a fitted model requires 24 months of additional data on average to have a comparable level of accuracy to a model fitted to data recorded at a precision of 0.1 °C. This difference in length decreases to 16 months when a larger slope 0.20 °C dec^{-1} is used.

An analysis with a large number of variables as shown here is bound to have a medley of complex interactions between the various statistics being measured; however, much of the range seen in the results of the GLS models can be well explained by the influence of one independent variable, or two operating in concert, as we have shown above. The most important of these variables has clearly been length.

4. Discussion

The strongest finding of this analysis is that the accurate detection of long-term trends in time series primarily concerns the length of a dataset. But there is also a host of nuances resulting from time series length, the steepness of the decadal trend the model is asked to detect, the influence of the SD of a time series, the amount of missing values and the precision at which the data have been measured or recorded that interact with one-another and which must be considered.

Whereas time series with smaller variances (shown as SD in this study) generally produce model fits that are statistically significant (*i.e.* with decadal trends that are significantly different from 0 °C dec^{-1} at $p < 0.05$) and with smaller SE of the estimated trends after shorter lengths of time, we also see that increasing a time series' length beyond 25 years, but preferably beyond 30 years, will increase the likelihood of detecting a decadal temperature change even in very variable data sets. Measuring temperature change in highly variable coastal environments, such as those around the coast of South Africa and many temperate coastal environments globally (refs.), will therefore benefit from access to the longest possible time series available. This phenomenon is demonstrated in Figure 4, which uses color to show the time series binned by the three different coastal sections of South Africa (**?**). Of these three coastal sections the east coast is known to have the most stable thermal regime, with the south coast having the most variability. A good experiemental design must account for the location of study, and we have shown here that location may have a large effect, requiring for adequate planning to collect a long enough time series.

The detection of long term trends require long-term data. This finding is both positive and negative. The length of a time series needed for climate change research is firmly under the control of the investigator with sufficient foresight and perseverance to plan the installation and management of new instrument networks that will yield usable results only after about three-quarters of a typical academic career has passed. Should such data already exist, and considering the scarcity of such long-term records that are already yielding benefits today, we must ensure that these sources of data are managed and curated with great care and diligence as they are practically irreplaceable. For this reason, it is essential that we understand the inherent strengths and weaknesses of such existing sources of data so that we may fully maximize their utility and extract from them the model coefficients needed for climatic research, and know their accuracy to the best of our ability. There are many time series < 20 years in length that should be avoided, where possible, for trend analysis. These will mature with time and their maintenance need to be ensured going forward.

Aside from length, the most powerful time series have measurements that are taken regularly. The inclusion of too many missing values (`NA`s) in the data sets must be avoided. We have shown that permitting more than 2.5% `NA`s into our time series has a drastic and significant influence on the chance of committing a type I error (arriving at 'false positive,' *i.e.* detecting a trend when none exists) for time series with no or very gentle decadal trends. On the other hand, the inclusion of `NA`s in data sets with a decadal trend present tends to cause an increase in the probability of committing a type 2 error (*i.e.* finding 'false negatives'). Although our modern UTR data sets have fewer `NA`s than we should be concerned about – therefore with a low chance of committing type 1 or type 2 errors – the presence of `NA`s may seriously compromise some of the time series that are still being collected by hand using hand-held thermometers.

We have demonstrated clearly that as the steepness of an expected decadal trend increases, the ability for it to be modeled accurately increases, too. Our GLS model is generally not able to detect trends that are significantly different from 0 °C dec^{-1} unless a slope of 0.20 °C dec^{-1} exists. Very rarely were we able to produce significant model fits at shallower slopes. Finding significant trends at $<$ 0.05 °C dec^{-1} was not possible. Based on the relationship between SD and the added decadal trend, we see that time series with a SD of 1.5 °C (the bulk of the time series here) and a decadal trend of 0.10 °C dec^{-1} would require roughly 640 months of data before our GLS model regularly detects a significant trend ($p < 0.05$). This finding is somewhat discouraging as most global analyses of decadal SST change based on gridded SST products estimate a trend closer to 0.1 °C dec^{-1} (*e.g.* **?**). This means that the trends present in most time series representative of very variable coastal environments that exhibit the same variance structure as that of our data are probably unlikely to be detected as significant, even if they do indeed exist. In other words, the chance of committing a type 2 error is probably very real for such systems, unless time series $>$ 50 years are available.

As 50 year time series are very scarce, it is important to note that those measured at precisions of 0.1 °C to 0.001 °C require fewer months of data to detect long term trends. We have shown that time series measured at a low precision (0.5 °C) may require as much as an additional 24 months of data to accurately detect long-term trends. One of the motivators for this paper was to investigate the effect measurement precision has on a time series' ability to produce results useful for investigations of long-term climate change, and to validate the use of the low precision 0.5 °C thermometer data. Whereas the precision of these data is below the current standard of 0.1 °C required for climate change research (*ref. - I think it is a WMO report of the 2000s; I'll find it*), the length of the thermometer time series makes them a valuable asset. The average length of the thermometer time series in the SACTN, from which the 84 time series used in this study were drawn, is 349 months. The average length of the UTR time series is 167 months. Given this difference in the lengths of the time series, even after correcting for the negative effect of low measurement precision, the time series collected with thermometers are currently more useful for climate change research than the UTR time series within the SACTN.

We have reflected on the importance of the accuracy of the models, and not only on the importance of their of significance. Indeed, the *p*-value given for the slope in a model does not show how well the model detects the true trend in the data (known *a-priori* in this study); rather, it tells us if the detected trend is significantly different from 0 °C dec^{-1}. This is not particularly useful for applying the results of climate change research more broadly to biotic interests. That a long term trend does exist, may be accurately detected by a model and related to an observed change in the natural world – such as range expansion/contraction of coastal biota **??** – is more important than whether or not the model can show if that trend is significantly different from 0 °C dec^{-1} in a statistical sense.

We must mention that much of the meta-data pertaining to the older temperature records used here, such as the instrumentation used and the motivation behind the levels of precision at which the data were recorded, have over time been lost, highlighting the issues of staff rotation in government departments and the importance of implementing meta-data standards at a very early stage in any monitoring programme. The practical effect this has on our study is that we cannot verify which instruments have been used at which sites or if the same instruments have been used at the same sites over time (*e.g.* mercury to alcohol thermometers). Thus preventing us from measuring the accuracy of the data or the potential drift that may have occurred with the instruments used and quantifying that effect for inspection in this paper. We do know however that all time series sampled with thermometers were sampled only with thermometers, and *vice versa* for the UTR time series, ensuring that the precisions of the measured data used in this study are correct.

5. Conclusion

We draw several key conclusions:

1. The length of a time series has the largest effect on the accuracy and significance (p) of modelled trends however, natural variability (SD) also has a large effect on the significance (p) of the modelled trend.

2. There is a rapid increase in the accuracy and significance of modelled trends as time series lengths extend from 10 to 20 years. This improvement slows from 20 to 30 years, and as time series approach 40 years in length the accuracy of models becomes nearly exact. Modelled results from time series at or under 10 years in length should be interpreted with extreme caution.

3. Whereas time series with a reasonable decadal trend (i.e. 0.1 °C dec^{-1}) and 520 months in lengths will generally allow for perfect model accuracy, an additional 120 months is often required for the detected trend to be considered significant ($p \leq 0.05$).

4. The length of a time series required to detect a reasonable decadal trend (i.e. 0.1 °C dec^{-1}) may rapidly exceed 100 years when a large amount of variance is present.

5. The greater the decadal trend is within a time series, the more accurately it will be modelled regardless of the amount of variance in the time series.

6. There is a complicated relationship between the accuracy of a trend fitted to a time series and the %`NA` of that time series, which is exacerbated as the %`NA` increases. The modelled *p*-values for small or non-existent decadal trends move towards 0 whereas modelled *p*-values for larger trends move away from zero. Increasing thee potential to commit both type I and type II errors.

7. Filling `NA` values via linear interpolation drives the *p*-value of modelled slopes greatly towards zero and this effect is increased by the size of the decadal trend present in the data. It is therefore very important not to use linear interpolation to fill gaps in data.

8. A precision greater than 0.5 °C is not required to confidently detect the long-term trend in a time series however, precisions at or greater than 0.1 °C will reduce the length required to accurately detect a long term trend in the data if one does exist. This is an important consideration as many studies investigating the effects of climate change (e.g. **???**) do use lower precision 0.1 °C data.

9. Because time series with data precisions of 0.1 °C to 0.001 °C produce comparable results, lower precision data may be combined with newer higher precision data within the same time series without concern that the reduced overall data precision may have a negative impact on a models ability to detect decadal trends. Indeed, extending time series in this way will only serve to make them more dependable as length is the primary criteria through which one should initially assess a time series ability to detect climate change before refining ones assumptions with any statistical analyses.

10. Improving the precision of measurements to greater than 0.1 °C has almost no appreciable effect on a models ability to detect a long-term trend, provided that the reported effect size matches the level of precision by the instruments.

11. A time series with data precision greater than 0.1 °C is only necessary when an investigation requires that the decadal trend be known to an accuracy of 0.01 °C or greater (**?**).

We understand that time series of >30 years may be exceedingly rare. Therefore, while we move forward as a scientific community investigating the issues of climate change, the increasing length and continuity of any current and future time series must be ensured in order to construct and maintain a clear understanding of the trends in changing temperature that are occurring throughout Earth's oceans.

Acknowledgments. The authors would like to thank DAFF, DEA, EKZNW, KZNSB, SAWS and SAEON for their contributions of the raw data used in this study. Without it, this article and the SACTN would not be possible. This research was supported by NRF Grant (CPRR14072378735). The authors report no financial conflicts of interests. The data and analyses used in this paper may be found at https://github.com/schrob040/Trend_Analysis.

APPENDIX

Effects of linear interpolation

Many methods for interpolating missing values within a time series exist. As it is beyond the scope of this paper to investigate the differences in interpolation quality we have not gone through the pains of doing so. Our initial methodology called for linearly interpolating the missng data values (`NA`) found within our time series. The logic in doing so was that by only using time series with less than 15%NA the effect of linear interpolating would be minimal. In order to be thorough in our methodology we tested the effect that linear interpolation was having and found that it was significant. The result of the effect of the amount of linear interpolation that is present in a time series may be seen in Figure 5. The only difference between Figure 7 and Figure 5 is that the `NA` values in Figure 7 were left as is and in 5 they were filled with linear interpolation. One may immediately note in the figure showing the interpolated data how many more of the time series that have decadal trends larger than 0.1 °C dec^{-1} are significant compared to the non-interpolated data. It is for this reason that we strongly advise against the use of simple linear interpolation to fill missing gaps in time series.

LIST OF FIGURES

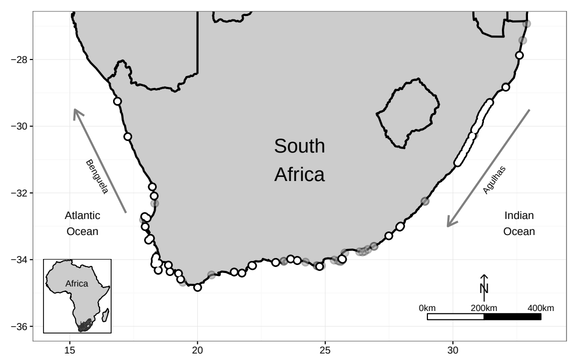


FIG. 1. The location of the 129 time series available for use in this study. The 84 time series used are shown as solid white circles and those not used are shown as opaque.

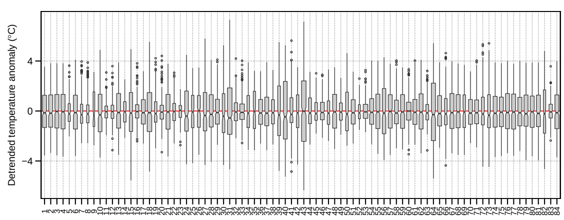


FIG. 2. Box and whisker plots of the base 84 time series used in this study after detrending but before changing the length, adding a decadal trend or rounding the data.

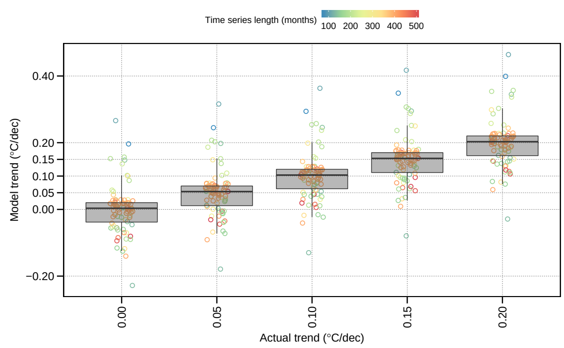


FIG. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to each time series. The box-and-whisker plots show the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, and the minima and maxima are indicated by the whiskers. Points indicate the spread of the actual data points and their colors are scaled according to the length of the time series.

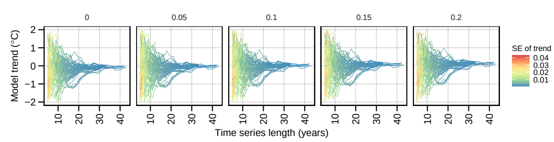


FIG. 4. The relationship between the length of a time series, the size of the modeled trend and its the standard error (SE). Each individual line shows the modeled trend for one of the 84 sites used in this analysis to which a model was fitted iteratively as the time series length was ‘grown’ from 5 years in length to the maximum duration available for the site.

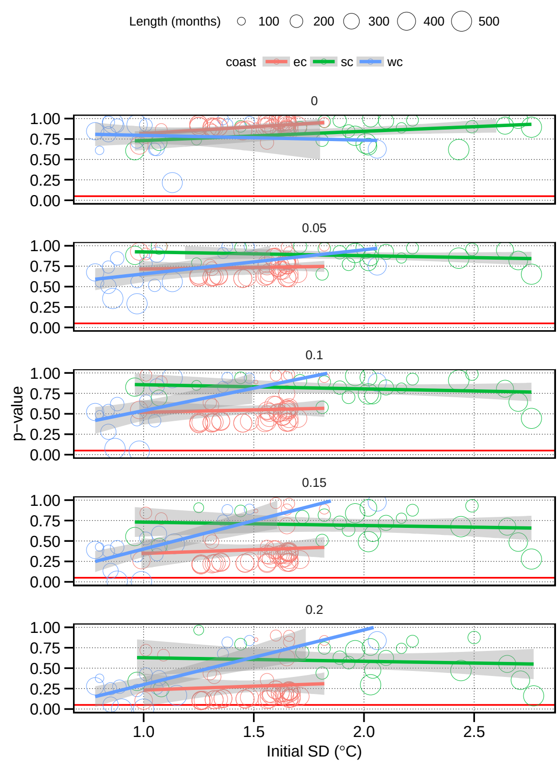
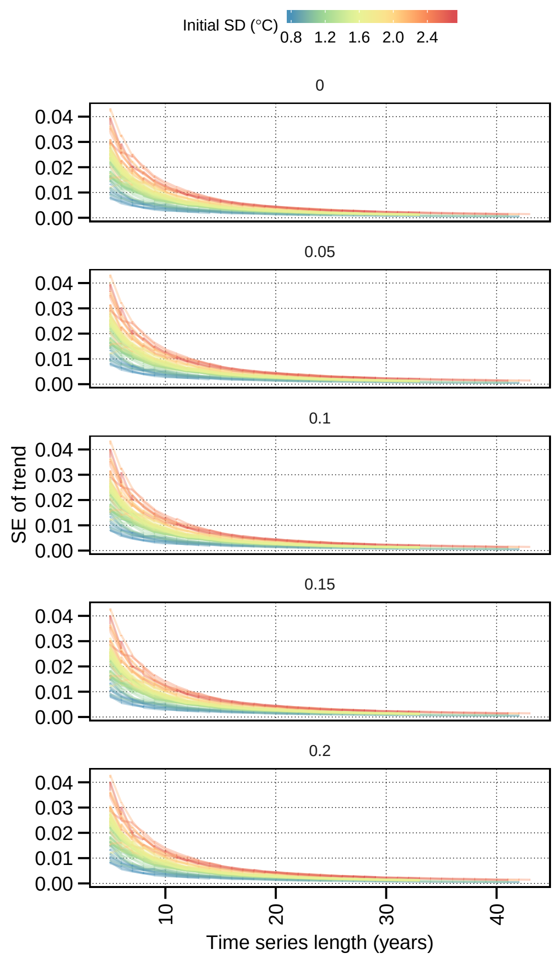


FIG. 5. The effect of the natural variation of a time series on the significance of the modelled trends estimated by the GLS. The size of the symbols are scaled proportionally to the time series length, with longer time series shown as larger circles.



452 FIG. 6. The relationship between the effect of the initial SD of a time series on the SE of a modelled trend,
453 controlled for by the length of the time series.

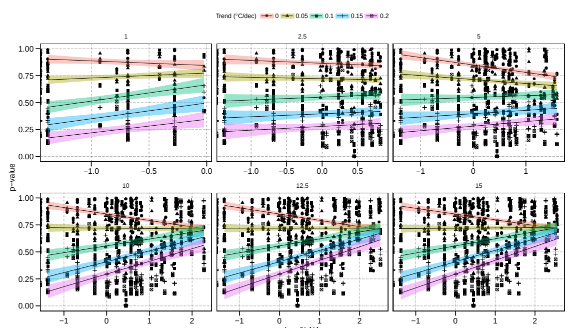


FIG. 7. The relationship between %NA and the significance of a fitted trend. Each panel shows the effect of an increasingly larger amount of missing values. The the fitted lines and 95% confidence intervals represent each of the five decadal trends assessed.

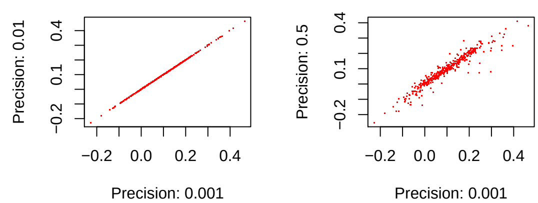


FIG. 8. The minimal effect of rounding from 0.001 °C to 0.01 °C may be seen in the panel on the right. The panel on the left shows that rounding from a precision of 0.001 °C to 0.5 °C has a more appreciable effect.

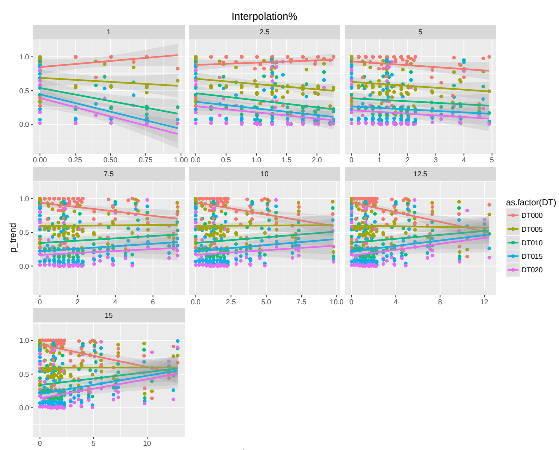


Fig. A1. Similar to Figure 7, this figure shows the effect missing data have on the significance of the slopes detected by GLS however; the missing values in the time series have been filled here via linear interpolation. The effect this has on the significance of the modeled trends is both immediate and dramatic. The behaviour of the quantity of interpolated data also differs from the effect of data left simply as `NA`. At lower levels of interpolation, missing data actually aid in the fitting of a more significant trend line. This phenomena reverses around 5%`NA` when the relationship becomes negative, meaning that as the amount of interpolated data increase, the significance of the fitted trend decrereases.