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 up 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 natural 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 has on trend detection. We determined that the variables contributing most to accurate trend detection in decreasing order are: time series length, the decadal trend, variance, amount of missing data and the precision of the measurements. Time series at least 25 years in length may be used tentatively for climate change research, but 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. *Say something too about the decadal trend, variance, amount of missing data and the precision. RWS - Like what? That greater decadal trends require fewer months of data to model accurately and significantly? That variance has a large effect on the significance of the trend detected, but not on the accuracy of the model? And that missing data erroneously increases the p-value of shallow trends, but decreases the p-value of steep trends, adding to the risk of committing type I and type II errors? That precision of 0.5C should be avoided unless the time series is quite long, and that precision of 0.1C is just about as good as 0.001C? Easy to add, but*

1. Introduction

The roughly 3,000 km of South Africa's coastline is bordered by the Benguela and Agulhas Currents (*e.g.* Roberts 2005; Hutchings et al. 2009), which, in combination with other nearshore processes, affect the country's marine coastal ecosystems (Santos et al. 2012). A thorough understanding of these coastal processes is provided by several physical variables, of which temperature is one of the main determinants (*e.g.* Blanchette et al. 2008; Tittensor et al. 2010; Couce et al. 2012). The statistical properties of *in situ* seawater temperature time series representing the whole coastline – such as the annual mean, minimum and maximum temperature, and the thermal range and variance characteristics – vary greatly among coastal sections due to the varying influence of the Benguela and Agulhas Currents. Based on these thermal properties, the coastline has been classified into a cool temperate west coast, a warm temperate south coast and tending towards sub-tropical along the east coast (Smit et al. 2013; Mead et al. 2013). That the ocean temperature of these regions is changing has been reported in recent years. For example, an increase of 0.55 °C to 0.7 °C dec^{-1} has been reported in the Agulhas Current (Rouault et al. 2009, 2010), while the southern Benguela has decreased in 0.5 °C dec^{-1} during some parts of the year (Rouault et al. 2010).

The aforementioned climate change trends were derived from gridded sea surface temperature (SST) records. Some sources warn of the pitfalls in using temperatures extracted from such gridded products close to the shore (*insert references here*), and a study by Smit et al. (2013) showed that SST data indeed 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.* Blanchette et al. 2008; Broitman et al. 2008; Tyberghein et al. 2012), 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.* Reynolds and Smith 1994; Brown et al. 1999; Martin et al. 2012), 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 was 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. Furthermore, the natural variability of each of the time series, which differ more-or-less predictably between coastlines variously affected by the Benguela and Agulhas Currents, was also entered into the analysis. 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 (Cliff et al. 1988). 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) (R Core Team 2013). 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 Kennedy et al. (2011) and used in IPCC (2013). 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.

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 Franzke (2012). Two broad approaches are widely used in climate change research (Stocker et al. 2013). 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}; Wilks 2011; Stocker et al. 2013). 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 (Scinocca et al. 2010; Wood 2006). 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 (Von Storch 1999; Santer et al. 2008). 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 (Wood 2006; Pinheiro and Bates 2006). 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; Pinheiro and Bates 2006):

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

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 (Hartmann et al. 2013). 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 influence 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 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 °C 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 °C 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 `NAs`, as suggested by the greater number of 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 °C 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 °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. Detecting temperature change in highly variable coastal environments, such as those around the coast of South Africa and many temperate coastal environments globally, will therefore benefit from access to the longest possible time series available. This phenomenon is demonstrated in Figure 5, which uses color to show the time series binned by the three different coastal sections of South Africa (Smit et al. 2013). Of these three coastal sections the east coast is known to have the most stable thermal regime (*i.e.* with the smallest variance), with the south coast having the greatest variance. Long time series at sites of low variance result in great improvements in our ability to detect significant climate change trends, and this effect is most obvious in data sets with steeper slopes in the temperature *vs.* time relationship. The selection of sites for long-term monitoring must therefore account for the location of study and necessitate adequate planning to collect a long enough time series.

The detection of long-term trends require long-term data, a fact that is already firmly established in climate change research (Ohring et al. 2005; IPCC 2013). The length of these time series 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 to detect decadal temperature trends, and know the accuracy of these estimates 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 generally 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 $0\,^{\circ}\mathrm{C\,dec^{-1}}$ unless a slope of $0.20\,^{\circ}\mathrm{C\,dec^{-1}}$ exists. Very rarely were we able to produce significant model fits at shallower slopes. Finding significant trends at $< 0.05\,^{\circ}\mathrm{C\,dec^{-1}}$ was not possible. Based on the relationship between SD and the added decadal trend, we see that time series with a SD of $1.5\,^{\circ}\mathrm{C}$ (the bulk of the time series here) and a decadal trend of $0.10\,^{\circ}\mathrm{C\,dec^{-1}}$ should consist of roughly 640 months of data before our GLS model would regularly be able to detect 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.* IPCC 2013). 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 coastal seawater temperature time series are probably 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. Based on the data presented here, we calculated 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. This is an important consideration as many studies investigating the effects of climate change (*e.g.* Grant et al. 2010; Scherrer and Körner 2010; Lathlean and Minchinton 2012) do use lower precision 0.1 °C data. Whereas the precision of much of our data is below the current standard of 0.1 °C required for climate change research (Ohring et al. 2005; Jarraud 2008), 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}. *How many of the models in the natural data produce accurate trend estimates but are not significant?* 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 (Bolton et al. 2012; Straub et al. 2016; Wernberg et al. 2016) – 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 also that much of the meta-data pertaining to the older temperature records used here have over time been lost. As with the bulk of the International Comprehensive Ocean-Atmosphere Data Set (ICOADS; Freeman et al. 2016), *in situ* coastal seawater temperature monitoring that started in the 1970s in South Africa was not developed with climate change research in mind, and comprehensive records that keep track of details of the instruments used, calibration, their turnover, change in monitoring methods and locations and so forth are not always available as per modern requirements (Aguilar et al. 2003). For studies of climate change *per se* this is a serious limitation and it prevents us from knowing anything about the accuracy of the instruments or potential issues of drift (stability) that may have occurred. 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. Moving forward with the further development of the SACTN and the establishment of a national standard of data collection and instrument maintenance, we are able to record and archive all these levels of pertienent meta-data, and allowing for the enforcement of SI traceability and the accurate measurement of instrument drift (Jarraud 2008). Nevertheless, the detrended anomaly time series used here were taken only for their variance properties, which we think accurately reflect that of the various coastal sections around the coast. They provide a strong backbone for semi-artificial time series, and we have shown how important insights about model fitting could be derived from these data.

5. Conclusion

We draw several key conclusions:

1. 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. Modeled trends from time series at or under 25 years in length should be interpreted with extreme caution.

2. For our variable coastal seawater, a time series of 520 month in length is required to detect a decadal trend in line with the global average (*i.e.* near 0.1 °C dec^{-1}) with perfect accuracy; however, an additional 120 months of data is often required for the detected trend to be considered significant ($p \leq 0.05$).

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

4. The larger the decadal trend within a time series, the more accurately it will be modeled regardless of the amount of variance in the time series.

5. There is a complicated relationship between the accuracy of a trend fitted to a time series and the `%NA` of that time series. As the `%NA` increases, so too does the change of committing type 1 (at gentle slopes) or type 2 errors (at steeper trends).

6. 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 of time required to accurately detect a long term trend if one does exist.

7. 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.

We understand that time series of >30 years may be exceedingly rare. Therefore, as we move forward as a scientific community investigating the issues of climate change, the continuity of any current time series of sufficient length must be ensured as these commodities are practically irreplacable.

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.

References

Aguilar, E., I. Auer, M. Brunet, T. Peterson, and J. Wieringa, 2003: Guidelines on climate metadata and homogenization. world climate programme data and monitoring wcdmp-no. 53, wmo-td no. 1186. *World Meteorological Organization, Geneva*.

Blanchette, C. A., C. Melissa Miner, P. T. Raimondi, D. Lohse, K. E. K. Heady, and B. R. Broitman, 2008: Biogeographical patterns of rocky intertidal communities along the Pacific coast of North America. *Journal of Biogeography*, **35 (9)**, 1593–1607.

Bolton, J. J., R. J. Anderson, A. J. Smit, and M. D. Rothman, 2012: South African kelp moving eastwards: the discovery of Ecklonia maxima (Osbeck) Papenfuss at De Hoop Nature Reserve on the south coast of South Africa. *African Journal of Marine Science*, **34 (1)**, 147–151.

Broitman, B. R., N. Mieszkowska, B. Helmuth, and C. A. Blanchette, 2008: Climate and recruitment of rocky shore intertidal invertebrates in the eastern North Atlantic. *Ecology*, **89 (11 Suppl)**, S81—-90.

Brown, O. B., P. J. Minnett, R. Evans, E. Kearns, K. Kilpatrick, A. Kumar, R. Sikorski, and A. Závody, 1999: MODIS Infrared Sea Surface Temperature Algorithm Algorithm Theoretical Basis Document Version 2.0. *University of Miami*, 31 098–33 149.

Cliff, G., S. F. J. Dudley, and B. Davis, 1988: Sharks caught in the protective gill nets off Natal, South Africa. 1. The sandbar shark Carcharhinus plumbeus (Nardo). *South African Journal of Marine Science*, **7 (1)**, 255–265.

Couce, E., A. Ridgwell, and E. J. Hendy, 2012: Environmental controls on the global distribution of shallow-water coral reefs. *Journal of Biogeography*, **39 (8)**, 1508–1523.

Franzke, C., 2012: Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *Journal of Climate*, **25 (12)**, 4172–4183, doi:10.1175/JCLI-D-11-00293.1.

Freeman, E., and Coauthors, 2016: ICOADS Release 3.0: a major update to the historical marine climate record. *International Journal of Climatology*, 1–22.

Grant, O. M., L. Tronina, J. C. Ramalho, C. Kurz Besson, R. Lobo-Do-Vale, J. Santos Pereira, H. G. Jones, and M. M. Chaves, 2010: The impact of drought on leaf physiology of Quercus suber L. trees: Comparison of an extreme drought event with chronic rainfall reduction. *Journal of Experimental Botany*, **61 (15)**, 4361–4371, doi:10.1093/jxb/erq239.

Hartmann, D., and Coauthors, 2013: Observations: Atmosphere and surface supplementary material. in: Climate change 2013: The physical science basis.

Hutchings, L., and Coauthors, 2009: The Benguela Current: An ecosystem of four components. *Progress in Oceanography*, **83 (1-4)**, 15–32, doi:10.1016/j.pocean.2009.07.046.

IPCC, 2013: *IPCC, 2013: climate change 2013: The Physical Science Basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Jarraud, M., 2008: Guide to meteorological instruments and methods of observation (wmo-no. 8). *World Meteorological Organisation: Geneva, Switzerland*.

Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011: Reassessing biases and other uncertainties in sea-surface temperature observations measured *in situ* since 1850, Part 1: measurement and sampling uncertainties. *Journal of Geophysical Research Atmospheres*, **116**.

Lathlean, J. A., and T. E. Minchinton, 2012: Manipulating thermal stress on rocky shores to predict patterns of recruitment of marine invertebrates under a changing climate. *Marine Ecology Progress Series*, **467**, 121–136, doi:10.3354/meps09996.

Martin, M., and Coauthors, 2012: Group for High Resolution Sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: A GHRSST multi-product ensemble (GMPE). *Deep Sea Research Part II: Topical Studies in Oceanography*, **77-80**, 21–30.

Mead, A., and Coauthors, 2013: Human-mediated drivers of change — impacts on coastal ecosystems and marine biota of South Africa. *African Journal of Marine Science*, **35 (3)**, 403–425.

Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, 2005: Satellite Instrument Calibration for Measuring Global Climate Change: Report of a Workshop. *Bulletin of the American Meteorological Society*, **86 (9)**, 1303–1313.

Pinheiro, J., and D. Bates, 2006: *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

R Core Team, 2013: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL http://www.r-project.org/.

Reynolds, R. W., and T. M. Smith, 1994: Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Journal of Climate*, **7 (6)**, 929–948.

Roberts, M. J., 2005: Chokka squid (Loligo vulgaris reynaudii) abundance linked to changes in South Africa's Agulhas Bank ecosystem during spawning and the early life cycle. *ICES Journal of Marine Science*, **62 (1)**, 33–55, doi:10.1016/j.icesjms.2004.10.002.

Rouault, M., P. Penven, and B. Pohl, 2009: Warming in the Agulhas Current system since the 1980's. *Geophysical Research Letters*, **36 (12)**, 1–5.

Rouault, M., B. Pohl, and P. Penven, 2010: Coastal oceanic climate change and variability from 1982 to 2009 around South Africa. *African Journal of Marine Science*, **32 (2)**, 237–246.

Santer, B. D., and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *International Journal of Climatology*, **28 (13)**, 1703–1722.

Santos, F., M. Gomez-Gesteira, M. DeCastro, and I. Alvarez, 2012: Differences in coastal and oceanic SST trends due to the strengthening of coastal upwelling along the Benguela current system. *Continental Shelf Research*, **34**, 79–86.

Scherrer, D., and C. Körner, 2010: Infra-red thermometry of alpine landscapes challenges climatic warming projections. *Global Change Biology*, **16 (9)**, 2602–2613, doi:10.1111/j.1365-2486.2009.02122.x.

Scinocca, J., D. B. Stephenson, T. C. Bailey, J. Austin, and Coauthors, 2010: Estimates of past and future ozone trends from multimodel simulations using a flexible smoothing spline methodology. *Journal of Geophysical Research: Atmospheres*, **115 (D3)**.

Smit, A. J., M. Roberts, R. J. Anderson, F. Dufois, S. F. J. Dudley, T. G. Bornman, J. Olbers, and J. J. Bolton, 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)**, doi:10.1371/journal.pone.0081944.

Stocker, T., and Coauthors, 2013: *IPCC, 2013: climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge University Press.

Straub, S. C., M. S. Thomsen, and T. Wernberg, 2016: The dynamic biogeography of the anthropocene: The speed of recent range shifts in seaweeds. *Seaweed Phylogeography*, Springer, 63–93.

Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, and B. Worm, 2010: Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466 (7310)**, 1098–1101, doi:10.1038/nature09329.

Tyberghein, L., H. Verbruggen, K. Pauly, C. Troupin, F. Mineur, and O. De Clerck, 2012: Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography*, **21 (2)**, 272–281.

Von Storch, H., 1999: Misuses of statistical analysis in climate research. *Analysis of Climate Variability*, Springer, 11–26.

Wernberg, T., and Coauthors, 2016: Climate-driven regime shift of a temperate marine ecosystem. *Science*, **353 (6295)**, 169–172.

Wilks, D. S., 2011: *Statistical methods in the atmospheric sciences*, Vol. 100. Academic press.

Wood, S., 2006: *Generalized additive models: an introduction with R*. CRC press.

LIST OF FIGURES

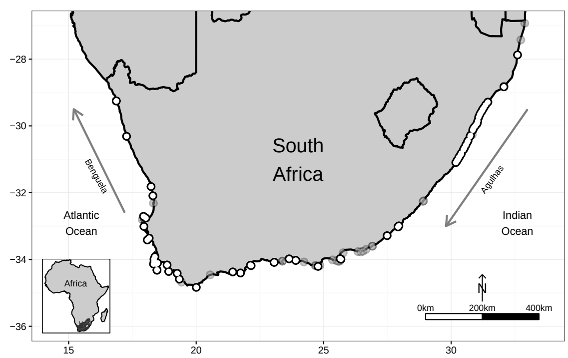


FIG. 1. The location of the 129 time series available for use in this study. The 84 time series actually used are shown as solid white circles and those not used are shown as opaque.

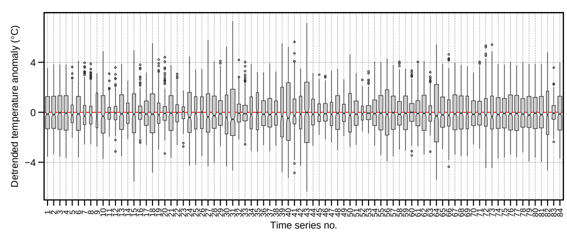


FIG. 2. Box and whisker plot summarizing the 84 base anomaly time series used in this study after detrending (*i.e.* the residuals after removing the linear trend using an ordinary least squares regression) but before adding a decadal trend or rounding the data. The plot indicates the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, the minima and maxima are indicated by the whiskers and the points are outliers.

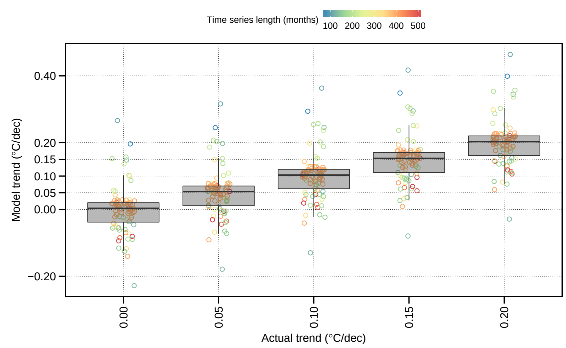


FIG. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to each time series. The box-and-whisker plots show the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, and the minima and 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.

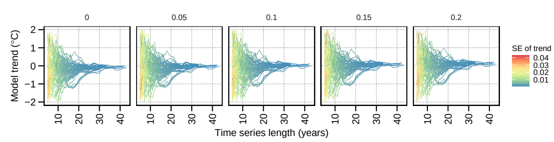


FIG. 4. The relationship between the length of a time series, the size of the modeled trend and its the standard error (SE). Each individual line shows the modeled trend for one of the 84 sites used in this analysis to which a model was fitted iteratively as the time series length was 'grown' from 5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

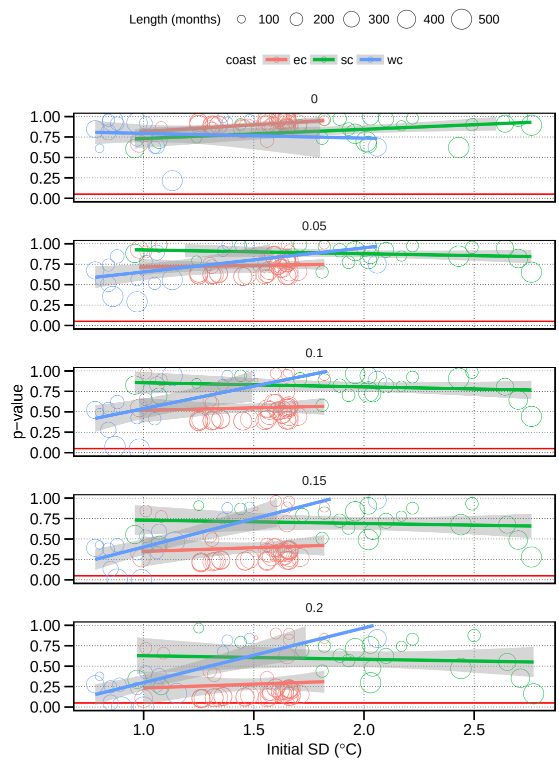
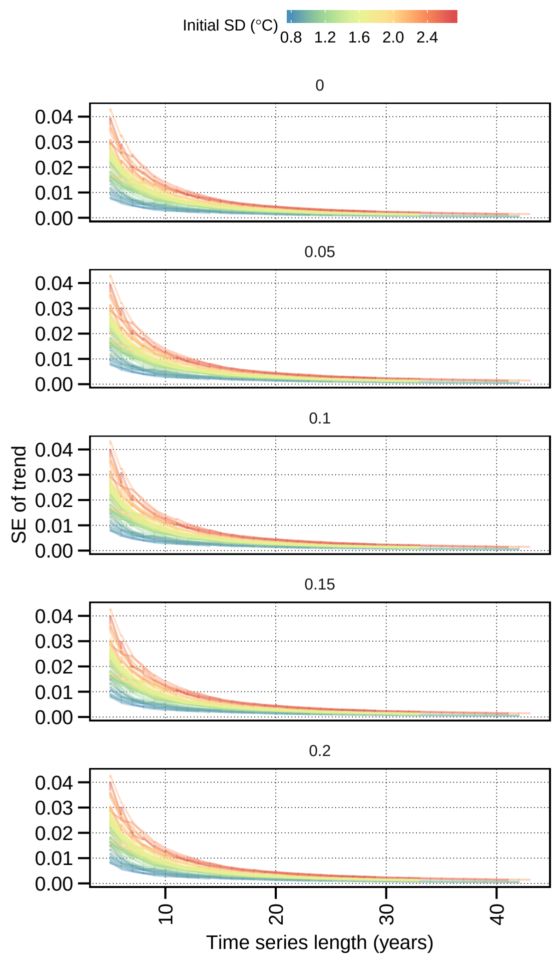


FIG. 5. The effect of the natural variation of a time series on the significance of the modelled trends estimated by the GLS. The size of the symbols are scaled proportionally to the time series length, with longer time series shown as larger circles.



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

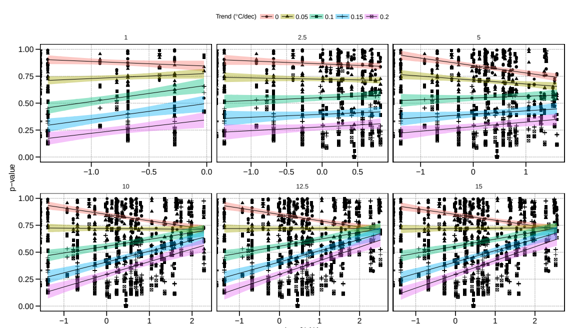


FIG. 7. The relationship between the percentage of missing values (%NA) and the significance of a fitted trend. Each panel shows the effect of an increasingly larger amount of missing values. The fitted lines and 95% confidence intervals represent each of the five decadal trends assessed.

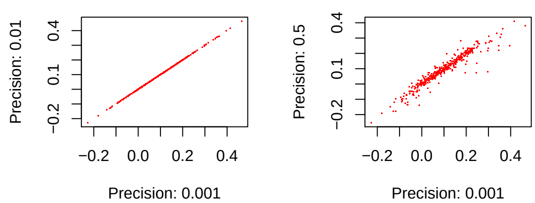


FIG. 8. The minimal effect of rounding from 0.001 °C to 0.01 °C may be seen in the panel on the right. The panel on the left shows that rounding from a precision of 0.001 °C to 0.5 °C has a more appreciable effect.