

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

4 Robert Schlegel* and Albertus Smit

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

7 **Corresponding author address:* Robert Schlegel, Department of Biodiversity and Conservation
8 Biology, University of the Western Cape, Bellville, Republic of South Africa.
9 E-mail: 3503570@myuwc.ac.za

ABSTRACT

10 In South Africa 129 *in situ* temperature time series of 1 to 43 years are
11 used for investigations of the thermal characteristics of coastal seawater. They
12 are comprised of temperature recordings at precisions ranging from 0.5 °C to
13 0.001 °C and collected with handheld thermometers or underwater tempera-
14 ture recorders (UTRs). Using the naturally occurring range of seasonal sig-
15 nals, variability and temperature trends for 84 of these time series, the length,
16 decadal trend and data precision of each time series were systematically var-
17 ied before fitting generalized least squares (GLS) models to study the effect
18 these variables have on trend detection. We determined that the variables con-
19 tributing most to accurate trend detection in decreasing order are: the length
20 of the time series, the decadal trend, variance, amount of missing data and the
21 precision of the measurements. We found that time series at least 20 years
22 in length may be used tentatively for climate change research, but that time
23 series >30 years in length are preferable. The implication is that long-running
24 thermometer time series in this dataset, and others around the world, are more
25 useful for decadal scale climate change studies than the shorter, more precise
26 UTR time series. It is important to note that due to the nature of the dataset
27 used in this study, instrument drift was not able to be quantified.

28 1. Introduction

29 The roughly 3,000 km of South Africa's coastline is bordered by the Benguela and Agulhas
30 currents (*e.g.* Roberts 2005; Hutchings et al. 2009), which, in combination with other nearshore
31 processes, affect the country's marine coastal ecosystems (Santos et al. 2012). A thorough under-
32 standing of these coastal processes is provided by several physical variables, of which temperature
33 is one of the main determinants (*e.g.* Blanchette et al. 2008; Tittensor et al. 2010; Couce et al.
34 2012). In order to ensure a true representation of organisms' biological thermal limits, nearshore
35 temperatures must be accurately recorded and monitored. Some sources warn of the pitfalls in
36 doing so *RWS: Add references here showing which sources say using SST for the coast is inappro-*
37 *priate*, and a study by Smit et al. (2013) showed that SST data have a warm bias as large as 6 °C
38 when compared to coastal *in situ* data. Nevertheless, a widespread approach in coastal ecologi-
39 cal research is to use satellite and/or model-generated temperature data as a representation of the
40 sea surface temperature (SST) along coastlines (*e.g.* Blanchette et al. 2008; Broitman et al. 2008;
41 Tyberghein et al. 2012), because apparently the dangers of applying gridded SSTs to the coast
42 are not widely known or in many places in the world there simply are no suitable *in situ* coastal
43 temperature time series available. It is for this reason that we strongly recommended the use of *in*
44 *situ* data to support research conducted within 400 m from the shoreline.

45 Where records of *in situ* coastal seawater temperature do exist, the reliability of many of these
46 datasets that could be used in place of the remotely-sensed SST data remains to be verified. Users
47 of SST data benefit from it being refined through a number of well documented validation and
48 quality control processes (*e.g.* Reynolds and Smith 1994; Brown et al. 1999; Martin et al. 2012),
49 whereas the standards and methods with which local *in situ* data from a single dataset are collected
50 and refined may differ greatly. For example, there are currently seven organizations and/or govern-

51 mental departments (hereafter referred to as bodies) contributing coastal seawater temperature data
52 to the South African Coastal Temperature Network (SACTN). These bodies use different methods
53 and instruments to collect their data as no national standard has been set. One consequence of this
54 methodological disparity is that two thirds of the data were sampled with hand-held thermometers
55 that are manually recorded at a data precision of 0.5 °C, as opposed to the current generation of
56 Underwater Temperature Recorders (UTRs) with an instrument precision of down to 0.001 °C. If
57 these *in situ* data are to be used together *in lieu* of the satellite-based SST data, it is important that
58 the characteristics of the contributing data sources are understood in terms of their ability to yield
59 useful, reliable and accurate long-term measurements for use in climate change studies.

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

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

72 2. Methods

73 a. Data Sources

74 Our study lies within the political borders of South Africa's coastline. The location of each point
75 of collection appears in Figure 1. Of these 129 time series, 43 are recorded with UTRs and the
76 other 86 with hand-held mercury thermometers. The oldest currently running time series began on
77 January 1st, 1972; there are 11 total time series that started in the 70s, 53 more started in the 80s,
78 34 began in the 90s, 18 in the 00s and 13 in the current decade.

79 The data are collected using two different methods and a variety of instruments. Hand-held
80 mercury thermometers (which are being phased out in favor of alcohol thermometers or electronic
81 instruments) are used in some instances at the shoreline, and represent seawater temperatures at
82 the surface. At other places, predominantly along the country's east coast, data are collected with
83 glass thermometers from small boats at the location of shark nets along the coast (Cliff et al. 1988).
84 Whereas both types of thermometers allow for a measurement precision of 0.1 °C, the recordings
85 are written down at a precision of 0.5 °C. Data at other localities are collected using delayed-mode
86 instruments that are permanently moored shallower than 10 m, but generally very close to the
87 surface below the low-water spring tide level.

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

94 0.1 °C and so have been rounded down to match this level of precision. Eight additional UTR time
95 series have older data that were recorded at a precision of 0.1 °C.

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

100 *b. Data Management*

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

110 All analyses and data management performed in this paper were conducted with R version 3.3.1
111 (2016-06-21) (R Core Team 2013). The script and data used to conduct the analyses and create
112 the tables and figures in this paper may be found at https://github.com/schrob040/Trend_Analysis.

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

117 At the time of this analysis, this useable daily dataset consisted of 84 time series, consisting of
118 819,499 days of data; these data were then binned further to the 26,924 monthly temperature
119 values available for use in this study.

120 *c. Systematic Analysis of Time Series*

121 We used the 84 time series simply for their variance properties (comprised of seasonal, inter-
122 annual, decadal and noise components), which reflect that of the thermal environment naturally
123 present along the roughly 3,000 km of South African coastline. Linear trends that may have been
124 present in each time series were removed prior to the ensuing analysis by applying an ordinary
125 least squares regression and keeping the detrended residuals as anomaly time series. In doing so
126 we avoided the need to simulate a series of synthetic time series, whose variance components may
127 not have been fully representative of that naturally present in coastal waters. These detrended
128 anomaly time series (henceforth simply called ‘time series’) represent a range of time scales from
129 72 to 519 months in duration.

130 To each of the 84 time series we artificially added linear decadal trends of 0.00°C to
131 $0.20^{\circ}\text{C dec}^{-1}$. In other words, we now had time series that captured the natural thermal vari-
132 abilities around the coast, but with their decadal trends known *a priori*. The range of decadal
133 trends was selected based around the global average of 0.124°C from Kennedy et al. (2011) and
134 used in IPCC (2013). Furthermore, in order to represent the instrumental precision of the instru-
135 ments used to collect these time series, we rounded each of these (84 time series \times 5 decadal
136 trends) to four levels of precision: 0.5°C , 0.1°C , 0.01°C and 0.001°C . Consequently, we had a
137 pool of 1,680 time series with which to work.

138 To gain further insight into the effect of time series length on trend detection, each time series
139 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 NAs 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 Franzke (2012). Two broad approaches are widely used in climate change research (Stocker et al. 2013).

163 The first group of models estimates linear trends, and although linearity may not reflect reality (*i.e.*
 164 trends are very frequently non-linear), these models do provide the convenience of producing an
 165 easy to understand decadal trend (*e.g.* $0.106^{\circ}\text{C dec}^{-1}$; Wilks 2011; Stocker et al. 2013). The other
 166 group accommodates non-linear trajectories of temperature through time by the use of higher-
 167 degree polynomial terms or non-parametric smoothing splines, but the inconvenience comes from
 168 not being able to easily compare models among sites (Scinocca et al. 2010; Wood 2006). Both
 169 groups of models can accommodate serially correlated residuals, which is often the cause for much
 170 criticism due to their effect on the uncertainty of the trend estimates (Von Storch 1999; Santer et al.
 171 2008). For example, Generalized Least Squares (GLS; yielding estimates of linear trends) and
 172 Generalized Additive Mixed Models (GAMM; non-linear fitting with no trend estimate provided)
 173 can both capture various degrees of serial autocorrelation (Wood 2006; Pinheiro and Bates 2006).
 174 Although our exploratory analysis assessed two parameterizations of each of the model groups, we
 175 opted to proceed here with a GLS equipped with a second-order autoregressive AR(2) correlation
 176 structure fitted using Restricted Maximum Likelihood (REML; Pinheiro and Bates 2006):

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

177 where the lag-2 autocorrelated residuals are given by

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

178 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 detrended 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 $0\text{ }^{\circ}\text{C dec}^{-1}$ 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 $0\text{ }^{\circ}\text{C dec}^{-1}$.

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 NAs 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 NAs 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 p -values

226 that approach 0.05). In other words, the inclusion of missing values results in time series with no
227 added decadal trend to veer away from $0^{\circ}\text{C dec}^{-1}$ towards a situation where they may erroneously
228 appear to display a trend. On the other hand, time series that do indeed have decadal trends tend
229 to produce fits that are not significantly different from $0^{\circ}\text{C dec}^{-1}$.

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

235 The effect of decreasing data measurement precision from 0.001°C to 0.5°C has almost no
236 appreciable effect on any of the measures of variance presented in this study. The effect of mea-
237 surement precision on the accuracy of the modeled slope, however, becomes very pronounced
238 going from 0.1°C to 0.5°C . This effect is larger on smaller decadal trends. For example, at a
239 trend of $0.05^{\circ}\text{C dec}^{-1}$, the deviation from the true value of models fitted to data with a precision
240 of 0.1°C is negligible; however, the accuracy of the fitted model on data recorded at a precision
241 of 0.5°C with a real trend of $0.05^{\circ}\text{C dec}^{-1}$ is 10.81% different on average (*i.e.* given a slope of
242 $0.05^{\circ}\text{C dec}^{-1}$ the model detects slopes of $0.05540^{\circ}\text{C dec}^{-1}$). This accuracy of the models im-
243 proves to an average difference of 6.44% with a slope of $0.10^{\circ}\text{C dec}^{-1}$, 2.24% at $0.15^{\circ}\text{C dec}^{-1}$
244 and decreases slightly to 2.30% at $0.20^{\circ}\text{C dec}^{-1}$. A precision of 0.5°C always provides clearly
245 less accurate modeled trends than at higher precisions; however, the current analysis did not high-
246 light one precision that consistently provides the most accurate estimate of the trends. This may
247 however become determinable in an analysis of synthetic data with variance structures that are
248 manipulated in a more consistent manner.

249 As the actual time series used to generate the data for this study are predominantly greater than
250 300 months in length and recorded at a data precision of 0.5°C , we would be remiss not to investi-
251 gate the interaction between the increase in accuracy provided by a lengthy time series, against the
252 decrease caused by a data precision of 0.5°C . In other words, at what point does a model fitted to a
253 longer time series, with less precise measurements (*e.g.* those taken by thermometers and reported
254 at a precision of 0.5°C), become as accurate as a time series with more precise measurements (*e.g.*
255 UTRs)? Figure 8 shows how varied the modeled trends become when a precision of 0.5°C is used,
256 and we see here that when these low resolution time series have a shallow slope of $0.05^{\circ}\text{C dec}^{-1}$,
257 a fitted model requires 24 months of additional data on average to have a comparable level of accu-
258 racy to a model fitted to data recorded at a precision of 0.1°C . This difference in length decreases
259 to 16 months when a larger slope $0.20^{\circ}\text{C dec}^{-1}$ is used.

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

265 4. Discussion

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

271 Whereas time series with smaller variances (shown as SD in this study) generally produce model
272 fits that are statistically significant (*i.e.* with decadal trends that are significantly different from
273 $0^{\circ}\text{C dec}^{-1}$ at $p < 0.05$) and with smaller SE of the estimated trends after shorter lengths of time,
274 we also see that increasing a time series' length beyond 25 years, but preferably beyond 30 years,
275 will increase the likelihood of detecting a decadal temperature change even in very variable data
276 sets. Measuring temperature change in highly variable coastal environments, such as those around
277 the coast of South Africa and many temperate coastal environments globally (refs.), will therefore
278 benefit from access to the longest possible time series available. This phenomenon is demonstrated
279 in Figure 4, which uses color to show the time series binned by the three different coastal sections
280 of South Africa (Smit et al. 2013). Of these three coastal sections the east coast is known to
281 have the most stable thermal regime, with the south coast having the most variability. A good
282 experimental design must account for the location of study, and we have shown here that location
283 may have a large effect, requiring for adequate planning to collect a long enough time series.

284 The detection of long term trends require long-term data. This finding is both positive and neg-
285 ative. The length of a time series needed for climate change research is firmly under the control of
286 the investigator with sufficient foresight and perseverance to plan the installation and management
287 of new instrument networks that will yield usable results only after about three-quarters of a typ-
288 ical academic career has passed. Should such data already exist, and considering the scarcity of
289 such long-term records that are already yielding benefits today, we must ensure that these sources
290 of data are managed and curated with great care and diligence as they are practically irreplaceable.
291 For this reason, it is essential that we understand the inherent strengths and weaknesses of such
292 existing sources of data so that we may fully maximize their utility and extract from them the
293 model coefficients needed for climatic research, and know their accuracy to the best of our ability.

294 There are many time series < 20 years in length that should be avoided, where possible, for trend
295 analysis. These will mature with time and their maintenance need to be ensured going forward.

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

306 We have demonstrated clearly that as the steepness of an expected decadal trend increases, the
307 ability for it to be modeled accurately increases, too. Our GLS model is generally not able to de-
308 tect trends that are significantly different from $0\text{ }^{\circ}\text{C dec}^{-1}$ unless a slope of $0.20\text{ }^{\circ}\text{C dec}^{-1}$ exists.
309 Very rarely were we able to produce significant model fits at shallower slopes. Finding significant
310 trends at $< 0.05\text{ }^{\circ}\text{C dec}^{-1}$ was not possible. Based on the relationship between SD and the added
311 decadal trend, we see that time series with a SD of $1.5\text{ }^{\circ}\text{C}$ (the bulk of the time series here) and a
312 decadal trend of $0.10\text{ }^{\circ}\text{C dec}^{-1}$ would require roughly 640 months of data before our GLS model
313 regularly detects a significant trend ($p < 0.05$). This finding is somewhat discouraging as most
314 global analyses of decadal SST change based on gridded SST products estimate a trend closer
315 to $0.1\text{ }^{\circ}\text{C dec}^{-1}$ (*e.g.* IPCC 2013). This means that the trends present in most time series repre-
316 sentative of very variable coastal environments that exhibit the same variance structure as that of
317 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

341 than whether or not the model can show if that trend is significantly different from $0^{\circ}\text{C dec}^{-1}$ in a
342 statistical sense.

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

354 *RWS: None of this text has been edited. I must still do so.*

355 5. Conclusion

356 We draw several key conclusions:

- 357 1. There is not a significant relationship between the goodness of fit (R^2) of a linear model to a
358 time series and the NA% of that time series when the NAs are filled in via linear interpolation.
359 This is an important finding as it means that, within reason, linear interpolation may be used
360 to fill gaps in a time series before applying any time series analysis methods.
- 361 2. Length has the largest effect on the goodness of fit (R^2) of the decadal trend and natural
362 variability (SD) has the largest effect on the significance (p) of the trend detected.

- 363 3. There is a predictable decrease in the goodness of fit (R^2) of a linear model to the trend line of
364 a time series as it extends from 10 to 20 years in length. The goodness of fit (R^2) then begins
365 to increase once the time series becomes roughly 30+ years long. Analyses of time series at
366 or under 10 years in length should be interpreted with extreme caution in spite of them often
367 having strong R^2 values.
- 368 4. Within the first decade of a time series, if the temperatures within the last few months move
369 strongly in the opposite direction from the prevailing trend, the linear model used to detect the
370 trendline may show an abrupt change in direction (*i.e.* a positive trend can become negative
371 and *vice versa*).
- 372 5. After the first decade of data, the changes detected in almost all trends for all 105 time series
373 become more gradual; however, many trend lines still change direction over the course of the
374 following two decades.
- 375 6. It is at these changes in direction that the p -values for the time series plummet, though gener-
376 ally they tend to follow the same pattern of becoming weaker and then slowly stronger over
377 time, as we see in the R^2 values.
- 378 7. There is a slight linear decrease in R^2 as the natural thermal variability (SD) of seawater
379 increases; however, the decrease in p -values is larger and more rapid.
- 380 8. A precision greater than 0.5 °C is not required to confidently detect the long-term trend in a
381 time series. This is an important consideration as many studies investigating the effects of
382 climate change (*e.g.* Grant et al. 2010; Scherrer and Körner 2010; Lathlean and Minchinton
383 2012) do use lower precision 0.1 °C data. That being said, a precision of 0.001 °C or 0.01 °C
384 is preferable over 0.5 °C. In fact, because the results from the higher precision of 0.001 °C

were almost identical to the 0.5 °C tests, the higher precision is only necessary when one needs to identify trends at a precision of 0.01 °C or greater (Karl et al. 2015). This finding means that older, lower precision data may be combined with newer higher precision data within the same time series without concern that the reduced overall data precision will have a large negative impact on the time series 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.

9. Decreasing the precision of measurements to greater than 0.1 °C has almost no appreciable effect on a time series 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, 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 contributing all 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

Blurb here about linear interpolation effects.... Figure 5. *RWS* - This figure still needs to be updated to the current standard. It is still a rough draft.

References

- 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.
- 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.

- 427 Grant, O. M., L. Tronina, J. C. Ramalho, C. Kurz Besson, R. Lobo-Do-Vale, J. Santos Pereira,
428 H. G. Jones, and M. M. Chaves, 2010: The impact of drought on leaf physiology of *Quercus*
429 *suber* L. trees: Comparison of an extreme drought event with chronic rainfall reduction. *Journal*
430 *of Experimental Botany*, **61** (15), 4361–4371, doi:10.1093/jxb/erq239.
- 431 Hartmann, D., and Coauthors, 2013: Observations: Atmosphere and surface supplementary mate-
432 rial. in: *Climate change 2013: The physical science basis*.
- 433 Hutchings, L., and Coauthors, 2009: The Benguela Current: An ecosystem of four components.
434 *Progress in Oceanography*, **83** (1-4), 15–32, doi:10.1016/j.pocean.2009.07.046.
- 435 IPCC, 2013: *IPCC, 2013: climate change 2013: The Physical Science Basis. Contribution of*
436 *working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate*
437 *Change*. Cambridge University Press.
- 438 Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface
439 warming hiatus. *Science*, **348** (6242), 1469–1472, doi:10.1126/science.aaa5632.
- 440 Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011: Reassessing biases
441 and other uncertainties in sea-surface temperature observations measured *in situ* since 1850, Part
442 1: measurement and sampling uncertainties. *Journal of Geophysical Research Atmospheres*,
443 **116**.
- 444 Lathlean, J. A., and T. E. Minchinton, 2012: Manipulating thermal stress on rocky shores to pre-
445 dict patterns of recruitment of marine invertebrates under a changing climate. *Marine Ecology*
446 *Progress Series*, **467**, 121–136, doi:10.3354/meps09996.

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

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

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

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

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

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

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

464 Scherrer, D., and C. Körner, 2010: Infra-red thermometry of alpine landscapes challenges climatic
 465 warming projections. *Global Change Biology*, **16 (9)**, 2602–2613, doi:10.1111/j.1365-2486.
 466 2009.02122.x.

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

470 Smit, A. J., M. Roberts, R. J. Anderson, F. Dufois, S. F. J. Dudley, T. G. Bornman, J. Olbers, and
 471 J. J. Bolton, 2013: A coastal seawater temperature dataset for biogeographical studies: Large
 472 biases between in situ and remotely-sensed data sets around the coast of South Africa. *PLoS*
 473 *ONE*, **8** (12), doi:10.1371/journal.pone.0081944.

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

477 Tittensor, D. P., C. Mora, W. Jetz, H. K. Lotze, D. Ricard, E. V. Berghe, and B. Worm, 2010:
 478 Global patterns and predictors of marine biodiversity across taxa. *Nature*, **466** (7310), 1098–
 479 1101, doi:10.1038/nature09329.

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

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

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

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

487	LIST OF FIGURES	
488	Fig. 1. The location of the 129 time series available for use in this study	25
489	Fig. 2. Box and whisker plots of the base 84 time series	26
490	Fig. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to a	
491	time series	27
492	Fig. 4. The relationship between time series length and the standard error (SE) of the modelled trend	28
493	Fig. 5. The effect of the natural variation of a time series on the significance of the modelled trend	29
494	Fig. 6. The relationship between the effect of initial SD on the SE of a modelled trend	30
495	Fig. 7. The relationship between %NA and the significance of a fitted trend	31
496	Fig. 8. The minimal effect of rounding from	32
497	Fig. A1. Similar to Figure 7, this figure shows the effect missing data have on the significance of the	
498	slopes detected by GLS however; the missing values in the time series have been filled here	
499	via linear interpolation. The effect this has on the significance of the modeled trends is both	
500	immediate and dramatic. The behaviour of the quantity of interpolated data also differs from	
501	the effect of data left simply as NA. At lower levels of interpolation, missing data actually aid	
502	in the fitting of a more significant trend line. This phenomena reverses around 5%NA when	
503	the relationship becomes negative, meaning that as the amount of interpolated data increase,	
504	the significance of the fitted trend decreereases.	33

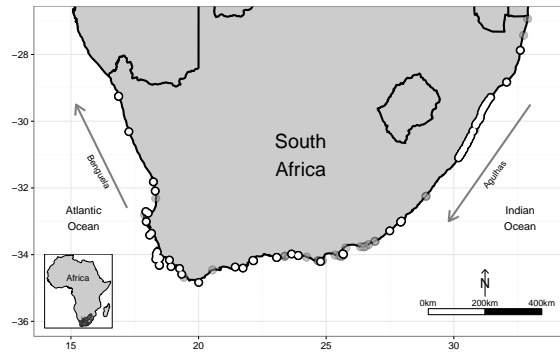
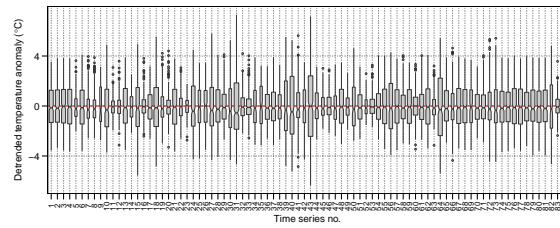


FIG. 1. The location of the 129 time series available for use in this study. The 84 time series used are shown as solid white circles and those not used are shown as opaque.



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

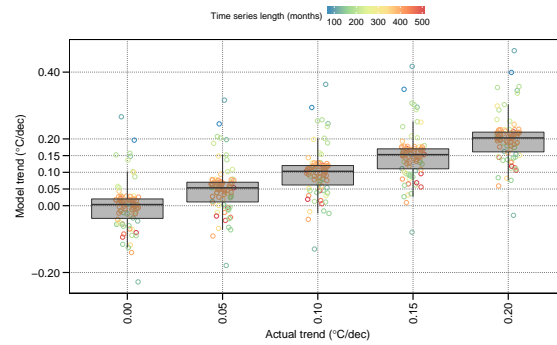
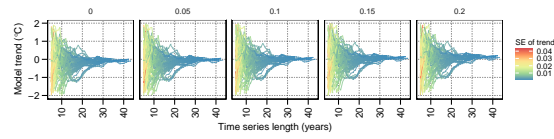


FIG. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to each time series. The box-and-whisker plots show the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, and the minima and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



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

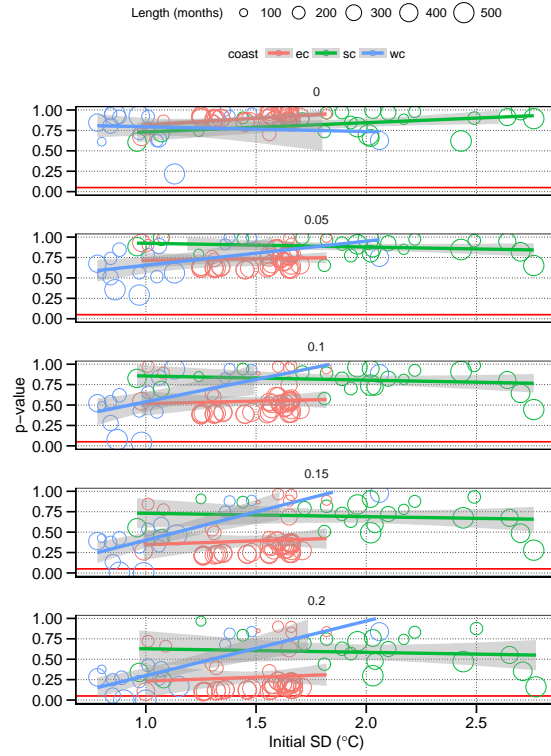


FIG. 5. The effect of the natural variation of a time series on the significance of the modelled trends estimated by the GLS. The size of the symbols are scaled proportionally to the time series length, with longer time series shown as larger circles.

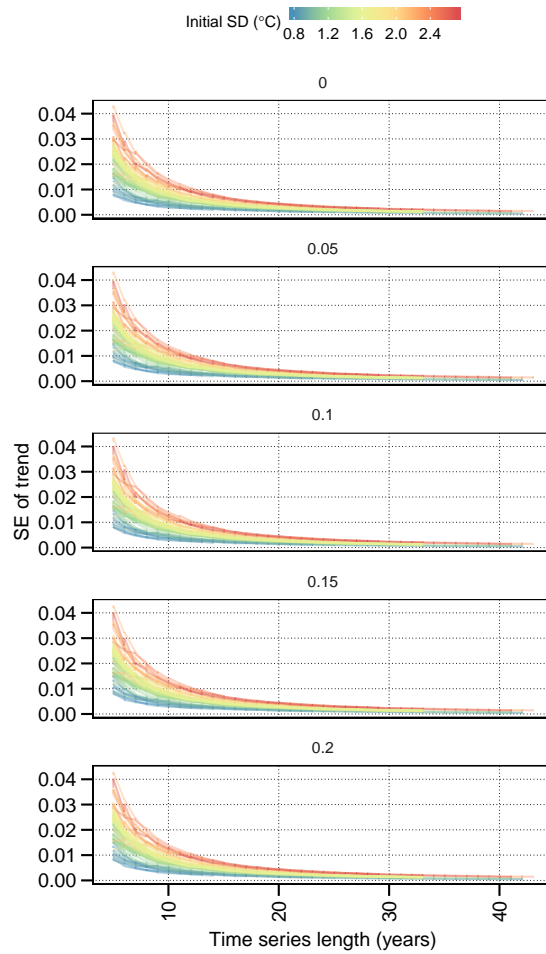


FIG. 6. The relationship between the effect of the initial SD of a time series on the SE of a modelled trend,
controlled for by the length of the time series.

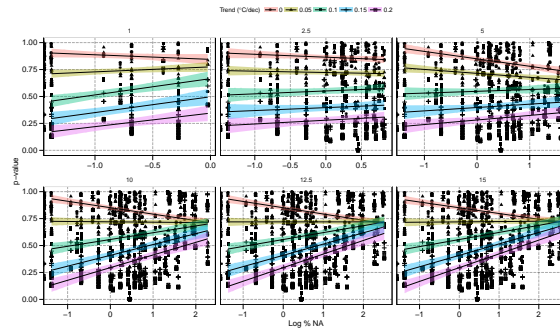


FIG. 7. The relationship between %NA and the significance of a fitted trend. Each panel shows the effect of an increasingly larger amount of missing values. The the fitted lines and 95% confidence intervals represent each of the five decadal trends assessed.

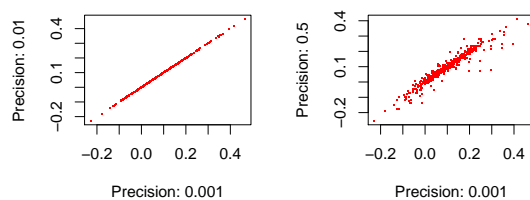
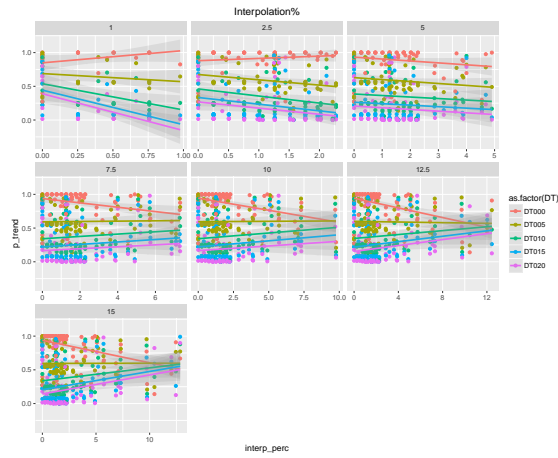


FIG. 8. The minimal effect of rounding from 0.001 °C to 0.01 °C may be seen in the panel on the right. The panel on the left shows that rounding from a precision of 0.001 °C to 0.5 °C has a more appreciable effect.



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