

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 up to 43 years are used
11 for investigations of the thermal characteristics of coastal seawater. They are
12 collected with handheld thermometers or underwater temperature recorders
13 (UTRs) and are recorded at precisions from 0.5 °C to 0.001 °C. Using the nat-
14 ural range of seasonal signals and variability for 84 of these time series, their
15 length, decadal trend and data precision were systematically varied before fit-
16 ting generalized least squares (GLS) models to study the effect these variables
17 have on trend detection. The variables that contributed most to accurate trend
18 detection in decreasing order were: time series length, decadal trend, variance,
19 percentage of missing data (%NA) and measurement precision. Time series
20 > 30 years in length are preferred and though larger decadal trends are mod-
21 eled more accurately, modeled significance (p -value) is largely affected by the
22 variance present. The risk of committing both type 1 and 2 errors increases
23 when $\geq 5\%$ NA is present. There is no appreciable effect on model accuracy
24 between measurement precision of 0.1 °C to 0.001 °C however, measurement
25 precisions of 0.5 °C require longer time series to give equally accurate model
26 results. The implication is that the thermometer time series in this dataset,
27 and others around the world, must be at least two years longer than their UTR
28 counterparts to be useful for decadal scale climate change studies. Further-
29 more, adding older lower precision UTR data to newer higher precision UTR
30 data within the same time series will increase their usefulness for this purpose.

31 1. Introduction

32 The roughly 3,000 km of South Africa's coastline is bordered by the Benguela and Agulhas
33 Currents (*e.g.* Roberts 2005; Hutchings et al. 2009), which, in combination with other nearshore
34 processes, affect the country's marine coastal ecosystems (Santos et al. 2012). A thorough under-
35 standing of these coastal processes is provided by several physical variables, of which temperature
36 is one of the main determinants (*e.g.* Blanchette et al. 2008; Tittensor et al. 2010; Couce et al.
37 2012). The statistical properties of *in situ* seawater temperature time series representing the whole
38 coastline – such as the annual mean, minimum and maximum temperature, and the thermal range
39 and variance characteristics – vary greatly among coastal sections due to the varying influence of
40 the Benguela and Agulhas Currents. Based on these thermal properties, the coastline has been
41 classified into a cool temperate west coast, a warm temperate south coast and tending towards
42 sub-tropical along the east coast (Smit et al. 2013; Mead et al. 2013). That the ocean temperature
43 of these regions is changing has been reported in recent years. For example, an increase of 0.55°C
44 to $0.7^{\circ}\text{C dec}^{-1}$ has been reported in the Agulhas Current (Rouault et al. 2009, 2010), while the
45 southern Benguela has decreased in $0.5^{\circ}\text{C dec}^{-1}$ during some parts of the year (Rouault et al.
46 2010).

47 The aforementioned climate change trends were derived from remotely-sensed gridded sea sur-
48 face temperature (SST) products. Whereas newer remotely-sensed gridded SST products are ap-
49 proaching high enough resolutions for use in coastal waters, older longer products that could be
50 used for the detection of long terms trends are not (*e.g.* Chao et al. 2009; Qiu et al. 2009; Vazquez-
51 Cuervo et al. 2013). A study by Smit et al. (2013) has also shown that remotely-sensed gridded
52 SST data have a warm bias as large as 6°C when compared to coastal *in situ* data. Nevertheless,
53 a widespread approach in coastal ecological research is to use satellite and/or model-generated

54 temperature data as a representation of SST along coastlines (*e.g.* Blanchette et al. 2008; Broitman
55 et al. 2008; Tyberghein et al. 2012), because apparently the dangers of applying gridded SSTs to
56 the coast are not widely known or in many places in the world there simply are no suitable *in situ*
57 coastal temperature time series available. It is for this reason that we strongly recommended the
58 use of *in situ* data to support research conducted within 400 m from the shoreline.

59 Where records of *in situ* coastal seawater temperature do exist, the reliability of many of these
60 datasets that could be used in place of the remotely-sensed SST data remains to be verified. Users
61 of remotely-sensed SST data benefit from it being refined through a number of well documented
62 validation and quality control processes (*e.g.* Reynolds and Smith 1994; Brown et al. 1999; Mar-
63 tin et al. 2012), whereas the standards and methods with which local *in situ* data from a single
64 dataset are collected and refined may differ greatly. For example, there are currently seven orga-
65 nizations and/or governmental departments (hereafter referred to as bodies) contributing coastal
66 seawater temperature data to the South African Coastal Temperature Network (SACTN). These
67 bodies use different methods and instruments to collect their data as no national standard has
68 been set. One consequence of this methodological disparity is that two thirds of the data were
69 sampled with hand-held thermometers that are manually recorded at a data precision of 0.5 °C,
70 as opposed to the current generation of underwater temperature recorders (UTRs) with an instru-
71 ment precision of down to 0.001 °C. If these *in situ* temperature data are to be used together *in*
72 *lieu* of remotely-sensed SST data, it is important that the characteristics of the contributing data
73 sources are understood in terms of their ability to yield useful, reliable and accurate long-term
74 measurements for use in climate change studies.

75 This prompted us to examine the 129 *in situ* time series that comprise the SACTN. The range
76 of measurement precisions and statistical characteristics of this dataset were used to guide a series
77 of enquiry-driven analyses into the suitability of the time series to yield statistically significant

78 and accurate assessments of decadal temperature change. The length, decadal trend and data
79 precision of each time series were adjusted in a systematic manner, and forms the core of our
80 analyses. Furthermore, the natural variability of each of the time series, which differ more-or-less
81 predictably between coastlines variously affected by the Benguela and Agulhas Currents, was also
82 entered into the analysis. Our aim was to assess the effect that each of these variables has on the
83 ability of a model to produce a robust estimate of time series decadal trend. The effect gaps in the
84 time series may have on the fitting of models was also investigated as many of the time series used
85 here have some missing data scattered throughout, which is unavoidable for a 20+ year time series
86 that is sampled by hand by a single technician at each site.

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

89 **2. Methods**

90 *a. Data Sources*

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

96 The data are collected using two different methods and a variety of instruments. Hand-held
97 mercury thermometers (which are being phased out in favor of alcohol thermometers or electronic
98 instruments) are used in some instances at the shoreline, and represent seawater temperatures at
99 the surface. At other places, predominantly along the country's east coast, data are collected with

100 glass thermometers from small boats at the location of shark nets along the coast (Cliff et al. 1988).
101 Whereas both types of thermometers allow for a measurement precision of 0.1 °C, the recordings
102 are written down at a precision of 0.5 °C. Data at other localities are collected using delayed-mode
103 instruments that are permanently moored shallower than 10 m, but generally very close to the
104 surface below the low-water spring tide level.

105 Over the last 40+ years the electronic instruments used to measure coastal seawater temperatures
106 have changed and improved. The previous standard was the Onset Hobo UTR with a thermal
107 precision of 0.01 °C. The new standard currently being phased in is the Starmon Mini UTR. These
108 devices have a maximum thermal precision of $0.001\text{ °C} \pm 0.025\text{ °C}$ (<http://www.star-oddi.com/>).
109 Of the 43 UTR time series in this dataset, 30 were recorded at a precision of 0.001 °C for their
110 entirety, five UTR time series include older data that were recorded at a precision of 0.01 °C or
111 0.1 °C and so have been rounded down to match this level of precision. Eight additional UTR time
112 series have data recorded at a precision of only 0.1 °C.

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

117 *b. Data Management*

118 Each of the seven bodies contributing data to this study have their own method of data for-
119 matting. Steps are being taken towards a national standard as we move towards replacing all the
120 thermometer recordings with UTR devices; however, as of the writing of this article, one does not
121 yet exist. Data from each organization were formatted to a project-wide comma-separated values
122 (CSV) format with consistent column headers before any statistical analyses were performed. This

123 allowed for the same methodology to be used across the entire dataset, ensuring consistent analy-
124 sis. Before analysing the data they were scanned for any values above 35 °C or below 0 °C. These
125 data points were changed to NA, meaning ‘not available’, before including them in the SACTN
126 dataset.

127 All analyses and data management performed in this paper were conducted with R version 3.3.1
128 (2016-06-21) (R Core Team 2013). The script and data used to conduct the analyses and create
129 the figures seen in this paper may be found at https://github.com/schrob040/Trend_Analysis.

130 Any time series with a temporal precision greater than one day were averaged into daily values
131 before being aggregated into the SACTN. A series of additional checks were then performed (*e.g.*
132 removing long stretches in the time series without associated temperature recordings) and time
133 series shorter than five calendar years, collected at depths greater than 10 m or missing more than
134 15% of their monthly values were removed. At the time of this analysis, this usable daily dataset
135 consisted of 84 time series, consisting of 819,499 days of data; these data were then binned further
136 to the 26,924 monthly temperature values available for use in this study.

137 *c. Systematic Analysis of Time Series*

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

145 anomaly time series (henceforth simply called ‘time series’) represent a range of time scales from
146 72 to 519 months in duration.

147 To each of the 84 time series we artificially added linear decadal trends of 0.00°C to
148 $0.20^{\circ}\text{C dec}^{-1}$. In other words, we now had time series that captured the natural thermal vari-
149 abilities around the coast, but with their decadal trends known *a priori*. The range of decadal
150 trends was selected based around the global average of 0.124°C from Kennedy et al. (2011) and
151 used in Stocker et al. (2013). Furthermore, in order to represent the instrumental precision of the
152 instruments used to collect these time series, we rounded each of these (84 time series \times 5 decadal
153 trends) to four levels of precision: 0.5°C , 0.1°C , 0.01°C and 0.001°C . Consequently, we had a
154 pool of 1,680 time series with which to work.

155 To gain further insight into the effect of time series length on trend detection, each time series
156 was first shortened to a minimum length of 5 years, starting in January so that the timing of the
157 seasonal signal for each time series would be equitable. After fitting the model (see *Time Series*
158 *Model*, below) to all 1,680 of the shortened time series, the next year of data for each time series
159 was added and the models fitted again. This process was iterated until the full length of each time
160 series was attained. For example, if a time series consisted of 12 full years of data, it would require
161 160 models (8 iterations of increasing length \times 5 decadal trends \times 4 levels of precision); similarly,
162 720 models would be applied to a 40 year time series. Considering the 84 time series available,
163 the total number of individual models required to capture each combination of variables quickly
164 increased to 36,220.

165 Our approach of fitting models to each of the semi-artificial time series that we generated allowed
166 us to study the effect that the relevant variables (time series length, natural variability, added
167 slope and level of measurement precision) has on the ability of the time series model to faithfully
168 detect the decadal thermal trend, which was known *a priori*. The primary results of interest in

169 these analyses were the significance (p -value) of the model fit, the accuracy of the decadal trend
170 determined by the GLS model as well as the error associated with the trend estimate.

171 *d. Time Series Model*

172 The selection of the appropriate model can greatly influence the ability to detect trends (Franzke
173 2012). Two broad approaches are widely used in climate change research (Stocker et al. 2013).
174 The first group of models estimates linear trends, and although linearity may not reflect reality (*i.e.*
175 trends are very frequently non-linear), these models do provide the convenience of producing an
176 easy to understand decadal trend (*e.g.* $0.106^{\circ}\text{C dec}^{-1}$; Wilks 2011; Stocker et al. 2013). The other
177 group accommodates non-linear trajectories of temperature through time by the use of higher-
178 degree polynomial terms or non-parametric smoothing splines, but the inconvenience comes from
179 not being able to easily compare models among sites (Wood 2006; Scinocca et al. 2010). Both
180 groups of models can accommodate serially correlated residuals, which is often the cause for much
181 criticism due to their effect on the uncertainty of the trend estimates (Von Storch 1999; Santer et al.
182 2008). For example, Generalized Least Squares (GLS; yielding estimates of linear trends) and
183 Generalized Additive Mixed Models (GAMM; non-linear fitting with no trend estimate provided)
184 can both capture various degrees of serial autocorrelation (Pinheiro and Bates 2006; Wood 2006).
185 Although our exploratory analysis assessed two parametrizations of each of the model groups, we
186 opted to proceed here with a GLS equipped with a second-order autoregressive AR(2) correlation
187 structure fitted using Restricted Maximum Likelihood (REML; Pinheiro and Bates 2006):

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

188 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 modelling 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 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 $0\text{ }^{\circ}\text{C dec}^{-1}$ 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 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 $0\text{ }^{\circ}\text{C dec}^{-1}$ 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 $0\text{ }^{\circ}\text{C dec}^{-1}$.

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

The effect of decreasing data measurement precision from $0.001\text{ }^{\circ}\text{C}$ to $0.5\text{ }^{\circ}\text{C}$ 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 $0.1\text{ }^{\circ}\text{C}$ to $0.5\text{ }^{\circ}\text{C}$. This effect is larger on smaller decadal trends. For example, at a trend of $0.05\text{ }^{\circ}\text{C dec}^{-1}$, the deviation from the true value of models fitted to data with a precision of $0.1\text{ }^{\circ}\text{C}$ is negligible; however, the accuracy of the fitted model on data recorded at a precision of $0.5\text{ }^{\circ}\text{C}$ with a real trend of $0.05\text{ }^{\circ}\text{C dec}^{-1}$ is 10.81% different on average (*i.e.* given a slope of $0.05\text{ }^{\circ}\text{C dec}^{-1}$ the model detects slopes of $0.055\text{ }^{\circ}\text{C dec}^{-1}$). This accuracy of the models improves to an average difference of 6.44% with a slope of $0.10\text{ }^{\circ}\text{C dec}^{-1}$, 2.24% at $0.15\text{ }^{\circ}\text{C dec}^{-1}$ and decreases slightly to 2.30% at $0.20\text{ }^{\circ}\text{C dec}^{-1}$. A precision of $0.5\text{ }^{\circ}\text{C}$ 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 $0.05^{\circ}\text{C dec}^{-1}$, 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 $0.20^{\circ}\text{C dec}^{-1}$ 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 $0\text{ }^{\circ}\text{C dec}^{-1}$ 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 time series with steeper decadal trends. 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; Stocker et al. 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

303 must ensure that these sources of data are managed and curated with great care and diligence as
304 they are practically irreplaceable. For this reason, it is essential that we understand the inherent
305 strengths and weaknesses of such existing sources of data so that we may fully maximize their
306 utility and extract from them the model coefficients needed to detect decadal temperature trends,
307 and know the accuracy of these estimates to the best of our ability. There are many time series
308 < 20 years in length that should be avoided, where possible, for trend analysis. These will mature
309 with time and their maintenance need to be ensured going forward.

310 Aside from length, the most powerful time series have measurements that are taken regularly.
311 The inclusion of too many missing values (NAs) in the data sets must be avoided. We have shown
312 that permitting 5% NAs or more into our time series has a drastic and significant influence on the
313 chance of committing a type 1 error (arriving at ‘false positive,’ *i.e.* detecting a trend when none
314 exists) for time series with no or very gentle decadal trends. On the other hand, the inclusion of NAs
315 in data sets with a decadal trend present tends to cause an increase in the probability of committing
316 a type 2 error (*i.e.* finding ‘false negatives’). Although our modern UTR data sets generally have
317 fewer NAs than we should be concerned about – therefore with a low chance of committing type 1
318 or type 2 errors – the presence of NAs may seriously compromise some of the time series that are
319 still being collected by hand using hand-held thermometers.

320 We have demonstrated clearly that as the steepness of an expected decadal trend increases, the
321 ability for it to be modeled accurately increases, too. Our GLS model is generally not able to detect
322 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. Very
323 rarely were we able to produce significant model fits at shallower slopes. Finding significant trends
324 at $< 0.05\text{ }^{\circ}\text{C dec}^{-1}$ was not possible. Based on the relationship between SD and the added decadal
325 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 decadal
326 trend of $0.10\text{ }^{\circ}\text{C dec}^{-1}$ should consist of roughly 640 months of data before our GLS model would

327 regularly be able to detect a significant trend ($p < 0.05$). This finding is somewhat discouraging
328 as most global analyses of decadal SST change based on gridded SST products estimate a trend
329 closer to $0.1\text{ }^{\circ}\text{C dec}^{-1}$ (*e.g.* Stocker et al. 2013). This means that the trends present in most time
330 series representative of very variable coastal environments that exhibit the same variance structure
331 as that of our data are probably unlikely to be detected as significant, even if they do indeed exist.
332 In other words, the chance of committing a type 2 error is probably very real for such systems,
333 unless time series > 50 years are available.

334 As 50 year coastal seawater temperature time series are probably very scarce, it is important to
335 note that those measured at precisions of $0.1\text{ }^{\circ}\text{C}$ to $0.001\text{ }^{\circ}\text{C}$ require fewer months of data to detect
336 long term trends. Based on the data presented here, we calculated that time series measured at a
337 low precision ($0.5\text{ }^{\circ}\text{C}$) may require as much as an additional 24 months of data to accurately detect
338 long-term trends. One of the motivators for this paper was to investigate the effect measurement
339 precision has on a time series' ability to produce results useful for investigations of long-term
340 climate change, and to validate the use of the low precision $0.5\text{ }^{\circ}\text{C}$ thermometer data. This is an
341 important consideration as many studies investigating the effects of climate change (*e.g.* Grant
342 et al. 2010; Scherrer and Körner 2010; Lathlean and Minchinton 2012) do use lower precision
343 $0.1\text{ }^{\circ}\text{C}$ data. Whereas the precision of much of our data is below the current standard of $0.1\text{ }^{\circ}\text{C}$
344 required for climate change research (Ohring et al. 2005; Jarraud 2008), the length of the ther-
345 mometer time series makes them a valuable asset. The average length of the thermometer time
346 series in the SACTN, from which the 84 time series used in this study were drawn, is 349 months.
347 The average length of the UTR time series is 167 months. Given this difference in the lengths of
348 the time series, even after correcting for the negative effect of low measurement precision, the time
349 series collected with thermometers are currently more useful for climate change research than the
350 UTR time series within the SACTN. Because time series with data precisions of $0.1\text{ }^{\circ}\text{C}$ to $0.001\text{ }^{\circ}\text{C}$

351 produce comparable results, newer higher precision UTR data may be combined with older lower
352 precision UTR data within the same time series without concern that the reduced overall data
353 precision may have a negative impact on a model's ability to detect decadal trends. Extending
354 time series in this way will serve to make them more dependable as length is the primary criterion
355 through which one should initially assess a time series ability to suggest the presence of climate
356 change before refining ones assumptions with any statistical analyses. A time series with data
357 precision greater than 0.1°C is therefore only necessary when an investigation requires that the
358 decadal trend be known to an accuracy of 0.01°C or greater (*e.g.* Karl et al. 2015).

359 It is important to take note of the accuracy of the models, not only to focus on the significance of
360 their results. Indeed, the p -value given for the slope in a model does not show how well the model
361 detects the true trend in the data (known *a-priori* in this study); rather, it tells us if the detected trend
362 is significantly different from $0^{\circ}\text{C dec}^{-1}$. This is not particularly useful for applying the results
363 of climate change research more broadly to biotic interests. For example, of the 1344 models (84
364 base time series \times 4 decadal trends \times 4 levels of precision) fitted to time series with decadal trends
365 $\geq 0.05^{\circ}\text{C dec}^{-1}$, 317 of these were accurate to within 10% of the decadal trend known *a priori*,
366 but not significant ($p \geq 0.05$). That a long term trend does exist, may be accurately detected by a
367 model and related to an observed change in the natural world – such as range expansion/contraction
368 of coastal biota (Bolton et al. 2012; Straub et al. 2016; Wernberg et al. 2016) – is more important
369 than whether or not the model can show if that trend is significantly different from $0^{\circ}\text{C dec}^{-1}$ in a
370 statistical sense.

371 We must mention also that much of the meta-data pertaining to the older temperature records
372 used here have over time been lost. As with the bulk of the International Comprehensive Ocean-
373 Atmosphere Data Set (ICOADS; Freeman et al. 2016), *in situ* coastal seawater temperature mon-
374 itoring that started in the 1970s in South Africa was not developed with climate change research

375 in mind, and comprehensive records that keep track of details of the instruments used, calibration,
376 their turnover, change in monitoring methods and locations and so forth are not always available
377 as per modern requirements (Aguilar et al. 2003). For studies of climate change *per se* this is a
378 serious limitation and it prevents us from knowing anything about the accuracy of the instruments
379 or potential issues of drift (stability) that may have occurred. We do know however that all time
380 series sampled with thermometers were sampled only with thermometers, and *vice versa* for the
381 UTR time series, ensuring that the precisions of the measured data used in this study are correct.
382 Moving forward with the further development of the SACTN and the establishment of a national
383 standard of data collection and instrument maintenance, we are able to record and archive all these
384 levels of pertinent meta-data, and allowing for the enforcement of SI traceability and the accurate
385 measurement of instrument drift (Jarraud 2008). Nevertheless, the detrended anomaly time series
386 used here were taken only for their variance properties, which we think accurately reflect that of
387 the various coastal sections around the coast. They provide a strong backbone for semi-artificial
388 time series, and we have shown how important insights about model fitting could be derived from
389 these data.

390 **5. Conclusion**

391 We draw several key conclusions:

- 392 1. There is a rapid increase in the accuracy and significance of modeled trends as time series
393 lengths extend from 10 to 20 years. This improvement slows from 20 to 30 years, and as time
394 series approach 40 years in length the accuracy of models becomes nearly exact. Modelled
395 trends from time series at or under 25 years in length should be interpreted with extreme
396 caution.

2. For our variable coastal seawater, a time series of 520 months in length is required to detect a decadal trend in line with the global average (*i.e.* near $0.1\text{ }^{\circ}\text{C dec}^{-1}$) 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 $0.1\text{ }^{\circ}\text{C dec}^{-1}$ 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 chance of committing type 1 (with gentle trends) or type 2 errors (with steeper trends).
6. A measurement precision greater than $0.5\text{ }^{\circ}\text{C}$ is not required to confidently detect the long-term trend in a time series; however, precisions at or greater than $0.1\text{ }^{\circ}\text{C}$ may reduce the length of time required to accurately detect a long term trend, if one does exist, by as much as two years.
7. Improving the precision of measurements to greater than $0.1\text{ }^{\circ}\text{C}$ 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 irreplaceable.

419 *Acknowledgments.* The authors would like to thank DAFF, DEA, EKZNW, KZNSB, SAWS and
420 SAEON for their contributions of the raw data used in this study. Without it, this article and the
421 SACTN would not be possible. This research was supported by NRF Grant (CPRR14072378735).
422 The authors report no financial conflicts of interests.

423 **References**

- 424 Aguilar, E., I. Auer, M. Brunet, T. Peterson, and J. Wieringa, 2003: Guidelines on climate metadata
425 and homogenization, wmo-td no. 1186 (wcdmp2). *World Meteorological Organization, Geneva*.
- 426 Blanchette, C. A., C. Melissa Miner, P. T. Raimondi, D. Lohse, K. E. K. Heady, and B. R. Broit-
427 man, 2008: Biogeographical patterns of rocky intertidal communities along the Pacific coast of
428 North America. *Journal of Biogeography*, **35** (9), 1593–1607.
- 429 Bolton, J. J., R. J. Anderson, A. J. Smit, and M. D. Rothman, 2012: South African kelp moving
430 eastwards: the discovery of *Ecklonia maxima* (Osbeck) Papenfuss at De Hoop Nature Reserve
431 on the south coast of South Africa. *African Journal of Marine Science*, **34** (1), 147–151.
- 432 Broitman, B. R., N. Mieszkowska, B. Helmuth, and C. A. Blanchette, 2008: Climate and re-
433 cruitment of rocky shore intertidal invertebrates in the eastern North Atlantic. *Ecology*, **89** (11
434 **Suppl**), S81–90.
- 435 Brown, O. B., P. J. Minnett, R. Evans, E. Kearns, K. Kilpatrick, A. Kumar, R. Sikorski, and
436 A. Závody, 1999: MODIS Infrared Sea Surface Temperature Algorithm Algorithm Theoretical
437 Basis Document Version 2.0. *University of Miami*, 31 098–33 149.
- 438 Chao, Y., Z. Li, J. D. Farrara, and P. Hung, 2009: Blending sea surface temperatures from multi-
439 ple satellites and in situ observations for coastal oceans. *Journal of Atmospheric and Oceanic*
440 *Technology*, **26** (7), 1415–1426, doi:10.1175/2009JTECHO592.1.

- 441 Cliff, G., S. F. J. Dudley, and B. Davis, 1988: Sharks caught in the protective gill nets off Natal,
442 South Africa. 1. The sandbar shark *Carcharhinus plumbeus* (Nardo). *South African Journal of*
443 *Marine Science*, **7** (1), 255–265.
- 444 Couce, E., A. Ridgwell, and E. J. Hendy, 2012: Environmental controls on the global distribution
445 of shallow-water coral reefs. *Journal of Biogeography*, **39** (8), 1508–1523.
- 446 Franzke, C., 2012: Nonlinear trends, long-range dependence, and climate noise properties of sur-
447 face temperature. *Journal of Climate*, **25** (12), 4172–4183, doi:10.1175/JCLI-D-11-00293.1.
- 448 Freeman, E., and Coauthors, 2016: ICOADS release 3.0: a major update to the historical marine
449 climate record. *International Journal of Climatology*, 1–22.
- 450 Grant, O. M., L. Tronina, J. C. Ramalho, C. Kurz Besson, R. Lobo-Do-Vale, J. Santos Pereira,
451 H. G. Jones, and M. M. Chaves, 2010: The impact of drought on leaf physiology of *Quercus*
452 *suber* L. trees: Comparison of an extreme drought event with chronic rainfall reduction. *Journal*
453 *of Experimental Botany*, **61** (15), 4361–4371, doi:10.1093/jxb/erq239.
- 454 Hartmann, D., and Coauthors, 2013: Observations: Atmosphere and surface supplementary mate-
455 rial. in: *Climate change 2013: The physical science basis*.
- 456 Hutchings, L., and Coauthors, 2009: The Benguela Current: An ecosystem of four components.
457 *Progress in Oceanography*, **83** (1-4), 15–32, doi:10.1016/j.pocean.2009.07.046.
- 458 Jarraud, M., 2008: Guide to meteorological instruments and methods of observation (wmo-no. 8).
459 *World Meteorological Organisation, Geneva*.
- 460 Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface
461 warming hiatus. *Science*, **348** (6242), 1469–1472, doi:10.1126/science.aaa5632.

- 462 Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker, 2011: Reassessing biases
463 and other uncertainties in sea-surface temperature observations measured *in situ* since 1850, Part
464 1: measurement and sampling uncertainties. *Journal of Geophysical Research Atmospheres*,
465 **116**.
- 466 Lathlean, J. A., and T. E. Minchinton, 2012: Manipulating thermal stress on rocky shores to pre-
467 dict patterns of recruitment of marine invertebrates under a changing climate. *Marine Ecology*
468 *Progress Series*, **467**, 121–136, doi:10.3354/meps09996.
- 469 Martin, M., and Coauthors, 2012: Group for High Resolution Sea Surface temperature (GHRSSST)
470 analysis fields inter-comparisons. Part 1: A GHRSSST multi-product ensemble (GMPE). *Deep*
471 *Sea Research Part II: Topical Studies in Oceanography*, **77-80**, 21–30.
- 472 Mead, A., and Coauthors, 2013: Human-mediated drivers of change — impacts on coastal ecosys-
473 tems and marine biota of South Africa. *African Journal of Marine Science*, **35 (3)**, 403–425.
- 474 Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, 2005: Satellite instrument cali-
475 bration for measuring global climate change: report of a workshop. *Bulletin of the American*
476 *Meteorological Society*, **86 (9)**, 1303–1313.
- 477 Pinheiro, J., and D. Bates, 2006: *Mixed-effects models in S and S-PLUS*. Springer Science &
478 Business Media.
- 479 Qiu, C., D. Wang, H. Kawamura, L. Guan, and H. Qin, 2009: Validation of AVHRR and TMI-
480 derived sea surface temperature in the northern South China Sea. *Continental Shelf Research*,
481 **29 (20)**, 2358–2366.
- 482 R Core Team, 2013: *R: A Language and Environment for Statistical Computing*. Vienna, Austria,
483 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.

Vazquez-Cuervo, J., B. Dewitte, T. M. Chin, E. M. Armstrong, S. Purca, and E. Alburquerque, 2013: An analysis of SST gradients off the Peruvian Coast: The impact of going to higher resolution. *Remote Sensing of Environment*, **131**, 76–84, doi:10.1016/j.rse.2012.12.010.

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.

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

529	LIST OF FIGURES	
530	Fig. 1. Of the 129 time series available for use in this study	27
531	Fig. 2. Box and whisker plots summarizing the anomaly time series of the base 84 data sets	28
532	Fig. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend	29
533	Fig. 4. The relationship between the length of a time series, the size of the modeled trend and its standard error (SE) . . .	30
534	Fig. 5. The effect of the natural variation found within the detrended residuals on the significance	31
535	Fig. 6. The relationship between the effect of Initial SD (Figure 5) on the SE of a modeled trend	32
536	Fig. 7. The relationship between the percentage of missing values (%NA) and the significance of a fitted trend	33
537	Fig. 8. The minimal effect of rounding from	34

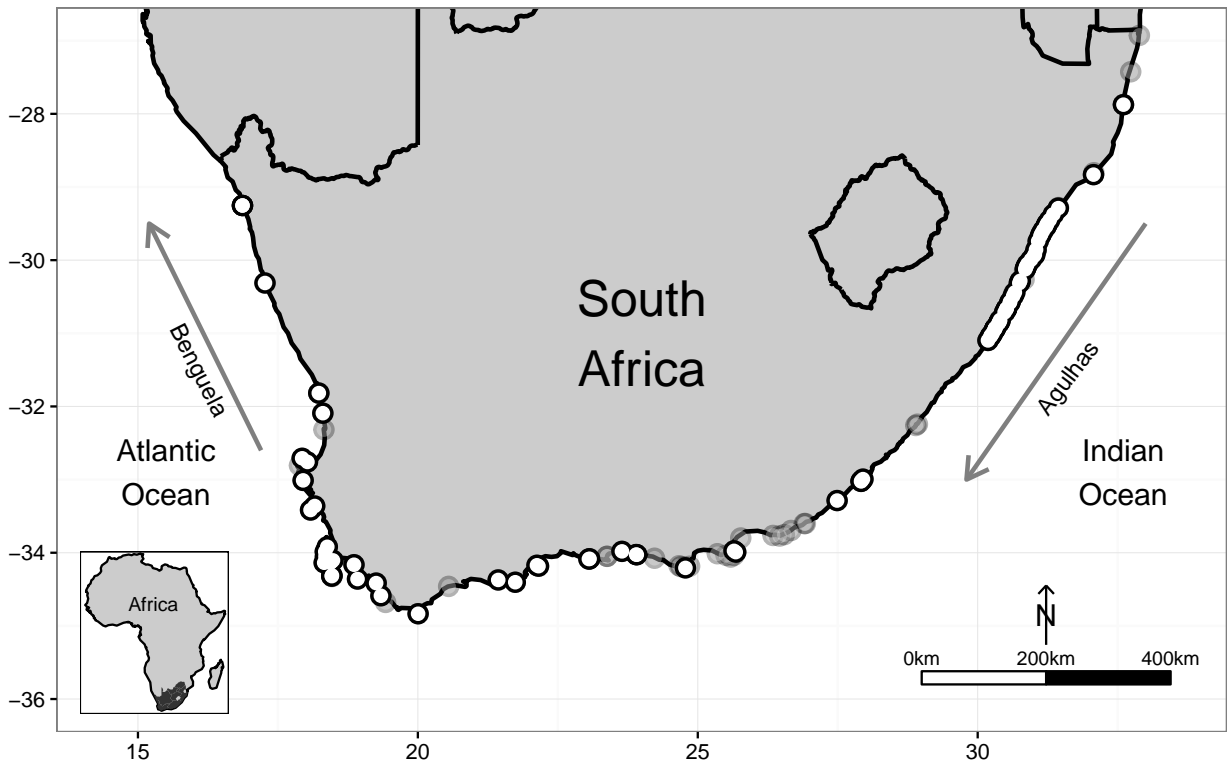


FIG. 1. Of the 129 time series available for use in this study, the location of the 84 time series used are shown as solid white circles and the unused shown as opaque.

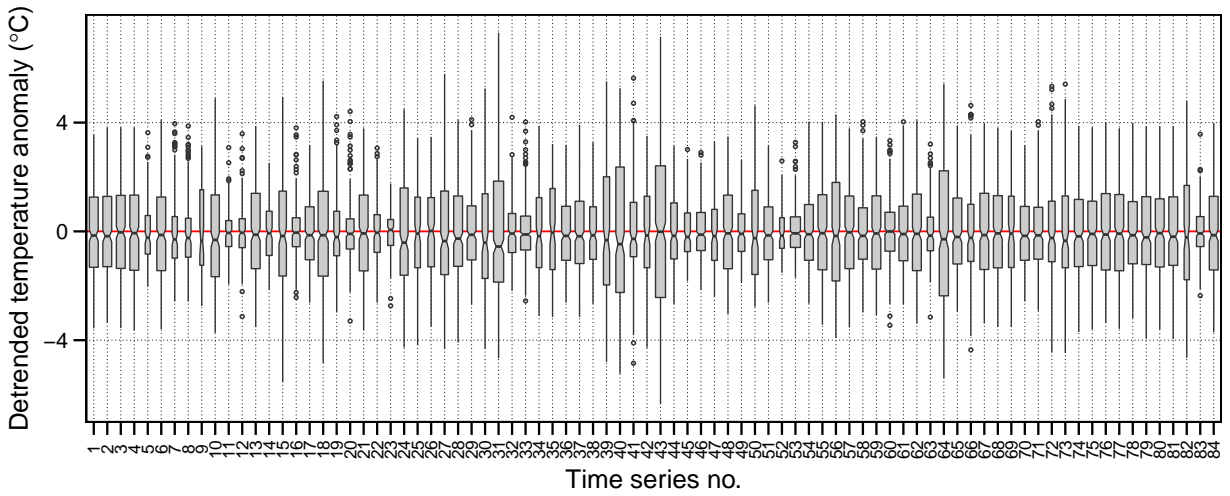


FIG. 2. Box and whisker plot summarizing the 84 base anomaly time series used in this study after detrending
(i.e. the residuals after removing the linear trend using an ordinary least squares regression) but before adding
a decadal trend or rounding the data. The plot indicates the first and third quartile 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.

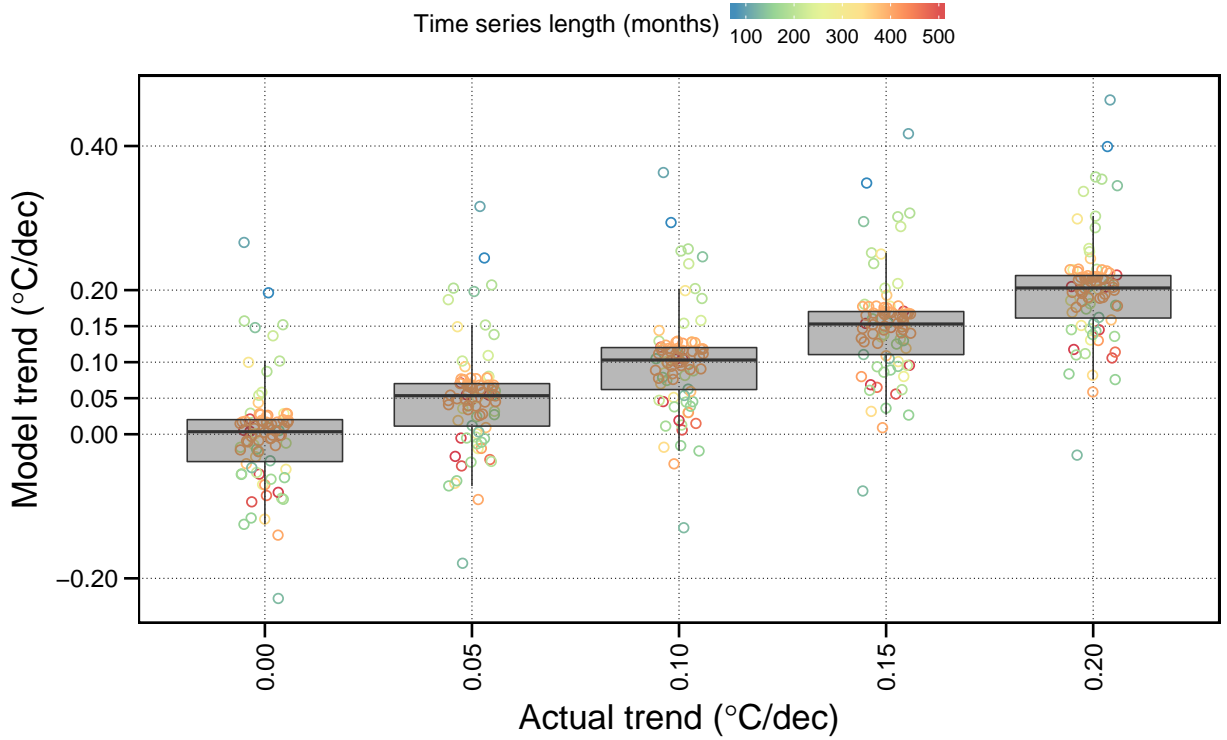


FIG. 3. The effect of time series length on the ability of the GLS model to accurately detect the trend added to each time series. The box and whisker plots show the first and third quartile as the extremities of the boxes, the median is shown as the horizontal line within each box, and the minima and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

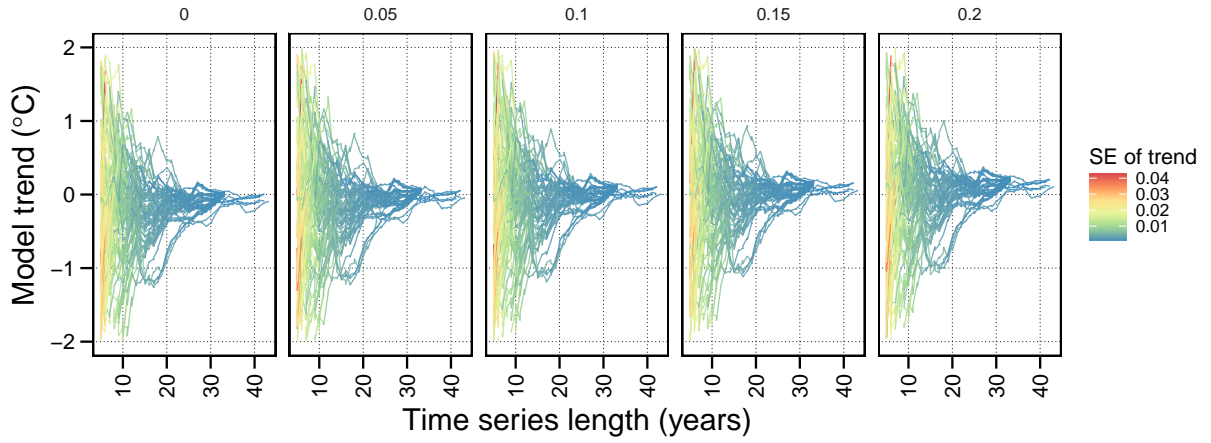


FIG. 4. The relationship between the length of a time series, the size of the modeled trend and its standard error (SE). Each individual line shows the modeled trend for one of the 84 sites used in this analysis to which a model was fitted iteratively as the time series length was ‘grown’ from 5 years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

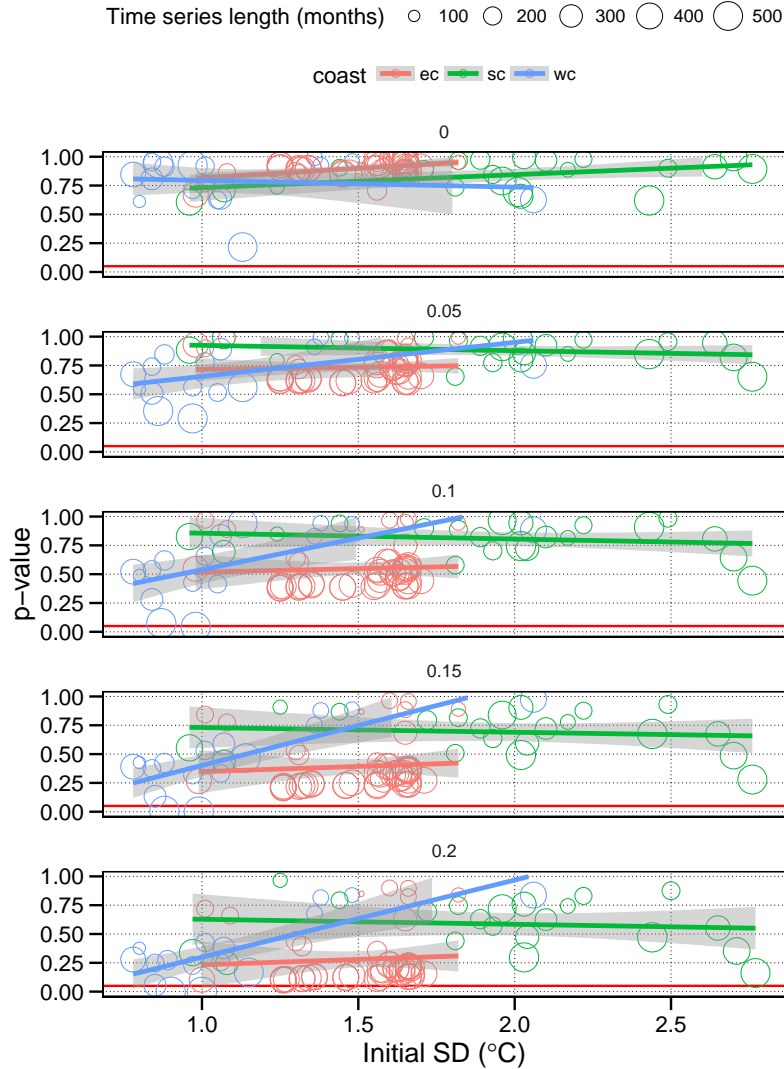


FIG. 5. The effect of the natural variation found within the detrended residuals before adding a decadal trend or rounding the data (Figure 2), shown here as the initial standard deviation (Initial SD), on the significance of the modeled trend. The size of the symbols denoting each time series are scaled proportionally to the time series length in months, with longer time series shown as larger circles. The time series from the different coastal sections of South Africa are represented in color. The east coast (ec) typically has the most stable thermal regime of the three coasts, with the south coast (sc) having the greatest amount of variance and the west coast (wc) consisting of areas with both high and low variance. Linear models with 95% confidence intervals shown here as gray ribbons have been fitted to each coastal sections to illustrate the interaction between the range of Initial SD values found in each group and the significance (p -value) of the models fitted to these time series. Each panel shows how these relationships change as the slope of the decadal trend increases.

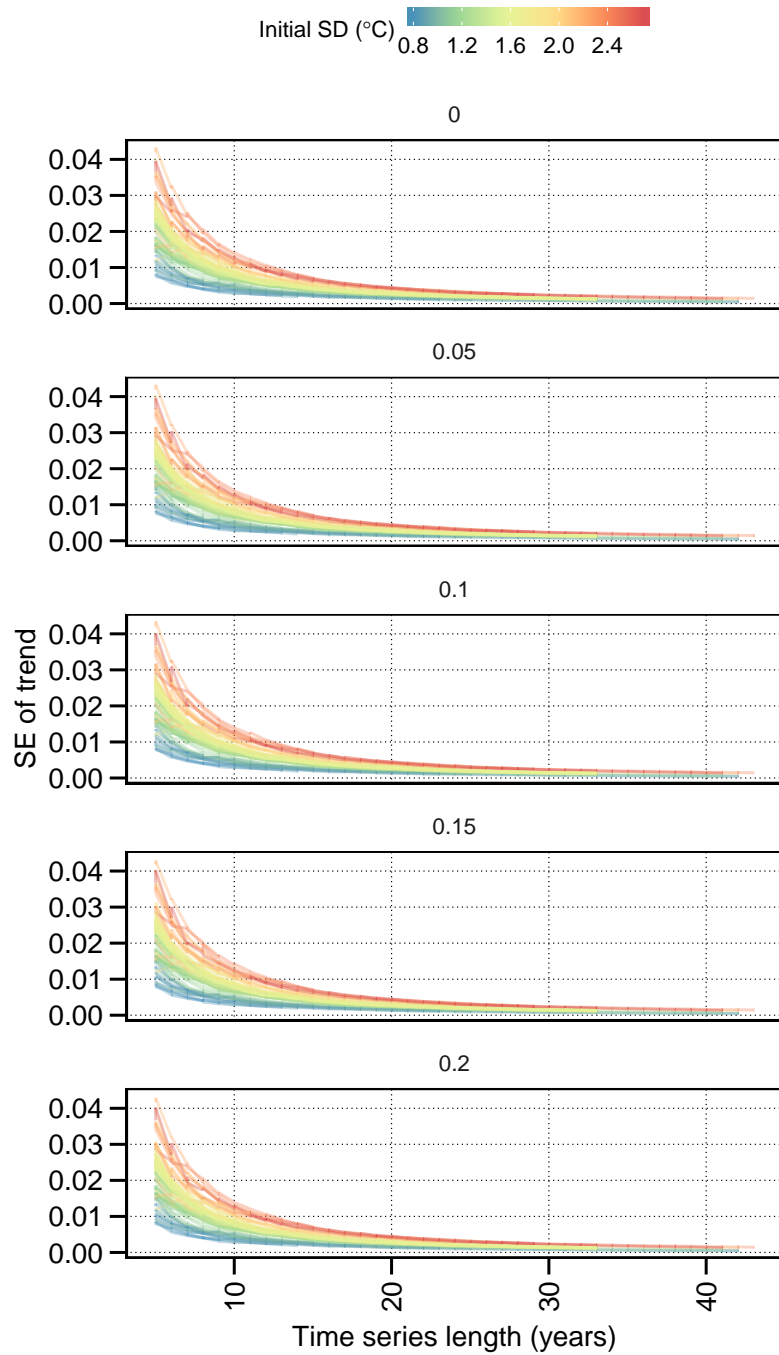


FIG. 6. The relationship between the effect of Initial SD (Figure 5; shown here in color), on the standard error (SE) of a modeled trend, controlled for by the length of the time series in years. The panels demonstrate the imperceivable effect increases in the slope of the decadal trend have on this relationship.

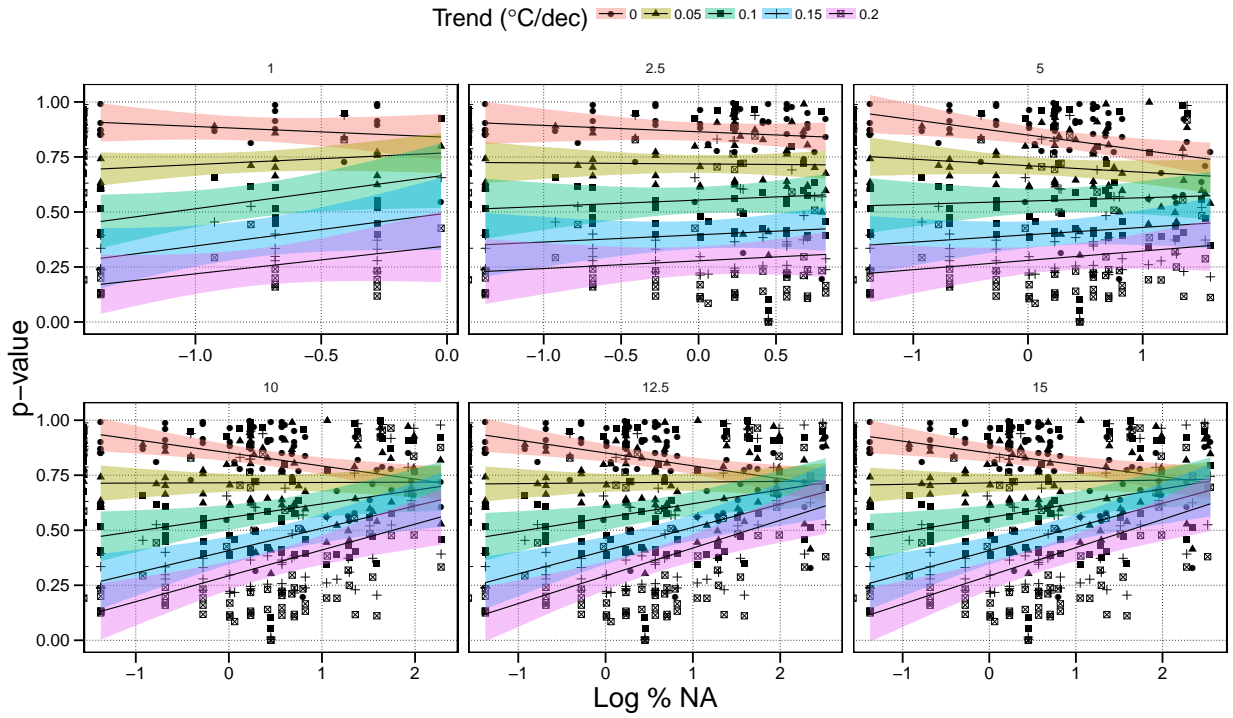
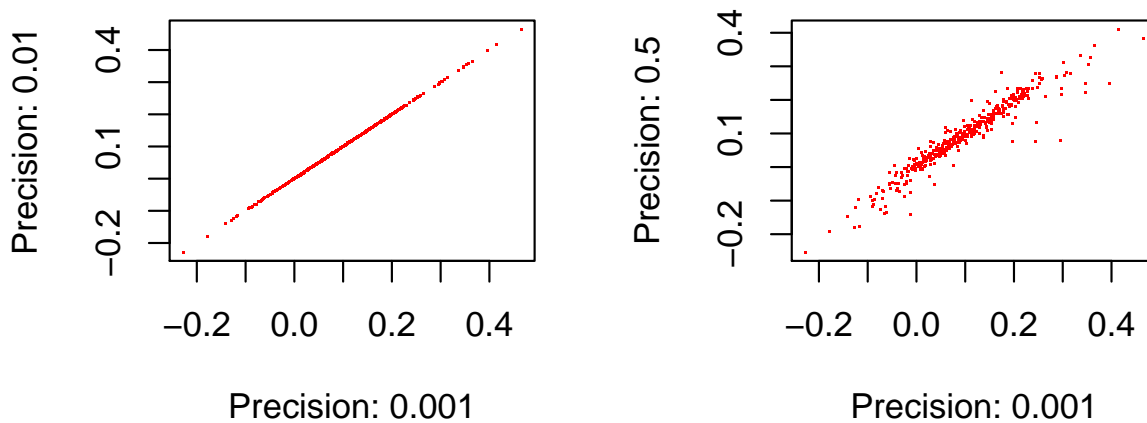


FIG. 7. The relationship between the percentage of missing values (%NA) and the significance of a modeled trend. Each panel shows the effect of an increasingly larger amount of missing values. The fitted lines and 95% confidence intervals represent each of the five decadal trends assessed.



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